2., ÜBERARBEITETE UND AKTUALISIERTE AUFLAGE

DATA DRIVEN CONSTRUCTION

NAVIGIEREN IM DATENZEITALTER IN DER BAUINDUSTRIE

+ ANWENDUNGSBEISPIELE KI UND LLM

Artem Boiko

MORE LANGUAGES ON THE SITE
DATADRIVENCONSTRUCTION.IO

DATA-DRIVEN CONSTRUCTION

Navigieren im Datenzeitalter in der Bauindustrie

Zweite, überarbeitete und ergänzte Auflage

ARTEM BOIKO

"Boiko ist der James Carville der IT – in des letzteren viel zitiertem: „It's the economy, stupid" muss für dieses famose Buch nur ein Wort getauscht werden. „It's the data, stupid." (nicht die Software) Und sich im Daten-Universum zurechtzufinden, ist eine schon aufs Griechische zurückgehende Weisheit der alten Römer bis heute gültig: „Navigare necesse est". Der Autor navigiert seine Leser mit sicherer Hand und unbeirrbarem Kompass durch alle Tiefen und Untiefen des Daten-Ozeans, obendrein mit umfassendem historischen Anspruch und nicht zuletzt mit höchst originellen Grafiken und nicht erst auf den zweiten Blick erkennbaren, gutem Humor. An Boikos Buch scheiden sich vielleicht die Geister – von euphorischer Zustimmung bis zu eher galligem Skeptizismus reicht das internationale Echo, das der zweiten deutschen Auflage des Buches gutgetan hat. Boiko ist ein origineller und undogmatischer Datendenker. Er beschenkt den Leser mit spannenden Einsichten und immer mutigen, auch provokanten Thesen die zum Weiterdenken inspirieren. Vorzügliche Medizin für die bundesdeutsche Krankheit des latenten Konsensualismus'. Obige lateinische Weisheit übrigens kennt eine Ergänzung: „vivere non est necesse." Sie stimmt für Boikos Betrachtungsweise der Datenwelt nicht – Daten leben und ihr Leben ist notwendig, um nicht zu sagen: entscheidend."

- Dr. Burkhard Talebitari, freier Redakteur – u. a. für das seit 2013 bei Ernst & Sohn jährlich erscheinende Journal: BIM.

"Artem Boikos Buch ist ein Meilenstein für die Demokratisierung der Digitalisierung im Bauwesen – und ein echter Gamechanger für kleine und mittelständische Unternehmen (KMU). Mit beeindruckender Klarheit zeigt es, wie komplexe Building Information Modeling (BIM)-Daten mithilfe schlanker Konverter in einfach nutzbare, flache Formate überführt werden können. Damit wird der Zugang zu wertvollen Projektdaten endlich für jene geöffnet, die bislang durch technische oder finanzielle Hürden ausgeschlossen waren. Besonders bahnbrechend: Durch den Einsatz moderner Open-Source-Low-Code- und No-Code-Tools können Unternehmen bereits heute – ganz ohne tiefgreifende Programmierkenntnisse – BIM-Daten effizient in ihre Geschäftsprozesse integrieren und gewinnbringend analysieren. Der teure Einsatz schwerfälliger kommerzieller Softwarepakete wird damit überflüssig. Dieses Buch ist ein Aufruf zum Handeln! Es ist ein wertvoller Wegweiser für alle, die die digitale Transformation im Bauwesen nicht nur verstehen, sondern aktiv gestalten wollen – pragmatisch, effizient und zukunftsweisend. Jetzt ist die Zeit reif, dieses Wissen gemeinsam in die Breite zu tragen und die Produktivität der Bauwirtschaft nachhaltig zu steigern."

- Dr. Michael Max Buehler, Professor für Baubetriebswirtschaft an der HTWG Konstanz, Co-Founder von Gemein-Werk Ventures, unabhängiger Direktor bei DevvStream und Instructor für Lean Construction an der Construction Akademie Deutschland.

„Das Buch "DataDrivenConstruction" ist einer der ersten Schritte über die Grenzen der gewohnten Welt der Bauherren mit ihren komplexen Planungs- und Managementsystemen hinaus, in der, so scheint es, die Komplexität und Sättigung der Daten nicht einmal eine Chance für eine radikale Vereinfachung und erhöhte Transparenz der Arbeit mit Baudaten bietet. In seinem Buch zeigt Artem in einfacher Sprache, welche Möglichkeiten die modernen Technologien für die Arbeit mit Daten uns eröffnen, und gibt buchstäblich konkrete Schritte vor, die Sie sofort in Ihrer Arbeit anwenden können. Ich empfehle jedem, der verstehen will, wohin sich Automatisierungssysteme im Bauwesen entwickeln werden, dieses Buch aufmerksam zu lesen, um zu erkennen, dass die Datenrevolution im Bauwesen bereits an unsere Tür klopft. Noch ist sie nur für Freaks interessant, aber in ein paar Jahren werden solche Ansätze und Software, wie BIM, allgegenwärtig sein!"

- Ihor Rogachew, Leiter des IMT-Kompetenzzentrums, BIM & Digitale Transformation bei RGD, und Gründer von InfraBIM.Pro.

"Alles, was ich sagen kann, ist: WOW! Die Art und Weise, wie Sie die Geschichte, den LLM, die Grafiken und die allgemeine Leichtigkeit des Verständnisses Ihrer Punkte integriert haben, ist wirklich bemerkenswert. Der Fluss des Buches ist erstaunlich. Es gibt so viele brillante Aspekte in diesem Buch; es ist wirklich ein echter Wendepunkt. Es ist eine großartige Informationsquelle, und ich lobe Sie für die Mühe und Leidenschaft, die Sie hineingesteckt haben. Herzlichen Glückwunsch zu einem so bemerkenswerten Werk. Ich könnte noch mehr sagen, aber es genügt, wenn ich sage, dass ich unglaublich beeindruckt bin!"

- Natasha Prinsloo, Digital Practice Lead bei energylab_.

"Ich kann das Buch DataDrivenConstruction sehr empfehlen, das sich, wie der Titel schon sagt, mit einem datengesteuerten Informationsmanagementansatz für AECO befasst. Ich verwende es derzeit, um eine Reihe von Diskussionen mit verschiedenen Gruppen zu initiieren. Ich finde, es ist ein sehr zugängliches Nachschlagewerk. Neben einem gründlichen Überblick über den historischen Kontext von AECO-Tools, Daten und einer Einführung in verschiedene Schlüsseltechnologien enthält das Buch eine Reihe sehr nützlicher Diagramme, die den Umfang von Datenquellen und Endnutzer-Artefakten mit Beispiel-Workflows skizzieren. Ich bin der Meinung, dass wir diese Art von Diagrammen bei der Entwicklung und Überwachung von Informationsstrategien und als Beitrag zu BEP's verstärkt brauchen - sie definieren das gesamte Unternehmensdatenmodell, auf das die Grenzen für ein PIM und AIM übertragen werden können."

- Paul Ransley, Principal Consultant bei Acmena und Systemintegrationsingenieur bei Transport for London.

"Daten sind das neue Öl", wie man so schön sagt. Daher sollten die Prospektoren oder Bergleute über die richtigen Werkzeuge und die richtige Einstellung verfügen, um aus dieser Ressource des 21. Jahrhunderts zu nutzen. Die Baubranche hat sich zu lange auf der schiefen Ebene von "3D-Informations"-basierten Prozessen bewegt, bei denen die Projektabwicklung auf den gebackenen Informationen eines anderen beruht (z. B. wurde das Torten- oder Balkendiagramm bereits erstellt), während die zugrundeliegenden "Daten" (z. B. die rohe Kalkulationstabelle) in der Lage sind, viel mehr zu liefern, vor allem, weil Multidatenfusion und KI unbegrenzte Potenziale bieten. Wenn Sie in der Baubranche tätig sind (oder lehren/forschen), ist dieses Buch Ihre beste - und bisher einzige - Ressource für die Navigation in der datengesteuerten Welt, in der wir uns befinden."

- Dr. Zulfikar Adamu, außerordentlicher Professor für strategische IT im Bauwesen an der LSBU, Großbritannien.

"Wie in dem Buch hervorgehoben wird, sind Informationen für den Bausektor von entscheidender Bedeutung, und wenn sie in zugänglichen Formaten vorliegen, erleichtert dies die genaue Entscheidungsfindung und beschleunigt die Projektfristen. Das Buch bietet einen neutralen und effizienten Ansatz, um auf diese Quelle zuzugreifen und sie für die Entscheidungsfindung zu nutzen. Die im Buch vorgestellte Methodik nutzt einen zeitgemäßen Ansatz, der auf künstlicher Intelligenz basierende Programmierung mit zugänglichen Open-Source-Tools kombiniert. Durch die Nutzung der Leistungsfähigkeit von KI und die Verwendung von Open-Source-Software zielt die Methodik darauf ab, die Automatisierung zu verbessern, Prozesse zu optimieren und die Zugänglichkeit und Zusammenarbeit in diesem Bereich zu fördern. Die Sprache des Buches ist klar und einfach zu verstehen."

- Dr. Salih Ofluoğlu, Dekan der Fakultät für Fine Arts and Architecture an der Antalya Bilim Universität und Organisator des Eurasischen BIM-Forums.

"Wenn "Daten das neue Öl" sind, müssen wir lernen, sie zu definieren, zu finden, abzubauen und zu veredeln, um sie wertvoll zu machen. Ich habe das Buch DataDrivenConstruction als sehr informativ und aufschlussreich empfunden. Das Buch bietet einen nützlichen historischen Hintergrund und erklärt die Arbeit mit Daten in einfacher Sprache. Für diejenigen, die sich für die digitale Transformation interessieren, vermittelt es ein gutes Verständnis von Daten - wie sie funktionieren, wie sie strukturiert sind und wie sie genutzt werden können."

- Ralph Montague, Direktor bei ArcDox, Direktor des BIM Coordinators Summit und Vorsitzender des BIM National Mirror Committee bei der National Standards Authority of Ireland

"Data Driven Construction vermittelt anschaulich die Grundlagen zum Informationsgestüten arbeiten mit Gebäudedaten. Ein Buch, das sich mit Informationsflüssen und grundlegenden wirtschaftlichen Konzepten beschäftigt und sich dadurch von anderen BIM Büchern abgrenzt, da es nicht nur die Sichtweise eines Softwareherstellers repräsentiert, sondern versucht grundlegende Konzepte zu vermitteln. Ein lesens- und sehenswertes Buch."

- Jakob Hirn, Geschäftsführer und Co-Founder der Build Informed GmbH, und Initiator des Innovationsforums "On Top With BIM".

"Für jeden in der Baubranche, vom Neuling bis zum erfahrenen Profi, ist dieses Buch eine echte Bereicherung! Es ist keine typische staubige Lektüre - es ist vollgepackt mit Erkenntnissen, Strategien und einer Prise Humor, um Sie bei der Stange zu halten. Von alten Datenerfassungsmethoden bis hin zu modernsten digitalen Technologien deckt es die Entwicklung der Datennutzung im Bauwesen ab. Es ist wie eine Zeitmaschine, die Sie durch die Evolution der Baudaten führt. Ob Sie nun Architekt, Ingenieur, Projektmanager oder Datenanalyst sind, dieses umfassende Handbuch wird Ihre Herangehensweise an Projekte revolutionieren. Machen Sie sich bereit, Prozesse zu optimieren, die Entscheidungsfindung zu verbessern und Projekte wie nie zuvor zu managen!"

- Pierpaolo Vergati, Dozent an der Universität Sapienza in Rom und leitender Projektmanager im Bauwesen bei Fintecna.

"Ich habe das Buch in einem Atemzug gelesen, in weniger als 6 Stunden. Die Verarbeitungsqualität des Buches ist ausgezeichnet, dichtes Hochglanzpapier, Farbschemata, eine angenehme Schriftart. Die große Anzahl praktischer Beispiele für die Arbeit mit LLM speziell für die Baubranche erspart Ihnen Monate, wenn nicht Jahre des Selbststudiums. Die Arbeitsbeispiele sind sehr vielfältig und reichen von einfach bis komplex, ohne dass Sie eine komplexe und teure Software kaufen müssen. Das Buch ermöglicht es Inhabern aller Unternehmen der Baubranche, einen neuen Blick auf ihre Geschäftsstrategie, die Digitalisierung und die Entwicklungsperspektiven zu werfen. Und für kleinere Unternehmen, die Effizienz mit erschwinglichen und kostenlosen Tools zu steigern."

- Mikhail Kosarev, Dozent und Berater für digitale Transformation in der Bauindustrie bei TIM-ASG.

"Das Buch "DATA DRIVEN CONSTRUCTION" ist ein Meilenstein für alle, die wissen wollen, wohin sich die Baubranche im Zeitalter der Daten entwickelt. Artem kratzt nicht nur an der Oberfläche; er taucht tief in die aktuellen Entwicklungen, Herausforderungen und vielversprechenden Möglichkeiten im Bauwesen ein. Was dieses Buch von anderen abhebt, ist seine Zugänglichkeit - Artem erklärt komplexe Ideen anhand von nachvollziehbaren Analogien, die den Inhalt leicht begreifbar machen. Ich empfand das Buch als unglaublich informativ und gleichzeitig fesselnd. Zusammenfassend lässt sich sagen, dass Artem eine wertvolle Ressource geschaffen hat, die nicht nur informiert, sondern auch inspiriert. Ob Sie nun ein erfahrener Fachmann oder ein Neuling im Bauwesen sind, dieses Buch wird Ihre Perspektive erweitern und Ihr Verständnis dafür vertiefen, wohin sich die Branche entwickelt. Äußerst empfehlenswert!"

- Moayad Saleh, Architekt und BIM Implementierungsmanager bei der TMM GROUP Gesamtplanungs GmbH.

"Ich muss sagen, dass Data-Driven Construction es wert ist, als Lehrbuch an Universitäten gelehrt zu werden, und dass es ein Buch ist, das einen wertvollen Beitrag zum sich entwickelnden Bereich BIM leisten wird. Das Buch "Data-Driven Construction" enthält ein technisches Glossar, das die Konzepte sehr gut erklärt. Themen, die extrem schwierig zu erklären sind, werden mit einer sehr schönen Bildsprache einfach und verständlich gemacht. Ich denke, dass das, was in den Bildern erklärt werden soll, dem Leser, wenn auch nur kurz, vermittelt werden sollte. Die Verständlichkeit mancher Bilder, also das Lesen des Bildes, erfordert eine gesonderte Information. Ich möchte auch sagen, dass ich gerne die wertvolle Arbeit von Artem Boiko in meinen Vorlesungen und Seminaren an Universitäten vorstelle."

- Dr. Ediz Yazicioglu, Inhaber von ArchCube und Dozent für Bauprojektmanagement an der Fakultät für Architektur der Technischen Universität Istanbul und der Medipol-Universität.

""Data-Driven Construction" von Artem Boiko ist ein beeindruckendes Werk, das in Zeiten ständig wachsender Technologien und Informationsmöglichkeiten eine fundierte Grundlage für die Baubranche bietet. Boiko gelingt es, komplexe Themen verständlich aufzubereiten und gleichzeitig visionäre Ideen einzubringen. Das Buch ist ein durchdachtes Kompendium, das nicht nur aktuelle Entwicklungen beleuchtet, sondern auch einen Ausblick auf zukünftige Innovationen gibt. Eine klare Empfehlung für alle, die sich mit datengetriebener Bauplanung und -ausführung auseinandersetzen möchten."

- Markus Eiberger, Lehrbeauftragter an der Hochschule für Technik Stuttgart, Senior Project Manager und stellvertretender Niederlassungsleiter der Konstruktionsgruppe Bauen, Vorstandsmitglied des BIM-Cluster Baden-Württemberg.

Zweite Auflage, April 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-9827303-1-8

REGISTRIERTES URHEBERWERK
ID-CHECK D9A2A2C55EAC5
WWW.PRIORMART.COM

REGISTRIERTES URHEBERWERK
ID-CHECK 8ECEC0192F499
WWW.PRIORMART.COM

Artem Boiko Urheberrecht

boikoartem@gmail.com
info@datadrivenconstruction.io

Kein Teil dieses Buches darf ohne schriftliche Genehmigung des Autors in irgendeiner Form oder mit irgendwelchen Mitteln, elektronisch oder mechanisch, einschließlich Fotokopien, Aufzeichnungen oder Informationsspeicher- und -abrufsystemen, vervielfältigt oder übertragen werden - mit Ausnahme der nichtkommerziellen Verbreitung in unveränderter Form. Das Buch wird kostenlos verteilt und darf zu persönlichen, Bildungs- oder Forschungszwecken mit anderen Nutzern geteilt werden, vorausgesetzt, dass die Urheberschaft und die Verweise auf das Original beibehalten werden. Der Autor behält sich alle nicht-proprietären Rechte an dem Text vor und übernimmt keine ausdrücklichen oder stillschweigenden Garantien. Im Buch erwähnte Unternehmen, Produkte und Namen können fiktiv sein oder als Beispiele dienen. Der Autor haftet nicht für die Folgen der Verwendung der gegebenen Informationen. Die in diesem Buch enthaltenen Informationen werden ohne Gewähr auf Vollständigkeit und Aktualität zur Verfügung gestellt. Der Autor haftet nicht für zufällige oder Folgeschäden, die sich aus der Verwendung der in diesem Buch enthaltenen Informationen, Codes oder Programme ergeben. Die in diesem Buch vorgestellten Code-Beispiele dienen nur zu Lehrzwecken. Der Leser benutzt sie auf eigenes Risiko. Der Autor empfiehlt, dass alle Softwarelösungen vor dem Einsatz in einer Produktionsumgebung getestet werden. Alle im Text erwähnten Warenzeichen und Produktnamen sind Warenzeichen, eingetragene Warenzeichen oder Dienstleistungsmarken der jeweiligen Unternehmen und sind Eigentum der jeweiligen Inhaber. Die Verwendung dieser Namen in diesem Buch bedeutet nicht, dass eine Beziehung zu den Eigentümern besteht oder dass sie von diesen unterstützt werden. Die Erwähnung von Produkten oder Dienstleistungen Dritter stellt keine Empfehlung dar und impliziert keine Befürwortung. Die in den Beispielen verwendeten Firmen- und Produktnamen können Marken der jeweiligen Eigentümer sein. Links zu Websites Dritter werden aus Gründen der Zweckmäßigkeit zur Verfügung gestellt und bedeuten nicht, dass der Autor die auf diesen Websites bereitgestellten Informationen gutheißt. Alle zitierten Statistiken, Zitate und Studien waren zum Zeitpunkt der Erstellung des Artikels aktuell. Die Daten können sich im Laufe der Zeit ändern

Dieses Buch wird unter einer Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Lizenz (CC BY-NC-ND 4.0) verbreitet. Sie dürfen es für nicht-kommerzielle Zwecke vervielfältigen und verbreiten, sofern Sie die Urheberschaft beibehalten und keine Änderungen vornehmen.

CC BY-NC-ND 4.0

© 2024 Artem Boiko. Erste Ausgabe.
© 2025 Artem Boiko. Zweite, überarbeitete und ergänzte Auflage.
Alle Rechte vorbehalten.

VORWORT ZUR ZWEITEN AUFLAGE

Dieses Buch ist das Ergebnis eines lebhaften Dialogs mit der Fachwelt. Es basiert auf zahlreichen Fachdiskussionen zum Thema Datenmanagement in der Bauwirtschaft, die auf verschiedenen Fachplattformen und Social-Media-Plattformen stattfanden. Diese Diskussionen bildeten die Grundlage für Artikel, Veröffentlichungen und Bildmaterial, die in der Fachwelt ein breites Echo hervorgerufen haben. Die Inhalte des Autors werden jedes Jahr auf verschiedenen Plattformen und in verschiedenen Sprachen millionenfach aufgerufen und bringen Fachleute aus dem Bereich der Digitalisierung des Bauwesens zusammen.

Innerhalb eines Jahres nach Erscheinen der ersten Auflage wurde das Buch von Experten aus mehr als 50 Ländern bestellt, von Brasilien und Peru bis Mauritius und Japan. Die zweite Auflage des Buches, das Sie jetzt in Händen halten, wurde auf der Grundlage von Expertenfeedback, Kritik an der ersten Auflage und Diskussionen in Fachkreisen überarbeitet und erweitert. Dank des Feedbacks wurde die zweite Auflage deutlich erweitert: Neue Kapitel über CAD (BIM) Technologien und die Erstellung effektiver ETL -Prozesse wurden hinzugefügt. Auch die Zahl der Praxisbeispiele und Fallstudien wurde deutlich erhöht. Besonders wertvoll sind die Rückmeldungen von Führungskräften aus der Bauindustrie, von Beratungsunternehmen und großen IT-Firmen, die sowohl vor als auch nach Erscheinen der ersten Version des Buches mit Fragen zur Digitalisierung und Interoperabilität an den Autor herangetreten sind. Viele von ihnen haben die im Buch beschriebenen Ansätze bereits angewandt oder planen, dies in naher Zukunft zu tun

Sie halten ein Buch in den Händen, das durch Diskussion und aktiven Meinungsaustausch entstanden ist. Fortschritt entsteht im Dialog, im Aufeinanderprallen von Meinungen und in der Offenheit für neue Ansätze. Ich danke Ihnen, dass Sie sich an diesem Dialog beteiligen. Ihre konstruktive Kritik ist die Grundlage für künftige Verbesserungen. Wenn der Text Fehler enthält oder wenn Sie Ideen und Vorschläge einbringen möchten, ist jede Rückmeldung willkommen. Kontaktinformationen finden Sie am Ende des Buches

WARUM IST DAS BUCH KOSTENLOS?

Dieses Buch wurde als offene Bildungsressource konzipiert, die darauf abzielt, moderne Ansätze für das Datenmanagement in der Bauindustrie zu verbreiten. Die erste Version des Buches diente als Grundlage für die Sammlung von Kommentaren und Vorschlägen aus der Fachwelt, die eine Verbesserung der Struktur und des Inhalts des Materials ermöglichten. Alle Kommentare, Vorschläge und Ideen wurden sorgfältig analysiert und in diese überarbeitete Version eingearbeitet. Ziel des Buches ist es, Baufachleuten zu vermitteln, wie wichtig es ist, mit Daten zu arbeiten: systematisch, bewusst und mit Blick auf den langfristigen Wert der Informationen. Der Autor hat Beispiele, Illustrationen und praktische Beobachtungen aus mehr als 10 Jahren Arbeit im Bereich der Digitalisierung des Bauwesens gesammelt. Das meiste Material stammt aus realen Projekten, Gesprächen mit Ingenieuren und Bauherren, der Teilnahme an internationalen Initiativen und Schulungsseminaren. Das Buch ist ein Versuch, die gesammelten Erfahrungen zu strukturieren und sie in einer zugänglichen Form zu teilen. Wenn Sie die weitere Verbreitung der Ideen des Buches unterstützen und ein bequemes Format zum Lesen, Arbeiten mit Beispielen und visuellem Material erhalten möchten, können Sie eine gedruckte Version erwerben.

NUTZUNGSRECHTE

Alle Materialien, Abbildungen und Fragmente dieses Buches dürfen in jedem Format und auf jedem Medium vervielfältigt, zitiert oder verwendet werden, vorausgesetzt, die Quelle wird Artem Boiko zugeschrieben und der Titel des Buches lautet "Data-Driven Construction". Wir danken Ihnen für die Achtung der Arbeit und die Verbreitung von Wissen.

Mit aufrichtiger Dankbarkeit widme ich dieses Buch meiner Familie, die mir von klein auf eine tiefe Liebe zum Bauwesen eingeflößt hat, meiner Heimatstadt im Bergbau, die mir eine Lektion in Sachen Widerstandsfähigkeit erteilt hat, und meiner Frau, einer Vermessungsingenieurin, deren unermüdliche Unterstützung meine ständige Inspiration war.

FÜR WEN DIESES BUCH BESTIMMT IST

Dieses Buch ist in einer verständlichen Sprache geschrieben und richtet sich an ein breites Spektrum von Lesern in der Baubranche - von Studenten und Anfängern, die die Grundlagen moderner Bauprozesse verstehen wollen, bis hin zu Fachleuten, die eine aktuelle Methodik für das Datenmanagement im Bauwesen benötigen. Egal, ob Sie Architekt, Ingenieur, Polier, Bauleiter oder Datenanalyst sind, dieser umfassende Leitfaden mit vielen einzigartigen Illustrationen und Grafiken bietet wertvolle Einblicke in die Nutzung von Daten im Unternehmen, um Prozesse zu optimieren und zu automatisieren, die Entscheidungsfindung zu verbessern und Bauprojekte auf verschiedenen Ebenen mit modernen Tools zu verwalten

Das Buch ist ein umfassender Leitfaden, der theoretische Grundlagen und praktische Empfehlungen für die Integration von Datenmanagementtechniken in Bauprozesse kombiniert. Der Schwerpunkt des Buches liegt auf der strategischen Nutzung von Informationen zur Optimierung von Abläufen, zur Automatisierung von Prozessen, zur Verbesserung der Entscheidungsfindung und zur effektiven Verwaltung von Projekten mithilfe moderner digitaler Werkzeuge.

Dieses Buch behandelt die theoretischen und praktischen Aspekte der Arbeit mit Informationen in der Bauindustrie. Anhand detaillierter Beispiele wird die Methodik der Aufgabenparametrisierung, der Anforderungserfassung, der Verarbeitung unstrukturierter und formatübergreifender Daten und deren Umwandlung in effektive Lösungen für Bauunternehmen untersucht.

Der Leser durchläuft sukzessive den Weg von der Anforderungsbildung und der Entwicklung grundlegender Datenmodelle bis hin zu komplexeren Prozessen der Integration heterogener Informationsquellen, der Erstellung von ETL -Prozessen, dem Aufbau von Informationspipelines und maschinellen Lernmodellen. Der sequenzielle Ansatz ermöglicht es, die Mechanismen der Organisation und Automatisierung von Geschäftsprozessen und Entscheidungsunterstützungssystemen in der Bauindustrie deutlich zu machen. Jeder Teil des Buches schließt mit einem praktischen Kapitel ab, das Schritt-für-Schritt-Anleitungen enthält, die eine sofortige Anwendung des erworbenen Wissens in realen Projekten ermöglichen.

ZUSAMMENFASSUNG VON TEILEN DES BUCHES

Dieses Buch ist um das Konzept der Datentransformation in der Wertschöpfungskette herum aufgebaut: von der Datenerfassung und Qualitätssicherung über die analytische Verarbeitung bis hin zur Gewinnung wertvoller praktischer Lösungen mit Hilfe moderner Werkzeuge und Methoden.

Teil 1: Digitale Entwicklung im Bauwesen - zeichnet den historischen Wandel der Datenverwaltung von Tontafeln zu modernen digitalen Systemen nach und analysiert die Entstehung modularer Systeme und die wachsende Bedeutung der Digitalisierung von Informationen im Kontext industrieller Revolutionen.

Teil 2: Informationsherausforderungen für die Bauindustrie - untersucht die Probleme der Datenfragmentierung, "Informationssilos", die Auswirkungen des HiPPO-Ansatzes auf die Entscheidungsfindung und die Grenzen proprietärer Formate und regt an, den Übergang zu KI und LLM-Ökosystemen zu erwägen.

Teil 3: Systematisierung von Daten im Bauwesen - bildet eine Typologie von Baudaten, beschreibt Methoden ihrer Organisation, die Integration in Unternehmenssysteme und erörtert die Schaffung von Kompetenzzentren für die Standardisierung von Informationsprozessen.

Teil 4: Datenqualitätssicherung - zeigt Methoden zur Umwandlung disparater Informationen in hochwertige, strukturierte Daten, einschließlich Datenextraktion aus verschiedenen Quellen, Validierung und Modellierung mit LLM.

Teil 5: Kosten- und Zeitkalkulationen - behandelt die Digitalisierung von Kosten- und Planungskalkulationen, die Automatisierung der Mengenermittlung aus CAD-Modellen (BIM), 4D-8D-Modellierungstechnologien und die ESG-Kalkulation von Bauprojekten.

Teil 6: CAD und BIM - analysiert kritisch die Entwicklung von Entwurfstechnologien, Fragen der Interoperabilität von Systemen, Trends zu offenen Datenformaten und die Aussichten für die Anwendung künstlicher Intelligenz im Entwurf.

Teil 7: Datenanalyse und Automatisierung - befasst sich mit den Grundsätzen der Informationsvisualisierung, den wichtigsten Leistungsindikatoren, ETL-Prozessen, Tools zur Workflow-Orchestrierung und der Anwendung von Sprachmodellen zur Automatisierung von Routineaufgaben.

Teil 8: Datenspeicherung und -verwaltung - befasst sich mit Datenspeicherformaten, Data-Warehouse- und Data-Lake-Konzepten, Datenverwaltungsprinzipien und neuen Ansätzen wie Vektordatenbanken und den Methoden DataOps und VectorOps.

Teil 9: Big Data und maschinelles Lernen - befasst sich mit dem Übergang zu objektiven Analysen auf der Grundlage historischer Daten, dem Internet der Dinge auf Baustellen und der Anwendung von Algorithmen des maschinellen Lernens zur Vorhersage von Projektkosten und -fristen.

Teil 10: Das Baugewerbe im Zeitalter digitaler Daten - wirft einen Blick auf die Zukunft des Baugewerbes und analysiert den Wechsel von der Kausalanalyse zur Arbeit mit Korrelationen, das Konzept der "Uberisierung" des Bauwesens und Strategien für die digitale Transformation.

What is meant by **data-driven construction**?

Led by CEO and project managers

A data-driven approach to construction is the **process** of implementing **connected and automated systems** that enable companies to **continuously improve** **project delivery** and operational efficiency while **reducing costs and risks**, ultimately creating a **sustainable competitive advantage.**

Focuses on real-time analytics

It's a mindset shift!

Streamline operations & enhance collaboration

The ultimate goal

EINFÜHRUNG

Wie lange kann Ihr Unternehmen in einer Welt wettbewerbsfähig bleiben, in der sich die Technologie rasant weiterentwickelt und jeder Aspekt des Geschäfts, von der Zeitplanung und Kostenkalkulation bis hin zur Risikoanalyse, durch maschinelle Lernmodelle automatisiert wird?

Das Baugewerbe, das es schon so lange gibt wie die Menschheit selbst, steht an der Schwelle zu einem revolutionären Wandel, der die Art und Weise, wie wir über das traditionelle Bauen denken, völlig verändern wird. Bereits in anderen Wirtschaftszweigen ändert die Digitalisierung nicht nur die Regeln, sondern verdrängt rücksichtslos Unternehmen vom Markt, die sich nicht an die neue Datenverarbeitungsumgebung angepasst haben und nicht in der Lage sind, die Entscheidungsgeschwindigkeit zu verbessern (Abb. 1).

Abb. 1 Die Geschwindigkeit der Entscheidungsfindung im Baugewerbe hängt häufiger als in anderen Branchen vom Faktor Mensch ab.

Das Bankwesen, der Einzelhandel, die Logistik und die Agrarwirtschaft bewegen sich rasch auf eine vollständige Digitalisierung zu, bei der Ungenauigkeiten und subjektive Meinungen keinen Platz mehr haben. Moderne Algorithmen sind in der Lage, enorme Datenmengen zu analysieren und den Kunden genaue Vorhersagen zu liefern - sei es die Wahrscheinlichkeit der Kreditrückzahlung, optimale Lieferwege oder Risikoprognosen.

Das Baugewerbe ist eine der letzten Branchen, die den unvermeidlichen Übergang von Lösungen, die auf der Meinung hoch bezahlter Spezialisten beruhen, zu datengesteuerten Lösungen vollzieht. Dieser Übergang wird nicht nur durch neue technologische Möglichkeiten vorangetrieben, sondern auch durch die gestiegenen Anforderungen des Marktes und der Kunden an Transparenz, Genauigkeit und Geschwindigkeit.

Robotisierung, Prozessautomatisierung, offene Daten und darauf basierende Prognosen sind nicht mehr nur Möglichkeiten, sondern unumgänglich. Die meisten Unternehmen der Baubranche, die noch vor kurzem gegenüber dem Kunden für die Kalkulation von Umfang, Kosten, Zeit der Projekte und Qualitätskontrolle

verantwortlich waren, laufen nun Gefahr, zu bloßen Ausführern von Aufträgen zu werden, die keine wichtigen Entscheidungen treffen (Abb. 2).

Dank der Fortschritte bei der Rechenleistung, den Algorithmen des maschinellen Lernens und dem demokratischen Zugang zu Daten ist es jetzt möglich, Daten aus verschiedenen Quellen automatisch zu kombinieren, was eine tiefere Prozessanalyse, Risikovorhersage und Kostenoptimierung in den Diskussionsphasen eines Bauprojekts ermöglicht. Diese Technologien haben das Potenzial, die Effizienz radikal zu verbessern und die Kosten in der gesamten Branche zu senken.

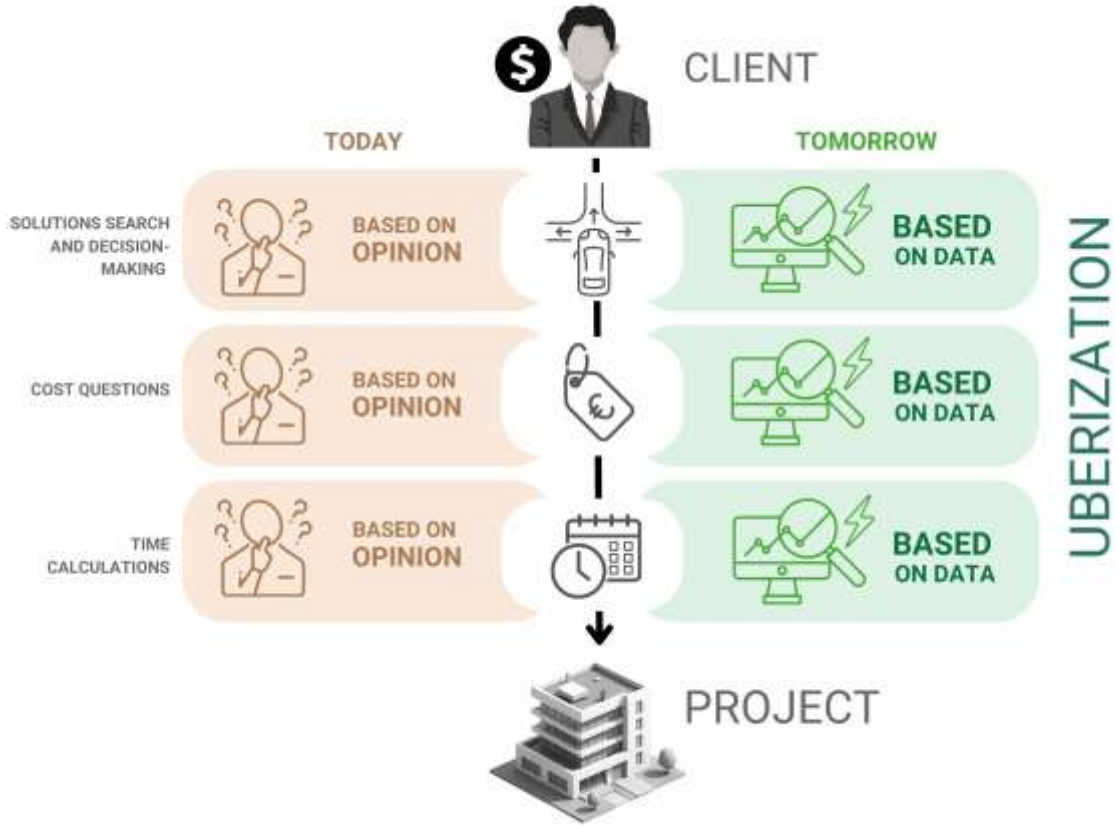


Abb. 2 Der Kunde ist nicht an übermäßigen menschlichen Faktoren auf dem Weg zur Verwirklichung seines Projekts interessiert.

Trotz aller Vorteile neuer Werkzeuge und Konzepte hinkt die Bauindustrie bei der Übernahme neuer Technologien weit hinter anderen Wirtschaftszweigen hinterher.

> Laut dem Bericht "IT Metrics Key Data 2017" liegt das Baugewerbe bei den IT-Ausgaben unter 19 anderen Branchen an letzter Stelle [1].

Das rasante Wachstum der Datenmenge und die Komplexität der Prozesse bereiten der Unternehmensführung zunehmend Kopfzerbrechen. Das Hauptproblem bei der Nutzung neuer Technologien besteht darin, dass die Daten trotz ihrer Fülle fragmentiert, unstrukturiert und oft nicht kompatibel zwischen verschiedenen Systemen und Softwareprodukten sind. Aus diesem Grund beschäftigen sich viele Unternehmen im Bausektor heute vor allem mit Fragen der Datenqualität, die nur durch die Einführung effizienter, automatisierter Verwaltungs- und Analysesysteme gelöst werden können.

> Laut einer von KPMG® durchgeführten Umfrage unter Bauleitern im Jahr 2023 [2] haben Projektmanagement-Informationssysteme (PMIS), fortgeschrittene und grundlegende Datenanalyse und Gebäudedatenmodellierung (BIM) das größte Potenzial zur Verbesserung des ROI von Projekten (Abb. 3).

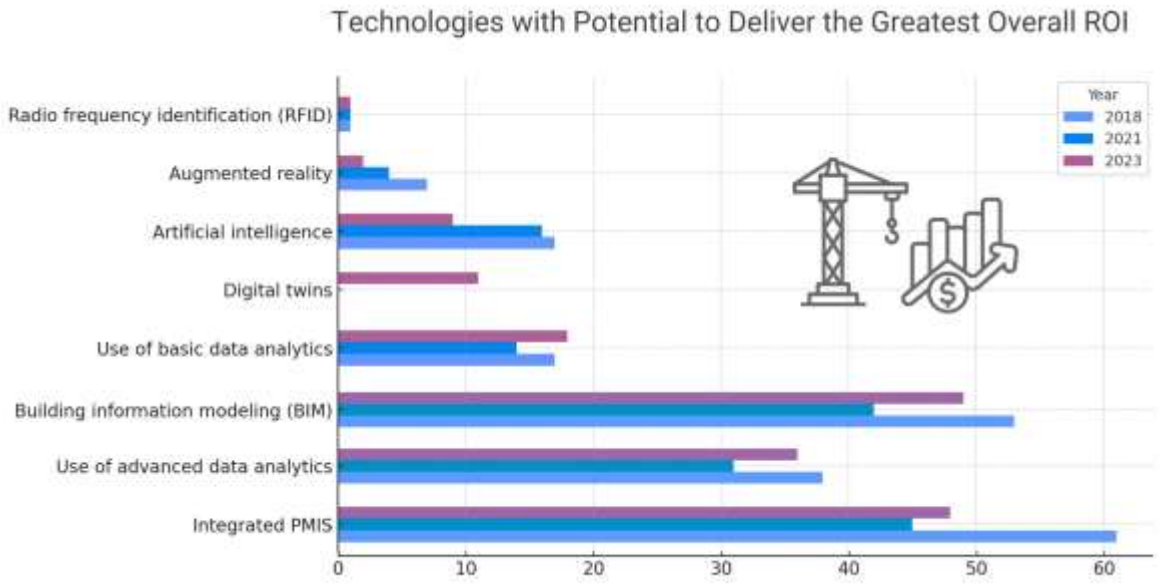


Abb. 3 Umfrage unter Managern von Bauunternehmen: Welche Technologien bieten die höchste Investitionsrendite (ROI) bei Investitionsprojekten? (basierend auf Materialien [2]).

Die Lösung für die Herausforderungen, die mit der Integration von Daten in Geschäftsprozesse verbunden sind, besteht in der Sicherstellung einer hohen Informationsqualität, der Verwendung geeigneter Datenformate und der Anwendung effektiver Methoden zur Erstellung, Speicherung, Analyse und Verarbeitung von Daten.

Die Erkenntnis des Wertes von Daten zwingt verschiedene Branchen dazu, sich von isolierten Anwendungen und komplexen bürokratischen Verwaltungsstrukturen zu lösen. Stattdessen verlagert sich der Schwerpunkt auf die Schaffung neuer Ansätze für die Informationsarchitektur, die Unternehmen in moderne datengesteuerte Unternehmen verwandeln. Früher oder später wird auch das Baugewerbe diesen Schritt vollziehen und von einer schrittweisen digitalen Evolution zu einer echten digitalen Revolution übergehen, die alle Unternehmen betrifft.

> Der Übergang zu datengesteuerten Geschäftsprozessen wird nicht einfach sein. Viele Unternehmen werden vor Herausforderungen stehen, weil die Führungskräfte nicht immer wissen, wie sie chaotische Datensätze nutzen können, um die Effizienz und das Unternehmenswachstum zu verbessern.

Dieses Buch taucht ein in die Welt der Daten, in der Informationen zu einer wichtigen strategischen Ressource werden, die die Effizienz und Nachhaltigkeit von Geschäftsprozessen bestimmt. Mit dem rasanten Wachstum von Informationen stehen Unternehmen vor neuen Herausforderungen. Die digitale Transformation ist nicht mehr nur ein Schlagwort - sie wird zu einer Notwendigkeit.

Abb. 4 Daten und Prozesse sind die Grundlage der Konstruktion.

Transformation zu verstehen bedeutet, das Komplexe in einfachen Worten erklären zu können. Deshalb ist das Buch in einer leicht verständlichen Sprache geschrieben und wird von Illustrationen des Autors begleitet, die eigens geschaffen wurden, um Schlüsselkonzepte klar zu erklären. Diese Diagramme, Schaubilder und Visualisierungen sollen Wahrnehmungsbarrieren beseitigen und den Stoff auch für diejenigen verständlich machen, denen solche Themen bisher zu komplex erschienen. Alle Illustrationen, Diagramme und Grafiken in diesem Buch wurden vom Autor erstellt und speziell für die Visualisierung der im Text beschriebenen Schlüsselkonzepte konzipiert.

> *Ein Bild sagt mehr als tausend Worte [3].*
>
> - Fred R. Barnard, englischer Illustrator, 1927.

Um die Theorie mit der Praxis zu verbinden, werden wir Werkzeuge der künstlichen Intelligenz (insbesondere Sprachmodelle) verwenden, die es Ihnen ermöglichen, Lösungen zu entwickeln, ohne dass Sie über tiefgreifende Programmierkenntnisse verfügen müssen. Wenn Sie sich an praktischem Material orientieren und mehr an der praktischen Arbeit mit Daten interessiert sind, können Sie den ersten einführenden Teil überspringen und direkt zum zweiten Teil des Buches übergehen, wo die Beschreibung konkreter Beispiele und Fälle beginnt.

Allerdings sollten Sie keine überzogenen Erwartungen in KI (Künstliche Intelligenz), maschinelles Lernen und LLM (Large Language Models) Tools im Allgemeinen setzen. Ohne qualitativ hochwertige Eingabedaten und ein tiefes Verständnis der Materie können selbst die fortschrittlichsten Algorithmen keine zuverlässigen und aussagekräftigen Ergebnisse liefern.

> Der CEO von Microsoft, Satya Nadella, warnt vor der Gefahr einer Blase im Bereich der künstlichen Intelligenz Anfang 2025 [4] und vergleicht den derzeitigen Hype mit der Dot-Com-Blase. Er betont, dass Behauptungen über das Erreichen von AGI-Meilensteinen (Artificial General Intelligence) ohne angemessene Begründung eine "bedeutungslose Manipulation von Metriken" sind. Nadella ist der Ansicht, dass der tatsächliche Erfolg der KI an ihrem Beitrag zum globalen BIP-Wachstum gemessen werden sollte, anstatt sich auf Schlagworte zu versteifen.

Hinter all den Schlagwörtern über neue Technologien und Konzepte verbirgt sich eine komplexe und mühsame Arbeit zur Sicherstellung der Datenqualität, zur Parametrisierung der Geschäftsprozesse und zur Anpassung der Werkzeuge an die realen Aufgaben.

> Ein datengesteuerter Ansatz ist kein Produkt, das man einfach herunterladen oder kaufen kann. Es ist eine Strategie, die aufgebaut werden muss. Sie beginnt mit einem neuen Blick auf die bestehenden Prozesse und Probleme und erfordert dann eine disziplinierte Bewegung in die gewählte Richtung.

Führende Softwareentwickler und Anwendungsanbieter werden nicht der Motor des Wandels in der Baubranche sein, denn für viele von ihnen ist ein datengesteuerter Ansatz eine Bedrohung für ihr etabliertes Geschäftsmodell.

> *Andere Branchen [im Gegensatz zum Baugewerbe], wie z. B. die Automobilindustrie, haben bereits einen radikalen und disruptiven Wandel durchlaufen, und ihre digitale Transformation ist in vollem Gange. Bauunternehmen müssen schnell und entschlossen handeln: Flinke Unternehmen werden enorme Gewinne erzielen, während für diejenigen, die zögern, die Risiken sehr hoch sind. Denken Sie nur an die Umwälzungen, die die digitale Fotografie in dieser Branche verursacht hat [5].*
>
> - Bericht des Weltwirtschaftsforums Shaping the Future of Construction, 2016

Diejenigen Unternehmen, die die Chancen und Vorteile des neuen Ansatzes rechtzeitig erkennen, werden einen nachhaltigen Wettbewerbsvorteil erlangen und in der Lage sein, sich zu entwickeln und zu wachsen, ohne von den Lösungen großer Anbieter abhängig zu sein.

Dies ist Ihre Chance, dem kommenden Sturm der Digitalisierung von Informationen nicht nur zu trotzen, sondern ihn zu beherrschen. In diesem Buch finden Sie nicht nur eine Analyse des aktuellen Stands der Branche, sondern auch konkrete Empfehlungen, wie Sie Ihre Prozesse und Ihr Unternehmen überdenken und umstrukturieren können, um in der neuen Ära des Bauwesens führend zu werden und Ihre berufliche Erfahrung zu verbessern.

Bei der digitalen Zukunft des Bauwesens geht es nicht nur um den Einsatz neuer Technologien und Programme, sondern um ein grundlegendes Überdenken der Datenverarbeitung und der Geschäftsmodelle.

Ist Ihr Unternehmen bereit für diesen strategischen Wandel?

INHALTSVERZEICHNIS

EINFÜHRUNG **1**

INHALTSVERZEICHNIS ..... **I**

II TEIL VON TONTAFELN ZUR DIGITALEN REVOLUTION: WIE SICH DIE INFORMATION IM BAUWESEN ENTWICKELT HAT **2**

KAPITEL 2.1. ENTWICKLUNG DER DATENNUTZUNG IN DER BAUINDUSTRIE 3

Die Geburt des Datenzeitalters im Bauwesen 3

Von Ton und Papyrus zur Digitaltechnik 4

Prozess als Werkzeug für datengesteuerte Erfahrungen 5

Digitalisierung von Bauprozessinformationen 7

KAPITEL 2.2. TECHNOLOGIEN UND MANAGEMENTSYSTEME IM MODERNEN BAUWESEN 11

Die digitale Revolution und das Aufkommen von modularen MRP/ERP -Systemen 11

Datenverwaltungssysteme: vom Data Mining bis zu geschäftlichen Herausforderungen 13

Unternehmensmyzel: Wie Daten mit Geschäftsprozessen verbunden sind 17

KAPITEL 2.3. DIE DIGITALE REVOLUTION UND DIE EXPLOSION DER DATEN 20

Der Beginn des Datenbooms als eine evolutionäre Welle 20

Die Menge der in einem modernen Unternehmen erzeugten Daten 22

Die Kosten der Datenspeicherung: der wirtschaftliche Aspekt 23

Die Grenzen der Datenakkumulation: von der Masse zur Bedeutung 25

Nächste Schritte: von der Datentheorie zum praktischen Wandel 27

III TEIL WIE DAS BAUGEWERBE IM DATENCHAOS ERTRINKT **28**

KAPITEL 3.1. DATENFRAGMENTIERUNG UND -SILOS 29

Je mehr Werkzeuge, desto effizienter das Unternehmen? 29

Datensilos und ihre Auswirkungen auf die Unternehmensleistung 31

Doppelarbeit und mangelnde Datenqualität als Folge der Uneinheitlichkeit 34

HiPPO oder die Gefahr von Meinungen bei der Entscheidungsfindung 36

Kontinuierliche Zunahme der Komplexität und Dynamik der Geschäftsprozesse 39

Die vierte industrielle Revolution (Industrie 4.0) und die fünfte industrielle Revolution (Industrie 5.0) im Bauwesen 42

KAPITEL 3.2. CHAOS IN ORDNUNG BRINGEN UND KOMPLEXITÄT REDUZIEREN 45

Redundanter Code und geschlossene Systeme als Hindernis für Produktivitätssteigerungen 45

Von Silos zu einem einzigen Data Warehouse 47

Integrierte Speichersysteme ermöglichen den Übergang zu KI-Agenten 49
Von der Datenerfassung zur Entscheidungsfindung: der Weg zur Automatisierung 51
Nächste Schritte: Das Chaos in ein überschaubares System verwandeln 53
IV TEIL DATENRAHMEN IN DEN GESCHÄFTSPROZESSEN DES BAUWESENS 55
KAPITEL 4.1. DATENTYPEN IN DER KONSTRUKTION 56
Die wichtigsten Datenarten im Baugewerbe 56
Strukturierte Daten 60
Relationale Datenbanken RDBMS und SQL-Abfragesprache 62
SQL - Abfragen in Datenbanken und neue Trends 64
Unstrukturierte Daten 66
Textdaten: zwischen unstrukturiertem Chaos und strukturiertem Chaos y 67
Semi-strukturierte und lose strukturierte Daten 68
Geometrische Daten und ihre Anwendung 69
CAD Daten: vom Entwurf bis zur Datenspeicherung 72
Das Aufkommen des BIM-Konzepts (BOM) und die Verwendung von CAD in den Prozessen 75
KAPITEL 4.2. HARMONISIERUNG UND STRUKTURIERUNG DER DATEN 81
Befüllung von Systemen mit Daten in der Bauindustrie 81
Datentransformation: die entscheidende Grundlage der modernen Unternehmensanalyse 83
Datenmodelle: Beziehungen in den Daten und Beziehungen zwischen Elementen 87
Proprietäre Formate und ihre Auswirkungen auf digitale Prozesse 91
Offene Formate verändern den Ansatz der Digitalisierung 95
Paradigmenwechsel: Open Source als Ende der Ära der Vorherrschaft der Softwarehersteller 96
Strukturierte offene Daten: die Grundlage der digitalen Transformation 99
KAPITEL 4.3. LLM UND IHRE ROLLE BEI DER DATENVERARBEITUNG UND DEN GESCHÄFTSPROZESSEN 102
LLM Chaträume: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok zur Automatisierung von Datenprozessen 102
Große LLM-Sprachmodelle: wie es funktioniert 103
Nutzung lokaler LLMs für sensible Unternehmensdaten 106
Volle Kontrolle über AI im Unternehmen und wie Sie Ihr eigenes LLM einsetzen können 108
RAG: Intelligente LLM -Assistenten mit Zugang zu Unternehmensdaten 110
KAPITEL 4.4. IDE MIT LLM-UNTERSTÜTZUNG UND ZUKÜNFTIGE PROGRAMMÄNDERUNGEN 112
Auswahl eines IDE: von LLM Experimenten bis zu Unternehmenslösungen 112

IDE mit LLM-Unterstützung und zukünftige Programmieränderungen 114
Python Pandas: ein unverzichtbares Werkzeug für die Arbeit mit Daten 115
DataFrame: universelles tabellarisches Datenformat 119
Nächste Schritte: Aufbau eines nachhaltigen Datenrahmens 122
V TEIL DATENQUALITÄT: ORGANISATION, STRUKTURIERUNG, MODELLIERUNG 124
KAPITEL 5.1. UMWANDLUNG VON DATEN IN EINE STRUKTURIERTE FORM 125
Lernen, wie man Dokumente, PDF, Bilder und Texte in strukturierte Formate umwandelt .. 125
Beispiel für die Umwandlung eines PDF -Dokuments in eine Tabelle 126
Konvertierung von JPEG- und PNG-Bildern in eine strukturierte Form 130
Umwandlung von Textdaten in eine strukturierte Form 133
Umwandlung von CAD-Daten (BIM) in eine strukturierte Form 136
Die Anbieter von CAD Lösungen bewegen sich in Richtung strukturierter Daten 141
KAPITEL 5.2. KLASSIFIZIERUNG UND INTEGRATION: EINE GEMEINSAME SPRACHE FÜR BAUDATEN 144
Die Geschwindigkeit der Entscheidungsfindung hängt von der Qualität der Daten ab 144
Standardisierung und Integration von Daten 145
Digitale Interoperabilität beginnt mit Anforderungen 148
Eine gemeinsame Sprache der Konstruktion: die Rolle von Klassifikatoren bei der digitalen Transformation 150
Masterformat, OmniClass, Uniclass und CoClass: die Entwicklung der Klassifikationssysteme 153
KAPITEL 5.3. DATENMODELLIERUNG UND KOMPETENZZENTRUM 158
Datenmodellierung: konzeptionelles, logisches und physisches Modell 158
Praktische Datenmodellierung im Kontext des Bauwesens 162
Erstellen einer Datenbank mit LLM 164
Exzellenzzentrum (CoE) für Datenmodellierung 166
KAPITEL 5.4. SYSTEMATISIERUNG DER ANFORDERUNGEN UND VALIDIERUNG DER INFORMATIONEN 170
Erfassung und Analyse von Anforderungen: Umwandlung von Mitteilungen in strukturierte Daten 170
Prozessablaufdiagramme und die Wirksamkeit von konzeptionellen Rahmen 174
Strukturierte Anforderungen und reguläre Ausdrücke RegEx 176
Datenerhebung für den Überprüfungsprozess 181
Überprüfung der Daten von und Ergebnisse der Überprüfung 183
Visualisierung der Ergebnisse der Überprüfung 188
Vergleich der Datenqualitätsprüfungen mit den Bedürfnissen des menschlichen Lebens .. 190

Nächste Schritte: Umwandlung von Daten in genaue Berechnungen und Pläne 192

VI TEIL KOSTEN- UND ZEITBERECHNUNGEN: EINBEZIEHUNG VON DATEN IN DEN BAUPROZESS 194

KAPITEL 6.1. KOSTENBERECHNUNGEN UND -VORANSCHLÄGE FÜR BAUPROJEKTE 195

Grundlagen der Konstruktion: Schätzung von Menge, Kosten und Zeit 195

Methoden zur Berechnung der geschätzten Kosten der Projekte 196

Ressourcenbasierte Methode Kalkulation und Kostenvoranschlag im Bauwesen 197

Datenbank der Bauressourcen: Katalog der Baumaterialien und -arbeiten 198

Zusammenstellung von Kalkulationen und Kalkulation von Arbeiten auf der Grundlage der Ressourcenbasis 199

Endgültige Projektkalkulation: von Schätzungen zu Budgets 204

KAPITEL 6.2. MENGENERMITTLUNG UND AUTOMATISCHE ERSTELLUNG VON KOSTENVORANSCHLÄGEN UND ZEITPLÄNEN 208

Übergang von 3D zu 4D und 5D: Verwendung volumetrischer und quantitativer Parameter 208

5D-Attribute und Abrufen von Attributvolumen aus CAD 208

QTO Quantity Take-Off: Gruppierung von Projektdaten nach Attributen 212

QTO-Automatisierung mit LLM und strukturierten Daten 217

QTO Berechnung des gesamten Projekts mit Hilfe von Gruppenregeln aus einem Excel-Tabellenblatt 221

KAPITEL 6.3. 4D, 6D -8D UND BERECHNUNG DER KOHLENDIOXIDEMISSIONEN CO_2 227

4D-Modell: Integration der Zeit in die Baukostenschätzung 227

Bauzeitenplan und seine Automatisierung auf der Grundlage von Kalkulationsdaten 228

Erweiterte Attributsebenen 6D -8D: von der Energieeffizienz bis zur Sicherheitsgarantie ... 230

Schätzung von CO_2 und Berechnung der Kohlendioxidemissionen von Bauprojekten 233

KAPITEL 6.4. BAU ERP UND PMIS SYSTEME 237

Bau-ERP-Systeme am Beispiel von Kalkulationen und Kostenvoranschlägen 237

PMIS: Mittler zwischen ERP und der Baustelle 242

Spekulation, Profit, Insellage und mangelnde Transparenz bei ERP und PMIS 243

Das Ende der Ära des geschlossenen ERP /PMIS: die Bauindustrie braucht neue Ansätze 246

Nächste Schritte: effiziente Nutzung von Projektdaten 248

VII TEIL CAD UND BIM: MARKETING, REALITÄT UND DIE ZUKUNFT DER KONSTRUKTIONSDATEN IM BAUWESEN 251

KAPITEL 7.1. ENTSTEHUNG VON BIM- KONZEPTEN IN DER BAUINDUSTRIE 252

Geschichte des Aufkommens von BIM und Open BIM als Marketingkonzepte von CAD-Anbietern 252

Die Realität von BIM: anstelle von integrierten Datenbanken - geschlossene modulare Systeme 255
Das Aufkommen des offenen Formats IFC in der Bauindustrie 257
IFC Formatproblem je nach geometrischem Kern 259
Erscheinungsbild im Aufbau des Themas Semantik und Ontologie 262
Warum semantische Technologien die Erwartungen in der Bauindustrie nicht erfüllen 264
KAPITEL 7.2. GESCHLOSSENE PROJEKTFORMATE UND INTEROPERABILITÄTSPROBLEME 268
Geschlossene Daten und sinkende Produktivität: die Sackgasse der CAD-Branche (BIM) 268
Der Mythos der Interoperabilität zwischen CAD-Systemen 270
Weiter zu USD und granulare Daten 274
KAPITEL 7.3. GEOMETRIE IM BAUWESEN: VON LINIEN ZU KUBIKMETERN 278
Wenn Linien zu Geld werden oder warum Bauherren Geometrie brauchen 278
Von Linien zu Volumen: Wie Fläche und Volumen zu Daten werden 278
Umstellung auf MESH, USD und Polygone: Verwendung der Tesselierung für Geometrie 281
LOD, LOI, LOMD - eindeutige Klassifizierung von Details in CAD (BIM) 282
Neue CAD-Normen (BIM) - AIA, BEP, IDS, LOD, COBie 285
KAPITEL 7.4. ENTWURFSPARAMETRIERUNG UND NUTZUNG VON LLM FÜR CAD-BETRIEB 290
Die Illusion der Einzigartigkeit von CAD-Daten (BIM): der Weg zu Analytik und offenen Formaten 290
Entwurf durch Parameter: die Zukunft von CAD und BIM 293
Entstehung von LLM in Design-CAD Datenverarbeitungsprozessen 296
Automatisierte Analyse von DWG -Dateien mit LLM und Pandas 299
Nächste Schritte: Übergang von geschlossenen Formaten zu offenen Daten 305
VIII TEIL DATENGESTEUERTE ENTSCHEIDUNGSFINDUNG, ANALYTIK, AUTOMATISIERUNG UND MASCHINELLES LERNEN 308
KAPITEL 8.1. DATENANALYTIK UND DATENGESTÜTZTE ENTSCHEIDUNGSFINDUNG 309
Daten als Ressource für die Entscheidungsfindung 309
Visualisierung von Daten: der Schlüssel zum Verständnis und zur Entscheidungsfindung 312
KPIs und ROI 314
Dashboards und Dashboards: Visualisierung von Metriken für ein effizientes Management 316
Datenanalyse und die Kunst, Fragen zu stellen 318
KAPITEL 8.2. DATENFLUSS OHNE MANUELLEN AUFWAND: WARUM ETL BENÖTIGT WIRD 321
ETL-Automatisierung: Kostensenkung und Beschleunigung der Datenverarbeitung 321

ETL Extract: Datenerfassung 325
ETL Transform: Anwendung von Validierungs- und Transformationsregeln 328
ETL Load: Visualisieren Sie Ergebnisse in Tabellen und Diagrammen 330
ETL Laden: Automatische Erstellung von PDF-Dokumenten 335
ETL Laden: automatische Erzeugung von Dokumenten aus FPDF 337
ETL Laden: Berichte und Laden in andere Systeme 341
ETL mit LLM: Visualisierung von Daten aus PDF -Dokumenten 342
KAPITEL 8.3. AUTOMATISCHER ETL FÖRDERER (PIPELINE) 347
Pipeline: Automatische ETL 347
Pipeline -ETL Datenvalidierungsprozess mit LLM 351
Pipeline -ETL: Überprüfung der Daten und Informationen der Projektelemente in CAD (BIM) 353
KAPITEL 8.4. ORCHESTRIERUNG VON ETL UND ARBEITSABLÄUFEN: PRAKTISCHE LÖSUNGEN 359
DAG und Apache Airflow: Automatisierung und Orchestrierung von Arbeitsabläufen 359
Apache Airflow: praktische Anwendung zur ETL-Automatisierung 360
Apache NiFi für Routing und Datenkonvertierung 364
n8n Low-Code, No-Code Prozessorchestrierung 365
Nächste Schritte: Umstellung von manuellen Vorgängen auf analytikbasierte Lösungen 368
IX TEIL SPEICHERUNG UND VERWALTUNG VON DATEN IM BAUWESEN 370
KAPITEL 9.1. DATENINFRASTRUKTUR: VON SPEICHERFORMATEN ZU DIGITALEN REPOSITORIES 371
Datenatome: die Grundlage eines effektiven Informationsmanagements 371
Informationsspeicherung: Dateien oder Daten 372
Speicherung großer Datenmengen: Analyse der gängigen Formate und ihrer Effektivität 374
Optimierung der Speicherung mit Apache Parquet 377
DWH: Data Warehouse data warehouses 379
Data Lake - Entwicklung von ETL zu ELT: von der traditionellen Bereinigung zur flexiblen Verarbeitung 381
Data Lakehouse Architektur: Synergie von Lagerhäusern und Data Lakes 383
CDE, PMIS, ERP oder DWH und Data Lake 385
KAPITEL 9.2. VERWALTUNG VON DATENLAGERN UND CHAOSVERMEIDUNG 389
Vektordatenbanken und die Bounding Box 389
Data Governance, Data Minimalism und Data Swamp 392
DataOps und VectorOps: neue Datenstandards 395

Die nächsten Schritte: von der chaotischen Speicherung zur strukturierten Speicherung ..397

X TEIL BIG DATA, MASCHINELLES LERNEN UND VORHERSAGEN 399

KAPITEL 10.1. BIG DATA UND DEREN ANALYSE 400

Big Data im Bauwesen: von der Intuition zur Vorhersagbarkeit 400

Hinterfragen der Machbarkeit von Big Data: Korrelation, Statistik und Datenstichproben ..401

Big Data: Analyse von Daten aus dem Datensatz der Millionen Baugenehmigungen in San Francisco 404

Beispiel für Big Data auf der Grundlage von CAD-Daten (BIM) 410

IoT Internet der Dinge und intelligente Verträge 414

KAPITEL 10.2. MASCHINELLES LERNEN UND VORHERSAGEN 418

Maschinelles Lernen und künstliche Intelligenz werden die Art und Weise, wie wir bauen, verändern 418

Von der subjektiven Einschätzung zur statistischen Prognose 420

Titanic-Datensatz: Hallo Welt in der Welt der Analytik Daten und Big Data 422

Maschinelles Lernen in Aktion: von Titanic-Passagieren zum Projektmanagement 428

Vorhersagen und Prognosen auf der Grundlage historischer Daten 432

Schlüsselkonzepte des maschinellen Lernens 434

KAPITEL 10.3. VORHERSAGE VON KOSTEN UND TERMINEN DURCH MASCHINELLES LERNEN 437

Ein Beispiel für den Einsatz von maschinellem Lernen zur Ermittlung von Projektkosten und -zeitplänen 437

Vorhersage von Projektkosten und -zeiten mittels linearer Regression 439

Vorhersage von Projektkosten und -zeiten mit dem K-Nächste-Nachbar-Algorithmus (k-NN) 442

Nächste Schritte: von der Speicherung zur Analyse und Vorhersage 446

XI TEIL DIE BAUINDUSTRIE IM DIGITALEN ZEITALTER. CHANCEN UND HERAUSFORDERUNGEN 449

KAPITEL 11.1. ÜBERLEBENSSTRATEGIEN: AUFBAU VON WETTBEWERBSVORTEILEN 450

Korrelationen statt Berechnungen: die Zukunft der Bauanalytik 450

Datengesteuerter Ansatz im Bauwesen: eine neue Ebene der Infrastruktur 453

Das digitale Büro der nächsten Generation: Wie KI den Arbeitsbereich verändert 455

Offene Daten und Uberisierung sind eine Bedrohung für das bestehende Baugeschäft 457

Ungelöste Probleme der Uberisierung als letzte Chance, Zeit für Transformation zu nutzen 460

KAPITEL 11.2. EIN PRAKTISCHER LEITFADEN ZUR UMSETZUNG EINES DATENGESTÜTZTEN ANSATZES 465

Von der Theorie zur Praxis: Ein Fahrplan für die digitale Transformation im Bauwesen 465

Legen des digitalen Fundaments: 1-5 Schritte zur digitalen Reife 467
Das Potenzial von Daten freisetzen: 5-10 Schritte zur digitalen Reife 471
Transformationsfahrplan: vom Chaos zum datengesteuerten Unternehmen..... 478
Bauen in der Industrie 5.0: Wie man Geld verdient, wenn man sich nicht mehr verstecken kann..... 480
SCHLUSSFOLGERUNG 484
ÜBER DEN AUTOR..... 487
WEITERE BEZIEHUNG 488
ÜBERSETZUNGSKOMMENTAR 488
ANDERE FÄHIGKEITEN UND KONZEPTE..... 489
GLOSSAR 493
LISTE DER REFERENZEN UND ONLINE-MATERIALIEN 500
THEMA-INDEX..... 516

MAXIMALER KOMFORT MIT DER GEDRUCKTEN VERSION

Sie sind im Besitz einer kostenlosen digitalen Version von **Data-Driven Construction**. Für bequemeres Arbeiten und schnellen Zugriff auf die Materialien empfehlen wir Ihnen, die **gedruckte Ausgabe zu verwenden**:

- **Immer zur Hand**: Das Buch im gedruckten Format wird zu einem zuverlässigen Arbeitsinstrument, das es Ihnen ermöglicht, in jeder Arbeitssituation schnell die richtigen Visualisierungen und Diagramme zu finden und zu verwenden
- **Hochwertige Illustrationen**: Alle Bilder und Grafiken in der gedruckten Ausgabe werden in höchster Qualität präsentiert
- **Schneller Zugriff auf Informationen**: einfache Navigation, die Möglichkeit, Notizen zu machen, Lesezeichen zu setzen und überall mit dem Buch zu arbeiten.

Mit dem Kauf der gedruckten Vollversion des Buches erhalten Sie ein praktisches Hilfsmittel für die bequeme und effiziente Arbeit mit Informationen: die Möglichkeit, bei alltäglichen Aufgaben schnell auf Anschauungsmaterial zurückzugreifen, die notwendigen Schemata schnell zu finden und Notizen zu machen. Darüber hinaus unterstützen Sie mit Ihrem Kauf die Verbreitung von offenem Wissen.

Sie können eine gedruckte Version des Buches bestellen unter: datadrivenconstruction.io/books

I TEIL

VON TONTAFELN ZUR DIGITALEN REVOLUTION: WIE SICH DIE INFORMATION IM BAUWESEN ENTWICKELT HAT

Der erste Teil des Buches untersucht die historische Entwicklung des Datenmanagements in der Bauindustrie, von primitiven Aufzeichnungen auf physischen Medien bis zu modernen digitalen Ökosystemen. Er analysiert den Wandel der Informationsmanagement-Technologien, das Aufkommen von ERP -Systemen und die Auswirkungen der Datenfragmentierung auf die Effizienz der Geschäftsprozesse. Besonderes Augenmerk wird auf den Prozess der Digitalisierung von Informationen und die wachsende Bedeutung objektiver Analysen anstelle subjektiver Expertenurteile gelegt. Das exponentielle Wachstum des Informationsvolumens, mit dem die moderne Bauindustrie konfrontiert ist, und die damit verbundenen Herausforderungen für Unternehmenssysteme werden eingehend untersucht. Die Positionierung der Bauindustrie im Kontext der vierten und fünften industriellen Revolution wird ebenso untersucht wie das Potenzial des Einsatzes künstlicher Intelligenz und datenzentrierter Ansätze zur Schaffung nachhaltiger Wettbewerbsvorteile.

KAPITEL 1.1.
ENTWICKLUNG DER DATENNUTZUNG IN DER BAUINDUSTRIE

Die Geburt des Datenzeitalters im Bauwesen

Vor etwa 10 000 Jahren, in der Jungsteinzeit, vollzog die Menschheit einen revolutionären Wandel in ihrer Entwicklung, indem sie die nomadische Lebensweise zugunsten einer sesshaften Lebensweise aufgab, was zur Entstehung der ersten primitiven Gebäude aus Lehm, Holz und Stein führte [6]. Von diesem Moment an beginnt die Geschichte des Bauwesens.

Mit der Entwicklung der Zivilisationen wurde die Architektur immer komplexer, was zu den ersten rituellen Tempeln und öffentlichen Gebäuden führte. Die zunehmende Komplexität der architektonischen Entwürfe erforderte von den Ingenieuren und Managern der Antike die Erstellung der ersten Aufzeichnungen und Berechnungen. Die ersten Aufzeichnungen auf Tontafeln und Papyri enthielten oft eine Beschreibung der Logik, die hinter der Berechnung der benötigten Baumaterialien, ihrer Kosten und der Berechnung der Bezahlung für die geleistete Arbeit stand [7]. Damit begann die Ära der Datennutzung im Bauwesen - lange vor dem Aufkommen der modernen digitalen Technologien (Abb. 1.1-1).

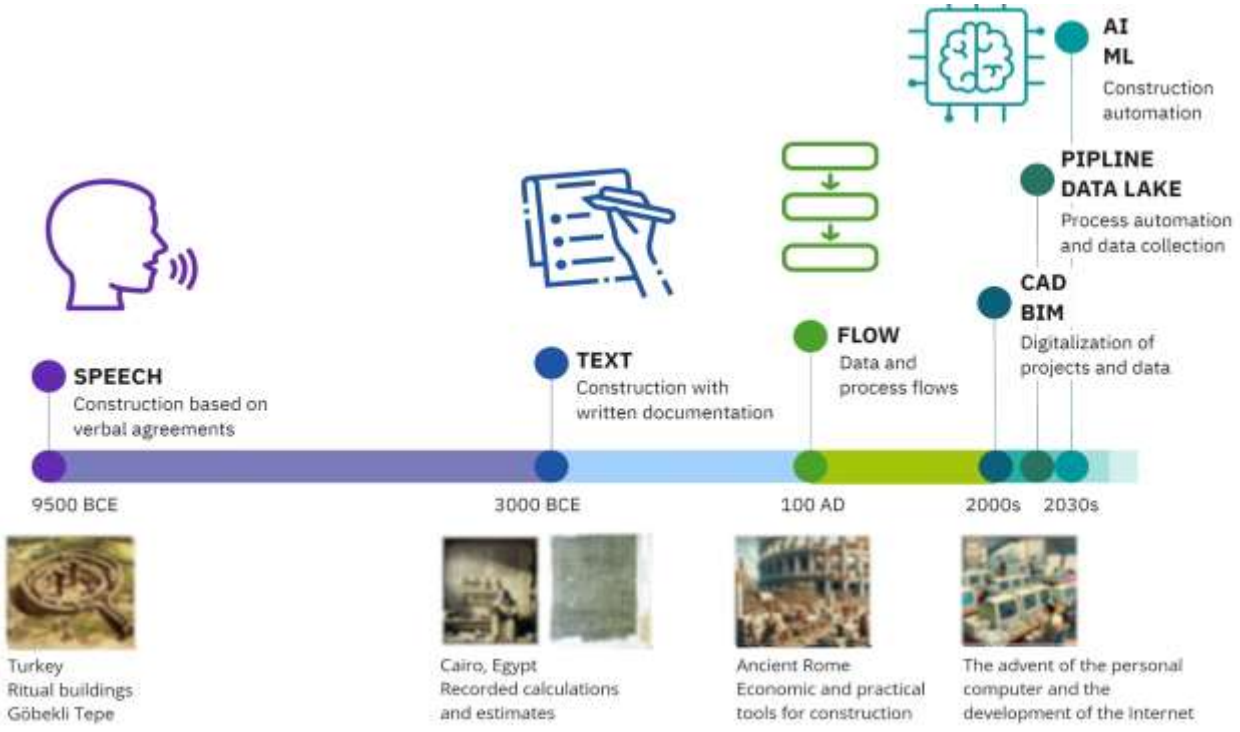


Abb. 1.1-1 Chronologie der Entwicklung der Informationstechnologie im Bauwesen: von der verbalen Information zur künstlichen Intelligenz.

Von Ton und Papyrus zur Digitaltechnik

Die ersten dokumentarischen Belege im Bauwesen stammen aus der Zeit des Pyramidenbaus, etwa 3000-4000 v. Chr. [7]. Seitdem hat das Führen von schriftlichen Aufzeichnungen den Fortschritt im Bauwesen erleichtert und begleitet und die Anhäufung und Systematisierung von Wissen ermöglicht, das in den folgenden 10 000 Jahren zu bedeutenden Innovationen in Bauverfahren und Architektur führte.

Die Verwendung der ersten physischen Medien im Bauwesen, wie Tontafeln, Papyrus von vor Tausenden von Jahren (Abb. 1.1-2) oder "A0"-Papier in den 1980er Jahren, zur Aufzeichnung von Daten war ursprünglich nicht dazu gedacht, diese Informationen auf neue Projekte anzuwenden. Der Hauptzweck solcher Aufzeichnungen bestand darin, den aktuellen Stand des Projekts zu dokumentieren, einschließlich der Berechnung der benötigten Materialien und der Kosten für die Arbeiten. Auch in der heutigen Welt garantiert die Verfügbarkeit digitaler Konstruktionsdaten und -modelle nicht immer deren Anwendung bei künftigen Projekten und dient oft hauptsächlich als Information für aktuelle Berechnungen der benötigten Materialien und Baukosten.

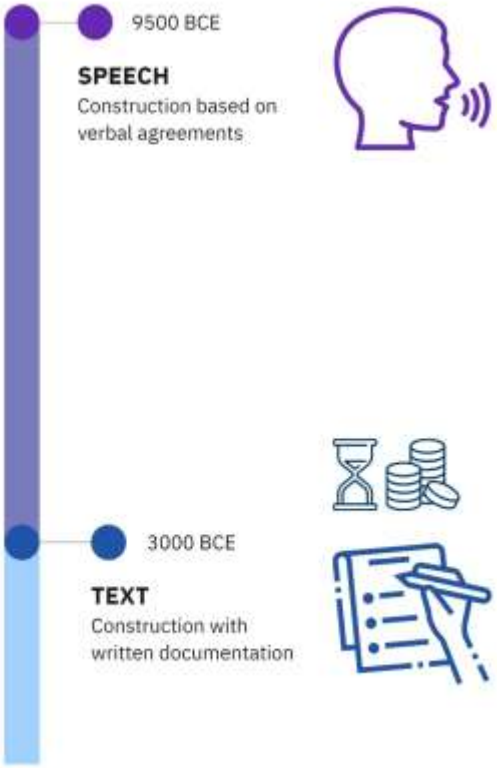


Abb. 1.1-2 Ein Papyrus aus dem 3. Jahrhundert v. Chr., in dem die Kosten für die Bemalung verschiedener Fenstertypen in einem königlichen Palast mit der Enkaustik-Technik beschrieben werden.

> Die Menschheit hat etwa 5.000 Jahre gebraucht, um von mündlichen Gesprächen zu schriftlichen Dokumenten im Bauprojektmanagement überzugehen, und ebenso lange, um von Papier zu digitalen Daten als Hauptressource für Planung und Kontrolle überzugehen.

So wie die Entwicklung der Handels- und Währungsbeziehungen die Entstehung der Schrift und der ersten Anwälte zur Beilegung von Streitigkeiten förderte, so führten die ersten Aufzeichnungen über Materialkosten und Arbeitsumfänge im Bauwesen zur Entstehung der ersten Manager im Baugewerbe, zu deren Aufgaben die Dokumentation, Überwachung und Verantwortung für wichtige Informationen über Projektzeitpläne und -kosten gehörten.

Heute spielen Daten eine viel wichtigere Rolle: Sie zeichnen nicht nur die getroffenen Entscheidungen auf, sondern werden auch zu einem Werkzeug für die Vorhersage und Modellierung der Zukunft. Dies ist die Grundlage, auf der der moderne Prozessansatz im Projektmanagement aufbaut - die Umwandlung gesammelter Erfahrungen in ein Entscheidungssystem, das auf strukturierten und überprüfbaren Daten basiert.

Prozess als Werkzeug für datengesteuerte Erfahrungen

Das Herzstück eines jeden Prozesses ist die Umwandlung von Erfahrungen aus der Vergangenheit in ein Instrument zur Planung der Zukunft. Erfahrung im modernen Sinne ist ein strukturierter Datensatz, der analysiert werden kann, um fundierte Vorhersagen zu treffen.

> Historische Daten bilden die Grundlage für Prognosen, da sie die Ergebnisse der geleisteten Arbeit klar aufzeigen und einen Einblick in die Faktoren geben, die diese Ergebnisse beeinflussen.

Nehmen wir ein konkretes Beispiel aus dem monolithischen Bau: Bei der Planung des Zeitplans für die Arbeiten werden in der Regel die Betonmenge, die Komplexität des Bauwerks und die Wetterbedingungen berücksichtigt. Nehmen wir an, dass ein bestimmter Bauleiter oder die historischen Daten des Unternehmens für die letzten drei Jahre (2023-2025) zeigen, dass das Gießen eines 200 m² großen monolithischen Bauwerks bei regnerischem Wetter zwischen 4,5 und 6 Tagen dauerte (Abb. 1.1-3). Diese kumulierten Statistiken bilden die Grundlage für die Vorhersage von Durchlaufzeiten und die Kalkulation von Ressourcen bei der Planung ähnlicher Arbeiten in zukünftigen Projekten. Auf der Grundlage dieser historischen Daten kann der Polier oder der Kalkulator eine fundierte, auf Erfahrung basierende Vorhersage über die Zeit treffen, die für die Ausführung ähnlicher Arbeiten im Jahr 2026 unter ähnlichen Bedingungen benötigt wird.

In diesem Fall der Zeitschätzungen fungiert der analytische Prozess als Mechanismus zur Umwandlung disparater Daten in strukturierte Erfahrungen und dann in ein präzises Planungsinstrument. Daten und Prozesse sind ein einziges Ökosystem, in dem das eine nicht ohne das andere existieren kann.

> *Zählen Sie, was zählbar ist, messen Sie, was messbar ist, und machen Sie das, was nicht messbar ist, messbar [8].*
>
> - Galileo Galilei

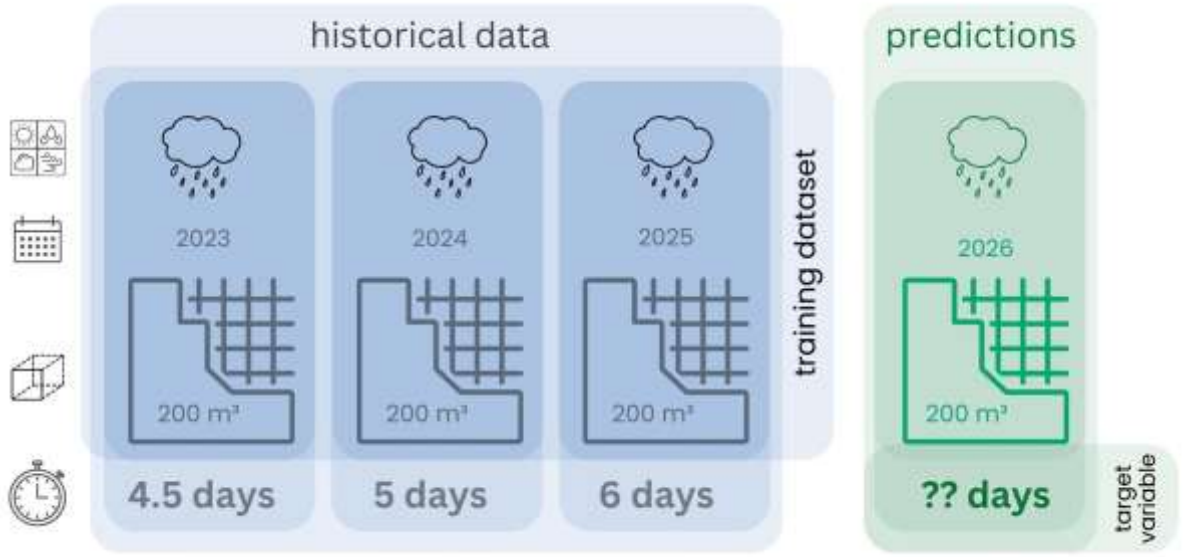


Abb. 1.1-3 Historische Daten dienen als Trainingsdatensatz für die Vorhersage eines der Werte in der Zukunft.

In der heutigen Unternehmenslandschaft wird die Datenanalyse zu einer entscheidenden Komponente für effektives Projektmanagement, Prozessoptimierung und strategische Entscheidungsfindung. Die Baubranche beherrscht nach und nach vier Schlüsselebenen der Analytik, die jeweils eine spezifische Frage beantworten und einzigartige Vorteile bieten (Abb. 1.1-4):

- **Deskriptive Analytik** - beantwortet die Frage "*Was ist passiert*?" und liefert historische Daten und Berichte über vergangene Ereignisse und Ergebnisse: In den letzten drei Jahren (2023-2025) dauerte es zwischen 4,5 und 6 Tagen, um eine 200 m² große monolithische Struktur bei Regenwetter zu gießen.
- **Diagnostische Analyse** - beantwortet die Frage "*Warum ist das passiert?", indem* die Ursachen der Probleme ermittelt werden: Die Analyse zeigt, dass sich die Betonierzeit des monolithischen Bauwerks aufgrund des regnerischen Wetters verlängert hat, was den Aushärtungsprozess des Betons verlangsamt hat.
- **Predictive Analytics** - vorausschauende Vorhersage möglicher Risiken und Vorlaufzeiten durch Beantwortung der Frage "*Was wird passieren*?": Auf der Grundlage historischer Daten wird vorhergesagt, dass das Schütten eines ähnlichen monolithischen Bauwerks von 200 m² bei Regenwetter im Jahr 2026 etwa 5,5 Tage dauern wird, wobei alle bekannten Faktoren und Trends berücksichtigt werden.
- **Die präskriptive Analytik** gibt automatische Empfehlungen und beantwortet die Frage "*Was ist zu tun*?", so dass die Unternehmen die optimalen Maßnahmen auswählen können: Um die Arbeiten zu optimieren, wird beispielsweise empfohlen, spezielle Zusatzstoffe zu verwenden, um die Aushärtung des Betons bei hoher Luftfeuchtigkeit zu beschleunigen; den Betoneinbau für die Zeiträume mit der geringsten Niederschlagswahrscheinlichkeit zu planen; provisorische Unterstände für das Bauwerk einzurichten, die die Arbeitszeit auch bei ungünstigen Wetterbedingungen auf 4 bis 4,5 Tage verkürzen.

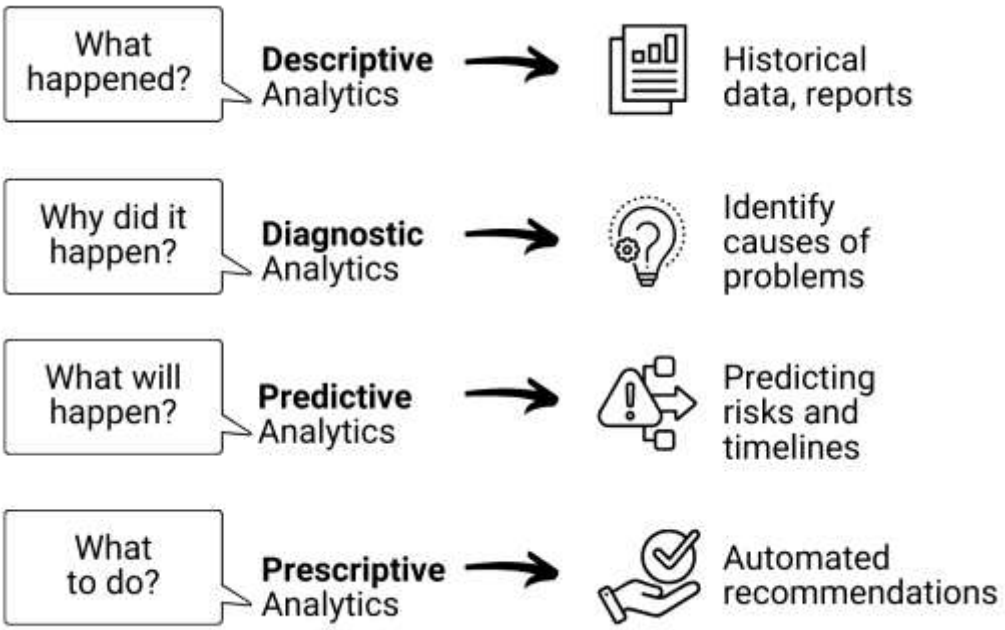


Abb. 1.1-4 Die Haupttypen der Analytik: von der Beschreibung der Vergangenheit bis zur automatisierten Entscheidungsfindung.

Eine umfassende digitale Transformation, die einen Übergang zu Systemanalysen und datengesteuertem Management impliziert, erfordert nicht nur Outsourcing, sondern auch die Bildung eines kompetenten internen Teams. Die wichtigsten Mitglieder eines solchen Teams sollten Produktmanager, Dateningenieure, Analysten und Entwickler sein, die eng mit den Geschäftsbereichen zusammenarbeiten (Abb. 4.3-9). Diese Zusammenarbeit ist notwendig, um intelligente analytische Fragen zu stellen und die Entscheidungsfindung der Unternehmen effektiv zu parametrisieren. In einer Informationsgesellschaft werden Daten nicht nur zu einem Hilfsmittel, sondern zur Grundlage für Prognosen und Optimierung.

Im Bauwesen verändert die digitale Transformation die Art und Weise, wie Anlagen geplant, verwaltet und betrieben werden, grundlegend. Dieser Prozess wird als Digitalisierung von Informationen bezeichnet, bei der alle Aspekte des Bauprozesses in eine für die Analyse geeignete digitale Form gebracht werden.

Digitalisierung von Bauprozessinformationen

Seit Jahrtausenden hat sich die Menge der im Bauwesen aufgezeichneten Informationen kaum verändert, doch in den letzten Jahrzehnten hat sie rapide zugenommen (Abb. 1.1-5).

> Laut der PwC-Studie® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] wurden 90% aller Daten weltweit in den letzten zwei Jahren (Stand: 2015) erstellt. Die meisten Unternehmen nutzen diese Daten jedoch nicht in vollem Umfang, da sie entweder in isolierten Systemen verbleiben oder einfach ohne echte Analyse archiviert werden.

Der Anstieg des Datenvolumens hat sich in den letzten Jahren noch beschleunigt und verdoppelt sich von

15 Zettabyte im Jahr 2015 auf 181 Zettabyte im Jahr 2025 [10]. Jeden Tag werden die Server von Bau- und Planungsunternehmen mit Projektdokumentationen, Arbeitsplänen, Berechnungen und Kalkulationen, Finanzberichten gefüllt. Für 2D/3D -Zeichnungen werden die Formate DWG, DXF und DGN verwendet, und für 3D-Modelle - RVT, NWC, PLN und IFC™. Textdokumente, Tabellen und Präsentationen werden in DOC, XLSX und PPT gespeichert. Neben Videos und Bildern von der Baustelle - in MPG und JPEG - werden auch Echtzeitdaten von IoT-Komponenten, RFID® Tags (Identifikation und Tracking) und BMS Gebäudemanagementsystemen (Überwachung und Steuerung) verwendet

Abb. 1.1-5 Parabolisches Datenwachstum 2010-2025 (basierend auf [10]).

Mit der rasanten Zunahme von Informationen steht die Bauindustrie vor der Notwendigkeit, Daten nicht nur zu sammeln und zu speichern, sondern auch ihre Überprüfung, Validierung, Messbarkeit und analytische Verarbeitung zu gewährleisten. Heute befindet sich die Branche in einer aktiven Phase der Informationsdigitalisierung - der systematischen Umwandlung aller Aspekte der Bautätigkeit in eine digitale Form, die sich für Analyse, Interpretation und Automatisierung eignet.

> Die Digitalisierung von Informationen bedeutet, dass Informationen über alle Einheiten und Elemente eines Bauprojekts und des Bauprozesses selbst - einschließlich derjenigen, die wir früher überhaupt nicht als Informationen betrachteten - in ein Datenformat umgewandelt werden, um die Informationen quantifizierbar und leicht analysierbar zu machen.

Im Kontext des Bauwesens bedeutet dies die Erfassung und Digitalisierung von Informationen über alle Elemente von Projekten und alle Prozesse - von der Bewegung von Maschinen und Menschen auf der Baustelle bis hin zu den Wetter- und Klimabedingungen auf der Baustelle, den aktuellen Materialpreisen und den Zinssätzen der Zentralbanken um analytische Modelle zu erstellen.

> *Wenn Sie das, worüber Sie sprechen, messen und in Zahlen ausdrücken können, dann wissen Sie etwas über dieses Thema. Wenn Sie es aber nicht quantitativ ausdrücken können, ist Ihr Wissen äußerst begrenzt und unbefriedigend. Es mag ein Ausgangspunkt sein, aber es ist nicht das Niveau wahrer wissenschaftlicher Erkenntnis. [11].*
>
> - W. Thomson (Lord Kelvin), 1824-1907, britischer Wissenschaftler

Die Digitalisierung von Informationen geht weit über den traditionellen Ansatz der Informationserfassung hinaus, bei dem nur grundlegende Metriken wie Arbeitsstunden oder tatsächliche Materialkosten erfasst wurden. Heute kann praktisch jedes Ereignis in einen Datenstrom umgewandelt werden, der sich für eine eingehende Analyse mit Hilfe fortschrittlicher Analysetools und maschineller Lerntechniken eignet. Die Baubranche hat einen grundlegenden Wandel von Papierzeichnungen, Excel-Tabellen und mündlichen Anweisungen hin zu digitalen Systemen vollzogen (Abb. 1.2-4), bei denen jedes Element eines Projekts zu einer Datenquelle wird. Sogar die Mitarbeiter - von den Ingenieuren bis zu den Bauarbeitern auf der Baustelle - werden jetzt als eine Sammlung digitaler Variablen und Datensätze betrachtet.

> Laut der KPMG-Studie "Familiar Challenges - New Approaches: Global Construction Survey 2023" entwickeln sich digitale Zwillinge, KI (AI) und Big Data zu den wichtigsten Faktoren für die Verbesserung der Projektrentabilität [2].

Moderne Technologien vereinfachen nicht nur die Informationserfassung und machen sie weitgehend automatisch, sondern senken auch die Kosten der Datenspeicherung radikal. Dies führt dazu, dass Unternehmen von einem selektiven Ansatz abrücken und lieber die gesamte Bandbreite an Informationen für eine spätere Analyse speichern (Abb. 2.1-5), was potenzielle Möglichkeiten für zukünftige Prozessoptimierungen eröffnet.

> Die Digitalisierung von Informationen und die Digitalisierung ermöglichen es, den verborgenen, bisher ungenutzten Wert von Informationen aufzudecken. Wenn sie richtig organisiert sind, können Daten wiederverwendet, neu interpretiert und in neue Dienste und Lösungen integriert werden.

In Zukunft wird die Digitalisierung von Informationen wahrscheinlich zu einer vollständigen Automatisierung des Dokumentenmanagements, zur Einführung selbstverwalteter Bauprozesse und zum Entstehen neuer Berufe führen - Baudatenanalysten, KI-Projektmanagementexperten und Digitalingenieure. Bauprojekte werden zu dynamischen Informationsquellen, und die Entscheidungsfindung wird nicht auf Intuition oder subjektiver Erfahrung beruhen, sondern auf zuverlässigen und reproduzierbaren digitalen Fakten

> *Informationen sind das Öl des 21. Jahrhunderts, und Analytik ist der Verbrennungsmotor [12].*
>
> - Peter Sondergaard, Senior Vice President, Gartner®

> Laut IoT Analytics 2024 [13] wird erwartet, dass die weltweiten Ausgaben für Datenmanagement und -analyse von 185,5 Mrd. USD im Jahr 2023 auf 513,3 Mrd. USD im Jahr 2030 drastisch ansteigen werden, was einer durchschnittlichen jährlichen Wachstumsrate von 16% entspricht. Allerdings wachsen nicht alle Komponenten in gleichem Maße: Die Analytik wächst schnell, während sich das Wachstum bei der Speicherung verlangsamt. Die Analytik wird das schnellste Wachstum im Datenmanagement-Ökosystem aufweisen: Es wird prognostiziert, dass dieser Bereich von 60,6 Milliarden US-Dollar im Jahr 2023 auf 227,9 Milliarden US-Dollar im Jahr 2030 ansteigen wird, was einer durchschnittlichen jährlichen Wachstumsrate von 27% entspricht.

Mit der zunehmenden Digitalisierung von Informationen und dem rasanten Wachstum des Informationsvolumens steht das Management von Bauprojekten und Unternehmen vor der Notwendigkeit, vielfältige, oft heterogene Daten systematisch zu speichern, zu analysieren und zu verarbeiten. Als Reaktion auf diese Herausforderung begann die Branche ab Mitte der 1990er Jahre mit einer massiven Umstellung auf die elektronische Erstellung, Speicherung und Verwaltung von Unterlagen - von Tabellenkalkulationen und Konstruktionsberechnungen bis hin zu Zeichnungen und Verträgen.

Herkömmliche Papierdokumente, die Unterschriften, physische Aufbewahrung, regelmäßige Überarbeitung und Archivierung in Schränken erfordern, werden allmählich durch digitale Systeme ersetzt, die Daten in strukturierter Form - in Datenbanken spezialisierter Anwendungen - speichern.

KAPITEL 1.2.
TECHNOLOGIEN UND MANAGEMENTSYSTEME IM MODERNEN BAUWESEN

Die digitale Revolution und das Aufkommen von modularen MRP/ERP - Systemen

Die Ära der modernen digitalen Datenspeicherung und -verarbeitung begann mit dem Aufkommen des Magnetbands in den 1950er Jahren, das die Möglichkeit eröffnete, große Mengen an Informationen zu speichern und zu nutzen. Der nächste Durchbruch war das Aufkommen von Diskettenlaufwerken, die das Datenmanagement in der Bauindustrie radikal veränderten.

Mit der Entwicklung von Data Warehousing sind zahlreiche Unternehmen in den Lösungsmarkt eingetreten und haben begonnen, modulare Software zur Erstellung, Speicherung und Verarbeitung von Daten sowie zur Automatisierung von Routineaufgaben zu entwickeln

> Das exponentielle Wachstum von Informationen und Werkzeugen hat zu einem Bedarf an integrierten, modularen Lösungen geführt, die nicht mit einzelnen Dateien arbeiten, sondern den Datenfluss über Prozesse und Projekte hinweg verwalten und kontrollieren.

Die ersten umfassenden Plattform-Tools mussten nicht nur Dokumente speichern, sondern auch alle Änderungsanträge und Vorgänge in Prozessen dokumentieren: wer sie initiierte, welchen Umfang der Antrag hatte und was schließlich als Wert oder Attribut erfasst wurde. Für diese Zwecke wurde ein System benötigt, das genaue Berechnungen und getroffene Entscheidungen nachverfolgen konnte (Abb. 1.2-1). Solche Plattformen waren die ersten MRP (Material Requirements Planning) und ERP (Enterprise Resource Planning) Systeme, die ab den frühen 1990er Jahren populär wurden [14]

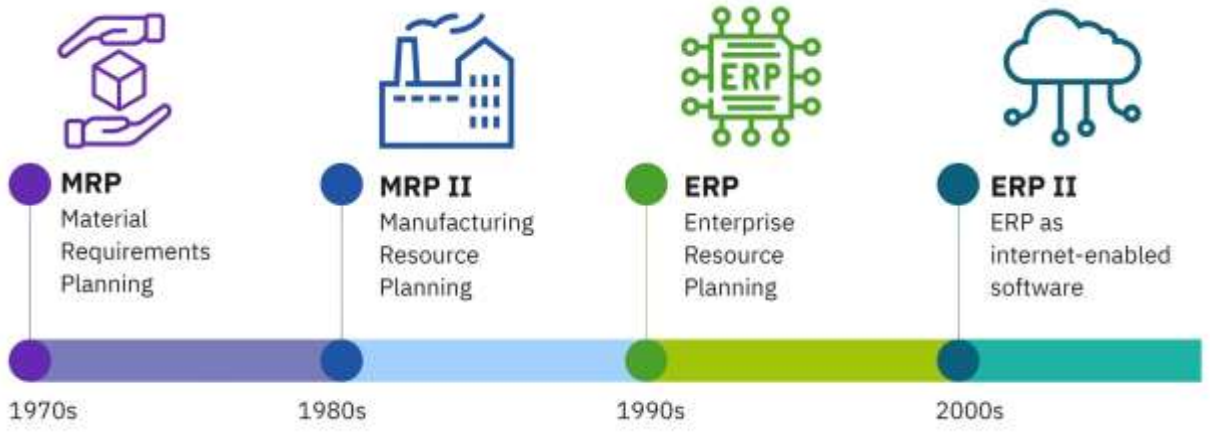


Abbildung 1.2-1 Fortschritte in der Datenspeichertechnologie führten in den 1980er Jahren zur Entstehung von ERP-Systemen.

Die ersten MRP - und ERP - Systeme legten den Grundstein für das Zeitalter der Digitalisierung im Geschäftsprozess- und Bauprojektmanagement. Modulare Systeme, die ursprünglich für die Automatisierung wichtiger Geschäftsprozesse konzipiert waren, wurden schließlich mit zusätzlichen, flexibleren und anpassungsfähigen Softwarelösungen integriert.

Diese zusätzlichen Lösungen wurden für die Datenverarbeitung und die Verwaltung von Projektinhalten entwickelt (Abb. 1.2-2). Sie ersetzten entweder bestimmte Module großer Systeme oder ergänzten diese effektiv, indem sie die Funktionalität des gesamten Systems erweiterten.

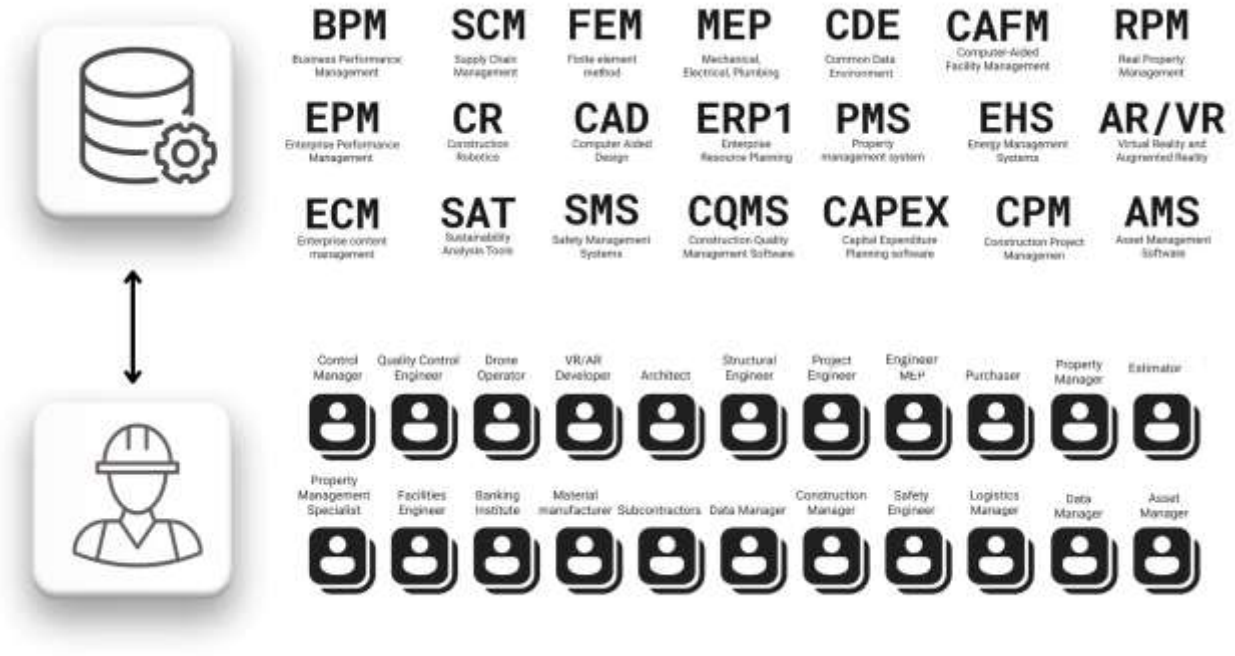


Abb. 1.2-2 Neue Softwarelösungen haben ein Heer von Managern in das Unternehmen gelockt, die den Datenfluss verwalten.

In den letzten Jahrzehnten haben die Unternehmen stark in modulare Systeme [15] investiert, da sie diese als langfristige integrierte Lösungen betrachten.

> Laut dem Software Path-Bericht für 2022 [16] liegt das durchschnittliche Budget pro Benutzer eines ERP -Systems bei 9.000 $. Im Durchschnitt nutzen etwa 26% der Mitarbeiter des Unternehmens solche Systeme. Bei einem Unternehmen mit 100 Benutzern belaufen sich die Gesamtkosten für die ERP-Implementierung also auf etwa 900.000 $.

Investitionen in proprietäre, geschlossene, modulare Lösungen sind vor dem Hintergrund der rasanten Entwicklung moderner, flexibler und offener Technologien immer weniger gerechtfertigt. Wenn solche Investitionen bereits getätigt wurden, ist es wichtig, die Rolle der bestehenden Systeme objektiv neu zu bewerten: ob sie langfristig unverzichtbar bleiben, oder ob ihre Funktionen überarbeitet und effizienter und transparenter umgesetzt werden können.

Eines der Hauptprobleme der heutigen modularen Datenplattformen besteht darin, dass sie die Datenverwaltung in geschlossenen Anwendungen zentralisieren. Dies hat zur Folge, dass Daten - das Kernvermögen eines Unternehmens - von bestimmten Softwarelösungen abhängig werden und nicht umgekehrt. Dies

schränkt die Wiederverwendung von Informationen ein, erschwert die Migration und verringert die geschäftliche Flexibilität in einer sich schnell verändernden digitalen Landschaft.

Wenn es wahrscheinlich ist, dass die Bedeutung oder Relevanz geschlossener modularer Architekturen in Zukunft abnehmen wird, ist es sinnvoll, die heute anfallenden Kosten als verlorene Kosten zu erkennen und sich auf einen strategischen Wechsel zu einem offeneren, skalierbaren und anpassungsfähigen digitalen Ökosystem zu konzentrieren.

> Proprietäre Software zeichnet sich dadurch aus, dass der Entwickler die ausschließliche Kontrolle über den Quellcode und die Benutzerdaten hat, die im Rahmen der Nutzung solcher Lösungen entstehen. Im Gegensatz zu Open-Source-Software haben die Nutzer keinen Zugriff auf die interne Struktur der Anwendung und können sie nicht unabhängig überprüfen, ändern oder an ihre Bedürfnisse anpassen. Stattdessen müssen sie Lizenzen erwerben, die ihnen das Recht geben, die Software innerhalb der vom Anbieter festgelegten Grenzen zu nutzen.

Ein moderner datenzentrierter Ansatz bietet ein anderes Paradigma: Daten sollten als ein wichtiger strategischer Vermögenswert betrachtet werden - unabhängig, dauerhaft und getrennt von spezifischen Softwarelösungen. Anwendungen wiederum werden zu reinen Datenwerkzeugen, die beliebig ausgetauscht werden können, ohne dass die Gefahr besteht, wichtige Informationen zu verlieren.

Die Entwicklung von ERP- und MRP-Systemen in den 1990er Jahren (Abb. 1.2-1) gab den Unternehmen leistungsfähige Instrumente für das Prozessmanagement an die Hand, hatte aber auch die unbeabsichtigte Folge, dass die Zahl der an der Aufrechterhaltung der Informationsflüsse beteiligten Personen erheblich zunahm. Anstatt betriebliche Aufgaben zu automatisieren und zu vereinfachen, führten diese Systeme häufig zu einem neuen Maß an Komplexität, Bürokratie und Abhängigkeit von internen IT-Ressourcen.

Datenverwaltungssysteme: vom Data Mining bis zu geschäftlichen Herausforderungen

Unternehmen stehen heute vor der Notwendigkeit, mehrere Datenverwaltungssysteme zu integrieren. Die Auswahl von Datenverwaltungssystemen, die kluge Verwaltung dieser Systeme und die Integration unterschiedlicher Datenquellen sind für die Unternehmensleistung von entscheidender Bedeutung.

Mitte der 2020er Jahre gibt es Hunderte (und in großen Bauunternehmen Tausende) von verschiedenen Systemen (Abb. 1.2-3), die harmonisch zusammenarbeiten müssen, damit alle Aspekte des Bauprozesses reibungslos und kohärent ablaufen.

> Laut der Deloitte-Studie aus dem Jahr 2016® "Data-Driven Management in Digital Capital Projects" nutzt der durchschnittliche Bauunternehmer täglich 3,3 Softwareanwendungen, aber nur 1,7 davon sind miteinander integriert [17].

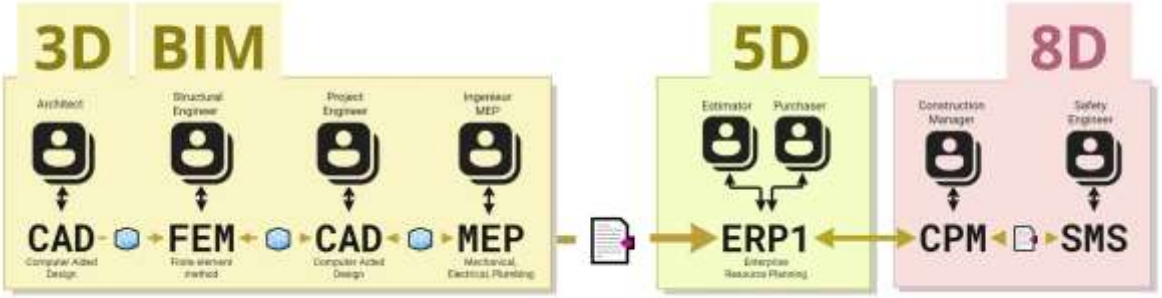


Abb. 1.2-3 Jedes Geschäftssystem erfordert ein professionelles Team und einen verantwortlichen Manager für das Qualitätsdatenmanagement.

Im Folgenden finden Sie eine Liste beliebter Systeme für mittlere und große Unternehmen der Baubranche, die für ein effektives Bauprojektmanagement eingesetzt werden:

- **ERP (Enterprise Resource Planning)** - ermöglicht die Integration von Geschäftsprozessen einschließlich Buchhaltung, Beschaffung und Projektmanagement.
- **CAPEX (Capital Expenditure Planning Software)** - wird für die Budgetierung und Verwaltung von Finanzinvestitionen in Bauprojekte verwendet und hilft bei der Ermittlung der Kosten von Anlagevermögen und Investitionen in langfristige Vermögenswerte.
- **CAD (Computer-Aided Design)** und **BIM (Building Information Modeling)** - werden verwendet, um detaillierte und genaue technische Zeichnungen und 3D -Modelle von Projekten zu erstellen. Der Schwerpunkt dieser Systeme liegt auf der Arbeit mit geometrischen Informationen.
- **MEP (Mechanical, Electrical, Plumbing)** - Technische Systeme, die mechanische, elektrische und sanitäre Komponenten umfassen und das interne "Kreislaufsystem" eines Projekts detailliert beschreiben.
- **GIS (Geografische Informationssysteme**) - wird für die Geländeanalyse und Planung verwendet, einschließlich Kartografie und Raumanalyse.
- **CQMS (Software für das Qualitätsmanagement im Bauwesen)** - stellt sicher, dass die Bauprozesse mit den festgelegten Normen und Vorschriften übereinstimmen und hilft, Mängel zu beseitigen.
- **CPM (Bauprojektmanagement)** - umfasst die Planung, Koordination und Kontrolle von Bauprozessen.
- **CAFM (Computer-Aided Facility Management)** - Gebäudemanagement- und Instandhaltungssysteme.
- **SCM (Supply Chain Management)** ist erforderlich, um den Material- und Informationsfluss zwischen Lieferanten und der Baustelle zu optimieren.
- **EPM (Enterprise Performance Management)** - zielt auf die Verbesserung von Geschäftsprozessen und Leistung ab.
- **AMS (Asset Management Software)** - zur Optimierung der Nutzung, Verwaltung und Wartung von Anlagen und Infrastrukturen während des gesamten Lebenszyklus der Anlagen.
- **RPM (Real Property Management)** - umfasst Aufgaben und Prozesse im Zusammenhang mit der Verwaltung und dem Betrieb von Gebäuden und Grundstücken sowie den damit verbundenen Ressourcen und Vermögenswerten.

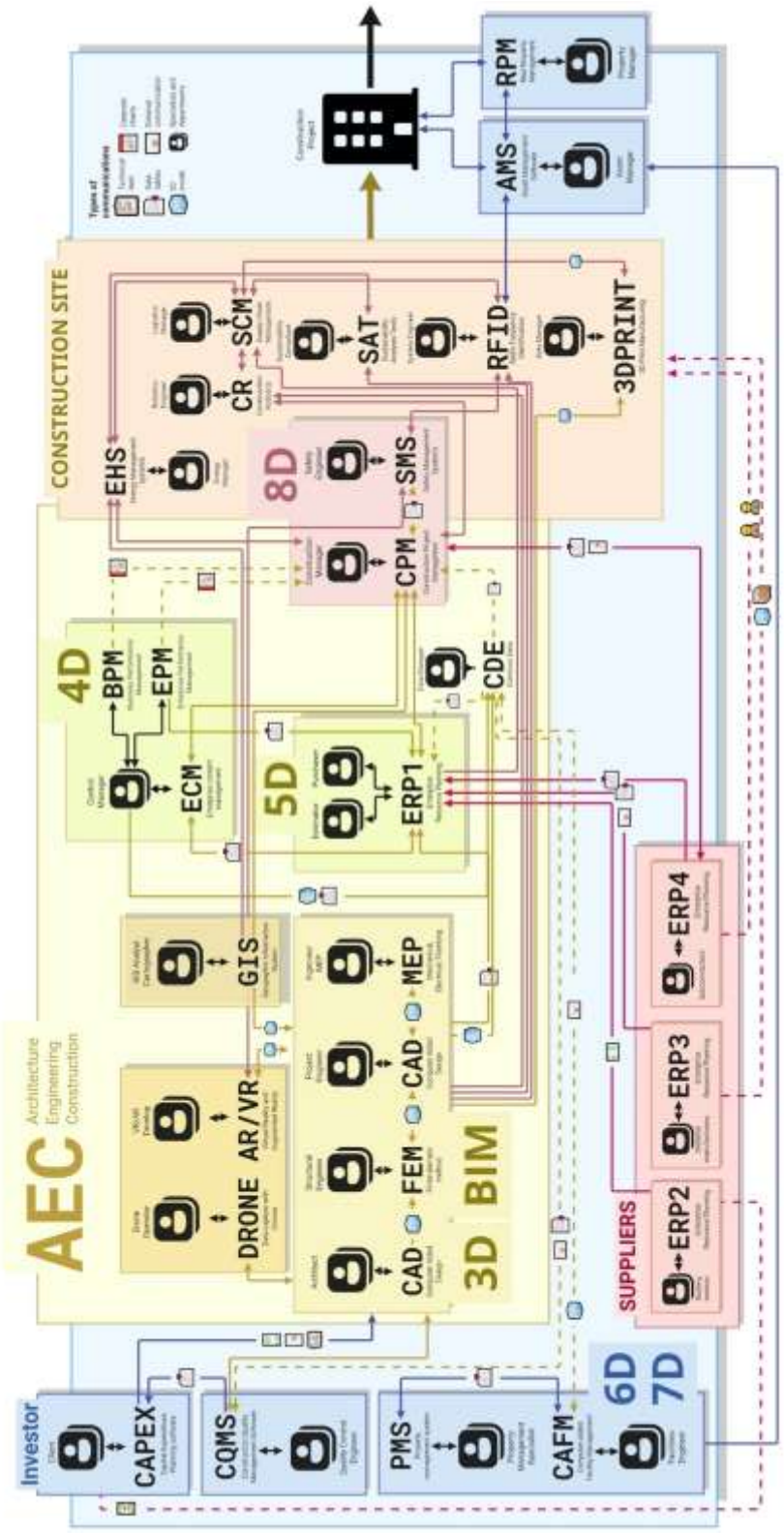


Abb. 1.2-4 Interkonnektivität von Systemen, die die Unternehmensprozesse mit dem Informationsfluss zwischen verschiedenen Abteilungen verbindet.

- **CAE (Computer-Aided Engineering)** - Computergestütztes Engineering, einschließlich Berechnungs- und Simulationssysteme wie Finite-Elemente-Analyse (FEA) und numerische Strömungsmechanik (CFD).
- **CFD (**Computational **Fluid** Dynamics**)** - Computational Fluid Dynamics, Modellierung von Flüssigkeits- und Gasströmungen. CAE-Unterkategorie.
- **CAPP (Computer-Aided Process Planning)** - Computergestützte Prozessplanung. Es wird verwendet, um Routen- und Prozesspläne zu erstellen.
- **CAM (Computer-Aided Manufacturing)** - computergestützte Fertigung, Erstellung von Steuerungsprogrammen für CNC-Maschinen.
- **PDM (Product Data Management)** - Product Data Management, ein System zur Speicherung und Verwaltung von technischer Dokumentation.
- **MES (Manufacturing Execution System)** ist ein System zur Steuerung von Fertigungsprozessen in Echtzeit.
- **PLM (Product Lifecycle Management)** - Lebenszyklusmanagement eines Projektelements, integriert PDM, CAPP, CAM und andere Systeme für eine vollständige Produktkontrolle von der Entwicklung bis zur Entsorgung.

Diese und viele andere Systeme, die eine Vielzahl von Softwarelösungen beinhalten, sind zu einem festen Bestandteil der modernen Bauindustrie geworden (Abb. 1.2-4). Im Kern handelt es sich bei diesen Systemen um spezialisierte Datenbanken mit intuitiven Schnittstellen, die eine effiziente Eingabe, Verarbeitung und Analyse von Informationen in allen Phasen der Planung und Konstruktion ermöglichen. Die Integration digitaler Werkzeuge untereinander trägt nicht nur zur Optimierung von Arbeitsabläufen bei, sondern verbessert auch die Genauigkeit von Entscheidungen erheblich, was sich positiv auf den Zeitplan und die Qualität der Projektdurchführung auswirkt.

Aber in der Hälfte der Fälle gibt es keine Integration. Laut Statistik ist nur jede zweite Anwendung oder jedes zweite System mit anderen Lösungen integriert [17]. Dies zeigt die anhaltende Fragmentierung der digitalen Umgebung und unterstreicht die Notwendigkeit, offene Standards und einheitliche Schnittstellen zu entwickeln, um einen durchgängigen Informationsaustausch innerhalb eines Bauprojekts zu gewährleisten.

Eine der größten Herausforderungen bei der Integration moderner Unternehmen ist nach wie vor die hohe Komplexität digitaler Systeme und die Anforderungen an die Nutzerkompetenz, die für eine effektive Informationsbeschaffung und -interpretation erforderlich ist. Für jedes im Unternehmen implementierte System wird ein Team von Spezialisten unter der Leitung eines Key Managers gebildet (Abb. 1.2-2).

> Der Key System Manager spielt eine entscheidende Rolle bei der richtigen Ausrichtung des Datenflusses und ist für die Qualität der endgültigen Informationen verantwortlich, so wie die ersten Manager vor Tausenden von Jahren für die auf Papyrus oder Tontafeln geschriebenen Zahlen verantwortlich waren.

Um aus disparaten Informationsflüssen ein Managementinstrument zu machen, ist die Fähigkeit zur systematischen Integration und Verwaltung von Daten unerlässlich. In dieser Architektur müssen die Manager als Elemente eines einzigen Netzwerks agieren - wie ein Myzel, das die einzelnen Teile des Unternehmens zu einem kohärenten lebenden Organismus verbindet, der sich anpassen und weiterentwickeln kann.

Unternehmensmyzel: Wie Daten mit Geschäftsprozessen verbunden sind

Der Prozess der Integration von Daten in Anwendungen und Datenbanken basiert auf der Zusammenführung von Informationen aus einer Vielzahl von Quellen, einschließlich verschiedener Abteilungen und Spezialisten (Abb. 1.2-4). Fachleute suchen nach relevanten Daten, verarbeiten sie und übertragen sie zur weiteren Verwendung in ihre Systeme und Anwendungen.

Jedes Unternehmenssystem, das aus einer Reihe von Werkzeugen, Technologien und Datenbanken besteht, ist ein Wissensbaum, der im Boden historischer Daten verwurzelt ist und wächst, um neue Früchte in Form von fertigen Lösungen zu tragen: Dokumente, Berechnungen, Tabellen, Diagramme und Dashboards (Abb. 1.2-5). Die Systeme in einem Unternehmen interagieren und kommunizieren miteinander wie Bäume in einem bestimmten Waldstück und bilden ein komplexes und gut strukturiertes System, das von erfahrenen Managern unterstützt und verwaltet wird.

> Das Informationsbeschaffungs- und -übertragungssystem eines Unternehmens funktioniert wie ein komplexes Waldnetz, das aus Bäumen (Systemen) und Myzelpilzen (Managern) besteht, die als Leiter und Recycler fungieren und dafür sorgen, dass die Informationen an die richtigen Systeme weitergeleitet werden und fließen. Dies trägt dazu bei, einen gesunden und effizienten Datenfluss und -vertrieb innerhalb des Unternehmens aufrechtzuerhalten.

Wie Wurzeln nehmen Experten in den ersten Phasen eines Projekts Rohdaten auf und verwandeln sie in Nährstoffe für das Ökosystem des Unternehmens. Daten- und Content-Management-Systeme (Abb. 1.2-4 - ERP, CPM, BIM, usw.) fungieren als leistungsfähige Informationsautobahnen, über die dieses Wissen durch alle Ebenen des Unternehmens zirkuliert.

Wie in der Natur, wo jedes Element des Ökosystems seine Rolle spielt, trägt in der Geschäftslandschaft eines Unternehmens jeder Prozessbeteiligte - vom Ingenieur bis zum Analysten - zum Wachstum und zur Fruchtbarkeit der Informationsumgebung bei. Diese systemischen "Datenbäume" (Abb. 1.2-5) sind nicht nur Mechanismen zum Sammeln von Informationen, sondern ein Wettbewerbsvorteil, der die Zukunftsfähigkeit eines Unternehmens sichert.

Waldökosysteme spiegeln erstaunlich genau wider, wie digitale Unternehmensstrukturen organisiert sind. Wie die stufenförmige Struktur eines Waldes - vom Unterholz bis zu den Baumwipfeln - weist die Unternehmensführung den Verantwortungsebenen und Funktionsbereichen Aufgaben zu.

> Tiefe und verzweigte Baumwurzeln sorgen für Widerstandsfähigkeit und Zugang zu Nährstoffen. In ähnlicher Weise unterstützen eine solide Organisationsstruktur und stabile Prozesse für die Arbeit mit Qualitätsdaten das gesamte Informationsökosystem eines Unternehmens und tragen zu dessen nachhaltigem Wachstum und Entwicklung bei, selbst in Zeiten von (windigen) Marktinstabilitäten und Krisen.

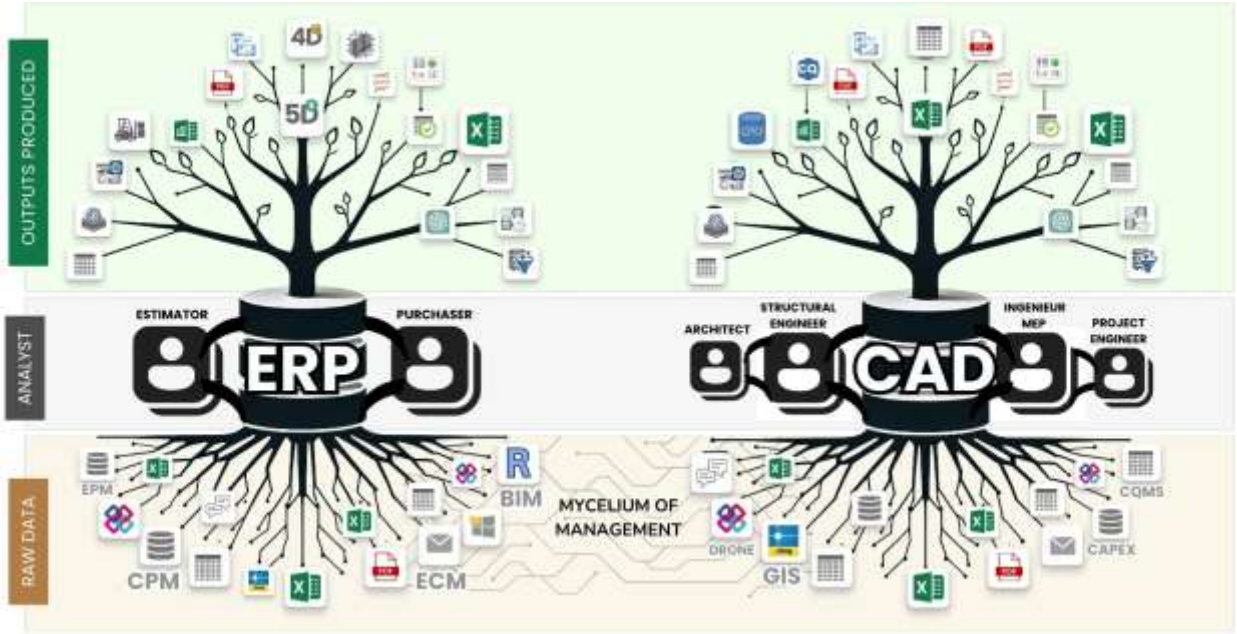


Abb. 1.2-5 Die Integration von Daten über verschiedene Systeme ist wie ein Myzel, das Manager und Spezialisten zu einem einzigen Informationsnetz verbindet.

Das moderne Verständnis von Größe in der Wirtschaft hat sich weiterentwickelt. Heute wird der Wert eines Unternehmens nicht nur durch seinen sichtbaren Teil - die "Kronen" in Form von Abschlussdokumenten und Berichten - bestimmt, sondern auch durch die Tiefe des "Wurzelsystems" der qualitativ gesammelten und systematisch verarbeiteten Daten. Je mehr Informationen gesammelt und verarbeitet werden können, desto höher wird der Geschäftswert. Unternehmen, die methodisch einen "Kompost" von bereits verarbeiteten Daten anhäufen und in der Lage sind, daraus nützliche Erkenntnisse zu gewinnen, gewinnen einen strategischen Vorteil

> Historische Informationen werden zu einer neuen Art von Kapital, das Wachstum, Prozessoptimierung und Wettbewerbsvorteile ermöglicht. In einer datengesteuerten Welt gewinnt nicht, wer mehr hat, sondern wer mehr weiß.

Für das Baugewerbe bedeutet dies den Übergang zu einem Projektmanagement in Echtzeit, bei dem alle Prozesse - von der Planung über die Beschaffung bis hin zur Koordinierung der Auftragnehmer - auf relevanten, tagesaktuellen Daten beruhen werden. Die Integration von Informationen aus verschiedenen Quellen (ERP -Systeme, CAD -Modelle, Sensoren IoT auf Baustellen, RFID) wird es ermöglichen, genauere Prognosen zu erstellen, schnell auf Änderungen zu reagieren und Verzögerungen zu vermeiden, die durch den Mangel an aktuellen Daten entstehen.

> Laut McKinsey & Company's Data-Driven Enterprise 2025 (McKinsey & Company®, 2022 [18]) werden sich erfolgreiche Unternehmen der Zukunft in allen wichtigen Aspekten ihrer Tätigkeit auf Daten stützen, von strategischen Entscheidungen bis hin zu betrieblichen Interaktionen.

Daten werden nicht mehr nur ein Analysewerkzeug sein, sondern zu einem integralen Bestandteil aller Geschäftsprozesse werden und für Transparenz, Kontrolle und Automatisierung der Verwaltung sorgen. Die

datengesteuerte Entwicklung unter wird es den Unternehmen ermöglichen, die Auswirkungen des menschlichen Faktors zu minimieren, die betrieblichen Risiken zu verringern und die Transparenz und Effizienz der Entscheidungsfindung zu erhöhen.

Das 21. Jahrhundert stellt das wirtschaftliche Paradigma auf den Kopf: Während Öl früher als "schwarzes Gold" bezeichnet wurde, weil es Maschinen und Transportmittel antrieb, werden historische Daten heute unter Zeitdruck komprimiert zu einer neuen strategischen Ressource, die nicht Maschinen, sondern Algorithmen zur Entscheidungsfindung antreibt, die das Geschäft vorantreiben werden.

KAPITEL 1.3.
DIE DIGITALE REVOLUTION UND DIE EXPLOSION DER DATEN

Der Beginn des Datenbooms als eine evolutionäre Welle

Die Baubranche erlebt eine beispiellose Informationsexplosion. Wenn man sich die Wirtschaft als einen Wissensbaum (Abb. 1.2-5) vorstellt, der von Daten gespeist wird, kann man die derzeitige Phase der Digitalisierung mit dem rasanten Wachstum der Vegetation während des Karbon vergleichen, einer Ära, in der die Biosphäre der Erde durch die rasche Anhäufung von Biomasse verändert wurde (Abb. 1.3-1).

Im Zuge der weltweiten Digitalisierung verdoppelt sich die Menge an Informationen im Baugewerbe jedes Jahr. Die moderne Technologie ermöglicht es, Daten im Hintergrund zu sammeln, in Echtzeit zu analysieren und in einem Umfang zu nutzen, der noch vor kurzem unmöglich schien.

> Nach dem von Gordon Moore (Mitbegründer von Intel®) formulierten Mooreschen Gesetz verdoppelt sich die Dichte und Komplexität integrierter Schaltkreise und die Menge der verarbeiteten und gespeicherten Daten etwa alle zwei Jahre [19].

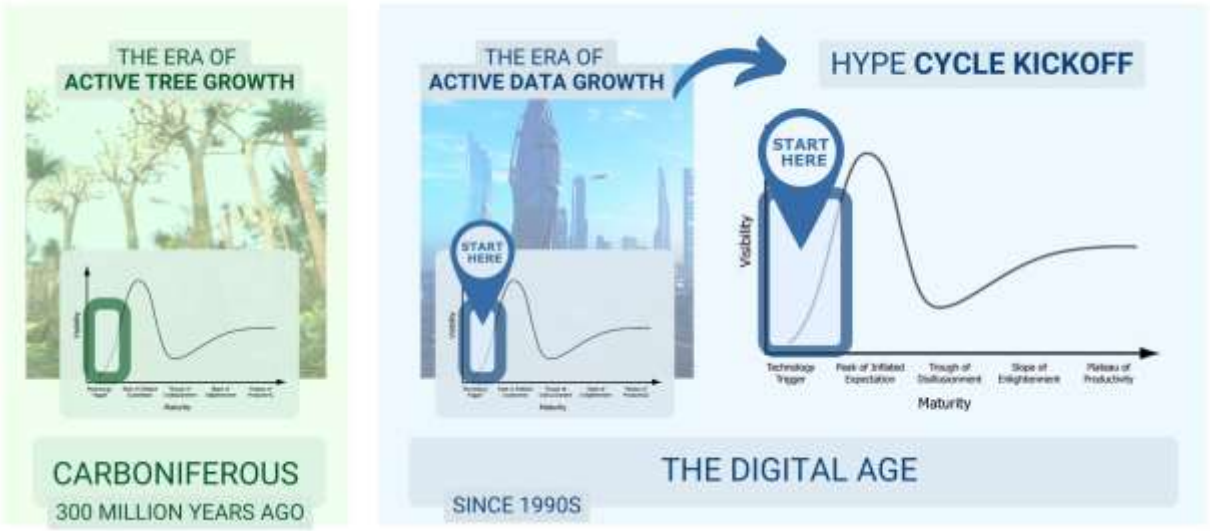


Abb. 1.3-1 Der Beginn der Digitalisierung hat zu einem exponentiellen Wachstum von Daten geführt, ähnlich wie die Vegetationswelle im Kohlezeitalter.

Während antike megalithische Bauwerke wie Göbekli Tepe (Türkei) kein dokumentiertes Wissen hinterließen, das sich zur Wiederverwendung eignet, ermöglichen es die digitalen Technologien heute, Informationen zu sammeln und wiederzuverwenden. Dies ist vergleichbar mit dem evolutionären Übergang von Sporenpflanzen zu Samenpflanzen (Angiospermen): Das Aufkommen des Samens führte zur weiten Verbreitung des Lebens auf dem Planeten. (Abb. 1.3-2).

Ebenso werden Daten aus früheren Projekten zu einer Art "digitalem Saatgut" - DNA-Wissensträgern, die skaliert und in neuen Projekten und Produkten verwendet werden können. Das Aufkommen moderner Werkzeuge der künstlichen Intelligenz - maschinelles Lernen und große Sprachmodelle (LLMs) wie ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok - ermöglicht es, Daten automatisch zu extrahieren, zu

interpretieren und in neuen Kontexten anzuwenden

> So wie Samen die Ausbreitung des Lebens auf einem zunächst leblosen Planeten revolutionierten, werden "Datensamen" zur Grundlage für das automatische Entstehen neuer Informationsstrukturen und neuen Wissens, so dass sich digitale Ökosysteme selbständig weiterentwickeln und an sich ändernde Nutzeranforderungen anpassen können.

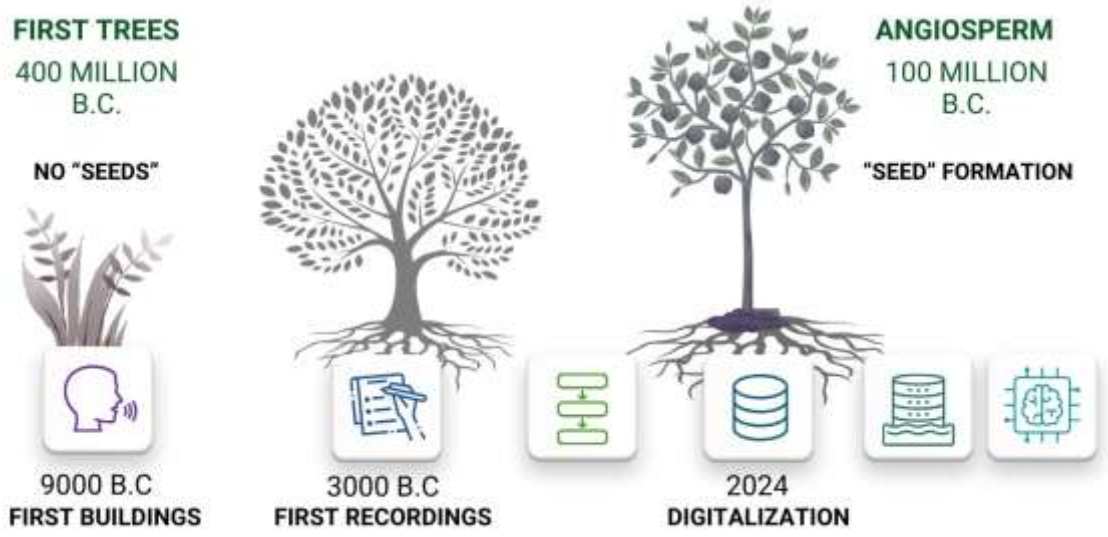


Abb. 1.3-2 Digitale "Datensamen" spielen die gleiche evolutionäre Rolle wie die Angiospermen, die Blütenpflanzen, die das Ökosystem der Erde verändert haben.

Wir stehen an der Schwelle zu einer neuen Ära im Bauwesen, in der die Explosion von Daten und die aktive Verbreitung von "Datensamen" - strukturierten Informationen aus vergangenen und laufenden Projekten - die Grundlage für die digitale Zukunft der Branche bilden. Ihre "Bestäubung" durch Big-Data-Sprachmodelle (LLMs) ermöglicht es uns, den digitalen Wandel nicht nur zu beobachten, sondern aktiv an der Schaffung selbstlernender, adaptiver Ökosysteme teilzunehmen. Dies ist keine Evolution - es ist eine digitale Revolution, in der Daten zum Hauptbaustein einer neuen Realität werden

> *Die Datenmenge im Baugewerbe nimmt aufgrund von Informationen aus verschiedenen Disziplinen während des gesamten Lebenszyklus von Bauprojekten dramatisch zu. Diese riesige Datenmenge hat die Bauindustrie in das Zeitalter von Big Data geführt [20].*
>
> - Prof. Hang Yang, Fachbereich Bauingenieurwesen und Architektur, Wuhan University of Technology, Wuhan,

Das Wachstum der Daten im Informationszeitalter erinnert an evolutionäre Prozesse in der Natur: So wie die Entwicklung der Wälder die alte Landschaft des Planeten veränderte, so verändert die derzeitige Informationsexplosion die Landschaft der gesamten Bauindustrie.

Die Menge der in einem modernen Unternehmen erzeugten Daten

In den letzten zwei Jahren wurden 90% aller weltweit vorhandenen Daten erzeugt [21]. Ab 2023 erzeugt jeder Mensch, einschließlich der Fachleute im Baugewerbe, etwa 1,7 Megabyte Daten pro Sekunde [22], und die Gesamtdatenmenge in der Welt wird im Jahr 2023 64 Zettabyte erreichen und bis 2025 voraussichtlich 180 Zettabyte oder 180*10^15 Megabyte übersteigen [23].

> Diese Informationsexplosion hat einen historischen Präzedenzfall - die Erfindung des Buchdrucks durch Johannes Gutenberg im fünfzehnten Jahrhundert. Jahrhundert. Nur fünfzig Jahre nach ihrer Einführung verdoppelte sich die Zahl der Bücher in Europa: In wenigen Jahrzehnten wurden so viele Bücher gedruckt, wie in den vorangegangenen 1.200 Jahren von Hand hergestellt worden waren [24]. Heute erleben wir ein noch schnelleres Wachstum: Die Datenmenge auf der Welt verdoppelt sich alle drei Jahre.

> Angesichts des derzeitigen Datenwachstums hat die Bauindustrie das Potenzial, in den nächsten Jahrzehnten so viele Informationen zu generieren, wie sie in ihrer gesamten bisherigen Geschichte angesammelt hat

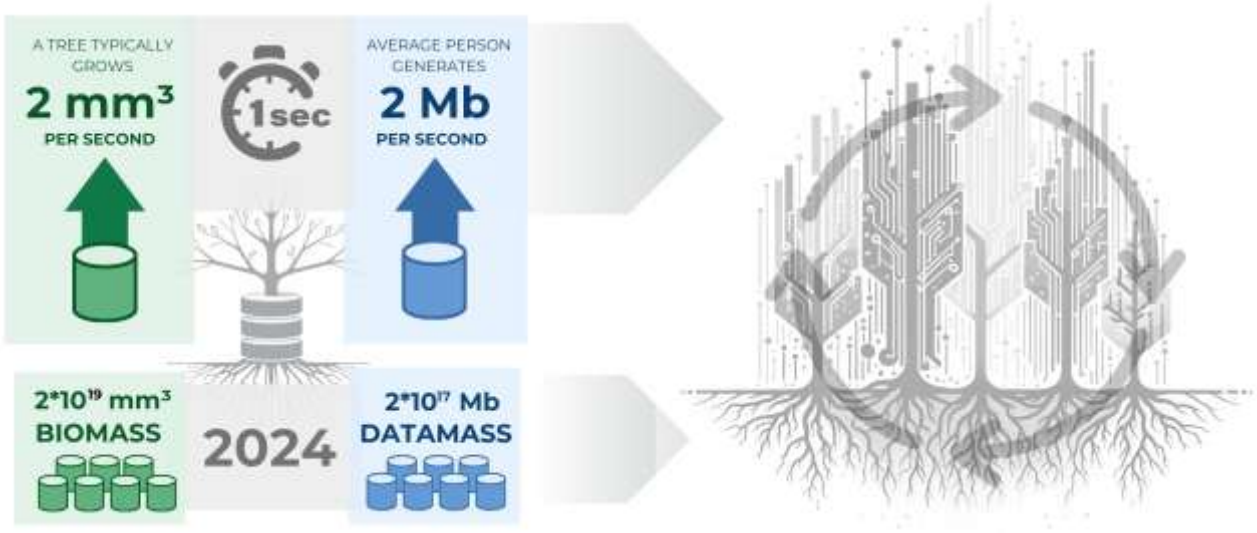


Abb. 1.3-3 Die tägliche Speicherung von Daten durch jeden Mitarbeiter auf den Servern des Unternehmens trägt zum ständigen Wachstum der Daten bei.

In der heutigen Welt des Bauwesens erzeugen selbst kleine Unternehmen täglich eine riesige Menge an Multiformat-Informationen, und der digitale Fußabdruck selbst eines kleinen Bauunternehmens kann Dutzende von Gigabytes pro Tag erreichen - von Modellen und Zeichnungen bis hin zu fotografischen Aufzeichnungen und Sensoren auf der Baustelle. Wenn wir davon ausgehen, dass jeder Techniker im Durchschnitt etwa 1,7 MB Daten pro Sekunde erzeugt, entspricht dies etwa 146 GB pro Tag oder 53 TB pro Jahr (Abb. 1.3-3).

Wenn ein Team von 10 Personen täglich nur 3 Stunden aktiv arbeitet, erreicht die kumulierte Informationsmenge pro Tag 180 Gigabyte (Abb. 1.3-4).

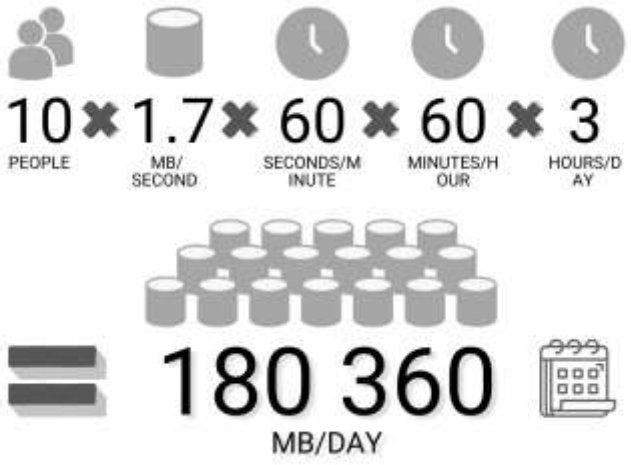


Abbildung 1.3-4 Ein Unternehmen mit 10 Mitarbeitern erzeugt täglich etwa 50-200 Gigabyte an Daten.

Wenn man davon ausgeht, dass 30% der Arbeitsdaten neu sind (der Rest wird überschrieben oder gelöscht), kann ein Unternehmen mit 10 Mitarbeitern pro Monat mehrere hundert Gigabyte an neuen Daten erzeugen (die tatsächlichen Zahlen hängen von der Art der Geschäftstätigkeit des Unternehmens ab)

Damit ist klar: Wir generieren nicht nur immer mehr Daten - wir stehen auch vor einem wachsenden Bedarf an deren effizienter Verwaltung, Speicherung und langfristiger Verfügbarkeit. Und während die Daten früher kostenlos auf lokalen Servern "liegen" konnten, gehen im Rahmen der digitalen Transformation immer mehr Unternehmen dazu über, Cloud-Lösungen als Basis ihrer Informationsinfrastruktur zu nutzen.

Die Kosten der Datenspeicherung: der wirtschaftliche Aspekt

In den letzten Jahren lagern immer mehr Unternehmen die Datenspeicherung in Cloud-Dienste aus. Wenn ein Unternehmen beispielsweise die Hälfte seiner Daten in der Cloud hostet, können seine Speicherkosten bei einem durchschnittlichen Preis von 0,015 USD pro Gigabyte pro Monat um 10-50 USD [25] pro Monat steigen.

Für ein kleines Unternehmen mit typischen Datenerzeugungsmustern können die Kosten für die Speicherung in der Cloud innerhalb weniger Jahre von Hunderten bis möglicherweise über tausend Dollar pro Monat reichen (Abb. 1.3-5), was eine potenziell erhebliche finanzielle Belastung darstellt.

> Laut der Forrester-Studie "Enterprises Outsource Data Storage as Complexity Grows" [26], für die 214 Entscheidungsträger aus dem Bereich Technologie-Infrastruktur befragt wurden Laut der Forrester-Studie "Enterprises Outsourcing Data Storage as Growing Complexity" [26], für die 214 Entscheidungsträger aus dem Bereich Technologie-Infrastruktur befragt wurden, lagert mehr als ein Drittel der Unternehmen die Datenspeicherung aus, um das wachsende Volumen und die Komplexität der Datenoperationen zu bewältigen, wobei fast zwei Drittel der Unternehmen ein abonnementbasiertes Modell bevorzugen.

Abb. 1.3-5 Die Verlagerung von Daten in die Cloud kann die monatlichen Speicherkosten um bis zu 2.000 Dollar erhöhen, selbst für ein Unternehmen mit nur 10 Mitarbeitern.

Die Situation wird noch komplizierter durch die beschleunigte Einführung von Cloud-Technologien wie CAD (BIM), CAFM, PMIS und ERP -Systemen, die die Kosten für die Datenspeicherung und -verarbeitung weiter erhöhen. Infolgedessen sind die Unternehmen gezwungen, nach Möglichkeiten zu suchen, die Kosten zu optimieren und die Abhängigkeit von Cloud-Anbietern zu verringern.

Seit 2023, mit der aktiven Entwicklung von großen Sprachmodellen (LLM), haben die Ansätze zur Datenspeicherung begonnen, sich zu ändern. Immer mehr Unternehmen denken darüber nach, die Kontrolle über ihre Daten zurückzugewinnen, da es sicherer und rentabler wird, Informationen auf ihren eigenen Servern zu verarbeiten.

In diesem Zusammenhang rückt der Trend weg von der Cloud-basierten Speicherung und Verarbeitung nur der notwendigen Daten hin zum lokalen Einsatz von LLM und KI -Lösungen für Unternehmen in den Vordergrund. Wie der CEO von Microsoft in einem seiner Interviews [27] betonte, werden KI-Agenten Prozesse in Datenbanken verwalten und die Funktionen verschiedener Systeme automatisieren, anstatt sich auf mehrere separate Anwendungen oder cloudbasierte SaaS-Lösungen zu verlassen, um verschiedene Aufgaben zu erfüllen.

> *[...] Die alte Herangehensweise an dieses [Datenverarbeitungs]-Problem war: Wenn man zurückdenkt, wie verschiedene Geschäftsanwendungen die Integration gehandhabt haben, wurden Konnektoren verwendet. Die Unternehmen verkauften Lizenzen für diese Konnektoren, und das Geschäftsmodell wurde um diese Konnektoren herum aufgebaut. SAP [ERP] ist eines der klassischen Beispiele: Man konnte nur auf SAP-Daten zugreifen, wenn man den richtigen Konnektor hatte. Es scheint mir also, dass sich etwas Ähnliches im Fall der [KI] Agenteninteraktion herausbilden wird [...]. Der Ansatz, den wir zumindest verfolgen, ist: Ich denke, dass das Konzept der Existenz von Geschäftsanwendungen im Zeitalter der [KI-]Agenten wahrscheinlich zusammenbrechen wird. Denn wenn man darüber nachdenkt, sind sie im Grunde genommen Datenbanken mit einem Haufen Geschäftslogik*
>
> - Satya Nadella, Microsoft CEO, Interview mit BG2 Channel, 2024. [28]

In diesem Paradigma geht der datengesteuerte LLM-Ansatz über klassische Systeme hinaus. Künstliche Intelligenz wird zu einem Vermittler zwischen dem Nutzer und den Daten (Abb. 2.2-3, Abb. 2.2-4), wodurch die Notwendigkeit mehrerer Zwischenschnittstellen entfällt und die Effizienz der Geschäftsprozesse erhöht wird. Mehr über diesen Ansatz für die Arbeit mit Daten erfahren Sie im Kapitel "Das Chaos in Ordnung bringen und die Komplexität reduzieren".

Während die Architektur der Zukunft noch Gestalt annimmt, sind die Unternehmen bereits mit den Folgen vergangener Entscheidungen konfrontiert. Die massive Digitalisierung der letzten Jahrzehnte, die mit der Einführung unterschiedlicher Systeme und der unkontrollierten Anhäufung von Daten einherging, hat zu einem neuen Problem geführt - der Informationsflut.

Die Grenzen der Datenakkumulation: von der Masse zur Bedeutung

Moderne Unternehmenssysteme entwickeln sich erfolgreich und funktionieren unter kontrolliertem Wachstum, wenn das Datenvolumen und die Anzahl der Anwendungen im Gleichgewicht mit den Fähigkeiten der IT-Abteilungen und Manager stehen. In den letzten Jahrzehnten hat die Digitalisierung jedoch zu einem unkontrollierbaren Anstieg des Datenvolumens und der Komplexität geführt, was einen Übersättigungseffekt im Informationsökosystem der Unternehmen zur Folge hatte.

Heutzutage sind Server und Speicher einem beispiellosen Zustrom unverarbeiteter und mehrfach formatierter Informationen ausgesetzt, die keine Zeit haben, sich in Kompost zu verwandeln und schnell irrelevant werden. Begrenzte Unternehmensressourcen können diese Flut nicht bewältigen, und die Daten sammeln sich in isolierten Silos an (so genannte "Silos "), die manuell bearbeitet werden müssen, um nützliche Informationen zu extrahieren.

> Wie ein mit Efeu bewachsener und von Schimmelpilzen bedeckter Wald leiden moderne Unternehmensmanagementsysteme daher oft unter Informationsüberlastung. Statt eines nährenden Informationshumus im Herzen des Unternehmensökosystems bilden sich isolierte Bereiche mit Daten unterschiedlicher Formate, was unweigerlich zu einer Abnahme der Gesamteffizienz der Geschäftsprozesse führt.

Auf die lange Phase des exponentiellen Datenwachstums der letzten 40 Jahre wird unweigerlich eine Sättigungs- und anschließende Abkühlungsphase folgen. Wenn die Speicherung an ihre Grenzen stößt, wird es zu einem qualitativen Wandel kommen: Daten werden nicht mehr nur ein Speicherobjekt sein, sondern eine strategische Ressource.

Mit der Entwicklung der künstlichen Intelligenz und des maschinellen Lernens haben die Unternehmen die Möglichkeit, die Kosten für die Informationsverarbeitung zu senken und vom quantitativen Wachstum zur qualitativen Nutzung der Daten überzugehen. In den nächsten zehn Jahren wird die Bauindustrie ihren Schwerpunkt von der Erzeugung von immer mehr Daten auf die Gewährleistung ihrer Struktur, Integrität und ihres analytischen Werts verlagern müssen.

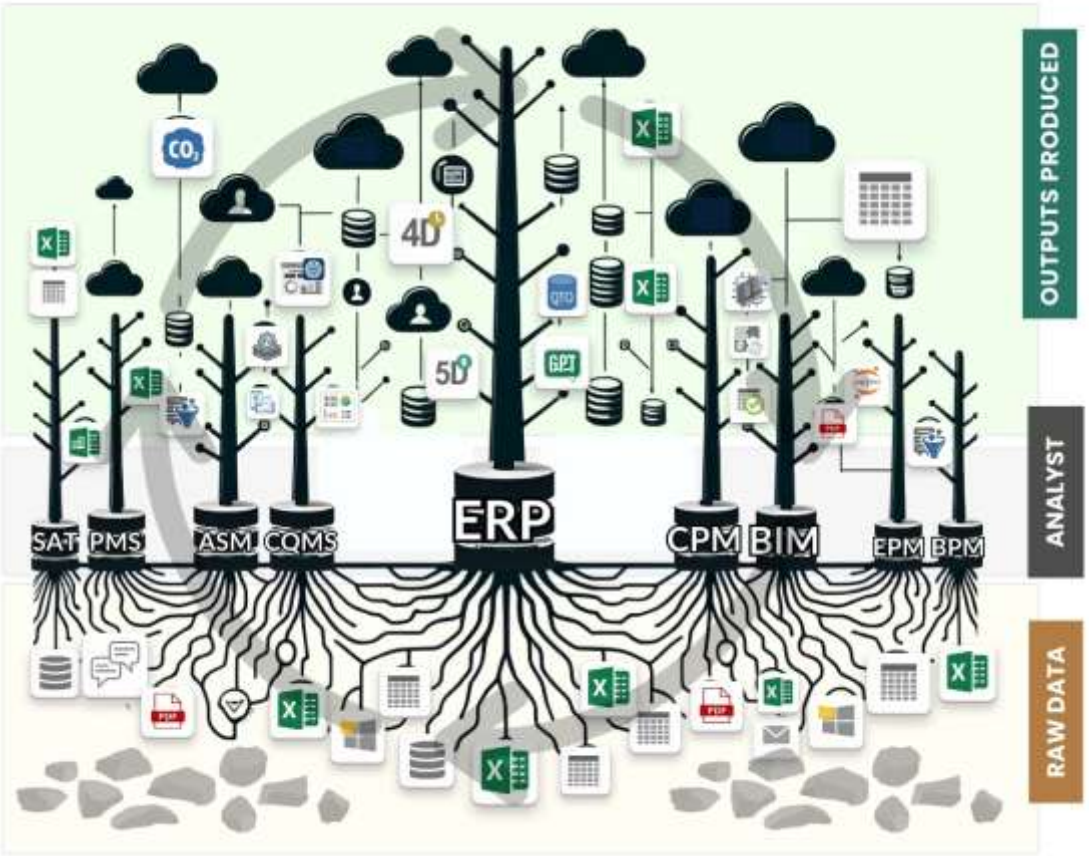


Abbildung 1.3-6 Isolierte Datenquellen verhindern den Informationsaustausch zwischen Datensystemen.

> Der Hauptwert liegt nicht mehr in der Menge der Informationen, sondern in der Fähigkeit, sie automatisch zu interpretieren und sie in angewandtes Wissen umzuwandeln, das für Managemententscheidungen nützlich ist. Damit Daten wirklich nützlich werden, müssen sie richtig verwaltet werden: Sie müssen gesammelt, überprüft, strukturiert, gespeichert und im Zusammenhang mit spezifischen Geschäftsaufgaben analysiert werden.

Der Datenanalyseprozess in einem Unternehmen ist vergleichbar mit dem Lebens- und Verfallszyklus von Bäumen in einem Wald und der Entstehung neuer junger und starker Bäume: reife Bäume sterben ab, zerfallen und werden zum Nährboden für neues Wachstum. Beendete und abgeschlossene Prozesse werden, wenn sie abgeschlossen sind, Teil des Informationsökosystems des Unternehmens und werden schließlich zum Informationshumus, der das zukünftige Wachstum neuer Systeme und Daten antreibt.

In der Praxis wird dieser Kreislauf jedoch häufig unterbrochen. Anstelle einer organischen Erneuerung entsteht ein schichtweises Chaos, ähnlich wie bei geologischen Schichten, bei dem neue Systeme ohne tiefgreifende Integration und Strukturierung auf alte Systeme aufgeschichtet werden. Infolgedessen entstehen disparate Informations-"Silos", die den Wissensaustausch behindern und die Datenverwaltung erschweren.

Nächste Schritte: von der Datentheorie zum praktischen Wandel

Die Entwicklung von Daten im Bauwesen ist eine Reise von Tontafeln zu modernen modularen Plattformen. Die Herausforderung besteht heute nicht im Sammeln von Informationen, sondern in der Schaffung eines Rahmens, der disparate und vielfältige Daten in eine strategische Ressource verwandelt. Unabhängig davon, ob Sie in der Rolle eines Unternehmensleiters oder eines Ingenieurs tätig sind, wird das Verständnis für den Wert von Daten und die Arbeit mit ihnen in Zukunft eine Schlüsselkompetenz sein.

Um diesen Teil zusammenzufassen, lohnt es sich, die wichtigsten praktischen Schritte hervorzuheben, die Ihnen helfen werden, die besprochenen Ansätze bei Ihren täglichen Aufgaben anzuwenden:

- Führen Sie ein persönliches Audit der Informationsflüsse durch
  - ☐ Erstellen Sie eine Liste mit allen Systemen und Anwendungen, mit denen Sie täglich arbeiten
  - ☐ Markieren Sie, wo Sie die meiste Zeit mit dem Suchen oder Überprüfen von Daten verbringen
  - ☐ Identifizieren Sie Ihre wichtigsten Informationsquellen
  - ☐ Analysieren Sie Ihre aktuelle Anwendungslandschaft auf Redundanz und doppelte Funktionen
- Streben Sie danach, die Prozesse nach dem Grad der analytischen Reife zu durchlaufen
  - ☐ Beginnen Sie Ihre Aufgaben mit einer beschreibenden Analyse (was ist passiert?)
  - ☐ Schrittweise Einführung einer Diagnose (warum ist das passiert?)
  - ☐ Überlegen Sie, wie Sie in Prozessen zu prädiktiven (was wird passieren?) und präskriptiven (was ist zu tun?) Analysen übergehen können.
- Beginnen Sie mit der Strukturierung Ihrer Arbeitsdaten
  - ☐ Einführung eines einheitlichen Systems für die Benennung von Dateien und Ordnern, die Sie bei Ihrer Arbeit häufig verwenden
  - ☐ Vorlagen für häufig verwendete Dokumente und Berichte erstellen
  - ☐ Regelmäßige Archivierung abgeschlossener Projekte mit einer klaren Struktur

Selbst wenn Sie nicht die gesamte Informationsinfrastruktur in Ihrem Team oder Unternehmen ändern können, sollten Sie mit Ihren eigenen Prozessen und kleinen Verbesserungen in Ihrer täglichen Arbeit beginnen. Denken Sie daran, dass der wahre Wert von Daten nicht in ihrer Menge liegt, sondern in der Fähigkeit, daraus verwertbare Erkenntnisse zu gewinnen. Selbst kleine, aber richtig strukturierte und analysierte Datensätze können erhebliche Auswirkungen haben, wenn sie in Entscheidungsprozesse integriert werden.

In den nächsten Teilen des Buches werden wir uns mit spezifischen Methoden und Werkzeugen für die Arbeit mit Daten befassen, Möglichkeiten zur Umwandlung unstrukturierter Informationen in strukturierte Datensätze untersuchen, Technologien zur Automatisierung von Analysen erforschen und detailliert aufzeigen, wie ein effektives Analyse-Ökosystem in einem Bauunternehmen aufgebaut werden kann.

II TEIL
WIE DAS BAUGEWERBE IM DATENCHAOS ERTRINKT

Der zweite Teil ist einer kritischen Analyse der Herausforderungen gewidmet, denen sich die Bauunternehmen im Umgang mit den zunehmenden Datenmengen gegenübersehen. Die Folgen der Informationsfragmentierung und das Phänomen der "Daten in Silos", das eine effiziente Entscheidungsfindung behindert, werden eingehend untersucht. Die Problematik des HiPPO -Ansatzes (Highest Paid Person's Opinion) und seine Auswirkungen auf die Qualität von Managemententscheidungen in Bauprojekten werden untersucht. Die Auswirkungen der dynamischen Geschäftsprozesse und ihrer zunehmenden Komplexität auf den Informationsfluss und die betriebliche Effizienz werden bewertet. Anhand konkreter Beispiele wird aufgezeigt, wie eine übermäßige Systemkomplexität die Kosten erhöht und die Flexibilität von Organisationen verringert. Besonderes Augenmerk wird auf die Einschränkungen gelegt, die durch proprietäre Formate entstehen, sowie auf die Aussichten für die Verwendung offener Standards in der Bauindustrie. Das Konzept des Übergangs zu KI- und LLM-basierten Software-Ökosystemen wird vorgestellt, die eine übermäßige Komplexität und technische Barrieren minimieren.

KAPITEL 2.1.
DATENFRAGMENTIERUNG UND -SILOS

Je mehr Werkzeuge, desto effizienter das Unternehmen?

Auf den ersten Blick mag es so aussehen, als ob mehr digitale Werkzeuge zu mehr Effizienz führen. In der Praxis ist dies jedoch nicht der Fall. Mit jeder neuen Lösung, sei es ein Cloud-Service, ein Altsystem oder ein weiterer Excel-Bericht, fügt ein Unternehmen seiner digitalen Landschaft eine weitere Schicht hinzu - eine Schicht, die oft nicht mit den anderen integriert ist (Abb. 2.1-1).

Daten lassen sich mit Kohle oder Öl vergleichen: Es dauert Jahre, bis sie sich angesammelt haben, verdichtet unter Schichten von Chaos, Fehlern, unstrukturierten Prozessen und vergessenen Formaten. Um wirklich nützliche Informationen daraus zu extrahieren, müssen sich Unternehmen buchstäblich durch Schichten veralteter Lösungen und digitales Rauschen wühlen.

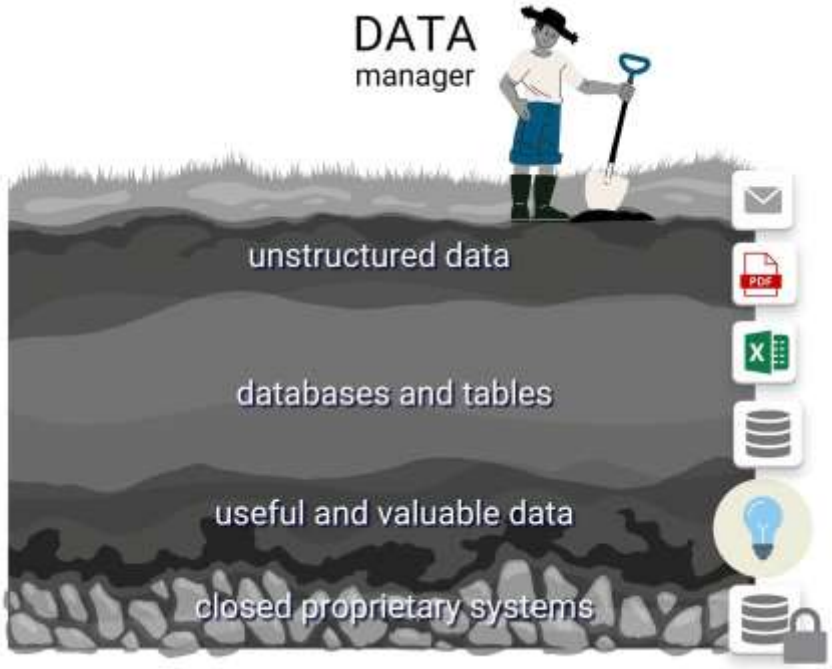


Abb. 2.1-1 Vielfältige Daten bilden kompartimentierte Schichten - selbst "goldene" Erkenntnisse gehen im geologischen Gestein der systemischen Komplexität verloren.

Jede neue Anwendung hinterlässt eine Spur: eine Datei, eine Tabelle oder ein ganzes isoliertes "Silo" auf dem Server. Eine Schicht ist Lehm (veraltete und vergessene Daten), eine andere ist Sand (disparate Tabellen und Berichte), und die dritte ist Granit (geschlossene proprietäre Formate, die nicht integriert werden können). Im Laufe der Zeit ähnelt die digitale Umgebung eines Unternehmens immer mehr einem Reservoir unkontrollierter Informationsansammlungen, in dem der Wert tief in den Servern des Unternehmens verloren geht.

Mit jedem neuen Projekt und jedem neuen System wird nicht nur die Infrastruktur komplexer, sondern auch

der Weg zu nützlichen Qualitätsdaten. Um an das wertvolle "Gestein" heranzukommen, müssen die Daten gründlich gesäubert, strukturiert, in sinnvolle Teile gruppiert und durch Analysen und Datenmodellierung strategisch wichtige Erkenntnisse gewonnen werden.

> *Daten sind ein wertvolles Gut, und sie werden länger halten als die Systeme [die die Daten verarbeiten] selbst [29].*
>
> - Tim Berners-Lee, Vater des World Wide Web und Schöpfer der ersten Website

Bevor Daten zu einer "wertvollen Sache" und einer zuverlässigen Entscheidungsgrundlage werden können, müssen sie sorgfältig aufbereitet werden. Erst die richtige Vorverarbeitung macht aus disparaten Daten strukturierte Erfahrung, nützlichen Informationshumus, der dann zum Prognose- und Optimierungsinstrument wird.

> Es ist ein Irrglaube, dass man perfekt saubere Daten braucht, um mit der Analyse beginnen zu können, aber in der Praxis ist die Fähigkeit, mit schmutzigen Daten zu arbeiten, ein wesentlicher Teil des Prozesses.

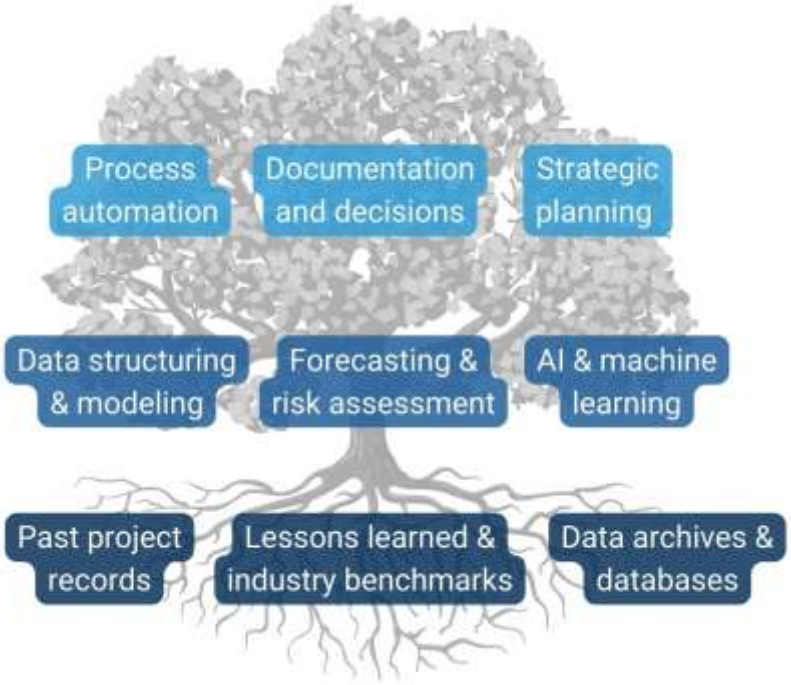


Abb. 2.1-2 Daten sind das Wurzelsystem und die Grundlage der Wirtschaft, die ihrerseits auf Entscheidungsprozessen basiert.

Im Zuge des technologischen Fortschritts muss sich auch Ihr Unternehmen weiterentwickeln und lernen, wie man aus Daten Werte schafft. So wie Öl- und Kohleunternehmen die Infrastruktur für die Gewinnung

von Mineralien aufbauen, müssen auch Unternehmen lernen, den Fluss neuer Informationen auf ihren eigenen Servern zu verwalten und wertvolle Erkenntnisse aus ungenutzten, unformatierten und veralteten Daten zu gewinnen und sie in eine strategische Ressource zu verwandeln.

Die Schaffung von Feldern (Data Warehouses) ist der erste Schritt. Selbst die leistungsfähigsten Tools lösen das Problem der Datenisolierung und der Daten in verschiedenen Formaten nicht, wenn die Unternehmen weiterhin in Silosystemen arbeiten. Wenn Daten getrennt voneinander vorhanden sind, ohne dass sie sich überschneiden und gemeinsam genutzt werden, sind Unternehmen mit einem "Datensilo"-Effekt konfrontiert. Anstelle einer einheitlichen, konsistenten Infrastruktur sind die Unternehmen gezwungen, Ressourcen für die Zusammenführung und Synchronisierung von Daten aufzuwenden.

Datensilos und ihre Auswirkungen auf die Unternehmensleistung

Stellen Sie sich vor, Sie bauen eine Wohnsiedlung, aber jedes Team hat sein eigenes Projekt. Die einen bauen Wände, die anderen verlegen Leitungen und wieder andere legen Straßen an, ohne sich abzusprechen. Das Ergebnis ist, dass die Rohre nicht mit den Öffnungen in den Wänden übereinstimmen, dass die Aufzugsschächte nicht mit den Stockwerken übereinstimmen und dass die Straßen abgebaut und neu verlegt werden müssen.

Diese Situation ist nicht nur ein hypothetisches Szenario, sondern Realität bei vielen modernen Bauprojekten. Aufgrund der großen Anzahl von General- und Subunternehmern, die mit unterschiedlichen Systemen und ohne eine einzige Koordinierungsstelle arbeiten, wird der Prozess zu einer Reihe von endlosen Genehmigungen, Nacharbeiten und Konflikten. All dies führt zu erheblichen Verzögerungen und einer Vervielfachung der Projektkosten.

Eine klassische Situation auf einer Baustelle ist einfach: Die Schalung ist fertig, aber die Lieferung der Bewehrung ist nicht rechtzeitig eingetroffen. Wenn man die Informationen in verschiedenen Systemen überprüft, sieht die Kommunikation in etwa so aus:

- **Der Vorarbeiter** auf der Baustelle am 20. schreibt an den Projektleiter: "*Wir sind mit dem Setzen der Schalung fertig, wo ist der Betonstahl?*"
- **Projektleiter** (PMIS) an die Beschaffungsabteilung**:** - "*Die Schalung ist fertig. In meinem System [PMIS] steht, dass der Betonstahl am 18. eintreffen sollte. Wo sind die Bewehrungsstäbe?"*
- **Spezialist für die Lieferkette** (ERP): - Unser ERP sagt, dass die Lieferung am 25. erfolgen wird.
- **Data Engineer** oder IT-Abteilung (zuständig für Integrationen): - In PMIS ist das Datum der 18., in ERP der 25. Es gibt keine OrderID-Verknüpfung zwischen ERP und PMIS, so dass die Daten nicht synchronisiert sind. Dies ist ein typisches Beispiel für eine Informationslücke.
- **Projektleiter** an Generaldirektor - "Die *Lieferung der Armaturen verzögert sich, die Baustelle steht, und es ist unklar, wer verantwortlich ist*".

Die Ursache für den Vorfall war die Isolierung von Daten in unterschiedlichen Systemen. Durch die Integration und Vereinheitlichung von Datenquellen, die Schaffung eines einzigen Informationsspeichers und die Automatisierung durch ETL -Werkzeuge (Apache NiFi, Airflow oder n8n) können die Silos zwischen den Systemen beseitigt werden. Diese und andere Methoden und Tools werden in späteren Abschnitten des Buches ausführlich besprochen.

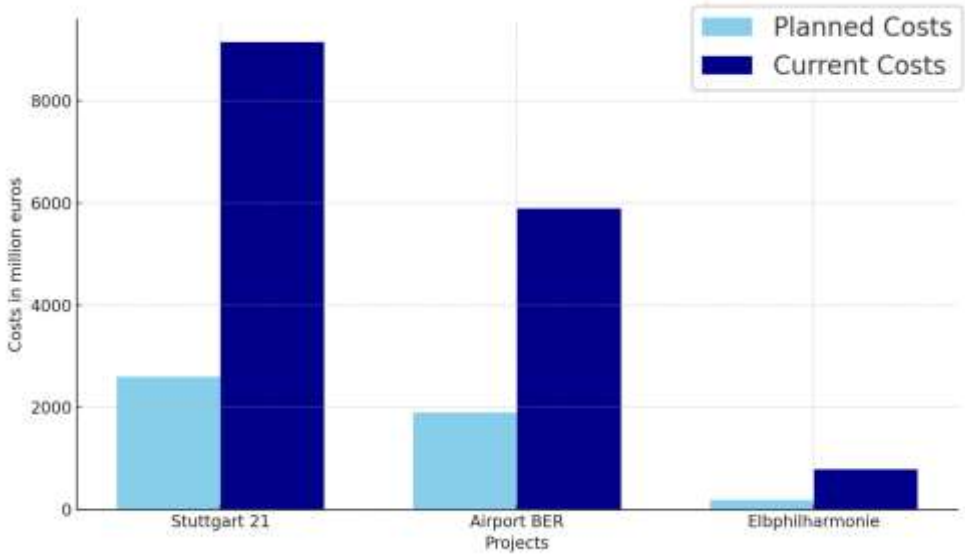


Abb. 2.1-3 Vergleich der geplanten und tatsächlichen Kosten von großen Infrastrukturprojekten in Deutschland.

Das Gleiche passiert mit Unternehmenssystemen: Zuerst werden Insellösungen geschaffen, und dann müssen riesige Budgets aufgewendet werden, um sie zu integrieren und zu harmonisieren. Wären die Daten- und Kommunikationsmodelle von Anfang an durchdacht, gäbe es überhaupt keinen Bedarf für eine Integration. Silo-Daten verursachen in der digitalen Welt ein Chaos, wie ein unkoordinierter Bauprozess.

> Laut der KPMG-Studie "Cue construction 4.0: Time to make or break" aus dem Jahr 2023 tauschen nur 36% der Unternehmen Daten effektiv abteilungsübergreifend aus, während 61% aufgrund von isolierten Datensilos ernsthafte Probleme haben [30].

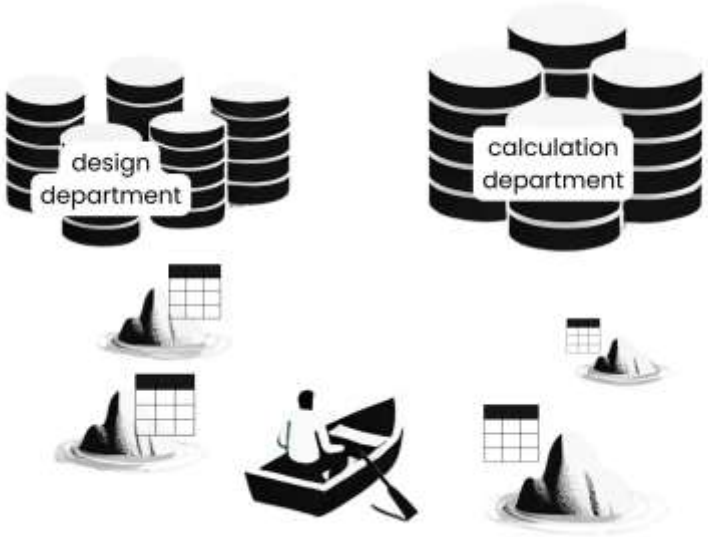


Abb. 2.1-4 Jahrelang sammeln sich schwer zu erfassende Daten in isolierten "Silos" an, mit dem Risiko, dass sie nie genutzt werden.

Unternehmensdaten sind in isolierten Systemen gespeichert, wie einzelne Bäume, die über die Landschaft

verstreut sind. Jeder Baum enthält wertvolle Informationen, aber der Mangel an Verbindungen zwischen ihnen verhindert die Schaffung eines einzigen, vernetzten Ökosystems. Diese Abschottung behindert den Datenfluss und schränkt die Fähigkeit des Unternehmens ein, das Gesamtbild zu erkennen. Die Verbindung dieser Silos ist ein äußerst langwieriger und komplexer Prozess, bei dem auf Managementebene ein Pilzmyzel wächst, um zu lernen, wie man einzelne Informationen zwischen den Systemen austauscht.

Laut einer WEF-Studie aus dem Jahr 2016 ist eines der Haupthindernisse für die digitale Transformation der Mangel an gemeinsamen Datenstandards und die Fragmentierung.

> *Die Baubranche ist eine der am stärksten fragmentierten Branchen der Welt und hängt vom reibungslosen Zusammenspiel aller Beteiligten in der Wertschöpfungskette ab [5].*
>
> - Weltwirtschaftsforum 2016: Die Zukunft des gestalten

Konstrukteure, Manager, Koordinatoren und Entwickler ziehen es oft vor, autonom zu arbeiten und die Komplexität der Koordination zu vermeiden. Diese natürliche Neigung führt zur Bildung von "Informationssilos", in denen Daten in separaten Systemen isoliert werden. Je mehr solcher isolierten Systeme es gibt, desto schwieriger ist es, sie dazu zu bringen, zusammenzuarbeiten. Im Laufe der Zeit erhält jedes System seine eigene Datenbank und eine spezialisierte Unterstützungsabteilung von Managern (Abb. 1.2-4), was die Integration weiter erschwert.

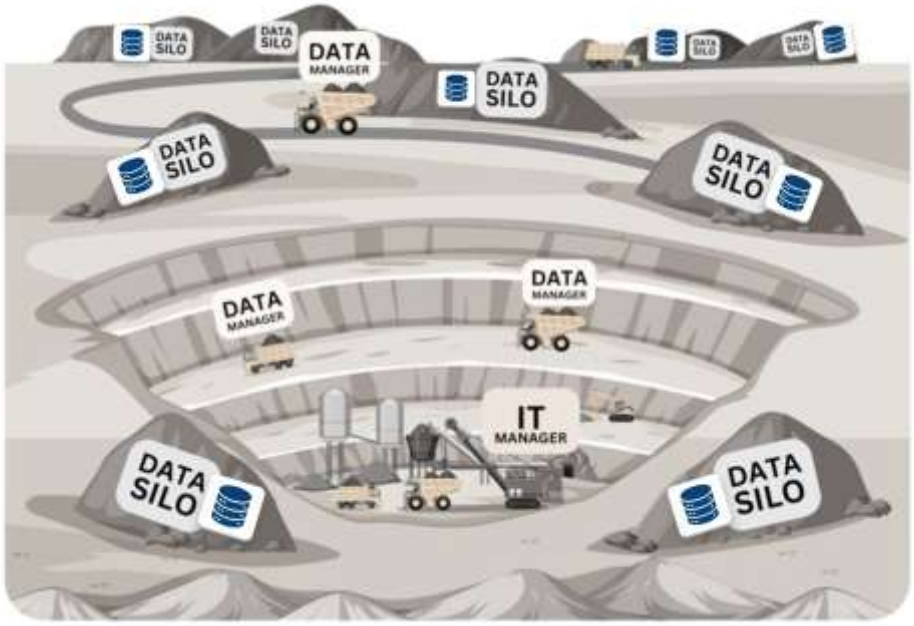


Abb. 2.1-5 Jedes System neigt dazu, sein eigenes einzigartiges Datensilo zu schaffen, das mit geeigneten Werkzeugen verarbeitet werden muss [31].

Der Teufelskreis bei Unternehmenssystemen sieht so aus: Unternehmen investieren in komplexe isolierte Lösungen, haben dann hohe Kosten für deren Integration zu tragen, und Entwickler, die sich der Komplexität

der Kombination von Systemen bewusst sind, ziehen es vor, in ihren geschlossenen Ökosystemen zu arbeiten. All dies erhöht die Fragmentierung der IT-Landschaft und erschwert die Migration zu neuen Lösungen (Abb. 2.1-5). Manager kritisieren Datensilos, analysieren aber selten deren Ursachen und wie man sie verhindern kann. Manager beschweren sich über veraltete IT-Systeme, deren Ersatz jedoch erhebliche Investitionen erfordert und selten die erwarteten Ergebnisse bringt. Infolgedessen verschlimmern selbst Versuche, das Problem zu bekämpfen, häufig die Situation.

Der Hauptgrund für diese Diskrepanz ist der Vorrang von Anwendungen vor Daten. Die Unternehmen entwickeln zunächst getrennte Systeme oder kaufen Standardlösungen von Anbietern und versuchen dann, sie zu vereinheitlichen, indem sie doppelte und inkompatible Speicher und Datenbanken einrichten.

Die Überwindung des Problems der Fragmentierung erfordert einen radikalen neuen Ansatz - Vorrang für Daten vor Anwendungen. Die Unternehmen müssen zunächst Strategien für die Datenverwaltung und Datenmodelle entwickeln und dann Systeme aufbauen oder Lösungen kaufen, die mit einem einzigen Satz von Informationen arbeiten, anstatt neue Barrieren zu schaffen.

> *Wir treten in eine neue Welt ein, in der Daten vielleicht wichtiger sind als Software.*
>
> - Tim O'Reilly, CEO von O'Reilly Media, Inc.

> Die Studie "Rethinking Construction: the path to improved productivity" (2016) des McKinsey Global Institute zeigt, dass die Bauindustrie bei der digitalen Transformation hinter anderen Branchen zurückbleibt [32]. Dem Bericht zufolge kann die Einführung eines automatisierten Datenmanagements und digitaler Plattformen die Produktivität erheblich steigern und Verluste aufgrund von Prozessinkonsistenzen verringern. Diese Notwendigkeit der digitalen Transformation wird auch im Bericht von Egan (UK, 1998) [33] betont, der die Schlüsselrolle integrierter Prozesse und eines kollaborativen Ansatzes im Bauwesen hervorhebt.

> Während in den letzten 10.000 Jahren das Hauptproblem für Datenverwalter der Mangel an Daten war, stehen Nutzer und Verwalter mit der Lawine von Daten und Datenverwaltungssystemen vor einem Problem - einem Überangebot an Daten, das es schwierig macht, rechtlich korrekte und hochwertige Informationen zu finden.

Die unterschiedlichen Datensilos führen unweigerlich zu dem ernsten Problem einer verminderten Datenqualität. Bei mehreren unabhängigen Systemen können dieselben Daten in verschiedenen Versionen vorliegen, oft mit widersprüchlichen Werten, was zusätzliche Schwierigkeiten für die Nutzer mit sich bringt, die feststellen müssen, welche Informationen relevant und zuverlässig sind.

Doppelarbeit und mangelnde Datenqualität als Folge der Uneinheitlichkeit

Aufgrund des Problems der Datensilos müssen Manager viel Zeit mit der Suche und dem Abgleich von Daten verbringen. Um sich gegen Qualitätsprobleme abzusichern, schaffen Unternehmen komplexe Strukturen für das Informationsmanagement, in denen eine Vertikale von Managern für die Suche, die Überprü-

fung und den Abgleich von Daten zuständig ist. Dieser Ansatz erhöht jedoch nur den bürokratischen Aufwand und verlangsamt die Entscheidungsfindung. Je mehr Daten vorhanden sind, desto schwieriger ist es, sie zu analysieren und zu interpretieren, vor allem, wenn es keinen einheitlichen Standard für die Speicherung und Verarbeitung der Daten gibt.

Mit der Fülle von Softwareanwendungen und -systemen, die im letzten Jahrzehnt wie Pilze aus dem Boden schossen, hat das Problem der Silos und der unangemessenen Datenqualität für die Endnutzer an Bedeutung gewonnen. Dieselben Daten, aber mit unterschiedlichen Werten, können nun in verschiedenen Systemen und Anwendungen zu finden sein (Abb. 2.1-6). Dies führt zu Schwierigkeiten für die Endnutzer, wenn sie versuchen zu bestimmen, welche Version der Daten unter den vielen verfügbaren relevant und korrekt ist. Dies führt zu Fehlern bei der Analyse und letztlich bei der Entscheidungsfindung.

Um sich gegen Probleme beim Auffinden der richtigen Daten abzusichern, schaffen die Unternehmensleiter eine mehrstufige Bürokratie von Prüfungsverantwortlichen. Ihre Aufgabe ist es, die erforderlichen Daten in Form von Tabellen und Berichten schnell zu finden, zu prüfen und zu übermitteln, wobei sie sich durch das Labyrinth der verschiedenen Systeme bewegen müssen.

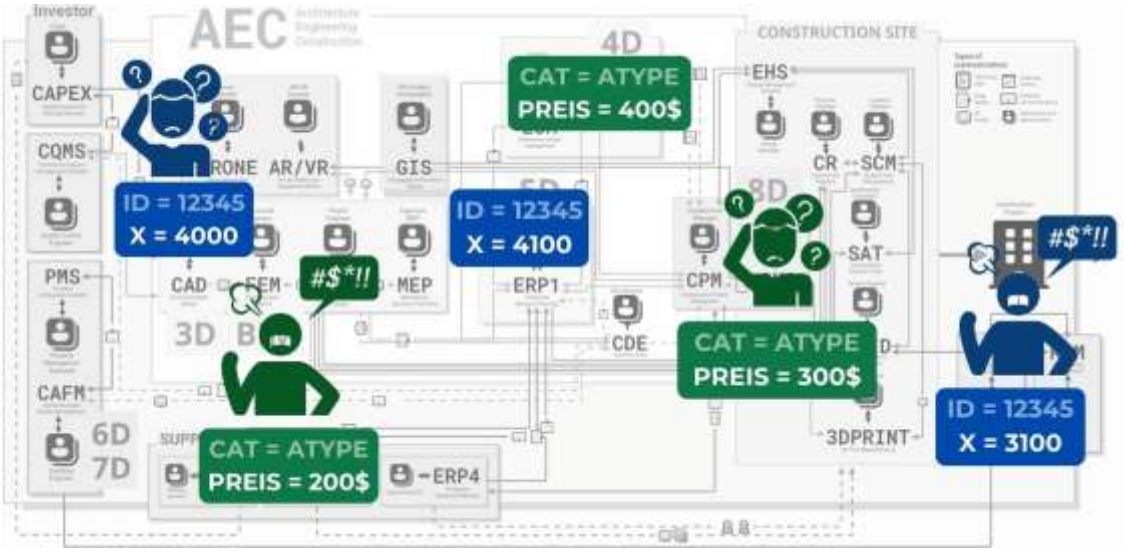


Abb. 2.1-6 Bei dem Versuch, die richtigen Daten zu finden, müssen die Verantwortlichen die Qualität und die rechtliche Zuverlässigkeit der Daten zwischen den verschiedenen Systemen sicherstellen.

In der Praxis führt dieses Modell jedoch zu neuen Komplexitäten. Wenn Daten manuell verwaltet werden und Informationen über viele nicht zusammenhängende Entscheidungen verstreut sind, wird jeder Versuch, genaue und aktuelle Informationen über eine Pyramide von Entscheidungsträgern (Abb. 2.1-7) zu erhalten, zu einem Engpass - zeitraubend und fehleranfällig.

Die Situation wird durch die Lawine digitaler Lösungen noch verschärft. Der Softwaremarkt wird weiterhin mit neuen Tools überschwemmt, die vielversprechend erscheinen. Aber ohne eine klare Strategie für das Datenmanagement lassen sich diese Lösungen nicht in ein einheitliches System integrieren, sondern schaffen stattdessen zusätzliche Schichten von Komplexität und Doppelarbeit. Anstatt die Prozesse zu vereinfachen, finden sich die Unternehmen in einer noch stärker fragmentierten und chaotischen Informationsumgebung wieder.

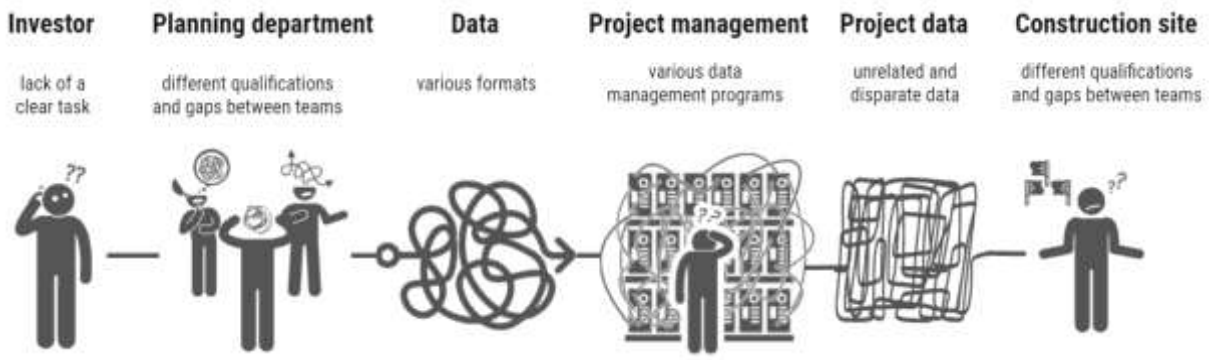


Abb. 2.1-7 Die Komplexität der Systeme und die Vielfalt der Datenformate führen zu einem Verlust an Konsistenz im Bauprozess.

All diese Probleme, die mit der Verwaltung einer Vielzahl unterschiedlicher Lösungen verbunden sind, bringen die Unternehmensleitung früher oder später zu einer wichtigen Erkenntnis: Es geht nicht um die Menge der Daten oder darum, das nächste "Einheitswerkzeug" zu finden, um sie zu verarbeiten. Der wahre Grund liegt in der Qualität der Daten und in der Art und Weise, wie das Unternehmen sie erstellt, empfängt, speichert und nutzt.

> Der Schlüssel zum nachhaltigen Erfolg liegt nicht in der Jagd nach neuen "magischen" Anwendungen, sondern im Aufbau einer Datenkultur im Unternehmen. Dies bedeutet, dass Daten als strategisches Gut behandelt werden und dass Datenqualität, -integrität und -relevanz auf allen Ebenen des Unternehmens Priorität haben.

Die Lösung des Dilemmas zwischen Qualität und Quantität liegt in der Schaffung einer einheitlichen Datenstruktur, die Doppelarbeit und Inkonsistenzen beseitigt und den Informationsfluss vereinheitlicht. Diese Architektur bietet eine einzige, zuverlässige Datenquelle, auf deren Grundlage fundierte, genaue und zeitnahe Entscheidungen getroffen werden können.

Andernfalls verlassen sich die Unternehmen nach wie vor häufig auf die subjektiven Meinungen und intuitiven Einschätzungen der HiPPO-Experten und nicht auf verlässliche Fakten. Im Baugewerbe, wo Fachwissen traditionell eine große Rolle spielt, ist dies besonders auffällig.

HiPPO oder die Gefahr von Meinungen bei der Entscheidungsfindung

Im Baugewerbe werden wichtige Entscheidungen traditionell auf der Grundlage von Erfahrung und subjektiven Einschätzungen getroffen. Ohne zeitnahe und zuverlässige Daten müssen die Unternehmensleiter blind handeln und sich auf die Intuition der bestbezahlten Mitarbeiter (HiPPO - Highest Paid Person's Opinion) statt auf objektive Fakten verlassen (Abb. 2.1-8).

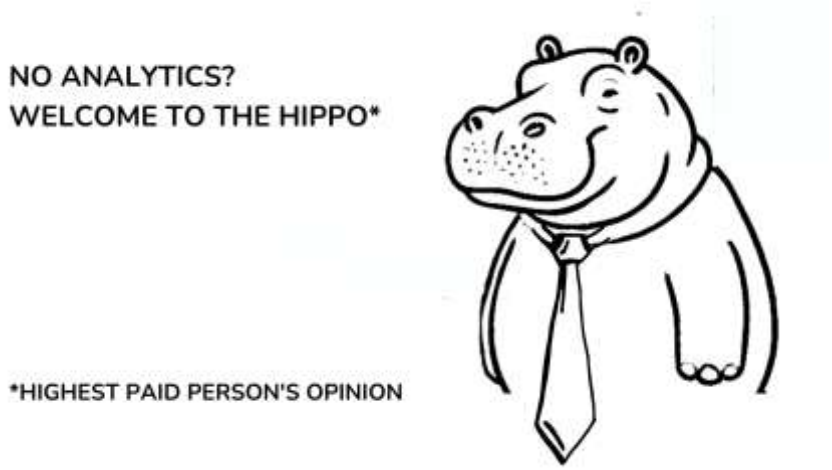


Abbildung 2.1-8 In Ermangelung von Analytik hängt das Geschäft von der subjektiven Meinung erfahrener Fachleute ab.

Ein solcher Ansatz mag in einem stabilen und sich langsam verändernden Umfeld gerechtfertigt sein, aber in einer Ära des digitalen Wandels wird er zu einem ernsthaften Risiko. Entscheidungen, die auf Intuition und Vermutungen beruhen, sind anfällig für Verzerrungen, beruhen oft auf unbegründeten Hypothesen und berücksichtigen nicht das komplexe Bild, das sich in den Daten widerspiegelt

> Was auf der Entscheidungsebene in einem Unternehmen als intelligente Debatte ausgegeben wird, hat oft keine konkrete Grundlage. Der Erfolg eines Unternehmens sollte nicht von der Autorität und dem Gehalt von Experten abhängen, sondern von der Fähigkeit, effektiv mit Daten zu arbeiten, Muster zu erkennen und fundierte Entscheidungen zu treffen.

Es ist wichtig, sich von der Vorstellung zu verabschieden, dass Autorität oder Erfahrung automatisch bedeutet, dass eine Entscheidung richtig ist. Der datengesteuerte Ansatz ist ein Wendepunkt: Daten und Analysen, nicht Position und Gehalt, sind jetzt die Grundlage für Entscheidungen. Big Data, maschinelles Lernen und visuelle Analysen ermöglichen es uns, Muster zu erkennen und uns auf Fakten statt auf Vermutungen zu stützen (Abb. 1.1-4).

> *Ohne Daten sind Sie nur ein weiterer Mensch mit einer Meinung [34].*
>
> - W. Edwards Deming, Wissenschaftler und Unternehmensberater

Moderne Datenmanagementmethoden sichern auch die Kontinuität des Wissens im Unternehmen. Klar beschriebene Prozesse, Automatisierung und ein systematisches Vorgehen ermöglichen es, auch Schlüsselrollen ohne Effizienzverlust zu übertragen.

Blindes Vertrauen in Daten kann jedoch auch zu schwerwiegenden Fehlern führen. Daten selbst sind nur eine Ansammlung von Zahlen. Ohne angemessene Analyse, Kontext und die Fähigkeit, Muster zu erkennen, haben sie keinen Wert und können keine Prozesse vorantreiben. Der Schlüssel zum Erfolg liegt nicht in der Wahl zwischen HiPPO-Intuition und Analytik, sondern in der Entwicklung intelligenter Werkzeuge, die disparate Informationen in handhabbare, fundierte Entscheidungen umwandeln.

> In einer digitalen Bauumgebung werden nicht Dienstalter und Rang in der Hierarchie zu entscheidenden Erfolgsfaktoren, sondern Reaktionsfähigkeit, Entscheidungsgenauigkeit und Ressourceneffizienz

Daten sind Werkzeuge, keine absoluten Wahrheiten. Sie sollten das menschliche Denken ergänzen, nicht ersetzen. Trotz der Vorteile der Analytik können Daten die menschliche Intuition und Erfahrung nicht vollständig ersetzen. Sie sollen helfen, genauere und fundiertere Entscheidungen zu treffen.

Wettbewerbsvorteile werden nicht nur durch die Einhaltung von Standards erzielt, sondern auch dadurch, dass man in der Lage ist, seine Konkurrenten durch die effiziente Nutzung von Ressourcen, die für alle gleich sind, zu übertreffen. In Zukunft werden Datenkenntnisse genauso wichtig sein wie früher Lese- und Schreibkenntnisse oder mathematische Fähigkeiten. Fachleute, die Daten analysieren und interpretieren können, werden in der Lage sein, präzisere Entscheidungen zu treffen und diejenigen zu verdrängen, die sich nur auf ihre persönliche Erfahrung verlassen (Abb. 2.1-9).

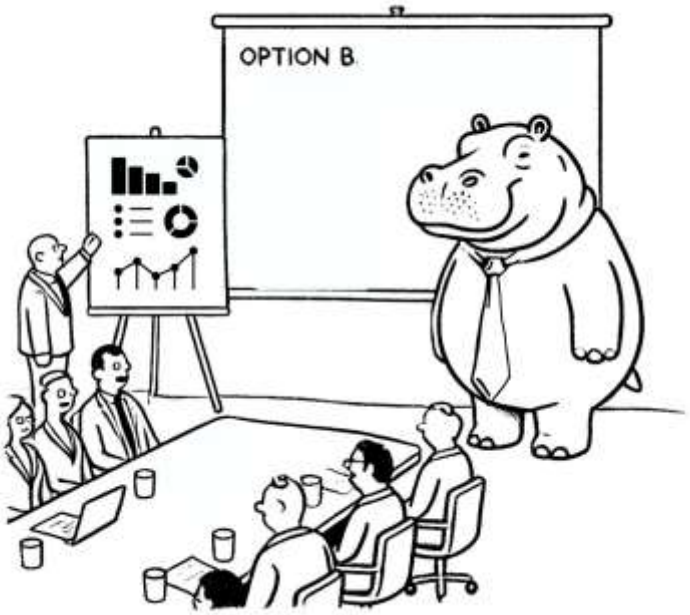


Abb. 2.1-9 Entscheidungen sollten auf einer objektiven Analyse beruhen und nicht auf der Meinung des höchstbezahlten Mitarbeiters.

Manager, Spezialisten und Ingenieure werden als Datenanalysten fungieren und die Struktur, Dynamik und Schlüsselindikatoren von Projekten untersuchen. Die Humanressourcen werden zu Elementen des Systems, die eine flexible, datengestützte Anpassung erfordern, um die Effizienz zu maximieren.

> *Die Fehler bei der Verwendung unzureichender Daten sind viel kleiner als bei der Verwendung keiner Daten [35].*
>
> - Charles Babbage, Erfinder der ersten analytischen Rechenmaschine

Das Aufkommen von Big Data und die Einführung von LLM (Large Language Models) haben nicht nur die Art und Weise, wie wir analysieren, sondern auch die Art der Entscheidungsfindung selbst radikal verändert. Während früher der Schwerpunkt auf der Kausalität lag (warum etwas passiert ist - diagnostische Analytik) (Abb. 1.1-4), rückt heute die Fähigkeit zur Vorhersage der Zukunft (prädiktive Analytik) und in Zukunft die präskriptive Analytik, bei der maschinelles Lernen und KI die beste Wahl im Entscheidungsprozess vorschlagen, in den Vordergrund.

> Laut der neuen SAP™-Studie "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36] wären 44% der Führungskräfte bereit, ihre bisherige Entscheidung auf der Grundlage von KI-Ratschlägen zu ändern, und 38% würden der KI vertrauen, dass sie in ihrem Namen Geschäftsentscheidungen trifft. Inzwischen gaben 74% der Führungskräfte an, dass sie KI-Ratschlägen mehr vertrauen als ihren Freunden und ihrer Familie, und 55% arbeiten in Unternehmen, in denen von KI abgeleitete Erkenntnisse herkömmliche Entscheidungsfindungsmethoden ersetzen oder häufig umgehen - insbesondere in Organisationen mit einem Jahresumsatz von über 5 Mrd. US-Dollar. Darüber hinaus nutzen 48% der Befragten täglich generative KI-Tools, 15% davon sogar mehrmals täglich.

Mit der Entwicklung von LLM und automatisierten Datenverwaltungssystemen stellt sich eine neue Herausforderung: Wie können Informationen effektiv genutzt werden, ohne dass ihr Wert im Chaos inkompatibler Formate und heterogener Quellen verloren geht, was durch die zunehmende Komplexität und Dynamik der Geschäftsprozesse noch verstärkt wird.

Kontinuierliche Zunahme der Komplexität und Dynamik der Geschäftsprozesse

Das Baugewerbe steht heute vor großen Herausforderungen beim Daten- und Prozessmanagement. Die größten Herausforderungen sind siloartige Informationssysteme, übermäßige Bürokratie und mangelnde Integration digitaler Werkzeuge. Diese Herausforderungen verschärfen sich noch, da die Geschäftsprozesse selbst immer komplexer werden - angetrieben durch Technologie, sich ändernde Kundenanforderungen und sich entwickelnde Vorschriften.

Die Einzigartigkeit von Bauprojekten ergibt sich nicht nur aus ihren technischen Besonderheiten, sondern auch aus den unterschiedlichen nationalen Normen und gesetzlichen Anforderungen in den verschiedenen Ländern (Abb. 4.2-10, Abb. 5.1-7). Dies erfordert eine flexible, individuelle Herangehensweise an jedes Projekt, die im Rahmen traditioneller modularer Kontrollsysteme nur schwer zu realisieren ist. Aufgrund der Komplexität der Prozesse und der großen Datenmengen wenden sich viele Unternehmen an Anbieter von Speziallösungen. Doch der Markt ist überladen - viele Start-ups bieten ähnliche Produkte an, die sich auf enge Aufgaben konzentrieren. Dadurch geht ein ganzheitlicher Ansatz für das Datenmanagement oft verloren.

> Die Anpassung an den ständigen Fluss neuer Technologien und Marktanforderungen wird zu einem entscheidenden Faktor für die Wettbewerbsfähigkeit. Bestehende proprietäre Anwendungen und modulare Systeme weisen jedoch eine geringe Anpassungsfähigkeit auf - Änderungen erfordern oft langwierige und kostspielige Überarbeitungen durch Entwickler, die die Besonderheiten der Bauprozesse nicht immer verstehen.

Die Unternehmen sehen sich als Geisel des technologischen Rückstands und warten auf neue Updates,

anstatt innovative integrierte Ansätze umgehend umzusetzen. Infolgedessen ist die interne Struktur von Bauunternehmen oft ein komplexes Ökosystem miteinander verbundener, hierarchischer und oft geschlossener Systeme, die durch ein mehrstufiges Netzwerk von Managern koordiniert werden (Abb. 2.1-10).

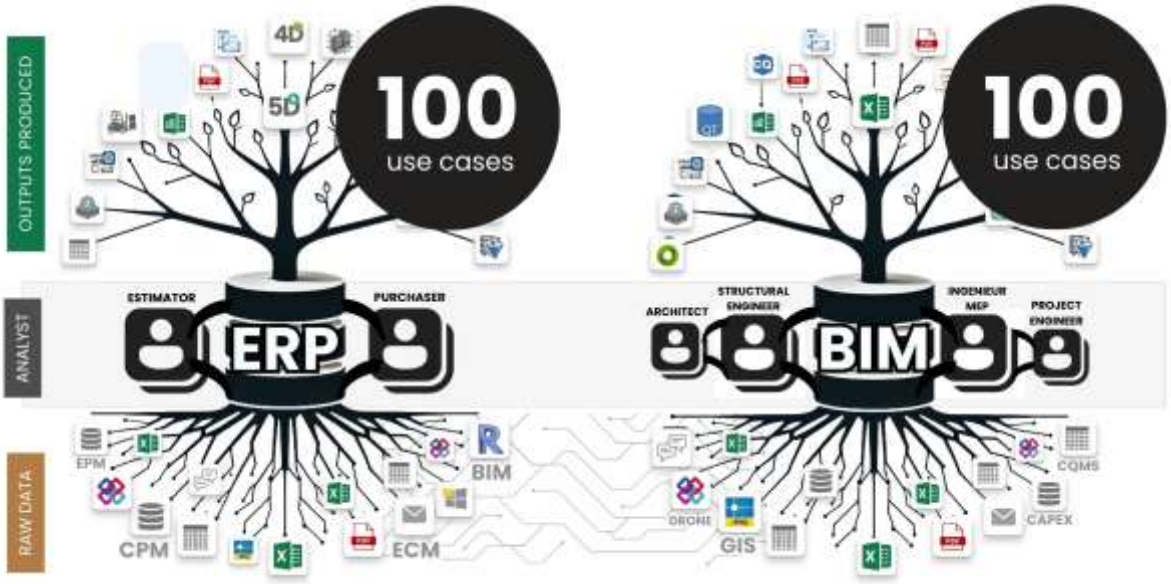


Abb. 2.1-10 Unternehmen bestehen aus miteinander verbundenen Systemen, deren Verknüpfung Prozesse bildet, die eine Automatisierung erfordern.

> Laut einer Umfrage des kanadischen Bauverbands und von KPMG Canada aus dem Jahr 2021 [37] glauben nur 25% der Unternehmen, dass sie sich im Vergleich zu ihren Mitbewerbern in einer bedeutenden oder anderen Position befinden, wenn es um die Einführung von Technologien oder digitalen Lösungen geht. Nur 23% der Befragten gaben an, dass ihre Lösungen in erheblichem oder hohem Maße datengesteuert sind. Gleichzeitig bezeichnete die Mehrheit der Umfrageteilnehmer die Nutzung einer Reihe anderer Technologien als rein experimentell oder gab zu, sie überhaupt nicht zu nutzen.

Diese Zurückhaltung gegenüber technologischen Experimenten zeigt sich besonders bei großen Infrastrukturprojekten, bei denen Fehler Millionen von Dollar kosten können. Selbst die fortschrittlichsten Technologien - digitale Zwillinge, prädiktive Analytik - stoßen oft auf Widerstand, nicht wegen ihrer Wirksamkeit, sondern wegen der mangelnden erwiesenen Zuverlässigkeit bei realen Projekten.

> Laut dem Bericht des Weltwirtschaftsforums (WEF) "Shaping the Future of Construction" [5] stößt die Einführung neuer Technologien im Bauwesen nicht nur auf technische Schwierigkeiten, sondern auch auf psychologische Barrieren auf Seiten der Kunden Die Einführung neuer Technologien im Bauwesen stößt nicht nur auf technische Schwierigkeiten, sondern auch auf psychologische Barrieren auf Seiten der Kunden [5]. Viele Bauherren befürchten, dass der Einsatz fortschrittlicher Lösungen ihre Projekte zu einem Experimentierfeld macht und sie zu "Versuchskaninchen" werden, wobei unvorhersehbare Folgen zu zusätzlichen Kosten und Risiken führen können.

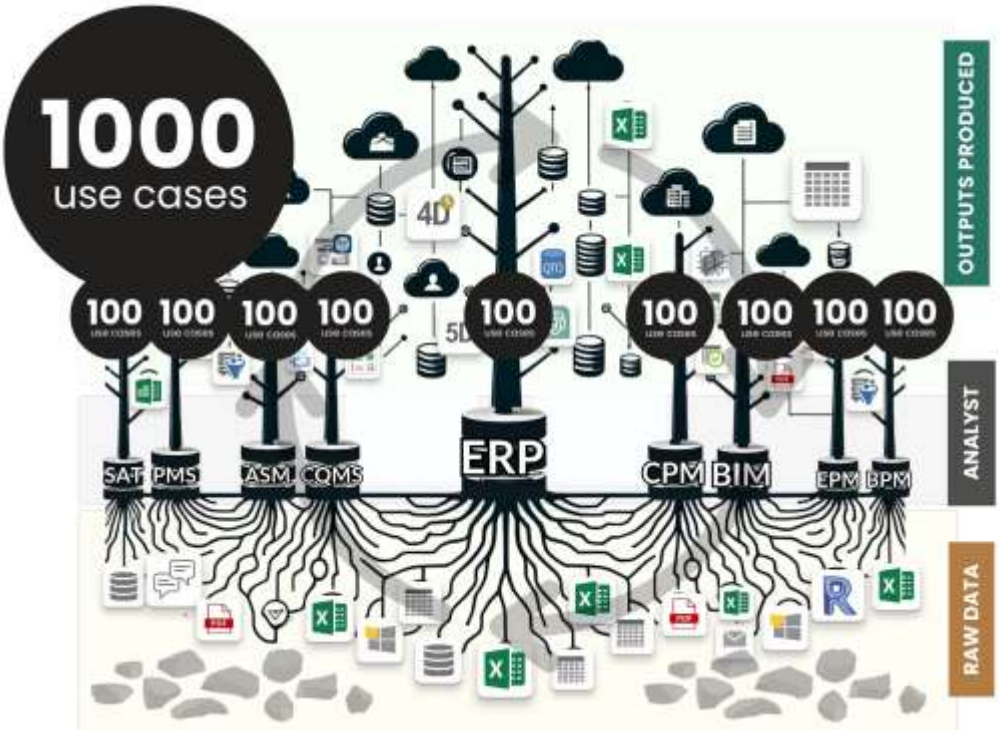


Abb. 2.1-11 Für jeden Datennutzungsfall bietet der Lösungsmarkt Anwendungen zur Optimierung und Automatisierung von Prozessen.

> Die Baubranche ist sehr vielfältig: verschiedene Projekte haben unterschiedliche Anforderungen, regionale Besonderheiten, gesetzliche Klassifizierungsvorschriften (Abb. 4.2-10), Berechnungsstandards (Abb. 5.1-7) usw. Daher ist es praktisch unmöglich, eine eigene universelle Anwendung oder ein System zu entwickeln, das all diesen Anforderungen und Projektspezifika perfekt gerecht wird.

Bei dem Versuch, die zunehmende Komplexität der Systeme und die Abhängigkeit von Softwareanbietern in den Griff zu bekommen, setzt sich immer mehr die Erkenntnis durch, dass der Schlüssel zu einem effizienten Datenmanagement nicht nur in der Offenheit und Standardisierung, sondern auch in der Vereinfachung der Prozessarchitektur selbst liegt. Die zunehmende Komplexität und Dynamik der Geschäftsprozesse erfordert neue Ansätze, bei denen die Priorität von der Datenakkumulation auf die Strukturierung und Organisation der Daten verlagert wird. Diese Verlagerung wird der nächste Schritt in der Entwicklung der Bauindustrie sein und das Ende der Ära der Dominanz der Softwareanbieter und den Beginn der Ära einer sinnvollen Informationsorganisation markieren.

Die Einsicht in die Grenzen von Einheitslösungen und die Anfälligkeit für zunehmende Komplexität führen zu einer Verlagerung der Prioritäten von geschlossenen Plattformen und Datenhortung hin zu Transparenz, Anpassungsfähigkeit und strukturierter Informationsverarbeitung. Dieses Umdenken spiegelt die breiteren Veränderungen in der globalen Wirtschaft und Technologie wider, die durch die Linse der so genannten "industriellen Revolutionen" beschrieben werden. Um zu verstehen, wohin sich das Bauwesen bewegt und welche Richtung es in Zukunft einschlagen wird, muss man die Stellung der Branche im Kontext der vierten und fünften industriellen Revolution betrachten - von der Automatisierung und Digitalisierung bis hin zur Personalisierung, offenen Standards und dem dienstleistungsbasierten Datenmodell.

Die vierte industrielle Revolution (Industrie 4.0) und die fünfte industrielle Revolution (Industrie 5.0) im Bauwesen

Technologische und wirtschaftliche Phasen sind theoretische Konzepte, die zur Beschreibung und Analyse der Entwicklung von Gesellschaft und Wirtschaft in verschiedenen Entwicklungsstadien verwendet werden. Sie können von verschiedenen Forschern und Experten unterschiedlich interpretiert werden.

- **Die vierte industrielle Revolution** (4IR oder Industrie 4.0) steht im Zusammenhang mit Informationstechnologie, Automatisierung, Digitalisierung und Globalisierung. Eines ihrer Schlüsselelemente ist die Schaffung proprietärer Softwarelösungen, d. h. spezialisierter digitaler Produkte, die für bestimmte Aufgaben und Unternehmen entwickelt wurden. Diese Lösungen werden oft zu einem wichtigen Teil der IT-Infrastruktur, sind aber ohne zusätzliche Änderungen nur schlecht skalierbar.
- **Die Fünfte Industrielle Revolution (5IR)** befindet sich jetzt in einem früheren Stadium der Konzeption und Entwicklung als die 4IR. Zu ihren Grundprinzipien gehört eine stärkere Personalisierung von Produkten und Dienstleistungen. Die 5IR ist eine Bewegung hin zu einer anpassungsfähigeren, flexibleren und stärker personalisierten Wirtschaftstätigkeit mit dem Schwerpunkt auf Personalisierung, Beratung und dienstleistungsorientierten Modellen. Ein Schlüsselaspekt der fünften Wirtschaftsform ist die Nutzung von Daten für die Entscheidungsfindung, die ohne den Einsatz offener Daten und offener Werkzeuge praktisch unmöglich ist (Abb. 2.1-12).

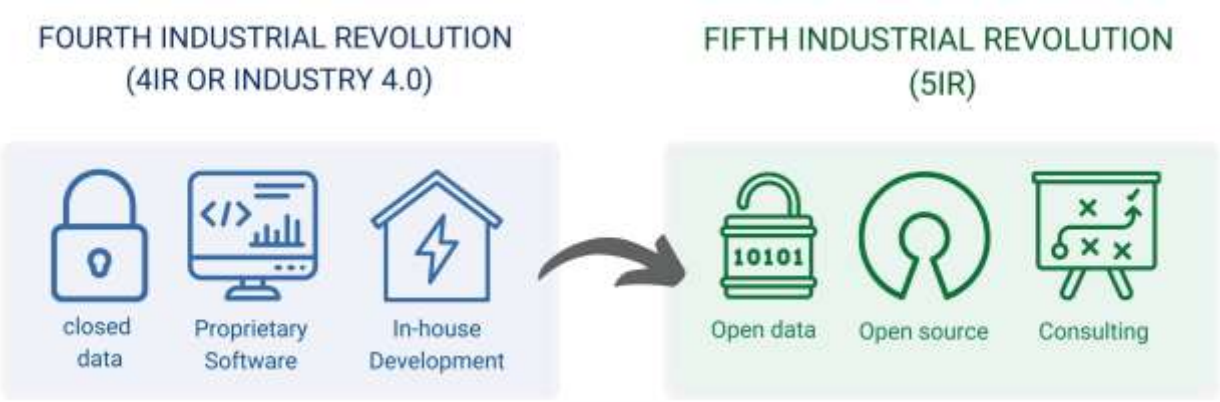


Abb. 2.1-12 Das vierte Muster konzentriert sich auf Lösungen, während sich das fünfte Muster auf Personalisierung und Daten konzentriert.

> Die Entwicklung einer Anwendung für Unternehmen der Baubranche, die in zehn oder hundert Organisationen eingesetzt werden kann, garantiert nicht, dass sie ohne erhebliche Änderungen und Erweiterungen auf andere Unternehmen, Regionen oder Länder übertragen werden kann. Die Wahrscheinlichkeit einer erfolgreichen Skalierung solcher Lösungen bleibt gering, da jedes Unternehmen einzigartige Prozesse, Anforderungen und Bedingungen hat, die möglicherweise individuelle Anpassungen erfordern.

Es ist wichtig zu verstehen, dass die erfolgreiche Integration technologischer Lösungen bereits heute einen zutiefst individuellen Ansatz für jeden Prozess, jedes Projekt und jedes Unternehmen voraussetzt. Das bedeutet, dass selbst nach der Entwicklung eines universellen Rahmens, Werkzeugs oder Programms eine

detaillierte Anpassung und Individualisierung erforderlich ist, um die einzigartigen Anforderungen und Bedingungen jedes spezifischen Unternehmens und Projekts zu erfüllen.

> Laut dem PwC-Bericht "Decoding the Fifth Industrial Revolution" [38] [38] verlassen sich in diesem Jahr etwa 50% der Führungskräfte in verschiedenen Branchen auf die Integration von Spitzentechnologie und menschlichem Fachwissen. Dieser Ansatz ermöglicht es ihnen, sich schnell an Änderungen im Produktdesign oder bei den Kundenanforderungen anzupassen und eine personalisierte Produktion zu schaffen.

Jeder Prozess erfordert die Entwicklung einer einzigartigen Funktion oder Anwendung, was angesichts der Größe der globalen Bauindustrie und der Vielfalt der Projekte zur Existenz einer riesigen Anzahl von Geschäftsfällen führt, die jedes Mal eine einzigartige Pipeline Logik darstellen (Abb. 2.1-13). Jeder dieser Fälle hat seine eigenen Besonderheiten und erfordert einen maßgeschneiderten Ansatz. Im Kapitel über maschinelles Lernen und das Parsing des Titanic-Datensatzes (Abb. 9.2-9) werden wir die Vielfalt möglicher Lösungen für ein und dasselbe analytische Problem im Zusammenhang mit verschiedenen Ansätzen genauer betrachten.

> **Pipeline** ist im Zusammenhang mit digitalen Prozessen eine Abfolge von Aktivitäten, Prozessen und Tools, die einen automatisierten oder strukturierten Daten- und Arbeitsfluss durch die verschiedenen Phasen des Projektlebenszyklus ermöglichen.

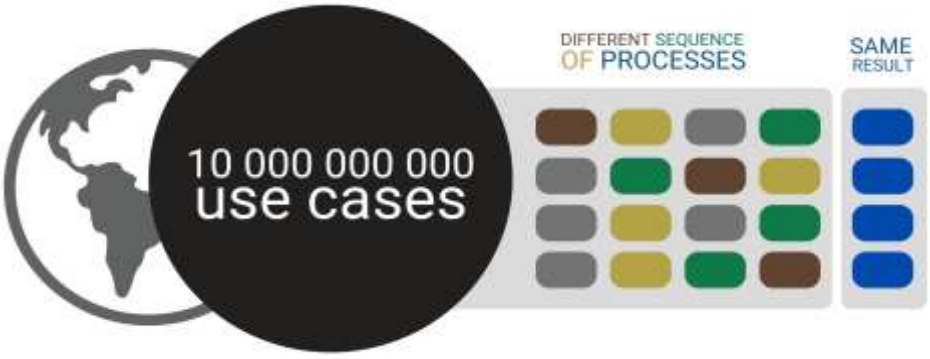


Abb. 2.1-13 Die Individualität und Variabilität von Geschäftsfällen macht den Versuch unmöglich, skalierbare, geschlossene Plattformen und Werkzeuge zu schaffen.

Unser Leben hat sich unter dem Einfluss der digitalen Transformation bereits in vielerlei Hinsicht verändert, und heute können wir vom Beginn einer neuen Phase in der wirtschaftlichen Entwicklung der Bauwirtschaft sprechen. In dieser "neuen Wirtschaft" wird der Wettbewerb nach anderen Regeln organisiert sein: Wer in der Lage ist, öffentliches Wissen und offene Daten effizient in nachgefragte Produkte und Dienstleistungen umzuwandeln, gewinnt unter den Bedingungen der fünften industriellen Revolution einen entscheidenden Vorteil.

> Wie die Wirtschaftswissenschaftlerin Kate Maskus in ihrem Buch "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39] feststellt 2012 [39] fest: "Wir leben in einer globalen Wissensökonomie, und die Zukunft gehört denjenigen, die es verstehen, wissenschaftliche Entdeckungen in Waren zu verwandeln".

Der Übergang zum fünften Wirtschaftsmodus bedeutet eine Verlagerung des Schwerpunkts von geschlossenen IT-Lösungen zu offenen Standards und Plattformen. Die Unternehmen werden beginnen, sich von traditionellen Softwareprodukten zugunsten von dienstleistungsorientierten Modellen zu lösen, bei denen Daten und nicht proprietäre Technologien zum wichtigsten Vermögenswert werden.

> Die Studie der Harvard Business School 2024 [40] zeigt den enormen wirtschaftlichen Wert von Open-Source-Software (Open Source Software, OSS). Der Studie zufolge ist OSS in 96% aller Softwarecodes enthalten, und manche kommerzielle Software besteht zu 99,9% aus OSS-Komponenten. Ohne OSS würden die Unternehmen 3,5 Mal mehr für Software ausgeben.

> Der Aufbau von Unternehmens-Ökosystemen wird, globalen Trends folgend, allmählich zu einem fünften wirtschaftlichen Paradigma übergehen, bei dem datenzentrierte Analyse- und Beratungsdienste eine höhere Priorität erhalten als isolierte, geschlossene Lösungen mit starr definierten Nutzungsszenarien.

Das Zeitalter der Digitalisierung wird das Kräfteverhältnis in der Branche verändern: Statt sich auf Anbieterlösungen zu verlassen, werden Unternehmen ihre Wettbewerbsfähigkeit auf ihre Fähigkeit zur effektiven Datennutzung stützen. Infolgedessen wird sich das Baugewerbe von starren Altsystemen zu flexiblen, anpassungsfähigen Ökosystemen entwickeln, in denen offene Standards und interoperable Tools die Grundlage des Projektmanagements bilden werden. Das Ende der Ära der Dominanz von Anwendungsanbietern wird ein neues Umfeld schaffen, in dem der Wert nicht durch den Besitz von geschlossenem Quellcode und spezialisierten Konnektoren definiert wird, sondern durch die Fähigkeit, Daten in einen strategischen Vorteil zu verwandeln.

KAPITEL 2.2.
CHAOS IN ORDNUNG BRINGEN UND KOMPLEXITÄT REDUZIEREN

Redundanter Code und geschlossene Systeme als Hindernis für Produktivitätssteigerungen

In den vergangenen Jahrzehnten wurde der technologische Wandel im IT-Bereich vor allem von Softwareanbietern vorangetrieben. Sie gaben den Kurs der Entwicklung vor und bestimmten, welche Technologien die Unternehmen übernehmen sollten und welche zurückgelassen werden sollten. In der Ära des Übergangs von Insellösungen zu zentralisierten Datenbanken und integrierten Systemen warben die Anbieter für lizenzierte Produkte, die die Kontrolle über den Zugang und die Skalierbarkeit ermöglichten. Später, mit dem Aufkommen von Cloud-Technologien und Software-as-a-Service-Modellen (SaaS), entwickelte sich diese Kontrolle zu einem Abonnementmodell, das die Nutzer als treue Kunden digitaler Dienste festigte.

Dieser Ansatz hat zu einem Paradoxon geführt: Trotz der noch nie dagewesenen Menge an erstelltem Programmcode wird nur ein kleiner Teil davon tatsächlich genutzt. Vielleicht gibt es hunderte oder tausende Male mehr Code als nötig, weil die gleichen Geschäftsprozesse in Dutzenden oder Hunderten von Programmen auf unterschiedliche Weise beschrieben und dupliziert werden, sogar innerhalb desselben Unternehmens. Gleichzeitig sind die Entwicklungskosten bereits bezahlt, und diese Kosten sind nicht mehr einbringbar. Dennoch fährt die Industrie fort, diesen Kreislauf zu reproduzieren, indem sie neue Produkte mit minimalem Mehrwert für den Endverbraucher schafft, und zwar häufiger unter dem Druck der Markterwartungen als der tatsächlichen Bedürfnisse.

> Nach dem Defence Acquisition University (DAU) Software Development Cost Estimating Guide [41] können die Kosten für die Softwareentwicklung in Abhängigkeit von mehreren Faktoren, darunter die Komplexität des Systems und die gewählte Technologie, erheblich schwanken. In der Vergangenheit lagen die Entwicklungskosten für 2008 bei etwa 100 $ pro Quellcodezeile (SLOC), während die Wartungskosten bis zu 4.000 $ pro SLOC betragen können.

Allein eine der Komponenten von CAD-Anwendungen - der geometrische Kern - kann mehrere zehn Millionen Codezeilen umfassen (Abb. 6.1-5). Eine ähnliche Situation ist bei ERP-Systemen zu beobachten (Abb. 5.4-4), auf deren Komplexität wir im fünften Teil des Buches zurückkommen werden. Bei näherer Betrachtung zeigt sich jedoch, dass ein Großteil dieses Codes keinen Mehrwert schafft, sondern lediglich als "Briefträger" fungiert, der mechanisch Daten zwischen der Datenbank, der API, der Benutzeroberfläche und anderen Tabellen im System bewegt. Trotz des weit verbreiteten Mythos über die entscheidende Bedeutung der so genannten Geschäftslogik ist die harte Realität viel prosaischer: Moderne Codebasen sind voll von veralteten Vorlagenblöcken (Legacy-Code), deren einziger Zweck darin besteht, den Datentransfer zwischen Tabellen und Komponenten zu gewährleisten, ohne die Entscheidungsfindung oder die Geschäftseffizienz zu beeinflussen.

Geschlossene Lösungen, die Daten aus verschiedenen Quellen verarbeiten, werden so unweigerlich zu verwirrenden "Spaghetti-Ökosystemen". Diese komplexen, miteinander verflochtenen Systeme können nur von einem Heer von Managern bewältigt werden, die in einem halbroutinemäßigen Modus arbeiten. Diese Organisation der Datenverwaltung ist nicht nur in Bezug auf die Ressourcen ineffizient, sondern schafft auch kritische Schwachstellen in den Geschäftsprozessen und macht das Unternehmen von einem engen Kreis von Spezialisten abhängig, die wissen, wie dieses technologische Labyrinth funktioniert.

Die kontinuierliche Zunahme der Codemenge, der Anzahl der Anwendungen und der zunehmenden Komplexität der von den Anbietern angebotenen Konzepte hat zu einem natürlichen Ergebnis geführt - einer Zunahme der Komplexität des IT-Ökosystems im Bauwesen. Dies hat dazu geführt, dass die praktische Umsetzung der Digitalisierung durch die Erhöhung der Anzahl der Anwendungen in der Branche unwirksam geworden ist. Softwareprodukte, die ohne Rücksicht auf die Bedürfnisse der Nutzer entwickelt werden, erfordern oft erhebliche Ressourcen für die Implementierung und den Support, bringen aber nicht die erwarteten Erträge.

> Laut der McKinsey-Studie "Increasing Construction Productivity" [42] betrug das weltweite Wachstum der Arbeitsproduktivität im Baugewerbe in den letzten zwei Jahrzehnten durchschnittlich nur 1% pro Jahr, verglichen mit einem Wachstum von 2,8% für die Weltwirtschaft insgesamt und 3,6% für das verarbeitende Gewerbe. In den Vereinigten Staaten hat sich die Arbeitsproduktivität pro Arbeitnehmer im Baugewerbe seit den 1960er Jahren halbiert [43].

Die zunehmende Komplexität der Systeme, die Isolierung und die Verschlossenheit der Daten haben die Kommunikation zwischen den Fachleuten beeinträchtigt, so dass die Bauindustrie zu den am wenigsten effizienten Branchen gehört (Abb. 2.2-1). bis 2040 auf 22 Billionen Dollar ansteigen, was erhebliche Effizienzsteigerungen erfordert.

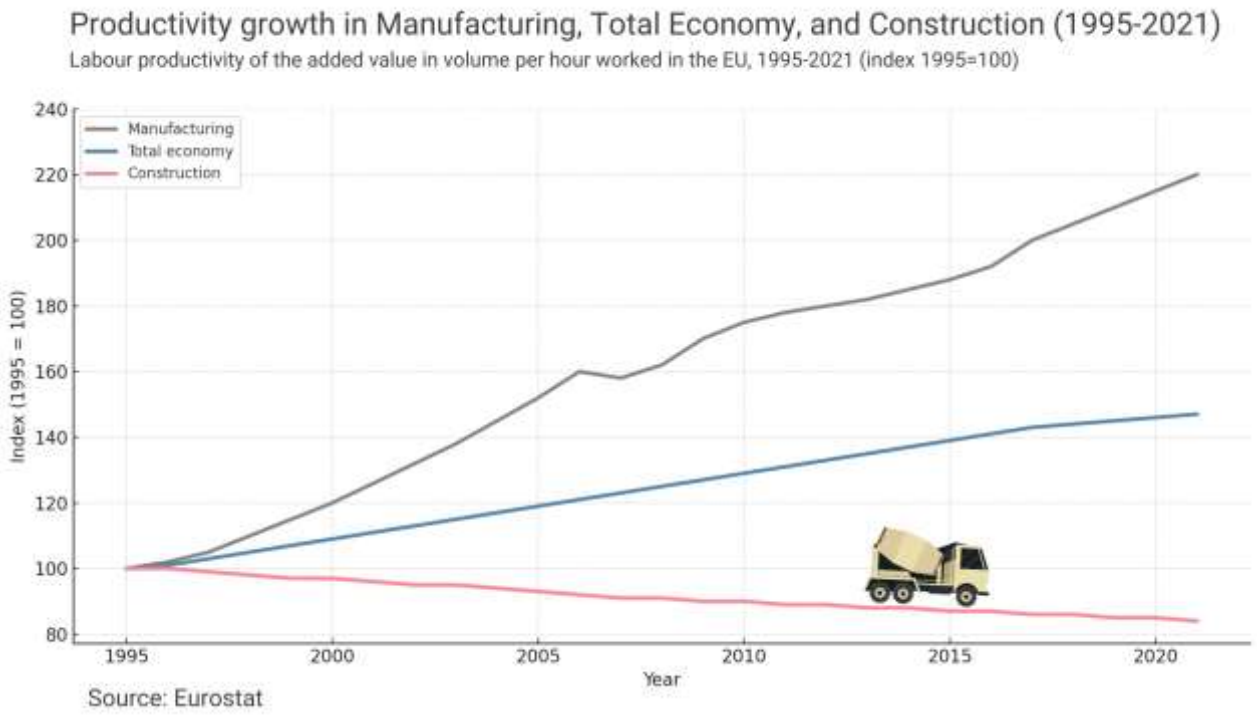


Abb. 2.2-1 Geschlossene und komplexe Daten und infolgedessen schlechte Kommunikation zwischen Fachleuten haben das Baugewerbe zu einem der am wenigsten effizienten Sektoren der Wirtschaft gemacht (basierend auf [44], [45]).

> Wie in der McKinsey-Studie (2024) "Ensuring construction productivity is no longer optional" betont wird, kann es sich das Baugewerbe angesichts der zunehmenden Ressourcenknappheit und des Bestrebens der Branche, ihre Wachstumsrate zu verdoppeln, nicht länger leisten, auf dem derzeitigen Produktivitätsniveau zu verharren [44]. Es wird prognostiziert, dass die weltweiten Baukosten von 13 Billionen Dollar im Jahr 2023 bis zum Ende des Jahrzehnts auf ein viel höheres Niveau ansteigen werden, was die Frage der Effizienz nicht nur relevant, sondern entscheidend macht.

Einer der wichtigsten Wege zur Verbesserung der Effizienz wird die unvermeidliche Vereinheitlichung und Vereinfachung der Anwendungsstrukturen und der Architekturen des Datenökosystems sein. Dieser Rationalisierungsansatz wird überflüssige Abstraktionsebenen und unnötige Komplexität beseitigen, die sich im Laufe der Jahre in den Unternehmenssystemen angesammelt haben.

Von Silos zu einem einzigen Data Warehouse

Je mehr Daten ein Unternehmen anhäuft, desto schwieriger wird es, daraus einen echten Nutzen zu ziehen. Aufgrund der fragmentierten Speicherung von Informationen in isolierten Silos gleichen die Geschäftsprozesse moderner Unternehmen dem Versuch von Baumeistern, einen Wolkenkratzer aus Materialien zu errichten, die in Tausenden von verschiedenen Lagern gelagert sind. Das Übermaß an Informationen erschwert nicht nur den Zugriff auf rechtlich relevante Informationen, sondern verlangsamt auch die Entscheidungsfindung: Jeder Schritt muss wiederholt überprüft und bestätigt werden.

> Jede Aufgabe oder jeder Prozess ist fest mit einer separaten Tabelle oder Datenbank verdrahtet, und der Datenaustausch zwischen Systemen erfordert komplexe Integrationen. Fehler und Unstimmigkeiten in einem System können zu Kettenfehlern in anderen Systemen führen. Falsche Werte, verspätete Aktualisierungen und doppelte Informationen zwingen die Mitarbeiter dazu, viel Zeit mit dem manuellen Abgleich und der Abstimmung von Daten zu verbringen. Infolgedessen verbringt das Unternehmen mehr Zeit damit, sich mit den Folgen der Fragmentierung zu befassen, als Prozesse zu entwickeln und zu optimieren

Dieses Problem ist universell: Einige Unternehmen kämpfen weiterhin mit dem Chaos, während andere die Lösung in der Integration finden, d. h. in der Verlagerung der Informationsflüsse in ein zentralisiertes Speichersystem. Stellen Sie sich das als eine große Tabelle vor, in der Sie alle Entitäten im Zusammenhang mit Aufgaben, Projekten und Objekten speichern können. Anstelle von Dutzenden unterschiedlicher Tabellen und Formate entsteht ein einziges zusammenhängendes Repository (Abb. 2.2-2), das dies ermöglicht:

- den Datenverlust zu minimieren;
- die Notwendigkeit einer ständigen Harmonisierung der Informationen entfällt;
- die Verfügbarkeit und Qualität der Daten zu verbessern;
- die analytische Verarbeitung und das maschinelle Lernen zu vereinfachen

Daten auf einen gemeinsamen Standard zu bringen bedeutet, dass die Informationen unabhängig von ihrer Quelle in ein einheitliches und maschinenlesbares Format umgewandelt werden. Eine solche Organisation der Daten ermöglicht es, ihre Integrität zu überprüfen, sie in Echtzeit zu analysieren und sie umgehend für Managemententscheidungen zu nutzen.

Das Konzept der integrierten Speichersysteme und ihre Anwendung in der Analytik und im maschinellen

Lernen werden im Kapitel "Big Data Storage und maschinelles Lernen" näher erläutert. Die Themen Datenmodellierung und -strukturierung werden in den Kapiteln "Umwandlung von Daten in eine strukturierte Form" und "Wie Standards das Spiel verändern: von zufälligen Dateien zu einem durchdachten Datenmodell" ausführlich behandelt.

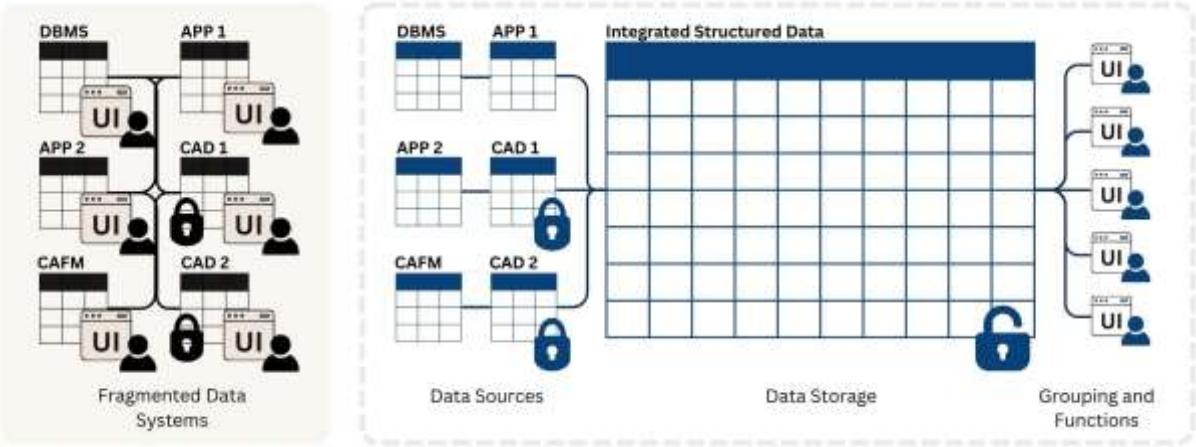


Abb. 2.2-2 Datenintegration beseitigt Silos, verbessert die Verfügbarkeit von Informationen und optimiert Geschäftsprozesse.

Sobald die Daten strukturiert und zusammengeführt sind, besteht der nächste logische Schritt darin, sie zu validieren. Mit einem einzigen integrierten Repository wird dieser Prozess erheblich vereinfacht: Es gibt nicht mehr mehrere inkonsistente Schemata, doppelte Strukturen und komplexe Beziehungen zwischen Tabellen. Alle Informationen werden an einem einzigen Datenmodell ausgerichtet, wodurch interne Inkonsistenzen beseitigt werden und der Validierungsprozess beschleunigt wird. Die Validierung von und die Sicherstellung der Datenqualität sind Eckpfeiler aller Geschäftsprozesse, die wir in den entsprechenden Kapiteln des Buches näher betrachten werden.

In der letzten Phase werden die Daten gruppiert, gefiltert und analysiert. Verschiedene Funktionen werden auf sie angewendet: Aggregation (Addition, Multiplikation), Berechnungen zwischen Tabellen, Spalten oder Zeilen (Abb. 2.2-4). Die Arbeit mit Daten wird zu einer Abfolge von Schritten: Sammlung, Strukturierung, Validierung, Umwandlung, analytische Verarbeitung und Weiterleitung an die endgültigen Anwendungen, in denen die Informationen zur Lösung praktischer Probleme verwendet werden. In den Kapiteln über ETL - Prozesse und den Datenpipeline-Ansatz werden wir mehr über den Aufbau solcher Szenarien, die Automatisierung von Schritten und den Aufbau von Verarbeitungsabläufen diskutieren.

Bei der digitalen Transformation geht es also nicht nur um die Vereinfachung des Umgangs mit Informationen. Es geht um die Beseitigung übermäßiger Komplexität bei der Datenverwaltung, um den Übergang vom Chaos zur Vorhersehbarkeit, von mehreren Systemen zu einem überschaubaren Prozess. Je geringer die Komplexität der Architektur ist, desto weniger Code ist zu ihrer Unterstützung erforderlich. Und in Zukunft könnte der Code als solcher ganz verschwinden und intelligenten Agenten Platz machen, die selbstständig Daten analysieren, systematisieren und umwandeln.

Integrierte Speichersysteme ermöglichen den Übergang zu KI-Agenten

Je weniger komplex die Daten und Systeme sind, desto weniger Code müssen Sie schreiben und pflegen. Und der einfachste Weg, Entwicklung einzusparen, ist, den Code ganz abzuschaffen und durch Daten zu ersetzen. Wenn die Entwicklung von Anwendungen von Code auf Datenmodelle umgestellt wird, kommt es zwangsläufig zu einer Verschiebung hin zu einem datenzentrierten (datengesteuerten) Ansatz, da hinter diesen Konzepten eine völlig andere Denkweise steht.

> Wenn man sich dafür entscheidet, die Daten in den Mittelpunkt zu stellen, beginnt man, ihre Rolle anders zu sehen. Daten sind nicht mehr nur "Rohmaterial" für Anwendungen - sie sind jetzt die Grundlage, um die herum Architektur, Logik und Interaktion aufgebaut werden.

Der traditionelle Ansatz für die Datenverwaltung beginnt in der Regel auf der Anwendungsebene und ähnelt in der Konstruktion einem schwerfälligen bürokratischen System: mehrstufige Genehmigungen, manuelle Prüfungen, endlose Versionen von Dokumenten durch entsprechende Softwareprodukte. Mit der Entwicklung digitaler Technologien werden immer mehr Unternehmen gezwungen sein, auf das Prinzip des Minimalismus umzusteigen - nur das zu speichern und zu nutzen, was wirklich notwendig ist und genutzt werden wird.

Die Logik der Minimierung ist von den Anbietern aufgegriffen worden. Um die Datenspeicherung und -verarbeitung zu vereinfachen, wird die Arbeit der Nutzer von Offline-Anwendungen und -Tools auf Cloud-Dienste und sogenannte SaaS-Lösungen verlagert.

> Das SaaS-Konzept (Software as a Service oder "Software als Dienstleistung") ist einer der wichtigsten Trends in der modernen IT-Infrastruktur, der es den Nutzern ermöglicht, über das Internet auf Anwendungen zuzugreifen, ohne dass sie Software auf ihren eigenen Computern installieren und warten müssen.

Einerseits hat SaaS die Skalierung, die Versionskontrolle und die Verringerung der Support- und Wartungskosten erleichtert, andererseits hat es den Nutzer neben der Abhängigkeit von der Logik einer bestimmten Anwendung auch vollständig von der Cloud-Infrastruktur des Anbieters abhängig gemacht. Fällt ein Dienst aus, kann der Zugang zu Daten und Geschäftsprozessen vorübergehend oder sogar dauerhaft blockiert sein. Darüber hinaus werden bei der Arbeit mit SaaS-Anwendungen alle Nutzerdaten auf den Servern des Anbieters gespeichert, was Sicherheits- und Compliance-Risiken birgt. Änderungen der Tarife oder Nutzungsbedingungen können ebenfalls zu erhöhten Kosten oder zur Notwendigkeit einer dringenden Migration führen.

Die Entwicklung von KI-Agenten (), LLM-Agenten () und datenzentrierten Ansätzen hat die Zukunft von Anwendungen in ihrer herkömmlichen Form und der SaaS-Ausführung () in Frage gestellt. Während Anwendungen und Dienste bisher für die Verwaltung von Geschäftslogik und die Verarbeitung von Daten erforderlich waren, können sich diese Funktionen mit dem Aufkommen von KI-Agenten auf intelligente Systeme verlagern, die direkt mit Daten arbeiten.

Aus diesem Grund werden in den IT-Abteilungen und auf der Managementebene zunehmend hybride Architekturen diskutiert, bei denen KI -Agenten und Lösungen vor Ort die Cloud-Dienste ergänzen und die Abhängigkeit von SaaS-Plattformen verringern.

> *Der von uns verfolgte Ansatz trägt der Tatsache Rechnung, dass sich traditionelle Geschäftsanwendungen oder SaaS -Anwendungen im Zeitalter der Agenten drastisch verändern können. Diese Anwendungen sind im Wesentlichen CRUD [create, read, update and delete] Datenbanken mit Geschäftslogik. In Zukunft wird diese Logik jedoch von KI-Agenten [46] übernommen werden.*
>
> - Satya Nadella, CEO von Microsoft, 2024.

Ein datenzentrierter Ansatz und der Einsatz von KI/LLM-Agenten können überflüssige Prozesse reduzieren und damit die Arbeitsbelastung der Mitarbeiter verringern. Wenn Daten richtig organisiert sind, lassen sie sich leichter analysieren, visualisieren und für die Entscheidungsfindung nutzen. Anstelle von endlosen Berichten und Überprüfungen erhalten Fachleute mit wenigen Klicks oder mit Hilfe von LLM-Agenten automatisch Zugang zu aktuellen Informationen in Form von fertigen Dokumenten und Dashboards.

Wir werden bei der Datenmanipulation durch Werkzeuge der künstlichen Intelligenz (AI) und LLM chats unterstützt. In den letzten Jahren gab es einen Trend weg von den traditionellen CRUD-Operationen (Erstellen, Lesen, Aktualisieren, Löschen) hin zur Verwendung von großen Sprachmodellen (LLMs) für die Datenverwaltung. LLMs sind in der Lage, natürliche Sprache zu interpretieren und automatisch entsprechende Datenbankabfragen zu generieren, was die Interaktion mit Datenmanagementsystemen vereinfacht (Abb. 2.2-3).

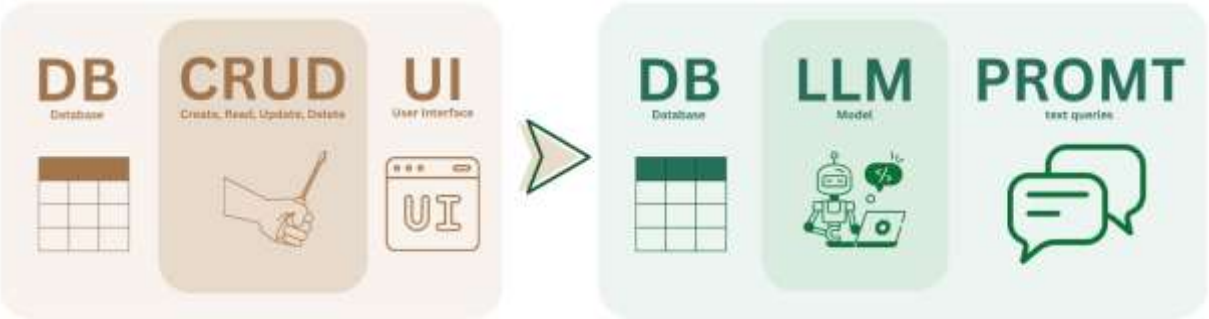


Abb. 2.2-3 KI wird Speicher- und Datenbanklösungen ersetzen und integrieren, wodurch traditionelle Anwendungen und CRUD -Operationen allmählich verdrängt werden.

> *In den nächsten 3-6 Monaten wird die KI 90% des Codes schreiben, und in 12 Monaten könnte fast der gesamte Code von der KI generiert werden [47].*
>
> - Dario Amodei, CEO von LLM Anthropic, März 2025.

Trotz der rasanten Entwicklung von KI-Entwicklungstools (z. B. GitHub Copilot) spielen auch im Jahr 2025 die Entwickler eine Schlüsselrolle in diesem Prozess. KI-Agenten werden immer mehr zu nützlichen Assistenten: Sie interpretieren automatisch Benutzeranfragen, generieren SQL- und Pandas-Abfragen (mehr dazu

in den folgenden Kapiteln) oder schreiben Code zur Analyse von Daten. Auf diese Weise ersetzt die künstliche Intelligenz allmählich die traditionellen Benutzeroberflächen von Anwendungen.

Die Verbreitung von Modellen künstlicher Intelligenz, wie z. B. Sprachmodelle, wird die Entwicklung hybrider Architekturen vorantreiben. Anstatt Cloud-Lösungen und SaaS -Produkte vollständig aufzugeben, werden wir möglicherweise die Integration von Cloud-Diensten mit lokalen Datenverwaltungssystemen erleben. So ermöglicht beispielsweise föderiertes Lernen leistungsstarke KI-Modelle, ohne dass sensible Daten in die Cloud verschoben werden müssen. Auf diese Weise können Unternehmen die Kontrolle über ihre Daten behalten und erhalten gleichzeitig Zugang zu fortschrittlichen Technologien.

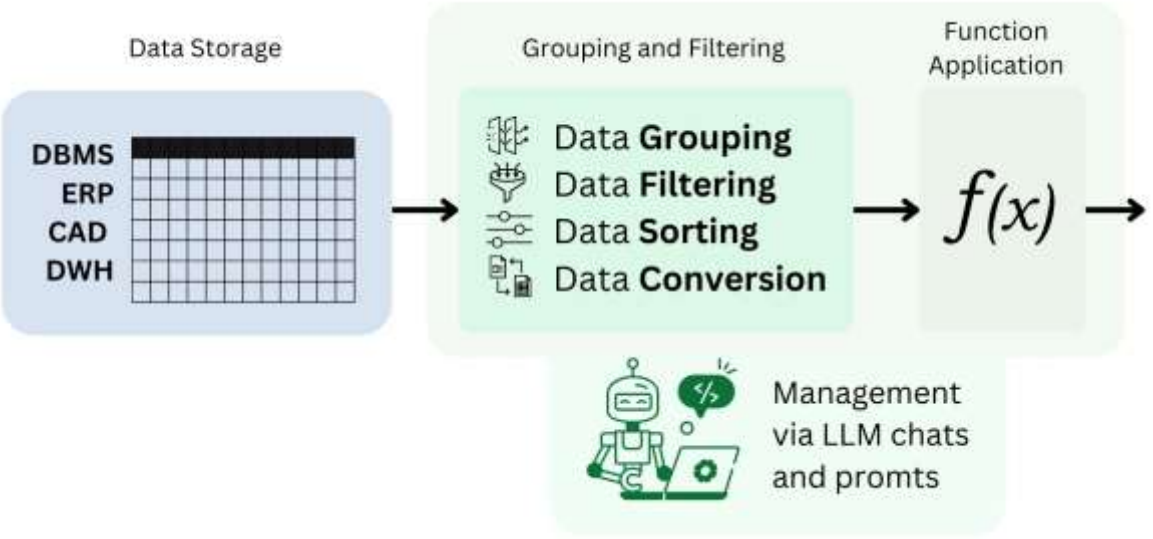


Abbildung 2.2-4 Die grundlegenden Operationen des Gruppierens, Filterns und Sortierens mit anschließender Funktionsanwendung werden von LLM chats übernommen.

Die Zukunft des Baugewerbes wird auf einer Kombination aus Vor-Ort-Lösungen, Cloud-Leistung und intelligenten Modellen beruhen, die zusammenarbeiten, um effiziente und sichere Datenverwaltungssysteme zu schaffen. LLM wird es Nutzern ohne tiefes technisches Wissen ermöglichen, mit Datenbanken und Data Warehouses zu interagieren, indem sie ihre Anfragen in natürlicher Sprache formulieren. Wir werden im Kapitel "LLM-Agenten und strukturierte Datenformate" mehr über LLM- und KI-Agenten und deren Funktionsweise sprechen.

Ordnungsgemäß organisierte Daten und einfache, benutzerfreundliche LLM-gestützte Analysetools erleichtern nicht nur die Arbeit mit Informationen, sondern tragen auch dazu bei, Fehler zu minimieren, die Effizienz zu steigern und Prozesse zu automatisieren.

Von der Datenerfassung zur Entscheidungsfindung: der Weg zur Automatisierung

In späteren Teilen des Buches werden wir uns ausführlich damit befassen, wie Spezialisten miteinander interagieren und wie Daten die Grundlage für Entscheidungsfindung, Automatisierung und betriebliche Effizienz bilden. Abbildung 2.2-5 enthält ein Beispieldiagramm, das die Abfolge der Datenverarbeitungsschritte

in einem datenzentrierten Ansatz zeigt. Dieses Diagramm veranschaulicht die "Continuous Improvement Pipeline" (), von der Teile später in diesem Buch ausführlich behandelt werden.

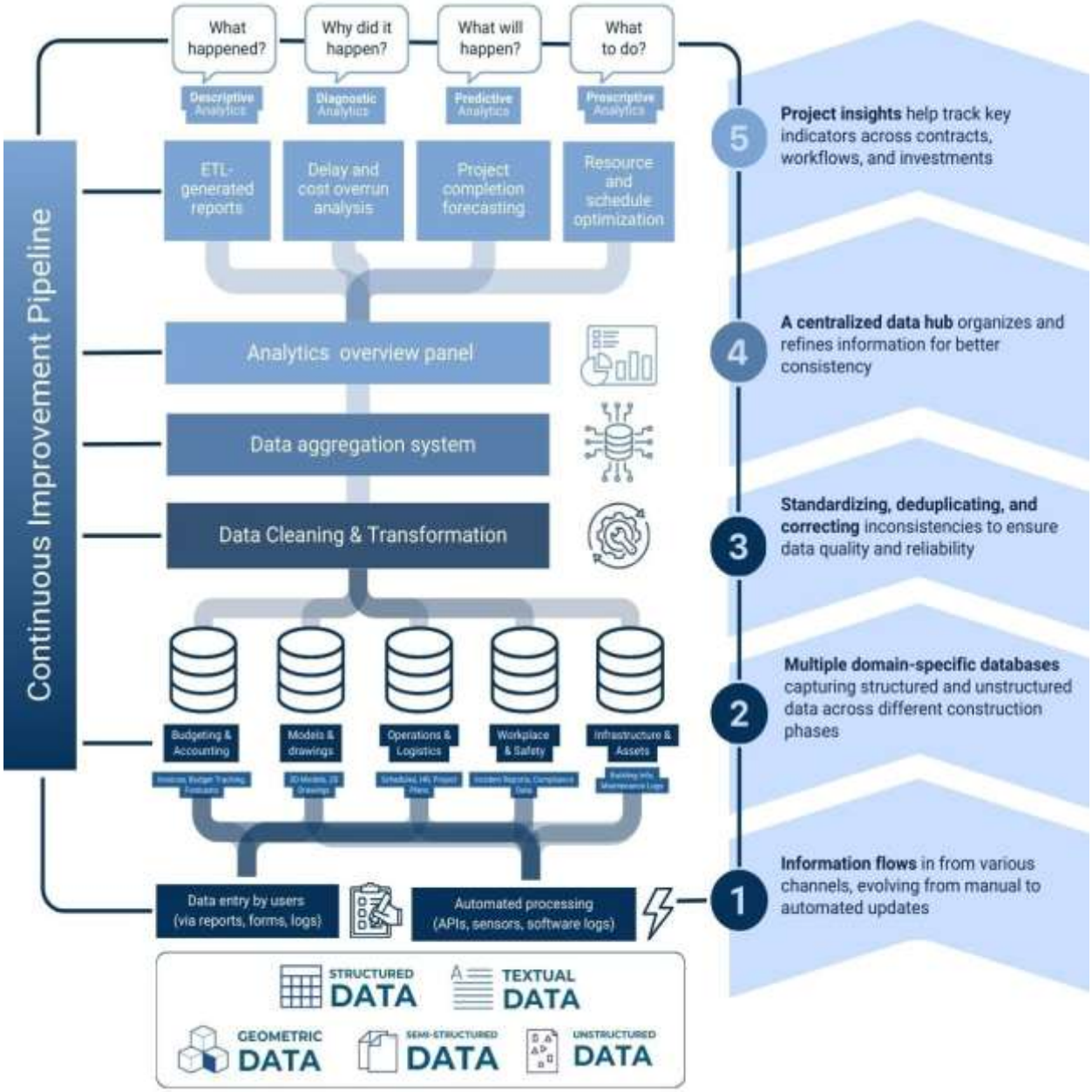


Abb. 2.2-5 Ein Beispiel für eine kontinuierliche Datenverbesserungspipeline: der Datenverarbeitungs- und -analysefluss in Bauprojekten.

Das System, das die Geschäftsprozesse eines mittelständischen Unternehmens beschreibt, ist nach einem mehrstufigen Prinzip aufgebaut. Es umfasst: Datenerfassung, Bereinigung, Aggregation, analytische Verarbeitung und Entscheidungsfindung auf der Grundlage der Ergebnisse. Wir werden alle diese Stufen im weiteren Verlauf des Buches untersuchen - sowohl in einem theoretischen Kontext als auch anhand praktischer

Beispiele:

- Auf der ersten Ebene findet die **Dateneingabe statt (**Abb. 3.1-1**)**. Die Informationen werden sowohl manuell (über Berichte, Formulare, Protokolle) als auch in automatisierter Form (von API, Sensoren, Softwaresystemen) empfangen. Die Daten können unterschiedlich strukturiert sein: geometrisch, textlich, unstrukturiert. In diesem Stadium besteht ein Bedarf an Standardisierung, Strukturierung und Vereinheitlichung der Informationsflüsse.
- Die nächste Ebene ist die **Datenverarbeitung und -umwandlung**. Sie umfasst die Prozesse der Bereinigung, des Entfernens von Duplikaten, der Korrektur von Fehlern und der Vorbereitung von Informationen für weitere Analysen (Abb. 4.2-5). Diese Stufe ist entscheidend, da die Qualität der Analysen direkt von der Sauberkeit und Genauigkeit der Daten abhängt.
- **Die Daten fließen** dann **in spezialisierte Tabellen, Datenrahmen oder Datenbanken,** die nach Funktionsbereichen unterteilt sind: Budgetierung und Buchhaltung, Modelle und Zeichnungen, Logistik, Sicherheit und Infrastruktur. Diese Unterteilung ermöglicht einen einfachen Zugang und eine Queranalyse der Informationen.
- Die Daten werden dann **aggregiert und in einem analytischen Dashboard** (Showcase) **angezeigt**. Hier werden deskriptive, diagnostische, prädiktive und präskriptive Analysemethoden angewandt. Dies hilft bei der Beantwortung von Schlüsselfragen (Abb. 1.1-4): was ist passiert, warum ist es passiert, was wird in Zukunft passieren und welche Maßnahmen müssen ergriffen werden. So kann das System beispielsweise Verzögerungen erkennen, den Abschluss von Projekten vorhersagen oder Ressourcen optimieren.
- Die letzte Ebene schließlich liefert **analytische Schlussfolgerungen und Schlüsselindikatoren**, die helfen, die Vertragserfüllung zu überwachen, Investitionen zu verwalten und Geschäftsprozesse zu verbessern (Abb. 7.4-2). Diese Informationen bilden die Grundlage für die Entscheidungsfindung und die Entwicklungsstrategie des Unternehmens.

Ähnlich verhält es sich mit dem Übergang von der Datenerhebung zur Nutzung der Daten im strategischen Management. In den folgenden Teilen des Buches werden wir jede Phase im Detail betrachten, wobei wir uns auf Datentypen, Datenverarbeitungstechniken, Analysewerkzeuge und reale Fälle, wie diese Ansätze in der Bauindustrie verwendet werden, konzentrieren.

Nächste Schritte: Das Chaos in ein überschaubares System verwandeln

In diesem Teil haben wir uns mit den Herausforderungen von Informationssilos befasst und die Auswirkungen übermäßiger Systemkomplexität auf die Unternehmensleistung untersucht, wobei wir den Übergang von der vierten zur fünften industriellen Revolution analysiert haben, bei der Daten und nicht Anwendungen im Mittelpunkt stehen. Wir haben gesehen, wie siloartige Informationssysteme den Wissensaustausch behindern und wie die anhaltende Komplexität der IT-Landschaft die Produktivität verringert und Innovationen in der Bauindustrie hemmt.

Um diesen Teil zusammenzufassen, lohnt es sich, die wichtigsten praktischen Schritte hervorzuheben, die Ihnen helfen werden, die besprochenen Ansätze bei Ihren täglichen Aufgaben anzuwenden:

- Visualisieren Sie Ihre Informationslandschaft
    - ☐ Erstellen Sie eine visuelle Karte der Datenquellen (Miro, Figma, Canva), mit denen Sie regelmäßig arbeiten
    - ☐ Fügen Sie die Systeme und Anwendungen, die Sie bei Ihrer Arbeit verwenden, zu dieser

Karte hinzu

- Identifizierung potenziell doppelter Funktionen und redundanter Lösungen
- Identifizierung kritischer Punkte, an denen es bei der Übertragung zwischen Systemen zu Datenverlusten oder -beschädigungen kommen kann

- Einführung personalisierter Datenverwaltungspraktiken
  - Verlagerung des Schwerpunkts von Anwendungen auf Daten als Schlüsselressource in Prozessen
  - Dokumentation von Datenquellen und Verarbeitungsmethoden zur Gewährleistung der Transparenz
  - Entwicklung von Mechanismen zur Bewertung und Verbesserung der Datenqualität
  - Streben Sie danach, dass Daten einmal eingegeben und immer wieder verwendet werden - dies ist die Grundlage für eine effiziente Prozessorganisation
- Fördern Sie einen datenzentrierten (datengesteuerten) Ansatz in Ihrem Team
  - die Verwendung von standardisierten und einheitlichen Formaten für den Peer-to-Peer-Datenaustausch vorschlagen
  - Regelmäßiges Ansprechen von Problemen im Zusammenhang mit der Datenqualität und -verfügbarkeit in Teamsitzungen
  - Lernen Sie Open Source Alternativen zu den Tools kennen, die Sie zur Lösung Ihrer Probleme verwenden

Fangen Sie klein an - wählen Sie einen bestimmten Prozess oder Datensatz, der für Ihre Arbeit entscheidend ist, und wenden Sie einen datenzentrierten Ansatz darauf an, indem Sie den Schwerpunkt von den Tools auf die Daten verlagern. Wenn Sie mit einem einzigen Pilotprojekt Erfolg haben, können Sie nicht nur praktische Erfahrungen sammeln, sondern Ihrem Team auch die Vorteile der neuen Methodik deutlich vor Augen führen. Wenn Sie bei der Durchführung der meisten dieser Schritte Fragen haben, können Sie sich an jeden aktuellen LLM wenden, um Klarheit und Unterstützung zu erhalten.

In den folgenden Teilen des Buches werden wir uns eingehender mit Techniken zur Datenstrukturierung und -harmonisierung befassen und praktische Ansätze zur Integration heterogener Informationen untersuchen. Besonderes Augenmerk wird auf den Übergang von disparaten Silos zu einheitlichen Datenökosystemen gelegt, die eine Schlüsselrolle bei der digitalen Transformation der Baubranche spielen.

III TEIL

DATENRAHMEN IN DEN GESCHÄFTSPROZESSEN DES BAUWESENS

Im dritten Teil wird ein umfassendes Verständnis der Typologie von Daten im Bauwesen und der Methoden ihrer effektiven Organisation entwickelt. Die Merkmale und Besonderheiten der Arbeit mit strukturierten, unstrukturierten, halbstrukturierten, textuellen und geometrischen Daten im Rahmen von Bauprojekten werden analysiert. Moderne Speicherformate und Protokolle für den Austausch von Informationen zwischen verschiedenen in der Industrie verwendeten Systemen werden besprochen. Es werden praktische Werkzeuge und Techniken zur Konvertierung von Daten in verschiedenen Formaten in eine einzige strukturierte Umgebung beschrieben, einschließlich der Integration von CAD (BIM) Daten. Es werden Ansätze zur Sicherstellung der Qualität von Daten durch Standardisierung und Validierung vorgeschlagen, die für die Genauigkeit von Bauberechnungen entscheidend sind. Praktische Aspekte des Einsatzes moderner Technologien (Python Pandas, LLM -Modelle) mit Code-Beispielen zur Lösung typischer Probleme in der Bauindustrie werden eingehend analysiert. Der Wert der Schaffung eines Kompetenzzentrums (CoE) als Organisationsstruktur für die Koordination und Standardisierung von Informationsmanagementansätzen wird begründet.

KAPITEL 3.1.
DATENTYPEN IN DER KONSTRUKTION

Die wichtigsten Datenarten im Baugewerbe

In der modernen Bauwirtschaft werden die Systeme, Anwendungen und Data Warehouses von aktiv mit Informationen und Daten unterschiedlicher Art und Formate gefüllt (Abb. 3.1-1). Schauen wir uns die wichtigsten Datentypen genauer an, die die Informationslandschaft eines modernen Unternehmens der Baubranche ausmachen:

- **Strukturierte** Daten: Diese Daten haben eine klare Organisationsstruktur, z. B. Excel-Tabellen und relationale Datenbanken.
- **Unstrukturierte** Daten: Das sind Informationen, die nicht nach strengen Regeln organisiert sind. Beispiele für solche Daten sind Texte, Videos, Fotos und Audioaufnahmen.
- **Lose strukturierte** Daten: Diese Daten nehmen eine Zwischenstellung zwischen strukturierten und unstrukturierten Daten ein. Sie enthalten Strukturelemente, aber diese Struktur ist nicht immer klar oder oft durch verschiedene Schemata beschrieben. Beispiele für halbstrukturierte Daten im Bauwesen sind: technische Spezifikationen, Projektdokumentation oder Fortschrittsberichte.
- **Textdaten**: Dazu gehört alles, was aus der mündlichen und schriftlichen Kommunikation stammt, z. B. E-Mails, Sitzungsprotokolle und Termine.
- **Geometrische** Daten: Diese Daten stammen aus CAD-Programmen, in denen Spezialisten geometrische Daten von Projektelementen zur Visualisierung, Bestätigung von Volumenwerten oder Kollisionsprüfung erstellen.

Es ist wichtig, darauf hinzuweisen, dass geometrische und textuelle (alphanumerische) Daten keine separate Kategorie darstellen, sondern in allen drei Datentypen enthalten sein können. Geometrische Daten können beispielsweise sowohl Teil strukturierter Daten (parametrische CAD-Formate) als auch unstrukturierter Daten (gescannte Zeichnungen) sein. Textdaten können in ähnlicher Weise sowohl in Datenbanken (strukturierte Daten) organisiert sein als auch als Dokumente ohne klare Struktur vorliegen.

Jede Art von Daten in einem Bauunternehmen ist ein einzigartiges Element im Mosaik der Informationsbestände des Unternehmens. Von unstrukturierten Daten wie Bildern von Baustellen und Audioaufzeichnungen von Besprechungen bis hin zu strukturierten Aufzeichnungen, einschließlich Tabellen und Datenbanken, spielt jedes Element eine wichtige Rolle bei der Gestaltung der Informationslandschaft des Unternehmens.

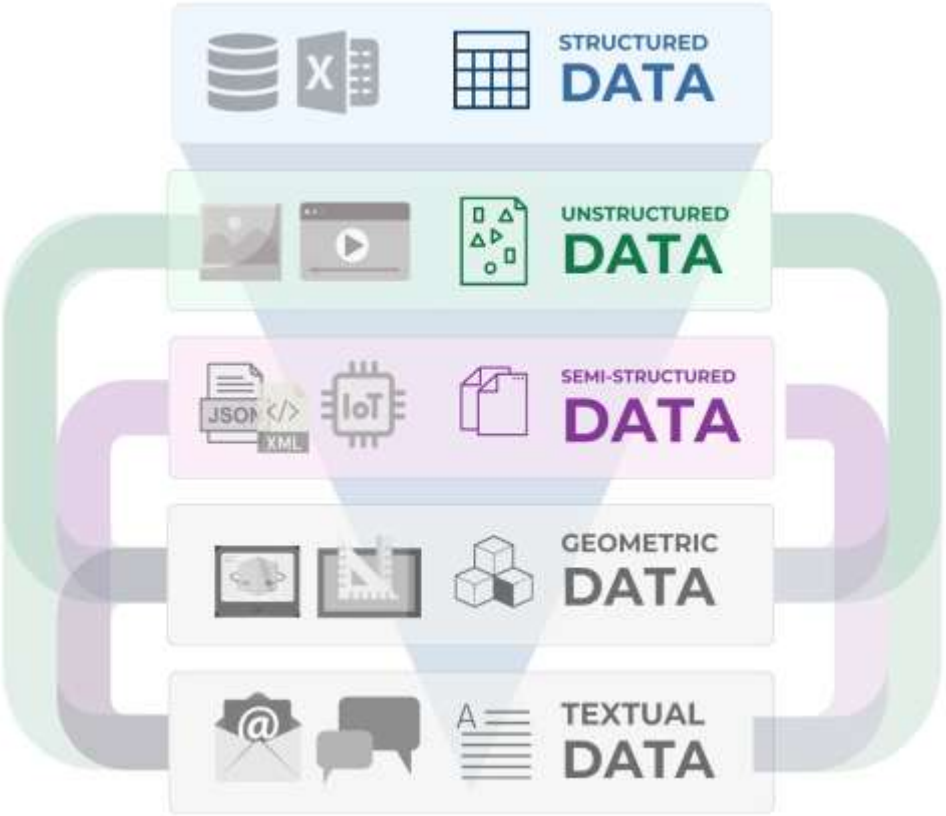


Abb. 3.1-1 Ingenieure und Datenmanager müssen lernen, mit allen Arten von Daten zu arbeiten, die in der Bauindustrie verwendet werden.

Nachfolgend finden Sie eine beispielhafte Auflistung einiger der im Bauwesen verwendeten Systeme und zugehörigen Datentypen (Abb. 3.1-2):

- **ERP** (Enterprise Resource Planning) - verarbeitet allgemein strukturierte Daten, um die Verwaltung von Unternehmensressourcen und die Integration verschiedener Geschäftsprozesse zu unterstützen.
- **CAD** (Computer-Aided Design) in Kombination mit **BIM** (Building Information Modeling) - nutzt geometrische und halbstrukturierte Daten zur Planung und Modellierung von Bauprojekten und gewährleistet so die Genauigkeit und Konsistenz der Informationen während der Planungsphase.
- **GIS** (Geographische Informationssysteme) - arbeitet mit geometrischen und strukturierten Daten, um kartographische Daten und räumliche Beziehungen zu erstellen und zu analysieren.
- **RFID** (Radio-Frequency Identification) - nutzt halbstrukturierte Daten zur effizienten Verfolgung von Materialien und Geräten auf einer Baustelle mittels Radiofrequenz-Identifikation.
- **ECM** (Engineering Content Management) ist ein System zur Verwaltung technischer Daten und Dokumentationen, einschließlich halbstrukturierter und unstrukturierter Daten wie technischer Zeichnungen und Konstruktionsunterlagen.

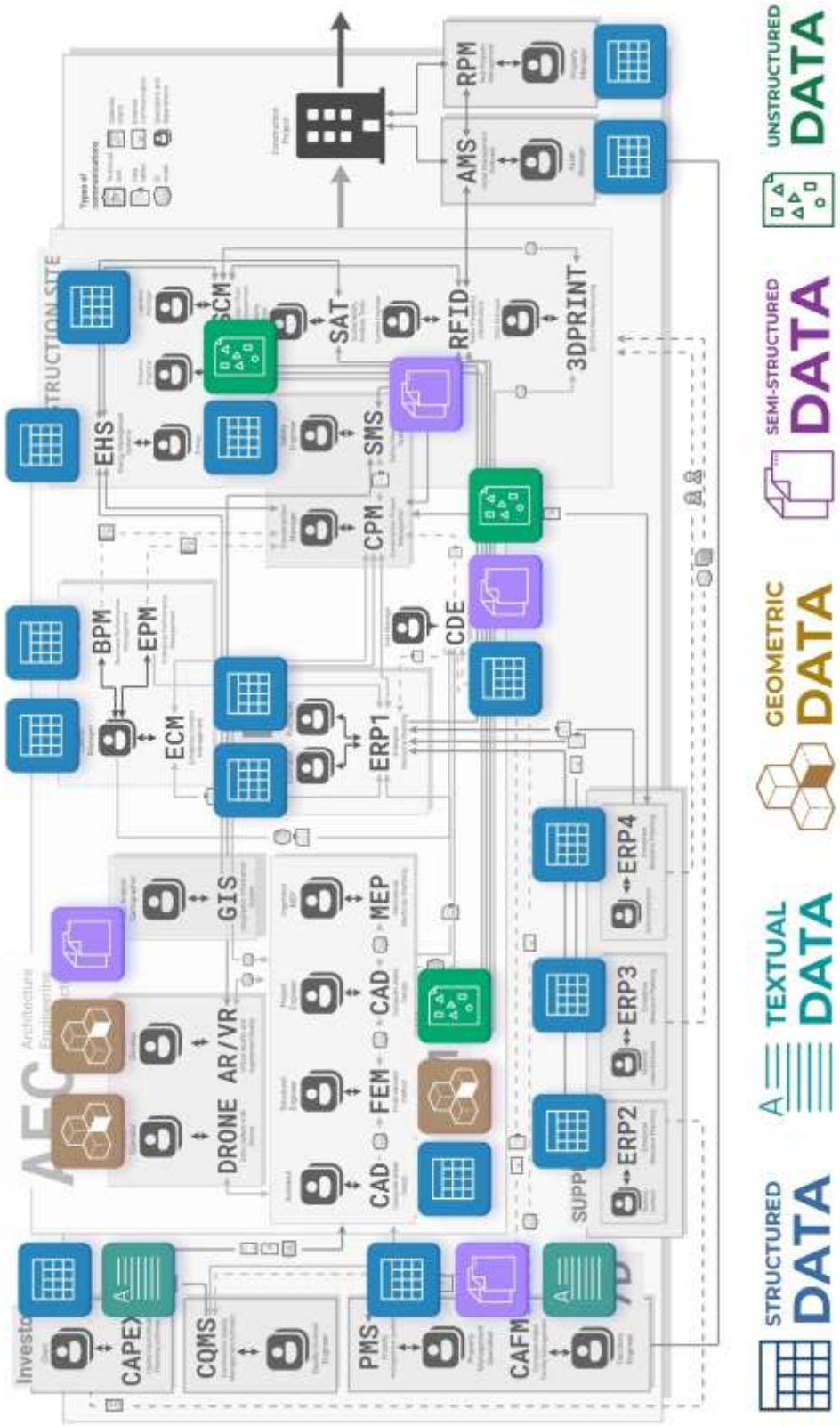


Abb. 3.1-2 Unterschiedliche Formate und Daten bevölkern verschiedene Systeme und müssen in eine für die komplexe Integration geeignete Form übersetzt werden.

Diese und viele andere Systeme des Unternehmens verwalten ein breites Spektrum an Daten, von strukturierten Tabellendaten bis hin zu komplexen geometrischen Modellen, und ermöglichen so eine integrierte Zusammenarbeit bei Entwurfs-, Planungs- und Baumanagementprozessen.

Im Beispiel eines vereinfachten Dialogs (Abb. 3.1-3) werden verschiedene Arten von Daten zwischen den Spezialisten des Bauprojekts ausgetauscht:

- **Architekt**: "*Unter Berücksichtigung der Wünsche des Kunden habe ich einen Sitzbereich auf dem Dach hinzugefügt. Bitte sehen Sie sich den neuen Entwurf an*" (Geometrische Daten - Modell).
- **Tragwerksplaner**: "*Das Projekt ist eingegangen. Ich berechne die Tragfähigkeit des Daches für den neuen Erholungsraum*" (strukturierte und halbstrukturierte Daten - Berechnungstabellen).
- **Einkaufsleiter**: "*Benötige Spezifikationen und Mengen von Materialien für den Erholungsraum, um den Einkauf zu organisieren*" (textuelle und halbstrukturierte Daten - Listen und Spezifikationen).
- **Ingenieur für Gesundheit und Sicherheit**: "*Ich habe Daten über den neuen Bereich erhalten. Ich bewerte die Risiken und aktualisiere den Sicherheitsplan*" (halbstrukturierte Daten - Dokumente und Pläne).
- **Spezialist für BIM** -Modellierung: "*Änderungen am Gesamtprojektmodell zur Anpassung der Arbeitsunterlagen*" (geometrische Daten und halbstrukturierte Daten).
- **Projektleiter**: "*Ich integriere den neuen Rastplatz in den Arbeitsplan. Ich aktualisiere die Zeitpläne und Ressourcen im Projektmanagementsystem*" (strukturierte und halbstrukturierte Daten - Zeitpläne und Pläne).
- **Spezialist für die Instandhaltung von Anlagen (FM)**: "*Ich bereite Daten für die künftige Instandhaltung des Erholungsgebiets vor und gebe sie in das Immobilienverwaltungssystem ein*" (strukturierte und halbstrukturierte Daten - Anweisungen und Instandhaltungspläne).

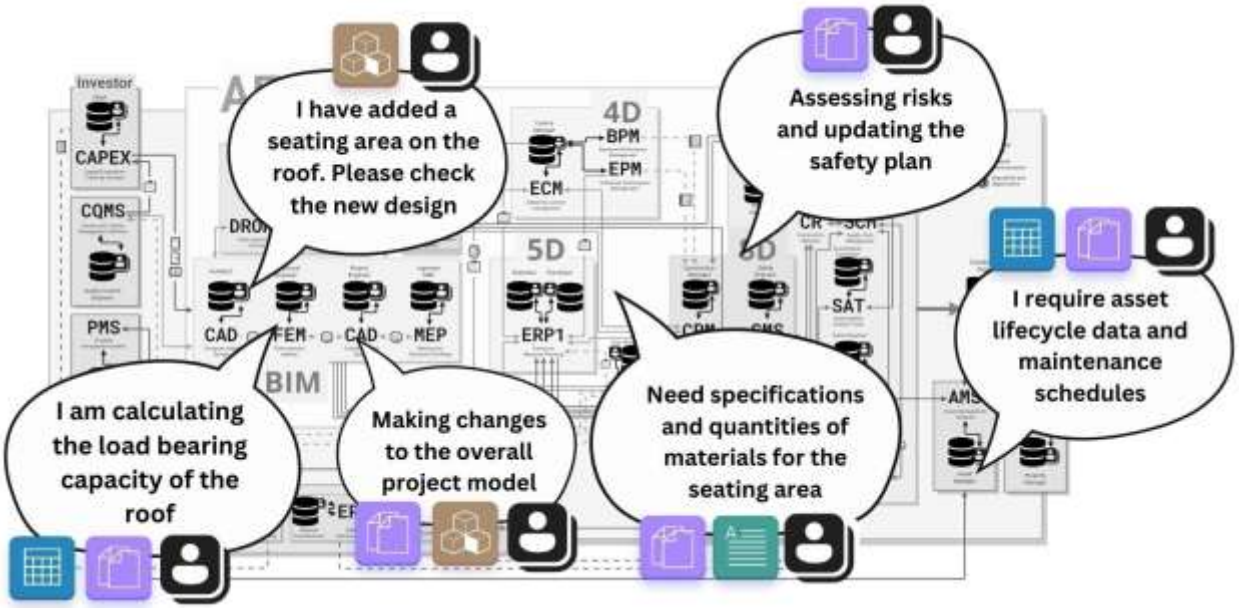


Abb. 3.1-3 Die Kommunikation zwischen Fachleuten findet sowohl auf Text- als auch auf Datenebene statt.

Jeder Fachmann arbeitet mit unterschiedlichen Datentypen, um eine effektive Zusammenarbeit im Team und den Projekterfolg sicherzustellen. Wenn Sie die Unterschiede zwischen strukturierten, halbstrukturierten und unstrukturierten Daten verstehen, können Sie die einzigartige Rolle erkennen, die jeder Datentyp in

digitalen Geschäftsprozessen spielt. Es ist nicht nur wichtig zu wissen, dass es verschiedene Formen von Daten gibt, sondern auch zu verstehen, wie, wo und warum sie verwendet werden.

> Vor nicht allzu langer Zeit schien die Idee, so unterschiedliche Daten zu kombinieren, zwar ehrgeizig, aber schwer zu realisieren. Heute ist sie bereits Teil der täglichen Praxis. Die Integration von Daten unterschiedlicher Schemata und Strukturen ist zu einem festen Bestandteil der Architektur moderner Informationssysteme geworden.

In den folgenden Kapiteln werden wir uns ausführlich mit den wichtigsten Standards und Ansätzen befassen, die es ermöglichen, strukturierte, halbstrukturierte und unstrukturierte Daten in einer einzigen kohärenten Ansicht zu kombinieren. Besonderes Augenmerk wird dabei auf strukturierte Daten und relationale Datenbanken als Hauptmechanismen für die Speicherung, Verarbeitung und Analyse von Informationen in der Bauindustrie gelegt.

Strukturierte Daten

In der Baubranche stammen Informationen aus vielen Quellen - Zeichnungen, Spezifikationen, Zeitpläne und Berichte. Um diesen Informationsfluss effektiv zu verwalten, müssen die Informationen strukturiert werden. Strukturierte Daten ermöglichen es Ihnen, die Informationen in einer praktischen, lesbaren und zugänglichen Form zu organisieren.

> Laut dem 5th Annual Construction Technology Report von JB Knowledge [17] verfolgen und bewerten 67% der Fachleute für Bauprojektmanagement die Arbeitsleistung manuell oder mithilfe von Tabellenkalkulationen.

Einige der gängigsten strukturierten Datenformate sind XLSX und CSV. Sie werden häufig für die Speicherung, Verarbeitung und Analyse von Informationen in Tabellenkalkulationen verwendet. In solchen Tabellenkalkulationen werden die Daten in Form von Zeilen und Spalten dargestellt, so dass sie leicht zu lesen, zu bearbeiten und zu analysieren sind

XLSX, ein von Microsoft entwickeltes **Format**, basiert auf der Verwendung von XML -Strukturen und wird mit Hilfe des ZIP-Algorithmus archiviert. Die wichtigsten Merkmale des Formats sind:

- Unterstützung für komplexe Formeln, Diagramme und Makros.
- Möglichkeit der Speicherung von Daten in verschiedenen Blättern sowie der Formatierung von Informationen.
- Optimiert für Microsoft Excel, aber kompatibel mit anderen Office-Suiten.

CSV-Format ist eine reine Textdatei, in der die Werte durch Kommas, Semikolons oder andere Begrenzungszeichen getrennt sind. Wichtigste Vorteile:

- Universelle Kompatibilität mit verschiedenen Programmen und Betriebssystemen.
- Einfacher Import/Export in Datenbanken und Analysesysteme.
- Einfache Bearbeitung auch in Texteditoren.

CSV unterstützt jedoch keine Formeln und Formatierungen, so dass seine Hauptanwendung der Datenaus-

tausch zwischen Systemen und die Massenaktualisierung von Informationen ist. Aufgrund seiner Vielseitigkeit und Plattformunabhängigkeit ist CSV zu einem beliebten Instrument für den Datentransfer in heterogenen IT-Umgebungen geworden.

Die beiden Formate XLSX und CSV dienen als Bindeglied zwischen verschiedenen Systemen, die mit strukturierten Daten arbeiten (Abb. 3.1-4). Sie sind besonders nützlich für Aufgaben, bei denen Lesbarkeit, manuelle Bearbeitung und grundlegende Kompatibilität wichtig sind

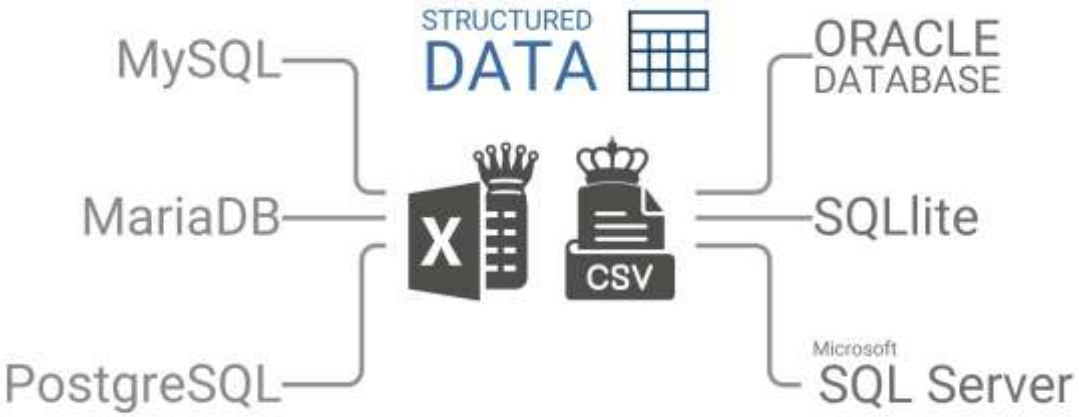


Abb. 3.1-4 Die Formate XLSX und CSV sind das Bindeglied zwischen verschiedenen Systemen, die mit strukturierten Daten arbeiten.

> Die Plattformunabhängigkeit macht CSV zum beliebtesten Format für die Datenübertragung in heterogenen IT-Umgebungen und Systemen.

XLSX und CSV sind jedoch nicht für Hochleistungsberechnungen oder die langfristige Speicherung großer Datenmengen geeignet. Modernere strukturierte Formate wie Apache Parquet, Apache ORC, Feather, HDF5 werden für solche Zwecke verwendet. Diese Formate werden im Kapitel "Speicherung von Big Data: Analyse gängiger Formate und ihrer Effektivität" in Teil 9 dieses Buches ausführlicher behandelt

In der Praxis wird Excel im XLSX-Format häufiger für kleinere Aufgaben und die Automatisierung von Routineprozessen eingesetzt. Komplexere Szenarien erfordern den Einsatz von Datenmanagementsystemen wie ERP, PMIS CAFM, CPM, SCM und anderen (Abb. 3.2-1). In diesen Systemen werden strukturierte Daten gespeichert, auf denen die Organisation und das Management der Informationsflüsse des Unternehmens beruhen.

Moderne Datenmanagement-Informationssysteme, die in der Bauindustrie eingesetzt werden, stützen sich auf strukturierte Daten, die in Form von Tabellen organisiert sind. Für die zuverlässige, skalierbare und ganzheitliche Verwaltung großer Informationsmengen setzen Anwendungs- und Systementwickler auf relationale Datenbankmanagementsysteme (RDBMS).

Relationale Datenbanken RDBMS und SQL-Abfragesprache

Um Daten effizient zu speichern, zu verarbeiten und zu analysieren, sind **relationale Datenbanken (RDBMS**) Datenspeichersysteme, die Informationen in Tabellen mit definierten Beziehungen zwischen ihnen organisieren.

> Die in Datenbanken (RDBMS) organisierten Daten sind nicht nur digitale Informationen, sondern bilden die Grundlage für Transaktionen und Interaktionen zwischen verschiedenen Systemen.

Im Folgenden werden einige der gängigsten relationalen Datenbankmanagementsysteme (RDBMS) vorgestellt (Abb. 3.1-5):

- **MySQL** (Open Source) ist eines der beliebtesten RDBMS, das Teil des LAMP-Stacks (Linux, Apache, MySQL, PHP /Perl/Python) ist. Es wird aufgrund seiner Einfachheit und hohen Leistung häufig in der Webentwicklung eingesetzt.
- **PostgreSQL** (Open Source) ist ein leistungsfähiges objektrelationales System, das für seine Zuverlässigkeit und seine erweiterten Funktionen bekannt ist. Es ist für komplexe Unternehmenslösungen geeignet.
- **Microsoft SQL Server** ist ein kommerzielles System von Microsoft, das aufgrund seiner Integration mit anderen Unternehmensprodukten und seines hohen Sicherheitsniveaus in Unternehmensumgebungen weit verbreitet ist.
- **Oracle Database** ist eines der leistungsstärksten und zuverlässigsten DBMS, das in großen Unternehmen und geschäftskritischen Anwendungen eingesetzt wird.
- **IBM DB2** - richtet sich an große Unternehmen und bietet hohe Leistung und Fehlertoleranz.
- **SQLite** (Open Source) ist eine leichtgewichtige eingebettete Datenbank, die sich ideal für mobile Anwendungen und eigenständige Systeme wie CAD-Designsoftware (BIM) eignet.

Die in der Baubranche gängigen Datenbankmanagementsysteme - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 und SQLite - arbeiten mit strukturierten Daten. Alle diese DBMS sind leistungsstarke und flexible Lösungen für die Verwaltung einer Vielzahl von Geschäftsprozessen und Anwendungen, von kleinen Websites bis hin zu großen Unternehmenssystemen (Abb. 3.2-1).

> Laut Statista [48] machen relationale Datenbankmanagementsysteme (RDBMS) im Jahr 2022 etwa 72% der insgesamt eingesetzten DBMS aus.

Rank			DBMS	Database Model	Open Source vs Commercial
Mar2025	Feb2025	Mar2024			
1.	1.	1.	**Oracle®**	Relational, Multi-model	Commercial
2.	2.	2.	**MySQL**	Relational, Multi-model	**Open Source**
3.	3.	3.	**Microsoft® SQL Server**	Relational, Multi-model	Commercial
4.	4.	4.	**PostgreSQL**	Relational, Multi-model	**Open Source**
5.	5.	5.	MongoDB	Document, Multi-model	**Open Source**
6.	7.	9.	**Snowflake®**	Relational	Commercial
7.	6.	6.	Redis®	Key-value, Multi-model	**Open Source**
8.	8.	7.	Elasticsearch®	Multi-model	**Open Source**
9.	9.	8.	**IBM Db2**	Relational, Multi-model	Commercial
10.	10.	10.	**SQLite**	Relational	**Open Source**
11.	11.	12.	Apache Cassandra®	Multi-model	**Open Source**
12.	12.	11.	**Microsoft Access®**	Relational	**Open Source**
13.	13.	17.	Databricks®	Multi-model	Commercial
14.	14.	13.	**MariaDB**	Relational, Multi-model	**Open Source**
15.	15.	14.	Splunk	Search engine	Commercial
16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
17.	17.	15.	**Microsoft Azure SQL**	Relational, Multi-model	Commercial

Abb. 3.1-5 Beliebtheit der Verwendung strukturierter Datenbanken (blau markiert) im DBMS-Ranking (basierend auf [49]).

Die Installation von Open-Source-Datenbanken ist recht einfach - auch ohne umfangreiche technische Kenntnisse. Open-Source-Systeme, wie PostgreSQL, MySQL oder SQLite, sind kostenlos erhältlich und funktionieren auf den meisten Betriebssystemen: Windows, macOS und Linux. Alles, was Sie tun müssen, ist, die offizielle Website des Projekts aufzurufen, das Installationsprogramm herunterzuladen und den Anweisungen zu folgen. In den meisten Fällen dauert die Installation nicht länger als 10-15 Minuten. Im vierten Teil des Buches werden wir eine solche Datenbank modellieren und erstellen (Abb. 4.3-8).

Wenn Ihr Unternehmen Cloud-Dienste nutzt (z. B. Amazon Web Services, Google Cloud oder Microsoft Azure), können Sie die Datenbank mit ein paar Klicks bereitstellen - die Plattform bietet Ihnen vorgefertigte Vorlagen für die Installation. Dank der Offenheit des Codes lassen sich solche Datenbanken leicht an Ihre Aufgaben anpassen, und eine große Gemeinschaft von Nutzern wird Ihnen immer helfen, eine Lösung für jedes Problem zu finden.

RDBMS bilden nach wie vor die Grundlage für eine Vielzahl von Geschäftsanwendungen und Analyseplattformen (Abb. 3.1-6), die es Unternehmen ermöglichen, Daten effizient zu speichern, zu verarbeiten und zu analysieren - und damit fundierte und zeitnahe Entscheidungen zu treffen.

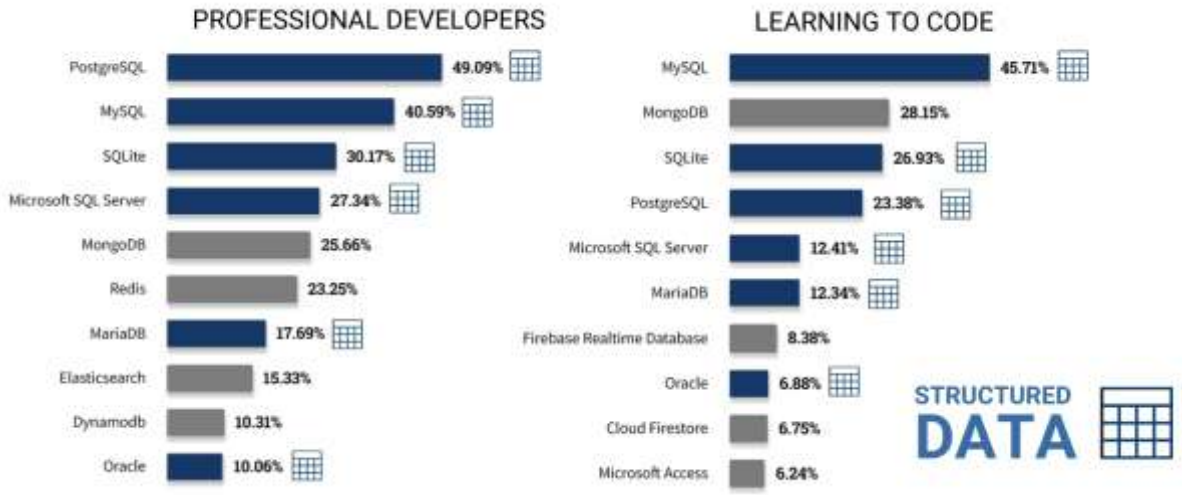


Abb. 3.1-6 Umfrage unter Entwicklern bei StackOverFlow (dem größten IT-Forum), welche Datenbanken sie im letzten Jahr verwendet haben und welche sie im nächsten Jahr verwenden wollen (RDBMS sind blau hervorgehoben) (basierend auf [50]).

RDBMS bieten Zuverlässigkeit, Datenkonsistenz, Transaktionsunterstützung und verwenden eine leistungsstarke Abfragesprache - SQL (Structured Query Language), die häufig in der Analytik eingesetzt wird und es ermöglicht, in Datenbanken gespeicherte Informationen leicht abzurufen, zu ändern und zu analysieren. SQL ist das wichtigste Werkzeug für die Arbeit mit Daten in relationalen Systemen.

SQL - Abfragen in Datenbanken und neue Trends

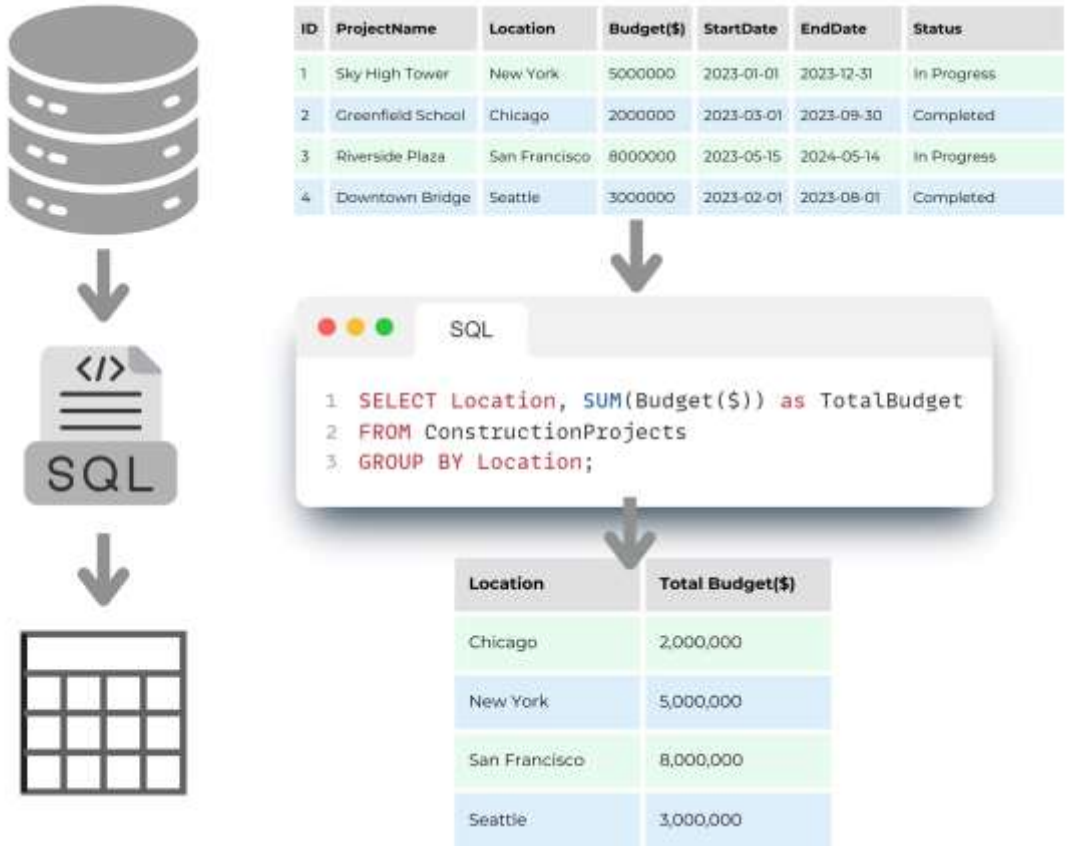
Der Hauptvorteil der SQL-Sprache, die häufig in relationalen Datenbanken verwendet wird, gegenüber anderen Arten der Informationsverwaltung (z. B. mit Hilfe klassischer Excel-Tabellen) ist die Unterstützung sehr großer Datenbankvolumina bei hoher Geschwindigkeit der Abfrageverarbeitung.

> Structured Query Language (SQL) ist eine spezielle Programmiersprache, die für die Speicherung, Verarbeitung und Analyse von Informationen in relationalen Datenbanken entwickelt wurde. SQL wird zur Erstellung, Verwaltung und zum Zugriff auf Daten verwendet und ermöglicht es, Informationen effizient zu finden, zu filtern, zu kombinieren und zu aggregieren. Sie ist ein wichtiges Werkzeug für den Datenzugriff und bietet eine bequeme und formalisierte Möglichkeit, mit Informationsspeichern zu interagieren.

Die Entwicklung von SEQUEL-SQL Systemen geht durch bedeutende Produkte und Unternehmen wie Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL und MySQL und gipfelt in der Entstehung von SQLite und MariaDB [51]. SQL bietet Tabellenkalkulationsfunktionen, die in Excel nicht zu finden sind, und macht die Datenverarbeitung skalierbarer, sicherer und einfacher zu automatisieren:

- **Erstellen und Verwalten von Datenstrukturen (DDL):** In SQL können Sie Tabellen in einer Datenbank erstellen, ändern und löschen, Verbindungen zwischen ihnen herstellen und Datenspeicherstrukturen definieren. Excel hingegen arbeitet mit festen Blättern und Zellen, ohne klar definierte Beziehungen zwischen Blättern und Datensätzen.

- **Datenmanipulation (DML):** Mit SQL können Sie Daten in großem Umfang und mit hoher Geschwindigkeit hinzufügen, ändern, löschen und abrufen, indem Sie komplexe Abfragen mit Filterung, Sortierung und Tabellen-Joins durchführen (Abb. 3.1-7). In Excel erfordert die Verarbeitung großer Datenmengen manuelle Aktionen oder spezielle Makros, was den Prozess verlangsamt und die Fehlerwahrscheinlichkeit erhöht.
- **Zugriffskontrolle (DCL):** SQL ermöglicht es, die Zugriffsrechte auf die Daten für verschiedene Benutzer zu differenzieren und die Möglichkeit, Informationen zu bearbeiten oder anzuzeigen, einzuschränken. In Excel hingegen wird der Zugriff entweder gemeinsam genutzt (bei der Übertragung einer Datei) oder erfordert komplexe Einstellungen mit der gemeinsamen Nutzung von Berechtigungen über Cloud-Dienste.



ID	ProjectName	Location	Budget($)	StartDate	EndDate	Status
1	Sky High Tower	New York	5000000	2023-01-01	2023-12-31	In Progress
2	Greenfield School	Chicago	2000000	2023-03-01	2023-09-30	Completed
3	Riverside Plaza	San Francisco	8000000	2023-05-15	2024-05-14	In Progress
4	Downtown Bridge	Seattle	3000000	2023-02-01	2023-08-01	Completed

Location	Total Budget($)
Chicago	2,000,000
New York	5,000,000
San Francisco	8,000,000
Seattle	3,000,000

Abb. 3.1-7 Beispiel für DML in SQL: schnelle Verarbeitung, Gruppierung und Aggregation mit wenigen Zeilen Code für die automatische Datenverarbeitung.

Excel erleichtert die Arbeit mit Daten durch seine visuelle und intuitive Struktur. Mit zunehmender Datenmenge nimmt jedoch die Leistung von Excel ab. Außerdem ist Excel bei der Datenmenge, die es speichern kann, an Grenzen gestoßen - maximal eine Million Zeilen - und die Leistung nimmt ab, lange bevor diese Grenze erreicht ist. Während Excel also für die Visualisierung und Bearbeitung kleiner Datenmengen vorzuziehen ist, eignet sich SQL besser für den Umgang mit großen Datensätzen.

Der nächste Schritt in der Entwicklung strukturierter Daten war das Aufkommen spaltenförmiger Datenbanken (Columnar Databases), die eine Alternative zu den traditionellen relationalen Datenbanken darstellen,

insbesondere wenn es um wesentlich größere Datenmengen und analytische Berechnungen geht. Im Gegensatz zu Zeilendatenbanken, in denen die Daten zeilenweise gespeichert werden, werden in spaltenorientierten Datenbanken die Informationen spaltenweise gespeichert. Im Vergleich zu klassischen Datenbanken ist dies möglich:

- Reduzieren Sie den Speicherplatz durch effiziente Komprimierung einheitlicher Daten in Spalten.
- Beschleunigung von analytischen Abfragen, da nur die benötigten Spalten gelesen werden und nicht die gesamte Tabelle.
- Optimieren Sie Big Data und Data Warehousing, z.B. Data Lakehouse Architecture.

In den folgenden Kapiteln dieses Buches - "DataFrame: ein universelles tabellarisches Datenformat" und "Datenspeicherformate und Arbeiten mit Apache Parquet: DWH -Data Warehouses und Data Lakehouse-Architektur " - werden wir mehr über kolumnare Datenbanken, Pandas DataFrame, Apache Parquet, HDF5 sowie über die Erstellung von darauf basierenden Big Data -Speichern zur Datenanalyse und -verarbeitung erfahren.

Unstrukturierte Daten

Obwohl die meisten Daten, die in Anwendungen und Informationssystemen verwendet werden, in strukturierter Form vorliegen, sind die meisten Informationen, die im Bauwesen erzeugt werden, unstrukturierte Daten - Bilder, Videos, Textdokumente, Tonaufnahmen und andere Formen von Inhalten. Dies gilt insbesondere für die Phasen des Baus, des Betriebs und der technischen Überwachung, in denen visuelle und textuelle Informationen überwiegen.

> **Bei unstrukturierten Daten** handelt es sich um Informationen ohne vordefiniertes Modell oder Struktur, die nicht in herkömmlichen Zeilen und Spalten wie in Datenbanken oder Tabellen organisiert sind.

Im Allgemeinen können unstrukturierte Daten in zwei Kategorien eingeteilt werden:

- Von Menschen erstellte unstrukturierte Daten, die verschiedene Arten von von Menschen erstellten Inhalten umfassen: Textdokumente, E-Mails, Bilder, Videos usw.
- Maschinengenerierte unstrukturierte Daten werden von Geräten und Sensoren erzeugt: Dazu gehören Protokolldateien, GPS-Daten, Ergebnisse aus dem Internet der Dinge (IoT) und andere Telemetriedaten z. B. von einer Baustelle.

Im Gegensatz zu strukturierten Daten, die bequem in Tabellen und Datenbanken organisiert werden können, erfordern unstrukturierte Daten zusätzliche Verarbeitungsschritte, bevor sie in Informationssysteme integriert werden können (Abb. 3.1-8). Der Einsatz von Technologien zur automatisierten Erfassung, Analyse und Umwandlung solcher Daten eröffnet neue Möglichkeiten, die Effizienz im Bauwesen zu verbessern, Fehler zu reduzieren und den Einfluss des menschlichen Faktors zu minimieren.

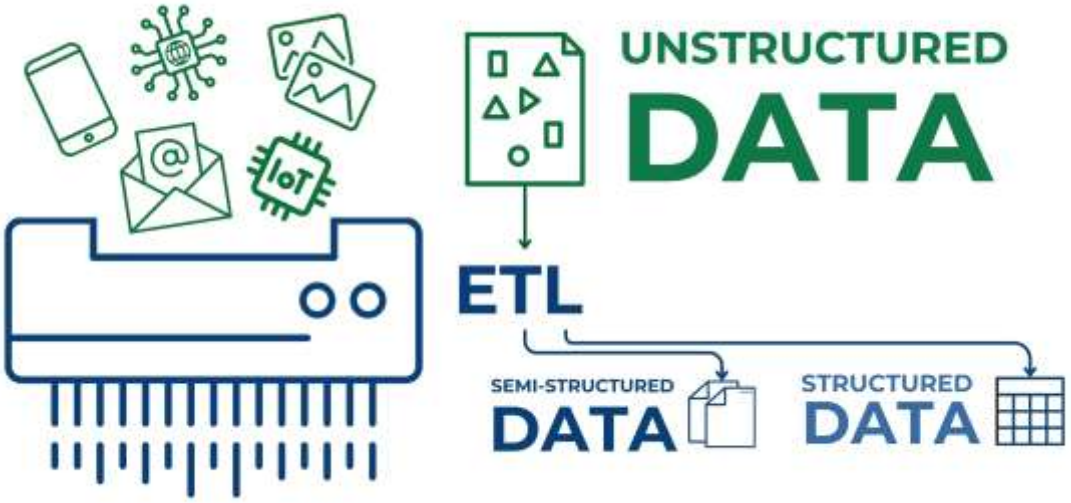


Abbildung 3.1-8 Die Verarbeitung von unstrukturierten Daten beginnt mit ihrer Umwandlung in halbstrukturierte und strukturierte Daten.

Unstrukturierte Daten machen bis zu 80% aller Informationen [52] aus, mit denen die Fachleute in den Unternehmen konfrontiert werden, weshalb wir in den folgenden Kapiteln des Buches ausführlich auf ihre Arten und ihre Verarbeitung anhand von Beispielen eingehen werden.

Um die Diskussion zu erleichtern, werden Textdaten gesondert kategorisiert. Obwohl es sich um eine Art von eher unstrukturierten Daten handelt, erfordern ihre Bedeutung und ihre Verbreitung in der Baubranche besondere Aufmerksamkeit.

Textdaten: zwischen unstrukturiertem Chaos und strukturiertem Chaos y

Textdaten im Baugewerbe umfassen ein breites Spektrum an Formaten und Arten von Informationen, von Papierdokumenten bis hin zu informellen Kommunikationsmethoden wie Briefen, Gesprächen, Arbeitskorrespondenz und mündlichen Besprechungen auf der Baustelle. Alle diese Textdaten enthalten wichtige Informationen für das Management von Bauprojekten, von Details zu Entwurfsentscheidungen und Planänderungen bis hin zu Diskussionen über Sicherheitsfragen und Verhandlungen mit Auftragnehmern und Kunden (Abb. 3.1-9).

Abb. 3.1-9 Textdaten, eine der am häufigsten verwendeten Informationsarten in der Kommunikation zwischen Projektteilnehmern.

Textuelle Informationen können sowohl formalisiert als auch unstrukturiert sein. Zu den formalisierten Daten gehören Word-Dokumente (.doc,.docx), PDF sowie Textdateien von Sitzungsprotokollen (.txt). Zu den informellen Daten gehören Messenger- und E-Mail-Korrespondenz, Sitzungsprotokolle (Teams, Zoom, Google Meet) und Audioaufnahmen von Diskussionen (.mp3,.wav), die in Text umgewandelt werden müssen.

Doch während schriftliche Dokumente wie formelle Anfragen, Vertragsbedingungen und E-Mails in der Regel bereits eine gewisse Struktur aufweisen, bleibt die mündliche Kommunikation und Arbeitskorrespondenz oft unstrukturiert, was ihre Analyse und Integration in Projektmanagement-Systeme erschwert.

Der Schlüssel zu einer effizienten Verwaltung von Textdaten liegt in ihrer Umwandlung in ein strukturiertes Format. Dadurch können die verarbeiteten Informationen automatisch in bestehende Systeme integriert werden, die bereits mit strukturierten Daten arbeiten.

Abbildung 3.1-10 Umwandlung von Textinhalten in strukturierte Daten.

Um textuelle Informationen effektiv nutzen zu können, müssen sie automatisch in eine strukturierte Form umgewandelt werden (Abb. 3.1-10). Dieser Prozess umfasst in der Regel mehrere Schritte:

- **Texterkennung (OCR**) - Umwandlung von Bildern von Dokumenten und Zeichnungen in ein maschinenlesbares Format.
- **Textanalyse (NLP**) - automatische Identifizierung von Schlüsselparametern (Daten, Beträge und Zahlen, die für das Projekt relevant sind).
- **Datenklassifizierung** - Kategorisierung von Informationen (Finanzen, Logistik, Risikomanagement).

Nach der Erkennung und Klassifizierung können die bereits strukturierten Daten in Datenbanken integriert und in automatisierten Berichts- und Verwaltungssystemen verwendet werden.

Semi-strukturierte und lose strukturierte Daten

Halbstrukturierte Daten enthalten ein gewisses Maß an Organisation, haben aber kein strenges Schema oder eine Struktur. Obwohl solche Informationen strukturierte Elemente enthalten (z. B. Daten, Namen von Mitarbeitern und Listen erledigter Aufgaben), kann das Format der Darstellung von Projekt zu Projekt oder sogar von einem Mitarbeiter zum anderen erheblich variieren. Beispiele für solche Daten sind Zeitprotokolle, Fortschrittsberichte und Zeitpläne, die in einer Vielzahl von Formaten dargestellt werden können.

Halbstrukturierte Daten sind leichter zu analysieren als unstrukturierte Daten, erfordern aber eine zusätzliche Verarbeitung für die Integration in standardisierte Projektmanagementsysteme.

> Die Arbeit mit halbstrukturierten Daten, die sich durch eine ständig wechselnde Struktur auszeichnen, stellt eine große Herausforderung dar. Dies liegt daran, dass die Variabilität der Datenstruktur separate individuelle Ansätze für die Verarbeitung und Analyse jeder Quelle halbstrukturierter Daten erfordert.

Doch während der Umgang mit unstrukturierten Daten viel Aufwand erfordert, lässt sich die Verarbeitung halbstrukturierter Daten mit relativ einfachen Methoden und Tools bewerkstelligen.

> **Schwach strukturierte** Daten ist ein allgemeinerer Begriff, der Daten mit minimaler oder unvollständiger Struktur beschreibt. Meistens handelt es sich dabei um Textdokumente, Chats, E-Mails, in denen zwar einige Metadaten (z. B. Datum, Absender) zu finden sind, der Großteil der Informationen jedoch in chaotischer Weise präsentiert wird.

Im Bauwesen sind lose strukturierte Daten in einer Vielzahl von Prozessen zu finden. Sie können zum Beispiel umfassen:

- Kostenvoranschläge und Kostenvoranschläge - Tabellen mit Angaben zu Materialien, Mengen und Kosten, aber ohne einheitliches Format.
- Zeichnungen und technische Diagramme - Dateien im PDF-Format oder DWG, die Textanmerkungen und Metadaten enthalten, aber keine streng festgelegte Struktur haben.
- Arbeitspläne - Daten aus MS Project, Primavera P6 oder anderen Systemen, die eine andere Exportstruktur haben können.
- CAD (BIM -Modelle) - enthalten Elemente der Struktur, aber die Datendarstellung hängt von der Software und dem Projektstandard ab.

Geometrische Daten, die von CAD-Systemen erzeugt werden, können auf die gleiche Weise wie halbstrukturierte Daten kategorisiert werden. Wir werden jedoch geometrische CAD-Daten (BIM) als einen separaten Datentyp klassifizieren, da sie, wie Textdaten, in Unternehmensprozessen oft als separater Datentyp behandelt werden können.

Geometrische Daten und ihre Anwendung

Während Metadaten über Projektelemente fast immer in Form von Tabellen, strukturierten oder locker strukturierten Formaten gespeichert werden, werden die geometrischen Daten von Projektelementen in den meisten Fällen mit speziellen CAD-Tools (Abb. 3.1-11) erstellt, die es ermöglichen, Projektelemente im Detail als eine Reihe von Linien (2D) oder geometrischen Körpern (3D) zu visualisieren.

Abb. 3.1-11 CAD Tools haben dazu beigetragen, geometrische Informationen von physischen Medien in eine Datenbank zu übertragen.

Bei der Arbeit mit Geometriedaten im Bauwesen und in der Architektur lassen sich drei Hauptanwendungen von Geometriedaten unterscheiden (Abb. 3.1-12):

- **Bestätigung von Volumina**: Geometrische Daten, die in CAD-Programmen (BIM) unter Verwendung spezieller geometrischer Kerne erzeugt werden, sind erforderlich, um automatisch und genau die Volumina und Abmessungen von Projektelementen zu bestimmen. Diese Daten umfassen automatisch berechnete Flächen, Volumen, Längen und andere wichtige Attribute, die für die Planung, Budgetierung und Bestellung von Ressourcen und Materialien erforderlich sind
- **Visualisierung des Projekts**: Bei Änderungen des Projekts ermöglicht die Visualisierung der Elemente die automatische Erstellung von aktualisierten Zeichnungen in verschiedenen Ebenen. Die Visualisierung des Projekts in der Anfangsphase hilft, das Verständnis zwischen allen Beteiligten zu beschleunigen, um Zeit und Ressourcen während des Bauprozesses zu sparen.
- **Überprüfung von Kollisionen**: Bei komplexen Bau- und Konstruktionsprojekten, bei denen das Zusammenspiel mehrerer Elementkategorien (z. B. Rohre und Wände) ohne "geometrische Konflikte" kritisch ist, spielt die Kollisionsprüfung eine entscheidende Rolle. Mit der Kollisionserkennungssoftware können Sie potenzielle geometrische Konflikte zwischen Projektelementen proaktiv erkennen und so kostspielige Fehler während des Bauprozesses vermeiden.

Seit den Anfängen der Ingenieurbüros, seit der Errichtung der ersten komplexen Bauwerke, lieferten die Bauingenieure geometrische Informationen in Form von Zeichnungen, Linien und flachen geometrischen Elementen (auf Papyrus, "A0"-Papier oder im DWG-Format, PDF, PLT), auf deren Grundlage die Vorarbeiter und die Kalkulatoren (Abb. 3.1-11), in den letzten Jahrtausenden mit Hilfe von Linealen und Winkelmessern attributive Volumina oder Mengen von Elementen und Elementgruppen sammelten. 3.1-11), in den letzten Jahrtausenden mit Hilfe von Linealen und Transportern die attributiven Volumina oder Mengen von Elementen und Gruppen von Elementen gesammelt haben.

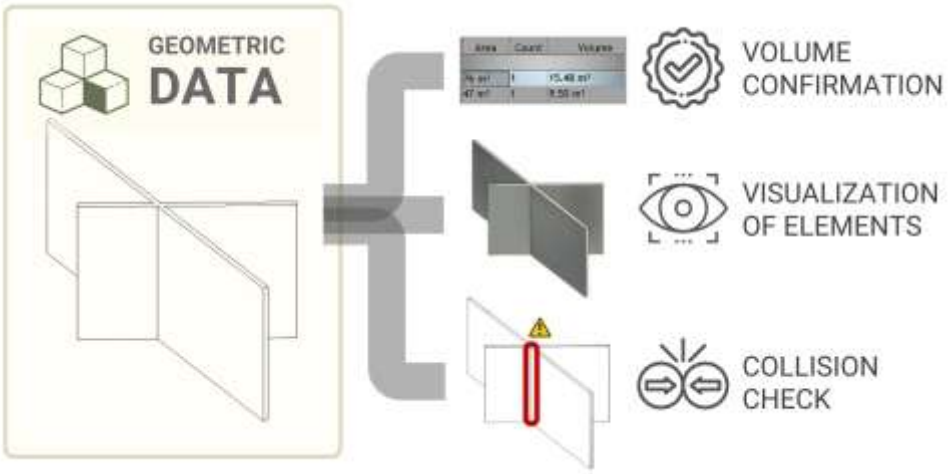


Abb. 3.1-12 Die Geometrie ist die Grundlage für die Ermittlung der volumetrischen Parameter der Elemente, die dann zur Berechnung der Kosten und des Zeitplans des Projekts verwendet werden.

Heute ist diese manuelle und zeitaufwändige Aufgabe dank der Entwicklung der volumetrischen Modellierung in modernen CAD-Werkzeugen (BIM) vollständig automatisiert, so dass mit Hilfe eines speziellen geometrischen Kerns automatisch die volumetrischen Attribute jedes Elements ermittelt werden können, ohne dass die volumetrischen Parameter manuell berechnet werden müssen.

Moderne CAD-Werkzeuge ermöglichen es auch, Projektelemente zu klassifizieren und zu kategorisieren, so dass Sie Spezifikationstabellen aus der Projektdatenbank hochladen können, um sie in verschiedenen Systemen zu verwenden, wie z. B. Kostenschätzung, Terminplanung oder CO_2-Berechnung (Abb. 3.1-13). Wir werden die Beschaffung von Spezifikationen, QTO-Tabellen und Mengen sowie praktische Beispiele im Kapitel "Beschaffung von Mengen und Quantifizierung" besprechen.

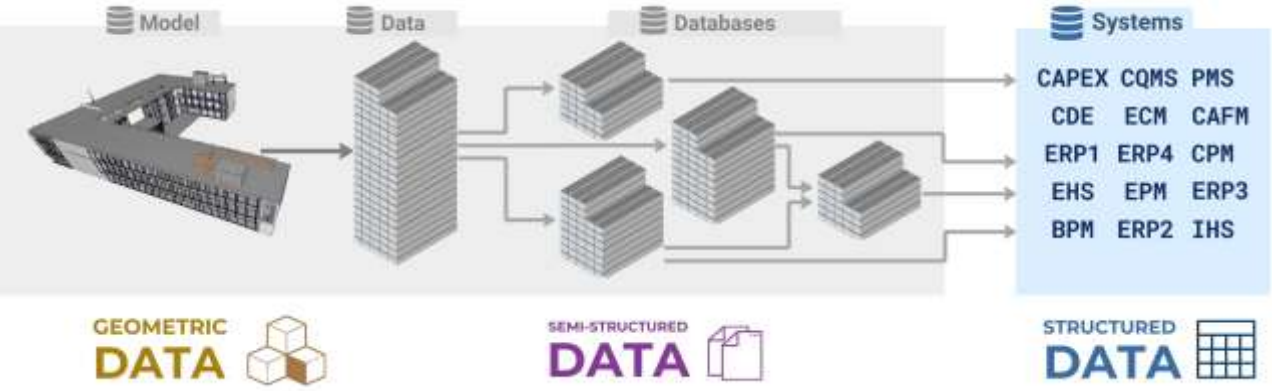


Abb. 3.1-13 CAD-Tools (BIM) speichern Daten in Datenbanken, die für die Integration und Interaktion mit anderen Systemen ausgelegt sind.

Aufgrund des geschlossenen Charakters der in der CAD-Umgebung verwendeten Datenbanken und Formate sind die in CAD-Lösungen erstellten Geometriedaten zu einer eigenen Informationsart geworden. Sie umfassen sowohl die Geometrie der Elemente als auch Metainformationen (strukturiert oder halbstrukturiert),

die in speziellen Dateien und Formaten enthalten sind.

CAD Daten: vom Entwurf bis zur Datenspeicherung

Moderne CAD- und BIM-Systeme speichern Daten in ihren eigenen, oft proprietären Formaten: DWG, DXF, RVT, DGN, PLN und andere. Diese Formate unterstützen sowohl 2D- als auch 3D-Darstellungen von Objekten, wobei nicht nur die Geometrie, sondern auch die mit den Objekten verbundenen Attribute erhalten bleiben. Hier sind die gängigsten Formate:

- **DWG** ist ein binäres Dateiformat, das zum Speichern von zweidimensionalen (und seltener dreidimensionalen) Entwurfsdaten und Metadaten verwendet wird.
- **DXF** ist ein Textformat für den Austausch von 2D- und 3D-Zeichnungen -Zeichnungen zwischen CAD-Systemen. Es enthält Geometrie-, Layer- und Attributdaten und unterstützt sowohl die ASCII- als auch die Binärdarstellung.
- **RVT** ist ein binäres Format für die Speicherung von CAD-Modellen, einschließlich 3D -Geometrie, Elementattribute, Beziehungen und Konstruktionsparameter.
- **IFC** ist ein offenes Textformat für den Austausch von Konstruktionsdaten zwischen CAD (BIM) Systemen. Es umfasst Geometrie, Objekteigenschaften und Informationen über ihre Beziehungen.

Neben diesen Formaten werden noch weitere verwendet: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Obwohl sie sich in Bezug auf ihren Zweck und den Grad der Offenheit unterscheiden (Abb. 3.1-14), können sie alle dasselbe Projektinformationsmodell in verschiedenen Formen darstellen. In komplexen Projekten werden diese Formate oft parallel verwendet, von der Zeichnung bis zur Koordinierung der Projektmodelle.

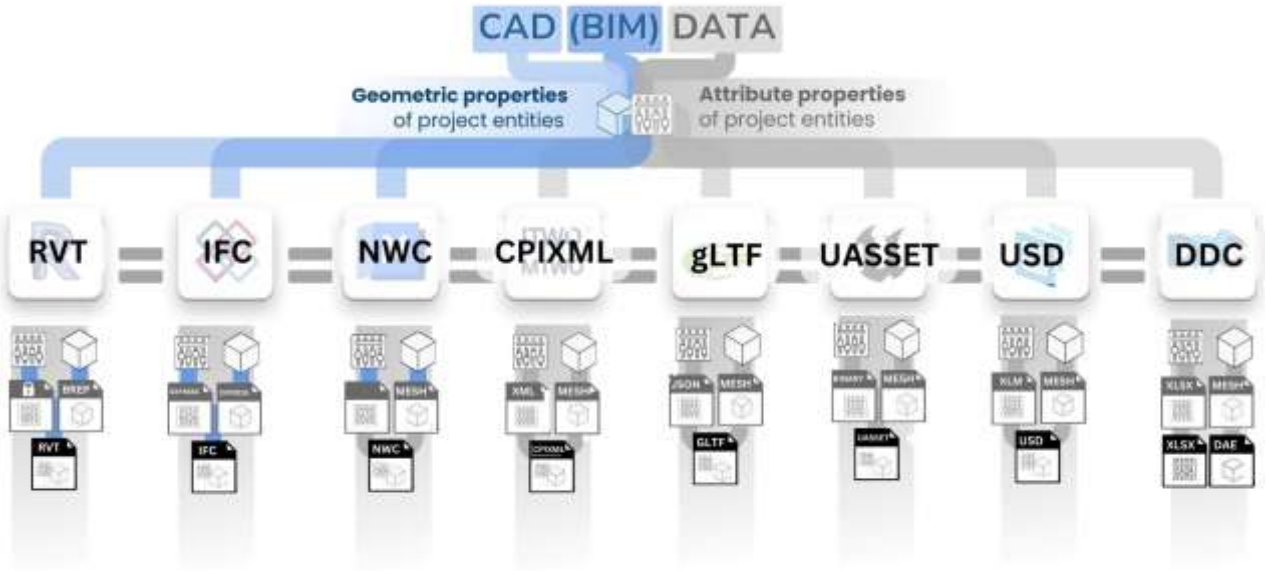


Abbildung 3.1-14 Gängige CAD-Speicherformate beschreiben die Geometrie durch BREP- oder MESH-Parameter, die durch Attributdaten ergänzt werden.

Alle oben genannten Formate ermöglichen es Ihnen, Daten über jedes Element eines Bauprojekts zu speichern, und alle oben genannten Formate enthalten zwei wichtige Arten von Daten:

- **Geometrische Parameter** - beschreiben die Form, Lage und Abmessungen eines Objekts. Die Geometrie und ihre Verwendung werden im sechsten Teil des Buches, der den CAD-Lösungen (BIM) gewidmet ist, eingehend behandelt;
- **Attributeigenschaften** - enthalten verschiedene Informationen: Materialien, Elementtypen, technische Merkmale, eindeutige Bezeichner und andere Eigenschaften, die Projektelemente haben können.

Attributdaten sind in modernen Projekten von besonderer Bedeutung, da sie die Betriebseigenschaften von Objekten definieren, technische und kalkulatorische Berechnungen ermöglichen und eine durchgängige Interaktion zwischen den Beteiligten in Planung, Bau und Betrieb ermöglichen. Zum Beispiel:

- Bei Fenstern und Türen: Art der Konstruktion, Art der Verglasung, Öffnungsrichtung (Abb. 3.2-1).
- Bei Wänden werden Informationen über Materialien, Wärmedämmung und Schallschutz aufgezeichnet.
- Für technische Systeme werden die Parameter von Rohrleitungen, Kanälen, Kabeltrassen und deren Verbindungen gespeichert.

Diese Parameter können sowohl in den CAD-(BIM -)Dateien selbst als auch in externen Datenbanken gespeichert werden - als Ergebnis des Exports, der Konvertierung oder des direkten Zugriffs auf interne CAD-Strukturen über Reverse-Engineering-Tools. Dieser Ansatz erleichtert die Integration von Planungsinformationen mit anderen Unternehmenssystemen und -plattformen

> **Reverse Engineering** im Kontext von CAD (BIM) ist der Prozess der Extraktion und Analyse der internen Struktur eines digitalen Modells, um dessen Logik, Datenstruktur und Abhängigkeiten ohne Zugang zu den ursprünglichen Algorithmen oder der Dokumentation neu zu erstellen.

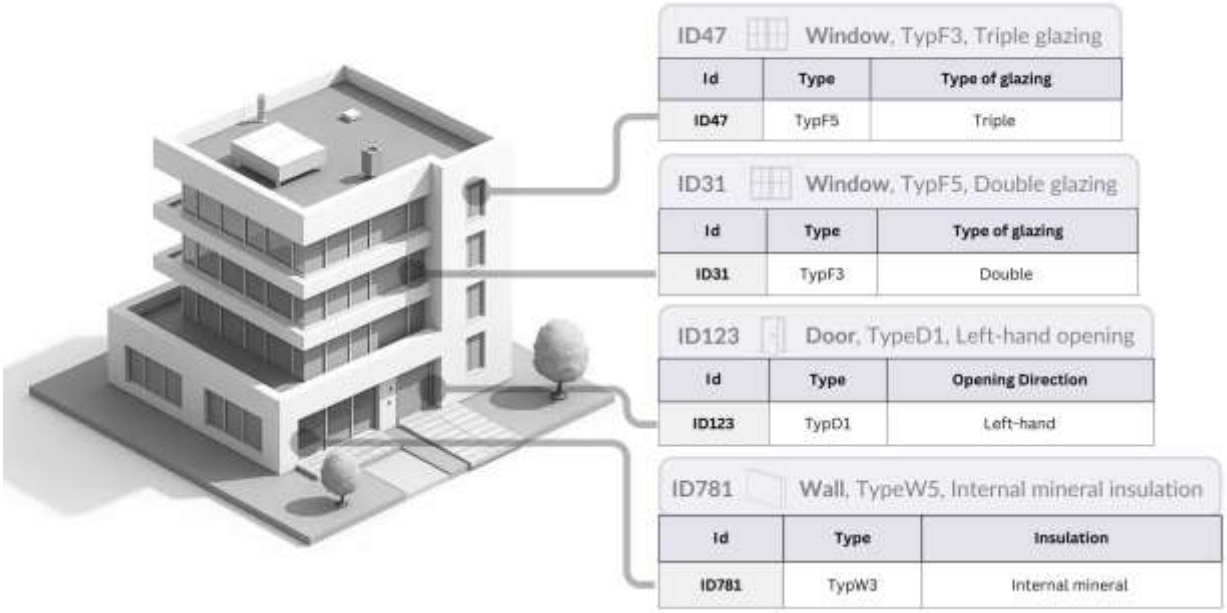


Abb. 3.1-15 Ein Projektelement beschreibt nicht nur eine parametrische oder polygonale Geometrie, sondern enthält auch Informationen über Parameter und Eigenschaften der Elemente.

Infolgedessen wird um jedes Element herum ein einzigartiger Satz von Parametern und Eigenschaften gebildet, der sowohl einzigartige Merkmale jedes Objekts (z. B. Bezeichner und Abmessungen) als auch gemeinsame Attribute für Gruppen von Elementen enthält. Auf diese Weise lassen sich nicht nur einzelne Elemente des Projekts analysieren, sondern auch zu logischen Gruppen zusammenfassen, die dann von anderen Spezialisten für ihre Aufgaben und Berechnungen in Systemen und Datenbanken verwendet werden können.

> Eine Entität ist ein konkretes oder abstraktes Objekt der realen Welt, das eindeutig identifiziert, beschrieben und in Form von Daten dargestellt werden kann.

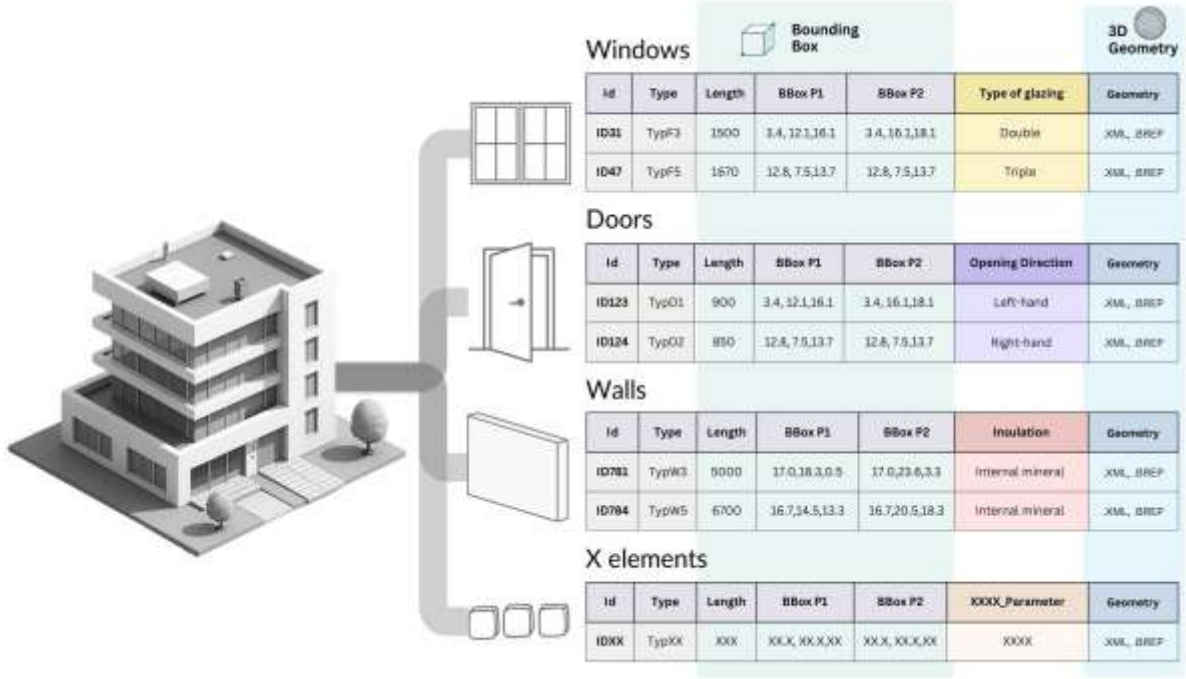


Abb. 3.1-16 Jedes Projektelement enthält Attribute, die entweder vom Konstrukteur eingegeben oder im Programm CAD berechnet werden.

In den letzten Jahrzehnten hat die Bauindustrie viele neue CAD-Formate (BIM) entwickelt, die die Erstellung, Speicherung und Übertragung von Daten vereinfachen. Diese Formate können geschlossen oder offen, tabellarisch, parametrisch oder grafisch sein. Ihre Vielfalt und Fragmentierung erschweren jedoch die Datenverwaltung in allen Phasen des Projektlebenszyklus erheblich. Eine Vergleichstabelle der wichtigsten Formate, die für den Informationsaustausch im Bauwesen verwendet werden, ist in Abbildung 3.1-17 dargestellt (vollständige Version über QR-Code erhältlich).

Um die Probleme der Interoperabilität und des Zugriffs auf CAD-Daten zu lösen, werden BIM-Manager (BIM) und Koordinatoren hinzugezogen, deren Aufgabe es ist, den Export zu steuern, die Datenqualität zu prüfen und Teile der CAD-(BIM-)Daten in andere Systeme zu integrieren.

Aufgrund des geschlossenen Charakters und der Komplexität der Formate ist es jedoch schwierig, diesen Prozess zu automatisieren, so dass Fachleute gezwungen sind, viele Vorgänge manuell durchzuführen,

ohne die Möglichkeit zu haben, vollwertige Inline-Datenverarbeitungsprozesse (Pipeline) aufzubauen.

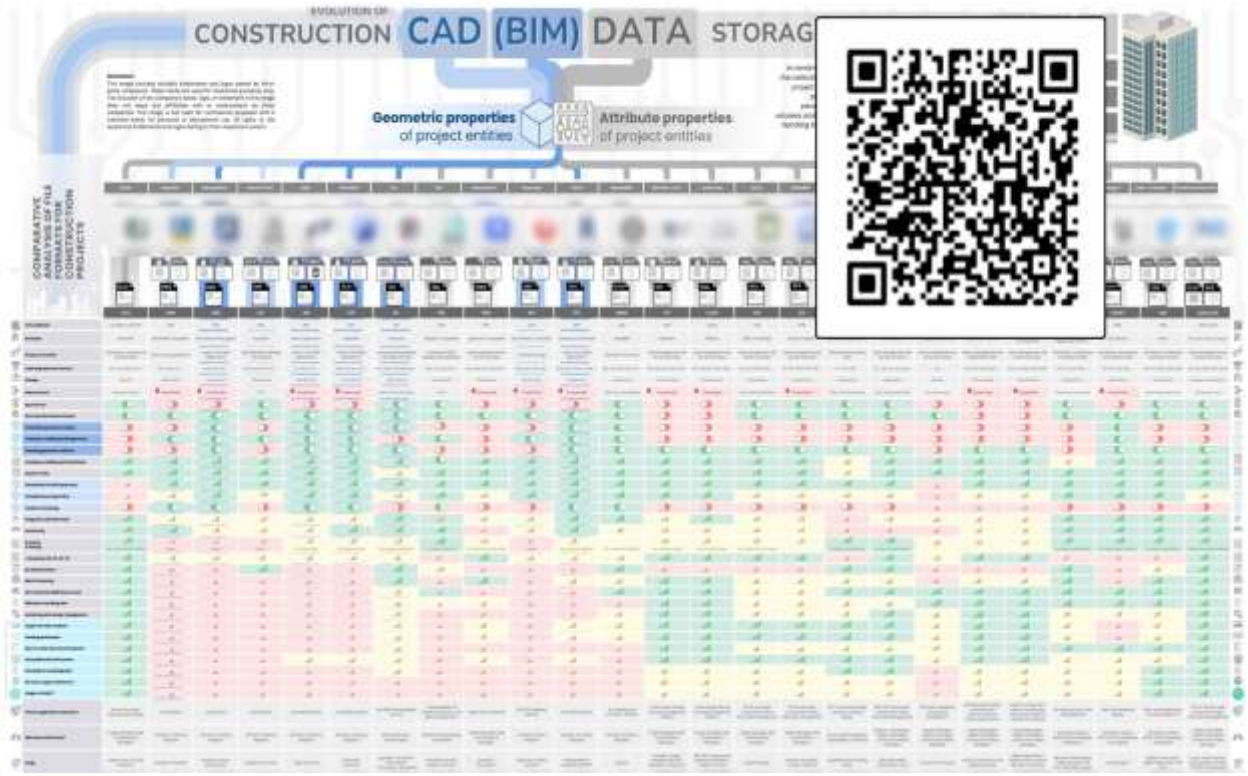


Abbildung 3.1-17 Tabelle zum Vergleich der wichtigsten Datenformate, in denen Projektelementinformationen gespeichert werden [53].

Um zu verstehen, warum es so viele verschiedene Datenformate gibt und warum die meisten von ihnen geschlossen sind, ist es wichtig, sich mit den Prozessen zu befassen, die innerhalb von CAD (BIM) Programmen ablaufen und die im sechsten Teil des Buches im Detail untersucht werden.

Eine zusätzliche Informationsebene zur Geometrie wurde von den Entwicklern von CAD-Systemen in Form des BIM-Konzepts (Building Information Modeling) eingeführt, ein Marketingbegriff, der seit 2002 in der Bauindustrie aktiv gefördert wird [54].

Das Aufkommen des BIM-Konzepts (BOM) und die Verwendung von CAD in den Prozessen

Das Konzept der Gebäudedatenmodellierung (BIM), das erstmals im BIM-Whitepaper [54] aus dem Jahr 2002 beschrieben wurde, entstand aus den Marketinginitiativen der CAD-Softwarehersteller. Es entstand aus den Marketinginitiativen der CAD-Softwareentwickler und war ein Versuch, die im Maschinenbau bereits gut etablierten Prinzipien an die Bedürfnisse der Bauindustrie anzupassen.

Die Inspiration für BIM kam vom Konzept der BOM (Bill of Materials) - Stückliste, das seit den späten 1980er Jahren in der Industrie weit verbreitet ist. Im Maschinenbau ermöglichte die Stückliste die Verknüpfung von Daten aus CAD-Systemen mit PDM- (Product Data Management), PLM- (Product Lifecycle Management) und ERP-Systemen und ermöglichte so eine ganzheitliche Verwaltung von Konstruktionsinformationen

während des gesamten Produktlebenszyklus (Abb. 3.1-8).

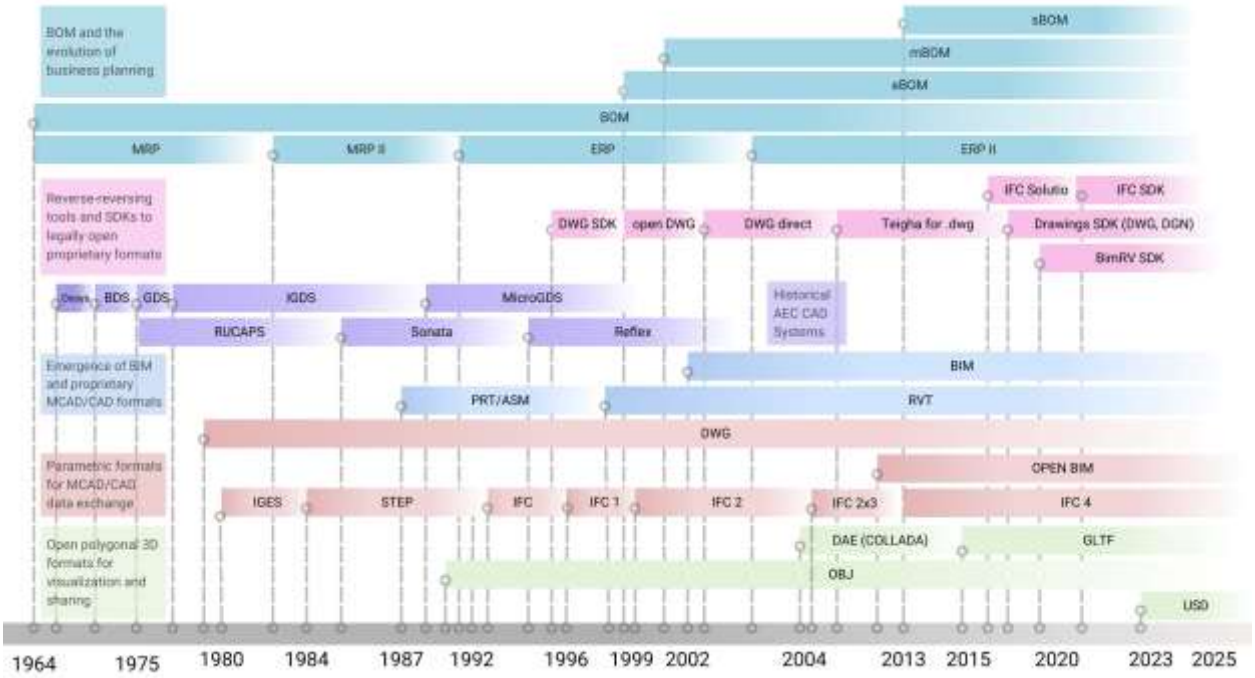


Abb. 3.1-18 Entwicklung von Spezifikationen (BOM), Informationsmodellierung (BIM) und digitalen Formaten in der Bauindustrie.

Die moderne Entwicklung des Stücklistenkonzepts hat zur Entstehung eines erweiterten Rahmens - XBOM (Extended BOM) - geführt, der nicht nur die Produktzusammensetzung, sondern auch Verhaltensszenarien, Betriebsanforderungen, Nachhaltigkeitsparameter und Daten für vorausschauende Analysen umfasst. XBOM erfüllt im Wesentlichen dieselbe Rolle wie BIM im Bauwesen: Beide Ansätze zielen darauf ab, das digitale Modell während des gesamten Lebenszyklus des Projekts zu einer Single Source of Truth für alle Projektbeteiligten zu machen.

Ein wichtiger Meilenstein bei der Einführung von Stücklisten im Bauwesen war die Einführung des ersten parametrischen CAD-Systems (MCAD), das speziell für die Bauindustrie entwickelt wurde, im Jahr 2002. Es wurde von dem Team entwickelt, das zuvor Pro-E® geschaffen hatte, ein revolutionäres MCAD-System für den Maschinenbau, das in den späten 1980er Jahren aufkam und zum Industriestandard geworden war [55].

Bereits in den späten 1980er Jahren war es das Ziel, die Einschränkungen [56] der damals existierenden CAD -Programme zu beseitigen. Hauptziel war es, den Arbeitsaufwand für die Änderung der Parameter von Konstruktionselementen zu reduzieren und die Aktualisierung des Modells auf der Basis von Daten außerhalb der CAD-Programme über eine Datenbank zu ermöglichen [57]. Die wichtigste Rolle sollte dabei die Parametrisierung spielen: das automatische Abrufen von Merkmalen aus der Datenbank und deren Verwendung zur Aktualisierung des Modells innerhalb der CAD-Systeme.

> Pro-E und das ihm zugrunde liegende Konzept der elementaren parametrischen Modellierung c BOM haben die Entwicklung des CAD - und MCAD - Marktes maßgeblich beeinflusst [58]. Seit 25 Jahren ist dieses Modell in der Industrie vertreten und viele moderne Systeme sind seine konzeptionellen Nachfolger geworden.

> *Das Ziel ist es, ein System zu schaffen, das flexibel genug ist, um den Ingenieur zu ermutigen, verschiedene Entwürfe in Betracht zu ziehen. Und die Kosten für Änderungen am Entwurf sollten so nahe bei Null liegen wie möglich. Herkömmliche CAD / CAM Software schränkt die Möglichkeit, kostengünstige Änderungen vorzunehmen, in unrealistischer Weise auf den Beginn des Konstruktionsprozesses ein [59].*
>
> - Samuel Heisenberg, Gründer der Parametric Technology Corporation®, Entwickler des MCAD -Produkts Pro-E und Lehrer des Erfinders eines CAD-Produkts im RVT-Format

Im Maschinenbau sind PDM-, PLM-, MRP- und ERP-Systeme zu wichtigen Plattformen geworden. Sie spielen eine zentrale Rolle im Daten- und Prozessmanagement, sammeln Informationen aus CAx-Systemen (CAD, CAM, CAE) und organisieren die Konstruktionsaktivitäten auf der Grundlage der Produktstruktur (BOM: eBOM, pBOM, mBOM) (Abb. 3.1-18). Diese Integration reduziert Fehler, vermeidet Datendopplungen und gewährleistet eine durchgängige Rückverfolgbarkeit vom Entwurf bis zur Produktion.

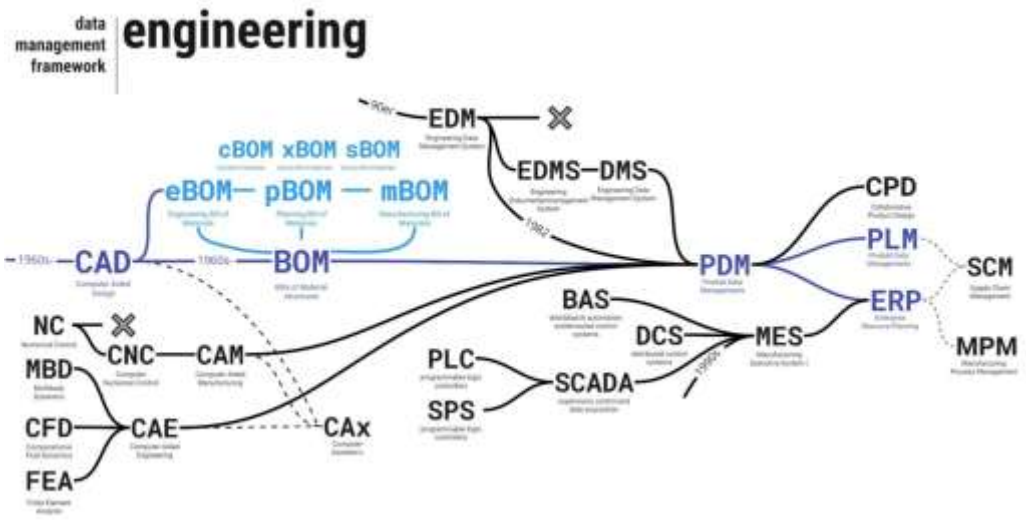


Abbildung 3.1-19 Historisch gesehen entstand die Stückliste in den 1960er Jahren als eine Möglichkeit, Daten aus CAx-Systemen zu strukturieren und an Steuerungssysteme weiterzugeben.

Der Kauf einer vom ehemaligen Pro-E-Team entwickelten und auf dem BOM-Ansatz basierenden CAD-Lösung durch einen der führenden Anbieter wurde fast unmittelbar durch die Veröffentlichung der BIM-Whitepaper-Serie (2002-2003)[60][61] markiert. Bereits ab Mitte der 2000er Jahre wurde das BIM-Konzept in der Baubranche aktiv gefördert, was das Interesse an parametrischer Software deutlich erhöhte. Die Popularität wuchs so schnell, dass die Konstruktionsgabel des Maschinenbauers Pro-E - parametrisches CAD, das von diesem Anbieter gefördert wird - die Konkurrenten im Bereich der Architektur- und Tragwerksplanung sogar verdrängt hat (Abb. 3.1-20). Anfang der 2020er Jahre hat er de facto die weltweite Vorherrschaft auf dem BIM-Markt (CAD) gefestigt [62].

Abb. 3.1-20 Popularität von Google-Suchanfragen (RVT versus IFC): parametrisches CAD, erstellt vom ehemaligen Pro-E-Team mit Stücklistenunterstützung -BIM hat in fast allen Ländern der Welt an Popularität gewonnen.

In den letzten 20 Jahren hat die Abkürzung BIM eine Vielzahl von Interpretationen erfahren, deren unterschiedliche Bedeutungen auf die ersten Marketingkonzepte zurückzuführen sind, die Anfang der 2000er Jahre aufkamen. Die ISO-Norm 19650, die eine wichtige Rolle bei der Popularisierung des Begriffs spielte, sicherte den Status von BIM als "wissenschaftlich fundierten" Ansatz für das Informationsmanagement. Im Text der Norm selbst, der sich dem Datenmanagement während des gesamten Lebenszyklus von Objekten unter Verwendung von BIM widmet, wird die Abkürzung BIM zwar erwähnt, aber nie klar definiert

Die ursprüngliche Website des Anbieters, die in den Jahren 2002[60] und 2003[61] eine Reihe von Whitepapers zu BIM veröffentlichte, reproduzierte Marketingmaterialien zu den Konzepten BOM (Bills of Materials) und PLM (Product Lifecycle Management), die bereits in den 1990er Jahren in der Maschinenbausoftware Pro-E verwendet wurden [63].

> *Die Gebäudedatenmodellierung, ein innovativer neuer Ansatz für die Planung, den Bau und die Verwaltung von Gebäuden, der 2002 von..... [Name des CAD-Anbieters] eingeführt wurde, hat die Denkweise von Fachleuten in der ganzen Welt darüber verändert, wie die Technologie für die Planung, den Bau und die Verwaltung von Gebäuden eingesetzt werden kann.*
>
> - BIM-Whitepaper, 2003 [61]

Diese frühen Veröffentlichungen verknüpften BIM direkt mit dem Konzept einer zentralisierten integrierten Datenbank. Wie im Whitepaper von 2003 dargelegt, handelt es sich bei BIM um ein Gebäudeinformationsmanagement, bei dem alle Aktualisierungen in einem einzigen Repository erfolgen, wodurch sichergestellt wird, dass alle Zeichnungen, Schnitte und Spezifikationen (BOM - Bills of Materials) synchronisiert werden.

> *BIM wird als Building Information Management beschrieben, bei dem alle Aktualisierungen und Änderungen in einer Datenbank erfolgen. Egal, ob es sich um Schemata, Schnitte oder Zeichnungen handelt, alles ist immer koordiniert, konsistent und auf dem neuesten Stand.*
>
> - Website eines CAD-Unternehmens Anbieter mit BIM Whitepaper, 2003 [54]

Die Idee, das Design über eine einzige integrierte Datenbank zu verwalten, wurde bereits in den 1980er Jahren in der Forschung breit diskutiert. So enthielt das BDS-Konzept von Charles Eastman [57] 43 Verweise auf den Begriff "Datenbank" (Abb. 6.1-2). Bis 2004 hatte sich diese Zahl fast halbiert auf 23 im Whitepaper von 2002 über BIM [64]. Und Mitte der 2000er Jahre war das Thema Datenbanken praktisch aus den Marketingmaterialien der Anbieter und der Digitalisierungsagenda im Allgemeinen verschwunden.

Obwohl ursprünglich die Datenbank und der Zugang zu ihr als Kern des BIM -Systems konzipiert war, hat sich der Schwerpunkt im Laufe der Zeit auf Geometrie, Visualisierung und 3D verlagert. Derselbe Hersteller, der 1994 den IFC-Standard einführte und 2002 das BIM-Whitepaper veröffentlichte, wies Anfang der 2000er Jahre in seinem Whitepaper ausdrücklich auf die Grenzen neutraler Formate wie IGES, STEP und IFC sowie auf die Notwendigkeit eines direkten Zugangs zu CAD-Datenbanken hin:

> *Verschiedene Anwendungen können inkompatibel sein und neu eingegebene Daten können ungenau sein [...]. Das Ergebnis der traditionellen computergestützten Konstruktion [CAD]: höhere Kosten, längere Markteinführungszeiten und geringere Produktqualität. Heute verwenden alle wichtigen Anwendungen Industriestandard-Schnittstellen für den Low-Level-Datenaustausch. Durch die Verwendung der alten IGES-Standards oder des neuen STEP-Formats [IFC ist de facto und de jure eine Kopie des STEP/IGES-Formats] für den Datenaustausch zwischen Anwendungen verschiedener Hersteller können die Benutzer eine gewisse Datenkompatibilität zwischen den besten Produkten der jeweiligen Branche erreichen. IGES und STEP funktionieren jedoch nur auf niedrigem Niveau und können keine Daten austauschen, die so umfangreich sind wie die Informationen, die von den heute führenden Anwendungen erzeugt werden [...]. Und obwohl diese und andere Standards fast täglich verbessert werden, werden sie in Bezug auf den Datenreichtum immer hinter den Produkten der heutigen Anbieter zurückbleiben. [...] Programme innerhalb einer Anwendung müssen in der Lage sein, den Datenreichtum auszutauschen und zu erhalten, ohne auf neutrale Übersetzer wie IGES, STEP [IFC] oder PATRAN zurückgreifen zu müssen. Stattdessen sollten Rahmenanwendungen in der Lage sein, direkt auf die zugrunde liegende CAD-Datenbank zuzugreifen, damit die Details und die Genauigkeit der Informationen nicht verloren gehen.*
>
> - Whitepaper eines CAD-Anbieters (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale", 2000 [65]

So wurde bereits in den 1980er und frühen 2000er Jahren die Datenbank und nicht die Formatdatei oder das neutrale IFC-Format als Schlüsselelement des digitalen Designs in der CAD-Umgebung angesehen. Es wurde vorgeschlagen, auf Übersetzer zu verzichten und den Anwendungen einen direkten Zugriff auf die Daten zu ermöglichen. In der Realität jedoch begann das Konzept von BIM Mitte der 2020er Jahre einer "Teile und Herrsche"-Strategie zu ähneln, bei der die Interessen von Softwareanbietern, die geschlossene geometrische Kerne verwenden, Vorrang vor der Entwicklung eines offenen Informationsaustauschs haben.

BIM wird heute als integraler Bestandteil der Bauindustrie wahrgenommen. Doch in den letzten zwei Jahrzehnten sind die Versprechen einer vereinfachten Zusammenarbeit und Datenintegration weitgehend unerfüllt geblieben. Die meisten Lösungen sind immer noch an geschlossene oder neutrale Formate und spezialisierte Werkzeuge gebunden. In Teil 6 des Buches "CAD und BIM: Marketing, Realität und die Zukunft von Konstruktionsdaten im Bauwesen" werden wir uns ausführlich mit der Geschichte von BIM, Open BIM und IFC sowie mit den Themen Interoperabilität und geometrische Kerne beschäftigen.

> Die Branche steht heute vor der großen Herausforderung, vom traditionellen Verständnis von CAD (BIM) als Modellierungswerkzeug zu dessen Nutzung als vollwertige Datenbank überzugehen. Dies erfordert neue Ansätze für die Arbeit mit Informationen, die Abkehr von der Abhängigkeit von geschlossenen Ökosystemen und die Einführung offener Lösungen.

Mit der Entwicklung von Reverse-Engineering-Tools, die den Zugriff auf CAD-Datenbanken ermöglichen, sowie der Verbreitung von Open-Source- und LLM-Technologien entfernen sich Anwender und Entwickler in der Baubranche zunehmend von den vagen Begriffen der Softwareanbieter. Stattdessen verlagert sich der Fokus auf das, was wirklich zählt: Daten (Datenbanken) und Prozesse.

> Hinter den trendigen Akronymen und Visualisierungen verbergen sich Standardpraktiken des Datenmanagements: Speicherung, Übertragung und Transformation - also der klassische ETL-Prozess (Extract, Transform, Load). Wie in anderen Branchen auch, erfordert die Digitalisierung des Bauwesens nicht nur Austauschstandards, sondern auch einen klar strukturierten Umgang mit heterogenen Informationen.

Um das Potenzial von CAD-Daten (BIM) voll ausschöpfen zu können, müssen Unternehmen ihr Informationsmanagement neu überdenken. Dies wird unweigerlich zu einem Schlüsselelement der digitalen Transformation führen - der Vereinheitlichung, Standardisierung und sinnvollen Strukturierung der Daten, mit denen Baufachleute täglich arbeiten.

KAPITEL 3.2.
HARMONISIERUNG UND STRUKTURIERUNG DER DATEN

Befüllung von Systemen mit Daten in der Bauindustrie

Ob Großkonzerne oder mittelständische Unternehmen, täglich sind Spezialisten damit beschäftigt, Softwaresysteme und Datenbanken mit verschiedenen Schnittstellen mit vielgestaltigen Informationen zu füllen (Abb. 3.2-1), die mit Hilfe von Managern miteinander kooperieren müssen. Es ist dieser Komplex von interagierenden Systemen und Prozessen, der letztlich Umsatz und Gewinn für das Unternehmen schafft.

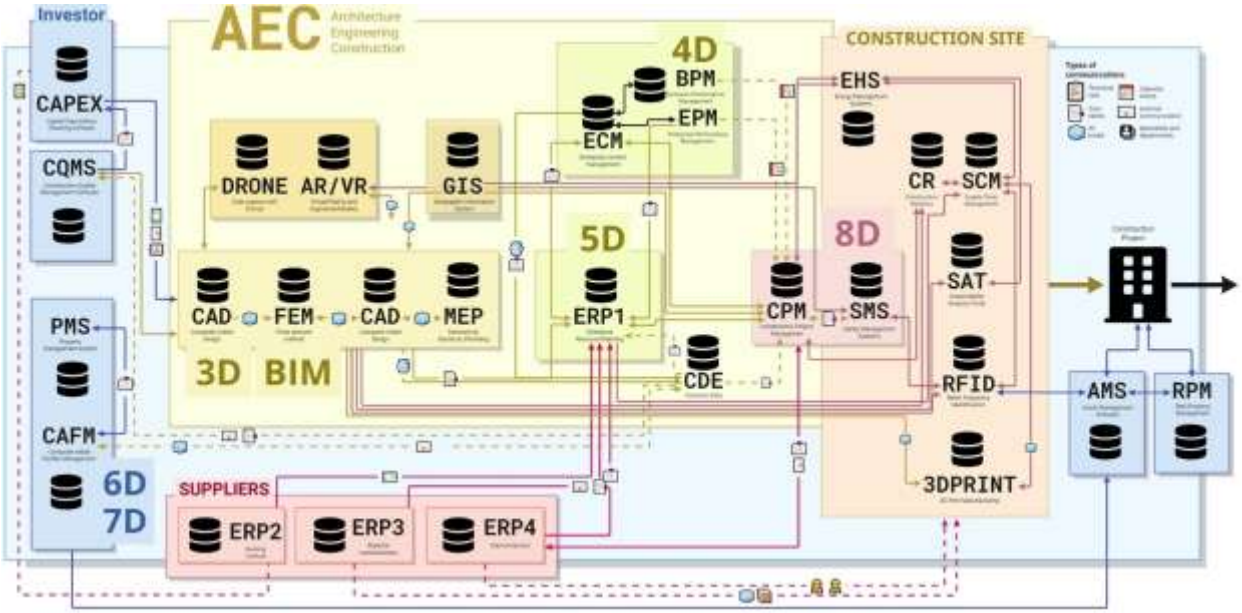


Abbildung 3.2-1 Praktisch jedes System oder jede Anwendung, die in der Baubranche verwendet wird, hat eine der beliebten RDBMS Datenbanken als Kern.

Jede der oben erwähnten Systemkategorien, die im Baugewerbe eingesetzt werden, arbeitet mit eigenen Datentypen, die der funktionalen Rolle dieser Systeme entsprechen. Um von der abstrakten Ebene zur konkreten Ebene zu gelangen, gehen wir von den Datentypen zu ihrer Darstellung als Formate und Dokumente über.

Zu der zuvor erstellten Liste von Systemen (Abb. 1.2-4) fügen wir nun die spezifischen Arten von Formaten und Dokumenten hinzu, mit denen sie häufig arbeiten:

- **Investor (CAPEX)**
  - Finanzdaten: Haushaltspläne, Ausgabenprognosen (strukturierte Daten).
  - Markttrenddaten: Marktanalysen (strukturierte und unstrukturierte Daten).
  - Rechtliche und vertragliche Daten: Verträge (Textdaten).

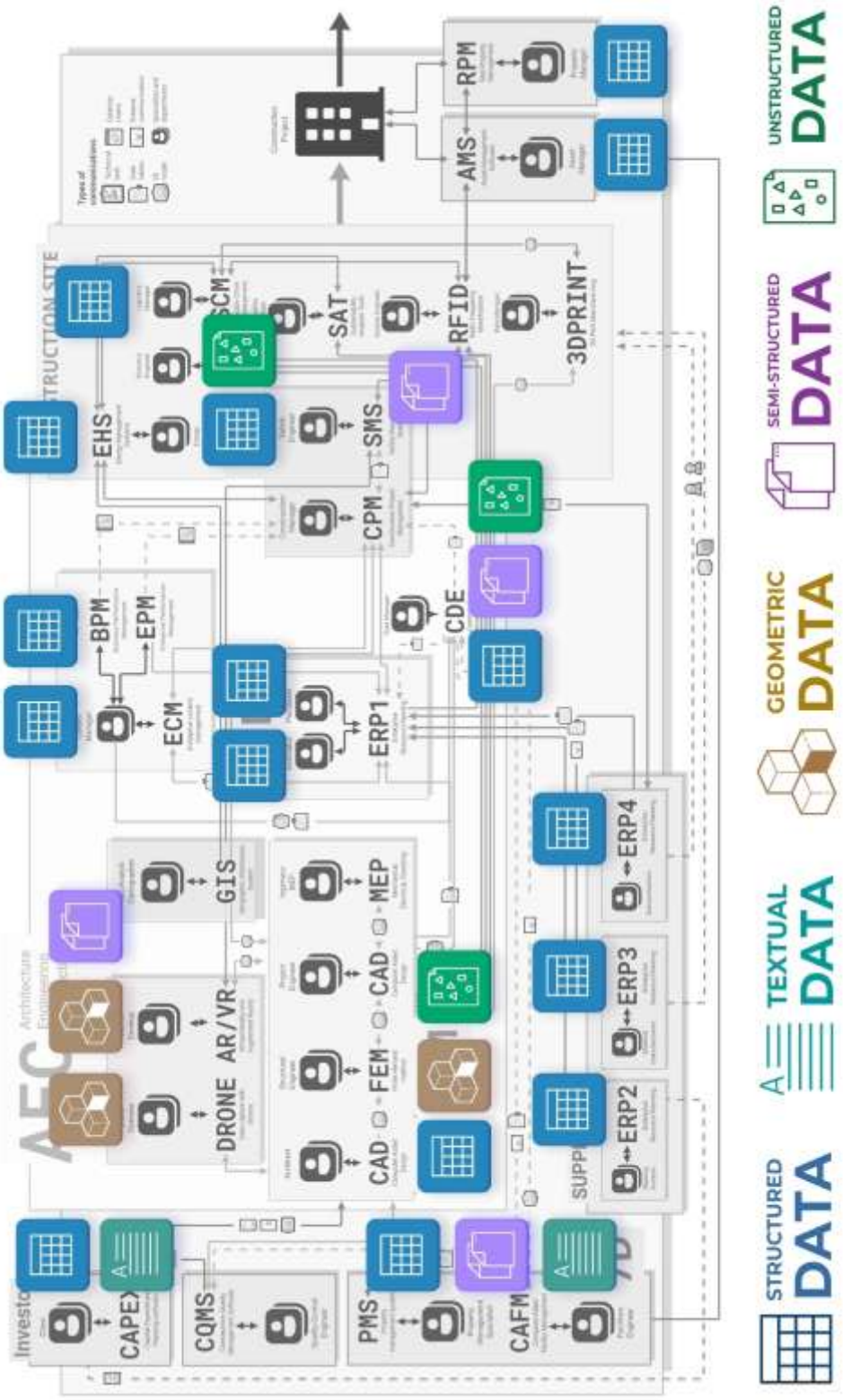


Abbildung 3.2-2 In der Baubranche werden viele Systeme mit unterschiedlichen Schnittstellen eingesetzt, die mit verschiedenen Arten von Daten arbeiten.

- **Managementsysteme (PMS, CAFM, CQMS)**
  - Projektdaten: Diagramme, Aufgaben (strukturierte Daten).
  - Daten zur Instandhaltung von Anlagen: Wartungspläne (Text und halbstrukturierte Daten).
  - Daten zur Qualitätskontrolle: Normen, Inspektionsberichte (textuelle und unstrukturierte Daten).
- **CAD, FEM und BIM**
  - Technische Zeichnungen: Architektur- und Strukturpläne (geometrische Daten, unstrukturierte Daten).
  - Gebäudemodelle: 3D -Modelle, Materialdaten (geometrische und semistrukturierte Daten).
  - Technische Berechnungen: Lastanalyse (strukturierte Daten).
- **Baustellenmanagementsysteme (EHS, SCM)**
  - Sicherheits- und Gesundheitsdaten: Sicherheitsprotokolle (textuelle und strukturierte Daten).
  - Daten der Lieferkette: Bestände, Bestellungen (strukturierte Daten).
  - Tagesberichte: Arbeitszeiten, Produktivität (strukturierte Daten).
- **Drohnen, AR/VR, GIS, 3D -drucken**
  - Geodaten: topografische Karten (geometrische und strukturierte Daten).
  - Daten in Echtzeit: Videos und Fotos (unstrukturierte Daten).
  - Modelle für den 3D -Druck: digitale Zeichnungen (geometrische Daten).
- **Zusätzliche Verwaltungssysteme (4D BPM, 5D ERP1)**
  - Zeit- und Kostendaten: Zeitpläne, Schätzungen (strukturierte Daten).
  - Änderungsmanagement: Aufzeichnungen über Projektänderungen (Text und strukturierte Daten).
  - Leistungsberichte: Erfolgsindikatoren (strukturierte Daten).
- **Datenintegration und -kommunikation (CDE, RFID, AMS, RPM)**
  - Datenaustausch: Austausch von Dokumenten, Datenmodelle (strukturierte und textuelle Daten).
  - RFID und Verfolgungsdaten: Logistik, Vermögensverwaltung (strukturierte Daten).
  - Überwachung und Kontrolle: Sensoren vor Ort (strukturierte und unstrukturierte Daten).

So arbeitet jedes System in der Bauindustrie - vom Baustellenmanagementsystem bis zur operativen Datenbank - mit einer eigenen Art von Informationen: strukturierte, textuelle, geometrische und andere. Die "Datenlandschaft", mit der die Fachleute täglich arbeiten müssen, ist äußerst vielfältig. Eine einfache Aufzählung von Formaten zeigt jedoch nicht die Komplexität der tatsächlichen Arbeit mit Informationen.

In der Praxis sehen sich die Unternehmen mit der Tatsache konfrontiert, dass Daten, selbst wenn sie aus Systemen abgerufen werden, nicht "so wie sie sind" verwendet werden können. Dies gilt insbesondere für Texte, Bilder, PDFs, CAD-Dateien und andere Formate, die mit Standardwerkzeugen nur schwer zu analysieren sind. Deshalb ist der nächste wichtige Schritt die Datentransformation - ein Prozess, ohne den Verarbeitung, Analyse, Visualisierung und Entscheidungsfindung nicht effektiv automatisiert werden können.

Datentransformation: die entscheidende Grundlage der modernen Unternehmensanalyse

Die meisten Unternehmen sehen sich heute mit einem Paradoxon konfrontiert: Etwa 80% ihrer täglichen Prozesse basieren immer noch auf klassisch strukturierten Daten - den bekannten Excel-Tabellen und rela-

tionalen Datenbanken (RDBMS) [66]. Gleichzeitig sind jedoch 80% der neuen Informationen, die in das digitale Ökosystem von Unternehmen gelangen, unstrukturiert oder lose strukturiert (Abb. 3.2-3) [52]. Dazu gehören Texte, Grafiken, Geometrie, Bilder, CAD -Modelle, Dokumentationen im PDF-Format, Audio- und Videoaufzeichnungen, elektronische Korrespondenz und vieles mehr.

Darüber hinaus wächst die Menge der unstrukturierten Daten weiterhin schnell - die jährliche Wachstumsrate wird auf 55-65% geschätzt [67]. Diese Dynamik führt zu ernsthaften Schwierigkeiten bei der Integration neuer Informationen in bestehende Geschäftsprozesse. Das Ignorieren dieses Stroms von Multiformat-Daten führt zur Bildung von Informationslücken und verringert die Verwaltbarkeit der gesamten digitalen Umgebung des Unternehmens.

Abb. 3.2-3 Die jährliche Zunahme unstrukturierter Daten führt zu Herausforderungen bei der Integration von Informationsströmen in Geschäftsprozesse.

> Das Ignorieren komplexer, unstrukturierter und unübersichtlicher, lose strukturierter Daten in Automatisierungsprozessen kann zu erheblichen Lücken in der Informationslandschaft eines Unternehmens führen. In der heutigen Welt unkontrollierbarer und lawinenartiger Informationsbewegungen müssen Unternehmen einen hybriden Ansatz für das Datenmanagement verfolgen, der effektive Methoden für den Umgang mit allen Arten von Daten umfasst.

Der Schlüssel zu einer effektiven Datenverwaltung liegt in der Organisation, Strukturierung und Klassifizierung verschiedenen Arten von Daten "Babel" (einschließlich unstrukturierter, textueller und geometrischer Formate, in strukturierte oder lose strukturierte Daten). Durch diesen Prozess werden chaotische Datensätze in geordnete Strukturen umgewandelt, die in Systeme integriert werden können und so eine Entscheidungsfindung auf der Grundlage dieser Daten ermöglichen (Abb. 3.2-4).

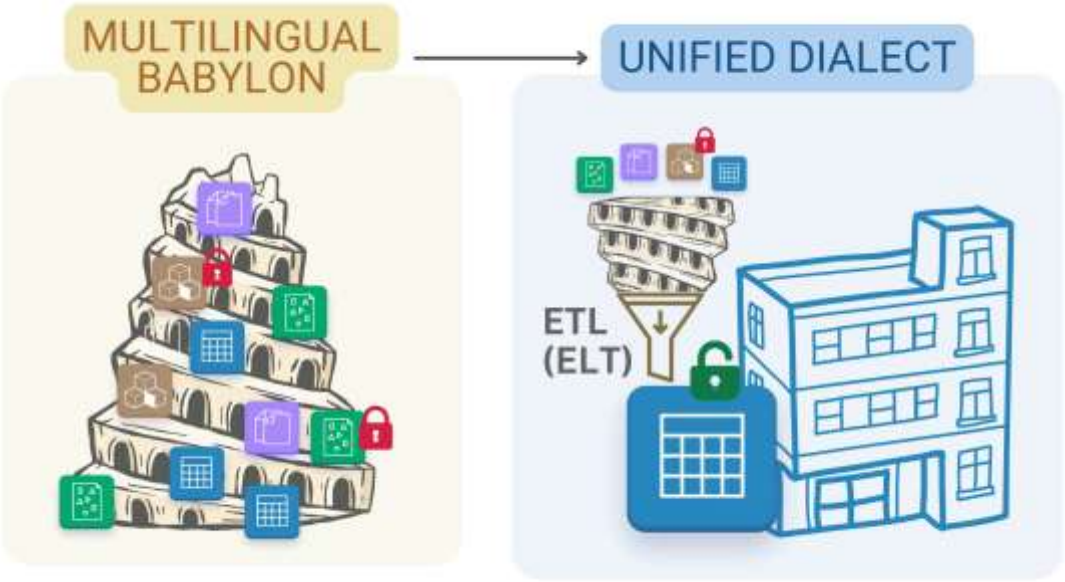


Abb. 3.2-4 Die Hauptaufgabe der Datenverwaltungsabteilungen besteht darin, das "Babylon" der vielfältigen und vielformatigen Daten in ein strukturiertes und kategorisiertes System zu überführen.

Eines der Haupthindernisse für eine solche Harmonisierung ist nach wie vor die geringe Interoperabilität zwischen den verschiedenen digitalen Plattformen - die "Silos", die wir in den vorangegangenen Kapiteln besprochen haben.

> Dem Bericht zufolge betont das National Institute of Standards and Technology (NIST, USA) [68], dass schlechte Datenkompatibilität zwischen verschiedenen Gebäudeplattformen zu Informationsverlusten und erheblichen Zusatzkosten führt. Allein im Jahr 2002 verursachten Probleme mit der Software-Interoperabilität im US-Investitionsbau Verluste in Höhe von 15,8 Milliarden Dollar pro Jahr, wobei zwei Drittel dieser Verluste von den Gebäudeeigentümern und -betreibern getragen werden, insbesondere während des Betriebs und der Wartung [68]. In der Studie wird auch festgestellt, dass die Standardisierung von Datenformaten diese Verluste verringern und die Effizienz während des gesamten Lebenszyklus von Gebäuden verbessern kann.

Laut der CrowdFlower-Studie von 2016 [69], für die 16 000 Datenwissenschaftler in aller Welt befragt wurden, bleiben "schmutzige" und vielgestaltige Daten das Hauptproblem. Laut dieser Studie sind die wertvollste Ressource nicht die endgültigen Datenbanken oder Modelle für maschinelles Lernen, sondern die Zeit, die für die Aufbereitung der Informationen aufgewendet wird.

Das Bereinigen, Formatieren und Organisieren nimmt bis zu 60% der Zeit eines Analysten und Datenmanagers in Anspruch. Fast ein Fünftel der Zeit wird für die Suche und Sammlung der richtigen Datensätze aufgewendet, die oft in Silos versteckt und für die Analyse unzugänglich sind. Und nur etwa 9 Prozent der Zeit wird direkt für Modellierung, Analyse, Vorhersagen und das Testen von Hypothesen verwendet. Der Rest entfällt auf Kommunikation, Visualisierung, Berichterstattung und die Suche nach unterstützenden Informationsquellen

Im Durchschnitt ist die Datenarbeit eines Managers wie folgt verteilt (Abb. 3.2-5):

- **Bereinigung und Organisation von Daten (60%):** Saubere und strukturierte Daten können die Arbeitszeit des Analysten erheblich reduzieren und die Erledigung von Aufgaben beschleunigen.
- **Datenerfassung (19%):** Eine große Herausforderung für Datenwissenschaftler ist das Auffinden relevanter Datensätze. Oft sind die Unternehmensdaten in chaotisch organisierten "Silos" gestapelt, was den Zugriff auf die benötigten Informationen erschwert.
- **Modellierung/Maschinelles Lernen (9%):** Wird oft durch unklare Unternehmensziele seitens der Kunden behindert. Das Fehlen eines klaren Leitbildes kann das Potenzial selbst des besten Modells zunichte machen.
- **Sonstige Aufgaben (5%):** Neben der Verarbeitung von Daten müssen Analysten recherchieren, Daten aus verschiedenen Blickwinkeln untersuchen, Ergebnisse durch Visualisierungen und Berichte kommunizieren und Empfehlungen zur Optimierung von Prozessen und Strategien geben.

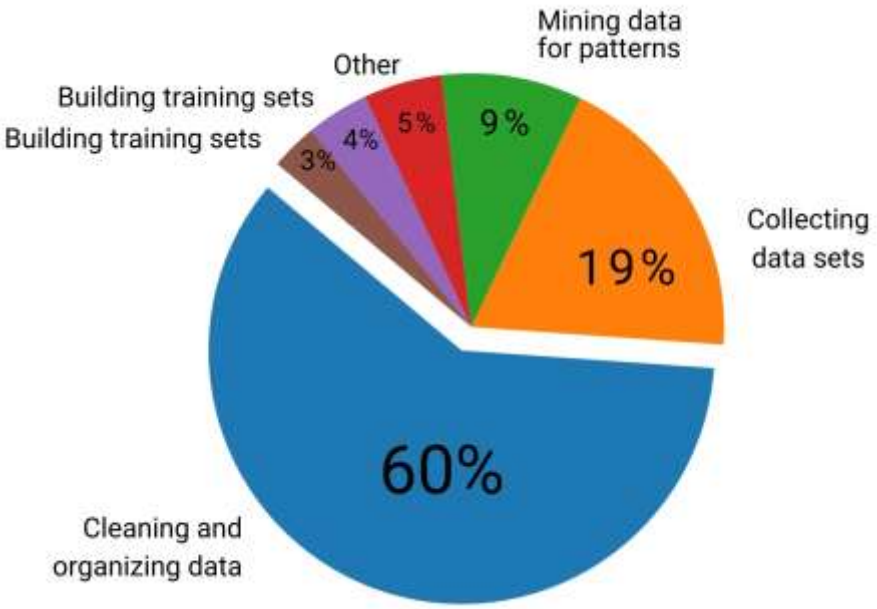


Abb. 3.2-5 Womit Datenmanager, die mit Daten arbeiten, die meiste Zeit verbringen (basierend auf [70]).

> Diese Schätzungen werden durch andere Studien gestützt. Laut der Xplenty-Studie, die 2015 im BizReport veröffentlicht wurde [71], verbringen Fachleute für Business Intelligence (BI) zwischen 50% und 90% ihrer Zeit (BI) mit der Vorbereitung von Daten für die Analyse.

Das Bereinigen, Validieren und Organisieren von Daten ist eine wichtige Grundlage für alle nachgelagerten Daten- und Analyseprozesse und nimmt bis zu 90% der Zeit von Datenwissenschaftlern in Anspruch.

Diese akribische Arbeit, die für den Endnutzer unsichtbar ist, ist von entscheidender Bedeutung. Fehler in den Rohdaten verzerren unweigerlich die Analysen, sind irreführend und können zu kostspieligen Managementfehlern führen. Deshalb werden Datenbereinigungs- und Standardisierungsprozesse - von der Beseitigung von Duplikaten und der Ergänzung von Lücken bis hin zur Harmonisierung von Maßeinheiten und der Anpassung an ein gemeinsames Modell - zu einem Eckpfeiler der modernen digitalen Strategie.

Die gründliche Umwandlung, Bereinigung und Standardisierung von Daten nimmt daher nicht nur den größten Teil der Zeit von Spezialisten in Anspruch (bis zu 80% der Arbeit mit Daten), sondern bestimmt auch die Möglichkeit ihrer effektiven Nutzung im Rahmen moderner Geschäftsprozesse. Mit der Datenorganisation und -bereinigung allein ist es jedoch nicht getan, um die Informationsflüsse in einem Unternehmen optimal zu steuern. In der Phase der Organisation und Strukturierung geht es um die Wahl eines geeigneten Datenmodells, das sich unmittelbar auf die Bequemlichkeit und Effizienz der Arbeit mit den Informationen in den nachfolgenden Verarbeitungsphasen auswirkt.

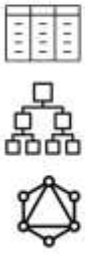
Da Daten und Unternehmensziele unterschiedlich sind, ist es wichtig, die Merkmale von Datenmodellen zu verstehen und die richtige Struktur auszuwählen oder zu erstellen. Je nach dem Grad der Strukturierung und der Art und Weise, wie die Beziehungen zwischen den Elementen beschrieben werden, gibt es drei Hauptmodelle: strukturierte, locker strukturierte und grafische Modelle. Jedes eignet sich für unterschiedliche Aufgaben und hat seine eigenen Stärken und Schwächen.

Datenmodelle: Beziehungen in den Daten und Beziehungen zwischen Elementen

Daten in Informationssystemen werden auf unterschiedliche Weise organisiert - je nach den Aufgaben und Anforderungen für die Speicherung, Verarbeitung und Übermittlung von Informationen. Der wesentliche Unterschied zwischen den Arten von Datenmodellen, der Form, in der Informationen gespeichert werden, ist der Grad der Strukturierung und die Art und Weise, in der die Beziehungen zwischen den Elementen beschrieben werden.

> Strukturierte Daten haben ein klares und wiederholbares Schema: Sie sind in Tabellen mit festen Spalten organisiert. Dieses Format bietet Vorhersehbarkeit, einfache Verarbeitung und Effizienz bei der Durchführung von SQL Abfragen, Filterung und Aggregation. Beispiele - Datenbanken (RDBMS), Excel, CSV.

Locker strukturierte Daten ermöglichen eine flexible Struktur: verschiedene Elemente können unterschiedliche Attribute enthalten und als Hierarchien gespeichert werden. Beispiele sind JSON, XML oder andere Dokumentenformate. Diese Daten sind praktisch, wenn es darum geht, verschachtelte Objekte und Beziehungen zwischen ihnen zu modellieren, aber andererseits erschweren sie die Datenanalyse und die Standardisierung (Abb. 3.2-6).



Data Model	Storage Format	Example
Relational	CSV, SQL	A table of doors in Excel
Hierarchical	JSON, XML	Nested door objects inside a room
Graph-based	RDF, GraphDB	Relationships between building elements

Abbildung 3.2-6 Ein Datenmodell ist eine logische Struktur, die beschreibt, wie Daten in einem System organisiert, gespeichert und verarbeitet werden.

Die Wahl des geeigneten Formats hängt von den Zielen ab:

- Wenn die Geschwindigkeit des Filterns und der Analyse wichtig ist - relationale Tabellen (SQL, CSV, RDBMS, spaltenförmige Datenbanken) reichen aus.
- Wenn Flexibilität in der Struktur erforderlich ist, ist es besser, JSON oder XML zu verwenden.
- Wenn die Daten komplexe Beziehungen aufweisen, bieten Graphdatenbanken Transparenz und Skalierbarkeit.

In klassischen relationalen Datenbanken (RDBMS) wird jede Entität (z. B. eine Tür) durch eine Zeile und ihre Eigenschaften durch Tabellenspalten dargestellt. Eine Tabelle mit Elementen der Kategorie "Türen" kann zum Beispiel die Felder ID, Höhe, Breite, Feuerbeständigkeit und Raum-ID enthalten, die den Raum angeben (Abb. 3.2-7).

In klassischen relationalen Datenbanken (RDBMS) werden Beziehungen in Form von Tabellen gebildet, wobei jeder Datensatz ein Objekt darstellt und die Spalten dessen Parameter repräsentieren. Im Tabellenformat sehen die Daten über Türen im Projekt wie folgt aus, wobei jede Zeile ein separates Element darstellt - eine Tür mit ihrer eindeutigen Kennung und ihren Attributen, und die Verbindung mit dem Raum erfolgt über den Parameter "Raum-ID".

Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes

Abb. 3.2-7 Informationen zu den drei Elementen der Kategorie "Türen" des Projekts in tabellarisch strukturierter Form.

In locker strukturierten Formaten wie JSON oder XML werden Daten in einer hierarchischen oder verschachtelten Form gespeichert, wobei Elemente andere Objekte enthalten können und ihre Struktur variieren kann. Dadurch können komplexe Beziehungen zwischen Elementen modelliert werden. Ähnliche Informationen über Türen im Projekt, die in strukturierter Form erfasst wurden (Abb. 3.2-7), werden in einem locker strukturierten Format (JSON) so dargestellt (Abb. 3.2-8), dass sie zu verschachtelten Objekten innerhalb von Räumen werden (Räume - ID), was die Hierarchie logisch widerspiegelt.

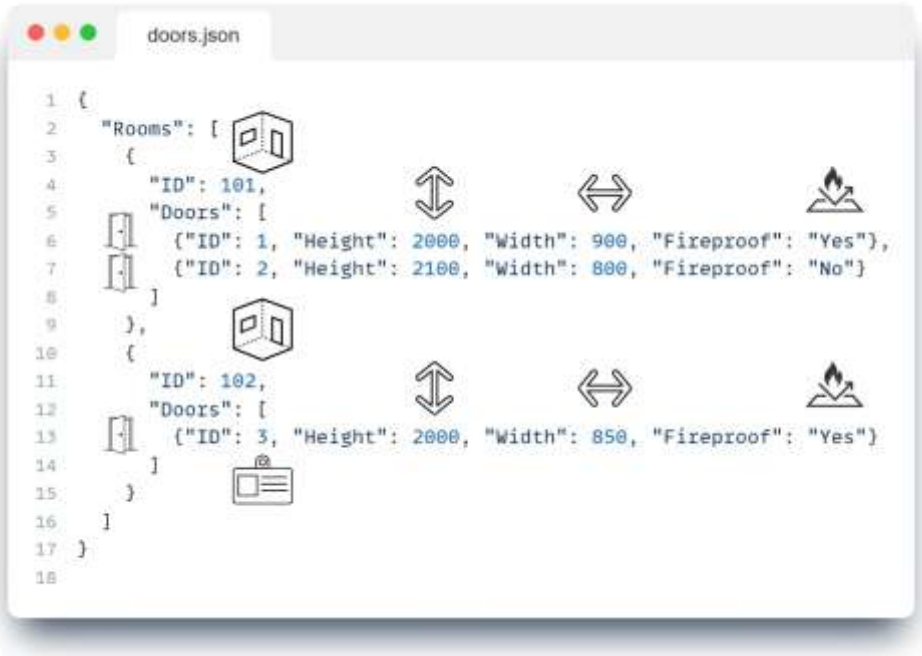


Abb. 3.2-8 Informationen über die Elemente der Kategorie "Türen" des Projekts im JSON-Format.

In einem Graphenmodell werden die Daten als Knoten (Vertices) und Verbindungen (Edges) zwischen ihnen dargestellt. So können Sie die komplexen Beziehungen zwischen Objekten und ihren Attributen visualisieren. Im Fall der Tür- und Raumdaten im Projekt sieht die grafische Darstellung von wie folgt aus:

- **Knoten (Nodes)** repräsentieren die wichtigsten Einheiten: Räume (Room 101, Room 102) und Türen (ID1001, ID1002, ID1003)
- **Rippen (Links)** zeigen die Beziehungen zwischen diesen Einheiten, z. B. die Zugehörigkeit einer Tür zu einem bestimmten Raum
- **Attribute** werden auf Knoten abgebildet und enthalten die Eigenschaften des Objekts (Höhe, Breite, Feuerbeständigkeit bei Türen)

Abb. 3.2-9 Projekttür-Informationen in der grafischen Ansicht.

Im Graphdatenmodell der Türbeschreibung sind jeder Raum und jede Tür separate Knoten. Die Türen sind mit den Räumen durch Kanten verbunden, die anzeigen, dass die Tür zu einem bestimmten Raum gehört. Die Attribute der Türen (Höhe, Breite, Feuerwiderstand) werden als Eigenschaften der entsprechenden Knoten gespeichert. Weitere Einzelheiten zu den Graphenformaten und zur Entstehung der Graphen-Semantik in der Baubranche werden im Kapitel "Entstehung von Semantik und Ontologie im Bauwesen" erörtert.

Graphdatenbanken sind dann sinnvoll, wenn nicht so sehr die Daten selbst wichtig sind, sondern die Beziehungen zwischen ihnen, wie z. B. in Empfehlungssystemen, Routingsystemen oder bei der Modellierung komplexer Beziehungen in Facility-Management-Projekten. Das Graphenformat vereinfacht die Erstellung neuer Beziehungen, da neue Datentypen zum Graphen hinzugefügt werden können, ohne die Speicherstruktur zu ändern. Im Vergleich zu relationalen Tabellen und strukturierten Formaten gibt es in einem Graphen jedoch keine zusätzliche Datenkonnektivität - die Übertragung von zweidimensionalen Datenbankdaten in einen Graphen erhöht nicht die Anzahl der Beziehungen und ermöglicht nicht den Erhalt neuer Informationen.

> Die Form und das Schema der Daten sollten auf den jeweiligen Anwendungsfall und die zu erfüllenden Aufgaben zugeschnitten sein. Um effektiv in Geschäftsprozessen zu arbeiten, ist es wichtig, jene Werkzeuge und Datenmodelle zu verwenden, die helfen, so schnell und einfach wie möglich Ergebnisse zu erhalten.

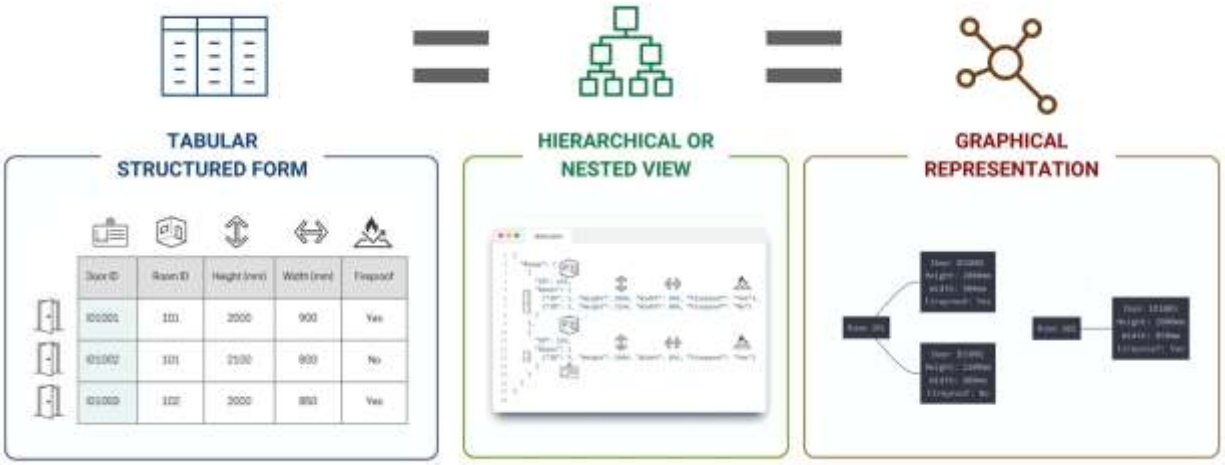


Abbildung 3.2-10 Dieselben Informationen über Projektelemente können in verschiedenen Formaten und mit unterschiedlichen Datenmodellen gespeichert werden.

> Die meisten großen Unternehmen stehen heute vor dem Problem einer übermäßigen Datenkomplexität. Jede von Hunderten oder Tausenden von Anwendungen verwendet ihr eigenes Datenmodell, was zu einer übermäßigen Komplexität führt - ein einzelnes Modell ist oft Dutzende Male komplexer als nötig, und die Gesamtheit aller Modelle ist Tausende Male komplexer. Diese übermäßige Komplexität behindert die Arbeit sowohl der Entwickler als auch der Endbenutzer erheblich.

Diese Komplexität schränkt die Entwicklung und Wartung der Systeme des Unternehmens erheblich ein. Jedes neue Element des Modells erfordert zusätzlichen Code, die Implementierung einer neuen Logik, gründliche Tests und die Anpassung an bestehende Lösungen. All dies erhöht die Kosten und verlangsamt die Arbeit des Automatisierungsteams im Unternehmen, wodurch selbst einfache Aufgaben zu kostspieligen und zeitaufwändigen Prozessen werden.

Die Komplexität betrifft alle Ebenen der Datenarchitektur. In relationalen Datenbanken äußert sie sich in der wachsenden Zahl von Tabellen und Spalten, die oft redundant sind. In objektorientierten Systemen wird die Komplexität durch die Vielzahl von Klassen und miteinander verknüpften Eigenschaften erhöht. In Formaten wie XML oder JSON äußert sich die Komplexität durch verwirrende verschachtelte Strukturen, eindeutige Schlüssel und inkonsistente Schemata.

> Die übermäßige Komplexität der Datenmodelle macht die Systeme nicht nur weniger effizient, sondern auch schwer verständlich für die Endnutzer und in Zukunft große Sprachmodelle und LLM-Agenten. Es ist das Problem des Verständnisses und der Komplexität von Datenmodellen und der Datenverarbeitung, das die Frage aufwirft, wie man Daten so einfach nutzbar machen kann, dass sie tatsächlich schnell nützlich werden.

Selbst wenn Datenmodelle klug ausgewählt werden, wird ihr Nutzen drastisch reduziert, wenn der Zugang zu den Daten eingeschränkt ist. Proprietäre Formate und geschlossene Plattformen behindern die Integration, erschweren die Automatisierung und entziehen die Kontrolle über geschützte Informationen. So entsteht nicht nur ein Silo mit neuen Daten, sondern ein verschlossenes Silo, auf das nur mit Genehmigung des Anbieters zugegriffen werden kann. Um das Ausmaß des Problems zu verstehen, ist es wichtig, genau zu betrachten, wie geschlossene Systeme digitale Prozesse im Bauwesen beeinflussen.

Proprietäre Formate und ihre Auswirkungen auf digitale Prozesse

Eine der größten Herausforderungen, mit denen Bauunternehmen bei der Digitalisierung konfrontiert sind, ist der begrenzte Zugang zu Daten. Dies erschwert die Integration von Systemen, mindert die Qualität der Informationen und erschwert die Organisation effizienter Prozesse. Die Verwendung proprietärer Formate und geschlossener Softwarelösungen ist häufig die Ursache für diese Schwierigkeiten.

Leider erlauben viele Programme, die in der Baubranche eingesetzt werden, bisher nur die Speicherung von Daten in proprietären Formaten oder in einem Cloud-Speicher, auf den nur über streng begrenzte Schnittstellen zugegriffen werden kann. Und nicht selten werden diese Lösungen in Abhängigkeit von noch geschlosseneren Systemen größerer Anbieter gebaut. Das hat zur Folge, dass selbst die Entwickler, die gerne offenere Architekturen anbieten würden, gezwungen sind, sich an die von den großen Anbietern diktierten Regeln zu halten.

> Während moderne Baudatenmanagementsysteme zunehmend offene Formate und Standards unterstützen (Abb. 3.1-5), bleiben CAD- (BIM)-basierte Datenbanken und damit verbundene ERP- und CAFM-Systeme isolierte proprietäre "Inseln" in der digitalen Landschaft der Branche (Abb. 3.2-11).

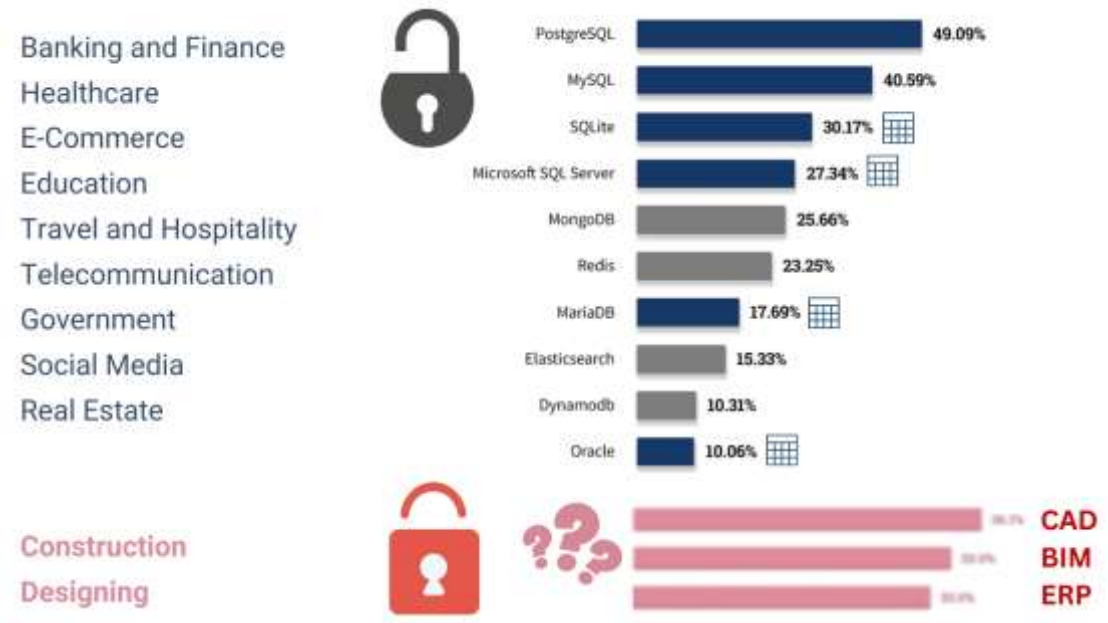


Abbildung 3.2-11 Der geschlossene und proprietäre Charakter von Daten schafft Hindernisse für die Datenintegration und den Datenzugang.

Geschlossene und monopolisierte Formate und Protokolle sind nicht nur ein Problem für die Bauindustrie. In vielen Wirtschaftszweigen begann der Kampf gegen geschlossene Standards und den begrenzten Zugang zu Daten mit einer Verlangsamung der Innovation (Abb. 3.2-12), dem Bestehen künstlicher Eintrittsbarrieren für neue Akteure und einer zunehmenden Abhängigkeit von großen Anbietern. Angesichts der raschen Zunahme der Bedeutung von Daten haben die Wettbewerbsbehörden einfach keine Zeit, auf die Herausforderungen der neuen digitalen Märkte zu reagieren, so dass geschlossene Formate und ein eingeschränkter Zugang zu Daten im Grunde zu digitalen "Grenzen" werden, die den Informationsfluss und das Wachstum behindern [63].

> *Wenn Maschinen alles produzieren, was wir brauchen, dann wird unsere Situation davon abhängen, wie diese Güter verteilt werden. Nur wenn der von Maschinen produzierte Reichtum verteilt wird, können alle ein Leben in Wohlstand genießen. Oder die meisten Menschen werden in bitterer Armut leben, wenn die Autobesitzer erfolgreich gegen die Umverteilung des Reichtums lobbyieren können. Bislang scheinen die Dinge in die zweite Richtung zu gehen, wobei die Technologie zu immer größerer Ungleichheit führt [72].*
>
> - Stephen Hawking, Astrophysiker, 2015

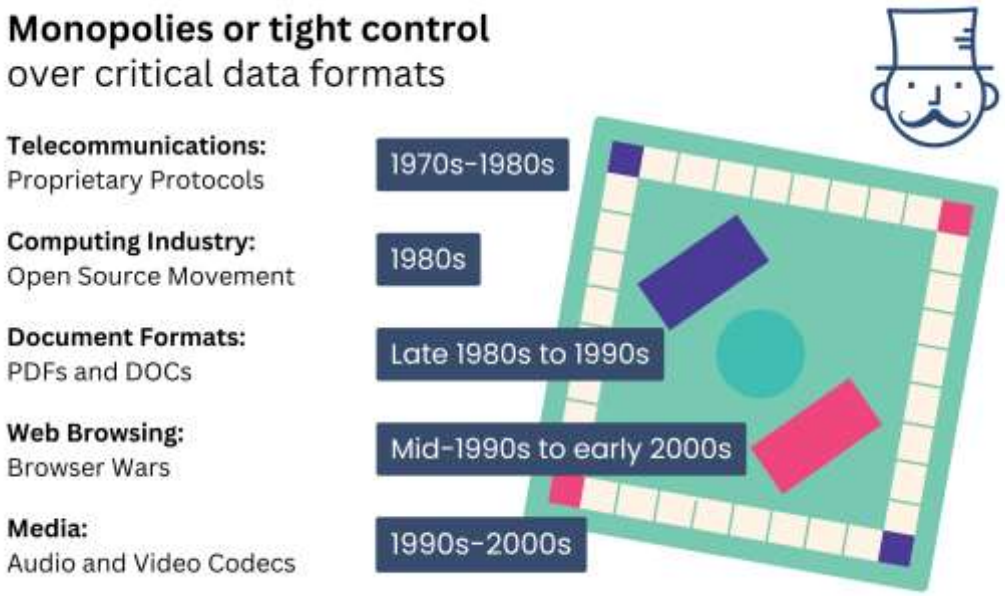


Abb. 3.2-12 Das Monopol auf die wichtigsten Datenformate und -protokolle ist kein ausschließliches Problem der Bauindustrie.

Infolgedessen sind Datenmanager, Datenanalysten, IT-Spezialisten und Entwickler, die Anwendungen für den Datenzugriff, die Datenverarbeitung und die Datenautomatisierung in der Bauindustrie erstellen, aufgrund des geschlossenen Zugangs zu Datenbanken -Programmen heute mit zahlreichen Abhängigkeiten von Softwareanbietern konfrontiert (Abb. 3.2-13). Diese Abhängigkeiten in Form von zusätzlichen Zugriffsebenen erfordern die Schaffung von Lösungen mit speziellen API -Anbindungen und speziellen Tools und Software.

> Eine API (Application Programming Interface) ist eine formalisierte Schnittstelle, über die ein Programm mit einem anderen interagieren und Daten und Funktionen austauschen kann, ohne auf den Quellcode zugreifen zu müssen. Eine API beschreibt, welche Anfragen ein externes System stellen kann, welches Format sie haben sollten und welche Antworten es erhalten wird. Sie ist ein standardisierter "Vertrag" zwischen Softwaremodulen.

Die große Anzahl von Abhängigkeiten von geschlossenen Lösungen führt dazu, dass die gesamte Code-Architektur und Geschäftsprozesslogik in einem Unternehmen zu einer "Spaghetti-Architektur" von Tools wird, die von der Politik des Softwareanbieters abhängt, der einen qualitativ hochwertigen Zugang zu den Daten bietet.

Die Abhängigkeit von geschlossenen Lösungen und Plattformen führt nicht nur zu einem Verlust an Flexibilität, sondern auch zu echten Geschäftsrisiken. Die Änderung von Lizenzbedingungen, die Sperrung des Zugangs zu Daten, die Änderung von Formaten oder der API-Struktur - all dies kann kritische Prozesse blockieren. Plötzlich stellt sich heraus, dass die Aktualisierung einer Tabelle die Überarbeitung eines ganzen Blocks von Integrationen und Konnektoren erfordert (Abb. 3.2-13), und jede groß angelegte Aktualisierung der Software oder ihres API-Anbieters wird zu einer potenziellen Bedrohung für die Stabilität des gesamten Unternehmenssystems.

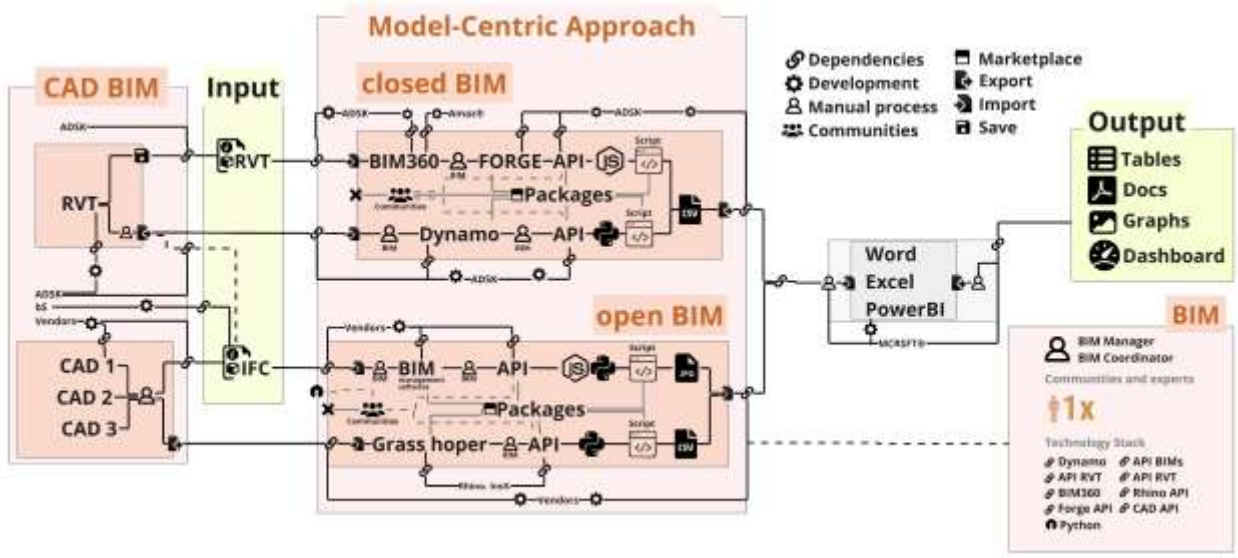


Abb. 3.2-13 Ein Beispiel für die große Anzahl von Abhängigkeiten in der CAD-Bearbeitung - Daten schaffen Barrieren für die Datenintegration im Ökosystem von Bauunternehmen.

> Entwickler und Systemarchitekten sind unter solchen Bedingungen gezwungen, nicht auf Vorfreude, sondern auf Überleben zu arbeiten. Anstatt neue Lösungen zu implementieren, passen sie sich an. Anstatt zu entwickeln, versuchen sie, die Kompatibilität aufrechtzuerhalten. Anstatt Prozesse zu automatisieren und zu beschleunigen, verbringen sie ihre Zeit mit dem Studium der nächsten geschlossenen Schnittstellen, der API-Dokumentation und der endlosen Wiederherstellung von Code.

Die Arbeit mit geschlossenen Formaten und Systemen ist nicht nur eine technische Herausforderung - sie ist ein strategischer Zwang. Trotz der offensichtlichen Möglichkeiten, die moderne Automatisierung, KI, LLM und prädiktive Analytik bieten, schöpfen viele Unternehmen ihr Potenzial nicht voll aus. Und die durch proprietäre Formate errichteten Barrieren (Abb. 3.2-13) verwehren den Unternehmen den Zugang zu ihren eigenen Daten. Dies ist vielleicht die Ironie der digitalen Transformation im Bauwesen.

> Datentransparenz und offene Systeme sind kein Luxus, sondern eine Voraussetzung für Schnelligkeit und Effizienz. Ohne Offenheit sind Geschäftsprozesse mit unnötiger Bürokratie, vielschichtigen Genehmigungsketten und einer zunehmenden Abhängigkeit vom HiPPO-Prinzip - Entscheidungen auf der Grundlage der Meinung der bestbezahlten Person zu treffen - gefüllt.

Dennoch zeichnet sich am Horizont ein Paradigmenwechsel ab. Trotz der Dominanz proprietärer Lösungen erkennen immer mehr Unternehmen die Grenzen der von der Vierten Industriellen Revolution inspirierten Architekturen. Heute verschiebt sich der Vektor hin zu den Grundsätzen der Fünften Revolution, bei der Daten als strategisches Gut, offene Schnittstellen (APIs) und echte Interoperabilität zwischen Systemen im Mittelpunkt stehen.

Dieser Übergang markiert eine Verlagerung weg von geschlossenen Ökosystemen hin zu flexiblen, modularen digitalen Architekturen, in denen offene Formate, Normen und ein transparenter Datenaustausch eine Schlüsselrolle spielen.

Offene Formate verändern den Ansatz der Digitalisierung

Die Baubranche war eine der letzten, die sich mit dem Problem geschlossener und geschützter Daten auseinandergesetzt hat. Im Gegensatz zu anderen Wirtschaftszweigen hat sich die Digitalisierung hier nur langsam entwickelt. Die Gründe dafür sind unter anderem der traditionell konservative Charakter der Branche, das Vorherrschen unterschiedlicher lokaler Lösungen und die tief verwurzelte papierbasierte Dokumentenverwaltung. Jahrzehntelang beruhten die wichtigsten Bauprozesse auf physischen Zeichnungen, Telefonaten und nicht synchronisierten Datenbanken. In diesem Zusammenhang wurden geschlossene Formate lange Zeit eher als Norm denn als Hindernis wahrgenommen.

Erfahrungen aus anderen Branchen zeigen, dass die Beseitigung von Hindernissen für geschlossene Daten zu einem Innovationsschub, einer beschleunigten Entwicklung und einem verstärkten Wettbewerb führt [73]. In der Wissenschaft ermöglicht der Austausch offener Daten, Entdeckungen zu beschleunigen und die internationale Zusammenarbeit zu entwickeln. In der Medizin kann er die Effizienz von Diagnose und Behandlung verbessern. In der Softwareentwicklung - zur Schaffung von Ökosystemen der gemeinsamen Entwicklung und schnellen Produktverbesserung.

> Laut dem McKinsey-Bericht "Open Data: Unlock Innovation and Productivity with Information Flow" 2013. [74] haben offene Daten das Potenzial, jährlich 3 bis 5 Billionen Dollar in sieben Schlüsselindustrien freizusetzen, darunter Bauwesen, Verkehr, Gesundheitswesen und Energie. Derselben Studie zufolge ermöglichen dezentrale Datenökosysteme großen Bauunternehmen und Auftragnehmern, die Kosten für Softwareentwicklung und -pflege zu senken und so die Einführung der Digitalisierung zu beschleunigen.

Der Übergang zu offenen Architekturen, der in anderen Wirtschaftszweigen längst begonnen hat, breitet sich allmählich auf das Baugewerbe aus. Große Unternehmen und öffentliche Auftraggeber und insbesondere die Finanzorganisationen, die Investitionen in Bauprojekte kontrollieren, fordern zunehmend die Verwendung offener Daten und den Zugang zum Quellcode von Berechnungen, Kalkulationen und Anwendungen. Von den Entwicklern wird nicht mehr nur erwartet, dass sie digitale Lösungen erstellen und die endgültigen Zahlen eines Projekts vorlegen - es wird erwartet, dass sie transparent, reproduzierbar und unabhängig von Drittanbietern sind.

> Die Verwendung von Open-Source-Lösungen gibt dem Kunden die Gewissheit, dass die Weiterentwicklung von Tools und Systemen nicht beeinträchtigt wird, selbst wenn externe Entwickler die Zusammenarbeit einstellen oder das Projekt verlassen. Einer der Hauptvorteile offener Daten besteht darin, dass sie die Abhängigkeit der Anwendungsentwickler von bestimmten Plattformen für den Datenzugriff beseitigen.

Wenn ein Unternehmen nicht vollständig auf proprietäre Lösungen verzichten kann, ist ein möglicher Kompromiss der Einsatz von Reverse-Engineering-Techniken. Diese rechtlich und technisch soliden Methoden ermöglichen es, geschlossene Formate in besser zugängliche, strukturierte und für die Integration geeignete Formate umzuwandeln. Dies ist besonders wichtig, wenn eine Verbindung zu Altsystemen hergestellt

oder Informationen von einer Softwarelandschaft in eine andere migriert werden sollen.

> Eines der besten Beispiele in der Geschichte des Übergangs zu offenen Formaten und der Anwendung von Reverse Engineering (legales Hacken von proprietären Systemen) im Bauwesen ist der Kampf um die Öffnung des DWG-Formats, das in computergestützten Konstruktionssystemen (CAD) weit verbreitet ist. Als Reaktion auf das Monopol eines Softwareanbieters schlossen sich 1998 die anderen 15 CAD-Anbieter zu einer neuen Allianz mit dem Namen "Open DWG" zusammen, um Entwicklern kostenlose und unabhängige Werkzeuge für die Arbeit mit dem DWG-Format (dem De-facto-Standard für die Übertragung von Zeichnungen) zur Verfügung zu stellen, ohne dass proprietäre Software oder geschlossene APIs erforderlich sind. Dieses Ereignis war ein Wendepunkt, der es Zehntausenden von Unternehmen ermöglichte, freien Zugang zu dem geschlossenen Format einer beliebten CAD-Lösung aus den späten 1980er Jahren bis heute zu erhalten und kompatible Lösungen zu schaffen, die den Wettbewerb auf dem CAD-Markt förderten [75]. Heute wird das "Open DWG"-SDK, das erstmals 1996 erstellt wurde, in fast allen Lösungen verwendet, in denen es möglich ist, das DWG-Format zu importieren, zu bearbeiten und zu exportieren, und zwar außerhalb der offiziellen Anwendung des DWG-Formatentwicklers.

Andere Tech-Giganten erzwingen ähnliche Umstellungen. Microsoft, einst ein Symbol für einen proprietären Ansatz, öffnete den Quellcode des.NET Framework, begann mit der Verwendung von Linux in der Azure-Cloud-Service-Infrastruktur und übernahm GitHub, um seine Position in der Open-Source-Gemeinschaft zu stärken. [76]. Meta (ehemals Facebook) hat quelloffene KI-Modelle , wie z. B. die Llama-Serie, veröffentlicht, um Innovation und Zusammenarbeit bei der Entwicklung von KI-Agenten zu fördern. CEO Mark Zuckerberg geht davon aus, dass Open-Source-Plattformen in den nächsten zehn Jahren den technologischen Fortschritt anführen werden [77].

> Open Source ist ein Softwareentwicklungs- und -verteilungsmodell, bei dem der Quellcode zur freien Nutzung, Untersuchung, Änderung und Verteilung offen ist.

Offene Daten und Open-Source-Lösungen sind nicht nur ein Trend, sondern die Grundlage der digitalen Nachhaltigkeit. Sie geben Unternehmen Flexibilität, Belastbarkeit, Kontrolle über ihre eigenen Entscheidungen und die Möglichkeit, digitale Prozesse zu skalieren, ohne von Herstellerrichtlinien abhängig zu sein. Und, was ebenso wichtig ist, sie geben Unternehmen die Kontrolle über die wertvollste Ressource des 21. Jahrhunderts zurück - ihre Daten.

Paradigmenwechsel: Open Source als Ende der Ära der Vorherrschaft der Softwarehersteller

Die Baubranche befindet sich in einem Wandel, der sich nicht auf die übliche Weise monetarisieren lässt. Das Konzept des datengesteuerten, datenzentrierten Ansatzes und der Einsatz von Open Source Tools führt zu einem Überdenken der Spielregeln, auf denen die Software-Giganten des Marktes stehen.

> Im Gegensatz zu früheren technologischen Umwälzungen wird dieser Übergang von den Anbietern nicht aktiv gefördert. Der Paradigmenwechsel bedroht ihre traditionellen Geschäftsmodelle, die auf Lizenzierung, Abonnements und Beratung basieren. Die neue Realität beinhaltet kein sofort einsatzbereites Produkt oder ein kostenpflichtiges Abonnement - sie erfordert eine Neuausrichtung der Prozesse und des Denkens.

Um Rechenzentrumslösungen auf der Grundlage offener Technologien zu verwalten und zu entwickeln, müssen die Unternehmen ihre internen Prozesse überdenken. Fachleute aus verschiedenen Abteilungen müssen nicht nur zusammenarbeiten, sondern auch die Art und Weise ihrer Zusammenarbeit neu überdenken.

Das neue Paradigma impliziert die Nutzung offener Daten und Open-Source-Lösungen, bei denen Werkzeuge, die auf künstlicher Intelligenz und großen Sprachmodellen (LLM) basieren, und nicht Programmierer eine besondere Rolle bei der Erstellung von Softwarecode spielen werden. Bereits Mitte 2024 werden mehr als 25 Prozent des neuen Codes bei Google mit KI erstellt [78]. In Zukunft wird die Codierung mit LLMs 80% der Arbeit in nur 20% der Zeit erledigen (Abb. 3.2-14).

> Laut der McKinsey-Studie 2020 [79] ersetzen GPUs aufgrund ihrer hohen Leistung und der Unterstützung durch moderne Open-Source-Tools zunehmend CPUs in der Analytik. Dies ermöglicht es Unternehmen, die Datenverarbeitung zu beschleunigen, ohne signifikante Investitionen in teure Software zu tätigen oder knappe Spezialisten einzustellen.

Führende Beratungsunternehmen wie McKinsey, PwC und Deloitte betonen die wachsende Bedeutung von offenen Standards, Open Source Anwendungen in allen Branchen.

> Laut dem PwC-Bericht Open Source Monitor 2019 [80] setzen 69% der Unternehmen mit 100 oder mehr Mitarbeitern bewusst Open-Source-Lösungen ein. Besonders aktiv wird OSS in großen Unternehmen genutzt: 71% der Unternehmen mit 200-499 Mitarbeitern, 78% in der Kategorie 500-1999 Mitarbeiter und bis zu 86% bei Unternehmen mit mehr als 2000 Mitarbeitern. Laut dem OSSRA 2023-Bericht von Synopsys enthielten 96% der analysierten Codebases Open-Source-Komponenten [81].

Die Rolle des Entwicklers wird in Zukunft nicht darin bestehen, manuell Code zu schreiben, sondern Datenmodelle und Flussarchitekturen zu entwerfen und KI-Agenten zu verwalten, die bei Bedarf die richtigen Berechnungen durchführen. Die Benutzeroberflächen werden minimalistisch und die Interaktion wird dialogbasiert. Die klassische Programmierung wird einem High-Level-Design und der Orchestrierung digitaler Lösungen weichen (Abb. 3.2-14). Aktuelle Trends - wie Low-Code-Plattformen (Abb. 7.4-6) und LLM-fähige Ökosysteme (Abb. 7.4-4) - werden die Kosten für die Entwicklung und Wartung von IT-Systemen erheblich senken.

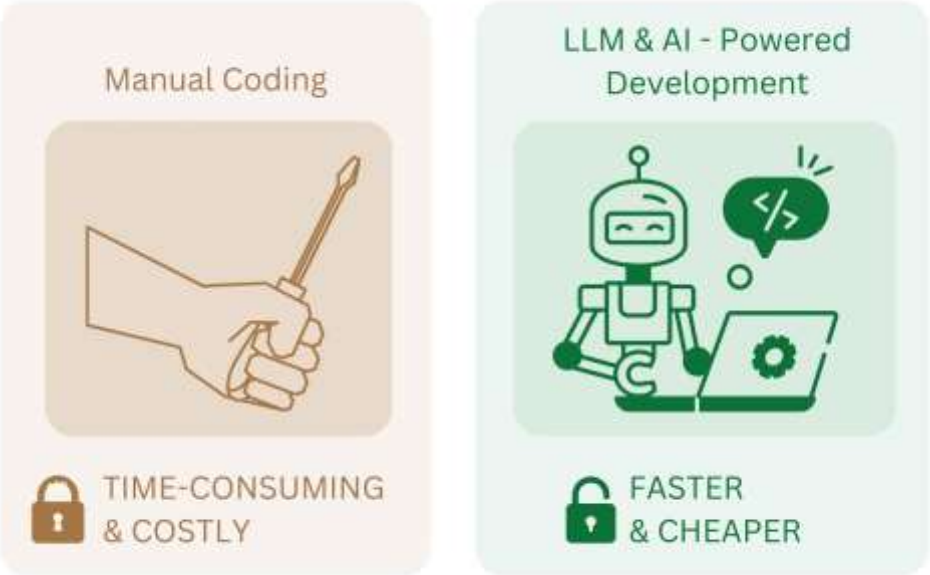


Abb. 3.2-14 Während heute Anwendungen manuell von Programmierern erstellt werden, wird in Zukunft ein erheblicher Teil des Codes von KI- und LLM-basierten Lösungen generiert werden.

Dieser Übergang wird anders sein als frühere, und es ist unwahrscheinlich, dass die großen Softwareanbieter ihn beschleunigen werden.

> In der Studie der Harvard Business School "The Value of Open Source Software" 2024 [40] wird der Gesamtwert von Open-Source-Software unter zwei Gesichtspunkten geschätzt. Wenn wir einerseits berechnen, wie viel es kosten würde, alle bestehenden Open-Source-Lösungen von Grund auf neu zu entwickeln, würde sich der Betrag auf etwa 4,15 Milliarden Dollar belaufen. Wenn wir uns andererseits vorstellen, dass jedes Unternehmen seine eigenen Analogien zu Open-Source-Lösungen entwickelt (was überall geschieht), ohne Zugang zu bestehenden Tools zu haben, dann würden sich die Gesamtkosten für Unternehmen auf kolossale 8,8 Billionen Dollar belaufen - das sind die Kosten der Nachfrage.

> Es ist nicht schwer zu erraten, dass kein großer Softwareanbieter daran interessiert ist, einen Softwaremarkt mit einem potenziellen Wert von 8,8 Billionen Dollar auf nur 4,15 Milliarden Dollar zu schrumpfen. Dies würde bedeuten, dass sich das Nachfragevolumen um mehr als das 2.000-fache verringern würde. Eine solche Umstellung ist für Anbieter, deren Geschäftsmodelle auf der jahrelangen Abhängigkeit der Kunden von geschlossenen Lösungen beruhen, schlicht unrentabel. Unternehmen, die erwarten, dass ihnen jemand eine bequeme und offene schlüsselfertige Lösung anbietet, könnten also enttäuscht werden - diese Anbieter werden einfach nicht auftauchen.

Der Übergang zu einer offenen digitalen Architektur bedeutet nicht, dass Arbeitsplätze oder Einnahmen verloren gehen. Im Gegenteil, er schafft die Voraussetzungen für flexible und anpassungsfähige Geschäftsmodelle, die den traditionellen Lizenz- und Boxsoftwaremarkt schließlich verdrängen können.

Statt des Verkaufs von Lizenzen - Dienstleistungen, statt geschlossener Formate - offene Plattformen,

statt der Abhängigkeit von einem Anbieter - Unabhängigkeit und die Möglichkeit, Lösungen für echte Bedürfnisse zu entwickeln. Diejenigen, die bisher einfach nur Werkzeuge benutzten, werden zu deren Mitautoren. Und diejenigen, die mit Daten, Modellen, Szenarien und Logik arbeiten können, werden sich im Zentrum der neuen digitalen Wirtschaft der Branche wiederfinden. Im abschließenden, zehnten Teil des Buches werden wir mehr über diese Veränderungen und die neuen Rollen, Geschäftsmodelle und Kooperationsformate sprechen, die sich rund um offene Daten entwickeln.

> Lösungen, die auf offenen Daten und offenem Code basieren, ermöglichen es Unternehmen, sich auf die Effizienz von Geschäftsprozessen zu konzentrieren, anstatt sich mit veralteten APIs und der Integration geschlossener Systeme herumzuschlagen. Ein bewusster Übergang zu einer offenen Architektur kann die Produktivität erheblich verbessern und die Abhängigkeit von Anbietern verringern.

Der Übergang zu einer neuen Realität bedeutet nicht nur eine Änderung der Ansätze für die Softwareentwicklung, sondern auch ein Überdenken des Prinzips der Arbeit mit Daten selbst. Im Mittelpunkt dieses Wandels steht nicht der Code, sondern die Information: ihre Struktur, Zugänglichkeit und Interpretierbarkeit. Hier rücken offene und strukturierte Daten in den Vordergrund und werden zu einem integralen Bestandteil der neuen digitalen Architektur.

Strukturierte offene Daten: die Grundlage der digitalen Transformation

Während in den vergangenen Jahrzehnten die Nachhaltigkeit von Unternehmen weitgehend von der Wahl der Softwarelösungen und der Abhängigkeit von bestimmten Anbietern bestimmt wurde, ist in der heutigen digitalen Wirtschaft die Datenqualität und die Fähigkeit, effektiv damit zu arbeiten, der entscheidende Faktor. Open-Source-Code ist ein wichtiger Bestandteil des neuen technologischen Paradigmas, aber sein Potenzial wird nur dann wirklich ausgeschöpft, wenn die Daten verständlich, organisiert und maschinenlesbar sind. Unter allen Arten von Datenmodellen werden strukturierte offene Daten zum Eckpfeiler einer nachhaltigen digitalen Transformation.

Der Hauptvorteil strukturierter offener Daten ist die eindeutige Interpretation und die Möglichkeit der automatisierten Verarbeitung. Dies ermöglicht erhebliche Effizienzsteigerungen sowohl auf der Ebene der einzelnen Vorgänge als auch in der gesamten Organisation.

> Laut dem Deloitte-Bericht "The Data Transfer Process in Enterprise Transformation" [82] ist die Zusammenarbeit mit der IT-Abteilung zur Verwaltung der Übertragung strukturierter Daten von entscheidender Bedeutung. Laut dem Bericht der britischen Regierung "Data Analytics and AI in Government Project Delivery" (2024) [83] ist die Beseitigung von Hindernissen für den Datenaustausch zwischen verschiedenen Projekten und Organisationen der Schlüssel zur Verbesserung der Effizienz des Projektmanagements. In dem Dokument wird betont, dass durch die Standardisierung von Datenformaten und die Einführung von Grundsätzen für offene Daten die Duplizierung von Informationen vermieden, Zeitverluste minimiert und die Genauigkeit von Prognosen verbessert werden können.

Für die Baubranche, in der traditionell ein hohes Maß an Fragmentierung und Formatvielfalt vorherrscht, spielen der strukturierte Vereinheitlichungsprozess und strukturierte offene Daten eine entscheidende Rolle

bei der Gestaltung kohärenter und handhabbarer Prozesse (Abb. 4.1-14). Sie ermöglichen es den Projektteilnehmern, sich auf die Verbesserung der Produktivität zu konzentrieren und nicht auf die Lösung technischer Probleme im Zusammenhang mit Inkompatibilitäten zwischen geschlossenen Plattformen, Datenmodellen und Formaten.

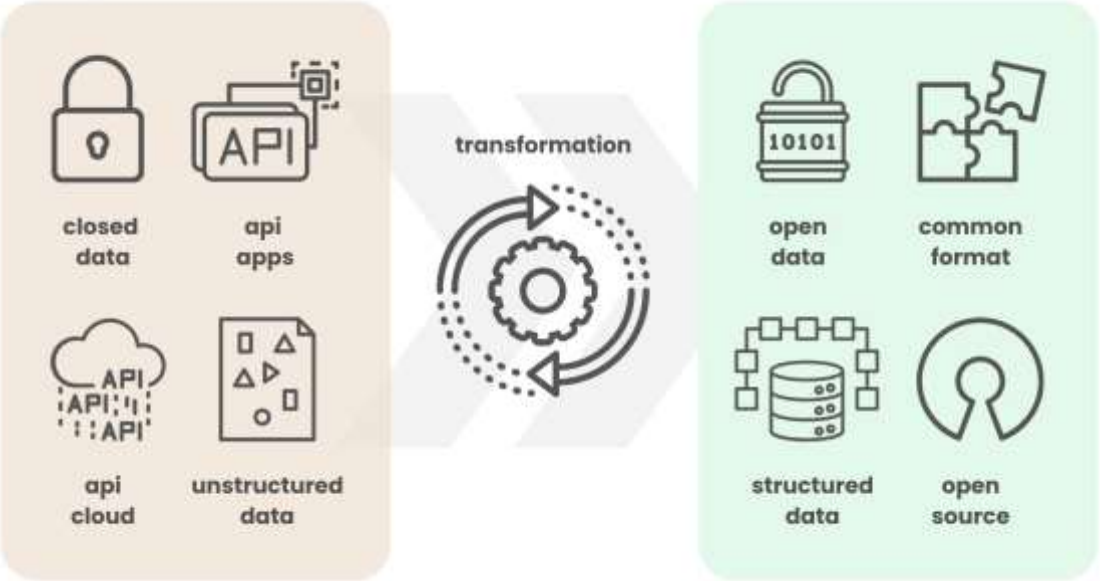


Abb. 3.2-15 Offene strukturierte Daten verringern die Abhängigkeit von Softwarelösungen und Plattformen und beschleunigen die Innovation.

Moderne technologische Hilfsmittel, auf die wir später in diesem Buch näher eingehen werden, ermöglichen nicht nur das Sammeln von Informationen, sondern auch deren automatische Bereinigung: Beseitigung von Duplikaten, Korrektur von Fehlern und Normalisierung von Werten. Dies bedeutet, dass Analysten und Ingenieure nicht mit unzusammenhängenden Dokumenten arbeiten, sondern mit einer organisierten Wissensbasis, die sich für Analysen, Automatisierung und Entscheidungsfindung eignet.

> *Machen Sie es so einfach wie möglich, aber nicht einfacher.*
>
> - Albert Einstein, theoretischer Physiker (die Urheberschaft des Zitats ist umstritten [84])

Die meisten Benutzeroberflächen für die Arbeit mit Daten können heute automatisch erstellt werden, ohne dass für jeden Geschäftsfall manuell Code geschrieben werden muss. Dies erfordert eine Infrastrukturschicht, die die Datenstruktur, das Modell und die Logik ohne zusätzliche Anweisungen versteht (Abb. 4.1-15). Es sind die strukturierten Daten, die diesen Ansatz möglich machen: Formulare, Tabellen, Filter und Ansichten können mit minimalem Programmieraufwand automatisch generiert werden.

Die wichtigsten benutzerkritischen Schnittstellen müssen unter Umständen noch manuell nachbearbeitet werden. In den meisten Fällen - und das sind zwischen 50 und 90 Prozent der Arbeitsszenarien - reicht jedoch die automatische Generierung von Anwendungen und Berechnungen ohne den Einsatz spezieller Anwendungen für diesen Zweck aus (Abb. 3.2-16), was die Entwicklungs- und Wartungskosten deutlich senkt, Fehler reduziert und die Implementierung digitaler Lösungen beschleunigt.

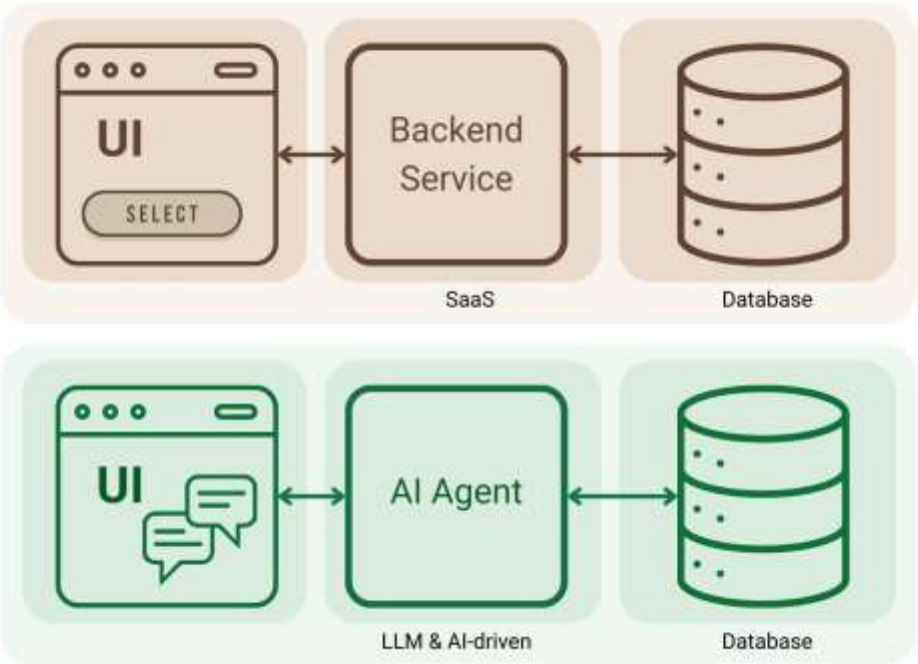


Abb. 3.2-16 Architekturmodelle für die Arbeit mit Daten: traditionelle Anwendungsarchitektur und KI -orientiertes Modell mit LLM.

Der Übergang von Architekturen, die auf einzelnen Anwendungen aufbauen, zu intelligent verwalteten Systemen auf der Grundlage von Sprachmodellen (LLM) ist der nächste Schritt der digitalen Evolution. In einer solchen Architektur werden strukturierte Daten nicht nur zu einem Speicherobjekt, sondern auch zur Grundlage für die Interaktion mit KI-Tools, die in der Lage sind, kontextabhängige Aktionen zu analysieren, zu interpretieren und zu empfehlen.

In den folgenden Kapiteln werden wir Beispiele aus der Praxis für die Implementierung einer Architektur auf der Grundlage offener strukturierter Daten betrachten und zeigen, wie Sprachmodelle zur automatischen Interpretation, Validierung und Verarbeitung von Daten eingesetzt werden. Anhand dieser Praxisbeispiele können Sie besser verstehen, wie die neue digitale Logik in der Praxis funktioniert - und welche Vorteile sie für Unternehmen bringt, die für die Transformation bereit sind.

KAPITEL 3.3.
LLM UND IHRE ROLLE BEI DER DATENVERARBEITUNG UND DEN GESCHÄFTSPROZESSEN

LLM Chaträume: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok zur Automatisierung von Datenprozessen

Das Aufkommen von Large Language Models (LLMs) war eine natürliche Erweiterung der Bewegung hin zu strukturierten offenen Daten und der Open-Source-Philosophie. Wenn Daten organisiert, zugänglich und maschinenlesbar werden, ist der nächste Schritt ein Werkzeug, das mit diesen Informationen interagieren kann, ohne dass komplexer Code geschrieben werden muss oder spezielle technische Kenntnisse erforderlich sind.

> LLMs sind ein direktes Produkt der Offenheit: große offene Datensätze, Publikationen und die Open-Source-Bewegung. Ohne offene wissenschaftliche Artikel, öffentlich verfügbare Textdaten und eine Kultur der kollaborativen Entwicklung gäbe es weder ChatGPT noch andere LLMs. Der LLM ist in gewissem Sinne ein "Destillat" des gesammelten digitalen Wissens der Menschheit, das durch die Prinzipien der Offenheit gesammelt und gebildet wird.

Moderne große Sprachmodelle (LLM - Large Language Models) wie ChatGPT ® (OpenAI), LlaMa ™ (Meta AI), Mistral DeepSeek™, Grok ™ (xAI), Claude ™ (Anthropic), QWEN™ bieten den Nutzern die Möglichkeit, Abfragen an Daten in natürlicher Sprache zu formulieren. Dies macht die Arbeit mit Informationen nicht nur für Entwickler, sondern auch für Analysten, Ingenieure, Planer, Manager und andere Fachleute zugänglich, die der Programmierung bisher fern standen

> **LLM** (Large Language Model) ist eine künstliche Intelligenz, die darauf trainiert ist, Texte auf der Grundlage riesiger Datenmengen aus dem gesamten Internet zu verstehen und zu generieren. Sie ist in der Lage, den Kontext zu analysieren, Fragen zu beantworten, Dialoge zu führen, Texte zu schreiben und Softwarecode zu generieren.

Wenn früher die Visualisierung, Verarbeitung oder Analyse von Daten die Kenntnis einer speziellen Programmiersprache erforderte: Python, SQL, R oder Scala, sowie die Fähigkeit, mit Bibliotheken wie Pandas, Polars oder DuckDB und vielen anderen zu arbeiten, hat sich die Situation ab 2023 radikal geändert. Jetzt kann der Benutzer einfach beschreiben, was er erhalten möchte - und das Modell selbst wird den Code generieren, ihn ausführen, eine Tabelle oder ein Diagramm anzeigen und das Ergebnis erklären. Zum ersten Mal seit Jahrzehnten hat die Entwicklung der Technologie nicht den Weg der Komplikation, sondern der radikalen Vereinfachung und Zugänglichkeit eingeschlagen.

Dieses Prinzip - "Daten mit Worten (Prompts) verarbeiten". - markiert eine neue Etappe in der Entwicklung der Arbeit mit Informationen und hebt die Erstellung von Lösungen auf eine noch höhere Abstraktionsebene. So wie es früher für die Benutzer nicht mehr notwendig war, die technischen Grundlagen des Internets zu verstehen, um Online-Shops zu betreiben oder Websites mit WordPress, Joomla und anderen Open-Source-

Baukastensystemen zu erstellen. Dies wiederum hat zu einem Boom bei digitalen Inhalten und Online-Geschäften geführt - heute können Ingenieure, Analysten und Manager Arbeitsabläufe ohne Kenntnisse von Programmiersprachen automatisieren. Erleichtert wird dies durch leistungsstarke LLMs - sowohl kostenlos als auch quelloffen, wie LLaMA, Mistral, Qwen, DeepSeek und andere - die fortgeschrittene Technologien einem möglichst breiten Publikum zugänglich machen.

Große LLM-Sprachmodelle: wie es funktioniert

Große Sprachmodelle (ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok) sind neuronale Netze, die auf großen Mengen von Textdaten aus dem Internet, Büchern, Artikeln und anderen Quellen trainiert werden. Ihre Hauptaufgabe ist es, den Kontext menschlicher Sprache zu verstehen und sinnvolle Antworten zu generieren.

Das moderne LLM basiert auf der Transformer-Architektur, die von Google-Forschern 2017 vorgeschlagen wurde [85]. Die Schlüsselkomponente dieser Architektur ist der Aufmerksamkeitsmechanismus, der es dem Modell ermöglicht, Beziehungen zwischen Wörtern unabhängig von ihrer Position im Text zu berücksichtigen.

> Der LLM-Lernprozess ähnelt im Entferntesten der Art und Weise, wie Menschen eine Sprache lernen, nur in millionenfacher Ausführung. Das Modell analysiert Milliarden von Beispielen von Wörtern und Ausdrücken und identifiziert Muster in der Struktur der Sprache und in der Logik der semantischen Übergänge. Der gesamte Text wird in Token - minimale semantische Einheiten (Wörter oder deren Teile) - zerlegt, die dann in Vektoren in einem mehrdimensionalen Raum umgewandelt werden (Abb. 8.2-2). Diese Vektordarstellungen ermöglichen es der Maschine, die verborgenen Beziehungen zwischen Konzepten zu "verstehen", anstatt den Text einfach als eine Folge von Symbolen zu betrachten.

Big Language Models sind nicht nur Werkzeuge zur Texterstellung. Sie sind in der Lage, Bedeutung zu erkennen, Verbindungen zwischen Konzepten zu finden und mit Daten zu arbeiten, auch wenn diese in unterschiedlichen Formaten vorliegen. Die Hauptsache ist, dass die Informationen in verständliche Modelle zerlegt und als Token dargestellt werden, mit denen das LLM arbeiten kann.

Der gleiche Ansatz kann auf Bauprojekte angewandt werden. Wenn wir uns ein Projekt als eine Art Text vorstellen, bei dem jedes Gebäude, Element oder Bauwerk ein Token ist, können wir beginnen, solche Informationen auf ähnliche Weise zu verarbeiten. Bauprojekte können mit Büchern verglichen werden, die in Kategorien, Kapiteln und Absatzgruppen gegliedert sind, die aus minimalen Token - Elementen eines Bauprojekts - bestehen (Abb. 3.3-1). Durch die Übersetzung von Datenmodellen in ein strukturiertes Format können wir strukturierte Daten auch in Vektorbasen übersetzen (Abb. 8.2-2), die eine ideale Quelle für maschinelles Lernen und Technologien wie LLM sind.

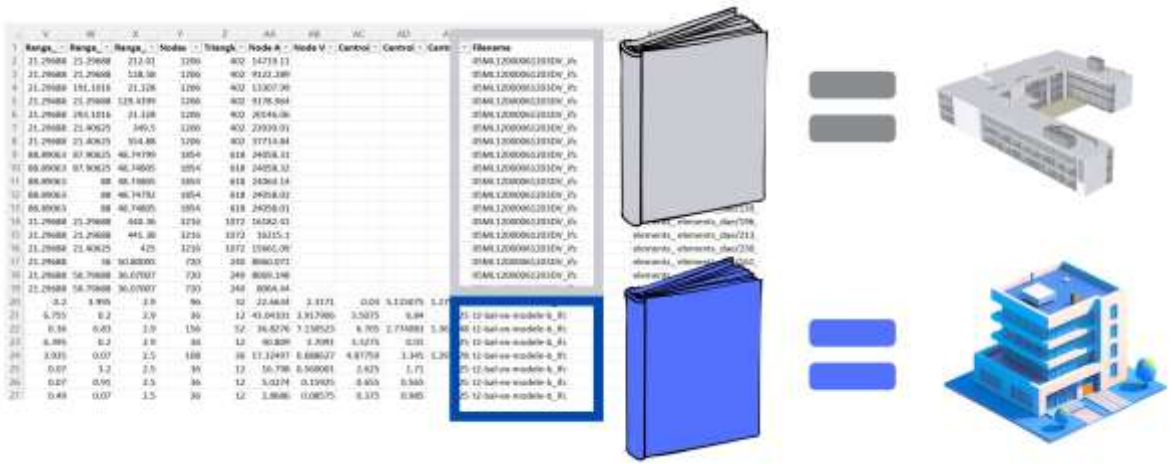


Abb. 3.3-1 Ein Bauprojektelement ist wie ein Token in einem Text: eine minimale Einheit, aus der Gruppen (Absätze) von Abschnitten (Kategorien) des gesamten Projekts gebildet werden.

Wenn ein Bauprojekt digitalisiert ist und seine Elemente als Token oder Vektoren dargestellt werden, ist es möglich, in natürlicher Sprache auf sie zuzugreifen, anstatt durch starre formale Abfragen. Hier kommt einer der Hauptvorteile des LLM ins Spiel - die Fähigkeit, die Bedeutung einer Abfrage zu verstehen und sie mit den relevanten Daten zu verknüpfen.

Der Ingenieur muss keine SQL -Abfrage oder Python-Code mehr schreiben, um die benötigten Daten zu erhalten - er kann einfach, mit dem Verständnis des LLM und der Datenstruktur, die Aufgabe in der üblichen Weise formulieren: "*Finde alle Stahlbetonstrukturen mit einer Betonklasse höher als B30 und berechne ihr Gesamtvolumen*". Das Modell erkennt die Bedeutung der Anfrage, wandelt sie in eine maschinenlesbare Form um, findet die Daten (gruppieren und transformieren) und gibt das Endergebnis zurück.

Dokumente, Tabellen, Projektmodelle werden in Vektordarstellungen umgewandelt (Einbettung) und in der Datenbank gespeichert. Wenn ein Benutzer eine Frage stellt, wird die Anfrage ebenfalls in einen Vektor umgewandelt und das System findet die relevantesten Daten. Auf diese Weise kann sich das LLM nicht nur auf sein eigenes trainiertes Wissen stützen, sondern auch auf aktuelle Unternehmensdaten, selbst wenn diese erst nach Abschluss des Modelltrainings erschienen sind.

> Einer der wichtigsten Vorteile des LLM im Bauwesen ist die Fähigkeit, Programmcode zu erzeugen. Anstatt die technische Aufgabe an einen Programmierer zu übergeben, können Spezialisten die Aufgabe in natürlicher Sprache beschreiben, und das Modell wird den notwendigen Code erstellen, der (durch Kopieren aus dem Chat) bei der Erstellung von Prozessautomatisierungscode verwendet werden kann. LLM -Modelle ermöglichen es Spezialisten ohne tiefe Programmierkenntnisse, zur Automatisierung und Verbesserung der Geschäftsprozesse des Unternehmens beizutragen.

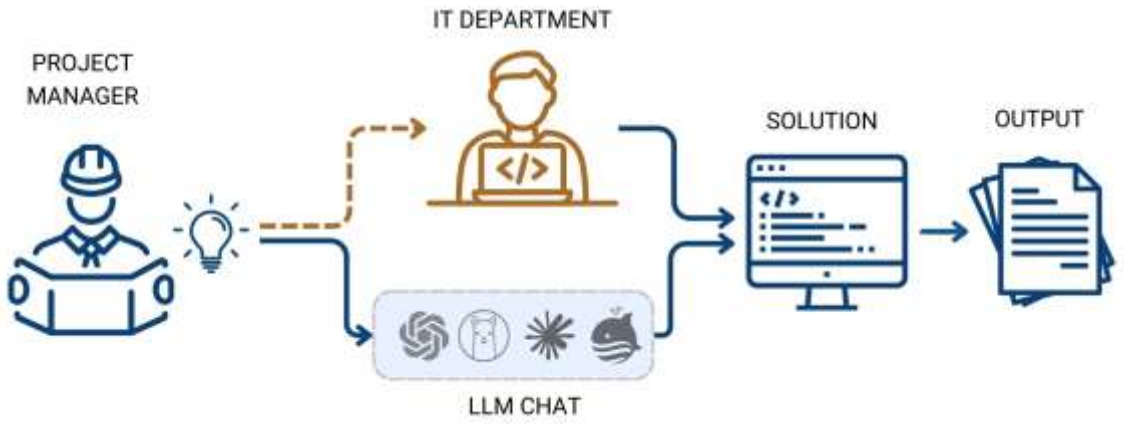


Abbildung 3.3-2 LLMs bieten den Nutzern die Möglichkeit, Code zu schreiben und Ergebnisse zu erhalten, ohne dass sie über Programmierkenntnisse verfügen müssen.

> Laut einer von Wakefield Research durchgeführten und von SAP gesponserten Studie aus dem Jahr 2024 [36], bei der 300 leitende Angestellte von Unternehmen mit einem Jahresumsatz von mindestens 1 Milliarde US-Dollar in den USA befragt wurden, vertrauen 52% der leitenden Angestellten auf KI, um Daten zu analysieren und Empfehlungen für die Entscheidungsfindung zu geben. Weitere 48% nutzen KI, um bisher nicht berücksichtigte Risiken zu identifizieren, und 47% nutzen KI, um alternative Pläne vorzuschlagen. Darüber hinaus nutzen 40 Prozent KI für die Entwicklung neuer Produkte, die Budgetplanung und die Marktforschung. Die Studie zeigt auch die positiven Auswirkungen von KI auf das Privatleben: 39% der Befragten berichten von einer besseren Work-Life-Balance, 38% von einer besseren psychischen Gesundheit und 31% von einem geringeren Stressniveau.

Trotz ihrer Leistungsfähigkeit bleiben LLM jedoch ein Instrument, das bewusst eingesetzt werden muss. Wie jede Technologie haben auch sie ihre Grenzen. Eines der bekanntesten Probleme sind die so genannten "Halluzinationen" - Fälle, in denen das Modell zwar eine plausible, aber faktisch falsche Antwort liefert. Daher ist es wichtig zu verstehen, wie das Modell funktioniert: welche Daten und Datenmodelle es fehlerfrei interpretieren kann, wie es Abfragen interpretiert und woher es seine Informationen bezieht. Es ist auch zu bedenken, dass das Wissen des LLM auf das Datum seiner Ausbildung beschränkt ist, und ohne eine Verbindung zu externen Daten berücksichtigt das Modell möglicherweise nicht die aktuellen Normen, Standards, Preise oder Technologien.

Die Lösung für diese Probleme liegt in der regelmäßigen Aktualisierung von Vektordatenbanken, der Anbindung an relevante Quellen und der Entwicklung autonomer KI - Agenten, die nicht nur Fragen beantworten, sondern proaktiv Daten zum Training nutzen, Aufgaben verwalten, Risiken erkennen, Optimierungsmöglichkeiten vorschlagen und die Projektleistung überwachen.

> Der Übergang zu LLM -Schnittstellen im Bauwesen ist nicht nur eine technologische Neuheit. Es ist ein Paradigmenwechsel, der die Barrieren zwischen Menschen und Daten beseitigt. Es geht um die Möglichkeit, mit Informationen so einfach zu arbeiten, wie wir miteinander reden - und trotzdem genaue, überprüfte und umsetzbare Ergebnisse zu erhalten.

Diejenigen Unternehmen, die früher als andere mit dem Einsatz solcher Werkzeuge beginnen, werden einen

erheblichen Wettbewerbsvorteil erlangen. Dazu gehören die Beschleunigung der Arbeit, die Senkung der Kosten und die Verbesserung der Qualität der Entwurfslösungen durch den schnellen Zugriff auf die Datenanalyse und die Möglichkeit, schnell Antworten auf komplexe Fragen zu finden. Aber es gibt auch Sicherheitsaspekte zu berücksichtigen. Die Nutzung von Cloud-basierten LLM -Diensten kann mit dem Risiko von Datenverlusten verbunden sein. Daher suchen Unternehmen zunehmend nach alternativen Lösungen, die es ihnen ermöglichen, LLM-Tools in ihrer eigenen Infrastruktur einzusetzen - vor Ort, mit vollem Schutz und voller Kontrolle über die Informationen.

Nutzung lokaler LLMs für sensible Unternehmensdaten

Das Erscheinen der ersten Chat-LLMs im Jahr 2022 markierte eine neue Etappe in der Entwicklung der künstlichen Intelligenz. Unmittelbar nach der breiten Einführung dieser Modelle stellte sich jedoch eine berechtigte Frage: Wie sicher ist es, unternehmensbezogene Daten und Abfragen in die Cloud zu übertragen? Die meisten Cloud-basierten Sprachmodelle speicherten den Kommunikationsverlauf und hochgeladene Dokumente auf ihren Servern, und für Unternehmen, die mit sensiblen Informationen umgehen, war dies ein ernsthaftes Hindernis für die Einführung von KI.

> Eine der nachhaltigsten und logischsten Lösungen für dieses Problem ist der Einsatz von Open Source LLM vor Ort, innerhalb der IT-Infrastruktur des Unternehmens. Im Gegensatz zu Cloud-Diensten funktionieren lokale Modelle ohne Internetverbindung, übertragen keine Daten auf externe Server und geben Unternehmen die volle Kontrolle über Informationen

> *Das beste offene Modell [Open Source LLM] ist derzeit in seiner Leistung vergleichbar mit geschlossenen Modellen [wie ChatGPT, Claude], allerdings mit einem Rückstand von etwa einem Jahr [77].*
>
> - Ben Cottier, leitender Forscher bei Epoch AI, einer gemeinnützigen Forschungsorganisation, 2024

Große Technologieunternehmen haben damit begonnen, ihre LLMs für die lokale Nutzung zur Verfügung zu stellen. Die quelloffene LLaMA-Serie von Meta und das schnell wachsende DeepSeek Projekt aus China waren Beispiele für den Übergang zu einer offenen Architektur. Daneben haben auch Mistral und Falcon leistungsstarke Modelle veröffentlicht, die frei von den Beschränkungen proprietärer Plattformen sind. Diese Initiativen haben nicht nur die Entwicklung der globalen KI beschleunigt, sondern auch datenschutzbewussten Unternehmen echte Alternativen in Bezug auf Unabhängigkeit, Flexibilität und Einhaltung von Sicherheitsvorschriften geboten.

In einem Unternehmensumfeld, insbesondere im Baugewerbe, ist der Datenschutz nicht nur eine Frage der Bequemlichkeit, sondern auch der Einhaltung von Vorschriften. Die Arbeit mit Ausschreibungsunterlagen, Kostenvoranschlägen, Zeichnungen und vertraulicher Korrespondenz erfordert strenge Kontrollen. Und hier bietet LLM die notwendige Sicherheit, dass die Daten innerhalb des Unternehmens bleiben.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
Data Control	Data is transmitted to third parties	Data remains within the company's network
License	Proprietary, paid	Open-source (Apache 2.0, MIT)
Infrastructure	Requires internet	Operates in an isolated environment
Customization	Limited	Full adaptation to company needs
Cost	Pay-per-token/request	One-time hardware investment + maintenance costs
Scalability	Easily scalable with cloud resources	Scaling requires additional local hardware
Security & Compliance	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
Performance & Latency	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
Integration	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
Updates & Maintenance	Automatically updated by provider	Requires manual updates and model retraining
Energy Consumption	Energy cost is covered by provider	High power consumption for inference and training
Offline Availability	Not available without an internet connection	Works completely offline
Inference Cost	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

Abb. 3.3-3 Lokale Modelle bieten vollständige Kontrolle und Sicherheit, während Cloud-basierte Lösungen eine einfache Integration und automatische Updates ermöglichen.

Die wichtigsten Vorteile der lokalen Open Source LLM:

- Vollständige Kontrolle über die Daten. Alle Informationen bleiben innerhalb des Unternehmens, wodurch unbefugter Zugriff und Datenverluste ausgeschlossen werden.
- Offline-Betrieb. Keine Abhängigkeit von der Internetverbindung, was besonders für isolierte IT-Infrastrukturen wichtig ist. Dies gewährleistet auch bei Sanktionen oder gesperrten Cloud-Diensten einen unterbrechungsfreien Betrieb.
- Flexibilität der Anwendung. Das Modell kann für die Texterstellung, die Datenanalyse, das Schreiben von Programmcode, die Designunterstützung und die Verwaltung von Geschäftsprozessen verwendet werden.
- Anpassung an die Unternehmensziele. LLM kann auf interne Dokumente trainiert werden, was

Ihnen erlaubt, die Besonderheiten der Arbeit des Unternehmens und seiner Branche zu berücksichtigen. Das lokale LLM kann mit CRM, ERP oder BI-Plattformen verbunden werden, wodurch Sie die Analyse von Kundenanfragen, die Erstellung von Berichten oder sogar Trendprognosen automatisieren können.

> Die Bereitstellung des kostenlosen und quelloffenen Modells von DeepSeek -R1-7B auf einem Server, auf den ein ganzes Team von Nutzern zugreifen kann, kostet mit 1000 Dollar pro Monat möglicherweise weniger als die Jahresgebühren für Cloud-APIs, wie ChatGPT oder Claude und ermöglicht es Unternehmen, die volle Kontrolle über ihre Daten zu übernehmen, ihre Übertragung ins Internet zu vermeiden und regulatorische Anforderungen wie die GDPR einzuhalten

In anderen Branchen ändern die lokalen LLMs bereits ihren Ansatz zur Automatisierung. Im Kundendienst beantworten sie häufige Kundenanfragen und verringern so die Arbeitsbelastung der Mitarbeiter. In Personalabteilungen analysieren sie Lebensläufe und wählen geeignete Bewerber aus. Im elektronischen Handel erstellen sie personalisierte Angebote, ohne die Daten der Nutzer preiszugeben.

Ein ähnlicher Effekt wird für den Bausektor erwartet. Dank der Integration von LLM mit Projektdaten und Standards ist es möglich, die Erstellung von Unterlagen zu beschleunigen, die Erstellung von Kostenvoranschlägen und vorausschauenden Kostenanalysen zu automatisieren. Die Verwendung von LLM in Verbindung mit strukturierten Tabellen und Datenrahmen wird zu einem besonders vielversprechenden Bereich.

Volle Kontrolle über AI im Unternehmen und wie Sie Ihr eigenes LLM einsetzen können

Mit modernen Werkzeugen können Unternehmen ein großes Sprachmodell (LLM) in nur wenigen Stunden lokal einsetzen. Dies ermöglicht eine vollständige Kontrolle über die Daten und die Infrastruktur, wodurch die Abhängigkeit von externen Cloud-Diensten entfällt und das Risiko von Informationsverlusten minimiert wird. Diese Lösung ist besonders wichtig für Organisationen, die mit sensiblen Projektdokumenten oder kommerziell sensiblen Daten arbeiten.

Je nach Aufgaben und Ressourcen stehen verschiedene Einsatzszenarien zur Verfügung, von sofort einsatzbereiten Lösungen bis hin zu flexibleren und skalierbaren Architekturen. Eines der einfachsten Tools ist Ollama, mit dem Sie Sprachmodelle buchstäblich mit einem Mausklick ausführen können, ohne dass Sie über tiefgreifende technische Kenntnisse verfügen müssen. Ein schneller Start mit Ollama:

1. Laden Sie die Distribution für Ihr Betriebssystem (Windows / Linux / macOS) von der offiziellen Website herunter: ollama.com
2. Installieren Sie das Modell über die Kommandozeile. Zum Beispiel für das Modell *Mistral*:

```
Ollama-Lauf Mistral
```

3. Nach der Ausführung des Modells ist es einsatzbereit - Sie können Textabfragen über das Terminal senden oder es in andere Tools integrieren. Starten Sie das Modell und führen Sie eine Abfrage aus:

```
ollama run mistral "Wie kann man eine Kalkulation mit allen Ressourcen für die In-
stallation einer 100 mm breiten Trennwand aus Gipskarton erstellen?".
```

Für diejenigen, die lieber in einer vertrauten visuellen Umgebung arbeiten, gibt es LM Studio, eine kostenlose Anwendung mit einer Oberfläche, die an ChatGPT erinnert

- Installieren Sie LM Studio, indem Sie das Distributionskit von der offiziellen Website - lmstudio.ai - herunterladen.
- Wählen Sie über den integrierten Katalog ein Modell aus (z. B. Falcon oder GPT-Neo-X) und laden Sie es herunter.
- Arbeiten Sie mit dem Modell über eine intuitive Schnittstelle, die an ChatGPT erinnert, aber vollständig lokalisiert ist.

	Developer	Parameters	GPU Requirements (GB)	Features	Best For
Mistral 7B	Mistral AI	7	8 (FP16)	Fast, supports multimodal tasks (text + images), fully open-source code	Lightweight tasks, mobile devices, laptops
LLaMA 2	Meta	7–70	16–48 (FP16)	High text generation accuracy, adaptable for technical tasks, CC-BY-SA license	Complex analytical and technical tasks
Baichuan 7B/13B	Baichuan Intelligence	7–13	8–16 (FP16)	Fast and efficient, great for large data processing, fully open-source code	Data processing, automating routine tasks
Falcon 7B/40B	Technology Innovation Institute (TII)	7–40	8–32 (FP16)	Open-source, high performance, optimized for fast work	Workloads with limited computational resources
DeepSeek-V3	DeepSeek	671	1543 (FP16) / 386 (4-bit)	Multilingual, 128K token context window, balanced speed and accuracy	Large enterprises, SaaS platforms, multitasking scenarios
DeepSeek-R1-7B	DeepSeek	7	18 (FP16) / 4.5 (4-bit)	Retains 92% of R1 capabilities in MATH-500, local deployment support	Budget solutions, IoT devices, edge computing

Abbildung 3.3-4 Vergleich beliebter lokaler, quelloffener LLM-Modelle - Modelle.

Die Wahl des Modells hängt von den Anforderungen an Geschwindigkeit, Genauigkeit und verfügbaren Hardwarekapazitäten ab (Abb. 3.3-4). Kleine Modelle wie Mistral 7B und Baichuan 7B eignen sich für leichte Aufgaben und mobile Geräte, während leistungsstarke Modelle wie DeepSeek -V3 erhebliche Rechenressourcen erfordern, aber hohe Leistung und Unterstützung für mehrere Sprachen bieten. In den kommenden Jahren wird der LLM-Markt rasch wachsen - wir werden immer mehr leichtgewichtige und spezialisierte

Modelle sehen. Anstelle von Allzweck-LLMs, die alle menschlichen Inhalte von abdecken, werden Modelle auftauchen, die auf ein enges Fachgebiet trainiert sind. So ist beispielsweise mit dem Aufkommen von Modellen zu rechnen, die ausschließlich für die Bearbeitung von technischen Berechnungen, Baukostenvoranschlägen oder Daten in CAD-Formaten konzipiert sind. Solche spezialisierten Modelle werden schneller, genauer und sicherer in der Anwendung sein - insbesondere in professionellen Umgebungen, in denen hohe Zuverlässigkeit und fachliche Tiefe wichtig sind.

Sobald das lokale LLM gestartet ist, kann es an die spezifischen Aufgaben des Unternehmens angepasst werden. Zu diesem Zweck wird die Technik der Feinabstimmung angewandt, bei der das Modell an internen Dokumenten, technischen Anweisungen, Vertragsvorlagen oder Projektunterlagen weiter trainiert wird.

RAG: Intelligente LLM -Assistenten mit Zugang zu Unternehmensdaten

Die nächste Stufe in der Entwicklung von LLM-Anwendungen in der Wirtschaft ist die Integration von Modellen mit tatsächlichen Echtzeit-Unternehmensdaten. Dieser Ansatz wird RAG (Retrieval-Augmented Generation) genannt - Retrieval-Augmented Generation. In dieser Architektur wird das Sprachmodell nicht nur zu einer Dialogschnittstelle, sondern zu einem vollwertigen intelligenten Assistenten, der in der Lage ist, durch Dokumente, Zeichnungen und Datenbanken zu navigieren und genaue, kontextbezogene Antworten zu geben.

> Der Hauptvorteil von RAG ist die Möglichkeit, unternehmensinterne Daten zu nutzen, ohne das Modell vorher trainieren zu müssen, und dabei eine hohe Genauigkeit und Flexibilität in der Informationsverarbeitung zu erhalten.

Die RAG Technologie kombiniert zwei Hauptkomponenten:

- Abruf: Das Modell stellt eine Verbindung zu Datenspeichern her - Dokumente, Tabellen, PDF -Dateien, Zeichnungen - und ruft die vom Benutzer angeforderten Informationen ab.
- Erweiterte Generierung: Auf der Grundlage der extrahierten Daten generiert das Modell eine genaue, fundierte Antwort, die den Kontext und die Besonderheit der Anfrage berücksichtigt.

Um LLM mit RAG-Unterstützung zu betreiben, sind einige Schritte zu beachten:

- **Datenaufbereitung**: Sammeln Sie die erforderlichen Dokumente, Zeichnungen, Spezifikationen, Tabellen. Sie können in verschiedenen Formaten und Strukturen vorliegen, von PDF bis Excel.
- **Indizierung und Vektorisierung**: Mit Hilfe von Tools wie LlamaIndex oder LangChain werden Daten in Vektordarstellungen umgewandelt, die es ermöglichen, semantische Verknüpfungen zwischen Textfragmenten zu finden (mehr über Vektordatenbanken und die Umwandlung großer Arrays in Vektordarstellungen, einschließlich CAD Projekte, in Teil 8).
- **Abfrage des Assistenten**: Sobald die Daten hochgeladen sind, können Sie dem Modell Fragen stellen, und es wird nach Antworten innerhalb des Unternehmensrahmens und nicht nach allgemeinem Wissen aus dem Internet suchen.

Angenommen, ein Unternehmen hat einen Ordner constructionsite_docs, in dem Verträge, Anweisungen,

Kostenvoranschläge und Tabellen gespeichert sind. Mit Hilfe eines Python-Skripts (Abb. 3.3-5) können wir diesen Ordner durchsuchen und eine Vektorindizierung aufbauen: Jedes Dokument wird in eine Reihe von Vektoren umgewandelt, die den semantischen Inhalt des Textes widerspiegeln. Dadurch werden die Dokumente zu einer Art "Landkarte der Bedeutungen", auf der das Modell effizient navigieren und Verbindungen zwischen Begriffen und Phrasen finden kann.

Das Modell "merkt" sich zum Beispiel, dass die Worte "Rückgabe" und "Reklamation" häufig in dem Abschnitt des Vertrags vorkommen, der sich auf die Lieferung von Materialien zur Baustelle bezieht. Wenn dann eine Frage gestellt wird - zum Beispiel: "Wie lang ist unsere Rückgabefrist?" (Abb. 3.3-5 - Zeile 11 des Codes) - analysiert das LLM interne Dokumente und findet die genaue Information, indem es sich wie ein intelligenter Assistent verhält, der den Inhalt aller Unternehmensdateien lesen und verstehen kann.

llama.py

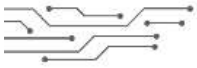
```
1  from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3  # Load documents from the folder
4  documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6  # Creating a vector index for semantic search
7  index = VectorStoreIndex.from_documents(documents)
8
9  # Integration with LLM (e.g. Llama 3)
10 query_engine = index.as_query_engine()
11 response = query_engine.query("What are the return terms in the contracts?")
12 print(response)
```

Abb. 3.3-5 LM liest einen Ordner mit Dateien - ähnlich der Art und Weise, wie ein Mensch einen Ordner öffnet und nach dem gewünschten Dokument sucht

Der Code kann auf jedem Computer ausgeführt werden, auf dem Python installiert ist. Wir werden im nächsten Kapitel mehr über die Verwendung von Python und IDEs zur Ausführung des Codes erfahren.

Der lokale Einsatz von LLM ist nicht nur ein Trend, sondern eine strategische Lösung für Unternehmen, die Wert auf Sicherheit und Flexibilität legen. Der Einsatz von LLM, ob auf lokalen Firmencomputern oder mit Online-Lösungen, ist jedoch nur der erste Schritt. Um LLM-Fähigkeiten auf reale Aufgaben anzuwenden, müssen Unternehmen Tools verwenden, die es ihnen ermöglichen, nicht nur Chat-Antworten zu empfangen, sondern auch die erstellte Logik in Form von Code zu speichern, der außerhalb des Kontexts der LLM-Nutzung ausgeführt werden kann. Dies ist wichtig für die Skalierung von Lösungen - richtig organisierte Prozesse ermöglichen die Anwendung von KI-Entwicklungen auf mehrere Projekte gleichzeitig oder sogar auf das gesamte Unternehmen.

In diesem Zusammenhang spielt die Wahl einer geeigneten Entwicklungsumgebung (IDE) eine wichtige Rolle. Moderne Programmierwerkzeuge ermöglichen nicht nur die Entwicklung von LLM-basierten Lösungen, sondern auch deren Integration in bestehende Geschäftsprozesse, wodurch sie zu automatisierten ETL -Pipeline werden



KAPITEL 3.4.
IDE MIT LLM-UNTERSTÜTZUNG UND ZUKÜNFTIGE PROGRAMMÄNDERUNGEN

Auswahl eines IDE: von LLM Experimenten bis zu Unternehmenslösungen

Wenn Sie in die Welt der Automatisierung, Datenanalyse und künstlichen Intelligenz eintauchen - insbesondere bei der Arbeit mit großen Sprachmodellen (LLMs) - ist es entscheidend, die richtige integrierte Entwicklungsumgebung (IDE) zu wählen. Diese IDE wird Ihr Hauptarbeitswerkzeug sein: der Ort, an dem der von der LLM generierte Code ausgeführt wird, sowohl auf einem lokalen Computer als auch innerhalb des Unternehmensnetzwerks. Die Wahl der IDE bestimmt nicht nur die Bequemlichkeit Ihrer Arbeit, sondern auch, wie schnell Sie in der Lage sein werden, von experimentellen LLM-Abfragen zu vollwertigen, in reale Geschäftsprozesse eingebetteten Lösungen überzugehen.

> **IDE** (Integrated Development Environment) ist ein vielseitiger Baustein auf Ihrem Computer für die Prozessautomatisierung und Datenverarbeitung. Anstatt eine Säge, einen Hammer, eine Bohrmaschine und andere Werkzeuge separat aufzubewahren, haben Sie ein Gerät, das alles kann - schneiden, befestigen, bohren und sogar die Qualität der Materialien prüfen. IDE für Programmierer ist ein einziger Raum, in dem Sie Code schreiben (in Analogie zum Bauwesen - Zeichnungen erstellen), seine Arbeit testen (Montage von Gebäudemodellen), Fehler finden (wie die Überprüfung der Festigkeit von Strukturen im Bauwesen) und das fertige Projekt ausführen können (Inbetriebnahme des Hauses).

Ein Überblick über gängige IDEs:

- **PyCharm®** (JetBrains) ist eine leistungsstarke professionelle IDE für Python. Aufgrund der großen Anzahl integrierter Funktionen ist sie gut für ernsthafte Projekte geeignet. Allerdings ist die grundlegende Unterstützung für interaktive Jupyter-Dateien (IPYNB) nur in der kostenpflichtigen Version verfügbar, und Anfänger könnten die Benutzeroberfläche als überwältigend empfinden.

> Eine Datei mit der Erweiterung IPYNB (Interactive Python Notebook) ist ein Format für interaktive Jupyter® Notebooks (Abb. 3.4-1), in dem Code, Visualisierungen und Erklärungen in einem einzigen Dokument zusammengefasst sind. Dieses Format ist ideal für die Erstellung von Berichten, Analysen und Trainingsszenarien.

- **VS Code®** (Microsoft) ist ein schnelles, flexibles und anpassbares Werkzeug mit kostenloser IPYNB-Unterstützung und vielen Plug-ins. Sowohl für Anfänger als auch für Profis geeignet. Ermöglicht die Integration von GitHub Copilot und Sprachmodell-Plugins, was es zu einer guten Wahl für KI und Data Science-Projekte macht.
- **Jupyter Notebook** - Ein klassisches und beliebtes Mittel zum Experimentieren und Lernen. Es ermöglicht das Schreiben von Code, das Hinzufügen von Erklärungen und die Visualisierung von Ergebnissen in einer einzigen Oberfläche (Abb. 3.4-1). Ideal für das schnelle Testen von Hypothesen, die Arbeit mit LLM und die Erstellung reproduzierbarer wilder Datenanalyseschritte. Zur Verwaltung

von Abhängigkeiten und Bibliotheken empfehlen wir die Verwendung von Anaconda Navigator, einer visuellen Schnittstelle zur Verwaltung der Python-Umgebung.

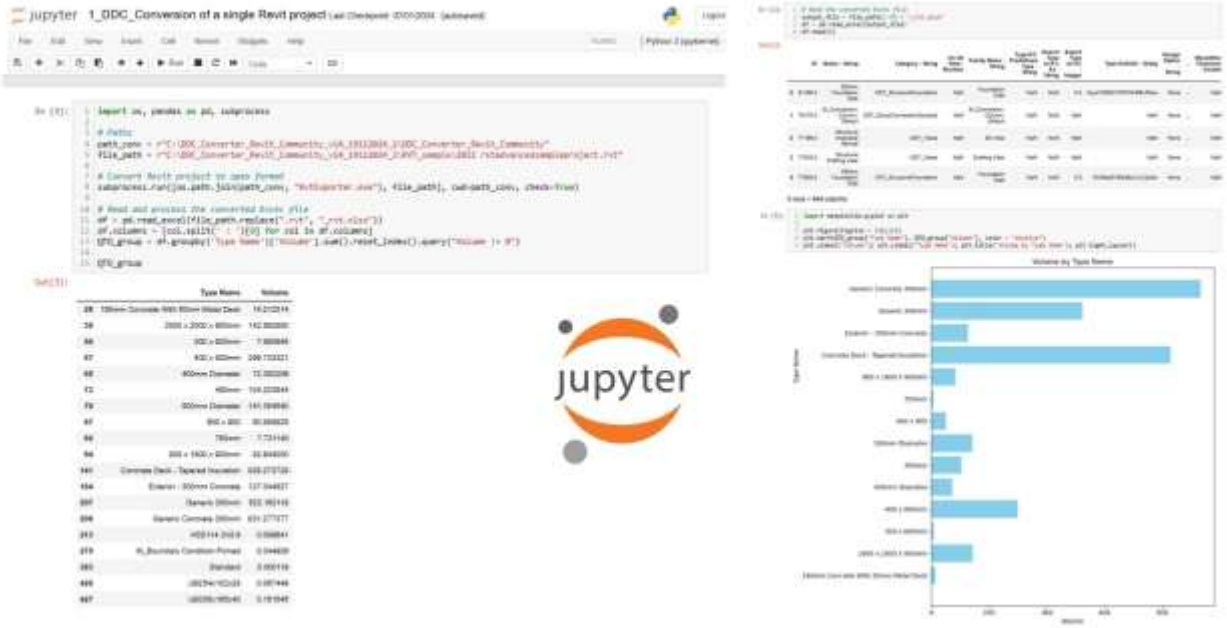


Abb. 3.4-1 Jupyter Notebook ist eines der bequemsten und beliebtesten Werkzeuge zur Erstellung von Pipeline-Prozessen.

- **Google Collab**™ (und die Kaggle-Plattform (Abb. 9.2-5)) ist eine cloudbasierte Alternative zu Jupyter, die kostenlosen GPU/TPU-Zugang bietet. Es ist eine großartige Lösung für den Einstieg - keine lokale Softwareinstallation und die Möglichkeit, direkt über einen Browser zu arbeiten. Es unterstützt die Integration mit Google Drive und seit kurzem auch mit Gemini (Googles LLM).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
Complexity	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

Abbildung 3.4-2 IDE-Vergleich: Jupyter Notebook eines der bequemsten und einfachsten Werkzeuge zur Erstellung von Pipeline Prozessen.

Die Wahl der IDE hängt von Ihren Aufgaben ab. Wenn Sie schnell mit AI arbeiten wollen, versuchen Sie Jupyter Notebook oder Google Collab. Für ernsthafte Projekte ist es besser, PyCharm oder VS Code zu verwenden. Die Hauptsache ist, dass Sie loslegen können. Mit modernen Tools können Sie Ihre Experimente schnell in funktionierende Lösungen umwandeln.

Alle oben beschriebenen IDEs ermöglichen die Erstellung von Datenverarbeitungspipelines, d. h. Ketten von Codeblockmodulen (die vom LLM generiert werden könnten), von denen jedes beispielsweise für einen anderen Schritt zuständig ist:

- analytische Szenarien,
- Ketten der Informationsextraktion aus Dokumenten,
- automatische Antworten auf der Grundlage der RAG,
- Erstellung von Berichten und Visualisierungen.

Dank des modularen Aufbaus kann jeder Schritt als eigener Block dargestellt werden: Laden der Daten→ Filtern→ Analysieren→ Visualisieren→ Exportieren der Ergebnisse. Diese Blöcke können wiederverwendet, - angepasst und zu neuen Ketten zusammengesetzt werden, wie ein Konstruktor, nur für Daten.

Für Ingenieure, Manager und Analysten eröffnet dies die Möglichkeit, Entscheidungslogik in Form von Code zu dokumentieren, der mit LLM generiert werden kann. Dieser Ansatz hilft, Routineaufgaben zu beschleunigen, typische Abläufe zu automatisieren und wiederholbare Prozesse zu schaffen, bei denen jeder Schritt klar dokumentiert und für alle Teammitglieder transparent ist.

Die automatisierten ETL-Tools Pipelines (Abb. 7.2-3), Apache Airflow (Abb. 7.4-4), Apache NiFi (Abb. 7.4-5) und n8n (Abb. 7.4-6) zum Aufbau von Logikbausteinen für die Prozessautomatisierung werden in Teil 7 und Teil 8 des Buches ausführlicher behandelt.

IDE mit LLM-Unterstützung und zukünftige Programmieränderungen

Die Integration von künstlicher Intelligenz in Entwicklungsprozesse verändert die Programmierlandschaft. Moderne Umgebungen sind nicht mehr nur Texteditoren mit Syntaxhervorhebung - sie verwandeln sich in intelligente Assistenten, die in der Lage sind, die Projektlogik zu verstehen, Code zu vervollständigen und sogar zu erklären, wie ein bestimmtes Codefragment funktioniert. Es kommen Produkte auf den Markt, die KI nutzen, um die Grenzen der konventionellen Entwicklung zu verschieben:

- **GitHub Copilot** (integriert mit VS Code, PyCharm): KI -Assistent, der Code auf der Grundlage von Kommentaren oder Teilbeschreibungen generiert und textliche Hinweise in fertige Lösungen umwandelt.
- **Cursor** (ein Fork von VS Code mit AI -kernel): ermöglicht nicht nur das Schreiben von Code zu beenden, sondern auch Fragen an das Projekt zu stellen, nach Abhängigkeiten zu suchen und aus der Codebasis zu lernen.
- **JetBrains AI Assistant**: ein Plugin für JetBrains IDE (einschließlich PyCharm) mit der Funktion der Erklärung von komplexem Code, Optimierung und Testerstellung.
- **Amazon CodeWhisperer**: ein Analogon zu Copilot mit dem Schwerpunkt auf Sicherheit und Unterstützung für Amazons AWS-Dienste.

Die Programmierung wird in den kommenden Jahren einen dramatischen Wandel durchmachen. Der

Schwerpunkt wird sich vom routinemäßigen Schreiben von Code auf das Entwerfen von Modellen und Datenarchitekturen verlagern - Entwickler werden stärker in das Systemdesign einbezogen, während die KI Vorlagenaufgaben übernehmen wird: Codegenerierung, Tests, Dokumentation und grundlegende Funktionen. Die Zukunft der Programmierung ist eine Zusammenarbeit zwischen Menschen und KI, bei der Maschinen die technische Routine übernehmen und Menschen sich auf die Kreativität konzentrieren.

Die Programmierung in natürlicher Sprache wird zum Alltag werden. Die Personalisierung von IDEs wird ein neues Niveau erreichen - Entwicklungsumgebungen werden lernen, sich an den Arbeitsstil und die Unternehmen des Nutzers anzupassen, indem sie Muster vorhersehen, kontextbezogene Lösungen anbieten und aus früheren Projekten lernen

> Damit wird die Rolle des Entwicklers nicht abgeschafft, aber sie wird sich radikal verändern: vom Schreiben von Code zum Wissens-, Qualitäts- und Prozessmanagement. Diese Entwicklung wird sich auch auf Business Intelligence auswirken, wo die Erstellung von Berichten, Visualisierungen und Anwendungen zur Entscheidungsunterstützung zunehmend durch die Generierung von Code und Logik mit Hilfe von KI und LLM, Chat und Agentenschnittstellen erfolgen wird.

Sobald ein Unternehmen LLM-Chats eingerichtet und eine geeignete Entwicklungsumgebung ausgewählt hat, besteht der nächste wichtige Schritt darin, die Daten zu organisieren. Dieser Prozess umfasst die Extraktion von Informationen aus unterschiedlichen Quellen, ihre Bereinigung, ihre Umwandlung in eine strukturierte Form und ihre Integration in Unternehmenssysteme.

Bei einem modernen datenzentrierten Ansatz für die Datenverwaltung besteht ein Hauptziel darin, Daten in eine einzige universelle Form zu bringen, die mit einer großen Anzahl von Werkzeugen und Anwendungen kompatibel ist. Um Strukturierungsprozesse und strukturierte Daten zu handhaben, werden spezialisierte Bibliotheken benötigt. Eine der leistungsfähigsten, flexibelsten und beliebtesten ist die Pandas-Bibliothek für Python. Mit ihr lassen sich Tabellendaten bequem verarbeiten: filtern, gruppieren, bereinigen, anhängen, aggregieren und auswerten.

Python Pandas: ein unverzichtbares Werkzeug für die Arbeit mit Daten

Pandas nimmt einen besonderen Platz in der Welt der Datenanalyse und Automatisierung ein. Es ist eine der beliebtesten und am weitesten verbreiteten Bibliotheken der Programmiersprache Python [86], die für die Arbeit mit strukturierten Daten entwickelt wurde.

Eine Bibliothek ist wie ein Satz von vorgefertigten Werkzeugen: Funktionen, Module, Klassen. So wie man auf einer Baustelle nicht jedes Mal einen Hammer oder eine Wasserwaage neu erfinden muss, ermöglichen es Bibliotheken in der Programmierung, Probleme schnell zu lösen, ohne grundlegende Funktionen und Lösungen neu erfinden zu müssen.

> Pandas ist eine Open-Source-Python-Bibliothek , die leistungsstarke und intuitive Datenstrukturen bietet, insbesondere DataFrame, ein universelles Format für die Arbeit mit Tabellen. Pandas ist ein Schweizer Messer für Analysten, Ingenieure und Entwickler, die mit Daten arbeiten.

Python ist eine hochentwickelte Programmiersprache mit einer einfachen Syntax, die in den Bereichen Analytik, Automatisierung, maschinelles Lernen und Webentwicklung eingesetzt wird. Ihre Beliebtheit ist auf die Lesbarkeit des Codes, die plattformübergreifende Natur und das reichhaltige Ökosystem an Bibliotheken zurückzuführen. Bis heute wurden mehr als 137.000 Open-Source-Pakete für Python erstellt [87], und diese Zahl steigt fast täglich weiter an. Jede dieser Bibliotheken ist eine Art Repository mit vorgefertigten Funktionen: von einfachen mathematischen Operationen bis hin zu komplexen Tools für die Bildverarbeitung, Big-Data-Analyse, neuronale Netze und die Integration mit externen Diensten.

> Mit anderen Worten: Stellen Sie sich vor, Sie haben freien und offenen Zugang zu Hunderttausenden von sofort einsatzbereiten Softwarelösungen - Bibliotheken und Tools, die Sie direkt in Ihre Geschäftsprozesse einbetten können. Es ist wie ein riesiger Katalog von Anwendungen für die Automatisierung, Analyse, Visualisierung, Integration und mehr - und alle sind sofort nach der Installation von Python verfügbar.

Pandas ist eines der beliebtesten Pakete im Python-Ökosystem. Im Jahr 2022 erreichte die durchschnittliche Anzahl der Downloads der Pandas-Bibliothek 4 Millionen pro Tag (Abb. 3.4-3), während diese Zahl bis Anfang 2025 auf 12 Millionen Downloads pro Tag anstieg, was die wachsende Beliebtheit und die weit verbreitete Verwendung in der Datenanalyse und im LLM-Chat widerspiegelt [86]

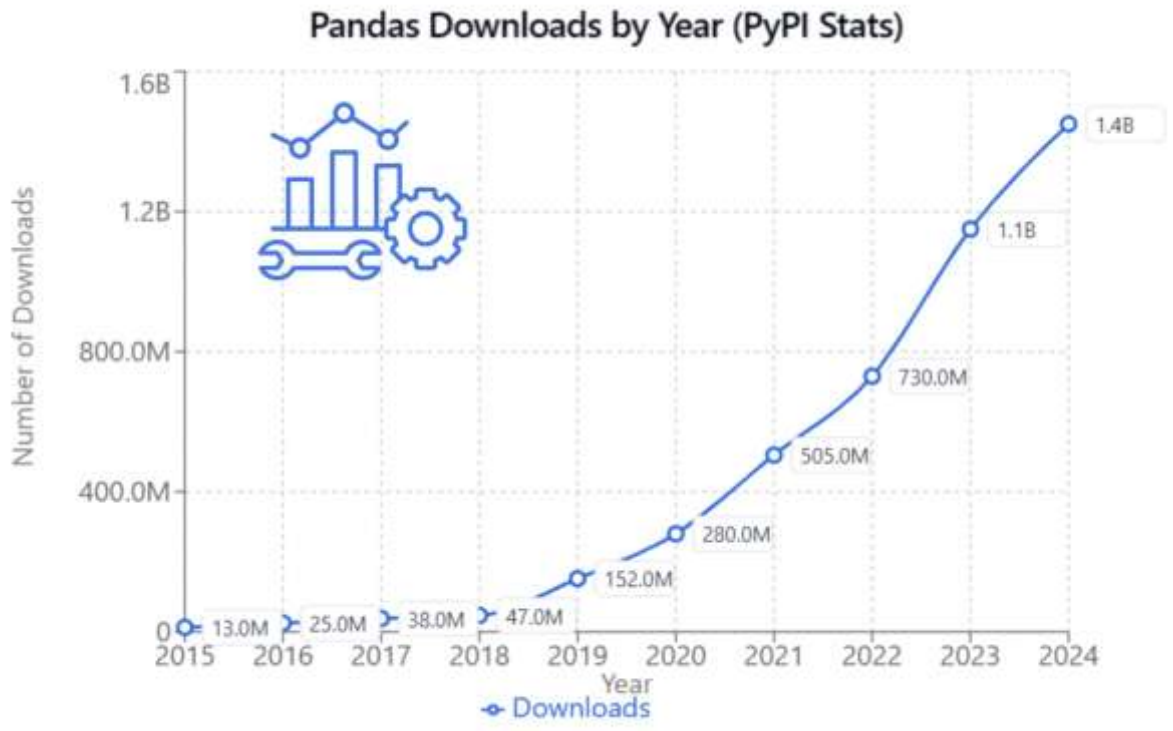


Abbildung 3.4-3 Pandas ist eine der am häufigsten heruntergeladenen Bibliotheken. Im Jahr 2024 überstieg die Zahl der jährlichen Downloads 1,4 Milliarden.

Die Abfragesprache in der Pandas-Bibliothek ähnelt in ihrer Funktionalität der SQL-Abfragesprache, die wir

im Kapitel "Relationale Datenbanken und SQL-Abfragesprache" besprochen haben.

> In der Welt der Analytik und des strukturierten Datenmanagements zeichnet sich Pandas durch seine Einfachheit, Schnelligkeit und Leistungsfähigkeit aus und bietet den Nutzern eine breite Palette von Werkzeugen zur effektiven Analyse und Verarbeitung von Informationen.

Beide Tools - SQL und Pandas - bieten leistungsstarke Funktionen zur Datenmanipulation, insbesondere im Vergleich zu herkömmlichem Excel. Sie unterstützen Operationen wie Auswahl und Filterung (Abb. 3.4-4), wobei der einzige Unterschied darin besteht, dass SQL für die Arbeit mit relationalen Datenbanken optimiert ist, während Pandas die Daten im RAM verarbeitet, wodurch es auf jedem Computer laufen kann, ohne dass Datenbanken erstellt und eine separate Infrastruktur eingerichtet werden muss.

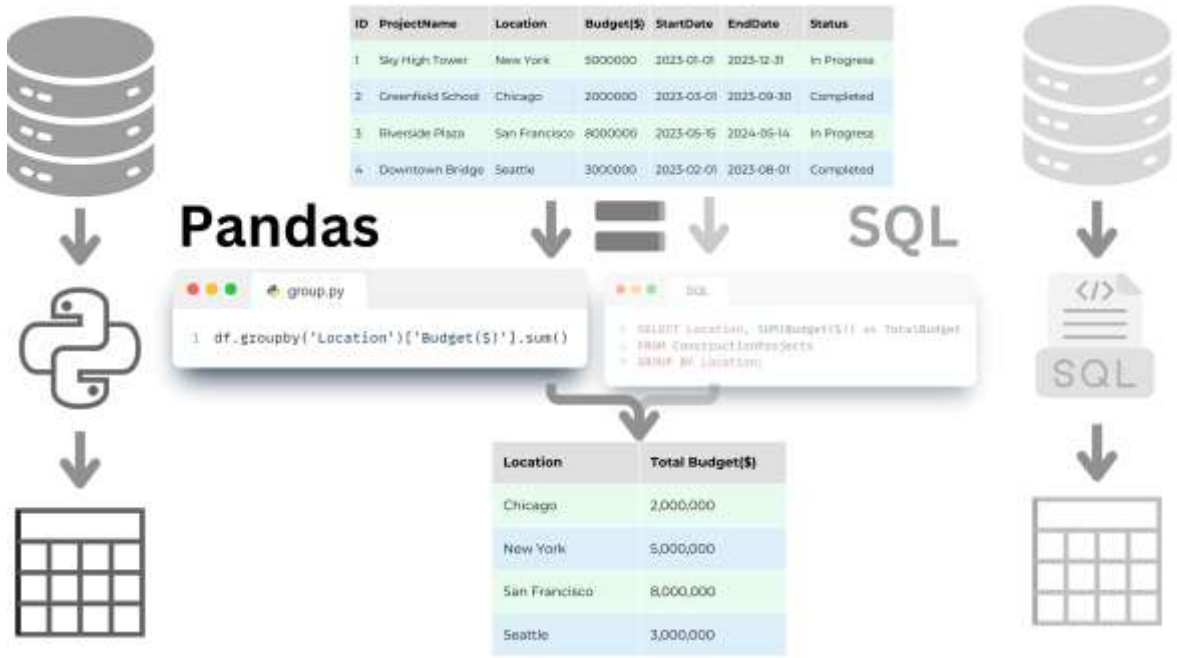


Abb. 3.4-4 Pandas ist im Gegensatz zu SQL flexibel und kann mit einer Vielzahl von Datenformaten arbeiten, die nicht auf Datenbanken beschränkt sind.

Pandas wird häufig für die wissenschaftliche Forschung, die Prozessautomatisierung, die Erstellung von Pipelines (einschließlich ETL) und die Datenmanipulation in Python bevorzugt, während SQL ein Standard für die Datenbankverwaltung ist und häufig in Unternehmensumgebungen zur Verarbeitung großer Datenmengen verwendet wird.

> Mit der Pandas-Bibliothek der Programmiersprache Python können Sie nicht nur grundlegende Operationen wie das Lesen und Schreiben von Tabellen durchführen, sondern auch komplexere Aufgaben wie das Zusammenführen von Daten, das Gruppieren von Daten und das Durchführen komplexer analytischer Berechnungen.

Heute wird die Bibliothek Pandas nicht nur in der akademischen Forschung und der Unternehmensanalyse

verwendet, sondern auch in Verbindung mit LLM -Modellen. So hat beispielsweise die Meta®-Abteilung (Facebook™) bei der Veröffentlichung des neuen Open-Source-Modells LlaMa 3.1 im Jahr 2024 besonderes Augenmerk auf die Arbeit mit strukturierten Daten gelegt und als einen der ersten Fälle in ihrer Veröffentlichung genau die Verarbeitung strukturierter Datenrahmen (Abb. 3.4-5) im CSV-Format und die Integration mit der Pandas-Bibliothek direkt im Chat vorgesehen.

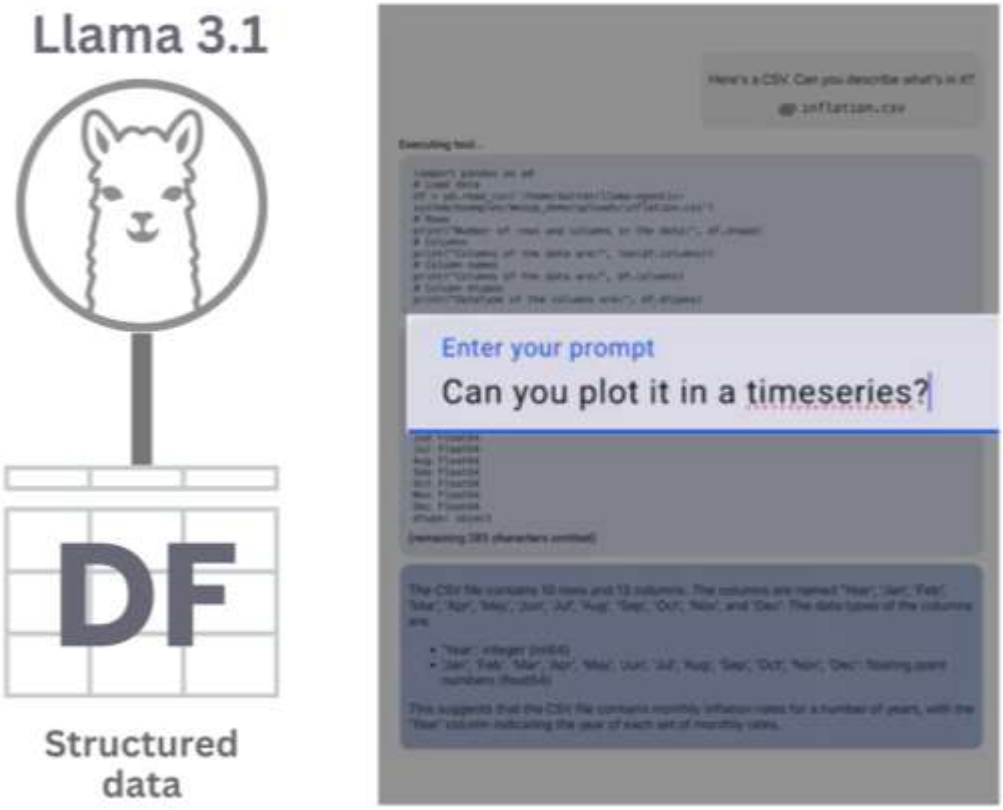


Abb. 3.4-5 Einer der ersten und wichtigsten Fälle des Meta-Teams, der in LlaMa 3.1 im Jahr 2024 vorgestellt wurde, war die Erstellung von Anwendungen mit Pandas.

> *Pandas ist ein unverzichtbares Werkzeug für Millionen von Datenwissenschaftlern, die Daten für generative KI verarbeiten und aufbereiten. Die Beschleunigung von Pandas ohne Code-Änderungen ist ein großer Schritt nach vorn. Datenwissenschaftler werden in der Lage sein, Daten in Minuten statt in Stunden zu verarbeiten und um Größenordnungen mehr Daten zum Trainieren generativer KI-Modelle zu erhalten [88].*
>
> - Jensen Huang, Gründer und CEO von NVIDIA

Mit Pandas ist es möglich, Datensätze zu verwalten und zu analysieren, die weit über die Möglichkeiten von Excel hinausgehen. Während Excel in der Regel bis zu 1 Million Datenzeilen verarbeiten kann, kann Pandas problemlos Datensätze (Abb. 9.1-2, Abbildung 9.1-10) mit Dutzenden von Millionen von Zeilen verarbeiten

[89]. Diese Fähigkeit erlaubt es den Nutzern, anspruchsvolle Datenanalysen und Visualisierungen auf großen Datensätzen durchzuführen, die tiefe Einblicke liefern und datengestützte Entscheidungsfindungen erleichtern. Darüber hinaus wird Pandas stark von der Community unterstützt [90]: Hunderte von Millionen von Entwicklern und Analysten weltweit (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) verwenden es täglich online oder offline und bieten eine große Anzahl von sofort einsetzbaren Lösungen für jedes Geschäftsproblem.

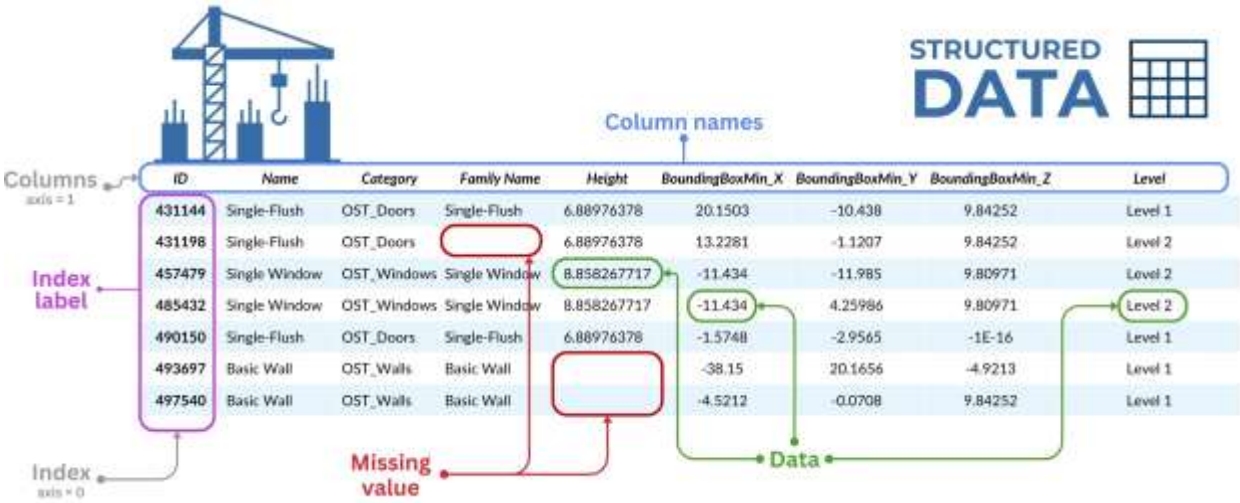
Das Herzstück der meisten analytischen Prozesse in Python ist eine strukturierte Form von Daten namens DataFrame, die von der Pandas-Bibliothek bereitgestellt wird. Es handelt sich dabei um ein leistungsstarkes und flexibles Werkzeug für die Organisation, Analyse und Visualisierung von Tabellendaten.

DataFrame: universelles tabellarisches Datenformat

DataFrame ist die zentrale Struktur in der Pandas-Bibliothek, bei der es sich um eine zweidimensionale Tabelle handelt (Abb. 3.4-6), bei der die Zeilen den einzelnen Objekten oder Datensätzen und die Spalten ihren Eigenschaften, Parametern oder Kategorien entsprechen. Diese Struktur ähnelt visuell Excel-Tabellen, ist aber in Bezug auf Flexibilität, Skalierbarkeit und Funktionalität weit überlegen.

> Ein **DataFrame** ist eine Methode zur Darstellung und Verarbeitung von Tabellendaten, die im RAM des Computers gespeichert sind.

DataFrame ist eine Methode zur Darstellung und Verarbeitung von tabellarischen Daten, die im Arbeitsspeicher des Computers gespeichert sind. In einer Tabelle können die Zeilen z. B. die Elemente eines Bauprojekts und die Spalten ihre Eigenschaften darstellen: Kategorien, Abmessungen, Koordinaten, Kosten, Begriffe usw. Außerdem kann eine solche Tabelle sowohl Informationen über ein einziges Projekt (Abb. 4.1-13) als auch Daten über Millionen von Objekten aus Tausenden von verschiedenen Projekten enthalten (Abb. 9.1-10). Dank der vektorisierten Pandas-Operationen ist es einfach, solche Informationsmengen mit hoher Geschwindigkeit zu filtern, zu gruppieren und zu aggregieren.



ID	Name	Category	Family Name	Height	BoundingBoxMin_X	BoundingBoxMin_Y	BoundingBoxMin_Z	Level
431144	Single-Flush	OST_Doors	Single-Flush	6.88976378	20.1503	-10.438	9.84252	Level 1
431198	Single-Flush	OST_Doors		6.88976378	13.2281	-1.1207	9.84252	Level 2
457479	Single Window	OST_Windows	Single Window	8.858267717	-11.434	-11.985	9.80971	Level 2
485432	Single Window	OST_Windows	Single Window	8.858267717	-11.434	4.25986	9.80971	Level 2
490150	Single-Flush	OST_Doors	Single-Flush	6.88976378	-1.5748	-2.9565	-1E-16	Level 1
493697	Basic Wall	OST_Walls	Basic Wall		-38.15	20.1656	-4.9213	Level 1
497540	Basic Wall	OST_Walls	Basic Wall		-4.5212	-0.0708	9.84252	Level 1

Abbildung 3.4-6 Bauprojekt als DataFrame ist eine zweidimensionale Tabelle mit Elementen in Zeilen und Attributen in Spalten.

Nvidia schätzt, dass bereits heute bis zu 30 Prozent aller Rechenressourcen für die Verarbeitung strukturierter Daten - Dataframes - verwendet werden, und dieser Anteil nimmt weiter zu.

> *Die Datenverarbeitung ist das, was wahrscheinlich ein Drittel der weltweiten Computerarbeit in jedem Unternehmen ausmacht. Die Datenverarbeitung und die Daten der meisten Unternehmen liegen in DataFrame im Tabellenformat*
>
> - Jensen Huang, CEO von Nvidia [91]

Lassen Sie uns einige Schlüsselfunktionen von DataFrame in Pandas auflisten:

- Spalten: In DataFrame sind die Daten in Spalten organisiert, die jeweils einen eindeutigen Namen haben. Attributspalten können Daten verschiedener Typen enthalten, ähnlich wie Spalten in Datenbanken oder Spalten in Tabellen.
- **Pandas Series** ist eine eindimensionale Datenstruktur in Pandas, ähnlich einer Liste oder Spalte in einer Tabelle, wobei jeder Wert einem anderen Index entspricht

> Pandas Series verfügt über mehr als 400 Attribute und Methoden, die die Arbeit mit Daten unglaublich flexibel machen. Sie können eine der vierhundert verfügbaren Funktionen direkt auf eine Spalte anwenden, mathematische Operationen durchführen, Daten filtern, Werte ersetzen, mit Daten, Strings und mehr arbeiten. Darüber hinaus unterstützt Series vektorisierte Operationen, die die Verarbeitung großer Datensätze im Vergleich zu zyklischen Berechnungen erheblich beschleunigen. So können Sie beispielsweise problemlos alle Werte mit einer Zahl multiplizieren, fehlende Daten ersetzen oder komplexe Transformationen durchführen, ohne komplizierte Schleifen schreiben zu müssen.

- **Zeilen**: in DataFrame können mit eindeutigen Werten indiziert werden. Dieser Index ermöglicht es Ihnen, die Daten in bestimmten Zeilen schnell zu ändern und anzupassen.
- **Index**: Wenn Sie einen DataFrame erstellen, weist Pandas standardmäßig jeder Zeile einen Index von 0 bis N-1 zu (wobei N die Anzahl aller Zeilen im DataFrame ist). Der Index kann jedoch geändert werden, um spezielle Bezeichnungen wie Daten oder eindeutige Merkmale aufzunehmen.
- **Die Indizierung von** Zeilen in einem DataFrame bedeutet, dass jeder Zeile ein eindeutiger Name oder eine Bezeichnung zugewiesen wird, der/die als DataFrame-Index bezeichnet wird.
- Datentypen**:** DataFrame unterstützt eine Vielzahl von Datentypen, darunter: `int`, `float`, `bool`, `datetime64` und `obect` für Textdaten. Jede DataFrame-Spalte hat ihren eigenen Datentyp, der bestimmt, welche Operationen mit ihrem Inhalt durchgeführt werden können.
- **Datenoperationen**: DataFrame unterstützt eine breite Palette von Operationen zur Datenverarbeitung, einschließlich Aggregation (`groupby`), Merge (`merge` und `join`), Concatenation (`concat`), Split-Apply-Combine und viele andere Datentransformationstechniken.
- **Größenmanipulation**: DataFrame ermöglicht das Hinzufügen und Entfernen von Spalten und Zeilen, so dass es eine dynamische Struktur ist, die je nach den Bedürfnissen Ihrer Datenanalyse geändert werden kann.
- **Visualisierung von Daten:** Mithilfe integrierter Visualisierungstechniken oder über Schnittstellen

zu beliebten Datenvisualisierungsbibliotheken wie Matplotlib oder Seaborn kann DataFrame leicht in Diagramme und Tabellen umgewandelt werden, um Daten grafisch darzustellen.

- **Dateneingabe und -ausgabe:** Pandas bietet Funktionen zum Lesen, Importieren und Exportieren von Daten in verschiedene Dateiformate wie CSV, Excel, JSON, HTML und SQL, wodurch DataFrame zu einem zentralen Knotenpunkt für die Datensammlung und -verteilung werden kann.

Im Gegensatz zu CSV und XLSX bietet Pandas DataFrame mehr Flexibilität und Leistung bei der Arbeit mit Daten: Es kann große Mengen an Informationen im Arbeitsspeicher verarbeiten, unterstützt erweiterte Datentypen (einschließlich Datumsangaben, logische Werte und Zeitreihen) und bietet umfangreiche Funktionen zum Filtern, Aggregieren, Zusammenführen und Visualisieren von Daten. Während CSV keine Informationen über Datentypen und -strukturen speichert und XLSX oft mit Formatierungen überfrachtet ist und eine geringe Skalierbarkeit aufweist, bleibt DataFrame die optimale Wahl für schnelle Analysen, Prozessautomatisierung und Integration mit KI -Modellen (Abb. 3.4-7). In den folgenden Kapiteln wird jeder dieser Datenaspekte im Detail untersucht, außerdem werden in Teil 8 des Buches ähnliche Formate wie Parquet, Apache Orc, JSON, Feather, HDF5 und Data Warehouses im Detail besprochen (Abb. 8.1-2).

	XLSX	CSV	Pandas DataFrame
Storage	Tabular	Tabular	Tabular
Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
Compression	Built-in	None	None (in-memory)
Performance	Low	Medium	High (memory dependent)
Complexity	High (formatting, styles)	Low	Low
Data Type Support	Limited	Very limited	Extended
Scalability	Low	Low	Medium (memory limited)

Abb. 3.4-7 DataFrame ist die optimale Wahl für die Datenmanipulation mit hoher Leistung und erweiterter Datentypunterstützung.

Dank ihrer Flexibilität, Leistungsfähigkeit und Benutzerfreundlichkeit sind die Pandas-Bibliothek und das DataFrame-Format zum De-facto-Standard in der Python-Datenanalyse geworden. Sie sind ideal sowohl für die Erstellung einfacher Berichte als auch für den Aufbau komplexer analytischer Pipelines, insbesondere in Verbindung mit LLM-Modellen.

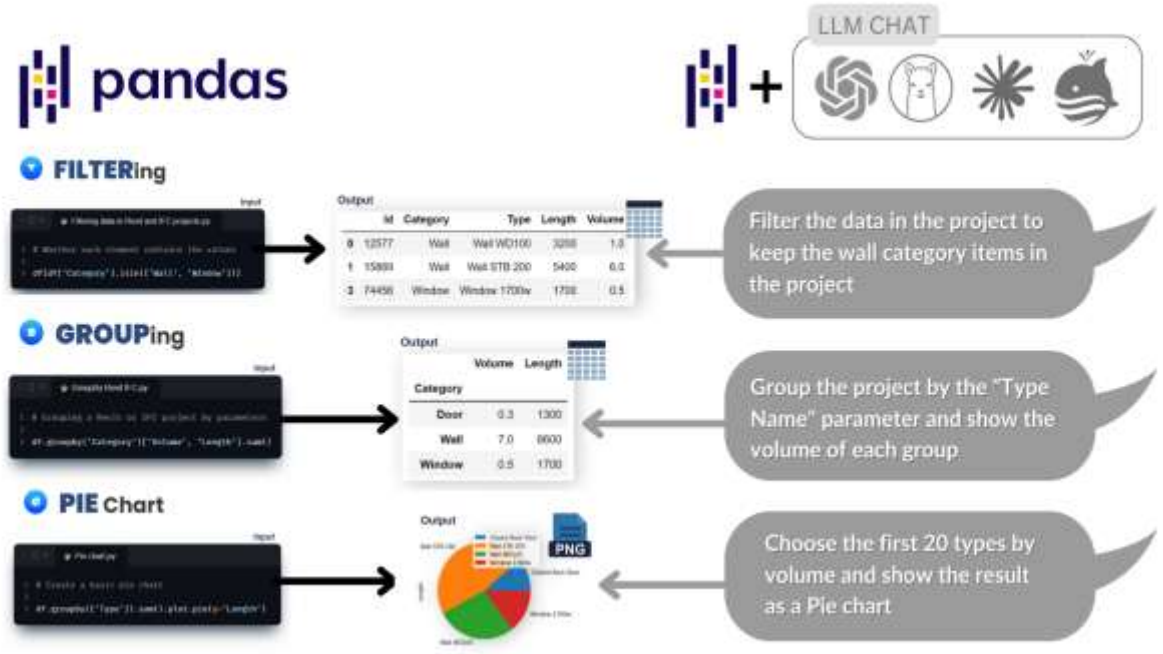


Abb. 3.4-8 LLMs vereinfachen die Interaktion mit Pandas: Anstelle von Code genügt eine Textabfrage.

Heute wird Pandas aktiv in LLM-basierten Chatrooms wie ChatGPT, LlaMa, DeepSeek, QWEN und anderen verwendet. Wenn ein Modell eine Abfrage zur Tabellenverarbeitung, Datenvalidierung oder Analyse erhält, generiert es in vielen Fällen Code, der genau die Pandas-Bibliothek verwendet. Dies macht DataFrame zu einer natürlichen "Sprache" für die Darstellung von Daten in Dialogen mit KI (Abb. 3.4-8).

Moderne Datentechnologien wie Pandas erleichtern die Analyse, Automatisierung und Integration von Daten in Geschäftsprozesse. Sie liefern schnell Ergebnisse, reduzieren die Arbeitsbelastung von Spezialisten und sorgen für wiederholbare Abläufe.

Nächste Schritte: Aufbau eines nachhaltigen Datenrahmens

In diesem Teil befassten wir uns mit den wichtigsten Datentypen, die in der Bauindustrie verwendet werden, lernten die verschiedenen Formate für ihre Speicherung kennen und analysierten die Rolle moderner Werkzeuge, einschließlich LLM und IDEs, bei der Verarbeitung von Informationen. Wir lernten, dass ein effektives Datenmanagement die Grundlage für eine fundierte Entscheidungsfindung und die Automatisierung von Geschäftsprozessen ist. Unternehmen, die in der Lage sind, ihre Daten zu strukturieren und zu organisieren, erlangen einen erheblichen Wettbewerbsvorteil in den Phasen der Datenverarbeitung und -umwandlung.

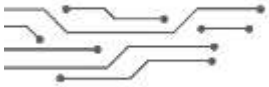
Um diesen Teil zusammenzufassen, lohnt es sich, die wichtigsten praktischen Schritte hervorzuheben, die Ihnen helfen werden, die besprochenen Ansätze bei Ihren täglichen Aufgaben anzuwenden:

- Führen Sie ein Datenaudit Ihrer Prozesse durch
  - Machen Sie eine Bestandsaufnahme aller Datentypen, die Sie in Ihren Projekten verwenden

  - ☐ Bestimmen Sie, welche Datentypen und -modelle für Ihre Geschäftsprozesse am wichtigsten sind.
  - ☐ Ermittlung von Problembereichen, in denen Informationen häufig unstrukturiert, schlecht strukturiert oder unzugänglich sind
- Beginn der Ausarbeitung einer Datenverwaltungsstrategie
  - ☐ politische Fragen und Standards für den Umgang mit verschiedenen Datentypen aufwerfen
  - ☐ Analysieren Sie, welche Ihrer Arbeitsabläufe durch die Umwandlung unstrukturierter Daten in strukturierte Daten verbessert werden können
  - ☐ Erstellen Sie eine Richtlinie zur Datenspeicherung und zum Datenzugriff, die Sicherheit und Vertraulichkeit berücksichtigt.
- Installation und Beherrschung grundlegender Tools für die Arbeit mit Daten
  - ☐ Wählen Sie eine geeignete IDE, die Ihren Aufgaben entspricht (z.B. VS Code oder Jupyter Notebook installieren)
  - ☐ Versuchen Sie, eine lokale LLM einzurichten, um Ihre persönlichen Daten vertraulich zu behandeln.
  - ☐ Experimentieren Sie mit der Pandas-Bibliothek, um XLSX-Tabellendaten zu verarbeiten
  - ☐ Beschreiben Sie dem LLM die typischen Aufgaben, die Sie in Tabellenkalkulationsprogrammen oder Datenbanken erledigen und bitten Sie den LLM, die Arbeit mit Pandas zu automatisieren

Die Anwendung dieser Schritte wird es Ihnen ermöglichen, Ihre Herangehensweise an die Arbeit mit Daten schrittweise zu verändern und von disparaten, unstrukturierten Datensätzen zu einem einheitlichen Ökosystem überzugehen, in dem Daten zu einem zugänglichen und verständlichen Gut werden. Fangen Sie klein an - erstellen Sie Ihren ersten DataFrame in Pandas, führen Sie einen lokalen LLM aus, automatisieren Sie Ihre erste Routineaufgabe mit Python (z. B. Excel-Tabellen).

Der vierte Teil des Buches befasst sich mit Datenqualität, Datenorganisation, -strukturierung und -modellierung. Wir werden uns auf Methoden konzentrieren, die disparate Datenquellen - von PDFs und Texten bis hin zu Bildern und CAD-Modellen - in strukturierte Datensätze umwandeln, die sich für Analysen und Automatisierung eignen. Wir werden auch untersuchen, wie Datenanforderungen formalisiert werden, wie konzeptionelle und logische Modelle in Konstruktionsprojekten aufgebaut werden und wie moderne Sprachmodelle (LLMs) bei diesem Prozess helfen können.



IV TEIL
DATENQUALITÄT: ORGANISATION, STRUKTURIERUNG, MODELLIERUNG

Der vierte Teil befasst sich mit Methoden und Technologien zur Umwandlung disparater Informationen in strukturierte Datensätze von hoher Qualität. Die Prozesse zur Formulierung und Dokumentation von Datenanforderungen als Grundlage für eine effektive Informationsarchitektur in Bauprojekten werden ausführlich diskutiert. Es werden praktische Methoden zur Extraktion strukturierter Informationen aus verschiedenen Quellen (PDF -Dokumente, Bilder, Textdateien, CAD -Modelle) mit Beispielen für die Umsetzung vorgestellt. Die Verwendung von regulären Ausdrücken (RegEx) und anderen Werkzeugen zur automatischen Validierung und Überprüfung von Daten wird analysiert. Der Prozess der Datenmodellierung auf konzeptioneller, logischer und physischer Ebene wird Schritt für Schritt beschrieben, wobei die Besonderheiten des Baugewerbes berücksichtigt werden. Konkrete Beispiele für den Einsatz von Sprachmodellen (LLM) zur Automatisierung der Prozesse der Strukturierung und Validierung von Informationen werden aufgezeigt. Es werden wirksame Ansätze zur Visualisierung von Analyseergebnissen vorgeschlagen, die die Verfügbarkeit von analytischen Informationen für alle Ebenen des Bauprojektmanagements erhöhen

KAPITEL 4.1.
UMWANDLUNG VON DATEN IN EINE STRUKTURIERTE FORM

Im Zeitalter der datengesteuerten Wirtschaft werden Daten zur Grundlage der Entscheidungsfindung und nicht mehr zum Hindernis. Anstatt Informationen ständig an jedes neue System und seine Formate anzupassen, versuchen Unternehmen zunehmend, ein einziges strukturiertes Datenmodell zu bilden, das als universelle Quelle der Wahrheit für alle Prozesse dient. Moderne Informationssysteme sind nicht auf Formate und Schnittstellen ausgelegt, sondern auf die Bedeutung der Daten - denn die Struktur mag sich ändern, aber die Bedeutung der Informationen bleibt viel länger gleich.

> Der Schlüssel zur effektiven Arbeit mit Daten liegt nicht darin, sie endlos zu konvertieren und umzuwandeln, sondern sie von Anfang an richtig zu organisieren: eine universelle Struktur zu schaffen, die Transparenz, Automatisierung und Integration in allen Phasen des Projektlebenszyklus ermöglicht.

Der herkömmliche Ansatz zwingt bei jeder neuen Plattformimplementierung zu manuellen Anpassungen: Daten migrieren, Attributnamen ändern und Formate anpassen. Diese Schritte verbessern nicht die Qualität der Daten selbst, sondern verschleiern nur Probleme und schaffen einen Teufelskreis aus endlosen Transformationen. Infolgedessen werden Unternehmen von bestimmten Softwarelösungen abhängig und die digitale Transformation verlangsamt sich.

In den folgenden Kapiteln befassen wir uns mit der richtigen Strukturierung von Daten und der Erstellung universeller Modelle, der Minimierung der Plattformabhängigkeit und der Konzentration auf das Wesentliche - Daten als strategische Ressource, um die herum nachhaltige Prozesse aufgebaut werden.

Lernen, wie man Dokumente, PDF, Bilder und Texte in strukturierte Formate umwandelt

Bei Bauprojekten liegt die überwiegende Mehrheit der Informationen in unstrukturierter Form vor: technische Dokumente, Leistungsverzeichnisse, Zeichnungen, Spezifikationen, Zeitpläne, Protokolle. Ihre Vielfalt - sowohl im Format als auch im Inhalt - macht die Integration und Automatisierung schwierig.

> Der Prozess der Konvertierung in strukturierte oder halbstrukturierte Formate kann je nach Art der Eingabedaten und den gewünschten Verarbeitungsergebnissen variieren.

Die Umwandlung von Daten aus unstrukturierter in strukturierte Form ist sowohl eine Kunst als auch eine Wissenschaft. Dieser Prozess variiert je nach Art der Eingabedaten und dem Zweck der Analyse und nimmt oft einen großen Teil der Arbeit des Dateningenieurs (Abb. 3.2-5) und des Analysten in Anspruch, mit dem Ziel, einen sauberen, organisierten Datensatz zu erzeugen.

Abb. 4.1-1 Umwandlung eines unstrukturierten gescannten Dokuments in ein strukturiertes Tabellenformat.

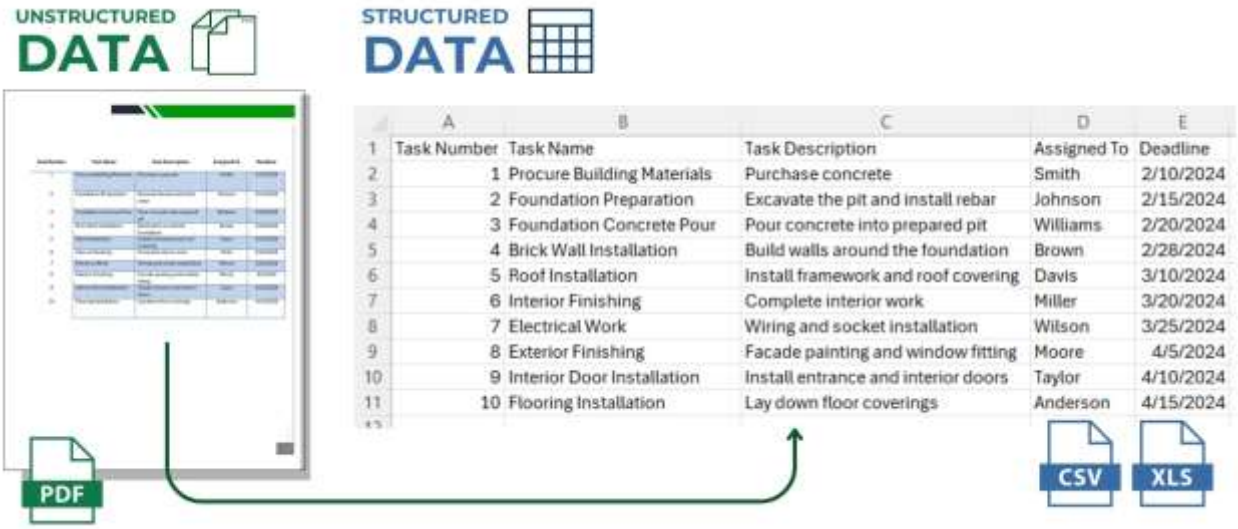
Die Umwandlung von Dokumenten, PDF, Bildern und Texten in ein strukturiertes Format (Abb. 4.1-1) ist ein schrittweiser Prozess, der die folgenden Schritte umfasst:

- Extrahieren): In diesem Schritt wird ein Quelldokument oder ein Bild mit unstrukturierten Daten geladen. Dabei kann es sich zum Beispiel um ein PDF -Dokument, ein Foto, eine Zeichnung oder einen Schaltplan handeln.
- **Datentransformation** (Transform): Es folgt der Schritt der Umwandlung unstrukturierter Daten in ein strukturiertes Format. Dies kann zum Beispiel die Erkennung und Interpretation von Text aus Bildern mittels optischer Zeichenerkennung (OCR) oder anderer Verarbeitungsmethoden beinhalten.
- **Laden und Speichern von Daten** (Load): Der letzte Schritt besteht darin, die verarbeiteten Daten in verschiedenen Formaten wie CSV, XLSX, XML, JSON für die weitere Arbeit zu speichern, wobei die Wahl des Formats von spezifischen Anforderungen und Präferenzen abhängt.

Dieser als ETL (Extract, Transform, Load) bezeichnete Prozess spielt eine Schlüsselrolle bei der automatisierten Datenverarbeitung und wird im Kapitel "ETL und Pipeline: Extract, Transform, Load" näher erläutert. Als nächstes werden wir uns Beispiele dafür ansehen, wie Dokumente unterschiedlicher Formate in strukturierte Daten umgewandelt werden.

Beispiel für die Umwandlung eines PDF -Dokuments in eine Tabelle

Eine der häufigsten Aufgaben bei Bauprojekten ist die Verarbeitung von Spezifikationen im PDF-Format. Um den Übergang von unstrukturierten Daten in ein strukturiertes Format zu demonstrieren, betrachten wir ein praktisches Beispiel: Extrahieren einer Tabelle aus einem PDF-Dokument und Konvertieren in das CSV- oder Excel-Format (Abb. 4.1-2).



	A	B	C	D	E
1	Task Number	Task Name	Task Description	Assigned To	Deadline
2	1	Procure Building Materials	Purchase concrete	Smith	2/10/2024
3	2	Foundation Preparation	Excavate the pit and install rebar	Johnson	2/15/2024
4	3	Foundation Concrete Pour	Pour concrete into prepared pit	Williams	2/20/2024
5	4	Brick Wall Installation	Build walls around the foundation	Brown	2/28/2024
6	5	Roof Installation	Install framework and roof covering	Davis	3/10/2024
7	6	Interior Finishing	Complete interior work	Miller	3/20/2024
8	7	Electrical Work	Wiring and socket installation	Wilson	3/25/2024
9	8	Exterior Finishing	Facade painting and window fitting	Moore	4/5/2024
10	9	Interior Door Installation	Install entrance and interior doors	Taylor	4/10/2024
11	10	Flooring Installation	Lay down floor coverings	Anderson	4/15/2024

Abb. 4.1-2 Im Gegensatz zu PDF sind die Formate CSV und XLSX weit verbreitet und lassen sich leicht in verschiedene Datenmanagementsysteme integrieren.

LLM-Sprachmodelle, wie ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN vereinfachen die Arbeit von Datenwissenschaftlern mit Daten erheblich, da sie den Bedarf an tiefem Lernen von Programmiersprachen verringern und es ermöglichen, viele Aufgaben mit Textabfragen zu lösen.

Anstatt Zeit damit zu verbringen, im Internet nach Lösungen zu suchen (in der Regel auf der StackOverFlow-Website oder in thematischen Foren und Chats) oder sich an Datenverarbeitungsspezialisten zu wenden, können wir die Möglichkeiten moderner Online- oder lokaler LLMs nutzen. Es genügt, eine Anfrage zu stellen, und das Modell stellt einen fertigen Code für die Konvertierung von PDF -Dokumenten in ein Tabellenformat bereit.

- Senden Sie die folgende Textanfrage an ein beliebiges LLM -Modell (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN oder ein anderes):

> Bitte schreiben Sie einen Code, um Text aus einer PDF -Datei zu extrahieren, die eine Tabelle enthält. Der Code sollte den Dateipfad als Argument nehmen und die extrahierte Tabelle als DataFrame zurückgeben⏎

- Die Antwort des LLM -Modells wird in den meisten Fällen in Form von Code in Python erfolgen, da diese Sprache für die Datenverarbeitung, Automatisierung und die Arbeit mit verschiedenen Dateiformaten weit verbreitet ist:

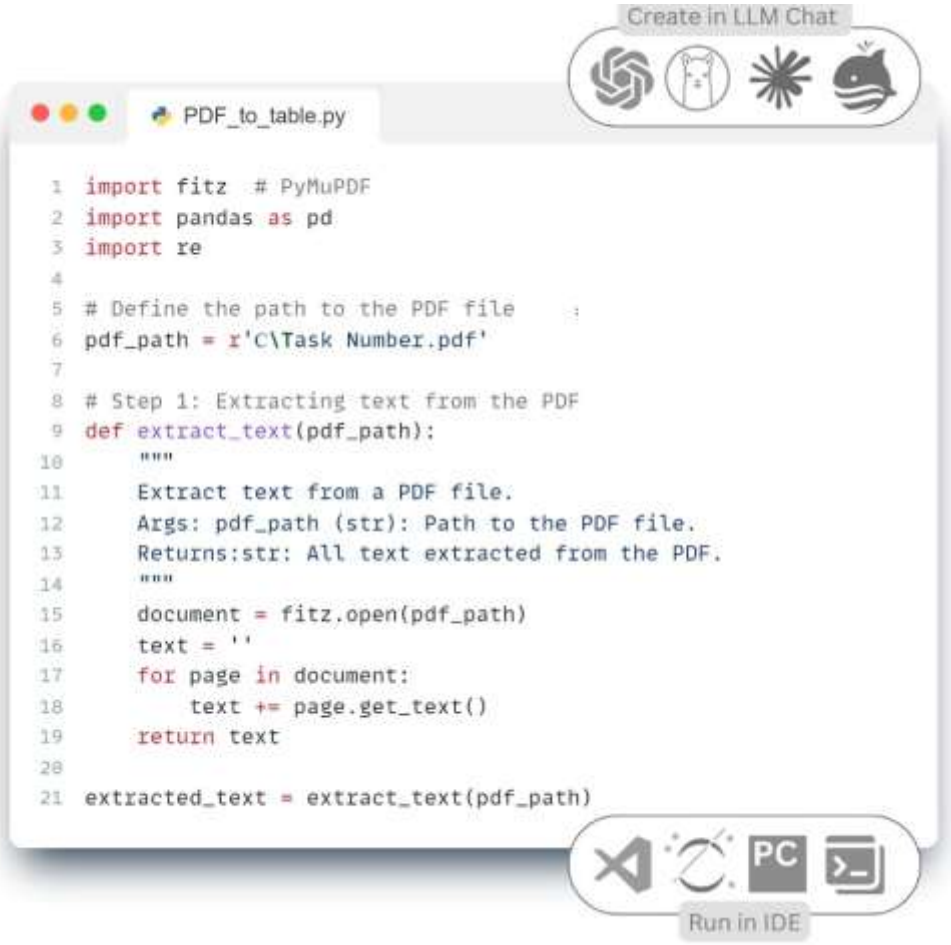


Abb. 4.1-3 Die LLM-Antwort in Form des Python-Codes und seiner Bibliotheken und Pakete (Pandas, Fitz) extrahiert Text aus einer PDF -Datei.

Dieser Code (Abb. 4.1-3) kann offline in einer der oben genannten gängigen IDEs ausgeführt werden: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse mit PyDev-Plugin, Thonny, Wing IDE, IntelliJ IDEA mit Python-Plugin, JupyterLab oder beliebte Online-Tools: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Im Schritt "Konvertieren" verwenden wir die beliebte Pandas-Bibliothek (die wir im Kapitel "Python Pandas: ein unverzichtbares Werkzeug für die Arbeit mit Daten" ausführlich besprochen haben), um den extrahierten Text in den DataFrame einzulesen und den DataFrame in einer CSV-Tabellendatei oder XLXS zu speichern:

> Ich benötige Code, der die resultierende Tabelle aus einer PDF -Datei in einen DataFrame konvertiert. Fügen Sie außerdem Code zum Speichern des DataFrame in einer CSV-Datei hinzu. ⏎

Die Antwort von LLM:

Create in LLM Chat

PDF_to_table.py

```python
# Step 2: Converting the extracted text to a DataFrame
def text_to_df(text):
    """
    Convert extracted text to a pandas DataFrame.
    Args: text (str): Extracted text from a PDF file.
    Returns: pandas.DataFrame: DataFrame representation of the text.
    """
    lines = text.split('\n')
    data = []
    for line in lines:
        if line.strip():  # Check if line is not empty
            data.append(re.split(r'\s{2,}', line))

    # Determine the number of columns based on the longest row
    num_columns = max(len(row) for row in data)

    # Create DataFrame with dynamic columns
    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
    return df

df = text_to_df(extracted_text)

# Save the DataFrame to a CSV file
csv_path = 'C:\extracted_table.csv'
df.to_csv(csv_path, index=False)
```

Run in IDE

Abb. 4.1-4 Konvertieren der extrahierten Tabelle von PDF zu DataFrame und Speichern der Tabelle in einer CSV -Datei.

Tritt während der Codeausführung ein Fehler auf (Abb. 4.1-3, Abb. 4.1-4) - z.B. wegen fehlender Bibliotheken oder eines falschen Dateipfades - kann der Fehlertext einfach zusammen mit dem Quellcode kopiert und dem LLM -Modell erneut vorgelegt werden. Das Modell analysiert die Fehlermeldung, erklärt, was das Problem ist, und schlägt Korrekturen oder zusätzliche Schritte vor.

> Auf diese Weise wird die Interaktion mit der KI LLM zu einem vollständigen Zyklus →Test→ Feedback→ Korrektur - ohne die Notwendigkeit tiefer technischer Kenntnisse.

Mit einer einfachen Textabfrage in LLM chat und einem Dutzend Zeilen Python, die wir lokal in jeder IDE ausführen können, haben wir ein PDF-Dokument in ein tabellarisches CSV-Format konvertiert, das im Gegensatz zu einem PDF-Dokument leicht maschinenlesbar ist und schnell in jedes Datenverwaltungssystem integriert werden kann.

Wir können diesen Code (Abb. 4.1-3, Abbildung 4.1-4), indem wir ihn aus einem beliebigen LLM-Chatroom kopieren, auf Dutzende oder Tausende neuer PDF-Dokumente auf dem Server anwenden und so den Prozess der Umwandlung eines Stroms unstrukturierter Dokumente in ein strukturiertes CSV-Tabellenformat automatisieren.

Doch PDF-Dokumente enthalten nicht immer Text, sondern sind meist gescannte Dokumente, die als Bilder verarbeitet werden müssen. Obwohl Bilder von Natur aus unstrukturiert sind, ermöglicht uns die Entwicklung und Anwendung von Erkennungsbibliotheken, ihren Inhalt zu extrahieren, zu verarbeiten und zu analysieren, so dass wir diese Daten in unseren Geschäftsprozessen voll nutzen können.

Konvertierung von JPEG- und PNG-Bildern in eine strukturierte Form

Bilder sind eine der häufigsten Formen von unstrukturierten Daten. Im Baugewerbe und in vielen anderen Branchen wird eine riesige Menge an Informationen in Form von gescannten Dokumenten, Schemata, Fotos und Zeichnungen gespeichert. Solche Daten enthalten wertvolle Informationen, können aber nicht direkt verarbeitet werden, wie z. B. in einer Excel-Tabelle oder einer Datenbank. Bilder enthalten eine Menge komplexer Informationen, da ihr Inhalt, ihre Farben und Texturen sehr unterschiedlich sind und eine spezielle Verarbeitung erforderlich ist, um nützliche Informationen zu extrahieren.

Die Schwierigkeit bei der Verwendung von Bildern als Datenquelle liegt in der fehlenden Struktur. Bilder vermitteln keine Bedeutung in einer direkten, leicht quantifizierbaren Art und Weise, die ein Computer sofort verstehen oder verarbeiten kann, wie es eine Excel-Tabelle oder eine Datenbanktabelle tut. Um unstrukturierte Bilddaten in eine strukturierte Form umzuwandeln, müssen spezielle Bibliotheken verwendet werden, die in der Lage sind, die in ihnen enthaltenen visuellen Informationen zu interpretieren (Abb. 4.1-5).

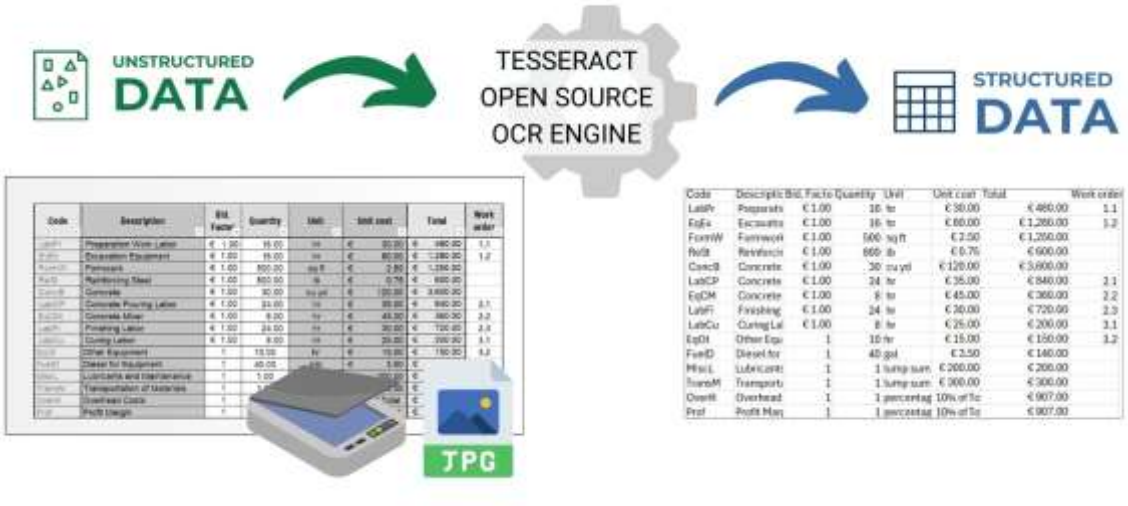


Abb. 4.1-5 Die Umwandlung gescannter Dokumente und Bilder in strukturierte Formate ist mit speziellen OCR-Tools möglich.

Die OCR (Optical Character Recognition) Technologie wird verwendet, um Text aus Bildern zu extrahieren. Sie ermöglicht es, Buchstaben und Zahlen in gescannten Dokumenten, Fotos und PDF -Dateien zu erkennen und sie in bearbeitbaren und maschinenlesbaren Text umzuwandeln. OCR-Technologien werden schon seit langem in der Dokumentenautomatisierung eingesetzt und lassen sich heute problemlos in beliebige Geschäftsprozesse und Python -Anwendungen integrieren. Eines der beliebtesten OCR-Tools ist Tesseract, ein Open-Source-Programm, das ursprünglich von HP™ entwickelt wurde und jetzt von Google™ unterstützt wird. Es unterstützt über 100 Sprachen und hat eine hohe Erkennungsgenauigkeit.

Bitten wir den LLM chat, einen Beispielcode zu schreiben, um Daten aus einer gescannten oder fotografierten Tabelle auf strukturierte Weise abzurufen.

- Senden Sie eine Textanfrage an LLM chat (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN oder andere):

> Schreiben Sie Code, um ein JPEG-Bild, das eine Tabelle enthält, in eine DataFrame-Tabelle zu konvertieren ⏎

Die Antwort von LLM wird in den meisten Fällen die Verwendung der Pytesseract-Bibliothek zur Texterkennung in Bildern vorschlagen:

Create in LLM Chat

group.py

```
1  import pytesseract
2  from PIL import Image
3  import pandas as pd
4
5  # Load the image from file
6  image_path = '/mnt/data/ConstructiontDDC,
7  image = Image.open(image_path)
8
9  # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head()  # Show only the first few rows for brevity
22
```

Run in IDE

Abb. 4.1-6 Konvertieren Sie den aus einer Tabelle mit Bildern oder Fotos extrahierten Text in eine strukturierte tabellarische Darstellung.

In diesem Beispiel - der aus dem LLM abgeleitete Code (Abb. 4.1-6) verwendet die pytesseract-Bibliothek (Tesseract für Python), um ein Bild mittels OCR (optische Zeichenerkennung) in Text umzuwandeln, und die Pandas-Bibliothek, um diesen Text in eine strukturierte Form, d. h. einen DataFrame, umzuwandeln.

> Der Konvertierungsprozess umfasst in der Regel eine Vorverarbeitung zur Verbesserung der Bildqualität, nach der verschiedene Algorithmen zur Mustererkennung, Merkmalsextraktion oder Objekterkennung angewendet werden. Als Ergebnis werden unstrukturierte visuelle Informationen in strukturierte Daten umgewandelt.

Obwohl PDF und Bilder die wichtigsten Quellen für unstrukturierte Informationen sind, ist der wahre Champion, was das Volumen angeht, der in E-Mails, Chats, Meetings und Messenger generierte Text. Diese Daten sind nicht nur zahlreich, sondern auch verstreut, informalisiert und extrem schlecht strukturiert.

Umwandlung von Textdaten in eine strukturierte Form

Neben PDF-Dokumenten mit Tabellen (Abb. 4.1-2) und gescannten Versionen von Tabellenformularen (Abb. 4.1-5) wird ein großer Teil der Informationen in der Projektdokumentation in Textform dargestellt. Dabei kann es sich sowohl um zusammenhängende Sätze in Textdokumenten als auch um fragmentarische Aufzeichnungen handeln, die über Zeichnungen und Pläne verstreut sind. Unter den Bedingungen der modernen Datenverarbeitung besteht eine der häufigsten Aufgaben darin, solche Texte in ein strukturiertes Format umzuwandeln, das für die Analyse, Visualisierung und Entscheidungsfindung geeignet ist.

Im Mittelpunkt dieses Prozesses steht die Taxonomie, ein Klassifizierungssystem, mit dem Informationen auf der Grundlage gemeinsamer Merkmale in Kategorien und Unterkategorien eingeteilt werden können.

> **Eine Taxonomie** ist eine hierarchische Klassifikationsstruktur, die zur Gruppierung und Organisation von Objekten dient. Im Zusammenhang mit der Textverarbeitung dient sie als Grundlage für die systematische Einteilung von Objekten in semantische Kategorien, wodurch die Analyse vereinfacht und die Qualität der Datenverarbeitung verbessert wird.

Die Erstellung einer Taxonomie wird von den Schritten der Entitätsextraktion, Kategorisierung und Kontextualisierung begleitet. Um den Prozess der Extraktion von Informationen aus Textdaten zu modellieren, sind die folgenden Schritte denen ähnlich, die wir bereits bei der Strukturierung von Daten aus PDF-Dokumenten angewendet haben:

- Extrahieren): Sie müssen Textdaten analysieren, um Informationen über Verzögerungen und Änderungen im Projektzeitplan zu extrahieren.
- **Kategorisierung und Klassifizierung** (Transofrm): Kategorisierung der erhaltenen Informationen, z. B. Gründe für Verspätungen und Fahrplanänderungen.
- **Integration** (Load): Am Ende bereiten wir strukturierte Daten für die Integration in externe Datenverwaltungssysteme vor.

Stellen Sie sich eine Situation vor: Wir haben einen Dialog zwischen einem Projektleiter und einem Ingenieur, die Probleme mit dem Zeitplan diskutieren. Unser Ziel ist es, die Schlüsselelemente (Gründe für die Verzögerung, Anpassungen des Zeitplans) zu extrahieren und sie in strukturierter Form darzustellen (Abb. 4.1-7).

Wir führen die Extraktion anhand der erwarteten Schlüsselwörter durch, erstellen einen DataFrame, um die Datenextraktion zu simulieren, und nach der Transformation eine neue DataFrame-Tabelle, die Spalten für Datum, Ereignis (z. B. Grund für die Verzögerung) und Aktion (z. B. Zeitplanänderung) enthält.

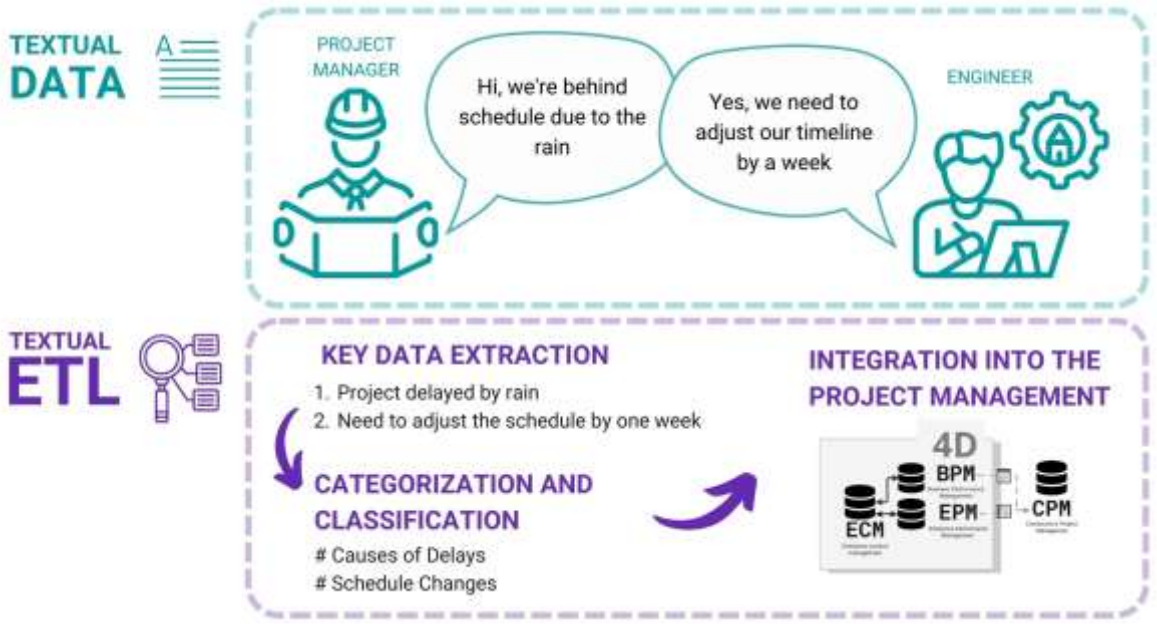


Abbildung 4.1-7 Hervorhebung von Schlüsselinformationen aus dem Text über die Notwendigkeit, Zeitpläne anzupassen und Änderungen in das Projektmanagementsystem zu integrieren.

Hier ist der Code zur Lösung des Problems mit einer Textabfrage in einem der Sprachmodelle wie in den vorherigen Beispielen.

- Senden Sie eine Textanfrage an einen beliebigen LLM Chat-Raum:

> Ich habe ein Gespräch zwischen einem Manager: "Hallo, wir sind wegen des Regens in Verzug" und einem Ingenieur: "Ja, wir müssen den Termin um eine Woche verschieben". Ich benötige ein Skript, das zukünftige ähnliche Textdialoge analysiert, daraus die Gründe für Verzögerungen und notwendige Terminanpassungen extrahiert und aus diesen Daten einen DataFrame generiert. Der DataFrame soll dann in einer CSV -Datei gespeichert werden. ⏎

- Die Antwort des LLM enthält in der Regel Python -Code mit regulären Ausdrücken (re - Regex) und die Pandas (pd) Bibliothek:

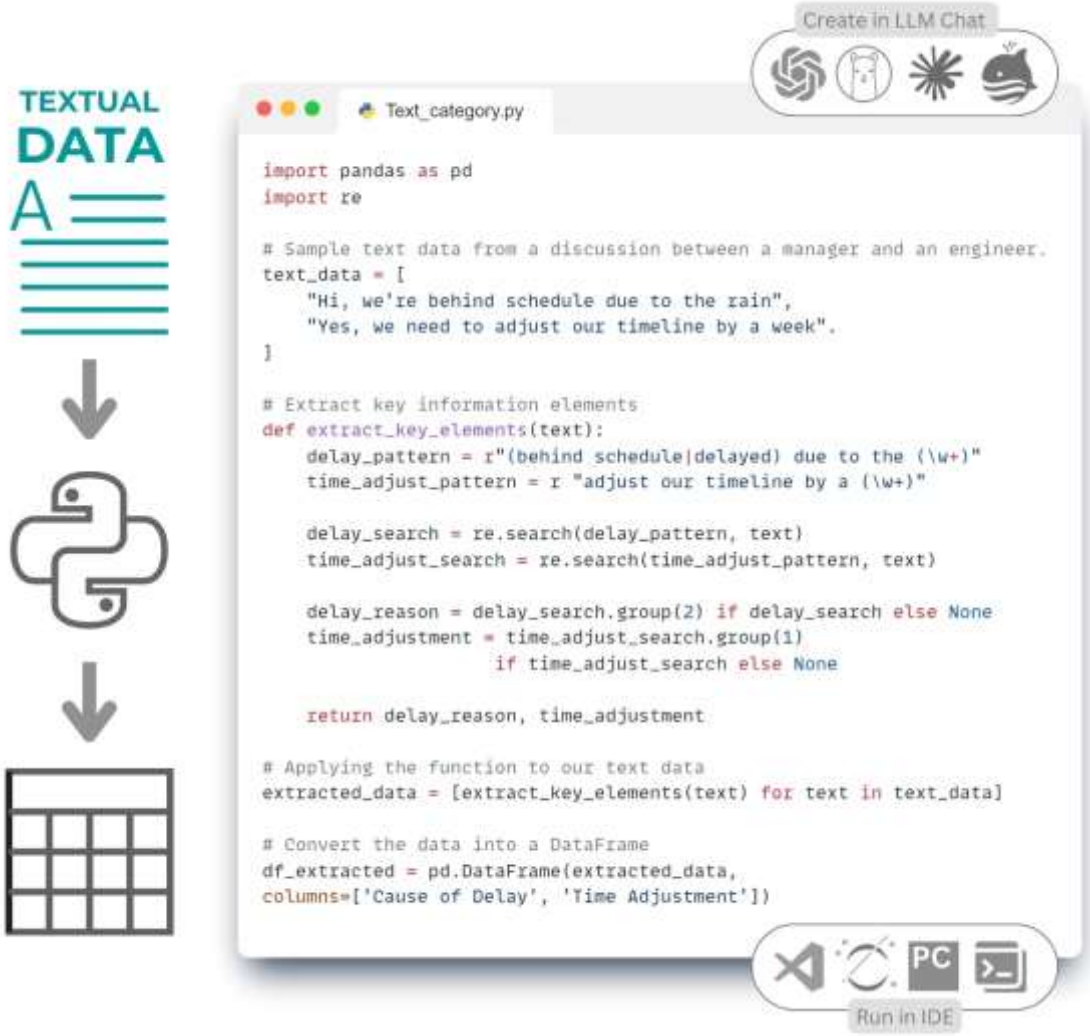
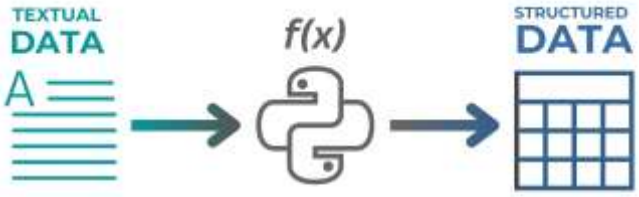


Abbildung 4.1-8 Hervorhebung von Schlüsselinformationen aus dem Text über die Notwendigkeit der Anpassung von Fristen in einer Tabelle.

In diesem Beispiel (Abb. 4.1-7) werden Textdaten, die die Korrespondenz zwischen einem Projektmanager und einem Ingenieur enthalten, analysiert, um spezifische Informationen zu identifizieren und zu extrahieren, die das Management zukünftiger Projekte mit ähnlichen Dialogen beeinflussen können. Mit Hilfe von regulären Ausdrücken (mehr über reguläre Ausdrücke erfahren Sie im Kapitel "Strukturierte Anforderungen und reguläre Ausdrücke RegEx ") werden die Ursachen für Projektverzögerungen und notwendige Anpassungen des Zeitplans anhand von Mustern ermittelt. Die in diesem Beispiel geschriebene Funktion extrahiert entweder die Ursache der Verzögerung oder die Zeitanpassung aus den Zeichenketten auf der Grundlage der Muster: Hervorhebung des Wortes nach "because of" als Ursache der Verzögerung oder des Wortes nach "by" als Zeitanpassung.

Wird in einer Zeile eine witterungsbedingte Verspätung erwähnt, wird "Regen" als Ursache angegeben; wird in einer Zeile eine Fahrplananpassung für einen bestimmten Zeitraum erwähnt, wird dieser Zeitraum als Zeitanpassung extrahiert (Abb. 4.1-9). Fehlt eines dieser Wörter in einer Zeile, so ergibt sich für die entsprechende Attributspalte der Wert "Keine".



Text string	Cause of Delay	Time Adjustment
0	rain	None
1	None	week

Abb. 4.1-9 Die nach der Codeausführung als DataFrame erhaltene Übersichtstabelle enthält Informationen über das Vorhandensein von Verzögerungen und notwendige Zeitanpassungen.

Die Strukturierung und Parametrisierung von Bedingungen aus dem Text (Dialog, Brief, Dokument) ermöglicht es, Verzögerungen am Bau schnell zu beseitigen: z.B. kann der Mangel an Arbeitern das Arbeitstempo bei schlechtem Wetter beeinträchtigen, so dass Unternehmen, die die Verzögerungsparameter aus den Dialogen (Abb. 4.1-9) zwischen dem Vorarbeiter auf der Baustelle und dem Projektleiter im Voraus kennen, die Mannschaft im Falle einer ungünstigen Prognose verstärken können.

Die Konvertierung von Dokumenten und Bildern in ein strukturiertes Format kann mit relativ einfachen, offenen und kostenlosen kategoriebasierten Tools erfolgen.

Die Kategorisierung von Elementen ist ebenfalls ein wichtiger Bestandteil der Arbeit mit Projektdaten, insbesondere im Zusammenhang mit der Verwendung von CAD-Software (BIM).

Umwandlung von CAD-Daten (BIM) in eine strukturierte Form

Die Strukturierung und Kategorisierung von CAD-Daten (BIM) ist eine komplexere Aufgabe, da die in CAD-(BIM)-Datenbanken gespeicherten Daten fast immer in geschlossenen oder komplexen parametrischen Formaten vorliegen und häufig gleichzeitig geometrische Datenelemente (halbstrukturierte) und Metainformationselemente (halbstrukturierte oder strukturierte Daten) enthalten.

Native Datenformate in CAD (BIM) Systemen sind in der Regel geschützt und für die direkte Nutzung unzugänglich, es sei denn, es handelt sich um spezielle Software oder API - Schnittstellen des Entwicklers selbst (Abb. 4.1-10). Diese Isolierung von Daten bildet geschlossene Speichersilos, die den freien Austausch von Informationen einschränken und die Schaffung durchgängiger digitaler Prozesse im Unternehmen hemmen.

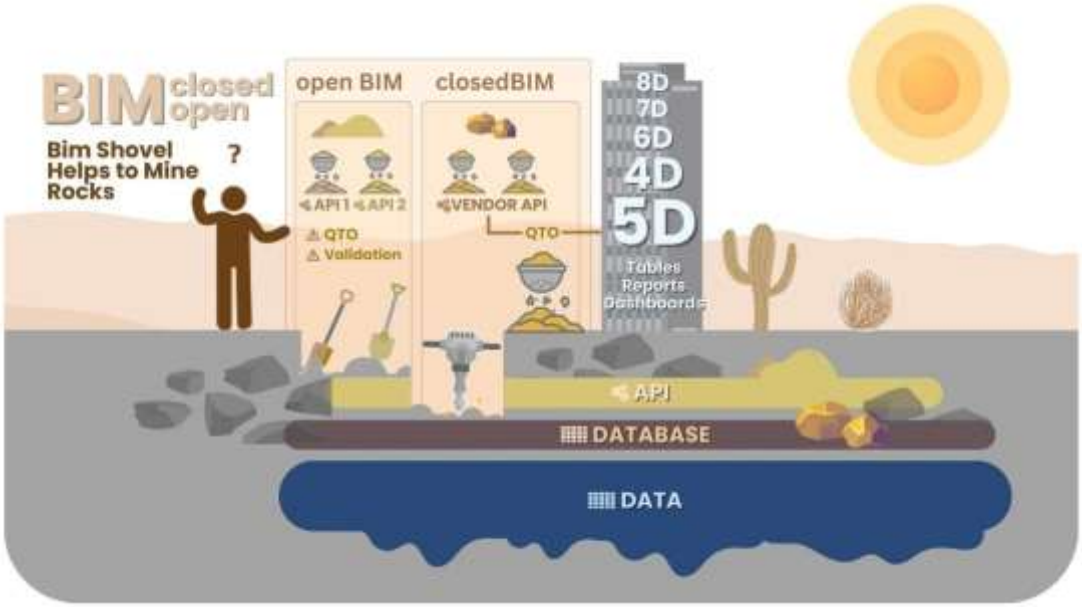


Abbildung 4.1-10 CAD-Fachleute (BIM) können über API -Verbindungen oder Anbieter-Tools auf native Daten zugreifen.

In speziellen CAD (BIM)-Formaten werden die Informationen über die Eigenschaften und Attribute der Projektelemente in einem hierarchischen Klassifizierungssystem gesammelt, in dem sich die Entitäten mit den entsprechenden Eigenschaften wie die Früchte eines Obstbaums in den jüngsten Knoten der Datenklassifizierungszweige befinden (Abb. 4.1-11).

Die Datenextraktion aus solchen Hierarchien ist auf zwei Arten möglich: entweder manuell, indem man auf jeden Knoten klickt, wie bei der Bearbeitung eines Baumes, und ausgewählte Äste von Kategorien und Typen mit einer Axt abschneidet. Alternativ dazu bietet die Verwendung von Anwendungsprogrammierschnittstellen (APIs) einen effizienteren, automatisierten Ansatz zum Abrufen und Gruppieren von Daten, um sie schließlich in eine strukturierte Tabelle zur Verwendung in anderen Systemen umzuwandeln.

Verschiedene Tools wie Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC oder Open-Source-Lösungen, wie IfcOpSh oder IFCjs für das IFC-Format, können verwendet werden, um strukturierte Datentabellen aus CAD-Projekten (BIM) zu extrahieren.

Moderne Datenexport- und Konvertierungstools ermöglichen es, zur Vereinfachung der Datenverarbeitung und -aufbereitung den Inhalt von CAD-Modellen in zwei Hauptkomponenten zu trennen: Geometrieinformationen und Attributdaten (Abb. 4.1-13) - Metainformationen, die die Eigenschaften der Konstruktionselemente beschreiben (Abb. 3.1-16). Diese beiden Datenschichten sind durch eindeutige Bezeichner miteinander verbunden, wodurch es möglich ist, jedes Element mit Geometriebeschreibung (über Parameter oder Polygone) genau auf seine Attribute abzubilden: Name, Material, Fertigstellungsgrad, Kosten, usw. Dieser Ansatz gewährleistet die Integrität des Modells und ermöglicht eine flexible Nutzung der Daten sowohl für die Visualisierung (geometrische Modelldaten) als auch für Analyse- oder Verwaltungsaufgaben (strukturiert oder lose strukturiert), wobei mit beiden Datenarten getrennt oder parallel gearbeitet werden kann.

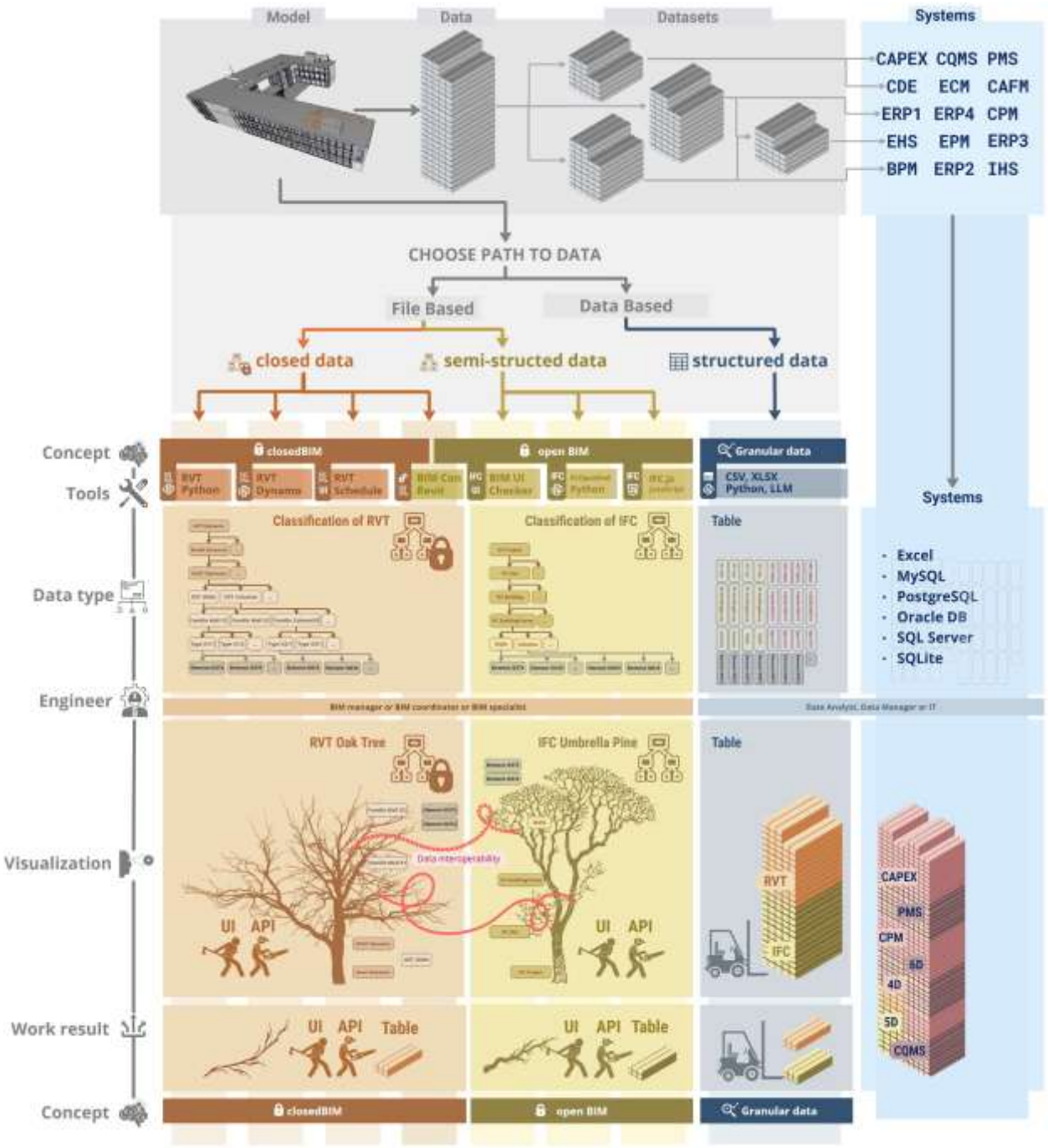


Abb. 4.1-11 Die Sicht auf Informationen aus CAD-Datenbanken (BIM) wird dem Nutzer in Form von Klassifikationsbäumen präsentiert.

Mit der Entwicklung von Reverse-Engineering-Technologien und dem Aufkommen von SDKs (Software Development Kit) für die Konvertierung von CAD-Daten ist die Verfügbarkeit und Konvertierung von Daten aus geschlossenen CAD-Softwareformaten (BIM) viel einfacher geworden. Es ist nun möglich, Daten aus geschlossenen Formaten legal und sicher in universelle Formate zu konvertieren, die für die Analyse und Verwendung in anderen Systemen geeignet sind. Die Geschichte der ersten Reverse-Engineering-Tools ("Open

DWG") und der Kampf um die Vorherrschaft über die Formate der CAD-Anbieter wurde im Kapitel "Strukturierte Daten: die Grundlage der digitalen Transformation" behandelt.

> Reverse-Engineering-Tools ermöglichen den rechtmäßigen Abruf von Daten aus geschlossenen proprietären Formaten, indem sie Informationen aus dem gemischten CAD (BIM)-Format in die vom Benutzer benötigten Datentypen und -formate aufschlüsseln, so dass sie leichter verarbeitet und analysiert werden können.

Mit Hilfe von Reverse Engineering und direktem Zugriff auf Informationen aus CAD-Datenbanken macht Informationen zugänglich und ermöglicht offene Daten und offene Werkzeuge sowie die Datenanalyse mit Standardwerkzeugen, die Erstellung von Berichten, Visualisierungen und die Integration mit anderen digitalen Systemen (Abb. 4.1-12).

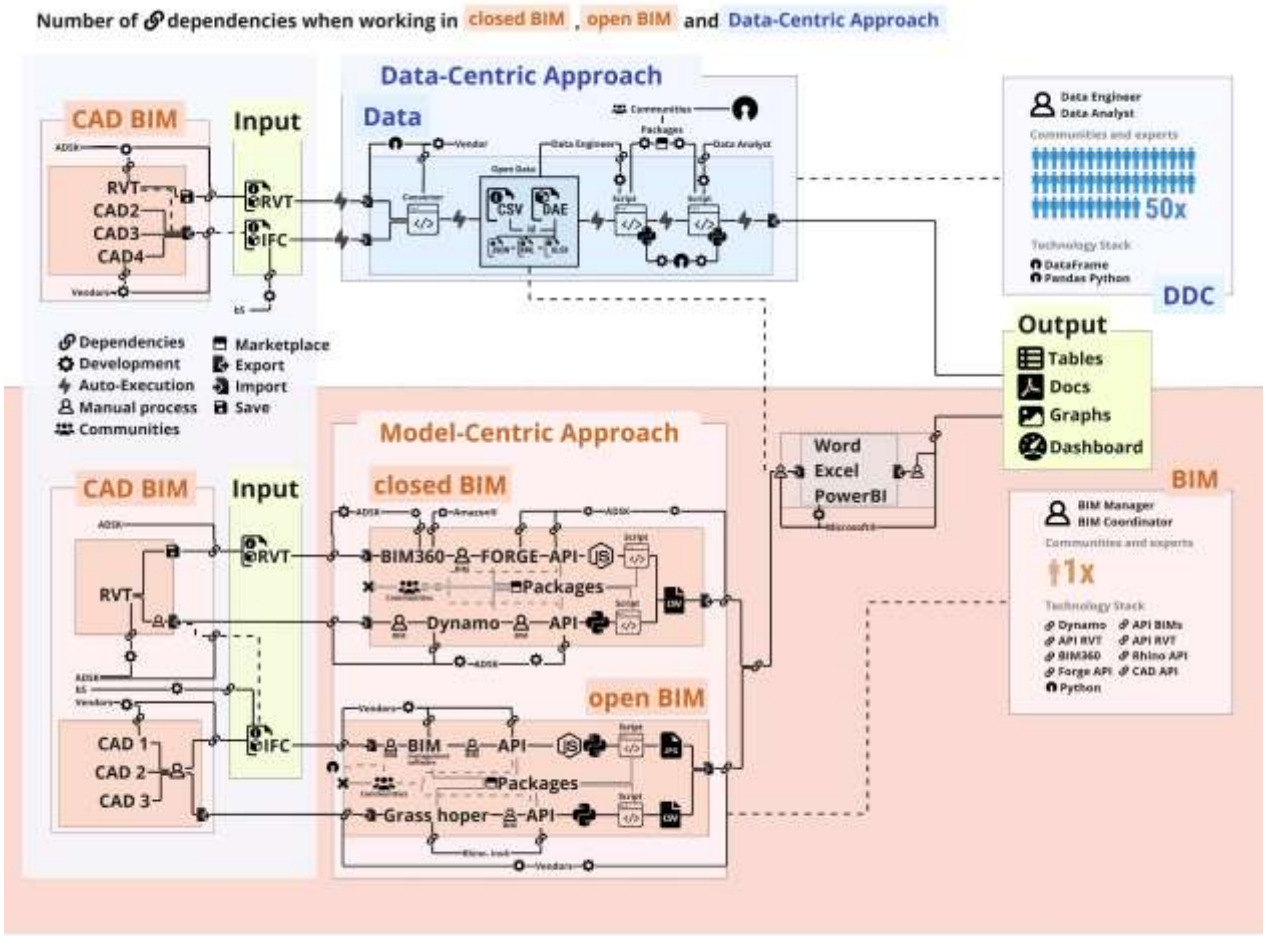


Abb. 4.1-12 Der direkte Zugriff auf CAD-Daten minimiert die Abhängigkeiten von Softwareplattformen und führt zu einem datenzentrierten Ansatz.

Seit 1996 für das DWG-Format, seit 2008 für das DGN-Format und seit 2018 für RVT ist es möglich, ursprünglich geschlossene CAD-Datenformate mit Hilfe von Reverse-Engineering-Werkzeugen bequem und effizient in beliebige andere Formate, einschließlich strukturierter Formate, zu konvertieren (Abb. 4.1-13).

Heute verwenden fast alle großen CAD (BIM) und großen Ingenieurbüros in der Welt SDKs - Reverse-Engineering-Tools, um Daten aus geschlossenen CAD (BIM) Herstellerformaten zu extrahieren [92].

Abb. 4.1-13 Der Einsatz von Reverse-Engineering-Werkzeugen ermöglicht die Konvertierung von CAD (BIM) Programmdatenbanken in beliebige Datenmodelle.

Die Konvertierung von Daten aus geschlossenen, proprietären Formaten in offene Formate und die Trennung von gemischten CAD (BIM)-Formaten in Geometrie- und Metainformations-Attributdaten vereinfacht die Arbeit mit den Daten und macht sie für die Analyse, Manipulation und Integration mit anderen Systemen verfügbar (Abb. 4.1-14).

> Bei der heutigen Arbeit mit CAD-Daten (BIM) sind wir an einem Punkt angelangt, an dem wir nicht mehr die Erlaubnis von CAD- (BIM-) Anbietern einholen müssen, um auf Informationen aus CAD-Formaten zuzugreifen.

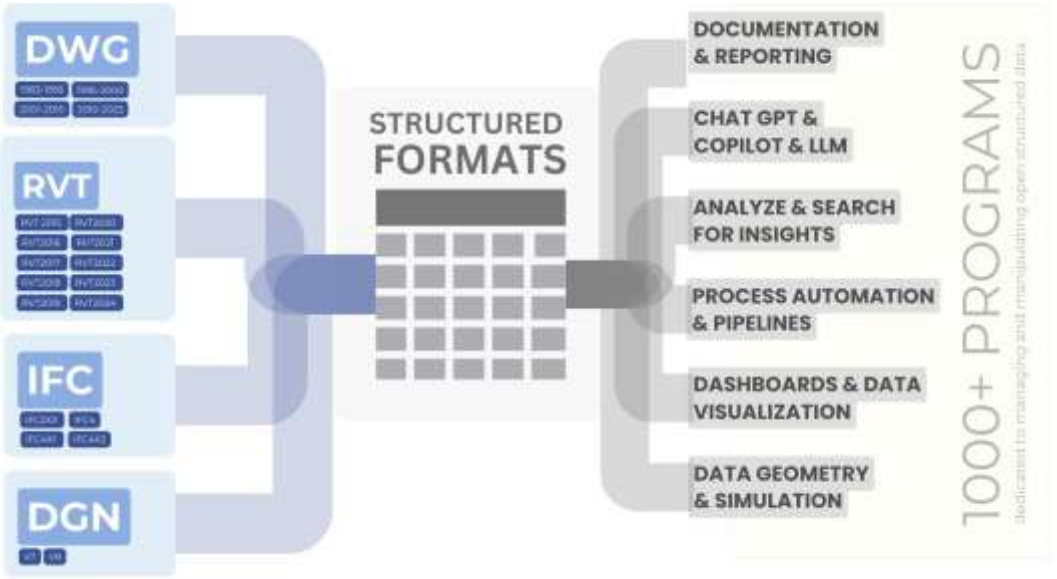


Abb. 4.1-14 Moderne SDK-Tools ermöglichen die legale Konvertierung von Daten aus proprietären CAD-Datenbankformaten (BIM).

Die aktuellen Trends in der Verarbeitung von CAD-Konstruktionsdaten werden weiterhin von den wichtigsten Marktteilnehmern - CAD - geprägt, die daran arbeiten, ihre Position in der Datenwelt zu stärken und neue

Formate und Konzepte zu schaffen.

Die Anbieter von CAD Lösungen bewegen sich in Richtung strukturierter Daten

Ab 2024 vollzieht sich in der Planungs- und Baubranche ein bedeutender technologischer Wandel bei der Nutzung und Verarbeitung von Daten. Anstelle des freien Zugangs zu Konstruktionsdaten konzentrieren sich die Anbieter von CAD -Systemen auf die Förderung der nächsten neuen Konzepte. Ansätze wie BIM (seit 2002) und Open BIM (seit 2012) weichen allmählich modernen technologischen Lösungen, die CAD-Anbieter zu fördern beginnen [93]:

- Umstellung auf die Verwendung "granularer" Daten, die ein effizientes Informationsmanagement und einen Übergang zur Datenanalyse ermöglichen
- Entstehung des Formats USD und Einführung des Entity-Component-System-Ansatzes (ECS) für eine flexible Datenorganisation
- Aktive Nutzung der künstlichen Intelligenz in der Datenverarbeitung, Prozessautomatisierung und Datenanalyse
- Entwicklung der Interoperabilität - verbesserte Interaktion zwischen verschiedenen Programmen, Systemen und Datenbanken

Jeder dieser Aspekte wird im sechsten Teil des Buches "CAD und BIM: Marketing, Realität und die Zukunft der Planungsdaten im Bauwesen" ausführlicher behandelt. In diesem Kapitel wollen wir nur kurz den allgemeinen Vektor des Wandels skizzieren: Die großen CAD-Anbieter versuchen nun, die Art und Weise, wie Planungsinformationen strukturiert werden, zu überdenken. Eine der wichtigsten Neuerungen ist die Abkehr vom klassischen dateibasierten Speichermodell zugunsten einer granularen, analytisch orientierten Datenarchitektur, die einen kontinuierlichen Zugriff auf einzelne Modellkomponenten ermöglicht [93].

> Im Wesentlichen geht es darum, dass die Industrie umständliche, spezialisierte und parametrische Formate, die geometrische Kerne erfordern, zugunsten universellerer, maschinenlesbarer und flexiblerer Lösungen abschafft.

Ein solcher Treiber des Wandels ist das USD-Format (Universal Scene Description), das ursprünglich in der Computergrafikindustrie entwickelt wurde, aber dank der Entwicklung der NVIDIA Omniverse (und Isaac Sim)-Plattform für Simulationen und Visualisierungen [93] bereits in technischen Anwendungen anerkannt ist. Im Gegensatz zum parametrischen IFC bietet USD eine einfachere Struktur und ermöglicht die Beschreibung von Geometrie und Objekteigenschaften im JSON-Format (Abb. 4.1-15), was die Informationsverarbeitung erleichtert und die Integration in digitale Prozesse beschleunigt. Das neue Format ermöglicht die Speicherung von Geometrie (zusätzlich zu BREP -NURBS - mehr Details in Teil 6 des Buches) in Form von MESH-Polygonen und von Objekteigenschaften in JSON, was es für automatisierte Prozesse und die Arbeit in Cloud-Ökosystemen bequemer macht [94].

Einige CAD- und ERP-Anbieter verwenden bereits ähnliche Formate (z. B. NWD, SVF, CP2, CPIXML), aber die meisten von ihnen sind nach wie vor geschlossen und für die externe Nutzung nicht verfügbar, was die Möglichkeiten der Datenintegration und Wiederverwendung einschränkt. In diesem Zusammenhang kann USD die gleiche Rolle spielen wie seinerzeit DXF - eine offene Alternative zu proprietären Formaten wie DWG.

General Information			Comparison / Notes
Year of format creation	1991	2016	IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar	IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF	IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK	ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK	ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD	Different alliances for each format
Year of Alliance Formation	1994	2023	The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co	ADSK and Co actively promotes both formats in bS (IAI) since the introduction

Purpose and Usage			Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification	IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing	IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction	USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera	USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

Abb. 4.1-15 Das Format USD ist ein Versuch der CAD-Anbieter, die Forderung nach Interoperabilität und Unabhängigkeit der Konstruktionsdaten von geometrischen Kernen zu erfüllen.

Der Übergang der wichtigsten Entwickler zu offenen und vereinfachten Formaten wie USD, GLTF, OBJ, XML (geschlossenes NWD, CP2, SVF, SVF2, CPIXML) und ähnlichen Formaten (Abb. 3.1-17) spiegelt den globalen Trend und die Nachfrage der Industrie nach Datenvereinfachung und besserer Zugänglichkeit wider. In den kommenden Jahren ist mit einer allmählichen Abkehr von komplexen parametrischen Standards und Formaten mit Abhängigkeit von geometrischen Kernen zugunsten von einfacheren und besser strukturierten Lösungen zu rechnen. Dieser Übergang wird die Digitalisierung der Bauindustrie beschleunigen, die Prozessautomatisierung erleichtern und den Datenaustausch vereinfachen.

Trotz der strategischen Pläne der CAD -Anbieter, neue offene Formate zu fördern, können Fachleute aus der Baubranche auch ohne CAD-Tools (BIM) vollen Zugriff auf Daten aus geschlossenen CAD-Systemen erhalten, indem sie Reverse-Engineering-Tools verwenden.

All diese Trends führen unweigerlich zu einem Wechsel von sperrigen, monolithischen 3D-Modellen zu universellen, strukturierten Daten und zur Verwendung von Formaten, die sich in anderen Branchen schon lange bewährt haben. Sobald Projektteams beginnen, CAD-Modelle nicht mehr nur als visuelle Objekte oder eine Reihe von Dateien zu betrachten, sondern als Datenbanken, die Wissen und Informationen enthalten, ändert sich der Ansatz für Design und Management dramatisch.

Sobald die Teams gelernt haben, wie man strukturierte Daten aus Dokumenten, Texten, Zeichnungen und CAD-Modellen extrahiert, und Zugang zu Datenbanken haben, ist der nächste wichtige Schritt die Datenmodellierung und Qualitätssicherung. Dieser Schritt bestimmt weitgehend die Geschwindigkeit der Verarbeitung und Umwandlung von Informationen, die letztendlich zur Entscheidungsfindung bei bestimmten Anwendungsaufgaben verwendet werden.

KAPITEL 4.2.
KLASSIFIZIERUNG UND INTEGRATION: EINE GEMEINSAME SPRACHE FÜR BAUDATEN

Die Geschwindigkeit der Entscheidungsfindung hängt von der Qualität der Daten ab

Die heutige Architektur von Konstruktionsdaten unterliegt einem grundlegenden Wandel. Die Branche bewegt sich weg von sperrigen, isolierten Modellen und geschlossenen Formaten hin zu flexibleren, maschinenlesbaren Strukturen, die auf Analysen, Integration und Prozessautomatisierung ausgerichtet sind. Der Übergang zu neuen Formaten allein garantiert jedoch noch keine Effizienz - die Qualität der Daten selbst steht zwangsläufig im Mittelpunkt des Interesses.

> Auf den Seiten dieses Buches sprechen wir viel über Formate, Systeme und Prozesse. Aber all diese Bemühungen sind bedeutungslos ohne ein Schlüsselelement: Daten, denen man vertrauen kann. Die Datenqualität ist der Eckpfeiler der Digitalisierung, auf den wir in den folgenden Abschnitten zurückkommen werden.

Moderne Bauunternehmen - vor allem große - verwenden Dutzende und manchmal Tausende von verschiedenen Systemen und Datenbanken (Abb. 4.2-1). Diese Systeme müssen nicht nur regelmäßig mit neuen Informationen befüllt werden, sondern auch effektiv miteinander interagieren. Alle neuen Daten, die durch die Verarbeitung eingehender Informationen entstehen, werden in diese Umgebungen integriert und dienen der Lösung bestimmter Geschäftsaufgaben.

Und während in der Vergangenheit Entscheidungen über bestimmte Geschäftsaufgaben von Topmanagern - so genannten HiPPOs (Abb. 2.1-9) - auf der Grundlage von Erfahrung und Intuition getroffen wurden, ist dieser Ansatz heute, angesichts der stark gestiegenen Informationsmenge, umstritten. Die automatisierte Analytik, die mit Echtzeitdaten arbeitet, tritt an seine Stelle.

> "Traditionell-manuelle" Geschäftsprozessdiskussionen auf Führungsebene werden sich in Richtung operativer Analytik verlagern, die schnelle Antworten auf Geschäftsanfragen erfordert.

Die Zeiten, in denen Buchhalter, Vorarbeiter und Kalkulatoren tage- und wochenlang manuell erstellte Berichte, Übersichtstabellen und Projektdatenvitrinen anfertigten, gehören der Vergangenheit an. Heute werden Geschwindigkeit und Aktualität der Entscheidungsfindung zu einem Schlüsselfaktor für den Wettbewerbsvorteil.

Abb. 4.2-1 Im Gegensatz zu anderen Branchen, in denen dies innerhalb von Stunden oder Minuten geschieht, dauert es im Baugewerbe Tage, bis Berechnungen und Entscheidungen getroffen werden.

Der Hauptunterschied zwischen dem Baugewerbe und den digital fortgeschritteneren Branchen (Abb. 4.2-1) besteht in der geringen Datenqualität und Standardisierung. Veraltete Ansätze bei der Erzeugung, Übertragung und Verarbeitung von Informationen verlangsamen die Prozesse und verursachen Chaos. Das Fehlen einheitlicher Datenqualitätsstandards behindert die Umsetzung einer durchgängigen Automatisierung

> Eine der größten Herausforderungen ist nach wie vor die schlechte Qualität der Eingabedaten sowie das Fehlen formalisierter Verfahren für deren Aufbereitung und Validierung. Ohne verlässliche und konsistente Daten kann es keine effektive Integration zwischen Systemen geben. Dies führt zu Verzögerungen, Fehlern und erhöhten Kosten in jeder Phase des Projektlebenszyklus.

In den folgenden Abschnitten des Buches gehen wir im Detail darauf ein, wie Sie die Datenqualität verbessern, Prozesse standardisieren und den Weg von Informationen zu hochwertigen, validierten und konsistenten Daten verkürzen können.

Standardisierung und Integration von Daten

Eine wirksame Datenverwaltung erfordert eine klare Standardisierungsstrategie. Nur mit klaren Anforderungen an die Datenstruktur und -qualität kann die Datenvalidierung automatisiert, der manuelle Aufwand reduziert und die fundierte Entscheidungsfindung in allen Phasen eines Projekts beschleunigt werden.

In der täglichen Praxis muss ein Bauunternehmen jeden Tag Hunderte von Dateien verarbeiten: E-Mails, PDF

-Dokumente, CAD-Konstruktionsdateien, Daten von IOT-Sensoren, die in die Geschäftsprozesse des Unternehmens integriert werden müssen.

Der Wald des Ökosystems von Datenbanken und Werkzeugen eines Unternehmens (Abb. 4.2-2) muss lernen, Nährstoffe aus den eingehenden Multiformat-Daten abzuleiten, um die vom Unternehmen gewünschten Ergebnisse zu erzielen.

> Um den Datenfluss effektiv zu bewältigen, müssen Sie nicht unbedingt ein Heer von Managern einstellen, sondern zunächst strenge Anforderungen und Standards für die Daten entwickeln und geeignete Werkzeuge zur automatischen Validierung, Harmonisierung und Verarbeitung der Daten einsetzen.

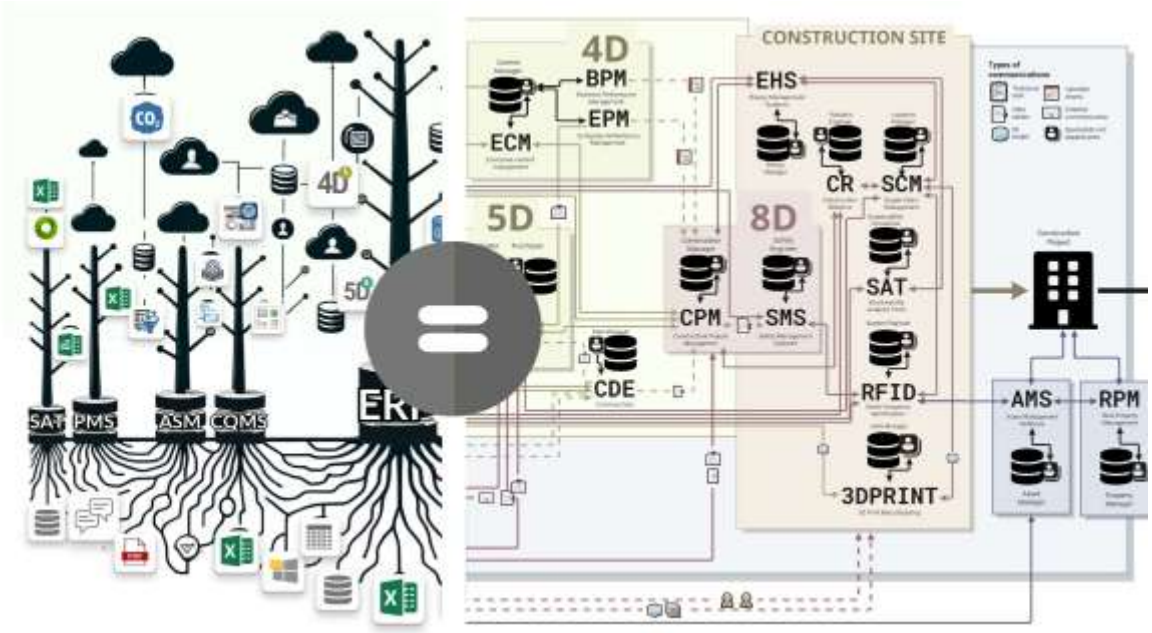


Abb. 4.2-2 Um ein gesundes Unternehmensökosystem zu gewährleisten, müssen die Systeme qualitativ hochwertig und rechtzeitig mit Ressourcen ausgestattet sein.

Um den Prozess der Datenvalidierung und -harmonisierung (für die anschließende automatische Integration) zu automatisieren, sollte man damit beginnen, die Mindestanforderungen an die Daten für jedes spezifische System zu beschreiben. Diese Anforderungen definieren:

- Was genau müssen Sie besorgen?
- In welcher Form (Struktur, Format)?
- Welche Attribute sind obligatorisch?
- Welche Toleranzen in Bezug auf Genauigkeit und Vollständigkeit sind akzeptabel?

Datenanforderungen beschreiben Kriterien für die Qualität, Struktur und Vollständigkeit der empfangenen und verarbeiteten Informationen. So ist beispielsweise bei Texten in PDF -Dokumenten auf eine korrekte

Formatierung nach Industriestandards zu achten (Abb. 7.2-14 - Abb. 7.2-16). Objekte in CAD -Modellen müssen korrekte Attribute (Abmessungen, Codes, Verknüpfungen zu Klassifikatoren) aufweisen (Abb. 7.3-9, Abb. 7.3-10). Und für Vertragsscans sind eindeutige Daten und die Fähigkeit zur automatischen Extraktion des Betrags und der Schlüsselbegriffe wichtig (Abb. 4.1-7 - Abb. 4.1-10).

> Die Formulierung von Datenanforderungen und die automatische Überprüfung ihrer Einhaltung ist einer der zeitaufwändigsten, aber wichtigsten Schritte. Es ist der zeitaufwändigste Schritt in Geschäftsprozessen.

Wie in Teil 3 dieses Buches erwähnt, verbringen Fachleute für Business Intelligence (BI) zwischen 50 und 90% ihrer Zeit mit der Datenaufbereitung und nicht mit der Analyse (Abb. 3.2-5). Dieser Prozess umfasst die Sammlung, Überprüfung, Validierung, Harmonisierung und Strukturierung von Daten.

Laut einer Umfrage aus dem Jahr 2016 [95] gaben Datenwissenschaftler aus den verschiedensten Bereichen an, dass sie den größten Teil ihrer Arbeitszeit (etwa 80%) damit verbringen, das zu tun, was sie am wenigsten mögen (Abb. 4.2-3): vorhandene Datensätze zu sammeln und sie zu organisieren (zu vereinheitlichen, zu strukturieren). Somit bleiben weniger als 20% ihrer Zeit für kreative Aufgaben, wie das Auffinden von Mustern und Regelmäßigkeiten, die zu neuen Erkenntnissen und Entdeckungen führen.

Abb. 4.2-3 Die Überprüfung von und die Sicherstellung der Datenqualität ist der kostspieligste, zeitaufwändigste und komplexeste Schritt bei der Vorbereitung von Daten für die Integration in andere Systeme.

Erfolgreiches Datenmanagement in einem Bauunternehmen erfordert einen umfassenden Ansatz, der die Parametrisierung von Aufgaben, die Formulierung von Anforderungen an die Datenqualität und den Einsatz geeigneter Werkzeuge zur automatisierten Validierung umfasst.

Digitale Interoperabilität beginnt mit Anforderungen

Mit der zunehmenden Anzahl digitaler Systeme in Unternehmen steigt auch der Bedarf an Datenkonsistenz zwischen diesen Systemen. Manager, die für verschiedene IT-Systeme verantwortlich sind, sind oft nicht in der Lage, mit der zunehmenden Menge an Informationen und der Vielfalt der Formate Schritt zu halten. Unter diesen Umständen sind sie gezwungen, Spezialisten damit zu beauftragen, Daten in einer Form zu erstellen, die für die Verwendung in anderen Anwendungen und Plattformen geeignet ist.

Dies wiederum erfordert, dass Ingenieure und Mitarbeiter, die an der Datengenerierung beteiligt sind, sich an eine Vielzahl von Anforderungen anpassen, oft ohne Transparenz und ohne ein klares Verständnis davon, wo und wie die Daten in Zukunft verwendet werden. Das Fehlen standardisierter Ansätze für den Umgang mit Informationen führt zu Ineffizienzen und erhöhten Kosten während der Überprüfungsphase, die aufgrund der Komplexität und der nicht standardisierten Natur der Daten oft manuell erfolgt.

> Das Problem der Datenstandardisierung ist nicht nur eine Frage der Bequemlichkeit oder der Automatisierung. Es ist ein direkter finanzieller Verlust. Einem IBM-Bericht aus dem Jahr 2016 zufolge beläuft sich der jährliche Verlust durch schlechte Datenqualität in den USA auf 3,1 Billionen US-Dollar [96]. Darüber hinaus zeigen Studien des MIT und anderer analytischer Beratungsunternehmen, dass die Kosten einer schlechten Datenqualität bis zu 15-25% des Umsatzes eines Unternehmens ausmachen können [97].

Unter diesen Bedingungen ist es von entscheidender Bedeutung, klar definierte Datenanforderungen und Beschreibungen darüber zu haben, welche Parameter in welchem Format und mit welchem Detaillierungsgrad in den erstellten Objekten enthalten sein sollten. Ohne eine Formalisierung dieser Anforderungen ist es unmöglich, die Datenqualität und Kompatibilität zwischen Systemen und Projektphasen zu gewährleisten (Abb. 4.2-4).

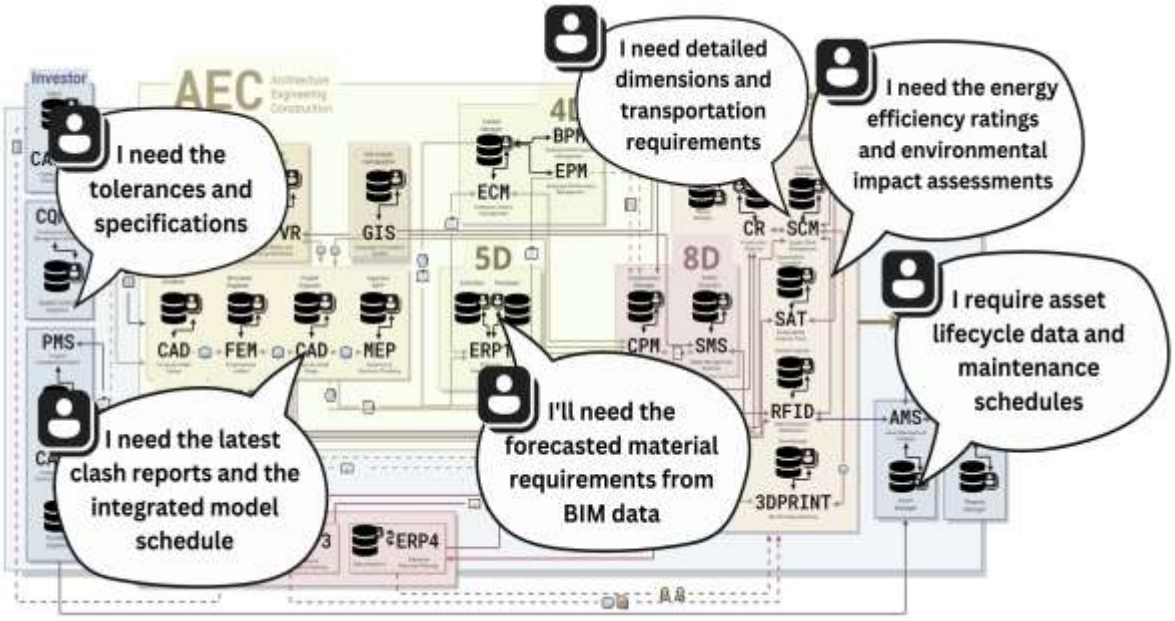


Abb. 4.2-4 Das Geschäft basiert auf der Interaktion verschiedener Rollen, von denen jede bestimmte Parameter und Werte benötigt, die für die Erfüllung der Geschäftsziele entscheidend sind.

Um die richtigen Datenanforderungen formulieren zu können, müssen Sie die Geschäftsprozesse auf der Datenebene verstehen. Bauprojekte unterscheiden sich in Art, Umfang und Anzahl der Beteiligten, und jedes System - sei es die Modellierung (CAD (BIM)), die Terminplanung (ERP 4D), die Kalkulation (ERP 5D) oder die Logistik (SCM) - erfordert seine eigenen einzigartigen Parameter für die Eingaben (Input-Entity-Elemente).

Je nach diesem Bedarf müssen die Unternehmensleiter entweder neue Datenstrukturen entwerfen, um die Anforderungen zu erfüllen, oder bestehende Tabellen und Datenbanken anpassen. Die Qualität der erstellten Daten hängt direkt davon ab, wie präzise und korrekt die Anforderungen formuliert sind (Abb. 4.2-5).

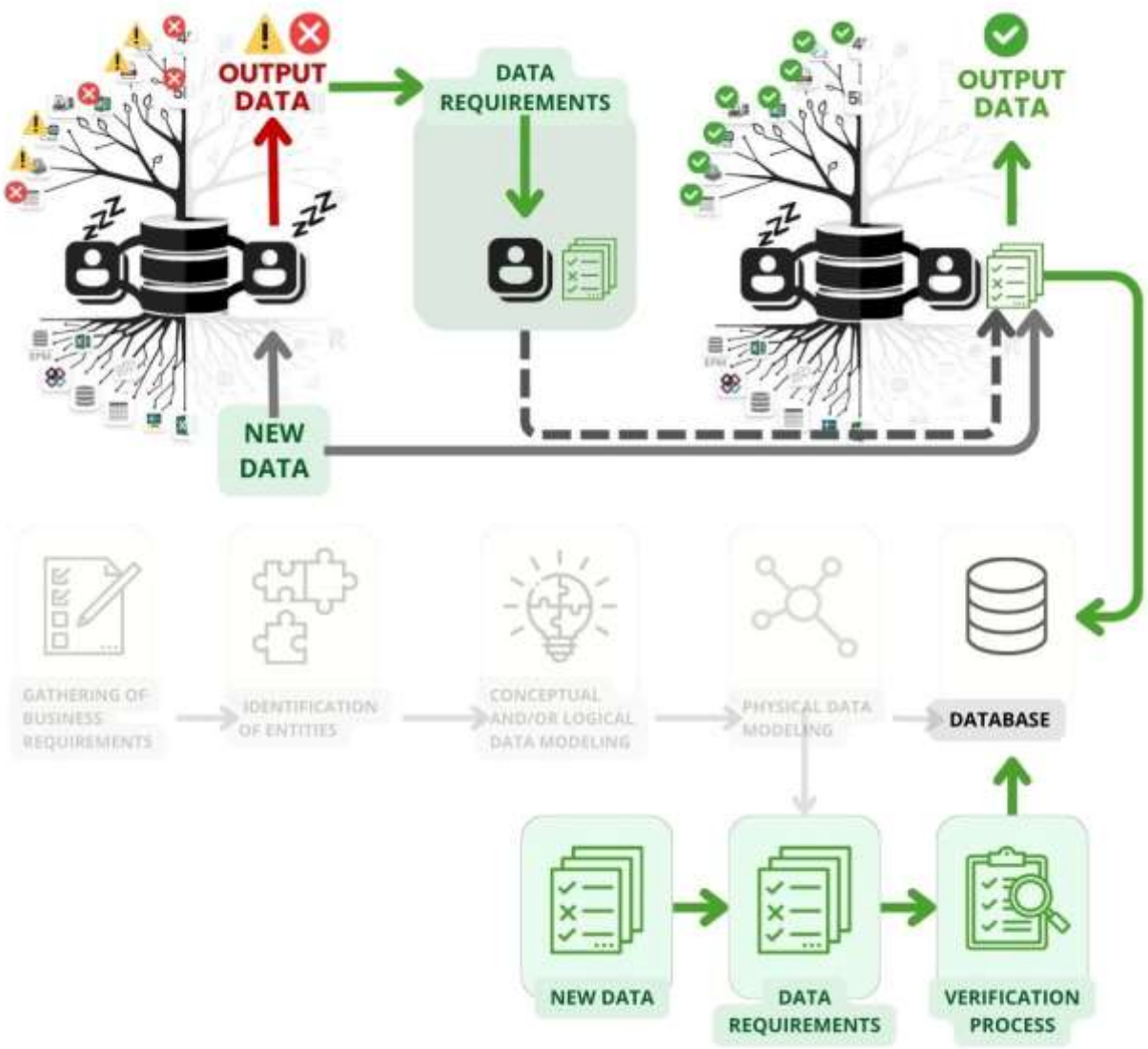


Abbildung 4.2-5 Die Datenqualität hängt von der Qualität der Anforderungen ab, die für bestimmte Datenverwendungsfälle erstellt werden.

Da jedes System seine eigenen spezifischen Datenanforderungen hat, besteht der erste Schritt bei der Formulierung allgemeiner Anforderungen darin, alle an den Geschäftsprozessen beteiligten Elemente zu kategorisieren. Dies bedeutet, dass die Objekte in Klassen und Klassengruppen unterteilt werden müssen, die bestimmten Systemen oder Anwendungsaufgaben entsprechen. Für jede dieser Gruppen werden separate Anforderungen an Datenstruktur, Attribute und Qualität entwickelt.

In der Praxis steht die Umsetzung dieses Konzepts jedoch vor einer großen Herausforderung: dem Fehlen einer gemeinsamen Sprache für die Gruppierung von Daten. Unterschiedliche Klassifizierungen, doppelte Bezeichner und inkompatible Formate führen dazu, dass jedes Unternehmen, jede Software und sogar jedes Projekt seine eigenen, isolierten Datenmodelle und Klassen bildet. Das Ergebnis ist ein digitaler "Turm zu Babel", bei dem die Übertragung von Informationen zwischen Systemen eine mehrfache Konvertierung in die richtigen Datenmodelle und -klassen erfordert, die häufig manuell erfolgt. Dieses Hindernis kann nur durch den Übergang zu universellen Klassifizierern und standardisierten Anforderungskatalogen überwunden werden.

Eine gemeinsame Sprache der Konstruktion: die Rolle von Klassifikatoren bei der digitalen Transformation

Im Zusammenhang mit der Digitalisierung und Automatisierung von Prüf- und Bearbeitungsprozessen spielen Klassifizierungssysteme eine besondere Rolle - eine Art "digitales Wörterbuch", das für Einheitlichkeit bei der Beschreibung und Parametrisierung von Objekten sorgt. Klassifikatoren bilden die "gemeinsame Sprache", die es ermöglicht, Daten nach ihrer Bedeutung zu gruppieren und Daten zwischen verschiedenen Systemen, Verwaltungsebenen und Phasen des Projektlebenszyklus zu integrieren.

Die greifbarste Auswirkung von Klassifizierern liegt in der Wirtschaftlichkeit des Lebenszyklus von Gebäuden, wobei der wichtigste Aspekt die Optimierung der langfristigen Betriebskosten ist. Studien zeigen, dass die Betriebskosten bis zu 80% der Gesamtbetriebskosten eines Gebäudes ausmachen und damit dreimal höher sind als die anfänglichen Baukosten (Abb. 4.2-6) [98]. Dies bedeutet, dass die Entscheidung über die künftigen Kosten weitgehend in der Planungsphase getroffen wird

> Aus diesem Grund sollten die Anforderungen der Betriebsingenieure (CAFM, AMS, PMS, RPM) zum Ausgangspunkt für die Erstellung von Datenanforderungen während der Entwurfsphase werden (Abb. 1.2-4). Diese Systeme sollten nicht als letzte Phase des Projekts betrachtet werden, sondern als integraler Bestandteil des gesamten digitalen Ökosystems des Projekts, vom Konzept bis zum Rückbau

Ein modernes Klassifizierungssystem ist nicht nur ein System von Codes zur Gruppierung. Er ist ein Mechanismus für das gegenseitige Verständnis zwischen Architekten, Ingenieuren, Kalkulatoren, Logistikern, Wartungs- und IT-Systemen. So wie der Autopilot eines Autos Objekte auf der Straße eindeutig und genau erkennen muss, müssen digitale Bausysteme und ihre Nutzer das gleiche Projektelement über die Elementklasse für verschiedene Systeme eindeutig interpretieren.

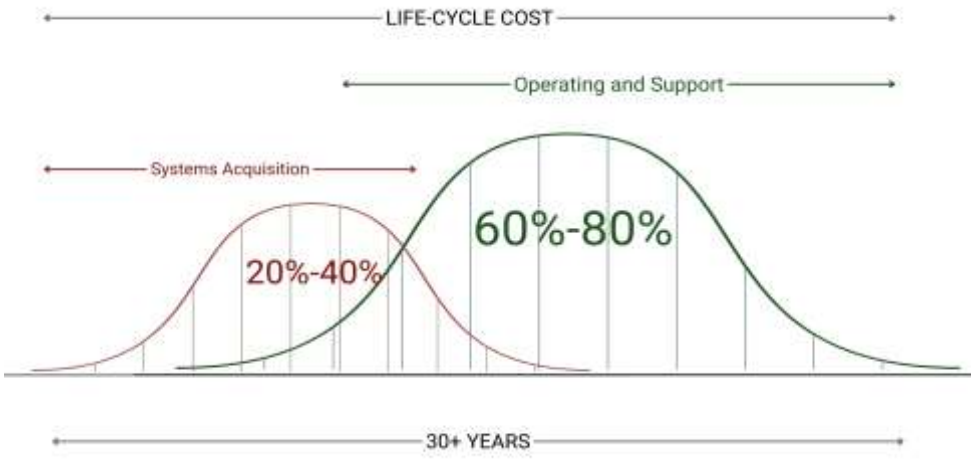


Abb. 4.2-6 Die Betriebs- und Wartungskosten übersteigen die Baukosten um das Dreifache und machen 60-80% der gesamten Lebenszykluskosten eines Gebäudes aus (nach [99]).

Der Entwicklungsstand der Klassifikatoren steht in direktem Zusammenhang mit dem Grad der Digitalisierung eines Unternehmens und seiner digitalen Reife. Unternehmen mit einem geringen digitalen Reifegrad sind mit fragmentierten Daten, inkompatiblen Informationssystemen und infolgedessen mit inkompatiblen und ineffizienten Klassifikatoren konfrontiert. In solchen Unternehmen kann ein und dasselbe Element in verschiedenen Systemen oft unterschiedliche Gruppierungskennungen haben, was die endgültige Integration erheblich erschwert und eine Prozessautomatisierung unmöglich macht.

So kann beispielsweise ein und dasselbe Fenster in einem Projekt im CAD-Modell, im Kalkulations- und im Wartungssystem (Abb. 4.2-7) aufgrund der mehrdimensionalen Wahrnehmung der Elemente durch die verschiedenen Prozessbeteiligten unterschiedlich bezeichnet werden. Für den Kalkulator sind in der Fensterkategorie Element, Volumen und Kosten wichtig, für den Wartungsdienst - Verfügbarkeit und Wartbarkeit, für den Architekten - ästhetische und funktionale Eigenschaften. Infolgedessen kann ein und dasselbe Element unterschiedliche Parameter erfordern.

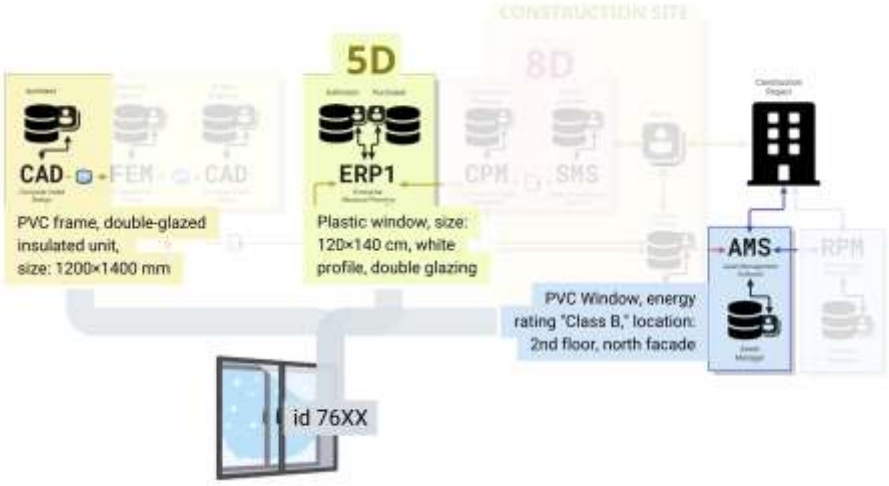


Abb. 4.2-7 Bei inkonsistenter Klassifizierung zwischen den Systemen verliert ein Element bei jedem Schritt seines Übergangs zu einem anderen System einige Attributinformationen.

Aufgrund der Schwierigkeit, die Klassifizierung von Bauelementen eindeutig zu definieren, ordnen Fachleute aus verschiedenen Bereichen demselben Element oft unvereinbare Klassen zu. Dadurch geht eine einheitliche Sicht auf das Objekt verloren, was ein nachträgliches manuelles Eingreifen erfordert, um die verschiedenen Klassifizierungssysteme zu harmonisieren und die Konsistenz zwischen den von verschiedenen Spezialisten definierten Typen und Klassen herzustellen.

Infolge dieser Inkonsistenz kann die Betriebsdokumentation, die die Beschaffungsabteilung (ERP) erhält, wenn ein Bauelement von einem Hersteller gekauft wird, oft nicht korrekt mit der Klassifizierung dieses Elements auf der Baustelle verknüpft werden (PMIS, SCM). Infolgedessen werden kritische Informationen wahrscheinlich nicht in die Infrastruktur- und Anlagenverwaltungssysteme (CAFM, AMS) integriert, was zu ernsthaften Problemen bei der Inbetriebnahme sowie bei der späteren Wartung (AMS, RPM) oder dem Austausch des Elements führt.

> In Unternehmen mit einem hohen digitalen Reifegrad spielen Klassifikatoren die Rolle eines Nervensystems, das alle Informationsflüsse vereint. Ein und dasselbe Element erhält eine eindeutige Kennung, so dass es ohne Verzerrung oder Verlust zwischen CAD, ERP, AMS und CAFM -Systemen und ihren Klassifikatoren übertragen werden kann.

Um effektive Klassifikatoren zu erstellen, muss man verstehen, wie die Daten verwendet werden. Ein und derselbe Ingenieur kann ein Element in verschiedenen Projekten unterschiedlich benennen und klassifizieren. Nur durch das Sammeln von Nutzungsstatistiken im Laufe der Jahre kann ein stabiles Klassifizierungssystem entwickelt werden. Maschinelles Lernen hilft dabei: Algorithmen analysieren Tausende von Projekten (Abb. 9.1-10) und ermitteln durch maschinelles Lernen wahrscheinliche Klassen und Parameter (Abb. 10.1-6). Die automatische Klassifizierung ist besonders wertvoll in Umgebungen, in denen eine manuelle Klassifizierung aufgrund des Datenvolumens nicht möglich ist. Automatische Klassifizierungssysteme werden in der Lage sein, grundlegende Kategorien auf der Grundlage von minimal bestückten Artikelparametern zu unterscheiden (weitere Einzelheiten im neunten und zehnten Teil des Buches).

Die entwickelten Klassifizierungssysteme werden zu Katalysatoren für die weitere Digitalisierung und schaffen die Grundlage für die Digitalisierung:

- Automatisierte Schätzung von Projektkosten und -fristen.
- Vorausschauende Analyse von potenziellen Risiken und Konflikten
- Optimierung von Beschaffungsprozessen und Logistikketten
- Erstellung digitaler Zwillinge von Gebäuden und Strukturen
- Integration mit Systemen für intelligente Städte und das Internet der Dinge

> Die Zeit für den Wandel ist begrenzt - mit der Entwicklung des maschinellen Lernens und der Computer-Vision-Technologien wird das Problem der automatischen Klassifizierung, das jahrzehntelang unlösbar war, in den kommenden Jahren gelöst werden, und Bau- und Planungsunternehmen, die sich nicht rechtzeitig anpassen, laufen Gefahr, das Schicksal der von digitalen Plattformen verdrängten Taxiflotten zu wiederholen.

Mehr über die Automatisierung von Kalkulation und Terminplanung sowie über Big Data und maschinelles Lernen erfahren Sie im fünften und neunten Teil des Buches. Das Risiko einer Wiederholung des Schicksals

der Taxiflotten und der Uberisierung der Baubranche wird im zehnten Teil des Buches ausführlich diskutiert.

Um die Schlüsselrolle von Klassifikatoren bei der digitalen Transformation der Bauindustrie zu verstehen, ist es notwendig, sich mit der Geschichte ihrer Entwicklung zu befassen. Der historische Kontext ermöglicht es uns zu erkennen, wie sich die Klassifizierungsansätze entwickelt haben und welche Trends ihren aktuellen Stand bestimmen.

Masterformat, OmniClass, Uniclass und CoClass: die Entwicklung der Klassifikationssysteme

Historisch gesehen haben sich die Bauelement- und Arbeitsklassifizierer in drei Generationen entwickelt, die jeweils den Stand der verfügbaren Technologie und die aktuellen Bedürfnisse der Branche in einem bestimmten Zeitraum widerspiegeln (Abb. 4.2-8):

- **Erste Generation** (Anfang der 1950er bis Ende der 1980er Jahre) - Papierverzeichnisse, hierarchische Klassifikatoren, die lokal verwendet werden (z. B. Masterformat, SfB).
- **Die zweite Generation** (Ende der 1990er bis Mitte der 2010er Jahre) sind Tabellenkalkulationen und strukturierte Datenbanken, die in Excel und Access implementiert sind (ASTM E 1557, OmniClass, Uniclass 1997).
- **Dritte Generation** (2010er Jahre bis heute) - digitale Dienste und APIs -Schnittstellen, Integration mit CAD (BIM), Automatisierung (Uniclass 2015, CoClass).

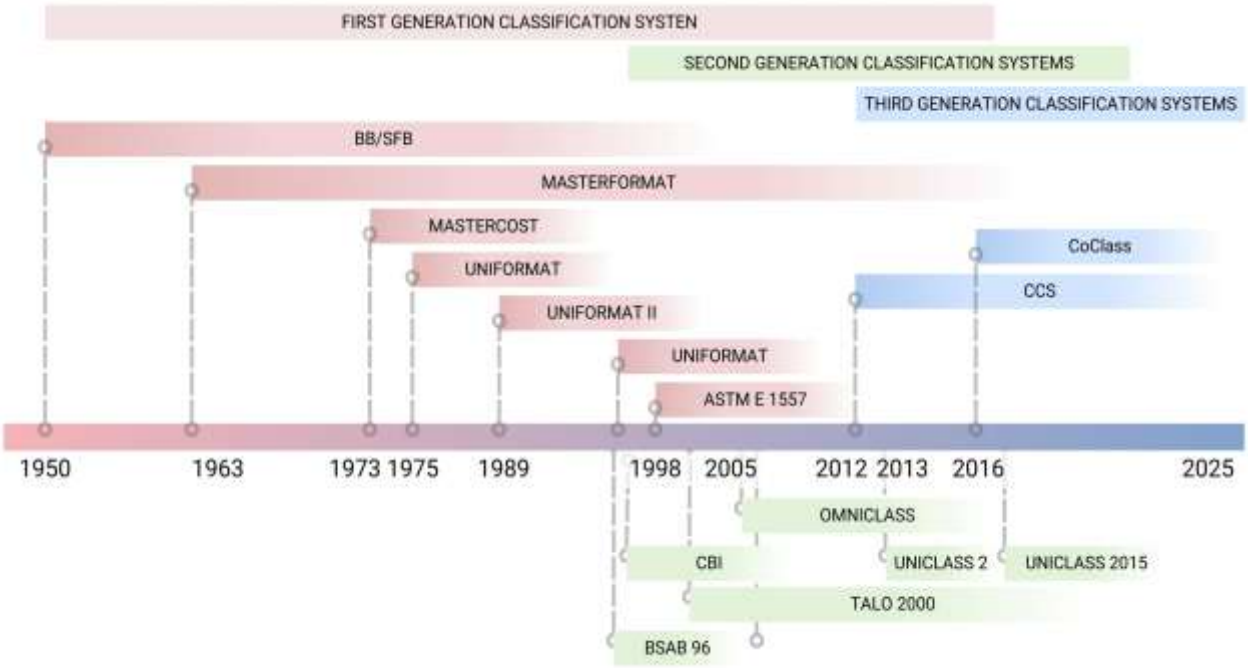


Abb. 4.2-8 Drei Generationen von Klassifikatoren für das Baugewerbe.

In den letzten Jahrzehnten hat sich die hierarchische Komplexität (Abb. 4.2-9) von Klassifikatoren verringert: Während frühe Systeme wie OmniClass bis zu 7 Verschachtelungsebenen verwendeten, um 6887 Klassen zu beschreiben, beschränken sich moderne Lösungen wie CoClass auf 3 Ebenen mit 750 Klassen. Dies erleichtert die Arbeit mit den Daten, ohne dass die notwendige Granularität verloren geht. Uniclass 2015, das im Vereinigten Königreich häufig als Standard verwendet wird, fasst 7210 Klassen in nur 4 Ebenen zusammen, was für CAD-Projekte und das öffentliche Auftragswesen praktisch ist.

Classifier	Table / Objects	Number of classes	Nesting depth
OmniClass	Table 23 Products	6887	7 levels
Uniclass 2015	Pr — Products	7210	4 levels
CoClass, CCS	Components	750	3 levels

Abb. 4.2-9 Mit jeder neuen Generation von Klassifikatoren nimmt die Komplexität der Kategorisierung drastisch ab.

In den Kalkulationssystemen der verschiedenen Länder kann selbst ein typisches Element wie eine Betonfundamentwand aufgrund unterschiedlicher Klassifizierungen auf sehr unterschiedliche Weise beschrieben werden (Abb. 4.2-10). Diese Unterschiede spiegeln die nationalen Baupraktiken, die verwendeten Messsysteme, die Ansätze zur Materialklassifizierung sowie die in den einzelnen Ländern geltenden rechtlichen und technischen Anforderungen wider.

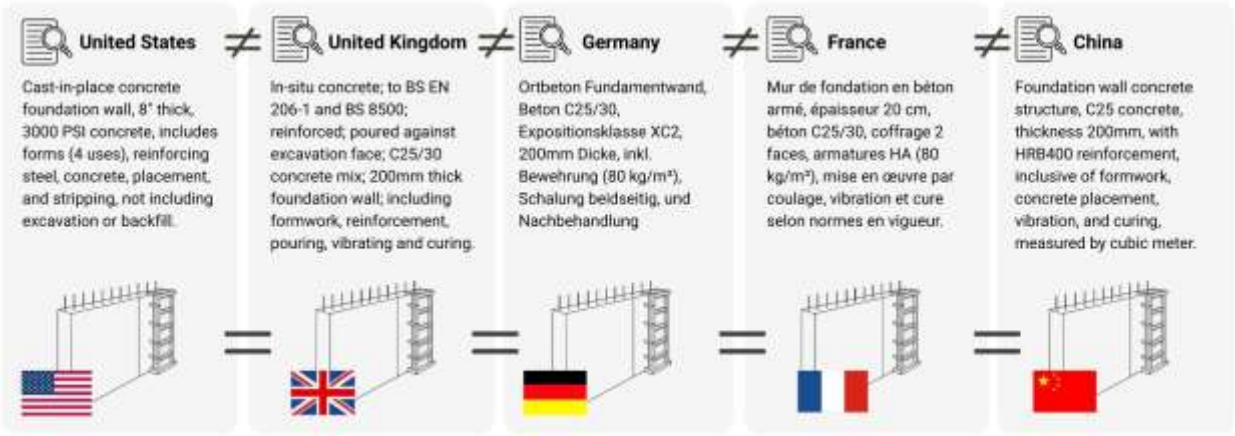


Abb. 4.2-10 Dasselbe Element wird in Projekten in verschiedenen Ländern durch unterschiedliche Beschreibungen und Klassifizierungen verwendet.

Die unterschiedlichen Klassifizierungen der gleichen Elemente erschweren die internationale Zusammenarbeit und machen den Vergleich von Kosten und Arbeitsumfang bei internationalen Projekten zeitaufwändig und manchmal fast unmöglich. Derzeit gibt es keinen universellen Klassifikator auf globaler Ebene - jedes Land oder jede Region entwickelt seine eigenen Systeme auf der Grundlage lokaler Normen, Sprache und Geschäftskultur:

- **CCS** (Dänemark): Cost Classification System - ein System zur Klassifizierung der Kosten während des gesamten Lebenszyklus einer Anlage (Planung, Bau, Betrieb). Der Schwerpunkt liegt auf der Betriebs- und Wartungslogik, umfasst aber auch das Budget- und Ressourcenmanagement.
- **NS 3451** (Norwegen): kategorisiert Einrichtungen nach Funktion, Konstruktionselementen und Lebenszyklusphasen. Wird für Projektmanagement, Kostenabschätzung und langfristige Planung verwendet.
- **MasterFormat** (USA): ein System zur Strukturierung von Baubeschreibungen in Abschnitte (z. B. Beton, Elektrik, Ausbau). Der Schwerpunkt liegt auf Disziplinen und Arbeitsarten und nicht auf funktionalen Elementen (im Gegensatz zu UniFormat).
- **Uniclass 2** (UK): einer der detailliertesten Klassifikatoren, der im öffentlichen Auftragswesen und bei BIM-Projekten verwendet wird. Vereint Daten über Objekte, Arbeiten, Materialien und Räume in einem einzigen System.
- **OmniClass**: ein internationaler Standard (entwickelt von CSI in den USA) zur Verwaltung von Objektinformationen von Bauteilbibliotheken bis zu elektronischen Spezifikationen. Geeignet für die langfristige Datenspeicherung, kompatibel mit CAD (BIM) und anderen digitalen Werkzeugen.
- **COBie**: Construction-Operation Building Information Exchange ist ein internationaler Standard für den Austausch von Daten zwischen den Phasen Planung, Bau und Betrieb. Enthalten in BS 1192-4:2014 als Teil des Konzepts "BIM -model ready for use". Konzentriert sich auf den Informationstransfer (z. B. Ausrüstungsspezifikationen, Garantien, Kontakte zu Auftragnehmern).

Die Globalisierung des Baugewerbes wird wahrscheinlich zu einer allmählichen Vereinheitlichung der Klassifizierungssysteme für Bauelemente führen, was die Abhängigkeit von lokalen nationalen Normen erheblich verringern wird. Dieser Prozess könnte sich analog zur Entwicklung der Internet-Kommunikation entwickeln, bei der universelle Datenübertragungsprotokolle schließlich die unterschiedlichen lokalen Formate verdrängen und die globale Interoperabilität der Systeme gewährleisten.

Ein alternativer Entwicklungspfad könnte ein direkter Übergang zu automatischen Klassifizierungssystemen auf der Grundlage von Technologien des maschinellen Lernens sein. Diese Technologien, die heute hauptsächlich im Bereich des autonomen Verkehrs entwickelt werden, haben ein erhebliches Potenzial für die Anwendung auf große CAD-Konstruktionsdatensätze (Abb. 10.1-6).

> Heute ist die Situation nicht mehr nur auf die nationale Gruppierung von Klassifizierern beschränkt. Aufgrund der vielen Besonderheiten, die auf nationaler Ebene nicht berücksichtigt werden, muss jedes Unternehmen die Kategorien von Elementen und Ressourcen, mit denen es arbeitet, vereinheitlichen und standardisieren.

In der Regel beginnt dieser Prozess im Kleinen - mit lokalen Objekttabellen oder internen Kennzeichnungssystemen. Das strategische Ziel ist jedoch der Übergang zu einer gemeinsamen Sprache für die Beschreibung aller Elemente, die nicht nur innerhalb des Unternehmens, sondern auch außerhalb des Unternehmens verständlich ist - idealerweise harmonisiert mit internationalen oder branchenspezifischen Klassifikatoren (Abb. 4.2-8). Dieser Ansatz erleichtert die Integration mit externen Partnern und digitalen Systemen und fördert die Bildung einheitlicher End-to-End-Prozesse innerhalb des Lebenszyklus von Objekten.

Bevor man zur Automatisierung und zu skalierbaren IT-Systemen übergeht, ist es notwendig, entweder Klassifikatoren auf nationaler Ebene zu verwenden oder eine eigene, logische und eindeutige Struktur zur Identifizierung von Elementen aufzubauen. Jedes Objekt - sei es ein Fenster (Abb. 4.2-11), eine Tür oder ein technisches System - muss so beschrieben werden, dass es in jedem digitalen System des Unternehmens

eindeutig erkannt werden kann. Dies ist entscheidend für den Übergang von flachen Zeichnungen zu digitalen Modellen, die sowohl die Planungsphase als auch den Betrieb von Gebäuden abdecken.

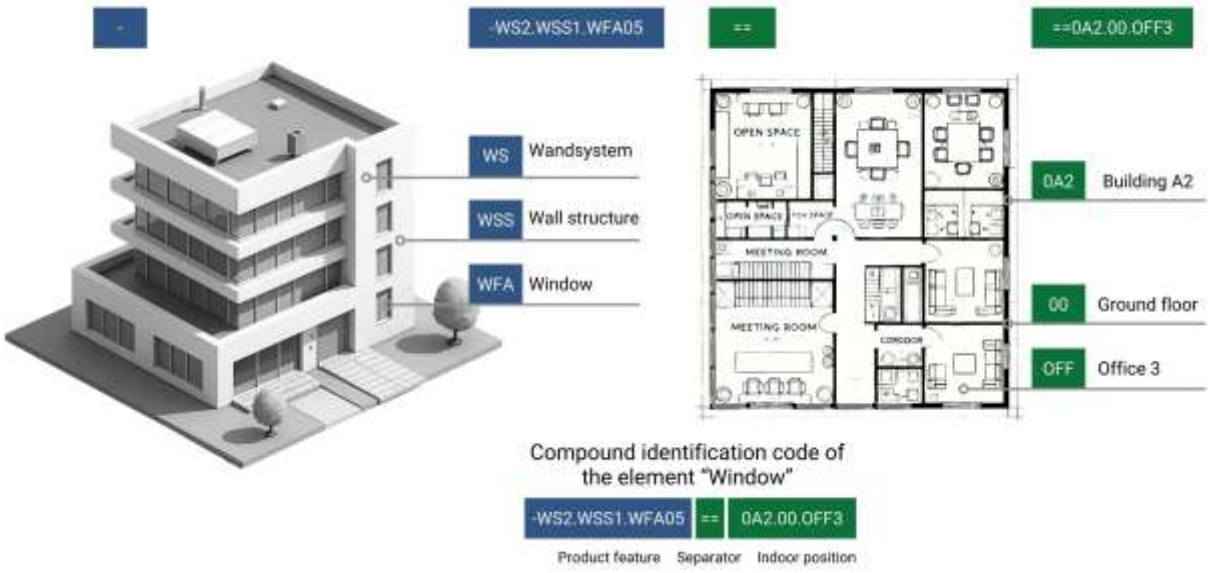


Abb. 4.2-11 Beispiel eines zusammengesetzten Fensterelement-Identifikators auf der Grundlage der Klassifizierung und der Position im Gebäude.

Ein Beispiel für interne Klassifikatoren könnte die Entwicklung eines zusammengesetzten Identifikationscodes sein (Abb. 4.2-11). Ein solcher Code kombiniert mehrere Informationsebenen: den funktionalen Zweck des Elements (z. B. "Fenster in der Wand"), seinen Typ und den genauen räumlichen Bezug - Gebäude A2, Stockwerk 0, Raum 3. Eine solche mehrstufige Struktur ermöglicht die Schaffung eines einheitlichen Navigationssystems durch digitale Modelle und Dokumentationen, insbesondere in den Phasen der Datenüberprüfung und -umwandlung, in denen eine eindeutige Gruppierung von Elementen erforderlich ist. Die eindeutige Erkennung von Elementen gewährleistet die Konsistenz zwischen den Abteilungen und verringert das Risiko von Doppelarbeit, Fehlern und Informationsverlusten.

Ein gut aufgebauter Klassifikator ist nicht nur ein technisches Dokument, sondern die Grundlage des digitalen Ökosystems eines Unternehmens:

- gewährleistet die Datenkompatibilität zwischen den Systemen;
- reduziert die Kosten für die Suche und Verarbeitung von Informationen;
- erhöht die Transparenz und Verwaltbarkeit;
- schafft eine Grundlage für Skalierung und Automatisierung.

Die standardisierte Beschreibung von Objekten durch die Verwendung nationaler Klassifikatoren oder eigener zusammengesetzter Identifikationscodes wird zur Grundlage für konsistente Daten, zuverlässigen Informationsaustausch und die anschließende Implementierung intelligenter Dienste - von der automatisierten Beschaffung bis zu digitalen Zwillingen.

Nach Abschluss der Strukturierungsphase von Multiformat-Daten und der Auswahl des Klassifikators, der zur Erkennung und Gruppierung der Elemente verwendet werden soll, besteht der nächste Schritt darin, die Daten korrekt zu modellieren. Dieser Prozess umfasst die Identifizierung von Schlüsselparametern, den Aufbau einer logischen Datenstruktur und die Beschreibung der Beziehungen zwischen den Elementen.

KAPITEL 4.3.
DATENMODELLIERUNG UND KOMPETENZZENTRUM

Datenmodellierung: konzeptionelles, logisches und physisches Modell

Eine effiziente Verwaltung von Daten (die wir zuvor strukturiert und kategorisiert haben) ist ohne eine gut durchdachte Speicher- und Verarbeitungsstruktur nicht möglich. Um den Zugang und die Konsistenz der Informationen in der Speicher- und Verarbeitungsphase zu gewährleisten, verwenden die Unternehmen die Datenmodellierung, eine Methode, die es ihnen ermöglicht, Tabellen, Datenbanken und Verknüpfungen zwischen ihnen entsprechend den geschäftlichen Anforderungen zu gestalten.

> Die Datenmodellierung ist das Fundament, auf dem jedes digitale Ökosystem aufgebaut ist. Ohne eine Beschreibung der Systeme, der Anforderungen und der Datenmodellierung wissen oder verstehen Ingenieure und Fachleute, die Daten erstellen, nicht, wo die von ihnen erstellten Daten verwendet werden sollen.

Wie beim Bau eines Gebäudes, bei dem man ohne Plan nicht mit dem Verlegen von Ziegeln beginnen kann, ist für die Erstellung eines Data-Warehouse-Systems eine klare Vorstellung davon erforderlich, welche Daten verwendet werden, wie sie verknüpft werden und wer mit ihnen arbeiten wird. Ohne eine Beschreibung der Prozesse und Anforderungen verlieren die Ingenieure und Fachleute, die die Daten erstellen, den Überblick darüber, wo und wie die Daten in Zukunft verwendet werden.

Das Datenmodell dient als Brücke zwischen Unternehmen und IT. Es ermöglicht die Formalisierung von Anforderungen, die Strukturierung von Informationen und die Erleichterung der Kommunikation zwischen den Beteiligten. In diesem Sinne ist die Datenmodellierung vergleichbar mit der Arbeit eines Architekten, der nach den Plänen des Kunden einen Bauplan entwickelt und diesen dann an die Bauarbeiter - Datenbankadministratoren und Entwickler - zur Umsetzung (Datenbankerstellung) weitergibt.

Daher muss jedes Bauunternehmen neben der Strukturierung und Kategorisierung von Elementen und Ressourcen (Abb. 4.2-11) auch die Kunst des "Aufbaus" von Datenbanken (Tabellen) beherrschen und lernen, wie man Verbindungen zwischen ihnen herstellt, so als ob man die Ziegelsteine zu einer zuverlässigen und starken Wissenswand aus den Daten des Unternehmens verbindet. Zu den Schlüsselkonzepten der Datenmodellierung (Abb. 4.3-1) gehören:

- **Entitäten** sind Objekte, über die Daten gesammelt werden müssen. In der frühen Entwurfsphase kann eine Entität ein einzelnes Element sein (z. B. "Tür"), im Kalkulationsmodell kann es eine Gruppe von Elementen sein, die kategorisiert werden (z. B. "Innentüren").
- **Attribute** sind Merkmale von Einheiten, die wichtige Details beschreiben: Abmessungen, Eigenschaften, Montagekosten, Logistik und andere Parameter.
- **Beziehungen (Verknüpfungen)** - zeigen, wie Entitäten miteinander interagieren. Sie können von einem der folgenden Typen sein: "einer zu einem", "viele zu einem", "viele zu vielen".
- **ER-Diagramme** (Entity-Relationship-Diagramme) sind visuelle Diagramme, die Entitäten, Attribute und die Beziehungen zwischen ihnen zeigen. ER-Diagramme können konzeptionell, logisch und physisch sein - jedes spiegelt eine andere Detailstufe wider.

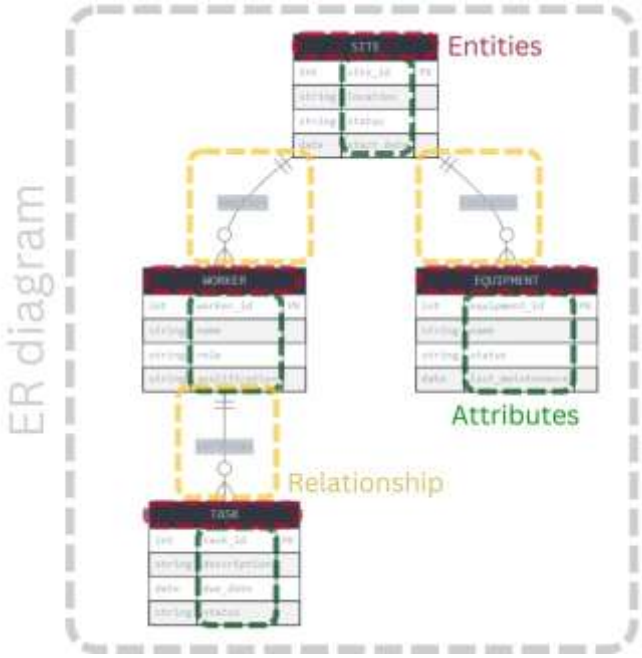


Abb. 4.3-1 ER-Diagramm einer konzeptionellen Datenbankstruktur mit Entitäten, Attributen und Beziehungen.

Der Prozess der Gestaltung von Daten und der Definition von Beziehungen zwischen ihnen wird traditionell in drei Hauptmodelle unterteilt. Jedes von ihnen erfüllt bestimmte Funktionen, die sich im Detaillierungs- und Abstraktionsgrad der Darstellung der Datenstruktur unterscheiden:

- **Konzeptuelles Datenmodell**: Dieses Modell beschreibt die wichtigsten Entitäten und ihre Beziehungen, ohne auf die Details der Attribute einzugehen. Es wird normalerweise in der Anfangsphase der Planung verwendet. In dieser Phase können wir anhand von Datenbanken und Systemen die Beziehungen zwischen verschiedenen Abteilungen und Fachleuten skizzieren.

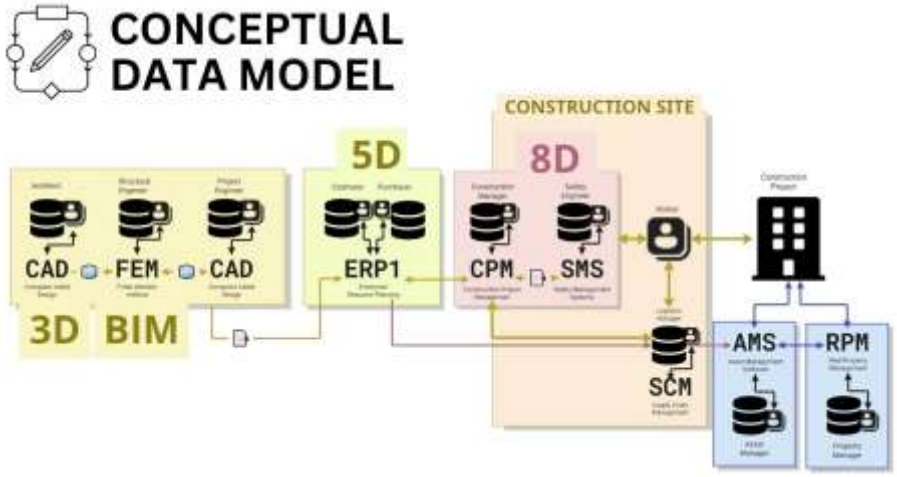


Abb. 4.3-2 Das konzeptionelle Diagramm beschreibt den Inhalt des Systems: eine Darstellung der Beziehungen auf hoher Ebene, ohne technische Details.

- **Logisches Datenmodell**: Auf der Grundlage des konzeptionellen Modells enthält das logische Datenmodell detaillierte Beschreibungen von Entitäten, Attributen, Schlüsseln und Beziehungen, die Geschäftsinformationen und Regeln abbilden.

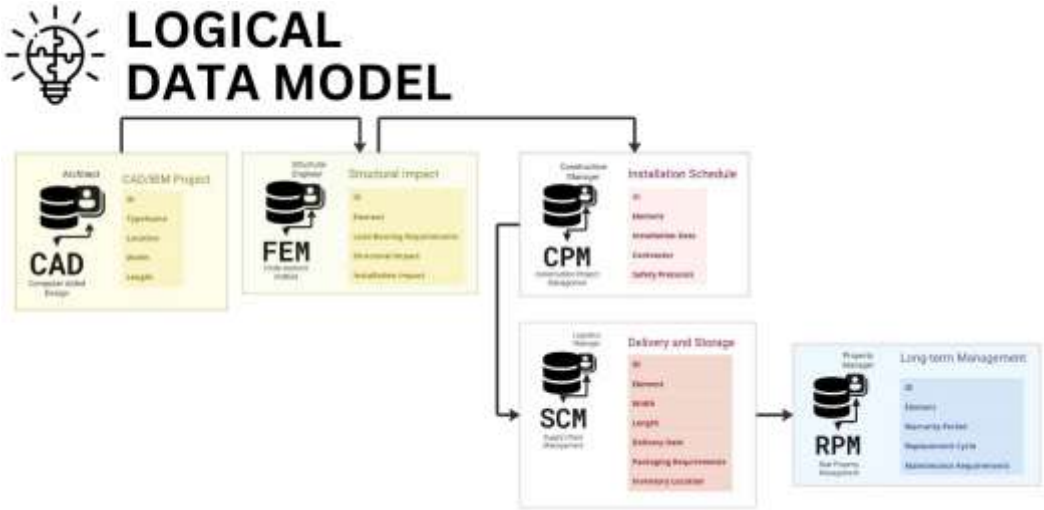


Abbildung 4.3-3 Logisches Datenmodell enthält Details zu Datentypen, Beziehungen und Schlüsseln, jedoch ohne Systemimplementierung.

- **Physisches Datenmodell**: Dieses Modell beschreibt die notwendigen Strukturen für die Implementierung einer Datenbank, einschließlich Tabellen, Spalten und Beziehungen. Es konzentriert sich auf die Datenbankleistung, Indizierungsstrategien und die physische Speicherung, um den physischen Einsatz von Datenbanken zu optimieren.

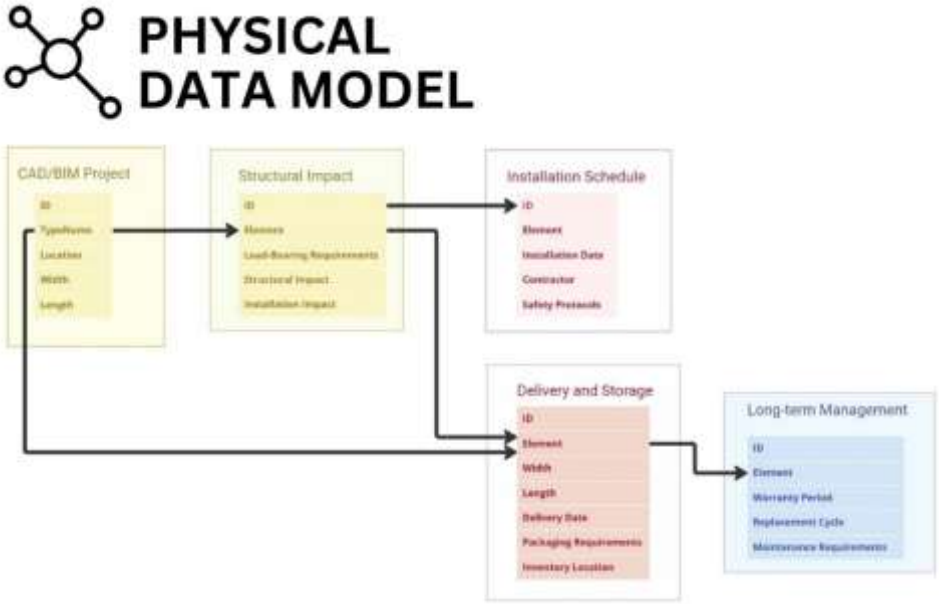


Abbildung 4.3-4 Das physische Datenmodell definiert, wie das System implementiert wird, einschließlich der Tabellen und spezifischen Details der Datenbank.

> Bei der Entwicklung von Datenbanken und der Gestaltung von tabellarischen Beziehungen spielt das Verständnis der Abstraktionsebenen eine wichtige Rolle beim Aufbau einer effektiven Systemarchitektur.

Eine effektive Datenmodellierungsmethodik ermöglicht es Ihnen, die Geschäftsziele mit der technischen Umsetzung zu verbinden, wodurch die gesamte Prozesskette transparenter und überschaubarer wird. Die Datenmodellierung ist keine einmalige Aufgabe, sondern ein Prozess mit aufeinander folgenden Schritten (Abb. 4.3-5):

- **Erfassen der Geschäftsanforderungen:** Die wichtigsten Aufgaben, Ziele und Informationsflüsse werden definiert. Dies ist die Phase der aktiven Interaktion mit Experten und Nutzern.
- **Identifizierung von Entitäten:** Die wichtigsten Objekte, Kategorien und Datentypen, die in dem künftigen System berücksichtigt werden müssen, werden hervorgehoben.
- **Entwicklung eines konzeptionellen und logischen Modells:** Zunächst werden die wichtigsten Einheiten und ihre Beziehungen erfasst, dann die Attribute, Regeln und die detaillierte Struktur.
- **Physische Modellierung:** Die technische Umsetzung des Modells wird entworfen: Tabellen, Felder, Beziehungen, Beschränkungen, Indizes.
- **Erstellung der Datenbank:** Der letzte Schritt ist die Implementierung des physischen Modells in das gewählte DBMS, die Durchführung von Tests und die Vorbereitung des Betriebs.

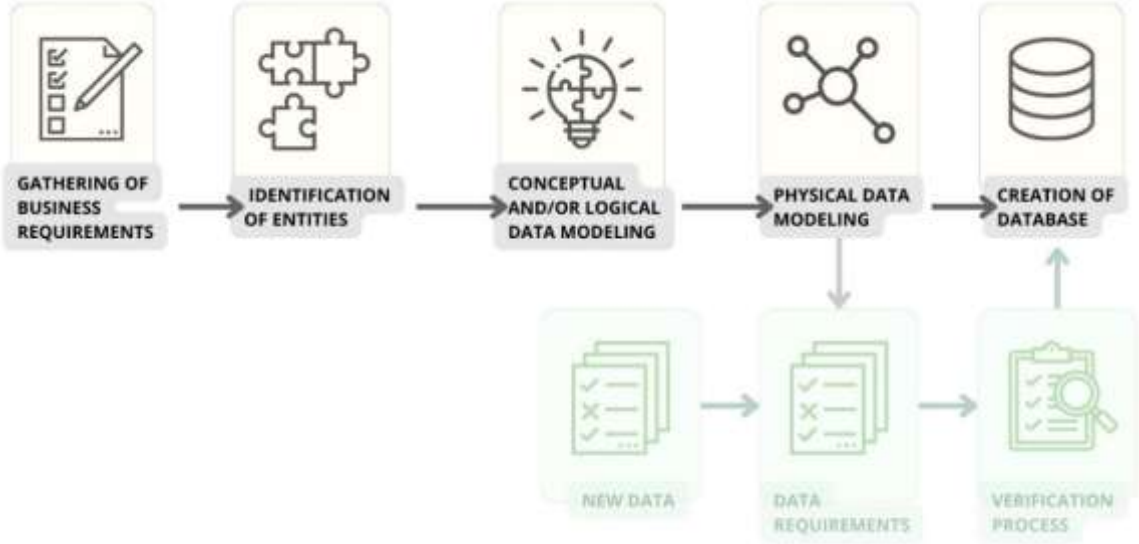


Abb. 4.3-5 Die Erstellung von Datenbanken und Datenmanagementsystemen für Geschäftsprozesse beginnt mit der Anforderungserstellung und Datenmodellierung.

Richtig konzipierte Datenmodellierungsprozesse ermöglichen transparente Informationsflüsse, was besonders bei komplexen Projekten wie dem Bauprojekt- oder Baustellenmanagement wichtig ist. Sehen wir uns an, wie der Übergang von einem konzeptionellen Modell zu einem logischen Modell und dann zu einem physischen Modell dazu beitragen kann, Prozesse zu rationalisieren.

Praktische Datenmodellierung im Kontext des Bauwesens

Nehmen wir eine Baustellenmanagementaufgabe als Beispiel für die Datenmodellierung und setzen wir die Anforderungen des Bauleiters in ein strukturiertes Logikmodell um. Ausgehend von den grundlegenden Anforderungen des Baustellenmanagements definieren wir Schlüsselentitäten für: Baustelle (SITE), Arbeiter (WORKER), Ausrüstung (EQUIPMENT), Aufgaben (TASK) und Ausrüstungsnutzung (EQUIPMENT_USAGE). Jede Entität enthält eine Reihe von Attributen, die wichtige Merkmale widerspiegeln. Für TASK könnte dies zum Beispiel eine Beschreibung der Aufgabe, Fälligkeitsdatum, Status, Priorität sein; für WORKER könnte dies der Name, seine Rolle auf der Baustelle, die aktuelle Beschäftigung usw. sein.

Das Logikmodell stellt Beziehungen zwischen diesen Einheiten her und zeigt, wie sie in realen Arbeitsprozessen miteinander interagieren (Abb. 4.3-6). Zum Beispiel zeigt die Beziehung zwischen Standort und Arbeitern, dass viele Arbeiter an einem Standort arbeiten können, während die Beziehung zwischen Arbeitern und Aufgaben widerspiegelt, dass ein Arbeiter mehrere Aufgaben ausführen kann.

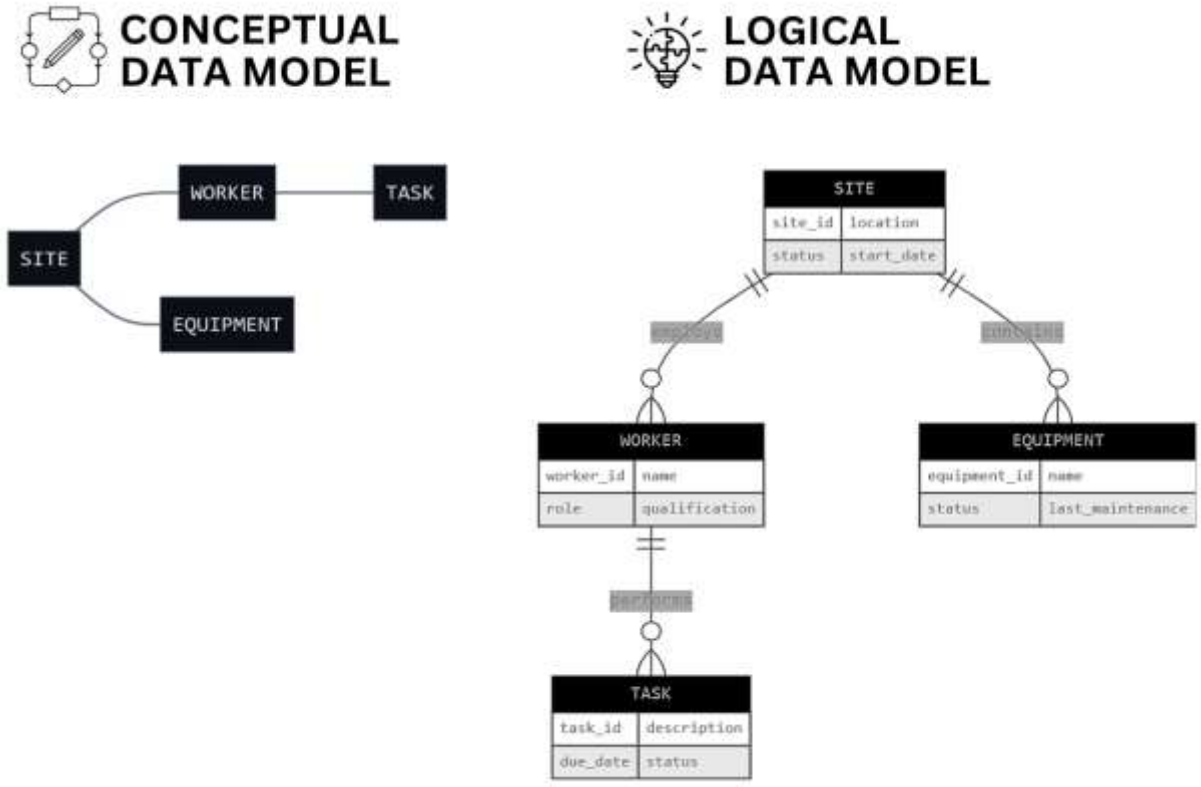


Abb. 4.3-6 Konzeptuelles und logisches Datenmodell, das aus den Anforderungen des Poliers zur Beschreibung der Baustellenprozesse entsteht.

Beim Übergang zum physischen Modell kommen technische Implementierungsdetails hinzu: spezifische Datentypen (VARCHAR, INT, DATE), Primär- und Fremdschlüssel für Beziehungen zwischen Tabellen und Indizes zur Optimierung der Datenbankleistung (Abb. 4.3-7).

So sollten beispielsweise für Status spezifische Typen mit möglichen Werten definiert und Indizes für Schlüsselfelder wie status und worker_id hinzugefügt werden, um die Suchleistung zu verbessern. So wird aus einer logischen Beschreibung des Systems ein konkreter Datenbankimplementierungsplan, der erstellt und implementiert werden kann.

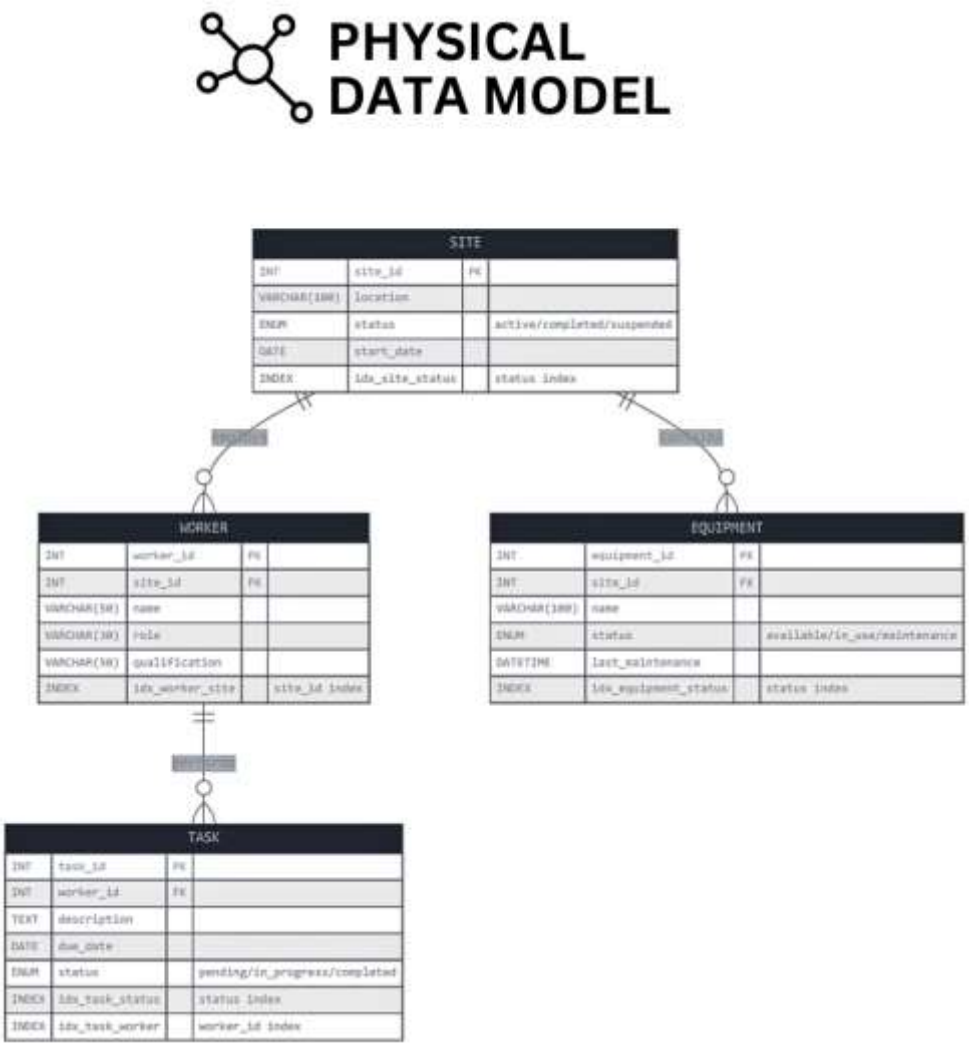


Abbildung 4.3-7 Das physische Datenmodell beschreibt die Entitäten einer Baustelle durch die minimal erforderlichen Parameter.

> Das physische Modell unterscheidet sich häufig vom logischen Modell. Im Durchschnitt verteilt sich die Modellierungszeit wie folgt: Etwa 50% entfallen auf das konzeptionelle Modell (Erfassen von Anforderungen, Diskussion von Prozessen, Identifizierung von Entitäten), 10% auf das logische Modell (Spezifizierung von Attributen und Beziehungen) und 40% auf das physische Modell (Implementierung, Testen, Anpassung an DBMS).

Diese Ausgewogenheit ist darauf zurückzuführen, dass in der konzeptionellen Phase die Grundlage für die Datenstruktur gelegt wird, während im logischen Modell lediglich Beziehungen und Attribute festgelegt werden. Das physische Modell erfordert die meisten Ressourcen, da in dieser Phase die Daten in spezifische Plattformen und Werkzeuge implementiert werden

Erstellen einer Datenbank mit LLM

Nach der Erstellung eines Datenmodells und der Beschreibung der Entitäten durch Parameter sind wir bereit, Datenbanken zu erstellen - Speicher, in denen wir die Informationen speichern, die nach der Strukturierungsphase zu bestimmten Prozessen kommen.

Versuchen wir, am Beispiel der Programmiersprache Python ein Beispiel für eine einfache, aber funktionale Datenbank mit einem Minimum an Code mit SQLite zu erstellen. Relationale Datenbanken wurden im Kapitel "Strukturierte relationale Datenbanken und die Abfragesprache SQL " ausführlich besprochen.

> SQLite ist eine leichtgewichtige, quelloffene, einbettbare, plattformübergreifende Datenbank, die grundlegende SQL-Funktionen unterstützt. SQLite ist weit verbreitet in CAD (BIM), ERP, EPM und ist in fast jedem Smartphone vorhanden.

Bitten wir das LLM-Modell, eine kleine SQLlite-Datenbank für uns zu erstellen, die auf dem oben erstellten physischen Modell (Abb. 4.3-4) für die Systeme FEM (strukturelle Auswirkungen) und CPM (Montageplan) basiert.

- Senden Sie eine Textanfrage an LLM chat (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN oder andere):

> Erstellen Sie SQLlite-Datenbanken für die Systeme "Constructive Impact" und "Installation Schedule" anhand der Abbildung im Anhang (fügen Sie das Schema als Bild in den Chat ein (Abb. 4.3-4)⏎

Die Antwort von LLM:

Create in LLM Chat

Database SQLite

```python
import sqlite3

# Connect to SQLite database (or create if it doesn't exist)
conn = sqlite3.connect('construction.db')

# Create a cursor object using the cursor() method
cursor = conn.cursor()

# Create table - Structural Impact
cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    LoadBearingRequirements TEXT,
                    StructuralImpact TEXT,
                    InstallationImpact TEXT
                )''')

# Create table - Installation Schedule
cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    InstallationDate DATE,
                    Contractor TEXT,
                    SafetyProtocols TEXT
                )''')

# Commit the changes
conn.commit()

# Close the connection
conn.close()

print("Database and tables created successfully.")
```

Run in IDE

Abbildung 4.3-8 Der generierte Code erstellt aus dem hochgeladenen Bild Datenbanken und Tabellen, in die wir nun Daten laden können.

Dieser Datenbankcode kann in der Python-Umgebung mit SQLite-Unterstützung ausgeführt werden, die normalerweise in der Standard-Python-Umgebung enthalten ist. Er kann auf jedem lokalen Python-Rechner ausgeführt werden, der in der Lage ist, Skripte auszuführen und Dateien zu erstellen. Nach der Ausführung erstellt dieser Code eine Datei auf Ihrer Festplatte - eine SQLite-Datenbank mit dem Namen construction.db (ähnlich einer Excel-Datei mit Arbeitsblättern), die die Tabellen "Structural Impact" und "Installation Schedule" enthält.

Sobald eine neue **construction.db-Datenbank** in diesen Tabellen erstellt wurde, können Daten über SQL _Abfragen oder Import zu diesen Tabellen hinzugefügt werden, was Ihnen später die Erstellung einer automatischen Datenverarbeitung ermöglicht. Daten können aus CSV-Dateien, Excel-Tabellen in die SQLite-Datenbank importiert oder aus anderen Datenbanken und Repositories über die API exportiert werden.

Um eine nachhaltige Datenmodellierung und wirksame Datenbankverwaltungsprozesse zu etablieren, benötigt ein Unternehmen eine klar definierte Strategie sowie eine Koordination zwischen technischen und geschäftlichen Teams. Bei disparaten Projekten und mehreren Datenquellen ist es oft schwierig, Konsistenz, Standardisierung und Qualitätskontrolle auf allen Ebenen zu gewährleisten. Eine wichtige Lösung kann darin bestehen, innerhalb des Unternehmens ein spezielles Kompetenzzentrum für Datenmodellierung (COE) einzurichten.

Exzellenzzentrum (CoE) für Datenmodellierung

Angesichts der Tatsache, dass Daten zu einem der wichtigsten strategischen Vermögenswerte werden, müssen Unternehmen mehr tun, als nur Informationen korrekt zu sammeln und zu speichern - es ist wichtig zu lernen, wie man Daten systematisch verwaltet. Das Kompetenzzentrum für Klassifizierung und Datenmodellierung (CoE) ist eine Struktureinheit, die die Konsistenz, Qualität und Effizienz der gesamten Datenverarbeitung in der Organisation sicherstellt.

> Das Centre of Excellence (CoE) ist der Kern der fachlichen Unterstützung und die methodische Grundlage für die digitale Transformation in einem Unternehmen. Es baut eine datengesteuerte Kultur auf und ermöglicht es Organisationen, Prozesse zu entwickeln, die Entscheidungen auf der Grundlage strukturierter, validierter und repräsentativer Daten treffen, statt auf Intuition oder lokalen Informationen.

Ein Exzellenz-Rechenzentrum wird in der Regel aus funktionsübergreifenden Teams gebildet, die nach dem Prinzip der "zwei Pizzen" arbeiten. Dieses von Jeff Bezos vorgeschlagene Prinzip bedeutet, dass das Team so groß sein sollte, dass es mit zwei Pizzen gefüttert werden kann, d. h. nicht mehr als 6-10 Personen. Dieser Ansatz hilft, übermäßige Bürokratie zu vermeiden und erhöht die Flexibilität der Arbeit. Dem CoE-Team sollten Mitarbeiter mit unterschiedlichen technischen Fähigkeiten angehören, von der Datenanalyse über maschinelles Lernen bis hin zu Fachwissen in bestimmten Geschäftsbereichen. Mit ihrem fundierten technischen Wissen sollten die Dateningenieure nicht nur Prozesse optimieren und Daten modellieren, sondern auch die Kollegen unterstützen, indem sie Zeit für Routineaufgaben einsparen (Abb. 4.3-9).

So wie in der Natur die Widerstandsfähigkeit von Ökosystemen durch die biologische Vielfalt gewährleistet wird, werden in der digitalen Welt Flexibilität und Anpassungsfähigkeit durch eine Vielfalt von Ansätzen für den Umgang mit Daten erreicht. Diese Vielfalt muss jedoch durch gemeinsame Regeln und Konzepte untermauert werden.

Das Centre of Excellence (CoE) kann mit den "klimatischen Bedingungen" eines Waldökosystems verglichen werden, die bestimmen, welche Arten von Daten gedeihen und welche automatisch aussortiert werden. Indem das CoE ein günstiges "Klima" für Qualitätsdaten schafft, erleichtert es die natürliche Auswahl der besten Verfahren und Methoden, die dann zu Standards für die Organisation werden.

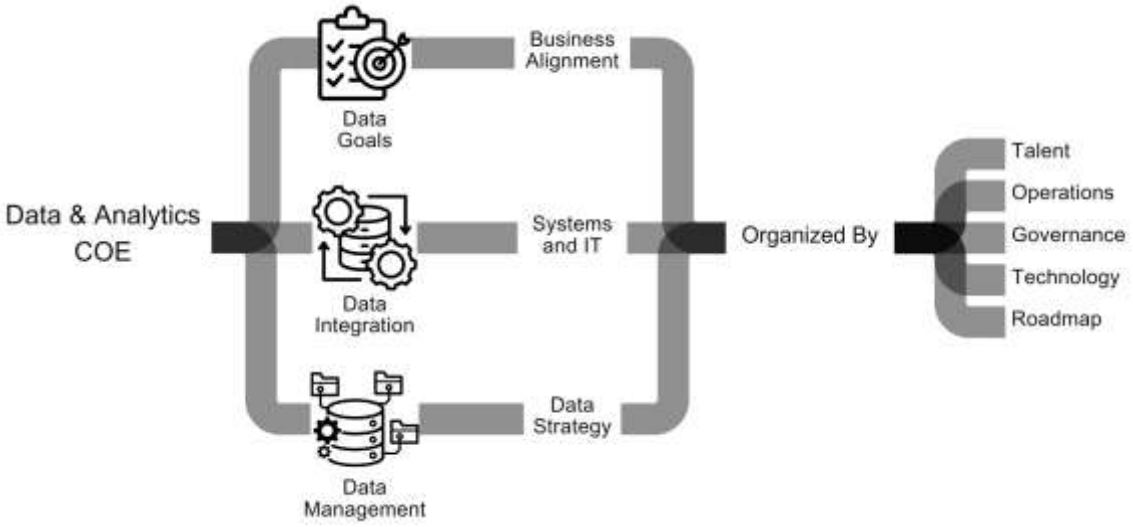


Abb. 4.3-9 Das Centre of Excellence (CoE) für Daten und Analytik bündelt Fachwissen zu wichtigen Aspekten der Datenverwaltung, -integration und -strategie.

Um die Integrationszyklen zu beschleunigen und bessere Ergebnisse zu erzielen, sollte der CoE seinen Mitgliedern ein ausreichendes Maß an Autonomie bei der Entscheidungsfindung bieten. Dies ist besonders wichtig in einem dynamischen Umfeld, in dem Versuch und Irrtum, ständiges Feedback und häufige Freigaben erhebliche Vorteile bringen können. Diese Autonomie ist jedoch nur dann wirksam, wenn es eine klare Kommunikation und Unterstützung durch die Unternehmensleitung gibt. Ohne strategische Vision und Koordination auf höchster Ebene kann selbst das kompetenteste Team bei der Umsetzung seiner Initiativen auf Hindernisse stoßen.

Das COE oder die Geschäftsleitung des Unternehmens ist dafür verantwortlich, dass der Datenmodellierungsansatz nicht auf ein oder zwei Projekte beschränkt bleibt, sondern in das gesamte Informationsmanagement- und Geschäftsprozessmanagementsystem eingebettet wird.

Das Centre of Expertise (CoE) ist neben den Aufgaben im Zusammenhang mit der Datenmodellierung und der Data Governance für die Entwicklung gemeinsamer Standards und Ansätze für die Einrichtung und den Betrieb der Dateninfrastruktur zuständig. Darüber hinaus fördert es eine Kultur der kontinuierlichen Verbesserung, der Prozessoptimierung und der effizienten Nutzung von Daten in der Organisation (Abb.

4.3-10).

Der systematische Ansatz für das Daten- und Modellmanagement innerhalb des CoE lässt sich grob in mehrere Schlüsselbereiche unterteilen:

- **Standardisierung von Prozessen und Modell-Lebenszyklus-Management:** CoE entwickelt und implementiert Methoden zur Vereinheitlichung der Erstellung und Verwaltung von Datenmodellen. Dazu gehören: die Einrichtung von Strukturvorlagen, Qualitätskontrollmethoden und Versionskontrollsystemen zur Gewährleistung der Datenkontinuität in allen Arbeitsphasen.
- **Rollenmanagement und Zuweisung von Zuständigkeiten:** Das COE definiert die Schlüsselrollen im Datenmodellierungsprozess. Jedem Projektteilnehmer werden klar definierte Rollen und Verantwortungsbereiche zugewiesen, wodurch die Teamarbeit erleichtert und das Risiko von Dateninkonsistenzen verringert wird.
- **Qualitätskontrolle und Audit:** Eine wirksame Verwaltung von Baudaten erfordert eine kontinuierliche Überwachung ihrer Qualität. Es werden automatisierte Mechanismen zur Überprüfung der Daten, zur Ermittlung von Fehlern und fehlenden Attributen eingeführt.
- **Metadaten- und Informationsarchitekturmanagement:** CoE ist für die Schaffung eines einheitlichen Systems von Klassifizierungs- und Identifikatoren, Benennungs- und Entitätsbeschreibungsstandards zuständig, die für die Integration zwischen Systemen entscheidend sind.

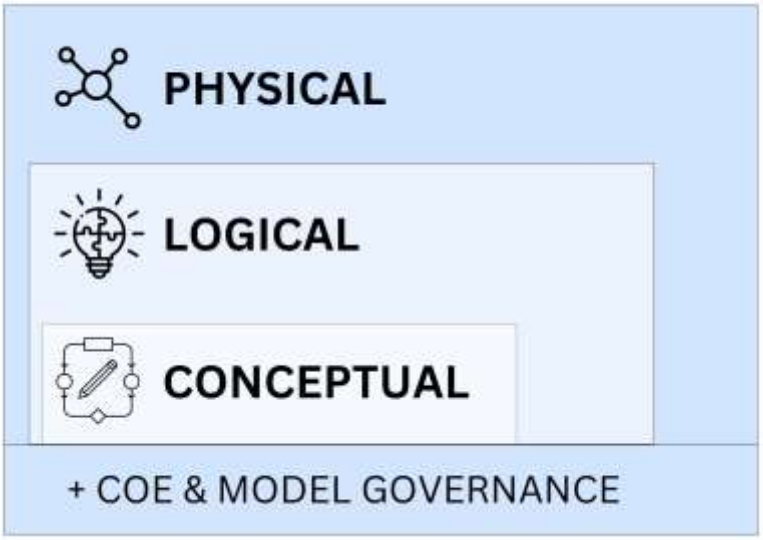


Abb. 4.3-10 Datenmodellierung und Datenqualitätsmanagement ist eine der größten Herausforderungen des CoE

Das Centre of Excellence (CoE) für Daten ist nicht nur eine Gruppe von Experten, sondern ein systemischer Mechanismus, der eine neue datengesteuerte Kultur schafft und einen einheitlichen Ansatz für die Arbeit mit Daten im gesamten Unternehmen gewährleistet. Durch kompetente Integration von Modellierungsprozessen in das gesamte Informationsmanagementsystem, Standardisierung, Klassifizierung und Datenqualitätskontrolle hilft das CoE den Unternehmen, ihre Produkte und Geschäftsprozesse kontinuierlich zu verbessern, schneller auf Marktveränderungen zu reagieren und fundierte Entscheidungen auf der Grundlage zuverlässiger Analysen zu treffen.

Solche Zentren sind besonders effektiv, wenn sie mit modernen DataOps-Prinzipien kombiniert werden - unter ein Schritt, der die kontinuierliche Bereitstellung, Automatisierung und Qualitätskontrolle von Daten

gewährleistet. Wir werden in Teil 8, im Kapitel "Moderne Datentechnologien in der Bauwirtschaft", mehr über DataOps sprechen.

In den folgenden Kapiteln werden wir von der Strategie zur Praxis übergehen - lassen Sie uns bedingt in ein Rechenzentrum "transformieren": Wir werden uns mehrere Beispiele ansehen, wie die Aufgabenparametrisierung, die Anforderungserfassung und der automatische Validierungsprozess ablaufen.

KAPITEL 4.4.
SYSTEMATISIERUNG DER ANFORDERUNGEN UND VALIDIERUNG DER INFORMATIONEN

Erfassung und Analyse von Anforderungen: Umwandlung von Mitteilungen in strukturierte Daten

Das Sammeln und Verwalten von Anforderungen ist der erste Schritt zur Gewährleistung der Datenqualität. Trotz der Entwicklung digitaler Werkzeuge werden die meisten Anforderungen immer noch auf unstrukturierte Weise formuliert: durch Briefe, Sitzungsprotokolle, Telefonate und mündliche Diskussionen. Diese Form der Kommunikation erschwert die Automatisierung, Validierung und Wiederverwendung von Informationen. In diesem Kapitel befassen wir uns mit der Frage, wie textliche Anforderungen in formale Strukturen übersetzt werden können, um sicherzustellen, dass die Geschäftsanforderungen transparent und systematisch sind.

> Die Studie "Data Quality: Best Practices for Accurate Insights" von Gartner unterstreicht die entscheidende Bedeutung der Datenqualität für erfolgreiche Daten- und Analyseinitiativen [100]. Sie stellen fest, dass schlechte Datenqualität Unternehmen im Durchschnitt mindestens 12,9 Millionen US-Dollar pro Jahr kostet und dass zuverlässige, qualitativ hochwertige Daten für die Schaffung eines datengesteuerten Unternehmens unerlässlich sind.

Das Fehlen strukturierter Anforderungen führt dazu, dass ein und dasselbe Element (Entität) und seine Parameter in verschiedenen Systemen in unterschiedlichen Varianten gespeichert sein können. Dies mindert nicht nur die Effizienz der Prozesse, sondern führt auch zu Zeitverschwendung, doppelten Informationen und der Notwendigkeit, Daten erneut zu validieren, bevor sie verwendet werden können. So kann schon ein einziges Versäumnis - ein verlorener Parameter oder ein einziges falsch beschriebenes Element - die Entscheidungsfindung verlangsamen und eine ineffiziente Ressourcennutzung verursachen.

> *In Ermangelung eines Nagels war das Hufeisen verloren.*
> *In Ermangelung eines Hufeisens war das Pferd verloren.*
> *In Ermangelung eines Pferdes war der Reiter verloren.*
> *In Ermangelung eines Reiters war eine Nachricht verloren.*
> *Ohne eine Botschaft war die Schlacht verloren.*
> *In Ermangelung einer Schlacht ging ein Königreich verloren.*
> *Alles wegen eines fehlenden Nagels im Hufeisen.*
>
> - Sprichwort [101]

Die Analyse und Erfassung von Anforderungen an den Prozess der Befüllung und Speicherung von Daten beginnt mit der Identifizierung aller Beteiligten. So wie der sprichwörtliche Verlust eines einzelnen Nagels zu einer Kette kritischer Konsequenzen führt, kann in der Wirtschaft der Verlust eines einzelnen Beteiligten, einer übersehenen Anforderung oder auch nur eines Parameters nicht nur einen einzelnen Geschäftspro-

zess, sondern das gesamte Ökosystem eines Projekts und die Organisation als Ganzes erheblich beeinträchtigen. Daher ist es von entscheidender Bedeutung, auch solche Elemente, Parameter und Rollen zu identifizieren, die auf den ersten Blick unbedeutend erscheinen (), sich aber später als entscheidend für die Nachhaltigkeit des Unternehmens erweisen können.

Stellen wir uns vor, dass ein Unternehmen ein Projekt hat, bei dem der Kunde eine neue Anfrage stellt - "*ein zusätzliches Fenster auf der Nordseite des Gebäudes hinzufügen*". An dem kleinen Prozess "*Antrag des Kunden auf Hinzufügung eines neuen Fensters zum laufenden Projekt*" sind der Architekt, der Kunde, der CAD-Spezialist (BIM), der Bauleiter, der Logistikmanager, der ERP-Analyst, der Ingenieur für Qualitätskontrolle, der Sicherheitsingenieur, der Kontrollmanager und der Immobilienmanager beteiligt.

> Selbst an einem kleinen Prozess können Dutzende von verschiedenen Spezialisten beteiligt sein. Jeder Prozessbeteiligte muss die Anforderungen der Spezialisten verstehen, mit denen er auf der Datenebene verbunden ist.

Auf der Textebene (Abb. 4.4-1) läuft die Kommunikation zwischen dem Auftraggeber und den Spezialisten in der Prozesskette wie folgt ab:

- **Kunde:** "*Wir haben beschlossen, auf der Nordseite ein zusätzliches Fenster für eine bessere Beleuchtung einzubauen. Kann das realisiert werden*?"
- **Architekt**: "*Sicher, ich werde das Projekt überarbeiten, um das neue Fenster mit einzubeziehen, und aktualisierte CAD-Pläne (BIM) schicken*".
- **CAD-Spezialist (BIM):** *"Ich habe ein neues Projekt erhalten. Ich aktualisiere das CAD (BIM)-Modell mit dem zusätzlichen Fenster und gebe nach Abstimmung mit dem FEM-Ingenieur die genaue Position und die Abmessungen des neuen Fensters an*".
- **Bauleiter**: "*Wir haben ein neues Projekt erhalten. Wir passen die Termine für die Installation von 4D an und informieren alle betroffenen Subunternehmer*".
- **Gebäudetechniker (CAFM**): *"Ich werde die 6D-Daten des neuen Fensters in das CAFM-System für das zukünftige Facility Management und die Wartungsplanung eingeben*."
- **Logistikleiter**: "*Ich benötige die Abmessungen und das Gewicht des neuen Fensters, um die Lieferung des Fensters an die Baustelle zu organisieren*".
- **ERP** -analyst: "*Ich benötige die Umfangstabellen und den genauen Fenstertyp für die 5D Budgetaktualisierung in unserem ERP-System, um die Kosten des neuen Fensters in der Gesamtprojektschätzung zu berücksichtigen*."
- **Ingenieur für Qualitätskontrolle**: "*Sobald die Fensterspezifikationen fertig sind, werde ich sicherstellen, dass sie unseren Qualitäts- und Materialstandards entsprechen.*"
- **Sicherheitsingenieur**: "*Ich werde die Sicherheitsaspekte des neuen Fensters bewerten, mit besonderem Augenmerk auf die Einhaltung der Vorschriften und die Evakuierung im Rahmen der 8D-Regelung "*.
- **Leiter der Steuerung**: "*Auf der Grundlage des genauen Arbeitsumfangs von ERP werden wir unseren 4D-Zeitplan aktualisieren, um die Installation des neuen Fensters zu berücksichtigen, und die neuen Daten im Content-Management-System des Projekts speichern*."
- **Arbeiter (Installateur)**: *"Sie brauchen eine Anleitung für die Installation, die Montage und den Zeitplan der Arbeiten. Gibt es darüber hinaus besondere Sicherheitsvorschriften, die ich beachten muss?*"
- **Hausverwalter**: "*Nach der Installation werde ich die Garantie- und Wartungsinformationen für die langfristige Verwaltung dokumentieren*".
- **Vermögensverwalter**: "*Geräteingenieur, bitte senden Sie die endgültigen Daten für die Anlagenverfolgung und das Lebenszyklusmanagement.*"

- **Bauherr**: "*Moment, vielleicht habe ich es eilig und das Fenster wird nicht gebraucht. Vielleicht sollte ich einen Balkon bauen*".

In solchen Szenarien, die häufig vorkommen, löst selbst eine kleine Änderung eine Kettenreaktion zwischen mehreren Systemen und Rollen aus. In diesem Fall erfolgt fast die gesamte Kommunikation in der Anfangsphase in Textform: E-Mails, Chats, Sitzungsprotokolle (Abb. 4.4-1).

In einem solchen textbasierten Kommunikationssystem für ein Bauprojekt ist ein System zur rechtlichen Bestätigung und Aufzeichnung aller Datenaustauschvorgänge und aller getroffenen Entscheidungen sehr wichtig. Damit soll sichergestellt werden, dass jede Entscheidung, Anweisung oder Änderung rechtsgültig und nachvollziehbar ist, wodurch das Risiko künftiger "Missverständnisse" verringert wird

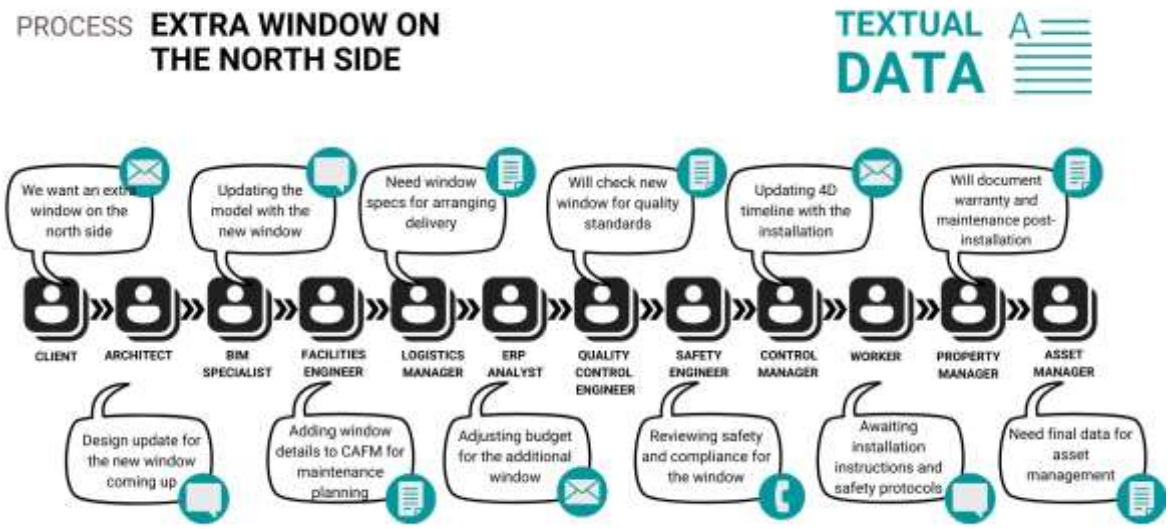


Abb. 4.4-1 Die Kommunikation zwischen Auftraggeber und Auftragnehmer in der Anfangsphase eines Projekts enthält häufig mehrformatige Textdaten.

> Das Fehlen einer rechtlichen Kontrolle und Validierung von Entscheidungen in den relevanten Systemen eines Bauprojekts kann zu ernsthaften Problemen für alle Beteiligten führen. Jede Entscheidung, Bestellung oder Änderung, die ohne ordnungsgemäße Dokumentation und Validierung getroffen wird, kann zu Streitigkeiten (und Rechtsstreitigkeiten) führen.

Die rechtliche Konsolidierung aller Entscheidungen in Textform kann nur durch eine große Anzahl unterzeichneter Dokumente gewährleistet werden, die auf den Schultern der Verwaltung lasten, die verpflichtet ist, alle Transaktionen aufzuzeichnen. Wenn jeder Beteiligte für jede Aktion ein Dokument unterschreiben muss, verliert das System an Flexibilität und wird zu einem bürokratischen Labyrinth. Fehlende Transaktionsbestätigungen verzögern nicht nur die Projektdurchführung, sondern können auch zu finanziellen Verlusten und einer Verschlechterung der Beziehungen zwischen den Beteiligten führen, bis hin zu rechtlichen Problemen.

Ein solcher Transaktionsgenehmigungsprozess, der in der Regel mit einer textbasierten Diskussion beginnt, entwickelt sich in den folgenden Phasen allmählich zu einem mehrformatigen Dokumentenaustausch (Abb. 4.4-2), was die Kommunikation, die bisher nur über Text erfolgte, erheblich erschwert. Ohne klar definierte Anforderungen ist eine Automatisierung solcher Prozesse, die mit Multiformat-Daten und einer Vielzahl von

textlichen Anforderungen gefüllt sind, kaum möglich.

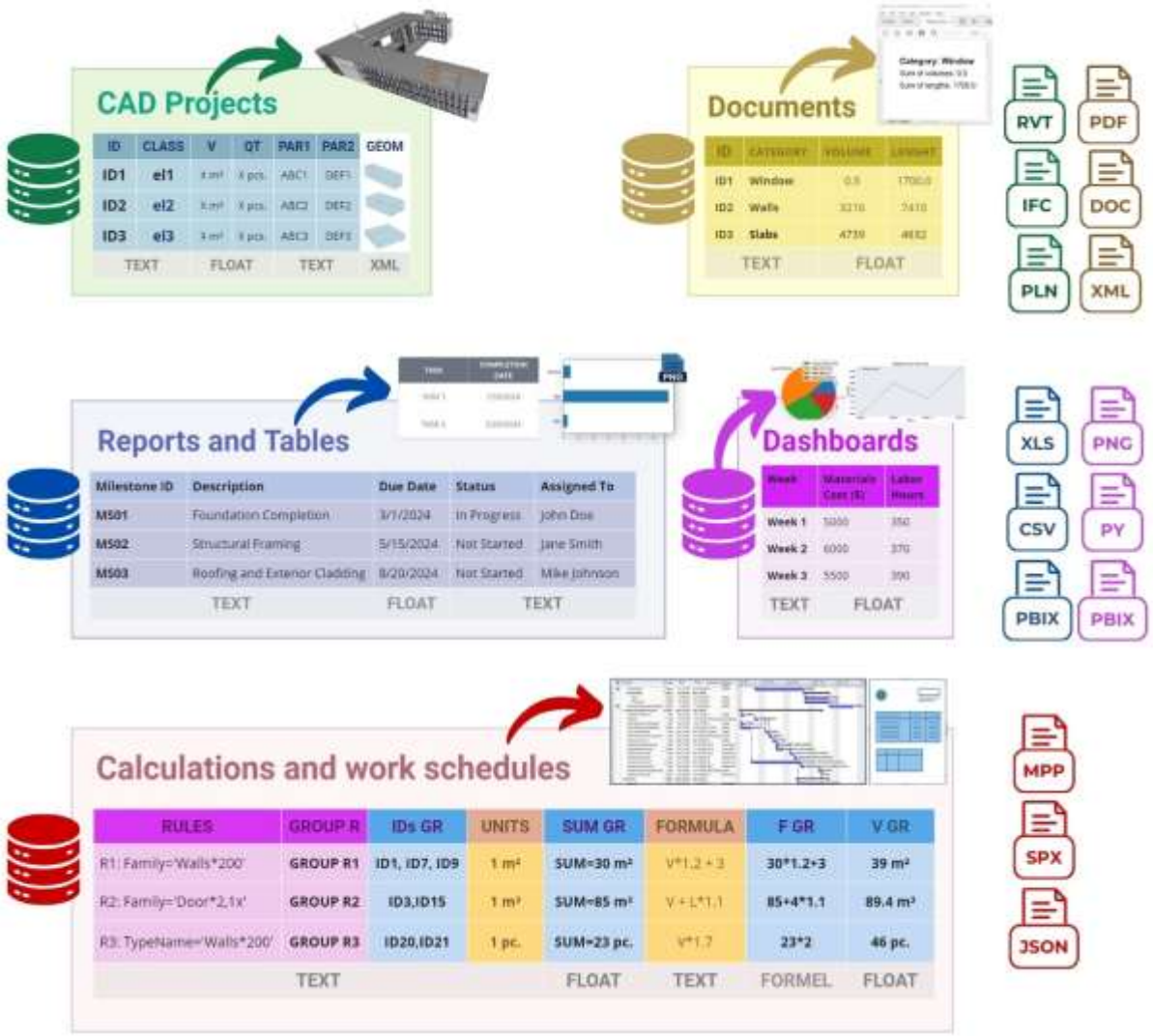


Abb. 4.4-2 Jedes System in der Landschaft der Bauunternehmen dient als Quelle für rechtlich relevante Dokumente in unterschiedlichen Formaten.

Die Textkommunikation erfordert, dass sich jeder Fachmann entweder mit dem gesamten Schriftverkehr vertraut macht oder regelmäßig an allen Sitzungen teilnimmt, um sich über den aktuellen Stand des Projekts zu informieren.

Um diese Einschränkung zu überwinden, ist ein Übergang von der textuellen Kommunikation zu einem strukturierten Anforderungsmodell notwendig. Dies ist nur durch systematische Analyse, Prozessvisualisierung und Beschreibung von Wechselwirkungen in Form von Flussdiagrammen und Datenmodellen möglich (Abb. 4.4-3). Wie bei der Datenmodellierung (Abb. 4.3-7) wurde von der kontextuell-ideellen Ebene auf die konzeptionelle Ebene übergegangen, indem die von den Beteiligten verwendeten Systeme und Werkzeuge sowie die Verbindungen zwischen ihnen hinzugefügt wurden.

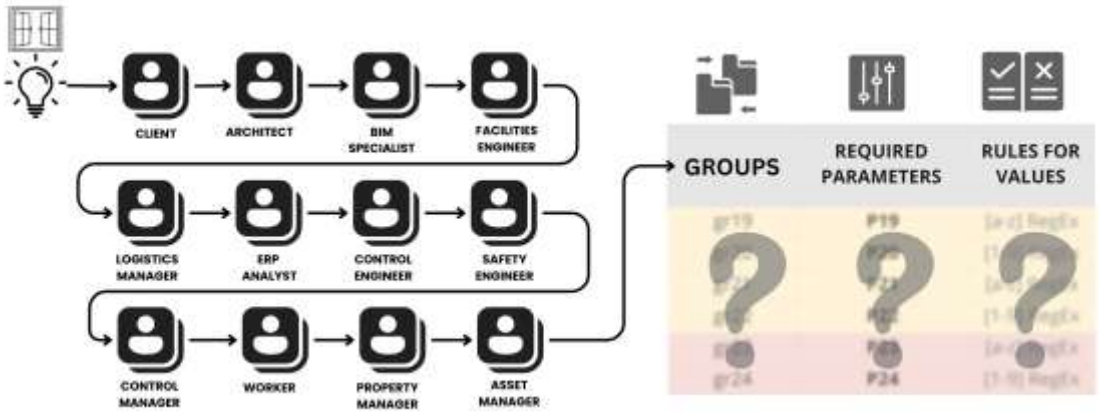


Abb. 4.4-3 Um zu lernen, wie man den Validierungsprozess verwaltet und automatisiert, ist es notwendig, die Prozesse zu visualisieren und die Anforderungen zu strukturieren.

Der erste Schritt zur Systematisierung von Anforderungen und Beziehungen besteht darin, alle Verknüpfungen und Beziehungen mit Hilfe von konzeptionellen Flussdiagrammen zu visualisieren. Die konzeptionelle Ebene erleichtert es nicht nur allen Prozessbeteiligten, die gesamte Prozesskette zu verstehen, sondern zeigt auch klar auf, warum und für wen die Daten (und Anforderungen) bei jedem Prozessschritt benötigt werden.

Prozessablaufdiagramme und die Wirksamkeit von konzeptionellen Rahmen

Um die Kluft zwischen traditionellen und modernen Ansätzen des Datenmanagements zu überbrücken, müssen Unternehmen bewusst von fragmentierten Textbeschreibungen zu strukturierten Prozessdarstellungen übergehen. Die Entwicklung von Daten - von Tontafeln zu digitalen Ökosystemen - erfordert neue Denkwerkzeuge. Ein solches Werkzeug ist die konzeptionelle Modellierung mit Flussdiagrammen. Die Erstellung visueller Diagramme - Flussdiagramme, Prozessdiagramme, Interaktionsdiagramme - ermöglicht den Projektteilnehmern zu erkennen, wie sich ihre Handlungen und Entscheidungen auf das gesamte Entscheidungssystem auswirken.

> Wenn Prozesse nicht nur die Speicherung von Daten, sondern auch deren Analyse oder Automatisierung erfordern, müssen Sie sich mit dem Thema der Erstellung einer konzeptionellen und visuellen Anforderungsschicht auseinandersetzen.

In unserem Beispiel (Abb. 4.4-1) kann jeder Spezialist Teil eines kleinen Teams sein, aber auch einer größeren Abteilung angehören, die bis zu einem Dutzend Experten unter der Leitung eines Generaldirektors umfasst. Jede Abteilung verwendet eine spezialisierte Anwendungsdatenbank (Abb. 1.2-4 z. B. ERP, CAD, MEP, CDE, ECM, CPM usw.), die regelmäßig mit eingehenden Informationen aktualisiert wird, die für die Erstellung von Dokumenten, die Erfassung des rechtlichen Status von Entscheidungen und die Verwaltung von Prozessen erforderlich sind.

> Der Transaktionsprozess ähnelt der Arbeit der antiken Manager vor 4.000 Jahren, als Tontafeln und Papyrus verwendet wurden, um Entscheidungen rechtlich zu bestätigen. Der Unterschied zwischen modernen Systemen und ihren Vorgängern aus Ton und Papier besteht darin, dass moderne Methoden zusätzlich den Prozess der Umwandlung von Textinformationen in digitale Form zur weiteren automatischen Verarbeitung in anderen Systemen und Werkzeugen umfassen.

Die Visualisierung des Prozesses in Form von konzeptionellen Flussdiagrammen hilft dabei, jeden Schritt und die Interaktionen zwischen den verschiedenen Rollen zu beschreiben, wodurch ein komplexer Arbeitsablauf klar und einfach wird.

> Die Visualisierung von Prozessen stellt sicher, dass die Prozesslogik transparent und für alle Teammitglieder zugänglich ist.

Der gleiche kommunikative Prozess zum Hinzufügen eines Fensters zu einem Projekt, der in Form von Text, Nachrichten (Abb. 4.4-1) und Blockdiagramm beschrieben wurde, ähnelt dem konzeptionellen Modell, das wir im Kapitel über die Datenmodellierung (Abb. 4.4-4) besprochen haben.

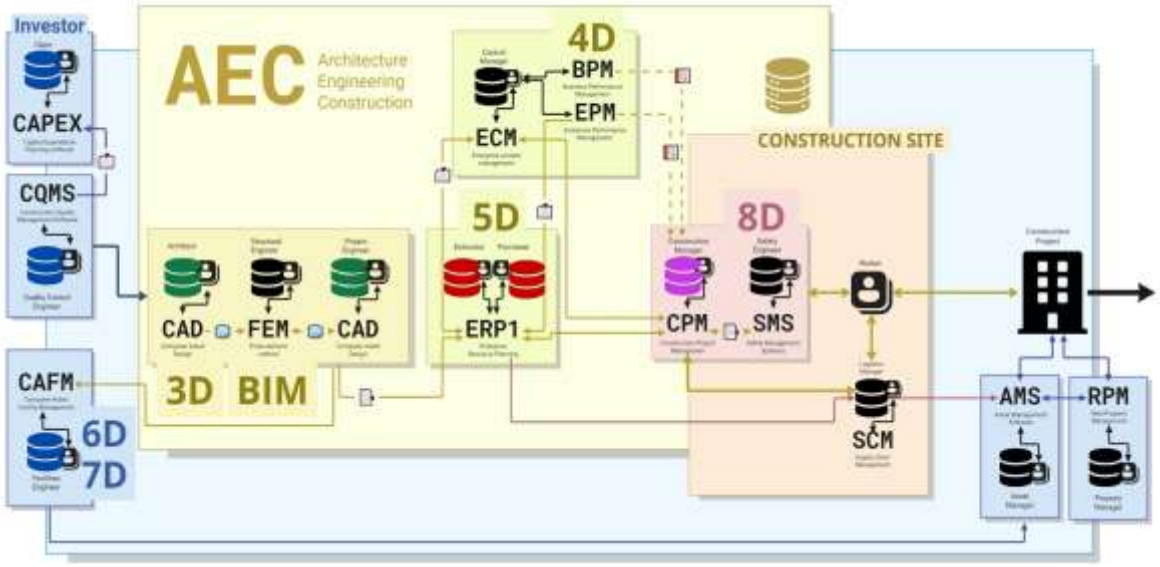


Abb. 4.4-4 Das konzeptionelle Diagramm zeigt die Projektteilnehmer als Nutzer einer Datenbank, deren Abfragen die verschiedenen Systeme miteinander verbinden.

Obwohl konzeptionelle Diagramme ein wichtiger Schritt sind, beschränken sich viele Unternehmen auf diese Ebene und glauben, dass ein visuelles Diagramm ausreicht, um die Prozesse zu verstehen. Dadurch entsteht die Illusion der Überschaubarkeit: Manager können auf einem solchen Flussdiagramm leichter das große Ganze erkennen und die Verbindungen zwischen den Beteiligten und den Phasen sehen. Allerdings vermitteln solche Schemata keine klare Vorstellung davon, welche Daten für jeden Teilnehmer erforderlich sind, in welchem Format sie übermittelt werden sollten und welche Parameter und Attribute für die Realisierung der Automatisierung zwingend erforderlich sind. Ein konzeptionelles Flussdiagramm ähnelt eher einer Routenkarte: Es zeigt an, wer mit wem interagiert, aber es verrät nicht, was bei diesen Interaktionen übertragen

wird.

> Selbst wenn ein Prozess auf konzeptioneller Ebene mit Hilfe von Flussdiagrammen detailliert beschrieben wird, ist dies keine Garantie für seine Wirksamkeit. Die Visualisierung vereinfacht häufig die Arbeit von Managern, da sie den Prozess anhand von Schritt-für-Schritt-Berichten leichter verfolgen können. Für Datenbankingenieure ist die konzeptionelle Darstellung jedoch möglicherweise nicht klar genug und vermittelt kein klares Verständnis dafür, wie der Prozess auf der Ebene der Parameter und Anforderungen zu implementieren ist.

In dem Maße, wie wir uns auf komplexere Datenökosysteme zubewegen, wird die anfängliche Implementierung von konzeptionellen und visuellen Werkzeugen entscheidend, um sicherzustellen, dass die Datenprozesse nicht nur effizient sind, sondern auch mit den strategischen Zielen der Organisation übereinstimmen. Um diesen Prozess der Fensterergänzung (Abb. 4.4-1) vollständig auf die Ebene der Datenanforderungen zu übertragen, müssen wir eine Ebene tiefer gehen und die konzeptionelle Visualisierung des Prozesses auf die logische und physische Ebene der Daten, der erforderlichen Attribute und ihrer Grenzwerte übertragen.

Strukturierte Anforderungen und reguläre Ausdrücke RegEx

Bis zu 80% der in Unternehmen anfallenden Daten liegen in unstrukturierten oder halbstrukturierten Formaten vor [52] - Texte, Dokumente, Briefe, PDF -Dateien, Konversationen. Solche Daten (Abb. 4.4-1) sind schwer zu analysieren, zu verifizieren, zwischen Systemen zu übertragen und für die Automatisierung zu nutzen.

Um Handhabbarkeit, Transparenz und automatische Validierung zu gewährleisten, ist es notwendig, textuelle und halbstrukturierte Anforderungen in wohldefinierte, strukturierte Formate zu übersetzen. Der Strukturierungsprozess betrifft nicht nur die Daten (die wir in den ersten Kapiteln dieses Teils des Buches ausführlich besprochen haben), sondern auch die Anforderungen selbst, die die Projektteilnehmer in der Regel während des gesamten Projektlebenszyklus in freier Textform formulieren, oft ohne daran zu denken, dass diese Prozesse automatisiert werden können.

> So wie wir bereits Daten von einer unstrukturierten textuellen Form in eine strukturierte Form umgewandelt haben, werden wir im Anforderungsworkflow textuelle Anforderungen in ein strukturiertes Format der "logischen und physikalischen Schicht" umwandeln.

Als Teil des Beispiels für das Hinzufügen eines Fensters (Abb. 4.4-1) besteht der nächste Schritt darin, die Datenanforderungen in Tabellenform zu beschreiben. Wir werden die Informationen für jedes von den Projektteilnehmern verwendete System strukturieren, indem wir Schlüsselattribute und ihre Grenzwerte angeben

Betrachten wir zum Beispiel ein solches System (Abb. 4.4-5) - das Construction Quality Management System (CQMS), das vom Qualitätskontrollingenieur auf der Seite des Auftraggebers verwendet wird. Mit seiner Hilfe prüft er, ob ein neues Element des Projekts - in diesem Fall das "neue Fenster" - den festgelegten Normen und Anforderungen entspricht.

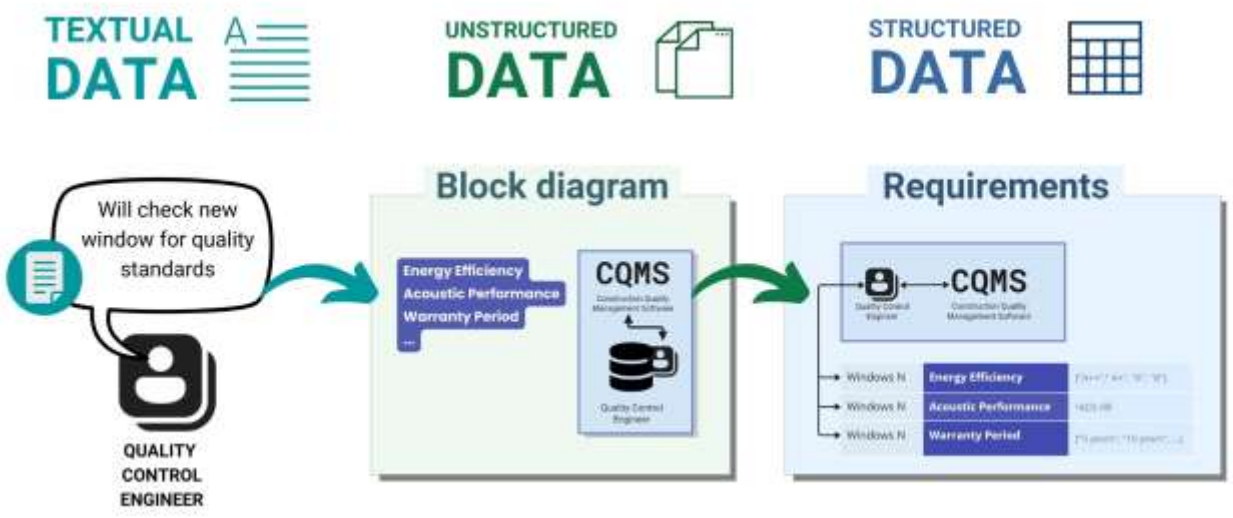


Abb. 4.4-5 Die Umwandlung von textlichen Anforderungen in ein Tabellenformat mit Beschreibungen der Entitätsattribute vereinfacht das Verständnis für andere Spezialisten.

Betrachten wir als Beispiel einige wichtige Anforderungen für die Attribute von Einheiten des Typs "Fenstersysteme" im CQMS -System (Abb. 4.4-6): Energieeffizienz, akustische Leistung und Garantiezeit. Jede Kategorie umfasst bestimmte Normen und Spezifikationen, die bei der Planung und dem Einbau von Fenstersystemen berücksichtigt werden müssen.

Abb. 4.4-6 Der Qualitätskontrolleur sollte neue Fensterelemente auf Energieeffizienz, Schalldämmung und Garantiestandards prüfen.

Die Datenanforderungen, die ein Qualitätskontrolleur in Form einer Tabelle festlegt, haben beispielsweise folgende Grenzwerte:

- **Die Energieeffizienzklasse von Fenstern** reicht von "A++", der höchsten Effizienzklasse, bis zu "B", dem niedrigsten annehmbaren Niveau, und diese Klassen werden durch eine Liste annehmbarer Werte dargestellt **["A++", "A+", "A", "A", "A", "B"].**

- **Die Schalldämmung von Fenstern**, die in Dezibel gemessen wird und ihre Fähigkeit zur Reduzierung des Straßenlärms angibt, wird durch den regelmäßigen Ausdruck \d{2}dB definiert**.**
- **Das Attribut "Gewährleistungsfrist"** für die Entität "Fenstertyp" beginnt bei fünf Jahren und legt diesen Zeitraum als Mindestzeitraum fest, der bei der Auswahl eines Produkts zulässig ist; es werden auch Werte für die Gewährleistungsfrist wie **["5 Jahre", "10 Jahre" usw.]** oder die logische Bedingung "**>5 (Jahre)**" angegeben**.**

Gemäß den gesammelten Anforderungen, innerhalb der festgelegten Attribute, werden neue Fensterkategorien oder Klassenelemente mit Noten unter "B", wie "C" oder "D", die Energieeffizienzprüfung nicht bestehen. Die Schalldämmung von Fenstern in Daten oder Dokumenten für den QS-Ingenieur muss mit einer zweistelligen Zahl, gefolgt von der Nachsilbe "dB", gekennzeichnet sein, wie z. B. "35 dB" oder "40 dB", und Werte außerhalb dieses Formats, wie z. B. "9 D B" oder "100 Dezibel", werden nicht akzeptiert (da sie nicht dem Muster für RegEx-Strings entsprechen). Die Garantiezeit muss mit mindestens "5 Jahren" beginnen, und Fenster mit kürzeren Garantiezeiten wie "3 Jahre" oder "4 Jahre" erfüllen nicht die Anforderungen, die der Qualitätsingenieur im Tabellenformat beschrieben hat.

Um solche Attribut-Parameterwerte im Validierungsprozess gegen Grenzwerte aus Anforderungen zu prüfen, verwenden wir entweder eine Liste zulässiger Werte (["A", "B", "C"]), Wörterbücher (["A": "H1", "H2"; "B": W1", "W2"]), logische Operationen (z. B. ">", "<", "<=", ">=" "==") für numerische Werte) und reguläre Ausdrücke (für String- und Textwerte wie im Attribut "Acoustic Performance"). Reguläre Ausdrücke sind ein äußerst wichtiges Werkzeug bei der Arbeit mit String-Werten.

> **Reguläre Ausdrücke** (RegEx) werden in Programmiersprachen, einschließlich Python (Re-Bibliothek), verwendet, um Zeichenketten zu finden und zu verändern. Regex ist wie ein Detektiv in der Welt der Zeichenketten, der in der Lage ist, Textmuster in Texten präzise zu erkennen.

In regulären Ausdrücken werden Buchstaben direkt durch die entsprechenden Buchstaben des Alphabets beschrieben, während Zahlen durch das Sonderzeichen \d dargestellt werden können, das einer beliebigen Ziffer von 0 bis 9 entspricht. Eckige Klammern werden verwendet, um einen Bereich von Buchstaben oder Ziffern anzugeben, z. B. [a-z] für jeden Kleinbuchstaben des lateinischen Alphabets oder [0-9], was \d entspricht. Für nicht numerische und nicht buchstabenbezogene Zeichen werden \D bzw. \W verwendet.

Beliebte RegEx-Anwendungsfälle (Abb. 4.4-7):

- **Überprüfung der E-Mail-Adresse:** Um zu überprüfen, ob eine Zeichenfolge eine gültige E-Mail-Adresse ist, können Sie die Vorlage "^ **[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$**" verwenden**.**
- **Datumsextraktion**: "\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}**.\d{4}\b**" kann verwendet werden, um das Datum aus einem Text im Format TT.MM.JJJJJ zu extrahieren.
- Überprüfung von Telefonnummern: Um Telefonnummern im Format +49(000)000-0000 zu überprüfen, sieht das Muster wie **"\+\d{2}\(\d{3}\)\d{3}-\d{4}"** aus**.**

Durch die Übersetzung der Anforderungen eines QS-Ingenieurs in das Format der Attribute und ihrer Grenzwerte (Abb. 4.4-6) haben wir sie von ihrem ursprünglichen Textformat (Gespräche, Briefe und behördliche Dokumente) in eine organisierte und strukturierte Tabelle umgewandelt, die es ermöglicht, alle eingehenden Daten (z. B. neue Elemente der Kategorie Fenster) automatisch zu prüfen und zu analysieren. Das Vorhandensein von Anforderungen ermöglicht es, nicht geprüfte Daten automatisch zu verwerfen und die geprüften Daten automatisch an die Systeme zur weiteren Verarbeitung zu übermitteln.

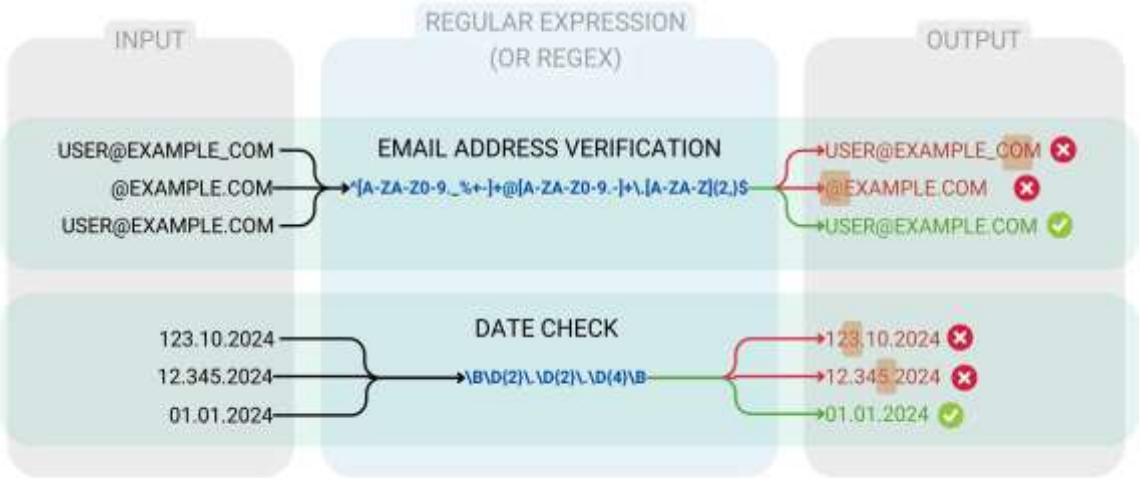


Abbildung 4.4-7 Die Verwendung von regulären Ausdrücken ist ein äußerst wichtiges Werkzeug bei der Validierung von Textdaten.

Wenn wir nun von der konzeptionellen zur logischen Ebene der Arbeit mit Anforderungen übergehen, werden wir alle Anforderungen aller Spezialisten in unserem Prozess der Installation eines neuen Fensters (Abb. 4.4-4) in eine geordnete Liste im Attributformat umwandeln und diese Listen mit den notwendigen Attributen zu unserem Flussdiagramm für jeden Spezialisten hinzufügen (Abb. 4.4-8).

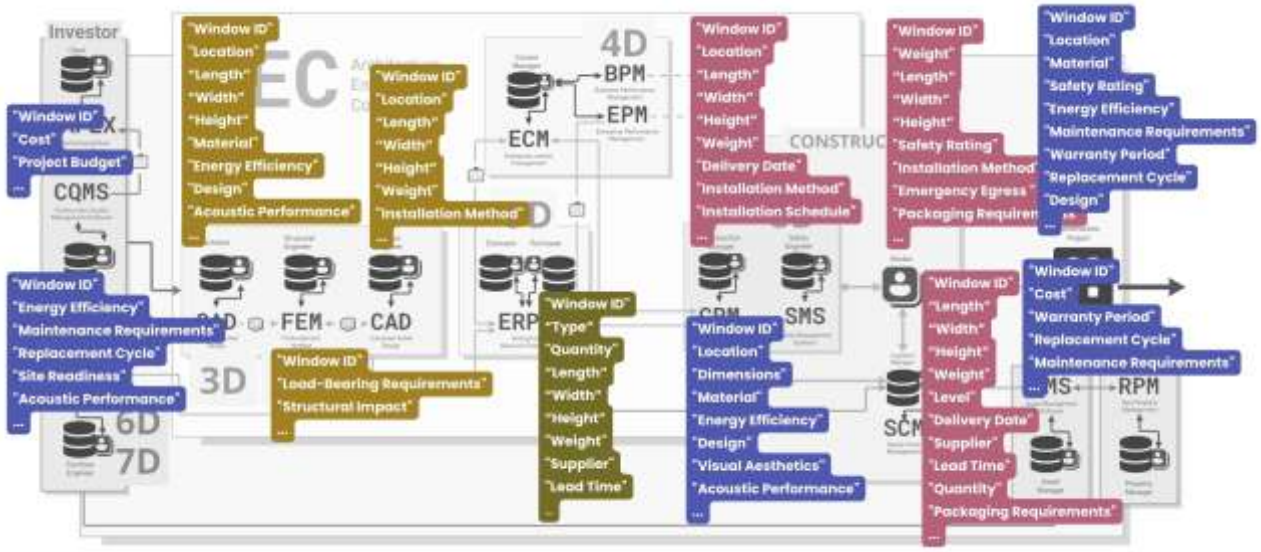


Abbildung 4.4-8 Auf der Ebene des logischen Prozesses werden die Attribute, die von jedem Spezialisten bearbeitet werden, den jeweiligen Systemen hinzugefügt.

Indem wir alle Attribute zu einer gemeinsamen Prozesstabelle hinzufügen, wandeln wir die Informationen, die zuvor als Text und Dialog auf der konzeptionellen Ebene (Abb. 4.4-1) dargestellt wurden, in die strukturierte und systematische Form von Tabellen auf der physischen Ebene um (Abb. 4.4-9).

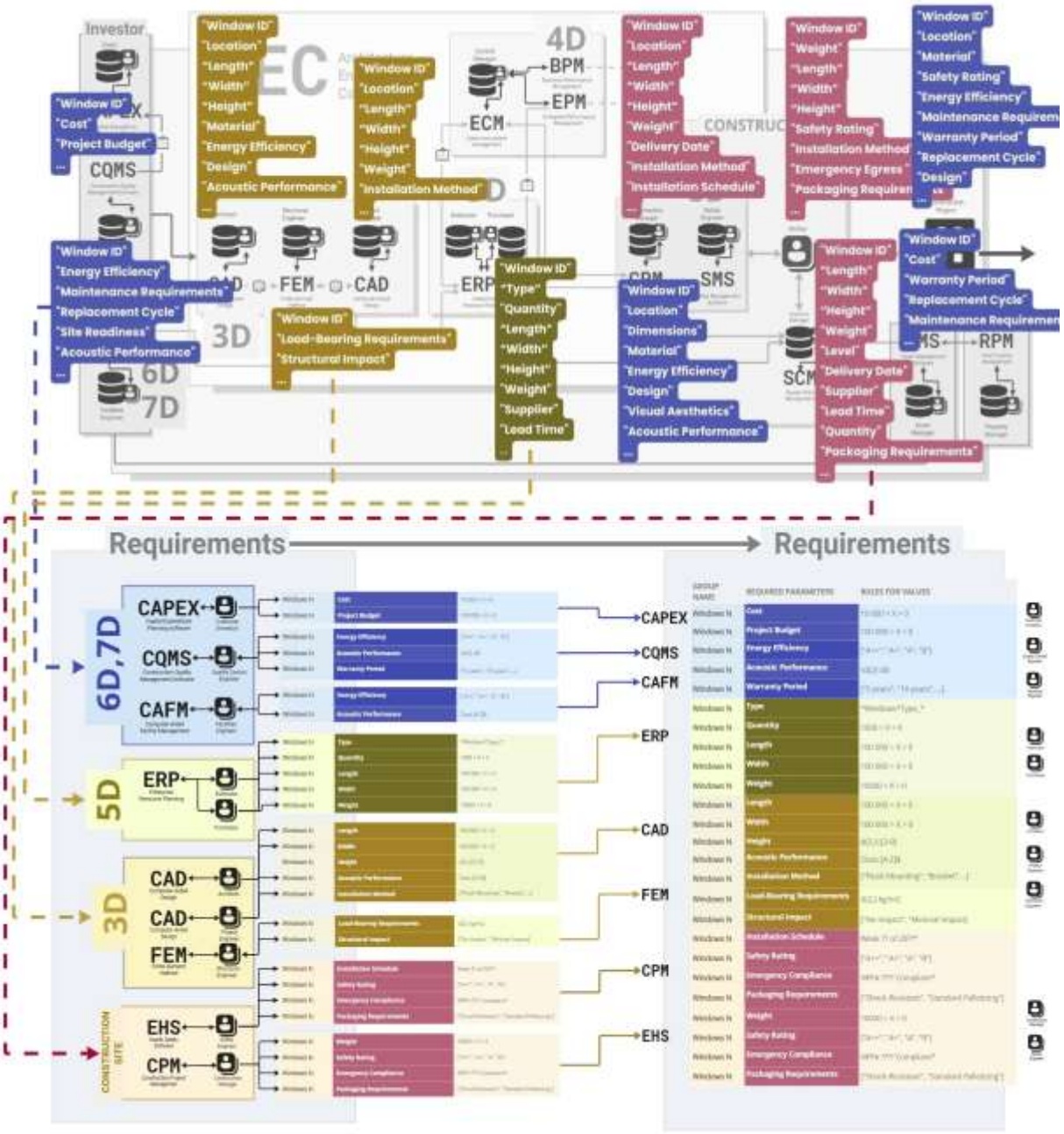


Abbildung 4.4-9 Die Umwandlung des unstrukturierten Fachdialogs in strukturierte Tabellen hilft, die Anforderungen von auf der physischen Ebene zu verstehen.

Nun müssen die Datenanforderungen an die Spezialisten weitergegeben werden, die Informationen für bestimmte Systeme erstellen. Wenn Sie zum Beispiel in einer CAD-Datenbank arbeiten, sollten Sie, bevor Sie mit der Modellierung von Elementen beginnen, alle notwendigen Parameter auf der Grundlage der Endverwendungsszenarien der Daten sammeln. Dies beginnt in der Regel mit der Betriebsphase, gefolgt von der Baustelle, der Logistikabteilung, der Kalkulationsabteilung, der Abteilung für statische Berechnungen und so weiter. Erst wenn Sie die Anforderungen all dieser Bereiche berücksichtigt haben, können Sie mit der Datenerstellung beginnen - auf der Grundlage der gesammelten Parameter. Auf diese Weise können Sie in Zukunft die Überprüfung und Übertragung von Daten entlang der Kette automatisieren.

Wenn neue Daten die Anforderungen erfüllen, werden sie automatisch in das Datenökosystem des Unternehmens integriert und gelangen direkt zu den Benutzern und Systemen, für die sie bestimmt sind. Die Überprüfung der Daten anhand von Attributen und deren Werten stellt sicher, dass die Informationen die erforderlichen Qualitätsstandards erfüllen und für die Verwendung in Unternehmensszenarien bereit sind.

Die Datenanforderungen sind definiert, und bevor mit der Prüfung begonnen werden kann, müssen die zu prüfenden Daten erstellt, beschafft oder gesammelt werden, oder der aktuelle Stand der Informationen in Datenbanken muss erfasst werden, damit sie im Prüfungsverfahren verwendet werden können.

Datenerhebung für den Überprüfungsprozess

Bevor die Validierung beginnen kann, muss sichergestellt werden, dass die Daten in einer für den Validierungsprozess geeigneten Form vorliegen. Dies bedeutet nicht, dass die Informationen einfach nur verfügbar sind, sondern dass sie aufbereitet werden müssen: Die Daten müssen gesammelt und von unstrukturierten, lose strukturierten, textlichen und geometrischen Formaten in eine strukturierte Form umgewandelt werden. Dieser Prozess wird in den vorangegangenen Kapiteln ausführlich beschrieben, in denen Methoden zur Umwandlung verschiedener Datentypen erörtert wurden. Als Ergebnis aller Transformationen haben die eingehenden Daten die Form offener strukturierter Tabellen (Abb. 4.1-2, Abb. 4.1-9, Abb. 4.1-13).

Mit den Anforderungen und strukturierten Tabellen mit den erforderlichen Parametern und Grenzwerten (Abb. 4.4-9) können wir mit der Validierung der Daten beginnen - entweder als einzelner automatisierter Prozess (Pipeline) oder als schrittweise Validierung jedes eingehenden Dokuments.

Um die Prüfung zu starten, ist es erforderlich, entweder eine neue Datei als Eingabe zu erhalten oder den aktuellen Zustand der Daten zu fixieren - einen Schnappschuss zu erstellen oder aktuelle und eingehende Daten zu exportieren oder eine Verbindung zu einer externen oder internen Datenbank herzustellen. Im vorliegenden Beispiel wird ein solcher Schnappschuss erstellt, indem CAD-Daten von automatisch in ein strukturiertes Format konvertiert werden, das z. B. am Freitag, dem 29. März 2024, um 23:00 Uhr aufgezeichnet wird, nachdem alle Konstrukteure nach Hause gegangen sind.

Abbildung 4.4-10 CAD-Datenbank-Snapshot (BIM), der die aktuellen Attributinformationen für ein neues Objekt der Klasse "Fenster" in der aktuellen Version des Projektmodells zeigt.

Dank der Reverse-Engineering-Werkzeuge, die im Kapitel "Übersetzen von CAD-Daten (BIM) in eine strukturierte Form" besprochen werden, können diese Informationen aus verschiedenen CAD-(BIM-)Werkzeugen

und Editoren in separaten Tabellen organisiert (Abb. 4.4-11) oder in einer gemeinsamen Tabelle zusammengefasst werden, die verschiedene Abschnitte des Projekts kombiniert (Abb. 9.1-10).

Eine solche Tabellen-Datenbank enthält eindeutige Bezeichner von Fenstern und Türen (ID-Attribut), Typenbezeichnungen (TypeName), Abmessungen (Breite, Länge), Materialien (Material) sowie Energie- und Schallleistungsindikatoren und andere Merkmale. Eine solche Tabelle, die im CAD-Programm (BIM) ausgefüllt wird, wird von einem Planungsingenieur aus verschiedenen Abteilungen und Dokumenten zusammengestellt und bildet ein Informationsmodell des Projekts.

Abbildung 4.4-11 Strukturierte Daten aus CAD Systemen können eine zweidimensionale Tabelle sein, deren Spalten die Attribute der Elemente bezeichnen.

Echte CAD-Projekte (BIM) umfassen Zehn- oder Hunderttausende von Elementen (Abb. 9.1-10). Elemente in CAD-Formaten werden automatisch nach Typ und Kategorie kategorisiert - von Fenstern und Türen bis hin zu Decken, Platten und Wänden. Eindeutige Bezeichner (z. B. native ID, die automatisch von der CAD-Lösung vergeben wird) oder Typattribute (Typname, Typ, Familie) ermöglichen die Verfolgung desselben Objekts in verschiedenen Systemen. So kann beispielsweise ein neues Fenster an der Nordwand eines Gebäudes in allen relevanten Systemen des Unternehmens durch einen einzigen Bezeichner "W-NEW" eindeutig identifiziert werden.

Während die Namen und Bezeichner von Entitäten in allen Systemen einheitlich sein sollten, können sich die mit diesen Entitäten verbundenen Attribute und Werte je nach Nutzungskontext erheblich unterscheiden. Architekten, Bauingenieure, Bau-, Logistik- und Instandhaltungsfachleute nehmen die gleichen Elemente auf unterschiedliche Weise wahr. Jeder von ihnen stützt sich auf seine eigenen Klassifizierungsmerkmale, Normen und Ziele: Einige betrachten ein Fenster unter rein ästhetischen Gesichtspunkten, indem sie seine Form und Proportionen bewerten, während andere es unter technischen oder betrieblichen Gesichtspunkten betrachten, indem sie die Wärmeleitfähigkeit, die Einbaumethode, das Gewicht oder die Wartungsanforderungen analysieren. Daher ist es wichtig, bei der Modellierung von Daten und der Beschreibung von Elementen die Vielseitigkeit ihrer Verwendung zu berücksichtigen und die Konsistenz der Daten zu gewährleisten, wobei die Besonderheiten der Branche zu berücksichtigen sind.

Für jede Rolle in den Unternehmensprozessen gibt es spezialisierte Datenbanken mit einer eigenen Benutzeroberfläche - von der Planung und Berechnung bis hin zu Logistik, Installation und Gebäudebetrieb (Abb. 4.4-12). Jedes dieser Systeme wird von einem professionellen Team von Fachleuten über eine spezielle Benutzeroberfläche oder über Datenbankabfragen verwaltet, wobei hinter der Summe aller Entscheidungen, die über die eingegebenen Werte am Ende der Kette getroffen werden, der Systemmanager oder Abteilungsleiter steht, der gegenüber seinen Partnern, die andere Systeme bedienen, für die rechtliche Gültigkeit und Qualität der eingegebenen Daten verantwortlich ist.

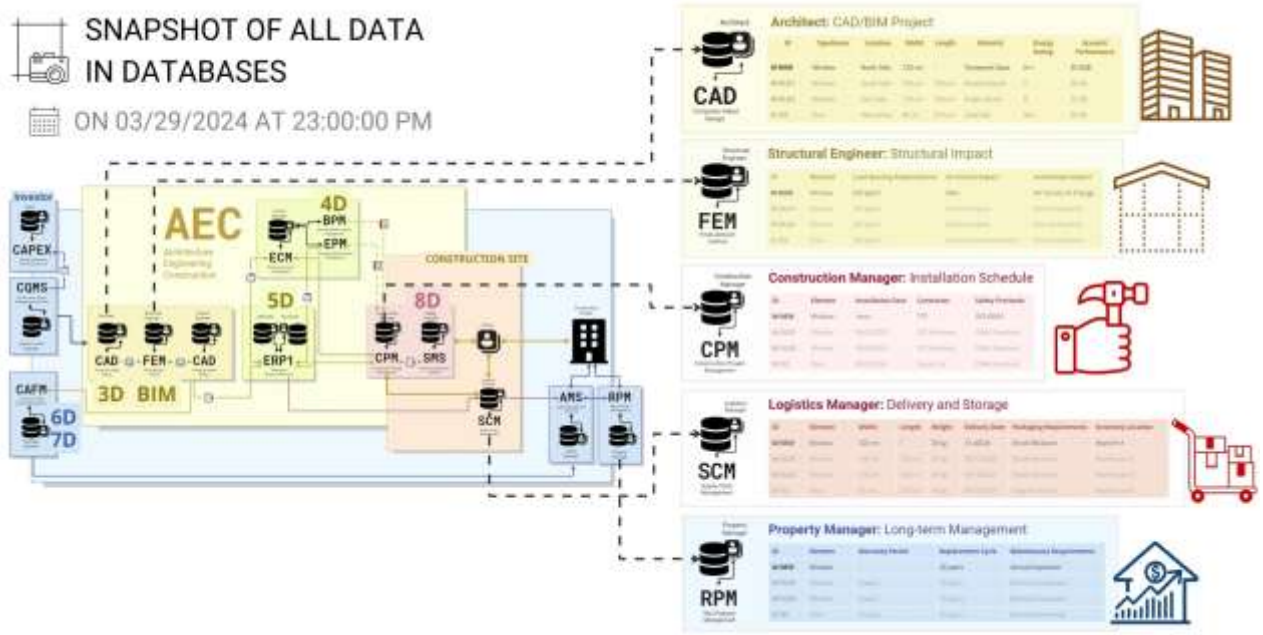


Abbildung 4.4-12 Ein und dieselbe Entität hat in verschiedenen Systemen denselben Bezeichner, aber unterschiedliche Attribute, die nur in diesem System wichtig sind.

Nachdem wir die Sammlung von strukturierten Anforderungen und Daten auf logischer und physischer Ebene organisiert haben, müssen wir einen Prozess einrichten, um die Daten aus verschiedenen eingehenden Dokumenten und unterschiedlichen Systemen automatisch mit den zuvor gesammelten Anforderungen abzugleichen.

Überprüfung der Daten von und Ergebnisse der Überprüfung

Alle neu in das System eingehenden Daten - seien es Dokumente, Tabellen oder Datenbankeinträge von Bauherrn, Architekt, Ingenieur, Polier, Logistiker oder Hausverwalter - müssen anhand der zuvor formulierten Anforderungen validiert werden (Abb. 4.4-9). Der Validierungsprozess ist entscheidend: Fehler in den Daten können zu falschen Berechnungen, Terminverzögerungen und sogar zu finanziellen Verlusten führen. Um solche Risiken zu minimieren, sollte ein systematisches und wiederholbares, iteratives Datenvalidierungsverfahren organisiert werden.

> Um neue Daten, die in das System eingehen - unstrukturiert, textuell oder geometrisch - zu validieren, müssen sie in ein locker strukturiertes oder strukturiertes Format umgewandelt werden. Der Validierungsprozess muss dann die Daten anhand einer vollständigen Liste der erforderlichen Attribute und ihrer zulässigen Werte überprüfen.

Die Konvertierung verschiedener Datentypen: Text, Bilder, PDF -Dokumente und gemischte CAD (BIM) Daten in eine strukturierte Form wurde im Kapitel "Konvertierung von Daten in eine strukturierte Form" ausführlich besprochen.

Ein Beispiel ist eine Tabelle aus einem CAD (BIM) Projekt (Abb. 4.4-11). Sie enthält halbstrukturierte Geometriedaten und strukturierte Attributinformationen zu Projektelementen (Abb. 3.1-14) - z. B. ein Element der Klasse "Fenster".

Zur Validierung vergleichen wir die Attributwerte (Abb. 4.4-11) mit den Referenzgrenzwerten, die von Experten in Form einer Anforderung definiert wurden (Abb. 4.4-9). Aus der abschließenden Vergleichstabelle (Abb. 4.4-13) wird ersichtlich, welche Werte akzeptabel sind und welche korrigiert werden müssen, bevor die Daten außerhalb von CAD-Anwendungen (BIM) verwendet werden können.

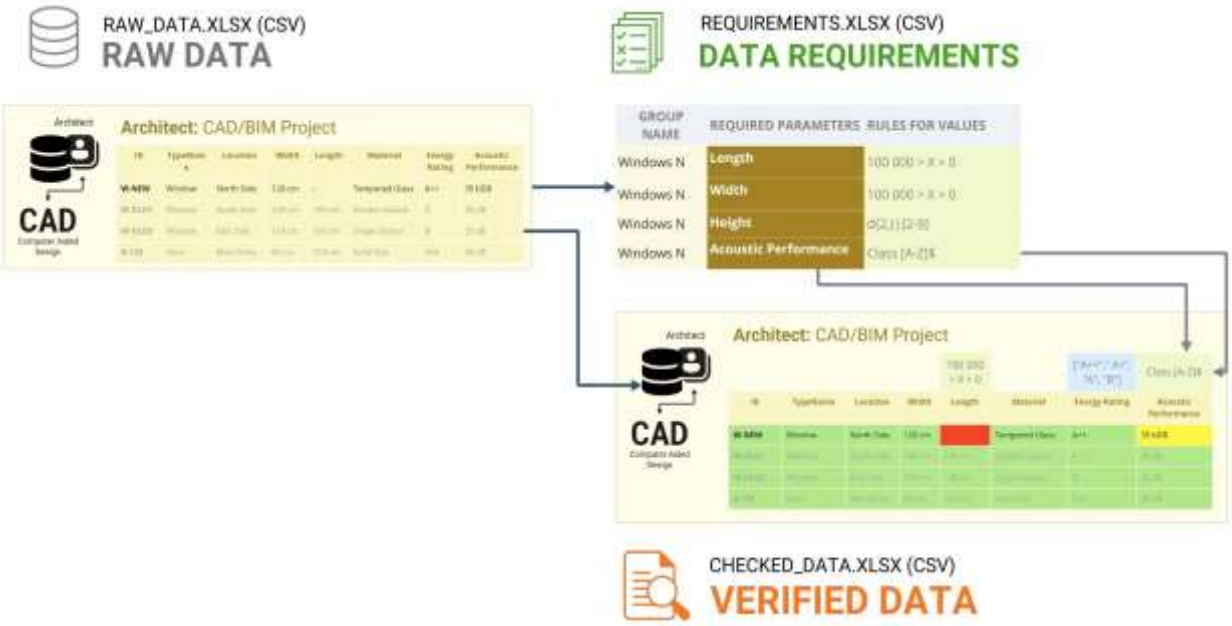


Abbildung 4.4-13 Die abschließende Validierungstabelle hebt die Attributwerte für die neue Entität der Klasse "Windows" hervor, auf die Sie achten sollten.

Bei der Implementierung einer ähnlichen Lösung unter Verwendung der Pandas-Bibliothek, die wir bereits im Kapitel "Pandas: Ein unverzichtbares Werkzeug für die Datenanalyse" beschrieben haben, werden wir Daten aus einer tabellarischen Datei, die aus einer CAD-Datei (BIM) (RVT, IFC, DWG, NWS, DGN) (Abb. 4.4-11) extrahiert wurde, mit den Anforderungen aus einer anderen tabellarischen Anforderungsdatei (Abb. 4.4-9) validieren.

Um den Code zu erhalten, müssen wir in der Eingabeaufforderung für LLM beschreiben, dass wir die Daten aus der Datei **raw_data.xlsx** (ein vollständiger Datensatz aus der CAD-Datenbank (BIM)) laden, sie **prüfen** und das Ergebnis in einer neuen Datei **checked_data.xlsx** speichern müssen (Abb. 4.4-13).

Lassen Sie uns den Code mit LLM erhalten, ohne die Pandas-Bibliothek zu erwähnen:

> Schreiben Sie einen Code, um die Tabelle aus der Datei raw_data.xlsx zu prüfen und anhand der folgenden Prüfregeln zu validieren: Die Werte der Spalten "Breite" und "Länge" sind größer als Null, "Energiekennzahl" ist in der Liste ["A++", "A+", "A", "B"] enthalten, und "Akustische Leistung" ist eine Variable, die wir später angeben werden - mit der letzten hinzugefügten Validierungsspalte, und speichern Sie die endgültige Tabelle in einer neuen Excel-Datei checked_data.xlsx ⏎

Die Antwort von LLM wird ein kurzes Beispiel für Python-Code beschreiben, der durch nachfolgende Aufforderungen verfeinert und erweitert werden kann:

Create in LLM Chat

Validation.py

```python
import pandas as pd

# Load raw data from Excel file
df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)

# Adding a 'Requirements Met' column that checks all conditions
df_data['Requirements Met'] = (
    (df_data['Width'] > 0) &
    (df_data['Length'] > 0) &
    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
    (df_data['Acoustic Performance'] >= min_acoustic_performance)
)

# Save the results to a new Excel file
df_data.to_excel('checked_data.xlsx', index=False)
```

Run in IDE

Abbildung 4.4-14 Der vom LLM- Modell generierte Code überprüft den konvertierten CAD (BIM) Entwurf gegen die Attributanforderungen in Form von Randwerten.

Der vom LLM-Sprachmodell generierte Code kann in jeder gängigen IDE oder Online-Tool verwendet werden: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse mit PyDev-Plugin, Thonny, Wing IDE, IntelliJ IDEA mit Python-Plugin, JupyterLab oder beliebte Online-Tools Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Die Ausführung des Codes (Abb. 4.4-14) zeigt, dass die "Entity-Elemente" W-OLD1, W-OLD2, D-122 (und andere Elemente) aus der CAD-Datenbank (BIM) die Attributanforderungen erfüllen: Breite und Länge sind größer als Null, und die Energieeffizienzklasse ist einer der Listenwerte "A++", "A", "B", "C" (Abb. 4.4-15).

Das von uns benötigte und kürzlich hinzugefügte Element W-NEW, das für die neue Elementklasse 'Window' auf der Nordseite verantwortlich ist, entspricht nicht den Anforderungen (Attribut 'Requrments Met'), da seine Länge Null ist (ein Wert von '0,0' wird von unserer 'Width'>0-Regel als inakzeptabel angesehen) und es keine Energieeffizienzklasse angibt.

CHECKED_DATA.XLSX (CSV)

VERIFIED DATA

	ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

Abbildung 4.4-15 Überprüfung identifiziert Entitäten, die den Verifizierungsprozess nicht durchlaufen haben, und fügt den Ergebnissen ein neues Attribut mit den Werten "Falsch" oder "Wahr" hinzu.

In ähnlicher Weise überprüfen wir die Konsistenz aller Projektelemente (Entitäten) und der erforderlichen Attribute für jedes der Systeme, Tabellen oder Datenbanken in allen Daten, die wir von verschiedenen Spezialisten erhalten (Abb. 4.4-1), während des Prozesses des Hinzufügens eines Fensters zum Projekt.

In der abschließenden Tabelle ist es zweckmäßig, die Ergebnisse der Prüfung zur Visualisierung farblich hervorzuheben: Attribute, die die Prüfung erfolgreich bestanden haben, sind grün markiert, gelb - Werte mit unkritischen Abweichungen, und rot - kritische Abweichungen (Abb. 4.4-16).

Als Ergebnis der Validierung (Abb. 4.4-16) erhalten wir eine Liste vertrauenswürdiger und validierter Elemente mit ihren Bezeichnern, bei denen überprüft wurde, ob sie die Attributanforderungen erfüllen. Validierte Elemente bieten die Gewissheit, dass diese Elemente die angegebenen Standards und Spezifikationen für alle Systeme erfüllen, die am Prozess des Hinzufügens von Elementen der Klasse Window oder einer anderen Klasse beteiligt sind (wir werden mehr über die Automatisierung der Datenvalidierung und die Erstellung eines automatisierten ETL-Prozesses im Kapitel "Automatisierung von ETL und Datenvalidierung" besprechen).

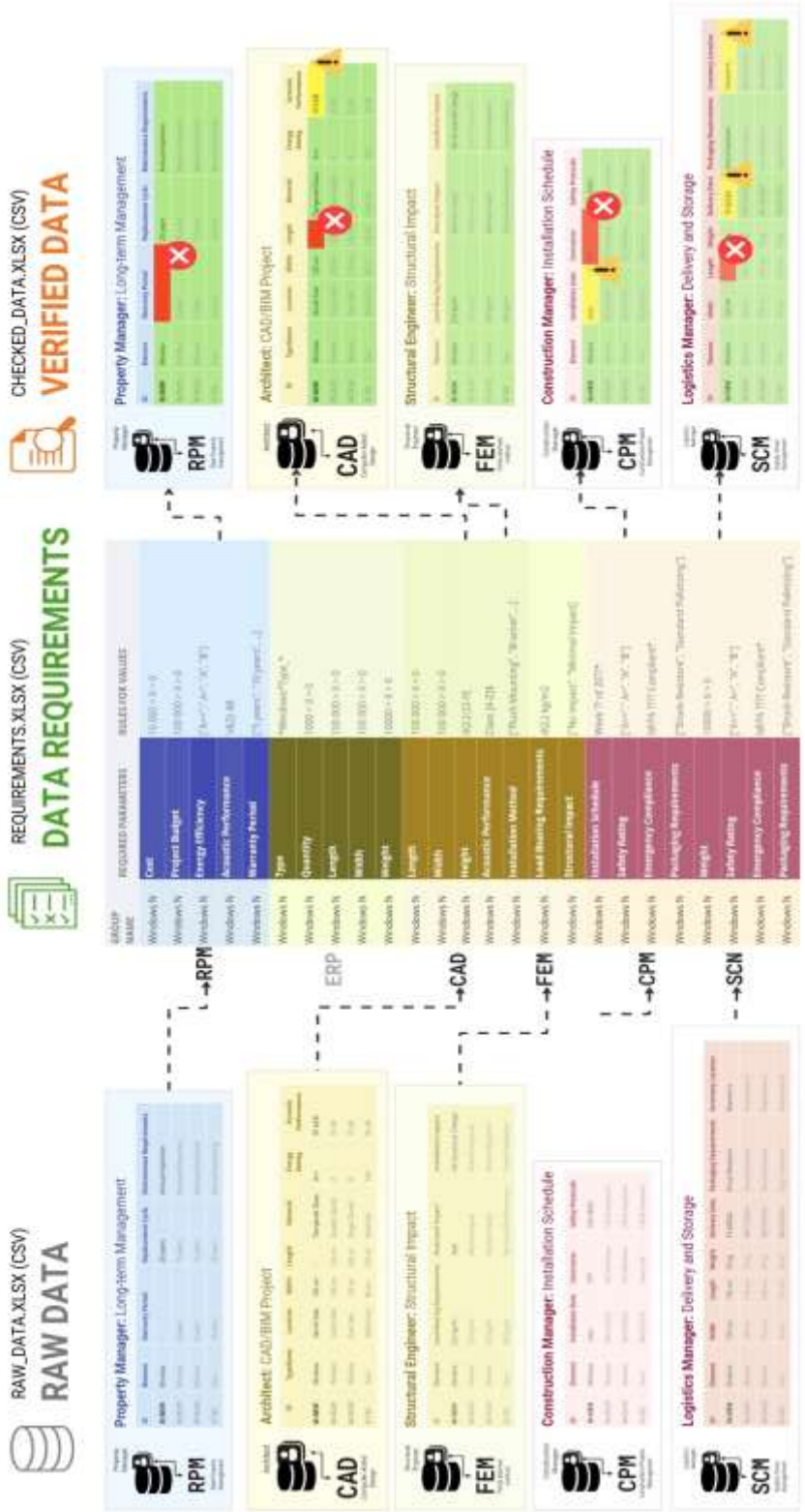


Abb. 4.4-16 Anhand des Ergebnisses der für alle Systeme durchgeführten Prüfung lässt sich feststellen, welche Daten nicht den Anforderungen des Unternehmens entsprechen.

Entitäten, die erfolgreich validiert wurden, erfordern in der Regel nicht viel Aufmerksamkeit. Sie können ohne Hindernisse in die nächsten Phasen der Verarbeitung und Integration in andere Systeme übergehen. Im Gegensatz zu den "Qualitäts"-Elementen sind die Elemente, die die Validierung nicht bestehen, von größtem Interesse. Informationen über solche Abweichungen sind von entscheidender Bedeutung: Sie sollten nicht nur in tabellarischen Berichten, sondern auch mit Hilfe verschiedener Visualisierungswerkzeuge vermittelt werden. Eine grafische Darstellung der Überprüfungsergebnisse hilft, den Gesamtzustand der Datenqualität schnell zu beurteilen, Problembereiche zu erkennen und umgehend Korrektur- oder Abhilfemaßnahmen zu ergreifen.

Visualisierung der Ergebnisse der Überprüfung

Die Visualisierung ist ein wichtiges Instrument zur Interpretation der Inspektionsergebnisse. Zusätzlich zu den üblichen zusammenfassenden Tabellen kann es Informationstafeln, Diagramme und automatisch generierte PDF-Dokumente enthalten, die Projektelemente nach ihrem Prüfstatus gruppieren. Eine Farbcodierung kann hier eine unterstützende Rolle spielen: Grün kann auf Elemente hinweisen, die erfolgreich validiert wurden, gelb auf Elemente, die zusätzliche Aufmerksamkeit erfordern, und rot auf Elemente, die kritische Fehler aufweisen oder bei denen wichtige Daten fehlen.

In unserem Beispiel (Abb. 4.4-1) analysieren wir Daten aus jedem System Schritt für Schritt: vom CAD (BIM) über die Objektverwaltung bis hin zu Logistik und Montageplänen (Abb. 4.4-16). Im Anschluss an die Prüfung werden für jeden Fachmann automatisch individuelle Warnhinweise oder Berichtsdokumente generiert, z.B. im PDF-Format (Abb. 4.4-17). Sind die Daten korrekt, erhält der Spezialist eine kurze Nachricht: "Danke für die Zusammenarbeit". Werden Unstimmigkeiten festgestellt, wird ein detaillierter Bericht mit dem Wortlaut "Dieses Dokument listet Elemente, ihre Bezeichner, Attribute und Werte auf, die nicht auf ihre Übereinstimmung geprüft wurden" versandt.

Abbildung 4.4-17 Validierung und automatische Berichterstellung beschleunigen den Prozess des Auffindens und Verstehens von Datenmängeln für den Fachmann, der die Daten erstellt.

Dank des automatisierten Validierungsprozesses wird - sobald ein Fehler oder eine Datenlücke entdeckt wird - eine sofortige Benachrichtigung in Form einer Chat-Nachricht, einer E-Mail oder eines PDF-Dokuments an die Person gesendet, die für die Erstellung oder Bearbeitung der betreffenden Entitäten und ihrer Attribute verantwortlich ist (Abb. 4.4-18), mit einer Liste der Elemente und Attributbeschreibungen, die nicht validiert wurden.

Abb. 4.4-18 Automatische Inspektionsberichte erleichtern das Verständnis von Fehlern und beschleunigen die Arbeit bei der Vervollständigung der Projektdaten.

Wenn beispielsweise ein Dokument im Immobilienverwaltungssystem (nach der Strukturierung) eintrifft, bei dem das Attribut "Gewährleistungsfrist" falsch ausgefüllt ist, erhält der Immobilienverwalter eine Warnmeldung mit einer Liste von Attributen, die überprüft und korrigiert werden müssen.

Ebenso führen Mängel im Installationsplan oder in den Logistikdaten dazu, dass ein automatischer Bericht erstellt wird und beispielsweise eine Chat-Benachrichtigung oder eine E-Mail mit den Ergebnissen der Inspektion an den zuständigen Fachmann geschickt wird.

Zusätzlich zu den PDF-Dokumenten und den Diagrammen mit den Ergebnissen können Dashboards und interaktive 3D-Modelle (Abb. 7.1-6, Abb. 7.2-12) erstellt werden, die Elemente mit fehlenden Attributen hervorheben, so dass die Benutzer die 3D-Geometrien der Elemente visuell nutzen können, um die Qualität und Vollständigkeit dieser Elemente im Projekt zu filtern und zu bewerten.

> Die Visualisierung von Inspektionsergebnissen in Form von automatisch generierten Dokumenten, Grafiken oder Dashboards vereinfacht die Dateninterpretation erheblich und erleichtert die effektive Kommunikation zwischen den Projektbeteiligten.

Der Prozess der automatischen Überprüfung von Daten aus verschiedenen Systemen und Informationsquellen kann mit einer fundierten Entscheidungsfindung im Alltag verglichen werden. So wie Unternehmen in der Bauindustrie viele Variablen berücksichtigen - von der Zuverlässigkeit der Eingangsdaten bis hin zu ihren Auswirkungen auf den Zeitplan, die Kosten und die Qualität der Projektrealisierung, so wägt auch der Mensch bei wichtigen Entscheidungen, z. B. bei der Wahl seines Wohnorts, eine ganze Reihe von Faktoren ab: Verkehrsanbindung, Infrastruktur, Kosten, Sicherheit, Lebensqualität. All diese Überlegungen bilden ein System von Kriterien, die die Grundlage für die endgültigen Entscheidungen bilden, die unser Leben ausmachen.

Vergleich der Datenqualitätsprüfungen mit den Bedürfnissen des menschlichen Lebens

Trotz der ständigen Weiterentwicklung von Methoden und Instrumenten zur Kontrolle der Datenqualität bleibt das Grundprinzip der Einhaltung der Informationsvorschriften unverändert. Dieses Prinzip ist in die Grundlage eines ausgereiften Managementsystems eingebaut, sei es in der Wirtschaft oder im Alltag.

> Der Prozess der iterativen Datenvalidierung ähnelt dem Entscheidungsfindungsprozess, mit dem jeder täglich konfrontiert ist. In beiden Fällen stützen wir uns auf Erfahrungen, Daten und neue Informationen, sobald diese verfügbar sind. Und immer mehr Entscheidungen im Leben und im Beruf - von strategischen bis hin zu alltäglichen Entscheidungen - werden auf der Grundlage von Daten getroffen.

Bei der Wahl eines Wohnorts oder eines Lebenspartners zum Beispiel bilden wir intuitiv eine Tabelle mit Kriterien und Merkmalen, anhand derer wir Alternativen vergleichen (Abb. 4.4-19). Diese Merkmale - seien es persönliche Eigenschaften einer Person oder Parameter einer Immobilie - stellen Attribute dar, die die endgültige Entscheidung beeinflussen.

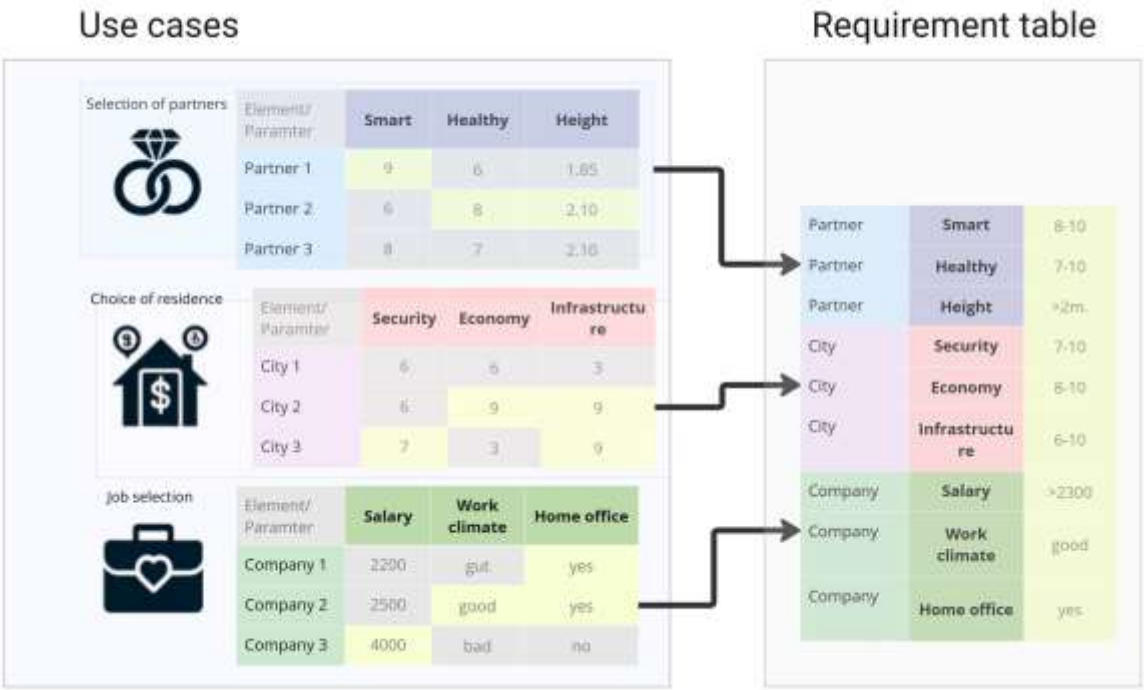


Abb. 4.4-19 Die Wahl des Wohnsitzes, der Beschäftigung oder der Partnerschaft hängt von den individuellen Anforderungen an die Attribute ab.

Die Verwendung strukturierter Daten und eines formalisierten Ansatzes zur Beschreibung von Anforderungen (Abb. 4.4-20) trägt zu fundierteren und sachkundigeren Entscheidungen sowohl im beruflichen als auch im privaten Bereich bei.

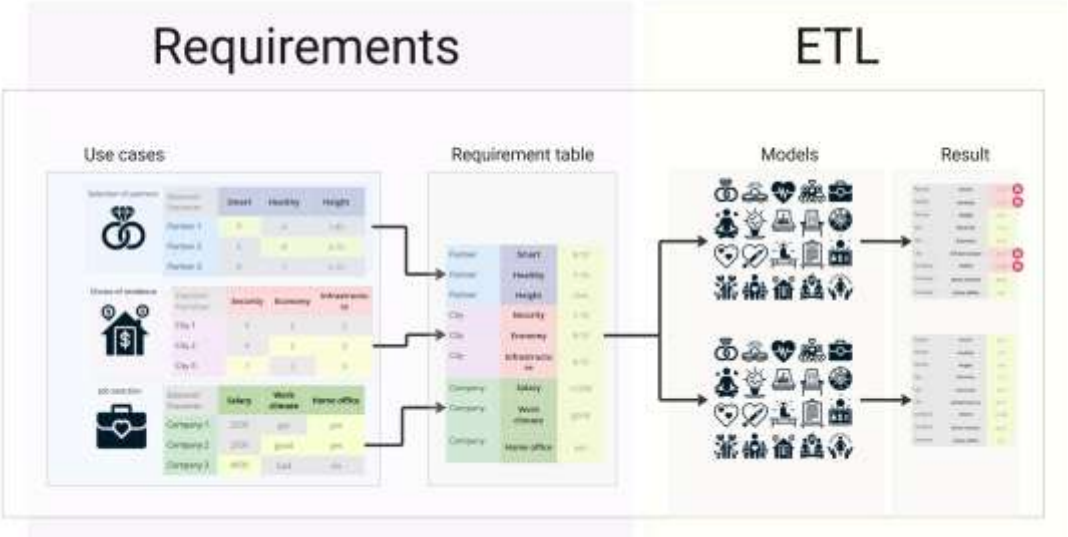


Abb. 4.4-20 Die Formalisierung von Anforderungen ermöglicht eine Systematisierung der Lebenswahrnehmung und der unternehmerischen Entscheidungen.

Der Ansatz der datengesteuerten Entscheidungsfindung ist nicht ausschließlich ein Geschäftsinstrument. Er fügt sich auch nahtlos in den Alltag ein und folgt gemeinsamen Datenverarbeitungsschritten (Abb. 4.4-21), ähnlich dem ETL-Prozess (Extract, Transform, Load), den wir bereits zu Beginn dieses Teils bei der Strukturierung von Daten besprochen haben und den wir im Kontext der Aufgabenautomatisierung im siebten Teil des Buches näher betrachten werden:

- **Daten als Grundlage (Auszug**): In jedem Bereich - sei es im Beruf oder im Privatleben - sammeln wir Informationen. Im Geschäftsleben können es Berichte, Zahlen, Marktdaten sein; im Privatleben können es persönliche Erfahrungen, Ratschläge von geliebten Menschen, Feedback, Beobachtungen sein.
- **Bewertungskriterien (Transform): Die** gesammelten Informationen werden auf der Grundlage vordefinierter Kriterien interpretiert. Im Berufsleben sind dies Leistungsindikatoren (KPIs), Budgetbeschränkungen und Normen; im Privatleben Parameter wie Preis, Bequemlichkeit, Zuverlässigkeit, Ausstrahlung usw.
- **Vorhersage und Risikoanalyse (Load):** Die letzte Phase umfasst die Entscheidungsfindung auf der Grundlage der Analyse der umgewandelten Daten und des Vergleichs möglicher Folgen. Dies ist vergleichbar mit Geschäftsprozessen, bei denen die Daten eine Geschäftslogik und einen Risikofilter durchlaufen.

> Die Entscheidungen, die wir treffen - von trivialen Vorlieben wie der Frage, was wir zum Frühstück essen, bis hin zu wichtigen Lebensereignissen wie der Wahl des Berufs oder des Lebenspartners - sind naturgemäß das Ergebnis der Verarbeitung und Bewertung von Daten.

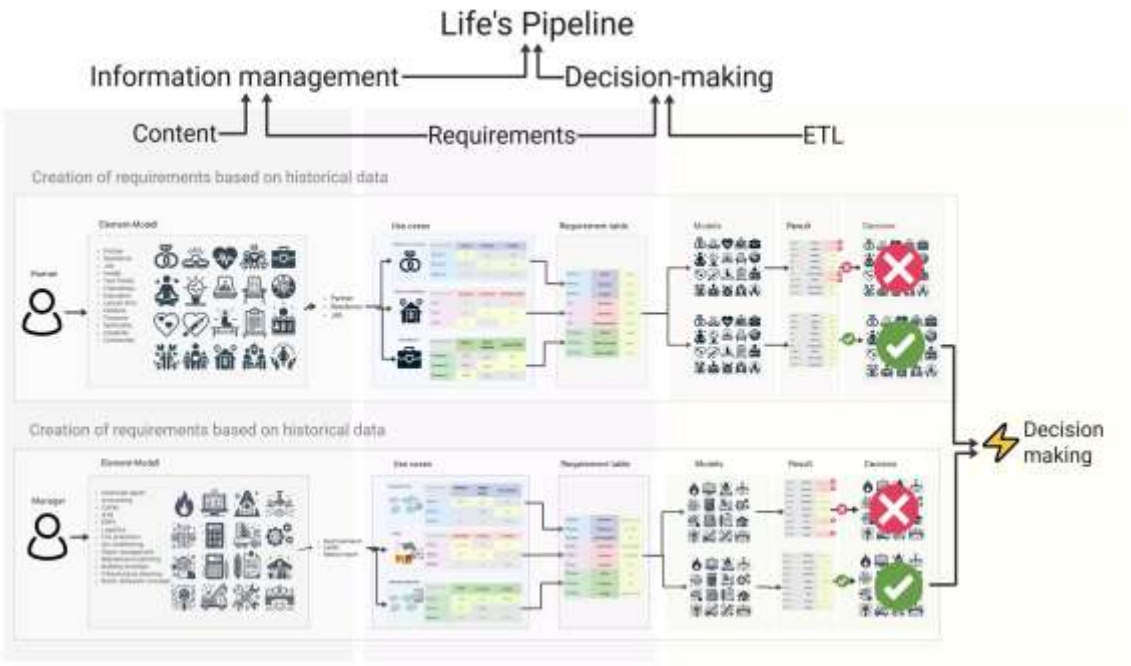


Abb. 4.4-21 Die Wirtschaft und das Leben im Allgemeinen sind eine Reihe von datengesteuerten Entscheidungen, bei denen die Qualität der für die Entscheidungsfindung verwendeten Daten ein Schlüsselfaktor ist.

Alles in unserem Leben ist miteinander verbunden, und so wie lebende Organismen, einschließlich des Menschen, den Gesetzen der Natur folgen, sich weiterentwickeln und an veränderte Bedingungen anpassen, so spiegeln auch menschliche Prozesse, einschließlich der Art und Weise, wie wir Daten sammeln und analysieren, diese natürlichen Prinzipien wider. Die enge Beziehung zwischen Natur und menschlicher Tätigkeit bestätigt nicht nur unsere Abhängigkeit von der Natur, sondern auch unseren Wunsch, die in Millionen von Jahren der Evolution geschliffenen Gesetze bei der Schaffung von Datenarchitekturen, Prozessen und Systemen für die Entscheidungsfindung anzuwenden.

Neue Technologien, insbesondere im Bauwesen, sind ein Paradebeispiel dafür, wie sich die Menschheit immer wieder von der Natur inspirieren lässt, um bessere, nachhaltigere und effizientere Lösungen zu schaffen.

Nächste Schritte: Umwandlung von Daten in genaue Berechnungen und Pläne

In diesem Teil haben wir uns mit der Umwandlung unstrukturierter Daten in ein strukturiertes Format, der Entwicklung von Datenmodellen und der Organisation von Prozessen zur Überprüfung der Informationsqualität in Bauprojekten beschäftigt. Die Verwaltung, Standardisierung und Klassifizierung von Daten ist ein grundlegender Prozess, der einen systematischen Ansatz und ein klares Verständnis der Geschäftsanforderungen erfordert. Die in diesem Teil behandelten Techniken und Werkzeuge ermöglichen eine zuverlässige Integration zwischen verschiedenen Systemen während des gesamten Lebenszyklus eines Objekts.

Um diesen Teil zusammenzufassen, wollen wir die wichtigsten praktischen Schritte hervorheben, die Ihnen helfen werden, die besprochenen Ansätze bei Ihren täglichen Aufgaben anzuwenden:

- Beginnen Sie mit der Systematisierung der Anforderungen
  - ☐ Erstellen Sie ein Verzeichnis von Attributen und Parametern für Schlüsselelemente Ihrer Projekte und Prozesse
  - ☐ Dokumentieren Sie die Grenzwerte für jedes Attribut
  - ☐ Visualisierung von Prozessen und Beziehungen zwischen Klassen, Systemen und Attributen mithilfe von Flussdiagrammen (z. B. in Miro, Canva, Visio)
- Automatisieren Sie die Datenkonvertierung
  - ☐ Prüfen Sie, welche Ihrer Dokumente, die häufig in Prozessen verwendet werden, mit Hilfe von OCR-Bibliotheken digitalisiert und in Tabellenform umgewandelt werden können
  - ☐ Informieren Sie sich über Reverse-Engineering-Tools zur Extraktion von Daten aus CAD (BIM)
  - ☐ Versuchen Sie, einen automatischen Datenabruf von Dokumenten oder Formaten, die Sie bei Ihrer Arbeit häufig verwenden, in einem Tabellenkalkulationsformular einzurichten
  - ☐ Automatische Konvertierungen zwischen verschiedenen Datenformaten einrichten
- Schaffung einer Wissensbasis für die Kategorisierung
  - ☐ Entwicklung eines internen oder Verwendung eines bestehenden Elementklassifizierers, der mit Industriestandards übereinstimmt
  - ☐ Dokumentieren Sie die Zusammenhänge zwischen den verschiedenen Klassifizierungssystemen
  - ☐ Diskutieren Sie mit Ihrem Team über die Verwendung eines einheitlichen Systems zur Identifizierung und eindeutigen Klassifizierung von Elementen
  - ☐ Beginnen Sie mit dem Aufbau eines Prozesses zur automatischen Validierung von Daten - sowohl der Daten, mit denen Sie innerhalb des Teams arbeiten, als auch der Daten, die an externe Systeme weitergegeben werden.

Mit diesen Ansätzen können Sie die Qualität Ihrer Daten erheblich verbessern und deren Weiterverarbeitung und Transformation vereinfachen. In den folgenden Teilen des Buches werden wir uns damit befassen, wie Sie bereits strukturierte und aufbereitete Daten für automatisierte Kalkulationen, Kostenschätzungen, Terminplanung und Bauprojektmanagement nutzen können.

V TEIL

KOSTEN- UND ZEITBERECHNUNGEN: EINBEZIEHUNG VON DATEN IN DEN BAUPROZESS

Der fünfte Teil ist den praktischen Aspekten der Nutzung von Daten zur Optimierung der Kalkulation und Planung von Bauprojekten gewidmet. Die ressourcenbasierte Methode der Kostenkalkulation und die Automatisierung von Kalkulationsprozessen werden eingehend analysiert. Methoden zur automatisierten Erfassung von Mengengerüsten) aus CAD (BIM)-Modellen und deren Integration in Kalkulationssysteme werden betrachtet. 4D und 5D Modellierungstechnologien für die Terminplanung und das Baukostenmanagement werden mit konkreten Anwendungsbeispielen untersucht. Eine Analyse der erweiterten Informationsschichten 6D -8D, die einen integrierten Ansatz zur Bewertung der Nachhaltigkeit, des Betriebs und der Sicherheit von Immobilienobjekten bieten, wird vorgestellt. Die Methoden zur Berechnung von Carbon Footprint und ESG -Indikatoren von Bauprojekten im Kontext moderner Umweltanforderungen und -standards werden ausführlich diskutiert. Die Möglichkeiten und Grenzen traditioneller ERP- und PMIS- -Systeme im Bauprozessmanagement werden kritisch bewertet, mit einer Analyse ihrer Auswirkungen auf die Preistransparenz. Es werden Aussichten für den Übergang von geschlossenen Lösungen zu offenen Standards und flexiblen Datenanalysewerkzeugen prognostiziert, die eine größere Effizienz der Bauprozesse gewährleisten können.

KAPITEL 5.1.
KOSTENBERECHNUNGEN UND -VORANSCHLÄGE FÜR BAUPROJEKTE

Grundlagen der Konstruktion: Schätzung von Menge, Kosten und Zeit

Unter den vielen Geschäftsprozessen, die die Nachhaltigkeit eines Unternehmens in der Bauindustrie bestimmen, sind - wie schon vor Tausenden von Jahren - die Prozesse zur genauen Schätzung der Anzahl der Elemente, der Projektkosten und der Lieferzeit von besonderer Bedeutung (Abb. 5.1-1).

Die Entwicklung der Schrift war das Ergebnis eines Komplexes von Faktoren, darunter die Notwendigkeit, wirtschaftliche Transaktionen, Handel und Ressourcenmanagement in frühen Gesellschaften aufzuzeichnen. Die ersten rechtlich bedeutsamen Dokumente, Tontafeln mit Berechnungen von Materialkosten und Arbeitslöhnen, wurden im Zusammenhang mit Handel und Bauwesen verwendet. Diese Tafeln hielten die Verpflichtungen der Parteien bei der Errichtung von Bauwerken fest und wurden als Beweis für Vereinbarungen, Geld- und Handelsbeziehungen aufbewahrt.

> Seit Jahrtausenden ist die Vorgehensweise bei der Kalkulation weitgehend unverändert geblieben: Berechnungen wurden manuell durchgeführt, wobei man sich auf die Erfahrung und Intuition des Kalkulationsingenieurs verließ. Mit dem Aufkommen modularer ERP-Systeme und CAD-Tools begann sich der traditionelle Ansatz der Mengen-, Kosten- und Zeitschätzung jedoch rasch zu verändern. Mit den heutigen digitalen Technologien können die wichtigsten Zeit- und Kostenberechnungen vollständig automatisiert werden, was eine höhere Genauigkeit, Schnelligkeit und Transparenz bei der Ressourcenplanung von Bauprojekten ermöglicht.

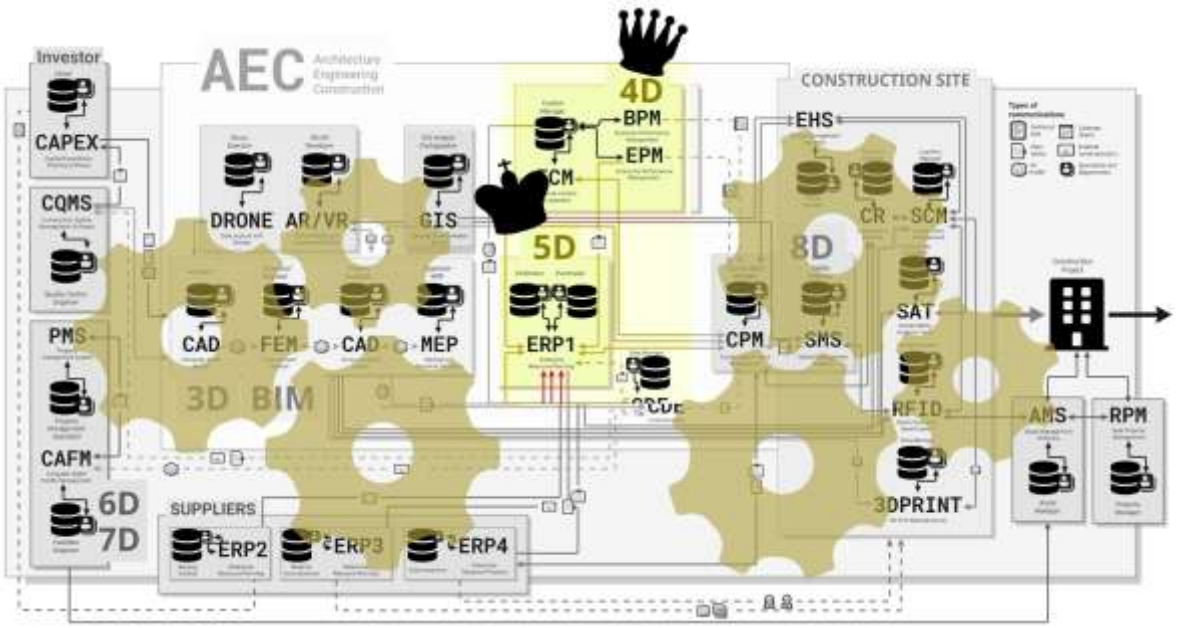


Abb. 5.1-1 Von den vielen verschiedenen Systemen sind die Werkzeuge, die für Volumen-, Kosten- und Zeitmetriken zuständig sind, die wichtigsten im Unternehmen.

Der Schwerpunkt der Bauunternehmen liegt auf genauen Zeit- und Kostendaten. Diese wiederum hängen vom Material- und Arbeitsaufwand ab, und ihre Transparenz wirkt sich auf die Rentabilität aus. Die Komplexität der Kalkulationsprozesse und ihre mangelnde Transparenz führen jedoch häufig zu höheren Projektkosten, Terminüberschreitungen und sogar zum Konkurs.

> Laut dem KPMG-Bericht "Familiar Problems - New Approaches" (2023) werden nur 50% der Bauprojekte fristgerecht abgeschlossen, und 87% der Unternehmen berichten über eine verstärkte Kontrolle der Wirtschaftlichkeit von Investitionsprojekten. Die Hauptprobleme stehen im Zusammenhang mit dem Mangel an qualifiziertem Personal und der Schwierigkeit der Risikovorhersage [2].

Historische Kalkulations- und Prozesszeitdaten werden während des Baus vergangener Projekte während der gesamten Lebensdauer des Bauunternehmens gesammelt und in Datenbanken verschiedener Systeme (ERP, PMIS BPM, EPM, etc.) eingegeben.

> Qualitativ hochwertige historische Kalkulationsdaten sind ein wichtiger Wettbewerbsvorteil für ein Bauunternehmen und wirken sich direkt auf seine Überlebensfähigkeit aus.

Schätz- und Kalkulationsabteilungen in Bau- und Ingenieurunternehmen werden eingerichtet, um historische Daten zu Projektkalkulationen zu sammeln, zu speichern und zu aktualisieren. Ihre Hauptaufgabe besteht darin, die Erfahrungen des Unternehmens zu sammeln und zu systematisieren, wodurch die Genauigkeit der Schätzung von Umfang, Zeitplan und Kosten neuer Projekte im Laufe der Zeit verbessert werden kann. Dieser Ansatz trägt dazu bei, Fehler bei künftigen Berechnungen auf der Grundlage der Praxis und der Ergebnisse bereits durchgeführter Projekte zu minimieren.

Methoden zur Berechnung der geschätzten Kosten der Projekte

Kalkulationsspezialisten verwenden eine Vielzahl von Schätzmethoden, die sich jeweils auf eine bestimmte Art von Daten, die Verfügbarkeit von Informationen und den Detaillierungsgrad des Projekts konzentrieren. Zu den gängigsten gehören:

- **Ressourcenbasierte Methode**: Schätzung der voraussichtlichen Kosten eines Projekts auf der Grundlage einer detaillierten Analyse aller erforderlichen Ressourcen wie Material, Ausrüstung und Arbeitskräfte. Diese Methode erfordert eine detaillierte Auflistung aller Aufgaben und der für die Durchführung jeder Aufgabe erforderlichen Ressourcen, gefolgt von einer Berechnung ihrer Kosten. Diese Methode ist sehr genau und wird häufig für Kostenschätzungen verwendet.
- **Parametrische Methode**: Verwendung statistischer Modelle zur Kostenschätzung auf der Grundlage von Projektparametern. Dies kann die Analyse der Kosten pro Maßeinheit, z. B. Gebäudefläche oder Arbeitsumfang, und die Anpassung dieser Kosten an die spezifischen Projektbedingungen beinhalten. Diese Methode ist besonders in der Anfangsphase wirksam, wenn noch keine detaillierten Informationen vorliegen.
- Einheitsmethode (Einheitskostenmethode): Berechnet die geschätzten Kosten eines Projekts auf der Grundlage von Kosten pro Maßeinheit (z. B. pro Quadratmeter oder Kubikmeter). Auf diese Weise lassen sich die Kosten verschiedener Projekte oder Projektteile schnell und einfach vergleichen und analysieren.
- Expertenurteil (Delphi-Methode): basiert auf den Meinungen von Experten, die ihre Erfahrungen und

Kenntnisse nutzen, um den Wert eines Projekts zu schätzen. Dieser Ansatz ist nützlich, wenn keine genauen Ausgangsdaten verfügbar sind oder das Projekt einzigartig ist.

Es ist erwähnenswert, dass die parametrische Methode und die Experteneinschätzungen an maschinelle Lernmodelle angepasst werden können. Dies ermöglicht die automatische Erstellung von Projektkosten- und Zeitplanprognosen auf der Grundlage von Trainingsmustern. Beispiele für die Anwendung solcher Modelle werden im Kapitel "Beispiel für die Verwendung von maschinellem Lernen zur Ermittlung von Projektkosten und -terminen" näher erläutert (Abb. 9.3-5).

Dennoch ist die ressourcenbasierte Methode nach wie vor die beliebteste und in der Praxis am weitesten verbreitete Methode. Sie liefert nicht nur eine genaue Einschätzung der geschätzten Kosten, sondern ermöglicht auch die Berechnung der Dauer einzelner Vorgänge auf der Baustelle und des gesamten Projekts als Ganzes (weitere Einzelheiten im Kapitel "Bauzeitpläne und 4D -Projektdaten").

Ressourcenbasierte Methode Kalkulation und Kostenvoranschlag im Bauwesen

Die ressourcenbasierte Kostenrechnung ist eine Methode des betrieblichen Rechnungswesens, bei der die Kosten eines Projekts auf einer direkten Erfassung aller beteiligten Ressourcen beruhen. Im Bauwesen beinhaltet dieser Ansatz eine detaillierte Analyse und Bewertung aller für die Durchführung der Arbeiten erforderlichen Material-, Arbeits- und technischen Ressourcen.

> Die ressourcenbasierte Methode bietet ein hohes Maß an Transparenz und Genauigkeit bei der Budgetplanung, da sie sich auf die tatsächlichen Preise der Ressourcen zum Zeitpunkt der Schätzung konzentriert. Dies ist besonders wichtig in einem instabilen wirtschaftlichen Umfeld, in dem Preisschwankungen die Gesamtkosten eines Projekts erheblich beeinflussen können.

In den folgenden Kapiteln werden wir den Prozess der ressourcenbasierten Kalkulation im Detail betrachten. Zum besseren Verständnis seiner Prinzipien in der Konstruktion ziehen wir eine Analogie mit der Kalkulation eines Abendessens in einem Restaurant. Der Restaurantleiter erstellt bei der Planung des Abends eine Liste der benötigten Produkte, berücksichtigt die Zubereitungszeit der einzelnen Gerichte und multipliziert dann die Kosten mit der Anzahl der Gäste. Im Bauwesen läuft der Prozess ähnlich ab: Für jede Kategorie von Projektelementen (Objekten) werden detaillierte Kostenvoranschläge erstellt Rezepte, und die Gesamtkosten des Projekts werden durch Aufsummierung aller Kosten in einer Gesamtrechnung - dem endgültigen Kostenvoranschlag nach Kategorie - ermittelt.

Die wichtigste und erste Phase des ressourcenbasierten Ansatzes ist die Erstellung der ersten Datenbank des Unternehmens. In der ersten Phase der Kalkulation wird eine strukturierte Liste aller Gegenstände, Materialien, Arbeitsarten und Ressourcen erstellt, die dem Unternehmen im Rahmen seiner Bauprojekte zur Verfügung stehen - vom Nagel im Lager bis zur Beschreibung der Personen über ihre Qualifikation und ihren Stundensatz. Diese Informationen werden in einer einzigen "Datenbank für Baumaterialien und -hilfsmittel" systematisiert - ein tabellarisches Register mit Angaben zu Namen, Eigenschaften, Maßeinheiten und aktuellen Preisen. Diese Datenbank ist die wichtigste und primäre Informationsquelle für alle nachfolgenden Ressourcenberechnungen - sowohl für die Kosten als auch für den Zeitplan der Arbeiten.

Datenbank der Bauressourcen: Katalog der Baumaterialien und -arbeiten

Eine Datenbank oder Tabelle der Bauhilfsmittel und -materialien - enthält detaillierte Informationen über jedes Element, das in einem Bauprojekt verwendet werden kann - ein Produkt, ein Gegenstand, ein Material oder eine Dienstleistung, einschließlich der Bezeichnung, der Beschreibung, der Maßeinheit und der Einheitskosten, die in strukturierter Form erfasst werden. In dieser Tabelle finden Sie alles, von verschiedenen Arten von Brennstoffen und Materialien, die in Projekten verwendet werden, bis hin zu detaillierten Listen von Spezialisten in Form verschiedener Kategorien mit Beschreibungen von Stundensätzen (Abb. 5.1-2).

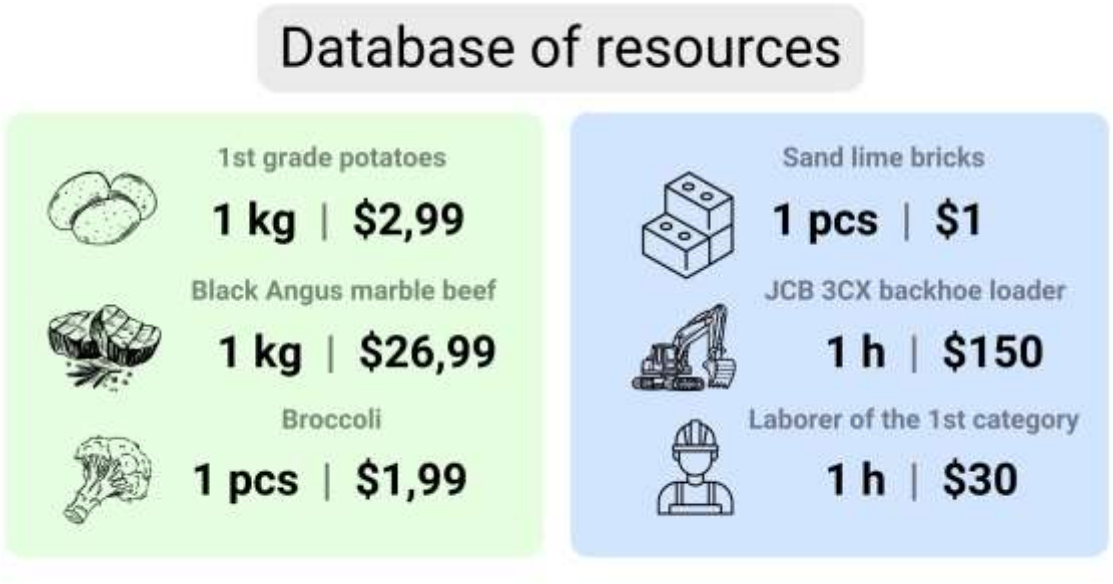


Abbildung 5.1-2 Eine Ressourcentabelle ist eine Zutatenliste, die ein Material und eine Dienstleistung mit einem Stückpreis beschreibt.

> Die "Ressourcendatenbank" ist vergleichbar mit dem Produktkatalog eines Online-Shops, in dem jeder Artikel mit einer detaillierten Beschreibung seiner Eigenschaften versehen ist. Dies erleichtert den Kalkulatoren die Auswahl der richtigen Ressourcen (wie die Auswahl von Produkten beim Hinzufügen zum Warenkorb), die für die Berechnung bestimmter Bauprozesse in Form von Kalkulationen benötigt werden (endgültige Bestellung im Online-Shop).

Eine Ressourcendatenbank kann man sich auch als eine Liste aller Zutaten in einem Restaurantkochbuch vorstellen. Jedes Baumaterial, jede Ausrüstung und jede Dienstleistung ist vergleichbar mit den Zutaten, die in Rezepten verwendet werden. Eine "Ressourcendatenbank" ist eine detaillierte Liste aller Zutaten - Baumaterialien und Dienstleistungen - einschließlich ihrer Kosten pro Einheit: Stück, Meter, Stunde, Liter, usw.

Neue Entitätselemente können der Tabelle "Construction Resource Databases" auf zwei Arten hinzugefügt werden - manuell (Abb. 5.1-3) oder automatisch durch die Integration mit den Bestandsverwaltungssystemen oder Lieferantendatenbanken des Unternehmens.

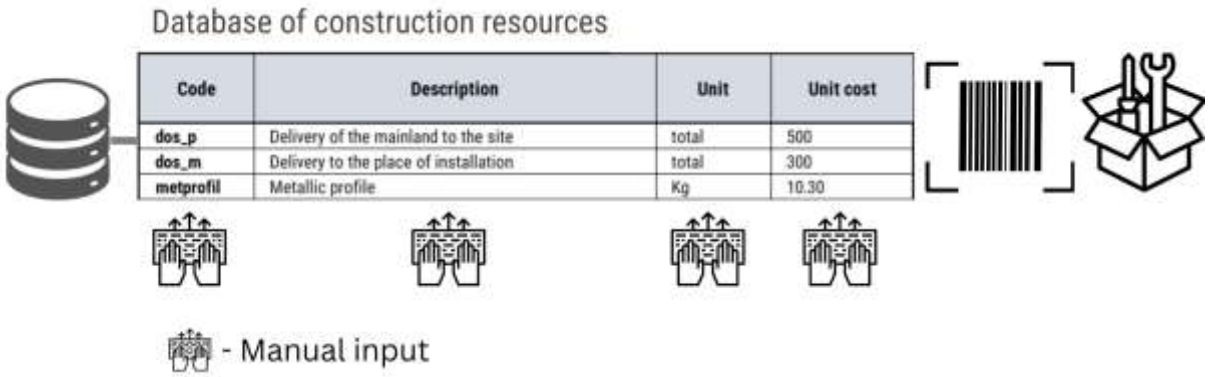


Abb. 5.1-3 Die Ressourcendatenbank wird manuell ausgefüllt oder übernimmt automatisch Daten aus anderen Datenbanken.

Ein typisches mittelgroßes Bauunternehmen verwendet eine Datenbank mit Tausenden und manchmal Zehntausenden von Artikeln mit detaillierten Beschreibungen, die in Bauprojekten verwendet werden können. Diese Daten werden dann automatisch in Verträgen und Projektunterlagen verwendet, um die Zusammensetzung der Arbeiten und Prozesse genau zu beschreiben

> Um mit sich ändernden Marktbedingungen wie der Inflation Schritt zu halten, wird das Attribut "Stückkosten" für jedes Produkt (Ware oder Dienstleistung) in der Ressourcendatenbank (Abb. 5.1-3) regelmäßig manuell oder durch automatisches Herunterladen aktueller Preise aus anderen Systemen oder Online-Plattformen aktualisiert.

Die Aktualisierung der Stückkosten einer Ressource kann monatlich, vierteljährlich oder jährlich erfolgen - je nach Art der Ressource, der Inflation und dem externen Wirtschaftsklima. Solche Aktualisierungen sind notwendig, um die Genauigkeit der Berechnungen und Schätzungen aufrechtzuerhalten, da diese Grundelemente den Ausgangspunkt für die Arbeit der Kostenschätzer bilden. Aktuelle Daten werden verwendet, um Schätzungen, Budgets und Zeitpläne zu erstellen, die die tatsächlichen Marktbedingungen widerspiegeln und das Risiko von Fehlern in späteren Projektkalkulationen verringern.

Zusammenstellung von Kalkulationen und Kalkulation von Arbeiten auf der Grundlage der Ressourcenbasis

Nachdem Sie die "Construction Resource Database" (Abb. 5.1-3) mit Mindesteinheiten gefüllt haben, können Sie damit beginnen, Berechnungen zu erstellen, die für jeden Prozess oder jede Arbeit auf der Baustelle für bestimmte Maßeinheiten berechnet werden: zum Beispiel für einen Kubikmeter Beton, einen Quadratmeter Gipskartonwand, pro Meter Bordstein oder pro Fenstereinbau.

Um beispielsweise eine 1 m² große Ziegelmauer zu errichten (Abb. 5.1-4), werden nach den Erfahrungen aus früheren Projekten etwa 65 Ziegel (Entität "Silikat-Ziegel") zu einem Preis von 1 $ pro Stück (Attribut "Kosten pro Stück") benötigt, also insgesamt 65 $. Außerdem ist es meiner Erfahrung nach erforderlich, 10 Minuten lang Baumaschinen (Entität "JCB 3CX Loader") einzusetzen, die die Steine in der Nähe des Arbeitsbereichs platzieren. Da die Miete für das Gerät 150 $ pro Stunde kostet (), würden 6 Minuten Einsatz etwa 15 $ kosten. Darüber hinaus wird ein Ziegelsteinleger für 2 Stunden benötigt, mit einem Stundensatz von

$30 und einem Gesamtbetrag von $60.

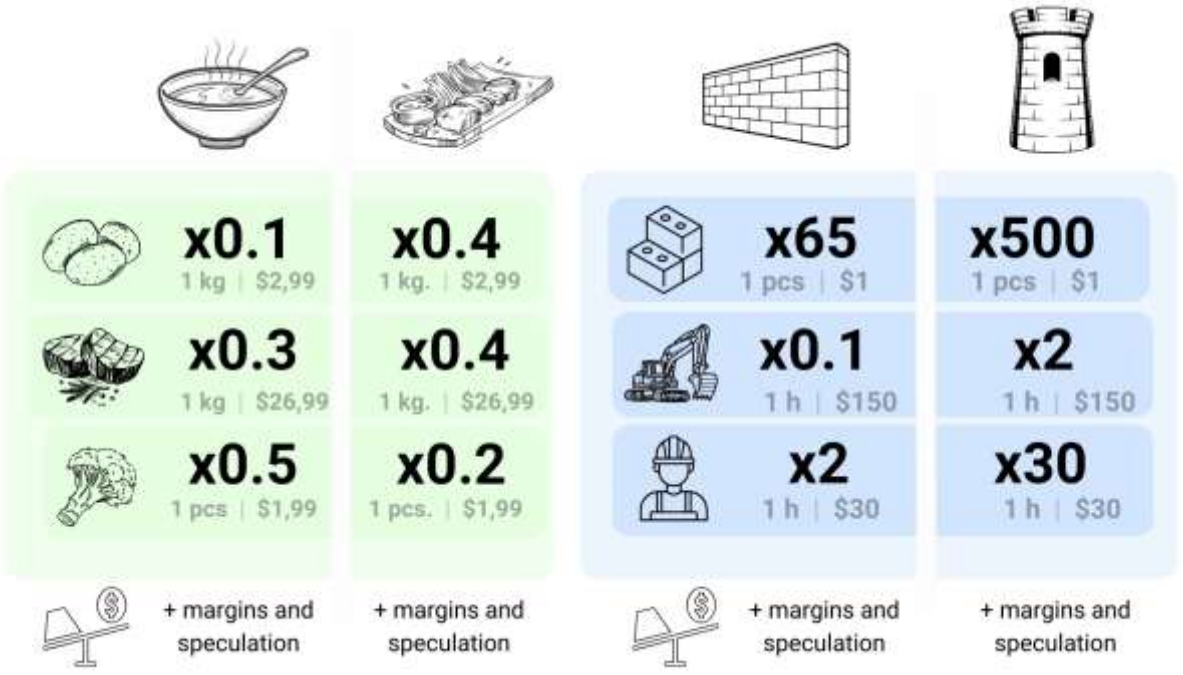


Abb. 5.1-4 Die Kostenberechnungen enthalten eine detaillierte Aufstellung der für die Ausführung der Arbeiten und Prozesse erforderlichen Baumaterialien und Dienstleistungen.

Die Zusammensetzung der Berechnungen (so genannte "Rezepte") beruht auf den Erfahrungen, die das Unternehmen bei der Durchführung einer großen Zahl ähnlicher Arbeiten gesammelt hat. Diese praktischen Erfahrungen werden in der Regel durch Rückmeldungen von der Baustelle gesammelt. Insbesondere sammelt der Polier Informationen direkt auf der Baustelle, indem er die tatsächlichen Arbeitskosten, den Materialverbrauch und die Feinheiten der technischen Abläufe erfasst. In Zusammenarbeit mit der Kalkulationsabteilung werden diese Informationen dann iterativ verfeinert: Prozessbeschreibungen werden verfeinert, der Ressourcenmix wird angepasst und die Kalkulationen werden aktualisiert, um die tatsächlichen Daten der letzten Projekte widerzuspiegeln.

> So wie ein Rezept die Zutaten und Mengen beschreibt, die für die Zubereitung eines Gerichts benötigt werden, enthält ein Kostenvoranschlag eine detaillierte Liste aller Baumaterialien, Ressourcen und Dienstleistungen, die für die Ausführung eines bestimmten Auftrags oder Prozesses erforderlich sind.

Regelmäßig durchgeführte Arbeiten ermöglichen es Arbeitern, Vorarbeitern und Kalkulatoren, sich an der erforderlichen Menge an Ressourcen zu orientieren: Material, Kraftstoff, Arbeitszeit und andere Parameter, die zur Durchführung einer Arbeitseinheit erforderlich sind (Abb. 5.1-5). Diese Daten werden in die Kalkulationssysteme in Form von Tabellen eingegeben, in denen jede Aufgabe und jeder Vorgang durch die Mindestelemente der Ressourcenbasis (mit ständig aktualisierten Preisen) beschrieben wird, was die Genauigkeit der Berechnungen gewährleistet.

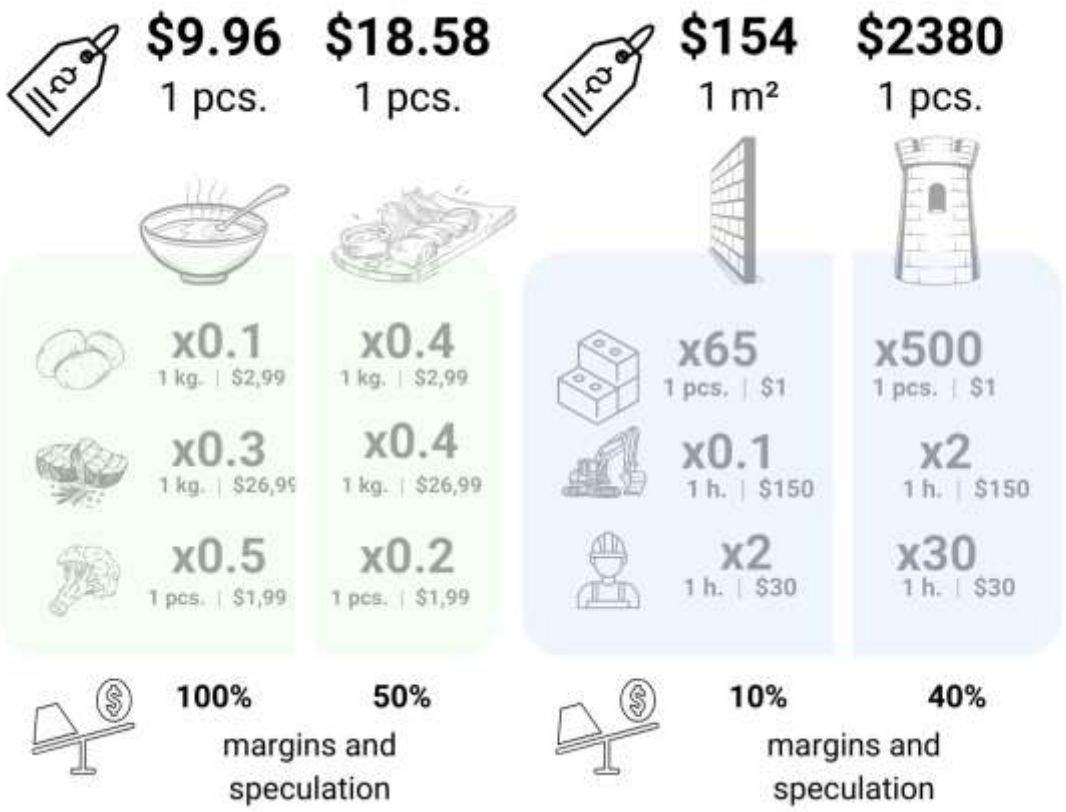


Abb. 5.1-5 Die Stückpreise werden für jeden Auftrag erfasst, wobei das Volumenattribut der Einheit mit ihrer Menge multipliziert und ein Gewinnprozentsatz hinzugefügt wird.

Um die Gesamtkosten jedes Prozesses oder jeder Tätigkeit (Kalkulationsobjekt) zu erhalten, wird das Kostenattribut mit seiner Anzahl und seinen Faktoren multipliziert. Die Koeffizienten können verschiedene Faktoren berücksichtigen, wie z. B. die Komplexität der Arbeit, regionale Merkmale, Inflationsrate, potenzielle Risiken (erwarteter Gemeinkostenanteil) oder Spekulation (zusätzlicher Gewinnfaktor).

> Der Kalkulator setzt als Analytiker die Erfahrungen und Empfehlungen des Poliers in standardisierte Schätzungen um, die Bauprozesse durch Ressourceneinheiten in tabellarischer Form beschreiben. Im Wesentlichen besteht die Aufgabe des Kalkulators darin, die von der Baustelle kommenden Informationen zu sammeln und durch Parameter und Koeffizienten zu strukturieren.

Die endgültigen Kosten pro Arbeitseinheit (z. B. Quadrat- oder Kubikmeter oder eine Installation einer einzelnen Einheit) umfassen also nicht nur die direkten Material- und Arbeitskosten, sondern auch Firmenaufschläge, Gemeinkosten, Versicherungen und andere Faktoren (Abb. 5.1-6)

Gleichzeitig müssen wir uns nicht mehr um die tatsächlichen Preise in (Rezept-)Berechnungen kümmern, da die tatsächlichen Preise immer in der "Ressourcenbasis" (Zutatentabelle) enthalten sind. Auf der Ebene der Berechnungen werden die Daten aus der Ressourcenbasis automatisch in die Tabelle geladen (z. B. nach Artikelcode oder seiner eindeutigen Kennung), die die Beschreibung und den tatsächlichen Preis pro

Einheit laden, die wiederum automatisch von Online-Plattformen oder Online-Shops für Baumaterialien geladen werden können. Der Kalkulator auf der Ebene der Kalkulationen muss die Arbeit oder den Prozess nur durch das Attribut "Menge der Ressourcen" und zusätzliche Faktoren beschreiben.

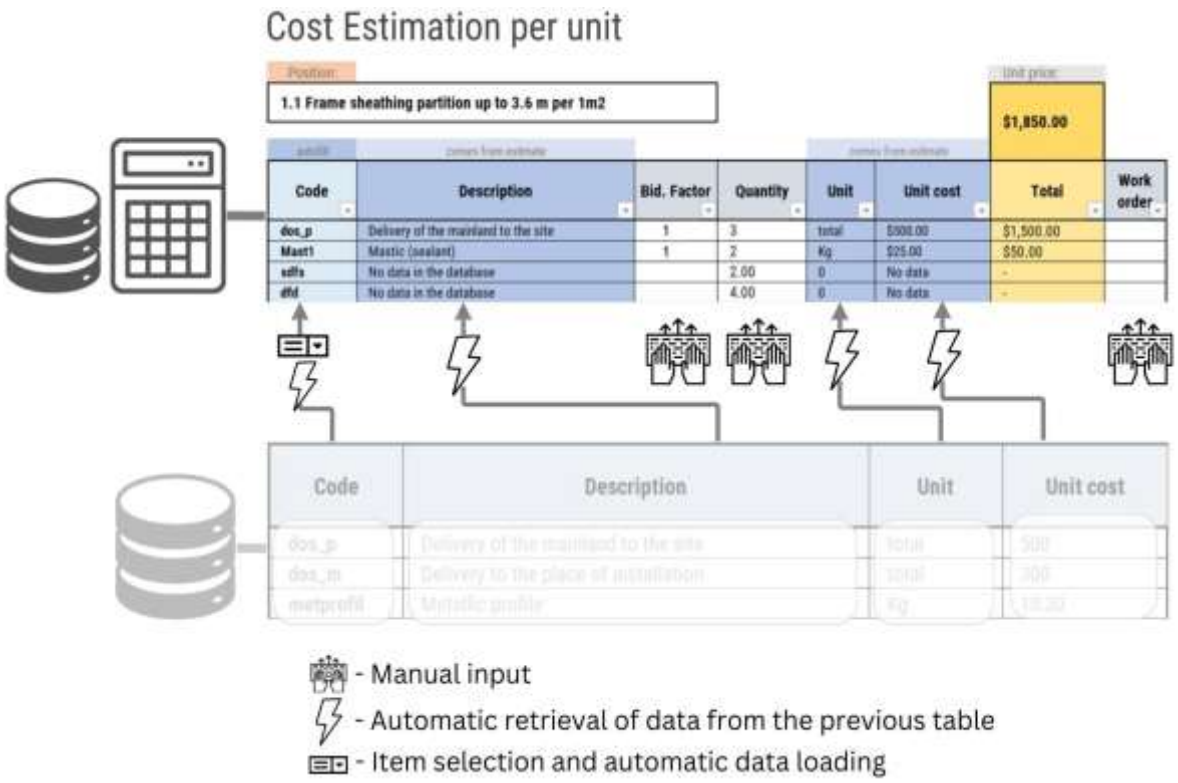


Abb. 5.1-6 In der Phase der Berechnung der Stückkosten der Arbeit werden nur die Attribute der Anzahl der benötigten Ressourcen ausgefüllt, alles andere wird automatisch aus der Ressourcendatenbank geladen.

Die erstellten Auftragskalkulationen werden in Form von Vorlagentabellen für typische Projekte gespeichert, die direkt mit der Datenbank für Bauhilfsmittel und Materialien verknüpft sind. Diese Vorlagen stellen standardisierte Rezepte für sich wiederholende Arbeitstypen für künftige Projekte dar, wodurch die Einheitlichkeit der Berechnungen im gesamten Unternehmen gewährleistet wird.

Wenn sich die Kosten einer Ressource in der Datenbank ändern (Abb. 5.1-3) - sei es manuell oder automatisch durch das Herunterladen aktueller Marktpreise (z.B. bei Inflation) - werden die Aktualisierungen sofort in allen verknüpften Kalkulationen berücksichtigt (Abb. 5.1-6). Das bedeutet, dass nur die Ressourcenbasis geändert werden muss, während die Kalkulationsvorlagen und Schätzungen im Laufe der Zeit unverändert bleiben. Dieser Ansatz gewährleistet die Stabilität und Reproduzierbarkeit der Berechnungen bei eventuellen Preisschwankungen, die nur in einer relativ einfachen Ressourcentabelle berücksichtigt werden (Abb. 5.1-3).

Für jedes neue Projekt wird eine Kopie der Standardkalkulationsvorlage erstellt, die es Ihnen ermöglicht, Änderungen vorzunehmen und Aktivitäten an spezifische Anforderungen anzupassen, ohne die ursprüngliche, vom Unternehmen angenommene Vorlage zu verändern. Dieser Ansatz bietet Flexibilität bei der Anpassung der Berechnungen: Sie können die Besonderheiten der Baustelle und die Wünsche des Kunden berücksichtigen, Risiko- oder Rentabilitätskoeffizienten (Spekulationen) einführen - all das, ohne die Standards des Unternehmens zu verletzen. Dies hilft dem Unternehmen, ein Gleichgewicht zwischen Gewinnmaximierung, Kundenzufriedenheit und Erhaltung seiner Wettbewerbsfähigkeit zu finden.

In einigen Ländern sind solche Kalkulationsvorlagen, die sich über Jahrzehnte angesammelt haben, auf nationaler Ebene standardisiert und werden Teil der nationalen Baukalkulationssystemstandards (Abb. 5.1-7).

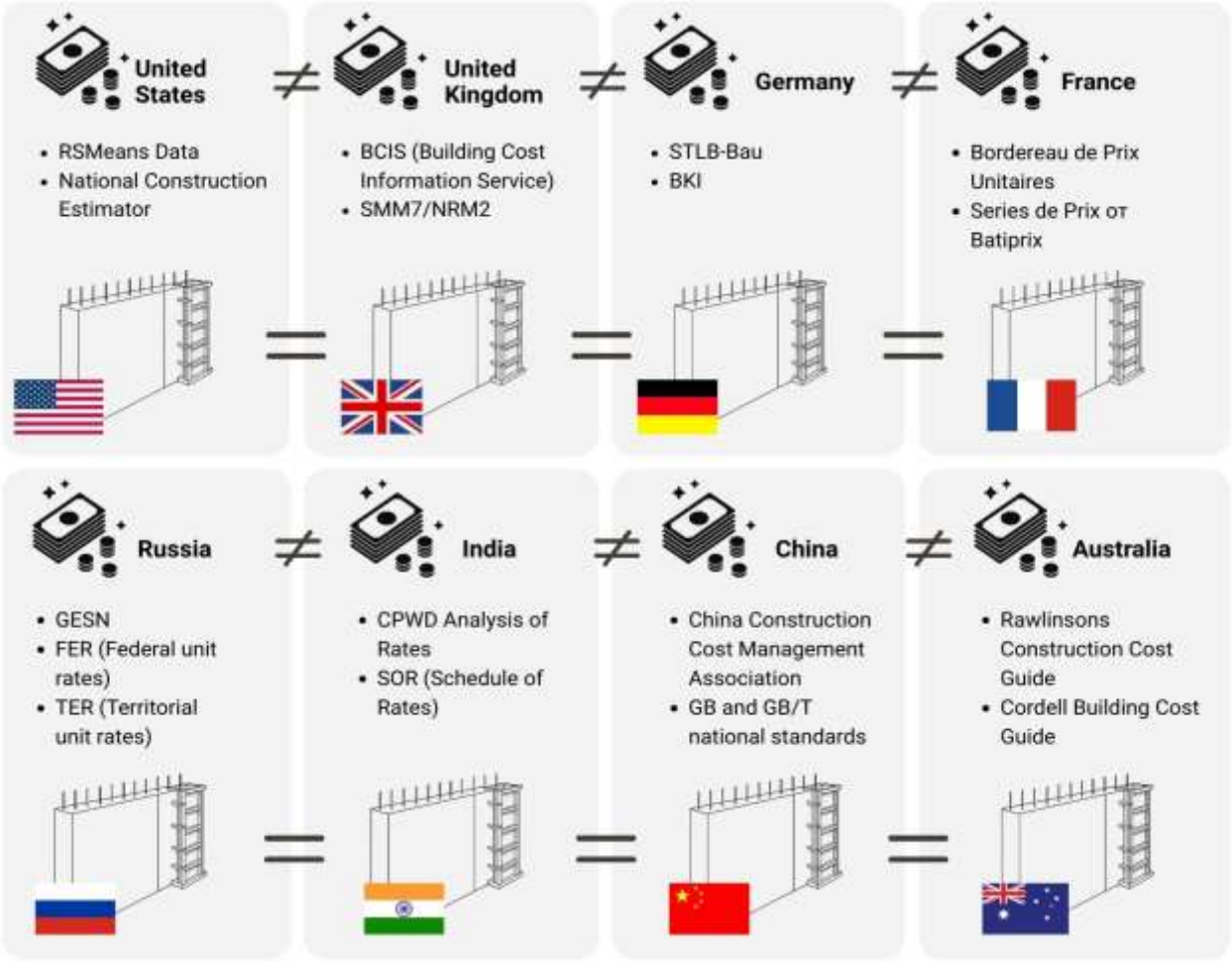


Abb. 5.1-7 Verschiedene Länder auf der ganzen Welt haben ihre eigenen Kalkulationsregeln mit ihren eigenen (präskriptiven) Kompendien und Normen für Bauarbeiten zur Kalkulation desselben Elements.

Solche standardisierten Ressourcenschätzungsgrundlagen (Abb. 5.1-7) sind für Marktteilnehmer verpflichtend zu verwenden, insbesondere bei öffentlich finanzierten Projekten. Eine solche Standardisierung gewährleistet Transparenz, Vergleichbarkeit und Fairness bei der Preisbildung und den vertraglichen Verpflichtungen für den Kunden

Endgültige Projektkalkulation: von Schätzungen zu Budgets

Staatliche und branchenspezifische Schätzungsnormen spielen in der Baupraxis der verschiedenen Länder eine unterschiedliche Rolle. Während einige Länder die strikte Einhaltung einer einzigen Norm vorschreiben, verfolgen die meisten entwickelten Volkswirtschaften einen flexibleren Ansatz. In den Marktwirtschaften dienen die staatlichen Baunormen in der Regel nur als Grundlinie. Die Bauunternehmen passen diese Normen an ihre Betriebsmodelle an oder überarbeiten sie vollständig und ergänzen sie um ihre eigenen, individuellen Faktoren. Diese Anpassungen spiegeln die Erfahrung des Unternehmens, die Effizienz des Ressourcenmanagements und häufig auch Faktoren wider, bei denen zum Beispiel die Spekulationsgewinne eines Unternehmens berücksichtigt werden können.

> Infolgedessen können das Ausmaß des Wettbewerbs, die Marktnachfrage, die Zielmargen und sogar die Beziehungen zu bestimmten Kunden zu erheblichen Abweichungen von den standardisierten Normen führen. Diese Praxis bietet dem Markt Flexibilität, erschwert aber auch den transparenten Vergleich von Angeboten verschiedener Auftragnehmer und führt in dieser Phase des Kalkulationsprozesses ein Element der spekulativen Preisgestaltung in die Bauindustrie ein.

Nachdem die Kalkulationsvorlagen für die einzelnen Aktivitäten und Prozesse erstellt wurden - oder, was häufiger der Fall ist, einfach von den Standardschätzungen der Regierung kopiert wurden (Abb. 5.1-7), wobei Koeffizienten hinzugefügt wurden, um die "Besonderheiten" eines bestimmten Unternehmens widerzuspiegeln, besteht der letzte Schritt darin, die Kosten jedes Postens mit dem entsprechenden Attribut des Arbeitsumfangs oder der Prozesse im neuen Projekt zu multiplizieren.

> Bei der Berechnung der Gesamtkosten eines Neubauprojekts ist der wichtigste Schritt die Zusammenfassung der Kosten aller Kalkulationspositionen, multipliziert mit dem Volumen dieser Arbeitspositionen im Projekt.

Um die Gesamtkosten des Projekts zu ermitteln, beginnen wir in unserem vereinfachten Beispiel mit der Berechnung der Kosten für den Bau eines Quadratmeters Wand und multiplizieren die Kosten für dessen Berechnung (z.B. Arbeit "1m² Standardinstallation von Wandelementen") mit der Gesamtzahl der Quadratmeter Wände im Projekt (z.B. das Attribut "Fläche" oder "Menge" (Abb. 5.1-8) einer Entität vom Typ "Wandelemente" aus dem CAD des Projekts oder die Berechnungen des Vorarbeiters).

Auf ähnliche Weise berechnen wir die Kosten für alle Elemente des Projekts (Abb. 5.1-8): Wir nehmen die Kosten pro Arbeitseinheit und multiplizieren sie mit dem Volumen eines bestimmten Elements oder einer Gruppe von Elementen im Projekt. Der Kalkulator muss nur noch die Anzahl dieser Elemente, Aktivitäten oder Prozesse im Projekt in Form von Volumen oder Menge eingeben. Auf diese Weise lässt sich automatisch eine vollständige Baukostenschätzung erstellen.

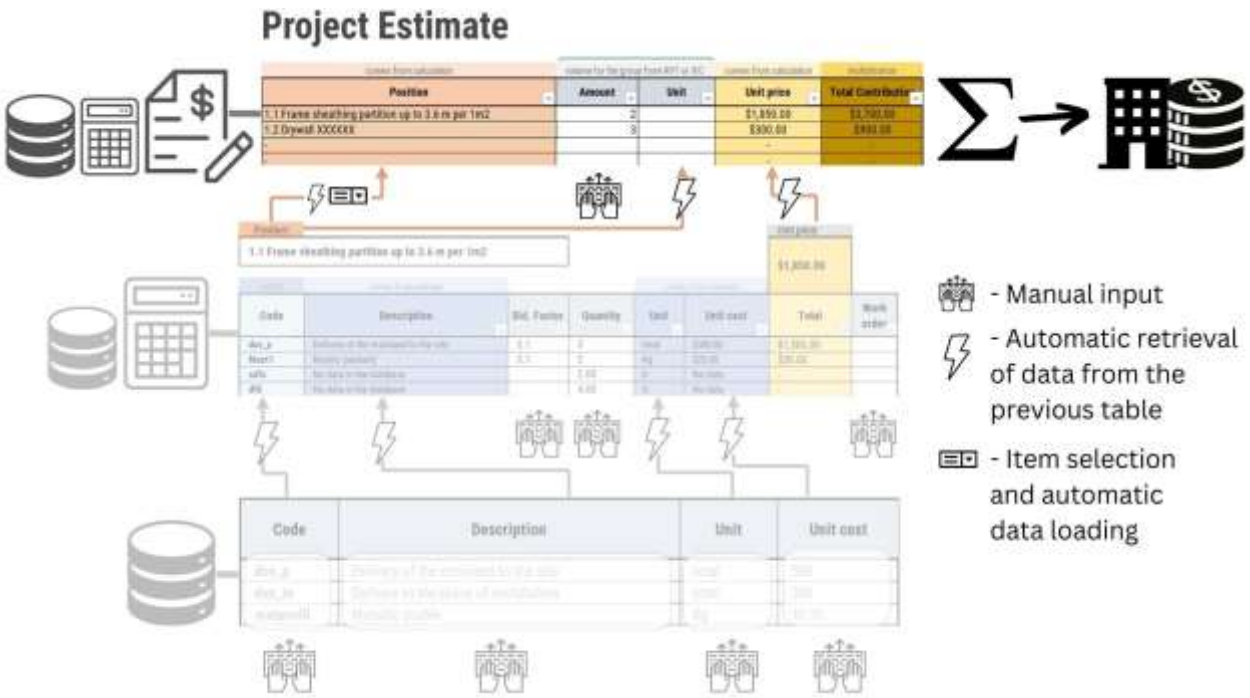


Abb. 5.1-8 In der Phase der Erstellung eines Kostenvoranschlags geben wir nur den Arbeitsumfang ein.

Wie bei den Kalkulationen laden wir auf dieser Ebene automatisch fertig kalkulierte Positionen hoch (aus der Kalkulationsvorlage oder neue, die aus der Vorlage kopiert und bearbeitet wurden), die automatisch die aktuellen Kosten pro Arbeitseinheit mit sich bringen (die automatisch aus der Ressourcendatenbank aktualisiert werden (Abb. 5.1-8 untere Tabelle)). Dementsprechend werden bei einer Änderung der Daten in der Ressourcendatenbank oder den Kalkulationstabellen die Daten in der Kalkulation automatisch für den aktuellen Tag aktualisiert, ohne dass die Kalkulation oder die Kalkulation selbst geändert werden muss.

In einem Restaurant werden die endgültigen Kosten einer Veranstaltung auf ähnliche Weise berechnet und entsprechen den endgültigen Kosten des gesamten Abendessens, wobei sich die Kosten der einzelnen Gerichte multipliziert mit der Anzahl der Gäste zu den Gesamtkosten der Rechnung addieren (Abb. 5.1-9). Und genau wie im Baugewerbe können sich die Rezepte für das Kochen in einem Restaurant über Jahrzehnte hinweg nicht ändern. Im Gegensatz zu den Preisen, wo sich die Kosten für die Zutaten stündlich ändern können.

> So wie ein Restaurantbesitzer die Kosten für jede Mahlzeit mit der Anzahl der Portionen und Personen multipliziert, um die Gesamtkosten der Veranstaltung zu ermitteln, addiert der für die Kostenschätzung zuständige Manager die Kosten aller Projektkomponenten, um zu einem vollständigen Baukostenvoranschlag zu gelangen.

So werden für jede Aktivität im Projekt die endgültigen Kosten ermittelt (Abb. 5.1-9), die - multipliziert mit dem Attributvolumen der dieser Aktivität entsprechenden Einheit - die Kosten von Gruppen von Aktivitäten ergeben, aus denen sich die endgültigen Kosten des gesamten Projekts ergeben.

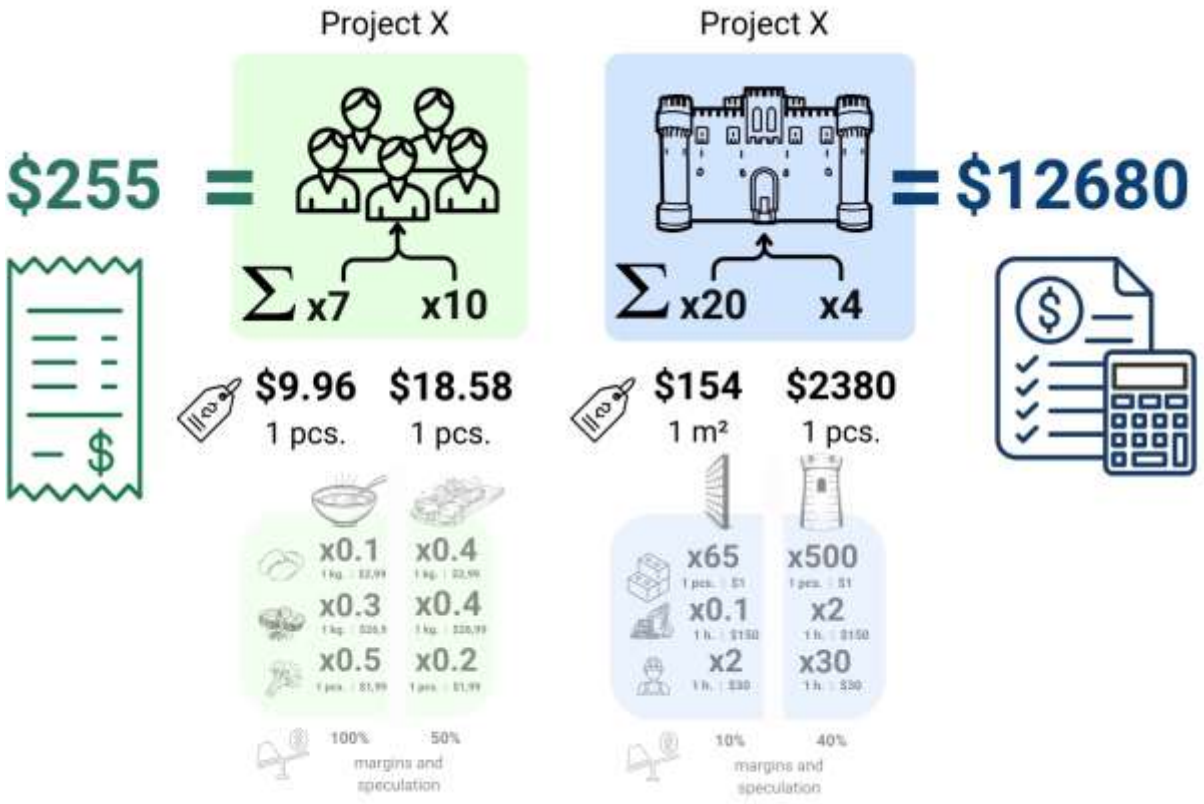


Abbildung 5.1-9 Der endgültige Kostenvoranschlag wird berechnet, indem die Arbeitskostenattribute der einzelnen Elemente mit ihren Umfangsattributen addiert werden.

Die Gesamtprojektkosten (Abb. 5.1-8) vermitteln ein finanzielles Bild des Projekts, das es Kunden, Investoren oder Finanzorganisationen ermöglicht, das Gesamtbudget und die finanziellen Ressourcen zu verstehen, die für die Durchführung des Projekts an einem bestimmten Tag und unter Berücksichtigung der aktuellen Preise erforderlich sind.

Und wenn die Prozesse der Zusammenstellung von Ressourcengrundlagen, Berechnungen und Schätzungen (Prozessrezepte) bereits seit Zehntausenden von Jahren ausgearbeitet, halbautomatisiert und verfeinert und auf staatlicher Ebene aufgezeichnet wurden, dann ist die automatische Beschaffung von qualitativen Informationen über den Umfang und die Menge der Elemente für die letzte Stufe der endgültigen Schätzung - heute noch ein Engpass in den Prozessen aller Berechnungen der Kosten- und Zeitattribute des Projekts und im Allgemeinen des Gesamtbudgets des Projekts.

> Seit Tausenden von Jahren besteht die traditionelle Methode zur Berechnung von Volumen in der manuellen Messung von Volumen und Menge anhand von flachen Zeichnungen. Mit dem Aufkommen des digitalen Zeitalters haben die Unternehmen entdeckt, dass Volumen- und Mengeninformationen nun automatisch aus den in CAD-Modellen enthaltenen geometrischen Daten extrahiert werden können, was die Jahrtausende alten Methoden zur Gewinnung quantitativer Daten revolutioniert.

Moderne Ansätze zur Prozessabschätzung und -kalkulation beinhalten die automatische Extraktion von volumetrischen und quantitativen Attributen aus CAD-Datenbanken, die hochgeladen und unter mit dem Kalkulationsprozess verbunden werden können, um aktuelle Projektgruppenvolumina in jeder Phase der Planung bis hin zum Betrieb zu erhalten.

KAPITEL 5.2.
MENGENERMITTLUNG UND AUTOMATISCHE ERSTELLUNG VON KOSTENVORANSCHLÄGEN UND ZEITPLÄNEN

Übergang von 3D zu 4D und 5D: Verwendung volumetrischer und quantitativer Parameter

Wenn die Kalkulationstabellen mit den beschriebenen Prozessen über die Ressourcen (Abb. 5.1-8) vorliegen, besteht der nächste Schritt darin, automatisch die Volumen- oder Mengenparameter für eine Gruppe von Elementen zu erhalten, die für die Berechnungen und für die endgültige Schätzung benötigt werden.

Volumeneigenschaften von Projektelementen - z.B. Wände oder Platten - können automatisch aus CAD-Datenbanken extrahiert werden. In CAD-Programmen erstellte parametrische Objekte werden mit Hilfe eines Geometriekerns in numerische Werte für Länge, Breite, Fläche, Volumen und andere Parameter umgewandelt. Der Prozess der Gewinnung von Volumina auf der Grundlage von 3D-Geometrie wird im nächsten, sechsten Teil (Abb. 6.3-3), der der Arbeit mit CAD gewidmet ist, ausführlicher behandelt (BIM). Neben den Volumina kann auch die Anzahl ähnlicher Elemente aus der Datenbank des CAD-Modells ermittelt werden, indem die Objekte nach Kategorien und Eigenschaften gefiltert und gruppiert werden. Diese Parameter, die eine Gruppierung ermöglichen, bilden die Grundlage für die Verknüpfung von Projektelementen durch Ressourcenberechnungen mit Kalkulationen, endgültigen Kostenvoranschlägen und dem Budget für das gesamte Projekt.

So wird das aus dem 3D (CAD)-Modell extrahierte Datenmodell durch neue Parameterschichten ergänzt, die als 4D und 5D bezeichnet werden. In den neuen Attributschichten für Entitäten, 4D (Zeit) und 5D (Kosten), werden 3D-Geometriedaten als Quelle für die Attributwerte von Entitätsvolumen verwendet.

- **4D** ist eine Parameter-Informationsschicht, die den 3D-Parametern der Elemente Informationen über die Dauer der Bauarbeiten hinzufügt. Diese Daten sind für die Planung von Zeitplänen und die Verwaltung von Projektzeitplänen unerlässlich
- **5D** ist die nächste Ausbaustufe des Datenmodells, bei der die Elemente mit Kostenmerkmalen ergänzt werden. Dadurch wird den geometrischen Informationen ein finanzieller Aspekt hinzugefügt: die Kosten für Materialien, Arbeiten und Ausrüstung, was Budgetberechnungen, Rentabilitätsanalysen und Kostenmanagement während des Bauprozesses ermöglicht.

Kosten- und 3D, 4D und 5D Attributdaten von Gruppen von Projekteinheiten werden in ähnlicher Weise beschrieben wie Berechnungen in modularen ERP-, PIMS -Systemen (oder Excel -ähnlichen Tools) und werden für die automatische Kosten- und Budgetplanung sowohl einzelner Gruppen als auch des gesamten Projektbudgets verwendet.

5D-Attribute und Abrufen von Attributvolumen aus CAD

Bei der Erstellung des endgültigen Kostenvoranschlags eines Bauprojekts, dessen Vorbereitung wir in den vorangegangenen Kapiteln besprochen haben (Abb. 5.1-8), werden die Volumenattribute für jede Kategorie

von Projektelementen entweder manuell erfasst oder aus den von der CAD-Software bereitgestellten Spezifikationen der Volumenattribute extrahiert.

> Bei der traditionellen manuellen Methode der Mengenermittlung analysieren Vorarbeiter und Kalkulatoren Zeichnungen, die seit Tausenden von Jahren als Linien auf Papier und seit 30 Jahren in digitalen Formaten wie PDF (PLT) oder DWG vorliegen. Auf der Grundlage ihrer beruflichen Erfahrung messen sie die erforderlichen Arbeits- und Materialmengen, oft mit Lineal und Winkelmesser. Diese Methode erfordert einen erheblichen Arbeits- und Zeitaufwand sowie eine besondere Aufmerksamkeit für Details.

Die Bestimmung von Umfangsattributen auf diese Weise kann je nach Umfang des Projekts einige Tage bis mehrere Monate dauern. Da alle Messungen und Berechnungen manuell durchgeführt werden, besteht außerdem die Gefahr menschlicher Fehler, die zu ungenauen Daten führen können, was sich wiederum auf Fehler bei der Schätzung von Projektzeit und -kosten auswirkt, für die das gesamte Unternehmen verantwortlich gemacht wird.

Moderne Methoden, die auf der Verwendung von CAD-Datenbanken basieren, vereinfachen die Berechnung von Volumen erheblich. In CAD-Modellen enthält die Geometrie von Elementen bereits Volumenattribute, die automatisch berechnet (über den Geometriekern (Abb. 6.3-3)) und in Tabellenform dargestellt oder exportiert werden können.

In einem solchen Szenario fragt die Kalkulationsabteilung den CAD-Konstrukteur nach Daten über die Mengen- und Volumeneigenschaften der Projektelemente. Diese Daten werden als Tabellenkalkulationen exportiert oder direkt in Kalkulationsdatenbanken - sei es Excel, ERP oder PMIS - integriert. Dieser Prozess beginnt oft nicht mit einer formellen Anfrage, sondern mit einem kurzen Dialog zwischen dem Kunden (Initiator) und dem Architekten Kalkulator des Bau- oder Planungsunternehmens. Im Folgenden wird anhand eines vereinfachten Beispiels gezeigt, wie aus der alltäglichen Kommunikation eine strukturierte Tabelle für automatische Berechnungen (QTO) entsteht:

- Kunde - "*Ich möchte ein weiteres Stockwerk an das Gebäude anbauen, in der gleichen Konfiguration wie das erste Stockwerk*".
- Architekt (CAD) - "*Hinzufügen einer dritten Etage, die Konfiguration ist die gleiche wie die der zweiten Etage*". Und nach dieser Nachricht sendet eine neue CAD-Version des Projekts an den Kalkulator.
- *Estimator* führt automatisch Gruppierungen und Berechnungen durch (ERP, PMIS, Excel) - "*Ich lasse das Projekt durch eine Excel-Tabelle mit QTO-Regeln (ERP, PMIS) laufen, ermittle die Volumina nach Kategorie für die neue Etage und erstelle einen Kostenvoranschlag*"

Als Ergebnis wird der Textdialog in eine Tabellenstruktur mit Gruppierungsregeln umgewandelt:

Element	Kategorie	Boden
Decke	OST_Slabs	3
Säule	OST_StructuralColumns	3

Element	Kategorie	Boden
Treppe	OST_Stairs	3

Nach dem Prozess der automatischen Gruppierung des CAD-Modells des Konstrukteurs gemäß den QTO-Regeln des Kalkulators und der automatischen Multiplikation der Volumina mit den Ressourcenberechnungen (Abb. 5.1-8) erhalten wir die folgenden Ergebnisse, die an den Kunden gesendet werden:

Element	Band	Boden	Preis pro Einheit.	Gesamtkosten
Decke	420 м²	3	150 €/м²	**63 000 €**
Säule	4 Stk.	3	2450 €/Stück.	**9 800 €**
Treppe	2 Stk.	3	4.300 €/Stück.	**8 600 €**
GESAMT:	-	-	-	**81 400 €**

- Der Kunde - "*Danke, das ist ziemlich viel, wir müssen ein paar Zimmer abbauen*". Und der Zyklus wiederholt sich viele Male.

Dieses Szenario kann sich viele Male wiederholen, insbesondere in der Genehmigungsphase, in der der Kunde eine sofortige Rückmeldung erwartet. In der Praxis können sich solche Prozesse jedoch über Tage oder sogar Wochen hinziehen. Dank der Einführung automatischer Gruppierungs- und Berechnungsregeln können Tätigkeiten, die früher viel Zeit in Anspruch nahmen, heute in wenigen Minuten erledigt werden. Die automatische Erfassung von Mengen durch Gruppierungsregeln beschleunigt nicht nur die Berechnungen und Kostenvoranschläge, sondern verringert durch die Minimierung des menschlichen Faktors auch die Fehlerwahrscheinlichkeit und sorgt für eine transparente und genaue Bewertung der Projektkosten.

> Wenn die Anforderungen der Schätzungsabteilung bereits bei der Erstellung des 3D-Modells im CAD-System berücksichtigt wurden (was in der Praxis noch selten der Fall ist) und die Namen, Bezeichner der Elementgruppen und ihre Klassifizierungsattribute in Form von Parametern festgelegt werden, die mit den Strukturen der Schätzungsgruppen und -klassen übereinstimmen, dann können die volumetrischen Attribute automatisch und ohne zusätzliche Transformationen in die Schätzungssysteme übertragen werden.

Die automatische Extraktion von volumetrischen Attributen aus dem CAD in Form von Spezifikationstabellen ermöglicht es, schnell aktuelle Daten über die Kosten einzelner Arbeiten und des gesamten Projekts zu erhalten (Abb. 5.2-1). Indem nur die CAD-Datei mit den Projektvolumina im Kalkulationsprozess oder im

Kalkulationssystem aktualisiert wird, kann das Unternehmen den Kostenvoranschlag unter Berücksichtigung der neuesten Änderungen schnell neu berechnen und so eine hohe Genauigkeit und Konsistenz aller nachfolgenden Berechnungen gewährleisten.

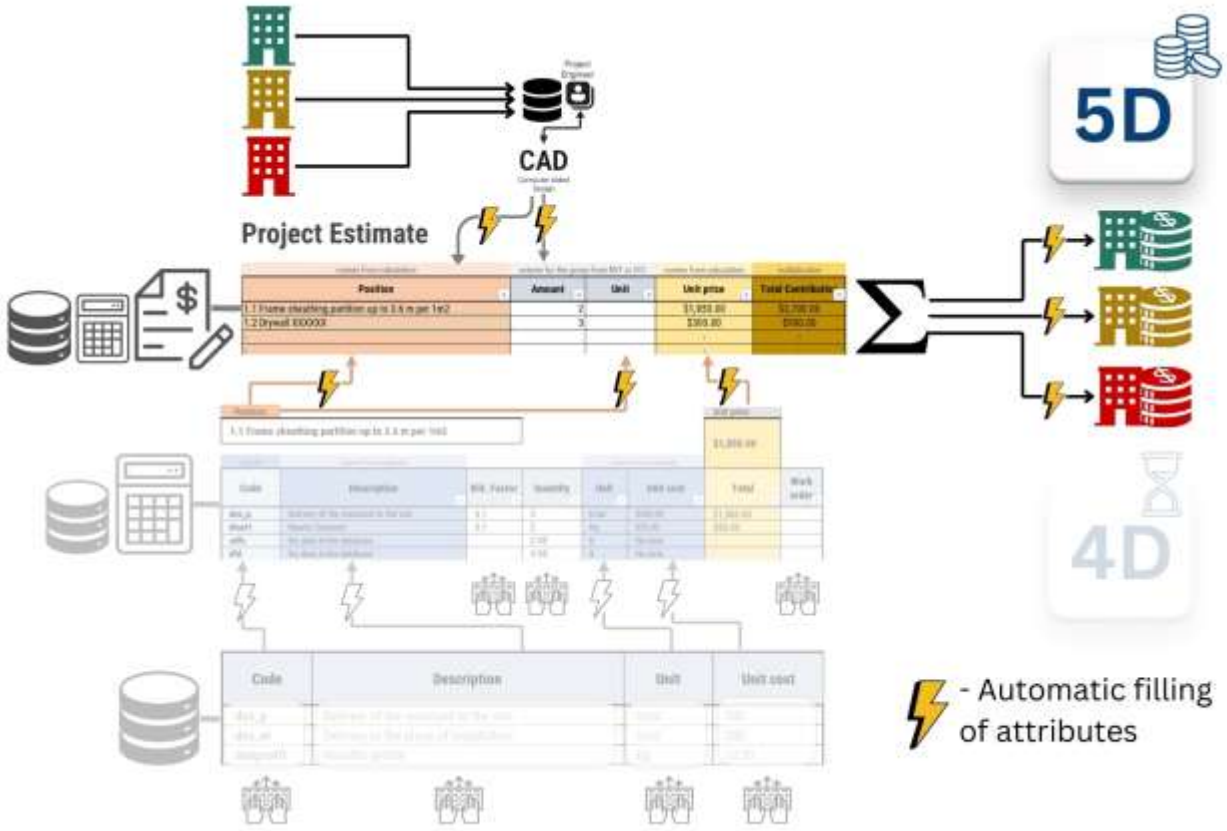


Abb. 5.2-1 Volumenattribute aus CAD-Tabellen oder Datenbanken werden automatisch in die Kalkulation übernommen, so dass Sie die Gesamtkosten des Projekts sofort berechnen können.

Mit der zunehmenden Komplexität von Investitionsprojekten wird die Berechnung des Gesamtbudgets und die Analyse der Gesamtkosten von Projekten in einem solchen Szenario (Abb. 5.2-1) - zu einem wichtigen Instrument für eine fundierte Entscheidungsfindung.

> Laut der Accenture-Studie Creating More Value through Capital Projects (2024) [20] integrieren führende Unternehmen aktiv Datenanalysen in digitale Initiativen und nutzen historische Informationen, um Ergebnisse vorherzusagen und zu optimieren. Untersuchungen zeigen, dass immer mehr Eigentümer und Betreiber Big-Data-Analysen einsetzen, um Markttrends vorherzusagen und die wirtschaftliche Tragfähigkeit zu bewerten, bevor mit der Planung begonnen wird. Dies wird durch die Analyse von Datenlagern aus einem bestehenden Projektportfolio erreicht. Darüber hinaus setzen 79 Prozent der Betreiber "robuste" prädiktive Analysen ein, um die Projektleistung zu bewerten und betriebliche Entscheidungen in Echtzeit zu unterstützen.

Ein modernes, effektives Management von Bauprojekten ist untrennbar mit der Verarbeitung und Analyse großer Informationsmengen in allen Phasen der Planung und den Prozessen, die der Planung vorausgehen,

verbunden. Der Einsatz von Data Warehouses, Ressourcenberechnungen, prädiktiven Modellen und maschinellem Lernen ermöglicht nicht nur die Minimierung von Risiken bei Berechnungen, sondern auch strategische Entscheidungen über die Projektfinanzierung in den frühen Phasen der Planung. Im neunten Teil des Buches werden wir mehr über Data Warehouses und prädiktive Modelle sprechen, die die Berechnungen ergänzen werden.

Die automatische Ermittlung der volumetrischen Parameter von Elementen aus CAD-Projekten, die für die Erstellung von Kostenvoranschlägen erforderlich sind, erfolgt mit Hilfe der Gruppierungswerkzeuge QTO (Quantity Take-Off). Die QTO-Tools arbeiten mit der Gruppierung aller Projektobjekte nach speziellen Elementkennungen oder Elementattributparametern unter Verwendung von Spezifikationen und Tabellen, die in der CAD-Datenbank erstellt wurden.

QTO Quantity Take-Off: Gruppierung von Projektdaten nach Attributen

QTO (Quantity Take-Off) im Bauwesen ist der Prozess der Extraktion der Mengen von Elementen, die für die Realisierung eines Projekts erforderlich sind. In der Praxis ist die Mengenermittlung oft ein halbmanueller Prozess, der die Erfassung von Daten aus verschiedenen Quellen beinhaltet: PDF-Dokumente, DWG-Zeichnungen und digitale CAD-Modelle.

Bei der Arbeit mit Daten aus CAD-Datenbanken wird der QTO-Prozess als eine Abfolge von Filter-, Sortier-, Gruppierungs- und Aggregationsoperationen realisiert. Modellelemente werden nach den Parametern von Klassen, Kategorien und Typen ausgewählt, woraufhin ihre quantitativen Attribute - wie Volumen, Fläche, Länge oder Menge - gemäß der Berechnungslogik zusammengefasst werden (Abb. 5.2-2).

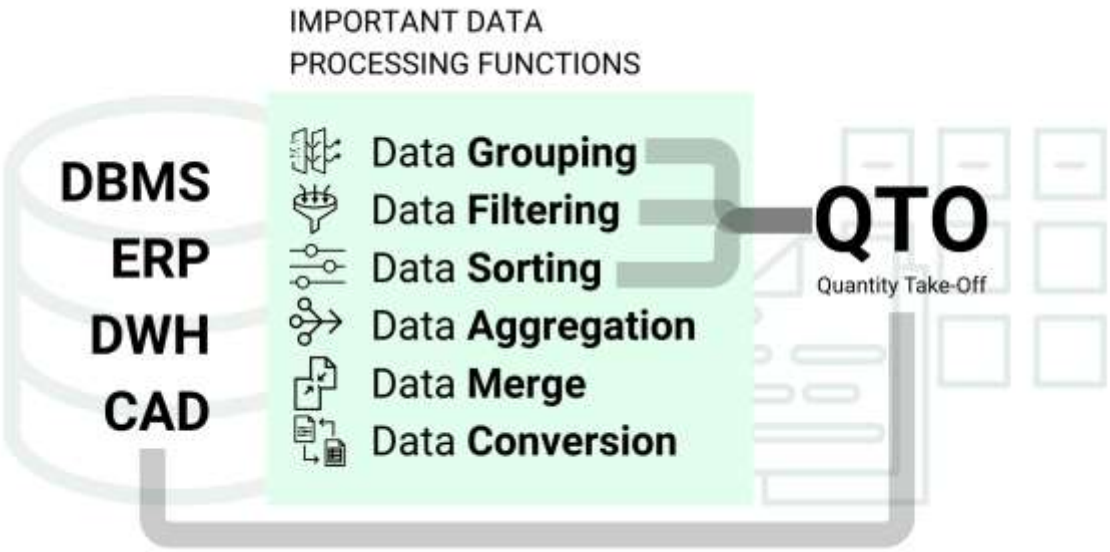


Abb. 5.2-2 Datengruppierung und -filterung sind die am häufigsten verwendeten Funktionen in Datenbanken und Data Warehouses.

Der QTO-Prozess (Filterung und Gruppierung) ermöglicht die Systematisierung von Daten, die Erstellung von Spezifikationen und die Vorbereitung von Eingangsinformationen für die Berechnung von Schätzungen,

Einkäufen und Arbeitsplänen. Die Grundlage der QTO ist die Klassifizierung der Elemente nach der Art der zu messenden Attribute. Für jedes Element oder jede Gruppe von Elementen wird der entsprechende quantitative Messparameter ausgewählt. Zum Beispiel:

- **Attribut Länge** (Bordsteinkante - in Metern)
- **Flächenattribut** (Gipskartonarbeiten - in Quadratmetern)
- **Volumenattribut** (Betonwerk - in Kubikmetern)
- **Mengenattribut** (Fenster - pro Stück)

Zusätzlich zu den rechnerisch aus der Geometrie generierten Volumenkennwerten werden in den Berechnungen nach der QTO-Gruppierung häufig Überschreitungsfaktoren (Abb. 5.2-12 z.B. 1,1 zur Berücksichtigung von 10 Prozent für Logistik und Einbau) eingesetzt - Korrekturwerte, die Verluste, Einbau, Lagerung oder Transport berücksichtigen. Auf diese Weise lässt sich der tatsächliche Materialverbrauch genauer vorhersagen und sowohl Engpässe als auch Überbevorratung auf der Baustelle vermeiden.

> Ein automatisiertes Mengenabnahmeverfahren (QTO) ist unerlässlich, um genaue Berechnungen und Schätzungen zu erstellen, menschliche Fehler bei Mengenangaben zu reduzieren und eine Über- oder Unterbestellung von Materialien zu vermeiden.

Als Beispiel für den QTO-Prozess betrachten wir einen häufigen Fall, in dem es notwendig ist, aus der CAD-Datenbank eine Tabellenspezifikation der Volumina nach Elementtypen für eine bestimmte Kategorie oder Klasse von Elementen anzuzeigen. Wir gruppieren alle Projektelemente nach Typ aus der CAD-Projektwandkategorie und fassen die Volumenattribute für jeden Typ zusammen, um das Ergebnis als QTO-Volumentabelle zu präsentieren (Abb. 5.2-3).

Im Beispiel eines typischen CAD-Projekts (Abb. 5.2-3) sind alle Elemente der Wandkategorie in der CAD-Datenbank nach Wandtypen gruppiert, z.B. "Lamelle 11,5", "MW 11,5" und "STB 20,0", und haben klar definierte Volumenattribute, die in metrischen Würfeln dargestellt werden.

Das Ziel des Managers, der an der Schnittstelle zwischen Konstrukteuren und Berechnungsspezialisten steht, ist es, eine automatisierte Tabelle der Volumina nach Elementtyp in der ausgewählten Kategorie zu erhalten. Nicht nur für ein bestimmtes Projekt, sondern auch in einer universellen Form, die auf andere Projekte mit einer ähnlichen Modellstruktur anwendbar ist. Dadurch ist der Ansatz skalierbar und die Daten können ohne doppelten Aufwand wiederverwendet werden.

> Vorbei sind die Zeiten, in denen erfahrene Planer und Kalkulatoren sich mit einem Lineal bewaffneten und sorgfältig jede Linie auf Papier oder PDF -Plänen abmaßen - eine Tradition, die sich in den letzten Jahrtausenden nicht geändert hat. Mit der Entwicklung der 3D -Modellierung, bei der die Geometrie jedes Elements jetzt direkt mit automatisch berechneten volumetrischen Attributen verknüpft ist, ist der Prozess der Bestimmung von Volumen und QTO -Mengen automatisiert worden.

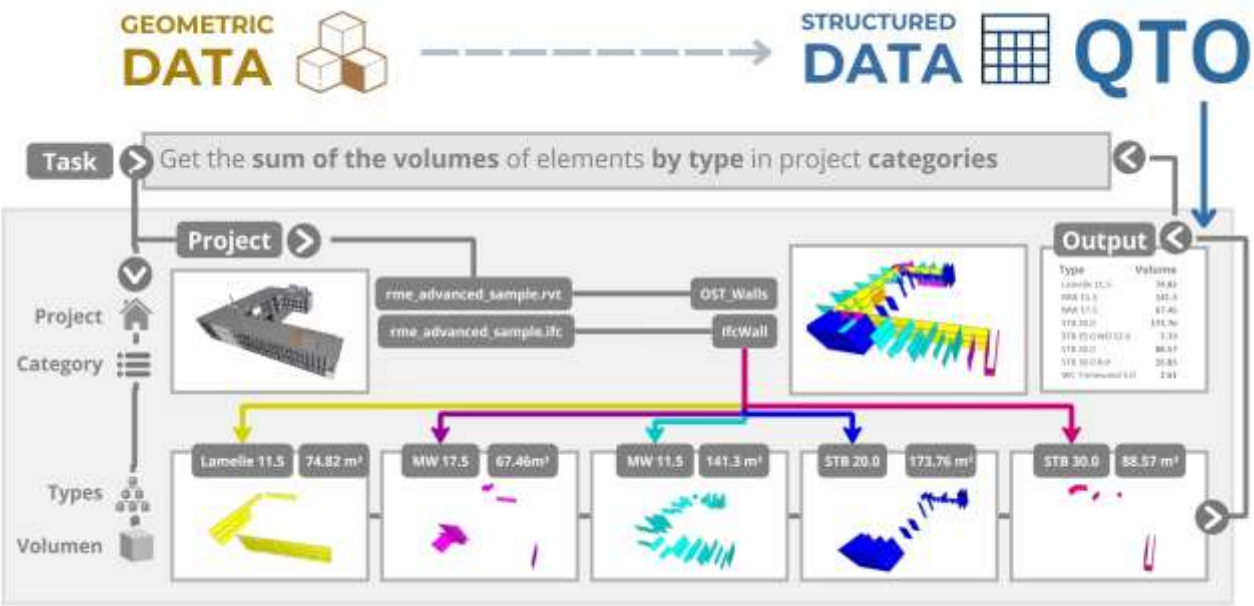


Abbildung 5.2-3 Um QTO Volumen- und Mengenattribute aus einem Projekt zu erhalten, müssen Projektelemente gruppiert und gefiltert werden.

In unserem Beispiel besteht die Aufgabe darin, "eine Kategorie von Wänden in einem Projekt auszuwählen, alle Elemente nach Typ zu gruppieren und Informationen über Volumenattribute in einem strukturierten, tabellarischen Format darzustellen", so dass diese Tabelle von Dutzenden anderer Spezialisten für Berechnungen von Kalkulationen, Logistik, Arbeitsplänen und anderen Geschäftsfällen verwendet werden kann (Abb. 6.1-3).

Aufgrund des geschlossenen Charakters der CAD-Daten kann heute nicht jeder Fachmann den direkten Zugriff auf die CAD-Datenbank nutzen (die Gründe und Lösungen für das Zugriffsproblem werden im sechsten Teil des Buches ausführlich erläutert). Daher müssen viele auf spezialisierte BIM-Werkzeuge zurückgreifen, die auf den Konzepten von Open BIM und Closed BIM [63] basieren. Bei der Arbeit mit spezialisierten BIM-Tools oder direkt in der CAD-Softwareumgebung kann die Tabelle mit den QTO-Ergebnissen (Quantity Take-Off) auf unterschiedliche Weise erstellt werden - je nachdem, ob eine manuelle Schnittstelle oder eine Softwareautomatisierung verwendet wird.

Bei der Verwendung der Benutzeroberfläche einer CAD-(BIM-)Software reichen beispielsweise etwa 17 Aktionen (Schaltflächenklicks) aus, um eine fertige Tabelle mit den Volumen zu erhalten (Abb. 5.2-4). Allerdings muss der Benutzer die Modellstruktur und die Funktionen der CAD-(BIM)-Software gut verstehen.

Wenn die Automatisierung über Programmcode oder über Plug-ins und API-Tools innerhalb von CAD-Programmen erfolgt, verringert sich die Anzahl der manuellen Schritte zur Ermittlung der Volumentabellen, aber es müssen je nach Bibliothek oder Tool 40 bis 150 Zeilen Code geschrieben werden:

- **IfcOpSh (offenes BIM**) oder **Dynamo IronPython (geschlossenes BIM)** - ermöglichen es Ihnen, eine QTO-Tabelle aus einem CAD-Format oder CAD-Programm in nur ~40 Zeilen Code zu erhalten.
- **IFC_js (open BIM**) - erfordert etwa 150 Zeilen Code, um umfangreiche Attribute aus dem IFC-Modell zu extrahieren.

- **Schnittstelle CAD-Tools (BIM**) - ermöglicht es Ihnen, das gleiche Ergebnis manuell, in 17 Mausklicks zu erhalten.

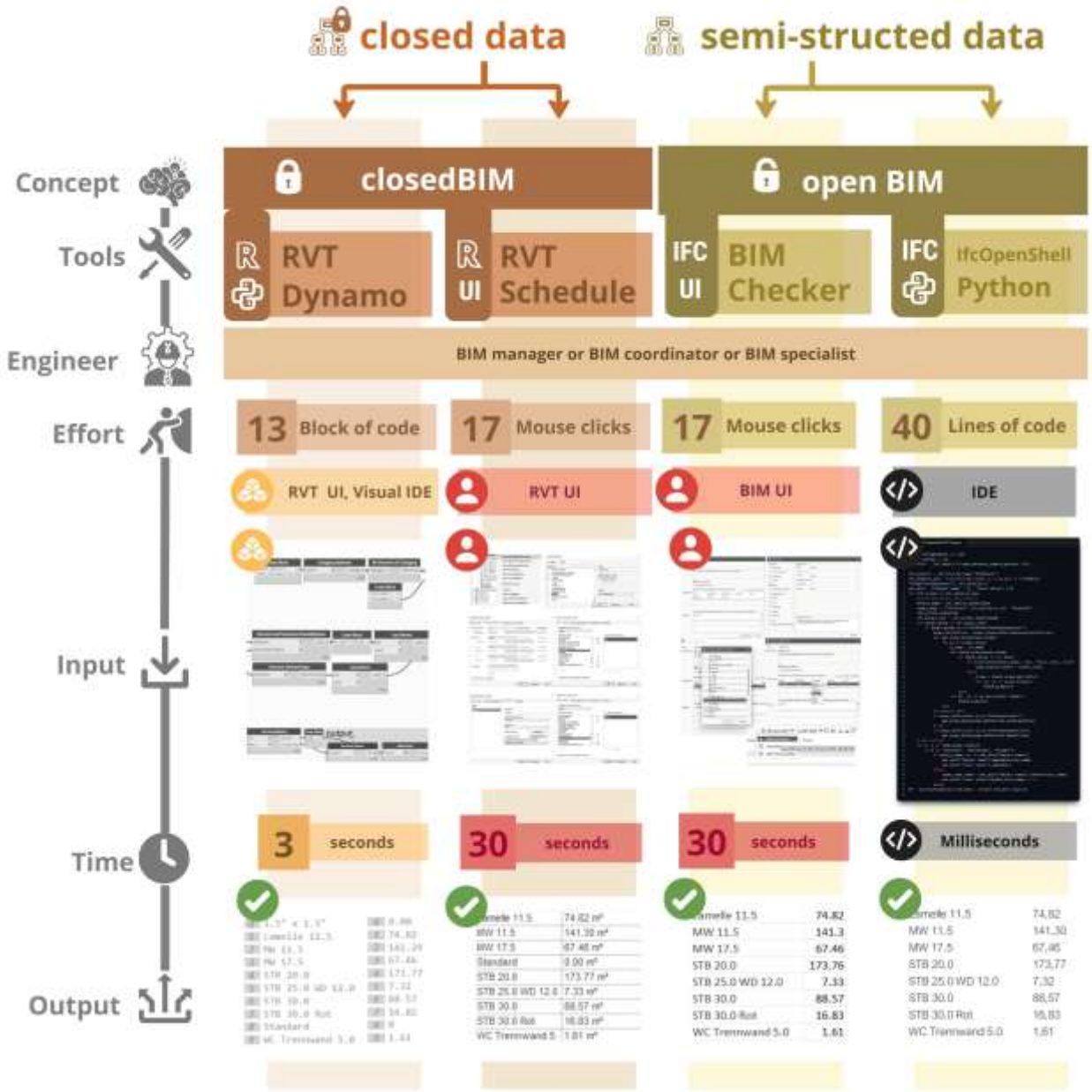


Abbildung 5.2-4 CAD (BIM) Designer und Manager verwenden 40 bis 150 Codezeilen oder ein Dutzend Tastenanschläge, um QTO Tabellen zu erstellen.

Das Ergebnis ist das gleiche - eine strukturierte Tabelle mit Volumenattributen für eine Gruppe von Elementen. Der einzige Unterschied ist der Arbeitsaufwand und das erforderliche technische Fachwissen des Benutzers (Abb. 5.2-4). Moderne Werkzeuge beschleunigen den QTO-Prozess im Vergleich zur manuellen Erfassung von Mengen erheblich und verringern die Fehlerwahrscheinlichkeit. Sie ermöglichen es, Daten direkt aus dem Projektmodell zu extrahieren, wodurch die Notwendigkeit entfällt, Volumina manuell aus

Zeichnungen neu zu berechnen, wie es in der Vergangenheit der Fall war.

Unabhängig von der verwendeten Methode - ob Open BIM oder Closed BIM - ist es möglich, eine identische QTO - Tabelle mit Projektelementvolumen zu erhalten (Abb. 5.2-4). Bei der Arbeit mit Projektdaten in CAD - (BIM-) Konzepten sind die Anwender jedoch auf spezialisierte Werkzeuge und APIs von Anbietern angewiesen (Abb. 3.2-13). Dies schafft zusätzliche Abhängigkeiten und erfordert das Erlernen einzigartiger Datenschemata, während der direkte Zugriff auf die Daten eingeschränkt wird.

Aufgrund des geschlossenen Charakters von CAD-Daten erschwert die Beschaffung von QTO-Tabellen und anderen Parametern die Automatisierung von Berechnungen und die Integration mit externen Systemen. Durch die Verwendung von Werkzeugen für den direkten Zugriff auf Datenbanken und die Übertragung von CAD -Daten des Projekts mit Hilfe von Reverse-Engineering-Werkzeugen in ein offenes strukturiertes Datenrahmenformat (Abb. 4.1-13) kann eine identische QTO -Tabelle mit nur einer Zeile Code erhalten werden (Abb. 5.2-5 - Variante mit granularen Daten).

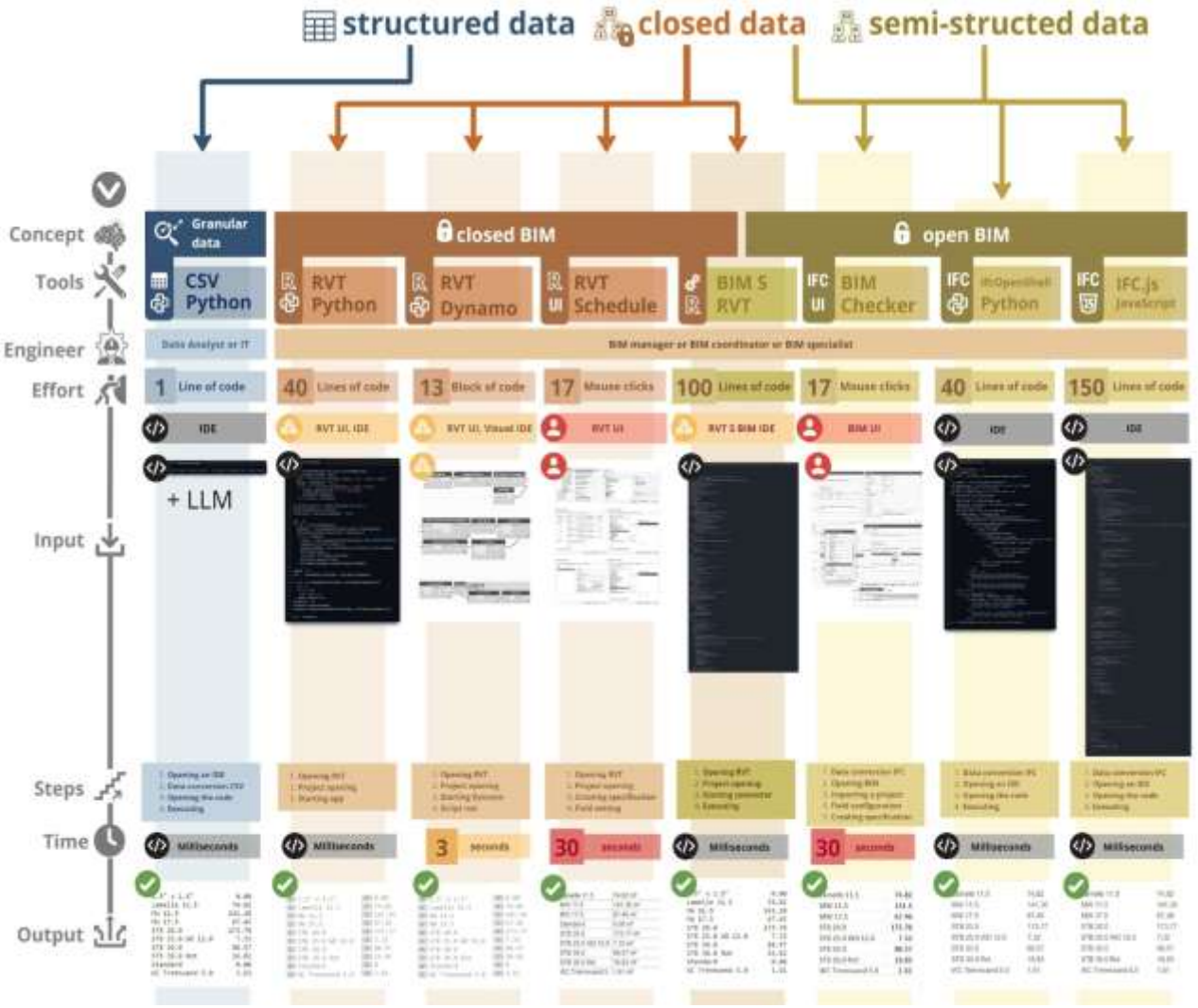


Abbildung 5.2-5 Unterschiedliche Werkzeuge erzeugen die gleichen Ergebnisse in Form von Attributtabellen der Projekteinheiten, jedoch mit unterschiedlichem Arbeitsaufwand.

Bei der Verwendung offener strukturierter Daten aus CAD Projekten, wie im Kapitel "Umwandlung von CAD (BIM) Daten in strukturierte Form" erwähnt, wird der Gruppierungsprozess, QTO, stark vereinfacht.

Ansätze, die auf der Verwendung offener strukturierter Daten oder dem direkten Zugriff auf CAD-Modelldatenbanken beruhen, sind frei von den Marketing-Zwängen, die mit dem Akronym BIM verbunden sind. Sie stützen sich auf bewährte Werkzeuge, die in anderen Branchen seit langem eingesetzt werden (Abb. 7.3-10 ETL-Prozess).

> Laut der McKinsey-Studie "Open Data: Unleash Innovation and Productivity with Streaming Information" [102], die 2013 durchgeführt wurde, kann die Nutzung offener Daten Einsparungen in Höhe von 30 bis 50 Milliarden Dollar pro Jahr in den Bereichen Design, Engineering, Beschaffung und Bau von Energieerzeugungsanlagen ermöglichen. Dies entspricht einer Einsparung von 15 Prozent der Baukapitalkosten.

Die Arbeit mit offen strukturierten (granularen) Daten vereinfacht die Informationsbeschaffung und -verarbeitung, verringert die Abhängigkeit von spezialisierten BIM-Plattformen und öffnet den Weg zur Automatisierung, ohne dass proprietäre Systeme oder parametrische und komplexe Datenmodelle aus CAD-Formaten verwendet werden müssen.

QTO-Automatisierung mit LLM und strukturierten Daten

Die Übersetzung unstrukturierter Daten in eine strukturierte Form verbessert die Effizienz verschiedener Prozesse erheblich: Sie vereinfacht die Datenverarbeitung (Abb. 4.1-1, Abb. 4.1-2) und beschleunigt den Validierungsprozess, indem sie die Anforderungen klar und transparent macht, wie wir bereits in den vorangegangenen Kapiteln diskutiert haben. In ähnlicher Weise erleichtert die Übersetzung von CAD-Daten (BIM) in eine strukturierte offene Form (Abb. 4.1-12, Abb. 4.1-13) den Prozess der Attributgruppierung und den QTO-Prozess.

Die Attributtabelle QTO hat eine strukturierte Form, so dass wir bei der Verwendung von strukturierten CAD-Daten mit einem einzigen Datenmodell arbeiten (Abb. 5.2-5), wodurch die Notwendigkeit entfällt, Projektdatenmodelle und Gruppierungsregeln zu konvertieren und auf einen gemeinsamen Nenner zu bringen. So können wir Daten mit nur einer Zeile Code nach einem oder mehreren Attributen gruppieren. Im Gegensatz dazu erfordert die Verarbeitung in Open BIM und Closed BIM, wo Daten in halbstrukturierten, parametrischen oder geschlossenen Formaten gespeichert sind, Dutzende oder sogar Hunderte von Codezeilen und die Verwendung der API zur Interaktion mit Geometrie- und Attributinformationen.

- Beispiel für die Gruppierung eines strukturierten QTO-Projekts nach einem Attribut. Textabfrage in einem LLM Chatroom (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN oder einem anderen):

> Ich habe CAD -Projekt als DataFrame - bitte filtern Sie die Projektdaten, um Elemente mit "Typ" Parameter enthält nur "Typ 1 Wert zu erhalten. ⏎

- Die Antwort von LLM wird sehr wahrscheinlich in Form von Python-Code unter Verwendung von Pandas erfolgen:

Abb. 5.2-6 Eine Zeile Code, die mit LLM geschrieben wurde, ermöglicht es Ihnen, ein gesamtes CAD Projekt nach dem Attribut "Typ" zu gruppieren und die gewünschte Gruppe von Elementen zu erhalten.

Dank der einfachen Struktur des zweidimensionalen DataFrame müssen wir das LLM Schema und das Datenmodell nicht erklären, was die Interpretationsschritte verkürzt und die Erstellung der endgültigen Lösungen beschleunigt. Früher erforderte das Schreiben von selbst einfachem Code das Erlernen von Programmiersprachen, aber jetzt ermöglichen uns moderne Sprachmodelle (LLMs) die automatische Übersetzung von Prozesslogik in Code, wenn wir mit strukturierten Daten unter Verwendung von Textabfragen arbeiten.

> LLM Automatisierung und Sprachmodelle können Fachleute, die mit CAD (BIM) Datengruppierung und -verarbeitung arbeiten, von der Notwendigkeit befreien, Programmiersprachen oder BIM-Tools zu erlernen, indem sie die Möglichkeit bieten, Probleme durch Textabfragen zu lösen.

Dieselbe Abfrage - Gruppierung aller Projektelemente aus der Kategorie "Wände" und Berechnung der Volumina für jeden Typ (Abb. 5.2-5) - die in einer CAD-Umgebung (BIM) 17 Klicks in der Schnittstelle oder das Schreiben von 40 Codezeilen erfordert, sieht in offenen Datenverarbeitungswerkzeugen (z. B. SQL oder Pandas) wie eine einfache und intuitive Abfrage aus:

- Mit einer einzigen Zeile in Pandas:

```
df[df['Kategorie'].isin(['OST_Wände'])].groupby('Typ')['Volumen'].sum()
```

Code-Entschlüsselung: Entnehmen Sie aus df (DataFrame) die Elemente, deren Attribut-Spalte "Kategorie" den Wert "OST_Walls" hat, gruppieren Sie alle erhaltenen Elemente nach der Attribut-Spalte "Typ" und summieren Sie für die erhaltene Gruppe von Elementen das Attribut "Volumen".

- Gruppierung eines strukturierten Projekts aus dem CAD mit Hilfe von SQL:

```
SELECT Typ, SUM(Volumen) AS GesamtVolumen
FROM Elemente
WHERE Kategorie = 'OST_Wände'
GROUP BY Typ;
```

- Mit Hilfe von LLM können wir eine Gruppierungsanfrage an die Projektdatenbank als einfache Textreferenz schreiben - eine Eingabeaufforderung (Abb. 5.2-7):

> Für den Projektdatenrahmen gruppieren Sie die Elemente nach dem Parameter "Typ", aber nur für Elemente, bei denen der Parameter "Kategorie" gleich "OST_Walls" oder "OST_Columns" ist, und fassen Sie bitte den Spaltenparameter "Volumen" für die resultierende zusammen ⏎

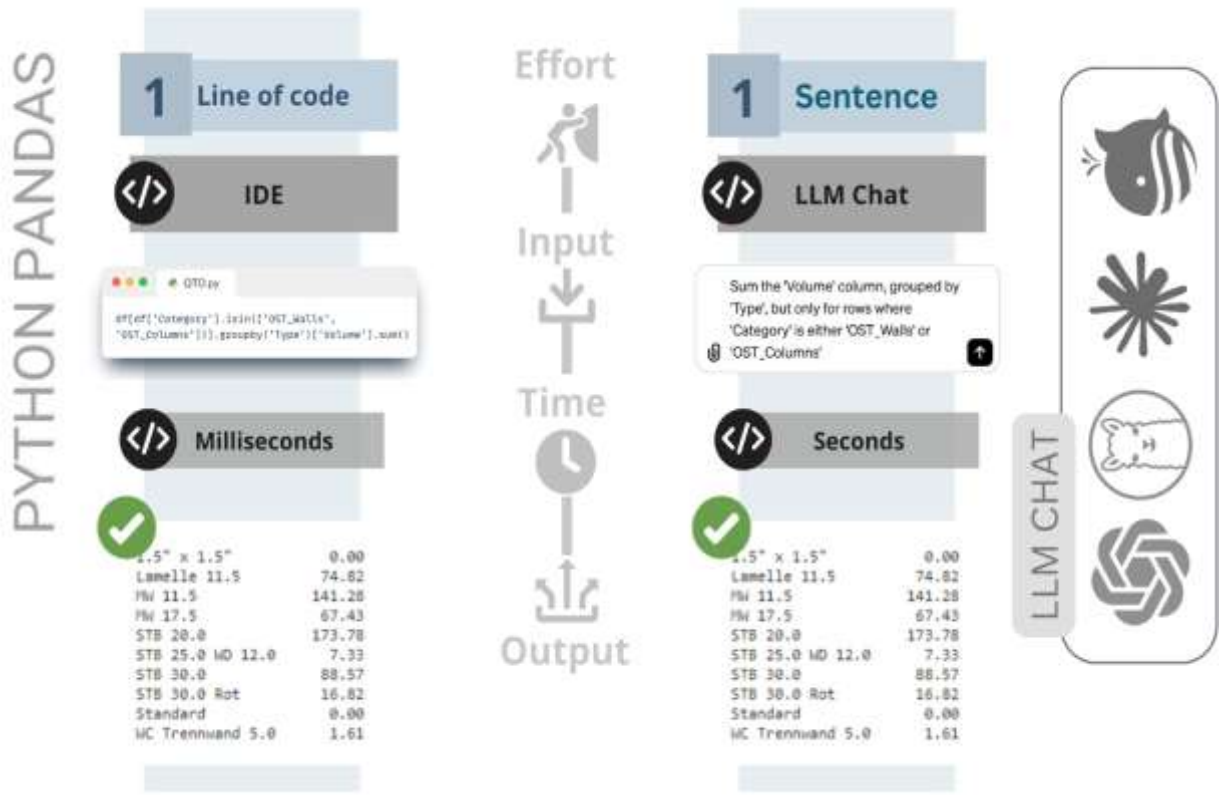


Abbildung 5.2-7 Mit SQL, Pandas und LLM lässt sich die Datenverarbeitung nun mit wenigen Zeilen Code und Textabfragen automatisieren.

Die Gewinnung von QTO aus CAD-Daten unter Verwendung von LLM-Tools (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) verändert die herkömmlichen Methoden zur Extraktion von Attributinformationen, quantitativen und volumetrischen Daten für einzelne Objekte und Objektgruppen grundlegend.

> Jetzt können selbst Projektmanager, Kalkulations- oder Logistikspezialisten, die keine tiefgreifenden Kenntnisse in der Konstruktion haben und keine spezialisierte CAD-Software - (BIM-) besitzen, mit Zugang zur CAD-Datenbank das Gesamtvolumen der Elemente der Kategorie Wände oder anderer Objekte in Sekundenschnelle ermitteln, indem sie einfach eine Abfrage schreiben oder diktieren.

Bei Textabfragen (Abb. 5.2-8) verarbeitet der LLM-Agent des Modells die Anfrage des Benutzers, eine bestimmte Funktion auf einen oder mehrere Parameter - Spalten der Tabelle - anzuwenden. Als Ergebnis erhält der Benutzer in der Kommunikation mit dem LLM entweder einen neuen Spaltenparameter mit neuen Werten oder einen bestimmten Wert nach Gruppierung

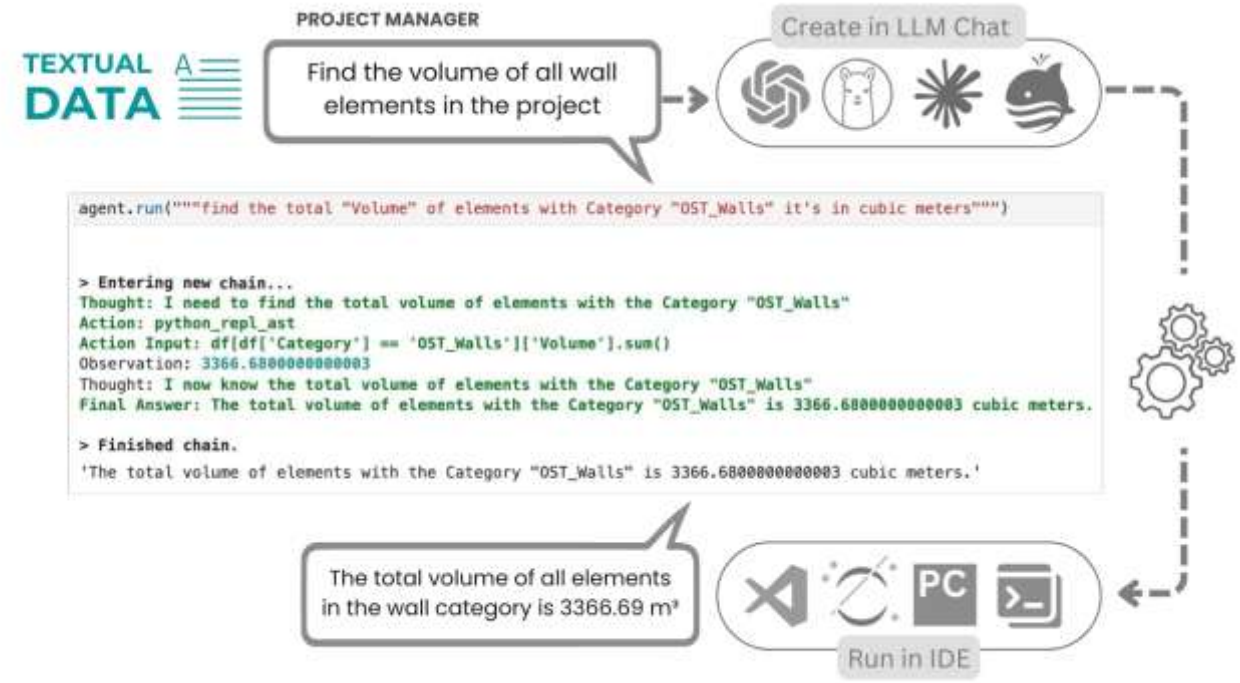


Abb. 5.2-8 LLM Modell, das mit strukturierten Daten arbeitet, versteht aus dem Kontext einer Textabfrage, nach welchen Gruppierungen und Attributen der Benutzer fragt.

Wenn es notwendig ist, Mengen für nur eine Gruppe von Elementen zu erhalten, ist es ausreichend, eine einfache QTO-Abfrage (Abb. 5.2-7) auf den CAD-Modelldaten durchzuführen. Bei der Berechnung eines Budgets oder eines Kostenvoranschlags für ein ganzes Projekt, das aus vielen Gruppen von Elementen besteht, ist es jedoch oft notwendig, Mengen für alle Arten von Elementen (Klassen0) zu extrahieren, wobei jede Kategorie von Elementen separat verarbeitet wird - mit Gruppierung nach relevanten Attributen.

In der Praxis von Schätzern und Gutachtern werden individuelle Gruppierungs- und Berechnungsregeln für verschiedene Arten von Objekten verwendet. So werden z.B. Fenster in der Regel nach Stockwerken oder Zonen gruppiert (Gruppierungsparameter - Attribut Ebene, Räume) und Wände - nach Material oder Bauart

(Parameter Material, Typ). Um den Gruppierungsprozess zu automatisieren, werden solche Regeln im Voraus in Form von Gruppierungsregeltabellen beschrieben. Diese Tabellen dienen als Konfigurationsvorlagen, die festlegen, welche Attribute in den Berechnungen für jede Gruppe von Elementen im Projekt verwendet werden sollen.

QTO Berechnung des gesamten Projekts mit Hilfe von Gruppenregeln aus einem Excel-Tabellenblatt

In realen Bauprojekten ist es oft notwendig, innerhalb einer Gruppe von Elementen gleichzeitig eine Aggregation nach mehreren Attributen durchzuführen. Wenn man beispielsweise mit der Kategorie "Fenster" arbeitet (wobei das Attribut Kategorie Werte wie OST_Windows oder IfcWindows enthält), können die Elemente nicht nur nach dem Typ gruppiert werden - z. B. nach dem Wert im Feld Typname oder Typ - sondern auch nach zusätzlichen Merkmalen, wie z. B. dem im entsprechenden Attribut angegebenen Wärmeleitfähigkeitsniveau. Diese mehrdimensionale Gruppierung ermöglicht genauere Ergebnisse für eine bestimmte Gruppe. In ähnlicher Weise können bei der Berechnung von Wand- oder Bodenkategorien beliebige Kombinationen von Attributen - wie Material, Ebene, Boden, Feuerwiderstand und andere Parameter - als Filter oder Gruppierungskriterien verwendet werden (Abb. 5.2-9).

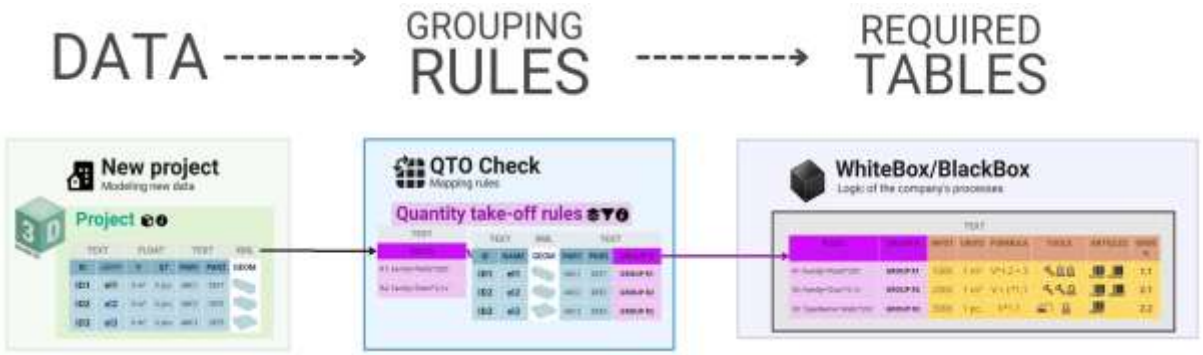


Abbildung 5.2-9 Für jede Gruppe oder Kategorie von Einheiten in einem Projekt gibt es eine andere Gruppierungsformel, die aus einem oder mehreren Kriterien besteht.

Der Prozess der Definition solcher Gruppierungsregeln ähnelt dem Prozess der Erstellung von Datenanforderungen, der im Kapitel "Erstellen von Anforderungen und Qualitätsprüfung data " (Abb. 4.4-5) beschrieben wurde, wo wir die Arbeit mit Datenmodellen im Detail besprochen haben. Solche Gruppierungs- und Berechnungsregeln stellen die Genauigkeit und Relevanz der Ergebnisse sicher, um die Gesamtattribute der Menge oder des Volumens einer Kategorie von Entitäten automatisch zu berechnen, wobei alle notwendigen Bedingungen berücksichtigt werden, die in Berechnungen und Kalkulationen berücksichtigt werden müssen.

- Das folgende Codebeispiel filtert die Projekttabelle so, dass der resultierende Datensatz nur Entitäten enthält, bei denen die Attributspalte "Kategorie" die Werte "OST_Windows" oder "IfcWindows" und gleichzeitig die Attributspalte "Typ" den Wert "Typ 1" enthält:

> Ich habe ein DataFrame Projekt - filtern Sie die Daten so, dass nur Elemente im Datensatz verbleiben, deren Attribut "Kategorie" die Werte "OST_Windows" oder "IfcWindows" enthält und gleichzeitig das Attribut "Typ" den Wert "Typ 1 enthält ⏎

Die Antwort von LLM:

Abbildung 5.2-10 Mit einer einzigen Codezeile, ähnlich der Excel-Formel, können Sie alle Projektelemente nach mehreren Attributen gruppieren.

Der resultierende Code (Abb. 5.2-10) nach der Übersetzung der CAD-Daten in strukturierte offene Formate (Abb. 4.1-13) kann in einer der beliebten IDEs (integrierten Entwicklungsumgebungen), die wir oben im Offline-Modus erwähnt haben, ausgeführt werden: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse mit PyDev-Plugin, Thonny, Wing IDE, IntelliJ IDEA mit Python-Plugin, JupyterLab oder beliebte Online-Tools: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Um die Projekteinheiten im QTO DataFrame-Formular unter der Kategorie "Windows" nur mit einem bestimmten Wärmeleitfähigkeitswert abzurufen, können wir die folgende Abfrage an das LLM verwenden:

> Ich habe ein DataFrame Projekt - die Daten sollen so gefiltert werden, dass nur Datensätze, deren "Kategorie" die Werte "OST_Windows" oder "IfcWindows" enthält, im Datensatz verbleiben, und gleichzeitig soll die Spalte "ThermalConductivity" einen Wert von 0, haben ⏎

- Die Antwort von LLM:

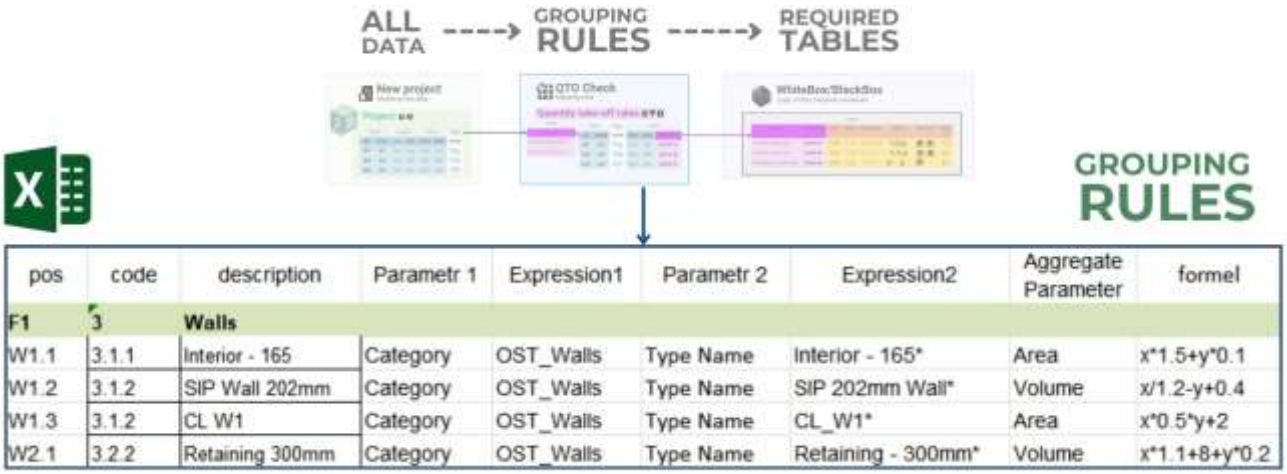
Abbildung 5.2-11 Die extrem einfache Abfragesprache Pandas Python ermöglicht es Ihnen, QTOs für eine beliebige Anzahl von Projekten gleichzeitig auszuführen.

In der vom LLM erhaltenen Antwort (Abb. 5.2-11) wird die logische Bedingung "&" verwendet, um zwei Kriterien zu kombinieren: den Wert der Wärmeleitfähigkeit und die Zugehörigkeit zu einer der beiden Kategorien. Die Methode "isin" prüft, ob der Wert der Attributspalte "Kategorie" in der vorgegebenen Liste enthalten ist.

Bei Projekten mit einer großen Anzahl von Elementgruppen mit unterschiedlicher Gruppierungslogik - für jede Kategorie von Projektelementen (z.B.: Fenster, Türen, Decken) sollten individuelle Gruppierungsregeln aufgestellt werden, die zusätzliche Koeffizienten oder Formeln zur Berechnung der Gesamtattribute enthalten können. Diese Formeln (Abb. 5.2-12 Attribut "formel", z.B. x-Wert der Menge und y-Volumen der Gruppe) und Koeffizienten berücksichtigen z.B. die einzigartigen Eigenschaften jeder Gruppe:

- %-Zuschläge zum Materialvolumen zur Berücksichtigung von Überschreitungen
- Festgelegte zusätzliche Menge an Material
- Anpassungen in Bezug auf mögliche Risiken und Berechnungsfehler in Form von Formeln

Sobald die Filter- und Gruppierungsregeln in Form von Parameterformeln für jede Positionskategorie formuliert wurden, können sie als zeilenweise Tabelle gespeichert werden - zum Beispiel im Excel-Format (Abb. 5.2-12). Durch die Speicherung dieser Regeln in strukturierter Form kann der Prozess der Extraktion, Filterung und Gruppierung von Projektdaten vollständig automatisiert werden. Anstatt viele einzelne Abfragen manuell zu schreiben, liest das System einfach die Parametertabelle und wendet die entsprechenden Regeln auf das Modell (den gesamten Projektdatenrahmen (Abb. 4.1-13)) an, wodurch endgültige QTO-Tabellen für jede Kategorie von Projektelementen erzeugt werden.



pos	code	description	Parametr 1	Expression1	Parametr 2	Expression2	Aggregate Parameter	formel
F1	3	Walls						
W1.1	3.1.1	Interior - 165	Category	OST_Walls	Type Name	Interior - 165*	Area	x*1.5+y*0.1
W1.2	3.1.2	SIP Wall 202mm	Category	OST_Walls	Type Name	SIP 202mm Wall*	Volume	x/1.2-y+0.4
W1.3	3.1.2	CL W1	Category	OST_Walls	Type Name	CL_W1*	Area	x*0.5*y+2
W2.1	3.2.2	Retaining 300mm	Category	OST_Walls	Type Name	Retaining - 300mm*	Volume	x*1.1+8+y*0.2

Abbildung 5.2-12 QTO-Attributgruppierung Tabelle legt Regeln für die Gruppierung von Projektelementen fest, die eine genaue Gesamtzahl und einen genauen Umfang für jede Kategorie liefern.

Die gesammelten Regeln ermöglichen es Ihnen, das gesamte Projekt zu gruppieren und alle notwendigen Berechnungen durchzuführen, einschließlich Anpassungen der Volumenattribute. Als Ergebnis werden die Volumina auf das "echte Volumen" gebracht, das für Berechnungen und Kalkulationen verwendet wird, und nicht die, die ursprünglich in der Entwurfsphase im CAD-Modell waren.

> Bei der automatischen Erstellung von QTO Volumentabellen für das gesamte Projekt muss die Anwendung alle Kategorien der Tabelle mit den Gruppierungsregeln durchgehen, die Gruppierungsattribute nehmen, alle Projektelemente nach ihnen gruppieren und das Volumenattribut für diese Gruppe aggregieren, wobei es zusätzlich mit einem Verfeinerungsfaktor oder -koeffizienten multipliziert wird.

Bitten wir LLM, für uns einen Code für eine solche Lösung zu schreiben, wobei der Code zwei Tabellen laden müsste - eine Tabelle mit Gruppierungsregeln (Abb. 5.2-12) und eine Tabelle mit Daten aus dem Projekt selbst (Abb. 4.1-13), und dann die Gruppierungsregeln anwenden, die Elemente nach den gegebenen Regeln gruppieren, die aggregierten Werte berechnen und die Ergebnisse in einer neuen Excel-Datei speichern.

- Senden Sie eine Textanfrage an den LLM-Chat:

> Ich benötige Code, um Projektdaten aus der Datei "basic_sample_project.xlsx" zu lesen und dann Regeln aus "Grouping_rules_QTO.xlsx", mit denen alle Daten nach "Parameter 1" und "Parameter 2" gruppiert, "Aggregate Parameter" aggregiert, nach "Expression2" gefiltert, Berechnungen aus "Formel1" durchgeführt und die QTO-Tabelle in "QTQ_table2.xlsx gespeichert werden⏎

Die Antwort von LLM:

Create in LLM Chat

Validation.py

```python
import pandas as pd

# Reading CSV and Excel files with project data and grouping rules respectively
df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])

# Cleaning numerical columns in the dataframe by removing non-numeric characters
numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
df[numeric_cols] = df[numeric_cols
].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+.\d+)')[0].fillna(0)))

# Define a function to process each row in the grouping rules dataframe
def process_row(row):
    # Group by specified parameters and calculate sum and count for the aggregate parameter
    group = df.groupby([row["Parameter 1"], row["Parameter 2"]
    ])[row["Aggregate Parameter"]].agg(['sum', 'count'])
    group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]

    # Perform calculation defined in the rules
    group['After Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
    return group.reset_index()

# Apply the function to each row in the grouping rules dataframe and concatenate the results
dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)

# Save the processed data to an Excel file using the OpenPyXL engine
dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)
```

Run in IDE

Abbildung 5.2-13 Python-Code Pandas verarbeitet Konstruktionsdaten aus einem CAD-Projekt und gruppiert sie nach in Excel definierten Regeln.

Das Endergebnis der Codeausführung (Abb. 5.2-13) wird eine Entity-Group-Tabelle sein, die nicht nur die generalisierten Volumenattribute aus dem ursprünglichen CAD- (BIM-) Modell enthält, sondern auch ein neues reales Volumenattribut, das alle Anforderungen für die korrekte Erstellung von Berechnungen und Schätzungen berücksichtigt (Beispiel Abb. 5.2-14).

QTO TABLE FOR ALL GROUPS OF ELEMENTS
ASSEMBLED BY RULES FROM EXCEL TABLE

	Category_x	Type Name	Volumetric parameter	Amount	Id´s	After Calculation	Measure
0	OST_Walls	Interior - 165 Partitior	17	3	424922, 425745,	25.8	m2
1	OST_Walls	SIP 202mm Wall - con	21.95	4	198694, 428588,	13.9	m3
2	OST_Walls	CL_W1	230	6	493612, 493697,	692	m2
3	OST_Walls	Retaining - 300mm Cc	57.93	10	599841, 599906,	72.7	m3

Abbildung 5.2-14 Das Attribut "Nach Berechnung" wird der Übersichtstabelle hinzugefügt, nachdem ein Code ausgeführt wurde, der automatisch das tatsächliche Volumen berechnet.

Der resultierende Code (Abb. 5.2-13) kann in einer der beliebten IDEs (die wir oben erwähnt haben) ausgeführt und auf eine beliebige Anzahl bestehender oder neu eingehender Projekte (RVT, IFC, DWG, NWS, DGN usw.) angewendet werden, seien es einige wenige Projekte oder vielleicht Hunderte von Projekten in verschiedenen Formaten in strukturierter Form (Abb. 5.2-15).

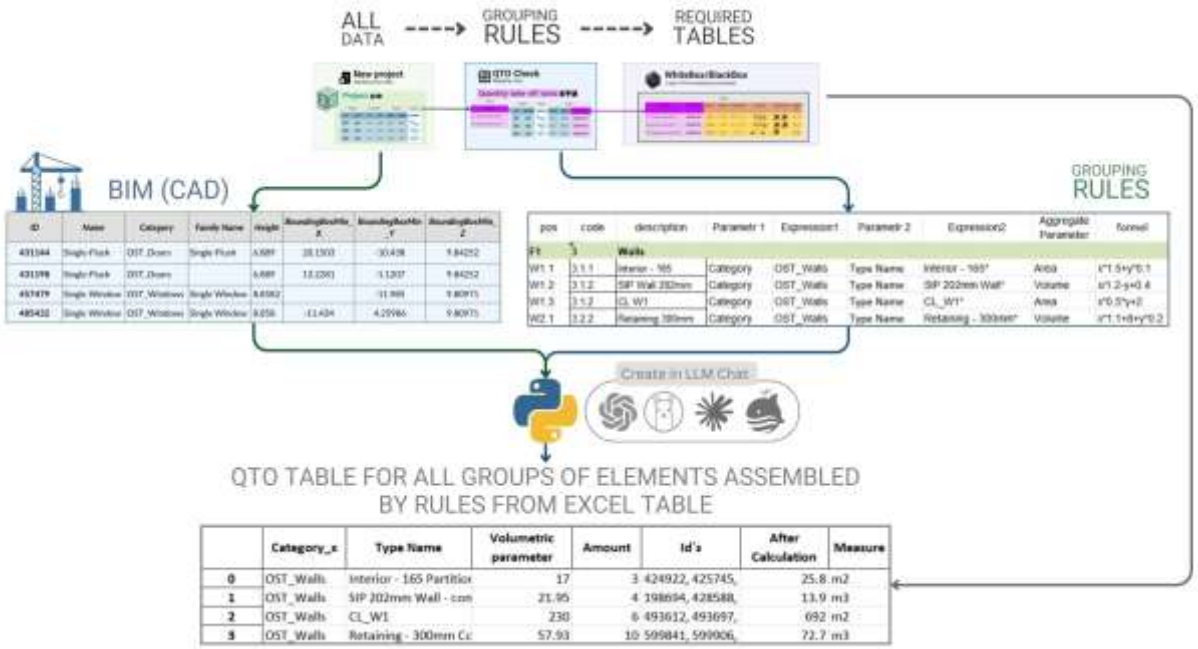


Abbildung 5.2-15 Der automatische Konstruktionsdaten-Gruppierungsprozess verknüpft BIM-Daten (CAD) mit QTO-Tabellen über Regeln aus einer Excel-Tabelle.

Der maßgeschneiderte und parametrisierte Prozess der volumetrischen Datenerfassung (Abb. 5.2-15) ermöglicht die vollautomatische Erfassung von Daten zu quantitativen Attributen und Volumina von Projektelementen für die weitere Arbeit mit ihnen, einschließlich Kostenabschätzung, Logistik, Arbeitspläne und Berechnung des CO2-Fußabdrucks und anderer analytischer Aufgaben.

Nachdem wir uns mit Werkzeugen vertraut gemacht haben, die es uns ermöglichen, Projektelemente nach bestimmten Attributen zu organisieren und zu gruppieren, sind wir nun in der Lage, gruppierte und gefilterte Projekte in verschiedene Unternehmensberechnungen und Geschäftsszenarien zu integrieren.

KAPITEL 5.3.
4D, 6D -8D UND BERECHNUNG DER KOHLENDIOXIDEMISSIONEN CO_2

4D-Modell: Integration der Zeit in die Baukostenschätzung

Neben der Kostenermittlung ist eine der wichtigsten Anwendungen von Planungsdaten im Bauwesen die Ermittlung von Zeitparametern, sowohl für einzelne Baumaßnahmen als auch für das gesamte Projekt. Die im vorangegangenen Kapitel "Kalkulationen und Schätzungen für Bauprojekte" ausführlich besprochene Methode der ressourcenbasierten Schätzung und die dazugehörige Kalkulationsdatenbank werden häufig als Grundlage für eine automatisierte Zeitberechnung und die Erstellung eines Zeitplans für die Ausführung der Arbeiten verwendet.

Beim ressourcenbasierten Ansatz werden nicht nur die Materialkosten, sondern auch die Zeitressourcen berücksichtigt. In der Kalkulation kann jedem Vorgang ein Arbeitsauftragsattribut (Abb. 5.3-1 - Arbeitsauftragsparameter) und der mit der Ausführung dieses Vorgangs verbundene Zeit- und Kostenaufwand zugeordnet werden. Diese Parameter sind besonders wichtig für die Beschreibung von Vorgängen, die keinen festen Marktpreis haben und nicht direkt käuflich sind - wie der Einsatz von Baumaschinen, die Beschäftigung von Arbeitskräften oder logistische Prozesse (die normalerweise in Stunden ausgedrückt werden). In solchen Fällen werden die Kosten nicht von der Beschaffungsabteilung, sondern direkt vom ausführenden Unternehmen auf der Grundlage interner Normen oder Produktionsraten ermittelt (Abb. 5.3-1).

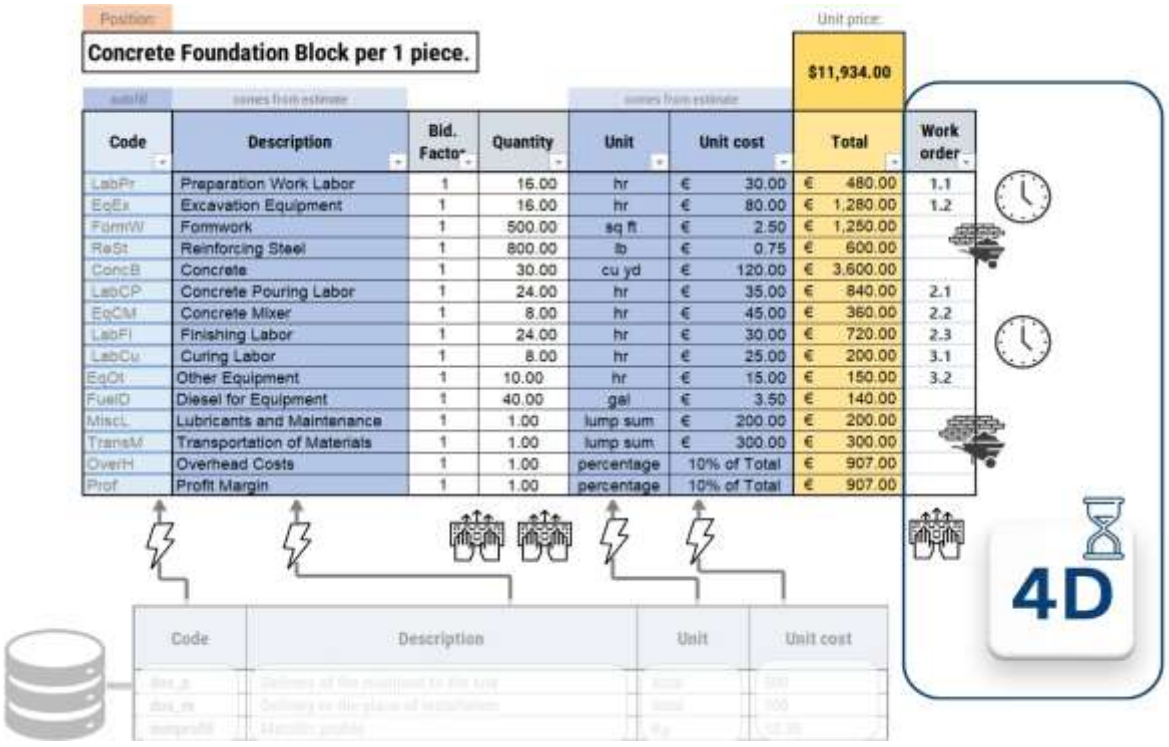


Abb. 5.3-1 Die Berechnung der Arbeit bei der ressourcenbasierten Schätzungsmethode beinhaltet die zeitliche Planung der Arbeitsstunden.

Die Berechnungen auf der Kalkulationsebene umfassen also nicht nur die Kraftstoff- und Materialkosten

(Einkaufskosten), sondern auch die von Fahrern, Technikern und Hilfskräften auf der Baustelle aufgewendete Zeit. In dem angegebenen Beispiel (Abb. 5.3-1) ist die Kostentabelle eine Kalkulation der Kosten für den Einbau eines Fundamentblocks, einschließlich der einzelnen Arbeitsschritte wie Vorbereitung, Einbau des Rahmens und Betonieren, sowie der erforderlichen Materialien und Arbeitskräfte. Einzelne Arbeitsschritte, wie z. B. die Vorbereitungsarbeiten, können jedoch keine Materialkosten, aber erhebliche vorübergehende Arbeitskosten, ausgedrückt in Arbeitsstunden, beinhalten.

Zur Planung der Arbeitsreihenfolge (für den Arbeitsplan) auf der Baustelle wird das Attribut "Arbeitsauftrag" manuell in die Berechnungstabelle eingefügt (Abb. 5.3-1). Dieses Attribut wird in einer zusätzlichen Spalte nur für Positionen angegeben, deren Maßeinheit in Zeit (Stunde, Tag) ausgedrückt ist. Dieses Attribut wird zusätzlich zu Arbeitscode, Beschreibung, Menge, Mengeneinheit (Parameter "Einheit") und Kosten angegeben. Die numerische Reihenfolge (Parameter "Arbeitsauftrag") der Tätigkeiten ermöglicht es Ihnen, die Reihenfolge festzulegen, in der die Aufgaben auf der Baustelle ausgeführt werden, und sie für die Terminplanung zu verwenden.

Bauzeitenplan und seine Automatisierung auf der Grundlage von Kalkulationsdaten

Der Bauzeitenplan ist eine visuelle Darstellung des Arbeitsplans und der Prozesse, die im Rahmen der Projektdurchführung durchgeführt werden sollen. Er wird auf der Grundlage detaillierter Ressourcenberechnungen erstellt (Abb. 5.3-1), wobei jede Aufgabe zusätzlich zu den Kosten der Ressourcen nach Zeit und Reihenfolge geplant wird.

Im Gegensatz zu Durchschnittsansätzen, bei denen die Zeitschätzungen auf typischen Stunden für die Installation von Materialien oder Ausrüstungen beruhen, basiert die Planung bei der ressourcenbasierten Methode auf den tatsächlichen Daten der Kalkulation. Jeder Posten der Schätzung, der sich auf die Arbeitskosten bezieht, basiert auf dem angewandten Kalender, der die tatsächlichen Bedingungen der Ressourcennutzung während des Arbeitszeitraums berücksichtigt. Die Anpassung der produktiven Stunden durch Koeffizienten auf der Kalkulationsebene (Abb. 5.3-1 Parameter "Bid. Factor") ermöglicht die Berücksichtigung von Produktivitätsunterschieden und saisonalen Besonderheiten, die sich auf den Zeitplan der Arbeit auswirken.

Um die Start- und Endtermine für den Bauzeitplan im Gantt-Diagramm zu bestimmen, nehmen wir die Werte des Zeitattributs für jede Position aus der Fundamentblockkalkulation und multiplizieren sie mit der Anzahl der Blöcke (in diesem Fall die Anzahl der Betonfundamentblöcke). Diese Berechnung ergibt die Dauer der einzelnen Aufgaben. Anschließend werden diese Dauern auf einer Zeitachse, beginnend mit dem Startdatum des Projekts, in einem Diagramm dargestellt. Der Parameter "Arbeitsauftrag" für Prozesse ermöglicht es uns außerdem zu verstehen, ob der Arbeitsprozess parallel ("Arbeitsauftrag" z.B. 1.1-1.1) oder sequentiell (1.1-1.2) abläuft.

> **Das Gantt-Diagramm** ist ein grafisches Werkzeug für Planung und Projektmanagement, das Aufgaben als horizontale Balken auf einer Zeitachse darstellt. Jeder Balken stellt die Dauer der Aufgabe, ihren Beginn und ihr Ende dar.

Ein Zeitplan oder ein Gantt-Diagramm hilft Projektmanagern und Arbeitern, klar zu erkennen, wann und in welcher Reihenfolge die verschiedenen Bauphasen durchgeführt werden sollen, um eine effiziente Nutzung

der Ressourcen und die Einhaltung der Fristen zu gewährleisten.

Lassen Sie uns einen Zeitplan für die Installation von drei Betonfundamentblöcken unter Verwendung der Berechnungen aus der obigen Tabelle erstellen. Benutzen Sie die Kostentabelle (Abb. 5.3-1) aus dem obigen Beispiel und bitten Sie das LLM, die Installation der drei Fundamentblöcke zu planen, z.B. für den ersten Mai 2024.

Um eine Kalkulation an LLM zu senden, können wir die Kalkulationstabelle im XLSX-Format hochladen oder einfach einen Screenshot eines JPEG-Bildes der Kalkulation direkt in den LLM-Chat einfügen (Abb. 5.3-1). Der LLM findet selbstständig eine Bibliothek zur Visualisierung des Tabellenbildes und fügt durch Multiplikation der Zeitattribute der Aktivitäten aus der Tabelle mit ihrem Volumen alle Daten in ein Diagramm ein.

- Senden Sie eine Textanfrage an LLM:

> Erstellen Sie einen Arbeitsplan in Form eines Gantt-Diagramms unter Verwendung der Werte aus der Zeittabelle (Abb. 5.3-1 als JPEG im Anhang) für die Installation von 3 Fundamentblöcken. Die Arbeiten für jeden Block sind nacheinander auszuführen. Der Beginn der Arbeiten soll auf den 01.05.2024 festgelegt werden. ⏎

- Die Antwort von LLM

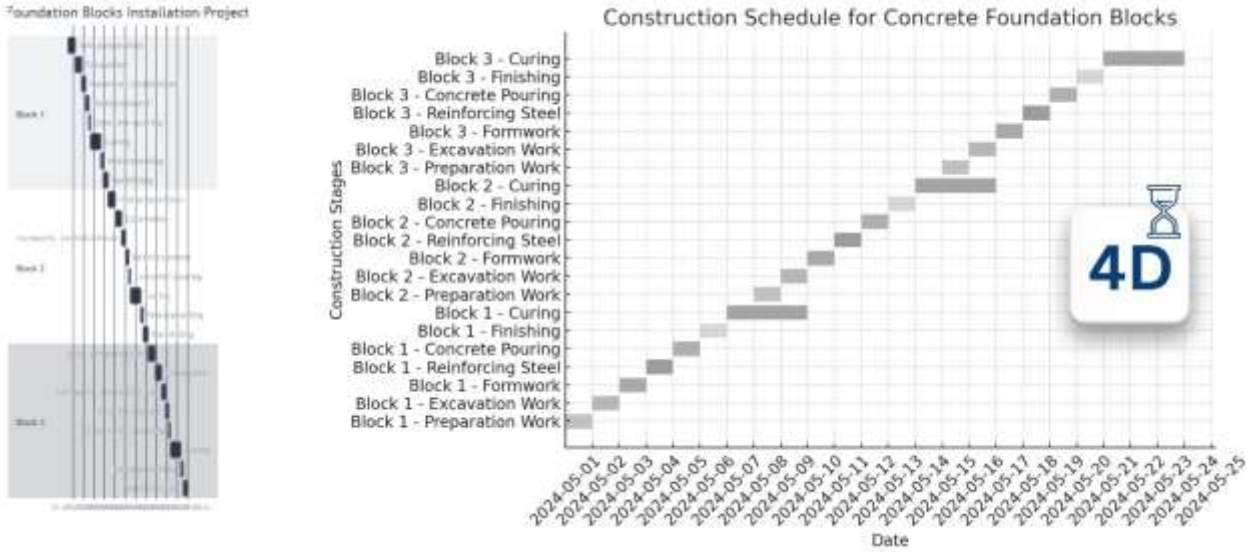


Abb. 5.3-2 Automatisch generiert durch mehrere LLMs Gantt-Diagramm zeigt die Bauabschnitte von drei Betonblöcken, entsprechend den Bedingungen aus der Eingabeaufforderung.

Die sich daraus ergebende Grafik (Abb. 5.3-2) ist ein Zeitdiagramm, in dem jeder horizontale Balken einem bestimmten Arbeitsschritt am Fundamentblock entspricht und die Abfolge der Vorgänge (Parameter "Arbeitsreihenfolge") anzeigt, wie z. B. Vorbereitung, Aushub, Schalungseinbau, Bewehrung, Betonieren und Fertigstellung, d. h. die Vorgänge, die in den Berechnungen mit Zeitparametern und Reihenfolge ausgefüllt wurden.

Ein solcher Zeitplan (Abb. 5.3-2) berücksichtigt keine Zwänge im Zusammenhang mit Arbeitstagen, Schichten oder Arbeitszeitnormen, sondern dient lediglich der konzeptionellen Visualisierung des Prozesses. Ein genauer Zeitplan, der die Parallelität der Arbeit widerspiegelt, kann durch entsprechende Hinweise oder zusätzliche Anweisungen im Chatroom ergänzt werden

Mit Hilfe einer einzigen Kalkulation (Abb. 5.3-1) ist es dank der Volumenattribute aus der 3D-Geometrie möglich, sowohl die Projektkosten über automatisierte Schätzungen automatisch abzuschätzen als auch gleichzeitig die zeitlichen Eigenschaften von Gruppen in Form von Tabellen oder Grafiken für verschiedene Projektvarianten zu berechnen (Abb. 5.3-3).

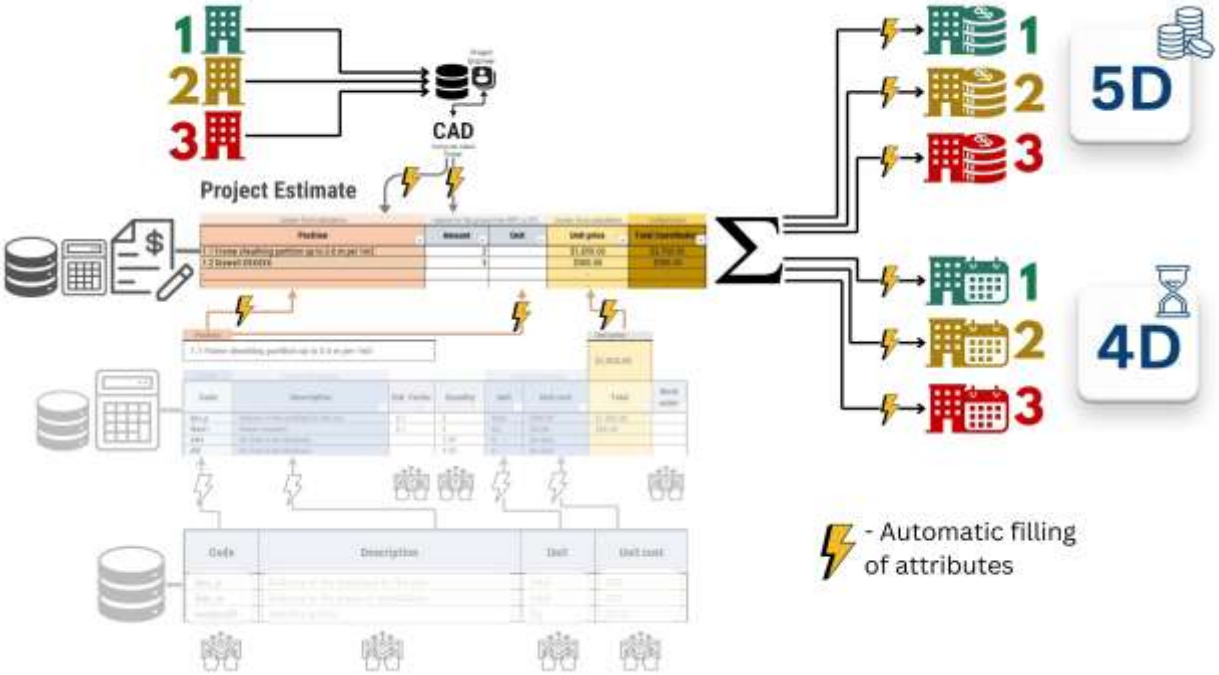


Abb. 5.3-3 Automatische Berechnung, ermöglicht sofortige und automatische Kosten- und Zeitprognosen für verschiedene Projektoptionen.

Moderne modulare ERP -Systeme (Abb. 5.4-4), die Daten aus CAD -Modellen laden, verwenden ähnliche automatische Zeitberechnungsmethoden, die den Entscheidungsprozess erheblich verkürzen. Dies ermöglicht eine sofortige und genaue Planung von Arbeitsplänen und die Berechnung der Gesamtzeit, die für die Durchführung aller Aufgaben bei der Projektrealisierung erforderlich ist, unter Berücksichtigung der realen Preise.

Erweiterte Attributsebenen 6D -8D: von der Energieeffizienz bis zur Sicherheitsgarantie

6D, 7D und 8D sind erweiterte Ebenen der Informationsmodellierung, die jeweils zusätzliche Schichten von Attributen zum umfassenden Projektinformationsmodell beitragen, deren Grundlage die Attribute des 3D -Modells mit ihrer Anzahl und ihrem Umfang sind. Jede zusätzliche Ebene steuert spezifische Parameter bei, die für eine weitere Gruppierung oder Identifizierung in anderen Systemen benötigt werden, wie z. B.

Property Management Systeme (PMS), Computer-Aided Facilities Management (CAFM), Construction Project Management (CPM) und Safety Management Systeme (SMS).

Abbildung 5.3-4 Die Attribute 6D, 7D und 8D im Dateninformationsmodell erweitern die Berücksichtigung verschiedener Aspekte des Projekts, von der Energieeffizienz bis zur Sicherheit.

- In **6D** werden zusätzlich zur Projektdatenbank (oder Datenrahmen (Abb. 4.1-13)) mit geometrischen und volumetrischen Attributen der Elemente Informationen (Attributspalten) zur ökologischen Nachhaltigkeit hinzugefügt. Dazu gehören Informationen zur Energieeffizienz, zum Kohlenstoff-Fußabdruck, zur Wiederverwertbarkeit von Materialien und zum Einsatz umweltfreundlicher Technologien. Anhand dieser Daten lassen sich die Umweltauswirkungen des Projekts bewerten, Projektentscheidungen optimieren und die Ziele für nachhaltige Entwicklung (ESG) erreichen.
- Die Attribute **von 7D** ergänzen die für die Verwaltung der Gebäudewartung erforderlichen Attribute. Dabei handelt es sich um Daten zu Wartungsplänen, Lebenszyklen von Komponenten, technischer Dokumentation und Reparaturhistorie. Diese Informationen stellen sicher, dass das Modell in Instandhaltungssysteme (CAFM, AMS) integriert werden kann, eine effiziente Planung der Instandhaltung und des Austauschs von Ausrüstungsgegenständen ermöglicht und Unterstützung während des gesamten Lebenszyklus der Anlage bietet.
- **8D** zusätzliche Attributschicht, - enthält Informationen zur Sicherheit - sowohl in der Bauphase als auch während des späteren Betriebs. Das Modell enthält Maßnahmen zur Personensicherheit, Notfallanweisungen, Evakuierungssysteme und Brandschutzanforderungen. Die Integration dieser Daten in das digitale Modell hilft, Risiken im Voraus zu berücksichtigen und architektonische, technische und organisatorische Lösungen zu entwickeln, die den Anforderungen an Sicherheit und Gesundheitsschutz Rechnung tragen.

In strukturierter Tabellenform stellen die Ebenen 4D bis 8D zusätzliche Attribute in Form von Spalten mit ausgefüllten Werten dar (Abb. 5.3-5), die zu den bereits ausgefüllten 3D -Modellattributen wie Name, Kategorie, Typ und volumetrische Merkmale hinzugefügt werden. Die Werte in den Attributsebenen 6D, 7D und 8D enthalten zusätzliche textliche und numerische Daten wie z. B. Recyclinganteil, CO2-Fußabdruck, Garantiezeit, Austauschzyklus, Installationsdatum, Sicherheitsprotokolle usw.

ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO2e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO2e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO2e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO2e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

Abbildung 5.3-5 6D -8D fügen dem Dateninformationsmodell, das bereits geometrische und volumetrische Attribute aus dem 3D -Modell enthält, Attributebenen hinzu.

Für unser neues Fenster (Abb. 4.4-1) kann das Element mit dem Bezeichner W-NEW (Abb. 5.3-5) die folgenden 3D -8D Attribute haben:

3D -Attribute - Geometrische Informationen aus CAD-Systemen:

- "Typbezeichnung" - Element "Fenster"
- "Breite" - 120 cm
- Zusätzlich können Sie die "Bounding Box " Punkte eines Elements oder dessen "Geometrie BREP / MESH " als separates Attribut hinzufügen

Attribute von 6D - ökologische Nachhaltigkeit:

- Recyclingquote von 90 Prozent
- "Kohlenstoff-Fußabdruck - 1.622 kg CO_2

Attribute 7D - Objektverwaltungsdaten:

- "Garantiezeit" - 8 Jahre
- "Replacement Cycle" ist 20 Jahre alt
- "Wartung" - jährlich erforderlich

Attribute von 8D - Gewährleistung der sicheren Nutzung und des Betriebs von Gebäuden:

- Fenster "Installiert" - von der Firma "XYZ Windows
- "Sicherheitsstandard" - entspricht der ISO 45001

Alle in einer Datenbank oder einem Datensatz (Abb. 5.3-5) erfassten Parameter werden von Fachleuten in verschiedenen Abteilungen zur Gruppierung, Suche oder Berechnung benötigt. Eine solche mehrdimensionale, attributbasierte Beschreibung von Projektobjekten liefert ein vollständiges Bild ihres Lebenszyklus, ihrer betrieblichen Anforderungen und vieler anderer Aspekte, die für die Planung, den Bau und den Betrieb von Projekten erforderlich sind.

Schätzung von CO_2 und Berechnung der Kohlendioxidemissionen von Bauprojekten

Neben dem Thema der Nachhaltigkeit von Bauprojekten auf Stufe 6D (Abb. 5.3-5) konzentriert sich das moderne Bauwesen auf die ökologische Nachhaltigkeit von Projekten, wobei einer der wichtigsten Aspekte die Bewertung und Minimierung der Kohlendioxid-CO_2 -Emissionen ist, die während der Lebenszyklusphasen eines Projekts (z. B. Herstellung und Installation) entstehen.

> Die Schätzung und Berechnung der Kohlenstoffemissionen von Baumaterialien ist ein Prozess, bei dem die gesamten Kohlenstoffemissionen durch Multiplikation der volumetrischen Eigenschaften eines Elements oder einer Gruppe von Elementen, die in einem Projekt verwendet werden, mit einem geeigneten Kohlenstoffemissionsfaktor für die jeweilige Kategorie bestimmt werden.

Die Berücksichtigung von Kohlenstoffemissionen bei der Bewertung von Bauprojekten als Teil der umfassenderen ESG-Kriterien (Umwelt, Soziales und Unternehmensführung) verleiht der Analyse eine neue Komplexitätsebene. Dies ist für den Bauherrn/Investor besonders wichtig, um eine entsprechende Zertifizierung wie LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) oder DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen) zu erhalten. Die Erlangung einer dieser Zertifizierungen kann die Marktfähigkeit einer Immobilie erheblich steigern, die Inbetriebnahme vereinfachen und die Einhaltung der Anforderungen nachhaltigkeitsorientierter Mieter (ESG) sicherstellen. Je nach Projektanforderungen können auch HQE (Haute Qualité Environnementale, der französische Green-Building-Standard), WELL (WELL Building Standard, mit Schwerpunkt auf Gesundheit und Komfort der Nutzer) und GRESB (Global Real Estate Sustainability Benchmark) zum Einsatz kommen

> Umwelt, Soziales und Unternehmensführung **ESG** (Umwelt, Soziales und Unternehmensführung) ist eine breite Palette von Grundsätzen, die zur Bewertung der Unternehmensführung sowie der sozialen und ökologischen Auswirkungen eines Unternehmens sowohl intern als auch extern verwendet werden können.

ESG, ursprünglich in den frühen 2000er Jahren von Finanzfonds entwickelt, um Anlegern Informationen über umfassende Umwelt-, Sozial- und Governance-Kriterien zur Verfügung zu stellen, hat sich zu einem Schlüsselindikator für die Bewertung von Unternehmen und Projekten, einschließlich Bauprojekten, entwickelt. Nach Untersuchungen großer Beratungsunternehmen werden Umwelt-, Sozial- und Governance-Überlegungen (ESG) zu einem integralen Bestandteil der Bauindustrie.

> Laut EY (2023) "The Path to Carbon Neutrality" reduzieren Unternehmen, die ESG -Prinzipien aktiv umsetzen, nicht nur langfristige Risiken, sondern steigern auch die Effizienz ihrer Geschäftsmodelle, was bei der globalen Umgestaltung der Märkte besonders wichtig ist [103]. Der PwC-Bericht "ESG Awareness" stellt fest, dass das Bewusstsein der Unternehmen für die Bedeutung der ESG -Faktoren zwischen 67% und 97% liegt, wobei die meisten Organisationen diese Trends als entscheidend für die künftige Nachhaltigkeit ansehen [104], und dass die Unternehmen größtenteils erheblichen Druck seitens der Stakeholder verspüren, ESG-Prinzipien zu integrieren.

Die Integration von ESG -Prinzipien in Bauprojekte trägt somit nicht nur zur Erlangung internationaler Nachhaltigkeitszertifikate wie LEED, BREEAM, DGNB bei, sondern sichert auch die langfristige Nachhaltigkeit und Wettbewerbsfähigkeit von Unternehmen in der Branche.

Einer der wichtigsten Faktoren, die den gesamten CO_2-Fußabdruck eines Bauprojekts beeinflussen, sind die Produktions- und Logistikphasen von Baumaterialien und -komponenten. Die auf der Baustelle verwendeten Materialien haben oft einen entscheidenden Einfluss auf die gesamten CO_2-Emissionen, insbesondere in den frühen Phasen des Projektlebenszyklus - von der Gewinnung der Rohstoffe bis zur Anlieferung auf der Baustelle.

Die Berechnung der Emissionen nach Kategorie oder Art des Bauelements erfordert die Verwendung von Referenz-Kohlenstoffemissionsfaktoren, die die Menge an CO_2 widerspiegeln, die bei der Herstellung der verschiedenen Materialien entsteht. Zu diesen Materialien gehören Beton, Ziegel, recycelter Stahl, Aluminium und andere. Diese Werte werden in der Regel aus seriösen Quellen und internationalen Datenbanken wie dem britischen ICE 2015 (Inventory of Carbon and Energy) und der US EPA 2006 (U.S. Environmental Protection Agency) [105] entnommen. Die folgende Tabelle (Abb. 5.3-6) fasst die Basisemissionsfaktoren für eine Reihe gängiger Baumaterialien zusammen. Für jedes Material sind zwei Schlüsselparameter angegeben: die spezifischen CO_2-Emissionen (in Kilogramm pro Kilogramm Material) und die Umrechnungsfaktoren von Volumen in Gewicht (in Kilogramm pro Kubikmeter), die für die Integration der Berechnungen in das Designmodell und die Verknüpfung mit der QTO-Datengruppierung erforderlich sind.

Carbon Emitted in Production

Material	Abbreviated	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K1)	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K2)	Coefficient m3 to kg Kg / m3 (K3)
Concrete	**Concrete**	0.12	0.12	**2400**
Concrete block	**Concrete_block**	0.13**	0.14	**2000**
Brick	**Brick**	0.24	0.32	**2000**
Medium density fiberboard (MDF)	**MDF**	0.39*	0.32	**700**
Recycled steel (avg recy content)	**Recycled_steel**	0.47	0.81	**7850**
Glass (not including primary mfg.)	**Glass**	0.59	0.6	**2500**
Cement (Portland, masonry)	**Cement**	0.95	0.97	**1440**
Aluminum (virgin)	**Aluminum**	12.79	16.6	**2700**

Abbildung 5.3-6 Menge des Kohlenstoffs, der bei der Herstellung verschiedener Baumaterialien emittiert wird, gemäß der britischen ICE- und der US EPA-Datenbank.

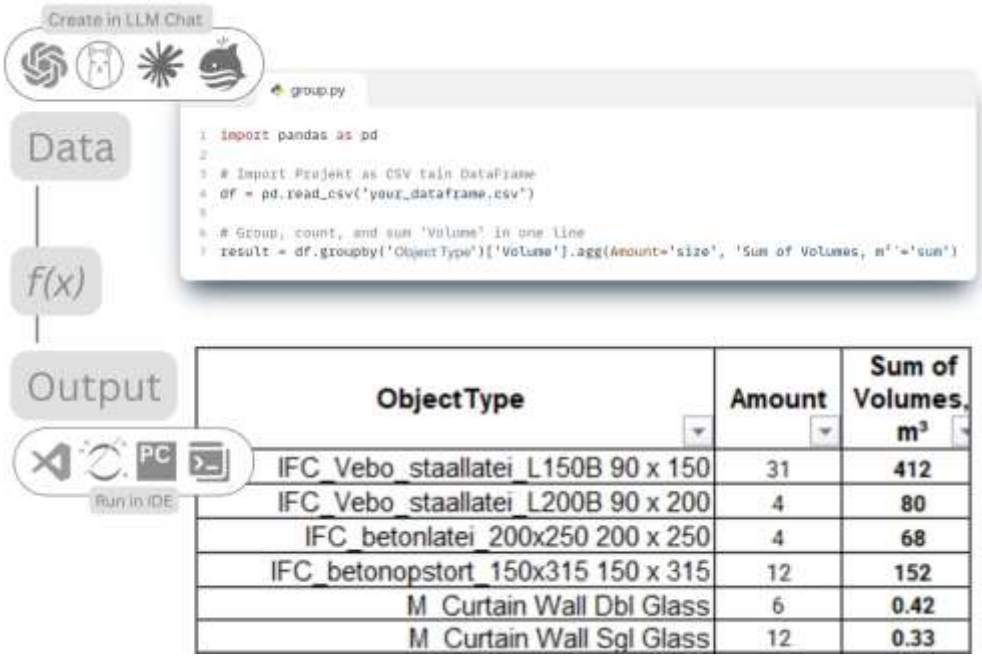
Um die gesamten CO_2-Emissionen für ein Projekt zu berechnen, ist es wie bei den 4D und 5D Berechnungen notwendig, die Attributvolumina jeder Gruppe von Einrichtungen zu bestimmen. Dies kann mit Hilfe quantitativer Analysewerkzeuge (QTO) erfolgen, indem die Attributvolumina in Kubikmetern ermittelt werden, wie im Abschnitt über die Mengenermittlung ausführlich beschrieben. Diese Volumina werden dann mit den entsprechenden Koeffizienten für das Attribut "CO_2-Prozessemissionen" der einzelnen Materialgruppen multipliziert.

- Extrahieren wir automatisch die Tabelle der Volumina nach Elementtyp aus dem CAD (BIM) Projekt, indem wir alle Projektdaten gruppieren, wie bereits in den vorherigen Kapiteln geschehen.

Um diese Aufgabe auszuführen, verweisen wir auf LLM.

> Bitte gruppieren Sie die DataFrame-Tabelle aus dem CAD (BIM) Projekt nach dem Spaltenparameter "Objektname" (oder "Typ") und zeigen Sie die Anzahl der Elemente in jeder Gruppe, und fassen Sie den Parameter "Volumen" für alle Elemente des Typs zusammen. ⏎

Die Antwort von LLM:



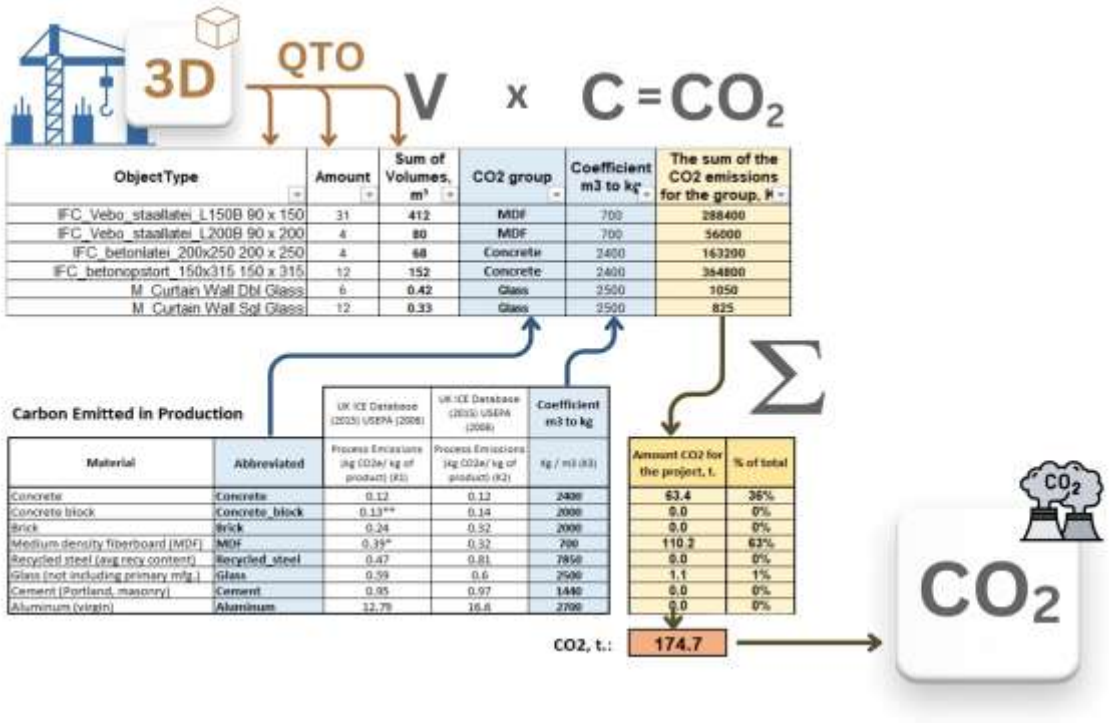
ObjectType	Amount	Sum of Volumes, m³
IFC_Vebo_staallatei_L150B 90 x 150	31	412
IFC_Vebo_staallatei_L200B 90 x 200	4	80
IFC_betonlatei_200x250 200 x 250	4	68
IFC_betonopstort_150x315 150 x 315	12	152
M_Curtain Wall Dbl Glass	6	0.42
M_Curtain Wall Sgl Glass	12	0.33

Abbildung 5.3-7 Der generierte Code in LLM hat die Projekteinheiten für uns nach Typ (ObjectType) mit dem zusammengefassten Attribut "Volume" gruppiert.

Um die Berechnung der gesamten CO_2-Emissionen für das gesamte Projekt zu automatisieren, richten Sie einfach eine automatische Datenzuordnung in der Tabelle ein oder ordnen Sie die Elementtypen (Abb. 5.3-7) manuell den entsprechenden Materialtypen (Abb. 5.3-6) aus der Emissionsfaktortabelle zu. Die fertige Tabelle mit Emissionsfaktoren und Formeln sowie der Code zum Abrufen von Volumina aus CAD-Formaten (BIM) und zur Automatisierung der CO_2-Bestimmung sind auf GitHub zu finden, indem man nach "CO_2_calculating-the-embodied-carbon. DataDrivenConstruction." [106].

So ermöglicht die Datenintegration nach der Gruppierung von QTO-Elementen aus der CAD-Datenbank die automatische Berechnung der Kohlendioxidemissionen (Abb. 5.3-8) für verschiedene Entwurfsoptionen. Dadurch ist es möglich, die Auswirkungen verschiedener Materialien in unterschiedlichen Varianten zu analysieren und nur solche Lösungen auszuwählen, die die CO_2-Emissionsanforderungen Kunden erfüllen, um bei der Inbetriebnahme des Gebäudes ein bestimmtes Zertifikat zu erhalten

Die Schätzung der CO_2-Emissionen durch die Multiplikation von Faktoren mit den Volumina gruppierter Projektelemente ist ein typisches Beispiel für eine Aufgabe im Rahmen des Prozesses eines Bauunternehmens zur Erlangung eines ESG-Ratings (z. B. einer LEED-Zertifizierung) für eine Einrichtung.



ObjectType	Amount	Sum of Volumes, m³	CO2 group	Coefficient m3 to kg	The sum of the CO2 emissions for the group, kg
IFC_Vebo_staallatei_L150B 90 x 150	31	412	MDF	700	288400
IFC_Vebo_staallatei_L200B 90 x 200	4	80	MDF	700	56000
IFC_betonlatei_200x250 200 x 250	4	68	Concrete	2400	163200
IFC_betonopstort_150x315 150 x 315	12	152	Concrete	2400	364800
M_Curtain Wall Dbl Glass	6	0.42	Glass	2500	1050
M_Curtain Wall Sgl Glass	12	0.33	Glass	2500	825

Carbon Emitted in Production

Material	Abbreviated	UK ICE Database (2023) USEPA (2008) Process Emissions (kg CO2e/ kg of product) (R1)	UK ICE Database (2023) USEPA (2008) Process Emissions (kg CO2e/ kg of product) (R2)	Coefficient m3 to kg kg / m3 (R3)	Amount CO2 for the project, t.	% of total
Concrete	Concrete	0.12	0.12	2400	63.4	36%
Concrete block	Concrete_block	0.13**	0.14	2000	0.0	0%
Brick	Brick	0.24	0.32	2000	0.0	0%
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700	110.2	63%
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850	0.0	0%
Glass (not including primary mfg.)	Glass	0.59	0.6	2500	1.1	1%
Cement (Portland, masonry)	Cement	0.85	0.97	1440	0.0	0%
Aluminum (virgin)	Aluminum	12.79	16.6	2700	0.0	0%

Abbildung 5.3-8 Die Integration von QTO-Gruppen aus CAD-Datenbanken bietet Genauigkeit und Automatisierung bei der Ableitung von Schätzungen der endgültigen CO_2-Emissionen.

In ähnlicher Weise können wir durch die Definition von Elementgruppenvolumina Berechnungen für Materialkontrolle und Logistik, Qualitätsüberwachung und -management, Energiemodellierung und -analyse und eine Vielzahl anderer Aufgaben durchführen, um einen neuen Attributstatus (Parameter in der Tabelle) sowohl für einzelne Elementgruppen als auch für das gesamte Projekt zu erhalten.

Wenn die Anzahl solcher Berechnungsprozesse im Unternehmen zunimmt, stellt sich die Frage nach der Notwendigkeit, diese Berechnungen zu automatisieren und die Ergebnisse der Berechnungen in die Prozesse und Datenverwaltungssysteme des Unternehmens zu implementieren.

Aufgrund der Komplexität einer Komplettlösung lagern mittlere und große Unternehmen der Baubranche diese Automatisierung an ERP (oder PMIS) Systementwicklungsunternehmen aus. Die Entwicklungsunternehmen erstellen ein einziges umfassendes modulares System für Großkunden, um viele verschiedene Informationsebenen zu verwalten, einschließlich Material- und Ressourcenberechnungen.

KAPITEL 5.4.
BAU ERP UND PMIS SYSTEME

Bau-ERP-Systeme am Beispiel von Kalkulationen und Kostenvoranschlägen

Modulare ERP-Systeme integrieren verschiedene Attribut-(Informations-)Schichten und Datenflüsse in ein einziges umfassendes System, das es den Projektmanagern ermöglicht, Ressourcen, Finanzen, Logistik und andere Aspekte eines Projekts auf synchronisierte Weise innerhalb einer einzigen Plattform zu verwalten. Ein Bau-ERP-System fungiert als "Gehirn" von Bauprojekten, vereinfacht sich wiederholende Prozesse durch Automatisierung und sorgt für Transparenz und Kontrolle während des gesamten Bauprozesses.

> **Bau-ERP-Systeme** (Enterprise Resource Planning) sind umfassende Softwarelösungen, die für die Verwaltung und Optimierung verschiedener Aspekte des Bauprozesses entwickelt wurden. Kernstück von Bau-ERP-Systemen sind Module zur Verwaltung der Kosten- und Terminplanung, was sie zu einem wichtigen Instrument für eine effiziente Ressourcenplanung macht.

ERP-Module -Systeme ermöglichen es den Nutzern, Daten strukturiert einzugeben, zu verarbeiten und zu analysieren, die verschiedene Aspekte eines Projekts abdecken, z. B. Material- und Arbeitskostenabrechnung, Gerätenutzung, Logistikmanagement, Personalwesen, Kontakte und andere Bauaktivitäten.

Einer der Funktionsblöcke des Systems ist das Modul für die Automatisierung der Geschäftslogik - BlackBox/WhiteBox, das die Rolle des Prozesskontrollzentrums spielt.

Die BlackBox /WhiteBox ermöglicht es Fachleuten, die ein ERP -System verwenden, über Zugriffsrechte verschiedene Aspekte des Unternehmens flexibel zu verwalten, die bereits von anderen Benutzern oder Administratoren vorkonfiguriert wurden. Im Zusammenhang mit ERP-Systemen beziehen sich die Begriffe *BlackBox* und *WhiteBox* auf den Grad der Transparenz und Kontrollierbarkeit der internen Logik des Systems:

- **BlackBox** ("Schwarzer Kasten") - der Benutzer interagiert mit dem System über die Schnittstelle, ohne Zugriff auf die interne Logik der Prozessausführung. Das System führt selbständig Berechnungen durch, die auf vordefinierten Regeln beruhen, die dem Endbenutzer verborgen bleiben. Er gibt Daten ein und erhält das Ergebnis, ohne zu wissen, welche Attribute oder Koeffizienten darin verwendet wurden.
- **WhiteBox** ("weiße Box") - die Prozesslogik ist zur Ansicht, Anpassung und Änderung verfügbar. Fortgeschrittene Benutzer, Administratoren oder Integratoren können Datenverarbeitungsalgorithmen, Berechnungsregeln und Interaktionsszenarien zwischen Projekteinheiten manuell definieren.

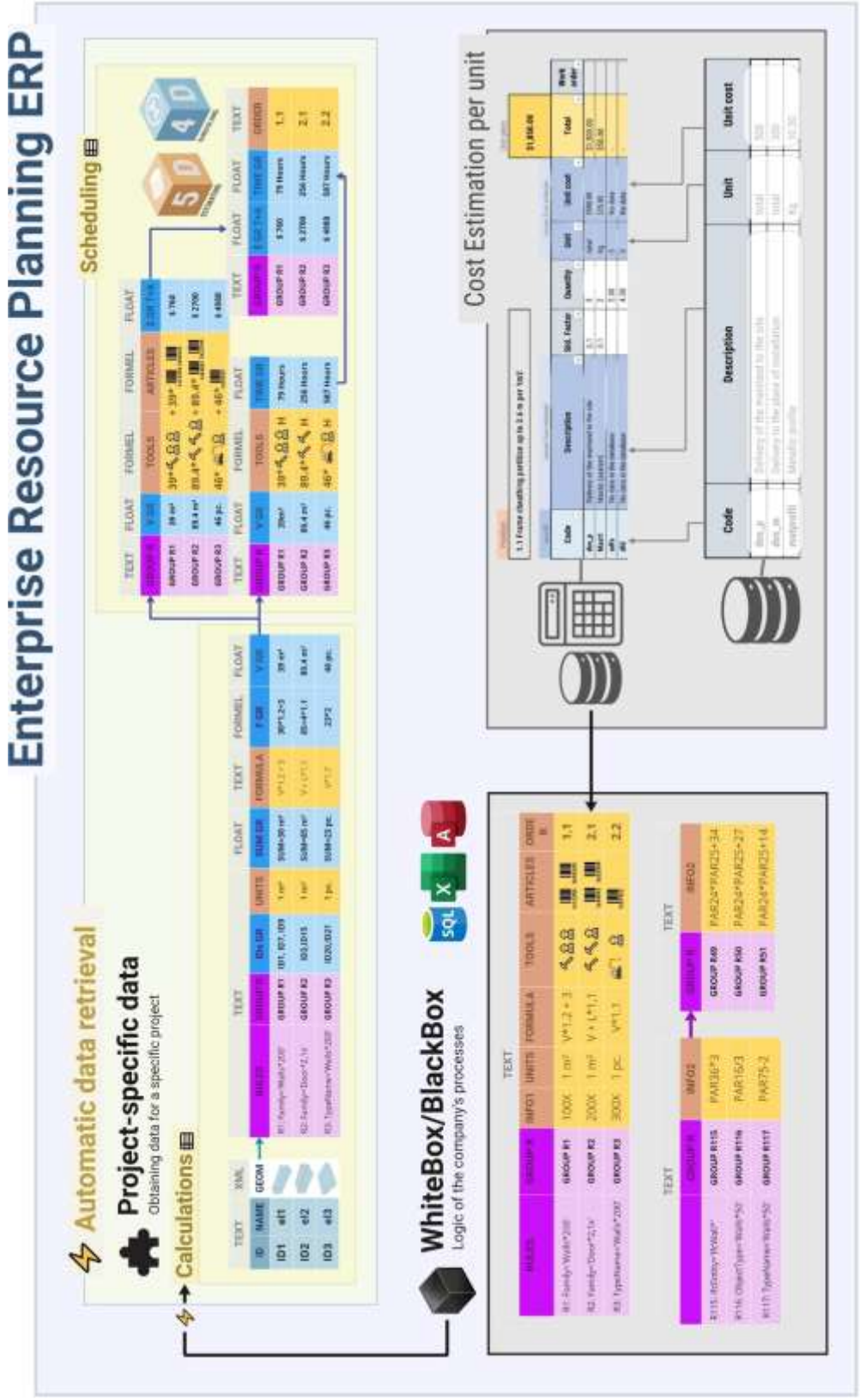


Abb. 5.4-1 Architektur des ERP-Systems für das Bauwesen -system, um Kostenvoranschläge und Arbeitspläne durch manuelles Ausfüllen der Volumenattribute zu erhalten.

Ein Beispiel: Ein erfahrener Benutzer oder Administrator legt eine Regel fest: welche Attribute in einem Kostenvoranschlag miteinander multipliziert oder nach einem bestimmten Merkmal gruppiert werden sollen und wo das Endergebnis gespeichert werden soll. Später laden weniger geschulte Fachleute, wie z. B. Kalkulationsingenieure, einfach neue Daten über die Benutzeroberfläche in das ERP-System - und erhalten fertige Kostenvoranschläge, Zeitpläne oder Spezifikationen, ohne Code schreiben oder die technischen Details der Logik verstehen zu müssen.

> In den vorangegangenen Kapiteln wurden die Berechnungs- und Logikmodule im Zusammenhang mit der LLM-Interaktion erörtert. In einer ERP-Umgebung finden solche Berechnungen und Transformationen in Modulen statt, die sich hinter einer Oberfläche aus Schaltflächen und Formularen verbergen.

Im folgenden Beispiel (Abb. 5.4-1) hat der Administrator des ERP -Systems im BlackBox /WhiteBox -Modul Regeln für den Abgleich von Attributen von Entitäten aus Schätzungen mit Attributen für die Gruppierung von QTO definiert. Dank dieses BlackBox/WhiteBox-Moduls, das (vom Manager oder Administrator) konfiguriert wurde, erhält der Benutzer (Kalkulator oder Ingenieur) durch manuelles Hinzufügen eines Mengen- oder Volumenattributs über die ERP-Benutzeroberfläche automatisch die fertigen Kostenvoranschläge und Arbeitspläne. Auf diese Weise werden die in den vorangegangenen Kapiteln besprochenen Kalkulations- und Kostenvoranschlagsprozesse mit Hilfe von Code innerhalb des ERP zu einer halbautomatischen Förderanlage.

Durch die Verbindung dieses halbautomatischen Prozesses mit den volumetrischen Attributen der CAD-Modelle (BIM) (Abb. 4.1-13), z. B. durch das Laden des CAD-Projekts in das dafür vorkonfigurierte ERP-Modul, wird der Datenfluss zu einem synchronisierten Mechanismus, der in der Lage ist, den Wert einzelner Elementgruppen oder des gesamten Projekts als Reaktion auf etwaige Änderungen während der Planungsphase beim Laden des CAD-Modells in das ERP selbständig und sofort zu aktualisieren.

Um einen automatisierten Datenfluss (Abb. 5.4-2) zwischen CAD (BIM) und ERP Systemen zu schaffen, müssen die grundlegenden Prozesse und Anforderungen für Daten aus CAD (BIM) Modelldatenbanken strukturiert definiert werden, wie wir bereits im obigen Kapitel "Anforderungen und Datenqualitätssicherung " diskutiert haben. Dieser Prozess in ERP ist in ähnliche Schritte unterteilt:

- **Erstellung von Validierungsregeln (1),** die eine wichtige Rolle bei der Gewährleistung der Richtigkeit der Daten spielen, die in das ERP -System eingegeben werden. Validierungsregeln dienen als Filter, die Entitäten und ihre Attribute validieren, so dass nur die Elemente, die die Anforderungen erfüllen, in das System gelangen können. Mehr über Verifizierung und Validierung erfahren Sie im Kapitel "Erstellen von Anforderungen und Validierung Datenqualität ".
- Dann **findet ein Verifizierungsprozess (2)** innerhalb von ERP statt, der bestätigt, dass alle Projektentitätselemente mit ihren Attributen und Werten korrekt erstellt wurden und für die nächsten Verarbeitungsschritte bereit sind.
- Wenn es Probleme mit unvollständigen Attributdaten gibt, **wird ein Bericht (3) erstellt** und das Projekt zusammen mit Anweisungen zur Korrektur zur Überarbeitung geschickt, bis es für die nächste Iteration bereit ist.
- Sobald die Projektdaten validiert und verifiziert sind, werden sie in einem anderen ERP-Modul **(4)** verwendet**, um Mengenabnahmetabellen (QTO) zu erstellen**, die Mengenattribute für Entitäts-

gruppen, Materialien und Ressourcen entsprechend den zuvor erstellten Regeln (WhiteBox/Black-Box) erzeugen.

- Gruppierte Daten nach Abgleichsregeln oder QTO werden automatisch **in Berechnungen (z.B. Kosten und Zeit) integriert (5).**
- Im letzten Schritt des ERP -Systems **generiert** der Anwender durch Multiplikation der Scope-Attribute aus der QTO-Tabelle mit den Attributen der Prozesstabellen (z.B. geschätzte Positionen) **automatisch Berechnungsergebnisse (6)** (z.B. Kostenschätzungen, Arbeitspläne oder CO_2-Emissionen) für jede Entitätsgruppe und für das Projekt als Ganzes.

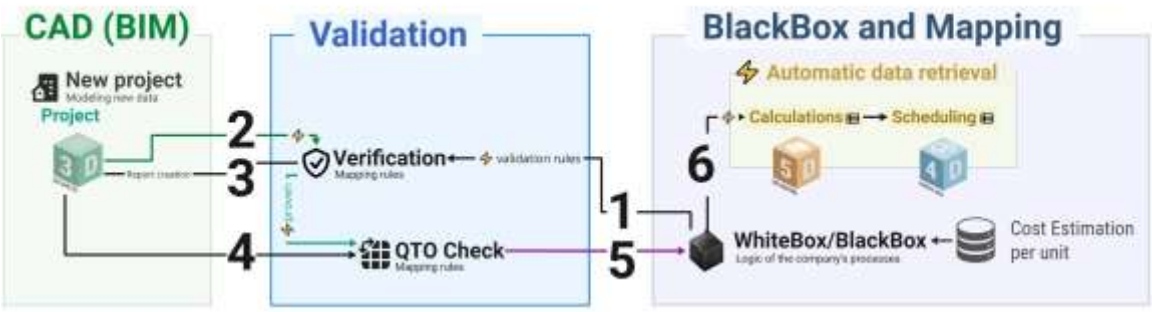


Abb. 5.4-2 Architektur eines Bau-ERP -Systems mit CAD (BIM), von der Erstellung von Validierungsregeln (1) bis zur automatischen Berechnung von Kosten und Arbeitsplänen (5-6).

In einem modularen ERP -System werden die Prozesse mithilfe einer Software integriert, die eine Benutzeroberfläche enthält. Hinter der Benutzeroberfläche befindet sich das Back-End, in dem strukturierte Tabellen Daten verarbeiten, indem sie verschiedene Operationen ausführen, die vom Manager oder Administrator vorkonfiguriert wurden. Im Ergebnis erhält der Benutzer dank der vordefinierten und angepassten Automatisierungslogik (in den Modulen BlackBox /WhiteBox) halbautomatisch erstellte Dokumente, die seinen Aufgaben entsprechen

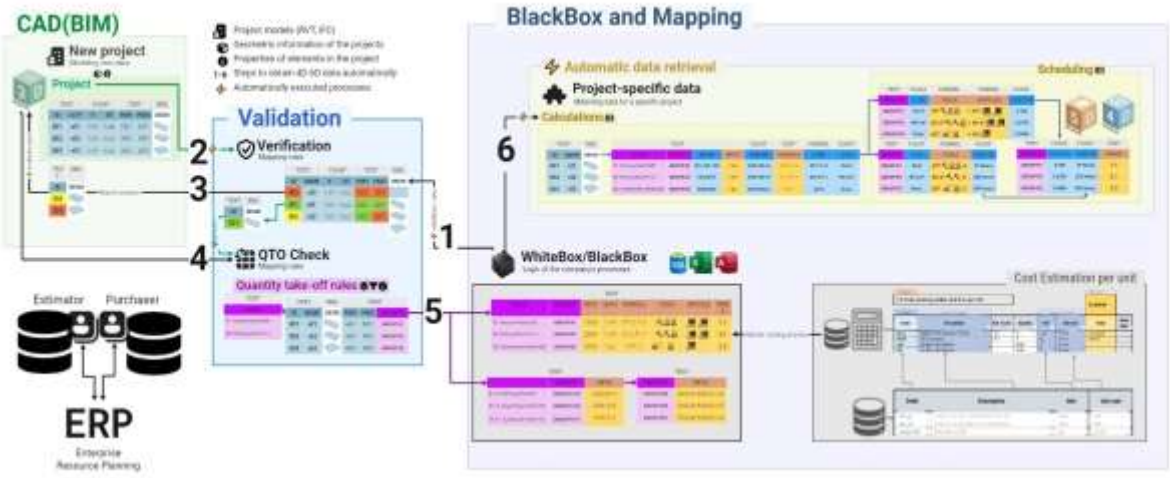


Abb. 5.4-3 ERP -System hilft Managern und Benutzern, sich zwischen Fachtabellen zu bewegen, um neue Daten zu generieren.

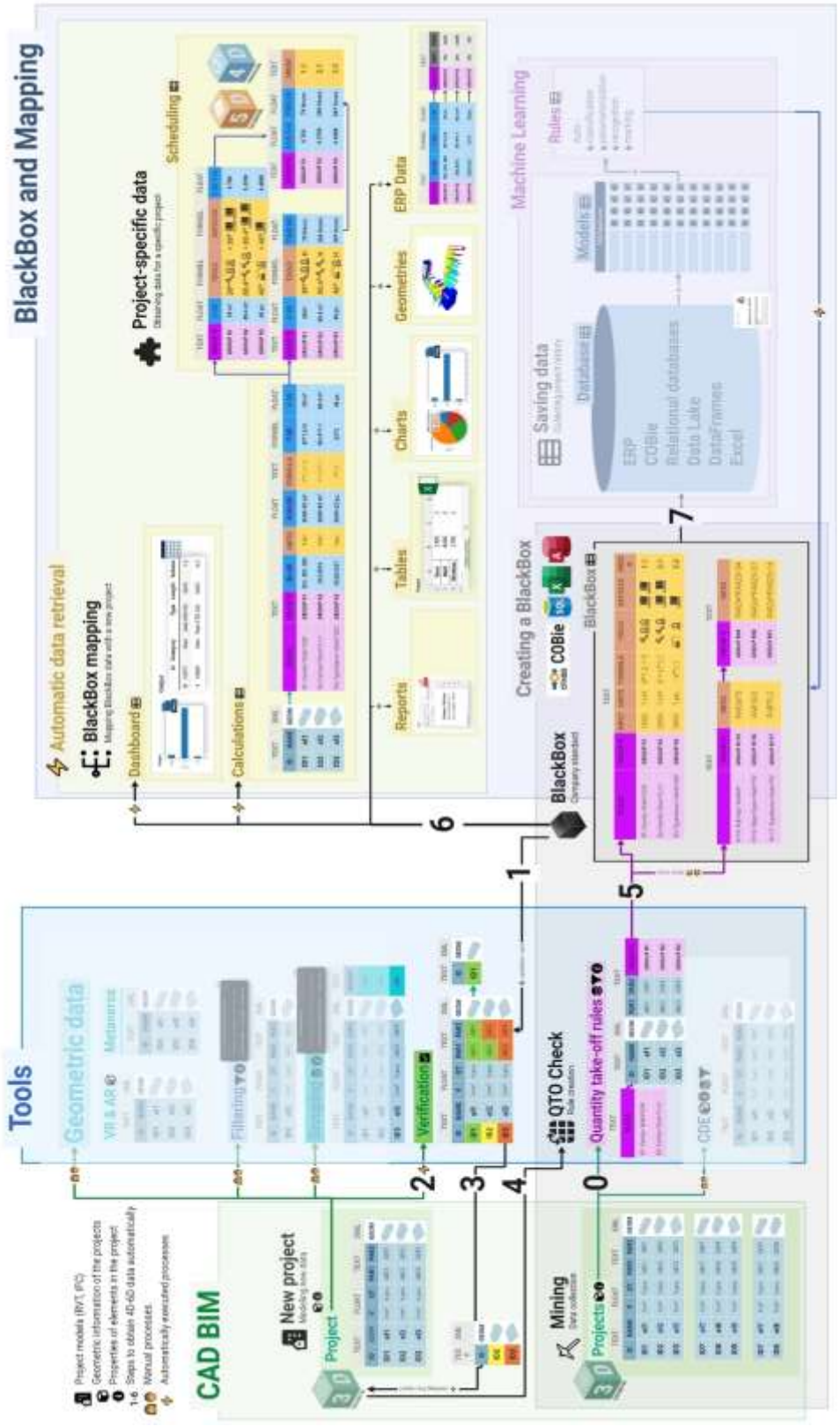


Abb. 5.4-4 ERP -System ist mit Analysetools integriert und automatisiert den Entscheidungsprozess des Unternehmens.

In ähnlicher Weise sind die Prozesse in ERP -Systemen von der Planung bis zur endgültigen Berechnung (Schritte 1-6, Abbildung 5.4-3) eine Kette miteinander verbundener Schritte, die letztendlich für Transparenz, Effizienz und Genauigkeit in der Planung sorgen.

Moderne ERP-Systeme für das Bauwesen umfassen nicht nur Module zur Kosten- und Zeitplanberechnung, sondern auch Dutzende von anderen vorkonfigurierten Modulen, die in der Regel das Dokumentenmanagement, die Verfolgung des Projektfortschritts, das Vertragsmanagement, die Lieferkette und die Logistik sowie die Integration mit anderen Geschäftssystemen und Plattformen abdecken. Die integrierten Analysewerkzeuge von ERP ermöglichen es den Anwendern, die Erstellung von Dashboards zur Überwachung von Projekt-KPIs (KPI - Key Performance Indicators) zu automatisieren. Dies ermöglicht ein zentralisiertes und konsistentes Management aller Aspekte eines Bauprojekts, wobei versucht wird, eine große Anzahl von Anwendungen und Systemen auf einer einzigen Plattform zu kombinieren.

In Zukunft werden ERP -Analysen in Kombination mit maschinellem Lernen eingesetzt, um die Genauigkeit zu erhöhen und den Prozess der Berechnung zukünftiger Projektattribute zu optimieren. Die in Big Data (Abb. 5.4-4) analysierten und gesammelten Daten und Attribute aus ERP-Systemen werden in Zukunft die Grundlage für die Erstellung von Vorhersagemodellen bilden, die potenzielle Verzögerungen, Risiken oder z. B. mögliche Änderungen der Materialkosten genau vorhersagen können.

Als Alternative zu ERP wird im Baugewerbe häufig PMIS (Project Management Information System) eingesetzt, ein Projektmanagementsystem, das für die detaillierte Steuerung von Aufgaben auf der Ebene eines einzelnen Bauprojekts konzipiert ist.

PMIS: Mittler zwischen ERP und der Baustelle

Im Gegensatz zu ERP, das die gesamte Kette der Geschäftsprozesse eines Unternehmens abdeckt, konzentriert sich PMIS auf die Verwaltung eines bestimmten Projekts, die Überwachung von Zeitplänen, Budgets, Ressourcen und Dokumentation.

> PMIS (Project Management Information System) ist eine Software für die Verwaltung von Bauprojekten, die für die Planung, Verfolgung, Analyse und Berichterstattung über alle Aspekte eines Projekts entwickelt wurde.

PMIS ermöglicht Ihnen die Verwaltung von Dokumenten, Zeitplänen und Budgets. Auf den ersten Blick mag PMIS wie ein Duplikat von ERP erscheinen, doch der entscheidende Unterschied liegt in der Verwaltungsebene:

- **ERP** ist auf die Geschäftsprozesse des Unternehmens als Ganzes ausgerichtet: Kosten-, Vertrags-, Beschaffungs-, Personal- und Ressourcenmanagement auf Unternehmensebene.
- **PMIS** konzentriert sich auf das individuelle Projektmanagement und bietet eine detaillierte Planung, Änderungskontrolle, Berichterstattung und Teilnehmerkoordination.

In vielen Fällen verfügen die ERP-Systeme bereits über ausreichende Funktionen, und die Implementierung von PMIS ist eher eine Frage der Bequemlichkeit und der Präferenz des Unternehmens. Viele Auftragnehmer und Kunden nutzen PMIS nicht, weil es notwendig ist, sondern weil es vom Anbieter oder einem Großkunden vorgeschrieben wird, der Daten auf einer bestimmten Plattform zusammenfassen möchte.

Es sollte erwähnt werden, dass es in der internationalen Terminologie für das Bauprojektmanagement andere separate populäre Begriffe wie PLM (Product Lifecycle Management) und EPC und EPC-M (Engineering, Procurement and Construction Management) - Vertragsmethoden in der Bauindustrie - gibt.

Wenn ein Unternehmen bereits ERP mit Projektmanagementmodulen verwendet, kann die Einführung von PMIS eine unnötige Verknüpfung sein, die die Funktionalität dupliziert. Wenn die Prozesse jedoch nicht automatisiert sind und die Daten fragmentiert sind, kann PMIS ein bequemeres und leichter zu pflegendes Werkzeug sein.

Spekulation, Profit, Insellage und mangelnde Transparenz bei ERP und PMIS

Trotz der äußerlichen Einfachheit von Schnittstellen und Verfahren sind ERP- und PMIS-Systeme für das Bauwesen in den meisten Fällen geschlossene und unflexible Lösungen. Solche Systeme werden in der Regel als vorkonfiguriertes Softwarepaket von einem einzigen Anbieter geliefert, mit begrenztem Zugang zu internen Datenbanken und Prozesslogik.

CAD-(BIM-) Anbieter übernehmen zunehmend die Entwicklung und Kontrolle solcher Systeme, da ihre Datenbanken die von ERP-Systemen benötigten Informationen enthalten: quantitative und volumetrische Attribute von Projektelementen. Anstatt jedoch den Zugriff auf diese Daten in einem offenen oder maschinenlesbaren Format zu ermöglichen, bieten die Anbieter nur begrenzte Benutzerszenarien und eine geschlossene Verarbeitungslogik an - vordefiniert in BlackBox-Modulen. Dies schränkt die Flexibilität des Systems ein und verhindert, dass es an spezifische Projektbedingungen angepasst werden kann.

> Begrenzte Datentransparenz ist nach wie vor eine der größten Herausforderungen bei digitalen Prozessen im Bauwesen. Geschlossene Datenbankarchitekturen, mangelnder Zugang zu vollständigen Attributsätzen von Bauelementen, die Konzentration auf BlackBox-Automatisierungsmodule und das Fehlen offener Schnittstellen erhöhen die Risiken der Dokumentenbürokratie erheblich. Solche Einschränkungen führen zu Engpässen im Entscheidungsprozess, erschweren die Überprüfung von Informationen und öffnen Tür und Tor für Datenverstecke oder Spekulationen in ERP/PMIS-Systemen. Die Benutzer erhalten in der Regel nur einen eingeschränkten Zugang - sei es eine abgespeckte Schnittstelle oder eine teilweise API - ohne die Möglichkeit, direkt mit den primären Datenquellen zu interagieren. Dies ist besonders kritisch, wenn es um Parameter geht, die automatisch aus CAD-Projekten generiert werden, wie z. B. Volumen, Flächen und Mengen, die für QTO-Berechnungen verwendet werden.

Anstatt nach Effizienz durch Prozessautomatisierung, offene Daten, Senkung der Transaktionskosten und Schaffung neuer Geschäftsmodelle zu streben, konzentrieren sich viele Bauunternehmen auf die Verwaltung externer Parameter - Manipulationsfaktoren, Anpassungsfaktoren und Berechnungsmethoden, die die Projektkosten in geschlossenen ERP/PMIS-Plattformen beeinflussen. Dies schafft Raum für Spekulationen, verzerrt die tatsächlichen Produktionskosten und verringert das Vertrauen zwischen allen am Bauprozess Beteiligten

Im Bauwesen ergibt sich der Gewinn aus der Differenz zwischen den Einnahmen aus einem abgeschlossenen Projekt und den variablen Kosten, zu denen Planung, Material, Arbeit und andere direkte Kosten gehören, die direkt mit der Durchführung des Projekts zusammenhängen. Der Schlüsselfaktor, der den Wert dieser Kosten beeinflusst, ist jedoch nicht nur die Technologie oder die Logistik, sondern auch die Geschwindigkeit und Genauigkeit der Berechnungen und die Qualität der Managemententscheidungen innerhalb des Unternehmens.

Das Problem wird durch die Tatsache verschärft, dass in den meisten Bauunternehmen die Kostenberechnungsprozesse nicht nur für die Kunden, sondern auch für die Mitarbeiter selbst, die nicht zu den Kalkulations- oder Finanzabteilungen gehören, intransparent bleiben. Diese Abgeschlossenheit begünstigt die Bildung einer privilegierten Gruppe von Spezialisten innerhalb des Unternehmens - Träger von "Finanzwissen", die das alleinige Recht haben, Attribute und Korrekturfaktoren in ERP/PMIS -Systemen zu bearbeiten. Diese Mitarbeiter können zusammen mit den Unternehmensleitern die finanzielle Logik des Projekts tatsächlich kontrollieren.

Schätzer werden unter solchen Bedingungen zu "Finanzjongleuren", die zwischen der Maximierung des Unternehmensgewinns und der Notwendigkeit, einen wettbewerbsfähigen Preis für den Kunden zu erzielen, abwägen müssen. Gleichzeitig müssen sie eklatante und grobe Manipulationen vermeiden, um den Ruf des Unternehmens nicht zu schädigen. In dieser Phase werden Koeffizienten festgelegt, um überhöhte Mengen oder Kosten für Materialien und Arbeiten zu verbergen.

Das wichtigste Mittel zur Steigerung der Effizienz und Rentabilität von Unternehmen im Baugewerbe ist daher nicht die Automatisierung und Beschleunigung von Entscheidungsprozessen, sondern die Spekulation mit Material- und Arbeitspreisen (Abb. 5.4-5). Die Überbewertung der Bau- und Materialkosten erfolgt durch die "graue" Buchhaltung in geschlossenen ERP /PMIS - Systemen, indem die Prozentsätze über den durchschnittlichen Marktpreisen für Materialien oder Bauvolumen mit Hilfe von Koeffizienten aufgebläht werden (Abb. 5.1-6), die im Kapitel "Zusammenstellung der Berechnungen und Berechnung der Baukosten auf der Grundlage der Ressourcenbasis" erörtert wurden.

Infolgedessen erhält der Kunde eine Kalkulation, die nicht die tatsächlichen Kosten oder den tatsächlichen Umfang der Arbeiten widerspiegelt, sondern eine Ableitung vieler versteckter interner Koeffizienten ist. Gleichzeitig sind die Subunternehmer in dem Bemühen, die vom Generalunternehmer zu niedrig angesetzten Preise einzuhalten, oft gezwungen, billigere und minderwertige Materialien zu kaufen, was die endgültige Qualität der Bauleistung verschlechtert.

Der spekulative Prozess der Suche nach Gewinnen aus dem Nichts schadet sowohl den Kunden, die unzuverlässige Daten erhalten, als auch den Führungskräften, die gezwungen sind, immer mehr Spekulationsmodelle zu finden

Je größer das Projekt, desto höher ist der bürokratische Aufwand für das Daten- und Prozessmanagement. Hinter jedem Schritt und jedem Modul verbergen sich oft undurchsichtige Koeffizienten und Zuschläge, die in Berechnungsalgorithmen und internen Verfahren eingebettet sind. Dies erschwert nicht nur die Rechnungsprüfung, sondern verzerrt auch das finanzielle Bild des Projekts erheblich. Bei großen Bauprojekten

führen solche Praktiken oft zu einer Vervielfachung (manchmal bis zum Zehnfachen) der Endkosten, während die tatsächlichen Mengen und Kosten außerhalb der effektiven Kontrolle des Auftraggebers bleiben (Abb. 2.1-3 Vergleich der geplanten und tatsächlichen Kosten großer Infrastrukturprojekte in Deutschland.

> Laut dem Bericht Imagining the Digital Future of Construction von McKinsey & Company (2016) werden große Bauprojekte im Durchschnitt 20% später als geplant und bis zu 80% über dem Budget abgeschlossen [107].

Die Kalkulations- und Budgetierungsabteilungen werden zum am meisten gehüteten Glied innerhalb eines Unternehmens. Der Zugang zu ihnen ist selbst für interne Spezialisten streng begrenzt, und aufgrund der geschlossenen Logik und der Datenbankstrukturen ist es unmöglich, die Effizienz von Projektentscheidungen objektiv und ohne Verzerrungen zu bewerten. Die fehlende Transparenz führt dazu, dass Unternehmen gezwungen sind, nicht die Prozesse zu optimieren, sondern durch "kreatives" Management von Zahlen und Faktoren ums Überleben zu kämpfen (Abb. 5.3-1, Abb. 5.1-6 - z.B. der Parameter "Bid. Factor").

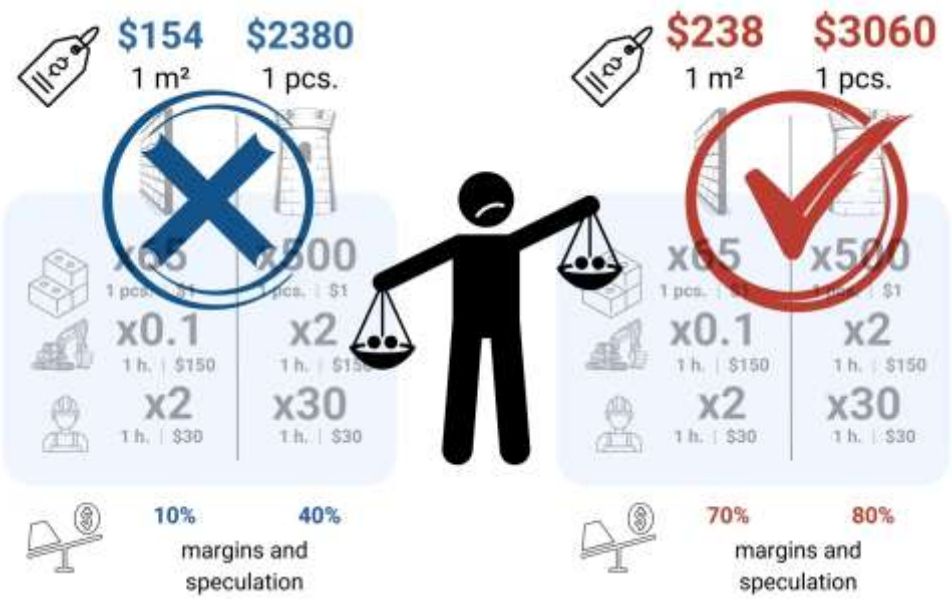


Abb. 5.4-5 Spekulationsquoten auf Abrechnungsebene sind der Hauptgewinn der Unternehmen und die Kunst des Jonglierens zwischen Arbeitsqualität und Reputation.

All dies stellt den weiteren Einsatz von geschlossenen ERP/PMIS -Systemen im Bauwesen in Frage. Vor dem Hintergrund der digitalen Transformation und steigender Transparenzanforderungen seitens der Kunden (Abb. 10.2-3) ist es unwahrscheinlich, dass die Projektrealisierung langfristig von proprietären Lösungen abhängig bleibt, die die Flexibilität einschränken, die Integration behindern und die Geschäftsentwicklung hemmen.

Und so günstig es für Bauunternehmen auch sein mag, mit Datensilos und undurchsichtigen Daten in geschlossenen Datenbanken zu arbeiten - die Zukunft der Bauindustrie wird unweigerlich einen Übergang zu offenen Plattformen, maschinenlesbaren und transparenten Datenstrukturen und vertrauensbasierter Automatisierung mit sich bringen. Dieser Wandel wird von der Unternehmensspitze vorangetrieben werden - unter dem Druck von Kunden, Regulierungsbehörden und der Gesellschaft, die zunehmend Rechenschaftspflicht, Nachhaltigkeit, Transparenz und wirtschaftliche Machbarkeit fordern.

Das Ende der Ära des geschlossenen ERP /PMIS: die Bauindustrie braucht neue Ansätze

Die Verwendung sperriger modularer ERP/PMIS -Systeme, die aus Dutzenden von Millionen Codezeilen bestehen, macht jede Änderung an ihnen extrem schwierig. In diesem Fall wird der Übergang zu einer neuen Plattform in Gegenwart von bereits für das Unternehmen vorkonfigurierten Modulen, Zehntausenden von Artikeln in Ressourcendatenbanken (Abb. 5.1-3) und Tausenden von vorgefertigten Berechnungen (Abb. 5.1-6) zu einem kostspieligen und langwierigen Prozess. Je mehr Code und Legacy-Architektur - desto höher die interne Ineffizienz, und jedes neue Projekt macht die Sache nur noch schlimmer. In vielen Unternehmen werden die Datenmigration und die Integration neuer Lösungen zu einem mehrjährigen Epos, begleitet von ständiger Nacharbeit und endloser Suche nach Kompromissen. Das Ergebnis ist oft eine Rückkehr zu alten, vertrauten Plattformen, trotz deren Einschränkungen.

Wie im deutschen Schwarzbuch-Bericht [108] über Systemfehler im Baudatenmanagement hervorgehoben wird, sind die Fragmentierung von Informationen und das Fehlen eines zentralisierten Ansatzes für ihre Verwaltung eine Hauptursache für Ineffizienz. Ohne Standardisierung und Integration verlieren die Daten ihren Wert und werden eher zu einem Archiv als zu einem Managementinstrument.

> Eine der Hauptursachen für den Verlust von Datenqualität ist die unzureichende Planung und Kontrolle von Bauprojekten, die oft zu erheblichen Kostensteigerungen führt. Der Abschnitt 'Focus: The Cost Explosion' des Black Book analysiert die wichtigsten Faktoren, die zu diesen unerwünschten Folgen beitragen. Dazu gehören unzureichende Bedarfsanalysen, fehlende Machbarkeitsstudien und eine unkoordinierte Planung, die zu zusätzlichen Kosten führt, die hätten vermieden werden können.

> In einem ausgereiften IT-Ökosystem eines Unternehmens ist der Austausch eines veralteten Systems vergleichbar mit dem Austausch einer tragenden Säule in einem bereits errichteten Gebäude. Es reicht nicht aus, die alte zu entfernen und eine neue zu installieren - es ist wichtig, dies so zu tun, dass das Gebäude stabil bleibt, die Decken nicht einstürzen und die gesamte Kommunikation weiterhin funktioniert. Genau hier liegt die Schwierigkeit: Jeder Fehler kann schwerwiegende Folgen für das gesamte Unternehmenssystem haben.

Dennoch führen die Entwickler großer ERP-Produkte für das Baugewerbe die Menge des geschriebenen Codes weiterhin als Argument für ihre Plattform an. Auf Fachkonferenzen hört man immer noch Sätze wie: ″Man bräuchte 150 Mannjahre, um ein solches System neu zu erstellen″, obwohl sich hinter dem Großteil der Funktionalität solcher Systeme Datenbanken und recht einfache Funktionen für die Arbeit mit Tabellen verbergen, die in eine spezielle feste Benutzeroberfläche verpackt sind. In der Praxis wird das Codevolumen von "150 Mannjahren" eher zu einer Belastung als zu einem Wettbewerbsvorteil. Je mehr Code, desto höher die Supportkosten, desto schwieriger die Anpassung an neue Bedingungen und desto höher die Eintrittsschwelle für neue Entwickler und Kunden.

> Viele modulare Bausysteme ähneln heute schwerfälligen und veralteten "Frankenstein-Konstruktionen", bei denen jede unbedachte Änderung zu Ausfällen führen kann. Jedes neue Modul erhöht die Komplexität eines bereits überladenen Systems und verwandelt es in ein Labyrinth, das nur wenige Spezialisten verstehen können, was die Wartung und Modernisierung noch schwieriger macht.

Die Komplexität wird auch von den Entwicklern selbst wahrgenommen, die in regelmäßigen Abständen ein Refactoring vornehmen, d. h. die Architektur überarbeiten, um dem Aufkommen neuer Technologien Rechnung zu tragen. Doch selbst wenn regelmäßig Refactoring durchgeführt wird, nimmt die Komplexität unweigerlich zu. Die Architekten solcher Systeme gewöhnen sich an die wachsende Komplexität, aber für neue Benutzer und Spezialisten wird sie zu einem unüberwindbaren Hindernis. Infolgedessen konzentriert sich das gesamte Fachwissen in den Händen einiger weniger Entwickler, und das System ist nicht mehr skalierbar. Kurzfristig sind solche Experten nützlich, aber langfristig werden sie Teil des Problems.

> *Unternehmen werden auch weiterhin "kleine" Daten mit ihren "großen" Daten integrieren, und es ist töricht zu glauben, dass eine einzige Anwendung - egal wie teuer oder robust sie ist - alles bewältigen kann [109].*
>
> - Phil Simon, Gastgeber des Podcasts "Gespräche über Zusammenarbeit

Es stellt sich die berechtigte Frage: Brauchen wir wirklich solch schwerfällige und geschlossene Systeme zur Berechnung von Kosten und Zeitplan von Arbeiten in Form von Tabellen, wenn in anderen Branchen analytische Werkzeuge mit offenen Daten und transparenter Logik ähnliche Aufgaben schon lange bewältigen?

Gegenwärtig sind geschlossene modulare Plattformen in der Bauindustrie immer noch gefragt, vor allem wegen der Besonderheiten der Kostenrechnung (Abb. 5.1-7). Solche Systeme werden häufig für graue oder undurchsichtige Systeme verwendet, so dass die tatsächlichen Kosten vor dem Kunden verborgen werden können. In dem Maße, in dem die Branche digital reift, vor allem die Kunden, und sich in das so genannte "Uber-Zeitalter" bewegt, werden die Zwischenhändler, d. h. die Bauunternehmen mit ihren ERP-Systemen, ihre Bedeutung bei der Zeit- und Kostenberechnung verlieren. Dies wird das Gesicht der Bauindustrie für immer verändern. Lesen Sie mehr im letzten Teil des Buches und im Kapitel "Bau 5.0: Wie man Geld verdient, wenn man sich nicht mehr verstecken kann".

> Tausende von Legacy-Lösungen, die sich in den letzten 30 Jahren angesammelt haben und in deren Entwicklung Tausende von Mannjahren investiert wurden, werden in Kürze verschwinden. Der Übergang zu einer offenen, transparenten und flexiblen Datenverwaltung ist unvermeidlich. Die Frage ist nur, welche Unternehmen in der Lage sein werden, sich an diese Veränderungen anzupassen, und welche in der Geiselhaft des alten Modells bleiben werden.

Eine ähnliche Situation ist im Bereich der CAD-Werkzeuge (BIM -) zu beobachten, deren Daten heute die volumetrischen Parameter von Planungseinheiten in ERP/PMIS -Systemen füllen. Ursprünglich die Idee von BIM (entwickelt im Jahr 2002 [110]) auf dem Konzept einer einzigen integrierten Datenbank, aber in der

Praxis erfordert die Arbeit mit BIM heute eine ganze Reihe spezialisierter Software und -Formate. Was eigentlich die Planung und das Baumanagement vereinfachen sollte, hat sich zu einer weiteren Schicht proprietärer Lösungen entwickelt, die die Integration erschweren und die Flexibilität des Unternehmens einschränken.

Nächste Schritte: effiziente Nutzung von Projektdaten

In diesem Teil haben wir gezeigt, wie strukturierte Daten die Grundlage für genaue Kosten- und Terminberechnungen für Bauprojekte bilden. Die Automatisierung der QTO-, Terminplanungs- und Kalkulationsprozesse reduziert die Arbeitskosten und verbessert die Genauigkeit der Ergebnisse erheblich.

Um diesen Teil zusammenzufassen, lohnt es sich, die wichtigsten praktischen Schritte hervorzuheben, die Ihnen helfen werden, die besprochenen Ansätze auf Ihre täglichen Aufgaben anzuwenden. Diese Ansätze sind universell - sie sind sowohl für die digitale Transformation eines Unternehmens als auch für die tägliche Arbeit der mit Berechnungen befassten Fachleute nützlich:

- Automatisieren Sie Routineberechnungen
  - ☐ Versuchen Sie, Standardkalkulationen zu finden, auf die Sie sich bei Ihrer Arbeit beziehen können.
  - ☐ Analysieren Sie, welche Methoden zur Kostenermittlung oder Kalkulation von Arbeiten oder Prozessen auf der Baustelle in Ihrem Land verwendet werden (Abb. 5.1-7).
  - ☐ Wenn Sie mit einem CAD-System arbeiten - lernen Sie die Funktionen der automatischen Extraktion von Spezifikationen und QTO-Daten in Ihrer CAD-Software (BIM-) kennen.
  - ☐ Verwenden Sie den LLM, um Codeentwürfe zur Automatisierung von Berechnungen zu schreiben.
- Entwickeln Sie Ihre eigenen Tools für QTO
  - ☐ Skripte oder Tabellen zur Automatisierung der Volumenzählung erstellen
  - ☐ Standardisierung von Kategorien und Gruppen von Elementen für einen einheitlichen Ansatz bei der Bewertung
  - ☐ Dokumentation der Berechnungsmethode, um die Reproduzierbarkeit der Ergebnisse in neuen Projekten zu gewährleisten
- Integrieren Sie verschiedene Aspekte des Projekts in Ihre Arbeit
  - ☐ Wenn Sie mit modularen Systemen arbeiten, versuchen Sie, Ihre Prozesse nicht nur in Form von Diagrammen oder Tabellen zu visualisieren, sondern auch auf der Ebene der Daten - insbesondere in Form von Tabellen
  - ☐ Beherrschen Sie die automatische Zusammenführung von Daten aus CAD-Datenbanken mit Berechnungen - mit Python-Code unter Verwendung von Gruppierung, Filterung und Aggregation
  - ☐ Erstellung klarer Visualisierungen von QTO-Gruppen zur Präsentation komplexer Informationen für Kollegen und Kunden

Diese Schritte werden dazu beitragen, ein nachhaltiges Berechnungssystem aufzubauen, das auf der Automatisierung und Standardisierung von Daten beruht. Dieser Ansatz wird die Genauigkeit verbessern und die Routine der täglichen Berechnungen verringern.

Die folgenden Kapitel befassen sich mit den technischen Aspekten von CAD - (BIM-) Produkten und den Gründen, warum CAD-Datenbanken immer noch schwer in die Geschäftsprozesse von Unternehmen zu integrieren sind. Wenn Sie jetzt nicht an der Geschichte der BIM-Implementierung im Bauwesen, der Entwicklung der CAD-Tools und den technischen Aspekten der Arbeit mit diesen Technologien interessiert sind, können Sie direkt zum siebten Teil des Buches "Datengestützte Entscheidungsfindung" übergehen.

MAXIMALER KOMFORT MIT DER GEDRUCKTEN VERSION

Sie sind im Besitz einer kostenlosen digitalen Version von **Data-Driven Construction**. Für bequemeres Arbeiten und schnellen Zugriff auf die Materialien empfehlen wir Ihnen, die **gedruckte Ausgabe** zu verwenden:

- **Immer zur Hand**: Das Buch im gedruckten Format wird zu einem zuverlässigen Arbeitsinstrument, das es Ihnen ermöglicht, in jeder Arbeitssituation schnell die richtigen Visualisierungen und Diagramme zu finden und zu verwenden

- **Hochwertige Illustrationen**: Alle Bilder und Grafiken in der gedruckten Ausgabe werden in höchster Qualität präsentiert

- **Schneller Zugriff auf Informationen**: einfache Navigation, die Möglichkeit, Notizen zu machen, Lesezeichen zu setzen und überall mit dem Buch zu arbeiten.

Mit dem Kauf der gedruckten Vollversion des Buches erhalten Sie ein praktisches Hilfsmittel für die bequeme und effiziente Arbeit mit Informationen: die Möglichkeit, bei alltäglichen Aufgaben schnell auf Anschauungsmaterial zurückzugreifen, die notwendigen Schemata schnell zu finden und Notizen zu machen. Darüber hinaus unterstützen Sie mit Ihrem Kauf die Verbreitung von offenem Wissen.

Bestellen Sie eine gedruckte Version des Buches unter: datadrivenconstruction.io/books

VI TEIL

CAD UND BIM: MARKETING, REALITÄT UND DIE ZUKUNFT DER KONSTRUKTIONSDATEN IM BAUWESEN

Der sechste Teil des Buches enthält eine kritische Analyse der Entwicklung der CAD und BIM -Technologien und ihrer Auswirkungen auf die Datenmanagementprozesse im Bauwesen. Er zeichnet den historischen Wandel des BIM-Konzepts von der ursprünglichen Idee einer integrierten Datenbank bis hin zu den aktuellen, von Softwareanbietern beworbenen Marketingkonstrukten nach. Die Auswirkungen proprietärer Formate und geschlossener Systeme auf die Effizienz der Projektdatenverarbeitung und die Gesamtleistung der Bauindustrie werden bewertet. Die Probleme der Kompatibilität verschiedener CAD-Systeme und die Schwierigkeiten ihrer Integration in die Geschäftsprozesse von Bauunternehmen werden im Detail analysiert. Aktuelle Trends in Richtung vereinfachter offener Datenformate, wie z.B. USD, und ihre möglichen Auswirkungen auf die Branche werden diskutiert. Es werden alternative Ansätze zur Extraktion von Informationen aus geschlossenen Systemen vorgestellt, einschließlich Reverse-Engineering-Techniken. Die Aussichten für die Anwendung von künstlicher Intelligenz und maschinellem Lernen zur Automatisierung von Entwurfs- und Datenanalyseprozessen im Bauwesen werden analysiert. Es werden Prognosen für die Entwicklung von Entwurfstechnologien formuliert, die sich an den tatsächlichen Bedürfnissen der Nutzer und nicht an den Interessen der Softwarehersteller orientieren.

KAPITEL 6.1.
ENTSTEHUNG VON BIM- KONZEPTEN IN DER BAUINDUSTRIE

Ursprünglich war dieser sechste Teil, der sich mit CAD (BIM) befasst, in der ersten Version des Buches nicht enthalten. Die Themen proprietäre Formate, geometrische Kernel und geschlossene Systeme sind zu technisch, überladen mit Details und scheinbar nutzlos für diejenigen, die einfach nur verstehen wollen, wie man mit Daten arbeitet. Rückmeldungen und Bitten, die erste Version des Buches zu vervollständigen, haben jedoch gezeigt, dass es unmöglich ist, wirklich zu verstehen, warum die von den Anbietern propagierten Konzepte die Arbeit mit Informationen erschweren und den Übergang zum offenen parametrischen Design behindern, ohne die Komplexität des Innenlebens von CAD-Systemen, Geometriekernen, die Vielfalt der Formate und inkompatiblen Speicherschemata für dieselben Daten zu verstehen. Aus diesem Grund hat dieser Teil einen eigenen Platz in der Struktur des Buches eingenommen. Wenn CAD (BIM) für Sie keine Priorität hat, können Sie direkt zum nächsten Teil übergehen - "TEIL VII: Datengestützte Entscheidungsfindung, Analytik, Automatisierung und maschinelles Lernen".

Geschichte des Aufkommens von BIM und Open BIM als Marketingkonzepte von CAD- Anbietern

Mit dem Aufkommen digitaler Daten in den 1990er Jahren wurde die Computertechnologie nicht nur in die Geschäftsprozesse, sondern auch in die Planungsprozesse eingeführt, was zu Konzepten wie CAD (Computer-Aided Design Systems) und später BIM (Building Information Modeling

Wie jede Innovation sind sie jedoch nicht der Endpunkt der Entwicklung. Konzepte wie BIM sind zu einem wichtigen Meilenstein in der Geschichte der Bauindustrie geworden, aber früher oder später werden sie besseren Werkzeugen und Ansätzen weichen, die den Herausforderungen der Zukunft besser gerecht werden.

> Überwältigt vom Einfluss der CAD-Anbieter und verwirrt von der Komplexität seiner eigenen Umsetzung, wird das Konzept von BIM, das 2002 aufkam, sein dreißigjähriges Bestehen wohl nicht mehr erleben, so wie ein Rockstar, der aufblühte, aber schnell wieder verblasste. Der Grund dafür ist einfach: Die Anforderungen der Datenwissenschaftler ändern sich schneller, als sich die CAD-Anbieter darauf einstellen können.

Angesichts des Mangels an qualitativ hochwertigen Daten verlangen die Fachleute der Baubranche heute plattformübergreifende Interoperabilität und Zugang zu offenen Daten aus CAD- Projekten, um deren Analyse und Verarbeitung zu vereinfachen. Die Komplexität von CAD-Daten und die verwirrende Verarbeitung von CAD-Daten wirkt sich negativ auf alle am Bauprozess Beteiligten aus: Konstrukteure, Projektmanager, Bauarbeiter auf der Baustelle und letztlich auch der Kunde.

Statt eines vollwertigen Datensatzes für den Betrieb erhalten Kunden und Investoren heute Container in CAD- Formaten, die komplexe geometrische Kerne, das Verständnis von Datenschemata, eine jährlich aktualisierte API -Dokumentation und spezialisierte CAD-Software (BIM) für die Arbeit mit den Daten erfordern. Gleichzeitig bleibt ein Großteil der Planungsdaten ungenutzt.

In der heutigen Planungs- und Bauwelt führt die Komplexität des Zugriffs auf CAD-Daten zu einem Over-Engineering des Projektmanagements. Mittlere und große Unternehmen, die mit CAD-Daten arbeiten oder BIM -Lösungen entwickeln, sind entweder gezwungen, enge Beziehungen zu CAD-Anbietern -Lösungen zu unterhalten, um über APIs auf die Daten zuzugreifen, oder die Beschränkungen der CAD-Anbieter zu umgehen, indem sie teure SDK-Konverter für das Reverse-Engineering verwenden, um offene Daten zu erhalten [75].

> Der Ansatz der proprietären Daten ist veraltet und entspricht nicht mehr den Anforderungen des heutigen digitalen Umfelds. Die Zukunft wird die Unternehmen in zwei Typen unterteilen: diejenigen, die offene Daten effektiv nutzen, und diejenigen, die den Markt verlassen.

Das Konzept BIM (Building Information Modeling) tauchte in der Baubranche mit der Veröffentlichung des Whitepapers BIM [54] eines der großen CAD-Anbieter im Jahr 2002 auf und entstand, ergänzt durch das Maschinenbaukonzept BOM (Bills of Materials), aus dem parametrischen Ansatz zur Erstellung und Bearbeitung von Projektdaten (Abb. 6.1-1). Der parametrische Ansatz für die Erstellung und Bearbeitung von Konstruktionsdaten wurde als einer der ersten im Pro-E-System für die Konstruktion im Maschinenbau (MCAD) umgesetzt. Dieses System wurde zum Prototyp [111] für viele moderne CAD -Lösungen, einschließlich derer, die heute in der Bauindustrie verwendet werden.

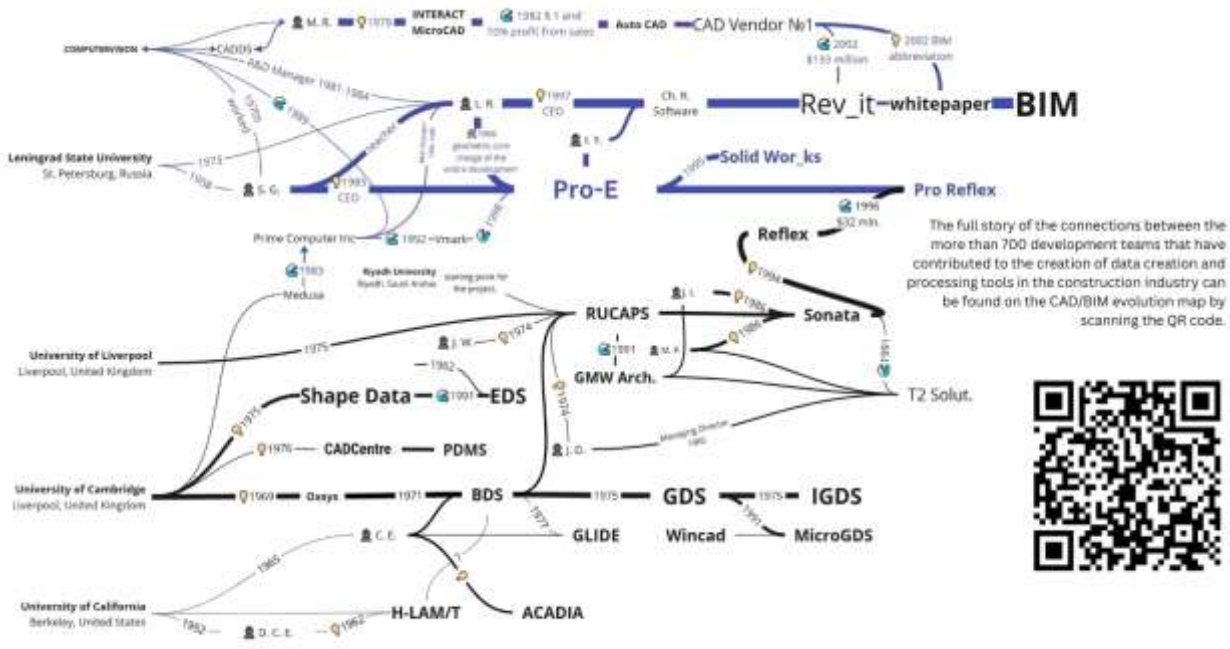


Abb. 6.1-1 Übersicht über die Geschichte des BIM-Konzepts und ähnlicher Konzepte.

Journalisten und AEC-Berater, die bis Anfang der 2000er Jahre für CAD-Tools -Anbieter warben, richteten ihre Aufmerksamkeit ab 2002 auf das Whitepaper BIM. Das Whitepaper BIM 2002-2004 und die in den Jahren 2002, 2003, 2005 und 2007 veröffentlichten Artikel spielten eine Schlüsselrolle bei der Popularisierung des BIM-Konzepts in der Baubranche [112].

> *Building Information Modelling ist eine Strategie..... [Name des CAD-Anbieters] zur Anwendung von Informationstechnologie in der Bauindustrie.*
>
> - BIM-Whitepaper, 2002 [60]

Mitte der 2000er Jahre begannen "Forscher", das 2002 vom CAD-Anbieter veröffentlichte Konzept BIM- mit früheren wissenschaftlichen Arbeiten wie Charles Eastmans BDS zu verknüpfen, die zur Grundlage für Systeme wie GLIDE, GBM, BPM und RUCAPS wurden. In seinem Pionierwerk Building Description System (1974) legte Charles Eastman die theoretischen Grundlagen der modernen Informationsmodellierung. Der Begriff "Datenbank " taucht in seinem Werk 43 Mal auf (Abb. 6.1-2) - so oft wie kein anderer, mit Ausnahme des Wortes "Gebäude".

> Eastmans Grundgedanke war, dass alle Informationen über ein Gebäude - von der Geometrie bis zu den Eigenschaften der Elemente und deren Beziehungen untereinander - in einer einzigen strukturierten Datenbank gespeichert werden sollten. Aus dieser Datenbank können Zeichnungen, Spezifikationen, Berechnungen und die Einhaltung von Vorschriften automatisch erstellt und analysiert werden. Eastman kritisierte Zeichnungen ausdrücklich als veraltete und redundante Kommunikationsmethode und verwies auf die Verdoppelung von Informationen, Probleme bei der Aktualisierung und die Notwendigkeit manueller Aktualisierungen bei Änderungen. Stattdessen schlug er ein einziges digitales Modell in einer Datenbank vor, bei dem jede Änderung nur einmal vorgenommen wird und sich automatisch auf alle Ansichten auswirkt.

Es ist bemerkenswert, dass Eastman in seinem Konzept die Visualisierung nicht in den Mittelpunkt gestellt hat. In seinem System standen Informationen im Mittelpunkt: Parameter, Beziehungen, Attribute, Analyse- und Automatisierungsmöglichkeiten. Zeichnungen waren in seinem Verständnis nur eine der Formen der Darstellung von Daten aus der Datenbank, nicht die primäre Quelle von Entwurfsinformationen.

Im ersten BIM-Whitepaper des führenden CAD-Anbieters wurde der Begriff "Datenbank " so oft verwendet wie in Charles Eastmans BDS - 23 Mal [60] auf sieben Seiten und war nach "Building", "Information", "Modeling" und "Design" eines der beliebtesten Wörter in dem Dokument. Bis 2003 taucht der Begriff "Datenbank" in ähnlichen Dokumenten jedoch nur noch zweimal auf [61], und Ende der 2000er Jahre war das Thema Datenbanken aus der Diskussion über Designdaten praktisch verschwunden. Infolgedessen wurde das Konzept einer "einzigen integrierten Datenbank für visuelle und quantitative Analysen" nie vollständig verwirklicht.

So hat sich die Bauindustrie von Charles Eastmans fortschrittlichem BDS-Konzept mit seiner Betonung von Datenbanken und Samuel Geisbergs Ideen über die automatische Aktualisierung von Konstruktionsdaten aus Datenbanken im Maschinenbau-Produkt Pro-E (dem Vorläufer der heute im Bauwesen verbreiteten CAD -Lösungen) zu dem aktuell vermarkteten BIM entwickelt, in dem die Datenverwaltung durch Datenbanken kaum erwähnt wird, obwohl dies das Konzept hinter dem ursprünglichen theoretischen

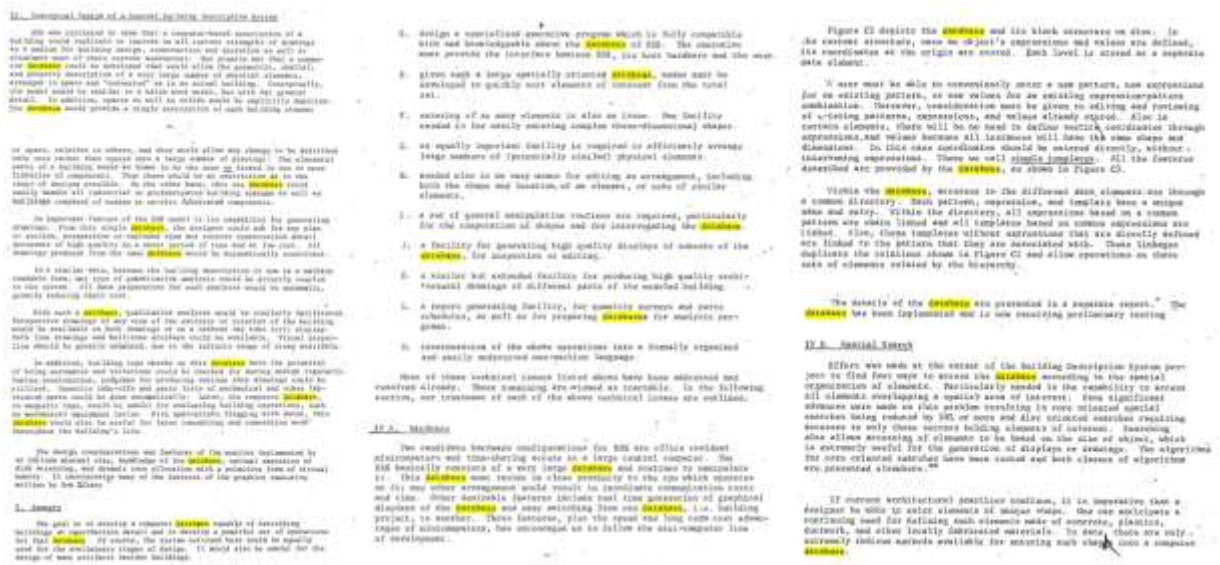


Abbildung 6.1-2 In dem von Charles Eastman 1974 beschriebenen BDS-Konzept wurde der Begriff "Datenbank" (gelb unterlegt) 43 Mal verwendet.

BDS und ähnliche Konzepte bis in die 2000er Jahre wurden als digitale Datenbank von Gebäuden und nicht als Visualisierungswerkzeug entwickelt. BIM wurde im Jahr 2002 zu einem Planungsinstrument, bei dem die Datenbank in den Hintergrund trat. Was haben wir beim Übergang von BDS und ähnlichen Konzepten in den 1990er Jahren zu BIM bis Mitte der 2010er Jahre verloren?

- Offene Datenbanken: BDS und andere ähnliche Konzepte legten den Schwerpunkt auf die Analytik, BIM auf die Gestaltung.
- Flexibilität bei der Arbeit mit Daten: BDS legte den Schwerpunkt auf die Datenanalyse, BIM auf Prozesse, die auf unklaren Daten beruhen.
- Transparenz: Der BDS sollte eine offene integrierte Datenbank sein, während CAD-Anbieter im Bereich BIM ihre Datenbanken vollständig geschlossen haben und seit 20 Jahren erfolglos gegen Reverse-Engineering-Tools kämpfen, die proprietäre Formate öffnen.

In den letzten 30 Jahren hatten Konstrukteure nie Zugang zu einer "integrierten Datenbank", und nach zwanzig Jahren Marketingeuphorie rund um BIM -Werkzeuge beginnt die Bauindustrie nun, die Folgen dieser Modeerscheinung zu erkennen.

Die Realität von BIM: anstelle von integrierten Datenbanken - geschlossene modulare Systeme

Anstatt sich auf die Daten zu konzentrieren, sie zu strukturieren und in einheitliche Prozesse zu integrieren, sind die Nutzer von CAD - (BIM-) Systemen gezwungen, mit einer fragmentierten Reihe von proprietären Lösungen zu arbeiten, von denen jede ihre eigenen Spielregeln diktiert:

- **Die einheitliche Datenbank**, die im ersten BIM-Whitepaper diskutiert wurde, **ist ein Mythos geblieben.** Trotz lautstarker Behauptungen ist der Zugang zu den Daten immer noch begrenzt und auf geschlossene Systeme verteilt.
- **BIM -Modelle sind eher zu einem geschlossenen Ökosystem** als zu einem Werkzeug geworden.

Anstelle eines transparenten Informationsaustauschs sind die Nutzer gezwungen, für Abonnements zu zahlen und proprietäre APIs zu verwenden.

- **Die Daten gehören den Anbietern, nicht den Nutzern.** Projektinformationen sind in proprietären Formaten oder Cloud-Diensten eingeschlossen und nicht in offenen und unabhängigen Formaten verfügbar.

Konstrukteure und Projektleiter haben oft keinen Zugriff auf die CAD-Datenbank -systeme und auch nicht auf das Format, in dem ihre eigenen Projektdaten gespeichert sind. Dies macht es unmöglich, Informationen schnell zu überprüfen oder Anforderungen an die Datenstruktur und -qualität zu formulieren (Abb. 6.1-3). Der Zugriff auf diese Daten erfordert eine ganze Reihe spezialisierter Programme, die über APIs und Plug-ins miteinander verbunden sind, was zu einer übermäßigen Bürokratisierung der Prozesse in der Bauwirtschaft führt. Inzwischen werden diese Daten gleichzeitig von Dutzenden von Informationssystemen und Hunderten von Spezialisten genutzt.

> *Wir müssen in der Lage sein, all diese Daten zu verwalten [CAD (BIM)], sie digital zu speichern und Software für das Lebenszyklus- und Prozessmanagement zu verkaufen, denn **auf jeden Ingenieur** [Designer], der etwas [in CAD-Software] erstellt, **kommen zehn Personen**, die mit diesen Daten arbeiten" [41].*
>
> - CEO von CAD - dem Anbieter, der das BIM-Konzept entwickelt hat, 2005.

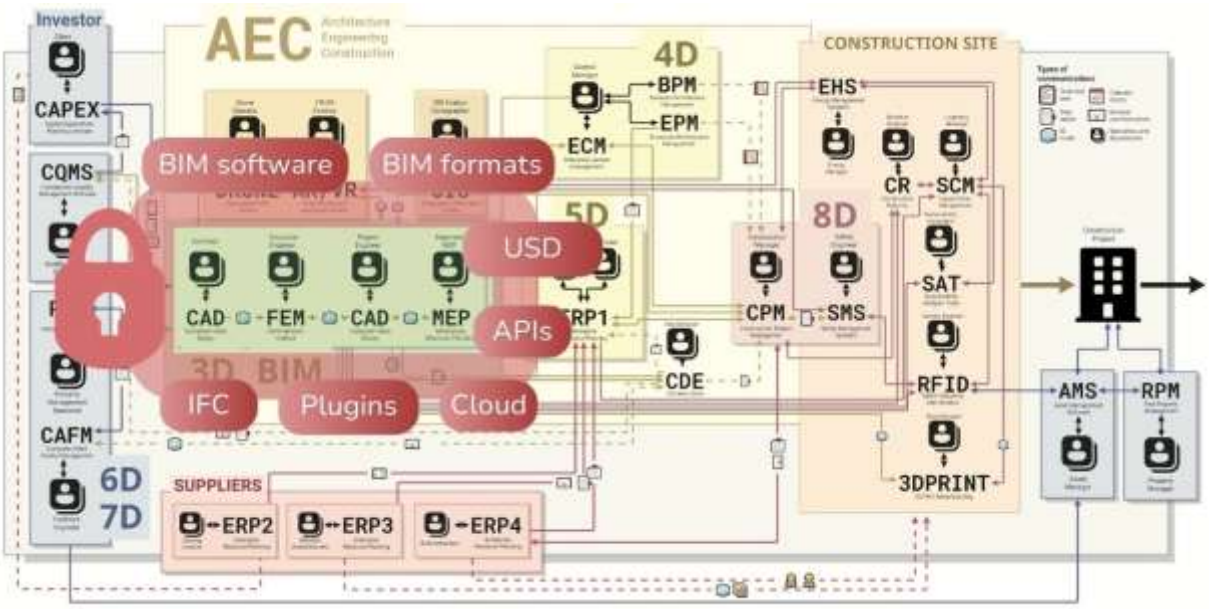


Abb. 6.1-3 CAD- (BIM-) Datenbanken sind nach wie vor eines der letzten geschlossenen Systeme für IT-Abteilungen und Datenverwalter im Ökosystem der Baubranche.

> Wenn sich herausstellt, dass BIM eher ein Mittel zur Kommerzialisierung von Datenbanken als ein vollwertiges Datenbankmanagement-Tool ist, stellt sich eine logische Frage: Wie können wir die Kontrolle über die Daten zurückgewinnen? Die Antwort liegt in der Verwendung offener Datenstrukturen, bei denen der Nutzer und nicht der Softwareanbieter zum Eigentümer der Informationen wird.

Anwender und Lösungsentwickler im Baugewerbe werden sich, wie ihre Kollegen in anderen Branchen auch, unweigerlich von der vagen Terminologie der Softwareanbieter lösen, die in den letzten 30 Jahren vorherrschte, und sich auf die Schlüsselaspekte der Digitalisierung - "Daten" und "Prozesse" - konzentrieren.

In den späten 1980er Jahren wurde der Schwerpunkt der digitalen Entwicklung im Bauwesen auf den Datenzugang und das Projektinformationsmanagement gelegt. Im Laufe der Zeit hat sich der Schwerpunkt jedoch verlagert. Anstatt transparente und zugängliche Ansätze für die Arbeit mit Daten zu entwickeln, wurden das IFC-Format und das offene BIM-Konzept aktiv gefördert, um die Aufmerksamkeit der Fachleute von den Themen des Projektdatenmanagements abzulenken.

Das Aufkommen des offenen Formats IFC in der Bauindustrie

Das so genannte offene Format IFC (Industry Foundation Classes) ist ein Standard, der die Interoperabilität zwischen verschiedenen CAD-Systemen (BIM -) gewährleisten soll. Seine Entwicklung fand im Rahmen von Organisationen statt, die von großen CAD-Anbietern gegründet und kontrolliert wurden. Auf der Grundlage des IFC-Formats entwickelten zwei CAD- -Unternehmen im Jahr 2012 das Marketingkonzept OPEN BIM[63].

> **IFC** (Industry Foundation Classes) ist ein offener Standard für den Datenaustausch in der Bauindustrie, der die Interoperabilität zwischen verschiedenen CAD - (BIM-) Systemen gewährleisten soll.

Open BIM - Konzept beinhaltet die Arbeit mit Informationen aus CAD-Datenbanken und den Austausch von Informationen zwischen Systemen über ein offenes Format für den Austausch von CAD-Daten - IFC.

> *Das Open BIM-Programm ist eine Marketingkampagne, die von... [1 CAD-Anbieter],... [2 CAD-Anbieter] und anderen Unternehmen initiiert wurde, um die globale, koordinierte Förderung des OPEN BIM-Konzepts in der gesamten AEC-Branche zu fördern und zu erleichtern, wobei den Programmteilnehmern eine einheitliche Kommunikation und ein gemeinsames Branding zur Verfügung stehen.*
>
> - Von der Website des CAD-Anbieters, OPEN BIM-Programm, 2012 [113]

IFC wurde in den späten 1980er Jahren von der Technischen Universität München aus dem Maschinenbauformat STEP adaptiert und später von einem großen Konstruktionsunternehmen und einem bedeutenden CAD- Anbieter zur Gründung der IAI (Industry Alliance for Interoperability) im Jahr 1994 [114] angemeldet (Abb. 6.1-4). Das Format IFC wurde entwickelt, um die Interoperabilität zwischen verschiedenen CAD-

Systemen zu gewährleisten, und basierte auf den Grundsätzen des Maschinenbau-Formats STEP, das seinerseits aus dem IGES-Format hervorging, das 1979 von einer Gruppe von CAD-Anwendern und -Anbietern mit Unterstützung des NIST (The National Institute of Standards and Technology) und des US-Verteidigungsministeriums [115] entwickelt wurde.

Die komplexe Struktur von IFC, seine enge Abhängigkeit vom geometrischen Kern sowie Unterschiede in der Implementierung des Formats durch verschiedene Softwarelösungen haben jedoch zu vielen Problemen bei der praktischen Anwendung geführt. Ähnliche Schwierigkeiten - Detailverluste, Einschränkung der Genauigkeit und die Notwendigkeit, Zwischenformate zu verwenden - hatten die Fachleute des Maschinenbaus früher bei der Arbeit mit den Formaten IGES und STEP, aus denen IFC hervorgegangen ist.

Abb. 6.1-4 Karte der Verbindungen zwischen Entwicklungsteams und CAD-Produkten (BIM) [116].

> Im Jahr 2000 veröffentlicht derselbe CAD -Anbieter, der das Format IFC registriert und die Organisation IAI (später bS) gegründet hat, das Whitepaper "Integrated Design and Manufacturing: Benefits and Rationale" [65]. In dem Papier wird betont, wie wichtig es ist, beim Austausch zwischen Programmen innerhalb desselben Systems die volle Datengranularität beizubehalten, ohne neutrale Formate wie IGES oder STEP [identisch mit IFC] zu verwenden. Stattdessen wurde vorgeschlagen, dass die Anwendungen direkten Zugriff auf die zugrunde liegende CAD-Datenbank haben sollten, um einen Verlust der Genauigkeit der Informationen zu vermeiden.

Im Jahr 2002 kauft derselbe CAD-Anbieter das Produkt parametrische Stückliste (Abb. 3.1-18, mehr Details in des dritten Teils) und bildet auf dessen Grundlage das BIM-Konzept. Infolgedessen werden für den Austausch von Bauprojektdaten nur geschlossene CAD-Formate oder das IFC-Format (STEP) verwendet, über dessen Grenzen der CAD-Anbieter selbst im Jahr 2000 schrieb, als er dieses Format in die Bauindustrie einführte.

Eine detaillierte Geschichte der Interaktion von mehr als 700 Entwicklungsteams, die an der Schaffung von Werkzeugen zur Erstellung und Verarbeitung von Konstruktionsdaten beteiligt waren, wird in der Karte "The Evolution of CAD (BIM)" [116] [116].

Die offene Form IFC besteht aus einer geometrischen Beschreibung der Konstruktionselemente und einer Beschreibung der Metainformationen. Für die Darstellung der Geometrie im IFC-Format werden verschiedene Methoden verwendet, z. B. CSG und Swept Solids: Die parametrische Darstellung BREP hat sich jedoch als führender Standard für die Übertragung der Elementgeometrie im IFC-Format durchgesetzt, da dieses Format beim Export aus CAD- (BIM-)-Programmen unterstützt wird und beim IFC-Import zurück in CAD-Programme eine mögliche Bearbeitung der Elemente ermöglicht.

IFC Formatproblem je nach geometrischem Kern

In den meisten Fällen, wenn die Geometrie in IFC parametrisch definiert ist (BREP), ist es unmöglich, geometrische Eigenschaften wie Volumen oder Fläche von Projektelementen nur mit einer IFC-Datei zu visualisieren oder abzurufen, da für die Arbeit mit und die Visualisierung der Geometrie in diesem Fall ein Geometriekern (Abb. 6.1-5) erforderlich ist, der zunächst fehlt.

Geometry kernel ist eine Softwarekomponente, die grundlegende Algorithmen für die Erstellung, Bearbeitung und Analyse von geometrischen Objekten in CAD (CAD), BIM und anderen technischen Anwendungen bietet. Er ist verantwortlich für die Konstruktion von 2D- und 3D-Geometrie sowie für Operationen wie boolesche Operationen, Glättung, Überschneidungen, Transformationen und Visualisierung.

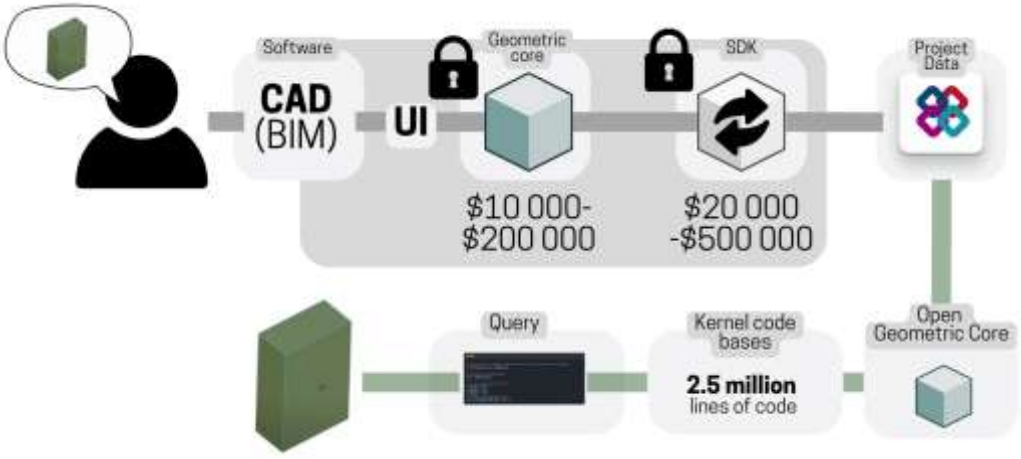


Abbildung 6.1-5 Die Erstellung von Geometrien mit CAD- Software erfolgt heute über proprietäre Geometrie-Kernel und SDKs, die oft nicht von CAD-Anbietern stammen.

Jedes CAD-Programm und jedes Programm, das mit parametrischen oder IFC-Formaten arbeitet, hat einen eigenen oder gekauften geometrischen Kern. Und wenn es mit primitiven Elementen im IFC -BREP -Format keine Probleme gibt und diese Elemente in Programmen mit verschiedenen geometrischen Kernen ähnlich dargestellt werden können, gibt es neben den Problemen mit verschiedenen Engines von geometrischen Kernen auch genügend Elemente, die ihre eigenen Besonderheiten für eine korrekte Darstellung haben. Diese Problematik wird in der internationalen Studie "A reference study of IFC software support", veröffentlicht 2019 [117], ausführlich diskutiert.

> *Dieselben standardisierten Datensätze führen zu widersprüchlichen Ergebnissen, wobei nur wenige gemeinsame Muster gefunden wurden, und es wurden ernsthafte Probleme bei der Unterstützung der Norm festgestellt [IFC], was wahrscheinlich auf die sehr hohe Komplexität des Standarddatenmodells zurückzuführen ist. Die Standards selbst tragen hier eine Mitschuld, da sie oft einige Details undefiniert lassen, mit hohen Freiheitsgraden und verschiedenen möglichen Interpretationen. Sie erlauben eine hohe Komplexität bei der Organisation und Speicherung von Objekten, was einem effektiven universellen Verständnis, eindeutigen Implementierungen und einer konsistenten Datenmodellierung nicht förderlich ist [117].*
>
> - Referenzstudie zur IFC-Softwareunterstützung, 2021

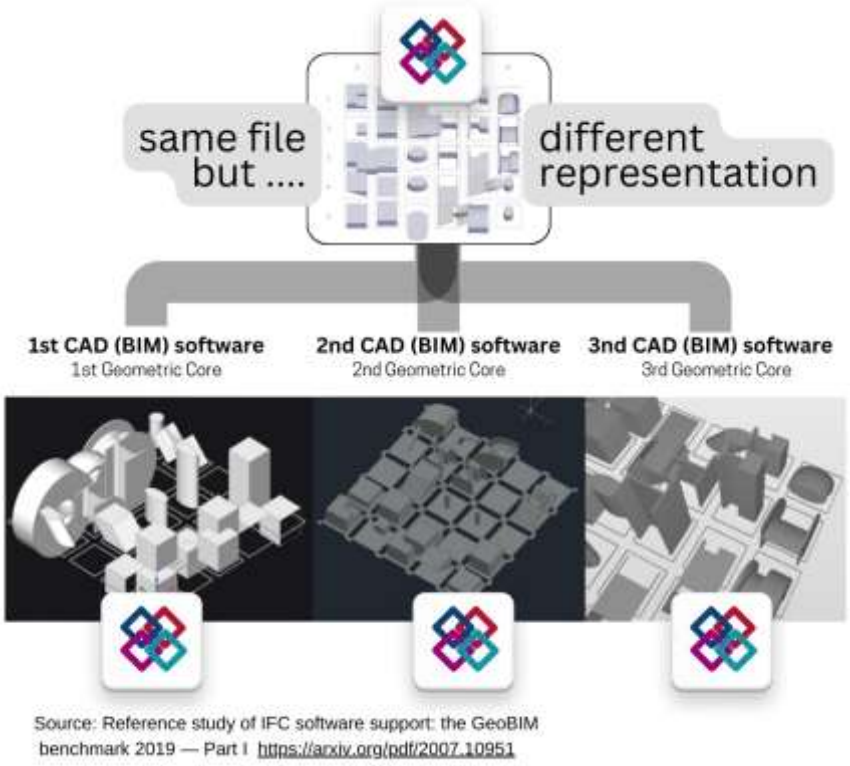


Abb. 6.1-6 Verschiedene geometrische Kerne ergeben unterschiedliche Darstellungen derselben parametrisch beschriebenen Geometrie (basierend auf [117]).

Das korrekte Verständnis "bestimmter Bestimmungen" ist den zahlenden Mitgliedern spezieller Organisationen zugänglich, die IFC entwickeln. Wer also Zugang zu wichtigem Wissen über bestimmte Eigenschaften von IFC haben will, wird versuchen, mit großen CAD- Anbietern zusammenzuarbeiten oder durch eigene Forschung zu einer qualitativen Betrachtung der Eigenschaften zu gelangen

> *Sie stolpern über eine Frage zum Import und Export von Daten über das IFC-Format und fragen andere Anbieter: "Warum gibt es in der IFC-Datei Informationen über die parametrische Übertragung von Räumen? In der offenen Spezifikation steht nichts darüber". Antwort von "sachkundigeren" europäischen Anbietern: "Ja, es wird nicht gesagt, aber es ist erlaubt".*
>
> - Aus dem Interview mit dem CAD 2021-Entwickler [118]

IFC beschreibt die Geometrie durch parametrische Primitive, enthält aber keinen eingebetteten Kern - seine Rolle wird vom CAD-Programm erfüllt, das die Geometrie durch den Geometriekern kompiliert. Der Geometriekern führt die mathematischen Berechnungen durch und definiert die Schnittpunkte, während die IFC nur die Daten für ihre Interpretation liefert. Wenn die IFC falsche Flächen enthält, können verschiedene Programme mit unterschiedlichen Geometriekernen diese entweder ignorieren oder Fehler produzieren, je nach Kernel.

> Um mit dem IFC-Format arbeiten zu können, ist es daher notwendig, die Hauptfrage zu beantworten, auf die es schwierig ist, eine eindeutige Antwort zu finden: Welches Werkzeug mit welchem geometrischen Kern sollte verwendet werden, um die Datenqualität zu erhalten, die das Projekt ursprünglich im CAD-Programm hatte, aus dem IFC gewonnen wurde?

Probleme mit der Datenqualität und die Komplexität des Formats IFC lassen eine direkte Nutzung der Konstruktionsdaten für die Prozessautomatisierung, Analyse und Datenverarbeitung nicht zu, was die Entwickler oft zu der unvermeidlichen Notwendigkeit führt, geschlossene CAD -Lösungen mit "Qualitäts"-Zugang zu den Daten zu verwenden[63], worüber der Hersteller selbst schrieb, der IFC 1994 registrierte [65].

Alle Besonderheiten des Mappings und der Generierung von IFC -Parametern im Geometriekern können nur von großen Entwicklerteams realisiert werden, die Erfahrung in der Arbeit mit Geometriekernen haben. Daher kommt die derzeitige Praxis der IFC-Formatbesonderheiten und -komplexität in erster Linie den CAD-Anbietern zugute und hat viel mit der Strategie der großen Softwareanbieter "übernehmen, erweitern, zerstören" gemein, wenn die wachsende Komplexität des Standards tatsächlich Barrieren für kleine Marktteilnehmer schafft [94].

Die Strategie der großen Anbieter bei einer solchen Strategie kann darin bestehen, offene Standards zu adaptieren, proprietäre Erweiterungen und Funktionen hinzuzufügen, um eine Abhängigkeit der Nutzer von ihren Produkten zu schaffen und so die Konkurrenz zu verdrängen.

> Das IFC-Format, das als universelle Brücke zwischen verschiedenen CAD- (BIM-)-Systemen gedacht ist, erfüllt in Wirklichkeit die Rolle eines Indikators für Kompatibilitätsprobleme zwischen den geometrischen Kernen verschiedener CAD-Plattformen, ähnlich wie das STEP-Format, aus dem es ursprünglich hervorgegangen ist.

Infolgedessen ist heute eine vollständige und qualitativ hochwertige Implementierung der IFC Ontologie für große CAD-Anbieter machbar, die erhebliche Ressourcen investieren können, um alle Entitäten und deren Zuordnung zu ihrem eigenen internen Geometriekern zu unterstützen, der für IFC als Standard nicht existiert. Große Anbieter sind auch in der Lage, technische Details von Merkmalen untereinander abzustimmen, die selbst den aktivsten Teilnehmern an den Organisationen zur Entwicklung des IFC-Formats möglicherweise nicht zur Verfügung stehen.

> Für kleine unabhängige Teams und Open-Source-Projekte, die sich um die Entwicklung interoperabler Formate bemühen, wird das Fehlen eines eigenen Geometriekerns zu einem ernsten Problem. Ohne ihn ist es praktisch unmöglich, all die verschiedenen Feinheiten und Nuancen zu berücksichtigen, die mit dem plattformübergreifenden Datenaustausch verbunden sind.

Mit der Entwicklung des parametrischen IFC-Formats und des offenen BIM-Konzepts hat sich die Diskussion in der Baubranche über die Rolle der Ontologie und der Semantik im Daten- und Prozessmanagement intensiviert.

Erscheinungsbild im Aufbau des Themas Semantik und Ontologie

Dank der Ideen des semantischen Internets Ende der 1990er und der Bemühungen von Organisationen, die an der Entwicklung des IFC-Formats beteiligt waren, gehören Semantik und Ontologien zu den Schlüsselelementen der Normung, die bis Mitte der 2020er Jahre in der Baubranche diskutiert wird.

> Semantische Technologien dienen der Vereinheitlichung, Standardisierung und Veränderung großer Mengen heterogener Daten und der Durchführung komplexer Suchvorgänge.

OWL (Web Ontology Language), dargestellt als RDF-Graphen -triplets (Resource Description Framework) (Abb. 6.1-7), wird zur Speicherung semantischer Daten verwendet. OWL bezieht sich auf Graphen-Datenmodelle, auf deren Typen wir im Kapitel "Datenmodelle: Datenbeziehungen und Beziehungen zwischen Elementen" näher eingegangen sind.

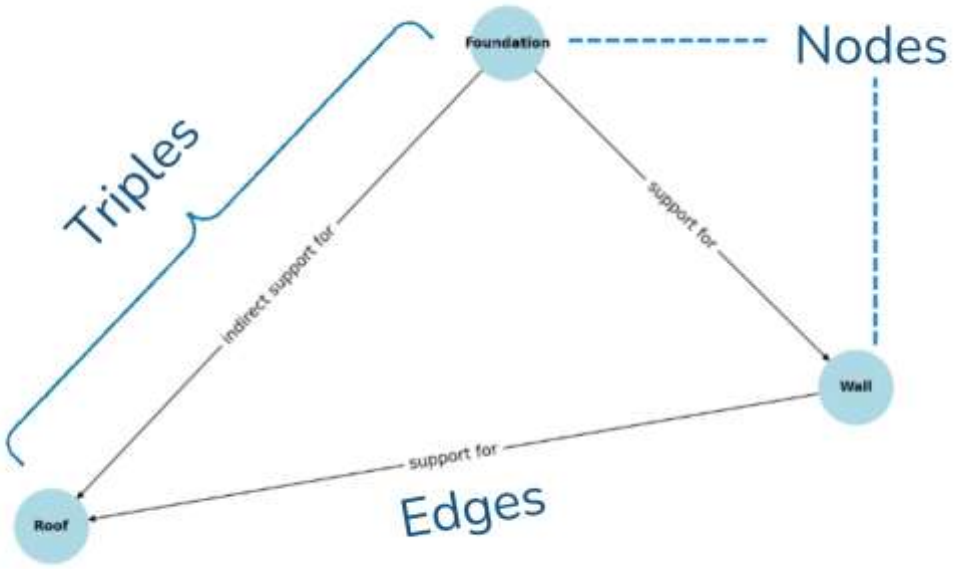


Abb. 6.1-7 RDF-Datenmodell: Knoten, Kanten und Tripel zur Veranschaulichung der Beziehungen zwischen den Bausteinen.

Theoretisch können durch die logische Inferenz von Risonern (automatische logische Inferenzsoftware) neue Aussagen aus Ontologien abgeleitet werden. Wenn zum Beispiel in der Ontologie des Bauwesens festgehalten ist, dass "ein Fundament eine Stütze für eine Wand ist" und "eine Wand eine Stütze für ein Dach ist" (Abb. 6.1-7), kann der Risoner automatisch ableiten, dass "ein Fundament eine Stütze für ein Dach ist".

Ein solcher Mechanismus ist für die Optimierung der Datenanalyse nützlich, da er es vermeidet, alle Abhängigkeiten explizit vorzuschreiben. Er schafft jedoch kein neues Wissen, sondern identifiziert und strukturiert lediglich bereits bekannte Fakten.

> Die Semantik schafft an sich keine neue Bedeutung oder neues Wissen und ist anderen Datenspeicherungs- und -verarbeitungstechnologien in dieser Hinsicht nicht überlegen. Die Darstellung von Daten aus relationalen Datenbanken als Triplets macht sie nicht aussagekräftiger. Das Ersetzen von Tabellen durch Graphenstrukturen mag für die Vereinheitlichung von Datenmodellen, den einfachen Abruf und die sichere Bearbeitung nützlich sein, aber es macht die Daten nicht "intelligenter" - der Computer beginnt nicht, ihren Inhalt besser zu verstehen.

Logische Beziehungen in Daten können ohne komplexe semantische Technologien organisiert werden (Abb. 6.1-8). Traditionelle relationale Datenbanken (SQL) sowie die Formate CSV oder XLSX ermöglichen den Aufbau ähnlicher Abhängigkeiten. In einer spaltenbasierten Datenbank können Sie beispielsweise ein Feld "Dachunterstützung" hinzufügen und das Dach automatisch mit dem Fundament verknüpfen, wenn Sie eine Wand erstellen. Dieser Ansatz wird ohne die Verwendung von RDF, OWL, Graphen oder Risonern umgesetzt und bleibt eine einfache und effiziente Lösung für die Speicherung und Analyse von Daten.

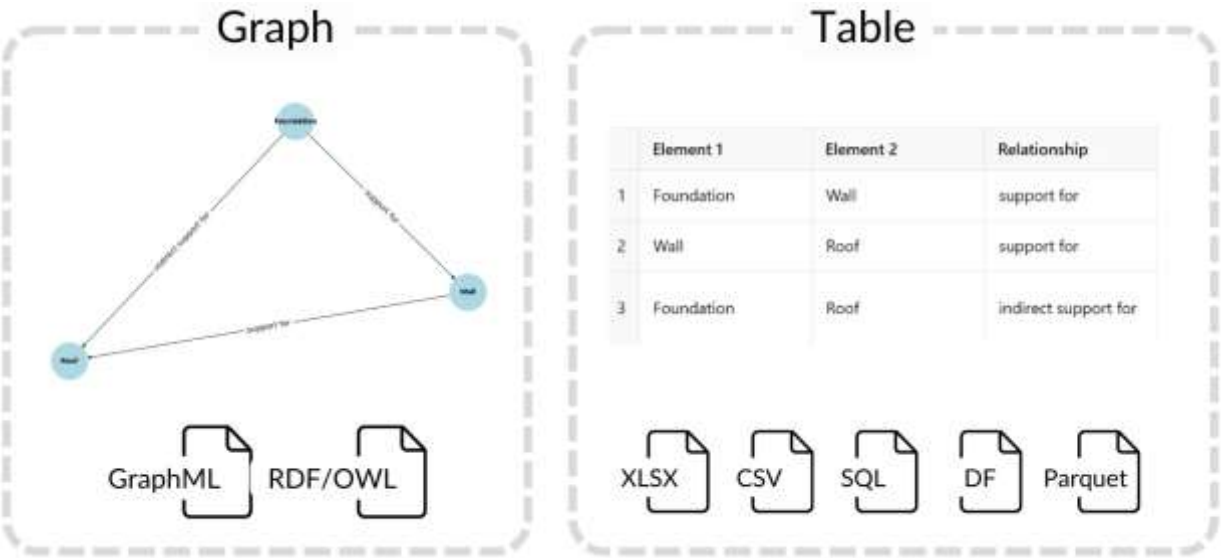


Abb. 6.1-8 Vergleich von Diagramm- und Tabellendatenmodellen zur Darstellung der gleichen logischen Beziehungen.

Die Entscheidung einiger großer Bauunternehmen und der Organisation zur Entwicklung des IFC-Formats [94], dem Konzept des semantischen Webs zu folgen, das in den späten 1990er Jahren vielversprechend erschien, hat die Entwicklung von Normen in der Bauindustrie erheblich beeinflusst.

Das Paradoxe ist jedoch, dass das Konzept des semantischen Webs, das ursprünglich für das Internet gedacht war, nicht einmal in seiner ursprünglichen Umgebung weit verbreitet ist. Trotz der Entwicklung von RDF und OWL ist das vollwertige semantische Web nicht in seiner ursprünglichen Konzeption erschienen, und seine Schaffung ist bereits unwahrscheinlich.

Warum semantische Technologien die Erwartungen in der Bauindustrie nicht erfüllen

Andere Branchen sind mit den Grenzen der Technologien zur Nutzung der Semantik konfrontiert. In der Spieleindustrie haben sich Versuche, Spielobjekte und ihre Interaktionen durch Ontologien zu beschreiben, aufgrund der hohen Änderungsdynamik als unwirksam erwiesen. Infolgedessen wurden einfachere Datenformate wie XML und JSON sowie algorithmische Lösungen bevorzugt. Ähnlich war die Situation im Immobiliensektor: Aufgrund regionaler Unterschiede in der Terminologie und häufiger Marktänderungen erwies sich die Verwendung von Ontologien als zu komplex, während einfache Datenbanken und Standards wie RETS [119] den Herausforderungen des Datenaustauschs besser gewachsen waren.

> Technische Schwierigkeiten, wie die Komplexität des Markups, der hohe Arbeitsaufwand für die Unterstützung und die geringe Motivation der Entwickler, verlangsamten die Einführung des semantischen Webs und in anderen Bereichen der Wirtschaft. RDF (Resource Description Framework) wurde nicht zu einem Massenstandard, und Ontologien erwiesen sich als zu komplex und wirtschaftlich nicht gerechtfertigt.

Die ehrgeizige Idee, ein globales semantisches Web zu schaffen, konnte daher nicht verwirklicht werden. Obwohl einige Elemente der Technologie, wie z. B. Ontologien und SPARQL, ihren Weg in Unternehmenslösungen gefunden haben, wurde das ursprüngliche Ziel, eine einzige umfassende Datenstruktur zu schaffen, nicht erreicht.

Das Konzept eines Internets, in dem Computer in der Lage sind, Inhalte zu verstehen, hat sich als technisch schwierig und kommerziell unrentabel erwiesen. Aus diesem Grund haben die Unternehmen, die diese Idee unterstützten, ihre Verwendung schließlich auf einzelne nützliche Werkzeuge reduziert und RDF und OWL für hochspezialisierte Unternehmensbedürfnisse und nicht für das Internet als Ganzes übrig gelassen. Eine Analyse der Google Trends (Abb. 6.1-9) über die letzten 20 Jahre lässt vermuten, dass es für das semantische Web möglicherweise keine Perspektiven mehr gibt.

> *Es gibt keinen Grund, Entitäten unnötig zu multiplizieren. Wenn es mehrere logisch konsistente Erklärungen für ein Phänomen gibt, die es gleichermaßen gut erklären, sollte man, wenn alle anderen Dinge gleich sind, die einfachste von ihnen vorziehen.*
>
> Rasiermesser

Hier stellt sich eine logische Frage: Warum sollte man überhaupt Tripletts, Risoner und SPARQL in den Projekten verwenden, wenn man Daten mit gängigen strukturierten Abfragen (SQL, Pandas, Apache®) verarbeiten kann? In Unternehmensanwendungen ist SQL der Standard für die Arbeit mit Datenbanken. SPARQL hingegen erfordert komplexe Graphenstrukturen und spezialisierte Software und stößt, den Trends bei Google zufolge, nicht auf das Interesse von Entwicklern.

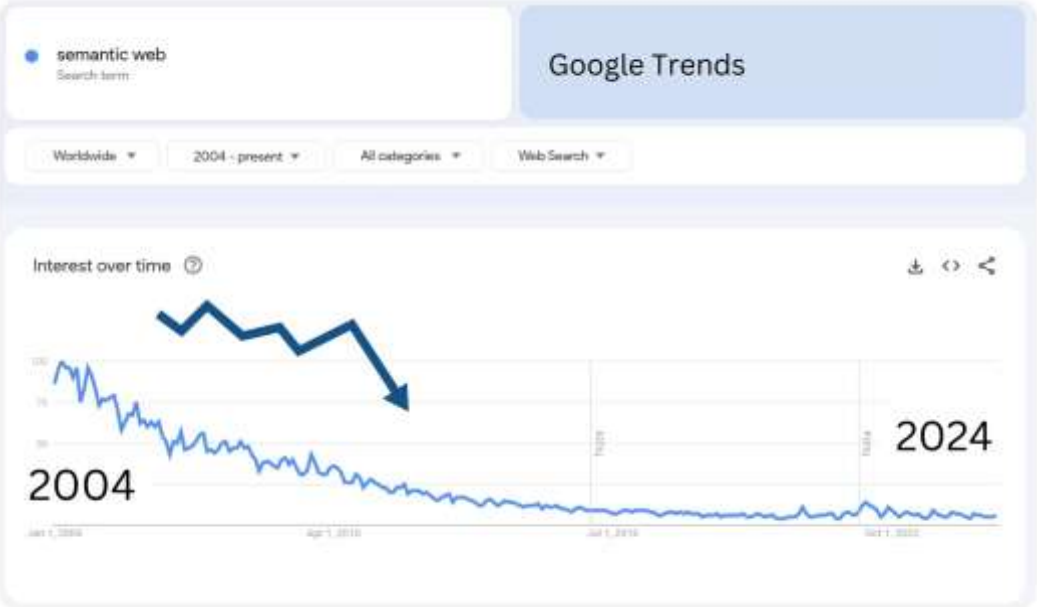


Abb. 6.1-9 Interesse an "semantischen Internet"-Abfragen laut Google-Statistiken.

Graphdatenbanken und Klassifikationsbäume können in einigen Fällen nützlich sein, aber ihre Anwendung

ist für die meisten alltäglichen Aufgaben nicht immer gerechtfertigt. Daher sind die Erstellung von Wissensgraphen und die Verwendung von semantischen Webtechnologien nur dann sinnvoll, wenn es darum geht, Daten aus verschiedenen Quellen zusammenzuführen oder komplexe logische Schlussfolgerungen zu ziehen.

> Der Wechsel von Tabellen zu grafischen Datenmodellen verbessert die Suche und vereinheitlicht den Informationsfluss, macht die Daten aber nicht aussagekräftiger für Maschinen. Die Frage ist nicht, ob semantische Technologien eingesetzt werden sollten, sondern wo sie wirklich einen Unterschied machen. Bevor Sie Ontologie, Semantik und Graphdatenbanken in Ihrem Unternehmen einführen, sollten Sie herausfinden, welche Unternehmen diese Technologien bereits erfolgreich einsetzen und wo sie gescheitert sind.

Trotz ehrgeiziger Erwartungen haben sich die semantischen Technologien nie zu einer Universallösung für die Strukturierung von Daten in der Bauwirtschaft entwickelt. In der Praxis haben diese Technologien nicht zu einer Universallösung geführt, sondern nur neue Komplexitäten hinzugefügt. Diese Bemühungen erinnern an die nicht realisierten Ambitionen des Konzepts des semantischen Internets, bei dem die Erwartungen die Realität weit übertrafen.

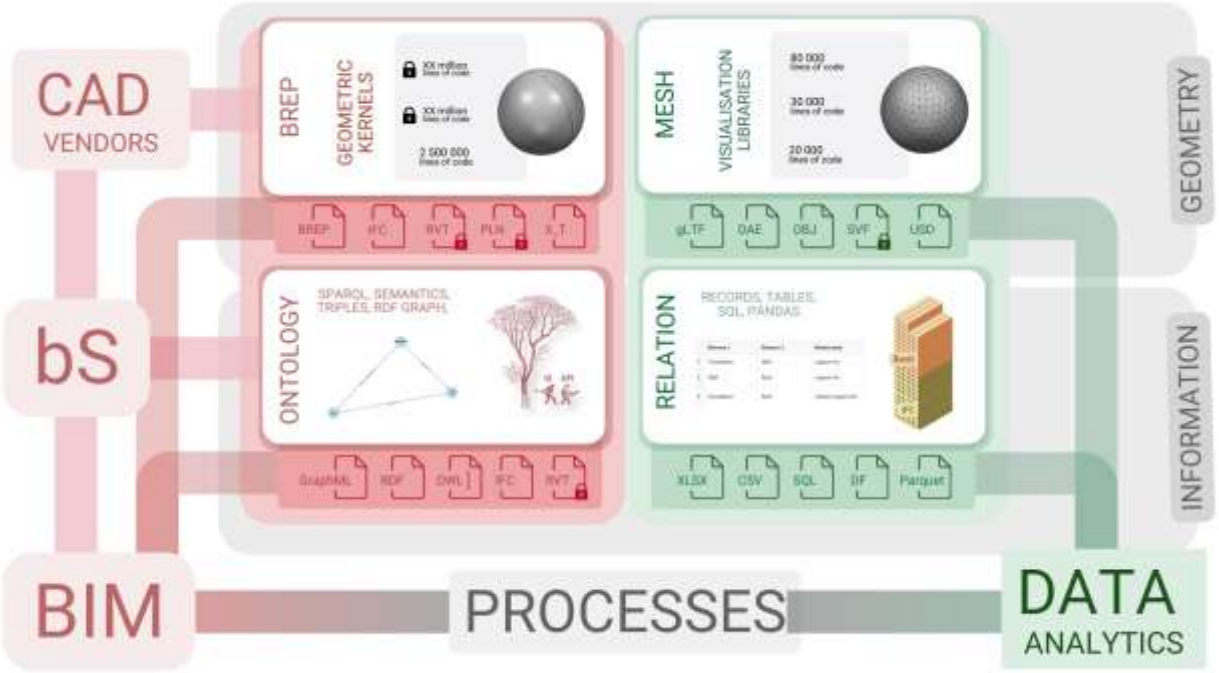


Abb. 6.1-10 Geometrie und Information in Bauprozessen: von komplexen CAD- und BIM-Systemen zu vereinfachten Daten für die Analytik.

Während in der IT-Branche die Misserfolge des semantischen Webs durch das Aufkommen neuer Technologien (Big Data, IoT, maschinelles Lernen, AR/VR) ausgeglichen wurden, gibt es in der Baubranche keine solchen Gelegenheiten.

Zusätzlich zu den Herausforderungen, die die Verwendung von Konzepten zur Kommunikation von Datenbeziehungen zwischen Projektelementen mit sich bringt, bleibt ein grundlegendes Problem bestehen - die

Verfügbarkeit dieser Daten selbst. Die Baubranche wird immer noch von geschlossenen Systemen beherrscht, was die Arbeit mit Daten, den Informationsaustausch und die Verbesserung der Prozesseffizienz erschwert.

Es ist die Geschlossenheit der Daten, die zu einem der Haupthindernisse für die Entwicklung digitaler Lösungen im Bauwesen wird. Anders als in der IT-Branche, wo offene und harmonisierte Datenformate zum Standard geworden sind, verwendet im CAD-Sektor (BIM) jede Software ihr eigenes Format, wodurch geschlossene Ökosysteme entstehen und die Nutzer künstlich eingeschränkt werden.

KAPITEL 6.2.
GESCHLOSSENE PROJEKTFORMATE UND INTEROPERABILITÄTSPROBLEME

Geschlossene Daten und sinkende Produktivität: die Sackgasse der CAD-Branche (BIM)

Die proprietäre Natur der CAD -Systeme hat dazu geführt, dass jedes Programm sein eigenes, einzigartiges Datenformat hat, das entweder geschlossen und von außen unzugänglich ist - RVT, PLN, DWG, NDW, NWD, SKP, oder in halbstrukturierter Form durch einen ziemlich komplexen Konvertierungsprozess verfügbar ist - JSON, XML (CPIXML), IFC, STEP und ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, usw...

Die verschiedenen Datenformate, in denen dieselben Daten über dieselben Projekte gespeichert werden können, unterscheiden sich nicht nur in ihrer Struktur, sondern enthalten auch unterschiedliche Versionen des internen Markups, was die Entwickler berücksichtigen müssen, um die Anwendungskompatibilität zu gewährleisten. So lässt sich beispielsweise ein CAD-Format aus dem Jahr 2025 in einem CAD-Programm aus dem Jahr 2026 öffnen, aber dasselbe Projekt wird sich niemals in allen Versionen des CAD-Programms öffnen lassen, die möglicherweise vor 2025 verfügbar waren.

> Da ein Softwareanbieter im Baugewerbe keinen direkten Zugang zu den Datenbanken anbietet, erstellt er oft ein eigenes Format und eigene Werkzeuge, die ein Fachmann (Konstrukteur oder Datenverwalter) verwenden muss, um auf Daten zuzugreifen, sie zu importieren und zu exportieren.

Infolgedessen erhöhen die Anbieter von Basis-CAD (BIM) und verwandten Lösungen (z. B. ERP/PMIS)) ständig die Preise für die Nutzung der Produkte, und die normalen Nutzer sind gezwungen, in jeder Phase des Datentransfers nach Formaten [63] eine "Provision" zu zahlen: für die Verbindung, den Import, den Export und die Arbeit mit Daten, die die Nutzer selbst erstellt haben.

Die Kosten für den Zugriff auf Daten im Cloud-Speicher von beliebten CAD - (BIM-)-Produkten werden im Jahr 2025 $1 pro Transaktion erreichen [120], und Abonnements für Bau-ERP -Produkte für mittelständische Unternehmen erreichen fünf- und sechsstellige Beträge pro Jahr [121].

> Das Wesen moderner Bausoftware besteht darin, dass nicht die Automatisierung oder Effizienzsteigerung, sondern die Fähigkeit der Ingenieure, eine bestimmte hochspezialisierte Software zu verstehen, die Qualität und die Kosten der Datenverarbeitung von Bauprojekten sowie die Gewinne und das langfristige Überleben von Unternehmen, die Bauprojekte durchführen, beeinflusst.

Der fehlende Zugang zu den Datenbanken CAD -systemen, die in Dutzenden von anderen Systemen und Hunderten von Prozessen verwendet werden [63], und die daraus resultierende mangelnde Qualität der Kommunikation zwischen den einzelnen Fachleuten hat dazu geführt, dass die Bauindustrie zu einem der ineffizientesten Wirtschaftszweige in Bezug auf die Produktivität geworden ist [44].

In den letzten 20 Jahren der CAD- (BIM-)-Entwurfsanwendungen, des Aufkommens neuer Systeme (ERP), neuer Bautechnologien und -materialien ist die Produktivität des gesamten Baugewerbes um 20% gesunken (Abb. 2.2-1), während die Gesamtproduktivität aller Wirtschaftszweige, die keine größeren Probleme beim Zugriff auf Datenbanken und marketingähnliche BIM-Konzepte haben, um 70% gestiegen ist (96% in der verarbeitenden Industrie) [122].

Abb. 6.2-1 Aufgrund der Isolierung und Komplexität der Projektdaten, von denen in der Bauindustrie Dutzende von Abteilungen und Hunderte von Prozessen abhängen, ist die Geschwindigkeit der Entscheidungsfindung um ein Vielfaches langsamer als in anderen Branchen.

Es gibt aber auch vereinzelte Beispiele für alternative Ansätze zur Schaffung von Interoperabilität zwischen CAD-Lösungen. Europas größtes Bauunternehmen zeigt mit dem 2018 gestarteten SCOPE-Projekt [123], wie es möglich ist, über die klassische Logik von CAD- (BIM-) Systemen hinauszugehen. Anstatt zu versuchen, IFC zu unterwerfen oder sich auf proprietäre Geometriekerne zu verlassen, nutzen die SCOPE-Entwickler APIs und SDKs Reverse Engineering, um Daten aus verschiedenen CAD-Programmen zu extrahieren, sie in neutrale Formate wie OBJ oder CPIXML zu konvertieren, die auf dem einzigen Open Source Geometriekernel OCCT basieren, und sie anschließend auf Hunderte von Geschäftsprozessen von Bau- und Planungsunternehmen anzuwenden. Trotz der Fortschrittlichkeit der Idee stoßen solche Projekte jedoch an die Grenzen und die Komplexität freier Geometrie-Kernel und bleiben Teil geschlossener Ökosysteme eines Unternehmens, die die Logik von Monovendor-Lösungen reproduzieren.

Aufgrund der Beschränkungen geschlossener Systeme und der unterschiedlichen Datenformate sowie des Mangels an wirksamen Werkzeugen für ihre Vereinheitlichung sehen sich Unternehmen, die mit CAD-Formaten arbeiten müssen, mit der Anhäufung erheblicher Datenmengen mit unterschiedlichem Strukturierungs- und Schließungsgrad konfrontiert. Diese Daten werden nicht richtig genutzt und verschwinden in Archiven, wo sie für immer vergessen und ungenutzt bleiben.

> Daten, die mit erheblichem Aufwand in der Entwurfsphase gewonnen wurden, sind aufgrund ihrer Komplexität und ihres geschlossenen Charakters für die weitere Verwendung unzugänglich.

Infolgedessen sahen sich die Entwickler in der Baubranche in den letzten 30 Jahren immer wieder mit demselben Problem konfrontiert: Jedes neue geschlossene Format oder jede proprietäre Lösung führt dazu, dass eine Integration mit bestehenden offenen und geschlossenen CAD-Systemen erforderlich ist. Diese ständigen Versuche, die Interoperabilität zwischen verschiedenen CAD- und BIM-Lösungen zu gewährleisten, verkomplizieren das Datenökosystem nur, anstatt zu seiner Vereinfachung und Standardisierung beizutragen.

Der Mythos der Interoperabilität zwischen CAD-Systemen

Während Mitte der 1990er Jahre die Hauptrichtung der Interoperabilitätsentwicklung im CAD-Umfeld das Aufbrechen des proprietären DWG-Formats war - was im Sieg der Open-DWG-Allianz [75] und der tatsächlichen Öffnung des populärsten Zeichnungsformats für die gesamte Baubranche gipfelte, hat sich der Schwerpunkt bis Mitte der 2020er Jahre verschoben. Ein neuer Trend gewinnt in der Baubranche an Fahrt: Zahlreiche Entwicklungsteams konzentrieren sich darauf, sogenannte "Brücken" zwischen geschlossenen CAD-Systemen (Closed BIM), IFC-Format und offenen Lösungen (Open BIM) zu schaffen. Die meisten dieser Initiativen basieren auf der Verwendung des IFC-Formats und des OCCT-Geometriekerns, die eine technische Brücke zwischen unterschiedlichen Plattformen bilden. Dieser Ansatz wird als vielversprechender Weg angesehen, der den Datenaustausch und die Interoperabilität von Softwaretools erheblich verbessern kann.

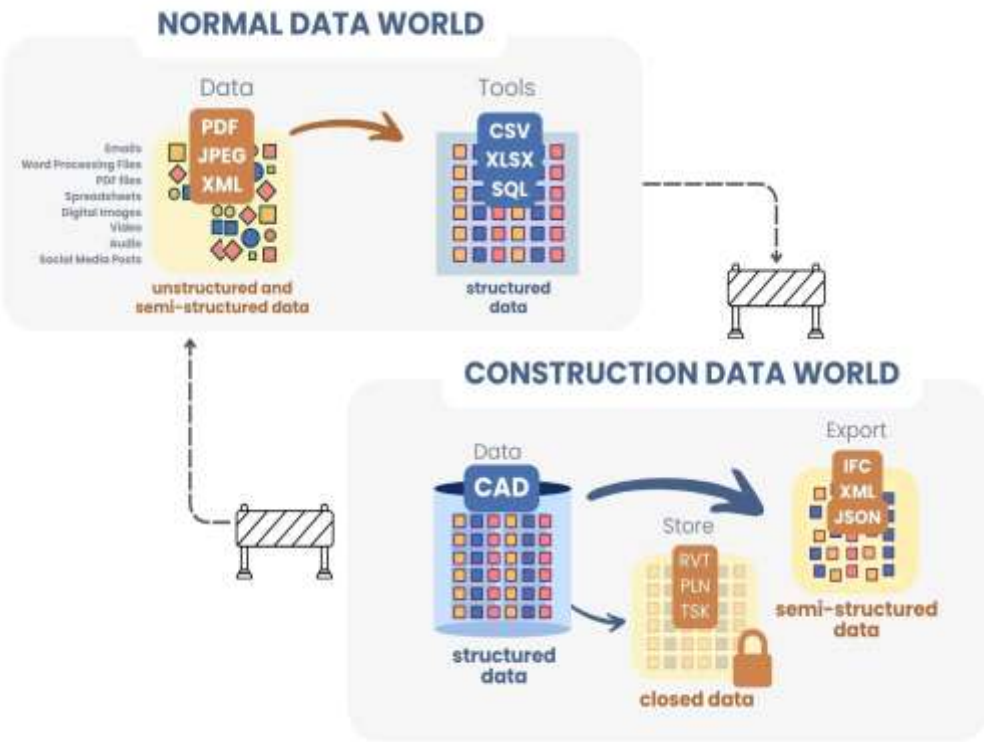


Abbildung 6.2-2 Während andere Branchen mit offenen Daten arbeiten, muss die Bauindustrie mit geschlossenen oder lose strukturierten CAD-Formaten arbeiten (BIM).

Ein solcher Ansatz hat historische Parallelen. In den 2000er Jahren versuchten Entwickler, die Dominanz des größten Anbieters von Grafikeditoren (2D World) zu überwinden, indem sie eine nahtlose Integration zwischen seiner proprietären Lösung und der freien Open-Source-Software - eine Alternative zu GIMP - zu schaffen (Abb. 6.2-3). Damals wie heute geht es in der Konstruktion darum, eine Brücke zwischen geschlossenen und offenen Systemen zu schlagen und dabei komplexe Parameter, Ebenen und interne Softwarelogik zu erhalten.

Die Nutzer suchten jedoch nach einfachen Lösungen - flache, offene Daten ohne übermäßige Komplexität von Layern und Programmparametern (Analogien zum geometrischen Kern im CAD). Die Nutzer suchten nach einfachen und offenen Datenformaten, frei von übermäßiger Logik. JPEG, PNG und GIF wurden zu solchen Formaten in der Grafik. Heute werden sie in sozialen Netzwerken, auf Websites und in Anwendungen verwendet - sie sind einfach zu verarbeiten und zu interpretieren, unabhängig von der Plattform oder dem Softwareanbieter.

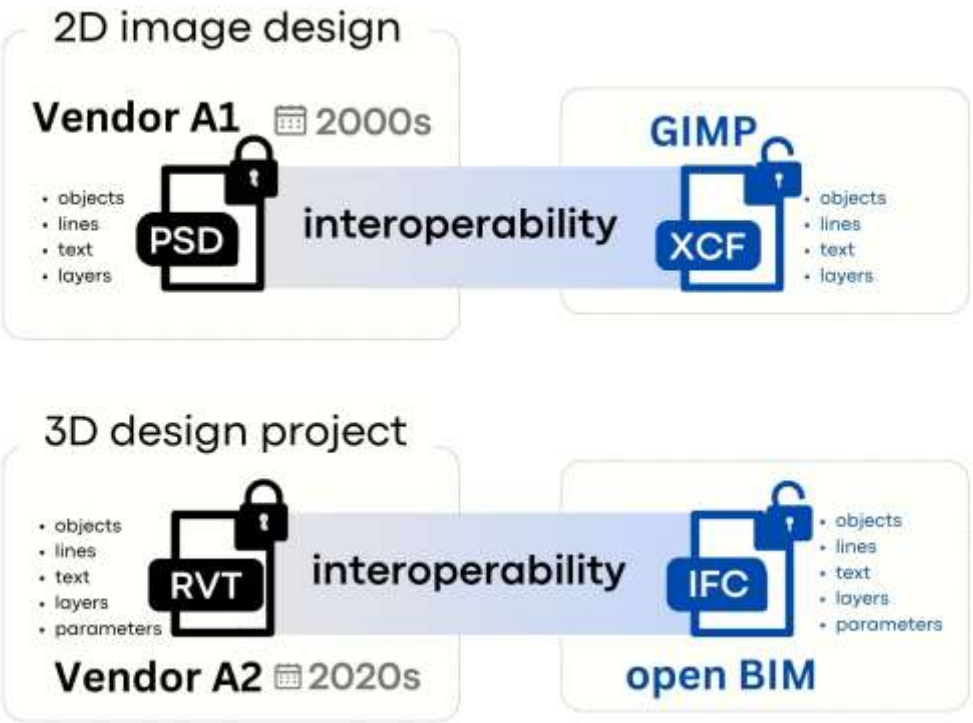


Abbildung 6.2-3 Die Interoperabilität von Datenformaten im Bauwesen ist vergleichbar mit den Versuchen, das proprietäre Produkt eines beliebten Herstellers und das Open-Source-Produkt GIMP in den 2000er Jahren zusammenzuführen.

Infolgedessen verwendet heute fast niemand in der Bildbearbeitungsbranche geschlossene Formate wie PSD oder das offene XCF für Anwendungen, soziale Netzwerke wie Facebook und Instagram oder als Inhalt auf Websites. Stattdessen werden für die meisten Aufgaben flache und offene JPEG-, PNG- und GIF-Formate verwendet, die einfach zu verwenden und weitgehend kompatibel sind. Offene Formate wie JPEG

und PNG haben sich aufgrund ihrer Vielseitigkeit und breiten Unterstützung zum Standard für die gemeinsame Nutzung von Bildern entwickelt, da sie auf einer Vielzahl von Plattformen leicht zu verwenden sind. Ein ähnlicher Übergang ist bei anderen Austauschformaten wie Video und Audio zu beobachten, wo universelle Formate wie MPEG und MP3 aufgrund ihrer Komprimierungseffizienz und breiten Kompatibilität hervorgehoben werden. Diese Entwicklung hin zur Standardisierung hat die gemeinsame Nutzung und Wiedergabe von Inhalten und Informationen vereinfacht, so dass sie für alle Nutzer über verschiedene Plattformen hinweg zugänglich sind (Abb. 6.2-4).

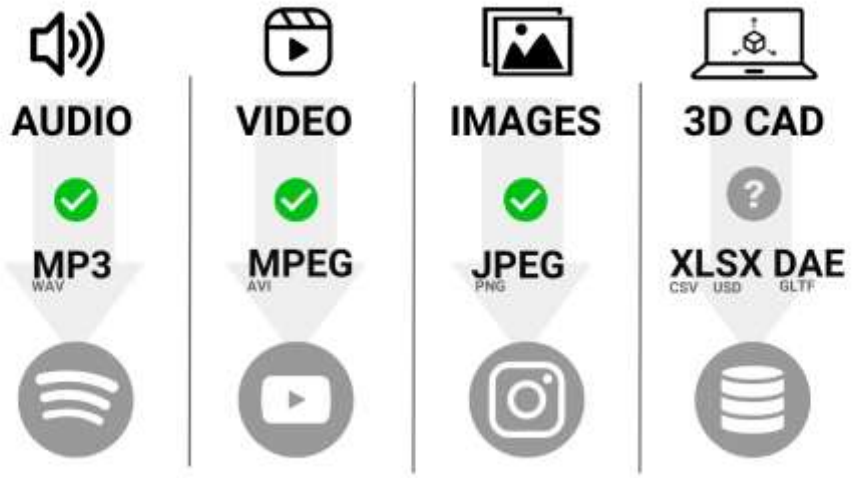


Abbildung 6.2-4 Vereinfachte Formate ohne komplexe Bearbeitungsfunktionen haben sich für den Austausch und die Nutzung von Daten durchgesetzt.

Ähnliche Prozesse finden bei der 3D-Modellierung statt. Einfache und offene Formate wie USD, OBJ, glTF, DAE, DXF, SQL und XLSX werden in Projekten zunehmend für den Datenaustausch außerhalb der CAD-Umgebung verwendet (BIM). In diesen Formaten werden alle erforderlichen Informationen, einschließlich Geometrie und Metadaten, gespeichert, ohne dass eine komplexe BREP-Struktur, Geometriekerne oder herstellerspezifische interne Klassifikatoren verwendet werden müssen. Proprietäre Formate wie NWC, SVF, SVF2, CPIXML und CP2, die von führenden Softwareanbietern bereitgestellt werden, erfüllen ähnliche Funktionen, bleiben aber im Gegensatz zu offenen Standards geschlossen.

> Es ist bemerkenswert (und es lohnt sich, noch einmal daran zu erinnern, wie bereits im vorherigen Kapitel erwähnt), dass diese Idee - die Ablehnung von neutralen und parametrischen Zwischenformaten wie IGES, STEP und IFC - bereits im Jahr 2000 von dem großen CAD-Anbieter unterstützt wurde, der das BIM-Whitepaper erstellt und das IFC-Format 1994 registriert hatte. In dem Whitepaper 2000 "Integrated Design and Manufacturing" [65] betont der CAD-Anbieter die Bedeutung des nativen Zugriffs auf die CAD-Datenbank innerhalb der Softwareumgebung, ohne die Notwendigkeit, zwischengeschaltete Übersetzer und parametrische Formate zu verwenden, um die Vollständigkeit und Genauigkeit der Informationen zu gewährleisten.

Die Bauindustrie hat sich noch nicht auf Werkzeuge für den Zugriff auf CAD-Datenbanken oder deren erzwungenes Reverse Engineering oder auf die Annahme eines gemeinsamen vereinfachten Datenformats zur Verwendung außerhalb von CAD-Plattformen (BIM) geeinigt. Viele große Unternehmen in Mitteleuropa

und im deutschsprachigen Raum, die im Bausektor tätig sind, verwenden beispielsweise das CPIXML -Format in ihren ERP -Systemen [121]. Dieses proprietäre Format, das eine Art XML ist, kombiniert CAD-(BIM)-Projektdaten, einschließlich Geometrie- und Metadaten, in einer einzigen organisierten, vereinfachten Struktur. Große Bauunternehmen entwickeln auch eigene neue Formate und Systeme, wie im SCOPE-Projekt, das wir im vorherigen Kapitel besprochen haben

Die geschlossene Logik parametrischer CAD-Formate oder komplexer parametrischer Dateien IFC (STEP) ist in den meisten Geschäftsprozessen überflüssig. Die Anwender suchen nach vereinfachten und flachen Formaten wie USD, CPIXML, XML &OBJ, DXF, glTF, SQLlite, DAE &XLSX, die alle notwendigen Elementinformationen enthalten, aber nicht mit redundanter BREP-Geometrielogik, Abhängigkeit von Geometriekernen und internen Klassifikationen spezifischer CAD und BIM -Produkte belastet sind (Abb. 6.2-5).

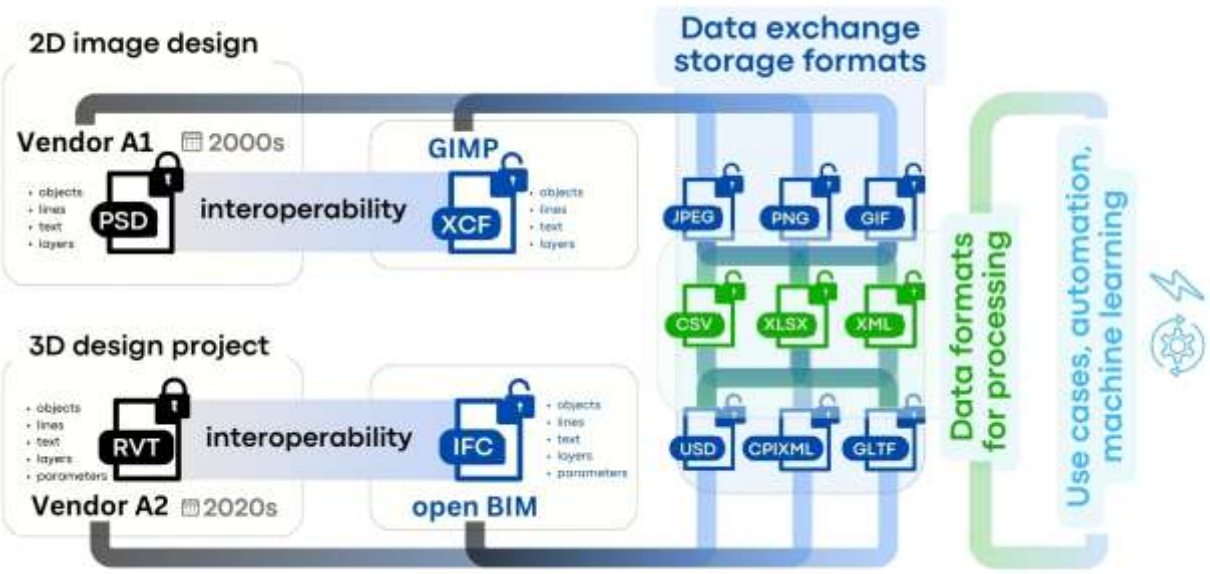


Abb. 6.2-5 Für die meisten Anwendungsfälle wählen die Benutzer die einfachsten möglichen Formate, die von den Programmen der Hersteller unabhängig sind.

Das Aufkommen von flachen Bildformaten wie JPEG, PNG und GIF, die von der redundanten Logik der internen Engines der Anbieter befreit sind, hat die Entwicklung von Tausenden von interoperablen Lösungen für die Verarbeitung und Nutzung von Grafiken vorangetrieben. Dies hat zu einer Vielzahl von Anwendungen geführt, von Retuschier- und Filterwerkzeugen bis hin zu sozialen Netzwerken wie Instagram, Snapchat und Canva, wo diese vereinfachten Daten genutzt werden können, ohne an einen bestimmten Softwareentwickler gebunden zu sein.

> Die Standardisierung und Vereinfachung der CAD-Formate wird das Entstehen zahlreicher neuer benutzerfreundlicher und unabhängiger Werkzeuge für die Arbeit mit Bauprojekten fördern.

Die Abkehr von der komplexen Logik der Herstelleranwendungen, die an geschlossene Geometriekerne gebunden sind, hin zu universellen offenen Formaten, die auf Bibliotheken mit vereinfachten Elementen

basieren, schafft die Voraussetzungen für eine flexiblere, transparentere und effizientere Datenverarbeitung. Dies eröffnet auch den Zugang zu Informationen für alle am Bauprozess Beteiligten - vom Planer über den Kunden bis hin zum Wartungsdienst.

Dennoch ist es sehr wahrscheinlich, dass die Anbieter von CAD in den kommenden Jahren versuchen werden, die Debatte über Interoperabilität und den Zugang zu CAD-Datenbanken wieder zu verschieben. Es wird bereits um "neue" Konzepte gehen - wie granulare Daten, intelligente Graphen, "föderierte Modelle", digitale Zwillinge in Cloud-Repositories - sowie um die Schaffung von Branchenallianzen und Standards, die den Weg von BIM und Open BIM weitergehen. Trotz der attraktiven Terminologie können solche Initiativen wieder zu Instrumenten werden, um die Nutzer in proprietären Ökosystemen zu halten. Ein Beispiel ist die aktive Förderung des Formats USD (Universal Scene Description) als "neuer Standard" für die plattformübergreifende CAD-(BIM-)Zusammenarbeit ab 2023.

Weiter zu USD und granulare Daten

Das Entstehen der AOUSD-Allianz [124] im Jahr 2023 markiert eine wichtige Wende in der Bauindustrie. Wir erleben den Beginn einer neuen, von CAD-Anbietern geprägten Realität im Umgang mit Konstruktionsdaten durch mehrere bedeutende Veränderungen. Die erste große Veränderung betrifft die Wahrnehmung von CAD -Daten. Fachleute, die in den frühen Phasen des Entwurfsprozesses tätig sind, erkennen zunehmend, dass die Erstellung eines Entwurfs in einer CAD-Umgebung nur ein Ausgangspunkt ist. Die während des Entwurfsprozesses erzeugten Daten bilden schließlich die Grundlage für die Analyse, den Betrieb und die Verwaltung von Objekten. Das bedeutet, dass sie in Systemen zugänglich und nutzbar sein müssen, die über die traditionellen CAD-Werkzeuge hinausgehen.

Parallel dazu findet eine Revolution in der Vorgehensweise der führenden Entwickler statt. Der branchenführende CAD-Anbieter, der das BIM-Konzept und das IFC-Format entwickelt hat, vollzieht eine unerwartete Wende in seiner Strategie. Ab 2023 wird sich das Unternehmen von der traditionellen Speicherung von Daten in separaten Dateien verabschieden, sich auf die Arbeit mit granularen (normalisierten und strukturierten) Daten konzentrieren und zu einem datenzentrierten Ansatz übergehen [125].

> Die Anbieter folgen den historischen Trends anderer Branchen: Die meisten Nutzer brauchen keine geschlossenen CAD-Formate (ähnlich wie PSD) oder komplexe parametrische IFC-Dateien (ähnlich wie GIMP mit Ebenenlogik). Sie brauchen einfache Objektbilder, die in CAFM (Construction Instagram), ERP (Facebook) und tausenden anderen Prozessen verwendet werden können, die mit Excel Tabellen und PDF Dokumenten gefüllt sind.

Die aktuellen Trends in der Bauindustrie bereiten möglicherweise die Bühne für eine allmähliche Abkehr von parametrischen und komplexen Formaten zugunsten der universelleren und unabhängigeren Formate USD, GLTF, DAE, OBJ (mit Metainformationen sowohl innerhalb des Hybrids als auch in separaten strukturierten oder lose strukturierten Formaten). Historisch führende Unternehmen, darunter auch große Designfirmen, die Mitte der 1990er Jahre IFC aktiv förderten, werben jetzt offen für das neue USD-Format [93] und betonen dessen Einfachheit und Vielseitigkeit (Abb. 6.2-6). Die massenhafte Akzeptanz von USD in Produkten, die GLTF-Kompatibilität und die aktive Integration in Tools wie Blender, Unreal Engine und Omniverse zeigen das Potenzial für den Beginn eines neuen Paradigmas für die Arbeit mit Daten. Neben der Popularität von lokalisierten Lösungen wie dem europäischen flachen USD-Format CPIXML, das in den gängigen europäischen ERP-Systemen verwendet wird, könnte die Position des USD in Mitteleuropa stärken. Organisationen,

die an der Entwicklung des IFC-Formats beteiligt sind, passen ihre Strategie bereits an das USD-Format an [126], was die Unvermeidbarkeit des Wandels nur bestätigt.

Technical Specifications			Comparison / Notes
File Structure	Monolithic file	Uses ECS and linked data	IFC stores all data in one file; USD uses Entity-Component-System and linked data for modularity and flexibility
Data Structure	Complex semantics, parametric geometry	Flat format, geometry in MESH, data in JSON	IFC is complex and parametric; USD is simpler and uses flat data
Geometry	Parametric, dependent on BREP	Flat, MESH (triangular meshes)	IFC uses parametrics; USD uses meshes for simplified processing.
Properties	Complex structure of semantic descriptions	Properties in JSON, easy access	Properties in USD are easier to use thanks to JSON
Export/Import	Complex implementation, dependent on third-party SDKs	Easy integration, wide support	USD integrates more easily and is supported in many products
Format Complexity	High, requires deep understanding	Low, optimized for convenience	The time required to understand the structure of the file and the information stored in it
Performance	Can be slow when processing large models	High performance in visualization and processing	USD is optimized for speed and efficiency. Simulations, machine learning, AI, smart cities will be held in the Nvidia Omniverse
Integration with 3D Engines	Limited	High, designed for graphics engines	USD excels with native support for real-time visualization platforms
Support outside CAD Software	BlenderBIM, IfcOpenShell	Unreal Engine, Unity, Blender, Omniverse	USD is widely supported in graphics tools
Cloud Technology Support	Limited	Well-suited for cloud services and online collaboration	USD is optimized for cloud solutions
Ease of Integration into Web Applications	Difficult to integrate due to size and complexity	Easy to integrate, supports modern web technologies	USD is preferable for web applications
Change Management	Versions through separate files	Versioning built into the format core	IFC handles changes via separate files, while USD embeds versioning directly into its structure
Collaboration Support	Supports data exchange between project participants	Designed for collaborative work on complex scenes	USD provides efficient collaboration through layers and variations
Learnability	Steep learning curve due to complexity	Easier to master thanks to a clear structure	USD is easier to learn and implement

Abbildung 6.2-6 Vergleich der technischen Spezifikationen von IFC und USD-Formaten.

> Vor diesem Hintergrund hat USD das Potenzial, zum De-facto-Standard zu werden und verspricht, viele derzeitige Beschränkungen zu überwinden, die in erster Linie mit der Komplexität der bestehenden CAD-Formate - (BIM-) und der Abhängigkeit ihrer Interpretation von geometrischen Kerneln zusammenhängen.

Statt parametrischer und komplexer CAD -Formate und IFC - werden sich vereinfachte Datenformate USD, gLTF, DAE, OBJ mit Element-Meta-Informationen in CSV, XLSX, JSON, XML aufgrund ihrer Einfachheit und Flexibilität in der Bauindustrie durchsetzen.

Die aktuellen Veränderungen in der Baubranche sehen auf den ersten Blick wie ein technologischer Durchbruch aus, der mit dem Übergang vom veralteten IFC zum moderneren USD verbunden ist. Allerdings sollte man bedenken, dass derselbe CAD-Anbieter, der IFC entwickelt hat, bereits im Jahr 2000 über dessen Probleme und die Notwendigkeit des Zugangs zur Datenbank schrieb [65] und nun aktiv den Übergang zu einem neuen Standard - USD - fördert.

> Hinter einer weiteren Fassade von "offenen Daten" (USD) und "neuen" Konzepten für ein granulares Datenmanagement durch Cloud-basierte Anwendungen, die von CAD-Anbietern () beworben werden, kann sich die Absicht der Anbieter verbergen, das Projektdatenmanagement zu monopolisieren, so dass sich die Nutzer in einer Situation wiederfinden, in der die Wahl des Formats mehr mit Unternehmensinteressen als mit den tatsächlichen Bedürfnissen zusammenhängt.

Eine Analyse der wichtigsten Fakten [93] zeigt, dass der Hauptzweck dieser Änderungen weniger in der Bequemlichkeit der Benutzer als vielmehr in der Aufrechterhaltung der Kontrolle über Ökosysteme und Datenflüsse zugunsten der Hersteller liegt, die in 40 Jahren nie in der Lage waren, Zugang zu CAD-Datenbanken zu gewähren.

Vielleicht ist es jetzt an der Zeit, dass die Unternehmen nicht mehr auf neue Konzepte von Softwareanbietern warten, sondern sich auf die Eigenentwicklung in Richtung Datenzentrierung konzentrieren. Nachdem sich die Branche durch Reverse-Engineering-Tools von Datenzugriffsproblemen befreit hat, kann sie sich unabhängig von neuen Konzepten auf moderne, kostenlose und bequeme Tools für die Arbeit und Analyse von Daten umstellen.

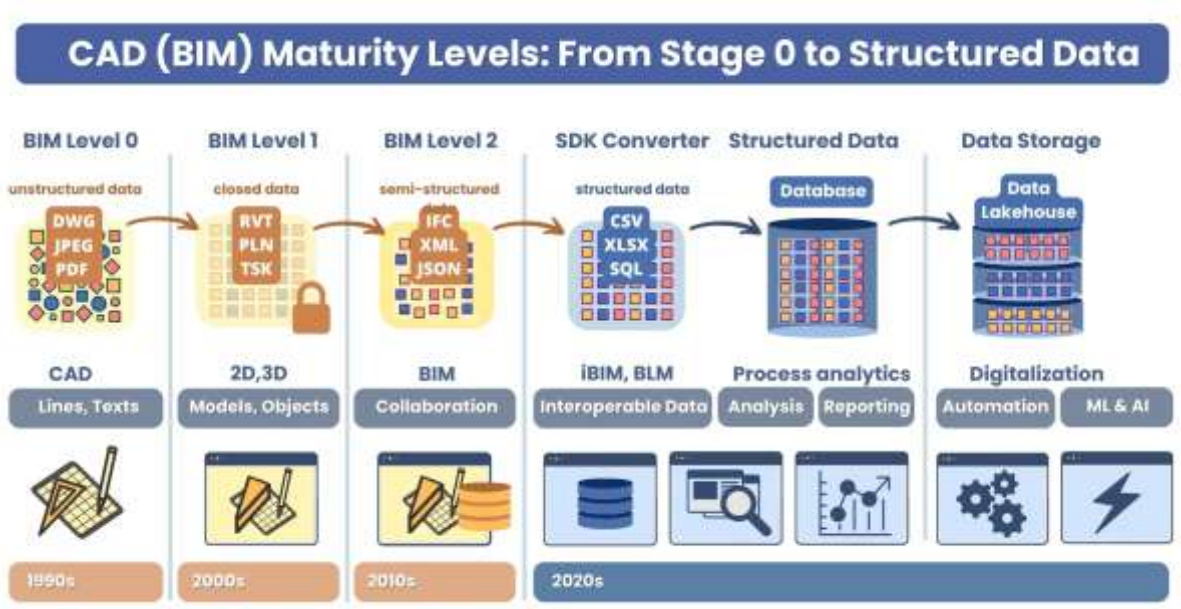


Abb. 6.2-7 CAD-Reifegrad (BIM): von unstrukturierten Daten zu strukturierten Daten und Repositories.

Der Zugang zu Datenbanken, offenen Daten und Formaten wird unweigerlich zum Standard in der Bauindustrie werden, unabhängig von den Versuchen der Anbieter, den Prozess aufzuhalten - es ist nur eine

Frage der Zeit (Abb. 6.2-7). Das Tempo dieses Übergangs kann erheblich gesteigert werden, wenn sich immer mehr Fachleute mit offenen Formaten, Datenbankwerkzeugen und verfügbaren Reverse-Engineering-SDKs vertraut machen, die einen direkten Zugriff auf CAD-Daten -Systeme ermöglichen [92].

> Die Zukunft liegt in offenen, vereinheitlichten und analytisch zugänglichen Daten. Um die Abhängigkeit von Anbieterlösungen zu vermeiden und nicht in Geiselhaft geschlossener Ökosysteme zu geraten, werden Bau- und Ingenieurbüros früher oder später auf Offenheit und Unabhängigkeit setzen und Formate und Lösungen wählen müssen, die volle Kontrolle über die Daten bieten.

Die Daten, die heute in der Baubranche entstehen, werden in Zukunft eine wichtige Ressource für Geschäftsentscheidungen sein. Sie werden der strategische "Treibstoff" sein, der die Entwicklung und Effizienz von Bauunternehmen vorantreibt. Die Zukunft des Baugewerbes liegt in der Fähigkeit, mit Daten zu arbeiten, und nicht in der Wahl der Datenformate oder -modelle.

Um den Unterschied zwischen den offenen Formaten USD, glTF, DAE, OBJ und den proprietären parametrischen CAD-Formaten zu verstehen, ist es wichtig, eines der komplexesten und wichtigsten Datenelemente in der Visualisierung und den Konstruktionsberechnungen zu betrachten - die Geometrie und ihre Erzeugungsprozesse. Und um zu verstehen, wie geometrische Daten zur Grundlage für die Analytik und Berechnungen im Bauwesen werden, ist es notwendig, die Mechanismen der Geometrieerzeugung, -transformation und -speicherung genauer zu betrachten.

KAPITEL 6.3.
GEOMETRIE IM BAUWESEN: VON LINIEN ZU KUBIKMETERN

Wenn Linien zu Geld werden oder warum Bauherren Geometrie brauchen

Geometrie im Bauwesen ist nicht nur eine Visualisierung, sondern auch die Grundlage für genaue quantitative Berechnungen. Im Projektmodell ergänzt die Geometrie die Listen der Elementparameter (Abb. 3.1-16) um wichtige volumetrische Kennwerte wie Länge, Fläche und Volumen. Diese volumetrischen Parameterwerte werden automatisch mit den Geometriekernen berechnet und sind der Ausgangspunkt für Schätzungen, Zeitpläne und Ressourcenmodelle. Wie wir bereits in Teil 5 dieses Buches und im Kapitel "Kalkulation und Schätzung von Bauprojekten" erörtert haben, sind es die volumetrischen Parameter von Objektgruppen aus CAD -Modellen, die die Grundlage für moderne ERP-, PMIS -Systeme bilden/ Die Geometrie spielt nicht nur in der Entwurfsphase eine grundlegende Rolle, sondern auch im Management der Projektdurchführung, der Terminkontrolle, der Budgetierung und des Betriebs. So wie vor Tausenden von Jahren beim Bau der ägyptischen Pyramiden die Genauigkeit eines Projekts von Längenmaßen wie Bögen und Ellen abhing, wirkt sich heute die Genauigkeit der Geometrieinterpretation in CAD -Programmen direkt auf das Ergebnis aus: vom Budget und den Fristen bis hin zur Auswahl der Auftragnehmer und der Lieferlogistik

> In einem wettbewerbsintensiven und budgetbeschränkten Umfeld wird die Genauigkeit der volumetrischen Berechnungen, die direkt von der Geometrie abhängt, zu einem Überlebensfaktor. Moderne ERP -Systeme hängen direkt von korrekten volumetrischen Eigenschaften ab, die aus CAD - und BIM -Modellen gewonnen werden. Aus diesem Grund ist die genaue geometrische Beschreibung von Elementen nicht nur eine Visualisierung, sondern ein Schlüsselinstrument für das Management von Kosten und Bauzeit.

In der Vergangenheit war die Geometrie die wichtigste Sprache der technischen Kommunikation. Von Linien auf Papyrus bis zu digitalen Modellen dienten Zeichnungen und geometrische Darstellungen als Mittel zum Informationsaustausch zwischen Konstrukteuren, Vorarbeitern und Kalkulatoren. Vor dem Aufkommen von Computern wurden Berechnungen manuell mit Lineal und Winkelmesser durchgeführt. Heute ist diese Aufgabe dank der volumetrischen Modellierung automatisiert: Die geometrischen Kerne der CAD-Software wandeln Linien und Punkte in dreidimensionale Körper um, aus denen alle erforderlichen Merkmale automatisch extrahiert werden.

Bei der Arbeit in CAD -Programmen erfolgt die Erstellung von geometrischen Elementen für Berechnungen über die Benutzeroberfläche der CAD- (BIM-) -Programme. Zur Umwandlung von Punkten und Linien in volumetrische Körper wird der geometrische Kern verwendet, der die Hauptaufgabe erfüllt - die Umwandlung von Geometrie in volumetrische Modelle, aus denen die volumetrischen Eigenschaften des Elements nach der Annäherung automatisch berechnet werden.

Von Linien zu Volumen: Wie Fläche und Volumen zu Daten werden

In der technischen Praxis werden Volumen und Flächen aus geometrischen Oberflächen berechnet, die analytisch oder durch parametrische Modelle wie NURBS (nonuniform rational B-splines) im Rahmen von BREP

(boundary element representation) beschrieben werden.

> NURBS (Non-Uniform Rational B-Splines) ist eine mathematische Methode zur Beschreibung von Kurven und Oberflächen, während BREP ein Rahmen für die Beschreibung der gesamten dreidimensionalen Geometrie eines Objekts ist, einschließlich seiner Grenzen, die mit NURBS definiert werden können.

Trotz der Genauigkeit von BREP und NURBS erfordern sie leistungsstarke Rechenressourcen und komplexe Algorithmen. Die direkte Berechnung aus solch mathematisch genauen Beschreibungen ist jedoch oft rechnerisch schwierig, so dass in der Praxis fast immer eine Tesselierung - die Umwandlung von Oberflächen in ein Gitter aus Dreiecken - verwendet wird, was die nachfolgenden Berechnungen vereinfacht. Tesselierung ist die Aufteilung einer komplexen Oberfläche in Dreiecke oder Polygone. In CAD /CAE Umgebungen wird diese Methode zur Visualisierung, Volumenberechnung, Kollisionssuche, zum Export in Formate wie MESH und zur Kollisionsanalyse verwendet. Ein Beispiel aus der Natur ist eine Bienenwabe, bei der eine komplexe Form in ein regelmäßiges Gitter zerlegt wird (Abb. 6.3-1).

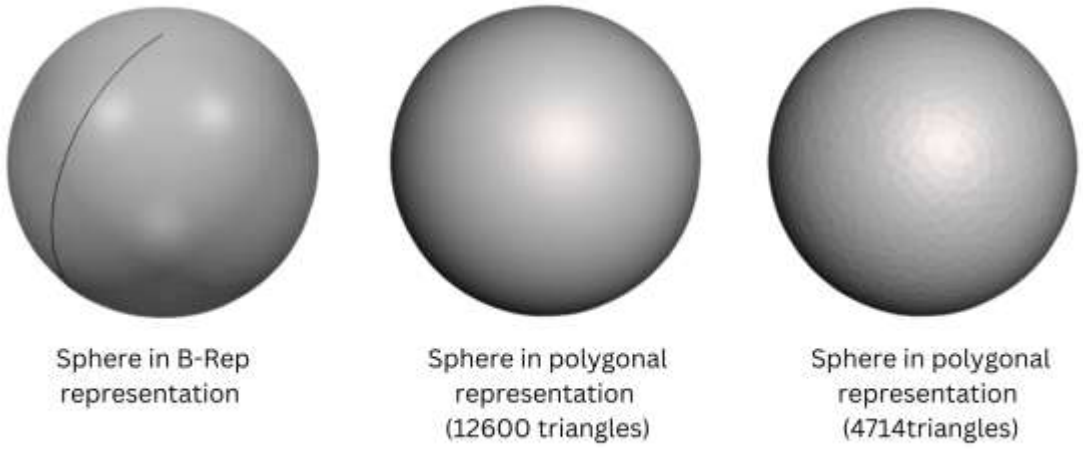


Abb. 6.3-1 Dieselbe Kugel in parametrischer Beschreibung BREP und polygonaler Darstellung mit unterschiedlicher Anzahl von Dreiecken.

BREP (NURBS), das in CAD verwendet wird, ist kein grundlegendes Modell der Geometrie. Es wurde als praktisches Werkzeug zur Darstellung von Kreisen und rationalen Splines und zur Minimierung der Speicherung von Geometriedaten entwickelt. Es hat jedoch Einschränkungen - zum Beispiel die Unfähigkeit, die Sinuskurve, die schraubenförmigen Linien und Flächen zugrunde liegt, genau zu beschreiben, und die Notwendigkeit, komplexe geometrische Kerne zu verwenden.

Im Gegensatz dazu zeichnen sich Dreiecksnetze und die Tesselierung parametrischer Formen durch Einfachheit, effiziente Speichernutzung und die Fähigkeit zur Verarbeitung großer Datenmengen aus (Abb. 6.3-2). Diese Vorteile ermöglichen es, bei der Berechnung geometrischer Formen auf komplexe und teure geometrische Kerne und die darin eingebetteten zig Millionen Codezeilen zu verzichten.

> In den meisten Fällen von Gebäuden spielt es keine Rolle, wie genau die volumetrischen Merkmale definiert werden - durch parametrische Modelle (BREP, IFC) oder durch Polygone (USD, glTF, DAE, OBJ). Die Geometrie bleibt die Form der Annäherung: ob durch NURBS oder MESH, es ist immer eine ungefähre Beschreibung der Form.

Die als Polygone oder BREP (NURBS) definierte Geometrie bleibt bis zu einem gewissen Grad nur eine Möglichkeit der Annäherung durch eine ungefähre Beschreibung einer kontinuierlichen Form. Genauso wie Fresnel-Integrale keinen exakten analytischen Ausdruck haben, ist die Diskretisierung der Geometrie durch Polygone oder NURBS immer eine Annäherung, genauso wie dreieckige MESH.

Parametrische Geometrie im BREP -Format ist vor allem dort notwendig, wo es auf minimale Datengröße ankommt und es möglich ist, ressourcenintensive und teure Geometriekerne für deren Verarbeitung und Darstellung zu verwenden. Meist ist dies charakteristisch für Entwickler von CAD -Programmen, die zu diesem Zweck in ihren Produkten Geometriekerne von MCAD -Anbietern einsetzen. In diesem Fall werden auch innerhalb dieser Programme BREP-Modelle im Zuge der Tesselierung zur Visualisierung und Berechnung oft in Dreiecke umgewandelt (ähnlich wie PSD-Dateien in JPEG vereinfacht werden).

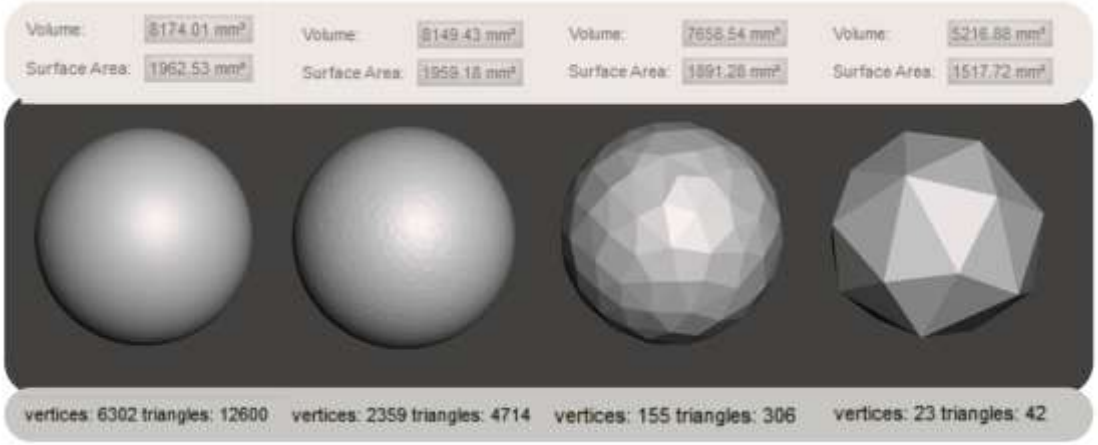


Abb. 6.3-2 Differenz der volumetrischen Merkmale in Figuren mit unterschiedlicher Anzahl von Polygonen.

Sowohl das polygonale MESH als auch das parametrische BREP haben ihre eigenen Vorteile und Grenzen, aber das Ziel ist das gleiche - die Geometrie unter Berücksichtigung der Aufgaben des Benutzers zu beschreiben. Letztendlich hängt die Genauigkeit eines geometrischen Modells nicht nur von der Methode seiner Darstellung ab, sondern auch von den Anforderungen einer bestimmten Aufgabe.

> Bei den meisten Konstruktionsproblemen ist der Bedarf an parametrischer Geometrie und komplexen geometrischen Kernen überflüssig.

Bei jeder einzelnen Aufgabe der Berechnungsautomatisierung lohnt es sich zu überlegen, ob die Bedeutung der parametrischen Geometrie von CAD-Entwicklern, die an der Förderung und dem Verkauf ihrer eigenen Softwareprodukte interessiert sind, übertrieben wird.

Umstellung auf MESH, USD und Polygone: Verwendung der Tesselierung für Geometrie

In der Bauindustrie ist es wichtig, beim Streaming, bei der Entwicklung von Systemen, Datenbanken oder bei der Automatisierung von Prozessen, die mit Konstruktionsinformationen und Merkmalsgeometrien arbeiten, eine Unabhängigkeit von bestimmten CAD-Editoren und Geometriekernen anzustreben.

> Das Austauschformat, das sowohl in den Berechnungsabteilungen als auch auf der Baustelle zu verwenden ist, sollte nicht auf einem speziellen CAD- (BIM-) Programm basieren. Geometrische Informationen sollten in dem Format direkt durch Tesselierung dargestellt werden, ohne Bezug auf den geometrischen Kern oder die CAD-Architektur.

Parametrische Geometrie aus CAD kann als Zwischenquelle betrachtet werden, aber nicht als Grundlage für ein universelles Format. Die meisten parametrischen Beschreibungen (einschließlich BREP und NURBS) werden in jedem Fall zur weiteren Verarbeitung in polygonale MESH umgewandelt. Wenn das Ergebnis dasselbe ist (Tesselierung und Polygone) und der Prozess einfacher ist, liegt die Wahl auf der Hand. Dies ist vergleichbar mit der Wahl zwischen Graphen-Ontologien und strukturierten Tabellen (die wir im vierten Teil besprochen haben): übermäßige Komplexität ist selten gerechtfertigt (Abb. 3.2-10, Abb. 6.1-8).

Offene Formate wie OBJ, STL, glTF, SVF, CPIXML, USD und DAE verwenden eine universelle Dreiecksnetzstruktur, was ihnen erhebliche Vorteile verleiht. Diese Formate zeichnen sich durch eine hervorragende Interoperabilität aus - sie sind mit Hilfe verfügbarer Open-Source-Bibliotheken leicht zu lesen und zu visualisieren, ohne dass komplexe spezialisierte Geometriekerne mit Millionen von Codezeilen erforderlich sind (Abb. 6.3-3). Diese vielseitigen Geometrieformate werden in Anwendungen eingesetzt, die von relativ einfachen Küchenplanungswerkzeugen bei IKEA™ bis zu komplexen Objektvisualisierungssystemen in Kino- und VR -Anwendungen reichen. Ein wichtiger Vorteil ist die Verfügbarkeit einer großen Anzahl von kostenlosen und quelloffenen Bibliotheken für die Arbeit mit diesen Formaten, die für die meisten Plattformen und Programmiersprachen verfügbar sind.

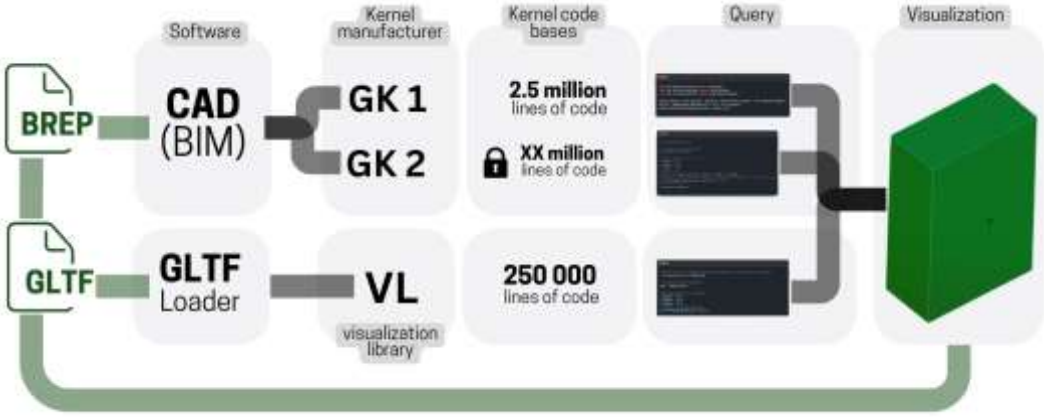


Abbildung 6.3-3 Die gleiche Geometriedarstellung wird durch die Verwendung von parametrischen Formaten und Geometriekernen oder durch die Verwendung von triangulierten Formaten und Open-Source-Visualisierungsbibliotheken erreicht.

Nicht nur die Nutzer selbst, sondern auch die CAD -Anbieter haben Probleme mit der Interpretation ausländischer parametrischer CAD-Formate oder offener IFC aufgrund unterschiedlicher geometrischer Kerne. In der Praxis verwenden ausnahmslos alle CAD -Anbieter das Reverse-Engineering-SDK für den Datentransfer zwischen Systemen, und keiner von ihnen verlässt sich auf Formate wie IFC oder USD [93] für Interoperabilitätszwecke.

> Anstatt Konzepte zu verwenden, die von Allianzen von CAD- Anbietern gefördert werden, die sie selbst nicht nutzen, ist es für Entwickler und Nutzer von CAD-Lösungen produktiver, sich darauf zu konzentrieren, die Vorteile jedes Ansatzes in einem bestimmten Kontext zu verstehen und je nach Anwendungsfall die eine oder andere Art von Geometrie zu wählen. Die Wahl zwischen verschiedenen geometrischen Darstellungen ist ein Kompromiss zwischen Genauigkeit, Berechnungseffizienz und den praktischen Anforderungen einer bestimmten Aufgabe.

Die Komplexität, die mit der Verwendung geometrischer Kerne verbunden ist, die der Bauindustrie traditionell von großen Anbietern bei der Verarbeitung von Entwurfsdaten auferlegt werden, erweist sich oft als zu groß. Das auf der MESH-Geometrie basierende USD-Format kann für die Branche zu einer Art "Büchse der Pandora" werden, indem es Entwicklern neue Möglichkeiten eröffnet, den Datenaustausch zu organisieren - außerhalb des Rahmens der für CAD-Anbieter typischen IFC- und parametrischen BREP-Strukturen.

Bei näherer Betrachtung der Struktur von USD, DAE, gLTF, OBJ usw. wird deutlich, dass es einfachere, offene Formate gibt, die es ermöglichen, die Übertragung und Nutzung geometrischer Informationen effizient zu organisieren, ohne auf komplexe Parametrik und geschlossene geometrische Kerne zurückgreifen zu müssen. Dieser Ansatz senkt nicht nur die technische Einstiegshürde für Entwickler, sondern begünstigt auch die Entwicklung flexibler, skalierbarer und wirklich offener Lösungen für das digitale Bauen.

LOD, LOI, LOMD - eindeutige Klassifizierung von Details in CAD (BIM)

Zusätzlich zu den geometrischen Darstellungsformaten bieten die CAD - (BIM-) -Methoden in einer Welt, in der verschiedene Branchen unterschiedliche Detaillierungsgrade und Datentiefen verwenden, ihre eigenen, einzigartigen Klassifizierungssysteme, die den Ansatz zur Information über Gebäudemodelle strukturieren.

Eines der Beispiele für neue Ansätze zur Standardisierung ist die Einführung von Ebenen der Modellentwicklung, die den Grad der Bereitschaft und Zuverlässigkeit sowohl der grafischen als auch der Informationskomponenten widerspiegeln. Für die Differenzierung des Informationsgehalts bei der Arbeit mit CAD - (BIM-) Daten erschienen LOD (Level Of Detail) - Detailgrad des grafischen Teils des Modells, und LOI (Level Of Information) - Grad der Datenausarbeitung. Darüber hinaus wurde für den integrierten Ansatz das Konzept der LOA (Level of Accuracy) eingeführt - die Genauigkeit der dargestellten Elemente und LOG (Level of Geometry) zur Bestimmung der Genauigkeit der grafischen Darstellung.

Die Detaillierungsgrade (LOD) werden durch Zahlen von 100 bis 500 angegeben, die den Grad der Modellentwicklung widerspiegeln. LOD 100 ist ein konzeptionelles Modell mit allgemeinen Formen und Abmessungen. LOD 200 umfasst genauere Abmessungen und Formen, jedoch mit bedingter Detailgenauigkeit. LOD 300 ist ein detailliertes Modell mit genauen Abmessungen, Formen und Elementpositionen. LOD 400 enthält detaillierte Informationen, die für die Herstellung und den Einbau von Elementen erforderlich sind. LOD 500 spiegelt den tatsächlichen Zustand der Anlage nach dem Bau wider und wird für Betrieb und War-

tung verwendet. Diese Ebenen beschreiben die Struktur der CAD-(BIM)-Modellinformationssättigung in verschiedenen Phasen des Lebenszyklus, einschließlich 3D, 4D, 5D und darüber hinaus.

In realen Projekten ist der hohe Detaillierungsgrad (LOD400) oft zu hoch und es reicht aus, LOD100-Geometrie oder sogar flache Zeichnungen zu verwenden, während der Rest der Daten entweder rechnerisch oder aus verwandten Elementen gewonnen werden kann, die möglicherweise keine eindeutige Geometrie haben. Beispielsweise haben Räume und Raumelemente (Raumelementkategorien) möglicherweise keine visuelle Geometrie, enthalten aber erhebliche Mengen an Informationen und Datenbanken, um die herum viele Geschäftsprozesse aufgebaut sind.

> Es ist daher wichtig, den erforderlichen Detaillierungsgrad vor Beginn des Entwurfs klar zu definieren. Für 4D -7D Anwendungsfälle sind sogar DWG-Zeichnungen und minimale LOD100-Geometrie oft ausreichend. Die Hauptaufgabe im Anforderungsprozess besteht darin, ein Gleichgewicht zwischen dem Detailreichtum und der praktischen Anwendbarkeit des Modells zu finden.

Wenn wir CAD-(BIM-)Daten als Datenbank betrachten (was sie ja auch sind), ist die Beschreibung der Modellsättigung durch neue Akronyme im Grunde nichts anderes als eine schrittweise Datenmodellierung für Informationssysteme, beginnend mit der konzeptionellen Ebene und endend mit der physischen (Abb. 6.3-4), die im dritten und vierten Teil des Buches ausführlich diskutiert wurde. Jede Erhöhung von LOD und LOI bedeutet die Hinzufügung von Informationen, die für neue Aufgaben benötigt werden: Berechnungen, Bauleitung, Betrieb und ist gekennzeichnet durch die sukzessive Anreicherung des Modells mit zusätzlichen Informationsschichten (3D -8D) in Form von verschiedenen Parametern, die wir im fünften Teil des Buches diskutiert haben.

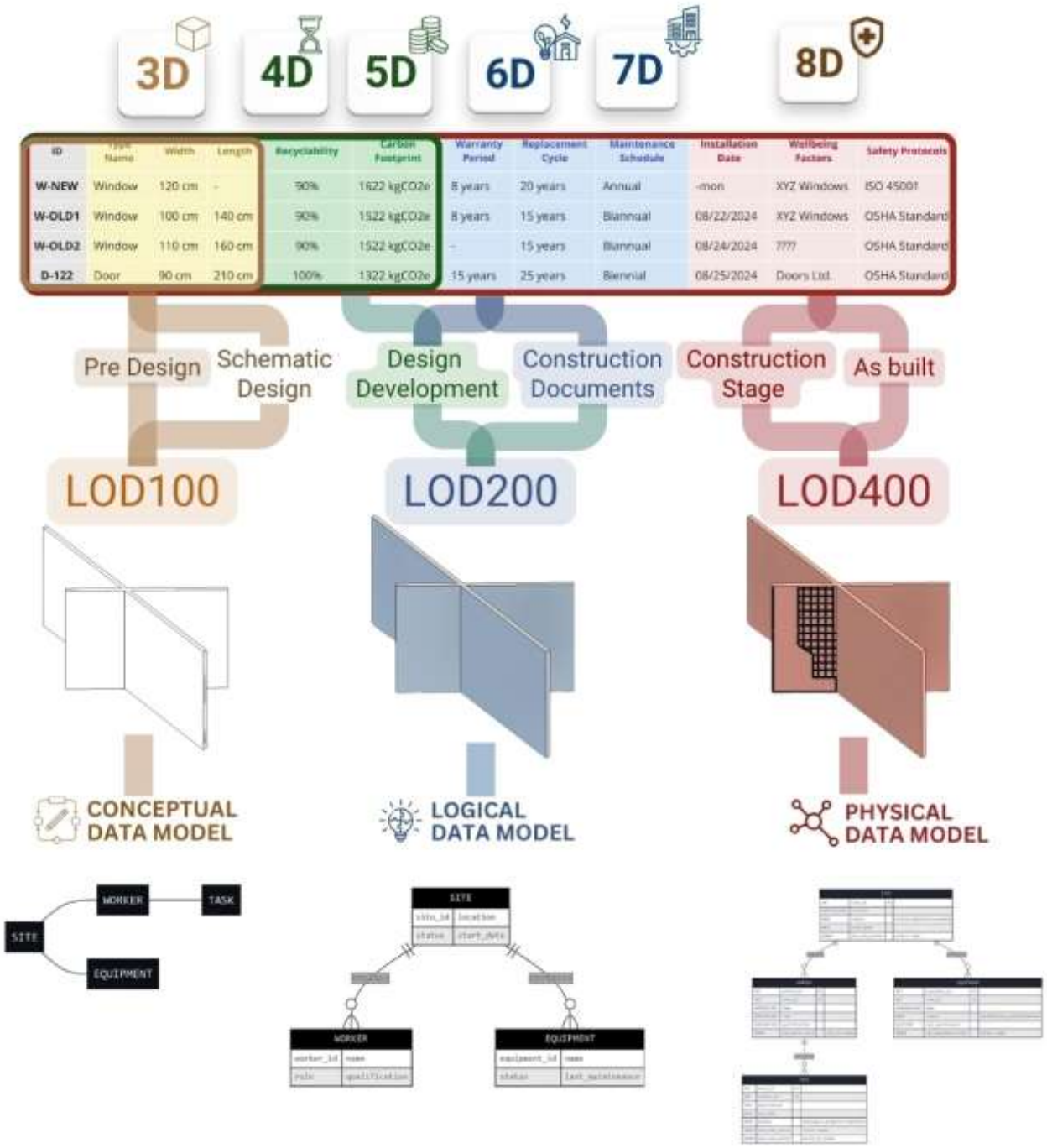


Abbildung 6.3-4 Der Prozess der Konkretisierung eines Projekts ist identisch mit der Datenmodellierung von einem konzeptionellen zu einem physischen Datenmodell.

Die Geometrie ist nur ein Teil der Konstruktionsdaten, deren Bedarf bei Bauprojekten nicht immer gerechtfertigt ist. Das Hauptproblem bei der Arbeit mit CAD -Daten ist nicht so sehr die Visualisierung der Modelle, sondern vielmehr die Frage, wie die Daten aus diesen Modellen außerhalb von CAD- (BIM -) Programmen genutzt werden können.

Mitte der 2000er Jahre sah sich die Bauindustrie mit einer noch nie dagewesenen Herausforderung konfrontiert, da die Datenmenge in den Verwaltungs- und Datenverarbeitungssystemen rapide anstieg, insbesondere die Daten aus den CAD-Abteilungen (BIM). Dieser dramatische Anstieg des Datenvolumens überraschte die Unternehmensleiter und sie waren nicht auf die wachsenden Anforderungen an die Datenqualität und das Datenmanagement vorbereitet.

Neue CAD-Normen (BIM) - AIA, BEP, IDS, LOD, COBie

Unter Ausnutzung des fehlenden offenen Zugangs zu den CAD-Datenbanken und des begrenzten Wettbewerbs auf dem Datenverarbeitungsmarkt sowie unter Ausnutzung der Marketingkampagnen im Zusammenhang mit dem neuen Akronym BIM haben Organisationen, die an der Entwicklung von Konzepten für den Umgang mit CAD-Daten beteiligt sind, damit begonnen, neue Standards und Konzepte zu schaffen, die de jure auf eine Verbesserung der Datenverwaltungspraktiken abzielen sollten.

> Obwohl fast alle Initiativen, die direkt oder indirekt von CAD-Anbietern und -Entwicklern (BIM) unterstützt werden, auf die Optimierung von Arbeitsabläufen abzielen, haben sie zu einer Fülle von Standards geführt, die von verschiedenen Interessengruppen befürwortet werden, was in der Bauindustrie zu einer gewissen Unklarheit und Verwirrung über Datenprozesse führt.

Im Folgenden werden einige der neuen Datenstandards neben LOD, LOI, LOA, LOG aufgeführt, die in den letzten Jahren in der Bauindustrie entstanden sind:

- **BEP** (BIM Execution Plan) - beschreibt die Integration und Nutzung von CAD (BIM) in einem Projekt und definiert Methoden und Prozesse für die Datenverarbeitung.
- **EIR-Dokument /AIA** (Informationsanforderungen des Auftraggebers) - wird vom Auftraggeber vor der Ausschreibung erstellt und enthält die Anforderungen an den Auftragnehmer zur Erstellung und Bereitstellung von Informationen. Es dient als Grundlage für den BEP im jeweiligen Projekt.
- **AIM** (Asset Information Model) ist ein Teil des BIM-Prozesses. Sobald das Projekt geliefert und abgeschlossen ist, wird das Datenmodell als Asset Information Model oder AIM bezeichnet. Der Zweck des AIM ist die Verwaltung, die Instandhaltung und der Betrieb der realisierten Anlage.
- **IDS** (Information Delivery Specification) - definiert die Anforderungen von und welche Daten in welchem Format in den verschiedenen Phasen eines Bauprojekts benötigt werden.
- **iLOD** ist der Detaillierungsgrad LOD, mit dem Informationen im BIM -Modell dargestellt werden. Er definiert, wie detailliert und vollständig die Informationen im Modell sind, von grundlegenden geometrischen Darstellungen bis hin zu detaillierten Spezifikationen und Daten.
- **eLOD** - LOD Detaillierungsgrad der einzelnen Elemente in einem CAD-Modell (BIM). Er definiert den Umfang, in dem jedes Element modelliert wird, und die damit verbundenen Informationen wie Abmessungen, Materialien, Leistungsmerkmale und andere relevante Attribute.
- **APS** (Platform Services) und andere Produkte von großen CAD-Anbietern (BIM) - beschreiben die Werkzeuge und die Infrastruktur, die für die Erstellung von verknüpften und offenen Datenmodellen erforderlich sind.

Obwohl der erklärte Zweck der Einführung von CAD-(BIM)-Standards - wie LOD, LOI, LOA, LOG, BEP, EIR, AIA,

AIM, IDS, iLOD, eLOD - darin besteht, die Qualität der Datenverwaltung zu verbessern und die Automatisierungsmöglichkeiten zu erweitern, führt ihre Anwendung in der Praxis häufig zu einer übermäßigen Komplexität und Fragmentierung der Prozesse. Betrachtet man das CAD-(BIM-)Modell als eine Art Datenbank, so wird deutlich, dass viele dieser Standards die in anderen Branchen seit langem etablierten und effektiven Ansätze für die Arbeit mit Informationssystemen duplizieren. Anstelle von Vereinfachung und Vereinheitlichung schaffen solche Initiativen oft zusätzlichen terminologischen Aufwand und behindern die Umsetzung wirklich offener und flexibler Lösungen.

Bemerkenswert ist, dass viele dieser neuen Konzepte die Modellierungs- und Datenvalidierungsverfahren ersetzen, die in den ersten Teilen des Buches ausführlich besprochen wurden und in anderen Wirtschaftszweigen seit langem verwendet werden. Im Bauwesen hingegen läuft der Standardisierungsprozess oft in die entgegengesetzte Richtung - es werden neue Datenbeschreibungsformate, neue Standards und neue Konzepte zur Datenvalidierung geschaffen, die nicht immer zu einer echten Einheitlichkeit und praktischen Anwendbarkeit führen. Die Folge: Statt die Verarbeitung zu vereinfachen und zu automatisieren, sieht sich die Branche mit einer zusätzlichen Regulierungs- und Bürokratieebene konfrontiert (Abb. 6.3-1), was einer Effizienzsteigerung nicht immer förderlich ist.

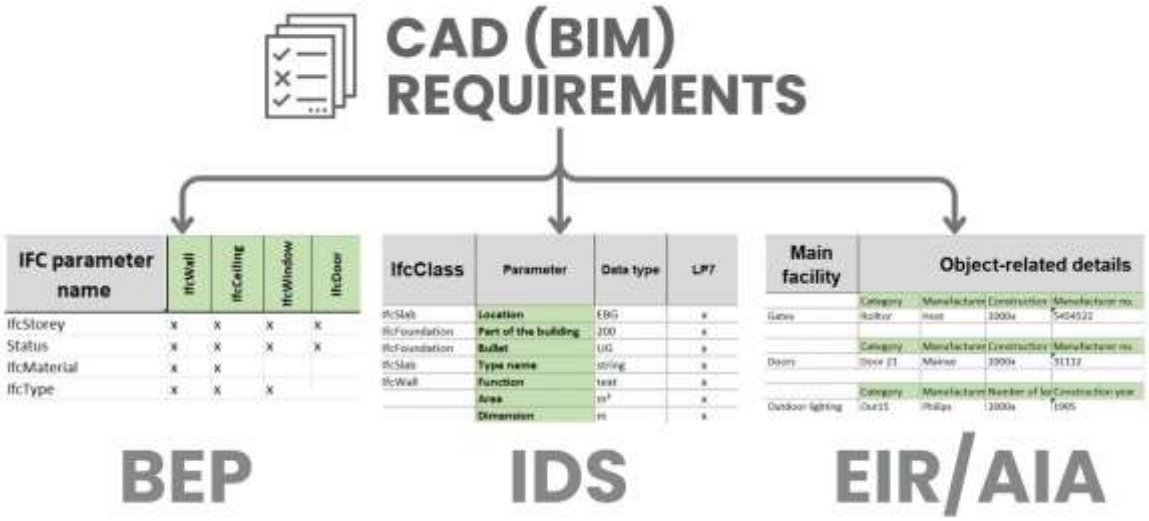


Abb. 6.3-1 Die Anforderungen an den Daten- und Informationsgehalt werden auf die Beschreibung von Attributen und deren Grenzwerten reduziert, die mit Hilfe von Tabellen beschrieben werden.

> Anstatt die Datenverarbeitung zu vereinfachen, führen neue Konzepte im Zusammenhang mit CAD-(BIM-)Daten häufig zu zusätzlicher Komplexität und Streitigkeiten bereits im Stadium der Interpretation und der grundlegenden Definitionen.

Eines der jüngsten Beispiele für neue Konzepte ist das IDS-Format (eingeführt im Jahr 2020), das es ermöglicht, die Anforderungen für die Attributzusammensetzung eines Informationsmodells im offenen BIM-Konzept zu beschreiben. IDS-Anforderungen beschreiben Informationen über Attribute und ihre Grenzwerte in Form einer strukturierten Tabelle (Excel oder MySQL), die dann in das Markup eines halbstrukturierten

XML-Formats übersetzt wird, das von XML in die spezielle Abkürzung IDS umbenannt wurde.

> Im Gegensatz zu der von Anbietern propagierten und von BIM und Open BIM unterstützten Ansicht, dass die Datenverarbeitung im Bauwesen aufgrund der Verwendung spezialisierter Werkzeuge wie CAD und BIM einzigartig ist, unterscheiden sich die Datenformate und Datenverwaltungspraktiken dieser Branche nicht von denen anderer Branchen.

Die Anzahl der Anforderungen für Projekte und CAD (BIM)-Formate kann durch die Verwendung einer einzigen Anforderungstabelle mit Attributspalten, die im Kapitel "Übersetzen von Anforderungen in strukturierte Form" beschrieben wird, vereinfacht werden, ohne dass die ursprünglich strukturierten Anforderungen in nicht tabellarische Formate übersetzt werden müssen (IDS wird zunächst über eine Tabelle beschrieben).

Der vereinfachte Ansatz (Abb. 6.3-2), der Spalten für Entitätsbezeichner, Eigenschaften und Grenzwerte enthält, die in früheren Kapiteln ausführlich besprochen wurden (Abb. 4.4-9, Abbildung 4.4-16, Abbildung 7.3-10), macht eine Konvertierung der Anforderungen in das IDS-XML-Format überflüssig. Diese Methode bietet einen direkten, weniger schwerfälligen und transparenteren Mechanismus zur Kontrolle der Datenqualität. Sie stützt sich auf weit verbreitete Tools, von regulären Ausdrücken (RegEx) bis hin zu Dataframes, Pandas und Standard-ETL -payplanes-, genau wie sie von Fachleuten in anderen Wirtschaftszweigen für die Arbeit mit Daten verwendet werden.

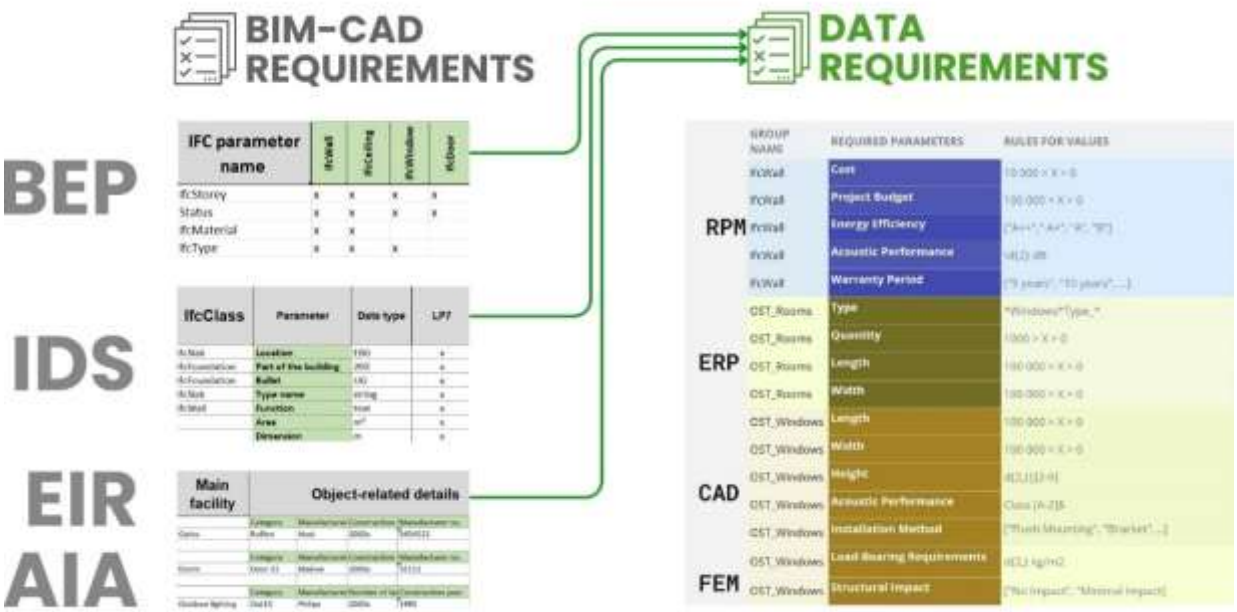


Abb. 6.3-2 In anderen Branchen werden die Datenanforderungen auf eine strukturierte Beschreibung von Attributen und deren Grenzwerten vereinfacht.

Im Laufe der Zeit entstehen im Baugewerbe aufgrund der geschlossenen Natur der Daten immer mehr neue Ansätze und Techniken, um diese vielfältigen Daten zu kontrollieren und zu verwalten, obwohl die Daten in Bauprojekten im Wesentlichen dieselben sind wie in anderen Bereichen. Während andere Branchen erfolgreich mit standardisierten Ansätzen zur Datenverarbeitung ausgekommen sind, entwickelt das Baugewerbe weiterhin neue und einzigartige Datenformate, Anforderungen und Validierungskonzepte.

> Die Methoden und Instrumente zur Erhebung, Aufbereitung und Analyse von Daten im Bauwesen sollten sich nicht grundlegend von denen unterscheiden, die von Fachleuten in anderen Wirtschaftszweigen verwendet werden.

Die Branche hat ein ausgeprägtes terminologisches Ökosystem entwickelt, das eine kritische Reflexion und Neubewertung erfordert:

- Das STEP-Format wird unter dem neuen Namen IFC positioniert, ergänzt um eine Konstruktionskategorisierung, ohne dabei die Einschränkungen des STEP-Formats selbst zu berücksichtigen.
- Das parametrische Format IFC wird in Datenkommunikationsprozessen verwendet, obwohl ein einheitlicher geometrischer Kern für die Visualisierung und Berechnung fehlt.
- Der Zugang zu Datenbanken CAD -systemen wird unter dem Begriff "BIM " beworben, ohne dass auf die Besonderheiten dieser Datenbanken und den Zugang zu ihnen eingegangen wird.
- Die Anbieter fördern die Interoperabilität durch die Formate IFC und USD, oft ohne sie in die Praxis umzusetzen, indem sie ein kostspieliges Reverse Engineering anwenden, mit dem sie sich selbst schwer tun.
- Die Begriffe LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD werden allgemein verwendet, um dieselben Entitätsparameter zu beschreiben, ohne Bezug auf Modellierungs- und Verifizierungswerkzeuge, die in anderen Branchen schon lange verwendet werden.

Das Baugewerbe zeigt, dass all das, auch wenn es seltsam klingt, im Baugewerbe möglich ist - vor allem, wenn das Hauptziel darin besteht, jede Stufe der Datenverarbeitung durch den Verkauf von spezialisierten Dienstleistungen und Software zu monetarisieren. Aus geschäftlicher Sicht ist daran nichts auszusetzen. Ob solche Akronyme und Ansätze im Zusammenhang mit CAD (BIM) jedoch wirklich einen Mehrwert schaffen und professionelle Prozesse vereinfachen, bleibt eine offene Frage.

> Im Baugewerbe funktioniert ein solches System, weil die Branche selbst den größten Teil ihrer Spekulationsgewinne in diesem Labyrinth von Systemen und Abkürzungen macht. Unternehmen, die an transparenten Prozessen und offenen Daten interessiert sind, sind selten. Diese komplexe Situation wird wahrscheinlich auf unbestimmte Zeit fortbestehen - bis Kunden, Auftraggeber, Investoren, Banken und Private Equity anfangen, klarere und fundiertere Ansätze für das Informationsmanagement zu fordern.

In der Branche hat sich eine übermäßige Anzahl von Akronymen angesammelt, jedoch alle in unterschiedlichem Maße dieselben Prozesse und Datenanforderungen beschreiben. Ihr tatsächlicher Nutzen für die Vereinfachung der Arbeitsabläufe bleibt fraglich.

Während Konzepte und Marketing-Akronyme kommen und gehen, werden die Prozesse zur Validierung der Datenanforderungen selbst für immer ein integraler Bestandteil der Geschäftsprozesse bleiben. Anstatt immer mehr spezialisierte Formate und Vorschriften zu schaffen, sollte die Baubranche auf Instrumente zurückgreifen, die sich in anderen Bereichen wie Finanzen, Industrie und IT bereits bewährt haben.

Die Fülle von Begriffen, Akronymen und Formaten erweckt den Anschein, dass die digitalen Konstruktionsprozesse sehr gut durchdacht sind. Hinter den Marketingkonzepten und der komplexen Terminologie verbirgt sich jedoch oft eine einfache, aber unbequeme Wahrheit: Daten sind nach wie vor schwer zugänglich, schlecht dokumentiert und starr an bestimmte Softwarelösungen gebunden.

Um aus diesem Teufelskreis von Akronymen und Formaten um der Formate willen herauszukommen, muss man die CAD-Systeme (BIM) nicht als magische Informationsmanagement-Tools betrachten, sondern als das, was sie wirklich sind - spezialisierte Datenbanken. Und erst durch dieses Prisma kann man verstehen, wo das Marketing aufhört und die wirkliche Arbeit mit Informationen beginnt.

KAPITEL 6.4.
ENTWURFSPARAMETRIERUNG UND NUTZUNG VON LLM FÜR CAD-BETRIEB

Die Illusion der Einzigartigkeit von CAD-Daten (BIM): der Weg zu Analytik und offenen Formaten

Moderne CAD (BIM)-Plattformen haben den Ansatz für das Management von Planungs- und Konstruktionsinformationen erheblich verändert. Während diese Werkzeuge früher hauptsächlich zur Erstellung von Zeichnungen und 3D-Modellen verwendet wurden, fungieren sie heute als vollwertige Projektdaten-Repositories. Im Rahmen des Konzepts der "Single Source of Truth" wird das parametrische Modell immer mehr zur Haupt- und oft zur einzigen Quelle von Projektinformationen, wodurch seine Integrität und Relevanz während des gesamten Lebenszyklus des Projekts gewährleistet wird.

> Der Hauptunterschied zwischen CAD - (BIM -) Plattformen und anderen Systemen zur Verwaltung von Baudaten besteht darin, dass für den Zugriff auf die Informationen (die einzige Quelle der Wahrheit) spezielle Werkzeuge und APIs erforderlich sind. Diese Datenbanken sind nicht universell im herkömmlichen Sinne: Anstelle einer offenen Struktur und flexiblen Integration sind sie eine geschlossene Umgebung, die fest mit einer bestimmten Plattform und einem bestimmten Format verdrahtet ist.

Trotz der Komplexität der Arbeit mit CAD -Daten gibt es eine wichtigere Frage, die über die technische Umsetzung hinausgeht: Was sind CAD-Datenbanken (BIM) wirklich? Um diese Frage zu beantworten, ist es notwendig, über die üblichen Akronyme und Konzepte hinauszugehen, die von Softwareentwicklern vorgegeben werden. Stattdessen lohnt es sich, sich auf das Wesentliche der Arbeit mit Projektinformationen zu konzentrieren: Daten und ihre Verarbeitung.

> Der Geschäftsprozess im Bauwesen beginnt nicht mit der Arbeit in CAD - oder BIM - Werkzeugen, sondern mit der Formulierung der Projektanforderungen und der Datenmodellierung. Zunächst werden die Aufgabenparameter definiert: die Liste der Objekte, ihre ursprünglichen Merkmale und Grenzwerte, die bei der Lösung einer bestimmten Aufgabe berücksichtigt werden müssen. Erst dann werden Modelle und Elemente in CAD (BIM)-Systemen auf der Grundlage der festgelegten Parameter erstellt

Der Prozess, der der Erstellung von Informationen in CAD - (BIM-) Datenbanken vorausgeht, ist völlig identisch mit dem Prozess der Datenmodellierung, der im vierten Teil des Buches und im Kapitel "Datenmodellierung: konzeptionelles, logisches und physikalisches Modell" ausführlich behandelt wurde (Abb. 4.3-1).

So wie wir bei der Datenmodellierung Anforderungen für die Daten erstellen, die wir später in der Datenbank verarbeiten wollen, erstellen Manager für CAD-Datenbanken Konstruktionsanforderungen in Form von mehreren Tabellenspalten oder Listen von Schlüssel-Wert-Paaren (Abb. 6.4-1, Schritte 1-2). Und erst auf der Grundlage dieser Ausgangsparameter erstellt (bzw. verfeinert) der Konstrukteur mit Hilfe der API automatisch oder manuell Objekte in CAD- (BIM) Datenbanken (Schritte 3-4), die anschließend nochmals

auf Übereinstimmung mit den Ausgangsanforderungen überprüft werden (Schritte 5-6). Dieser Prozess - Definition→ Erstellung→ Validierung→ Anpassung (Schritte 2-6) - wird iterativ wiederholt, bis die Datenqualität, genau wie bei der Datenmodellierung, das gewünschte Niveau für das Zielsystem - Dokumente, Tabellen oder Dashboards - erreicht (Schritt 7).

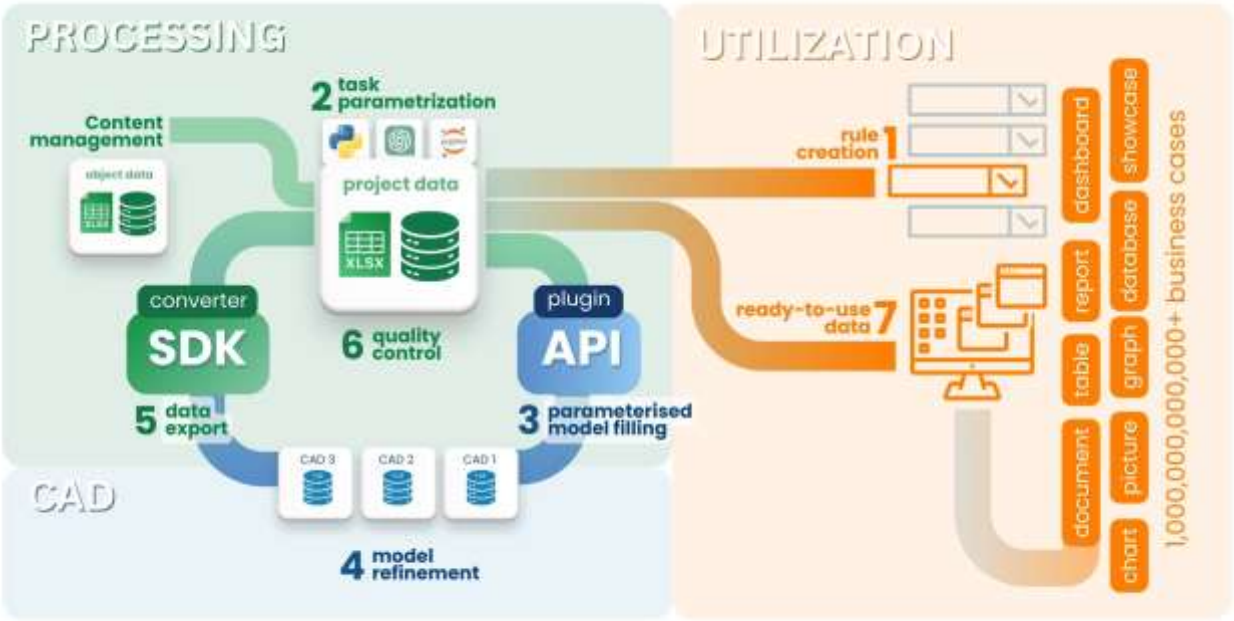


Abb. 6.4-1 Der Zyklus der Informationssättigung von Datenbanken für Geschäftsprozesse bei der Durchführung von Bauprojekten.

Betrachtet man CAD (BIM) als einen Mechanismus zur Parameterübergabe in Form eines Satzes von Schlüssel-Wert-Paaren, die auf der Grundlage von außerhalb der Entwurfsumgebung definierten Anforderungen generiert werden (Abb. 6.4-1, Schritte 1-2), so verschiebt sich der Schwerpunkt der Diskussion von spezifischen Softwarelösungen und ihren Einschränkungen zu grundlegenderen Aspekten - Datenstruktur, Datenmodelle und Datenanforderungen. Im Wesentlichen geht es um die Parametersättigung der Datenbank und den klassischen Datenmodellierungsprozess (Schritte 2-3 und 5-6). Der einzige Unterschied besteht darin, dass dieser Prozess aufgrund des geschlossenen Charakters von CAD-Datenbanken und der Besonderheiten der verwendeten Formate durch den Einsatz spezialisierter BIM-Tools begleitet wird. Es stellt sich die Frage: Worin besteht die Einzigartigkeit von BIM, wenn es in anderen Branchen keine ähnlichen Ansätze gibt?

In den letzten 20 Jahren wurde BIM als mehr als nur eine einzelne Datenquelle positioniert. Das CAD -BIM-Paket wird häufig als parametrisches Werkzeug mit einer inhärent integrierten Datenbank [64] vermarktet, das in der Lage ist, die Prozesse der Planung, Modellierung und des Lebenszyklusmanagements von Bauprojekten zu automatisieren. In der Realität ist BIM jedoch eher ein Werkzeug, um die Benutzer auf der Plattform des Anbieters zu halten, als eine praktische Methode für das Daten- und Prozessmanagement.

Dies hat zur Folge, dass CAD- (BIM-)-Daten innerhalb ihrer Plattformen isoliert sind und Projektinformationen hinter proprietären APIs und Geometriekernen verborgen sind. Dies hat den Nutzern die Möglichkeit genommen, unabhängig auf Datenbanken zuzugreifen und Daten zu extrahieren, zu analysieren, zu automatisieren und an andere Systeme zu übertragen, um so die Ökosysteme der Anbieter zu umgehen.

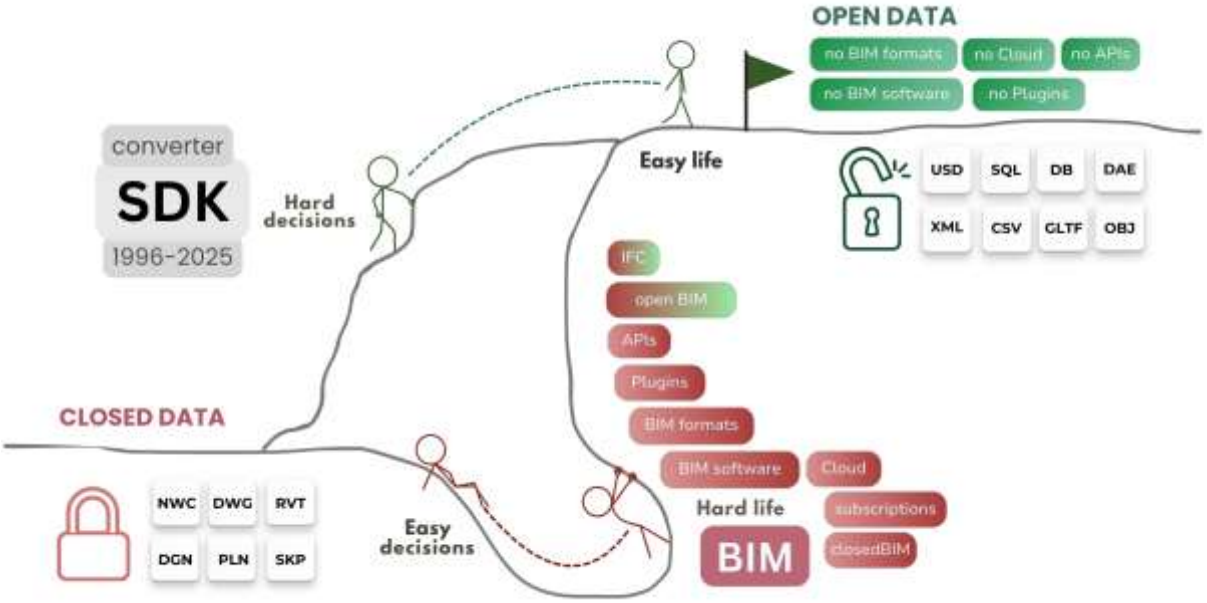


Abb. 6.4-2 Im Bauwesen erfordern moderne Formate anspruchsvolle Geometriekerne, eine jährlich aktualisierte API und spezielle Lizenzen für CAD -(BIM -) Software.

Unternehmen, die mit modernen CAD-Werkzeugen arbeiten, sollten denselben Ansatz für die Arbeit mit Daten verwenden, den ausnahmslos alle CAD-Anbieter selbst in der Praxis anwenden: Datentransformation mit Hilfe von SDK - Reverse-Engineering-Werkzeugen, gegen die die CAD-Anbieter seit 1995 ankämpfen [75]. Mit vollem Zugriff auf die CAD-Datenbank und unter Verwendung von Reverse-Engineering-Werkzeugen können wir [127] einen flachen Satz von Entitäten mit Attributen erhalten und sie in jedes geeignete offene Format exportieren (Abb. 6.4-2), einschließlich der Geometrie und der Parameter von Konstruktionselementen. Dieser Ansatz ändert das Paradigma der Arbeit mit Informationen grundlegend - von einer dateiorientierten zu einer datenzentrierten Architektur:

- Datenformate wie RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET und andere enthalten identische Informationen über Elemente desselben Projekts. Dies bedeutet, dass die Kenntnis eines bestimmten Formats und seines Schemas kein Hindernis für die Arbeit mit den Daten selbst sein sollte.
- Daten aus beliebigen Formaten können zu einer einzigen offenen, strukturierten und granularen Struktur kombiniert werden (Abb. 9.1-10), die die MESH-Dreiecksgeometrie und die Eigenschaften aller Objektentitäten enthält, ohne die Beschränkungen geometrischer Kerne.
- Data analytics strebt nach Universalität: Mit offenen Daten können Sie mit Projektdaten arbeiten, unabhängig vom verwendeten Format.
- Minimierung der Abhängigkeit von APIs und Anbieter-Plugins: Die Arbeit mit Daten hängt nicht mehr von API-Kenntnissen ab.

Wenn und CAD -Datenanforderungen in einfach zu analysierende, strukturierte Darstellungsformate umge-

wandelt werden, sind Entwickler nicht mehr auf spezifische Datenschemata und geschlossene Ökosysteme angewiesen.

Entwurf durch Parameter: die Zukunft von CAD und BIM

Kein Bauprojekt der Welt hat jemals in einem CAD-Programm begonnen. Bevor eine Zeichnung oder ein Modell im CAD Gestalt annimmt, durchläuft es die Konzeptionsphase (Abb. 6.4-1, Stufen 1-2), in der der Fokus auf den Parametern liegt, die die Grundidee und Logik des zukünftigen Objekts definieren. Diese Phase entspricht der konzeptionellen Ebene in der Datenmodellierung (Abb. 4.3-6). Die Parameter können allein im Kopf des Designers existieren, idealerweise werden sie jedoch in Form von strukturierten Listen, Tabellen oder in Datenbanken gespeichert (Abb. 6.4-3), was Transparenz, Reproduzierbarkeit und eine weitere Automatisierung des Designprozesses ermöglicht.

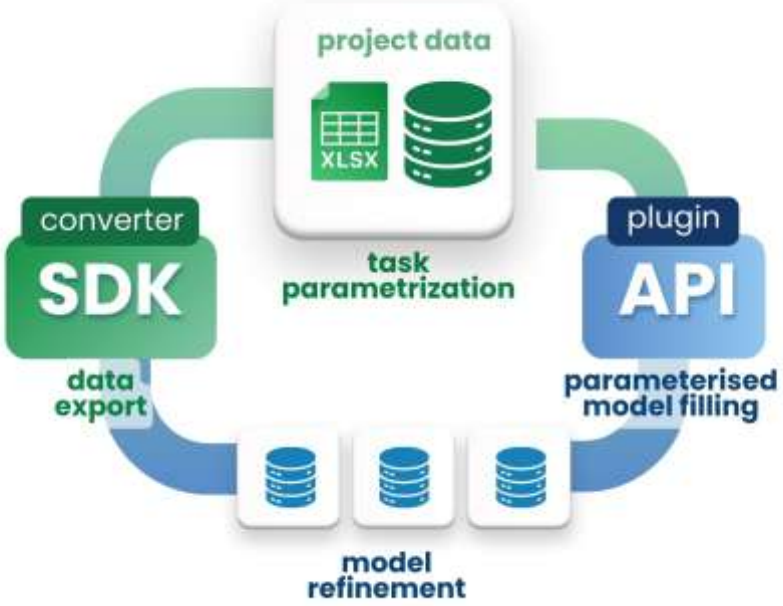


Abb. 6.4-3 Der Entwurfsprozess ist ein iterativer Prozess, bei dem die CAD-Datenbank mit Informationen von außen anhand von Anforderungen aus der Wertschöpfungskette gefüllt wird.

Bevor mit der eigentlichen CAD-Modellierung (der logischen und physischen Phase der Datenmodellierung (Abb. 4.3-7)) begonnen wird, ist es wichtig, die Randparameter zu definieren, die als Grundlage für das Projekt dienen. Diese Attribute werden, wie auch andere Anforderungen, ganz am Ende der Datennutzungskette (z.B. Systeme) gesammelt und durch sie werden bereits die Einschränkungen, Ziele und Schlüsseleigenschaften der zukünftigen Objekte im Projekt definiert.

Die Modellierung selbst kann mit Hilfe von parametrischen Modellierungswerkzeugen (Abb. 6.4-3) zu 60-100% automatisiert werden, wenn die Anforderungen gut definiert sind. Sobald das Projekt in Form von Parametern beschrieben ist, wird seine Gestaltung technisch machbar, z.B. mit Hilfe von visuellen Programmiersprachen wie Grasshopper Dynamo, eingebettet in moderne CAD -Umgebungen oder freien Lösungen in Blender, UE, Omniverse.

Abb. 6.4-4 Die meisten typisierten Projekte werden bereits heute dank parametrischer Programmierwerkzeuge vollautomatisch erstellt.

Schon heute werden große industrielle und typisierte Projekte nicht mehr von der Designabteilung entworfen, sondern mit Hilfe parametrischer Werkzeuge und visueller Programmierung. Dies ermöglicht es, ein Modell auf der Grundlage von Daten und nicht auf der Grundlage der subjektiven Entscheidungen eines bestimmten Designers oder Managers zu erstellen.

> *Der Inhalt geht dem Design voraus. Design ohne Inhalt ist kein Design, sondern Dekoration [128].*
>
> - Jeffrey Zeldman, Web-Designer und Unternehmer

Der Prozess beginnt nicht mit einer Zeichnung oder 3D -Modellierung, sondern mit der Formulierung von Anforderungen. Die Anforderungen bestimmen, welche Elemente im Projekt verwendet werden und welche Daten an andere Abteilungen und Systeme übertragen werden müssen. Erst das Vorhandensein strukturierter Anforderungen macht es möglich, Modelle regelmäßig (z.B. sogar alle 10 Minuten) automatisch zu überprüfen, ohne den Konstrukteur von seiner Arbeit abzulenken.

> Vielleicht wird das CAD- (BIM-) System in Zukunft nur noch eine Schnittstelle zum Befüllen der Datenbank sein, und es wird keine Rolle mehr spielen, in welchem CAD-Tool die Modellierung (physikalische Ebene).

Auch im Maschinenbau wird die 3D-Modellierung häufig verwendet, ist aber kein notwendiges oder obligatorisches Element des Projekts. In den meisten Fällen reicht die klassische 2D-Dokumentation aus, auf deren Grundlage das erforderliche Informationsmodell erstellt wird. Dieses Modell wird aus nach Industriestandards strukturierten Komponenten zusammengesetzt und enthält alle notwendigen Informationen zum

Verständnis der Konstruktion und der Produktionsorganisation. Aus dem Fabrikinformationsmodell wird dann ein Fabrikinformationsmodell erstellt, dem spezifische Produkte und Flussdiagramme hinzugefügt werden, die sich bereits an den Bedürfnissen der Technologen orientieren. Der gesamte Prozess kann ohne unnötige Komplexität organisiert werden, ohne das System mit 3D-Grafiken zu überfrachten, wo sie keine wirklichen Vorteile bieten.

> Es ist wichtig zu verstehen, dass das 3D-Modell selbst und das CAD -System nicht die Hauptrolle spielen sollten - es ist lediglich ein Werkzeug für die quantitative und geometrische Analyse. Alle anderen Parameter, außer der Geometrie, die das Objekt beschreiben, sollten nach Möglichkeit außerhalb der CAD-Umgebung gespeichert und verarbeitet werden (BIM).

Design durch Parameter ist nicht nur ein Trend, sondern die unvermeidliche Zukunft der Bauindustrie. Anstatt komplexe 3D-Modelle manuell zu erstellen, werden die Konstrukteure mit Daten arbeiten, diese validieren und Prozesse automatisieren, wodurch sich das Bauwesen der Welt der Programmierung annähert. Mit der Zeit werden die Konstruktionsprozesse auf den Grundsätzen der Softwareentwicklung aufgebaut sein:

- Anforderungen erstellen → Modell erstellen→ Auf den Server hochladen→ Validieren Änderungen→ Pull-Anfrage
- Pull-Requests führen automatisch Modellprüfungen gegen Anforderungen durch, die vor oder während des Designs als Teil des Pull-Requests erstellt wurden.
- Nach der Überprüfung der Datenqualität und der Freigabe werden die Änderungen in das Projekt, die gemeinsame Datenbank oder automatisch in andere Systeme übertragen

Im Maschinenbau beginnen solche Konstruktionsänderungen bereits jetzt mit der Erstellung einer Änderungsmitteilung. Ein ähnliches System erwartet die Bauindustrie: Die Planung wird ein iterativer Prozess sein, bei dem jeder Schritt durch parametrische Anforderungen unterstützt wird. Ein solches System wird es den Konstrukteuren ermöglichen, automatisierte Prüfungen und automatisierte Pull-Anforderungen für bestimmte Anforderungen zu erstellen.

> Der Designer der Zukunft ist in erster Linie ein Datenverwalter und kein manueller Modellierer. Seine Aufgabe ist es, das Projekt mit parametrischen Elementen zu füllen, wobei die Geometrie nur eines der Attribute ist.

Das Verständnis für die Bedeutung der Datenmodellierung, der Klassifizierung und der Standardisierung, die in den vorangegangenen Kapiteln des Buches ausführlich erörtert wurden, wird bei der Umgestaltung eine wichtige Rolle spielen. Die Gestaltungsvorschriften der Zukunft werden als Schlüssel-Wert-Parameterpaare in Form von XLSX oder XML -Schemata formalisiert.

> Die Zukunft des Baugewerbes liegt in der Sammlung von Daten, ihrer Analyse, ihrer Validierung und der Automatisierung von Prozessen mit Hilfe von Analysewerkzeugen. BIM (oder CAD) ist nicht das Endziel, sondern nur eine Stufe der Entwicklung. Wenn die Fachleute erkennen, dass sie direkt mit Daten arbeiten können und die traditionellen CAD-Tools umgehen, wird der Begriff "BIM" selbst allmählich den Konzepten zur Verwendung strukturierter und granularer Bauprojektdaten weichen.

Einer der Schlüsselfaktoren, die diesen Wandel beschleunigen, ist das Aufkommen großer Sprachmodelle (LLM) und der darauf basierenden Werkzeuge. Diese Technologien verändern die Art und Weise, wie Konstruktionsdaten gehandhabt werden, und ermöglichen den Zugang zu Informationen, ohne dass tiefgreifende Kenntnisse über APIs oder Anbieterlösungen erforderlich sind. Mit LLMs wird der Prozess der Erstellung einer Anforderung und die Interaktion mit CAD-Daten intuitiv und zugänglich.

Entstehung von LLM in Design-CAD Datenverarbeitungsprozessen

Neben der Entwicklung von Werkzeugen für den Zugang zu CAD-Datenbanken und offenen und vereinfachten CAD -Formaten revolutioniert das Aufkommen von LLM -Werkzeugen (Large Language Models) die Verarbeitung von Konstruktionsdaten. Während in der Vergangenheit der Zugang zu Informationen hauptsächlich über komplexe Schnittstellen erfolgte und Programmierkenntnisse und Wissen über API erforderte, ist es jetzt möglich, mit Daten über natürliche Sprache zu interagieren.

Ingenieure, Manager und Planer ohne technischen Hintergrund können die notwendigen Informationen aus den Projektdaten erhalten, indem sie Abfragen in normaler Sprache formulieren. Unter der Voraussetzung, dass die Daten strukturiert und zugänglich sind (Abb. 4.1-13), genügt es, eine Frage in LLM chat zu stellen wie: "*Zeige in einer Tabelle mit Gruppierung nach Typ alle Wände mit einem Volumen von mehr als 10 Kubikmetern*" - und das Modell wird diese Abfrage automatisch in SQL oder Code in Pandas umwandeln und eine zusammenfassende Tabelle, ein Diagramm oder sogar ein fertiges Dokument erzeugen.

Nachfolgend finden Sie einige Beispiele aus der Praxis, wie LLM -Modelle mit Konstruktionsdaten interagieren, die in verschiedenen CAD- (BIM-)-Formaten dargestellt sind.

- Beispiel einer Abfrage im LLM Chat zu einem CAD-Projekt im RVT Format nach Konvertierung (Abb. 4.1-13) in einen tabellarischen Datenrahmen (CHATGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN oder andere):

> Gruppieren Sie die Daten im Datenrahmen aus der RVT-Datei nach "Typname", wenn Sie den Parameter "Volumen" summieren, und zeigen Sie die Anzahl der Elemente in der Gruppe an. Und bitte zeigen Sie all dies als horizontales Histogramm ohne Nullwerte.

LLM-Antwort als horizontales Balkendiagramm (PNG-Format):

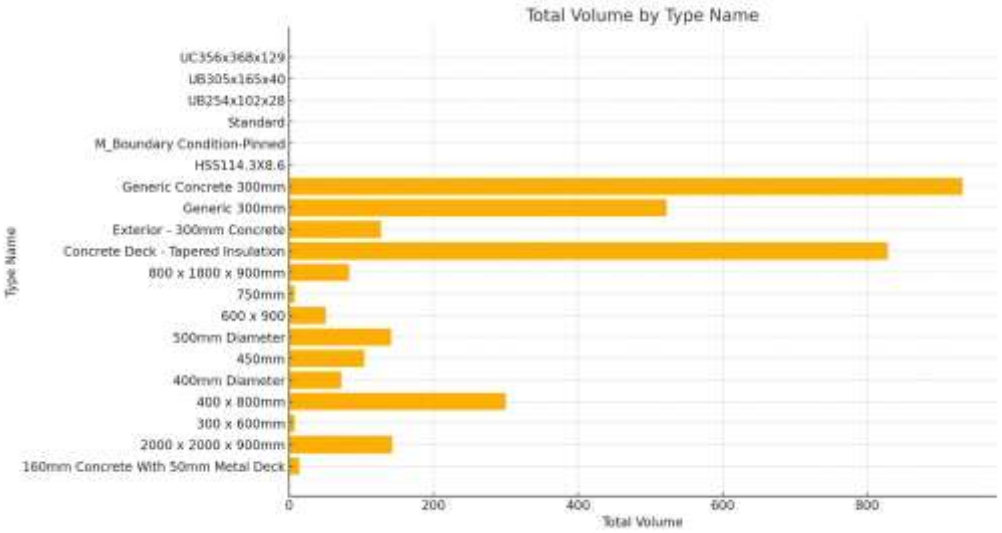


Abbildung 6.4-5 Anstelle von 17 Mausklicks oder 40 Codezeilen durch die Verwendung von Plug-ins, rufen wir in LLM die Tabelle QTO sofort durch eine Textabfrage ab.

Um eine QTO Tabelle der Wandtypen mit Gesamtfläche und Menge aus der Kategorie "Wände" zu erstellen, formulieren wir eine Textabfrage für LLM -chat:

> Nehmen Sie nur die Elemente aus dem Projektdatenrahmen, die "OST_Walls" im Parameter "Category" haben, gruppieren Sie sie nach "Type Name", summieren Sie den Wert der Spalte "Area", addieren Sie die Menge und zeigen Sie sie in der Tabelle an, indem Sie Nullwerte entfernen.

LLM-Antwort in Form einer fertigen QTO-Tabelle:

Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

Abb. 6.4-6 Die Erstellung einer QTO Tabelle in natürlicher Sprache liefert das gleiche Qualitätsergebnis wie die Verwendung von CAD - (BIM-) Tools.

Lassen Sie uns das Projekt im IFC Format abfragen, nachdem wir es in einen tabellarischen Datenrahmen konvertiert haben, und eine ähnliche Textabfrage in einen beliebigen LLM Chat eingeben:

> Nehmen Sie nur die Elemente aus dem Projekt, die Werte für Level 1 und Level 2 im Parameter "Parent" haben, und nehmen Sie die Elemente, die IfcSlab-Werte im Parameter "Category" haben, gruppieren Sie diese Elemente dann nach dem Parameter "ObjectType", summieren Sie die Werte im Parameter "PSet_RVT _Dimensions Area" und zeigen Sie sie als Kreisdiagramm an.

- LLM-Antwort als fertiges Kreisdiagramm von Elementgruppen aus IFC-Daten:

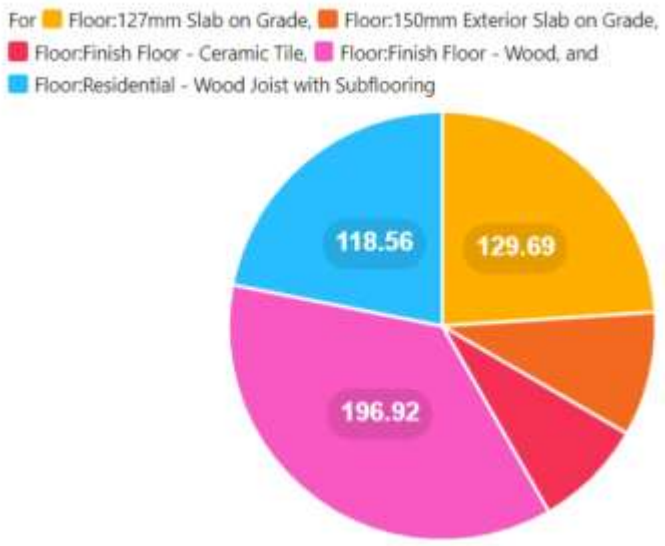


Abbildung 6.4-7 Das Ergebnis einer Abfrage von IFC Daten in einem strukturierten Format kann jede Art von Grafik sein, die für das Verständnis der Daten geeignet ist.

Hinter jeder der vorgefertigten Lösungen (Abb. 6.4-5 - Abb. 6.4-7) steht ein Dutzend Zeilen Python-Code, der die Pandas-Bibliothek verwendet. Der resultierende Code kann aus dem LLM-Chatroom kopiert und in jeder lokalen oder Online-IDE verwendet werden, um identische Ergebnisse außerhalb des LLM-Chatrooms zu erhalten.

Im gleichen LLM-Chat können wir nicht nur mit Projekten arbeiten, die aus 3D-CAD-Formaten (BIM) stammen, sondern auch mit flachen Zeichnungen im DWG-Format, zu denen wir den LLM-Chat abfragen können, um zum Beispiel Daten zu Elementgruppen in Form von Linien oder 3D-Geometrien nach der Konvertierung in eine strukturierte Form anzuzeigen.

Automatisierte Analyse von DWG -Dateien mit LLM und Pandas

Die Verarbeitung von Daten aus DWG -Dateien war aufgrund des unstrukturierten Charakters der Informationen schon immer eine komplexe Aufgabe, die spezialisierte Software und häufig eine manuelle Analyse erforderte. Mit der Entwicklung von künstlicher Intelligenz und LLM-Tools ist es jedoch möglich geworden, viele Schritte dieses heute meist manuellen Prozesses zu automatisieren. Betrachten wir eine echte Pipeline von Anfragen an LLM (in diesem Beispiel ChatGPT), um mit DWG-Zeichnungen zu arbeiten, die es Ihnen ermöglichen, mit dem Projekt zu arbeiten:

- Filtern von DWG-Daten nach Ebene, ID und Koordinaten
- Visualisieren Sie die Geometrie der Elemente
- Zeichnungen automatisch auf der Grundlage von Parametern mit Anmerkungen versehen

- Wandpolylinien auf die horizontale Ebene ausdehnen
- Erstellen Sie interaktive 3D -Visualisierungen von flächigen Daten
- Strukturieren und Analysieren von Konstruktionsdaten ohne komplexe CAD -Werkzeuge

In unserem Fall beginnt der Prozess der Erstellung der Pipeline mit der sequenziellen Codegenerierung durch den LLM. Zunächst wird eine Abfrage, die die Aufgabe beschreibt, generiert. ChatGPT generiert Python -Code, der ausgeführt und analysiert wird, wobei das Ergebnis im Chatroom angezeigt wird. Wenn das Ergebnis nicht den Erwartungen entspricht, wird die Anfrage korrigiert und der Prozess wiederholt

> **Pipeline** ist eine Abfolge von automatisierten Schritten zur Verarbeitung und Analyse von Daten. In einem solchen Prozess nimmt jeder Schritt Daten als Eingabe, führt Umwandlungen durch und leitet das Ergebnis an den nächsten Schritt weiter.

Nachdem das gewünschte Ergebnis erzielt wurde, wird der Code aus LLM kopiert und in Form von Blöcken in den Code einer beliebigen IDE eingefügt, in unserem Fall auf der Platform Kaggle.com. Die resultierenden Codefragmente werden in einer einzigen Pipeline zusammengefasst, die den gesamten Prozess - vom Laden der Daten bis zu ihrer endgültigen Analyse - automatisiert. Dieser Ansatz ermöglicht eine schnelle Entwicklung und Skalierung von Analyseprozessen ohne tiefgreifende Programmierkenntnisse. Der vollständige Code aller unten aufgeführten Fragmente kann zusammen mit Beispielabfragen auf der Plattform Kaggle.com gefunden werden, indem man nach "DWG Analyse with ChatGPT | DataDrivenConstruction" sucht [129].

Beginnen wir den Prozess der Arbeit mit DWG-Daten nach der Konvertierung in eine strukturierte Form (Abb. 4.1-13), mit einem klassischen Schritt - der Gruppierung und Filterung aller Zeichnungsdaten, die für unsere Aufgabe erforderlich sind, nämlich Wandelemente, insbesondere Polylinien (der Parameter 'ParentID' ermöglicht die Gruppierung von Linien in Gruppen), die im Parameter (Datenrahmenspalte) "Layer" einen String-Wert mit der folgenden Buchstabenkombination (RegEx) haben - "wall".

- Um den Code für eine ähnliche Aufgabe und das Ergebnis in Form eines Bildes zu erhalten, sollten Sie die folgende Abfrage in LLM schreiben:

> Prüfen Sie zunächst, ob der von DWG erhaltene Datenrahmen die definierten Spalten enthält: 'Layer', 'ID', 'ParentID' und 'Point'. Filtern Sie dann die IDs aus der Spalte "Layer" heraus, die die Zeichenfolge "wall" enthalten. Suchen Sie die Elemente in der Spalte "ParentID", die mit diesen Bezeichnern übereinstimmen. Definieren Sie eine Funktion zum Bereinigen und Aufteilen der Daten in der Spalte "Punkt". Dazu gehört das Entfernen von Klammern und das Aufteilen der Werte in "x"-, "y"- und "z"-Koordinaten. Plotten Sie die Daten mit matplotlib. Zeichnen Sie für jede eindeutige "ParentID" eine separate Polylinie, die die "Point"-Koordinaten verbindet. Achten Sie darauf, dass der erste und der letzte Punkt möglichst miteinander verbunden sind. Legen Sie die entsprechenden Beschriftungen und Titel fest und stellen Sie sicher, dass die x- und y-Achsen gleich skaliert sind.

- Die Antwort LLM gibt Ihnen ein fertiges Bild, hinter dem sich der Python-Code verbirgt, der es erzeugt hat:

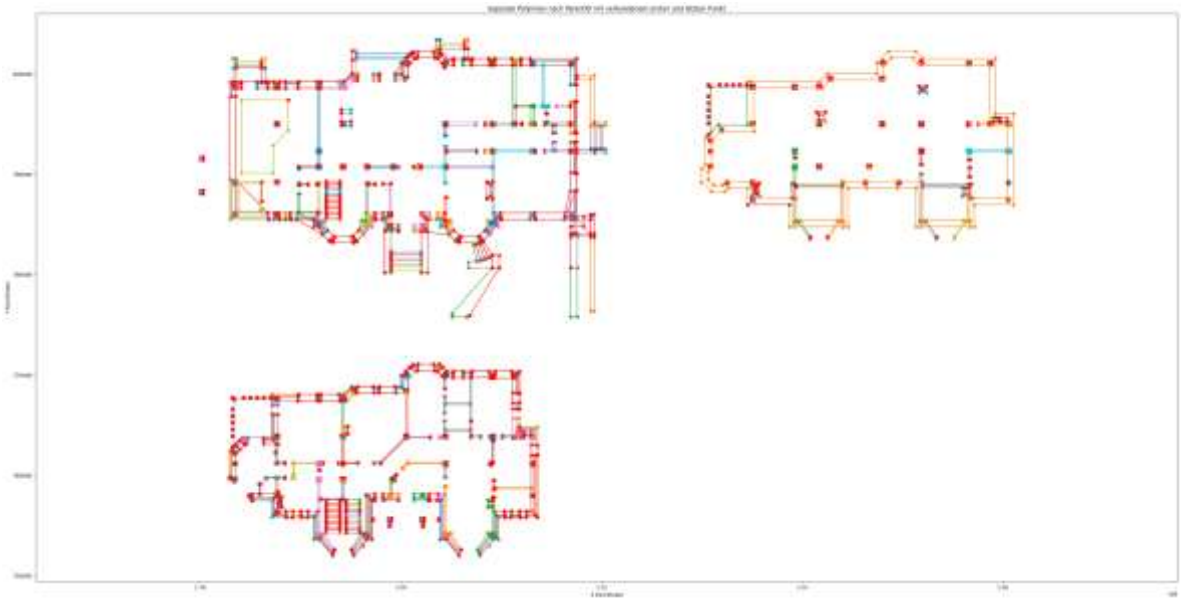


Abb. 6.4-8 Der LLM Code extrahierte alle Linien des Layers "Wand" aus der DWG Datei, löschte ihre Koordinaten und konstruierte die Polylinien mit Hilfe einer der Python Bibliotheken.

Fügen wir nun zu den Linien den Flächenparameter hinzu, den jede Polylinie in ihren Eigenschaften (in einer der Spalten des Datenrahmens) hat:

> Ermitteln Sie nun nur eine "ParentID" von jeder Polylinie - suchen Sie diese ID in der Spalte "ID", nehmen Sie den Wert "Area", teilen Sie ihn durch 1.000.000 und fügen Sie diesen Wert dem Diagramm hinzu

Die LLM-Antwort zeigt ein neues Diagramm, in dem jede Polylinie eine Beschriftung mit ihrer Fläche hat:

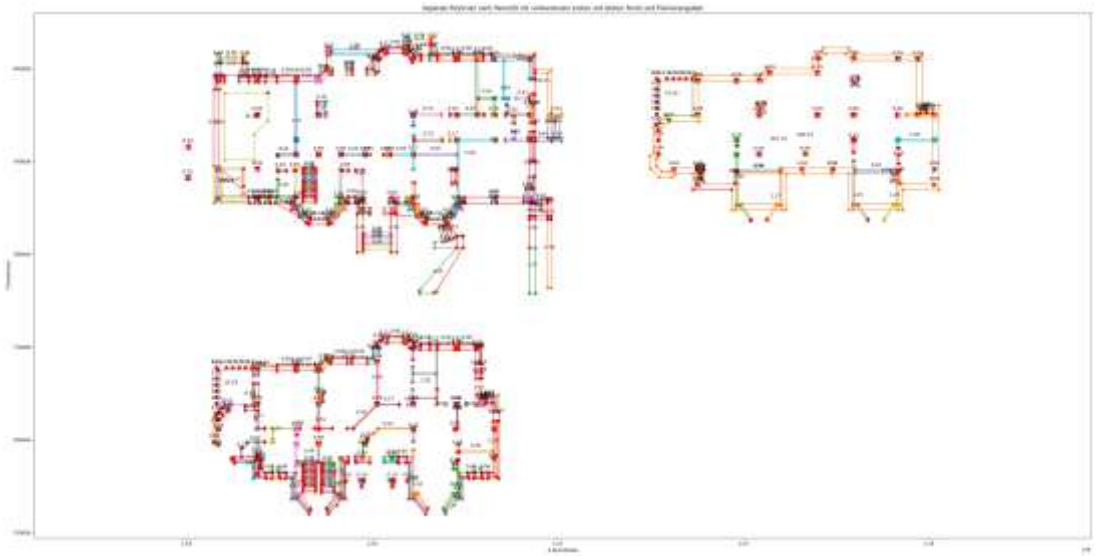


Abb. 6.4-9 LLM hat einen Code hinzugefügt, der die Flächenwerte für jede Polylinie nimmt und sie dem Bild mit Linienvisualisierung hinzufügt.

Dann wandeln wir jede Polylinie in eine horizontale Linie um, fügen eine parallele Linie in einer Höhe von 3000 mm hinzu und verbinden sie zu einer einzigen Ebene, um auf diese Weise die Anordnung der Oberflächen der Wandelemente darzustellen:

> Sie müssen alle Elemente aus der Spalte "Layer" mit dem Wert "wall" nehmen. Nehmen Sie diese IDs als Liste aus der Spalte "ID" und suchen Sie diese IDs aus dem gesamten Datenrahmen in der Spalte "ParentID". Alle Elemente sind Linien, die zu einer einzigen Polylinie kombiniert werden. Jede Linie hat eine andere x- und y-Geometrie als der erste Punkt in der Spalte "Punkt". Sie müssen jede Polylinie der Reihe nach nehmen und vom Punkt 0,0 horizontal die Länge jedes Segments der Polylinie zeichnen. die Länge jedes Segments der Polylinie in eine Linie. Zeichnen Sie dann genau die gleichen Linien, nur 3000 höher, und verbinden Sie alle Punkte in einer Ebene.

Die LLM-Antwort wird einen Code ausgeben, mit dem Sie Wandzeichnungen in der Ebene erstellen können:

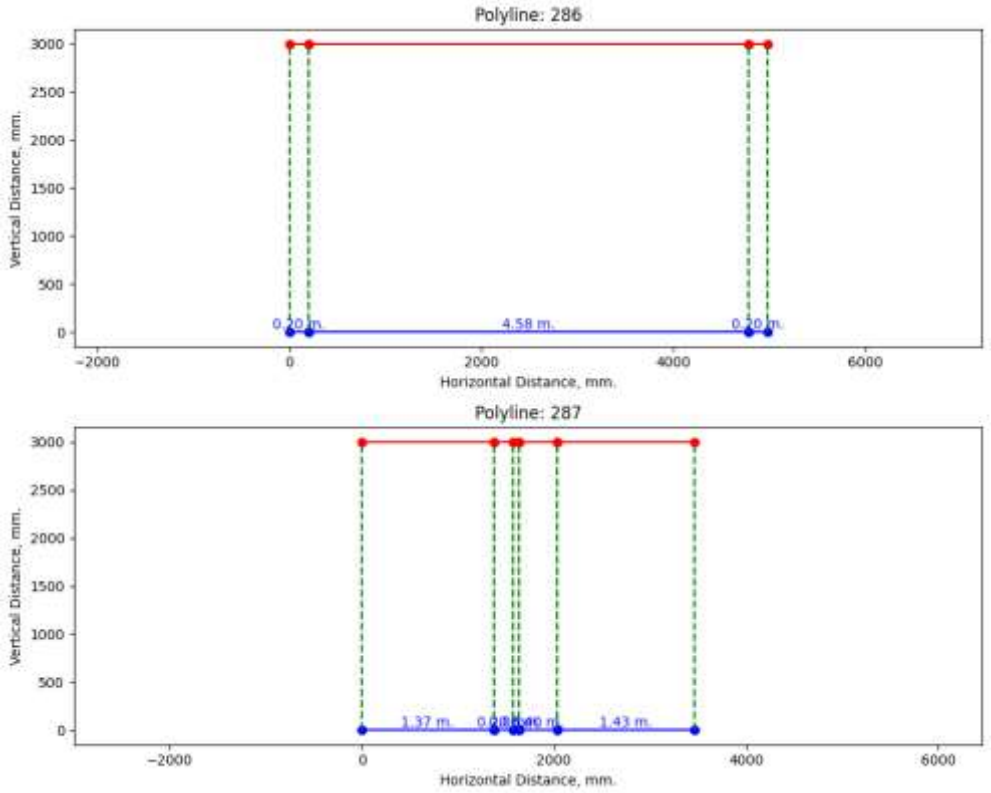


Abb. 6.4-10 Wir verwandeln jede Polylinie mit Hilfe von Prompts in ein Layout, das die Wandflächen direkt im LLM-Chat visualisiert.

Gehen wir nun von der 2D-Projektion zur 3D-Projektion über und modellieren wir Wände aus flachen Linien, indem wir die oberen und unteren Schichten von Polylinien verbinden:

> Visualisieren Sie Wandelemente in 3D, indem Sie Polylinien in den Höhen z = 0 und z = 3000 mm verbinden. Erstellen Sie eine geschlossene Geometrie, die die Wände des Gebäudes darstellt. Verwenden Sie das Matplotlib-3D-Diagrammwerkzeug.

- LLM wird eine interaktive 3D -Grafik erzeugen, in der jede Polylinie als eine Reihe von Ebenen dargestellt wird. Der Benutzer kann sich mit einer Computermaus frei zwischen den Elementen bewegen und das Modell im 3D-Modus erkunden, indem er den Code aus dem Chat in die IDE kopiert:

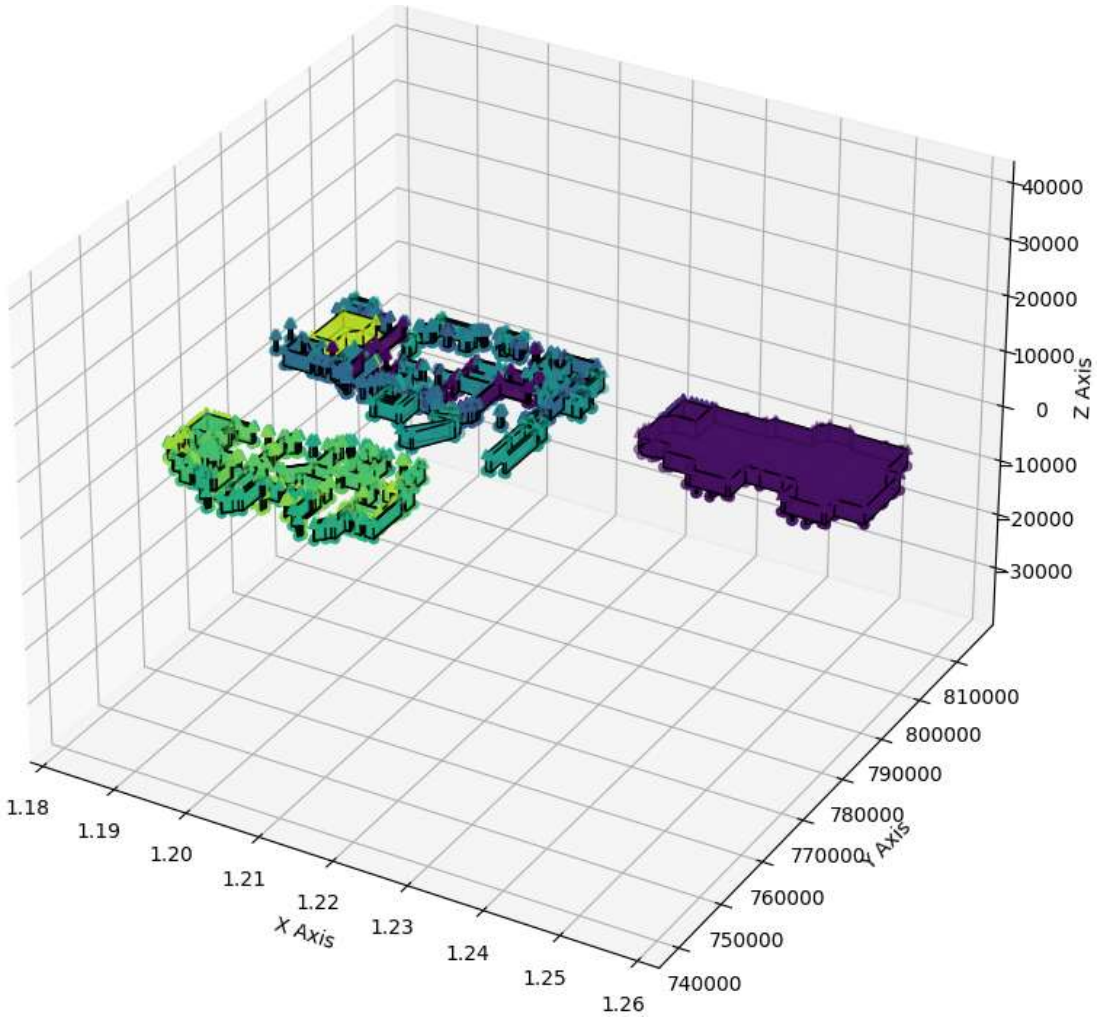


Abb. 6.4-11 LLM half bei der Erstellung von Code [129], um flache Zeichnungslinien in einer 3D-Ansicht zu visualisieren, die im 3D-Viewer innerhalb der IDE erkundet werden kann.

Um eine logische und reproduzierbare Pipeline aufzubauen - von der anfänglichen Konvertierung und dem Laden der DWG -Datei bis hin zum Endergebnis - empfiehlt es sich, den generierten LLM -Block Code nach jedem Schritt in die IDE zu kopieren. Auf diese Weise können Sie das Ergebnis nicht nur im Chat überprüfen, sondern es auch sofort in Ihrer Entwicklungsumgebung ausführen. Auf diese Weise können Sie den Prozess sequentiell aufbauen und ihn nach Bedarf debuggen und anpassen.

Sie finden den vollständigen Pipeline Code aller Fragmente (Abb. 6.4-8 bis Abb. 6.4-11) zusammen mit Beispielabfragen auf der Kaggle-Plattform.com, indem Sie nach "DWG Analyse with ChatGPT | DataDriven-Construction" suchen. [129]. Auf Kaggle können Sie nicht nur den Code und die verwendeten Prompts an-

sehen, sondern auch die gesamte Pipeline mit den ursprünglichen DWG-Datenrahmen in der Cloud kostenlos kopieren und testen, ohne zusätzliche Software oder die IDE selbst installieren zu müssen.

Der in diesem Kapitel vorgestellte Ansatz ermöglicht es Ihnen, die Prüfung, Verarbeitung und Erstellung von Dokumenten auf der Grundlage von DWG -Projekten vollständig zu automatisieren. Die entwickelte Pipeline eignet sich sowohl für die Verarbeitung einzelner Zeichnungen als auch für die Stapelverarbeitung von Dutzenden, Hunderten und Tausenden von DWG-Dateien mit automatischer Erstellung der erforderlichen Berichte und Visualisierungen für jedes Projekt.

Der Prozess kann sequentiell und transparent organisiert werden: zunächst werden die Daten aus der CAD -Datei automatisch in das XLSX-Format konvertiert, dann in einen Datenrahmen geladen, gefolgt von Gruppierung, Prüfung und Ergebnisgenerierung - all dies wird in einem einzigen Jupyter-Notebook oder Python -Skript in jeder gängigen IDE implementiert. Bei Bedarf kann der Prozess durch die Integration mit Projektdokumentations-Managementsystemen leicht erweitert werden: CAD-Dateien können automatisch nach bestimmten Kriterien abgerufen werden, die Ergebnisse können an das Speichersystem zurückgegeben werden und die Benutzer können per E-Mail oder Messenger benachrichtigt werden, wenn die Ergebnisse bereitstehen.

> Die Verwendung von LLM chats und Agenten für die Arbeit mit Entwurfsdaten reduziert die Abhängigkeit von spezialisierten CAD -Programmen und ermöglicht die Analyse und Visualisierung von architektonischen Entwürfen ohne die Notwendigkeit einer manuellen Interaktion mit der Schnittstelle - ohne Mausklicks und das Erinnern an eine komplexe Menüführung.

Mit jedem Tag, der vergeht, wird die Baubranche mehr und mehr von LLM, granular strukturierten Daten, DataFrames und kolumnaren Datenbanken hören. Vereinheitlichte zweidimensionale DataFrames, die aus verschiedenen Datenbanken und CAD-Formaten gebildet werden, werden der ideale Treibstoff für moderne Analysewerkzeuge sein, die von Spezialisten in anderen Branchen aktiv genutzt werden.

Der Automatisierungsprozess selbst wird erheblich vereinfacht - statt die API geschlossener Nischenprodukte zu studieren und komplexe Skripte zu schreiben, um Parameter zu analysieren oder zu transformieren, wird es nun ausreichen, eine Aufgabe in Form einer Reihe einzelner Textbefehle zu formulieren, die in die erforderliche Pipeline oder den Workflow-Prozess für die gewünschte Programmiersprache eingefügt werden, die kostenlos auf fast jedem Gerät läuft. Kein Warten mehr auf neue Produkte, Formate, Plug-ins oder Updates von CAD- (BIM-) Toolanbietern. Ingenieure und Bauherren werden in der Lage sein, selbstständig mit Daten zu arbeiten, indem sie einfache, kostenlose und leicht verständliche Werkzeuge verwenden, die von LLM Chats und Agenten unterstützt werden.

Nächste Schritte: Übergang von geschlossenen Formaten zu offenen Daten

Bei der Arbeit mit den Designdaten der Zukunft ist es unwahrscheinlich, dass jemand wirklich die geometrischen Kerne proprietärer Tools verstehen oder Hunderte von inkompatiblen Formaten mit denselben Informationen lernen muss. Ohne zu verstehen, warum der Übergang zu offenen strukturierten Daten wichtig ist, ist es jedoch schwierig, für den Einsatz neuer kostenloser Tools, offener Daten und Ansätze zu argumentieren, die wahrscheinlich nicht von Softwareanbietern gefördert werden.

In diesem Kapitel haben wir die wichtigsten Merkmale von CAD (BIM)-Daten, ihre Grenzen und Möglichkei-

ten erörtert und festgestellt, dass Ingenieure und Designer trotz der Marketingversprechen der Anbieter täglich mit Schwierigkeiten bei der Extraktion, Übertragung und Analyse von Konstruktionsdaten konfrontiert. Das Verständnis der Architektur dieser Systeme und das Kennenlernen von alternativen Ansätzen - basierend auf offenen Formaten und Automatisierung durch LLM - kann das Leben selbst für einen einzelnen Fachmann, ganz zu schweigen von Unternehmen, sehr viel einfacher machen. Um diesen Teil zusammenzufassen, lohnt es sich, die wichtigsten praktischen Schritte hervorzuheben, die Ihnen helfen werden, die in diesem Abschnitt besprochenen Ansätze auf Ihre täglichen Aufgaben anzuwenden:

- Erweitern Sie Ihr Toolkit für die Arbeit mit Projektdaten
  - ☐ Informieren Sie sich über verfügbare Plug-Ins und Dienstprogramme zur Extraktion von Daten aus den von Ihnen verwendeten CAD-Systemen - (BIM-)
  - ☐ Untersuchen Sie verfügbare SDKs und APIs, die es Ihnen ermöglichen, die Datenextraktion aus geschlossenen Formaten zu automatisieren, ohne dass Sie manuell spezielle Software öffnen müssen.
  - ☐ Beherrschen von Grundkenntnissen im Umgang mit offenen nichtparametrischen Geometrieformaten (OBJ, glTF, USD, DAE) und entsprechenden Open-Source-Bibliotheken
  - ☐ Versuchen Sie, ein System für die Speicherung von Projekt-Metadaten getrennt von der Geometrie außerhalb von CAD (BIM) Lösungen zu finden, um die Analyse und Integration mit anderen Systemen zu vereinfachen
  - ☐ Verwenden Sie LLM, um Datenkonvertierungsprobleme zwischen Formaten zu automatisieren.
- Erstellen Sie Ihre eigenen Prozesse für den Umgang mit Projektinformationen
  - ☐ Beginnen, Aufgaben und Modellierungsanforderungen durch Parameter und deren Werte in einfachen und strukturierten Formaten zu beschreiben
  - ☐ Erstellen einer persönlichen Bibliothek mit Skripten oder Codeblöcken für häufig ausgeführte Vorgänge
- Fördern Sie die Verwendung offener Standards in Ihrer Arbeit
  - ☐ Aufforderung an Kollegen und Partner, Daten in offenen Formaten auszutauschen, die nicht durch das Ökosystem der Softwareanbieter eingeschränkt sind
  - ☐ Aufzeigen der Vorteile der Verwendung strukturierter Daten anhand konkreter Beispiele
  - ☐ Diskussionen über Probleme mit geschlossenen Formaten und mögliche Lösungen anstoßen

Selbst wenn Sie die Politik Ihres Unternehmens in Bezug auf CAD - (BIM-) Plattformen nicht ändern können, wird ein persönliches Verständnis der Grundsätze der Arbeit mit Projektdaten in offenen Formaten es Ihnen ermöglichen, die Effizienz Ihrer Arbeit erheblich zu steigern. Indem Sie Ihre eigenen Werkzeuge und Methoden für die Extraktion und Umwandlung von Daten aus verschiedenen Formaten entwickeln, optimieren Sie nicht nur Ihre Arbeitsabläufe, sondern gewinnen auch die Flexibilität, um die Einschränkungen von Standardsoftwarelösungen zu umgehen.

VII TEIL

DATENGESTEUERTE ENTSCHEIDUNGSFINDUNG, ANALYTIK, AUTOMATISIERUNG UND MASCHINELLES LERNEN

Der siebte Teil befasst sich mit Datenanalyse und Prozessautomatisierung in der Bauindustrie. Es wird erörtert, wie Daten zur Grundlage für die Entscheidungsfindung werden, und es werden die Grundsätze der Visualisierung von Informationen für eine effektive Analyse erläutert. Ausführlich werden Key Performance Indicators (KPI), Methoden zur Bewertung des Return on Investment (ROI) und die Erstellung von Dashboards zur Projektüberwachung beschrieben. Besonderes Augenmerk wird auf ETL-Prozesse (Extract, Transform, Load) und deren Automatisierung mit Hilfe von Pipelines (Pipeline) gelegt, um disparate Daten in strukturierte Informationen für die Analyse zu verwandeln. Workflow-Orchestrierungstools wie Apache Airflow, Apache NiFi und n8n, die den Aufbau automatisierter Datenpipelines ohne tiefgreifende Programmierkenntnisse ermöglichen, werden diskutiert. Große Sprachmodelle (Large Language Models, LLMs) und ihre Verwendung zur Vereinfachung der Datenanalyse und zur Automatisierung von Routineaufgaben spielen eine wichtige Rolle

KAPITEL 7.1.
DATENANALYTIK UND DATENGESTÜTZTE ENTSCHEIDUNGSFINDUNG

Nach den Schritten des Sammelns, Strukturierens, Bereinigens und Verifizierens der Informationen ist ein kohärenter und analysierbarer Datensatz entstanden. In den vorangegangenen Teilen des Buches ging es um die Systematisierung und Strukturierung heterogener Quellen - von PDF-Dokumenten und Textaufzeichnungen von Sitzungen bis hin zu CAD-Modellen und Geometriedaten. Der Prozess der Überprüfung und des Abgleichs der Informationen mit den Anforderungen verschiedener Systeme und Klassifikatoren sowie die Beseitigung von Duplikaten und Inkonsistenzen wird ausführlich beschrieben.

Alle mit diesen Daten durchgeführten Berechnungen (dritter und vierter Teil des Buches) - von einfachen Transformationen bis zu Berechnungen von Zeit-, Kosten- und ESG-Indikatoren (fünfter Teil) - sind aggregierte Analyseaufgaben. Sie bilden die Grundlage, um den aktuellen Stand eines Projekts zu verstehen, seine Parameter zu bewerten und dann Entscheidungen zu treffen. Im Ergebnis werden die Daten durch die Berechnungen von einer Reihe disparater Datensätze zu einer handhabbaren Ressource, die in der Lage ist, wichtige Geschäftsfragen zu beantworten.

In den vorangegangenen Kapiteln wurden die Datenerfassungs- und Qualitätskontrollprozesse für die Verwendung in typischen Geschäftsfällen und Prozessen, die für die Bauindustrie spezifisch sind, detailliert beschrieben. Die Analytik in diesem Kontext ähnelt in vielerlei Hinsicht den Anwendungen in anderen Branchen, weist aber eine Reihe von Besonderheiten auf.

In den folgenden Kapiteln wird der Prozess der Datenanalyse detailliert beschrieben, einschließlich der Automatisierungsschritte von der anfänglichen Erfassung von Informationen und ihrer Umwandlung bis hin zu ihrer anschließenden Übertragung auf Zielsysteme und Dokumente. Zunächst wird ein theoretischer Teil vorgestellt, der sich auf ausgewählte Aspekte der datengesteuerten Entscheidungsfindung konzentriert. In den folgenden Kapiteln folgt dann der praktische Teil, der sich auf die Automatisierung und den Aufbau der ETL -Pipeline bezieht.

Daten als Ressource für die Entscheidungsfindung

Die datengesteuerte Entscheidungsfindung ist häufig ein iterativer Prozess und beginnt mit der systematischen Sammlung von Informationen aus einer Vielzahl von Informationsquellen. Wie ein natürlicher Kreislauf fallen einzelne Datenelemente und ganze Informationssysteme allmählich zu Boden und sammeln sich in den Informationsspeichern der Unternehmen an (Abb. 1.3-2). Im Laufe der Zeit verwandeln sich diese Daten, wie herabgefallene Blätter und Äste, in wertvolles Material. Das Myzel der Dateningenieure und -analysten organisiert und bereitet Informationen für die künftige Nutzung vor und verwandelt gefallene Daten und Systeme in wertvollen Kompost, aus dem neue Triebe und neue Systeme wachsen (Abb. 1.2-5).

Der Trend zur weit verbreiteten Nutzung von Analytik in verschiedenen Branchen markiert den Beginn einer neuen Ära, in der die Arbeit mit Daten zur Grundlage der beruflichen Tätigkeit wird (Abb. 7.1-1). Es ist wichtig, dass sich die Fachleute der Baubranche auf diese Veränderungen einstellen und bereit sind, in eine neue Ära einzutreten - die Ära der Daten und der Analytik

Das manuelle Verschieben von Daten zwischen Tabellen und die manuelle Durchführung von Berechnungen gehören allmählich der Vergangenheit an und machen der Automatisierung, der Datenflussanalyse, der Analytik und dem maschinellen Lernen Platz. Diese Werkzeuge werden zu Schlüsselelementen moderner Entscheidungsunterstützungssysteme.

In McKinseys Buch "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] wird eine im Jahr 2022 durchgeführte Studie mit 1.330 leitenden Angestellten aus verschiedenen Regionen, Branchen und Funktionsbereichen zitiert [130] zitiert eine Studie, die 2022 mit 1.330 Führungskräften aus verschiedenen Regionen, Branchen und Funktionsbereichen durchgeführt wurde. Demnach nutzen 70 Prozent der Führungskräfte fortschrittliche Analysen, um ihre eigenen Ideen zu entwickeln, und 50 Prozent setzen künstliche Intelligenz ein, um Entscheidungsprozesse zu verbessern und zu automatisieren.

Abb. 7.1-1 Die Datenanalyse und die Analytik sind die wichtigsten Instrumente zur Beschleunigung der Entscheidungsfindung in einem Unternehmen.

Die Datenanalyse durchdringt wie das sich ausbreitende Myzel den Humus vergangener Entscheidungen, hilft dabei, die einzelnen Systeme zu verbinden und führt die Manager zu wertvollen Erkenntnissen. Dieses Wissen nährt, wie die Nährstoffe von verrotteten Datensystembäumen, neue Entscheidungen im Unternehmen und führt zu effektiven Veränderungen und qualitativem Informationswachstum, wie neue Triebe und Sprossen, die aus einem reichen und gesunden Boden hervorgehen (Abb. 1.2-5).

Zahlen haben eine wichtige Geschichte zu erzählen. Sie zählen darauf, dass Sie ihnen eine klare und überzeugende Stimme geben [131].

- Stephen Few, Experte für Datenvisualisierung

In mittleren und kleinen Unternehmen ist die Extraktion und Aufbereitung von Informationen für weitere Analysen heute ein äußerst arbeitsintensiver Prozess (Abb. 7.1-2), vergleichbar mit dem Kohlebergbau des 18. Jahrhunderts. Bis vor kurzem war die Arbeit der Datengewinnung und -aufbereitung eher Abenteurern vorbehalten, die in einer hochspezialisierten Nische mit einem kleinen und begrenzten Satz von Werkzeugen für die Arbeit mit verschiedenen Arten von Daten aus unstrukturierten, locker strukturierten, gemischten und geschlossenen Quellen arbeiteten.

Entscheidungsträger und Manager sind oft unerfahren im Umgang mit heterogenen Daten und Systemen, müssen aber dennoch Entscheidungen auf der Grundlage dieser Daten treffen. Infolgedessen glich die datengestützte Entscheidungsfindung in der modernen Bauindustrie in den letzten Jahrzehnten weniger einem automatisierten Prozess als vielmehr der mehrtägigen Handarbeit eines Bergmanns in den frühen Kohleminen.

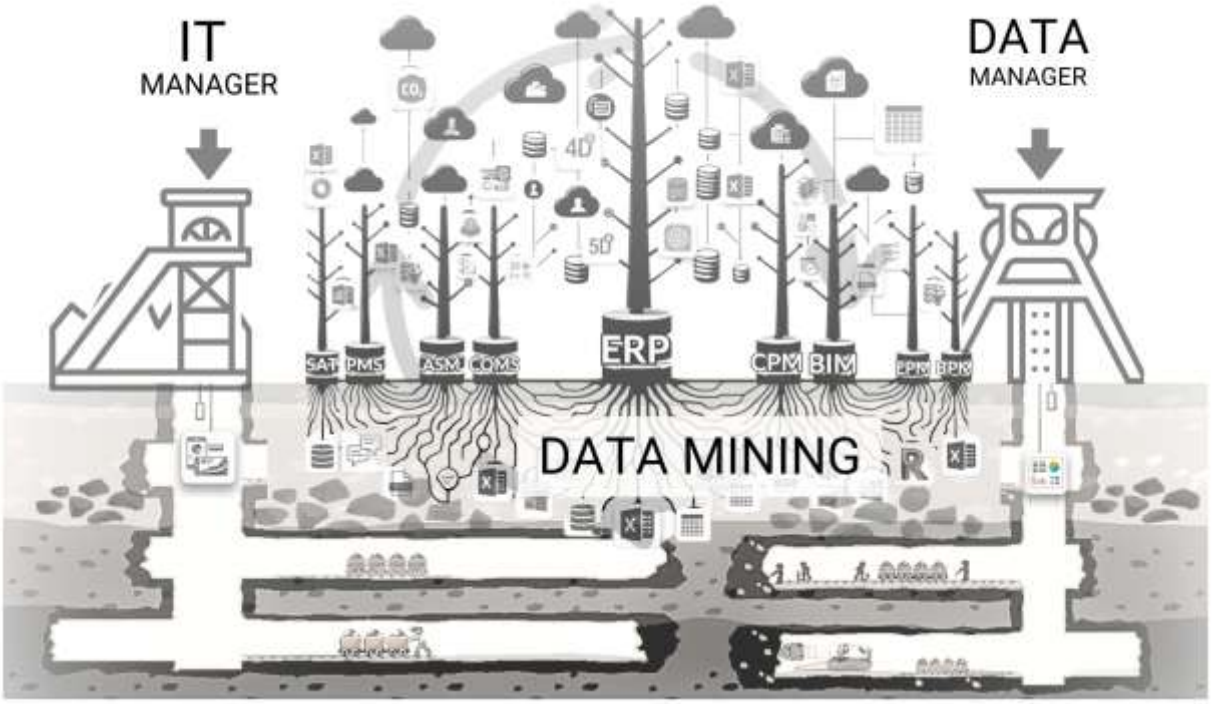


Abb. 7.1-2 Im Data-Mining-Prozess durchlaufen die Experten einen komplexen Weg der Datenaufbereitung - von der Bereinigung bis zur Strukturierung für die anschließende Analyse.

Obwohl die modernen Methoden der Datenextraktion in der Bauindustrie sicherlich fortschrittlicher sind als die primitiven Techniken der Bergleute des 12. Jahrhunderts, handelt es sich immer noch um eine komplexe und risikoreiche Aufgabe, die erhebliche Ressourcen und Fachkenntnisse erfordert, die sich nur große Unternehmen leisten können. Die Extraktion und Analyse von Daten aus dem angesammelten Erbe vergangener Projekte wurde bis vor kurzem überwiegend von großen, technologisch fortschrittlichen Unternehmen durchgeführt, die seit Jahrzehnten kontinuierlich Daten sammeln und speichern

Früher spielten technologisch ausgereifte Unternehmen, die seit Jahrzehnten Daten sammeln, die führende Rolle in der Analytik. Heute ändert sich die Situation: Der Zugang zu Daten und Datenverarbeitungswerkzeugen wird demokratisiert - ehemals komplexe Lösungen sind nun für jeden kostenlos verfügbar.

Die Anwendung von Analysen ermöglicht es Unternehmen, genauere und fundiertere Entscheidungen zu treffen in Echtzeit. Die folgende Fallstudie veranschaulicht, wie historische Daten helfen können, finanziell solide Entscheidungen zu treffen:

- **Projektleiter** - "*Der Durchschnittspreis für Beton in der Stadt liegt jetzt bei 82€ /m³, wir haben 95 €/m³ im Kostenvoranschlag.*"
- **Schätzer** - "*Bei früheren Projekten lag die Überschreitung bei etwa 15%, also habe ich mich zurückgehalten*".
- **Datenmanager oder Kontrollingenieur auf Kundenseite** - "*Schauen wir uns die Analysen für die letzten drei Ausschreibungen an*".

Nach der Analyse des DataFrame von vergangenen Projekten erhalten wir:

- **Durchschnittlicher tatsächlicher Einkaufspreis:** 84,80 /m³€
- **Durchschnittliche Ausgabenüberschreitungsquote:** +4,7%
- **Empfohlene Rate im Kostenvoranschlag:** ~ 85 /m³€

Eine solche Entscheidung basiert nicht mehr auf subjektiven Gefühlen, sondern auf spezifischen historischen Statistiken, was dazu beiträgt, Risiken zu verringern und die Aussagekraft des Angebots zu erhöhen. Die Analyse von Daten aus vergangenen Projekten wird zu einer Art "organischem Dünger", aus dem neue, genauere Lösungen keimen.

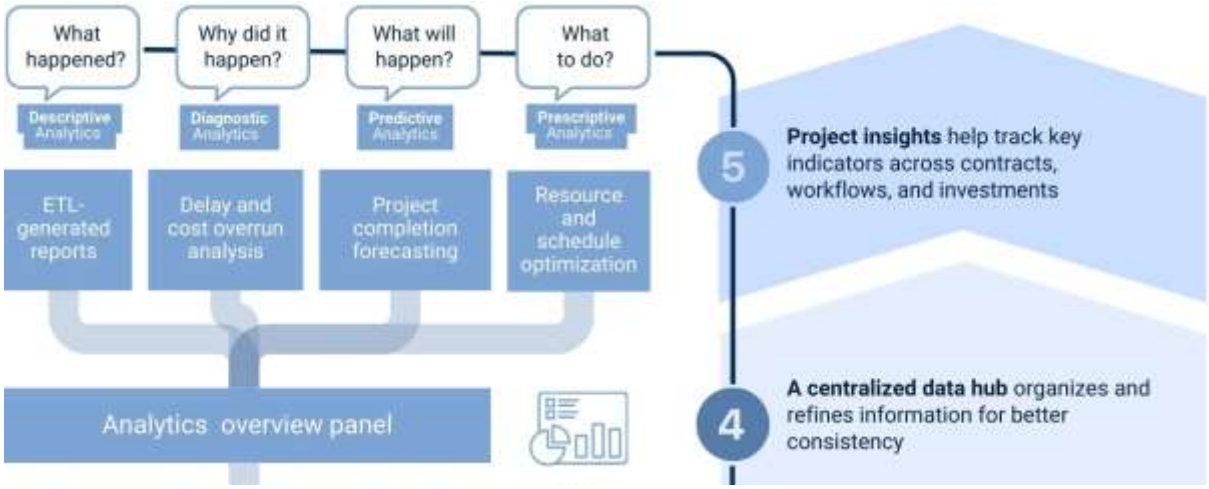


Abbildung 7.1-3 Die Datenanalyse beantwortet drei Schlüsselfragen: Was ist passiert, warum ist es passiert, und was sollte als nächstes getan werden.

Entscheidungsträger und Manager sehen sich häufig mit der Notwendigkeit konfrontiert, mit heterogenen Daten und Systemen zu arbeiten, ohne über ausreichende technische Kenntnisse zu verfügen. In solchen Situationen ist die Visualisierung, einer der ersten und wichtigsten Schritte im Analyseprozess, eine wichtige Hilfe für das Verständnis der Daten. Sie ermöglicht es, Informationen in einer visuellen und verständlichen Form darzustellen.

Visualisierung von Daten: der Schlüssel zum Verständnis und zur Entscheidungsfindung

In der heutigen Bauindustrie, in der Projektdaten durch Komplexität und mehrstufige Strukturen gekenn-

zeichnet sind, spielt die Visualisierung eine Schlüsselrolle. Die Visualisierung von Daten ermöglicht es Projektleitern und Ingenieuren, komplexe Muster und Trends zu erkennen, die in großen, heterogenen Datenmengen verborgen sind.

> Die Visualisierung von Daten erleichtert das Verständnis des Projektstatus: Ressourcenzuweisung, Kostentrends oder Materialverbrauch. Grafiken und Diagramme machen komplexe und trockene Informationen zugänglich und verständlich und ermöglichen es Ihnen, schnell die wichtigsten Bereiche zu identifizieren, die Aufmerksamkeit erfordern, und potenzielle Probleme zu erkennen.

Die Visualisierung von Daten erleichtert nicht nur die Interpretation von Informationen, sondern ist ein entscheidender Schritt im Analyseprozess und bei der fundierten Entscheidungsfindung des Managements, in*dem sie hilft, di*e Fragen *"Was i*st *passiert?*" und "Wie ist es passiert?" zu beantworten (Abb. 2.2-5). (Abb. 2.2-5).

> *Grafiken sind visuelle Hilfsmittel zur Lösung von logischen Problemen [132].*
>
> - Jacques Bertin, "Grafiken und grafische Informationsverarbeitung

Bevor sie wichtige Entscheidungen treffen, greifen Projektmanager eher auf visuelle Darstellungen von Daten zurück als auf trockene und schwer zu interpretierende Zahlen aus Tabellenkalkulationen oder Textnachrichten.

> Daten ohne Visualisierung sind wie Baumaterialien, die wahllos auf einer Baustelle verstreut sind: Ihr Potenzial ist unklar. Erst wenn sie klar visualisiert werden, wie ein Haus aus Ziegeln und Beton, wird ihr Wert deutlich. Solange das Haus nicht gebaut ist, lässt sich nicht sagen, ob aus dem Materialhaufen eine kleine Hütte, eine Luxusvilla oder ein Wolkenkratzer entstehen wird.

Unternehmen verfügen über Daten aus verschiedenen Systemen (Abb. 1.2-4 bis Abbildung 2.1-10), Finanztransaktionen und umfangreiche Textdaten. Die Nutzung dieser Daten zum geschäftlichen Nutzen ist jedoch oft eine Herausforderung. In solchen Situationen wird die Visualisierung zu einem wichtigen Instrument, um die Bedeutung der Daten zu vermitteln und die Informationen in Form von Dashboards, Grafiken und Diagrammen so zu präsentieren, dass sie für jeden Experten verständlich sind.

> Die PwC-Studie "What Students Need to Succeed in a Fast Changing Business World" (2015) unterstreicht [9], dass erfolgreiche Unternehmen über die Datenanalyse hinausgehen und aktiv interaktive Visualisierungstools wie Diagramme, Infografiken und analytische Dashboards zur Unterstützung der Entscheidungsfindung einsetzen. Dem Bericht zufolge hilft die Datenvisualisierung den Kunden, die Geschichte zu verstehen, die die Daten durch Grafiken, Diagramme, Dashboards und interaktive Datenmodelle erzählen.

Der Prozess der Umwandlung von Informationen in visuelle grafische Formen wie Diagramme, Schaubilder und Diagramme verbessert das Verständnis und die Interpretation von Daten durch das menschliche Gehirn

(Abb. 7.1-4). Dies ermöglicht es Projektmanagern und Analytikern, komplexe Szenarien schneller zu bewerten und fundierte Entscheidungen zu treffen, die auf visuell erkennbaren Trends und Mustern statt auf Intuition beruhen.

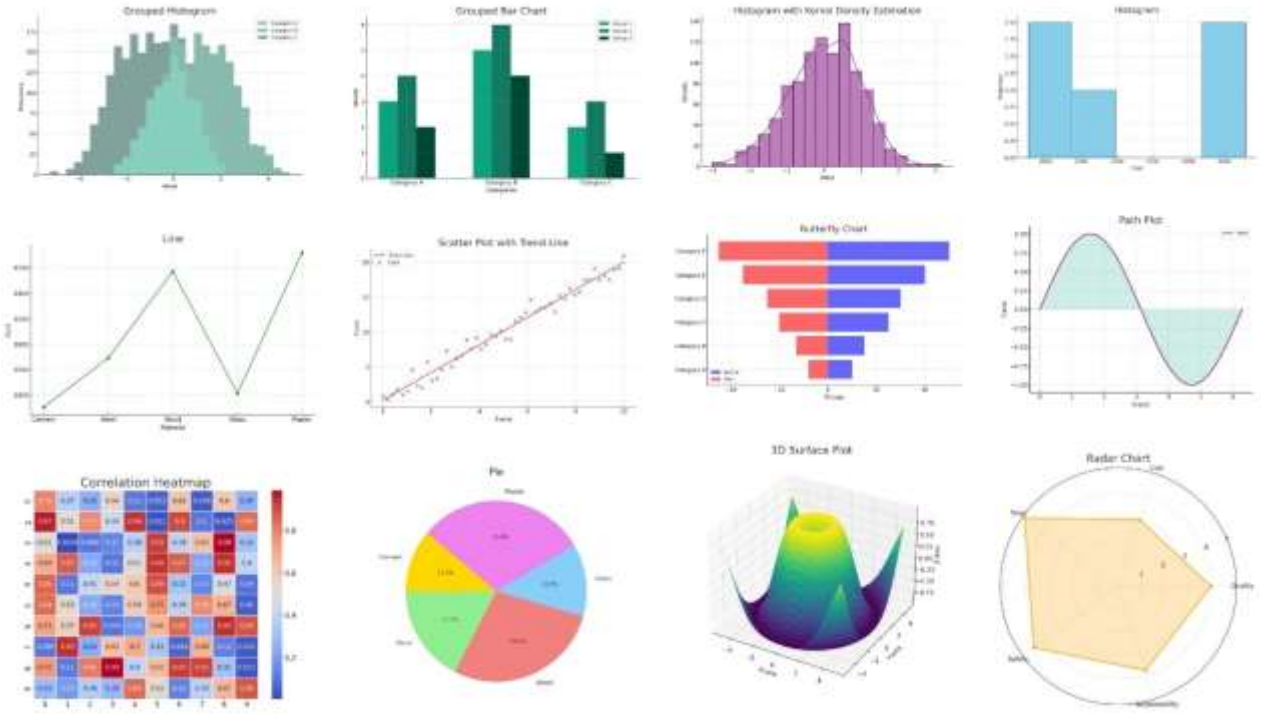


Abb. 7.1-4 Verschiedene Arten der Visualisierung sollen dem menschlichen Gehirn helfen, die trockenen Informationen von Zahlen besser zu verstehen und zu verarbeiten.

Die Erstellung von Visualisierungen aus Daten und die Verwendung verschiedener kostenloser Visualisierungsbibliotheken werden im nächsten Kapitel über ETL -Prozesse ausführlicher behandelt.

Die Visualisierung wird zu einem integralen Bestandteil der Arbeit mit Daten in der Bauindustrie - sie hilft nicht nur, Daten zu "sehen", sondern auch, ihre Bedeutung im Kontext von Managementaufgaben zu verstehen. Damit die Visualisierung jedoch wirklich nützlich ist, muss im Vorfeld festgelegt werden, was genau visualisiert werden soll und welche Metriken für die Bewertung der Projektleistung wirklich wichtig sind. An dieser Stelle kommen Leistungskennzahlen wie KPIs und ROI ins Spiel. Ohne sie besteht selbst bei den schönsten Dashboards die Gefahr, dass sie nur "Informationsrauschen" sind.

KPIs und ROI

In der heutigen Bauindustrie spielen das Management von Leistungsindikatoren (KPI und ROI) und deren Visualisierung durch Berichte und Dashboards eine Schlüsselrolle bei der Verbesserung der Produktivität und der Effizienz des Projektmanagements.

Wie in jedem Unternehmen ist es auch im Baugewerbe notwendig, die Messgrößen, mit denen Erfolg, Rentabilität und Leistung gemessen werden, klar zu definieren. Bei der Beschaffung von Daten über verschiedene Prozesse muss ein datengesteuertes Unternehmen zunächst lernen, die wichtigsten **KPIs (Key Performance Indicators)** zu ermitteln - quantitative Messgrößen, die das Ausmaß der Erreichung strategischer und operativer Ziele widerspiegeln.

Zur Berechnung eines KPI wird in der Regel eine Formel verwendet (Abb. 7.1-5), die Ist- und Plan-Indikatoren enthält. Um beispielsweise einen individuellen KPI für ein Projekt, einen Mitarbeiter oder einen Prozess zu berechnen, teilen Sie die Ist-Leistung durch die geplante Leistung und multiplizieren das Ergebnis mit 100%.

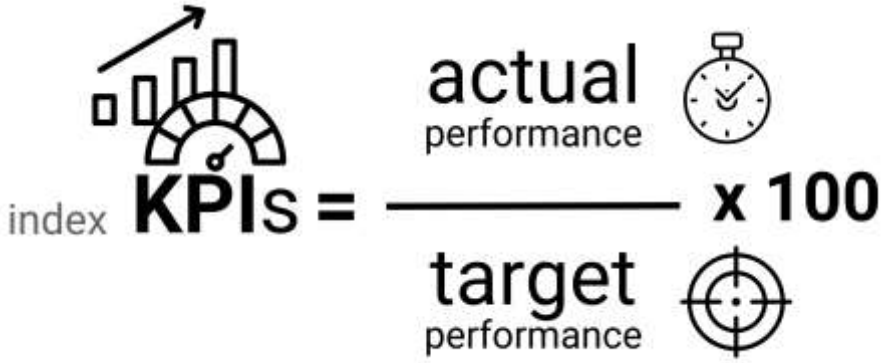


Abbildung 7.1-5 KPIs werden verwendet, um den Erfolg eines Projekts oder Prozesses bei der Erreichung der wichtigsten Ziele zu messen.

Auf der Ebene des Standorts können detailliertere KPIs verwendet werden:

- **Zeitplanung der wichtigsten Meilensteine** (Gründung, Installation, Fertigstellung) - ermöglicht die Kontrolle der Einhaltung der Arbeitspläne.
- **Prozentsatz der Materialüberschreitung** - hilft bei der Verwaltung der Beschaffung und der Minimierung von Verschwendung.
- **Anzahl der ungeplanten Maschinenstillstände** - wirkt sich auf Produktivität und Kosten aus.

Die Wahl der falschen Messgrößen kann zu falschen "*Was tun?*"-Entscheidungen führen (Abb. 2.2-5). Wenn sich ein Unternehmen beispielsweise nur auf die Kosten pro Quadratmeter konzentriert, aber die Kosten für Änderungen nicht berücksichtigt, können Einsparungen bei den Materialien zu schlechterer Qualität und höheren Kosten bei zukünftigen Projekten führen.

Bei der Festlegung von Zielen ist es wichtig, dass klar ist, was gemessen werden soll. Vage Formulierungen führen zu falschen Schlussfolgerungen und erschweren die Kontrolle. Sehen wir uns Beispiele für erfolgreiche und erfolglose KPIs im Bauwesen an.

Gute KPIs:

- *"Bis zum Ende des Jahres soll der Anteil der Renovierungsarbeiten um* 10 Prozent *reduziert werden*.
- "*Steigerung der Geschwindigkeit der Fassadenmontage um 15 Prozent ohne Qualitätseinbußen*

bis zum nächsten Quartal"

- "*Reduzierung der Maschinenstillstandszeiten um 20% durch Optimierung der Arbeitspläne bis Ende des Jahres*".

Diese Metriken sind eindeutig messbar und haben spezifische Werte und Zeitrahmen.

Schlechte KPIs:

- "*Wir werden schneller bauen*" (Wie viel schneller? Was bedeutet "schneller"?).
- "*Wir werden die Qualität der konkreten Arbeit verbessern*" (Wie genau wird die Qualität gemessen?)
- "*Wir werden die Interaktion mit den Auftragnehmern vor Ort verbessern*" (Anhand welcher Kriterien wird die Verbesserung sichtbar?)

> Ein guter KPI ist ein KPI, der gemessen und objektiv bewertet werden kann. Im Bauwesen ist dies besonders wichtig, denn ohne klare Indikatoren ist es unmöglich, die Leistung zu überwachen und stabile Ergebnisse zu erzielen.

Neben dem KPI gibt es eine weitere Kennzahl zur Bewertung der Wirksamkeit von Investitionen: **ROI (Return on Investment)** - ein Indikator für die Kapitalrendite, der das Verhältnis zwischen Gewinn und investierten Mitteln wiedergibt. Der ROI hilft zu beurteilen, ob die Einführung neuer Methoden, Technologien oder Werkzeuge gerechtfertigt ist: von digitalen Lösungen und Automatisierung (z. B. Abbildung 7.3-2) bis hin zur Verwendung neuer Baumaterialien. Dieser Indikator hilft dabei, fundierte Entscheidungen über weitere Investitionen auf der Grundlage ihrer tatsächlichen Auswirkungen auf die Unternehmensrentabilität zu treffen

> Im Zusammenhang mit dem Management von Bauprojekten kann der ROI (Return on Investment) als einer der wichtigsten Leistungsindikatoren (KPI) verwendet werden, wenn das Ziel eines Unternehmens darin besteht, die Investitionsrendite eines Projekts, einer Technologie oder einer Prozessverbesserung zu messen. Wenn zum Beispiel eine neue Baumanagementtechnik eingeführt wird, kann der ROI zeigen, wie sehr sie die Rentabilität verbessert hat.

Die regelmäßige Messung von KPIs und ROI auf der Grundlage von Daten aus verschiedenen Quellen, wie Materialverbrauch, Arbeitsstunden und Kosten, ermöglicht es dem Projektmanagement, Ressourcen effektiv zu verwalten und schnelle Entscheidungen zu treffen. Die langfristige Speicherung dieser Daten ermöglicht es, zukünftige Trends zu analysieren und Prozesse zu optimieren.

Zur Visualisierung von KPIs, ROIs und anderen Metriken werden verschiedene Diagramme und Grafiken verwendet, die in der Regel zu Dashboards zusammengefasst werden.

Dashboards und Dashboards: Visualisierung von Metriken für ein effizientes Management

Zur Visualisierung von Indikatoren und Metriken wird eine Vielzahl von Diagrammen und Grafiken verwendet, die in der Regel zu Datenvitrinen und Dashboards zusammengefasst werden. Diese Dashboards bieten

einen zentralen Überblick über den Status eines Projekts oder von Teilen eines Projekts und zeigen Schlüsselindikatoren an (idealerweise in Echtzeit). Aktuelle und ständig aktualisierte Dashboards ermöglichen es dem Team, schnell auf Veränderungen zu reagieren.

> **Dashboards** sind Instrumente zur Visualisierung quantitativer Bewertungen, die für alle Projektteilnehmer leicht zugänglich und verständlich sind.

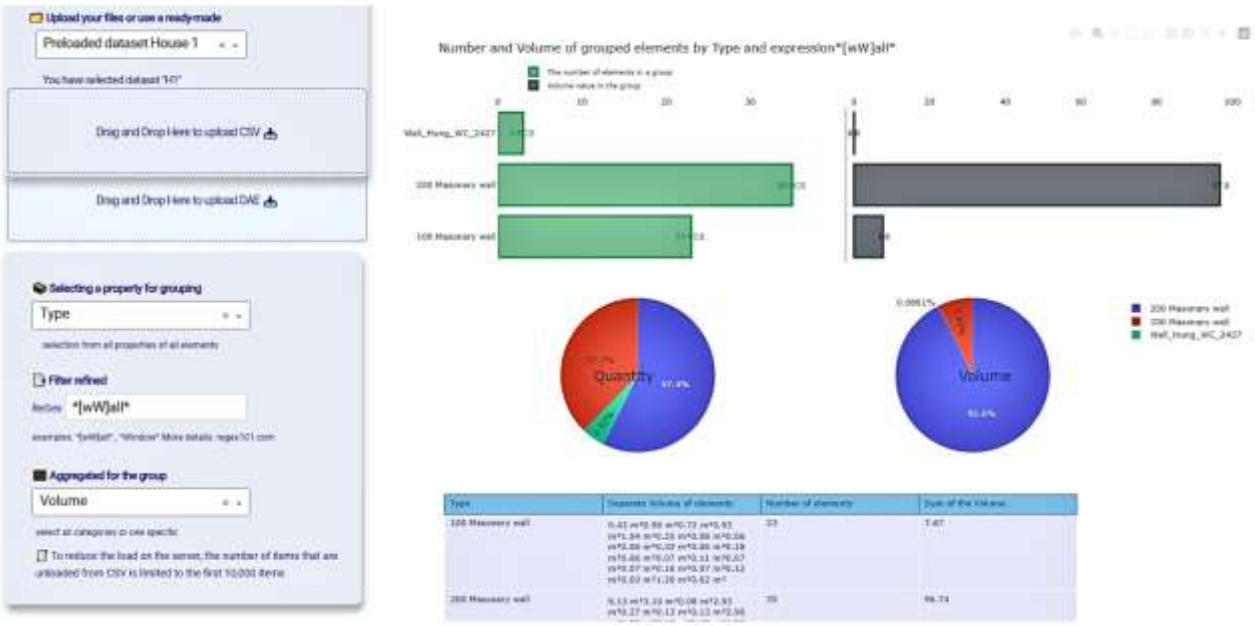


Abb. 7.1-6 Die Verwaltung von KPIs und deren Visualisierung durch Dashboards ist der Schlüssel zur Verbesserung der Projektproduktivität und -effizienz.

Hier sind einige Beispiele für gängige Tools, mit denen Sie Dashboards erstellen können:

- **Power BI** ist ein Tool von Microsoft zur Erstellung interaktiver Berichte und Dashboards.
- **Tableau und Google Data Studio** sind leistungsstarke Tools für die Visualisierung von Daten und die Erstellung von Dashboards, ohne dass Sie Code schreiben müssen.
- **Plotly** (Abb. 7.1-6, Abbildung 7.2-12) ist eine Bibliothek zur Erstellung interaktiver Diagramme, und Dash ist ein Framework zur Erstellung von Webanwendungen für die Datenanalyse. Sie können in Kombination verwendet werden, um interaktive Dashboards zu erstellen.
- **Zahlreiche Python-Bibliotheken** (Abb. 7.2-9 - Abbildung 7.2-11) - Python verfügt über zahlreiche quelloffene und kostenlose Bibliotheken für die Datenvisualisierung, wie z. B. Matplotlib, Seaborn, Plotly, Bokeh und andere. Diese können verwendet werden, um Diagramme zu erstellen und sie mit Hilfe von Frameworks wie Flask oder Django in eine Webanwendung zu integrieren.
- **JavaScript-Bibliotheken**: ermöglicht es Ihnen, interaktive Dashboards mit Open-Source-JavaScript-Bibliotheken wie D3.js oder Chart.js zu erstellen und in eine Webanwendung zu integrieren.

Um die KPIs zu bewerten und Dashboards zu erstellen, benötigen Sie aktuelle Daten und einen klaren Zeitplan für die Sammlung und Analyse der Informationen.

Generell bilden KPI, ROI und Dashboards in der Baubranche die Grundlage für einen analytischen Ansatz im Projektmanagement. Sie helfen nicht nur dabei, den aktuellen Status zu überwachen und zu bewerten, sondern liefern auch wertvolle Erkenntnisse für künftige Planungs- und Optimierungsprozesse - Prozesse, die direkt von der Interpretation der Daten und dem Stellen der richtigen und rechtzeitigen Fragen abhängen.

Datenanalyse und die Kunst, Fragen zu stellen

Die Interpretation der Daten ist die letzte Phase der Analyse, in der die Informationen einen Sinn ergeben und zu "sprechen" beginnen. Hier werden die Antworten auf die wichtigsten Fragen formuliert: *"Was ist zu tun?"* und *"Wie ist zu tun?"* (Abb. 2.2-5). In dieser Phase lassen sich die Ergebnisse zusammenfassen, Muster erkennen, Ursache-Wirkungs-Beziehungen herstellen und Schlussfolgerungen auf der Grundlage von Visualisierungen und statistischen Analysen ziehen.

> *Vielleicht ist die Zeit nicht mehr fern, in der man erkennen wird, dass man, um ein vollwertiger Bürger eines der großen neuen komplexen Weltstaaten zu werden, die sich jetzt entwickeln, genauso gut rechnen können muss, um in Durchschnittswerten, Maxima und Minima zu denken, wie man heute lesen und schreiben kann* [133].
>
> - Samuel S. Wilkes, zitiert in einer Präsidentenrede vor der American Statistical Association im Jahr 1951

> Laut dem von der britischen Regierung veröffentlichten Bericht "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) [83] kann die Implementierung von Analytik Daten und künstlicher Intelligenz (KI) die Projektmanagementprozesse erheblich verbessern, die Genauigkeit von Zeit- und Kostenprognosen erhöhen sowie Risiken und Unsicherheiten verringern. In dem Papier wird hervorgehoben, dass öffentliche Einrichtungen, die fortschrittliche Analysetools einsetzen, bei Infrastrukturinitiativen eine höhere Leistung erzielen.

Das moderne Baugeschäft, das in dem wettbewerbsintensiven und margenschwachen Umfeld der vierten industriellen Revolution operiert, kann mit militärischen Operationen verglichen werden. Hier hängen das Überleben und der Erfolg des Unternehmens von der Schnelligkeit der Beschaffung von Ressourcen und Qualitätsinformationen ab - und damit von rechtzeitigen und fundierten Entscheidungen (Abb. 7.1-7).

Wenn die Datenvisualisierung die "Intelligenz" ist, die den Überblick verschafft, dann ist die Datenanalyse die "Munition", die zum Handeln benötigt wird. Sie beantwortet die Fragen: *"Was ist zu tun?"* und *"Wie ist es zu tun?"* und bildet die Grundlage für die Erlangung eines Wettbewerbsvorteils auf dem Markt.

Analytik verwandelt disparate Daten in strukturierte und aussagekräftige Informationen, die als Grundlage für Entscheidungen dienen.

Die Aufgabe von Analysten und Managern besteht nicht nur darin, Informationen zu interpretieren, sondern auch fundierte Entscheidungen zu treffen, Trends zu erkennen, Beziehungen zwischen verschiedenen Datentypen zu bestimmen und sie entsprechend den Zielen und Besonderheiten des Projekts zu kategorisieren. Mit Hilfe von Visualisierungswerkzeugen und statistischen Analysemethoden verwandeln sie Daten in einen strategischen Wert für das Unternehmen.

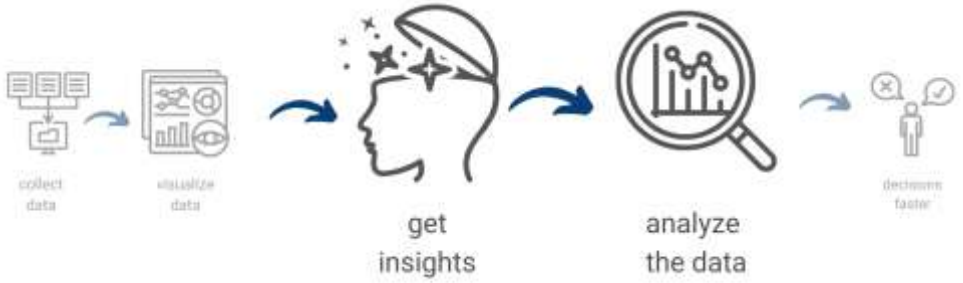


Abb. 7.1-7 Es ist die Datenanalyse, die die gesammelten Informationen letztlich zu einer Quelle für die Entscheidungsfindung macht.

Um im Analyseprozess wirklich fundierte Entscheidungen treffen zu können, muss man lernen, die Fragen, die an die Daten gestellt werden, richtig zu formulieren. Die Qualität dieser Fragen wirkt sich direkt auf die Tiefe der gewonnenen Erkenntnisse und damit auf die Qualität der Managemententscheidungen aus.

> *Die Vergangenheit existiert nur insofern, als sie in den Aufzeichnungen der Gegenwart präsent ist. Und was diese Aufzeichnungen darstellen, wird durch die Fragen bestimmt, die wir stellen. Es gibt keine andere Geschichte als diese [134].*
>
> - John Archibald Wheeler, Physiker 1982

Die Kunst, tiefgehende Fragen zu stellen und kritisch zu denken, ist eine wesentliche Fähigkeit bei der Arbeit mit Daten. Die meisten Menschen neigen dazu, einfache, oberflächliche Fragen zu stellen, deren Beantwortung wenig Aufwand erfordert. Eine echte Analyse beginnt jedoch mit aussagekräftigen und durchdachten Fragen, die verborgene Zusammenhänge und Ursache-Wirkungs-Beziehungen in Informationen aufdecken können, die möglicherweise hinter mehreren Ebenen der Argumentation verborgen sind.

> Laut der Studie "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135] erfordert eine erfolgreiche digitale Transformation Investitionen in Analysefähigkeiten, Change-Management-Programme und die Abstimmung von Geschäftszielen mit IT-Initiativen. Unternehmen, die eine datengesteuerte Kultur schaffen, sollten in Datenanalysekapazitäten investieren und Change-Management-Programme auflegen, um neue Denkweisen, Verhaltensweisen und Arbeitsweisen einzuführen.

Ohne Investitionen in die Entwicklung einer analytischen Kultur, die Verbesserung von Dateninstrumenten und die Ausbildung von Fachleuten werden Unternehmen weiterhin Gefahr laufen, Entscheidungen auf der Grundlage veralteter oder unvollständiger Informationen zu treffen - oder sich auf die subjektiven Meinungen von HiPPO-Managern zu verlassen (Abb. 2.1-9).

Die Erkenntnis der Relevanz und der Notwendigkeit, Analysen und Dashboards ständig zu aktualisieren, führt unweigerlich dazu, dass das Management die Bedeutung der Automatisierung von Analyseprozessen erkennt. Die Automatisierung erhöht die Geschwindigkeit der Entscheidungsfindung, verringert den Einfluss des menschlichen Faktors und gewährleistet die Relevanz der Daten. Mit dem exponentiellen Wachstum der Informationsmengen wird Geschwindigkeit nicht nur zu einem Wettbewerbsvorteil, sondern zu einem Schlüsselfaktor für nachhaltigen Erfolg.

Die Automatisierung von Datenanalyse- und Verarbeitungsprozessen im Allgemeinen ist untrennbar mit dem Thema ETL (Extract, Transform, Load) verbunden. So wie im Automatisierungsprozess Daten transformiert werden müssen, werden im ETL-Prozess Daten aus verschiedenen Quellen extrahiert, entsprechend den notwendigen Anforderungen transformiert und zur weiteren Verwendung in Zielsysteme geladen.

KAPITEL 7.2.
DATENFLUSS OHNE MANUELLEN AUFWAND: WARUM ETL BENÖTIGT WIRD

ETL-Automatisierung: Kostensenkung und Beschleunigung der Datenverarbeitung

Wenn die wichtigsten Leistungsindikatoren (KPIs) trotz steigender Datenmengen und wachsender Teamgröße nicht mehr wachsen, kommt die Unternehmensleitung unweigerlich zu der Erkenntnis, dass die Prozesse automatisiert werden müssen. Früher oder später wird diese Erkenntnis zum Anreiz für die Einführung einer komplexen Automatisierung, deren Hauptziel darin besteht, die Komplexität der Prozesse zu verringern, die Verarbeitung zu beschleunigen und die Abhängigkeit vom menschlichen Faktor zu reduzieren.

> Laut der McKinsey-Studie "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136] haben Unternehmen, die Streaming-Data-Architekturen einsetzen, einen erheblichen Vorteil, da sie Informationen in Echtzeit analysieren können. Streaming-Technologien ermöglichen die direkte Analyse von Echtzeitnachrichten und die Anwendung der vorausschauenden Wartung in der Fertigung durch die Analyse von Sensordaten in Echtzeit.

> Prozessvereinfachung ist Automatisierung, bei der traditionelle manuelle Funktionen durch Algorithmen und Systeme ersetzt werden.

Das Thema Automatisierung, oder besser gesagt "die Minimierung der Rolle des Menschen in der Datenverarbeitung", ist für jedes Unternehmen ein unumkehrbarer und hochsensibler Prozess. Fachleute in jedem Berufsfeld zögern oft, ihre Methoden und Feinheiten der Arbeit gegenüber anderen Optimierern vollständig offenzulegen, da sie sich des Risikos bewusst sind, in einem sich schnell entwickelnden technologischen Umfeld ihren Arbeitsplatz zu verlieren.

> *Wenn Sie sich Feinde machen wollen, versuchen Sie, die Dinge zu ändern [137].*
>
> - Woodrow Wilson, Rede vor einem Kongress von Geschäftsleuten, Detroit, 1916

Trotz der offensichtlichen Vorteile der Automatisierung haben viele Unternehmen immer noch einen hohen Anteil an manueller Arbeit in ihrer täglichen Praxis, insbesondere im Bereich der technischen Daten. Um die aktuelle Situation zu veranschaulichen, betrachten wir ein typisches Beispiel für die sequentielle Datenverarbeitung innerhalb solcher Prozesse.

Die manuelle Datenverarbeitung lässt sich am Beispiel der Interaktion mit Informationen aus CAD-Datenbanken veranschaulichen. Die traditionelle Datenverarbeitung ("manueller" ETL -Prozess) in CAD (BIM)-Abteilungen zur Erstellung von Attributtabellen oder zur Erstellung von Dokumentationen auf der Basis von Konstruktionsdaten läuft in folgender Reihenfolge ab (Abb. 7.2-1):

1. Manuelle **Extraktion (Extract**): Der Benutzer öffnet das Projekt manuell, indem er die CAD-

Anwendung (BIM) startet (Abb. 7.2-1 Schritt 1).

2. **Überprüfung**: Der nächste Schritt umfasst in der Regel die manuelle Ausführung mehrerer Plug-ins oder Hilfsanwendungen, um die Daten vorzubereiten und ihre Qualität zu bewerten (Abb. 7.2-1 Schritt 2-3).
3. Manuelle **Transformation (Transform):** Nach der Vorbereitung beginnt die Datenverarbeitung, die eine manuelle Bedienung verschiedener Softwaretools erfordert, in denen die Daten für das Hochladen vorbereitet werden (Abb. 7.2-1 Schritt 4).
4. Manueller **Upload (Load):** Manueller Upload von konvertierten Daten in externe Systeme, Datenformate und Dokumente (Abb. 7.2-1 Schritt 5).

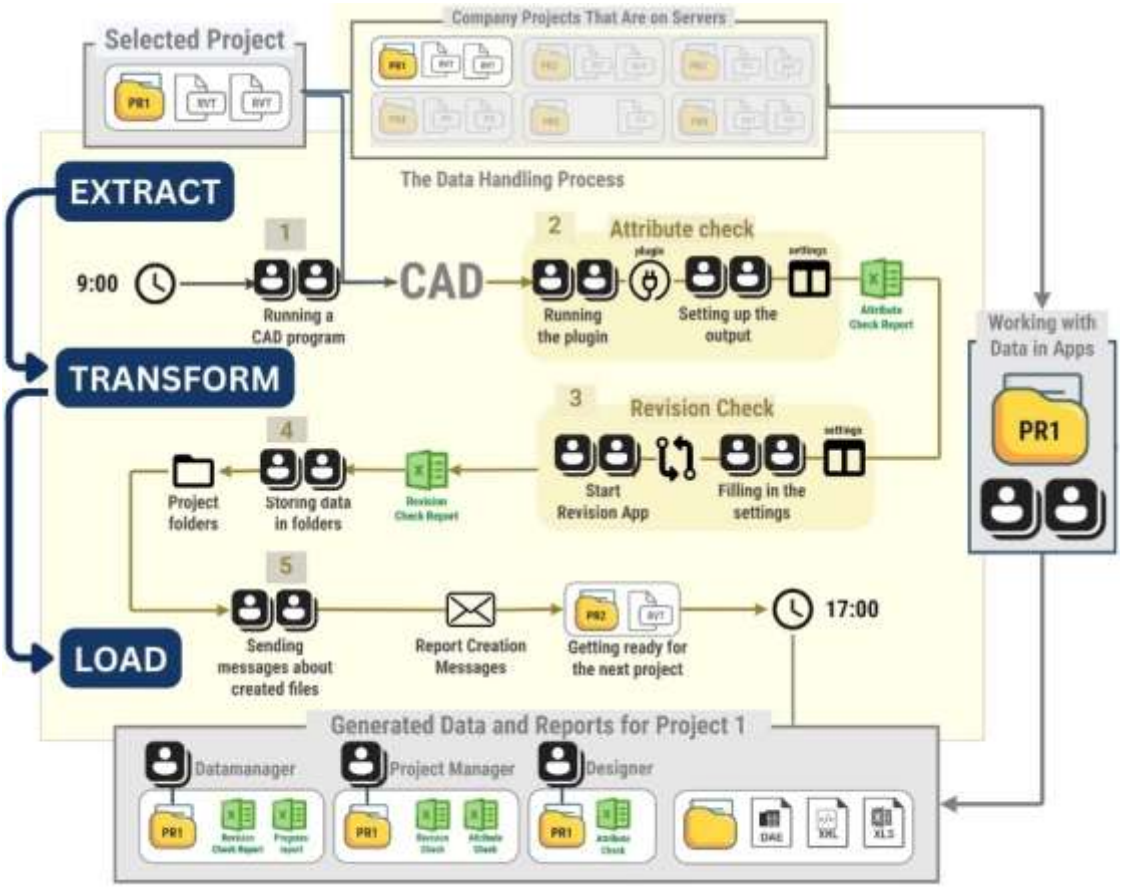


Abb. 7.2-1 Die herkömmliche manuelle ETL-Verarbeitung ist durch die Wünsche und physischen Fähigkeiten des einzelnen Technikers begrenzt.

Ein solcher Workflow ist ein Beispiel für einen klassischen ETL -Prozess - Extraktion, Transformation und Laden (ETL). Im Gegensatz zu anderen Branchen, in denen automatische ETL-Pipelines längst zum Standard gehören, dominiert im Baugewerbe immer noch die manuelle Arbeit, die die Prozesse verlangsamt und die Kosten erhöht

> **ETL (Extract, Transform, Load**) ist der Prozess des Extrahierens von Daten aus verschiedenen Quellen, des Transformierens in das gewünschte Format und des Ladens in das Zielsystem zur weiteren Analyse und Nutzung.

ETL ist ein Prozess, der drei Schlüsselkomponenten der Datenverarbeitung umfasst: Extrahieren, Transformieren, und Laden (Abb. 7.2-2):

- Extract - Extrahieren von Daten aus verschiedenen Quellen (Dateien, Datenbanken, API).
- **Transform** - Datenbereinigung, Aggregation, Normalisierung und logische Verarbeitung.
- **Load** - Laden Sie strukturierte Informationen in ein Data Warehouse, einen Bericht oder ein BI-System.

Zuvor wurde das Konzept der ETL nur gelegentlich gestreift: bei der Umwandlung eines unstrukturierten gescannten Dokuments in ein strukturiertes Tabellenformat (Abb. 4.1-1), im Zusammenhang mit der Formalisierung von Anforderungen zur Systematisierung der Wahrnehmung von Lebens- und Geschäftsprozessen (Abb. 4.4-20) und bei der Automatisierung der Datenvalidierung und Datenverarbeitung aus CAD-Lösungen. Betrachten wir nun ETL im Zusammenhang mit typischen Arbeitsabläufen genauer.

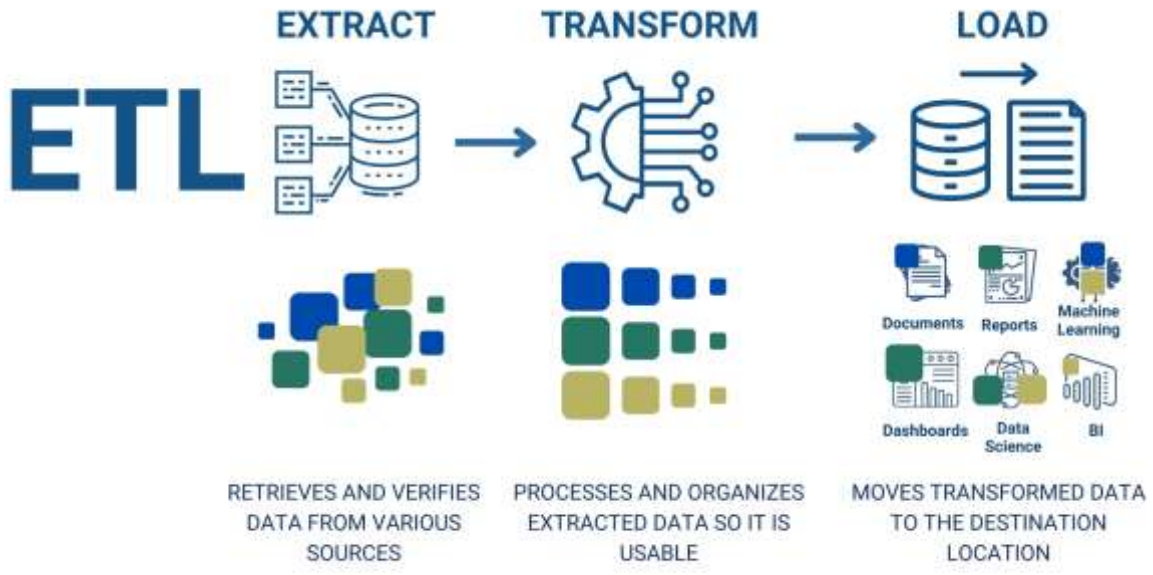


Abbildung 7.2-2 ETL automatisiert sich wiederholende Datenverarbeitungsaufgaben.

Ein manueller oder halbautomatischer ETL -Prozess setzt einen Manager oder Techniker voraus, der alle Schritte manuell durchführt - von der Datenerfassung bis zur Berichterstellung. Ein solcher Prozess nimmt viel Zeit in Anspruch, insbesondere wenn die Arbeitszeiten begrenzt sind (z. B. 9:00 bis 17:00 Uhr).

Unternehmen versuchen oft, das Problem der geringen Effizienz und der langsamen Geschwindigkeit zu lösen, indem sie modulare integrierte Lösungen kaufen (ERP, PMIS, CPM, CAFM, usw.), die dann von externen Anbietern und Beratern weiterentwickelt werden. Diese Anbieter und Drittentwickler werden jedoch oft

zu einem kritischen Abhängigkeitspunkt: Ihre technischen Beschränkungen wirken sich direkt auf die Leistung des gesamten Systems und des Unternehmens aus, wie in den vorangegangenen Kapiteln über proprietäre Systeme und Formate beschrieben. Die Probleme, die durch Fragmentierung und Abhängigkeit entstehen, wurden im Kapitel "Wie Bauunternehmen im Datenchaos ertrinken" ausführlich erörtert.

> Wenn ein Unternehmen nicht bereit ist, eine große modulare Plattform von einem der Anbieter zu implementieren, sucht es nach alternativen Möglichkeiten der Automatisierung. Eine davon ist die Entwicklung eigener modularer offener ETL -Förderer, bei denen jede Stufe (Extraktion, Transformation, Validierung, Laden) als Skripte implementiert wird, die nach einem Zeitplan ausgeführt werden.

In der automatisierten Version desselben ETL-Workflows (Abb. 7.2-1) sieht der Arbeitsprozess wie ein modularer Code aus, der mit der Verarbeitung von Daten und deren Umwandlung in eine offene, strukturierte Form beginnt. Sobald die strukturierten Daten eingegangen sind, werden verschiedene Skripte oder Module automatisch und nach einem bestimmten Zeitplan ausgeführt, um Änderungen zu überprüfen, umzuwandeln und Nachrichten zu senden (Abb. 7.2-3).

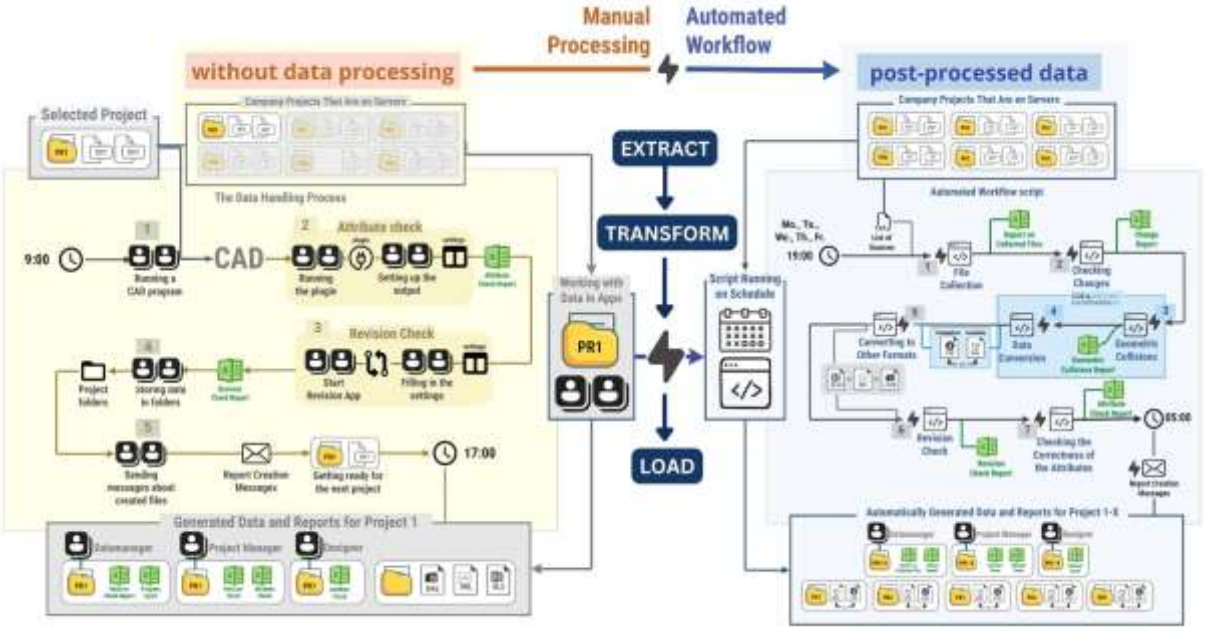


Abb. 7.2-3 Links die manuelle Bearbeitung, rechts ein automatischer Prozess, der im Gegensatz zur traditionellen manuellen Bearbeitung nicht durch die Fähigkeiten des Anwenders begrenzt ist.

In einem automatisierten Arbeitsablauf wird die Datenverarbeitung durch die Vorverarbeitung der ET(L)-Daten vereinfacht: Strukturierung und Vereinheitlichung.

> Bei den traditionellen Verarbeitungsverfahren arbeiten die Fachleute mit den Daten "wie sie sind" - so wie sie aus den Systemen oder der Software abgerufen werden. Bei automatisierten Prozessen durchlaufen die Daten dagegen oft zunächst eine ETL -Pipeline, wo sie in eine einheitliche Struktur und ein Format gebracht werden, das für die weitere Verwendung und Analyse geeignet ist.

Nehmen wir ein praktisches ETL-Beispiel, das den im Kapitel "Validierung von Daten und Validierungsergebnisse" beschriebenen Prozess der Validierung von Datentabellen demonstriert (Abb. 4.4-13). Hierfür verwenden wir die Pandas-Bibliothek in Verbindung mit dem LLM für automatisierte Datenanalyse- und Verarbeitungsprozesse.

ETL Extract: Datenerfassung

Die erste Phase des ETL-Prozesses - Extract - beginnt mit dem Schreiben von Code zum Sammeln von Datensätzen, die weiter geprüft und verarbeitet werden sollen. Zu diesem Zweck werden alle Ordner des Produktionsservers gescannt, Dokumente eines bestimmten Formats und Inhalts gesammelt und anschließend in eine strukturierte Form umgewandelt. Dieser Prozess wird in den Kapiteln "Umwandlung von unstrukturierten und textuellen Daten in eine strukturierte Form" und "Umwandlung von CAD-Daten (BIM) in eine strukturierte Form" ausführlich behandelt (Abb. 4.1-1 - Abb. 4.1-12).

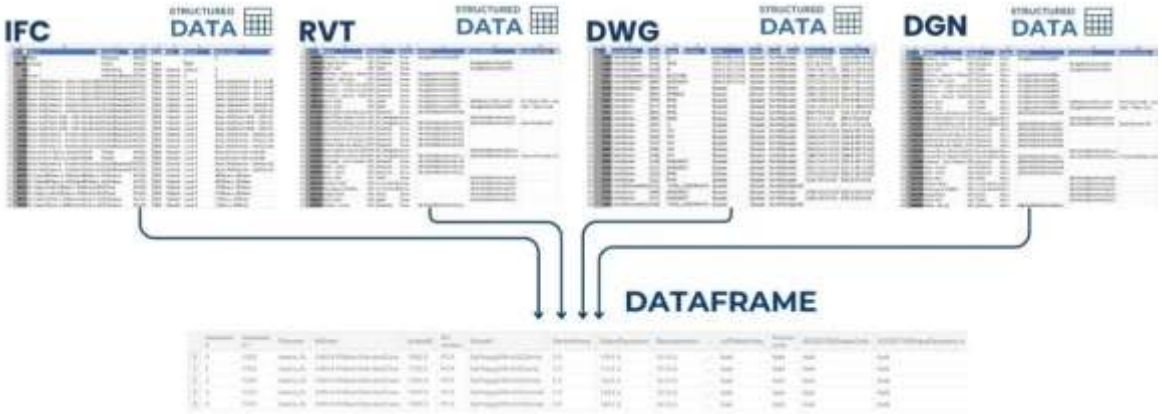


Abb. 7.2-4 Konvertieren Sie CAD-Daten (BIM) in einen großen Datenrahmen, der alle Projektabschnitte enthält.

Als anschauliches Beispiel verwenden wir den Datenladeschritt Extract und erhalten eine Tabelle aller CAD-(BIM-)-Projekte (Abb. 7.2-4) unter Verwendung von Reverse-Engineering-fähigen Konvertern [138für RVT- und IFC-Formate, um strukturierte Tabellen aus allen Projekten zu erhalten und sie in einer großen DataFrame-Tabelle zu kombinieren.

Create in LLM Chat

Extract data from CAD-BIM.py

```python
import os
import subprocess
import time
import pandas as pd

path_conv = r'C:\DDC_2023\\'  # Where RvtExporter.exe|IfcExporter.exe is located
path = r'C:\IFCprojects\\'    # Where Revit|IFC files are stored

def convert_and_wait(path_conv, exporter_name, file_path, extension):
    # Start the conversion process
    subprocess.Popen([os.path.join(path_conv, exporter_name),
    file_path], cwd=path_conv)
    output_file = os.path.join(path,
    f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")

# Conversion process for RVT and IFC files
for file in os.listdir(path):
    full_path = os.path.join(path, file)
    if file.endswith('.ifc'):
        convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
    elif file.endswith('.rvt'):
        convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')

# Combine converted Excel files into one dataframe
df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
if f.endswith('.xlsx')], ignore_index=True)
```

Run in IDE

Abb. 7.2-5 Konvertierung mit Python Code und SDK Reverse Engineering Tool für RVT und IFC Dateien in einen großen strukturierten (df) DataFrame.

Pandas DataFrame kann Daten aus einer Vielzahl von Quellen laden, darunter CSV-Textdateien, Excel-Tabellen, JSON - und XML - Dateien, Big-Data-Speicherformate wie Parquet und HDF5, und aus MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle und anderen Datenbanken. Darüber hinaus unterstützt Pandas das Laden von Daten aus APIs, Webseiten, Cloud-Diensten und Speichersystemen wie Google BigQuery, Amazon Redshift und Snowflake.

- Um einen Code zu schreiben, der eine Verbindung zu Datenbanken herstellt und Informationen aus diesen sammelt, senden Sie eine ähnliche Textanfrage an den LLM Chatroom (CHATGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN oder einen anderen):

Bitte schreiben Sie ein Beispiel für die Verbindung zu MySQL und die Konvertierung von Daten in ⏎

Die Antwort von LLM:

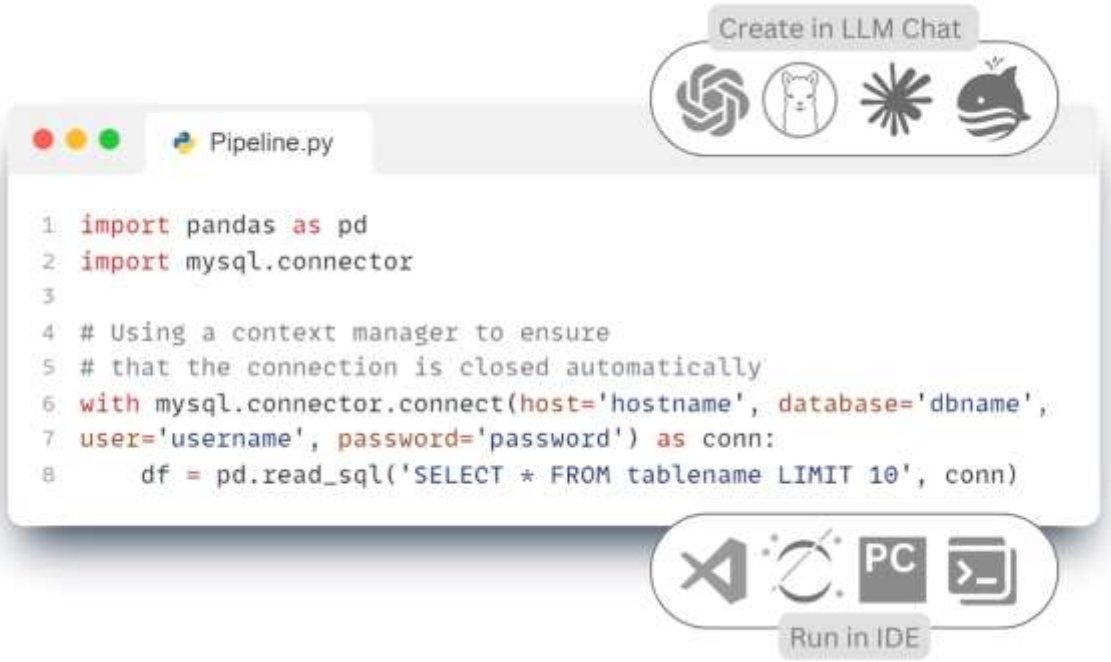


Abbildung 7.2-6 Beispiel für die Verbindung über Python zu einer MySQL-Datenbank und den Import von Daten aus der MySQL-Datenbank in einen DataFrame.

Der resultierende Code (Abb. 7.2-5, Abb. 7.2-6) kann in einer der oben genannten IDEs (integrierten Entwicklungsumgebungen) im Offline-Modus ausgeführt werden: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse mit PyDev-Plugin, Thonny, Wing IDE, IntelliJ IDEA mit Python-Plugin, JupyterLab oder beliebte Online-Tools: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Durch das Laden der Multiformat-Daten in die Variable "df" (Abb. 7.2-5 - Zeile 25; Abbildung 7.2-6 - Zeile 8) haben wir die Daten in das Pandas DataFrame Format konvertiert, eine der beliebtesten Strukturen für die Datenverarbeitung, die eine zweidimensionale Tabelle mit Zeilen und Spalten ist. Auf andere Speicherformate, die in ETL -Pipelines verwendet werden, wie Parquet, Apache ORC, JSON, Feather, HDF5, sowie moderne Data Warehouses wird im Kapitel "Datenspeicherung und -verwaltung in der Bauwirtschaft" näher eingegangen (Abb. 8.1-2).

Nach der Phase der Datenextraktion und -strukturierung (Extract) wird ein einzelnes Informationsfeld gebildet (Abb. 7.2-5, Abb. 7.2-6), das für die weitere Verarbeitung bereit ist. Bevor diese Daten jedoch in die Zielsysteme geladen oder für Analysen verwendet werden, müssen ihre Qualität, Integrität und Übereinstimmung mit den festgelegten Anforderungen sichergestellt werden. In dieser Phase findet die Datentransformation (Transform) statt - ein wichtiger Schritt, um die Zuverlässigkeit der nachfolgenden Schlussfolgerungen und Entscheidungen zu gewährleisten.

ETL Transform: Anwendung von Validierungs- und Transformationsregeln

Im Schritt Transformieren werden die Daten verarbeitet und transformiert. Dieser Prozess kann die Überprüfung der Korrektheit, die Normalisierung, das Auffüllen fehlender Werte und die Validierung mit Hilfe automatisierter Tools umfassen

> Laut der PwC-Studie "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] bewegen sich moderne Wirtschaftsprüfungsunternehmen weg von der stichprobenartigen Datenüberprüfung und hin zur Analyse riesiger Informationsmengen mithilfe automatisierter Tools. Dieser Ansatz ermöglicht es nicht nur, Unstimmigkeiten in der Berichterstattung zu erkennen, sondern auch Empfehlungen zur Optimierung von Geschäftsprozessen zu geben.

Im Bauwesen können ähnliche Methoden z. B. für die automatische Validierung von Entwurfsdaten, die Qualitätskontrolle im Bauwesen und die Leistungsbewertung von Auftragnehmern eingesetzt werden. Ein Werkzeug zur Automatisierung und Beschleunigung der Datenverarbeitung ist die Verwendung von regulären Ausdrücken (RegEx) in der Datenumwandlungsphase (Transform) des ETL-Prozesses. Mit RegEx können Sie Datenstrings effizient validieren, Inkonsistenzen aufdecken und die Integrität von Informationen bei minimalem Ressourcenverbrauch sicherstellen. Weitere Einzelheiten zu RegEx (Abb. 4.4-7) wurden im Kapitel "Übersetzen von Anforderungen in strukturierte Form" erörtert.

Betrachten wir ein praktisches Beispiel: Im Property Object Management System (RPM) legt der Verwalter die Anforderungen für die Schlüsselattribute der Objekte fest (Abb. 7.2-7). In der Transformationsphase müssen die folgenden Parameter validiert werden:

- Überprüfung der Formate der Objektkennungen (Attribut "ID")
- Kontrolle der Werte der Ersatzgewährleistungsfrist (Attribut "Gewährleistungsfrist")
- Überprüfung des Zyklus für den Austausch von Elementen (Attribut "Wartungsanforderungen")

Abbildung 7.2-7 Überprüfung der Qualität beginnt mit der Festlegung von Anforderungen für Attribute und deren Grenzwerte.

Um Grenzwerte für die Parametervalidierung festzulegen, nehmen wir beispielsweise an, dass wir aus wissen, dass gültige Werte für das Attribut "ID" nur die Zeichenketten "W-NEW", "W-OLD1" oder "D-122" oder

ähnliche Werte umfassen können, bei denen das erste Zeichen ein Buchstabe ist, gefolgt von einem Bindestrich und dann drei alphabetischen Zeichen "NEW", "OLD" oder einer beliebigen dreistelligen Zahl (Abb. 7.2-7). Der folgende reguläre Ausdruck (RegEx) kann verwendet werden, um diese Bezeichner zu validieren:

```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

Diese Vorlage stellt sicher, dass alle Bezeichner in den Daten den angegebenen Kriterien entsprechen. Wenn ein Wert die Prüfung nicht besteht, protokolliert das System einen Fehler. Um Python -Code zu erstellen, um die Daten umzuwandeln und die resultierenden Daten zur Erstellung einer Ergebnistabelle zu verwenden, formulieren Sie einfach eine Abfrage im LLM-Chat.

Textanfrage an LLM:

> Schreiben Sie einen Code zur Validierung der Spalten des DataFrame mit Hilfe regulärer Ausdrücke, der Identifikatoren im Format 'W-NEW' oder 'W-OLD' über RegEx, Energieeffizienz mit den Buchstaben 'A' bis 'G', Garantiezeit und Austauschzyklus mit numerischen Werten in Jahren überprüft. ⏎

Die Antwort von LLM:

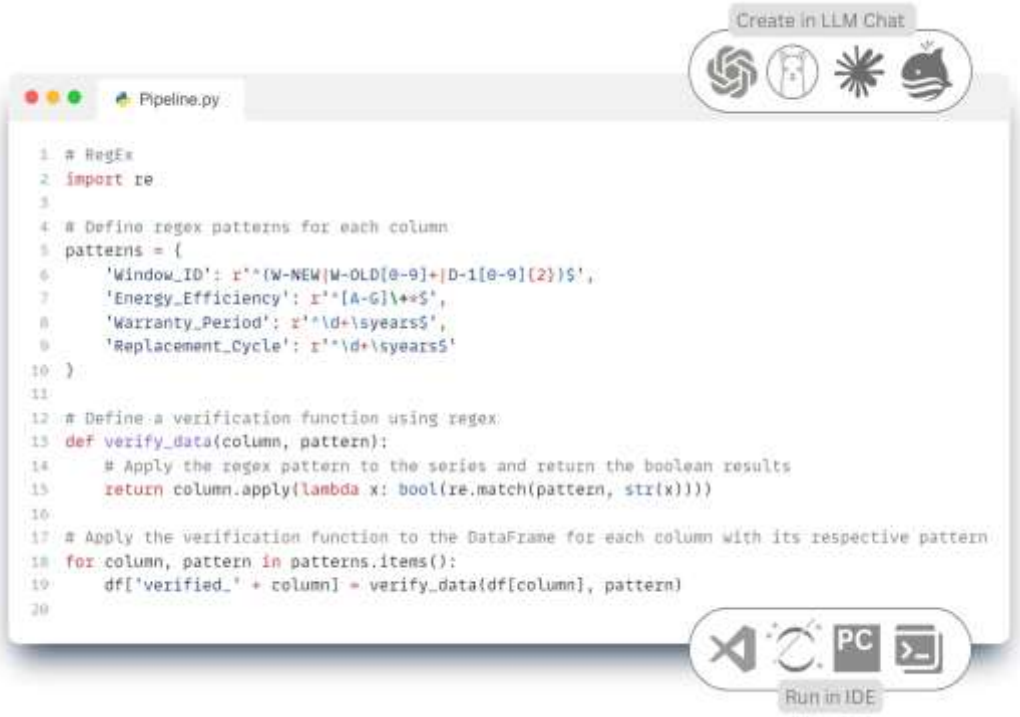


Abbildung 7.2-8 Der Code automatisiert den Validierungsprozess durch Anwendung von RegEx-Vorlagen auf die Spalten der Datenrahmenparameter.

Der obige, automatisch generierte Python -Code (Abb. 7.2-8) verwendet die Bibliothek "re" (RegEx regular

expressions), um eine Funktion zu definieren, die jedes Attribut eines Datenelements in einem DataFrame überprüft. Für jede angegebene Spalte (Attribut) wendet die Funktion ein RegEx-Muster an, um zu prüfen, ob jeder Datensatz mit dem erwarteten Format übereinstimmt, und fügt die Ergebnisse als neue Werte (False/True) in die neue Attributspalte des DataFrame ein.

> Eine solche automatische Validierung gewährleistet die formale Datenkonformität und kann als Teil eines Qualitätskontrollsystems während der Transformationsphase verwendet werden.

Nach erfolgreichem Abschluss des Schritts Transform und der Qualitätsprüfung können die Daten in die Zielsysteme hochgeladen werden. Transformierte und validierte Daten können in CSV, JSON, Excel, Datenbanken und andere Formate zur weiteren Verwendung hochgeladen werden. Je nach Aufgabe können die Ergebnisse auch in Berichten, Diagrammen oder analytischen Dashboards dargestellt werden.

ETL Load: Visualisieren Sie Ergebnisse in Tabellen und Diagrammen

Nach Abschluss der Phase Transform, in der die Daten strukturiert und überprüft wurden, ist die letzte Phase Load, in der die Daten sowohl in das Zielsystem geladen als auch zur Analyse visualisiert werden können. Die visuelle Darstellung der Daten ermöglicht es, Abweichungen schnell zu erkennen, Verteilungen zu analysieren und die wichtigsten Schlussfolgerungen allen Projektteilnehmern mitzuteilen, auch denen, die keinen technischen Hintergrund haben.

Anstatt Informationen in Form von Tabellen und Zahlen darzustellen, können wir Infografiken, Diagramme und Dashboards (Armaturenbretter) verwenden. Eines der gebräuchlichsten und flexibelsten Werkzeuge zur Visualisierung strukturierter Daten in Python ist die Matplotlib-Bibliothek (Abb. 7.2-9, Abbildung 7.2-10). Mit ihr lassen sich statische, animierte und interaktive Diagramme erstellen, und sie unterstützt eine Vielzahl von Diagrammtypen.

- Um die Ergebnisse der Attributprüfung aus dem RPM-System (Abb. 7.2-7) zu visualisieren, können Sie die folgende Abfrage an das Sprachmodell verwenden:

> Schreiben Sie einen Code zur Visualisierung der DataFrame-Daten, oben (Abb. 7.2-7), mit einem Histogramm für die Ergebnisse, um die Häufigkeit der Fehler im Attribut zu zeigen ⏎

- LLM-Antwort in Codeform und fertige Visualisierung der Ergebnisse der Codeausführung direkt im LLM-Chatroom:

Create in LLM Chat

Pipeline.py

```
# Re-importing necessary libraries for visualization
import matplotlib.pyplot as plt
import seaborn as sns

# Visualization 1: Bar Chart
plt.figure(figsize=(10, 6))
df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
plt.title('Data Verification Summary - Bar Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Run in IDE

Abb. 7.2-9 Visualisierung der Ergebnisse des Schritts Transform der Überprüfung von Attributwerten aus dem RPM-System (Abb. 7.2-7) als Histogramm im Schritt Load.

Es gibt viele quelloffene und kostenlose Visualisierungsbibliotheken, mit denen Sie strukturierte Daten in verschiedenen Formaten darstellen können. Fahren wir mit der Visualisierung der Ergebnisse mit einer anderen Art von Diagramm mit der folgenden Aufforderung im Chat fort:

> Dieselben Daten als Liniendiagramm aufzeichnen ⏎

Die Antwort von LLM:

Create in LLM Chat

Pipeline.py

```
# Visualization 3: Line Chart
df_visual.plot(kind='line', marker='o', figsize=(10, 6))
plt.title('Data Verification Trend - Line Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```

Run in IDE

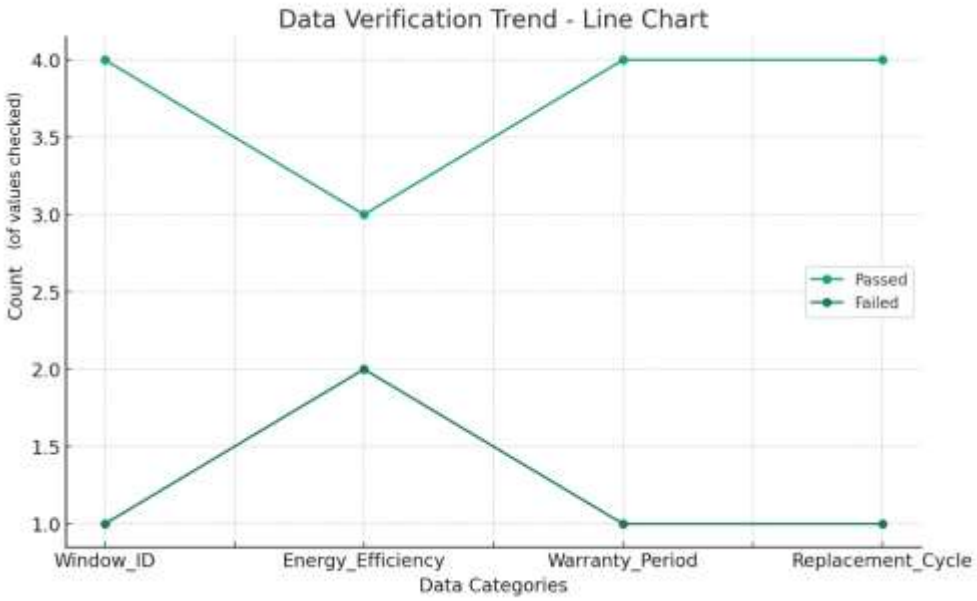


Abbildung 7.2-10 Visualisierung der Validierungsdaten (Abb. 7.2-8) als Liniendiagramm unter Verwendung der Matplotlib-Bibliothek.

Es gibt viele quelloffene und kostenlose Visualisierungsbibliotheken wie z.B.:

- Seaborn - für statistische Diagramme (Abb. 7.2-11)
- Plotly - für interaktive Webvisualisierungen (Abb. 7.2-12, Abb. 7.1-6)
- Altair - für deklarative Visualisierung
- Dash oder Streamlit - zur Erstellung vollständiger Dashboards

Die Kenntnis spezifischer Bibliotheken für die Visualisierung ist nicht erforderlich - moderne Werkzeuge, einschließlich LLM, ermöglichen die automatische Generierung von Code für Diagramme und ganze Anwendungen auf der Grundlage der Aufgabenbeschreibung.

Die Wahl des Tools hängt von den Zielen des Projekts ab: ob es sich um einen Bericht, eine Präsentation oder ein Online-Dashboard handelt. Die Open-Source-Bibliothek Seaborn eignet sich zum Beispiel besonders gut für die Arbeit mit kategorischen Daten und hilft, Muster und Trends zu erkennen.

- Um die Seaborn-Bibliothek in Betrieb zu sehen, können Sie entweder das LLM direkt bitten, die gewünschte Bibliothek zu benutzen oder eine entsprechende Textanfrage an das LLM unter senden:

  > Heatmap für anzeigen ⏎

- LLMs Antwort in Form von Code und einem fertigen Diagramm, dessen Plotting-Code nun in die IDE kopiert werden kann, und das Diagramm selbst kann kopiert oder gespeichert werden, um es in ein Dokument einzufügen:

Abbildung 7.2-11 Visualisierung der Ergebnisse der Validierung (Abb. 7.2-8) der Daten unter Verwendung der Seaborn-Bibliothek.

Für diejenigen, die einen interaktiven Ansatz bevorzugen, gibt es Tools, mit denen Sie dynamische Diagramme und Panels mit der Möglichkeit zur Interaktion erstellen können. Die Plotly-Bibliothek (Abb. 7.1-6, Abbildung 7.2-12) bietet die Möglichkeit, hochgradig interaktive Diagramme und Panels zu erstellen, die in Webseiten eingebettet werden können und es dem Benutzer ermöglichen, in Echtzeit mit den Daten zu interagieren.

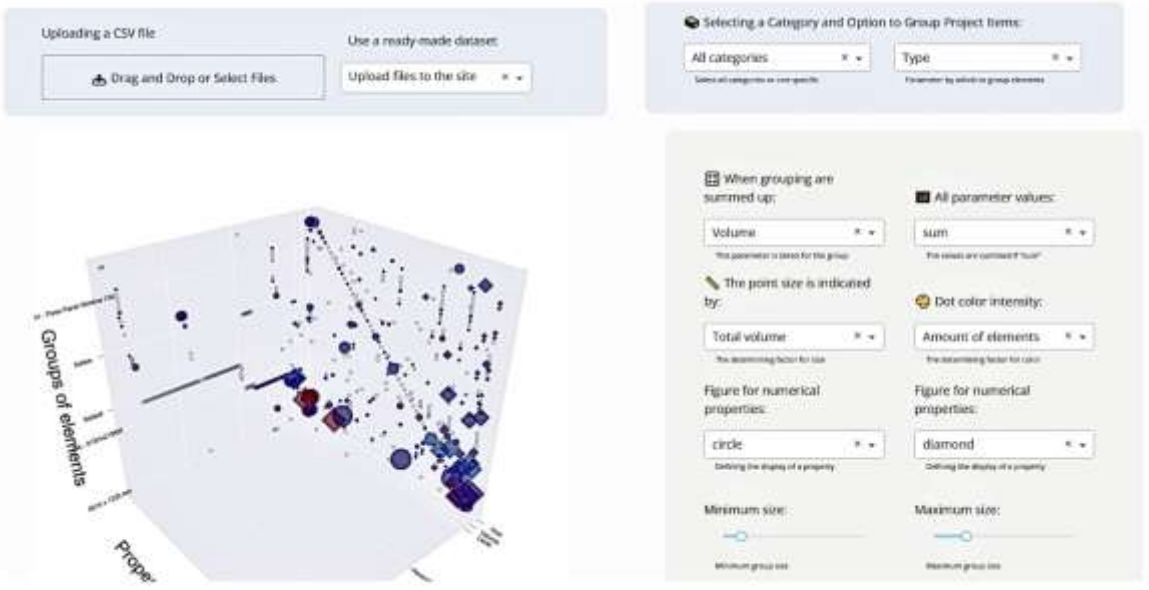


Abb. 7.2-12 Interaktive 3D -Visualisierung von Elementattributen aus einem CAD- (BIM-) Projekt unter Verwendung der Plotly-Bibliothek.

Die spezialisierten Open-Source-Bibliotheken Bokeh, Dash und Streamlit bieten eine bequeme Möglichkeit zur Darstellung von Daten, ohne dass tiefgreifende Kenntnisse der Webentwicklung erforderlich sind. Bokeh eignet sich für komplexe interaktive Diagramme, Dash wird für die Erstellung vollwertiger analytischer Dash-

boards verwendet, und Streamlit ermöglicht die schnelle Erstellung von Webanwendungen für die Datenanalyse.

Mit solchen Visualisierungstools können Entwickler und Analysten ihre Ergebnisse effektiv an Kollegen und Stakeholder weitergeben, was eine intuitive Interaktion mit den Daten ermöglicht und die Entscheidungsfindung vereinfacht.

ETL Laden: Automatische Erstellung von PDF-Dokumenten

In der Phase des Datenladens können Sie die Daten nicht nur visualisieren und in Tabellen oder Datenbanken hochladen, sondern auch automatisch Berichte mit den erforderlichen Grafiken, Diagrammen und analytischen Schlüsselindikatoren erstellen, die der Manager oder Spezialist erhält, der die Ergebnisse des Audits erwartet. Automatisierte Berichte können sowohl Kommentare als auch Textinterpretationen von Daten und Visualisierungselementen - Tabellen, Diagramme - enthalten.

- Um einen PDF -Bericht mit einem Histogramm (Abb. 7.2-9) und einer Beschreibung der Analyse auf der Grundlage der in den vorangegangenen Kapiteln durchgeführten Validierung zu erstellen, genügt es, eine Anfrage in einer Dialogfortsetzung mit dem LLM zu formulieren, z. B.:

  > Schreiben Sie einen Code, um eine PDF -Datei mit einem Histogramm und einer Beschreibung der Ergebnisse der obigen Datenvalidierung (im Chat) zu erstellen, und schreiben Sie eine Textwarnung, dass einige Kategorien nicht validiert wurden und dass fehlende ausgefüllt werden müssen ⏎

- LLMs Antwort als Code und fertiges PDF mit Ergebnissen:

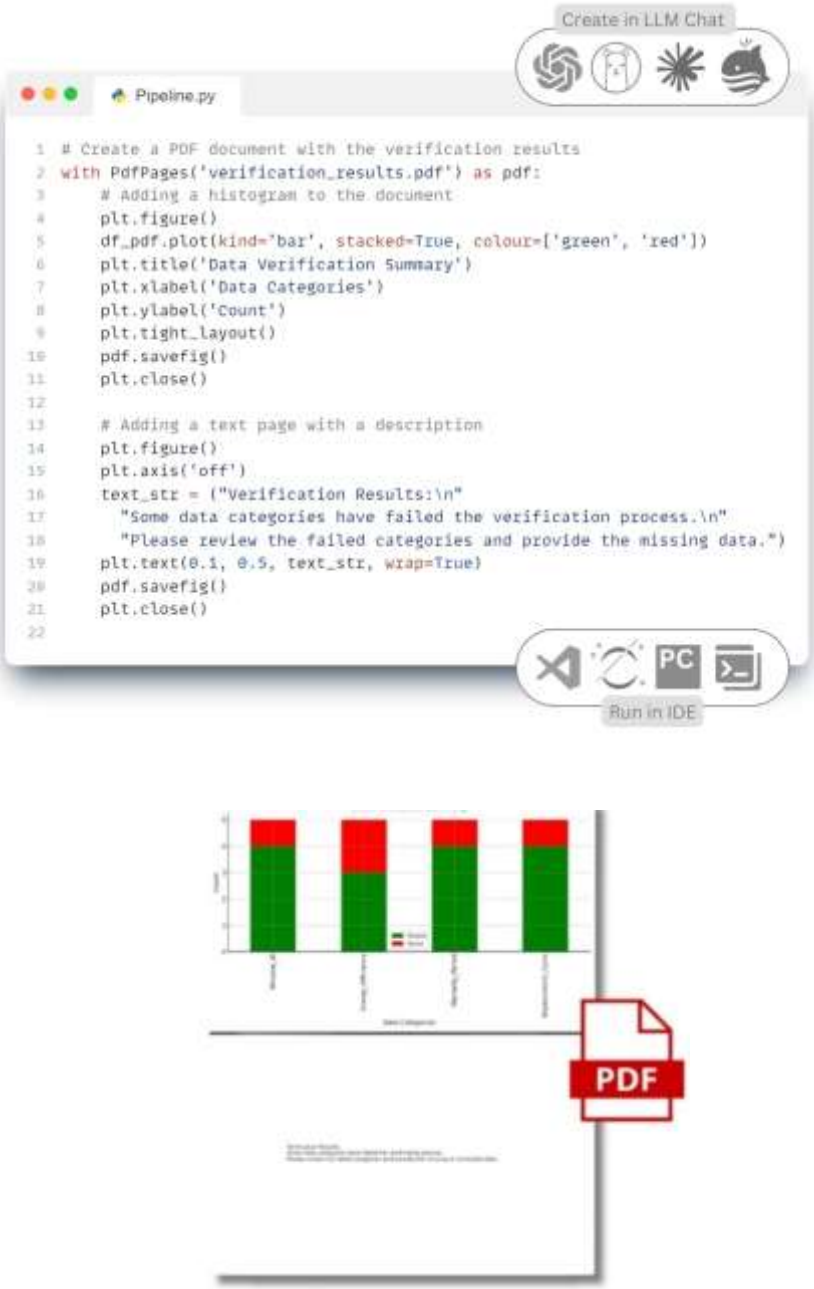


Abbildung 7.2-13 Der automatisierte Code erstellt ein PDF -Dokument, das ein Histogramm mit Testdaten und einen Text mit Testergebnissen enthält.

Eine automatisch geschriebene Lösung mit nur 20 Zeilen Code unter Verwendung von LLM erstellt sofort das gewünschte PDF (oder DOC) Dokument mit einer Visualisierung in Form eines Attributhistogramms (Abb. 7.2-13), das die Anzahl der Daten zeigt, die die Validierung bestanden und nicht bestanden haben, und mit einem zusätzlichen Textblock, der die Ergebnisse und Empfehlungen weitere Maßnahmen zusammenfasst.

Die automatisierte Dokumentenerstellung ist ein Schlüsselelement der Lastphase, insbesondere in einer Projektumgebung, in der die Geschwindigkeit der Berichterstattung und die Genauigkeit entscheidend sind.

ETL Laden: automatische Erzeugung von Dokumenten aus FPDF

Automatisierung des Berichtswesens in der ETL-Phase Load ist ein wichtiger Schritt in der Datenverarbeitung, insbesondere wenn die Ergebnisse von Analysen in einem leicht verständlichen Format dargestellt werden müssen. In der Baubranche ist dies häufig für Fortschrittsberichte, Projektdatenstatistiken, Qualitätssicherungsberichte oder Finanzdokumentationen relevant.

Eines der bequemsten Werkzeuge für solche Aufgaben ist die Open-Source-Bibliothek, FPDF, die sowohl für Python als auch für PHP verfügbar ist.

> Die **Open-Source-FPDF-Bibliothek** bietet eine flexible Möglichkeit, Dokumente durch Code zu generieren, der das Hinzufügen von Kopfzeilen, Text, Tabellen und Bildern ermöglicht. Die Verwendung von Code anstelle der manuellen Bearbeitung reduziert Fehler und beschleunigt die Erstellung von Berichten im PDF-Format.

Einer der wichtigsten Schritte bei der Erstellung eines PDF -Dokuments ist das Hinzufügen von Überschriften und des Haupttextes in Form von Kommentaren oder Beschreibungen. Bei der Erstellung eines Berichts ist es jedoch nicht nur wichtig, Text hinzuzufügen, sondern auch, ihn richtig zu strukturieren. Überschriften, Einzüge, Zeilenabstände - all dies wirkt sich auf die Lesbarkeit des Dokuments aus. Mit FPDF können Sie Formatierungsparameter festlegen, die Anordnung der Elemente steuern und den Stil des Dokuments individuell gestalten.

FPDF ist im Prinzip sehr ähnlich wie HTML. Diejenigen, die bereits mit HTML vertraut sind, können mit FPDF problemlos PDF Dokumente beliebiger Komplexität erzeugen, da die Codestruktur dem HTML-Markup sehr ähnlich ist: Überschriften, Text, Bilder und Tabellen werden auf ähnliche Weise hinzugefügt. Diejenigen, die mit HTML nicht vertraut sind, brauchen sich keine Sorgen zu machen - sie können LLM verwenden, das ihnen sofort hilft, den Code zu erstellen, um das gewünschte Dokumentenlayout zu erzeugen.

- Das folgende Beispiel zeigt, wie man einen Bericht mit einer Kopfzeile und einem Text erzeugt. Die Ausführung dieses Codes in einer IDE mit Python-Unterstützung erzeugt eine PDF -Datei, die die gewünschte Kopfzeile und den Text enthält:

```
from fpdf import FPDF     # Importiere die FPDF-Bibliothek
pdf = FPDF()    # PDF erstellen -document
pdf.add_page()     # Eine Seite hinzufügen

pdf.set_font("Arial", style='B', size=16)  # Schriftart setzen: Arial, fett, Größe 16
pdf.cell(200, 10, "Projektbericht", ln=True, align='C')  # Eine Überschrift erstellen und zentrieren
pdf.set_font("Arial", size=12)  # Ändern Sie die Schriftart in normale Arial, Größe 12
pdf.multi_cell(0, 10, "Dieses Dokument enthält Daten zu den Ergebnissen der Projektdateiprüfung...")  # Mehrzeiligen Text hinzufügen
pdf.output(r "C:\Berichte\Bericht.pdf") # PDF speichern -file
```

Abbildung 7.2-14 Mit ein paar Zeilen Python-Code können wir automatisch das benötigte PDF Textdokument erzeugen.

Bei der Erstellung von Berichten ist es wichtig zu berücksichtigen, dass die Daten, aus denen das Dokument gebildet wird, selten statisch sind. Überschriften, Textblöcke (Abb. 7.2-14) werden oft dynamisch gebildet und erhalten Werte in der Transformationsphase im ETL-Prozess.

> Mit Hilfe des Codes können Sie Dokumente erstellen, die aktuelle Informationen enthalten: Projektname, Datum der Berichterstellung sowie Informationen über die Teilnehmer oder den aktuellen Status. Durch die Verwendung von Variablen im Code können Sie diese Daten automatisch an den gewünschten Stellen in den Bericht einfügen, so dass eine manuelle Bearbeitung vor dem Versenden völlig überflüssig ist.

Neben einfachem Text und Überschriften nehmen Tabellen einen besonderen Platz in der Projektdokumentation ein. Fast jedes Dokument enthält strukturierte Daten: von Objektbeschreibungen bis hin zu Inspektionsergebnissen. Die automatische Generierung von Tabellen auf der Grundlage von Daten aus der Transform -Stufe ermöglicht es nicht nur, den Prozess der Dokumentenerstellung zu beschleunigen, sondern auch, Fehler bei der Übertragung von Informationen zu minimieren. FPDF ermöglicht das Einfügen von Tabellen in PDF -Dateien (als Text oder Bild), wobei Zellgrenzen, Spaltengrößen und Schriftarten festgelegt werden können (Abb. 7.2-15). Dies ist besonders praktisch bei der Arbeit mit dynamischen Daten, wenn die Anzahl der Zeilen und Spalten je nach den Aufgaben des Dokuments variieren kann.

Das folgende Beispiel zeigt, wie die Erstellung von Tabellen, z.B. mit Materiallisten, Schätzungen oder Parametertestergebnissen, automatisiert werden kann:

```
data = [
    ["Artikel", "Menge", "Preis"],   # Spaltenüberschriften
    ["Concrete", "10 m³", "$ 500."],   # Daten der ersten Zeile
    ["Armierungseisen", "2 Tonnen", "$ 600"],   # Daten der zweiten Zeile.
    ["Backstein", "5.000 Stück", "$ 750."], # Daten der dritten Zeile.
]

pdf = FPDF ()   # PDF erstellen -document
pdf.add_page() # Eine Seite hinzufügen
pdf.set_font("Arial", size=12)   # Schriftart festlegen
```

```
for row in data: #   Tabellenzeilen durchsuchen
    for item in row:   # Gehe durch die Zellen in der Zeile
        pdf.cell(60, 10, item, border=1)   # Erzeugt eine Zelle mit einem Rahmen, Breite 60 und
Höhe 10
    pdf.ln()   # Gehe zur nächsten Zeile
pdf.output(r "C:\reports\table.pdf")   # PDF speichern -file
```

Item	Quantity	Price
Concrete	10 m³	$500
Rebar	2 t.	$600
Brick	5000 pcs.	$750

Abbildung 7.2-15 Sie können nicht nur Text, sondern auch beliebige Tabelleninformationen aus dem Transformationsschritt in PDF automatisch erzeugen.

In realen Berichtsszenarien sind Tabellen in der Regel dynamisch generierte Informationen, die in der Phase der Datentransformation gewonnen werden. Im gegebenen Beispiel (Abb. 7.2-15) wird die Tabelle in statischer Form in das PDF -Dokument eingefügt: Die Daten für das Beispiel wurden im Datenwörterbuch (erste Zeile des Codes) platziert, unter realen Bedingungen wird eine solche Datenvariable automatisch nach z.B. der Gruppierung des Datenrahmens ausgefüllt.

In der Praxis werden solche Tabellen oft auf der Grundlage von strukturierten Daten erstellt, die aus verschiedenen dynamischen Quellen stammen: Datenbanken, Excel -Dateien, API -Schnittstellen oder Ergebnisse analytischer Berechnungen. Meistens werden die Daten in der Transform (ETL)-Phase aggregiert, gruppiert oder gefiltert - und erst dann in Summen in Form von Diagrammen oder zweidimensionalen Tabellen transformiert, die in Berichten angezeigt werden. Das bedeutet, dass sich der Tabelleninhalt je nach den gewählten Parametern, dem Analysezeitraum, den Projektfiltern oder den Benutzereinstellungen ändern kann.

> Die Verwendung von dynamischen Datenrahmen und Datensätzen in der Phase Transform macht den Berichtsprozess in der Phase Load so flexibel, skalierbar und leicht wiederholbar wie möglich, ohne dass manuelle Eingriffe erforderlich sind.

Neben Tabellen und Text unterstützt FPDF auch das Hinzufügen von Diagrammen zu tabellarischen Daten, was es Ihnen ermöglicht, mit Matplotlib oder anderen Visualisierungsbibliotheken, die wir oben besprochen haben, erzeugte Bilder in den Bericht einzubetten. Beliebige Grafiken, Diagramme und Schaubilder können mit Hilfe des Codes in das Dokument eingefügt werden.

- Mit der Python-Bibliothek FPDF fügen wir ein mit Matplotlib vorgeneriertes Diagramm in das PDF-Dokument ein:

```
import matplotlib.pyplot as plt  # Importieren Sie Matplotlib, um Plots zu erstellen.

fig, ax = plt.subplots()  # Erstellen Sie die Figur und die Achsen des Diagramms
categories = ["Concrete", "Rebar", "Brick"]  # Namen der Kategorien
Werte = [50000, 60000, 75000]  # Werte der Kategorien
ax.bar(categories, values)  # Ein Balkendiagramm erstellen
plt.ylabel("Wert,$.")  # Vorzeichen der Y-Achse
plt.title("Kostenverteilung")  # Einen Titel hinzufügen
plt.savefig(r "C:\berichte\chart\chart\chart.png")  # Das Diagramm als Bild speichern

pdf = FPDF ()  # PDF erstellen -document
pdf.add_page() # Eine Seite hinzufügen
pdf.set_font("Arial", size=12)  # Schriftart festlegen
pdf.cell(200, 10, "Kostentabelle", ln=True, align='C')  # Eine Kopfzeile hinzufügen

pdf.image(r "C:\berichte\chart\chart\chart.png", x=10, y=30, w=100)  # Bild in PDF
einfügen (x, y - Koordinaten, w - Breite)
pdf.output(r "C:\Berichte\Karte_Bericht.pdf")  # Speichern der PDF-Datei
```

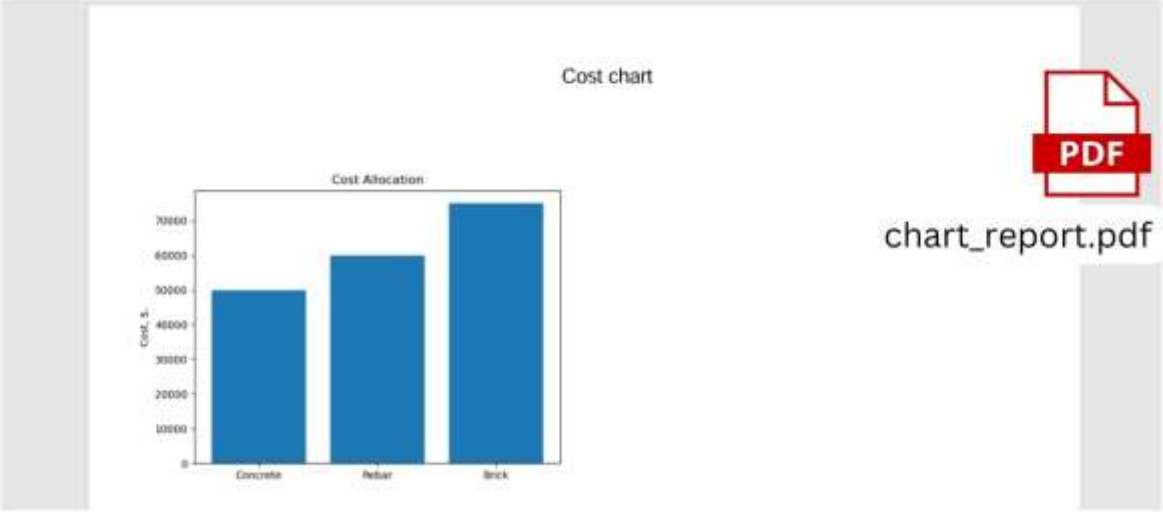


Abb. 7.2-16 Mit einem Dutzend Codezeilen können Sie ein Diagramm erstellen, es speichern und dann in ein PDF-Dokument einfügen.

Mit FPDF wird der Prozess der Dokumentenerstellung und -logik transparent, schnell und bequem. In den Code integrierte Vorlagen ermöglichen die Erstellung von Dokumenten mit aktuellen Daten, so dass die Notwendigkeit des manuellen Ausfüllens entfällt.

ETL-Automatisierung - statt zeitaufwändiger manueller Berichterstellung können sich Fachleute auf die Analyse von Daten und das Treffen von Entscheidungen konzentrieren, anstatt das richtige Tool für die Arbeit mit einem bestimmten Datensilo mit einer übersichtlichen Benutzeroberfläche auszuwählen.

Damit bietet die FPDF-Bibliothek ein flexibles Werkzeug für die automatisierte Erstellung von Dokumenten jeglicher Komplexität - von kurzen technischen Berichten bis hin zu komplexen analytischen Zusammenfassungen mit Tabellen und Diagrammen, wodurch nicht nur der Dokumentenfluss beschleunigt, sondern auch die Fehlerwahrscheinlichkeit bei der manuellen Dateneingabe und Formatierung erheblich reduziert werden kann.

ETL Laden: Berichte und Laden in andere Systeme

In der Lastphase wurden die Ergebnisse in Form von Tabellen, Grafiken und PDF-Abschlussberichten erstellt, die den festgelegten Anforderungen entsprachen. Darüber hinaus ist es möglich, diese Daten in maschinenlesbare Formate (z. B. CSV) zu exportieren, was für die Integration mit externen Systemen wie ERP, CAFM, CPM, BI-Plattformen und anderen Unternehmens- oder Branchenlösungen erforderlich ist. Zusätzlich zu CSV können Uploads in XLSX, JSON, XML oder direkt in Datenbanken, die den automatischen Informationsaustausch unterstützen, durchgeführt werden.

- Um den entsprechenden Code zur Automatisierung des Ladeschritts zu generieren, fragen Sie einfach die LLM -Schnittstelle ab, zum Beispiel: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude oder QWEN:

> Schreiben Sie einen Code, um einen Bericht über die Ergebnisse der Datenüberprüfung in DataFrame zu erstellen. Dabei werden die Spalten mit dem Präfix "verified_" gezählt und in "Passed" und "Failed" umbenannt, fehlende Werte werden durch 0 ersetzt, und nur die Zeilen, die alle Überprüfungen bestanden haben, werden in eine CSV-Datei exportiert.

- Die Antwort von LLM:

Abbildung 7.2-17 Validierte Daten, die im Transformationsschritt aus dem endgültigen Datenrahmen gewonnen wurden, werden zur Integration in andere Systeme in eine CSV -Datei exportiert.

Der obige Code (Abb. 7.2-17) implementiert die letzte Phase des ETL -Prozesses - das Laden, bei dem die verifizierten Daten im CSV-Format gespeichert werden, das mit den meisten externen Systemen und Datenbanken kompatibel ist. Damit haben wir den gesamten Zyklus des ETL -Prozesses abgeschlossen, einschließlich der Extraktion, Umwandlung, Visualisierung, Dokumentation und des Exports der Daten in die von uns benötigten Systeme und Formate, was die Reproduzierbarkeit, Transparenz und Automatisierung der Arbeit mit Informationen gewährleistet.

Die ETL-Pipeline kann sowohl für die Verarbeitung einzelner Projekte als auch für umfangreiche Anwendungen eingesetzt werden - bei der Analyse von Hunderten oder Tausenden von eingehenden Daten in Form von Dokumenten, Bildern, Scans, CAD-Projekten, Punktwolken, PDF-Dateien oder anderen Quellen, die aus verteilten Systemen stammen. Die Fähigkeit, den Prozess vollständig zu automatisieren, macht ETL nicht nur zu einem technischen Verarbeitungswerkzeug, sondern zur Grundlage einer digitalen Bauinformationsinfrastruktur.

ETL mit LLM: Visualisierung von Daten aus PDF -Dokumenten

Es ist an der Zeit, einen vollwertigen ETL-Prozess zu entwickeln, der alle wichtigen Phasen der Datenverarbeitung in einem einzigen Szenario abdeckt - Extraktion, Transformation und Laden. Lassen Sie uns eine automatisierte ETL-Pipeline aufbauen, die es Ihnen ermöglicht, PDF-Dokumente ohne manuelle Arbeit zu verarbeiten - Daten aus Dokumenten zu extrahieren, zu visualisieren, zu analysieren und an andere Systeme zu übertragen.

Der ETL-Prozess in unserem Beispiel wird durch Prompts beschrieben, die dem Sprachmodell (LLM) alle ETL-Prozesse mit einer Beschreibung des zu erzielenden Endergebnisses erklären müssen. In diesem Fall besteht die Aufgabe darin, alle PDF-Dateien im angegebenen Ordner und seinen Unterordnern zu finden, relevante Informationen aus ihnen zu extrahieren - z. B. Materialnamen, Mengen und Kosten - und das Ergebnis als strukturierte Tabelle (DataFrame) zur weiteren Analyse zu präsentieren

- Erste Textabfrage im LLM zur automatischen Extraktion von Daten aus mehreren PDF-Dokumenten und Erstellung eines Datenrahmens der Daten für den Extraktionsschritt:

> Schreiben Sie einen Code zum Extrahieren von Materialinformationen aus PDF-Dateien in einem bestimmten Ordner und seinen Unterordnern. Die Daten in der PDF-Datei umfassen Materialname, Menge und Kosten. Das Ergebnis soll in einem DataFrame gespeichert werden. ⏎

Die Antwort von LLM:

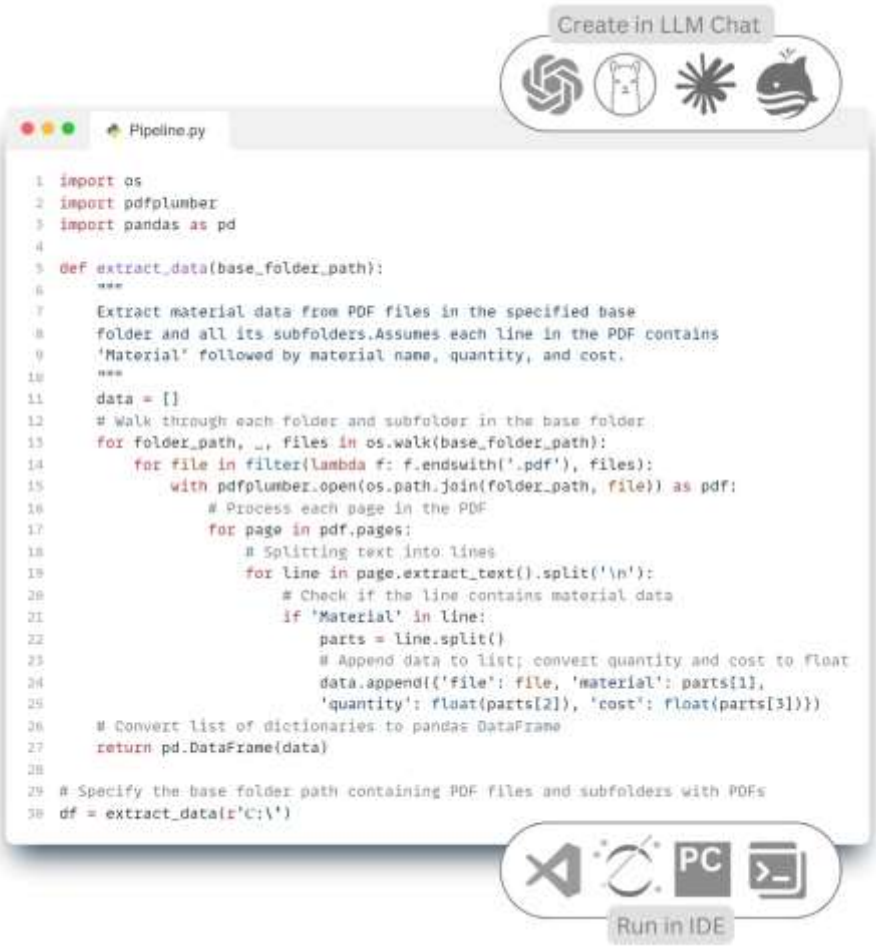


Abbildung 7.2-18 LLM erstellt Python -Code zum Extrahieren von Daten aus PDF -Dateien in einem bestimmten Ordner und allen seinen Unterordnern.

Die LLM-Antwort (Abb. 7.2-18) ist ein vorgefertigtes Python -Skript, das automatisch alle Ordner durchsucht, die gefundenen PDF -Dateien öffnet, Textinformationen aus ihnen extrahiert und in eine Tabelle umwandelt. Der resultierende Chat-Code kann in einer der gängigen IDEs PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse mit PyDev-Plugin, Thonny, Wing IDE, IntelliJ IDEA mit Python-Plugin, JupyterLab oder beliebten Online-Tools Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker ausgeführt werden.

Alternativ, um den Erfassungsprozess zu vereinfachen, können wir, anstatt den Code aus dem LLM zu kopieren und den Code in der IDE zu verwenden, auch ein Dutzend PDF -Dateien direkt in den LLM-Chatroom hochladen (Abb. 7.2-19) und eine Tabelle als Ausgabe erhalten, ohne den Code sehen oder ausführen zu müssen. Das Ergebnis der Ausführung dieses Codes wird eine Tabelle mit den von uns ausgewählten Attributen sein.

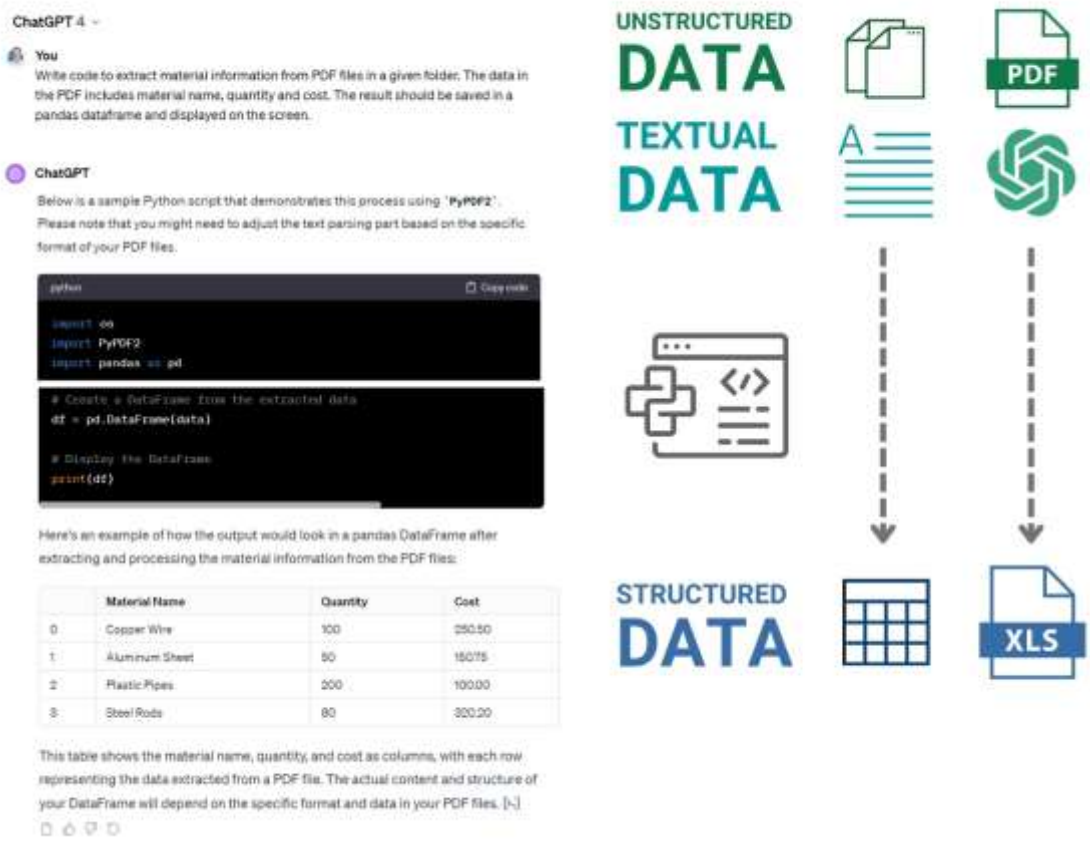


Abbildung 7.2-19 Das Ergebnis der Ausführung von Code in LLM, der Daten aus PDF -Dateien in einer strukturierten Datenrahmenansicht mit ausgewählten Attributen extrahiert.

Im nächsten Schritt bitten wir um ein Sprachmodell für die gewonnenen Daten - zum Beispiel, um die Kosten und das Volumen des Materialverbrauchs zu vergleichen und einige Beispielvisualisierungen zu erstellen, die als Grundlage für weitere Analysen dienen.

Bitten Sie in einem fortlaufenden Gespräch mit dem LLM darum, einige Diagramme aus den Tabellen zu erstellen, die im Schritt Transformieren erstellt wurden (Abb. 7.2-18):

> Visualisieren Sie die Gesamtkosten und die Menge der einzelnen Materialien im Data-Frame (Abb. 7.2-18⏎

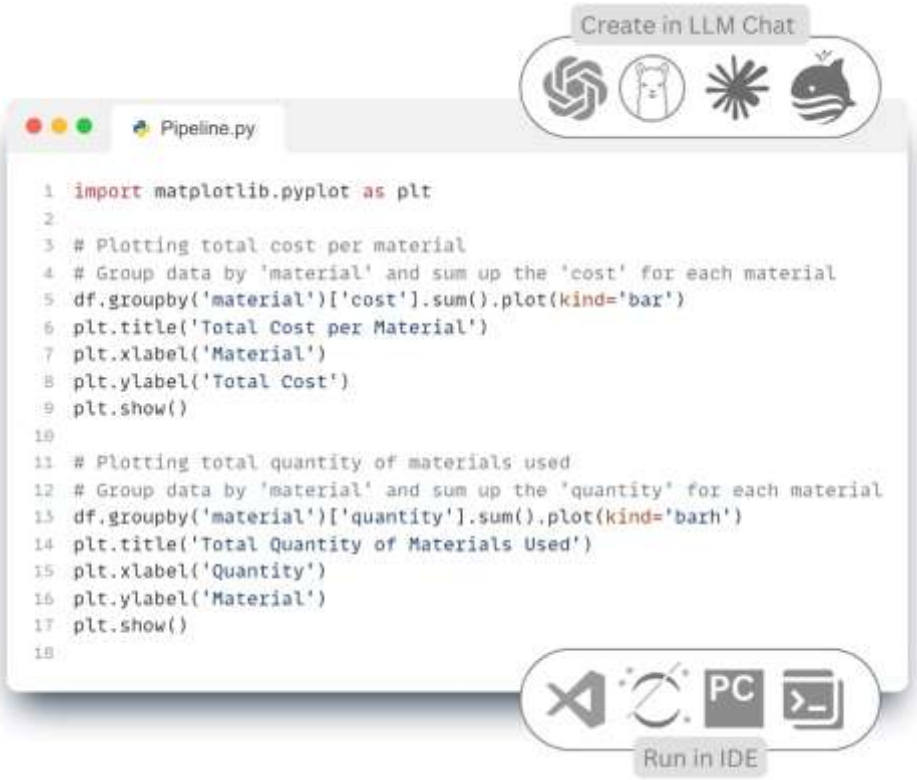


Abbildung 7.2-20 LLM-Antwort -Modelle als Python-Code zur Visualisierung von Daten aus einem Datenrahmen unter Verwendung der matplotlib-Bibliothek.

LLM generiert automatisch Python -Code (Abb. 7.2-20) unter Verwendung der matplotlib-Bibliothek und führt ihn aus. Nach der Ausführung dieses Codes erhalten wir Diagramme der Kosten und des Materialverbrauchs in Bauprojekten direkt im Chat (Abb. 7.2-21), was die analytische Arbeit stark vereinfacht.

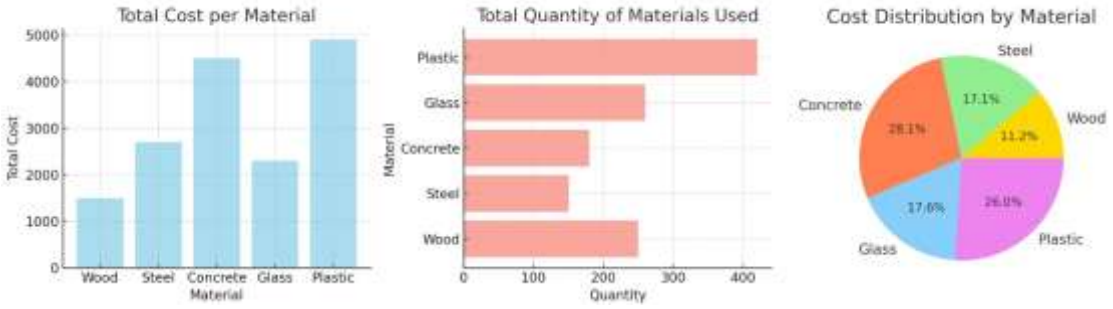


Abb. 7.2-21 Visualisierung der LLM-Antwort als Graphen basierend auf den im DataFrame gesammelten Daten.

Unterstützung bei der Entwicklung von Ideen für das Schreiben von ETL Code, die Analyse und Ausführung von Code und die Visualisierung von Ergebnissen ist durch einfache Textabfragen in LLM möglich, ohne dass die Grundlagen der Programmierung erlernt werden müssen. Das Aufkommen von KI Tools wie LLM verändert definitiv die Herangehensweise an die Programmierung und Automatisierung der Datenverarbeitung (Abb. 7.2-22).

> Laut dem PwC-Bericht "What is the real value of artificial intelligence for your business and how can you capitalise on it?" (2017) [139] werden Prozessautomatisierung und Produktivitätssteigerungen die Haupttreiber des Wirtschaftswachstums sein. Und es wird erwartet, dass Produktivitätsverbesserungen zwischen 2017 und 2030 mehr als 55% des gesamten KI-getriebenen BIP-Wachstums ausmachen werden."

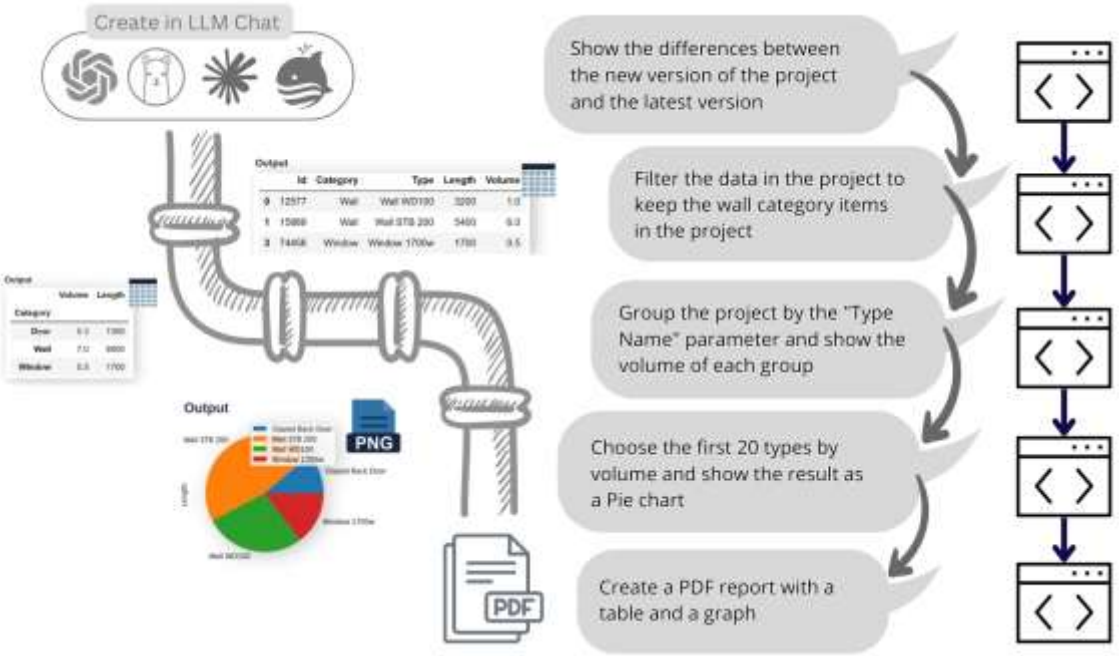


Abbildung 7.2-22 AI LLM hilft bei der Erstellung von Code-Entwürfen, die auf zukünftige Projekte angewendet werden, ohne dass ein LLM erforderlich ist.

> Mit Werkzeugen wie ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok sowie offenen Daten und Open-Source-Software können wir Prozesse automatisieren, die früher nur mit spezialisierten, teuren und schwer zu wartenden modularen proprietären Systemen durchgeführt wurden.

Für das Baugewerbe bedeutet dies, dass Unternehmen, die als erste automatisierte Pipeline -Datenprozesse einführen, von erheblichen Vorteilen profitieren werden, von einer verbesserten Effizienz des Projektmanagements über geringere finanzielle Verluste bis hin zur Abschaffung fragmentierter Anwendungen und isolierter Datenlager.

Die beschriebene Logik der Ausführung von Geschäftsaufgaben im ETL-Prozess ist ein entscheidender Teil der Automatisierung von Analyse- und Datenverarbeitungsprozessen, die eine spezifische Variante eines umfassenderen Konzepts sind - Pipelines (Pipelines).

KAPITEL 7.3.
AUTOMATISCHER ETL FÖRDERER (PIPELINE)

Pipeline: Automatische ETL

Der ETL-Prozess wird traditionell für die Verarbeitung von Daten in analytischen Systemen verwendet, die sowohl strukturierte als auch unstrukturierte Quellen umfassen. In der heutigen digitalen Umgebung wird jedoch zunehmend ein breiterer Begriff verwendet - Pipeline (Förderband), der jede sequentielle Verarbeitungskette beschreibt, bei der der Output einer Stufe zum Input für die nächste wird.

Dieser Ansatz gilt nicht nur für Daten, sondern auch für andere Arten der Automatisierung: Aufgabenbearbeitung, Gebäudeberichte, Integration in Software und digitale Arbeitsabläufe (Abb. 7.3-1).

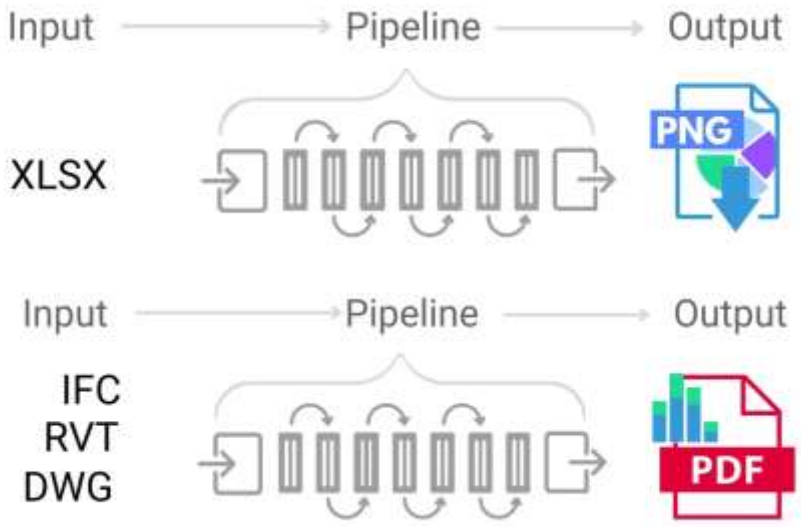


Abbildung 7.3-1 Pipeline ist eine Verarbeitungssequenz, bei der der Ausgang einer Stufe zum Eingang der nächsten Stufe wird.

Die Verwendung von Pipelines ist eines der Hauptelemente der Automatisierung, insbesondere bei der Arbeit mit großen Mengen heterogener Daten. Die Pipeline-Architektur ermöglicht es, komplexe Verarbeitungsschritte in einem modularen, konsistenten und überschaubaren Format zu organisieren, was die Lesbarkeit erhöht, die Wartung des Codes vereinfacht und inkrementelles Debugging sowie skalierbare Tests ermöglicht.

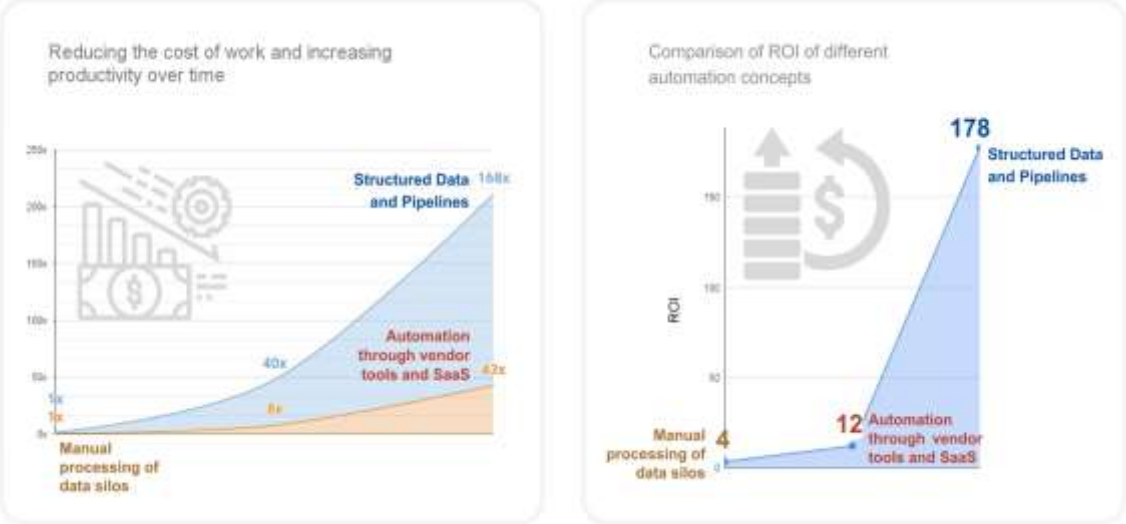


Abb. 7.3-2 ROI Pipeline Datenvalidierungsprozess reduziert die Ausführungszeit um das Zehn- bis Hundertfache im Vergleich zur Verarbeitung mit klassischen Tools [74].

Im Gegensatz zur manuellen Arbeit in proprietären Systemen (ERP, PMIS, CAD, etc.) können Sie mit Pipelining die Geschwindigkeit von Aufgaben deutlich erhöhen (Abb. 7.3-2), sich wiederholende Arbeiten vermeiden und den Start von Prozessen zum richtigen Zeitpunkt automatisieren (Abb. 7.3-3).

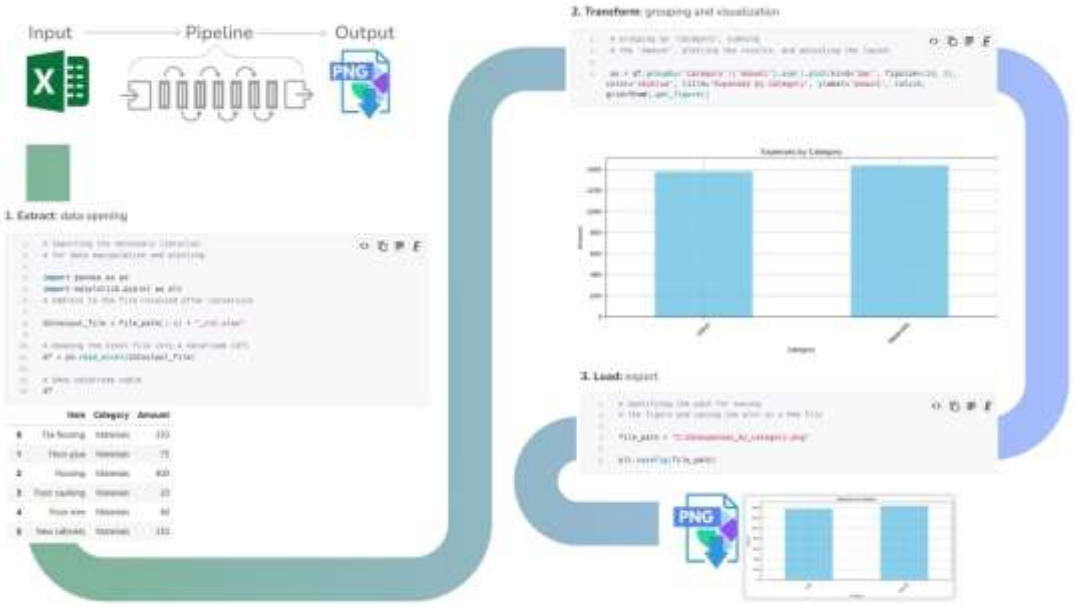


Abb. 7.3-3 ETL-Beispiel Pipeline zum automatischen Abrufen eines Diagramms aus tabellarischen Daten in einer XLSX-Datei ohne Excel zu öffnen.

Für die Verarbeitung von Streaming-Daten und den Aufbau einer automatisierten Pipeline, ähnlich dem ETL-Prozess, müssen Sie im Voraus die Datenquellen sowie den Zeitrahmen für die Erfassung der Daten festlegen - entweder für einen bestimmten Geschäftsprozess oder für das gesamte Unternehmen.

> Bei Bauprojekten stammen die Daten aus vielen heterogenen Quellen mit unterschiedlichen Aktualisierungsintervallen. Um eine zuverlässige Datenpräsentation zu schaffen, ist es wichtig, aufzuzeichnen, wann die Informationen abgerufen und aktualisiert werden. Dies ermöglicht eine rechtzeitige Entscheidungsfindung und verbessert die Effizienz des Projektmanagements.

Eine Möglichkeit besteht darin, den Zusammenstellungsprozess zu einem festen Zeitpunkt zu starten - zum Beispiel um 19:00 Uhr, am Ende des Arbeitstages. Zu diesem Zeitpunkt wird das erste Skript, das für die Aggregation von Daten aus verschiedenen Systemen und Speichern zuständig ist, aktiviert (Abb. 7.3-4 Schritt 1). Es folgt die automatische Verarbeitung und Umwandlung der Daten in ein strukturiertes, für die Analyse geeignetes Format (Abb. 7.3-4 Schritt 2-4). In der letzten Phase werden unter Verwendung der aufbereiteten Daten automatisch Berichte, Dashboards und andere in den vorangegangenen Kapiteln beschriebene Produkte erstellt (Abb. 7.3-4 Schritt 6-7). Als Ergebnis liegen den Managern bereits um 05:00 Uhr morgens aktuelle Projektstatusberichte in dem gewünschten Format vor (Abb. 7.3-5).

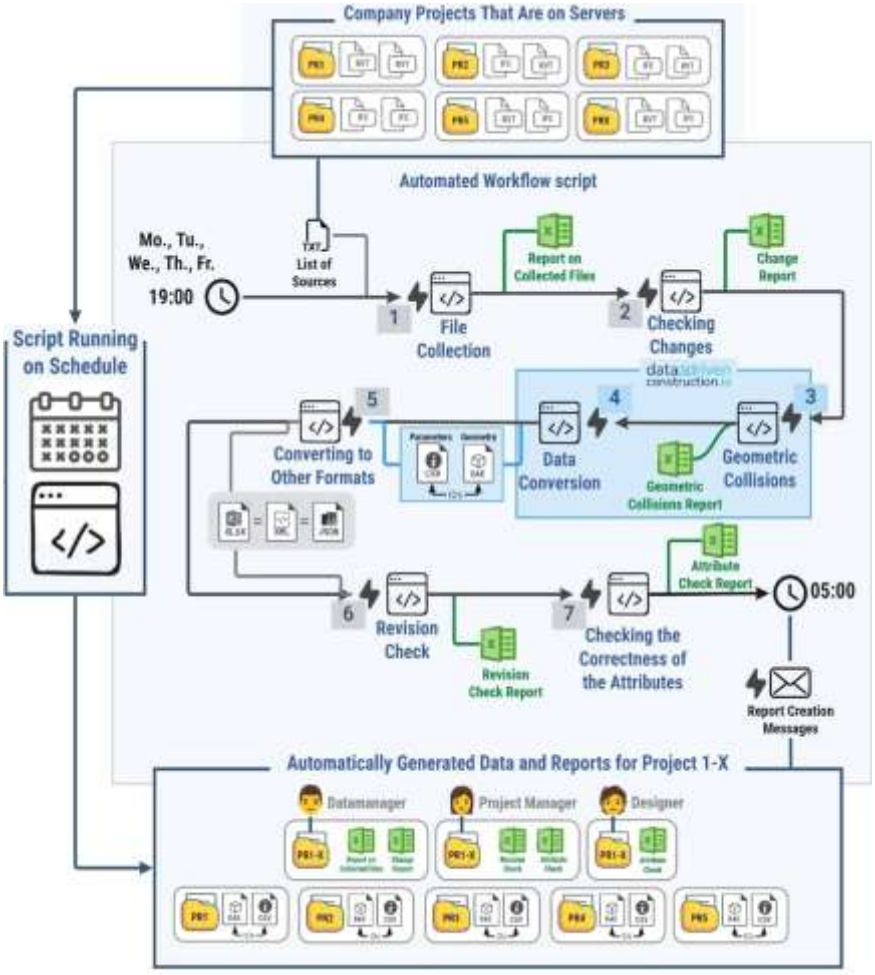


Abb. 7.3-4 Die abends automatisch erfassten Daten in der Pipeline werden über Nacht verarbeitet, so dass den Managern am Morgen aktuelle Berichte und neue Berichte vorliegen.

> Rechtzeitige Datenerfassung, Definition von Kennzahlen, Automatisierung von Transformationsprozessen und Visualisierung durch Dashboards sind Schlüsselelemente einer erfolgreichen datengesteuerten Entscheidungsfindung.

Solche automatisierten Prozesse (Abb. 7.3-4) können völlig autonom ausgeführt werden: Sie laufen nach einem Zeitplan, verarbeiten Daten unbeaufsichtigt und können entweder in der Cloud oder auf dem unternehmenseigenen Server bereitgestellt werden (Abb. 7.3-5). Dadurch können solche ETL-Pipelines in die bestehende IT-Infrastruktur integriert werden, wobei die Kontrolle über die Daten erhalten bleibt und eine flexible Skalierung möglich ist.

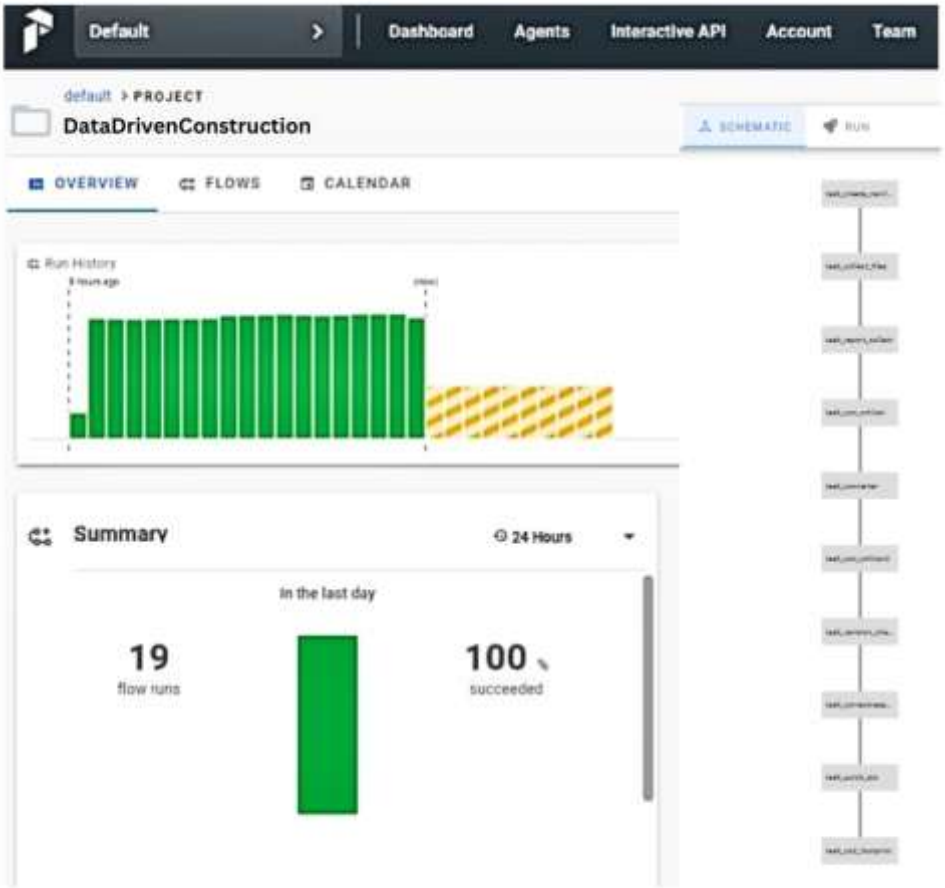


Abb. 7.3-5 Automatische ETL -Förderer -Prozesse (Abb. 7.3-4) auf der Prefect-Plattform, bei denen 10 Python-Skripte werktäglich abwechselnd nach 19:00 Uhr ausgeführt werden.

Die Automatisierung von Arbeitsabläufen steigert nicht nur die Produktivität des Teams, indem sie Zeit für sinnvollere und weniger routinemäßige Aufgaben freisetzt, sondern dient auch als wichtiger erster Schritt zur Integration von Technologien der künstlichen Intelligenz (KI) in Geschäftsprozesse, auf die wir im Kapitel über vorausschauende Analysen und maschinelles Lernen näher eingehen werden.

Pipeline -ETL Datenvalidierungsprozess mit LLM

In den vorangegangenen Kapiteln über die Erstellung von Datenanforderungen und die Automatisierung von ETL haben wir den Prozess der Datenaufbereitung, -umwandlung, -validierung und -visualisierung schrittweise aufgegliedert. Diese Aktivitäten wurden als separate Codeblöcke implementiert (Abb. 7.2-18 - Abbildung 7.2-20), die jeweils eine bestimmte Aufgabe erfüllen.

Jetzt haben wir das nächste Ziel - diese Elemente in einer einzigen, kohärenten und automatisierten Pipeline der Datenverarbeitung - einer Pipeline, ETL -Pipeline - zu kombinieren, in der alle Phasen (Laden, Validierung, Visualisierung, Export) nacheinander in einem einzigen automatisch ausführbaren Skript ausgeführt werden.

Im folgenden Beispiel wird ein kompletter Datenverarbeitungszyklus implementiert: vom Laden der CSV-Quelldatei →bis zur Überprüfung der Struktur und der Werte mittels regulärer Ausdrücke →Berechnung der Ergebnisse→ Erstellung eines visuellen Berichts im PDF-Format.

- Sie können die folgende Textabfrage an den LLM verwenden, um den entsprechenden Code abzurufen:

> Bitte schreiben Sie ein Codebeispiel, das Daten aus CSV lädt, DataFrame mit regulären Ausdrücken validiert, Identifikatoren im Format 'W-NEW' oder 'W-OLD', Energieeffizienz mit den Buchstaben 'A' bis 'G', Garantiezeit und Austauschzyklus mit numerischen Werten in Jahren prüft und am Ende einen Bericht mit einer Zählung der bestandenen und nicht bestandenen Werte erstellt, ein PDF mit einem Histogramm der Ergebnisse erzeugt und eine Textbeschreibung hinzufügt. ⏎

Die Antwort von LLM:

Create in LLM Chat

Pipeline.py

```python
# Importing necessary libraries
import pandas as pd
import re
import matplotlib.pyplot as plt
from matplotlib.backends.backend_pdf import PdfPages


# Load data
df = pd.read_csv('path_to_csv_file.csv')

# Define patterns and verify data
patterns = {
    'Window_ID': r'^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$',
    'Energy_Efficiency': r'^[A-G]\+*$',
    'Warranty_Period': r'^\d+\syears$',
    'Replacement_Cycle': r'^\d+\syears$'
}

for column, pattern in patterns.items():
    df['verified_' + column] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))

# Create report
report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
report.columns = ['Passed', 'Failed']

# Generate PDF
pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
with PdfPages(pdf_path) as pdf:
    # Plot and save bar chart
    report.plot(kind='bar', stacked=True, color=['green', 'red'])
    plt.title('Data Verification Summary')
    plt.xlabel('Data Categories')
    plt.ylabel('Count')
    plt.tight_layout()
    pdf.savefig()
    plt.close()

    # Add text page
    plt.figure()
    plt.axis('off')
    text_str = "Verification Results Summary"
    plt.text(0.1, 0.5, text_str, wrap=True)
    pdf.savefig()
    plt.close()

pdf_path
```

Run in IDE

Abb. 7.3-6 Pipeline (ETL) automatisiert den gesamten Zyklus der Datenverarbeitung: vom Herunterladen und der Validierung bis zur Erstellung eines strukturierten Berichts im PDF-Format.

Der automatisierte Code (Abb. 7.3-6) im LLM-Chatroom oder im DIE wird nach dem Kopieren des Codes die Daten aus der CSV -Datei unter Verwendung der angegebenen regulären Ausdrücke validieren, einen Bericht über die Anzahl der bestandenen und nicht bestandenen Datensätze erstellen und dann die Validierungsergebnisse als PDF -Datei speichern.

Diese ETL -Förderstruktur, bei der jeder Schritt - vom Laden der Daten bis zur Berichterstellung - als separates Modul implementiert ist, gewährleistet Transparenz, Skalierbarkeit und Reproduzierbarkeit. Die Darstellung der Validierungslogik als leicht lesbarer Python-Code macht den Prozess nicht nur für Entwickler, sondern auch für Spezialisten in den Bereichen Datenmanagement, -qualität und -analyse transparent und verständlich.

> Ansatz von Pipeline zur Automatisierung der Datenverarbeitung ermöglicht es Ihnen, Prozesse zu standardisieren, ihre Wiederholbarkeit zu erhöhen und die Anpassung an neue Projekte zu vereinfachen. So entsteht eine einheitliche Methodik für die Analyse von Daten, unabhängig von der Quelle oder der Art der Aufgabe - ob es sich nun um Compliance-Tests, die Erstellung von Berichten oder die Datenübertragung an externe Systeme handelt.

Eine solche Automatisierung reduziert menschliche Fehler, verringert die Abhängigkeit von proprietären Lösungen und erhöht die Genauigkeit und Zuverlässigkeit der Ergebnisse, so dass sie sich sowohl für operative Analysen auf Projektebene als auch für strategische Analysen auf Unternehmensebene eignen.

Pipeline -ETL: Überprüfung der Daten und Informationen der Projektelemente in CAD (BIM)

Daten aus CAD-Systemen und Datenbanken (BIM) gehören zu den anspruchsvollsten und am dynamischsten aktualisierten Datenquellen im Bereich der Bauunternehmen. Diese Anwendungen beschreiben das Projekt nicht nur anhand der Geometrie, sondern ergänzen es auch durch mehrere Schichten von Textinformationen: Volumina, Materialeigenschaften, Raumzuordnungen, Energieeffizienzwerte, Toleranzen, Lebenserwartungen und andere Attribute.

> Die Attribute, die den Objekten in den CAD-Modellen zugewiesen werden, werden in der Konstruktionsphase gebildet und bilden die Grundlage für weitere Geschäftsprozesse, einschließlich Kalkulation, Zeitplanung, Lebenszyklusbewertung und Integration mit ERP und CAFM -Systemen, wobei die Effizienz der Prozesse weitgehend von der Qualität der Daten abhängt, die aus den Konstruktionsabteilungen stammen.

Der traditionelle Ansatz zur Attributvalidierung in CAD- (BIM-) Modellen beinhaltet die manuelle Validierung (Abb. 7.2-1), die bei großen Modellmengen zu einem langwierigen und kostspieligen Prozess wird. In Anbetracht des Umfangs und der Anzahl moderner Bauprojekte und ihrer regelmäßigen Aktualisierungen wird der Prozess der Datenvalidierung und -umwandlung unhaltbar und unbezahlbar.

Generalunternehmer und Projektmanager müssen große Mengen von Projektdaten verarbeiten, darunter mehrere Versionen und Fragmente derselben Modelle. Die Daten kommen von den Planungsbüros in RVT-, DWG-, DGN-, IFC-, NWD- und anderen Formaten (Abb. 3.1-14) und müssen regelmäßig auf ihre Übereinstimmung mit den Branchen- und Unternehmensstandards überprüft werden

Die Abhängigkeit von manuellen Tätigkeiten und spezialisierter Software macht den Datenvalidierungsprozess zu einem Engpass in Arbeitsabläufen, die sich auf Daten aus unternehmensweiten Modellen beziehen. Durch Automatisierung und die Verwendung strukturierter Anforderungen kann diese Abhängigkeit beseitigt und die Geschwindigkeit und Zuverlässigkeit der Datenvalidierung drastisch erhöht werden (Abb. 7.3-7).

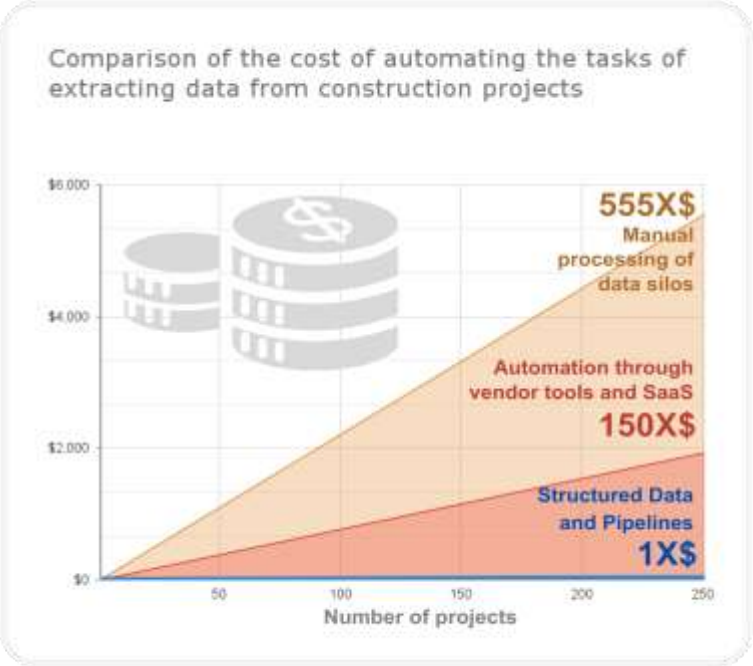


Abb. 7.3-7 Die Automatisierung erhöht die Geschwindigkeit der Datenüberprüfung und -verarbeitung, was die Arbeitskosten um das Dutzendfache reduziert [140].

Der Prozess der CAD-Datenvalidierung umfasst die Datenextraktion (ETL-Phase Extract) aus verschiedenen geschlossenen (RVT, DWG, DGN, NWS, usw.) oder offenen halbstrukturierten und parametrischen Formaten (IFC, CPXML, USD).) oder offenen halbstrukturierten und parametrischen Formaten (IFC, CPXML, USD), wobei auf jedes Attribut und seine Werte (Transformationsphase) mit Hilfe von regulären Ausdrücken RegEx (Abb. 7.3-8) Regeltabellen angewendet werden können, ein Verfahren, das wir im vierten Teil des Buches ausführlich behandelt haben.

Die Erstellung eines PDF-Fehlerberichts von und erfolgreich validierten Datensätzen sollte mit der Ausgabe (Schritt Laden) in strukturierten Formaten abgeschlossen werden, die nur validierte Entitäten berücksichtigen, die für weitere Prozesse verwendet werden können.

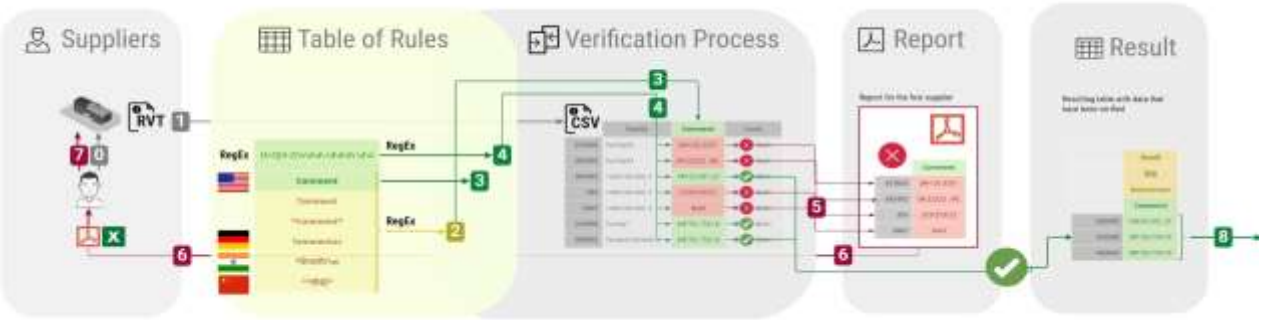


Abbildung 7.3-8 Datenvalidierungsprozess von den Projektdatenanbietern bis zum Abschlussbericht, der mit regulären Ausdrücken validiert wird.

Die Automatisierung der Validierung von Daten aus CAD-Systemen (BIM) mit strukturierten Anforderungen und dem Streaming neuer Daten, die über ETL-Pipelines (Abb. 7.3-9) verarbeitet werden, reduziert die Notwendigkeit einer manuellen Beteiligung am Validierungsprozess (jeder der Validierungs- und Datenanforderungsprozesse wurde in den vorherigen Kapiteln besprochen).

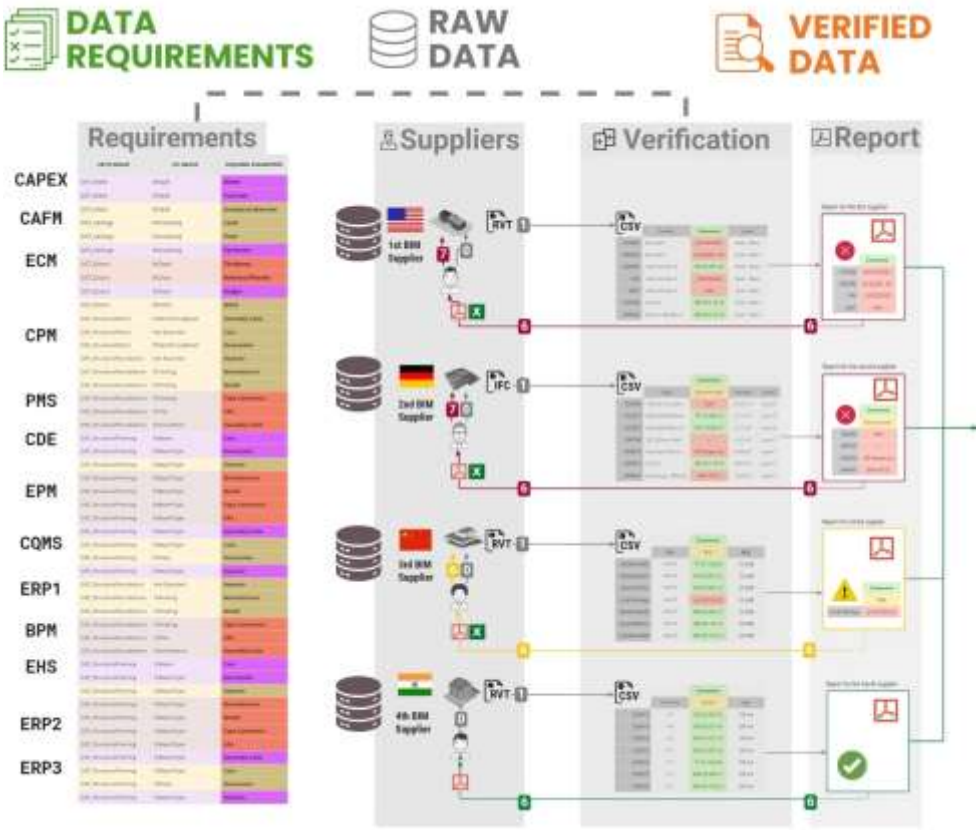


Abb. 7.3-9 Die Automatisierung der Datenvalidierung durch ETL vereinfacht die Verwaltung von Bauprojekten, indem sie die Prozesse beschleunigt.

Traditionell kann die Validierung von Modellen, die von Auftragnehmern und CAD (BIM)-Spezialisten bereitgestellt werden, Tage bis Wochen dauern. Mit der Einführung automatisierter ETL-Prozesse kann dies jedoch auf wenige Minuten reduziert werden. In einer typischen Situation stellt der Auftragnehmer fest: *"Das Modell ist validiert und entspricht den Anforderungen."* Mit dieser Aussage beginnt eine Kette der Überprüfung der vom Auftragnehmer behaupteten Datenqualität:

- Projektleiter - "*Der Auftragnehmer sagt: 'Das Modell wurde getestet, alles ist in Ordnung'*".
- Data Manager - Ladevalidierung:
  - Ein einfaches Skript in Pandas erkennt einen Verstoß in Sekundenschnelle. Automatisierung beseitigt Streitigkeiten:
    - Kategorie: OST_StructuralColumns, Parameter: FireRating IS NULL.
    - Generieren Sie eine Liste von Verstoß-IDs→ Export in Excel/PDF.

  Ein einfaches Skript in Pandas erkennt den Verstoß innerhalb von Sekunden:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"]  # Filtern
issues = df[df["FireRating"].isnull()]  # Leere Werte
issues[[["ElementID"]].to_excel("fire_rating_issues.xlsx") # IDs exportieren
```

- Datenmanager an Projektmanager - *"Eine Überprüfung von zeigt, dass in 18 Spalten der Parameter FireRating nicht ausgefüllt ist."*
- Projektleiter an Auftragnehmer - "Das *Modell wird zur Überarbeitung zurückgegeben: der Parameter FireRating ist obligatorisch, ohne ihn ist eine Abnahme nicht möglich"*

Das Ergebnis ist, dass das CAD-Modell nicht validiert werden muss, dass die Automatisierung Streitigkeiten ausschließt und dass der Auftragnehmer fast sofort einen strukturierten Bericht mit einer Liste von IDs problematischer Elemente erhält. Auf diese Weise wird der Validierungsprozess transparent, wiederholbar und vor menschlichen Fehlern geschützt (Abb. 7.3-10).

> Mit diesem Ansatz wird der Datenvalidierungsprozess zu einer technischen Funktion und nicht zu einer manuellen Qualitätskontrolle. Dies steigert nicht nur die Produktivität, sondern ermöglicht es auch, dieselbe Logik auf alle Projekte des Unternehmens anzuwenden, was eine durchgängige digitale Transformation der Prozesse vom Design bis zum Betrieb ermöglicht.

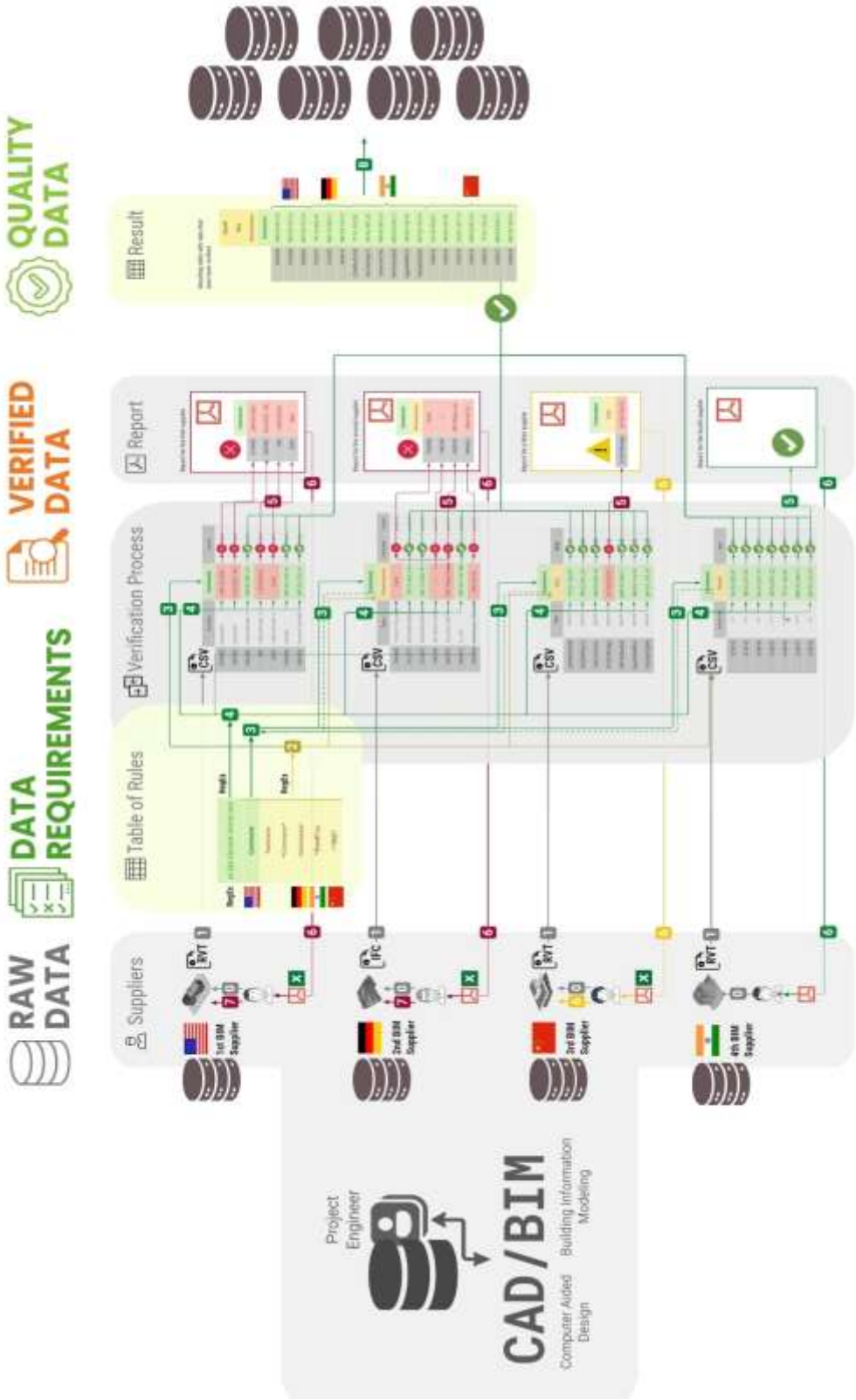


Abb. 7.3-10 Durch die Automatisierung der Prüfung von Elementattributen werden menschliche Fehler vermieden und die Fehlerwahrscheinlichkeit verringert.

Durch den Einsatz von automatisierten Pipelines (Abb. 7.3-10) können Systemanwender, die Qualitätsdaten von CAD- (BIM-) Systemen erwarten, sofort die benötigten Ausgabedaten - Tabellen, Dokumente, Bilder - erhalten und diese schnell in ihre Arbeitsaufgaben integrieren.

> Die Automatisierung von Steuerung, Verarbeitung und Analyse führt zu einem Wandel in der Art und Weise, wie das Bauprojektmanagement angegangen wird, insbesondere die Interoperabilität verschiedener Systeme, ohne den Einsatz komplexer und teurer modularer proprietärer Systeme oder geschlossener Anbieterlösungen.

Während Konzepte und Marketing-Akronyme kommen und gehen, werden die Prozesse zur Validierung der Datenanforderungen selbst für immer ein integraler Bestandteil der Geschäftsprozesse bleiben. Anstatt immer mehr spezialisierte Formate und Standards zu schaffen, sollte die Bauindustrie auf Tools zurückgreifen, die sich in anderen Branchen bereits bewährt haben. Heute gibt es leistungsstarke Plattformen für die Automatisierung der Datenverarbeitung und der Prozessintegration, die es den Unternehmen ermöglichen, den Zeitaufwand für Routinevorgänge erheblich zu reduzieren und Fehler beim Extrahieren, Transformieren und Laden zu minimieren.

Eines der beliebtesten Beispiele für Lösungen zur Automatisierung und Orchestrierung von ETL -Prozessen ist Apache Airflow, das die Organisation komplexer Berechnungsprozesse und die Verwaltung von ETL-Pipelines ermöglicht. Neben Airflow werden auch andere ähnliche Lösungen wie Apache NiFi für Daten-Routing und -Streaming und n8n für die Automatisierung von Geschäftsprozessen aktiv genutzt.

KAPITEL 7.4.
ORCHESTRIERUNG VON ETL UND ARBEITSABLÄUFEN: PRAKTISCHE LÖSUNGEN

DAG und Apache Airflow: Automatisierung und Orchestrierung von Arbeitsabläufen

Apache Airflow ist eine freie und quelloffene Plattform, die für die Automatisierung, Orchestrierung und Überwachung von Arbeitsabläufen (ETL -Förderer) entwickelt wurde.

Die Arbeit mit großen Datenmengen ist tagtäglich erforderlich:

- Herunterladen von Dateien aus verschiedenen Quellen - Extrahieren von (z. B. von Lieferanten oder Kunden).
- Umwandlung dieser Daten in das gewünschte Format - Umwandlung (Strukturierung, Bereinigung und Validierung)
- Senden von Ergebnissen zur Überprüfung und Erstellen von Berichten - Laden von (Hochladen in erforderliche Systeme, Dokumente, Datenbanken oder Dashboards).

Die manuelle Ausführung solcher ETL-Prozesse nimmt viel Zeit in Anspruch und birgt das Risiko menschlicher Fehler. Eine Änderung in der Datenquelle oder ein Fehler in einem der Schritte kann zu Verzögerungen und falschen Ergebnissen führen.

Mit Automatisierungstools wie Apache Airflow können Sie einen zuverlässigen ETL -Förderer aufbauen, Fehler minimieren, die Verarbeitungszeit verkürzen und sicherstellen, dass die Daten in jeder Phase korrekt sind. Das Herzstück von Apache Airflow ist das Konzept des DAG (Directed Acyclic Graph) - ein gerichteter azyklischer Graph, in dem jede Aufgabe (Operator) mit anderen Abhängigkeiten verbunden ist und streng in einer bestimmten Reihenfolge ausgeführt wird. DAG eliminiert Zyklen, was eine logische und vorhersehbare Struktur der Aufgabenausführung ermöglicht.

Airflow kümmert sich um die Orchestrierung - die Verwaltung von Abhängigkeiten zwischen Aufgaben, die Steuerung von Ausführungsplänen, die Verfolgung des Status und die automatische Reaktion auf Ausfälle. Dieser Ansatz minimiert manuelle Eingriffe und gewährleistet die Zuverlässigkeit des gesamten Prozesses.

> **Task Orchestrator** ist ein Werkzeug oder System zur Verwaltung und Steuerung der Aufgabenausführung in komplexen Computer- und Informationsumgebungen. Es erleichtert den Prozess der Bereitstellung, Automatisierung und Verwaltung der Aufgabenausführung, um die Leistung zu verbessern und die Ressourcen zu optimieren.

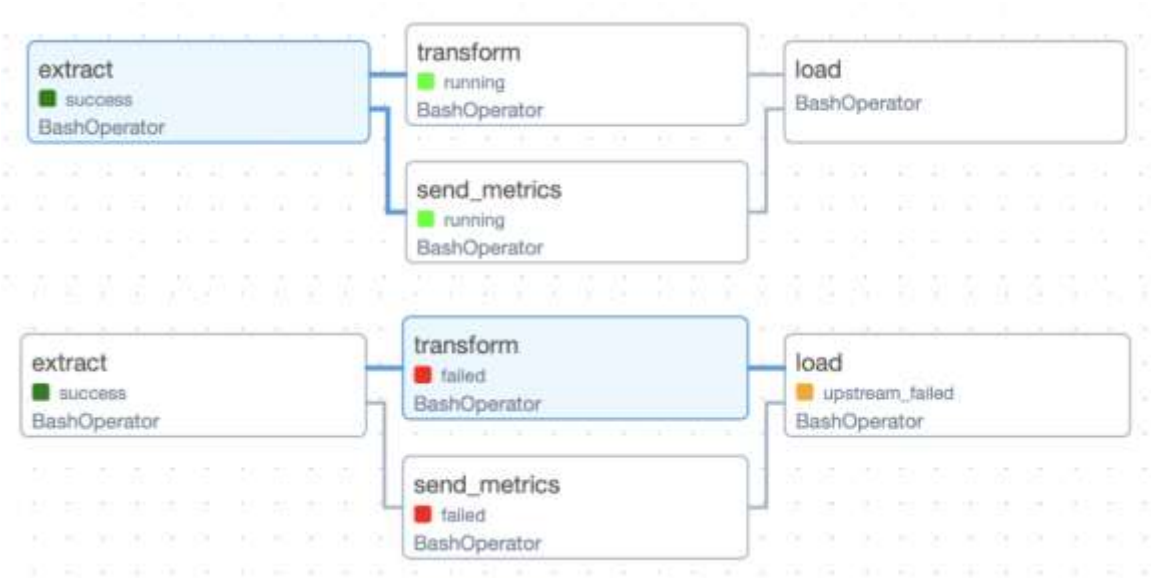


Abbildung 7.4-1 Apache Airflow bietet eine benutzerfreundliche Oberfläche, auf der Sie die DAG -ETL visualisieren, Ausführungsprotokolle, den Startstatus von Aufgaben und vieles mehr anzeigen können.

Airflow wird häufig für die Orchestrierung und Automatisierung von verteilter Datenverarbeitung, Datenverarbeitung, ETL (Extract, Transform, Load), Prozessmanagement, Aufgabenplanung und andere Datenszenarien verwendet. Standardmäßig verwendet Apache Airflow SQLite als Datenbank.

Ein Beispiel für einen einfachen DAG, ähnlich wie ETL, besteht aus Tasks - Extract, Transform und Load. In dem Graphen, der über die Benutzeroberfläche gesteuert wird (Abb. 7.4-1), ist die Reihenfolge der Ausführung von Tasks (Codefragmenten) festgelegt: z. B. wird zuerst Extract ausgeführt, dann Transform (und sending_metrics), und Load Task schließt die Arbeit ab. Wenn alle Aufgaben abgeschlossen sind, gilt der Datenladevorgang als erfolgreich.

Apache Airflow: praktische Anwendung zur ETL-Automatisierung

Apache Airflow wird häufig zur Organisation komplexer Datenverarbeitungsprozesse verwendet und ermöglicht den Aufbau flexibler ETL -Förderer. Apache Airflow kann entweder über eine Weboberfläche oder programmatisch über Python-Code betrieben werden (Abb. 7.4-2). In der Weboberfläche (Abb. 7.4-3) können Administratoren und Entwickler DAGs visuell überwachen, Aufgaben ausführen und die Ausführungsergebnisse analysieren.

Mit DAG können Sie eine klare Reihenfolge der Aufgaben festlegen, Abhängigkeiten zwischen ihnen verwalten und automatisch auf Änderungen in den Quelldaten reagieren. Betrachten wir ein Beispiel für die Verwendung von Airflow zur Automatisierung der Berichtsverarbeitung (Abb. 7.4-2).

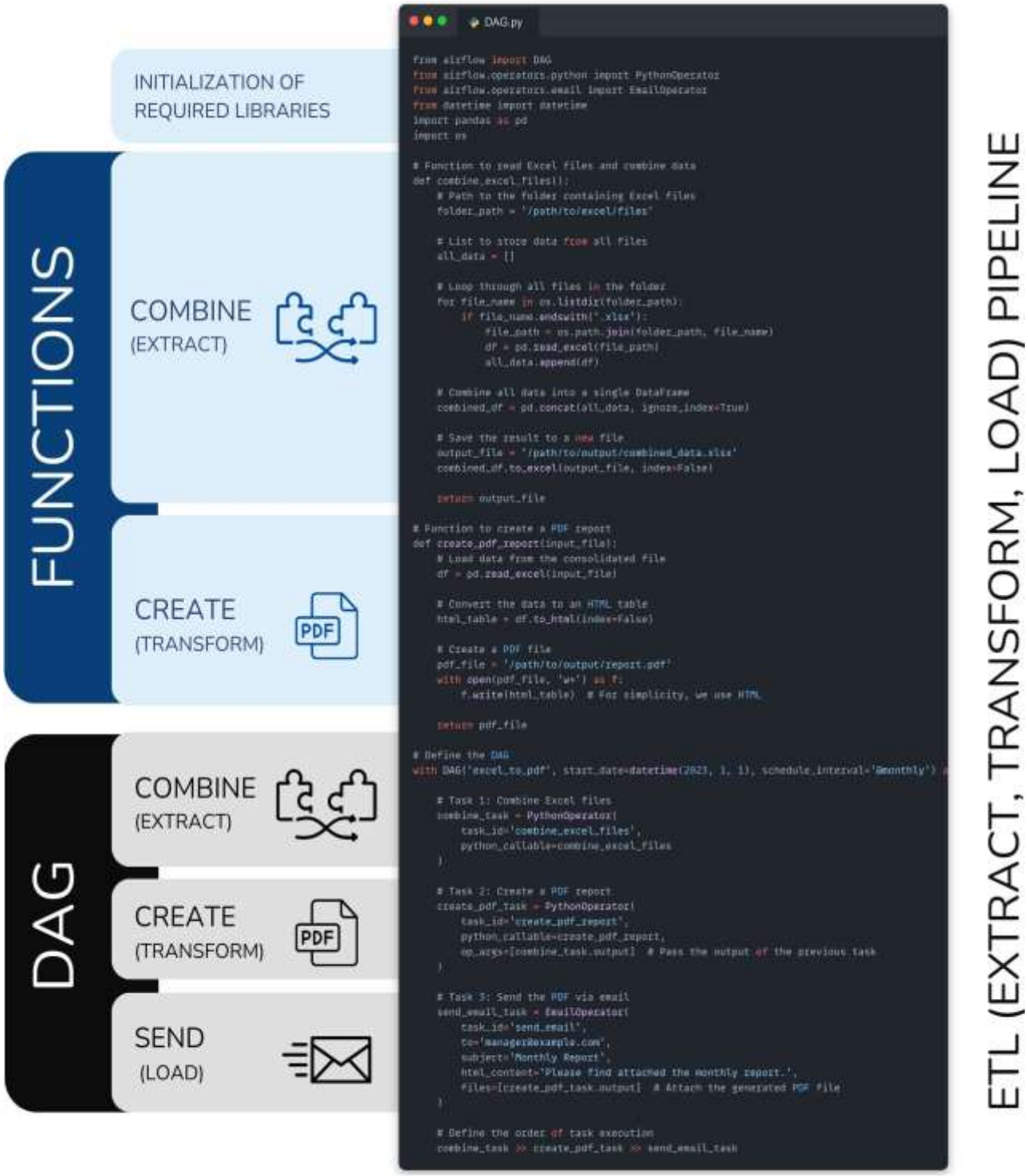


Abb. 7.4-2 ETL-Konzept -Förderer für die Datenverarbeitung mit Apache Airflow.

In diesem Beispiel (Abb. 7.4-2) wird die DAG betrachtet, die Schlüsselaufgaben innerhalb des ETL -Förderers wahrnimmt:

- **Lesen von Excel -Dateien (Extract):**
  - Sequentielle Durchquerung aller Dateien in einem bestimmten Verzeichnis.
  - Lesen der Daten aus jeder Datei mit Hilfe der Pandas-Bibliothek.
  - Zusammenführen aller Daten in einen einzigen DataFrame.
- **PDF -Dokument erstellen (Transform):**
  - Transformieren Sie den zusammengeführten DataFrame in eine HTML -Tabelle.
  - Speichern Sie die Tabelle als PDF (in der Demoversion - über HTML).
- **Versenden eines Berichts per E-Mail (Load):**
  - Wenden Sie EmailOperator an, um ein PDF -Dokument per E-Mail zu versenden.
- **Konfigurieren der DAG:**
  - Festlegen der Reihenfolge der Aufgaben: Daten extrahieren→ Bericht erstellen→ Senden.
  - Zuweisung eines Startzeitplans (@monatlich - erster Tag jedes Monats).

Das automatisierte ETL -Beispiel (Abb. 7.4-2) zeigt, wie man Daten aus Excel -Dateien sammelt, ein PDF -Dokument erstellt und per E-Mail versendet. Dies ist nur einer von vielen möglichen Anwendungsfällen für Airflow. Dieses Beispiel kann an jede spezifische Aufgabe angepasst werden, um die Datenverarbeitung zu vereinfachen und zu automatisieren.

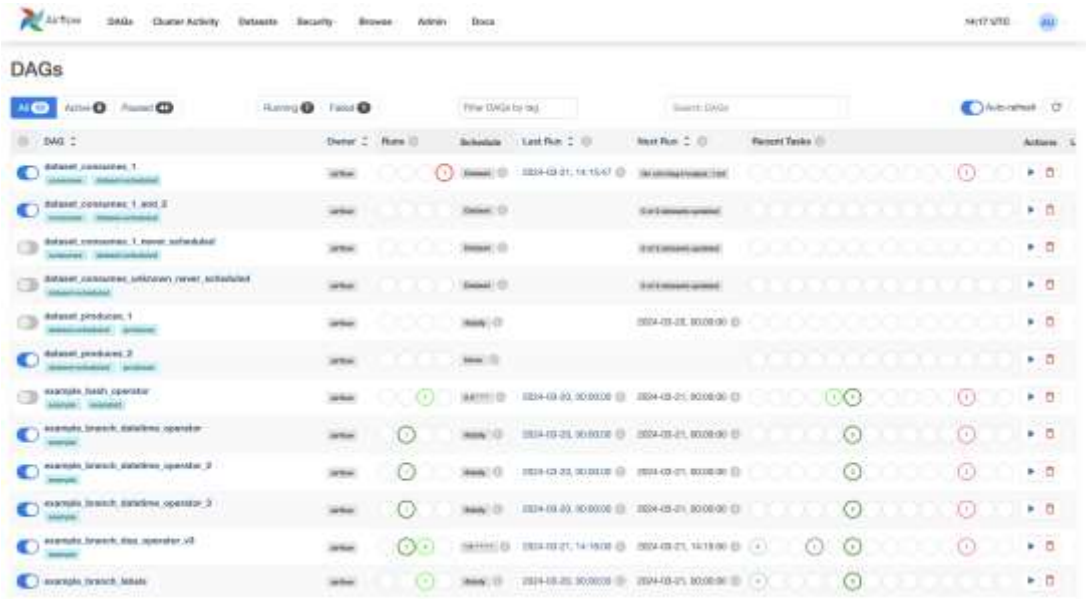


Abb. 7.4-3 Übersicht über alle DAGs in der Umgebung mit Informationen über die letzten Läufe.

Die Apache Airflow-Webschnittstelle (Abb. 7.4-3) bietet eine umfassende visuelle Umgebung für die Verwaltung von Daten-Workflows. Sie zeigt DAGs als interaktive Graphen an, bei denen die Knoten Aufgaben und die Kanten Abhängigkeiten zwischen ihnen darstellen, was die Verfolgung komplexer Daten-Workflows erleichtert. Die Benutzeroberfläche enthält ein Dashboard mit Informationen über den Status der Aufgabenausführung, den Verlauf der Ausführung, detaillierte Protokolle und Leistungsmetriken. Adminis-

tratoren können über eine intuitive Benutzeroberfläche Aufgaben manuell starten, fehlgeschlagene Vorgänge neu starten, DAGs anhalten und Umgebungsvariablen konfigurieren.

Eine solche Architektur kann durch Datenvalidierung, Benachrichtigungen über den Ausführungsstatus, Integration mit externen APIs oder Datenbanken ergänzt werden. Airflow erlaubt es, die DAG flexibel anzupassen: neue Aufgaben hinzufügen, ihre Reihenfolge ändern, Ketten kombinieren - was sie zu einem effektiven Werkzeug für die Automatisierung komplexer Datenverarbeitungsprozesse macht. Wenn Sie DAG in der Airflow-Weboberfläche (Abb. 7.4-3, Abb. 7.4-4) ausführen, können Sie den Status der Aufgaben überwachen. Das System verwendet eine farbliche Kennzeichnung:

- Grün - die Aufgabe wurde erfolgreich abgeschlossen.
- Gelb - der Prozess ist im Gange.
- Rot - ein Fehler bei der Durchführung der Aufgabe.

Im Falle von Fehlern (z. B. fehlende Datei oder defekte Datenstruktur) leitet das System automatisch eine Benachrichtigung ein.

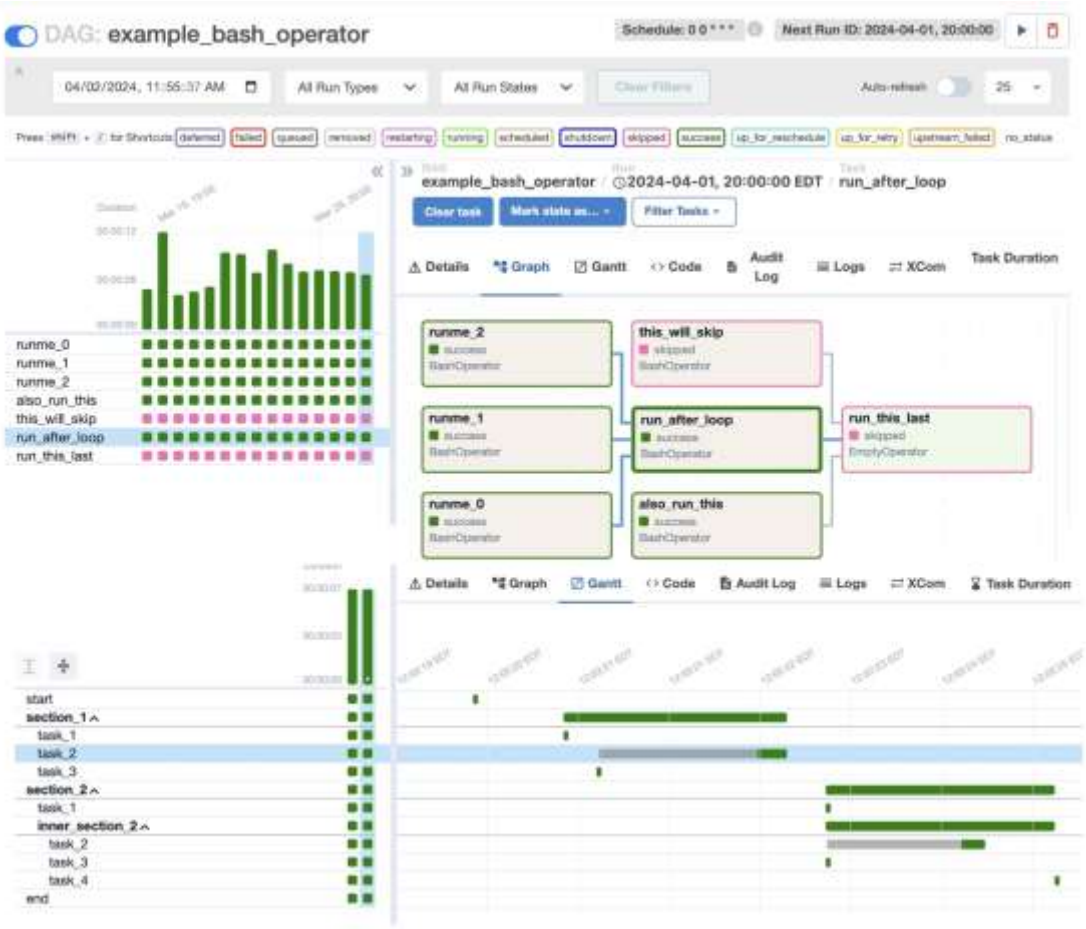


Abbildung 7.4-4 Apache Airflow vereinfacht die Problemdiagnose, die Prozessoptimierung und die Zusammenarbeit im Team bei komplexen Datenverarbeitungspipelines erheblich.

> Apache Airflow ist praktisch, weil es Routineaufgaben automatisiert, so dass sie nicht mehr manuell ausgeführt werden müssen. Es sorgt für Zuverlässigkeit, indem es die Prozessausführung überwacht und sofortige Fehlermeldungen liefert. Die Flexibilität des Systems macht es einfach, neue Aufgaben hinzuzufügen oder bestehende zu ändern und die Arbeitsabläufe an sich ändernde Anforderungen anzupassen.

Neben Apache Airflow gibt es ähnliche Tools für die Orchestrierung von Arbeitsabläufen. Das quelloffene und kostenlose Prefect (Abb. 7.3-5) bietet beispielsweise eine einfachere Syntax und lässt sich besser in Python integrieren. Luigi, das von Spotify entwickelt wurde, bietet ähnliche Funktionen und funktioniert gut mit Big Data. Ebenfalls erwähnenswert sind Kronos und Dagster, die moderne Ansätze für die Erstellung von Pipelines mit Schwerpunkt auf Modularität und Skalierbarkeit bieten. Die Wahl des Task-Orchestrierungswerkzeugs hängt von den spezifischen Anforderungen des Projekts ab, aber sie alle helfen bei der Automatisierung komplexer ETL Datenprozesse

Besonders hervorzuheben ist Apache NiFi, eine Open-Source-Plattform, die für das Streaming und Routing von Daten entwickelt wurde. Im Gegensatz zu Airflow, dessen Schwerpunkt auf der Stapelverarbeitung und dem Abhängigkeitsmanagement liegt, konzentriert sich NiFi auf die Datenumwandlung in Echtzeit und das flexible Routing zwischen Systemen.

Apache NiFi für Routing und Datenkonvertierung

Apache NiFi ist eine leistungsstarke Open-Source-Plattform, die für die Automatisierung des Datenflusses zwischen verschiedenen Systemen entwickelt wurde. Ursprünglich wurde sie 2006 von der Nationalen Sicherheitsagentur der USA (NSA) unter dem Namen "Niagara Files" für den internen Gebrauch entwickelt. Im Jahr 2014 wurde das Projekt als Open-Source-Projekt an die Apache Software Foundation übertragen und Teil ihrer Initiativen zum Technologietransfer [141].

Apache NiFi wurde entwickelt, um Daten in Echtzeit zu sammeln, zu verarbeiten und zu übertragen. Im Gegensatz zu Airflow, das mit Stapelverarbeitungsaufgaben arbeitet und genau definierte Zeitpläne erfordert, arbeitet NiFi in einem Stromverarbeitungsmodus, so dass Daten kontinuierlich zwischen verschiedenen Diensten übertragen werden können.

> Apache NiFi ist ideal für die Integration mit IoT Geräten, Bausensoren, Überwachungssystemen und z.B. für die Stream-Validierung von CAD Formaten auf einem Server, wo eine sofortige Reaktion auf Datenänderungen erforderlich sein kann.

Mit eingebauten Filter-, Transformations- und Routing-Tools ermöglicht NiFi die Standardisierung von Daten (Transform), bevor sie an Speicher- oder Analysesysteme übertragen werden (Load). Einer seiner Hauptvorteile ist die eingebaute Sicherheitsunterstützung und Zugangskontrolle, die es zu einer zuverlässigen Lösung für den Umgang mit sensiblen Informationen macht.

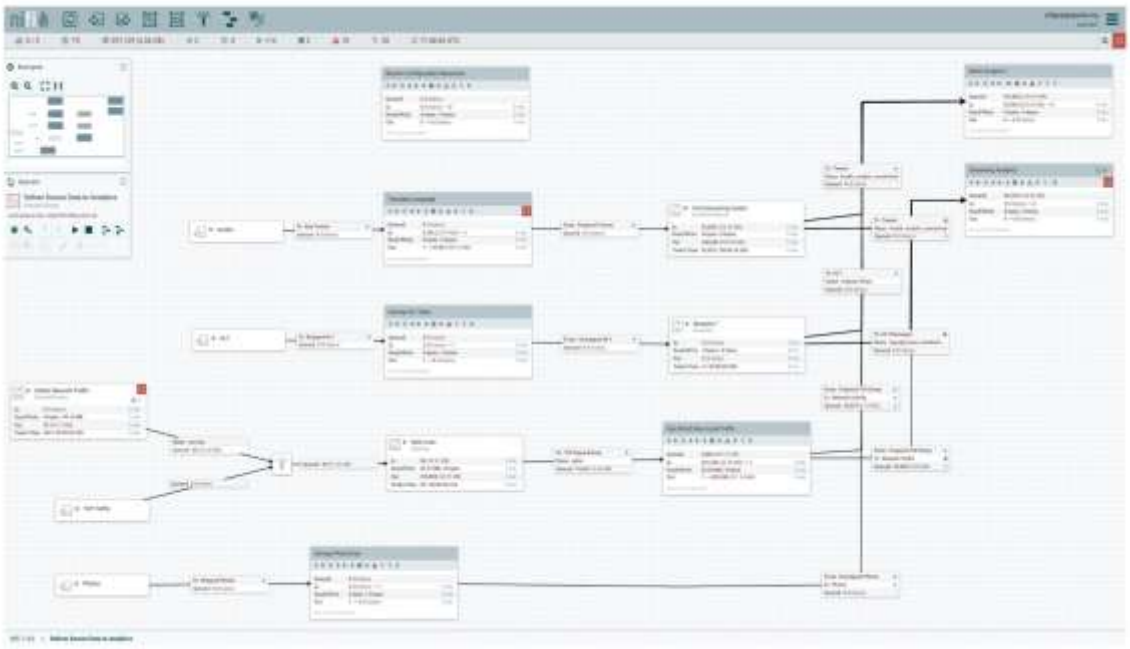


Abbildung 7.4-5 Grafische Darstellung des Datenflusses in der Apache-NiFi-Schnittstelle.

Apache NiFi bewältigt effizient Echtzeit-Datenströme, Filter- und Routing-Aufgaben. Es ist ideal für technisch intensive Szenarien, bei denen eine stabile Informationsübertragung zwischen Systemen und ein hoher Durchsatz wichtig sind.

Wenn jedoch das Hauptziel darin besteht, verschiedene Dienste zu integrieren, Routinevorgänge zu automatisieren und Arbeitsabläufe ohne tiefe Programmierkenntnisse schnell einzurichten, sind Lösungen mit einer niedrigen Einstiegsschwelle und maximaler Flexibilität gefragt. Ein solches Tool ist n8n - eine Low-Code /No-Code Plattform, die sich auf Geschäftsautomatisierung und visuelle Orchestrierung von Prozessen konzentriert.

n8n Low-Code, No-Code Prozessorchestrierung

n8n ist eine Open Source Low-Code / No-Code Plattform für den Aufbau automatisierter Workflows, die sich durch Benutzerfreundlichkeit, Flexibilität und die Möglichkeit der schnellen Integration mit einer breiten Palette externer Dienste auszeichnet.

> **No-Code** ist eine Methode zur Erstellung digitaler Produkte ohne das Schreiben von Code. Alle Elemente des Prozesses - von der Logik bis zur Schnittstelle - werden ausschließlich mit Hilfe von visuellen Tools realisiert. No-Code-Plattformen richten sich an Nutzer ohne technischen Hintergrund und ermöglichen die schnelle Erstellung von Automatisierungen, Formularen, Integrationen und Webanwendungen. Beispiel: Ein Nutzer richtet über eine Drag-and-Drop-Schnittstelle ohne Programmierkenntnisse den automatischen Versand von Benachrichtigungen oder die Integration von Google Sheets ein.

Mit Open-Source- und lokalen Bereitstellungsmöglichkeiten gibt n8n Unternehmen bei der Automatisierung und Erstellung von ETL-Pipelines die vollständige Kontrolle über ihre Daten und gewährleistet gleichzeitig Sicherheit und Unabhängigkeit von Cloud-Anbietern.

Im Gegensatz zu Apache Airflow, das auf Berechnungsaufgaben mit starrer Orchestrierung ausgerichtet ist und Python-Kenntnisse voraussetzt, bietet n8n einen visuellen Editor, der die Erstellung von Skripten ohne Kenntnisse von Programmiersprachen ermöglicht (Abb. 7.4-6). Obwohl die Schnittstelle die Erstellung von automatisierten Prozessen ohne das Schreiben von Code ermöglicht (No-Code), können Benutzer in komplexeren Szenarien ihre eigenen JavaScript und Python -Funktionen hinzufügen, um die Fähigkeiten zu erweitern (Low-Code).

> **Low-Code** ist ein Ansatz für die Softwareentwicklung, bei dem die grundlegende Logik einer Anwendung oder eines Prozesses über eine grafische Oberfläche und visuelle Elemente erstellt wird und Programmcode nur zur Anpassung oder Erweiterung der Funktionalität verwendet wird. Low-Code-Plattformen ermöglichen es, die Entwicklung von Lösungen erheblich zu beschleunigen, indem nicht nur Programmierer, sondern auch Geschäftsanwender mit grundlegenden technischen Kenntnissen einbezogen werden. Ein Beispiel: ein Benutzer kann einen Geschäftsprozess aus vorgefertigten Blöcken aufbauen und bei Bedarf sein eigenes Skript in JavaScript oder Python hinzufügen.

Obwohl n8n als Plattform mit einer niedrigen Einstiegsschwelle positioniert ist, sind grundlegende Programmierkenntnisse, Verständnis für Webtechnologien und Fähigkeiten im Umgang mit API erforderlich. Die Flexibilität des Systems ermöglicht die Anpassung an ein breites Spektrum von Aufgaben - von der automatisierten Datenverarbeitung bis zur Integration mit Messengern, IoT -Geräten und Cloud-Diensten.

Die wichtigsten Merkmale und Vorteile des n8n:

- Open-Source- und lokale Bereitstellungsoptionen gewährleisten eine vollständige Datenkontrolle, die Einhaltung von Sicherheitsvorschriften und die Unabhängigkeit von Cloud-Anbietern.
- **Integration mit über 330 Diensten**, darunter CRM, ERP, E-Commerce, Cloud-Plattformen, Messenger und Datenbanken.
- **Szenario-Flexibilität:** von einfachen Benachrichtigungen bis zu komplexen Ketten mit API-Verarbeitung von -Anfragen, Entscheidungslogik und Anbindung von KI -Diensten.
- **Unterstützung für JavaScript und Python**: Benutzer können bei Bedarf benutzerdefinierten Code einbetten und so die Automatisierungsfunktionen erweitern.
- **Intuitive visuelle Schnittstelle:** ermöglicht die schnelle Konfiguration und Visualisierung aller Prozessschritte.

Plattformen der Low-Code-Klasse bieten Tools für die Erstellung digitaler Lösungen mit minimalem Code und sind damit ideal für Teams, die über kein tiefes technisches Know-how verfügen, aber Prozesse automatisieren müssen.

Im Bauwesen kann n8n zur Automatisierung einer Vielzahl von Prozessen verwendet werden, wie z. B. zur Integration mit Projektmanagement-Systemen, zur Überprüfung von Abläufen, zum Schreiben von sofort einsetzbaren Berichten und Briefen, zur automatischen Aktualisierung von Materialbestandsdaten, zum

Senden von Benachrichtigungen über den Aufgabenstatus an Teams und vieles mehr. Eine maßgeschneiderte Pipeline in n8n kann manuelle Vorgänge um ein Vielfaches reduzieren, die Fehlerwahrscheinlichkeit verringern und die Entscheidungsfindung für die Projektdurchführung beschleunigen.

> Sie können aus fast zweitausend vorgefertigten, kostenlosen und quelloffenen n8n Pipeline, verfügbar unter: n8n.io/workflows, wählen, um sowohl Bauabläufe als auch persönliche Aufgaben zu automatisieren und Routinearbeiten zu reduzieren.

Nehmen Sie zum Beispiel eine der vorgefertigten Pipeline-Vorlagen, die kostenlos unter n8n.io [142] erhältlich ist und automatisch Antwortentwürfe in Gmail erstellt (Abb. 7.4-6), die Nutzern helfen, die eine große Menge an E-Mails erhalten oder Schwierigkeiten beim Verfassen von Antworten haben.

Diese n8n "Gmail AI Auto-Responder: Erstellen von Antwortentwürfen auf eingehende E-Mails"-Vorlage (Abb. 7.4-6) analysiert eingehende E-Mails mit LLM von ChatGPT, bestimmt, ob eine Antwort erforderlich ist, erzeugt einen Entwurf von ChatGPT und konvertiert den Text in HTML und fügt ihn der Nachrichtenkette in Gmail hinzu. Dabei wird die E-Mail nicht automatisch versendet, so dass Sie die Antwort manuell bearbeiten und genehmigen können. Die Einrichtung dauert etwa 10 Minuten und umfasst die OAuth-Konfiguration der Gmail-API und die OpenAI-API-Integration. Das Ergebnis ist eine bequeme und kostenlose Lösung zur Automatisierung der routinemäßigen E-Mail-Kommunikation, ohne die Kontrolle über den Inhalt der E-Mails zu verlieren.

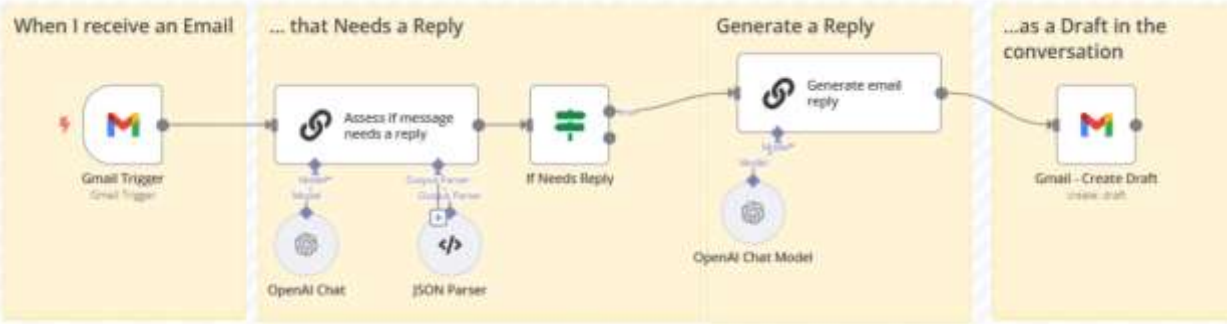


Abb. 7.4-6 Automatisierter Prozess zur Erstellung von E-Mail-Antworten mit n8n.

Ein weiteres Beispiel für die Automatisierung mit n8n ist die Suche nach günstigen Angeboten auf dem Immobilienmarkt [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", sammelt täglich relevante Angebote, die den vorgegebenen Kriterien entsprechen, mit Hilfe der Zillow API. Es berechnet automatisch die wichtigsten Investitionskennzahlen (Cash on Cash ROI, monatlicher Cashflow, Anzahlung), aktualisiert Google Sheets und sendet einen zusammenfassenden Bericht per E-Mail (Abb. 7.4-7), so dass Investoren Zeit sparen und schnell auf die besten Angebote reagieren können.

Abb. 7.4-7 Automatisierter Prozess zur Bewertung der Investitionsattraktivität von Immobilien.

Seine Flexibilität und Erweiterbarkeit machen das n8n zu einem wertvollen Werkzeug für Unternehmen, die mit relativ einfachen und kostenlosen Open-Source-Tools die digitale Transformation vorantreiben und ihre Wettbewerbsfähigkeit auf dem Markt steigern wollen.

Tools wie Apache NiFi, Airflow und n8n können als drei Ebenen der Datenverarbeitung betrachtet werden (Abb. 7.4-8). NiFi verwaltet den Datenfluss und sorgt für die Bereitstellung und Umwandlung der Daten, Airflow orchestriert die Aufgabenausführung, indem es Daten in Verarbeitungspipelines zusammenfasst, und n8n automatisiert die Integration mit externen Diensten und verwaltet die Geschäftslogik.

	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Abb. 7.4-8 Apache Airflow, Apache NiFi und n8n können als drei komplementäre Schichten einer modernen Datenmanagement-Architektur betrachtet werden.

Zusammen bilden diese kostenlosen und quelloffenen Tools ein Beispiel für ein effektives Ökosystem für das Daten- und Prozessmanagement in der Bauindustrie, das es Unternehmen ermöglicht, Informationen für die Entscheidungsfindung und Prozessautomatisierung zu nutzen.

Nächste Schritte: Umstellung von manuellen Vorgängen auf analytikbasierte Lösungen

Bauunternehmen arbeiten heute in einem Umfeld großer Unsicherheit: schwankende Materialpreise, verspätete Lieferungen, Arbeitskräftemangel und knappe Projektfristen. Der Einsatz von analytischen Dashboards, ETL -Förderern und BI-Systemen hilft den Unternehmen, Problembereiche schnell zu erkennen, die Ressourceneffizienz zu bewerten und Veränderungen vorherzusagen bevor sie zu finanziellen Verlusten führen.

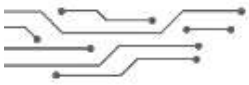
Um diesen Teil zusammenzufassen, lohnt es sich, die wichtigsten praktischen Schritte hervorzuheben, die Ihnen helfen werden, die besprochenen Technologien bei Ihren täglichen Aufgaben anzuwenden:

- Implementierung von Datenvisualisierungen und analytischen Dashboards
  - ☐ Beherrschen des Prozesses der Erstellung von Dashboards zur Überwachung der wichtigsten Leistungsindikatoren (KPIs)
  - ☐ Nutzen Sie Visualisierungstools für Ihre Daten (Power BI, Tableau, Matplotlib, Plotly)

- Automatisierung der Datenverarbeitung durch ETL -Prozesse
  - Einrichtung einer automatischen Datenerfassung aus verschiedenen Quellen (Dokumentation, Tabellen, CAD) über ETL-Prozesse
  - Datenumwandlung (z. B. Überprüfung mit regulären Ausdrücken oder Berechnung) mit Python-Skripten organisieren
  - Versuchen Sie, automatische PDF- (oder DOC-) Berichte mit der FPDF-Bibliothek einzurichten, indem Sie Daten aus Excel-Dateien verwenden oder Informationen aus anderen PDF-Dokumenten extrahieren
- Verwendung von Sprachmodellen (LLM) für die Automatisierung
  - Verwendung großer Sprachmodelle (LLMs), um Code zu generieren, der bei der Extraktion und Analyse von Daten aus unstrukturierten Dokumenten hilft
  - Machen Sie sich mit dem n8n-Automatisierungstool vertraut und erkunden Sie die vorgefertigten Vorlagen und Fallstudien auf der Website. Bestimmen Sie, welche Prozesse Ihrer Arbeit mit dem No-Code/Low-Code-Ansatz vollständig automatisiert werden können.

Ein analytischer Ansatz für die Daten- und Prozessautomatisierung reduziert nicht nur den Zeitaufwand für Routinevorgänge, sondern verbessert auch die Qualität der Entscheidungsfindung. Unternehmen, die visuelle Analysetools und ETL-Pipelines einsetzen, erhalten die Möglichkeit, schnell auf Veränderungen zu reagieren

Die Automatisierung von Geschäftsprozessen mit Tools wie n8n, Airflow und NiFi ist nur der erste Schritt zur digitalen Reife. Der nächste Schritt ist die qualitativ hochwertige Speicherung und Verwaltung der Daten, die die Grundlage für die Automatisierung bilden. In Teil 8 befassen wir uns eingehend mit der Frage, wie Bauunternehmen eine nachhaltige Datenspeicherarchitektur aufbauen können, die von einem Chaos von Dokumenten und Dateien in verschiedenen Formaten zu zentralisierten Speicher- und Analyseplattformen übergeht.



VIII TEIL

SPEICHERUNG UND VERWALTUNG VON DATEN IM BAUWESEN

Teil 8 befasst sich mit modernen Datenspeicher- und -verwaltungstechnologien für das Baugewerbe. Er analysiert effiziente Formate für den Umgang mit großen Informationsmengen - von einfachen CSV und XLSX bis zu den leistungsfähigeren Apache Parquet und ORC mit einem detaillierten Vergleich ihrer Möglichkeiten und Grenzen. Die Konzepte von Data Warehouses (DWH), Data Lakes) und deren hybride Lösungen (Data Lakehouse), sowie die Prinzipien von Data Governance) und Data Minimalism) werden diskutiert. Die Problematik des Datensumpfes) und Strategien zur Vermeidung von Chaos in Informationssystemen werden ausführlich behandelt. Neue Ansätze für die Arbeit mit Daten werden vorgestellt, darunter Vektordatenbanken und ihre Anwendung in der Konstruktion durch das Konzept der Bounding Box. In diesem Teil wird auch auf die Methoden DataOps und VectorOps als neue Standards für die Organisation von Daten-Workflows eingegangen.

KAPITEL 8.1.
DATENINFRASTRUKTUR: VON SPEICHERFORMATEN ZU DIGITALEN REPOSITORIES

Datenatome: die Grundlage eines effektiven Informationsmanagements

Alles im Universum besteht aus kleinsten Bausteinen - Atomen und Molekülen, und im Laufe der Zeit kehren alle lebenden und nicht lebenden Dinge unweigerlich in diesen Ausgangszustand zurück. In der Natur vollzieht sich dieser Prozess mit erstaunlicher Geschwindigkeit, die wir versuchen, auf die vom Menschen gesteuerten Prozesse zu übertragen.

> Im Wald werden alle lebenden Organismen schließlich in eine nahrhafte Substanz umgewandelt, die als Grundlage für neue Pflanzen dient. Diese Pflanzen wiederum werden zur Nahrung für neue Lebewesen, die aus denselben Atomen bestehen, die vor Millionen von Jahren das Universum geschaffen haben.

In der Geschäftswelt ist es auch wichtig, komplexe, vielschichtige Strukturen in ihre grundlegendsten, minimal verarbeiteten Einheiten zu zerlegen - ähnlich wie Atome und Moleküle in der Natur. Auf diese Weise können Datenatome effizient gespeichert und verwaltet werden, so dass sie zu einer reichhaltigen, fruchtbaren Grundlage werden, die zu einer Schlüsselressource für das Wachstum der Analytik und die Entscheidungsqualität wird.

Abb. 8.1-1 Die Analyse und Entscheidungsfindung basiert auf wiederverwendeten Daten, die einmal verarbeitet und gespeichert wurden.

Musikalische Kompositionen bestehen aus Noten, die sich zu komplexen Musikstücken zusammenfügen, während Wörter aus einer primitiven Einheit, dem Buchstaben-Laut, entstehen. Ob in der Natur, der Wissenschaft, der Wirtschaft, der Kunst oder der Technik, die Welt zeigt eine bemerkenswerte Einheit und Harmonie in ihrem Streben nach Zerstörung, Struktur, Kreislauf und Schöpfung. In ähnlicher Weise werden Prozesse in Kalkulationssystemen in winzige strukturierte Einheiten - Ressourcenpositionen - auf der Ebene der Kalkulation und der Zeitpläne zerlegt. Diese Einheiten werden dann wie Notizen verwendet, um komplexere Berechnungen und Zeitpläne zu erstellen. Das gleiche Prinzip wird von computergestützten Entwurfssystemen verwendet, in denen komplexe Architektur- und Ingenieurprojekte aus Grundelementen - einzelnen Elementen und Bibliothekskomponenten - aufgebaut werden, aus denen ein vollständiges 3D-Modell des Projekts eines komplexen Gebäudes oder einer Struktur erstellt wird.

> Das der Natur und der Wissenschaft innewohnende Konzept der Zyklizität und Struktur spiegelt sich auch in der modernen Welt der Daten wider. So wie in der Natur alle Lebewesen zu Atomen und Molekülen zurückkehren, so neigen auch in der Welt der modernen Datenverarbeitungswerkzeuge die Informationen dazu, zu ihrer primitivsten Form zurückzukehren.

Die kleinsten Elemente mit ihrer endlichen Unteilbarkeit sind die Grundbausteine der Geschäftsprozesse. Es ist wichtig, von Anfang an sorgfältig zu überlegen, wie diese winzigen Bausteine aus verschiedenen Quellen gesammelt, strukturiert (in Atome zerlegt) und gespeichert werden können. Bei der Organisation und Speicherung von Daten geht es nicht nur darum, sie in ihre Einzelteile zu zerlegen. Ebenso wichtig ist es, dafür zu sorgen, dass sie integriert und strukturiert gespeichert werden, damit die Daten bei Bedarf leicht abgerufen, analysiert und für die Entscheidungsfindung genutzt werden können.

Um Informationen effizient zu verarbeiten, müssen das Format und die Methoden der Datenspeicherung sorgfältig ausgewählt werden - so wie der Boden für das Wachstum von Bäumen vorbereitet werden muss. Die Datenspeicher müssen so organisiert werden, dass eine hohe Qualität und Aktualität der Informationen gewährleistet ist und redundante oder irrelevante Daten eliminiert werden. Je besser dieser "Informationsboden" strukturiert ist, desto schneller und genauer können die Nutzer die richtigen Daten finden und analytische Probleme lösen.

Informationsspeicherung: Dateien oder Daten

Data Warehouses ermöglichen es Unternehmen, Informationen aus verschiedenen Systemen zu sammeln und zu kombinieren und so ein einziges Zentrum für weitere Analysen zu schaffen. Die gesammelten historischen Daten ermöglichen nicht nur eine tiefere Analyse von Prozessen, sondern auch die Identifizierung von Mustern, die die Unternehmensleistung beeinflussen können.

Nehmen wir an, ein Unternehmen arbeitet an mehreren Projekten gleichzeitig. Ein Ingenieur möchte wissen, wie viel Beton bereits gegossen wurde und wie viel noch gekauft werden muss. Bei einem herkömmlichen Ansatz müsste er manuell den Server durchsuchen und mehrere Tabellen mit Kostenvoranschlägen öffnen, sie mit den Bescheinigungen über abgeschlossene Arbeiten vergleichen und die aktuellen Lagerbestände überprüfen. Dies dauert Stunden oder sogar Tage. Auch mit ETL-Prozessen und automatischen Skripten bleibt die Aufgabe halb-manuell: Der Ingenieur muss immer noch manuell den Pfad zu Ordnern oder bestimmten Dateien auf dem Server angeben. Dies schmälert den Gesamteffekt der Automatisierung, da sie weiterhin wertvolle Arbeitszeit in Anspruch nimmt.

Bei der Umstellung auf Datenmanagement arbeitet der Techniker nicht mehr mit dem Dateisystem des Servers, sondern erhält Zugang zu einer einheitlichen Speicherstruktur, in der die Informationen in Echtzeit aktualisiert werden. Eine einzige Abfrage - in Form eines Codes, einer SQL -Abfrage oder sogar eines LLM -Agentenaufrufs - kann sofort genaue Daten über aktuelle Salden, geleistete Arbeit und bevorstehende Lieferungen liefern, wenn die Daten im Voraus aufbereitet und in einem Data Warehouse zusammengefasst wurden, in dem man nicht durch Ordner wandern, Dutzende von Dateien öffnen und Werte manuell vergleichen muss.

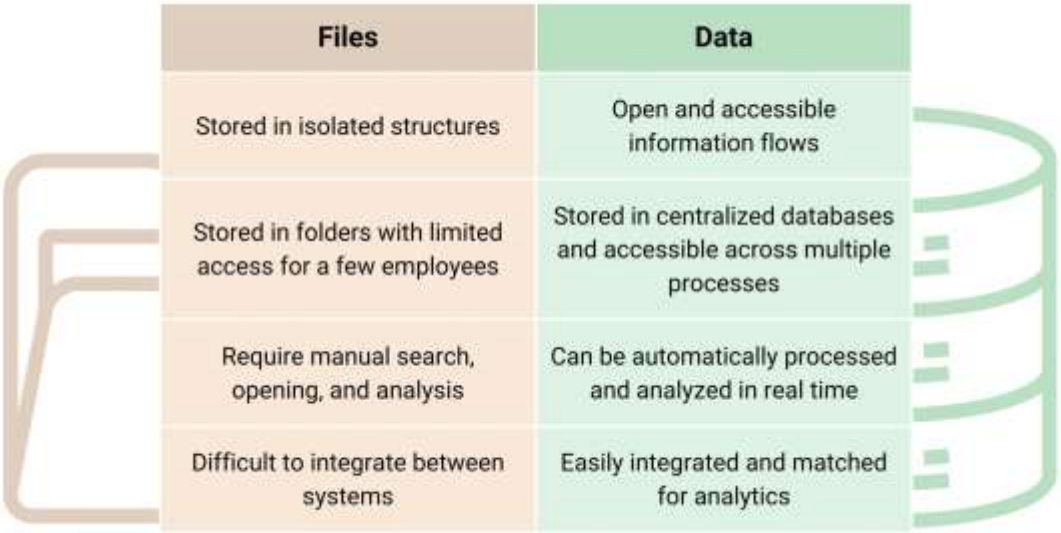
Lange Zeit verwendeten Bauunternehmen PDF -Dokumente, DWG -Zeichnungen, RVT -Modelle und Hunderte und Tausende von Excel -Tabellen und andere unterschiedliche Formate, die in bestimmten Ordnern auf den Servern des Unternehmens gespeichert sind, was die Suche nach Informationen, deren Überprüfung und Analyse erschwert. Infolgedessen werden die Dateien, die nach Abschluss von Projekten übrig bleiben, meist zurück auf den Server in Archivierungsordner verschoben, die in Zukunft praktisch nicht mehr genutzt werden. Eine solche traditionelle dateibasierte Datenspeicherung verliert mit zunehmendem Datenfluss an Bedeutung, da sie anfällig für menschliche Fehler ist.

> **Eine Datei** ist lediglich ein isolierter Behälter, in dem Daten gespeichert werden. Dateien werden für Menschen und nicht für Systeme erstellt, so dass sie manuell geöffnet, gelesen und interpretiert werden müssen. Beispiele dafür sind eine Excel -Tabelle, ein PDF -Dokument oder eine CAD -Zeichnung, die speziell in einem bestimmten Werkzeug geöffnet werden muss, um die gewünschten Informationen abzurufen. Ohne strukturiertes Abrufen und Verarbeiten bleiben die darin enthaltenen Informationen ungenutzt.
>
> **Daten** wiederum sind maschinenlesbare Informationen, die automatisch verknüpft, aktualisiert und analysiert werden. In einem einzelnen Data Warehouse (z. B. Datenbank, DWH oder Data Lake) werden Informationen in Form von Tabellen, Datensätzen und Beziehungen dargestellt. Dies ermöglicht eine einheitliche Speicherung, automatisierte Abfragen, Wertanalysen und Echtzeitberichte.

Die Verwendung von Daten anstelle von Dateien (Abb. 8.1-1) eliminiert den manuellen Suchprozess und vereinheitlicht die Verarbeitung. Unternehmen, die einen solchen Ansatz bereits umgesetzt haben, erlangen einen Wettbewerbsvorteil durch den schnellen Zugriff auf Informationen und die Fähigkeit, diese schnell in Geschäftsprozesse zu integrieren.

> Der Übergang von der Verwendung von Dateien zu Daten ist ein unvermeidlicher Wandel, der die Zukunft der Bauindustrie prägen wird.

Jedes Unternehmen in der Baubranche wird vor einer wichtigen Entscheidung stehen: Entweder es speichert die Informationen weiterhin in verstreuten Dateien und Silos, die von Menschen mit Hilfe von Spezialprogrammen gelesen werden müssen, oder es wandelt sie in den ersten Phasen der Verarbeitung in strukturierte Daten um und schafft so eine einzige integrierte digitale Grundlage für ein automatisiertes Projektmanagement.



Files	Data
Stored in isolated structures	Open and accessible information flows
Stored in folders with limited access for a few employees	Stored in centralized databases and accessible across multiple processes
Require manual search, opening, and analysis	Can be automatically processed and analyzed in real time
Difficult to integrate between systems	Easily integrated and matched for analytics

Abbildung 8.1-1 Entwicklung des Informationsflusses: von isolierten Dateien zu integrierten Daten.

Angesichts der explosionsartigen Zunahme von Informationen werden die herkömmlichen Methoden zur Speicherung und Verarbeitung von Dateien immer weniger effizient. Im Baugewerbe, wie auch in anderen Sektoren, reicht es nicht mehr aus, sich auf disparate Aktenordner mit unterschiedlichen Dateiformaten oder unverbundene Datenbanken zu verlassen.

Unternehmen, die im digitalen Zeitalter wettbewerbsfähig bleiben wollen, werden sich unweigerlich auf integrierte digitale Plattformen verlegen, Big-Data-Technologien und automatisierte Analysesysteme nutzen.

Der Übergang von der dateibasierten Speicherung zu datengesteuerten Arbeitsabläufen erfordert ein Umdenken beim Informationsmanagement und eine bewusste Auswahl von Formaten, die sich für die weitere Integration in zentralisierte Repositories eignen. Diese Wahl wird bestimmen, wie effizient Daten verarbeitet werden können, wie schnell auf sie zugegriffen werden kann und wie einfach sie in die digitalen Prozesse eines Unternehmens integriert werden können.

Speicherung großer Datenmengen: Analyse der gängigen Formate und ihrer Effektivität

Speicherformate spielen eine Schlüsselrolle für die Skalierbarkeit, Zuverlässigkeit und Leistung der Analyseinfrastruktur. Für die Analyse und Verarbeitung von Daten - wie Filterung, Gruppierung und Aggregation - verwendeten unsere Beispiele Pandas DataFrame - eine beliebte Struktur für die Arbeit mit Daten im RAM.

Pandas DataFrame verfügt jedoch nicht über ein eigenes Speicherformat, so dass die Daten nach Abschluss der Verarbeitung in eines der externen Formate exportiert werden - meist CSV oder XLSX. Diese Tabellenformate sind einfach auszutauschen und mit den meisten externen Systemen kompatibel, haben aber eine Reihe von Einschränkungen: geringe Speichereffizienz, fehlende Komprimierung und schlechte Unterstützung für die Versionierung:

- **CSV** (Comma-Separated Values): ein einfaches Textformat, das von verschiedenen Plattformen

und Tools unterstützt wird. Es ist einfach zu verwenden, unterstützt aber keine komplexen Datentypen und keine Komprimierung.

- **XLSX** (Excel Open XML Spreadsheet): ein Microsoft Excel-Dateiformat, das anspruchsvolle Funktionen wie Formeln, Diagramme und Styling unterstützt. Es ist zwar für die manuelle Datenanalyse und -visualisierung nützlich, aber nicht für die Verarbeitung umfangreicher Daten optimiert.

Neben den beliebten Tabellenformaten XLSX und CSV gibt es mehrere gängige Formate für die effiziente Speicherung strukturierter Daten (Abb. 8.1-2), die je nach den spezifischen Anforderungen an die Datenspeicherung und -analyse einzigartige Vorteile bieten:

- **Apache Parquet**: ein spaltenförmiges Dateiformat zur Datenspeicherung, das für die Verwendung in Datenanalysesystemen optimiert ist. Es bietet effiziente Datenkompressions- und Kodierungsschemata und ist daher ideal für komplexe Datenstrukturen und die Verarbeitung großer Datenmengen.
- **Apache ORC** (Optimised Row Columnar): Ähnlich wie Parquet bietet ORC eine hohe Komprimierung und effiziente Datenspeicherung. Es ist für umfangreiche Lesevorgänge optimiert und eignet sich gut für die Speicherung von Data Lakes.
- **JSON** (JavaScript Object Notation): Obwohl JSON im Vergleich zu binären Formaten wie Parquet oder ORC nicht so effizient in der Datenspeicherung ist, ist es sehr leicht zugänglich und einfach zu verarbeiten, was es ideal für Skripte macht, bei denen Lesbarkeit und Webkompatibilität wichtig sind.
- **Feather**: ein schnelles, leichtgewichtiges und einfach zu verwendendes, analytisch orientiertes, binäres, spaltenförmiges Datenspeicherformat. Es wurde entwickelt, um Daten effizient zwischen Python (Pandas) und R zu übertragen, was es zu einer ausgezeichneten Wahl für Projekte macht, an denen diese Programmierumgebungen beteiligt sind.
- **HDF5** (Hierarchical Data Format Version 5): Entwickelt für die Speicherung und Organisation großer Datenmengen. Es unterstützt eine breite Palette von Datentypen und ist gut geeignet für die Arbeit mit komplexen Datensammlungen. HDF5 ist aufgrund seiner Fähigkeit, große Datenmengen effizient zu speichern und darauf zuzugreifen, besonders in der wissenschaftlichen Datenverarbeitung beliebt.

	XLSX	CSV	Apache Parquet	HDF5	Pandas DataFrame
Storage	Tabular	Tabular	Columnar	Hierarchical	Tabular
Usage	Office tasks, data presentation	Simple data exchange	Big data, analytics	Scientific data, large volumes	Data analysis, manipulation
Compression	Built-in	None	High	Built-in	None (in-memory)
Performance	Low	Medium	High	High	High (memory dependent)
Complexity	High (formatting, styles)	Low	Medium	Medium	Low
Data Type Support	Limited	Very limited	Extended	Extended	Extended
Scalability	Low	Low	High	High	Medium (memory limited)

Abb. 8.1-2 Vergleich von Datenformaten mit den wichtigsten Unterschieden in Bezug auf Speicherung und Verarbeitung.

Um eine vergleichende Analyse der in der Lastphase des ETL-Prozesses verwendeten Formate durchzuführen, wurde eine Tabelle mit den Dateigrößen und Lesezeiten erstellt (Abb. 8.1-3). Für die Studie wurden Dateien mit identischen Daten verwendet: Die Tabelle enthielt 10.000 Zeilen und 10 Spalten, die mit Zufallswerten gefüllt waren.

Die folgenden Speicherformate sind in der Studie enthalten: CSV, Parquet, XLSX und HDF5 sowie ihre komprimierten Versionen in ZIP-Archiven. Die Rohdaten wurden mit der NumPy-Bibliothek erzeugt und als Pandas DataFrame-Struktur dargestellt. Der Testprozess umfasste die folgenden Schritte:

- Dateispeicherung: Der Datenrahmen wird in vier verschiedenen Formaten gespeichert: CSV, Parquet, XLSX, und HDF5. Jedes Format hat einzigartige Merkmale in der Art und Weise, wie es Daten speichert, was sich auf die Dateigröße und die Lesegeschwindigkeit auswirkt.
- ZIP-Dateikomprimierung: Um die Wirksamkeit der Standardkomprimierung zu analysieren, wurde jede Datei zusätzlich in ein ZIP-Archiv komprimiert.
- Lesen von Dateien (ETL - Laden): Die Lesezeit wurde für jede Datei nach dem Entpacken aus dem ZIP-Archiv gemessen. So lässt sich die Geschwindigkeit des Datenzugriffs nach dem Entpacken aus dem Archiv abschätzen.

Es ist wichtig zu beachten, dass Pandas DataFrame nicht direkt für die Analyse der Größe oder der Lesezeit verwendet wurde, da es kein eigenständiges Speicherformat darstellt. Es diente lediglich als Zwischenstruktur für die Erzeugung und anschließende Speicherung von Daten in verschiedenen Formaten.

	Original Size (MB)	ZIP Size (MB)	Read Time (s)
CSV	18.82	8.73	0.48
Parquet	9.67	9.22	0.17
XLSX	12.59	12.59	41.73
HDF5	8.50	7.62	0.13

Abb. 8.1-3 Vergleich der Speicherformate nach Größe und Lesegeschwindigkeit.

CSV und HDF5 Dateien weisen (Abb. 8.1-3) eine hohe Komprimierungseffizienz auf, die ihre Größe beim Packen in ZIP erheblich reduziert, was besonders in Szenarien nützlich sein kann, die eine Speicheroptimierung erfordern. XLSX Dateien hingegen sind praktisch nicht komprimierbar und ihre Größe in ZIP bleibt vergleichbar mit der des Originals, was sie für die Verwendung in großen Datenmengen oder in Umgebungen, in denen die Geschwindigkeit des Datenzugriffs wichtig ist, weniger günstig macht. Darüber hinaus ist die Lesezeit für XLSX im Vergleich zu anderen Formaten deutlich höher, so dass es für schnelle Datenlesevorgänge weniger geeignet ist. Apache Parquet hat aufgrund seiner spaltenförmigen Struktur eine hohe Leistung für analytische Aufgaben und große Datenmengen gezeigt.

Optimierung der Speicherung mit Apache Parquet

Eines der beliebtesten Formate für die Speicherung und Verarbeitung von Big Data ist Apache Parquet. Dieses Format wurde speziell für die spaltenorientierte Speicherung entwickelt (ähnlich wie Pandas), wodurch der Speicherbedarf erheblich reduziert und die Geschwindigkeit analytischer Abfragen erhöht werden kann. Im Gegensatz zu herkömmlichen Formaten wie CSV und XLSX unterstützt Parquet eine native Komprimierung und ist für Big-Data-Systeme wie Spark, Hadoop und Cloud-Speicher optimiert.

> Zu den Hauptmerkmalen von Parquet gehört die Unterstützung von Datenkomprimierung und -kodierung, was die Speichergröße erheblich reduziert und das Lesen von Daten beschleunigt, da direkt mit den gewünschten Spalten und nicht mit allen Datenzeilen gearbeitet wird.

Als anschauliches Beispiel dafür, wie einfach es ist, den notwendigen Code zur Konvertierung von Daten in

Apache Parquet zu erhalten, wollen wir die LLM verwenden.

- Senden Sie eine Textanfrage an den LLM Chat (CHATGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

> Schreiben Sie Code, um Daten aus Pandas DataFrame in Apache Parquet zu speichern. ⏎

- Die Antwort von LLM:

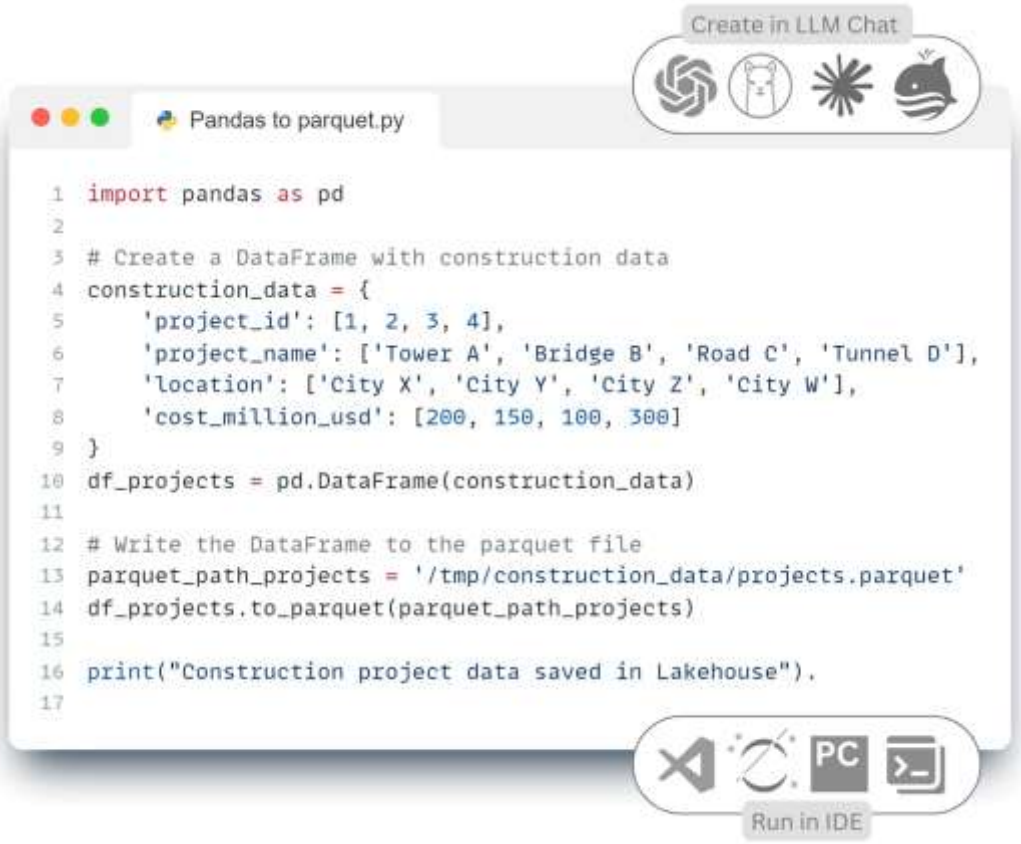


Abbildung 8.1-4 Übertragung von Dataframe-Daten aus dem RAM in ein speichereffizientes Format Apache Parquet mit ein paar Zeilen Python.

Nächstes Beispiel: Simulieren wir den ETL-Prozess mit im Format Parquet gespeicherten Daten, um Projekte nach einem bestimmten Wert eines der Attribute "cost_million_usd" zu filtern (Abb. 8.1-4).

- Während der Chat weitergeht, senden Sie eine Textanfrage an LLM:

> Schreiben Sie einen Code, um die Daten in einer Tabelle zu filtern und nur die Projekte (Tabellenzeilen) aus den Apache Parquet-Daten zu speichern, deren Kosten (Parameter cost_million_usd) 150 Millionen Dollar übersteigen. ⏎

Die Antwort von LLM:

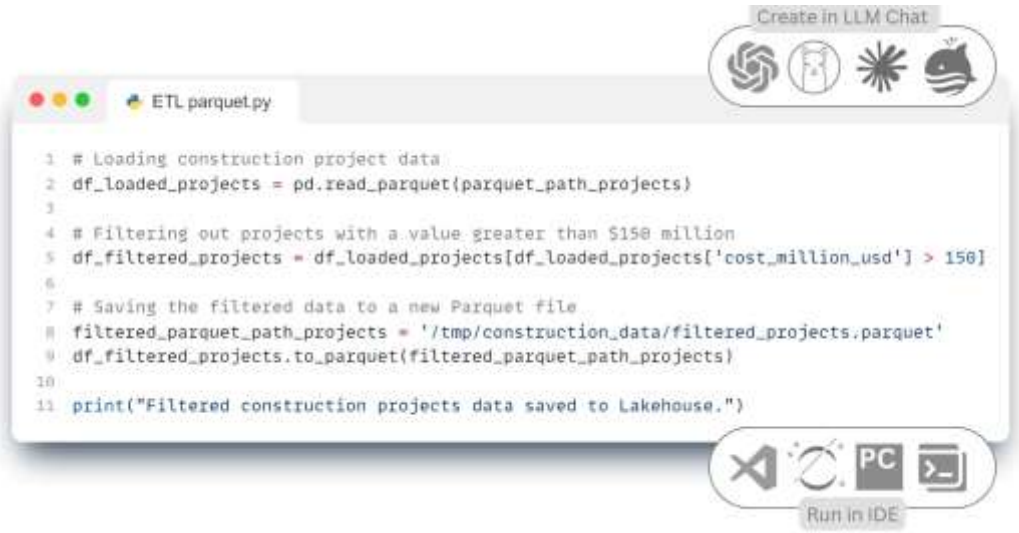


Abbildung 8.1-5 Der ETL-Prozess sieht bei der Arbeit mit Daten im Format Apache Parquet genauso aus wie bei anderen strukturierten Formaten.

Durch die Verwendung des Parquet-Formats (im Gegensatz zu XLSX, CSV usw.) reduziert die Menge der gespeicherten Informationen erheblich und beschleunigt die Suchvorgänge. Damit eignet es sich hervorragend für die Speicherung und Analyse von Daten. Parquet lässt sich in verschiedene Verarbeitungssysteme integrieren und ermöglicht einen effizienten Zugriff in hybriden Architekturen.

Ein effizientes Speicherformat ist jedoch nur ein Element einer umfassenden Datenerfahrung. Um eine nachhaltige und skalierbare Umgebung zu schaffen, ist eine gut durchdachte Datenverwaltungsarchitektur erforderlich. Systeme der Klasse DWH (Data Warehouse) erfüllen diese Funktion. Sie ermöglichen die Aggregation von Daten aus heterogenen Quellen, die Transparenz von Geschäftsprozessen und die Möglichkeit komplexer Analysen mit BI-Tools und maschinellen Lernalgorithmen.

DWH: Data Warehouse data warehouses

So wie das Format Parquet für die effiziente Speicherung großer Informationsmengen optimiert ist, ist das Data Warehouse für die Integration und Strukturierung von Daten optimiert, um Analysen, Prognosen und Managemententscheidungen zu unterstützen.

In den Unternehmen von heute stammen die Daten aus vielen verschiedenen Quellen: ERP, CAFM, CPM, CRM-Systeme, Buchhaltung und Lagerverwaltung, digitale CAD-Modelle von Gebäuden, IoT-Sensoren und andere Lösungen. Um ein ganzheitliches Bild zu erhalten, reicht es nicht aus, Daten einfach nur zu sammeln - sie müssen organisiert, standardisiert und in einem einzigen Repository zentralisiert werden. Genau das

leistet DWH - ein zentrales Speichersystem, mit dem Sie Informationen aus verschiedenen Quellen zusammenführen, strukturieren und für Analysen und strategisches Management verfügbar machen können.

> **DWH (Data Warehouse**) ist ein zentrales Data-Warehouse-System, das Informationen aus verschiedenen Quellen zusammenfasst, strukturiert und für Analysen und Berichte zur Verfügung stellt.

In vielen Unternehmen sind die Daten über verschiedene Systeme verstreut, die wir bereits in den ersten Teilen des Buches besprochen haben (Abb. 1.2-4). Ein DWH integriert diese Quellen und gewährleistet vollständige Transparenz und Zuverlässigkeit der Informationen. Ein DWH Data Warehouse ist eine spezialisierte Datenbank (eine große Datenbank), die Daten aus verschiedenen Quellen sammelt, verarbeitet und speichert. Die Hauptmerkmale eines DWH sind:

- **Verwendung von ETL -Prozessen** (Extract, Transform, Load) - Extrahieren von Daten aus Quellen, Bereinigen, Transformieren, Laden in das Repository und Automatisieren dieser Prozesse, die im siebten Teil des Buches behandelt wurden.
- **Datengranularität** - Daten in DWH können sowohl in aggregierter Form (zusammenfassende Berichte) als auch in granularer Form (Rohdaten) gespeichert werden. Ab 2024 sind es die CAD- Anbieter, die von granularen Daten sprechen [125], was vielleicht darauf hindeutet, dass sich die Branche auf den Übergang zu einer spezialisierten Cloud-Speicherung für den Umgang mit digitalen Gebäudemodelldaten vorbereitet.
- **Unterstützende Analytik und prädiktive** - Data Warehouses bilden die Grundlage für BI-Tools, Big Data -Analysen und maschinelles Lernen.

Das DWH dient als Grundlage für Business Intelligence und ermöglicht die Analyse der wichtigsten Leistungsindikatoren, die Vorhersage von Verkäufen, Einkäufen und Kosten sowie automatisierte Berichte und Datenvisualisierung (Abb. 8.1-6).

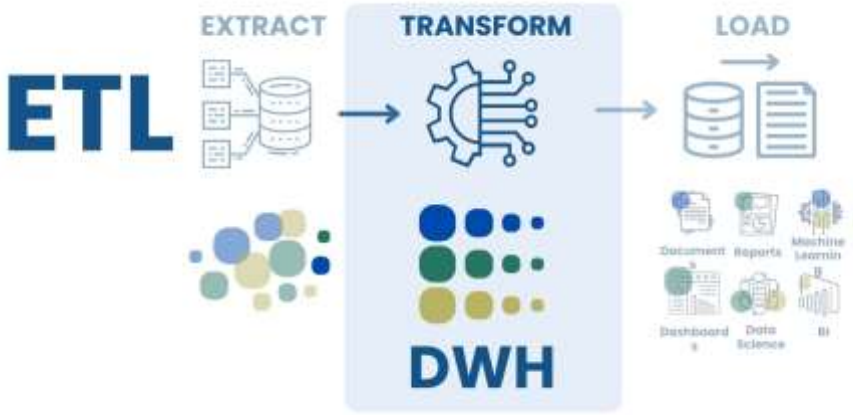


Abb. 8.1-6 In einem ETL -Prozess kann das DWH als zentrales Repository dienen, in dem die aus verschiedenen Systemen extrahierten Daten transformiert und ausgelagert werden.

Das DWH spielt eine Schlüsselrolle bei der Integration, Bereinigung und Strukturierung von Informationen

und bietet eine solide Grundlage für Business Intelligence und Entscheidungsprozesse. In der heutigen Umgebung, in der die Datenmengen schnell wachsen und die Datenquellen immer vielfältiger werden, erfordert der traditionelle DWH-Ansatz zur Informationsspeicherung jedoch häufig eine Erweiterung in Form von ELT und Data Lake

Data Lake - Entwicklung von ETL zu ELT: von der traditionellen Bereinigung zur flexiblen Verarbeitung

Klassische DWH - Data Warehouses, die für die Speicherung strukturierter Daten in einem für analytische Abfragen optimierten Format konzipiert wurden, stießen bei der Handhabung unstrukturierter Daten und der Skalierbarkeit an ihre Grenzen. Als Antwort auf diese Herausforderungen sind Data Lakes) entstanden, die eine flexible Speicherung großer Mengen heterogener Daten ermöglichen.

> **Data Lake** bietet einen alternativen DWH -Ansatz, der die Arbeit mit unstrukturierten, halbstrukturierten und Rohdaten ohne ein vorheriges starres Schema ermöglicht. Diese Speichermethode ist häufig für die Datenverarbeitung in Echtzeit, maschinelles Lernen und erweiterte Analysen relevant. Im Gegensatz zu DWH, bei dem die Daten vor dem Laden strukturiert und aggregiert werden, ermöglicht Data Lake die Speicherung von Informationen in ihrer Rohform und bietet somit Flexibilität und Skalierbarkeit

Die Enttäuschung über herkömmliche Data Warehouses (RDBMS, DWH) und das Interesse an "Big Data" haben zur Entstehung von Data Lakes geführt, bei denen die Daten statt in ein komplexes ETL einfach in ein locker strukturiertes Repository hochgeladen werden und die Verarbeitung in der Analysephase erfolgt:

- In herkömmlichen Data Warehouses werden die Daten in der Regel vorverarbeitet, transformiert und bereinigt (ETL - Extract, Transform, Load), bevor sie in das Warehouse geladen werden (Abb. 8.1-6). Dies bedeutet, dass die Daten strukturiert und für spezifische zukünftige Analysen und Berichtsaufgaben optimiert werden. Der Schwerpunkt liegt dabei auf der Aufrechterhaltung einer hohen Abfrageleistung und Datenintegrität. Dieser Ansatz kann jedoch kostspielig und weniger flexibel sein, was die Integration neuer Datentypen und sich schnell ändernder Datenschemata angeht.
- Data Lakes hingegen sind für die Speicherung großer Mengen von Rohdaten in ihrem ursprünglichen Format konzipiert (Abb. 8.1-7). Das ETL-Verfahren (Extract, Transform, Load), wird durch das ELT-Verfahren (Extract, Load, Transform) ersetzt, bei dem die Daten zunächst "wie sie sind" in das Warehouse geladen werden und erst dann nach Bedarf umgewandelt und analysiert werden können. Dies bietet mehr Flexibilität und die Möglichkeit, heterogene Daten zu speichern, einschließlich unstrukturierter Daten wie Text, Bilder und Protokolle.

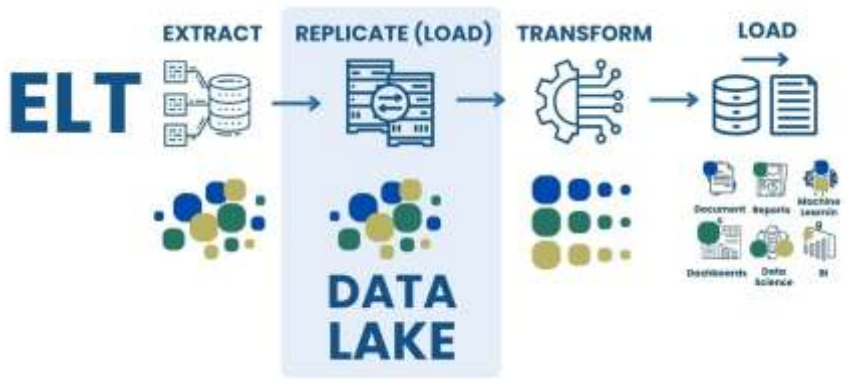


Abb. 8.1-7 Im Gegensatz zu ETL verwendet Data Lake ELT, bei dem die Informationen zunächst in "roher" Form hochgeladen werden und die Transformation in der Upload-Phase erfolgt.

Traditionelle Data Warehouses konzentrieren sich auf die Vorverarbeitung von Daten, um eine hohe Abfrageleistung zu gewährleisten, während Data Lakes die Flexibilität in den Vordergrund stellen: Sie speichern Rohdaten und wandeln sie nach Bedarf um (Abb. 8.1-8).

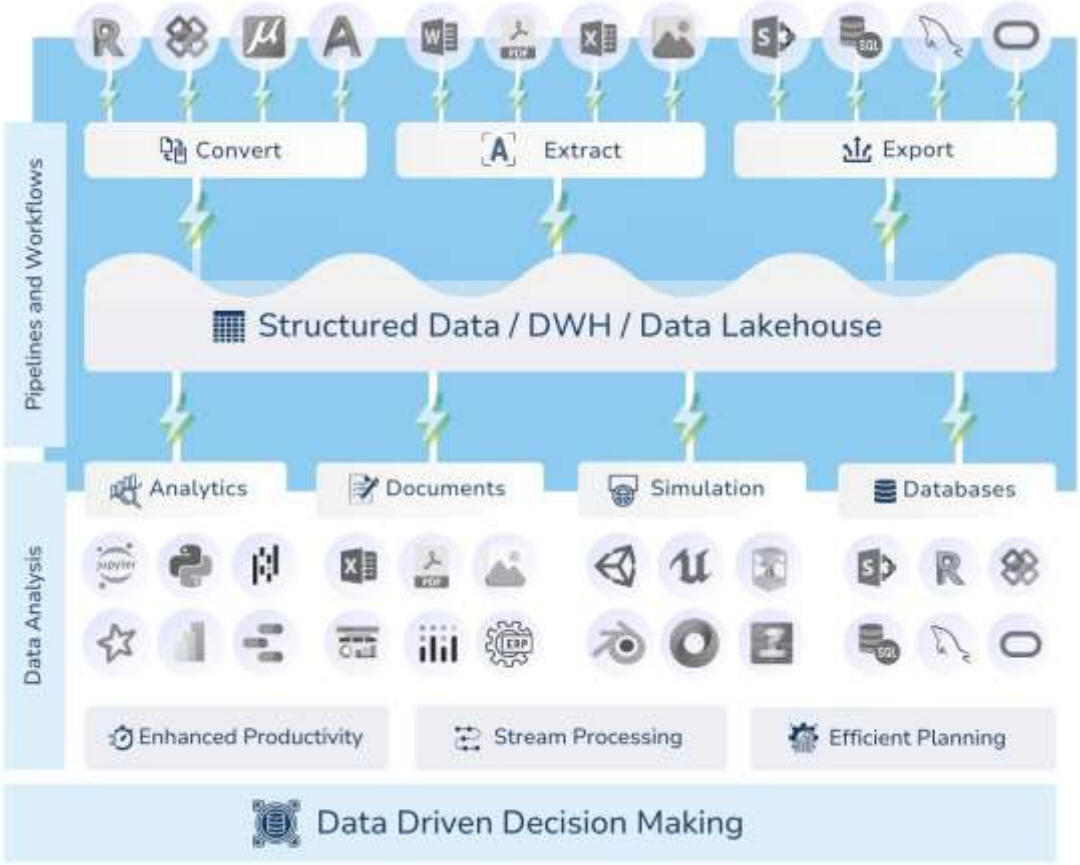


Abb. 8.1-8 Moderne Speicherkonzepte zielen darauf ab, alle Arten von Daten für Entscheidungszwecke zu speichern und zu verarbeiten.

Doch trotz aller Vorteile sind Data Lakes nicht ohne Nachteile. Das Fehlen einer strengen Struktur und die Komplexität der Informationsverwaltung können zu einem Chaos führen, in dem Daten doppelt vorhanden

sind, sich widersprechen oder irrelevant werden. Darüber hinaus ist die Suche und Analyse von Daten in einem solchen Repository mit erheblichem Aufwand verbunden, insbesondere wenn es sich um heterogene Informationen handelt. Um diese Einschränkungen zu überwinden und die besten Eigenschaften von traditionellen Data Warehouses und Data Lakes zu kombinieren, wurde die Data Lakehouse-Architektur entwickelt.

Data Lakehouse Architektur: Synergie von Lagerhäusern und Data Lakes

Um die besten Eigenschaften von DWH (strukturiert, verwaltbar, leistungsstarke Analysen) und Data Lake (Skalierbarkeit, Umgang mit heterogenen Daten) zu kombinieren, wurde der Data Lakehouse-Ansatz entwickelt. Diese Architektur kombiniert die Flexibilität von Data Lakes mit den leistungsstarken Verarbeitungs- und Verwaltungstools, die für herkömmliche Warehouses typisch sind, und schafft ein Gleichgewicht zwischen Speicherung, Analyse und maschinellem Lernen. Data Lakehouse ist eine Synthese aus Data Lakes und Data Warehouses und kombiniert die Flexibilität und Skalierbarkeit der ersteren mit der Verwaltbarkeit und Abfrageoptimierung der letzteren.

> **Data Lakehouse** ist ein architektonischer Ansatz, der versucht, die Flexibilität und Skalierbarkeit von Data Lakes mit der Verwaltbarkeit und Abfrageleistung von Data Warehouses zu kombinieren (Abb. 8.1-9).

Zu den wichtigsten Merkmalen von Data Lakehouse gehören:

- **Offenes Datenspeicherformat:** Die Verwendung offener Formate für die Datenspeicherung, wie Apache Parquet, bietet Effizienz und optimierte Abfragen.
- Schreibgeschütztes **Schema:** Im Gegensatz zum traditionellen Ansatz eines schreibgeschützten Schemas in DWH unterstützt Lakehouse ein schreibgeschütztes Schema, das mehr Flexibilität bei der Verwaltung der Datenstruktur ermöglicht.
- **Flexibel und skalierbar**: Unterstützt die Speicherung und Analyse von strukturierten und unstrukturierten Daten und bietet eine hohe Abfrageleistung durch Optimierung auf Speicherebene.

Data Lakehouse bietet eine Kompromisslösung, die die Vorteile beider Ansätze vereint und damit ideal für moderne Analyse-Workloads ist, die Flexibilität bei der Datenverarbeitung erfordern.

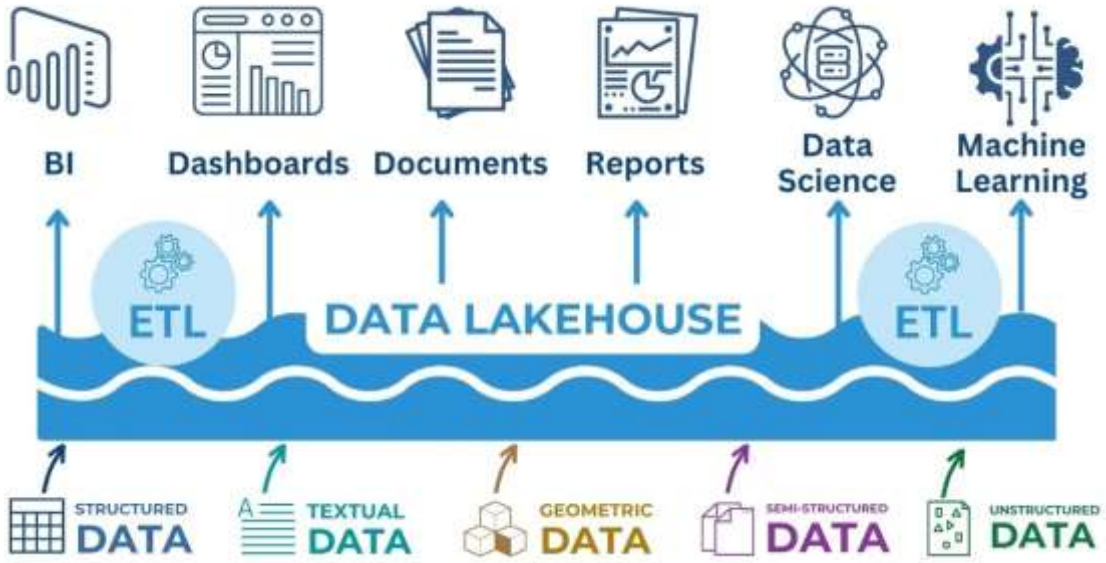


Abb. 8.1-9 Data Lakehouse ist die nächste Generation von Speichersystemen, die komplexe und sich ständig ändernde Anforderungen erfüllen.

Die Idee hinter modernen Data Warehouses scheint einfach: Wenn alle Daten an einem Ort sind, lassen sie sich leichter analysieren. In der Praxis läuft jedoch nicht alles so glatt. Stellen Sie sich vor, ein Unternehmen beschließt, die üblichen Buchhaltungs- und Verwaltungssysteme (ERP, PMIS, CAFM oder andere) vollständig aufzugeben und sie durch einen riesigen Datensee zu ersetzen, zu dem jeder Zugang hat. Was wird geschehen? Höchstwahrscheinlich wird es zu einem Chaos kommen: Daten werden dupliziert, widersprüchlich sein, und wichtige Informationen werden verloren gehen oder beschädigt werden. Selbst wenn der Datensee nur für Analysen verwendet wird, wird er ohne angemessene Verwaltung ernsthaft gefährdet sein:

- Daten sind schwer zu verstehen: In herkömmlichen Systemen haben die Daten eine klare Struktur, aber in einem See sind sie nur eine riesige Ansammlung von Dateien und Tabellen. Um etwas zu finden, muss Spezialist herausfinden, wofür jede Zeile und Spalte zuständig ist.
- Daten können ungenau sein: Wenn viele Versionen derselben Informationen an einem Ort gespeichert sind, ist es schwierig zu wissen, welche Version aktuell ist. Infolgedessen werden Entscheidungen auf der Grundlage veralteter oder fehlerhafter Daten getroffen.
- Es ist schwierig, Daten für die Arbeit aufzubereiten: Die Daten müssen nicht nur gespeichert, sondern auch in einer geeigneten Form präsentiert werden - in Form von Berichten, Grafiken, Tabellen. In traditionellen Systemen geschieht dies automatisch, aber in Data Lakes erfordert es zusätzliche Verarbeitung.

Infolgedessen hat jedes Data-Warehousing-Konzept seine eigenen Merkmale, Verarbeitungsansätze und Geschäftsanwendungen. Herkömmliche Datenbanken konzentrieren sich auf Transaktionsvorgänge, Data Warehouses (DWH) bieten eine Struktur für Analysen, Data Lakes (Data Lake) speichern Informationen in Rohform, und Hybrid Warehouses (Data Lakehouse) kombinieren die Vorteile von DWH und Data Lake (Abb. 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

Abb. 8.1-10 DWH, Data Lake und Data Lake House: wesentliche Unterschiede bei Datentypen, Nutzungsszenarien, Verarbeitungsmethoden und Speicheransätzen.

Die Entscheidung für eine Speicherarchitektur ist ein komplexer Prozess, der von den Geschäftsanforderungen, dem Informationsvolumen und den Analyseanforderungen abhängt. Jede Lösung hat ihre Vor- und Nachteile: DWH sorgt für Struktur, Data Lake für Flexibilität, und Lakehouse bietet ein Gleichgewicht zwischen beiden. Unternehmen sind selten auf eine einzige Datenarchitektur beschränkt.

> Unabhängig von der gewählten Architektur sind automatisierte Datenverwaltungssysteme den manuellen Methoden deutlich überlegen. Sie minimieren menschliche Fehler, beschleunigen die Informationsverarbeitung und gewährleisten Transparenz und Nachvollziehbarkeit der Daten in allen Phasen der Geschäftsprozesse.

Während zentralisierte Data Warehouses in vielen Bereichen der Wirtschaft bereits zum Industriestandard geworden sind, ist die Situation im Baugewerbe nach wie vor fragmentiert. Die Daten sind hier über verschiedene Plattformen (CDE, PMIS, ERP usw.) verteilt, was die Erstellung eines einheitlichen Bildes der Vorgänge erschwert und Architekturen erfordert, die in der Lage sind, diese Quellen zu einer ganzheitlichen, analytisch nutzbaren digitalen Umgebung zu kombinieren.

CDE, PMIS, ERP oder DWH und Data Lake

Einige Bau- und Ingenieurbüros verwenden bereits das Konzept des Common Data Environment (CDE) gemäß ISO 19650. Im Wesentlichen erfüllt das CDE die gleichen Funktionen wie ein Data Warehouse (DWH) in anderen Branchen: Zentralisierung von Informationen, Versionskontrolle, Zugang zu validierten Informationen.

> **Eine gemeinsame Datenumgebung (CDE**) ist ein zentraler digitaler Raum, der für die Verwaltung, Speicherung, gemeinsame Nutzung und Zusammenarbeit von Projektinformationen in allen Phasen des Lebenszyklus einer Einrichtung genutzt wird. CDE wird häufig mit cloudbasierten Technologien implementiert und in CAD-Systeme (BIM) integriert.

Im Finanzsektor, im Einzelhandel, in der Logistik und in der Industrie werden bereits seit Jahrzehnten zentralisierte Datenverwaltungssysteme eingesetzt, die Informationen aus verschiedenen Quellen zusammenführen, ihre Relevanz kontrollieren und Analysen bereitstellen. CDE führt diese Grundsätze weiter, indem es sie an die Herausforderungen der Gebäudeplanung und des Lebenszyklusmanagements anpasst.

Wie DWH strukturiert auch CDE die Daten, erfasst Änderungen und bietet einen zentralen Zugriffspunkt auf verifizierte Informationen. Mit der Verlagerung in die Cloud und der Integration mit Analysetools werden die Unterschiede zwischen den beiden Systemen immer weniger deutlich. Wenn man zu CDE granulare Daten hinzufügt, deren Konzept von CAD -Anbietern seit 2023[93, 125] diskutiert wird, kann man noch mehr Parallelen zum klassischen DWH erkennen.

Im Kapitel "ERP-Systeme im Bauwesen und PMIS " haben wir uns bereits mit PMIS (Project Management Information System) und ERP (Enterprise Resource Planning) befasst. Bei Bauprojekten arbeiten CDE und PMIS zusammen: CDE dient als Ablage für Daten wie Zeichnungen, Modelle und Projektdokumentation, während PMIS Prozesse wie die Kontrolle von Terminen, Aufgaben, Ressourcen und Budgets verwaltet.

ERP, das für die Verwaltung des gesamten Unternehmens (Finanzen, Beschaffung, Personal, Produktion) zuständig ist, kann mit PMIS integriert werden und bietet Kosten- und Budgetkontrolle auf Unternehmensebene. Für Analysen und Berichte kann das DWH genutzt werden, um Daten aus CDE, PMIS und ERP zu sammeln, zu strukturieren und zu aggregieren, um finanzielle KPIs (ROI) auszuwerten und Muster zu erkennen. Der Data Lake (DL) wiederum kann DWH ergänzen, indem er rohe und unstrukturierte Daten (z. B. Protokolle, Sensordaten, Bilder) speichert. Diese Daten können verarbeitet und zur weiteren Analyse in das DWH geladen werden.

> So konzentrieren sich CDE und PMIS auf das Projektmanagement, ERP auf die Geschäftsprozesse und DWH und Data Lake auf Analysen und Datenoperationen.

Vergleicht man CDE, PMIS und ERP Systeme mit DWH und Data Lake, so zeigen sich signifikante Unterschiede in Bezug auf Herstellerunabhängigkeit, Kosten, Integrationsflexibilität, Datenunabhängigkeit, Geschwindigkeit der Anpassung an Veränderungen und analytische Fähigkeiten (Abb. 8.1-11). Traditionelle Systeme wie CDE, PMIS und ERP sind oft an bestimmte Anbieterlösungen und Standards gebunden, was sie weniger flexibel macht und ihre Kosten aufgrund von Lizenzen und Support erhöht. Darüber hinaus sind die Daten in solchen Systemen oft in proprietären, geschlossenen Formaten gekapselt, was ihre Nutzung und Analyse einschränkt.

		CDE, PMIS, ERP	DWH, Data Lake
	Vendor Dependency	High (tied to specific solutions and standards of vendors)	Low (flexibility in tool and platform choice)
	Integration Flexibility	Limited (integration depends on vendor solutions)	High (easily integrates with various data sources)
	Cost	High (licensing and support costs)	Relatively lower (use of open technologies and platforms)
	Data Independence	Low (data often locked in proprietary formats)	High (data stored in open and accessible formats)
	Adaptability to Changes	Slow (changes require vendor approval and integration)	Fast (adaptation and data structure modification without intermediaries)
	Analytical Capabilities	Limited (dependent on vendor-provided solutions)	Extensive (support for a wide range of analytical tools)

Abb. 8.1-11 DWH und Data Lake bieten mehr Flexibilität und Datenunabhängigkeit als Systeme wie CDE, PMIS und ERP.

Im Gegensatz dazu bieten DWH und Data Lake mehr Flexibilität bei der Integration mit verschiedenen Datenquellen, und ihre Verwendung offener Technologien und Plattformen trägt zur Senkung der Gesamtbetriebskosten bei. Außerdem unterstützen DWH und Data Lake eine breite Palette von Analysetools, was die Analyse- und Verwaltungsfunktionen verbessert.

> Mit der Entwicklung von Reverse-Engineering-Tools für CAD-Formate und dem Zugang zu CAD-Anwendungsdatenbanken wird die Frage immer akuter: Wie gerechtfertigt ist es, weiterhin geschlossene, isolierte Plattformen zu verwenden, wenn die Konstruktionsdaten einem breiten Spektrum von Spezialisten zur Verfügung stehen müssen, die in Dutzenden von Auftragnehmern und Konstruktionsunternehmen arbeiten?

Diese anbieterspezifische Technologieabhängigkeit kann die Flexibilität der Datenverwaltung erheblich einschränken, die Reaktion auf Projektänderungen verlangsamen und eine effektive Zusammenarbeit zwischen den Beteiligten verhindern.

Traditionelle Ansätze zur Datenverwaltung - einschließlich DWH, Data Lake, CDE und PMIS - haben sich in erster Linie auf die Speicherung, Strukturierung und Verarbeitung von Informationen konzentriert. Mit der Entwicklung der künstlichen Intelligenz und des maschinellen Lernens wächst jedoch der Bedarf an neuen Methoden zur Organisation von Daten, die nicht nur aggregieren, sondern auch komplexe Zusammenhänge

aufzeigen, verborgene Muster finden und sofortigen Zugriff auf die wichtigsten Informationen bieten.

Vektordatenbanken - ein neuer Speichertyp, der für hochdimensionale Einbettungen optimiert ist - beginnen, eine besondere Rolle in dieser Richtung zu spielen.

KAPITEL 8.2.
VERWALTUNG VON DATENLAGERN UND CHAOSVERMEIDUNG

Vektordatenbanken und die Bounding Box

Vektordatenbanken sind eine neue Klasse von Datenspeichern, die nicht nur Daten speichern, sondern auch die Suche nach Bedeutungen, den Vergleich von Objekten anhand ihrer semantischen Nähe und die Schaffung intelligenter Systeme ermöglichen: von Empfehlungen bis hin zur automatischen Analyse und Kontexterstellung. Im Gegensatz zu herkömmlichen Datenbanken, die sich auf exakte Übereinstimmungen konzentrieren, finden Vektordatenbanken ähnliche Objekte auf der Grundlage von Attributen - selbst wenn es keine exakte Übereinstimmung gibt

> **Eine** Vektordatenbank ist ein spezieller Datenbanktyp, der Daten als mehrdimensionale Vektoren speichert, die jeweils bestimmte Merkmale oder Eigenschaften darstellen. Diese Vektoren können eine unterschiedliche Anzahl von Dimensionen haben, je nach Komplexität der Daten (in einem Fall können es einige wenige Dimensionen sein, in einem anderen — Tausende).

Der Hauptvorteil von Vektordatenbanken ist die Suche nach semantischer Relevanz und nicht nach der exakten Übereinstimmung von Werten. Anstelle von SQL- und Pandas -Abfragen mit "gleich"- oder "enthält"-Filtern wird die Suche nach den nächsten Nachbarn (k-NN) im Merkmalsraum verwendet (wir werden im nächsten Teil des Buches mehr über k-NN sprechen).

Mit der Entwicklung von LLM (Large Language Models) und generativen Modellen beginnt sich die Interaktion mit Datenbanken zu verändern. Es ist jetzt möglich, Daten in natürlicher Sprache abzufragen, Dokumente semantisch zu durchsuchen, automatisch Schlüsselbegriffe zu extrahieren und kontextuelle Beziehungen zwischen Objekten herzustellen - und das alles ohne SQL-Kenntnisse oder Kenntnisse der Tabellenstruktur. Dies wurde im Abschnitt "LLMs und ihre Rolle in der Datenverarbeitung und in Geschäftsprozessen" ausführlicher erörtert.

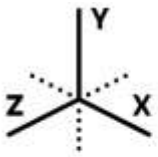
Es ist jedoch wichtig zu wissen, dass LLMs die Informationen nicht automatisch strukturieren und ordnen. Das Modell durchforstet einfach die Daten und findet die relevantesten Daten auf der Grundlage des Kontexts der Abfrage. Wenn die Daten nicht vorbereinigt oder umgewandelt wurden, ist die Tiefensuche wie der Versuch, eine Antwort im digitalen "Müll" zu finden - es kann funktionieren, aber die Qualität der Ergebnisse wird geringer sein. Ideal ist es, wenn die Daten strukturiert werden können (z. B. durch Übersetzung von Dokumenten in Markdown) und in eine Vektordatenbank geladen werden. Dies erhöht die Genauigkeit und Relevanz der Ergebnisse erheblich.

> Ursprünglich wurden Vektordatenbanken im Bereich des maschinellen Lernens eingesetzt, aber heute finden sie immer mehr Anwendungen außerhalb dieses Bereichs - in Suchmaschinen, bei der Personalisierung von Inhalten und bei intelligenten Analysen.

Eines der offensichtlichsten Beispiele für den Vektoransatz in der Konstruktion ist die Bounding Box (Begrenzungsquader). Dabei handelt es sich um eine geometrische Konstruktion, die die Grenzen eines Objekts im dreidimensionalen Raum beschreibt. Die Bounding Box wird durch die minimalen und maximalen X-, Y-

und Z-Koordinaten definiert und bildet eine "Box" um das Objekt. Mit dieser Methode können Sie die Größe und Platzierung eines Elements abschätzen, ohne die gesamte Geometrie analysieren zu müssen.

Jede Bounding Box kann als Vektor in einem mehrdimensionalen Raum dargestellt werden: z.B. [x, y, z, Breite, Höhe, Tiefe] - bereits 6 Dimensionen (Abb. 8.2-1).



Bounding Box

	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Abb. 8.2-1 Bounding Box -Element-Koordinateninformationen und ihre Position im Projektmodell sind analog zu einer Vektordatenbank.

Diese Datendarstellung erleichtert viele Aufgaben, z. B. die Überprüfung von Schnittpunkten zwischen Objekten, die Planung der räumlichen Verteilung von Gebäudeelementen und die Durchführung automatischer Berechnungen. Bounding Box kann als Brücke zwischen komplexen 3D-Modellen und traditionellen Vektordatenbanken dienen, so dass Sie die Vorteile beider Ansätze bei der Architektur- und Ingenieursmodellierung effektiv nutzen können

> Bounding Box ist eine "Vektorisierung der Geometrie", und Embedding (eine Methode zur Umwandlung von etwas Abstraktem) ist eine "Vektorisierung der Bedeutung". Beide Ansätze ermöglichen es, von der manuellen Suche zur intelligenten Suche überzugehen, sei es bei 3D - Objekten in einem Projektmodell oder Konzepten in einem Text.

Die Suche nach Objekten im Projekt (z. B. "Finde alle Fenster mit einer Breite > 1,5 m") ähnelt der Suche nach den nächsten Nachbarn (k-NN) in einer Vektordatenbank, wobei die Kriterien eine "Zone" im Merkmalsraum definieren. (wir werden im nächsten Teil über maschinelles Lernen mehr über die k-NN-Nächste-Nachbarn-Suche sprechen) (Abb. 8.2-2). Wenn wir den Bounding-Box-Attributen zusätzliche Parameter (Material, Gewicht, Produktionszeit) hinzufügen, wird die Tabelle zu einem hochdimensionalen Vektor, bei dem jedes Attribut eine neue Dimension darstellt. Dies ist näher an modernen Vektorbasen, bei denen die Dimensionen in Hunderten oder Tausenden gezählt werden (z. B. Einbettung von neuronalen Netzen).

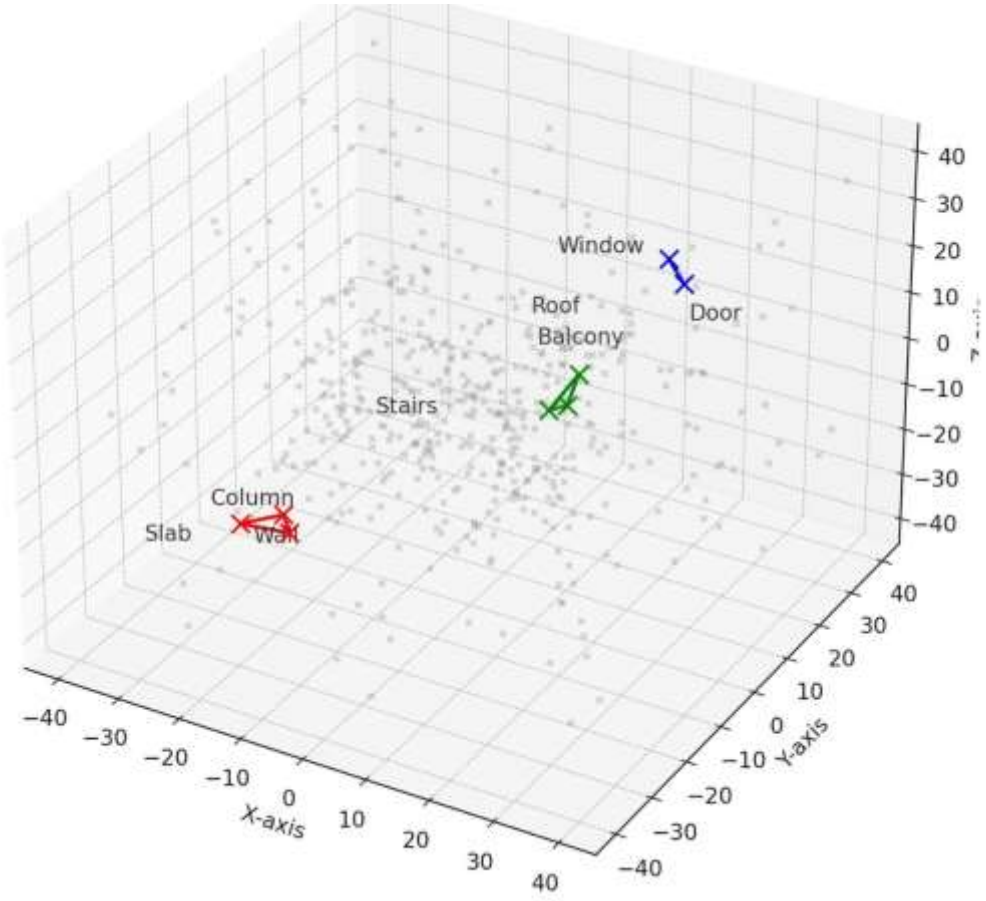


Abb. 8.2-2 Suchen Sie mit Hilfe von Vektordatenbanken nach Objekten im Projekt.

Der in der Bounding Box verwendete Ansatz ist nicht nur auf geometrische Objekte, sondern auch auf die Text- und Sprachanalyse anwendbar. Vektordarstellungen von Daten werden bereits aktiv in der natürlichen Sprachverarbeitung (NLP) eingesetzt. So wie Objekte in einem Bauprojekt nach ihrer räumlichen Nähe gruppiert werden können (Abb. 8.2-2), können Wörter in einem Text nach ihrer semantischen und kontextuellen Nähe analysiert werden.

Zum Beispiel werden die Wörter "Architekt", "Konstruktion" und "Design" im Vektorraum nebeneinander stehen, weil sie eine ähnliche Bedeutung haben. Im LLM ermöglicht dieser Mechanismus eine automatische, nicht manuell vorzunehmende Kategorisierung:

- Das Thema eines Textes identifizieren
- Durchführung semantischer Suchen im Inhalt von Dokumenten
- Automatische Annotationen und Textzusammenfassungen erstellen
- Synonyme und verwandte Begriffe finden

Vektordatenbanken ermöglichen die Analyse von Text und die Suche nach verwandten Begriffen in der gleichen Weise, wie Bounding Box die Analyse von räumlichen Objekten in 3D-Modellen ermöglicht. Das Bounding-Box-Beispiel von Projektelementen hilft zu verstehen, dass die Vektordarstellung kein rein "künst-

liches" Konzept aus dem ML ist, sondern eine natürliche Art der Strukturierung von Daten zur Lösung angewandter Probleme, sei es die Suche nach Spalten in einem CAD-Projekt oder semantisch ähnliche Bilder in einer Datenbank.

> Spezialisten, die mit Datenbanken arbeiten, sollten den Vektorspeichern Aufmerksamkeit schenken. Ihre Verbreitung deutet auf eine neue Phase in der Datenbankentwicklung hin, in der klassische relationale Systeme und KI -orientierte Technologien beginnen, sich zu verflechten und hybride Lösungen der Zukunft zu bilden.

Nutzer, die komplexe und umfangreiche KI -Anwendungen entwickeln, werden spezielle Datenbanken für die Vektorsuche verwenden. Gleichzeitig werden diejenigen, die nur separate KI-Funktionen für die Integration in bestehende Anwendungen benötigen, eher die eingebauten Vektorsuchfunktionen in den von ihnen bereits verwendeten Datenbanken (PostgreSQL, Redis) wählen.

Obwohl Systeme wie DWH, Data Lake, CDE, PMIS, Vektordatenbanken und andere unterschiedliche Ansätze für die Datenspeicherung und -verwaltung bieten, wird ihre Effektivität nicht nur durch ihre Architektur bestimmt, sondern auch dadurch, wie gut die Daten selbst organisiert und verwaltet werden. Selbst bei der Verwendung moderner Lösungen - seien es Vektordatenbanken, klassische relationale DBMS oder Data Lake-ähnliche Warehouses - kann das Fehlen klarer Regeln für die Verwaltung, Strukturierung und Aktualisierung von Daten zu denselben Schwierigkeiten führen, mit denen Benutzer konfrontiert sind, die mit disparaten Dateien und Daten in verschiedenen Formaten arbeiten.

Ohne Data Governance) können selbst die leistungsfähigsten Lösungen chaotisch und unstrukturiert werden, so dass Data Lakes zu Datensümpfen werden). Um dies zu vermeiden, müssen Unternehmen nicht nur die richtige Speicherarchitektur wählen, sondern auch Strategien für Datenminimalismus), Zugriffsmanagement und Qualitätskontrolle implementieren, um Daten in ein effektives Entscheidungsinstrument zu verwandeln.

Data Governance, Data Minimalism und Data Swamp

Das Verständnis und die Umsetzung der Konzepte der Data Governance, des Datenminimalismus und der Vermeidung von Datensümpfen sind der Schlüssel zur erfolgreichen Verwaltung von Data Warehouses und zur Erzielung von Geschäftswert (Abb. 8.2-3).

> Laut einer Studie von Gartner (2017) scheitern 85% der Big-Data-Projekte, und einer der Hauptgründe dafür ist eine unzureichende Datenqualität und Data Governance [144].

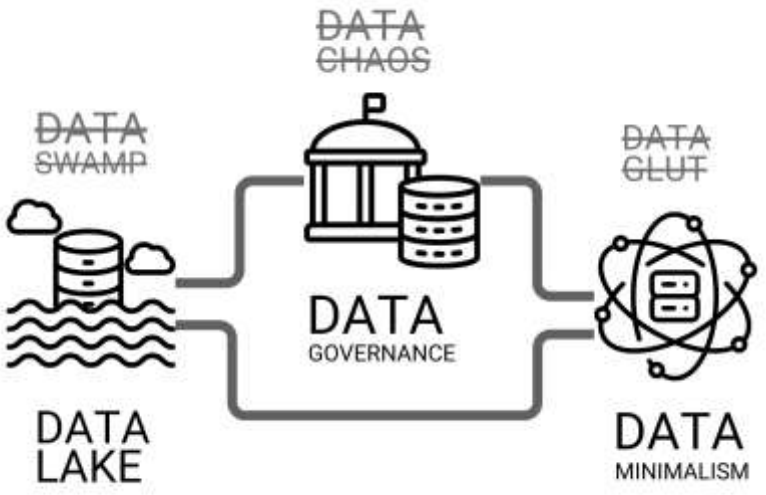


Abb. 8.2-3 Einige der wichtigsten Aspekte der Data Governance sind Data Governance und Data Minimalism.

Data Governance (Data Governance) ist ein grundlegender Bestandteil des Datenmanagements, der sicherstellt, dass Daten in allen Geschäftsprozessen angemessen und effektiv genutzt werden. Dabei geht es nicht nur um die Festlegung von Regeln und Verfahren, sondern auch um die Gewährleistung der Verfügbarkeit, Zuverlässigkeit und Sicherheit von Daten:

- Definition und Klassifizierung von Daten: Eine klare Definition und Klassifizierung von Entitäten ermöglicht es Organisationen zu verstehen, welche Entitäten im Unternehmen benötigt werden, und zu bestimmen, wie sie verwendet werden sollen.
- Zugriffsrechte und Verwaltung: Durch die Entwicklung von Richtlinien und Verfahren für den Zugriff auf Daten und deren Verwaltung wird sichergestellt, dass nur befugte Benutzer auf bestimmte Daten zugreifen können.
- Schutz der Daten vor externen Bedrohungen: Der Schutz der Daten vor externen Bedrohungen ist ein wichtiger Aspekt des Datenmanagements. Dazu gehören nicht nur technische Maßnahmen, sondern auch die Schulung der Mitarbeiter in den Grundlagen der Informationssicherheit.

Datenminimalismus (Data Minimalism) ist ein Ansatz zur Reduzierung von Daten auf die wertvollsten und aussagekräftigsten Attribute und Einheiten in der Formation (Abb. 8.2-4), wodurch die Kosten gesenkt und die Datennutzung verbessert werden:

- Vereinfachung der Entscheidungsfindung: Die Reduzierung der Anzahl der Objekte und ihrer Attribute auf die wichtigsten vereinfacht die Entscheidungsfindung, da weniger Zeit und Ressourcen für die Analyse und Verarbeitung der Daten benötigt werden.
- Konzentration auf das Wesentliche: Durch die Auswahl der relevantesten Entitäten und Attribute können Sie sich auf die Informationen konzentrieren, die für das Unternehmen wirklich wichtig sind, und Rauschen und unnötige Daten eliminieren.
- Effiziente Ressourcenzuweisung: Die Datenminimierung ermöglicht eine effizientere Ressourcenzuweisung, wodurch die Kosten für die Datenspeicherung und -verarbeitung gesenkt und die Datenqualität und -sicherheit verbessert werden.

Die Logik der Arbeit mit Daten sollte nicht mit ihrer Erstellung als solcher beginnen (Abb. 8.2-4), sondern mit

dem Verständnis der zukünftigen Szenarien für die Verwendung dieser Daten, noch bevor der Generierungsprozess beginnt. Dieser Ansatz ermöglicht es, im Voraus die Mindestanforderungen für Attribute, ihre Typen und Grenzwerte zu definieren. Diese Anforderungen bilden die Grundlage für die Erstellung korrekter und stabiler Entitäten im Informationsmodell. Ein vorheriges Verständnis der Zwecke und Verwendungszwecke der Daten trägt zur Bildung einer für die Analyse geeigneten Struktur bei. Weitere Einzelheiten über Ansätze zur Datenmodellierung auf konzeptioneller, logischer und physischer Ebene wurden im Kapitel "Datenmodellierung: konzeptionelles, logisches und physisches Modell" erörtert.

> In den traditionellen Geschäftsprozessen von Bauunternehmen gleicht die Datenverarbeitung eher dem Abladen von Daten in einem Sumpf, in dem Daten zunächst erstellt werden und dann von Spezialisten versucht wird, sie in andere Systeme und Werkzeuge zu integrieren.

Der Datensumpf () ist das Ergebnis einer unkontrollierten Sammlung und Speicherung von Daten ohne angemessene Organisation, Strukturierung und Verwaltung, was zu unstrukturierten, schwer zu nutzenden und wenig wertvollen Daten führt.

Wie lässt sich verhindern, dass der Informationsfluss zu einem Sumpf wird?

- **Verwaltung der Datenstruktur:** Die Strukturierung und Kategorisierung der Daten von trägt dazu bei, eine Datenflut zu verhindern, indem sie geordnet und leicht zugänglich gemacht werden.
- **Verstehen und Interpretieren von** Daten: Eine klare Beschreibung der Datenherkunft, -veränderung und -bedeutung gewährleistet, dass die Daten richtig verstanden und interpretiert werden.
- **Aufrechterhaltung der Datenqualität**: Regelmäßige Datenpflege und -bereinigung tragen dazu bei, die Datenqualität, die Relevanz und den Wert für Analysen und Geschäftsprozesse zu erhalten.

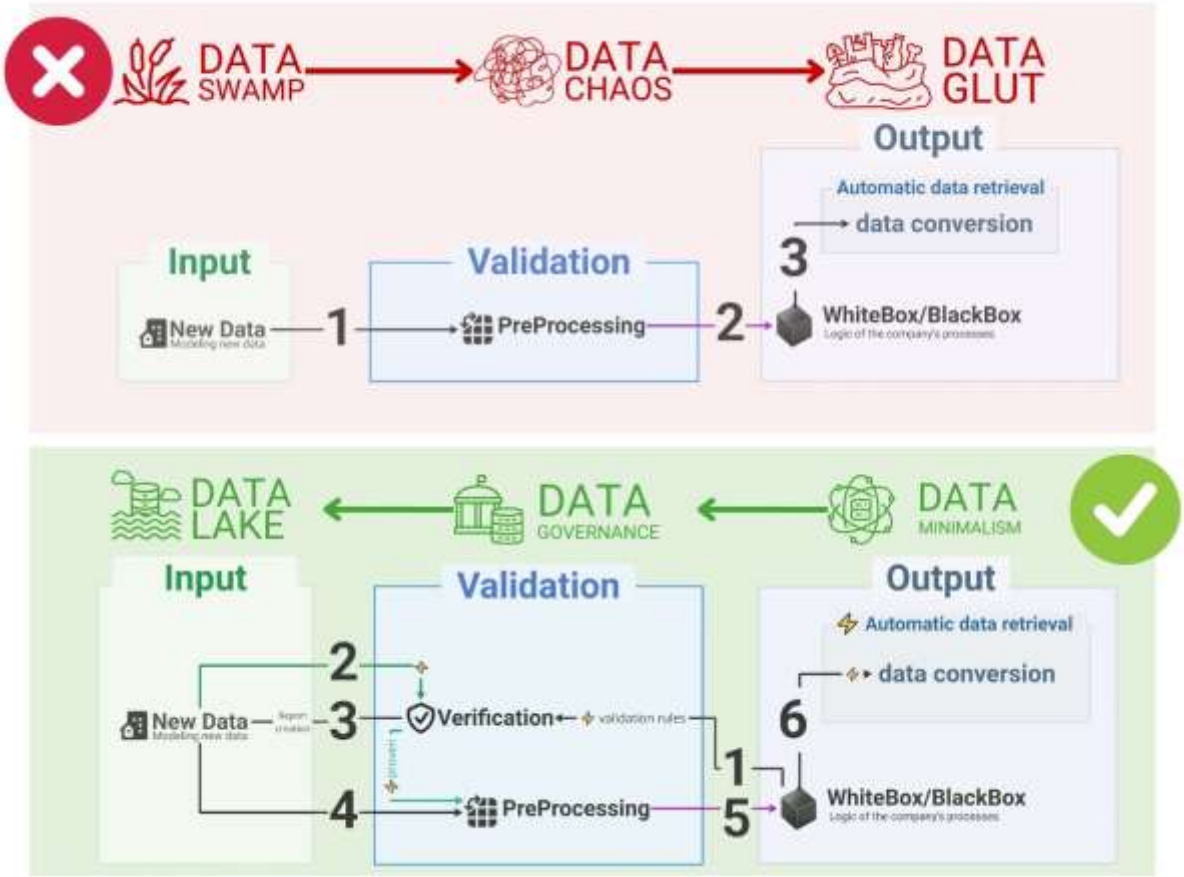


Abbildung 8.2-4 Um Unordnung im Data Warehouse zu vermeiden, sollten Sie den Datenerstellungsprozess mit der Erfassung von Attributanforderungen beginnen.

> Indem sie die Grundsätze der Data Governance und des Datenminimalismus in die Datenverwaltungsprozesse integrieren und aktiv verhindern, dass Data Warehouses zu Datensümpfen werden, können Unternehmen das Potenzial ihrer Daten maximieren.

Die nächste Stufe in der Entwicklung der Arbeit mit Daten ist nach der Lösung der Fragen der Verwaltung und des Minimalismus die Standardisierung der automatischen Verarbeitung, der Qualitätssicherung und der Umsetzung von Methoden, die Daten für die Analyse, Umwandlung und Entscheidungsfindung nutzbar machen. Dies tun die Methoden DataOps und VectorOps, die zu wichtigen Werkzeugen für Unternehmen werden, die mit Big Data und maschinellem Lernen arbeiten.

DataOps und VectorOps: neue Datenstandards

Während Data Governance für die Kontrolle und Organisation von Daten zuständig ist, sorgt DataOps für deren Genauigkeit, Konsistenz und reibungslosen Fluss innerhalb des Unternehmens. Dies ist besonders wichtig für eine Reihe von Geschäftsfällen im Bauwesen, wo Daten kontinuierlich generiert werden und zeitnah verarbeitet werden müssen. In Situationen, in denen beispielsweise Gebäudeinformationsmodelle, Projektanforderungen und analytische Berichte innerhalb eines einzigen Arbeitstages zwischen verschiedenen

Systemen synchronisiert werden müssen, kann DataOps eine Schlüsselrolle spielen. Es ermöglicht Ihnen den Aufbau stabiler und wiederholbarer Datenverarbeitungsprozesse, wodurch das Risiko von Verzögerungen und des Verlusts der Relevanz von Informationen verringert wird.

Data Governance allein reicht nicht aus - es ist wichtig, dass die Daten nicht nur gespeichert, sondern auch aktiv im täglichen Betrieb genutzt werden. Hier kommt DataOps ins Spiel - eine Methodik, die sich auf Automatisierung, Integration und kontinuierlichen Datenfluss konzentriert.

> **DataOps** konzentriert sich auf die Verbesserung der Zusammenarbeit, Integration und Automatisierung von Datenflüssen in Unternehmen. Die Übernahme von DataOps-Praktiken fördert die Datengenauigkeit, -konsistenz und -verfügbarkeit, was für datenzentrierte Anwendungen entscheidend ist.

Die wichtigsten Tools im DataOps-Ökosystem sind Apache Airflow (Abb. 7.4-4) für die Workflow-Orchestrierung und Apache NiFi (Abb. 7.4-5) für das Routing und die Umwandlung von Datenflüssen. Zusammen ermöglichen diese Technologien flexible, zuverlässige und skalierbare Datenpipelines zur automatischen Verarbeitung, Steuerung und Integration von Informationen zwischen Systemen (weitere Einzelheiten im Kapitel "Automatischer ETL -Pipeline "). Bei der Implementierung des DataOps-Ansatzes in Bauprozesse sind vier grundlegende Aspekte zu beachten:

1. **Menschen und Werkzeuge sind wichtiger als Daten**: Siloed data repositories kann als ein großes Problem angesehen werden, aber die Realität ist komplexer. Neben der Datenfragmentierung spielen auch die Isolierung der Teams und die von ihnen verwendeten unterschiedlichen Werkzeuge eine wichtige Rolle. Im Baugewerbe arbeiten Spezialisten aus verschiedenen Disziplinen mit Daten: Dateningenieure und -analysten, BI- und Visualisierungsteams sowie Projektmanagement- und Qualitätsexperten. Jeder von ihnen hat eine andere Arbeitsweise, so dass es wichtig ist, ein Ökosystem zu schaffen, in dem Daten frei zwischen den Beteiligten fließen und eine einzige, konsistente Version der Informationen bereitgestellt wird.
2. **Automatisieren Sie die Prüfung und Fehlererkennung**: Baudaten enthalten immer Fehler, seien es Ungenauigkeiten in Modellen, Berechnungsfehler oder veraltete Spezifikationen. Durch regelmäßiges Testen der Daten und Beseitigung wiederkehrender Fehler kann die Datenqualität erheblich verbessert werden. Als Teil von DataOps müssen Sie automatisierte Kontrollen und Validierungsmechanismen implementieren, die die Korrektheit der Daten überwachen, Fehler analysieren und Muster erkennen sowie Systemfehler in jedem Arbeitsablauf erfassen und beheben. Je höher der Grad der automatisierten Validierung ist, desto höher ist die Datenqualität insgesamt und desto geringer ist die Wahrscheinlichkeit von Fehlern in den letzten Phasen.
3. **Daten sollten auf die gleiche Weise getestet werden wie Programmcode**: Die meisten Bauanwendungen basieren auf der Datenverarbeitung, aber ihre Kontrolle wird oft sekundären Rollen überlassen. Wenn Modelle für maschinelles Lernen auf ungenauen Daten trainiert werden, führt dies zu falschen Vorhersagen und finanziellen Verlusten. Im Rahmen von DataOps sollten Daten der gleichen strengen Prüfung unterzogen werden wie Softwarecode: Logikprüfungen, Stresstests, Bewertung des Verhaltens der Modelle bei veränderten Eingabewerten. Nur validierte und zuverlässige Daten können als Grundlage für Managemententscheidungen verwendet werden.
4. **Datenbeobachtung ohne Leistungseinbußen**: Die Datenüberwachung ist nicht nur eine Sammlung von Metriken, sondern ein strategisches Qualitätsmanagementinstrument. Damit DataOps effektiv arbeiten kann, muss die Beobachtbarkeit in allen Phasen der Datenverarbeitung eingebaut werden, vom Entwurf bis zum Betrieb. Gleichzeitig ist es wichtig, dass die Überwachung das System nicht

ausbremst. Bei Bauprojekten ist es von entscheidender Bedeutung, nicht nur Daten zu sammeln, sondern dies auch so zu tun, dass die Arbeit der Fachleute (z. B. der Planer), die die Daten erstellen, in keiner Weise gestört wird. Dieses Gleichgewicht ermöglicht es Ihnen, die Datenqualität zu kontrollieren, ohne die Produktivität zu beeinträchtigen.

DataOps ist keine zusätzliche Belastung für Datenwissenschaftler, sondern das Rückgrat ihrer Arbeit. Durch die Implementierung von DataOps können Bauunternehmen von einer chaotischen Datenverwaltung zu einem effizienten Ökosystem übergehen, in dem Daten für das Unternehmen arbeiten.

VectorOps wiederum stellt die nächste Stufe in der Entwicklung von DataOps dar und konzentriert sich auf die Verarbeitung, Speicherung und Analyse multidimensionaler Vektordaten (die im vorigen Kapitel behandelt wurden). Dies ist besonders relevant für Bereiche wie digitale Zwillinge, neuronale Netzmodelle und semantische Suche, die allmählich auch in der Baubranche Einzug halten. VectorOps stützt sich auf Vektordatenbanken, um multidimensionale Darstellungen von Objekten effizient zu speichern, zu indizieren und zu durchsuchen.

> **VectorOps** ist der nächste Schritt nach DataOps und konzentriert sich auf die Verarbeitung, Analyse und Verwendung von Vektordaten in der Konstruktion. Im Gegensatz zu DataOps, das sich auf den Datenfluss, die Konsistenz und die Qualität konzentriert, konzentriert sich VectorOps auf die Verwaltung der multidimensionalen Objektdarstellungen, die für maschinelles Lernen erforderlich sind.

Im Gegensatz zu herkömmlichen Ansätzen ermöglicht VectorOps genauere Objektbeschreibungen, was für digitale Zwillinge, generative Designsysteme und die automatische Fehlererkennung bei in das Vektorformat konvertierten CAD-Daten entscheidend ist. Die kombinierte Implementierung von DataOps und VectorOps bildet eine solide Grundlage für die skalierbare, automatisierte Arbeit mit großen Informationsmengen - von klassischen Tabellen bis hin zu semantisch umfangreichen räumlichen Modellen

Die nächsten Schritte: von der chaotischen Speicherung zur strukturierten Speicherung

Herkömmliche Ansätze für den Aufbau von Datenspeichern führen oft zur Schaffung von disparaten "Informationssilos", in denen wichtige Erkenntnisse für die Analyse und Entscheidungsfindung unzugänglich sind. Moderne Speicherkonzepte wie Data Warehouse, Data Lake und deren Mischformen ermöglichen es, disparate Informationen zu vereinheitlichen und sie zentral für Daten-Streaming und Business Intelligence zur Verfügung zu stellen. Es ist nicht nur wichtig, die richtige Speicherarchitektur zu wählen, sondern auch Data Governance) und Data Minimalism) zu implementieren, um zu verhindern, dass Speichereinrichtungen zu unkontrollierbaren Datensümpfen werden).

Um diesen Teil zusammenzufassen, lohnt es sich, die wichtigsten praktischen Schritte hervorzuheben, die Ihnen helfen werden, die besprochenen Konzepte auf Ihre täglichen Aufgaben anzuwenden:

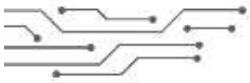
- Auswahl effizienter Datenspeicherformate
  - ☐ Umstellung von CSV und XLSX auf effizientere Formate (Apache Parquet, ORC) zur Speicherung großer Datenmengen
  - ☐ Implementierung eines Systems zur Versionierung von Daten, um Änderungen zu verfolgen

- Verwendung von Metadaten zur Beschreibung der Struktur und Herkunft von Informationen

- Schaffung einer einheitlichen Unternehmensdatenarchitektur
  - Vergleichen Sie verschiedene Speicherarchitekturen: RDBMS, DWH und Data Lake. Wählen Sie diejenige, die Ihre Anforderungen an Skalierbarkeit, Quellenintegration und analytische Verarbeitung am besten erfüllt.
  - Entwerfen Sie eine Prozesslandkarte für das Extrahieren, Laden und Umwandeln von Daten (ETL) aus verschiedenen Quellen für Ihre Aufgaben. Verwenden Sie Visualisierungstools wie Miro, Lucidchart oder Draw.io, um die wichtigsten Schritte und Integrationspunkte zu visualisieren.
- Implementierung von Data-Governance-Praktiken und Datenminimalismus
  - Befolgen Sie den Ansatz des Datenminimalismus - speichern und verarbeiten Sie nur, was wirklich wertvoll ist.
  - Umsetzung der Data-Governance-Grundsätze - Festlegung der Verantwortung für Daten, Gewährleistung von Qualität und Transparenz
  - Erfahren Sie mehr über Datenverwaltungsrichtlinien und DataOps-Konzepte, VectorOps
  - Definition von Datenqualitätskriterien und Verfahren für die Datenvalidierung innerhalb von DataOps

Eine gut organisierte Datenspeicherung schafft die Grundlage für die Zentralisierung der analytischen Prozesse eines Unternehmens. Der Übergang von der chaotischen Anhäufung von Dateien zu strukturierten Speichern ermöglicht es, Informationen in ein strategisches Gut zu verwandeln, das hilft, fundierte Entscheidungen zu treffen und die Effizienz der Geschäftsprozesse zu verbessern.

Sobald die Prozesse der Datenerfassung, -umwandlung, -analyse und strukturierten Speicherung automatisiert und standardisiert sind, ist die nächste Stufe der digitalen Transformation die vollständige Verarbeitung von Big Data.



IX TEIL
BIG DATA, MASCHINELLES LERNEN UND VORHERSAGEN

Der neunte Teil befasst sich mit Big Data, maschinellem Lernen und prädiktiver Analytik in der Baubranche. Er erforscht den Übergang von der intuitiven Entscheidungsfindung zu objektiven Analysen auf der Grundlage historischer Daten. Anhand praktischer Beispiele wird die Big-Data-Analyse im Bauwesen demonstriert - von der Analyse des Datensatzes für Baugenehmigungen in San Francisco bis zur Verarbeitung von CAD - Projekten mit Millionen von Elementen. Besonderes Augenmerk wird auf Methoden des maschinellen Lernens für die Vorhersage von Kosten und Terminen von Bauprojekten gelegt, wobei lineare Regression und k-nearest neighbours Algorithmen ausführlich diskutiert werden. Es wird gezeigt, wie strukturierte Daten die Grundlage für Vorhersagemodelle bilden, um Risiken zu bewerten, Ressourcen zu optimieren und die Effizienz des Projektmanagements zu verbessern. Der Teil enthält auch Empfehlungen zur Auswahl repräsentativer Datenstichproben und erklärt, warum für effektive Analysen nicht immer große Datensätze erforderlich sind.

KAPITEL 9.1.
BIG DATA UND DEREN ANALYSE

Big Data im Bauwesen: von der Intuition zur Vorhersagbarkeit

Für den Begriff "Big Data" gibt es keine eindeutige Definition. Der Begriff tauchte ursprünglich auf, als die Menge der Informationen die Möglichkeiten der traditionellen Methoden zu ihrer Verarbeitung zu übersteigen begann. Heute ist das Volumen und die Komplexität der Daten in vielen Branchen, einschließlich des Bauwesens, so stark gestiegen, dass sie nicht mehr in den lokalen Speicher von Computern passen und den Einsatz neuer Technologien zu ihrer Verarbeitung erfordern.

Das Wesentliche bei der Arbeit mit Big Data ist nicht nur die Speicherung und Verarbeitung, sondern auch die Vorhersagefähigkeit. In der Baubranche öffnet Big Data den Weg von intuitiven Entscheidungen, die auf der subjektiven Interpretation von Tabellen und Visualisierungen beruhen (wie bereits erwähnt), zu fundierten Prognosen, die sich auf reale Beobachtungen und Statistiken stützen.

Entgegen der landläufigen Meinung geht es bei der Arbeit mit Big Data nicht darum, "eine Maschine dazu zu bringen, wie ein Mensch zu denken", sondern darum, mathematische Modelle und Algorithmen anzuwenden, um riesige Datenmengen zu analysieren, damit Muster erkannt, Ereignisse vorhergesagt und Prozesse optimiert werden können.

> **Big Data** ist keine kalte Welt der Algorithmen, die ohne menschlichen Einfluss ist. Im Gegenteil: Big Data arbeitet mit unseren Instinkten, Fehlern und unserer Kreativität zusammen. Es ist die Unvollkommenheit des menschlichen Denkens, die es uns ermöglicht, untypische Lösungen zu finden und Durchbrüche zu erzielen.

Mit der Entwicklung der digitalen Technologie hat die Bauindustrie begonnen, aktiv Datenverarbeitungstechniken zu nutzen, die aus dem IT-Sektor stammen. Dank Werkzeugen wie Pandas und Apache Parquet können strukturierte und unstrukturierte Daten kombiniert werden, was den Zugang zu Informationen vereinfacht und die Verluste bei der Analyse verringert, während große Datensätze aus Dokumenten oder CAD-Projekten (Abb. 9.2-10 - Abb. 9.2-12) es ermöglichen, Daten in allen Phasen des Projektlebenszyklus zu sammeln, zu analysieren und vorherzusagen.

Big Data hat einen transformativen Einfluss auf die Bauindustrie und kann diese auf vielfältige Weise beeinflussen. Die Anwendung von Big-Data-Technologien führt zu Ergebnissen in einer Reihe von Schlüsselbereichen, wie zum Beispiel den folgenden:

- **Analyse des Investitionspotenzials** - Vorhersage der Rentabilität und der Amortisationszeiten von Projekten auf der Grundlage von Daten aus früheren Anlagen.
- **Vorausschauende Wartung** - Erkennung wahrscheinlicher Geräteausfälle, bevor sie tatsächlich auftreten, was die Ausfallzeiten reduziert.
- **Optimierung der Lieferkette** - Vorhersage von Störungen und Verbesserung der logistischen Effizienz.
- **Energieeffizienzanalysen** - Unterstützung bei der Planung von Niedrigenergiegebäuden.

- **Sicherheitsüberwachung** - Einsatz von Sensoren und tragbaren Geräten zur Überwachung der Standortbedingungen.
- **Qualitätskontrolle** - Echtzeitüberwachung der Einhaltung von Prozessstandards.
- **Personalverwaltung** - Leistungsanalyse und Vorausschätzung des Personalbedarfs.

Es gibt kaum einen Bereich im Bauwesen, in dem Datenanalyse und Vorhersagen nicht gefragt sind. Der Hauptvorteil von Vorhersagealgorithmen ist ihre Fähigkeit, selbst zu lernen und sich ständig zu verbessern, wenn sich Daten ansammeln.

> In naher Zukunft wird künstliche Intelligenz Bauherren nicht nur unterstützen, sondern auch wichtige Entscheidungen treffen - von Planungsprozessen bis hin zu Fragen des Gebäudebetriebs.

Mehr darüber, wie Vorhersagen erstellt und Lernmodelle verwendet werden, erfahren Sie im nächsten Teil des Buches, "Maschinelles Lernen und Vorhersagen".

Der Übergang zu einer vollwertigen Arbeit mit Big Data erfordert eine Änderung des eigentlichen Analyseansatzes. Während sich die klassischen Systeme, die wir bisher betrachtet haben, auf Ursache-Wirkungs-Beziehungen konzentrierten, verlagert sich der Schwerpunkt der Big-Data-Analyse auf die Suche nach statistischen Mustern und Korrelationen, die es uns ermöglichen, verborgene Beziehungen zu erkennen und das Verhalten von Objekten auch ohne vollständiges Verständnis aller Faktoren vorherzusagen.

Hinterfragen der Machbarkeit von Big Data: Korrelation, Statistik und Datenstichproben

Traditionell beruhte die Konstruktion auf subjektiven Hypothesen und persönlichen Erfahrungen. Die Ingenieure nahmen - mit einer gewissen Wahrscheinlichkeit - an, wie sich das Material verhalten würde, welchen Belastungen das Bauwerk standhalten würde und wie lange das Projekt dauern würde. Diese Annahmen wurden in der Praxis getestet, oft auf Kosten von Zeit, Ressourcen und zukünftigen Risiken.

> Mit dem Aufkommen von Big Data ändert sich die Herangehensweise dramatisch: Entscheidungen werden nicht mehr auf der Grundlage intuitiver Ahnungen getroffen, sondern als Ergebnis der Analyse umfangreicher Datensätze. Das Bauen wird allmählich von einer Kunst der Intuition zu einer präzisen Wissenschaft der Vorhersage.

Der Übergang zur Idee der Nutzung von Big Data wirft unweigerlich eine wichtige Frage auf: Wie kritisch ist die Datenmenge und wie viele Informationen sind für zuverlässige prädiktive Analysen wirklich erforderlich? Der weit verbreitete Glaube, dass "je mehr Daten, desto höher die Genauigkeit", sich in der Praxis nicht immer als statistisch gültig erweist.

> Bereits 1934 wies der Statistiker Jerzy Neumann nach [145], dass der Schlüssel zur Genauigkeit statistischer Schlussfolgerungen nicht so sehr in der Menge der Daten als vielmehr in ihrer Repräsentativität und der Zufälligkeit der Stichproben liegt.

Dies gilt insbesondere für die Baubranche, in der große Datenmengen mit IoT -Sensoren, Scannern, Überwachungskameras, Drohnen und sogar Multiformat-CAD -Modellen erfasst werden, was das Risiko von blinden Flecken, Ausreißern und Datenverzerrungen erhöht.

Nehmen wir ein Beispiel für die Überwachung des Straßenzustands. Ein vollständiger Datensatz aller Straßenabschnitte kann X GB umfassen und die Verarbeitung etwa einen Tag in Anspruch nehmen. Gleichzeitig würde eine Zufallsstichprobe, die nur jeden 50. Straßenabschnitt umfasst, nur X/50 GB umfassen und in einer halben Stunde verarbeitet werden, wobei die Genauigkeit der Schätzungen für bestimmte Berechnungen ähnlich ist (Abb. 9.1-1).

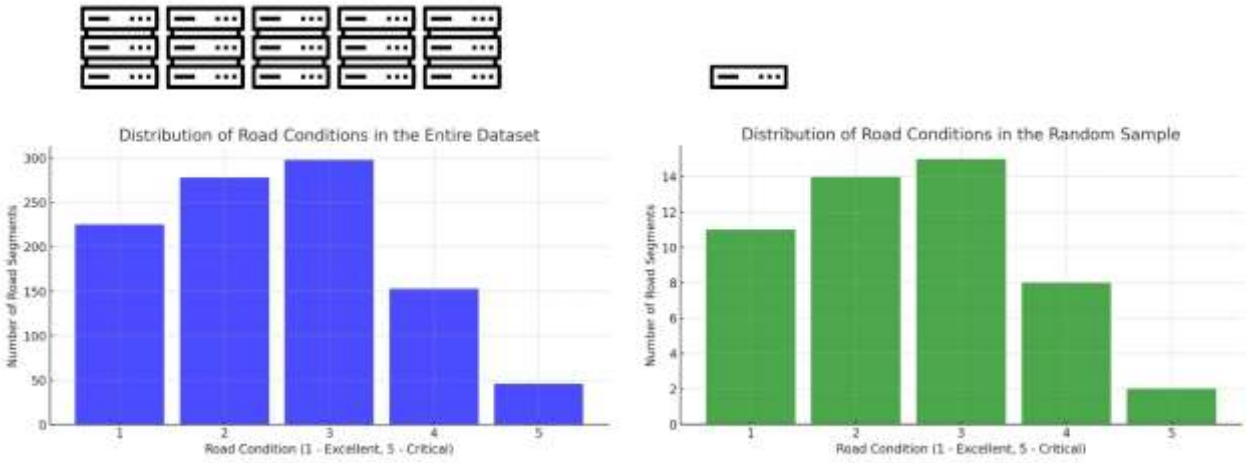


Abbildung 9.1-1 Histogramme des Fahrbahnzustands: Der vollständige Datensatz und die Stichprobe zeigen identische Ergebnisse.

Der Schlüssel zu einer erfolgreichen Datenanalyse liegt also oft nicht in der Menge der Daten, sondern in der Repräsentativität der Stichprobe und der Qualität der verwendeten Verarbeitungsmethoden. Der Übergang zu Stichproben und einem selektiveren Ansatz erfordert ein Umdenken in der Bauindustrie. In der Vergangenheit folgten die Unternehmen der Logik "je mehr Daten, desto besser" und glaubten, dass die Erfassung aller möglichen Indikatoren die Genauigkeit maximieren würde.

Dieser Ansatz erinnert an einen beliebten Irrglauben aus dem Projektmanagement: "Je mehr Spezialisten ich anziehe, desto effektiver wird die Arbeit sein". Wie bei den Humanressourcen sind jedoch die Qualität und die Instrumente wichtiger als die Quantität. Ohne die Wechselbeziehungen (Korrelationen) zwischen Daten oder Projektteilnehmern zu berücksichtigen, kann eine Erhöhung der Menge nur zu Rauschen, Verzerrungen, Doppelarbeit und unnötiger Verschwendung führen.

Letztendlich stellt sich oft heraus, dass es viel produktiver ist, über einen kleineren, aber qualitativ aufbereiteten Datensatz zu verfügen, der stabile und vernünftige Prognosen ermöglicht, als sich auf massive, aber chaotische Informationen mit vielen widersprüchlichen Signalen zu verlassen.

Ein übermäßiges Datenvolumen garantiert nicht nur keine größere Genauigkeit, sondern kann auch zu verzerrten Schlussfolgerungen führen, weil Rauschen, redundante Merkmale, versteckte Korrelationen und irrelevante Informationen vorhanden sind. Unter solchen Umständen steigt das Risiko einer Überanpassung der Modelle und die Zuverlässigkeit der Analyseergebnisse sinkt.

Im Bauwesen besteht eine große Herausforderung im Umgang mit Big Data darin, die optimale Menge und Qualität der Daten zu bestimmen. Bei der Überwachung des Zustands von Betonkonstruktionen beispielsweise kann die Verwendung von Tausenden von Sensoren und die Erfassung von Informationen pro Minute das Speicher- und Analysesystem überfordern. Wenn man jedoch eine Korrelationsanalyse durchführt und die 10% aussagekräftigsten Sensoren auswählt, kann man eine fast identische Vorhersagegenauigkeit erzielen und dabei ein Vielfaches, manchmal ein Zehn- oder Hundertfaches, an Ressourcen einsparen.

Die Verwendung einer kleineren Teilmenge von Daten reduziert sowohl die erforderliche Speichermenge als auch die Verarbeitungszeit, was die Kosten für die Speicherung und Analyse von Daten erheblich senkt und Stichproben oft zu einer idealen Lösung für die prädiktive Analyse macht, insbesondere bei großen Infrastrukturprojekten oder bei Arbeiten in Echtzeit. Letztendlich wird die Effizienz von Bauprozessen nicht durch die Menge der gesammelten Daten bestimmt, sondern durch die Qualität ihrer Analyse. Ohne einen kritischen Ansatz und eine sorgfältige Analyse können die Daten zu falschen Schlussfolgerungen führen.

Nach einer gewissen Datenmenge liefert jede neue Informationseinheit immer weniger nützliche Ergebnisse. Anstatt endlos Informationen zu sammeln, ist es wichtig, sich auf ihre Repräsentativität und die Methoden der Analyse zu konzentrieren (Abb. 9.2-2).

Dieses Phänomen wird von Allen Wallis [146] gut beschrieben, der den Einsatz statistischer Methoden am Beispiel der Prüfung zweier alternativer Projektilentwürfe der US-Marine illustriert.

Die Marine testete zwei alternative Projektilkonstruktionen (A und B), indem sie eine Reihe von gepaarten Schüssen durchführte. In jeder Runde erhält A eine 1 oder 0, je nachdem, ob seine Leistung besser oder schlechter als die von B ist, und umgekehrt. Der statistische Standardansatz beinhaltet die Durchführung einer festen Anzahl von Versuchen (z. B. 1000) und die Ermittlung des Gewinners auf der Grundlage einer prozentualen Verteilung (z. B. wenn A in mehr als 53% der Fälle eine 1 erhält, wird es als das beste Ergebnis angesehen). Als Allen Wallis ein solches Problem mit (Navy) Captain Garrett L. Schuyler diskutierte, wandte dieser ein, dass ein solcher Test, um Allens Geschichte zu zitieren, nutzlos sein könnte. Wäre ein kluger und erfahrener Ordonnanzoffizier wie Schuyler vor Ort gewesen, hätte er nach den ersten paar hundert [Schüssen] gesehen, dass das Experiment nicht abgebrochen werden muss, entweder weil die neue Methode eindeutig minderwertig ist oder weil sie eindeutig besser ist als das, was man sich erhofft hatte [146].

- Statistische Forschungsgruppe der U.S.-Regierung an der Columbia University, während des Zweiten Weltkriegs

Dieses Prinzip ist in verschiedenen Branchen weit verbreitet. In der Medizin werden beispielsweise klinische

Studien zu neuen Medikamenten an Zufallsstichproben von Patienten durchgeführt, wodurch statistisch signifikante Ergebnisse erzielt werden können, ohne dass das Medikament an der gesamten Bevölkerung der Erde getestet werden muss. In Wirtschaft und Soziologie werden repräsentative Umfragen durchgeführt, um die Meinung der Gesellschaft widerzuspiegeln, ohne dass alle Menschen im Land befragt werden müssen.

So wie Regierungen und Forschungsorganisationen kleine Bevölkerungsgruppen befragen, um allgemeine gesellschaftliche Trends zu verstehen, können Unternehmen in der Baubranche Stichproben nutzen, um das Projektmanagement effektiv zu überwachen und Prognosen zu erstellen (Abb. 9.1-1).

> *Big Data kann die Herangehensweise an die Sozialwissenschaft verändern, wird aber den statistischen Hausverstand nicht ersetzen [147].*
>
> - Thomas Landsall-Welfair, "Predicting the nation's current mood", Significance v. 9 (4), 2012

Aus der Perspektive der Ressourceneinsparung ist es bei der Erhebung von Daten für zukünftige Vorhersagen und Entscheidungen wichtig, die Frage zu beantworten: Ist es sinnvoll, erhebliche Ressourcen für die Erhebung und Verarbeitung riesiger Datensätze aufzuwenden, wenn ein viel kleinerer und billigerer Testdatensatz verwendet werden kann, der schrittweise vergrößert werden kann? Die Wirksamkeit von Stichproben zeigt, dass Unternehmen die Kosten für die Erhebung und das Training von Modellen um das Zehn- oder sogar Tausendfache senken können, wenn sie Datenerhebungsmethoden wählen, die keine umfassende Abdeckung erfordern, aber dennoch eine ausreichende Genauigkeit und Repräsentativität bieten. Dieser Ansatz ermöglicht es auch kleinen Unternehmen, mit deutlich weniger Ressourcen und Datenmengen Ergebnisse zu erzielen, die mit denen großer Konzerne vergleichbar sind, was für Unternehmen wichtig ist, die mit geringen Ressourcen Kosten optimieren und fundierte Entscheidungen beschleunigen wollen. In den folgenden Kapiteln werden Beispiele für Analytik und prädiktive Analytik auf der Grundlage öffentlicher Datensätze unter Verwendung von Big-Data-Tools vorgestellt.

Big Data: Analyse von Daten aus dem Datensatz der Millionen Baugenehmigungen in San Francisco

Die Arbeit mit offenen Datensätzen bietet eine einzigartige Gelegenheit, die in den vorangegangenen Kapiteln erörterten Grundsätze in die Praxis umzusetzen: sorgfältige Auswahl von Merkmalen, repräsentative Stichproben, Visualisierung und kritische Analyse. In diesem Kapitel werden wir untersuchen, wie komplexe Phänomene wie die Bautätigkeit in einer Großstadt mithilfe offener Daten untersucht werden können - insbesondere die über eine Million Baugenehmigungsdatensätze in San Francisco

Öffentlich verfügbare Daten zu über einer Million Baugenehmigungen (Abb. 9.1-2) (Datensätze in zwei Datensätzen im CSV-Format) vom "San Francisco Department of Buildings" [148] ermöglichen es uns, die rohe CSV -Tabelle zu verwenden, um nicht nur die Bautätigkeit in der Stadt zu analysieren, sondern auch die jüngsten Trends und die Geschichte der Bauindustrie San Franciscos in den letzten 40 Jahren, von 1980 bis 2019, kritisch zu untersuchen.

Die Code-Beispiele, die zur Erstellung der Datensatz-Visualisierungen verwendet wurden (Abb. 9.1-3- Abbildungen 9.1-8), sowie visuelle Grafiken mit Code, Erklärungen und Kommentaren können auf der Kaggle-

Plattform durch die Suche nach "San Francisco" gefunden werden. Construction Sector 1980-2019" [149].

count 1.137695e+06

Building Permits on or after January 1, 2013 — June 13, 2020 — 2,237 Views

Building Permits before January 1, 2013 — June 13, 2020 — 880 Views

	permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date
	07/01/1998	repair stucco	complete	07/07/1998	07/01/1998	07/01/1998	07/07/1998
	12/13/2004	reroofing	expired	01/24/2006	12/13/2004	12/13/2004	NaN
	02/18/1992	install auto fire spks.	complete	06/29/1992	02/18/1992	03/18/1992	06/29/1992

	permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	94123.0	(37.7963468780498, -122.432264144357 4)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0	(37.7292585180008386, -122.464424566746 2)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0	(37.7950600255297 4, -122.395932244618 05)

Abb. 9.1-2 Die Datensätze enthalten Informationen über erteilte Baugenehmigungen mit verschiedenen Objektattributen.

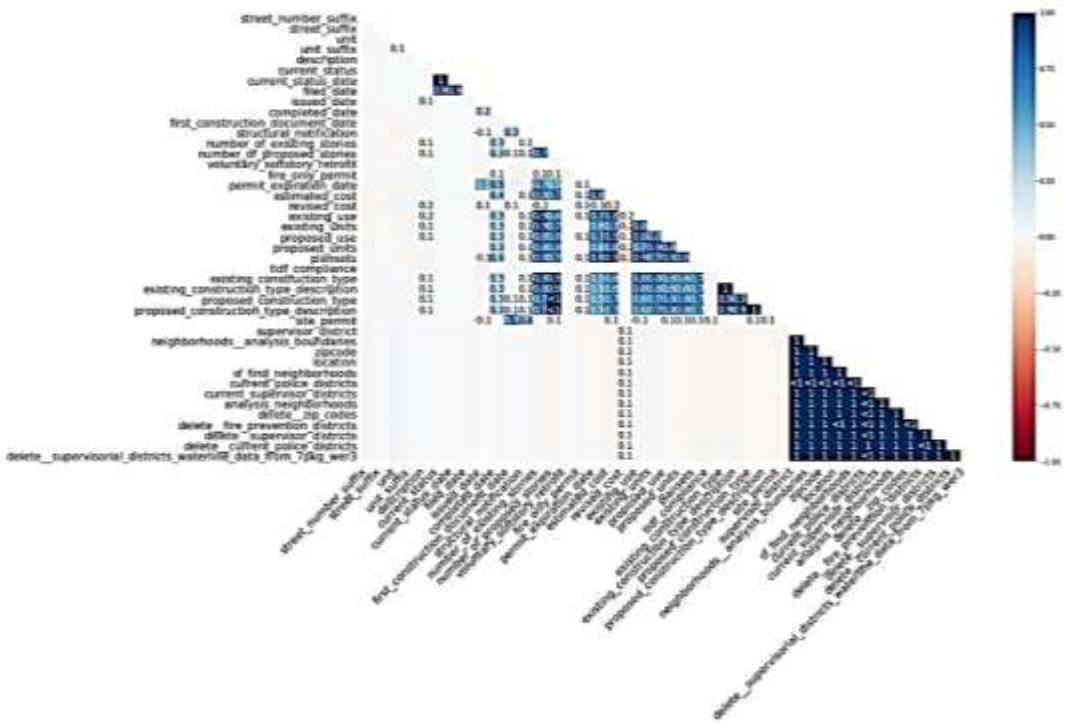


Abb. 9.1-3 Eine Heatmap (Pandas und Seaborn), die alle Attribute eines Datensatzes visualisiert und hilft, Beziehungen zwischen Attributpaaren zu erkennen.

Aus der vom San Francisco Department of Buildings zur Verfügung gestellten Tabelle (Abb. 9.1-2) lassen sich keine Trends oder Schlussfolgerungen ablesen. Trockene Zahlen in tabellarischer Form sind keine Grundlage für die Entscheidungsfindung. Um die Daten visuell verständlich zu machen, sollten sie, wie in

den Kapiteln zur Datenvisualisierung ausführlich besprochen, mit den verschiedenen Bibliotheken visualisiert werden, die im siebten Teil des Buches zum Thema "ETL und Visualisierung der Ergebnisse als Grafiken" besprochen werden.

Durch die Analyse von Daten unter Verwendung von Pandas DataFrame und Python-Visualisierungsbibliotheken über den Wert von 1.137.695 Genehmigungen [148] können wir feststellen, dass die Bautätigkeit in San Francisco eng mit den Wirtschaftszyklen verbunden ist, insbesondere mit der boomenden Technologiebranche im Silicon Valley (Abb. 9.1-4).

Wirtschaftliche Auf- und Abschwünge haben einen erheblichen Einfluss auf die Anzahl und den Wert von Bauprojekten. So fiel beispielsweise der erste Höhepunkt der Bautätigkeit mit dem Elektronikboom Mitte der 1980er Jahre zusammen (verwendet wurden Pandas und Matplotlib), und die nachfolgenden Spitzen und Rückgänge waren mit der Dot-Com-Blase und dem Technologieboom der letzten Jahre verbunden.

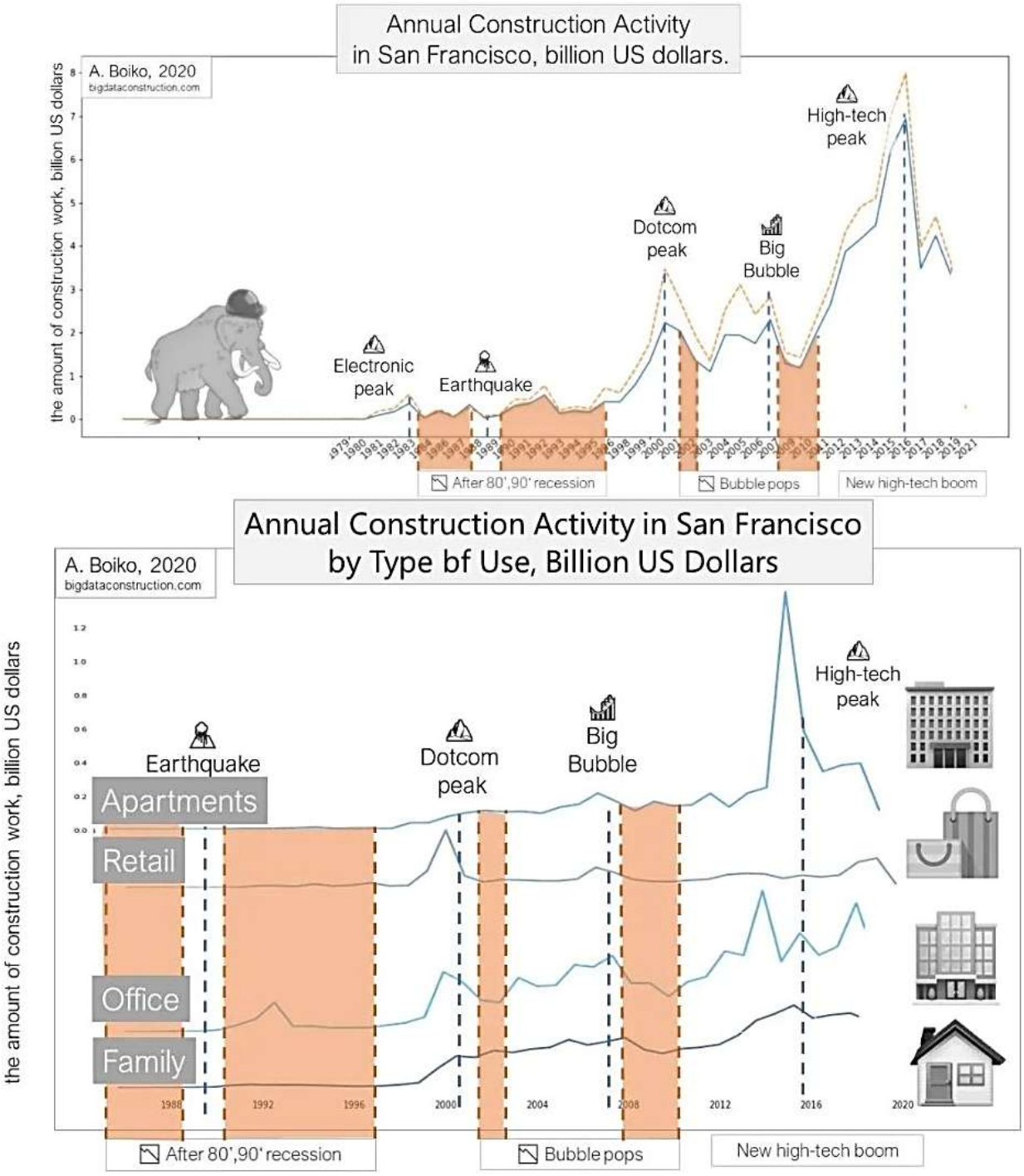


Abb. 9.1-4 Im Immobiliensektor von San Francisco korrelieren die Investitionen mit der technologischen Entwicklung des Silicon Valley.

Datenanalysen deuten darauf hin, dass in San Francisco der Großteil der 91,5 Milliarden Dollar, die in den letzten zehn Jahren in den Bau und die Sanierung von Gebäuden investiert wurden - fast 75% - im Stadtzentrum (Abb. 9.1-5 - unter Verwendung von Pandas und der Visualisierungsbibliothek Folium) und in einem Radius von 2 km um das Stadtzentrum konzentriert ist, was die höhere Investitionsdichte in diesen zentralen Zonen widerspiegelt.

Die durchschnittlichen Kosten für Baugenehmigungen variieren je nach Stadtviertel beträchtlich, wobei Anträge im Stadtzentrum dreimal so viel kosten wie solche außerhalb des Stadtzentrums, was auf die höheren Kosten für Land, Arbeit und Material sowie auf strenge Bauvorschriften zurückzuführen ist, die teurere Materialien zur Verbesserung der Energieeffizienz vorschreiben.

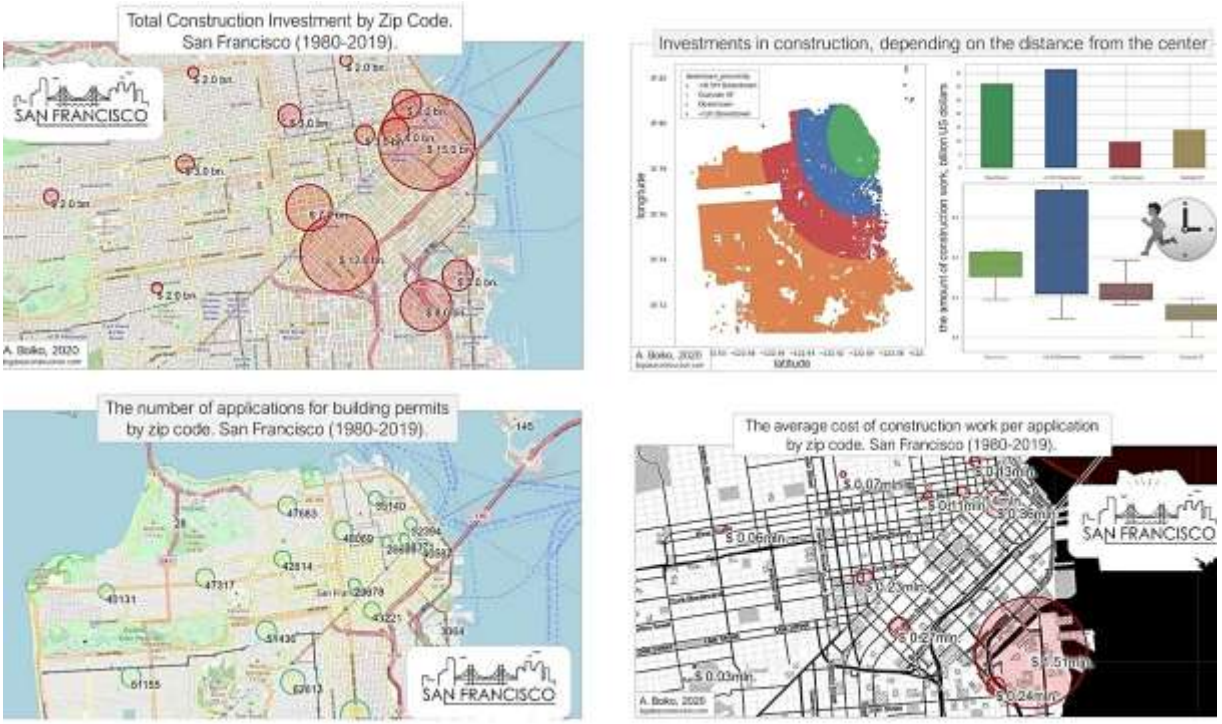


Abbildung 9.1-5 In San Francisco konzentrieren sich 75 Prozent der Bauinvestitionen (91,5 Milliarden Dollar) auf die Innenstadt.

Der Datensatz ermöglicht auch die Berechnung der durchschnittlichen Reparaturpreise nicht nur nach Haustyp, sondern auch nach Stadtvierteln und einzelnen Adressen (Postleitzahlen). In San Francisco zeigt die Dynamik der Renovierungskosten für die verschiedenen Arten von Renovierungen und Wohnungen unterschiedliche Trends (Abb. 9.1-6 - verwendet wurden Pandas und Matplotlib). Küchenrenovierungen sind deutlich teurer als Badrenovierungen: Die durchschnittliche Küchenrenovierung in einem Einfamilienhaus kostet etwa 28.000 $, in einem Zweifamilienhaus dagegen 25.000 $.

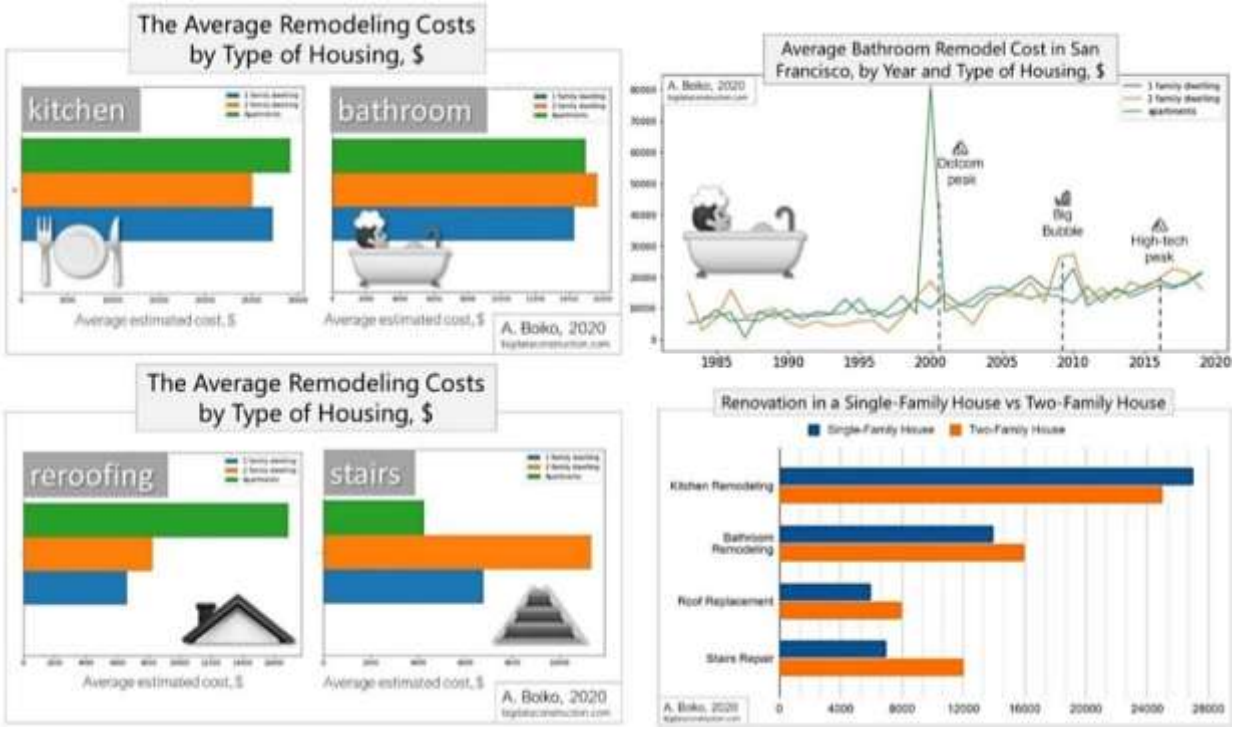


Abbildung 9.1-6 In SF kosten Küchenrenovierungen fast doppelt so viel wie Badrenovierungen, und Hausbesitzer müssen 15 Jahre lang jeden Monat$ 350 zur Seite legen, um die Kosten für größere Reparaturen zu decken.

Die Inflation der Baukosten in San Francisco im Laufe der Jahre lässt sich anhand einer Analyse der nach Wohnungstyp und Jahr gruppierten Daten nachvollziehen (Abb. 9.1-7 - verwendet von Pandas und Seaborn), die einen stetigen Anstieg der durchschnittlichen Reparaturkosten seit 1990 zeigt und kurzfristige Dreijahreszyklen bei den Kosten für die Reparatur von Mehrfamilienhäusern erkennen lässt.

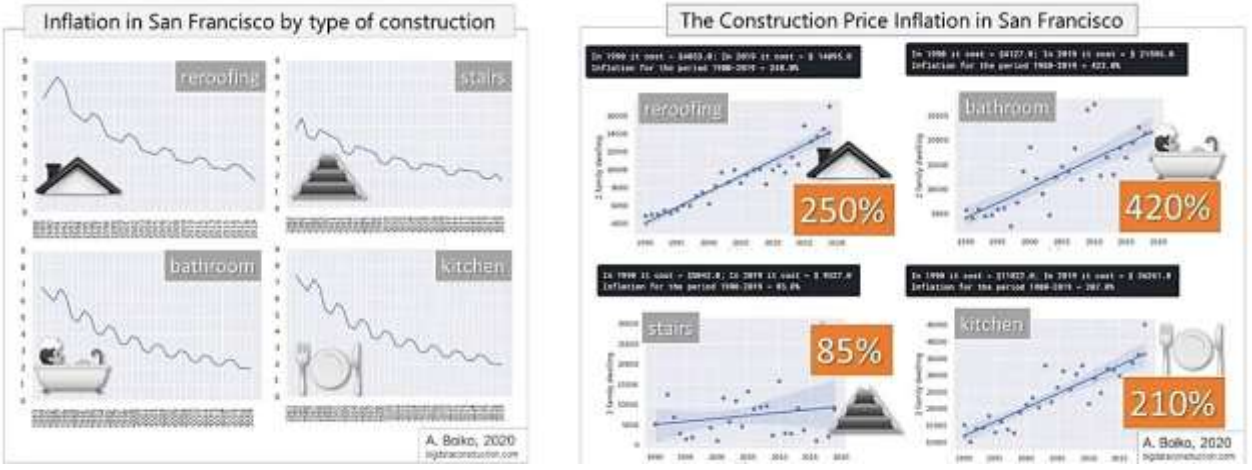


Abbildung 9.1-7 Von 1980 bis 2019 haben sich die Kosten für die Renovierung von Badezimmern in SF verfünffacht, während sich der Preis für die Renovierung von Dächern und Küchen verdreifacht hat und die Renovierung von Treppenhäusern nur um 85% gestiegen ist.

Eine Untersuchung der öffentlichen Daten des Bauamtes von San Francisco (Abb. 9.1-3) zeigt, dass die Baukosten in der Stadt extrem variabel und oft unvorhersehbar sind und von einer Vielzahl von Faktoren beeinflusst werden. Zu diesen Faktoren gehören das Wirtschaftswachstum, technologische Innovationen und die besonderen Anforderungen der verschiedenen Wohnformen.

In der Vergangenheit erforderten solche Analysen tiefgreifende Kenntnisse in der Programmierung und Analytik. Mit dem Aufkommen der LLM- -Tools ist der Prozess jedoch für ein breites Spektrum von Fachleuten in der Bauindustrie zugänglich und verständlich geworden, von Ingenieuren in den Konstruktionsabteilungen bis hin zur Geschäftsleitung.

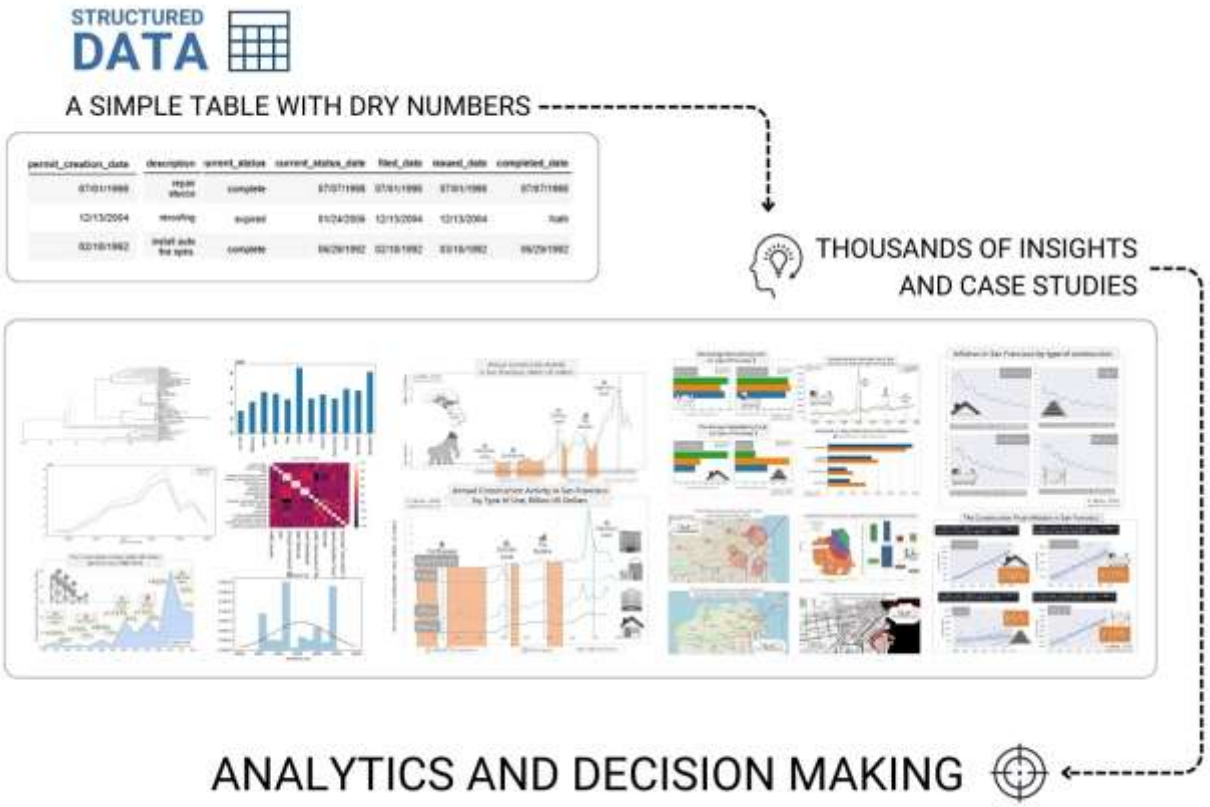


Abb. 9.1-8 Der Übergang zu visuell verständlichen Daten ermöglicht eine automatisierte Entscheidungsfindung durch die Erkennung verborgener Muster.

So wie wir die Daten aus dem Tabellendatensatz der "San Francisco Building Authority" analysiert haben, können wir jeden beliebigen Datensatz visualisieren und analysieren - von Bildern und Dokumenten bis hin zu IoT-Daten oder Daten aus abgeleiteten CAD-Datenbanken.

Beispiel für Big Data auf der Grundlage von CAD-Daten (BIM)

Im folgenden Beispiel wird ein großer Datensatz mit Daten aus verschiedenen CAD-Tools (BIM) analysiert. Zum Sammeln und Erstellen des großen Datensatzes wurde ein spezieller automatischer Web-Crawler (Skript) verwendet, der so konfiguriert wurde, dass er automatisch nach Entwurfsdateien auf Websites

sucht und diese sammelt, die kostenlose Architekturmodelle in den Formaten RVT und IFC anbieten. Innerhalb weniger Tage konnte der Crawler 4.596 IFC-Dateien, 6.471 RVT-Dateien und 156.024 DWG-Dateien finden und herunterladen[149].

Nach dem Sammeln von Projekten in den Formaten RVT und IFC in verschiedenen Versionen und deren Konvertierung in ein strukturiertes CSV-Format mit Hilfe der kostenlosen Reverse Engineering SDKs wurden fast 10 Tausend RVT- und IFC-Projekte in einer großen Apache Parquet-Tabellendatei gesammelt und zur Analyse in Pandas DataFrame hochgeladen (Abb. 9.1-9).

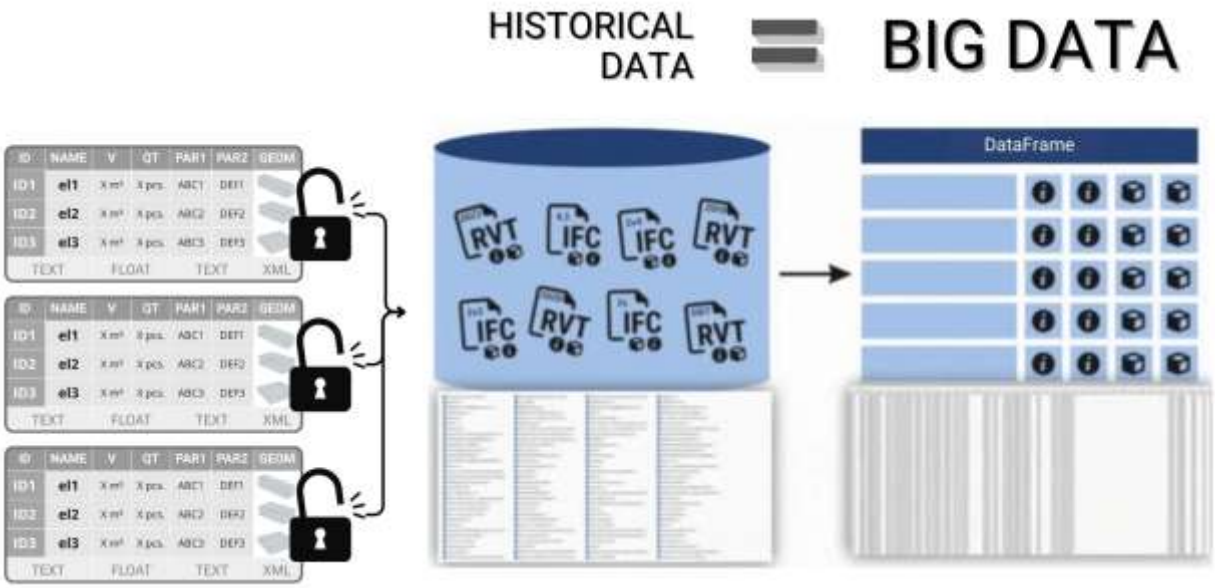


Abbildung 9.1-9 Strukturierte Daten Projektdaten ermöglichen es Ihnen, eine beliebige Anzahl von Projekten in einer einzigen zweidimensionalen Tabelle zusammenzufassen.

Die Daten aus dieser groß angelegten Sammlung enthalten folgende Informationen: Der IFC-Dateisatz enthält etwa 4 Millionen Entitäten (Zeilen) und 24.962 Attribute (Spalten), und der RVT-Dateisatz, der aus etwa 6 Millionen Entitäten (Zeilen) besteht, enthält 27.025 verschiedene Attribute (Spalten).

Diese Informationssätze (Abb. 9.1-10) umfassen Millionen von Elementen, für die jeweils zusätzlich die Koordinaten der Geometrie der Bounding Box (ein Rechteck, das die Grenzen eines Objekts im Projekt definiert) und Bilder jedes Elements im PNG-Format sowie die Geometrie im offenen XML-Format - DAE (Collada) erstellt und einer gemeinsamen Tabelle hinzugefügt wurden.

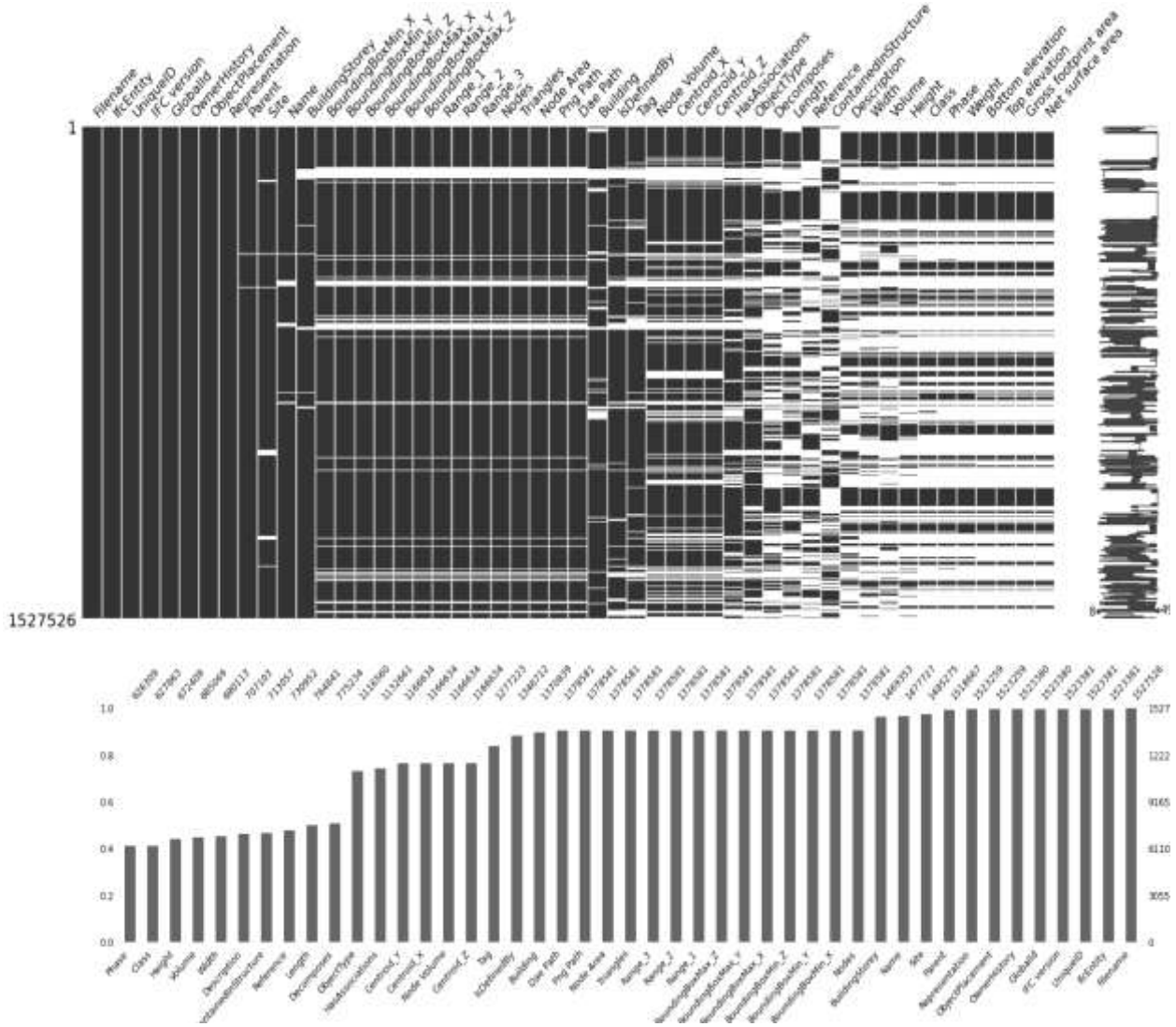


Abb. 9.1-10 Sabset von 1,5 Millionen Elementen und Visualisierung (missingno library) der Belegung der ersten 100 Attribute als Histogramm.

Auf diese Weise erhielten wir alle Informationen über mehrere zehn Millionen Elemente aus 4.596 IFC-Projekten und 6.471 RVT-Projekten, bei denen alle Attribute/Eigenschaften aller Entity-Elemente und ihre Geometrie (Bounding Box) in eine strukturierte Form einer einzigen Tabelle (DataFrame) übersetzt wurden (Abb. 9.1-10 - Daten auf DataFrame-Populationen erscheinen als Histogramme).

> Die während des Analyseprozesses aufgezeichneten Histogramme (Abb. 9.1-10, Abb. 9.2-6, Abb. 9.2-7) ermöglichen eine schnelle Beurteilung der Datendichte und der Häufigkeit des Auftretens von Werten in den Spalten. Dies gibt einen ersten Einblick in die Verteilung der Merkmale, das Vorhandensein von Ausreißern und die potenzielle Nützlichkeit einzelner Attribute für die Analyse und den Aufbau von Modellen für maschinelles Lernen.

Ein Beispiel für die praktische Nutzung dieses Datensatzes (Abb. 9.1-10) ist das Projekt "5000 IFC und RVT ". [149], verfügbar auf der Kaggle-Plattform. Es präsentiert Jupyter Notebook mit einer kompletten Pipeline-

Lösung: von der Datenvorverarbeitung und -analyse bis hin zur Visualisierung der Ergebnisse unter Verwendung von Python-Bibliotheken - Pandas, Matplotlib, Seaborn, Folium und anderen (Abb. 9.1-11).

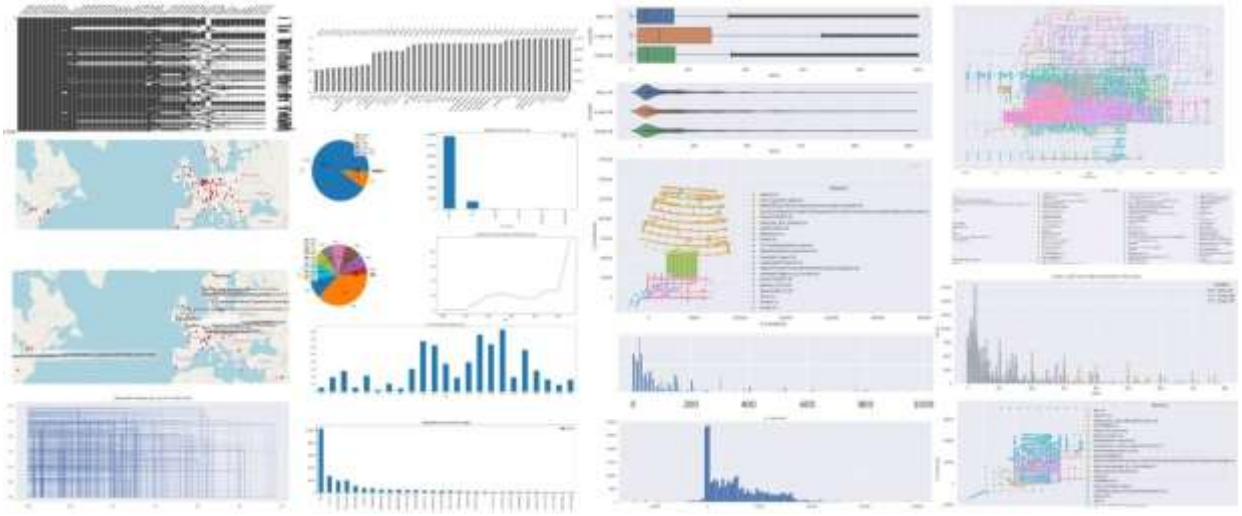


Abbildung 9.1-11 Beispiele für die Analyse von Daten aus CAD-Formaten (BIM) unter Verwendung von Python-Visualisierungsbibliotheken und der Pandas-Bibliothek.

Auf der Grundlage von Metainformationen ist es möglich, festzustellen, in welchen Städten bestimmte Projekte entwickelt wurden, und dies auf einer Karte darzustellen (z. B. mithilfe der Folium-Bibliothek). Darüber hinaus ermöglichen Zeitstempel in den Daten die Untersuchung von Mustern, wann Dateien gespeichert oder bearbeitet wurden: nach Wochentag, Tageszeit und Monat.

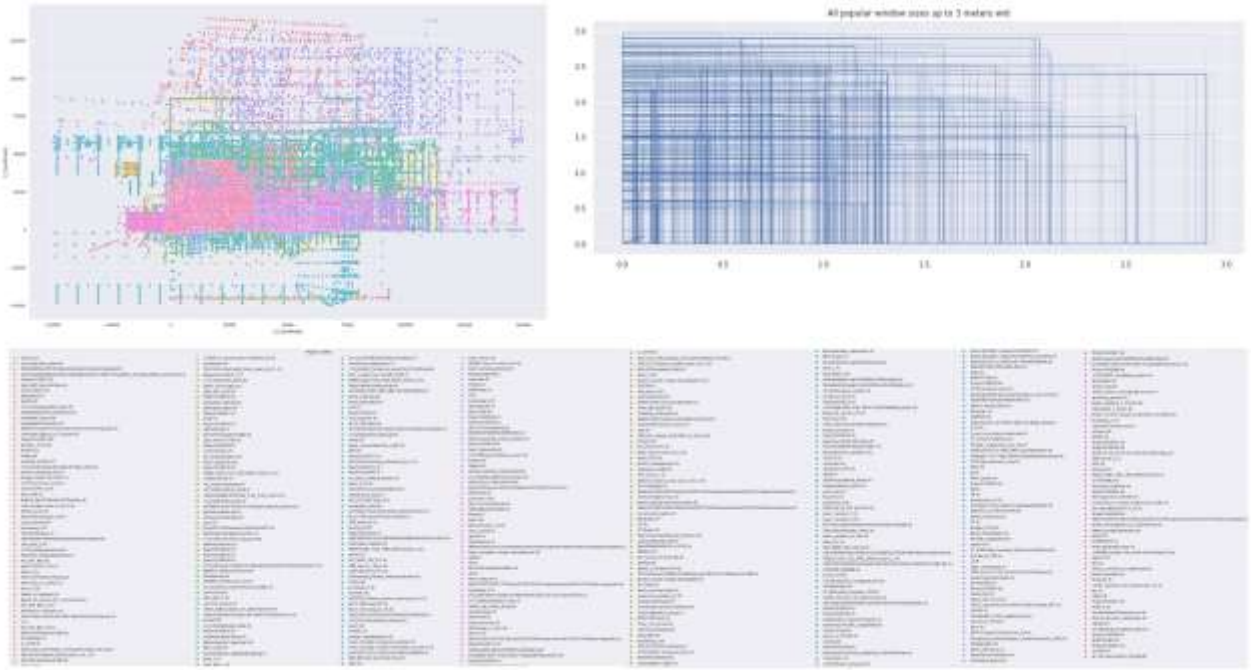


Abb. 9.1-12 Visualisierung der geometrischen Position aller Säulen und der Abmessungen aller Fenster bis zu 3 Metern in Projekten aus der Liste am unteren Rand der Karte.

Geometrische Parameter in Form von Bounding Boxen, die aus den Modellen extrahiert werden, eignen sich

auch für aggregierte Analysen. Abbildung 9.1-12 zeigt beispielsweise zwei Diagramme: das linke zeigt die Verteilung der Abstände zwischen den Säulen für alle Projekte relativ zum Nullpunkt, und das rechte zeigt die Abmessungen aller Fenster bis zu einer Höhe von 3 Metern in einer Stichprobe von Zehntausenden von Fensterelementen (nach Gruppierung des gesamten Datensatzes durch den Parameter "Kategorie" mit dem Wert "OST_Windows", "IfcWindows").

Der Pipeline-Analysecode für dieses Beispiel und der Datensatz selbst sind auf der Kaggle-Website unter dem Titel "5000 IFC und RVT | DataDrivenCo-nstruction.io projects" [149] verfügbar. Diese fertige Pipeline zusammen mit dem Datensatz kann kopiert und kostenlos online auf Kaggle oder offline in einer der gängigen IDEs ausgeführt werden: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse mit PyDev-Plugin, Thonny, Wing IDE, IntelliJ IDEA mit Python-Plugin, JupyterLab oder beliebte Online-Tools Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

> Die analytischen Erkenntnisse, die aus der Verarbeitung und Untersuchung riesiger Mengen strukturierter Daten gewonnen werden, werden eine entscheidende Rolle bei den Entscheidungsprozessen in der Bauindustrie spielen.

Mit dieser Art von Informationsanalyse auf der Grundlage früherer Projekte können die Fachleute beispielsweise den Material- und Arbeitsbedarf effektiv vorhersagen und die Konstruktionslösungen vor Baubeginn optimieren

Während jedoch Entwurfsdaten oder Baugenehmigungen relativ statische Informationen sind, die sich nur langsam ändern, wird der Bauprozess selbst schnell mit einer Vielzahl von Sensoren und IoT -Geräten gesättigt: Kameras, automatisierte Überwachungssysteme, die Daten in Echtzeit übertragen - all dies macht die Baustelle zu einer dynamischen digitalen Umgebung, in der Daten in Echtzeit analysiert werden müssen.

IoT Internet der Dinge und intelligente Verträge

IoT Das Internet der Dinge stellt eine neue Welle der digitalen Transformation dar, bei der jedes Gerät seine eigene IP-Adresse erhält und Teil eines globalen Netzwerks wird. IoT ist ein Konzept, bei dem physische Objekte mit dem Internet verbunden werden, um Daten zu sammeln, zu verarbeiten und zu übertragen. Im Bauwesen bedeutet dies die Möglichkeit, Bauprozesse in Echtzeit zu steuern, die Materialverschwendung zu minimieren, den Verschleiß von Geräten vorherzusagen und die Entscheidungsfindung zu automatisieren.

> Laut dem CFMA-Artikel "Preparing for the Future with Connected Construction" [150] wird die Bauindustrie in den nächsten zehn Jahren einen tiefgreifenden digitalen Wandel durchlaufen, der in dem Konzept des vernetzten Bauens gipfeln wird - einer vollständig integrierten und automatisierten Baustelle.

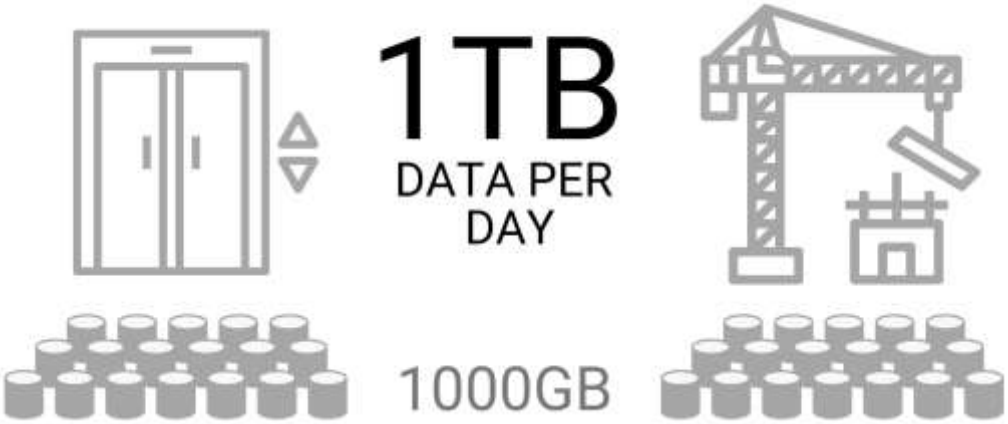


Abb. 9.1-13 IOT- oder Baustellendatengeräte können pro Tag Terabytes an Daten erzeugen und übertragen.

Eine digitale Baustelle bedeutet, dass alle Elemente des Bauwesens - von der Planung und Logistik bis hin zur Arbeitsausführung und Qualitätskontrolle auf der Baustelle mit Hilfe von stationären Kameras und Quadrocoptern - in ein einziges dynamisches digitales Ökosystem integriert werden. In Teil 7 dieses Buches haben wir uns bereits mit den Möglichkeiten von Apache NiFi (Abb. 7.4-5) befasst, einem kostenlosen und quelloffenen Tool, das das Streaming von Daten in Echtzeit ermöglicht - von der Sammlung aus verschiedenen Quellen bis zur Übertragung an Speicher- oder Analyseplattformen.

Daten über Baufortschritt, Materialverbrauch, Gerätezustand und Sicherheit werden in Echtzeit an Analysesysteme übermittelt (Abb. 9.1-13). So können potenzielle Risiken vorhergesagt, auf Abweichungen zeitnah reagiert und die Baustellenabläufe optimiert werden. Wichtige Bestandteile einer digitalen Baustelle sind:

- IoT -Sensoren - Überwachung von Umweltparametern, Überwachung von Baumaschinen und Kontrolle der Arbeitsbedingungen.
- Digitale Zwillinge - virtuelle Modelle von Gebäuden und Infrastrukturen, um mögliche Abweichungen vorherzusagen und Fehler zu vermeiden.
- Automatisierte Logistiksysteme - Lieferkettenmanagement in Echtzeit zur Reduzierung von Ausfallzeiten und Kosten.
- Robotersysteme im Bauwesen - der Einsatz autonomer Maschinen zur Ausführung von Routine- und gefährlichen Aufgaben.

Die Robotisierung und Einsatz von IoT und das Konzept der vernetzten Baustelle (Construction) werden nicht nur die Effizienz steigern und die Kosten senken, sondern auch eine neue Ära der Sicherheit, des nachhaltigen Bauens und des vorausschauenden Projektmanagements einläuten.

> RFID (Radio Frequency Identification) Tags sind ebenfalls eine der wichtigsten Komponenten des IoT. Sie werden verwendet, um Materialien, Maschinen und sogar Personal auf einer Baustelle zu identifizieren und zu verfolgen, was die Transparenz und Kontrolle der Projektressourcen erhöht.

Die RFID -Technologie dient der automatischen Erkennung von Objekten anhand von Funksignalen. Sie besteht aus drei Schlüsselelementen:

- RFID -Anhänger (passiv oder aktiv) - enthalten eine eindeutige Kennung und werden an Materialien, Werkzeugen oder Maschinen angebracht.
- Scanner sind Geräte, die Informationen von Etiketten lesen und an das System übertragen.
- Zentrale Datenbank - speichert Informationen über den Standort, den Status und die Bewegung von Objekten.

Anwendung von RFID im Bauwesen:

- Automatische Materialbuchhaltung - Etiketten an Transportbetonprodukten, Bewehrung oder Sandwichpaneel-Paketen ermöglichen eine Bestandskontrolle und verhindern Diebstahl.
- Kontrolle der Arbeit des Personals - RFID - Mitarbeiterausweise erfassen die Zeiten des Schichtbeginns und -endes und ermöglichen so eine Aufzeichnung der Arbeitszeiten.
- Überwachung der Ausrüstung - RFID - System verfolgt die Bewegung der Ausrüstung, verhindert Ausfallzeiten und verbessert die Effizienz der Logistik.

Ergänzt wird dieses Technologiepaket durch Blockchain-basierte Smart Contracts, die Zahlungen, Lieferkontrolle und Vertragserfüllung ohne Zwischenhändler automatisieren und so das Risiko von Betrug und Verzögerungen verringern.

> In Ermangelung eines gemeinsamen Datenmodells sind intelligente Verträge heute einfach ein Code, auf den sich die Teilnehmer einigen. Mit einem datenzentrierten Ansatz ist es jedoch möglich, ein gemeinsames Modell der Vertragsparameter zu erstellen, es in einer Blockchain zu kodieren und die Erfüllung der Bedingungen zu automatisieren.

In einem System zur Verwaltung der Lieferkette kann ein intelligenter Vertrag beispielsweise die Lieferung einer Sendung anhand von IoT-Sensoren und RFID-Etiketten nachverfolgen und die Zahlung automatisch überweisen, sobald die Sendung eingetroffen ist. In ähnlicher Weise kann ein intelligenter Vertrag auf einer Baustelle den Abschluss einer Arbeitsphase - wie das Verlegen von Bewehrungsstäben oder das Gießen eines Fundaments - auf der Grundlage von Daten von Drohnen oder Bausensoren erfassen und automatisch die nächste Zahlung an den Auftragnehmer veranlassen, ohne dass manuelle Kontrollen und Papierzertifikate erforderlich sind.

Doch trotz neuer Technologien und der Bemühungen internationaler Normungsorganisationen verkompliziert eine Fülle konkurrierender Normen die IoT-Landschaft.

> Laut einer 2017 veröffentlichten Cisco-Studie [151] brechen fast 60% der Internet-of-Things-Initiativen (IoT) in der Proof-of-Concept-Phase ab, und nur 26% der Unternehmen betrachten ihre IoT-Projekte als vollständig erfolgreich. Außerdem erreicht ein Drittel der abgeschlossenen Projekte nicht die gesteckten Ziele und wird auch nach der Umsetzung nicht als erfolgreich anerkannt.

Einer der Hauptgründe ist die mangelnde Interoperabilität zwischen Plattformen, die Daten von verschiedenen Sensoren verarbeiten. Infolgedessen bleiben die Daten in separaten Lösungen isoliert. Eine Alternative zu diesem Ansatz, wie auch in anderen ähnlichen Fällen (die wir in diesem Buch behandelt haben), ist eine Architektur, die auf den Daten selbst als primärem Gut aufbaut.

> IoT-Sensoren spielen nicht nur bei der Überwachung des technischen Zustands von Anlagen eine wichtige Rolle, sondern auch bei der prädiktiven Analyse zur Verringerung von Risiken auf der Baustelle und zur Verbesserung der Gesamtprozessleistung durch Vorhersage von Ausfällen und Abweichungen.

Die von IoT-Sensoren und RFID-Etiketten gesammelten Daten können in Echtzeit von Algorithmen des maschinellen Lernens verarbeitet werden, die Anomalien erkennen und Ingenieure im Voraus vor möglichen Störungen warnen können. Dies kann von Mikrorissen in Betonstrukturen bis hin zu untypischen Pausen beim Betrieb von Turmdrehkränen reichen, die auf technisches Versagen oder Verstöße gegen Vorschriften hinweisen. Darüber hinaus können fortschrittliche Algorithmen zur Verhaltensanalyse Verhaltensmuster erfassen, die beispielsweise auf körperliche Ermüdung des Personals hindeuten, und so das proaktive Management von Sicherheit und Wohlbefinden der Mitarbeiter auf der Baustelle verbessern.

Im Baugewerbe geschehen Unfälle und Ausfälle - ob von Maschinen oder Menschen - selten plötzlich. Meist gehen ihnen kleine Abweichungen voraus, die unbemerkt bleiben. Predictive Analytics und maschinelles Lernen ermöglichen es, diese Signale frühzeitig zu erkennen, noch bevor kritische Folgen auftreten.

Während Dokumente, Projektdateien und Daten von IoT-Geräten und RFID-Etiketten den digitalen Fußabdruck von Bauprojekten bilden, kann maschinelles Lernen helfen, nützliche Erkenntnisse daraus zu gewinnen. Mit dem Wachstum der Daten und der Demokratisierung des Datenzugriffs ergeben sich für die Baubranche neue Möglichkeiten in den Bereichen Analytik, prädiktive Analytik und künstliche Intelligenz.

KAPITEL 9.2.
MASCHINELLES LERNEN UND VORHERSAGEN

Maschinelles Lernen und künstliche Intelligenz werden die Art und Weise, wie wir bauen, verändern

Die Datenbanken der verschiedenen Systeme im Baugewerbe - mit ihrer unweigerlich verfallenden und immer komplexeren Infrastruktur - werden zum Nährboden für zukünftige Lösungen. Die Server der Unternehmen sind wie ein Wald reich an einer Biomasse wichtiger Informationen, die oft unterirdisch in den Eingeweiden von Ordnern und Servern verborgen sind. Die Massen von Daten aus den verschiedenen Systemen, die heute entstehen - nach der Nutzung, nachdem sie auf den Boden des Servers gefallen sind und nach Jahren der Fossilisierung - werden in Zukunft maschinelles Lernen und Sprachmodelle befeuern. Firmeninterne Chatrooms (z.B. eine eigene Instanz des lokal konfigurierten ChatGPT, LlaMa, Mistral, DeepSeek) werden auf diesen firmeninternen Modellen aufbauen, die einen zentralen Speicher nutzen, um schnell und bequem Informationen abzurufen und die notwendigen Grafiken, Dashboards und Dokumente zu erstellen.

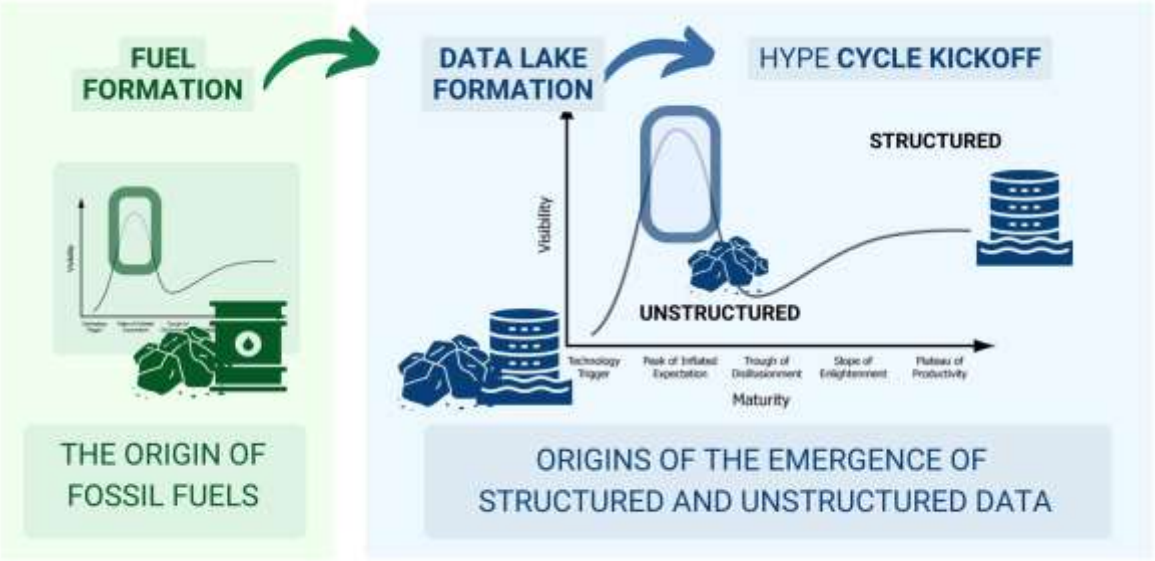


Abb. 9.2-1 So wie sich Bäume in Kohle verwandeln, so verwandeln sich auch Informationen im Laufe der Zeit unter dem Druck von Zeit und Analytik in wertvolle Geschäftsenergie.

Durch die Versteinerung von Pflanzenmasse in Verbindung mit Druck und Temperatur entsteht eine homogene und einzigartig strukturierte homogene Masse von Bäumen verschiedener Arten, die zu unterschiedlichen Zeiten gelebt haben - Holzkohle [152]. In gleicher Weise bilden Informationen, die auf Festplatten in unterschiedlichen Formaten und zu unterschiedlichen Zeiten unter dem Druck der Analyseabteilungen und der Temperatur des Qualitätsmanagements aufgezeichnet wurden, schließlich eine homogen strukturierte Masse wertvoller Informationen (Abb. 9.2-1).

Diese Informationsschichten (oder häufiger isolierte Nuggets) entstehen durch eine sorgfältige Datenorganisation durch erfahrene Analysten, die nach und nach wertvolle Informationen aus scheinbar lange irrelevanten Daten extrahieren.

In dem Moment, in dem diese ausgereiften Datenschichten nicht mehr nur in Berichten "verbrannt" werden, sondern in Geschäftsprozessen zirkulieren, Entscheidungen bereichern und Prozesse verbessern, ist das Unternehmen bereit für den nächsten Schritt - den Übergang zu maschinellem Lernen und künstlicher Intelligenz (Abb. 9.2-2).

> Maschinelles Lernen (ML - Machine learning) ist eine Klasse von Methoden zur Lösung von Problemen der künstlichen Intelligenz. Algorithmen des maschinellen Lernens erkennen Muster in großen Datensätzen und nutzen sie, um selbst zu lernen. Jeder neue Datensatz ermöglicht es den mathematischen Algorithmen, sich zu verbessern und an die gewonnenen Informationen anzupassen, wodurch die Genauigkeit der Empfehlungen und Vorhersagen ständig verbessert werden kann.

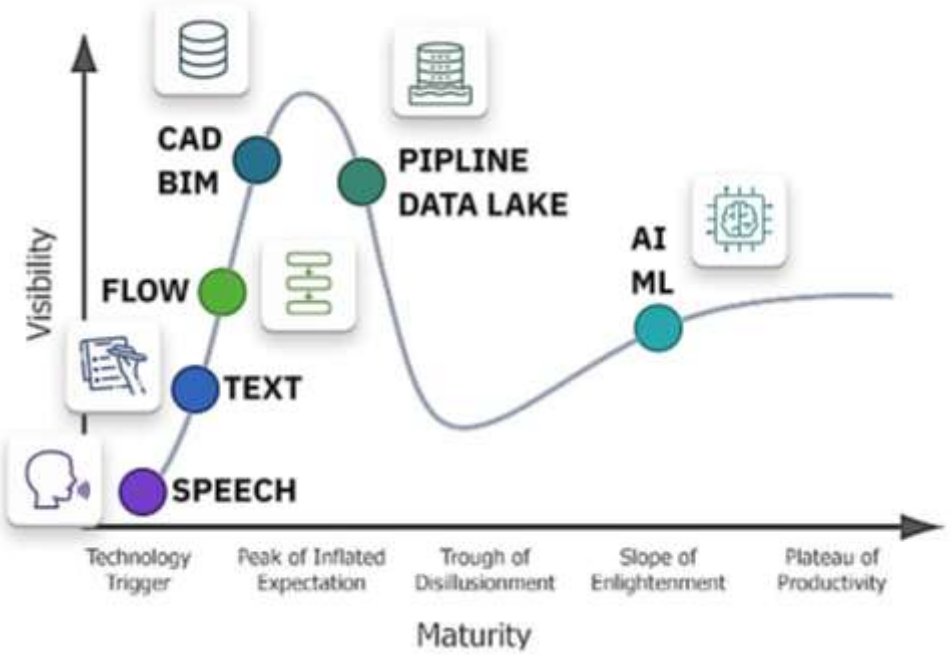


Abb. 9.2-2 Das Verblassen von Datenerstellungstechnologien und der Einsatz von Analysetools öffnet die Tür zum Thema maschinelles Lernen.

Wie der einflussreiche CEO des größten Investmentfonds der Welt (der wichtige Anteile an fast allen der größten Bausoftwareunternehmen sowie an den Unternehmen besitzt, die den größten Immobilienbestand der Welt besitzen [55]) in einem Interview 2023 sagte - maschinelles Lernen wird die Welt des Bauwesens verändern.

> *KI hat ein enormes Potenzial. Sie wird die Art, wie wir arbeiten und wie wir leben, verändern. KI und Robotik werden die Art und Weise verändern, wie wir arbeiten und bauen, und wir werden in der Lage sein, KI und Robotik als Mittel zur Steigerung der Produktivität zu nutzen [153].*
>
> - CEO des größten Investmentfonds der Welt, Interview, September 2023.

Maschinelles Lernen (ML) funktioniert durch die Verarbeitung großer Datenmengen, wobei statistische Techniken eingesetzt werden, um Aspekte des menschlichen Denkens nachzuahmen. Die meisten Unternehmen verfügen jedoch nicht über solche Datensätze, und wenn doch, dann sind sie oft nicht ausreichend gekennzeichnet. Hier können semantische Technologien und Transfer Learning helfen, eine Technik, die es ML ermöglicht, mit kleinen Datenmengen effektiver umzugehen, und deren Durchführbarkeit in früheren Kapiteln dieses Teils erörtert wurde.

Das Wesen des Transferlernens besteht darin, dass man nicht jede Aufgabe von Grund auf neu lernen muss, sondern Wissen aus verwandten Bereichen nutzen kann. Man muss sich darüber im Klaren sein, dass Muster und Entdeckungen aus anderen Branchen angepasst und in der Baubranche angewendet werden können. So helfen beispielsweise Methoden zur Optimierung von Logistikprozessen, die im Einzelhandel entwickelt wurden, die Effizienz des Lieferkettenmanagements im Bauwesen zu verbessern. Big-Data-Analysen, die im Finanzwesen aktiv genutzt werden, können auf Kostenprognosen und Risikomanagement bei Bauprojekten angewendet werden. Und die in der Industrie entwickelten Computer-Vision- und Robotik-Technologien finden bereits Anwendung in der automatisierten Qualitätskontrolle, der Sicherheitsüberwachung und der Verwaltung von Baustelleneinrichtungen.

> Transferlernen ermöglicht nicht nur eine schnellere Einführung von Innovationen, sondern auch eine Senkung der Kosten für ihre Entwicklung, indem die bereits gesammelten Erfahrungen anderer Branchen genutzt werden.

Abb. 9.2-3 Künstliche Intelligenz Technologien und Robotik werden die Hauptantriebskraft der Zukunft sein, um die Produktivität in der Bauindustrie zu steigern.

Das menschliche Denken ist nach einem ähnlichen Prinzip organisiert: Wir bauen auf zuvor erworbenem Wissen auf, um neue Probleme zu lösen (Abb. 4.4-19, Abb. 4.4-20, Abb. 4.4-21). Beim maschinellen Lernen funktioniert dieser Ansatz ebenfalls - indem wir das Datenmodell vereinfachen und eleganter gestalten, können wir die Komplexität des Problems für ML-Algorithmen reduzieren. Dies wiederum verringert den Bedarf an großen Datenmengen und senkt die Rechenkosten.

Von der subjektiven Einschätzung zur statistischen Prognose

Die Zeiten, in denen strategische Entscheidungen von der Intuition einzelner Manager abhingen (Abb. 9.2-4), gehören der Vergangenheit an. In einem zunehmend wettbewerbsorientierten und anspruchsvollen wirtschaftlichen Umfeld wird ein subjektiver Ansatz zu riskant und ineffizient. Unternehmen, die sich weiterhin auf persönliche Meinungen verlassen, anstatt die Daten von objektiv zu analysieren, verlieren die Fähigkeit, schnell auf Veränderungen zu reagieren.

Das Wettbewerbsumfeld erfordert Genauigkeit und Wiederholbarkeit auf der Grundlage von Daten, statistischen Mustern und berechenbaren Wahrscheinlichkeiten. Entscheidungen können nicht mehr auf Gefühlen beruhen, sondern müssen auf Korrelationen, Trends und Vorhersagemodellen beruhen, die aus der Analytik und dem maschinellen Lernen abgeleitet werden. Dies ist nicht nur ein Wechsel der Werkzeuge - es ist ein Wechsel in der Logik des Denkens: von Annahmen zu Beweisen, von subjektiven Wahrscheinlichkeiten zu statistisch berechneten Abweichungen, von Gefühlen zu Fakten.

Abb. 9.2-4 Die Ära der Entscheidungen, die von HiPPO (der Meinung des bestbezahlten Mitarbeiters) getroffen werden, wird mit dem Aufkommen von Big Data und maschinellem Lernen der Vergangenheit angehören.

Manager, die sich früher ausschließlich auf ihr eigenes Gefühl verlassen haben, werden unweigerlich mit einer neuen Realität konfrontiert: Autorität bestimmt nicht mehr die Entscheidungen. Im Zentrum des Managements stehen heute Systeme, die Millionen von Parametern und Vektoren analysieren, versteckte Muster erkennen und optimale Strategien vorschlagen.

Der Hauptgrund, warum Unternehmen heute immer noch davon absehen, ML einzusetzen, ist die mangelnde Transparenz. Die meisten Modelle arbeiten als "Blackbox" für Manager, ohne zu erklären, wie genau sie zu ihren Schlussfolgerungen kommen. Das führt zu Problemen: Algorithmen können Stereotypen verstärken und sogar humorvolle Situationen schaffen, wie im Fall des Chatbots von Microsoft, der sich schnell zu einem toxischen Kommunikationsinstrument entwickelte [154].

> In Deep Thinking reflektiert Garry Kasparov, ehemaliger Schachweltmeister, über seine Niederlage gegen den IBM Big Blue Computer [155]. Er argumentiert, dass der wahre Wert der KI nicht darin liegt, die menschliche Intelligenz zu kopieren, sondern darin, unsere Fähigkeiten zu ergänzen. KI sollte Aufgaben übernehmen, bei denen der Mensch schwach ist, während der Mensch Kreativität einbringt. Computer haben die traditionelle Herangehensweise an die Analyse von Schachspielen verändert. Anstatt faszinierende Geschichten über Partien zu erfinden, bewerten Computerschachprogramme jeden Zug unparteiisch, nur auf der Grundlage seiner tatsächlichen Stärke oder Schwäche. Kasparow stellt fest, dass die menschliche Tendenz, Ereignisse als zusammenhängende Geschichten und nicht als einzelne Aktionen zu sehen, oft zu falschen Schlussfolgerungen führt - nicht nur im Schach, sondern im Leben allgemein.

Wenn Sie vorhaben, maschinelles Lernen für Vorhersagen und Analysen zu nutzen, ist es daher wichtig, die grundlegenden Prinzipien zu verstehen - wie Algorithmen funktionieren und wie Daten verarbeitet werden, bevor Sie damit beginnen, Werkzeuge für maschinelles Lernen und AI bei Ihrer Arbeit einzusetzen. Der beste Weg für den Einstieg ist praktische Erfahrung.

Eines der geeignetsten Werkzeuge für eine erste Einführung in das Thema maschinelles Lernen und Vorhersage ist das Jupyter Notebook und der beliebte klassische Titanic-Datensatz, der eine visuelle Einführung in die wichtigsten Methoden der Datenanalyse und der Erstellung von ML-Modellen bietet.

Titanic-Datensatz: Hallo Welt in der Welt der Analytik Daten und Big Data

Eines der bekanntesten Beispiele für den Einsatz von ML in der Datenanalyse ist die Analyse des Titanic-Datensatzes, der häufig zur Untersuchung der Überlebenswahrscheinlichkeit der Passagiere verwendet wird. Das Erlernen dieser Tabelle ist vergleichbar mit dem "Hello World"-Programm beim Erlernen von Programmiersprachen.

Beim Untergang der RMS Titanic im Jahr 1912 kamen 1502 von 2224 Menschen ums Leben. Der Titanic-Datensatz enthält nicht nur Informationen darüber, ob ein Passagier überlebt hat, sondern auch Attribute wie: Alter, Geschlecht, Ticketklasse und andere Parameter. Dieser Datensatz ist kostenlos verfügbar und kann auf verschiedenen Offline- und Online-Plattformen geöffnet und analysiert werden.

```
Link zum Titanic-Datensatz:
https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv
```

Im Kapitel "LLM-fähige IDEs und zukünftige Änderungen in der Programmierung" haben wir bereits über Jupyter Notebook gesprochen - eine der beliebtesten Entwicklungsumgebungen für Datenanalyse und maschinelles Lernen. Kostenlose Cloud-Analoga von Jupyter Notebook sind die Plattformen Kaggle und Google Collab, die es Ihnen ermöglichen, Python-Code ohne Installation von Software auszuführen und kostenlosen Zugang zu Rechenressourcen zu erhalten.

> Kaggle ist die größte Wettbewerbsplattform für Datenanalyse und maschinelles Lernen mit einer integrierten Code-Ausführungsumgebung. Im Oktober 2023 hat Kaggle über 15 Millionen Nutzer [156] aus 194 Ländern.

Laden Sie den Titanic-Datensatz auf der Kaggle-Plattform (Abb. 9.2-5) herunter und verwenden Sie ihn, um den Datensatz (eine Kopie davon) zu speichern und Python Code mit vorinstallierten Bibliotheken direkt in einem Browser auszuführen, ohne eine spezielle IDE installieren zu müssen.

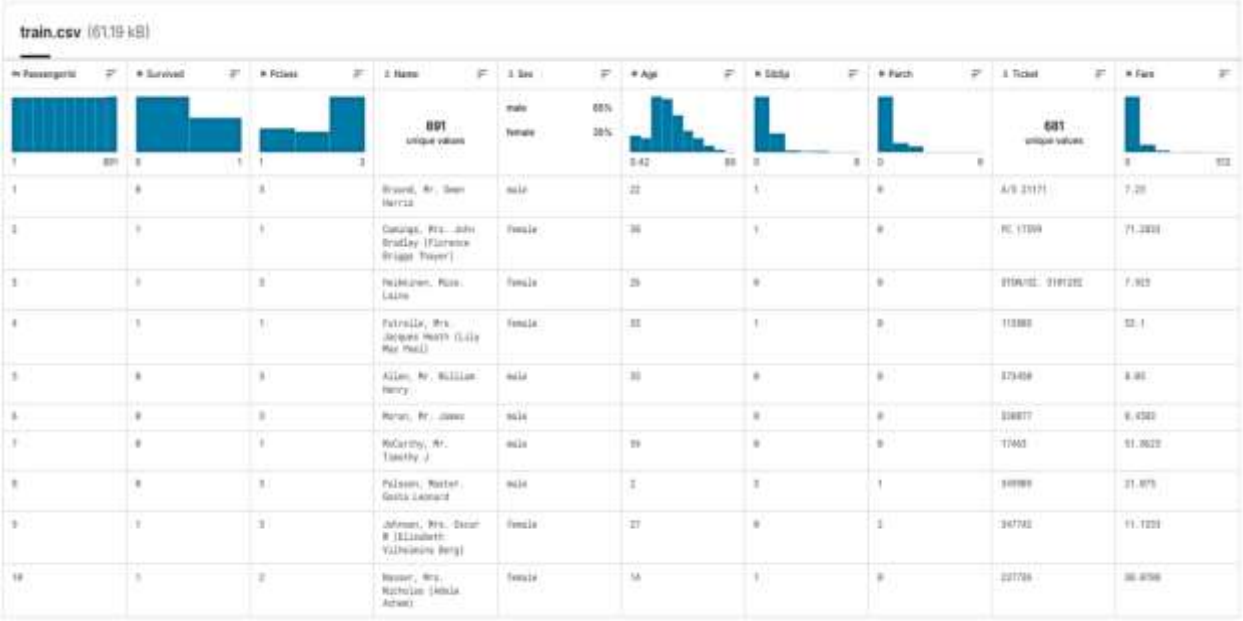


Abb. 9.2-5 Titanic-Tischstatistiken - der beliebteste Trainingsdatensatz für das Erlernen von Datenanalyse und maschinellem Lernen.

Der Titanic-Datensatz enthält Daten über die 2.224 Passagiere an Bord der *RMS Titanic* zum Zeitpunkt ihres Untergangs im Jahr 1912. Der Datensatz wird in zwei separaten Tabellen dargestellt, einer Trainings- (train.csv) und einer Teststichprobe (test.csv), so dass er sowohl für das Training von Modellen als auch für die Bewertung ihrer Genauigkeit anhand neuer Daten verwendet werden kann.

Der Trainingsdatensatz enthält sowohl Attribute - Attribute der Passagiere (Alter, Geschlecht, Ticketklasse und andere) - als auch Informationen darüber, wer überlebt hat (Spalte mit binären Werten "Überlebt"). Der Trainingsdatensatz (Abb. 9.2-6 - Datei train.csv) wird zum Trainieren des Modells verwendet. Der Testdatensatz (Abb. 9.2-7 - Datei test.csv) enthält nur Passagierattribute ohne Informationen über die Überlebenden (ohne eine einzige Spalte "Survivor"). Der Testdatensatz dient dazu, das Modell an neuen Daten zu testen und seine Genauigkeit zu bewerten.

Wir haben also fast identische Attribute von Passagieren in den Trainings- und Testdatensätzen. Der einzige wesentliche Unterschied besteht darin, dass wir im Testdatensatz eine Liste von Fahrgästen haben, die die Spalte "Survivor" nicht aufweisen - die Zielvariable, deren Vorhersage wir mit Hilfe verschiedener mathematischer Algorithmen erlernen wollen. Nach der Erstellung des Modells können wir die Ausgabe unseres Modells mit dem realen Parameter "Survivor" aus dem Testdatensatz vergleichen, den wir zur Bewertung der Ergebnisse heranziehen werden.

Hauptspalten der Tabelle, Passagierparameter im Trainings- und Testdatensatz:

- **PassengerId** - eindeutige Kennung des Passagiers
- **Überlebt** - 1, wenn der Fahrgast überlebt hat, 0, wenn er tot ist (im Testsatz nicht verfügbar)
- **Pclass** - Fahrscheinklasse (1, 2 oder 3)

- **Name** - Name des Passagiers
- **Geschlecht** - Geschlecht des Fahrgastes (männlich/weiblich)
- **Alter**
- **SibSp** - Anzahl der Brüder/Schwestern oder Ehepartner an Bord
- **Parch** - Anzahl der Eltern oder Kinder an Bord
- **Ticket** - Ticketnummer
- **Fahrpreis** - Ticketpreis
- **Kabine** - Kabinennummer (viele Daten fehlen)
- **Eingeschifft** ist der Hafen der Einschiffung (C = Cherbourg, Q = Queenstown, S = Southampton).

Um fehlende Daten in beiden Tabellen zu visualisieren, können Sie die missingno-Bibliothek (Abb. 9.2-6, Abb. 9.2-7) verwenden, die fehlende Werte in Form eines Histogramms anzeigt, wobei weiße Felder fehlende Daten anzeigen. Diese Visualisierung ermöglicht eine schnelle Bewertung der Datenqualität vor der Verarbeitung.

titanic_missingno.py

```
1  import pandas as pd
2  import missingno as msno
3  import matplotlib.pyplot as plt
4
5  # Load the dataset
6  train = pd.read_csv('train.csv')
7
8  # Visualize missing values
9  msno.matrix(train)
10 plt.show()
```

Run in IDE

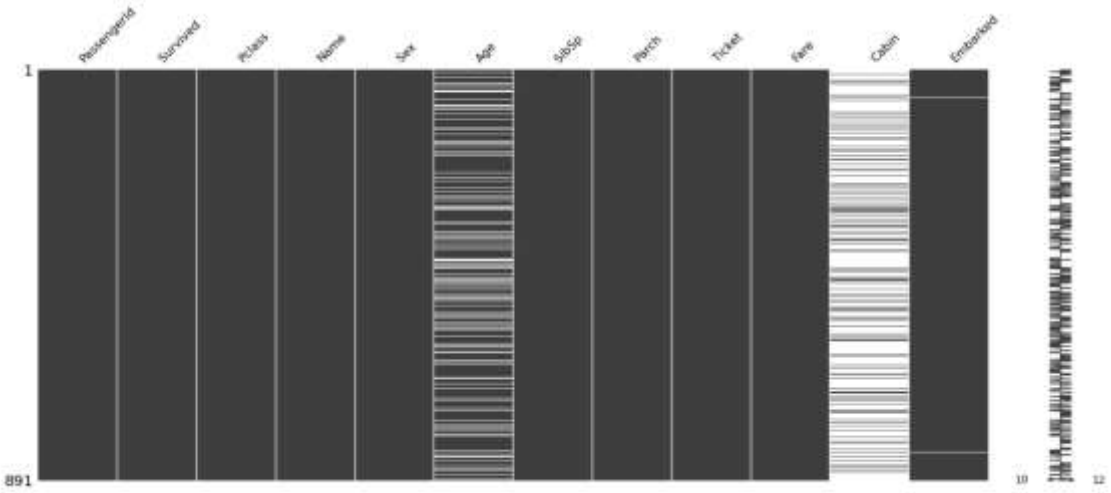


Abb. 9.2-6 Ein paar Zeilen Code werden verwendet, um die fehlenden Daten im Titanic-Trainingsdatensatz zu visualisieren, wobei der Schlüsselparameter für das Training der Parameter "Überlebt" ist.

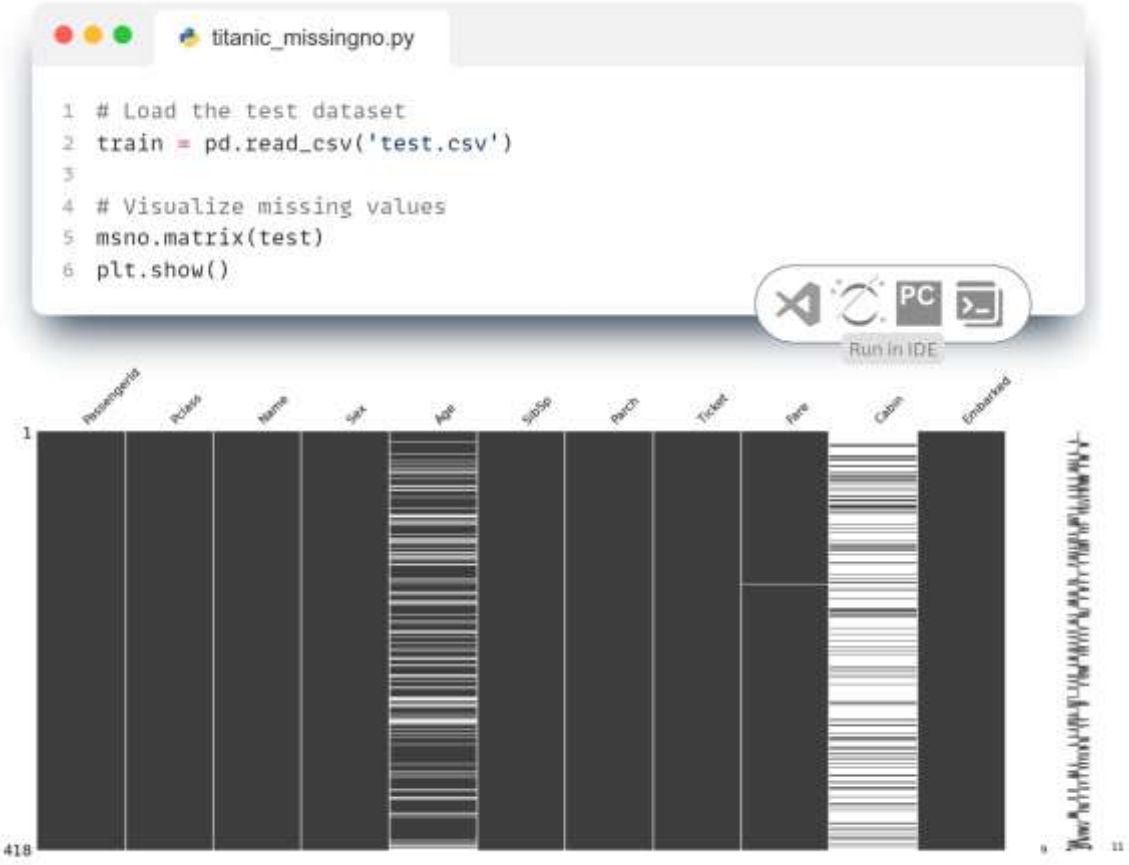


Abbildung 9.2-7 Visualisierung von fehlenden Daten im Titanic-Testdatensatz, der nur Passagiermerkmale ohne Informationen enthält.

Bevor Hypothesen formuliert und Vorhersagen auf der Grundlage des Datensatzes getroffen werden, helfen visuelle Analysen dabei, wichtige Muster in den Daten zu erkennen, ihre Qualität zu bewerten und mögliche Abhängigkeiten zu identifizieren. Es gibt viele Visualisierungstechniken, die Ihnen helfen können, den Titanic-Datensatz besser zu verstehen. Sie können Verteilungsdiagramme verwenden, um die Altersgruppen der Passagiere zu analysieren, Überlebensdiagramme nach Geschlecht und Klasse und Matrizen mit fehlenden Daten, um die Qualität der Informationen zu bewerten und die Daten zu verstehen.

Bitten wir LLM, uns bei der Visualisierung der Daten aus dem Titanic-Datensatz zu helfen, indem wir die folgende Textanfrage an ein beliebiges LLM-Modell (CHATGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN oder ein anderes) senden:

> Bitte zeigen Sie einige einfache Diagramme für den Titanic-Datensatz. Laden Sie den Datensatz selbst herunter und zeigen Sie die ⏎

LLM-Antwort in Form von vorgefertigtem Code und Diagrammen zur Visualisierung der Datensatzparameter

Create in LLM Chat

titanic_plot.py

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
file_path = "/mnt/data/titanic.csv"
df = pd.read_csv(file_path)

# Convert column names to lowercase for consistency
df.columns = df.columns.str.lower()

# Basic Plots
plt.figure(figsize=(10, 5))

# Bar Chart: Survival Count
plt.subplot(1, 3, 1)
df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
plt.title("Survivors vs. Non-Survivors")
plt.xlabel("Survived (0 = No, 1 = Yes)")
plt.ylabel("Count")

# Histogram: Age Distribution
plt.subplot(1, 3, 2)
df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")

# Scatter Plot: Fare vs. Age
plt.subplot(1, 3, 3)
plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
plt.title("Fare vs. Age")
plt.xlabel("Age")
plt.ylabel("Fare")

# Show the plots
plt.tight_layout()
plt.show()
```

Run in IDE

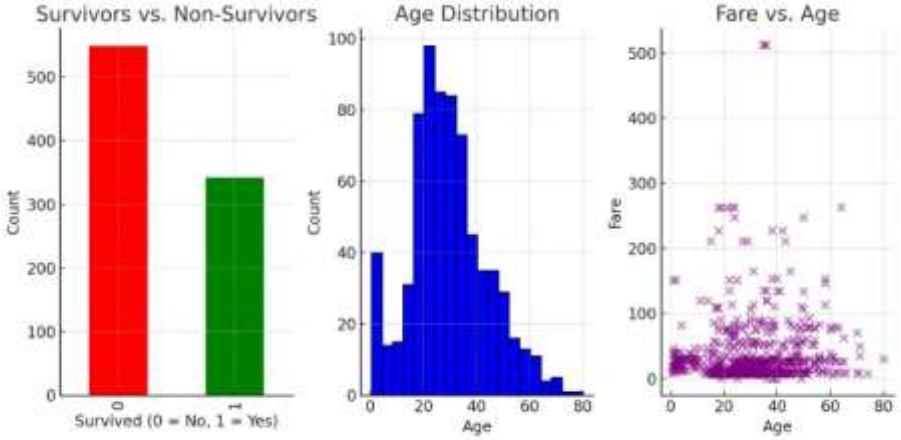


Abbildung 9.2-8 LLM hilft Ihnen, eine sofortige Visualisierung Ihrer Daten zu erhalten.

Die Datenvisualisierung ist ein wichtiger Schritt, um den Datensatz für die anschließende Konstruktion eines

maschinellen Lernmodells vorzubereiten, was nur durch das Verständnis der Daten erreicht werden kann.

Maschinelles Lernen in Aktion: von Titanic-Passagieren zum Projektmanagement

Die Haupthypothese, die zur Erforschung des maschinellen Lernsystems auf der Grundlage des Titanic-Datensatzes verwendet wurde, ist, dass bestimmte Gruppen von Passagieren eine höhere Überlebenschance hatten.

> Die kleine Tabelle der Titanic-Passagiere ist weltweit bekannt geworden, und Millionen von Menschen nutzen sie zum Trainieren, Experimentieren und Testen von Modellen, um herauszufinden, welche Algorithmen und Hypothesen auf der Grundlage des Trainingsdatensatzes für die Titanic-Passagiere das genaueste Überlebensvorhersagemodell erstellen.

Die Attraktivität des Titanic-Datensatzes liegt in seiner Kompaktheit: Mit mehreren hundert Zeilen und zwölf Spalten (Abb. 9.2-6) bietet er reichlich Gelegenheit zur Analyse. Der Datensatz ist relativ einfach ein klassisches Beispiel für eine binäre Klassifikationslösung, bei der das Ziel des Problems - das Überleben - im praktischen Format 0 oder 1 ausgedrückt wird.

> John Wheeler in "It from Bit" [7] argumentiert, dass das Universum auf binären Entscheidungen beruht. Auch ein Unternehmen, das von Menschen geführt wird, die aus Molekülen bestehen, basiert auf einer Reihe von binären Entscheidungen.

Darüber hinaus basieren die Daten auf einem realen historischen Ereignis, was sie im Gegensatz zu künstlich erstellten Beispielen für die Forschung wertvoll macht. Allein auf der Kaggle-Plattform, einer der größten Data Pipeline und ETL, nahmen 1.355.998 Personen an den auf dem Titanic-Datensatz basierenden Herausforderungen teil und entwickelten 53.963 einzigartige Data Pipeline-Lösungen [157] (Abb. 9.2-9).

> Es scheint unglaublich, aber nur 1000 Zeilen Daten über die Passagiere der Titanic mit 12 Parametern sind zu einem Feld für Millionen von Hypothesen, logischen Ketten und einzigartigen Data-Pipelines geworden. Aus einem kleinen Datensatz entstehen endlose Erkenntnisse, Hypothesen und Interpretationen - von einfachen Überlebensmodellen bis hin zu komplexen Ensembles, die verborgene Muster und komplexe Labyrinthe der Argumentation berücksichtigen.

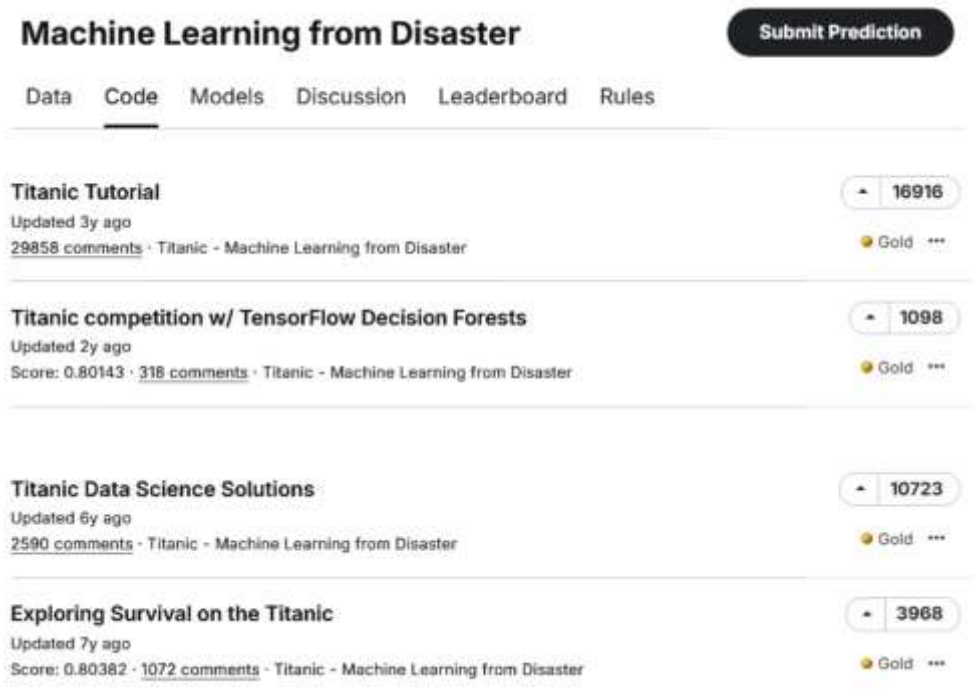


Abb. 9.2-9 Die ersten fünf Lösungen von insgesamt 53.963 fertigen und quelloffenen Pipeline-Lösungen. Allein auf Kaggle [157] haben bereits fast 1,5 Millionen Menschen versucht, dieses Problem zu lösen.

Wenn selbst eine so kleine Tabelle Millionen von eindeutigen Lösungen generieren kann (Abb. 9.2-9), was können wir dann über reale industrielle Baudatensätze sagen, bei denen die Parameter in Zehntausenden gemessen werden?

> Ein Standard-CAD -Entwurf eines relativ kleinen Gebäudes enthält Zehntausende von Elementen mit Tausenden von Parametern - von geometrischen Merkmalen bis hin zu Kosten- und Zeitattributen. Stellen Sie sich vor, wie viele potenzielle Erkenntnisse, Beziehungen, Vorhersagen und Managementhypothesen in den Daten aller Projekte Ihres Unternehmens verborgen sind, die im Laufe der Jahre gesammelt wurden. Historische Projektdaten sind nicht nur ein Archiv - sie sind das lebendige Gedächtnis einer Organisation, ihr digitaler Fußabdruck, der analysiert werden kann, um eine große Anzahl einzigartiger Hypothesen zu erstellen.

Das Wichtigste ist, dass Sie nicht darauf warten müssen, dass sich die Kaggle-Community für Ihr Unternehmen oder Ihre Daten interessiert. Sie können mit dem arbeiten, was Sie heute haben: Führen Sie Analysen mit Ihren eigenen Daten durch, trainieren Sie Modelle mit Ihren eigenen Daten, identifizieren Sie Wiederholungen, Anomalien und Muster. Wo früher jahrelange Experimente und teure Beratung nötig waren, genügen jetzt Initiative, ein LLM, ein offener Umgang mit Daten und die Bereitschaft zu lernen.

- Um einen Algorithmus für maschinelles Lernen zu entwickeln, der die Überlebensrate von Fahrgästen auf der Grundlage des Trainingsdatensatzes train.csv vorhersagen kann, bitten wir LLM, dieses Problem für uns zu lösen:

> Erstellen Sie auf der Grundlage des Trainingsdatensatzes der Titanic-Passagiere ein maschinelles Lernmodell zur Vorhersage der Überlebensfähigkeit. ⏎

Die Antwort von LLM:

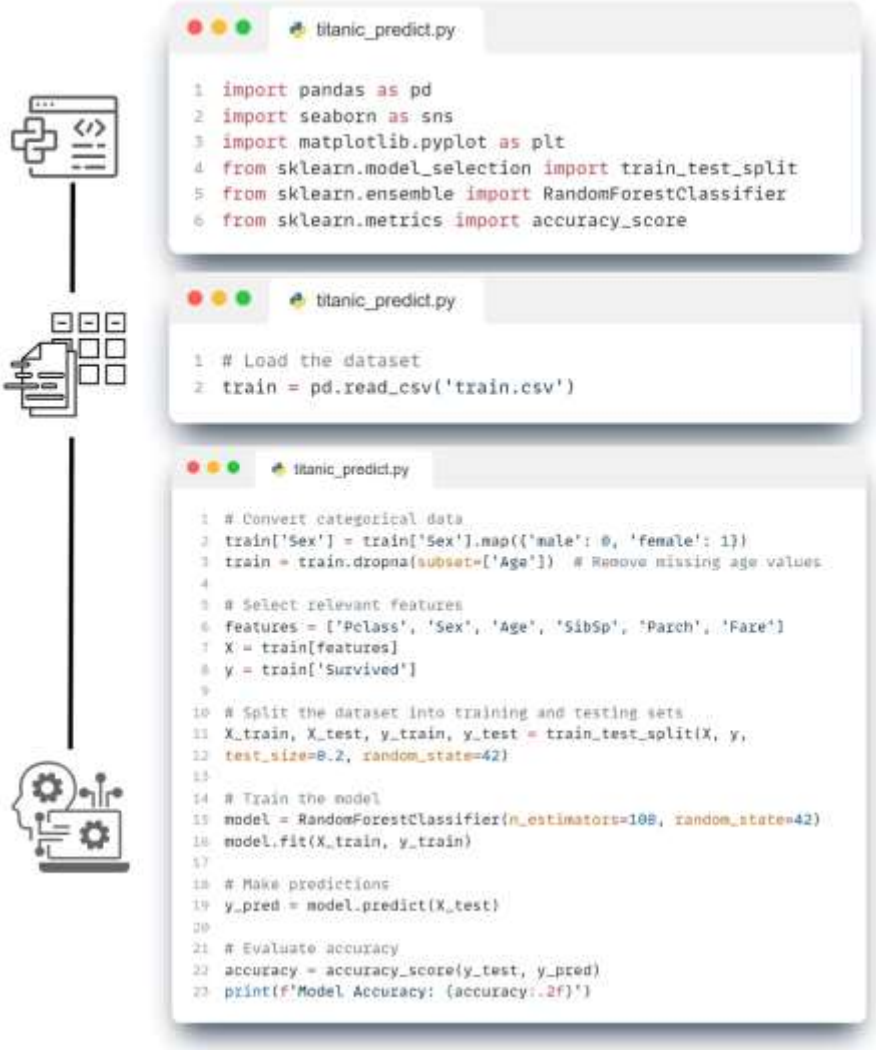


Abb. 9.2-10 LLM erstellte eine Vorhersage der Überlebenden der Titanic mit Hilfe des Algorithmus für maschinelles Lernen Random Forest.

Der resultierende Code von LLM (Abb. 9.2-10) lädt die Daten der Titanic-Passagiere, bereinigt sie, konvertiert kategorische Variablen (z. B. das Geschlecht in ein numerisches Format) und trainiert das Modell mit dem RandomForestClassifier-Algorithmus, um vorherzusagen, ob ein Passagier überlebt hat oder nicht (wir werden in den folgenden Kapiteln mehr über beliebte Algorithmen sprechen).

Der Code trennt die Trainingsdaten in Trainings- und Testdatensätze auf (auf der Kaggle-Website wurden bereits vorgefertigte test.csv (Abb. 9.2-7) und train.csv (Abb. 9.2-6) für das Training erstellt), dann wird das Modell anhand der Trainingsdaten trainiert und anhand der Testdaten getestet, um zu sehen, wie gut ein bestimmtes Vorhersagemodell ist. Nach dem Training werden die Testdaten aus test.csv (mit realen Daten über die Überlebenden und Nicht-Überlebenden) in das Modell eingespeist, und es sagt voraus, wer überlebt hat und wer nicht. In unserem Fall liegt die Genauigkeit des maschinellen Lernmodells, das wir erhalten haben, bei etwa 80%, was zeigt, dass es die Muster recht gut erfasst.

> Das maschinelle Lernen kann mit einem Kind verglichen werden, das versucht, einen rechteckigen Klotz in ein rundes Loch zu stecken. In der Anfangsphase probiert der Algorithmus viele Ansätze aus und stößt dabei auf Fehler und Unstimmigkeiten. Dieser Prozess mag ineffizient erscheinen, aber er bietet wichtige Lernmöglichkeiten: Durch die Analyse jedes Fehlers verbessert das Modell seine Vorhersagen und trifft immer genauere Entscheidungen.

Dieses Modell (Abb. 9.2-10) kann nun verwendet werden, um die Überlebensrate neuer Passagiere vorherzusagen. Wenn Sie es zum Beispiel mit Passagierinformationen füttern, indem Sie die Funktion model.predict Parametern: "männlich", "3. Klasse", "25 Jahre alt", "keine Verwandten an Bord", wird das Modell eine Vorhersage treffen - dass der Passagier mit 80%iger Wahrscheinlichkeit die Katastrophe nicht überleben wird, wenn er 1912 auf der Titanic war (Abb. 9.2-11).

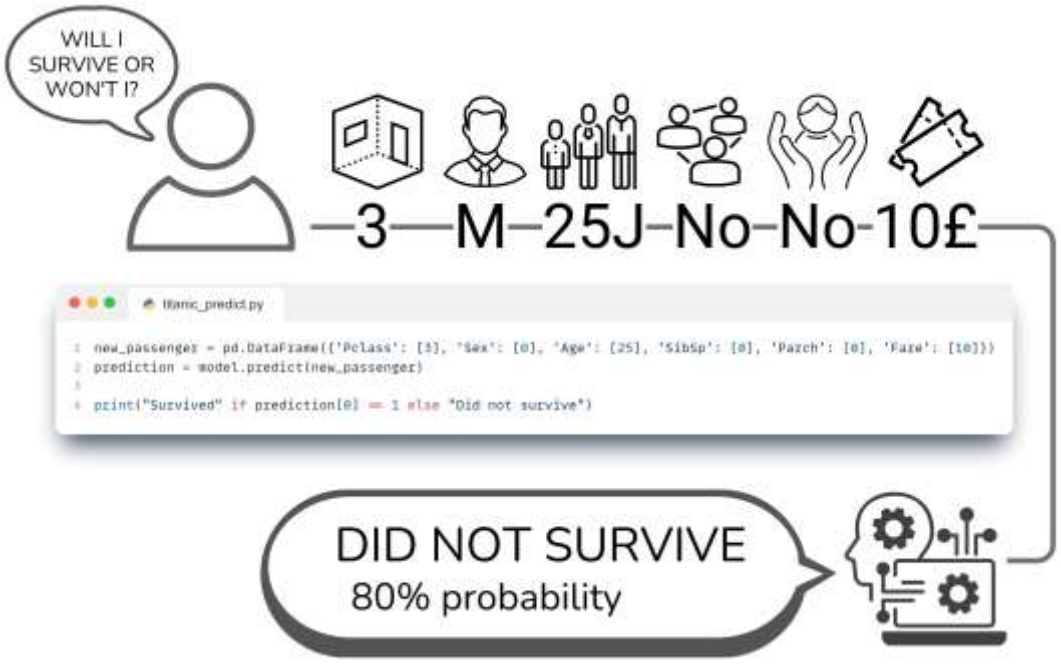


Abbildung 9.2-11 Das oben erstellte Modell kann nun mit einer Wahrscheinlichkeit von 80% vorhersagen, ob ein neuer Passagier der Titanic überleben wird oder nicht.

Das Modell zur Vorhersage des Überlebens der Titanic-Passagiere veranschaulicht ein viel umfassenderes Konzept: Tausende von Fachleuten im Baugewerbe treffen täglich ähnliche "duale" Entscheidungen - über Leben oder Tod einer Entscheidung, eines Projekts, eines Kostenvoranschlags, eines Werkzeugs, Gewinn oder Verlust, Sicherheit oder Risiko. Wie im Beispiel der Titanic, wo das Ergebnis von Faktoren (Geschlecht,

Alter, Klasse) abhing, wird im Baugewerbe jeder Aspekt der Entscheidung von vielen eigenen Faktoren und Variablen (Tabellenspalten) beeinflusst: Materialkosten, Qualifikation der Arbeiter, Fristen, Wetter, Logistik, technische Risiken, Kommentare und Hunderttausende anderer Parameter.

> In der Baubranche folgt das maschinelle Lernen den gleichen Grundsätzen wie in anderen Bereichen: Modelle werden auf historischen Daten - aus Projekten, Verträgen, Kostenvoranschlägen - trainiert, um verschiedene Hypothesen zu testen und die effektivsten Lösungen zu finden. Dieser Prozess ist vergleichbar mit dem Lernen eines Kindes durch Versuch und Irrtum: Mit jedem Zyklus passen sich die Modelle an und werden genauer.

Die Nutzung gesammelter Daten eröffnet neue Horizonte für das Bauwesen. Anstelle arbeitsintensiver manueller Berechnungen können Modelle trainiert werden, die wichtige Merkmale künftiger Projekte mit hoher Genauigkeit vorhersagen können. Auf diese Weise verwandelt Predictive Analytics die Baubranche in einen Raum, in dem man nicht nur planen, sondern auch Entwicklungen sicher vorhersagen kann.

Vorhersagen und Prognosen auf der Grundlage historischer Daten

Die über die Projekte des Unternehmens gesammelten Daten eröffnen die Möglichkeit, Modelle zu erstellen, die in der Lage sind, die Kosten- und Zeitmerkmale zukünftiger, noch nicht realisierter Objekte vorherzusagen - ohne zeitaufwändige manuelle Berechnungen und Vergleiche. Auf diese Weise können Bewertungsprozesse erheblich beschleunigt und vereinfacht werden, da sie sich nicht auf subjektive Annahmen, sondern auf solide mathematische Prognosen stützen.

Zuvor, im vierten Teil des Buches, haben wir uns eingehend mit den traditionellen Methoden der Projektkostenschätzung befasst, einschließlich der ressourcenbasierten Methode, und auch parametrische und Expertenansätze erwähnt. Diese Methoden sind nach wie vor relevant, aber in der modernen Praxis werden sie allmählich mit Werkzeugen der statistischen Analyse und des maschinellen Lernens angereichert, die die Genauigkeit und Reproduzierbarkeit der Schätzungen erheblich verbessern können.

Die Prozesse der manuellen und halbautomatischen Berechnung von Preisen und Zeitattributen werden in Zukunft durch die Meinungen und Vorhersagen von ML-Modellen ergänzt, die in der Lage sind, historische Daten zu analysieren, versteckte Muster zu finden und fundierte Entscheidungen vorzuschlagen. Neue Daten und Szenarien werden automatisch aus bereits verfügbaren Informationen generiert - ähnlich wie Sprachmodelle (LLM) Texte, Bilder und Code auf der Grundlage von Daten erstellen, die über Jahre hinweg aus offenen Quellen gesammelt wurden [158].

> So wie sich der Mensch heute auf Erfahrung, Intuition und interne Statistiken verlässt, um künftige Ereignisse einzuschätzen, wird die Zukunft von Bauprojekten in den kommenden Jahren zunehmend durch eine Kombination aus gesammeltem Wissen und mathematischen Modellen des maschinellen Lernens bestimmt werden.

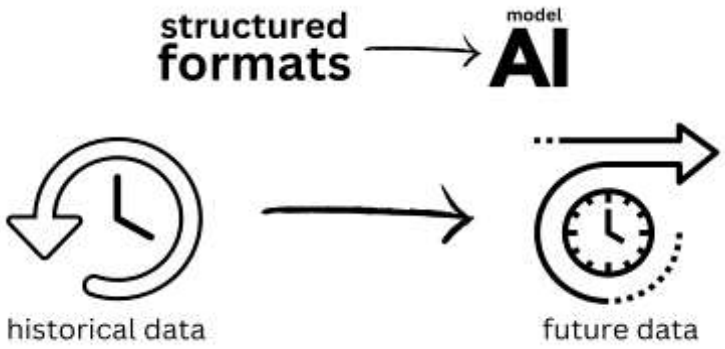


Abb. 9.2-12 Qualitative und strukturierte historische Unternehmensdaten sind das Material, auf dem maschinelle Lernmodelle und Vorhersagen aufgebaut werden.

Betrachten wir ein einfaches Beispiel: die Vorhersage des Preises eines Hauses auf der Grundlage seiner Fläche, Grundstücksgröße, Anzahl der Zimmer und geografischen Lage. Ein Ansatz besteht darin, ein klassisches Modell zu erstellen, das diese Parameter analysiert und den erwarteten Preis berechnet (Abb. 9.2-13). Dieser Ansatz erfordert eine genaue und bekannte Formel im Voraus, was in der Praxis praktisch unmöglich ist.

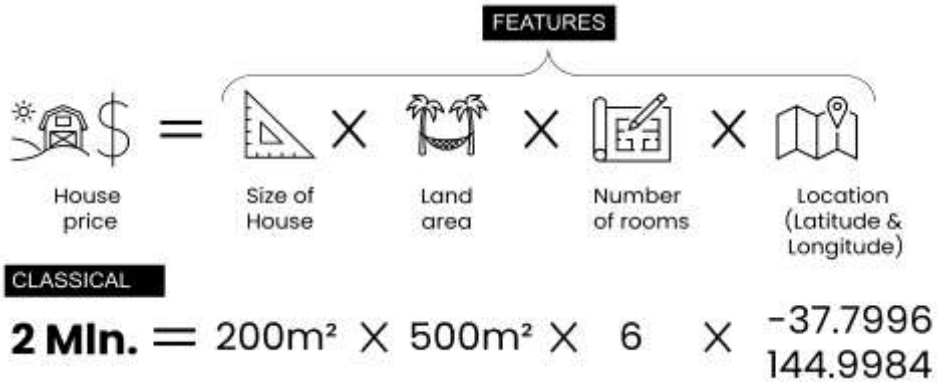


Abb. 9.2-13 Ein klassischer Algorithmus kann verwendet werden, um den Wert eines Hauses zu schätzen, wobei eine feste Formel gefunden werden muss.

Das maschinelle Lernen macht die manuelle Suche nach Formeln überflüssig und ersetzt sie durch trainierte Algorithmen, die unabhängig voneinander Abhängigkeiten erkennen und um ein Vielfaches genauer sind als vorgegebene Gleichungen. Alternativ dazu können wir einen Algorithmus für maschinelles Lernen erstellen, der ein Modell auf der Grundlage eines vorherigen Verständnisses des Problems und historischer Daten, die möglicherweise unvollständig sind, erstellt (Abb. 9.2-14).

> Am Beispiel der Preisgestaltung ermöglicht das maschinelle Lernen die Erstellung verschiedener Arten von mathematischen Modellen, die keine Kenntnis des genauen Mechanismus der Kostenbildung erfordern. Das Modell "lernt" aus den Daten früherer Projekte und passt sich an reale Muster zwischen Bauparametern, Kosten und Terminen an.

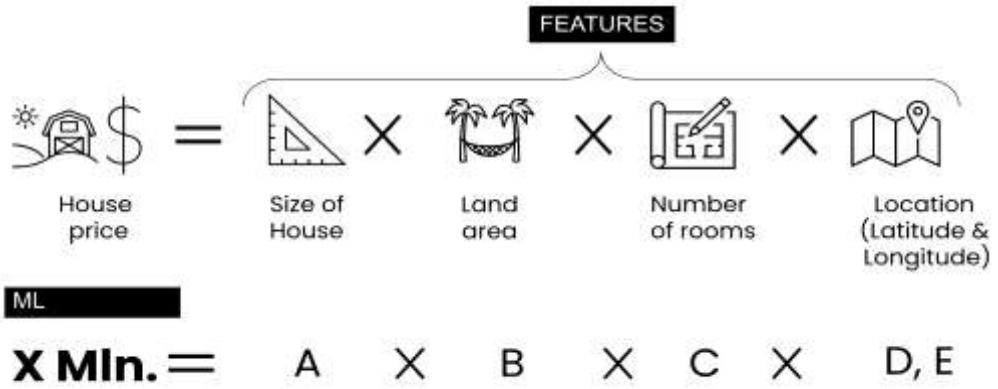


Abb. 9.2-14 Im Gegensatz zur klassischen formelbasierten Schätzung wird der Algorithmus für maschinelles Lernen auf historischen Daten trainiert.

Im Rahmen des überwachten maschinellen Lernens enthält jedes Projekt im Trainingsdatensatz sowohl Input-Attribute (z. B. Kosten und Zeit ähnlicher Gebäude) als auch erwartete Output-Werte (z. B. Kosten oder Zeit). Ein ähnlicher Datensatz wird verwendet, um ein maschinelles Lernmodell zu erstellen und anzupassen (Abb. 9.2-15). Je größer der Datensatz und je höher die Qualität der darin enthaltenen Daten ist, desto genauer wird das Modell und desto genauer werden die Vorhersageergebnisse sein.

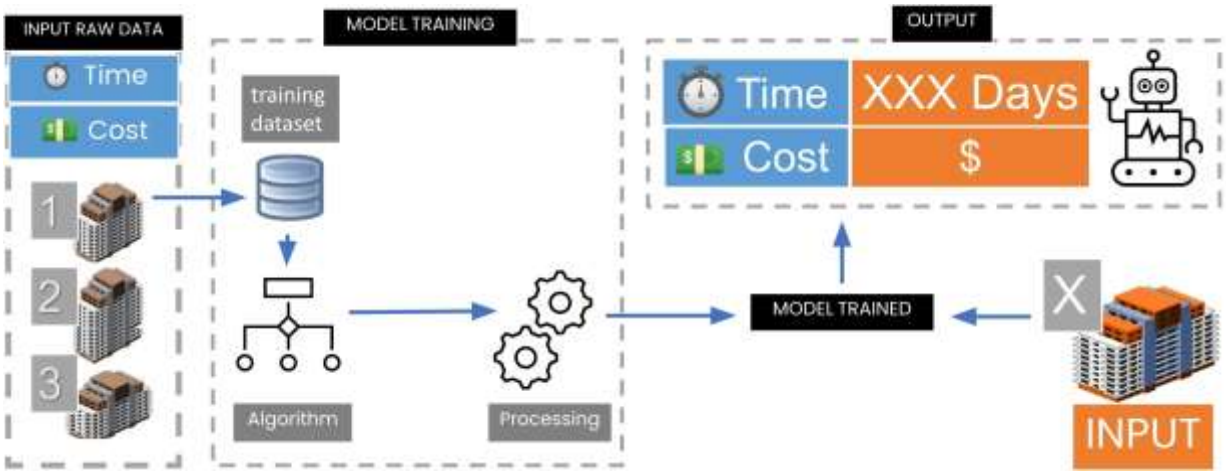


Abbildung 9.2-15 Ein ML-Modell, das auf Kosten- und Zeitplandaten vergangener Projekte trainiert wurde, bestimmt die Kosten und den Zeitplan eines neuen Projekts mit einer bestimmten Wahrscheinlichkeit.

Sobald das Modell erstellt und für die Schätzung des Baus eines neuen Projekts trainiert ist, geben Sie dem Modell einfach neue Attribute für das neue Projekt vor, und das Modell wird mit einer gewissen Wahrscheinlichkeit Schätzergebnisse auf der Grundlage der zuvor erlernten Muster liefern.

Schlüsselkonzepte des maschinellen Lernens

Maschinelles Lernen ist keine Magie, es geht nur um Mathematik, Daten und das Finden von Mustern. Es

hat keine wirkliche Intelligenz, sondern ist ein Programm, das auf Daten trainiert wird, um Muster zu erkennen und Entscheidungen ohne ständige menschliche Beteiligung zu treffen.

Das maschinelle Lernen verwendet eine Reihe von Schlüsselkonzepten zur Beschreibung seiner Struktur (Abb. 9.2-16):

- **Labels** sind Zielvariablen oder Attribute (der Parameter "Survivor" im Titanic-Datensatz), die das Modell vorhersagen soll. Beispiel: Baukosten (z. B. in Dollar), Dauer der Bauarbeiten (z. B. in Monaten).
- **Merkmale** sind unabhängige Variablen oder Attribute, die als Input für das Modell dienen. In einem Prognosemodell werden sie zur Vorhersage von Kennzeichnungen verwendet. Beispiele: Grundstücksgröße (in Quadratmetern), Anzahl der Stockwerke eines Gebäudes, Gesamtgeschossfläche eines Gebäudes (in Quadratmetern), geografischer Standort (Breiten- und Längengrad), Art der beim Bau verwendeten Materialien. Die Anzahl der Merkmale bestimmt auch die Dimensionalität der Daten.
- **Ein Modell** ist ein Satz verschiedener Hypothesen, von denen eine die zu prognostizierende oder zu approximierende Zielfunktion annähert. Beispiel: Modell des maschinellen Lernens, das Regressionsanalysetechniken zur Vorhersage der Kosten und des Zeitplans von Bauvorhaben verwendet.
- **Lernalgorithmus Ein Lernalgorithmus** ist der Prozess der Suche nach der besten Hypothese in einem Modell, die genau der Zielfunktion entspricht, unter Verwendung eines Satzes von Trainingsdaten. Beispiel: Ein linearer Regressions-, KNN- oder Random-Forest-Algorithmus, der Kosten- und Bauzeitdaten analysiert, um Beziehungen und Muster zu erkennen.
- **Training** - Während des Trainingsprozesses analysiert der Algorithmus die Trainingsdaten und findet Muster, die der Beziehung zwischen den Eingabeattributen und den Zielbezeichnungen entsprechen. Das Ergebnis dieses Prozesses ist ein trainiertes maschinelles Lernmodell, das zur Vorhersage bereit ist. Beispiel: Ein Prozess, bei dem ein Algorithmus historische Baudaten (Kosten, Zeit, Anlagenmerkmale) analysiert, um ein Vorhersagemodell zu erstellen.

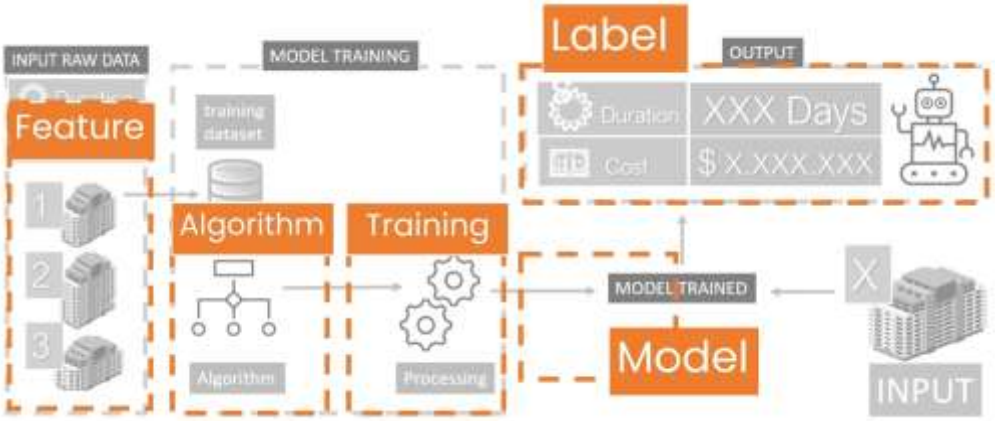


Abbildung 9.2-16 ML verwendet Kennzeichnungen und Attribute, um Modelle zu erstellen, die mithilfe von Algorithmen auf Daten trainiert werden, um Ergebnisse vorherzusagen.

Maschinelles Lernen existiert nicht isoliert, sondern ist Teil eines breiteren Ökosystems analytischer Diszip-

linen wie Statistik, Datenbanken, Data Mining, Mustererkennung, Big-Data-Analytik und künstliche Intelligenz. Abbildung 9.2-17 zeigt, wie sich diese Bereiche überschneiden und gegenseitig ergänzen und so einen umfassenden Rahmen für moderne Entscheidungs- und Automatisierungssysteme bieten.

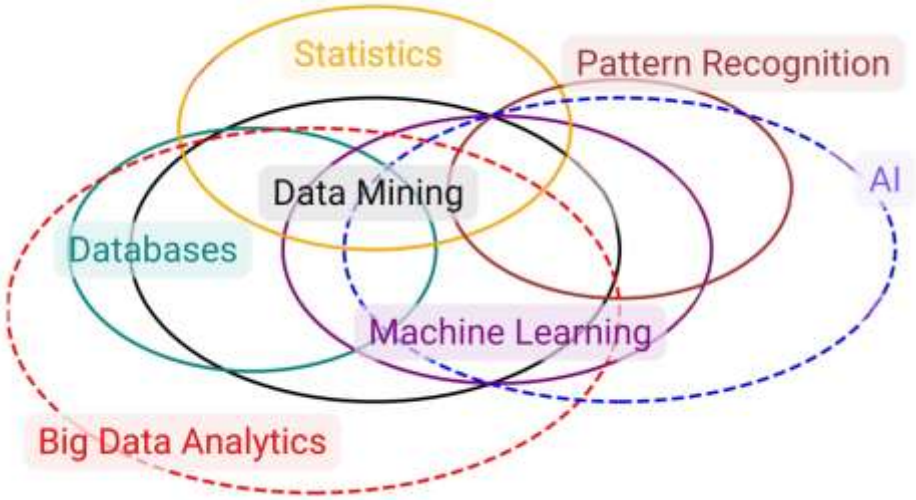


Abb. 9.2-17 Die Beziehung zwischen den verschiedenen Bereichen der Datenanalyse: Statistik, maschinelles Lernen, künstliche Intelligenz, Big Data, Mustererkennung und Data Mining.

> Das Hauptziel des maschinellen Lernens besteht darin, Computer mit der Fähigkeit auszustatten, automatisch und ohne menschliches Zutun Wissen zu erwerben und ihre Handlungen entsprechend anzupassen [159].

In Zukunft wird die Rolle des Menschen also nur noch darin bestehen, die Maschine mit kognitiven Fähigkeiten auszustatten - er wird die Bedingungen, Gewichte und Parameter festlegen, und das maschinelle Lernmodell wird den Rest erledigen.

Im nächsten Kapitel werden wir konkrete Beispiele für die Anwendung des Algorithmus betrachten. Anhand von realen Tabellen und vereinfachten Modellen wird gezeigt, wie die Prognose Schritt für Schritt aufgebaut wird.

KAPITEL 9.3.
VORHERSAGE VON KOSTEN UND TERMINEN DURCH MASCHINELLES LERNEN

Ein Beispiel für den Einsatz von maschinellem Lernen zur Ermittlung von Projektkosten und -zeitplänen

Die Schätzung von Bauzeit und -kosten ist einer der wichtigsten Prozesse in der Tätigkeit eines Bauunternehmens. Traditionell wurden solche Schätzungen von Experten auf der Grundlage von Erfahrungen, Nachschlagewerken und gesetzlichen Datenbanken vorgenommen. Mit der digitalen Transformation und der zunehmenden Datenverfügbarkeit ist es nun jedoch möglich, Modelle des maschinellen Lernens (ML) einzusetzen, um die Genauigkeit und Automatisierung solcher Schätzungen zu verbessern.

Die Einführung von maschinellem Lernen in den Prozess der Kosten- und Zeitberechnung von Bauvorhaben ermöglicht nicht nur eine effizientere Planung, sondern wird auch zum Ausgangspunkt für die Integration intelligenter Modelle in andere Geschäftsprozesse - vom Risikomanagement bis zur Optimierung von Logistik und Beschaffung.

> Es ist wichtig, schnell feststellen zu können, wie lange der Bau eines Projekts dauern wird und wie hoch die Gesamtkosten sein werden. Diese Fragen zur Projektdauer und zu den Kosten stehen seit den Anfängen des Baugewerbes traditionell im Vordergrund der Überlegungen sowohl von Bauherren als auch von Bauunternehmen.

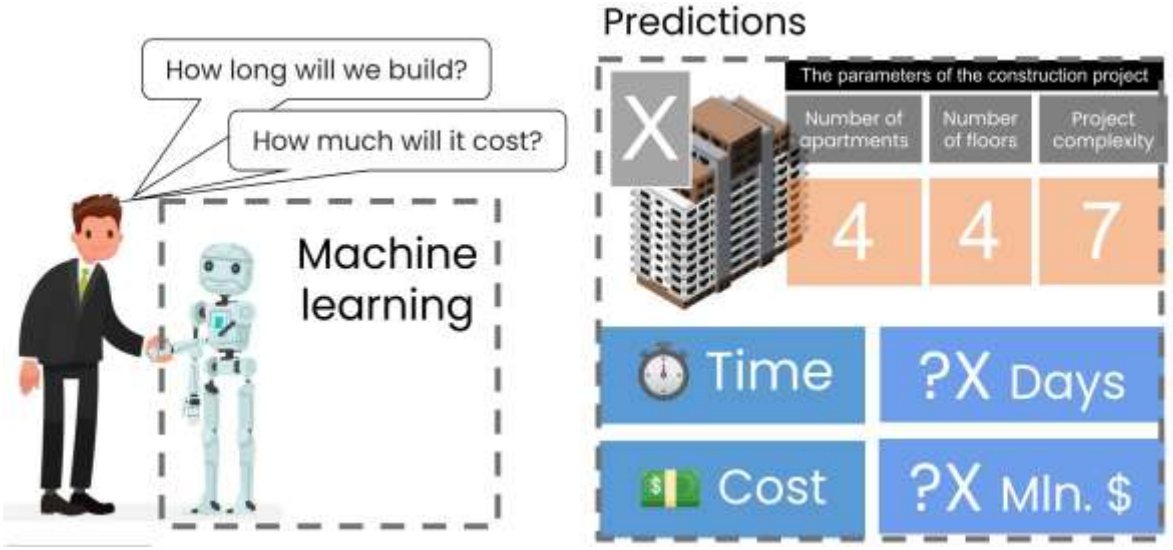


Abb. 9.3-1 Bei Bauprojekten sind die Schnelligkeit und Qualität der Bauzeit- und Kostenschätzung entscheidende Erfolgsfaktoren.

Im folgenden Beispiel werden Schlüsseldaten aus vergangenen Projekten extrahiert und zur Entwicklung eines maschinellen Lernmodells verwendet, das es uns ermöglicht, das Modell zur Schätzung der Kosten und des Zeitplans für neue Bauprojekte mit neuen Parametern zu verwenden (Abb. 9.3-1).

Betrachten wir drei Projekte mit drei Schlüsselattributen: die Anzahl der Wohnungen (wobei 100 Wohnungen der Einfachheit halber der Zahl 10 entsprechen), die Anzahl der Stockwerke und ein bedingtes Maß für die Komplexität des Baus auf einer Skala von 1 bis 10, wobei 10 der höchste Komplexitätswert ist. Beim maschinellen Lernen wird der Prozess der Umwandlung und Vereinfachung von Werten wie 100 zu 10 oder 50 zu 5 als "Normalisierung" bezeichnet.

> Beim maschinellen Lernen werden unterschiedliche numerische Daten auf eine gemeinsame Skala gebracht, um die Verarbeitung und Analyse zu erleichtern. Dieser Prozess ist besonders wichtig, wenn die Daten unterschiedliche Maßstäbe und Maßeinheiten haben.

Angenommen, das erste Projekt (Abb. 9.3-2) hatte 50 Wohnungen (nach Normalisierung 5), 7 Stockwerke und eine Komplexitätsbewertung von 2, was eine relativ einfache Konstruktion bedeutete. Das zweite Projekt hatte bereits 80 Wohnungen, 9 Stockwerke und einen relativ komplexen Entwurf. Unter diesen Bedingungen dauerte der Bau des ersten und zweiten Wohngebäudes 270 bzw. 330 Tage, und die Gesamtkosten des Projekts beliefen sich auf 4,5 bzw. 5,8 Mio. $.

Abbildung 9.3-2 Ein Beispiel für eine Reihe vergangener Projekte, die zur Schätzung von Zeit und Kosten für das zukünftige Projekt X verwendet werden sollen.

Beim Aufbau eines maschinellen Lernmodells für solche Daten besteht die Hauptaufgabe darin, kritische Attribute (oder Labels) für die Vorhersage zu identifizieren, in diesem Fall die Bauzeit und die Kosten. Mit einem kleinen Datensatz werden wir Informationen über frühere Bauprojekte verwenden, um neue Projekte zu planen: Mit Hilfe von Algorithmen des maschinellen Lernens müssen wir die Kosten und die Bauzeit eines neuen Projekts X auf der Grundlage gegebener Attribute des neuen Projekts vorhersagen, z. B. 40 Wohnungen, 4 Stockwerke und eine relativ hohe Projektkomplexität von 7 (Abb. 9.3-2). In einer realen Umgebung kann die Anzahl der Eingabeparameter viel größer sein und von einigen Dutzend bis zu Hunderten von Faktoren reichen. Dazu können gehören: Art der Baumaterialien, Klimazone, Qualifikationsniveau der Bauunternehmer, Verfügbarkeit von Versorgungseinrichtungen, Art des Fundaments, Zeitpunkt des Baubeginns, Kommentare der Vorarbeiter usw.

> Um ein prädiktives Modell für maschinelles Lernen zu erstellen, müssen wir einen Algorithmus wählen, der es erstellt. Ein Algorithmus beim maschinellen Lernen ist wie ein mathematisches Rezept, das dem Computer beibringt, wie er Vorhersagen treffen (in der richtigen Reihenfolge der Parameter mischen) oder Entscheidungen auf der Grundlage von Daten treffen kann.

Zur Analyse von Daten über vergangene Bauprojekte und zur Vorhersage von Zeit und Kosten zukünftiger Projekte (Abb. 9.3-2) kann ein beliebter Algorithmus für maschinelles Lernen verwendet werden:

- **Lineare Regression (Lineare Regression):** Dieser Algorithmus versucht, eine direkte Beziehung zwischen Attributen zu finden, zum Beispiel zwischen der Anzahl der Stockwerke und den Baukosten. Ziel des Algorithmus ist es, eine lineare Gleichung zu finden, die diese Beziehung am besten beschreibt und die Vorhersagen ermöglicht.
- **Algorithmus k-nearest neighbours (k-NN):** Dieser Algorithmus vergleicht ein neues Projekt mit früheren Projekten, die in Größe oder Komplexität ähnlich waren. Der k-NN klassifiziert die Daten auf der Grundlage, welche der k (Anzahl) Trainingsbeispiele ihnen am nächsten sind. Im Kontext der Regression ist das Ergebnis der Mittelwert oder Median der k nächsten Nachbarn.
- **Entscheidungsbäume**: Ein Vorhersagemodell, das Daten anhand einer Baumstruktur in Teilmengen unterteilt, die auf verschiedenen Bedingungen basieren. Jeder Knoten des Baums steht für eine Bedingung oder Frage, die zu einer weiteren Unterteilung der Daten führt, und jedes Blatt steht für die endgültige Vorhersage oder das Ergebnis. Der Algorithmus unterteilt die Daten in kleinere Gruppen auf der Grundlage verschiedener Merkmale, z. B. zunächst nach der Anzahl der Geschichten, dann nach der Komplexität und so weiter, um eine Vorhersage zu treffen.

Werfen wir einen Blick auf die Algorithmen des maschinellen Lernens zur Schätzung der Kosten eines neuen Projekts am Beispiel von zwei beliebten Algorithmen: der linearen Regression und dem K-nearest neighbours-Algorithmus.

Vorhersage von Projektkosten und -zeiten mittels linearer Regression

Die lineare Regression ist ein grundlegender Algorithmus zur Datenanalyse, der den Wert einer Variablen auf der Grundlage einer linearen Beziehung zu einer oder mehreren anderen Variablen vorhersagt. Dieses Modell geht davon aus, dass eine direkte lineare Beziehung zwischen der abhängigen Variablen und einer oder mehreren unabhängigen Variablen besteht, und das Ziel des Algorithmus ist es, diese Beziehung zu finden.

> Die Einfachheit und Klarheit der linearen Regression hat sie zu einem beliebten Werkzeug in einer Vielzahl von Bereichen gemacht. Bei einer einzelnen Variablen geht es bei der linearen Regression darum, die am besten passende Linie durch die Datenpunkte zu finden.

Die lineare Regression findet die beste gerade Linie (rote Linie), die die Abhängigkeit zwischen der Eingangsvariablen X und der Ausgangsvariablen Y annähert. Diese Linie ermöglicht die Vorhersage von Y-Werten für neue X-Werte auf der Grundlage der ermittelten linearen Beziehung (Abb. 9.3-3).

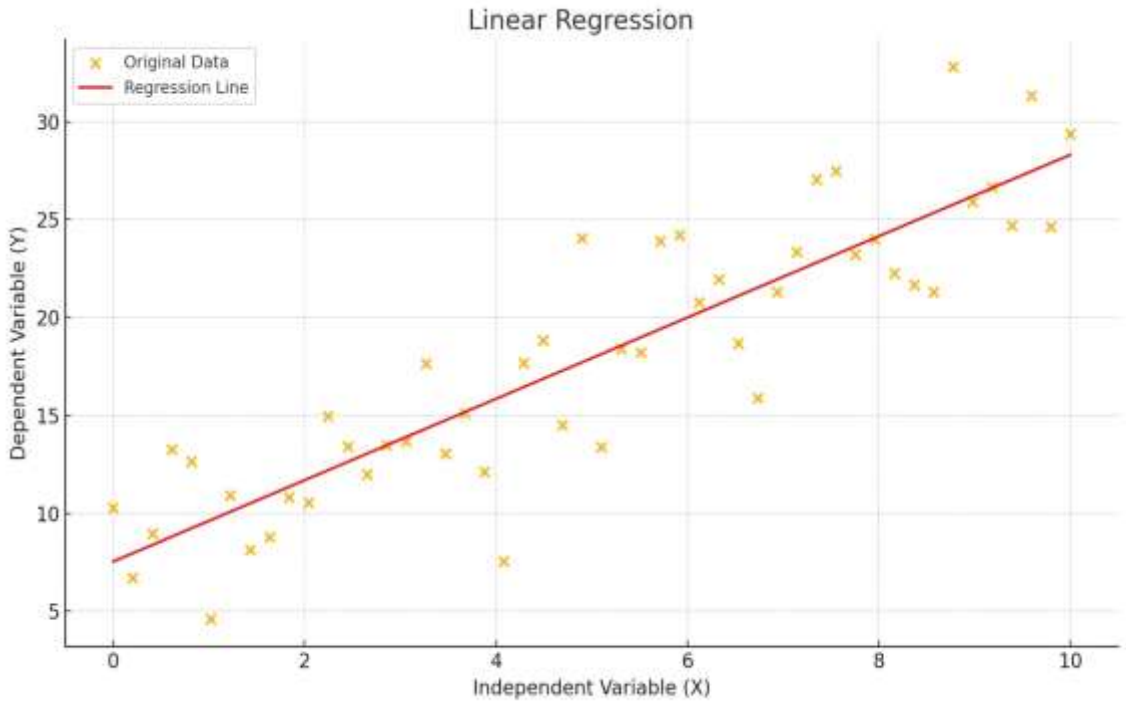


Abb. 9.3-3 Das Prinzip der linearen Regression besteht darin, die beste gerade Linie durch die Trainingswerte zu finden.

Diese Linie wird durch eine Gleichung dargestellt, bei der die Eingabe eines Wertes der unabhängigen Variablen (X) einen vorhergesagten Wert der abhängigen Variablen (Y) ergibt. Dieses Verfahren ermöglicht eine effiziente Vorhersage von Y aus bekannten Werten von X unter Verwendung einer linearen Beziehung zwischen ihnen. Ein Beispiel für die Ermittlung einer solchen statistisch gemittelten Linie ist in der Bewertung der Baugenehmigungsdaten von San Francisco zu sehen (Abb. 9.1-7), wo die Inflation mithilfe der linearen Regression für verschiedene Arten von Einrichtungen berechnet wurde.

Lassen Sie uns die Projektdatentabelle (Abb. 9.3-2 aus dem vorherigen Kapitel) direkt in den LLM laden und ihn bitten, ein einfaches maschinelles Lernmodell für uns zu erstellen.

- Senden Sie eine Textanfrage an LLM chat (CHATGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

> Wir müssen die Konstruktion eines einfachen maschinellen Lernmodells zeigen, um die Kosten und die Zeit für die Realisierung eines neuen Projekts X vorherzusagen (Abb. 9.3-2 als beigefügtes Bild) ⏎

Die Antwort von LLM:

Create in LLM Chat

Linear regression.py

```
1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost
```

Run in IDE

Result

```
(array([238.44444444]), array([3042337.77777778]))
```

Abbildung 9.3-4 LLM wählte die lineare Regression, um ein maschinelles Lernmodell zur Vorhersage von Projektkosten und -zeiten zu erstellen.

Der LLM erkannte automatisch die Tabelle aus dem angehängten Bild und konvertierte die Daten von einem visuellen Format in ein Tabellenfeld (Abb. 9.3-4 - Zeile 6). Dieses Array wurde als Grundlage für die Erstellung von Merkmalen und Beschriftungen verwendet, aus denen ein Modell für maschinelles Lernen erstellt wurde (Abb. 9.3-4 - 17. bis 22. Zeile), das eine lineare Regression verwendete.

Unter Verwendung eines einfachen linearen Regressionsmodells, das auf einem "extrem kleinen" Datensatz trainiert wurde, wurden Vorhersagen für ein neues hypothetisches Bauprojekt mit der Bezeichnung Projekt

X gemacht. In unserem Problem ist dieses Projekt durch 40 Wohnungen, 4 Stockwerke und eine Komplexitätsstufe von 7 gekennzeichnet (Abb. 9.3-2).

Wie durch ein lineares Regressionsmodell auf der Grundlage eines begrenzten und kleinen Datensatzes für das neue Projekt X vorhergesagt (Abb. 9.3-4 - Zeile 24-29):

- **Die Bauzeit** wird etwa 238 Tage betragen (238.4444444)
- Die **Gesamtausgaben** belaufen sich auf etwa$ 3.042.338 (3042337,777)

Um die Projektkostenhypothese weiter zu untersuchen, ist es sinnvoll, mit verschiedenen Algorithmen und Methoden des maschinellen Lernens zu experimentieren. Lassen Sie uns daher dieselben Kosten- und Zeitwerte für ein neues Projekt X auf der Grundlage eines kleinen Satzes historischer Daten mit Hilfe des Algorithmus K-Nearest Neighbours (k-NN) vorhersagen.

Vorhersage von Projektkosten und -zeiten mit dem K-Nächste-Nachbar-Algorithmus (k-NN)

Wir verwenden den k-Nächste-Nachbarn-Algorithmus (k-NN) als zusätzlichen Prädiktor zur Schätzung der Kosten und der Dauer eines neuen Projekts. Der K-Nearest Neighbours (k-NN)-Algorithmus ist ein überwachtes maschinelles Lernverfahren (supervised machine learning) sowohl für die Klassifizierung als auch für die Regression. Wir haben den k-NN-Algorithmus auch schon im Zusammenhang mit der Suche in Vektordatenbanken (Abb. 8.2-2) erörtert, wo er zum Auffinden der nächstgelegenen Vektoren (z. B. Texte, Bilder oder technische Beschreibungen) verwendet wird. Bei diesem Ansatz wird jedes Projekt als ein Punkt in einem mehrdimensionalen Raum dargestellt, wobei jede Dimension einem bestimmten Attribut des Projekts entspricht.

In unserem Fall werden wir die drei Attribute eines jeden Projekts als Punkte in einem dreidimensionalen Raum darstellen (Abb. 9.3-5). So wird unser nächstes Projekt X in diesem Raum mit den Koordinaten (x=4, y=4, z=7) lokalisiert. Es sei darauf hingewiesen, dass unter realen Bedingungen die Anzahl der Punkte und die Dimensionalität des Raums um Größenordnungen größer sein können.

> Der K-NN-Algorithmus (k-nearest neighbours) arbeitet mit der Messung des Abstands zwischen dem gewünschten Projekt X und den Projekten in der Trainingsdatenbank. Durch den Vergleich dieser Abstände ermittelt der Algorithmus die Projekte, die dem Punkt des neuen Projekts X am nächsten liegen.

Wenn zum Beispiel das zweite Projekt (x=8, y=9, z=6) aus unserem ursprünglichen Datensatz viel weiter von X entfernt ist (Abb. 9.3-5) als die anderen Projekte, kann es von der weiteren Analyse ausgeschlossen werden. Folglich können nur die beiden (k=2) nächstgelegenen Projekte für die Berechnungen herangezogen werden, auf deren Grundlage dann der Durchschnittswert ermittelt wird.

Eine solche Methode ermöglicht es, durch eine Nachbarschaftssuche die Ähnlichkeiten zwischen Projekten zu bewerten, was wiederum dazu beiträgt, Schlussfolgerungen über die voraussichtlichen Kosten und den Zeitplan eines neuen Projekts auf der Grundlage ähnlicher, bereits durchgeführter Projekte zu ziehen.

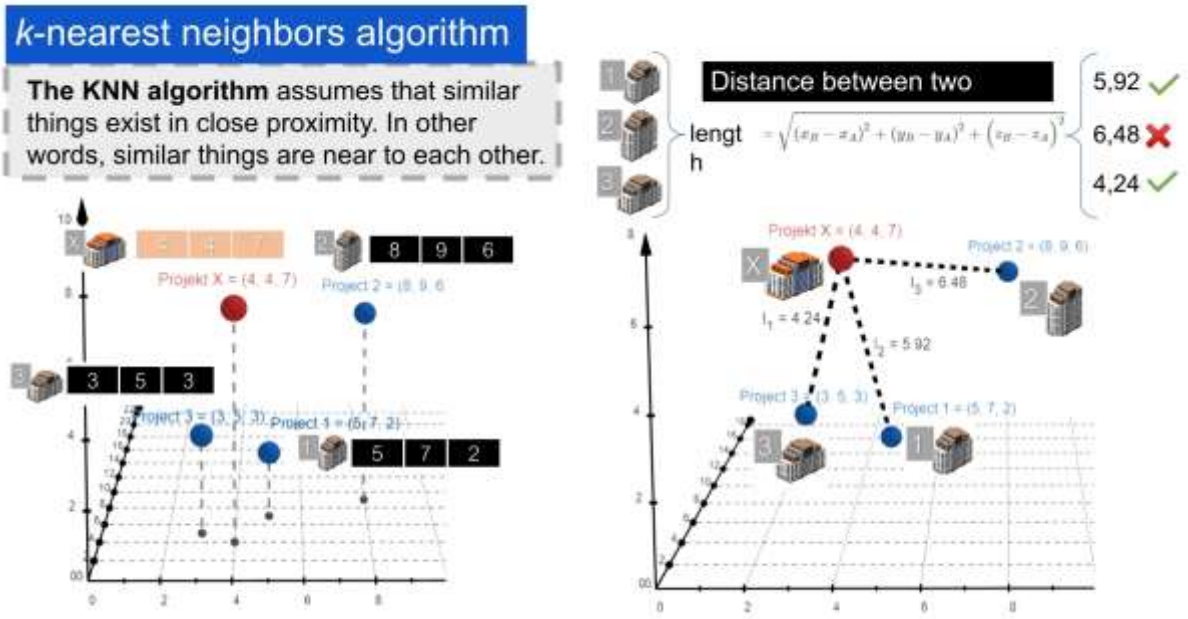


Abb. 9.3-5 Beim K-NN-Algorithmus werden Projekte als Punkte in einem mehrdimensionalen Raum dargestellt, und die nächstgelegenen Projekte werden auf der Grundlage von Entfernungen für die Ähnlichkeitsbewertung und Vorhersage ausgewählt.

Die Arbeit von k-NN umfasst mehrere wichtige Schritte:

- **Datenvorbereitung:** Zuerst werden Trainings- und Testdatensätze geladen. Die Trainingsdaten dienen zum "Trainieren" des Algorithmus, die Testdaten zur Überprüfung seiner Effizienz.
- **Auswahl des Parameters K:** Es wird eine Zahl K gewählt, die angibt, wie viele nächste Nachbarn (Datenpunkte) im Algorithmus berücksichtigt werden sollen. Der Wert von "K" ist sehr wichtig, da er das Ergebnis beeinflusst.
- Klassifizierungsprozess und Regression für Testdaten:
  - **Abstandsberechnung:** Für jedes Element aus den Testdaten wird der Abstand zu jedem Element aus den Trainingsdaten berechnet (Abb. 9.3-5). Hierfür können verschiedene Abstandsmessmethoden verwendet werden, wie z. B. der euklidische Abstand (die gängigste Methode), der Manhattan-Abstand oder der Hamming-Abstand.
  - **Sortierung und Auswahl der K nächsten Nachbarn:** Nach der Berechnung der Abstände werden diese sortiert und K Punkte ausgewählt, die dem Testpunkt am nächsten liegen.
  - **Bestimmung der Klasse oder des Wertes eines Testpunktes: Handelt** es sich um eine Klassifizierungsaufgabe, wird die Klasse des Testpunktes auf der Grundlage der häufigsten Klasse unter K ausgewählten Nachbarn bestimmt. Handelt es sich um eine Regressionsaufgabe, wird der Mittelwert (oder ein anderes Maß für die zentrale Tendenz) der Werte der K Nachbarn berechnet.
- **Abschluss des Prozesses:** Sobald alle Testdaten klassifiziert sind oder Vorhersagen für sie getroffen wurden, ist der Prozess abgeschlossen.

Der Algorithmus k-nearest neighbours (k-NN) hat sich in vielen praktischen Anwendungen bewährt und ist eines der wichtigsten Instrumente im Arsenal der Spezialisten für maschinelles Lernen. Dieser Algorithmus ist aufgrund seiner Einfachheit und Effizienz beliebt, insbesondere bei Aufgaben, bei denen die Beziehungen zwischen den Daten leicht zu interpretieren sind.

In unserem Beispiel wurden nach Anwendung des K-Nächste-Nachbarn-Algorithmus die beiden Projekte (aus unserer kleinen Stichprobe) mit dem geringsten Abstand zu Projekt X ermittelt (Abb. 9.3-5). Auf der Grundlage dieser Projekte bestimmt der Algorithmus den Durchschnitt ihres Preises und ihrer Bauzeit. Nach der Analyse (Abb. 9.3-6) kommt der Algorithmus durch Mittelwertbildung der nächsten Nachbarn zu dem Schluss, dass das Projekt X ungefähr$ 3.800.000 $ kosten wird und etwa 250 Tage für die Fertigstellung benötigt.

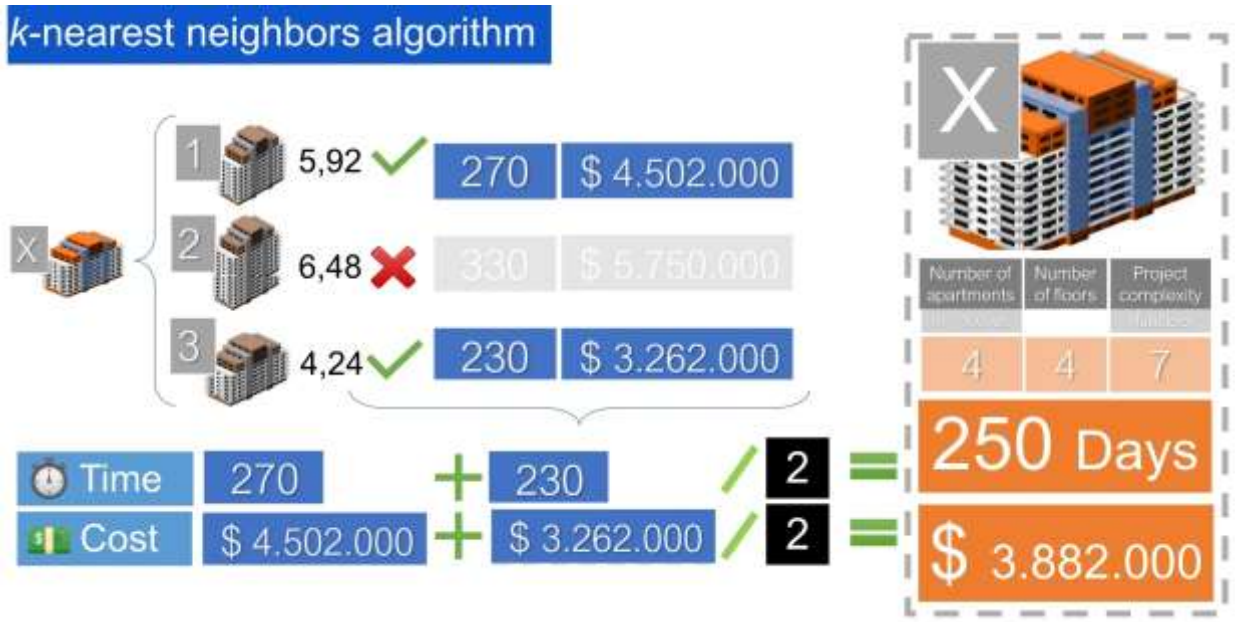


Abbildung 9.3-6 Der Algorithmus "K-Nächste Nachbarn" () ermittelt die Kosten und den Zeitplan von Projekt X, indem er die beiden nächstgelegenen Projekte der Stichprobe analysiert.

Der k-Nearest Neighbors (k-NN) Algorithmus ist besonders beliebt bei Klassifizierungs- und Regressionsaufgaben, wie z.B. bei Empfehlungssystemen, wo er verwendet wird, um Produkte oder Inhalte auf der Grundlage von Präferenzen vorzuschlagen, die den Interessen eines bestimmten Nutzers entsprechen. Darüber hinaus wird der k-NN-Algorithmus häufig in der medizinischen Diagnostik zur Klassifizierung von Arten von Krankheiten auf der Grundlage von Patientensymptomen, in der Mustererkennung und im Finanzsektor zur Bewertung der Kreditwürdigkeit von Kunden eingesetzt.

> Selbst mit begrenzten Daten können Modelle des maschinellen Lernens nützliche Vorhersagen liefern und die analytische Komponente des Bauprojektmanagements erheblich verbessern. Wenn historische Daten erweitert und bereinigt werden, ist es möglich, zu anspruchsvolleren Modellen überzugehen - zum Beispiel unter Berücksichtigung der Art der Konstruktion, des Standorts, der Jahreszeit des Baubeginns und anderer Faktoren.

In unserer vereinfachten Aufgabe wurden drei Attribute für die Visualisierung im 3D-Raum verwendet, aber reale Projekte umfassen im Durchschnitt Hunderte oder Tausende von Attributen (siehe den Datensatz aus dem Kapitel "Ein Beispiel für CAD-basierte Big Data (BIM)"), was die Dimensionalität des Raums und die Komplexität der Darstellung von Projekten als Vektoren erheblich erhöht (Abb. 9.3-7).

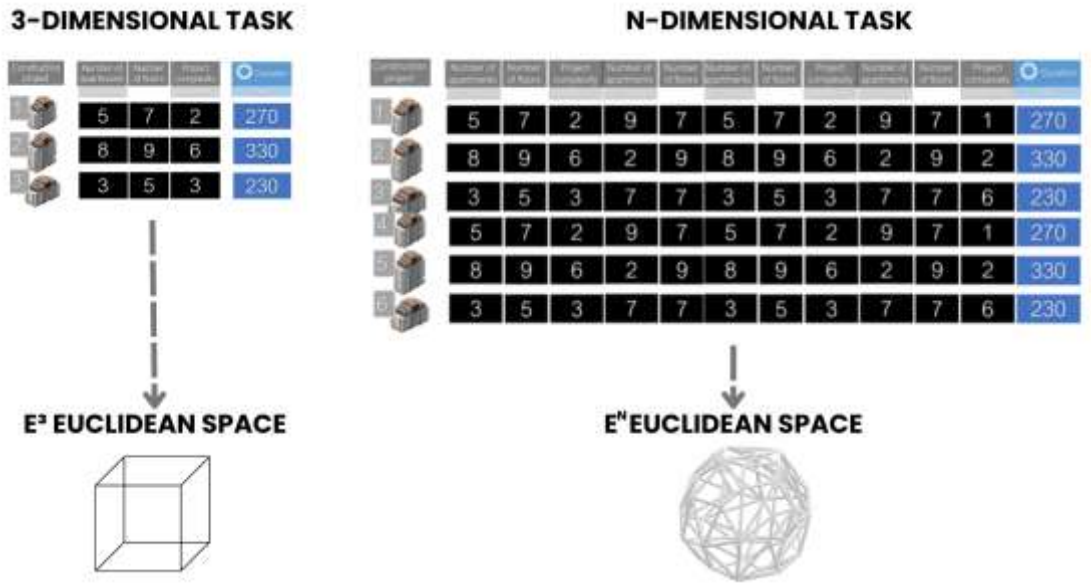


Abbildung 9.3-7 In dem vereinfachten Beispiel wurden drei Attribute für die 3D-Visualisierung verwendet, während reale Projekte mehr haben.

Die Anwendung verschiedener Algorithmen auf denselben Datensatz für das Projekt X, das 40 Wohnungen, 4 Stockwerke und die Komplexitätsstufe 7 umfasst, ergab unterschiedliche Prognosewerte. Der lineare Regressionsalgorithmus sagte eine Fertigstellungszeit von 238 Tagen und Kosten von$ 3.042.338 voraus (Abb. 9.3-4), während der k-NN-Algorithmus 250 Tage und$ 3.882.000 voraussagte (Abb. 9.3-6).

Die Genauigkeit der Vorhersagen, die mit Hilfe von Modellen des maschinellen Lernens erzielt werden, hängt direkt von der Menge und Qualität der Eingabedaten ab. Je mehr Projekte am Training beteiligt sind und je vollständiger und genauer ihre Merkmale (Attribute) und Ergebnisse (Labels) dargestellt werden, desto höher ist die Wahrscheinlichkeit, zuverlässige Vorhersagen mit minimalen Fehlerwerten zu erhalten.

Techniken zur Datenvorverarbeitung spielen in diesem Prozess eine wichtige Rolle:

- Normalisierung, um die Merkmale auf einen gemeinsamen Maßstab zu bringen;
- Erkennung und Eliminierung von Ausreißern und Beseitigung von Modellverzerrungen;
- Kodierung von kategorischen Attributen, um die Bearbeitung von Textdaten zu ermöglichen;

- Auffüllen von fehlenden Werten, um die Robustheit des Modells zu erhöhen.

Darüber hinaus werden Kreuzvalidierungsverfahren eingesetzt, um die Verallgemeinerbarkeit des Modells und seine Robustheit gegenüber neuen Datensätzen zu bewerten, um eine Überanpassung zu erkennen und die Zuverlässigkeit der Vorhersage zu verbessern.

> *Das Chaos ist eine Ordnung, die es zu entschlüsseln gilt [160].*
>
> - José Saramago, "Der Doppelgänger"

Auch wenn Sie den Eindruck haben, dass sich das Chaos Ihrer Aufgaben nicht formal beschreiben lässt, sollten Sie wissen, dass jedes Ereignis in der Welt und insbesondere Bauprozesse mathematischen Gesetzmäßigkeiten unterliegen, die unter Umständen die Unterstützung der Berechnung von Werten nicht durch strenge Formeln, sondern mit Hilfe von Statistiken und historischen Daten benötigen.

Sowohl herkömmliche Schätzungen der Kalkulationsabteilung als auch Modelle des maschinellen Lernens sind zwangsläufig mit Unsicherheiten und potenziellen Fehlerquellen behaftet. Wenn jedoch ausreichend hochwertige Daten zur Verfügung stehen, können maschinelle Lernmodelle eine vergleichbare und manchmal sogar höhere Vorhersagegenauigkeit aufweisen als Expertenurteile.

> Das maschinelle Lernen dürfte sich zu einem zuverlässigen ergänzenden Instrument für Analysen entwickeln, mit dem Berechnungen verfeinert, alternative Szenarien vorgeschlagen und verborgene Abhängigkeiten zwischen Projektparametern ermittelt werden können. Solche Modelle werden nicht den Anspruch erheben, universell zu sein, aber sie werden bald einen wichtigen Platz in Berechnungen und Entscheidungsprozessen einnehmen. Die Technologien des maschinellen Lernens werden die Beteiligung von Ingenieuren, Kalkulatoren und Analytikern nicht ausschließen, sondern im Gegenteil ihre Möglichkeiten erweitern, indem sie einen zusätzlichen Blickwinkel auf der Grundlage historischer Daten bieten.

Wenn es richtig in die Geschäftsprozesse von Bauunternehmen integriert wird, hat maschinelles Lernen das Potenzial, ein wichtiges Element im Entscheidungsunterstützungssystem des Managements zu werden - nicht als Ersatz für Menschen, sondern als Erweiterung ihrer professionellen Intuition und technischen Logik.

Nächste Schritte: von der Speicherung zur Analyse und Vorhersage

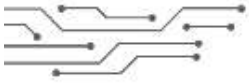
Moderne Ansätze für die Arbeit mit Daten beginnen die Art und Weise, wie Entscheidungen in der Bauindustrie getroffen werden, zu verändern. Der Übergang von intuitiven Bewertungen zur objektiven Datenanalyse verbessert nicht nur die Genauigkeit, sondern eröffnet auch neue Möglichkeiten zur Optimierung von Prozessen. Um diesen Teil zusammenzufassen, lohnt es sich, die wichtigsten praktischen Schritte hervorzuheben, die Ihnen helfen werden, die besprochenen Methoden bei Ihren täglichen Aufgaben anzuwenden:

- Aufbau einer nachhaltigen Speicherinfrastruktur

- Versuchen Sie, unterschiedliche Dokumente und Projektdaten in einem einzigen Tabellenmodell zusammenzufassen und die wichtigsten Informationen für die weitere Analyse in einem einzigen Datenrahmen zusammenzufassen.
- Verwendung effizienter Datenspeicherformate - z. B. spaltenförmige Formate wie Apache Parquet anstelle von CSV oder XLSX - insbesondere für die Datensätze, die in Zukunft möglicherweise zum Trainieren von Modellen für maschinelles Lernen verwendet werden könnten
- Einrichtung eines Systems zur Versionierung von Daten, um Änderungen während des Projekts zu verfolgen

- Implementierung von Analyse- und Automatisierungswerkzeugen
  - Beginnen Sie mit der Analyse historischer Projektdaten - anhand von Unterlagen, Modellen und Schätzungen, um Muster, Trends und Anomalien zu erkennen.
  - Master-ETL-Prozesse (Extract, Transform, Load) zum automatischen Laden und Aufbereiten von Daten
  - Lernen Sie, wie man mit verschiedenen kostenlosen Python-Visualisierungsbibliotheken wichtige Metriken visualisiert
  - Anwendung von statistischen Methoden und Zufallsstichproben, um repräsentative und reproduzierbare analytische Ergebnisse zu erzielen
- Zunehmende Reife im Umgang mit Daten
  - Lernen Sie einige grundlegende Algorithmen des maschinellen Lernens anhand einfacher und überschaubarer Beispiele wie dem Titanic-Datensatz
  - Analysieren Sie aktuelle Prozesse und ermitteln Sie, wo Sie von einer starren Ursache-Wirkungs-Logik zu statistischen Methoden der Vorhersage und Schätzung übergehen können
  - Beginnen Sie, Daten als strategisches Gut und nicht als Nebenprodukt zu behandeln: Bauen Sie Entscheidungsprozesse auf Datenmodellen und nicht auf spezifischen Softwarelösungen auf.

Bauunternehmen, die den Wert von Daten erkannt haben, treten in eine neue Entwicklungsphase ein, in der der Wettbewerbsvorteil nicht durch die Menge der Ressourcen, sondern durch die Geschwindigkeit der Entscheidungsfindung auf der Grundlage von Analysen bestimmt wird.



MAXIMALER KOMFORT MIT DER GEDRUCKTEN VERSION

Sie sind im Besitz einer kostenlosen digitalen Version von **Data-Driven Construction**. Für bequemeres Arbeiten und schnellen Zugriff auf die Materialien empfehlen wir Ihnen, die **gedruckte Ausgabe zu verwenden**:

- **Immer zur Hand**: Das Buch im gedruckten Format wird zu einem zuverlässigen Arbeitsinstrument, das es Ihnen ermöglicht, in jeder Arbeitssituation schnell die richtigen Visualisierungen und Diagramme zu finden und zu verwenden
- **Hochwertige Illustrationen**: Alle Bilder und Grafiken in der gedruckten Ausgabe werden in höchster Qualität präsentiert
- **Schneller Zugriff auf Informationen**: einfache Navigation, die Möglichkeit, Notizen zu machen, Lesezeichen zu setzen und überall mit dem Buch zu arbeiten.

Mit dem Kauf der gedruckten Vollversion des Buches erhalten Sie ein praktisches Hilfsmittel für die bequeme und effiziente Arbeit mit Informationen: die Möglichkeit, bei alltäglichen Aufgaben schnell auf Anschauungsmaterial zurückzugreifen, die notwendigen Schemata schnell zu finden und Notizen zu machen. Darüber hinaus unterstützen Sie mit Ihrem Kauf die Verbreitung von offenem Wissen.

Bestellen Sie eine gedruckte Version des Buches unter: datadrivenconstruction.io/books

X TEIL

DIE BAUINDUSTRIE IM DIGITALEN ZEITALTER. CHANCEN UND HERAUSFORDERUNGEN

Der abschließende zehnte Teil ist ein umfassender Blick auf die Zukunft der Bauindustrie im Zeitalter der digitalen Transformation. Er analysiert den Übergang von der Kausalanalyse zur Arbeit mit Big-Data-Korrelationen. Es werden Parallelen zwischen der Entwicklung der bildenden Kunst und der Entwicklung der Datenarbeit im Bauwesen gezogen, um zu zeigen, wie sich die Branche von der detaillierten Kontrolle zu einem ganzheitlichen Verständnis der Prozesse bewegt. Das Konzept der "Uberisierung " der Bauindustrie wird untersucht, bei der Datentransparenz und automatisierte Berechnungen traditionelle Geschäftsmodelle radikal verändern können, indem sie die Notwendigkeit von Zwischenhändlern beseitigen und die Möglichkeiten für Spekulationen verringern. Ungelöste Fragen, wie die universelle Klassifizierung von Elementen, werden ausführlich erörtert, um den Bauunternehmen Zeit zu geben, sich an das neue Umfeld anzupassen. Der Teil schließt mit konkreten Empfehlungen für die Gestaltung einer digitalen Transformationsstrategie, die die Analyse von Schwachstellen und den Ausbau von Dienstleistungen umfasst, um in einer sich wandelnden Branche wettbewerbsfähig zu bleiben.

KAPITEL 10.1.
ÜBERLEBENSSTRATEGIEN: AUFBAU VON WETTBEWERBSVORTEILEN

Korrelationen statt Berechnungen: die Zukunft der Bauanalytik

Aufgrund der rasanten Digitalisierung von Informationen (Abb. 1.1-5) durchläuft das moderne Bauwesen einen grundlegenden Wandel, bei dem Daten nicht nur ein Werkzeug, sondern ein strategisches Gut werden, das die traditionellen Ansätze des Projekt- und Geschäftsmanagements grundlegend verändern kann.

Seit Tausenden von Jahren stützt sich die Bautätigkeit auf deterministische Methoden, d. h. auf präzise Berechnungen, Detaillierung und strenge Kontrolle der Parameter. In den ersten Jahrhunderten nach Christus wendeten römische Ingenieure mathematische Prinzipien beim Bau von Aquädukten und Brücken an. Im Mittelalter strebten Architekten nach idealen Proportionen für gotische Kathedralen, und in der Industrialisierung des 20. Jahrhunderts entstanden Systeme standardisierter Normen und Vorschriften, die zur Grundlage des Massenbaus wurden

Heute verlagert sich der Vektor der Entwicklung von der Suche nach reinen Ursache-Wirkungs-Beziehungen zur probabilistischen Analyse, der Suche nach Korrelationen und verborgenen Mustern. Die Branche tritt in eine neue Phase ein - Daten werden zu einer Schlüsselressource, und die darauf basierende Analytik ersetzt intuitive und lokal optimierte Ansätze.

Abb. 10.1-1 Verborgene Potenziale von Konstruktionsdaten: Vorhandene Kalkulationen im Unternehmen sind nur die Spitze des Eisbergs, die dem Management zur Analyse zur Verfügung stehen.

> Das Informationssystem eines Unternehmens ist wie ein Eisberg (Abb. 10.1-1): nur ein kleiner Teil des Datenpotenzials ist für die Unternehmensleitung sichtbar, während der Hauptwert in der Tiefe verborgen ist. Es ist wichtig, Daten nicht nur im Hinblick auf ihre derzeitige Nutzung zu bewerten, sondern auch im Hinblick auf die Möglichkeiten, die sie in Zukunft eröffnen werden. Diejenigen Unternehmen, die lernen, verborgene Muster zu erkennen und neue Erkenntnisse aus Daten zu gewinnen, werden einen nachhaltigen Wettbewerbsvorteil erlangen

Bei der Suche nach verborgenen Mustern und der sinnvollen Nutzung von Daten geht es nicht nur um die Arbeit mit Zahlen, sondern um einen kreativen Prozess, der abstraktes Denken und die Fähigkeit erfordert, das Gesamtbild hinter disparaten Elementen zu sehen. In diesem Sinne kann die Entwicklung der Arbeit mit Daten mit der Entwicklung der bildenden Kunst verglichen werden (Abb. 10.1-2).

Die Entwicklung des Bauwesens ähnelt in bemerkenswerter Weise der Entwicklung der bildenden Künste. In beiden Fällen hat sich die Menschheit von primitiven Methoden zu hoch entwickelten Visualisierungs- und Analysetechniken entwickelt. In prähistorischen Zeiten nutzten die Menschen Höhlenzeichnungen und primitive Werkzeuge, um alltägliche Aufgaben zu lösen. Während des Mittelalters und der Renaissance nahm der Grad der Verfeinerung in Architektur und Kunst dramatisch zu. Im frühen Mittelalter hatten sich die Bauwerkzeuge von einer einfachen Axt zu umfangreichen Werkzeugsätzen entwickelt, die das Wachstum des technischen Wissens symbolisierten.

Das Zeitalter des Realismus war die erste Revolution in der bildenden Kunst: Die Künstler lernten, die kleinsten Details wiederzugeben und ein Höchstmaß an Plausibilität zu erreichen. Im Bauwesen war das Gegenstück zu dieser Zeit die präzise Ingenieurtechnik, detaillierte Zeichnungen und streng geregelte Berechnungen, die für Jahrhunderte zur Grundlage der Konstruktionspraxis wurden.

> Später veränderte der Impressionismus die Wahrnehmung der künstlerischen Realität: Statt die Form buchstäblich wiederzugeben, begannen die Künstler, Stimmungen, Licht und Dynamik einzufangen, um einen Gesamteindruck und nicht absolute Genauigkeit wiederzugeben. In ähnlicher Weise bewegt sich das maschinelle Lernen in der Gebäudeanalyse weg von starren logischen Modellen hin zur Mustererkennung und zu probabilistischen Mustern, die es ermöglichen, verborgene Abhängigkeiten in Daten zu "sehen", die bei der klassischen Analyse unzugänglich sind. Dieser Ansatz erinnert an die Bauhaus-Ideen von Minimalismus und Funktionalität, bei denen die Bedeutung (Funktion) wichtiger ist als die Form. Das Bauhaus versuchte, das Überflüssige zu entfernen, auf Ornamente zu verzichten, um Klarheit, Nützlichkeit und Massenattraktivität zu erreichen. Die Dinge sollten verständlich und nützlich sein, ohne Übermaß - die Ästhetik entstand aus der Logik des Designs und des Zwecks.

Mit dem Aufkommen der Fotografie im späten 19. Jahrhundert erhielt die Kunst ein neues Werkzeug, mit dem sie die Realität mit nie dagewesener Genauigkeit erfassen konnte und das die Einstellung zur bildenden Kunst auf den Kopf stellte. In ähnlicher Weise führt die industrielle Revolution im 21. Jahrhundert im Bauwesen zum Einsatz von Robotertechnologien, Lasern, IoT, RFID und Konzepten wie Connected Construction, bei denen sich die Erfassung einzelner Parameter zu einer skalierbaren intelligenten Erfassung der gesamten Realität der Baustelle entwickelt hat.

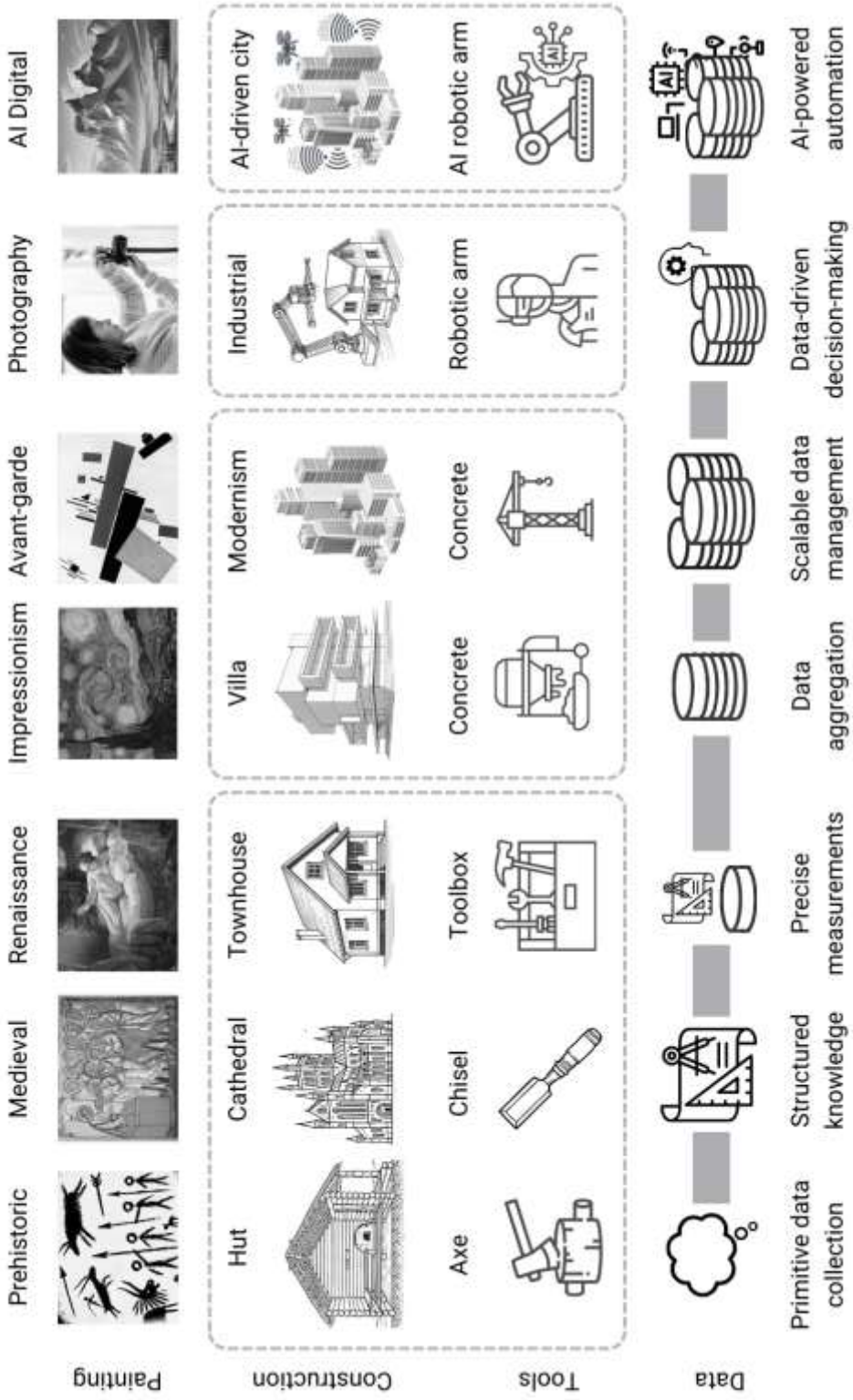


Abb. 10.1-2 Era of Fine Art Evolutions steht im Einklang mit den Entwicklungen Arbeit mit Daten in der Bauindustrie.

So wie die bildende Kunst mit dem Aufkommen der KI-Tools und LLM ein Umdenken erfährt, erlebt die Bau-

branche heute einen weiteren Quantensprung: Intelligente Systeme, die durch künstliche Intelligenz (KI) angetrieben werden, LLM-Chats ermöglichen die Vorhersage, Optimierung und Generierung von Lösungen mit minimalem menschlichen Eingriff.

Die Rolle von Daten in der Planung und Verwaltung hat sich radikal verändert. Während Wissen früher verbal und empirisch vermittelt wurde - so wie die Realität bis ins 19. Jahrhundert durch handgemalte Bilder erfasst wurde - liegt der Schwerpunkt heute auf der vollständigen digitalen Erfassung des "Bildes" der Konstruktion. Mit Hilfe von Algorithmen des maschinellen Lernens wird dieses digitale Bild in eine impressionistische Darstellung der Bauwirklichkeit umgewandelt - kein exaktes Abbild, sondern ein generalisiertes, probabilistisches Verständnis der Prozesse.

> Wir nähern uns rasch einer Ära, in der die Prozesse des Entwerfens, Bauens und Betreibens von Gebäuden nicht nur erweitert, sondern weitgehend durch Systeme der künstlichen Intelligenz gesteuert werden. So wie moderne digitale Kunst ohne Pinsel entsteht - mit Hilfe von Textanweisungen und generativen Modellen, werden die architektonischen und ingenieurtechnischen Lösungen der Zukunft durch vom Nutzer festgelegte Schlüsselabfragen und Parameter geprägt sein.

Im 21. Jahrhundert sind der Zugang zu Daten, ihre Interpretation und die Qualität der Analysen für den Projekterfolg unverzichtbar geworden. Und der Wert der Daten wird nicht durch ihre Menge bestimmt, sondern durch die Fähigkeit der Fachleute, sie zu analysieren, zu überprüfen und in Maßnahmen umzusetzen.

Datengesteuerter Ansatz im Bauwesen: eine neue Ebene der Infrastruktur

In der Geschichte der Menschheit hat jeder dieser Technologiesprünge zu grundlegenden Veränderungen in Wirtschaft und Gesellschaft geführt. Heute erleben wir eine neue Welle des Wandels, die in ihrem Ausmaß mit der industriellen Revolution des 19. Jahrhunderts vergleichbar ist. Doch während vor hundert Jahren mechanische Kräfte und Energietechnologien die Haupttriebkräfte des Wandels waren, sind es heute Daten und künstliche Intelligenz.

Agenten für maschinelles Lernen, LLM und KI verändern das Wesen von Anwendungen und machen herkömmliche Software-Stacks (die im zweiten Teil des Buches behandelt werden) überflüssig (Abb. 2.2-3). Die gesamte Datenlogik ist in KI-Agenten und nicht in hart kodierten Geschäftsregeln angesiedelt (Abb. 2.2-4).

> Im Datenzeitalter werden die traditionellen Ansichten über Anwendungen grundlegend verändert. Wir bewegen uns auf ein Modell zu, bei dem sperrige, modulare Unternehmenssysteme unweigerlich offenen, leichtgewichtigen, maßgeschneiderten Lösungen Platz machen werden.

> *In Zukunft wird nur noch die zugrunde liegende Datenstruktur übrig bleiben, und die gesamte Interaktion mit ihr wird über Agenten erfolgen, die direkt mit der Datenbank arbeiten. Ich glaube wirklich, dass der gesamte Anwendungsstapel verschwinden wird, weil es einfach keinen Bedarf dafür gibt, wenn künstliche Intelligenz direkt mit der zugrunde liegenden Datenbank interagiert. Ich habe meine gesamte Karriere damit verbracht, in SaaS zu arbeiten - Unternehmen aufzubauen, in ihnen zu arbeiten, und um ehrlich zu sein, würde ich jetzt wahrscheinlich kein neues SaaS-Unternehmen gründen. Und ich würde wahrscheinlich auch nicht in SaaS-Unternehmen investieren. Die Situation ist zu unsicher. Das soll nicht heißen, dass es in Zukunft keine Softwareunternehmen geben wird, nur werden sie ganz anders aussehen. Künftige Systeme werden Datenbanken mit Geschäftslogik sein, die in [AI] Agenten eingebracht werden. Diese Agenten werden mit mehreren Datenbeständen gleichzeitig arbeiten und nicht auf eine einzige Datenbank beschränkt sein. Die gesamte Logik wird in die KI-Schicht verlagert [46].*
>
> - Matthew Berman, Geschäftsführer Forward Future

Der Hauptunterschied des neuen Paradigmas liegt in der Minimierung des technologischen Ballasts. Anstelle monumentaler komplexer und geschlossener Softwaresysteme werden wir flexible, offene und schnell anpassbare Module erhalten, die buchstäblich im Datenfluss "leben" (Abb. 7.4-1 - Apache Airflow, NiFi). Die Architektur des künftigen Prozessmanagements sieht den Einsatz von Mikroanwendungen vor - kompakte, zweckgebundene Tools, die sich grundlegend von massiven und geschlossenen ERP-, PMIS-, CDE-, CAFM-Systemen unterscheiden. Die neuen Agenten werden so anpassungsfähig, integriert und geschäftsspezifisch wie möglich sein (z.B. Low-Code/No-Code Abb. 7.4-6).

> *Die gesamte Geschäftslogik geht an diese [AI] Agenten, und diese Agenten führen CRUD [Create, Read, Update, and Delete]-Operationen auf mehreren Repositories durch, d.h. sie unterscheiden nicht, welches Backend verwendet wird. Sie aktualisieren mehrere Datenbanken, und die gesamte Logik landet in der sogenannten KI-Ebene. Und sobald sich die gesamte Logik in der KI-Ebene befindet, werden die Leute anfangen, Backends zu ersetzen. Wir sehen bereits einen ziemlich hohen Prozentsatz an Marktgewinnen bei Dynamics-Backends und der Nutzung von Agenten, und wir werden uns aggressiv in diese Richtung bewegen und versuchen, alles zusammenzubringen. Sei es im Kundenservice oder in anderen Bereichen, zum Beispiel nicht nur im CRM, sondern auch in unseren Finanz- und Betriebslösungen. Denn die Menschen wollen mehr KI-gesteuerte Geschäftsanwendungen, bei denen die Logikebene durch KI und KI-Agenten gesteuert werden kann. [...]. Eines der aufregendsten Dinge für mich ist Excel mit Python, was vergleichbar ist mit GitHub mit Copilot. Das heißt, was wir gemacht haben: Wenn Sie Excel haben, sollten Sie es einfach öffnen, Copilot starten und anfangen, damit zu spielen. Es geht nicht mehr nur darum, die verfügbaren Zahlen zu verstehen - es wird von selbst einen Plan erstellen. So wie der GitHub Copilot-Arbeitsbereich einen Plan erstellt und ihn dann ausführt, ist es wie bei einem Datenanalysten, der Excel als Werkzeug zur Visualisierung von Zeilen und Spalten für die Analyse verwendet. Copilot nutzt also Excel als Tool mit all seinen Möglichkeiten, da es Daten generieren kann und über einen Python-Interpreter verfügt.*
>
> - Satya Nadella, CEO, Microsoft, Interview mit BG2 Channel Dezember 2024. [28]

Der Wandel, den wir in der Logik der Büroanwendungen erleben - der Übergang von modularen, geschlossenen Systemen zu KI-Agenten, die direkt mit offenen Daten arbeiten - ist nur Teil eines viel größeren Prozesses. Es geht nicht nur um die Veränderung von Schnittstellen oder der Softwarearchitektur: Die Veränderungen werden sich auf die grundlegenden Prinzipien der Arbeitsorganisation, der Entscheidungsfindung und der Unternehmensführung auswirken. Im Bauwesen wird dies zu einer datengesteuerten Logik führen, in der Daten zum Kernstück von Prozessen von der Planung über das Ressourcenmanagement bis zur Bauüberwachung werden.

Das digitale Büro der nächsten Generation: Wie KI den Arbeitsbereich verändert

Vor fast einem Jahrhundert erlebte die Menschheit bereits eine ähnliche technologische Revolution. Der Übergang von der Dampfmaschine zum Elektromotor dauerte mehr als vier Jahrzehnte, führte aber letztlich zu einem beispiellosen Produktivitätswachstum - vor allem durch die Dezentralisierung der Energiekapazität und die Flexibilität der neuen Lösungen. Dieser Wandel veränderte nicht nur den Lauf der Geschichte, indem er den Großteil der Bevölkerung vom Land in die Städte brachte, sondern legte auch den Grundstein für die moderne Wirtschaft. Die Geschichte der Technologie ist eine Reise von körperlicher Arbeit zu Automatisierung und intelligenten Systemen. So wie der Traktor unzählige Bodenbearbeitungsmaschinen ersetzt hat, so ersetzt die moderne Digitaltechnik die traditionellen bürogestützten Baumanagementmethoden (Abb. 10.1-3). Noch zu Beginn des 20. Jahrhunderts bearbeitete der Großteil der Weltbevölkerung das Land von Hand, bis in den 1930er Jahren die Mechanisierung der Arbeit durch Maschinen und Traktoren begann.

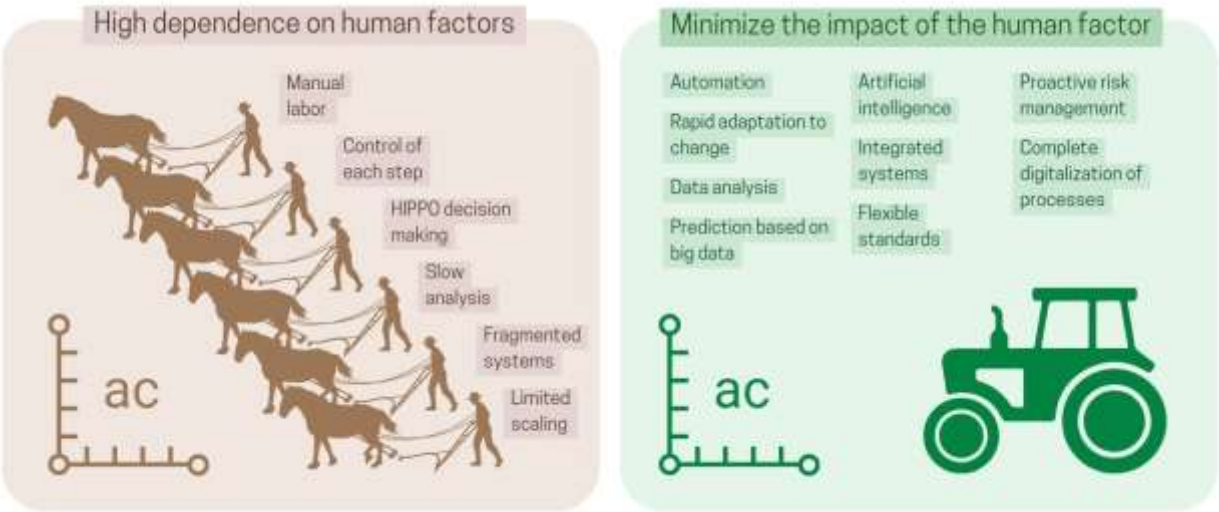


Abb. 10.1-3 So wie der Traktor Anfang des 20. Jahrhunderts Dutzende von Menschen ersetzte, wird das maschinelle Lernen im 21. Jahrhundert traditionelle Geschäfts- und Projektmanagementmethoden ersetzen.

So wie die Menschheit vor hundert Jahren von der Bewirtschaftung einzelner Parzellen mit primitiven Werkzeugen zur groß angelegten Landwirtschaft mit Hilfe von Maschinen überging, machen wir heute den Übergang von der Verarbeitung disparater "Silos" von Informationen zur Arbeit mit Datenarrays mit Hilfe von leistungsstarken "Traktoren" - ETL-Pipeline und Algorithmen der künstlichen Intelligenz.

Wir stehen an der Schwelle zu einem ähnlichen Sprung - allerdings auf der digitalen Ebene: von der traditionellen, manuellen Unternehmensführung zu datengesteuerten Modellen.

Der Weg zu einer vollwertigen datengesteuerten Architektur wird Zeit, Investitionen und organisatorische Anstrengungen erfordern. Aber dieser Weg eröffnet nicht nur eine schrittweise Verbesserung, sondern einen qualitativen Sprung hin zu mehr Effizienz, Transparenz und Handhabbarkeit der Bauprozesse. Voraussetzung dafür ist die systematische Einführung digitaler Werkzeuge und die Abkehr von überholten Geschäftspraktiken.

> Aufgabenparametrisierung, ETL, LLM, IoT-Komponenten, RFID, Tokenisierung, Big Data und maschinelles Lernen werden das traditionelle Bauwesen in ein **datengesteuertes Bauwesen** verwandeln, in dem jedes Detail des Projekts und des Baugeschäfts durch Daten gesteuert und optimiert wird.

Früher brauchte man Tausende von Arbeitsstunden, um Informationen zu analysieren. Jetzt werden diese Aufgaben von Algorithmen und LLMs übernommen, die disparate Datensätze mithilfe von Eingabeaufforderungen in strategische Quellen verwandeln. In der Welt der Technik geschieht das Gleiche wie in der Landwirtschaft: Wir bewegen uns von der Hacke zum automatisierten Agrarbusiness. Auch die Büroarbeit im Baugewerbe bewegt sich von Excel-Dateien und manuellen Zusammenfassungen hin zu einem intelligenten System, in dem Daten gesammelt, bereinigt, strukturiert und in Erkenntnisse umgewandelt werden.

Schon heute sollten Unternehmen damit beginnen, Informationsfelder durch Qualitätsdatenerfassung und Informationsstrukturierung zu "kultivieren", sie mit Bereinigungs- und Normalisierungswerkzeugen zu "düngen" und sie dann in Form von prädiktiven Analysen und automatisierten Lösungen zu "ernten". Wenn ein moderner Landwirt mit einer Maschine in der Lage ist, hundert Ackerbauern zu ersetzen, werden intelligente Algorithmen in der Lage sein, den Mitarbeitern die Routine abzunehmen und sie in die Rolle von strategischen Managern der Informationsflüsse zu versetzen.

Es ist jedoch wichtig zu verstehen, dass die Schaffung einer wirklich datengesteuerten -Organisation kein schneller Prozess ist. Es handelt sich um eine langfristige strategische Ausrichtung, ähnlich wie bei der Schaffung eines neuen Standorts für die Anpflanzung eines neuen Waldes (Abb. 1.2-5) von Systemen, wobei jeder "Baum" in diesem Ökosystem ein separater Prozess, eine Kompetenz oder ein Werkzeug ist, das Zeit braucht, um zu wachsen und sich zu entwickeln. Wie bei einem echten Wald hängt der Erfolg nicht nur von der Qualität des Pflanzmaterials (Technologie) ab, sondern auch vom Boden (Unternehmenskultur), dem Klima (Geschäftsumfeld) und der Pflege (Systemansatz).

Die Unternehmen werden sich nicht mehr ausschließlich auf geschlossene, sofort einsatzbereite Lösungen verlassen können. Im Gegensatz zu früheren Phasen der technologischen Entwicklung wird der derzeitige Übergang zu offenem Datenzugang, künstlicher Intelligenz und Open Source von den großen Anbietern wahrscheinlich nicht unterstützt, da er ihre etablierten Geschäftsmodelle und Haupteinnahmequellen direkt bedroht.

Wie aus der Studie der Harvard Business School [40] hervorgeht, die bereits im Kapitel über die vierte und fünfte technologische Revolution erörtert wurde, würden sich die Kosten für die Neuerstellung der am häufigsten verwendeten Open-Source-Lösungen für alle Unternehmen auf etwa 4,15 Milliarden Dollar belaufen. Wenn wir uns jedoch vorstellen, dass jedes Unternehmen seine eigenen Alternativen ohne Zugang zu bestehenden Open-Source-Tools entwickeln würde, wie es in den letzten Jahrzehnten geschehen ist, könnten die Gesamtkosten für die Unternehmen kolossale 8,8 Billionen Dollar erreichen - das ist der Preis für die irrationale Nachfrage, mit dem der Softwaremarkt bewertet werden kann.

Der technologische Fortschritt wird unweigerlich zu einem Überdenken der etablierten Geschäftsmodelle führen. Während Unternehmen früher mit komplexen, undurchsichtigen Prozessen und verschlossenen Daten Geld verdienen konnten, wird dieser Ansatz mit der Entwicklung von KI und Analytik immer weniger praktikabel.

> Infolge der Demokratisierung des Zugangs zu Daten und Werkzeugen könnte der traditionelle Software-Verkaufsmarkt erheblich schrumpfen. Gleichzeitig wird jedoch ein neuer Markt wachsen - der Markt für digitales Know-how, Anpassung, Integration und Lösungsdesign. Hier wird der Wert nicht aus dem Verkauf von Lizenzen kommen, sondern aus der Fähigkeit, flexible, offene und anpassungsfähige digitale Prozesse aufzubauen. So wie die Elektrifizierung und das Aufkommen von Traktoren neue Industrien hervorgebracht haben, wird auch die Anwendung von Big Data, KI und LLM völlig neue Horizonte für das Baugewerbe eröffnen, die nicht nur technologische Investitionen, sondern auch einen tiefgreifenden Wandel von Denkweisen, Prozessen und Organisationsstrukturen erfordern werden. Und die Unternehmen und Fachleute, die dies erkennen und heute damit beginnen, werden die Marktführer von morgen sein.

In einer Welt, in der offene Daten zu einem wichtigen Gut werden, wird die Verfügbarkeit von Informationen ein entscheidender Faktor sein. Investoren, Kunden und Aufsichtsbehörden werden zunehmend Transparenz fordern, und Algorithmen des maschinellen Lernens werden in der Lage sein, Diskrepanzen bei Schätzungen, Zeitplänen und Kosten automatisch zu erkennen. Dies schafft die Voraussetzungen für eine neue Phase der digitalen Transformation, die uns allmählich zur "Uberisierung " der Baubranche führt.

Offene Daten und Uberisierung sind eine Bedrohung für das bestehende Baugeschäft

Das Bauwesen entwickelt sich zu einem Prozess des Informationsmanagements. Je genauer, hochwertiger und vollständiger die Daten sind, desto effizienter sind Entwurf, Berechnung, Kostenvoranschlag, Errichtung und Betrieb von Gebäuden. In Zukunft werden nicht mehr Kran, Beton und Bewehrung die wichtigste Ressource sein, sondern die Fähigkeit, Informationen zu sammeln, zu analysieren und zu nutzen.

> In Zukunft werden die Kunden von Bauunternehmen - Investoren und Bauherren - unweigerlich den Wert offener Daten und Analysen historischer Daten nutzen. Dies wird Möglichkeiten eröffnen, die Berechnung von Projektzeitplänen und -kosten zu automatisieren, ohne dass die Bauunternehmen in Kalkulationsfragen involviert werden, was dazu beitragen wird, die Kosten zu kontrollieren und überflüssige Kosten schneller zu erkennen.

Stellen Sie sich eine Baustelle vor, auf der Laserscanner, Quadrocopter und Photogrammetriesysteme genaue Echtzeitdaten über die verbauten Betonmengen sammeln. Diese Informationen werden automatisch in einfache flache MESH -Modelle mit Metadaten umgewandelt, wobei umständliche CAD (BIM) -Systeme umgangen werden, ohne Abhängigkeiten von komplexen geometrischen Kernen, ERP oder PMIS. Diese von der Baustelle gesammelten Daten werden zentral in ein einziges strukturiertes Repository übertragen, das dem Kunden für unabhängige Analysen zur Verfügung steht. Dort werden reale Preise von verschiedenen Bauunternehmen hochgeladen und Parameter, die von der Kreditfinanzierungsrate bis hin zu sich dynamisch verändernden Faktoren wie Wetterbedingungen, Börsennotierungen für Baumaterialien, Logistiktari-

fen und statistischen saisonalen Schwankungen der Arbeitspreise reichen. Unter diesen Bedingungen werden Abweichungen zwischen Planung und tatsächlichen Materialmengen sofort deutlich, so dass eine Manipulation der Schätzungen sowohl in der Planungsphase als auch bei der Übergabe des Projekts unmöglich ist. Die Transparenz des Bauprozesses wird also nicht durch ein Heer von Aufsichtspersonen und Managern erreicht, sondern durch objektive digitale Daten, die den menschlichen Faktor und die Möglichkeit von Spekulationen minimieren.

Diese Art der Datenkontrolle wird in Zukunft eher von Datenmanagern auf der Kundenseite übernommen werden (Abb. 1.2-4 CQMS manager). Dies gilt insbesondere für Kalkulationen und Projektschätzungen: Wo es früher eine ganze Abteilung von Schätzern gab, wird es morgen bereits maschinelles Lernen und Prognosewerkzeuge geben, die Preisgrenzen festlegen, in die sich die Bauunternehmen einfügen müssen.

> *Angesichts der Fragmentierung der [Bau-]Industrie, in der die meisten Systeme und Teilsysteme von KMU geliefert werden, muss die digitale Strategie vom Kunden ausgehen. Die Kunden müssen die Bedingungen und Mechanismen schaffen, um die digitalen Fähigkeiten der Lieferkette freizusetzen (20).*
>
> - Andrew Davis und Giuliano Denicol, Accenture "Mehr Wert schaffen durch Kapitalprojekte"

Eine solche Offenheit und Transparenz von Daten stellt eine Bedrohung für Bauunternehmen dar, die es gewohnt sind, mit undurchsichtigen Prozessen und verwirrenden Berichten Geld zu verdienen, bei denen sich Spekulationen und versteckte Kosten hinter komplexen und geschlossenen Formaten und modularen proprietären Datenplattformen verbergen können. Daher sind Bauunternehmen, ebenso wie Anbieter, die für Open-Source-Lösungen werben, wahrscheinlich nicht daran interessiert, offene Daten vollständig in ihre Geschäftsprozesse zu integrieren. Wenn die Daten verfügbar und für den Kunden einfach zu verarbeiten sind, können sie automatisch überprüft werden, was die Möglichkeit einer Überschätzung der Mengen und der Manipulation von Schätzungen ausschließt.

> Laut dem Bericht des Weltwirtschaftsforums "Shaping the Future of Construction" (2016) [5] bleibt eine der wichtigsten Herausforderungen für die Branche die passive Rolle des Kunden. Dennoch sind es die Kunden, die mehr Verantwortung für das Ergebnis von Projekten übernehmen sollten - von der frühzeitigen Planung über die Auswahl nachhaltiger Interaktionsmodelle bis hin zur Leistungsüberwachung. Ohne die aktive Beteiligung der Bauherren ist ein systemischer Wandel in der Bauindustrie nicht möglich.

Der Verlust der Kontrolle über Mengen- und Kostenkalkulationen hat in den letzten 20 Jahren bereits andere Branchen verändert und es den Kunden ermöglicht, ihre Ziele direkt, ohne Zwischenhändler, an zu richten. Digitalisierung und Datentransparenz haben viele traditionelle Geschäftsmodelle verändert, wie z. B. Taxifahrer mit dem Aufkommen von Uber (Abb. 10.1-4), Hoteliers mit dem Aufkommen von Airbnb und Einzelhändler und Geschäfte mit dem Aufstieg von Amazon sowie Banken mit dem Aufkommen von Neo-Banken und dezentralisierten Fintech-Ökosystemen, in denen der direkte Zugang zu Informationen und Automatisierung von Zeit- und Kostenberechnungen die Rolle der Vermittler deutlich reduziert haben.

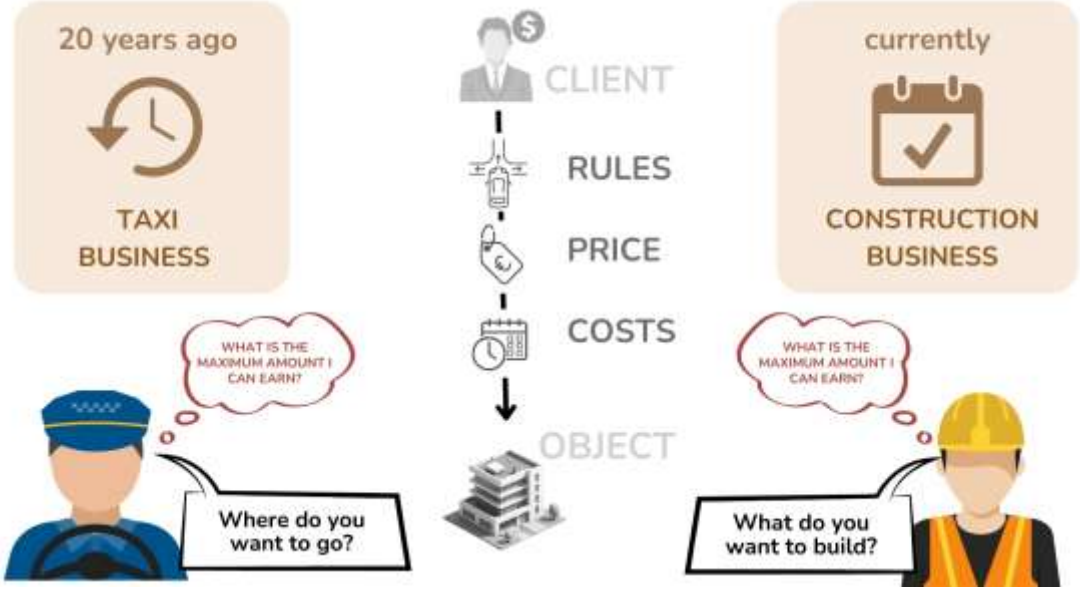


Abb. 10.1-4 Das Baugewerbe wird mit der Uberisierung konfrontiert werden, mit der Taxifahrer, Hoteliers und Einzelhändler vor 10 Jahren konfrontiert waren.

Der Prozess der Demokratisierung des Zugangs zu Daten und Werkzeugen für ihre Verarbeitung ist unvermeidlich, und mit der Zeit werden offene Daten zu allen Projektkomponenten zu einer Anforderung der Kunden und zu einem neuen Standard werden. Daher wird die Einführung offener Formate und transparenter Berechnungen von Investoren, Kunden, Banken und Private-Equity-Fonds (Private Equity) vorangetrieben werden - also von denjenigen, die die Endnutzer der errichteten Anlagen sind und die Anlage dann jahrzehntelang betreiben.

> Großinvestoren, Bauherren und Banken fordern bereits Transparenz in der Baubranche. Laut der Accenture-Studie "Creating more value through capital projects" (2020) [20] werden transparente und zuverlässige Daten zu einem entscheidenden Faktor für Investitionsentscheidungen im Bauwesen. Wie Experten betonen, ist ein vertrauenswürdiges und effektives Projektmanagement ohne Transparenz nicht möglich, insbesondere in Krisenzeiten. Darüber hinaus gehen Eigentümer und Auftragnehmer zunehmend dazu über, Verträge abzuschließen, die Anreize für die gemeinsame Nutzung von Daten und kollaborative Analysen bieten. Dies spiegelt die wachsenden Anforderungen von Investoren, Banken und Regulierungsbehörden an Rechenschaftspflicht und Transparenz wider.

> Die Bewegung des Investors, des Kunden von der Idee bis zum fertigen Gebäude, wird in Zukunft wie eine Fahrt mit dem Autopiloten sein - ohne Fahrer in Form einer Baufirma, verspricht unabhängig von Spekulation und Unsicherheit zu werden.

Das Zeitalter der offenen Daten und der Automatisierung wird das Baugeschäft unweigerlich verändern, so wie es im Bankwesen, im Handel, in der Landwirtschaft und in der Logistik bereits geschehen ist. In diesen Branchen weichen die Rolle der Vermittler und die traditionellen Geschäftspraktiken der Automatisierung und Robotisierung und lassen keinen Raum für ungerechtfertigte Aufschläge und Spekulationen.

Die Daten und Prozesse in allen menschlichen Wirtschaftstätigkeiten unterscheiden sich nicht von denen, mit denen die Fachleute in der Baubranche umgehen müssen. Langfristig könnten die Bauunternehmen, die

heute den Markt beherrschen, indem sie die Standards für Preise und Dienstleistungsqualität festlegen, ihre Rolle als wichtige Vermittler zwischen dem Kunden und seinem Bauprojekt verlieren.

Ungelöste Probleme der Uberisierung als letzte Chance, Zeit für Transformation zu nutzen

Doch kehren wir zu den Realitäten des Baugewerbes zurück. Während selbstfahrende Autos, dezentralisierte Finanzsysteme und auf künstlicher Intelligenz basierende Lösungen in einigen Wirtschaftszweigen auftauchen, ist ein erheblicher Teil der Bauunternehmen nach wie vor papierbasierte Organisationen, in denen wichtige Entscheidungen eher auf der Grundlage der Intuition und Erfahrung einzelner Spezialisten getroffen werden.

In diesem Paradigma kann ein modernes Bauunternehmen mit einem 20 Jahre alten Taxi-Unternehmen verglichen werden, das Ressourcen, Routen und Lieferzeiten kontrolliert und für den Zeitplan und die Kosten der "Reise" verantwortlich ist - von der Projektidee (Logistik und Installationsprozess) bis zur Übergabe des Projekts. So wie GPS (im Bauwesen IoT, RFID) und Algorithmen des maschinellen Lernens bei Zeit-/Kostenberechnungen einst den Bereich des Transportwesens veränderten, haben Daten, Algorithmen und KI -Agenten das Potenzial, das Baumanagement zu verändern - von intuitiven Einschätzungen zu vorausschauenden, geführten Modellen. In den letzten 20 Jahren haben viele Branchen - Finanzwesen, Landwirtschaft, Einzelhandel und Logistik - erlebt, wie die Fähigkeit, durch die Undurchsichtigkeit von Daten zu spekulieren, allmählich verschwand. Preise, Lieferkosten oder Finanztransaktionen werden automatisch und statistisch fundiert berechnet - in nur wenigen Sekunden auf digitalen Plattformen.

Mit Blick auf die Zukunft müssen Bauunternehmen erkennen, dass die Demokratisierung des Datenzugangs und der Analysewerkzeuge den traditionellen Ansatz zur Schätzung von Projektkosten und -zeiträumen durchbrechen und die Möglichkeit, auf undurchsichtigen Mengen- und Preisdaten zu spekulieren, beseitigen wird.

> Wie das Fahren auf einer geregelten Straße ohne Fahrereingriff werden die Bauprozesse der Zukunft zunehmend einem "Uber"-System ähneln - mit automatischer Zeit- und Kostenschätzung, transparenter Aufgabenverteilung und minimaler Abhängigkeit von menschlichen Faktoren. Dies wird die Art der "Reise" von der Idee bis zur Realisierung verändern und sie vorhersehbarer, kontrollierbarer und datengesteuerter machen.

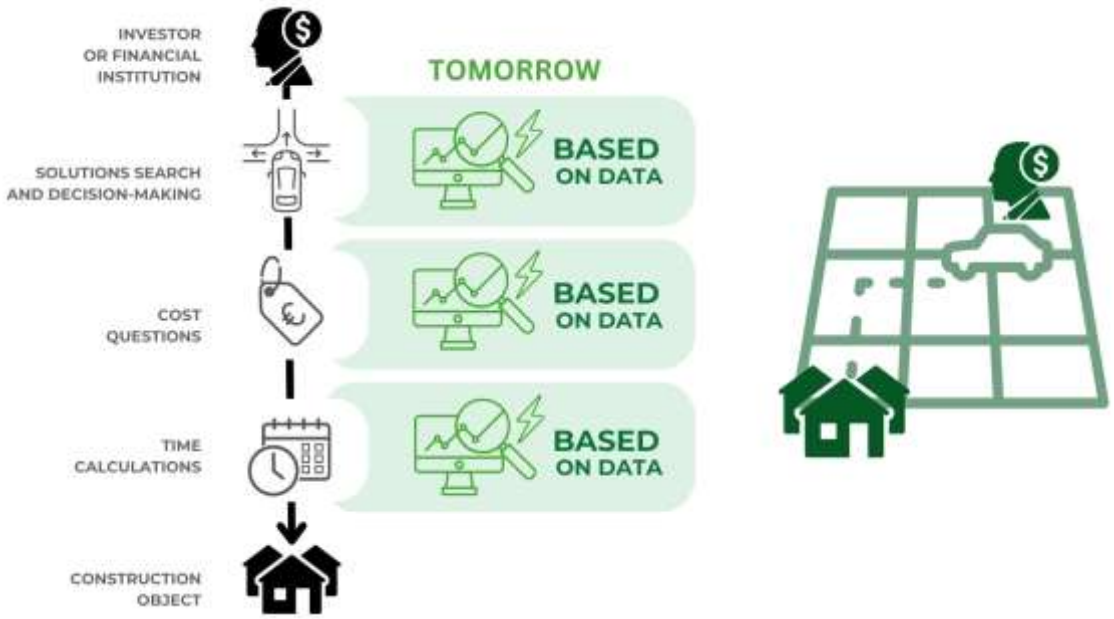


Abb. 10.1-5 Die Kosten und die "Reisezeit" während der Bauphase werden mit Hilfe von maschinellem Lernen und statistischen Tools ermittelt.

Mit der schrittweisen Einführung neuer Vorschriften und Anforderungen in fast allen Ländern der Welt, die die Übermittlung von CAD- (BIM)-Modellen an Bauherren oder Banken, die Bauprojekte finanzieren, vorschreiben, haben Auftraggeber und Kunden die Möglichkeit, unabhängig voneinander für Transparenz bei Kosten- und Umfangskalkulationen zu sorgen. Dies ist insbesondere für Großkunden und Investoren wichtig, die über ausreichende Kompetenzen und Werkzeuge verfügen, um Mengen zeitnah zu analysieren und Marktpreise zu überwachen. Für Unternehmen, die große Standardprojekte - Geschäfte, Bürogebäude, Wohnkomplexe - realisieren, werden solche Praktiken zum Standard.

Da der Informationsgehalt von Modellen immer vollständiger und standardisierter wird, verschwindet die Möglichkeit der Manipulation und Spekulation praktisch. Die digitale Transformation verändert allmählich die Spielregeln in der Baubranche, und Unternehmen, die sich nicht an diese Veränderungen anpassen, könnten vor ernsthaften Herausforderungen stehen.

> Zunehmender Wettbewerb, technologische Störungen und schrumpfende Gewinnspannen haben das Potenzial, die Nachhaltigkeit von Unternehmen zu beeinträchtigen. Angesichts der begrenzten Liquidität setzen immer mehr Branchenakteure auf Automatisierung, Analytik und Datentechnologien, um die Effizienz und Prozesstransparenz zu verbessern. Diese Tools werden zu einer wichtigen Ressource, um in einem sich verändernden wirtschaftlichen Umfeld wettbewerbsfähig zu bleiben.

Vielleicht sollten wir nicht auf äußere Umstände warten, die uns zu dringendem Handeln zwingen - es ist viel effektiver, sich schon heute vorzubereiten, indem wir die digitalen Kompetenzen stärken, moderne Lösungen implementieren und eine datenzentrierte Kultur aufbauen.

Eines der letzten großen technologischen Hindernisse für die groß angelegte digitale Transformation der Bauindustrie, die in den kommenden Jahren jedes Unternehmen betreffen wird, ist das Problem der automatischen Klassifizierung von Elementen von Bauprojekten.

> Ohne eine zuverlässige, genaue und skalierbare Klassifizierung ist es unmöglich, die Grundlage für eine vollwertige Analyse, Prozessautomatisierung und ein Lebenszyklusmanagement mithilfe von KI und Vorhersagemodellen zu schaffen. Solange die Klassifizierung von Objekten noch von der manuellen Interpretation durch erfahrene Spezialisten - Vorarbeiter, Konstrukteure, Kalkulatoren - abhängt, hat die Baubranche noch ein Zeitfenster. Diese Zeit kann genutzt werden, um sich auf die unvermeidlichen Veränderungen vorzubereiten: steigende Anforderungen an die Transparenz, die Demokratisierung von Werkzeugen und Daten sowie das Aufkommen automatischer Klassifizierungssysteme, die die Spielregeln radikal verändern werden.

Die Aufgabe der automatischen Klassifizierung von Elementen der Bauwelt ist in ihrer Komplexität vergleichbar mit der Objekterkennung in unbemannten Fahrsystemen, die eine der größten Herausforderungen darstellt. Stellen wir uns ein unbemanntes Auto vor, das von Punkt A nach Punkt B fährt (Abb. 10.1-5). Die derzeitigen automatischen Fahrsysteme sind mit dem Problem der Klassifizierung von Objekten, die von Lidars und Kameras erkannt werden, überfordert. Es reicht nicht aus, dass ein Auto ein Hindernis oder eine Landmarke "sieht", sondern es muss auch erkennen können, ob es sich um einen Fußgänger, ein Straßenschild oder einen Mülleimer handelt.

Eine ähnliche grundsätzliche Herausforderung stellt sich für die gesamte Bauindustrie. Projektelemente - wie Fenster, Türen oder Stützen - können in Unterlagen erfasst, in CAD-Modellen dargestellt, auf der Baustelle fotografiert oder in Punktwolken aus Laserscans erkannt werden. Ihre visuelle oder grobe geometrische Erkennung reicht jedoch nicht aus, um ein wirklich automatisiertes Projektmanagementsystem aufzubauen. Es muss sichergestellt werden, dass jedes Element akkurat und konsistent in einen Typ kategorisiert wird, der in allen nachfolgenden Prozessen eindeutig identifizierbar ist - von Kostenvoranschlägen und Spezifikationen bis hin zur Logistik, Bestandskontrolle und vor allem dem Betrieb (Abb. 4.2-6).

In dieser Phase - dem Übergang von der Erkennung zur sinnvollen Klassifizierung - entsteht eines der größten Hindernisse. Auch wenn digitale Systeme technisch in der Lage sind, Objekte in Modellen und auf der Baustelle zu unterscheiden und zu identifizieren, liegt die Hauptschwierigkeit in der korrekten und kontextstabilen Definition des Elementtyps für unterschiedliche Softwareumgebungen.. So kann beispielsweise eine Tür vom Konstrukteur in einem CAD-Modell als Element der Kategorie "Tür" bezeichnet werden, aber bei der Übertragung in ein ERP- oder PMIS-System kann sie falsch getippt werden, entweder aufgrund eines Fehlers des Konstrukteurs oder aufgrund von Inkonsistenzen zwischen den Systemen. Darüber hinaus verliert das Element bei Datenexporten und -importen oft einige wichtige Attribute oder verschwindet aus der Buchhaltung des Systems. Dies führt zu einer Lücke im Datenfluss und untergräbt das Prinzip der durchgängigen Digitalisierung von Bauprozessen. Dadurch entsteht eine kritische Lücke zwischen "sichtbarer" und "verständlicher" semantischer Bedeutung, die die Datenintegrität untergräbt und die Automatisierung von Prozessen während des gesamten Lebenszyklus eines Bauprojekts erheblich erschwert.

Die Lösung des Problems der universellen Klassifizierung von Bauelementen mithilfe von Big Data und Technologien des maschinellen Lernens (Abb. 10.1-6) wird ein Katalysator für einen branchenweiten Wandel sein - und vielleicht eine unerwartete Entdeckung für viele Bauunternehmen. Ein einheitliches, lernfähiges Klassifizierungssystem wird die Grundlage für skalierbare Analysen, digitales Management und die Einführung von KI in die tägliche Praxis von Bauunternehmen bilden.

NVIDIA und andere Technologieführer bieten bereits in anderen Branchen Lösungen an, mit denen große Mengen an Text- und Bildinformationen automatisch kategorisiert und strukturiert werden können.

> Das NeMo Curator-Modell von NVIDIA [161] beispielsweise ist auf die automatische Klassifizierung und Kategorisierung von Daten in vordefinierte Kategorien spezialisiert und spielt eine Schlüsselrolle bei der Optimierung von Informationsverarbeitungspipelines für die Feinabstimmung und das Vortraining generativer KI-Modelle. Die Cosmos-Plattform wird auf realen Video- und 3D-Szenen [162] trainiert und bietet eine Grundlage für autonome Systeme und digitale Zwillinge, die bereits im NVIDIA-Ökosystem entwickelt werden. NVIDIA Omniverse, das sich bis 2025 zum führenden Tool für die Arbeit mit dem USD-Format entwickelt hat, einer universellen Szenenbeschreibung, die schließlich das IFC-Format bei der Übertragung von Designinformationen ersetzen könnte. Zusammen mit Isaac Sim, einem Simulator für Roboterprozesse [163], stellen Lösungen wie NeMo Curator, Cosmos und Omniverse eine neue Stufe der Automatisierung dar: von der Datenbereinigung und -filterung bis hin zur Erstellung von Trainingssätzen, der Modellierung von Objekteigenschaften und dem Robotertraining auf der Baustelle. Und alle diese Tools sind kostenlos und quelloffen, was die Hürden für die Einführung in der Ingenieur- und Baupraxis erheblich senkt.

Die automatische Klassifizierung von Daten auf der Ebene von strukturierten Tabellen ist nicht so schwierig, wie es auf den ersten Blick erscheinen mag. Wie wir im vorigen Kapitel gezeigt haben (Abb. 9.1-10), ist es möglich, fehlende oder falsche Klassenwerte auf der Grundlage ähnlicher Parameter anderer Elemente auszugleichen, wenn es akkumulierte historische Daten gibt. Wenn Elemente mit ähnlichen Merkmalen bereits in mehreren abgeschlossenen Projekten korrekt klassifiziert wurden, kann das System mit hoher Wahrscheinlichkeit einen passenden Wert für ein neues oder unvollständiges Element vorschlagen (Abb. 10.1-6). Eine solche Logik, die auf gemittelten Werten und einer Kontextanalyse basiert, kann bei der Massenverarbeitung von tabellarischen Daten aus Kostenvoranschlägen, Spezifikationen oder CAD-Modellen besonders effektiv sein.

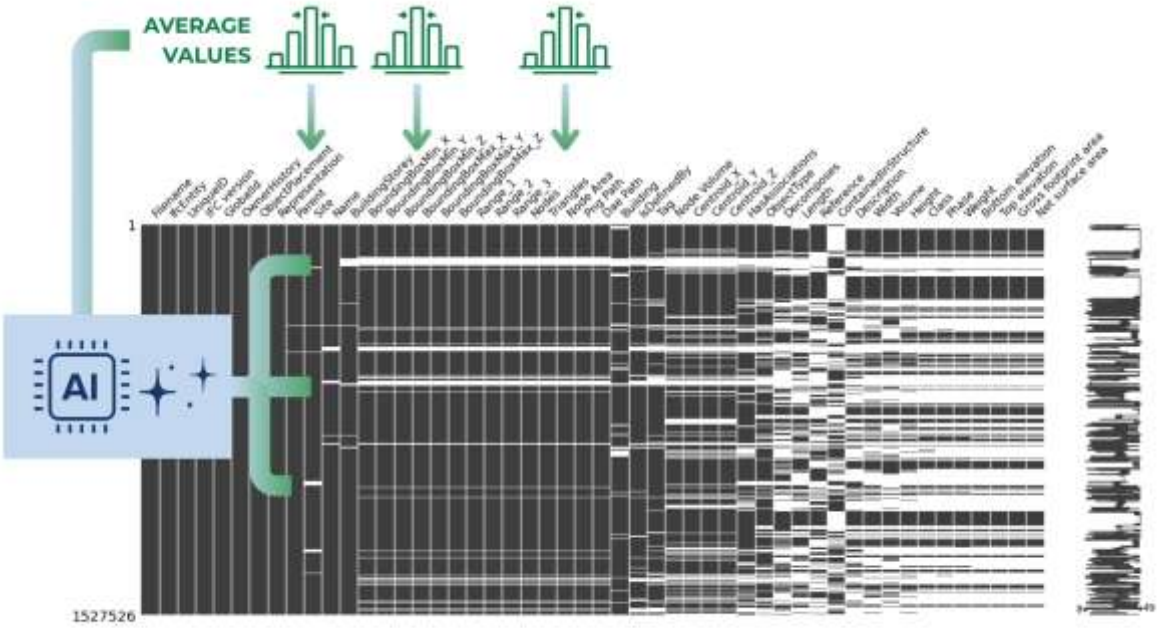


Abb. 10.1-6 Mit Hilfe des maschinellen Lernens können Sie automatisch Durchschnittswerte für nicht ausgefüllte (weiße Kästchen) Tabellenparameter auf der Grundlage früherer Projekte ermitteln.

> Vor dem Hintergrund dieser rasanten Fortschritte beim maschinellen Lernen ist klar: Es ist naiv zu glauben, dass das Problem der automatischen Klassifizierung von Bauelementen im Jahr 2025 noch lange ungelöst sein wird. Ja, die aktuellen Algorithmen sind noch nicht ganz ausgereift, insbesondere bei unvollständigen oder heterogenen Daten, aber das Zeitfenster für Anpassungen schließt sich rasch.

Unternehmen, die bereits in das Sammeln, Bereinigen und Organisieren ihrer Daten investieren und ETL-Automatisierungstools einsetzen, werden im Vorteil sein. Alle anderen laufen Gefahr, ins Hintertreffen zu geraten - so wie Unternehmen einst an den Herausforderungen der digitalen Transformation in der Transport- und Finanzbranche scheiterten

Wer weiterhin auf manuelle Datenverwaltung und traditionelle Methoden der Kosten- und Zeitkalkulation setzt, läuft Gefahr, sich in die Lage der Taxiflotten der 2000er Jahre zu versetzen, die nicht in der Lage sind, sich bis Anfang der 2020er Jahre an das Zeitalter der mobilen Apps und der automatischen Routenberechnung anzupassen.

KAPITEL 10.2.
EIN PRAKTISCHER LEITFADEN ZUR UMSETZUNG EINES DATENGESTÜTZTEN ANSATZES

Von der Theorie zur Praxis: Ein Fahrplan für die digitale Transformation im Bauwesen

Die Baubranche tritt allmählich in eine neue Entwicklungsphase ein, in der vertraute Prozesse zunehmend durch digitale Plattformen und transparente Interaktionsmodelle ergänzt und teilweise sogar ersetzt werden. Dies stellt die Unternehmen nicht nur vor Herausforderungen, sondern auch vor große Chancen. Diejenigen Organisationen, die bereits heute eine langfristige digitale Strategie aufbauen, werden ihre Position am Markt nicht nur halten, sondern auch ausbauen können, indem sie den Kunden moderne Ansätze und zuverlässige, technologisch gestützte Lösungen anbieten.

> Es ist wichtig zu erkennen, dass das Wissen über Konzepte und Technologien nur ein Ausgangspunkt ist. Manager und Fachleute stehen vor der praktischen Frage, wo sie mit der Umsetzung beginnen und wie sie theoretische Ideen in realen Wert umsetzen können. Darüber hinaus stellt sich zunehmend die Frage, worauf das Geschäft basieren soll, wenn die traditionellen Methoden der Kosten- und Zeitrechnung vom Kunden jederzeit geändert werden können.

Die Antwort liegt wahrscheinlich nicht so sehr in der Technologie, sondern in der Herausbildung einer neuen Berufskultur, in der die Arbeit mit Daten als integraler Bestandteil der täglichen Praxis wahrgenommen wird. Es ist die mangelnde Aufmerksamkeit für digitale Technologie und Innovation, die das Baugewerbe in den letzten Jahrzehnten in einen schweren Rückstand gebracht hat [43].

> Nach Angaben von McKinsey belaufen sich die FuE-Ausgaben im Baugewerbe auf weniger als 1% des Umsatzes, während diese Zahl in der Automobil- und Luftfahrtindustrie 3,5-4,5% erreicht. In ähnlicher Weise liegen die IT-Kosten im Baugewerbe bei weniger als 1% des Gesamtumsatzes [107].

Infolgedessen sinkt nicht nur der Automatisierungsgrad, sondern auch die Arbeitsproduktivität im Baugewerbe, und im Jahr 2020 wird der Bauarbeiter weniger produzieren als vor einem halben Jahrhundert (Abb. 10.2-1)

Derartige Produktivitätsprobleme im Bausektor sind in den meisten Industrie- und Entwicklungsländern anzutreffen (die Produktivität im Baugewerbe ist in 16 von 29 OECD-Ländern gesunken (Abb. 2.2-1)) und deuten nicht nur auf einen Mangel an Technologie hin, sondern auch auf die Notwendigkeit von systemischen Veränderungen in den Bereichen Management, Ausbildung und Innovation.

> Der Erfolg der digitalen Transformation hängt nicht so sehr von der Anzahl und der Verfügbarkeit von Tools ab, sondern von der Fähigkeit der Unternehmen, ihre Prozesse zu überprüfen und eine Kultur zu entwickeln, die offen für Veränderungen ist. Nicht die Technologie selbst ist der Schlüssel, sondern die Menschen und Prozesse, die ihre effektive Nutzung gewährleisten, kontinuierliches Lernen unterstützen und die Akzeptanz neuer Ideen fördern.

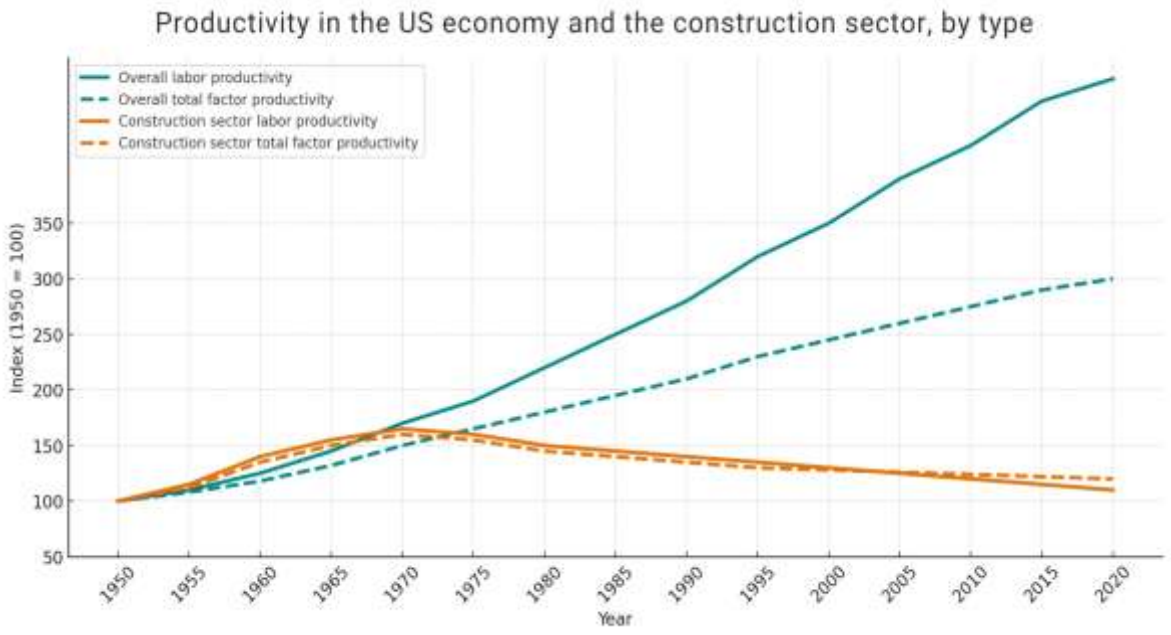


Abb. 10.2-1 Das Paradoxon der Arbeitsproduktivität und der Gesamtressourcenproduktivität in der US-Wirtschaft und im Bausektor (1950-2020) (basierend auf [43]).

In den ersten Teilen des Buches wurde das Modell des Unternehmensumfelds mit einem Waldökosystem verglichen (Abb. 2.1-2, Abbildung 1.2-4, Abbildung 1.3-2). In einem gesunden Wald spielen periodische Brände trotz ihrer zerstörerischen Kraft eine Schlüsselrolle bei der langfristigen Erneuerung. Sie befreien den Boden von alter Vegetation, geben gespeicherte Nährstoffe zurück und schaffen Platz für neues Leben. Einige Pflanzenarten haben sich sogar so entwickelt, dass sich ihre Samen nur öffnen, wenn sie hohen Feuertemperaturen ausgesetzt sind - ein natürlicher Mechanismus, der den idealen Zeitpunkt für die Keimung bietet.

In ähnlicher Weise können Krisen in der Wirtschaft wie ein "kontrollierter Burnout" wirken und das Entstehen neuer Ansätze und Unternehmen fördern, die nicht an veraltete Systeme gebunden sind. Solche Phasen zwingen dazu, ineffiziente Praktiken aufzugeben, wodurch Ressourcen für Innovationen frei werden. So wie ein Wald nach einem Brand mit Pionierpflanzen beginnt, so bildet auch ein Unternehmen nach einer Krise neue, flexible Prozesse, die die Grundlage für eine ausgereifte Informationsumgebung bilden.

Unternehmen, denen es gelingt, diese "Signalfeuer" richtig zu interpretieren und ihre disruptive Energie in konstruktive Veränderungen umzuwandeln, werden ein neues Leistungsniveau erreichen - mit transparenteren, anpassungsfähigeren Datenprozessen, die die natürliche Fähigkeit einer Organisation zur Erneuerung und zum Wachstum verbessern.

> Der wachsende Einfluss der künstlichen Intelligenz und des maschinellen Lernens auf das Geschäftsumfeld ist nicht mehr zu bezweifeln. Es handelt sich nicht nur um einen vorübergehenden Trend, sondern um eine strategische Notwendigkeit. Unternehmen, die KI ignorieren, riskieren den Verlust ihrer Wettbewerbsfähigkeit in einem Markt, der zunehmend Innovation und Flexibilität fördert.

Die Zukunft gehört denjenigen, die KI nicht nur als Werkzeug sehen, sondern als Chance, jeden Aspekt ihres Unternehmens zu überdenken - von der Optimierung der Prozesse bis hin zu Managemententscheidungen.

Legen des digitalen Fundaments: 1-5 Schritte zur digitalen Reife

In diesem Kapitel gehen wir auf die Roadmap für die digitale Transformation ein und zeigen die wichtigsten Schritte auf, die zur Umsetzung eines datengesteuerten Ansatzes erforderlich sind, der sowohl die Unternehmenskultur als auch das Informationsökosystem des Unternehmens verändern kann.

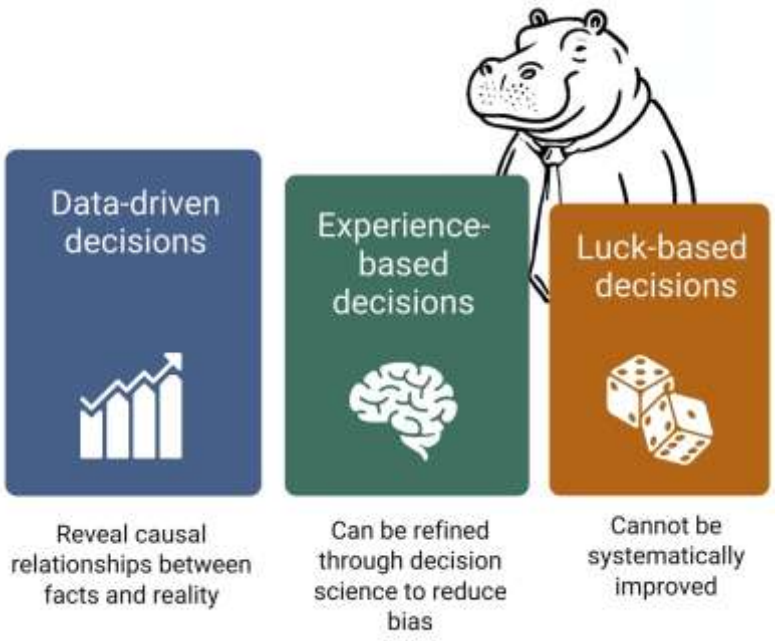


Abbildung 10.2-2 Kontrollierte Aktualisierung und Strategieauswahl: Fall, Erfahrung oder Daten.

Laut der McKinsey-Studie "Why Digital Strategies Fail" (2018) gibt es mindestens fünf Gründe [164], warum Unternehmen die Ziele der digitalen Transformation nicht erreichen

- **Unklare Definitionen**: Führungskräfte und Manager haben unterschiedliche Vorstellungen davon, was "digital" bedeutet, was zu Missverständnissen und Unstimmigkeiten führt.
- **Falsches Verständnis der digitalen Wirtschaft**: Viele Unternehmen unterschätzen das Ausmaß der Veränderungen, die die Digitalisierung für Geschäftsmodelle und Branchendynamik mit sich bringt (Abb. 10.1-6).
- **Ignorieren von Ökosystemen**: Unternehmen konzentrieren sich auf einzelne Technologielösungen (Datensilos) und übersehen die Notwendigkeit, sich in breitere digitale Ökosysteme zu integrieren (Abb. 2.2-2, Abbildung 4.1-12).
- **Unterschätzung der Digitalisierung durch die Konkurrenten**: Die Manager berücksichtigen nicht, dass auch die Konkurrenten aktiv digitale Technologien einsetzen, was zu einem Verlust von Wettbewerbsvorteilen führen kann.
- **Verpassen der Dualität der Digitalisierung**: CEOs delegieren die Verantwortung für die digitale

Transformation an andere Führungskräfte, was die Kontrolle bürokratisiert und den Veränderungsprozess verlangsamt.

Die Bewältigung dieser Herausforderungen erfordert ein klares Verständnis und eine Abstimmung der digitalen Strategien auf allen Ebenen des Unternehmens. Vor der Entwicklung einer digitalen Strategie ist es wichtig, die Ausgangssituation zu verstehen. Viele Organisationen neigen dazu, neue Tools und Plattformen einzuführen, ohne ein vollständiges Bild des aktuellen Zustands zu haben.

Schritt 1: Führen Sie ein Audit Ihrer aktuellen Systeme und Daten durch.

Vor der Änderung von Prozessen ist es wichtig zu verstehen, was bereits vorhanden ist. Die Durchführung eines Audits ermöglicht es Ihnen, Schwachstellen im Datenmanagement zu erkennen und zu verstehen, welche Ressourcen genutzt werden können. Ein Audit ist eine Art "Röntgenaufnahme" Ihrer Geschäftsprozesse. Sie ermöglicht es Ihnen, Risikobereiche zu identifizieren und festzustellen, welche Daten für Ihr Projekt oder Ihr Unternehmen von entscheidender und welche von untergeordneter Bedeutung sind.

Haupthandlungen:

- Stellen Sie Ihre IT-Umgebung dar (in Draw.io, Lucidchart, Miro, Visio oder Canva). Führen Sie die verwendeten Systeme auf (ERP, CAD, CAFM, CPM, SCM und andere), die an Ihren Prozessen beteiligt sind und die wir im Kapitel "Technologien und Managementsysteme im modernen Bauwesen" besprochen haben (Abb. 1.2-4)
- Bewerten Sie die Datenqualität für jedes System im Hinblick auf die Häufigkeit von Duplikaten, möglichen fehlenden Werten und Formatinkonsistenzen in jedem System.
- Identifizieren Sie "Schmerzpunkte" - Stellen, an denen Prozesse zusammenbrechen können oder häufig manuelle Eingriffe erfordern - Importe, Exporte und zusätzliche Validierungsprozesse.

> Wenn Sie wollen, dass das Team den Berichten vertraut, müssen Sie sicherstellen, dass die Daten von Anfang an korrekt sind.

Ein Audit der Datenqualität zeigt Ihnen, welche Daten:

- Weiterer Entwicklungsbedarf (automatische Reinigungsprozesse oder zusätzliche Transformationen müssen eingerichtet werden)
- Sie sind "Müll", der nur die Systeme verstopft und den man loswerden kann, indem man sie nicht mehr in Prozessen verwendet.

Es ist möglich, ein solches Audit selbst durchzuführen. Manchmal ist es jedoch sinnvoll, einen externen Berater - vor allem aus anderen Branchen - hinzuzuziehen: Ein frischer Blickwinkel und die Unabhängigkeit von den "Eigenheiten" des Bauwesens helfen dabei, den Status quo nüchtern zu bewerten und die typischen Fallstricke der Voreingenommenheit gegenüber bestimmten Lösungen und Technologien zu vermeiden.

Schritt 2: Ermittlung der wichtigsten Standards für die Datenharmonisierung.

Nach dem Audit müssen gemeinsame Regeln für die Arbeit mit den Daten aufgestellt werden. Wie wir im Kapitel "Standards: Von zufälligen Dateien zu einem intelligenten Datenmodell" erörtert haben, wird dies dazu beitragen, siloartige Datenflüsse zu beseitigen.

> Ohne einen einheitlichen Standard wird jedes Team weiterhin "auf seine Weise" arbeiten, und Sie werden einen "Zoo" von Integrationen aufrechterhalten, in dem bei jeder Konvertierung Daten verloren gehen.

Haupthandlungen:

- Wählen Sie die Datenstandards für den Informationsaustausch zwischen Systemen:
  - Für tabellarische Daten können dies strukturierte Formate wie CSV, XLSX oder effizientere Formate wie Parquet sein.
  - Für den Austausch von lose strukturierten Daten und Dokumenten: JSON oder XML
- Beherrschen Sie die Arbeit mit Datenmodellen:
  - Beginnen Sie mit der Parametrisierung der Aufgaben auf der Ebene des konzeptionellen Datenmodells - wie im Kapitel "Datenmodellierung: konzeptionelles, logisches und physisches Modell" beschrieben (Abb. 4.3-2)
  - Wenn Sie tiefer in die Logik des Geschäftsprozesses eindringen, gehen Sie dazu über, die Anforderungen mithilfe von Parametern in den logischen und physischen Modellen zu formalisieren (Abb. 4.3-6).
  - Identifizieren Sie die wichtigsten Entitäten, ihre Attribute und Beziehungen innerhalb von Prozessen, und visualisieren Sie diese Beziehungen - sowohl zwischen Entitäten als auch zwischen Parametern (Abb. 4.3-7)
- Verwenden Sie reguläre Ausdrücke (RegEx), um Daten zu validieren und zu standardisieren (Abb. 4.4-7), wie wir im Kapitel "Strukturierte Anforderungen und reguläre Ausdrücke RegEx" besprochen haben. RegEx ist kein komplexes, aber ein äußerst wichtiges Thema bei der Erstellung von Anforderungen auf der Ebene der physischen Datenmodelle.

Ohne Standards auf Datenebene und Prozessvisualisierung ist es unmöglich, eine konsistente und skalierbare digitale Umgebung zu schaffen. Denken Sie daran: "Schlechte Daten sind teuer." Und die Kosten von Fehlern steigen, je komplexer ein Projekt oder eine Organisation wird. Die Vereinheitlichung von Formaten, die Definition von Benennungs-, Struktur- und Validierungsregeln ist eine Investition in die Stabilität und Skalierbarkeit künftiger Lösungen.

Schritt 3: Implementieren Sie DataOps und automatisieren Sie Prozesse.

Ohne eine klar definierte Architektur werden Unternehmen unweigerlich mit uneinheitlichen Daten in isolierten Informationssystemen konfrontiert sein. Die Daten sind nicht integriert, werden an mehreren Orten dupliziert und sind teuer in der Pflege.

> Stellen Sie sich vor, dass Daten Wasser sind und die Datenarchitektur das komplexe System von Rohrleitungen ist, das dieses Wasser von der Speicherquelle bis zum Verwendungsort transportiert. Es ist die Datenarchitektur, die bestimmt, wie Informationen gesammelt, gespeichert, umgewandelt, analysiert und an die Endnutzer oder Anwendungen geliefert werden.

DataOps (Data Operations) ist eine Methodik, die die Sammlung, Bereinigung, Validierung und Verwendung

von Daten in einen einzigen automatisierten Prozessablauf integriert, wie wir in Teil 8 des Buches ausführlich besprochen haben.

Haupthandlungen:

- Erstellen und Anpassen von ETL -Förderern zur Automatisierung von Prozessen:
  - Extract: Organisieren Sie die automatische Datenerfassung aus PDF-Dokumenten (Abb. 4.1-2, Abb. 4.1-5, Abb. 4.1-7), Excel-Tabellen, CAD-Modellen (Abb. 7.2-4), ERP-Systemen und anderen Quellen, mit denen Sie arbeiten.
  - Transform: Einrichtung automatischer Prozesse zur Umwandlung von Daten in ein einheitliches strukturiertes Format und zur Automatisierung von Berechnungen, die außerhalb geschlossener Anwendungen durchgeführt werden (Abb. 7.2-8)
  - Laden Sie: Versuchen Sie, einen automatischen Daten-Upload in Übersichtstabellen, Dokumente oder zentrale Repositories zu erstellen (Abb. 7.2-9, Abb. 7.2-13, Abb. 7.2-16)
- Automatisieren Sie die Berechnungs- und QTO (Quantity Take-Off) Prozesse, wie im Kapitel "QTO Quantity Take-Off: Gruppierung von Projektdaten nach Attributen" beschrieben:
  - Richten Sie die automatische Extraktion von Volumina aus CAD-Modellen ein, indem Sie APIs, Plug-ins oder Reverse-Engineering-Tools verwenden (Abb. 5.2-5).
  - Erstellen Sie Regeln für die Gruppierung von Elementen für verschiedene Klassen nach Attributen in Form von Tabellen (Abb. 5.2-12)
  - Versuchen Sie, häufig wiederkehrende Mengen- und Kostenberechnungen außerhalb von modularen geschlossenen Systemen zu automatisieren (Abb. 5.2-15)
- Verwenden Sie Python und Pandas, um Daten zu verarbeiten, wie im Kapitel "Python Pandas: ein unverzichtbares Werkzeug für die Arbeit mit Daten" beschrieben:
  - Wenden Sie DataFrame an, um mit XLSX-Dateien zu arbeiten und die Verarbeitung von Tabellendaten zu automatisieren (Abb. 3.4-6)
  - Automatisieren Sie die Aggregation und Transformation von Informationen durch verschiedene Python-Bibliotheken
  - Verwenden Sie den LLM, um das Schreiben vorgefertigter Codeblöcke und ganzer Pipelines zu vereinfachen (Abb. 7.2-18)
  - Versuchen Sie, eine Pipeline in Python zu erstellen, die Fehler oder Anomalien findet und eine Benachrichtigung an die zuständige Person (z. B. den Projektleiter) sendet (Abb. 7.4-2).

Die Automatisierung auf der Grundlage von DataOps-Prinzipien ermöglicht Ihnen den Übergang von der manuellen und fragmentierten Datenverarbeitung zu nachhaltigen und wiederholbaren Prozessen. Dadurch werden nicht nur die Mitarbeiter entlastet, die sich täglich mit denselben Transformationen befassen, sondern auch die Zuverlässigkeit, Skalierbarkeit und Transparenz des gesamten Informationssystems drastisch erhöht.

Schritt 4: Schaffung eines offenen Ökosystems für die Datenverwaltung.

Trotz der Entwicklung geschlossener modularer Systeme und ihrer Integration mit neuen Werkzeugen stehen die Unternehmen vor einem ernsten Problem: Die zunehmende Komplexität dieser Systeme übersteigt ihren Nutzen. Die ursprüngliche Idee, eine einzige proprietäre Plattform zu schaffen, die alle Geschäftsprozesse abdeckt, hat zu einer übermäßigen Zentralisierung geführt, bei der jede Änderung erhebliche Ressourcen und Zeit für die Anpassung erfordert.

Wie wir im Kapitel "Unternehmensmyzel: Wie Daten Geschäftsprozesse verbinden" erörtert haben, erfordert eine effektive Datenverwaltung ein offenes und einheitliches Ökosystem, das alle Informationsquellen miteinander verbindet.

Schlüsselelemente des Ökosystems:

- Wählen Sie einen geeigneten Datenspeicher:
  - Für Tabellen und Berechnungen verwenden Sie Datenbanken - zum Beispiel PostgreSQL oder MySQL (Abb. 3.1-7)
  - Für Dokumente und Berichte können Cloud-Speicher (Google Drive, OneDrive) oder Systeme, die das JSON-Format unterstützen, geeignet sein.
  - Informieren Sie sich über die Möglichkeiten von Data Warehouse, Data Lakes und anderen Tools zur zentralen Speicherung und Analyse großer Informationsmengen (Abb. 8.1-8)
- Implementierung von Lösungen für den Zugang zu geschützten Daten:
  - Wenn Sie proprietäre Systeme verwenden, konfigurieren Sie den Zugang zu diesen Systemen über API oder SDK, um Daten für die externe Verarbeitung zu erhalten (Abb. 4.1-2)
  - Machen Sie sich mit den Möglichkeiten von Reverse-Engineering-Tools vertraut für CAD-Formate (Abb. 4.1-13)
  - Einrichtung von ETL-Pipelines, die in regelmäßigen Abständen Daten aus Anwendungen oder Servern sammeln, sie in offene, strukturierte Formate konvertieren und in Repositories speichern (Abb. 7.2-3)
  - innerhalb des Teams zu diskutieren, wie der Zugang zu Daten ohne proprietäre Software ermöglicht werden kann
  - Denken Sie daran: Daten sind wichtiger als Schnittstellen. Es sind die Struktur und die Verfügbarkeit von Informationen, nicht spezifische Werkzeuge für die Benutzeroberfläche, die einen langfristigen Wert schaffen
- Überlegen Sie, ob Sie ein Kompetenzzentrum (CoE) für Daten einrichten wollen, wie wir es im Kapitel "Kompetenzzentrum (CoE) für Datenmodellierung" besprochen haben, oder wie Sie auf andere Weise Datenexpertise bereitstellen können (Abb. 4.3-9)

Das Datenmanagement-Ökosystem schafft einen einheitlichen Informationsraum, in dem alle Projektbeteiligten mit konsistenten, aktuellen und verifizierten Informationen arbeiten. Es ist die Basis für skalierbare, flexible und zuverlässige digitale Prozesse

Das Potenzial von Daten freisetzen: 5-10 Schritte zur digitalen Reife

Neben der technischen Integration ist ein wichtiger Faktor für die erfolgreiche Umsetzung digitaler Lösun-

gen deren Akzeptanz durch die Endnutzer. Die Einbindung von Kunden oder Nutzern in die Leistungsmessung ist sowohl eine Herausforderung für die Verbesserung der Nutzererfahrung als auch für das Management des Wandels im Unternehmen. Wenn eine Lösung nicht in einen vertrauten Arbeitsablauf passt oder keine wirklichen Benutzer- oder Kundenprobleme löst, wird sie nicht genutzt, und keine noch so großen zusätzlichen Maßnahmen und Anreize werden das ändern.

Die Umwandlung ist ein iterativer Prozess, der auf der Analyse von Benutzerinteraktionsdaten mit neuen Prozessen basiert, mit häufigen Testzyklen, ständigem Feedback und Verfeinerungen.

Schritt 5: Aufbau einer Datenkultur, Schulung der Mitarbeiter und Sammlung von Feedback

Selbst das fortschrittlichste System funktioniert nicht ohne das Engagement der Mitarbeiter. Sie müssen ein Umfeld schaffen, in dem die Daten täglich genutzt werden und das Team ihren Wert versteht.

> Der von der britischen Regierung im Jahr 2024 veröffentlichte Bericht "Data Analytics and AI in Government Project Delivery" [83] stellt fest, dass die Ausbildung von Fachkräften mit den erforderlichen Kompetenzen in der Datenverarbeitung und -interpretation für die erfolgreiche Umsetzung von data analytics and AI entscheidend ist.

Der Mangel an Fachwissen im Bereich der Datenanalyse ist eines der Hauptprobleme, die den digitalen Wandel behindern. Führungskräfte sind an etablierte Routinen gewöhnt: vierteljährliche Zyklen, nach Prioritäten geordnete Initiativen und traditionelle Wege, Projekte voranzutreiben. Veränderungen erfordern eine besondere Führungspersönlichkeit - hoch genug im Rang, um Einfluss zu haben, aber nicht so hoch, dass sie die Zeit und Motivation hat, ein langfristiges Transformationsprojekt zu leiten.

Haupthandlungen:

- Erkennen der Notwendigkeit, von subjektiven Entscheidungen, die auf der Meinung eines hochbezahlten Mitarbeiters beruhen (HiPPO), zu einer Entscheidungskultur überzugehen, die auf Fakten und Daten beruht, wie im Kapitel "HiPPO oder die Gefahr der Meinung bei der Entscheidungsfindung" (Abb. 2.1-9) erörtert.
- Organisieren Sie systematische Schulungen:
  - Durchführung von Schulungen zur Nutzung strukturierter Daten und Einladung von Experten aus anderen Branchen, die nicht voreingenommen gegenüber Produkten und Konzepten sind, die heute in der Bauindustrie verbreitet sind
  - Diskutieren Sie mit Kollegen über Datenanalyseansätze und -tools und erlernen Sie selbstständig die praktische Arbeit mit Tools wie Python, Pandas und LLM (Abb. 4.1-3, Abb. 4.1-6)
  - Erstellen einer Bibliothek von Tutorials (vorzugsweise mit kurzen Videos) zum Thema Datenstrukturierung (Abb. 3.2-15) und Erstellung von Datenmodellen (Abb. 4.3-7)
- Nutzung moderner Lerntechnologien:
  - Verwendung von Sprachmodellen (LLMs) zur Unterstützung von Code- und Datenmanipulationen, einschließlich Codegenerierung, Refactoring und Analyse, sowie zur Verarbeitung und Interpretation von Tabelleninformationen (Abb. 3.4-1)
  - Erkunden Sie, wie LLM-generierter Code angepasst und in eine vollständige Pipeline-Lösung integriert werden kann, wenn Sie in einer Offline-Entwicklungsumgebung (IDE) arbeiten (Abb. 4.4-14, Abb. 5.2-13)

> Wenn ein Manager seine Entscheidungen weiterhin auf die "altmodische Art" trifft, kann keine noch so gute Schulung die Mitarbeiter davon überzeugen, die Analytik ernst zu nehmen.

Der Aufbau einer Datenkultur ist ohne kontinuierliches Feedback unmöglich. Durch Feedback können Sie Lücken in Prozessen, Werkzeugen und Strategien erkennen, die sich nicht durch interne Berichte oder formelle KPIs Metriken aufdecken lassen. Komplementäre Kommentare von Nutzern Ihrer Lösungen bieten keinen praktischen Nutzen. Gerade kritisches Feedback ist wertvoll, vor allem wenn es auf konkreten Beobachtungen und Fakten beruht. Die Beschaffung solcher Informationen ist jedoch mühsam: Sie müssen Prozesse aufbauen, in denen die Teilnehmer - sowohl interne als auch externe - ihre Kommentare (ggf. auch anonym) ohne Verzerrung und ohne Angst, dass sich ihre Meinung auf ihre eigene Arbeit auswirken könnte, mitteilen können. Es ist wichtig, dass sie dies ohne Verzerrung und ohne Angst vor negativen Konsequenzen für sich selbst tun.

> *Jedes Lernen ist letztlich selbstlernend [165].*
>
> - Milton Friedman, amerikanischer Wirtschaftswissenschaftler und Statistiker

Die Einführung von Analyseinstrumenten sollte mit einer regelmäßigen Überprüfung ihrer Wirksamkeit in der Praxis einhergehen (ROI, KPIs), was nur durch strukturiertes Feedback von Mitarbeitern, Kunden und Partnern erreicht werden kann. Auf diese Weise können Unternehmen nicht nur die Wiederholung von Fehlern vermeiden, sondern sich auch schneller an Veränderungen im Umfeld anpassen. Das Vorhandensein eines Mechanismus zur Sammlung und Analyse von Feedback ist eines der Zeichen für die Reife einer Organisation, die sich von gelegentlichen digitalen Initiativen zu einem nachhaltigen Modell der kontinuierlichen Verbesserung entwickelt (Abb. 2.2-5).

Schritt 6: Von Pilotprojekten zur Ausweitung der Aktivitäten

> *Wähle Schlachten, die groß genug sind, um wichtig zu sein, und klein genug, um zu gewinnen.*
>
> - Jonathan Kozol

Die digitale Transformation "auf einmal und überall" zu starten, ist äußerst riskant. Ein effektiverer Ansatz ist es, mit Pilotprojekten zu beginnen und erfolgreiche Praktiken schrittweise auszubauen.

Haupthandlungen:

- Wählen Sie das richtige Projekt für das Pilotprojekt:
  - ☐ Definieren Sie ein spezifisches Geschäftsziel oder einen Prozess mit messbaren Ergebnissen (KPI, ROI) (Abb. 7.1-5)
  - ☐ Wählen Sie einen ETL-Automatisierungsprozess, z. B. automatische Datenvalidierung oder Workload-Berechnung (QTO) mit Python und Pandas (Abb. 5.2-10

    - Festlegung klarer Erfolgskennzahlen (z. B. Verkürzung der Zeit für die Erstellung von Prüfspezifikationen oder Datenvalidierungsberichten von einer Woche auf einen Tag)
- Wählen Sie iterative Ansätze:
    - Beginnen Sie mit einfachen Datenkonvertierungsprozessen und erstellen Sie Streaming-Konvertierungen von Multiformat-Daten in die Formate, die Sie für Ihre Prozesse benötigen (Abb. 4.1-2, Abb. 4.1-5)
    - Erhöhen Sie schrittweise die Komplexität der Aufgaben und erweitern Sie die Automatisierung der Prozesse, indem Sie eine vollständige Pipeline in der IDE auf der Grundlage von dokumentierten Codeblöcken bilden (Abb. 4.1-7, Abb. 7.2-18).
    - Dokumentieren und dokumentieren Sie (vorzugsweise mit kurzen Videos) erfolgreiche Lösungen und teilen Sie diese mit Kollegen oder in professionellen Gemeinschaften
- Entwickeln Sie Vorlagen und begleitende Dokumentation, um solche Lösungen zu vervielfältigen, damit sie von Ihren Kollegen (oder Mitgliedern der Berufsgemeinschaft, einschließlich der Nutzer sozialer Medien) effektiv genutzt werden können.

Ein schrittweiser "Roll-up" ermöglicht es Ihnen, die hohe Qualität der Änderungen beizubehalten und nicht in das Chaos paralleler Implementierungen zu verfallen. Die Strategie "vom Kleinen zum Großen" minimiert die Risiken und ermöglicht es Ihnen, aus kleinen Fehlern zu lernen, ohne sie zu kritischen Problemen werden zu lassen.

Die Umstellung von einem Projektansatz, bei dem die Mitarbeiter nur teilweise einbezogen werden, auf die Bildung ständiger Teams (z. B. Kompetenzzentren - CoE) trägt dazu bei, die nachhaltige Entwicklung des Produkts auch nach der Freigabe seiner ersten Version sicherzustellen. Solche Teams unterstützen nicht nur bestehende Lösungen, sondern verbessern sie auch weiter.

Dadurch verringert sich die Abhängigkeit von langwierigen Genehmigungen: Die Teammitglieder sind befugt, Entscheidungen in ihrem Verantwortungsbereich zu treffen. Infolgedessen sind die Manager von der Notwendigkeit befreit, Mikromanagement zu betreiben, und die Teams können sich auf die Schaffung echter Werte konzentrieren.

> Die Entwicklung neuer Lösungen ist kein Sprint, sondern ein Marathon. Erfolgreich sind diejenigen, die zunächst auf eine langfristige und konsequente Arbeit setzen.

Es ist wichtig zu erkennen, dass die Technologie eine ständige Weiterentwicklung erfordert. Die Investition in die langfristige Entwicklung technologischer Lösungen ist die Grundlage für eine erfolgreiche Tätigkeit.

Schritt 7: Verwendung offener Datenformate und Lösungen

Wie in den Kapiteln über modulare Plattformen (ERP, PMIS, CAFM, CDE usw.) erörtert, ist es wichtig, sich auf offene und universelle Datenformate zu konzentrieren, die die Unabhängigkeit von Anbieterlösungen gewährleisten und die Verfügbarkeit von Informationen für alle Beteiligten verbessern.

Haupthandlungen:

- Übergang von geschlossenen Formaten zu offenen Formaten:
    - Verwenden Sie offene Formate anstelle von proprietären Formaten oder finden Sie einen

Weg, um einen automatischen Upload oder eine Konvertierung von geschlossenen Formaten in offene Formate einzurichten (Abb. 3.2-15).

- ☐ Implementierung von Werkzeugen für die Arbeit mit Parquet, CSV, JSON, XLSX, die Austauschstandards zwischen den meisten modernen Systemen sind (Abb. 8.1-2)
- ☐ Wenn die Arbeit mit 3D-Geometrie eine wichtige Rolle in Ihren Prozessen spielt, sollten Sie die Verwendung offener Formate wie USD, glTF, DAE oder OBJ in Betracht ziehen (Abb. 3.1-14).

- Nutzung von Vektordatenbanken Datenbanken zur effizienten Analyse und Suche nach Informationen:
  - ☐ Verwenden Sie Bounding Box und andere Methoden zur Vereinfachung der 3D -Geometrie (Abb. 8.2-1)
  - ☐ Überlegen Sie, wo Sie die Vektorisierung von Daten - die Umwandlung von Texten, Objekten oder Dokumenten in numerische Darstellungen - einsetzen können (Abb. 8.2-2)
- Anwendung von Big-Data-Analysetools:
  - ☐ Organisieren Sie die Speicherung der gesammelten historischen Daten (z.B. PDF, XLSX, CAD) in geeigneten Formaten für die Analyse (Apache Parquet, CSV, ORC) (Abb. 8.1-2).
  - ☐ Beginnen Sie mit der Anwendung grundlegender statistischer Methoden und arbeiten Sie mit repräsentativen Stichproben - oder machen Sie sich zumindest mit den grundlegenden Prinzipien der Statistik vertraut (Abb. 9.2-5)
  - ☐ Implementierung und Erlernen von Werkzeugen zur Visualisierung von Daten und der Beziehungen zwischen Daten, um die Ergebnisse der Analyse zu visualisieren. Ohne eine gute Visualisierung ist es unmöglich, die Daten selbst oder die auf ihnen basierenden Prozesse vollständig zu verstehen (Abb. 7.1-4).

Die Umstellung auf offene Datenformate und die Einführung von Werkzeugen zur Analyse, Speicherung und Visualisierung von Informationen schafft die Grundlage für eine nachhaltige und unabhängige digitale Verwaltung. Dies verringert nicht nur die Abhängigkeit von Anbietern, sondern gewährleistet auch einen gleichberechtigten Zugang zu Daten für alle Beteiligten.

Schritt 8: Beginn der Implementierung von maschinellem Lernen für Vorhersagen

In vielen Unternehmen haben sich riesige Datenmengen angesammelt - eine Art "Informationsgeysir", die noch ungenutzt sind. Diese Daten wurden in Hunderten oder Tausenden von Projekten gesammelt, aber oft nur einmal oder gar nicht in weiteren Prozessen verwendet. Dokumente und Modelle, die in geschlossenen Formaten und Systemen gespeichert sind, werden oft als veraltet und unbrauchbar empfunden. In Wirklichkeit sind sie jedoch eine wertvolle Ressource - die Grundlage für die Analyse von Fehlern, die Automatisierung von Routinevorgängen und die Entwicklung innovativer Lösungen für die automatische Klassifizierung und Merkmalserkennung in zukünftigen Projekten.

Die größte Herausforderung besteht darin, zu lernen, wie man diese Daten extrahiert und in verwertbare Erkenntnisse umwandelt. Wie im Kapitel über maschinelles Lernen und Vorhersagen erörtert, haben Techniken des maschinellen Lernens das Potenzial, die Genauigkeit von Schätzungen und Vorhersagen in einer Vielzahl von baubezogenen Prozessen erheblich zu verbessern. Die vollständige Nutzung der gesammelten Daten eröffnet die Möglichkeit, die Effizienz zu verbessern, Risiken zu verringern und nachhaltige digitale

Prozesse aufzubauen.

Haupthandlungen:

- Beginnen Sie mit einfachen Algorithmen:
  - Versuchen Sie, die lineare Regression - unter Verwendung von Hinweisen aus dem LLM - anzuwenden, um die wiederkehrende Leistung in Datensätzen vorherzusagen, in denen Abhängigkeiten von einer großen Anzahl von Faktoren nicht oder nur minimal vorhanden sind (Abb. 9.3-4)
  - Überlegen Sie, in welchen Phasen Ihrer Prozesse der k-NN-Algorithmus (k-Nearest Neighbours) theoretisch eingesetzt werden könnte - zum Beispiel für Klassifizierungsaufgaben, die Bewertung der Objektähnlichkeit oder die Vorhersage auf der Grundlage historischer Analogien (Abb. 9.3-5).
- Sammeln und strukturieren Sie Daten, um Modelle zu trainieren:
  - Sammeln Sie historische Projektdaten an einem Ort und in einem einzigen Format (Abb. 9.1-10)
  - Arbeit an der Qualität und Repräsentativität der Trainingsmuster durch automatische ETL (Abb. 9.2-8)
  - Lernen Sie, die Daten in Trainings- und Testdatensätze aufzuteilen, wie wir es im Beispiel des Titanic-Datensatzes getan haben (Abb. 9.2-6, Abbildung 9.2-7)
- Erwägung der Ausweitung der Anwendung von Techniken des maschinellen Lernens auf ein breites Spektrum von Aufgaben, von der Vorhersage von Projektfristen bis zur Optimierung der Logistik, des Ressourcenmanagements und der Früherkennung potenzieller Probleme

Maschinelles Lernen ist ein Werkzeug, mit dem archivierte Daten in einen wertvollen Vermögenswert für Vorhersagen, Optimierung und fundierte Entscheidungen verwandelt werden können. Beginnen Sie mit kleinen Datensätzen (Abb. 9.2-5) und einfachen Modellen und steigern Sie allmählich die Komplexität.

Schritt 9: Integration des IoT und moderner Datenerfassungstechnologien

Die Welt des Bauens wird immer digitaler: Jedes Baufoto, jeder Beitrag des Teams ist bereits Teil eines größeren Prozesses der Parametrisierung und Tokenisierung der Realität. So wie GPS einst die Logistik veränderte, verändern IoT, RFID und automatische Datenerfassungstechnologien die Bauindustrie. Wie im Kapitel "IoT Internet der Dinge und intelligente Verträge" erörtert, ist die digitale Baustelle mit Sensoren und automatischer Überwachung die Zukunft der Branche.

Haupthandlungen:

- Implementieren Sie IoT -Geräte, RFID -Etiketten und beschreiben Sie die damit verbundenen Prozesse:
  - Evaluieren Sie, in welchen Bereichen oder Projektphasen eines Projekts der größte Return on Investment (ROI) erzielt werden kann - zum Beispiel bei der Überwachung von Temperatur, Vibration, Feuchtigkeit oder
  - Erwägen Sie den Einsatz von RFID zur Verfolgung von Materialien, Werkzeugen und Geräten in der gesamten Lieferkette

- ☐ Überlegen Sie, wie die gesammelten Daten in ein einziges Informationssystem, wie z. B. Apache NiFi, für die automatische Echtzeitverarbeitung und -analyse integriert werden können (Abb. 7.4-5)

- Einrichtung eines Echtzeit-Überwachungssystems:
  - ☐ Entwicklung von Dashboards zur Verfolgung wichtiger Prozess- oder Projektmetriken unter Verwendung von Visualisierungstools wie Streamlit, Flask oder Power BI)
  - ☐ Richten Sie automatische Benachrichtigungen ein, um kritische Abweichungen vom Plan oder von Normen zu signalisieren (Abb. 7.4-2)
  - ☐ Bewertung des Potenzials für eine vorausschauende Wartung der Ausrüstung auf der Grundlage der gesammelten Daten und der identifizierten Muster (Abb. 9.3-6)
- Kombinieren Sie Daten aus verschiedenen Quellen:
  - ☐ Beginnen Sie mit der Visualisierung des Datenmodells auf der physischen Ebene - spiegeln Sie die Struktur der Informationsflüsse und Schlüsselparameter aus CAD-Systemen, IoT-Geräten und ERP-Plattformen wider (Abb. 4.3-1)
  - ☐ Beginnen Sie mit der Erstellung einer groben Skizze einer einheitlichen Plattform für die Datenanalyse und die Unterstützung von Managemententscheidungen. Erfassen Sie Schlüsselfunktionen, Datenquellen, Benutzer und voraussichtliche Anwendungsszenarien (Abb. 4.3-7).

Je eher Sie damit beginnen, reale Prozesse mit der digitalen Welt zu verbinden, desto eher können Sie sie mit Daten steuern - effizient, transparent und in Echtzeit.

Schritt 10. Bereiten Sie sich auf zukünftige Veränderungen in der Branche vor

Bauunternehmen stehen unter ständigem Druck des äußeren Umfelds: Wirtschaftskrisen, Technologiesprünge, Änderungen der Rechtsvorschriften. Wie ein Wald, der Regen, Schnee, Dürre und sengender Sonne trotzen muss, leben Unternehmen unter Bedingungen der ständigen Anpassung. Und so wie Bäume durch ein tiefes Wurzelsystem widerstandsfähig gegen Frost und Trockenheit werden, so bleiben nur diejenigen Organisationen lebens- und wettbewerbsfähig, die über ein solides Fundament automatisierter Prozesse, die Fähigkeit zur Antizipation von Veränderungen und flexible Anpassungsstrategien verfügen.

Wie im Kapitel "Überlebensstrategien: Aufbau von Wettbewerbsvorteilen" erwähnt, tritt die Bauindustrie in eine Phase des radikalen Wandels ein. Die Interaktion zwischen Auftraggeber und Auftragnehmer bewegt sich in Richtung eines Modells der Überzeugung, bei dem Transparenz, Vorhersehbarkeit und digitale Werkzeuge die traditionellen Ansätze ersetzen. In dieser neuen Realität gewinnen nicht die größten, sondern die flexibelsten und technologisch ausgereiftesten Unternehmen.

Haupthandlungen:

- Analyse der Schwachstellen von Unternehmen im Zusammenhang mit offenen Daten:
  - ☐ Beurteilen Sie, inwiefern die Demokratisierung des Datenzugangs im Rahmen der Uberisierung verheerende Auswirkungen auf Ihren Wettbewerbsvorteil und Ihr Unternehmen haben könnte (Abb. 10.1-5)
  - ☐ Überlegen Sie sich eine Strategie für den Übergang von undurchsichtigen und isolierten Prozessen zu Geschäftsmodellen, die auf offenen Lösungen, Systeminteroperabilität und

Datentransparenz beruhen (Abb. 2.2-5).

- Entwickeln Sie eine langfristige digitale Strategie:
  - Bestimmen Sie, ob Sie eine führende Rolle bei der Innovation anstreben oder ein "Aufholszenario" bevorzugen, bei dem Sie Ihre Ressourcen schonen werden
  - Beschreiben Sie die Etappen: kurzfristig (Prozessautomatisierung, Datenstrukturierung), mittelfristig (Implementierung von LLM und ETL), langfristig (digitale Ökosysteme, zentralisierte Repositories)
- Denken Sie über eine Erweiterung Ihres Dienstleistungsangebots nach:
  - Erwägen Sie das Angebot neuer Dienstleistungen (mit Schwerpunkt auf Energieeffizienz, ESG, Datendienste). Wir werden im nächsten Kapitel mehr über neue Geschäftsmodelle sprechen
  - Positionieren Sie sich als zuverlässiger Technologiepartner, der den gesamten Lebenszyklus einer Anlage von der Planung bis zum Betrieb unterstützt. Das Vertrauen in Sie sollte auf einem systematischen Ansatz, transparenten Prozessen und der Fähigkeit beruhen, nachhaltige technologische Lösungen anzubieten.

In einem sich wandelnden Umfeld gewinnen nicht diejenigen, die einfach nur auf den Wandel reagieren, sondern diejenigen, die proaktiv sind. Flexibilität, Offenheit und digitale Reife sind die Grundlagen für Nachhaltigkeit beim Aufbau von morgen.

Transformationsfahrplan: vom Chaos zum datengesteuerten Unternehmen

Der folgende Plan kann als erste Richtschnur dienen - ein Ausgangspunkt für die Gestaltung Ihrer eigenen datengesteuerten digitalen Transformationsstrategie:

- **Audit und Standards**: Analyse des aktuellen Stands, Vereinheitlichung der Daten
- **Datenstrukturierung und -klassifizierung**: Automatisieren Sie die Umwandlung unstrukturierter und lose strukturierter Daten
- **Automatisieren Sie Gruppierungen**, Berechnungen und Kalkulationen: Verwenden Sie Open-Source-Tools und -Bibliotheken für die Automatisierung
- **Ökosystem und COE**: Aufbau eines internen Teams zur Bildung eines einheitlichen Datenökosystems im Unternehmen
- **Kultur und Lernen**: Weg von HiPPO -Lösungen hin zu datengesteuerten Lösungen
- **Pilotprojekte, Feedback und Skalierung**: iterativ vorgehen: neue Methoden in begrenztem Umfang testen, gültiges Feedback einholen und Lösungen schrittweise ausbauen.
- **Offene Formate**: Verwendung universeller und offener Formate zur Unabhängigkeit von Softwareanbietern
- **Maschinelles Lernen**: Einbettung von ML-Algorithmen in Prozesse zur Vorhersage und Optimierung
- **IoT und die digitale Baustelle**: Moderne Datenerfassungstechnologien in Prozesse integrieren
- **Strategische Anpassung**: Vorbereitung auf künftige Veränderungen in der Branche

> Denken Sie vor allem daran, dass "Daten allein ein Unternehmen nicht verändern: Es sind die Menschen, die wissen, wie man mit diesen Daten arbeitet, die es verändern". Legen Sie Wert auf Kultur, transparente Prozesse und die Verpflichtung zu kontinuierlicher Verbesserung

Ein Systemansatz ermöglicht den Übergang von isolierten digitalen Initiativen zu einem vollwertigen datengesteuerten Managementmodell, bei dem Entscheidungen nicht auf Intuition oder Annahmen beruhen, sondern auf Daten, Fakten und mathematisch berechneten Wahrscheinlichkeiten. Bei der digitalen Transformation der Baubranche geht es nicht nur um die Einführung von Technologien, sondern um die Gestaltung eines Geschäftsökosystems, in dem Projektinformationen nahtlos und iterativ zwischen verschiedenen Systemen übertragen werden. Dabei sorgen Algorithmen des maschinellen Lernens für automatisierte, kontinuierliche Analysen, Prognosen und Prozessoptimierung. In einer solchen Umgebung werden Spekulationen und versteckte Daten irrelevant - es bleiben nur bewährte Modelle, transparente Berechnungen und vorhersehbare Ergebnisse.

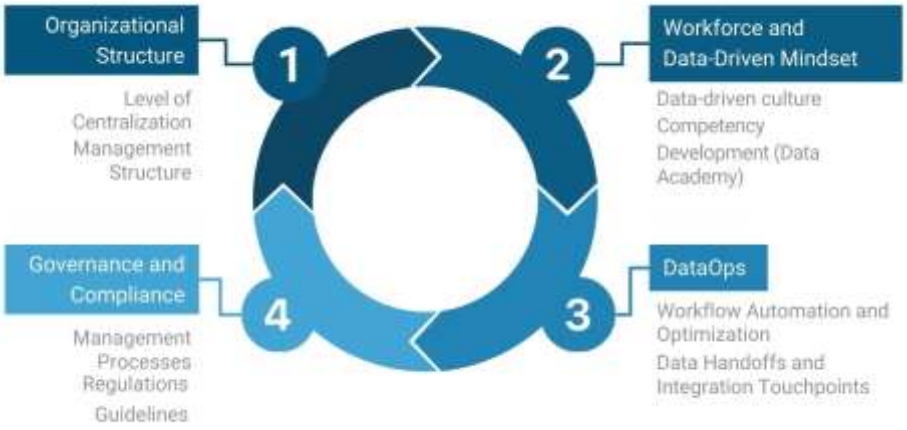


Abb. 10.2-3 Schlüsselelemente eines erfolgreichen Datenmanagements auf Unternehmensebene.

Jeder Teil des Buches entspricht einer bestimmten Phase der Datenverarbeitung und -analyse in Bauprojekten (Abb. 2.2-5). Wenn Sie zu einem der zuvor besprochenen Themen zurückkehren und es aus einem ganzheitlichen Verständnis des Datenverwendungsflusses heraus betrachten möchten, können Sie sich auf die Titel der Teile in Abb. 10.2-4 beziehen.

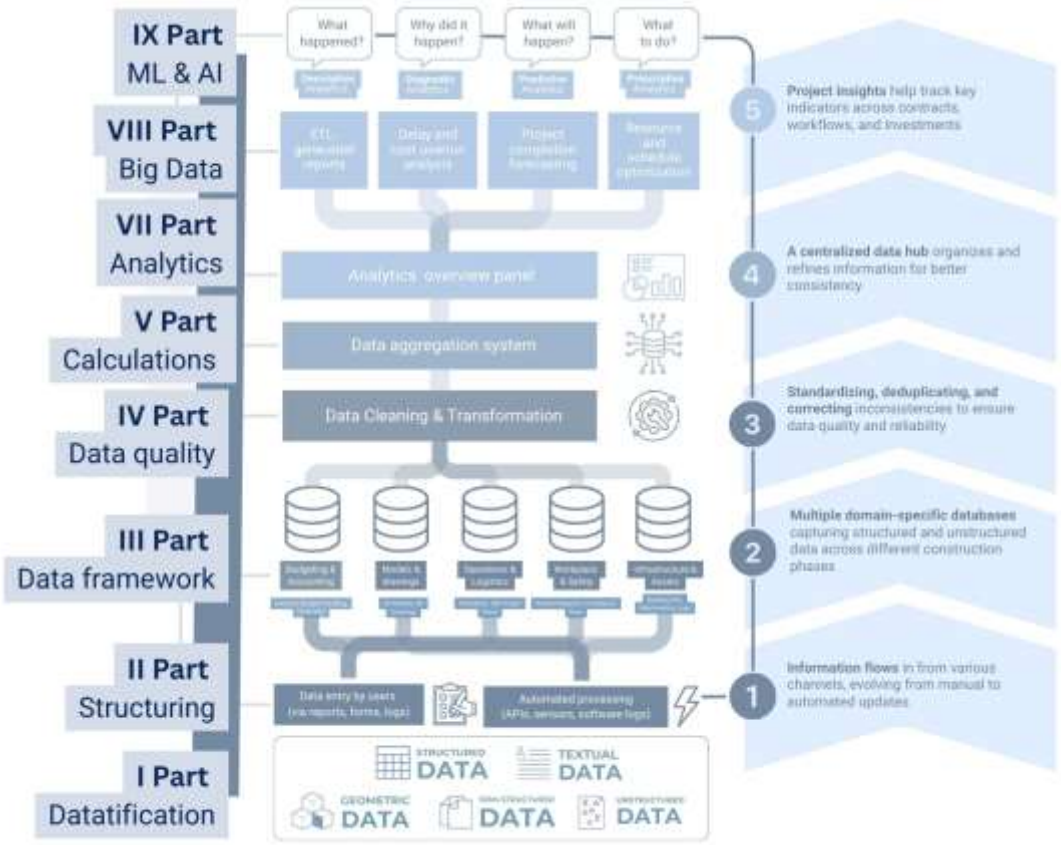


Abb. 10.2-4 Teile des Buches im Kontext der Datenverarbeitungspipeline (Abb. 2.2-5): von der Digitalisierung von Informationen bis zu Analytik und künstlicher Intelligenz.

Unabhängig von der Größe, dem technologischen Reifegrad oder dem Budget Ihres Unternehmens können Sie schon heute mit der Umstellung auf einen datengesteuerten Ansatz beginnen. Selbst kleine Schritte in die richtige Richtung werden mit der Zeit zu Ergebnissen führen.

> Die datengestützte Transformation ist kein einmaliges Projekt, sondern ein kontinuierlicher, iterativer Verbesserungsprozess, der die Einführung neuer Tools, die Umgestaltung von Prozessen und die Entwicklung einer Kultur der datengestützten Entscheidungsfindung umfasst.

Bauen in der Industrie 5.0: Wie man Geld verdient, wenn man sich nicht mehr verstecken kann

Lange Zeit haben die Bauunternehmen mit der Undurchsichtigkeit der Prozesse Geld verdient. Das Hauptgeschäftsmodell war die Spekulation - die Überschätzung der Materialkosten, des Arbeitsumfangs und der prozentualen Aufschläge in geschlossenen ERP - und PMIS - Systemen, die für externe Audits unzugänglich sind. Der eingeschränkte Zugang der Kunden und ihrer Vertrauenspersonen zu den ursprünglichen Projektdaten schuf den Boden für Machenschaften, bei denen es fast unmöglich wurde, die Zuverlässigkeit der Berechnungen zu überprüfen.

Dieses Modell verliert jedoch rasch an Bedeutung. Mit der Demokratisierung des Datenzugangs, dem Aufkommen von LLMs, dem Aufkommen offener Daten und ETL-Automatisierungstools bewegt sich die Branche auf einen neuen Arbeitsstandard zu.

> Infolgedessen ist Intransparenz nicht länger ein Wettbewerbsvorteil - bald wird sie zu einer Last, von der man sich nur schwer trennen kann. Transparenz wird von einer Option zu einer Voraussetzung, um auf dem Markt zu bestehen.

Mit wem werden die Kunden - Banken, Investoren, physische Kunden, Private Equity, staatliche Kunden - in der neuen digitalen Realität zusammenarbeiten? Die Antwort liegt auf der Hand: mit denjenigen, die in der Lage sind, nicht nur das Ergebnis zu liefern, sondern auch die Rechtfertigung für jeden Schritt auf dem Weg dorthin. In dem Maße, wie das Volumen offener Daten wächst, werden Partner und Kunden Unternehmen wählen, die Transparenz, Genauigkeit und Vorhersagbarkeit der Ergebnisse garantieren.

Vor diesem Hintergrund entstehen neue Geschäftsmodelle, die auf Datenmanagement und Vertrauen statt auf Spekulation beruhen:

- **Verkauf von Prozessen anstelle von Quadratmetern**: Der wichtigste Vermögenswert wird Vertrauen und Effizienz anstelle von diskontierten konkreten Vereinbarungen. Der wichtigste Wert wird die Vorhersehbarkeit des Ergebnisses auf der Grundlage zuverlässiger und überprüfter Daten sein. Moderne Unternehmen werden nicht das Bauobjekt als solches verkaufen, sondern:
  - genaue Fristen und transparente Arbeitspläne;
  - vernünftige Schätzungen, gestützt durch Berechnungen;
  - vollständige digitale Rückverfolgbarkeit und Kontrolle in allen Phasen des Projekts.
- **Engineering und Analytik als Dienstleistung**: das Modell "Data-as-a-Service " (eine Möglichkeit, den Nutzern fertige Daten über das Internet als Dienstleistung zu liefern), bei der jedes Projekt Teil einer digitalen Datenkette wird und der geschäftliche Wert in der Fähigkeit liegt, diese Kette zu verwalten. Die Unternehmen verwandeln sich in intelligente Plattformen, die Lösungen auf der Grundlage von Automatisierung und Analyse anbieten:
  - Automatisierte und transparente Erstellung von Kostenvoranschlägen und Plänen;
  - Risiko- und Zeiteinschätzung auf der Grundlage von Algorithmen des maschinellen Lernens;
  - Berechnung von Umweltindikatoren (ESG, CO_2, Energieeffizienz);
  - Erstellung von Berichten aus geprüften offenen Quellen.
- **Produktivierung des technischen Know-hows**: Die Entwicklungen des Unternehmens können innerhalb des Unternehmens mehrfach genutzt und als separates Produkt vertrieben werden - so entsteht eine zusätzliche Einnahmequelle durch digitale Dienstleistungen. In dem neuen Umfeld erstellen Unternehmen nicht nur Projekte, sondern auch digitale Assets:
  - Bibliotheken mit Komponenten und Kalkulationsvorlagen;
  - automatische Prüfmodule;
  - Open-Source-Plugins und -Skripte (Verkauf von Beratung) für die Arbeit mit Daten.

- **Ein neuer Unternehmenstyp: der Data-Driven Integrator**: ein Marktteilnehmer, der nicht von bestimmten Softwareanbietern oder modularen Systemen abhängt und nicht auf eine einzige Software-Schnittstelle "festgelegt" ist. Er arbeitet frei mit Daten - und baut seine Wettbewerbsfähigkeit darauf auf. Das Bauunternehmen der Zukunft ist nicht nur ein Auftragnehmer, sondern ein Informationsintegrator, der in der Lage ist, die folgenden Funktionen für den Kunden zu übernehmen:
    - Kombinieren Sie Daten aus unterschiedlichen Quellen und führen Sie Analysen durch;
    - Gewährleistung von Transparenz und Glaubwürdigkeit der Prozesse;
    - Beratung bei der Optimierung von Geschäftsprozessen;
    - Entwicklung von Tools, die im Ökosystem für offene Daten funktionieren, LLM, ETL und Pipelines.

Industrie 5.0 (Abb. 2.1-12) markiert das Ende der "Ära der manuellen Mittelwertbildung" und der abendlichen Treffen zwischen CEOs und der Kalkulations- und Buchhaltungsabteilung. Alles, was bisher im Verborgenen lag - Berechnungen, Schätzungen, Mengenangaben - wird offen, überprüfbar und auch für Nichtfachleute verständlich. Diejenigen, die sich als Erste neu orientieren, werden die Gewinner sein. Alle anderen werden in der neuen digitalen Wirtschaft des Baugewerbes außen vor bleiben.

SCHLUSSFOLGERUNG

Das Baugewerbe steht am Beginn einer Ära grundlegender Veränderungen. Von den ersten Aufzeichnungen auf Tontafeln bis hin zu den riesigen Mengen digitaler Daten, die von Projektservern und Baustellen fließen, spiegelt die Geschichte des Informationsmanagements im Bauwesen stets den Reifegrad der Technologie seiner Zeit wider. Heute, mit dem Aufkommen der Automatisierung, offener Formate und intelligenter Analysesysteme, steht die Branche nicht vor einer allmählichen Entwicklung, sondern vor einer raschen digitalen Transformation.

Wie in anderen Wirtschaftszweigen müssen auch im Bauwesen nicht nur die Instrumente, sondern auch die Arbeitsgrundsätze überdacht werden. Unternehmen, die früher den Markt diktierten und als Hauptvermittler zwischen dem Kunden und dem Projekt fungierten, verlieren ihre einzigartige Position. Vertrauen und die Fähigkeit, mit Daten zu arbeiten - vom Sammeln und Strukturieren bis hin zum Analysieren, Prognostizieren und Automatisieren von Entscheidungen - rücken in den Vordergrund.

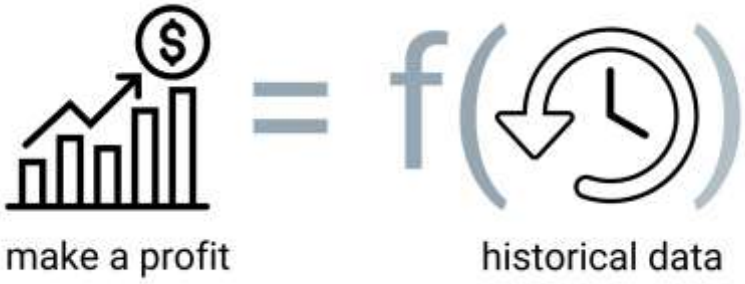


Abb. 10.2-1 Strukturierte historische Daten sind der Treibstoff für ein effizientes und kontrollierbares Unternehmen.

In diesem Buch werden die wichtigsten Grundsätze des Datenmanagements in der Baubranche erläutert - von der Prüfung und Standardisierung bis hin zur Prozessautomatisierung, dem Einsatz von Visualisierungstools und der Implementierung intelligenter Algorithmen. Wir haben uns angesehen, wie Sie auch mit begrenzten Ressourcen eine funktionierende Datenarchitektur aufbauen und Entscheidungen auf der Grundlage nachprüfbarer Fakten statt auf der Basis von Intuition treffen können. Die Arbeit mit Daten ist nicht mehr nur die Aufgabe der IT-Abteilung, sondern wird zur Grundlage der Managementkultur, die über die Flexibilität, Anpassungsfähigkeit und langfristige Nachhaltigkeit des Unternehmens entscheidet.

Die Anwendung von Technologien des maschinellen Lernens, von automatischen Verarbeitungssystemen, digitalen Zwillingen und offenen Formaten ermöglicht es bereits heute, den menschlichen Faktor dort zu eliminieren, wo er früher entscheidend war. Das Bauwesen bewegt sich in Richtung Autonomie und Kontrollierbarkeit, wobei der Weg von der Idee bis zur Projektrealisierung mit der Navigation im Autopilot-Modus verglichen werden kann: ohne Abhängigkeit von subjektiven Entscheidungen, ohne die Notwendigkeit manueller Eingriffe in jeder Phase, aber mit vollständiger digitaler Rückverfolgbarkeit und Kontrolle (Abb. 10.2-2).

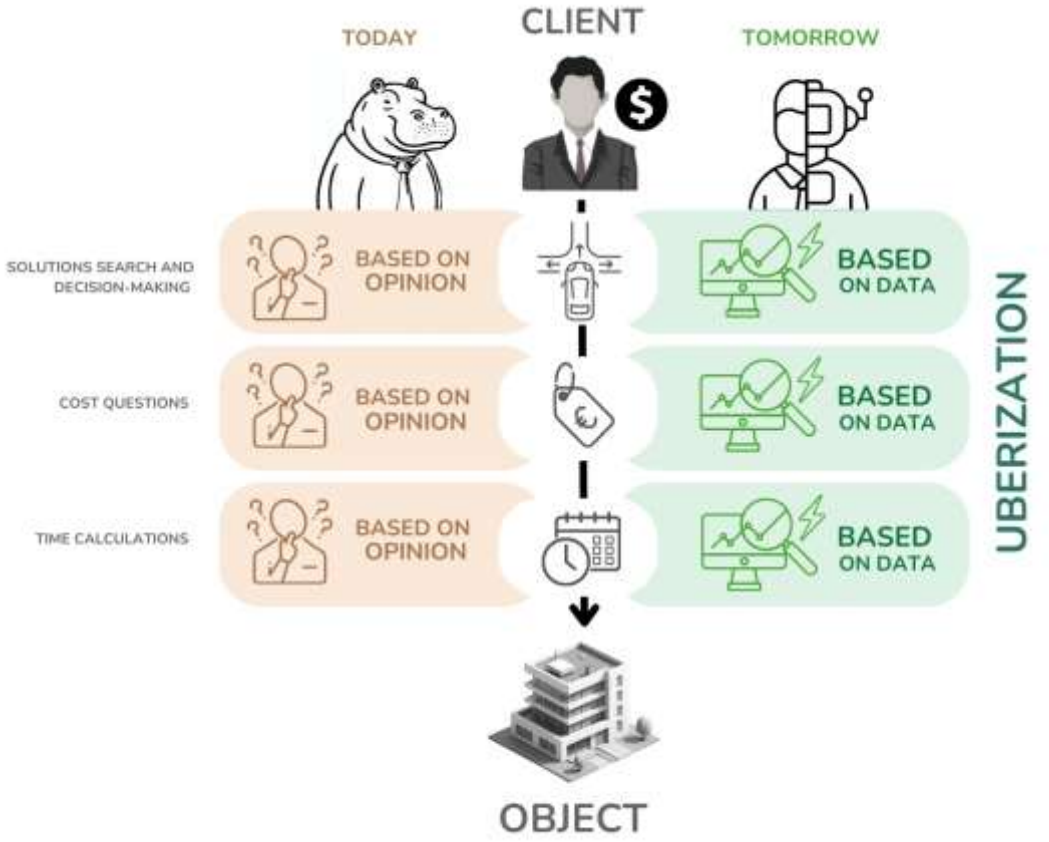


Abb. 10.2-2 Die Verlagerung der Entscheidungsfindung von der Meinung wichtiger Experten (HiPPO) zur Datenanalyse wird in erster Linie vom Kunden gefördert.

Durch das Erlernen der in diesem Buch vorgestellten Methoden, Prinzipien und Werkzeuge werden Sie in der Lage sein, in Ihrem Unternehmen datengesteuerte statt intuitive Entscheidungen zu treffen. Sie werden auch in der Lage sein, Modulketten in LLMs auszuführen, fertige ETL-Pipelines in Ihre Entwicklungsumgebung (IDE) zu kopieren und Daten automatisch zu verarbeiten, um die benötigten Informationen in der von Ihnen gewünschten Form zu erhalten. Aufbauend auf den Kapiteln zu Big Data und maschinellem Lernen werden Sie in der Lage sein, komplexere Szenarien zu implementieren, indem Sie neue Erkenntnisse aus historischen Daten gewinnen und Algorithmen des maschinellen Lernens anwenden, um Ihre Prozesse vorherzusagen und zu optimieren.

Offene Daten und Prozesse werden die Grundlage für genauere Schätzungen der Projektkosten und -fristen bilden und Bauunternehmen davon abhalten, auf undurchsichtigen Daten zu spekulieren. Dies ist sowohl eine Herausforderung als auch eine Chance für die Branche, ihre Rolle zu überdenken und sich an ein neues Umfeld anzupassen, in dem Transparenz und Effizienz zu zentralen Erfolgsfaktoren werden.

> Die Bereitschaft, Wissen aufzunehmen und in die Praxis umzusetzen, ist der Schlüssel zum Erfolg im Zeitalter der digitalen Transformation.

Unternehmen, die dies zuerst erkennen, werden im neuen digitalen Wettbewerb einen Vorteil haben. Aber es ist wichtig zu erkennen, dass Daten allein nichts ändern werden. Viele Menschen werden ihre Denkweise ändern müssen, und das erfordert einen Anreiz. Ihr Unternehmen muss die Art und Weise, wie es Daten

weitergibt, neu überdenken

Die Menschen, die das Unternehmen verändern, sind diejenigen, die wissen, wie man mit diesen Daten arbeitet, sie interpretiert, sie zur Optimierung einsetzt und auf ihrer Grundlage eine neue Prozessarchitektur schafft.

Wenn Sie diese Zeilen lesen, sind Sie bereit für Veränderungen und schon einen Schritt weiter. Danke, dass Sie diesen Weg gewählt haben. Willkommen in der Ära der digitalen Transformation!

ÜBER DEN AUTOR

Mein Name ist Artem Boiko. Meine Reise auf der Baustelle begann 2007 - mit einem Job als Bergmann in einem Ölschieferbergwerk in meiner Heimatstadt, während ich an der Bergbauuniversität St. Petersburg studierte und mich auf Bergbau und Untertagebau spezialisierte. Auf der Rückseite dieses Buches sehen Sie einen Sprengmeister in der Grube, in der wir Hunderte von Ölschieferwürfeln abgebaut und gesprengt haben. Mein beruflicher Werdegang hat mich in viele verschiedene Richtungen geführt, von der Arbeit im Berg- und Untertagebau bis hin zum Industriekletterer, Dachdecker und Installateur von Aufzugsanlagen. Ich hatte die Ehre, an Projekten unterschiedlicher Größenordnung beteiligt zu sein, vom Bau von Privathäusern bis hin zu großen Industrieanlagen in verschiedenen Regionen der Welt.

Im Laufe der Zeit hat sich meine Arbeit von der physischen Konstruktion zu Informationsmanagement und digitalen Prozessen verlagert. Seit 2013 habe ich in verschiedenen Positionen in kleinen, mittleren und großen Bauunternehmen in mehreren Regionen Deutschlands gearbeitet, vom Konstrukteur bis zum Datenmanagement-Manager. Was das Datenmanagement betrifft, so besteht meine Erfahrung in der Arbeit mit Daten in verschiedenen ERP-Systemen, CAD (BIM), MEP, FEM, CMS. Ich habe mich mit Optimierung, Prozessautomatisierung sowie Analyse, maschinellem Lernen und Datenverarbeitung in der Planungs-, Kalkulations- und Ausführungsphase von Bauarbeiten in Industrie-, Wohnungs-, Infrastruktur- und Versorgungsbauunternehmen befasst.

Ich arbeite seit 2003 mit Open-Source-Software und offenen Daten. In dieser Zeit habe ich viele Webprojekte umgesetzt - von Websites und Online-Shops bis hin zu vollwertigen Webanwendungen mit Open-Source-Lösungen und offenen CMS. Diese Plattformen, die in vielerlei Hinsicht modernen Gebäude-ERPs ähneln, zeichnen sich durch modulare Architektur, hohe Anpassungsfähigkeit und Zugänglichkeit aus. Diese Erfahrung legte den Grundstein für meinen professionellen Ansatz - ein Fokus auf Open-Source-Technologien und eine Kultur der kollaborativen Entwicklung. Der Respekt vor Open Source und dem freien Austausch von Wissen ist etwas, das ich in der Baubranche zu fördern versuche. Meine Arbeit zur Verbesserung der Datenzugänglichkeit in der Baubranche hat sich in der Gründung mehrerer Social-Media-Communities niedergeschlagen, in denen die Offenheit von Daten und die Verwendung von Open Source im Bauwesen diskutiert werden, sowie in der Gründung mehrerer Start-ups, die Lösungen für den Zugang zu Daten aus verschiedenen geschlossenen Systemen und Plattformen entwickeln.

Meinen Beitrag zur Fachwelt leiste ich durch die Teilnahme als Referent an Konferenzen zu den Themen CAD-Interoperabilität (BIM), ERP, 4D-5D, LLM Machine Learning und Künstliche Intelligenz sowie durch die Veröffentlichung von Artikeln in europäischen Publikationen der Baubranche. Eine meiner bemerkenswerten Leistungen ist die Erstellung der "BIM History" [111], eine umfassende Übersicht über wichtige Softwarelösungen für das Datenmanagement in der Bauindustrie. Meine siebenteilige Artikelserie "BIM Development and Lobbying Games" (BIM-Entwicklung und Lobbying-Spiele), die in mehrere Sprachen übersetzt wurde, ist weithin als Versuch anerkannt worden, die verborgene Dynamik digitaler Standards zu beleuchten.

So bin ich vom Gesteinsabbau zum Abbau und zur Systematisierung von Baudaten gekommen. Ich bin immer offen für einen fachlichen Dialog, neue Ideen und gemeinsame Projekte. Ich bin dankbar für jedes Feedback und freue mich auf Ihre Nachrichten oder darauf, Sie unter meinen Social-Media-Followern zu sehen. Vielen Dank, dass Sie dieses Buch bis zum Ende gelesen haben! Ich würde mich freuen, wenn dieses Buch Ihnen hilft, das Thema Daten in der Bauwirtschaft besser zu verstehen.

WEITERE BEZIEHUNG

Die Meinungen der Leser spielen eine wichtige Rolle bei der Weiterentwicklung der Veröffentlichungen und bei der Auswahl der Schwerpunktthemen. Besonders wertvoll sind Kommentare darüber, welche Ideen sich als nützlich erwiesen haben und welche Zweifel aufgeworfen haben und weiterer Klärung oder Quellenangaben bedürfen. Das Buch enthält ein breites Spektrum an Material und Analysen, von denen einige kontrovers oder subjektiv erscheinen mögen. Sollten Sie bei der Lektüre Ungenauigkeiten, falsch zitierte Quellen, logische Ungereimtheiten oder Tippfehler entdecken, würde ich mich über Ihre Kommentare, Gedanken oder Kritik freuen, die Sie an boikoartem@gmail.com senden können. Oder über Nachrichten auf LinkedIn: linkedin.com/in/boikoartem/.

Ich wäre sehr dankbar für jede Erwähnung des Buches "Data-Driven Construction" in den sozialen Medien - das Teilen des Leseerlebnisses hilft, offene Daten und Tools zu verbreiten und unterstützt meine Arbeit.

ÜBERSETZUNGSKOMMENTAR

Dieses Buch wurde mit der Technologie der künstlichen Intelligenz übersetzt. Dadurch konnte der Übersetzungsprozess wesentlich beschleunigt werden. Wie bei jedem technologischen Vorgang können jedoch Fehler oder Ungenauigkeiten auftreten. Wenn Ihnen etwas auffällt, das falsch oder fehlerhaft übersetzt zu sein scheint, schicken Sie mir bitte eine E-Mail. Ihre Kommentare werden dazu beitragen, die Qualität der Übersetzung zu verbessern.

COMMUNITY

Es ist ein Ort, an dem Sie frei Fragen stellen und Ihre Probleme und Lösungen mitteilen können:

DataDrivenConstruction.io: https://datadrivenconstruction.io

LinkedIn: https://www.linkedin.com/company/78381569

Twitter: https://twitter.com/datadrivenconst

Telegram: https://t.me/datadrivenconstruction

YouTube: https://www.youtube.com/@datadrivenconstruction

ANDERE FÄHIGKEITEN UND KONZEPTE

Zusätzlich zu den Schlüsselprinzipien der Arbeit mit Daten in der Bauindustrie behandelt DataDriven-Construction eine breite Palette von zusätzlichen Konzepten, Programmen und Fähigkeiten, die für den datengesteuerten Fachmann unerlässlich sind. Einige von ihnen werden nur in einem Überblick vorgestellt, sind aber für die Praxis entscheidend.

Der interessierte Leser kann auf der Website DataDrivenConstruction.io Links zu weiteren Materialien zu Schlüsselkompetenzen finden. Diese Materialien umfassen die Arbeit mit Python und Pandas, den Aufbau von ETL -Prozessen, Beispiele für die Datenverarbeitung in CAD-Projekten im Bauwesen, Big-Data-Systeme und moderne Ansätze zur Visualisierung und Analyse von Baudaten.

Bei der Erstellung des Buches "DataDrivenConstruction" und aller Fallstudien wurden viele Open-Source-Tools und Software. verwendet. Der Autor bedankt sich bei den Entwicklern und Co-Autoren der folgenden Lösungen:

- Python und Pandas - das Rückgrat der Datenmanipulation und Automatisierung
- Scipy, NumPy, Matplotlib und Scikit-Learn - Bibliotheken für Datenanalyse und maschinelles Lernen
- SQL und Apache Parquet - Werkzeuge für die Speicherung und Verarbeitung großer Mengen von Baudaten
- Open Source CAD (BIM) offene Datenwerkzeuge in offenen Formaten
- N8n, Apache Airflow, Apache NiFi - Systeme zur Orchestrierung und Automatisierung von Arbeitsabläufen
- DeepSeek, LlaMa, Mistral - Open Source LLM

Besonderer Dank gilt allen Diskussionsteilnehmern zum Thema offene Daten und Tools in Fachgemeinschaften und sozialen Netzwerken, deren Kritik, Kommentare und Ideen dazu beigetragen haben, den Inhalt und die Struktur dieses Buches zu verbessern.

Verfolgen Sie die Entwicklung des Projekts auf der Website DataDrivenConstruction.io, wo nicht nur Buchaktualisierungen und Korrekturen veröffentlicht werden, sondern auch neue Kapitel, Tutorials und praktische Beispiele für die Anwendung der beschriebenen Techniken.

MAXIMALER KOMFORT MIT DER GEDRUCKTEN VERSION

Sie sind im Besitz einer kostenlosen digitalen Version von **Data-Driven Construction**. Für bequemeres Arbeiten und schnellen Zugriff auf die Materialien empfehlen wir Ihnen, die **gedruckte Ausgabe zu verwenden**:

- **Immer zur Hand**: Das Buch im gedruckten Format wird zu einem zuverlässigen Arbeitsinstrument, das es Ihnen ermöglicht, in jeder Arbeitssituation schnell die richtigen Visualisierungen und Diagramme zu finden und zu verwenden

- **Hochwertige Illustrationen**: Alle Bilder und Grafiken in der gedruckten Ausgabe werden in höchster Qualität präsentiert

- **Schneller Zugriff auf Informationen**: einfache Navigation, die Möglichkeit, Notizen zu machen, Lesezeichen zu setzen und überall mit dem Buch zu arbeiten.

Mit dem Kauf der gedruckten Vollversion des Buches erhalten Sie ein praktisches Hilfsmittel für die bequeme und effiziente Arbeit mit Informationen: die Möglichkeit, bei alltäglichen Aufgaben schnell auf Anschauungsmaterial zurückzugreifen, die notwendigen Schemata schnell zu finden und Notizen zu machen. Darüber hinaus unterstützen Sie mit Ihrem Kauf die Verbreitung von offenem Wissen.

Bestellen Sie eine gedruckte Version des Buches unter: datadrivenconstruction.io/books

EINZIGARTIGE GELEGENHEIT FÜR STRATEGISCHE POSITIONIERUNG

Wir bieten Ihnen an, Werbematerialien in der kostenlosen Version von DataDriven-Construction zu platzieren. Die kostenpflichtige Version der Publikation hat bereits im ersten Jahr nach ihrer Veröffentlichung die Aufmerksamkeit von Fachleuten aus mehr als 50 Ländern der Welt - von Lateinamerika bis zum asiatisch-pazifischen Raum - auf sich gezogen. Um individuelle Bedingungen der Zusammenarbeit zu besprechen und detaillierte Informationen über Platzierungsmöglichkeiten zu erhalten, füllen Sie bitte das Feedback-Formular auf dem offiziellen Portal datadriven-construction.io aus oder schreiben Sie an die am Ende des Buches aufgeführten Kontakte.

DIE BUCHKAPITEL SIND UNTER DATADRIVENCONSTRUCTION.IO VERFÜGBAR

Sie können die Kapitel von Data-Driven Construction auf der Data-Driven Construction-Website lesen, wo Abschnitte des Buches nach und nach veröffentlicht werden, damit Sie die benötigten Informationen schnell finden und bei Ihrer Arbeit nutzen können. Sie finden dort auch viele andere Veröffentlichungen zu ähnlichen Themen sowie Anwendungsbeispiele und Lösungen, die Ihnen helfen, Ihre Fähigkeiten zu entwickeln und Daten im Bauwesen anzuwenden.

AKTUELLE VERSIONEN DES BUCHES DOWNLOAD VON DER OFFIZIELLEN WEBSITE

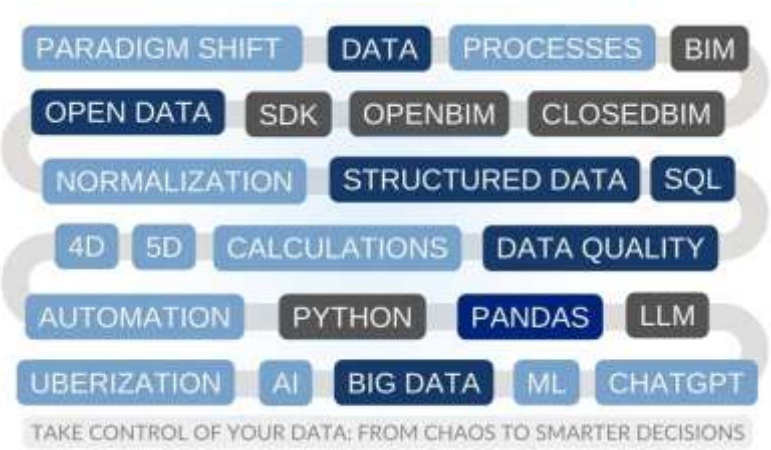
Aktuelle und neueste Versionen des DataDrivenConstruction-Buches stehen unter datadrivenconstruction.io zum Download bereit. Wenn Sie Updates mit neuen Kapiteln des Buches, praktischen Tipps oder Rezensionen neuer Anwendungen erhalten möchten, abonnieren Sie den Newsletter:

- Sie werden die ersten sein, die die neuen Abschnitte des Buches kennenlernen
- Erhalten Sie praktische Fallstudien und Ratschläge zu Analytik und Automatisierung in der Bauindustrie
- Halten Sie sich über Trends, Veröffentlichungen und Anwendungsbeispiele auf dem Laufenden

Gehen Sie zu datadrivenconstruction.io um zu abonnieren!

DATENGESTÜTZTES BAUEN:
BERATUNG, WORKSHOPS UND SCHULUNGEN

Die Schulungs- und Beratungsprogramme von DataDrivenConstruction haben Dutzenden von führenden Bauunternehmen auf der ganzen Welt geholfen, ihre Effizienz zu steigern, Kosten zu senken und die Qualität ihrer Lösungen zu verbessern. Zu den Kunden von DataDrivenConstruction gehören einige der größten Akteure auf dem Milliardenmarkt, darunter Bau-, Beratungs- und IT-Unternehmen.



Warum sollten Sie uns wählen?

- **Relevanz:** Gespräche über die wichtigsten Trends und Erkenntnisse in der Branche
- **Praktisch:** Unterstützung von Fachleuten bei der effizienten Lösung alltäglicher Aufgaben durch PoC.
- **Persönlicher Ansatz:** auf Ihr Unternehmen zugeschnitten, um den Nutzen von Schulung und Beratung zu maximieren

Die Hauptschwerpunkte des DataDrivenConstruction-Teams:

- **Datenqualitätsmanagement:** Unterstützung bei der Parametrisierung von Aufgaben, Erfassung von Anforderungen, Validierung und Vorbereitung von Daten für die automatische Verarbeitung.
- **Data Mining - Extraktion und Strukturierung von Daten:** Einrichtung von ETL-Prozessen und Extraktion von Daten aus E-Mails, PDF, Excel, Bildern und anderen Quellen.
- **BIM und CAD analytics**: Sammeln, Strukturieren und Analysieren von Informationen aus RVT-Dateien, IFC, DWG und anderen CAD-Formaten (BIM).
- **Analytik und Datenumwandlung:** Umwandlung disparater Informationen in strukturierte Daten, Analytik, Erkenntnisse und Entscheidungen.
- **Datenintegration und Prozessautomatisierung:** von der automatischen Dokumentenerstellung bis zur Integration mit internen Systemen und externen Datenbanken.

Wenden Sie sich an DataDrivenConstruction.io, um zu erfahren, wie Ihr Unternehmen durch Automatisierung greifbare Geschäftsergebnisse erzielen kann.

GLOSSAR

AI (Artificial Intelligence) - Künstliche Intelligenz; die Fähigkeit von Computersystemen, Aufgaben auszuführen, die normalerweise menschliche Intelligenz erfordern, wie Mustererkennung, Lernen und Entscheidungsfindung.

Apache Airflow ist eine Open-Source-Plattform für die Workflow-Orchestrierung, mit der Sie Workflows und ETLs mithilfe von DAGs (gerichteten azyklischen Graphen) programmatisch erstellen, planen und verfolgen können.

Apache NiFi ist ein Tool zur Automatisierung von Datenflüssen zwischen Systemen, das sich auf die Weiterleitung und Umwandlung von Daten spezialisiert.

Apache Parquet ist ein effizientes Dateiformat für die Speicherung spaltenförmiger Daten, das für die Verwendung in Big-Data-Analysesystemen optimiert wurde. Es bietet erhebliche Komprimierung und schnelle Verarbeitung.

API (Application Programming Interface) - eine formalisierte Schnittstelle, die es einem Programm ermöglicht, mit einem anderen zu interagieren, ohne auf den Quellcode zugreifen zu müssen, wobei Daten und Funktionen über standardisierte Anfragen und Antworten ausgetauscht werden.

Attribut - ein Merkmal oder eine Eigenschaft eines Objekts, die seine Merkmale beschreibt (z. B. Fläche, Volumen, Kosten, Material).

Datenbanken sind organisierte Strukturen zur Speicherung, Verwaltung und zum Zugriff auf Informationen, die für eine effiziente Datenabfrage und -verarbeitung genutzt werden.

BEP (BIM Execution Plan) - Ein Implementierungsplan für Building Information Modelling, der die Ziele, Methoden und Prozesse für die Implementierung von BIM in einem Projekt definiert.

Big Data - Datenmengen von erheblichem Umfang, Vielfalt und Aktualisierungsrate, die spezielle Technologien für die Verarbeitung und Analyse erfordern.

BI (Business Intelligence) - Business **Intelligence**; die Prozesse, Technologien und Werkzeuge zur Umwandlung von Daten in aussagekräftige Informationen für die Entscheidungsfindung.

BIM (Building Information Modeling) - Building Information Modelling; der Prozess der Erstellung und Verwaltung digitaler Darstellungen der physischen und funktionalen Merkmale von Bauprojekten, die nicht nur 3D-Modelle, sondern auch Informationen über Leistung, Materialien, Zeitplan und Kosten enthalten.

BlackBox/WhiteBox - Ansätze zum Verständnis des Systems: im ersten Fall ist die interne Logik verborgen, nur Eingänge und Ausgänge sind sichtbar; im zweiten Fall ist die Verarbeitung transparent und für die Analyse verfügbar.

Bounding Box ist ein geometrisches Konstrukt, das die Grenzen eines Objekts im dreidimensionalen Raum durch die minimalen und maximalen X-, Y- und Z-Koordinaten beschreibt und einen "Kasten" um das Objekt herum bildet.

BREP (Boundary Representation) ist eine geometrische Darstellung von Objekten, die sie durch die Grenzen von Oberflächen definiert.

CAD (Computer-Aided Design) ist ein computergestütztes Konstruktionssystem, das zur Erstellung, Bearbeitung und Analyse von genauen Zeichnungen und 3D-Modellen in der Architektur, im Bauwesen, im Ingenieurwesen und in anderen Branchen verwendet wird.

CAFM (Computer-Aided Facility Management) ist eine Software für das Immobilien- und Infrastrukturmanagement, die Raumplanung, Anlagenverwaltung, Instandhaltung und Kostenüberwachung umfasst.

CDE (Common Data Environment) - ein zentraler digitaler Raum für die Verwaltung, Speicherung, gemeinsame Nutzung und Zusammenarbeit mit Projektinformationen in allen Phasen des Lebenszyklus einer Anlage.

Ein **Exzellenzzentrum (Centre of Excellence, CoE**) ist eine spezialisierte Struktur innerhalb einer Organisation, die für die Entwicklung eines bestimmten Wissensgebiets, die Entwicklung von Standards und bewährten Verfahren, die Schulung des Personals und die Unterstützung bei der Einführung von Innovationen zuständig ist.

CoClass ist ein modernes Klassifizierungssystem für Bauelemente der dritten Generation.

Ein konzeptionelles Datenmodell ist eine High-Level-Darstellung grundlegender Entitäten und ihrer Beziehungen ohne detaillierte Attribute, die in den ersten Phasen des Datenbankdesigns verwendet wird.

CRM (Customer Relationship Management) ist ein System zur Verwaltung von Kundenbeziehungen, das zur Automatisierung von Vertriebs- und Serviceprozessen eingesetzt wird.

DAG (Directed Acyclic Graph) ist ein gerichteter azyklischer Graph, der in Datenorchestrierungssystemen (Airflow, NiFi) verwendet wird, um Aufgabensequenzen und Abhängigkeiten zu bestimmen.

Dash ist ein Python-Framework zur Erstellung interaktiver webbasierter Datenvisualisierungen.

Dashboard - Ein **Dashboard**, das wichtige Leistungsindikatoren und Metriken in Echtzeit visuell darstellt.

Der datenzentrierte Ansatz ist eine Methode, die Daten den Vorrang vor Anwendungen oder Softwarecode gibt und sie zum zentralen Gut des Unternehmens macht.

Data Governance - eine Reihe von Praktiken, Prozessen und Richtlinien, die die angemessene und effektive Nutzung von Daten innerhalb einer Organisation gewährleisten, einschließlich Zugangs-, Qualitäts- und Sicherheitskontrollen.

Data Lake ist ein Speicher, in dem große Mengen von Rohdaten in ihrem ursprünglichen Format gespeichert werden, bis sie verwendet werden.

Data Lakehouse ist ein architektonischer Ansatz, der die Flexibilität und Skalierbarkeit von Data Lakes (Data **Lake**) mit der Verwaltbarkeit und Leistung von Data Warehouses (DWH) kombiniert.

Data-Driven Construction ist ein strategischer Ansatz, bei dem jede Phase des Lebenszyklus einer Anlage - von der Planung bis zum Betrieb - durch automatisierte, vernetzte Systeme unterstützt wird. Dieser Ansatz ermöglicht kontinuierliches, faktenbasiertes Lernen, verringert die Unsicherheit und ermöglicht es Unternehmen, eine nachhaltige Branchenführerschaft zu erreichen.

Data-Driven Integrator - ein Unternehmen, das darauf spezialisiert ist, Daten aus unterschiedlichen Quellen zu kombinieren und zu analysieren, um Managemententscheidungen zu treffen.

Datengesteuerter Ansatz - eine Methodik, bei der Daten als strategisches Gut betrachtet werden und Entscheidungen auf der Grundlage einer objektiven Analyse von Informationen und nicht auf der Grundlage subjektiver Meinungen getroffen werden.

Datenminimalismus - ein Ansatz zur Reduzierung der Daten auf das Wertvollste und Sinnvollste, der eine vereinfachte Verarbeitung und Analyse der Informationen ermöglicht.

Datensumpf - Eine verstreute Masse von unstrukturierten Daten, die entsteht, wenn Informationen unkontrolliert und ohne angemessene Organisation gesammelt und gespeichert werden.

DataOps ist eine Methodik, die DevOps-Prinzipien, Daten und Analysen kombiniert und sich auf die Verbesserung der Zusammenarbeit, Integration und Automatisierung von Datenflüssen konzentriert.

Die Digitalisierung von **Informationen** ist der Prozess der Umwandlung aller Aspekte der Bautätigkeit in eine digitale Form, die für die Analyse, Interpretation und Automatisierung geeignet ist.

DataFrame - Eine zweidimensionale tabellarische Datenstruktur in der Pandas-Bibliothek, bei der Zeilen einzelne Datensätze oder Objekte und Spalten deren Merkmale oder Attribute darstellen.

Deskriptive Analyse - Analyse historischer Daten, um zu verstehen, was in der Vergangenheit passiert ist.

Diagnostische Analyse - Analyse von Daten, um festzustellen, warum etwas passiert ist.

Ein Gantt-Diagramm ist ein Projektplanungsinstrument, das Aufgaben als horizontale Balken auf einer Zeitachse darstellt und es Ihnen ermöglicht, die Reihenfolge und Dauer der Arbeiten zu visualisieren.

DWH (Data Warehouse) ist ein zentralisiertes Data-Warehouse-System, das Informationen aus verschiedenen Quellen zusammenfasst, strukturiert und für Analysen und Berichte zur Verfügung stellt.

ESG (Environmental, Social, Governance) - eine Reihe von Kriterien zur Bewertung der ökologischen, sozialen und Governance-Auswirkungen eines Unternehmens oder Projekts.

ELT (Extrahieren, Laden, Transformieren) ist ein Prozess, bei dem Daten zunächst aus Quellen extrahiert und in ein Repository geladen und dann für Analysezwecke transformiert werden.

ETL (Extrahieren, Transformieren, Laden) ist der Prozess, bei dem Daten aus verschiedenen Quellen extrahiert, in das gewünschte Format umgewandelt und zur Analyse in den Zielspeicher geladen werden.

ER-Diagramm (Entity-Relationship) - ein visuelles Diagramm, das Entitäten, ihre Attribute und die Beziehungen zwischen ihnen darstellt und bei der Datenmodellierung verwendet wird.

ERP (Enterprise Resource Planning) ist ein umfassendes modulares Warenwirtschaftssystem, das zur Verwaltung und Optimierung verschiedener Aspekte des Bauprozesses eingesetzt wird.

Merkmale - Beim maschinellen Lernen unabhängige Variablen oder Attribute, die als Eingaben für ein Modell verwendet werden.

Physisches Datenmodell - eine detaillierte Darstellung der Datenbankstruktur, einschließlich Tabellen, Spalten, Datentypen, Schlüssel und Indizes, optimiert für ein bestimmtes DBMS.

FPDF ist eine Python-Bibliothek zur Erstellung von PDF-Dokumenten.

Geometric Core ist eine Softwarekomponente, die grundlegende Algorithmen für die Erstellung, Bearbeitung und Analyse von geometrischen Objekten in CAD-, BIM- und anderen technischen Anwendungen bietet.

HiPPO (Highest Paid Person's Opinion) - ein Ansatz zur Entscheidungsfindung, der auf der Meinung der höchstbezahlten Person im Unternehmen und nicht auf objektiven Daten beruht.

IDE (Integrated Development Environment) - integrierte Entwicklungsumgebung, ein umfassendes Werkzeug zum Schreiben, Testen und Debuggen von Code (z. B. PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) ist eine Spezifikation für die Bereitstellung von Informationen, die die Datenanforderungen in verschiedenen Phasen eines Projekts definiert.

IFC (Industry Foundation Classes) ist ein BIM-Datenaustauschformat, das die Interoperabilität zwischen verschiedenen Softwarelösungen gewährleistet.

Industrie 5.0 ist ein industrielles Entwicklungskonzept, das die Möglichkeiten der Digitalisierung, Automatisierung und künstlichen Intelligenz mit dem menschlichen Potenzial und der ökologischen Nachhaltigkeit verbindet.

Bei der Datenintegration werden Daten aus verschiedenen Quellen in einem einzigen, kohärenten System kombiniert, um eine einheitliche Sicht auf die Informationen zu erhalten.

Informationssilos sind isolierte Datenspeichersysteme, die keine Informationen mit anderen Systemen austauschen und so die effiziente Nutzung von Daten behindern.

IoT (Internet der Dinge) ist das Konzept, physische Objekte mit dem Internet zu verbinden, um Daten zu sammeln, zu verarbeiten und zu übertragen.

k-NN (k-Nearest Neighbors) ist ein Algorithmus für maschinelles Lernen, der Objekte auf der Grundlage der Ähnlichkeit mit den nächsten Nachbarn in der Trainingsstichprobe klassifiziert.

Kaggle ist eine Plattform für Datenanalyse und Wettbewerbe zum maschinellen Lernen.

Kalkulation - Berechnung der Kosten von Bauarbeiten oder Prozessen für eine bestimmte Maßeinheit (z. B. 1 m² Gipskartonwand, 1 m³ Beton).

KPIs (Key Performance Indicators) sind wichtige Leistungsindikatoren, quantifizierbare Messgrößen, die zur Bewertung des Erfolgs eines Unternehmens oder eines bestimmten Projekts verwendet werden.

Labels - Beim maschinellen Lernen die Zielvariablen oder Attribute, die das Modell vorhersagen soll.

Lernalgorithmus - Der Prozess der Suche nach der besten Hypothese in einem Modell, die einer Zielfunktion entspricht, unter Verwendung eines Satzes von Trainingsdaten.

Lineare Regression - Eine statistische Methode zur Modellierung der Beziehung zwischen einer abhängigen Variablen und einer oder mehreren unabhängigen Variablen.

LLM (Large Language Model) - Large Language Model, eine künstliche Intelligenz, die darauf trainiert ist, Text aus großen Datenmengen zu verstehen und zu generieren, und die in der Lage ist, den Kontext zu analysieren und Programmiercode zu schreiben.

LOD (Level of Detail/Development) - der Detailgrad des Modells, der den Grad der geometrischen Genauigkeit und den Informationsgehalt bestimmt.

Ein logisches Datenmodell ist eine detaillierte Beschreibung von Entitäten, Attributen, Schlüsseln und Beziehungen, die Geschäftsinformationen und -regeln widerspiegelt und eine Zwischenstufe zwischen dem konzeptionellen und dem physischen Modell darstellt.

Maschinelles Lernen - Eine Klasse von Techniken der künstlichen Intelligenz, die es Computersystemen ermöglichen, ohne ausdrückliche Programmierung aus Daten zu lernen und Vorhersagen zu treffen.

Masterformat ist ein Klassifizierungssystem der ersten Generation, das zur Strukturierung von Baubeschreibungen in Abschnitte und Disziplinen verwendet wird.

MEP (Mechanical, Electrical, Plumbing) - Gebäudetechnische Systeme, die mechanische, elektrische und sanitärtechnische Komponenten umfassen.

Mesh ist eine Netzdarstellung von 3D-Objekten, die aus Scheitelpunkten, Kanten und Flächen besteht.

Modell - Beim maschinellen Lernen ein Satz verschiedener Hypothesen, von denen eine die zu prognostizierende oder zu approximierende Zielfunktion annähert.

Bei der Datenmodellierung handelt es sich um den Prozess der strukturierten Darstellung von Daten und ihren Beziehungen zur Implementierung in Informationssysteme, einschließlich der konzeptionellen, logischen und physischen Ebene.

n8n ist ein Open-Source-Tool für die Automatisierung von Arbeitsabläufen und die Integration von Anwendungen durch einen Low-Code-Ansatz.

Normalisierung - beim maschinellen Lernen der Prozess, bei dem unterschiedliche numerische Daten auf eine gemeinsame Skala gebracht werden, um die Verarbeitung und Analyse zu erleichtern.

Reverse-Engineering - der Prozess der Untersuchung des Geräts, der Funktionsweise und der Herstellungstechnologie eines Objekts durch Analyse seiner Struktur, seiner Funktionen und seines Betriebs. Im Zusammenhang mit Daten - Extraktion von Informationen aus proprietären Formaten zur Verwendung in offenen Systemen.

OCR (Optical Character Recognition) ist eine Technologie zur **optischen** Zeichenerkennung, die Textbilder (gescannte Dokumente, Fotos) in ein maschinenlesbares Textformat umwandelt.

OmniClass ist ein internationaler Klassifizierungsstandard der zweiten Generation für das Informationsmanagement im Bauwesen.

Ontologie - Ein System von Begriffsverknüpfungen, das einen bestimmten Wissensbereich formalisiert.

Open Source - ein Modell für die Entwicklung und Verbreitung von quelloffener Software, die zur freien Nutzung, Untersuchung und Veränderung zur Verfügung steht.

Open BIM ist das Konzept des offenen BIM, das die Verwendung offener Standards und Formate für den Datenaustausch zwischen verschiedenen Softwarelösungen beinhaltet.

Offene Standards - öffentlich verfügbare Spezifikationen für eine bestimmte Aufgabe, die es verschiedenen Systemen ermöglichen, miteinander zu interagieren und Daten auszutauschen.

Pandas ist eine Open-Source-Python-Bibliothek für die Datenverarbeitung und -analyse, die DataFrame- und Series-Datenstrukturen für die effiziente Verarbeitung von tabellarischen Informationen bereitstellt.

Das Paradigma der offenen Daten ist ein Ansatz für die Datenverarbeitung, bei dem Informationen zur Nutzung, Wiederverwendung und Verbreitung durch jedermann frei zugänglich gemacht werden.

Die parametrische Methode ist eine Methode zur Schätzung von Bauprojekten, bei der statistische Modelle verwendet werden, um den Wert auf der Grundlage von Projektparametern zu schätzen.

PIMS (Project Information Model) ist ein digitales System, mit dem alle Projektinformationen organisiert, gespeichert und gemeinsam genutzt werden können.

Pipeline - Eine Abfolge von Datenverarbeitungsprozessen, von der Extraktion und Transformation bis zur Analyse und Visualisierung.

PMIS (Project Information Management System) ist ein Projektmanagementsystem, das für eine detaillierte Aufgabensteuerung auf der Ebene eines einzelnen Bauprojekts konzipiert ist.

Predictive Analytics ist ein Teilbereich der Analytik, der statistische Methoden und maschinelles Lernen einsetzt, um zukünftige Ergebnisse auf der Grundlage historischer Daten vorherzusagen.

Prescriptive Analytics - Ein Bereich der Analytik, der nicht nur zukünftige Ergebnisse vorhersagt, sondern auch optimale Maßnahmen vorschlägt, um die gewünschten Ergebnisse zu erzielen.

Proprietäre Formate sind geschlossene, von einem bestimmten Unternehmen kontrollierte Datenformate, die die Möglichkeit des Informationsaustauschs einschränken und die Abhängigkeit von einer bestimmten Software erhöhen.

QTO (Quantity Take-Off) ist der Prozess, bei dem die Mengen der Elemente aus den Planungsunterlagen extrahiert werden, um die für die Realisierung des Projekts erforderlichen Materialmengen zu berechnen.

Qualitätsmanagementsystem - ein Qualitätsmanagementsystem, das sicherstellt, dass Prozesse und Ergebnisse den festgelegten Anforderungen entsprechen.

RAG (Retrieval-Augmented Generation) ist eine Methode, die die generativen Fähigkeiten von Sprachmodellen mit der Extraktion relevanter Informationen aus Unternehmensdatenbanken kombiniert und so die Genauigkeit und Relevanz der Antworten verbessert.

RDBMS (Relational Database Management System) ist ein relationales Datenbankmanagementsystem, das Informationen in Form von miteinander verknüpften Tabellen verwaltet.

RegEx (Regular Expressions) ist eine formalisierte Sprache für die Suche und Verarbeitung von Zeichenketten, mit der Sie Vorlagen für die Überprüfung von Textdaten auf die Einhaltung bestimmter Kriterien festlegen können.

Die Regression ist eine statistische Methode zur Analyse der Beziehung zwischen Variablen.

CO_2-Berechnungen sind eine Methode zur Schätzung der Kohlendioxidemissionen, die mit der Herstellung und Verwendung von Baumaterialien und -verfahren verbunden sind.

Die Ressourcenmethode ist eine Methode zur Erstellung von Kostenvoranschlägen auf der Grundlage einer detaillierten Analyse aller für die Durchführung der Bauarbeiten erforderlichen Ressourcen (Material, Arbeit, Ausrüstung).

RFID (Radio Frequency Identification) ist eine Technologie zur automatischen Identifizierung von Objekten mit Hilfe von Funksignalen, die zur Verfolgung von Materialien, Maschinen und Personal eingesetzt wird.

ROI (Return on Investment) ist ein Indikator, der das Verhältnis zwischen Gewinn und investierten Mitteln widerspiegelt und zur Bewertung der Wirksamkeit von Investitionen dient.

SaaS (Software as a Service**)** ist ein Modell für Software als Dienstleistung, bei dem die Anwendungen von einem Anbieter gehostet und den Benutzern über das Internet zur Verfügung gestellt werden.

SCM (Supply Chain Management) - Lieferkettenmanagement, das die Koordinierung und Optimierung aller Prozesse von der Materialbeschaffung bis zur Auslieferung der fertigen Produkte umfasst.

Datensilos sind isolierte Informationsspeicher innerhalb eines Unternehmens, die nicht mit anderen Systemen integriert sind, was den Datenaustausch erschwert und ineffizient ist.

SQL (Structured Query **Language)** ist eine strukturierte Abfragesprache, die für die Arbeit mit relationalen Datenbanken verwendet wird.

SQLite ist ein leichtgewichtiges, einbettbares, plattformübergreifendes DBMS, das keinen separaten Server benötigt und grundlegende SQL-Funktionen unterstützt, die in mobilen Anwendungen und eingebetteten Systemen weit verbreitet sind.

Strukturierte Daten - Informationen, die in einem bestimmten Format mit einer klaren Struktur organisiert sind, wie z. B. in relationalen Datenbanken oder Tabellen.

Lose strukturierte Daten - Informationen mit teilweiser Organisation und flexibler Struktur, wie z. B. JSON oder XML, bei denen verschiedene Elemente unterschiedliche Sätze von Attributen enthalten können.

Eine Entität ist ein konkretes oder abstraktes Objekt der realen Welt, das eindeutig identifiziert, beschrieben und in Form von Daten dargestellt werden kann.

Überwachtes Lernen - Eine Art des maschinellen Lernens, bei dem ein Algorithmus auf partitionierten Daten trainiert wird, wobei das gewünschte Ergebnis für jedes Beispiel bekannt ist.

Die Taxonomie ist ein hierarchisches Klassifizierungssystem, das dazu dient, Elemente auf der Grundlage gemeinsamer Merkmale systematisch zu kategorisieren.

Der Titanic-Datensatz ist ein beliebter Datensatz zum Trainieren und Testen von Modellen für maschinelles Lernen.

Training - Der Prozess, bei dem ein Algorithmus für maschinelles Lernen Daten analysiert, um Muster zu erkennen und ein Modell zu erstellen.

Transferlernen ist eine Technik des maschinellen Lernens, bei der ein für eine Aufgabe trainiertes Modell als Ausgangspunkt für eine andere Aufgabe verwendet wird.

Datenumwandlung - Der Prozess der Änderung des Formats, der Struktur oder des Inhalts von Daten zur späteren Verwendung.

Datenanforderungen - formalisierte Kriterien, die die Struktur, das Format, die Vollständigkeit und die Qualität von Informationen definieren, die zur Unterstützung von Geschäftsprozessen erforderlich sind.

Die Uberisierung der Baubranche ist der Prozess der Umwandlung traditioneller Geschäftsmodelle im Bauwesen unter dem Einfluss digitaler Plattformen, die eine direkte Interaktion zwischen Kunden und Auftragnehmern ohne Zwischenhändler ermöglichen.

Uniclass ist ein Klassifizierungssystem für Gebäudeelemente der zweiten und dritten Generation, das im Vereinigten Königreich weit verbreitet ist.

USD (Universal Scene Description) ist ein Datenformat, das für die Computergrafik entwickelt wurde, aber aufgrund seiner einfachen Struktur und seiner Unabhängigkeit von geometrischen Kernen auch in technischen Systemen Anwendung findet.

Bei der Datenvalidierung werden Informationen anhand festgelegter Kriterien und Anforderungen überprüft, um die Richtigkeit, Vollständigkeit und Konsistenz der Daten zu gewährleisten.

Vektordatenbank - Ein spezieller Datenbanktyp, der Daten als mehrdimensionale Vektoren für eine effiziente semantische Suche und einen Objektvergleich speichert.

Die Vektordarstellung (Einbettung) ist eine Methode zur Umwandlung von Daten in mehrdimensionale numerische Vektoren, die es Maschinenalgorithmen ermöglicht, Informationen effizient zu verarbeiten und zu analysieren.

VectorOps ist eine Methodik, die sich auf die Verarbeitung, Speicherung und Analyse multidimensionaler Vektordaten konzentriert, die insbesondere in Bereichen wie digitale Zwillinge und semantische Suche relevant sind.

Visualisierung - Grafische Darstellung von Daten zur besseren Wahrnehmung und Analyse von Informationen.

Die alphabetische Kategorisierung der Begriffe erfolgte nach ihren englischen Bezeichnungen.

LISTE DER REFERENZEN UND ONLINE-MATERIALIEN

[1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12. Dezember 2016. [Online]. Verfügbar: https://www.gartner.com/en/documents/3530919. [Datum des Zugriffs: 1. März 2025].

[2] KPMG, "Vertraute Herausforderungen - neue Ansätze. 2023 Global Construction Survey", 1. Januar 2023. [Online]. Verfügbar: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf. [Zugriff: 5. März 2025].

[3] F. R. Barnard, "Ein Bild sagt mehr als tausend Worte", 10 Mapi 1927. [Online]. Verfügbar: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Datum der Bezugnahme: 15. März 2025].

[4] M. Bastian, "Microsoft-CEO Satya Nadella sagt, die Selbstbehauptung, AGI sei "unsinniges Benchmark-Hacking"," 21. Februar 2025. [Online]. Verfügbar: https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/. [Datum des Zugriffs: 15. März 2025].

[5] W. E. Forum, "Forum Zukunft des Bauens gestalten - Eine Landschaft im Wandel:", 1. Januar 2016. [Online]. Verfügbar: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Datum des Zugriffs: 2. März 2025].

[6] C. D. Gillespie, "Clay: Earth's entanglement in the age of clay", 2024. [Online]. Verfügbar: https://ufl.pb.unizin.org/imos/chapter/clay/.

[7] "Papyrus aus dem 3. Jahrhundert v. Chr. Sprache ist Griechisch", 2024. [Online]. Available: https://www.facebook.com/429710190886668/posts/595698270954525.

[8] "Überwachung: Nutzung der zur Verfügung stehenden Instrumente", 1980. [Online]. Verfügbar: https://pubmed.ncbi.nlm.nih.gov/10246720/. [Datum der Ansprache: 15. März 2025].

[9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15. Februar 2015. [Online]. Verfügbar: https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf. [Datum des Zugriffs: 15. März 2025].

[10] Skanska USA, "Herbstmarkttrends im Bauwesen", 2. November 2023. [Online]. Available: https://x.com/SkanskaUSA/status/1720167220817588714.

[11] "Oxford Essential Quotations (4th ed.)", Oxford University Press, 2016. [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236. [Datum des Zugriffs: 1. März 2025].

[12] "Zitat: Sondergaard über Datenanalyse," [Online]. Verfügbar: https://www.causeweb.org/cause/resources/library/r2493. [Datum des Zugriffs: 15. März 2025].

[13] "Wie das weltweite Interesse an KI den Markt für Datenmanagement ankurbelt", 28. Mai 2024. [Online]. Verfügbar: https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/. [Datum des Zugriffs: 15. März 2025].

[14] И. McCue, "ERP-Geschichte", 2024. [Online]. Verfügbar: https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml.

[15] erpscout, "ERP-Preis: Was kostet ein ERP-System?"," [Online]. Verfügbar: https://erpscout.de/en/erp-costs/. [Datum des Zugriffs: 15. März 2025].

[16] softwarepath, "Was 1.384 ERP-Projekte über die Auswahl von ERP sagen (ERP-Report 2022)", 18. Januar 2022. [Online]. Verfügbar: https://softwarepath.com/guides/erp-report. [Datum des Zugriffs: 15. März 2025].

[17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16. Dezember 2016. [Online]. Verfügbar: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf. [Datum der Adresse: 1. März 2025].

[18] McKinsey, "Das datengesteuerte Unternehmen des Jahres 2025", 28. Januar 2022. [Online]. Verfügbar: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025. [Datum der Bezugnahme: 22. Mai 2024].

[19] Wikipedia, "Moore's law," [Online]. Verfügbar: https://en.wikipedia.org/wiki/Moore%27s_law. [Datum des Zugriffs: 15. März 2025].

[20] Accenture, "Mehr Wert schaffen mit Kapitalprojekten", 1. Januar 2020. [Online]. Verfügbar: https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf. [Datum des Zugriffs: 3. März 2024].

[21] Б. Marr, "Wie viele Daten erzeugen wir jeden Tag? The Mind-Blowing Stats Everyone Should Read," 2018. [Online]. Verfügbar: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read.

[22] "Wie viele Daten werden täglich produziert?", 2024. [Online]. Verfügbar: https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/.

[23] T. Sullivan, "AI and the global 'datasphere': how much information will humanity will have by 2025?", 2024. [Online]. Verfügbar unter: https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By- 2025.html.

[24] Statista, "Gesamtzahl der gedruckten Bücher, die in verschiedenen Regionen Westeuropas in jedem halben Jahrhundert zwischen 1454 und 1800 hergestellt wurden", [Internet]. [Online]. Verfügbar: https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/. [Datum der Bezugnahme: 1. März 2025.]

[25] "Preisbeispiele", 2024. [Online]. Verfügbar unter: https://cloud.google.com/storage/pricing-examples.

[26] M. Ashare, "Unternehmen lagern Datenspeicherung aus, da die Komplexität steigt", 10. Mai 2024. [Online]. Verfügbar: https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/. [Datum des Zugriffs: 15. März 2025].

[27] JETSOFTPRO, "SaaS ist tot? Microsoft CEO's Shocking Prediction Explained," 13 Jan. 2025. [Online]. Verfügbar: https://jetsoftpro.com/blog/saas-is-dead/.

[28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12. Dezember 2024. [Online]. Verfügbar: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Datum des Zugriffs: 15. März 2025].

[29] GoodReads, "Tim Berners-Lee," [Online]. Verfügbar: https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than. [Datum des Zugriffs: 15. März 2025].

[30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1. Januar 2023. [Online]. Verfügbar: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Zugriff: 5. März 2025].

[31] И. Deininger, B. Koch, R. Bauknecht, and M. Langhans, "Using digital models for decarbonising a production site: An example of connecting a building model, a production model and an energy model," 2024. [Online]. Verfügbar: https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Study_of_Connecting_the_Building_Model_Production_Model_and_Energy_Model.

[32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY", 1. Februar 2017. [Online]. Verfügbar: https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf.

[33] Taskforce Bauwesen an den stellvertretenden Premierminister, "Rethinking Construction", 1. Oktober 2014. [Online]. Verfügbar: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.

[34] Forbes, "Without An Opinion, You're Just Another Person With Data," 15 Mar. 2016. [Online]. Verfügbar unter: https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/. [Datum des Zugriffs: 15. März 2025].

[35] Wikiquote, "Charles Babbage," [Online]. Verfügbar: https://en.wikiquote.org/wiki/Charles_Babbage. [Datum des Zugriffs: 15. März 2025].

[36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves," 12 Mar. 2025. [Online]. Verfügbar: https://news.sap.com/2025/03/new-research-executive-trust-ai/. [Datum des Zugriffs: 15. März 2025].

[37] The Canadian Construction Association und KPMG in Kanada, 2021, "Construction in a digital world", 1. Mai 2021. [Online]. Verfügbar: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Datum des Zugriffs: 5. März 2025].

[38] ZCS, "Decoding the Fifth Industrial Revolution"," [Online]. Verfügbar unter: https://www.pwc.in/decoding-the-fifth-industrial-revolution.html. [Datum des Zugriffs: 15. März 2025].

[39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.

[40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "Der Wert von Open Source Software", 24. Januar 2024. [Online]. Verfügbar: https://www.hbs.edu/faculty/Pages/item.aspx?num=65230. [Datum des Zugriffs: 15. März 2025].

[41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook," 1 September 2008. [Online]. Verfügbar unter: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf.

[42] McKinsey, "Verbesserung der Produktivität im Bauwesen", [Online]. Verfügbar unter: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity. [Datum der Ansprache: 15. März 2025].

[43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector", 19 Jan. 2023. [Online]. Verfügbar: https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/. [Datum des Zugriffs: 1. März 2025].

[44] McKinsey, "Produktivität im Baugewerbe ist keine Option mehr", 9. August 2024. [Online]. Verfügbar: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional. [Datum des Zugriffs: 5. März 2025].

[45] ING Group, "Nachlassende Produktivität im Baugewerbe treibt die Baukosten in die Höhe", 12. Dezember 2022. [Online]. Verfügbar: https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/. [Datum der Ansprache: 15. März 2025].

[46] M. Berman, "Microsoft CEO's Shocking Prediction: 'Agents Will Replace ALL Software'," 19. Dezember 2024. [Online]. Verfügbar: https://www.youtube.com/watch?v=uGOLYz2pgr8. [Datum des

Zugriffs: 15. März 2025].

[47] Business Insider, "Anthropic's CEO says that in 3 to 6 months, AI will be writing 90% of the code software developers were in charge of," 15 Mar 2025. [Online]. Verfügbar: https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3. [Datum des Zugriffs: 30. März 2025].

[48] Statista, "Popularitätsvergleich der Datenbankmanagementsysteme (DBMS) weltweit ab Juni 2024, nach Kategorie," Juni 2024. [Online]. Verfügbar: https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/. [Datum der Adresse: 15. März 2025].

[49] DB-Engines, "DB-Engines Ranking", [Online]. Verfügbar: https://db-engines.com/en/ranking. [Datum der Bezugnahme: 15. März 2025].

[50] "Stack Overflow Developer Survey 2023", 2024. [Online]. Verfügbar: https://survey.stackoverflow.co/2023/.

[51] "SQL", 2024. [Online]. Verfügbar: https://en.wikipedia.org/wiki/SQL.

[52] "Strukturierte und unstrukturierte Daten: Was ist der Unterschied?", 2024. [Online]. Verfügbar: https://www.ibm.com/blog/structured-vs-unstructured-data/.

[53] DataDrivenConstruction, "PDF-VERGLEICH VON DATENFORMATEN FÜR BAUPROJEKTE", 23 Apr. 2024. [Online]. Verfügbar: https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf.

[54] Website "Building Information Modeling Whitepaper", 2003. [Online]. Verfügbar unter: https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112.

[55] A. Boiko, "Lobbying-Kriege und BIM-Entwicklung. Teil 5: BlackRock ist der Herr aller Technologien. Wie Konzerne Open-Source-Code kontrollieren", 2024. [Online]. Verfügbar: https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/.

[56] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 14 11 2011. [Online]. Verfügbar: https://isicad.net/articles.php?article_num=14805. [Datum des Vortrags: 02 2025].

[57] C. Eastman und A. Cthers, "Eastman, Charles; und Cthers," September 1974. [Online]. Verfügbar: https://files.eric.ed.gov/fulltext/ED113833.pdf. [Datum des Zugriffs: 15. März 2025].

[58] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 11. November 2011. [Online]. Verfügbar: https://isicad.net/articles.php?article_num=14805. [Datum der Bezugnahme: 15. März 2025].

[59] D. Weisberg, "Geschichte des CAD", 12. Dezember 2022. [Online]. Verfügbar: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Datum der Bezugnahme: 15. März 2025].

[60] ADSK, "Weißbuch Gebäudedatenmodellierung", 2002. [Online]. Verfügbar unter: https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Datum der Bezugnahme: 15. März 2025].

[61] ADSK, "White Paper Building Information Modeling in der Praxis", [Online]. Verfügbar unter: https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Datum des Zugriffs: 15. März 2025].

[62] A. Boiko, "Lobbying-Kriege und BIM-Entwicklung. Teil 2: open BIM VS closed BIM. Europa VS der Rest der Welt," 2024. [Online]. Verfügbar: https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/.

[63] A. Boiko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs," 2024. [Online]. Verfügbar: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.

[64] ADSK, "Whitepaper BIM," 2002. [Online]. Verfügbar: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Datum der Adresse: 15. März 2025].

[65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale", [Online]. Verfügbar unter: https://web.archive.org/web/20010615093351/http:/www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Datum des Zugriffs: 15. März 2025].

[66] M. Shacklett, "Structured and unstructured data: Die wichtigsten Unterschiede", 2024. [Online]. Verfügbar: https://www.datamation.com/big-data/structured-vs-unstructured-data/.

[67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Verfügbar unter: https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/.

[68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the," 2004. [Online]. Verfügbar: https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf. [Datum des Zugriffs: 02 2025].

[69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Verfügbar: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Datum des Zugriffs: 15. März 2025].

[70] Analyticsindiamag, "6 Most Time Consuming Tasks For Data Scientists", 15. Mai 2019. [Online]. Verfügbar: https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/.

[71] BizReport, "Report: Data scientists spend a bulk of time clean up," 06 July 2015. [Online]. Verfügbar: https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html. [Datum des Zugriffs: 5. März 2025].

[72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27. Juli 2015. [Online]. Verfügbar: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Datum des Zugriffs: 15. März 2025].

[73] Б. Cyphers und K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [Online]. Verfügbar unter: https://www.eff.org/wp/interoperability-and-privacy.

[74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 1. Oktober 2013. [Online]. Verfügbar: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information. [Datum des Zugriffs: 15. März 2025].

[75] A. Boiko, "Der Kampf um offene Daten in der Bauindustrie. Die Geschichte von AUTOLISP, intelliCAD, openDWG, ODA und openCASCADE," 15 05 2024. [Online]. Verfügbar: https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393. [Datum des Vortrags: 16 02 2025].

[76] Wikipedia, "Microsoft und Open Source", [Online]. Verfügbar: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Datum des Zugriffs: 15. März 2025].

[77] TIME, "The Gap Between Open and Closed AI Models Might Be Shrinking. Here's Why That Matters," 5. November 2024. [Online]. Verfügbar: https://time.com/7171962/open-closed-ai-models-epoch/. [Datum des Zugriffs: 15. März 2025].

[78] The Verge, "Mehr als ein Viertel des neuen Codes bei Google wird von KI erzeugt", 29. Oktober 2024. [Online]. Available: https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024. [Datum des Zugriffs: 15. März 2025].

[79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15. Dezember 2020. [Online]. Verfügbar unter: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-processing. [Zugriff: 15. März 2025].

[80] PWC, "PwC Open Source Monitor 2019", 2019. [Online]. Available: https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf. [Accessed 15 Mar. 2025].

[81] Travers Smith, "Das offene Geheimnis: Open-Source-Software", 2024. [Online]. Verfügbar: https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/. [Datum des Zugriffs: 15. März 2025].

[82] Deloitte, "Der Datentransferprozess bei Unternehmenstransformationen", 2021. [Online]. Verfügbar: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf. [Datum der Bezugnahme: 15. März 2025].

[83] gov.uk, "Data Analytics and AI in Government Project Delivery," 20 Mar. 2024. [Online]. Verfügbar: https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery. [Datum des Zugriffs: 5. März 2025].

[84] Zitat Ursprung: "Alles sollte so einfach wie möglich gemacht werden, aber nicht einfacher", 13. Mai 2011. [Online]. Verfügbar: https://quoteinvestigator.com/2011/05/13/einstein-simple/. [Datum des Zugriffs: 15. März 2025].

[85] "Transformer (Architektur des tiefen Lernens)," [Online]. Verfügbar: https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture). [Datum des Zugriffs: 15. März 2025].

[86] "Python-Pakete Download-Statistiken", 2024. [Online]. Verfügbar: https://www.pepy.tech/projects/pandas.

[87] Interview Bit, "Top 10 Python-Bibliotheken", 2023. [Online]. Verfügbar: https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more. [Datum des Zugriffs: 30. März 2025].

[88] "NVIDIA und HP verbessern Datenwissenschaft und generative KI auf Workstations", 7. März 2025. [Online]. Verfügbar: https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations. [Datum des Zugriffs: 15. März 2025].

[89] P. Orac, "Wie man einen DataFrame mit Millionen von Zeilen in Sekunden verarbeitet", 2024. [Online]. Verfügbar: https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds.

[90] Ç. Uslu, "Was ist Kaggle?", 2024. [Online]. Verfügbar: https://www.datacamp.com/blog/what-is-kaggle.

[91] "NVIDIA CEO Jensen Huang Keynote auf der COMPUTEX 2024", 2. Juni 2024. [Online]. Verfügbar: https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w. [Zugriff am 15. März 2025].

[92] "Mitglieder: Gründer und korporative Mitglieder", 2024. [Online]. Verfügbar: https://www.opendesign.com/member-showcase.

[93] A. Boiko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts," 24 November 2024. [Online]. Verfügbar: https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a. [Datum des Zugriffs: 23. Februar 2025].

[94] A. Boiko, "Die Post-BIM-Welt. Der Übergang zu Daten und Prozessen und die Frage, ob die Bauindustrie Semantik, Formate und Interoperabilität braucht", 20. Dezember 2024. [Online]. Verfügbar: https://boikoartem.medium.com/the-post-bim-world-7e35b7271119. [Datum des Vortrags: 23. Februar 2025].

[95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE," 2016. [Online]. Available: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Datum des Zugriffs: 23. Februar 2025].

[96] Harvard Business Review, "Bad Data Costs the U.S. $3 Trillion Per Year", 22. September 2016. [Online]. Verfügbar unter: https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

[97] Delpha, "Auswirkungen der Datenqualität", 1. Januar 2025. [Online]. Verfügbar: https://delpha.io/impacts-of-data-quality/.

[98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects," [Online]. Verfügbar: https://www.wbdg.org/resources/design-for-maintainability. [Datum des Zugriffs: 15. März 2025].

[99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", April 2014. [Online]. Verfügbar: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf. [Datum des Zugriffs: 15. März 2025].

[100] Gartner, "Datenqualität: Best Practices für genaue Einblicke", 1. Januar 2025. [Online]. Verfügbar: https://www.gartner.com/en/data-analytics/topics/data-quality.

[101] "For Want of a Nail," [Online]. Verfügbar: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Datum des Zugriffs: 15. März 2025].

[102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", Oktober 2013. [Online]. Verfügbar unter: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [Datum des Zugriffs: 15. März 2025].

[103] EY, "Der Weg zur Kohlenstoffneutralität", 10. März 2023. [Online]. Verfügbar: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Datum des Zugriffs: 15. März 2025].

[104] PWC, "ESG-Bewusstsein", 1. Juli 2024. [Online]. Verfügbar: https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf. [Datum des Zugriffs: 15. März 2025].

[105] G. Hammond, "Verkörperter Kohlenstoff - Das Inventar von Kohlenstoff und Energie (ICE)", 2024. [Online]. Verfügbar unter: https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-

BSRIA-ICE-guide.pdf.

[106] "CO2_Berechnung des verkörperten Kohlenstoffs", 2024. [Online]. Verfügbar: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.

[107] McKinsey, "Imagining Construction's Digital Future", 24. Juni 2016. [Online]. Verfügbar: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future. [Datum des Zugriffs: 25. Februar 2025].

[108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10. Oktober 2024. [Online]. Verfügbar: https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/. [Abrufdatum: 15. März 2025].

[109] SAS, "Data Lake und Data Warehouse - kennen Sie den Unterschied", [Online]. Verfügbar: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Datum der Bezugnahme: 15. März 2025].

[110] ADSK, "Gebäudedatenmodellierung", 2002. [Online]. Verfügbar unter: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Datum der Adresse: 15. März 2025].

[111] A. Boiko, "BIM History Map", 2024. [Online]. Verfügbar unter: https://bigdataconstruction.com/history-of-bim/.

[112] A. S. Borkowski, "Definitionen von BIM durch Organisationen und Normen", 27. Dezember 2023. [Online]. Verfügbar: https://encyclopedia.pub/entry/53149. [Datum der Ansprache: 5. März 2025].

[113] CAD-Anbieter, "OPEN BIM-Programm", 2012. [Online]. Verfügbar: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Datum der Bezugnahme: 30. März 2025].

[114] Wikipedia, "Industry Foundation Classes," [Online]. Verfügbar: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Datum des Zugriffs: 15. März 2025].

[115] Wikipedia, "IGES," [Online]. Verfügbar unter: https://en.wikipedia.org/wiki/IGES. [Datum des Zugriffs: 30. März 2025].

[116] A. Boiko, "Geschichte von CAD (BIM)", 15. Dezember 2021. [Online]. Verfügbar: https://miro.com/app/board/o9J_laML2cs=/. [Datum der Ansprache: 24. Februar 2025].

[117] T. K. K. A. O. F. B. C. E. L. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Referenzstudie zur IFC-Softwareunterstützung: der GeoBIM-Benchmark 2019 - Teil I," 8 Jan. 2021. [Online]. Verfügbar: https://arxiv.org/pdf/2007.10951. [Datum des Zugriffs: 5. März 2025].

[118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM," 13. April 2021. [Online]. Verfügbar: https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be.

[Datum des Vortrags: 5. März 2025].

[119] D. Ares, "RETS in Real Estate: Warum es entscheidend für Effizienz und Wachstum ist", 17. Dezember 2024. [Online]. Verfügbar: https://www.realpha.com/blog/rets-importance-in-real-estate-explained. [Datum des Zugriffs: 5. März 2025].

[120] "Flex-Token-Kosten", 2024. [Online]. Verfügbar unter: https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex.

[121] A. Boiko, "Vergessen Sie BIM und demokratisieren Sie den Zugang zu Daten (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024. [Online]. Verfügbar: https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Bauprozessmanagment-und-Bay-Bauindustrie.pdf.

[122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh, und F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [Online]. Verfügbar: https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle.

[123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle", 1. Januar 2018. [Online]. Verfügbar unter: https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html. [Datum des Zugriffs: 2. März 2025].

[124] Apple.com, "Pixar, Adobe, Apple und NVIDIA bilden Allianz für OpenUSD, um offene Standards für 3D-Inhalte voranzutreiben", 1. August 2023. [Online]. Verfügbar: https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/. [Zugriff: 2. März 2025].

[125] AECmag, "ADSKs granulare Datenstrategie", 25. Juli 2024. [Online]. Verfügbar: https://aecmag.com/technology/autodesks-granular-data-strategy/. [Datum des Zugriffs: 15. März 2025].

[126] A. Boiko, "The Age of Change: IFC is a thing of past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts," 24 11 2024. [Online]. Verfügbar: https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a. [Datum des Beitrags: 23 fevryall 2025].

[127] A. Boiko, "ENG BIM Cluster 2024 | Der Kampf um Daten und die Anwendung von LLM und ChatGPT im Bauwesen", 7. August 2024. [Online]. Verfügbar: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [Zugriff: 15. März 2025].

[128] "Jeffrey Zeldman präsentiert", 6. Mai 2008. [Online]. Verfügbar: https://zeldman.com/2008/05/06/content-precedes-design/. [Datum des Zugriffs: 15. März 2025].

[129] A. Boiko, "DWG Analyse mit ChatGPT | DataDrivenConstruction," 5 Mar. 2024. [Online]. Verfügbar: https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction. [Datum der Bezugnahme: 15. März 2025].

[130] McKinsey, "The McKinsey guide to outcompeting in the age of digital and AI", 2023. [Online]. Verfügbar unter: https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired. [Datum der Veröffentlichung: 30. März 2025].

[131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31. März 2016. [Online]. Verfügbar: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/. [Datum des Zugriffs: 15. März 2025].

[132] J. Bertin, "Grafik und grafische Informationsverarbeitung", 8. September 2011. [Online]. Verfügbar: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Datum des Zugriffs: 15. März 2025].

[133] CauseWeb, "Wells/Wilks über statistisches Denken", [Online]. Verfügbar: https://www.causeweb.org/cause/resources/library/r1266. [Datum des Zugriffs: 15. März 2025].

[134] Ministrymagazine, "Wie die Wissenschaft die Schöpfung entdeckte", Januar 1986. [Online]. Verfügbar: https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation. [Datum der Bezugnahme: 15. März 2025].

[135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23. Mai 2017. [Online]. Available: https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation. [Accessed 15 May 2024].

[136] "Wie man eine Datenarchitektur aufbaut, um Innovationen voranzutreiben - heute und morgen", 3. Juni 2020. [Online]. Verfügbar: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow. [Datum des Vortrags: 15. März 2025].

[137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630. [Datum der Ansprache: 15. März 2025].

[138] "Konvertoren", 2024. [Online]. Verfügbar: https://datadrivenconstruction.io/index.php/convertors/.

[139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1. Januar 2017. [Online]. Verfügbar: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf. [Datum des Zugriffs: 18. Februar 2025].

[140] "Pipeline im Bau", 2024. [Online]. Verfügbar: https://datadrivenconstruction.io/index.php/pipeline-in-construction/.

[141] Wikipedia, "Apache NiFi", 1. Januar 2025. [Online]. Verfügbar: https://de.wikipedia.org/wiki/Apache_NiFi. [Datum des Zugriffs: 5. März 2025].

[142] n8n, "Gmail AI Auto-Responder: Entwurfsantworten auf eingehende E-Mails erstellen", 1. Mai 2024. [Online]. Verfügbar: https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-

incoming-emails/. [Zugriff am 15. März 2025].

[143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail", 1. März 2025. [Online]. Verfügbar: https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/. [Datum des Zugriffs: 15. März 2025].

[144] B. T. O'Neill, "Misserfolgsraten für Analyse-, KI- und Big-Data-Projekte = 85% - igitt," 1. Januar 2025. [Online]. Verfügbar: https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/.

[145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.

[146] T. J. S. a. J. S. Jesse Perla, "A Problem that Stumped Milton Friedman," Quantitative Economics with Julia, 1 Jan. 2025. [Online]. Verfügbar: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Datum des Zugriffs: 1. Mai 2024].

[147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.

[148] A. Boiko, "San Francisco. Baugewerbe 1980-2019", 2024. [Online]. Verfügbar unter: https://www.kaggle.com/search?q=San+Francisco.+Bau+Sektor+1980-2019.

[149] A. Boiko, "Kaggle: RVT IFC reicht 5000 Projekte ein", 2024. [Online]. Verfügbar: https://www.kaggle.com/datasets/artemboiko/rvtifc-projects.

[150] CFMA, "Preparing for the Future with Connected Construction," [Online]. Verfügbar unter: https://cfma.org/articles/preparing-for-the-future-with-connected-construction. [Zugriff am 15. März 2025].

[151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing," 22. Mai 2017. [Online]. Verfügbar: https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2017/m05/cisco-survey-reveals-close-to-three-fourths-of-iot-projects-are-failing.html.

[152] "Erforderliche Bedingungen für die Konservierung von Pflanzenfossilien", 2024. [Online]. Verfügbar: https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html.

[153] "BlackRock's Fink über Anleihen, Fusionen und Übernahmen, die US-Rezession und die Wahl: Vollständiges Interview," 2023. [Online]. Verfügbar: https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video.

[154] cio, "12 berühmte KI-Katastrophen", 02. Oktober 2024. [Online]. Verfügbar: https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html. [Datum des Zugriffs: 15. März 2025].

[155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.

[156] Wikipedia, "Kaggle," 1. Januar 2025. [Online]. Verfügbar: https://en.wikipedia.org/wiki/Kaggle. [Datum des Zugriffs: 15. März 2025].

[157] Kaggle, "Titanic - Maschinelles Lernen aus der Katastrophe", 1. Januar 2025. [Online]. Verfügbar: https://www.kaggle.com/competitions/titanic/overview. [Zugriff am 10. März 2025].

[158] Ш. Johri, "Creating ChatGPT: From Data to Dialogue", 2024. [Online]. Verfügbar unter: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/.

[159] П. Domingos, "Ein paar nützliche Dinge über maschinelles Lernen", 2024. [Online]. Verfügbar: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

[160] J. Saramago, "Zitierfähiges Zitat", [Online]. Verfügbar: https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered. [Datum des Zugriffs: 17. März 2025].

[161] NVIDIA, "Verbessern Sie Ihre Trainingsdaten mit den neuen NVIDIA NeMo Curator Classifier-Modellen", 19. Dezember 2024. [Online]. Verfügbar: https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/. [Zugriff am 25. März 2025].

[162] "NVIDIA kündigt die Veröffentlichung von Cosmos World Foundation Models und Physical AI Data Tools an", 18. März 2025. [Online]. Verfügbar: https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools. [Datum des Zugriffs: 25. März 2025].

[163] NVIDIA, "NVIDIA Isaac Sim," [Im Internet]. Verfügbar unter: https://developer.nvidia.com/isaac/sim. [Zugriff am 25. März 2025].

[164] M. Quarterly, "Warum digitale Strategien scheitern", 25 Jan. 2018. [Online]. Verfügbar unter: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail. [Datum des Zugriffs: 15. März 2025].

[165] M. J. Perry, "Meine Lieblingszitate von Milton Friedman", 17. November 2006. [Online]. Verfügbar unter: https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/. [Datum des Zugriffs: 1. März 2025].

[166] J. A. Wheeler, "Information, Physik, Quanten: die Suche nach Verbindungen", 1990.

[169] A. Boiko, "Lobbying-Kriege und BIM-Entwicklung. Teil 5: BlackRock ist der Herr aller Technologien. Wie Konzerne Open-Source-Code kontrollieren", 2024. [Online]. Verfügbar: https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d.

[170] T. Krijnen und J. Beetz, "A SPARQL query engine for binary-formatted IFC building models," *Advanced Engineering Informatics*, 2024.

[171] "Anzahl der Unternehmen im Bausektor im Vereinigten Königreich im Jahr 2021, nach Unternehmensgröße," 2024. [Online]. Verfügbar: https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/.

[172] "5.000 IFC&RVT-Projekte", 2024. [Online]. Verfügbar: https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io.

[173] M. Popova, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe", 2008. [Online]. Verfügbar: https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/. [Datum der Bezugnahme: Februar 2025].

[174] *Lobbying Wars Over Data in Construction | Techno-Feudalismus und die Geschichte der verborgenen Vergangenheit von BIM.* [Film]. Deutschland: Artem Boiko, 2023.

[175] A. Boiko, "CHATGPT WITH REVIT AND IFC | Automatic retrieval of documents and data from projects," 16 Nov. 2023. [Online]. Verfügbar: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Datum des Zugriffs: 2. März 2025].

[176] M. &. Company, "Three new mandates for capturing a digital transformation's full value," 22 Jan. 2022. [Online]. Verfügbar: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value. [Datum des Zugriffs: 15. Februar 2025].

[177] KPMG, "Bauwesen in einer digitalen Welt", 1. Mai 2021. [Online]. Verfügbar unter: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-de.pdf. [Datum des Zugriffs: 5. April 2024].

[178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17. März 2023. [Online]. Verfügbar: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Datum des Zugriffs: 15. Februar 2025].

[179] O. Business, "Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models,'" 10 Jan. 2025. [Online]. Verfügbar: https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models. [Datum des Zugriffs: 15. März 2025].

[180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," 23 Mar. 2016. [Online]. Available: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/. [Datum des Zugriffs: 15. März 2025].

[181] Amt für Auswärtige Angelegenheiten, Commonwealth und Entwicklung des Vereinigten Königreichs, "Digitale Entwicklungsstrategie 2024 bis 2030", 18. März 2024. [Online]. Verfügbar: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030. [Datum des Zugriffs: 15. März 2025].

[182] "Vision und Strategie in der Bauplanungsbranche", 7. November 2003. [Online]. Verfügbar: https://web.archive.org/web/20030711125527/http:/usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112. [Datum der Ansprache: 5. März 2025].

[183] M. Bocharov, "Informationsmodellierung", März 2025. [Online]. Verfügbar: https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5. [Datum des Vortrags: 15. März 2025].

[184] "Integriertes Design-Through-Manufacturing: Vorteile und Gründe", 2000. [Online]. Verfügbar: https://web.archive.org/web/20010615093351/http:/www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Datum der Bezugnahme: 25. März 2025].

[185] CAD Vendor, "Open BIM Programme is a marketing campaign", 12 Mar. 2012. [Online]. Verfügbar: https://web.archive.org/web/20120827193840/http:/www.graphisoft.com/openbim/. [Datum der Adresse: 30. März 2025].

THEMA-INDEX

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477
AIA, 289
AIM, 3, 289
AMS, 14, 84, 151, 153
Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494
Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494
Apache ORC, 62, 330, 378
Apache-Parkett, 62, 67, 380, 381, 494
API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

B

BDS, 258, 259
Große Daten, 9, 67, 245, 383
BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 389, 413, 416, 448, 462, 466, 475, 476, 492, 494, 497
BlackBox, 240, 242, 243.
BMS, 8
Bokeh, 320, 337
BOM, 76, 77, 79, 257, 263
Bounding Box, 234, 373, 392, 393, 394, 414, 415, 480
BREP, 142, 234, 263, 264, 276, 283, 284

C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497
CAE, 16, 283
CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473
CAM, 16, 78
CAPEX, 14, 82
CDE, 84, 175, 388, 389, 390
ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421
Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443
CO_2, 72
CO_2, 229, 234, 235, 236, 237, 238, 239, 243
COBie, 156, 289, 292.
CoE, 56, 168, 169, 170, 477
Kopilot, 114, 116, 459
CPIXML, 143, 272, 273, 276, 277, 279, 285, 296
CPM, 14, 17, 62, 166, 175, 233, 326, 473
CQMS, 14, 84, 177, 178, 462
CRM, 109, 369, 459.
CRUD, 51, 459
CSG, 263
CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 354, 356, 373, 377, 378, 379, 380, 401, 407, 414, 474, 480

D

DAE, 276, 277, 278, 280, 281, 284, 285, 414
DAG, 362, 363, 365, 366
Dash, 320, 336, 337.
Datenverwaltung, 373, 395, 396, 398, 400, 401
Datensee, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
Data Lakehouse, 67, 373, 386, 387, 388
Datenminimalismus, 373, 395, 396, 400, 401
Daten-Sumpf, 373, 395, 397, 401
Data Warehouse, 382, 383, 400
Daten-as-a-Service, 487
datengesteuert, 50, 170, 460, 461, 484, 486
DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475
DataOps, 170, 373, 398, 399, 400, 401, 475
DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
DGN, 8, 140, 186, 227, 357
DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390
DXF, 8, 73, 277

E

ECM, 58, 175
ECS, 142
EIR, 289
eLOD, 289
ELT, 384, 385
EPM, 14, 166, 198
ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
ESG, 196, 235, 236, 238
ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497
Auszug, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Feder, 62, 123, 330, 378
FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
GIS, 58
GLTF, 143, 278
Google Sheets, 368, 370
Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380
HiPPO, 29, 37, 95, 424, 477, 484, 490
HTML, 123, 340, 365, 370

I

IDS, 289, 290, 291.
IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497
IGES, 262, 263, 276.
iLOD, 289
IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484
ISO 19650, 388

J

JavaScript, 320, 369, 378
JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480
Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
k-NN, 392, 393, 442, 445, 446, 447, 448
KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494
LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 166, 187, 219, 220, 220, 221, 222, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494
Last, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475
LOD, 287, 289
LOI, 287
LOMD, 287
Niedriger Code, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
MCAD, 77, 78, 257, 284
MEP, 14, 175, 492
MASCHE, 234, 283, 284, 285, 296, 462
Microsoft SQL, 65
Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
MRP, 11, 12
MS Projekt, 70
MySQL, 63, 64, 65, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372
NLP, 69, 394
No-Code, 368, 369
NURBS, 142, 282, 283, 284, 285
NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285
OCCT, 273
OCR, 69, 128, 132, 134
OmniClass, 154, 155, 156.
offene BIM, 142, 216, 219, 256, 261, 278, 291
Open Source, 45, 55, 97, 98, 107, 108, 273, 275, 494
EULE, 267, 268, 269.

P

Pandas, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494
Parkett, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480
PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497
PDM, 16
PHP, 63, 339
Pipeline, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476
PLM, 16, 246
PLN, 8, 272, 296
Plotly, 320, 336, 337, 372
PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487
PMS, 84, 151, 233
PostgreSQL, 63, 64, 65, 329, 395
Power BI, 320, 372, 482
privates Beteiligungskapital, 464, 487
Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479
QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 381, 429, 443

R

RAG, 111, 116
RDBMS, 63, 64, 65, 82, 89.
RDF, 267, 268, 269
RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474
RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482
ROI, 311, 317, 319, 321, 351, 370, 389, 479
DREHZAHL, 14, 84, 151, 331, 333, 334
RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329, 357, 376, 414, 415, 417, 497

S

SaaS, 24, 50, 51, 52, 458
UMFANG, 273, 277
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
Seaborn, 123, 320, 336, 337, 408, 412
SPARQL, 269
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494
SQLite, 63, 64, 65, 166, 167, 296, 329, 363
STUFE, 261, 262, 263, 266, 272, 276, 277, 292
Streamlit, 336, 337
SVF, 142, 276, 285

T

Transformation, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

U

Uniclass, 154, 155, 156.
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

V

VectorOps, 373, 398, 400, 401
VR, 84, 271, 285,

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474

ENTDECKEN SIE, WIE DATEN DIE BAUINDUSTRIE REVOLUTIONIEREN

Was ist enthalten

- Über 100 bedeutende datenbezogene Themen im AEC
- Mehr als 300 einzigartige Visualisierungen
- Mehr als 50 reale Business Cases
- Anwendungsorientierte LLM-und KI-Lösungen
- Codebeispiele und einsatzbereite Arbeitsabläufe

Themen

- Datengetriebene Konstruktion
- Digitale Transformation im AEC
- Datenanalyse und Automatisierung
- Datenqualität und -management
- CAD, BIM und Dateninteroperabilität
- LLMs, KI und ML im Bauwesen
- Prognose der Projektkosten und -zeiten

Für wen

- Bauprojektmanager
- Architekten und Bauingenieure
- BIM-Koordinatoren und Datenverwalter
- Digitale Transformation leitet ein
- IT-und KI-Fachleute im AEC
- Stadtplaner und Berater für Nachhaltigkeit
- Architektur-und Ingenieurstudierende
- Pädagogen und Wissenschaftler

Webseite

- www.datadrivenconstruction.io

Kommentare zur ersten Auflage:

«Boiko ist der James Carville der IT – in des letzteren viel zitiertem: „It's the economy, stupid" muss für dieses famose Buch nur ein Wort getauscht werden. „It's the data, stupid." (nicht die Software) Und sich im Daten-Universum zurechtzufinden, ist eine schon aufs Griechische zurückgehende Weisheit der alten Römer bis heute gültig: „Navigare necesse est". Der Autor navigiert seine Leser mit sicherer Hand und unbeirrbarem Kompass durch alle Tiefen und Untiefen des Daten-Ozeans, obendrein mit umfassendem historischen Anspruch und nicht zuletzt mit höchst originellen Grafiken...»

— Dr. Burkhard Talebitari

«Artem Boikos Buch ist ein Meilenstein für die Demokratisierung der Digitalisierung im Bauwesen – und ein echter Gamechanger für kleine und mittelständische Unternehmen (KMU). Mit beeindruckender Klarheit zeigt es, wie komplexe Building Information Modeling (BIM)-Daten mithilfe schlanker Konverter in einfach nutzbare, flache Formate überführt werden können. Damit wird der Zugang zu wertvollen Projektdaten endlich für jene geöffnet, die bislang durch technische oder finanzielle Hürden ausgeschlossen waren...»

— Dr. Michael Max Buehler

Dieser umfassende Leitfaden befähigt sowohl Fachleute als auch Einsteiger, sich in der dynamischen Welt des datengestützten Bauens zurechtzufinden. Von den Grundlagen des Datenmanagements über fortschrittliche digitale Workflows und KI-Tools bis hin zu praxisorientierten Anwendungen – dieses Buch ist Ihr Kompass für intelligentere, schnellere und effizientere Bauprozesse.

ISBN: 978-3-00078-230-5

9 783000 782305