建設業界におけるデータ
時代のナビゲーション

データ駆動型建設

建設業界におけるデータ
時代のナビゲーション

AIとLLMの利用事例

Artem Boiko



DATADRIVENCONSTRUCTION
ウェブサイトにある他の言語の本









データ駆動型建設

建設業界におけるデータ時代のナビゲーション

改訂第 2 版

アルテム・ボイコ

「ボイコは IT 界のジェームス・カーヴィルである。後者の名言 "It's the economy, stupid"（愚かなのは経済だ）において、この名著に置き換えるべき言葉はただひとつ。「データだよ、バカ」。(データの宇宙で自分の進むべき道を見つけるには、古代ローマ人がギリシャ語にまでさかのぼった格言「Navigare necesse est」が今日でも有効だ）。著者は確かな手腕と揺るぎない羅針盤、包括的な歴史的アプローチはもちろんのこと、非常に独創的なグラフィック、そして一見しただけではわからないユーモアのセンスで、データ海洋のあらゆる深みと浅瀬を読者にナビゲートする。ボイコの本に対する国際的な反応は、陶酔的な賛辞から、むしろ辛辣な懐疑論まで多岐にわたる。ボイコは独創的で独断的でないデータ思想家である。彼は読者に刺激的な洞察と、常に勇気ある、挑発的でさえある論文を提示し、さらなる思考を促す。潜在的合意主義というドイツの病に対する優れた薬である。ちなみに、上記のラテン語のことわざには補語がある。ボイコのデータ世界へのアプローチには当てはまらない。" データは生きており、その生命は必要であり、重要であるとは言わない。

- ブルクハルト・タレビタリ博士、フリーランス編集者。2013 年からアーンスト＆ゾーンが毎年発行するジャーナル「BIM 」を担当。

「アルテム・ボイコ氏の著書は、建設業界におけるデジタル化の民主化にとって画期的なものであり、中小企業にとって真のゲーム・チェンジャーとなる。特に画期的なのは、最新のオープンソースのローコードツールやノーコードツールを使うことで、企業はすでに、深いプログラミングの知識がなくても、効率的にデータをビジネスプロセスに統合し、有益に分析できるようになっていることだ。これにより、面倒な商用ソフトウェア・パッケージの高価な使用は不要になる。本書は行動への呼びかけである！建設業界のデジタルトランスフォーメーションを理解したいだけでなく、現実的かつ効率的に、そして将来を見据えた方法で、積極的にそれを実現したいと考えているすべての人にとって、本書は貴重なガイドとなる。今こそ、この知識を共有し、建設業界の生産性を持続的に向上させるために協力する時である」

-Dr. Michael Max Buehler, HTWG Konstanz の建設管理教授、GemeinWerk Ventures の共同経営者、DevvStream の独立取締役。

「DataDrivenConstruction の本は、複雑な設計や管理システムを持つ通常の建設業者の世界の枠を超えた最初のステップの一つである。アルテムは本書の中で、データを扱う現代のテクノロジーが私たちの前にどのような機会を開いてくれるかをわかりやすく示し、文字通り、あなたの仕事にすぐに応用できる具体的なステップを示している。建設業界におけるオートメーション・システムの行く末を理解したい人は、ぜひ本書を熟読し、建設業界におけるデータ革命がすでに私たちのドアをノックしていることを実感してほしい。今はオタクにしか興味がないが、数年後には BIM（）のように、このようなアプローチやソフトウェアがユビキタスになっているだろう！」

- RGD の IMT コンピテンスセンター、BIM & デジタルトランスフォーメーションの責任者であり、InfraBIM.Pro の創設者である Ihor Rogachew 氏。

「"Data Driven Construction" は、建築データを使った情報化施工の基本を生き生きと伝えている。情報の流れと基本的な経済概念を扱っているため、他の BIM 本とは一線を画している。ソフトウェアメーカーの視点を代弁しているだけでなく、基本的な概念を伝えようとしているからだ。読んで見る価値のある本である"」

- ヤコブ・ヒルン、Build Informed GmbH の CEO 兼共同創立者、イノベーション・フォーラム「On Top With BIM 」の発起人。

「"DataDrivenConstruction"は、タイトルにあるように、AECO のためのデータ主導の情報管理アプローチを扱った本で、非常にお勧めです。私は現在、様々なグループと議論を始めるためにこの本を使用しています。非常にわかりやすい参考書だと思います。AECO におけるツールの歴史的背景、データ、いくつかの重要な技術の紹介の徹底的な概要だけでなく、この本には、データソースとエンドユーザーの成果物の範囲をワークフローのサンプルとともに概説する、非常に有用な図が数多く含まれています。これらの図は、情報戦略を策定・監視し、BEP に貢献する際にもっと必要なものであり、PIM や AIM の境界線を重ね合わせることができる全体的な企業データモデルを定義するものである」

- ポール・ランズリー、アクメナ社主席コンサルタント、ロンドン交通局システム・インテグレーション・エンジニア。

「データが新しい石油"であるならば、我々はそれを定義し、見つけ、採掘し、精製し、価値あるものにすることを学ぶ必要がある。DataDrivenConstruction という本は、とても有益で洞察に富んでいる。この本は有益な歴史的背景を提供し、データを扱うことを平易な言葉で説明している。デジタルトランスフォーメーションに興味がある人にとっては、データの仕組み、構造、利用方法など、データについてよく理解することができる」

- ラルフ・モンタギュー、ArcDox 社ディレクター、BIM コーディネーターサミット・ディレクター、アイルランド国家規格庁 BIM ナショナルミラー委員会委員長。

「本書で強調されているように、情報は建設部門にとって極めて重要な資産であり、アクセス可能なフォーマットで提供されることで、正確な意思決定が容易になり、プロジェクトのスケジュールも短縮される。本書は、意思決定においてこの情報源にアクセスし、活用するための中立的かつ効率的なアプローチを提供している。本書で紹介する手法は、人工知能主導のプログラミングと、アクセス可能なオープンソースツールを組み合わせた現代的なアプローチを活用している。AI のパワーを活用し、オープンソースソフトウェアを活用することで、この方法論は自動化を強化し、プロセスを最適化し、この分野でのアクセシビリティとコラボレーションを促進することを目的としている。本書の言葉は明快で分かりやすい」

- アンタルヤ・ビリム大学美術・建築学部長、ユーラシア BIM フォーラム主催者。

"私が言えることは、すごいということです！歴史、法学修士号（ ）、グラフィック、そして全体的なわかりやすさ。本の流れも素晴らしい。この本には素晴らしい面がたくさんあり、純粋にゲームチェンジャーです。素晴らしい情報源であり、あなたがこの本に費やした努力と情熱を称賛したい。このような素晴らしい作品を世に送り出したことを祝福したい。まだまだ書き足りないが、一言で言えば、私は信じられないほど感動している！」

- energylab_のデジタル・プラクティス・リード、ナターシャ・プリンスルー。

「新人からベテランまで、建設業界に携わるすべての人にとって、本書は画期的な一冊だ！典型的な埃っぽい読み物ではなく、洞察、戦略、そしてユーモアが満載で、読む者を飽きさせない。古代のデータ記録方法から最先端のデジタル技術まで、建設におけるデータ活用の進化を網羅している。まるでタイムマシンで建設データの進化をたどるようだ。建築家、エンジニア、プロジェクト・マネージャー、データ・アナリストのいずれであっても、この包括的なガイドはプロジェクトへの取り組み方に革命をもたらすだろう。プロセスを最適化し、意思決定を強化し、かつてないほどプロジェクトを管理する準備をしよう！」

- ピエールパオロ・ヴェルガティ、ローマ・サピエンツァ大学講師、フィンテクナ社シニア・コンストラクション・プロジェクト・マネージャー。

「私はこの本を 6 時間足らずで一気に読んだ。本の質は素晴らしく、光沢のある紙、配色、心地よいフォント。建設業界に特化した LLM（ ）を使った作業方法に関する実践的な事例が数多く掲載されているため、自習に費やす数ヶ月、いや数年を節約できるだろう。複雑で高価なソフトウェアを購入することなく、単純なものから複雑なものまで、非常に多様な事例が掲載されている。この本を読めば、建設業界のあらゆる企業の経営者は、自社の事業戦略、デジタル化、発展の見込みを改めて見直すことができるだろう。また、中小企業にとっては、手ごろな価格の無料ツールで効率を上げることができる。」

- ミハイル・コサレフ、TIM-ASG 講師、建設業界のデジタルトランスフォーメーションに関するコンサルタント。

「アルテム・ボイコによる "Data-Driven Construction " は、絶え間なくテクノロジーと情報の可能性が拡大する時代の建設業界に確かな基盤を提供する印象的な作品である。ボイコは、複雑なトピックを理解しやすい形で提示しながら、先見性のあるアイデアも紹介している。本書は、現在の発展に焦点を当てるだけでなく、将来の革新の展望をも提供する、よく考えられた大要である。本書は、データ主導の建設計画と実行に取り組みたいすべての人に強く勧められる。」

- マルクス・アイベルガー、シュトゥットガルト応用科学大学講師、Konstruktionsgruppe Bauen シニア・プロジェクト・マネージャー兼副支社長、BIM クラスター・バーデン・ヴュルテンベルク協会理事。

「データ駆動型建設 " は、大学で教科書として教えられるに値する本であり、発展途上の BIM 分野に貴重な貢献をする本であると言うべきだろう。Data-Driven Construction』には技術用語集があり、コンセプトがよく説明されている。説明するのが非常に難しいトピックが、非常に美しいビジュアル言語によってシンプルに理解できるようになっている。ビジュアルで説明しようとすることは、たとえ簡潔であっても読者に表現されるべきだと思います。ビジュアルによっては理解しやすさ、言い換えれば、ビジュアルを読むには別の情報が必要になる。また、私はアルテム・ボイコの貴重な作品を大学での講義やセミナーで紹介できることをうれしく思っている。」

- アーキ・キューブのオーナーであり、イスタンブール工科大学建築学科とメディポル大学で建設プロジェクト・マネジメントの講師を務めるエディズ・ヤジシオグル博士。

「データは新しい石油である」と言われるように、この 21 世紀の資源から価値を引き出すために、試掘者や採掘者は適切なツールと考え方を持つべきである。建設業界はあまりにも長い間、「3D 情報」ベースのプロセスという滑りやすい坂道を登ってきた。そこでは、プロジェクトの納品は誰かが焼き上げた情報（例えば、円グラフや棒グラフはすでにプロットされている）に基づいて行われるが、その根底にある「データ」（例えば、生のスプレッドシート）はもっと多くのものを提供することができる。もしあなたが建設業を営んでいる（あるいは教育／研究している）のであれば、本書はデータ主導の世界をナビゲートするための最良の（そして今のところ唯一の）リソースとなるだろう」

- ズルフィカール・アダム博士、英国 LSBU の建設における戦略的 IT の准教授。

第 2 版、2025 年 4 月
© 2025｜アルテム・ボイコ｜カールスルーエ

ISBN 9783982625584







アルテム・ボイコ著作権

boikoartem@gmail.com
info@datadrivenconstruction.io

本書のいかなる部分も、著者の書面による許可なく、複写、記録、情報記憶・検索システムなど、電子的または機械的な手段を問わず、いかなる形式または手段によっても複製または転送することを禁じます。本書は無料で配布され、個人的、教育的、研究的な目的で他のユーザーと自由に共有することができます。著者は本文に関するすべての非独占的権利を保持し、明示または黙示の保証を行うものではない。本書で言及されている企業、製品、名称は架空のもの、または例として使用されている場合があります。著者は、与えられた情報の使用によるいかなる結果に対しても責任を負わない。本書に含まれる情報は、完全性や妥当性を保証することなく「現状のまま」提供される。著者は、本書に含まれる情報、コード、またはプログラムの使用から生じる偶発的または結果的な損害について責任を負いません。本書で紹介されているコード・サンプルは、教育目的のみのものです。読者は、自己の責任においてそれらを使用する。著者は、本番環境で使用する前に、すべてのソフトウェア・ソリューションをテストすることを推奨します。本文中に記載されている商標および製品名は、各社の商標、登録商標またはサービスマークであり、各社の所有物です。本書におけるこれらの名称の使用は、それらの所有者との関係や推奨を意味するものではありません。また、第三者の製品またはサービスについて言及することは、推奨を意味するものではありません。例で使用されている会社名や製品名は、その所有者の商標である場合があります。第三者のウェブサイトへのリンクは便宜上提供されているものであり、著者がそれらのサイトで提供されている情報を支持していることを意味するものではない。引用されている統計、引用、研究はすべて執筆時点のものです。データは時間の経過とともに変化する可能性があります

本書はクリエイティブ・コモンズ 表示-非営利-改変禁止 4.0 国際ライセンス（CC BY-NC-ND 4.0）の下で配布されています。非営利目的であれば、著作権を保持し、変更を加えない限り、複製・頒布することができます。

CC BY-NC-ND 4.0

© 2024 アルチョム・ボイコ初版。
© 2025 アルチョム・ボイコ改訂第2版。
無断複写・転載を禁じます。

第 2 版への序文

本書は、専門家コミュニティとの活発な対話の成果である。本書は、様々な専門家のプラットフォームやソーシャルメディア上で行われた、建設業界におけるデータ管理に関する数多くの専門的な議論に基づいている。これらの議論は、記事、出版物、映像資料の基礎となり、専門家コミュニティで幅広い反響を呼んだ。著者のコンテンツは、様々なプラットフォームや言語で毎年数百万ビューを集め、建設のデジタル化分野の専門家を結びつけている。

初版発行から 1 年も経たないうちに、ブラジルやペルーからモーリシャスや日本まで、50 カ国以上の専門家から注文が殺到した。現在お手元にある第 2 版は、専門家からのフィードバック、第 1 版に対する批判、専門家集団での議論に基づいて改訂・増補されたものである。その結果、CAD (BIM)の技術や、効果的な ETL -プロセスの作成に関する新しい章が追加されました。また、実践的な事例やケーススタディの数も大幅に増加した。特に価値があるのは、本書の初版が出版される前と後の両方において、デジタル化と相互運用性（）に関する質問を著者に投げかけてきた、建設業界、コンサルティング会社、大手 IT 企業のリーダーたちからのフィードバックである。その多くは、本書で紹介されているアプローチをすでに適用しているか、近い将来に適用する予定である

あなたが手にしているのは、議論と活発な意見交換を通じて作られた本である。進歩は対話の中で、意見の衝突の中で、そして新しいアプローチへの開放性の中で生まれる。この対話に参加していただき、ありがとうございます。皆様からの建設的なご批判は、今後の改善の基礎となります。文章に誤りがある場合、あるいはアイデアや提案を共有したい場合、どのようなご意見でも歓迎いたします。連絡先は巻末に記載されています

なぜこの本は無料なのか？

本書は、建設業界におけるデータ管理の最新のアプローチを普及させることを目的とした、オープンな教育リソースとして構想された。本書の最初のバージョンは、専門家コミュニティからのコメントや提案を集めるための基礎となり、それによって教材の構成や内容を改善することができた。すべてのコメント、提案、アイデアは慎重に分析され、この改訂版に取り入れられている。本書の目的は、建設専門家が、体系的に、意識的に、そして情報の長期的価値に目を向けながら、データを扱うことの重要性を理解することにある。著者は、10 年以上にわたって建設業のデジタル化に携わってきた経験から、実例や図解、実践的な考察を集めた。この資料の大半は、実際のプロジェクト、エンジニアや開発者との議論、国際的な取り組みやトレーニング・セミナーへの参加から生まれたものである。本書は、蓄積された経験を体系化し、わかりやすい形で共有する試みである。本書のアイデアのさらなる普及を支援し、事例やビジュアル資料を読み、作業するのに便利なフォーマットを手に入れたい方は、印刷版をお求めください。

使用権

本書のすべての資料、図版、断片は、出典をアルテム・ボイコとし、本書のタイトルを「データ駆動型建設」とすることを条件に、いかなる形式、いかなる媒体においても複製、引用、使用することができる。労働の尊重と知識の普及に感謝する。

、幼い頃から建設への深い愛情を植え付けてくれた家族、レジリエンスを教えてくれた故郷の鉱山の町、そして揺るぎないサポートで常に私を鼓舞してくれた測量士の妻に捧げるのは、心からの感謝の気持ちである。

この本は誰のためにあるのか

本書は、現代の建設プロセスの基本を理解したい学生や初心者から、建設におけるデータ管理の最新手法を必要とする専門家まで、建設業界の幅広い読者を対象として、わかりやすい言葉で書かれている。建築家、エンジニア、現場監督、施工管理者、データアナリストのいずれであっても、ユニークなイラストや図表を多用した本書は、ビジネスにおけるデータ活用の方法について貴重な洞察を提供し、プロセスの最適化・自動化、意思決定の改善、最新のツールを用いたさまざまなレベルでの建設プロジェクトの管理を実現する

本書は、データ管理技術を建設プロセスに統合するための理論的基礎と実践的な提言を組み合わせた包括的なガイドである。本書は、業務の最適化、プロセスの自動化、意思決定の改善、最新のデジタルツールを使った効果的なプロジェクト管理など、情報の戦略的活用に焦点を当てている。

本書は、建設業界における情報活用の理論的・実践的側面をカバーする。詳細な事例を通して、タスクのパラメータ化、要件収集、非構造化データおよびマルチフォーマットデータの処理、そして建設会社のための効果的なソリューションへの変換の方法論を探求している。

読者は、要件の形成や基本的なデータモデルの開発から、異種情報ソースの統合、ETL - プロセスの作成、情報パイプラインの構築、機械学習モデルといった、より複雑なプロセスへと順次進んでいく。順を追ったアプローチにより、建設業界におけるビジネスプロセスと意思決定支援システムの組織化と自動化のメカニズムを明確に示すことができる。本書の各章の最後には、習得した知識を実際のプロジェクトに即座に適用できるよう、ステップ・バイ・ステップの手順が記載された実践的な章が設けられている。

あらすじ

本書は、バリューチェーンにおけるデータ変換の概念を中心に構成されている。データの収集と品質保証から、分析処理、そして最新のツールと方法論を用いた価値ある実践的ソリューションの抽出まで。

パート 1：建設におけるデジタル進化 - 粘土板から現代のデジタルシステムまでのデータ管理の歴史的変遷をたどり、モジュラーシステムの出現と産業革命の文脈における情報デジタル化の重要性の高まりを分析する。

第 2 部：建設業界における情報の課題 - データの断片化、「情報のサイロ化」、HiPPO アプローチが意思決定に影響、独自フォーマットの限界などの問題を探り、AI や LLM エコシステムへの移行を示唆する。

第 3 部：建設におけるデータの体系化 - 建設データの類型化を行い、その整理方法、企業システムとの統合を説明し、情報プロセスの標準化のためのコンピテンスセンターの設立について論じる。

パート 4：データ品質保証 - さまざまなソースからのデータ抽出、検証、LLM を使用したモデリングなど、異種情報を高品質な構造化データにするための方法論を明らかにします。

パート 5：コストと時間の計算 - コストと計画計算のデジタル化、CAD（BIM）モデルからの数量取得の自動化、4D-8D モデリング技術、建設プロジェクトの ESG 計算を扱う。

パート 6：CAD と BIM - 設計技術の進化、システムの相互運用性の問題、オープンデータ形式への傾向、設計への人工知能適用の展望を批判的に分析する。

パート 7：データ分析と自動化 - 情報可視化の原則、主要業績評価指標、ETL プロセス、ワークフロー・オーケストレーション・ツール、定型作業を自動化するための言語モデルの適用について見ていきます。

パート 8：データストレージと管理 - データストレージのフォーマット、データウェアハウスとデータレイクのコンセプト、データ管理の原則、ベクターデータベースや DataOps と VectorOps の方法論を含む新しいアプローチについて説明します。

パート 9：ビッグデータと機械学習 - 過去のデータに基づく客観的分析への移行、建設現場におけるモノのインターネット、プロジェクトのコストとスケジュールを予測するための機械学習アルゴリズムの応用に焦点を当てる。

第 10 部：デジタルデータ時代の建設業界 - 因果分析から相関分析への移行、建設業の「ユーバー化」の概念、デジタル変革の戦略などを分析しながら、建設業界の未来について考察する。

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



はじめに

テクノロジーが急速に進化し、タイミングやコスト計算からリスク分析に至るまで、ビジネスのあらゆる側面が機械学習モデル（）によって自動化されつつある世界で、貴社はいつまで競争力を維持できるだろうか？

人類と同じくらい長い間存在してきた建設業界は、革命的な変化の入り口に立っており、従来の建設に対する考え方を完全に変えることが約束されている。経済の他の分野ではすでに、デジタル化は単にルールを変えるだけでなく、新たなデータ処理環境に適応できず、意思決定のスピードを改善できない企業を市場から冷酷に追い出している（図 1）。



図 1 建設業界における意思決定のスピードは、他の業界に比べて人的要因に左右されることが多い。

銀行、小売、ロジスティクス、アグリビジネスは、不正確な情報や主観的な意見がもはや通用しない完全なデジタル化に向けて急速に進んでいる。最新のアルゴリズムは膨大な量のデータを分析し、ローン返済の可能性、最適な配送ルート、リスク予測など、正確な予測を顧客に提供することができる。

建設業は、高給取りの専門家の意見に基づくソリューションから、データ主導のソリューションへの移行が避けられない最後の産業のひとつである。この移行は、新しい技術的能力だけでなく、透明性、正確性、スピー

ドに対する市場や顧客の要求の高まりによっても推進されている。

ロボット化 、プロセスの自動化、オープンデータ とそれに基づく予測 - これらはすべて、もはや可能性ではなく、必然となっている。建設業界の大半の企業（ ）は最近、発注者に対し、プロジェクトの量、コスト、時間、品質管理を計算する責任を負っていたが、現在では、重要な意思決定を行わず、単なる注文の実行者に成り下がる危険性がある（図 2）。

コンピューティング能力、機械学習アルゴリズムの進歩、データへのアクセスの民主化により、 、さまざまなソースからのデータを自動的に組み合わせることが可能になり、建設プロジェクトの検討段階において、より深いプロセス分析、リスク予測、コスト最適化を実現できるようになった。これらの技術は、建設セクター全体の効率を根本的に改善し、コストを削減する可能性を秘めている。

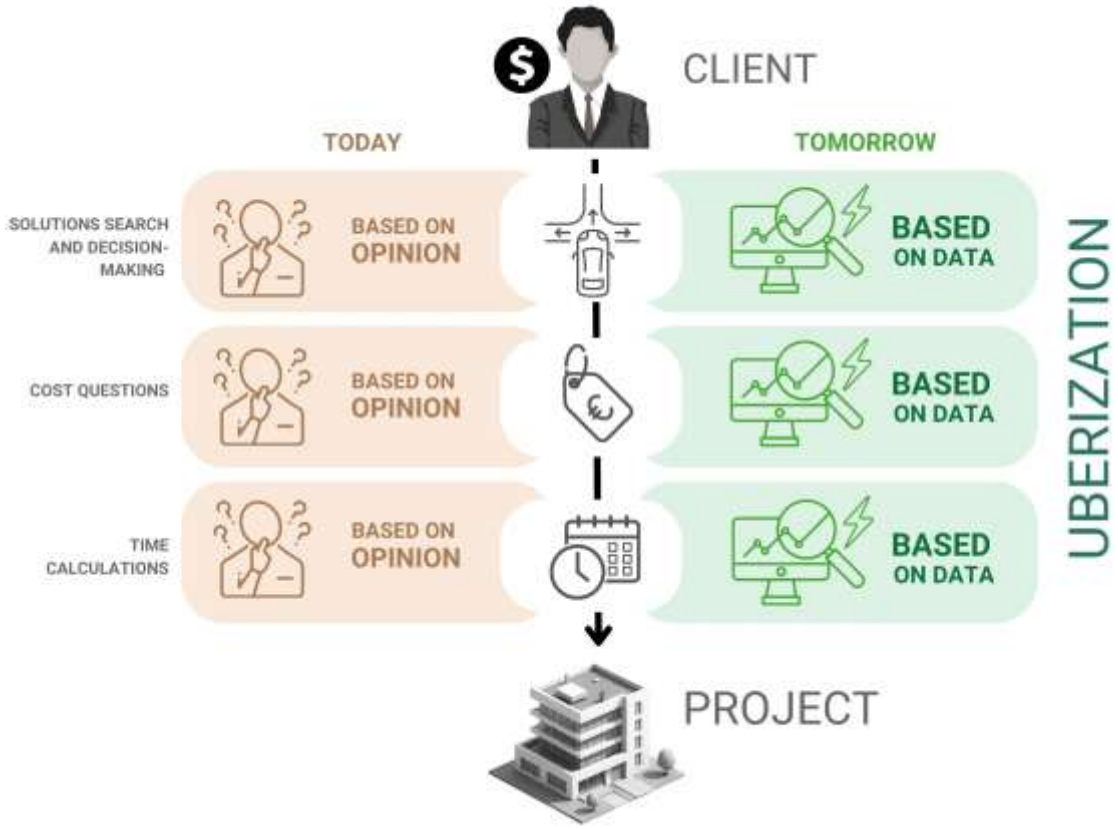


図 2 クライアントは、プロジェクト実現に向けた過度なヒューマン・ファクターには興味がない。

新しいツールやコンセプトにはあらゆる利点があるにもかかわらず、建設業界は新技術の採用において他の経済部門に大きく遅れをとっている。

> IT Metrics Key Data 2017」レポートによると、建設業界の IT 支出は他の 19 業界中最下位である[1]
> 。

データ量の急増とプロセスの複雑化は、企業経営にとって頭痛の種となっている。新しいテクノロジーを利用する際の主な問題は、データが豊富にあるにもかかわらず、断片的で構造化されておらず、異なるシステムやソフトウェア製品間で互換性がないことが多いことだ。このため、建設部門の多くの企業は現在、主にデータ品質の問題 を懸念している。この問題は、効率的で自動化された管理・分析システム を導入することでしか解決できない。

> KPMG®が 2023 年に建設管理者を対象に実施した調査[2]によると、プロジェクト管理情報システム（PMIS ）、高度および基本的なデータ分析 、建築情報モデリング（BIM ）は、プロジェクトの ROI を改善する最大の可能性を秘めている（図 3）。

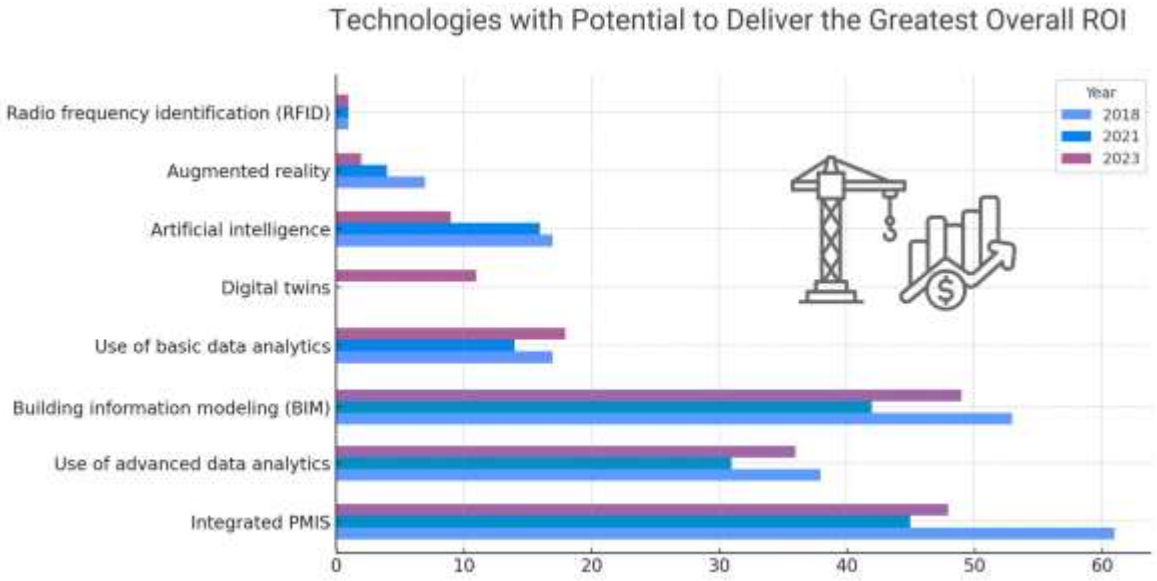


図 3 建設会社の経営者を対象としたアンケート調査：設備投資プロジェクトにおいて最も投資対効果（ROI）の高い技術は？(資料[2]に基づく）。

ビジネスプロセスへのデータ統合に伴う課題の解決策は、高品質の情報を確保し、適切なデータ形式を使用し、データの作成、保存、分析、処理に効果的な方法を適用することである。

データの価値が認識されたことで、さまざまな業界がサイロ化したアプリケーションや複雑な官僚的管理構造からの脱却を迫られている。その代わりに、情報アーキテクチャへの新たなアプローチを構築し、企業を最新のデータ駆動型企業へと変革することに焦点が移りつつある。遅かれ早かれ、建設業界自体もこのステップを踏み、緩やかなデジタル進化から、すべての企業に影響を与える真のデジタル革命へと移行するだろう。

> データ主導のビジネスプロセスへの移行は容易ではない。多くの企業は、経営陣が混沌としたデータセットを効率性と事業成長の向上のためにどのように活用すべきかを必ずしも理解していないため、課題に直面するだろう。

本書は、情報がビジネス・プロセスの効率性と持続可能性を左右する重要な戦略資源となりつつあるデータの世界を掘り下げる。情報の急速な増加に伴い、企業は新たな課題に直面している。デジタルトランスフォーメーションはもはや単なる流行語ではなく、必要不可欠なものとなりつつある。



図 4 データとプロセスは建設の基礎である。

変容を理解するということは、複雑なことを簡単な言葉で説明できるということである。そのため、本書はわかりやすい言葉で書かれており、重要な概念を明確に説明するために特別に作成された著者のイラストが添えられている。これらの図、チャート、ビジュアライゼーションは、知覚の障壁を取り除き、これまでこのようなトピックを複雑すぎると考えていた人でも理解できるようにデザインされている。本書に掲載されているイラスト、図表、画像はすべて著者が作成したもので、本文で説明されている主要概念を視覚化するために特別にデザインされている。

> 一枚の写真は千の言葉に値する[3]。
>
> - フレッド・R・バーナード、イギリスのイラストレーター、1927 年。

理論と実践を結びつけるために、深いプログラミングの知識がなくてもソリューションを開発できる人工知能ツール（特に言語モデル）を使用します。もしあなたが実践的な内容を志向し、データを使った実践的な作業に興味があるのであれば、最初の入門部分は飛ばして、具体的な例やケースの解説が始まる第 2 部分に直行しても構いません。

しかし、AI（人工知能）、機械学習、LLM（大規模言語モデル）ツール全般に過度な期待を寄せてはいけません。質の高い入力データと主題に対する深い理解がなければ、最先端のアルゴリズムであっても信頼できる意味のある結果を提供することはできません。

> マイクロソフトのサティア・ナデラ最高経営責任者（CEO）は、2025 年初頭の人工知能のバブルのリスクについて、現在の誇大広告をドットコムバブルと比較しながら警告している。彼は、適切な正当化なしに AGI（Artificial General Intelligence：人工知能）のマイルストーンに到達したという主張は「無意味な測定基準の操作」であると強調している。ナデラ氏は、AI の真の成功は、バズワードを過度に強調するのではなく、世界の GDP 成長への貢献によって測られるべきだと考えている。

新しいテクノロジーやコンセプトに関するあらゆるバズワードの背後には、データ品質（ ）、ビジネス・プロセスのパラメータ化、実際のタスクへのツールの適応を確実にするための複雑で骨の折れる作業がある。

> データ主導のアプローチは、ただダウンロードしたり購入したりできる製品ではない。構築しなければならない戦略なのだ。既存のプロセスや問題を見直すことから始まり、選択した方向に規律正しく進むことが求められる。

大手ソフトウェア開発会社やアプリケーション・ベンダーは、建設業界の変革のエンジンにはならないだろう。彼らの多くにとって、データ主導のアプローチは、既存のビジネスモデルに対する脅威なのだ。

> 建設業とは異なる）他の業界、例えば自動車業界は、すでに急進的かつ破壊的な変化を遂げており、デジタル・トランスフォーメーションはかなり進んでいる。建設企業は、迅速かつ果断に行動する必要がある。機敏な企業は大きな報酬を得るだろうが、躊躇する企業にとっては深刻なリスクとなる。デジタル写真がこの業界にもたらした激変を考えてみてほしい[5]。
>
> - 世界経済フォーラム報告書『Shaping the Future of Construction』（2016 年

新しいアプローチの機会と利点をタイムリーに認識した企業は、持続可能な競争優位性を獲得し、大手ベンダーのソリューションに依存することなく発展・成長することができるだろう。

これは、来るべき情報のデジタル化の嵐を乗り切るだけでなく、その嵐をコントロールするチャンスでもある。本書では、業界の現状分析だけでなく、建設業界の新時代をリードし、プロフェッショナルとしての経験を高めるために、プロセスやビジネスを見直し、再構築するための具体的な提案も紹介する。

建設におけるデジタルの未来とは、単に新しい技術やプログラムを使うことではなく、データの取り扱いやビジネスモデルを根本的に見直すことである。

あなたの会社はこの戦略的変化に対応できているだろうか？

目次

はじめに 1

目次 I

II 部　粘土板からデジタル革命へ：建設における情報の進化 2

章 2.1.　建設業界におけるデータ活用の進化 3

建設業界におけるデータ時代の誕生 3

粘土とパピルスからデジタル技術へ 4

データ主導の経験のためのツールとしてのプロセス 5

建設プロセス情報のデジタル化 7

章 2.2.　近代建設における技術と管理システム 11

デジタル革命とモジュラー型 MRP/ERP の出現 - システム 11

データ管理システム：データマイニングからビジネス課題まで 13

企業の菌糸体：データはどのようにビジネス・プロセスにつながるか 18

章 2.3.　デジタル革命とデータの爆発 21

進化の波としてのデータブームの始まり 21

現代企業で生成されるデータ量 23

データ保存のコスト：経済的側面 25

データ蓄積の最前線：質量から意味へ 27

次のステップ：データ理論から実践的変革へ 30

III 部　建設業はいかにしてデータの混乱に溺れているか 32

章 3.1.　データの断片化とサイロ化 33

ツールが多ければ多いほど、ビジネスの効率は上がるのか？ 33

データのサイロ化と企業業績への影響 35

重複、データ品質の欠如 不統一の結果である。 40

HiPPO、あるいは意思決定における意見の危険性 42

ビジネス・プロセスの複雑さとダイナミズムの継続的増加 45

第四次産業革命 （インダストリー4.0）と建設業における第五次産業革命（インダストリー5.0）。 48

章 3.2. 混沌を秩序に変え、複雑さを軽減する 52

生産性向上の障害となる冗長なコードと閉じたシステム 52

サイロから単一のデータウェアハウスへ 54

統合ストレージシステムが AI への移行を可能に エージェント 57

データ収集から意思決定まで：自動化への道 60

次のステップ：混乱を管理可能なシステムに変える 62

IV 部 建設ビジネスプロセスにおけるデータフレームワーク **64**

章 4.1. 建築におけるデータ型 65

建設業界で最も重要なデータタイプ 65

構造化データ 70

リレーショナル・データベース RDBMS および SQL クエリ言語 72

SQL -データベースにおけるクエリと新しいトレンド 74

非構造化データ 77

テキストデータ：非構造化カオスと構造化カオスの間 y 78

半構造化データおよび疎構造化データ 80

幾何学的データ とその応用 81

CAD データ：設計からデータ保存まで 83

BIM（BOM）コンセプトの出現 および工程における CAD の使用 87

章 4.2. データのハーモナイゼーションと構造化 95

建設業界におけるデータ充填システム 95

データ変換：現代のビジネス分析の重要な基盤 98

データモデル：データ内の関係と要素間の関係 102

独自フォーマットとデジタルプロセスへの影響 106

オープン・フォーマットがデジタル化へのアプローチを変える 110

パラダイムシフト：オープンソース ソフトウェアベンダー優位の時代の終焉として 113

構造化されたオープンデータ：デジタルトランスフォーメーションの基盤 116

章 4.3. LLM データ処理とビジネスプロセスにおける役割 119

LLM チャットルーム: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok データプロセスを自動化する。 119

大規模 LLM 言語モデル: その仕組み 120

ローカル LLM の利用 機密性の高い企業データのため 123

社内における AI の完全制御と独自の LLM の展開方法 126

RAG: インテリジェント LLM -企業データにアクセスできるアシスタント 129

章 4.4. IDE WITH LLM サポート および将来のプログラム変更 131

IDE の選択：LLM 実験からビジネス・ソリューションまで 131

LLM をサポートする IDE および将来のプログラミングの変更 134

Python Pandas: データ操作に欠かせないツール 135

DataFrame: 汎用表データフォーマット 139

次のステップ：持続可能なデータフレームワークの構築 143

V 部 データ品質：整理、構造化、モデリング **145**

章 5.1. データの構造化 146

文書、PDF 、画像、テキストを構造化されたフォーマットに変換する方法を学ぶ。 146

PDF -文書を表に変換する例 148

JPEG、PNG 画像を構造化形式に変換する 152

テキストデータを構造化フォームに変換する 155

CAD データ (BIM) の構造化形式への変換 158

CAD ソリューション・ベンダーは構造化データに向かっている 165
章 5.2. 分類と統合：建設データの共通言語 168
意思決定のスピードはデータの質に左右される 168
データの標準化と統合 169
デジタル相互運用性は要件から始まる 172
建設の共通言語：デジタル変革における分類器の役割 175
Masterformat, OmniClass, Uniclass and CoClass: 分類システムの進化 178
章 5.3. データモデリングとセンター・オブ・エクセレンス 184
データモデリング：概念モデル、論理モデル、物理モデル 184
建設における実践的データモデリング 188
LLM を使ったデータベースの作成 191
データモデリング・センター・オブ・エクセレンス（CoE） 193
章 5.4. 要件の体系化と情報の検証 197
要件収集と分析： コミュニケーションを構造化データに変換する。 197
プロセス・フローチャートと概念フレームワークの有効性 202
構造化要件 および RegEx 正規表現 204
データ収集 検証プロセス用 210
データの検証と検証結果 213
検証結果の視覚化 218
データ品質チェックの比較 人間の生活ニーズとの比較 220
次のステップ：データを正確な計算と計画に変える 222
VI 部 コストと時間の計算：建設プロセスにデータを組み込む **224**
章 6.1. 建設プロジェクトのコスト計算と見積もり 225
建設の基本：数量、コスト、時間の見積もり 225
プロジェクトの推定コストの計算方法 227

Resource-based method Costing and Estimating 建設業における原価計算と見積 228

建設資源のデータベース：建設資材と工事のカタログ 228

計算の取りまとめ およびリソース・ベースに基づく工事の原価計算 230

プロジェクトの最終的なコスト計算：見積もりから予算まで 235

章 6.2. 数量テイクオフ、見積もりとスケジュールの自動作成 239

3D から 4D、5D への移行：体積パラメーターと定量パラメーターの使用 239

5D 属性 および CAD からの属性ボリュームの取得 240

QTO Quantity Take-Off: 属性によるプロジェクトデータのグループ化 243

LLM および構造化データ を使用した QTO 自動化 250

QTO Excel スプレッドシートからグループルールを使用してプロジェクト全体の計算を行う。 255

章 6.3. 4D, 6D -8D および二酸化炭素排出量の計算 CO_2 262

4D モデル：建設見積もりに時間を組み込む 262

原価計算データに基づく建設スケジュールとその自動化 263

拡張属性レイヤー 6D -8D: エネルギー効率から安全保証まで 266

CO_2 を推定し、建設プロジェクトからの二酸化炭素排出量を計算する。 269

章 6.4. CONSTRUCTION ERP AND PMIS SYSTEMS 274

建設 ERP - 計算と見積もりを例にしたシステム 274

PMIS: ERP と建設現場の中間的存在。 280

投機、利益、偏狭さ、透明性の欠如 ERP と PMIS における透明性の欠如 281

クローズド ERP 時代の終焉 /PMIS: 建設業界は新たなアプローチを必要としている 284

次のステップ：プロジェクト・データの効率的活用 287

VII 部 CAD AND BIM: マーケティング、現実、そして建設における設計データの未来 290

章 7.1. EMBERGENCE OF BIM- 建設業界におけるコンセプト 291

CAD- ベンダーのマーケティングコンセプトとしての BIM とオープン BIM の出現の歴史 291

BIM の現実: 統合データベースの代わりに - 閉鎖されたモジュールシステム 295

建設業界におけるオープンフォーマット IFC の出現 297

IFC ジオメトリックコアによるフォーマットの問題 300

、オントロジー 304

建設業界でセマンティック技術が期待に応えられない理由 306

章 7.2. クローズド・プロジェクトのフォーマットと相互運用性の問題 310

閉ざされたデータと生産性の低下：CAD 業界の行き詰まり (BIM) 310

CAD システム間の相互運用性神話 312

米ドルへ と粒度データ m 317

章 7.3. 建築における幾何学：線から立方メートルまで 322

線がお金に変わるとき、あるいはなぜ建築家にジオメトリーが必要なのか？ 322

線から体積へ：面積と体積がデータになるまで 323

MESH、USD、ポリゴンへの移行：ジオメトリにテッセレーション を使う 325

LOD, LOI, LOMD - CAD (BIM) における ディテールのユニークな分類。 327

新しい CAD 標準（BIM ） - AIA, BEP, IDS, LOD, COBie 329

章 7.4. CAD 操作のための LLM の設計パラメータ化と使用 335

CAD データの一意性という幻想 (BIM): 分析とオープンフォーマットへの道 335

パラメータによるデザイン：CAD の未来 と BIM 338

デザイン CAD における LLM の出現 データ処理プロセス 342

LLM と Pandas を用いた DWG - ファイルの自動解析 345

次のステップ：クローズド・フォーマットからオープン・データへ 352

VIII 部 データ主導の意思決定、分析、自動化、機械学習 **354**

章 8.1. データ分析とデータに基づく意思決定 355

意思決定における資源としてのデータ 355

データの視覚化：理解と意思決定の鍵 359

KPI および ROI 361

ダッシュボードとダッシュボード：効果的なマネジメントのための指標の可視化 364
データ分析 と質問の技術 365
章 8.2. 手作業に頼らないデータフロー：なぜエテルが必要なのか 368
ETL 自動化: コスト削減とデータ処理のスピードアップ 368
ETL 抽出: データ収集 373
ETL Transform: バリデーションと変換ルールの適用 376
ETL Load: の結果を図表で可視化する。 379
ETL ロード: PDF ドキュメントの自動作成 384
ETL ロード: FPDF から文書を自動生成 386
ETL ロード: レポーティングと他システムへのロード 391
ETL with LLM: PDF からのデータの可視化 -documents 393
章 8.3. 自動 ETL コンベヤ (パイプライン) 399
パイプライン: 自動 ETL コンベア データ 399
パイプライン -ETL LLM によるデータ検証プロセス 403
パイプライン -ETL: CAD (BIM) におけるプロジェクト要素のデータと情報の検証。 406
章 8.4. ETL とワークフローのオーケストレーション：実践的ソリューション 413
DAG および Apache Airflow: ワークフローの自動化とオーケストレーション 413
Apache Airflow: ETL 自動化の実践的アプリケーション 414
ルーティングとデータ変換のための Apache NiFi 419
n8n Low-Code, No-Code プロセス・オーケストレーション 420
次のステップ：手作業から分析ベースのソリューションへ 424
IX 部 建設におけるデータの保存と管理 426
章 9.1. データ・インフラ：保存形式からデジタル・レポジトリまで 427
データアトムズ：効果的な情報管理の基盤 427
情報保管：ファイルまたはデータ 428

ビッグデータ・ストレージ：一般的なフォーマットとその有効性の分析 430

Apache Parquet でストレージを最適化 433

DWH: データウェアハウス データウェアハウス 436

データレイク - ETL の進化 から ELT: 従来のクリーニングから柔軟な処理へ 438

データレイクハウス・アーキテクチャ: ウェアハウスとデータレイクの相乗効果 440

CDE 、PMIS 、ERP または DWH およびデータレイク 443

章 9.2. データウェアハウスの管理と混乱防止 447

ベクターデータベースとバウンディングボックス 447

データガバナンス）、データミニマリズム）、データスワンプ） 451

DataOps と VectorOps: 新しいデータ標準 454

次のステップ：カオス・ストレージから構造化ストレージへ 456

X 部 ビッグデータ、機械学習、予測 458

章 10.1. ビッグデータとその分析 459

建設におけるビッグデータ：直感から予測可能性へ 459

ビッグデータの実現可能性を問う：相関、統計、データサンプリング 460

ビッグデータ：サンフランシスコの 100 万件の建築許可データセットからのデータ分析 464

CAD データに基づくビッグデータの例 (BIM) 470

IoT モノのインターネットとスマートコントラクト 475

章 10.2. 機械学習と予測 479

機械学習 と人工知能は、我々の建築方法を変えるだろう。 479

主観的評価から統計的予測へ 482

タイタニック・データセット：アナリティクスの世界でハロー・ワールド データとビッグデータ 483

機械学習の実例: タイタニックの乗客からプロジェクト管理まで 490

過去のデータに基づく予測・予想 495

機械学習の主要概念 498

章 10.3. 機械学習によるコストと納期の予測 501

機械学習（）を使ってプロジェクトのコストとスケジュールを求める例 501

線形回帰を用いたプロジェクト・コストと時間の予測 503

K-nearest neighbour algorithm (k-NN) を用いたプロジェクトコストと時間の予測 506

次のステップ：貯蔵から分析・予測へ 511

XI 部 デジタル時代の建設業界。機会と課題 **514**

章 11.1. 生存戦略：競争優位の構築 515

計算ではなく相関：建設アナリティクスの未来 515

建設におけるデータ主導のアプローチ：インフラの新たなレベル 519

次世代デジタルオフィス：AI がワークスペースをどう変えるか 521

オープンデータとウーバライゼーションは既存の建設ビジネスにとって脅威である 524

ユーバー化の未解決問題 変革のために時間を使う最後のチャンスとして 527

章 11.2. データ主導のアプローチを導入するための実践的ガイド 532

理論から実践へ：建設業におけるデジタルトランスフォーメーションのロードマップ 532

デジタル基盤の構築：デジタル成熟への 1-5 ステップ 534

データの可能性を解き放つ：デジタル成熟への 5〜10 ステップ 540

変革のロードマップ：混沌からデータ駆動型へ 企業 548

インダストリー5.0 の構築：隠れることができなくなったときにお金を稼ぐ方法 551

結論 **553**

著者について **556**

さらなる関係 **558**

翻訳解説 **558**

その他のスキルとコンセプト **560**

用語集 **564**

参考文献とオンライン資料のリスト **572**

サブジェクト・インデックス 593

印刷版で最大限の利便性を

データ駆動型建設』の無料デジタル版をご覧いただけます。より便利な作業や資料への迅速なアクセスのためには、印刷版のご利用をお勧めします：



■いつでも手元に：印刷された本書は信頼できる作業ツールとなり、どのような作業状況でも適切なビジュアライゼーションやダイアグラムを素早く見つけて使用することができます。

■高品質のイラスト：印刷版のすべての画像とグラフィックは最高の品質で表示されます。

■情報への素早いアクセス：簡単なナビゲーション、メモやブックマークを作成する機能、どこでも本で作業する機能。

印刷された完全版を購入することで、快適で効率的な情報作業のための便利なツールを手に入れることができます。日々の作業でビジュアル資料を素早く利用し、必要なスキームを素早く見つけ、メモを取ることができます。さらに、ご購入はオープンな知識の普及をサポートします。

印刷版のご注文は下記まで： datadrivenconstruction.io/books



I 部

粘土板からデジタル革命へ：建設における情報の進化

本書の第 1 部では、物理メディア上の原始的な記録から現代のデジタル・エコシステムに至るまで、建設業界におけるデータ管理の歴史的変遷を検証している。情報管理技術の変遷、ERP システムの出現、データの断片化がビジネスプロセスの効率に与える影響について分析している。特に、情報のデジタル化のプロセスと、主観的な専門家の判断に代わる客観的分析の重要性の高まりに注目している。現代の建設業界が直面している情報量の急激な増大と、それに伴う企業システムの課題についても詳細に検討する。第 4 次、第 5 次産業革命の文脈における建設産業の位置づけを探るとともに、持続可能な競争優位性を生み出すための人工知能（）とデータ中心アプローチの活用の可能性を探る。

章 1.1.

建設業界におけるデータ活用の進化

建設業界におけるデータ時代の誕生

約 10,000 年前の新石器時代、人類はその発展において革命的な転換を遂げ、遊牧生活を捨てて定住生活に移行した。この瞬間から、建設業の歴史が始まる。

文明が発展するにつれ、建築はますます複雑になり、最初の儀式用神殿や公共建築物が生まれた。建築設計の複雑化に伴い、古代の技術者や管理者は最初の記録や計算を作成する必要があった。粘土板やパピルスに記された最初の記録には、必要な建築資材の量やその費用、作業に対する報酬の計算の背後にある論理の記述がしばしば含まれていた[7]。こうして、現代のデジタル技術が登場するはるか以前から、建設におけるデータ利用の時代が始まった（図 1.1-1）。

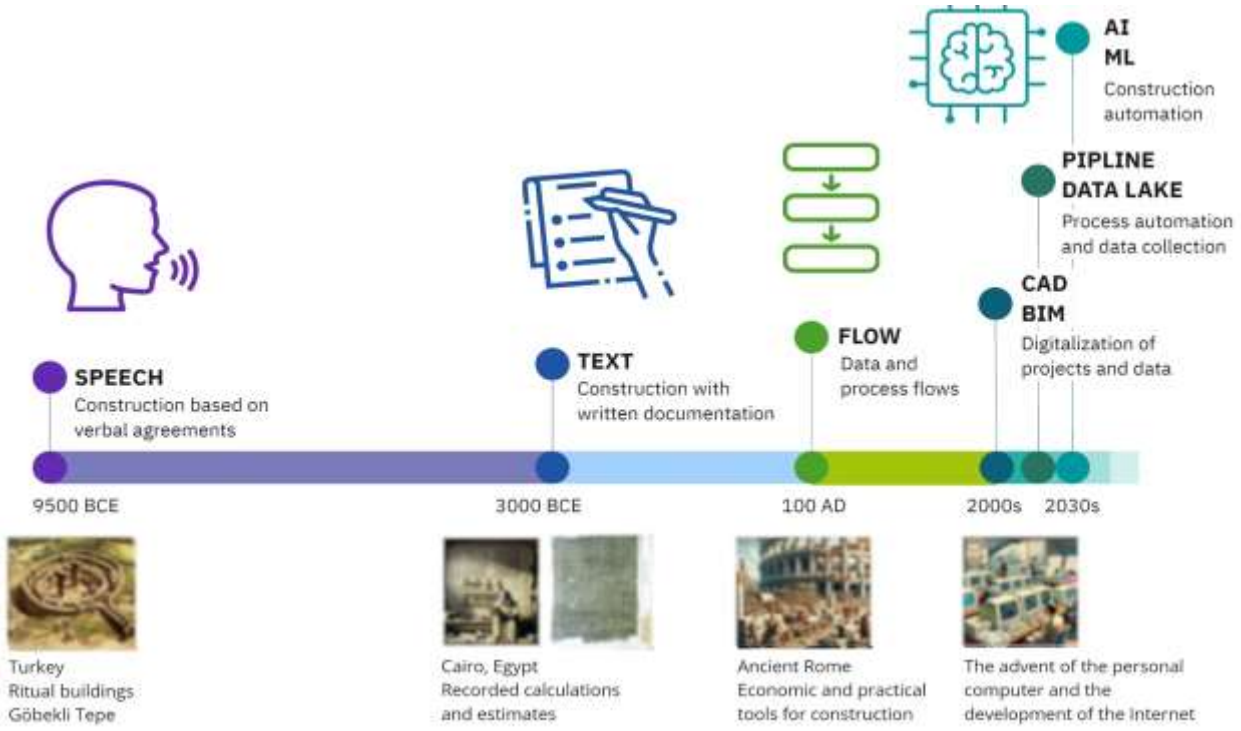


図 1.1-1 建設における情報技術開発の年表：言語情報から人工知能へ.

粘土とパピルスからデジタル技術へ

建設業における最初の文書による証拠は、紀元前 3000〜4000 年頃のピラミッド建造の時代にさかのぼる[7]。それ以来、文字による記録の保存は、建設業界における進歩を促進し、それに伴い、知識の蓄積と体系化を可能にし、その後 1 万年以上にわたって、建設方法と建築における重要な革新につながった。

数千年前の粘土板やパピルス（図 1.1-2）、1980 年代の「A0」紙など、建設における最初の物理的媒体を使用してデータを記録することは、本来、これらの情報を新しいプロジェクトに適用することを意図したものではなかった。このような記録の主な目的は、必要な材料や工事費用の計算を含む、プロジェクトの現状を詳細に記録することであった。同様に、今日の世界では、デジタルの設計データやモデルが利用可能であっても、将来のプロジェクトへの適用が必ずしも保証されるわけではなく、主に必要な材料や工事費の現在の計算のための情報として機能することが多い。

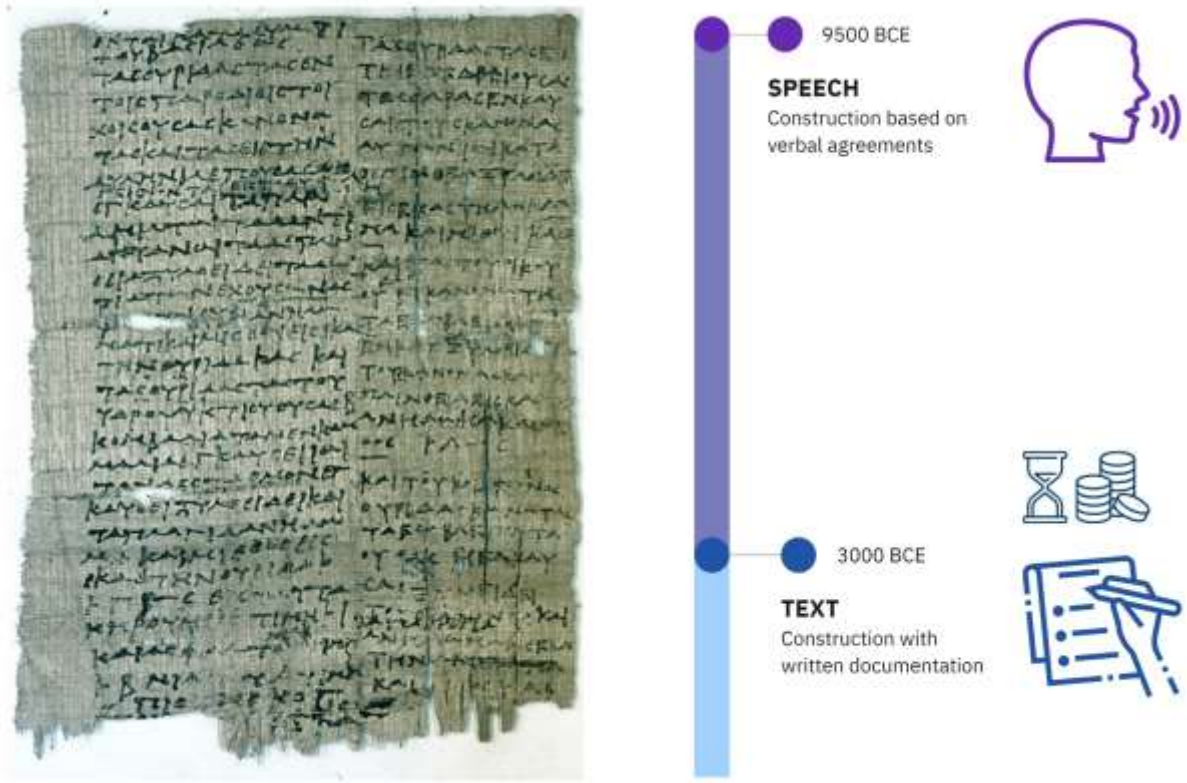


図 1.1-2 紀元前 3 世紀のパピルスで、王宮のさまざまな種類の窓をエンカウスティック技法で塗装するための費用を記したもの。

> 建設プロジェクト管理において、人類が口頭での会話から文書に移行するのに約 5,000 年かかり、計画と管理の主要リソースとして紙からデジタルデータに移行するのにも同じ時間がかかった。

貿易や貨幣関係の発展が、紛争を解決するための文字や最初の弁護士の出現を促したように、建設における資材コストや作業範囲の最初の記録が、建設業界における最初のマネージャーの出現につながった。

今日、データはより重要な役割を担っている。意思決定を記録するだけでなく、未来を予測し、モデル化するためのツールにもなっている。蓄積された経験を、構造化された検証可能なデータに基づく意思決定システムに変える。

データ主導の経験のためのツールとしてのプロセス

あらゆるプロセスの核心は、過去の経験を将来計画のためのツールに変換することである。現代的な意味での経験とは、情報に基づいた予測を行うために分析できる、構造化されたデータの集合である。

> フォーキャスティングの基礎となるのは過去のデータであり、それは実行された仕事の結果を明確に示し、その結果に影響を与える要因についての洞察を提供するものだからである。

一枚岩の建設から具体的な例を挙げてみよう。通常、工事の時期を計画する際には、コンクリートの量、構造の複雑さ、天候が考慮される。ある現場監督や会社の過去 3 年間（2023〜2025 年）の過去のデータによると、雨天時に 200m^2の一枚岩構造物の打設に 4.5〜6 日かかったとしよう（図 1.1-3）。将来のプロジェクトで同様の作業を計画する際、リードタイムを予測し、リソースをコスト計算する基礎となるのは、こうした統計の蓄積である。この過去のデータに基づき、現場監督や積算担当者は、経験に基づいて、2026 年に同じような条件で将来同様の工事を完了するのに必要な時間を予測することができる。

この時間の見積もりの場合、分析プロセスは、ばらばらのデータを構造化された経験に変換し、正確な計画ツールに変換するメカニズムとして機能する。データとプロセスは 1 つのエコシステムであり、一方が他方なしに存在することはできない。

> 数えられるものを数え、測れるものを測り、測れないものを測れるようにする[8]。
>
> - ガリレオ・ガリレイ

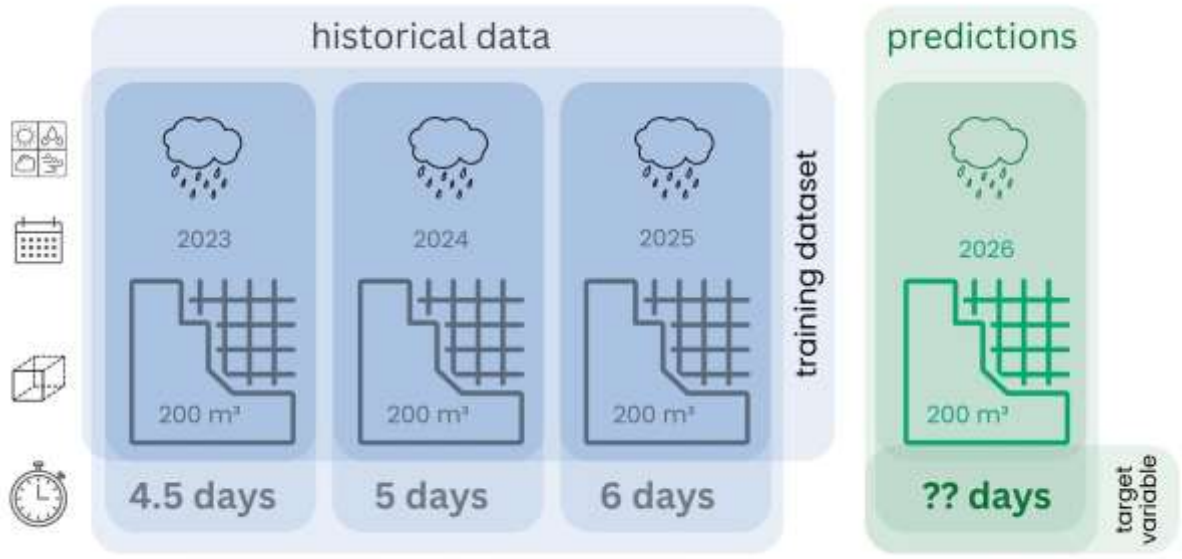


図 1.1-3 過去のデータは、将来のある値を予測するための訓練データセットとして機能する。

今日のビジネスシーンにおいて、データ分析（）は、効果的なプロジェクト管理、プロセスの最適化、戦略的意思決定の重要な要素となりつつある。建設業界は、アナリティクスの 4 つの主要レベル（）を徐々にマスターしつつあり、それぞれが特定の質問に答え、独自のメリットを提供している（図 1.1-4）：

- 記述的アナリティクス - 「何が起こったのか」という問いに答え、過去の出来事や結果に関する履歴データとレポートを提供する。過去 3 年間（2023～2025 年）、雨天時に 200m^{2}のモノリス構造物を打設するのに 4.5～6 日かかった。
- 診断分析 - 問題の原因を特定することで、「なぜこのようなことが起こったのか」という疑問に答える。分析によると、雨天のためにモノリス構造の打設時間が長くなり、コンクリートの硬化プロセスが遅くなった。
- 予測分析 - 「何が起こるか」という問いに答えることで、起こりうるリスクとリードタイムを予測するフォワードルッキングな分析。過去のデータに基づき、2026 年の雨天時に同様の 200m^{2}のモノリス構造物を打設する場合、既知のすべての要因と傾向を考慮すると、約 5.5 日かかると予測される。
- プリスクリプティブ・アナリティクス - 自動化された推奨を提供し、「何をすべきか」という問いに答え、企業が最適な行動を選択できるようにする：例えば、作業を最適化するために、高湿度条件下でコンクリートの硬化を促進する特別な添加剤を使用する、降水確率が最も低い期間に打設を計画する、構造物に仮設シェルターを手配し、悪天候でも作業時間を 4～4.5 日に短縮する、などが推奨される。

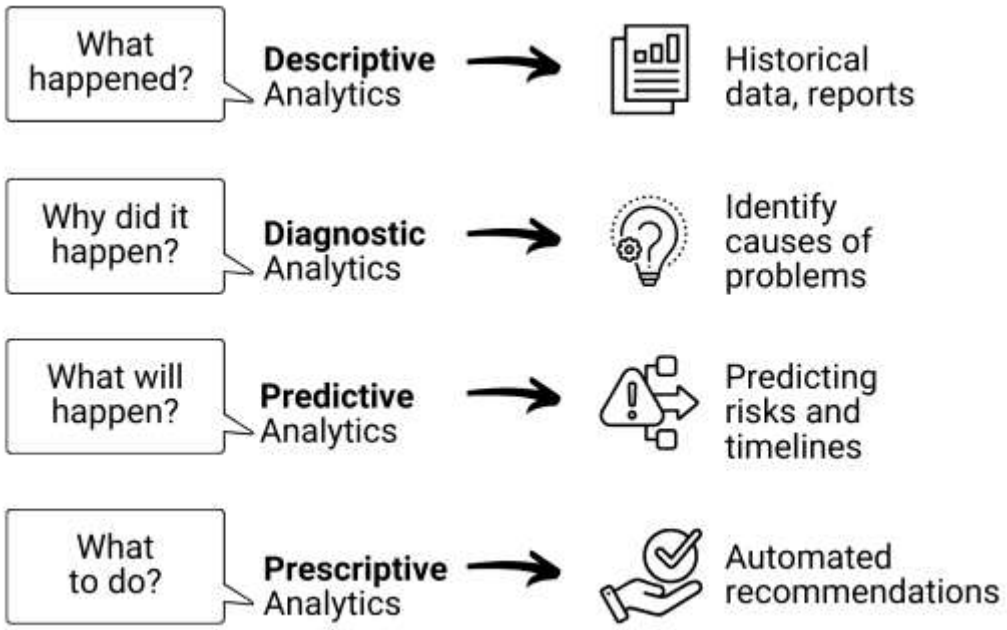


図 1.1-4 アナリティクスの主な種類: 過去の記述から自動意思決定まで。

本格的なデジタルトランスフォーメーションは、システム分析とデータ主導型経営への移行を意味するが、そのためには、単なるアウトソーシングではなく、有能な社内チームの形成が必要である。このようなチームの主要メンバーは、プロダクト・マネジャー、データ・エンジニア、アナリスト（ ）、開発者であるべきで、彼らはビジネス部門と緊密に協力しながら仕事をする（図 4.3-9）。インテリジェントな分析的質問を行い、ビジネス上の意思決定タスクを効果的にパラメータ化するためには、このようなコラボレーションが必要である。情報化社会では、データは単なる補助ツールではなく、予測と最適化の基礎となる。

建設業界では、デジタルトランスフォーメーションによって、施設の設計、管理、運営の方法が根本的に変わりつつある。このプロセスは情報のデジタル化と呼ばれ、建設プロセスのあらゆる側面がデジタル化され、分析に適したデジタル形式となる。

建設プロセス情報のデジタル化

何千年もの間、建築に記録される情報量はほとんど変わっていなかったが、ここ数十年で急速に増加した（図 1.1-5）。

> PwC の調査® 「マネージド・データ」によると。What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9]によると、世界の全データの 90%は過去 2 年間に作成されたものである（2015 年時点）。しかし、ほとんどの企業は、サイロ化されたシステムにとどまっているか、実際に分析されることなく単にアーカイブされているため、これらのデータを十分に活用していない。

データ量の増加は近年加速するばかりで、2015 年の 15 ゼタバイトから 2025 年には 181 ゼタバイトへと倍増している[10]。建設会社や設計会社のサーバーは毎日、プロジェクト文書、作業スケジュール、計算・算出、財務報告書で埋め尽くされている。2D/3D -図面には DWG 、DXF 、DGN フォーマットが、3D モデルには RVT 、NWC 、PLN 、IFC™ フォーマットが使用されている。テキスト文書、表、プレゼンテーションは、DOC、XLSX 、PPT で保存されます。建設現場からのビデオや画像（MPG と JPEG）に加え、IoT コンポーネント、RFID® タグ（識別と追跡）、BMS ビル管理システム（監視と制御）からのリアルタイム・データも使用します

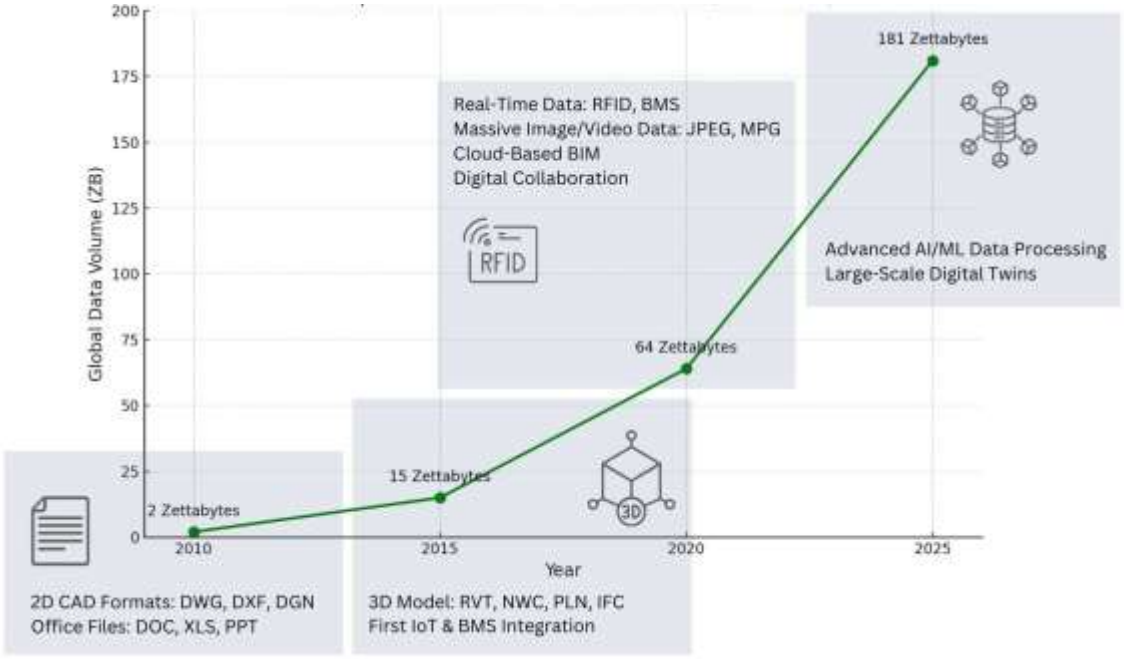


図 1.1-5 Parabolic data growth 2010-2025 （[10] に基づく）。

情報の急速な増加に伴い、建設業界は、データを収集・保存するだけでなく、その検証、妥当性確認、測定可能性、分析処理を確実に行う必要性に直面している。今日、建設業界は、情報のデジタル化、すなわち建設活動のあらゆる側面を、分析、解釈、自動化に適したデジタル形式に体系的に変換する段階を積極的に進めている。

> 情報のデジタル化とは、建設プロジェクトのあらゆる主体や要素、そして建設プロセスそのものに関する情報を、以前はまったく情報だと考えていなかったものも含めて、データ形式に変換し、情報を定量化し、分析しやすくすることである。

建設という文脈では、分析モデルを生成するために、建設現場の機械や人の動きから、建設現場の天候や気候条件、現在の資材価格、中央銀行の金利に至るまで、プロジェクトのあらゆる要素やあらゆるプロセスに関する情報を取得し、デジタル化することを意味する。

> あなたが話していることを測定し、それを数字で表すことができれば、あなたはそのテーマについて何かを知っていることになる。しかし、それを定量的に表すことができなければ、その知識は極めて限定的で満足のいくものではない。それは出発点かもしれないが、真の科学的知識のレベルではない。[11].
>
> - W.トムソン（ケルビン卿）、1824-1907、イギリスの科学者

情報のデジタル化は、工数や実際の材料費といった基本的な指標のみが記録されていた従来の情報収集のアプローチをはるかに超えている。今日では、事実上あらゆる事象を、高度な分析ツールや機械学習技術を用いた詳細な分析に適したデータの流れに変換することができる。建設業界は、紙の図面、エクセルのスプレッドシート、口頭による指示から、プロジェクトのあらゆる要素がデータソースとなるデジタルシステム（図 1.2-4）へと、根本的な転換を遂げた。エンジニアから現場の建設作業員に至るまで、従業員でさえもデジタル変数とデータセットの集合体と見なされるようになった。

> KPMG の「Familiar Challenges - New Approaches: Global Construction Survey 2023」によると、デジタルツイン、AI（人工知能）、ビッグデータ 、プロジェクトの収益性を向上させる重要な推進力として浮上している[2]。

最新のテクノロジーは、情報収集を単純化し、ほぼ自動化するだけでなく、データ保存のコストを根本的に削減する。その結果、企業は選択的なアプローチから脱却し、後の分析のために一連の情報全体を保存することを好むようになっている（図 2.1-5）。

> 情報のデジタル化とデジタル化によって、これまで未開拓だった情報の隠れた価値を掘り起こすことが可能になる。適切に整理されれば、データは再利用され、再解釈され、新しいサービスやソリューションに統合される。

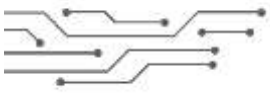
将来的には、情報のデジタル化によって、文書管理の完全自動化、自己管理型建設プロセスの導入、新しい職業（ 建設データアナリスト、AI プロジェクト管理専門家、デジタルエンジニア）の出現が予想される。建設プロジェクトはダイナミックな情報源となり、意思決定は直感や主観的な経験ではなく、信頼性と再現性のあるデジタルな事実に基づいて行われるようになるだろう

> 情報は 21 世紀の石油であり、アナリティクス（）は内燃機関である[12]。
>
> - ピーター・ソンダーガード、ガートナー上級副社長®

> IoT Analytics 2024 [13]によると、データ管理とアナリティクスに対する世界の支出は、2023 年の 1,855 億ドルから 2030 年には 5,133 億ドルへと、年平均成長率 16%で飛躍的に伸びると予想されている。アナリティクス（）は急成長しているが、ストレージの成長は鈍化している。アナリティクスは、データ管理エコシステムの中で最も急成長する。2023 年の 606 億ドルから 2030 年には 2279 億ドルに成長すると予測され、年平均成長率は 27%である。

情報のデジタル化が加速し、情報量が急増する中、建設プロジェクトや企業経営は、多様で、しばしば異質なデータを体系的に保存、分析、処理する必要に迫られている。この課題に対応するため、1990 年代半ば以降、業界では、スプレッドシートや設計計算書から図面や契約書に至るまで、ドキュメントの電子的な作成、保存、管理への大規模な移行が始まった。

署名、物理的な保管、定期的な改訂、キャビネットへの保管を必要とする従来の紙文書は、構造化された方法でデータを保存するデジタル・システム、つまり特殊なアプリケーションのデータベースに徐々に取って代わられつつある。



章 1.2.

近代建設における技術と管理システム

デジタル革命とモジュラー型 MRP/ERP の出現 - システム

現代のデジタル・データの保存と処理の時代は、1950 年代の磁気テープの登場によって始まった。次のブレークスルーはディスク・ドライブの出現で、建設業界におけるデータ管理のアプローチを根本的に変えた。

データウェアハウジングの発展に伴い、 、多くの企業がソリューション市場に参入し、データの作成、保存、処理、ルーチンタスクの自動化のためのモジュラーソフトウェアの開発に着手した

> 情報やツールの急激な増加により、個々のファイルに対応するのではなく、プロセスやプロジェクト全体のデータの流れを管理・制御する、統合されたモジュール式のソリューションが必要とされています。

最初の包括的なプラットフォーム・ツールは、ドキュメントを保存するだけでなく、プロセスにおけるすべての変更要求と操作を文書化する必要があった。誰がそれを開始したのか、要求の範囲は何だったのか、最終的に値や属性として記録されたものは何だったのか。これらの目的のためには、正確な計算と決定を追跡できるシステムが必要であった（図 1.2-1）。そのようなプラットフォームが、1990 年代初頭から普及した最初の MRP （Material Requirements Planning）と ERP （Enterprise Resource Planning）システムである[14]

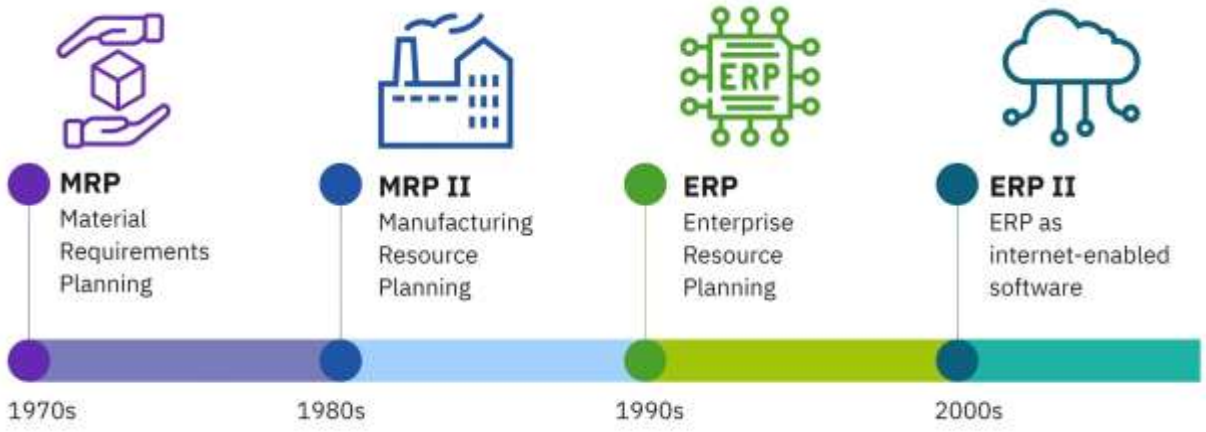


図 1.2-1 データ・ストレージ技術の進歩により、1980 年代に ERP -システムが登場した。

最初の MRP - そして ERP - システムは、ビジネスプロセスと建設プロジェクト管理のデジタル化時代の基礎を築いた。もともと主要なビジネス・プロセスを自動化するために設計されたモジュラー・システムは、やがて、より柔軟で適応性の高い追加的なソフトウェア・ソリューションと統合された。

これらの追加ソリューションは、データ処理とプロジェクト・コンテンツ管理のために設計され（図 1.2-2）、大規模システム の特定のモジュールを置き換えるか、あるいはシステム全体の機能を拡張して効果的に補完するものであった。

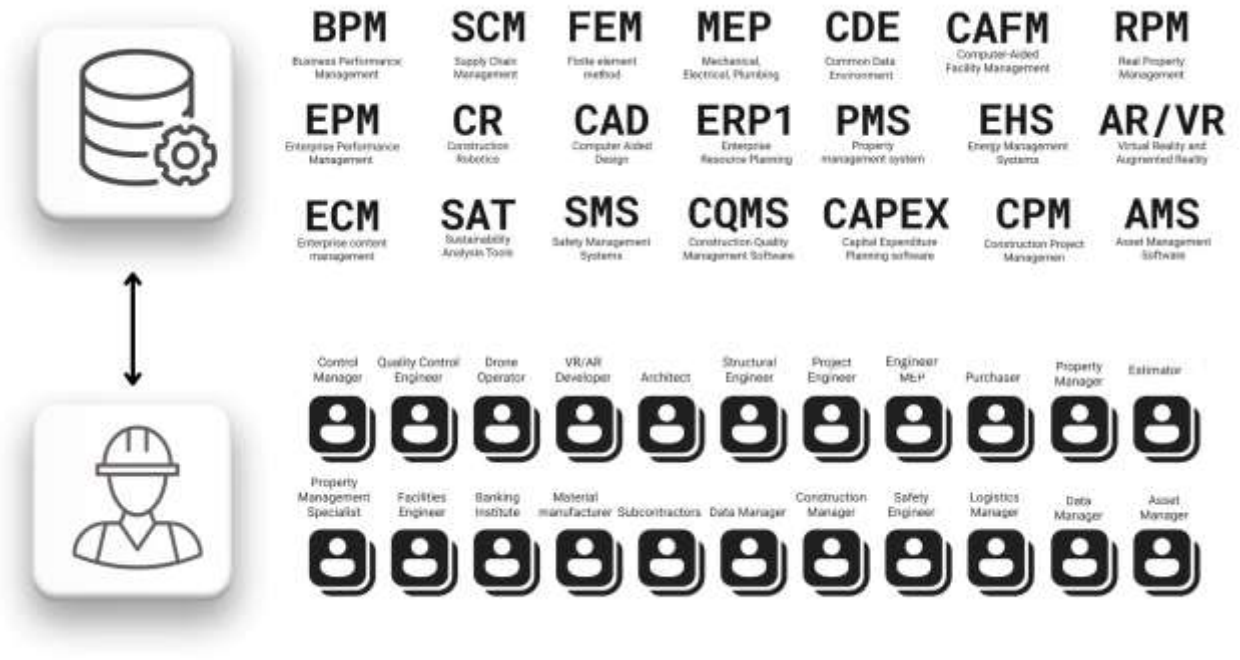


図 1.2-2 新しいソフトウエア・ソリューションは、データ・フローを管理するマネージャー軍団をビジネスに引き込んだ。

過去数十年間、企業はモジュラー・システム[15]を長期的な統合ソリューションとして捉え、多額の投資を行ってきた。

> 2022 年の Software Path レポート[16]によると、ERP -システムのユーザー一人当たりの平均予算は 9,000 ドルである。平均して、企業の従業員の約 26%がこのようなシステムを使用している。したがって、100 人のユーザーを抱える組織の場合、ERP 導入の総コストは約 900,000 ドルに達する。

近代的で柔軟なオープン・テクノロジーの急速な発展を背景に、プロプライエタリでクローズドなモジュール式ソリューションへの投資は、ますます正当化されなくなってきている。そのような投資がすでに行われているのであれば、既存のシステムの役割を客観的に再評価することが重要である。長期的に必要不可欠であり続

けるのか、あるいは、その機能を見直し、より効率的かつ透明性の高い形で実施することができるのかどうか。

今日のモジュール式データ・プラットフォームの主な問題のひとつは、データ管理を閉じたアプリケーションに一元化していることだ。その結果、企業の中核資産であるデータは、特定のソフトウェア・ソリューションに依存するようになり、むしろその逆になってしまう。これは情報の再利用を制限し、 への移行を複雑にし、急速に変化するデジタル環境におけるビジネスの俊敏性を低下させる。

クローズド・モジュラー・アーキテクチャーの関連性や妥当性が将来的に低下する可能性が高いのであれば、現在発生しているコストをサンクコストとして認識し、よりオープンで拡張性と適応性に優れたデジタル・エコシステムへの戦略的転換に注力することは理にかなっている。

> プロプライエタリ・ソフトウェアは、そのようなソリューションの使用の一部として作成されたソースコードとユーザーデータを開発者が排他的に管理することを特徴とする。オープンソースソフトウェアとは異なり、ユーザーはアプリケーションの内部構造にアクセスすることはできず、独自にレビューしたり、変更したり、ニーズに合わせたりすることはできません。その代わり、ベンダーが設定した範囲内でソフトウェアを使用する権利を付与するライセンスを購入する必要がある。

データは、独立した、耐久性のある、特定のソフトウェア・ソリューションから切り離された、主要な戦略的資産と見なされるべきである。アプリケーションは、重要な情報を失うリスクなしに自由に交換できる単なるデータツールとなる。

1990 年代の ERP と MRP システムの開発（図 1.2-1）は、企業にプロセス管理のための強力なツールを提供したが、同時に、情報の流れの維持に携わる人の数を大幅に増やすという、予期せぬ結果をもたらした。これらのシステムは、業務タスクを自動化し簡素化する代わりに、新たなレベルの複雑さ、官僚主義、社内 IT リソースへの依存を生み出すことが多かった。

データ管理システム：データマイニングからビジネス課題まで

今日の企業は、複数のデータ管理システムを統合する必要に迫られている。データ管理システムを選択し、これらのシステムを賢く管理し、異種のデータソースを統合することは、ビジネスパフォーマンスにとって非常に重要になっています。

2020 年代半ばには、建設プロセスのあらゆる側面を円滑に一貫性を持って進めるために、調和して機能しなけ

ればならないさまざまなシステム（図 1.2-3）が何百（大規模な建設会社では何千）も見つかるだろう。

> デロイトの 2016 年の調査® 「デジタル・キャピタル・プロジェクトにおけるデータ駆動管理」によると、平均的な建設専門家は毎日 3.3 つのソフトウェア・アプリケーションを使用しているが、そのうち互いに統合されているのは 1.7 つだけである[17]。

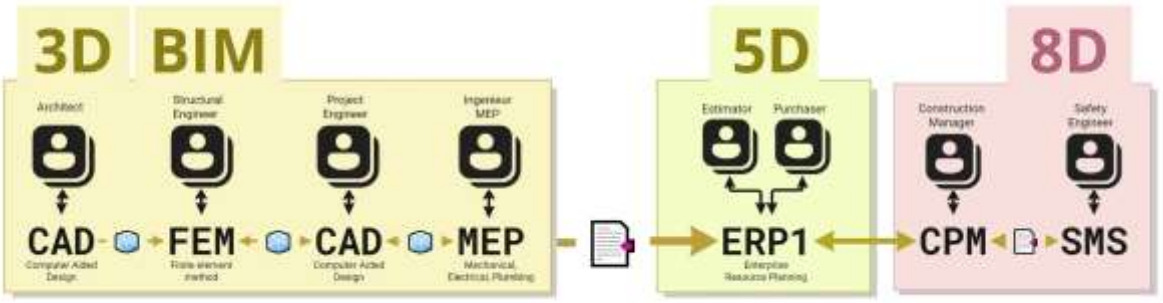


図 1.2-3 どのようなビジネス・システムにも、質の高いデータ管理のための専門チームと責任ある管理者が必要である。

以下は、建設業界の中規模から大規模の企業で、効果的な建設プロジェクト管理に使用されている一般的なシステムのリストです：

- ERP (企業資源計画) - 会計、調達、プロジェクト管理などのビジネスプロセスの統合を提供。
- CAPEX (Capital Expenditure Planning Software) - 建設プロジェクトの予算編成と財務投資に使用され、固定資産と長期資産への投資のコストを決定するのに役立つ。
- CAD （コンピューター支援設計）と BIM （ビルディング・インフォメーション・モデリング）は、詳細で正確な技術図面とプロジェクトの 3D （モデル）を作成するために使用されます。これらのシステムの焦点は、幾何学的情報を扱うことにある。
- MEP (機械、電気、配管) - 機械、電気、配管コンポーネントを含むエンジニアリングシステムで、プロジェクトの内部「循環」システムの詳細を示す。
- GIS (地理情報システム) - 地図作成と空間分析を含む、地形分析と計画に使用される。
- CQMS (建設品質管理ソフトウェア) - 建設工程が確立された基準や規則に準拠していることを確認し、欠陥を排除するのに役立つ。
- CPM (建設プロジェクト管理) - 建設プロセスの計画、調整、管理を含む。
- CAFM (Computer-Aided Facility Management) - ビル管理とメンテナンスシステム。

- SCM（サプライ・チェーン・マネジメント）は、サプライヤーと建設現場間の資材や情報の流れを最適化するために必要である。
- EPM (エンタープライズ・パフォーマンス・マネジメント) - ビジネス・プロセスとパフォーマンスの改善を目的とする。
- AMS (資産管理ソフトウェア) - 資産のライフサイクルを通じて、設備やインフラの使用、管理、メンテナンスを最適化するために使用される。
- RPM (Real Property Management) - 建物や土地、また関連する資源や資産の管理・運営に関する業務やプロセスが含まれる。

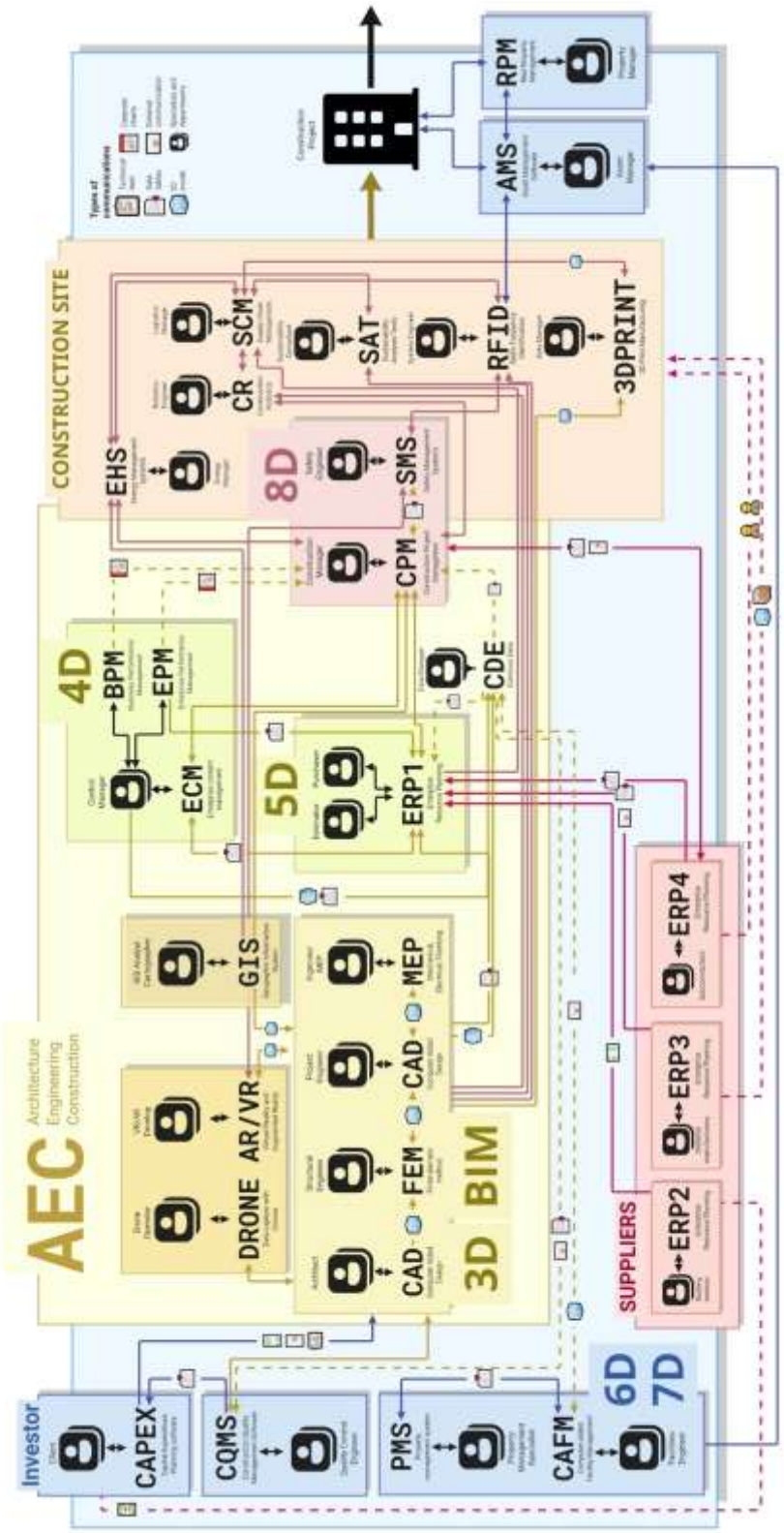


図 1.2-4 企業のプロセスと異なる部門間の情報の流れをつなぐシステムの相互接続性。

- CAE (Computer-Aided Engineering) - コンピュータ支援エンジニアリング。有限要素解析（FEA）や数値流体力学（CFD）などの計算・シミュレーションシステムを含む。
- CFD (Computational Fluid Dynamics) - 計算流体力学、流体や気体の流れのモデリング。CAE サブカテゴリ.
- CAPP (Computer-Aided Process Planning) - コンピュータ支援工程計画。ルートマップやプロセスマップの作成に使用される。
- CAM (Computer-Aided Manufacturing) - コンピュータ支援製造、CNC 機械の制御プログラム作成。
- PDM (Product Data Management) - 製品データ管理、技術文書を保存・管理するシステム。
- MES（Manufacturing Execution System）は、リアルタイムの製造工程管理システムである。
- PLM (Product Lifecycle Management) - プロジェクト要素のライフサイクル管理。PDM 、CAPP、CAM およびその他のシステムを統合し、開発から廃棄に至るまで製品を完全に管理。

このようなシステムをはじめ、さまざまなソフトウェア・ソリューショ ンを組み込んだ多くのシステムが、現代の建設業界に不可欠な要素となってい る（図 1.2-4）。このようなシステムの中核は、直感的なインターフェイスを備えた専用データベースであり、設計と施工のあらゆる段階で情報の効率的な入力、処理、分析を実現する。デジタルツール同士の統合は、作業プロセスの最適化に役立つだけでなく、意思決定の精度を大幅に向上させ、プロジェクト実施のタイミングと品質に好影響を与える。

しかし、半数のケースでは統合されていない。統計によると、他のソリューションと統合されているアプリケーションやシステムは、わずか 2 件に 1 件しかない[17]。これは、デジタル環境の断片化が続いていることを示しており、建設プロジェクト内でのエンド・ツー・エンドの情報交換を確実にするために、オープンスタンダードと統一インターフェースを開発する必要性を強調している。

現代の企業にとって、統合における主な課題のひとつは、デジタル・システムの高い複雑性と、効果的な情報検索と解釈に必要なユーザー能力の要件であることに変わりはない。キー・マネージャー（図 1.2-2）を筆頭とするスペシャリストからなるチームが、ビジネスに導入される各システムをサポートするために結成される。

> 数千年前の最初のマネージャーがパピルスや粘土板に書かれた数字に責任を負っていたように、キーとなるシステム・マネージャーはデータの流れを正しい方向に導く決定的な役割を果たし、最終的な情報の質に責任を負う。

バラバラな情報の流れを経営ツールに変えるには、データを体系的に統合・管理する能力が不可欠である。このアーキテクチャーでは、マネジャーはひとつのネットワークの要素として行動しなければならない。それは

、企業の各部分を、適応し進化することのできる首尾一貫した生命体につなぐ菌糸のようなものである。

企業の菌糸体：データはどのようにビジネス・プロセスにつながるか

データをアプリケーションやデータベースに統合するプロセスは、さまざまな部門や専門家を含む、さまざまなソースからの情報の集約に基づいている（図 1.2-4）。スペシャリストは、関連するデータを検索し、それを処理し、システムやアプリケーションに転送してさらに利用する。

ツール、テクノロジー、データベースのセットからなる各企業のシステムは、過去のデータという土壌に根を張り、文書、計算、表、グラフ、ダッシュボードといった完成されたソリューションという形で新たな実を結ぶために成長する知識の木である（図 1.2-5）。企業内のシステムは、特定の森の木々のように、相互に影響し合い、コミュニケーショ ンを取りながら、専門家であるマネジャーによってサポートされ、管理される、複雑で構造化された システムを構成している。

> 企業の情報検索・転送システムは、木（システム）と菌糸茸（管理者）からなる複雑な森林ネットワークのように機能し、導線とリサイクラーの役割を果たし、情報が転送され、適切なシステムに流れるようにする。これにより、社内の健全で効率的なデータの流れと流通を維持することができる。

専門家は、根のように、プロジェクトの初期段階で生データを吸収し、それを企業エコシステムの栄養分に変える。データとコンテンツの管理システム（図 1.2-4 - ERP 、CPM 、BIM など）は、強力な情報ハイウエイとして機能し、このハイウエイを通じて、この知識が企業のあらゆるレベルを循環する。

生態系の各要素がそれぞれの役割を果たす自然界と同じように、企業のビジネス環境においても、エンジニアからアナリスト に至るまで、各プロセスの参加者が情報環境の成長と肥沃化に貢献している。このようなシステマティックな「データツリー」（図 1.2-5）は、単なる情報収集の仕組みではなく、企業の持続可能性を保証する競争上の優位性である。

森の生態系は、デジタル企業構造がどのように組織化されているかを驚くほど正確に反映している。森林の下草から梢までの階層構造のように、コーポレート・ガバナンスは責任レベルと機能部門にタスクを割り当てる。

深く枝分かれした木の根は、弾力性と養分へのアクセスを提供する。同様に、堅固な組織構造と質の高いデータを扱うための安定したプロセスは、企業の情報エコシステム全体を支え、（強風による）市場の不安定な時期や危機的状況においても、企業の持続可能な成長と発展に貢献する。

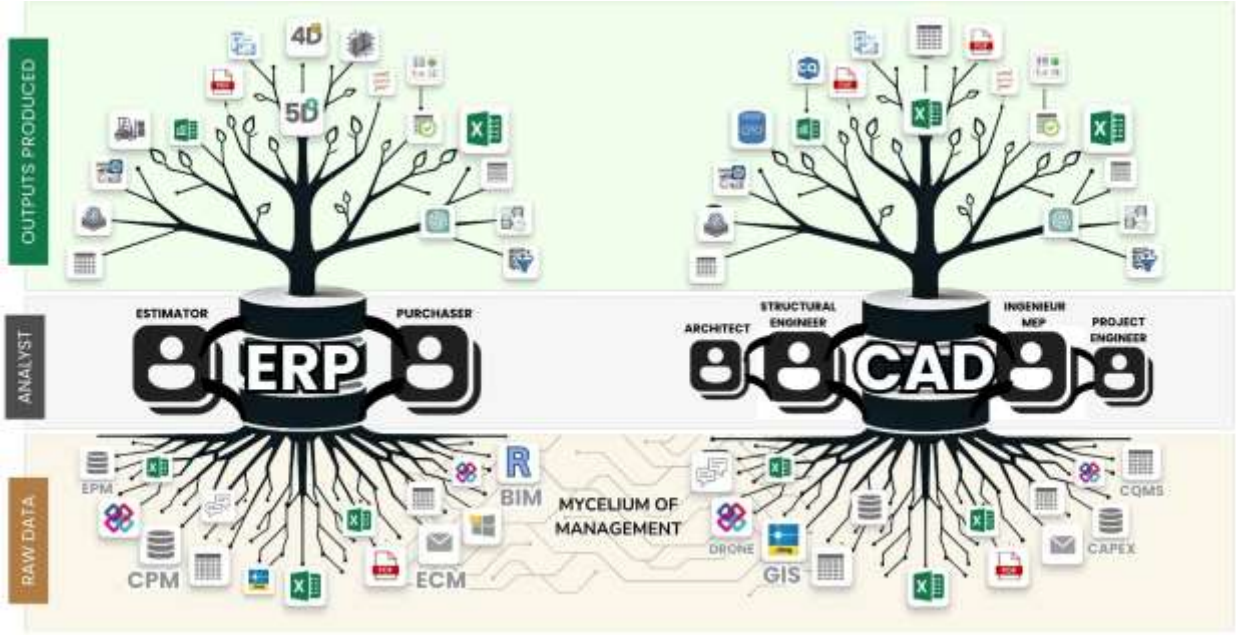


図 1.2-5 異なるシステムを通してデータを統合することは、管理者と専門家を一つの情報ネットワークにつなぐ菌糸のようなものである。

ビジネスにおける規模の現代的理解は進化した。今日、企業の価値は目に見える部分、つまり最終的な書類や報告書という形の「王冠」だけでなく、定性的に収集され体系的に処理されたデータの「根幹システム」の深さによっても決まる。より多くの情報を収集し、処理することができればできるほど、ビジネス価値は高くなる。すでに処理されたデータの「堆肥」を計画的に蓄積し、そこから有用な洞察を引き出すことができる企業は、戦略的優位性を獲得する

過去の情報は、成長、プロセスの最適化、競争上の優位性を可能にする新たな資本となりつつある。データ主導の世界では、誰がより多く持っているかではなく、誰がより多く知っているかが勝者となる。

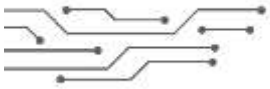
建設業界にとって、これはリアルタイムのプロジェクト管理への移行を意味し、設計や調達から請負業者の調

整に至るまで、すべてのプロセスが関連する日々更新されるデータに基づいて行われるようになる。さまざまな情報源（ERP -システム、CAD -モデル、センサー 建設現場の IoT 、RFID ）からの情報を統合することで、より正確な予測、変化への迅速な対応、最新データ不足による遅延の回避が可能になる。

> マッキンゼー・アンド・カンパニーの『データ駆動型エンタープライズ 2025』（McKinsey & Company®, 2022 [18]）によると、将来成功する企業は、戦略的意思決定から業務上のやりとりに至るまで、事業のあらゆる重要な側面でデータに依存するようになる。

データは単なる分析ツールではなくなり、 、すべてのビジネス・プロセスの不可欠な一部となり、経営の透明性、管理、自動化を実現する。 の下でのデータ主導の動きにより、組織は人的要因の影響を最小限に抑え、業務上のリスクを軽減し、意思決定の透明性と効率を高めることができる。

21 世紀は経済のパラダイムをひっくり返そうとしている。かつて石油は、機械や輸送の動力源として "黒い黄金 "と呼ばれていたが、今日では、時間的なプレッシャーの中で圧縮された過去のデータが新たな戦略的資源となりつつあり、機械ではなく、ビジネスを推進する意思決定アルゴリズムに電力を供給している。



章 1.3.

デジタル革命とデータの爆発

進化の波としてのデータブームの始まり

建設業界は、かつてない情報爆発に見舞われている。ビジネスを、データを栄養源とする知識の木（図 1.2-5）と考えるなら、デジタル化の現段階は、バイオマスの急速な蓄積によって地球の生物圏が変容した時代（図 1.3-1）である石炭紀の植生の急成長にたとえることができる。

世界的なデジタル化に伴い、建設業界の情報量は毎年倍増している。現代のテクノロジーは、データをバックグラウンドで収集し、リアルタイムで分析し、少し前までは不可能と思われた規模で活用することを可能にしている。

> ゴードン・ムーア（インテル®の共同創設者）が定式化したムーアの法則によれば、集積回路の密度と複雑さ、および処理・保存されるデータ量は、およそ 2 年ごとに倍増する[19]。

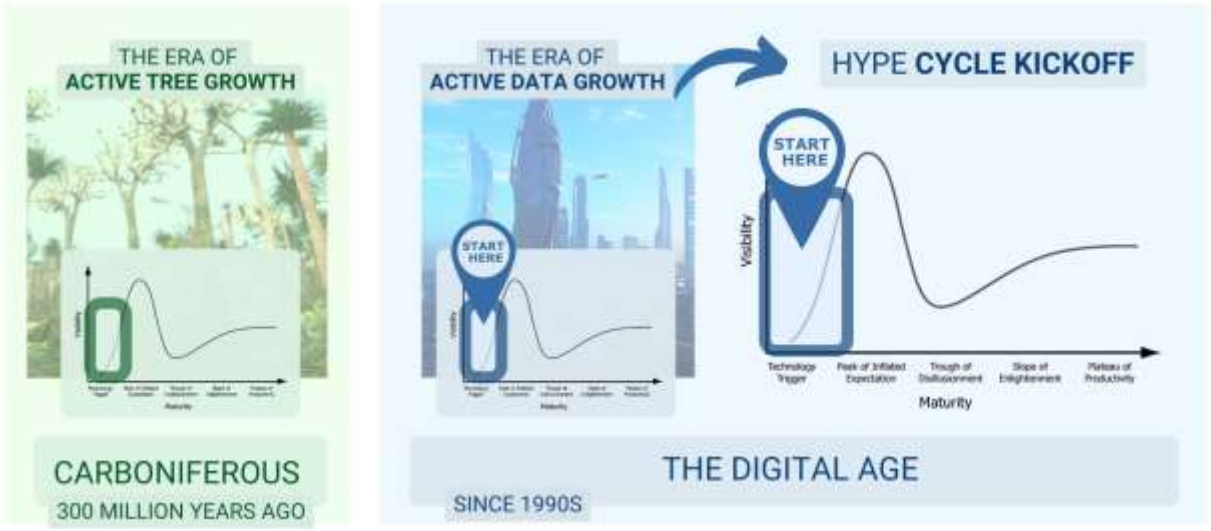


図 1.3-1 デジタル化の始まりは、石炭時代の植生の急増のように、データの急激な増加をもたらした。

ギョベクリ・テペ（トルコ）のような古代の巨石建造物は、再利用に適した知識を文書化したものを残さなかったが、今日ではデジタル技術によって情報の蓄積と再利用が可能になった。これは、胞子植物から種子植物（被子植物）への進化の変遷に例えることができる。種子の出現は、地球上に生命が広く普及するきっかけと

なった。(図 1.3-2）。

同様に、過去のプロジェクトのデータは、一種の "デジタルの種"、つまり、新しいプロジェクトやプロ ドキュメントでスケールアップして使用することができる DNA ナレッジキャリアとなる。ChatGPT、LlaMa、Mistral、Claude、DeepSeek、QWEN、Grok などの機械学習や大規模言語モデル（LLM）といった最新の人工知能ツールの出現により、データを自動的に抽出、解釈し、新たな文脈に適用することが可能になった

> 種が、最初は生命のなかった惑星に生命を広める革命を起こしたように、「データの種」は、新しい情報構造と知識を自動的に出現させる基礎となりつつあり、デジタル・エコシステムが独立して進化し、変化するユーザーの要求に適応することを可能にしている。

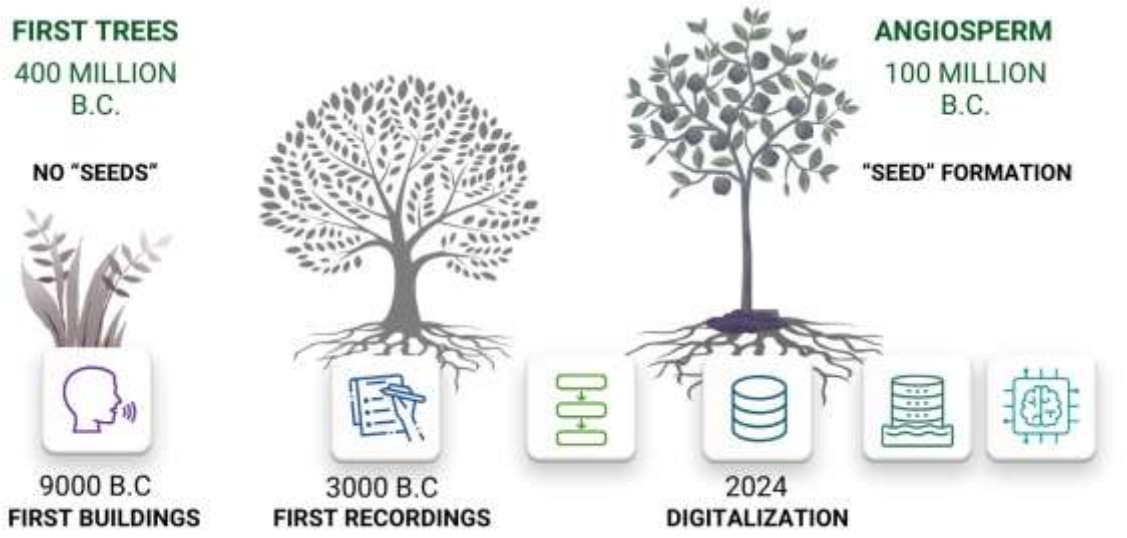


図 1.3-2 デジタルの「データ・シード」は、地球の生態系を変えた被子植物と同じ進化の役割を果たす。

データの爆発的な増大と、「データの種」（過去や進行中のプロジェクトから得られた構造化された情報）の積極的な普及が、業界のデジタルな未来の基盤を形成している。ビッグデータの言語モデル（LLM）を通じた「受粉」によって、私たちはデジタルの変化を観察するだけでなく、自己学習し適応するエコシステムの創造に積極的に参加することができる。これは進化ではなく、データが新しい現実の主要な構成要素になりつつあるデジタル革命なの。

> 建設プロジェクトのライフサイクルを通じて様々な分野から情報が得られるため、建設業界のデータ量は飛躍的に増加している。この膨大なデータの蓄積が、建設業界をビッグデータの時代へと押し上げている[20]。
>
> - Hang Yang 教授、武漢理工大学土木建築学部、武漢、

森林の発達が太古の地球の景観を変えたように、現在の情報爆発は建設業界全体の景観を変えつつある。

現代企業で生成されるデータ量

この 2 年間で、世界に存在する全データの 90%が作成された [21]。2023 年の時点で、建設業界の専門家を含む一人一人が 1 秒間に約 1.7 メガバイトのデータを生成しており[22]、世界のデータ総量は 2023 年に 64 ゼタバイトに達し、2025 年には 180 ゼタバイト、すなわち 180*10^15 メガバイトを超えると予測されている[23]。

> 15 世紀にヨハネス・グーテンベルクが印刷機を発明したのだ。導入後わずか 50 年で、ヨーロッパにおける書籍の数は倍増し、数十年の間に、それまでの 1,200 年間に手作業で作成されたのと同じ数の書籍が印刷された[24]。世界のデータ量は 3 年ごとに倍増している。

> 現在のデータ増加率を考えると、建設業界は今後数十年の間に、これまでの歴史が蓄積してきたのと同じだけの情報を生み出す可能性がある

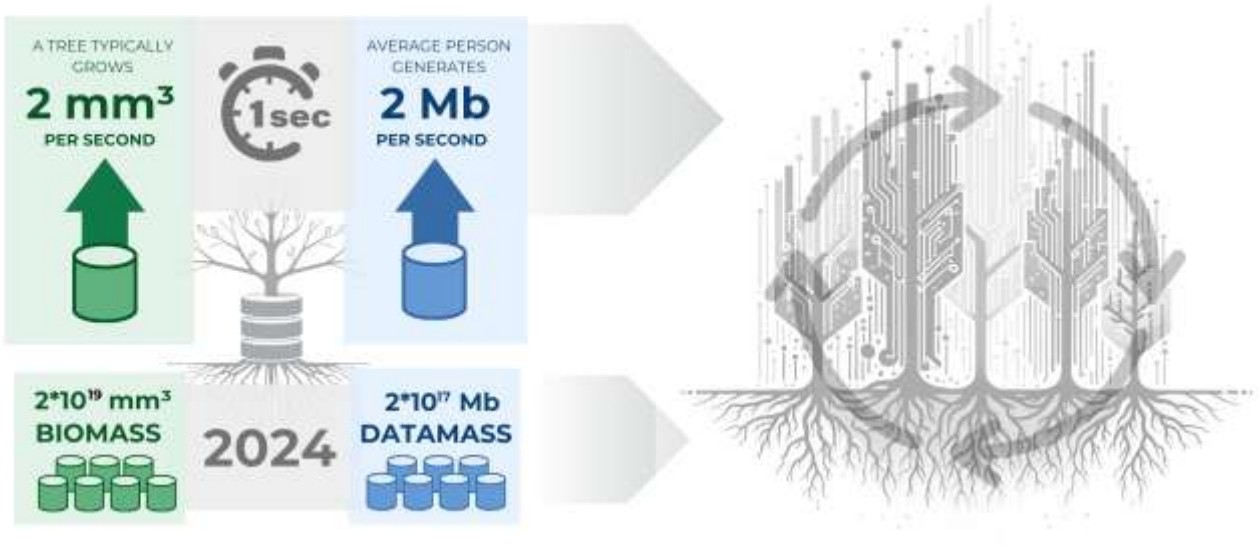


図 1.3-3 社員一人ひとりが日々、社内のサーバーにデータを保存しているため、データは常に増え続けている。

今日の建設ビジネスの世界では、小さな会社でも日々膨大な量のマルチフォーマット情報が生成され、模型や図面から写真記録、現場のセンサーに至るまで、デジタルフットプリントは 1 日あたり数十ギガバイトに達する。各技術者が 1 秒間に平均約 1.7MB のデータを生成すると仮定すると、これは 1 日あたり約 146GB、1 年あたり 53TB に相当する（図 1.3-3）。

10 人のチームが毎日 3 時間だけアクティブに働くと、1 日に生成される情報の累積量は 180 ギガバイトに達する（図 1.3-4）。

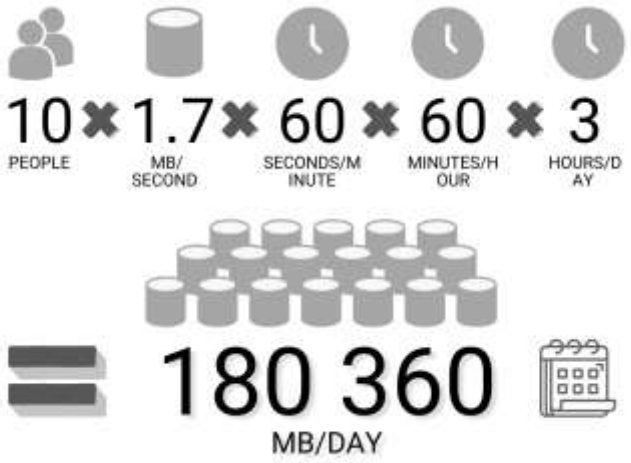


図 1.3-4 10 人の企業が 1 日に生成するデータは約 50〜200 ギガバイト。

作業データの 30%が新規データ（残りは上書きまたは削除）であると仮定すると、10 人規模の企業では、月に数百ギガバイトの新規データが作成されることになる（実際の数字は、その企業のビジネスの種類によって異なる）

このように、私たちはますます多くのデータを生み出しているだけでなく、その効率的な管理、保管、長期的な可用性に対するニーズが高まっていることは明らかです。そして、以前はデータはローカルのサーバーにコストをかけずに「眠る」ことができたが、デジタルトランスフォーメーションの文脈では、ますます多くの企業が情報インフラの基盤としてクラウド・ソリューションを使い始めている。

データ保存のコスト：経済的側面

近年、データ・ストレージをクラウド・サービスにアウトソーシングする企業が増えている。例えば、ある企業がデータの半分をクラウドでホスティングする場合、1 ギガバイトあたり月額平均 0.015 ドルで、ストレージ・コストは毎月 10〜50 ドル増加する可能性がある[25]。

典型的なデータ生成パターンを持つ小規模企業にとって、クラウド・ストレージのコストは、 数年で、月額数

百ドルから潜在的には千ドル以上（図 1.3-5）に達する可能性があり、大きな財務的負担となる可能性がある。

> テクノロジーインフラの意思決定者 214 人を対象とした Forrester 社の調査「複雑化するデータストレージをアウトソーシングする企業」[26]によると、企業の 3 分の 1 以上が、データ運用の量と複雑さの増大に対処するためにストレージをアウトソーシングしており、約 3 分の 2 の企業がストレージのアウトソーシングを好んでいる。[26]によると、3 分の 1 以上の組織が、データ運用の量と複雑さの増大に対処するためにストレージをアウトソーシングしており、3 分の 2 近くの企業がサブスクリプションベースのモデルを好んでいる。



図 1.3-5 データをクラウドに移行すると、従業員が 10 人程度の企業であっても、毎月のストレージ・コストが最大 2,000 ドル増加する可能性がある。

CAD （BIM ）、CAFM 、PMIS 、ERP などのクラウド技術の採用が加速しており、データ保管や処理コストがさらに増加しているため、状況はさらに複雑になっている。その結果、企業はコストを最適化し、クラウドプロバイダーへの依存度を下げる方法を模索せざるを得なくなっている。

2023 年以降、大規模言語モデル（LLM ）の開発が活発に行われるようになり、データ保存に対するアプローチも変わり始めている。自社のサーバーで情報を処理した方が安全で収益性が高くなるため、データの管理を取り戻そうと考える企業が増えている。

この文脈では、クラウドベースのストレージや必要なデータのみの処理から、企業向け LLM や AI -ソリューションのローカル展開を支持する傾向が前面に出てくる。マイクロソフトの CEO がインタビュー[27]で指摘したように、さまざまなタスクを実行するために複数の別々のアプリケーションやクラウドベースの SaaS ソリューションに頼るのではなく、AI エージェントがデータベース内のプロセスを管理し、さまざまなシステムの機能を自動化する。

> [この（データ処理の）問題に対する昔のアプローチは、さまざまなビジネス・アプリケーションがどのように統合を処理していたかを思い返すと、コネクタを使っていた。企業はコネクターのライセンスを販売し、それを中心にビジネスモデルが形成された。SAP [ERP]はその典型的な例のひとつで、適切なコネクターがなければ SAP のデータにアクセスできない。ですから、エージェントとのインタラクション AI ）においても、同じようなことが起こると思います。ビジネス・アプリケーションの存在という概念は、おそらく[AI]エージェントの時代に崩壊すると思います。というのも、考えてみれば、それらは本質的に、ビジネス・ロジックの束が入ったデータベースだからです
>
> - 2024 年、マイクロソフト CEO、サティア・ナデラ、BG2 チャンネルとのインタビュー。[28]

このパラダイムでは、データ駆動型 LLM アプローチは古典的なシステムを超える。人工知能がユーザーとデータの仲介役となり（図 2.2-3、図 2.2-4）、複数の仲介インターフェースの必要性を排除し、ビジネスプロセスの効率を高める。データを扱うこのアプローチについては、「カオスを秩序に変え、複雑さを軽減する」の章で詳しく説明する。

未来のアーキテクチャーはまだ具体化していないが、企業はすでに過去の決断の結果に直面している。ここ数十年の大規模なデジタル化は、異種システムの導入やデータの無秩序な蓄積を伴い、情報過多という新たな問題を引き起こしている。

データ蓄積の最前線：質量から意味へ

現代の企業システムは、データ量やアプリケーションの数が IT 部門や管理者の能力と釣り合っていれば、管理された成長の下でうまく発展し、機能している。しかし、ここ数十年、デジタル化によってデータの量と複雑さが制御不能なまでに増大し、企業の情報エコシステムに過飽和状態が生じている。

今日、サーバーやストレージ設備には、堆肥に変わる暇もなく、急速に無関係になりつつある、未処理でマルチフォーマット化された情報が、かつてないほど大量に押し寄せている。限られた企業リソースではこの大洪

水に対処できず、データは孤立したサイロ（いわゆる「サイロ」）に蓄積され、有用な情報を抽出するためには手作業による処理が必要となっている。

> その結果、ツタが生い茂り、カビに覆われた森のように、現代の企業管理システムはしばしば情報過多の状態に陥っている。企業のエコシステムの中心で情報の腐葉土を養う代わりに、異なるフォーマットのデータが孤立した領域が形成され、必然的にビジネス・プロセスの全体的な効率を低下させる。

過去 40 年にわたる指数関数的なデータ増加の長い期間は、必然的に飽和とそれに続く冷却段階を迎える。ストレージが限界に達すると、質的な変化が起こる。データはもはや単なるストレージ・オブジェクトではなく、戦略的リソースとなる。

人工知能と機械学習の発展により、企業は情報処理コストを削減し、データの定量的な増加から定性的な利用へと移行する機会を得ている。今後 10 年間で、建設業界はより多くのデータを作成することから、その構造、完全性、分析価値を確保することに重点を移さなければならないだろう。

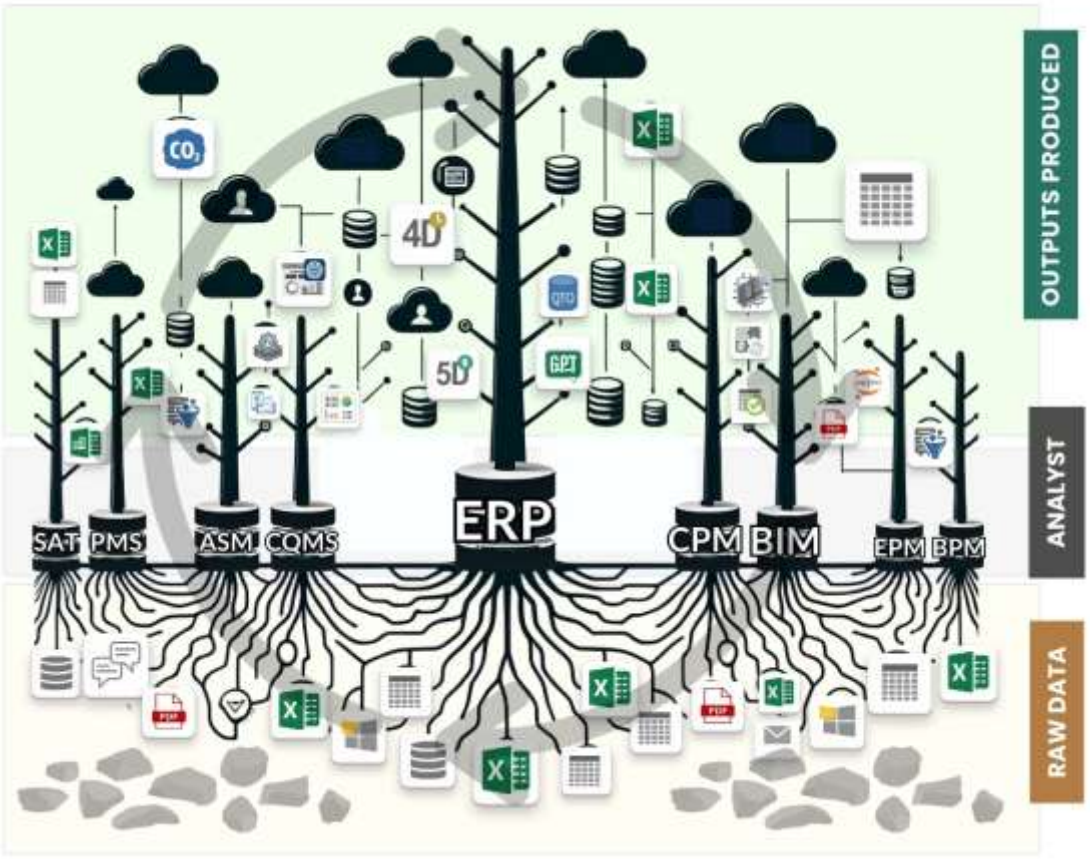


図 1.3-6 孤立したデータソースは、データシステム間の情報共有を妨げる。

> 主な価値はもはや情報量ではなく、それを自動的に解釈し、経営上の意思決定に役立つ応用知識に変える能力にある。データが真に有用なものとなるためには、収集、検証、構造化、保存、特定のビジネス・タスクの文脈での分析など、適切に管理されなければならない。

成熟した木は枯れ、分解し、新たな成長の温床となる。成熟した木は枯れ、分解し、新たな成長の温床となる。完成し、完了したプロセスは、企業の情報エコシステムの一部となり、最終的には新たなシステムやデータの将来の成長を促進する情報の腐葉土となる。

しかし実際には、このサイクルはしばしば破綻する。有機的な更新の代わりに、地層のように層状のカオスが形成され、深い統合や構造化がなされないまま、古いシステムの上に新しいシステムが重ねられていく。その結果、バラバラの情報の「サイロ」が出現し、知識の循環を妨げ、データ管理を複雑にしてしまう。

次のステップ：データ理論から実践的変革へ

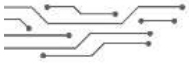
建設におけるデータの進化は、粘土板から最新のモジュール式プラットフォームへの旅である。今日の課題は、情報を収集することではなく、異質で多様なデータを戦略的リソースに変えるフレームワークを構築することである。あなたの役割が会社の経営者であれエンジニアであれ、データの価値を理解し、それをどのように扱うかは、今後重要なスキルとなるだろう。

この部分を要約すると、考察したアプローチを日常業務に応用するのに役立つ主な実践的ステップを強調する価値がある：

- ■ 情報フローの個人監査の実施
  - ☐ 日常的に使用しているすべてのシステムとアプリケーションのリストを作成する。
  - ☐ データの検索や再確認に最も時間を費やす場所に印をつける。
  - ☐ 主要な情報源を特定する
  - ☐ 現在のアプリケーションの状況を分析し、冗長性や機能の重複をチェックする。
- ■ 分析成熟度別にプロセスを進めるよう努力する。
  - ☐ （何が起こったのか？）
  - ☐ 徐々に診断導入（なぜこうなったのか？）
  - ☐ プロセスにおいて、予測的分析（何が起こるか）と処方的分析（何をすべきか）にどのように移行できるかを考える。
- ■ 作業データの構造化を開始する
  - ☐ 仕事でよく使うファイルやフォルダの名前を統一する。
  - ☐ 頻繁に使用する文書やレポートのテンプレートを作成
  - ☐ 完了したプロジェクトを明確な構造で定期的にアーカイブする。

チームや会社の情報インフラ全体を変えることができなくても、まずは自分自身のプロセスや日常業務の小さな改善から始めてみよう。データの本当の価値は、その量ではなく、そこから実用的な洞察を引き出す能力にあることを忘れてはならない。小さくても適切に構造化され、分析されたデータセットであれば、意思決定プロセスに組み込むことで大きな影響を与えることができる。

本書の次のパートでは、データを扱うための具体的な方法とツールに移り、非構造化情報を構造化されたセットに変換する方法、アナリティクスの自動化技術、建設会社における効果的なアナリティクス・エコシステムの構築方法について詳しく説明します。





II 部

建設業はいかにしてデータの混乱に溺れているか

第 2 部では、増加するデータ量に対応するために建設会社が直面する課題を批判的に分析する。情報の断片化がもたらす結果と、効果的な意思決定を妨げる「サイロ化したデータ」現象について詳しく検証する。HiPPO -approach (Highest Paid Person's Opinion)の問題点と、建設プロジェクトにおける経営判断の質への影響について調査している。ダイナミックなビジネスプロセスとその複雑化が、情報の流れと業務効率に与える影響について評価する。システムの過度な複雑化がいかにコストを増大させ、組織の柔軟性を低下させるかについて、具体的な例を挙げて説明する。特に、独自のフォーマットによって生じる制約と、建設業界におけるオープンスタンダードの活用の可能性に注目する。過剰な複雑さと技術的障壁を最小化する、AI と LLM ベースのソフトウェア・エコシステムへの移行というコンセプトが提示されている（）。

章 2.1.

データの断片化とサイロ化

ツールが多ければ多いほど、ビジネスの効率は上がるのか？

一見すると、デジタルツールが増えれば効率が上がるように思えるかもしれない。しかし、実際にはそうではない。クラウド・サービスであれ、レガシー・システムであれ、また別のエクセル・レポートであれ、新しいソリューションを導入するたびに、企業はデジタル・ランドスケープに新たなレイヤーを追加することになる。

データは石炭や石油に例えることができる。混沌、エラー、構造化されていないプロセス、忘れ去られたフォーマットの層の下に圧縮され、蓄積されるのに何年もかかる。そこから真に有用な情報を引き出すには、企業は文字通り、時代遅れのソリューションやデジタル・ノイズの層をかき分けていかなければならない。

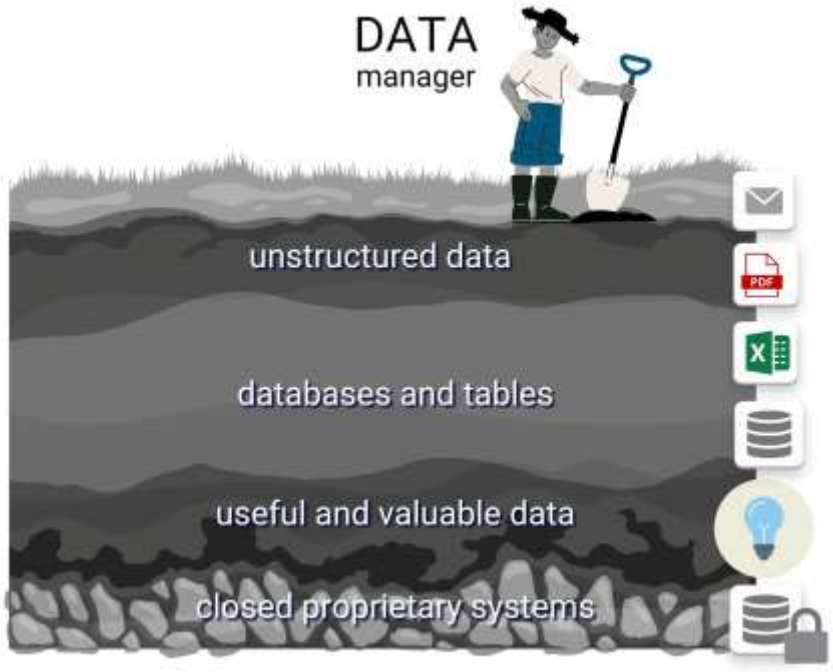


図 2.1-1 多様なデータは区分された層を形成する。「黄金」の洞察も、システム的に複雑な地質学的岩石の中では失われてしまう。

すべての新しいアプリケーションは、ファイル、テーブル、あるいはサーバー上の孤立した「サイロ」といっ

た痕跡を残す。ある層は粘土（時代遅れで忘れられたデータ）、別の層は砂（バラバラのテーブルやレポート）、3 つ目は花崗岩（統合できない閉じた独自フォーマット）だ。時間が経つにつれて、企業のデジタル環境はますます、制御されない情報の蓄積の貯水池のようになり、企業のサーバーの奥深くで価値が失われていく。

新しいプロジェクトや新しいシステムを導入するたびに、インフラが複雑になるだけでなく、有用な質の高いデータへの道筋も複雑になっていく。価値ある「岩石」にたどり着くには、ディープ・クレンジング、情報の構造化、「チャンク化」、意味のある塊へのグループ化、アナリティクスとデータモデリングによる戦略的に重要な洞察の抽出が必要となる。

> データは貴重なものであり、（データを処理する）システムそのものよりも長持ちする[29]。
>
> - ティム・バーナーズ＝リー、ワールド・ワイド・ウェブの父、最初のウェブサイトの作者

データが "価値あるもの "となり、信頼できる意思決定の基礎となるには、入念な準備が必要である。適切な前処理によって、バラバラのデータが構造化された経験、有用な情報の腐葉土となり、予測や最適化のツールとなるのである。

> 分析を始めるには完璧にクリーンなデータが必要だという誤解があるが、実際には、ダーティなデータを扱えるようにすることはプロセスの重要な部分である。

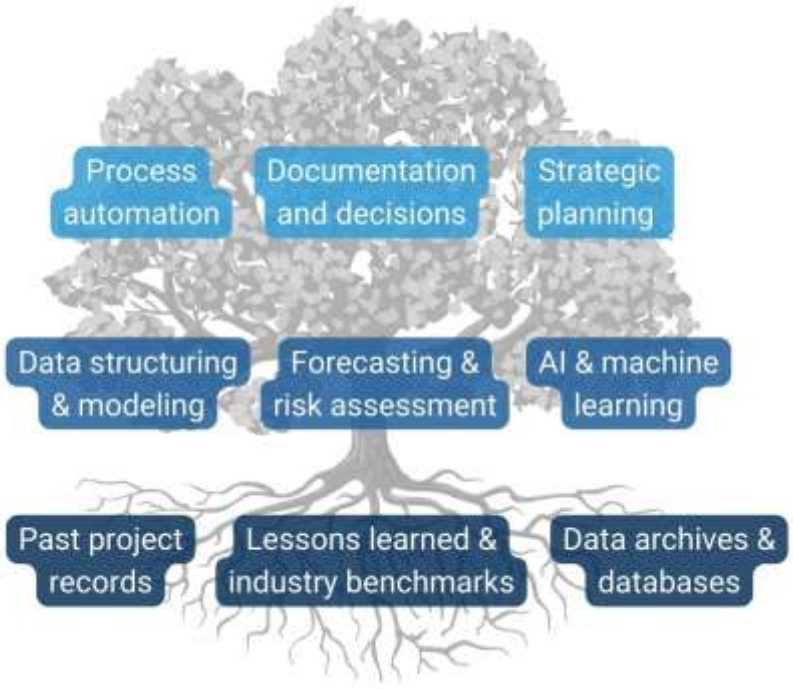


図 2.1-2 データはビジネスの根幹であり、基礎である。

テクノロジーが進歩し続ける中、ビジネスもまた前進し、データから価値を生み出す方法を学ばなければならない。石油会社や石炭会社が鉱物を抽出するためのインフラを構築しているように、企業もまた、自社のサーバー上の新しい情報の流れを管理する方法を学び、未使用、未フォーマット、時代遅れのデータから貴重な洞察を引き出し、戦略的資源に変える必要がある。

フィールド（データウェアハウス ）を作成することが第一歩である。最も強力なツールであっても、企業がサイロ化されたシステムで運用し続ければ、データの分離とマルチフォーマット・データの問題を解決することはできない。データが互いに分離して存在し、情報を交差・共有しない場合、企業は「データサイロ」効果に直面する。単一の一貫したインフラストラクチャの代わりに、企業はデータのマージと同期にリソースを費やすことを余儀なくされる。

データのサイロ化と企業業績への影響

団地を建設していると想像してほしい。あるチームは壁を造り、あるチームは通信設備を敷設し、あるチームは互いに確認することなく道路を敷いている。その結果、配管と壁の開口部が合わなかったり、エレベーターのシャフトと階数が合わなかったり、道路を解体して敷き直さなければならなかったりする。

このような状況は、単なる仮定のシナリオではなく、現代の多くの建設プロジェクトの現実である。多数のゼネコンや下請け業者が異なるシステムで作業し、単一の調整センターがないため、プロセスは終わりのない承認、手直し、衝突の連続となる。これらすべてが、大幅な遅延とプロジェクトコストの増大につながっている。

建設現場での典型的な状況は単純なものだ。型枠の準備は整ったが、鉄筋の搬入が間に合わない。さまざまなシステムで情報を確認すると、おおよそ次のようなやり取りがある：

- 20 日、建設現場の現場監督がプロジェクト・マネージャーにこう書き送った。"型枠の設置は終わったが、鉄筋はどこだ？"
- プロジェクト・マネジャー（PMIS ）から調達部門へ：「型枠の準備が整いました。私のシステム（PMIS）では、鉄筋は 18 日に到着することになっています。鉄筋はどこですか？
- サプライチェーン・スペシャリスト（ERP）：- ERP によると、納品は 25 日です。
- データエンジニアまたは IT 部門（統合担当）：- PMIS では日付は 18 日ですが、ERP では 25 日です。ERP と PMIS の間に OrderID のリンクがないため、データが同期されていません。これは情報ギャップの典型的な例です。
- プロジェクト・マネージャーからジェネラル・マネージャーへ - 「建具の納品が遅れ、現場は立ちっぱなしで、誰が責任を取るのかが不明確です。

インシデントの原因は、異種システムにおけるデータの分離にあった。データソースを統合・統一し、単一の情報リポジトリを作成し、ETL ツール（Apache NiFi、Airflow、n8n）を使って自動化することで、システム間のサイロをなくすことができる。これらの方法やツールについては、本書の後のセクションで詳しく説明する。

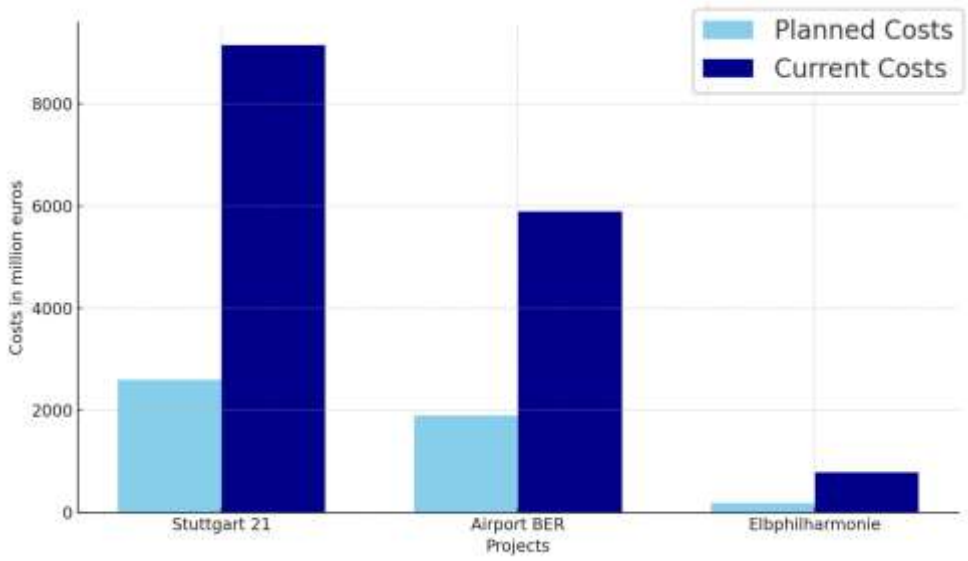


図 2.1-3 ドイツにおける主要インフラプロジェクトの計画コストと実績コストの比較。

企業システムでも同じことが起こる。孤立したソリューションが最初に作られ、それらを統合し調和させるために莫大な予算を費やさなければならない。データとコミュニケーション・モデルが最初から考え抜かれていれば、統合の必要性はまったくなかっただろう。サイロ化されたデータは、調整されていない建設プロセスのように、デジタルの世界に混乱をもたらす。

> KPMG の 2023 年調査「Cue construction 4.0: Time to make or break」によると、部門間で効果的にデータを共有している企業はわずか 36%で、61%は孤立したデータの「サイロ」による深刻な問題に直面している[30]。

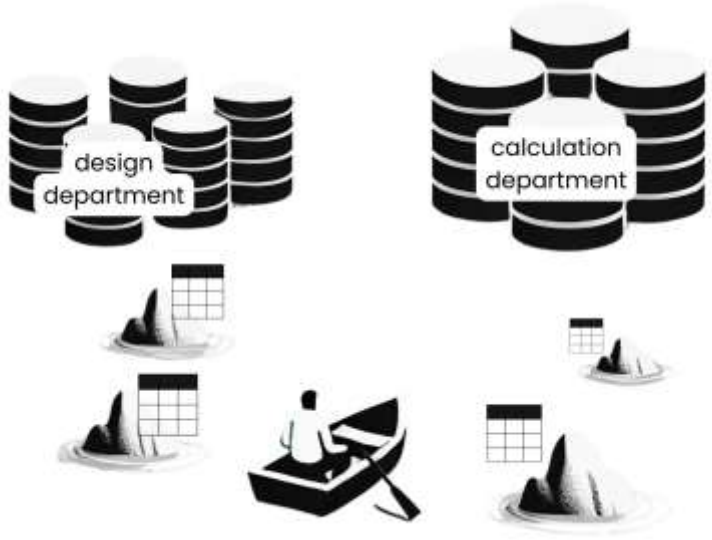


図 2.1-4 長年蓄積された収集困難なデータは、使われることのないまま、隔離されたストレージ「サイロ」に蓄積される。

企業のデータは、風景に散在する個々の木のように、孤立したシステムに保存されている。それぞれに貴重な情報が含まれているが、システム間の接続がないため、相互接続された単一のエコシステムを構築することができない。このサイロ化がデータの流れを妨げ、組織の全体像を把握する能力を制限している。これらのサイロを接続する、システム間で個々の情報を転送する方法を学ぶために、管理レベルでキノコの菌糸を成長させる非常に長く複雑なプロセスである。

2016 年の WEF の調査によると、デジタルトランスフォーメーションの主な障壁の 1 つは、共通のデータ標準の欠如と断片化である。

> 建設業界は、世界で最も細分化された業界のひとつであり、バリューチェーンにおけるすべての参加者の円滑な相互作用に依存している[5]。
>
> - 世界経済フォーラム 2016：未来を形作る

デザイナー、マネージャー、コーディネーター、開発者は、複雑な調整を避け、自律的に仕事をすることを好む。この自然な傾向は、データが別々のシステム内に隔離された情報「サイロ」を生み出すことにつながる。このような孤立したシステムが多ければ多いほど、それらを連携させるのは難しくなる。時が経つにつれ、各システムは独自のデータベースを持ち、マネージャーによる専門的なサポート部門を持つようになり（図 1.2-4

）、統合はさらに複雑になる。

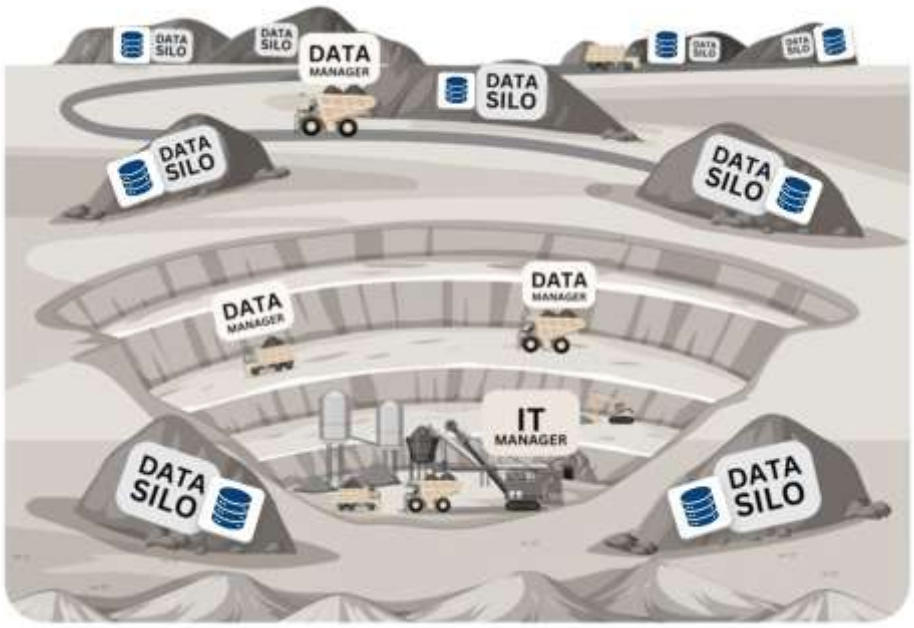


図 2.1-5 各システムは、適切なツールによって処理される必要のあるデータの、独自のサイロを作りがちである[31]。

企業は複雑な iso ソリューションに投資し、その統合のために高いコストに直面し、開発者はシステムを組み合わせることの複雑さに気づき、閉じたエコシステムでの作業を好む。こうしたことが、IT ランドスケープの断片化を助長し、新しいソリューショ ンへの移行を難しくしている（図 2.1-5）。マネジャーはデータのサイロ化を批判するばかりで、その原因や防止策を分析することはほとんどない。マネジャーは時代遅れの IT システムに不満を漏らすが、IT システムのリプレースには多額の投資が必要で、期待通りの結果が得られることはほとんどない。その結果、問題に対処しようとしても、かえって問題を悪化させることが多い。

断絶の主な原因は、データよりもアプリケーションを優先していることだ。企業はまず、別々のシステムを開発するか、ベンダーから既製のソリューションを購入し、その後、重複して互換性のないストレージやデータベースを作成することで、それらを統一しようとする。

断片化の問題を克服するには、アプリケーションよりもデータを優先するという、抜本的な新しいアプローチが必要だ。企業はまず、データ管理戦略とデータモデルを開発し、新たな障壁を作るのではなく、単一の情報で機能するシステムを構築するか、ソリューションを購入しなければならない。

> ソフトウェアよりもデータの方が重要かもしれない、新しい世界に突入しつつある。
>
> - ティム・オライリー、オライリー・メディア社 CEO

マッキンゼー・グローバル・インスティテュートの調査「建設業再考：生産性向上への道」（2016年）は、建設業がデジタルトランスフォーメーションにおいて他のセクターに遅れをとっていることを示している[32]。同報告書によると、自動化されたデータ管理とデジタルプラットフォームの採用により、生産性を大幅に向上させ、プロセスの不整合に伴う損失を減らすことができる。このデジタルトランスフォーメーションの必要性は、イーガン（英国、1998 年）の報告書[33]でも強調されており、建設における統合プロセスと協調的アプローチの重要な役割を強調している。

その結果、過去 1 万年間、データ管理者にとっての主な問題はデータの不足であったが、データとデータ管理システムが雪崩を打ったことで、ユーザーと管理者はデータの過剰という問題に直面し、法的に正しく質の高い情報を見つけることが難しくなった。

データのサイロ化が進むと、必然的にデータ品質の低下という深刻な問題が生じる 。複数の独立したシステムにより、同じデータが異なるバージョンで存在する可能性があり、しばしば相反する値を持つため、どの情報が適切で信頼できるかを判断する必要のあるユーザーにとって、さらなる困難が生じる。

重複、データ品質の欠如 不統一の結果である。

データ・サイロの問題により、管理者はデータの検索と調整に多大な時間を費やさざるを得ない。品質問題を回避するため、企業は複雑な情報管理構造を構築し、縦割りの管理職がデータの検索、検証、照合を担当する。しかし、このアプローチは官僚主義を増長し、意思決定を遅らせるだけである。データが多ければ多いほど、分析や解釈は難しくなり、特にデータの保存や処理に統一された基準がない場合はなおさらだ。

過去 10 年間に雨後の筍のように増えたソフトウェア・アプリケーションやシステムの多さに伴い、サイロ化や不適切なデータ品質 の問題は、エンド・ユーザーにとってますます大きな意味を持つようになっている。同じデータであっても、異なる値が異なるシステムやアプリケーションに存在することがある（図 2.1-6）。このためエンドユーザーは、数多くあるデータの中からどのバージョンが適切で正しいかを判断することが難しくな

っている。これは分析、ひいては意思決定におけるエラーにつながる。

正しいデータが見つからないという問題に備えるため、企業のマネージャーは、検証マネージャーという何段階もの官僚機構を作る。彼らの仕事は、異種システムの迷路を通り抜けながら、必要なデータを表やレポートの形で素早く見つけ、チェックし、送信できるようにすることである。

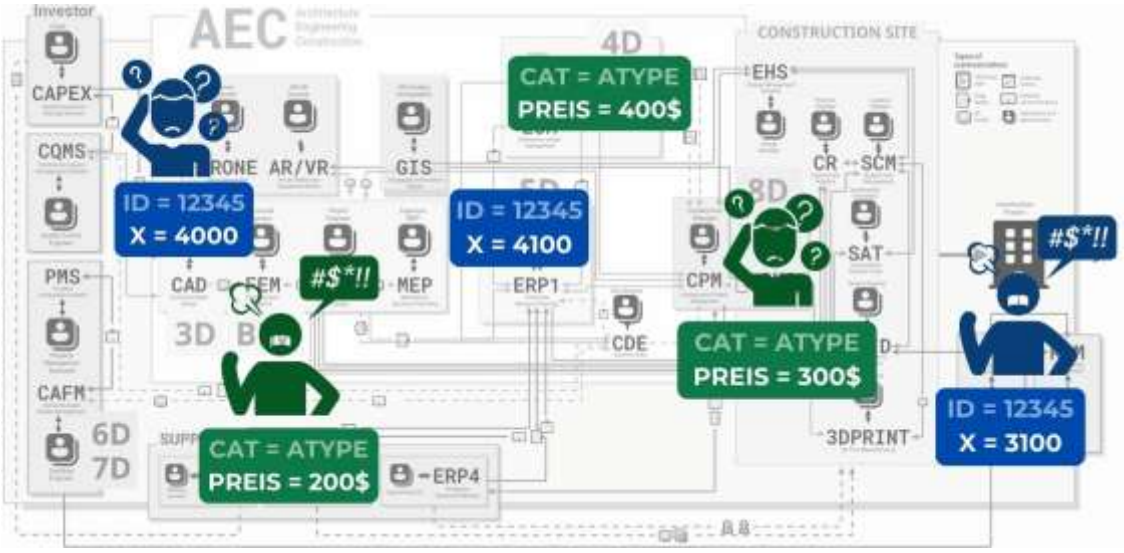


図 2.1-6 適切なデータを見つけようとする場合、管理者は異なるシステム間のデータの品質と法的信頼性を保証しなければならない。

しかし実際には、このモデルは新たな複雑さを生む。データが手作業で管理され、情報が多くの無関係な意思決定に散らばっている場合、意思決定者のピラミッド（図 2.1-7）を通して正確で最新の情報を得ようとする試みはすべてボトルネックになる。

この状況は、デジタル・ソリューションの雪崩によって悪化している。ソフトウェア市場には、有望と思われる新しいツールが氾濫し続けている。しかし、 、明確なデータ管理戦略がなければ、これらのソリューションは統一されたシステムに統合されることはなく、かえって複雑さと重複のレイヤーを増やしてしまう。その結果、企業はプロセスを簡素化するどころか、さらに断片化された混沌とした情報環境に身を置くことになる。

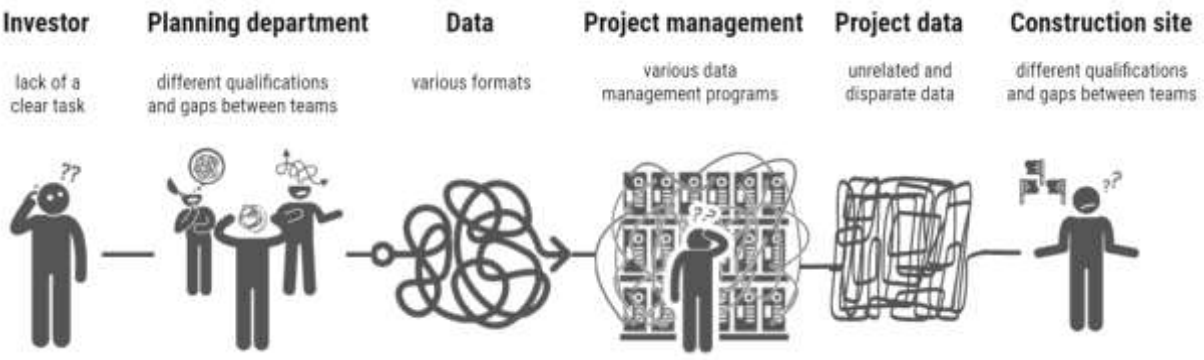


図 2.1-7 システムの複雑さとデータ形式の多様性は、建設プロセスにおける一貫性の喪失につながる。

多数の異種ソリューションの管理に関連するこれらの問題はすべて、遅かれ早かれ企業経営者に重要な気づきをもたらす。本当の理由は、データの質と、組織がそれをどのように作成し、受信し、保存し、使用するかにある。

> 持続可能な成功の鍵は、新しい「魔法の」アプリケーションを追い求めることではなく、社内にデータ文化を構築することにある。つまり、データを戦略的資産として扱い、データの品質、完全性、関連性を組織の全レベルで優先させることである。

質対量のジレンマの解決策は、重複を排除し、矛盾をなくし、情報の流れを統一する統一データ構造の構築にある。このアーキテクチャは、情報に基づいた正確でタイムリーな意思決定を行うための、信頼できる単一のデータソースを提供する。

そうでなければ、いまだによくあることだが、企業は信頼できる事実よりも、HiPPO 専門家の主観的な意見や直感的な評価に依存し続けることになる。伝統的に専門知識が重要な役割を果たす建設業界では、これは特に顕著である。

HiPPO、あるいは意思決定における意見の危険性

従来、建設業界では、重要な決定は経験と主観的な判断に基づいて行われてきた。タイムリーで信頼できるデータがなければ、会社のマネジャーは、客観的な事実ではなく、高給取り社員の直感（HiPPO - Highest Paid Person's Opinion）に頼って、やみくもに行動せざるを得ない（図 2.1-8）。

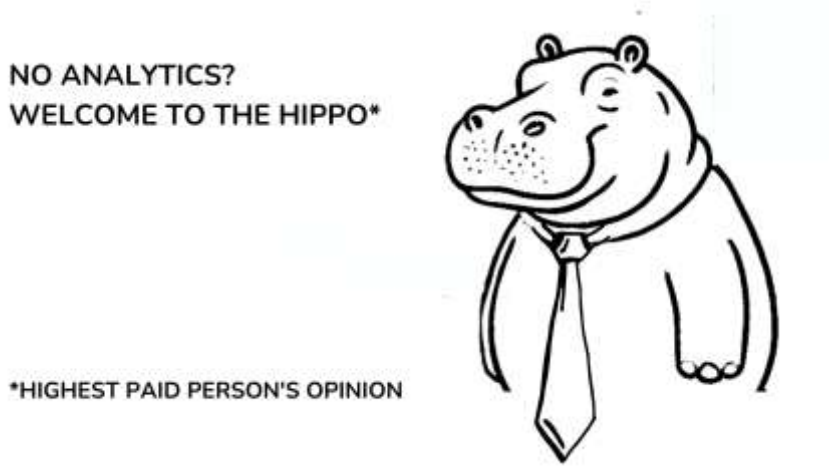


図 2.1-8 アナリティクスがない場合 ビジネスは経験豊富な専門家の主観的な意見に左右される

。

このアプローチは、安定していて変化の遅い環境では正当化されるかもしれないが、デジタルトランスフォーメーションの時代には深刻なリスクとなる。直感や当て推量に基づく決断は歪みを生じやすく、多くの場合、裏付けのない仮説に基づいており、データに反映された複雑な全体像を考慮していない

> 企業内の意思決定レベルで知的な議論として流布しているものは、具体的な何ものにも基づいていないことが多い。企業の成功は、専門家の権威や給与に左右されるべきではなく、データを効果的に扱い、パターンを特定し、情報に基づいた意思決定を行う能力に左右されるべきなのだ。

権威や経験があれば自動的に意思決定が正しいという概念を捨てることが重要だ。データ主導のアプローチは、ゲームチェンジャーである。地位や給与ではなく、データとアナリティクス（）が、今や意思決定の基礎となっている。ビッグデータ、機械学習、ビジュアル分析により、パターンを特定し、当て推量ではなく事実に頼ることができる（図 1.1-4）。

> データがなければ、あなたは意見を持つただの一人に過ぎない[34]。
>
> - W・エドワーズ・デミング、学者、経営コンサルタント

最新のデータ管理手法により、社内の知識の継続性も確保される。明確に記述されたプロセス、自動化、体系的なアプローチにより、効率を損なうことなく、重要な役割を引き継ぐことが可能です。

しかし、データへの盲目的な信頼は深刻なエラーを招くこともある。データそのものは数字の集まりに過ぎない。適切な分析、文脈、パターンを特定する能力がなければ、データには何の価値もなく、プロセスを推進することもできない。成功への鍵は、 、HiPPO の直感とアナリティクスのどちらを選ぶかではなく、異種情報を管理可能で情報に基づいた意思決定に変換するインテリジェントなツールを構築することにある。

> デジタル化された建設環境において、決定的な成功要因となるのは年功序列や序列ではなく、対応力、意思決定の正確さ、資源効率である

データは道具であって、絶対的な真実ではない。データは人間の思考を補完するものであって、それに取って代わるものではない。アナリティクスのメリット（ ）にもかかわらず、データは人間の直感や経験に完全に取って代わることはできない。データの役割は、より正確な情報に基づいた意思決定を支援することである。

競争上の優位性は、単に基準を満たすだけでなく、誰にとっても同じ資源を効率的に活用することで競合他社を凌駕することで達成される。将来、データ・スキルは、かつての読み書きや数学のスキルと同じくらい重要になるだろう。データを分析・解釈できるプロフェッショナルは、より正確な意思決定を行うことができるようになり、個人的な経験のみに頼るプロフェッショナルに取って代わるだろう（図 2.1-9）。

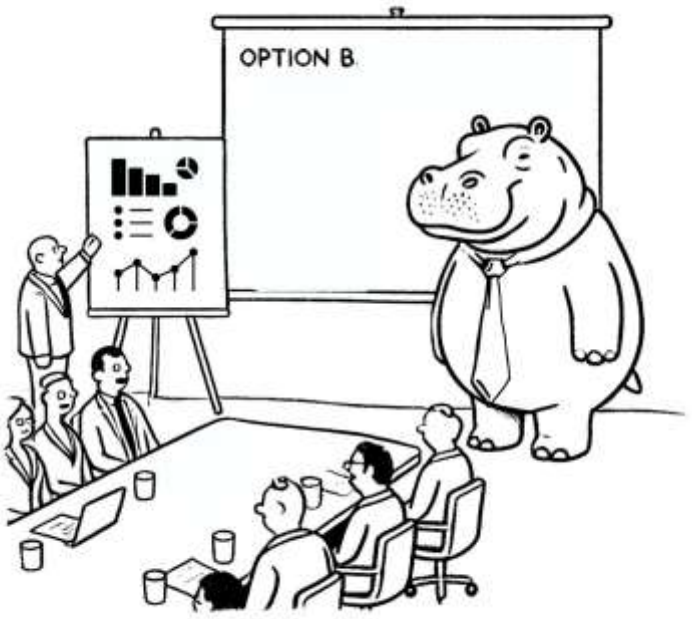


図 2.1-9 意思決定は客観的分析に基づくべきであり、高給取りの意見に基づくべきでない。

マネージャー、スペシャリスト、エンジニアは、 データアナリストとして、プロジェクトの構造、ダイナミク

ス、主要指標を研究する。人材はシステムの要素となり、効率を最大化するために柔軟なデータ主導のカスタマイズが必要となる。

> 不十分なデータを使用した場合の誤差は、データを使用しなかった場合よりもはるかに小さい[35]。
>
> - チャールズ・バベッジ、最初の分析計算機の発明者

ビッグデータの出現と LLM（大規模言語モデル）の導入は、分析方法だけでなく、意思決定の本質をも根本的に変えた。以前は因果関係（なぜ何かが起こったのか-診断アナリティクス ）に焦点が当てられていたが（図 1.1-4）、今日では未来を予測する能力（予測アナリティクス）、そして将来的には、機械学習と AI が意思決定プロセスにおける最適な選択を提案する処方アナリティクスが前面に出てきている。

> SAP™の新しい調査「New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves」2025 [36]によると、経営幹部の 44%は AI のアドバイスに基づいて以前の意思決定を変更したいと考え、 、38%は AI にビジネス上の意思決定を代行してもらいたいと考えている。一方、エグゼクティブの 74%は、友人や家族よりも AI のアドバイスを信頼すると回答しており、55%は、AI 由来の洞察が従来の意思決定手法に取って代わるか、しばしば迂回するような企業、特に年商 50 億ドル以上の組織で働いている。さらに、回答者の 48%がジェネレーティブ AI ツールを日常的に使用しており、そのうち 15%は 1 日に何度も使用している。

LLM や自動化されたデータ管理システムの発達に伴い、新たな課題が生じている。それは、互換性のないフォーマットや異種ソースの混沌の中で、いかにして価値を失うことなく情報を効果的に利用するかということであり、ビジネス・プロセスの複雑さとダイナミクスの増大がそれを補完している。

ビジネス・プロセスの複雑さとダイナミズムの継続的増加

今日の建設業界は、データとプロセスの管理において深刻な課題に直面している。主な課題は、サイロ化した情報システム、過剰な官僚主義、デジタルツール間の統合不足である。こうした課題は、テクノロジーや顧客要件の変化、規制の進化によってビジネス・プロセス自体が複雑化するにつれて、ますます深刻化している。

建設プロジェクトの独自性は、その技術的な特殊性だけでなく、各国の基準や規制要件の違いにも起因する（

図 4.2-10、図 5.1-7）。このため、プロジェクトごとに柔軟で個別化されたアプローチが必要となるが、従来のモジュール式制御システムでは実現が難しい。プロセスが複雑でデータ量も多いため、多くの企業は専門的なソリューションを提供するベンダーに頼っている。しかし、市場は過負荷に陥っており、多くの新興企業が類似した製品を提供し、狭いタスクに集中している。その結果、データ管理への全体的なアプローチが失われがちだ。

> 新しい技術や市場要求の絶え間ない流れに適応することは、競争力にとって重要な要素になりつつある。しかし、既存の独自のアプリケーションやモジュラー・システムは、適応性が低く、変更があれば、建設プロセスの詳細を必ずしも理解していない開発者による、長くてコストのかかる修正が必要になることが多い。

企業は、技術的な遅れの人質となり、革新的な統合アプローチを迅速に実施する代わりに、新しいアップデートを待つことになる。その結果、建設組織の内部構造は、多くの場合、管理職のマルチレベルのネットワークを通じて調整される、相互接続された階層的で、しばしば閉鎖的なシステムの複雑な生態系となっている（図 2.1-10）。

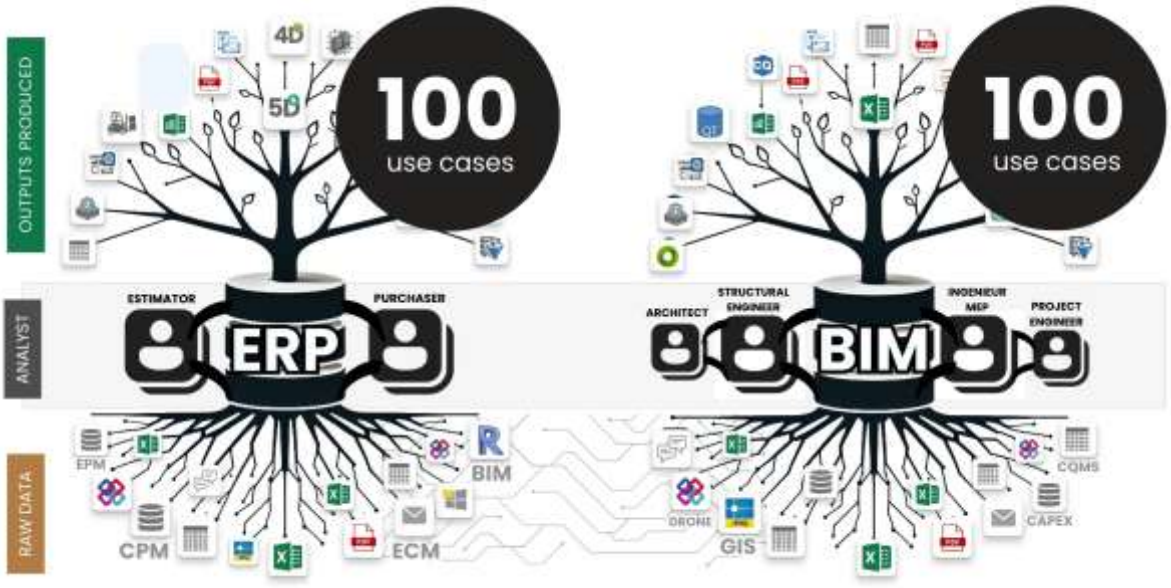


図 2.1-10 企業は相互接続されたシステムで構成されており、その相互接続が自動化を必要とするプロセスを形成している。

> カナダ建設協会と KPMG カナダが 2021 年に実施した調査[37]によると、テクノロジーの導入やデジタル・ソリューションに関して、競合他社と比較して自社が重要な、あるいは異なる立場にあると考える企業はわずか 25%に過ぎない。また、自社のソリューションがデータ主導型であると回答したのは、わずか 23%であった。同時に、調査回答者の大半は、他のさまざまなテクノロジーの利用を純粋に実験的なものと位置づけたり、まったく利用していないことを認めたりしている。

このような技術実験への消極的な姿勢は、ミスが数百万ドルの損失につながりかねない大規模インフラ・プロジェクトにおいて特に顕著である。デジタル・ツインや予測分析（ ）といった最先端の技術でさえ、その有効性ではなく、実際のプロジェクトでの信頼性が証明されていないために抵抗を受けることが多い。

> 世界経済フォーラム（WEF）の報告書「Shaping the Future of Construction」[5]によると、建設における新技術の導入は、技術的な困難だけでなく、顧客側の心理的な障壁にも直面している。[建設における新技術の導入は、技術的な難しさだけでなく、顧客の心理的な障壁にも直面している。多くの顧客は、先進的なソリューションの使用によって、自分たちのプロジェクトが実験場となり、「モルモット」になることを恐れており、予測不可能な結果が追加コストやリスクにつながる可能性がある。

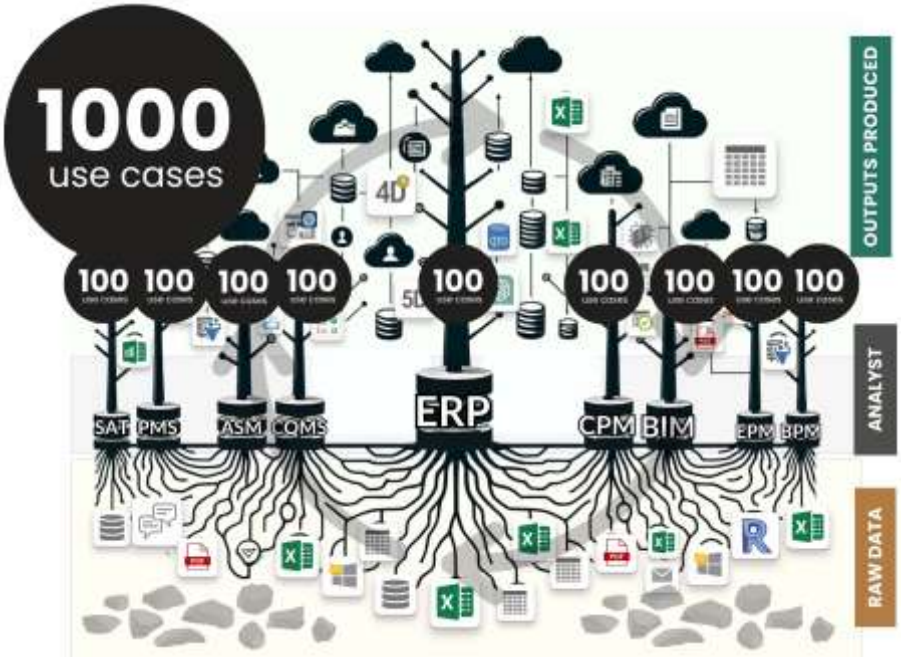


図 2.1-11 各データのユースケースに対して、ソリューション市場はプロセスの最適化と自動化のためのアプリケーションを提供する。

> 建設業界は非常に多様である。プロジェクトによって、、地域の特殊性、法定の分類規制（図 4.2-10）、計算基準（図 5.1-7）などが異なる。そのため、これらすべての要件（）とプロジェクトの特殊性に完璧に適合する独自の汎用アプリケーションやシステムを作成することは、現実的に不可能です。

システムの複雑化とソフトウェア・ベンダーへの依存に対処するため、効果的なデータ管理の鍵は、オープン化と標準化だけでなく、プロセス・アーキテクチャ自体の簡素化であることがますます認識されるようになっている。ビジネス・プロセスの複雑化とダイナミズムの高まりは、データの蓄積から構造化と整理へと優先順位をシフトさせる新たなアプローチを必要としている。このシフトこそが、建設業界の発展における次のステップであり、ソフトウェア・ベンダー優位の時代の終わりと、意味のある情報整理の時代の始まりを告げるものである。

画一的なソリューションの限界と複雑化への脆弱性が認識されたことで、クローズドなプラットフォームとデータの蓄積から、透明性、適応性、構造化された情報の取り扱いへと優先順位がシフトしている。この考え方の転換は、いわゆる「産業革命」のレンズを通して語られる、世界経済とテクノロジーにおけるより広範な変化を反映している。建設業の方向性と将来の方向性を理解するためには、自動化とデジタル化から、パーソナライゼーション、オープンスタンダード、サービスベースのデータモデルまで、第 4 次産業革命と第 5 次産業革命の文脈における建設業の位置づけを検討する必要がある。

第四次産業革命（インダストリー4.0）と建設業における第五次産業革命（インダストリー5.0）。

技術段階と経済段階は、社会と経済の発展段階を説明し分析するために用いられる理論的概念である。研究者や専門家によって解釈が異なることもある。

- 第 4 次産業革命（4IR またはインダストリー4.0）は、情報技術、自動化、デジタル化、グローバル化に関連している。その重要な要素のひとつは、独自のソフトウェア・ソリューション、すなわち特定のタスクや企業向けに設計された特殊化されたデジタル製品の創出である。これらのソリューションは、IT インフラストラクチャの重要な一部となることが多いが、追加修正なしでは拡張性に乏しい。
- 第五次産業革命（5IR）は現在、4IR よりも早い段階で概念化され、発展している。その基本原則には、製品やサービスのパーソナライゼーションの向上が含まれる。5IR は、パーソナライゼーション

、コンサルティング、サービス指向モデルに焦点を当てた、より適応性が高く、柔軟でパーソナライズされた経済活動を目指す動きである。第 5 の経済モードの重要な側面は、意思決定のためのデータの利用であり、これはオープンデータとオープンツールの利用なしには事実上不可能である（図 2.1-12）。



図 2.1-12 第 4 のパターンはソリューションに焦点を当て、第 5 のパターンはパーソナライゼーションとデータに焦点を当てる。

> 10 社または 100 社の組織で使用される建設業界の企業向けのアプリケーションを作成しても、大幅な修正や拡張を行わない限り、他の企業、地域、国への拡張が成功する保証はありません。各組織には固有のプロセス、要件、条件があり、個別対応が必要になる可能性があるからだ。

今日すでに、技術的ソリューションの統合を成功させるためには、各プロセス、プロジェクト、企業に対して深く個別化されたアプローチが必要であることを理解することが重要である。つまり、普遍的なフレームワークやツール、プログラムが開発された後でも、特定の企業やプロジェクトに固有の要件（ ）や条件を満たすために、詳細な適合やカスタマイズが必要になるということだ。

> PwC のレポート「第 5 次産業革命の解読」[38]によると[38]によると、今年、さまざまな業界の上級管理職の約 50%が、高度なテクノロジーと人間の専門知識の統合に依存している。このアプローチにより、製品設計や顧客要求の変化に迅速に対応し、パーソナライズされた生産を実現することができる。

各プロセスは、独自の機能やアプリケーションを開発する必要があり、グローバルな建設業界の規模とプロジェクトの多様性を考慮すると、膨大な数のビジネスケースが存在することになり、その都度、独自のパイプライン （図 2.1-13）ロジックを表現することになる。このようなケースには、それぞれ固有の特徴があり、カス

タマイズされたアプロ ーチが必要です。機械学習とタイタニック・データセットの解析の章（図 9.2-9）では、同じ解析問題に対する様々な可能な解決策を、様々なアプローチで詳しく見ていく。

> デジタルプロセスの文脈におけるパイプライン とは、プロジェクトのライフサイクルのさまざまな段階を通じて、データと作業の自動化または構造化された流れを可能にする一連の活動、プロセス、ツールのことである。

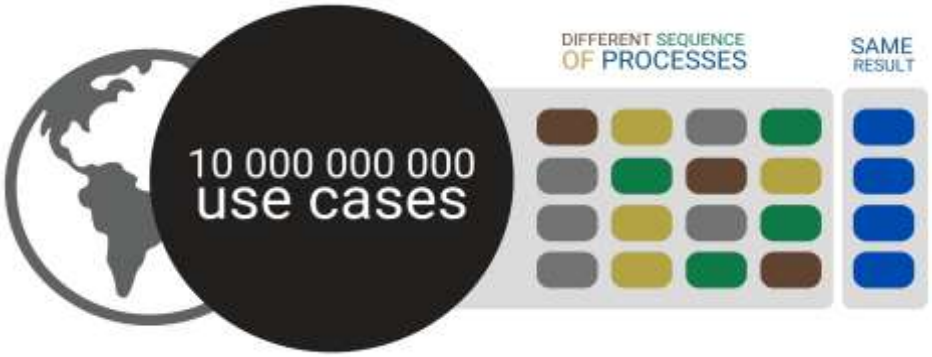


図 2.1-13 ビジネスケースの個別性と多様性は、スケーラブルなクローズド・プラットフォームやツールの作成を不可能にしている。

デジタルトランスフォーメーションの影響で、私たちの生活はすでに様々な形で変化しており、今日、建設業界の経済発展における新たなステージの到来について語ることができる。この "新しい経済 "では、競争は異なるルールに従って組織化される。公共の知識やオープンデータ、要求される製品やサービスに効率的に変えることができる者は、第 5 次産業革命の条件下で重要な優位性を獲得する。

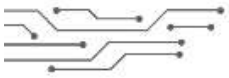
> 経済学者のケイト・マスカスは、著書『私的権利と公的問題：21 世紀のグローバル知的財産権経済』[39]の中で次のように述べている。2012 年 [39]、「私たちはグローバルな知識経済の中に生きており、未来は科学的発見を商品に変える方法を知っている人々のものである」と。

第 5 の経済モードへの移行は、クローズドな IT ソリューションからオープンな標準とプラットフォームへの焦点の移行を意味する。企業は従来のソフトウェア製品から、専有技術ではなくデータが主な資産となるサービス指向モデルに移行し始めるだろう。

> ハーバード・ビジネス・スクール 2024 の調査[40]は、オープンソース ソフトウェア（OSS、Open Source Software）の巨大な経済価値を示している。この調査によると、OSS は全ソフトウェアコードの 96%に存在し、一部の商用ソフトウェアは 99.9%が OSS コンポーネントで構成されている。OSS がなければ、企業はソフトウェアに 3.5 倍のコストをかけることになる。

> 世界的なトレンドに従って、企業のエコシステムを構築することは、徐々に第 5 の経済パラダイムへと移行していくだろう。そこでは、データ中心のアナリティクスとコンサルティング・サービスが、厳密に定義された利用シナリオを持つ孤立したクローズド・ソリューションよりも優先されるようになるだろう。

デジタル化の時代は、業界のパワーバランスを変えるだろう。ベンダーのソリューションに依存する代わりに、企業はデータを効果的に活用する能力を競争力の基盤とする。その結果、建設業界は、レガシーな硬直したシステムから、オープンスタンダードと相互運用可能なツールがプロジェクト管理の基盤となる、柔軟で適応力のあるエコシステムへと移行する。アプリケーション・ベンダー優位の時代の終焉は、クローズド・ソース・コードや特殊なコネクターの所有によってではなく、データを戦略的優位性に変える能力によって価値が定義される新しい環境を生み出すだろう。



章 2.2.

混沌を秩序に変え、複雑さを軽減する

生産性向上の障害となる冗長なコードと閉じたシステム

過去数十年間、IT 分野の技術革新は主にソフトウェア・ベンダーによって推進されてきた。ベンダーは開発の方向性を決め、企業がどの技術を採用し、どの技術を取り残すべきかを決定してきた。サイロ化されたソリューションから一元化されたデータベースや統合システムへと移行する時代には、ベンダーはライセンス製品を推進し、アクセスや拡張性のコントロールを提供した。その後、クラウド技術と SaaS（Software as a Service）モデルの登場により、このコントロールはサブスクリプション・モデルへと進化し、ユーザーはデジタル・サービスの忠実な顧客として定着した。

このアプローチは、パラドックスを生んでいる。前例のない量のプログラム・コードが作成されているにもかかわらず、実際に使われているのはそのごく一部なのだ。同じ企業内であっても、同じビジネス・プロセスが何十、何百ものプログラムにさまざまな方法で記述され、複製されているからだ。同時に、開発コストはすでに支払われており、これらのコストは回収不可能である。にもかかわらず、業界はこのサイクルを繰り返し、エンドユーザーにとって最小限の付加価値しかない新製品を作り続けている。

> 国防調達大学（DAU）のソフトウェア開発コスト見積もりガイド[41]によると、ソフトウェア開発コストは、システムの複雑さや選択した技術など、いくつかの要因によって大きく変動する。歴史的に、2008 年の開発コストは、ソースコード 1 行（SLOC）あたり約 100 ドルであったが、保守コストは SLOC あたり 4,000 ドルまで上昇する可能性がある。

CAD アプリケーションのコンポーネントのひとつであるジオメトリ・コアだけでも、数千万行のコードを持つことがある（図 6.1-5）。同じような状況は ERP システムにも見られる（図 5.4-4）。この複雑さについての議論は、本書の第 5 部で再び取り上げる。しかし、よく見てみると、このコードの多くは付加価値を与えるものではなく、データベース、API、ユーザー・インターフェース、システム内の他のテーブルの間で機械的にデータを移動させる「郵便配達人」の役割を果たしているに過ぎないことがわかる。いわゆるビジネス・ロジックの重要性についての俗説にもかかわらず、厳しい現実はもっと平凡だ。現代のコード・ベースは、意思決定やビジネス効率に影響を与えることなく、テーブルやコンポーネント間のデータ転送を確実にすることだけが目的の、時代遅れのテンプレート・ブロック（レガシー・コード）でいっぱいなのだ。

その結果、さまざまなソースからのデータを処理するクローズド・ソリューションは、必然的に混乱した「ス

パゲッティ・エコシステム」になってしまう。これらの複雑に絡み合ったシステムは、半定常的に働く管理者の軍隊によってのみ処理される。このようなデータ管理の組織は、リソースの面で非効率的であるだけでなく、ビジネス・プロセスに致命的な脆弱性を生み出し、この技術的迷路の仕組みを理解している狭い範囲の専門家に依存することになる。

コード量の継続的な増加、アプリケーション数の増加、ベンダーが提供するコンセプトの複雑化は、当然の結果として、建設業界における IT エコシステムの複雑化を招いた。このため、業界におけるアプリケーションの数を増やすことによるデジタル化の実践は、非効率的なものとなっている。ユーザーのニーズを十分に考慮せずに作成されたソフトウェア製品は、実装とサポートに多大なリソースを必要とするものの、期待された利益をもたらさないことが多い。

> マッキンゼーの研究「建設業の生産性向上」[42]によると、過去 20 年間、建設業における世界の労働生産性の伸びは、世界経済全体の 2.8%、製造業の 3.6%に対し、年平均 1%に過ぎなかった。[42]によれば、過去 20 年間、世界経済全体の労働生産性の伸びが 2.8%、製造業の伸びが 3.6%であったのに対し、建設業の労働生産性の伸びは年平均 1%にとどまっている。米国では、労働者一人当たりの建設労働生産性は、1960 年代から半減している[43]。

システムの複雑化、孤立化、データの閉鎖化により、専門家間のコミュニケーションが損なわれ、建設業界は最も効率の悪い業界のひとつとなっている（図 2.2-1）。2040 年には 22 兆ドルに達するため、大幅な効率化が

必要となる。

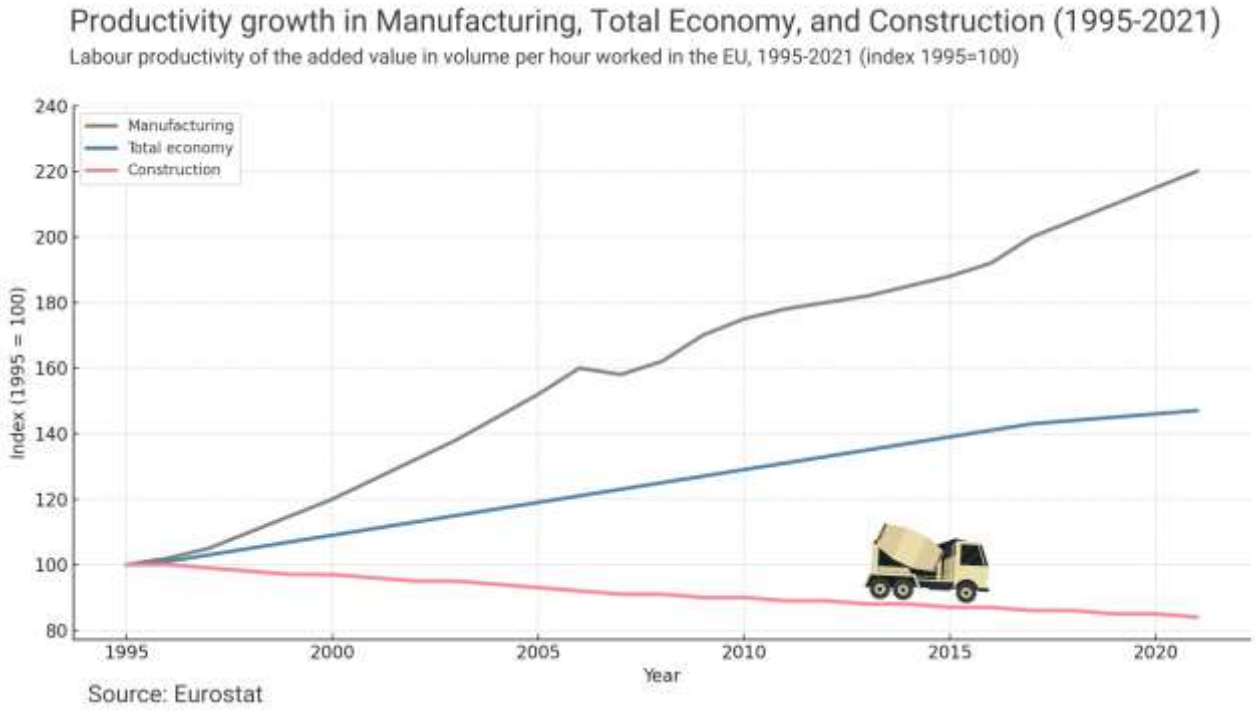


図 2.2-1 閉鎖的で複雑なデータ、そしてその結果としての専門家間の不十分なコミュニケーションは、建設業界を経済の中で最も効率の悪い部門のひとつへと導いた（[44]、[45]に基づく）。

> マッキンゼー（2024 年）の調査「建設業の生産性確保はもはやオプションではない」で強調されているように、資源不足の深刻化と業界の成長率倍増の推進により、建設業はもはや現在の生産性レベルにとどまっている余裕はない[44]。世界の建設コストは、2023 年の 13 兆ドルから、10 年後までにははるかに高いレベルにまで上昇すると予測されており、効率性の問題は、単に関連性があるだけでなく、極めて重要なものとなっている。

効率性を向上させる重要な方法のひとつは、アプリケーション構造とデータ・エコシステム・アーキテクチャの必然的な統一と簡素化である。合理化へのこのアプローチは、企業システムに長年蓄積された抽象化の冗長な層や不必要な複雑性を排除する。

サイロから単一のデータウェアハウスへ

組織がデータを蓄積すればするほど、そこから真の価値を引き出すのは難しくなる。孤立したサイロに情報が断片的に保管されるため、現代企業のビジネス・プロセスは、何千もの倉庫に保管された材料から高層ビルを

建てようとする建設業者のようなものとなっている。過剰な情報は、法的に適切な情報へのアクセスを困難にするだけでなく、意思決定を遅らせる。

> 各タスクやプロセスは別々のテーブルやデータベースにハードワイヤードされており、システム間のデータ交換には複雑な統合が必要である。あるシステムでのエラーや不整合は、他のシステムで連鎖的な障害を引き起こす可能性がある。不正確な値、更新の遅れ、重複した情報により、従業員は手作業によるデータの照合や調整に多大な時間を費やさざるを得なくなる。その結果、組織は、プロセスの開発や最適化よりも、分断の結果への対処に多くの時間を費やすことになる

この問題は普遍的なものである。ある企業はカオスと格闘し続け、ある企業は統合に解決策を見出す。それは、タスク、プロジェクト、オブジェクトに関連するあらゆるエンティティを保存できる、ひとつの大きなテーブルだと考えてほしい。何十ものバラバラのテーブルやフォーマットの代わりに、一つのまとまったリポジトリ（図 2.2-2）が出現し、それが可能になる：

- データ損失を最小限に抑える；
- 常に情報を調和させる必要がなくなる；
- データの可用性と質を向上させる；
- 分析処理と機械学習の簡素化

データを共通標準にするということは、情報源に関係なく、情報が統一された機械可読形式に変換されることを意味する。このようなデータの整理によって、データの完全性をチェックし、リアルタイムで分析し、経営上の意思決定に迅速に利用することができる。

統合ストレージ・システムの概念と、アナリティクスや機械学習への応用については、「ビッグデータ・ストレージと機械学習」の章で詳しく説明する。データモデリングと構造化については、「データを構造化する」

と「標準がゲームを変える：ランダムファイルから精巧なデータモデルへ」の章で詳しく説明する。

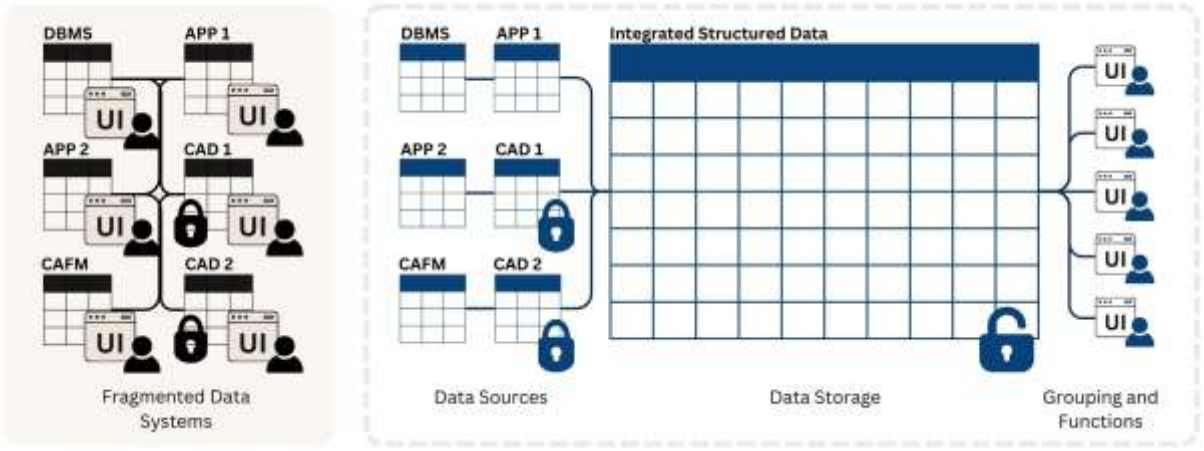


図 2.2-2 データ統合はサイロをなくし、情報の可用性を高め、ビジネスプロセスを最適化する。

データが構造化され、マージされたら、次の論理的なステップはそれを検証することである。単一の統合リポジトリを使えば、このプロセスは大幅に簡素化される。矛盾した複数のスキーマ、重複した構造、テーブル間の複雑なリレーションシップはもう必要ない。すべての情報が単一のデータモデルに整合されるため、内部的な不整合がなくなり、検証プロセスがスピードアップする。 を検証し、データ品質を保証する は、すべてのビジネス・プロセスの基礎となる側面である。

最終段階で、データはグループ化され、フィルターにかけられ、分析される。集計（加算、乗算）、表、列、行間の計算など、さまざまな関数が適用される（図 2.2-4）。データを扱うことは、収集、構造化、検証、変換、分析処理、そして最終的なアプリケーションへのオフロードという一連のステップとなり、そこで情報は実用的な問題を解決するために使用される。このようなシナリオの構築、ステップの自動化、処理フローの構築については、ETL -プロセスとデータパイプラインアプローチの章で詳しく説明する。

したがって、デジタルトランスフォーメーションとは、単に情報の取り扱いを簡素化することではない。データ管理における過剰な複雑性を排除し、混沌から予測可能性へ、複数のシステムから管理可能なプロセスへと移行することである。アーキテクチャの複雑性が低ければ低いほど、それをサポートするために必要なコードは少なくなる。そして将来的には、コードというものは完全に姿を消し、独立してデータを分析し、体系化し、変換するインテリジェント・エージェントに取って代わられるかもしれない。

統合ストレージシステムが AI への移行を可能に エージェント

データやシステムが複雑でなければないほど、コードを書いたり保守したりする必要は少なくなる。そして、開発を節約する最も簡単な方法は、コードを完全に取り除き、データに置き換えることである。アプリケーションのコード開発がコードからデータモデルに移行するとき、必然的にデータ中心（データ駆動型 ）アプローチにシフトする。なぜなら、これらのコンセプトの背後にはまったく異なる考え方があるからだ。

> データを中心にして仕事をすることを選ぶと、その役割が違って見えてくる。データはもはやアプリケーションのための単なる「原料」ではなく、アーキテクチャ、ロジック、インタラクションを構築するための基盤なのだ。

データ管理に対する従来のアプローチは、通常、アプリケーション・レベルから始まり、構築においては、煩雑な官僚主義的システムに似ている。デジタル技術の発展により、ますます多くの企業がミニマリズムの原則、つまり本当に必要なものだけを保存し、使用するという原則に切り替えざるを得なくなるだろう。

最小化の論理は、ベンダーにも取り入れられている。データの保存と処理を簡素化するため、ユーザーの作業はオフラインのアプリケーションやツールから、クラウドサービスやいわゆる SaaS ソリューションへと移行している。

> SaaS のコンセプト （Software as a Service、「サービスとしてのソフトウェア」）は、現代の IT インフラにおける重要なトレンドのひとつであり、ユーザーは自分のコンピューターにソフトウェアをインストールしたり保守したりすることなく、インターネット経由でアプリケーションにアクセスすることができる。

一方では、SaaS スケーリング、バージョン管理を容易にし、サポートとメンテナンスのコストを削減したが、他方では、特定のアプリケーションのロジックに依存するだけでなく、ユーザーをプロバイダーのクラウドインフラに完全に依存させることになった。サービスがダウンすると、データやビジネス・プロセスへのアクセスが一時的に、あるいは永久にブロックされる可能性がある。さらに、SaaS アプリケーションで作業する際のすべてのユーザーデータはプロバイダーのサーバーに保存されるため、セキュリティや法令遵守のリスクが生じる。また、料金や利用規約の変更により、コストが増加したり、早急な移行が必要になったりすることもある。

AI 、LLM - エージェントとデータ中心アプローチの発展は、従来の形態のアプリケーションと SaaS の実行の将

来に疑問を投げかけている。これまでアプリケーションやサービスはビジネスロジックを管理し、データを処理する必要があったが、AI エージェントの出現により、これらの機能はデータを直接扱うインテリジェントなシステムに移行する可能性がある。

ハイブリッド・アーキテクチャが IT 部門や経営レベルで議論されるようになっているのはこのためで、AI エージェントやオンプレミス・ソリューションがクラウド・サービスを補完し、SaaS プラットフォームへの依存を軽減する。

> 我々のアプローチは、従来のビジネスアプリケーションや SaaS（）アプリケーションは、エージェントの時代に劇的に変化する可能性があることを認識している。これらのアプリケーションは、基本的にビジネスロジックを持つ CRUD [作成、読み込み、更新、削除]データベースである。しかし将来的には、このロジックは AI エージェント [46]に引き継がれることになる。
>
> - 2024 年、マイクロソフト CEO、サティア・ナデラ。

データ中心のアプローチと AI/LLM エージェントの活用は、冗長なプロセスを削減し、従業員の作業負荷を軽減することができる。データが適切に整理されれば、分析、視覚化、意思決定への応用が容易になる。終わりのないレポートやチェックの代わりに、スペシャリストは数回クリックするだけで、あるいは LLM エージェントの助けを借りて、準備の整ったドキュメントやダッシュボードの形で自動的に最新の情報にアクセスできるようになる。

人工知能ツール （AI ）や LLM チャットによってデータ操作を支援する。近年、伝統的な CRUD 操作 (create, read, update, delete)から、大規模言語モデル(LLM)をデータ管理に利用する傾向がある。LLM は自然言語を解釈し、適切なデータベースクエリを自動生成することができるため、データ管理システムとのインタラクションが簡素化される（図 2.2-3）。

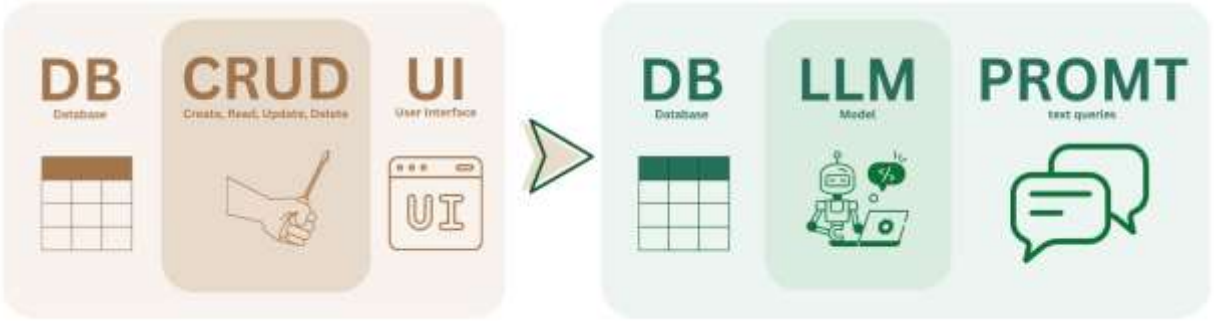


図 2.2-3 ストレージとデータベースのソリューションに取って代わり、統合され、従来のアプリケーションや CRUD -操作に徐々に取って代わる。

> 今後 3〜6 カ月で、AI がコードの 90%を書くようになり、12 カ月後にはほぼすべてのコードが AI によって生成される可能性がある[47]。
>
> - LLM Anthropic の CEO、ダリオ・アモデイ、2025 年 3 月。

AI 開発ツール（GitHub Copilot など）の急速な発展にもかかわらず、2025 年においても、開発者がこのプロセスにおいて重要な役割を果たしている。AI、ユーザーの クエリを自動的に解釈したり、SQL や Pandas クエリを生成したり（これについては次の章で詳しく説明する）、データを分析するコードを書いたりするなど、ますます便利なアシスタントになりつつある。このように、人工知能は徐々に従来のアプリケーション・ユーザー・インターフェースに取って代わりつつある。

言語モデルなどの人工知能モデル の普及は、ハイブリッド・アーキテクチャーの開発を促進するだろう。クラウド・ソリューションや SaaS 製品を完全に放棄するのではなく、クラウド・サービスとローカル・データ管理システムの統合が進むかもしれない。例えば、連携学習は、機密データをクラウドに移すことなく、強力な AI モデルを可能にする。このようにして、企業は高度なテクノロジーへのアクセスを得ながら、データのコン

トロールを維持することができる。

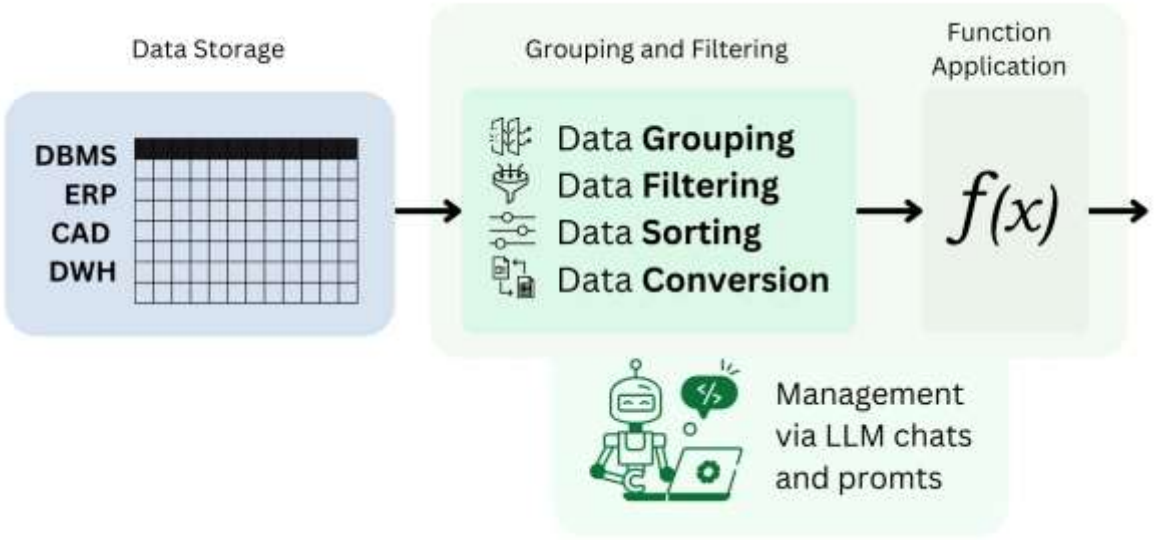


図 2.2-4 グループ化、フィルタリング、並べ替えの基本操作と、それに続く関数適用は、LLM チャットで処理される。

建設業界の未来は、効率的で安全なデータ管理システムを構築するために、オンプレミス・ソリューション、クラウド・パワー、インテリジェント・モデルを組み合わせることに基づいている。LLM は、深い専門知識を持たないユーザーでも、自然言語でクエリーを作成することで、データベースやデータウェアハウスと対話することを可能にする。LLM と AI エージェントとその仕組みについては、「LLM エージェントと構造化データフォーマット」の章で詳しく説明する。

適切に整理されたデータと、シンプルで使いやすい LLM 対応分析ツールは、情報の取り扱いを容易にするだけでなく、エラーを最小限に抑え、効率を高め、プロセスを自動化するのに役立つ。

データ収集から意思決定まで：自動化への道

本書の後の部分では、専門家同士がどのように相互作用し、データがどのように意思決定、自動化、業務効率の基礎となるかを詳しく見ていく。図 2.2-5 は、データ中心アプローチにおけるデータ処理ステップのシーケンス を示す図の例である。この図は継続的改善パイプライン（Continuous Improvement Pipeline） を示している。

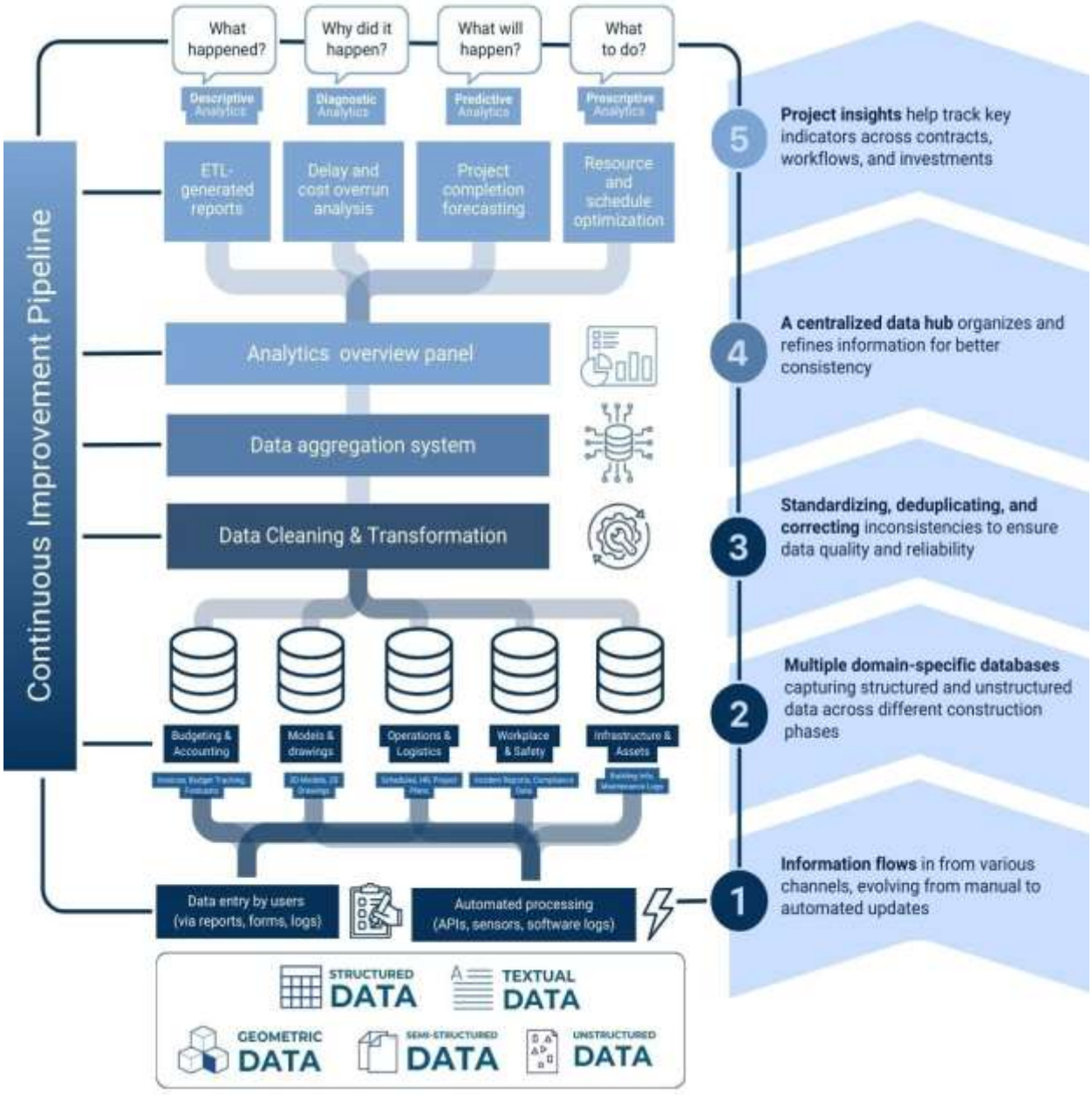


図 2.2-5 継続的なデータ改善パイプラインの一例：建設プロジェクトにおけるデータ処理と分析の流れ。

中堅企業のビジネスプロセスを記述するシステムは、マルチレベルの原則に基づいて構築されている。データ収集、クリーニング、集計、分析処理、結果に基づく意思決定などである。本書の後半では、これらのすべての段階について、理論的な文脈と実践的な例を通じて、：

- 最初のレベルでは、データ入力が行われる（図 3.1-1）。情報は、手作業（レポート、フォーム、ログ

）でも、自動化された形（API、センサー、ソフトウェア・システムから）でも受け取られる。データの構造は、幾何学的、テキスト、非構造化などさまざまである。この段階では、情報の流れの標準化、構造化、統一化が必要である。

- 次のレベルはデータ処理と変換である。これには、クリーニング、重複の除去、エラーの修正、さらなる分析に向けた情報の準備などのプロセスが含まれる（図 4.2-5）。アナリティクス（）の質はデータの清浄度と正確さに直接依存するため、この段階は極めて重要である。
- そしてデータは、予算・会計、モデル・図面、ロジスティクス、セキュリティ、インフラといった機能分野ごとに分けられた専用のテーブル、データフレーム、データベースに流れ込む。この分割により、情報へのアクセスや相互分析が容易になります。
- そしてデータは集計され、分析ダッシュボード（ショーケース）に表示されます。ここでは、記述的分析、診断的分析、予測的分析、処方的分析の手法が適用されます（ ）。これにより、「何が起こったのか」「なぜ起こったのか」「将来何が起こるのか」「どのようなアクションをとる必要があるのか」という重要な質問（図 1.1-4）に答えることができる。例えば、遅延の特定、プロジェクトの完了予測、リソースの最適化などである。
- 最後に、最後のレベルは、契約履行状況の監視、投資管理、ビジネス・プロセスの改善に役立つ分析的結論と主要指標を生成する（図 7.4-2）。この情報は、意思決定と企業の発展戦略の基礎となる。

同様に、データは収集されたものから戦略的経営に活用されるものへと変化する。本書の以下のパートでは、データの種類、データ処理技術、アナリティクス・ツール（ ）、そしてこれらのアプローチが建設業界でどのように活用されているかという実例に焦点を当てながら、各ステージについて詳しく見ていく。

次のステップ：混乱を管理可能なシステムに変える

このパートでは、情報のサイロ化がもたらす課題を探り、システムの過度な複雑性が業績に与える影響について考察した。第 4 次産業革命から、アプリケーションではなくデータが中心となる第 5 次産業革命への移行を分析した。サイロ化した情報システムが知識共有の障壁となっていること、IT 環境の複雑化が続いていることが、建設業界の生産性を低下させ、イノベーションを阻害していることを明らかにした。

この部分を要約すると、考察したアプローチを日常業務に応用するのに役立つ主な実践的ステップを強調する価値がある：

- 情報の可視化
    - ☐ データソースのビジュアルマップを作成する（Miro、Figma、Canva）。

- ☐ 仕事で使用しているシステムやアプリケーションをこのマップに追加する
- ☐ 重複する可能性のある機能や冗長なソリューションを特定する
- ☐ システム間の伝送中にデータの損失や破損が発生する可能性のある重要なポイントを特定する。

■ パーソナライズされたデータ管理の実践

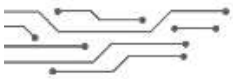
- ☐ プロセスにおける重要な資産として、アプリケーションからデータへと焦点を移す
- ☐ データソースと処理方法を文書化し、透明性を確保する。
- ☐ データの質を評価し、改善するメカニズムを開発する
- ☐ データは一度入力すれば繰り返し使用されるように努める - これが効率的なプロセス編成の基本である

■ チーム内でデータ中心（データドリブン）のアプローチを推進する。

- ☐ ピアツーピアのデータ交換に標準化された統一フォーマットの使用を提案する。
- ☐ データ品質と可用性に関する問題をチームミーティングで定期的に提起する。
- ☐ オープンソースを知る 問題解決のために使用しているツールの代替となる。

小さく始める - 業務にとって重要な特定のプロセスやデータセットを 1 つ選び、データ中心のアプローチを適用し、ツールからデータに焦点を移す。1 回のパイロットで成功を収めることで、実地経験が得られるだけでなく、新しい方法論の利点をチームに明確に示すことができる。これらのステップのほとんどを完了する際、疑問があれば、最新の LLM に説明と支援を求めることができる。

本書の次のパートでは、データの構造化とハーモナイゼーションのテクニックをより詳しく見ていき、異種情報を統合するための実践的なアプローチを探っていく。特に、建設業界のデジタルトランスフォーメーションにおいて重要な役割を果たす、異種サイロから統一されたデータエコシステムへの移行に注目する。





III 部

建設ビジネスプロセスにおけるデータフレームワーク

第 3 部では、建設におけるデータの類型とその効果的な整理方法について包括的に理解する。建設プロジェクトにおける構造化データ、非構造化データ、半構造化データ、テキストデータ、幾何学的データの特徴や取り扱い方について分析する。業界で使用される様々なシステム間で情報を交換するための最新の保存形式とプロトコルについて概説する。CAD (BIM) データを統合する方法など、マルチフォーマットデータを単一の構造化環境に変換するための実用的なツールと技術について説明する。建設計算の精度に不可欠な標準化と検証を通じて、 データの品質を確保するためのアプローチを提案する。建設業界における典型的な問題を解決するためのコード例を用いた最新技術（Python Pandas, LLM -models）の使用に関する実践的な側面が詳細に分析されている。情報管理アプローチの調整と標準化のための組織構造として、コンピテンスセンター（CoE ）を設立することの価値を立証する。

章 3.1.

建築におけるデータ型

建設業界で最も重要なデータタイプ

現代の建設業界では、 企業のシステム、アプリケーショ ン、データウェアハウスは、様々な種類と形式の情報とデータで活発に埋め尽くされている （図 3.1-1）。建設業を営む現代企業の情報状況を形成する主なデータの種類を詳しく見てみましょう：

- 構造化データ: このデータは明確な組織構造を持っている。例えば、エクセルのスプレッドシート やリレーショナル・データベースなど。
- 非構造化データ: 厳密なルールに従って整理されていない情報である。このようなデータの例としては、テキスト、ビデオ、写真、録音などがある。
- 疎構造化データ: これらのデータは、構造化データと非構造化データの中間的な位置を占めている。これらのデータには構造の要素が含まれているが、その構造は必ずしも明確ではなく、また異なるスキーマによって記述されていることも多い。建設における半構造化データの例としては、技術仕様書、プロジェクト文書、進捗報告書などがある。
- テキストデータ: 電子メール、会議やアポイントメントの記録など、口頭や書面によるコミュニケーションから得られるものを含む。
- 幾何学データ: このデータは、専門家が視覚化、体積値の確認、衝突チェックのためにプロジェクト要素の幾何学データを作成する CAD プログラム から得られます。

幾何学的データとテキスト（英数字）データは別個のカテゴリーではなく、3 種類のデータすべてに存在しうることに注意することが重要である。例えば、幾何学データ は、構造化データ（パラメトリック CAD フォーマット）と非構造化データ（スキャンした図面）の両方に含まれる可能性がある。テキストデータも同様に、データベース（構造化データ）に整理されることもあれば、明確な構造を持たない文書として存在することもある。

建設会社における各種データは、会社の情報資産のモザイクを構成するユニークな要素です。建設現場の画像や会議の録音などの非構造化データから、テーブルやデータベースなどの構造化レコードに至るまで

、各要素は企業の情報景観を形成する上で重要な役割を果たしている。

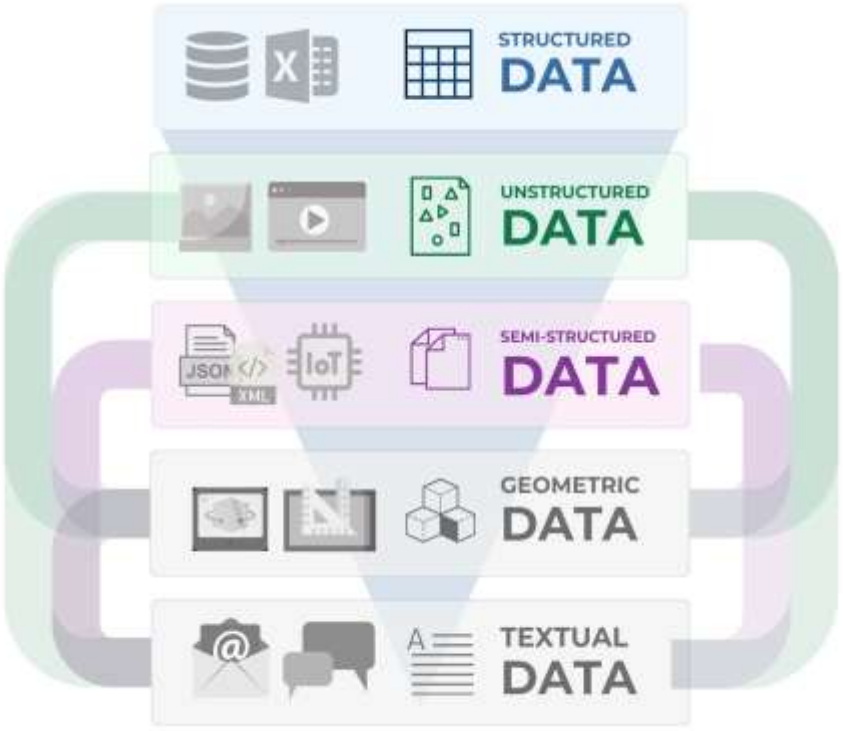


図 3.1-1 エンジニアとデータ管理者は、建設業界で使用されるあらゆる種類のデータを扱うことを学ばなければならない。

以下は、建設現場で使用されるシステムと関連データタイプのサンプルリストである（図 3.1-2）：

- ERP (エンタープライズ・リソース・プランニング) - 企業のリソースを管理し、様々なビジネスプロセスを統合するために、一般的に構造化されたデータを扱う。
- CAD （コンピューター支援設計）と BIM （ビルディング・インフォメーション・モデリング）の組み合わせ - 幾何学的および半構造化データを使用して建設プロジェクトを設計・モデリングし、設計段階における情報の正確性と一貫性を確保する。
- GIS (地理情報システム) - 幾何学的データや構造化データを扱い、地図データや空間的関係を作成・分析する。
- RFID (Radio-Frequency Identification) - 半構造化データを使用し、無線周波数識別を使用して建設現場の資材や設備を効率的に追跡する。
- ECM （Engineering Content Management）は、技術図面や設計文書などの半構造化データや非構造化データを含む、エンジニアリングデータや文書を管理するためのシステムである。

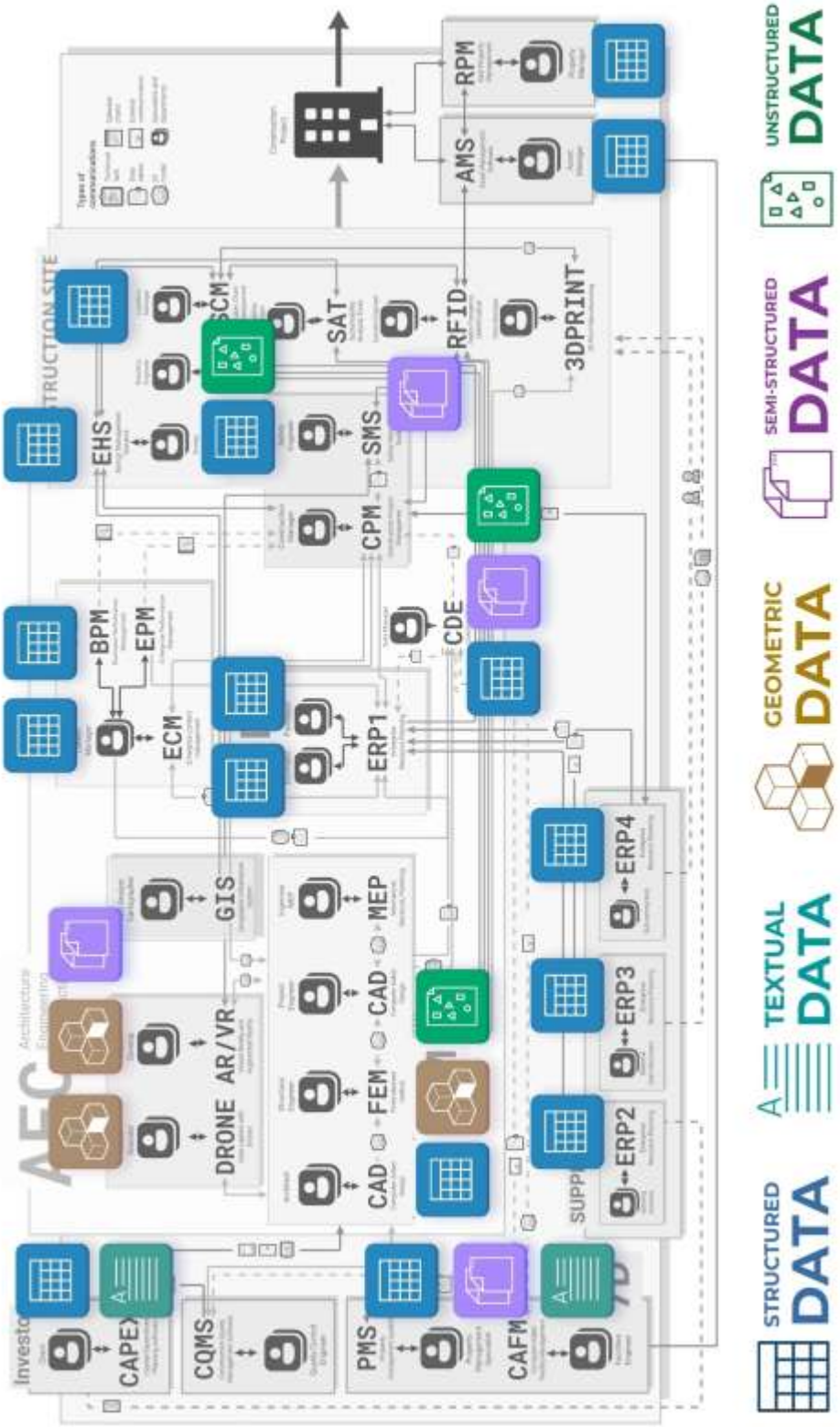


図 3.1-2 異なるフォーマットとデータが異なるシステムに入力されるため、複雑な統合に適した形に変換する必要がある。

これらのシステムおよび同社の他の多くのシステムは、構造化された表データから複雑な形状モデルまで、幅広いデータを管理し、設計、計画、建設管理プロセスにおける統合コラボレーションを提供する。

簡略化した対話の例（図 3.1-3）では、建設プロジェクトの専門家の間で、さまざまな種類のデータが交換されている：

- Architect: "クライアントの要望を考慮し、屋上に客席を設けました。新しいデザインをご覧ください。
- 構造エンジニア：「プロジェクトを受け取りました。新しいレクリエーションエリアの屋根の耐荷重を計算しています」（構造化データおよび半構造化データ - 計算表）。
- 購買担当者：「レクリエーションエリア用の資材を購入するため、その仕様と数量が必要です」（テキストデータおよび半構造化データ - リストと仕様書）。
- 安全衛生エンジニア：「新しいエリアのデータを受け取った。リスクを評価し、安全計画を更新している」（半構造化データ-文書と計画）。
- BIM のスペシャリスト -モデリング：「作業ドキュメントを調整するためにプロジェクトモデル全体に変更を加える」（幾何学データおよび半構造化データ）。
- プロジェクトマネージャー：「新しい休憩所を作業スケジュールに組み込んでいます。プロジェクト管理システムのスケジュールとリソースを更新しています。
- 施設メンテナンス・スペシャリスト（FM）：「レクリエーション・エリアの将来のメンテナンスのためのデータを作成し、資産管理システムに入力する」（構造化データおよび半構造化データ-指示書およびメンテナンス計画書）。

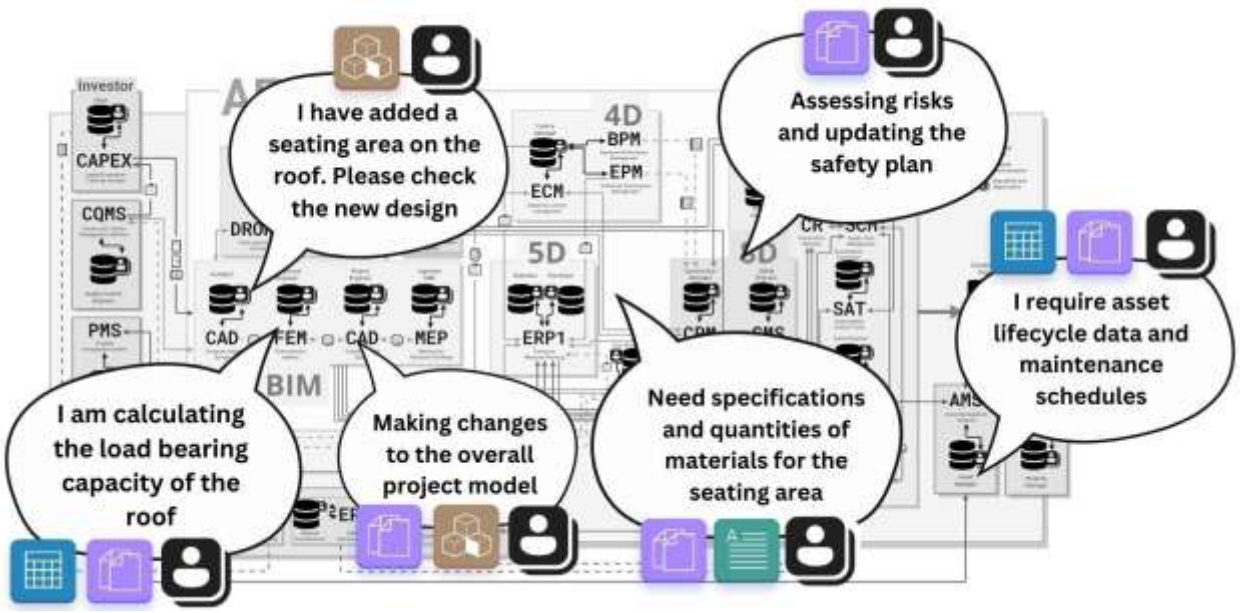


図 3.1-3 専門家同士のコミュニケーションは、テキストレベルとデータレベルの両方で行われる。

効果的なチームコラボレーションとプロジェクトの成功を実現するために、プロフェッショナルはそれぞれ異なるタイプのデータを扱います。構造化データ、半構造化データ、非構造化データの違いを理解することで、デジタル・ビジネス・プロセスにおいてそれぞれのタイプが果たす独自の役割を認識することができます。データにはさまざまな形式があることを知るだけでなく、それらがどのように、どこで、なぜ使われるのかを理解することも重要です。

> 少し前までは、このような多様なデータを組み合わせるというアイデアは野心的ではあったが、実現は難しいように思われた。今日、それはすでに日常的な実践の一部となっている。異なるスキーマや構造のデータを統合することは、現代の情報システム・アーキテクチャにとって不可欠な要素となっている。

以下の章では、構造化データ、半構造化データ、非構造化データを単一の一貫したビューに統合することを可能にする、主要な標準とアプローチについて詳しく見ていきます。特に、建設業界における情報の保存、処理、分析の主なメカニズムとして、構造化データとリレーショナル・データベースに注目する。

構造化データ

建設業界では、図面、仕様書、スケジュール、報告書など、多くの情報源から情報がもたらされる。このような情報の流れを効果的に管理するには、情報を構造化する必要があります。構造化データ 、便利で読みやすく、アクセスしやすい形で情報を整理することができます。

> JB ナレッジの第 5 回年次建設技術レポート[17]によると、建設プロジェクト管理専門家の 67%が、手作業またはスプレッドシートを使って作業実績を追跡・評価している。

最も一般的な構造化データ形式には、XLSX や CSV などがある。これらは、スプレッドシートにおける情報の保存、処理、分析に広く使用されている。このようなスプレッドシートでは、データは行と列の形で表示されるため、読み取り、編集、分析が容易である。

XLSX は、マイクロソフトが作成したフォーマットで、XML 構造の使用に基づき、ZIP アルゴリズムを使用してアーカイブされる。このフォーマットの主な特徴

- 複雑な数式、チャート、マクロをサポート。
- 情報をフォーマットするだけでなく、異なるシートにデータを保存する能力。
- Microsoft Excel 用に最適化されていますが、他のオフィススイートとも互換性があります。

CSV フォーマット は、カンマやセミコロンなどの区切り文字で値を区切ったプレーンテキストファイルです。主な利点

- 様々なプログラムやオペレーティングシステムとのユニバーサルな互換性。
- データベースや分析システムへのインポート/エクスポートが容易。
- テキストエディタでも簡単に処理できる。

しかし、CSV は数式やフォーマットをサポートしていないため、主な用途はシステム間のデータ交換や情報の大量更新である。CSV はその汎用性とプラットフォーム非依存性により、異種 IT 環境におけるデータ転送のための一般的なツールとなっている。

XLSX と CSV の 2 つのフォーマットは、構造化データを扱う異なるシステム間のリンクとして機能する（図 3.1-4）。読みやすさ、手作業による編集、基本的な互換性が重視される作業では、特に有用である

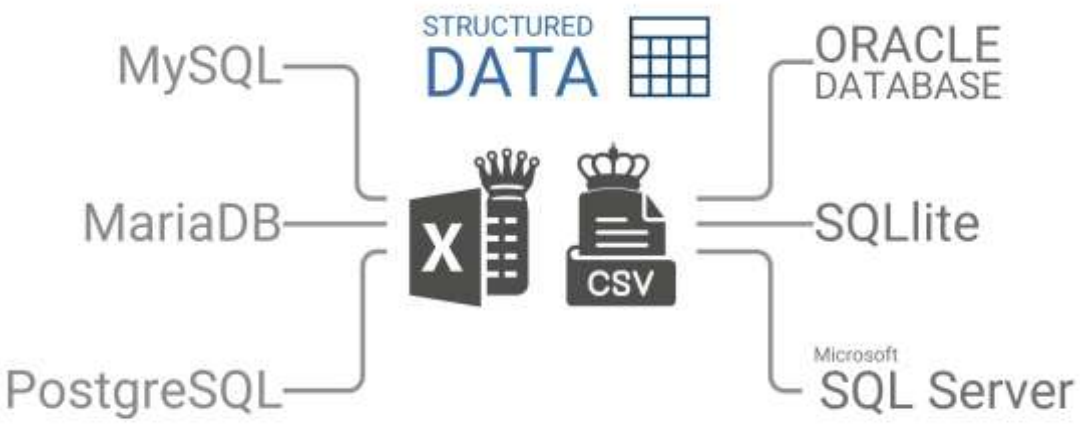


図 3.1-4 XLSX と CSV フォーマットは、構造化データを扱うさまざまなシステム間をつなぐものです。

> プラットフォームの独立性により、CSV は異種 IT 環境やシステムにおけるデータ転送のための最も一般的なフォーマットとなっている。

しかし、XLSX や CSV は、ハイパフォーマンス・コンピューティングや大量のデータの長期保存用に設計されていない。Apache Parquet, Apache ORC, Feather, HDF5 のような、より近代的な構造化フォーマットがそのような目的に使用されている。これらのフォーマットについては、本書のパート 9 の「ビッグデータの保存：一般的なフォーマットとその有効性の分析」の章で詳しく説明する

実際には、XLSX フォーマット（）の Excel（）が、小規模なタスクや定型的なプロセスの自動化によく使用される。より複雑なシナリオでは、ERP 、PMIS CAFM 、CPM 、SCM などのデータ管理システムの使用が必要となる（図 3.2-1）。これらのシステムには構造化されたデータが保存され、それを基に企業の情報フローが整理・管理される。

建設業界で使用される最新のデータ管理情報システムは、テーブルの形で整理された構造化データに依存している。信頼性が高く、 スケーラブルで、大量の情報を総合的に管理するために、アプリケーションやシステム開発者は、リレーショナルデータベース管理システム（RDBMS ）を利用している。

リレーショナル・データベース RDBMS および SQL クエリ言語

データを効率的に保存、処理、分析するために、リレーショナル・データベース（RDBMS ）は、情報を定義されたリレーションシップを持つテーブルに編成するデータ保存システムである。

> データベース（RDBMS ）に整理されたデータは、単なるデジタル情報ではなく、異なるシステム間のトランザクションや相互作用の基礎となる。

ここでは、最も一般的なリレーショナル・データベース管理システム（RDBMS）をいくつか紹介する（図 3.1-5）：

- MySQL (オープンソース) は、LAMP スタック (Linux, Apache, MySQL, PHP /Perl/Python) の一部で、最も人気のある RDBMS のひとつです。MySQL は、そのシンプルさと高いパフォーマンスにより、ウェブ開発で広く使用されています。
- PostgreSQL (オープンソース) は、信頼性と高度な機能で知られる強力なオブジェクトリレーショナルシステムです。複雑なエンタープライズソリューションに適しています。
- Microsoft SQL Server は、マイクロソフト社の商用システムで、他の企業製品との統合性と高いセキュリティレベルにより、企業環境で広く使用されている。
- オラクル・データベースは、大企業やミッション・クリティカルなアプリケーションで使用されている、最も強力で信頼性の高い DBMS の 1 つです。
- IBM DB2 - 大企業向けで、高いパフォーマンスと耐障害性を提供。
- SQLite (オープンソース) は軽量の組み込みデータベース で、モバイルアプリケーションや CAD 設計ソフト (BIM) などのスタンドアロンシステムに最適です。

MySQL 、PostgreSQL 、Microsoft SQL Server、Oracle® Database、IBM® DB2、SQLite など、建設業でよく使われるデータベース管理システムは、構造化データを扱います。これらの DBMS はすべて、小規模なウェブサイトから大規模なエンタープライズシステムまで、幅広いビジネスプロセスとアプリケーションを管理するための強力で柔軟なソリューションです（図 3.2-1）。

> Statista [48]によると、リレーショナルデータベース管理システム（RDBMS）は、2022 年に使用される DBMS 全体の約 72%を占める。



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

図 3.1-5 DBMS ランキングにおける構造化データベース（青字）使用の人気度（[49]に基づく）

。

オープンソースデータベースをインストールするのは非常に簡単である。PostgreSQL、MySQL、SQLite などのオープンソースシステムは無料で入手でき、Windows、macOS、Linux などのほとんどのオペレーティングシステムで動作します。必要なのは、プロジェクトの公式ウェブサイトに行き、インストーラをダウンロードして指示に従うだけです。ほとんどの場合、インストールには 10〜15 分もかかりません。本書の第 4 部では、このようなデータベースをモデル化して作成します（図 4.3-8）。

クラウドサービス（Amazon Web Services、Google Cloud、Microsoft Azure など）を利用している企業であれば、クリック数回でデータベースを導入することができる。コードのオープン性のおかげで、このようなデータベースは、あなたのタスクのためにカスタマイズするのが簡単です。

RDBMS は、企業が効率的にデータを保存、処理、分析し、その結果、情報に基づいたタイムリーな意思決定を行うことを可能にする、多くのビジネス・アプリケーションや分析プラットフォーム（図 3.1-6）の基盤であ

り続けている。

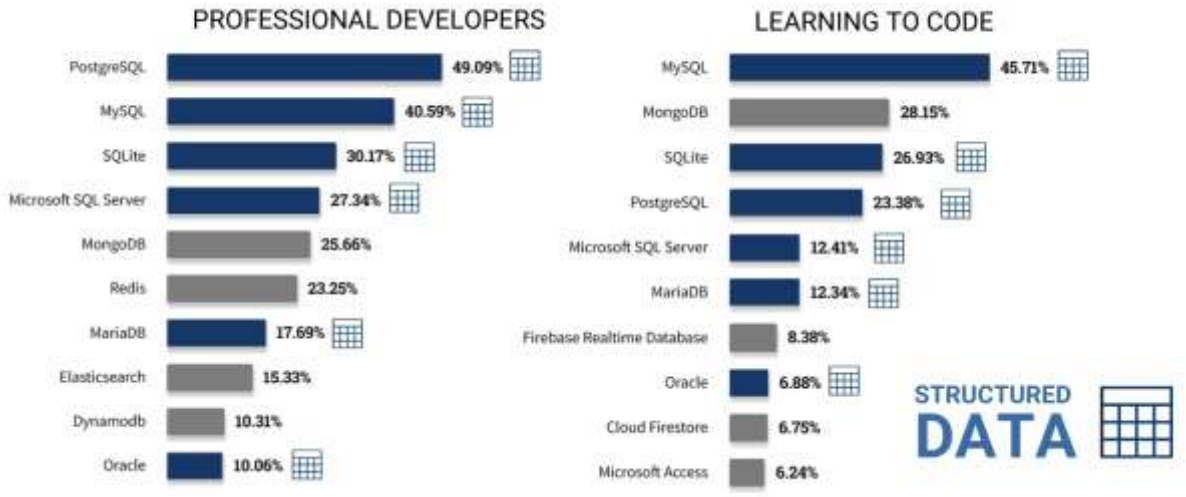


図 3.1-6 StackOverFlow（IT フォーラム最大手）の開発者を対象に、昨年使ったデータベースと来年使いたいデータベースを調査（RDBMS は青でハイライト）（[50]に基づく）。

RDBMS は、信頼性、データの一貫性、トランザクション・サポートを提供し、強力なクエリ言語である SQL (Structured Query Language)を使用します。SQL は分析でよく使用され、データベースに格納された情報を簡単に取得、変更、分析することができます。SQL は、リレーショナル・システムでデータを扱うための主要なツールです。

SQL -データベースにおけるクエリと新しいトレンド

リレーショナル・データベースでよく使われる SQL 言語（）の主な利点は、他のタイプの情報管理（たとえば、従来の Excel スプレッドシート（ ）を使用した場合）と比べて、非常に大量のデータベースを高速なクエリ処理でサポートできることである。

> 構造化クエリー言語（Structured Query Language：SQL）は、リレーショナル・データベースに情報を保存、処理、分析するために設計された特殊なプログラミング言語です。SQL はデータの作成、管理、アクセスに使用され、情報の検索、フィルタリング、結合、集約を効率的に行うことができます。SQL はデータにアクセスするための重要なツールであり、情報ストアにアクセスするための便利で形式化された方法を提供します。

SEQUEL-SQL システムの進化は、Oracle、IBM DB2、Microsoft SQL Server、SAP、PostgreSQL 、MySQL などの重要な製品や企業を経て、SQLite や MariaDB [51] の登場へと結実する。SQL は、Excel にはない表計算機能を提供し、データ操作をよりスケーラブルに、セキュアに、自動化しやすくする：

- データ構造の作成と管理（DDL）：SQL では、データベース内のテーブルを作成、変更、削除し、テーブル間のリンクを確立し、データ格納構造を定義することができます。一方、エクセル では、固定されたシートとセルを扱います。シートとデータセットの間に、 明確に定義された関係はありません。
- データ操作（DML）：SQL 、フィルタリング、ソート、テーブル結合などの複雑なクエリを実行することで、大量のデータを高速に追加、変更、削除、取得することができる（図 3.1-7）。エクセル では、大量の情報を処理するには、手作業や特別なマクロが必要で、処理速度が遅くなり、エラーの確率も高くなります。
- アクセス制御（DCL）：SQL では、ユーザーごとにデータへのアクセス権を区別し、情報の編集や閲覧を制限することができる。一方、Excel では、アクセスは共有されるか（ファイル転送時）、クラウドサービスを介した権限共有で複雑な設定が必要となる。

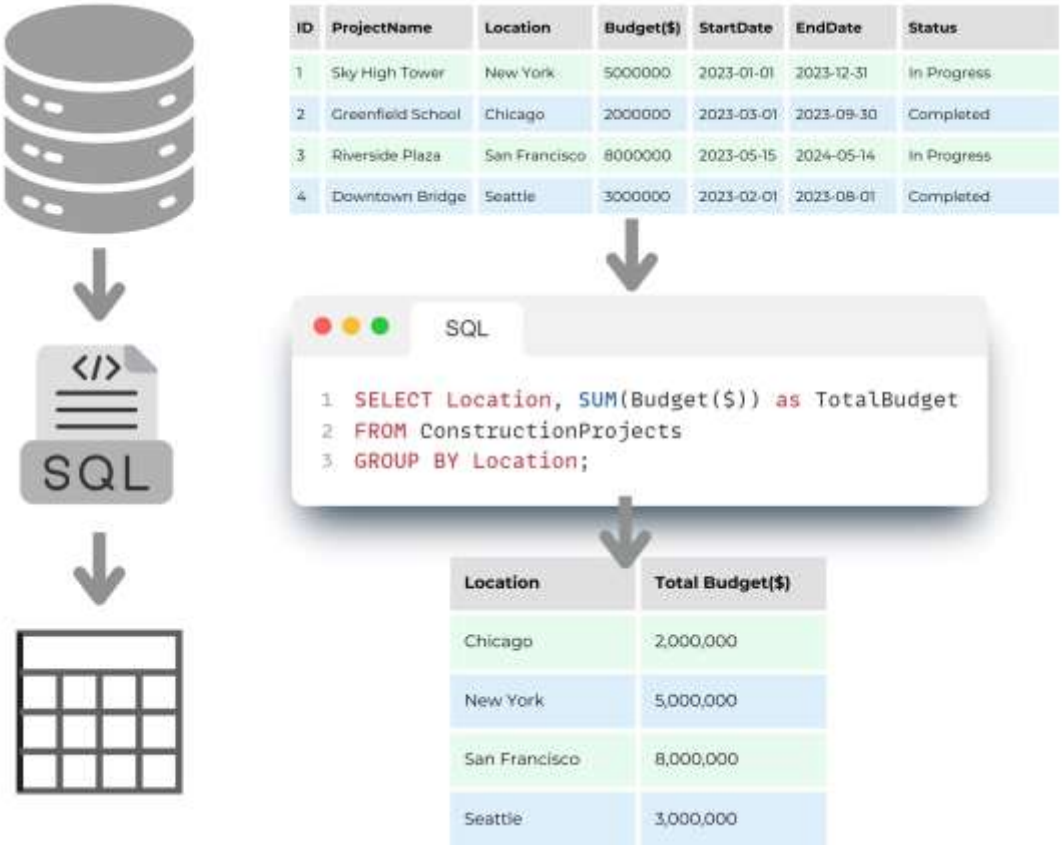


図 3.1-7 SQL における DML の例: 自動データ処理のための数行のコードによる高速処理、グループ化、集計。

エクセル その視覚的で直感的な構造により、データ作業が容易になる。しかし、データ量が増えるにつれ、エクセルのパフォーマンスは低下する。また、エクセルは保存できるデータ量に制限があり、最大 100 万行までしか保存できません。そのため、少量のデータを視覚化して操作するにはエクセルの方が適しているように見えるが、大規模なデータセットを扱うには SQL の方が適している 。

構造化データの発展における次の段階は、カラム型データベース（Columnar Databases）の登場であった。カラム型データベースは、従来のリレーショナル・データベース（ ）に代わるものであり、特にデータ量が著しく大きく、分析計算が必要な場合に適している。行ごとにデータが格納される行データベースとは異なり、カラム型データベースは列ごとに情報を記録する。従来のデータベースと比較すると、以下のような利点がある：

- カラム内の均一なデータを効率的に圧縮することで、ストレージ容量を削減します。
- テーブル全体ではなく、必要なカラムのみを読み込むため、分析クエリを高速化。
- ビッグデータ とデータウェアハウスの最適化, e.g. Data Lakehouse Architecture.

カラム型データベース、Pandas DataFrame 、Apache Parquet 、HDF5 、そしてデータ分析・加工を目的とした、これらに基づいたビッグデータ ストアの作成については、本書の次の章「DataFrame: a universal tabular data format」と「Data storage formats and working with Apache Parquet: DWH -data warehouses and Data Lakehouse architecture 」で詳しく説明します。

非構造化データ

アプリケーションや情報システムで使用されるデータのほとんどは構造化された形式であるが、建設現場で生成される情報のほとんどは、画像、ビデオ、テキスト文書、音声記録、その他の形式のコンテンツといった非構造化データの形式である。これは、特に建設、操業、技術監督の段階で顕著であり、視覚的・文字的情報が優位を占めている。

> 非構造化データとは、あらかじめ定義されたモデルや構造を持たない情報のことで、データベースやテーブルのように伝統的な行や列に整理されていない。

一般論として、非構造化データは 2 つのカテゴリーに分類できる：

- 人間が生成した非構造化データには、テキスト文書、電子メール、画像、ビデオなど、人間が生成したさまざまな種類のコンテンツが含まれる。
- 機械が生成する非構造化データは、デバイスやセンサーによって作成される。例えば、ログファイル、GPS データ、モノのインターネットの結果（IoT ）、建設現場からのその他の遠隔測定情報などである。

表やデータベースに便利に整理される構造化データとは異なり、非構造化データは、情報システムに統合する前に、さらなる処理段階を必要とする（図 3.1-8）。このようなデータの収集、分析、変換を自動化する技術を利用することで、建設効率を向上させ、エラーを減らし、人的要因の影響を最小限に抑える新たな機会が開かれる。

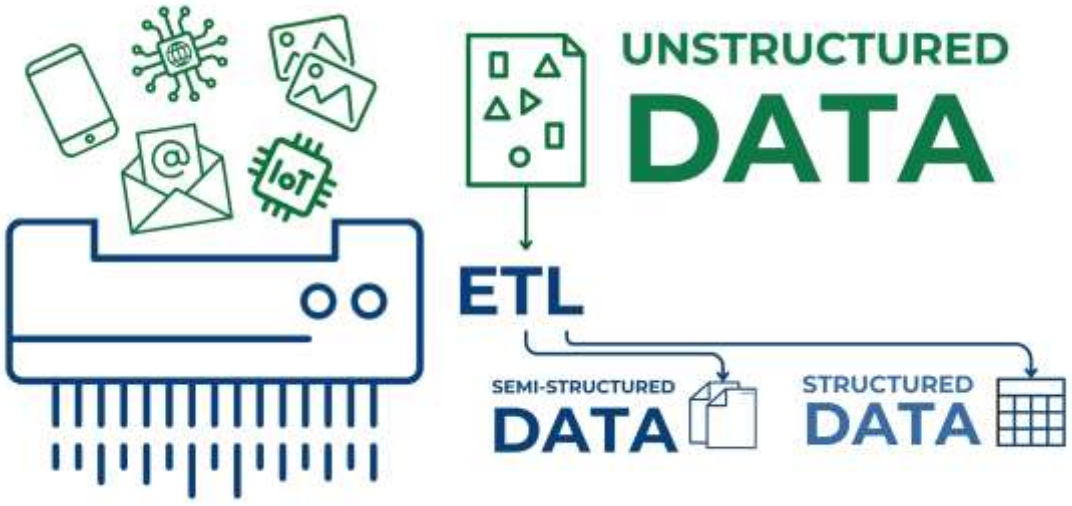


図 3.1-8 非構造化データの処理は、半構造化データと構造化データへの変換から始まる。

非構造化データは、企業で専門家が遭遇する全情報[52]の最大 80%を占めているため、本書の次の章では、その種類と処理について例を挙げて詳しく説明する。

議論を容易にするため、テキストデータは別に分類されている。テキスト・データは非構造化データの一種であるが、建設業界におけるその重要性と普及率には特別な注意が必要である。

テキストデータ：非構造化カオスと構造化カオスの間 y

建設業界におけるテキストデータ は、紙の文書から、手紙、会話、作業連絡、建設現場での口頭での打ち合わせなどの非公式なコミュニケーション手段に至るまで、幅広い形式と種類の情報を含んでいる。このようなテキストデータはすべて、設計の決定や計画の変更の詳細から、安全問題の議論、請負業者や発注者との交渉に至るまで、建設プロジェクトを管理する上で重要な情報を含んでいる（図 3.1-9）。



図 3.1-9 テキスト・データ プロジェクト参加者間のコミュニケーションに使われる最も一般的な情報タイプのひとつ。

テキスト情報には、形式化されたものと構造化されていないものがある。形式化されたデータには、Word 文書（.doc、.docx）、PDF（ ）、会議議事録のテキストファイル（.txt）などがある。非構造化データには、メッセンジャーや電子メールのやり取り、会議の議事録（Teams、Zoom、Google Meet）、テキストへの変換が必要なディスカッションの音声録音（.mp3、.wav）などがある。

しかし、正式な依頼書や契約条件、電子メールなどの文書には、すでにある程度の構造があるのが普通だが、口頭でのコミュニケーションや業務上のやり取りは構造化されていないことが多く、分析やプロジェクト管理システムへの統合が難しい。

テキストデータを効果的に管理する鍵は、構造化フォーマットに変換することである。これにより、すでに構造化データを扱っている既存のシステムに、処理された情報を自動的に統合することができる。



図 3.1-10 テキストコンテンツを構造化データに変換する。

テキスト情報を効果的に利用するためには、それを自動的に構造化された形に変換する必要がある（図 3.1-10 ）。このプロセスには通常、いくつかの段階があります：

- テキスト認識 (OCR) - 文書や図面の画像を機械が読み取り可能なフォーマットに変換する。

- テキスト分析（NLP ） - 主要なパラメータ（日付、金額、プロジェクトに関連する数字）を自動的に識別します。
- データ分類 - 情報を分類する（財務、ロジスティクス、リスク管理）。

認識と分類の後、 、すでに構造化されたデータをデータベースに統合し、自動化された報告書や管理システムで使用することができる。

半構造化データおよび疎構造化データ

半構造化データ ある程度整理されているが、厳密なスキーマや構造を持っていない。このような情報には、構造化された要素（例えば、 日付、従業員名、完了したタスクのリスト）が含まれるが、表示形式はプロジェクトごと、あるいは従業員ごとにかなり異なる場合がある。このようなデータの例としては、タイムログ、進捗報告書、スケジュールなどがあり、さまざまな形式で表示することができる。

半構造化データ は、非構造化データよりは分析しやすいが、標準化されたプロジェクト管理システムに統合するための追加処理が必要である。

> 常に変化する構造を特徴とする半構造化データを扱うことは、大きな困難を伴う。これは、データ構造の多様性により、半構造化データの各ソースの処理と分析に個別のアプローチが必要になるためである。

しかし、非構造化データを扱うには多大な労力を必要とするが、半構造化データの処理は比較的簡単な方法とツールで行うことができる。

> 弱構造化データ（Weakly structured data） は、最小限の、あるいは不完全な構造を持つデータを表す、より一般的な用語である。多くの場合、テキスト文書、チャット、電子メールなどであり、いくつかのメタデータ（日付、送信者など）は見つかるが、ほとんどの情報は無秩序な方法で表示される。

建設業界では、疎構造化データはさまざまなプロセスで見られる。例えば、以下のようなものがある：

- 見積書 および見積書 - 材料、数量、コストに関するデータを含む表だが、統一された書式はない。
- 図面およびエンジニアリング・ダイアグラム - PDF または DWG のファイルで、テキスト注釈とメタデ

ータを含むが、厳密には固定された構造を持たない。

- 作業スケジュール - MS Project 、Primavera P6 または他のシステムからのデータ。
- CAD (BIM -models) - 構造の要素を含むが、データ表現はソフトウェアとプロジェクト標準に依存する。

CAD システムによって作成される幾何学データ は、半構造化データと同じように分類することができる。しかし、幾何学的 CAD（BIM）データは、テキストデータと同様に、企業プロセスにおいて別個のデータタイプとして扱われることが多いため、ここでは別個のデータタイプとして分類する。

幾何学的データ とその応用

プロジェクト要素に関するメタデータは、ほとんどの場合、表や構造化された形式、あるいは緩やかな構造化された形式で保存されますが、プロジェクト要素の幾何学的データは、 ほとんどの場合、特別な CAD ツール（図 3.1-11）を使って作成され、プロジェクト要素を線の集合（2D）または幾何学的形状（3D ）として詳細に可視化することができます。



図 3.1-11 CAD ツールは、幾何学的情報を物理的メディアからデータベース形式へ移行するのに役立っている。

建設や建築において幾何学的データを扱う場合、幾何学的データの 3 つの主な用途を特定することができる（図 3.1-12）：

- 体積の確認：特殊な幾何学カーネル を使用して CAD プログラム （BIM ）内で生成された幾何学データ は、プロジェクト要素の体積と寸法を自動的かつ正確に決定するために必要である。このデータに

は、自動的に計算された面積、体積、長さ、および資源や材料の計画、予算、発注に必要なその他の重要な属性が含まれる

- プロジェクトの視覚化: プロジェクトに変更が生じた場合、要素の視覚化により、異なる平面で更新された図面を自動生成することができます。初期段階でのプロジェクトの視覚化は、すべての参加者間の理解を早め、建設プロセスにおける時間とリソースの節約に役立ちます。
- 衝突チェック: 複雑な建設プロジェクトやエンジニアリングプロジェクトでは、「幾何学的な衝突」なしに複数のカテゴリー（パイプや壁など）の要素を相互作用させることが重要であり、衝突チェックは重要な役割を果たします。 衝突検出ソフトウェアを使用することで、プロジェクト要素間の潜在的な幾何学的衝突を事前に特定することができ、建設プロセス中のコストのかかるエラーを防ぐことができます。

工学設計事務所のごく初期から、最初の複雑な構造物の建設時から、構造エンジニアは、図面、線、平 面幾何学的要素（パピルス、「A0」紙、または DWG 、PDF 、PLT 形式）の形で幾何学的情報を提供し、それに基づいて現場監[illegible] と積算担当者（図 3.1-11） は、過去数千年にわたり、定規と分度器を使って、要素や要素群の属性的な体積や数量を収集した。3.1-11) 、過去千年にわたり、定規や運搬具を用いて、元素や元素群の属性的な体積や数 量を収集した。

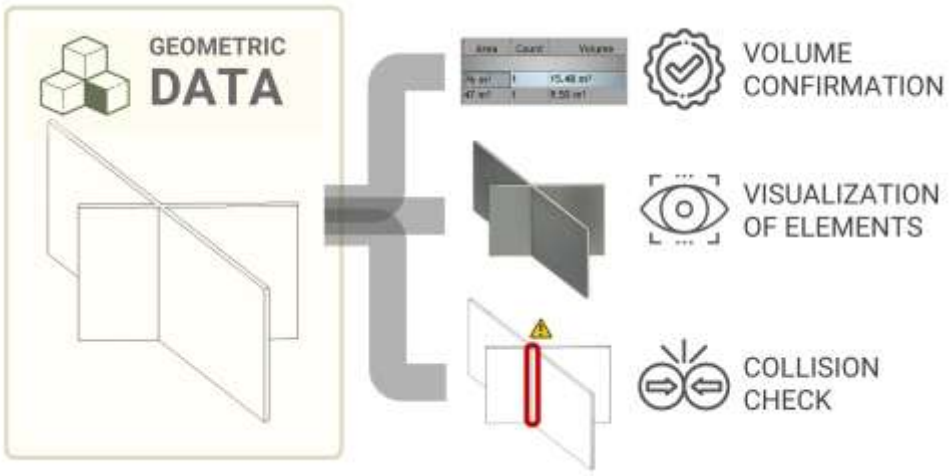


図 3.1-12 ジオメトリは、要素の容積パラメータを求めるための基礎であり、このパラメータは 、プロジェクトのコストとタイミングを計算するために使用される。

今日、この手作業で時間のかかる作業は、最新の CAD ツール (BIM)におけるボリュメトリック・モデリングの

出現により、完全自動化によって解決されています。このツールは、特別な幾何学的カーネルの助けを借りて、手動でボリュメトリック・パラメータを計算する必要なく、あらゆる要素のボリュメトリック属性を自動的に取得することができます。

また、最新の CAD ツール では、プロジェクト要素を分類・カテゴライズすることができ、プロジェクト データベースから仕様表をアップロードして、積算、スケジューリング、CO_2 計算 などの様々なシステムで使用することができます（図 3.1-13）。仕様の取得、QTO テーブル と数量の取得、実践例については、「数量の取得と定量化」の章で説明します。

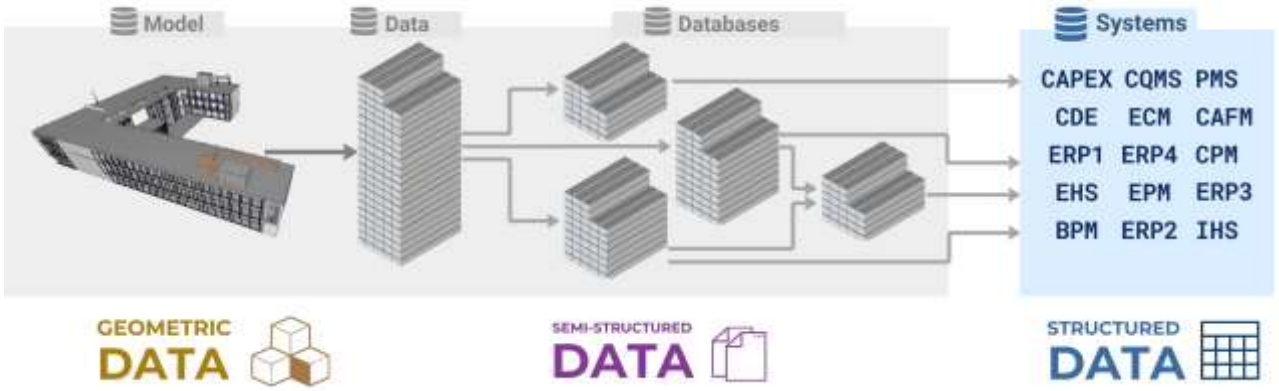


図 3.1-13 CAD ツール （BIM ）は、他のシステムと統合し相互作用するように設計されたデータベースにデータを格納する。

CAD 環境で使用されるデータベースやフォーマットの閉鎖的な性質により、CAD ソリューションで作成される幾何学データは、実際には別のタイプの情報になっています。これは、要素の形状とメタ情報（構造化または半構造化）の両方を組み合わせたもので、特殊なファイルやフォーマットに格納されています。

CAD データ：設計からデータ保存まで

最新の CAD および BIM システムは、多くの場合、DWG、DXF、RVT、DGN、PLN などの独自のフォーマットでデータを保存します。DWG、DXF、RVT、DGN、PLN などです。これらのフォーマットは、オブジェクトの 2D と 3D の両方をサポートし、ジオメトリだけでなく、オブジェクトに関連する属性も保持します。ここでは、最も一般的なものを紹介します：

- DWG は、2 次元（およびあまり一般的でない 3 次元）の設計データとメタデータを保存するために使用されるバイナリファイル形式です。

- DXF は、CAD -システム間で 2D および 3D -図面を交換するためのテキスト形式です。ジオメトリ、レイヤ、属性データを含み、ASCII とバイナリ表現の両方をサポートしています。
- RVT は、3D -形状、要素属性、関係、設計パラメータを含む CAD 保存するためのバイナリ・フォーマットです。
- IFC は、CAD （BIM ）システム間で建設データを交換するためのオープン・テキスト・フォーマットです。これには、ジオメトリ、オブジェクトのプロパティ、およびそれらの関係に関する情報が含まれます。

これらに加えて、PLN、DB1、SVF、NWC、CPIXML、BLEND、BX3、USD、XLSX、DAE などのフォーマットも使用される。これらのフォーマッ トは、目的やオープン性のレベル（図 3.1-14）は異なりますが、同じプロジェクト情報モデルを異なる形 で表現することができます。複雑なプロジェクトでは、図面作成からプロジェクトモデルの調整まで、これらのフォーマ ットが並行して使用されることがよくあります。

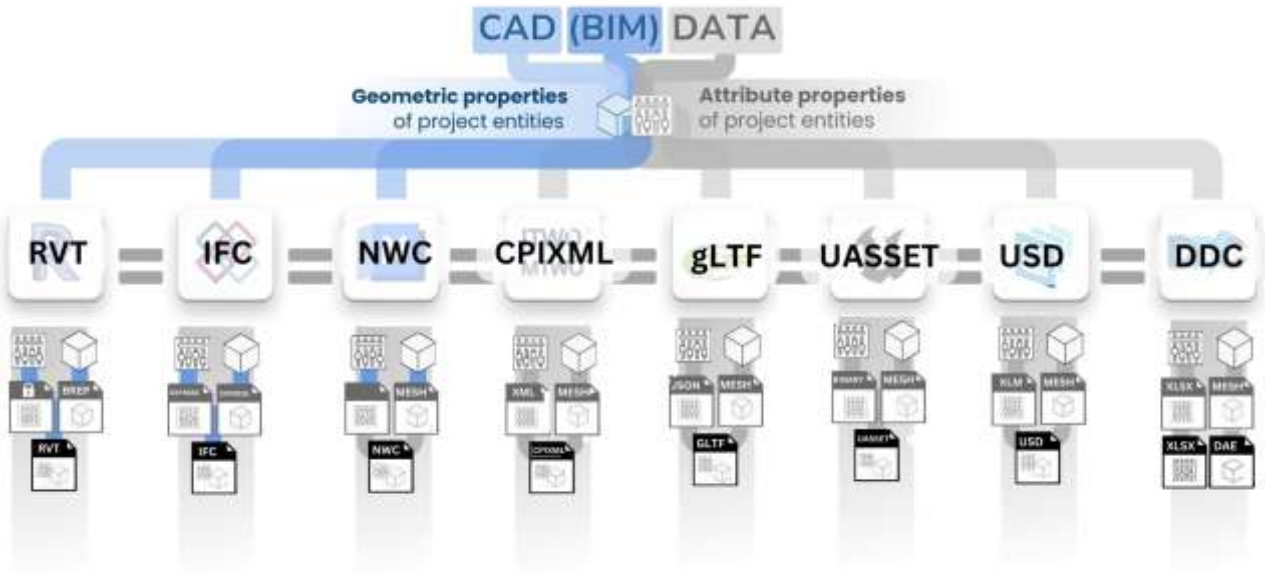


図 3.1-14 一般的な CAD 保存形式は、属性データによって補足された BREP または MESH パラメータによってジオメトリを記述します。

上記のフォーマットはすべて、建設プロジェクトの各要素に関するデータを保存することができ、上記のフォーマットはすべて、2 つの重要なタイプのデータを含んでいる：

- 幾何学パラメータ - オブジェクトの形状、位置、寸法を表す。ジオメトリとその使用法については、CAD（BIM ）ソリューションに特化した本書の第 6 部で詳しく説明する；
- 属性プロパティ - プロジェクト要素が持つ可能性のある材料、要素タイプ、技術的特性、一意の識別子

、その他のプロパティなど、さまざまな情報を含む。

属性データは、オブジェクトの動作特性を定義し、エンジニアリングとコスト計算を可能にし、設計、建設、運用の参加者間のエンドツーエンドの相互作用を提供するため、現代のプロジェクトでは特に重要である。例えば

- 窓とドアの場合：構造の種類、グレージングの種類、開口方向（図 3.2-1.）
- 壁については、材料、断熱材、音響性能に関する情報が記録されている。
- エンジニアリング・システムについては、パイプライン、ダクト、ケーブル・ルート、およびそれらの接続部のパラメータが保存される。

これらのパラメータは、CAD-(BIM)ファイル自体と、外部データベースの両方に保存することができます - エクスポート、変換、またはリバースエンジニアリングツール（）による内部 CAD 構造への直接アクセスの結果として。このアプローチは、他の企業システムやプラットフォームとの設計情報の統合を促進する

> CAD（BIM）の文脈におけるリバースエンジニアリングとは、元のアルゴリズムやドキュメントにアクセスすることなく、デジタルモデルのロジック、データ構造、依存関係を再現するために、その内部構造を抽出して分析するプロセスのことである。

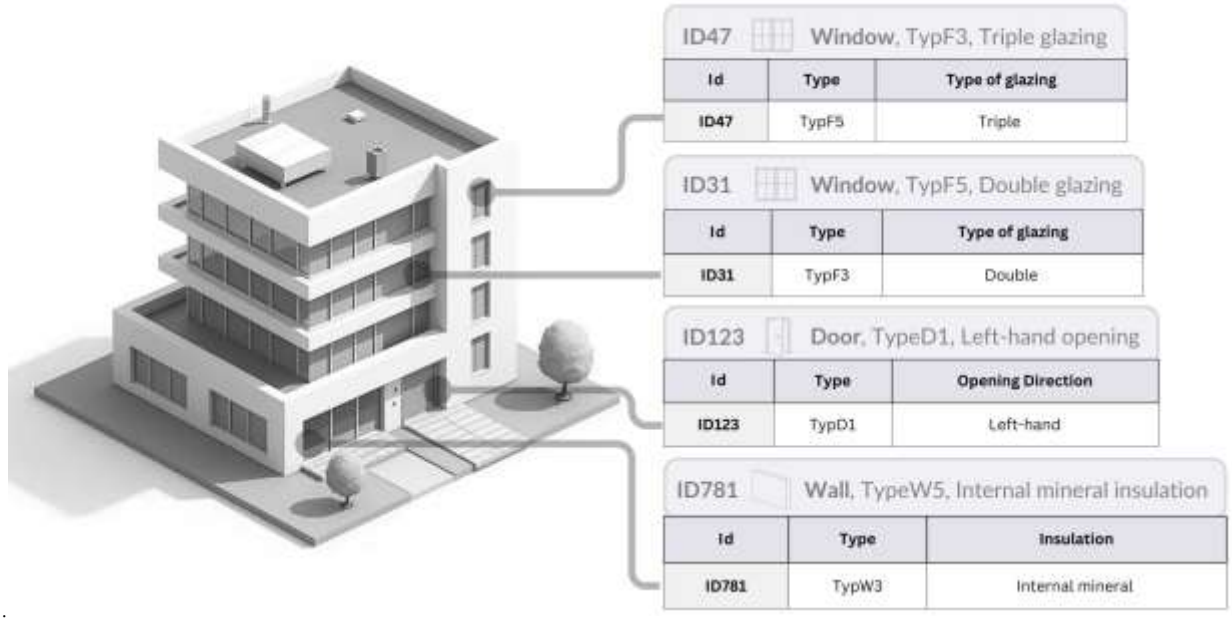


図 3.1-15 プロジェクト要素は、パラメトリックあるいはポリゴン形状の記述に加えて、要素のパラメー タとプロパティに関する情報を含んでいる 。

その結果、各オブジェクトの固有の特性（識別子や寸法など）と、要素グループの共通属性の両方を含む、固有のパラメータとプロパティのセットが各要素の周りに形成されます。これにより、プロジェクトの個々の要素-実体を分析できるだけでなく、それらを論理的なグループにまとめることができます。

> エンティティとは、現実世界の具体的または抽象的な物体で、一意に識別、記述、データの形で表現できるもの。

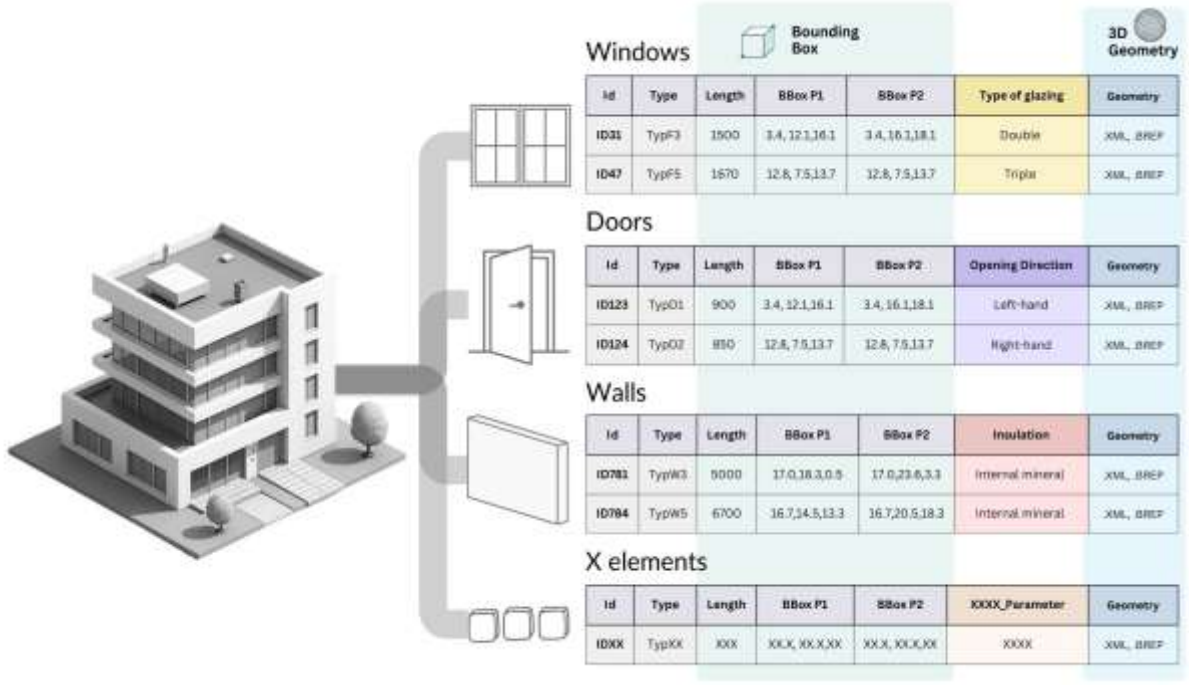


図 3.1-16 各プロジェクト要素には、設計者が入力するか、CAD プログラム内で計算された属性が含まれています。

過去数十年の間に、建設業界はデータの作成、保存、転送を簡素化する新しい CAD（BIM ）フォーマットを数多く開発してきました。これらのフォーマットは、クローズドであったりオープンであったり、表形式であったりパラメトリックであったりグラフィカルであったりする。しかし、その多様性と断片化は、プロジェクトのライフサイクルのあらゆる段階におけるデータ管理を著しく複雑にしている。図 3.1-17（フルバージョンは QR コードで入手可能）に、建設現場で情報交換に使用される主なフォーマットの比較表を示す。

相互運用性（）と CAD データへのアクセス）の問題を解決するために、BIM 管理者（BIM ）とコーディネーターが関与しており、その任務は、エクスポートを管理し、データ品質をチェックし、CAD（BIM）データの一

部を他のシステムに統合することである。

しかし、フォーマットが閉鎖的で複雑なため、このプロセスを自動化することは難しく、本格的なインラインのデータ処理プロセス（パイプライン）を構築する能力を持たない専門家は、多くの作業を手作業で行わざるを得ない。

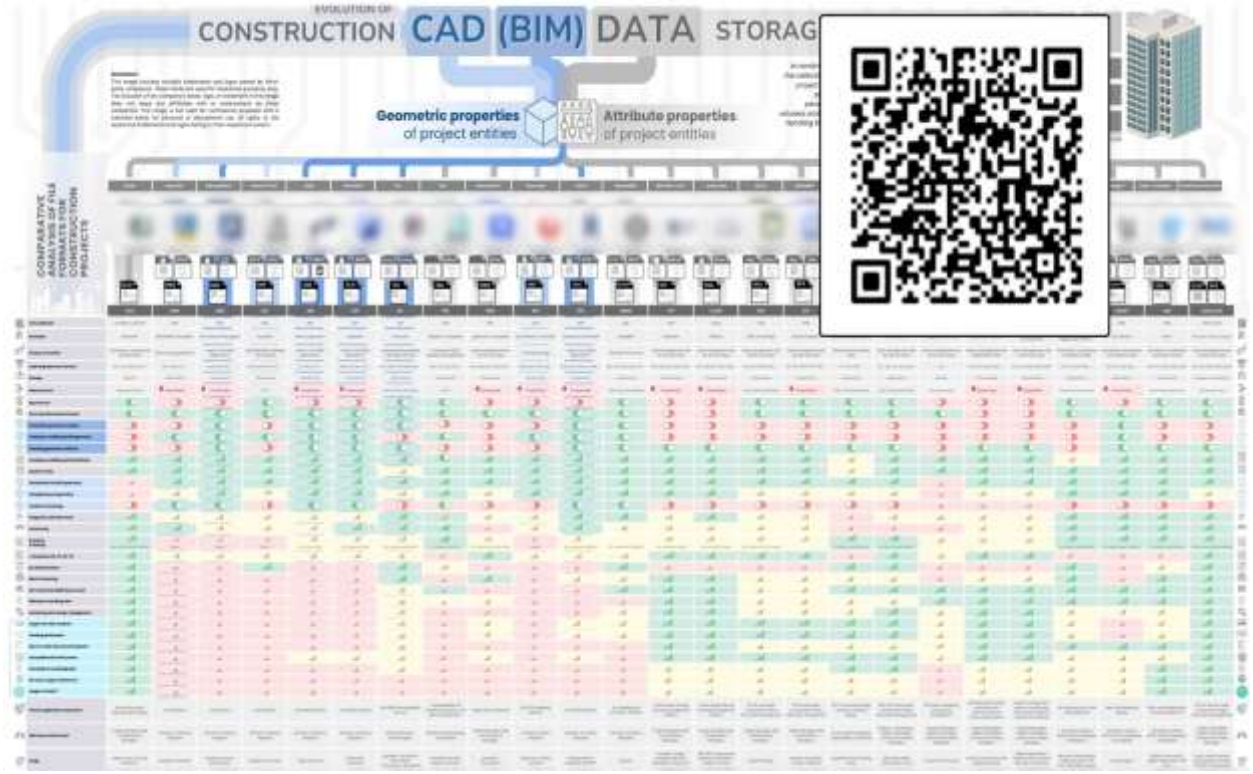


図 3.1-17 プロジェクト要素情報が格納される主なデータ形式を比較した表 [53]。

なぜこれほど多くの異なるデータ形式が存在し、そのほとんどがクローズドなのかを理解するためには、CAD (BIM) プログラムの内部で行われているプロセスを掘り下げることが重要である。本書の第 6 部では、このプロセスについて詳しく説明する。

ジオメトリに追加される情報レイヤーは、CAD システム開発者によって、BIM コンセプト （ビルディング・インフォメーション・モデリング）という形で導入された。このマーケティング用語は、2002 年以来、建設業界で積極的に推進されている[54]。

BIM（BOM）コンセプトの出現 および工程における CAD の使用

ビルディング・インフォメーション・モデリング（BIM ）の概念は、2002 年の BIM ホワイトペーパー[54]で初

めて概説されたもので、CAD ソフトウェアメーカー のマーケティングイニシアティブに端を発している。BIM は、CAD ソフトウェア開発者のマーケティング・イニシアティブから生まれたもので、機械工学ですでに確立された原則を建設業界のニーズに適応させようとする試みであった。

BIM の着想は、1980 年代後半から産業界で広く使われてきた BOM（Bill of Materials：部品表）の概念に由来する。機械工学では、BOM は CAD システムからのデータを PDM（製品データ管理）、PLM（製品ライフサイクル管理）、ERP システムとリンクさせることを可能にし、製品ライフサイクル全体を通してエンジニアリング情報の全体的な管理を提供した（図 3.1-8）。

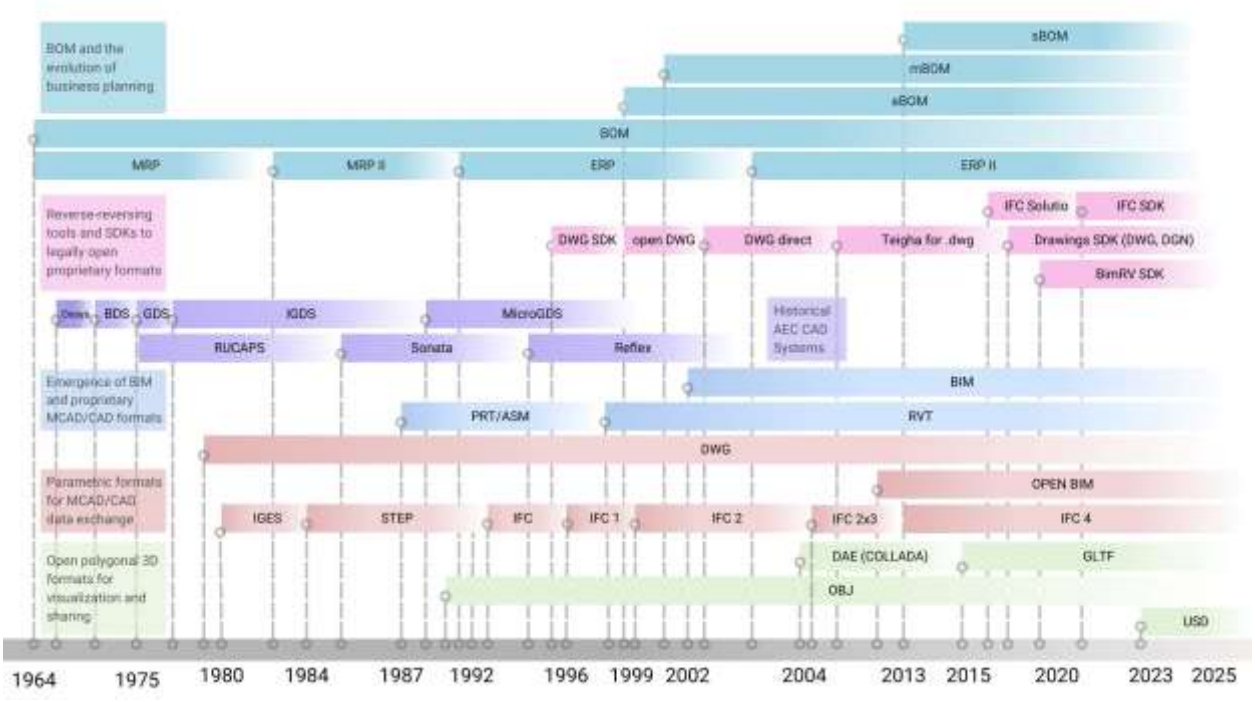


図 3.1-18 エンジニアリング建設業界における仕様書（BOM ）、情報モデリング（BIM ）、デジタルフォーマットの進化。

BOM コンセプトの現代的な発展により、製品構成だけでなく、行動シナリオ、運用要件、持続可能性パラメーター、予測分析用データなどを含む拡張フレームワーク、XBOM（拡張 BOM）が登場した。XBOM は本質的に、建設における BIM（ ）と同じ役割を果たす。どちらのアプローチも、プロジェクトのライフサイクルを通じて、デジタルモデルをすべてのプロジェクト参加者にとっての「真実の情報源（Single Source of Truth）」に変えようとするものである。

建設業における BOM の出現における重要なマイルストーンは、2002 年に建設業に特化した初のパラメトリック CAD（MCAD）が導入されたことである。これは、1980 年代後半に登場し、業界標準となった機械工学用の

革新的な MCAD システムである Pro-E®を開発したチームによって開発された[55]。

1980 年代末にはすでに、当時既存の CAD プログラムの制限 [56] をなくすことが目標となっていた。主な目的は、設計要素のパラメータを変更するために必要な労力を削減し、データベースを介して CAD プログラム外のデータに基づいてモデルを更新できるようにすることであった[57]。そのために最も重要な役割を果たすのがパラメータ化であり、データベースから特性を自動的に取得し、CAD システム内のモデルを更新するために使用することであった。

> Pro-E とその基礎となる要素パラメトリック・モデリング C BOM のコンセプトは、CAD - そして MCAD - 市場の発展に大きな影響を与えた [58]。このモデルは 25 年間業界に存在し、多くの最新システムがその概念的後継者となっている。

> 目標は、エンジニアが異なる設計を容易に検討できるような柔軟なシステムを作ることである。そして、設計の変更にかかるコストは可能な限りゼロに近づけなければならない。従来の CAD / CAM ソフトウェアは、非現実的なことに、設計プロセスのごく初期にのみ、安価な変更を行うことを制限している[59]。
>
> - Parametric Technology Corporation® の創設者であり、MCAD 製品 Pro-E の開発者であり、RVT フォーマットを使用した CAD 製品の作成者の教師である Samuel Heisenberg 氏は、次のように述べています。

機械工学では、PDM、PLM、MRP、ERP システムが重要なプラットフォームとなっている。これらのシステムは、CAx システム（CAD、CAM、CAE）から情報を収集し、製品構造（BOM：eBOM、pBOM、mBOM）に基づいて設計活動を組織化し、データとプロセスの管理において中心的な役割を果たします（図 3.1-18）。この統合により、エラーを削減し、データの重複を回避し、設計から製造までのエンドツーエンドのトレーサビリティを確保します。

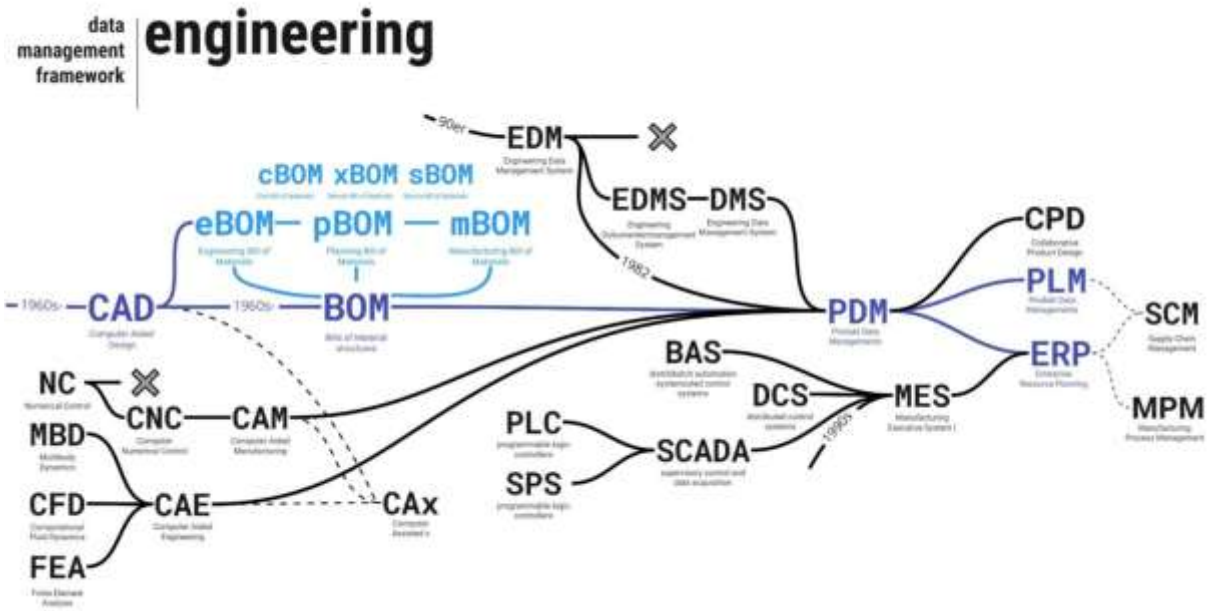


図 3.1-19 歴史的には、BOM は 1960 年代に CAx システムからのデータを構造化して制御システムに渡す方法として登場した。

旧 Pro-E チームが開発し、BOM アプローチに基づく CAD ソリューションを大手ベンダーの 1 社が購入したことは、BIM ホワイトペーパーシリーズ （2002〜2003 年）[60][61]の出版によってほぼ即座に明らかになった。すでに 2000 年代半ばから、BIM のコンセプトは建設業界で積極的に推進され始め、パラメトリックソフトウェアへの関心が著しく高まった。その人気は急速に高まり、機械工学（ Pro-E）の建設フォーク（このベンダーが推進するパラメトリック CAD）は、実際に建築・構造設計分野で競合他社を駆逐した（図 3.1-20）。2020 年代初頭までには、BIM（CAD）市場における世界的な優位性を事実上固めた[62]。



図 3.1-20 Google 検索クエリ人気（RVT 対 IFC）：旧 Pro-E チームが作成したパラメトリック CAD BOM サポート付き -BIM 世界のほとんどの国で人気を博している。

過去 20 年の間に、BIM（ ）という略語は多くの解釈を獲得してきたが、その複数の意味は、2000 年代初頭に登場した初期のマーケティングコンセプトに根ざしている。この用語の普及に重要な役割を果たした ISO 19650 規格は、情報管理に対する「科学に基づく」アプローチとしての BIM の地位を実際に確保した。しかし、BIM を用いたオブジェクトのライフサイクル全般にわたるデータ管理に特化したこの規格の本文自体では、BIM という略語は言及されているものの、明確に定義されたことはない

2002 年[60]と 2003 年[61]に BIM に関する一連のホワイトペーパーを発表したこのベンダーのオリジナルのウェブサイトは、実は 1990 年代に Pro-E 機械工学ソフトウェアで以前使用されていた BOM（部品表）と PLM（製品ライフサイクル管理）の概念に関するマーケティング資料を再現したものであった[63]。

> ビルディング・インフォメーション・モデリング（BIM）は、2002 年に.....【CAD ベンダーの会社名】によって導入された、建築物の設計、建設、管理に対する革新的な新しいアプローチであり、世界中の業界専門家が、建築物の設計、建設、管理にどのように技術を適用できるかについての考え方を変えた。
>
> - BIM ホワイトペーパー, 2003 [61].

これらの初期の出版物は、BIM（ ）を集中型統合データベースの概念に直接結びつけている。2003 年のホワイトペーパーで述べられているように、BIM は、すべての更新が単一のリポジトリで行われる建築情報管理であ

り、すべての図面、カット、仕様書（BOM - 部品表）が同期されていることを保証する。

> BIM は、すべての更新とすべての変更がデータベースで行われる建築情報管理と表現される。そのため、回路図、断面図、シート図にかかわらず、すべてが常に調整され、一貫性があり、最新の状態に保たれます。
>
> - CAD 会社のウェブサイト BIM ホワイトペーパーを持つベンダー、2003 [54].

単一の統合されたデータベースでデザインを管理するという考え方は、1980 年代の研究において早くから広く議論されてきた。例えば、チャールズ・イーストマンの BDS コンセプト[57]には、「データベース」という用語への 43 の言及が含まれていた（図 6.1-2）。2004 年までに、この数はほぼ半減し、2002 年の BIM に関するホワイトペーパー [64]では 23 になった。そして 2000 年代半ばまでには、データベースの話題は、ベンダーのマーケティング資料やデジタル化アジェンダ全般から事実上消えていた。

BIM システムの中核として当初考えられていたのはデータベースとそれへのアクセスであったが、時が経つにつれ、その重点はジオメトリ、視覚化、3D に移ってきた。1994 年に IFC 標準の登録者であり、2002 年に BIM ホワイトペーパーを発表したまさにそのベンダーは、2000 年代初頭のホワイトペーパーで、IGES、STEP、IFC のような中立フォーマットの限界と、CAD データベースへの直接アクセスの必要性を明確に指摘している：

> 異なるアプリケーションは互換性がなく、再入力されたデータは不正確かもしれない［...］。従来のコンピュータ支援設計[CAD]の結果：コスト高、市場投入までの時間、製品品質の低下。今日、主要なアプリケーションはすべて、低レベルのデータ交換に業界標準のインターフェースを使用しています。旧来の IGES 規格や新しい STEP（IFC は STEP/IGES フォーマットの事実上および実質上のコピー）を使用して異なるベンダーのアプリケーション間でデータを交換することにより、ユーザーはベスト・オブ・ブリードの製品間である程度のデータ互換性を実現することができます。しかし、IGES と STEP は低いレベルでしか機能せず、今日の主要なアプリケーションによって生成される情報ほど豊富なデータを交換することはできません。また、これらの標準やその他の標準は、ほぼ毎日改善されていますが、データの豊富さという点では、今日のベンダー製品に常に遅れをとるでしょう。[アプリケーション内のプログラムは、IGES、STEP [IFC]、PATRAN などの中立的なトランスレータに頼ることなく、データの豊かさを交換し、保持できなければなりません。その代わりに、フレームワーク・アプリケーションは、情報の詳細さと正確さが失われないように、基礎となる CAD データベースに直接アクセスできなければなりません。
>
> - CAD ベンダーのホワイトペーパー（IFC、BIM ）「統合された設計と製造：利点と根拠」、2000 年 [65].

このように、1980 年代から 2000 年代初頭にかけてすでに、CAD 環境におけるデジタル設計の重要な要素は、フォーマット・ファイルや中立的な IFC フォーマットではなく、データベースであると考えられていた。トランスレータは廃止され、アプリケーションがデータに直接アクセスできるようになることが提案された。しかし現実には、2020 年代半ばまでに、BIM（）のコンセプトは、オープンな情報交換の開発よりも、クローズドな幾何学カーネルを使用するソフトウェアベンダーの利益が優先される「分割統治」戦略に似てきた。

今日、BIM は建設業界に不可欠なものとして認識されている。しかし、過去 20 年間、簡素化されたコラボレーションとデータ統合の約束は、ほとんど実現されていない。ほとんどのソリューションは、いまだにクローズ

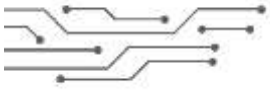
ドまたは中立的なフォーマットと特殊なツールに縛られている。本書「CAD と BIM：マーケティング、現実、そして建設における設計データの未来」のパート 6 では、BIM の歴史、オープン BIM、IFC、そして相互運用性とジオメトリックカーネルの問題について詳しく見ていきます。

> 今日、業界は、CAD（BIM） のモデリングツールとしての従来の理解から、本格的なデータベースとしての利用へと移行するという重要な課題に直面している。これには、情報を扱う新しいアプローチが必要であり、閉じたエコシステムへの依存を捨て、オープンなソリューションを導入する必要がある。

CAD データベースへのアクセスを可能にするリバースエンジニアリングツールの開発や、オープンソースや LLM 技術の普及により、建設業界のユーザーや開発者は、ソフトウェアベンダーの曖昧な用語からますます離れつつある。その代わりに、本当に重要なもの、つまりデータ（データベース）とプロセスに焦点が移りつつある。

> 流行の頭字語やビジュアライゼーションの背後には、保存、転送、変換といった標準的なデータ管理手法がある。すなわち、古典的な ETL プロセス （Extract, Transform, Load ） である。他の産業と同様、建設業のデジタル化には、交換標準だけでなく、異種情報の明確な構造化された取り扱いが必要である。

CAD（BIM ） データの可能性を十分に活用するために、企業は情報管理へのアプローチを再考する必要がある。これは必然的に、デジタルトランスフォーメーションの重要な要素である、建設専門家が日常的に扱うデータの統一、標準化、意味のある構造化につながる。



章 3.2.

データのハーモナイゼーションと構造化

建設業界におけるデータ充填システム

大企業であろうと中堅企業であろうと、専門家たちは日々、さまざまなインターフェイスを持つソフトウェアシステムやデータベースに、多種多様な情報を詰め込み（図 3.2-1）、管理者の助けを借りながら、互いに協力し合わなければならない。相互作用するシステムとプロセスの複合体が、最終的に企業の収益と利益を生み出すのである。

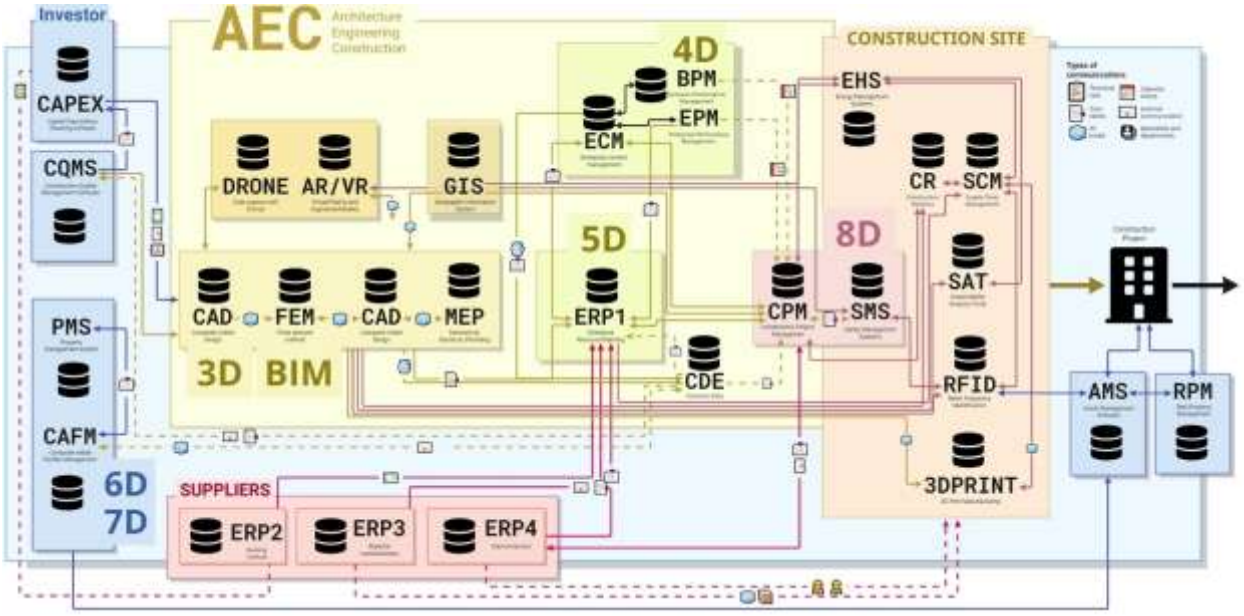


図 3.2-1 建設業で使用されるほぼすべてのシステムやアプリケーションは、その中核に一般的な RDBMS データベースのいずれかを備えています。

前述した建設産業で適用されるシステムの各カテゴリーは、これらのシステムの機能的役割に対応する独自のデータ型で動作する。抽象的なレベルから具体的なレベルに移行するために、データ型からフォーマットや文書としての表現に移行する。

先に示したシステムのリスト（図 1.2-4）に、今度は、そのシステムがよく扱う特定の種類のフォーマットと文書を追加する：

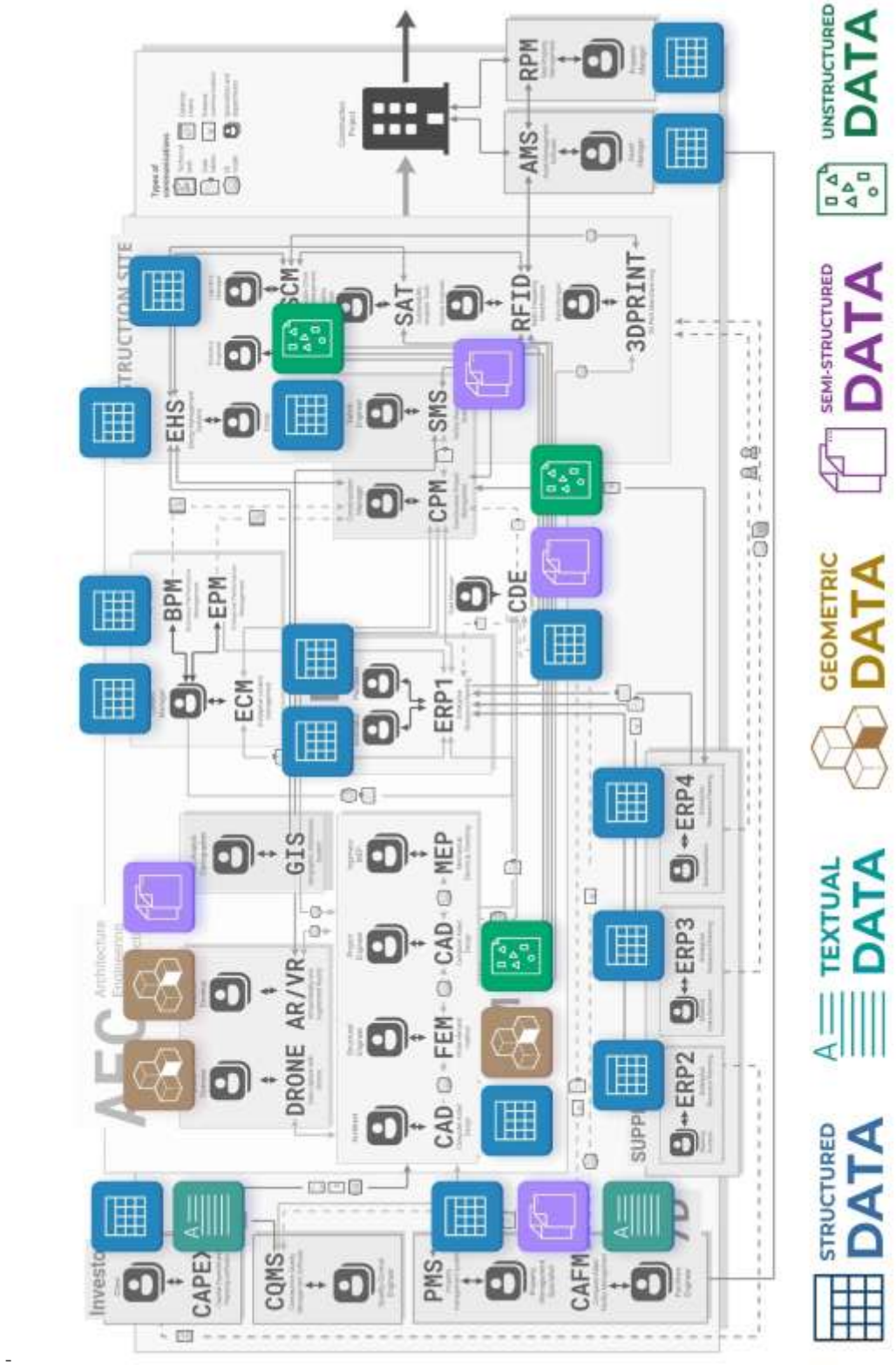


図 3.2-2 建設業界では、さまざまなタイプのデータを扱う、さまざまなインタフェースを持つ多くのシステムが使用されている。

- 投資家 (CAPEX)
  - ▪財務データ：予算、支出予測（構造化データ）。
  - ▪市場動向データ：市場分析（構造化データおよび非構造化データ）。
  - ▪法的および契約データ：契約（テキストデータ）。
- マネジメントシステム (PMS, CAFM, CQMS)
  - ▪ プロジェクトデータ：グラフ、タスク（構造化データ）。
  - ▪ 施設メンテナンスデータ：メンテナンス計画（テキストおよび半構造化データ）。
  - ▪ 品質管理データ：規格、検査報告書（テキストデータおよび非構造化データ）。
- CAD, FEM および BIM
  - ▪ 技術図面：建築図面、構造図面（幾何学データ、非構造化データ）。
  - ▪ 建物モデル：3D -モデル、材料データ（幾何学的および半構造化データ）。
  - ▪ 工学計算：荷重解析（構造化データ）。
- 建設現場管理システム（EHS、SCM）
  - ▪ 安全衛生データ：安全プロトコル（テキストデータおよび構造化データ）。
  - ▪ サプライチェーンデータ：在庫、注文（構造化データ）。
  - ▪ 日報：労働時間、生産性（構造化データ）。
- ドローン、AR/VR 、GIS、3D - プリント
  - ▪ ジオデータ：地形図（幾何学的データと構造化データ）。
  - ▪ リアルタイムデータ：ビデオや写真（非構造化データ）。
  - ▪ 3D プリント用モデル：デジタル図面（幾何学データ）。
- 追加管理システム (4D BPM, 5D ERP1)
  - ▪ 時間とコストのデータ：スケジュール、見積もり （構造化データ）。
  - ▪ 変更管理：プロジェクトの変更記録（テキストと構造化データ）。
  - ▪ パフォーマンス報告：成功の指標（構造化データ）。
- データ統合と通信（CDE, RFID, AMS, RPM)
  - ▪データ交換：文書交換、データモデル（構造化データとテキストデータ）。
  - ▪RFID および追跡データ：物流、資産管理（構造化データ）。

▪監視と制御：現場でのセンサー （構造化データと非構造化データ）。

このように、現場管理システムから業務用データベースまで、建設業界の各システムは、構造化された情報、テキスト情報、幾何学的な情報など、それぞれのタイプの情報を使って動作している。専門家が日常的に扱わなければならない「データ・ランドスケープ」は、極めて多様である。しかし、単純に形式を列挙しただけでは、実際の情報作業の複雑さは見えてこない。

実際、企業は、システムから取得したデータであっても、「そのまま」使える状態ではないという事実に直面している。これは特に、テキスト、画像、PDF、CAD ファイルなど、標準的なツールでは分析が困難な形式について当てはまります。このため、次の重要なステップはデータ変換です。このプロセスなしでは、処理、分析、視覚化、意思決定を効果的に自動化することはできません。

データ変換：現代のビジネス分析の重要な基盤

今日、ほとんどの企業はパラドックスに直面している。日常的なプロセスの約 80%は、いまだに古典的な構造化データ、つまり使い慣れたエクセルスプレッドシート（ ）やリレーショナルデータベース（RDBMS）に依存している[66]。しかし同時に、企業のデジタル・エコシステムに入ってくる新しい情報の 80%は、非構造化または疎構造化されている（図 3.2-3）[52]。これには、テキスト、グラフィック、ジオメトリ、画像、CAD - モデル、PDF の文書、音声やビデオの記録、電子的な通信文などが含まれる。

さらに、非構造化データの量は急速に増え続けており、その年間成長率は 55〜65%と推定されている[67]。このようなダイナミクスは、新しい情報を既存のビジネスプロセスに統合する上で深刻な困難を引き起こす。このような多形式データの流れを無視することは、情報格差の形成につながり、企業のデジタル環境全体の管理性を低下させる。



図 3.2-3 非構造化データが年々増加しているため、ストリーミング情報をビジネスプロセスに統合する際に課題が生じます。

> オートメーション・プロセスにおいて、複雑な非構造化データや混乱しやすい疎構造化データを無視すると、企業の情報ランドスケープに大きなギャップが生じる可能性がある。制御不能で雪崩のように情報が移動する今日の世界では、企業はあらゆるタイプのデータに対処する効果的な方法を含む、データ管理へのハイブリッド・アプローチを採用する必要がある。

効果的なデータ管理の鍵は、「バベル」（非構造化、テキスト形式、幾何学的形式を含み、構造化または疎構造化されたデータ）と呼ばれる 異なるタイプのデータを整理し、構造化し、分類することにある。このプロセスは、混沌としたデータセットをシステムに統合するための組織化された構造に変換し、それによってそれらに基づく意思決定を可能にする（図 3.2-4）。

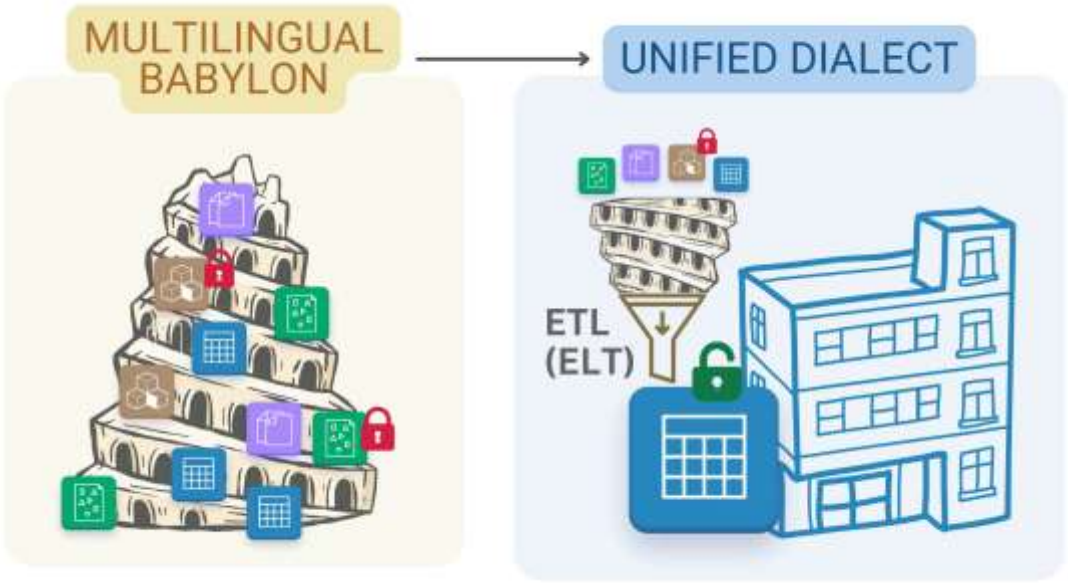


図 3.2-4 データ管理部門の主な仕事は、多様で多形式のデータの「バビロン」を、構造化され分類されたシステムに変換することである。

このようなハーモナイゼーションの主要な障害のひとつは、異なるデジタル・プラットフォーム間の相互運用性の低さである。

> 報告書によると、米国国立標準技術研究所（NIST、米国）は、異なるビル・プラットフォーム間のデータ互換性の低さが、情報の損失と多大な追加コストにつながることを強調している[68]。2002年だけでも、ソフトウェアの相互運用性の問題により、米国の資本建設における損失は年間 158 億ドルに上り、この損失の 3 分の 2 は、特に運用と保守の間、建物の所有者と運営者が負担している[68]。この研究では、データフォーマットの標準化によって、こうした損失を削減し、施設のライフサイクル全体を通じて効率を向上させることができるとも指摘している。

世界中の 16,000 人のデータサイエンティストを対象とした 2016 年の CrowdFlower の調査[69]によると、主な問題は依然として「汚れた」多形式のデータである。この調査によると、最も価値のあるリソースは、最終的なデータベースや機械学習モデル ではなく、情報の準備に費やされた時間である。

分析担当者（ ）やデータマネジャーは、データのクリーニング、フォーマット、整理に最大 60 パーセントの時間を費やしている。5 分の 1 近くは、適切なデータセットの検索と収集に費やされる。データセットはサイロに隠されていることが多く、分析にはアクセスできない。また、モデリング、分析、予測、仮説の検証に直接費やされる時間は全体の 9%程度に過ぎない。残りの時間は、情報伝達、視覚化、レポート作成、情報源の調査に費やされる

平均して、マネージャーのデータ作業は以下のように配分される（図 3.2-5）：

- データのクリーニングと整理 (60%): クリーンで構造化されたデータを持つことで、 アナリストの作業時間を大幅に短縮し、タスクを完了するプロセスをスピードアップすることができる。
- データ収集 (19%): データサイエンスの専門家にとっての大きな課題は、関連するデータセットを見つけることである。多くの場合、企業のデータは無秩序に整理された「サイロ」に積み重ねられており、必要な情報にアクセスするのは難しい。
- モデリング／機械学習（9%）：顧客側のビジネス目標が明確でないために、しばしば妨げられる。明確なミッション・ステートメントの欠如は、最高のモデルであってもその可能性を否定しかねない。
- その他の業務（5%）：アナリストは、データ処理に加えて、リサーチ、さまざまな角度からのデータ調査、ビジュアライゼーションやレポートによる結果の伝達、プロセスや戦略の最適化の提案などを行わなければならない。

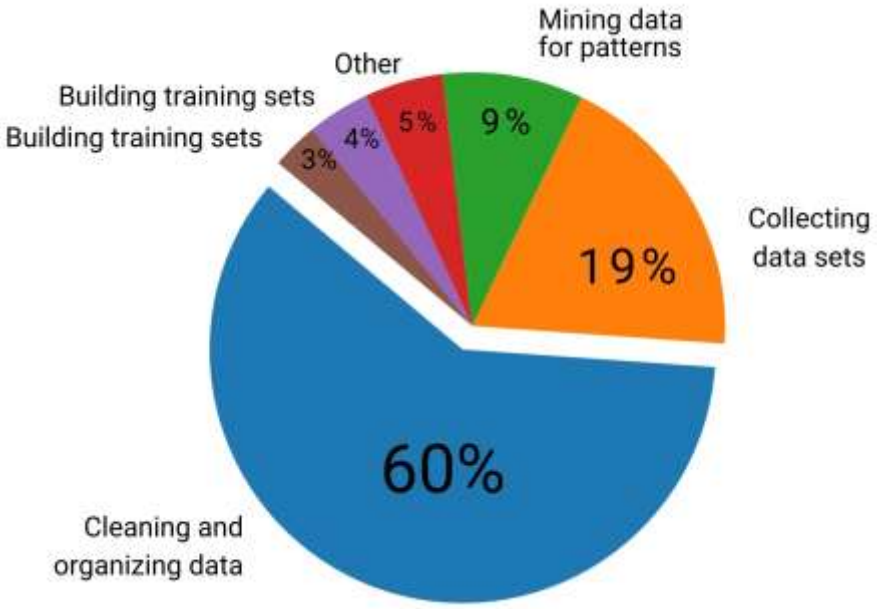


図 3.2-5 データを扱うデータ管理者が最も時間を費やすこと（[70]に基づく）。

> これらの推定は、他の研究によっても裏付けられている。2015 年に BizReport に掲載された Xplenty の調査[71]によると、ビジネスインテリジェンス（BI）専門家の時間 （BI）の 50%から 90%は、分析のためのデータ準備に費やされている。

データのクリーニング、検証、整理は、すべての下流データおよび分析プロセスの重要な基盤であり、データサイエンティストの時間の 90%を占める。

エンドユーザーには見えないこの骨の折れる作業が重要なのだ。生データのエラーは必然的に分析を歪め、誤解を招き、コストのかかる管理ミスにつながる。そのため、データクリーニングと標準化プロセス（ ）-重複の排除や抜け漏れの記入から、測定単位の調和や共通モデルへの整合まで-は、現代のデジタル戦略の礎石となりつつある。

したがって、データの徹底的な変換、クリーニング、標準化は、専門家の時間の大部分（データを扱う作業の最大 80%）を占めるだけでなく、最新のビジネス・プロセスの枠組みの中でデータを効果的に利用できるかどうかも左右する。しかし、データの整理とクリーニングだけで、企業の情報フローを最適に管理する作業が完了するわけではない。整理と構造化の段階で、適切なデータモデルを選択することが、その後の処理段階における情報操作の利便性と効率性に直接影響する。

データとビジネスの目的は異なるため、データモデルの特性を理解し、適切な構造を選択または作成できるようにすることが重要である。構造化の程度や要素間の関係の記述方法によって、構造化、疎構造化、グラフィ

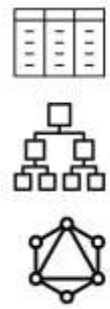
カルという 3 つの主要なモデルがある。それぞれ異なるタスクに適しており、長所と短所がある。

データモデル：データ内の関係と要素間の関係

情報システムにおけるデータは、情報の保存、処理、送信のタスクや要件に応じて、さまざまな方法で整理される。情報が格納される形式であるデータモデルのタイプ間の重要な違いは、構造化の程度と要素間の関係を記述する方法である。

> 構造化データ は、明確で反復可能なスキーマを持っている：それは、固定されたカラムを持つテーブルとして編成されている。このフォーマットは、SQL クエリ、フィルタリング、集計を実行する際に、予測可能性、処理の容易さ、効率性を提供する。例-データベース（RDBMS）、Excel 、CSV 。

ゆるやかに構造化されたデータ は、柔軟な構造を可能にする。異なる要素は、異なる属性セットを含み、階層として保存することができる。例としては、JSON 、XML などの文書形式がある。このようなデータは、入れ子になったオブジェクトやオブジェクト間の関係をモデ ル化する必要がある場合に便利ですが、その反面、データ分析や標準化が複雑になります（図 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

図 3.2-6 データ・モデルとは、システムにおいてデータがどのように整理され、保存され、処理されるかを記述する論理構造である。

適切な形式の選択は目的によって異なる：

- フィルタリングと分析のスピードが重要な場合は、 - リレーショナルテーブル（SQL 、CSV、RDBMS

、カラム型データベース）で十分です。

- 構造の柔軟性が必要な場合は、JSON または XML を使用することをお勧めします。
- データに複雑な関係がある場合、グラフデータベース は可視性と拡張性を提供する。

古典的なリレーショナル・データベース（RDBMS）では、各エンティティ（例えばドア）は行で表 され、そのプロパティはテーブルの列で表される。例えば、カテゴリー "ドア "の項目の表は、ID、高さ、幅、耐火性、および部屋を示す部屋 ID のフィールドを含むことができる（図 3.2-7）。

古典的なリレーショナル・データベース（RDBMS ）では、関係はテーブルの形で形成され、各レコードはオブジェクトを表し、列はそのパラメータを表します。表形式では、プロジェクトのドアに関するデータは以下のようになり、各行が個別の要素、つまり一意の識別子と属性を持つドアを表し、部屋との接続はパラメータ "部屋 ID "を通じて行われます。



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

図 3.2-7 プロジェクトの「ドア」カテゴリーの 3 つの要素に関する情報を表形式で構造化したもの。

JSON や XML のような疎構造化フォーマットでは、データは階層化または入れ子構造で保存され、要素に他のオブジェクトが含まれたり、その構造が変化したりする。これにより、要素間の複雑な関係をモデル化することができる。構造化された形式（図 3.2-7）で記録されたプロジェクト内のドアに関する同様の情報は、疎 構造化された形式（JSON ）で表現され（図 3.2-8）、Rooms（Rooms-ID）内のネストされたオブジェクトとなり、論理的 に階層を反映する。

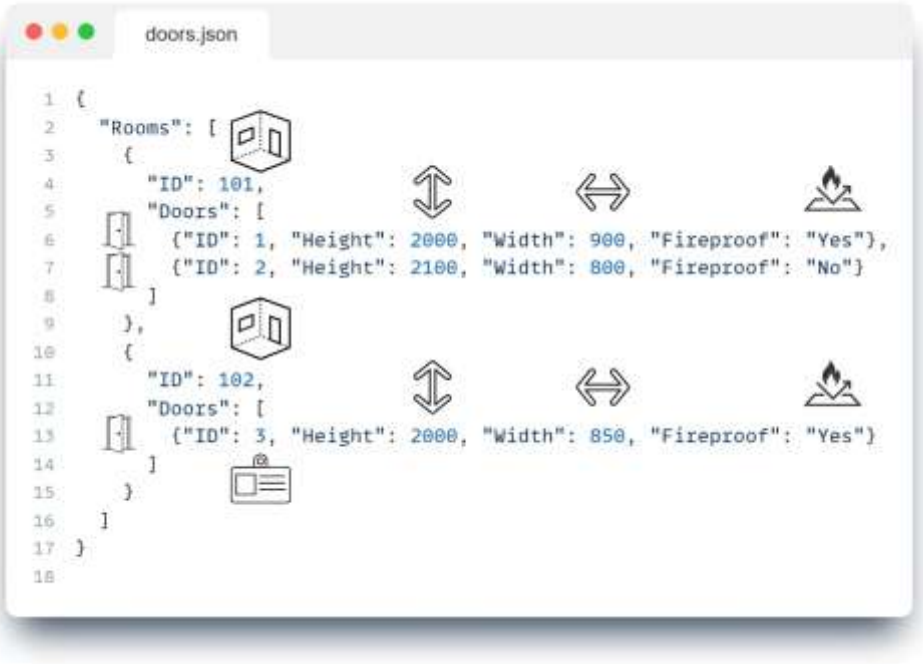


図 3.2-8 プロジェクトの「ドア」カテゴリーの要素に関する JSON 形式の情報.

グラフモデルでは、データはノード（頂点）とノード間のリンク（エッジ）として表現される。これにより、オブジェクトとその属性間の複雑な関係を視覚化することができます。プロジェクトのドアと部屋のデータの場合、 のグラフ表現は以下のようになります：

- ノード(節点)は主なエンティティを表す：部屋(101 号室、102 号室)とドア(ID1001、ID1002、ID1003)
- リブ（リンク）は、これらのエンティティ間の関係を示す。
- 属性はノードにマッピングされ、エンティティのプロパティ（高さ、幅、ドアの耐火性）を含む。

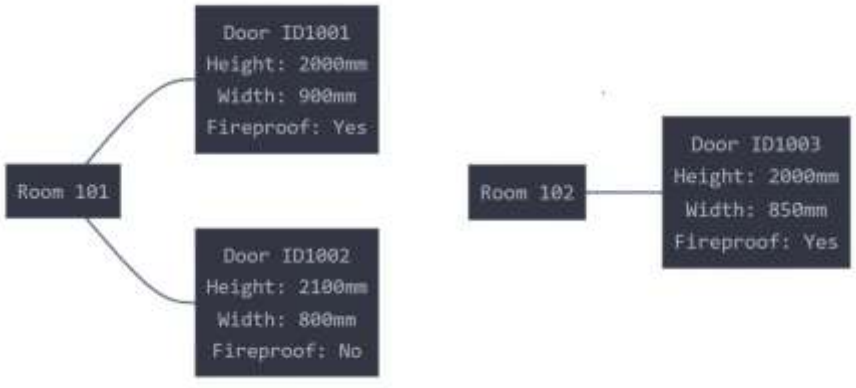


図 3.2-9 グラフ表示のプロジェクト・ドア・エンティティ情報

ドア記述のグラフ・データ・モデルでは、各部屋と各ドアは別々のノードである。ドアはエッジを通して部屋にリンクされ、ドアが特定の部屋に属することを示す。ドアの属性（高さ、幅、耐火性）は、対応するノードのプロパティとして格納される。グラフ・フォーマット の詳細と、建設業界でグラフ・セマンティクスがどのように登場したかについては、「建設業界における セマンティクスとオントロジーの登場」の章で説明する。

グラフ・データベース は、レコメンダー・システムやルーティング・システム、施設管理プロジェクトにおける複雑な関係のモデリングなど、重要なのはデータそのものではなく、データ間の関係である場合に有効である。グラフ・フォーマットは、ストレージ構造を変更することなく新しいデータタイプをグラフに追加できるため、新しい関係の作成が簡単になる。しかし、リレーショナル・テーブルや構造化フォーマットと比較すると、グラフではデータの接続性が追加されることはない。2 次元データベースのデータをグラフに転送しても、関係の数は増えないし、新しい情報を得ることもできない。

> データの形式とスキーマは、特定のユースケースと実行されるタスクに合わせて調整されるべきである。ビジネス・プロセスで効果的に作業するためには、できるだけ迅速かつ容易に結果を得られるようなツールやデータ・モデルを使用することが重要である。

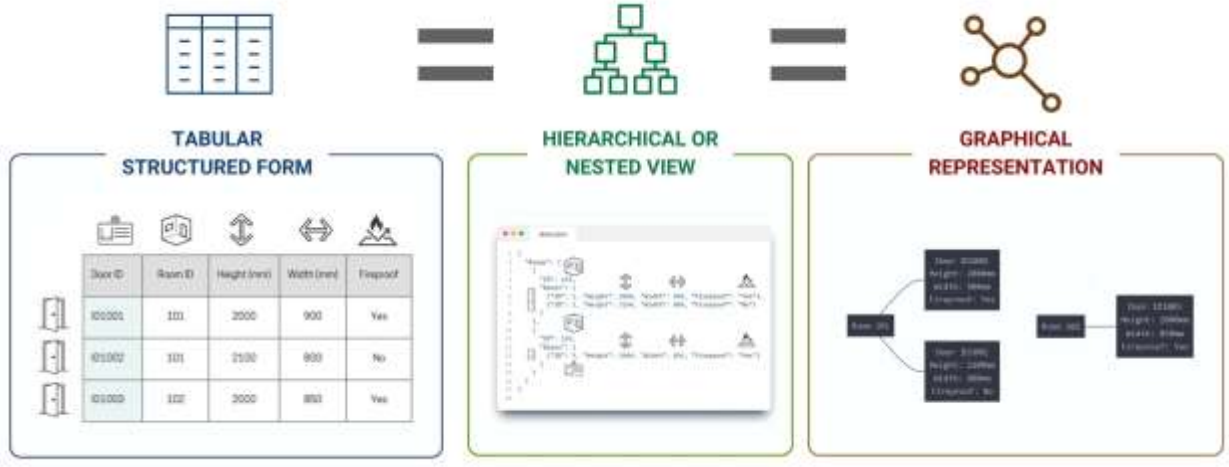


図 3.2-10 プロジェクト要素に関する同じ情報は、異なるデータモデルを使用して異なるフォーマットで保存することができます。

> 今日、ほとんどの大企業は過剰なデータの複雑さという問題に直面している。何百、何千というアプリケーションのそれぞれが独自のデータモデルを使用しているため、過度な複雑さが生じています。個々のモデルは必要以上に複雑であることが多く、すべてのモデルを集約すると何千倍も複雑になります。このような過剰な複雑さは、開発者とエンドユーザーの両方の作業を大幅に妨げます。

このような複雑さは、会社のシステムの開発とメンテナンスに深刻な制約を課す。モデルに新しい要素が追加されるたびに、コードの追加、新しいロジックの実装、徹底的なテスト、既存のソリューションへの適合が必要になる。このようなことはすべて、コストを増加させ、会社の自動化チームの作業を遅らせ、単純な作業でさえもコストと時間のかかるプロセスに変えてしまう。

複雑さはデータアーキテクチャのあらゆるレベルに影響する。リレーショナル・データベースでは、テーブルやカラムの数が増え、しばしば冗長になる。オブジェクト指向システムでは、クラスや相互に関連するプロパティの多重化によって複雑さが増す。XML や JSON のようなフォーマットでは、複雑さは、紛らわしい入れ子構造、ユニーク・キー、一貫性のないスキーマによって現れる。

> データモデルの過度な複雑さは、システムの効率を低下させるだけでなく、エンドユーザーの理解を困難にし、将来的には大規模な言語モデルや LLM エージェント。データモデルとデータ処理の理解と複雑さの問題は、「データを使いやすくして、実際にすぐに役立つようにするにはどうすればいいか」という問題を提起している。

データモデルを賢く選択しても、データへのアクセスが制限されていれば、その有用性は劇的に低下する。独自フォーマットやクローズド・プラットフォームは、統合を妨げ、自動化を複雑にし、独自情報のコントロールを奪う。問題の規模を理解するためには、クローズド・システムが建設業のデジタル・プロセスにどのような影響を与えるかを正確に検討することが重要だ。

独自フォーマットとデジタルプロセスへの影響

建設会社がデジタル化の過程で直面する主な課題のひとつは、データへのアクセスが限られていることだ。そのため、システムの統合が難しく、情報の質が低下し、効率的なプロセスの編成が難しくなる。独自のフォーマットや閉鎖的なソフトウェア・ソリューションの使用が、こうした困難の根底にあることが多い。

残念なことに、これまで建設業界で使用されてきた多くのプログラムでは、ユーザーがデータを独自のフォーマットやクラウドストレージに保存することしかできず、それらは厳密に限定されたインターフェイスを介してしかアクセスできない。また、これらのソリューションは、大手ベンダーのさらにクローズドなシステムに依存して構築されていることも珍しくない。その結果、よりオープンなアーキテクチャを提供したい開発者でさえ、大手ベンダーの決めたルールに従わざるを得なくなっている。

> 最新の建設データ管理システムは、オープンなフォーマットや標準をサポートするようになってきていますが（図 3.1-5）、CAD（BIM ）ベースのデータベースや、関連する ERP や CAFM システムは、業界のデジタル景観において孤立した独自の「島」のままです（図 3.2-11）。



図 3.2-11 データの閉鎖的で専有的な性質は、データの統合とアクセスに障害をもたらす。

閉鎖的で独占的なフォーマットやプロトコルは、建設業界だけの問題ではない。経済の多くの分野で、閉鎖的な規格やデータへのアクセス制限との戦いは、技術革新の鈍化（図 3.2-12）、新規参入者に対する人為的な障壁の存在、大手サプライヤーへの依存の深化から始まっている。データの重要性が急速に高まる中、競争当局には新たなデジタル市場がもたらす課題に対応する時間的余裕はなく、その結果、閉鎖的なフォーマットやデータへのアクセス制限は、本質的に情報と成長の流れを制約するデジタルの「国境」となっている[63]。

> もし機械が私たちに必要なものすべてを生産するのであれば、私たちの状況はこれらの財がどのように分配されるかにかかっている。機械が生産した富が分配された場合にのみ、誰もが豊かな生活を享受できる。あるいは、自動車所有者が富の再分配に反対するロビー活動を成功させれば、ほとんどの人が貧困に喘ぐことになるだろう。今のところ、物事は 2 番目の方向に進んでいるようで、テクノロジーによって不平等がますます拡大している[72]。
>
> - スティーブン・ホーキング、宇宙物理学者、2015 年

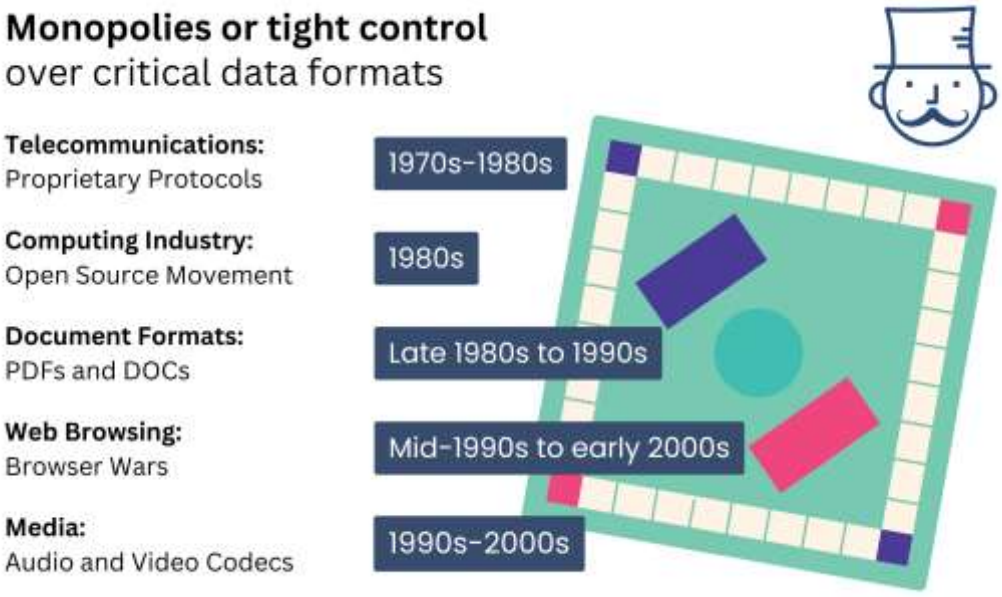


図 3.2-12 主要なデータ形式やプロトコルの独占的所有権は、建設業界だけの問題ではない。

その結果、データベース（ ）へのアクセスが閉ざされ、デー タマネージャー、データアナリスト（ ）、IT 専門家、開発者は、今日、 建設業界のデータアクセス、処理、自動化のためのアプリケーションを作成する際、ソフ トウェアベンダーへの多くの依存に直面している（図 3.2-13）。このような依存関係は、アクセス・レイヤーの追加という形で、特別な API - コネクションや特別なツールやソフトウェアを備えたソリューションの作成を必要とする。

> API（アプリケーション・プログラミング・インターフェース）とは、あるプログラムが他のプログラムと相互作用し、ソースコードにアクセスすることなくデータや機能を交換できる、形式化されたインターフェースのことである。API は、外部システムがどのようなリクエストをして、どのようなフォーマットで、どのようなレスポンスを受け取るかを記述する。これは、ソフトウェア・モジュール間の標準化された「契約」である。

クローズド・ソリューションへの依存が多いため、企業内のコード・アーキテクチャとビジネス・プロセス・ロジック全体が、データへの質の高いアクセスを提供するというソフトウェア・ベンダーの方針に依存したツールの「スパゲッティ・アーキテクチャ」になってしまう。

クローズドなソリューションやプラットフォームへの依存は、柔軟性を失うだけでなく、実際のビジネスリスクにもつながる。ライセンス条件の変更、データへのアクセスの閉鎖、フォーマットや API 構造の変更 - これらはすべて、重要なプロセスをブロックする可能性がある。突然、1 つのテーブルを更新するために、統合やコネクタのブロック全体を作り直す必要があることが判明し（図 3.2-13）、ソフトウェアやその API ベンダーの大規模な更新は、企業全体のシステムの安定性に対する潜在的な脅威となる。

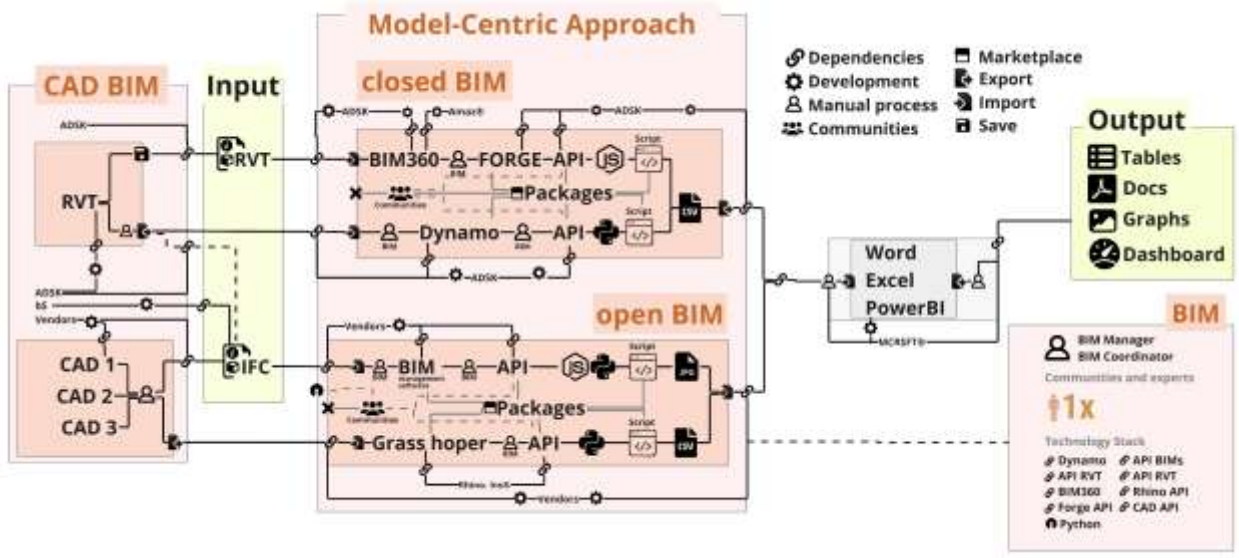


図 3.2-13 CAD 処理における依存関係の多さの一例 -データは、建設会社のエコシステムにおけるデータ統合の障壁となる。

> このような状況に置かれた開発者やシステム・アーキテクトは、期待するためではなく、生き残るために働くことを余儀なくされる。新しいソリューションを導入するのではなく、適応させる。開発する代わりに、互換性を維持しようとする。プロセスを自動化しスピードアップする代わりに、次のクローズド・インターフェースの研究、API ドキュメント、そして終わりのないコードの再構築に時間を費やす。

閉鎖的なフォーマットやシステムでの作業は、単なる技術的な課題ではなく、戦略的な制約でもある。現代の自動化、AI（）、LLM（）、予測分析（）が提供する明らかな機会にもかかわらず、多くの企業はその可能性を十分に実現できていない。また、独自のフォーマット（図 3.2-13）によって築かれた障壁は、企業が自社のデータにアクセスすることを妨げている。これこそ、建設業におけるデジタルトランスフォーメーションの皮肉と言えるかもしれない。

> データの透明性とオープンシステムは贅沢品ではなく、スピードと効率性の必須条件である。オープンでなければ、ビジネス・プロセスは不必要な官僚主義で満たされ、何層もの承認チェーンが存在し、HiPPO 原則 （最も高い報酬を得た人の意見に基づいて意思決定を行う）への依存が高まる。

とはいえ、パラダイム・シフトが起こりつつある。プロプライエタリなソリューションが支配的であるにもかかわらず、第 4 次産業革命に触発されたアーキテクチャの限界に気づく企業が増えつつある。今日、ベクトルは、戦略的資産としてのデータ、オープン・インターフェース（API）、システム間の真の相互運用性が中心となる第 5 次革命の原則へとシフトしている。

この移行は、クローズドなエコシステムから、オープンなフォーマット、標準、透明なデータ交換が鍵となる、柔軟でモジュール化されたデジタル・アーキテクチャへの移行を意味する。

オープン・フォーマットがデジタル化へのアプローチを変える

建設業界は、クローズドで専有的なデータの問題に最後に取り組んだ業界のひとつである。他の経済分野とは異なり、この業界ではデジタル化の進展が遅れている。その理由としては、この業界の伝統的な保守的な性質、ローカルでバラバラのソリューションが普及していること、紙ベースの文書管理が根強いことなどが挙げられる。何十年もの間、主要な建設プロセスは、物理的な図面、電話、同期化されていないデータベースに依存

していた。このような背景から、クローズド・フォーマットは長い間、障害というよりもむしろ規範として認識されてきた。

他の産業での経験から、クローズドデータに対する障壁を取り除くことは、イノベーションの急増、開発の加速、競争の激化につながることが示されている[73]。科学においては、オープンデータの交換は発見を加速し、国際協力を発展させる。医療では、診断と治療の効率を向上させることができる。ソフトウェア工学では、共創と迅速な製品改善のエコシステムを構築することができる。

> マッキンゼーの報告書「オープンデータ：Unlock Innovation and Productivity with Information Flow" 2013.[74]によると、オープンデータ は、建設、運輸、ヘルスケア、エネルギーを含む 7 つの主要産業において、年間 3 兆ドルから 5 兆ドルを解き放つ可能性を秘めている。同調査によると、分散型データ・エコシステムにより、大手建設会社や請負業者はソフトウェア開発とメンテナンスのコストを削減でき、デジタル導入を加速できる。

経済の他分野で始まって久しいオープン・アーキテクチャへの移行は、建設業界にも徐々に広がりつつある。大企業や公的顧客、特に建設プロジェクトへの投資を管理する金融機関は、オープンデータの利用や、計算・算出・アプリケーションのソースコードへのアクセスをますます求めるようになっている。開発者はもはや、デジタル・ソリューションを作成し、プロジェクトの最終的な数値を示すことだけを期待されているのではなく、透明性、再現性、サードパーティのアプリケーション・ベンダーからの独立性を求められているのだ。

> オープンソース・ソリューション（）を使用することで、外部の開発者が共同作業を中止したりプロジェクトから離脱したりしても、ツールやシステムのさらなる開発能力に影響を与えないという安心感が顧客にもたらされる。オープンデータの主な利点の 1 つは、アプリケーション開発者がデータにアクセスするための特定のプラットフォームへの依存を排除できることである。

企業がプロプライエタリなソリューションを完全に放棄できない場合、可能な妥協案はリバース・エンジニアリング技術の使用である。これらの合法的かつ技術的に 健全な方法によって、クローズドなフォーマットを、よりアクセスしやすく、構造化され、統合に適したものに変換することができる。これは、レガシー・システムとの接続や、あるソフトウェア・ランドスケープから別のソフトウェア・ランドスケープへの情報の移行が必要な場合に特に重要である。

> 建設業界におけるオープン・フォーマットへの移行とリバース・エンジニアリング（プロプライエタリ・システムの合法的ハッキング）の応用の歴史において、最も輝かしい例のひとつが、コンピュータ支援設計システム（CAD ）で広く使われている DWG フォーマット（）のオープン化を求める闘いの歴史である。1998 年、あるソフトウェアベンダーの独占に対抗して、他の 15 社の CAD ベンダーが「オープン DWG」と呼ばれる新しい同盟を結成し、プロプライエタリなソフトウェアやクローズドな API（）を必要とせずに、DWG フォーマット（図面転送の事実上の標準）を扱うためのフリーで独立したツールを開発者に提供した。この出来事は、1980 年代後半から今日に至るまで、何万もの企業が人気のある CAD ソリューションのクローズド・フォーマットに無料でアクセスできるようになり、CAD 市場における競争を促進する互換性のあるソリューションを生み出す転機となった[75]。今日、1996 年に初めて作成された「オープン DWG」SDK は、DWG フォーマット開発者の公式アプリケーショ ン以外でも、DWG フォーマットのインポート、編集、エクスポートが可能なほぼすべてのソリューションで使用されている。

他のハイテク大手も同様の変革を余儀なくされている。かつてはプロプライエタリなアプローチの象徴であったマイクロソフトは、.NET Framework のソースコードを公開し、Azure クラウドサービスのインフラで Linux を使い始め、オープンソースコミュニティでの地位を強化するために GitHub を買収した 。[76].メタ（旧フェイスブック）は、AI エージェント開発におけるイノベーションとコラボレーションを促進するため、ラマ・シリーズなどのオープンソース AI モデル をリリースした。マーク・ザッカーバーグ最高経営責任者（CEO）は、オープンソースの プラットフォームが、今後 10 年間の技術進歩をリードすると予想している[77]。

> オープンソース は、ソフトウェアの開発および配布モデルであり、ソースコードは自由に使用、研究、変更、配布できるように公開されている。

オープンデータとオープンソース・ソリューションは、単なるトレンドではなく、デジタルの持続可能性の基盤になりつつある。オープンデータとオープンソース・ソリューションは、企業に柔軟性、回復力、独自の意思決定に対するコントロール、そしてベンダーのポリシーに依存することなくデジタル・プロセスを拡張する能力を与える。そして同様に重要なことは、21 世紀で最も価値ある資源であるデータの管理権を企業に取り戻すことである。

パラダイムシフト：オープンソース ソフトウェアベンダー優位の時代の終焉として

建設業界は、通常の方法では収益化できない転換期を迎えている。データ主導、データ中心のアプローチ、 オープンソース ツールの使用というコンセプトは、市場のソフトウェア大手が拠って立つゲームのルールを再考することにつながっている。

> これまでの技術革新とは異なり、この移行はベンダーが積極的に推進するものではない。パラダイムシフトは、ライセンス、サブスクリプション、コンサルティングに基づく従来のビジネスモデルを脅かすものだからだ。新しい現実には、すぐに使える製品や有料のサブスクリプションは関係ない。

オープン・テクノロジーに基づくデータセンター・ソリューションを管理・開発するためには、企業は社内のプロセスを見直す必要がある。さまざまな部門の専門家が協力し合うだけでなく、連携方法も見直さなければならないだろう。

新しいパラダイムは、オープンデータとオープンソース・ソリューションの利用を意味し、プログラマーではなく、人工知能と大規模言語モデル（LLM）に基づくツールが、ソフトウェアコードの作成において特別な役割を果たすことになる。すでに 2024 年半ばまでに、グーグルでは新しいコードの 25%以上が AI で作成されている[78]。将来、LLM を使ったコーディングは、作業の 80%をわずか 20%の時間でこなすようになるだろう（図 3.2-14）。

> McKinsey の 2020 年調査[79]によると、GPU はその高いパフォーマンスと最新のオープンソースツールによるサポートにより、分析において CPU に取って代わることが増えている。これにより企業は、高価なソフトウェアや希少な専門家の雇用に多額の投資をすることなく、データ処理を高速化することができる。

マッキンゼー、PwC、デロイトなどの大手コンサルティング会社は、オープン・スタンダード、オープン・ソース アプリケーションの重要性が業界全体で高まっていることを強調している。

> PwC Open Source Monitor 2019」レポート[80]によると、従業員 100 人以上の企業の 69%が意識的に利用している。OSS は特に大企業で積極的に利用されている：従業員 200〜499 人の企業では 71%、従業員 500〜1999 人のカテゴリーでは 78%、従業員 2000 人以上の企業では 86%に達する。シノプシスの OSSRA 2023 レポートによると、分析対象となったコードベースの 96%にオープンソースコンポーネント [81]が含まれている。

これからの開発者の役割は、手作業でコードを書くことではなく、データモデルやフロー・アーキテクチャを設計し、オンデマンドで適切な計算を行う AI エージェントを管理することだ。ユーザー・インターフェースはミニマルになり、インタラクションはダイアログ・ベースになる。古典的なプログラミングは、デジタル・ソリューションのハイレベルな設計とオーケストレーションに取って代わられるだろう（図 3.2-14）。ローコードプラットフォーム（図 7.4-6）や LLM 対応エコシステム（図 7.4-4）といった現在のトレンドは、IT システムの開発・保守コストを大幅に削減する。

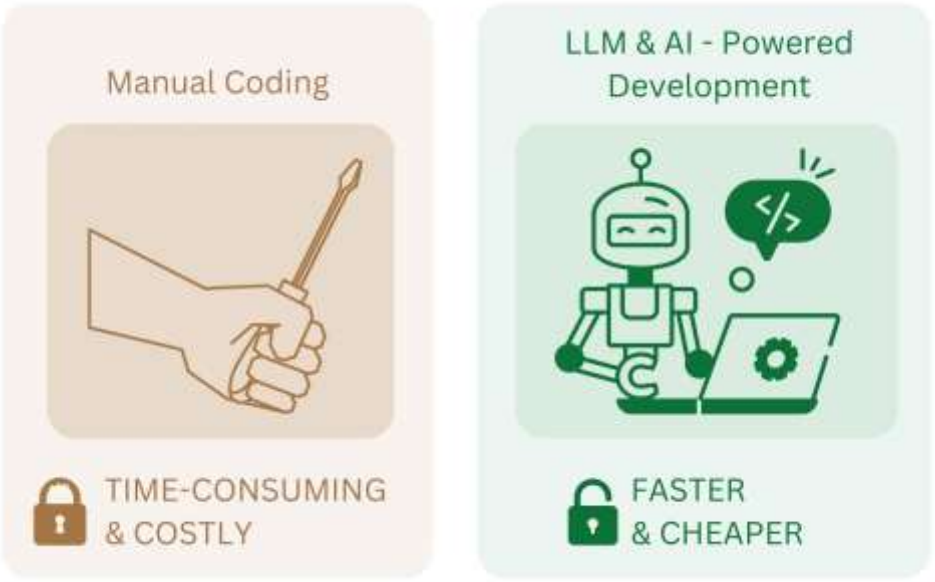


図 3.2-14 今日、アプリケーションはプログラマーによって手作業で作成されているが、将来的にはコードのかなりの部分が AI や LLM ベースのソリューション によって生成されるようになるだろう。

この移行はこれまでとは異なり、大手ソフトウェア・ベンダーが触媒になることはないだろう。

> ハーバード・ビジネス・スクールの研究「オープン・ソース・ソフトウェアの価値」2024 [40]では、オープン・ソース・ソフトウェアの総価値を 2 つの視点から見積もっている。一方では、既存のすべてのオープンソース・ソリューションをゼロから構築するのに必要な金額を計算すると、その金額は約 41 億 5000 万ドルになる。一方、既存のツールにアクセスすることなく、各企業が独自にオープンソース・ソリューションの類似品を開発することを想像すると（どこでも起こっていることだが）、ビジネスの総コストは 8 兆 8,000 億ドルという途方もない金額に達することになる--これが需要のコストである。

> 潜在的な市場価値が 8 兆 8000 億ドルもあるソフトウェア市場をわずか 41 億 5000 万ドルまで縮小させることに、大手ソフトウェア・ベンダーが興味を示さないことは想像に難くない。これは、需要量を 2,000 倍以上に減らすことを意味する。このような変革は、クローズドなソリューションへの顧客依存を長年維持することでビジネスモデルを構築してきたベンダーにとって、単純に採算が合わない。つまり、便利でオープンなターンキー・ソリューションを誰かが提供してくれると期待していた企業は失望するかもしれない。

オープンデジタルアーキテクチャーへの移行は、雇用や収益の損失を意味しない。それどころか、柔軟で適応力のあるビジネスモデルのための条件を作り出し、最終的には従来のライセンスや箱入りのソフトウェア市場に取って代わるかもしれない。

ライセンスの販売の代わりにサービス、クローズドなフォーマットの代わりにオープンプラットフォーム、ベンダーへの依存の代わりに独立性と真のニーズに対応したソリューションを構築する能力。これまでツールを使うだけだった人たちは、ツールの共同制作者になることができる。そして、データ、モデル、シナリオ、ロジックを扱うことができる人は、業界の新しいデジタル経済の中心にいることに気づくだろう。このような変化や、オープンデータをめぐってどのような新しい役割、ビジネスモデル、協力形態が生まれつつあるかについては、本書の最後の第 10 部で、さらに詳しくお話しする。

> オープン・データとオープン・コードに基づくソリューションによって、企業は時代遅れの API と格闘したり、クローズドなシステムを統合したりするよりも、ビジネス・プロセスの効率化に集中できるようになる。オープン・アーキテクチャへの意識的な移行は、生産性を大幅に向上させ、ベンダーへの依存を減らすことができる。

新しい現実への移行は、単にソフトウェア開発のアプローチを変えるだけでなく、データを扱う原理そのもの

を見直すことでもある。この変革の中心にあるのはコードではなく情報であり、その構造、アクセシビリティ、解釈可能性である。そこで、オープンで構造化されたデータが前面に押し出され、新しいデジタル・アーキテクチャーの不可欠な一部となる。

構造化されたオープンデータ：デジタルトランスフォーメーションの基盤

過去数十年間、ビジネスの持続可能性は、ソフトウェア・ソリューションの選択と特定のベンダーへの依存によってほぼ決定されていたが、今日のデジタル経済では、重要な要素はデータの品質とそれを効果的に扱う能力である。オープンソース・コードは新しい技術パラダイムの重要な一部であるが、その潜在能力は、データが理解可能で、整理され、機械可読である場合にのみ真に解き放たれる。あらゆる種類のデータモデルの中で、構造化されたオープンデータは、持続可能なデジタルトランスフォーメーションの礎石となりつつある。

構造化されたオープンデータの主な利点は、明確な解釈と自動処理の可能性である。これにより、個々の業務レベルでも、組織全体でも、大幅な効率化が可能になる。

> Deloitte のレポート "The Data Transfer Process in Enterprise Transformation"[82]によると、IT 部門と協力して構造化データを管理することは非常に重要である。[82]によると、構造化されたデータの転送を管理するために IT と協力することは非常に重要である。英国政府の報告書「Data Analytics and AI in Government Project Delivery」（2024 年）[83]によると、異なるプロジェクトや組織間のデータ共有の障壁を取り除くことが、プロジェクト管理の効率性を向上させる鍵である。この文書では、データ形式の標準化とオープンデータの原則の導入により、情報の重複を避け、時間のロスを最小限に抑え、予測の精度を向上させることができると強調している。

伝統的に断片化とフォーマットの多様性が高い建設業界にとって、構造化された統一プロセスと構造化されたオープンデータ は、首尾一貫した管理可能なプロセスを形成する上で重要な役割を果たす（図 4.1-14）。これにより、プロジェクト参加者は、クローズドなプラットフォーム、データモデル、フォーマット間の非互換性に関連する技術的な問題を解決することよりも、生産性を向上させることに集中することができる。

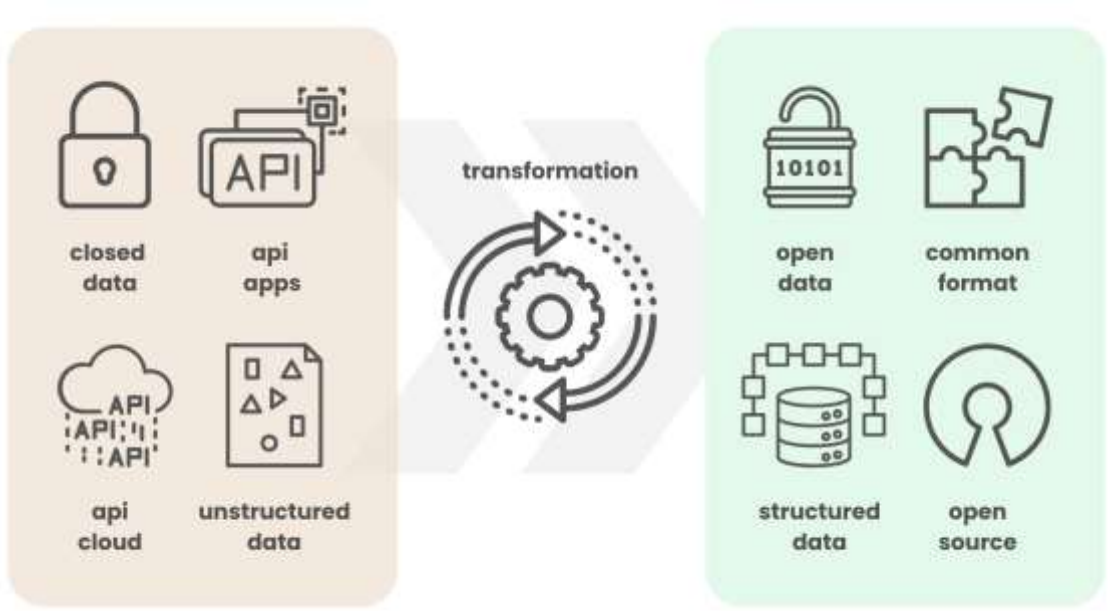


図 3.2-15 オープンな構造化データは、ソフトウェア・ソリューションやプラットフォームへの依存を減らし、イノベーションを加速する。

本書の後半で詳しく説明する最新のテクノロジー・ツールは、情報を収集するだけでなく、重複の排除、エラーの修正、値の正規化など、情報を自動的にクリーニングすることができる。これは、アナリストやエンジニアが、バラバラのドキュメントを扱うのではなく、分析、自動化、意思決定に適した、整理された知識ベースを扱うことを意味する。

> できる限りシンプルに、しかしこれ以上シンプルにならないように。
>
> - アルベルト・アインシュタイン、理論物理学者（引用の作者については議論がある[84]）。

今日、データを扱うためのほとんどのユーザー・インターフェースは、ビジネス・ケースごとに手作業でコードを記述する必要なく、自動的に作成することができます。そのためには、追加の指示なしにデータ構造、モデル、ロジックを理解するインフ ラ層が必要です（図 4.1-15）。このアプローチを可能にするのは、構造化されたデータです。フォーム、テーブル、フィルター、ビューは、最小限のプログラミング作業で自動的に生成することができます。

最も重要なユーザー・クリティカル・インターフェースは、まだ手作業による改良が必要かもしれない。しか

し、ほとんどの場合（これは作業シナリオの 50%から 90%）、この目的のために特別なアプリケーショ ン（ ）を使用しなくても、アプリケーションと計算の自動生成で十分である（図 3.2-16）。これにより、開発コストとメンテナンスコストが大幅に削減され、エラーが減少し、デジタルソリューションの実装がスピードアップする。

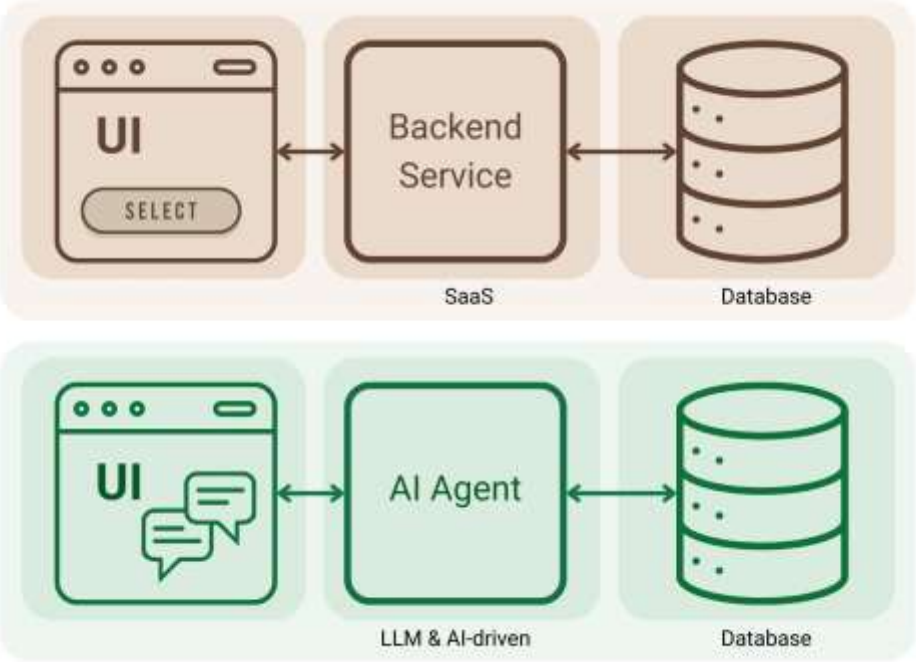
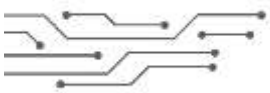


図 3.2-16 データを扱うためのアーキテクチャ・モデル：従来のアプリケーション・アーキテクチャと、LLM を用いた AI -指向モデル 。

個々のアプリケーションで構築されたアーキテクチャから、言語モデル（LLM）に基づくインテリジェントに管理されたシステムへの移行は、デジタル進化の次のステップである。このようなアーキテクチャーでは、構造化されたデータは単に保存の対象となるだけでなく、文脈に基づいて分析、解釈、行動を推奨できる AI ツールとのインタラクションの基盤となる。

以下の章では、オープンな構造化データに基づくアーキテクチャを実装した実例を見て、言語モデルがデータを自動的に解釈、検証、処理するためにどのように適用されているかを紹介する。これらの実践的なケースは、新しいデジタル・ロジックが実際にどのように機能するのか、そして変革の準備が整った企業にどのようなメリットをもたらすのか、理解を深めるのに役立つだろう。



章 3.3.

llm データ処理とビジネスプロセスにおける役割

LLM チャットルーム: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok データプロセスを自動化する。

ラージ・ランゲージ・モデル（LLM）の出現は、構造化されたオープンデータとオープンソースの哲学に向けた動きの自然な延長であった。データが整理され、アクセス可能になり、機械が読めるようになれば、次のステップは、複雑なコードを書いたり、専門的な技術的知識を持つことなく、この情報と対話できるツールである。

> LLM は、大規模なオープンデータセット、出版物、オープンソース運動など、オープン性の直接的な産物です。オープンな学術論文、公開されたテキストデータ、共同開発の文化がなければ、ChatGPT やその他の LLM は存在しなかっただろう。LLM は、ある意味で、オープンの原則によって集められ、教育された、人類が蓄積したデジタル知識の「蒸留物」なのです。

ChatGPT ®（OpenAI）、LlaMa ™（Meta AI ）、Mistral DeepSeek™ 、Grok ™（xAI）、Claude ™（Anthropic ）、QWEN™ のような最新の大規模言語モデル（LLM - Large Language Models）は、自然言語でデータに対するクエリを定式化する能力をユーザーに提供する。これにより、開発者だけでなく、アナリスト、エンジニア、プランナー、マネージャーなど、これまでプログラミングから遠ざかっていた専門家にとっても、情報を扱うことが身近になる

> LLM (Large Language Model)は、インターネット上から収集した膨大なデータに基づいてテキストを理解し、生成するように訓練された人工知能である。文脈を分析し、質問に答え、対話し、文章を書き、ソフトウェアコードを生成することができる。

以前は、データの可視化、処理、分析には、特殊なプログラミング言語（Python 、SQL 、R、Scala など）の知識と、Pandas 、Polars、DuckDB などのライブラリを扱う能力が必要でしたが、2023 年以降、状況は根本的に変わりました。そして、モデル自身がコードを生成し、実行し、表やグラフを表示し、結果を説明する。ここ数十年で初めて、テクノロジーの発展は複雑化ではなく、根本的な単純化とアクセシビリティの道を歩む

ことになった。

この原則-「言葉（プロンプト）でデータを処理する」。- は、情報操作の進化における新たな段階を示し、ソリューションの創造を事実上、さらに抽象度の高いレベルに引き上げた。かつて、ユーザーがオンラインショップを運営したり、ワードプレスやジュームラなどのオープンソースのモジュラーシステムを使ってウェブサイトを作ったりするのに、インターネットの技術的な基盤を理解する必要がなくなったのと同じように（avtor the book は 2005 年以来、教育やエンジニアリングのオンラインプラットフォームなど、このようなシステムに取り組んできた）。- その結果、デジタル・コンテンツとオンライン・ビジネスがブームとなり、今日、エンジニア、アナリスト、マネージャーは、プログラミング言語の知識がなくてもワークフローを自動化できるようになった。これは、LLaMA、Mistral、Qwen、DeepSeek などのフリーおよびオープンソースの強力な LLM によって促進されている。

大規模 LLM 言語モデル: その仕組み

ビッグ言語モデル（ChatGPT 、LlaMa 、Mistral 、Claude 、DeepSeek 、QWEN 、Grok ）は、インターネット、書籍、記事などの膨大なテキストデータに対して学習されたニューラルネットワークである。その主なタスクは、人間の音声の文脈を理解し、意味のある応答を生成することである。

現代の LLM は、Google の研究者が 2017 年に提案した Transformer アーキテクチャに基づいている[85]。このアーキテクチャの重要なコンポーネントはアテンションメカニズムであり、これによってモデルはテキスト内の位置に関係なく単語間の関係を考慮することができる。

> LLM の学習プロセス（）は、人間が言語を学習する方法に似ているが、その規模は数百万倍である。このモデルは何十億もの単語や表現の例を分析し、言語の構造や意味遷移の論理におけるパターンを特定する。テキスト全体はトークン、つまり最小の意味単位（単語やその部分）に分解され、多次元空間のベクトルに変換される（図 8.2-2）。これらのベクトル表現により、機械はテキストを単に記号列として扱うのではなく、概念間の隠れた関係を「理解」することができる。

ビッグ・ランゲージ・モデルは、単にテキストを生成するためのツールではない。意味を認識し、概念間のつながりを見つけ、異なるフォーマットで提示されたデータでも扱うことができる。重要なのは、情報が理解可能なモデルに分解され、LLM が扱うことのできるトークンとして表現されることである。

同じアプローチが建設プロジェクトにも適用できる。プロジェクトを一種のテキストと考え、それぞれの建物

、要素、工事をトークンとすれば、同じような方法でそのような情報を処理し始めることができる。建設プロジェクトは、建設プロジェクトの要素である最小限のトークンから構成されるカテゴリ、章、段落のグループに整理された書籍に例えることができます（図 3.3-1）。データモデルを構造化フォーマットに変換することで、構造化データをベクトルベース（図 8.2-2） に変換することもできます。ベクトルベースは、機械学習 や LLM などの技術にとって理想的なソースです。

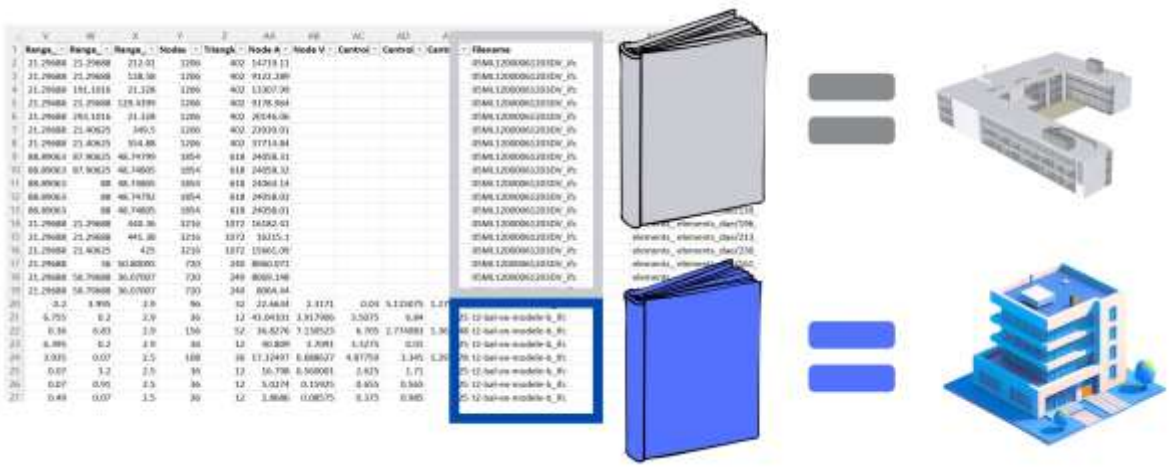


図 3.3-1 建設プロジェクト要素とは、文章におけるトークンのようなもので、プロジェクト全体を構成す るセクション（カテゴリー）のグループ（段落）を構成する最小単位である。

建設プロジェクトがデジタル化され、その要素がトークンやベクトルとして表現されていれば、形式的なクエリではなく、自然言語でアクセスすることが可能になる。クエリの意味を理解し、関連データにリンクさせる能力である。

エンジニアは、必要なデータを取得するために SQL（ ）や Python コード）を記述する必要はありません。LLM とデータ構造を理解し、「B30 以上のコンクリート等級を持つ鉄筋コンクリート構造物をすべて検索し、その総体積を計算する」という通常の方法でタスクを定式化するだけです。モデルはクエリの意味を認識し、それを機械可読形式に変換し、データを見つけ（グループ化し、変換する）、最終結果を返す。

文書、表、プロジェクトモデルはベクトル表現に変換され（エンベッディング ）、データベースに格納される。ユーザーが質問をすると、そのクエリもベクトルに変換され、システムが最も関連性の高いデータを探し出す。これによって LLM は、自身の学習した知識だけでなく、実際の企業データ（たとえそれがモデルの学習終了後に出現したものであっても）にも依存することができる。

建設における LLM の最も重要な利点の一つは、プログラムコードを生成する能力である。技術的なタスクをプログラマーに渡す代わりに、専門家がタスクを自然言語で記述すると、モデルが必要なコードを作成し、プロセス自動化コードの作成に（チャットからコピーして）使用することができます。LLM -モデルにより、深いプログラミング知識を持たない専門家でも、会社のビジネスプロセスの自動化と改善に貢献することができます。

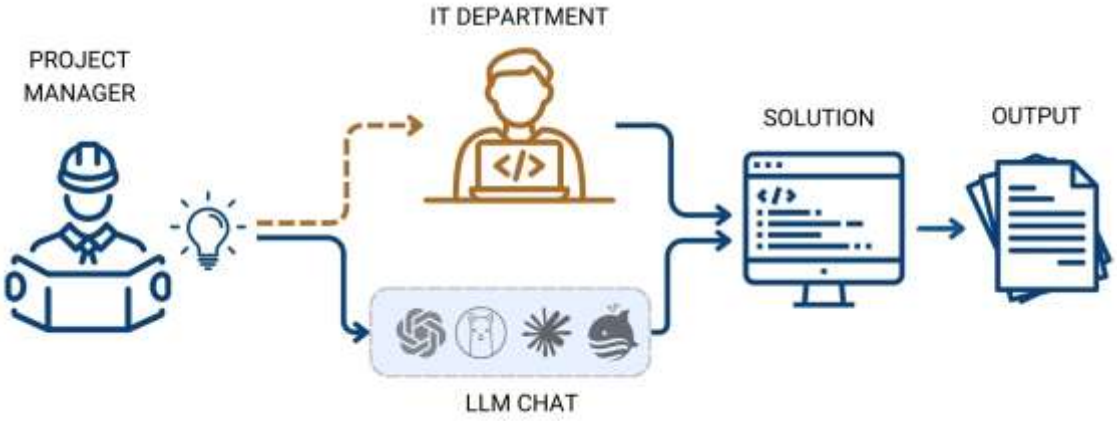


図 3.3-2 LLM ユーザーはプログラミングのスキルを必要とせずにコードを書き、結果を得ることができる。

> Wakefield Research が 2024 年に実施し、SAP がスポンサーとなった調査[36]によると、米国で年間売上高が 10 億ドル以上の企業の上級幹部 300 人を対象に行った調査では、上級幹部の 52%が、データを分析し、意思決定のための提言を提供する AI を信頼している。また、48%が AI を利用してこれまで把握できなかったリスクを特定し、47%が AI を利用して代替案を提案している。さらに、40%が新製品開発、予算計画、市場調査に AI を活用している。この調査では、AI が私生活に与えるポジティブな影響も示されており、回答者の 39%がワークライフバランスの改善、38%がメンタルヘルスの改善、31%がストレスレベルの低下を報告している。

しかし、LLM はその強力さゆえに、意識的に使用することが重要なツールであることに変わりはない。どんな技術にも限界がある。最もよく知られている問題のひとつは、いわゆる「幻覚」である。モデルが自信を持って、もっともらしいが事実としては正しくない答えを出すケースである。したがって、モデルがどのように機能するのかを理解することが重要である。どのようなデータやデータモデルを誤りなく解釈できるのか、クエリーをどのように解釈するのか、どこから情報を得るのか、などである。また、LLM の知識は訓練された時点

のものに限られ、外部データとの接続がなければ、モデルは現在の基準、標準、価格、技術を考慮しない可能性があることも忘れてはならない。

これらの問題の解決策は、ベクトル・データベースを定期的に更新し、関連ソースに接続し、自律型 AI を開発することである。このエージェントは、質問に答えるだけでなく、積極的にデータをトレーニングに利用し、タスクを管理し、リスクを特定し、最適化オプションを提案し、プロジェクトのパフォーマンスを監視する。

> 建設における LLM -インターフェースへの移行は、単なる技術的な目新しさではない。人とデータの間の障壁を取り除くパラダイム・シフトなのだ。私たちがお互いに話すのと同じくらい簡単に情報を扱うことができ、しかも正確で検証可能、実用的な結果を得ることができるのです。

このようなツールを他社よりも早く使い始めた企業は、大きな競争優位を得ることができる。これには、作業のスピードアップ、コスト削減、データ分析への迅速なアクセスや複雑な質問に対する答えを素早く見つける能力によるデザイン・ソリューションの品質向上などが含まれる。しかし、セキュリティの問題も考慮しなければならない。クラウドベースの LLM -サービスの利用は、データ漏洩のリスクを伴う可能性がある。そのため、組織は LLM ツールを自社のインフラに導入し、ローカルで情報を完全に保護・管理できる代替ソリューションを求めるようになっている。

ローカル LLM の利用 機密性の高い企業データのため

2022 年、最初のチャット LLM が登場し、人工知能 の発展に新たな段階が訪れた。しかし、これらのモデルが広く採用された直後、「企業関連のデータやクエリをクラウドに転送するのはどの程度安全なのか」という正当な疑問が生じた。ほとんどのクラウドベースの言語モデルは、通信履歴やアップロードされた文書をサーバーに保存しており、機密情報を扱う企業にとって、これは AI 導入の重大な障壁となっていた 。

> この問題に対する最も持続可能で論理的な解決策のひとつが、オープンソース LLM を企業の IT インフラ内にローカルに展開することである。クラウド・サービスとは異なり、ローカル・モデルはインターネット接続なしで動作し、データを外部サーバーに転送せず、企業は情報を完全にコントロールできる

> 最高のオープンモデル［Open Source LLM ］は、現在、クローズドモデル［ChatGPT、Claude ］に匹敵する性能を持っているが、約 1 年のタイムラグがある[77]。
>
> - 非営利研究組織エポック AI の主任研究員、ベン・コティエ（2024 年

大手テクノロジー企業は、自社の LLM を現地で利用できるようにし始めた。Meta のオープンソース LLaMA シリーズや、中国の急成長中の DeepSeek プロジェクトは、オープン・アーキテクチャへの移行の一例である。また、ミストラル（Mistral） やファルコン（Falcon）も、プロプライエタリなプラットフォームの制約から解放された強力なモデルをリリースしている。これらのイニシアチブは、グローバルな AI の開発を加速させただけでなく、プライバシーを重視する企業に、独立性、柔軟性、セキュリティコンプライアンスのための真の選択肢を与えた。

企業環境、特に建設業界では、データ保護は単なる利便性の問題ではなく、法令遵守の問題です。入札書類、見積書、図面、機密文書などを扱うには、厳格な管理が必要です。そこで、現地の LLM 、データが会社周辺に留まることを保証するために必要な保証を提供します。

	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

図 3.3-3 ローカルモデルは完全な制御とセキュリティを提供し、クラウドベースのソリューションは容易な統合と自動更新を提供する。

ローカルのオープンソースの主な利点 LLM ：

- データの完全管理。すべての情報は社内に留まり、不正アクセスやデータ漏洩を防ぎます。
- オフライン操作。インターネット接続に依存しないため、孤立した IT インフラにとって特に重要です。また、制裁措置やクラウドサービスがブロックされた場合でも、中断のないオペレーションを保証します。
- アプリケーションの柔軟性。このモデルは、テキスト生成、データ分析、プログラムコード作成、設

計支援、ビジネスプロセス管理に使用できる。

- 企業目的への適応 LLM は、社内文書に基づいてトレーニングすることができ、企業の業務や業界の特徴を考慮することができます。ローカルの LLM は、CRM 、ERP 、BI プラットフォーム に接続することができ、顧客からの問い合わせの分析、レポートの作成、さらにはトレンド予測の自動化を可能にします。

> DeepSeek の無料オープンソースモデル -R1-7B をサーバー上に展開し、ユーザーチーム全体がアクセスできるようにすれば、月額 1000 ドルのコストで、ChatGPT や Claude のようなクラウド API の年間利用料よりも低コストで済む可能性があり、企業はデータを完全に管理できるようになり、インターネットへの転送が不要になり、GDPR のような規制要件に準拠できるようになる

他の業界では、地元の LLM（ ）がすでに自動化へのアプローチを変えている。サポート・サービスでは、頻繁な顧客からの問い合わせに対応し、オペレーターの仕事量を減らしている。人事部では、履歴書を分析し、適切な候補者を選ぶ。e コマースでは、ユーザーデータを明かすことなく、パーソナライズされたオファーを生成する。

建設分野でも同様の効果が期待できる。LLM をプロジェクトデータや標準と統合することで、書類作成の迅速化、見積もり作成や予測コスト分析の自動化が可能になる。構造化テーブルやデータフレームと連携した LLM の利用は、特に有望な分野となりつつある。

社内における AI の完全制御と独自の LLM の展開方法

最新のツールを使えば、企業は大規模な言語モデル（LLM）をわずか数時間でローカルに展開できる。これにより、データとインフラストラクチャを完全に制御できるため、外部のクラウドサービスに依存する必要がなくなり、情報漏えいのリスクを最小限に抑えることができます。このソリューションは、機密性の高いプロジェクト文書や商業上の機密データを扱う組織に特に適しています。

タスクやリソースに応じて、すぐに使えるソリューションからより柔軟でスケーラブルなアーキテクチャまで、さまざまな導入シナリオが用意されている。最も簡単なツールの一つは Ollama で、深い技術的知識がなくても、文字通りワンクリックで言語モデルを実行することができます。Ollama のクイック・スタート

1. お使いのオペレーティング・システム（Windows / Linux / macOS）に対応したディストリビューションを公式ウェブサイトからダウンロード： ollama.com

2. コマンドラインでモデルをインストールする。例えば、Mistral モデルの場合：

```
オラマ・ラン・ミストラル
```

3. ターミナル経由でテキストクエリを送信したり、他のツールに統合することができます。モデルを実行し、クエリを実行します：

```
ollama run mistral "100mm 幅の石膏ボードパーティションウォールを設置するための作業に関するすべてのリソースを含む計算書を作成するには？
```

使い慣れたビジュアル環境での作業を好む人には、ChatGPT を彷彿とさせるインターフェイスを持つ無料アプリケーション、LM Studio がある

- 公式サイト lmstudio.ai から配布キットをダウンロードし、LM Studio をインストールする。
- 内蔵のカタログからモデル（例：ファルコンまたは GPT-Neo-X）を選択し、ダウンロードします。
- ChatGPT を彷彿とさせる直感的なインターフェイスでモデルを操作できますが、完全にローカライズされています。

	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

図 3.3-4 一般的なローカル・オープンソース LLM -モデルの比較。

モデルの選択は、速度、精度、および利用可能なハードウェア機能の要件に依存する（図 3.3-4）。Mistral 7B や Baichuan 7B のような小型モデルは、軽量タスクやモバイル機器に適しています。一方、DeepSeek -V3 のような強力なモデルは、多大な計算リソースを必要としますが、高いパフォーマンスと多言語のサポートを提供します。今後数年間で、LLM 市場は急速に成長し、軽量で特殊化されたモデルがますます増えていくだろう。すべての ヒューマンコンテンツをカバーする汎用の LLM の代わりに、狭いドメインの専門知識に基づいて訓練されたモデルが登場するだろう。例えば、工学的な計算、建設の見積もり、CAD 形式のデータなどを扱うためだけに設計されたモデルの出現が予想される。このような専門化されたモデルは、より速く、より正確で、より安全に使用することができる。特に、高い信頼性と主題の深さが重要視される専門的な環境ではそうである。

ローカルの LLM が立ち上げられたら、それを企業固有のタスクに適合させることができる。この目的のために、内部文書、技術指示書、契約書テンプレート、プロジェクト文書でモデルをさらに学習させる微調整技法が使用される。

RAG: インテリジェント LLM -企業データにアクセスできるアシスタント

ビジネスにおける LLM アプリケーションの進化における次の段階は、実際のリアルタイムの企業データとモデルの統合である。このアプローチは、RAG（Retrieval-Augmented Generation）-検索-拡張-生成と呼ばれる。このアーキテクチャでは、言語モデルは単なる対話インターフェースではなく、文書、図面、データベースをナビゲートし、正確で文脈に沿った回答を提供できる本格的なインテリジェント・アシスタントになる。

> RAG の主な利点は、高い精度と柔軟な情報の取り扱いを維持しながら、モデルを事前に訓練することなく、社内のデータを利用できることである。

RAG テクノロジーは、2 つの主要コンポーネントを組み合わせている：

- 検索： モデルは、データストア（文書、表、PDF - ファイル、図面）に接続し、ユーザーの要求に応じて関連情報を検索します。
- 拡張された生成：抽出されたデータに基づいて、モデルはクエリのコンテキストと特異性を考慮して、正確で情報に基づいた応答を生成します。

LLM with RAG support を実行するには、いくつかのステップを踏む必要がある：

- データの準備：必要な文書、図面、仕様書、表を収集する。PDF から Excel まで、さまざまな形式や構造のものがある。
- 索引付けとベクトル化：LlamaIndex や LangChain などのツールを使って、データをベクトル表現に変換し、テキスト断片間の意味的リンクを見つけることができる（ベクトルデータベース 、CAD プロジェクトを含む大規模配列をベクトル表現に変換することについては、第 8 回で詳しく説明する）。
- アシスタントに問い合わせる：データがアップロードされたら、モデルに質問することができる。 、インターネットから収集した一般的な知識ではなく、企業の枠組みの中で答えを検索する。

ある会社に constructionsite_docs フォルダがあり、契約書、指示書、見積書 、表が保存されているとする。Python スクリプト（図 3.3-5）を使って、このフォルダをスキャンし、ベクトル索引を構築することができる。各文書は、テキストの意味内容を反映したベクトルの集合に変換される。これによって文書は、モデルが効率的にナビゲートして用語や語句間のつながりを見つけることができる、一種の「意味の地図」となる。

たとえば、契約書の建設現場への資材の発送に関する部分に「返品」や「苦情」という言葉がよく出てくるこ

とを、このモデルは「記憶」している。そして、例えば、"返品期間はどのくらいですか？"といった質問があれば、それを答える。(図 3.3-5-コードの 11 行目） - LLM は社内文書を分析し、正確な情報を見つけ、すべての企業ファイルの内容を読み取り、理解することができる知的アシスタントのように動作します。

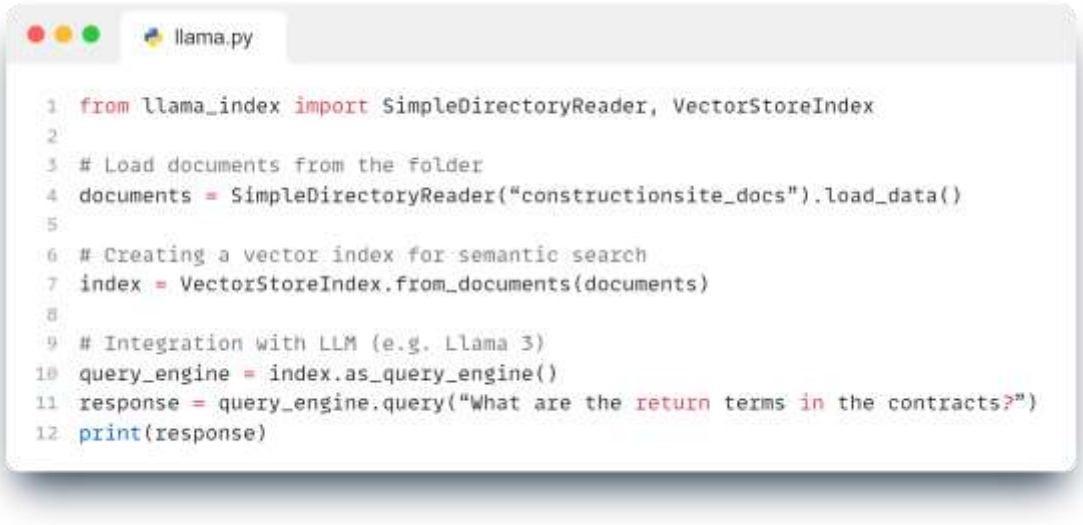
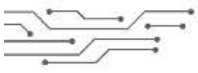


図 3.3-5 LM はファイルのあるフォルダーを読み込む。人がフォルダーを開いて必要な文書を探すのと同様である

このコードは、Python がインストールされたどのコンピューターでも実行できる 。Python と IDE を使ってコードを実行する方法については、次の章で詳しく説明します。

LLM のローカル展開は単なるトレンドではなく、セキュリティと柔軟性を重視する企業にとって戦略的なソリューションである。しかし、LLM（ ）の配備は、会社のローカル・コンピューター上であれ、オンライン・ソリューションの利用であれ、最初の一歩に過ぎない。LLM の機能を実際のタスクに適用するためには、企業はチャットの応答を受信するだけでなく、LLM を使用するコンテキスト外で実行できるコードの形で作成されたロジックを保存できるツールを使用する必要があります。これは、ソリューションの拡張にとって重要である。適切に組織化されたプロセスにより、AI 開発 、一度に複数のプロジェクトに適用したり、企業全体に適用したりすることができる。

この文脈では、適切な開発環境（IDE）の選択が重要な役割を果たす。最新の プログラミングツールは、LLM ベースのソリューションを開発できるだけでなく、それらを既存のビジネスプロセスに統合し、自動化された ETL -Pipeline にすることもできる



章 3.4.

IDE with llm サポート および将来のプログラム変更

IDE の選択：LLM 実験からビジネス・ソリューションまで

自動化、データ分析、人工知能の世界に飛び込むとき、特に大規模な言語モデル（LLM）を扱うときは、適切な統合開発環境（IDE）を選択することが重要です。この IDE は、LLM によって生成されたコードがローカルコンピューター上でも企業ネットワーク内でも実行される場所であり、主要な作業ツールとなります。IDE の選択は、作業の利便性を左右するだけでなく、実験的な LLM クエリから、実際のビジネス・プロセスに組み込まれた本格的なソリューションにどれだけ早く移行できるかも左右します。

> IDE（統合開発環境）は、プロセス・オートメーションとデータ処理のための、コンピュータ上の多用途ビルディング・ブロックです。ノコギリ、ハンマー、ドリル、その他の工具を別々に保管する代わりに、切断、固定、穴あけ、さらには材料の品質チェックまで、すべてをこなせるデバイスが 1 つあればいい。プログラマーにとっての IDE とは、コードを書き（建築に例えれば、図面を作成する）、その作業をテストし（建築モデルの組み立て）、エラーを見つけ（建築中の構造物の強度をチェックするように）、完成したプロジェクトを実行する（家の試運転をする）ことができる一つの空間である。

人気の IDE の概要：

- PyCharm® (JetBrains) は、Python 用の強力なプロフェッショナル IDE です。多くの機能が内蔵されているため、本格的なプロジェクトに適しています。しかし、インタラクティブな Jupyter ファイル (IPYNB)の基本的なサポートは有料版でしか利用できず、初心者はインターフェイスに圧倒されるかもしれません。

> IPYNB（Interactive Python Notebook ）という拡張子を持つファイルは、コード、ビジュアリゼーション、説明が 1 つの文書にまとめられたインタラクティブな Jupyter®ノートブック（図 3.4-1 ）用のフォーマットです。このフォーマットは、レポート、分析、 、トレーニングシナリオの構築に最適です。

- VS Code® (Microsoft)は、無料の IPYNB サポートと多くのプラグインを備えた、高速かつ柔軟でカスタマイズ可能なツールです。初心者にもプロフェッショナルにも適している。GitHub Copilot と言語モデルプラグインの統合が可能で、AI やデータサイエンスプロジェクトに最適です。
- Jupyter Notebook - 実験と学習のための古典的で人気のある選択肢。Jupyter Notebook は、コードを書いたり、説明を加えたり、結果を視覚化したりすることを一つのインターフェイスで可能にする（図 3.4-1）。仮説の迅速な検証、LLM での作業、再現可能な 野生のデータ解析ステップの作成に最適です。依存関係やライブラリを管理するには、Python 環境を管理するためのビジュアル・インターフェースである Anaconda Navigator の使用を推奨する。

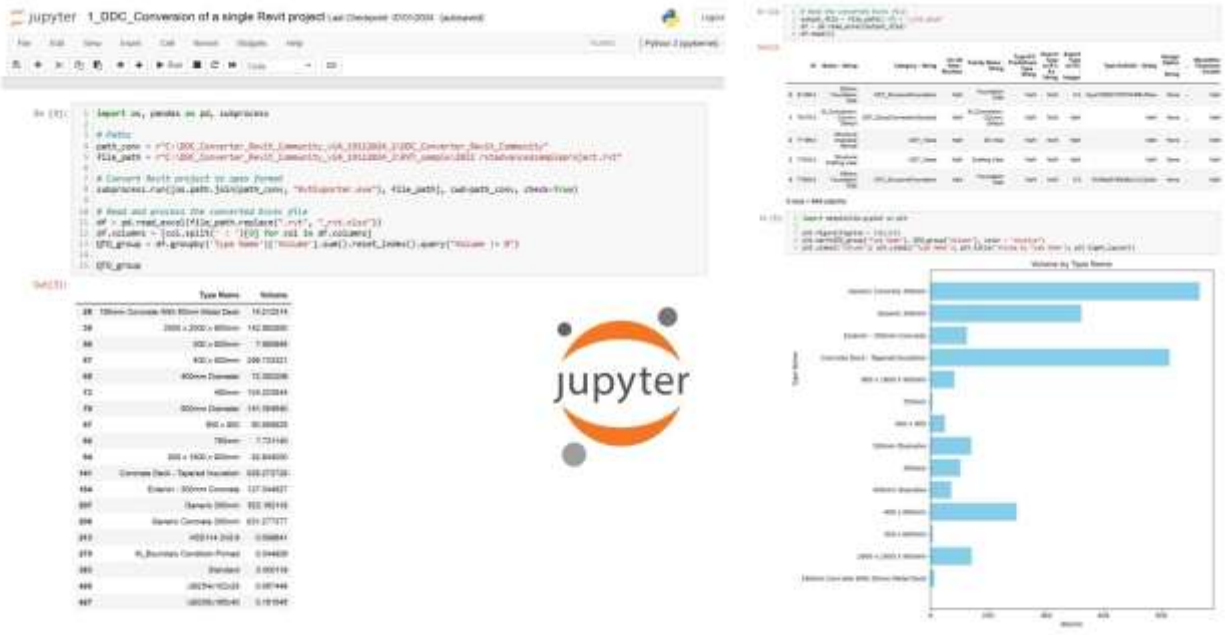


図 3.4-1 Jupyter Notebook Pipeline プロセスを作成するための最も便利で人気のあるツールの一つ。

- Google Collab™ （および Kaggle プラットフォーム （図 9.2-5））は、GPU/TPU への無料アクセスを提供する Jupyter に代わるクラウドベースのツールです。ローカルにソフトウェアをインストールする必要がなく、ブラウザから直接作業できるため、研究を始めるには最適なソリューションです。Google ドライブとの統合をサポートし、最近では Gemini（Google の LLM ）とも統合されている。

	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

図 3.4-2 IDE の比較：Jupyter Notebook Pipeline プロセスを作成するための最も便利で簡単なツールの 1 つ。

IDE の選択はあなたのタスクによる。AI で素早く作業を始めたいなら、Jupyter Notebook や Google Collab を試してみよう。本格的なプロジェクトには、PyCharm や VS Code を使うのが良いだろう。主なことは、始めることだ。最新のツールを使えば、実験を素早く実用的なソリューションに変えることができる。

上述したすべての IDE では、データ処理パイプラインを作成することができる。つまり、コードブロックモジュール（LLM によって生成される可能性がある）のチェーンで、それぞれが異なるステージを担当する、といった具合だ：

- 分析シナリオ
- 文書からの情報抽出の連鎖、
- RAG に基づく自動応答、
- レポートやビジュアライゼーションの作成

モジュール構造により、各ステップを個別のブロックとして表現することができます： データのロード→ フィルタリング→ 分析→ 視覚化→ 結果のエクスポート。これらのブロックは、再利用することができ、また、データ用のコンストラクタのように、新しいチェーンに適合させたり、組み立てたりすることができます。

エンジニア、マネージャー、アナリストにとっては、LLM で生成できるコードの形で意思決定ロジックを文書化できる可能性が広がる。このアプローチは、ルーチン・タスクのスピードアップ、典型的なオペレーションの自動化、すべてのステップが明確に文書化され、チーム・メンバー全員に透過的な再現可能なプロセスの作成に役立つ。

自動化された ETL Pipelines（図 7.2-3）、Apache Airflow（図 7.4-4）、Apache NiFi（図 7.4-5）、n8n（図 7.4-6、プロセス自動化のためのロジックのブロックを構築するためのツールについては、本書のパート 7 とパート 8 で詳しく説明する。

LLM をサポートする IDE および将来のプログラミングの変更

開発プロセスへの人工知能 の統合は、プログラミングの状況を変えつつある。最新の環境は、もはやシンタックスハイライトを備えた単なるテキストエディタではなく、プロジェクトのロジックを理解し、コードを完成させ、特定のコード断片がどのように機能するかを説明することさえできるインテリジェントなアシスタントに変わりつつある。AI を利用して従来の開発の限界を押し広げる製品が市場に登場している：

- GitHub Copilot (VS Code、PyCharm と統合): AI - コメントや部分的な記述に基づいてコードを生成し、テキストのヒントを完成したソリューションに変えるアシスタント。
- Cursor（AI -kernel による VS Code のフォーク）：コードを書き終えるだけでなく、プロジェクトに質問したり、依存関係を探したり、コードベースから学んだりすることができる。
- JetBrains AI Assistant: JetBrains IDE（PyCharm を含む）用のプラグインで、複雑なコードの説明、最適化、テスト作成などの機能があります。
- Amazon CodeWhisperer：Copilot の類似品で、セキュリティと Amazon の AWS サービスのサポートに重点を置いている。

プログラミングは今後数年で劇的な変化を遂げるだろう。主な焦点は、定型的なコード記述からモデルやデータ・アーキテクチャの設計に移るだろう。開発者はシステム設計により関与するようになり、AI はコード生成、テスト、文書化、基本機能といったテンプレート作業を引き継ぐだろう。プログラミングの未来は、人間と AI のコラボレーションであり、機械が技術的なルーチンを引き継ぎ、人間は創造性に集中する。

自然言語プログラミングは日常的なものになるだろう。IDE のパーソナライゼーションは新たなレベルに達するだろう。開発環境は、パターンを予測し、状況に応じた解決策を提供し、過去のプロジェクトから学習することで、ユーザーの作業スタイルや企業に適応することを学ぶだろう

> これは開発者の役割を廃止するものではないが、コードを書くことから知識、品質、プロセス管理へと根本的に変容させるだろう。この進化はビジネス・インテリジェンスにも影響し、レポートやビジュアライゼーション、意思決定支援アプリケーションの作成は、AI や LLM、チャット、エージェント・インターフェースの助けを借りてコードやロジックを生成することでますます行われるようになるだろう。

企業が LLM チャットをセットアップし、適切な開発環境を選択したら、次の重要なステップはデータの整理である。このプロセスでは、異種ソースから情報を抽出し、クリーニングし、構造化された形に変換し、企業システムに統合する。

データ管理に対する最新のデータ中心アプローチでは、重要な目標は、データを多数のツールやアプリケーションと互換性のある単一の普遍的な形式にすることである。構造化プロセスや構造化データを扱うためには、専門的なライブラリが必要である。最も強力で、柔軟性があり、人気のあるライブラリの 1 つは、Python 用の Pandas ライブラリ です 。このライブラリを使うと、表形式のデータを便利に処理することができます：フィルタリング、グループ化、クリーン化、追加、集計、レポート。

Python Pandas: データ操作に欠かせないツール

Pandas は、データ分析と自動化の世界で特別な位置を占めている。Pandas は Python プログラミング言語[86]の中で最も人気があり広く使われているライブラリの一つで、構造化データを扱うために設計されている。

ライブラリとは、関数、モジュール、クラスといった既製の道具一式のようなものだ。建設現場で毎回ハンマーや水平器を発明する必要がないように、プログラミングでもライブラリを使えば、基本的な機能や解決策を再発明することなく、素早く問題を解決することができる。

> Pandas はオープンソースの Python ライブラリ で、高性能で直感的なデータ構造、特にテーブルを扱うための汎用フォーマットである DataFrame を提供しています。Pandas は、データを扱うアナリスト、エンジニア、開発者にとってのスイスナイフです。

Python は、アナリティクス、自動化、機械学習、ウェブ開発で活発に使用されている、シンプルな構文を持つ高水準プログラミング言語である。その人気の理由は、コードの読みやすさ、クロスプラットフォーム性、

ライブラリの豊富なエコシステムにある。現在までに、137,000 以上のオープンソースパッケージが Python 用に作成されており[87]、この数はほぼ毎日増え続けている。そのようなライブラリはそれぞれ、単純な数学演算から画像処理、ビッグデータ解析、ニューラルネットワーク、外部サービスとの統合のための複雑なツールまで、既製の関数の一種のリポジトリである。

> 言い換えれば、何十万ものすぐに使えるソフトウェアソリューション、つまりビジネスプロセスに直接組み込めるライブラリやツールに、無料でオープンアクセスできると想像してください。自動化、分析、視覚化、統合などのためのアプリケーションの巨大なカタログのようなものです。

Pandas は Python エコシステムで最も人気のあるパッケージの一つである。2022 年、Pandas ライブラリ の平均ダウンロード数は 1 日あたり 400 万件に達したが（図 3.4-3）、2025 年初頭にはこの数字は 1 日あたり 1200 万件まで増加しており、これは人気の高まりとデータ分析および LLM チャット[86]における広範な利用を反映している

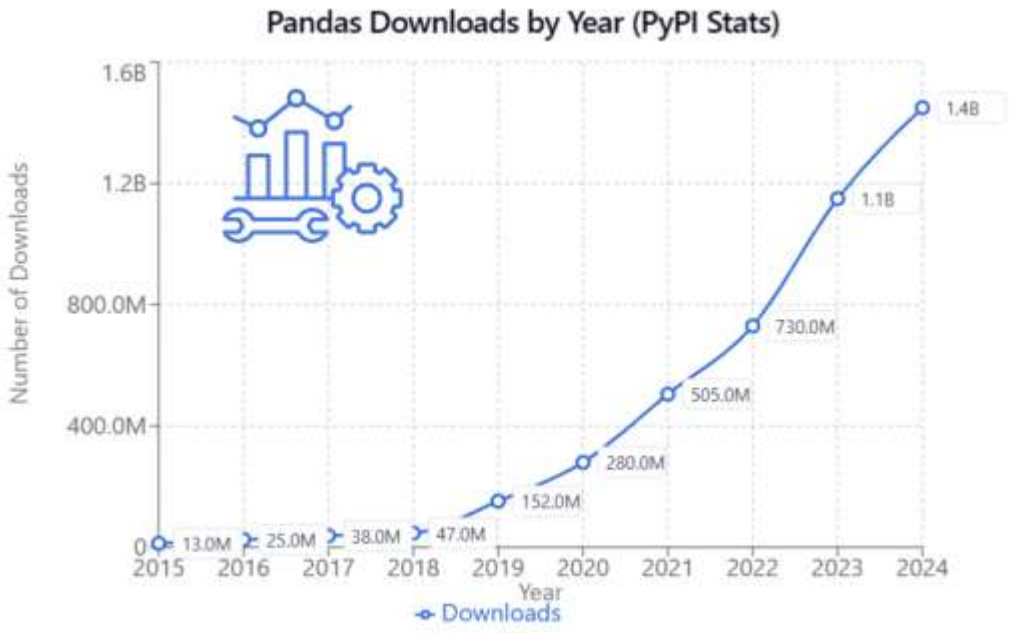


図 3.4-3 Pandas は最もダウンロードされているライブラリの一つである。2024 年には年間ダウンロード数が 14 億を超えた。

Pandas ライブラリ のクエリー言語は、「リレーショナル・データベース と SQL クエリー言語」の章で説明し

た SQL クエリー言語 と機能が似ています。

> アナリティクス と構造化データ管理の世界において、Pandas は、そのシンプルさ、スピード、パワーで際立っており、情報を効果的に分析・処理するための幅広いツールをユーザーに提供しています。

SQL と Pandas の両ツールは、特に従来の Excel と比較して、強力なデータ操作機能を提供します。これらのツールは、選択、フィルタリングなどの操作をサポートしています（図 3.4-4）。唯一の違いは、SQL がリレーショナル・データベースの操作に最適化されているのに対し、Pandas は RAM 上でデータを処理するため、データベースを作成したり、別のインフラを導入したりする必要がなく、どのコンピュータでも実行できることです。



図 3.4-4 Pandas は、SQL とは異なり、データベースに限らず様々なデータ形式を柔軟に扱うことができる。

Pandas は、科学研究、プロセスの自動化、パイプラインの作成（ETL を含む）、Python でのデータ操作によく使われ、SQL はデータベース管理の標準であり、大量のデータを扱う企業環境でよく使われる。

> Python プログラミング言語 の Pandas ライブラリ 、テーブルの読み書きのような基本的な操作だけでなく、データのマージ、データのグループ化、複雑な分析計算の実行など、より複雑なタスクも実行できる。

今日、Pandas ライブラリは、学術研究やビジネス分析だけでなく、LLM - モデルと組み合わせても使用されています。例えば、Meta®部門（Facebook™）は、2024 年に新しいオープンソースモデル LlaMa 3.1 を公開する際、構造化データの処理に特別な注意を払い、そのリリースの重要かつ最初のケースの 1 つを、まさに CSV 形式 の構造化データフレーム（図 3.4-5）の処理と、チャットでの Pandas ライブラリとの直接統合にしました。

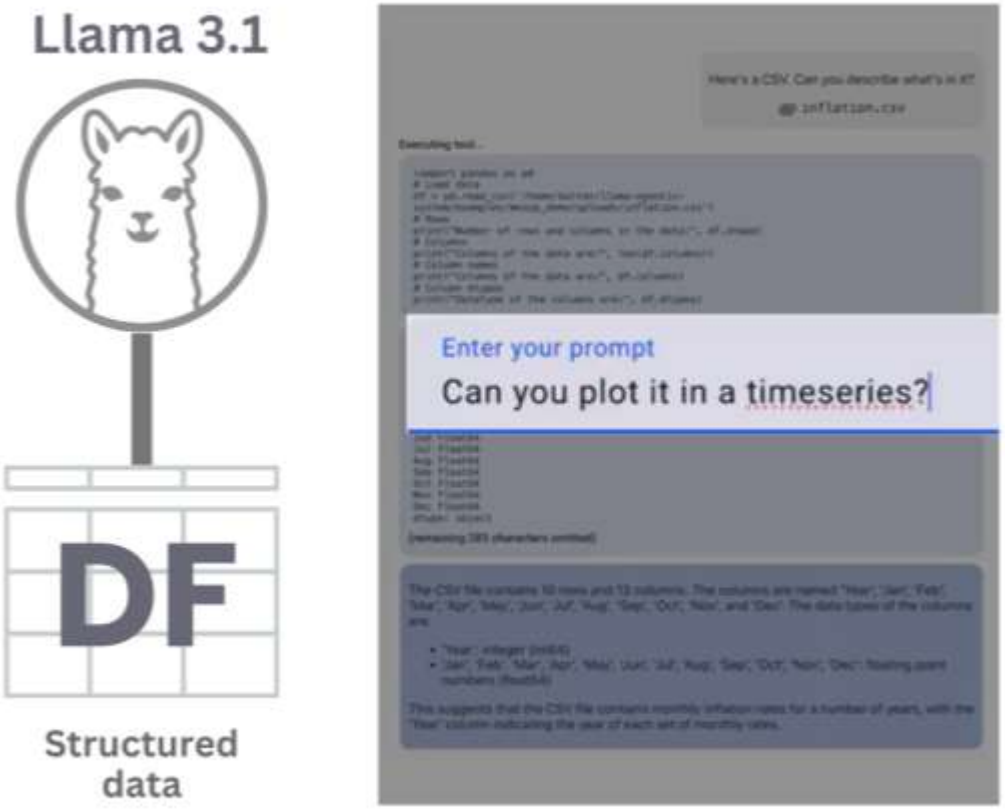


図 3.4-5 2024 年に LlaMa 3.1 で発表された Meta チームの最初の、そして主要なケースのひとつが、Pandas を使ったアプリケーションの構築だった。

> Pandas は、何百万人ものデータサイエンティストが生成 AI のためにデータを処理し、準備するために不可欠なツールです。コード変更ゼロで Pandas を高速化することは、大きな前進となる。データサイエンティストは、数時間ではなく数分でデータを処理できるようになり、生成的 AI モデルを訓練するために桁違いのデータを取得できるようになる[88]。
>
> - ジェンセン・フアン、NVIDIA 創業者兼 CEO

Pandas を使えば、Excel の能力をはるかに超えたデータセットの管理・分析が可能です。Excel は通常 100 万行までのデータを扱うことができますが、Pandas は 数千万行を含むデータセット（図 9.1-2、図 9.1-10）を簡単に扱うことができます[89]。この機能により、ユーザは大規模なデータセットに対して高度なデータ分析と可視化を行うことができ、深い洞察を提供し、データ駆動型の意思決定を容易にします。さらに、Pandas は強力なコミュニティサポート [90]を持っています。世界中の何億人もの開発者やアナリスト (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) が毎日オンラインまたはオフラインで Pandas を使用しており、あらゆるビジネス問題に対してすぐに使えるソリューションを数多く提供しています。

ほとんどの Python 分析プロセスの中心には、Pandas ライブラリ が提供する DataFrame と呼ばれる構造化されたデータ形式があります。これは、表形式のデータを整理、分析、可視化するための強力で柔軟なツールです。

DataFrame: 汎用表データフォーマット

DataFrame は Pandas ライブラリの中心的な構造で、行が個々のオブジェクトやレコードに対応し、列がその特性やパラメータ、カテゴリに対応する 2 次元の表（図 3.4-6）です。この構造は視覚的には Excel のスプレッドシートに似ていますが、柔軟性、拡張性、機能性においてはるかに優れています。

> DataFrame は、コンピュータの RAM に保存された表形式のデータを表現し、処理する方法である。

DataFrame は、コンピュータの RAM に保存された表形式のデータを表現し、処理する方法です。表では、行は建設プロジェクトの要素、列はその特性（カテゴリー、寸法、座標、コスト、用語など）を表します。さらに、このような表には、1 つのプロジェクトに関する情報（図 4.1-13）も、数千の異なるプロジェクトの数百万のオブジェクトに関するデータ（図 9.1-10）も含めることができます。ベクトル化された Pandas 操作のおかげで、このような大量の情報を高速にフィルタリングし、グループ化し、集約することが簡単にできます。

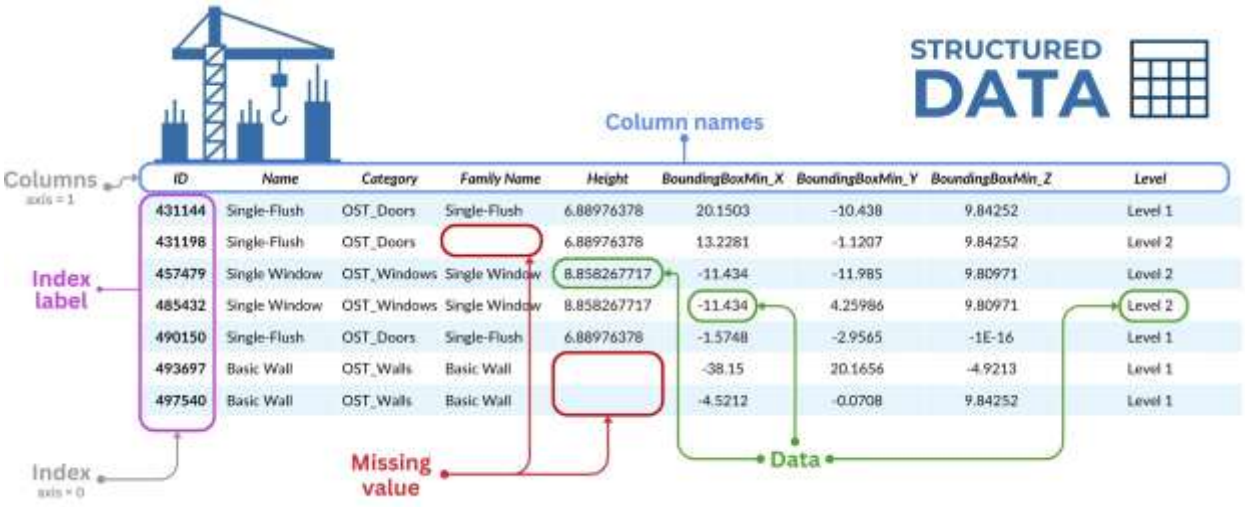


図 3.4-6 DataFrame としての建設プロジェクト は、行に要素、列に属性を持つ二次元テーブルです。

Nvidia の推定によると、今日すでに、全コンピューティング・リソースの最大 30%が構造化データ（データフレーム）の処理に使用されており、この割合は増加の一途をたどっている。

> データ処理は、おそらく世界のコンピューティングの 3 分の 1 がどの企業でも行っていることだ。ほとんどの企業のデータ処理とデータは、表形式の DataFrame にある
>
> - ジェンセン・フアン、エヌビディア CEO[91]

Pandas における DataFrame の主な機能をいくつか挙げてみよう：

- カラム：DataFrame では、データはカラムに編成され、それぞれに固有の名前が付けられます。属性カラムは、データベースのカラムやテーブルのカラムと同様に、異なるタイプのデータを含むことができます。

- Pandas シリーズは、Pandas における一次元のデータ構造であり、テーブルのリストやカラムに似ています。

> Pandas シリーズには 400 以上の属性とメソッドがあり、データを驚くほど柔軟に扱うことができます。利用可能な 400 の関数のいずれかを列に直接適用したり、数学的演算を実行したり、データをフィルタリングしたり、値を置換したり、日付や文字列を操作したりすることができます。さらに、Series はベクトル化された演算をサポートしており、大規模なデータセットの処理を、周期的な計算に比べて大幅に高速化します。例えば、複雑なループを記述することなく、すべての値に数値を掛けたり、欠損データを置換したり、複雑な変換を適用したりすることができます。

- 行：DataFrame の 、一意な値でインデックスを付けることができます。このインデックスにより、特定の行のデータを素早く変更・調整することができます。

- インデックス： デフォルトでは、DataFrame Pandas を作成すると、各行に 0 から N-1 までのインデックスが割り当てられます（N は DataFrame の全行の数です）。ただし、インデックスを変更することで、日付や一意の特徴などの特別な指定を含めることができます。

- DataFrame 、行にインデックスを付けるということは、各行に一意な名前やラベルを付けるということであり、これを DataFrame インデックスと呼びます。

- データ型: DataFrame は様々なデータ型をサポートしている：int`、`float`、`bool`、`datetime64`、テキストデータ用の `obect` などです。DataFrame の各カラムは、その内容に対してどのような操作が可能かを決定する独自のデータ型を持っています。

- データ操作：DataFrame は、集計（`groupby`）、結合（`merge`と`join`）、連結（`concat`）、分割-適用-結合、その他多くのデータ変換技術を含む、データ処理のための幅広い操作をサポートしている。

- サイズの操作: DataFrame では、列の追加や削除、 行の追加が可能で、データ分析のニーズに応じてダイナミックに構造を変更できます。

- データの可視化： 組み込みの可視化テクニックを使ったり、Matplotlib や Seaborn のような一般的なデータ可視化ライブラリと連携したりすることで、DataFrame を簡単にグラフやチャートに変換し、データをグラフィカルに表示することができる。

- データの入出力: Pandas は、CSV 、Excel 、JSON 、HTML 、SQL などの様々なファイルフォーマットへのデータの読み込みと書き出しを行う関数を提供しており、DataFrame をデータ収集と配布の中

心的なハブとする可能性があります。

CSV や XLSX とは異なり、Pandas DataFrame は、データを扱う際の柔軟性とパフォーマンスに優れています。RAM 内で大量の情報を扱うことができ、拡張データ型（日付、論理値、時系列を含む）をサポートし、データのフィルタリング、集約、結合、視覚化のための広範な機能を提供します。CSV はデータ型や構造に関する情報を保存しておらず、XLSX は書式設定に負荷がかかることが多く、スケーラビリテ ィが低い一方で、DataFrame は高速分析（ ）、プロセスの自動化、AI モデルとの統合に最適な選択肢であり続けています（図 3.4-7）。また、本書の第 8 部では、Parquet 、Apache Orc、JSON 、Feather 、HDF5 、データウェアハウスなどの類似フォーマットについても詳しく説明します（図 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)

図 3.4-7 DataFrame は、高性能で高度なデータ型をサポートする、データ操作の最適な選択です。

Pandas ライブラリと DataFrame フォーマットは、その柔軟性、パワー、使いやすさのおかげで、Python データ分析のデファクトスタンダードとなっています。これらは、特に LLM モデルと組み合わせて、簡単なレポートの作成と複雑な分析パイプラインの構築の両方に理想的な 。

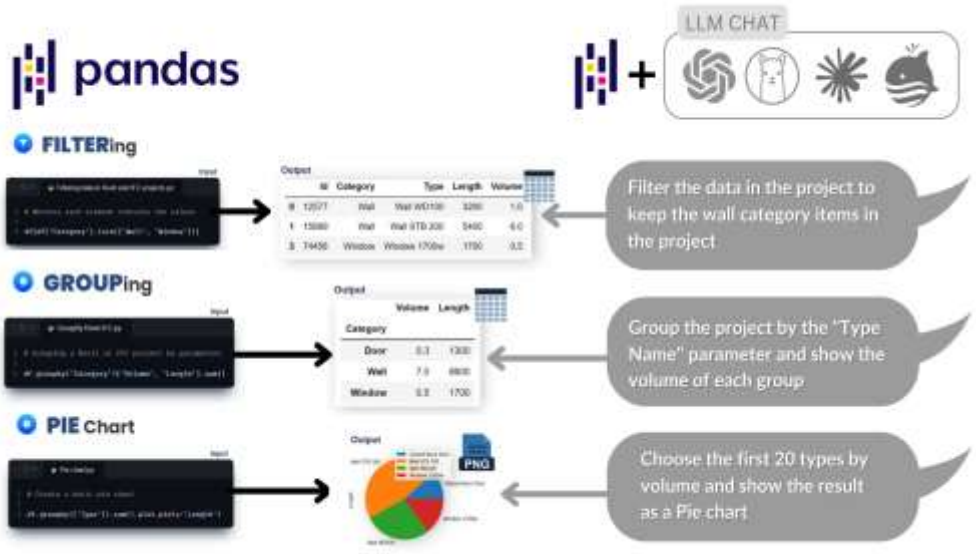


図 3.4-8 LLM は Pandas とのやりとりを簡素化します：コードの代わりにテキストクエリで十分です。

今日、Pandas は ChatGPT 、LlaMa 、DeepSeek 、QWEN などの LLM ベースのチャットルームで活発に使用されています。多くの場合、モデルがテーブル処理、データ検証、分析に関連するクエリを受け取ると、Pandas ライブラリを使用して正確にコードを生成します。これにより、DataFrame は AI との対話においてデータを表現するための自然な " 言語 " となる（図 3.4-8）。

Pandas のような最新のデータテクノロジーは、データの分析、自動化、ビジネスプロセスへの統合を容易にします。結果を迅速に提供し、専門家の作業負荷を軽減し、再現可能なオペレーションを保証します。

次のステップ：持続可能なデータフレームワークの構築

このパートでは、建設業界で使用される主なデータタイプに注目し、それらを保存するためのさまざまなフォーマットについて学び、LLM 、IDE を含む最新ツールの情報処理における役割について分析した。効果的なデータ管理は、情報に基づいた意思決定とビジネスプロセスの自動化の基盤であることを学んだ。データを構造化し、整理することができる組織は、データ処理と変換の段階で大きな競争上の優位性を得ることができる。

この部分を要約すると、考察したアプローチを日常業務に応用するのに役立つ主な実践的ステップを強調する価値がある：

- プロセスのデータ監査を実施する

- [ ] プロジェクトで使用するすべてのデータタイプのインベントリーを作成する。
- [ ] ビジネスプロセスにとって最も重要なデータタイプとモデルを決定する
- [ ] 情報が構造化されていなかったり、構造化が不十分であったり、アクセスできないことが多い問題領域を特定する。

■ データ管理戦略の策定を開始する

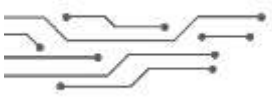
- [ ] さまざまなデータタイプを扱うための政策上の問題と基準を提起する。
- [ ] 非構造化データを構造化データに変換することで、どのワークフローを改善できるかを分析します。
- [ ] セキュリティと機密性を考慮したデータ保管とアクセスポリシーを作成する。

■ データを扱うための基本的なツールをインストールし、使いこなす

- [ ] タスクに適した IDE を選択する（例：VS Code や Jupyter Notebook をインストールする）
- [ ] あなたの個人情報を秘密厳守で取り扱うために、地元の LLM を設立してみてください。
- [ ] Pandas ライブラリ を使って XLSX 表形式データを処理する実験を開始する。
- [ ] スプレッドシートツールやデータベースで扱う典型的なタスクを LLM に説明し、LLM に Pandas を使って作業を自動化するよう依頼する。

これらのステップを適用することで、データを扱うアプローチを徐々に変革し、バラバラで構造化されていないデータセットから、データがアクセス可能で理解しやすい資産となる統一されたエコシステムへと移行することができます。小さく始める - Pandas で最初の DataFrame を作成する、ローカル LLM を実行する、Python を使って最初のルーチンタスクを自動化する（例：Excel スプレッドシート）。

本書の第 4 部では、データ品質 、データ整理、構造化、モデリングに焦点を当てる。PDF やテキストから画像や CAD モデルまで、異種データソースを分析や自動化に適した構造化データセットに変換する方法論に焦点を当てる。また、データ要件がどのように形式化されるのか、建設プロジェクトにおいて概念モデルや論理モデルがどのように構築されるのか、このプロセスにおいて最新の言語モデル（LLM）がどのように役立つのかについても探求する。





IV 部

データ品質：整理、構造化、モデリング

第 4 部では、異種情報を高品質な構造化データセットに変換するための方法論と技術に焦点を当てる。建設プロジェクトにおける効果的な情報アーキテクチャの基礎となる、データ要件の形成と文書化のプロセスについて詳しく説明する。様々なソース（PDF - 文書、画像、テキストファイル、CAD -モデル）から構造化された情報を抽出する実践的な方法を、実施例とともに紹介する。正規表現 （RegEx ）や、データの自動検証と検証のための他のツールの使用について分析する。概念的、論理的、物理的レベルでのデータモデリングのプロセスを、建設業界の特殊性を考慮しながら段階的に説明する。言語モデル（LLM ）を使用して、情報の構造化と検証のプロセスを自動化する具体例を示す。また、分析結果を視覚化するための効果的なアプローチを提案し、建設プロジェクトマネジメントのあらゆるレベルにおける分析情報の可用性を高めている

章 4.1.

データの構造化

データ駆動型経済の時代において、データは障害物ではなく、意思決定の基礎となりつつある。企業は、常に新しいシステムやそのフォーマットに情報を適合させるのではなく、すべてのプロセスで普遍的な真実のソースとして機能する単一の構造化データモデルを形成することを求めるようになっている。最新の情報システムは、フォーマットやインターフェイスではなく、データの意味を中心に設計されている。構造は変わっても、情報の意味はずっと変わらないからだ。

> データを効果的に扱うための鍵は、データを際限なく変換・変換することではなく、最初から正しく整理することである。つまり、プロジェクトのライフサイクルのすべての段階で、透明性、自動化、統合を提供できる普遍的な構造を作ることである。

従来のアプローチでは、新しいプラットフォームを導入するたびに、データの移行、属性名の変更、フォーマットの調整といった手作業による調整を余儀なくされていた。これらのステップでは、データそのものの品質が向上するわけではなく、問題を覆い隠すだけで、終わりのない変換の悪循環を生み出してしまう。その結果、企業は特定のソフトウェア・ソリューションに依存するようになり、デジタルトランスフォーメーションは減速する。

次の章では、データを正しく構造化する方法、そして普遍的なモデルを作成し、プラットフォームへの依存を最小限に抑え、最も重要なこと、つまり持続可能なプロセスを構築するための戦略的リソースとしてのデータに焦点を当てる方法を見ていく。

文書、PDF 、画像、テキストを構造化されたフォーマットに変換する方法を学ぶ。

建設プロジェクトでは、技術文書、作業明細書、図面、仕様書、スケジュール、プロトコルなど、大半の情報が構造化されていない形で存在する。形式も内容も多様であるため、統合や自動化は困難である。

> 構造化形式や半構造化形式への変換プロセスは、入力データの種類や希望する処理結果によって異なる場合がある。

非構造化データから構造化データへの変換は芸術であると同時に科学でもある。このプロセスは入力データの種類や分析の目的によって異なるが、データエンジニア（図 3.2-5）やアナリスト（ ）の作業の大部分を占めることが多い。



図 4.1-1 構造化されていないスキャン文書を構造化された表形式に変換する。

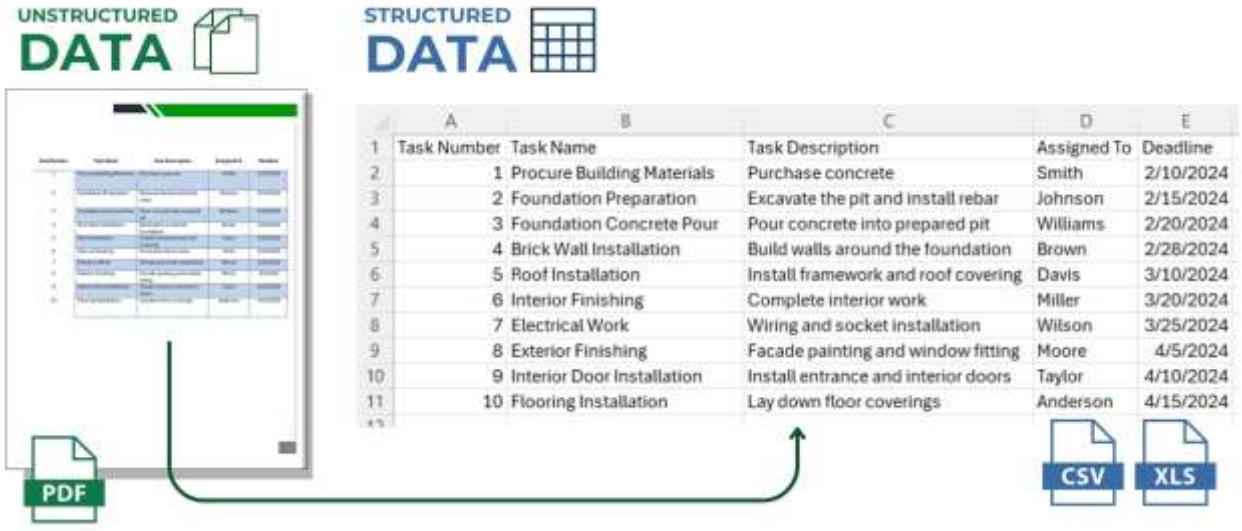
文書・PDF・画像・テキストを構造化形式に変換する（図 4.1-1）のは、ステップバイステップのプロセスであり、以下の手順を含みます：

- 抽出）：このステップでは、非構造化データを含むソース文書または画像が読み込まれる。これは例えば、PDF -文書、写真、図面、回路図などです。
- データ変換（Transform）：構造化されていないデータを構造化された形式に変換するステップ。例えば、光学式文字認識（OCR）やその他の処理方法を用いて、画像からテキストを認識・解釈することが含まれる。
- データのロードと保存（Load）：最後のステップでは、CSV、XLSX、XML、JSON などのさまざまなフォーマットで処理されたデータを保存し、さらに作業を行う。

ETL （Extract, Transform, Load ）と呼ばれるこのプロセスは、自動データ処理において重要な役割を果たし、「ETL とパイプライン: Extract, Transform, Load」の章で詳しく説明する。次に、様々な形式のドキュメントがどのように構造化データに変換されるかの例を見ていきます。

PDF -文書を表に変換する例

建設プロジェクトで最も一般的な作業の 1 つは、PDF 形式の仕様書を処理することです。非構造化データから構造化形式への移行を示すために、PDF 文書から表を抽出し、CSV または Excel 形式 （図 4.1-2）に変換することを実例として考えてみましょう。



	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

図 4.1-2 PDF とは異なり、CSV と XLSX 形式は広く普及しており、さまざまなデータ管理システムに容易に統合できます。

ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN のような LLM 言語モデル は、データサイエンティストがデータを扱う方法を大幅に簡素化し、プログラミング言語のディープラーニングの必要性を減らし、多くのタスクをテキストクエリで解決できるようにする。

したがって、インターネット（通常は StackOverFlow のウェブサイトやテーマ別のフォーラムやチャット）で解決策を探したり、データ処理の専門家に問い合わせたりすることに時間を費やす代わりに、最新のオンラインまたはローカルの LLM の機能を利用することができます。クエリーを尋ねるだけで、PDF -文書を表形式に変換するためのコードが提供されます。

- LLM -モデル（CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN またはその他）に以下のテキストリクエストを送る：

> 表を含む PDF -ファイルからテキストを抽出するコードを書いてください。コードは引数としてファイルパスを取り、抽出された表を DataFrame として返すべきです⏎

- LLM -モデル・レスポンスは、ほとんどの場合、Python のコード形式となる。この言語は、データ処理、自動化、さまざまなファイル形式を扱うために広く使用されているからである：

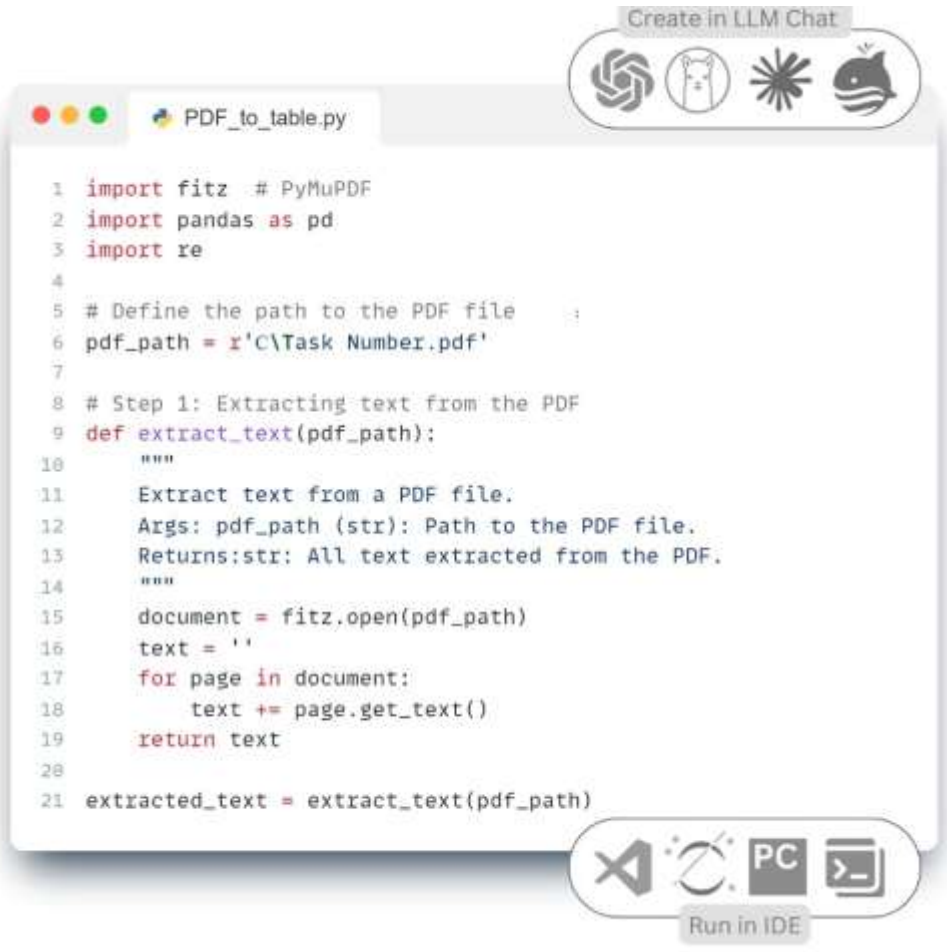


図 4.1-3 Python コード とそのライブラリとパッケージ（Pandas、Fitz）の形の LLM 応答は、PDF -ファイルからテキストを抽出する。

このコード(図 4.1-3)は、PyCharm、Visual Studio Code (VS Code)、Jupyter Notebook 、Spyder、Atom、Sublime Text、Eclipse with PyDev plugin、Thonny、Wing IDE、IntelliJ IDEA with Python plugin 、JupyterLab、または人気のあるオンラインツール: Kaggle.com、Google Collab、Microsoft Azure Notebooks、Amazon SageMaker。

- Convert " ステップでは、有名な Pandas ライブラリ （" Python Pandas: an indispensable tool for working with data " の章で詳しく説明しました）を使って、抽出されたテキストを DataFrame に読み込み、DataFrame を CSV テーブルファイル または XLXS に保存します：

  PDF -ファイルから結果の表を DataFrame に変換するコードが必要です。また、DataFrame を CSV ファイルに保存するコードも追加してください。 ⏎

LLM の回答 ：

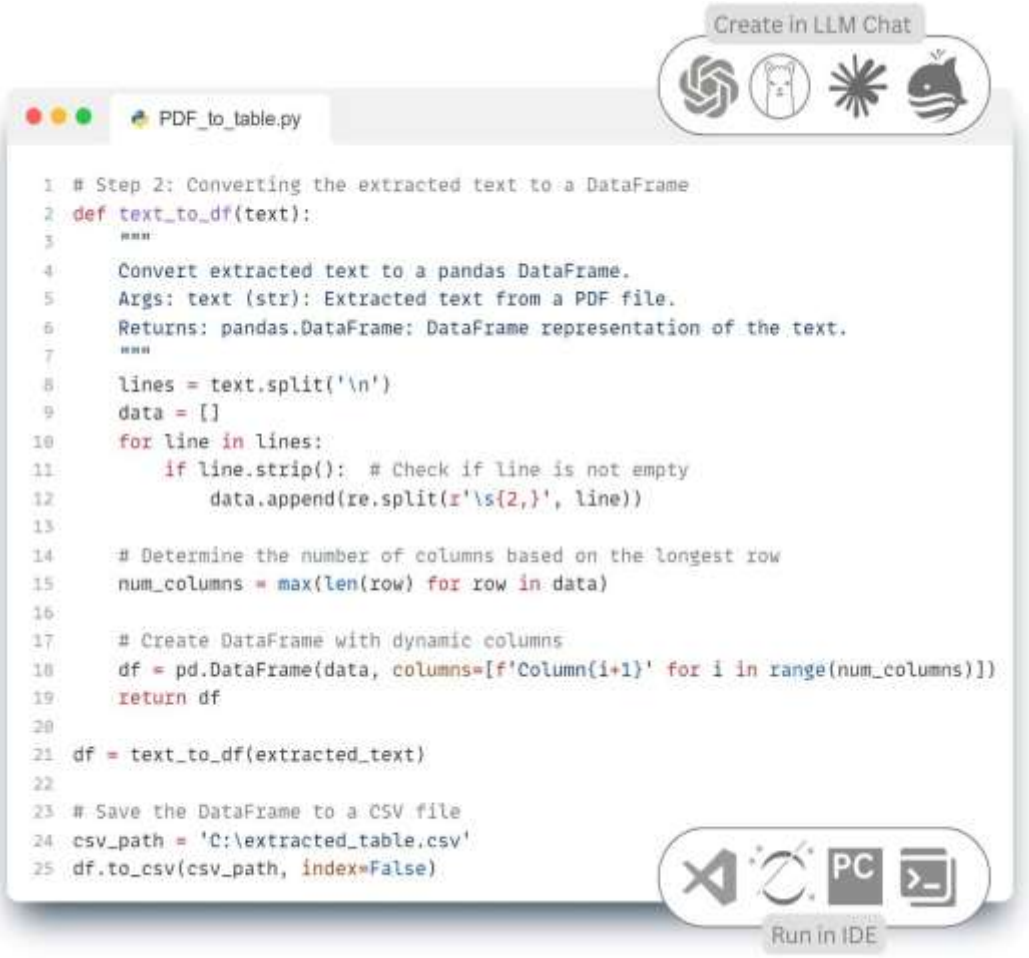


図 4.1-4 PDF から抽出した表を DataFrame に変換し、CSV - ファイルに保存。

コード実行中にエラーが発生した場合（図 4.1-3, 図 4.1-4）、例えばライブラリの不足やファイルパスの間違いなどが原因ですが、エラー・テキストをソース・コードと一緒にコピーし、LLM -モデルに再提出するだけです。モデルはエラーメッセージを分析し、何が問題かを説明し、修正や追加ステップを提案します。

> このようにして、AI LLM とのインタラクションは、リクエスト→ レスポンス→ テスト→ フィードバック→ 修正という完全なサイクルとなり、深い技術的知識を必要としない。

LLM チャットのプレーンテキストクエリと、どの IDE でもローカルで実行できる Python の数十行を使って、PDF -文書を表形式の CSV に変換した。これは PDF 文書とは異なり、簡単に機械で読むことができ、あらゆる

データ管理システムに素早く統合できる。

このコード（図 4.1-3, 図 4.1-4）を、LLM の任意のチャットルームからコピーすることによって、サーバー上の何万も何千もの新しい PDF 文書に適用することができます。これによって、構造化されていない文書のストリームを構造化された CSV 表形式へ変換するプロセスを自動化することができます。

しかし、PDF 文書 、必ずしもテキストが含まれているとは限りません。多くの場合、画像として処理する必要があるスキャン文書です。画像は本質的に非構造化ですが、認識ライブラリの開発と応用によって、その内容を抽出、処理、分析することが可能になり、ビジネスプロセスでこのデータをフル活用できるようになります。

JPEG、PNG 画像を構造化形式に変換する

画像は非構造化データの最も一般的な形態の一つである。建設業界をはじめとする多くの業界では、膨大な量の情報が、スキャンされた文書、回路図、写真、図面などの形で保存されている。このようなデータには貴重な情報が含まれているが、Excel スプレッドシート やデータベース のように直接処理することはできない。画像は、その内容、色、質感が多様であるため、多くの複雑な情報を含んでおり、有用な情報を抽出するためには特別な処理が必要である。

画像をデータソースとして使用することの難しさは、構造がないことである。画像は、Excel のスプレッドシート（ ）やデータベースのテーブルのように、コンピュータがすぐに理解したり処理したりできるような、直接的で簡単に数値化できる方法では意味を伝えません。構造化されていない画像データを構造化された形に変換するには、画像に含まれる視覚情報を解釈できる特別なライブラリを使用しなければならない（図 4.1-5）。

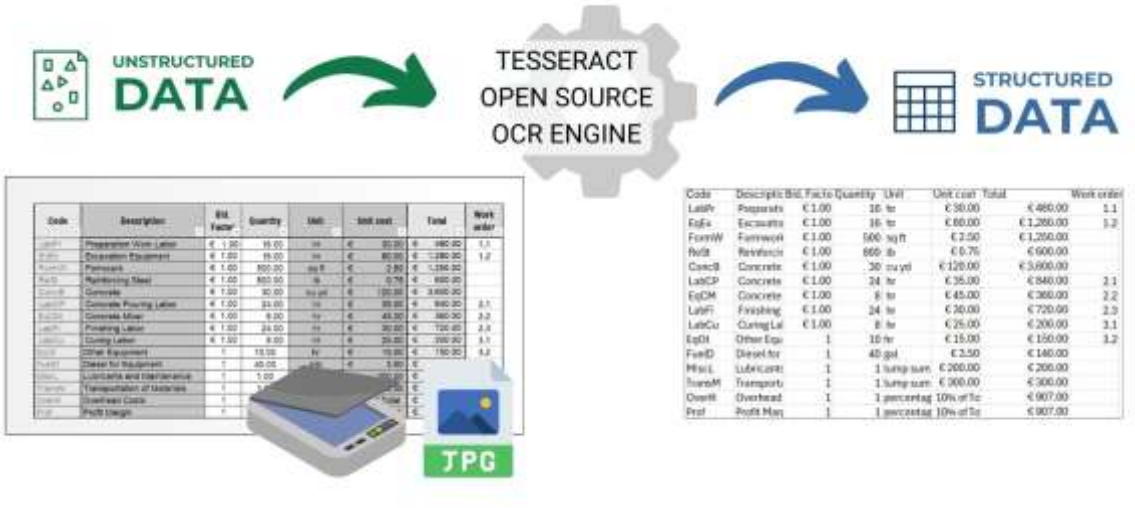


図 4.1-5 スキャンした文書や画像を構造化された形式に変換するには、特別な OCR ツールが必要です。

OCR (Optical Character Recognition) 技術は、画像からテキストを抽出するために使用されます。ドキュメントスキャン、写真、PDF -ファイル内の文字や数字を認識し、編集可能で機械が読めるテキストに変換することができます。OCR テクノロジーは古くから文書の自動化に使用されており、今日ではあらゆるビジネスプロセスや Python -アプリケーションに簡単に統合されています。最も人気のある OCR ツールの一つは Tesseract であり、オープンソースである。 、元々は HP™によって開発され、現在は Google™によってサポートされている。100 以上の言語をサポートし、認識精度も高い。

LLM チャットに、スキャンまたは撮影したテーブルから構造化された方法でデータを取り出すサンプルコードを書いてもらおう。

- LLM チャット（ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN またはその他）にテキストリクエストを送る：

> テーブルを含む JPEG 画像を DataFrame テーブルに変換するコードを書く ⏎

LLM の回答は、ほとんどの場合、画像中のテキストを認識するために Pytesseract ライブラリを使用することを提案します：

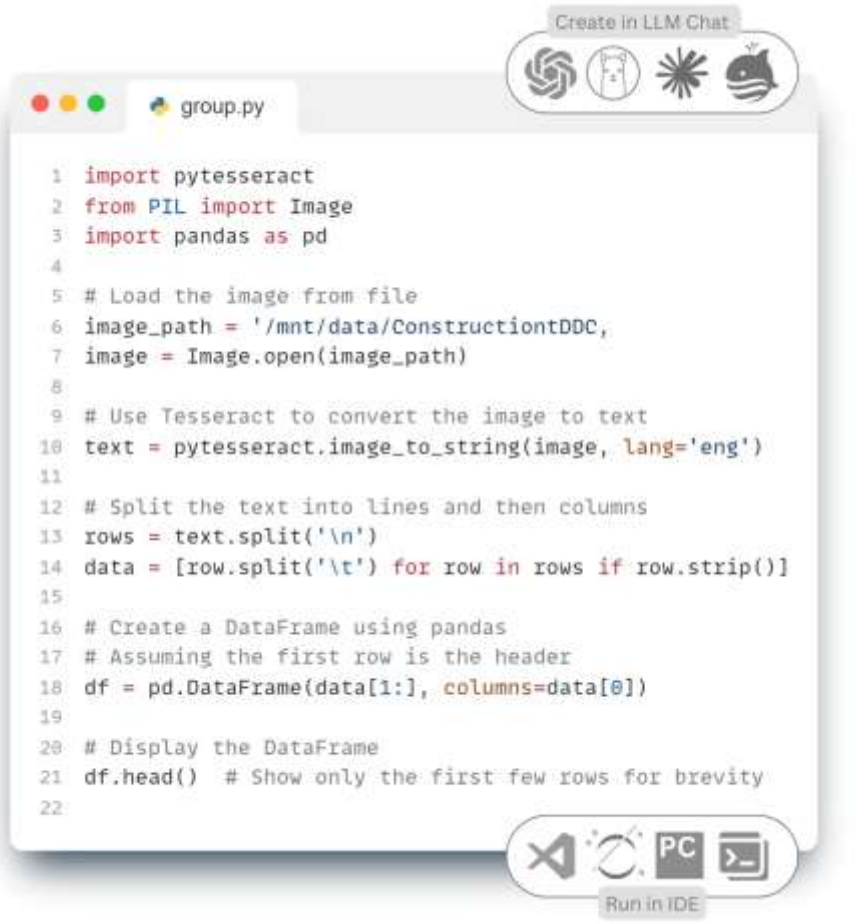


図 4.1-6 画像や写真の表から抽出したテキストを、構造化された表形式表現に変換する。

この例では、LLM に由来するコード（図 4.1-6）は、pytesseract ライブラリ（Tesseract for Python ）を使って OCR （光学式文字認識）を使って画像をテキストに変換し、Pandas ライブラリ を使ってこのテキストを構造化された形式、すなわち DataFrame に変換しています。

> 変換プロセスには通常、画質を向上させるための前処理が含まれ、その後、パターン検出、特徴抽出、オブジェクト認識のためのさまざまなアルゴリズムが適用される。その結果、非構造化視覚情報が構造化データに変換される。

PDF や画像は非構造化情報の重要な情報源であるが、量的な面で真の チャンピオンは、電子メール、チャット

、会議、メッセンジャーで生成されたテキストである。このデータは単に数が多いだけでなく、散在し、非公式化され、構造化が極めて不十分である。

テキストデータを構造化フォームに変換する

表を含む PDF 文書（図 4.1-2）や表形式のスキャン版（図 4.1-5）に加え、プロジェクト文書に含まれる情報の大部分はテキスト形式で提示される。それは、テキスト文書にまとまった文章であることもあれば、図面や設計図に散在する断片的な記録であることもある。近代的なデータ処理環境では、このようなテキストを、分析、視覚化、意思決定に適した構造化された形式に変換することが、最も一般的な作業のひとつである。

このプロセスの中心となるのが分類法であり、共通の特徴に基づいて情報をカテゴリーとサブカテゴリーに整理することを可能にする分類システムである。

> タクソノミーとは、対象をグループ化し、整理するために用いられる階層的な分類構造のことである。テキスト処理の文脈では、項目を意味的なカテゴリーに体系的に分類するための基礎となり、分析を簡素化し、データ処理の質を向上させる。

タクソノミの作成には、エンティティの抽出、分類、および文脈化のステップが伴います。テキストデータから情報を抽出するプロセスをモデル化するために、以下のステップは、PDF 文書からデータを構造化するために既に適用したステップと同様である：

- Extract): テキストデータを分析して、プロジェクトスケジュールの遅延や変更に関する情報を抽出する必要がある。
- Categorisation and Classification (Transofrm)：遅延やスケジュール変更の理由など、受け取った情報を分類する。
- 統合 (Load): 最後に、外部のデータ管理システムに統合するための構造化データを準備します。

プロジェクトマネジャーとエンジニアの間で、スケジュールに関する問題を議論する対話があるとします。私たちのゴールは、重要な要素（遅延の理由、スケジュール調整）を抽出し、構造化された方法で提示することです（図 4.1-7）。

予想されるキーワードに基づいて抽出を行い、データ抽出をシミュレートするために DataFrame を作成し、変換後に日付、イベント（例：遅延の理由）、アクション（例：スケジュール変更）の列を含む新しい

DataFrame テーブルを作成してみましょう。

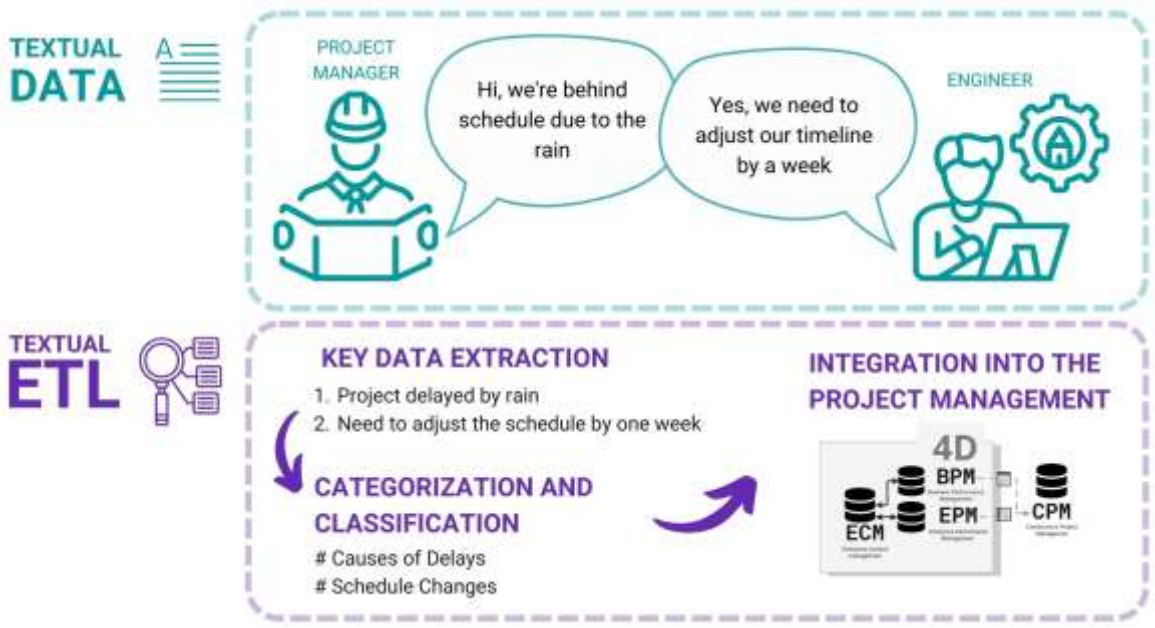


図 4.1-7 タイムラインを調整し、変更をプロジェクト管理システムに統合する必要性について、本文中の重要な情報をハイライトする。

前の例と同じように、言語モデルの 1 つにテキストクエリを使用して問題を解くコードを示します。

LLM チャットルームにテキストリクエストを送信してください：

> こんにちは、雨のせいで予定より遅れています」というマネージャーと、「はい、納期を 1 週間調整する必要があります」というエンジニアの会話があります。今後発生する同じようなテキスト対話を分析し、そこから遅延の理由と必要な納期調整を抽出し、このデータから DataFrame を生成するスクリプトが必要です。DataFrame は CSV ファイルに保存する。 ⏎

- LLM からのレスポンスには通常、正規表現 (re - Regex) と Pandas (pd) ライブラリを使用した Python - コードが含まれる ：

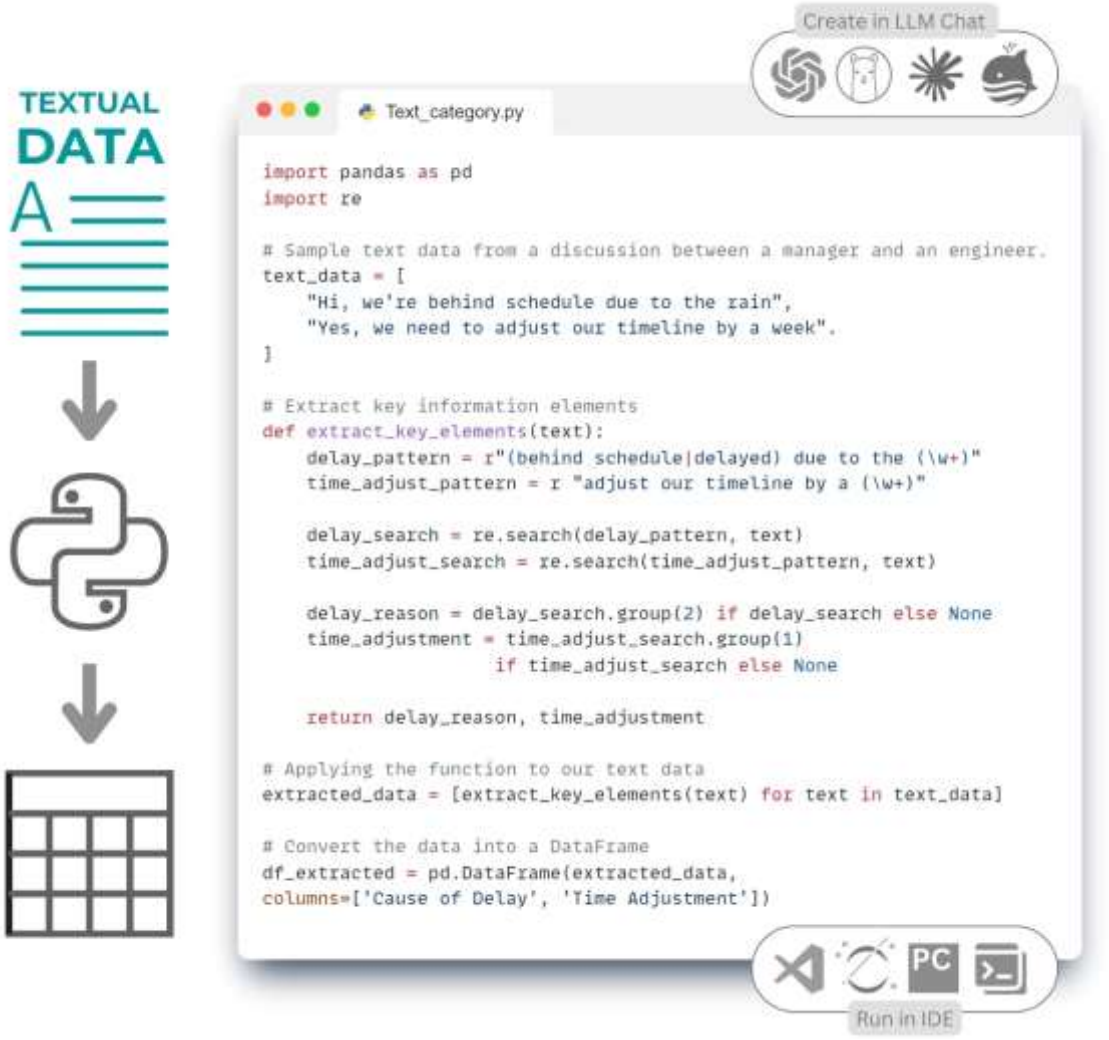
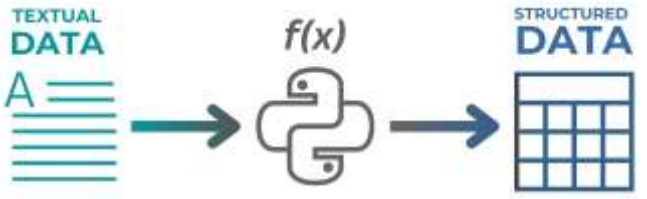


図 4.1-8 締切日を調整する必要性について、本文中の重要な情報を強調表示した表。

この例（図 4.1-7）では、プロジェクトマネジャーとエンジニアの間のやり取りを含むテキストデー タを分析し、同じようなやり取りがある将来のプロジェクトの管理に影響を与える可能性のある特定の情 報を特定し、抽出している。正規表現 (正規表現については「構造化要件 と RegEx 正規表現 」の章で詳しく説明する) を使って、プロジェクトの遅延の原因や、タイムスケジュールの必要な調整をパターンによって特定する。この例で書かれている関数は、パターンに基づいて、文字列から遅延の原因または時間調整のいずれかを抽出します。"because of" の後の単語を遅延の原因として選択するか、"by" の後の単語を時間調整として選択します。

もし、ある行が天候による遅延に言及していれば、「雨」が原因として特定される。もし、ある行が特定の期間のスケジュール調整に言及していれば、その期間は時間調整として抽出される（図 4.1-9）。行にこれらの単語がない場合、対応する属性カラムの値は "None "となる。



Text string	Cause of Delay	Time Adjustment
0	rain	None
1	None	week

図 4.1-9 コード実行後、DataFrame として得られるサマリー表には、遅延の有無と必要な時間調整に関する情報が含まれる。

例えば、作業員の不足は、悪天候時の作業ペースに影響を与える可能性がある。そのため、企業は、建設現場の現場監督とプロジェクト対話（図 4.1-9）から遅延パラメータを事前に知ることで、不利な予報の場合に作業員を強化することができる。

文書や画像の構造化フォーマットへの変換は、比較的シンプルでオープンでフリーの分類ベースのツールを使って実現できる。

特に CAD ソフトウェア（BIM ）を使用する場合、要素の分類もプロジェクトデータを扱う際の重要な要素である。

CAD データ (BIM) の構造化形式への変換

CAD データ (BIM) の構造化と分類は、より複雑なタスクである。なぜなら、CAD (BIM) データベースから保存されるデータは、ほとんどの場合、閉じた、または複雑なパラメトリック形式であり、多くの場合、幾何学的データ要素 (半構造化データ) とメタ情報要素 (半構造化データまたは構造化データ) を同時に組み合わせているからである。

CAD （BIM ）システムのネイティブ・データ・フォーマットは通常保護されており、特別なソフトウェア または API - 開発者自身のインターフェイス - がない限り、直接使用することはできない（図 4.1-10）。このようなデータの分離は、情報の自由な交換を制限し、企業におけるエンド・ツー・エンドのデジタル・プロセスの創出を阻害する、閉じたストレージ・サイロを形成する。

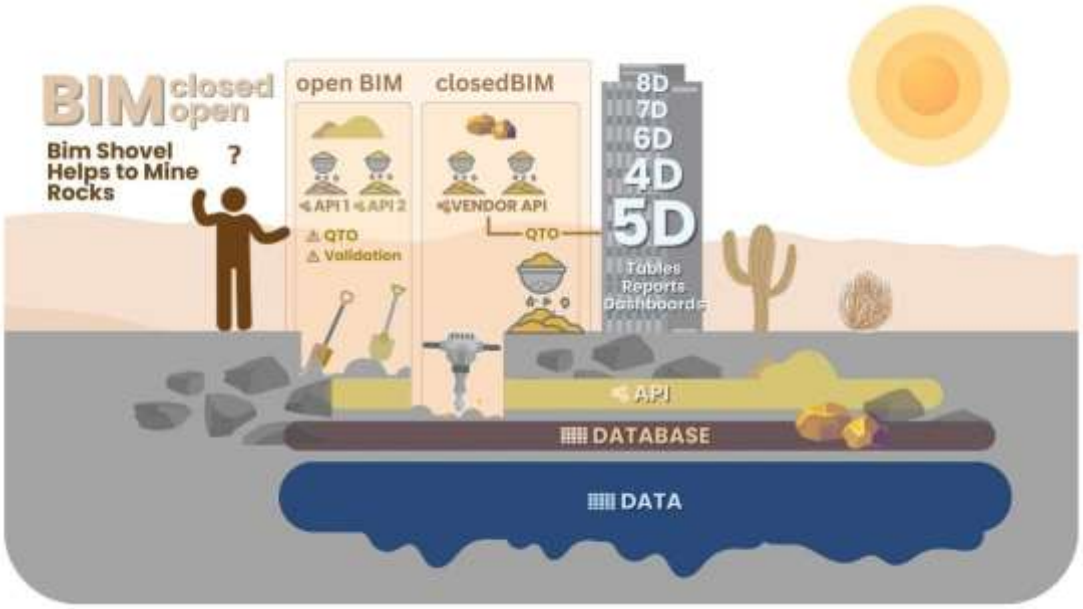


図 4.1-10 CAD 専門家（BIM ）は、API -接続またはベンダー・ツールを通じてネイティブ・データにアクセスできる。

特別な CAD （BIM ）フォーマットでは、プロジェクト要素の特性や属性に関する情報は、階層的な分類システム に集められ、対応する特性を持つエンティティは、果樹の果実のように、 データ分類の枝の最も新しいノードに配置される（図 4.1-11）。

このような階層からのデータ抽出は、2 つの方法で可能である。1 つは、手作業で、ツリーを処理するように各ノードをクリックし、選択したカテゴリーやタイプの枝を斧で切り落とす方法である。あるいは、アプリケーション・プログラミング・インターフェース（API）を使用することで、より効率的で自動化されたアプローチでデータを取り出し、グループ化し、最終的には他のシステムで使用するための構造化されたテーブルに変換することができる。

Dynamo、pyRvt、Pandamo（Pandas + Dynamo）、ACC、または IFC フォーマット用の IfcOpSh や IFCjs のようなオープンソースソリューション （）のようなさまざまなツールを使用して、CAD （BIM ）プロジェクトから構造化データテーブルを抽出することができます。

最新のデータエクスポート・変換ツールは、CAD モデルの内容を、形状情報と属性データ（図 4.1-13）、つま

り設計要素の特性を記述するメタ情報（図 3.1-16）の 2 つの主要コンポーネントに分割することで、データ処理と準備を簡素化することができます。これらの 2 つのデータレイヤーは、一意の識別子によってリンクされたままであり、そのおかげで、（パラメータやポリゴンを介した）ジオメトリ記述を持つ各要素を、その属性（名称、材料、完成段階、コストなど）に正確に対応付けることができます。このアプローチにより、モデルの完全性が保証され、ビジュアライゼーション（形状モデルデータ）と分析・管理タスク（構造化または疎構造化）の両方で、2 種類のデータを別々に、または並行して、柔軟に使用することができます。

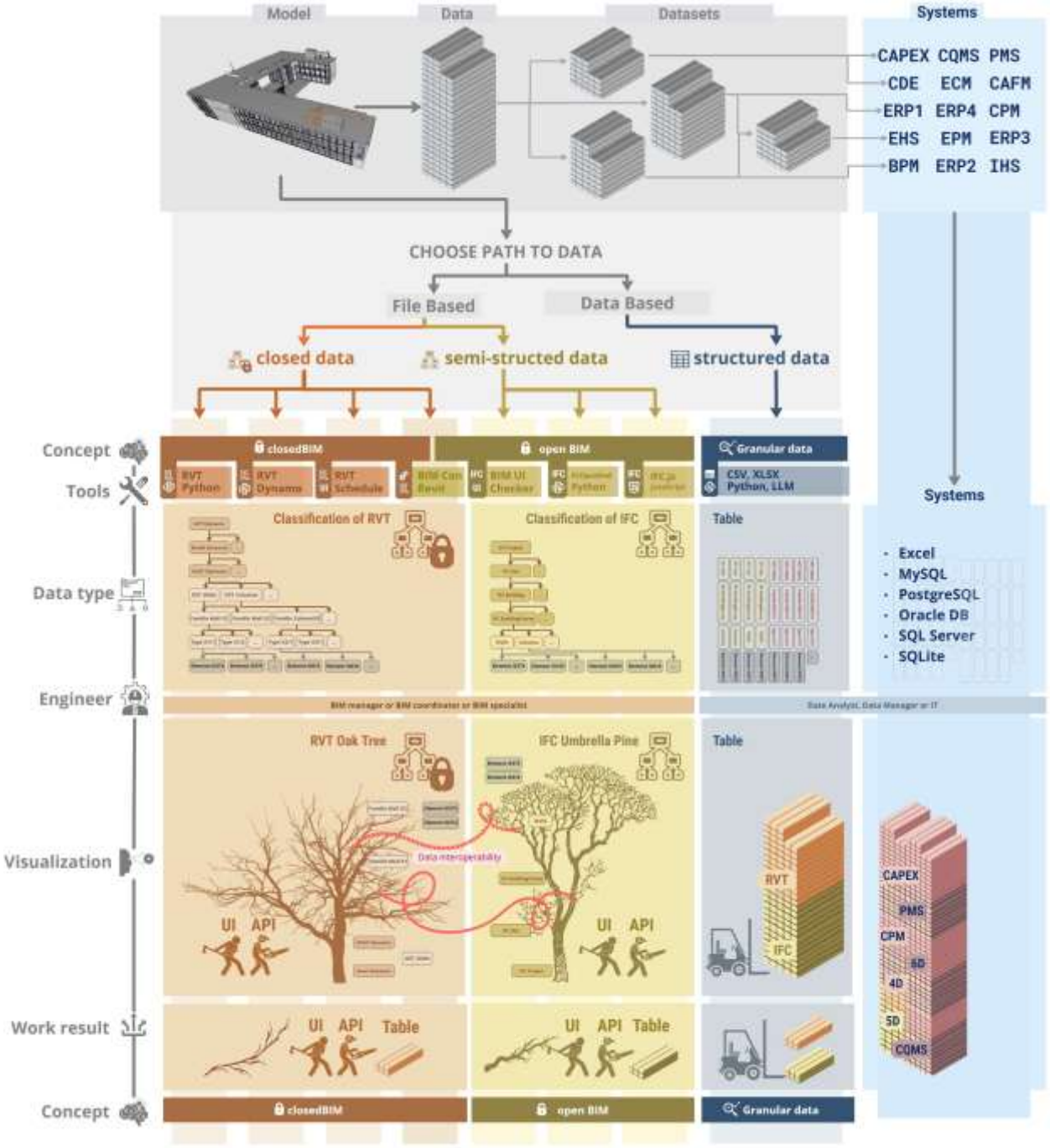


図 4.1-11 CAD データベース （BIM ）からの情報のビューは、分類ツリー の形でユーザーに提示される。

リバースエンジニアリング技術の発展 と、CAD データ変換のための SDK （ソフトウェア開発キット）の登場により、クローズド CAD ソフトウェアフォーマット （BIM ）からのデータの入手と変換が非常に容易になりま

した。現在では、クローズドフォーマット（ ）からのデータを、他のシステムでの解析や使用に適したユニバーサルフォーマットに合法的かつ安全に変換することが可能です。最初のリバースエンジニアリングツール（「オープン DWG」） の歴史と、CAD ベンダーのフォーマットをめぐる覇権争いについては、「構造化データ: デジタル変革の基盤」の章で述べた。

> リバースエンジニアリングツールは、閉じた独自フォーマットから合法的にデータを取り出すことを可能にし、混在する CAD (BIM) フォーマットの情報を、ユーザーが必要とするデータタイプとフォーマットに分解し、処理と分析を容易にします。

リバースエンジニアリングと CAD データベースからの情報への直接アクセスを使用する 情報へのアクセスを可能にし、オープンデータとオープンツール、標準ツールを使用したデータ分析、レポート作成、視覚化、他のデジタルシステムとの統合を可能にする（図 4.1-12）。

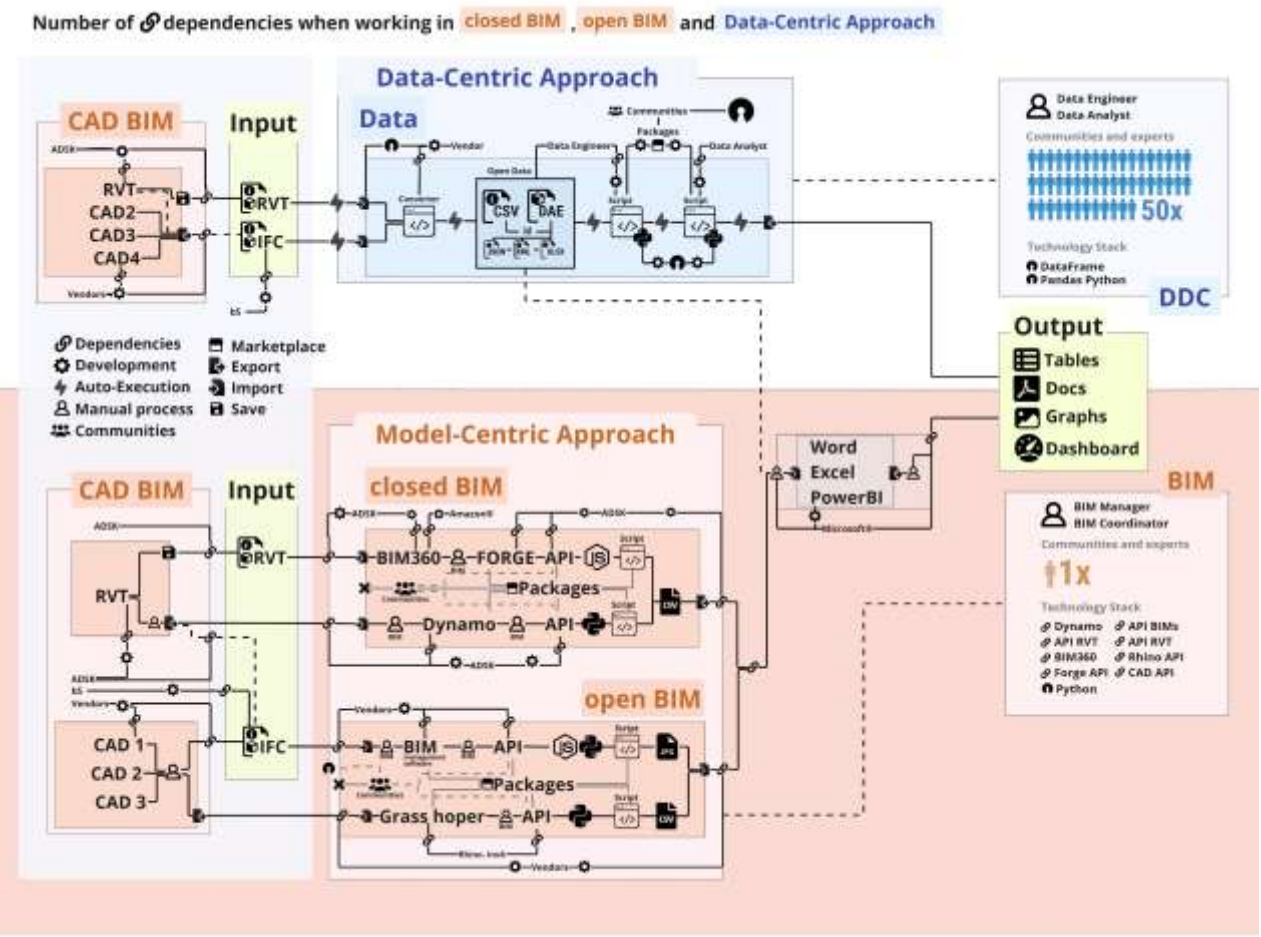


図 4.1-12 CAD データへの直接アクセスは、ソフトウェア・プラットフォームへの依存を最小化し、データ中心のアプローチに向かいます。

1996 年以降、DWG 形式 、2008 年以降、DGN 形式 、2018 年以降、RVT リバースエンジニアリングツールの助けを借りて、当初はクローズドだった CAD データ形式を、構造化形式を含むその他の形式 に便利かつ効率的に変換することが可能になっています（図 4.1-13）。今日、世界のほぼすべての主要な CAD （BIM ）および大手エンジニアリング企業は、SDK - リバースエンジニアリングツールを使用して、クローズド CAD（BIM）ベンダーのフォーマットからデータを抽出しています[92]。



図 4.1-13 リバース・エンジニアリング・ツールの使用により、CAD (BIM)のプログラム・データベースを任意の便利なデータ・モデルに変換することができる。

データをクローズドな独自フォーマットからオープンフォーマットに変換し、混在する CAD (BIM)フォーマットを幾何学的データとメタ情報属性データに分離することで、作業プロセスが簡素化され、分析、操作、他のシステムとの統合が可能になります（図 4.1-14）。

> 今日の CAD データ （BIM ）を使った作業では、CAD（BIM）ベンダーに許可を求めなくても、CAD フォーマットの情報にアクセスできるところまで来ている。

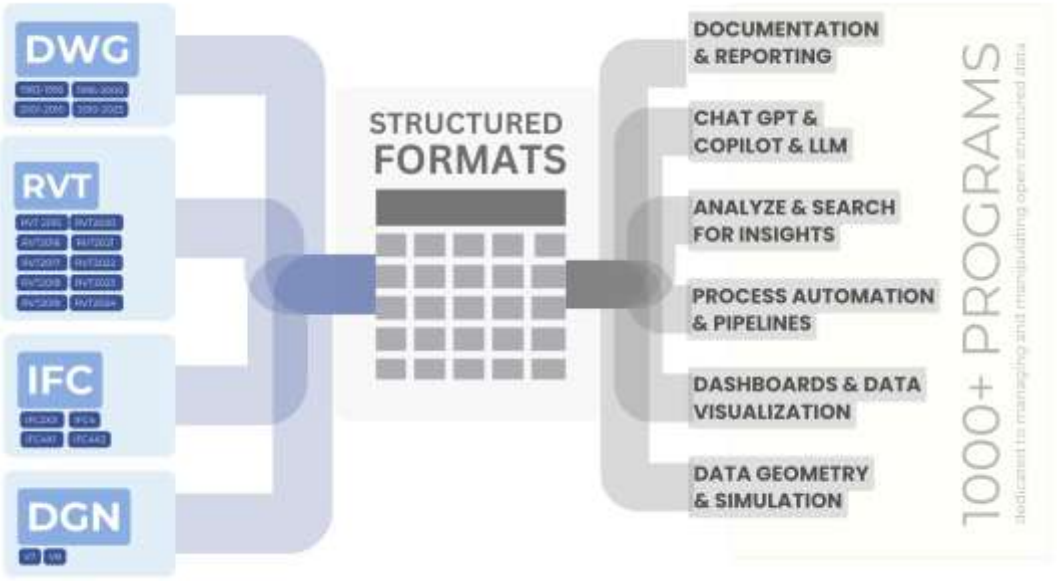


図 4.1-14 最新の SDK ツール は、独自の CAD データベース形式 （BIM ）からのデータの合法的な変換を可能にします。

CAD 設計データ処理の現在のトレンドは、市場の主要プレーヤーである CAD - ベンダーによって形成され続け

ている。ベンダーは、データの世界での地位を強化し、新しいフォーマットやコンセプトを生み出すために努力している。

CAD ソリューション・ベンダーは構造化データに向かっている

2024 年以降、設計・建設業界は、データの使用と処理において大きな技術的転換を迫られている。設計データへの自由なアクセスの代わりに、CAD -システムベンダーは次の新しいコンセプトの推進に注力している。BIM（2002 年作成）やオープン BIM （2012 年作成）といったアプローチは、CAD ベンダーが推進し始めた最新の技術ソリューションに徐々に道を譲りつつある[93]：

- 効率的な情報管理とデータ分析への移行を可能にする「きめ細かい」データの使用への移行。
- USD フォーマットの出現と、柔軟なデータ編成のための Entity-component-system アプローチ (ECS) の導入。
- データ処理、プロセス自動化、データ分析における人工知能の積極的活用
- 相互運用性の開発 - 異なるプログラム、システム、データベース間の相互作用の改善

これらの各側面については、本書の第 6 部「CAD and BIM: marketing, reality and the future of design data in construction」で詳しく説明する。この章では、変化の一般的なベクトルについて簡単に概説するにとどめる。主要な CAD ベンダーは現在、設計情報の構造化方法を見直そうとしている。主要なシフトの一つは、古典的なファイルベースのストレージモデルを放棄し、個々のモデルコンポーネントへの継続的なアクセスを提供する、粒度の細かい分析指向のデータアーキテクチャを採用することである[93]。

> 業界は、幾何学的なカーネルを必要とする、煩雑で特殊なパラメトリック・フォーマットを段階的に廃止し、より普遍的で、機械が読みやすく、柔軟性のあるソリューション を選ぼうとしているのだ。

このような変化の原動力のひとつが、USD（Universal Scene Description）フォーマットであり、もともとはコンピュータグラフィックス業界で開発されたものですが、シミュレーションとビジュアライゼーションのための NVIDIA Omniverse（および Isaac Sim）プラットフォームの開発により、エンジニアリングアプリケーションではすでに認知されています[93]。パラメトリック IFC とは異なり、USD はよりシンプルな構造を提供し、JSON フォーマット （図 4.1-15）でジオメトリとオブジェクトのプロパティを記述することができます。この新しいフォーマットでは、ジオメトリ（BREP -NURBS - 詳細は本書の第 6 部を参照 ） を MESH ポリゴンの

形式で、オブジェクトプロパティを JSON の形式で格納することができ、自動化プロセスやクラウドエコシステムでの作業により便利です [94]。

CAD や ERP のベンダーの中には、すでに同様のフォーマット（NWD、SVF 、CP2、 CPIXML など） を使用しているところもあるが、そのほとんどがクローズドなままで外部利用ができないため、データ統合や再利用の可能性が制限されている。このような状況において、USD はかつて DXF が果たしたような役割、つまり DWG のようなプロプライエタリなフォーマットに代わるオープンな役割を果たすことができる。

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

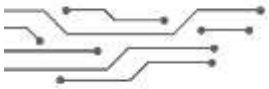
図 4.1-15 USD 形式は、CAD ベンダーが相互運用性 と幾何学カーネルからの設計データの独立性の要求に応えようとする試みである。

主要な開発者がオープンで簡素化された USD 、GLTF 、OBJ 、XML（クローズド NWD、CP2、SVF、SVF2、CPIXML） および同様のフォーマット（図 3.1-17）に移行していることは、データの簡素化とアクセシビリティの向上を求める世界的な傾向と業界の需要を反映している。今後数年間で、幾何学的カーネル（）に依存した複雑なパラメトリック標準やフォーマットから、より軽量で構造化されたソリューションへと徐々に移行していくことが予想されます。この移行は、建設業界のデジタル化を加速し、プロセスの自動化を促進し、データ交換を簡素化する。

CAD -ベンダーが新しいオープンフォーマットを推進する戦略的計画を立てているにもかかわらず、建設業界の専門家は、リバースエンジニアリングツール を使用することで、CAD（BIM ）ツールを必要とせずに、クローズド CAD システムのデータに完全にアクセスすることもできる。

これらのトレンドはすべて、必然的に、かさばるモノリシックな 3D モデルから、普遍的で構造化されたデータへの移行、そして他業界で長い間実績のあるフォーマットの使用につながります。プロジェクトチームが CAD モデルを単なるビジュアル・オブジェクトやファイル・セットとしてではなく、知識と情報を含むデータベースとして捉え始めると、設計と管理へのアプローチは劇的に変化する。

チームが文書、テキスト、図面、CAD モデルから構造化データを抽出する方法を学び、データベースにアクセスできるようになったら、次の重要なステップはデータモデリングと品質保証である。このステップこそが、最終的に特定のアプリケーション・タスクにおける意思決定に使用される情報の処理と変換のスピードを大きく左右する。



章 4.2.

分類と統合：建設データの共通言語

意思決定のスピードはデータの質に左右される

今日のデザイン・データ・アーキテクチャは、根本的な変化を遂げつつある。業界は、かさばる孤立したモデルや閉じたフォーマットから、分析、統合、プロセスの自動化に焦点を当てた、より柔軟で機械が読みやすい構造へと移行しています。しかし、新しいフォーマットへの移行だけでは効率性は保証されません。

> 本書のページでは、フォーマット、システム、プロセスについて多くを語っている。しかし、これらの努力は、ある重要な要素、すなわち信頼できるデータがなければ意味がない。データ品質はデジタル化の礎石であり、この点については次のパートで繰り返し説明する。

現代の建設会社（特に大企業）は、何十、時には何千もの異なるシステムやデータベースを使用している（図 4.2-1）。これらのシステムは、定期的に新しい情報で満たされるだけでなく、相互に効果的に作用し合わなければならない。入ってくる情報を処理した結果生成される新しいデータはすべて、これらの環境に統合され、特定のビジネス・タスクの解決に役立てられる。

また、かつては、特定のビジネス・タスクに関する意思決定は、トップ・マネジャー、いわゆる HiPPO（図 2.1-9）が経験と直感に基づいて行っていたが、情報量が急増した今日、このアプローチは議論の的になりつつある。リアルタイムのデータを扱う自動分析 がこれに取って代わりつつある。

> 「従来の手作業による」エグゼクティブ・レベルのビジネス・プロセスの議論は、ビジネス・クエリへの迅速な応答が求められるオペレーション・アナリティクス（）へとシフトしていくだろう。

経理担当者、現場監督、見積担当者が、 、手作業で作成したレポートやサマリー表、プロジェクトデータのショーケースを何日も何週間もかけて見ていた時代は過去のものとなった。今日、意思決定のスピードと適時性が競争優位の重要な要素になりつつある。



図 4.2-1 建設業界は、数時間あるいは数分で意思決定が行われる他の業界とは異なり、計算と意思決定に日数を要する。

建設業界と、よりデジタル化が進んだ業界との主な違い（図 4.2-1）は、データの質と標準化のレベルの低さである。情報の生成、送信、処理に対する時代遅れのアプローチは、プロセスを遅らせ、混乱を引き起こす。統一されたデータ品質基準（）の欠如は、エンド・ツー・エンドの自動化の実現を妨げている

> 主な課題のひとつは、入力データの質の低さと、その準備と検証のための正式なプロセスの欠如である。信頼できる一貫性のあるデータがなければ、システム間の効果的な統合はできない。これは、プロジェクトのライフサイクルのあらゆる段階で、遅延、エラー、コスト増につながる。

本書の以下のセクションでは、データ品質を向上させ、プロセスを標準化し、情報から品質、検証済み、一貫性のあるデータまでの道のりを短縮する方法について詳しく見ていきます。

データの標準化と統合

効果的なデータ管理には、明確な標準化戦略が必要である。データ構造と品質に関する明確な要件があって初めて、データ検証を自動化し、手作業を減らし、プロジェクトのあらゆる段階で情報に基づいた意思決定を加速することができる。

日々の実務では、建設会社は毎日何百ものファイルを処理しなければならない。電子メール、PDF -文書、CAD 設計ファイル、IOT センサーからの データなどである。これらは会社のビジネス・プロセスに統合する必要がある。

データベースとツールからなる企業のエコシステムの森（図 4.2-2）は、企業が望む結果を達成するために、入力されるマルチフォーマットデータから栄養素を導き出すことを学ばなければならない。

> データの流れに効果的に対処するためには、必ずしも管理職を雇う必要はない。まず、データに関する厳格な要件と標準を策定し、適切なツールを使って自動的に検証、調和、処理する必要がある。

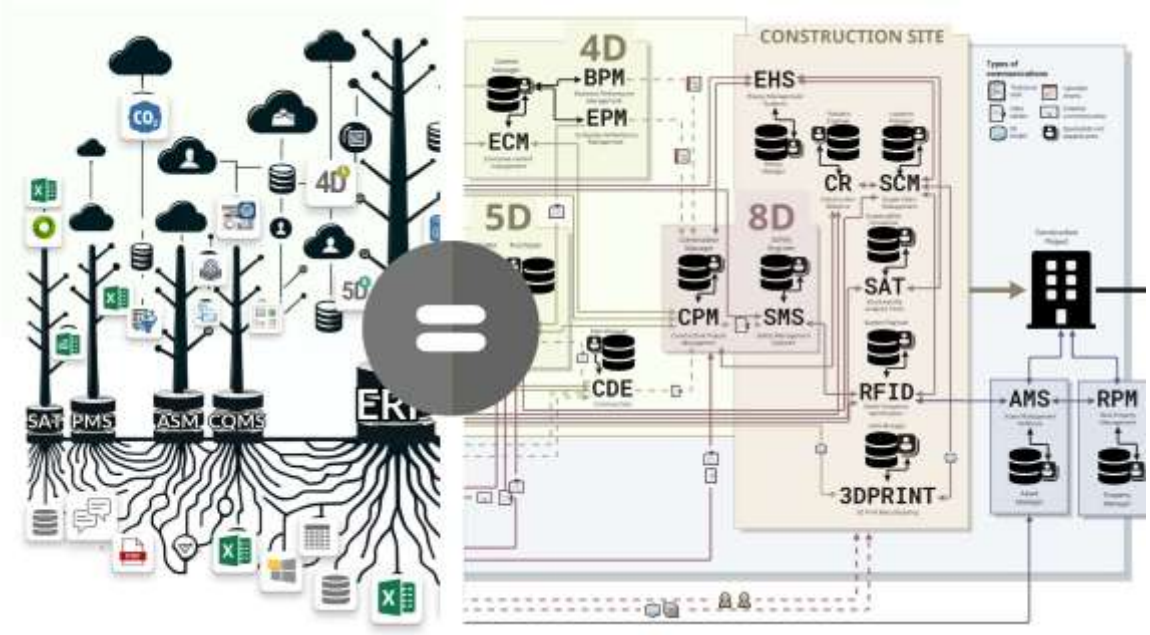


図 4.2-2 健全な企業エコシステムを確保するには、そのシステムの質とタイムリーなリソーシングが必要である。

データの妥当性確認と整合化プロセスを自動化するためには（その後の自動統合のために）、各特定システムに最低限必要なデータ要件を記述することから始めるべきである。これらの要件は で定義される：

- 具体的に何を手に入れる必要があるのか？
- どのような形（構造、形式）で？

- どのような属性が必須なのか？
- 正確さと完全性において、どの程度の許容範囲があるか？

データ要件は、受信し処理される情報の品質、構造、および完全性に関する基準を記述する。例えば、PDF -文書中のテキストについては、業界標準に従って正確にフォーマットされていることが重要である（図 7.2-14 - 図 7.2-16）。CAD -モデルのオブジェクトは、正しい属性（寸法、コード、分類子へのリンク）を持たなければならない（図 7.3-9、図 7.3-10）。また、契約スキャンでは、明確な日付と、金額と主要な用語を自動的に抽出する機能が重要である（図 4.1-7〜図 4.1-10）。

> データ要件を策定し、そのコンプライアンスを自動的にチェックすることは、最も時間がかかるが重要なステップのひとつである。ビジネスプロセスにおいて最も時間のかかるステップである。

本書のパート 3 で述べたように、ビジネスインテリジェンス（BI）担当者の時間の 50%から 90%は、分析よりもデータの準備に費やされています（図 3.2-5）。このプロセスには、データ収集、検証、妥当性確認、整合化、構造化が含まれる。

2016 年の調査[95]によると、様々な広範な分野のデータサイエンティストは、作業時間のほとんど（約 80%）を、最もやりたくないこと（図 4.2-3）、つまり既存のデータセットを収集し、それらを整理（統一、構造化）することに費やしていると回答している。したがって、新たな洞察や発見につながるパターンや規則性を見出すといった創造的な作業に割ける時間は、全体の 20%にも満たない。

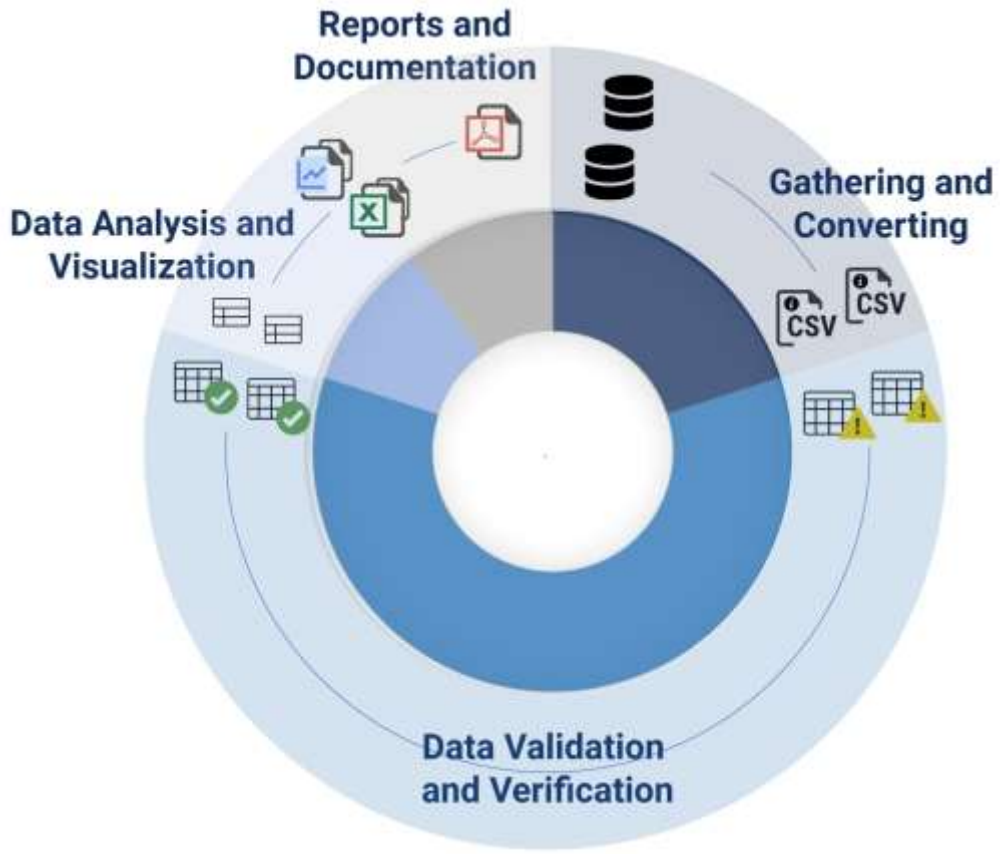


図 4.2-3 を検証し、データ品質を確保すること は、他のシステムに統合するためのデータを準備する上で、最もコストと時間のかかる複雑なステップである。

建設会社でデータ管理を成功させるには、タスクのパラメータ化、データ品質要件の策定、自動検証のための適切なツールの使用など、包括的なアプローチが必要です。

デジタル相互運用性は要件から始まる

企業内のデジタル・システムの数が増えるにつれて、それらのシステム間でデータの一貫性を保つ必要性も高まっています。さまざまな IT システムを担当する管理者は、情報量の増加やフォーマットの多様化に対応しきれないことが多い。そのような状況では、他のアプリケーションやプラットフォームで使用するのに適した形式でデータを作成するよう、専門家に依頼せざるを得なくなります。

このため、データ作成に携わるエンジニアやスタッフは、多くの要件に対応しなければならず、多くの場合、透明性や、データが将来どこでどのように使用されるのかについての明確な理解もないままである。情報の取り扱いに対する標準化されたアプローチの欠如は、データの複雑さと標準化されていない性質のために手作業

になりがちな検証段階での非効率とコスト増につながる。

> データの標準化 の問題は、単なる利便性や自動化の問題ではない。直接的な金銭的損失である。2016 年の IBM の報告書によると、米国における データ品質の低さによる年間損失は 3.1 兆ドルである[96]。さらに、マサチューセッツ工科大学（MIT）や他の分析コンサルティング会社の研究によると、 データ品質の低さがもたらすコストは、企業の収益の 15〜25%にも上ることがある[97]。

このような状況下では、どのようなパラメータを、どのような形式で、どの程度の詳細さで、作成されたオブジェクトに含めるべきかについて、データ要件と説明を明確に定義することが重要になる。このような要件を正式に定義しなければ、データ品質とシステム間およびプロジェクト・フェーズ間の互換性を保証することは不可能である（図 4.2-4）。

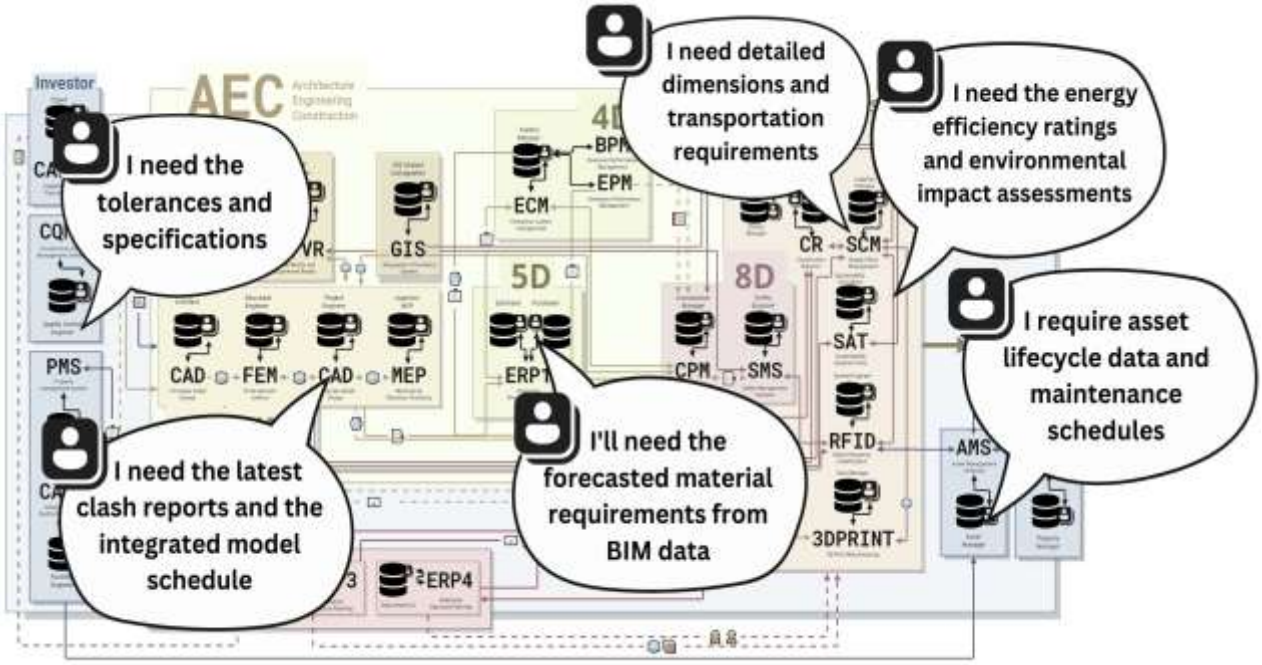


図 4.2-4 ビジネスはさまざまな役割の相互作用の上に成り立っており、それぞれの役割は、ビジネス目標を達成するために重要な、特定のパラメータや値を必要とする。

正しいデータ要件を策定するためには、 、データレベルでビジネスプロセスを理解する必要があります。建設プロジェクトは、その種類、規模、参加者数において様々であり、モデリング（CAD（BIM ））、スケジューリング（ERP 4D）、原価計算（ERP 5D）、ロジスティクス（SCM）など、各システムは、入力（入力エンティティ-要素）に独自のパラメータを必要とします。

これらのニーズに応じて、ビジネス・マネジャーは要件を満たす新しいデータ構造を設計するか、既存のテー

ブルやデータベースを適合させなければならない。作成されるデータの質は、要件がいかに正確かつ正しく策定されているかに直接依存する（図 4.2-5）。

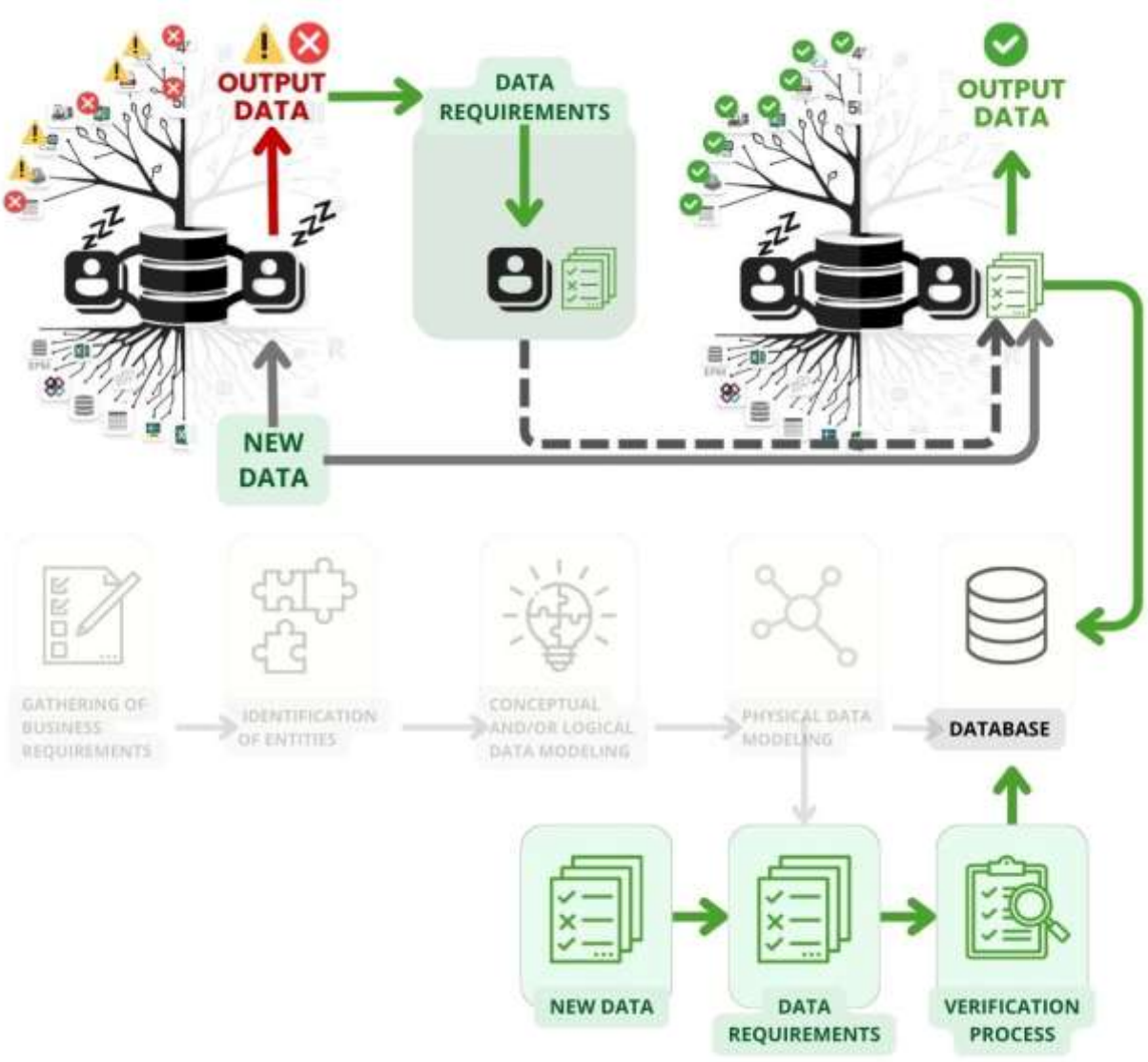


図 4.2-5 データ品質は、特定のデータ使用ケースに対して作成される要件の品質に依存する。

各システムにはそれぞれ固有のデータ要件があるため、一般的な要件を策定するための最初のステップは、ビジネスプロセスに関わるすべての要素を分類することである。これは、オブジェクトを特定のシステムやアプリケーションタスクに対応するクラスやクラスのグループに分ける必要があることを意味する。このようなグループごとに、データ構造、属性、品質に関する個別の要件を策定する。

しかし実際には、データをグループ化するための共通言語がないという大きな課題に直面している。バラバラの分類、重複した識別子、互換性のないフォーマットにより、各企業、各ソフトウェア、さらには各プロジェクトが、それぞれ独自の孤立したデータモデルやクラスを形成している。その結果、システム間で情報を転送するには、適切なデータモデルやクラスに何度も変換する必要があり、多くの場合手作業で行われる、デジタルの「バベルの塔」となってしまう。この障壁は、普遍的な分類器と標準化された要件セットに移行することによってのみ克服できる。

建設の共通言語：デジタル変革における分類器の役割

検査・加工プロセスのデジタル化と自動化において、分類システム（ ）は、オブジェクトの記述とパラメータ化の統一性を保証する「デジタル辞書」のような特別な役割を果たします。分類器は、データを意味ごとにグループ化し、異なるシステム間、管理レベル間、プロジェクトライフサイクルのフェーズ間で統合することを可能にする「共通言語」を形成する。

分級機の最も具体的な影響は、建物のライフサイクルの経済性であり、最も重要な側面は、長期的な運用コストの最適化である。調査によれば、運用コストは、建築物の総所有コストの最大 80%を占め、これは初期建設コストの 3 倍である（図 4.2-6）[98]。このことは、将来のコストに関する決定は、設計段階でほぼ形成されることを意味する

> このため、オペレーション・エンジニア（CAFM 、AMS 、PMS 、RPM ）からの要件（ ）が、設計段階でデータ要件を生成するための出発点となるべきである（図 1.2-4）。これらのシステムは、プロジェクトの最終段階としてではなく、コンセプトから解体まで、プロジェクトのデジタル・エコシステム全体の不可欠な部分としてとらえるべきである

最新の分類器 は、単なるグループ分けのためのコード体系ではない。建築家、エンジニア、積算担当者、ロジスティクス担当者、メンテナンス担当者、IT システムの相互理解のためのメカニズムである。自動車の自動操縦装置が道路上の物体を一義的に正確に認識しなければならないように、デジタル建設システムとそのユーザーは、同じプロジェクト要素を、要素クラスを通じて、異なるシステムに対して一義的に解釈しなければならない。

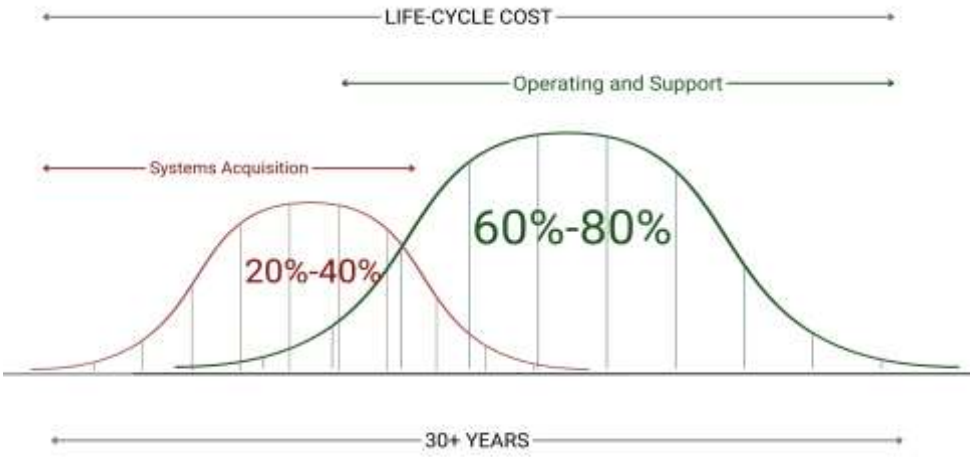


図 4.2-6 運用維持費は建設費の 3 倍を上回り、建物のライフサイクルコスト全体の 60〜80%を占める（[99]による）。

分類器の開発レベルは、企業のデジタル化の深度やデジタル成熟度と直接相関している。デジタル成熟度が低い組織は、断片化されたデータ、互換性のない情報システム、そしてその結果としての互換性のない非効率的な分類器に直面しています。このような企業では、同じ要素が異なるシステムで異なるグルーピング識別子を持つことがしばしばあり、これが最終的な統合を決定的に妨げ、プロセスの自動化を不可能にしている。

例えば、プロジェクトの同じ窓が、CAD（ ）モデル、見積もり、メンテナンスシステム（図 4.2-7）において異なるラベルが付けられることがある。窓のカテゴリー要素では、積算担当者にとっては体積とコストが重要であり、保守サービス担当者にとっては可用性と保守性が重要であり、建築家にとっては美観と機能特性が重要である。その結果、同じ要素でも異なるパラメータが必要になることがある。

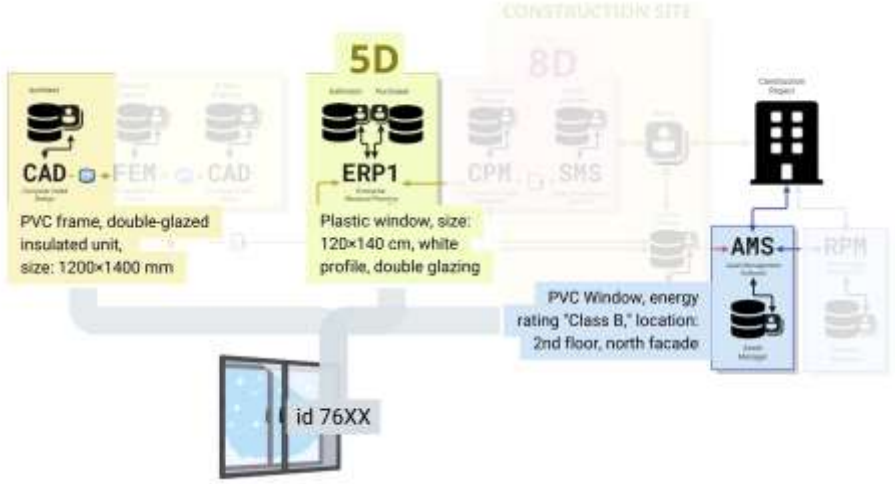


図 4.2-7 システム間の一貫性のない分類 要素は、他のシステムへの移行の各段階で、いくつかの属性情報を失う。

建築要素の分類を明確に定義することが困難なため、異なる分野の専門家が同じ要素に互換性のないクラスを割り当てることがよくあります。その結果、対象物の統一された見解が失われ、異なる分類システムを調和させ、異なる専門家によって定義されたタイプやクラス間の整合性を確立するために、手作業による介入が必要となる。

このような不整合の結果、建設資材をメーカーから購入する際に調達部門（ERP）が受け取る業務文書が、建設現場におけるその資材の分類と正しくリンクされないことが多い（PMIS 、SCM）。その結果、重要な情報がインフラや資産管理システム（CAFM、AMS）に統合されない可能性が高くなり、試運転時や、その後のメンテナンス（AMS、RPM）、要素の交換時に深刻な問題が生じる。

> デジタル成熟度の高い企業では、分類器はすべての情報の流れを統合する神経系の役割を果たします。同じ要素に一意の識別子が付与されるため、CAD 、ERP 、AMS 、CAFM の各システムとその分類器の間で、歪みや損失なしに転送することができます。

効果的な分類器を構築するには、データがどのように使用されているかを理解する必要がある。同じエンジニアでも、プロジェクトが異なれば、要素の名前も分類も異なるかもしれません。長年にわたる使用統計を収集することによってのみ、安定した分類システ ムを開発することができる 。機械学習は、これを支援する。アルゴリズムが何千ものプロジェクトを分析し（図 9.1-10）、機械学習によって可能性の高いクラスとパラメーターを特定する（図 10.1-6）。自動分類は、データ量が多いために手作業による分類が不可能な環境では、特に価値がある。自動分類システムは、最小限の項目パラメータに基づいて、基本的なカテゴリーを区別すること

ができるようになる（詳細は第 9 回と第 10 回で述べる）。

開発された分級機システムは、さらなるデジタル化の触媒となり、その基盤を作る：

- プロジェクトのコストとスケジュールの自動見積もり。
- 潜在的なリスクとコンフリクトの予測分析
- 調達プロセスと物流チェーンの最適化
- 建物や構造物のデジタルツインを作成する
- スマートシティやモノのインターネット・システムとの統合

> 変革のための時間は限られている。機械学習（）とコンピューター・ビジョン技術の発展により、何十年も解決できなかった自動分類の問題（）は、今後数年で解決されるだろう。時間内に適応できなかった建設会社や設計会社は、デジタル・プラットフォームに駆逐されたタクシー会社の運命を繰り返すリスクがある。

原価計算とスケジューリングの自動化、ビッグデータと 機械学習については、本書の第 5 部と第 9 部で詳しく取り上げる。第 10 部では、タクシーフリートの運命と建設業界のウーバー化が繰り返されるリスクについて詳述する。

建設業界のデジタル変革における分類機の重要な役割を理解するには、その進化の歴史に目を向ける必要がある。歴史的な背景を知ることで、分類へのアプローチがどのように進化してきたのか、また、どのようなトレンドが現在の状況を定義しているのかを理解することができる。

Masterformat, OmniClass, Uniclass and CoClass: 分類システムの進化

歴史的に、建設要素と作業分類器は 3 つの世代で進化しており、それぞれが特定の時代における利用可能な技術レベルと業界の現在のニーズを反映している（図 4.2-8）：

- 第一世代（1950 年代前半〜1980 年代後半） - 紙のディレクトリ、ローカルで使用される階層型分類器（Masterformat、SfB など）。
- 第 2 世代（1990 年代後半から 2010 年代半ば）は、Excel（）と Access（ASTM E 1557, OmniClass, Uniclass 1997）で実装されたスプレッドシートと構造化データベースである。

- 第 3 世代（2010 年代から現在） - デジタルサービスと API - インターフェース、CAD（BIM）との統合、自動化（Uniclass 2015、CoClass）。

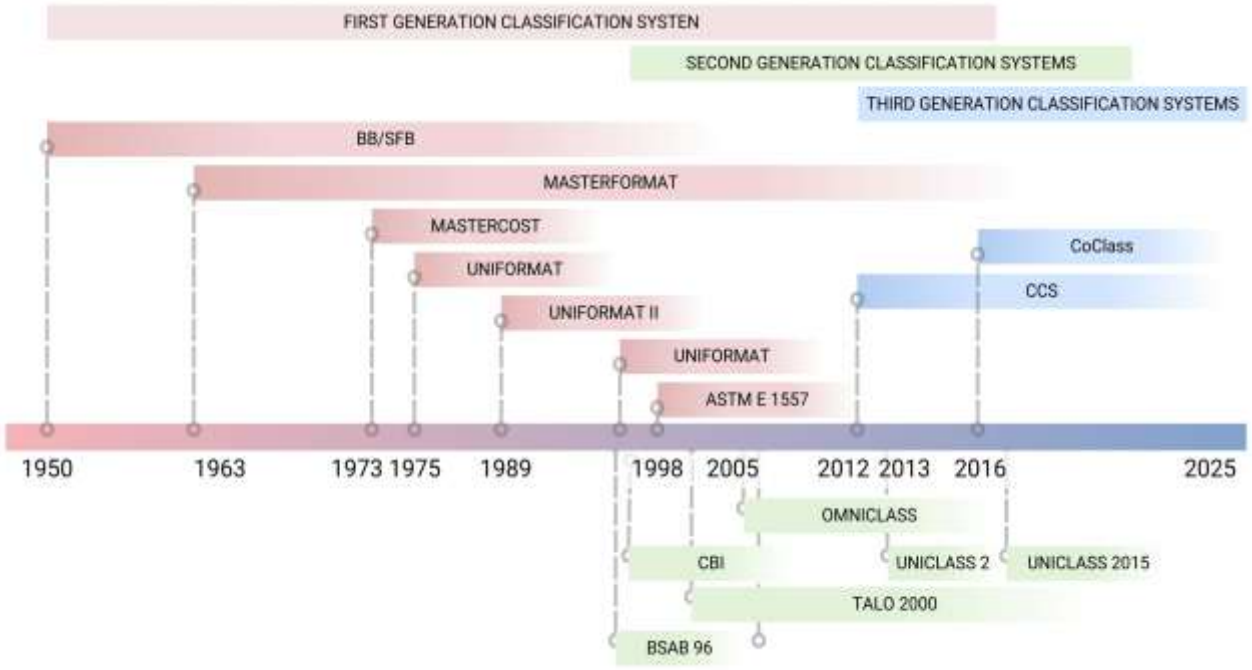


図 4.2-8 三世代の建設業用分類器。

OmniClass のような初期のシステムは、6887 のクラスを記述するために最大 7 つの入れ子レベルを使用していましたが、CoClass のような最新のソリューションは、750 のクラスで 3 つのレベルに制限されています。これにより、必要な粒度を維持しながら、データを扱うことが容易になりました。英国で標準としてよく使用されている Uniclass 2015 は、7210 のクラスをわずか 4 つのレベルにまとめており、CAD プロジェクトや公共調達に便利です。

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

図 4.2-9 分類器の世代が新しくなるたびに、分類の複雑さは飛躍的に減少している。

各国の建設積算システムでは、コンクリート基礎壁のような典型的な要素であっても、分類が異なるため、非常に異なる方法で記述されることがある（図 4.2-10）。これらの違いは、各国の建設慣行、使用される測定システム、材料分類のアプローチ、各国で施行されている規制や技術要件を反映しています。

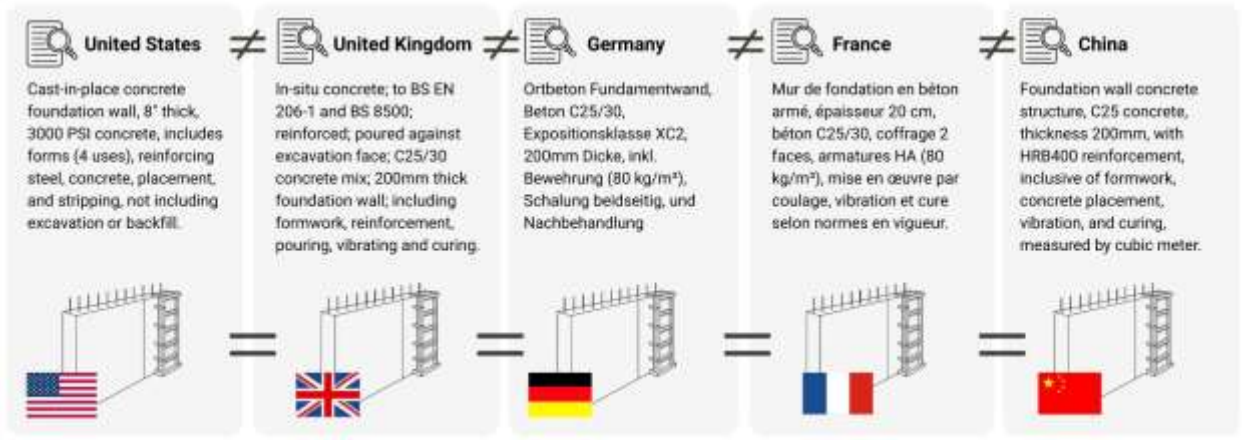


図 4.2-10 同じエレメントが、異なる国のプロジェクトでは、異なる記述と分類によって使用される。

同じ要素でも多様な分類があるため、国際協力が複雑になり、国際プロジェクトにおけるコストや作業範囲の比較に時間がかかり、時にはほとんど不可能になる。現在のところ、世界レベルで共通の分類法はなく、それぞれの国や地域が、その土地の規範、言語、ビジネス文化に基づいて独自のシステムを構築している：

- CCS（デンマーク）：コスト分類システム - 施設のライフサイクル（設計、建設、運転）を通して、 コストを分類するシステム。運転と保守の論理に重点が置かれているが、予算や資源管理も含まれる。
- NS 3451（ノルウェー）：施設を機能、設計要素、ライフサイクル・ステージごとに分類。プロジェクト管理、コスト見積もり、長期計画に使用される。
- マスターフォーマット（MasterFormat）（米国）：工事仕様書をセクション（コンクリート、電気、

仕上げなど）に分けて構造化するシステム。UniFormat とは異なり）機能要素よりも分野や作業タイプに重点を置く。

- Uniclass 2 (UK): 最も詳細な分類法のひとつで、公共調達や BIM プロジェクトで使用される。オブジェクト、作品、材料、空間に関するデータを単一のシステムに統合。
- OmniClass: コンポーネント・ライブラリから電子仕様書までのオブジェクト情報を管理するための国際標準（米国 CSI が開発）。長期間のデータ保存に適しており、CAD (BIM)やその他のデジタルツールと互換性がある。
- COBie: Construction-Operation Building information exchange は、設計・施工・運用の各段階におけるデータ交換のための国際規格。BIM -model ready for use」のコンセプトの一部として BS 1192-4:2014 に含まれる。情報伝達（設備仕様、保証、請負業者の連絡先など）に重点を置いている。

建設業界のグローバル化は、建築要素分類システムの段階的な統一につながる可能性が高い。このプロセスは、普遍的なデータ転送プロトコルが最終的にローカルな形式を置き換え、システムのグローバルな相互運用性を確保するという、インターネット通信の進化の流れに沿って発展する可能性がある。

別の発展経路としては、機械学習技術に基づく自動分類システムへの直接的な移行が考えられる。このような技術は、今日、主に自律輸送の分野で開発されているが、大規模な CAD 設計データセットに適用できる大きな可能性を秘めている（図 10.1-6）。

> 今日、このような状況は、クラシファイヤーが国ごとにまとまっていることだけにとどまらない。国レベルでは考慮されない多くの特殊性があるため、各企業は自社が扱う要素や資源のカテゴリーを統一し、標準化しなければならない。

原則として、このプロセスは、オブジェクトのローカルな表や社内のラベリングシステムなど、小規模なものから始める。しかし、戦略的な目標は、すべての要素を記述するための共通言語に移行することである。この共通言語は、社内だけでなく社外でも理解できるものであり、理想的には、国際的な分類基準や業界の分類基準と調和したものである（図 4.2-8）。このアプローチは、外部パートナーやデジタル・システムとの統合を容易にし、モノのライフサイクルの中で統一されたエンド・ツー・エンドのプロセスの形成を促進する。

自動化と拡張可能な IT システムに移行する前に、国家レベルの分類器を使用するか、独自の論理的で曖昧さのない要素識別構造を構築する必要がある。窓（図 4.2-11）であろうと、ドアであろうと、 エンジニアリング・システムであろうと、あらゆるオブジェクトは、どの企業のデジタル・システムでも紛れもなく認識できるように記述されなければならない。これは、平面の図面からデジタル・モデルへの移行において、建築物の設計

段階と運用の両方をカバーする上で非常に重要である。

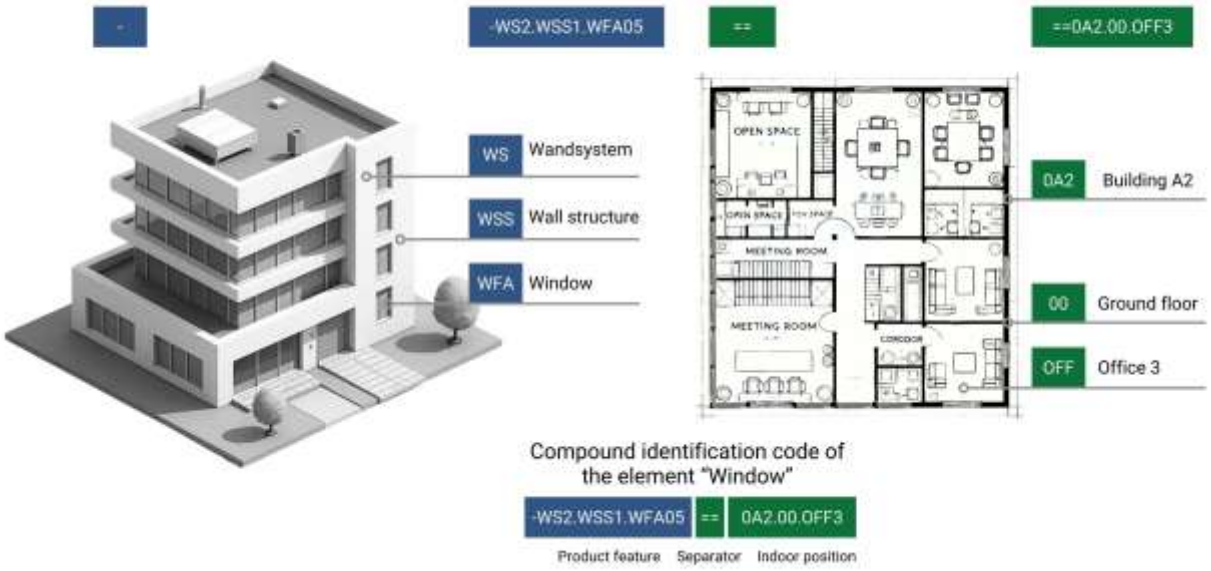


図 4.2-11 分類および建物内の位置に基づく複合窓建物要素識別子の例。

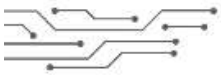
内部分類法の一例として、複合識別コード（図 4.2-11）の開発が考えられる。このようなコードは、要素の機能的目的（「壁の窓」など）、タイプ、正確な空間参照（建物 A2、フロア 0、部屋 3）など、いくつかのレベルの情報を組み合わせたものである。このようなマルチレベルの構造により、特にデータの検証や変換の段階で、要素の明確なグループ化が必要となる、デジタルモデルやドキュメントを通じた統一されたナビゲーションシステムを構築することができます。要素を明確に認識することで、部門間の一貫性が確保され、重複、エラー、情報損失のリスクが軽減されます。

よく構築された分類器 は単なる技術文書ではなく、企業のデジタル・エコシステムの基盤である：

- システム間のデータの互換性を確保する；
- 情報の検索と処理のコストを削減する；
- 透明性と管理性が向上する；
- は、スケーリングと自動化のための基盤を構築する。

国家分類や独自の複合識別コードを使用した対象物の標準化された記述は、一貫性のあるデータ、 信頼性のある情報交換、そして自動調達からデジタル・ツインに至るインテリジェント・サービスの実装の基礎となる。

マルチフォーマットデータの構造化段階を完了し、要素の認識とグループ化に使用する分類器を選択したら、次のステップはデータを正しくモデル化することである。このプロセスでは、重要なパラメータを特定し、論理的なデータ構造を構築し、要素間の関係を記述します。



章 4.3.

データモデリングとセンター・オブ・エクセレンス

データモデリング：概念モデル、論理モデル、物理モデル

データの効果的な管理（先に構造化と分類を行った）は、よく考えられた保管と処理の構造なしには不可能である。保存と処理の段階で情報へのアクセスと一貫性を確保するために、企業はデータモデリングという方法論を用いて、ビジネス要件に従ってテーブル、データベース、およびそれらの間のリンクを設計する。

> データモデリングは、あらゆるデジタル・エコシステムを構築する基盤である。システム、要件、データモデリングの記述がなければ、データを作成するエンジニアや専門家は、作成したデータがどこで使用されるかを知らないし、理解することもできない。

建築物を建てるとき、計画なしにレンガを積み始めることができないように、データウェアハウスシステムを構築するには、どのようなデータを使用し、どのようにリンクさせ、誰がそれを扱うのかを明確に理解する必要がある。プロセスや要件の説明がなければ、データを作成するエンジニアや専門家は、データが将来どこでどのように使用されるかを見失ってしまう。

データ・モデルはビジネスと IT の架け橋となる。それは、、情報を構造化し、利害関係者間のコミュニケーションを促進する。この意味で、データ・モデリングは、建築家の仕事に似ている。建築家は顧客の計画に従って建築計画を作成し、それを施工者（データベース管理者や開発者）に渡して実装（データベース作成）してもらう。

したがって、すべての建設会社は、要素やリソースを構造化・分類することに加えて（図 4.2-11）、データベース（テーブル）を「構築」する技術を習得し、あたかもレンガを繋ぎ合わせて会社のデータから知識の信頼できる強固な壁を作るかのように、データベース間のリンクを作成する方法を学ばなければならない。データモデリング（図 4.3-1）の主要概念は以下の通りである：

- エンティティは、データを収集しなければならない対象である。設計の初期段階では、エンティティは単一の要素（たとえば「ドア」）であることもあれば、見積もりモデルでは、カテゴリーごとにグループ化された要素のグループ（たとえば「内部ドア」）であることもあります。
- 属性とは、重要な詳細（寸法、特性、組立コスト、物流、その他のパラメータ）を記述するエンティ

ティの特性である。

- リレーションシップ（リンク） - エンティティ同士がどのように相互作用するかを示す。関係には以下のタイプがある：「1 対 1"、"多対 1"、"多対多"。
- ER 図（Entity-Relationship diagram）は、 エンティティ、属性、およびそれらの間の関係を示す視覚的な図である。ER 図には、概念図、論理図、物理図があり、それぞれ異なる詳細レベルを反映します。

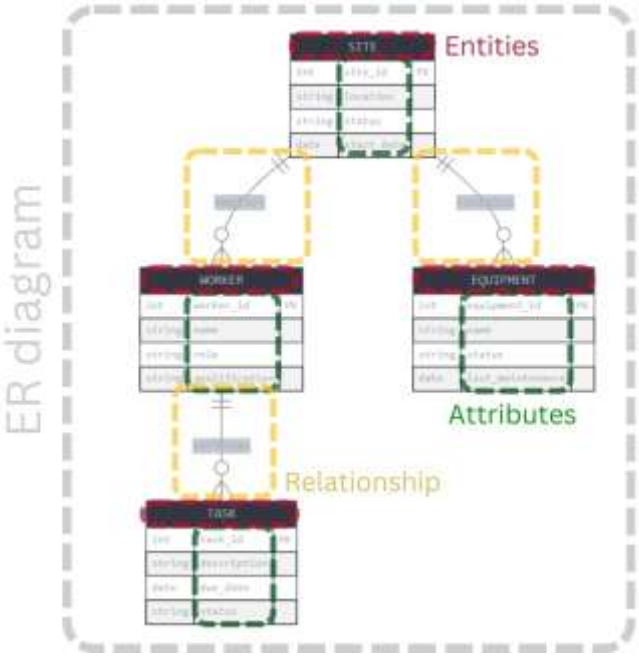


図 4.3-1 エンティティ、属性、関係を持つ概念データベース構造の ER 図。

データを設計し、データ間の関係を定義するプロセスは、伝統的に 3 つの主要なモデルに分けられる。それぞれ特定の機能を果たすもので、データ構造を表現する際の詳細度や抽象度が異なる：

- 概念的データモデル: このモデルは、属性の詳細には立ち入らずに、主なエンティティとその関係を記述する。通常、計画の初期段階で使用される。この段階では、データベースやシステムからスケッチし、異なる部門や専門家の関係を示すことができる。

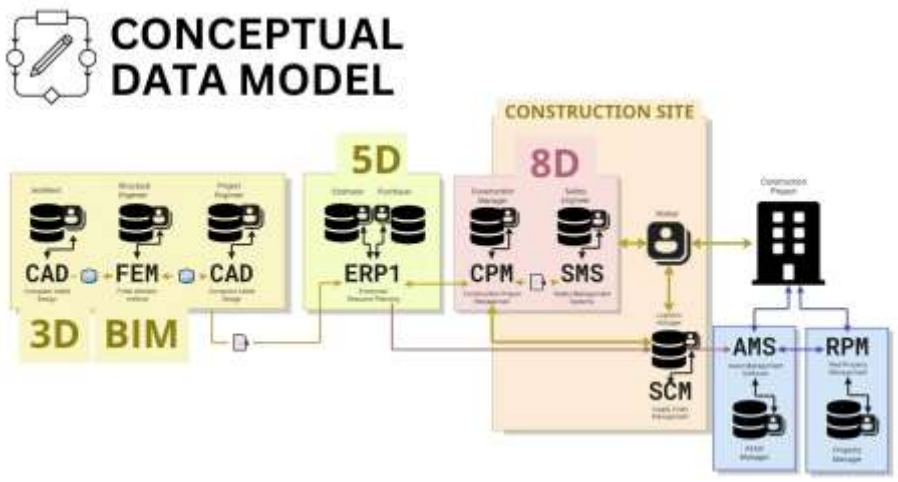


図 4.3-2 概念図はシステムの内容を説明するもので、技術的な詳細は含まず、関係を高レベルで表現したものである。

- 論理データモデル: 概念モデルに基づき、論理データモデルにはエンティティ、属性、キー、関係の詳細な記述が含まれ、ビジネス情報とルールがマッピングされる。

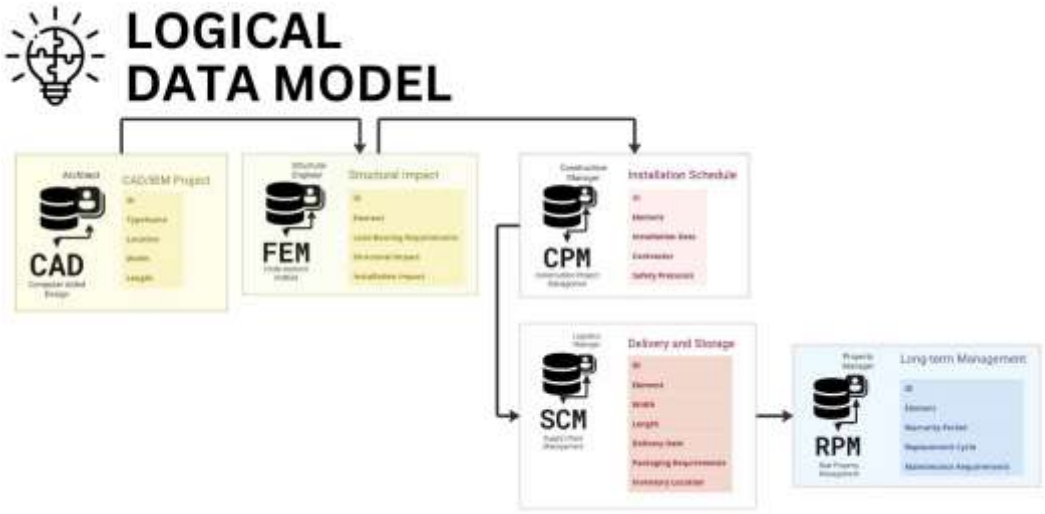


図 4.3-3 論理データモデル データ型、リレーションシップ、キーの詳細を示すが、システム実装はない。

- Physical Data Model: このモデルは、テーブル、カラム、リレーションシップなど、データベースを実装するために必要な構造を記述している。データベースの物理的な配置を最適化するために、データベースのパフォーマンス、インデックス戦略、物理ストレージに焦点を当てている。

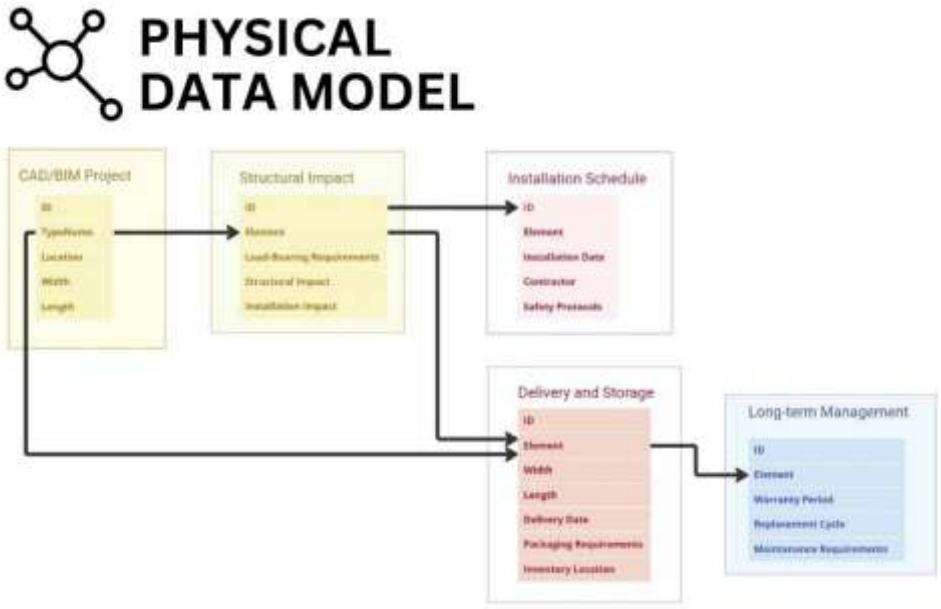


図 4.3-4 物理データモデル は、テーブルやデータベースの具体的な詳細を含め、システムがどのように実装されるかを定義する。

> データベースを設計し、表形式の関係を設計する際、抽象化レベルを理解することは、効果的なシステム・アーキテクチャを構築する上で重要な役割を果たす。

効果的なデータモデリング手法によって、ビジネス目標と技術的な実装を組み合わせることができ、プロセスチェーン全体がより透明で管理しやすくなります。データモデリングは単発の作業ではなく、連続したステップを含むプロセスである（図 4.3-5）：

- ビジネス要件の収集：主要なタスク、目的、情報の流れが定義される。これは、専門家やユーザーとの積極的な対話の段階です。
- エンティティの識別：将来のシステムで考慮すべき主なオブジェクト、カテゴリー、データタイプが強調されている。
- 概念モデルと論理モデルの開発：まず主要なエンティティとその関係を把握し、次に属性、ルール、詳細な構造を把握する。
- テーブル、フィールド、リレーションシップ、制約、インデックス。
- データベースの作成：最後のステップは、選択した DBMS に物理モデルを実装し、テストを実施し、運用の準備をすることです。

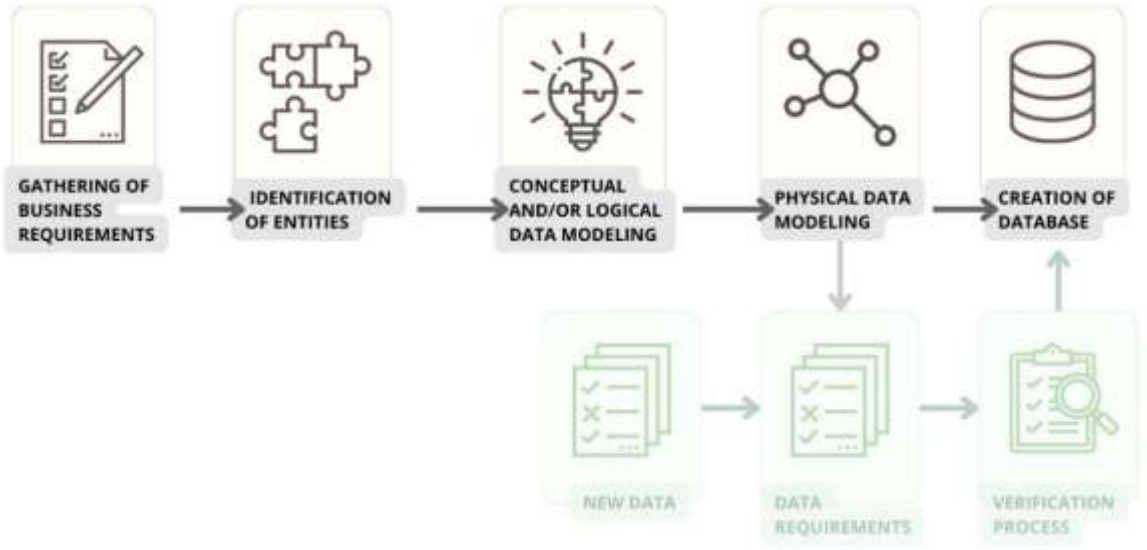


図 4.3-5 ビジネス・プロセスのためのデータベースとデータ管理システムの作成は、要件生成とデータ・モデリングから始まる。

適切に設計されたデータモデリングプロセスは、透明性のある情報の流れを可能にし、これは建設プロジェクトや現場管理のような複雑なプロジェクトでは特に重要である。概念モデルから論理モデル、そして物理モデルへと移行することで、どのようにプロセスを合理化できるかを見てみよう。

建設における実践的データモデリング

データモデリングの例として建設現場管理タスクを取り上げ、 現場監督の要件を構造化ロジックモデルに変換してみましょう。現場管理の基本的なニーズに基づいて、現場（SITE）、作業員（WORKER）、設備（EQUIPMENT）、タスク（TASK）、設備使用状況（EQUIPMENT_USAGE）のキー・エンティティを定義する。各エンティティには、重要な特性を反映する一連の属性が含まれる。例えば、TASK の場合、タスクの説明、期限、ステータス、優先度などである。WORKER の場合、名前、サイトでの役割、現在の雇用などである。

ロジックモデルは、これらのエンティティ間の関係を確立し、実際の作業プロセスにおいてそれらが どのように相互作用するかを示す（図 4.3-6）。例えば、サイトとワーカーの関係は、多くのワーカーが 1 つのサイトで働くことができることを示し、ワーカーとタスクの関係は、1 人のワーカーが複数のタスクを実行することができることを反映しています。

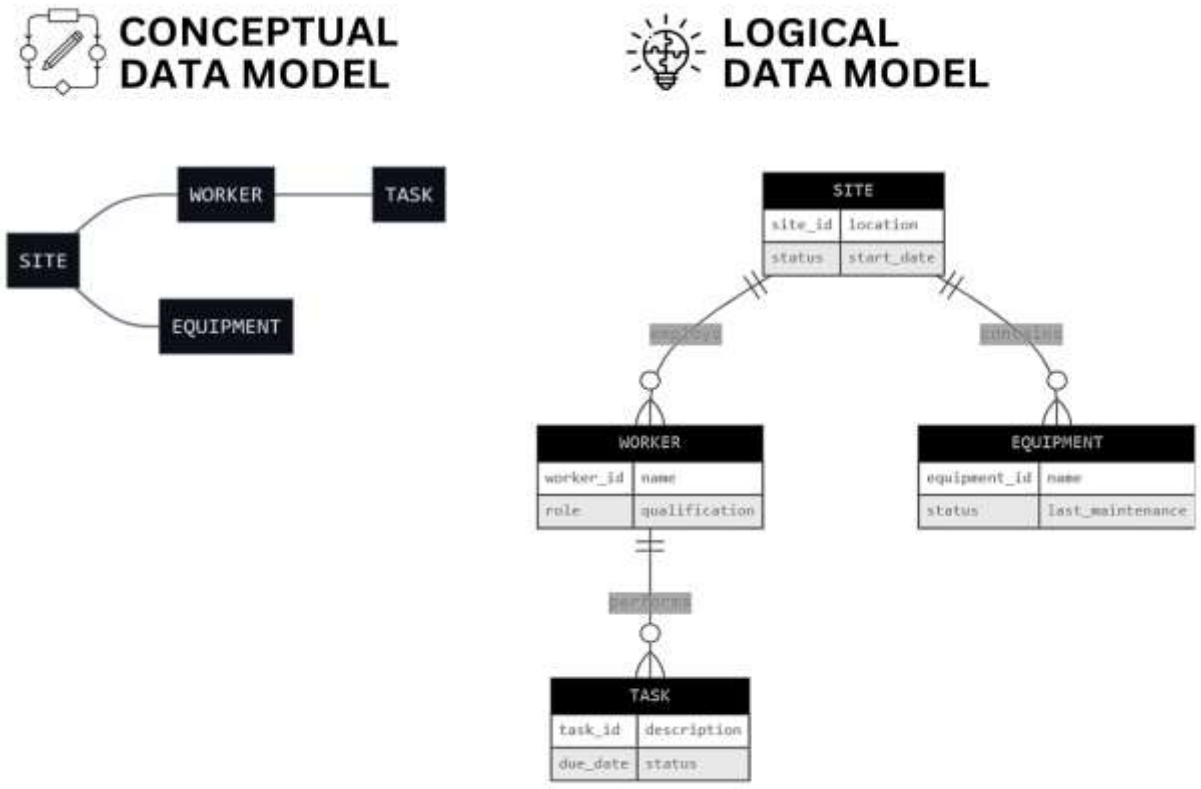


図 4.3-6 建設現場のプロセスを記述するための現場監督の要求によって生成された概念的・論理的データモデル。

物理モデルに移行すると、特定のデータ型（VARCHAR、INT、DATE）、テーブル間のリレーションシップのための主キーと外部キー、データベースのパフォーマンスを最適化するためのインデックスなど、技術的な実装の詳細が追加されます（図 4.3-7）。

例えば、ステータスに特定の型と可能な値を定義し、ステータスやワーカーID のような重要なフィールドにインデックスを追加して検索性能を向上させる必要があります。こうすることで、システムの論理的な記述が具体的なデータベース実装計画に変わり、すぐに作成・実装できるようになります。

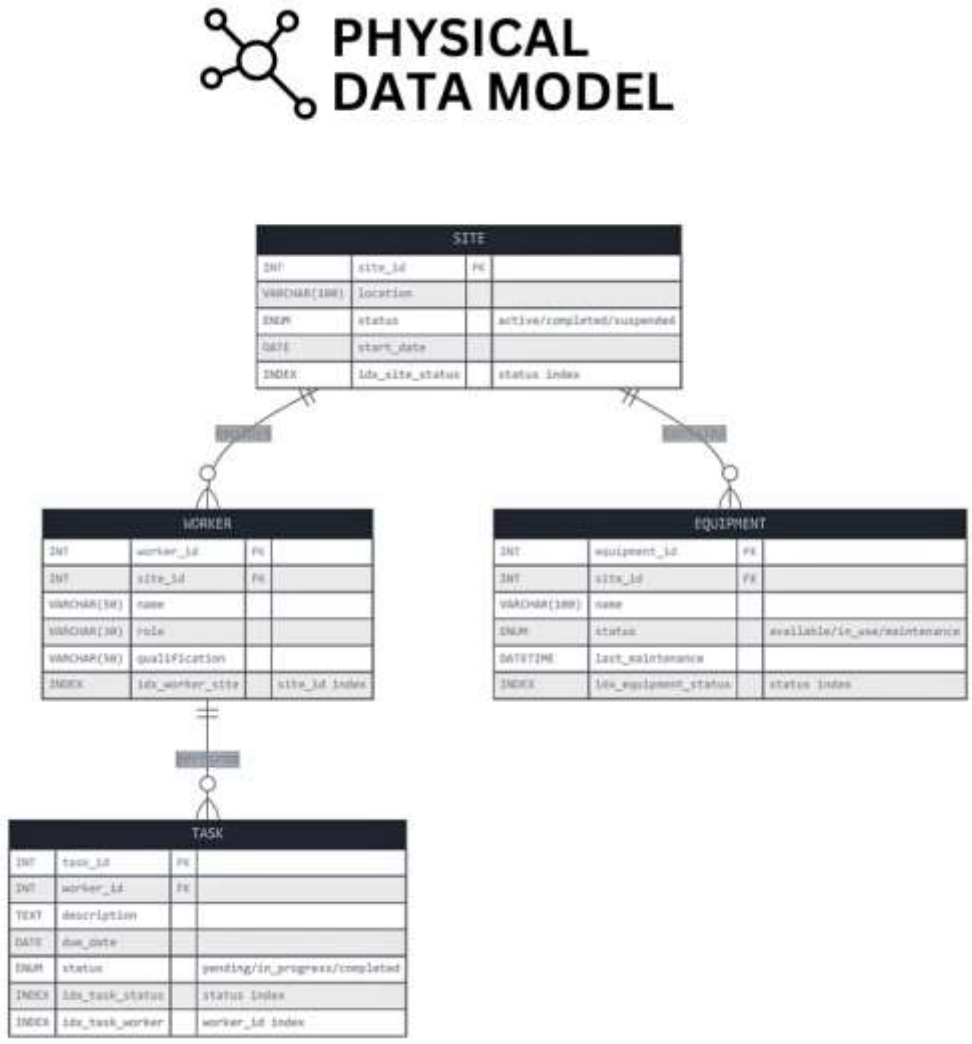


図 4.3-7 物理データモデル は、必要最小限のパラメータを通して建設現場の実体を記述する。

> 物理モデルは論理モデルとは異なることが多い。平均的なモデリング時間の配分は以下の通りである。約 50%が概念モデル（要求の収集、プロセスの議論、エンティティの特定）、10%が論理モデル（属性と関係の特定）、40%が物理モデル（実装、テスト、DBMS への適合）に費やされる。

このバランスは、概念段階ではデータ構造の基礎が築かれ、論理モデルでは関係性と属性のみが指定されるという事実によるものである。物理モデルは、データを特定のプラットフォームやツールに実装する段階であるため、最も多くのリソースを必要とする

LLM を使ったデータベースの作成

データモデルと、パラメータによるエンティティの記述ができれば、データベース（ストレージ）を作成する準備が整う。

プログラミング言語 Python を例に、SQLite を使って、最小限のコードでシンプルだが機能的なデータベースの例を作ってみよう。リレーショナル・データベース については、"構造化リレーショナル・データベースと SQL クエリ言語 " の章で詳しく説明しました。

> SQLite は軽量、オープンソース、組み込み可能、クロスプラットフォームデータベース であり、SQL のコア機能 をサポートしている。SQLite は、CAD (BIM)、ERP 、EPM で広く使用されており、ほとんどすべてのスマートフォンに搭載されています。

FEM (構造的影響) システムと CPM (組立スケジュール) システムについて、上記で作成した物理モデル (図 4.3-4) に基づいて、LLM モデルに小さな SQLlite データベースを作成してもらいましょう。

- LLM チャット（ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN またはその他）にテキストリクエストを送る：

> 添付の図から「建設的影響」と「設置スケジュール」システムの SQLlite データベースを作成する（スキーマを画像としてチャットに貼り付ける（図 4.3-4

LLM の回答 ：

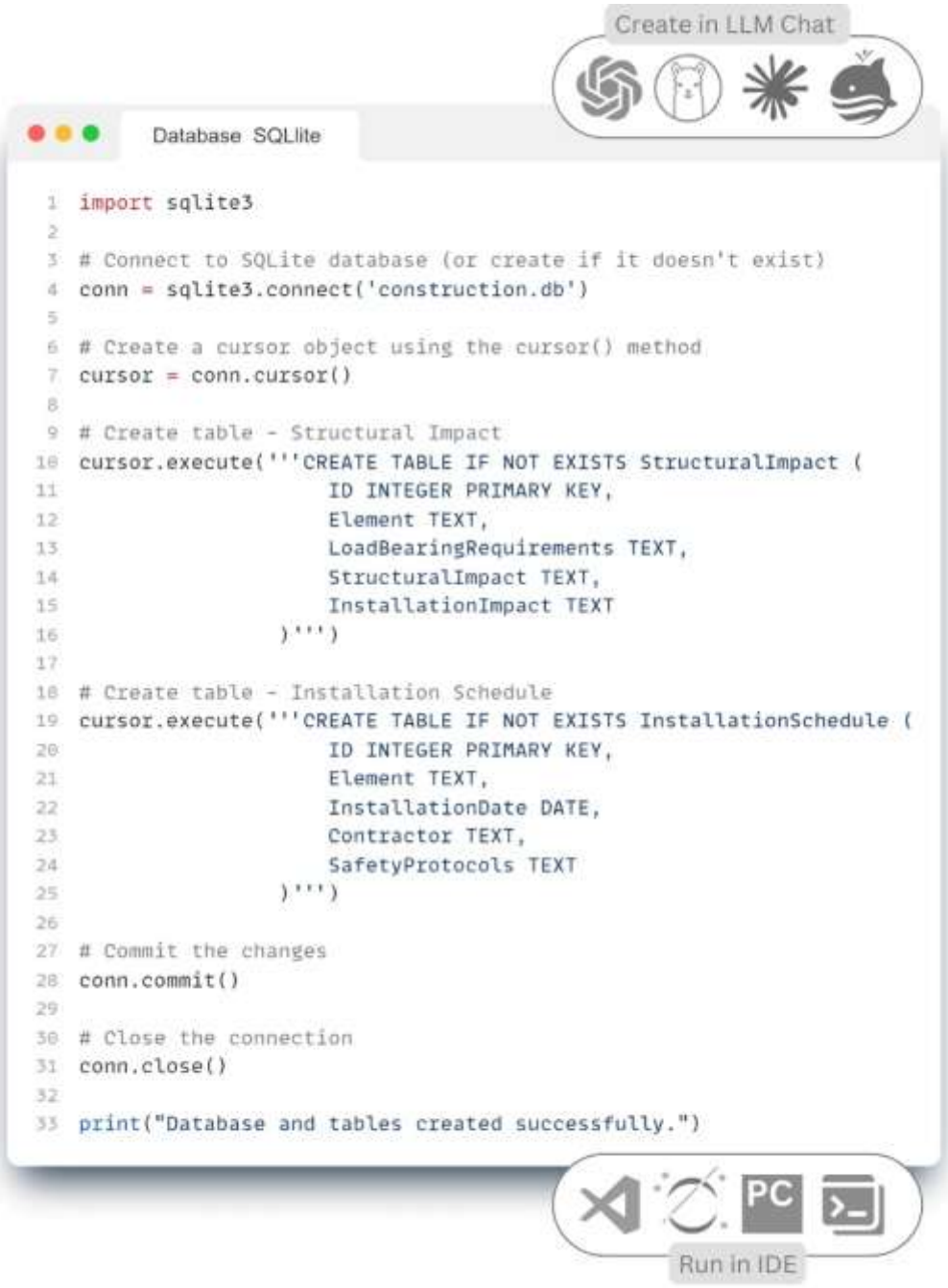


図 4.3-8 生成されたコードは、アップロードされた画像からデータベースとテーブルを作成し、そこにデータをロードする。

このデータベースコードは、通常標準の Python 環境に含まれている SQLite サポート 、Python 環境 。スクリプトの実行やファイルの作成が可能なローカルの Python マシンであれば、どこでも実行できます。このコードが実行されると、ハードドライブにファイルが作成されます。SQLite データベース の名前は construction.db（ワークシートを持つ Excel ファイル に似ています）。

これらのテーブルに新しい construction.db データベースが作成されると、SQL _クエリやインポートによってこれらのテーブルにデータを追加することができます。データは、CSV ファイル 、Excel スプレッドシートから SQLite データベースにインポートしたり、API を介して他のデータベースやリポジトリからエクスポートすることができます。

持続可能なデータモデリングと効果的なデータベース管理プロセスを確立するためには、企業には明確に定義された戦略と、技術チームとビジネスチーム間の調整が必要です。異なるプロジェクトや複数のデータソースでは、すべてのレベルで一貫性、標準化、品質管理を確保することは困難な場合が多い。一つの重要な解決策は、社内にデータモデリング専門のセンター・オブ・エクセレンス（COE）を設立することであろう。

データモデリング・センター・オブ・エクセレンス（CoE）

データが重要な戦略的資産のひとつとなりつつある現在、企業は単に情報を正しく収集・保管するだけでなく、データを体系的に管理する方法を学ぶことが重要である。Classification and Data Modelling Centre of Excellence (CoE) は、組織内のすべてのデータ取り扱いの一貫性、品質、効率を保証する構造的なユニットである。

> センター・オブ・エクセレンス（CoE）は、専門家によるサポートの中核であり、企業のデジタルトランスフォーメーションのための方法論的基盤である。データ駆動型の文化を構築し、直感やローカルな情報ではなく、構造化され、検証された代表的なデータに基づいて意思決定を行うプロセスを構築することを可能にする。

データ・センター・オブ・エクセレンスは通常、「ピザ 2 枚の原則」に従って働く部門横断チームから形成される。この原則は、ジェフ・ベゾスが提唱したもので、チームの規模はピザ 2 枚でまかなえる程度、つまり 6〜10 人以下とすることを意味する。このアプローチは、過剰な官僚主義を避け、仕事の柔軟性を高めるのに役立つ。CoE チームには、データ分析（）や機械学習（）から特定のビジネス分野の専門知識まで、さまざまな

技術スキルを持つ社員を含めるべきである。データエンジニアは、その深い技術的知識によって、プロセスの最適化やデータのモデリングを行うだけでなく、ルーティンワークの時間を短縮することによって同僚をサポートする必要がある（図 4.3-9）。

> 自然界において生態系の回復力が生物多様性によって確保されているように、デジタル世界においても柔軟性と適応性は、データを扱うための多様なアプローチによって達成される。しかし、この多様性は共通のルールと概念に支えられていなければならない。

センター・オブ・エクセレンス（CoE ）は、森林生態系の「気候条件」に例えることができる。質の高いデータにとって好ましい "気候 "を作り出すことで、CoE はベストプラクティスや方法論の自然な選択を促し、それが組織の標準となる。

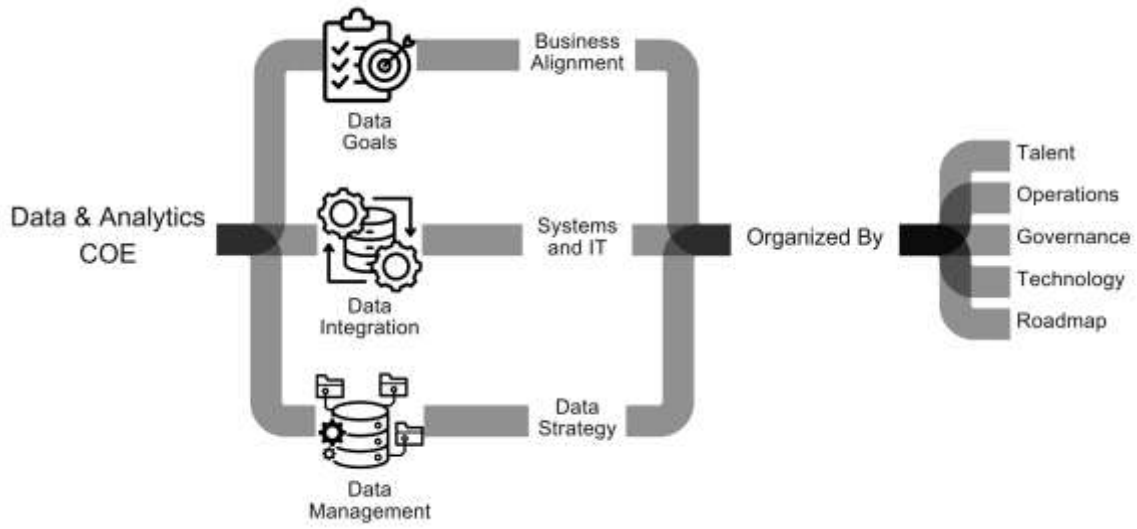


図 4.3-9 データとアナリティクスのためのセンター・オブ・エクセレンス（CoE ）は、データ管理、統合、戦略の重要な側面に関する専門知識を結集している。

統合サイクルを加速し、より良い結果を得るために、CoE は、意思決定において十分な自律性をメンバーに与えるべきである。これは、試行錯誤、絶え間ないフィードバック、頻繁なリリースが大きな利益をもたらすダイナミックな環境においては特に重要である。しかし、この自主性は、上級管理職からの明確なコミュニケーションとサポートがある場合にのみ有効である。戦略的ビジョンとトップレベルの調整がなければ、どんなに有能なチームであっても、イニシアチブを実行する上で障害に直面する可能性がある。

> データモデリングのアプローチが 1 つか 2 つのプロジェクトに限定されるのではなく、全体的な情報管理とビジネスプロセス管理システムに組み込まれるようにする責任があるのは、企業の COE または上級管理職である。

専門家センター（CoE）は、データモデリングとデータガバナンスに関する業務に加えて、データインフラストラ クチャの展開と運用に関する共通の標準とアプローチを開発する責任を負う。さらに、組織における継続的改善、プロセスの最適化、データの効率的利用の文化を醸成する（図 4.3-10）。

CoE におけるデータとモデルの管理に対する体系的なアプローチは、いくつかの重要なブロックに大別できる：

- プロセスおよびモデルライフサイクル管理の標準化： CoE データモデルの作成と管理を統一するための方法論を開発し、実施する。これには、構造テンプレート、品質管理方法、バージョン管理システムの確立が含まれ、作業の全段階にわたってデータの継続性を確保する。
- 役割管理と責任分担： COE は、データモデリングプロセスにおける主要な役割を定義する。各プロジェクト参加者には、明確に定義された役割と責任領域が割り当てられ、チームワークを促進し、データの不整合のリスクを低減します。
- 品質管理と監査：建設データの効果的な管理には、その品質を継続的に監視する必要がある。データをチェックし、エラーや属性の欠落を特定するための自動化されたメカニズムが導入されている。
- メタデータと情報アーキテクチャの管理：CoE は、システム間の統合に不可欠な、分類 と識別子、命名、実体記述標準の統一システムを構築する責任を負う。

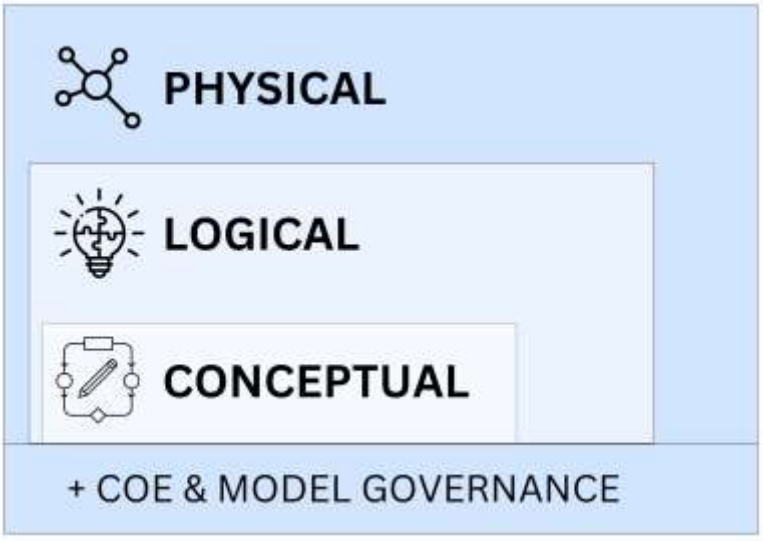


図 4.3-10 データモデリングとデータ品質管理は、CoE の主要課題の一つである

データのためのセンター・オブ・エクセレンス（CoE ）は、単なる専門家集団ではなく、新しいデータ駆動型文化を創造し、全社的なデータ作業への統一的なアプローチを保証する体系的なメカニズムである。CoE は、モデリング・プロセスの情報管理システム全体への適切な統合、標準化、分類、データ品質管理 を通じて、企業が製品やビジネス・プロセスを継続的に改善し、市場の変化に迅速に対応し、信頼性の高いアナリティクスに基づいて十分な情報に基づいた意思決定を行えるよう支援する。

このようなセンターは、近代的な DataOps の原則 - の下で、データの継続的なデリバリー、自動化、品質管理を保証する動き と組み合わせると特に効果的である。DataOps については、第 8 部「建設業界における最新のデータ技術」の章で詳しく説明する。

次の章では、戦略から実践へ、つまり条件付きでデータセンターに「変身」してみよう。タスクのパラメータ化、要件収集 、自動検証プロセスがどのように行われるか、いくつかの例を見てみよう。



章 4.4.

要件の体系化と情報の検証

要件収集と分析： コミュニケーションを構造化データに変換する。

要件の収集と管理は、データ品質を確保するための第一歩である。デジタル・ツールが発達したとはいえ、ほとんどの要件は、手紙、会議の議事録、電話、口頭での議論など、構造化されていない方法で策定されている。このようなコミュニケーション形態では、情報の自動化、検証、再利用が困難である。この章では、テキスト形式の要件を形式的な構造に変換し、ビジネス要件の透明性と体系性を確保する方法について説明する。

> ガートナー社の調査「データ品質：正確な洞察のためのベストプラクティス」は、データとアナリティクスのイニシアチブを成功させるためには、データ品質 が極めて重要であることを強調している[100]。ガートナーは、データ品質が低いと、少なくとも年間平均 1,290 万ドル（約 12 億円）のコストが組織にかかると指摘し、信頼性の高い高品質のデータは、データ主導型企業を構築するために不可欠であると述べている。

構造化された要件がないため、同じ要素（エンティティ）とそのパラメータが、異なるシステムで異なるバリエーションで保存される可能性がある。これはプロセスの効率を低下させるだけでなく、時間の浪費、情報の重複、データを使用する前の再確認の必要性をもたらします。その結果、たった 1 つの漏れ、つまり 1 つのパラメータの紛失や 1 つの誤った記述要素でさえ、意思決定を遅らせ、非効率的なリソースの使用を引き起こす可能性がある。

> 釘がなくて蹄鉄がなくなった。
> 蹄鉄がなくて馬がいなくなった。
> 馬のために、騎手は失われた。
> 騎手がいなかったために、メッセージが失われた。
> メッセージのために、戦いは失われた。
> 戦いのために王国は失われた。
> すべては蹄鉄の釘がなかったために。
>
> - ことわざ[101]

データの充填と保存のプロセスの分析と要件の収集は、すべての利害関係者を特定することから始まります。一本の釘の損失が連鎖的に重大な結果をもたらすということわざがあるように、ビジネスでは、一人の利害関係者の損失、見落とされた要件、あるいはたった一つのパラメーターの損失が、個々のビジネスプロセスだけでなく、プロジェクトのエコシステム全体や組織全体に大きな影響を与える可能性がある。したがって、一見すると取るに足らないような要素、パラメータ、役割（）であっても、後にビジネスの持続可能性にとって重要であることが判明する可能性があるものを特定することが極めて重要である。

ある企業が、クライアントから「建物の北側に窓を追加してほしい」という新たな要求を突きつけられたプロジェクトがあるとしよう。現在のプロジェクトに新しい窓を追加したいというクライアントの要望」という小さなプロセスには、建築家、クライアント、CAD スペシャリスト (BIM)、コンストラクション・マネージャー、ロジスティクス・マネージャー、ERP -アナリスト、品質管理エンジニア、安全エンジニア、コントロール・マネージャー、プロパティ・マネージャーが関与する。

> 小さなプロセスであっても、何十人もの異なるスペシャリストが関与する可能性がある。各プロセス参加者は、データレベルでつながっているスペシャリスト（）の要求を理解しなければならない。

テキストレベル（図 4.4-1）では、クライアントとプロセスチェーンのスペシャリストとのコミュニケーションは以下のように行われる：

- お客様：「採光をよくするために、北側に窓を追加することにしました。実現できますか？"

- 建築家: 「もちろん、新しい窓を含むようにプロジェクトを修正し、更新した CAD 図面 (BIM)を送ります」。
- CAD スペシャリスト (BIM)：「新しいプロジェクトを受注しました。CAD（BIM）モデルを追加窓で更新し、FEM エンジニアと調整した後、新しい窓の正確な位置と寸法を提供します」。
- コンストラクション・マネージャー：「新しいプロジェクトが入りました。4D の設置日を調整し、関連するすべての下請け業者に通知しています」。
- 設備エンジニア（CAFM ）："新しいウィンドウの 6D データを CAFM システムに入力し、将来の設備管理とメンテナンス計画に役立てます。"
- ロジスティクス・マネージャー：「現場への窓の配送を手配するため、新しい窓の寸法と重量が必要です。
- ERP -アナリスト：「プロジェクト全体の見積もりに新しいウィンドウのコストを反映させるため、ERP システムの 5D 予算更新のスコープテーブルと正確なウィンドウタイプが必要です。
- 品質管理エンジニア：「窓の仕様が出来上がったら、それが当社の品質と材料の基準を満たしているか確認します」。
- セーフティ・エンジニア：「8D スキーム（ ）に基づくコンプライアンスと避難に特に重点を置き、新しい窓の安全面を評価します。
- コントロール・マネージャー「ERP からの正確な作業範囲に基づいて、4D タイムライン を更新し、新しいウィンドウの設置を反映させ、プロジェクトのコンテンツ管理システムに新しいデータを保存します。"
- 作業員（インストーラー）：「設置、組み立て、作業のタイミングについての指示が必要です。また、特別に守らなければならない安全規則が導入されていますか？"
- 不動産管理者：「設置後、長期管理のために保証とメンテナンス情報を文書化します。
- アセットマネージャー「設備エンジニア、資産追跡とライフサイクル管理のための最終データを送ってください。
- クライアント：「ちょっと待って、急いでるから窓は必要ないかも。バルコニーにしようかな」。

頻繁に起こるこのようなシナリオでは、小さな変更でさえ、複数のシステムや役割の間で連鎖反応を引き起こす。この場合、初期段階（ ）におけるコミュニケーションは、電子メール、チャット、会議の議事録など、ほとんどすべてテキスト形式となる（図 4.4-1）。

このような建設プロジェクトにおけるテキストベースのコミュニケーション・システムでは、すべてのデータ交換操作と決定事項の法的確認と記録システムが非常に重要である。これは、すべての決定、指示、変更が法

的に有効であり、追跡可能であることを保証し、将来の「誤解」のリスクを減らすためである

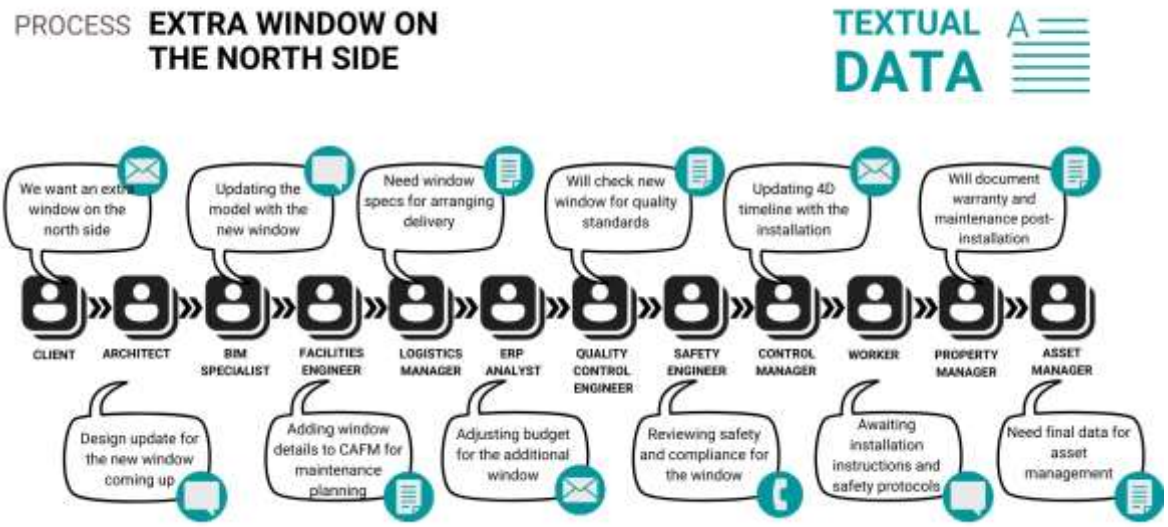


図 4.4-1 プロジェクトの初期段階におけるクライアントとコントラクター間のコミュニケーションには、しばしばマルチフォーマットのテキストデータが含まれる。

> 建設プロジェクトの関連システムにおいて、決定事項の法的管理と妥当性確認が欠けていると、関係者全員に深刻な問題が生じる可能性がある。適切な文書化と検証なしに行われたすべての決定、命令、変更は、紛争（および訴訟）につながる可能性がある。

文字によるコミュニケーションにおけるすべての決定事項の法的な統合は、多数の署名された文書によってのみ保証され、すべての取引を記録する義務のある経営陣の肩にかかることになる。その結果、すべての参加者がすべての行動に対して文書に署名する必要がある場合、システムは柔軟性を失い、官僚的な迷路となる。取引確認の欠如は、プロジェクト実施を遅らせるだけでなく、金銭的損失や参加者間の関係悪化、さらには法的問題にまで発展する可能性がある。

このような取引承認プロセスは、通常、テキストベースの議論から始まり、次第に次の段階でマルチフォーマットの文書交換へと発展し（図 4.4-2）、テキストだけで行われていたコミュニケーションを著しく複雑にしている。要件が明確に定義されていなければ、マルチフォーマットデータと大量のテキスト要件で埋め尽くされたこのようなプロセスを自動化することは、ほとんど不可能となる。

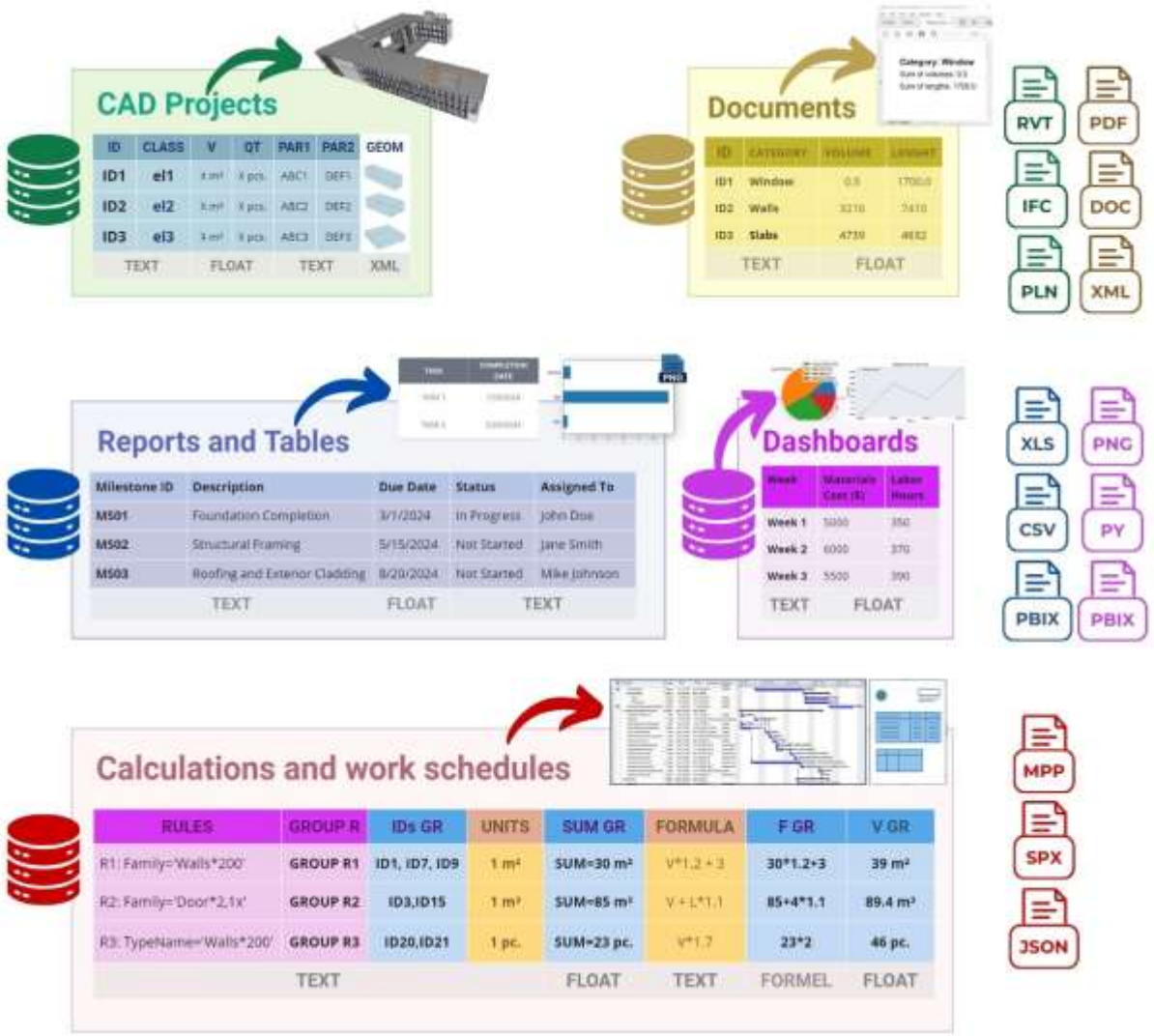


図 4.4-2 建設会社の各システムは、さまざまな形式の法的関連文書の情報源となる。

テキスト・コミュニケーションでは、各専門家が全対応文書に精通するか、定期的にすべての会議に出席し、プロジェクトの現状を理解する必要がある。

この限界を克服するためには、テキストによるコミュニケーションから構造化された要求モデルへの移行が必要である。これは、体系的な分析、プロセスの視覚化、およびフローチャートとデータモデル（図 4.4-3）の形式による相互作用の記述によってのみ可能である。データモデリング（図 4.3-7）と同様に、参加者が使用するシステムとツール、およびそれらの間のリンクを追加することで、文脈的アイデアレベルから概念的レベルへと移行した。

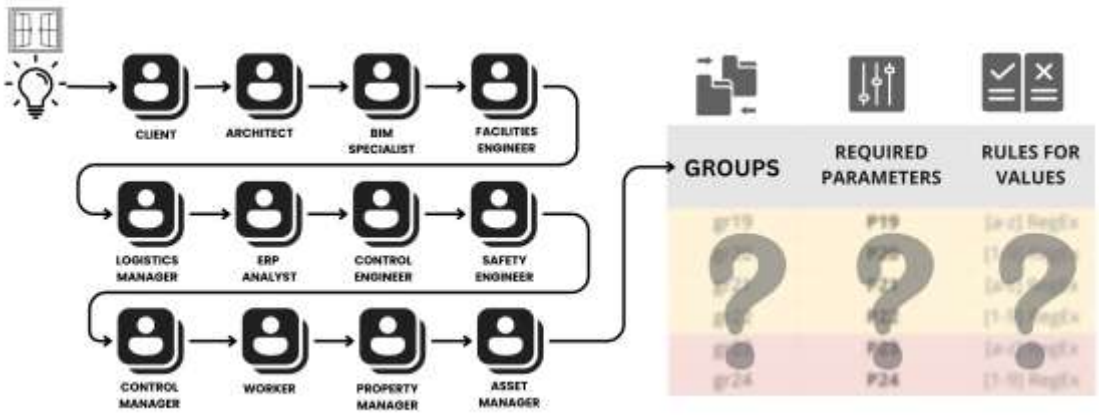


図 4.4-3 検証プロセスを管理し、自動化する方法を学ぶには、プロセスを可視化し、要件を構造化する必要がある。

要件と関係を体系化する最初のステップは、概念フローチャートを使用してすべてのリンクと関係を視覚化することである。概念レベル は、すべてのプロセス参加者がプロセスチェーン全体を理解しやすくするだけでなく、各プロセスステップでデータ（および要件）がなぜ、誰のために必要なのかを明確に示す。

プロセス・フローチャートと概念フレームワークの有効性

データ管理に対する従来のアプローチと最新のアプローチのギャップを埋めるために、企業は断片的なテキスト記述から構造化されたプロセス表現へと意識的にシフトする必要がある。粘土板からデジタル・エコシステムへのデータの進化には、新しい思考ツールが必要だ。そして、そのようなツールの 1 つが、フローチャートを使った概念モデリングである。フローチャート、プロセス図、相互作用図といった視覚的な図を作成することで、プロジェクト参加者は、自分の行動や決定が意思決定システム全体にどのような影響を与えるかを実感することができる。

> データの保存だけでなく、分析や自動化が必要なプロセスでは、概念的で視覚的な要件レイヤーの作成に着手する必要がある。

この例（図 4.4-1）では、各スペシャリストは小規模なチームの一員であることもあれば、ゼネラル・マネジャーの管理下にある、最大十数名のエキスパートを擁する大規模な部門の一員であることもあります。各部門は、専用のアプリケーション・データベース（図 1.2-4 例：ERP 、CAD 、MEP 、CDE 、ECM 、CPM など）を使用し、文書の作成、決定事項の法的ステータスの記録、プロセスの管理に必要な情報が定期的に更新される。

この取引プロセスは、粘土板やパピルスを使って合法的に意思決定を確認していた 4,000 年前の古代の管理職の仕事と似ている。現代のシステムと粘土や紙で作られた以前のシステムとの違いは、現代の方法には、さらに他のシステムやツールで自動処理するために、テキスト情報をデジタル形式に変換するプロセスが含まれていることである。

概念的なフローチャートの形でプロセスを視覚化することで、各ステップや異なる役割間の相互作用を説明しやすくなり、複雑なワークフローを明確かつシンプルにすることができる。

プロセスを可視化することで、プロセスのロジックが透明化され、チームメンバー全員がアクセスできるようになる。

テキスト、メッセージ（図 4.4-1）、ブロック図で説明した、プロジェクトにウィンドウを追加するためのコミュニケーショ ン・プロセスは、データ・モデリングの章で説明した概念モデル（図 4.4-4）と同様である。

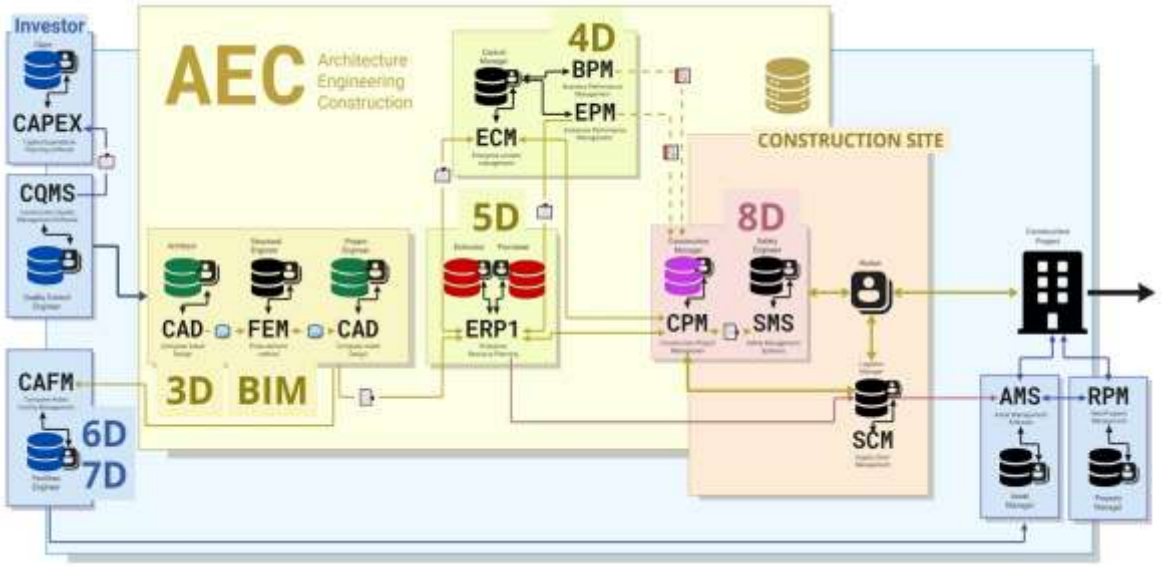


図 4.4-4 この概念図では、プロジェクト参加者がデータベースのユーザーであり、そのクエリが異なるシス テムをリンクしている。

概念図は重要なステップであるが、多くの企業は、プロセスを理解するには視覚的な図で十分だと考え、このレベルにとどまっている。管理者は、このようなフローチャートの方が全体像を把握しやすく、参加者やステージ間のつながりもわかりやすい。しかし、このような図式では、各参加者にどのようなデータが必要なのか、どのような形式で送信されるべきなのか、自動化を実現するためにどのようなパラメータや属性が必須なの

か、といった明確なアイデアは得られない。概念的なフローチャートは、どちらかというとルートマップのようなもので、誰が誰と相互作用しているのかを示しているが、これらの相互作用において何が転送されているのかは明らかにしていない。

> フローチャートを使ってプロセスを概念レベルで詳細に記述しても、その有効性が保証されるわけではない。可視化 、管理者の作業が簡素化され、ステップバイステップのレポートによってプロセスをより簡単に追跡できるようになることが多い。しかし、データベースエンジニアにとっては、概念的な表現が十分明確でなく、パラメータや要件レベルでのプロセスの実装方法を明確に理解できない可能性がある。

より複雑なデータエコシステムに移行するにつれ、データプロセスが効率的であるだけでなく、組織の戦 略的目標に沿ったものであることを保証するために、概念的・視覚的ツールの初期導入が重要になる。このウィンドウの追加プロセス（図 4.4-1）をデータ要件レベルに完全に変換するためには、一段深く踏み込んで、プロセスの概念的な視覚化をデータ、必要な属性、およびそれらの境界値の論理的・物理的なレベルに変換する必要がある。

構造化要件 および RegEx 正規表現

企業で作成されるデータの 80%は、テキスト、文書、手紙、PDF ファイル、会話など、非構造化または半構造化形式 [52] である。このようなデータ（図 4.4-1）は、分析、検証、システム間の転送、自動化での利用が困難である。

管理性、透明性、自動検証性を確保するためには、テキスト形式や半構造化形式の要件（ ）を、明確に定義された構造化形式に変換する必要がある。この構造化プロセスは、データ（本書の最初の章で詳しく説明した）だけでなく、プロジェクト参加者がプロジェクトのライフサイクル全体を通じて自由なテキスト形式で作成する要件自体にも関係する。

> 構造化されていないテキスト形式のデータを構造化された形式に変換したように、要件ワークフローでは、テキスト形式の要件（）を構造化された「論理層および物理層」形式に変換する。

ウィンドウを追加する例（図 4.4-1）の一部として、次のステップではデータ要件を表形式で記述する。プロジェクト参加者が使用する各システムの情報を、主要な属性とその境界値を指定することで構成する

例えば、このようなシステムの一つ（図 4.4-5）- 施主側の品質管理技術者が使用する建設品質管理システム（CQMS ）を考えてみよう。このシステムの助けを借りて、彼はプロジェクトの新しい要素、この場合は「新しい窓」 が、確立された基準と要件に適合しているかどうかをチェックする。

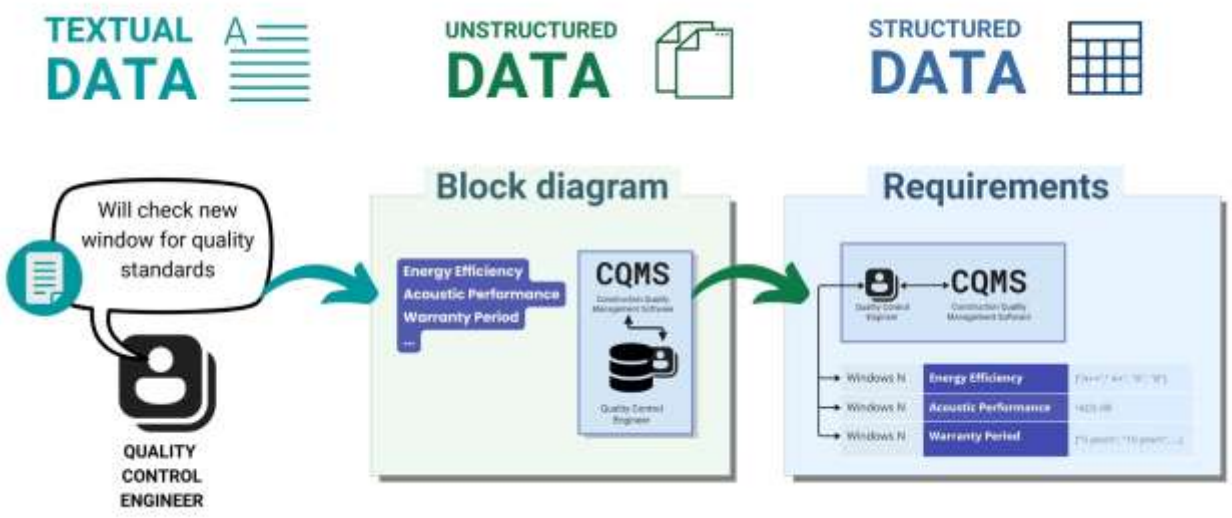


図 4.4-5 テキスト形式の要件を、エンティティの属性を記述した表形式に変換することで、他の専門家の理解が容易になります。

一例として、CQMS -system （図 4.4-6）の「窓システム」タイプのエンティティの属性に関する重要な要 件（エネルギー効率、音響性能、保証期間）を考えてみる。各カテゴリーには、窓システムを設計・施工する際に考慮する必要のある一定の基準や仕様が含まれる。



図 4.4-6 品質管理エンジニアは、新しいウィンドウ・タイプのエレメントについて、エネルギー効率、遮音性、 および保証基準を検査する。

品質管理エンジニアが表の形で指定するデータ要件には、例えば次のような境界値がある：

- 窓のエネルギー効率クラスは、最高効率を示す「A++」から、最低許容レベルとされる「B」まであり、これらのクラスは許容値のリスト［「A++」、「A+」、「A」、「A」、「B」］で表される。
- 窓の遮音性はデシベルで測定され、街頭騒音を低減する能力を示す。
- Window Type "エンティティの " Warranty Period " 属性は 5 年から始まり、この期間を製品選択時に許容される最小値として設定する。" 5 years"、" 10 years " などの保証期間値や、論理条件">5 (years) " も指定できる。

収集された要件によれば、確立された属性の範囲内において、「C」や「D」のような「B」を下回る等級を有する新しい窓カテゴリーまたはクラス要素は、エネルギー効率試験に合格しない。QA エンジニアに提出するデータまたは書類に記載される窓の遮音性能は、「35dB」や「40dB」のように、2 桁の数字の後に接頭辞「dB」を付けて表示するものとし、「9 D B」や「100 デシベル」のようなこの形式以外の値は認められない（RegEx 文字列のパターンを通過しないため）。保証期間は最低「5 年」から始まり、「3 年」や「4 年」のような短い保証期間のウィンドウは、品質エンジニアが表形式に記述した要件を満たしません。

このような属性パラメータ値を、バリデーションプロセスで要求事項の境界値と照合するために、許容値のリスト（["A", "B", "C"]）、辞書（["A": "H1", "H2"; "B" ：W1", "W2"]）、論理演算（例えば、数値の場合">", "<", "<=", ">=" "=="））、正規表現（"Acoustic Performance " 属性のような文字列やテキスト値の場合）。正規表現は、文字列値を扱うときに非常に重要なツールです。

> 正規表現 (RegEx) は、Python (Re ライブラリ) を含むプログラミング言語で、文字列の検索や修正に使用されます。Regex は文字列の世界の探偵のようなもので、テキストのパターンを正確に特定することができます。

正規表現では、文字は対応するアルファベット文字を使って直接記述し、数値は 0 から 9 までの任意の桁に対応する特殊文字␣d を使って表現することができます。例えば、[a-z]はラテンアルファベットの小文字、[0-9]は␣d と同じです。数字以外の文字や文字以外の文字には、それぞれ[D]と[W]を使用します。

よく使われる RegEx の使用例 （図 4.4-7）：

- 電子メールアドレスの検証：文字列が有効な電子メールアドレスかどうかを確認するには、"^ [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+.[a-zA-Z]{2,}$" テンプレートを使用することができます。
- Date Extraction: " \bdd{2}d{2}d{2}d{2}d{2}d{2}d{2}d{2}d{2}d{2}. \d{4}b" テンプレートを使って、DD.MM.YYYYYY 形式でテキストから日付を抽出できます。
- 電話番号の検証: +49(000)000-0000 形式の電話番号を検証する場合、パターンは" \+d{2}／(／d{3}／)／-d{4}" のようになります。

QA エンジニアの要求事項 を、属性とその境界値のフォーマット（図 4.4-6）に変換することにより、元のテキスト形式（会話、手紙、規制文書）から、整理され構造化された表に変換し、あらゆる入力データ（例えば、Window カテゴリの新しい 要素）を自動的にチェックし分析することを可能にした。要件が存在することで、チェックされていないデータは自動的に破棄され、チェックされたデータはさらなる処理のために自動的にシステムに転送される。

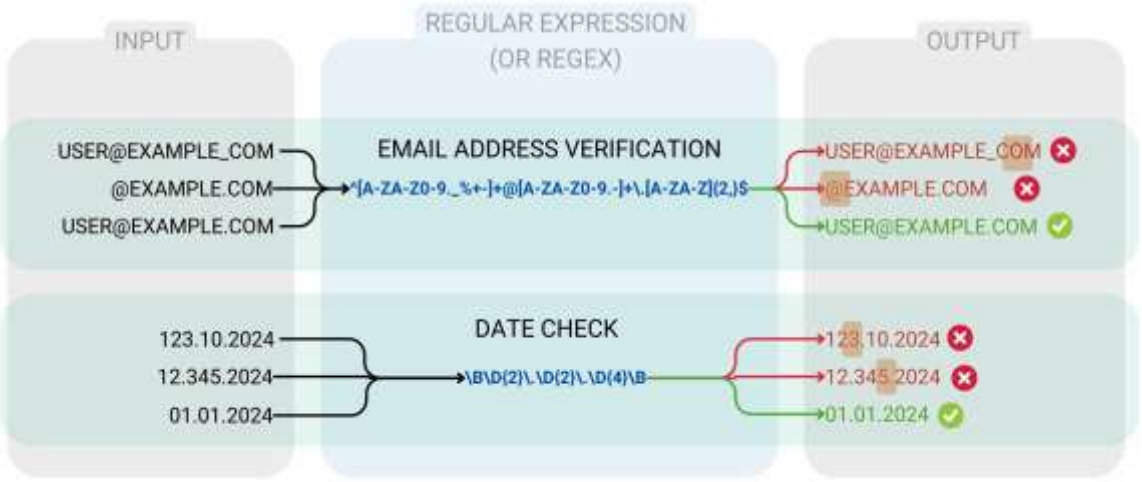


図 4.4-7 正規表現 の使用は、テキスト・データの検証プロセスにおいて極めて重要なツールである。

さて、要求を扱う概念的なレベルから論理的なレベルへと移行して、新しいウィンドウをインストー ルするプロセス（図 4.4-4）において、すべてのスペシャリストのすべての要求 を属性形式の順序付きリストに変換し、必要な属性を持つこれらのリストを各スペシャリストのフローチャートに追加します（図 4.4-8）。

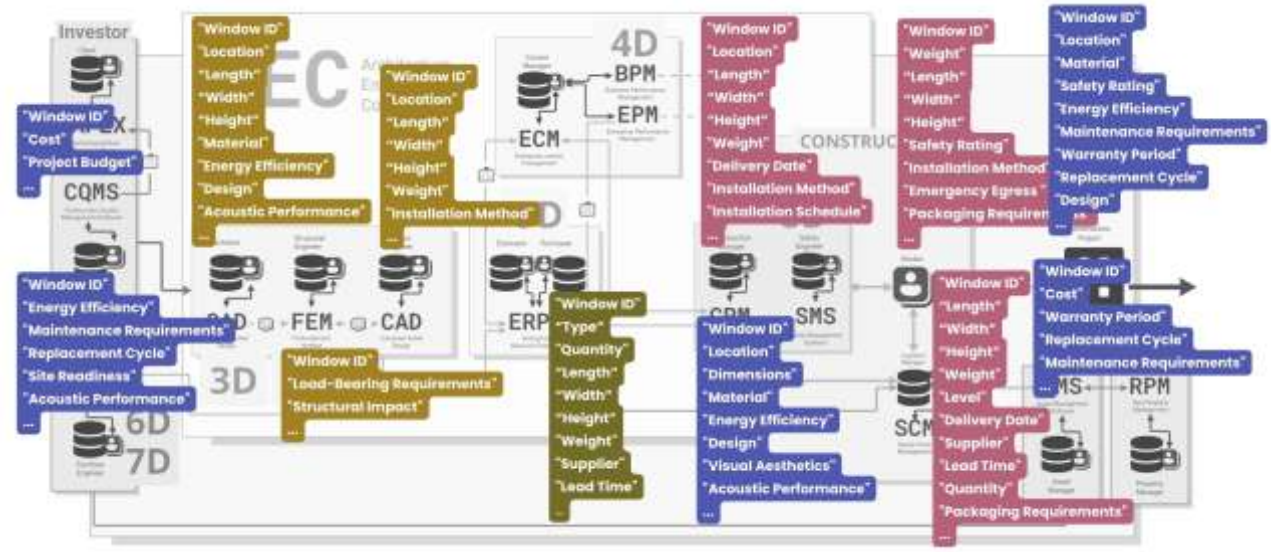


図 4.4-8 論理プロセスレベルでは、各スペシャリストが扱う属性がそれぞれのシステムに追加される。

すべての属性を 1 つの共通プロセス表に追加することで、これまで概念レベル（図 4.4-1）のテキストやダイアログとして提示されていた情報を、物理レベル表（図 4.4-9）の構造化された体系的な形に変換する。

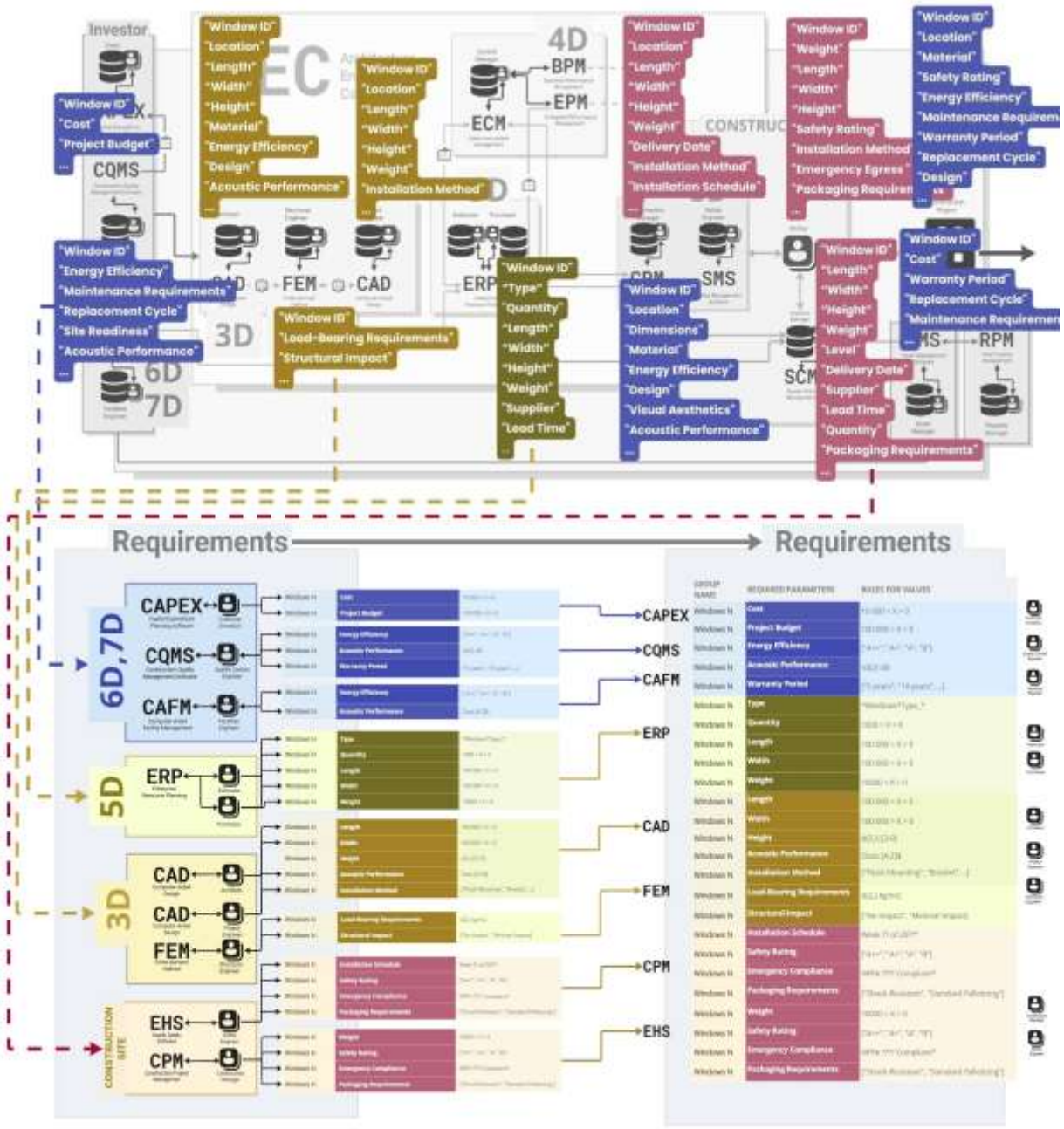


図 4.4-9 構造化されていない専門家の対話を構造化された表に変換することは、物理的なレベルで 要件を理解するのに役立つ。

ここで、データ要件を特定のシステム用の情報を作成する専門家に伝える必要がある。例えば、CAD データベースで作業する場合、要素のモデリングを開始する前に、データの最終使用シナリオに基づいて必要なパラメータをすべて収集する必要があります。これは通常、運用段階から始まり、建設現場、物流部門、見積部門、構造計算部門などが続きます。これらすべての分野の要件を考慮に入れて初めて、収集したパラメーターに基

づくデータの作成が可能になります。これにより、将来的にチェーンに沿ったデータの検証と転送を自動化することができる。

新しいデータが要件を満たすと、そのデータは自動的に企業のデータエコシステムに統合され、意図されたユーザーやシステムに直接提供されます。属性とその値に照らしてデータを検証することで、情報が要求される品質基準を満たし、企業のシナリオに適用する準備が整っていることを確認する。

データ要件が定義され、検証を開始する前に、検証対象のデータを作成、入手、または収集しなければならない。あるいは、検証プロセスに使用するために、データベース内の情報の現状を記録しなければならない。

データ収集 検証プロセス用

検証を開始する前に、検証プロセスに適した形でデータが利用可能であることを確認することが重要である。データを収集し、非構造化、疎構造化、テキスト形式、幾何学的形式から構造化された形式に変換しなければならない。このプロセスについては、前の章で詳しく説明し、さまざまなタイプのデータを変換する方法について述べた。すべての変換の結果、入力データはオープンな構造化テーブルの形をとる（図 4.1-2、図 4.1-9、図 4.1-13）。

要件と、必要なパラメータと境界値を含む構造化された表（図 4.4-9）があれば、データの検証を開始することができます。これは、単一の自動化されたプロセス（パイプライン ）として、または各入力文書のステップバイステップの検証として行うことができます。

チェックを開始するには、入力として新しいファイルを受信するか、データの現在の状態を修正する、つまりスナップショットを作成するか、現在のデータと受信データをエクスポートするか、外部または内部のデータベースへの接続を設定する必要があります。検討中の例では、このようなスナップショットは、 、すべての設計者が帰宅した後、たとえば 2024 年 3 月 29 日（金）23:00:00 に記録された CAD データを構造化フォーマットに自動的に変換することによって作成されます。



図 4.4-10 CAD データベースのスナップショット （BIM ）は、プロジェクトモデルの現在のバージョンにあるクラス「窓」の新 しいエンティティの現在の属性情報を示している。

CAD データ (BIM)の構造化形式への変換」の章で説明したリバースエンジニアリングツールのおかげで、異なる CAD（BIM）ツールやエディタからのこれらの情報は、別々のテーブルに整理したり（図 4.4-11）、プロジェクトの異なるセクションを組み合わせた 1 つの共通のテーブルにまとめることができます（図 9.1-10）。

このようなテーブル - データベースには、窓やドアの一意の識別子（ID 属性）、タイプ名（TypeName）、寸法（幅、長さ）、材料（Material）、エネルギーや音響性能の指標やその他の特性が表示されます。CAD プログラム （BIM ）に記入されたこのようなテーブルは、設計エンジニアによって様々な部門や文書から収集され、プロジェクトの情報モデルを形成します。



図 4.4-11 CAD システムからの構造化データ は、要素の属性を示す列を持つ二次元表とすることができる。

実際の CAD （BIM ）プロジェクトには、数万から数十万の要素が含まれる（図 9.1-10）。CAD フォーマット内

の要素は、窓やドアからスラブ、床版、壁まで、タイプやカテゴリーごとに自動的に分類されます。一意の識別子（CAD ソリューションによって自動的に設定されるネイティブ ID など）やタイプ属性（タイプ名、タイプ、ファミリー）によって、同じオブジェクトを異なるシステムで追跡することができます。例えば、建物の北側の壁にある新しい窓は、組織内のすべての関連システムにおいて、単一の識別子「W-NEW」によって一意に識別することができる。

> エンティティの名前と識別子は、すべてのシステムで一貫していなければならないが、これらのエンティティに関連する属性と値のセットは、使用するコンテキストによってかなり異なる場合がある。建築家、構造エンジニア、建設、物流、不動産メンテナンスの専門家は皆、同じ要素を異なる方法で認識している。純粋に美的観点から窓の形状やプロポーションを評価する人もいれば、エンジニアリングや運用の観点から、熱伝導率、設置方法、重量、メンテナンス要件などを分析する人もいます。したがって、データをモデリングして要素を記述する際には、その用途の多様性を考慮し、業界の特殊性を考慮しながらデータの一貫性を確保することが重要です。

会社の各プロセスにおける役割ごとに、設計から計算、ロジスティクス、据付、建物運営に至るまで、独自のユーザー・インタフェースを備えた専門データベースが存在する（図 4.4-12）。このような各システムは、専門家チームによって、特別なユーザー・インターフェイスを通じて、あるいはデータベース・クエリを通じて管理される。このようなシステム・チェーンの末端において、入力された値に基づいて行われるすべての決定の背後にあるのは、システム・マネージャーあるいは部門マネージャーであり、このマネージャーは、他のシステムにサービスを提供している取引先に対して、入力されたデータの法的妥当性と品質に責任を負っている。

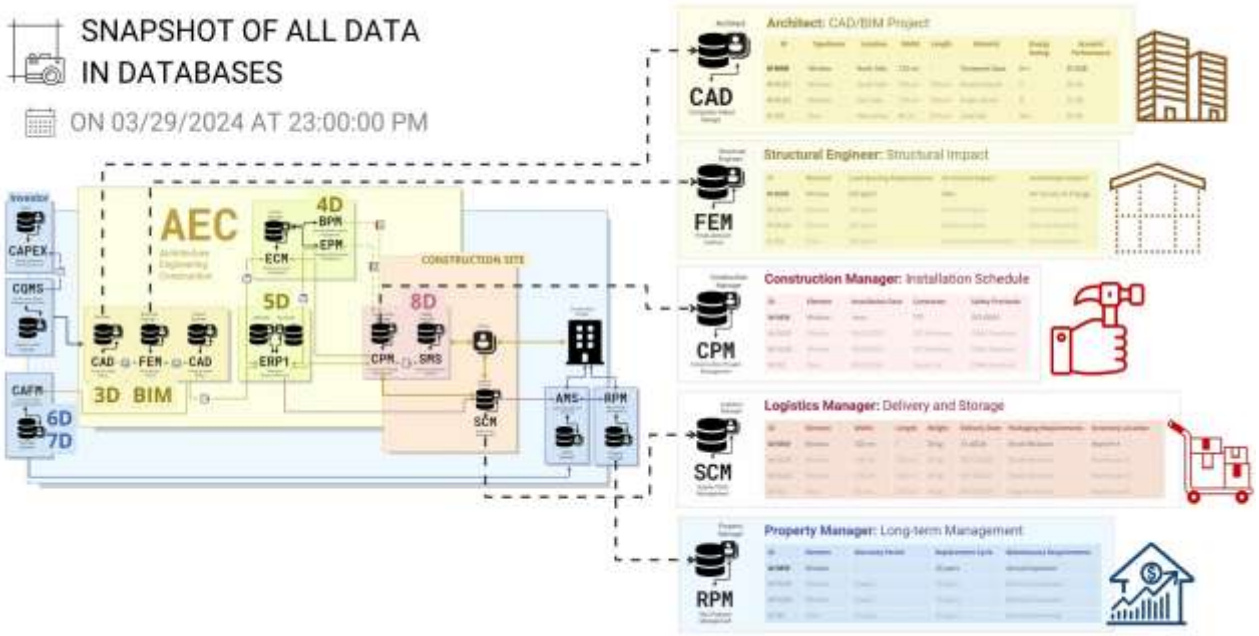


図 4.4-12 同じエンティティが異なるシステムで同じ識別子を持つが、そのシステムでのみ重要な異なる属性を持つ。

論理的・物理的なレベルで構造化された要件とデータの収集を整理したら、あとは、さまざまな受信文書やさまざまなシステムからのデータを、事前に収集した要件と自動的に照合するプロセスを設定するだけである。

データの検証と検証結果

発注者、建築家、エンジニア、現場監督、ロジスティシャン、プロパティマネージャーから の文書、テーブル、データベースエントリなど、システムに入力されるすべての新しいデー タは、先に策定した要件に照らして検証されなければならない（図 4.4-9）。データに誤りがあれば、誤った計算、スケジュールの遅延、さらには経済的損失につながる可能性がある。このようなリスクを最小化するために、体系的で反復可能なデータ検証手順を組織する必要がある。

> 非構造化データ、テキストデータ、幾何学的データなど、システムに入力される新しいデータを検証するには、そのデータを疎構造化または構造化フォーマットに変換しなければならない。検証プロセスでは、必要な属性とその許容値の完全なリストとデータを照合しなければならない。

テキスト、画像、PDF -文書、および CAD (BIM) の混合データなど、さまざまなタイプのデータを構造化フォー

ムに変換することについては、「データを構造化フォームに変換する」の章で詳しく説明しました。

例として、CAD （BIM ）プロジェクトから取得したテーブルがある（図 4.4-11）。このテーブルには、半構造化ジオメトリ・データと、プロジェクト・エンティティの構造化属性情報（図 3.1-14）が含まれています。

検証を行うために、属性値（図 4.4-11）を、専門家が要求事項の形で定義した参照境界値（図 4.4-9）と比較します。最終的な比較表（図 4.4-13）により、CAD アプリケーション（BIM ）以外でデータを使用する前に、どの値が許容可能で、どの値を修正する必要があるかが明確になります。

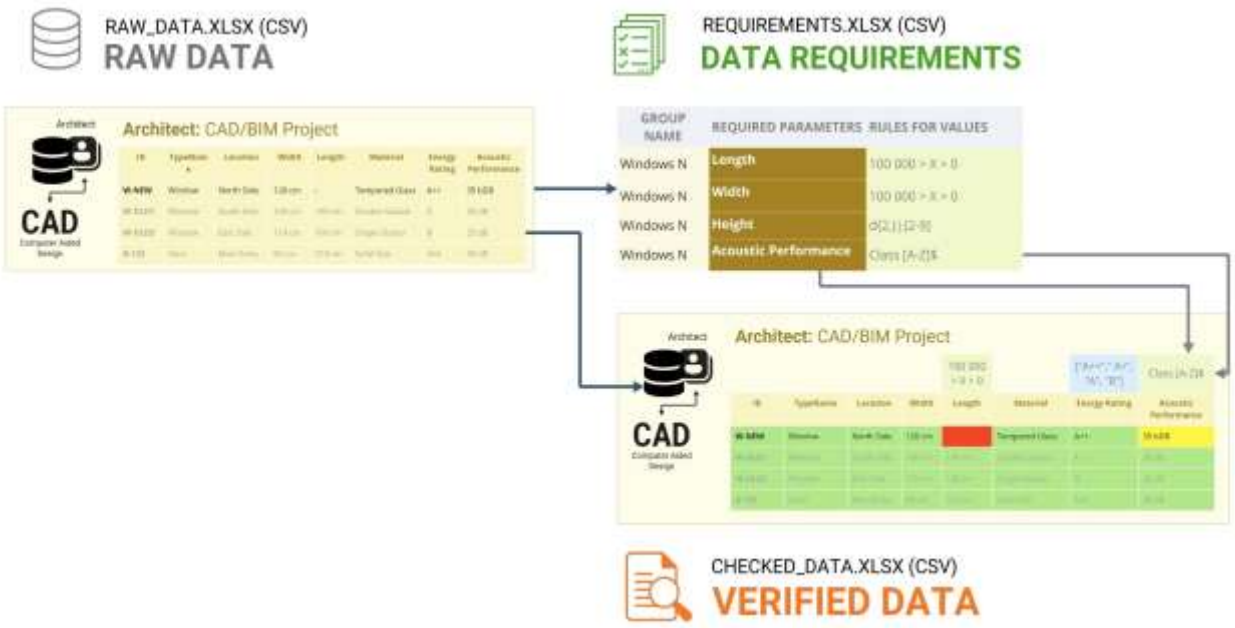


図 4.4-13 最後の検証表は、クラス "Windows "の新しいエンティティの属性値のうち、注意すべきものを強調表示します。

Pandas: An indispensable tool for data analysis （パンダス：データ分析に不可欠なツール）」 の章で前述した Pandas ライブラリを使用して同様のソリューションを実装すると、CAD ファイル（BIM ）（RVT 、IFC 、DWG 、NWS、DGN ）（図 4.4-11）から抽出した表形式ファイルのデータを、別の表形式要件ファイル（図 4.4-9）の要件を使用して検証します。

コードを取得するには、LLM のプロンプトに、raw_data.xlsx ファイル（CAD データベース （BIM ）のデーター式）から 、データを読み込んでチェックし、その結果を新しいファイル checked_data.xlsx に保存する必要があることを記述する必要がある（図 4.4-13）。

- Pandas ライブラリには触れずに、LLM を使ったコードを取得してみよう：

> raw_data.xlsx ファイルから表をチェックし、以下の検証ルールを使用して検証するコードを記述する：「幅」と「長さ」列の値がゼロより大きいこと、「エネルギー等級」が['A++', 'A+', 'A', 'B']リストに含まれていること、「音響性能」が後で指定する変数として含まれていること - 最後の検証列を追加し、最終表を新しい Excel ファイル checked_data.xlsx に保存する。 ⏎

LLM の応答には、Python コードの短い例が記述され、その後のプロンプト ：



図 4.4-14 LLM- モデルによって生成されたコードは、変換された CAD (BIM)の設計を、境界値の形で属性要件と照合します。

LLM 言語モデルによって生成されたコード は、一般的な IDE やオンラインツールで使用できます： PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse with PyDev plugin, Thonny, Wing IDE, IntelliJ IDEA with Python plugin, JupyterLab または一般的なオンラインツール Kaggle.com、Google Collab、Microsoft Azure Notebooks、Amazon SageMaker。

コードを実行すると（図 4.4-14）、CAD データベース （BIM ）の「実体要素」W-OLD1、W-OLD2、D-122（お

よびその他の 要素）が、属性要件である幅と長さがゼロより大きく、エネルギー効率クラスがリスト値「A++」、「A」、「B」、「C」のいずれかであることを示す（図 4.4-15）。

私たちが必要とし、最近追加した W-NEW 要素は、北側の新しい要素クラス'Window'を担っていますが、その長さがゼロ（'0.0'という値は、私たちの'Width'>0 ルールでは受け入れられないとみなされます）であり、エネルギー効率クラスを指定していないため、非準拠です（属性'Requrments Met'）。



	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

図 4.4-15 Checking 検証プロセスを経ていないエンティティを識別し、値に「False」または「True」を持つ新しい属性を結果に追加する。

同様に、プロジェクトにウィンドウを追加する過程で、異なるスペシャリストから受け取ったすべてのデータ（図 4.4-1）において、すべてのプロジェクト要素（エンティティ）と、各システム、テーブル、またはデータベースに必要な属性の整合性をチェックする。

最終表では、視覚化のためにチェック結果を色で強調するのが便利である。緑色はチェックに合格した属性、黄色はクリティカルでない乖離のある値、赤色はクリティカルな乖離を示すために使用される（図 4.4-16）。

検証の結果 (図 4.4-16) 、属性要件を満たすことが検証された、信頼され検証された要素のリストとその識別子が得られます。検証された要素は、Window クラスまたはその他のクラスの要素を追加するプロセスに関与するすべてのシステムに対して、これらの要素が指定された標準と仕様を満たしていることを保証します（データ検証の自動化と自動化された ETL プロセス の作成については、「ETL とデータ検証の自動化」の章で詳しく説明します）。

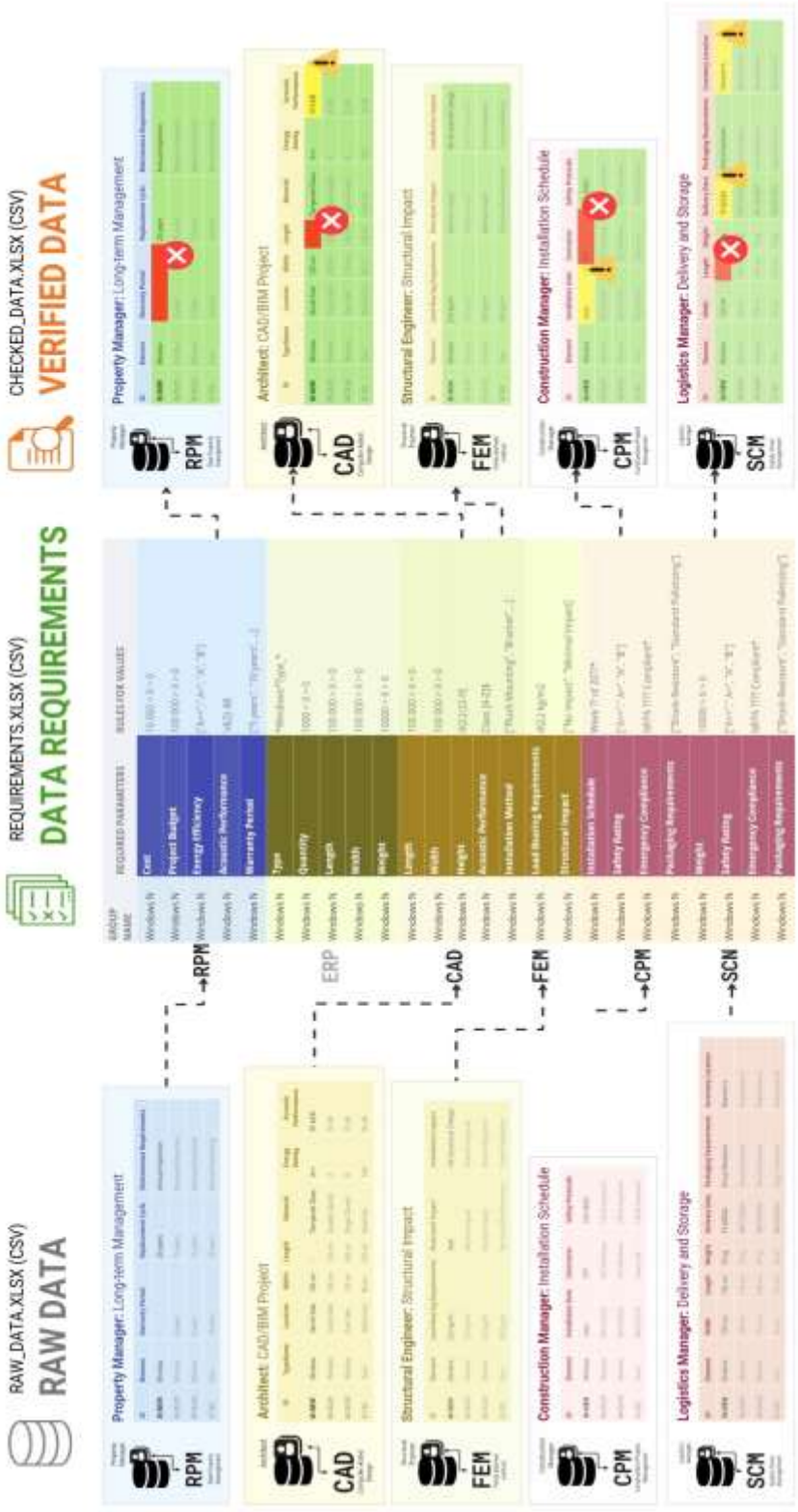


図 4.4-16 すべてのシステムに対して実施されたチェックの結果、どのデータが会社の要件を満たしていないかが判断できる。

バリデーションに成功したエンティティは、通常、それほど注意を払う必要はない。それらは、障害なく次の処理段階や他のシステムへの統合へと進む。品質」項目とは対照的に、最も関心があるのはバリデーションに失敗した項目である。このような逸脱に関する情報は非常に重要であり、表形式の報告書だけでなく、さまざまな視覚化ツールを用いて伝達されるべきである。検証結果をグラフィカルに表示することで、データ品質の全体的な状態を迅速に評価し、、問題箇所を特定し、速やかに是正措置や改善措置を講じることができる。

検証結果の視覚化

視覚化 は、検査結果を解釈するために不可欠なツールです。通常のサマリー表に加え、情報パネル、ダイアグラム、およびプロジェクト要素を検査ステータスごとにグループ化した自動生成 PDF ドキュメントを含めることができます。色分けは、ここで補助的な役割を果たすことができる。緑は検証に成功した項目、黄色は追加的な注意が必要な項目、赤は重大なエラーや重要なデータが欠落している項目を示すことができる。

私たちの例（図 4.4-1）では、CAD （BIM ）や物件管理から、ロジスティクスや設置スケジュールに至るまで、各システムのデータを段階的に分析している（図 4.4-16）。監査後、各専門家に対して、個別の警告または報告文書が、PDF 形式自動生成される（図 4.4-17）。データが正しければ，専門医は "Thank you for working together " という短いメッセージを受け取る。不一致が検出された場合、"This document lists elements, their identifiers, attributes and values that have not been checked for compliance " という文言の詳細なレポートが送られる。



図 4.4-17 バリデーションと自動レポート生成は、データを作成する専門家にとって、データの欠陥を発見し理解する プロセスをスピードアップする。

自動化された検証プロセスのおかげで、エラーやデータギャップが検出されるとすぐに、関連するエンティ ティとその属性を作成または処理する担当者に、チャットメッセージ、電子メール、または PDF -文書（図 4.4-18）の形で、検証されていない要素と属性の記述のリストとともに、即座に通知が送信されます。



図 4.4-18 自動検査レポートにより、エラーを把握しやすくなり、プロジェクトデータの完成作業がスピードアップする。

例えば、（構造化後に）「保証期間」属性が誤って記入された文書が不動産管理システムに到着した場合、不動産管理者は、チェックして修正する必要がある属性のリストを含むアラートを受け取る。

同様に、設置スケジュールやロジスティクスデータに不備があれば、自動レポートが作成され、例えばチャット通知や検査結果を記載したメールが関連する専門家に送信される。

結果を示す PDF -文書やグラフに加え、ダッシュボードやインタラクティブな 3D -モデル（図 7.1-6、図 7.2-12）を作成し、属性が欠落している要素を強調表示することができます。

> 検査結果を自動生成された文書、グラフ、またはダッシュボードの形で視覚化することで、データの解釈を大幅に簡素化し、プロジェクト参加者間の効果的なコミュニケーションを促進します。

様々なシステムや情報ソースからデータを自動的にチェックするプロセスは、日常生活における情報に基づいた意思決定に例えることができる。建設業界の企業が、入力データの信頼性から、プロジェクト実現のタイミング、コスト、品質への影響まで、多くの変数を考慮するように、重要な決定を下すとき、例えば、住む場所を選ぶとき、人は、交通の便、インフラ、コスト、安全性、生活の質など、あらゆる要素を総合的に判断する。これらの考慮事項はすべて、私たちの生活を構成する最終的な決断の基礎となる基準体系を形成している。

データ品質チェックの比較 人間の生活ニーズとの比較

データ品質管理手法やツールの絶え間ない発展にもかかわらず、情報コンプライアンスの基本原則は変わらない。この原則は、ビジネスであれ日常生活であれ、成熟した管理システムの基盤に組み込まれている。

> 反復的なデータ検証のプロセスは、誰もが日々直面する意思決定のプロセスによく似ている。どちらの場合も、私たちは経験、データ、そして入手可能になった新しい情報に頼っている。そして、戦略的なものから日常的な意思決定に至るまで、人生や仕事上の意思決定はますますデータに基づいて行われるようになっている。

例えば、住む場所や人生の伴侶を選ぶとき、私たちは直感的に、選択肢を比 較するための基準や特性の表を頭の中で作成する（図 4.4-19）。これらの特性は、人の個人的な資質であれ、物件のパラメータであれ、最終的な決定に影響を与える属性を表している。

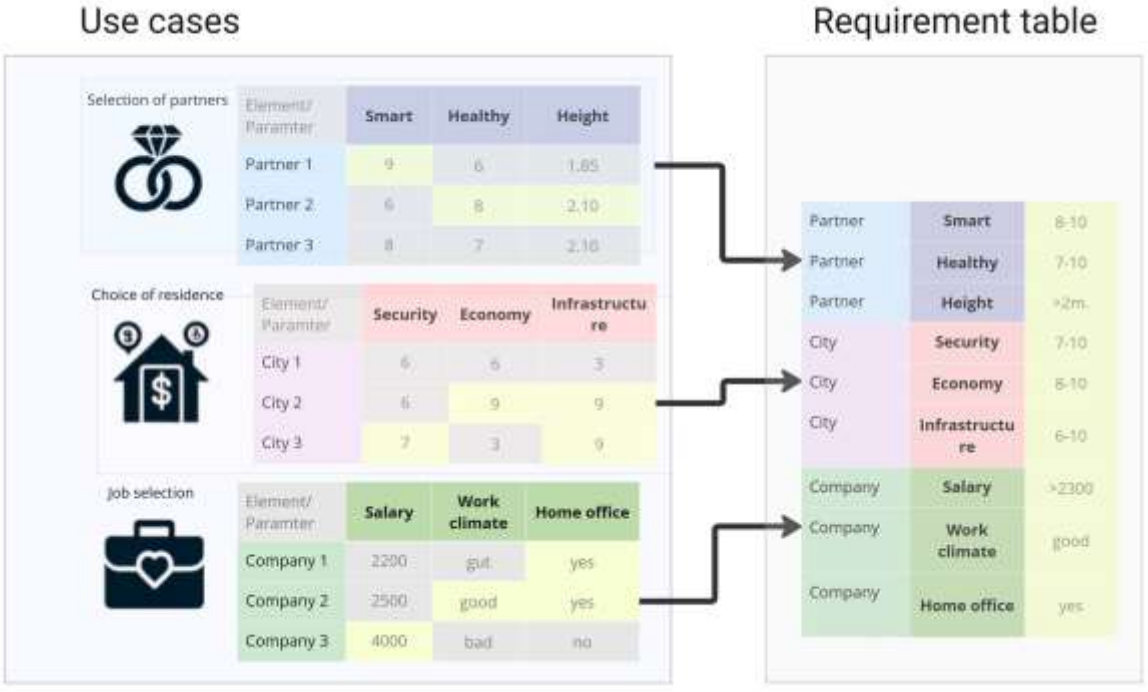


図 4.4-19 居住、雇用、パートナーシップの選択は、個人の属性要件に基づいている。

構造化されたデータの使用と、要求事項を記述するための形式化されたアプローチ（図 4.4-20）は、仕事と私生活の両方において、より多くの情報に基づいた、情報に基づいた選択に貢献する。

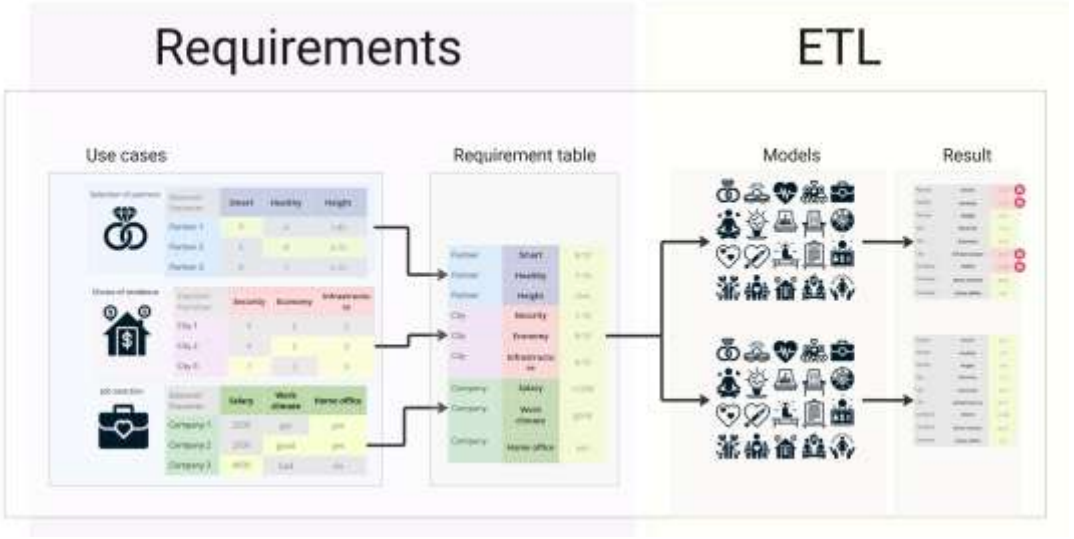


図 4.4-20 要件を形式化することで、生活やビジネス上の意思決定の認識を体系化することができる。

データ駆動型意思決定アプローチは、ビジネスツールに限定されるものではない。日常生活にもシームレスに統合され、ETL プロセス (Extract, Transform, Load)に似た一般的なデータ処理ステップ(図 4.4-21)に従って、データを構造化する：

- 基礎としてのデータ（抜粋）：仕事であれ私生活であれ、どんな分野でも私たちは情報を集める。ビジネスであれば、報告書、数字、市場データ、私生活であれば、個人的な経験、恋人からのアドバイス、フィードバック、観察などである。
- 評価基準（Transform）：収集された情報は、あらかじめ定義された基準に基づいて解釈される。仕事では、業績評価指標（KPI）、予算制約、規範などであり、個人生活では、価格、利便性、信頼性、カリスマ性などのパラメータである。
- 予測とリスク分析（Load）：最終段階では、変換されたデータを分析し、起こりうる結果を比較して意思決定を行う。これはビジネスプロセスに似ており、データはビジネスロジックとリスクフィルターを通過する。

朝食に何を食べるかというような些細な嗜好から、キャリアや人生のパートナーを選ぶというような人生の一大イベントまで、私たちが下す決断は、本質的にデータを処理し評価した結果である。

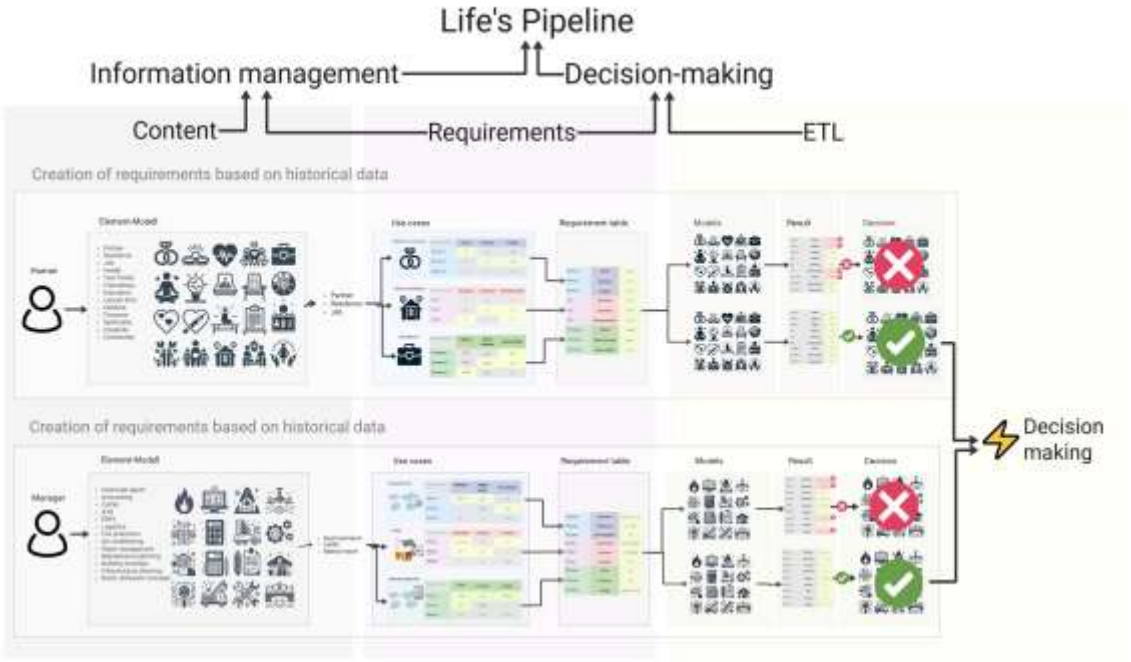


図 4.4-21 ビジネスと人生全般は、データ主導の意思決定の連続であり、意思決定に使用されるデータの質が重要な要素である。

人間を含む生物が自然の法則に従い、進化し、変化する状況に適応しているように、データの収集・分析方法を含む人間のプロセスも、このような自然の原理を反映している。自然と人間の活動の密接な関係は、私たちが自然に依存しているだけでなく、意思決定のためのデータ・アーキテクチャ、プロセス、システムを構築するために、何百万年もの進化によって磨かれた法則を適用したいという私たちの願望を裏付けている。

新しい技術、特に建設技術は、人類がいかに自然からインスピレーションを得て、より優れた、より持続可能で効率的な解決策を生み出してきたかの典型的な例である。

次のステップ：データを正確な計算と計画に変える

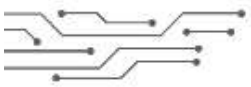
このパートでは、非構造化データを構造化フォーマットに変換する方法、データモデルを開発する方法、建設プロジェクトにおける情報の品質をチェックするプロセスを整理する方法について考察した。データ管理、標準化、分類は、体系的なアプローチとビジネス要件の明確な理解を必要とする基本的なプロセスである。このパートで取り上げる技術やツールは、オブジェクトのライフサイクル全体を通して、異なるシステム間の信頼性の高い統合を可能にします。

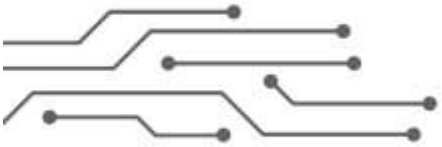
このパートをまとめるために、議論したアプローチを日常業務に応用するのに役立つ主な実践的ステップを紹

介しよう：

- 要件を体系化することから始める
  - ☐ プロジェクトや プロセスの主要要素の属性とパラメータのレジストリを作成する。
  - ☐ 各属性の境界値を文書化する
  - ☐ フローチャート（Miro、Canva、Visio など）を使って、プロセスやクラス、システム、属性間の関係を視覚化する。
- データ変換の自動化
  - ☐ プロセスで頻繁に使用される文書のうち、OCR ライブラリを使ってデジタル化し、表形式に変換できるものはどれかをチェックする。
  - ☐ リバースエンジニアリングツールをチェック CAD からデータを抽出する (BIM)
  - ☐ 仕事でよく使う文書やフォーマットから、表形式にデータを自動取得する設定をしてみる
  - ☐ 異なるデータ形式間の自動変換を設定する
- 分類のための知識ベースを作る
  - ☐ 業界標準に沿った 要素分類器を社内で開発するか、既存のものを使用する。
  - ☐ 異なる分類システム間の相互関係を文書化する。
  - ☐ 元素の識別と明確な分類のための統一されたシステムを使用することについて、チームで話し合う。
  - ☐ チーム内で扱うデータと外部システムに渡すデータの両方を自動的に検証するプロセスの構築を開始する。

これらのアプローチを使用することで、データの品質を大幅に向上させ、その後の処理や変換を簡素化することができます。本書の次のパートでは、すでに構造化され準備されたデータを、自動計算、コスト見積もり、スケジューリング、建設プロジェクト管理に応用する方法を見ていきます。





V 部

コストと時間の計算：建設プロセスにデータを組み込む

第 5 部では、建設プロジェクトのコスト計算とプランニングを最適化するためのデータ活用について、実践的な側面から解説する。原価積算のリソースベース手法（ ）と積算プロセスの自動化について詳しく分析している。CAD (BIM)-モデルから Quantity Take-Off)を自動取得する方法と、計算システムとの統合について検討する。時間スケジューリングと建設コスト管理のための 4D と 5D モデリング技法について、その具体的な適用例とともに検討する。不動産オブジェクトの持続可能性、運用、安全性を評価するための統合的なアプローチを提供する、拡張情報レイヤー 6D -8D の分析が示されている。最新の環境要件や基準に照らし合わせて、建設プロジェクトのカーボン・フットプリントや ESG の指標を算出する方法について詳しく考察している。建設プロセス管理における従来の ERP と PMIS システムの可能性と限界について、価格の透明性への影響を分析しながら批判的に評価する。クローズドなソリューションからオープンな標準や柔軟なデータ分析ツールへの移行が進み、建設プロセスの効率化が進むことが予測される。

章 5.1.

建設プロジェクトのコスト計算と見積もり

建設の基本：数量、コスト、時間の見積もり

建設業界において、企業の持続可能性を左右する数多くのビジネスプロセスの中で、特に重要なのは、数千年前もそうであったように、要素数、プロジェクトコスト、納期を正確に見積もるプロセスである（図 5.1-1）。

文字の発達は、初期社会における経済取引、貿易、資源管理を記録する必要性など、複合的な要因の結果であった。初めて法的に重要な文書となったのは、材料費と人件費を計算した粘土板で、貿易や建設の場面で使用された。これらの石版は、建造物の建設における当事者の義務を記録し、合意や金銭・貿易関係の証拠として保管された。

> 何千年もの間、積算のアプローチはほとんど変わっていません。積算エンジニアの経験と勘に頼って、手作業で計算が行われてきました。しかし、モジュール式 ERP システムと CAD ツールの登場により、数量、コスト、時間を見積もる従来のアプローチは急速に変化し始めた。今日のデジタル技術は、時間とコストの重要な計算を完全に自動化することを可能にし、建設プロジェクトのリソースプランニングにおいて、より高い精度、スピード、透明性を実現する。

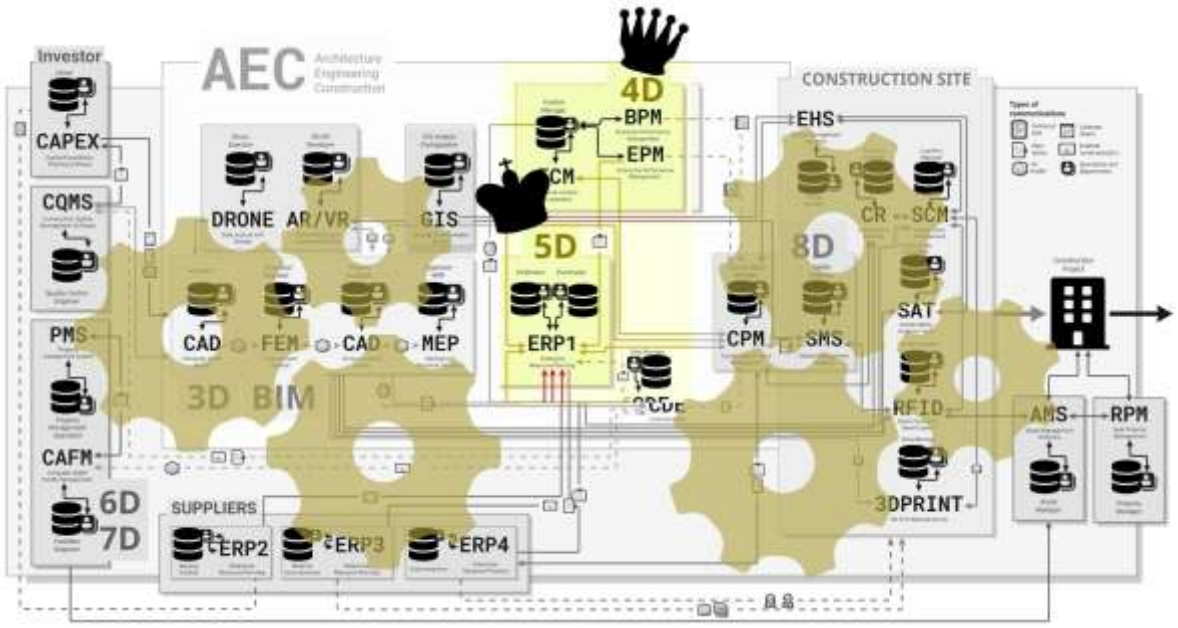


図 5.1-1 さまざまなシステムの中でも、数量、コスト、時間の指標を担当するツールは、ビジネスにおいて最も重要である。

建設会社が重視するのは、正確な時間とコストのデータである。これらは、使用する資材や労働力の量に左右され、その透明性は収益性に影響する。しかし、計算プロセスの複雑さや透明性の欠如は、しばしばプロジェクトコストの上昇、納期の遅れ、さらには倒産につながる。

> KPMG の報告書 "Familiar Problems - New Approaches"（2023 年）によると、建設プロジェクトが予定通りに完了するのは 50%に過ぎず、企業の 87%が資本プロジェクトの経済性に対する管理強化について報告している。主な問題は、有資格者の不足とリスク予測の難しさに関連している[2]。

過去の原価計算と工程時間のデータは、建設会社の生涯を通じて、過去のプロジェクトの建設中に収集され、様々なシステム（ERP 、PMIS BPM、EPM など）のデータベースに入力される。

> 良質な過去の原価計算データを持つことは、建設組織にとって大きな競争優位性であり、その存続に直接影響する。

建設会社やエンジニアリング会社の見積・原価計算部門は、プロジェクト計算に関する過去のデータを収集、保存、更新するために設置される。その主な役割は、会社の経験を蓄積し体系化することであり、これにより新規プロジェクトのスコープ、タイミング、コストの見積もり精度を時間の経過とともに向上させることができる。このアプローチは、すでに実施されたプロジェクトの実践と結果に基づき、将来の計算における誤差を

最小限に抑えるのに役立つ。

プロジェクトの推定コストの計算方法

積算の専門家には様々な積算方法があり、それぞれ特定の種類のデータ、入手可能な情報、プロジェクトの詳細度に重点を置いている。最も一般的なものは以下の通りである：

- リソースベース法: 材料、設備、労働力など、必要なすべてのリソースの詳細な分析に基づいて、プロジェクトの見積もりコストを見積もる。この方法では、すべてのタスクと各タスクの実行に必要なリソースの詳細なリストが必要であり、その後、それらのコストを計算する。この方法は精度が高く、コスト見積りに広く利用されている。
- パラメトリック法：統計モデルを用いて、プロジェクト・パラメータに基づくコストを見積もる。これには、建築面積や作業範囲などの測定単位あたりのコストを分析し、これらの値を特定のプロジェクト条件に適合させることが含まれる。この方法は、詳細な情報がまだ得られていない初期段階で特に有効である。
- 単位法（ユニットコスト法）：測定単位あたりのコスト（平方メートルや立方メートルあたりなど）に基づいて、プロジェクトの見積もりコストを計算します。これにより、異なるプロジェクトやプロジェクトの一部のコストを迅速かつ簡単に比較・分析することができます。
- 専門家の判断（デルファイ法）：専門家の意見に基づく。専門家は、 、経験と知識を駆使してプロジェクトの価値を見積もる。正確なベースラインデータが入手できない場合や、プロジェクトが特殊な場合に有効。

注目すべきは、パラメトリック手法と専門家の判断を機械学習モデルに適応できることである。これにより、トレーニングサンプルに基づいたプロジェクトコストとスケジュールの予測を自動生成することができます。このようなモデルの適用例は、「プロジェクトコストとスケジュールを求めるための機械学習 の使用例」の章（図 9.3-5）で詳しく説明しています。

とはいえ、リソースベース工法は、世界の実務において最も一般的で広く使用されている工法である。資源ベース工法は、見積もりコストの正確な評価を提供するだけでなく、建設現場やプロジェクト全体における各工程の工期を算出することもできる（詳細は「建設スケジュールと 4D -プロジェクトデータ」の章を参照）。

Resource-based method Costing and Estimating 建設業における原価計算と見積

資源ベース原価計算とは、管理会計の手法の一つで、プロジェクトに関わる全ての資源を直接的に計算することにより、プロジェクトの原価を算出するものである。建設業では、この手法により、作業遂行に必要なすべての資材、労働力、技術資源を詳細に分析・評価する。

> リソースベース法 は、見積もり時点におけるリソースの実際の価格に焦点を当てるため、予算計画において高い透明性と正確性を提供する。これは、価格の変動がプロジェクト全体のコストに大きな影響を与える不安定な経済環境においては特に重要である。

以下の章では、資源ベース原価計算プロセス（ ）について詳しく見ていく。その原理をよりよく理解するために、レストランでのディナーの原価計算を例にとって説明する。レストランのマネージャーは、ディナーの計画を立てる際、必要な商品のリストを作成し、各料理の調理時間を考慮した上で、コストにゲストの人数を掛けます。建設においても、このプロセスは同様である。プロジェクト要素（オブジェクト）の各カテゴリーについて、項目別に見積もりが作成される。 、プロジェクトの総コストは、カテゴリー別の最終見積もりである総請求書にすべてのコストを合計することによって決定される。

リソースベースド・アプローチの重要な初期段階は、会社の初期データベースの作成である。原価計算の最初の段階では、倉庫にある釘から、資格や時間給を通した人物の記述に至るまで、建設プロジェクトの中で会社が自由に使えるすべての品目、資材、作業の種類、資源の構造化されたリストが作成される。この情報は、名称、特性、測定単位、現在の価格などのデータを含む表形式の登録簿である「建設資源・資材データベース」に体系化される。このデータベースが、その後のすべての資源計算（工事のコストと時期の両方）の主要な情報源となる。

建設資源のデータベース：建設資材と工事のカタログ

建設資源・資材のデータベース（表）には、建設プロジェクトで使用される製品、品目、資材、サービスなどの各要素について、その名称、説明、測定単位、単価などの詳細情報が構造化された形で記録されている。この表には、プロジェクトで使用される燃料や資材の種類から、専門家の詳細なリストまで、さまざまなカテゴリーに分類され、時間給が記載されています（図 5.1-2）。

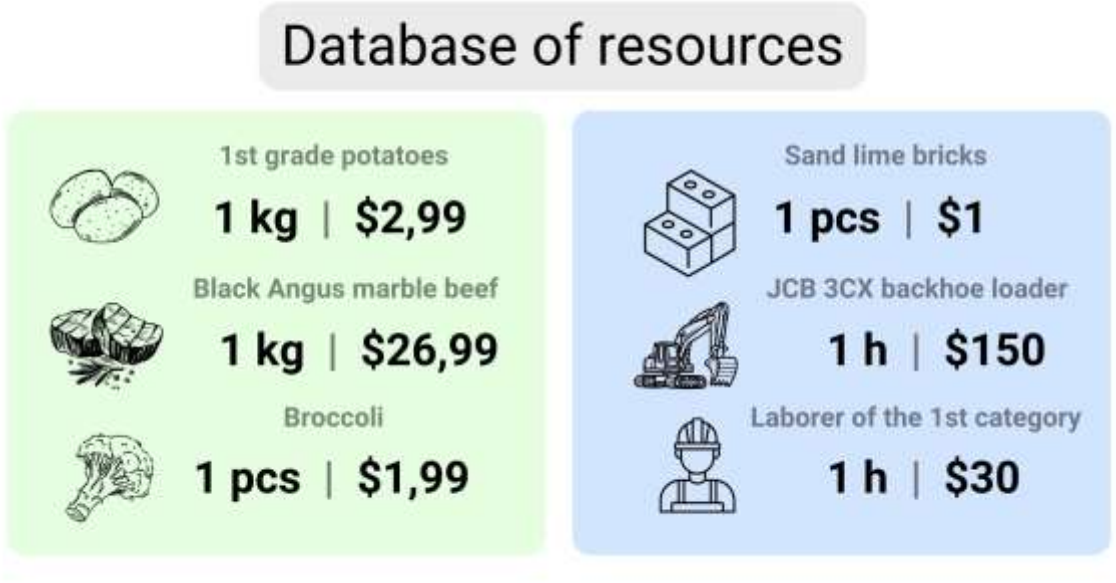


図 5.1-2 リソーステーブルとは、材料とサービスを単価で記述した成分表である。

"資源データベース "とは、オンラインショップの商品カタログのようなもので、各アイテムにはその属性の詳細な説明が記載されている。これにより、積算担当者は、具体的な工事工程を計算するために必要なリソースを、計算書（オンラインショップの最終注文）の形で適切に選択（買い物かごに入れる際に商品を選択するようなもの）しやすくなる。

資源データベースは、レストランの料理本に載っているすべての材料のリストと考えることもできる。それぞれの建築資材、設備、サービスは、レシピに使われる食材に似ている。「資源データベース "とは、すべての材料、つまり建築材料やサービスの詳細なリストであり、それらの単位あたりのコスト（ピース、メートル、時間、リットルなど）を含む。

新しいエンティティエレメントを "Construction Resource Databases "テーブルに追加するには、手動で追加する方法（図 5.1-3）と、会社の在庫管理システムやサプライヤーデータベースと統合して自動的に追加する方法がある。

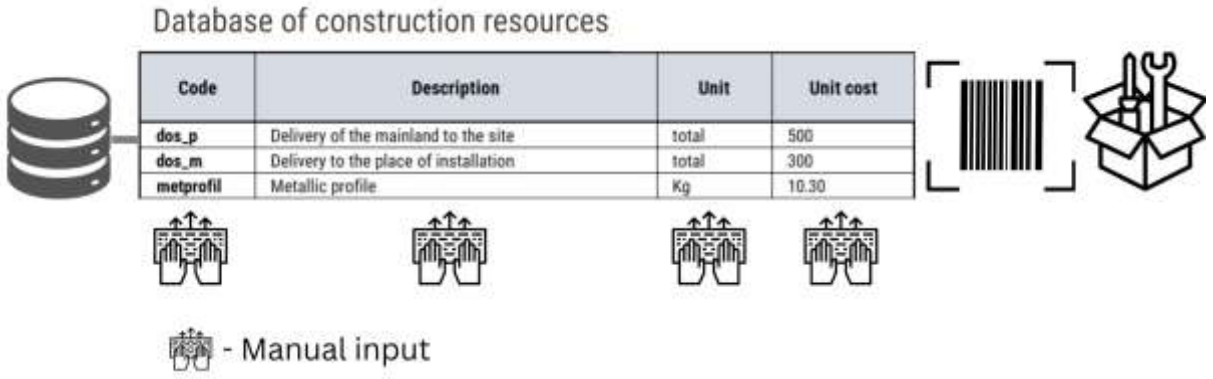


図 5.1-3 リソース・データベースは、手動で入力するか、他のデータベースからデータを自動的に採用する。

一般的な中規模の建設会社では、建設プロジェクトで使用可能な数千、場合によっては数万の項目と詳細な説明を含むデータベースを活用している。このデータは、契約書やプロジェクト文書に自動的に活用され、作業の構成や工程を正確に記述する

> インフレなどの市場環境の変化に対応するため、資源データベース（図 5.1-3）の各製品（商品またはサービス）の「単価」属性は、定期的に手動で更新されるか、他のシステムやオンラインプラットフォームから現在の価格を自動的にダウンロードして更新されます。

資源の単価の更新は、資源の性質、インフレ、外部経済情勢に応じて、月次、四半期、年次で行うことができる。このような更新は、コスト見積もり担当者の作業の出発点となる基本的な要素であるため、計算と見積もりの精度を維持するために必要である。最新のデータは、実際の市場状況を反映した見積もり、予算、スケジュールを作成するために使用され、その後のプロジェクト計算におけるエラーのリスクを軽減する。

計算の取りまとめ およびリソース・ベースに基づく工事の原価計算

建設資源データベース」（図 5.1-3）に最小単位のエンティティを入力したら、 、建設現場の各工程や作業について、特定の測定単位（例えば、コンクリート 1 立方メートル、石膏ボードの壁 1 平方メートル、縁石 1 メートル、窓の取り付け 1 枚など）で計算する計算の作成を開始できます。

例えば、1m^{2}のレンガの壁（図 5.1-4）を作るには、過去のプロジェクトの経験から、約 65 個のレンガ（実体「ケイ酸塩レンガ」）が必要で、1 個あたり 1 ドルのコスト（属性「1 個あたりのコスト」）、合計 65 ドルである。また、私の経験では、作業エリアの近くにレンガを置く建設機械（エンティティ「JCB 3CX ローダー」

）を 10 分間使用する必要がある。 、1 時間あたり 150 ドルのレンタル料がかかるので、6 分間の使用で約 15 ドルかかることになる。さらに、レンガを積む業者が 2 時間必要で、時給は 30 ドル、合計 60 ドルである。

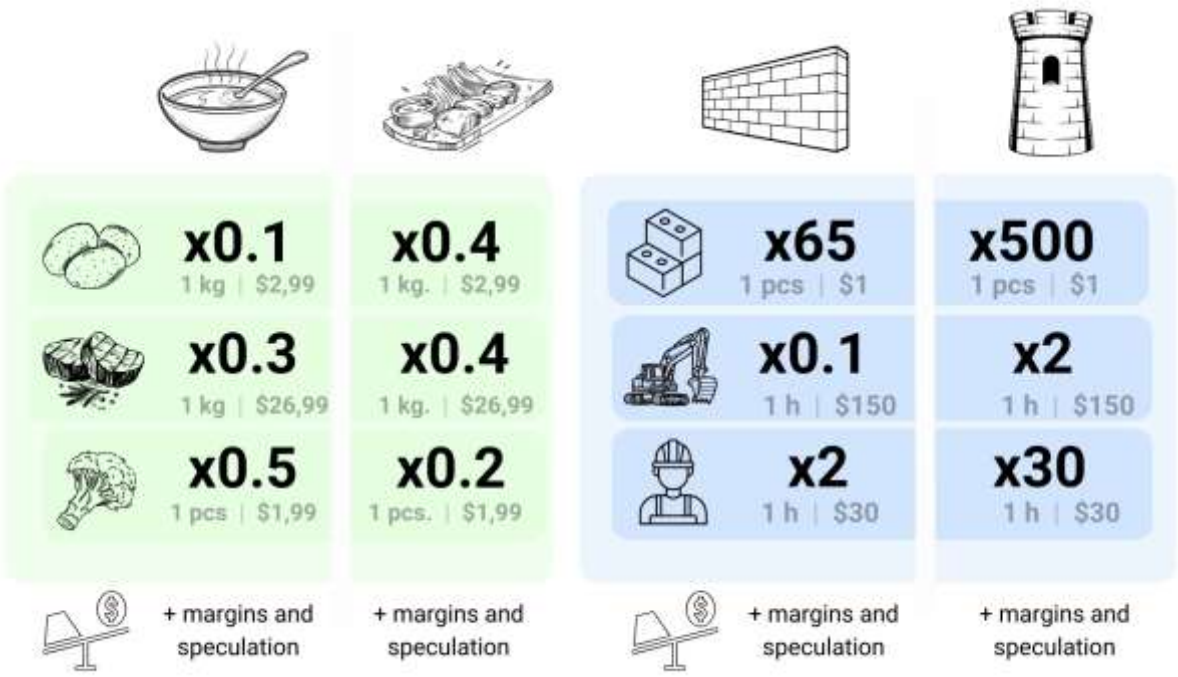


図 5.1-4 費用計算では、工事や工程の実行に必要な建設資材やサービスの詳細なリストが提供される。

計算の構成（いわゆる "レシピ"）は、同社が同様の作業を大量に行う過程で蓄積した過去の経験に基づいて形成される。この実務経験は通常、建設現場からのフィードバックを通じて蓄積される。特に、現場監督が建設現場で直接情報を収集し、実際の人件費、資材消費量、技術的操作のニュアンスを記録する。積算部門と協力しながら、この情報は繰り返し改良される。工程の説明は改良され、リソースの構成は調整され、積算は最近のプロジェクトの実績データを反映して更新される。

> レシピが料理を作るのに必要な材料と量を説明するように、見積もりシートは、特定の仕事やプロセスを完了するのに必要なすべての建設材料、資源、サービスの詳細なリストを提供します。

定期的に実施される作業により、作業員、現場監督、積算担当者は、必要な資源量、すなわち材料、燃料、労働時間、その他単位作業を実施するために必要なパラメータを把握することができる（図 5.1-5）。これらのデータは、表の形で見積システムに入力され、各作業と作業は、資源ベースの最小要素（常に更新される価格）

を通して記述され、これによって計算の正確さが保証される。

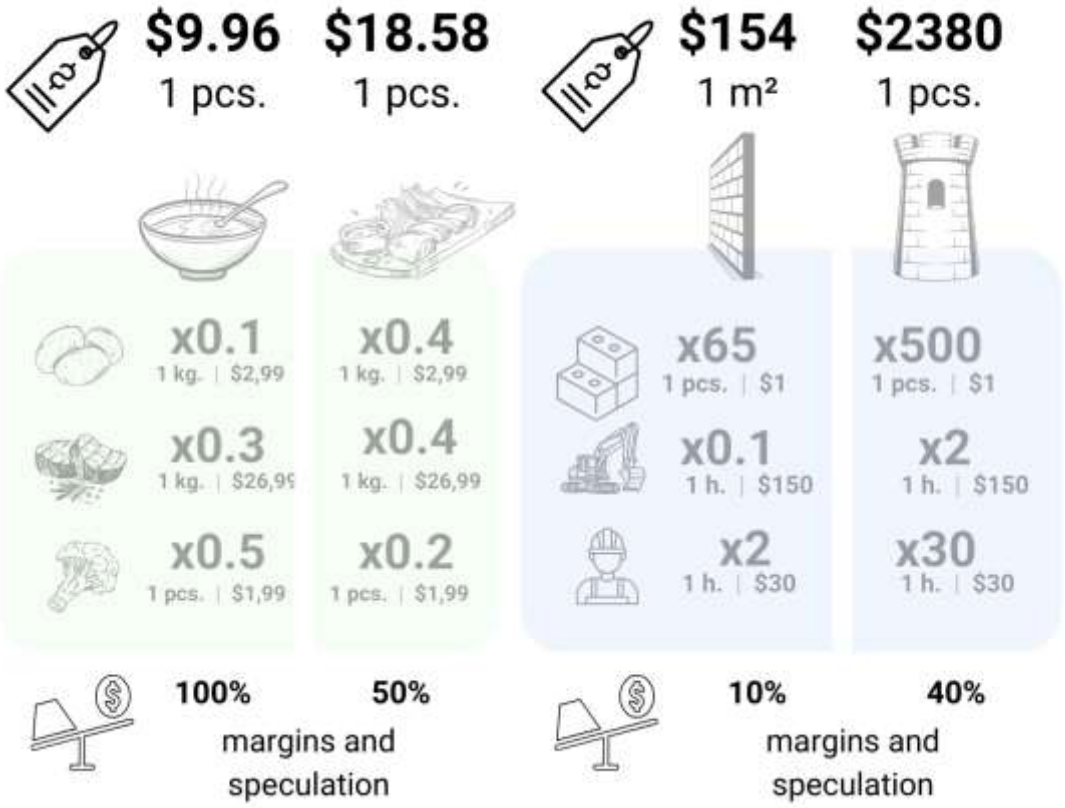


図 5.1-5 単価はジョブごとに収集され、エンティティのボリューム属性にその数量を乗じ、利益率が付加される。

各プロセスやアクティビティ（原価計算対象）の総コストを求めるには、コスト属性にその数と係数を掛ける。係数は、作業の複雑さ、地域特性、インフレ率、潜在的リスク（予想諸経費率）、投機（追加利益係数）など、さまざまな要素を考慮することができる。

> 積算担当者は分析担当者として、現場監督の経験や推奨事項を標準化された積算（）に変換し、表形式のリソースエンティティによって建設プロセスを記述する。要するに、積算担当者の仕事は、建設現場からもたらされる情報を収集し、パラメータや係数を用いて構造化することである。

従って、作業単位（例えば、平方メートルや立方メートル、あるいは 1 台の据付工事）あたりの最終的なコストには、材料や労働の直接費だけでなく、会社のマークアップ、諸経費、保険、その他の要素も含まれる（図 5.1-6）

同時に、実際の価格は常に "リソースベース"(食材テーブル)に反映されているため、(レシピ)計算で実際の価格を気にする必要がなくなります。計算のレベルでは リソースベースからのデータは自動的にテーブルにロードされる（例えば、アイテムコードやそのユニークな識別子によって）、これは説明と単位あたりの実際の価格をロードし、順番に建材のオンラインプラットフォームやオンラインショップから自動的にロードすることができます。 作業の計算レベルの見積り担当者は、「資源の量」という属性と追加要素を通して、作業やプロセスを記述するだけでよい。

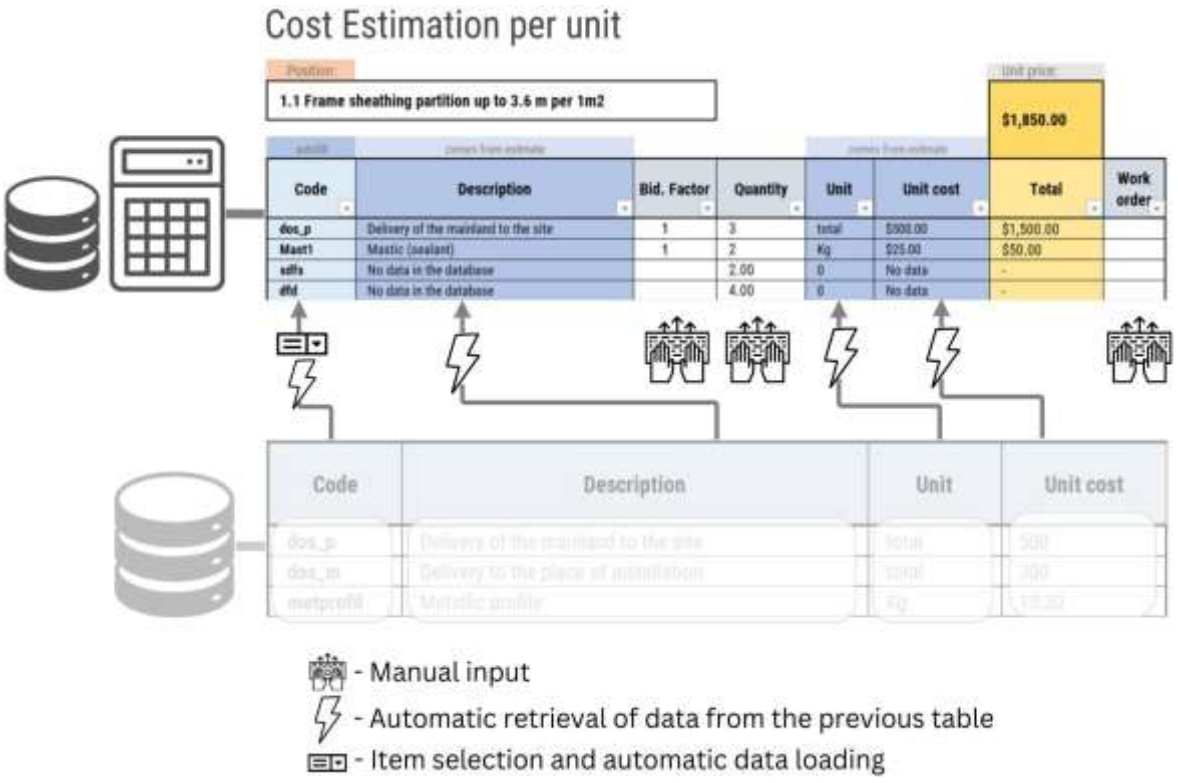


図 5.1-6 作業単価を計算する段階では、必要な資源数の属性のみが記入され、その他はすべて資源データベースから自動的に読み込まれる。

作成された作業原価は、典型的なプロジェクトのテンプレート・テーブルの形で保存され、建設資源と資材のデータベースに直接リンクされている。これらのテンプレートは、将来のプロジェクトで反復される作業タイプの標準化されたレシピを表し、全社的な計算の均一性を保証する。

データベース（図 5.1-3）において、資源コストが変更されると（手動であれ、現在の市場価格をダウンロードすることによる自動的なものであれ（例：インフレ状況））、その更新は、リンクされたすべての原価計算（図 5.1-6）に即座に反映される。つまり、資源ベースだけを変更すればよく、原価計算のテンプレートと見積もりは、長期間変更されない。このアプローチは、比較的単純な資源表（図 5.1-3）のみで説明される価格変動に

対する計算の安定性と再現性を保証する。

> 新しいプロジェクトごとに、標準的な原価計算テンプレートのコピーが作成されるため、会社で採用したオリジナルのテンプレートを変更することなく、特定の要件に合わせて活動を変更・調整することができます。このアプローチでは、建設現場の特殊性や顧客の要望を考慮したり、リスク係数や収益性（投機）係数を導入したりすることができます。これにより、会社は利益の最大化、顧客満足度、競争力の維持のバランスを取ることができます。

国によっては、数十年にわたって蓄積されたこのような原価計算テンプレートが、国レベルで標準化され、国の建設原価計算システム基準の一部となっている（図 5.1-7）。

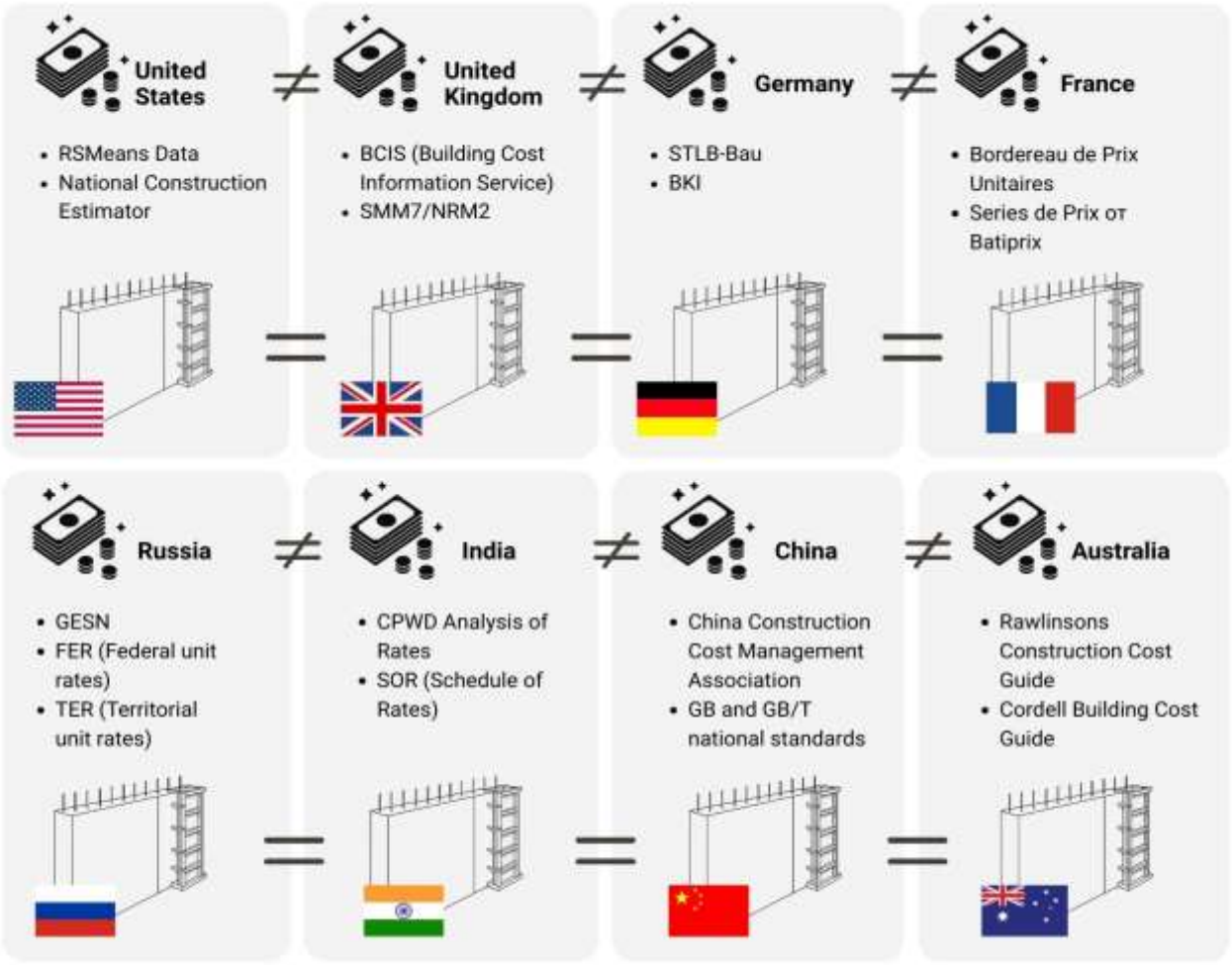


図 5.1-7 世界各国には、同じ要素を原価計算するために、独自の（規定的な）大要や建設工事基準による原価計算ルールがある。

このような標準化された資源見積ベース（図 5.1-7）は、特に公的資金によるプロジェクトの場合、すべての市場参加使用することが義務付けられている。このような標準化によって、価格形成と契約義務の透明性、比較可能性、公平性が確保される

プロジェクトの最終的なコスト計算：見積もりから予算まで

国や業界ごとの積算基準は、各国の建設実務において異なる役割を果たしている。一つの基準に厳格に従うことを要求する国もあるが、ほとんどの先進国はより柔軟なアプローチを採用している。市場経済国では、政府の建設基準は、通常、ベースラインとしてのみ機能する。建設会社は、これらの基準を自社の事業モデルに適合させるか、あるいは完全に改訂し、独自のカスタマイズ要素で補完する。このような調整には、企業の経験や資源管理の効率が反映されるほか、例えば、企業の投機的な利益が織り込まれることも多い。

> その結果、競争の度合い、市場の需要、目標マージン、さらには特定の顧客との関係によって、標準化された規範から大きく逸脱する可能性がある。このような慣行は、市場に柔軟性をもたらすが、同時に、異なる請負業者の入札を透明性をもって比較することを困難にし、計算プロセスのこの段階で、建設業界に投機的な価格設定の要素を持ち込むことになる。

個々の活動やプロセスの計算テンプレートが作成されたら（あるいは、より多くの場合、単に標準的な政府の見積もり（図 5.1-7）からコピーし、特定の企業の「特殊性」を反映するために係数を追加したもの）、最後のステップは、各項目のコストに、新しいプロジェクトの作業範囲またはプロセスの対応する属性を乗算することである。

> 新築プロジェクトの総工費を計算する場合、重要なステップは、すべての原価計算項目の原価をまとめることである。 、プロジェクトにおけるこれらの項目の工事量を乗じる。

プロジェクトの総コストを作成するために、この単純化した例では、まず 1 平方メートルの壁の建設コストを計算し、その計算コスト（例えば、作業「壁要素の 1m^2標準設置」）に、プロジェクト内の壁の総平方メートル数（例えば、プロジェクトの CAD から「壁要素」タイプのエンティティの「面積」または「数量」属性（図 5.1-8）、または現場監督の計算）を掛けます。

同様に、プロジェクトの全要素のコストを計算します（図 5.1-8）：単位作業あたりのコストをとり、プロジェ

クト内の特定の要素または要素グループのボリュームを掛けます。積算担当者は、 に、プロジェクトに含まれるこれらの要素、アクティビティ、またはプロセスの数を、ボリュームまたは数量の形式で入力するだけでよい。これにより、完全な工事見積もりを自動的に作成することができます。

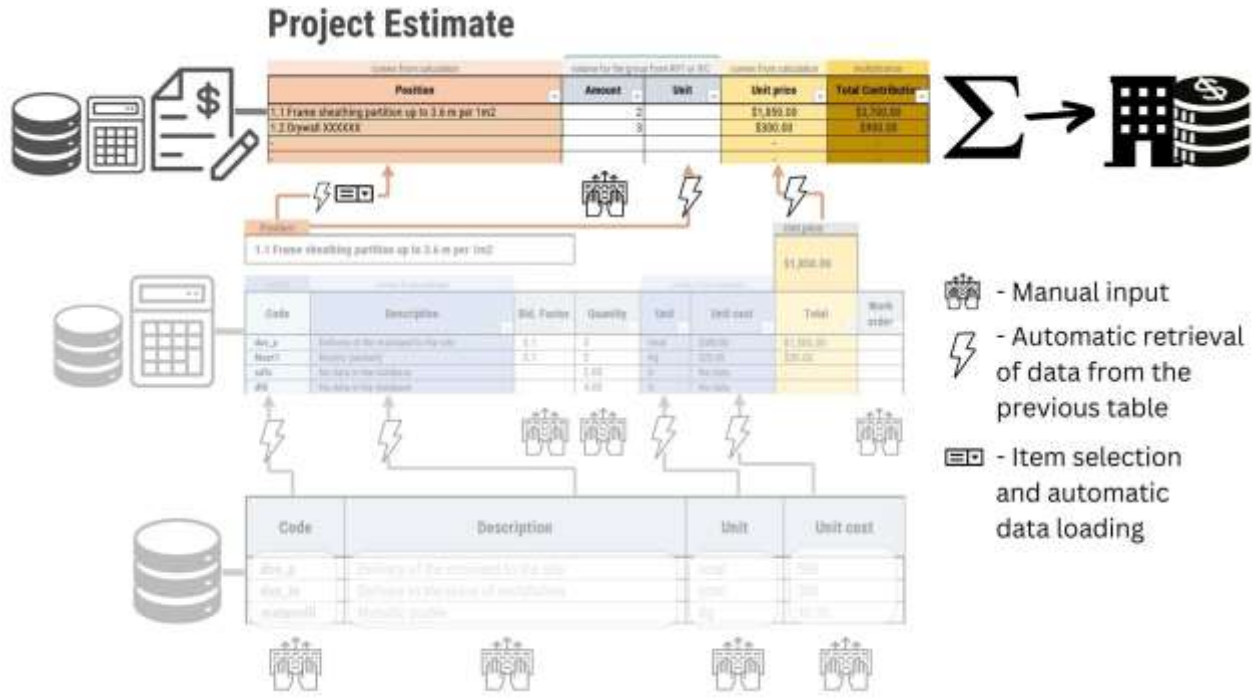


図 5.1-8 見積りを作成する段階では、作業範囲を入力するだけである。

計算の場合と同様に、このレベルでは、自動的に準備された計算項目（計算テンプレートから、またはテンプレートからコピーされ編集された新しいもの）をアップロードします。この計算項目は、自動的に現在の作業単価（これは資源データベースから自動的に更新されます（図 5.1-8 下表））を持ってきます。従って、資源データベースや原価計算テーブルのデータが変更された場合、原価計算や見積もりそのものを変更することなく、見積もりのデータも自動的にその日のものに更新されます。

レストランの場合、イベントの最終的なコストは、同様の方法で計算され、ディナー全体の最終的なコストに等しく、各料理のコストにゲストの人数を掛けたものが、小切手の合計コストになる（図 5.1-9）。また、建設業と同じように、レストランでの料理のレシピは何十年も変わらないかもしれない。食材の値段が刻々と変わる物価とは違う。

> レストランのオーナーが、各料理のコストに料理の数と人数をかけて、イベントの総コストを決定するように、積算担当者は、すべてのプロジェクトのコンポーネントのコストを合計して、完全な建設見積もりを作成します。

こうして、プロジェクト内の各アクティビティについて、その最終コストが決定され（図 5.1-9）、これに、このアクティビティに対応するエンティティの属性ボリュームを乗じることで、アクティビティ・グループのコストが得られ、そこからプロジェクト全体の最終コストが得られる。

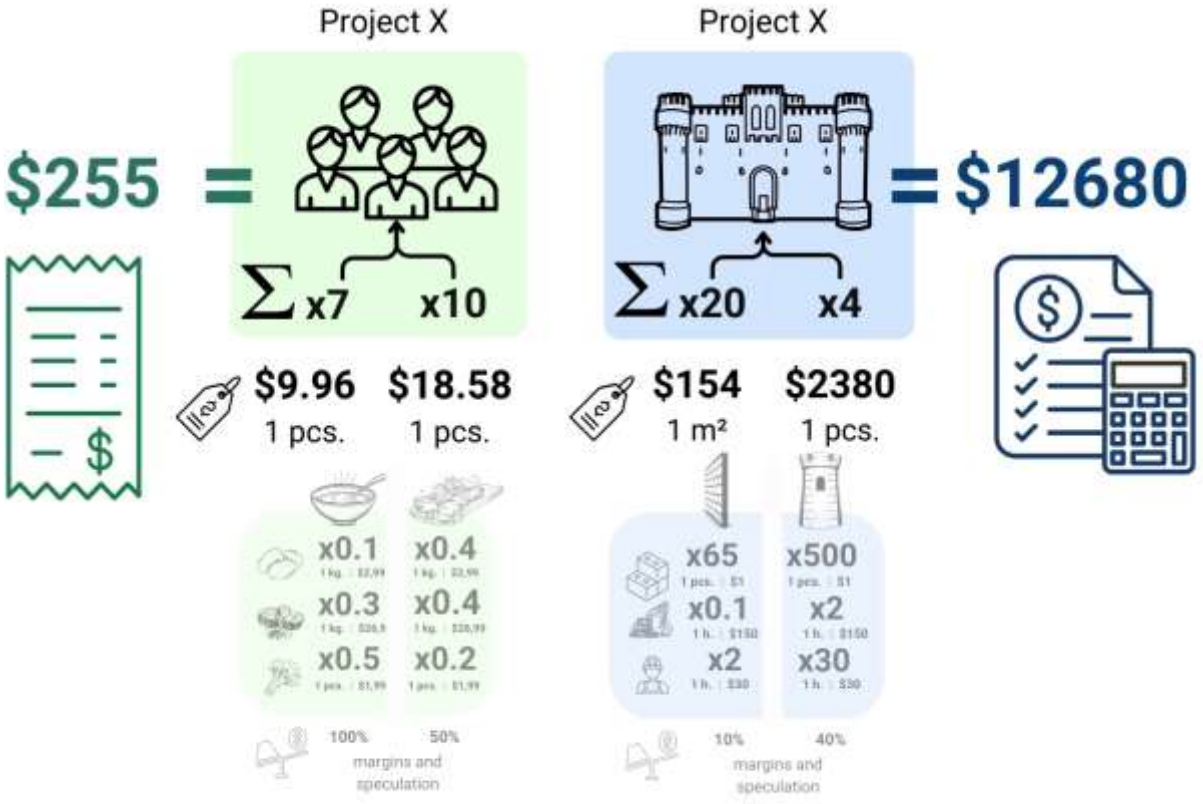


図 5.1-9 最終的な見積もりは、各要素の作業コスト属性をそのスコープ属性で合計することで計算されます。

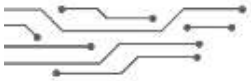
プロジェクトの総費用（図 5.1-8）は、プロジェクトの財務状況を示すもので、顧客、投資家、融資機関が、現在の物価を考慮した上で、任意の日にプロジェクトを実施するために必要な総予算と財源を理解することを可能にする。

そして、資源ベース、計算、見積もりをまとめるプロセス（プロセスレシピ）が、すでに何万年もかけて、半

自動化され、研ぎ澄まされ、国家レベルで記録されてきたとすれば、最終見積もりの最終段階（ ）で、要素の量と数量に関する定性的情報を自動的に入手することは、今日でも、プロジェクトのコストと時間の属性、そして一般的にプロジェクト全体の予算に関するすべての計算プロセスのボトルネックとなっている。

> 何千年もの間、体積を計算する伝統的な方法は、平面の図面を使って手作業で体積と数量を測定する方法であった。デジタル時代の到来により、企業は、CAD モデル に含まれる幾何学的データから体積と数量の情報を自動的に抽出できることを発見し、数千年来の定量データの取得方法に革命をもたらした。

工程見積もりと積算の現代的なアプローチには、CAD データベースから容積属性と定量属性を自動的に抽出することが含まれる。CAD データベースはアップロード可能で、 、設計から運転までのどの段階でも、最新のプロジェクト・グループ容積を得るために、積算プロセスに接続することができる。



章 5.2.

数量テイクオフ、見積もりとスケジュールの自動作成

3D から 4D、5D への移行：体積パラメーターと定量パラメーターの使用

資源を使った工程を記述した原価計算表（図 5.1-8）が手元にあれば、次のステップは、計算と最終見積りに必要な要素グループの体積または数量パラメータを自動的に取得することである。

壁やスラブなどのプロジェクト要素の体積特性を、CAD データベースから自動的に抽出することができます。CAD プログラムで作成されたパラメトリックオブジェクトは、ジオメトリカーネルによって、長さ、幅、面積、体積、その他のパラメータの数値に変換されます。3D ジオメトリに基づく体積の取得プロセスについては、次の第 6 部（図 6.3-3）、CAD（BIM ）を使った作業で詳しく説明します。ボリュームに加えて、類似要素の数は、カテゴリとプロパティによってオブジェクトをフィルタリングし、グループ化することによって、CAD-モデルデータベースから取得することもできます。グループ化を可能にするこれらのパラメータは、リソース計算を通じてプロジェクト要素を、プロジェクト全体の計算、最終見積もり、予算とリンクさせるための基礎となる。

このように、3D（CAD）モデルから抽出されたデータ・モデルは、4D および 5D（ ）と表記される新しいパラメータ・レイヤで拡張される。新しいエンティティ属性レイヤーの 4D（時間）と 5D（コスト）では、3D 幾何学データがエンティティボリューム属性値のソースとして使用されます。

- 4D は、要素の 3D パラメータに建設作業の期間に関する情報を追加するパラメータ情報レイヤです。このデータは、スケジュールを計画し、プロジェクトのスケジュールを管理するために不可欠です
- 5D は、データモデルの次の拡張レベルで、要素にコスト特性が付加されます。これにより、幾何学的な情報に、材料費、工事費、設備費といった財務的な側面が加わり、建設プロセスにおける予算計算、採算分析、コスト管理が可能になります。

プロジェクト・エンティティのグループのコストと 3D 、4D 、5D 属性データは、モジュラーERP、PIMS -システム（または Excel -類似ツール）の計算と同様の方法で記述され、個々のグループと完全なプロジェクト予算の両方の自動原価計算と予算計画に使用されます。

5D 属性 および CAD からの属性ボリュームの取得

前章で説明した建設プロジェクトの最終見積り を作成する際（図 5.1-8）、プロジェクト要素の各カテゴリのボリューム属性 を手作業で収集するか、CAD ソフト が提供するボリューム属性仕様から抽出する。

> 伝統的な手作業による数量計算では、現場監督や積算担当者が、何千年もの間、紙の上の線として、また過去 30 年間は PDF (PLT)や DWG などのデジタル形式で提示されてきた図面を分析する。プロの経験に基づき、定規や分度器を使って、必要な作業や材料の量を測定する。この方法は、かなりの労力と時間、そして細部への特別な注意を必要とする。

このようにしてスコープ属性を決定するには、プロジェクトの範囲にもよるが、数日から数ヶ月かかる。また、測定や計算はすべて手作業で行われるため、人為的なミスによりデータが不正確になるリスクがあり、その結果、プロジェクトの期間やコストの見積もりに誤差が生じ、会社全体の責任となる。

CAD データベースの使用に基づく最新の手法、体積計算を大幅に簡素化します。CAD モデルでは、要素のジオメトリにはすでに体積属性が含まれており、（ジオメトリカーネル（図 6.3-3）を介して）自動的に計算し、表形式で表示またはエクスポートすることができます。

このようなシナリオでは、積算部門が CAD 設計者にプロジェクト要素の数量と体積の特性に関するデータを要求します。このデータは、スプレッドシートとしてエクスポートされるか、直接原価計算データベース（Excel、ERP、PMIS ）に統合されます。このプロセスは、多くの場合、正式な依頼から始まるのではなく、発注者（イニシエーター）と建築家建設会社や設計会社の積算担当者との簡単な対話から始まります。以下は、日常的なコミュニケーションから自動計算のための構造化テーブル（QTO）がどのように形成されるかを示す簡略化した例である：

- 顧客 「1 階と同じ構成で、もう 1 フロア増築したい。
- 建築家（CAD） 「3 階を追加しますが、構成は 2 階と同じです。このメッセージの後、プロジェクトの新しい CAD バージョンが積算担当者に送信されます。
- エスティメーターが自動的にグルーピングと計算を行う（ERP、PMIS、Excel） -「QTO ルール（ERP、PMIS）を使って Excel スプレッドシートにプロジェクトを実行し、新しいフロアのカテゴリー別の数量を取得し、見積もりを作成します

その結果、テキスト対話はグループ化ルールを持つ表構造に変換される：

エレメント	カテゴリー	フロア
オーバーラップ	OST_Floors	3
コラム	OST_StructuralColumns	3
階段	OST_階段	3

設計者からの CAD モデルを、積算の QTO ルールに従って自動的にグループ化し、リソース計算によるボリュームの自動乗算を行うプロセスを経て（図 5.1-8）、以下の結果が得られ、顧客に送信される：

エレメント	ボリューム	フロア	1 台あたりの価格。	総費用
オーバーラップ	420 м²	3	150 €/м²	63 000 €
コラム	4 個	3	2450 ユーロ/個。	9 800 €
階段	2 個	3	4,300 ユーロ/個。	8 600 €
TOTAL だ：	-	-	-	81 400 €

- 顧客 「ありがとうございます、かなり多いので、何部屋か減らさせてください」。このサイクルが何繰り返される。

このシナリオは、特に顧客が即座のフィードバックを期待する承認段階では、何度も繰り返される可能性がある。しかし実際には、このようなプロセスは何日も、あるいは何週間も続くことがある。今日では、自動グループ化と計算ルールの導入により、以前はかなりの時間を要していた作業が数分で完了するようになった。グループ化ルールによる数量の自動取得は、計算と見積もりをスピードアップするだけでなく、人的要因を最小限に抑えることでエラーの可能性を減らし、プロジェクト・コストの透明で正確な評価を提供します。

CAD システムで 3D モデルを作成する際に、積算部門の要件が最初に考慮され（実際にはまだまれである）、要素グループの名前、識別子、およびそれらの分類属性が、積算グループおよびクラスの構造と一致するパラメータの形式で設定されている場合、ボリューム属性は、追加の変換を行うことなく、自動的に積算システムに転送することができる。

CAD から仕様表の形でボリューム属性を自動抽出することで、個々の工事やプロジェクト全体のコストに関する最新データを迅速に入手することができる（図 5.2-1）。計算プロセスや計算システムにおいて、プロジェクトボリュームを含む CAD ファイルのみを更新することで、最新の変更点を考慮した見積りの再計算を迅速に行うことができ、その後のすべての計算において高い精度と一貫性を確保することができる。

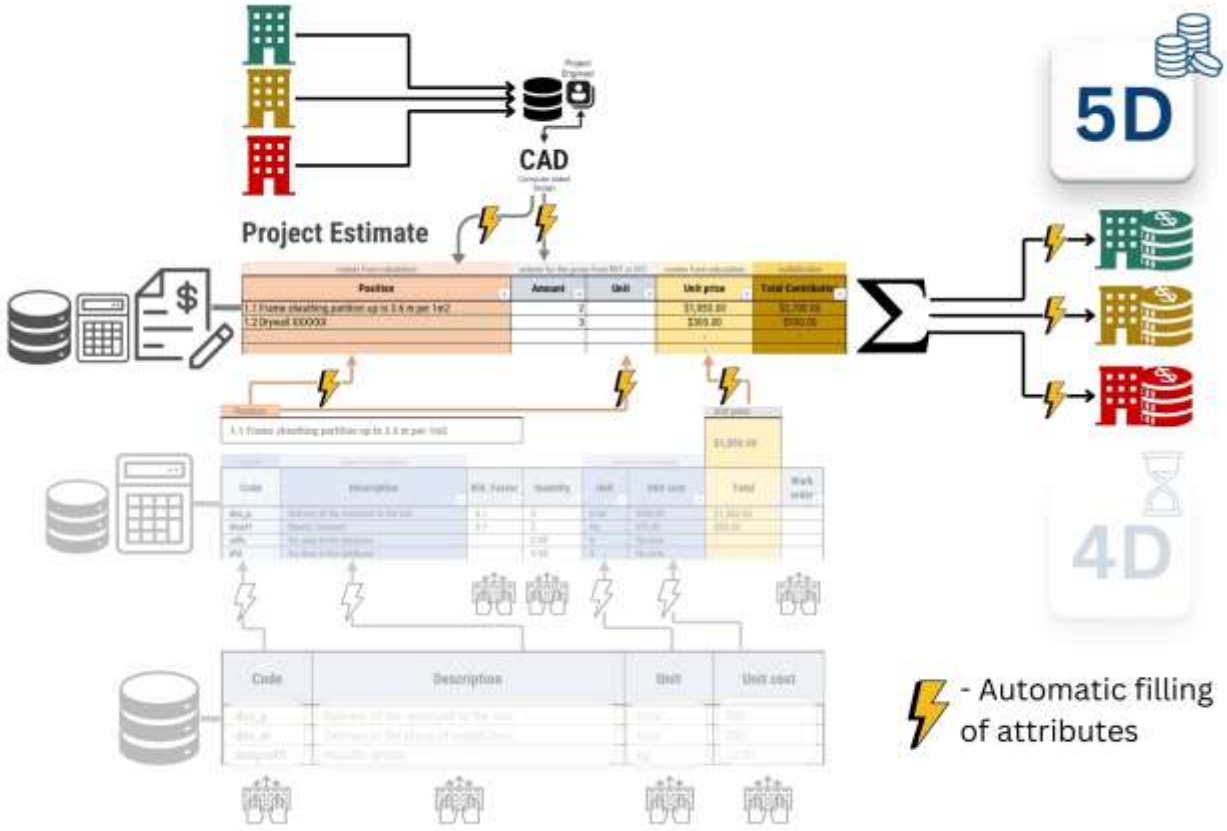


図 5.2-1 CAD テーブルまたはデータベースからのボリューム属性 が自動的に見積もりに入力されるため、プロジェクトの総コストを即座に計算することができます。

資本プロジェクトの複雑化に伴い、このようなシナリオ（図 5.2-1）の下で、全予算を計算し、 プロジェクトの総コストを分析することは、情報に基づいた意思決定のための重要なツールとなる。

> アクセンチュアの「Creating More Value through Capital Projects (2024)」調査[20]によると、大手企業はデータアナリティクスをデジタル施策に積極的に統合し、過去の情報を活用して成果を予測・最適化している。調査によると、設計開始前に市場動向を予測し、商業的実行可能性を評価するためにビッグデータ分析を適用するオーナー経営者が増えている。これは、既存のプロジェクト・ポートフォリオから、データ・ウェアハウスを分析することによって達成される。さらに、79%のオーナー・オペレーターが、プロジェクトのパフォーマンスを評価し、リアルタイムの業務上の意思決定をサポートするために「堅牢な」予測分析を導入している。

建設プロジェクトの現代の効果的な管理は、設計の全段階および設計に先立つプロセスにおける大量の情報の処理 と分析と表裏一体である。 データウェアハウス、資源計算 、予測モデル、機械学習 の活用により、計算上のリスクを最小化できるだけでなく、設計の初期段階でプロジェクトの資金調達に関する戦略的な意思決定を行うことができる。データウェアハウス と、計算を補完する予測モデルについては、本書の第 9 回で詳しく述べる。

CAD プロジェクト から、見積もり作成に必要な要素の体積パラメー タを自動的に取得するには、グルーピングツール QTO (Quantity Take-Off) を使用します。QTO ツールは、CAD データベースで作成された仕様とテーブルを使用して、特別な要素識別子または要素属性パラメータによってすべてのプロジェクトオブジェクトをグループ化することによって機能します。

QTO Quantity Take-Off: 属性によるプロジェクトデータのグループ化

建設における QTO（Quantity Take-Off）とは、プロジェクトの実現に必要な要素の数量を抽出するプロセスのことである。実際には、QTO は多くの場合、PDF 文書、DWG 図面、デジタル CAD モデルなど、さまざまなソースからのデータ収集を伴う半手動のプロセスです。

CAD データベースから抽出されたデータを扱う場合、QTO プロセスはフィルタリング、ソート、グループ化、集計操作のシーケンスとして実現されます。モデル要素は、クラス、カテゴリ、タイプのパラメータに従って選択され、その後、体積、面積、長さ、数量などの定量的属性が計算ロジックに従って要約されます（図 5.2-2

）。

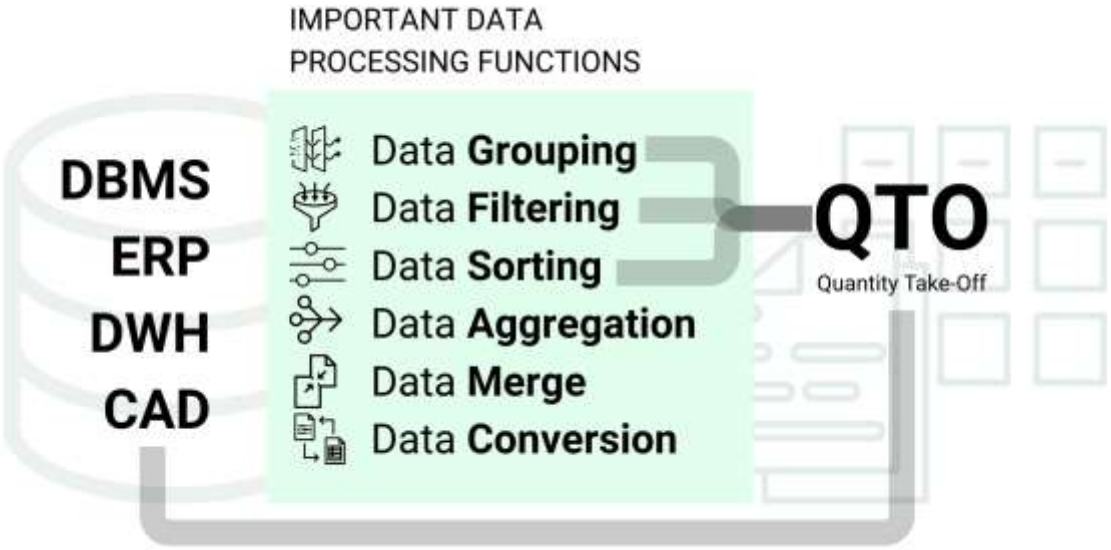


図 5.2-2 データのグループ化とフィルタリングは、データベースやデータウェアハウスに適用される最も一般的な機能です。

QTO（フィルタリングとグループ化）プロセスは、データを体系化し、仕様を形成し、見積もり、購入、作業スケジュールを計算するための入力情報を準備することができます。QTO の基本は、測定属性のタイプによる要素の分類です。各要素または要素グループに対して、対応する定量測定パラメータが選択されます。例えば

- 長さ属性（縁石 - メートル単位）
- 面積属性（石膏ボード工事-平方メートル）
- 体積属性（コンクリート工事-単位：立方メートル）
- 数量属性（窓-1 枚あたり）

ジオメトリに基づいて数学的に生成される容積特性に加えて、オーバーラン係数（図 5.2-12 例えば、ロジスティクスと据付のための 10%を考慮した 1.1）が QTO グループ化後の計算で適用されることが多い。これにより、資材の実際の消費量をより正確に予測することができ、建設現場での不足と過剰在庫の両方を避けることができる。

> 自動数量引取プロセス（QTO ）は、正確な の計算と見積もりを作成し、数量指定プロセスにおける人為的ミスを減らし、材料の過不足注文を防ぐために不可欠である。

QTO プロセスの一例として、CAD データベースから、あるカテゴリー（クラス）の要素タイプ別の体積をテーブル仕様で示す必要がある一般的なケースを考えてみよう。CAD のプロジェクト壁カテゴリーから、すべてのプロジェクト要素をタイプ別にグループ化し、各タイプの体積属性をまとめて、QTO 体積表 （図 5.2-3）として結果を示してみましょう。

典型的な CAD プロジェクトの例（図 5.2-3）では、CAD データベース 内のすべての壁カテゴリー要素は、壁タイプ別にグループ化されており、例えば "Lamelle 11.5"、"MW 11.5"、"STB 20.0 "のように、メトリックキューブで表現された明確な体積属性を持っている。

設計者と計算スペシャリストの接点に立つマネージャーの目標は、選択されたカテゴリーの要素タイプ別体積表を自動で作成することです。特定のプロジェクトだけでなく、同じようなモデル構造を持つ他のプロジェクトにも適用可能な普遍的な形式で。これにより、アプローチに拡張性を持たせることができ、労力を重複させることなくデータを再利用することができる。

> 経験豊富な設計者 や積算担当者 が定規で武装し、紙や PDF -図面上の一本一本の線を注意深く測定していた時代は終わった。3D -モデリングが発達し、各要素の形状が自動的に計算された容積属性に直接リンクされるようになったため、容積と QTO 数量を決定するプロセスが自動化された。

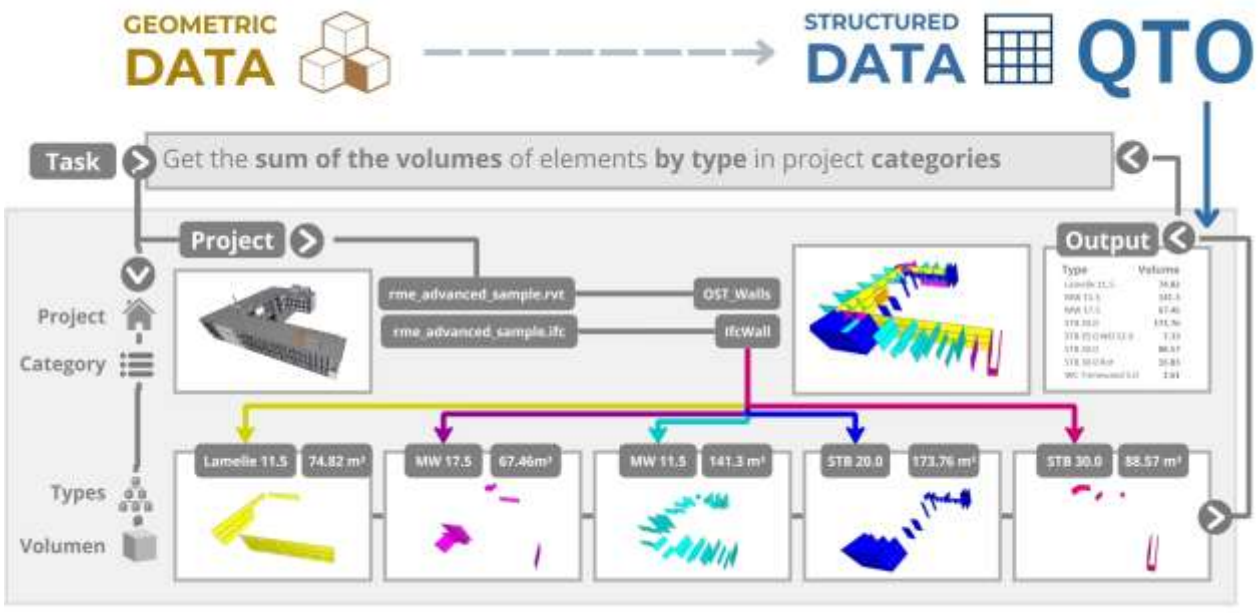


図 5.2-3 QTO プロジェクトから量と数量の属性を取得するには、プロジェクト要素をグループ化し、フィルタリングする必要があります。

この例では、「プロジェクト内の壁のカテゴリーを選択し、すべての要素をタイプ別にグループ化し、ボリューム属性に関する情報を構造化された表形式で提示する」ことで、この表を他の何十人ものスペシャリストが原価計算、ロジスティクス、作業スケジュール、その他のビジネスケースの計算に使用できるようにすることがタスクである（図 6.1-3）。

CAD データのクローズドな性質により、 今日、すべての専門家が CAD データベースへの直接アクセスを利用できるわけではない（アクセス問題の理由と解決策は、本書の第 6 部に詳述されている）。そのため、多くの人々は、オープン BIM とクローズド BIM [63]の概念に基づく特殊な BIM ツールに頼らざるを得ない。専門の BIM ツール（ ）で作業する場合、または CAD ソフトウェア環境で直接作業する場合、QTO（Quantity Take-Off）結果を含む表は、手動インターフェースまたはソフトウェア自動化のどちらを使用するかによって、異なる方法で生成することができます。

例えば、CAD(BIM)ソフトのユーザー・インタフェースを使用する場合、約 17 回の操作（ボタンクリック）を行うだけで、出来上がった出来高表を得ることができる（図 5.2-4）。ただし、ユーザーは CAD（BIM）ソフトのモデル構造や機能を十分に理解している必要があります。

自動化がプログラムコード、または CAD プログラム内のプラグインや API ツールによって適用される場合、ボリュームテーブルを取得するための手動ステップの数は減りますが、使用するライブラリやツールに応じて、40〜150 行のコードを記述する必要があります：

- IfcOpSh (オープン BIM) または Dynamo IronPython (クローズド BIM) - たった 40 行のコードで、CAD フォーマットまたは CAD プログラムから QTO テーブルを取得できます。
- IFC_js (オープン BIM) - IFC モデルから膨大な属性を抽出するために約 150 行のコードが必要です。
- インターフェース CAD ツール (BIM) - マウスのクリック数 17 回で同じ結果を手動で得ることができます。

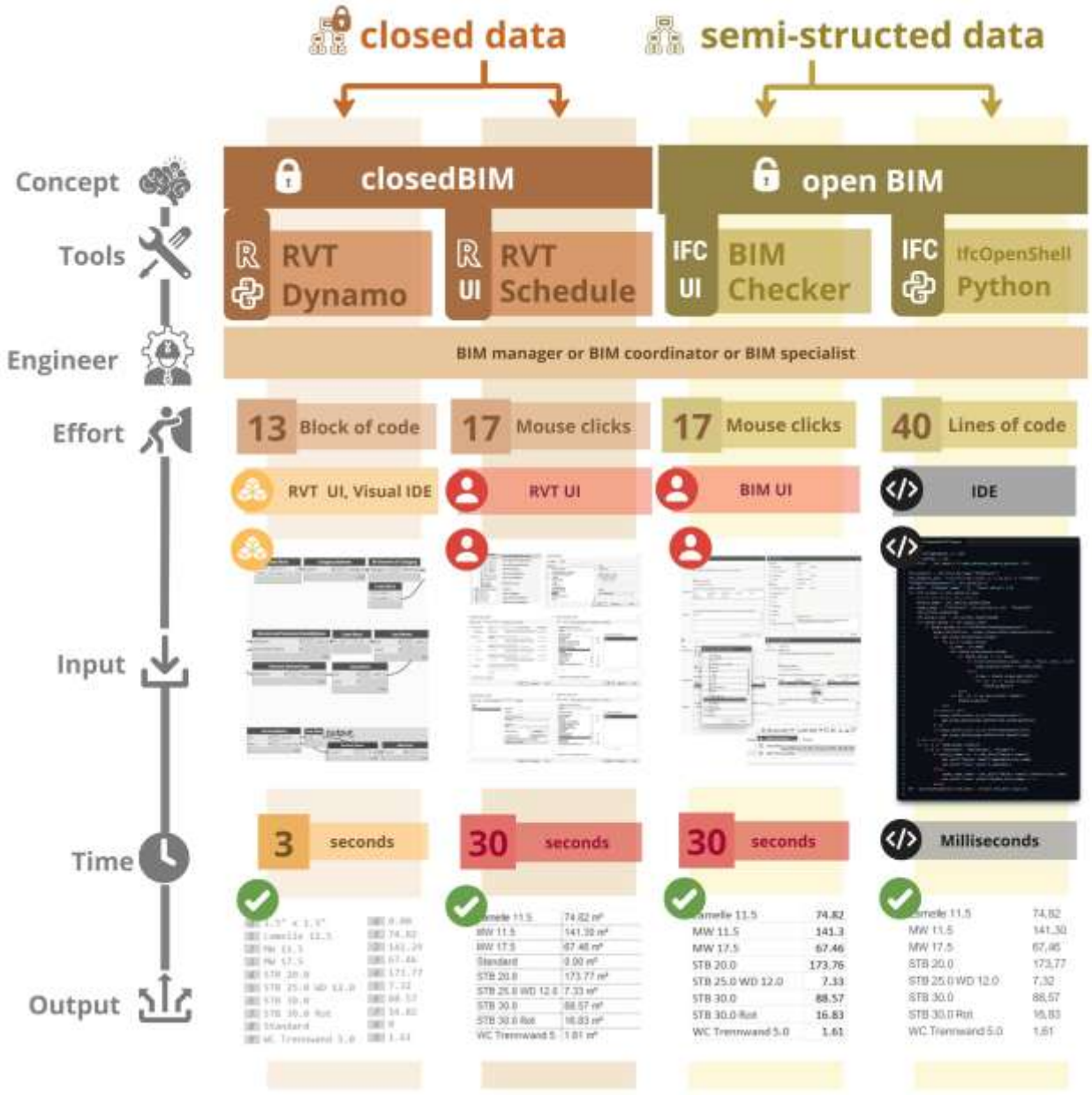


図 5.2-4 CAD (BIM) の設計者や管理者は、QTO テーブルを作成するために、40 ～ 150 行のコードまたは数十のキーストロークを使用する。

結果は同じである。つまり、要素グループのボリューム属性を持つ構造化テーブルである。唯一の違いは、人件費とユーザーが必要とする技術的専門知識のレベルである（図 5.2-4）。出来高の手作業による収集に関連して、最新のツールは QTO プロセスを大幅にスピードアップし、エラーの確率を低減する。プロジェクトモデルから直接データを抽出できるため、従来のように図面から手作業で出来高を再計算する必要がない。

オープン BIM（）であろうとクローズド BIM（）であろうと、使用する手法に関係なく、プロジェクト要素のボリュームを持つ同一の QTO（）テーブルを得ることは可能である（図 5.2-4）。しかし、CAD - (BIM-)のコンセプトでプロジェクトデータを扱う場合、ユーザーはベンダーが提供する専門ツールや API に依存する（図 3.2-13）。このため、データへの直接アクセスが制限される一方で、依存のレイヤーが増え、独自のデータスキーマを学習する必要があります。

CAD データは閉鎖的であるため、QTO テーブルやその他のパラメーターの取得は、計算の自動化や外部システムとの統合を複雑にする。 データベースに直接アクセスするツールを使用し、リバースエンジニアリングツールによってプロジェクトの CAD -data をオープンな構造化データフレーム形式に転送することにより（図 4.1-13）、わずか 1 行のコードで同一の QTO テーブルを取得することができます（図 5.2-5 - 細かいデータによるバリエーション）。

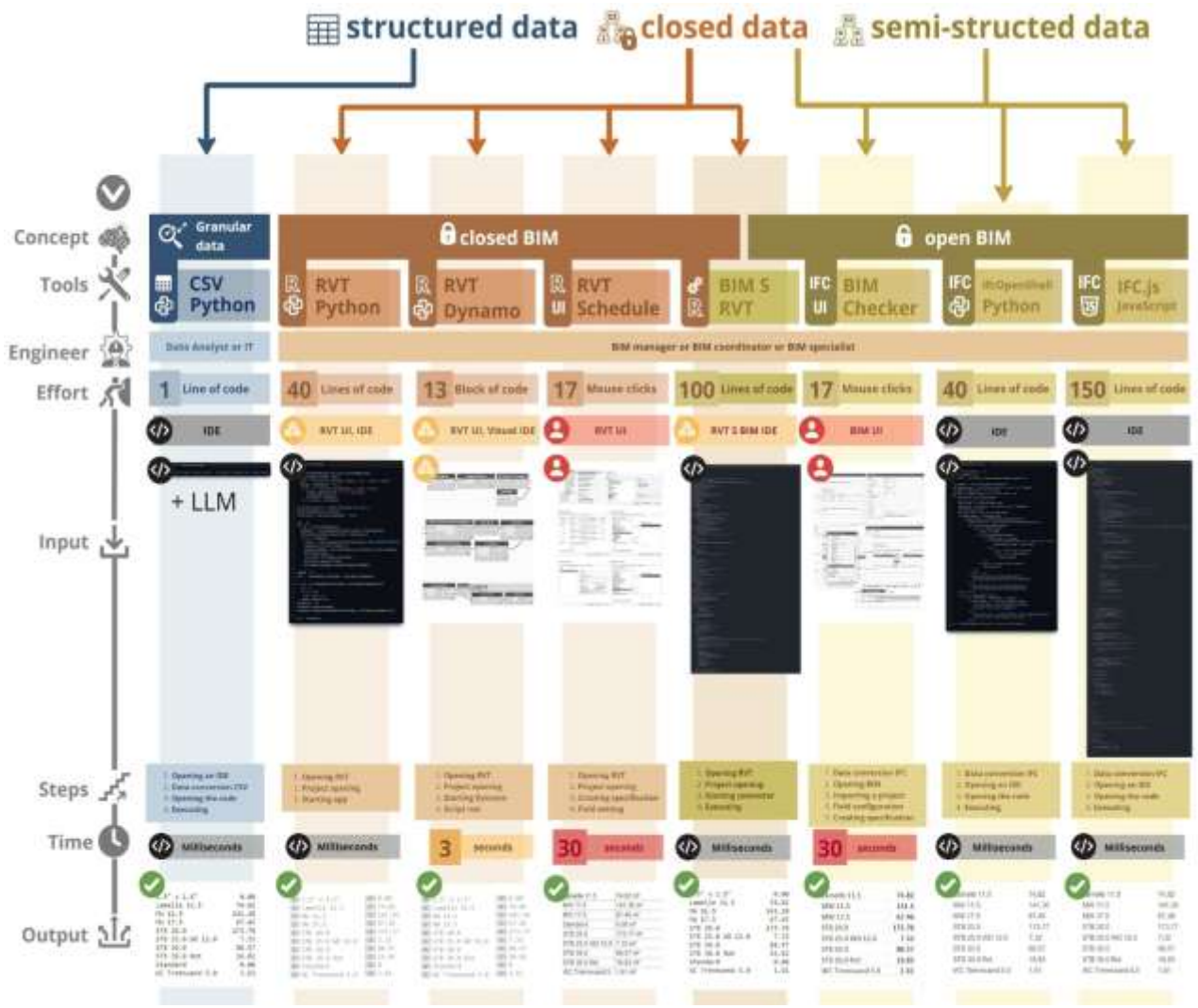


図 5.2-5 異なるツールは、プロジェクト・エンティティの属性表の形で同じ結果を出すが、人件費は異なる。

CAD（BIM）データの構造化形式への変換」の章で述べたように、CAD プロジェクトからオープンな構造化データを使用する場合、グループ化プロセスである QTO は大幅に簡素化されます。

オープンな構造化データの使用または CAD モデルデータベースへの直接アクセスに基づくアプロー チは、BIM（ ）という略語に関連するマーケティング上の制約から自由である。これらは、他の産業で長年使用されてきた実績のあるツールに依存している（図 7.3-10 ETL プロセス）。

> マッキンゼーの調査「オープンデータ」によるとストリーミング情報でイノベーションと生産性を引き出す"[102]によると、オープンデータを利用することで、発電施設の設計、エンジニアリング、調達、建設において、年間 300 億ドルから 500 億ドルのコスト削減の機会が生まれる可能性がある。これは、建設資本コストの 15%節約に相当する。

オープンな構造化（粒状）データで作業することで、情報の検索と処理が簡素化され、専門的な BIM プラットフォームへの依存が減り、専用システムや CAD フォーマットのパラメトリックで複雑なデータモデルを使用することなく、自動化への道が開かれる。

LLM および構造化データ を使用した QTO 自動化

構造化されていないデータを構造化された形に変換することは、様々なプロセスの効率を著しく改善する。前章で既に述べたように、データ処理を単純化し（図 4.1-1、図 4.1-2）、 要件を明確かつ透明化することによって検証プロセスをスピードアップする。同様に、CAD データ （BIM ）を構造化されたオープンフォームに変換すること（図 4.1-12、図 4.1-13）は、属性グルーピングプロセスと QTO プロセス を容易にします。

QTO 属性テーブルには構造化されたフォームがあり、構造化された CAD データを使用する場合、単一のデータモデル（図 5.2-5）で作業するため、プロジェクトのデータモデルやグループ化ルールを共通分母に変換・変換する必要がなくなります。これにより、1 行のコードで 1 つまたは複数の属性でデータをグループ化できます。対照的に、データが半構造化形式、パラメトリック形式、またはクローズド形式で保存されているオープン BIM やクローズド BIM では、処理に数十行から数百行のコードが必要であり、ジオメトリや属性情報と対話するために API を使用する必要があります。

- QTO 構造化プロジェクトを 1 つの属性でグループ化した例。LLM チャットルーム（ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN またはその他）でのテキストクエリ：

> CAD -project を DataFrame として持っているのですが、"Type " パラメータが "Type 1 値のみを含む項目を取得するように、プロジェクトデータをフィルタリングしてください。 ⏎

- LLM の答えは、Pandas を使った Python コード という形になる可能性が非常に高い：



図 5.2-6 LLM で書かれた 1 行のコードで、CAD プロジェクト全体を「Type」属性でグループ化し、目的の要素グループを取得することができます。

2D DataFrame のシンプルな構造のおかげで、LLM スキーマとデータモデルを説明する必要がありません。以前は、簡単なコードを書くにもプログラミング言語を学ぶ必要がありましたが、最新の言語モデル（LLM）により、テキストクエリを使用して構造化データを扱う際に、プロセスロジックを自動的にコードに変換できるようになりました。

> LLM 自動化と言語モデルは、CAD (BIM)データのグループ化と処理に携わる専門家が、テキストクエリを使用して問題を解決する機能を提供することで、プログラミング言語や BIM ツールを学ぶ必要性を完全に排除することができます。

CAD 環境（BIM ）では、インターフェイスを 17 回クリックするか、40 行のコードを書く必要がある同じクエリ（「壁」カテゴリーからすべてのプロジェクト要素をグループ化し、タイプごとに体積を計算する）でも、オープンなデータ処理ツール（SQL や Pandas など）では、シンプルで直感的なクエリに見えます：

- Pandas の 1 行で：

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

コード解読：df（DataFrame ）から、属性カラム "Category " の値が "OST_Walls " である要素を取り出し、得られたすべての要素を属性カラム "Type " でグループ化し、得られた要素グループの属性 "Volume " を合計する。

■ SQL を使用して CAD から取得した構造化プロジェクトをグループ化する ：

```
SELECT タイプ, SUM(Volume) AS TotalVolume
FROM 要素
WHERE Category = 'OST_Walls'
GROUP BY タイプ ;
```

■ LLM、単純なテキスト参照、つまりプロンプトとして、プロジェ クト・データベースにグループ化のリクエストを書き込むことができる（図 5.2-7）：

```
プロジェクトのデータフレームについて、'Type' パラメータで項目をグループ化する。ただし、'Category' パラメータが 'OST_Walls' または 'OST_Columns' である項目についてのみ化する。 ⏎
```

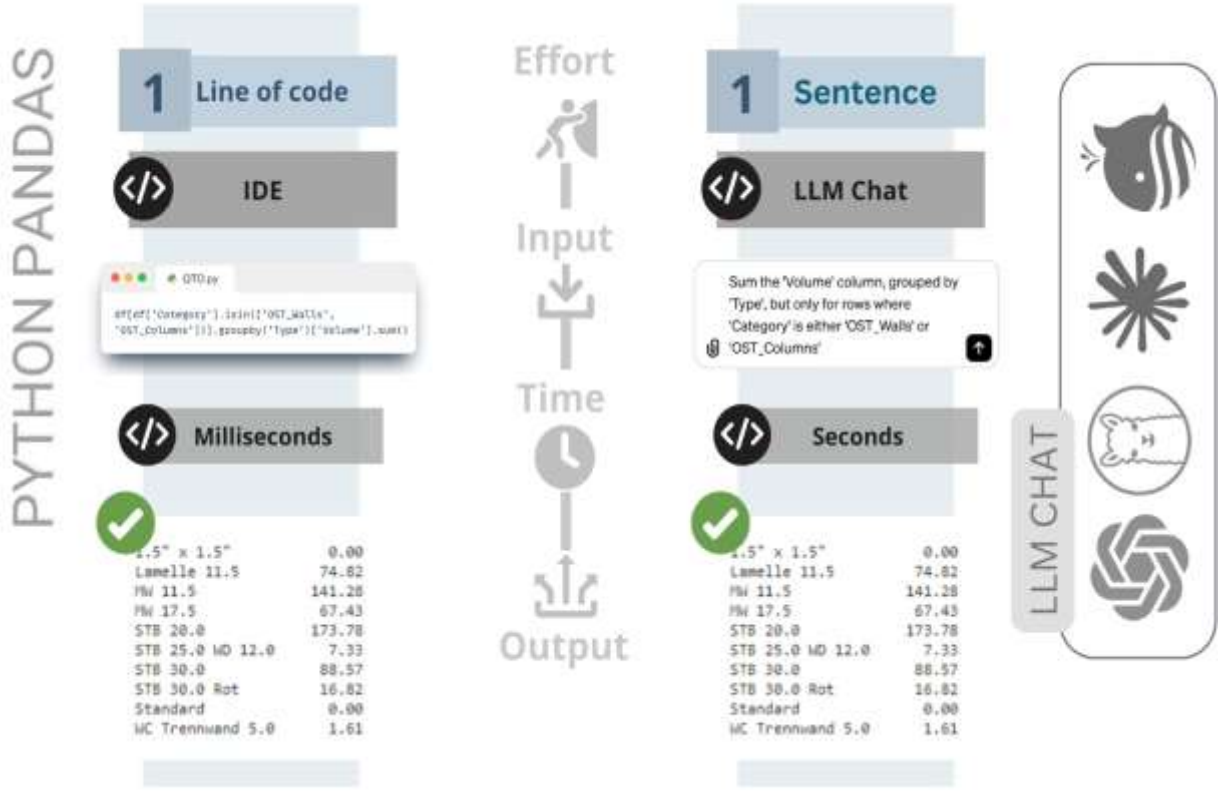


図 5.2-7 SQL、Pandas および LLM を使うことで、数行のコードとテキストクエリでデータ処理の自動化が可能になった。

LLM ツール（ChatGPT 、LLaMa、Mistral 、Claude 、DeepSeek 、QWEN 、Grok ）を使用して CAD データから QTO を取得することで、個々のオブジェクトやオブジェクトグループの属性情報、定量データ、体積データを抽出する従来の方法が劇的に変化します。

> 現在では、設計に関する深い知識を持たず、専門的な CAD ソフトウェア - (BIM-) ベンダーを持っていないプロジェクト・マネージャー、原価計算、ロジスティクスの専門家でも、CAD データベースにアクセスすれば、単にクエリーを書くか口述するだけで、数秒のうちに壁やその他のオブジェクトのカテゴリーの要素の総量を得ることができる。

テキストクエリ(Fig. 5.2-8)において、モデルの LLM エージェント、テーブルの 1 つ以上のパラメータ(カラム)に特定の関数を適用するユーザの要求を処理します。その結果、LLM と通信しているユーザは新しい値を持つ新しい列パラメータか、グループ化された後の特定の値を受け取ります

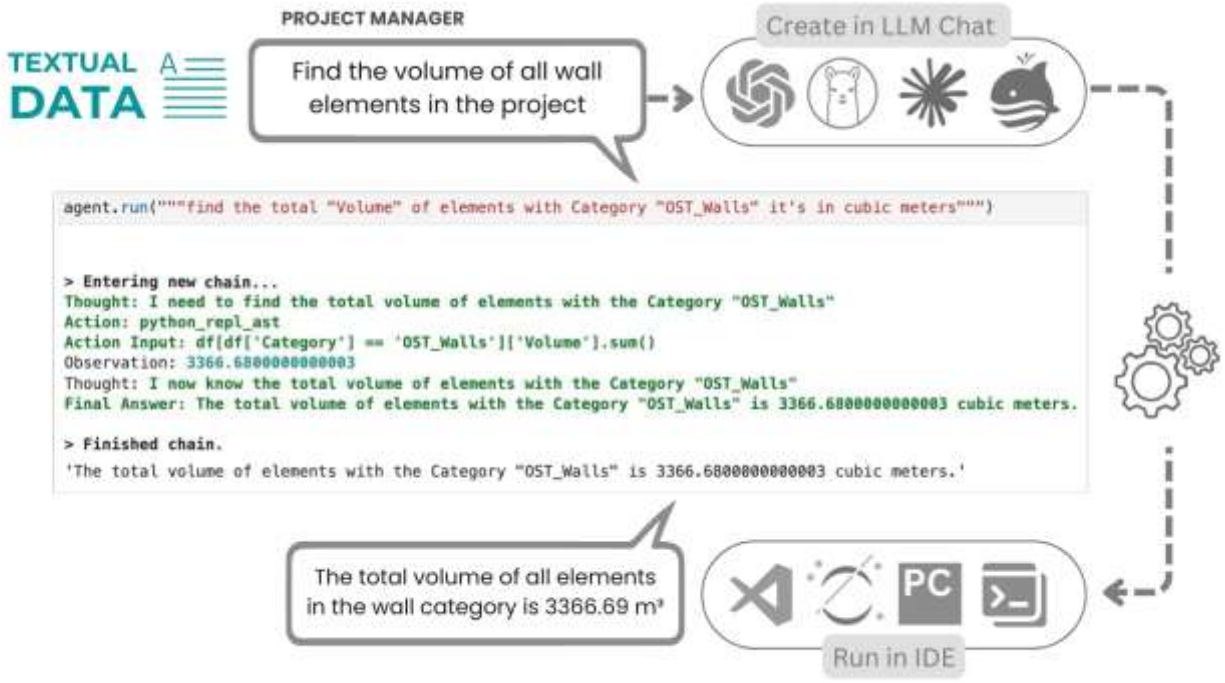


図 5.2-8 LLM 構造化データを扱うモデルは、テキストクエリのコンテキストから、ユーザーがどのようなグループ化と属性について質問しているかを理解します。

要素の 1 つのグループのみの数量を取得する必要がある場合は、CAD モデルデータに対して単純な QTO クエリ（図 5.2-7）を実行するだけで十分です。しかし、多くの要素グループからなるプロジェクト全体の予算や見積もりを計算する場合、多くの場合、すべてのタイプの要素（クラス 0）の数量を抽出する必要があり、要素の各カテゴリは関連する属性でグループ化され、別々に処理されます。

見積り担当者や査定担当者の実務では、異なるタイプのオブジェクトに対して個別のグループ化や計算ルールが使用されます。例えば、窓は通常、フロアやゾーンごとにグループ化され（グループ化パラメータ - 属性 Level, Rooms）、壁は材料や構造タイプごとにグループ化されます（パラメータ Material, Type）。グループ化プロセスを自動化するために、このようなルールは、 、グループ化ルールテーブルの形で事前に記述されます。これらのテーブルは、プロジェクト内の各要素グループの計算でどの属性を使用するかを定義する設定テンプレートとして機能します。

QTO Excel スプレッドシートからグループルールを使用してプロジェクト全体の計算を行う。

実際の建設プロジェクトでは、1 つの要素グループ内で複数の属性による集計を同時に実行する必要があることがよくあります。例えば、カテゴリ「Windows」（Category 属性に OST_Windows や IfcWindows のような値が含まれる）で作業する場合、要素をタイプ別（例えば、Type Name や Type フィールドの値）だけでなく、対応する属性で指定された熱伝導率レベルのような追加特性によってもグループ化することができます。この多次元的なグループ化により、特定のグループに対してより正確な結果を得ることができます。同様に、壁や床のカテゴリーを計算する場合、材料、レベル、床、耐火性、その他のパラメータなど、属性の任意の組み合わせをフィルタやグループ化の基準として使用することができます（図 5.2-9）。

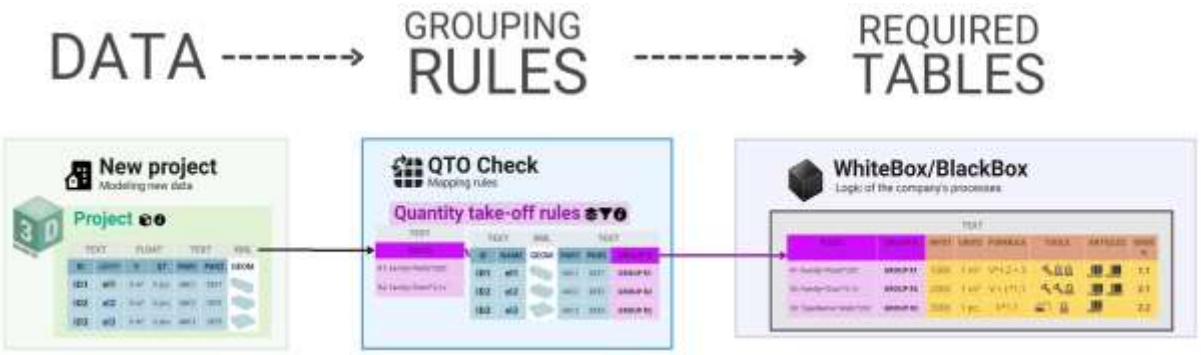


図 5.2-9 プロジェクト内のエンティティのグループやカテゴリーごとに、1 つ以上の基準からなる異なるグループ化の公式があります。

このようなグルーピング・ルールを定義するプロセスは、データ・モデルの作業について詳しく説明した「要件の作成と品質チェック データ」（図 4.4-5）の章で説明したデータ要件の作成プロセスと同様である。このようなグループ化ルールや計算ルールは、計算や算出で考慮しなければならない必要な条件をすべて考慮して、エンティティのカテゴリの量や体積の合計属性を自動的に算出するために、結果の正確性と妥当性を保証する。

- 以下のコード・サンプルは、"Category" 属性カラムに "OST_Windows" または "IfcWindows" という値が含まれ、同時に "Type" 属性カラムに "Type 1" という値が含まれるエンティティのみが、結果のデータセットに含まれるように、projects テーブルをフィルタリングします：

> DataFrame プロジェクトで、"Category "属性に "OST_Windows "または "IfcWindows "という値が含まれ、同時に "Type "属性に "Type 1 値が含まれる項目だけがデータセットに残るように、データをフィルタリングする。 ⏎

LLM の回答 ：



図 5.2-10 エクセルの計算式 に似た一行のコードで、すべてのプロジェクト・エンティティをいくつかの属性でグループ化することができます。

構造化されたオープンフォーマット（図 4.1-13）の CAD データ を変換した後のコード（図 5.2-10）は、上で述べた一般的な IDE（統合開発環境） でオフラインモードで実行できます。1-13)は、上で述べた一般的な IDE(統合開発環境)でオフラインモードで実行できます: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse with PyDev plugin, Thonny, Wing IDE, IntelliJ IDEA with Python plugin, JupyterLab または一般的なオンラインツール: Kaggle.com、Google Collab、Microsoft Azure Notebooks、Amazon SageMaker。

QTO DataFrame フォーム 、特定の熱伝導率値を持つ "Windows "カテゴリのプロジェクトエンティティのみを検索するには、LLM ：

> DataFrame プロジェクトで、"Category "に "OST_Windows "または "IfcWindows "の値を含むレコードだけがデータセットに残るようにデータをフィルタリングし、同時に ThermalConductivity カラムの値が.5 になるようにしています。 ⏎

LLM の回答 ：



図 5.2-11 非常にシンプルな Pandas クエリ言語 Python では、同時にいくつものプロジェクトに対して QTO を実行することができます。

LLM から受信した応答（図 5.2-11）では、論理条件"&"を使用して、熱伝導率値と 2 つのカテゴリのいずれかに属するという 2 つの条件を組み合わせている。isin "メソッドは、属性列 "Category "の値が提供されたリストに含まれているかどうかをチェックする。

異なるグループ化ロジックを持つ多数の要素グループを持つプロジェクトでは、プロジェクト実体の各カテゴリー（例：窓、ドア、スラブ）ごとに個別のグループ化ルールを確立する必要があり、これには追加の係数または総属性計算式が含まれる場合がある。これらの計算式（図 5.2-12 属性 "formel"、例：x-量の値、y-グループの体積）と係数は、例えば、各グループの特有な特性を考慮する：

- オーバーランを考慮した材料量への%加算
- 定量追加
- 想定されるリスクや計算ミスに関連した計算式の調整

フィルタリングとグループ化のルールが各項目分類のパラメータ式の形で定式化されたら、それを行 ごとの表として、例えばエクセル形式で保存することができます（図 5.2-12）。これらのルールを構造化された形で保存することで、プロジェクトデータの抽出、フ ィルタリング、グループ化のプロセスを完全に自動化することができます。手作業で多くのクエリーを書く代わりに、システムはパラメータテーブルを読み込み、適切なルールをモデル（プロジェクト全体のデータフレーム（図 4.1-13））に適用するだけで、プロジェクト要素のカテゴリーごとに最終的な QTO テーブルを生成します。



GROUPING RULES

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

図 5.2-12 QTO 属性グループ化表 は、プロジェクト要素をグループ化するためのルールを確立し、各カテゴリの正確な総数と範囲を提供する。

収集されたルールにより、プロジェクト全体をグループ化し、ボリューム属性の調整など、必要な計算をすべて実行することができます。その結果、ボリュームは、CAD モデル の設計段階に元々あったものではなく、計算や計算に使用される「実際のボリューム」にもたらされます。

> プロジェクト全体の QTO ボリューム・テーブルを自動的に作成するプロセスでは、アプリケーションはグループ化ルール・テーブルのすべてのカテゴリーを調べ、グループ化属性を取り出し、すべてのプロジェクト要素をその属性でグループ化し、そのグループのボリューム属性を集計し、さらに精製係数または係数を掛けなければなりません。

LLM に、このようなソリューションのためのコードを書いてもらうことにしよう。このコードでは、2 つのテーブル（グループ化ルールのテーブル（図 5.2-12）とプロジェクト自体のデータのテーブル（図 4.1-13）を読み込み、グループ化ルールを適用し、与えられたルールに従って項目をグループ化し、集計値を計算し、結果を新しい Excel ファイル に保存しなければならない。

- LLM チャットにテキストリクエストを送信する：

> ファイル 'basic_sample_project.xlsx' からプロジェクトデータを読み込み、'Grouping_rules_QTO.xlsx' からルールを作成して、すべてのデータを 'Parameter 1' と 'Parameter 2' でグループ化し、'Aggregate Parameter' で集計し、'Expression2' でフィルタリングし、'Formel1' から計算を実行し、QTO テーブルを 'QTQ_table2.xlsx に保存するコードが必要です。 ⏎

LLM の回答 ：



図 5.2-13 Python コード Pandas CAD プロジェクトからの設計データを処理し、Excel で定義されたルールに従ってグループ化します。

コード実行の最終結果（図 5.2-13）は、元の CAD- （BIM- ）モデルからの一般化されたボリューム属性だけでなく、 計算と見積もりを正しく作成するためのすべての 要件を考慮した新しい実ボリューム属性を含むエンティティグループテーブルとなります（例 図 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 Partitio	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

図 5.2-14 "After Calculation "属性は、実際の体積を自動的に計算するコードが実行された後、サマリー表に追加される。

出来上がったコード（図 5.2-13）は、一般的な IDE（前述した）の一つで実行することができ、既存または新規の着信プロジェクト（RVT 、IFC 、DWG 、NWS、DGN など）に適用することができます（図 5.2-15）。数個のプロジェクトでも、構造化されたさまざまな形式の数百のプロジェクトでもかまいません。

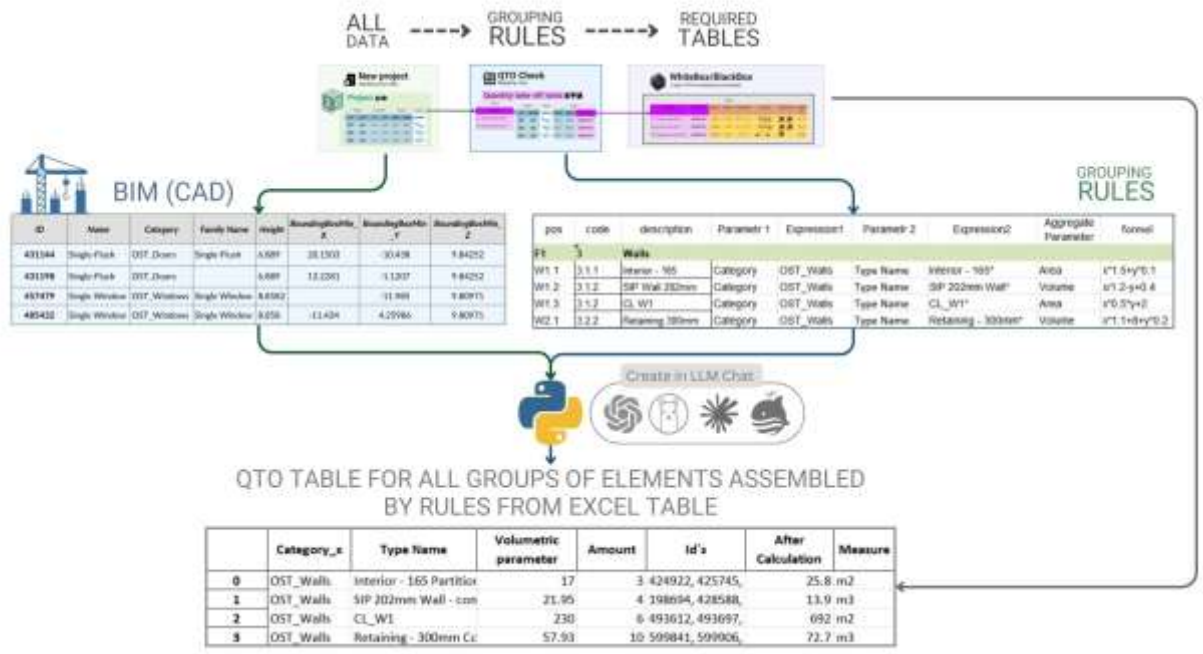
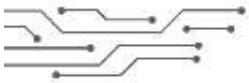


図 5.2-15 BIM データ (CAD) と QTO テーブル を、Excel スプレッドシート からのルールでリンクする。

カスタマイズされ、パラメータ化された容積データ収集プロセス（図 5.2-15）は、コスト見積もり、ロジスティクス、作業スケジュール、カーボン・フットプリント計算、その他の分析作業を含む、プロジェクト要素の更なる作業のための定量的属性と容積に関するデータの完全自動収集を可能にする。

プロジェクト要素のグループを特定の属性に従って簡単に整理し、グループ化できるツールを学んだので、グループ化されフィルタリングされたプロジェクトを、さまざまな会社の計算やビジネスシナリオと統合する準備が整った。



章 5.3.

4d, 6d -8d および二酸化炭素排出量の計算 co_2

4D モデル：建設見積もりに時間を組み込む

原価計算に加えて、建設における設計データの重要な用途の一つは、個々の建設作業とプロジェクト全体の両方について、時間パラメータを決定することである。前章「建設プロジェクトの計算と見積もり」で詳しく説明したリソースベースの見積もり方法と関連する計算データベースは、時間の自動計算と作業実行のスケジュール作成の基礎としてよく使用されます。

資源ベースのアプローチでは、材料費だけでなく時間資源も考慮する。原価計算では、各プロセスに作業指示属性（図 5.3-1-作業指示パラメータ）と、そのプロセスの実行に関連する時間とコストの量を割り当てることができる。これらのパラメータは、固定市場価格を持たず、直接購入できない業務、例えば、建設機械の使用、労働者の雇用、または物流プロセス（これらは通常、時間で表現される）を記述するために特に重要である。このような場合、コストは、調達部門によって決定されるのではなく、実施企業が社内の規範や生産率に基づいて直接決定する（図 5.3-1）。

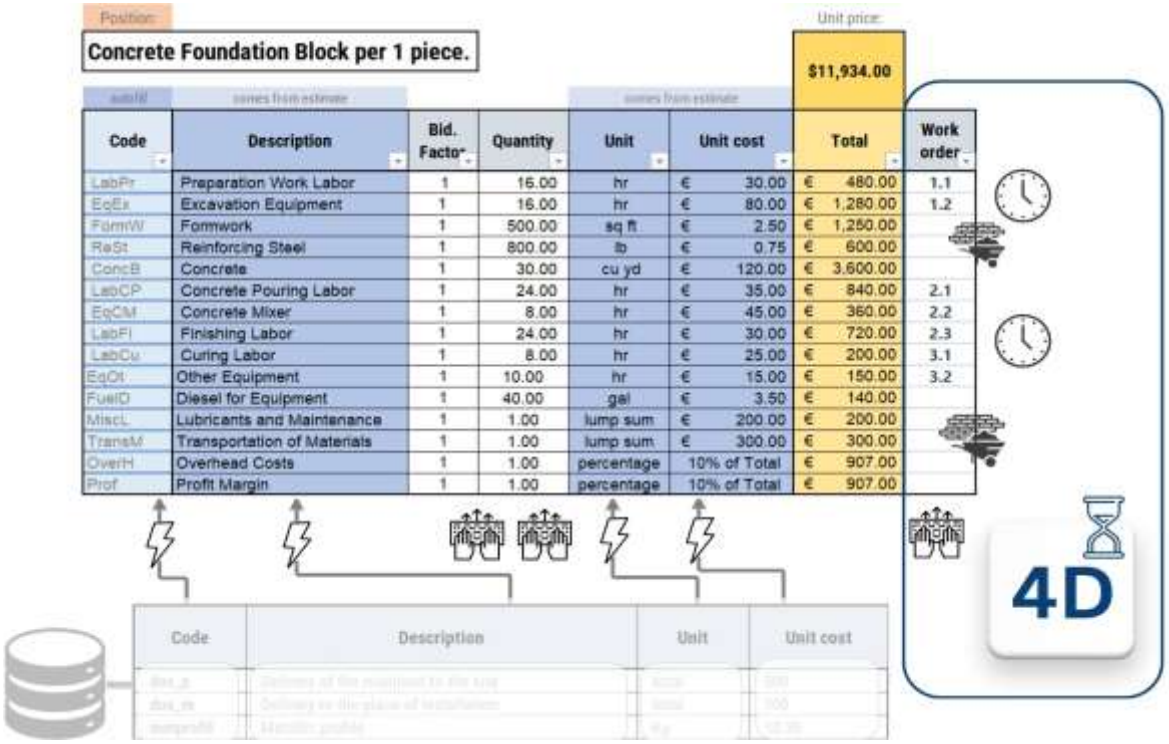


図 5.3-1 リソースベースの見積もり方法における作業の計算には、労働時間のタイミングが含まれる。

したがって、原価計算レベルの計算には、燃料費、材料費（購入費）だけでなく、運転手、技術者、補助作業員が現場で費やした時間も含まれる。例示された例（図 5.3-1）では、コスト表は、基礎ブロックを設置するコストの計算であり、準備作業、躯体設置、コンクリート打設といった作業の構成段階と、必要な材料と労務が含まれている。しかし、準備作業のような個々の作業では、材料費はかからないが、工数で表される一時的な人件費が多く含まれる場合がある。

建設現場での作業順序（作業スケジュール）を計画するために、計算表（図 5.3-1）に「作業順序」という属性を手作業で追加する。この属性は、単位が時間（時間、日）である項目に対してのみ、追加列で指定される。この属性は、作業コード、説明、数量、測定単位（パラメータ "Unit"）、コストに追加される。アクティビティの数値順序（パラメータ「作業順序」）により、建設現場でのタスクの実行順序を設定し、スケジューリングに使用することができます。

原価計算データに基づく建設スケジュールとその自動化

工事スケジュールは、プロジェクト実施の一環として実施される作業計画と工程を視覚的に表したものである

。これは、詳細なリソース計算（図 5.3-1）に基づいて作成され、各タスク-ジョブは、リソースのコストに加えて、時間と順序によってスケジューリングされます。

時間の見積もりは、材料や設備の設置にかかる典型的な時間に基づいて行われる平均的アプローチとは対照的に、資源ベース法では、原価計算における実際のデータに基づいて計画が行われる。人件費に関連する見積りの各項目は、作業期間中の資源使用の実態を考慮した適用カレンダーに基づいている。原価計算レベルでの係数による生産時間の調整（図 5.3-1 パラメータ「Bid.Factor」）により、生産性の違いや、作業のタイミングに影響する季節的な特殊性を考慮することが可能になる。

ガントチャートの工程表の工程開始日と工程終了日を決定するために、基礎ブロック積算から各項目の時間属性の値を取り出し、ブロック数（この場合はコンクリート基礎ブロックの数）を掛けます。この計算により、各作業の期間が得られる。そして、これらの期間をプロジェクト開始日を起点としたタイムライン上にプロットしてグラフを作成し、各タスクの開始と終了のタイミングを視覚的に表現します。プロセスの "作業順序 "パラメーターにより、作業プロセスが並列（"作業順序 "例：1.1-1.1）か、順次（1.1-1.2）かを理解することができます。

> ガントチャート は、計画とプロジェクト管理のためのグラフィカルなツールで、タスクをタイムライン上の横棒で表します。各バーはタスクの期間、開始と終了を表します。

スケジュール（ガントチャート）は、プロジェクトマネージャーと作業員が、建設の様々なフェーズをいつ、どのような順序で行うべきかを明確に理解するのに役立ち、リソースの効率的な使用と納期の遵守を保証します。

上表の計算を使用して、3 つのコンクリート基礎ブロックの設置スケジュールを示そう。上の例のコスト表(図 5.3-1)を使って、LLM 、3 つの基礎ブロック要素の据付を、例えば 2024 年 5 月 1 日□□□予定するよう、LLM 。

LLM に原価計算を送信するには、原価計算表を XLSX 形式 でアップロードするか、単純に原価計算の JPEG 画像のスクリーンショットを LLM のチャットに直接挿入します（図 5.3-1）。LLM は、独自に表画像を視覚化するライブラリを見つけ、表からアクティビティの時間属性にそのボリュームを乗じることで、すべてのデータをグラフに追加します。

- LLM にテキストリクエストを送信する：

> 3 つの基礎ブロックの据付について、時間関連表（付録の JPEG として図 5.3-1）の値を用いて、ガントチャートの形式で作業スケジュールを作成する。各ブロックの作業は順次行う。作業開始は 2024 年 5 月 1 日とする。 ⏎

LLM の回答

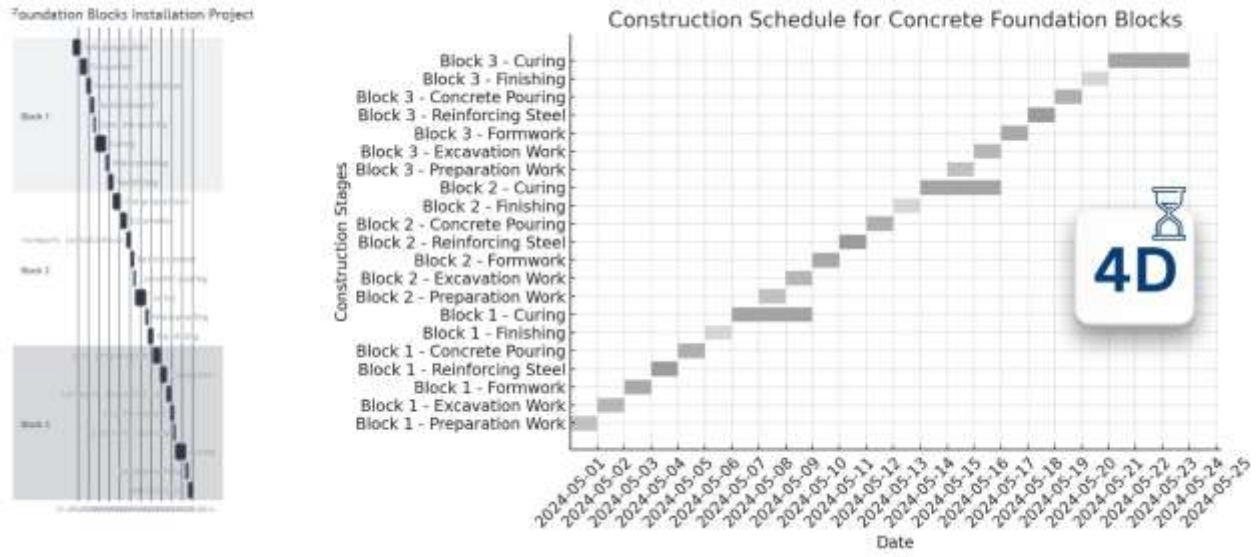


図 5.3-2 複数の LLM によって自動的に生成される ガントチャートは、プロンプトからの条件に従って、3 つのコンクリートブロックの建設段階を示す。

結果として得られるグラフ（図 5.3-2）は時間図であり、各横棒は基礎ブロックの作業のある段階に対応し、準備、掘削、型枠設置、補強、コンクリート打設、仕上げなどの一連の作業（パラメータ「作業順序」）を示している 、すなわち、計算において時間パラメータと順序が記入された工程を示している。

このようなスケジュール(図 5.3-2)は、作業日、シフト、作業時間基準に関する制約を考慮せず、プロセスの概念的な視覚化のみを目的としている。作業の並行性を反映する正確なスケジュールは、チャットルーム内の適切なプロムトや追加指示で補足することができます

3D -ジオメトリのボリューム属性により、単一のコスト計算（図 5.3-1）を使用して、自動見積もりによるプロジェクト・コストの見積もりと、同時に、さまざまなプロジェクト・バリエーション（図 5.3-3）について、グループの時間特性を表やグラフの形で計算することができます。

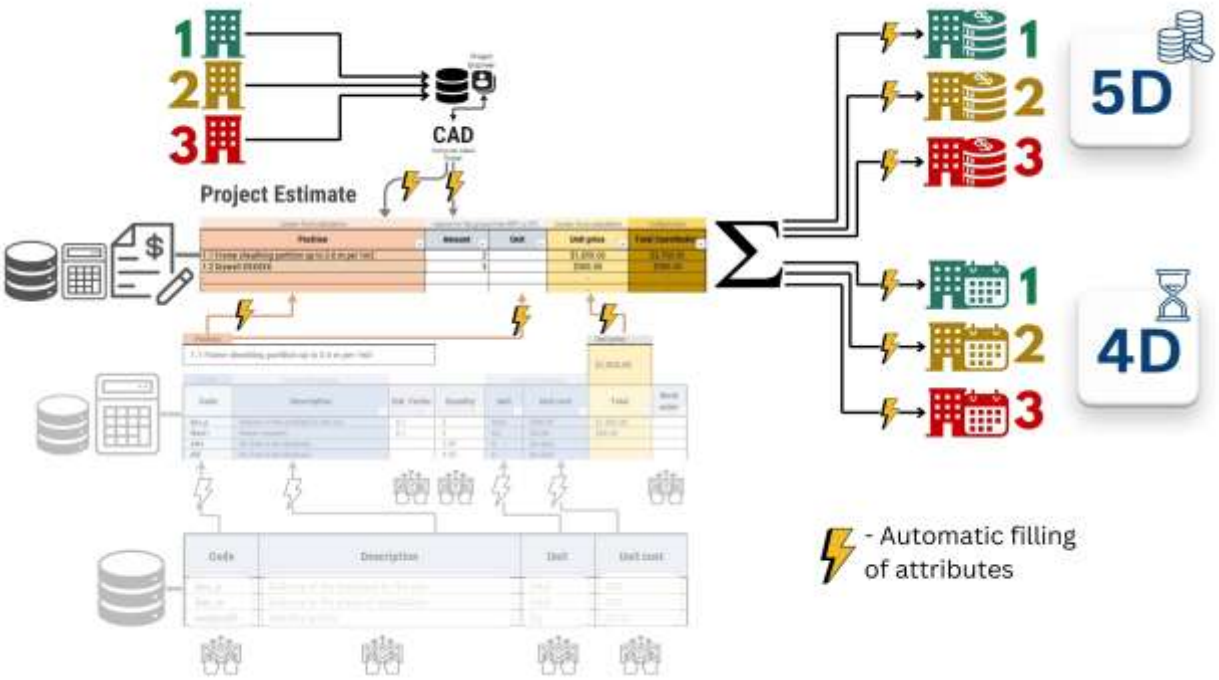


図 5.3-3 自動計算では、さまざまなプロジェクト・オプションについて、即座に自動的なコストと時間の予測を行うことができます。

最新のモジュール式 ERP -システム（図 5.4-4） CAD モデルからデータをロードする場合、意思決定プロセスを大幅に短縮する、同様の自動時間計算方法を使用します。これにより、作業スケジュールを即座に正確に計画し、実勢価格を考慮した上で、プロジェクト実現におけるすべてのタスクの完了に必要な総時間を計算することができます。

拡張属性レイヤー 6D -8D: エネルギー効率から安全保証まで

6D 、7D 、8D は、情報モデリングの拡張レベルであり、それぞれ包括的なプロジェ クト情報モデルに属性の追加レイヤーを提供する。各追加レイヤーは、例えば、不動産管理システム（PMS ）、コンピュータ支援設備管理（CAFM ）、建設プロジェクト管理（CPM ）、安全管理システム（SMS）のような他のシステムで、さらなるグループ化やさらなる識別のために必要とされる特定のパラメータを導入します。



図 5.3-4 データ情報モデルの属性 6D 、7D 、8D は、エネルギー効率から安全性まで、プロジェクトの様々な側面の検討を拡張する。

- 6D では、要素の幾何学的および容積的属性を持つプロジェクトデータベース（またはデータフレーム（図 4.1-13））に加えて、環境持続可能性に関する情報（属性列）が追加される。これには、エネルギー効率、カーボン・フットプリント、材料のリサイクル性、環境に優しい技術の使用に関する情報が含まれる。このデータにより、プロジェクトの環境への影響を評価し、プロジェクトの意思決定を最適化し、持続可能な開発目標（ESG）を達成することができる。
- 7D 属性は、建物の保守管理に必要な属性を補完する。これらは、メンテナンススケジュール、コンポーネントのライフサイクル、技術文書、修理履歴に関するデータである。この一連の情報により、モデルをメンテナンスシステム（CAFM、AMS）と統合し、メンテナンスの効率的な計画、設備の交換を可能にし、施設のライフサイクル全体を通してサポートを提供することができます。
- 8D 追加の属性レイヤーには、建設段階およびその後の操業中の安全に関する情報が含まれています。このモデルには、人員の安全対策、緊急時の指示、避難システム、防火要件などが含まれる。このデータをデジタルモデルに統合することで、事前にリスクを考慮し、安全衛生要件を考慮した建築、エンジニアリング、組織的ソリューションを開発することができます。

構造化された表形式では、レイヤー4D からレイヤー8D は、すでに入力されている 3D -名称、カテゴリー、タイプ、容積特性などのモデル属性に追加された、入力された値を持つ列（図 5.3-5）の形の追加属性を表す。属性レイヤー 6D 、7D および 8D の値には、リサイクル率、カーボンフットプリント、保証期間、交換サイクル、設置日、安全プロト コルなどの追加テキストデータおよび数値データが含まれる。



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

図 5.3-5 6D -8D データ情報モデルに属性レイヤーを追加する。このデータ情報モデルには、3D - モデルの幾何学的属性と体積属性がすでに含まれている。

新しいウィンドウ（図 4.4-1）の場合、識別子 W-NEW（図 5.3-5）を持つ要素は以下の 3D -8D 属性を持つことができます：

3D -attributes - CAD システムから得られる幾何学的情報：

- "タイプ名" - 要素 "Window"
- 幅" - 120 cm
- さらに、要素またはその「ジオメトリ BREP / MESH 」の「バウンディングボックス 」点を別の属性として追加できます。

6D の属性 - 環境の持続可能性：

- リサイクル率 90
- 「カーボンフットプリント - 1,622 kg CO_2

属性 7D - オブジェクト管理データ：

- 「保証期間" - 8 年
- 「リプレイスメント・サイクル」は 20 歳
- 「メンテナンス」-毎年必要

8D の属性 - 建物の安全な使用と運用を確保する：

- XYZ Windows "社による "Installed "ウィンドウ
- 「安全基準」 - ISO 45001 に準拠

データベースやデータセット（図 5.3-5）に記録されたすべてのパラメータは、グループ化、検索、計算のために、異なる部門の専門家によって必要とされる。このような多次元的な属性に基づくプロジェクト対象物の記述は、そのライフサイクルの全体像、運用要件、およびプロジェクトの設計、建設、運用に必要な他の多くの側面を提供する。

CO_2 を推定し、建設プロジェクトからの二酸化炭素排出量を計算する。

段階 6D （図 5.3-5）における建設プロジェクトの持続可能性の話題と並行して、現代の建設は、プロジェクトの環境持続可能性に焦点を当てており、重要な側面の一つは、プロジェクトのライフサイクル段階（例えば、製造と設置）で発生する二酸化炭素 CO_2 排出量の評価と最小化になる。

> 建築資材の炭素排出量の推定と算定は、プロジェクトで使用される要素または要素グループの体積属性に、そのカテゴリーに適した炭素排出係数を乗じることによって、総炭素排出量を決定するプロセスである。

建設プロジェクトの評価において、より広範な ESG 基準（環境、社会、ガバナンス）の一部として炭素排出を考慮することは、分析に新たなレベルの複雑さを加えることになる。これは、LEED® (Leadership in Energy and Environmental Design)、BREEAM® (Building Research Establishment Environmental Assessment Method)、DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen)などの関連認証を取得する際に、クライアントである投資家にとって特に重要である。これらの認証のいずれかを取得することで、物件の市場性を大幅に高め、試運転を簡素化し、持続可能性を重視するテナント（ESG）の要求事項への適合を確保することができる。プロジェクトの要件によっては、HQE（Haute Qualité Environnementale、フランスのグリーンビルディング基準）、WELL（WELL Building Standard、利用者の健康と快適性に焦点を当てた基準）、GRESB（Global Real Estate Sustainability Benchmark）も利用できる

> 環境・社会・ガバナンス ESG (environ-mental, social and governance)は、コーポレート・ガバナンス、社会、環境への影響を社内外から評価するための広範な原則である。

ESG は、もともと 2000 年代初頭に金融ファンドが投資家に広範な環境・社会・ガバナンス基準に関する情報を提供するために開発したもので、建設プロジェクトを含む企業とプロジェクトの両方を評価するための重要な指標へと発展してきた。大手コンサルティング会社の調査によると、環境・社会・ガバナンス（ESG）への配慮は、建設業界にとって不可欠な要素になりつつある。

> EY（2023 年）の「カーボン・ニュートラルへの道」によると、ESG -原則を積極的に導入している企業は、長期的なリスクを削減するだけでなく、ビジネスモデルの効率性も高めており、これは市場のグローバルな変革において特に重要である[103]。PwC の ESG 意識調査報告書によると、ESG -要因の重要性に対する企業の意識は 67%から 97%であり、ほとんどの組織がこれらの傾向を将来の持続可能性の鍵と見なしている[104]。

> したがって、建設プロジェクトに ESG の原則を取り入れることは、LEED 、BREEAM、DGNB といった国際的な持続可能性認証の取得に貢献するだけでなく、業界の企業の長期的な持続可能性と競争力を確保することにもなる。

建設プロジェクトの全体的なカーボンフットプリントに影響を与える最も重要な要因の一つは、建設資材や部品の生産・物流段階である。現場で使用される資材は、多くの場合、総 CO_2排出量に決定的な影響を与える 、特にプロジェクトのライフサイクルの初期段階（原材料の採取から建設現場への搬入まで）において。

建築要素のカテゴリーやタイプ別に排出量を計算するには、異なる材料の生産から発生する CO_2の量を反映した参照炭素排出係数を使用する必要がある。このような材料には、コンクリート、レンガ、再生鋼、アルミなどが含まれる。これらの値は一般的に、UK ICE 2015（Inventory of Carbon and Energy）や US EPA 2006（U.S. Environmental Protection Agency）のような評判の良い情報源や国際的なデータベースから抽出される[105]。以下の表（図 5.3-6）は、一般的な建築材料のベースライン排出係数をまとめたものである。各材料について、比 CO_2排出量（材料 1kg 当たりキログラム）と体積-重量換算係数（1 立方メートル当たりキログラム）の 2 つの重要なパラメータが提供されており、これは計算を設計モデルに統合し、QTO データグループとリンクさせるために必要である。



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

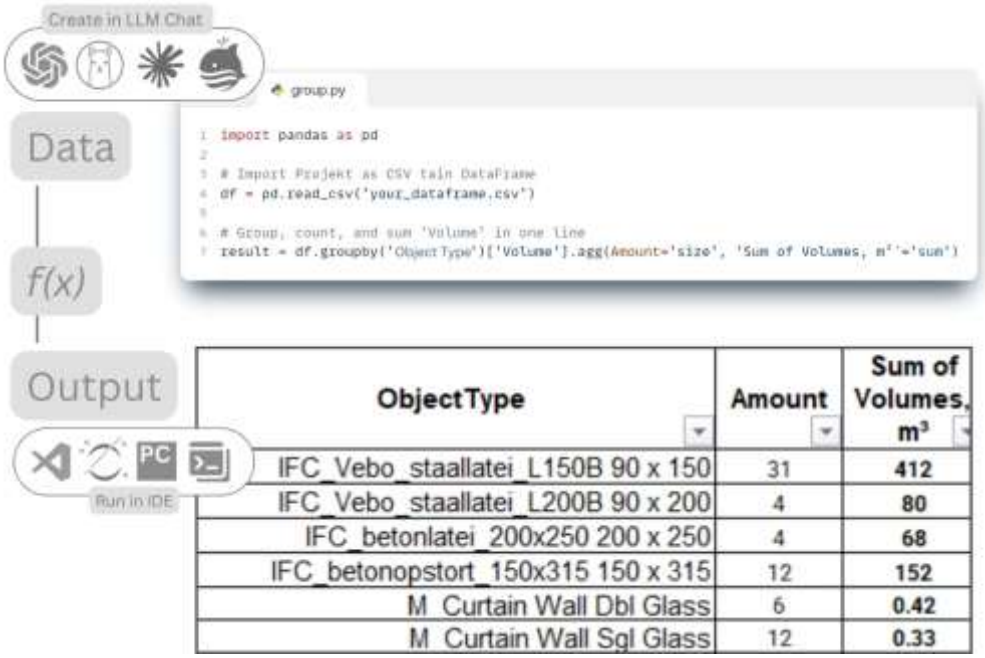
図 5.3-6 英国 ICE および米国 EPA データベースによる、さまざまな建築材料の生産において排出される炭素の量。

プロジェクトの CO_2 総排出量を計算するためには、4D や 5D 計算と同様に、各施設グループの属性量を決定する必要がある。これは、定量的分析ツール(QTO)を用いて行うことができる。定量的分析ツールの詳細については、 Quantity take-off の項を参照。これらの体積に、各物質グループの "CO_2 プロセス排出 "属性の適切な係数を乗じる。

- 前章ですでに行ったように、すべてのプロジェクトデータをグループ化して、CAD (BIM) プロジェクトから要素タイプ別の体積表を自動的に抽出してみましょう。このタスクを実行するために、LLM を参照しましょう。

> CAD (BIMDataFrame テーブル、列パラメータ「オブジェクト名」（または「タイプ」）でグループ化し、各グループのアイテム数を表示し、タイプ内の全アイテムの「体積」パラメータをまとめてください。 ⏎

LLM の回答 ：



ObjectType	Amount	Sum of Volumes, m^3
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

図 5.3-7 LLM の生成コード は、プロジェクト・エンティティをタイプ（ObjectType）別にグループ化し、「Volume」という属性でまとめている。

プロジェクト全体の総 CO_2 排出量の計算を自動化するには、テーブル内の自動データマッピングを設定するか、排出係数テーブルから要素タイプ（図 5.3-7）と対応する材料タイプ（図 5.3-6）を手動で関連付けるだけです。排出係数と計算式を含む完成した表、および CAD フォーマットから体積を取得するコード (BIM)、および CO_2決定を自動化するコードは、"CO_2_calculating-the-embodied-carbon." で検索して GitHub で見つけることができる。DataDrivenConstruction " で検索できる。[106].

このように、 CAD データベース から QTO 要素をグループ化した後のデータ統合により、さまざまな設計オプションに対する二酸化炭素排出量の自動計算（図 5.3-8）が可能になる。これにより、さまざまなバリエーションにおけるさまざまな材料の影響を分析し、顧客の CO_2排出量要件（ ）を満たすソリューションのみを選択し、建物の試運転時に特定の証明書を取得することが可能になる

> グループ化されたプロジェクト要素の量に係数を乗じることで、CO_2排出量 を見積もることは、建設会社が施設の ESG 格付け （例えば、LEED 認証 ）を取得する過程における作業の典型的な例である。

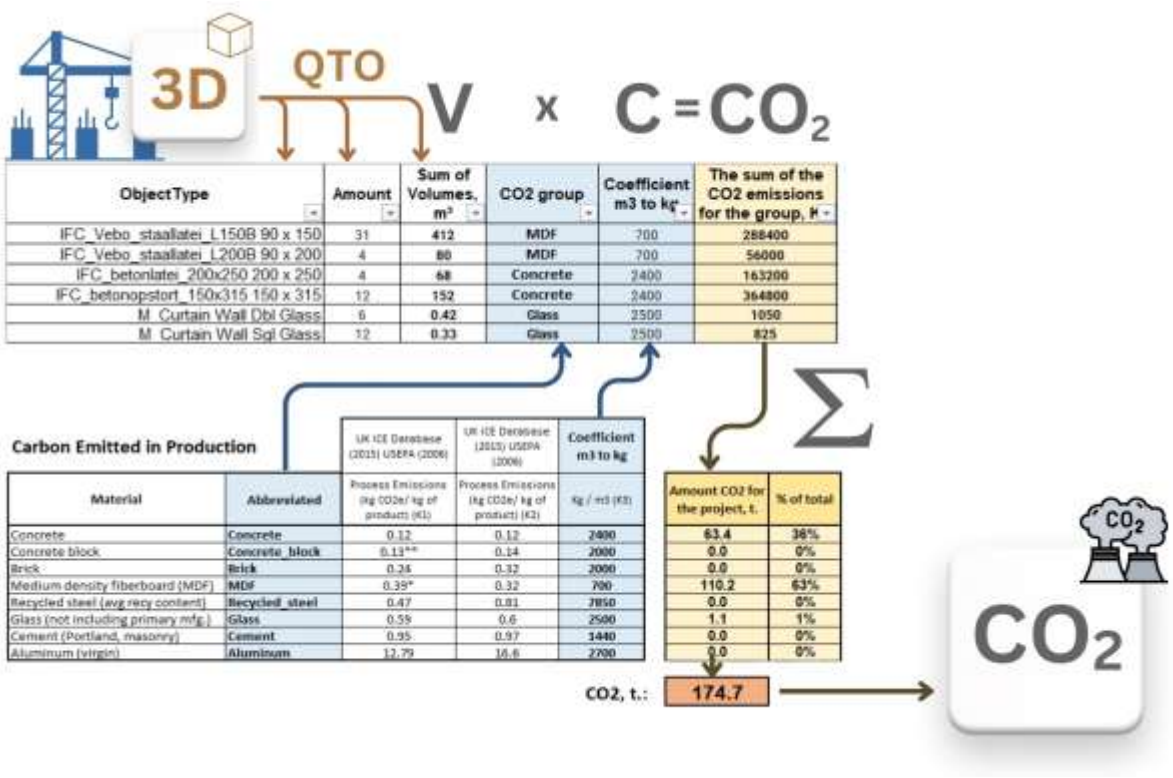


図 5.3-8 CAD データベースからの QTO グループの統合 最終的な CO_2排出量の推定値を導き出す際の精度と自動化を提供.

同様に、要素グループのボリュームを定義することで、材料管理とロジスティクス、品質モニタリングと管理、エネルギー・モデリングと分析、その他多くのタスクのための計算を実行し、個々の要素グループとプロジェクト全体の両方の新しい属性ステータス（表のパラメータ）を得ることができます。

社内でこのような計算プロセスが増え始めると、その計算を自動化し、計算結果を社内のプロセスやデータ管理システムに導入する必要性について疑問が生じる。

完全なソリューションは複雑であるため、建設業界の中・大企業は、このような自動化を ERP （または PMIS ）システム開発会社に委託している。開発会社は、資材やリソースの計算を含む多くの異なる情報レイヤーを管理するために、大規模な顧客向けに単一の包括的なモジュラーシステムを作成する。

章 5.4.

construction erp and pmis systems

建設 ERP - 計算と見積もりを例にしたシステム

モジュール型 ERP システムは、様々な属性（情報）レイヤーとデータフローを単一の包括的なシステムに統合し、プロジェクトマネージャーは、単一のプラットフォーム内で、リソース、財務、ロジスティクス、およびプロジェクトのその他の側面を同期的に管理することができます。建設 ERP システムは、建設プロジェクトの「頭脳」として機能し、自動化によって反復的なプロセスを簡素化し、建設プロセス全体の透明性とコントロールを提供します。

> 建設 ERP -システム（企業資源計画）は、建設プロセスの様々な側面を管理し、最適化するために設計された包括的なソフトウェアソリューションです。建設 ERP システムの中核には、原価計算とスケジューリングを管理するモジュールがあり、効率的なリソースプランニングのための重要なツールとなっています。

ERP モジュール -システムは、ユーザーが構造化された方法でデータを入力、処理、分析することを可能にし、材料費や人件費の会計、設備の利用、物流管理、人事、連絡先、その他の建設活動を含むプロジェクトのさまざまな側面をカバーする。

システムの機能ブロックのひとつは、ビジネスロジックの自動化モジュールである BlackBox/WhiteBox で、プロセスコントロールセンターの役割を果たす。

BlackBox /WhiteBox、ERP - システムを使用する専門家が、他のユーザーや管理者によって事前に設定されたビジネスの様々な側面を、アクセス権によって柔軟に管理することを可能にします。ERP システムの文脈では、BlackBox と WhiteBox という用語は、システム内部ロジックの透明性と制御性のレベルを指す：

- BlackBox ("ブラックボックス") - プロセス実行の内部ロジックにアクセスすることなく、ユーザーはインターフェースを通じてシステムと対話する。システムは、エンドユーザーには見えない定義済みのルールに基づいて、独自に計算を実行する。ユーザーはデータを入力し、内部でどのような属性や係数が使用されたかを知ることなく結果を得る。
- WhiteBox ("ホワイトボックス") - プロセスロジックの表示、カスタマイズ、変更が可能です。上級ユーザー、管理者、インテグレーターは、データ処理アルゴリズム、計算ルール、プロジェクトエンティティ間の相互作用シナリオを手動で定義することができます。

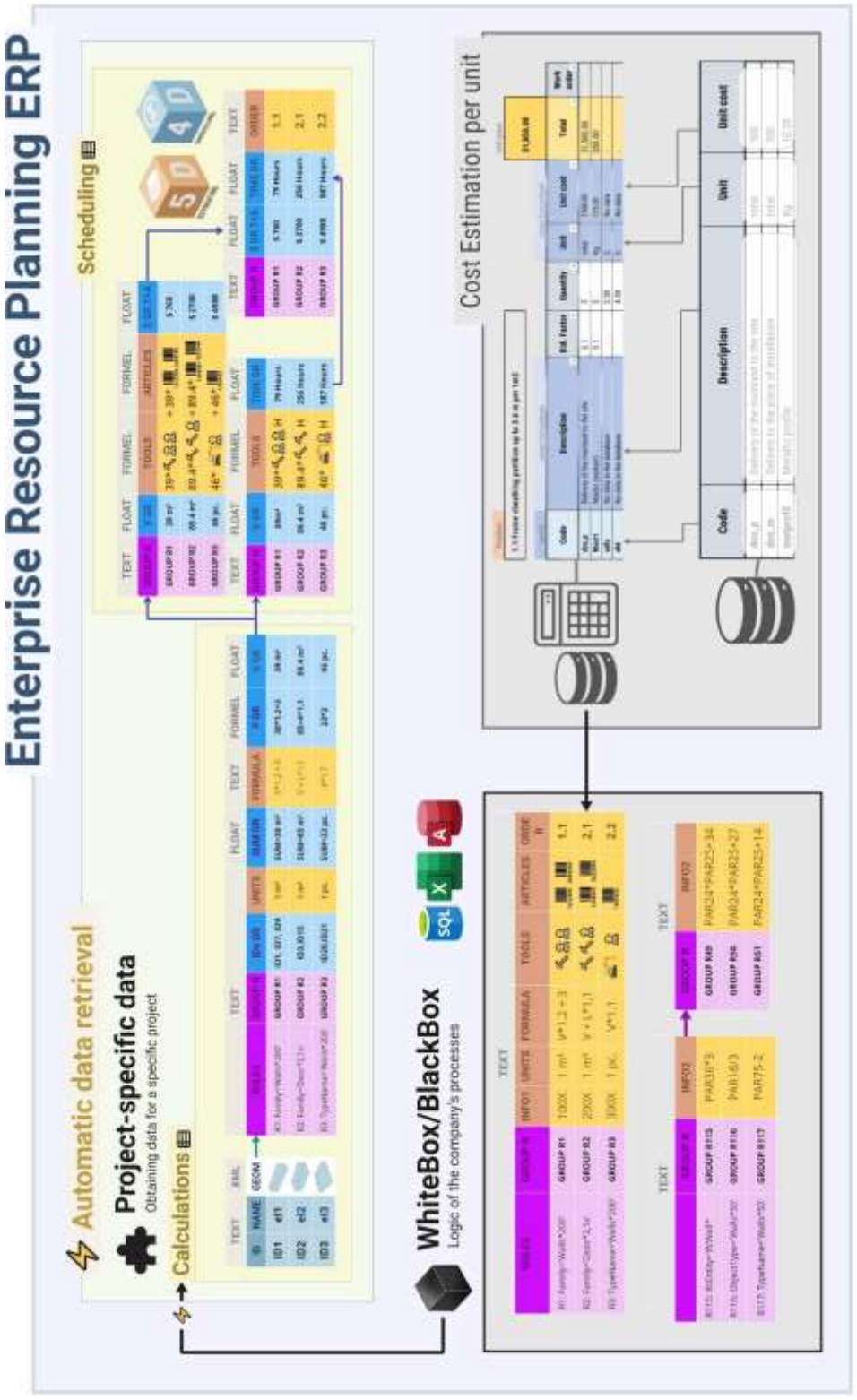


図 5.4-1 建設業 ERP のアーキテクチャ - システムでは、ボリューム属性を手入力することで、見積もりと作業スケジュールを取得する。

例えば、経験豊富なユーザーや管理者が、見積もりのどの属性を掛け合わせるか、または特定の特性でグループ化するか、そして最終結果をどこに記録するかといったルールを設定します。その後、見積エンジニアのようなあまり訓練を受けていない専門家は、ユーザーインターフェイスを介して ERP に新しいデータをロードするだけで、コードを書いたり、ロジックの技術的な詳細を理解したりすることなく、見積、スケジュール、または仕様を作成することができます。

> 前の章では、計算モジュールと論理モジュールについて、LLM とのインタラクションの文脈で説明した。ERP 環境では、このような計算や変換は、ボタンやフォームのインターフェースの後ろに隠れたモジュール内で行われる。

以下の例（図 5.4-1）では、BlackBox /WhiteBox モジュールの ERP - システムの管理者が、見積もりから得たエンティティの属性と QTO をグループ化するための属性をマッチングするためのルールを定義しています。この BlackBox/WhiteBox モジュールが（マネージャーまたは管理者によって）設定されたおかげで、ユーザー（見積担当者またはエンジニア）は、ERP のユーザーインターフェースを介して数量または量の属性を手動で追加することで、完成した 見積もりと作業スケジュールを自動的に受け取ることができます。このようにして、ERP 内部でコードの助けを借りて、前章で説明した計算と見積もり作成プロセスが、半自動化されたコンベアになる。

この半自動化プロセスを、CAD （BIM ）モデル（図 4.1-13）からの容積属性に接続することで、例えば、この目的のために予め設定されている ERP モジュール（ ）に CAD プロジェクトをロードすることで、データフローは、CAD モデルを ERP にロードする際に、設計段階での変更に応じて、個々の要素グループまたはプロジェクト全体の値を自律的かつ即座に更新できる同期化されたメカニズムに変わります。

CAD （BIM ）と ERP システムの間で自動化されたデータフロー（図 5.4-2）を構築するためには、CAD（BIM）モデル・データベースからのデータに関する基本的なプロセスと要件 を構造化された方法で定義する必要があります。ERP におけるこのプロセスは、同様のステップに分けられる：

- バリデーション・ルールの作成(1) ERP -システムに入力されるデータの正確性を保証する重要な役割を果たす。検証ルールは、 エンティティとその属性を検証するフィルタの役割を果たし、 要件をパスした項目だけがシステムに入力されるようにします。検証と妥当性確認については、"Creating requirements and validating data quality " の章で詳しく説明しています。
- 次に、ERP 内で検証プロセス(2)が行われ、すべてのプロジェクト・エンティティの要素とその属性と値が正しく作成され、次の処理ステップに進む準備ができていることが確認される。

- 不完全な属性データに問題がある場合、レポート(3)が作成され、修正指示とともにプロジェクトは次の反復の準備が整うまで修正のために送られる。
- プロジェクトデータが検証・確認されると、別の ERP モジュール (4)で使用され、事前に作成されたルール(WhiteBox/BlackBox)に従って、エンティティグループ、材料、リソースの数量属性を作成する数量テイクオフテーブル(QTO)が作成されます。
- マッチングルールや QTO によってグループ化されたデータは、自動的に計算（コストや時間など）と統合される(5)。
- ERP -システムの最後のステップでは、ユーザは、QTO テーブル のスコープ属性とプロセステーブルの属性（例えば、見積項目） を掛け合わせることにより、各エンティティグループおよびプロジェクト全体の計算結果（6）（例えば、コスト 見積、作業スケジュール、または CO_2 排出量 ）を自動的に生成する。

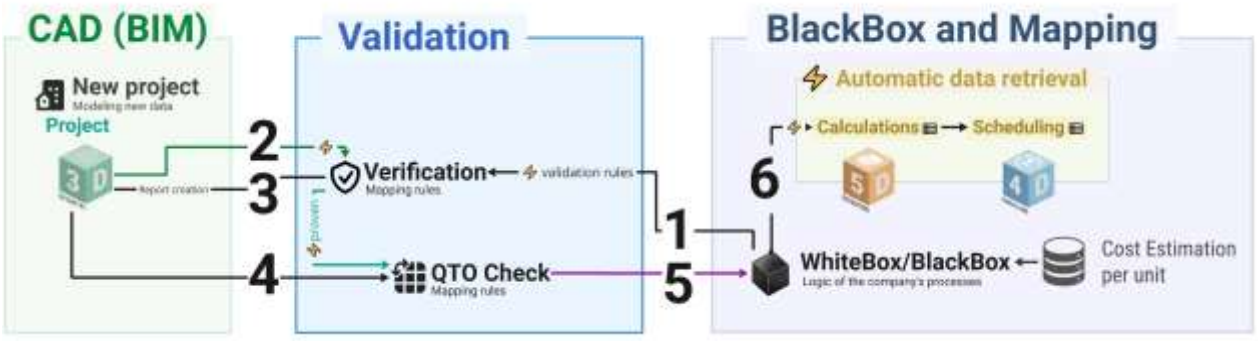


図 5.4-2 検証ルールの作成（1）からコストと作業スケジュールの自動計算（5-6）までの、CAD（BIM ）による建設 ERP -システムのアーキテクチャ。

モジュール型 ERP システムでは、ユーザー・インターフェースを含むソフトウ ェアを使ってプロセスが統合される。インターフェイスの背後にはバックエンドがあり、構造化されたテーブルが、マネージャーや管理者によって事前に設定された様々なオペレーションを実行することでデータを処理します。その結果、ユーザーは、事前に定義され、カスタマイズされた自動化ロジック（BlackBox /WhiteBox のモジュール）のおかげで、自分のタスクに合った半自動的に作成された文書を受け取ることができます

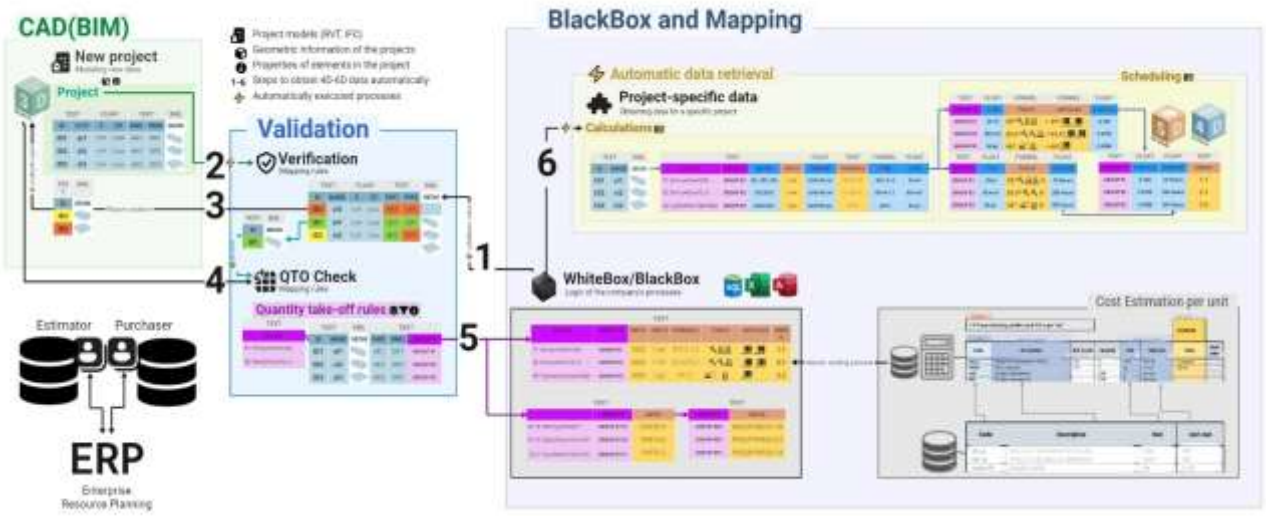


図 5.4-3 ERP - システムは、管理者とユーザーが新しいデータを生成するために、専門家テーブル間を移動するのを助ける。

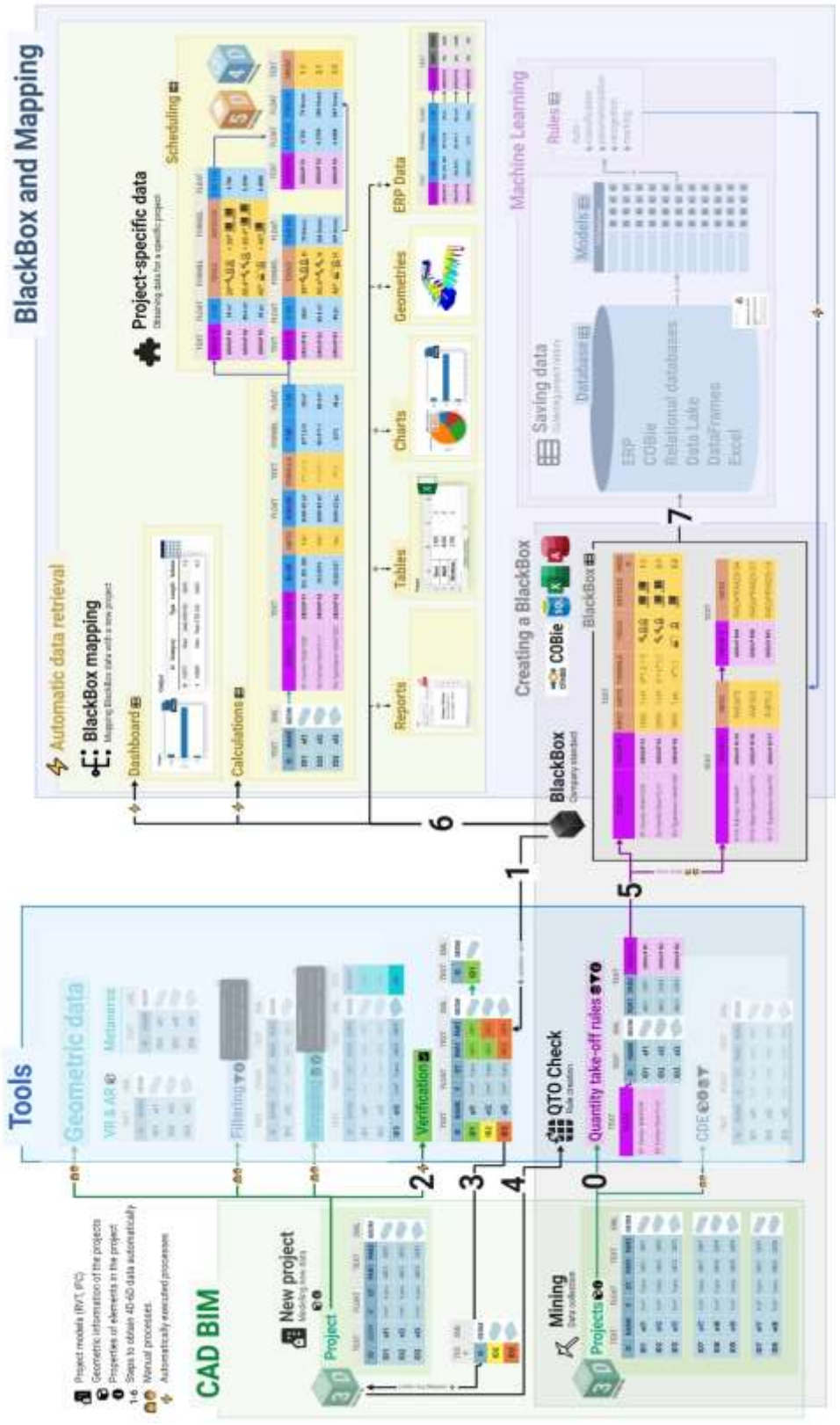


図 5.4-4 ERP -システムは分析ツールと統合され、企業の意思決定プロセスを自動化する。

同様に、ERP -システムにおける、開始から最終計算までのプロセス（ステップ 1-6 図 5.4-3）は、相互に関連するステップの連鎖であり、最終的に計画の透明性、効率性、正確性を提供する。

最新の建設 ERP -システムには、コストとスケジュールの計算モジュールだけでなく、その他の数十の設定済みモジュールが含まれており、通常、文書管理、プロジェクト進捗追跡、契約管理、サプライチェーン、ロジスティクス、および他のビジネスシステムやプラットフォームとの統合をカバーしています。ERP の統合された分析ツールにより、ユーザーはプロジェクト KPI（KPI - 主要業績評価指標）を監視するためのダッシュボードの作成を自動化することができます。これにより、建設プロジェクトのあらゆる側面を一元的かつ一貫性をもって管理することができ、多数のアプリケーションやシステムを単一のプラットフォームに統合することができる。

将来的には、ERP -analytics が機械学習と組み合わされて使用され、精度が向上し、将来のプロジェク ト属性の計算プロセスが最適化される。ビッグ・データ （図 5.4-4）において ERP システムから分析・収集されたデータと属性は、将来、潜在的な遅延やリスク、例えば材料費の変化の可能性を正確に予測できる予測モデルを作成するための基礎となる。

ERP に代わるものとして、建設業界では PMIS (Project Management Information System)を使用することが多い。PMIS は、個々の建設プロジェクト・レベルでの詳細なタスク管理のために設計されたプロジェクト管理システムである。

PMIS: ERP と建設現場の中間的存在。

企業のビジネスプロセス全体をカバーする ERP とは異なり、PMIS は特定のプロジェクトの管理に重点を置き、スケジュール、予算、リソース、ドキュメントを監視する。

> PMIS (Project Management Information System) は、プロジェクトのあらゆる側面を計画、追跡、分析、報告するために設計された建設プロジェクト管理ソフトウェアです。

PMIS では、文書、スケジュール、予算を管理することができます。一見すると、PMIS は ERP と重複したソリューションに見えるかもしれませんが、重要な違いは管理レベルです：

- ERP は、企業レベルのコスト、契約、調達、人事、資源管理など、企業全体のビジネス・プロセスに焦点を当てている。
- PMIS 個々のプロジェクト管理に重点を置き、詳細な計画、変更管理、報告、参加者の調整を行う。

多くの場合、既に十分な機能を持っているのは ERP -システムであり、PMIS の導入は利便性と会社の好みの問題となる。多くの請負業者や顧客が PMIS を使用しているのは、それが必要だからではなく、特定のプラットフォーム上でデータを集約したいベンダーや大口顧客によって押し付けられたからである。

建設プロジェクト・マネジメントの国際的な用語には、PLM （Product Lifecycle Management）や EPC、EPC-M（Engineering, Procurement and Construction Management）といった、建設業界における契約方式とは別の一般的な概念があることに触れておく必要がある。

もし、企業がすでにプロジェクト管理モジュールを備えた ERP を使用している場合、PMIS の導入は、機能が重複する不要なリンクかもしれない。しかし、プロセスが自動化されておらず、データが断片化されている場合、PMIS はより便利でメンテナンスが容易なツールとなる。

投機、利益、偏狭さ、透明性の欠如

ERP と PMIS における透明性の欠如

建設 ERP と PMIS -システムは、外見上はインターフェイスや手順がシンプルであるにもかかわらず、ほとんどの場合、閉鎖的で柔軟性に欠けるソリューションである。このようなシステムは通常、単一のベンダーから設定済みのソフトウェアパッケージとして提供され、内部データベースやプロセスロジックへのアクセスは制限されている。

CAD-(BIM-) ベンダーのデータベースには、ERP システムで必要とされる情報、つまりプロジェクト要素の量的属性や容積属性が含まれているため、こうしたシステムの開発と管理を引き継ぐケースが増えている。しかし、ベンダーは、このデータへのアクセスをオープンな形式や機械可読形式で提供する代わりに、限られたユーザーシナリオと、BlackBox モジュール内で事前に定義されたクローズドな処理ロジックしか提供していない。このため、システムの柔軟性が低下し、特定のプロジェクト条件に適合させることができません。

> 限られたデータの透明性は、建設におけるデジタルプロセスの重要な課題の 1 つである。クローズドなデータベース・アーキテクチャ、建築要素の属性の完全なセットへのアクセスの欠如、Black-Box 自動化モジュールへのフォーカス、オープン・インターフェースの欠如は、文書官僚主義のリスクを大幅に増大させる。このような制限は、意思決定プロセスにボトルネックを生み出し、情報の検証を困難にし、ERP/PMIS システム内でのデータ隠蔽や憶測の扉を開いてしまう。ユーザーは通常、主要なデータソースと直接やり取りすることなく、限定的なアクセスしかできない。これは、QTO 計算に使用される体積、面積、数量など、CAD プロジェクトから自動的に生成されるパラメータに関しては特に重要です。

その結果、多くの建設会社は、プロセスの自動化、オープンデータ、取引コストの削減、新たなビジネスモデルの構築を通じて効率化を追求する代わりに、外部パラメーターの管理、つまりクローズドな ERP/PMIS プラットフォーム（ ）におけるプロジェクトコストに影響を与える要因、調整要因、計算方法の操作に注力している。これは、投機の余地を生み、実質的な生産コストを歪め、建設プロセスのすべての参加者間の信頼を低下させる

> 建設業では、完成したプロジェクトから得られる収益と、設計費、材料費、労務費、その他プロジェクトの実施に直接関係する費用を含む変動費との差額として利益が形成される。しかし、これらのコストの価値を左右する重要な要因は、技術や物流だけでなく、計算のスピードや正確さ、企業内の経営判断の質にもある。

この問題は、ほとんどの建設会社において、コスト計算プロセスが、顧客だけでなく、見積部門や財務部門に属さない従業員自身にとっても、不透明なままであるという事実によって、さらに深刻化している。このような閉鎖性は、ERP/PMIS（ ）システムの属性や補正係数を編集する排他的な権利を持つ「財務専門知識」の担い手である特権的な専門家グループの社内形成を助長する。これらの従業員は、企業のトップとともに、プロジェクトの財務ロジックを実際にコントロールすることができる。

> このような状況では、見積もり担当者は「財務ジャグラー」となり、会社の利益を最大化することと、顧客に対して競争力のある価格を維持する必要性の間でバランスを取ることになる。同時に、会社の評判を落とさないために、露骨で重大な操作は避けなければならない。この段階で、材料や工事の過大な量やコストを隠すための係数が設定される。

その結果、建設業を営む企業の効率性と収益性を高めるための主なスキームは、意思決定プロセスの自動化や迅速化ではなく、資材や工事の価格への投機となっている（図 5.4-5）。ERP /PMIS - システムの灰色会計では、係数を用いて資材や工事量の平均市場価格に対するパーセンテージを水増しすることで、工事費や資材の過大計上が行われている（図 5.1-6）。「計算のまとめ と資源ベースに基づく工事費の計算」の章で説明した。

> その結果、顧客は実際のコストや作業範囲を反映しない計算書を受け取ることになるが、それは多くの隠れた内部係数の派生物である。同時に、下請け業者は、ゼネコンが設定した過小な見積もりに応えようとして、安価で低品質な資材を購入せざるを得ないことが多く、最終的な施工品質が悪化する。

何もないところから利益を求める投機的なプロセスは、信頼できないデータを受け取る顧客と、より多くの投機モデルを見つけることを余儀なくされる重役の双方を傷つける結果と

その結果、プロジェクトが大きくなればなるほど、データやプロセス管理における官僚主義的なレベルは高くなる。すべてのステップ、すべてのモジュールに、計算アルゴリズムや内部手続きに埋め込まれた不透明な係数や課徴金が隠されていることが多い。これは監査を困難にするだけでなく、プロジェクトの財務状況を大きく歪める。大規模な建設プロジェクト（ ）では、このような慣行によって、最終的なコストが何倍にも（時には 10 倍にも）膨れ上がることがよくあるが、その一方で、実際のボリュームとコストは、発注者の効果的なコントロールの外にとどまっている（図 2.1-3 ドイツにおける大規模インフラ・プロジェクトの計画コストと実績コストの比較）。

> マッキンゼー・アンド・カンパニーの報告書『Imagining the Digital Future of Construction』（ 2016 年）によると、大規模な建設プロジェクトの完成は平均して計画より 20%遅れ、予算は最大で 80%超過している[107]。

積算・予算管理部門は、企業内で最も警戒される部門になりつつある。社内の専門家であっても、これらの部署へのアクセスは厳しく制限されており、閉鎖的なロジックとデータベース構造により、プロジェクトの意思決定の効率性を歪みなく客観的に評価することは不可能である。透明性の欠如は、企業がプロセスを最適化するのではなく、数字や要因の「創造的な」管理によって生き残りをかけて戦うことを余儀なくされるという事実につながっている（図 5.3-1、図 5.1-6-たとえば「入札要因」というパラメーター）。

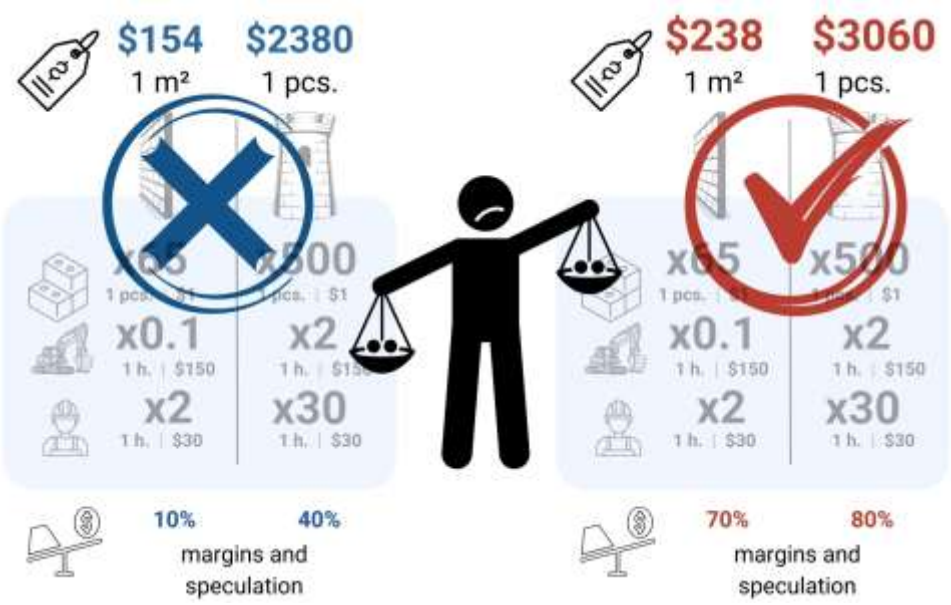


図 5.4-5 決済レベルの投機比率は、企業の主な利益であり、仕事の質と評判を両立させる術である。

こうしたことから、建設業界では、クローズドな ERP/PMIS（ ）システムの継続的な使用には疑問符がつく。デジタルトランスフォーメーションと、顧客側からの透明性への要求の高まり（図 10.2-3）を背景に、長期的なプロジェクト実現が、柔軟性を制限し、統合を妨げ、ビジネスの発展を妨げる、独自のソリューションに依存し続けるとは考えにくい。

また、建設会社にとって、クローズド・データベースのデータ・サイロや不透明なデータで作業することがいかに好ましいことであっても、必然的に、建設業界の将来は、オープン・プラットフォーム、機械可読で透明なデータ構造、信頼ベースの自動化への移行を伴うことになる。この変革は、説明責任、持続可能性、透明性、経済的実現可能性をますます求めるようになる顧客、規制当局、社会からのプレッシャーのもと、トップが推進することになるだろう。

クローズド ERP 時代の終焉 /PMIS: 建設業界は新たなアプローチを必要としている

何千万行ものコードで構成される、かさばるモジュール式 ERP/PMIS -システムの使用は、その変更を極めて困難にする。この場合、すでに企業用に設定されたモジュール、リソース・データベース（図 5.1-3）の何万もの記事、何千もの既成の計算（図 5.1-6）が存在する中で、新しいプラットフォームへの移行は、コストと時間の

かかるプロセスになってしまう。コードやレガシー・アーキテクチャーが増えれば増えるほど、社内の非効率レベルは高くなり、新しいプロジェクトが増えるたびに事態は悪化するばかりだ。多くの企業では、データ移行と新しいソリューションの統合が、絶え間ない手直しと妥協点の果てしない模索を伴う、数年にわたる壮大なプロジェクトになっている。その結果、限界があるにもかかわらず、使い慣れた古いプラットフォームに戻ることが多い。

建設データ管理のシステム的失敗に関するドイツのブラックブック報告書[108]で強調されているように、情報の断片化とその管理への集中的アプローチの欠如は、非効率の主な原因である。標準化と統合がなければ、データはその価値を失い、管理ツールというよりむしろアーカイブになってしまう。

> データ品質損失の主な原因 は、建設プロジェクトの不適切な計画と管理であり、これはしばしば大幅なコスト増につながる。ブラック・ブックの「フォーカス：コスト爆発」のセクションでは、こうした望ましくない結果をもたらす主な要因を分析している。これには、不適切なニーズ分析、フィージビリティ・スタディの欠如、回避できたはずの追加コストにつながる調整不足の計画などが含まれる。

> 成熟した企業の IT エコシステムでは、旧式のシステムを交換することは、すでに建設された建物の耐力柱を交換することに匹敵する。単に古いものを取り外して新しいものを設置するだけでは不十分で、建物が安定し、天井が崩れず、すべての通信が機能し続けるような方法で行うことが重要だ。ここが難しいところで、ミスがあれば会社のシステム全体に深刻な影響を及ぼしかねない。

それにもかかわらず、建設業界向け大規模 ERP 製品の開発者は、コードの記述量を自社のプラットフォームを支持する論拠として使い続けている。専門的な会議では、「このようなシステムを作り直すには 150 人年かかる」というようなフレーズを耳にすることがあるが、このようなシステムの機能のほとんどは、データベースと、テーブルを操作するためのかなり単純な機能を隠し、特別に固定されたユーザー・インターフェースにパッケージ化されているという事実にもかかわらずである。実際には、「150 人年」というコード量は、競争上の優位性どころか、むしろ重荷になってしまう。コードが多ければ多いほど、サポートにかかるコストが高くなり、新しい状況に適応するのが難しくなり、新しい開発者や顧客にとって参入の敷居が高くなる。

> 今日のモジュール式建築システムの多くは、煩雑で時代遅れの「フランケンシュタイン構造」に似ており、不注意な変更が故障につながりかねない。新しいモジュールが追加されるたびに、すでに過負荷になっているシステムの複雑さが増し、一部の専門家しか理解できない迷宮のようなシステムになってしまう。

この複雑さは、開発者自身によっても実感される。開発者は定期的にリファクタリング（新しい技術の出現を考慮してアーキテクチャを修正すること）を中断する。しかし、リファクタリングが定期的に行われても、複雑さは必然的に増大する。このようなシステムのアーキテクトは、複雑さの増大には慣れるが、新しいユーザーや専門家にとっては乗り越えられない壁となる。その結果、すべての専門知識が少数の開発者の手に集中し、システムはスケーラブルでなくなる。短期的にはそのような専門家は有用だが、長期的には問題の一部となる。

> 組織は今後も「小さな」データをビッグデータと統合していくだろうし、どんなに高価で堅牢なアプリケーションでも、1 つのアプリケーションですべてを処理できると考えるのは愚かなことだ[109]。
>
> - フィル・サイモン（Conversations About Collaboration ポッドキャストホスト

他の業界では、オープンデータと透明性のあるロジックを備えた分析ツールを使って、同じようなタスクに対処してきたのに。

現在、クローズド・モジュラー・プラットフォームは、主に原価計算の特殊性から、建設業界ではまだ需要がある（図 5.1-7）。このようなシステムは、グレーまたは不透明なスキームを実行するために使用されることが多く、実際のコストを顧客から隠すことができます。しかし、顧客が中心となって業界がデジタル的に成熟し、いわゆる「ユビキタス時代」に移行するにつれ、仲介業者、すなわち ERP を持つ建設会社は、時間とコストの計算における重要性を失っていくだろう。これによって建設業界の様相は一変するだろう。詳しくは、本書の最後の部分、「建設 5.0：隠せなくなったときにお金を稼ぐ方法」の章をお読みください。

> 過去 30 年間に蓄積され、開発に何千人もの工数が費やされたレガシー・ソリューションは、急速に姿を消し始めるだろう。オープンで透明性が高く、柔軟なデータ管理への移行は避けられない。唯一の問題は、どの企業がこうした変化に適応し、どの企業が古いモデルの人質のまま残るかということだ。

同様の状況は、CAD（BIM -）ツールの分野でも見られ、そのデータは今日、ERP/PMIS -システムにおける設計エンティティの容積パラメータを埋めている。当初、BIM のアイデア（2002 年に開発された[110]）は、単一の統合データベースのコンセプトに基づいていたが、実際には、今日、BIM で作業するには、一連の特殊なソフトウェアと 形式が必要である。設計と建設管理を簡素化するはずのものが、統合を複雑にし、ビジネスの柔軟性を低下させる、独自のソリューションの別のレイヤーになってしまった。

次のステップ：プロジェクト・データの効率的活用

このパートでは、構造化されたデータが、建設プロジェクトの正確なコストとスケジュールの計算の基礎となることを示した。QTO、スケジューリング、見積もりプロセスを自動化することで、人件費を削減し、結果の精度を大幅に向上させることができる。

このパートを要約すると、議論したアプローチを日常業務に適用するのに役立つ主な実践的ステップを強調する価値がある。これらのアプローチは普遍的なものであり、企業のデジタルトランスフォーメーションにも、計算に携わる専門家の日常業務にも役立つ：

- ルーチン計算の自動化
  - [ ] 自分の仕事に関連づけられる標準的な仕事の原価計算を見つけるようにする。
  - [ ] あなたの国の建設現場で、工事や工程の原価計算や計算にどのような方法が使われているかを分析しなさい（図 5.1-7）。
  - [ ] CAD システム（）で作業されている方は、CAD（BIM- ）ソフトウェアの仕様書と QTO データの自動抽出機能を学んでください。
  - [ ] LLM を使って、計算を自動化するためのドラフトコードを書く。
- QTO 用の独自ツールの開発
  - [ ] 台数カウントを自動化するスクリプトやテーブルを作成する

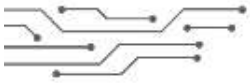
- [ ] 評価の一貫したアプローチのために、カテゴリーと要素グループを標準化する。
- [ ] 新しいプロジェクトで結果の再現性を確保するため、計算方法を文書化する。

■ プロジェクトのさまざまな側面を自分の仕事に統合する

- [ ] モジュラー・システムを使用している場合、プロセスを図や表としてだけでなく、データ・レベル、特に表の形でも視覚化するようにします。
- [ ] グループ化、フィルタリング、集計を使用した Python コードにより、CAD データベースから抽出したデータと計算との自動結合をマスターする。
- [ ] 複雑な情報を同僚や顧客に提示するために、QTO グループの明確なビジュアライゼーションを作成する。

これらのステップは、自動化とデータの標準化に基づいた持続可能な計算システムの構築に役立つ。このアプローチは、精度を向上させ、日常的な計算問題を軽減する。

以下の章では、CAD - (BIM-)製品の技術的側面と、CAD データベースが企業のビジネスプロセスに統合するのがまだ難しい理由に焦点を当てている。建設業における BIM 導入の歴史、CAD ツールの進化、これらの技術を使った作業の技術的な側面に今興味がなければ、そのまま本書の第 7 部「データ駆動型意思決定」に進むことができる。



印刷版で最大限の利便性を

データ駆動型建設』の無料デジタル版をご覧いただけます。より便利な作業や資料への迅速なアクセスのためには、印刷版のご利用をお勧めします：



■いつでも手元に：印刷された本書は信頼できる作業ツールとなり、どのような作業状況でも適切なビジュアライゼーションやダイアグラムを素早く見つけて使用することができます。

■高品質のイラスト：印刷版のすべての画像とグラフィックは最高の品質で表示されます。

■情報への素早いアクセス：簡単なナビゲーション、メモやブックマークを作成する機能、どこでも本で作業する機能。

印刷された完全版を購入することで、快適で効率的な情報作業のための便利なツールを手に入れることができます。日々の作業でビジュアル資料を素早く利用し、必要なスキームを素早く見つけ、メモを取ることができます。さらに、ご購入はオープンな知識の普及をサポートします。

印刷版のご注文はこちらから： datadrivenconstruction.io/books



VI 部

cad and bim: マーケティング、現実、そして建設における設計データの未来

本書の第 6 部では、CAD（）と BIM（）のテクノロジーの進化と、建設におけるデータ管理プロセスへの影響について批判的な分析を行う。統合データベースという当初のアイデアから、ソフトウェアベンダーが推進する現在のマーケティング構造まで、BIM コンセプトの歴史的変遷をたどっている。独自のフォーマットや閉鎖的なシステムが、プロジェクトデータの取り扱い効率や建設業界全体のパフォーマンスに与える影響について評価する。異なる CAD システムの互換性の問題や、建設会社のビジネスプロセスとの統合の難しさについても詳しく分析している。USD のような簡素化されたオープンデータ形式への現在の傾向と、それらが建設業界に与える潜在的な影響について議論する。リバースエンジニアリング技術（）を含む、クローズドシステムから情報を抽出するための代替アプローチも紹介する。人工知能（）や機械学習（）の建設における設計やデータ分析プロセスの自動化への応用の見通しについて分析する。ソフトウェアベンダーの利益ではなく、ユーザーの真のニーズを指向した設計技術の発展についての見通しが述べられている。

章 6.1.

Embergence of bim- 建設業界におけるコンセプト

もともと、CAD （BIM ）に特化したこの第 6 部は、この本の最初のバージョンには含まれていなかった。独自フォーマット、ジオメトリックカーネル 、クローズドシステムといったトピックは、専門的で詳細が多すぎるため、データの扱い方を理解したいだけの人には一見役に立たないように思われたからだ。しかし、CAD システムの複雑な内部構造や、ジオメトリカーネル 、様々なフォーマット、同じデータでも互換性のない保存方式を理解しなければ、ベンダーが推進するコンセプトがなぜ情報の取り扱いを難しくし、オープン・パラメトリック・デザインへの移行を妨げているのかを真に理解することは不可能であることが、本書の最初のバージョンへのフィードバックや説明を追加する要望からわかりました。そのため、このパートは本書の構成の中で独自の位置を占めている。CAD（BIM）が優先事項でない場合は、次のパート「PART VII：データ主導の意思決定、アナリティクス 、自動化、機械学習」に進んでください。

CAD- ベンダーのマーケティングコンセプトとしての BIM とオープン BIM の出現の歴史

1990 年代のデジタルデータの出現により、コンピュータ技術はビジネスプロセスだけでなく、設計プロセスにも導入され、CAD （コンピュータ支援設計システム）、後に BIM （ビルディング・インフォメーション・モデリング）などの概念につながった

しかし、どんな技術革新もそうであるように、それらは開発の終着点ではない。BIM のようなコンセプトは、建設業界の歴史において重要なマイルストーンとなったが、遅かれ早かれ、将来の課題によりよく対応する、より優れたツールやアプローチに道を譲るかもしれない。

> CAD ベンダーの影響力に圧倒され、それ自身の実装の複雑さに混乱し、2002 年に登場した BIM のコンセプトは、明るく輝きながらすぐに消えていったロックスターのように、30 周年を迎えるまで生きられないかもしれない。理由は簡単で、データサイエンティストの要求は、CAD ベンダーがそれに適応できるよりも速く変化しているからだ。

質の高いデータの不足に直面している今日の建設業界の専門家は、クロスプラットフォームの相互運用性と、

CAD（ ）プロジェクトからのオープンデータへのアクセスを求め、分析と処理を簡素化しています。CAD データの複雑さや処理の混乱は、設計者、プロジェクトマネージャー、現場の建設作業員、そして最終的には発注者など、建設プロセスに関わるすべての人に悪影響を及ぼします。

今日、顧客と投資家は、運用のための本格的なデータセットの代わりに、複雑な幾何学カーネル 、データスキーマの理解、毎年更新される API -文書、データを扱うための専門的な CAD ソフトウェア（BIM ）を必要とする CAD フォーマットのコンテナを受け取っている。同時に、設計データの多くは未使用のままである。

今日の設計と建設の世界では、CAD データへのアクセスの複雑さが、プロジェクト管理の過剰なエンジニアリングにつながる。CAD データを扱ったり、BIM ソリューションを開発したりする中堅・大企業は、オープンデータを入手するために、CAD ベンダー ソリューションと緊密な関係を維持して API を介してデータにアクセスするか、高価な SDK コンバータ を使用して CAD ベンダーの制限を回避してリバースエンジニアリング を行うことを余儀なくされている[75]。

> プロプライエタリー・データ・アプローチは時代遅れであり、もはや今日のデジタル環境の需要を満たしていない。今後は、オープンデータを効果的に活用する企業（ ）と、市場から撤退する企業の 2 つのタイプに分かれるだろう。

BIM （ビルディング・インフォメーション・モデリング）の概念は、2002 年に大手 CAD ベンダーの一つであるホワイトペー パー BIM [54] が発表したことにより建設業界に登場し、機械工学の概念である BOM （部品表）によって補完され、プロジェクトデータの作成と処理に対するパラメトリックアプローチに由来する（図 6.1-1）。設計データの作成と処理に対するパラメトリック・アプローチは、機械工学設計のための Pro-E システム（MCAD ）に最初に実装されたものの 1 つである。このシステムは今日の建設業界で使用されているものを含む、多くの近代的な CAD ソリューションのプロトタイプとなった [111]。

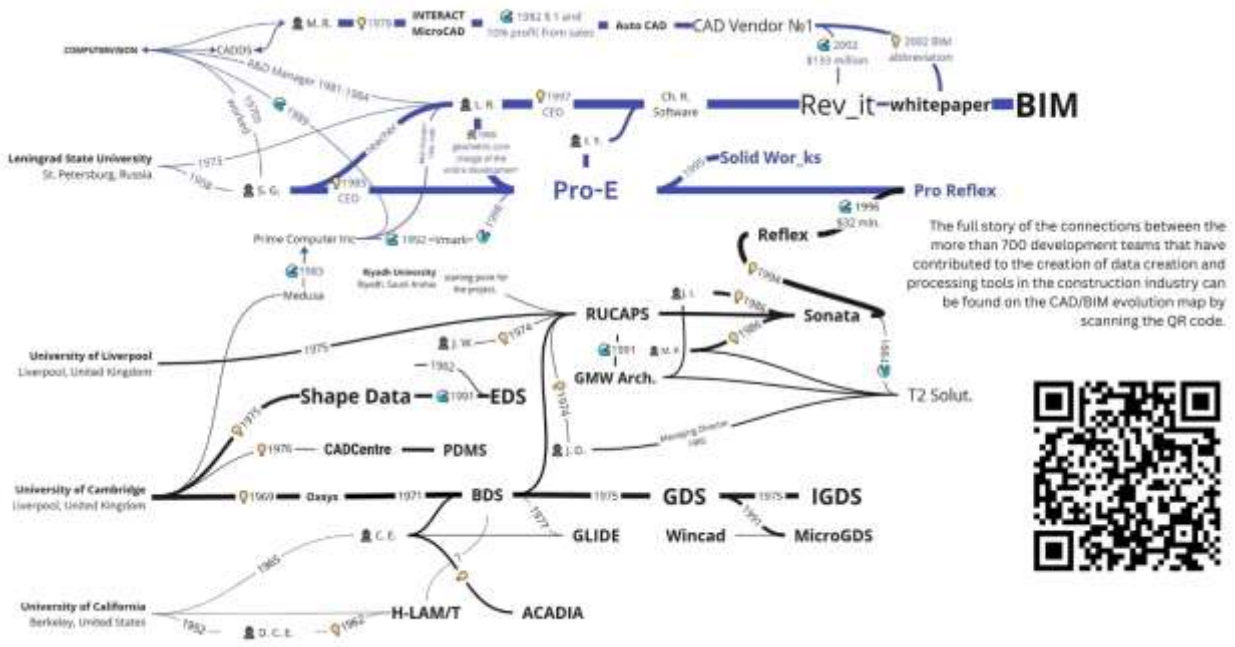


図 6.1-1 BIM 概念の歴史マップ および類似概念。

2000 年代初頭まで CAD ツール -ベンダーを宣伝していたジャーナリストや AEC コンサルタントは、2002 年からホワイトペーパーBIM に関心を移した 。建設業界に BIM の概念を普及させる上で重要な役割を果たしたのは、BIM ホワイトペーパー2002-2004 と、2002 年、2003 年、2005 年、2007 年に発表された記事であった[112]。

> ビルディング・インフォメーション・モデリングは、.....戦略である。[CAD ベンダーの会社名]は、情報技術を建設業界に応用するための戦略である。
>
> - BIM ホワイトペーパー, 2002 [60].

2000 年代半ばになると、"研究者 "たちは、2002 年に CAD- ベンダーによって発表された BIM- コンセプトを、GLIDE、GBM、BPM、RUCAPS などのシステムの基礎となったチャールズ・イーストマンの BDS のような以前の科学的著作とリンクさせ始めた。チャールズ・イーストマンは、その先駆的な著作「Building Description System」（1974 年）において、現代の情報モデリングの理論的基礎を築いた。イーストマンの著作には "データベース "という用語が 43 回も登場する（図 6.1-2）。

> イーストマンの重要なアイデアは、形状から要素の特性、それらの相互関係に至るまで、建物に関するすべての情報を単一の構造化データベースに保存すべきだというものだった。このデータベースから、図面、仕様書、計算書、法規適合性を自動的に生成し、分析することができる。イーストマンは、図面は時代遅れで冗長なコミュニケーション方法であると明確に批判し、情報の重複、更新の問題、変更が加えられた場合の手動更新の必要性を指摘した。イーストマンは、図面が時代遅れの冗長なコミュニケーション方法であることを明確に批判し、情報の重複や更新の問題、変更時の手作業による更新の必要性を指摘した。

イーストマンがビジュアライゼーションをコンセプトの中心に据えなかったことは注目に値する。彼のシステムの中心は情報であった。パラメータ、関係、属性、分析、自動化機能などである。彼の理解では、図面はデータベースからのデータを表示する形態のひとつに過ぎず、設計情報の主要なソースではなかった。

大手 CAD ベンダーの最初の BIM ホワイトペーパー では、「データベース 」というフレーズがチャールズ・イーストマンの BDS と同じ頻度で使用されており、7 ページにわたって 23 回[60]、「ビルディング」、「インフォメーション」、「モデリング」、「デザイン」に次いで、この文書で最もよく使われる単語の一つであった。しかし、2003 年までには、「データベース」という用語は同様の文書の中で 2 回しか登場しなくなり[61]、2000 年代後半には、データベースの話題はデザインデータの議論から事実上消えてしまった。その結果、「視覚的および定量的分析のための単一の統合データベース」というコンセプトが完全に実現されることはなかった。

このように、建設業界は、データベースを重視したチャールズ・イーストマンの先進的な BDS コンセプト 、機械工学製品 Pro-E（今日建設現場で使用されている一般的な CAD -ソリューションの前身）におけるデータベースから設計データを自動的に更新するというサミュエル・ガイスバーグのアイデアから、現在販売されている BIM に至るまで、これが当初の理論的なコンセプトであったにもかかわらず、データベースによるデータ管理についてはほとんど言及されていない。

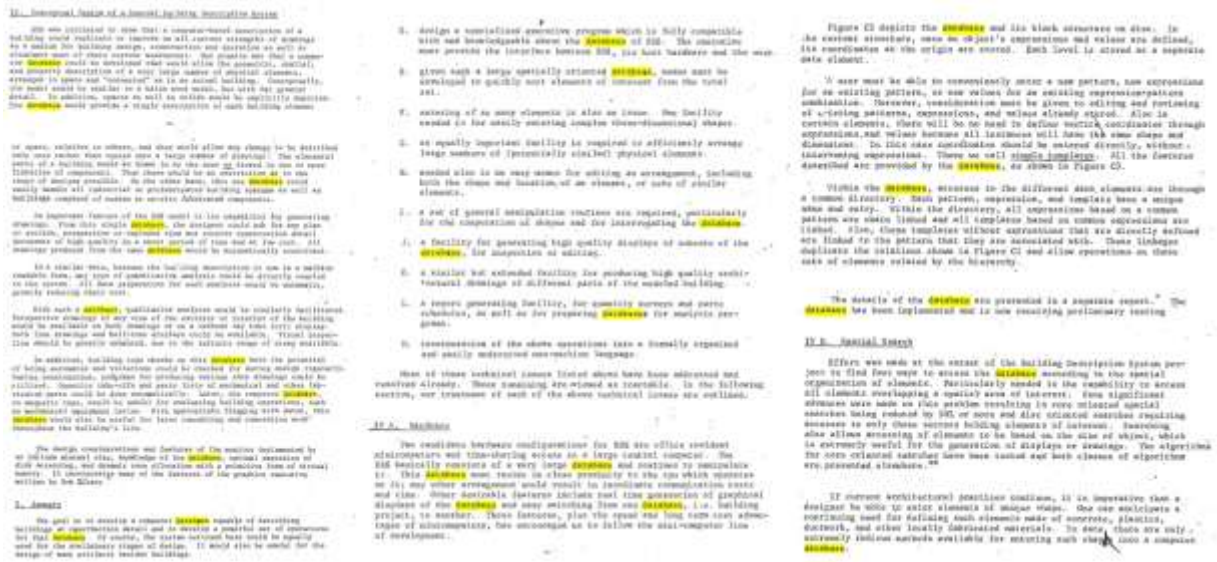


図 6.1-2 1974 年にチャールズ・イーストマンによって記述された BDS の概念 では、「データベース」（黄色で強調表示）というフレーズが 43 回使われている。

BDS と 2000 年代までの類似コンセプトは、ビジュアライゼーション・ツールとしてではなく、建物のデジタル・データベース として開発された。2002 年の BIM は、データベースが後回しにされた設計ツールとなった。1990 年代の BDS と同様のコンセプトから、2010 年代半ばまでの BIM への移行で、私たちは何を失ったのだろうか：

- オープンデータベース：BDS 、その他同様のコンセプトは分析を重視し、BIM はデザインを重視した。
- データを扱う柔軟性：BDS はデータ分析を重視し、BIM は曖昧なデータに基づくプロセスを重視した。
- 透明性： BDS は、オープンな統合データベース を意図していたが、BIM の CAD ベンダー は、自社のデータベースを完全にクローズドにし、プロプライエタリなフォーマットをオープン化するリバースエンジニアリングツール と 20 年間不首尾に終わっている 。

過去 30 年間、設計者は「統合データベース」にアクセスしたことがなく、BIM（ ）ツールにまつわる 20 年間のマーケティング陶酔の後、建設業界はこの流行の結果に気づき始めている。

BIM の現実: 統合データベースの代わりに - 閉鎖されたモジュールシステム

データに焦点を当て、それを構造化し、統一されたプロセスに統合する代わりに、CAD - (BIM-) システムのユ

ーザーは、それぞれが独自のゲームルールを指示する、断片化された独自のソリューションのセットで作業することを余儀なくされている：

- 最初の BIM 白書（）の主題であった単一データベース（）は、神話のままである。声高な主張にもかかわらず、データへのアクセスはまだ限定的で、閉鎖的なシステム間で分散している。
- BIM -モデルはツールというよりむしろ閉じたエコシステムになっている。透明な情報交換の代わりに、ユーザーはサブスクリプションにお金を払い、独自の API を使用することを余儀なくされている。
- データはユーザーのものではなく、ベンダーのものである。プロジェクト情報は、オープンで独立したフォーマットで利用できるのではなく、プロプライエタリなフォーマットやクラウドサービスに閉じ込められている。

設計エンジニアやプロジェクト・マネジャーは、CAD データベース（）にアクセスできないことが多く、また、自身のプロジェクト・データが保存されているフォーマットにもアクセスできない。このため、情報を迅速に検証したり、データ構造や品質に関する要件を策定したりすることができない（図 6.1-3）。このようなデータへのアクセスには、API（）やプラグインを介してリンクされた専門プログラム一式が必要であり、建設業界のプロセスを過度に官僚化することにつながる。一方、これらのデータは、何十もの情報システムと何百人もの専門家によって同時に利用されている。

> なぜなら、（CAD ソフトで）何かを作るエンジニア（設計者）1 人につき、そのデータを扱う人が 10 人いるからです」 [41]。
>
> - CAD の CEO - BIM コンセプトを生み出したベンダー 、2005 年。

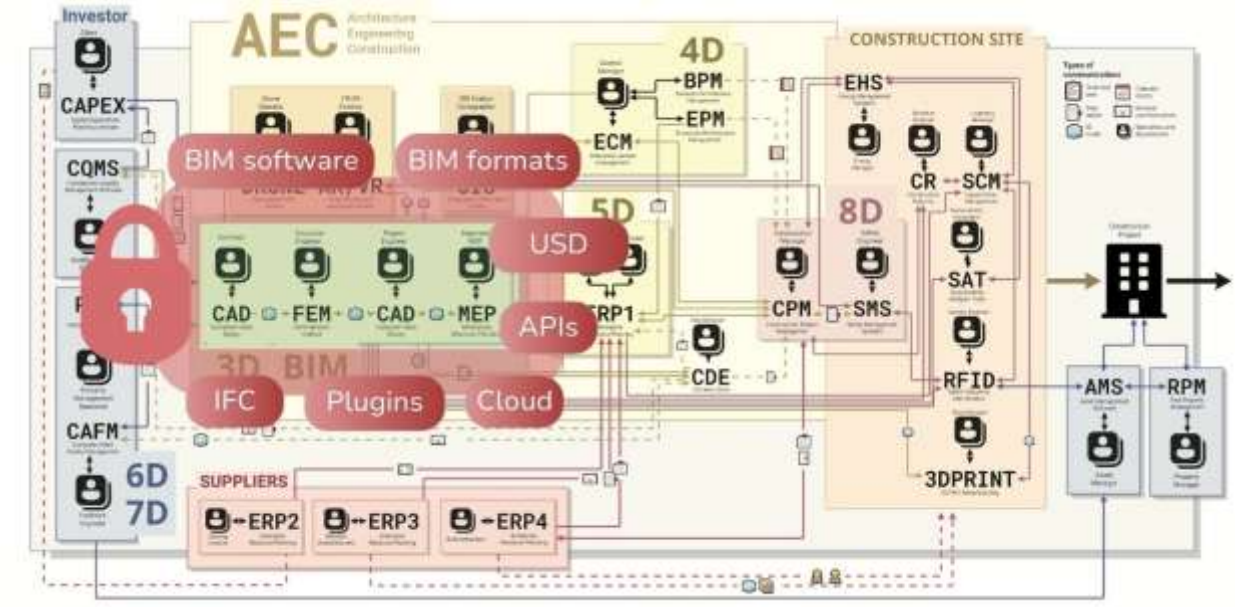


図 6.1-3 CAD- (BIM-) データベースは、建設ビジネスのエコシステムにおいて、IT 部門やデータ管理者にとって最後の閉じたシステムのひとつであり続けている。

> BIM が本格的なデータベース管理ツールというよりは、むしろデータベースを商業化する手段であることが明らかになると、論理的な疑問が生じる。その答えは、ソフトウェアベンダーではなくユーザーが情報の所有者となるオープンなデータ構造を使用することである。

建設業界のユーザーやソリューション開発者は、他の業界と同様、過去 30 年間主流だった曖昧なソフトウェア・ベンダー用語から脱却し、デジタル化の重要な側面である「データ」と「プロセス」に焦点を当てるようになるのは必至だ。

1980 年代後半、建設業におけるデジタル開発の主要分野は、データアクセスとプロジェクト情報管理の問題として提示されていた。しかし、時が経つにつれ、焦点は移り変わってきた。データを扱うための透明でアクセスしやすいアプローチを開発する代わりに、IFC フォーマット（）とオープン BIM コンセプト（）は、プロジェクト・データベース管理の話題から専門家の注意をそらす試みとして、積極的に推進された。

建設業界におけるオープンフォーマット IFC の出現

いわゆるオープンフォーマットである IFC （Industry Foundation Classes）は、異なる CAD（BIM -）システム間の相互運用性を確保するための標準として位置づけられている。その開発は、主要な CAD ベンダーが設立し

管理する組織の枠組みの中で行われた。IFC フォーマットに基づき、2012 年に CAD- の 2 社がマーケティングコンセプト OPEN BIM[63] を開発した。

> IFC (Industry Foundation Classes) は、建設業界におけるデータ交換のためのオープンスタンダードで、異なる CAD - (BIM-) システム間の相互運用性を確保するために設計されています。

オープン BIM - コンセプトは、CAD データベースの情報を使用し、CAD データ交換用のオープンフォーマットである IFC を介してシステム間で情報を交換することです。

> オープン BIM プログラムは、・・・[1 CAD ベンダー]、・・・[2 CAD ベンダー]、およびその他の企業によって開始されたマーケティングキャンペーンであり、AEC 業界全体で OPEN BIM コンセプトをグローバルに協調して推進することを奨励・促進するもので、プログラム参加者は一貫したコミュニケーションと共通のブランディングを利用できます。
>
> - CAD ベンダーのウェブサイト、OPEN BIM プログラム、2012 年 [113] より

IFC は、1980 年代後半にミュンヘン工科大学によって機械工学フォーマット STEP から転用され、その後、大手設計会社と大手 CAD ベンダー によって登録され、1994 年に IAI（Interoperability Industry Alliance for Interoperability）が設立されました[114]（図 6.1-4）。IFC フォーマットは、異なる CAD システム間の相互運用性 を確保するために開発されたもので、機械工学フォーマットである STEP の原則に基づいています。このフォーマットは、NIST（米国国立標準技術研究所）と米国国防総省の支援のもと、CAD ユーザーとベンダーのグループによって 1979 年に作成された IGES フォーマットに由来しています[115]。

しかし、IFC の複雑な構造（ ）、幾何学カーネル（ ）への緊密な依存性、および異なるソフトウェア・ソリューションによるフォーマットの実装の不一致により、その実用化には多くの問題がありました。以前、機械工学の専門家が IGES や STEP フォーマットを使用していた際にも、IFC と同じような問題（細部の損失、精度の制限、中間フォーマットの使用など）に直面しました。



図 6.1-4 開発チームと CAD 製品のリンクマップ (BIM) [116].

> 2000 年、IFC フォーマットを登録し、IAI（後の bS）組織を設立したのと同じ CAD - ベンダーが、ホワイトペーパー "Integrated Design and Manufacturing: Benefits and Rationale" を発行する。[65]. この論文では、IGES や STEP [IFC と同じ]のような中立フォーマットを使用せずに、同一システム内のプログラム間で交換する際に、データの粒度を完全に維持することの重要性が強調されています。その代わりに、情報の正確性が損なわれるのを防ぐために、アプリケーションは基礎となる CAD データベースに直接アクセスできるようにすることが提案された。

2002 年、同じ CAD ベンダー がパラメトリック BOM 製品を購入し（図 3.1-18、詳細は第 3 部 に記載）、それをベースに BIM コンセプト を形成している。その結果、建設プロジェクトデータの交換には、クローズドな CAD フォーマットか IFC フォーマット (STEP) しか使われなくなった。このフォーマットの限界については、2000 年にこのフォーマットを建設業界に持ち込んだ CAD ベンダー自身が書いている。

建設データを作成・処理するツールの作成に関わった 700 を超える開発チームの相互作用の詳細な歴史は、地図「CAD（BIM）の進化」[116]に示されている。[116].

オープンフォームの IFC、設計要素の幾何学的記述とメタ情報の記述から構成されます。CSG や Swept Solids など、IFC 形式でジオメトリを表現するためにさまざまな方法が使用されています。しかし、パラメトリック表現である BREP が、IFC 形式で要素ジオメトリを転送するための主要な標準となっています。この形式は、CAD- (BIM-) プログラムからエクスポートする際にサポートされており、IFC を CAD プログラムにインポートし直す際に要素を編集できる可能性があるためです。

IFC ジオメトリックコアによるフォーマットの問題

ほとんどの場合、IFC のジオメトリがパラメトリックに定義されている場合（BREP ）、IFC ファイルだけでプロジェクト・エンティティの体積や面積などのジオメトリ・プロパティを視覚化したり取得したりすることは不可能になります。この場合、ジオメトリを操作して視覚化するには、ジオメトリ・カーネル（図 6.1-5）が必要になりますが、これは初期状態では欠落しています。

> ジオメトリカーネルは、CAD（CAD ）、BIM 、その他のエンジニアリングアプリケーションで幾何オブジェクトを作成、編集、分析するための基本アルゴリズムを提供するソフトウェアコンポーネントです。2D および 3D ジオメトリを構築し、ブール演算、スムージング、交差、変換、視覚化などの操作を行います。

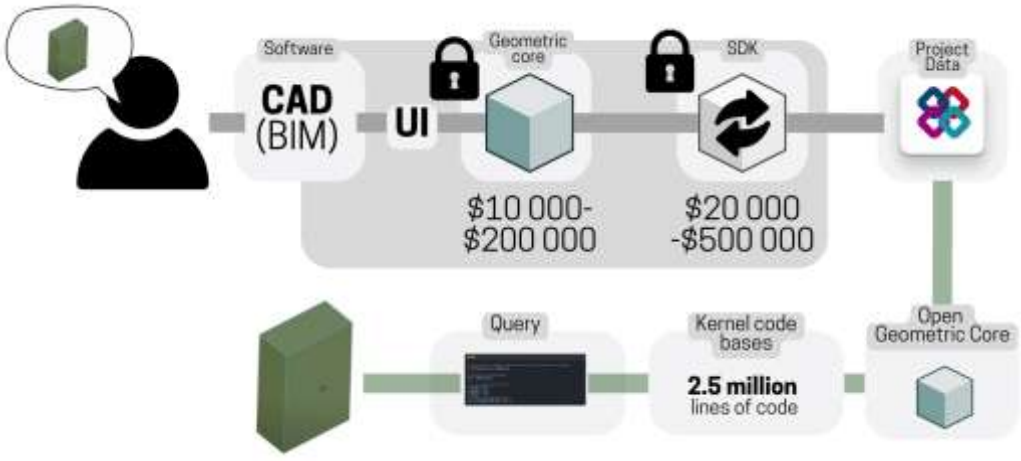


図 6.1-5 今日の CAD- ソフトウェアによるジオメトリの作成は、独自のジオメトリカーネルや SDK を介して行われる。

すべての CAD プログラム、パラメトリックまたは IFC フォーマットで動作するプログラムは、独自のまたは購入したジオメトリカーネルを持っています。そして、IFC -BREP フォーマットのプリミティブ要素であれば問

題はなく、異なるジオメトリカーネルを持つプログラムでもこれらの要素は同様に表示することができるが、ジオメトリカーネルの異なるエンジン の問題以外にも、正しく表示するために独自の特殊性を持つ要素が十分に存在する。この問題は、2019 年に発表された国際的な研究 " A reference study of IFC software support " [117]で詳細に議論されている。

> 同じ標準化されたデータセットでも、相反する結果が得られ共通するパターンはほとんど見いだせない。また、標準をサポートする上で深刻な問題が見つかっている[IFC]。おそらく標準のデータモデルが非常に複雑なためであろう。標準そのものにも責任の一端がある。というのも、標準はしばしば細部を未定義のままにしており、自由度が高く、さまざまな解釈が可能だからである。オブジェクトの組織化と保存において高い複雑性を許容しているため、効果的な普遍的理解、ユニークな実装、一貫性のあるデータ・モデリングには適していない[117]。
>
> - IFC ソフトウェア・サポートの参考調査、2021 年

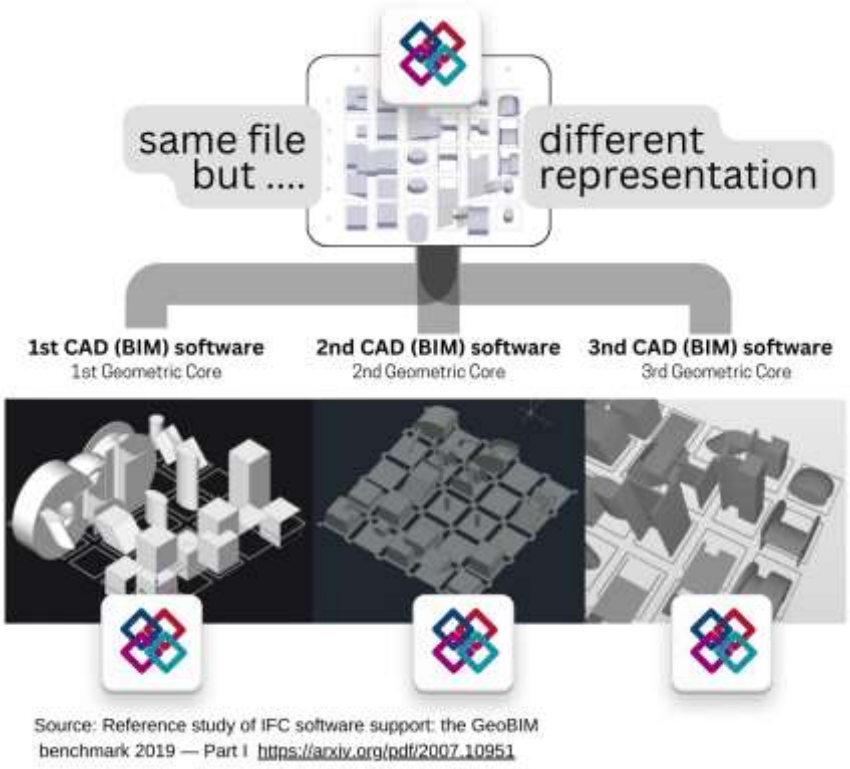


図 6.1-6 異なる幾何学カーネルは、パラメトリックに記述された同じ形状の異なる表現を与える（[117]に基づく）。

特定の条項」を正しく理解できるのは、IFC を開発する特別な組織の有料会員だけである。その結果、IFC の特定の機能に関する重要な知識を得たい人は、大手の CAD- ベンダーと協力するか、独自の研究によってその機能の定性的な考察に到達しようとすることになる

> IFC フォーマット（）を介したデータのインポートおよびエクスポートに関する質問に出くわし、ベンダー仲間に尋ねました：「なぜ IFC ファイルには、構内のパラメトリック転送に関する情報があるのですか？オープンな仕様書には、それについて何も書かれていません」。より詳しい」ヨーロッパのベンダーからの回答：「しかし、それは許可されている」。
>
> - CAD 2021 開発者 [118] のインタビューより

IFC はパラメトリック・プリミティブを通してジオメトリを記述しますが、埋め込まれたカーネルは含んでいません - その役割は、ジオメトリ・カーネルを通してジオメトリをコンパイルする CAD プログラムによって果たされます。ジオメトリカーネルは数学的計算を実行し、交点を定義しますが、IFC はその解釈のためのデータを提供するだけです。IFC に不正な面が含まれている場合、異なるジオメトリカーネルを持つ異なるプログラムは、カーネルに応じて、その面を無視するか、エラーを生成することができます。

> その結果、IFC フォーマットで作業するためには、明確な答えを見つけるのが難しい主要な質問に答える必要がある。

データ品質の問題 と IFC フォーマットの複雑さは、プロセスの自動化、分析、データ処理のために設計データを直接使用することを許さない。そのため、開発者はしばしば、データへの「品質」アクセスを備えたクローズド CAD -ソリューションを使用する必然的な必要性に迫られる[63]。これは、1994 年に IFC を登録したベンダー自身によって書かれている[65]。

ジオメトリカーネルにおける IFC パラメータのマッピングと生成のすべての特殊性は、ジオメトリカーネルでの作業経験を持つ大規模な開発チームによってのみ実現可能です。したがって、IFC フォーマットの特殊性と複雑性に関する現在の慣行は、主に CAD- ベンダーにとって有益なものであり、規格の複雑化が小規模な市場プレイヤ ーにとって障壁となる場合、「採用、拡張、破壊」という大規模ソフトウェア・ベンダーの戦略と多くの共通 点があります [94]。

このような戦略における大手ベンダーの戦略は、オープンスタンダードを適応させ、独自の拡張機能や機能を追加して、自社製品へのユーザー依存を作り出し、競合他社を締め出すしれない。

> IFC フォーマット は、異なる CAD- （BIM- ）システム間の普遍的な架け橋となることを意図しているが、現実には、異なる CAD プラットフォームの幾何学的コア間の互換性問題を示す指標としての役割を果たし、その元となった STEP フォーマットと同様である。

その結果、今日、 IFC オントロジーの完全かつ高品質な実装は、大規模な CAD ベンダー にとって実現可能である。大規模なベンダーは、すべてのエンティティをサポートし、独自の内部ジオメトリコア にマッピングするために多大なリソースを投資することができる。また、大規模なベンダーは、IFC フォーマットの開発組織に最も積極的に参加しているベンダ ーでさえ利用できないような機能の技術的な詳細を、ベンダー間で調整する能力も持っている。

> 相互運用可能なフォーマットの開発をサポートしようと努力している小規模な独立チームやオープンソースプロジェクト にとって、社内にジオメトリカーネルがないことは深刻な問題となります。カーネルがなければ、クロスプラットフォームのデータ交換に関連するさまざまな微妙なニュアンスをすべて考慮することは事実上不可能です。

IFC パラメトリック・フォーマット とオープン BIM コンセプト の発展に伴い、建設業界では、データとプロセス管理におけるオントロジー とセマンティクス の役割について議論が活発化している。

、オントロジー

1990 年代セマンティック・インターネットのアイデアや、IFC フォーマット（ ）の開発に携わった組織の努力により、セマンティクスとオントロジー（ ）は、2020 年代半ばまでに建設業界で議論される標準化の重要な要素の一部となった。

> セマンティック技術とは、異種データの大アレイを統一、標準化、修正し、複雑な検索を実行することである。

OWL （Web Ontology Language）は、RDF グラフ -トリプレット（Resource Description Framework）として表現され（図 6.1-7）、意味データを格納するために使用される。OWL はグラフ・データ・モデルを指し、その種類については「データ・モデル：データ関係と要素間の関係」の章で詳しく説明した。

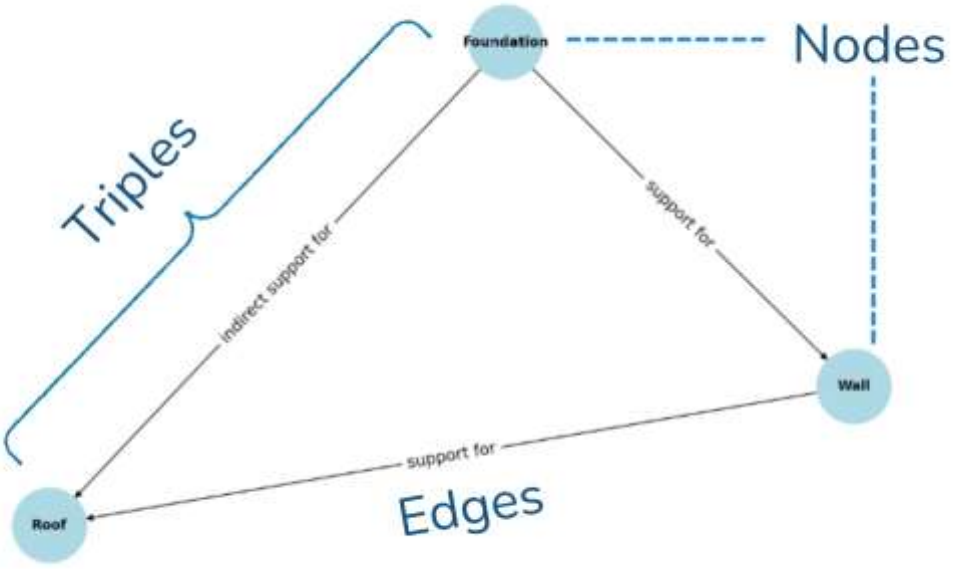


図 6.1-7 RDF データモデル：ノード、エッジ、トリプルは構成要素間の関係を示す。

理論的には、リゾナー（自動論理推論ソフト）の論理推論により、オントロジーから新たな記述を導出することができる。例えば、建築オントロジー に「基礎は壁の支えである」「壁は屋根の支えである」と記録されていれば（図 6.1-7）、リゾナーは自動的に「基礎は屋根の支えである」と推論することができる。

このようなメカニズムは、すべての依存関係を明示的に規定することを避けるため、データ分析を最適化するのに有用である。ただし、新しい知識を生み出すのではなく、すでに知られている事実を特定し、構造化するだけである。

> セマンティクスは、それ自体が新たな意味や知識を生み出すわけではないし、この点で他のデータ保存・処理技術より優れているわけでもない。リレーショナル・データベースのデータをトリプレットとして表現しても、より意味のあるデータにはならない。テーブルをグラフ構造に置き換えることは、データモデルを統一し、検索を容易にし、安全に編集するためには有効かもしれないが、データを「賢く」するものではない。

複雑なセマンティック技術を使わなくても、データの論理的関係を整理することができる（図 6.1-8）。従来のリレーショナル・データベース（SQL ）や CSV や XLSX フォーマットでも、同様の依存関係を構築することができる。例えば、列挙型データベースでは、「屋根のサポート」フィールドを追加し、壁を作成する際に屋根

と基礎を自動的に関連付けることができます。このアプローチは、RDF（）、OWL（）、グラフ、リゾナー（risoner）を使用せずに実装されており、データの保存と分析のためのシンプルで効率的なソリューションである。

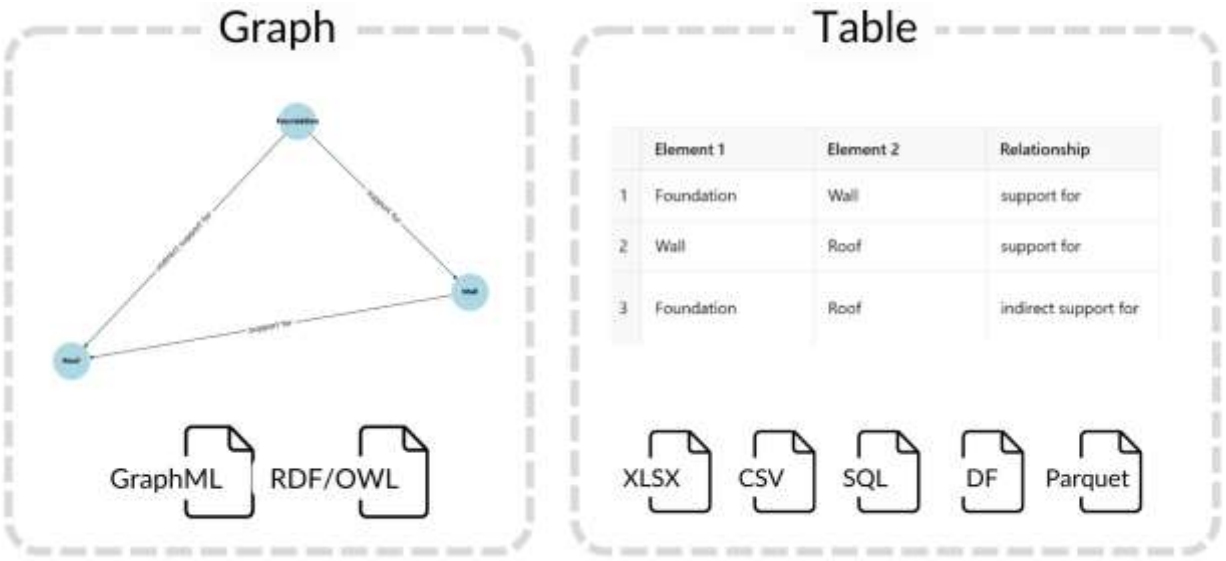


図 6.1-8 同じ論理関係を表すグラフと表のデータモデルの比較。

1990 年代後半に有望視されたセマンティック・ウェブのコンセプト を、多くの大手建設会社と IFC フォーマット開発組織 [94] が踏襲することを決定したことは、建設業界における標準の開発に大きな影響を与えた。

しかしながら、セマンティック・ウェブ という概念そのものが、元来インターネットのために意図されたものであるにもかかわらず、その本来の環境においてさえ広く採用されていないというパラドックスがある。RDF や OWL が開発されているにもかかわらず、本格的なセマンティック・ウェブ は当初の構想のまま登場しておらず、その誕生はすでに望み薄である。

建設業界でセマンティック技術が期待に応えられない理由

他の業界では、セマンティクスを利用する技術の限界に直面している 。ゲーム業界では、ゲームオブジェクトとそのインタラクションをオントロジー（） で記述しようとしても、変化のダイナミクスが大きいため、効果がないことが判明している。その結果、XML（）や JSON（）のような、よりシンプルなデータ形式とアルゴリズムによるソリューションが好まれるようになった。不動産分野でも状況は同様で、用語の地域差や市場の頻繁な変化により、オントロジーの使用は過度に複雑であることが判明し、RETS [119]のようなシンプルなデータベースや標準の方が、データ交換の課題に対処しやすかった。

> マークアップの複雑さ、労働集約的なサポート、開発者のモチベーションの低さといった技術的な困難が、セマンティック・ウェブ（）や他の経済分野での採用を遅らせた。RDF (Resource Description Framework)は大規模な標準にはならず、オントロジー は複雑すぎて経済的に正当化されないことが判明した。

その結果、グローバルなセマンティック・ウェブ（ ）を構築するという野心的なアイデアは実現しなかった。オントロジー（）や SPARQL（）など、セマンティック・ウェブ・テクノロジーのいくつかの要素は企業向けソリューションに採用されているものの、単一の包括的データ構造を構築するという当初の目標は達成されていない。

コンピュータがコンテンツの意味を理解できるインターネットというコンセプトは、技術的に難しく、商業的にも採算が合わないことが判明した。このため、このアイデアを支持した企業は、最終的にその使用を個々の有用なツールに絞り込み、RDF や OWL をインターネット全体ではなく、高度に専門化された企業のニーズに残したのである。過去 20 年間のグーグルトレンド（図 6.1-9）を分析すると、セマンティックウェブ の展望はもうないのかもしれない。

> 不必要に実体を増やす必要はない。ある現象について、論理的に矛盾のない説明がいくつかあり、それが同じようにうまく説明できるのであれば、他のすべての条件が同じであれば、その中で最も単純なものを選ぶべきである。
>
> 剃刀

一般的な構造化クエリー（SQL 、Pandas 、Apache® ）を使ってデータを処理できるのに、なぜトリプレットやライザー、SPARQL を使うのだろうか？エンタープライズ・アプリケーションでは、SQL はデータベースを扱うための標準である。それに対して SPARQL は、複雑なグラフ構造と専門的なソフトウェアを必要とし、Google のトレンドによれば、開発者の興味を惹かない。

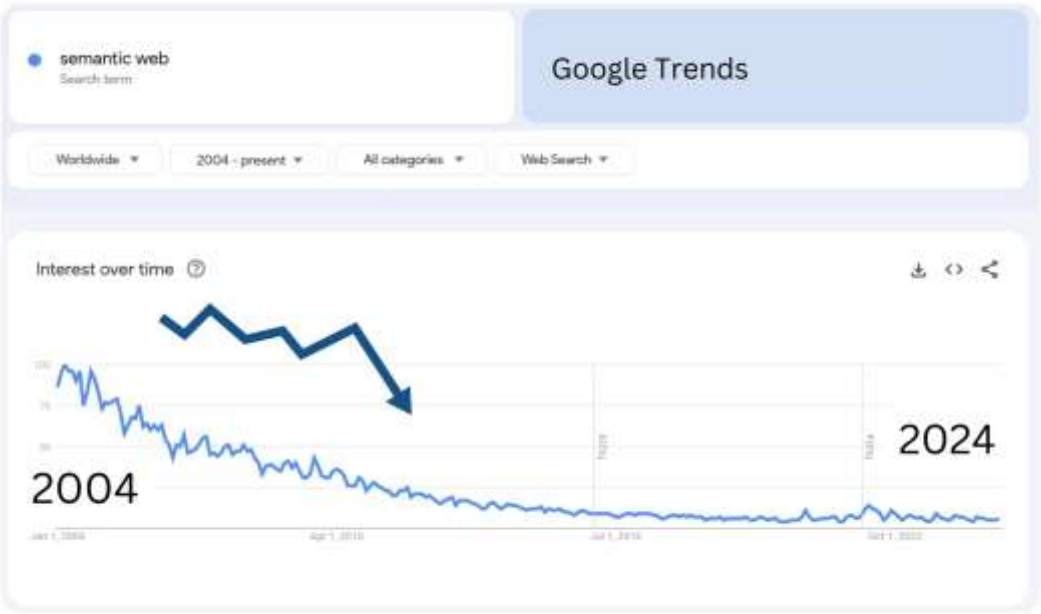


図 6.1-9 グーグル統計による「セマンティック・インターネット」クエリの関心度。

グラフデータベース（ ）や分類木は、場合によっては有用であるが、日常的なタスクの多くでは、その適用が必ずしも正当化されるわけではない。その結果、知識グラフの作成やセマンティック・ウェブ技術 の利用は、異なるソースからのデータを統一したり、複雑な論理的結論を実現する必要がある場合にのみ意味を持つ。

> 表からグラフデータモデルへの移行は、検索を改善し、情報の流れを一元化するが、データを機械にとってより意味のあるものにはしない。問題なのは、セマンティック技術を使うべきかどうかではなく、どこで本当に違いを生み出すかである。オントロジー 、セマンティクス 、グラフデータベース をあなたの会社で導入する前に、どの会社がすでにこれらの技術をうまく使っていて、どこで失敗したかを調べてください。

野心的な期待にもかかわらず、セマンティック技術は建設業界におけるデータ構造化の普遍的な解決策とはならなかった。実際のところ、これらの技術は普遍的な解決策をもたらすことなく、新たな複雑さを加えるだけであった。このような努力は、セマンティック・インターネットのコンセプトが実現しなかった野望と同じであり、期待が現実をはるかに超えていた。

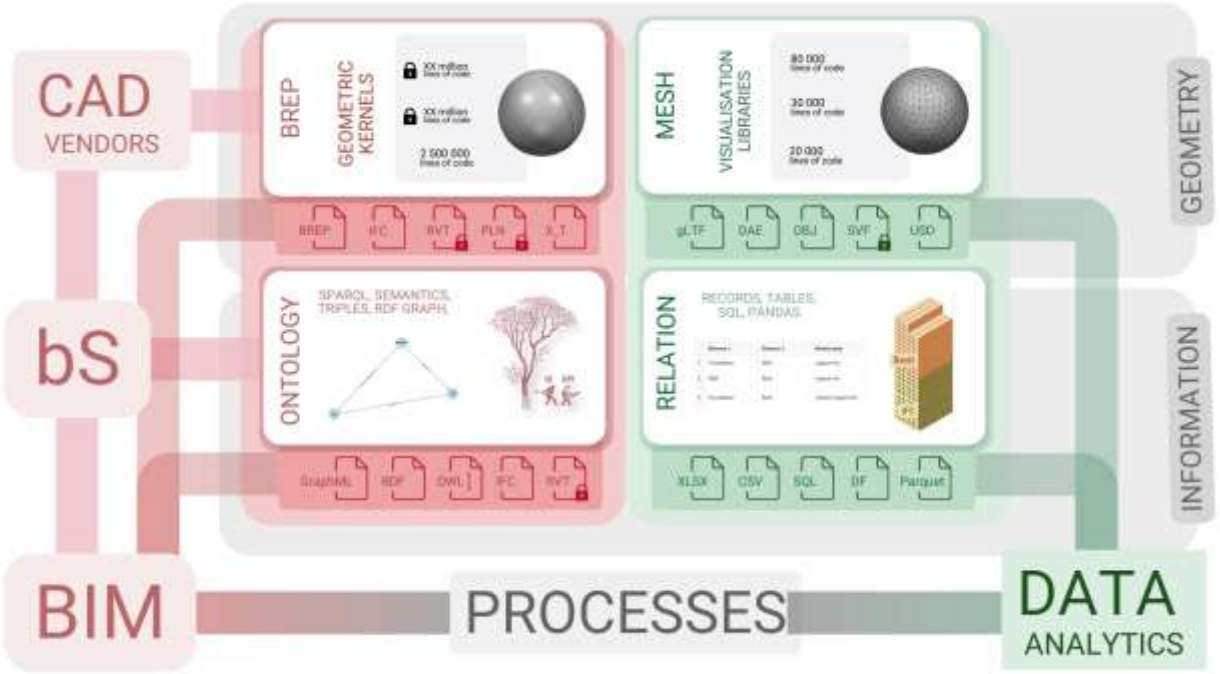


図 6.1-10 建設プロセスにおける幾何学と情報：複雑な CAD と BIM- システムから、分析用の簡素化されたデータ まで。

IT 業界では、セマンティック・ウェブ（）の失敗は、新しい技術（ビッグデータ、IoT 、機械学習 、AR/VR ）の出現によって相殺されてきたが、建設業界にはそのような機会はない。

プロジェクト要素間のデータ関係を伝達するためにコンセプトを使用するという課題に加え、データの可用性という根本的な問題も残っている。建設業界はいまだにクローズドなシステムが主流であり、データを扱うこと、情報を共有すること、プロセスの効率を改善することが難しい。

建設におけるデジタルソリューションの発展を妨げる重要な障壁のひとつとなりつつあるのは、データの閉鎖性である。オープンな 、統一されたデータ形式が標準となっている IT 業界とは異なり、CAD 分野では （BIM ）各ソフトウェアが独自の形式を使用しているため、閉鎖的なエコシステムが形成され、ユーザーを人為的に制限している。

章 6.2.

クローズド・プロジェクトのフォーマットと相互運用性の問題

閉ざされたデータと生産性の低下：CAD 業界の行き詰まり (BIM)

CAD -システムの専有的な性質が、各プログラムが独自のデータ形式を持つという事実につながっている。このデータ形式は、クローズドで外部からアクセスできないか、RVT 、PLN, DWG, NDW, NWD, SKP, あるいはかなり複雑な変換プロセスを通じて半構造化形式で利用可能 - JSON, XML (CPIXML), IFC, STEP and ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV,...など。.

同じプロジェクトに関する同じデータを保存できる異なるデータフォーマットは、構造が異なるだけでなく、内部マークアップのバージョンも異なるため、開発者はアプリケーションの互換性を確保するために考慮する必要がある。例えば、2025 年の CAD フォーマットは 2026 年の CAD プログラムで開くことができるが、同じプロジェクトが 2025 年以前に利用可能であったかもしれない CAD プログラムのすべてのバージョンで開くことは決してない。

> データベースへの直接アクセスを提供しないことで、建設業界のソフトウェア・プロバイダは、専門家（設計エンジニアやデータ・マネージャー）がデータにアクセスし、インポートおよびエクスポートするために使用しなければならない独自のフォーマットやツールを作成することが多い。

その結果、基本 CAD （BIM ）や関連ソリューション（例：ERP/PMIS ） ）のベンダーは、製品の使用料を絶えず値上げしており、一般ユーザーは、フォーマット[63]によるデータ転送の各段階で、ユーザー自身が作成したデータの接続、インポート、エクスポート、および作業に対して「手数料」を支払うことを余儀なくされている。

一般的な CAD - （BIM- ）製品からクラウドストレージのデータにアクセスするコストは、2025 年には 1 トランザクションあたり 1 ドルに達する[120]と予想され、中規模企業向けの建設 ERP -製品のサブスクリプションは、年間 5 桁および 6 桁の金額に達する[121]。

> 現代の建設ソフトウェアの本質は、自動化や効率化ではなく、特定の高度に専門化されたソフトウェアを理解するエンジニアの能力こそが、建設プロジェクトのデータ処理の質とコスト、そして建設プロジェクトを請け負う企業の利益と長期的な存続に影響を与えるということだ。

何十ものシステム、何百もの工程で使用されるシステムであるデータベース CAD [63]にアクセスできず、その結果、個々の専門家間の質の高いコミュニケーションが欠如しているため、建設業界は、生産性という点で、経済の中で最も非効率な部門のひとつに数えられている[44]。

CAD- （BIM- ）設計アプリケーション、新しいシステム（ERP ）、新しい建設技術や材料の出現の過去 20 年間で、 建設業界全体の生産性は 20%低下している（図 2.2-1）一方、データベースへのアクセスに大きな問題がない経済の全セクター とマーケティングのような BIM コンセプトの全体的な生産性は 70%増加している（製造業では 96%）[122]。



図 6.2-1 建設業界では、数十の部門と数百のプロセスが依存するプロジェクトデータが孤立し、複雑であるため、意思決定のスピードは他の業界に比べて数倍遅い。

しかし、CAD ソリューション間の相互運用性 を作り出すための代替アプローチの孤立した例もある。2018 年に開始された SCOPE プロジェクト [123]を持つヨーロッパ最大の建設会社は、CAD- (BIM-)システムの古典的な論理を超えることがいかに可能であるかを実証している。SCOPE の開発者は、IFC に従属しようとしたり、プロプライエタリなジオメトリカーネルに依存したりする代わりに、API や SDK リバースエンジニアリング を使

用して、さまざまなCADプログラムからデータを抽出し、唯一のオープンソース ジオメトリカーネルOCCTに基づいてOBJやCPIXMLなどの中立的なフォーマットに変換し、さらに建設会社や設計会社の何百ものビジネスプロセスに適用しています。しかし、このようなプロジェクトは、そのアイデアの先進性にもかかわらず、フリーのジオメトリカーネルの限界と複雑さに直面し、モノベンダーソリューションの論理を再現する 1 つの企業の閉じたエコシステムの一部にとどまっている。

クローズド・システムの限界やデータ・フォーマットの違い、そしてそれらを統一するための効果的なツールの欠如により、CAD フォーマットを扱う必要のある企業は、構造や閉鎖性の異なる大量のデータの蓄積に直面している。これらのデータは適切に使用されることなく、アーカイブに保管され、永遠に忘れ去られ、使用されないままになっている。

> 設計段階での多大な努力によって得られたデータは、その複雑さと閉鎖的な性質のために、それ以上使用することができなくなる。

その結果、この30年間、建設業界の開発者は何度も同じ問題に直面することを余儀なくされてきた。新しいクローズドフォーマットや独自のソリューションが登場するたびに、既存のオープンCADシステムやクローズドCADシステムと統合する必要性が生じるのだ。異なるCADやBIMソリューション間の相互運用性を確保しようとするこうした絶え間ない試みは、データ・エコシステムの簡素化や標準化に貢献するどころか、複雑化させるだけだ。

CADシステム間の相互運用性神話

1990年代半ばに、CAD環境における相互運用性開発の主要な方向性が、プロプライエタリなDWGフォーマットの打破であったとすれば、それはOpen DWGアライアンス[75]の勝利と、建設業界全体にとって最もポピュラーな図面フォーマットの実際のオープンに結実した。多くの開発チームが、クローズドCADシステム（クローズドBIM）、IFCフォーマット、オープンソリューション（オープンBIM）の間に、いわゆる「ブリッジ」を作ることに注力しているのだ。これらの取り組みのほとんどは、IFCフォーマットとOCCTジオメトリカーネルの使用に基づいており、異種プラットフォーム間の技術的な橋渡しを提供している。このアプローチは、ソフトウェアツールのデータ交換と相互運用性を大幅に向上させることができる有望な方向性と考えられている。

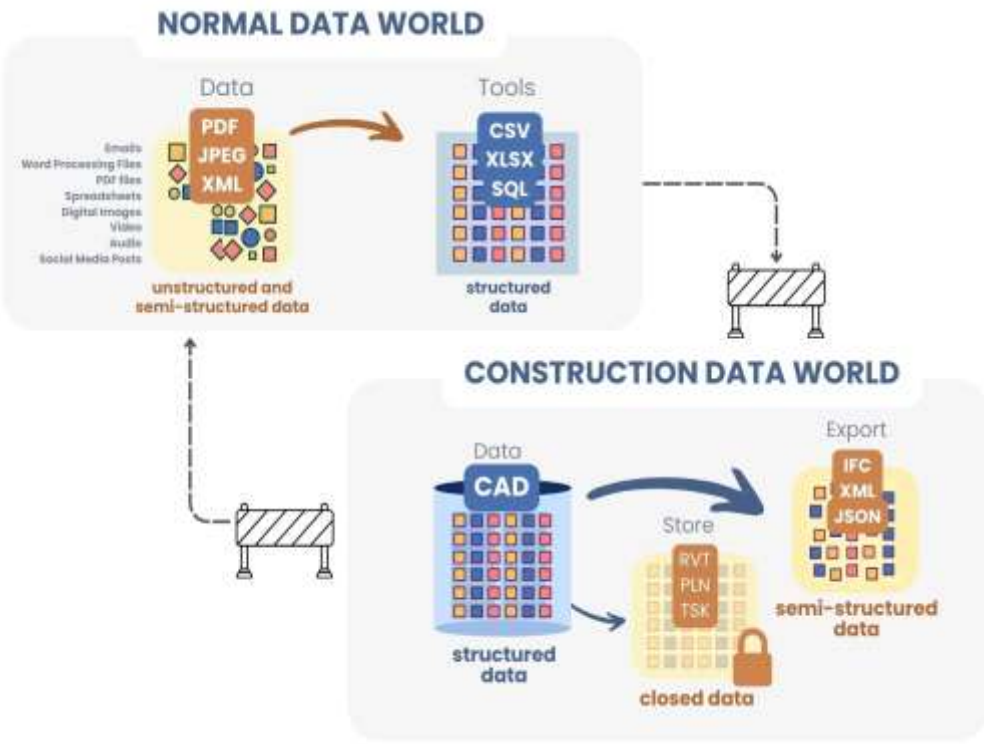


図 6.2-2 他の産業がオープンデータを扱う一方で、建設産業はクローズドな、あるいは緩く構造化された CAD フォーマット （BIM ）を使わざるを得ない。

このアプローチには歴史的な類似性がある。2000 年代、グラフィックエディターの最大手ベンダー（2D world ）の支配に打ち勝とうとした開発者たちは、そのプロプライエタリなソリューションと、GIMP に代わるフリーのオープンソース （図 6.2-3）とのシームレスな統合を実現しようとした。当時も今日と同じように、複雑なパラメータ、レイヤー、内部ソフトウェア・ロジックを維持しながら、クローズド・システムとオープン・システムの橋渡しをしようとする試みだった。

しかし、ユーザーが実際に求めていたのは、シンプルなソリューション、つまり、レイヤーやプログラム・パラメーター（CAD におけるジオメトリック・コアのアナログ）のような過度な複雑さのない、フラットでオープンなデータ だった。ユーザーは、過剰なロジックから解放された、シンプルでオープンなデータフォーマットを求めていた。JPEG、PNG、GIF は、グラフィックスにおいてそのようなフォーマットとなった。今日、JPEG、PNG、GIF はソーシャルネットワーク、ウェブサイト、アプリケーションで使用されている。

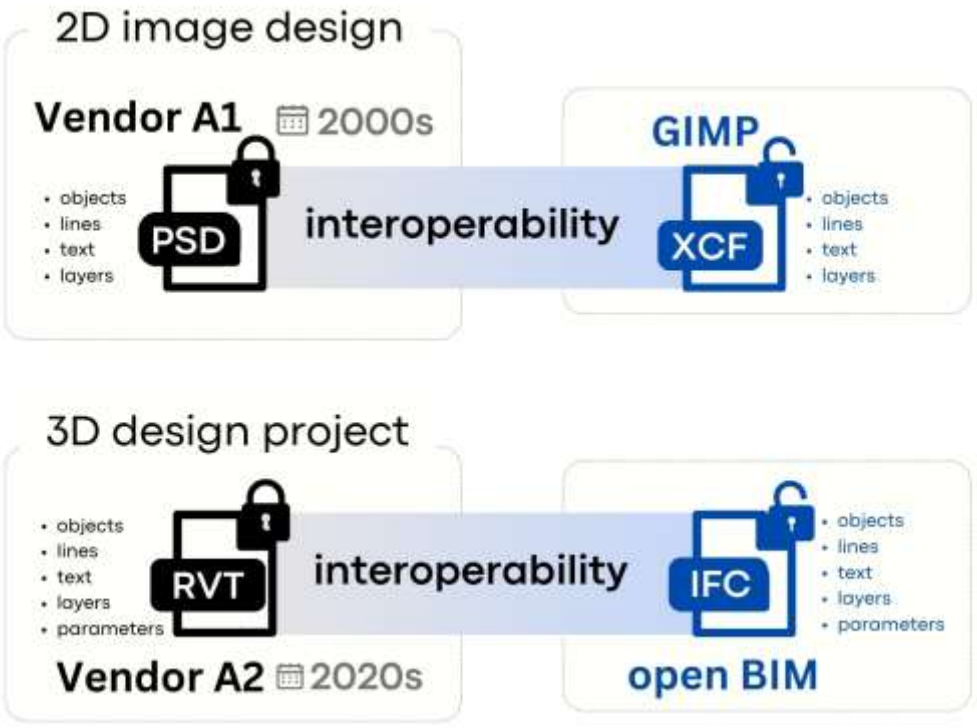


図 6.2-3 建設業におけるデータフォーマットの相互運用性は、2000 年代に一般的なベンダーのプロプライエタリな製品とオープンソースの GIMP を統合しようとした試みからの道筋に似ている。

その結果、今日のイメージング業界では、アプリケーション、Facebook や Instagram のようなソーシャルネットワーク、またはウェブサイトのコンテンツとして、PSD やオープン XCF のようなクローズドフォーマットを使用する人はほとんどいません。その代わりに、ほとんどのタスクでは、使いやすさと幅広い互換性のために、フラットでオープンな JPEG、PNG、GIF フォーマットを使用しています。JPEG や PNG のようなオープンフォーマットは、その汎用性と幅広いサポートにより、様々なプラットフォームで使いやすく、画像共有の標準となっている。同様の移行は、ビデオやオーディオといった他の交換フォーマットにも見られ、MPEG や MP3 といった普遍的なフォーマットが、その圧縮効率と幅広い互換性のために重視されている。このような標準化への動きにより、コンテンツや情報の共有や再生が簡素化され、複数のプラットフォームにまたがるすべてのユーザーがアクセスできるようになった（図 6.2-4）。

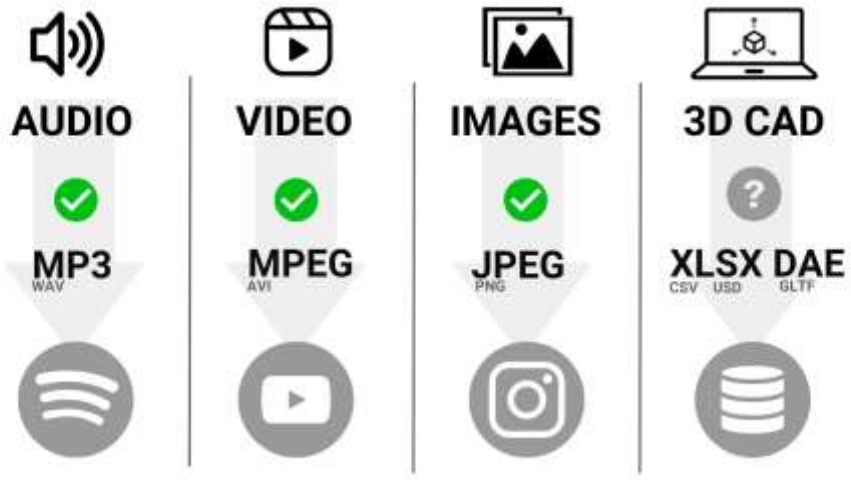


図 6.2-4 複雑な編集機能を持たない簡略化されたフォーマットが、データの共有と利用に普及している。

同様のプロセスは、3D モデリングでも行われている。USD 、OBJ 、glTF、DAE 、 DXF、SQL 、XLSX のようなシンプルでオープンなフォーマットが、CAD 環境外でのデータ交換のためにプロジェクトで使用されることが多くなっています （BIM ）。これらのフォーマットは、複雑な BREP 構造、ジオメトリカーネル またはベンダー固有の内部分類器を操作することなく、ジオメトリとメタデータを含む必要なすべての情報を保存します。大手ソフトウェアベンダーが提供する NWC 、SVF、SVF2、CPIXML 、CP2 などの独自フォーマットも同様の機能を果たしますが、オープンスタンダードとは異なり、クローズドなままです。

> IGES 、STEP 、IFC のような中間的で中立的なパラメトリック・フォーマットの否定というこの考え方が、BIM 白書 を作成し、1994 年に IFC フォーマットを登録した大手 CAD ベンダーによって 2000 年当時から支持されていたことは注目に値する（前章ですでに述べたように、もう一度思い起こす価値がある）。ホワイトペーパー2000「統合された設計と製造」[65]の中で、CAD ベンダーは次のように強調している。[65]において、CAD ベンダーは、情報の完全性と正確性を維持するために、中間トランスレータやパラメトリック形式を使用することなく、ソフトウェア環境内で CAD データベースにネイティブにアクセスすることの重要性を強調している。

建設業界は， CAD データベースにアクセスするためのツールや，その強制的なリバースエンジニアリング，あるいは，CAD プラットフォーム （BIM ）以外で使用するための共通の簡略化されたデータフォーマットの採用について，まだ合意していない．例えば、建設部門で事業を展開する中欧やドイツ語圏の大企業の多くは、ERP システムで CPIXML フォーマットを使用している [121]。この独自フォーマットは、XML の一種である で

あり、形状やメタデータを含む CAD（BIM）プロジェ クトデータを、単一の整理された簡略化された構造に統合している。前章で取り上げた SCOPE プロジェクト（）のように、大規模な建設会社も独自の新しいフォーマットとシステムを構築している

パラメトリック CAD フォーマットや複雑なパラメトリックファイル IFC (STEP)の閉じたロジックは、ほとんどのビジネスプロセスにおいて冗長です。ユーザーは、USD, CPIXML, XML &OBJ, DXF, glTF, SQLlite, DAE & XLSX のような簡素化されたフラットなフォーマットを求めています。XLSX のようなフラットフォーマットは、必要な要素情報をすべて含んでいますが、冗長な BREP ジオメトリロジック、ジオメトリカーネルへの依存、特定の CAD や BIM の内部分類などの負担はありません（図 6.2-5）。6.2-5).

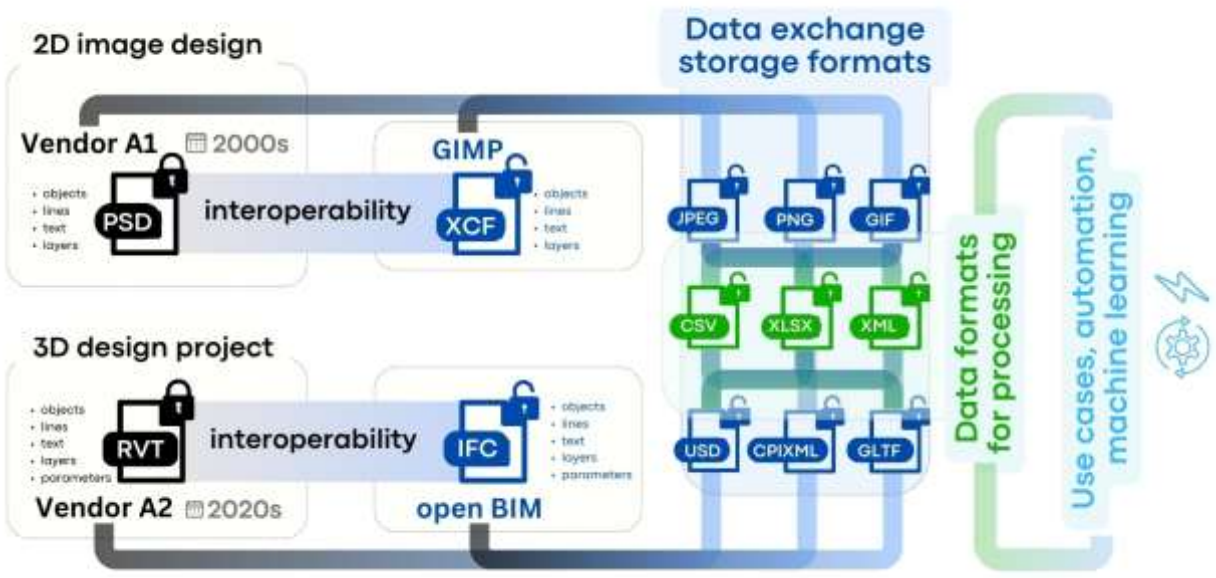


図 6.2-5 ほとんどのユースケースでは、ユーザーはベンダーのプログラムに依存しない、可能な限り単純な フォーマットを選択する。

JPEG、PNG、GIF のようなフラットな画像フォーマットの登場は、ベンダーの内部エンジンの冗長なロジックから解放され、グラフィックを処理し利用するための何千もの相互運用可能なソリューションの開発に拍車をかけた。これにより、レタッチやフィルタリングツールから、Instagram、Snapchat、Canva のようなソーシャルネットワークに至るまで、特定のソフトウェア開発者に縛られることなく、この単純化されたデータを利用できるアプリケーションが出現した。

> 設計 CAD の標準化と簡素化 - フォーマットは、建設プロジェクトで作業するための多くの新しいユーザーフレンドリーで独立したツールの出現を刺激する。

閉じたジオメトリカーネルに縛られたベンダーアプリケーションの複雑なロジックから、単純化された要素のライブラリに基づく普遍的なオープンフォーマットへの移行は、より柔軟で透明性の高い効率的なデータハンドリングのための前提条件を作り出します。これにより、設計者から顧客、保守サービスまで、建設プロセスに関わるすべての関係者が情報にアクセスできるようになります。

とはいえ、今後数年のうちに、CAD のベンダーが の相互運用性と CAD データベースへのアクセスに関する議論を再び転換させようとする可能性は高い。それはすでに、粒状データ、インテリジェント・グラフ、「連携モデル」、クラウド・リポジトリにおけるデジタル・ツインといった「新しい」概念や、BIM やオープン BIM の道を継続する業界アライアンスや標準の作成に関するものとなるだろう。魅力的な用語とは裏腹に、このような構想は、プロプライエタリなエコシステム内にユーザーを留まらせるためのツールとなる可能性がある。一例として、2023 年からのクロスプラットフォーム CAD（BIM）コラボレーションのための「新標準」としての USD （Universal Scene Description）フォーマットの積極的な推進が挙げられる。

米ドルへ と粒度データ m

2023 年の AOUSD アライアンス[124]の登場は、建設業界における重要な転機を意味する。私たちは、CAD ベンダーによって形成された、いくつかの重要な変化を通じて建設データを扱う新しい現実の始まりを目の当たりにしている。最初の大きな変化は、CAD データに対する認識に関するものである。概念設計の初期段階に携わる専門家たちは、CAD 環境で設計を行うことは出発点に過ぎないということを、ますます認識するようになっています。設計プロセスで生成されたデータは、最終的にはオブジェクトの分析、操作、管理の基礎となる。つまり、従来の CAD ツールの枠を超えたシステムで、データにアクセスし、使用できるようにしなければならないのです。

それと並行して、大手デベロッパーのアプローチにも革命が起きている。業界をリードする CAD- ベンダーであり、BIM コンセプト と IFC フォーマット の生みの親である同社は、その戦略において予想外の転機を迎えている。2023 年以降、同社は従来の個別ファイルでのデータ保存から脱却し、粒度の細かい（正規化され構造化された）データを扱うことに重点を置き、データ中心のアプローチに移行する[125]。

> ほとんどのユーザーは、クローズドな CAD フォーマット（PSD に似ている）や複雑なパラメトリック IFC ファイル（レイヤーロジックを持つ GIMP に似ている）を必要としていない。彼らが必要としているのは、CAFM（コンストラクション・インスタグラム）、ERP（フェイスブック）、その他 Excel スプレッドシートや PDF ドキュメントで埋め尽くされた何千ものプロセスで使用できるシンプルなオブジェクト画像です。

建設業界における現在のトレンドは、パラメトリックで複雑なフォーマットから、より普遍的で独立したフォーマットである USD、GLTF、DAE、OBJ（メタ情報は、ハイブリッド内と、個別の構造化または疎構造化フォーマットの両方で）へと、徐々にシフトする段階を設定する可能性があります。かつて 1990 年代半ばに IFC を積極的に推進した大手設計会社を含む歴史的リーダーたちは、現在、新しい USD フォーマット[93]を公然と推進し、そのシンプルさと多用途性を強調している（図 6.2-6）。製品への USD の大量採用、GLTF の互換性、Blender、Unreal Engine、Omniverse のようなツールへの積極的な統合は、データを扱うための新しいパラダイムの始まりの可能性を示しています。ヨーロッパで人気のある ERP（ ）で使用されている、ヨーロッパのフラットな USD フォーマットである CPIXML（ ）のようなローカライズされたソリューションの人気と並んで、中央ヨーロッパにおける USD の地位を強化する可能性があります。IFC フォーマット（ ）の開発に関与している組織は、すでにその戦略を USD に適応させている[126]。

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

図 6.2-6 IFC と USD フォーマットの技術仕様の比較.

> このような背景から、USD は事実上の標準になる可能性を秘めており、主に既存の CAD - （BIM-）フォーマットの複雑さと、その解釈の幾何学的カーネル への依存に関連する、多くの現在の限界を克服することが期待されている。

パラメトリックで複雑な CAD - フォーマットや IFC - の代わりに、単純化されたデータフォーマット USD, gLTF, DAE, OBJ と、要素のメタ情報を CSV, XLSX, JSON, XML で提供することで、そのシンプルさと柔軟性から、建設業界における地位を確立していくだろう。

建設業界における現在の変化は、一見すると、時代遅れの IFC から、より近代的な USD への移行に伴う技術的なブレークスルーのように見える。しかし、2000 年当時、IFC を開発した同じ CAD ベンダーが、その問題点とデータベースへのアクセスの必要性について書いており[65]、現在では新しい標準である USD への移行を積極的に推進していることを考慮する価値がある。

> オープンデータ」USD や、CAD ベンダーが推進し始めたクラウドベースのアプリケーションによる、きめ細かなデータ管理のための「新しい」コンセプトというまた別の見せかけの裏には、プロジェクトデータ管理を独占しようとするベンダーの意図が隠されているかもしれない。

主な事実[93]を分析すると、こうした変更の主な目的は、ユーザーの利便性というよりも、40 年間一度も CAD データベースへのアクセスを提供できなかったベンダーの利益のために、エコシステムとデータの流れを管理し続けることにあることがわかる 。

おそらく今こそ、企業はソフトウェア・ベンダーからの新しいコンセプトを待つのをやめ、データ中心の方向での自己開発に集中すべき時なのだろう。リバースエンジニアリングツールによってデータアクセスの問題から解放された業界は、新たなコンセプトを押し付けることなく、データを操作・分析するための現代的で自由かつ便利なツールへと独自に移行することができるだろう。

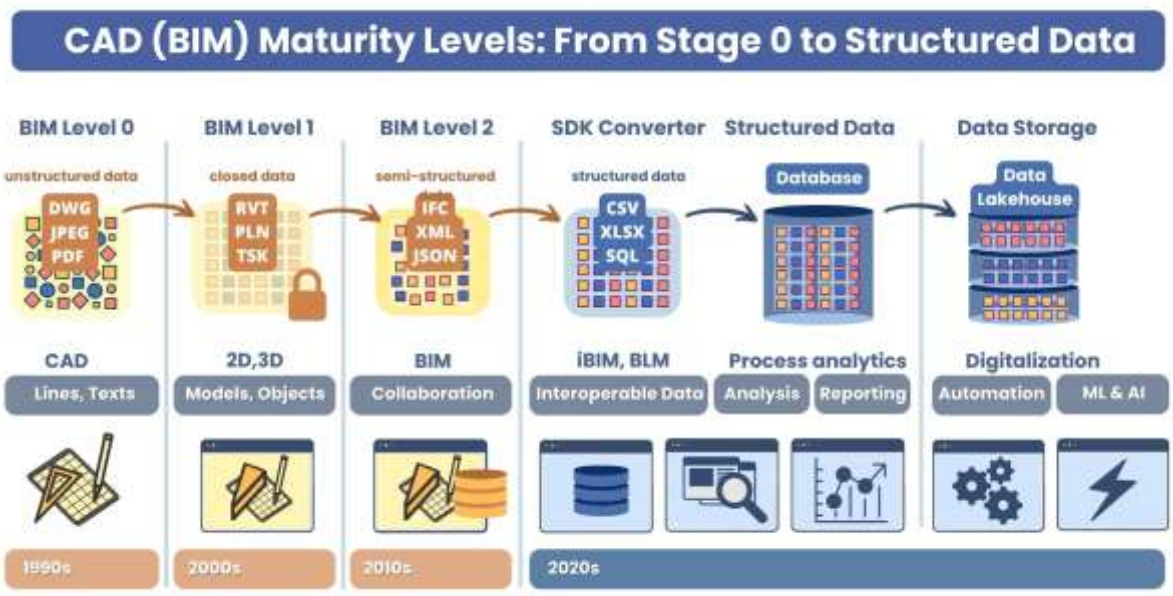


図 6.2-7 CAD 成熟度レベル (BIM): 非構造化データから構造化データおよびリポジトリへ。

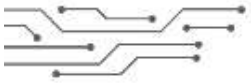
データベースへのアクセス 、オープンデータやフォーマットは、ベンダーがそのプロセスを遅らせよう としよ

うとも、建設業界では必然的に標準となる。より多くの専門家がオープン・フォーマット、データベース・ツール、利用可能なリバース・エンジニアリング SDK（）に精通し、CAD データ -システムへの直接アクセスが可能になれば、この移行のペースは大幅に増加する可能性がある[92]。

> 未来は、オープンで統一され、分析的にアクセス可能なデータ にある。ベンダーのソリューションへの依存を避け、クローズドなエコシステムの人質となることを避けるため、建設・エンジニアリング企業は遅かれ早かれ、オープン性と独立性に依存し、データを完全にコントロールできるフォーマットとソリューションを選択しなければならなくなるだろう。

今日、建設業界で生み出されているデータは、将来、ビジネス上の意思決定における重要なリソースとなるだろう。それは、建設会社の発展と効率性を促進する戦略的な「燃料」として機能する。建設業界の未来は、データ形式やモデルの選択ではなく、データを扱う能力にある。

オープンフォーマットの USD 、glTF、DAE 、OBJ とプロプライエタリなパラメトリック CAD フォーマットの違いを理解するには、ビジュアライゼーションと設計計算において最も複雑で重要なデータ要素のひとつであるジオメトリとその生成プロセスについて考えることが重要です。また、ジオメトリデータがどのように解析の基礎となるかを理解するためには、 、建設における計算を行うには、ジオメトリの生成、変換、保存のメカニズムを深く調べる必要がある。



章 6.3.

建築における幾何学：線から立方メートルまで

線がお金に変わるとき、あるいはなぜ建築家にジオメトリーが必要なのか？

建設におけるジオメトリは、視覚化だけでなく、正確な定量計算の基礎でもある。プロジェクトモデルにおいて、ジオメトリは、長さ、面積、体積などの重要な体積特性を要素パラメータのリスト（図 3.1-16）に補足します。これらの体積パラメータ値は、ジオメトリカーネル（ ）を使って自動的に計算され、見積もり、スケジュール、リソースモデルの出発点となります。本書の第 5 部、「 建設プロジェクトのコスト計算と見積もり」の章ですでに述べたように、最新の ERP、PMIS -システム/ジオメトリの基本的な役割は、設計段階だけでなく、プロジェクトの実施管理、スケジュール管理、予算管理、運用においても、CAD -モデルから得られるオブジェクト群の体積パラメータである。数千年前、エジプトのピラミッドを建設する際、プロジェクトの精度がキュービットやキュビットのような長さの尺度によって左右されたように、今日、CAD -プログラムにおける形状解釈の精度は、予算や納期から請負業者の選定や配送物流に至るまで、結果に直接影響する

> 競争が激しく、予算に制約のある環境では、形状に直接依存する容積計算の精度が、生き残り要因になる。最新の ERP（）システムは、CAD（）や BIM（）モデルから得られる正しい容積特性に直接依存している。だからこそ、要素の正確な幾何学的記述は、単なる視覚化ではなく、建設コストとタイミングを管理するための重要なツールなのだ。

歴史的に、幾何学はエンジニアリング・コミュニケーションの主要言語であった。パピルスに描かれた線からデジタルモデルに至るまで、図面や幾何学的表現は、設計者、現場監督、見積担当者間の情報交換の手段として役立ってきた。コンピュータが登場する以前は、定規や分度器を使って手作業で計算が行われていた。CAD ソフトウェアの幾何学カーネルは、線と点を三次元体に変換し、そこから必要な特性がすべて自動的に抽出されます。

CAD (BIM) プログラムのユーザーインターフェースを通じて、CAD プログラムで作業しながら、計算用の幾何学的要素を作成する。点および線を体積体に変換するために、幾何学カーネルが使用されます。このカーネルは、幾何学の体積モデルへの変換という重要なタスクを実行し、近似後にそこから要素の体積特性が自動的に計算されます。

線から体積へ：面積と体積がデータになるまで

エンジニアリングの実務では、体積と面積は、解析的に記述された幾何学的サーフェス（ ）から計算されるか、BREP（境界要素表現）の枠組みの中で NURBS （nonuniform rational B-splines）のようなパラメトリック・モデルを通して計算される。

> NURBS (Non-Uniform Rational B-Splines)は、曲線や曲面を記述する数学的な方法である。一方、BREP は、境界を含むオブジェクトの完全な 3 次元ジオメトリを記述するためのフレームワークであり、NURBS を使用して定義することができる。

BREP と NURBS は精度が高いにもかかわらず、強力な計算資源と複雑なアルゴリズムを必要とする。しかし、このような数学的に正確な記述から直接計算することは、しばしば計算上困難であるため、実際には、テッセレーション（サーフェスを三角形のグリッドに変換すること）がほとんどの場合使用され、その後の計算を簡素化します。テッセレーションとは 複雑なサーフェスを三角形や多角形に分割することである。CAD /CAE 環境では、この方法は視覚化、体積計算、衝突探索 、MESH のようなフォーマットへのエクスポート、衝突解析に使用されます。自然界の例としては、複雑な形状を規則的なグリッドに分解した蜂の巣があります（図 6.3-1 ）。

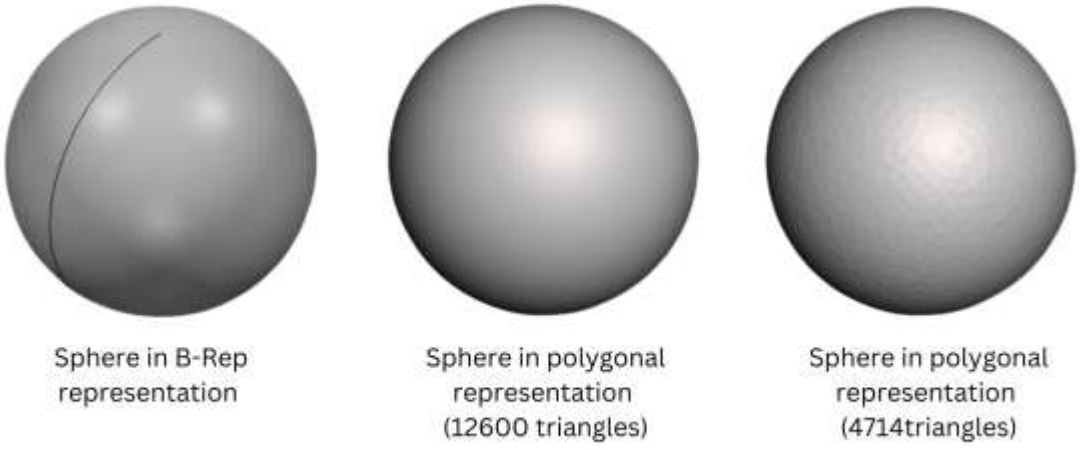


図 6.3-1 同じ球をパラメトリック記述した BREP と、三角形の数を変えたポリゴン表現。

BREP （NURBS ）は、CAD で使用されているが、形状の基本モデルではない。円や有理スプラインを表現し、ジオメトリデータの保存を最小限にするための便利なツールとして作られた。しかし、これには限界があり、

例えば、らせん状の線や曲面の根底にある正弦波を正確に記述できないことや、複雑な幾何学カーネル を使用する必要がある。

対照的に、三角メッシュとパラメトリック形状のテッセレーションは、シンプルで、メモリを効率的に使用し、大量のデータを処理できることが特徴です（図 6.3-2）。これらの利点により、幾何学的形状を計算する際に、複雑で高価な幾何学カーネル（）と、それに埋め込まれた何千万行ものコードを使わずに済む。

> ほとんどの建築物では、パラメトリックモデル（BREP、IFC）やポリゴン（USD、glTF、DAE、OBJ）によって、ボリュメトリック特性をどのように正確に定義するかは問題ではありません。ジオメトリは、NURBS であろうと MESH であろうと、常に形状の近似的な記述である。

多角形や BREP （NURBS）として定義される幾何学は、連続的な形状の近似的な記述で近似する方法にある程度とどまる。フレネル積分に厳密な解析式がないように、ポリゴンや NURBS による離散化は、三角メッシュと同様に、常に近似である。

BREP フォーマットのパラメトリックジオメトリは、主に最小限のデータサイズが重要で、その処理と表示にリソース集約的で高価なジオメトリカーネルを使用できる場合に必要です。多くの場合、これは CAD - プログラムの開発者に特徴的で、この目的のために MCAD - ベンダーのジオメトリカーネルを製品に適用します。この場合、これらのプログラム内でも、視覚化と計算のためのテッセレーション の過程で、BREP モデルはしばしば三角形に変換されます（PSD ファイルが JPEG に簡略化されるのと同様です）。

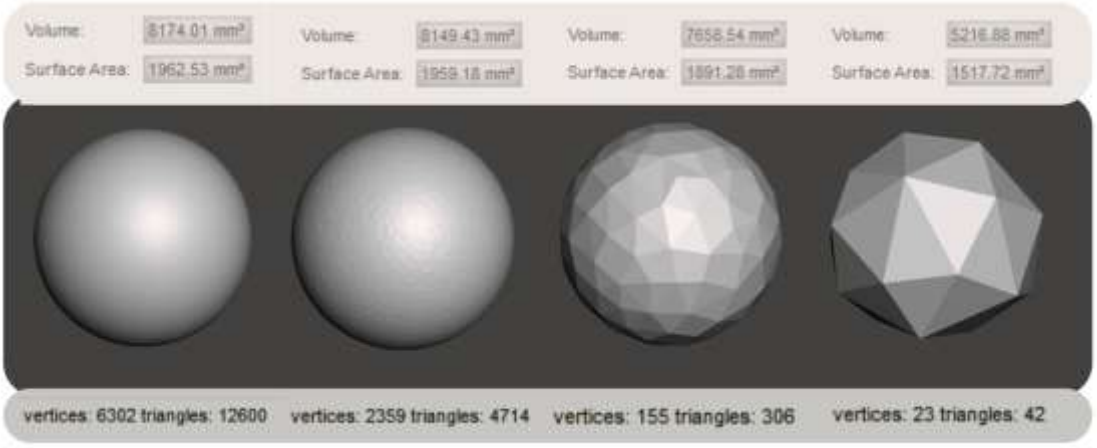


図 6.3-2 ポリゴン数の違いによる体積特性の違い。

ポリゴン MESH やパラメトリック BREP には、それぞれ長所と限界がありますが、ユーザーのタスクを念頭に置いて形状を記述するという目的は同じです。最終的に、幾何モデルの精度は、その表現方法だけでなく、特

定のタスクの要件にも依存します。

> ほとんどの建設問題では、パラメトリック幾何学や複雑な幾何学カーネルの必要性は冗長かもしれない。

それぞれの計算自動化タスクにおいて、パラメトリックジオメトリの重要性が、自社のソフトウェア製品の販売促進を目的とする CAD 開発者によって誇張されていないかどうかを検討する価値がある。

MESH、USD、ポリゴンへの移行：ジオメトリにテッセレーション を使う

建設業界では、設計情報やフィーチャー・ジオメトリを扱うストリーミング、システム開発、データベース、自動化プロセスにおいて、特定の CAD エディタやジオメトリ・カーネル からの独立性を確保することが重要です。

> 計算部門 と建設現場の両方で使用される交換フォーマットは、特定の CAD-（BIM-）プログラムに基づいてはならない。幾何学的情報は、幾何学的コア や CAD アーキテクチャーを参照することなく、テッセレーションによって直接フォーマットで表現されるべきである。

CAD からのパラメトリックジオメトリは、中間的なソースとして考えることはできるが、普遍的なフォーマットの基礎として考えることはできない。ほとんどのパラメトリック記述（BREP と NURBS を含む ）は、いずれにせよ、さらなる処理のためにポリゴン MESH に変換される。結果が同じ（テッセレーションとポリゴン）であり、処理がより単純であれば、その選択は明らかです。これは、グラフオントロジーと構造化テーブルの選択（第 4 回で取り上げた）に似ています。過剰な複雑さが正当化されることはほとんどありません（図 3.2-10, 図 6.1-8）。

OBJ 、STL、glTF、SVF 、CPIXML 、USD 、DAE などのオープン・フォーマットは、普遍的な三角形メッシュ構造を使用しているため、大きな利点があります。これらのフォーマットは相互運用性に優れており、数百万行のコードを含む複雑な特殊ジオメトリカーネルを必要とせず、利用可能なオープンソースライブラリを使用して簡単に読み取り、視覚化することができます（図 6.3-3）。これらの汎用的なジオメトリフォーマットは、IKEA™ の比較的単純なキッチンデザインツールから、映画や VR アプリケーションの複雑なオブジェクトビジュアライゼーションシステムまで、さまざまなアプリケーションで使用されています。重要な利点は、これらの形式を扱うためのフリーでオープンソースのライブラリが多数あり、ほとんどのプ ラットフォームとプログ

ラミング言語で利用できることです。

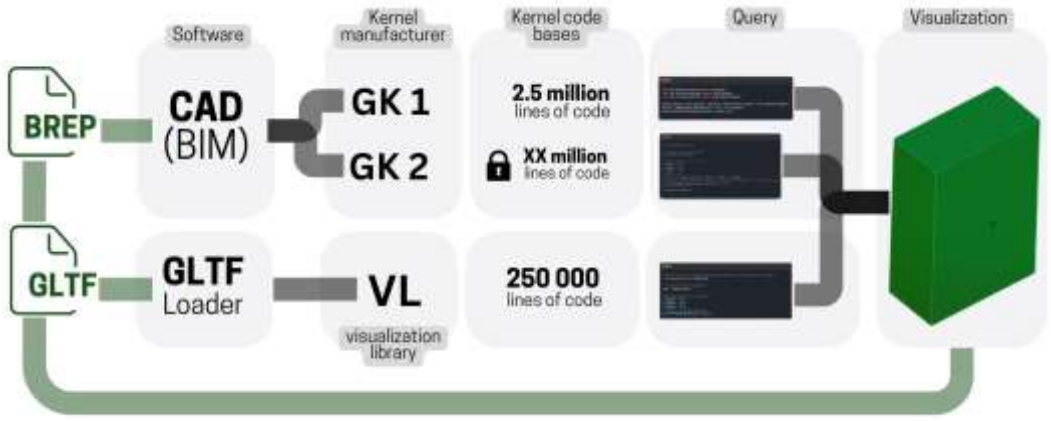


図 6.3-3 パラメトリックフォーマットとジオメトリカーネル の使用、または三角形フォーマットとオープンソースの視覚化ライブラリの使用によって、同じジオメトリ表現が実現される。

ユーザー自身だけでなく、CAD -ベンダーも、幾何学カーネルが異なるために、外国のパラメトリック CAD フォーマットやオープン IFC を解釈する際の問題に直面している。実際には、すべての CAD -ベンダは例外なく、リバース・エンジニアリング SDK を使用してシステム間のデータ転送を行っており、相互運用性の目的で IFC や USD [93]のようなフォーマットに依存しているところはありません。 。

> CAD- ベンダーのアライアンスが推進する、彼ら自身が使用していない概念を使用するのではなく CAD ソリューションの開発者とユーザーは、特定のコンテキストにおける各アプローチの利点を理解することに集中し、ユースケースに応じて 1 つまたは別のタイプの形状を選択する方が生産的です。異なる幾何学的表現を選択することは、精度、計算効率、および特定のタスクの実用的なニーズの間のトレードオフである。

従来、大手ベンダーが設計データを処理する際に建設業界に課してきた幾何学カーネルの使用に伴う複雑さは、しばしば冗長であることが判明する。MESH ジオメトリに基づく USD フォーマットは、業界にとって一種の "パンドラの箱 "となり、CAD ベンダーに典型的な IFC やパラメトリック BREP 構造の枠組みの外で、開発者がデータ交換を組織化するための新たな可能性を開くことができる。

USD、DAE、gLTF、OBJ などの構造をよく見てみると、複雑なパラメトリックやクローズドな幾何学カーネルに頼ることなく、幾何学的情報の転送と利用を効率的に整理できる、よりシンプルでオープンなフォーマット

があることがわかる。このアプローチは、開発者の技術的な参入の敷居を低くするだけでなく、デジタル建設のための柔軟でスケーラブルな、真にオープンなソリューションの開発を促進します。

LOD, LOI, LOMD - CAD (BIM) における ディテールのユニークな分類。

幾何学的な表現形式に加えて、さまざまな業界がさまざまな詳細レベルやデータの深さを使用する世界では、CAD - (BIM-) 方法論は、建築モデルに情報を提供するためのアプローチを構造化する、独自の分類システム を提供する。

標準化に対する新しいアプローチの一例として、グラフィカルコンポーネントと情報コンポーネントの両方の準備と信頼性の程度を反映する、モデル開発のレベルの導入がある。CAD - (BIM-) データを用いた作業における情報内容の差別化のために、LOD (Level Of Detail) - モデルのグラフィカルな部分の詳細レベル、および LOI (Level Of Information) - データの精緻化レベルが登場した。さらに、統合されたアプローチのために、LOA (Level of Accuracy) - 表現された要素の精度、LOG (Level of Geometry) - グラフィカル表現の精度を決定するための概念が導入された。

詳細度（LOD ）は、モデル開発の程度を反映し、100 から 500 までの数字で示される。LOD 100 は、一般的な形状と寸法を持つ概念的なモデルである。LOD 200 は、より正確な寸法と形状を含むが、条件付きの詳細である。LOD 300 は、正確な寸法、形状、要素位置を持つ詳細モデルです。LOD 400 は、要素の製作と設置に必要な詳細情報を含んでいます。LOD 500 は、建設後の施設の実際の状態を反映し、運用とメンテナンスに使用されます。これらのレベルは、3D 、4D 、5D およびそれ以降を含む、ライフサイクルのさまざまな段階における CAD（BIM）モデル情報の飽和の構造を表しています。

実際のプロジェクトでは、高い詳細レベル（LOD400）は過剰であることが多く、LOD100 ジオメトリやフラットな図面でも十分である。例えば、スペースや部屋要素（部屋要素カテゴリ）は、視覚的なジオメトリを持たないが、多くのビジネスプロセスが構築される重要な情報やデータベースを含んでいる場合がある。

> したがって、設計を開始する前に、必要な詳細レベルを明確に定義することが重要です。4D -7D のユースケースでは、DWG 図面 と最小限の LOD100 ジオメトリでも十分なことが多い。要求プロセスにおける重要なタスクは、モデルの豊富さと実用性のバランスを見つけることです。

要するに、CAD（BIM）データをデータベースと見なせば（実際そうなのだが）、新しい略語によるモデル飽和の説明は、概念レベルから始まり物理レベルで終わる、情報システムのための段階的なデータモデリングに

ほかならない（図 6.3-4）。本書の第 3 部と第 4 部で詳しく説明した。LOD と LOI が増加するごとに、計算、建設管理、運転といった新しいタスクに必要な情報が追加され、様々なパラメーターの形で追加情報レイヤー（3D -8D ）によってモデルが順次強化されていく。

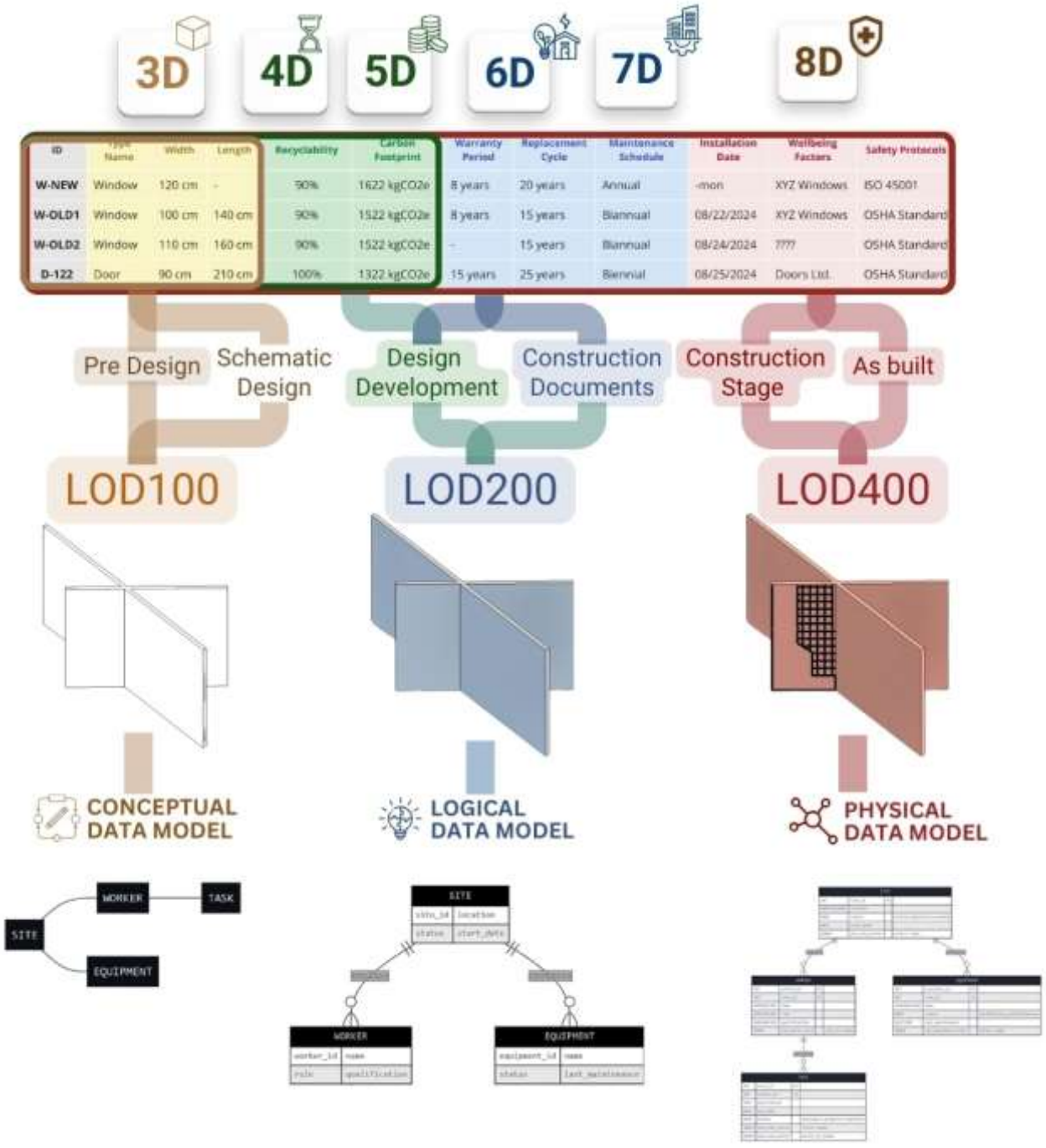


図 6.3-4 プロジェクトの詳細を具体化するプロセスは、概念的データモデルから物理的データモデルへのデータモデリングと同じである。

ジオメトリは設計データの一部に過ぎず、その必要性は建設プロジェクトでは必ずしも正当化されない。CAD（ ）データを扱う際の重要な問題は、モデルをどのように視覚化するかということよりも、これらのモデルからのデータを CAD（BIM ）プログラムの外でどのように利用できるかということである。

2000 年代半ばまでに、建設業界は、管理システムやデータ処理システム、特に CAD 部門 （BIM ）からのデータ量の急激な増加という前例のない課題に直面した。このデータ量の劇的な増加は、企業の管理者を驚かせ、データ品質と管理に対する要求の高まりに備えることができなかった。

新しい CAD 標準（BIM ） - AIA, BEP, IDS, LOD, COBie

CAD データベース（）へのオープンアクセスがないこと（）や、データ処理市場における競争が限られていることを利用し、BIM（ ）という新しい略語に関連したマーケティング・キャンペーンを利用して、CAD データを扱うアプローチの開発に携わる組織は、データ管理実務の改善を目的とした、事実上あるべき新しい標準や概念を作り始めた。

> CAD ベンダーやデベロッパー （BIM ）が直接的または間接的に支援するほとんどすべてのイニシアチブは、ワークフローの最適化を目的としているが、その結果、さまざまな利害関係者によって働きかけられた多数の標準が生まれ、建設業界をデータプロセスに関する曖昧さや混乱に導いている。

LOD、LOI、LOA、LOG に加えて、近年建設業界で登場した新しいデータ標準をいくつか挙げてみよう：

- BEP (BIM 実行計画) - CAD (BIM) をどのように統合し、プロジェクトで活用するかを説明し、データの取り扱い方法とプロセスを定義する。
- EIR 文書 /AIA （発注者の情報要件 ） - 入札募集前に発注者が作成し、請負業者が情報を準備・提供するための要件 が記載されている。各プロジェクトにおける BEP の基礎となる。
- AIM （資産情報モデル）は、BIM プロセス の一部である。プロジェクトが引き渡され、完成すると、データ・モデルは資産情報モデルまたは AIM と呼ばれる。AIM の目的は、実現された資産を管理、維持、運用することである。
- IDS (Information Delivery Specification) - の要件を定義し、建設プロジェクトのさまざまな段階で、どのようなデータが、どのような形式で必要とされるかを示す。

- iLOD は、BIM -モデルにおいて情報が表現される詳細 LOD のレベルである。これは、基本的な幾何学的表現から詳細な仕様やデータまで、モデル内の情報がどれだけ詳細で完全であるかを定義する。
- eLOD - LOD CAD モデル （BIM ）の個々の要素の詳細レベル。これは、各要素がモデル化される範囲と、寸法、材料、性能特性、その他の関連属性などの関連情報を定義します。
- APS (Platform Services) および主要 CAD ベンダーのその他の製品 (BIM) - リンクされたオープンデータモデルを作成するために必要なツールとインフラ について説明します。

CAD（BIM）標準（LOD、LOI、LOA、LOG、BEP、EIR、AIA、AIM、IDS、iLOD、eLOD など）を導入する宣言された目的は、データ管理の質を向上させ、自動化機能を拡張することであるが、実際には、それらの使用は、しばしばプロセスの過度の複雑化と断片化につながる。CAD(BIM)モデルを一種のデータベースと考えれば、これらの標準の多くが、情報システムとの連携において他業界で長年使われてきた効果的なアプローチと重複していることは明らかだ。このようなイニシアチブは、簡素化と統一化の代わりに、しばしばさらなる用語の負担を生み出し、真にオープンで柔軟なソリューションの実装を妨げている。

注目すべきは、これらの新しいコンセプトの多くが、本書の最初の部分で詳細に議論され、経済の他のセクターで長い間使用されてきたモデリングやデータ検証プロセスに実際に取って代わっていることである。一方、建設業界では、標準化プロセスはしばしば逆の方向に進む。新しいデータ記述フォーマット、新しい標準、データ検証のための新しいコンセプトが生み出されるが、これらは必ずしも実際の統一性や実用的な適用性にはつながらない。その結果、処理を簡素化・自動化する代わりに、業界は新たなレベルの規制や官僚主義に直面することになり（図 6.3-1）、これは必ずしも効率向上に資するものではない。

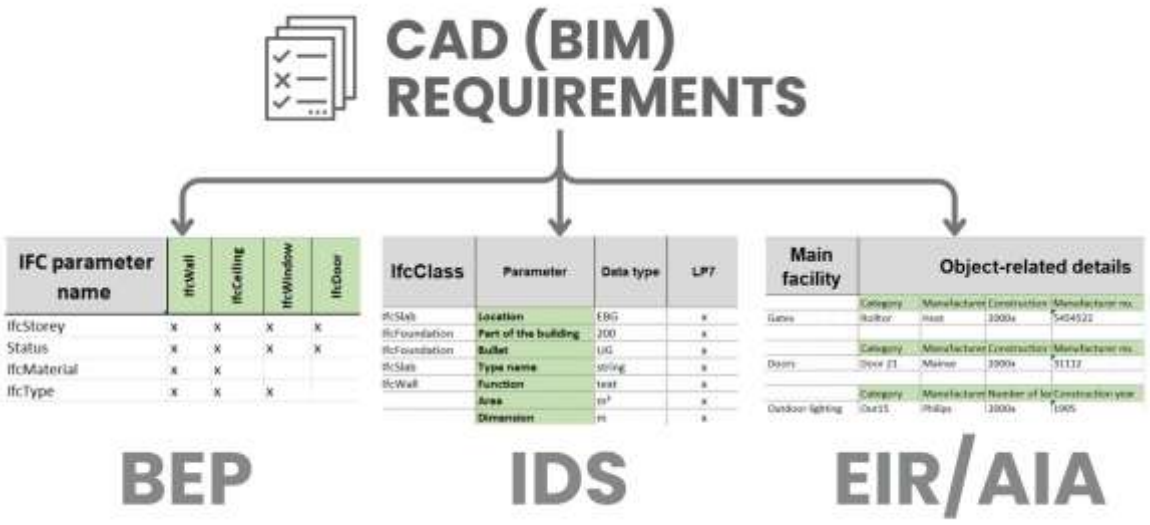


図 6.3-1 データと情報の内容要件は、属性とその境界値の記述に縮小され、表を使って記述される。

> CAD（BIM）データに関連する新しい概念は、データ処理を単純化する代わりに、解釈や基本的な定義の段階ですでにさらなる複雑さや紛争を生み出していることが多い。

新しい概念の最新の例の一つは、オープン BIM 概念（ ）における情報モデルの属性構 成の要件（ ）を記述することを可能にする IDS 形式（ ）（2020 年に導入）（ ）である。IDS 要件は、属性およびその境界値に関する情報を構造化テーブル（Excel または MySQL ）の形式で記述し、これを半構造化 XML 形式 のマークアップに変換する。

> Contrary to the view promoted by vendors and supported by BIM and open BIM, that data handling in construction is unique because of the use of specialised tools such as CAD and BIM, the data formats and data management practices of this industry are not different from those other industries

プロジェクトと CAD (BIM) フォーマットの要件数は、"要件を構造化されたフォームに変換する "の章で詳述されている、属性カラムを持つ単一の要件テーブルを使用することによって簡素化することができ、元々構造化された 要件をテーブル以外のフォーマットに変換する必要はありません（IDS は、最初にテーブルを介して記述されています）。

簡略化されたアプローチ（図 6.3-2）には、前の章（図 4.4-9、図 4.4-16、図 7.3-10）で詳細に説明したエンティティ識別子、プロパ ティ、および境界値の列が含まれており、要件を IDS-XML 形式に変換する必要がない。この方法は、データ品質管理のための直接的で、煩雑でなく、透明性の高いメカニズムを提供する。正規表現（RegEx）からデータフレーム、パンダ（Pandas）、標準的な ETL （ペ イプレーン）に至るまで、経済の他部門の専門家がデータを扱う際に使用するものと全く同じように、広く使用されて いるツールに依存する。

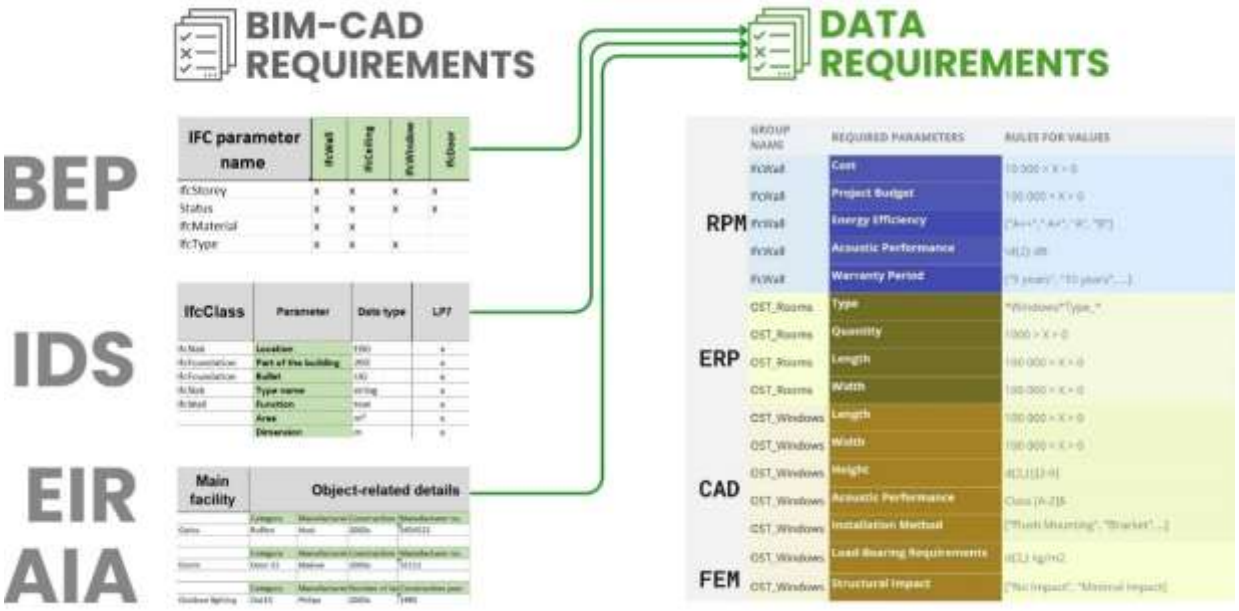


図 6.3-2 他の産業におけるデータ要件は、属性とその境界値の構造化された記述に単純化される

。

建設プロジェクトにおけるデータ は、基本的に他の分野と同じであるが、建設業界では、データの閉鎖的な性質のため、この多様なデータを管理・制御するための新しいアプローチや技術が、時間の経過とともに次々と登場している。他の業界では、データ処理に対する標準化されたアプローチで成功を収めていますが、建設業界では、新しくユニークなデータ形式、 要件、検証コンセプトが開発され続けています。

> 建設業におけるデータの収集、準備、分析に使用される手法やツールは、他の経済分野の専門家が使用するものと基本的に異なるものであってはならない。

業界は、批判的な考察と再評価を必要とする独特の用語生態系を構築してきた：

- STEP フォーマット（ ）は、STEP フォーマット自体の制限を考慮することなく、施工分類によって補完された新しい名称 IFC（ ）の下に位置づけられる。

- パラメトリックフォーマット IFC は、視覚化と計算に必要な統一された幾何学的コアがないにもかかわらず、データ通信プロセスで使用されている。
- データベースへのアクセス CAD -システムは、「BIM 」という用語の下で推進されており、これらのデータベースの具体的な内容やそれらへのアクセスについては議論されていない。
- ベンダーは、IFC（ ）や USD（ ）のフォーマットを通じて相互運用性（ ）を推進しているが、多くの場合、それを実践することなく、コストのかかるリバースエンジニアリングを行い、ベンダー自身が苦労している。
- LOD、LOI、LOA、LOG、BEP、EIR、AIA、AIM、IDS、iLOD、eLOD という用語は、他の業界で長く使用されてきたモデリングや検証ツールを参照することなく、同じエンティティパラメータを記述するために普遍的に使用されている。

特に、専門的なサービスやソフトウェアの販売を通じてデータ処理の各段階を収益化することが主な目的であればなおさらだ。ビジネスの観点からは、何も問題はない。しかし、CAD（BIM）に関連するこのような略語やアプローチが、本当に付加価値を与え、専門的なプロセスを簡素化するものなのかどうかは、まだ未解決の問題である。

> 建設業界では、このようなシステムが機能している。というのも、業界そのものが、このようなシステムや略語の迷路で投機的利益の大半を稼いでいるからだ。透明なプロセスとオープンデータに関心を持つ企業は稀である。この複雑な状況は、顧客、クライアント、投資家、銀行、プライベート・エクイティが、情報管理に対してより明確で情報に基づいたアプローチを求め始めるまで、おそらくいつまでも続くだろう。

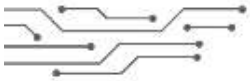
この業界には数多くの略語が蓄積されているが、それらはすべて、程度の差こそあれ、同じプロセスやデータ要件を説明している 。ワークフローを簡素化する上で、それらの本当の有用性には疑問が残る。

コンセプトやマーケティングの略語は生まれては消えていくが、データ要件の検証プロセス自体は、永遠にビジネスプロセスの不可欠な一部であり続けるだろう。建設業界は、 、より多くの専門的なフォーマットや規制を作成する代わりに、金融、工業、IT などの他の分野ですでに効果が実証されているツールに目を向けるべきである。

豊富な用語、頭字語、フォーマットは、デジタル構築プロセスが深く練り上げられたかのように錯覚させる。しかし、マーケティングの概念や複雑な専門用語には、単純だが不都合な真実が隠されていることが多い。データにアクセスするのが難しく、文書化が不十分で、特定のソフトウェア・ソリューションに固く縛られてい

るのだ。

頭字語やフォーマットのためのフォーマットという悪循環から抜け出すには、CAD (BIM) システムを魔法の情報管理ツールとしてではなく、その正体である特殊なデータベースとして見る必要がある。そして、このプリズムを通してこそ、マーケティングがどこで終わり、情報を使った実際の仕事がどこで始まるのかを理解することができるのである。



章 6.4.

CAD 操作のための llm の設計パラメータ化と使用

CAD データの一意性という幻想 (BIM): 分析とオープンフォーマットへの道

最新の CAD (BIM)プラットフォームは、設計と建設情報管理のアプローチを大きく変えました。以前は、これらのツールは主に図面や 3D モデルの作成に使用されていたが、今日では本格的なプロジェクト・データ・リポジトリとして機能している。単一真実情報源（Single Source of Truth）というコンセプトのもと、パラメトリック・モデルは、プロジェクトのライフサイクル全体を通じて、その完全性と妥当性を保証する、プロジェクト情報の主要な、そしてしばしば唯一の情報源となりつつある。

> CAD （BIM ）プラットフォームと他の建設データ管理システムとの主な違いは、情報（唯一の真実の源）にアクセスするために、特別なツールと API が必要なことである。これらのデータベースは、伝統的な意味での普遍的なものではありません。オープンな構造と柔軟な統合の代わりに、特定のプラットフォームとフォーマットに固定された、閉じた環境となっています。

CAD データベース（BIM ）とは一体何なのだろうか。この問いに答えるには、ソフトウェア開発者が押し付ける通常の略語や概念を超える必要がある。その代わりに、プロジェクト情報を扱うことの本質、すなわちデータとその処理に注目する価値がある。

> 建設におけるビジネスプロセスは、CAD（）や BIM（）ツールでの作業からではなく、プロジェクト要件の形成とデータモデリングから始まる。まず、タスクのパラメータが定義される。すなわち、特定のタスクを解決する際に考慮する必要があるエンティティのリスト、その初期特性、境界値である。その後、指定されたパラメータに基づいて、CAD（BIM）システムでモデルと要素が作成されます

CAD - (BIM-) データベースに情報を作成する前のプロセスは、本書の第 4 部「データモデリング：概念モデル、論理モデル、物理モデル」の章で詳しく説明したデータモデリングプロセスと完全に同じである（図 4.3-1 ）。

データモデリングにおいて、後にデータベースで処理するデータの要件（）を作成するように、CAD データベ

ースの管理者は、いくつかのテーブルの列またはキーと値のペアのリスト（）の形で設計要件（）を作成します（図 6.4-1、ステップ 1-2）。そして、API（）を使って自動的または手動でこれらの初期パラメータに基づいてのみ、設計者は CAD（）（BIM（））データベースでオブジェクトを作成し（またはむしろ改良し）（ステップ 3-4）、その後、それらが初期要件に準拠しているかどうかを再度チェックする（ステップ 5-6）。このプロセス（定義、→作成、→検証、→調整（ステップ 2-6））は、データモデリングと同様に、データ品質がターゲットシステム（文書、表、ダッシュボード）に望ましいレベルに達するまで繰り返し行われる（ステップ 7）。

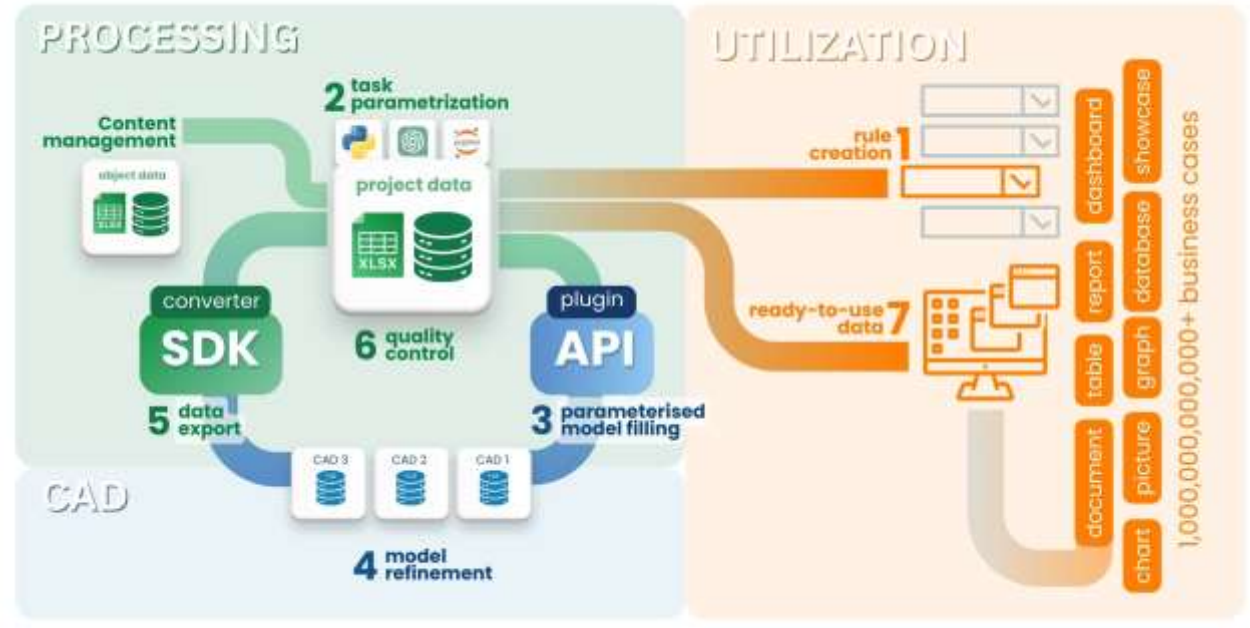


図 6.4-1 建設プロジェクトの実施におけるビジネスプロセスのためのデータベースの情報飽和のサイクル。

CAD (BIM) を、設計環境の外部で定義された要件に基づいて生成されたキーと値のペアのセットの形でパラメータを転送するメカニズムとして考える場合（図 6.4-1、ステップ 1-2）、議論の焦点は、特定のソフトウェアソリューションとその限界から、より基本的な側面、つまりデータ構造、データモデル、データ要件に移ります。本質的には、データベースのパラメータ飽和と古典的なデータモデリングプロセス（ステップ 2-3 と 5-6）について話している。唯一の違いは、CAD データベースの閉鎖的な性質と、使用されるフォーマットの特殊性により、このプロセスには特殊な BIM ツールの使用が伴うことです。他の業界には同様のアプローチがないとすれば、BIM の独自性とは何だろうか？

過去 20 年間、BIM 単なる単一のデータソース以上のものとして位置づけられてきた。CAD -BIM バンドルは、

建設プロジェクトの設計、モデリング、ライフサイクル管理のプロセスを自動化できる、本質的に統合されたデータベースを備えたパラメトリックツール[64]として販売されることが多い。しかし現実には、BIM はデータやプロセス管理の便利な方法というよりも、ユーザーをベンダーのプラットフォームに留めておくためのツールとなっている。

その結果、CAD- (BIM-) データはそれぞれのプラットフォーム内で孤立し、独自の API やジオメトリカーネルの背後にプロジェクト情報が隠されている。このため、ユーザーが独自に データベースにアクセスし、ベンダーのエコシステムをバイパスしてデータを抽出、分析、自動化、他のシステムに転送する能力が奪われている。

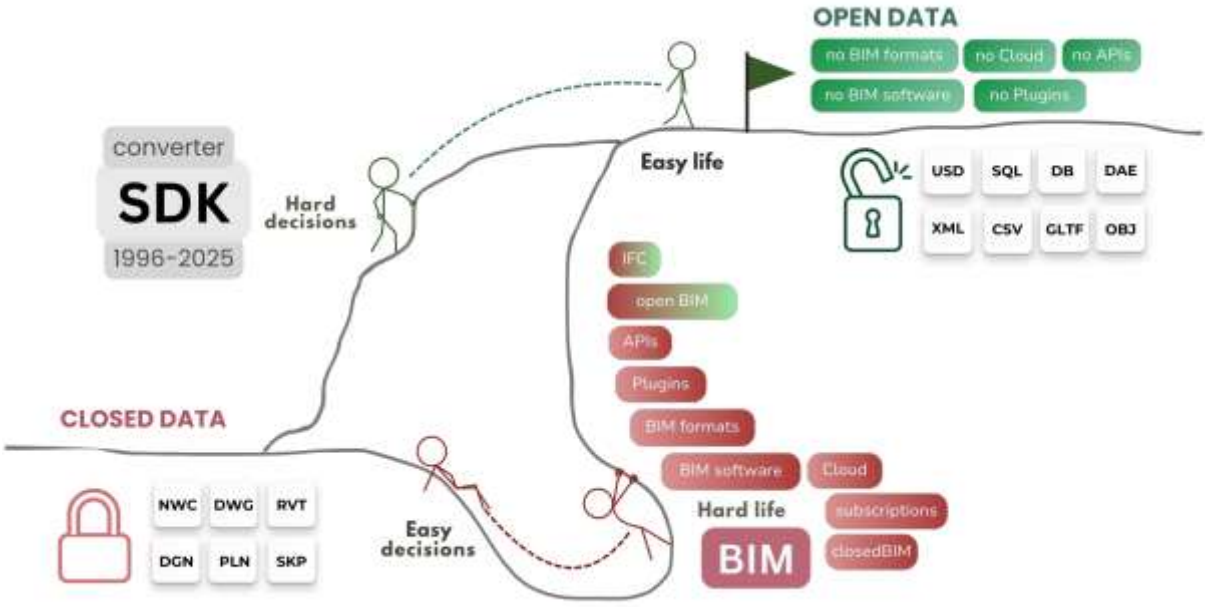


図 6.4-2 建築では、最新のフォーマットには洗練されたジオメトリカーネル 、毎年更新される API 、および CAD -（BIM -）ソフトウェア用の特別なライセンスが必要です。

最新の CAD ツールを使用する企業は、すべての CAD ベンダー自身が例外なく実際に使用している、データ を扱うための同じアプローチを使用する必要があります：SDK - リバース・エンジニアリング・ツール を使用したデータ変換です。CAD データベース（ ）にフルアクセスし、リバースエンジニアリングツールを使用することで、属性を持つエンティティのフラット セットを取得し[127]、設計要素のジオメトリとパラメータの両方を含む、任意の便利なオープンフォーマット（図 6.4-2）にエクスポートすることができます。このアプローチは 、ファイル指向からデータ中心のアーキテクチャへと、情報を扱うパラダイ ムを根本的に変えます：

- RVT 、IFC 、PLN 、DB1、CP2、CPIXML 、USD 、SQLite 、XLSX 、PARQUET などのデータフォーマ

ットは、同じプロジェクトの要素に関する同一の情報を含んでいる。つまり、特定のフォーマットやそのスキーマの知識が、データそのものを扱う際の障害になることはない。

- どのフォーマットからのデータも、幾何学カーネル の制約を受けずに、MESH の三角形ジオメトリ とすべてのオブジェクトエンティティのプロパティを含む、単一のオープンな構造化された粒状構造（図 9.1-10）に結合することができる。
- データ分析普遍性を追求している。オープンデータを使えば、、使用されている形式に関係なくプロジェクト・データを扱うことができる。
- API やベンダープラグインへの依存を最小化：データ操作はもはや API のスキルに依存しない。

、CAD -データ要件が分析しやすい構造化表現形式に変換されると開発者は特定のデータスキーマや閉じたエコシステムに依存しなくなる。

パラメータによるデザイン：CAD の未来 と BIM

世界中のどの建設プロジェクトも、CAD プログラムから始まったことはない。CAD で図面やモデルが形になる前に、それは概念化の段階（図 6.4-1、ステージ 1-2）を通過します。そこでは、将来のオブジェクトの基本的なアイデアとロジックを定義するパラメータに焦点が当てられます。この段階は、データモデリング（図 4.3-6）における概念レベルに相当します。パラメータは設計者の頭の中だけに存在することもありますが、理想的には、構造化されたリストやテーブルの形で配置されたり、データベースに保存されたりします（図 6.4-3）。

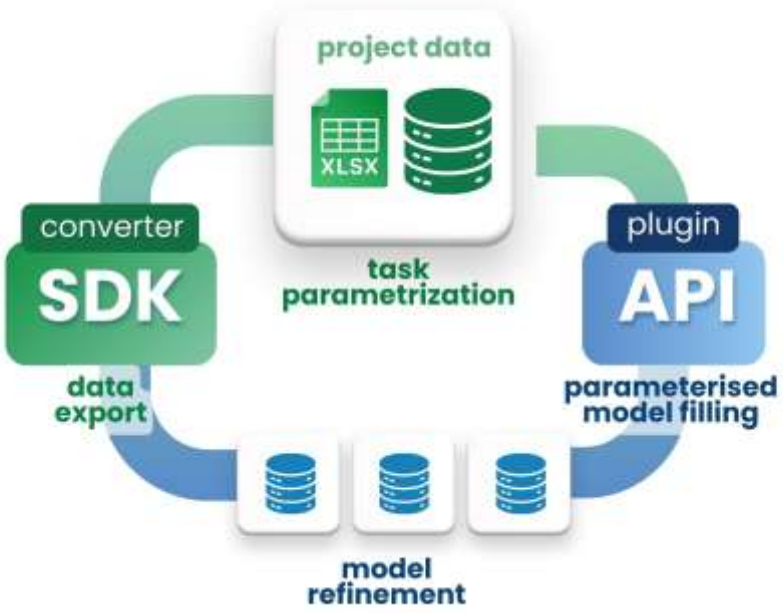


図 6.4-3 デザイン・プロセスは、バリュー・チェーンにおける要求を用いて、CAD データベースに外部からの情報を入力する反復プロセスである。

CAD モデリングそのもの（データモデリングの論理的・物理的フェーズ（図 4.3-7））を開始する前に、プロジェクトの基礎となる境界パラメータを定義することが重要です。これらの属性は、他の要件と同様に、データ利用チェーンの末端（システムなど）から収集され、これらを通じて、プロジェクトにおける将来のオブジェクトの制約、目標、主要な特性がすでに定義されています。

パラメトリックモデリングツール（図 6.4-3）を使えば、要件がきちんと定義されていれば、モデリング自体を 60〜100%完全に自動化することができる。例えば、Grasshopper Dynamo のようなビジュアルプログラミング言語や、最新の CAD -environments に組み込まれているもの、あるいは Blender、UE、Omniverse のようなフリーのソリューションがあります。

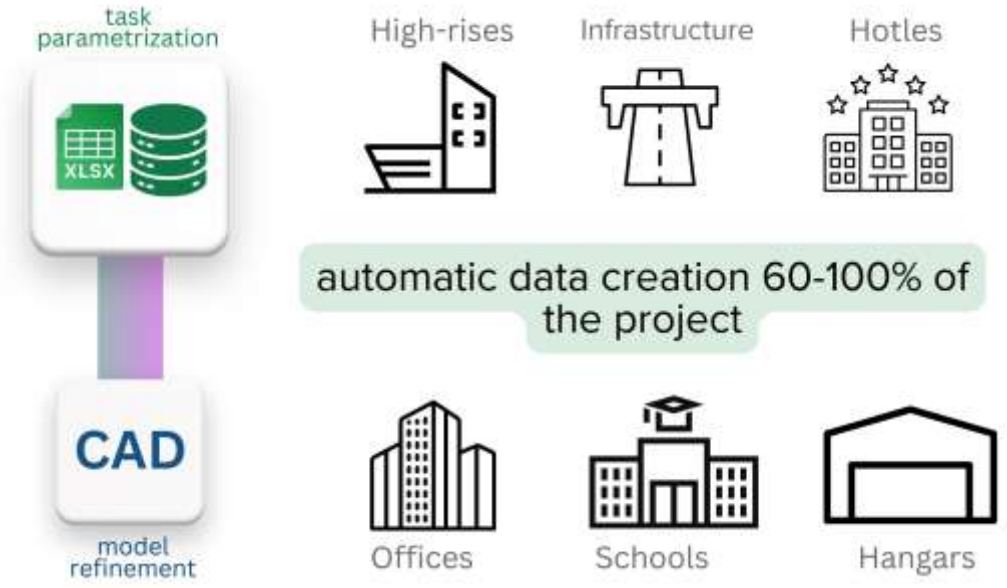


図 6.4-4 今日、パラメトリック・プログラミング・ツールのおかげで、類型化されたプロジェクトの大部分は、すでに完全自動で作成されている。

すでに今日、大規模な工業プロジェクトや類型化されたプロジェクトは、デザイン部門の手によってではなく、パラメトリック・ツールやビジュアル・プログラミングによって作られている。これによって、特定のデザイナーやマネージャーの主観的な判断ではなく、データに基づいてモデルを構築することが可能になる。

> コンテンツはデザインに先行する。内容のないデザインはデザインではなく、装飾である [128]。
>
> - ジェフリー・ゼルドマン（ウェブデザイナー、起業家

このプロセスは、図面や 3D モデリング（ ）から始まるのではなく、要件の形成から始まる。プロジェクトでどの要素を使用するか、どのデータを他の部門やシステムに転送する必要があるかを決定するのは、要件（ ）である。構造化された要件が存在するだけで、モデルを定期的に（たとえば 10 分ごとでも、設計者の作業を中断させることなく）自動的にチェックすることが可能になる。

> おそらく将来的には、CAD- (BIM-) システムは、データベースを埋めるための単なるインターフェイスになり、どの CAD ツールでモデリングが行わ（物理的なレベル）は問題にならなくなるだろう。

同様に、機械工学においても、3D モデリングはしばしば使用されるが、プロジェクトの必須要素ではない。ほとんどの場合、古典的な 2D 文書で十分であり、それを基に必要な情報モデルが作成される。このモデルは、業界標準に従って構造化されたコンポーネントから組み立てられ、設計と生産組織を理解するために必要なすべての情報を含んでいます。次に、工場情報モデルを作成するために、技術者のニーズにすでに方向づけられた特定の製品やフローチャートが追加されます。全プロセスは、不必要に複雑化することなく、3D グラフィックが実際の利点をもたらさないところでシステムに過負荷をかけることなく、整理することができる。

> 3D モデルそのものと CAD - システムが主役であってはならないことを理解することが重要である。実体を記述するジオメトリを除く他のすべてのパラメータは、可能であれば CAD 環境の外で保存および処理されるべきである（BIM ）。

パラメータによる設計は単なるトレンドではなく、建設業界の必然的な未来である。複雑な 3D モデル（ ）を手作業で作成する代わりに、設計者はデータを使って作業し、それを検証し、プロセスを自動化することで、建設はプログラミングの世界に近づいていくだろう。やがて、設計プロセスはソフトウェア開発の原則に基づいて構築されるようになるだろう：

- 要件の作成 → モデルの作成→ サーバーへのアップロード→ 検証 変更→ プルリクエスト
- Pull リクエストは、設計プロセスの前またはプロセス中に作成された要件に対して、モデル・チェックを自動的に実行します。
- データ品質チェック（ ）および承認後、変更はプロジェクト、共通データベースに実装されるか、他のシステムに自動的に転送される。

すでに機械工学では、このような設計変更は変更通知の作成から始まる。同様の仕組みが建設業界にも待ち受けている。設計は、各ステップがパラメトリック要件によってサポートされる反復プロセスとなる。このようなシステムにより、設計者は自動チェックや、特定の要件に対する自動プルリクエストを作成できるようになる 。

> 未来のデザイナーは、何よりもまずデータ・オペレーターであり、マニュアル・モデラーではない。彼の仕事は、ジオメトリが属性のひとつにすぎないパラメトリックなエンティティでプロジェクトを満たすことだ。

この変革において重要な役割を果たすのは、データモデリング、分類（ ）、標準化の重要性を理解することである。将来の設計規則は、XLSX（ ）または XML（ ）スキーマの形で、キーと値のパラメータ・ペアとして形式化される。

> 建設業界の未来とは、データを収集し、分析し、検証し、分析ツール を使ってプロセスを自動化することである。BIM （または CAD ）は最終目標ではなく、進化の段階に過ぎない。専門家が、従来の CAD ツールを迂回し、データを使って直接作業できることに気づけば、「BIM」という言葉自体も、構造化されたきめ細かな建設プロジェクトデータを使うという概念に徐々に変わっていくだろう。

この変革を加速させる重要な要因のひとつが、大規模言語モデル（LLM ）とそれに基づくツールの登場である。これらのテクノロジーは、設計データの扱い方を変え、 API やベンダーソリューションに関する深い知識を必要とせずに情報にアクセスできるようにします。LLM を使用すると、要件 を作成し、CAD データと対話するプロセスが直感的でアクセスしやすくなります。

デザイン CAD における LLM の出現 データ処理プロセス

CAD データベースアクセスツールやオープンで簡素化された CAD -フォーマットの開発に加え、LLM -ツール（大規模言語モデル）の出現は、設計データの処理に革命をもたらしている。以前は情報へのアクセスは主に複雑なインターフェイスを介しており、プログラミングスキルや API の知識が必要でしたが、現在では自然言語を使用してデータと対話することが可能です。

技術的背景を持たないエンジニア、マネージャー、プランナーは、通常の言語でクエリーを作成することで、プロジェクトデータから必要な情報を得ることができます。データが構造化され、アクセス可能であれば（図 4.1-13）、LLM で「10 立方メートル以上の容積を持つすべての壁を、タイプ別にグループ化したテーブルで表示する」というような質問をするだけで、このモデルは自動的にこのクエリを SQL または Pandas のコードに変換し、サマリーテーブル、グラフ、あるいは完成したドキュメントを生成します。

以下は、LLM -モデルが、異なる CAD- (BIM-) フォーマットで表現された設計データとどのように相互作用するかの実例である。

- 表形式のデータフレーム（CHATGP、LlaMa 、Mistral DeepSeek 、Grok 、Claude 、QWEN またはその他）への変換後（図 4.1-13）の RVT 形式の CAD プロジェクトへの LLM チャットでのクエリの例：

> RVT ファイル（）から得られた Dataframe のデータを、"Volume "パラメータを合計する際に "Type name "でグループ化し、グループ内の要素数を表示する。また、これをすべてゼロ値なしの水平ヒストグラムとして表示してください。

- LLM レスポンス 横棒グラフ（PNG 形式）：

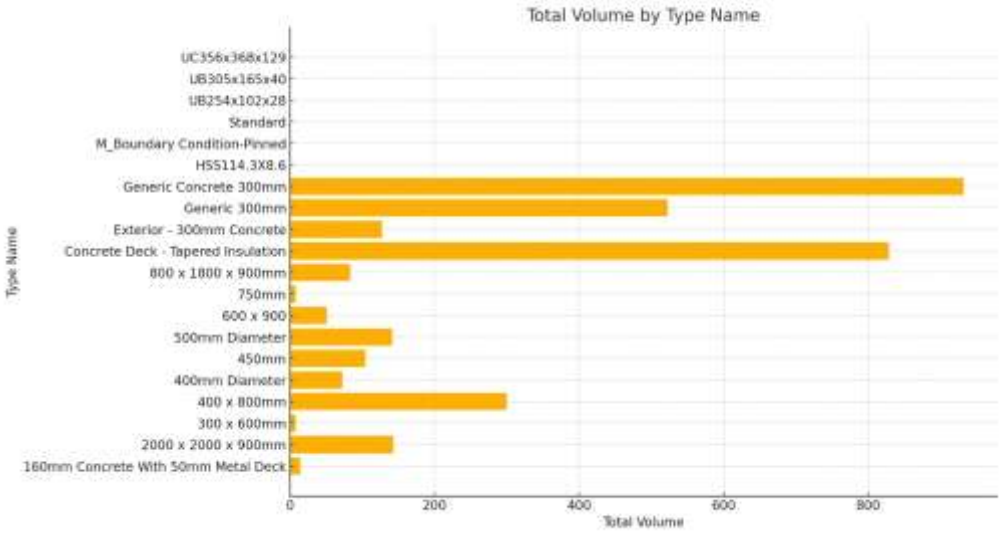


図 6.4-5 プラグインを使うことで、マウスを 17 回クリックしたり、コードを 40 行書いたりする代わりに、LLM では、テキストクエリによって QTO テーブルを即座に取得する。

- 壁」のカテゴリーから、総面積と数量を含む壁の種類の QTO テーブルを作成するために、LLM -

chat のテキストクエリを作成してみましょう：

> プロジェクトのデータフレームから、"Category "パラメータに "OST_Walls "を持つ項目のみを取り出し、"Type Name "でグループ化し、"Area "列の値を合計し、数量を加算し、NULL 値を削除して表に表示する。

- LLM の回答 QTO テーブルの完成形：

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

図 6.4-6 QTO テーブルを自然言語で作成すると、CAD - (BIM-) ツールを使用した場合と同じ品質の結果が得られる。

- 表形式のデータフレームに変換した後、IFC フォーマットのプロジェクトにクエリーを実行し、LLM チャットに同様のテキストクエリーを入力してみましょう：

> プロジェクトから、"Parent "パラメータに Level 1 と Level 2 の値を持つアイテムだけを取り出し、"Category "パラメータに IfcSlab の値を持つアイテムを取り出し、これらのアイテムを "ObjectType "パラメータでグループ化し、"PSet_RVT _Dimensions Area "パラメータの値を合計し、円グラフとして表示する。

LLM レスポンス IFC データから要素グループの完成円グラフとして ：

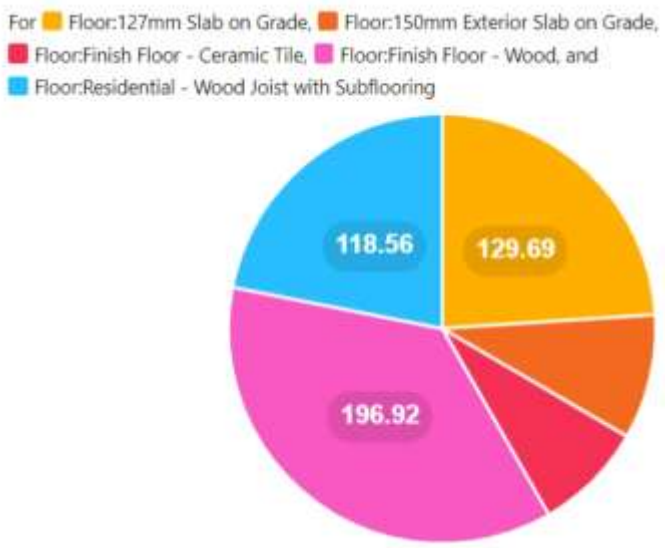


図 6.4-7 構造化フォーマットの IFC データのクエリーの結果は、データを理解するのに便利な任意のタイプのグラフにすることができます。

出来上がった各既成解（図 6.4-5〜図 6.4-7）の背後には、Pandas ライブラリを使った数十行の Python コードがあります。出来上がったコードは、LLM チャットルームからコピーして、LLM チャットルームの外でも同じ結果を得るために、ローカルまたはオンラインの IDE で使うことができます。

同じ LLM チャットで、3D CAD (BIM) フォーマットのプロジェクトだけでなく、DWG フォーマットの平面図面も扱うことができます。LLM チャットに問い合わせを行い、例えば、構造化されたフォームに変換した後の線や 3D ジオメトリの形の要素グループのデータを表示することができます。

LLM と Pandas を用いた DWG - ファイルの自動解析

DWG（）からのデータ処理プロセスは、情報の非構造化の性質上、常に複雑な作業であり、専門的なソフトウェアと、しばしば手作業による分析を必要としてきた。しかし、人工知能（）や LLM ツール（）の発達により、多くのステップを自動化できるようになった。LLM（この例では ChatGPT ）へのリクエストの実際のパイプライン 、プロジェクトで作業できる DWG 図面を考えてみましょう：

- DWG データ をレイヤー、ID、座標でフィルタリングする。
- エレメントの形状を視覚化する
- パラメータに基づいて図面に自動的に注釈を付ける
- 壁のポリラインを水平面に展開する
- インタラクティブな 3D - 平面データのビジュアライゼーションの作成
- 複雑な CAD（）を使わずに、建設データを構造化し、分析する。

我々の場合、Pipeline 構築プロセスは、LLM による逐次コード生成から始まる。まず、タスクを記述するクエリが生成される。ChatGPT 、Python -コードが生成され、実行・分析され、チャットルーム内で結果が表示される。結果が期待通りでない場合、リクエストは修正され、プロセスが繰り返される

> パイプライン は、データを処理し分析するために実行される一連の自動化されたステップである。このようなプロセスでは、各ステップはデータを入力として受け取り、変換を実行し、結果を次のステップに渡す。

目的の結果を得た後、コードは LLM からコピーされ、便利な IDE（我々の場合は Kaggle プラットフォーム.com）のブロックの形でコードに貼り付けられる。出来上がったコード断片は、単一のパイプラインにまとめられ、データのロードから最終的な分析までのプロセス全体が自動化される。このアプローチにより、深いプログラミングの専門知識がなくても、分析プロセスの迅速な開発とスケーリングが可能になります。以下の全てのフラグメントのフルコードとサンプルクエリは、Kaggle.com プラットフォームで "DWG Analyse with ChatGPT | DataDrivenConstruction" と検索すると見ることができます。[129].

DWG データを構造化フォームに変換した後（図 4.1-13）、DWG データを扱うプロセスを、古典的なステップである、壁要素、特にポリライン（パラメータ 'ParentID' で線をグループにまとめることができます）のタスクに必要なすべての図面データのグループ化とフィルタリングから始めましょう。

- 同様のタスクのコードと結果を画像で得るには、LLM で以下のクエリーを書く必要がある：

まず、DWG から取得したデータフレームに定義されたカラムが含まれているかどうかを確認する：Layer'、'ID'、'ParentID'、'Point'。次に'Layer'カラムから文字列'wall'を含む ID をフィルタリングする。ParentID' カラムから、これらの ID に一致するアイテムを見つける。Point' 列のデータをクリーニングし、分割する関数を定義する。これには、括弧を削除することと、値を 'x', 'y', 'z' 座標に分割することが含まれる。matplotlib を用いてデータをプロットする。一意の 'ParentID' ごとに、'Point' 座標を結ぶ個別の折れ線を引く。可能であれば、最初と最後の点がつながっていることを確認する。適切なラベルとタイトルを設定し、x 軸と y 軸のスケールが等しくなるようにします。

LLM 答えは、背後に生成された Python コード が隠されている既成の画像を提供します：

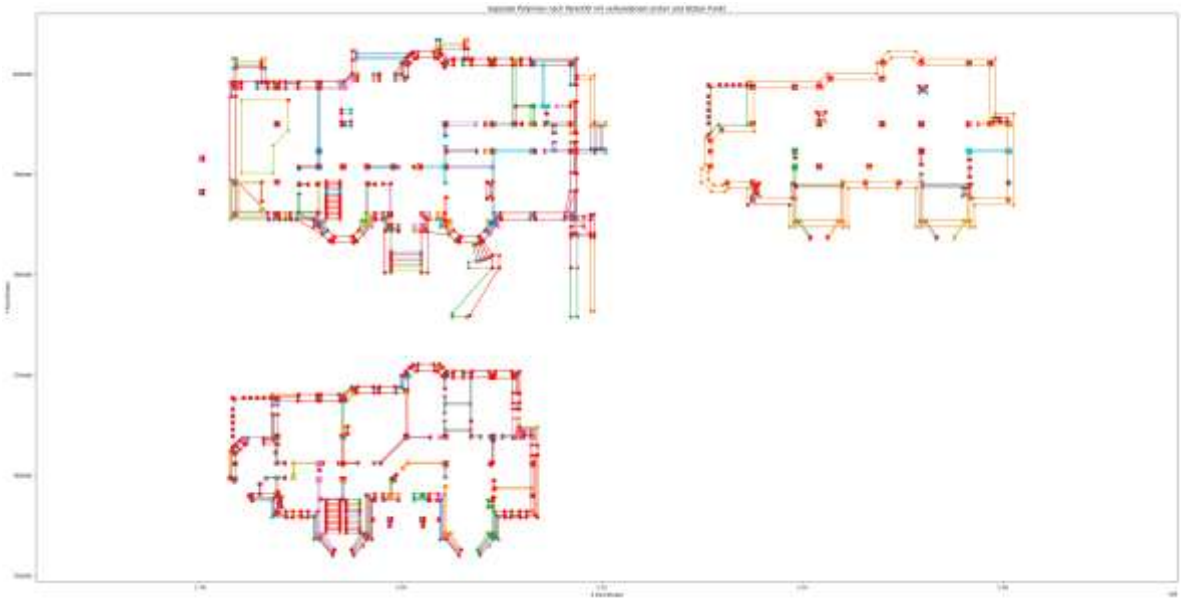


図 6.4-8 LLM コードは、DWG ファイルから「壁」レイヤーのすべての線を抽出し、その座標をクリアし、Python ライブラリの 1 つを使用してポリラインを構築します。

次に、各ポリラインのプロパティ（データフレームの列の 1 つ）にある面積パラメータをラインに追加してみましょう：

ここで、各ポリラインから "ParentID "を 1 つだけ取得する。"ID "列でこの ID を見つけ、"Area "の値を取り、1,000,000 で割って、この値をグラフに加える。

LLM レスポンス 、新しいグラフが表示され、各ポリラインにはその面積がキャプションで表示されます：

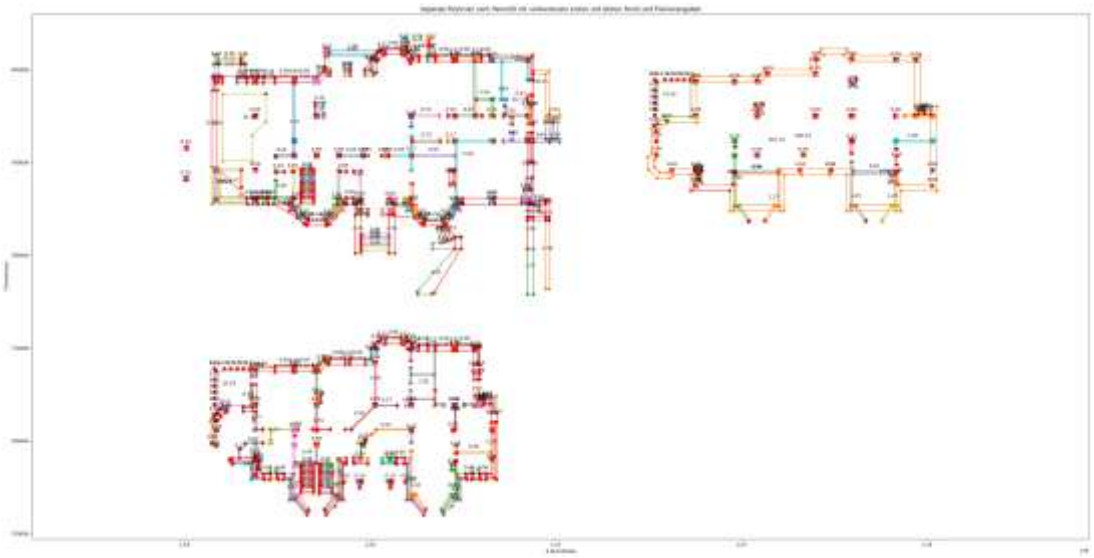


図 6.4-9 LLM は、各ポリラインの面積値を取得し、それを画像に追加して線を視覚化するコードを追加しました。

次に、各ポリラインを水平線に変換し、高さ 3000mm の平行線を追加して 1 つの平面につなぎ、このようにして壁要素の表面のレイアウトを示す：

> Layer "カラムから "wall "という値を持つすべての要素を取り出す必要がある。ID "列からこれらの ID をリストとして取り出し、"ParentID "列でデータフレーム全体からこれらの ID を見つける。すべての要素は、1 本のポリラインに結合されたラインです。各線は、"Point " 列の最初の点の異なる x, y ジオメトリを持つ。各ポリラインを順番に取り、点 0,0 から水平にポリラインの各セグメントの長さを描く。次に、まったく同じ線を 3000 だけ上に引き、すべての点を 1 つの平面に結ぶ。

LLM のレスポンス、平面上に壁の図面をプロットできるコードを出力します：

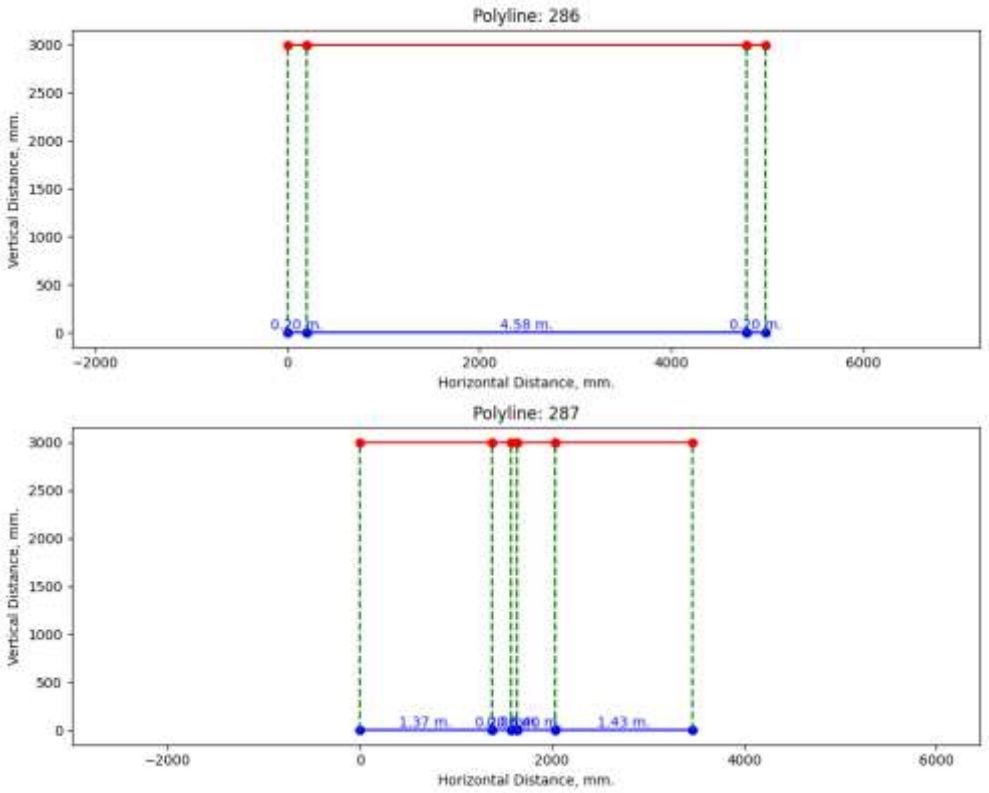


図 6.4-10 プロンプトを使った各ポリラインを、LLM チャットで直接壁面を視覚化するレイアウトにします。

それでは、2D 投影から 3D に移行しよう。ポリラインの上下のレイヤーを接続することで、平らな線から壁をモデリングする：

> 3D で壁要素を視覚化し、高さ z = 0 と z = 3000 mm でポリラインを接続します。建物の壁を表す閉じたジオメトリを作成します．Matplotlib 3D グラフ作成ツール を使います。

- LLM は、各ポリラインが平面の集合として表現されるインタラクティブな 3D グラフを生成する。ユーザーは、コンピュータのマウスを使って要素間を自由に移動することができ、チャットから IDE にコードをコピーすることで、3D モードでモデルを探索することができる：

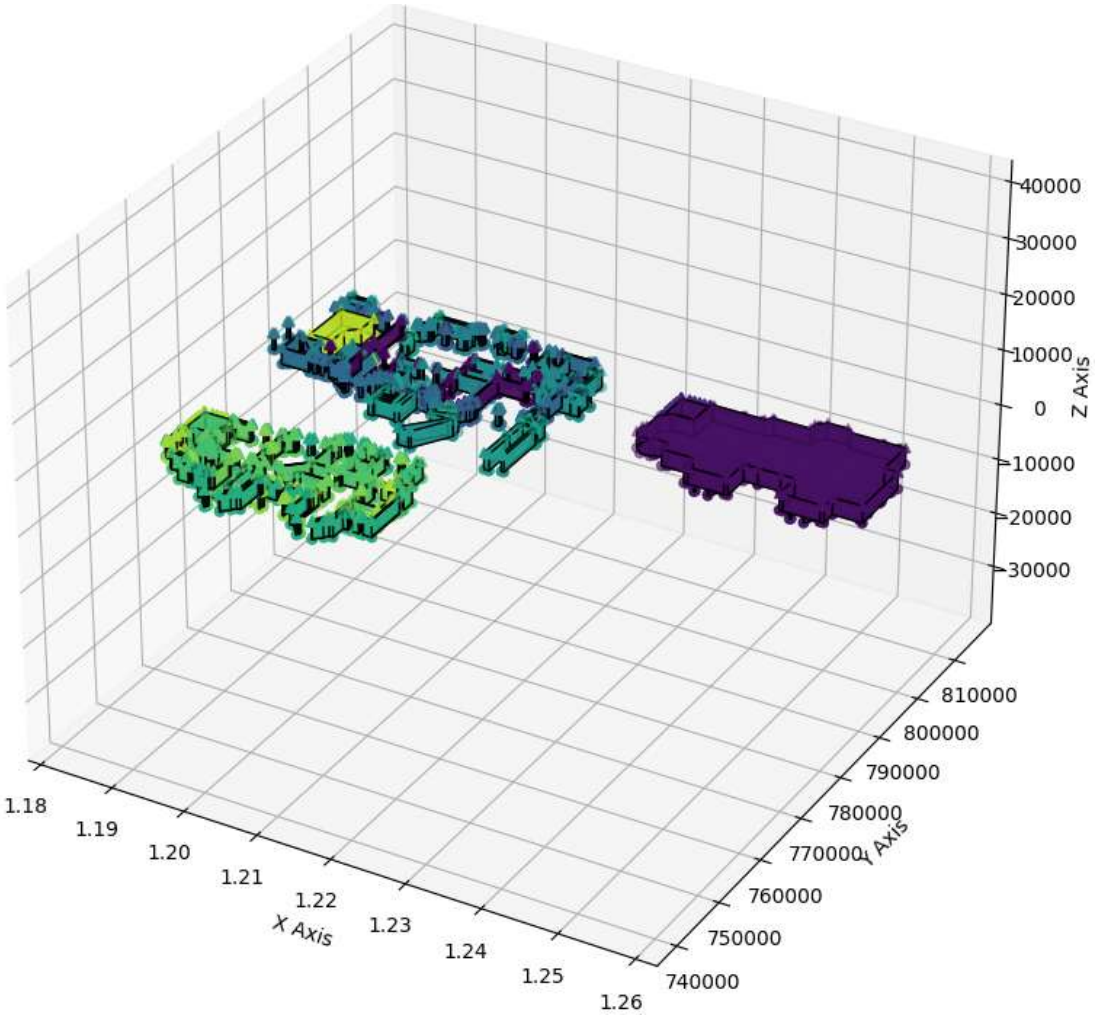


図 6.4-11 LLM は、IDE 内の 3D ビューアで探索できるように、平坦な描画線を 3D ビューに視覚化するコード[129]の構築を支援しました。

論理的で再現性のあるパイプライン を構築するために、最初の変換と DWG - ファイルの読み込みから最終結果まで、各ステップの後に、生成された LLM -コードのブロックを IDE にコピーすることをお勧めします。こうすることで、チャットで結果を確認するだけでなく、開発環境ですぐに実行することができます。こうすることで、プロセスを順次構築し、必要に応じてデバッグし、適応させることができます。

すべてのフラグメントの完全な Pipeline コード（図 6.4-8 から図 6.4-11）は、サンプルクエリとともに、

Kaggle プラットフォーム.com で "DWG Analyse with ChatGPT | DataDrivenConstruction " と検索することで見つけることができます。[129].Kaggle では、使用されているコードとプロンプトを表示できるだけでなく、追加のソフトウェアや IDE 自体をインストールすることなく、無料でクラウド上のオリジナルの DWG データフレームを使用してパイプライン全体をコピーしてテストすることができます。

本章で紹介するアプローチにより、DWG -プロジェクトに基づくドキュメントのチェック、処理、生成を完全に自動化することができます。開発されたパイプラインは、個々の図面の処理にも、数十、数百、数千の DWG ファイルのバッチ処理にも適しており、各プロジェクトに必要なレポートやビジュアライゼーションの自動生成も可能です。

まず、CAD -ファイルからのデータを自動的に XLSX フォーマット に変換し、次にデータフレームにロードし、グループ化、チェック、結果生成を行う。必要であれば、プロジェクト・ドキュメント管理システムとの統合により、プロセスを簡単に拡張することができる。指定された基準に従って CAD ファイルを自動的に検索し、結果をストレージ・システムに戻し、結果の準備ができたら、電子メールやメッセンジャーでユーザーに通知することができる。

> LLM チャットとエージェントを使用して設計データを扱うことで、専門的な CAD - プログラムへの依存を減らし、マウスクリックや複雑なメニューナビゲーションを覚えることなく、インターフェイスを手動で操作することなく、建築設計の解析と視覚化を実行できます。

日を追うごとに、建設業界では、LLM（）、粒状構造化データ、DataFrames（）、カラム型データベースについて耳にする機会が増えている。様々なデータベースや CAD フォーマットから形成される統一された二次元 DataFrames は、他業界の専門家が積極的に扱う最新の分析ツールの理想的な燃料となるだろう。

自動化プロセス自体が大幅に簡素化される。クローズドなニッチ製品の API を研究し、パラメータを分析または変換する複雑なスクリプトを書く代わりに、今では個々のテキストコマンドのセットという形でタスクを定式化するだけで十分である。このコマンドは、必要なプログラミング言語用の必要なパイプライン またはワークフロープロセスに組み込まれ、ほとんどすべてのデバイスで無料で実行される。CAD- (BIM-) ツールベンダーからの新製品、フォーマット、プラグイン、アップデートを待つ必要はもうない。エンジニアや建設業者は、LLM チャットやエージェントによって支援されながら、シンプルで、無料で、理解しやすいツールを使って、データを使って独自に作業することができるようになります。

次のステップ：クローズド・フォーマットからオープン・データへ

未来のデザインデータを扱うとき、プロプライエタリなツールの幾何学的なカーネルを理解したり、同じ情報を含む何百もの互換性のないフォーマットを学んだりする必要が本当にあるとは思えない。しかし、なぜオープンな構造化データへの移行が重要なのかを理解しなければ、新しいフリーツールやオープンデータ、ソフトウェアベンダーが推進しそうもないアプローチの使用を主張することは難しい。

この章では、CAD (BIM)データの主な特徴、その限界と機会、そしてベンダーのマーケティング上の約束にもかかわらず、エンジニアや設計者が毎日 設計情報の抽出、転送、分析において困難に直面していることについて述べてきました。これらのシステムのアーキテクチャーを理解し、オープン・フォーマットや LLM（ ）による自動化をベースとした代替アプローチについて学ぶことで、企業はおろか、一人の専門家にとっても生活がずっと楽になる可能性がある。本章のまとめとして、本章で取り上げたアプローチを日常業務に応用するための主な実践的ステップを紹介します：

- プロジェクトデータを扱うためのツールキットを増やす
  - お使いの CAD - (BIM-) システムからデータを抽出するために利用可能なプラグインやユーティリティをご覧ください。
  - 専用のソフトウェアを手動で開くことなく、閉じたフォーマットからのデータ抽出を自動化できる SDK や API を利用できます。
  - オープンなノンパラメトリック・ジオメトリ・フォーマット（OBJ、glTF、USD、DAE）および対応するオープンソース・ライブラリの基本的な操作スキルを習得する。
  - CAD (BIM) ソリューション以外のジオメトリとは別に、プロジェクトのメタデータを保存するシステムを考え、分析や他のシステムとの統合を簡素化する。
  - LLM を使用して、フォーマット間のデータ変換問題を自動化する。
- プロジェクト情報を扱う独自のプロセスを構築する
  - シンプルで構造化されたフォーマットで、パラメータとその値を通してタスクとモデリング要件を記述し始める。
  - 頻繁に実行される操作のためのスクリプトやコードブロックのパーソナルライブラリを作成する

。

■ オープンスタンダードの利用を推進する

- ☐ ソフトウェア・ベンダーのエコシステムに制限されないオープン・フォーマットでデータを共有するよう、同僚やパートナーを招待する。
- ☐ 構造化データを使用することの利点を、具体的な例を挙げて示す。
- ☐ クローズド・フォーマットの問題点と可能な解決策について議論を開始する。

CAD - (BIM-) プラットフォームに関する会社の方針を変えることができなくても、オープンフォーマットでプロジェクトデータを扱う原則を個人的に理解することで、作業効率を大幅に向上させることができます。異なるフォーマットからデータを抽出・変換するための独自のツールや方法を作成することで、ワークフローを最適化できるだけでなく、標準的なソフトウェアソリューションの制限を回避する柔軟性も得ることができます。



VII 部

データ主導の意思決定、分析、自動化、機械学習

第 7 部では、建設業界におけるデータ分析とプロセスの自動化に焦点を当てる。データがどのように意思決定の基礎となるかを論じ、効果的な分析のための情報可視化の原則を説明している。主要業績評価指標（KPI ）、投資収益率（ROI ）の評価方法、プロジェクトモニタリングのためのダッシュボード作成について詳しく説明する。特に、ETL プロセス （Extract, Transform, Load ）と、パイプライン（Pipeline ）を使用した自動化により、異種データを構造化された分析用情報に変換することに注目する。Apache Airflow 、Apache NiFi 、 n8n などのワークフロー・オーケストレーション・ツールは、深いプログラミング知識なしに自動化されたデータパイプラインを構築することができます。大規模言語モデル（LLM）と、データ分析を簡素化し、ルーチンタスクを自動化するためのその使用は、重要な役割を果たしている

章 7.1.

データ分析とデータに基づく意思決定

情報の収集、構造化、クリーニング、検証の段階を経て、首尾一貫した分析可能なデータセットが出現した。本書の前編では、PDF 文書や会議のテキスト記録から CAD モデルや幾何学的データまで、異種ソースの体系化と構造化を取り上げた。情報を様々なシステムや分類器の要件と照合・整合させ、重複や矛盾を排除するプロセスについて詳しく解説している。

単純な変換から、時間、コスト、ESG 指標の計算（第 5 部）に至るまで、このデータに対して実行されるすべての計算（本書の第 3 部、第 4 部）は、集約された分析作業である。これらは、プロジェクトの現状を理解し、そのパラメーターを評価し、意思決定を行うための基礎となる。その結果、データは、計算の結果、バラバラの記録の集合から、重要なビジネス上の質問に答えることができる管理可能なリソースに変わる。

これまでの章では、典型的なビジネスケースや建設業界特有のプロセスで使用するデータ収集と品質管理プロセスについて詳述してきた。この文脈におけるアナリティクスは、他の産業におけるアプリケーションと多くの点で類似しているが、いくつかの特殊な特徴がある。

以下の章では、最初の情報収集とその変換から、その後の対象システムや文書への転送までの自動化ステップを含む、データ分析プロセスについて詳述する。まず、理論的な部分として、データ駆動型意思決定に関する選択された側面に焦点を当てる。そして、次の章では、ETL -Pipeline の自動化と構築に関連する実践的なパートを紹介する。

意思決定における資源としてのデータ

データ主導の意思決定は、多くの場合、反復的なプロセスであり、様々な情報源から情報を体系的に収集することから始まる。自然のサイクルのように、個々のデータ要素や情報システム全体が徐々に土に落ち、企業の情報リポジトリに蓄積されていく（図 1.3-2）。時間の経過とともに、このデータは落ち葉や枝のように、価値ある素材へと変化していく。データ・エンジニアやアナリストの菌糸体は、情報を整理し、将来利用できるように準備し、落ちたデータやシステムを貴重な堆肥に変え、新しい芽や新しいシステムを育てる（図 1.2-5）。

様々な業界でアナリティクス（ ）が広く利用されるようになった傾向は、データを扱うことが専門家としての活動の基礎となる新しい時代の始まりを示すものである（図 7.1-1）。建設業界の専門家は、こうした変化に適応し、新しい時代（データとアナリティクスの時代）に移行する準備を整えることが重要です

> 手作業でデータをテーブル間で移動させ、手作業で計算を行うことは徐々に過去のものとなりつつあり、自動化、データフロー分析、アナリティクス、機械学習に取って代わられつつある。これらのツールは、最新の意思決定支援システムの重要な要素になりつつある。

> マッキンゼーの著書『再起動。McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130]では、2022 年に様々な地域、業界、機能分野の上級幹部 1,330 人を対象に実施された調査を引用している。[130]は、2022 年に様々な地域、業界、機能分野の上級幹部 1,330 人を対象に実施された調査を引用している。その結果によると、リーダーの 70%が高度なアナリティクスを使用して独自のアイデアを生み出し、50%が意思決定プロセスの改善と自動化のために人工知能を導入している。



図 7.1-1 データ分析 とアナリティクス は、企業における意思決定のスピードを上げるための主要なツールである。

データ分析 は、菌糸が広がるように、過去の意思決定の腐葉土に浸透し、個々のシス テムをつなぎ、マネジャーを価値ある洞察へと導く。この知識は、朽ち果てたデータシステムの樹木の栄養分のように、企業内の新たな意思決定に供給され、豊かで健康な土壌から芽吹く新芽のように、効果的な変化と質の高い情報の成長につながる（図 1.2-5）。

> 数字には語るべき重要なストーリーがある。彼らは、あなたが彼らに明確で説得力のある声を与えてくれることを期待している[131]。
>
> - スティーブン・フュー、データ・ビジュアライゼーション専門家

中堅・中小企業において、さらなる分析のために情報を抽出し準備する作業は、今日、18 世紀の石炭採掘に匹敵する極めて労働集約的なプロセスである（図 7.1-2）。つい最近まで、データマイニングと準備の作業は、非構造化、疎構造化、混合、およびクローズドなソースからのさまざまなタイプのデータを扱うための小さく限定されたツールセットで、高度に専門化されたニッチな分野で働く冒険家たちのためのものであった。

> 意思決定者や管理者は、異種データやシステムの扱いに不慣れなことが多いが、それらに基づいて意思決定を行う必要がある。その結果、過去数十年にわたり、現代の建設業界におけるデータ主導の意思決定は、自動化されたプロセスというよりは、初期の炭鉱における坑夫の数日にわたる手作業のようなものとなっている。

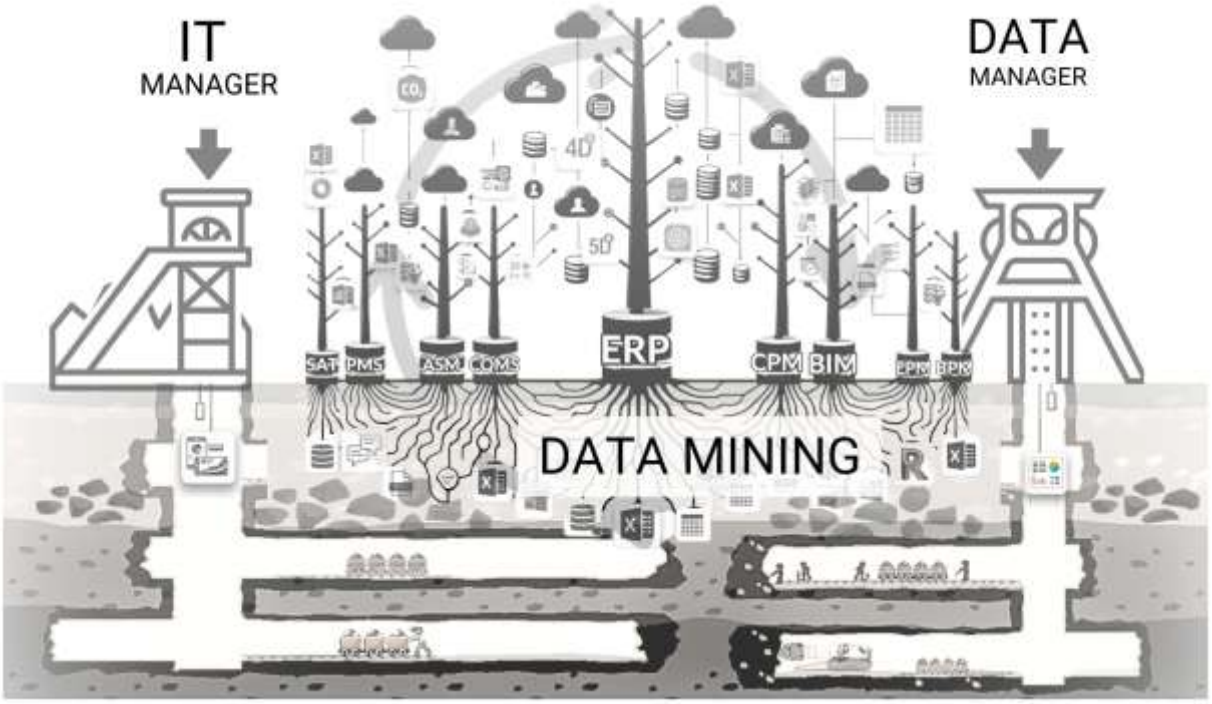


図 7.1-2 データマイニングのプロセスでは、専門家たちはデータのクリーニングからその後の分析に向けた構造化まで、複雑なデータ準備の過程を経る。

建設業界におけるデータ抽出の現代的な手法は、12 世紀の鉱山労働者の原始的な技術よりは確実に進歩しているものの、それでも複雑でリスクの高い作業であり、大企業にしかできないような多大なリソースと専門知識を必要とする。過去のプロジェクトで蓄積された遺産から データを抽出・分析するプロセスは、つい最近まで、何十年も一貫してデータを収集・保存してきた、技術的に進んだ大企業が主に行ってきた

> 以前は、アナリティクスで主導的な役割を担っていたのは、何十年もデータを蓄積してきた技術的に成熟した企業だった。今日、状況は変わりつつある。データとデータ処理ツールへのアクセスが民主化されつつあり、以前は複雑だったソリューションが誰でも無料で利用できるようになっている。

アナリティクスを適用することで、企業はより正確な情報に基づいた意思決定をリアルタイムで行うことができます。 。次のケーススタディは、過去のデータが財務的に健全な意思決定にどのように役立つかを示しています：

- プロジェクトマネージャー 「現在、市内のコンクリート平均価格は 82€ /m^3ですが、見積もりでは 95 €/m^3となっています。
- 見積もり担当者 「以前のプロジェクトでは、オーバーランが 15%くらいだったので、バックしました。
- データマネージャーまたは顧客側のコントロールエンジニア - "過去 3 件の入札の分析を見てみよう"

過去のプロジェクトの DataFrame を分析した結果、我々はこうなった：

- 平均実勢購入価格： 84.80 /m^3€
- 平均支出超過率：+4.7
- 見積もりの推奨レート： ~ 85 /m^3€

このような決定は、もはや主観的な感情ではなく、具体的な過去の統計に基づくことになり、リスクを減らし、入札の妥当性を高めるのに役立つ。過去のプロジェクトのデータを分析することは、一種の「有機肥料」となり、そこから新しい、より正確なソリューションが発芽する。

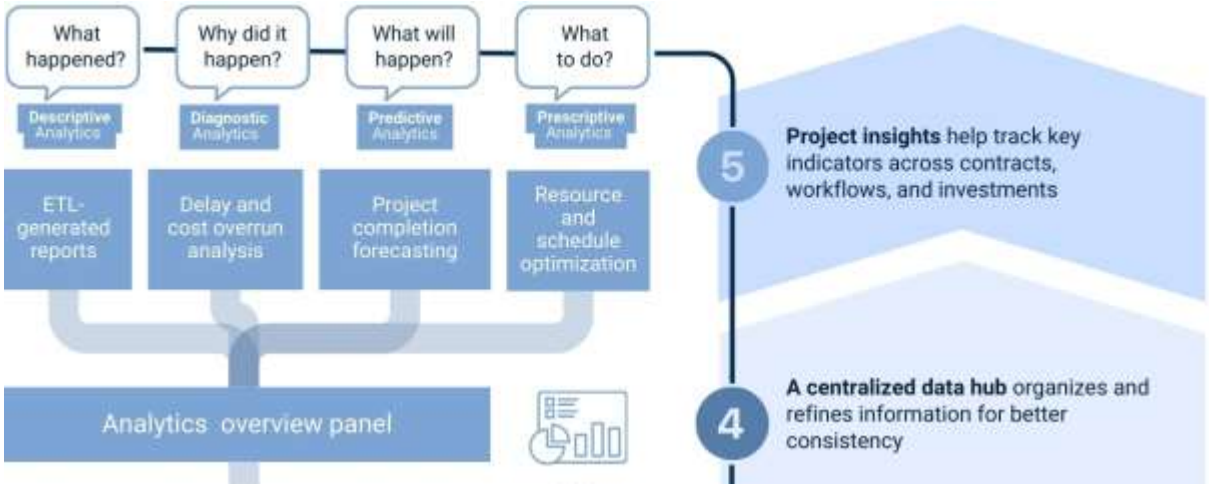


図 7.1-3 データ解析は 3 つの重要な質問に答える：何が起こったか、なぜ起こったか、次に何をすべきか。

意思決定者や管理者は、十分な技術的専門知識を持たないまま、異種データやシステムを扱う必要性にしばしば直面する。そのような状況において、分析プロセスにおける最初の、そして最も重要なステップのひとつである視覚化は、データを理解するための重要な助けとなる。視覚的で理解しやすい形で情報を提示することができる。

データの視覚化：理解と意思決定の鍵

プロジェクト・データが複雑かつマルチレベル構造であることが特徴である今日の建設業界では、視覚化が重要な役割を果たしている。 データを視覚化することで、プロジェクト・マネージャーやエンジニアは、大量の異種データに隠された複雑なパターンや傾向を視覚化することができます。

> データを視覚化することで、リソースの配分、コストの傾向、材料の使用状況など、プロジェクトの状況を理解しやすくなります。グラフやチャートは、複雑で乾燥した情報をアクセスしやすく理解しやすくし、注意を要する重要な領域をすばやく特定し、潜在的な問題を発見することを可能にします。

データの視覚化は、情報の解釈を容易にするだけでなく、「何が起こったのか」、「どのように起こったのか」という問いに答える助けとなり、分析プロセスや情報に基づいた経営上の意思決定における重要なステップとなる（図 2.2-5）。(図 2.2-5）。

> グラフィックは、論理的な問題を解くための視覚的なツールである[132]。
>
> - ジャック・ベルタン、"グラフィックスとグラフィカル情報処理

重要な決定を下す前に、プロジェクト・マネジャーはスプレッドシートやテキスト・メッセージの乾いた解釈しにくい数字よりも、視覚的に表現されたデータを使う傾向がある。

> 可視化されていないデータは、建設現場に無造作に散乱している建築資材のようなもので、その可能性は不明確だ。レンガとコンクリートで作られた家のように、明確に視覚化されて初めてその価値が明らかになる。家を建てるまでは、その資材の山が小さな小屋になるのか、豪華な別荘になるのか、超高層ビルになるのかはわからない。

企業は様々なシステム（図 1.2-4〜図 2.1-10）、金融取引、膨大なテキストデータからデータを得ている。しかし、これらのデータをビジネスに役立てることは困難な場合が多い。そのような状況では、データの意味を伝えるための視覚化が重要なツールとなり、ダッシュボード、グラフ、チャートなど、専門家であれば誰でも理解できる形式で情報を提示するのに役立ちます。

> PwC の調査「急速に変化するビジネス世界で成功するために学生に必要なこと」（2015 年）は、成功している企業はデータ分析にとどまらず、意思決定をサポートするためにグラフ、インフォグラフィックス、分析ダッシュボードなどのインタラクティブな視覚化ツールを積極的に活用していることを強調している[9]。同レポートによると、データ・ビジュアライゼーションは、チャート、グラフ、ダイアグラム、ダッシュボード、インタラクティブなデータモデルを通じて、顧客がデータが語るストーリーを理解するのに役立つ。

情報をチャート、グラフ、ダイアグラムのような視覚的な図形に変換するプロセスは、 人間の脳によるデータの理解と解釈を向上させる（図 7.1-4）。これにより、プロジェクトマネージャーやアナリストは、複雑なシナリオをより迅速に評価し、直感ではなく、視覚的に認識できる傾向やパターンに基づいて、情報に基づいた意思決定を行うことができる。

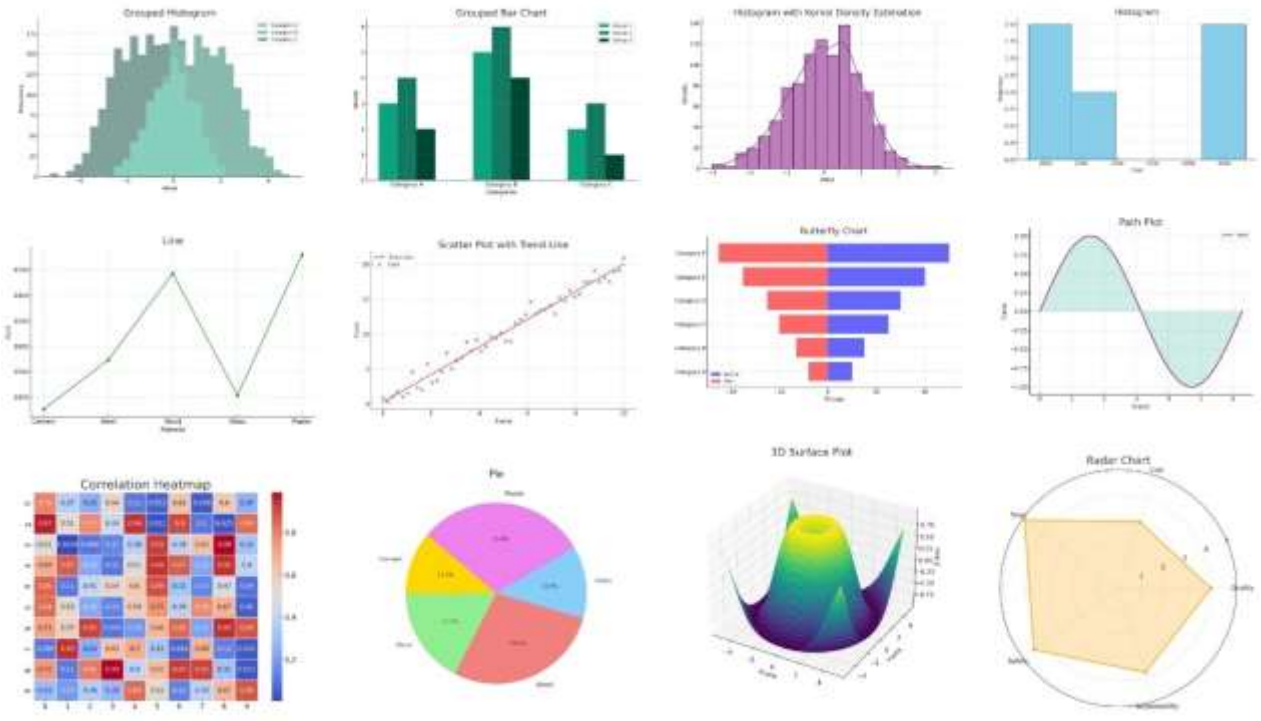


図 7.1-4 様々な種類の視覚化は、人間の脳が数字という乾燥した情報をよりよく理解し、意味を理解するのを助けるように設計されている。

データからのビジュアライゼーションの作成と、様々なフリーのビジュアライゼーション・ライブラリの使用については、次の ETL -プロセスの章で詳しく説明する。

ビジュアライゼーションは、建設業界でデータを扱う上で不可欠な要素になりつつある。データを「見る」だけでなく、管理業務の文脈でその意味を理解するのにも役立つ。しかし、ビジュアライゼーションを真に役立てるためには、具体的に何をビジュアライズすべきか、どのメトリクスがプロジェクトのパフォーマンスを評価する上で本当に重要なのかを事前に判断する必要がある。ここで、KPI や ROI などのパフォーマンス指標が登場する。これがなければ、どんなに美しいダッシュボードでも、単なる「情報のノイズ」になってしまう危険性があります。

KPI および ROI

今日の建設業界では、パフォーマンス指標（KPI 、ROI ）の管理と、レポートやダッシュボードによる可視化が、生産性とプロジェクト管理の効率を向上させる上で重要な役割を果たしている。

どのようなビジネスでもそうであるように、建設においても、成功、投資対効果、パフォーマンスを測定する指標を明確に定義することが必要である。様々なプロセスに関するデータを取得する際、データ主導型の組織は、まず主要な KPI （Key Performance Indicators：重要業績評価指標）を特定することを学ばなければならない。

KPI（） を計算するには、通常、実績指標と計画指標を含む計算式が使用される（図 7.1-5）。例えば、プロジェクト、従業員、またはプロセスの個別の KPI を計算するには、実績 を計画実績で割り、その結果に 100%を掛けます。

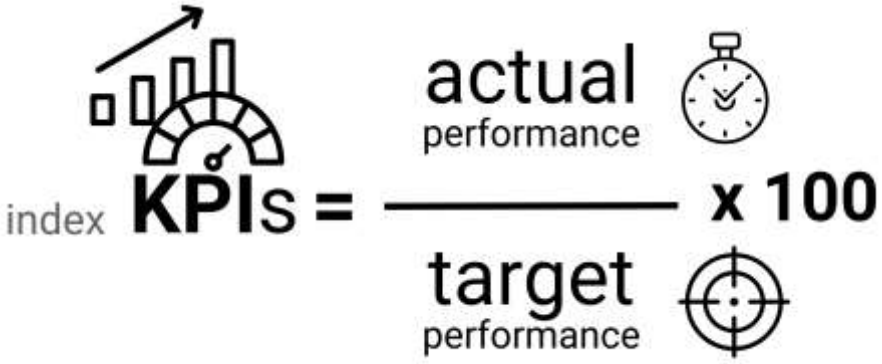


図 7.1-5 KPI は、プロジェクトまたはプロセスが主要な目標を達成する際の成功を測定するために使用される。

サイトレベルでは、より詳細な KPI を使用することができる メトリクス：

- 主要なマイルストーン（基礎、据付、仕上げ）のタイミング - 作業計画の遵守を管理することができる。
- 資材のオーバーラン率 - 調達を管理し、無駄を最小限に抑えるのに役立ちます。
- 予定外の機械停止回数 - 生産性とコストに影響。

誤った指標を選択すると、「何をすべきか」の判断を誤ることになる（図 2.2-5）。例えば、1 平方メートルあたりのコストにのみ注目し、改造にかかるコストを考慮しない場合、資材の節約によって品質が低下し、将来のプロジェクトにおけるコストが高くなる可能性がある。

目標を設定する際には、何を測定するのかを明確にすることが重要である。曖昧な表現は誤った結論を導き、

管理を複雑にする。建設業における KPI（）の成功例と失敗例を見てみよう。

良い KPI ：

- 「年末までに、模様替えの割合を 10%減らす。
- 「次の四半期までに、品質に妥協することなく、ファサードの施工速度を 15%向上させる。
- "年末までに作業スケジュールを最適化し、機械のダウンタイムを 20%削減する"

これらの指標は、明確に測定可能で、具体的な値と時間枠を持っている。

悪い KPI ：

- 「我々はより速く建設する」（どのくらい速いのか、"より速く "とはどういう意味なのか？）
- 「コンクリート工事の品質を向上させます」（品質とは具体的にどのように測られるのか？）
- 「現場での請負業者とのやり取りを改善する」（どのような基準で改善を示すのか？）

> 優れた KPI とは、、測定して客観的に評価できるものである。建設業では、明確な指標がなければ、パフォーマンスを監視し、安定した結果を得ることができないため、これは特に重要である。

KPI に加えて、投資の有効性を評価するための追加指標がある：ROI （投資利益率） - 利益と投資資金の比率を反映する投資収益率指標。ROI は、新しい方法、技術、ツールの導入が正当化されるかどうかを評価するのに役立つ。デジタルソリューションや自動化（例：図 7.3-2）から新しい建築資材の使用まで。この指標は、事業の収益性に与える実際の影響に基づき、さらなる投資について十分な情報に基づいた意思決定を行うのに役立つ

> 建設プロジェクトマネジメントの文脈では、ROI （投資利益率）は、企業の目標がプロジェクト、技術、プロセス改善の投資利益率を測定することである場合、主要業績評価指標（KPI ）の一つとして使用することができる。例えば、新しい建設管理技術を導入する場合、ROI は、それがどれだけ収益性を向上させたかを示すことができる。

材料消費量、労働時間、コストなど、さまざまなソースから収集したデータに基づき、KPI と ROI を定期的に測定することで、プロジェクト管理者はリソースを効果的に管理し、迅速な意思決定を行うことができる。このデータを長期的に保存することで、将来の傾向を分析し、プロセスを最適化することができます。

KPI、ROI、その他の指標を視覚化するために様々なチャートやグラフが使用され、、通常はダッシュボードに

まとめられます。

ダッシュボードとダッシュボード：効果的なマネジメントのための指標の可視化

指標や測定基準を視覚化するために、さまざまな図表が使用され、それらは通常、データ・ショーケースやダッシュボードにまとめられます。これらのダッシュボードは、プロジェクトまたはプロジェクトの一部のステータスを一元的に表示し、主要な指標を（理想的にはリアルタイムで）表示します。最新かつ継続的に更新されるダッシュボードにより、チームは変化に迅速に対応できる。

> ダッシュボード 定量的な評価を視覚化するツールで、プロジェクト参加者全員が簡単にアクセスでき、理解できるようにする。



図 7.1-6 KPI を管理し、ダッシュボードで可視化することは、プロジェクトの生産性と効率を向上させる鍵である。

ここでは、ダッシュボードを作成できる一般的なツールの例をいくつか紹介する：

- Power BI は、インタラクティブなレポートやダッシュボードを作成するための Microsoft のツールです。
- Tableau と Google Data Studio は、コードを書かずにデータを視覚化し、ダッシュボードを作成す

るための強力なツールである。

- Plotly (図 7.1-6、図 7.2-12)はインタラクティブなグラフを作成するためのライブラリであり、Dash はデータを分析するためのウェブアプリケーションを作成するためのフレームワークである。これらを組み合わせて使うことで、インタラクティブなダッシュボードを作成することができる。
- 数多くの Python ライブラリ （図 7.2-9 - 図 7.2-11） - Python には、Matplotlib 、Seaborn 、Plotly 、Bokeh など、データ視覚化のための多くの オープンソースで無償のライブラリがあります。これらを使ってグラフを作成し、Flask や Django などのフレームワークを使って Web アプリケーションに統合することができます。
- JavaScript Libraries: D3.js や Chart.js などのオープンソース JavaScript ライブラリを使用してインタラクティブなダッシュボードを作成し、Web アプリケーションに統合することができます。

> KPI を評価し、ダッシュボードを作成するには、最新のデータと、情報収集と分析のための明確なタイムテーブルが必要である。

一般的に、建設業界における KPI 、ROI 、ダッシュボードは、プロジェクト管理に対する分析的アプローチの基礎を形成する。ダッシュボードは、現状を監視・評価するのに役立つだけでなく、将来の計画や最適化プロセスのための貴重な洞察を提供する。

データ分析 と質問の技術

データの解釈は分析の最終段階であり、ここで情報は意味を持ち、「語り」始める。ここで重要な質問に対する答えが導き出される：「何をすべきか」「どのようにすべきか」である（図 2.2-5）。この段階では、結果を要約し、パターンを特定し、因果関係を確立し、視覚化と統計分析に基づいて結論を導き出すことができる。

> おそらく、現在発展しつつある新しい複雑な世界国家のひとつで有能な市民となるためには、現在読み書きできることが必要であるのと同じくらい、計算ができること、平均値や最大値、最小値で考えることが必要であることに気づく日はそう遠くないだろう[133]。

> - サミュエル・S・ウィルクス、1951 年アメリカ統計協会会長講演より引用

> 英国政府が発表した報告書「政府プロジェクトの実施におけるデータ分析と人工知能」（2024 年）[83]によると、分析 データと人工知能 （AI ）を導入することで、プロジェクト管理プロセスを大幅に改善し、時間とコスト予測の精度を高めるとともに、リスクと不確実性を低減することができる。この論文では、高度な分析ツールを使用する公共組織が、インフラストラクチャーの取り組みにおいてより高いパフォーマンスを達成していることを強調している。

第 4 次産業革命の激しい競争と低マージンの環境下で活動する現代の建設事業は、軍事作戦に例えることができる。そこでは、企業の存続と成功は、資源と質の高い情報をいかに早く入手するか、つまり、タイムリーで情報に基づいた意思決定にかかっている（図 7.1-7）。

データ可視化が概観を提供する「インテリジェンス」であるとすれば、データ分析 は行動に必要な「弾薬」である。データアナリティクスは、「何をすべきか」「どのようにすべきか」という問いに答え、市場における競争優位を獲得するための基盤を形成する。

> アナリティクスは、異種データを構造化し、意思決定の基礎となる意味のある情報に変える。

アナリストやマネジャーの仕事は、単に情報を解釈するだけでなく、情報に基づいた意思決定を行い、傾向を特定し、異なるタイプのデータ間の関係を判断し、プロジェクトの目標や詳細に従って分類することである。ビジュアライゼーション・ツールや統計分析手法を用いて、データを企業の戦略的資産に変えるのです。

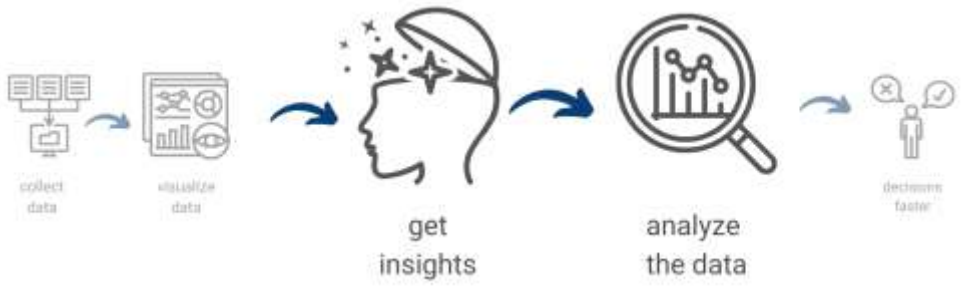


図 7.1-7 収集した情報を最終的に意思決定の材料に変えるのは、データ分析ある。

アナリティクス・プロセス（ ）において、真に情報に基づいた意思決定を行うためには、データに対して行う質問の正しい立て方を学ぶ必要がある。このような質問の質は、得られる洞察の深さ、ひいては経営上の意思決定の質に直接影響します。

> 過去は今日の記録の中に存在する限りにおいてのみ存在する。そして、これらの記録が何を表しているかは、私たちの問いかけによって決まる。この歴史以外に歴史は存在しない[134]。
>
> - ジョン・アーチボルト・ウィーラー、物理学者 1982 年

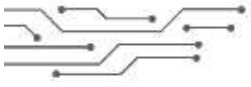
深い質問をし、批判的に考える技術は、データを扱う上で重要なスキルである。たいていの人は、答えを出すのにほとんど労力を必要としない、単純で表面的な質問をしがちである。しかし、真の分析は、意味のある、思慮深い質問から始まり、何重もの推論の背後に隠れているかもしれない情報の隠れた関係や因果関係を明らかにすることができる。

> Data-Driven Transformation: Accelerating at Scale Now」（BCG、2017 年）[135]という調査によると、デジタルトランスフォーメーションを成功させるには、アナリティクス能力への投資、チェンジマネジメントプログラム、ビジネス目標と IT イニシアチブの整合性が必要である。データ主導の文化を創造する企業は、データ分析能力に投資し、新しい考え方、行動、働き方を浸透させるための変更管理プログラムを立ち上げるべきである。

分析文化の発展、データツールの改善、スペシャリストの育成に投資しなければ、企業は時代遅れの情報や不完全な情報に基づいて、あるいは HiPPO マネジャーの主観的な意見に頼って意思決定を行うリスクを抱え続けることになる（図 2.1-9）。

> アナリティクス（）とダッシュボードを常に更新する必要性と関連性が認識されると、経営陣は必然的に分析プロセスの自動化の重要性を理解するようになる。自動化は意思決定のスピードを上げ、人的要因の影響を減らし、データの関連性を確保する。情報量が飛躍的に増大する中、スピードは単なる競争上の優位性ではなく、持続的な成功のための重要な要素となる。

一般的にデータ分析と処理プロセスの自動化は、ETL （Extract, Transform, Load ）と切っても切れない関係にある。自動化プロセスでデータを変換する必要があるように、ETL プロセスでもデータは様々なソースから抽出され、必要な要件に従って変換され、さらに使用するためにターゲットシステムにロードされる。



章 7.2.

手作業に頼らないデータフロー：なぜエテルが必要なのか

ETL 自動化: コスト削減とデータ処理のスピードアップ

データ量やチーム規模の増加にもかかわらず、主要業績評価指標（KPI ）が伸びなくなると、企業の経営陣は必然的にプロセスの自動化の必要性に気づくようになる。遅かれ早かれ、この認識は、複雑な自動化を開始する動機となる。その主な目的は、プロセスの複雑さを軽減し、処理をスピードアップし、 、人的要因への依存を減らすことである。

> マッキンゼーの研究「イノベーションを推進するデータ・アーキテクチャの構築方法 - 今日と明日」（2022 年）[136]によると、ストリーミング・データ・アーキテクチャを使用する企業は、リアルタイムで情報を分析できるため、大きな優位性を得ることができる。ストリーミング・テクノロジーは、リアルタイムのメッセージを直接分析し、リアルタイムのセンサー・データを分析することで、製造業における予知保全の適用を可能にする。

> プロセスの簡素化とは自動化であり、従来の手作業がアルゴリズムやシステムに取って代わられる。

自動化、いや「データ処理における人間の役割の最小化」という問題は、どの企業にとっても不可逆的で非常にデリケートなプロセスである。どのような専門分野のスペシャリストも、急速に進化する技術環境の中で職を失うリスクを認識しているため、自分たちの仕事のやり方や微妙な点をオプティマイザー仲間に完全に開示することをためらうことが多い。

> 敵を作りたければ、物事を変えようとすることだ[137]。
>
> - ウッドロー・ウィルソン、セールスマン議会での演説（1916 年、デトロイト

自動化のメリットは明らかであるにもかかわらず、多くの企業では、特にエンジニアリング・データの分野で、日常業務に占める手作業の割合が依然として高い。現状を説明するために、このようなプロセスにおけるシ

ーケンシャルなデータ処理の典型例を見てみよう。

手作業によるデータ処理は、CAD データベースから取得した情報とのやりとりを例に説明することができる。CAD (BIM)部門における従来のデータ処理（"手動 "ETL -プロセス）は、属性テーブルの作成や設計データに基づくドキュメンテーションの作成のために、以下の順序で行われる（図 7.2-1）：

1. 手動抽出（Extract ）：ユーザーが手動でプロジェクトを開く - CAD アプリケーション （BIM ）を起動することによって（図 7.2-1 手順 1）。
2. 検証：次のステップでは通常、いくつかのプラグインまたは ヘルパーアプリケーショ ンを手動で実行し、データを準備し、その品質を評価する（図 7.2-1 ステップ 2-3）。
3. 手作業による変換（Transform ）：準備の後、データ処理が始まるが、その、データをアップロードするための様々なソフトウェアツールを手作業で操作する必要がある（図 7.2-1 ステップ 4）。
4. 手動アップロード（Load ）：変換されたデータを外部システム、データフォーマット、文書に手動でアップロードする（図 7.2-1 ステップ 5）。

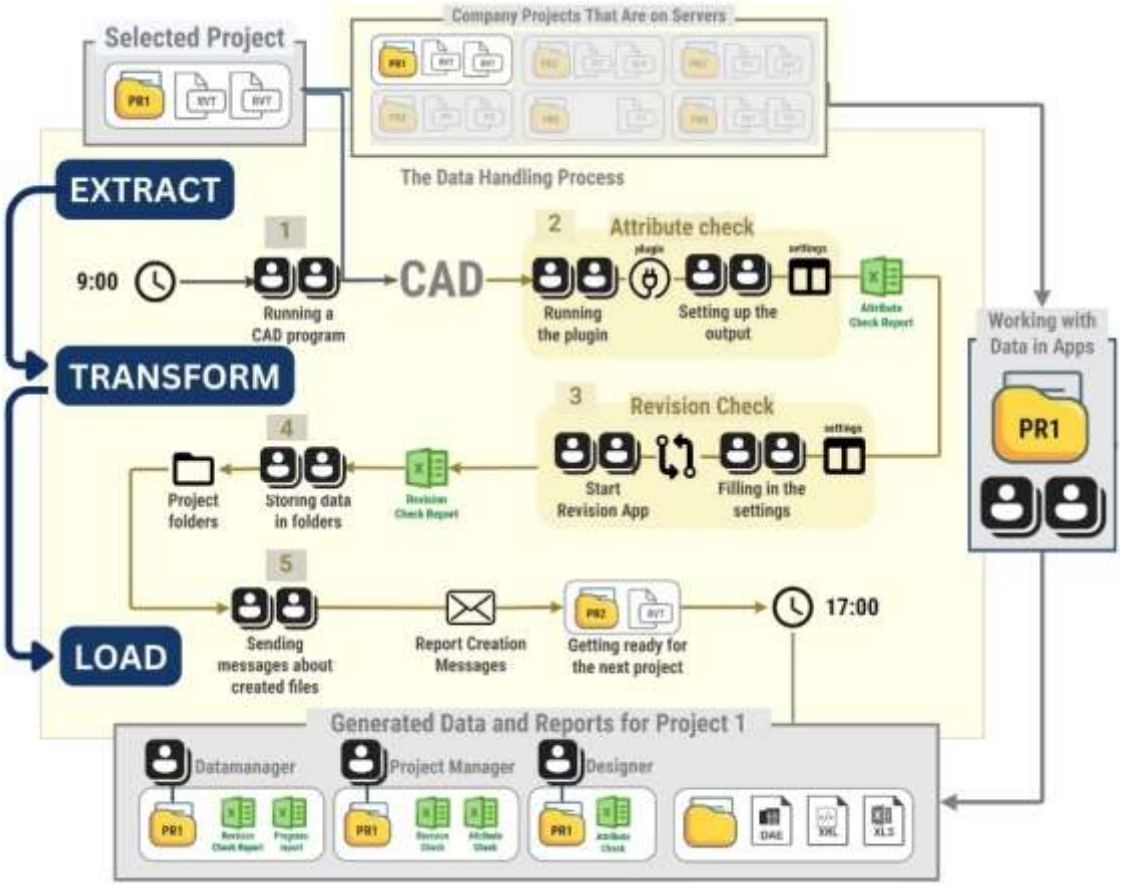


図 7.2-1 従来の手作業による ETL 処理は、個々の技術者の欲求と身体的能力によって制限される。

このようなワークフローは、古典的な ETL（ ）プロセスの一例であり、抽出、変換、ロード（ETL）である。他の業界では、自動 ETL パイプラインが標準となって久しいが、建設業界では未だに手作業が主流である

> ETL （Extract, Transform, Load ）とは、様々なソースからデータを抽出し、目的のフォーマットに変換し、さらに分析や使用のためにターゲットシステムにロードするプロセスである。

ETL は、データ処理の 3 つの主要な構成要素である Extract 、Transform 、Load を示すプロセスである（図 7.2-2）：

- Extract - さまざまなソース（ファイル、データベース、API ）からデータを抽出します。
- Transform - データクリーニング、集計、正規化、論理処理。
- ロード - 構造化された情報をデータウェアハウス、レポート、BI システムにロードします。

ETL の概念については、本書の前半で、構造化されていないスキャン文書を構造化された表形式に変換する場合（図 4.1-1）、生活プロセスとビジネスプロセスの両方の認識を体系化するために要件を形式化する場合（図 4.4-20）、CAD ソリューションからのデータ検証とデータ処理の自動化の場合など、時折触れる程度であった。次に、典型的なワークフローに即して、ETL をより詳細に見てみよう。

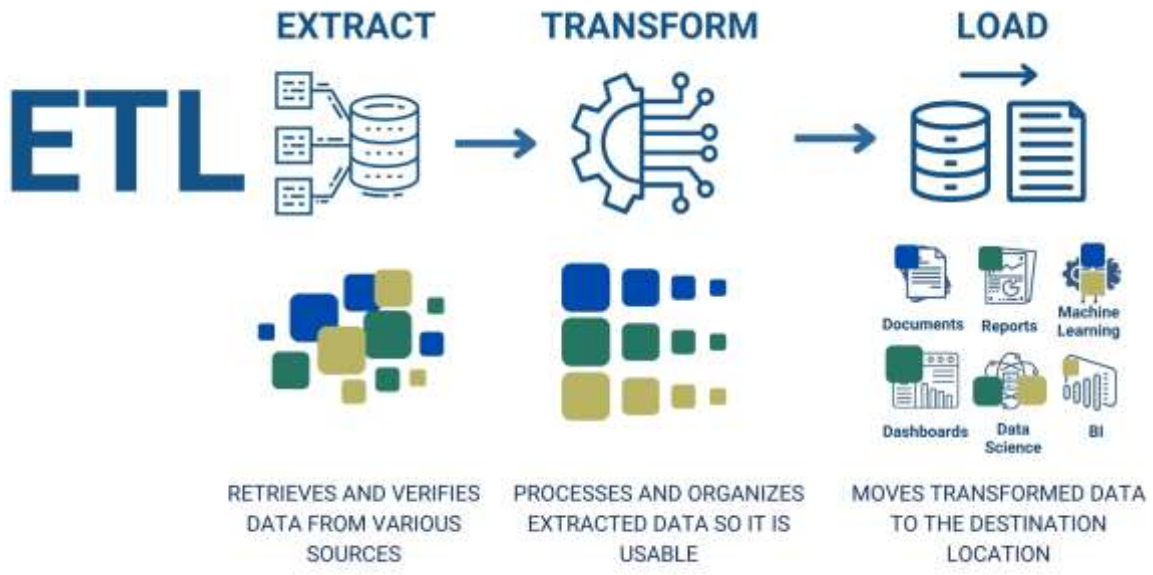


図 7.2-2 ETL 繰り返し行われるデータ処理作業を自動化する。

手動または半自動 ETL -プロセスは、データ収集からレポート作成まで、すべてのステップを手動で管理するマネージャーまたは技術者を意味する。このようなプロセスでは、特に作業時間が限られている場合（9:00〜17:00 など）、かなりの時間がかかる。

企業はしばしば、低効率と低速の問題を、モジュール化された統合ソリューション（ERP 、PMIS 、CPM 、CAFM 、など）を購入することで解決しようとし、それを外部のベンダーやコンサルタントがさらに開発する。しかし、これらのベンダーやサードパーティの開発者は、しばしば重要な依存ポイントになる。前章のプロプライエタリ・システムやフォーマットで詳述したように、彼らの技術的限界がシステム全体のパフォーマンスやビジネス全体に直接影響するのである。断片化と依存関係によって生じる問題については、「建設業はいかにしてデータ・カオスに溺れるか」の章で詳述した。

もし企業が、あるベンダーの大規模なモジュール式プラットフォームを導入する準備ができていない場合、自動化の代替方法を探し始める。その 1 つが、独自のモジュール式オープン ETL -コンベアを開発することである。ここでは、各ステージ（抽出、変換、検証、ロード）が、スケジュールに基づいて実行されるスクリプトとして実装されている。

同じ ETL ワークフローの自動化バージョン （図 7.2-1）では、作業プロセスは、データを処理してオープンな構造化フォームに変換することから始まるモジュラーコードのように見える。構造化されたデータが受信されると、様々なスクリプトやモジュールがスケジュールに基づいて自動的に実行され、変更をチェックし、変換し、メッセージを送信する（図 7.2-3）。

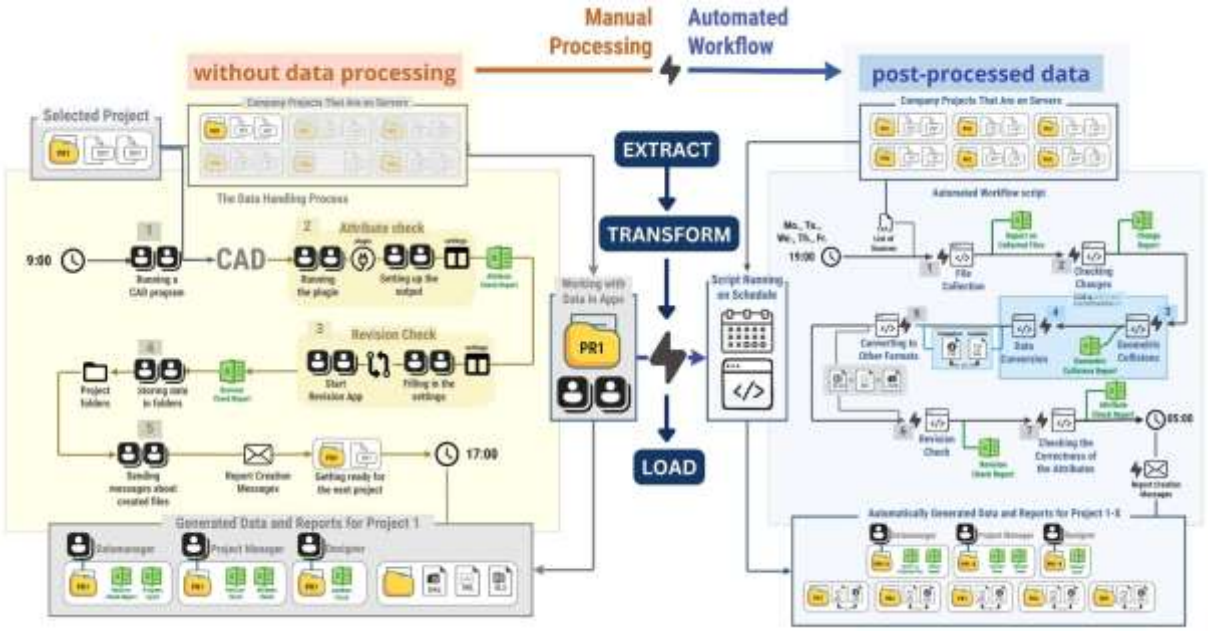


図 7.2-3 左は手動加工、右は従来の手動加工とは異なり、ユーザーの能力に制限されない自動加工。

自動化されたワークフローでは、ET(L)データの前処理（構造化と単一化）によってデータ処理が簡素化されます。

伝統的な処理方法では、専門家はシステムやソフトウェアから取得したデータを「そのまま」扱う。これとは対照的に、自動化されたプロセスでは、多くの場合、データはまず ETL -payplane を通過し、そこでさらなる利用や分析に適した一貫性のある構造と形式に変換される。

「データと検証結果の検証」の章（図 4.4-13）で説明したデータテーブルの検証プロセスを実演する、実用的な ETL の例を見てみましょう。そのために、Pandas ライブラリ と LLM を併用し、自動化されたデータ分析・処理処理を行います。

ETL 抽出: データ収集

ETL プロセスの第一段階である Extract) は、さらにチェックと処理を行うデータセットを収集するコードを書くことから始まる。そのために、本番サーバーのすべてのフォルダーをスキャンし、一定の形式と内容のドキュメントを収集し、それらを構造化された形式に変換する。このプロセスについては、「非構造化データとテキストデータを構造化形式へ変換する」と「CAD データ (BIM)を構造化形式へ変換する」の章で詳しく説明します（図 4.1-1 〜 図 4.1-12）。

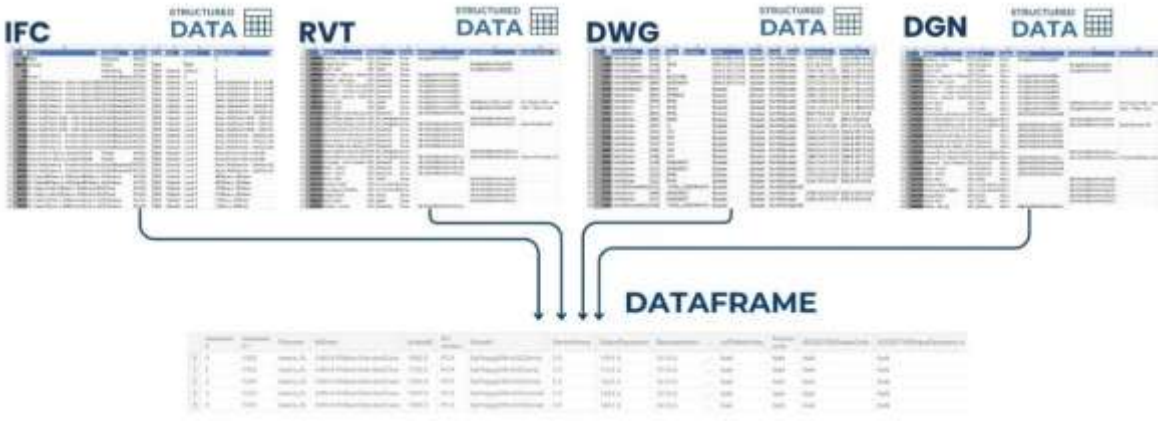


図 7.2-4 CAD データ (BIM)を、すべてのプロジェクトセクションを含む一つの大きなデータフレームに変換する。

例として、Extract データロードのステップを使用し、すべての CAD- （BIM- ）プロジェクトのテーブルを取得します（図 7.2-4）。RVT および IFC フォーマットのリバースエンジニアリング対応コンバータ [138] を使用して、すべてのプロジェクトから構造化テーブルを取得し、1 つの大きな DataFrame テーブル に結合します。



図 7.2-5 Python を使った変換 コードと SDK RVT 用リバースエンジニアリングツール と IFC ファイルを一つの大きな構造化(df)DataFrame に変換する。

Pandas DataFrame は、CSV テキストファイル 、Excel スプレッドシート 、JSON - や XML - ファイル、Parquet や HDF5 などのビッグデータストレージフォーマット、MySQL 、PostgreSQL 、SQLite 、Microsoft SQL Server、Oracle などのデータベースなど、様々なソースからデータをロードすることができます。さらに、Pandas は API 、Web ページ、クラウドサービス、Google BigQuery、Amazon Redshift、Snowflake などのストレージシステムからのデータロードもサポートしています。

- データベースに接続して情報を収集するコードを書くには、LLM チャット（ChatGP、LlaMa 、Mistral DeepSeek 、Grok 、Claude 、QWEN またはその他）に同様のテキストリクエストを送信する

：

> MySQL に接続し、データを 変換する例を書いてください。 ⏎

LLM の回答 ：

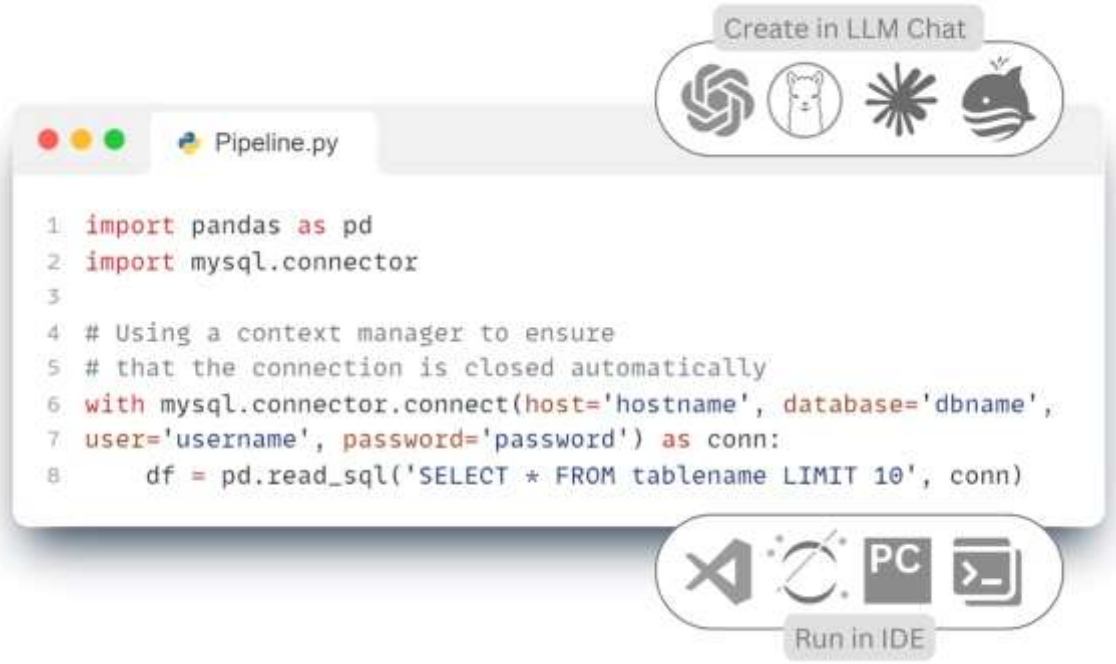


図 7.2-6 Python で MySQL データベース に接続し、MySQL データベースから DataFrame にデータをインポートする例。

出来上がったコード（図 7.2-5、図 7.2-6)は、上で述べた一般的な IDE(統合開発環境)でオフラインモードで実行できます: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse with PyDev plugin, Thonny, Wing IDE, IntelliJ IDEA with Python plugin, JupyterLab または一般的なオンラインツール: Kaggle.com、Google Collab、Microsoft Azure Notebooks、Amazon SageMaker。

多形式データを変数 "df" にロードすることで (図 7.2-5 - 行 25; 図 7.2-6 - 行 8)、データを Pandas DataFrame 形式に変換しました。この形式は、データ処理のための最も一般的な構造の 1 つで、行と列を持つ 2 次元テーブルです。Parquet 、Apache ORC 、JSON 、Feather 、HDF5 などの ETL -パイプラインで使用される他のストレージ形式や、最新のデータウェアハウスについては、「建設業界におけるデータストレージと管理」の章で詳しく説明します（図 8.1-2）。

データの抽出と構造化の段階（Extract）を経て、情報の単一配列が形成され（図 7.2-5、図 7.2-6）、さらなる処理の準備が整う。しかし、このデータを対象システムにロードしたり、分析に使用したりする前に、その品質、 完全性、および指定された要件への適合性を確保する必要がある。この段階でデータ変換（Transform ）が行われ、その後の結論や決定の信頼性を確保するための重要なステップとなる。

ETL Transform: バリデーションと変換ルールの適用

変換 ステップでは、データを処理し変換する。このプロセスには、正しさのチェック、正規化、欠損値の補充、自動化ツールによる検証などが含まれる

> PwC の調査「Data-Driven.What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9]によると、現代の監査法人は無作為のデータ検証から脱却し、自動化ツールを使って大量の情報を分析する方向に向かっている。このアプローチにより、報告書の矛盾を特定するだけでなく、ビジネス・プロセスの最適化のための提言を行うことも可能になっている。

建設業においても、設計データの自動検証、建設品質管理、請負業者のパフォーマンス評価などに、同様の手法を用いることができる。データ処理を自動化し高速化するツールの一つに、ETL プロセスのデータ変換フェーズ（Transform）で正規表現 （RegEx）を使用する方法がある。RegEx を使用すると、最小限のリソース消費で、データ文字列を効率的に検証し、不整合を検出して、情報の整合性を確保することができます。RegEx（図 4.4-7） の詳細については、 「要件を構造化形式に変換する」 の章で説明しました。

実際の例を考えてみよう。プロパティ・オブジェクト管理システム（RPM ）において、マネジャーはオブジェクトの主要な属性に関する要件（ ）を設定する（図 7.2-7）。変換段階では、以下のパラメータを検証する必要がある：

- オブジェクト識別子のフォーマット（属性 "ID"）の検証
- 交換保証期間値の管理（属性「保証期間）
- 要素の交換サイクルの検証（属性「メンテナンス要件）



図 7.2-7 検証 品質は、属性とその境界値に対する要求事項を設定することから始まる。

パラメータ検証の境界値を設定するために、例えば、 これまでの経験から、"ID "属性の有効な値は、文字列値 "W-NEW"、"W-OLD1"、"D-122"、あるいは、最初の文字が文字で、その後にダッシュが続き、3 つのアルファベット文字'NEW'、'OLD'、あるいは任意の 3 桁の数字が続く類似の値だけを含むことができると知っているとします（図 7.2-7）。これらの識別子を検証するには、以下の正規表現（RegEx）を使用することができる：

```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

このテンプレートは、データ内のすべての識別子が指定された条件に一致することを確認します。もしチェックにパスしない値があれば、システムはエラーを記録する。Python -コードを作成してデータを変換し、結果のデータを使用して結果テーブルを作成するには、LLM チャットでクエリを作成するだけです。

- テキストリクエストは LLM まで：

> 正規表現 を使用して、DataFrame のカラムを検証するコードを書く。このコードでは、RegEx を使用して、'W-NEW' または 'W-OLD' 形式の識別子、'A' から 'G' までの文字によるエネルギー効率、年単位の数値による保証期間と交換サイクルをチェックする。 ⏎

LLM の回答 ：

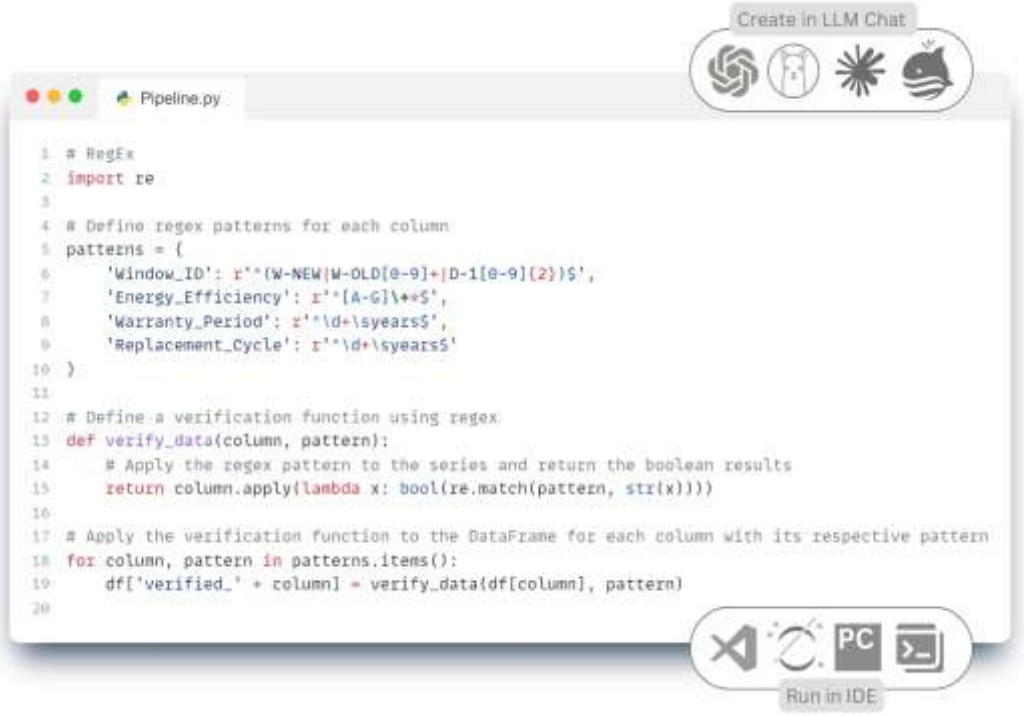


図 7.2-8 コードでは、RegEx テンプレートをデータフレームのパラメータ列に適用することで、検証プロセスを自動化します 。

上記の自動生成された Python -コード（図 7.2-8）は、"re" ライブラリ（RegEx 正規表現）を使用して、Data-Frame 内のデータ項目の各属性をチェックする関数を定義しています。指定された各列（属性）に対して、この関数は RegEx パターンを適用して、各レコードが期待される形式に適合しているかどうかをチェックし、その結果を新しい値（False/True）として DataFrame の新しい属性列に追加します。

> このような自動バリデーションは、正式なデータコンプライアンスを保証し、変換段階における品質管理システムの一部として使用することができる。

変換（Transform） ステップと品質チェックが正常に完了すると、データをターゲット・システムにアップロードする準備が整う。変換され検証されたデータは、CSV 、JSON 、Excel 、データベース、その他のフォーマットにアップロードして、さらに使用することができます。タスクによっては、結果をレポート、グラフ、分析ダッシュボードで表示することもできます。

ETL Load: の結果を図表で可視化する。

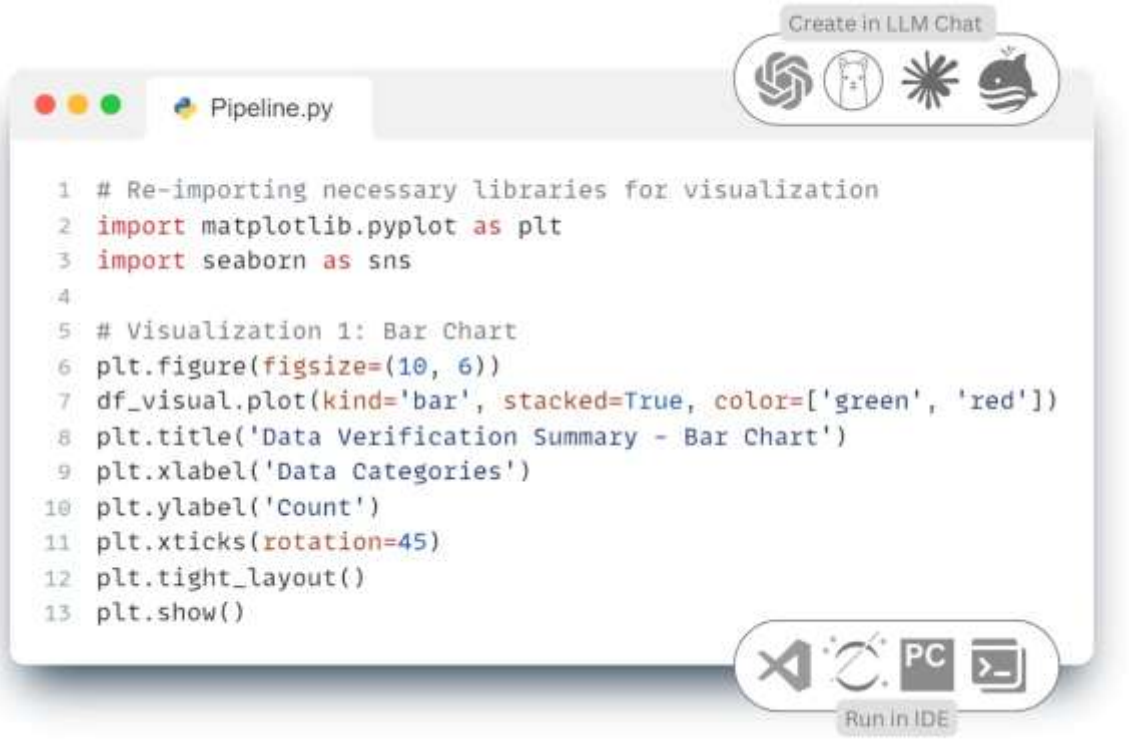
Transform 段階が完了し、データが構造化され、検証されると、最終段階は Load である。データを視覚的に表示することで、逸脱を素早く特定し、分布を分析し、技術的背景のない人を含むプロジェクト参加者全員に重要な結論を伝えることができる。

情報を表や数値で示す代わりに、インフォグラフィックス、グラフ、ダッシュボード（ダッシュボード）を使うことができます。Python で構造化データを視覚化するための最も一般的で柔軟なツールの 1 つ、Matplotlib ライブラリ （図 7.2-9、図 7.2-10）です。Matplotlib ライブラリでは、静的、アニメーション、インタラクティブなグラフを作成でき、さまざまな種類のグラフをサポートしています。

- RPM システムから属性チェックの結果を視覚化するには (図 7.2-7) 、言語モデル への次のクエリーを使用します：

> 上記の DataFrame データ （図 7.2-7）を視覚化するコードを書き、その結果をヒストグラムにして、 属性におけるエラーの頻度を示す。 ⏎

- コード形式での LLM のレスポンスと、コード実行の結果を LLM チャットルームで直接視覚化：



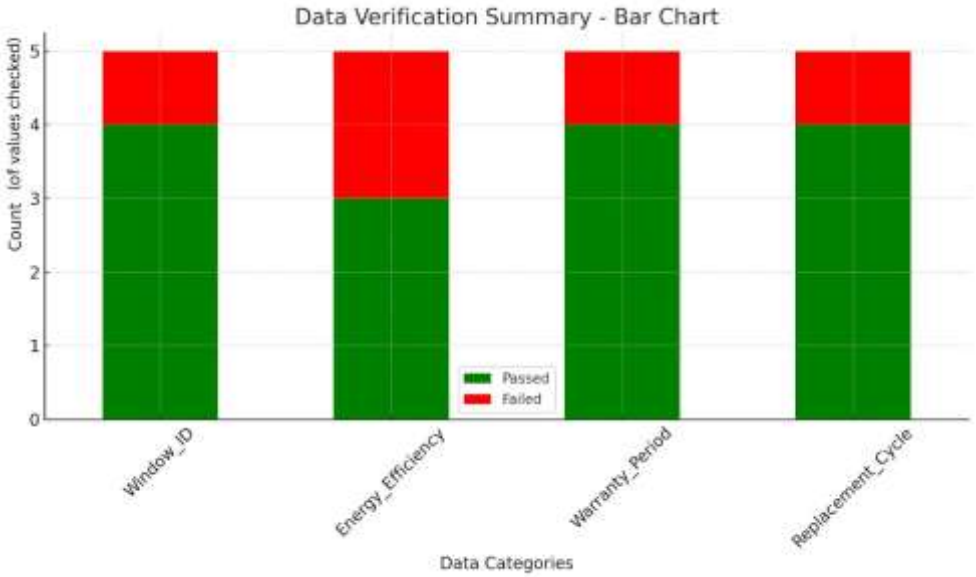
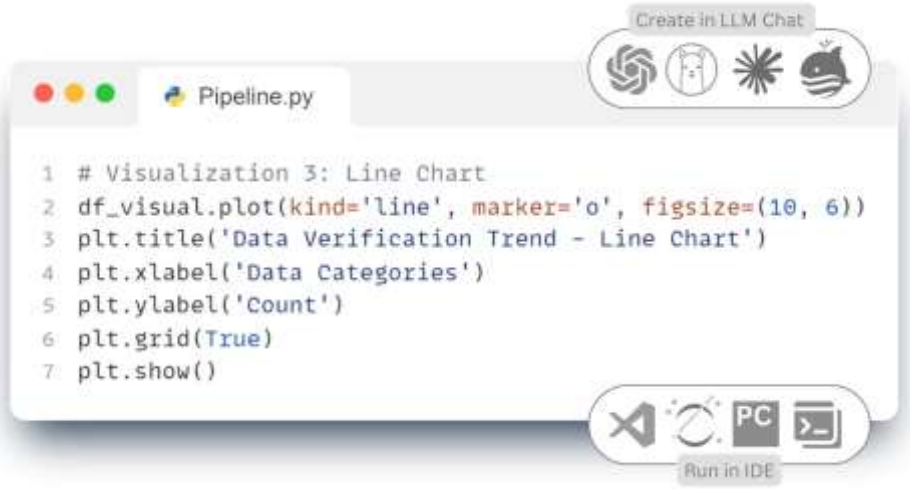


図 7.2-9 視覚化 RPM システム （図 7.2-7）から属性値をチェックする Transform ステップ の結果を、Load ステップでヒストグラムとして表示。

様々なフォーマットで構造化データを表現できるオープンソースでフリーの可視化ライブラリがたくさんある。チャット で次のプロンプトを表示しながら、別のタイプのグラフで結果を視覚化し続けよう：

> 同じデータを折れ線グラフとしてプロットする ⏎

LLM の回答 ：



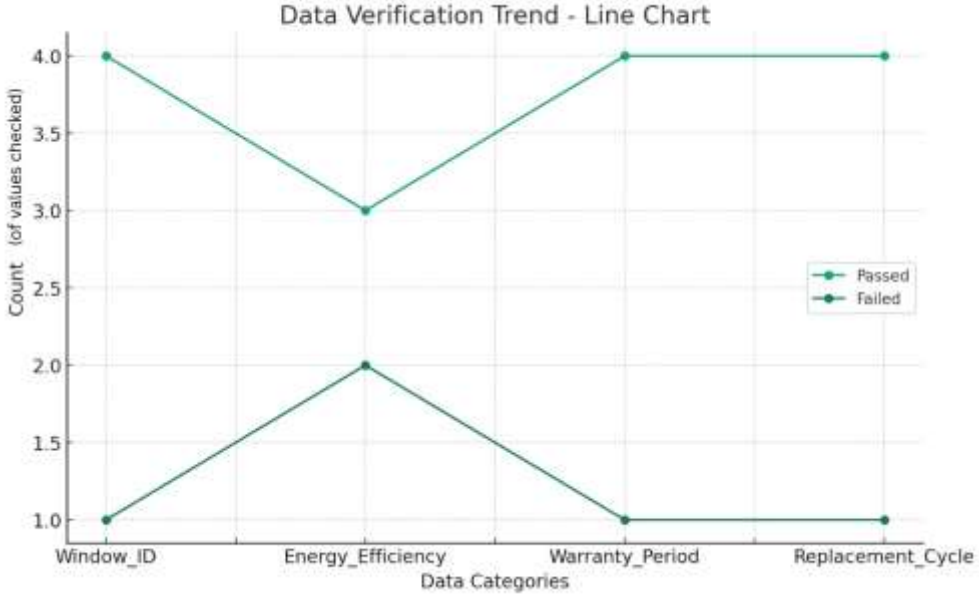


図 7.2-10 検証データ（図 7.2-8）の視覚化 Matplotlib ライブラリ を使って得られた線図。

次のようなオープンソースでフリーの可視化ライブラリがたくさんある：

- Seaborn - 統計グラフ用 (図 7.2-11)
- Plotly - インタラクティブなウェブ視覚化用 (図 7.2-12, 図 7.1-6)
- Altair - 宣言的視覚化
- Dash または Streamlit - 完全なダッシュボードを作成する。

LLM を含む最新のツールでは、タスク記述に基づいてグラフやアプリケーション全体のコードを自動生成することができます。

ツールの選択は、レポートなのか、プレゼンテーションなのか、オンラインダッシュボードなのか、といったプロジェクトの目的によって異なる。例えば、オープンソースライブラリ Seaborn は、カテゴリーデータを扱うのに特に適しており、パターンや傾向を特定するのに役立つ。

- シーボーン・ライブラリーの稼動状況を見るには 、LLM に直接希望するライブラリーの利用を依頼するか、同様のテキストリクエストを LLM に送ることができる：

  ヒートマップを表示 ⏎

- LLM の答えは、コードと完成したグラフの形で示され、そのプロットコードは IDE にコピーすることができ、グラフ自体はコピーするか、文書に貼り付けるために保存することができます ：

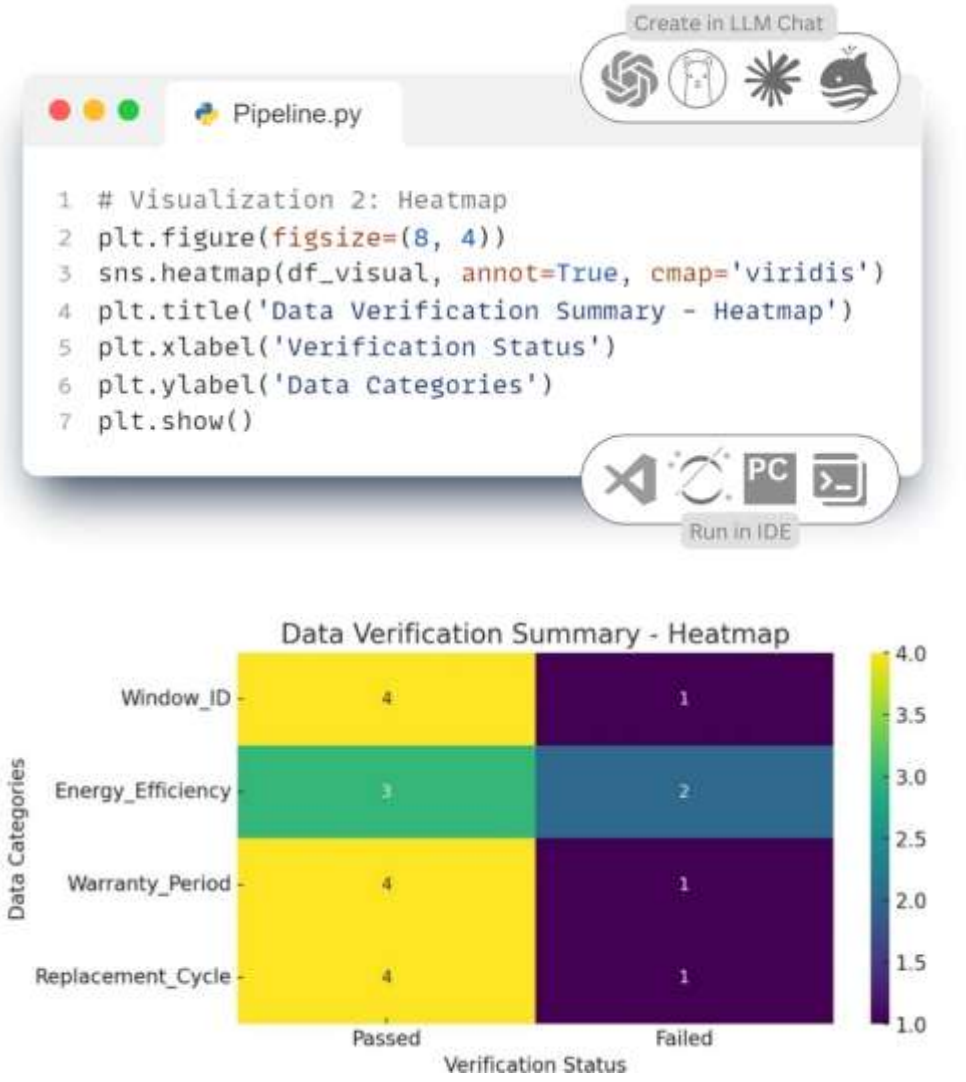


図 7.2-11 Seaborn ライブラリを使ったデータの検証結果（図 7.2-8）の視覚化.

インタラクティブなアプローチを好む人のために、インタラクティブな機能を持つダイナミックなダイアグラムやパネルを作成できるツールがある。Plotly ライブラリ (図 7.1-6、図 7.2-12)は、ウェブページに埋め込むことができ、ユーザーがリアルタイムでデータを操作できる、高度にインタラクティブなチャートやパネルを作成する機能を提供します。

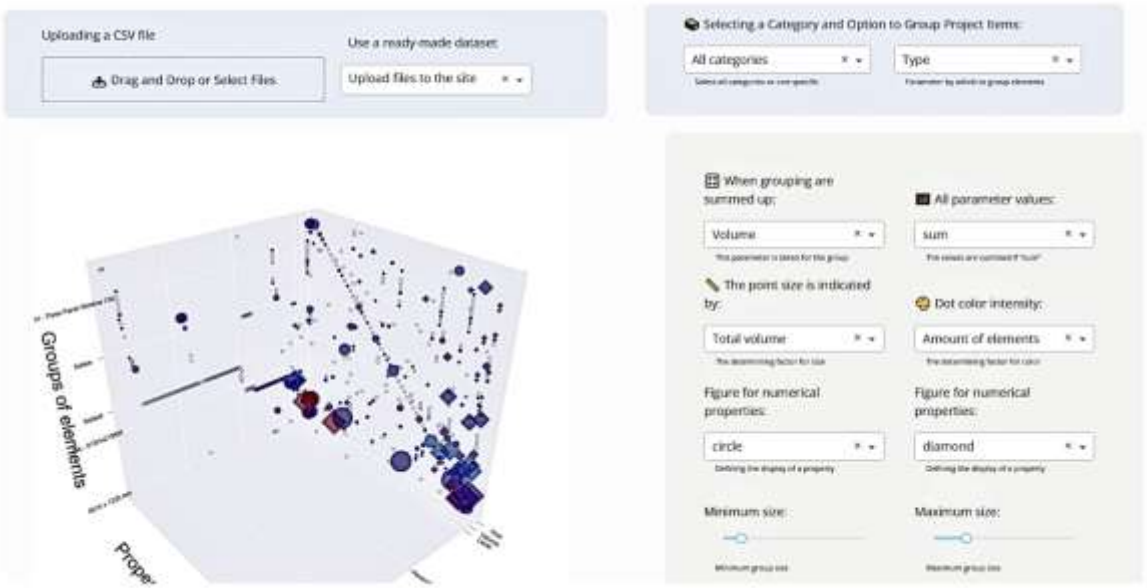


図 7.2-12 CAD- (BIM-) プロジェクトの要素属性を、Plotly ライブラリ を使ってインタラクティブに 3D -可視化。

専門的なオープンソースライブラリ Bokeh 、Dash 、Streamlit、ウェブ開発の深い知識を必要とせずにデータを表示する便利な方法を提供します。Bokeh は複雑なインタラクティブグラフに適しており、Dash は本格的な分析ダッシュボードを構築するために使用され、Streamlit はデータ分析のためのウェブアプリケーションを素早く作成することができます。

このような可視化ツールを使えば、開発者やアナリストは、 、同僚や利害関係者に結果を効果的に広めることができ、データとの直感的なインタラクションを可能にし、意思決定を簡素化することができる。

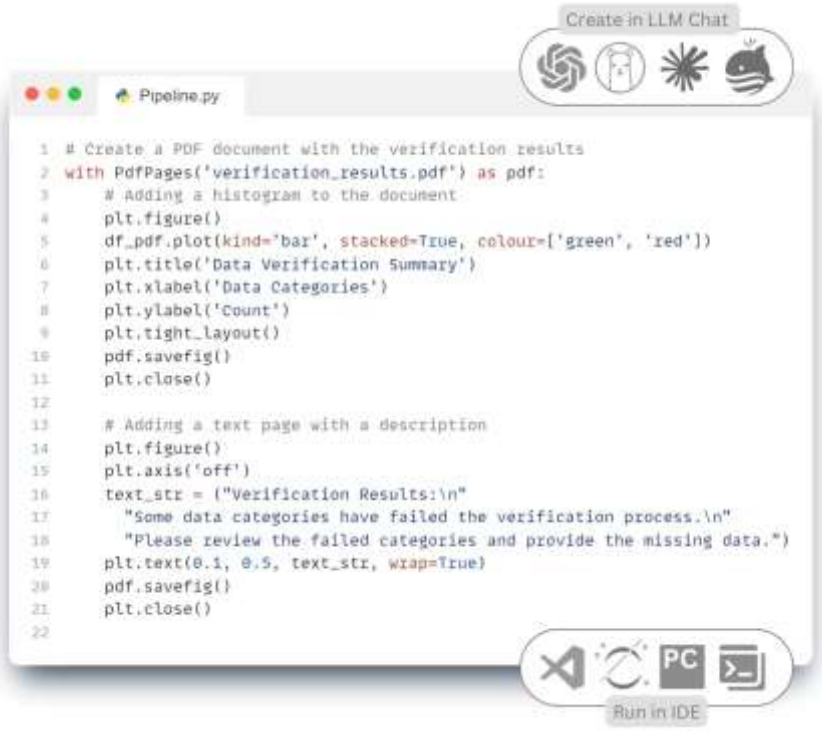
ETL ロード: PDF ドキュメントの自動作成

データのロード段階では、データを視覚化し、表やデータベースにアップロードするだけでなく、必要なグラフ、チャート、主要な分析指標を含むレポートを自動生成し、監査結果を待つマネージャーや専門家が受け取ることができます。自動化されたレポートには、データのコメントやテキスト解釈、表やグラフなどの視覚化要素を含めることができます。

- ヒストグラム（図 7.2-9）と、前章で実施した検証に基づく分析の説明を含む PDF -報告書を作成するには、LLM との対話の続きで、例えば 、要求を定式化すれば十分である：

> ヒストグラムと上記（チャットでの）データ検証結果の説明を含む PDF -ファイルを作成し、いくつかのカテゴリーが検証されておらず、欠落している埋める必要があることを警告するテキストを記述するコードを記述する。 ⏎

- コードとしての LLM の回答と、結果を含む完成した PDF ：



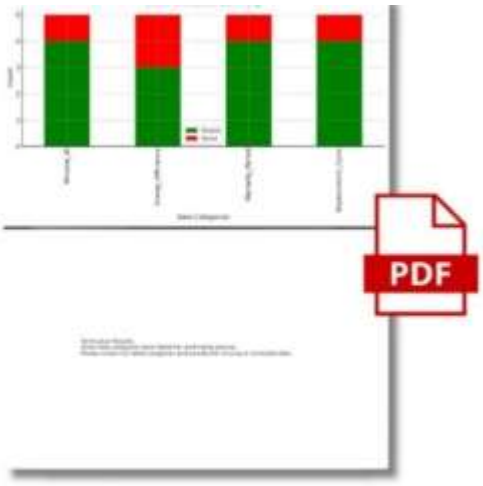


図 7.2-13 自動化されたコードは、テストデータのヒストグラムとテスト結果のテキストを含む PDF -文書を作成します。

LLM 、わずか 20 行のコードで自動的に記述されたソ リューションは、検証に合格したデータと不合格になったデータの数を属性ヒス トグラム（図 7.2-13）の形で視覚化し、結果を要約したテキスト・ブロックと今後の推奨追加した目的の PDF （または DOC）文書を即座に作成します。

文書作成の自動化は、ロード・ステージ の重要な要素である。特に、報告のスピードと正確さが重要なプロジェクト環境ではなおさらである。

ETL ロード: FPDF から文書を自動生成

ETL 段階でのレポーティングの自動化 ロード は、データ処理における重要なステップであり、特に、分析結果を伝達しやすく、理解しやすい形式で提示する必要がある場合に重要です。建設業界では、進捗報告書、プロジェクトデータの統計、品質保証報告書、財務文書などがこれに該当します。

このようなタスクのための最も便利なツールの 1 つは、Python と PHP の両方で利用可能なオープンソースライブラリ, FPDF です。

> オープンソースの FPDF ライブラリ は、コードを通じて文書を生成する柔軟な方法を提供し、ヘッダー、テキスト、表、画像を追加することができます。手作業による編集の代わりにコードを使用することで、エラーを減らし、PDF 形式でのレポート作成プロセスをスピードアップします。

PDF -文書を作成する際の重要なステップの一つは、コメントや説明の形で見出しやメインテキストを追加することです。しかし、レポートを作成する場合、テキストを追加するだけでなく、適切に構成することも重要です。見出し、インデント、行間 - これらすべてが文書の読みやすさに影響します。FPDF を使用すると、書式設定パラメータを設定し、要素の配置を制御し、文書のスタイルをカスタマイズすることができます。

FPDF は HTML と原理的に非常によく似ています。すでに HTML に慣れ親しんでいる方であれば、FPDF を使ってどのような複雑さの PDF 文書でも簡単に生成することができます。コード構造は HTML マークアップに非常によく似ており、ヘッダー、テキスト、画像、表が同様の方法で追加されるからです。HTML に不慣れな方でも心配はいりません - LLM を使えば、希望の文書レイアウトを生成するためのコードを即座に作成することができます。

- 次の例は、ヘッダーと本文を持つレポートを生成する方法を示しています。Python をサポートする IDE でこのコードを実行すると、目的のヘッダーとテキストを含む PDF -ファイルが作成されます：

```
from fpdf import FPDF  # FPDF ライブラリをインポート
pdf = FPDF()  # PDF を作成 -文書
pdf.add_page()  # ページを追加する

pdf.set_font("Arial", style='B', size=16) # フォントを設定：Arial、太字、サイズ 16
pdf.cell(200, 10, "Project Report", ln=True, align='C') # 見出しを作成し、中央に配置します。
pdf.set_font("Arial", size=12) # フォントを通常の Arial、サイズ 12 に変更します。
pdf.multi_cell(0, 10, "この文書には、プロジェクトファイルの検証結果のデータが含まれています...")# 複数行のテキストを追加
pdf.output(r "C:☆reportsreport.pdf") # PDF を保存 -file
```



図 7.2-14 数行の Python コードで、必要な PDF テキスト文書を自動生成することができます。

報告書を作成する際には、文書の元となるデータが静的であることは稀であることを考慮することが重要である。ヘッダ、テキストブロック（図 7.2-14）は、ETL プロセス の変換ステージ で値を受け取り、動的に形成されることが多い。

> このコードを使用すると、プロジェクト名、レポート作成日、参加者に関する情報、現在のステータスなど、最新の情報を含む文書を作成することができます。コード内で変数を使用することで、レポート内の必要な場所にこれらのデータを自動的に挿入することができ、送信前に手動で編集する必要が完全になくなります。

プロジェクト文書では、単純なテキストや見出しに加え、表が特別な位置を占める。オブジェクトの説明から検査結果まで、ほとんどすべての文書に構造化データが含まれています。Transform 段階のデータに基づいて表を自動生成することで、文書作成プロセスをスピードアップできるだけでなく、情報転送時のエラーも最小限に抑えることができます。FPDF では、 表を PDF - ファイルに （テキストまたは画像として） 挿入することができ、セルの枠線・列サイズ・フォントを設定することができます （図 7.2-15）。これは特に、動的データを扱う際に便利であり、行数・列数が文書タスクによって変わりうるときに利用できます。

- 次の例では、材料、見積もり、パラメータテスト結果のリストなどの表を自動作成する方法を示します：

```
data = [
[項目", "数量", "価格"], # 列見出し
["Concrete", "10 m³", "$ 500."], # 1 行目のデータ。
["Rebar", "2 tonnes", "$ 600"], # 2 行目のデータ.
[レンガ", "5,000 個", "$ 750."], # 3 行目のデータ。
]

pdf = FPDF () # PDF を作成 -document
pdf.add_page() # ページを追加する
pdf.set_font("Arial", size=12) # フォントを設定する
```

```
for row in data: # テーブルの行を検索する
for item in row: # 行内のセルを検索します。
pdf.cell(60, 10, item, border=1) # 幅 60、高さ 10 の枠線を持つセルを作成します。
pdf.ln() # 次の行に移動する
pdf.output(r "C: ¦ Publishertable.pdf") # PDF を保存 -file
```



図 7.2-15 変換ステップから、テキストだけでなく任意の表情報を PDF に自動生成することができます。

実際の報告シナリオでは、表は通常、データ変換段階で得られた情報を動的に生成したものです。この例（図 7.2-15）では、表は静的な形で PDF 文書に挿入されています：この例のデータはデータ辞書（コードの 1 行目）に配置されていますが、実際の状況では、このようなデータ変数は、データフレームのグループ化の後などに自動的に埋められます。

実際には、このような表は、さまざまな動的ソース（データベース、Excel -ファイル、API -インターフェース、分析計算の結果）から来る構造化データに基づいて構築されることが多い。多くの場合、ETL（Transform）の段階で、データは集約され、グループ化され、フィルタリングされ、その後、レポートに表示されるグラフや 2 次元テーブルの形で合計に変換される。つまり、表の内容は、選択したパラメータ、分析期間、プロジェクトフィルター、ユーザー設定によって変化します。

> Transform ステージでダイナミックなデータフレームとデータセットを使用することで、Load ステージでのレポーティングプロセスは、手作業による介入を必要とせず、可能な限り柔軟でスケーラブル、かつ簡単に繰り返し行えるようになっている。

表やテキストに加え、FPDF は表形式データのグラフの追加もサポートしています。これにより、Matplotlib や上記で説明したその他の可視化ライブラリで生成した画像をレポートに埋め込むことができます。グラフ、チャート、ダイアグラムはコードを使ってドキュメントに追加することができます。

- Python ライブラリ FPDF を使って、Matplotlib であらかじめ生成しておいたグラフを PDF 文書に追加してみましょう：

```
import matplotlib.pyplot as plt # Matplotlib インポートしてプロットを作成する。

fig, ax = plt.subplots() # グラフの図と軸を作成する
categories = ["Concrete", "Rebar", "Brick"] # カテゴリ名
values = [50000, 60000, 75000] # カテゴリーの値
ax.bar(categories, values) # 棒グラフを作成する
plt.ylabel("Value,$.") # Y 軸に符号を付ける
plt.title("Cost Distribution") # タイトルを追加する
plt.savefig(r "C:\reports\chart\chart\chart.png") # Save the chart as an image

pdf = FPDF () # PDF を作成 -document
pdf.add_page() # ページを追加する
pdf.set_font("Arial", size=12) # フォントを設定する
pdf.cell(200, 10, "Cost Chart", ln=True, align='C') # ヘッダを追加します。

pdf.image(r "C:㊟㊟㊟㊟.png", x=10, y=30, w=100) # 画像を PDF に挿入 (x, y - 座標、w - 幅)
pdf.output(r "C:¦チャート¦レポート.pdf") # PDF ファイルを保存する
```

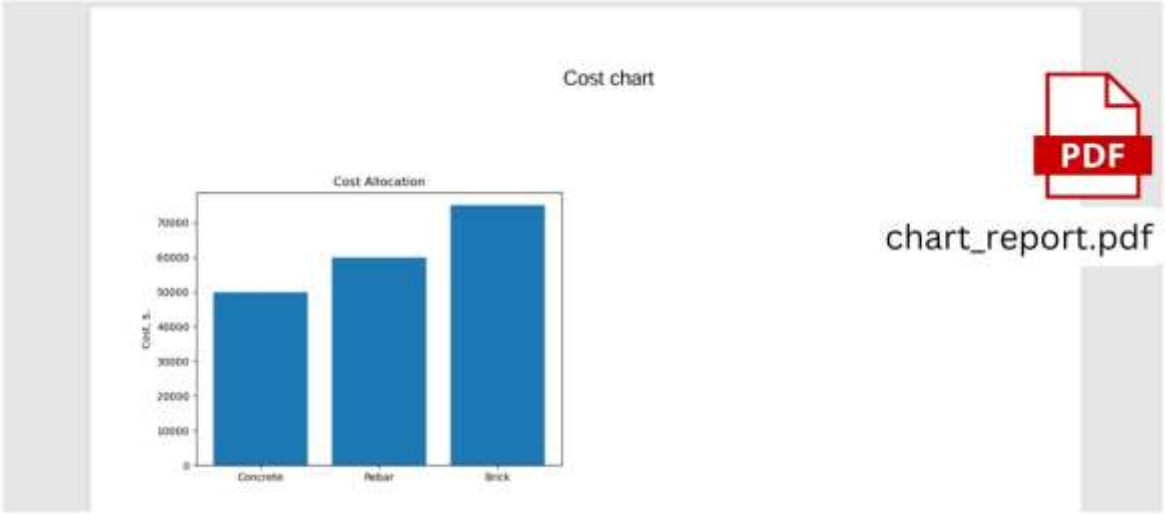


図 7.2-16 数十行の コ ー ド で、 グ ラ フ を生成 し て保存 し 、 それを PDF 文書に貼 り 付け る こ と ができ ます。

FPDF 、文書作成とロジックのプロセスが透明化され、迅速かつ便利になります。コードに組み込まれたテンプレートにより、最新のデータで文書を作成することができ、手入力の必要性がなくなります。

> ETL 自動化の使用 - 時間のかかる手作業によるレポーティングの代わりに、専門家は、明確なユーザーインターフェイスを持つ特定のデータサイロで作業するための適切なツールを選択するのではなく、データの分析と意思決定に集中することができます。

このように、FPDF ライブラリ は、短い技術レポートから表やグラフを含む複雑な分析サマリーまで、あらゆる複雑さのドキュメントを自動作成するための柔軟なツールを提供します。これにより、ドキュメントの流れをスピードアップできるだけでなく、手作業によるデータ入力や書式設定に伴うエラーの確率を大幅に減らすことができます。

ETL ロード: レポーティングと他システムへのロード

ロード段階では、結果は表、グラフ、および最終的な PDF レポートの形で生成され、確立された要件に従って作成されます。さらに、ERP、CAFM、CPM、BI プラットフォーム、その他の企業や業界のソリューションなどの外部システムとの統合に必要な、このデータを機械可読形式（CSV など）にエクスポートすることも可能です。CSV に加え、XLSX、JSON、XML、または自動情報交換をサポートするデータベースへの直接アップロードも可能です。

- ロード・ステップを自動化するための適切なコードを生成するには、LLM インターフェースに問い合わせるだけでよい：ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude または QWEN ：

> データ検証結果のレポートを DataFrame で生成するコードを書く。ここで、先頭に'verified_'が付いた列はカウントされ、'Passed'と'Failed'にリネームされ、欠損値は 0 に置き換えられ、すべての検証に合格した行のみが CSV -ファイルにエクスポートされる。

- LLM の回答 ：



図 7.2-17 最終データフレームからの変換ステップで得られた検証済みデータは、他のシステムとの統合のために、CSV - ファイルにエクスポートされる。

上記のコード（図 7.2-17）は、ETL -プロセスの最終段階であるロード（Load）を実装しており、検証されたデータは、ほとんどの外部システムやデータベースと互換性のある CSV 形式で保存される。このように、抽出、変換、可視化、文書化、必要なシステムやフォーマットへのデータのエクスポートを含む、ETL -プロセスの全サイクルが完了した。

ETL パイプラインは、単一プロジェクトの処理にも、数百または数千の文書、画像、スキャン、CAD プロジェクト、点群、PDF ファイル、または分散システムからのその他のソースの形で入力されるデータを分析する大規模なアプリケーションにも使用できます。プロセスを完全に自動化できるため、ETL は単なる技術的な処理ツールではなく、デジタル建設情報インフラの基盤となっています。

ETL with LLM: PDF からのデータの可視化 -documents

今こそ、抽出、変換、ロードというデータ処理の主要な段階を 1 つのシナリオでカバーする本格的な ETL プロセスの構築に移行する時です。PDF 文書を手作業なしで処理できる自動化された ETL パイプラインを構築し、文書からデータを抽出し、視覚化し、分析し、他のシステムに転送できるようにしましょう。

この例の ETL プロセスはプロンプトによって記述され、プロンプトは言語モデル（LLM）に対して、最終的な結果を説明しながらすべての ETL プロセスを説明しなければならない。この場合、タスクは、指定されたフォルダとそのサブフォルダ内のすべての PDF ファイルを検索し、そこから関連情報（材料名、数量、コストなど）を抽出し、その結果を構造化テーブル（DataFrame）として提示し、さらに分析することです

- LLM の最初のテキストクエリで、複数の PDF 文書からデータを自動的に抽出し、Extract ステップのためのデータのデータフレームを作成する：

> 与えられたフォルダとそのサブフォルダ内の PDF ファイルから材料情報を抽出するコードを記述する。PDF 内のデータには、材料名、数量、コストが含まれます。結果は DataFrame に保存されなければならない。 ⏎

LLM の回答 ：

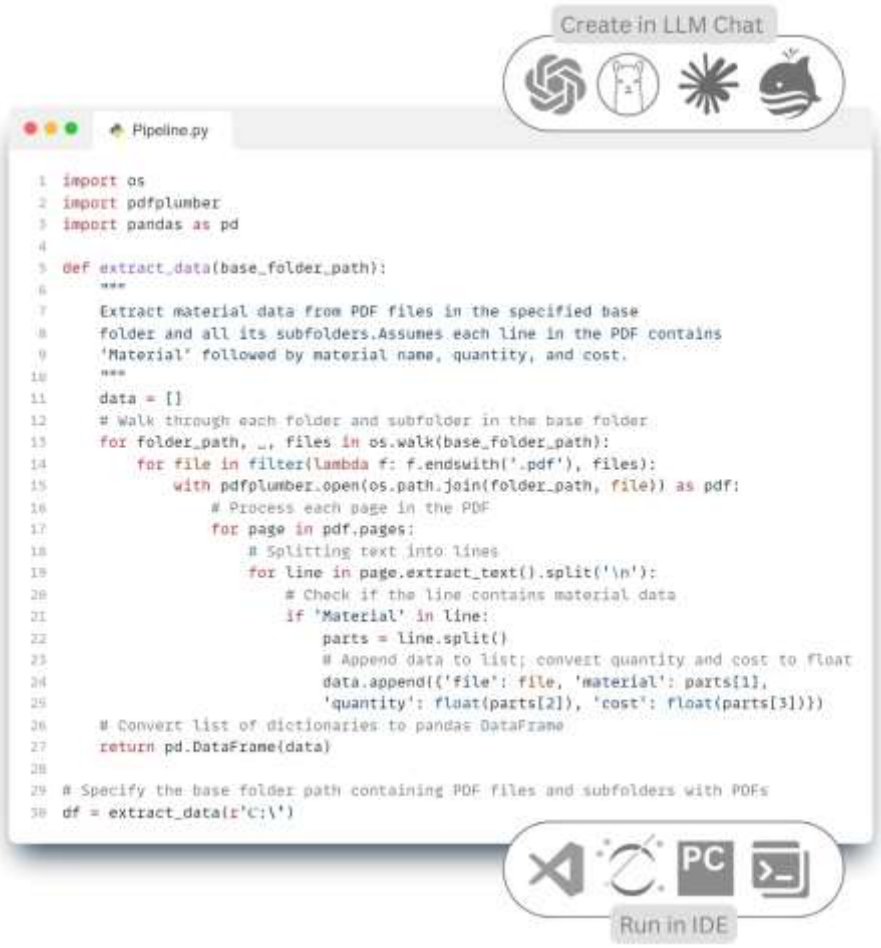


図 7.2-18 LLM は、特定のフォルダとそのすべてのサブフォルダ内の PDF - ファイルからデータを抽出する Python - コードを作成します。

LLM レスポンス （図 7.2-18） は、既製の Python スクリプトであり、これは自動的にすべてのフォルダをトラバースし、見つかった PDF -ファイルを開き、そこからテキスト情報を抽出して表に変換します。出来上がったチャットコードは、PyCharm、Visual Studio Code (VS Code)、Jupyter Notebook 、Spyder、Atom、Sublime Text、Eclipse with PyDev plugin、Thonny、Wing IDE、IntelliJ IDEA with Python plugin、JupyterLab、または人気のあるオンラインツール Kaggle.com、Google Collab、Microsoft Azure Notebooks、Amazon SageMaker で実行できます。

あるいは、コレクション処理を単純化するために、LLM からコードをコピーして IDE でコードを使用するのではなく、PDF - ファイル 12 個を直接 LLM チャットにアップロードして （図 7.2-19）、コ

ードを見る必要も実行する必要もなく、表を出力として得ることもできます。このコードの実行結果は、選択した属性の表になります。

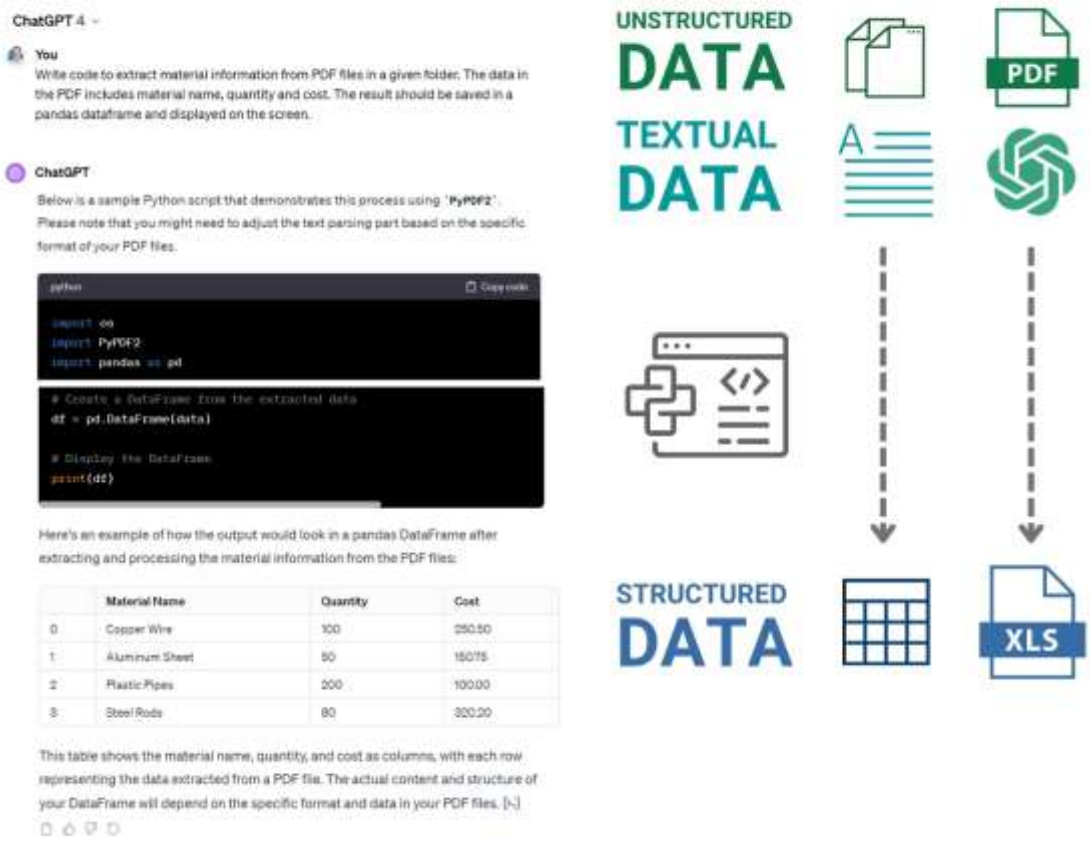


図 7.2-19 PDF - ファイルから抽出したデータを、選択した属性を持つ構造化データフレーム表示で表示する LLM のコード実行結果。

次のステップでは、得られたデータの言語モデルを求めます。例えば、材料の使用コストと使用量を比較し、さらなる分析の基礎となるいくつかのビジュアライゼーションの例を作成します。

- トランスフォームのステップで作成した表（図 7.2-18）から、いくつかのグラフをプロットするよう、LLM と引き続きチャットで依頼する：

> DataFrame （図 7.2-18）から各材料の総コストと数量を視覚化する↵

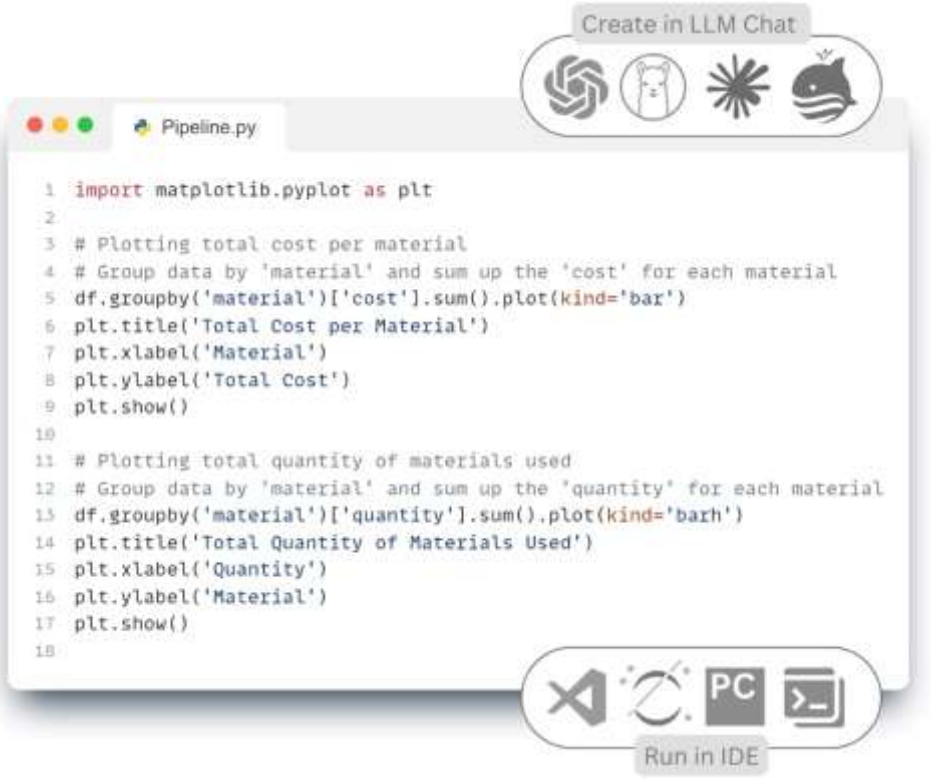


図 7.2-20 LLM 応答 -モデルを Python コードとして matplotlib ライブラリを使用してデータフレームからデータを可視化する。

LLM は、matplotlib ライブラリを使用して、Python -コード（図 7.2-20）を自動的に生成し、実行する。このコードを実行すると、建設プロジェクトのコストと材料使用量のグラフが直接チャットで表示され（図 7.2-21）、分析作業が大幅に簡素化される。

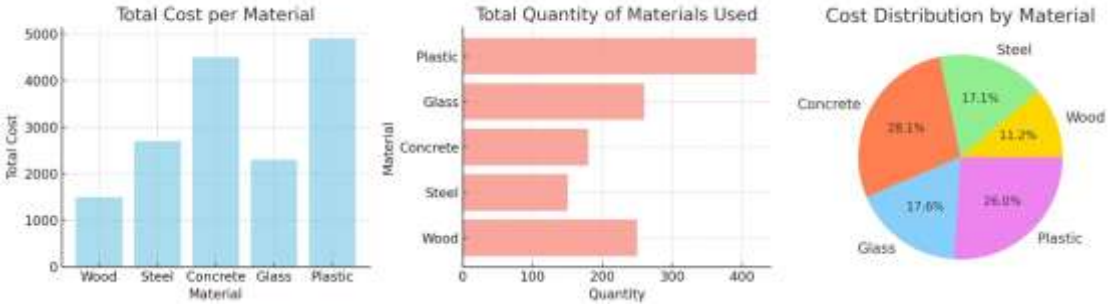


図 7.2-21 LLM レスポンスの視覚化 DataFrame で収集されたデータに基づくグラフとしての LLM レスポンス 。

ETL コードの記述、コードの分析と実行、結果の視覚化といったアイデアの開発支援は、プログラミングの基

礎を学ぶ必要なく、LLM のシンプルなテキストクエリを通じて利用できる。LLM のような AI ツールの出現は、プログラミングとデータ処理の自動化へのアプローチを確実に変えつつある（図 7.2-22）。

> PwC のレポート "What is the real value of artificial intelligence for your business and how can you capitalise on it? "によると、人工知能のビジネスにとっての本当の価値とは何か？(2017 年) [139]によると、プロセスの自動化と生産性の向上が経済成長の主な原動力になるという。そして、生産性の向上は、2017 年から 2030 年にかけて、AI による GDP 成長の 55%以上を占めると予想されている。"

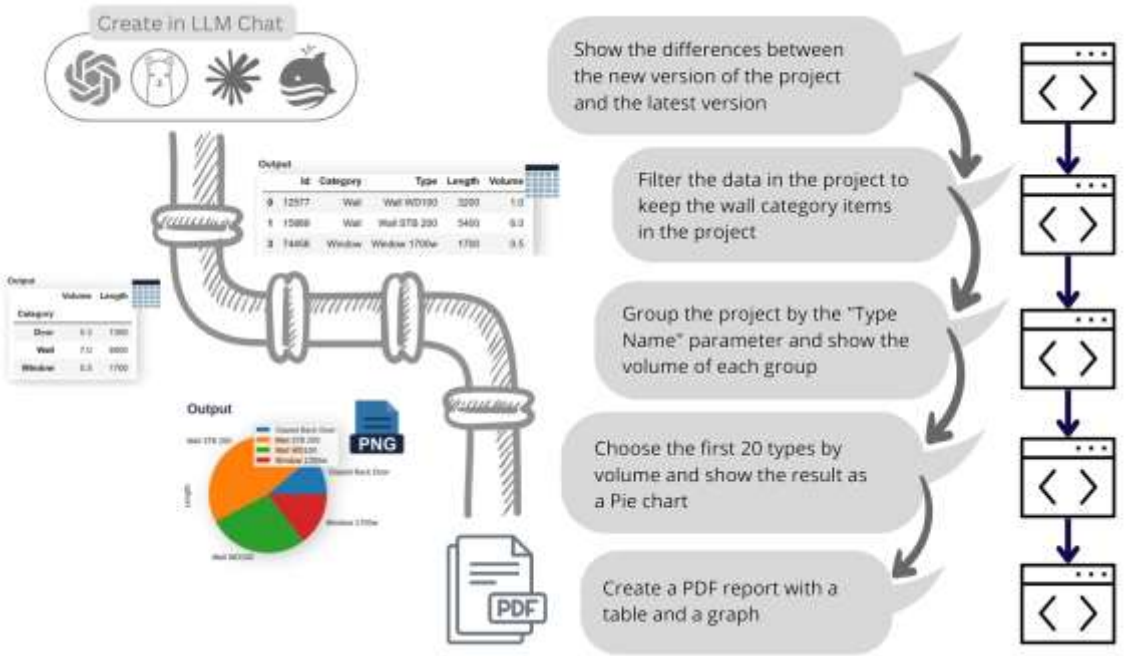


図 7.2-22 AI LLM は、LLM を必要とせずに将来のプロジェクトに適用されるドラフトコードを生成するのに役立つ。

> ChatGPT 、LlaMa 、Mistral 、Claude 、DeepSeek 、QWEN 、Grok のようなツールや、オープンデータ 、オープンソースソフトウェア を使うことで、以前は専門的で高コスト、メンテナンスが困難なモジュール式のプロプライエタリシステムでしかできなかったプロセスを自動化することができる。

建設業界においては、自動化されたパイプライン - データプロセスをいち早く導入した企業が、プロジェクト管理の効率化、財務損失の削減、断片化されたアプリケーションやサイロ化されたデータウェアハウスの排除など、大きなメリットを享受できることを意味する。

ETL プロセス）におけるビジネスタスク実行の記述されたロジックは、アナリティクス（）とデータ処理プロセスの自動化の重要な部分であり、より広範な概念であるパイプライン（Pipelines）の特定のバリエーションである。

章 7.3.

自動 etl コンベヤ (パイプライン)

パイプライン: 自動 ETL コンベア データ

ETL プロセス は、構造化ソースと非構造化ソースの両方を対象とする分析システムにおけるデータ処理に伝統的に使用されてきた。しかし、今日のデジタル環境では、より広範な用語であるパイプライン （コンベア ）が使用されるようになってきている。これは、あるステージの出力が次のステージの入力となる、あらゆるシーケンシャルな処理の連鎖を表す。

このアプローチは、データだけでなく、タスク処理、レポート作成、ソフトウェアとの統合、デジタルワークフロー （図 7.3-1）といった他のタイプの自動化にも適用される。

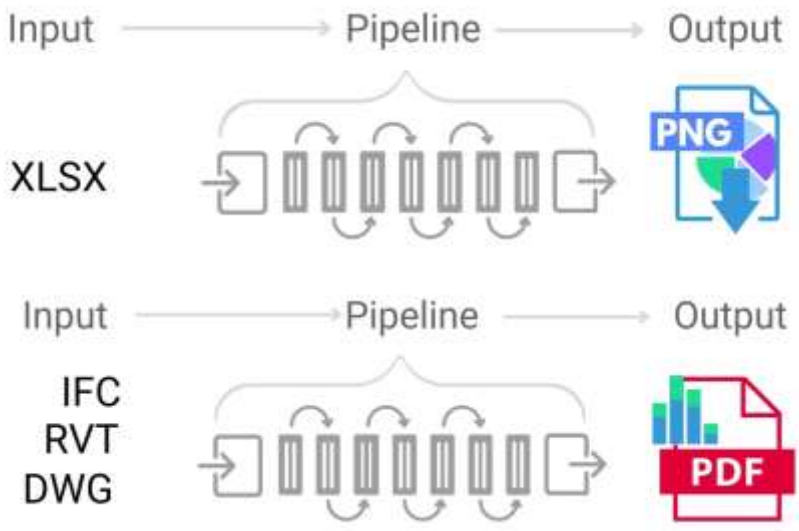


図 7.3-1 パイプライン は、ある段階の出力が次の段階の入力となる処理シーケンスである。

パイプライン の使用は、特に大量の異種データを扱う場合、自動化の主要な要素の 1 つである。パイプライン・アーキテクチャは、複雑な処理ステップを、モジュール化された一貫性のある管理しやすい形式で整理することを可能にし、可読性を高め、コードのメンテナンスを簡素化し、インクリメンタルなデバッグとスケーラブルなテストを可能にする。

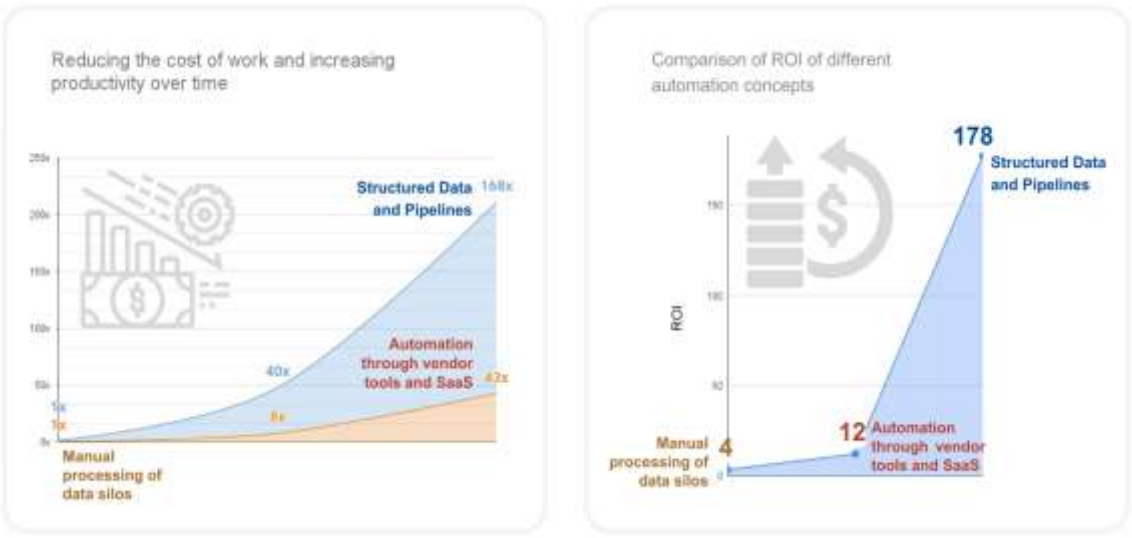


図 7.3-2 ROI パイプライン データ検証プロセスは、古典的なツール [74]を使った処理に比べ、実行時間を数十倍から数百倍に短縮する。

専用システム（ERP 、PMIS 、CAD 、その他）における手作業とは異なり、パイプラインを利用することで、作業のスピードを大幅に向上させ（図 7.3-2）、繰り返し作業を回避し、適切なタイミングでのプロセス開始を自動化することができる（図 7.3-3）。

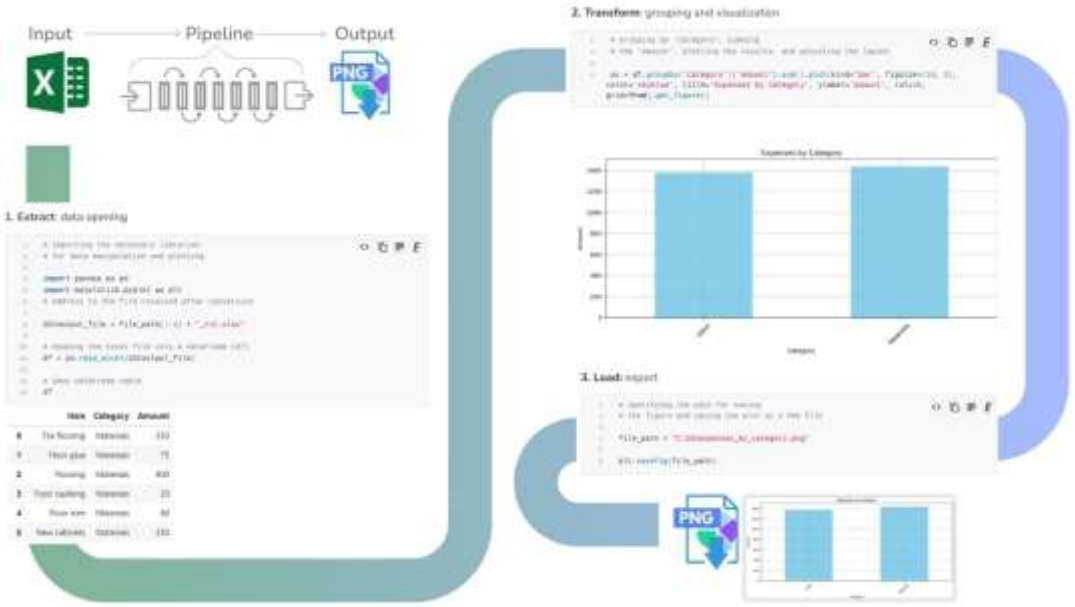


図 7.3-3 ETL の例 パイプライン エクセルを開かずに XLSX ファイル の表形式データからチャートを自動的に取り出すことについて.

ストリーミング・データを処理し、ETL プロセスと同様に自動化されたパイプラインを構築するには、事前にデータ・ソースと、それを収集する時間枠を特定する必要がある。

> 建設プロジェクトでは、データは更新間隔が異なる多くの異種ソースから取得される。信頼性の高いデータショーケースを作成するためには、情報の取得と更新のタイミングを記録することが重要です。これにより、タイムリーな意思決定が可能になり、プロジェクト管理の効率が向上します。

例えば、終業時刻の 19:00 などである。この時点で、さまざまなシステムやストレージからデータを集約する最初のスクリプトが起動する（図 7.3-4 ステップ 1）。続いて、データの自動処理と分析に適した構造化フォーマットへの変換が行われる（図 7.3-4 ステップ 2-4）。最終段階では、準備されたデータを使用して、レポート、ダッシュボード、および前の章で説明したその他の製品が自動的に生成される（図 7.3-4 ステップ 6-7）。その結果、管理者は、朝の 5 時までに、必要なフォーマットで最新のプロジェクト・ステータス・レポートを入手することができる（図 7.3-5）。

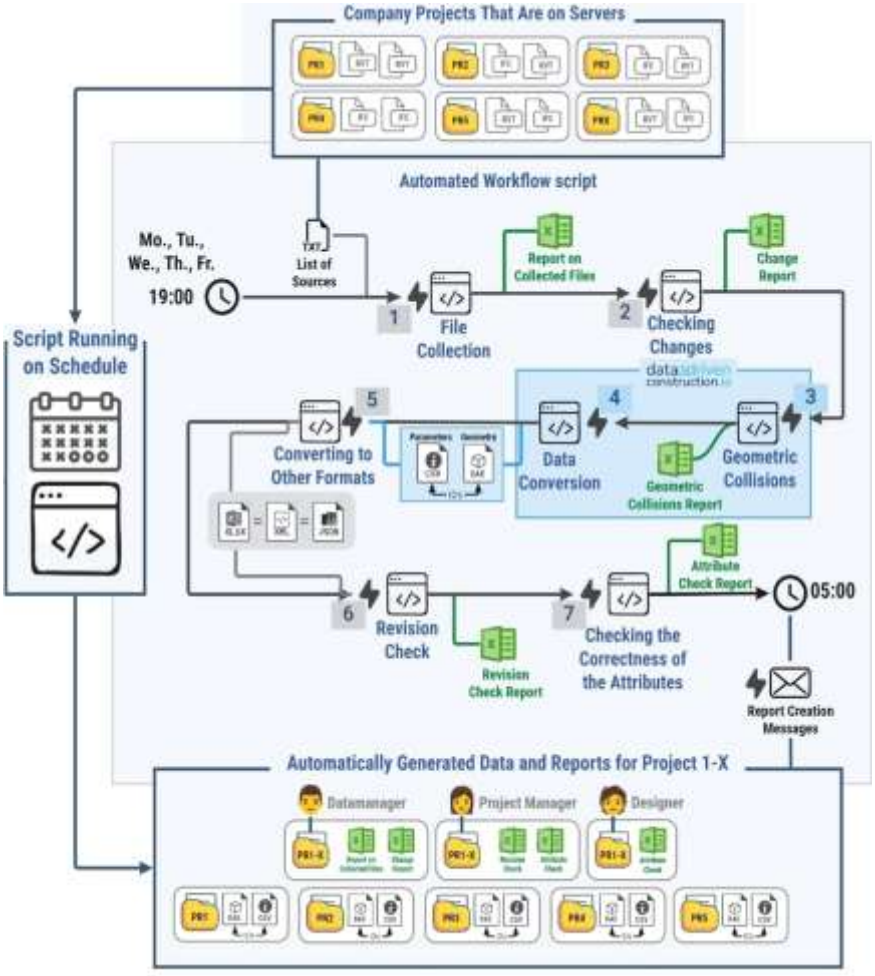


図 7.3-4 夕方に自動的に収集されるパイプライン のデータは、一晩中処理されるため、マネジャーは朝までに最新のレポートと新鮮なレポートを入手できる。

> タイムリーなデータ収集、KPI の定義、変革プロセスの自動化 、ダッシュボードによる可視化は、データ主導の意思決定を成功させるための重要な要素である。

このような自動化されたプロセス（図 7.3-4）は、完全な自律性で実行することができ る。スケジュールに従って実行され、無人でデータを処理し、クラウドまたは自社サーバーにデプロイ することができる（図 7.3-5）。このため、このような ETL パイプラインを既存の IT インフラに統合することができ、データに対する制御を維持し、柔軟なスケーリングを実現することができる。

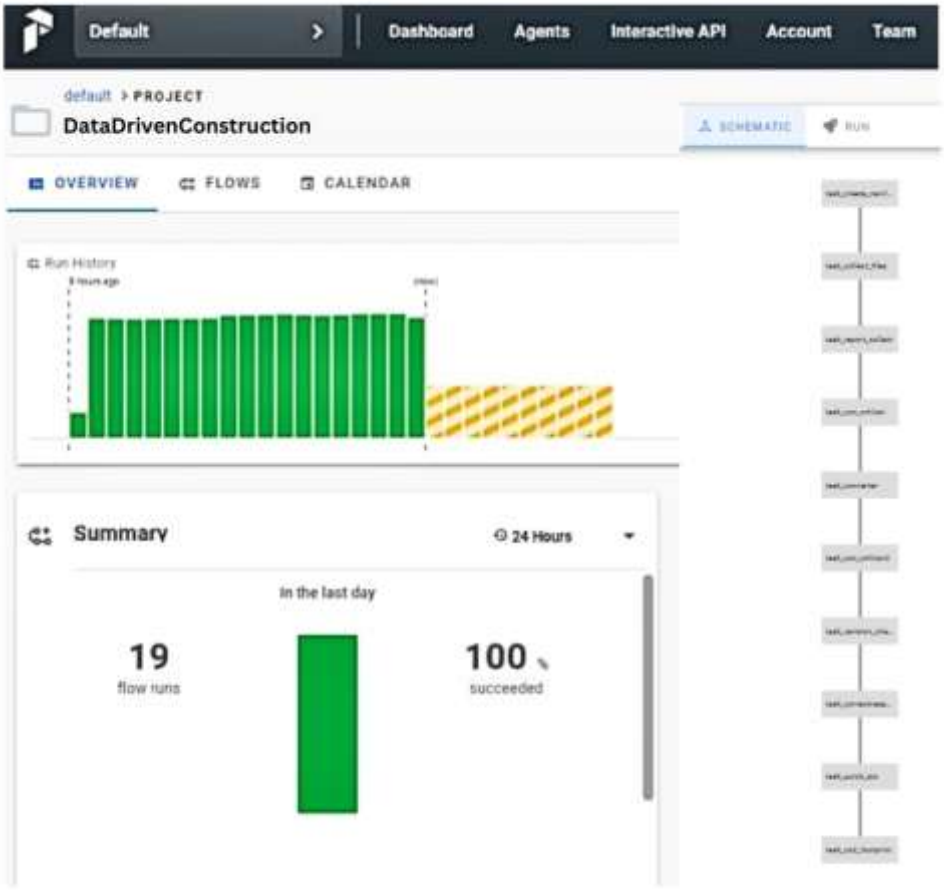


図 7.3-5 Prefect プラットフォーム上の自動 ETL -コンベア プロセス（図 7.3-4）では、毎日 19 時以降に 10 個の python スクリプトが交互に実行される。

ワークフローを自動化することで、より有意義でルーチンワークの少ないタスクに時間を割くことができるようになり、チームの生産性が向上するだけでなく、ビジネスプロセスに人工知能 (AI)の技術を取り入れるための重要な第一歩としても機能する。この技術については、予測分析と機械学習の章で詳しく説明する。

パイプライン -ETL LLM によるデータ検証プロセス

データ要件の作成と ETL の自動化に関する前章 では、データの準備、変換、検証、可視化のプロセスを段階的に分解した。これらのアクティビティは、それぞれ特定のタスクを実行する個別のコードブロック（図 7.2-18 ～図 7.2-20）として実装された。

次の目標は、これらの要素を 1 つの、首尾一貫した、データ処理の自動化されたパイプライン（パイプライン

、ETL -Pipeline -）に結合することである。パイプラインでは、すべてのステージ（ロード、検証、視覚化、エクスポート）が 1 つの自動実行可能なスクリプトで順次実行される。

以下の例では、完全なデータ処理サイクルが実装されます：ソース CSV ファイル →の読み込みから、正規表現を使用した構造と値のチェック →結果の計算→ PDF フォーマットでのビジュアルレポートの生成 まで。

- LLM に次のようなテキストクエリを実行すると、適切なコードを取得できます：

> CSV からデータをロードし、正規表現 を使用して DataFrame を検証し、'W-NEW' または 'W-OLD' 形式の識別子、'A' から 'G' までの文字によるエネルギー効率、年単位の数値による保証期間と交換サイクルをチェックし、最後に合格値と不合格値のカウントを含むレポートを作成し、結果のヒストグラムを含む PDF を生成し、テキストの説明を追加するコードサンプルを作成してください。 ⏎

LLM の回答 ：



図 7.3-6 パイプライン（ETL） データ処理の全サイクルを自動化：ダウンロードと検証から

PDF 形式の構造化レポートの作成まで.

LLM チャット・ルーム内または DIE 内の自動化コード（図 7.3-6）は、コードをコピーした後、指定された正規表現 を使って CSV -ファイルからデータを検証し、合格したレコード数と不合格になったレコード数のレポートを作成し、検証結果を PDF -ファイルとして保存します。

この ETL -コンベア構造は、データロードからレポート生成までの各ステップが独立したモジュールとして実装されており、透明性、拡張性、再現性を確保している。検証ロジックを読みやすい Python コード として提示することで、開発者だけでなく、データ管理、品質、分析の専門家にとってもプロセスが透明で理解しやすくなります。

> データ処理を自動化する Pipeline の - アプローチは、プロセスを標準化し、再現性を高め、新しいプロジェクトへの適応を簡素化します。これにより、コンプライアンステスト、レポート作成、外部システムへのデータ転送など、ソースやタスクの種類に関係なく、データを分析するための統一された手法が構築されます。

このような自動化により、人的ミスを減らし、独自のソリューションへの依存を減らし、結果の正確性と信頼性を高めることができるため、プロジェクトレベルの業務分析にも、企業レベルの戦略的分析いる。

パイプライン -ETL: CAD (BIM) におけるプロジェクト要素のデータと情報の検証。

CAD システムやデータベースからのデータ （BIM ）は、建設会社のビジネスにおいて、最も洗練され、動的に更新されるデータソースの一部です。これらのアプリケーションは、ジオメトリを使用してプロジェクトを記述するだけでなく、ボリューム、材料特性、部屋の割り当て、エネルギー効率レベル、許容差、寿命、その他の属性など、複数のレイヤーのテキスト情報で補完します。

> CAD -モデルのエンティティに割り当てられた属性は、設計段階で形成され、原価計算、スケジューリング、ライフサイクル評価、ERP および CAFM -システムとの統合を含む、さらなるビジネスプロセスの基礎となる。

CAD- （BIM-属性の検証に対する従来のアプローチは、手作業による検証を伴う（図 7.2-1）。現代の建設プロ

ジェクトの量と数、およびそれらの定期的な更新を考慮すると、データの検証および変換のプロセスは、持続不可能であり、手が届かなくなります。

> ゼネコンやプロジェクトマネージャーは、同じモデルの複数のバージョンや断片を含む、大量のプロジェクトデータを処理する必要性に直面している。これらのデータは、RVT、DWG、DGN、IFC、NWD、およびその他のフォーマットで設計組織から提供され（図 3.1-14）、業界標準や企業標準に準拠しているかどうかを定期的にチェックする必要があります

手作業と専用ソフトウェアに依存しているため、全社モデルのデータに関連するワークフローでは、データ検証プロセスがボトルネックになっている。自動化と構造化された要件の使用により、この依存性を排除し、データ検証のスピードと信頼性を飛躍的に向上させることができる（図 7.3-7）。

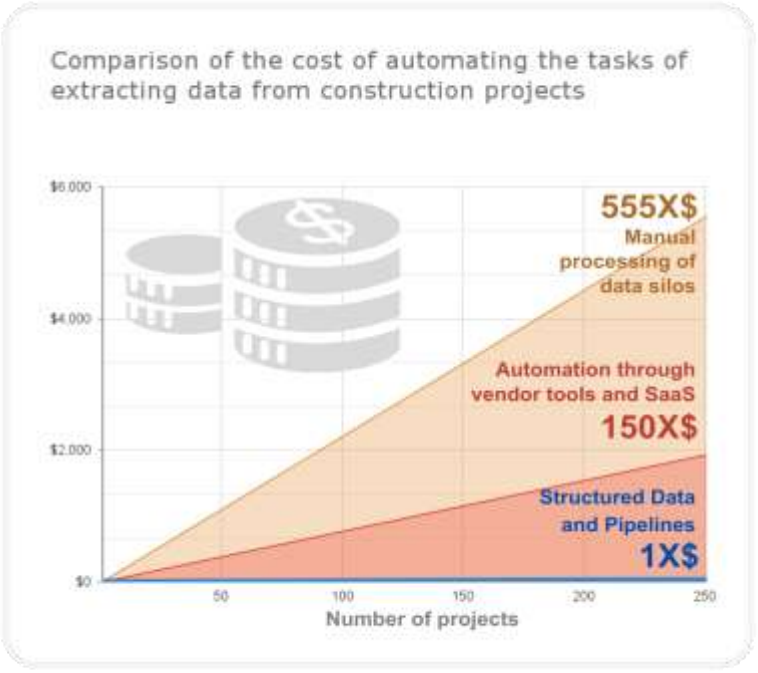


図 7.3-7 自動化によってデータの検証や処理のスピードが上がり、作業コストが数十分の 1 に削減される[140]。

CAD データ検証プロセスには、さまざまなクローズドフォーマット（RVT 、DWG 、DGN 、NWS など）またはオープンな半構造化フォーマットおよびパラメトリックフォーマット（IFC 、CPXML、USD）からのデータ抽出（ETL ステージ 抽出）が含まれます。)、またはオープンな半構造化フォーマットおよびパラメトリックフォーマット(IFC, CPXML, USD)で、正規表現 RegEx (図 7.3-8)を使用して、各属性とその値にルールテーブルを適

用することができます(Transform stage)。このプロセスについては、本書の第 4 部で詳しく説明しました。

と正常に検証されたレコードの PDF エラーレポートの作成は、さらなる処理に使用できる検証されたエンティティのみを考慮した構造化フォーマットでの出力（Load ステップ）で完了すべきである。

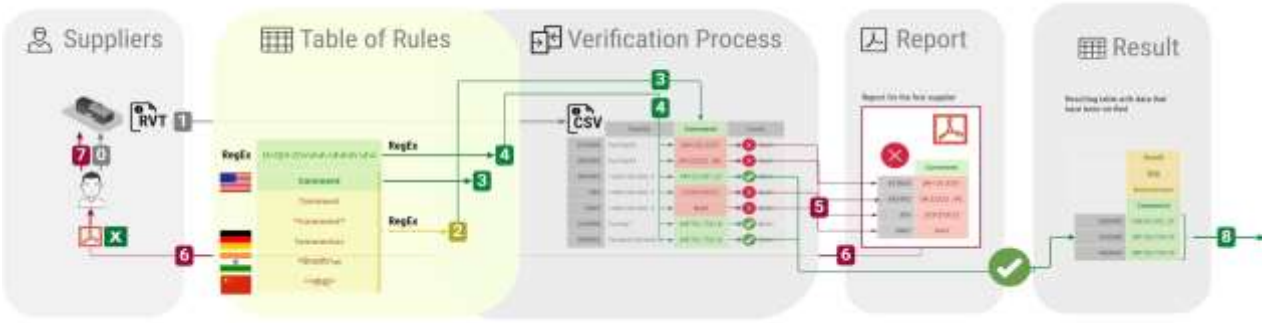


図 7.3-8 正規表現を使って検証されたプロジェクト・データ・プロバイダーから最終報告書までのデータ検証プロセス.

CAD システム (BIM)から構造化された要件でデータを検証し、ETL-Pipelines (図 7.3-9)を通して処理される新しいデータをストリーミングすることで、検証プロセスにおける手作業の必要性を減らすことができます（検証およびデータ要件の各プロセスについては、前の章で説明しました）。

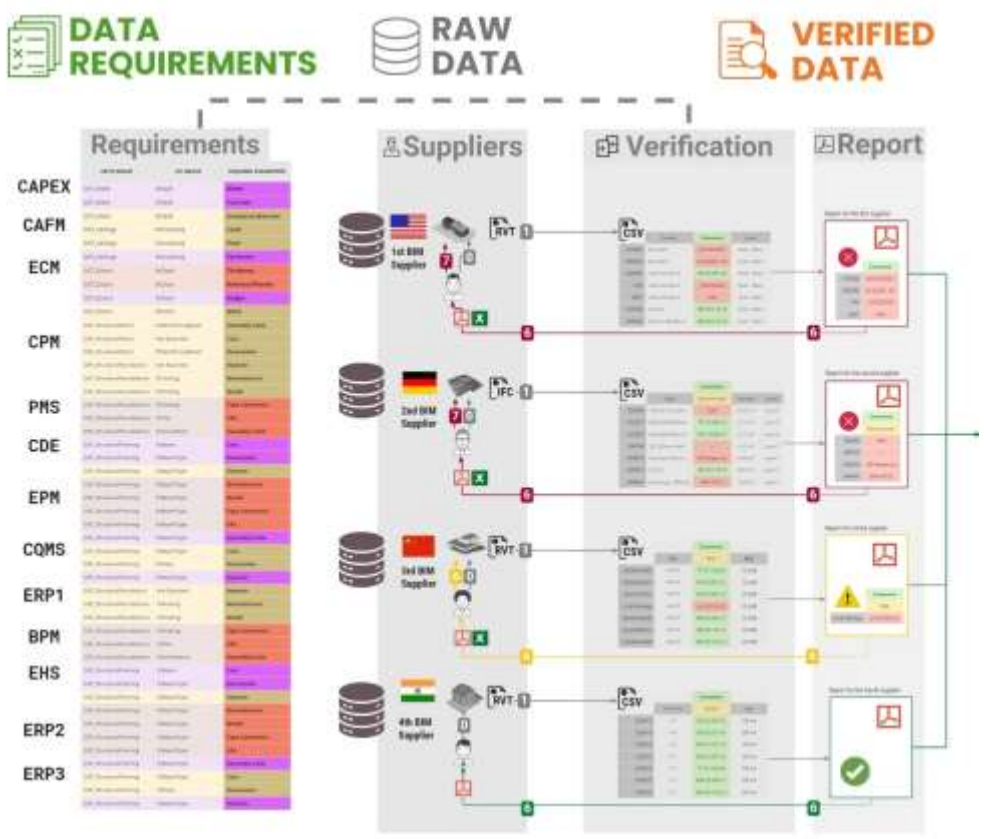


図 7.3-9 ETL によるデータ検証の自動化は、プロセスをスピードアップすることにより、建設プロジェクト管理を簡素化する。

従来、請負業者や CAD（BIM）の専門家から提供されたモデルの検証には、数日から数週間かかることがあった。しかし、自動化された ETL プロセスの導入により、これを数分に短縮することができる。典型的なケースでは、請負業者は次のように述べています：「モデルは検証され、適合しています。この発言から、請負業者のデータ品質に関する主張の検証の連鎖が始まる：

- プロジェクト・マネージャー 「請負業者は『モデルはテスト済みで、すべて問題ない』と言っています。
- データ・マネージャー - ロード・バリデーション：
  - Pandas のシンプルなスクリプトが数秒で違反を検出。自動化によって紛争はなくなります：
    - カテゴリ: OST_StructuralColumn、パラメータ: FireRating IS NULL。
    - 違反 ID のリストを作成→ Excel/PDF にエクスポート。

Pandas のシンプルなスクリプトは、数秒で侵入を検知する：

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"]. # フィルタリング
issues = df[df["FireRating"].isnull()] # 空の値 # 空の値
issues["ElementID"]].to_excel("fire_rating_issues.xlsx") # ID のエクスポート
```

- データマネージャーからプロジェクトマネージャーへ - 「をチェックしたところ、18 列に FireRating パラメータが設定されていませんでした。
- プロジェクトマネージャーから請負業者へ - 「モデルを修正のため返却します。

その結果、CAD モデルはバリデーションを受けることなく、自動化によって論争がなくなり、請負業者は問題のある要素の ID リストを含む構造化されたレポートをほぼ瞬時に受け取ることができる。このようにして、検証プロセスは透明化され、再現可能になり、ヒューマンエラーから保護される（図 7.3-10）。

> このアプローチは、データ検証プロセスを手作業による品質管理ではなく、エンジニアリング機能に変える。これにより生産性が向上するだけでなく、同社のすべてのプロジェクトに同じロジックを適用することが可能になり、設計から運用に至るまで、エンド・ツー・エンドのデジタル・トランスフォーメーションが可能になる。

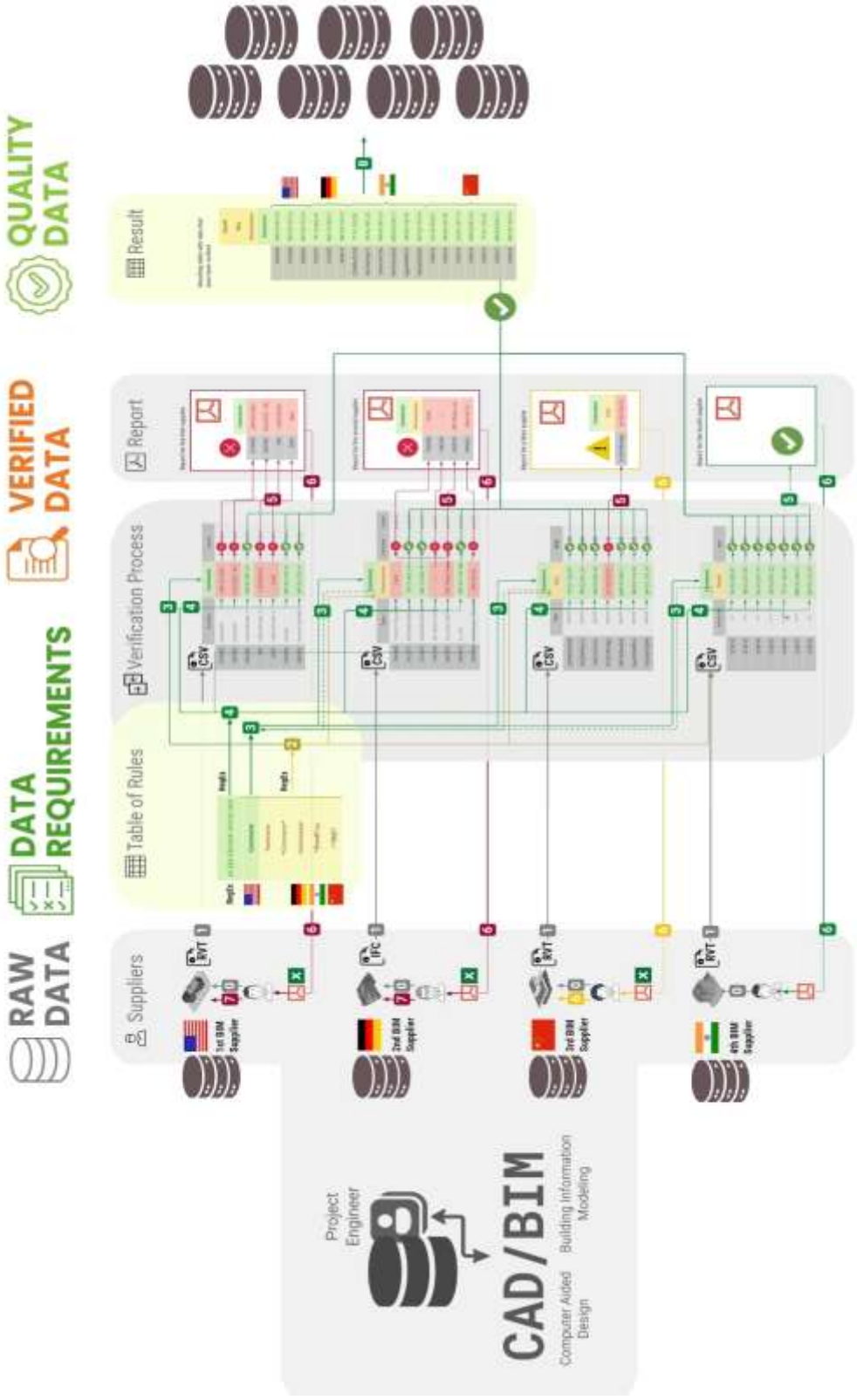

QUALITY DATA
VERIFIED DATA
DATA REQUIREMENTS
RAW DATA
Result
Report
Verification Process
Table of Rules
Suppliers
CAD/BIM
Computer Aided Design
Building Information Modeling
Project Engineer

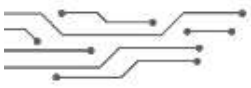
図 7.3-10 要素属性のチェックを自動化することで、人為的なミスをなくし、エラーの確率を下げます。

自動化されたパイプライン（図 7.3-10）を使用することで、CAD- (BIM-)システムからの高品質なデータを期待するシステムユーザーは、必要な出力データ（表、文書、画像）を即座に入手し、作業タスクに素早く統合することができます。

> 制御、処理、分析の自動化は、建設プロジェクト管理への取り組み方、特に複雑で高価なモジュール式の独自システム）やクローズド・ベンダー・ソリューション（）を使用しない、異なるシステムの相互運用性に変化をもたらしている。

コンセプトやマーケティングの略語は生まれては消えていくが、データ要件の検証プロセスそのものは、永遠にビジネスプロセスの不可欠な一部であり続けるだろう。建設業界は、より専門的なフォーマットや標準を作成するのではなく、他業界ですでにその有効性が実証されているツールに目を向けるべきである。今日、データ処理とプロセス統合を自動化する強力なプラットフォームがあり、これによって企業は、定型業務に費やす時間を大幅に短縮し、Extract 、Transform 、Load におけるエラーを最小限に抑えることができる。

ETL プロセスの自動化とオーケストレーションのためのソリューションの代表的な例として、 Apache Airflow があり、複雑な計算プロセスを組織化し、ETL パイプラインを管理することができる。Airflow と並んで、データルーティングとストリーミングのための Apache NiFi や、ビジネスプロセスの自動化のための n8n など、他の類似ソリューションも活発に使用されている。



章 7.4.

etl とワークフローのオーケストレーション：実践的ソリューション

DAG および Apache Airflow: ワークフローの自動化とオーケストレーション

Apache Airflow は、ワークフロー（ETL -コンベア）を自動化、オーケストレート、監視するために設計された、フリーでオープンソースのプラットフォーム である。

大量のデータを扱うことは日常的に要求される：

- さまざまなソースからファイルをダウンロード - （サプライヤーや顧客などから） 抽出。
- このデータを必要な形式に変換する - を変換する（構造化、クレンジング、検証）。
- 検証のために結果を送信し、レポートを作成する - をロードする（必要なシステム、文書、データベース、ダッシュボードにアップロードする。

このような ETL プロセスを手作業で実行すると、かなりの時間がかかり、人為的ミスのリスクが生じる。データソースの変更やステップの 1 つでの失敗は、遅延や不正確な結果を引き起こす可能性があります。

Apache Airflow のような自動化ツールは、信頼性の高い ETL -conveyor を構築し、エラーを最小化し、処理時間を短縮し、各ステージでデータが正しいことを保証します。Apache Airflow の中核にあるのは DAG （Directed Acyclic Graph）の概念で、各タスク（オペレータ）が他の依存関係に接続され、指定された順序で厳密に実行される有向非循環グラフです。DAG はサイクルを排除し、タスク実行の論理的で予測可能な構造を提供する。

タスク間の依存関係の管理、実行スケジュールの制御、ステータスの追跡、障害への自動対応など、Airflow がオーケストレーションの面倒を見ます。このアプローチにより、手作業による介入を最小限に抑え、プロセス全体の信頼性を確保します。

> Task Orchestrator は、複雑なコンピューティングと情報環境におけるタスク実行を管理・制御するために設計されたツールまたはシステムである。パフォーマンスを向上させ、リソースを最適化するために、タスクの実行を展開、自動化、管理するプロセスを容易にする。

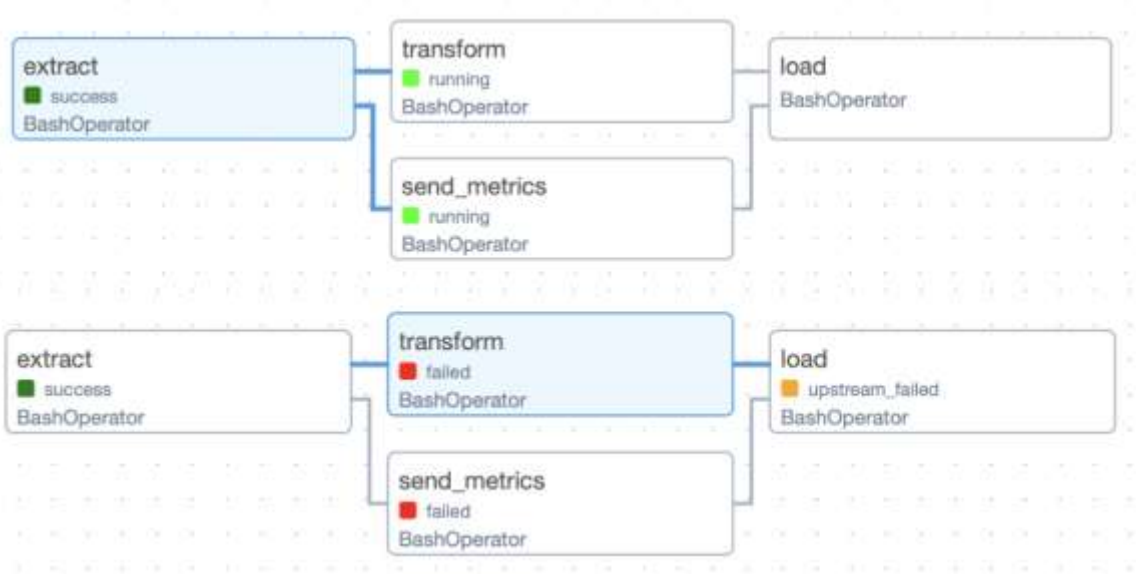


図 7.4-1 Apache Airflow は、DAG -ETL の可視化、実行ログの表示、タスクの起動ステータスなどが可能なユーザーフレンドリーなインターフェイスを提供します。

Airflow は、分散コンピューティング、データ処理、ETL （Extract, Transform, Load ）プロセス管理、タスクスケジューリング、およびその他のデータシナリオのオーケストレーションと自動化のために広く使用されています。デフォルトでは、Apache Airflow はデータベースとして SQLite を使用します。

ETL に似た単純な DAG の例は、Extract 、Transform 、Load というタスクで構成されている。このグラフはユーザーインターフェイス（図 7.4-1）で制御され、タスク（コードフラグメント）の実行順序が定義されている。すべてのタスクが完了すると、データのロード処理が成功したとみなされる。

Apache Airflow: ETL 自動化の実践的アプリケーション

Apache Airflow は、複雑なデータ処理プロセスを整理し、柔軟な ETL -コンベアを構築するために広く使用されている。Apache Airflow は、ウェブ・インターフェースまたは Python コード（図 7.4-2）を通じてプログラム的に実行することができます。ウェブインタフェース（図 7.4-3）では、管理者や開発者は DAG の視覚的な監視、タスクの実行、実行結果の分析を行うことができます。

DAG を使用すると、タスクの明確な順序を設定し、タスク間の依存関係を管理し、ソース データの変更に自動的に対応することができます。レポート処理の自動化に Airflow を使用する例を考えてみましょう（図 7.4-2）。

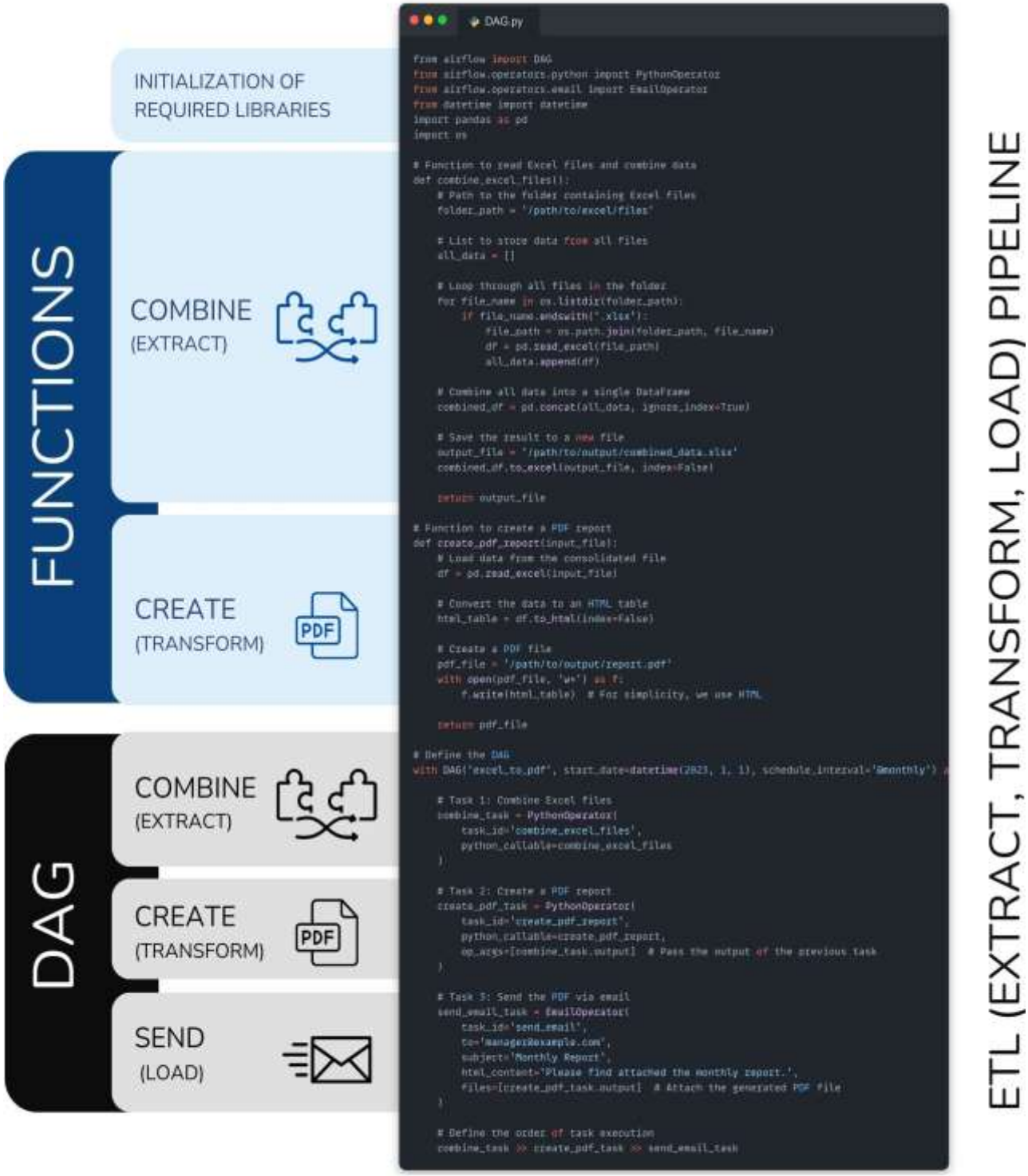


図 7.4-2 ETL のコンセプト -Apache Airflow を使ったデータ処理のためのコンベア.

この例（図 7.4-2）では、ETL -コンベア内で主要なタスクを実行する DAG について考える：

- Read Excel -files (Extract)：
  - 指定されたディレクトリ内のすべてのファイルを逐次走査。
  - pandas ライブラリを使用して各ファイルからデータを読み込む。
  - すべてのデータを単一の DataFrame にマージ。
- PDF -ドキュメントを作成（Transform ）：
  - マージされた DataFrame を HTML -table に変換。
  - 表を PDF として保存（デモ版では HTML 経由）。
- レポートを電子メールで送信する (Load)：
  - EmailOperator を適用して、PDF -document を電子メールで送信します。
- DAG の設定 ：
  - データの抽出→ レポートの作成→ 送信。
  - 起動スケジュールの割り当て（@monthly - 毎月 1 日）。

自動 ETL -例（図 7.4-2）は、Excel -ファイルからデータを収集し、PDF -文書を作成し、それを電子メールで送信する方法を示している。これは、Airflow で可能な多くの使用例の 1 つに過ぎません。この例は、データ処理を簡素化し自動化するために、どのような特定のタスクにも適合させることができます。

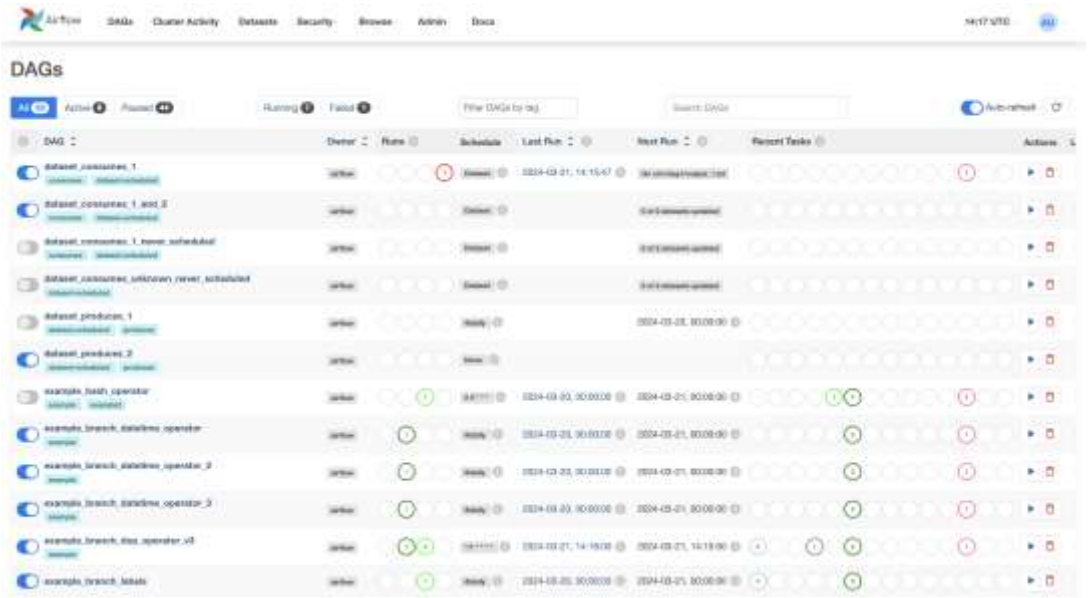


図 7.4-3 環境内のすべての DAG の概要 最近の実行に関する情報。

Apache Airflow Web インターフェース（図 7.4-3）は、データ・ワークフローを管理するための包括的なビジュアル環境を提供します。ノードはタスク、エッジはタスク間の依存関係を表すインタラクティブなグラフとして DAG を表示し、複雑なデータワークフローの追跡を容易にします。インターフェースには、タスクの実行状況、実行履歴、詳細なログ、パフォーマンスメトリクスの情報を含むダッシュボードが含まれます。管理者は、直感的なユーザーインターフェースを通じて、タスクの手動起動、失敗した操作の再開、DAG の一時停止、環境変数の設定を行うことができます。

このようなアーキテクチャは、データ検証、実行ステータスの通知、外部 API またはデータベースとの統合によって補完することができます。Airflow では、DAG を柔軟に適応させることができます。新しいタスクの追加、順番の変更、 チェーンの結合など、複雑なデータ処理プロセスを自動化するための効果的なツールとなります。Airflow Web インターフェース（図 7.4-3、図 7.4-4）で DAG を実行すると、タスクのステータスを監視できます。システムはカラー表示を使用します：

- 緑 - タスクが正常に完了した。
- 黄色 - 処理中。
- 赤 - タスク実行中のエラー。

障害が発生した場合（ファイルがない、データ構造が壊れているなど）、システムは自動的に通知の送信を開始する。

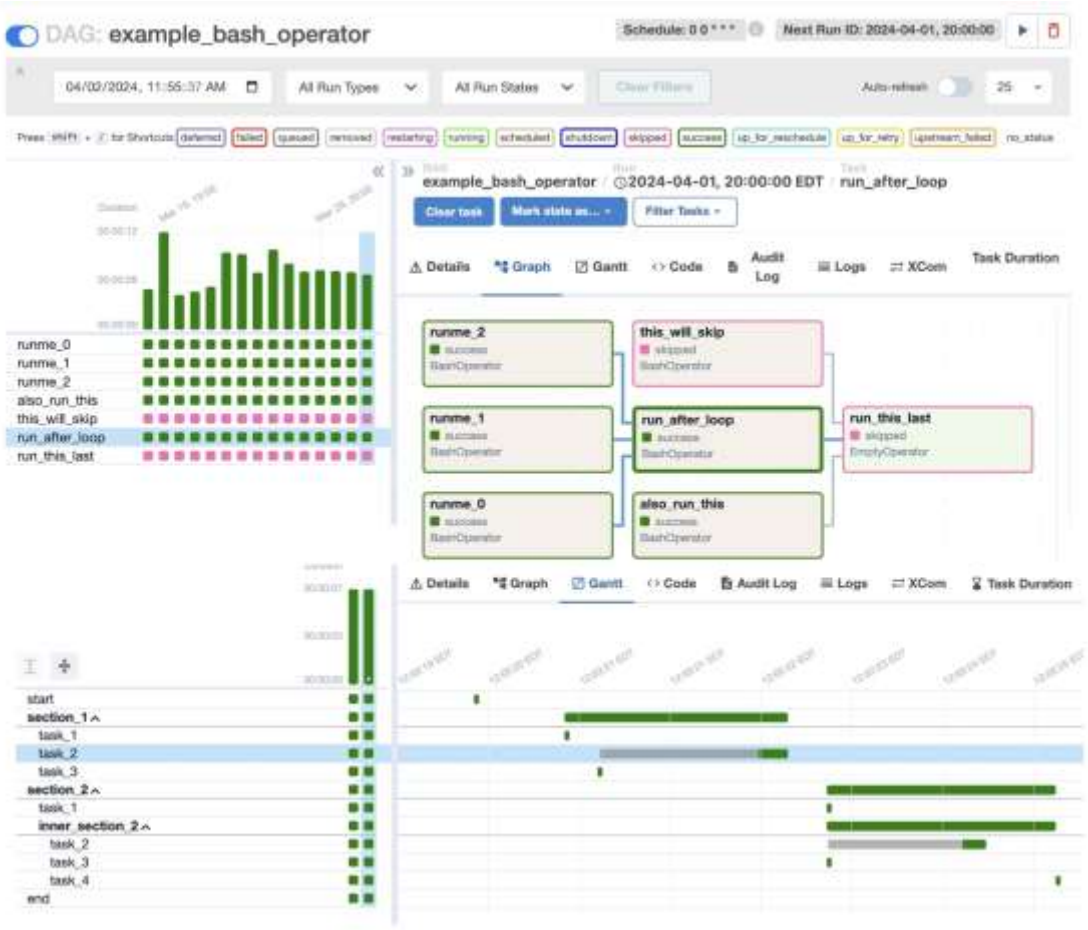


図 7.4-4 Apache Airflow は、複雑なデータ処理パイプラインの問題診断、プロセス最適化、チームコラボレーションを大幅に簡素化します。

> Apache Airflow は、ルーチンタスクを自動化するため、手動で実行する必要がなく便利です。プロセスの実行を監視し、エラーを即座に通知することで信頼性を確保します。システムの柔軟性により、新しいタスクの追加や既存のタスクの変更が容易であり、変化する要件に合わせてワークフローを適応させることができます。

Apache Airflow に加えて、ワークフローをオーケストレーショ ンするための同様のツールがある。例えば、オープンソースで無償の Prefect（図 7.3-5）は、よりシンプルな構文を提供し、Python とよりよく統合されます。Spotify によって開発された Luigi は、同様の機能を提供し、ビッグデータとうまく連動します。また、Kronos と Dagster も注目に値する。これらはモジュール性とスケーラビリティに重点を置き、Pipeline を構築するための現代的なアプローチを提供している。タスク・オーケストレーション・ツールの選択はプロジェクトの特定のニーズに依存するが、いずれも複雑な ETL データ・プロセスの自動化に役立つ

特に注目すべきは Apache NiFi である。オープンソースプラットフォーム で、データのストリーミングとルーティングのために設計されている。バッチ処理と依存関係管理に重点を置く Airflow とは異なり、NiFi はリアルタイム、オンザフライでのデータ変換とシステム間の柔軟なルーティングに重点を置いている。

ルーティングとデータ変換のための Apache NiFi

Apache NiFi は、異なるシステム間のデータフローを自動化するために設計された、強力なオープンソースプラットフォーム である。もともとは 2006 年に米国家安全保障局（NSA）が内部使用のために「Niagara Files」という名前で開発した。2014 年、このプロジェクトはオープンソース化され、Apache Software Foundation に移管され、彼らの技術移転イニシアチブの一部となった[141]。

Apache NiFi は、リアルタイムでデータを収集、処理、転送するように設計されている。バッチタスクで動作し、明確に定義されたスケジュールを必要とする Airflow とは異なり、NiFi はストリーム処理モードで動作し、異なるサービス間で継続的にデータを転送できる。

> Apache NiFi は、IoT デバイス、建設用センサー、監視システム、およびデータの変更に即座に対応する必要があるサーバー上の CAD フォーマットのストリーミング検証などとの統合に最適です。

フィルタリング、変換、ルーティングツールを内蔵した NiFi では、データを標準化（Transform ）してから、ストレージや分析システムに転送（Load）することができる。NiFi の主な利点のひとつは、内蔵のセキュリティ・サポートとアクセス・コントロールであり、機密情報を扱うための信頼性の高いソリューションとなっている。

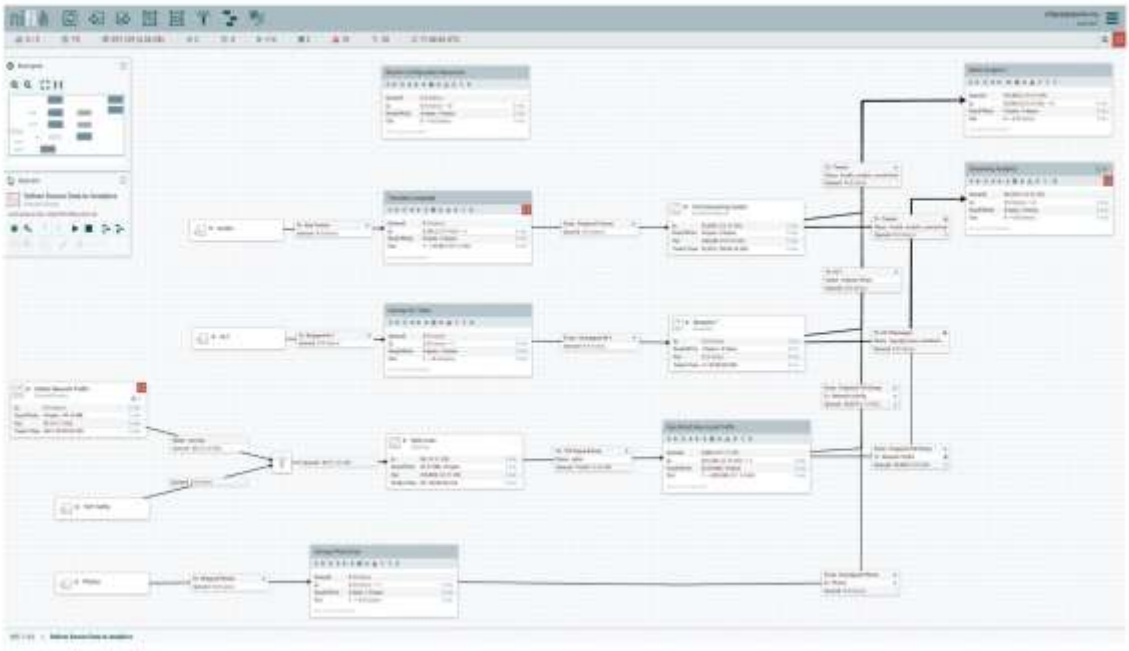


図 7.4-5 Apache NiFi インターフェースにおけるデータフローのグラフィカル表現.

Apache NiFi は、リアルタイムのデータストリーミング、フィルタリング、ルーティングタスクを効率的に処理します。システム間の安定した情報転送と高いスループットが重要な、技術集約的なシナリオに最適です。

しかし、さまざまなサービスを統合し、ルーチン業務を自動化し、深いプログラミングの知識がなくてもワークフローを迅速に設定することが主な目的である場合、参入の敷居が低く、最大限の柔軟性を備えたソリューションが求められる。そのようなツールの 1 つが n8n - ビジネスオートメーションとプロセスの視覚的オーケストレーションに焦点を当てた Low-Code /No-Code クラスのプラットフォームです。

n8n Low-Code, No-Code プロセス・オーケストレーション

n8n は、自動化されたワークフローを構築するためのオープンソース Low-Code / No-Code プラットフォームで、使いやすさ、柔軟性、幅広い外部サービスとの迅速な統合が特徴です。

No-Code は、コードを書かずにデジタル製品を作成する方法です。ロジックからインターフェイスまで、プロセスのすべての要素はビジュアルツールの助けを借りてのみ実現されます。No-Code プラットフォームは、技術的なバックグラウンドを持たないユーザーを対象としており、自動化、フォーム、統合、ウェブアプリケーションを素早く作成することができます。例：ユーザーは、プログラミングの知識がなくても、ドラッグ・アンド・ドロップのインターフェイスを使って、自動通知送信や Google Sheets の統合 。

オープンソースとローカル展開機能により、自動化と ETL パイプライン作成プロセスにおける n8n は、クラウドプロバイダーからのセキュリティと独立性を確保しながら、企業がデータを完全にコントロールできるようにする。

Apache Airflow は、厳密なオーケストレーショ ンを伴う計算タスク向けで、Python の知識を必要とするが、n8n は、プログラミング言語の知識を必要とせずにスクリプトを作成できるビジュアル・エディタを提供する（図 7.4-6）。そのインターフェイスは、コードを書かずに自動化されたプロセスを作成できる（No-Code ）が、より複雑なシナリオでは、ユーザーが独自の JavaScript や Python -関数を追加して機能を拡張できる（Low-Code ）。

ローコード（Low-Code） は、アプリケーションやプロセスの基本的なロジックをグラフィカル・インターフェースとビジュアル要素を使用して作成し、プログラムコードは機能のカスタマイズや拡張にのみ使用するソフトウェア開発へのアプローチです。ローコードプラットフォームでは、プログラマーだけでなく、基本的な技術スキルを持つビジネスユーザーも参加することで、ソリューションの開発を大幅に加速することができます。例：ユーザーは、既製のブロックからビジネス・プロセスを構築し、必要に応じて、JavaScript または Python で独自のスクリプトを追加することができます。

n8n は敷居の低いプラットフォームとして位置づけられているが、基本的なプログラミング知識、ウェブ技術の理解、API を扱うスキルが必要である。システムの柔軟性により、自動データ処理からメッセンジャー、IoT -デバイス、クラウドサービスとの統合まで、幅広いタスクに適応できる。

n8n を使用する主な特徴と利点：

- オープンソースとローカル展開オプションは、完全なデータコントロール、セキュリティコンプライ

アンス、クラウドプロバイダーからの独立性を保証します。

- CRM 、ERP 、e コマース、クラウドプラットフォーム、メッセンジャー、データベースを含む 330 以上のサービスとの統合。
- シナリオの柔軟性：単純な通知から、 -リクエストの API 処理、意思決定ロジック、AI -サービスとの接続による複雑なチェーンまで。
- JavaScript と Python のサポート: ユーザーは必要に応じてカスタムコードを埋め込むことができ、自動化機能を拡張できます。
- 直感的なビジュアル・インターフェース：すべてのプロセス・ステップを素早く設定し、視覚化することができます。

ローコード・クラスのプラットフォーム は、最小限のコードでデジタル・ソリューションを作成するためのツールを提供し、深い技術的専門知識はないがプロセスを自動化する必要があるチームに理想的である。

建設業界では、n8n を使用して、プロジェクト管理システムとの統合、ストリームチェック、すぐに使える レポートやレターの作成、資材在庫データの自動更新、チームへのタスクステータス通知の送信など、さまざまなプロセスを自動化することができます。n8n でカスタマイズされたパイプライン 、手作業を何倍にも減らし、エラーの可能性を減らし、プロジェクト実行のための意思決定をスピードアップすることができます。

> n8n.io/workflows で利用可能な、フリーでオープンソースの n8n Pipeline の約 2,000 の既成のものから選択することができ、建設ワークフローと個人的なタスクの両方を自動化し、ルーチン作業を削減することができます。

n8n.io[142]で無料で入手できる既製のパイプライン・テンプレート を使えば、Gmail で自動的に返信の草稿を作成することができ（図 7.4-6）、大量のメールを受け取ったり、返信を作成するのが難しいユーザーを助けることができる。

この n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" テンプレート（図 7.4-6）は、ChatGPT の LLM を使用して受信メールを分析し、応答が必要かどうかを判断し、ChatGPT から下書きを生成し、テキストを HTML に変換し、Gmail のメッセージチェーンに追加します。これは自動的にはメールを送信しないので、手動で編集し、応答を承認することができます。セットアップは約 10 分で完了し、Gmail API の OAuth 設定と OpenAI API の統合が含まれます。その結果、メールの内容のコントロールを失うことなく、日常的なメールコミュニケーションを自動化するための便利で無料のソリューションが実現します。

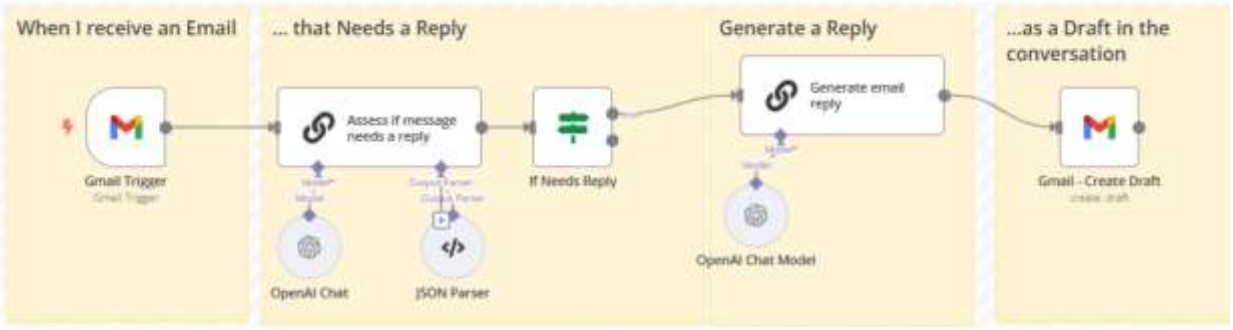


図 7.4-6 n8n を使用した電子メールの自動返信生成処理。

n8n を使った自動化のもう 1 つの例は、不動産市場でお得な情報を見つけることである [143]。N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail" は、Zillow API を使って、与えられた条件にマッチする関連するオファーを毎日収集する。主要な投資指標（キャッシュオンキャッシュ ROI、月次キャッシュフロー、頭金）を自動的に計算し、Google Sheets を更新し、サマリーレポートを電子メールで送信する（図 7.4-7）。



図 7.4-7 不動産の投資魅力を評価するための自動化されたプロセス。

その柔軟性と拡張性により、n8n は、比較的シンプルで無料のオープンソースツール を使ってデジタル変革を図り、市場での競争力を高めようとする企業にとって価値あるツールとなっている。

Apache NiFi 、Airflow、n8n などのツールは、データ処理の 3 つのレイヤーとして考えることができる（図 7.4-8）。NiFi はデータの流れを管理し、配信と変換を保証し、Airflow はデータを処理パイプラインに集約してタスク実行をオーケストレートし、n8n は外部サービスとの統合を自動化し、ビジネスロジックを管理する。




	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

図 7.4-8 Apache Airflow 、Apache NiFi 、n8n は、現代のデータ管理アーキテクチャの 3 つの補完的なレイヤーとして見ることができる。

これらの無料かつオープンソースのツールを組み合わせることで、建設業界におけるデータとプロセス管理のための効果的なエコシステムの一例を形成し、企業が意思決定とプロセスの自動化のために情報を活用できるようになる可能性がある。

次のステップ：手作業から分析ベースのソリューションへ

今日の建設企業は、資材価格の変動、納期の遅延、労働力不足、プロジェクトの納期厳守など、不確実性の高い環境で事業を展開している。分析ダッシュボード、ETL -コンベア、BI システムの活用は、企業が問題領域を迅速に特定し、リソース効率を評価し、財務的損失につながる前に変化を予測するのに役立ちます 。

この部分を要約すると、議論した技術を日常業務に応用するのに役立つ主な実践的ステップを強調する価値がある：

- データの視覚化と分析ダッシュボードの実装
  - 主要業績評価指標（KPI ）を監視するためのダッシュボード作成プロセスをマスターする。
  - データの可視化ツールを使用する（Power BI, Tableau, Matplotlib, Plotly ）。
- ETL -プロセスによるデータ処理の自動化
  - 様々なソース（ドキュメント、テーブル、CAD ）から ETL プロセスによる自動データ収集の設定

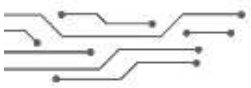
- ☐ Python スクリプトを使用したデータ変換（正規表現のチェックや計算など）の実行
- ☐ FPDF ライブラリを使用して、Excel ファイルからのデータを使用したり、他の PDF 文書から情報を抽出したりして、自動 PDF（または DOC）レポートを設定してみてください。

■ 自動化のための言語モデル（LLM ）の使用

- ☐ 大規模言語モデル（LLM）を使用して、非構造化文書からデータを抽出・分析するためのコードを生成する。
- ☐ n8n 自動化ツールに慣れ親しみ、ウェブサイトで既製のテンプレートやケーススタディを調べる。No-Code/Low-Code アプローチを使って、あなたの仕事のどのプロセスを完全に自動化できるかを判断する。

データとプロセスの自動化に対する分析的アプローチは、定型業務に費やす時間を短縮するだけでなく、意思決定の質を向上させる。ビジュアル・アナリティクス・ツール（）や ETL（）を導入する企業は、変化に迅速に対応する機会を得ることができる

n8n 、Airflow、NiFi のようなツールを使ってビジネスプロセスを自動化することは、デジタル成熟への第一歩に過ぎない。次のステップは、自動化を支えるデータそのものの質の高い保存と管理である。第 8 回では、建設会社が持続可能なデータ・ストレージ・アーキテクチャを構築し、ドキュメントやマルチフォーマット・ファイルのカオスから、一元化されたストレージと分析プラットフォームに移行する方法について詳しく見ていく。





VIII 部

建設におけるデータの保存と管理

パート 8 では、建設業界向けの最新のデータ保存・管理技術について解説する。シンプルな CSV や XLSX から、より高性能な Apache Parquet や ORC まで、大量の情報を扱うための効率的なフォーマットを分析し、その機能と限界を詳細に比較する。データウェアハウス (DWH)、データレイク)、およびそれらのハイブリッドソリューション (Data Lakehouse)の概念や、データガバナンス)、データミニマリズム)の原則についても議論する。データ沼（Data Swamp） ）の問題や、情報システムの混乱を防ぐための戦略についても詳しく取り上げている。ベクトル・データベース や、バウンディング・ボックス の概念による構築への応用など、データを扱うための新しいアプローチが紹介されている。このパートでは、データワークフローを組織化するための新しい基準として、DataOps と VectorOps の方法論にも触れている。

章 8.1.

データ・インフラ：保存形式からデジタル・レポジトリまで

データアトムズ：効果的な情報管理の基盤

宇宙に存在するすべてのものは、最小の構成要素である原子や分子から成り立ち、時間の経過とともに、生物も非生物も必然的にこの初期状態に戻る。自然界では、このプロセスは驚異的なスピードで起こるが、私たちはそれを人間がコントロールするプロセスに移そうとしている。

> 森の中では、どんな生物も最終的には栄養価の高い物質に変化し、新しい植物の基礎となる。これらの植物は、数百万年前に宇宙を創造したのと同じ原子で構成される新しい生物の餌となる。

ビジネスの世界では、複雑な多層構造を、自然界の原子や分子のように最も基本的で最小限の処理単位に分解することも重要です。これにより、データの原子が効率的に保存・管理され、アナリティクス（ ）の成長と意思決定の質を高めるための重要なリソースとなる、豊かで肥沃な基盤に変わります。



図 8.1-1 分析と意思決定は、一度処理され保存されたデータを再利用して行われる。

音楽は音符が組み合わさって複雑な楽曲を作り上げるが、言葉は文字と音という原始的な単位から生み出される。自然であれ、科学であれ、経済であれ、芸術であれ、技術であれ、世界は破壊、構造、循環、創造を追求する中で、驚くべき統一性と調和を示す。同様に、原価計算システムにおけるプロセスは、原価計算のレベル（ ）とスケジュールにおいて、小さな構造化された単位-資源項目-に分解される。これらの単位は、メモのように、より複雑な計算やスケジュールを形成するために使用される。同じ原理がコンピュータ支援設計システムにも使われており、複雑な建築やエンジニアリングのプロジェクトは、基本要素である個々の要素やライブラリコンポーネントから構築され、そこから複雑な建物や構造物のプロジェクトの完全な 3D -モデルが作成される。

> 自然や科学に内在する循環性や構造の概念は、現代のデータの世界にも反映されている。自然界ですべての生物が原子や分子に回帰するように、現代のデータ処理ツールの世界でも、情報は最も原始的な形に回帰する傾向がある。

有限不可分の最小要素は、ビジネスプロセスの基本構成要素である。様々なソースからこれらの小さな構成要素を収集し、構造化（原子に分解）し、保存する方法を最初から慎重に検討することが重要である。データの整理と保存は、単に構成要素に分解すればよいという問題ではない。データが必要なときにいつでも簡単に取り出し、分析し、意思決定に利用できるように、構造化された方法で統合し、保存することも同様に重要である。

情報を効率的に処理するためには、データ保存の形式と方法を注意深く選択しなければならない。冗長なデータや無関係なデータを排除し、情報の質と関連性を確保するために、データストアは整理されなければならない。この "情報の土壌 "がより良い構造であればあるほど、ユーザーはより速く、より正確に正しいデータを見つけ、分析的な問題を解決することができる。

情報保管：ファイルまたはデータ

データ・ウェアハウスは、企業がさまざまなシステムから情報を収集・結合し、さらなる分析のための単一のセンターを構築することを可能にする 。収集された履歴データは、プロセスのより深い分析を可能にするだけでなく、業績に影響を与える可能性のあるパターンの特定も可能にする。

ある会社が複数のプロジェクトを同時に進めているとしよう。あるエンジニアは、コンクリートがどれだけ打設され、まだどれだけ購入しなければならないかを把握したいと考えている。従来のやり方では、手作業でサ

ーバーを検索し、いくつかの見積もり表を開き、完成工事証明書と比較し、現在の在庫残高をチェックしなければならない。これには数時間から数日かかる。ETL プロセスや自動スクリプトを使っても、作業は半手動のままだ。エンジニアは、サーバー上のフォルダや特定のファイルへのパスを手動で指定しなければならない。このため、貴重な作業時間が取られ続け、自動化の全体的な効果が減少してしまう。

データ管理に切り替えると、サーバーのファイルシステムで作業する代わりに、エンジニアは情報がリアルタイムで更新される統合ストレージ構造にアクセスできるようになる。コード、SQL -クエリ、あるいは LLM -エージェント・コールのような単一のクエリで、現在の残高、実行された作業量、今後の納品に関する正確なデータを即座に提供することができる。もしデータが事前に準備され、データウェアハウスに統合されていれば、フォルダを探し回ったり、何十ものファイルを開いたり、手作業で値を比較したりする必要はない。

建設会社では長い間、PDF -文書、DWG -図面、RVT -モデル、何百、何千もの Excel -表、その他の異種フォーマットを使用していた。これらのファイルは、会社のサーバー上の特定のフォルダーに保存されるため、情報の検索、確認、分析が困難であった。その結果、プロジェクト終了後に残されたファイルは、ほとんどの場合、サーバーに戻され、アーカイブ用の保存フォルダに移されるが、このフォルダは実質的に将来使用されることはない。このような従来のファイルベースのデータ保管は、ヒューマンエラーに対する脆弱性から、データフローが増加するにつれて関連性を失っていく。

> ファイルとは、データが保存されている隔離された入れ物に過ぎない。ファイルは人間のために作られるものであり、システムのために作られるものではない。例としては、エクセルの -表、PDF の -文書、CAD の -図面などがあり、目的の情報にアクセスするには、特定のツールで特別に開く必要がある。構造化された検索と処理がなければ、その中の情報は使われないままである。
>
> データとは、機械が読み取り可能な情報であり、自動的にリンクされ、更新され、分析される。単一のデータウェアハウス（データベース、DWH またはデータレイク など）では、情報はテーブル、レコード、リレーションシップの形で表現される。これにより、統一されたストレージ、自動化されたクエリー、価値分析、リアルタイムのレポーティングが可能になる。

ファイルの代わりにデータを使う（図 8.1-1）ことで、手作業による検索プロセスがなくなり、処理が一元化される。このようなアプローチをすでに導入している企業は、情報へのアクセス速度が速く、ビジネス・プロセスに迅速に統合できるため、競争上の優位性を獲得している。

> ファイルからデータへの移行は、建設業界の将来を形作る必然的な変化である。

建設業界のすべての企業が、重要な選択に直面することになる。それは、特殊なプログラムを使用して人間が読み取らなければならない、バラバラのファイルやサイロに情報を保存し続けるか、あるいは、処理の最初の段階で情報を構造化データに変換し、自動化されたプロジェクト管理のための単一の統合デジタル基盤を作成するか、ということだ。

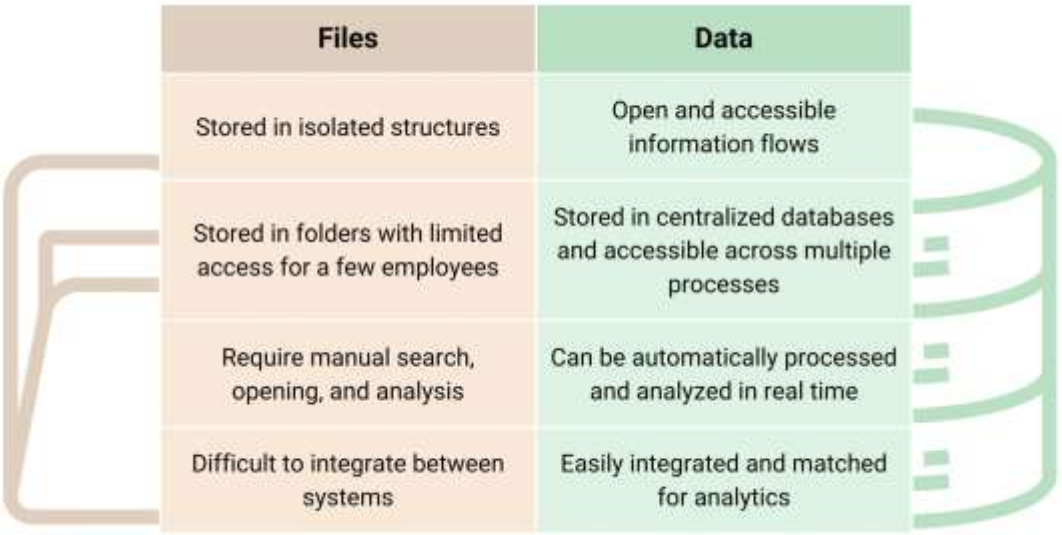


図 8.1-1 情報の流れの進化：孤立したファイルから統合されたデータへ。

情報の爆発的な増加に伴い、従来のファイルの保存・処理方法は、ますます効率的ではなくなってきています。他の分野と同様、建設業界においても、ファイル形式が異なるバラバラのファイルフォルダーや、接続されていないデータベースに頼るのではもはや十分ではありません。

デジタル時代に競争力を維持しようとする企業は、必然的に統合デジタル・プラットフォームへと移行し、ビッグデータ技術や自動分析システムを活用することになる。

ファイルベースのストレージからデータ主導のワークフローに移行するには、情報管理アプローチを再考し、集中型リポジトリへのさらなる統合に適したフォーマットを意識的に選択する必要がある。この選択によって、データをいかに効率的に処理し、いかに迅速にアクセスし、いかに簡単に企業のデジタル・プロセスに統合できるかが決まる。

ビッグデータ・ストレージ：一般的なフォーマットとその有効性の分析

ストレージフォーマットは、分析インフラのスケーラビリティ、信頼性、パフォーマンスにおいて重要な役割

を果たします。データの分析や処理（フィルタリング、グループ化、集約など）を行うために、我々の例では Pandas DataFrame を使用しています。

しかし、Pandas DataFrame は、独自の保存形式を持っていません。そのため、処理が完了すると、データは、外部形式（多くの場合、CSV または XLSX ）のいずれかにエクスポートされます。これらの表形式は、交換が簡単で、ほとんどの外部システムと互換性がありますが、 、保存効率が低い、圧縮ができない、バージョン管理のサポートが不十分であるなどの制約があります：

- CSV (Comma-Separated Values): 様々なプラットフォームやツールで広くサポートされているシンプルなテキストフォーマット。使いやすいが、複雑なデータ型や圧縮には対応していない。
- XLSX (Excel Open XML Spreadsheet): 数式、チャート、スタイリングなどの高度な機能をサポートする Microsoft Excel のファイルフォーマット。手作業でのデータ分析や視覚化には便利だが、大規模なデータ処理には最適化されていない。

表形式の XLSX（ ）と CSV（ ）のほかにも、構造化データを効率的に保存するための一般的なフォーマットがいくつかあります（図 8.1-2）：

- Apache Parquet: データ分析システムでの使用に最適化されたカラム型データストレージファイルフォーマット。効率的なデータ圧縮とエンコード方式を提供し、複雑なデータ構造やビッグデータ処理に最適。
- Apache ORC (Optimised Row Columnar): Parquet と同様、ORC は高圧縮と効率的なデータストレージを提供する。重い読み取り処理に最適化されており、データレイクの保存に適している。
- JSON (JavaScript Object Notation): JSON は、Parquet や ORC のようなバイナリ・フォーマットと比べると、データの保存という点では効率的とは言えないが、非常にアクセスしやすく、扱いやすいため、可読性とウェブ互換性が重要なスクリプトには理想的である。
- Feather: 高速、軽量、使いやすい分析指向のバイナリーカラム型データ保存フォーマット。Python (Pandas) と R の間で効率的にデータを転送するように設計されており、これらのプログラミング環境を含むプロジェクトに最適です。
- HDF5 (Hierarchical Data Format version 5): 大量のデータの保存と整理のために設計された。幅広いデータ型をサポートし、複雑なデータの集合を扱うのに適している。HDF5 は、大規模なデータセットを効率的に保存し、アクセスすることができるため、科学計算において特に人気があります。



	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)

図 8.1-2 データフォーマットの比較は、保存および処理面における主な違いを示している。

ETL プロセスの Load 段階で使用されるフォーマットの比較分析を行うため、ファイル・サイズと読 み取り時間を示す表が作成された（図 8.1-3）。表には、10,000 行とランダムな値で埋め尽くされた 10 列が含まれている。

保存形式は、CSV、Parquet、XLSX、HDF5、および ZIP アーカイブの圧縮版である。生データは NumPy ライブラリを使用して生成され、Pandas DataFrame 構造として表現された。テストプロセスは以下のステップで構成される：

- ファイル保存：データフレームは 4 つの異なるフォーマットで保存される： CSV, Parquet, XLSX, HDF5 。各フォーマットは、データの保存方法に独自の特徴があり、ファイル・サイズと読み込み速度に影響する。
- ZIP ファイル圧縮：標準的な圧縮の効果を分析するため、各ファイルをさらに ZIP アーカイブに圧縮した。
- ファイルの読み込み（ETL - Load）：ZIP から解凍した後の各ファイルの読み込み時間を測定。これに

より、アーカイブからの取り出し後のデータアクセス速度を推定することができる。

Pandas DataFrame は、それ自体がストレージフォーマットではないため、サイズや読み取り時間の分析には直接使用されていないことに注意することが重要である。これは、 、データを生成し、異なるフォーマットに保存するための中間構造としてのみ機能します。

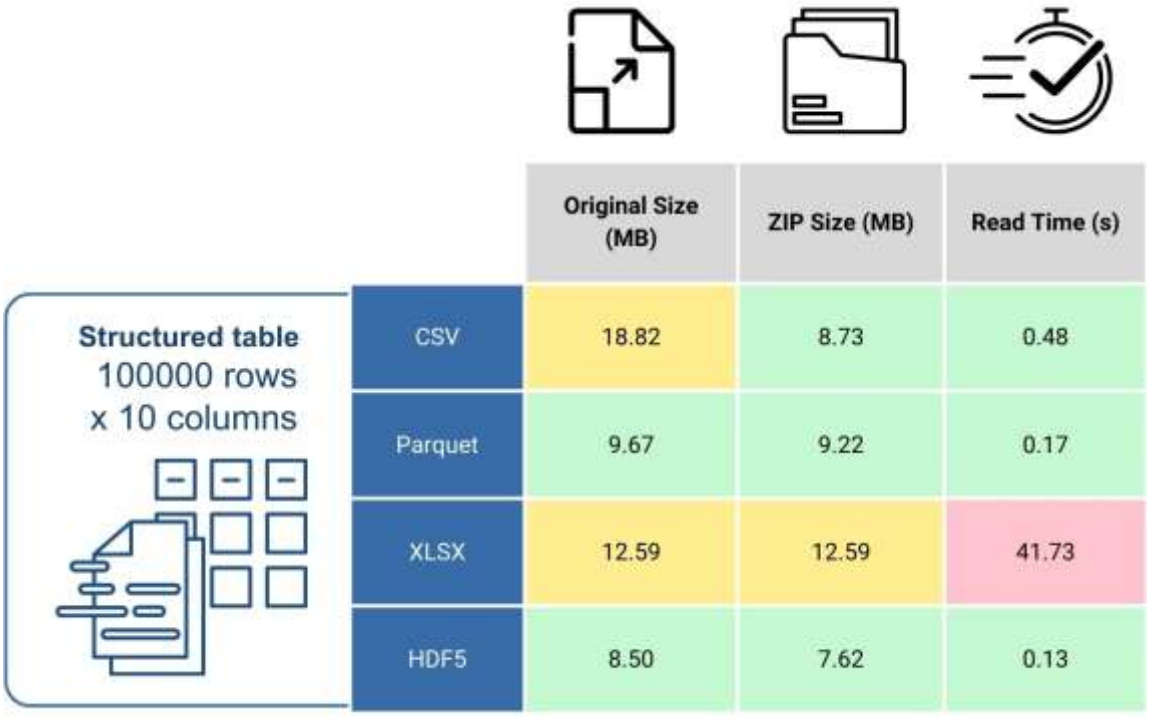


図 8.1-3 サイズと読み取り速度によるストレージフォーマットの比較。

CSV と HDF5 ファイルは高い圧縮効率を示し（図 8.1-3）、ZIP で圧縮するとサイズが大幅に小さくなる。一方、XLSX ファイルは事実上圧縮不可能であり、ZIP でのサイズはオリジナルと同程度のままであるため、大容量のデータやデータアクセスの速度が重要な環境での使用には不向きである。さらに、XLSX の読み取り時間は他のフォーマットと比較して著しく長く、高速なデータ読み取り操作には不向きです。Apache Parquet は、そのカラム構造により、分析タスクや大容量データに対して高いパフォーマンスを発揮します。

Apache Parquet でストレージを最適化

ビッグデータの保存と処理によく使われるフォーマットのひとつに、Apache Parquet がある。このフォーマットはカラム型ストレージ（Pandas に類似）専用に設計されており、メモリフットプリントを大幅に削減し

、分析クエリの速度を向上させることができます。CSV や XLSX といった従来のフォーマットとは異なり、Parquet はネイティブ圧縮をサポートし、Spark、Hadoop、クラウドストレージなどのビッグデータシステムに最適化されています。

> Parquet の主な機能には、データ圧縮とエンコーディングのサポートが含まれる。これは、ストレージサイズを大幅に削減し、データのすべての行ではなく、必要な列を直接操作することで、データの読み取り操作を高速化する。

データを Apache Parquet に変換するのに必要なコードを得るのがいかに簡単かを示す例証として、LLM を使ってみよう。

- LLM チャット（CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ）にテキストリクエストを送る：

> Pandas DataFrame から Apache Parquet にデータを保存するコードを書く。 ⏎

LLM の回答 ：

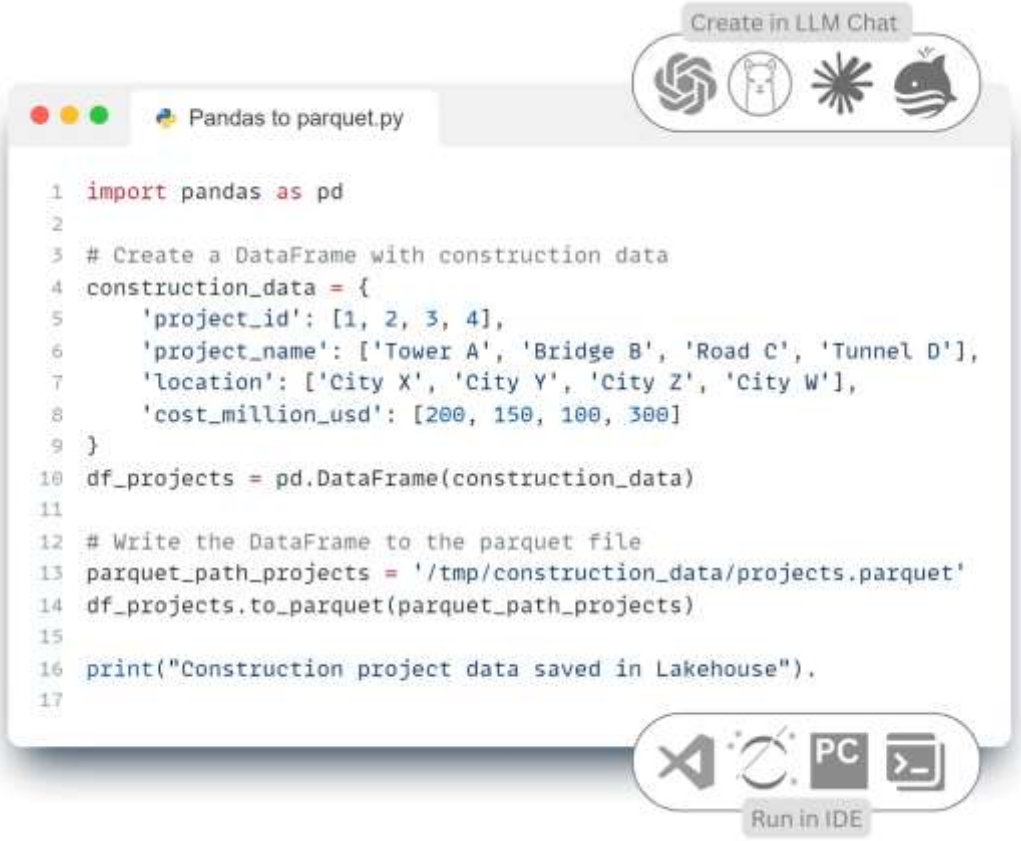


図 8.1-4 Dataframe データを RAM からストレージ効率の高い Apache Parquet フォーマットへ転送する Python を数行使って.

次の例: Parquet フォーマットで保存されたデータを使って、ETL プロセス をシミュレートし、属性のひとつである「cost_million_usd」の特定の値でプロジェクトをフィルタリングしてみましょう（図 8.1-4）。

チャットを続けながら、LLM にテキストリクエストを送る：

> テーブルのデータをフィルタリングし、Apache Parquet データ から、コスト（パラメータ cost_million_usd）が 1 億 5,000 万ドルを超えるプロジェクト（テーブルの行）のみを保存するコードを記述する。 ⏎

LLM の回答 ：

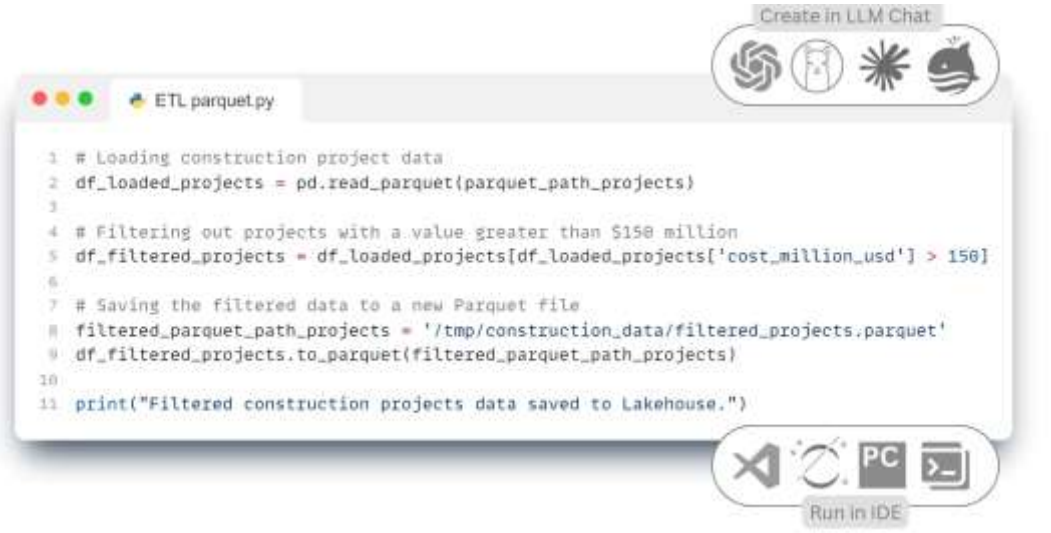


図 8.1-5 Apache Parquet フォーマットのデータを扱う場合の ETL プロセス は、他の構造化フォーマットと同じように見える。

Parquet フォーマット（XLSX、CSV などとの関連）を使用すると、、保存される情報量が大幅に削減され、検索操作が高速化されます。そのため、データの保存と分析の両方に優れています。Parquet は様々な処理システムと統合され、ハイブリッド・アーキテクチャでの効率的なアクセスを提供します。

しかし、効率的なストレージ・フォーマットは、完全なデータ体験の一要素に過ぎない。持続可能でスケーラブルな環境を構築するには、よく設計されたデータ管理アーキテクチャが必要である。DWH（データウェアハウス）クラスのシステムはこの機能を果たす。異種ソースからのデータの集約、ビジネスプロセスの透明性、BI ツールや機械学習アルゴリズムを使った複雑な分析の可能性を提供する。

DWH: データウェアハウス データウェアハウス

Parquet フォーマットが大量の情報を効率的に保存するために最適化されているように、Data Warehouse はデータを統合し構造化するために最適化されており、 アナリティクス 、予測、経営上の意思決定をサポートする。

今日の企業では、ERP、CAFM、CPM、CRM システム、会計、倉庫管理、建物のデジタル CAD モデル、IoT センサー、その他のソリューションなど、多くの異種ソースからデータがもたらされている。全体像を把握するためには、単にデータを収集するだけでは不十分で、単一のリポジトリで整理、標準化、一元管理する必要が

あります。これこそが DWH の役割であり、様々なソースからの情報を集約し、構造化し、アナリティクスや戦略的管理に利用できるようにする集中ストレージ・システムなのです。

> DWH (Data Warehouse) は、複数のソースからの情報を集約し、構造化し、分析 やレポーティングに利用できるようにする集中型データウェアハウスシステムである。

多くの企業では、データがさまざまなシステムに散在していることは、本書の最初の部分で説明した通りである（図 1.2-4）。DWH はこれらのソースを統合し、情報の完全な透明性と信頼性を確保する。DWH データウェアハウスは、複数のソースからデータを収集し、処理し、保存する特殊なデータベース（大規 模データベース）である。DWH の主な特徴は以下の通りである：

- ETL -プロセス（Extract, Transform, Load ）の使用 - ソースからデータを抽出し、クリーニングし、変換し、リポジトリにロードし、本書のパート 7 で説明したこれらのプロセスを自動化する。
- データの粒度 - DWH のデータは、集計形式（サマリーレポート）でも粒度形式（生データ）でも保存できる。2024 年以降、粒状データについて話し始めたのは CAD- ベンダーであり[125]、これはおそらく、業界がデジタル建築モデルデータを扱うための特別なクラウドストレージへの移行を準備していることを示している。
- 分析のサポート と予測 - データウェアハウス は、BI ツール、ビッグデータ - 分析、機械学習 の基盤となる。

DWH は、ビジネス・インテリジェンス の基盤として機能し、主要業績指標の分析、売上、仕入、コスト の予測、自動化されたレポーティングとデータの視覚化を可能にする（図 8.1-6）。

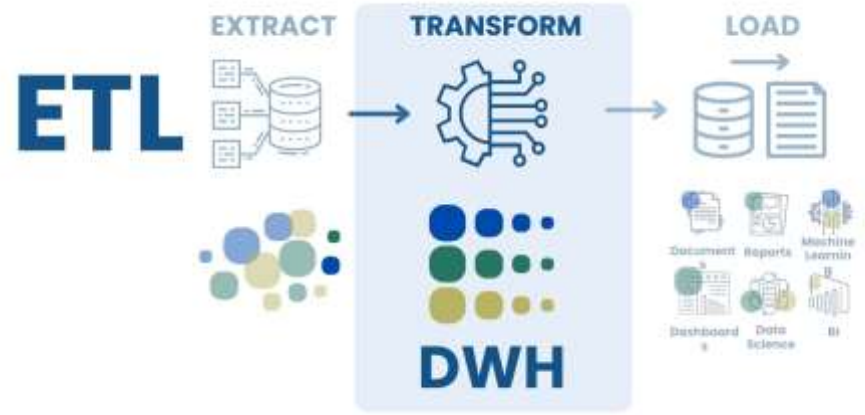


図 8.1-6 ETL -プロセスにおいて、DWH、様々なシステムから抽出されたデータが変換とオフロードの段階を経 る中央リポジトリとして機能することができる。

DWH は、情報の統合、クレンジング、構造化において重要な役割を果たし、ビジネスインテリジェンス や意思決定プロセスのための強固な基盤を提供する。しかし、データ量が急速に増加し、データソースがますます多様化している今日の環境では、情報ストレージに対する従来の DWH アプローチは、ELT やデータレイク のような形での拡張が必要になることが多い

データレイク - ETL の進化 から ELT: 従来のクリーニングから柔軟な処理へ

古典的な DWH - データウェアハウス は、構造化データを分析クエリに最適化された形式で保存するように設計されているが、非構造化データの取り扱いとスケーラビリティに限界があった。このような課題に対応するため、大量の異種データを柔軟に保存できるデータレイク が登場した。

> データレイク は、構造化されていないデータ、半構造化されたデータ、生のデータを、事前の厳密なスキーマなしに扱うことができる、代替の DWH - アプローチを提供する。この保存方法は、リアルタイムのデータ処理、機械学習（）、高度な分析（）に関連することが多い。データをロードする前に構造化・集約する DWH とは異なり、データレイクは情報を生のまま保存できるため、柔軟性と拡張性がある

従来のデータウェアハウス（RDBMS、DWH）への不満と「ビッグデータ」への関心が、データレイクの出現につながった。データレイクでは、複雑な ETL（ ）の代わりに、データを単純に緩く構造化されたリポジトリ

にアップロードし、分析段階で処理を行う：

- 従来のデータウェアハウスでは、データは通常、ウェアハウスにロードされる前に、前処理、変換、クリーニング（ETL - Extract, Transform, Load ）される（図 8.1-6）。つまり、データは構造化され、将来の特定のアナリティクス（ ）やレポーティングタスクのために最適化されます。重視されるのは、高いクエリ性能とデータの完全性を維持することです。しかし、このアプローチはコストがかかり、新しいデータタイプや急速に変化するデータスキーマを統合する柔軟性に欠ける可能性があります。
- 一方、データレイクは大量の生データをそのままの形式で保存するように設計されている（図 8.1-7）。ETL（Extract, Transform, Load ）、 のプロセスは、ELT （Extract, Load, Transform ）に取って代わられつつあり、データはまず「そのまま」ウェアハウスにロードされ、その後初めて必要に応じて変換・分析される。これにより、柔軟性が高まり、テキスト、画像、ログなどの非構造化データを含む異種データを保存できるようになる。

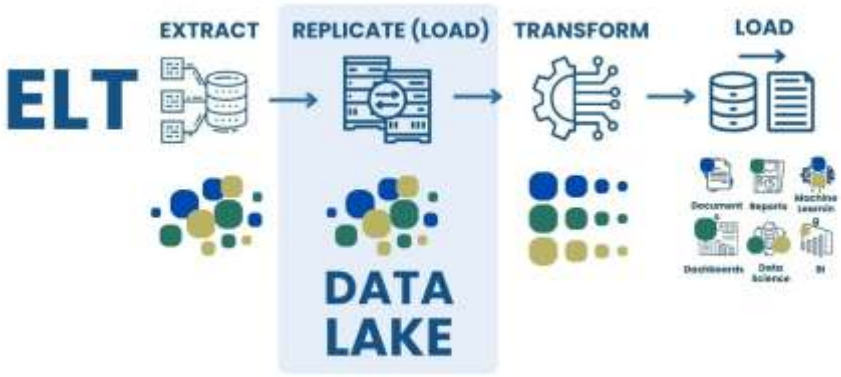


図 8.1-7 ETL とは異なり、データレイク では ELT を使用する。この場合、情報はまず「生」の形でアップロードされ、アップロードの段階で変換が行われる。

従来のデータウェアハウス（ ）は、高いクエリ性能を確保するためにデータの前処理に重点を置いているのに対し、データレイクは柔軟性を優先している。つまり、生のデータを保存し、必要に応じて変換する（図 8.1-8 ）。



図 8.1-8 現代のストレージ・コンセプトは、意思決定のためにあらゆる種類のデータを保存し、処理することを目指している。

しかし、あらゆる利点があるにもかかわらず、データレイクに欠点がないわけではない。厳密な構造がなく、情報管理が複雑なため、データが重複したり、互いに矛盾したり、無関係になったりするカオスにつながる可能性がある。さらに、このようなリポジトリでデータを検索・分析するには、特に異種情報を扱う場合にはかなりの労力を要する。このような制約を克服し、従来のデータウェアハウスとデータレイクの長所を組み合わせるために、データレイクハウス・アーキテクチャが開発された 。

データレイクハウス・アーキテクチャ: ウェアハウスとデータレイクの相乗効果

DWH （構造化、管理可能、高性能アナリティクス ）とデータレイク （スケーラビリティ、異種データの取り扱い）の最良の特徴を組み合わせるため、データレイクハウス・アプローチが開発された 。このアーキテクチャは、データレイクの柔軟性と、従来のウェアハウスに典型的な強力な処理・管理ツールを組み合わせたもので、ストレージ、アナリティクス、機械学習のバランスをうまくとっている。データレイクハウスは、データ

レイクとデータウェアハウス（ ）の統合であり、前者の柔軟性と拡張性に、後者の管理性とクエリの最適化を組み合わせたものである。

> データレイクハウス は、データレイクの柔軟性や拡張性と、データウェアハウスの管理性やクエリ性能を両立させようとするアーキテクチャアプローチである（図 8.1-9）。

Data Lakehouse の主な特徴は以下の通り：

- オープンデータストレージフォーマット：Apache Parquet のようなオープンフォーマットをデータストレージに使用することで、効率性と最適化されたクエリーを提供します。
- 読み取り専用スキーマ: DWH の書き込み専用スキーマという従来のアプローチとは対照的に、Lakehouse は読み取り専用スキーマをサポートしており、データ構造をより柔軟に管理することができます。
- 柔軟性と拡張性：構造化データおよび非構造化データの保存と分析をサポートし、ストレージレベルの最適化によって高いクエリ性能を提供します。

Data Lakehouse は、両アプローチの利点を組み合わせた妥協のないソリューションを提供し、データ処理に柔軟性を必要とする最新のアナリティクス・ワークロードに理想的です。

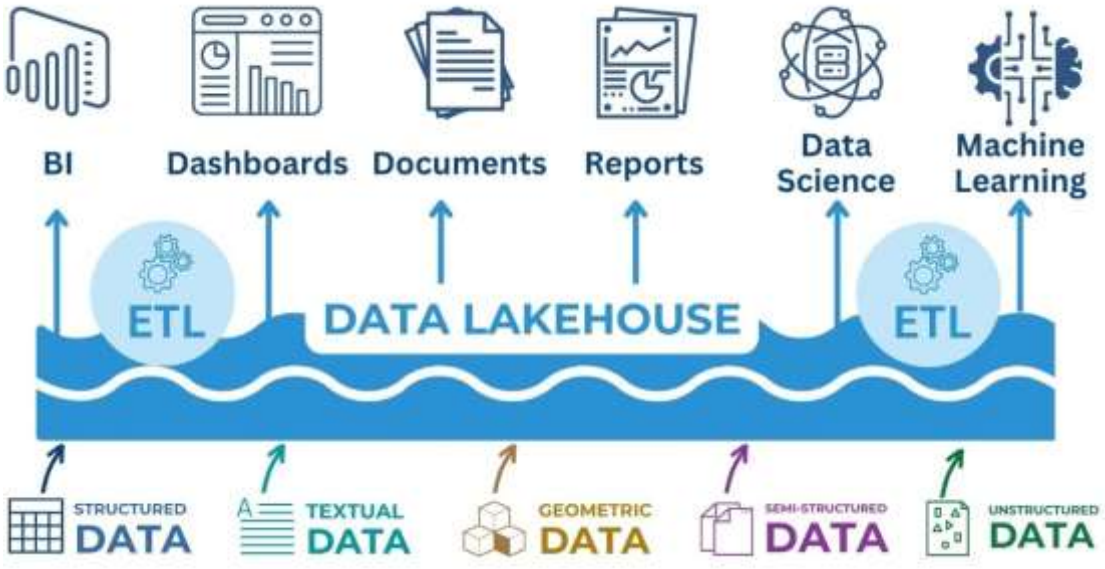


図 8.1-9 Data Lakehouse は、複雑で変化し続ける要件を満たすために設計された、次世代のストレージシステムです。

最新のデータウェアハウスの背景にある考え方は単純に見える。すべてのデータが一箇所にあれば、分析が容易になるからだ。しかし、実際にはすべてがそれほどスムーズではない。ある企業が、通常の会計システムや管理システム（ERP 、PMIS 、CAFM など）を完全に破棄し、誰もがアクセスできる巨大なデータレイクに置き換えることにしたとしよう。何が起こるだろうか？データは重複し、矛盾し、重要な情報が失われたり破損したりするだろう。データレイクがアナリティクス（ ）のためだけに使われるとしても、適切な管理がなければ、データレイクは大きく損なわれる：

- データを理解するのは難しい。従来のシステムでは、データは明確な構造を持っているが、レイクではファイルやテーブルの膨大な集積にすぎない。何かを見つけるためには、専門それぞれの行や列が何を担当しているのかを把握しなければならない。
- データが不正確である可能性がある：同じ情報の多くのバージョンが一箇所に保存されている場合、どのバージョンが最新かを知ることは難しい。その結果、古いデータや誤ったデータに基づいて意思決定が行われることになる。
- 仕事のためにデータを準備するのは難しい。データを保存するだけでなく、レポート、グラフ、表などの便利な形で提示しなければならない。従来のシステムでは、これは自動的に行われるが、データレイクでは追加の処理が必要だ。

その結果、データ・ウェアハウスのコンセプトはそれぞれ独自の特徴、処理アプローチ、ビジネス・アプリケーションを持っている。従来のデータベースはトランザクション処理に重点を置き、データウェアハウス （DWH）はアナリティクスのための構造 を提供し、データレイク（Data Lake）は情報を生の形で保存し、ハイブリッドウェアハウス（Data Lakehouse ）は DWH とデータレイクの利点を兼ね備えている（図 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

図 8.1-10 DWH 、データレイク 、データレイクハウス ：データの種類、利用シーン、処理方法
、保管方法の主な違い。

ストレージアーキテクチャの選択は、ビジネスニーズ、情報量、アナリティクスの要件によって複雑なプロセスとなる。それぞれのソリューションには長所と短所がある。DWH は構造を提供し、データレイク は柔軟性を提供し、レイクハウスはその 2 つのバランスを提供する。組織が単一のデータアーキテクチャに限定されることはほとんどありません。

> どのようなアーキテクチャを選択したかにかかわらず、自動化されたデータ管理システムは、手作業による方法に比べて格段に優れている。ヒューマンエラーを最小限に抑え、情報処理をスピードアップし、ビジネス・プロセスのあらゆる段階でデータの透明性とトレーサビリティを確保する。

一元化されたデータウェアハウスは、すでに経済の多くの分野で業界標準となっているが、建設業界では断片的な状況が続いている。 、データはさまざまなプラットフォーム（CDE、PMIS、ERP など）に分散しているため、現状を統一的に把握することが難しく、これらのソースを総合的かつ分析的に利用可能なデジタル環境に統合できるアーキテクチャが必要となる。

CDE 、PMIS 、ERP または DWH およびデータレイク

建設会社やエンジニアリング会社の中には、ISO 19650 に従った共通データ環境（CDE ）の概念を既に使用しているところもある。要するに、CDE は、他の産業におけるデータウェアハウス（DWH ）と同じ機能を果たすもので、情報の一元化、バージョン管理の提供、有効な情報へのアクセスの提供などである。

> 共通データ環境（CDE ）は、施設のライフサイクルの全段階を通じて、プロジェクト情報の管理、保存、共有、コラボレーションに使用される一元化されたデジタル空間である。CDE は多くの場合、クラウドベースの技術を用いて実装され、CAD (BIM)システムと統合されている。

金融、小売、ロジスティクス、工業の各分野では、数十年前から集中型データ管理システムを利用し、さまざまな情報源からの情報を組み合わせ、関連性を管理し、分析を行ってきた。CDE は、これらの原則を建築設計とライフサイクル管理の課題に適応させることで、さらに進化させています。

DWH のように、CDE はデータを構造化し、変更をキャプチャし、検証された情報への単一アクセスポイントを提供する。クラウドへの移行や分析ツールとの統合により、両者の違いはますます明確ではなくなりつつあ

る。CDE に、2023 年以来 CAD -ベンダーの間で議論されてきたコンセプトである粒状データ[93, 125] を加えると、古典的な DWH との類似点がさらに見えてくる。

建設 ERP と PMIS システム」の章では、すでに PMIS（プロジェクト管理情報システム）と ERP（統合基幹業務システム）について見てきました。建設プロジェクトでは、CDE（ ）と PMIS が連携しています。CDE は、図面、モデル、プロジェクト文書などのデータのリポジトリとして機能し、PMIS は、納期、タスク、リソース、予算の管理などのプロセスを管理します。

ERP は、ビジネス全体（財務、調達、人事、生産）の管理を担当し、PMIS と統合することで、企業レベルでのコストと予算の管理を行うことができる。アナリティクス とレポーティングのために、DWH を使用して、CDE 、PMIS、ERP からデータを収集、構造化、集約し、財務 KPI （ROI ）を評価し、パターンを特定することができる。一方、データレイク （DL）は、生データや非構造化データ（ログ、センサーデータ、画像など）を保存することで、DWH を補完することができる。このデータは処理され、さらなる分析のために DWH にロードすることができる。

> したがって、CDE と PMIS はプロジェクト管理に、ERP はビジネスプロセスに、DWH とデータレイク はアナリティクスとデータ運用に焦点を当てている。

CDE 、PMIS 、ERP システムと DWH 、データレイク を比較すると、ベンダーの独立性、コス ト、統合の柔軟性、データの独立性、変化への適応スピード、分析能力などの点で 大きな違いが見られる（図 8.1-11）。CDE 、PMIS、ERP のような従来のシステムは、特定のベンダーのソリューションや標準に縛られることが多く、柔軟性に欠け、ライセンスやサポートのためにコストがかさむ。さらに、このようなシステムのデータは、しばしば独自のクローズド・フォーマットでカプセル化されているため、その利用や分析が制限されている。

	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)

図 8.1-11 DWH とデータレイク は、CDE 、PMIS 、ERP のようなシステムよりも高い柔軟性とデータの独立性を提供する。

対照的に、DWH と Data Lake は、異なるデータソースとの統合においてより高い柔軟性を提供し、オープンなテクノロジーとプラットフォームの使用は、総所有コストの削減に役立つ。さらに、DWH とデータレイクは幅広い分析ツールをサポートしており、分析・管理能力を高めている。

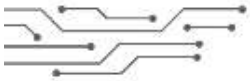
> CAD フォーマットのリバースエンジニアリングツールが開発され、CAD アプリケーション・データベースにアクセスできるようになるにつれ、疑問はますます深まっている。何十もの請負業者や設計組織で働く幅広い専門家が設計データを利用できるようにしなければならないのに、閉鎖的で孤立したプラットフォームを使い続けることが、どれほど正当化されるのだろうか？

このようなベンダー固有のテクノロジー依存は、データ管理の柔軟性を著しく制限し、プロジェクトの変更への対応を遅らせ、参加者間の効果的なコラボレーションを阻害する可能性がある。

DWH 、データレイク 、CDE 、PMIS など、データ管理に対する従来のアプローチは、情報の保存、構造化、処

理に主眼を置いてきた。しかし、人工知能（）や機械学習（）の発達に伴い、データを集約するだけでなく、複雑な関係を明らかにし、隠れたパターンを見つけ、最も関連性の高い情報に即座にアクセスできるような、新しいデータ整理方法の必要性が高まっている。

ベクトル・データベース - 高次元埋め込みに最適化された新しいタイプのストレージ - は、この方向で特別な役割を果たし始めている。



章 8.2.

データウェアハウスの管理と混乱防止

ベクターデータベースとバウンディングボックス

ベクターデータベースは、単にデータを保存するだけでなく、意味による検索、意味的近接性によるオブジェクトの比較、レコメンデーションから自動分析、コンテキストの生成まで、インテリジェントなシステムの構築を可能にする新しいクラスのリポジトリである。完全一致を重視する従来のデータベースとは異なり、ベクトルデータベースは属性に基づいて類似オブジェクトを検索。

> ベクトル・データベース は、データを多次元ベクトルとして格納する特殊なタイプのデータベースで、各ベクトルは特定の特性や品質を表す。これらのベクトルは、データの複雑さによって次元数が異なります（数次元の場合もあれば、数千次元の場合もあります）。

ベクトルデータベースの主な利点は、値の完全なマッチングではなく、意味的な関連性による検索です。SQL や Pandas の "equals "や "contains "フィルタによるクエリの代わりに、特徴空間における最近傍探索(k-NN)（k-NN については次のパートで詳しく説明する）が使用される。

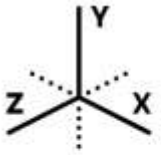
LLM (Large Language Models)と生成モデルの開発により、データベースとのインタラクションが変わり始めている。自然言語でデータを照会し、文書を意味的に検索し、キーワードを自動的に抽出し、オブジェクト間の文脈的関係を構築することが可能になった。これらはすべて、SQL の熟練やテーブル構造の知識がなくてもできる。これについては、「LLM とデータ処理およびビジネス・プロセスにおけるその役割」で詳しく説明した。

しかし、LLM が自動的に情報を構造化し、順序付けるわけではないことを認識することが重要である。LLM のモデルはデータ中を漂い、クエリのコンテキストに基づいて最も関連性の高いデータを探し出す。もしデータが事前にクリーニングや変換されていなければ、ディープ・サーチはデジタルの「ゴミ」の中から答えを見つけようとするようなものだ。理想的なのは、データを構造化し（例えば、ドキュメントを Markdown に翻訳する）、ベクトル・データベースにロードすることだ。これにより、アウトプットの精度と関連性が大幅に向上する。

> 当初、ベクトル・データベースは機械学習に使用されていたが、今日では検索エンジン、コンテンツのパーソナライズ、インテリジェント分析など、機械学習以外の分野での用途が増えつつある。

建築におけるベクトル・アプローチの最も明白な例のひとつが、バウンディング・ボックス （境界平行六面体）である。これは、3 次元空間におけるオブジェクトの境界を記述する幾何学的な構造である。バウンディングボックスは、X、Y、Z 座標の最小値と最大値によって定義され、オブジェクトを囲む「箱」を形成します。この方法により、ジオメトリ全体を解析することなく、要素のサイズや配置を推定することができます。

各バウンディングボックス は、多次元空間のベクトルとして表すことができる：例えば [x, y, z, width, height, depth] - すでに 6 次元（図 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

図 8.2-1 バウンディングボックス -要素座標情報とプロジェクトモデル内の位置は、ベクトルデータベースに類似している。

このデータ表現により、オブジェクト間の交差のチェック、建築要素の空間分布の計画、自動計算の実行など、多くのタスクが容易になります。バウンディングボックス、複雑な 3D モデルと従来のベクターデータベースの橋渡しの役割を果たし、建築やエンジニアリングのモデリングにおいて、両方のアプローチの利点を効果的に利用することができます

> バウンディングボックス は「ジオメトリのベクトル化」であり、エンベッディング （抽象的なものを変換する方法）は「意味のベクトル化」である。どちらのアプローチも、プロジェクトモデル内のオブジェクトであれ、テキスト内の概念であれ、3D 手動検索からインテリジェント検索への移行を可能にする。

プロジェクト内のオブジェクトの検索（例えば、"幅 1.5m 以上の窓をすべて見つける"）は、ベクトルデータベースにおける最近傍探索（k-NN ）に似ており、基準は特徴空間内の "ゾーン "を定義する。(k-NN 最近傍探索については次の機械学習のパートで詳しく説明します) (Fig. 8.2-2)。バウンディング・ボックスの属性に追加パラメータ（材料、重量、生産時間）を加えると、表は高次元ベクトルとなり、各属性が新たな次元となる。これは、次元が数百から数千単位でカウントされる現代のベクトル・ベースに近い（例えば、ニューラル・ネットワークの埋め込み）。

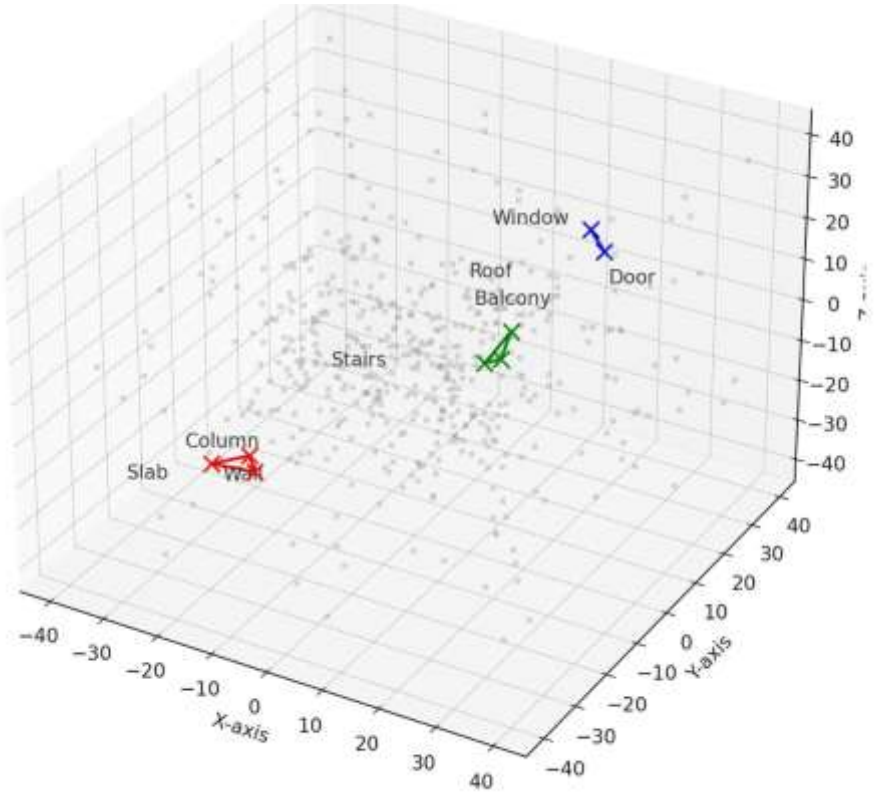


図 8.2-2 ベクターデータベースを使用してプロジェクト内のオブジェクトを検索します。

バウンディングボックス で使われているアプローチは、幾何学的なオブジェクトだけでなく、テキストや言語

解析にも応用できる。データのベクトル表現は、すでに自然言語処理（NLP ）で積極的に使用されている。建設プロジェクトのオブジェクトが空間的な近さに従ってグループ化できるように（図 8.2-2）、テキストの単語は意味的・文脈的な近さに従って分析できる。

例えば、「建築家」、「建設」、「デザイン」という単語は、似たような意味を持っているため、ベクトル空間で隣り合うことになる。LLM では、このメカニズムにより、手動による分類を必要としない自動的な分類が可能になる ：

- テキストの主題を特定する
- 文書内容の意味検索の実行
- 自動注釈とテキスト要約の生成
- 類義語と関連語を探す

ベクターデータベースは、バウンディングボックス（ ）が 3D -モデルにおける空間オブジェクトの分析に役立つのと同じように、テキストを分析し、関連する用語を見つけることを可能にします。プロジェクト要素のバウンディングボックスの例は、ベクトル表現が ML から純粋に 「人工的な」概念ではなく、CAD プロジェクト内の列の検索であれ、データベース内の意味的に近い画像であれ、応用問題を解決するための自然なデータ構造化の方法であることを理解するのに役立つ。

> データベースを扱う専門家は、ベクトルストアに注目すべきである。ベクターストアの普及は、データベース開発の新たな段階を示しており、古典的なリレーショナルシステムと AI（）指向のテクノロジーが絡み合い、未来のハイブリッドソリューションが形成され始めている。

複雑で大規模な AI（ ）アプリケーションを開発するユーザーは、ベクトル検索に特化したデータベースを使うだろう。同時に、既存のアプリケーションに統合するための個別の AI 機能のみを必要とするユーザーは、既に使用しているデータベース（PostgreSQL 、Redis）に内蔵されているベクトル検索機能を選択する可能性が高い。

DWH、データレイク、CDE、PMIS、ベクトル・データベースなどのシステムは、データの保存と管理に対してさまざまなアプローチを提供しているが、その有効性はアーキテクチャだけでなく、データそのものがいかにうまく整理・管理されているかによっても決まる。ベクターデータベース、古典的なリレーショナル DBMS 、データレイク型ウェアハウスなど、最新のソリューションを使用している場合でも、データを管理、構造化、更新するための明確なルールがないと、異種ファイルやマルチフォーマットデータを扱うユーザーが直面す

るのと同じ困難につながる可能性があります。

データ・ガバナンス）がなければ、どんなに強力なソリューションであっても無秩序で非構造化なものとなり、データレイクをデータ沼）に変えてしまう可能性がある。このような事態を避けるために、企業は適切なストレージアーキテクチャを選択するだけでなく、データを効果的な意思決定ツールに変えるために、データミニマリズム)、アクセス管理、品質管理戦略を導入する必要がある。

データガバナンス）、データミニマリズム）、データスワンプ）

データウェアハウスをうまく管理し、ビジネス価値を提供するためには、データガバナンス（Data Governance）、データミニマリズム（Data Minimalism）、データスワンプ（Data Swamp）の概念を理解し、実行することが鍵となります（図 8.2-3）。

> ガートナー社の調査（2017 年）によると、ビッグデータ・プロジェクトの 85%は失敗しており、その主な理由の 1 つはデータ品質とデータガバナンスが不十分なことである[144]。

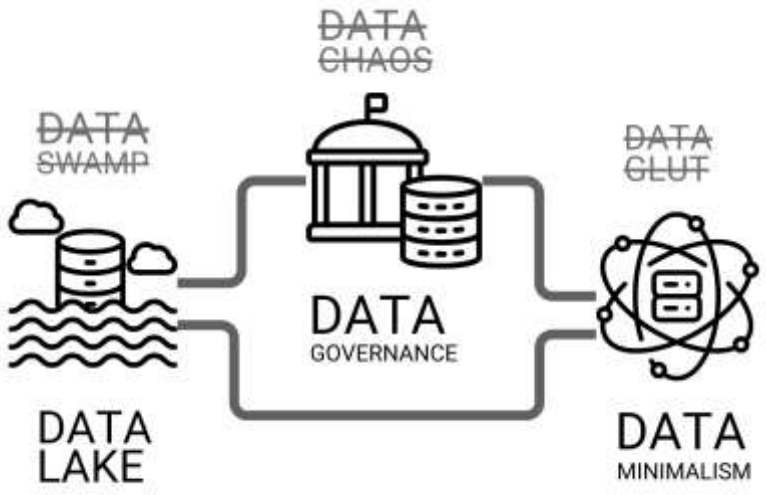


図 8.2-3 データガバナンスの重要な側面として、データガバナンス とデータミニマリズム がある。

データガバナンス（Data Governance データマネジメントの基本的な要素であり、あらゆるビジネスプロセスにおいてデータが適切かつ効果的に利用されることを保証するものである。ルールや手順を確立するだけでな

く、データの可用性、信頼性、セキュリティを確保することも重要である：

- データの定義と分類：エンティティの定義と分類を明確にすることで、組織は企業でどのようなエンティティが必要かを理解し、それらをどのように使用すべきかを決定することができる。
- アクセス権と管理：データへのアクセスと管理に関する方針と手順を策定することで、権限を与えられたユーザーのみが特定のデータにアクセスできるようにする。
- 外部からの脅威からデータを守る：外部からの脅威からデータを守ることは、データ管理の重要な側面である。これには、技術的な対策だけでなく、情報セキュリティの基本に関する従業員のトレーニングも含まれる。

データ・ミニマリズム（Data Minimalism とは、データを最も価値があり、意味のある属性とエンティティに絞り込むアプローチであり（図 8.2-4）、それによってコストを削減し、データの利用率を向上させる：

- 意思決定の簡素化：オブジェクトとその属性の数を最も関連性の高いものに絞り込むことで、データの分析と処理に必要な時間とリソースを削減し、意思決定を簡素化する。
- 重要なものに焦点を当てる：最も関連性の高いエンティティや属性を選択することで、ビジネスにとって本当に重要な情報に焦点を当てることができ、ノイズや不要なデータを排除することができる。
- 効率的なリソースの割り当て：データの最小化により、より効率的なリソースの割り当てが可能になり、データの保存と処理のコストが削減され、データの品質とセキュリティが向上します。

データを扱う論理は、そのようなデータを作成することから始めるべきで はなく（図 8.2-4）、生成プロセスを開始する前から、このデータを使用する将来のシナリオを 理解することから始めるべきである。このアプローチにより、属性に最低限必要な要件、そのタイプ、境界値を事前に定義することができる。これらの要件は、情報モデルにおいて正しく安定した実体を作成するための基礎となる。データの目的と用途の事前理解は、分析に適した構造の形成に貢献する。概念的、論理的、物理的レベルでのデータモデリングへのアプローチの詳細については、「データモデリング：概念的、論理的、物理的モデル」の章で説明した。

> 建設会社の伝統的なビジネスプロセスでは、データ処理は沼地にデータを捨てるようなもので、まずデータを作成し、それを専門家が他のシステムやツールに統合しようとすることが多い。

データ・スワンプ（Data Swamp とは、適切な整理、構造化、管理がなされないまま、データの収集と保存が無秩序に行われた結果、構造化されておらず、利用が困難で、価値の低いデータが生み出されることである。

情報の流れが泥沼化するのをどう防ぐか：

- データ構造管理： データが構造化され、分類されていることを確認することで、整然とし、容易にアクセスできるようにすることで、データの氾濫を防ぐことができる。
- データの理解と解釈：データの起源、修正、意味を明確に記述することで、データが正しく理解され、解釈されるようにする。
- データ品質の維持: 定期的なデータメンテナンスとクレンジングにより、アナリティクスとビジネスプロセスにおけるデータ品質、関連性、価値を維持することができます。

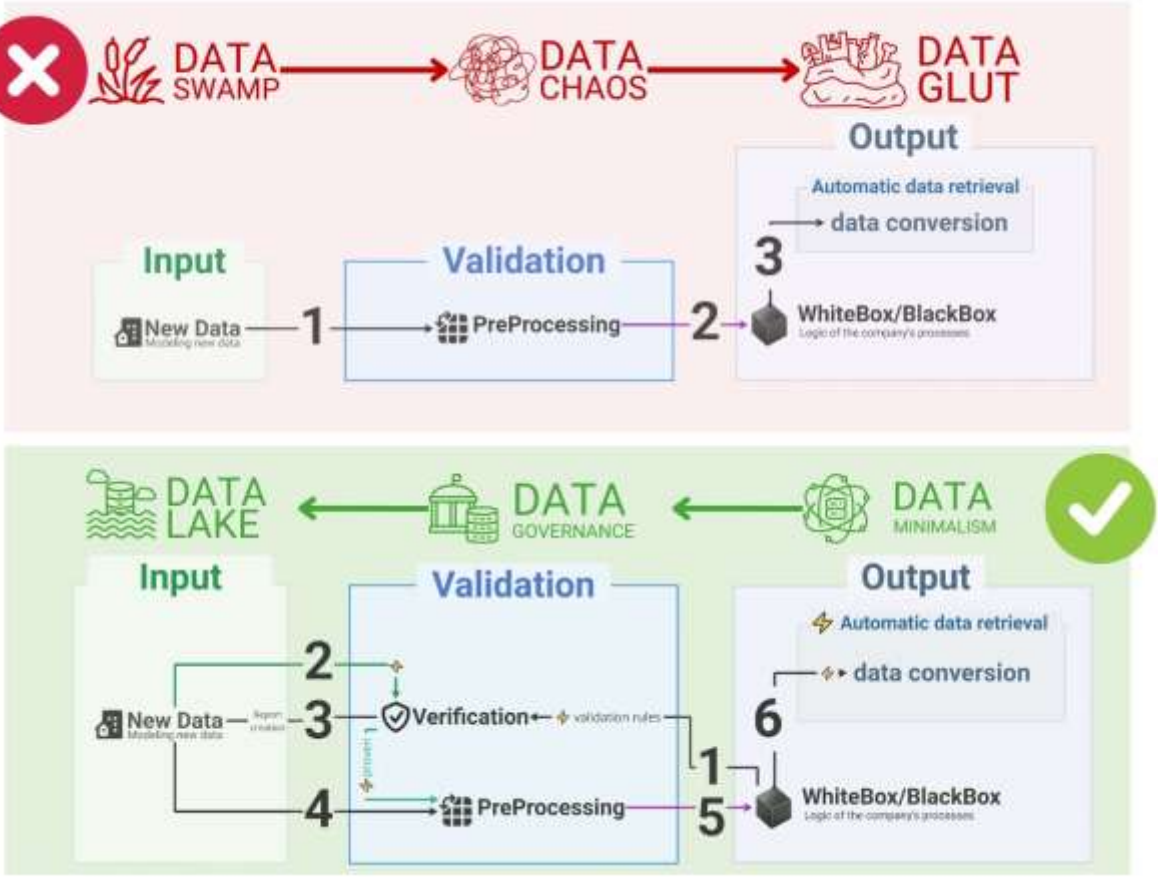


図 8.2-4 データ・ウェアハウスの混乱を避けるために、データ作成プロセスは属性要件を収集することから始めるべきです。

> データガバナンスとデータミニマリズムの原則をデータ管理プロセスに統合し、データウェアハウス（）がデータ沼になるのを積極的に防ぐことで、組織はデータの可能性を最大限に引き出すことができる。

データを扱う進化における次の段階は、管理とミニマリズムの問題を解決した後、自動処理の標準化、品質保

証、データを分析、変換、意思決定に使えるようにする手法の導入である。これは、DataOps と VectorOps の方法論が行っていることであり、ビッグデータと機械学習に取り組む企業にとって重要なツールになりつつある。

DataOps と VectorOps: 新しいデータ標準

Data Governance がデータのコントロールと整理を担当するのに対して、DataOps はデータの正確性、一貫性、社内でのスムーズな流れを確保するのを支援する。これは、データが継続的に生成され、タイムリーな処理が要求される建設業における多くのビジネスケースで特に重要である。例えば、建築情報モデル、プロジェクト 要件、分析レポートを 1 営業日以内に異なるシステム間で同期させる必要がある状況では、DataOps の役割が重要なカギとなります。安定した反復可能なデータ処理プロセスを構築し、情報の遅延や関連性喪失のリスクを低減することができます。

データガバナンスだけでは十分ではない。データを単に保存するだけでなく、日々の業務で積極的に活用することが不可欠である。そこで、自動化、統合、継続的なデータフローに焦点を当てた方法論である DataOps の出番となる。

> DataOps は、組織におけるデータフローのコラボレーション、統合、自動化の改善に焦点を当てている。DataOps のプラクティスを採用することで、データの正確性、一貫性、可用性が促進され、これはデータ中心のアプリケーションにとって重要である 。

DataOps エコシステムの主要ツールは、ワークフローオーケストレーションのための Apache Airflow （図 7.4-4）と、データフローのルーティングと変換のための Apache NiFi （図 7.4-5）である。これらの技術を組み合わせることで、柔軟性、信頼性、拡張性に優れたデータパイプラインを実現し、 システム間の情報を自動的に処理、制御、統合することができる（詳細は「自動 ETL -conveyor 」の章を参照）。建設プロセスに DataOps アプローチを導入する場合、4 つの基本的な側面を考慮することが重要である：

1. データよりも重要なのは人とツール：サイロ化したデータ・リポジトリ は大きな問題と見られるかもしれないが、現実はもっと複雑だ。データの断片化に加え、チームの孤立や、チームが使用するツールの違いも大きな役割を果たしている。建設業界では、データ・エンジニアやアナリスト（ ）、BI やビジュアライゼーション・チーム、プロジェクト管理や品質の専門家など、さまざまな分野の専門家がデータを扱っている。そのため、参加者間でデータが自由に流れ、単一の一貫したバージョンの情

報を提供するエコシステムを構築することが重要になる。

2. テストとエラー検出の自動化：建設データには、モデルの不正確さ、計算ミス、古い仕様など、常にエラーが含まれています。定期的にデータをテストし、繰り返し発生するエラーを排除することで、データ品質を大幅に向上させることができます。DataOps の一環として、データの正確性を監視し、エラーを分析し、パターンを特定し、すべてのワークフローでシステム障害を捕捉して対処する、自動化されたコントロールと検証メカニズムを実装する必要があります。自動化された検証の程度が高ければ高いほど、全体的なデータ品質が向上し、最終段階でエラーが発生する可能性が低くなる。
3. データはプログラムコードと同じようにテストされるべきである。ほとんどのビルアプリケーションはデータ処理に基づいているが、その制御は二次的な役割に任されることが多い。機械学習モデル が不正確なデータでトレーニングされると、誤った予測や金銭的損失につながる。DataOps では、データはソフトウェアコードと同じように厳密な精査を受けるべきである：ロジックチェック、ストレステスト、入力値が変化した場合のモデルの動作評価。検証された信頼できるデータのみが、経営判断の基礎として使用できる。
4. パフォーマンスを損なわないデータ観測性：データモニタリングは、 単なるメトリクスのコレクションではなく、戦略的な品質管理ツールである。DataOps が効果的に機能するためには、設計から運用に至るまで、データを扱うすべての段階で観測可能性を組み込む必要がある。同時に、モニタリングがシステムの速度を低下させないことも重要である。建設プロジェクトでは、データを収集するだけでなく、データを作成する専門家（設計者など）の作業を一切妨げないような方法で行うことが重要です。このバランスにより、生産性を損なうことなくデータの品質を管理することができる。

DataOps はデータサイエンティストにとって追加的な負担ではなく、彼らの仕事のバックボーンである。DataOps を導入することで、建設会社は混沌としたデータ管理から、データがビジネスのために機能する効率的なエコシステムへと移行することができる。

VectorOps、DataOps の進化における次のステージであり、（前章で説明した）多次元ベクトルデータの処理、保存、分析に焦点を当てている。これは特に、デジタル・ツイン、ニューラル・ネットワーク・モデル、セマンティック検索など、建設業界にも導入され始めている分野に関連している。VectorOps は、オブジェクトの多次元表現を効率的に保存、索引付け、検索するために、ベクトルデータベース 。

> VectorOps は、DataOps の次のステップであり、構築におけるベクトルデータの処理、分析、利用に焦点を当てている。データフロー、一貫性、品質に重点を置く DataOps とは異なり、VectorOps は機械学習に必要な多次元オブジェクト表現の管理に重点を置く 。

従来のアプローチとは異なり、VectorOps は、デジタルツイン、ジェネレーティブデザインシステム、およびベクトル形式に変換された CAD データの自動エラー検出に不可欠な、より正確なオブジェクト記述を実現します。DataOps と VectorOps を組み合わせて実装することで、古典的なテーブルからセマンティックでリッチな空間モデルまで、大量の情報をスケーラブルに自動化するための強固な基盤が形成されます

次のステップ：カオス・ストレージから構造化ストレージへ

データ・ストレージを構築する従来のアプローチでは、重要な洞察が分析や意思決定に利用できない「情報のサイロ化」が生じることが多い。データウェアハウス（Data Warehouse） 、データレイク（Data Lake） 、およびそれらのハイブリッドといった最新のストレージコンセプトは、異種の情報を統合し、データストリーミングやビジネスインテリジェンス（Business Intelligence） のために一元化された方法で利用できるようにする。適切なストレージ・アーキテクチャを選択するだけでなく、データ・ガバナンス（Data Governance ）やデータ・ミニマリズム（Data Minimalism ）を導入し、ストレージ設備が制御不能なデータ沼（Data Swamps）にならないようにすることも重要である。

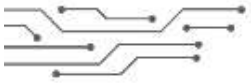
このパートを要約するために、説明したコンセプトを日々の仕事に応用するのに役立つ主な実践的ステップを挙げておこう：

- 効率的なデータ保存形式を選択する
  - 大量のデータを保存するために、CSV や XLSX からより効率的なフォーマット（Apache Parquet 、ORC）へ移行する。
  - データのバージョン管理システムを導入し、変更を追跡する
  - メタデータを使用して、情報の構造と出所を記述する。
- 統一された企業データ・アーキテクチャの構築
  - RDBMS、DWH、データレイクなど、さまざまなストレージ・アーキテクチャを比較。スケーラビリティ、ソース統合、分析処理のニーズに最適なものをお選びください。
  - 様々なソースからデータを抽出、ロード、変換（ETL）するためのプロセスマップをタスクに合わせて設計する。Miro、Lucidchart、Draw.io などの可視化ツールを使用して、主要なステップと統合ポイントを可視化する。
- データガバナンスの実践 とデータミニマリズム

- ☐ データミニマリズムのアプローチに従う - 本当に価値のあるものだけを保存し、処理する。
- ☐ データガバナンスの原則の導入 - データに対する責任を明確にし、品質と透明性を確保する。
- ☐ データ管理ポリシーと DataOps コンセプトの詳細, VectorOps
- ☐ DataOps におけるデータ検証のためのデータ品質基準 および手順を定義する。

整理されたデータ・ストレージは、企業の分析プロセスを一元化するための基盤を作ります。無秩序なファイルの集積から構造化されたストレージへの移行により、情報を戦略的資産に変え、十分な情報に基づいた意思決定とビジネス・プロセスの効率化を支援します。

データの収集、変換、分析、構造化保存のプロセスが自動化・標準化されたら、デジタル変革の次の段階はビッグデータの本格的な取り扱いだ。





IX 部

ビッグデータ、機械学習、予測

第 9 部では、建設業界におけるビッグデータ、機械学習、予測分析に焦点を当てる。直感的な意思決定から、過去のデータに基づく客観的な分析への移行を探る。サンフランシスコの建築許可データセットの解析から、CAD - 数百万の要素を持つプロジェクトの処理まで、建設におけるビッグデータ分析を実証するために実践的な例が用いられている。特に、建設プロジェクトのコストとタイミングを予測するための機械学習手法（ ）に注目し、線形回帰と k-nearest neighbors アルゴリズムについて詳しく説明する。構造化されたデータが、リスクを評価し、リソースを最適化し、プロジェクト管理の効率を向上させる予測モデルの基礎となることが示されている。また、代表的なデータサンプルの選択方法に関する推奨事項や、効果的な分析に大規模なデータセットが必ずしも必要でない理由についても説明しています。

章 9.1.
ビッグデータとその分析

建設におけるビッグデータ：直感から予測可能性へ

ビッグデータ」という言葉に厳密な定義はない。もともとは、情報量が従来の処理方法の能力を超え始めたときに登場した概念である。今日、建設業を含む多くの産業において、データの量と複雑さは、コンピュータのローカル・メモリに収まらないほど増大しており、それを処理するために新たなテクノロジーを使用する必要がある。

ビッグデータ活用の本質は、保存と処理だけでなく、予測能力にもある。建設業界では、ビッグデータは、表やビジュアライゼーションの主観的な解釈に基づく直感的な意思決定（前述）から、実際の観察や統計に裏打ちされた情報に基づく予測への道を開く。

一般に信じられていることとは異なり、ビッグデータを扱う目的は「機械に人間のような思考をさせる」ことではなく、パターンを特定し、事象を予測し、プロセスを最適化するために、膨大な量のデータを分析する数学的モデルとアルゴリズムを適用することである。

> ビッグデータは、人間の影響を排除した冷たいアルゴリズムの世界ではない。それどころか、ビッグデータは私たちの本能や過ち、創造性と連動している。人間の思考が不完全であるからこそ、私たちは非標準的な解決策を見出し、ブレークスルーを起こすことができるのだ。

デジタル技術の発達に伴い、建設業界は、IT 分野から生まれたデータ処理技術を積極的に活用し始めている。Pandas（） や Apache Parquet（） といったツールのおかげで、構造化データと非構造化データを組み合わせることができ、情報へのアク セスを簡素化し、分析にかかるロスを減らすことができる。また、文書や CAD プロジェ クトから得られる大規模なデータセット（図 9.2-10 - 図 9.2-12） によって、プロジェクトのライフサイクルの全段階でデータを収集、分 析、予測することができる。

ビッグデータは建設業界に変革的な影響を与え、様々な形で潜在的な影響を及ぼしている。ビッグデータ技術の応用は、例えば以下のような重要な分野で成果を上げている：

- 投資可能性分析 - 過去の施設のデータに基づき、プロジェクトの収益性と投資回収期間を予測。

- 予知保全 - 実際に故障が発生する前に、故障の可能性が高い機器を特定し、ダウンタイムを短縮する。
- サプライチェーンの最適化 - 混乱を予測し、物流効率を改善する。
- エネルギー効率分析 - 低エネルギー建築物の設計を支援。
- 安全監視 - センサーやウェアラブルデバイスを使用して現場の状況を監視する。
- 品質管理 - プロセス標準への準拠をリアルタイムで監視。
- 人員管理 - 業績分析と必要人員の予測。

建設業界において、データ分析 ）や予測が求められていない分野を見つけるのは難しい。予測アルゴリズムの主な利点は、自己学習能力があり、データが蓄積されるにつれて常に改善されることである。

> 近い将来、人工知能は単に建築業者を支援するだけでなく、設計プロセスから建物の運用問題に至るまで、重要な決定を下すようになるだろう。

予測がどのように生成され、学習モデルがどのように使用されるかについては、本書の次のパート「機械学習と予測」で詳しく説明する。

ビッグデータを本格的に扱うようになるには、アナリティクスへのアプローチそのものを変える必要がある。これまで私たちが考えてきた古典的なシステムが原因と結果の関係に焦点を当てていたのに対し、ビッグデータ解析は、すべての要因を完全に理解していなくても、隠れた関係を特定し、対象物の行動を予測することを可能にする統計的パターンと相関関係の探索へと焦点を移しつつある。

ビッグデータの実現可能性を問う：相関、統計、データサンプリング

伝統的に、建設は主観的な仮説と個人の経験に基づいていた。技術者たちは、材料がどのような挙動を示すか、構造物がどのような荷重に耐えられるか、プロジェクトがどれくらいの期間もつかを、ある程度の確率で想定していた。これらの仮定は、多くの場合、時間、資源、将来のリスクを犠牲にして、実際に検証された。

> ビッグデータの出現により、そのアプローチは劇的に変化している。意思決定はもはや、直感的なカンに基づいてなされるのではなく、大規模なデータセットを分析した結果としてなされるようになっている。建設は次第に直感の芸術ではなくなり、正確な予測の科学になりつつある。

ビッグデータの活用という考え方への移行は、必然的に重要な問題を提起する。それは、信頼性の高い予測分析に必要なデータ量と情報量は果たしてどの程度なのかということだ。データが多ければ多いほど精度が高くなる」という考え方が広く浸透しているが、実際に統計的に妥当であるとは限らない。

> 1934 年には早くも、統計学者のイエジー・ノイマンが、統計的推測の正確さの鍵は、データの量よりも、そのデータの代表性とサンプリングの無作為性にあることを証明した[145]。

特に建設業界では、IoT - センサー、スキャナー、監視カメラ、ドローン、さらにはマルチフォーマット CAD - モデルを使用して大量のデータが収集されるため、死角、異常値、データの歪みのリスクが高まる。

路面状態のモニタリングの例を考えてみよう。すべての道路区間の完全なデータセットには X GB かかり、処理に約 1 日かかるかもしれない。同時に、50 番目の道路区間だけを含むランダム・サンプルなら、X/50GB しかかからず、30 分で処理でき、しかも特定の計算については同じような推定精度が得られる（図 9.1-1）。

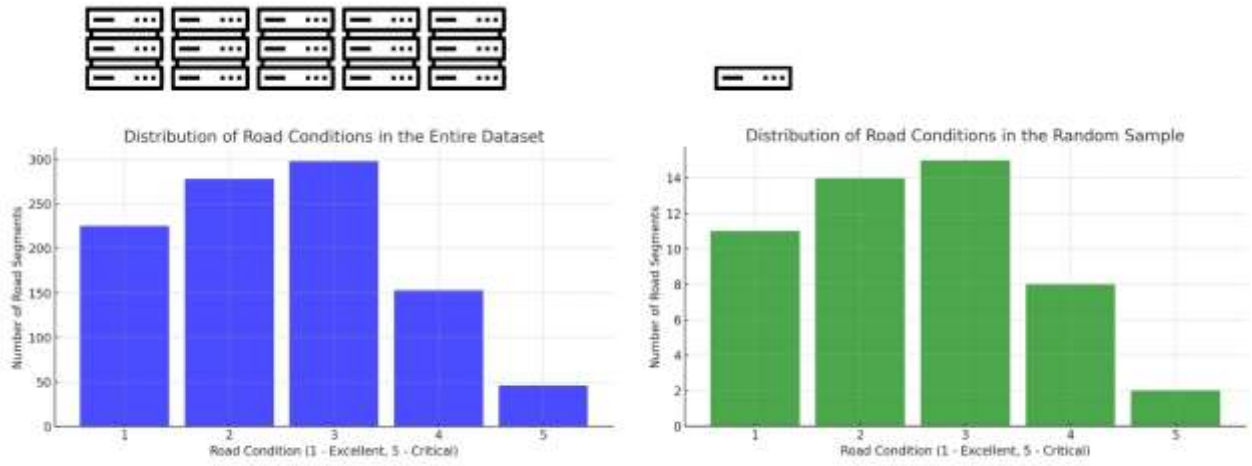


図 9.1-1 舗装状態ヒストグラム：全データセットと無作為抽出の結果は同じである。

したがって、データ分析を成功させる鍵は、データ量ではなく、サンプルの代表性と使用した処理方法の質にあることが多い。ランダムサンプリングとより選択的なアプローチへの移行は、建設業界における考え方の転換を必要とする。歴史的に、企業は「データは多ければ多いほど良い」という論理に従い、可能な限りの指標を網羅することが精度を最大化すると信じてきた。

このアプローチは、プロジェクトマネジメントでよく使われる誤解を想起させる：「スペシャリストを集めれば集めるほど、より効果的な仕事ができる」。しかし、人材と同様、重要なのは量よりも質とツールである。データやプロジェクト参加者間の相互関係（相関関係）を考慮せずに量を増やしても、ノイズや歪み、重複、不必要な無駄を生むだけだ。

結局のところ、多くの矛盾したシグナルを含む膨大だが無秩序な情報に頼るよりも、安定した妥当な予測を生み出すことのできる、小規模だが定性的に準備されたデータセットを持つ方がはるかに生産的であることがしばしば判明する。

> 過剰なデータ量は精度を保証しないだけでなく、ノイズ、冗長な特徴、隠れた相関関係、無関係な情報の存在により、結論が歪む可能性もある。このような状況では、モデルのオーバーフィッティングのリスクが高まり、分析結果の信頼性が低下する。

建設業界では、ビッグデータを扱う際の大きな課題は、最適なデータの量と質を決定することである 。例えば、コンクリート構造物の状態をモニタリングする場合、何千ものセンサーを使用し、毎分情報を収集すると、ストレージや分析システムを圧倒する可能性がある。しかし、相関分析を行い、最も情報量の多い 10%のセンサーを選択すれば、何倍も、時には何十倍も何百倍も少ないリソースで、ほぼ同じ精度の予測が可能になる。

より少ないデータのサブセットを使用することで、必要なストレージの量と処理時間の両方が削減され、データの保存と分析にかかるコストが大幅に削減されるため、特に大規模なインフラプロジェクトやリアルタイムで作業する場合、ランダムサンプリングが予測分析にとって理想的なソリューションになることがよくあります。最終的に、建設プロセスの効率は、収集したデータの量ではなく、その分析の質によって決まります。批判的なアプローチと慎重な分析がなければ、データは誤った結論を導きかねません。

> ある一定のデータ量を超えると、新たな情報単位が得られるたびに、有用な結果は少なくなっていく。情報を際限なく収集するのではなく、その代表性と分析方法に注目することが重要である（図 9.2-2）。

この現象はアレン・ウォリス[146]によってよく説明されており、彼は米海軍の 2 つの代替発射体設計をテストする例を用いて統計的手法の使用を説明している。

> 海軍は、一連のペアラウンドを実施することにより、2 つの代替発射体設計（A と B）をテストした。各ラウンドで、A は B より性能が良いか悪いかによって 1 または 0 を受け取り、逆もまた同様である。標準的な統計的アプローチでは、一定の試行回数（たとえば 1000 回）を行い、パーセンテージ分布に基づいて勝者を決定する（たとえば、A が 53%以上の確率で 1 を獲得した場合、それが最良とみなされる）。アレン・ウォリスが（海軍の）ギャレット・L・シューラー大尉とこのような問題について話し合ったとき、大尉は、アレンの説明を引用すれば、このようなテストは役に立たないかもしれないと異議を唱えた。もしシュイラーのような賢明で経験豊富な兵器将校がその場にいたならば、最初の数百[発]の後に、新しい方法が明らかに劣っているか、あるいは期待されたものより明らかに優れているかのどちらかであるため、実験を打ち切る必要がないことを見抜いただろう[146]。
>
> - コロンビア大学アメリカ政府統計研究グループ、第二次世界大戦期

この原理はさまざまな業界で広く使われている。例えば医学の分野では、新薬の臨床試験は無作為に抽出された患者を対象に行われる。経済学や社会学では、国民全員に聞き取り調査をすることなく、社会の意見を反映させるために代表調査が行われる。

政府や研究機関が社会の一般的な傾向を把握するために小規模な集団を調査するのと同じように、建設業界の企業は、プロジェクト管理のための効果的なモニタリングや予測の作成にランダムなデータサンプルを使用することができます（図 9.1-1）。

> ビッグデータは社会科学へのアプローチを変えるかもしれないが、統計的常識に取って代わることはない[147]。
>
> - トーマス・ランズオール＝ウェルフェア「国民の現在の気分を予測する」『Significance v.9(4), 2012 r.

資源節約の観点から、将来の予測や意思決定のためにデータを収集する場合、「段階的にスケールアップできるはるかに小規模で安価なテストデータセットを使用できるのに、膨大なデータセットの収集と処理に多大なリソースを費やすことに意味があるのか」という疑問に答えることが重要である。ランダムサンプリングの有

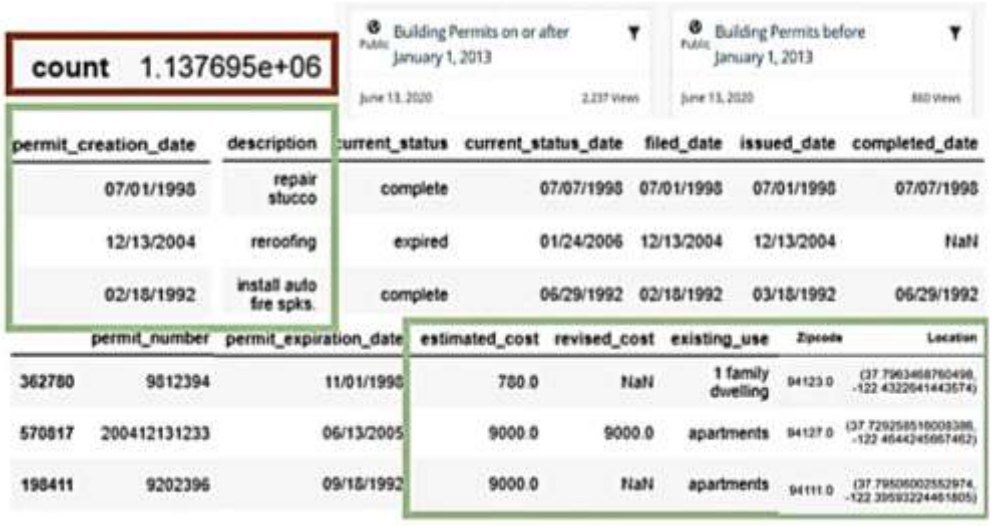
効性は、包括的なカバレッジを必要としないが、それでも十分な精度と代表性を提供するデータ収集方法を選択することで、企業はモデルの収集とトレーニングにかかるコストを数十倍から数千倍削減できることを示している。このアプローチにより、小規模の企業であっても、大幅に少ないリソースとデータ量で、大企業と同等の結果を得ることができる。以下の章では、ビッグデータツールを使用した公開データセットに基づく アナリティクスと予測アナリティクスの例を探る。

ビッグデータ：サンフランシスコの 100 万件の建築許可データセットからのデータ分析

オープンデータセットの利用は、前章で述べた原則、すなわち適切な特徴の選択、代表的なサンプリング、視覚化、批判的分析を実践するまたとない機会を提供する。本章では、大都市における建設活動のような複雑な現象を、オープンデータ、特にサンフランシスコにおける 100 万件以上の建築許可記録を用いてどのように調査できるかを探る。

San Francisco Department of Buildings"[148]から公開されている 100 万件以上の建築許可データ（図 9.1-2）（CSV 形式の 2 つのデータセットに記録されている。[148]により、生の CSV -表を使用して、市内の建設活動（ ）を分析するだけでなく、1980 年から 2019 年までの過去 40 年間におけるサンフランシスコの建設産業の最近の傾向と歴史を批判的に分析することができる。

データセットの視覚化（図 9.1-3-図 9.1-8）に使用したコード例、およびコード、説明、コメント付きの視覚的グラフは、Kaggle プラットフォーム で "San Francisco.建設セクター 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 860 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	9512394	11/01/1998	780.0	NaN	1 family dwelling	94123.0	(37.7963468760498, -122.432264144357 4)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0	(37.7292585160008386, -122.464424566746 2)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0	(37.7950600255297 4, -122.395932244618 05)

図 9.1-2 このデータセットには、さまざまなオブジェクト属性を持つ、発行された建築許可の情報が含まれている。

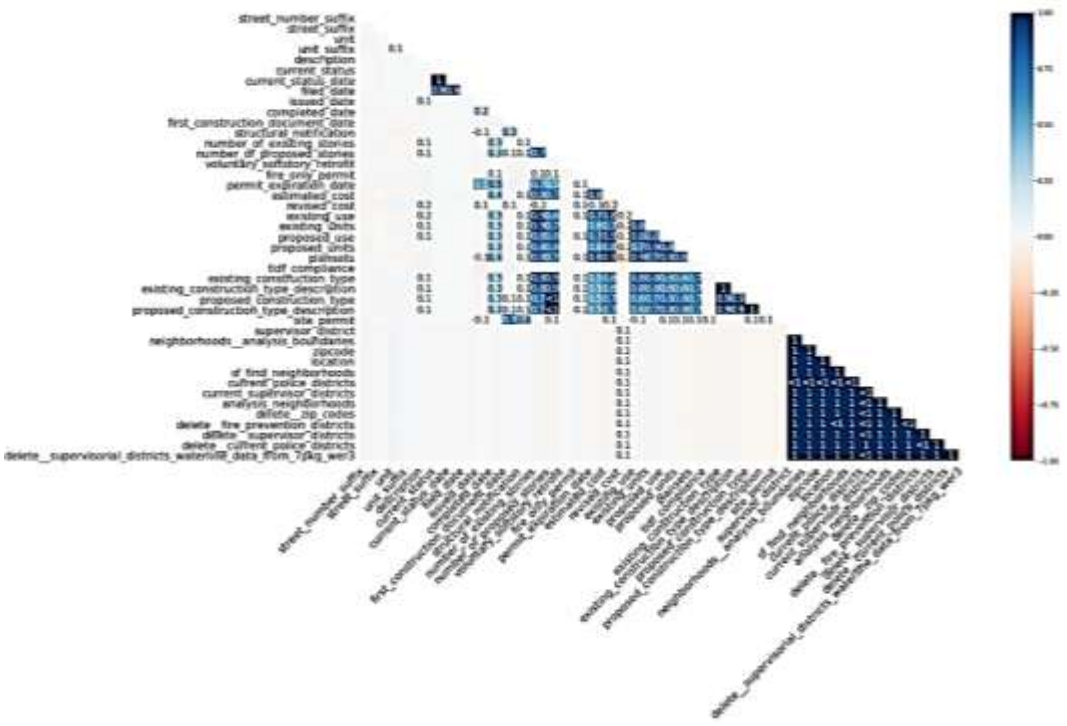


図 9.1-3 データセットのすべての属性を視覚化し、属性のペア間の関係を特定するのに役立つヒートマップ（Pandas および Seaborn ）。

サンフランシスコ市建築局が提供した表（図 9.1-2）からは、傾向や結論は明らかではない。表形式の乾燥した数字は意思決定の根拠にはならない。データを視覚的に理解できるようにするには、データの視覚化に関する

章で詳しく説明するように、本書の第 7 部「ETL 、結果をグラフとして視覚化する」で説明する様々なライブラリを使用して視覚化する必要がある。

Pandas DataFrame と Python の可視化ライブラリ を使って 1,137,695 件の許可証[148]のデータを分析した結果、サンフランシスコの建設活動は、景気サイクル、特に活況を呈しているシリコンバレーのテクノロジー産業と密接に関連していると結論づけることができる（図 9.1-4）。

景気の好不況は、建設プロジェクトの数や金額に大きな影響を与える。例えば、建設活動の最初のピークは、1980 年代半ばの電子機器ブームと重なり（Pandas と Matplotlib を使用）、その後のピークと衰退は、ドットコムバブルと近年の技術ブームに関連している。

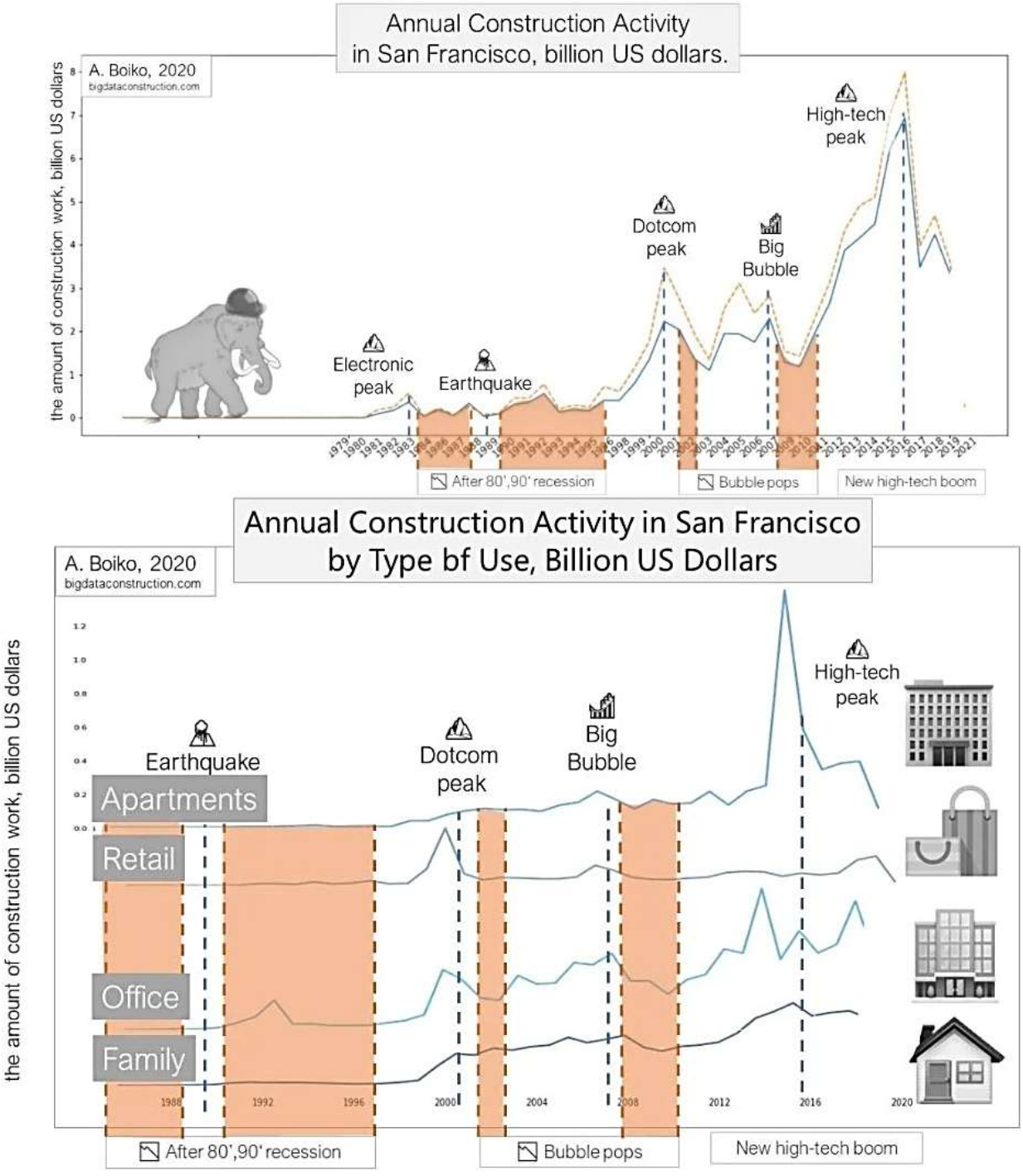


図 9.1-4 サンフランシスコの不動産セクターでは、投資はシリコンバレーの技術開発と相関関係にある。

データ分析によると、サンフランシスコでは、過去 10 年間に建設や再開発に投資された 915 億ドルの大部分（ほぼ 75%）が、市の中心部（図 9.1-5 - Pandas 、Folium 可視化ライブラリを使用）および中心部から半径

2km 以内に集中しており、これらの中心部における投資の密度が高いことを反映している。

建築許可にかかる平均費用は地域によってかなり差があり、市中心部での申請は、土地代、人件費、資材費、エネルギー効率を高めるためにより高価な資材を必要とする厳しい建築規制が高いため、市中心部以外での申請の 3 倍の費用がかかる。

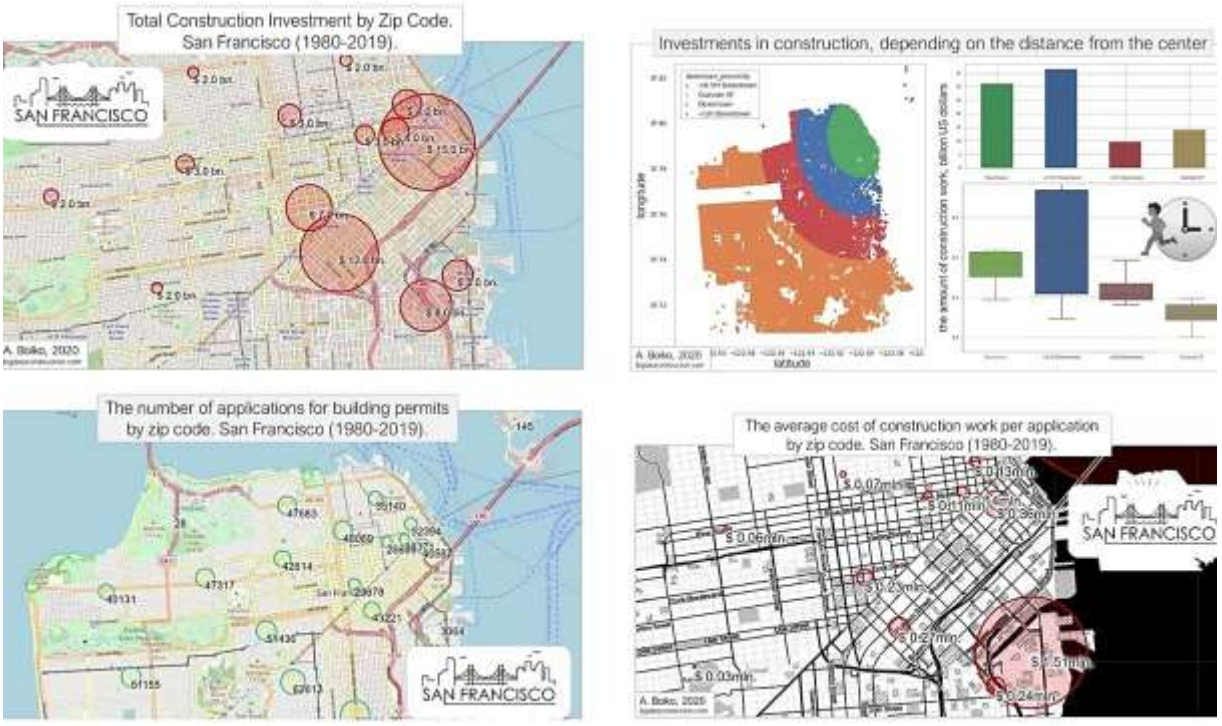


図 9.1-5 サンフランシスコでは、建設投資の 75%（915 億ドル）がダウンタウンに集中している。

また、このデータセットでは、住宅タイプ別だけでなく、市街地や個々の住所（ポストコード）別の平均修繕価格の算出も可能である。サンフランシスコでは、住宅の改修費の動態は、改修の種類や住宅ごとに異なる傾向を示している（図 9.1-6 - Pandas と Matplotlib を使用）。キッチンの改修はバスルームの改修よりも明らかに高額である。一戸建て住宅の平均的なキッチンの改修費用は約 28,000 ドルであるのに対し、二戸建て住宅では 25,000 ドルである。

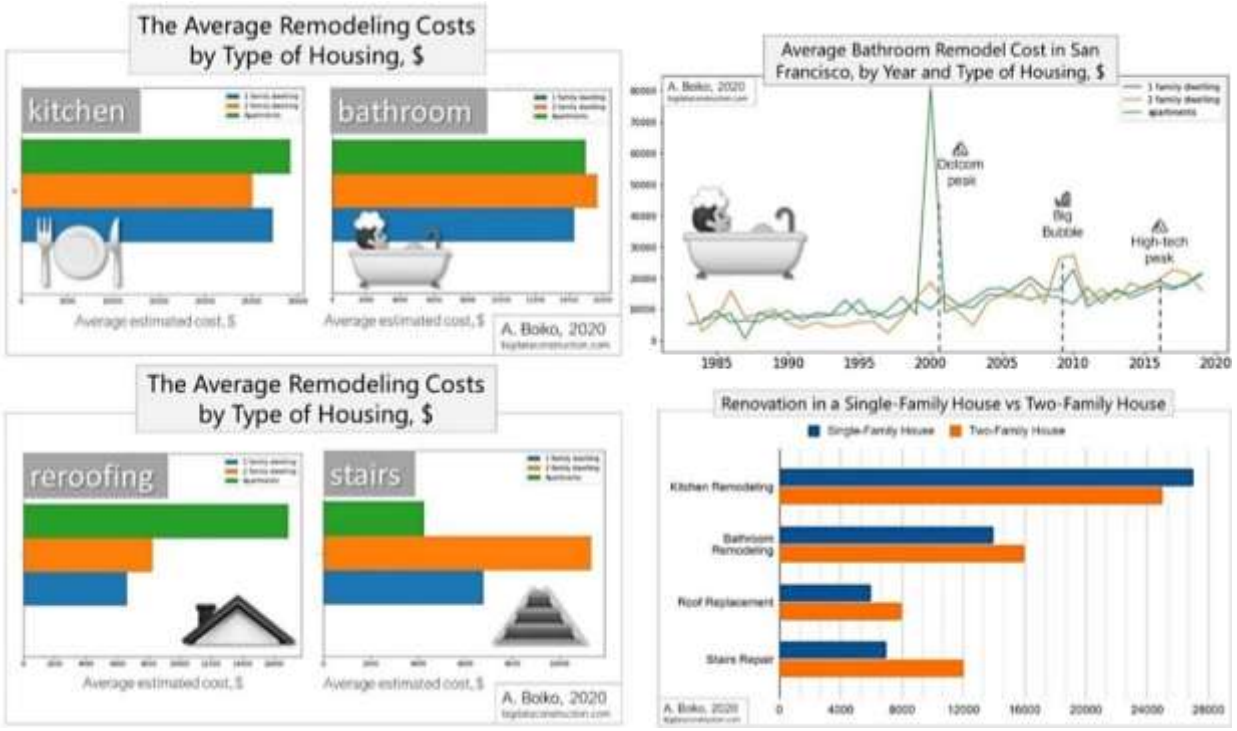


図 9.1-6 SF では、キッチンの改修費用はバスルームの改修費用のほぼ 2 倍で、住宅所有者は大規模修繕費用を賄うために 15 年間、毎月$ 350 を積み立てる必要がある。

サンフランシスコにおける建設費のインフレは、住宅タイプ別、年別に分類されたデータを分析することによって追跡することができる（図 9.1-7 - Pandas と Seaborn で使用）。それによると、1990 年以降、平均修繕費は着実に上昇しており、集合住宅の修繕費は 3 年周期で短期的に上昇していることがわかる。

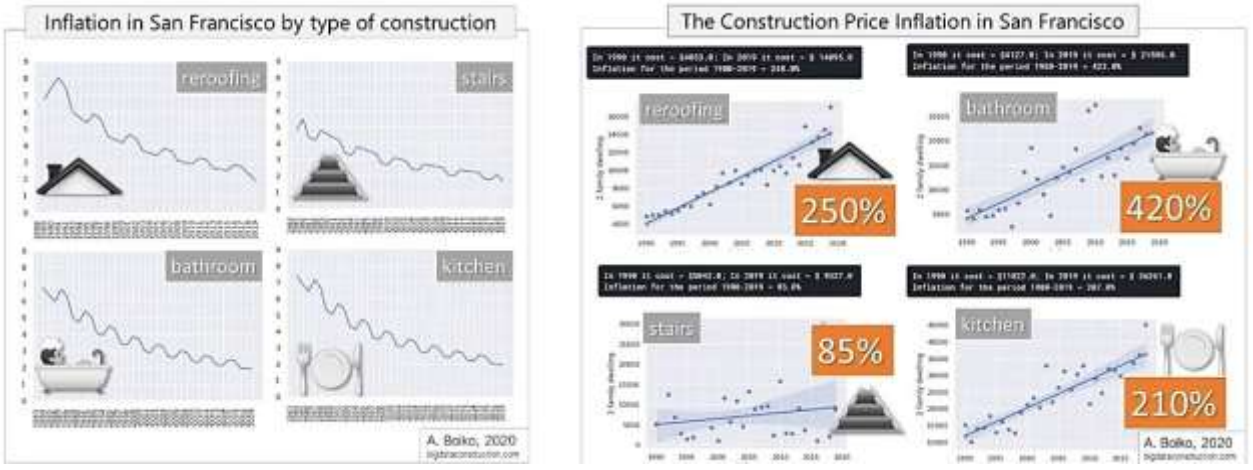


図 9.1-7 1980 年から 2019 年にかけて、SF における浴室の改修費用は 5 倍になったが、屋根とキッチンの改修費用は 3 倍になり、階段の改修費用は 85%しか増加していない。

サンフランシスコ市建築局が公表しているデータ（図 9.1-3）を調査した結果、市の建築費は様々な要因に影響され、非常に変動しやすく、しばしば予測不可能であることが明らかになった。これらの要因には、経済成長、技術革新、さまざまな住宅タイプに固有の要件（）が含まれる。

以前は、このような分析にはプログラミングや分析に関する深い知識が必要でした。しかし、LLM- ツールの登場により、このプロセスは、設計部門のエンジニアから上級管理職まで、建設業界の幅広い専門家にとって身近で理解しやすいものとなった。

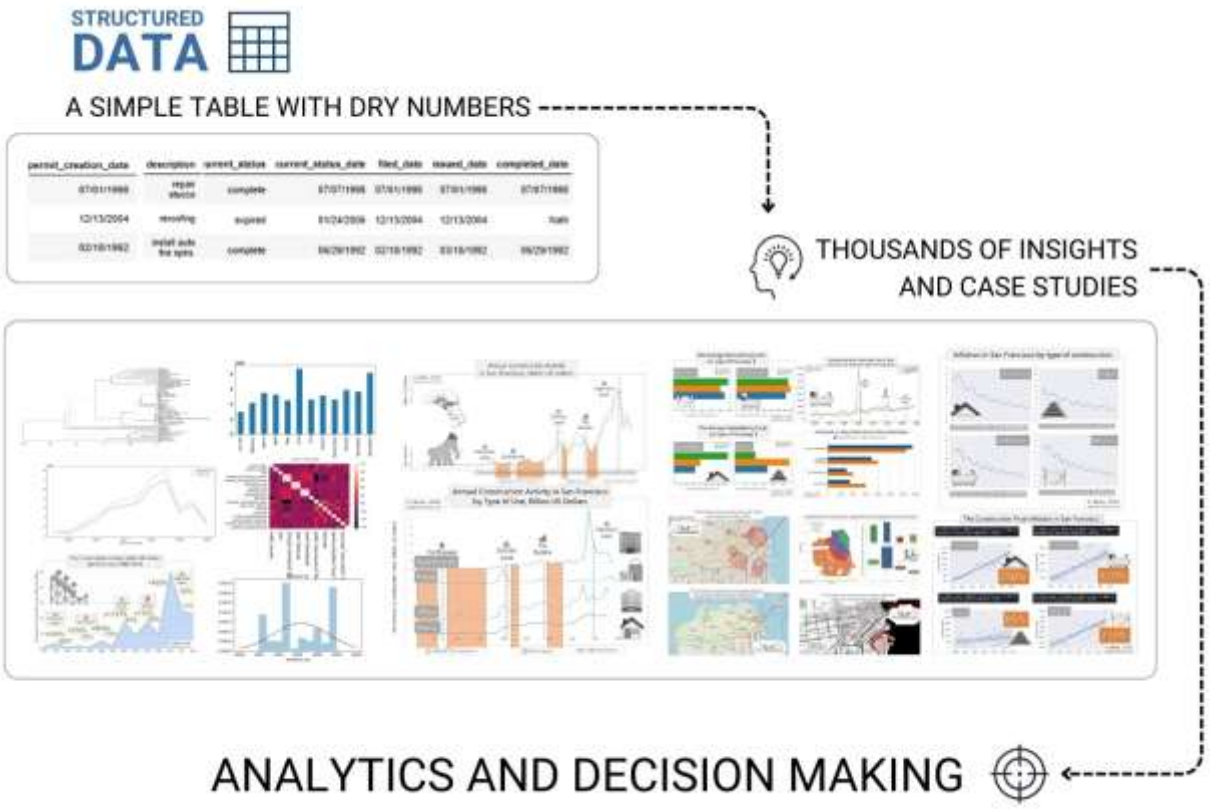


図 9.1-8 視覚的に理解しやすいデータへの移行は、隠れたパターンを認識することによる自動意思決定を可能にする。

サンフランシスコ建築局」の表形式データセットのデータを分析したように、画像や文書から IoT データ 、派生 CAD データベース まで、あらゆるデータセットを視覚化して分析することができる。

CAD データに基づくビッグデータの例 (BIM)

以下の例では、さまざまな CAD ツール （BIM ）からのデータを使用して、大規模なデータセットを分析する。

この大規模なデータセットを収集・作成するために、専用の自動ウェブクローラ（スクリプト）を使用し、RVT および IFC 形式で無料の建築モデルを提供するウェブサイトから設計ファイルを自動的に検索・収集するように設定した。数日間で、クローラーは 4,596 の IFC ファイルと 6,471 の RVT ファイル、156,024 の DWG ファイルを発見し、ダウンロードすることに成功した[149]。

RVT および IFC フォーマットの異なるバージョンのプロジェクトを収集し、無料の リバースエンジニアリング SDK を使用して構造化 CSV フォーマット に変換した後、ほぼ 1 万件の RVT および IFC プロジェクトを 1 つの大きな Apache Parquet テーブルファイル に収集し、分析のために Pandas DataFrame にアップロードした（図 9.1-9）。

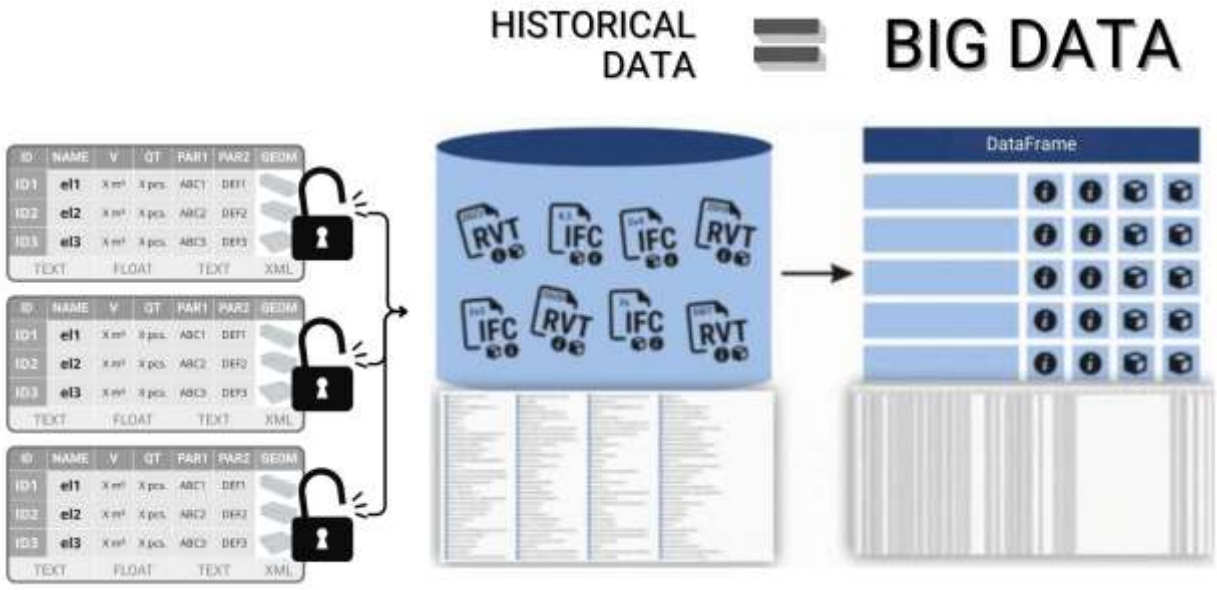


図 9.1-9 構造化データ プロジェクト・データでは、任意の数のプロジェクトを 1 つの 2 次元表にまとめることができます。

この大規模なコレクションからのデータには、以下の情報が含まれている：IFC ファイルセット 、約 400 万のエンティティ（行）と 24,962 の属性（列）が含まれ、RVT ファイルセット 、約 600 万のエンティティ（行）で構成され、27,025 の異なる属性（列）が含まれている。

これらの情報セット（図 9.1-10）は、何百万もの要素をカバーし、そのそれぞれについて、バウンディングボックス ジオメトリ（プロジェクト内のオブジェクトの境界を定義する長方形）の座標を追加で取得し、共通のテーブルに追加した。バウンディングボックス ジオメトリ（プロジェクト内のオブジェクトの境界を定義する長方形）の座標と、各要素の PNG 形式の画像、およびジオメトリのオープン XML 形式 - DAE （Collada） が作成された。

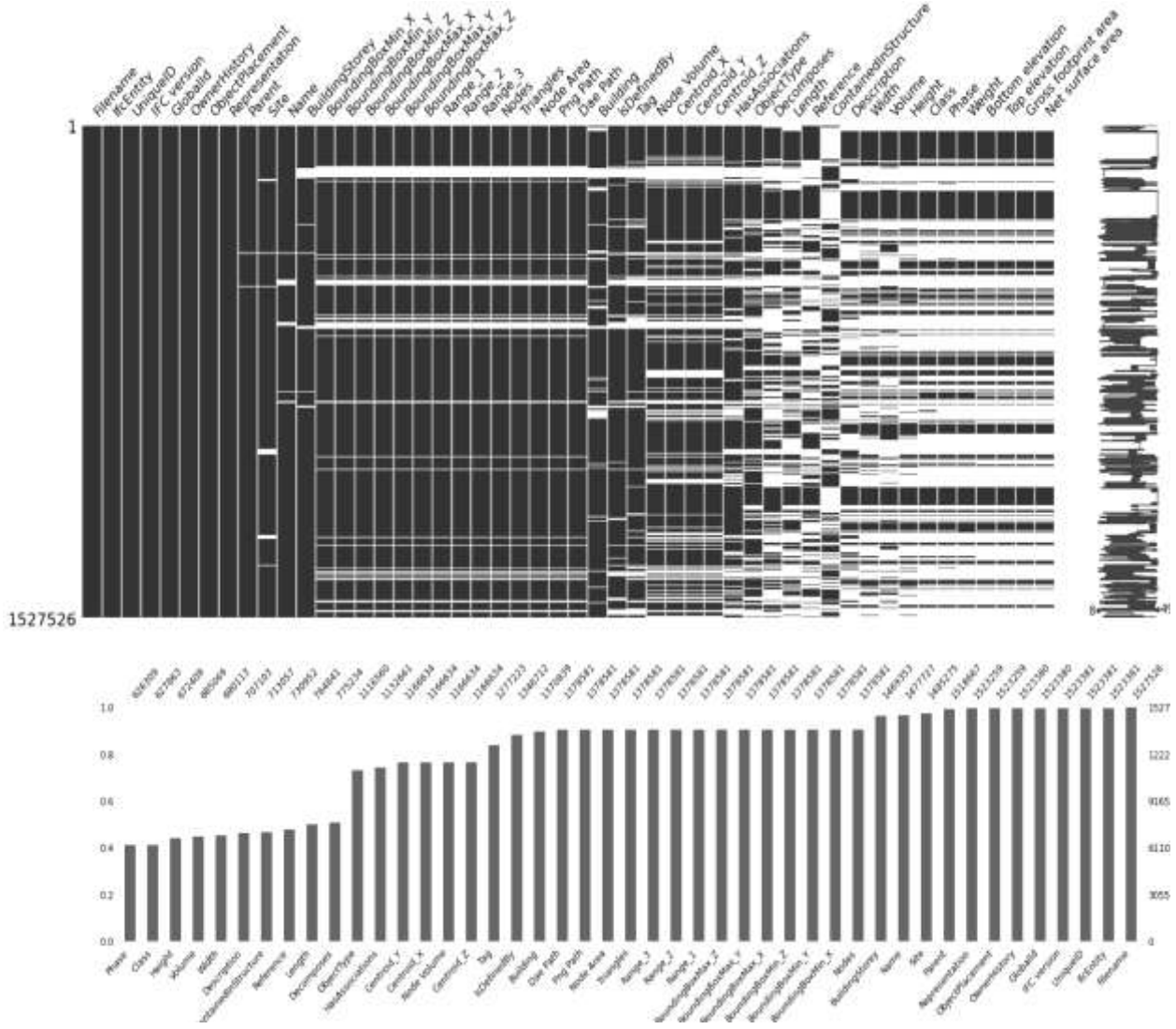


図 9.1-10 150 万要素のサブセットと、最初の 100 属性の占有率をヒストグラムとして視覚化（missingno library）。

そこで、4,596 の IFC プロジェクト（）と 6,471 の RVT プロジェクト（）から、数千万の要素に関するすべての情報を入手しました。このプロジェクトでは、すべてのエンティティ要素の属性-プロパティとそのジオメトリ（バウンディングボックス ）が、構造化された単一のテーブル（データフレーム ）に変換されました（図 9.1-10-データフレーム集団のデータはヒストグラムとして表示されます）。

分析プロセス中にプロットされるヒストグラム（図 9.1-10、図 9.2-6、図 9.2-7）により、列内のデータ密度と値の出現頻度を素早く評価することができる。これにより、特徴の分布、外れ値の存在、分析および機械学習モデル構築における個々の属性の潜在的な有用性についての最初の洞察が得られる。

このデータセット（図 9.1-10）の実用的な使用例として、"5000 IFC and RVT " プロジェクトがある。[149]、Kaggle プラットフォーム で利用可能である。このプロジェクトでは、Jupyter Notebook に、データの前処理と分析から、pandas、 matplotlib、seaborn、folium などの Python ライブラリを使った結果の視覚化まで、完全なパイプライン・ソリューションを提 供しています（図 9.1-11）。

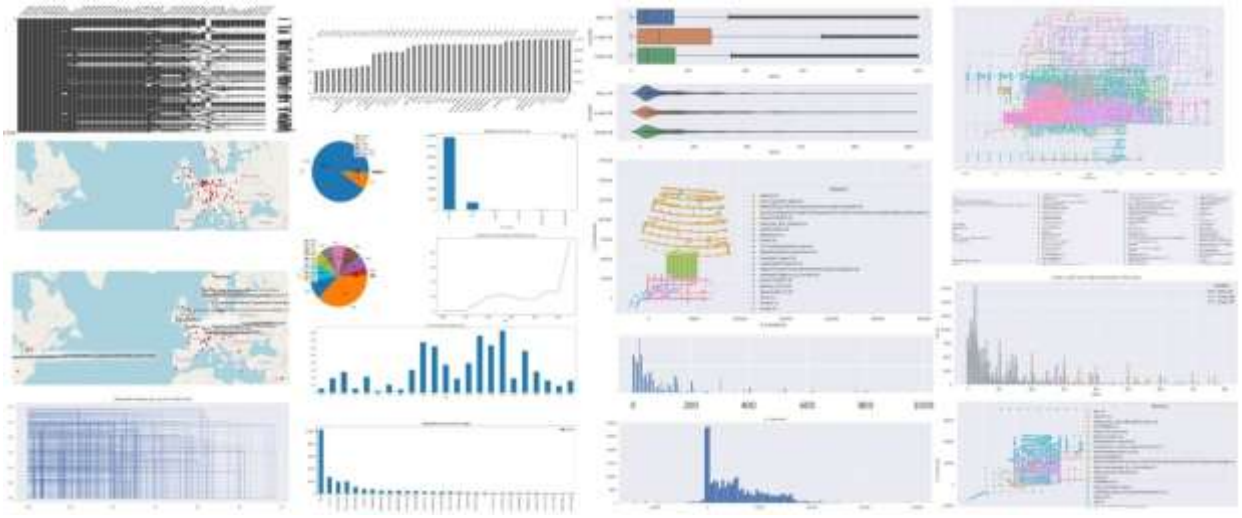


図 9.1-11 Python の可視化ライブラリと pandas ライブラリを使用した CAD フォーマット （ BIM ）からのデータ分析の例。

メタ情報に基づいて、特定のプロジェクトがどの都市で開発されたかを判断し、地図上に表示することが可能です（例：folium ライブラリを使用）。さらに、データ内のタイムスタンプにより、曜日別、時間帯別、月別など、ファイルが保存または編集された時間帯のパターンを調べることができます。

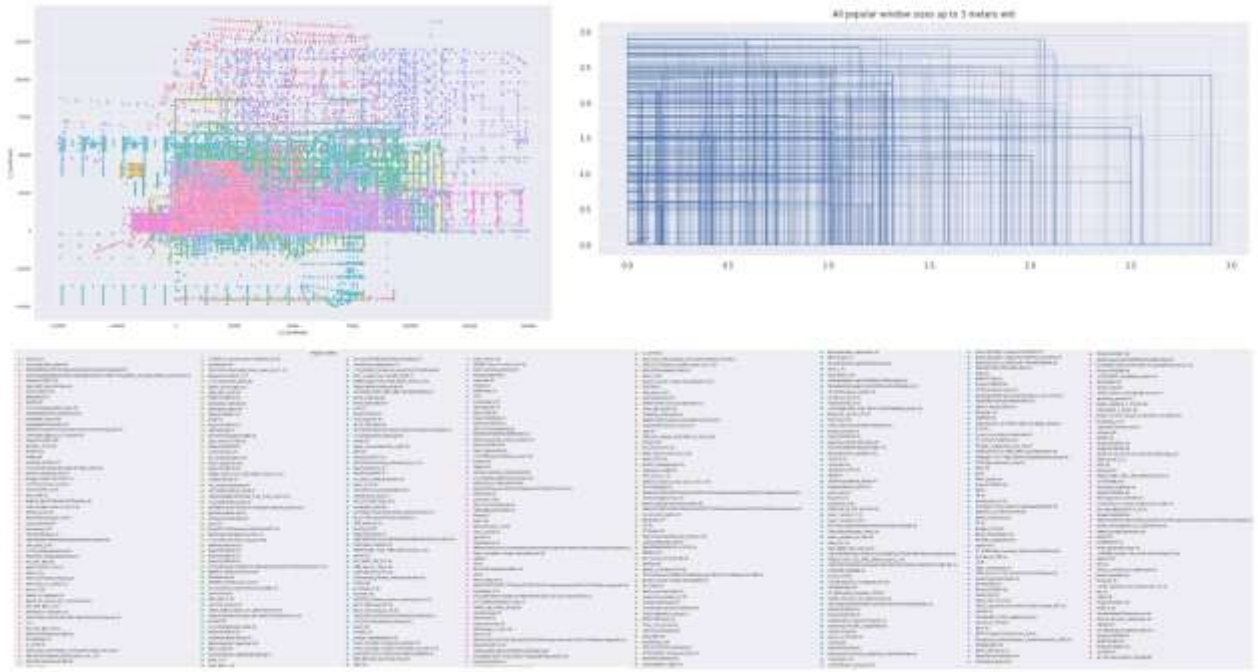


図 9.1-12 視覚化 チャート下部のリストから、すべての柱の幾何学的位置と、3 メートルまでの
すべての窓の寸法を表示。

モデルから抽出されたバウンディングボックス形式の幾何学パラメータは、集計分析にも適している。例えば、図 9.1-12 は、2 つのグラフを示している。左のグラフは、ゼロ点に対する全プロジェクト の柱間距離の分布を示し、右のグラフは、数万個の窓要素のサンプルにおける高さ 3m までの全窓の寸法を示 している（全データセットをパラメーター "Category " の値 "OST_Windows"、"IfcWindows " でグループ化した後）。

この例の Pipeline 解析コード とデータセット自体は、Kaggle のウェブサイト "5000 IFC and RVT | DataDrivenCo-nstruction.io projects" [149]で公開されています。この完成した Pipeline とデータセットは、コピーして Kaggle で無料でオンライン実行するか、人気のある IDE の 1 つでオフラインで実行することができます: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse with PyDev plugin, Thonny, Wing IDE, IntelliJ IDEA with Python plugin, JupyterLab または人気のあるオンラインツール Kaggle.com、Google Collab、Microsoft Azure Notebooks、Amazon SageMaker。

> 膨大な量の構造化データの処理と研究から得られる分析的洞察は、建設業界の意思決定プロセスにおいて重要な役割を果たすだろう。

過去のプロジェクトに基づくこの種の情報分析により、専門家は、例えば材料や労働力の必要量を効果的に予測し、建設開始前に設計ソリューションを最適化することができる

しかし、設計データや建築許可証が比較的ゆっくりと変化する静的な情報であるのに対し、建設プロセスそのものは、さまざまなセンサーや IoT - デバイス（カメラ、リアルタイムでデータを送信する自動監視システム）で急速に飽和しつつある。これらすべてが、建設現場をリアルタイムでデータを分析する必要がある動的なデジタル環境に変える。

IoT モノのインターネットとスマートコントラクト

IoT モノのインターネット（Internet of Things）は、あらゆるデバイスが独自の IP アドレスを取得し、グローバル・ネットワークの一部となるデジタル変革の新しい波を象徴している。IoT とは、物理的なモノをインターネットに接続し、データを収集、処理、送信する概念である。建設業界では、建設プロセスをリアルタイムで制御し、資材の浪費を最小限に抑え、設備の消耗を予測し、意思決定を自動化する能力を意味する。

> CFMA の記事 "Preparing for the Future with Connected Construction"[150]によると、建設業界は今後 10 年で大きなデジタル変革を遂げる。[150]によると、建設業界は今後 10 年で大きなデジタル変革を遂げ、コネクテッド・コンストラクション（完全に統合され自動化された建設現場）というコンセプトに結実する。

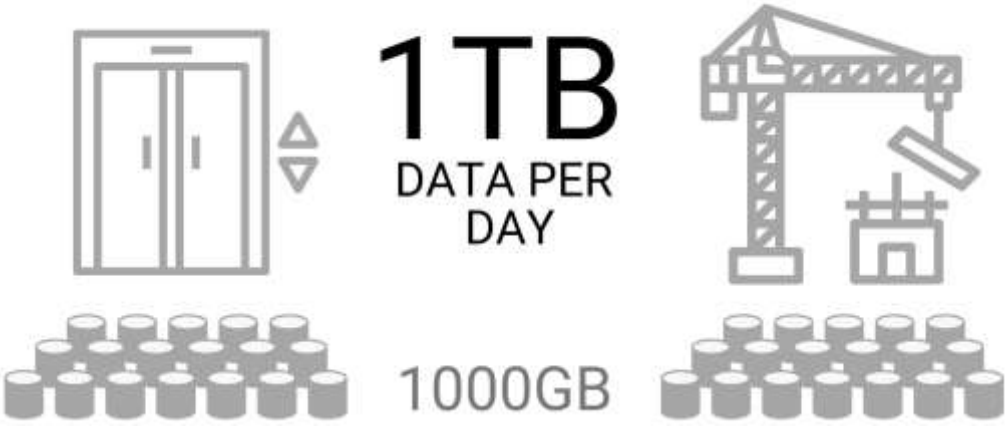


図 9.1-13 IoT や建設現場のデータ機器は、1 日にテラバイト単位のデータを生成し、送信することができる。

デジタル建設現場とは、固定カメラやクアッドロコプターを使った建設現場での計画やロジスティクスから、作業の実施や品質管理まで、建設に関わるすべての要素が単一のダイナミックなデジタル・エコシステムに統合されることを意味する。本書の第 7 部では、さまざまなソースからの収集からストレージや分析プラットフォームへの転送まで、リアルタイムのデータストリーミングを可能にする無償のオープンソースツール、

Apache NiFi（図 7.4-5）の機能をすでに見てきた。

工事の進捗状況、資材の消費量、設備の状態、安全性に関するデータは、分析シス テムにリアルタイムで送信される（図 9.1-13）。これにより、潜在的なリスクを予測し、逸脱に迅速に対応し、現場工程を最適化することができる。デジタル建設現場の主な構成要素には以下が含まれる：

- IoT -センサー - 環境パラメータの追跡、建設機械の監視、労働条件の管理。
- デジタルツイン - 建物やインフラの仮想モデルで、起こりうる逸脱を予測し、エラーを防ぐ。
- 自動物流システム - リアルタイムのサプライチェーン管理により、ダウンタイムとコストを削減。
- ロボット建設システム - 日常的な作業や危険な作業を行うための自律型機械の使用。

ロボット化 、IoT の 、コネクテッド・サイト（建設）デジタル建設現場コンセプトは、効率性を高め、コストを削減するだけでなく、安全性、持続可能な建設、予測可能なプロジェクト管理の新時代を切り開くだろう。

> RFID （Radio Frequency Identification）タグもまた、IoT の最も重要なコンポーネントのひとつである。RFID タグは、建設現場における資材、機械、さらには人員の識別と追跡に使用され、プロジェクト・リソースの透明性と管理を向上させる。

RFID - 技術は、無線信号を使用して物体を自動的に認識するために使用される。RFID は 3 つの重要な要素で構成されている：

- RFID -タグ（パッシブまたはアクティブ） - 一意の識別子を含み、材料、工具、機械に取り付けられる。
- スキャナーはタグから情報を読み取り、システムに送信する装置である。
- 集中データベース - 物体の位置、状態、移動に関する情報を保存。

建設業における RFID の応用 ：

- 自動材料計算 - 生コンクリート製品、補強材、サンドイッチパネルのパッケージにタグを付けることで、在庫を管理し、盗難を防ぐことができます。
- 人員の勤務管理 - RFID - 従業員バッジは、シフトの開始時刻と終了時刻を記録し、勤務時間の記録を提供する。
- 機器のモニタリング - RFID - システムは機器の動きを追跡し、ダウンタイムを防ぎ、物流効率を向上させます。

この技術群を補完するのが、仲介者を介さずに支払い、配送管理、契約順守を自動化し、詐欺や遅延のリスクを軽減するブロックチェーン・ベースのスマート・コントラクトである。

> 今日、共通のデータモデルがないため、スマートコントラクトは参加者が合意するコードに過ぎない。しかし、データ中心アプローチでは、契約パラメータの共通モデルを作成し、ブロックチェーンにエンコードし、条件の履行を自動化することができる。

例えば、サプライチェーン管理システムにおいて、スマート契約は IoT -センサーや RFID -タグから貨物の配送を追跡し、貨物が到着すると自動的に支払いを振り込むことができる。同様に、建設現場では、スマートコントラクトが、ドローンや建設センサーからのデータに基づいて、鉄筋の設置や基礎の打設などの作業段階の完了を記録し、手作業によるチェックや紙の証明書を必要とせずに、請負業者への次の支払いを自動的に開始することができる。

しかし、新しい技術や国際標準化団体の努力にもかかわらず、競合する規格の多さが IoT の状況を複雑にしている 。

> 2017 年に発表されたシスコの調査[151]によると、モノのインターネット構想（IoT ）のほぼ 60%は概念実証の段階で止まっており、IoT プロジェクトが完全に成功したと考えている企業はわずか 26 %である。さらに、完了したプロジェクトの 3 分の 1 は、掲げた目標を達成できず、実施後も成功したとは認識されていない。

主な理由のひとつは、異なるセンサーからのデータを処理するプラットフォーム間の相互運用性の欠如である。その結果、データは別々のソリューションの中で孤立したままになっている。本書で取り上げた）他の同様のケースと同様に、このアプローチに代わるものは、データそのものを主要な資産として構築されたアーキテクチャである。

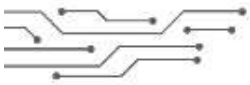
> IoT センサーは、設備の技術的な状態を監視するだけでなく、故障や逸脱を予測することで、建設現場でのリスクを軽減し、プロセス全体のパフォーマンスを向上させる予測分析においても重要な役割を果たしている。

IoT センサーや RFID タグによって収集されたデータは、機械学習アルゴリズムによってリアルタイムで処理され、異常を検知してエンジニアに潜在的な不具合を事前に警告することができる。これは、コンクリート構造物の微小なひび割れから、タワークレーンの運転における特徴的な一時停止まで、技術的な不具合や規制違反

を示すものである。さらに、高度な行動分析アルゴリズムにより、例えば作業員の肉体的疲労を示す行動パターンを捉えることができ、現場の安全性と従業員の福利厚生を事前に管理することができます。

建設業界では、機械であれ人であれ、事故や故障が突然起こることはほとんどない。通常、事故や故障の前には些細な逸脱があり、それが気づかれないことが多い。予測分析 ）と機械学習は、重大な結果が発生する前であっても、早い段階でこれらのシグナルを検出することを可能にします。

文書、プロジェクトファイル、IoT デバイスや RFID タグからのデータが建設プロジェクトのデジタルフットプリントを形成する一方で、機械学習はそこから有用な洞察を抽出するのに役立ちます。データの増大とデータアクセスの民主化により、建設業界はアナリティクス、予測分析、人工知能アプリケーションの新たな機会を獲得している。



章 9.2.

機械学習と予測

機械学習 と人工知能は、我々の建築方法を変えるだろう。

建設ビジネスにおけるさまざまなシステムのデータベースは、その腐敗が避けられず、ますます複雑化するインフラとともに、未来のソリューションの温床となりつつある。会社のサーバーは、まるで森のように、重要な情報のバイオマスに富んでおり、多くの場合、フォルダーやサーバーの地下に隠されている。現在作成されている様々なシステムから得られる大量のデータは、使用後、サーバーの底に落ちた後、何年も化石化した後、将来、機械学習 や言語モデルの燃料となるだろう。社内のチャット・ルーム（例えば、ローカルに設定された ChatGPT 、LlaMa 、Mistral 、DeepSeek の別インスタンス）は、集中ストレージを使用して、これらの社内モデル上に構築され、迅速かつ便利に情報を取得し、必要なグラフ、ダッシュボード、ドキュメントを生成する。

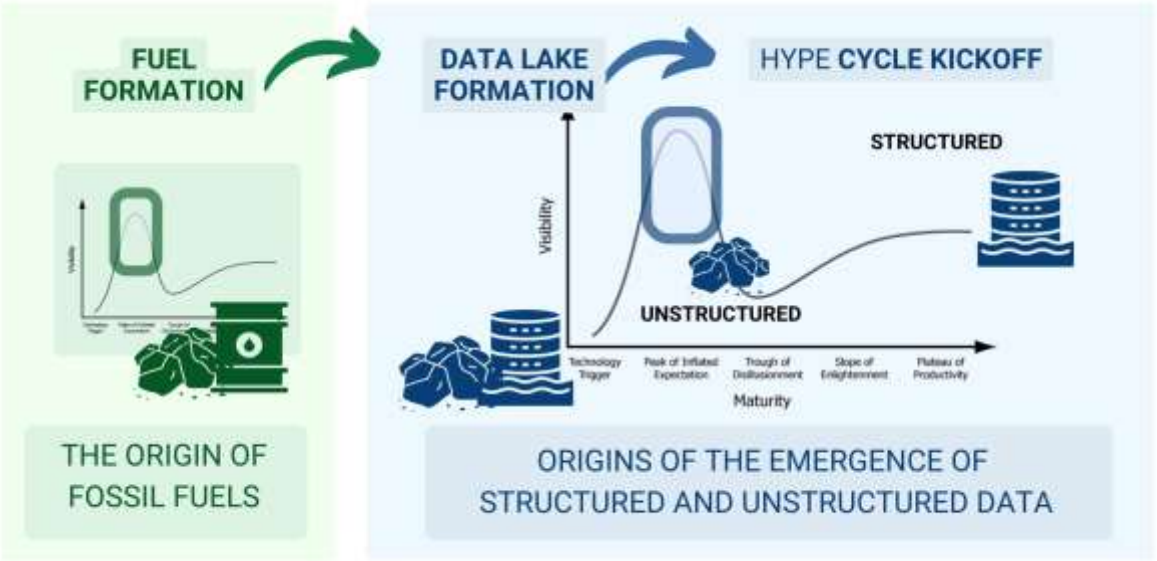


図 9.2-1 木が石炭に変わるように、情報もまた、時間と分析のプレッシャーの下で、時間の経過とともに価値あるビジネス・エネルギーに変わる。

圧力と温度とが組み合わさった植物塊の化石化は、異なる時代に生きていた異なる 種の樹木の、均質で独自に構造化された均質な塊、すなわち木炭を作り出す [152]。同じように、分析部門の圧力と品質管理の温度のもとで、異なるフォーマットで異なる時期にハードディスクに記録された情報は、最終的に価値ある情報の均質な

構造化された塊を形成する（図 9.2-1）。

このような情報の層（またはより多くの場合、孤立したナゲット）は、経験豊富なアナリストが丹念にデータを整理することによって作られる。

これらの成熟したデータレイヤーが、もはやレポートの中で「焼かれる」だけでなく、ビジネスプロセスを循環し始め、意思決定を豊かにし、プロセスを改善するようになった瞬間、企業は次のステップ、つまり機械学習と人工知能への移行の準備が整う（図 9.2-2）。

> 機械学習 (ML - Machine learning) は、人工知能の問題を解くための手法のクラスである 。機械学習アルゴリズム は、大規模なデータセットのパターンを認識し、それを使って自己学習する。新しいデータセットが増えるたびに、数学的アルゴリズムは得られた情報に従って改善・適応し、推奨や予測の精度を常に向上させることができる。

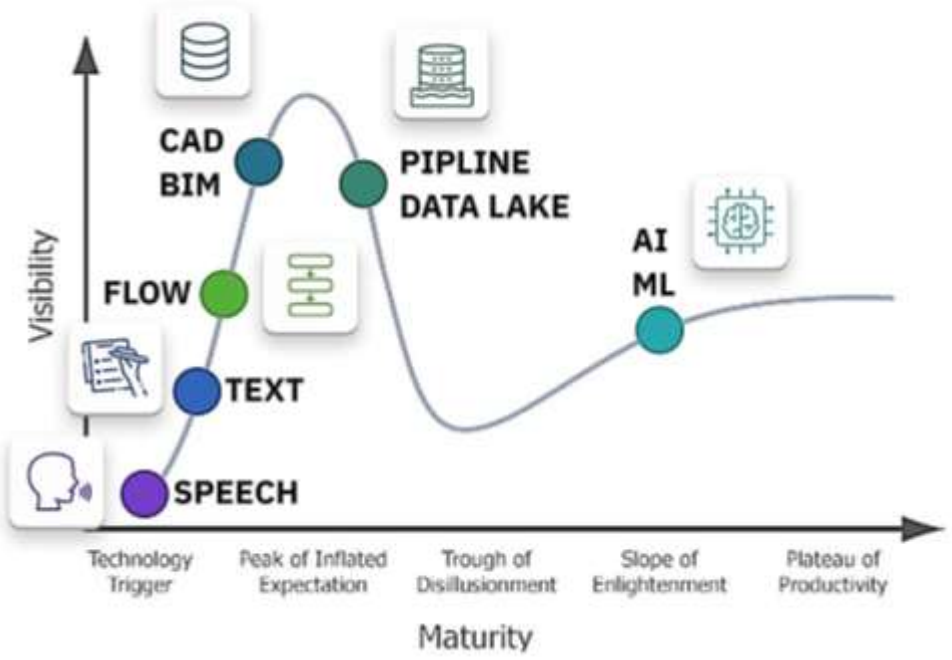


図 9.2-2 データ作成技術の衰退と分析ツールの応用 機械学習の話題への扉を開く 。

世界最大の投資ファンド（世界最大の建設ソフトウェア企業のほとんどすべてと、世界最大の不動産を所有する企業の主要株式を所有している[55]）の影響力のある CEO は、2023 年のインタビューで次のように述べている。- 機械学習は建設の世界を変えるだろう。

> AI は巨大な可能性を秘めている。それは私たちの働き方、生き方を変えるだろう。AI とロボティクスは、私たちの働き方や建築方法を変えるだろうし、はるかに大きな生産性を生み出す手段として AI とロボティクスを使うことができるようになるだろう[153]。
>
> - 世界最大の投資ファンドの CEO、インタビュー、2023 年 9 月。

機械学習 (ML)は、人間の思考を模倣する統計的手法を使って、大量のデータを処理することで機能する。しかし、ほとんどの企業はそのようなデータセットを持っておらず、持っていたとしても十分にラベル付けされていないことが多い。そこで、セマンティック技術と転移学習（ML が少量のデータを扱うときに、より効果的になる技術）が役に立つ。

トランスファー・ラーニングの本質は、各作業をゼロから学ぶのではなく、関連分野で得た知識を活用することである。他業界のパターンや発見を建設業界に適応・応用できることを認識する必要がある。例えば、小売業で開発された物流プロセスの最適化手法は、建設業のサプライチェーン・マネジメントの効率化に役立つ。金融業界で積極的に活用されているビッグデータ分析は、建設プロジェクトのコスト予測やリスク管理に応用できる。また、産業界で開発されているコンピューター・ビジョンやロボット工学の技術は、自動化された品質管理、安全監視、建設現場の設備管理にすでに応用されている。

> トランスファー学習は、イノベーションの導入を加速させるだけでなく、他産業ですでに蓄積された経験を活用することで、その開発コストを削減することも可能にする。



図 9.2-3 人工知能 技術とロボティクスは、建設業界の生産性を向上させる将来の主な原動力となるだろう。

人間の思考も似たような原理で組織化されている。新しい問題を解決するために、以前に獲得した知識をベースにするのである（図 4.4-19、図 4.4-20、図 4.4-21）。機械学習においても、このアプローチは有効です。データモデルを単純化し、よりエレガントにすることで、ML アルゴリズムにとって問題の複雑さを軽減することができます。その結果、大量のデータの必要性が減り、計算コストが削減される。

主観的評価から統計的予測へ

戦略的意思決定がマネジャー個人の直感に依存していた時代（図 9.2-4）は過去のものとなった。競争が激化し、厳しい経済環境の中で、主観的なアプローチはリスクが高すぎ、非効率になりつつある。 データ を客観的に分析する代わりに、個人の意見に依存し続ける企業は、変化に迅速に対応する能力を失う。

> 競争環境は、データ、統計的パターン、計算可能な確率に基づく正確性と再現性を求めている。意思決定はもはやフィーリングに基づくことはできず、アナリティクス（）や機械学習（）から得られる相関関係、傾向、予測モデルに基づいて行わなければならない。思い込みから証拠へ、主観的な確率から統計的に計算された偏差へ、感覚から事実へ。



図 9.2-4 HiPPO （最も高給取りの従業員の意見）によって意思決定される時代は、ビッグデータと機械学習の出現によって過去のものとなるだろう。

これまで自分の感覚だけに頼っていたマネジャーは、必然的に新しい現実に直面することになる。今やマネジメントの中心には、何百万ものパラメーターやベクトルを分析し、隠れたパターンを特定して最適な戦略を提案するシステムがある。

今日でも企業が ML の導入を避ける主な理由は、その透明性の欠如にある。ほとんどのモデルは、 、具体的にどのように結論に至るのかを説明することなく、管理者にとって「ブラックボックス」として働いている。これは問題につながる：アルゴリズムは固定観念を強化し、すぐに有害なコミュニケーションツールと化したマイクロソフトのチャットボットのケースのように、ユーモラスな状況を作り出すことさえある[154]。

> ディープ・シンキング』の中で、チェスの元世界チャンピオン、ガルリ・カスパロフは、IBM のビッグ・ブルー・コンピューターに敗れたことを振り返っている[155]。彼は、AI（）の真の価値は人間の知能をコピーすることではなく、人間の能力を補完することにあると主張する。AI は人間が苦手とするタスクをこなし、人間は創造性をもたらすべきだ。コンピューターはチェスを分析する従来のアプローチを変えた。コンピュータのチェスプログラムは、対局に関する魅力的なストーリーを作り出す代わりに、実際の強弱のみに基づいて、各手を公平に評価する。カスパロフは、個々の行動ではなく、出来事を首尾一貫したストーリーとしてとらえる人間の傾向は、チェスだけでなく、人生全般において、しばしば誤った結論を導くと指摘する。

したがって、機械学習を予測や分析に利用するつもりなら、機械学習ツール や AI を仕事で使い始める前に、アルゴリズムの仕組みやデータの処理方法など、その基本原理を理解しておくことが重要だ。始めるための最良の方法は、実地体験である。

機械学習 と予測のトピックの最初の導入に最も便利なツールの 1 つは、Jupyter Notebook 、人気のある古典的なタイタニックデータセットである。これは、データ分析と ML モデル構築の主要な方法を視覚的に紹介するものである。

タイタニック・データセット：アナリティクスの世界でハロー・ワールド データとビッグデータ

ML をデータ分析に使用した最も有名な例のひとつに、タイタニック号のデータセットの分析がある。この表を学ぶことは、プログラミング言語を学ぶときの「Hello World」プログラムに似ている。

1912 年に RMS タイタニック号が沈没し、2224 人中 1502 人が死亡した。タイタニック号のデータセットには、乗客が生存したかどうかの情報だけでなく、年齢、性別、チケットのクラスなどの属性も含まれている。こ

のデータセットは無料で利用可能で、さまざまなオフライン・オンライン・プラットフォームで開いて分析することができる。

> タイタニックデータセットへのリンク：
> https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv

LLM-enabled IDEs and future programming changes " の章では、Jupyter Notebook - データ分析と機械学習のための最もポピュラーな開発環境の 1 つ について説明した。Jupyter Notebook の無料のクラウド類似サービスとして、Kaggle と Google Collab プラットフォームがある。これらはソフトウェアをインストールすることなく Python コードを実行でき、コンピューティングリソースに無料でアクセスできる。

> Kaggle は、統合されたコード実行環境を持つ最大のデータ分析、機械学習コンペティションプラットフォームである。2023 年 10 月現在、Kaggle には 194 カ国から 1500 万人以上のユーザー[156]が参加している。

Kaggle プラットフォーム（図 9.2-5）でタイタニックのデータセットをダウンロードして使用することで、データセット（のコピー）を保存し、専用の IDE をインストールすることなく、プリインストールされたライブラリを使用した Python コードをブラウザで直接実行することができます。

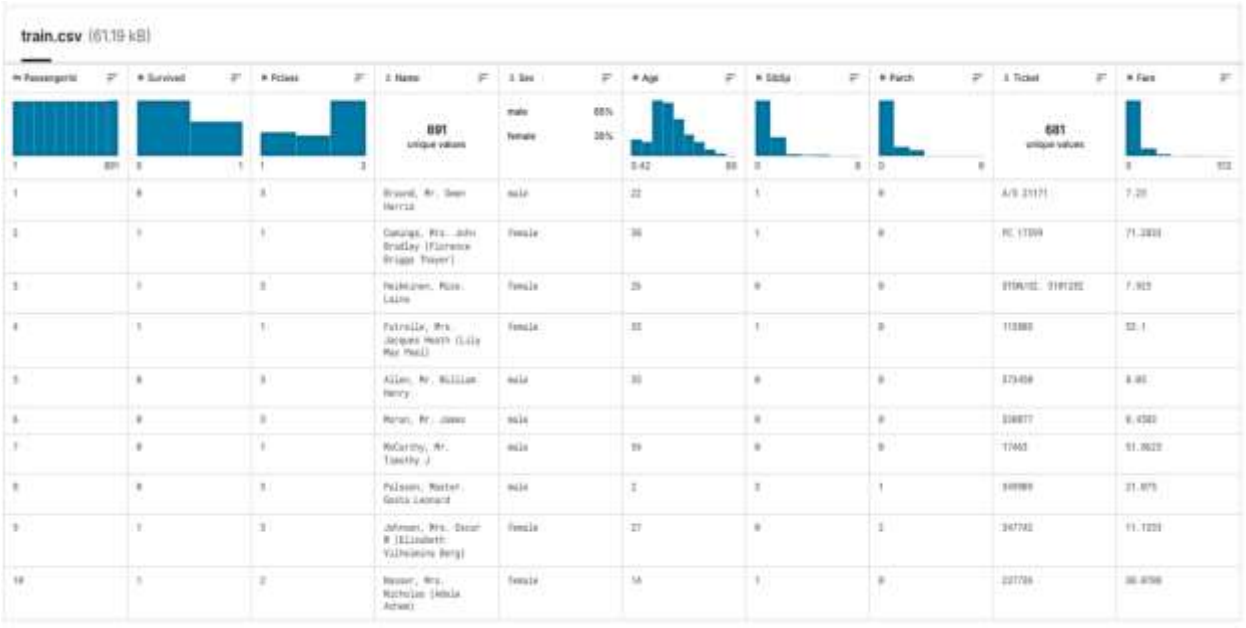


図 9.2-5 タイタニック・テーブルの統計 - データ分析および機械学習の学習用として最も一般的な学習データセット 。

Titanic データセットには、1912 年に難破した RMS タイタニック号の乗客 2,224 人のデータが含まれている。このデータセットは、トレーニング用（train.csv）とテスト用（test.csv）の 2 つの独立したテーブルとして提供されており、モデルのトレーニングと新しいデータでの精度評価の両方に使用することができます。

訓練データセットには、乗客の属性（年齢、性別、チケットクラス、その他）と、誰が生き残ったかという情報（バイナリ値 "Survived "の列）の両方が含まれている。訓練データセット（図 9.2-6 - train.csv ファイル）はモデルの訓練に使われる。テストデータセット（図 9.2-7-ファイル test.csv）には、生存者情報のない乗客属性のみが含まれる（「Survivor」という列が 1 つもない）。テストデータセットは、新しいデータでモデルをテストし、その精度を評価するためのものです。

このように、トレーニングデータセットとテストデータセットの乗客の属性はほぼ同じである。唯一の重要な違いは、テストデータセットには、我々が様々な数学的アルゴリズムを用いて予測することを学習したいターゲット変数である "Survivor "列を持たない乗客のリストがあるということです。そして、モデルを構築した後、テストデータセットからの実際のパラメータ "Survivor "とモデルの出力を比較することができます。

表の主な列、トレーニングデータセットとテストデータセットの乗客パラメータ：

- PassengerId - 一意の乗客識別子
- 生存 - 乗客が生存していれば 1、死亡していれば 0（テストセットでは使用不可）

- Pclass - チケットクラス（1、2 または 3）
- 氏名 - 同乗者の氏名
- 性別 - 乗客の性別（男性／女性）
- 年齢
- 兄弟姉妹 - 兄弟姉妹または配偶者の人数
- パーチ - 乗船している両親または子供の数
- チケット - チケット番号
- 運賃 - チケット料金
- キャビン - キャビン番号（多くのデータが欠落している）
- Embarked は乗船港（C＝シェルブール、Q＝クイーンズタウン、S＝サウサンプトン）。

両表の欠損データを視覚化するには、missingno ライブラリ（図 9.2-6、図 9.2-7）を使用することができる。これは欠損値をヒストグラムの形で表示し、白いフィールドが欠損データを示す。この視覚化により、処理前のデータ品質を素早く評価することができます。



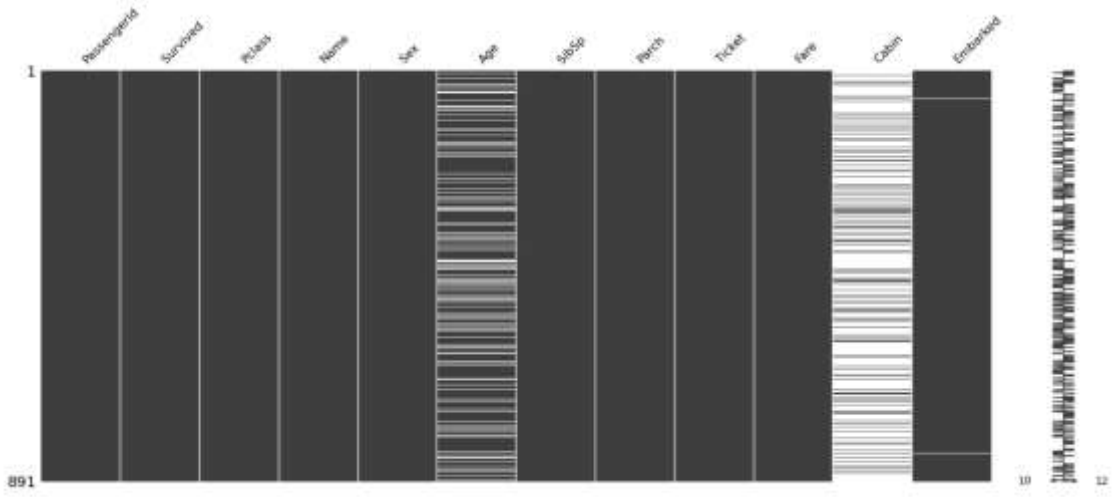


図 9.2-6 数行のコードを用いて、タイタニックのトレーニング・データセットにおける欠損データを視覚化する。

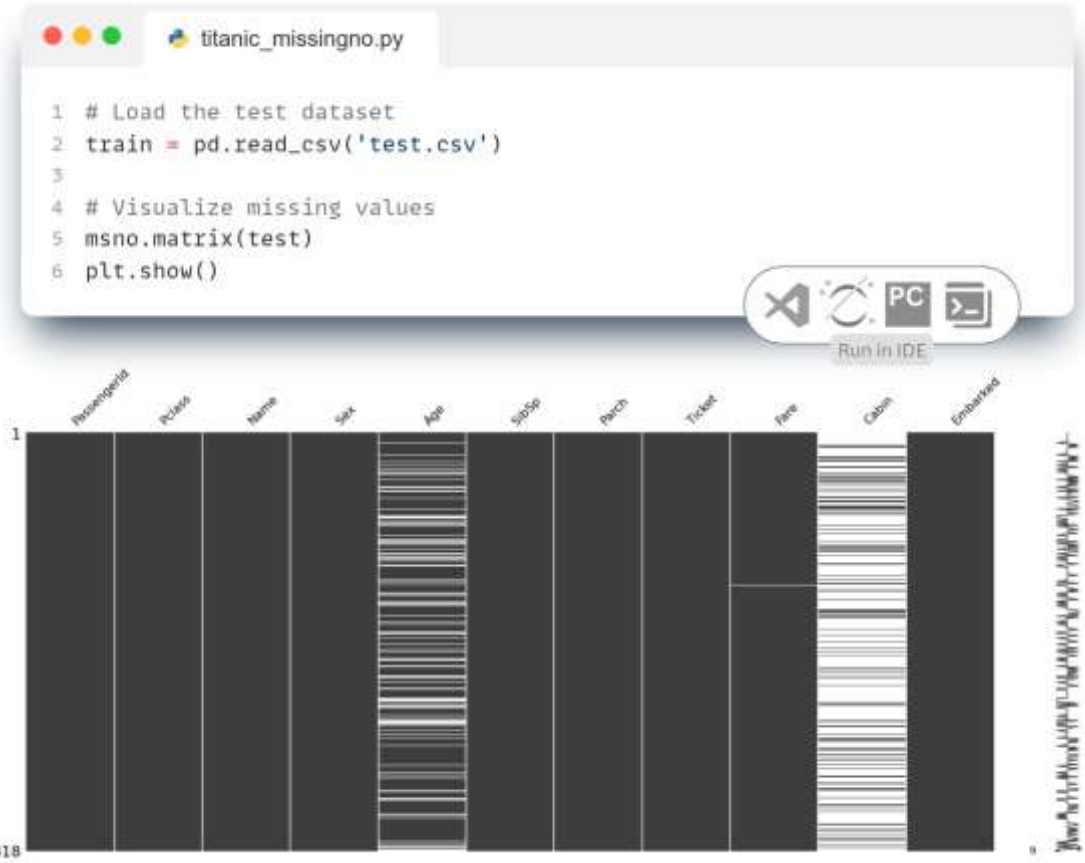


図 9.2-7 可視化 Titanic テストデータセットにおける欠損データの表示。このデータセットには
、情報のない乗客の特徴のみが含まれている。

データセットに基づいて仮説を立てたり予測を立てたりする前に、視覚的分析によってデータの主要なパターンを特定し、その質を評価し、可能性のある依存関係を特定する。タイタニック号のデータセットの理解を深めるのに役立つ可視化テクニックはたくさんあります。乗客の年齢層を分析する分布図、性別や階級別の生存チャート、欠損データ行列などを使って、情報の質を評価し、データを理解することができます。

- LLM モデル（CHATGP、LlaMa 、Mistral 、DeepSeek 、Grok 、Claude 、QWEN 、その他）に以下のテキストリクエストを送信することで、LLM 、タイタニックデータセットのデータを視覚化する手助けをしてもらおう：

  > タイタニックのデータセットについて、簡単なグラフを示してください。データセットを自分でダウンロードして，ください．

  - LLM レスポンス データセットのパラメーターを視覚化した既製のコードとグラフの形

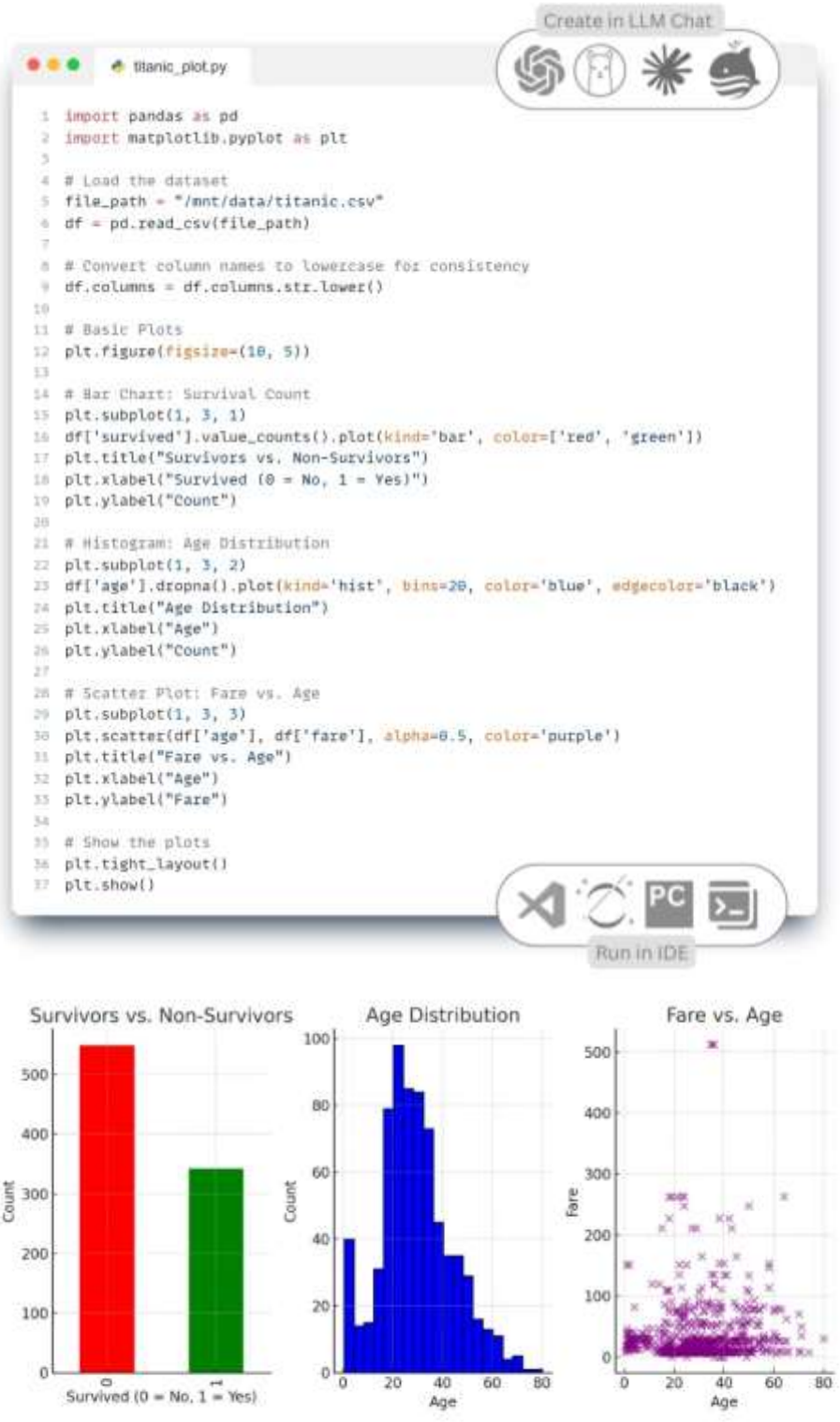


図 9.2-8 LLM データセットのデータを即座に視覚化することができます。

データの可視化は、その後の機械学習モデル 構築のためにデータセットを準備するための重要なステップである。

機械学習の実例: タイタニックの乗客からプロジェクト管理まで

タイタニック号のデータセットに基づき、機械学習フレームワーク を探求するために使用された主な仮説は、特定の乗客グループは生存の可能性が高いというものである。

> タイタニック号の乗客の小テーブルは世界中で人気があり、何百万人もの人々が、タイタニック号の乗客のトレーニングデータセットに基づいて、どのアルゴリズムや仮説が最も正確な生存予測モデルを構築するかを調べるためのトレーニング、実験、モデルテストに使用している。

タイタニックのデータセットの魅力は、そのコンパクトさにある。数百行と 12 列のデータセット（図 9.2-6）は、分析のための十分な機会を与えてくれる。このデータセットは、比較的単純に、問題の目標である生存が 0 か 1 かという便利な形式で表現される、バイナリ分類解 の古典的な例である。

> ジョン・ホイーラーは "It from Bit " の中で[7]では、宇宙は二進法の選択に基づいていると論じている。同様に、分子で構成される人々によって運営されるビジネスも、実際には一連の二進法の二者択一の上に成り立っている。

加えて、データは実際の歴史的な出来事に基づいているため、人工的に作成された例とは異なり、研究にとって価値がある。Kaggle プラットフォーム（ ）だけでも、最大級のデータパイプライン（ ）と ETL（） の 1 つであるタイタニック（Titanic）データセットに基づく課題に 1,355,998 人が参加し、53,963 のユニークなデータパイプラインソリューショ ンを開発した[157]（図 9.2-9）。

> 信じられないような話だが、タイタニック号の乗客に関するたった 1000 行のデータと 12 のパラメータが、何百万もの仮説、論理的連鎖、ユニークなデータ・パイプラインのフィールドとなった。単純な生存モデルから、隠れたパターンや複雑な推論の迷宮を考慮に入れた複雑なアンサンブルまで、小さなデータセットから無限の洞察、仮説、解釈が生まれる。

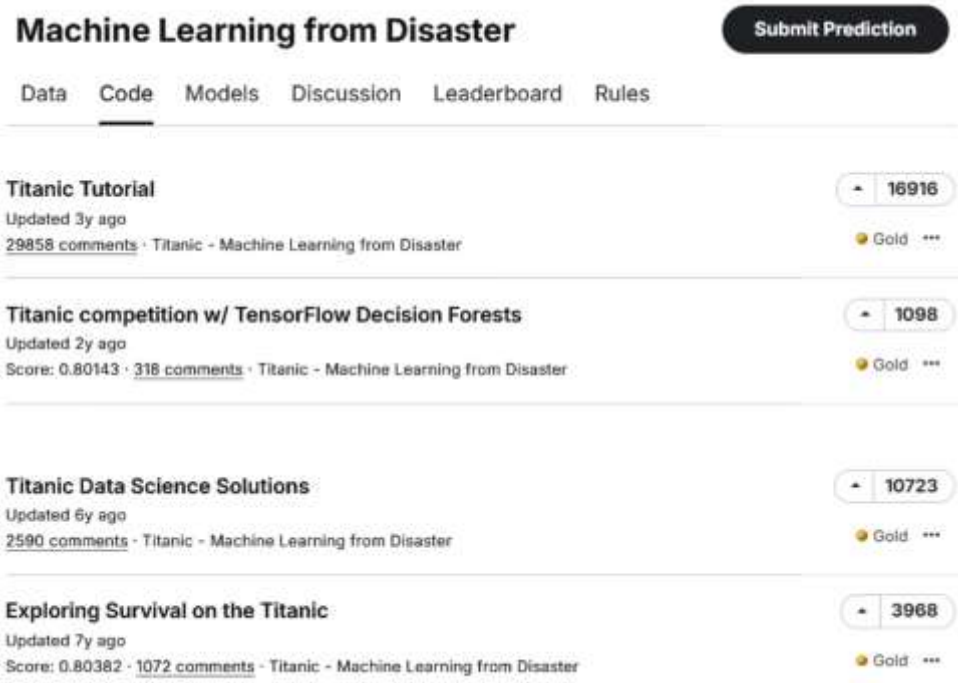


図 9.2-9 準備ができていてオープンソースの Pipeline ソリューションの合計 53,963 のうち、最初の 5 つのソリューション。Kaggle だけで、すでに 150 万人近くがこの問題を解こうとしている [157] 。

このような小さなテーブルでさえ、何百万ものユニークな解を生成できるのであれば（図 9.2-9）、何万ものパラメータが測定される実際の工業建設データセットについては、どう言えるだろうか。

> 標準的な CAD の -比較的小さな建物の設計には、幾何学的特性からコストや時間の属性に至るまで、何千ものパラメータを持つ何万ものエンティティが含まれています。長年にわたって収集された御社のすべてのプロジェクトのデータの中に、どれだけの潜在的な洞察、関係、予測、経営仮説が隠されているか想像してみてください。過去のプロジェクト・データは単なるアーカイブではなく、組織の生きた記憶であり、独自の仮説を数多く構築するために分析可能なデジタル・フットプリントなのです。

最も重要なことは、Kaggle コミュニティがあなたの会社やデータに関心を持つのを待つ必要はないということです。自社のデータで分析を実行し、自社のデータでモデルを訓練し、繰り返し、異常、 パターンを特定する。かつては何年もの実験と高価なコンサルティングが必要でしたが、今ではイニシアチブ、LLM、データへのオープンなアプローチ、学ぶ意欲があれば十分です。

train.csv 乗客訓練データセットに基づいて乗客の生存率を予測する機械学習アルゴリズム を構築するために、LLM にこの問題を解決してもらいましょう：

> タイタニック号の乗客のトレーニングデータセットに基づき、生存率を予測する機械学習モデル（）を構築する。

LLM の回答 ：

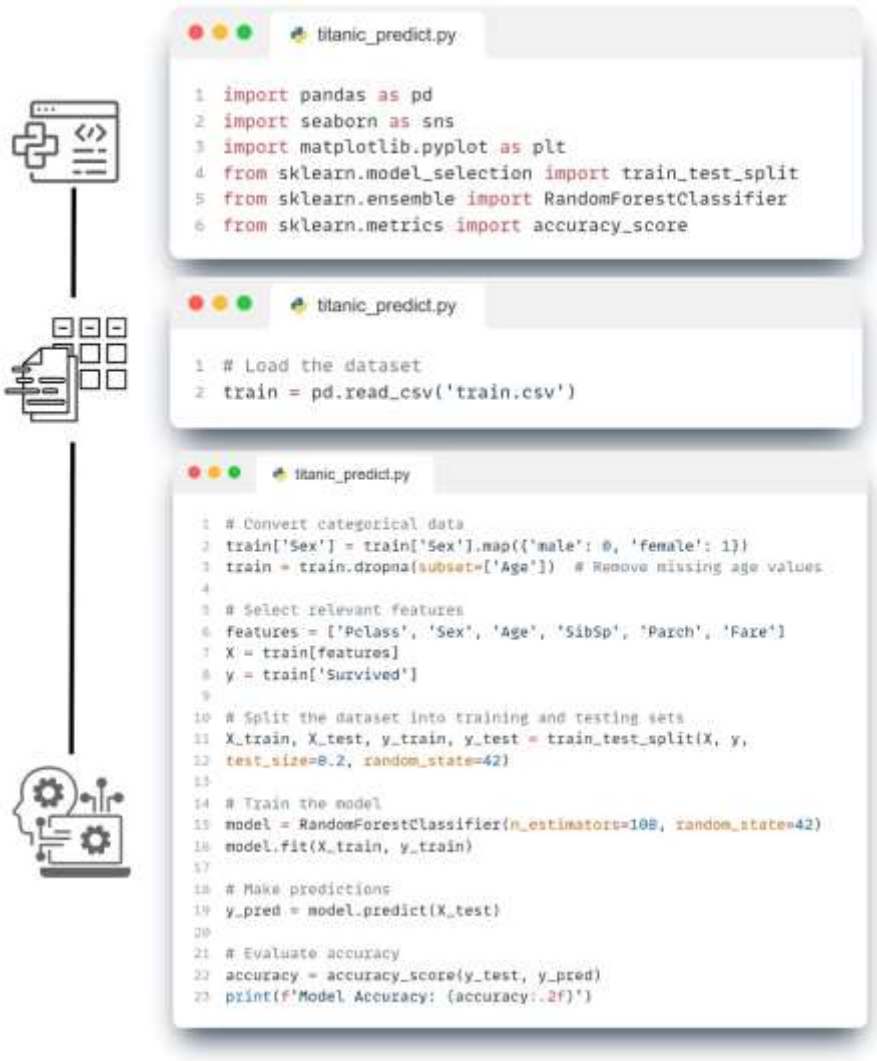


図 9.2-10 LLM は機械学習アルゴリズム Random Forest を用いてタイタニック号の生存者予測を構築した。

LLM （図 9.2-10）から得られるコードは、タイタニックの乗客データをロードし、それをクレンジングし、カテゴリ変数（例えば、性別）を数値フォーマットに変換し、RandomForestClassifier アルゴリズムを通してモデルを学習し、乗客が生存しているかどうかを予測します（次の章で一般的なアルゴリズムについて詳しく説明します）。

このコードでは、トレーニングデータをトレーニングセットとテストセットに分け（Kaggle のウェブサイト、トレーニング用に既製の test.csv（図 9.2-7）と train.csv（図 9.2-6）を作成済み）、トレーニングデータでモデルをトレーニングし、テストデータで特定の予測モデルがどの程度優れているかをテストする。トレーニングの後、test.csv のテストデータ（生き残った人、生き残らなかった人に関する実際のデータ）がモデルに投入され、生き残った人、生き残らなかった人を予測する。私たちの場合、 、得られた機械学習モデルの精度は約 80%で、パターンをかなりよく捉えていることがわかる。

> 機械学習 は、四角いブロックを丸い穴にはめようとする子供に例えることができる。初期段階では、アルゴリズムは多くのアプローチを試み、エラーや矛盾に遭遇する。このプロセスは非効率に見えるかもしれないが、重要な学習を提供する。各エラーを分析することで、モデルは予測を改善し、ますます正確な決定を下す。

このモデル（図 9.2-10）は、新しい乗客の生存率を予測するために使用することができます。例えば、model.predict 関数を使用して、パラメータを乗客情報を入力します：例えば、"男性"、"3 等席"、"25 歳"、"身寄りがない "というパラメータで model.predict 関数を使用して乗客情報を入力すると、モデルは予測を作成します - もし彼が 1912 年のタイタニック号にいた場合、80%の確率で乗客は惨事から生還できないでしょう（図 9.2-11）。

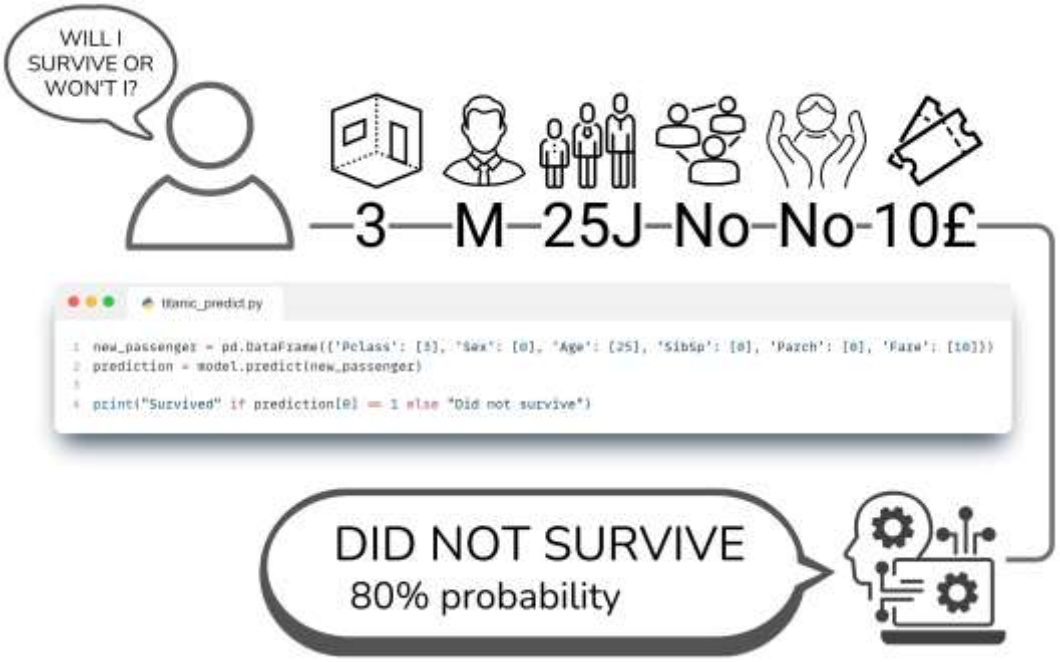


図 9.2-11 上で作成したモデルは、タイタニック号の乗客が新たに生存するかどうかを 80%の確率で予測できるようになった。

タイタニック号の乗客生存予測モデルは、もっと広い概念を示している。建設業界では毎日、何千人もの専門家が同じような「二重」の決断を下している。決断の生死、プロジェクト、見積もり、 、ツール、利益か損失か、安全かリスクか。タイタニック号の例で、結果が要因（性別、年齢、階級）に左右されたように、建設業では、意思決定の各側面が、多くの独自の要因や変数（表の列）に影響される：材料のコスト、作業員の資格、納期、天候、物流、技術的リスク、コメント、その他何十万ものパラメータ。

> 建設業界では、機械学習は他の分野と同じ原則に従っている。モデルは、プロジェクト、契約、見積もりなどの過去のデータに基づいて訓練され、さまざまな仮説を検証して最も効果的な解決策を見つける。このプロセスは、試行錯誤を繰り返しながら子供に教えるようなもので、サイクルを繰り返すごとにモデルは適応し、より正確になっていく。

蓄積されたデータの活用は、建設に新たな地平を開く。労働集約的な手作業による計算の代わりに、将来のプロジェクトの主要な特性を高い精度で予測できるモデルを訓練することができる。このように、予測分析 は、建設業界を、計画を立てるだけでなく、自信を持って発展を予測できる空間に変える。

過去のデータに基づく予測・予想

同社のプロジェクトで収集されたデータは、時間のかかる手作業による計算や比較をすることなく、まだ実現されていない将来の対象物のコストと時間の特性を予測できるモデルを構築する可能性を開く。これにより、主観的な仮定ではなく、数学的な予測に基づく評価プロセスを大幅にスピードアップし、簡素化することが可能になる。

先に、本書の第 4 部では、リソースベース法（ ）を含む、プロジェクトコスト見積りの伝統的な手法を詳細に検討し、パラメトリックアプローチやエキスパートアプローチについても言及した。これらの方法は、現在でも有効であるが、現代の実務においては、統計解析や機械学習（ ）のツールを用いて、見積りの精度と再現性を大幅に向上させることができるようになってきている。

価格や時間属性の手動計算や半自動計算のプロセスは、将来的には、過去のデータを分析し、隠れたパターンを見つけ、情報に基づいた意思決定を提案することができる ML モデルの意見や予測によって補完されるだろう 。
新しいデータやシナリオは、すでに利用可能な情報から自動的に生成されるようになるだろう。言語モデル（ LLM ）が、オープン・ソースから長年にわたって収集されたデータに基づいてテキスト、画像、コードを作成するのと似ている[158]。

> 今日、人間が経験や直感、内部統計に頼って将来の出来事を評価しているように、今後数年間で、建設プロジェクトの将来は、蓄積された知識と数学的機械学習モデル の組み合わせによってますます決定されるようになるだろう。

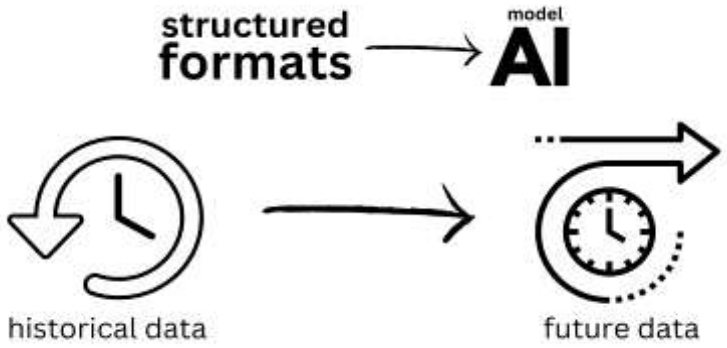


図 9.2-12 質的で構造化された過去の企業データは、機械学習モデル と予測を構築する材料となる。

単純な例として、面積、敷地面積、部屋数、地理的位置から住宅の価格を予測する場合を考えてみよう。一つのアプローチは、これらのパラメータを分析し、予想価格を計算する古典的モデルを構築することである（図 9.2-13）。このアプローチでは、事前に正確で既知の計算式が必要であり、実際の実務では現実的に不可能である。

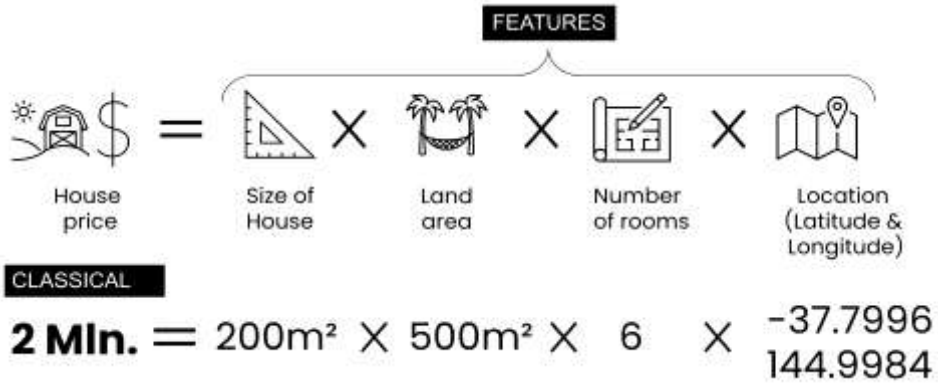


図 9.2-13 古典的なアルゴリズムを用いれば、求めるべき計算式が決まっている住宅の価値を推定することができる。

機械学習 は、手作業で数式を探す手間を省き、訓練されたアルゴリズムに置き換える。このアルゴリズ ムは、独立に依存関係を特定し、あらかじめ決められた数式よりも何倍も正確である。あるいは、機械学習アルゴリズム を作成しよう。このアルゴリズムは、問題の事前理解と不完全かもしれない過去のデータに基づいてモデルを生成する（図 9.2-14）。

価格設定を例にとると、機械学習は、コスト形成の正確なメカニズムに関する知識を必要としない、さまざまなタイプの数学的モデルの作成を可能にする。このモデルは、過去のプロジェクトのデータから「学習」し、建築パラメータ、コスト、納期の間の実際のパターンに適応する。

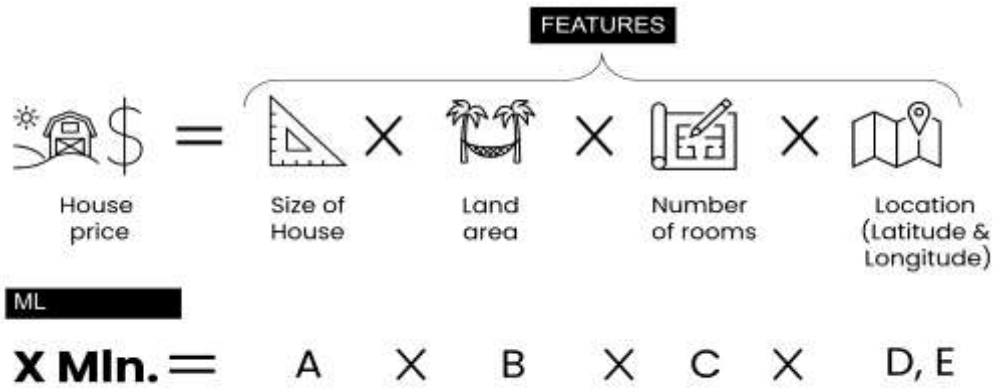


図 9.2-14 古典的な数式ベースの推定とは異なり、機械学習アルゴリズム は過去のデータに基づいて学習される。

教師あり機械学習の文脈では、、訓練データセットの各プロジェクトは、入力属性（例 えば、類似の建物のコストと時間）と期待される出力値（例 えば、コストや時間）の両方を含んでいる。同様のデータセットは、機械学習モデル （図 9.2-15）の作成とカスタマイズに使用される。データセットが大きく、その中のデータの質が高ければ高いほど、モデルはより正確になり、予測結果もより正確になる。

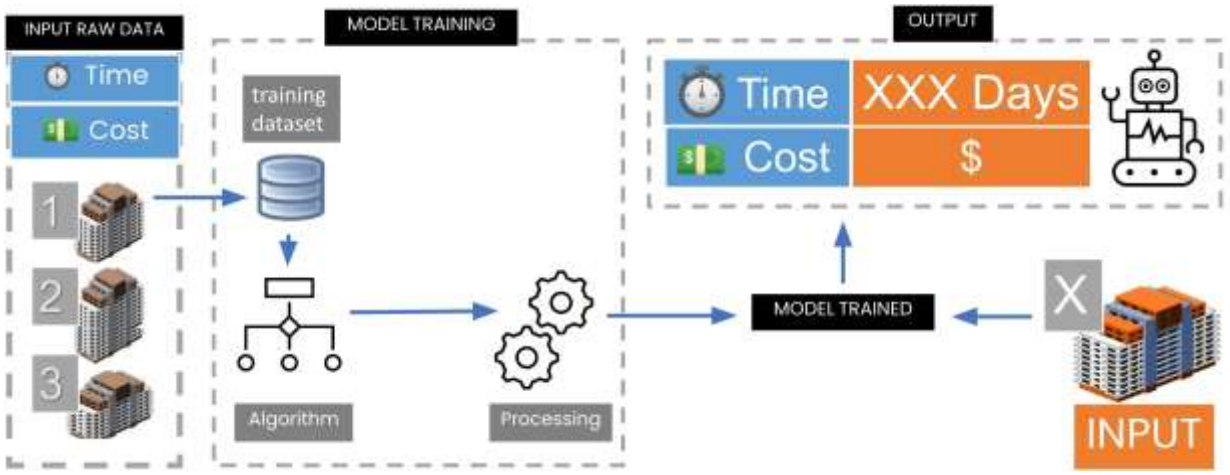


図 9.2-15 過去のプロジェクトのコストとスケジュールのデータで学習した ML モデルは、ある確率で新規プロジェクトのコストとスケジュールを決定する。

いったんモデルが作成され、新しいプロジェクトの建設を推定するように訓練されれば、新しいプロジェクトの新しい属性をモデルに与えるだけで、モデルは以前に学習したパターンに基づいて、ある程度の確率で推定結果を提供する。

機械学習の主要概念

機械学習は魔法ではなく、単なる数学であり、データであり、パターンを見つけるものだ。機械学習には本当の知性はなく、データに基づいて訓練されたプログラムがパターンを認識し、人間が常に関与することなく意思決定を行う。

機械学習は、その構造を説明するためにいくつかの重要な概念を用いる（図 9.2-16）：

- ラベルは、モデルが予測すべきターゲット変数または属性（タイタニック・データセットでは "Survivor "パラメータ）です。例：建設コスト（例：ドル）、建設工事の期間（例：月）。
- 特徴量はモデルの入力となる独立変数または属性である。予測モデルでは、ラベルを予測するために使用される。例：敷地面積（平方メートル）、建物の階数、建物の総床面積（平方メートル）、地理的位置（緯度と経度）、建設に使用される材料のタイプ。特徴の数はデータの次元性も決定する。
- モデルとは、異なる仮説の集合であり、そのうちの 1 つが、予測または近似される対象関数を近似する。例：機械学習モデル 、回帰分析技法を使用して建設コストと時期を予測する。
- 学習アルゴリズム 学習アルゴリズムとは、学習データのセットを使用して、ターゲット関数に正確に

一致するモデルの最良の仮説を見つけるプロセスである。例：関係やパターンを識別するためにコストや工期データを分析する線形回帰、KNN またはランダムフォレストアルゴリズム。

- トレーニング - トレーニング・プロセスでは、アルゴリズムはトレーニング・データを分析し、入力属性とターゲット・ラベルの間の関係に対応するパターンを見つける。このプロセスの結果は、学習済みの機械学習モデル（）であり、予測の準備が整っている。例：アルゴリズムが過去の建設データ（コスト、時間、設備特性）を分析し、予測モデルを作成するプロセス。

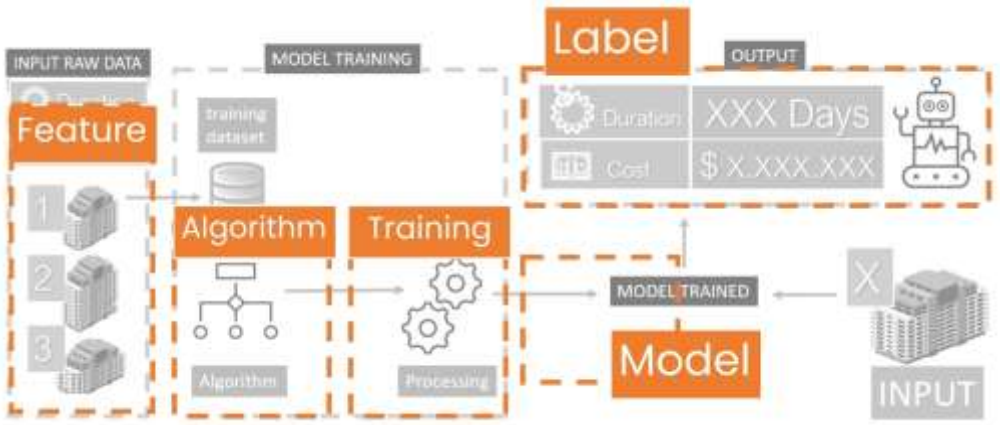


図 9.2-16 ML はラベルと属性を用いてモデルを作成し、アルゴリズムを用いてデータ上で学習させ、結果を予測する。

機械学習 は単独で存在するのではなく、統計学、データベース、データマイニング、パターン認識、ビッグデータ分析、人工知能を含む、より広範な分析分野のエコシステムの一部である。図 9.2-17 は、これらの分野がどのように重なり合い、補完し合って、現代の意思決定と自動化システムの包括的な枠組みを提供しているかを示している。

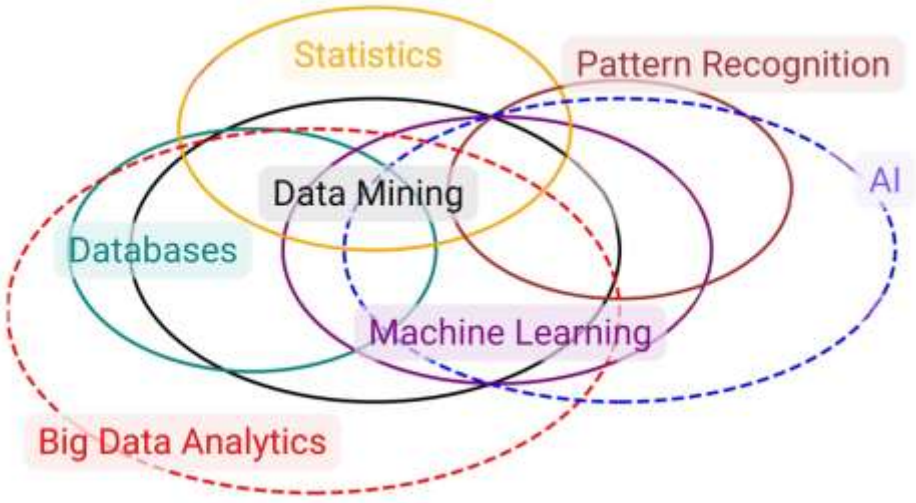
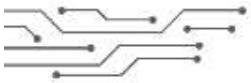


図 9.2-17 データ分析のさまざまな分野の関係：統計学、機械学習、人工知能、ビッグデータ、パターン認識、データマイニング。

> 機械学習 の主な目的は、人間の介入や支援なしに知識を自動的に学習し、それに応じて行動を調整する能力をコンピュータに与えることである[159]。

つまり、将来的に人間の役割は、機械に認知能力を提供することだけになるだろう。人間が条件、重み、パラメーターを設定し、あとは機械学習モデル 。

次の章では、アルゴリズムの具体的な適用例を検討する。実際の表や簡略化したモデルを用いて、予測がどのように段階的に構築されていくかを示す。



章 9.3.

機械学習によるコストと納期の予測

機械学習（）を使ってプロジェクトのコストとスケジュールを求める例

工期とコストの見積もりは、建設会社の活動において重要なプロセスのひとつである。従来、このような見積もりは、経験や参考書、規制データベースに基づいて専門家が行ってきた。しかし、デジタルトランスフォーメーションとデータの利用可能性の増加に伴い、機械学習（ML）モデル（ ）を使用して、このような見積もりの精度を向上させ、自動化することが可能になりました。

建設コストと建設時期の算出プロセスに機械学習（ ）を導入することで、より効率的なプランニングが可能になるだけでなく、リスク管理から物流や調達の最適化まで、他のビジネスプロセスにインテリジェントモデルを統合する出発点にもなる。

> プロジェクトの工期と総工費を迅速に判断することは重要である。このようなプロジェクトの期間とコストに関する疑問は、建設業界が誕生して以来、伝統的に顧客と建設会社の両方の頭の片隅にありました。

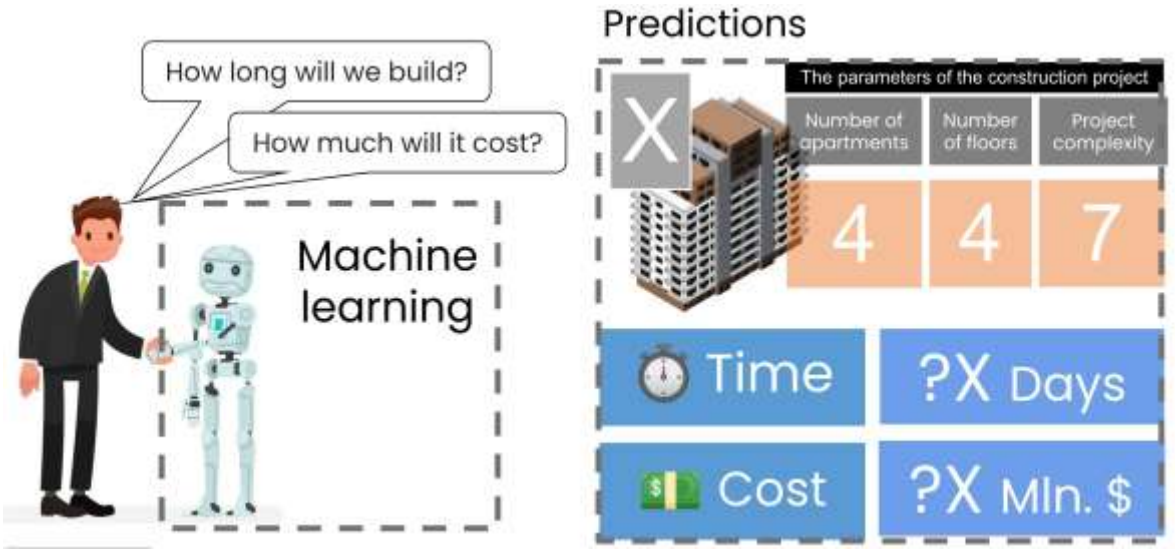


図 9.3-1 建設プロジェクトでは、工期とコストの見積もりのスピードと質が、成功の鍵となる。

以下の例では、過去のプロジェクトから主要なデータを抽出し、機械学習モデル を開発するために使用する。

このモデルを使用して、新しいパラメータ を持つ新しい建設プロジェクトのコストとタイミングを推定することができる（図 9.3-1）。

つの主要な属性を持つ 3 つのプロジェクトを考えてみよう：フラット数（ここで、100 フラットは、視覚化しやすいように 10 という数字に相当する）、階数、および 1 から 10 までのスケールでの建設の複雑さの条件付き尺度（ここで、10 は最も高い複雑さのスコアである）。機械学習では、100 を 10 に、50 を 5 に変換して単純化するプロセスを「正規化」と呼ぶ。

> 機械学習における正規化とは、処理と分析を容易にするために、異なる数値データを共通の尺度に合わせるプロセスのことである。このプロセスは、データの尺度や測定単位が異なる場合に特に重要である。

最初のプロジェクト（図 9.3-2）が、50 フラット（正規化後、5）、7 階建て、複雑さスコア 2、つまり比較的単純な構造であったとする。二つ目のプロジェクトは、80 フラット、9 階建て、比較的複雑な設計であった。このような条件下で、1 棟目と 2 棟目の建設に要した日数はそれぞれ 270 日と 330 日、総工費はそれぞれ 450 万ドルと 580 万ドルであった。



図 9.3-2 将来のプロジェクト X の時間とコストを見積もるために使われる、過去のプロジェクト一式の例。

このようなデータに対して機械学習モデル）を構築する場合、主なタスクは予測に重要な属性（またはラベル）、この場合は工期とコストを特定することである。機械学習アルゴリズム（ ）を使って、新しいプロジェク

ト X のコストと工期を、40 フラット、4 階建て、7 という相対的に高いプロジェクトの複雑さなど、新しいプロジェクトの与えられた属性に基づいて予測する必要がある（図 9.3-2）。現実世界では、入力パラメーターの数はもっと多く、数十から数百の要素に及ぶことがある。これには、建設資材の種類、気候帯、請負業者の資格レベル、ユーティリティの利用可能性、基礎の種類、工事開始時期、現場監督のコメントなどが含まれる。

> 予測機械学習モデル を作成するには、それを作成するアルゴリズムを選択する必要がある。機械学習におけるアルゴリズム（）とは、コンピューターに予測（パラメーターの正しい順序での組み合わせ）やデータに基づく意思決定の方法を教える数学的レシピのようなものである。

過去の建設プロジェクトのデータを分析し、将来のプロジェクトの期間とコストを予測するために（図 9.3-2）、よく使われる機械学習アルゴリズムを使用することができる ：

- 線形回帰 (Linear regression): このアルゴリズムは、属性間の直接的な関係、たとえば階数と建設コストの関係を見つけようとする。このアルゴリズムの目的は，予測を可能にするこの関係を最もよく記述する線形方程式を見つけることである．
- アルゴリズム k-nearest neighbours (k-NN): このアルゴリズムは、新しいプロジェクトを、サイズや複雑さが類似している過去のプロジェクトと比較する。k-NN は，k 個（個）の学習例のどれが最も近いかに基づいてデータを分類する．回帰の文脈では，結果は k 個の最近傍の平均または中央値となる．
- 決定木: 予測モデリング・モデルの一つで、ツリー構造を使って異なる条件に基づいてデータをサブセットに分割する。ツリーの各ノードは、データのさらなる分割につながる条件または質問を表し、各葉は最終的な予測または結果を表す。このアルゴリズム は、まずストーリーの数によって、次に複雑さによって、といったように、異なる特徴に基づいてデータをより小さなグループに分割し、予測を行う。

線形回帰と K-ニアレスト・ネイバーズ（K-nearest neighbours）アルゴリズムという 2 つのよく使われるアルゴリズムを例に、新しいプロジェクトのコストを見積もるための機械学習アルゴリズム（）を見てみよう。

線形回帰を用いたプロジェクト・コストと時間の予測

線形回帰 は、1 つまたは複数の他の変数との線形関係に基づいて変数の値を予測する基本的なデータ分析アルゴリズムである。このモデルは、従属変数と 1 つまたは複数の独立変数の間に直接的な線形関係があることを

仮定し、アルゴリズムの目的は、この関係を見つけることです。

> 線形回帰の単純さと明瞭さによって、さまざまな分野で人気のあるツールとなっている。1 つの変数を扱うとき，線形回帰は，データ・ポイントを通して最も適合する線を見つけることである．

線形回帰 は、入力変数 X と出力変数 Y の間の依存関係 を近似する最良の直線（赤線）を見つけます。この直線は、識別された線形関係に基づいて、新しい X 値に対する Y 値を予測することを可能にする（図 9.3-3）。

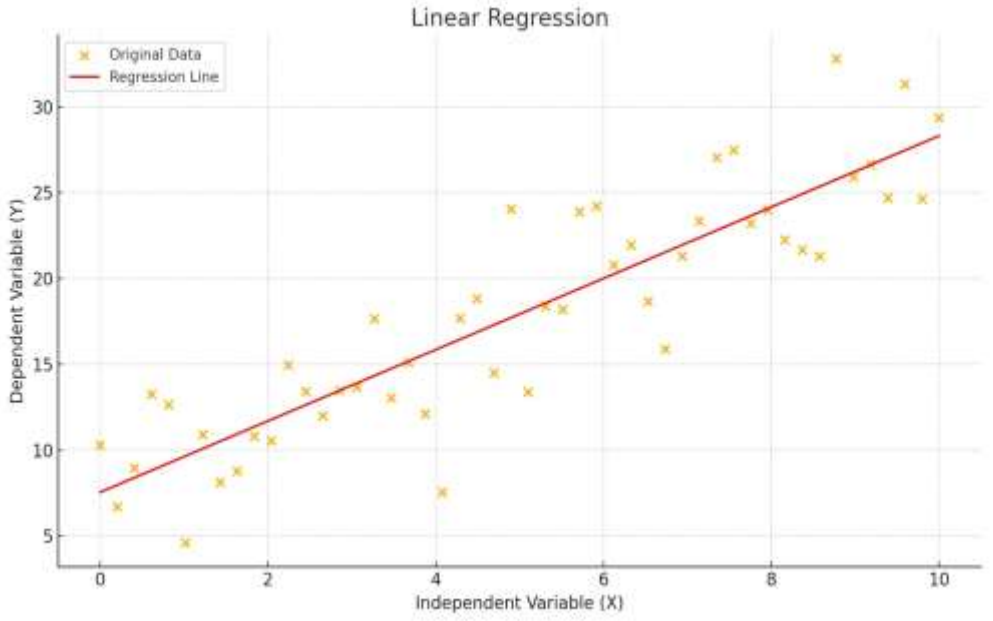


図 9.3-3 線形回帰の原理は、学習値を通る最良の直線を見つけることである。

この直線は、独立変数（X）の値を入力すると従属変数（Y）の予測値が得られる方程式で表される。このプロセスにより、両者の間の線形関係を使用して、X の既知の値から Y を効率的に予測することができます。このような統計的に平均化された直線を見つける例は、サンフランシスコの建築許可データ評価（図 9.1-7）で見ることができる。そこでは、線形回帰を使ったインフレ率がさまざまなタイプの施設について計算されている。

プロジェクト・データ表（前章の図 9.3-2）を LLM に直接ロードし、簡単な機械学習モデル を構築するよう依頼してみよう。

- LLM チャット（CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ）にテキストリクエストを送る：

> 新しいプロジェクト X を実現するためのコストと時間を予測するための、単純な機械学習モデル の構築を示す必要がある（図 9.3-2、添付画像）。⏎

LLM の回答 ：

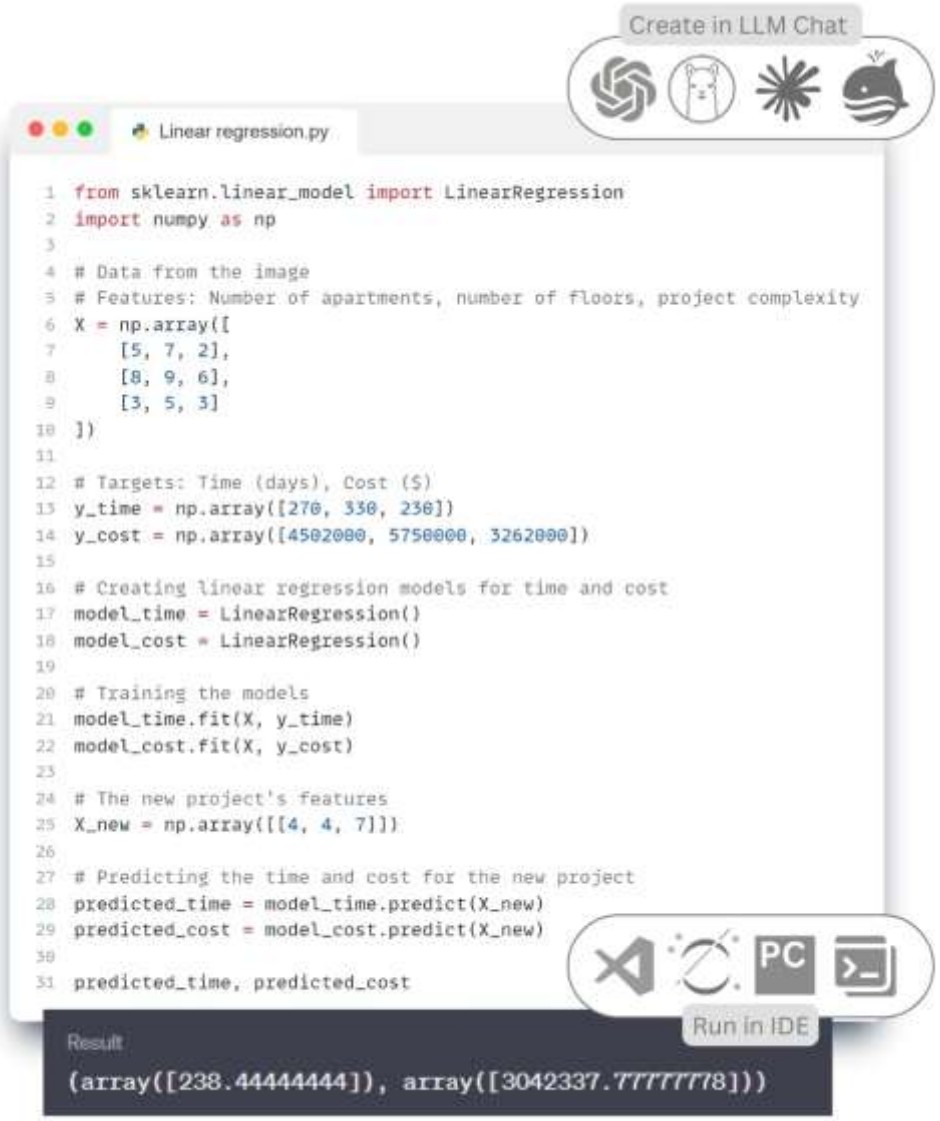


図 9.3-4 LLM プロジェクト・コストと時間を予測する機械学習モデル を作成するために、線形回帰を選択した。

LLM は、添付画像から表を自動的に認識し、データを視覚的フォーマットから表配列に変換した（図 9.3-4-6

行目）。この配列は、特徴量とラベルを作成するための基礎として使用され、そこから機械学習モデルが作成された（図 9.3-4-17〜22 行目）、線形回帰が使用された。

極めて小さい」データセットで学習させた基本的な線形回帰モデルを使用して、プロジェクト X（ ）とラベル付けされた新しい仮想の建設プロジェクトの予測を行った。我々の問題では、このプロジェクトは 40 フラット、4 階建て、複雑度 7 で特徴付けられる（図 9.3-2）。

新プロジェクト X の限定された小さなデータセットに基づく線形回帰モデルによって予測された通りである（図 9.3-4 - 24-29 行目）：

- 工期はおよそ 238 日（238.4444444）。
- 総支出は約$ 3,042,338 (3042337.777) となる。

プロジェクト・コストの仮説をさらに検討するためには、さまざまな機械学習アルゴリズ ムや手法（ ）を試してみることが有効である。そこで、K-Nearest Neighbours アルゴリズム（k-NN ）を使って、過去の小さなデータセットに基づき、新しいプロジェクト X の同じコストと時間の値を予測してみましょう。

K-nearest neighbour algorithm (k-NN) を用いたプロジェクトコストと時間の予測

新規プロジェクトのコストと期間を見積もるための追加予測因子として、k-Nearest Neighbours（k-NN ）アルゴリズムを使用する。 K-最近傍（k-NN）アルゴリズムは、分類 と回帰の両方に対応する教師あり機械学習（supervised machine learning）手法である。また、k-NN アルゴリズムについては、ベクトルデータベース検索（図 8.2-2） の文脈でも以前に説明しましたが、ここでは最も近いベクトル（テキスト、画像、技術的な説明など）を見つけるために使用されます。このアプローチでは、各プロジェクトは多次元空間の点として表現され、各次元はプロジェクトの特定の属性に対応する。

私たちの場合、各プロジェクトの 3 つの属性が与えられたら、それらを 3 次元空間の点として表します（図 9.3-5）。したがって、今度のプロジェクト X は、この空間に座標(x=4, y=4, z=7)で配置されることになる。現実の世界では、点の数と空間の次元は桁違いに大きい可能性があることに注意する必要がある。

> K-NN（k-nearest-neighbors）アルゴリズムは、希望するプロジェクト X とトレーニングデータベースのプロジェクト間の距離を測定することで動作する。これらの距離を比較することで、アルゴリズムは新しいプロジェクト X のポイントに最も近いプロジェクトを決定します。

たとえば、元のデータセットから 2 番目のプロジェクト（x=8, y=9, z=6）が、X（図 9.3-5）から他のプロジェク トよりもはるかに離れている場合、そのプロジェクトは以降の分析から除外することができる。その結果、2 つ (k=2) の最も近いプロジェクトだけが計算に使用され、それに基づいて平均値が決定される。

このような方法は、近隣探索を通じて、プロジェクト間の類似性を評価することができ、その結果、以前に実施された類似のプロジェクトに基づいて、新しいプロジェクトの可能性の高いコストと時期に関する結論を導き出すのに役立つ。

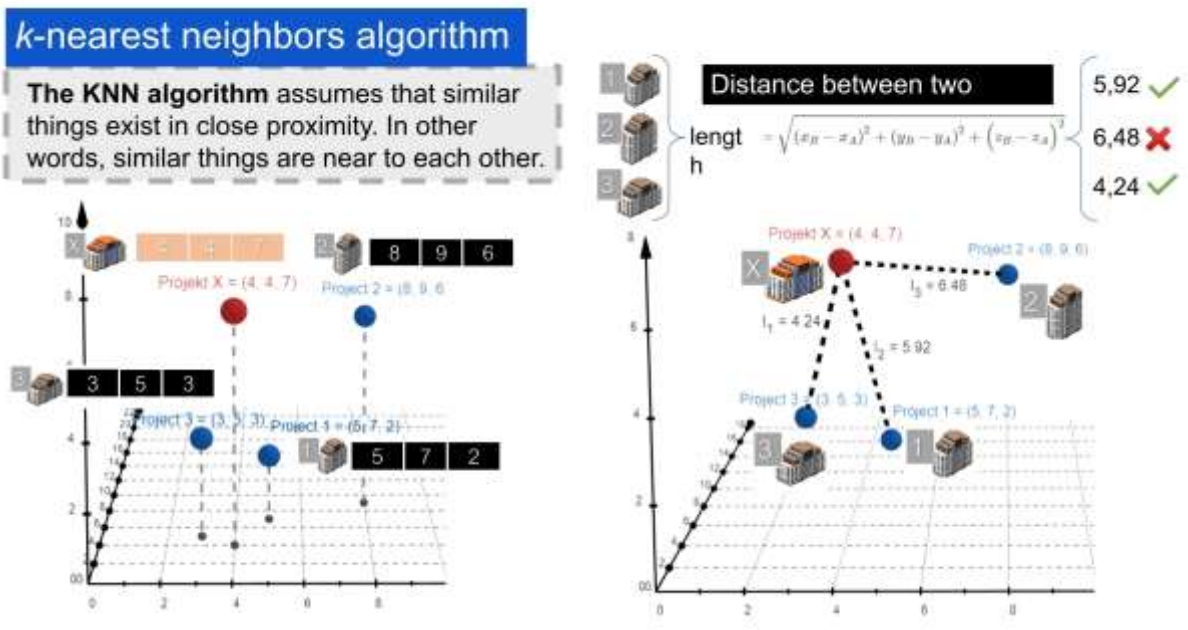


図 9.3-5 K-NN アルゴリズムでは、プロジェクトは多次元空間の点として表現され、類似性の評価と予測のために、距離に基 づいて最も近いプロジェクトが選択される。

k-NN の作業にはいくつかの重要なステップがある：

- データの準備：トレーニングデータとテストデータセットが最初にロードされる。訓練データはアルゴリズムを「訓練」するために使用され、テストデータはその効率をチェックするために使用される。
- パラメータ K の選択: 数値 K が選択され、アルゴリズムで考慮すべき最近傍（データポイント）の数を示します。K " の値は結果に影響するので非常に重要である。
- テストデータの分類プロセス 、回帰：
  - 距離の計算：テストデータの各要素について、トレーニングデータの各要素との距離を計算する（図 9.3-5）。ユークリッド距離（最も一般的な方法）、マンハッタン距離、ハミング距離など、さ

まざまな距離測定方法を使用できます。

- K 個の最近接点の並べ替えと選択：距離を計算した後、並べ替えを行い、テスト点に最も近い K 個の点を選択する。
- テストポイントのクラスまたは値の決定：分類タスク の場合、テストポイントのクラスは、選択された K 個の近傍の中で最も頻度の高いクラスに基づいて決定される。回帰タスクの場合は、K 個の近傍値の平均（または他の中心傾向の尺度）が計算される。

- プロセスの完了：すべてのテストデータが分類されるか、それに対する予測がなされれば、プロセスは完了する。

> アルゴリズム k-nearest neighbours (k-NN)は、多くの実用的なアプリケーションで効果的であり、機械学習の専門家の主要なツールの 1 つである。このアルゴリズムは、特にデータ間の関係が解釈しやすいタスクにおいて、その単純さと効率性から人気がある。

この例では、K-nearest neighbours アルゴリズムを適用した後、プロジェクト X と最短距離の 2 つのプロジェクトが（少数のサンプルの中から）特定された（図 9.3-5）。これらのプロジェクトに基づいて、アルゴ リズムは価格と工期の平均を決定する。分析の結果（図 9.3-6）、アルゴリズムは最近傍の平均をとることで、プロジェクト X のコストはおよそ$ 3,800,000 ドル、工期は約 250 日と結論づけた。

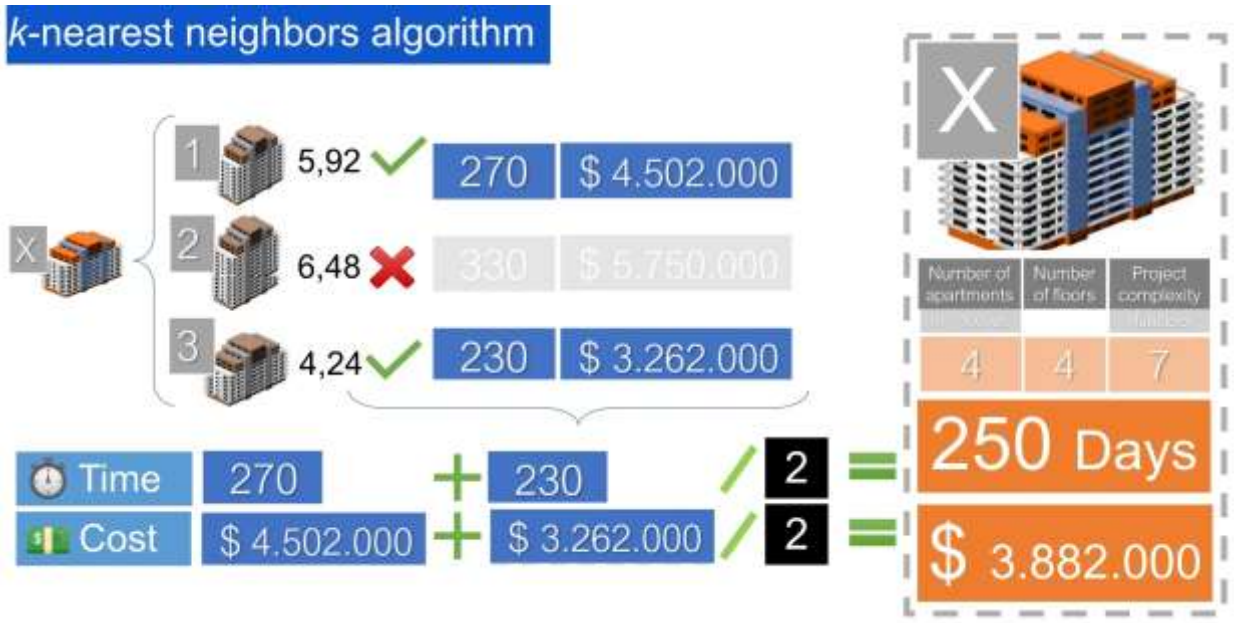


図 9.3-6 K-nearest neighbours アルゴリズムは、サンプルの中で最も近い 2 つのプロジェクトを分析することで、プロジェクト X のコストとスケジュールを決定する。

k- Nearest Neighbors (k-NN) アルゴリズムは、推薦システムなどの 分類や回帰タスクで特に人気があり、特定のユーザーの興味に似た嗜好に基づいて製品やコンテンツを提案するために使用される。さらに、k-NN は、患者の症状に基づいて 病気の種類を分類する医療診断や、パターン認識、顧客の信用度を評価する金融分野で広く使用されている。

> 限られたデータであっても、機械学習モデル（）は有用な予測を提供し、建設プロジェクト管理の分析要素を大幅に強化することができる。過去のデータが拡張され、クリーンアップされるにつれて、より洗練されたモデル（例えば、工事の種類、場所、着工時期、その他の要因を考慮したモデル）に移行することが可能になる。

我々の単純化したタスクでは、3D 空間で可視化するために 3 つの属性を使用したが、実際のプロジェ クトは、平均して数百から数千の属性を含んでいる（「CAD ベースのビッグデータの例 (BIM)」の章にあるデータセットを参照）。これは、空間の次元を著しく増加させ、プロジェクトをベ クトルとして表現する複雑さを増大させる（図 9.3-7）。

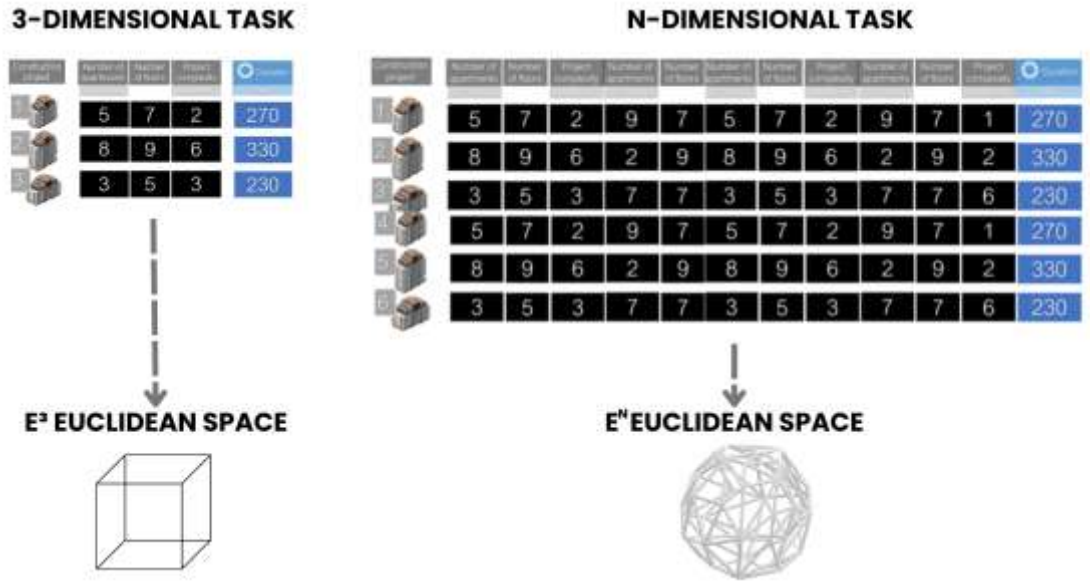


図 9.3-7 簡略化した例では、3D -ビジュアライゼーションに 3 つの属性を使用したが、実際のプロジェクトではもっと多くの属性がある。

40 フラット、4 階建て、複雑度レベル 7 のプロジェクト X について、同じデータセットに異なるアルゴリズムを適用すると、異なる予測値が得られた。 線形回帰アルゴリズムでは、完了日数 238 日、コスト$ 3,042,338 と予測され（図 9.3-4）、k-NN アルゴリズム では、完了日数 250 日、コスト$ 3,882,000 と予測された（図 9.3-6）。

機械学習モデル（ ）を用いて得られる予測の精度は、入力データの量と質に直接依存する。より多くのプロジェクトがトレーニングに関与し、その特性（属性）と結果（ラベル）がより完全かつ正確に表現されていればいるほど、最小限のエラー値で信頼性の高い予測が得られる確率は高くなる。

データの前処理技術は、このプロセスにおいて重要な役割を果たす：

- 特徴を共通の尺度に合わせるための正規化；
- 異常値の検出と除去、モデルの歪みの除去；
- テキストデータの操作を可能にするカテゴリー属性のコード化；
- 欠損値を埋め、モデルのロバスト性を高める。

さらに、交差検証法を用いてモデルの一般性と新しいデータセットに対する頑健性を評価し、オーバーフィッティングを検出して予測の信頼性を向上させる。

> カオスは解読されるべき秩序である［160］。
>
> - ジョゼ・サラマーゴ、"ダブル"

しかし、世の中のあらゆる出来事、特に建設プロセスには数学的な法則があり、厳密な数式ではなく、統計や過去のデータを用いて計算する必要がある。

従来の原価計算部門による見積もりも、機械学習モデル（ ）も、必然的に不確実性と潜在的なエラー源に直面する。しかし、十分な品質のデータがあれば、機械学習モデル、専門家の判断に匹敵する、時にはそれ以上の予測精度を示すことができる。

> 機械学習は、計算を改良し、代替シナリオを提案し、プロジェクト・パラメーター間の隠れた依存関係を特定することができる、分析のための信頼できる補完的なツールになる可能性が高い。このようなモデルは万能とは言えないが、近い将来、計算や意思決定プロセスにおいて重要な位置を占めるようになるだろう。機械学習技術は、エンジニア、見積もり担当者、アナリストの参加を排除するものではなく、逆に、過去のデータに基づく新たな視点を提供することで、彼らの能力を拡大するものである。

建設会社のビジネスプロセスに適切に統合されれば、機械学習は経営意思決定支援システムの重要な要素になる可能性を秘めている。それは人間の代わりとしてではなく、専門家としての直感や工学的論理の延長としてである。

次のステップ：貯蔵から分析・予測へ

データを扱う最新のアプローチは、建設業界における意思決定の方法を変え始めている。直感的な評価から客観的なデータ分析に移行することは、精度を向上させるだけでなく、プロセスを最適化する新たな機会を開くことにもなる。このパートを要約するために、考察した方法を日常業務に応用するのに役立つ主な実践的ステップを紹介する：

- ■ 持続可能な貯蔵インフラの確立
  - ☐ バラバラの文書やプロジェクトデータを 1 つの表モデルにまとめ、主要な情報を 1 つのデータフ

レームに集約し、さらなる分析に役立てる。

- [ ] 効率的なデータ保存形式、例えばCSVやXLSXの代わりにApache Parquetのようなカラムナー形式を使用する。
- [ ] プロジェクト全体を通して変更を追跡するためのデータバージョンシステムを確立する。

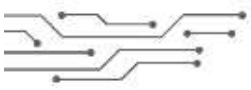
■ 分析・自動化ツールの導入

- [ ] 過去のプロジェクトデータ（文書、モデル、見積もり）の分析を開始し、パターン、傾向、異常を特定する。
- [ ] ETLプロセス（Extract, Transform, Load）をマスターし、データを自動的にロードし準備する。
- [ ] 様々な無料のPython可視化ライブラリを使用して、主要なメトリクスを可視化する方法を学びます。
- [ ] 代表的で再現性のある分析結果を得るために、統計的手法とランダムサンプリングの適用を開始する。

■ データを扱う成熟度の向上

- [ ] タイタニックデータセットのようなシンプルでわかりやすい例を使って、いくつかの基本的な機械学習アルゴリズムを学ぶ。
- [ ] 現在のプロセスを分析し、硬直的な因果関係ロジックから統計的手法による予測と推定に移行できる箇所を特定する。
- [ ] データを副産物ではなく戦略的資産として扱い始める：特定のソフトウェアソリューションではなく、データモデルを中心に意思決定プロセスを構築する。

データの価値に気づいた建設会社は、競争優位性がリソースの量ではなく、アナリティクスに基づく意思決定のスピードで決まるという、新たな発展段階を迎えている。



印刷版で最大限の利便性を

データ駆動型建設』の無料デジタル版をご覧いただけます。より便利な作業や資料への迅速なアクセスのためには、印刷版のご利用をお勧めします：



■いつでも手元に：印刷された本書は信頼できる作業ツールとなり、どのような作業状況でも適切なビジュアライゼーションやダイアグラムを素早く見つけて使用することができます。

■高品質のイラスト：印刷版のすべての画像とグラフィックは最高の品質で表示されます。

■情報への素早いアクセス：簡単なナビゲーション、メモやブックマークを作成する機能、どこでも本で作業する機能。

印刷された完全版を購入することで、快適で効率的な情報作業のための便利なツールを手に入れることができます。日々の作業でビジュアル資料を素早く利用し、必要なスキームを素早く見つけ、メモを取ることができます。さらに、ご購入はオープンな知識の普及をサポートします。

印刷版のご注文はこちらから：datadrivenconstruction.io/books



X 部

デジタル時代の建設業界。機会と課題

最後の第 10 部は、デジタルトランスフォーメーションの時代における建設業界の未来を包括的に考察する。因果関係の分析からビッグデータの相関関係への移行を分析している。ファインアートの進化と建設業におけるデータ作業の発展の間に類似点が描かれ、業界がいかに詳細な管理からプロセスの全体的な理解へと移行しつつあるかを示している。データの透明性と自動計算が従来のビジネスモデルを根本的に変え、仲介者の必要性をなくし、投機の機会を減らすという、建設業界の「ユーバー化」のコンセプトが探求される。 要素の普遍的な分類など、未解決の問題については、建設会社が新しい環境に適応するための時間を提供するため、詳細に議論されている。最後に、変化する業界において競争力を維持するために、脆弱性の分析やサービスの拡充を含むデジタルトランスフォーメーション戦略を策定するための具体的な提言で締めくくっている。

章 10.1.

生存戦略：競争優位の構築

計算ではなく相関：建設アナリティクスの未来

情報の急速なデジタル化（図 1.1-5）により、現代の建設は根本的な変革期を迎えており、データは単なるツールではなく、プロジェクトや事業管理に対する従来のアプローチを根本的に変える戦略的資産になりつつある。

何千年もの間、建設活動は、正確な計算、詳細設計、パラメータの厳格な管理といった決定論的手法に依存してきた。紀元後数世紀、ローマの技術者たちは水道橋や橋の建設に数学的原理を応用した。中世には、建築家がゴシック様式の大聖堂の理想的なプロポーションを追求し、20 世紀の工業化では、標準化された規範と規則のシステムが形成され、大量建設の基礎となった

今日、開発のベクトルは、厳密な因果関係の探求から、相関関係や隠れたパターンの探求といった確率論的分析へと移行しつつある。データは重要な資源となりつつあり、それに基づく分析（ ）は、直感的で局所的に最適化されたアプローチに取って代わりつつある。



図 10.1-1 工事データの隠された可能性：社内の既存の計算は、経営陣が分析できる氷山の一角にすぎない。

> 企業の情報システムは氷山のようなものである（図 10.1-1）。企業の経営陣が目にすることができるのは、データの潜在能力のごく一部であり、主な価値はその奥深くに隠されている。データを現在の利用方法だけでなく、将来それが解き放つ機会についても評価することが重要である。データから隠れたパターンを抽出し、新たな洞察を生み出す方法を学ぶ企業こそが、持続可能な競争優位性を生み出すのである

隠されたパターンを見つけ、データを理解することは、単に数字を扱うということではなく、抽象的な思考と、異質な要素の背後にある全体像を見る能力を必要とする創造的なプロセスである。この意味で、データを扱うことの進化は、視覚芸術の進化と比較することができる（図 10.1-2）。

建築の発展は、美術の発展と驚くほどよく似ている。どちらの場合も、人類は原始的な方法から高度な視覚化・分析技術へと進歩してきた。先史時代、人々は洞窟の絵や原始的な道具を使って日常の仕事を解決していた。中世とルネサンス期には、建築と芸術の洗練度が飛躍的に高まった。中世初期には、建築道具は単純な斧から、技術的知識の発展を象徴するような広範な道具一式へと進化していた。

写実主義の時代は、視覚芸術における最初の革命であった。芸術家たちは、細部まで再現する方法を学び、最

大限の説得力を獲得した。建築の分野では、この時代に対応するものとして、精密な工学技術、詳細な図面、厳格に規制された計算があり、これが何世紀にもわたって設計実務の基礎となった。

> その後、印象派は芸術的現実の認識そのものを変えた。芸術家たちは、文字通り形を描写する代わりに、絶対的な正確さよりも全体的な印象を反映することを目指し、雰囲気、光、ダイナミクスを捉えるようになった。同様に、建物分析における機械学習は、硬直した論理モデルからパターン認識や確率的パターンへと移行しており、古典的な分析ではアクセスできないデータの隠れた依存関係を「見る」ことを可能にしている。このアプローチは、形よりも意味（機能）を重視するバウハウスのミニマリズムと機能性の考え方を反映している。バウハウスは余計なものを排除し、明快さ、実用主義、大衆へのアピールのために装飾を捨てようとした。美学はデザインと目的の論理から生まれたのである。

19 世紀後半に写真が登場したことで、芸術はかつてない精度で現実を捉える新たなツールを手に入れ、視覚芸術に対する考え方がひっくり返った。同様に、建設においても、21 世紀の産業革命は、ロボット技術、レーザー、IoT 、RFID 、そして個々のパラメーターの収集が建設現場の現実を完全に把握するスケーラブルでインテリジェントなキャプチャへと進化した「コネクテッド・コンストラクション」のようなコンセプトの活用へとつながっている。

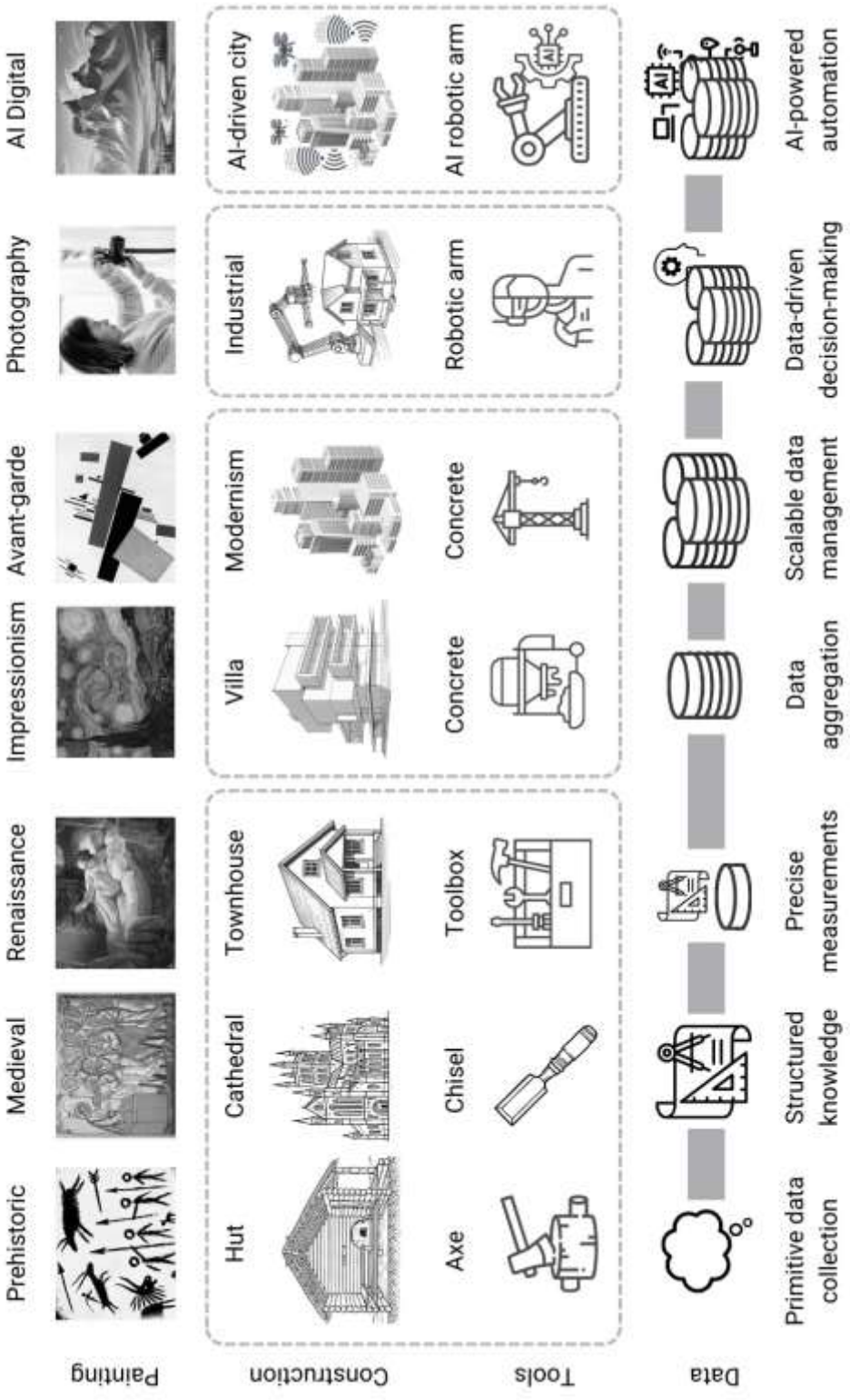


図 10.1-2 Era of Fine Art Evolutions は、建設業界におけるデータを扱うアプローチの発展と一致している。

今日、視覚芸術が AI ツール（ ）や LLM（ ）の登場によって再考を経験しているように、建設業界もまた新たな飛躍を経験している。人工知能（AI ）によって駆動されるインテリジェント・システム、LLM チャットは、人間の介入を最小限に抑えながら、ソリューションの予測、最適化、生成を可能にする。

設計と管理におけるデータの役割は根本的に変化した。19 世紀まで、現実が手描きの絵によって捉えられていたように、知識は口頭で経験的に伝達されるものであったが、今日では、施工の「絵」を完全にデジタル化することに焦点が当てられている。機械学習アルゴリズムの助けを借りて、このデジタル画像は建設現実の印象主義的表現に変換される。

> 私たちは、建築物の設計、建設、運用のプロセスが、単に拡張されるだけでなく、人工知能システム によって大きく駆動される時代に急速に近づいている。現代のデジタルアートが、テキストによるプロンプトと生成モデルを使って、絵筆を使わずに創作されるように、未来の建築とエンジニアリングのソリューションは、ユーザーが設定した主要なクエリーとパラメーターによって形作られるようになるだろう。

21 世紀において、データへのアクセス、その解釈、分析の質は、プロジェクトの成功に不可欠なものとなりつつある。そして、データの価値は、その量ではなく、それを分析し、検証し、行動に移す専門家の能力によって決まる。

建設におけるデータ主導のアプローチ：インフラの新たなレベル

人類の歴史において、このような技術の飛躍はそのたびに経済と社会に根本的な変化をもたらしてきた。今日、私たちは 19 世紀の産業革命に匹敵する規模の新たな変革の波を目の当たりにしている。しかし、100 年前の変化の主役は機械力とエネルギー技術であったのに対し、現在はデータと人工知能である。

機械学習 、LLM 、AI エージェントは、アプリケーションの本質を変え、従来のソフトウェアスタック（本書の第 2 部で説明）を不要にする（図 2.2-3）。すべてのデータロジックは、ハードコードされたビジネスルールではなく、AI エージェントに集約される（図 2.2-4）。

> データ時代において、従来のアプリケーションの見方は根本的に変わりつつある。かさばるモジュール式の企業システムが、必然的にオープンで軽量なカスタマイズ・ソリューションに取って代わられるモデルへと移行しつつある。

> 将来的には、基礎となるデータ構造だけが残り、それとのやりとりはすべて、データベースと直接作業するエージェントを通じて行われるようになるだろう。人工知能が基礎となるデータベースと直接やりとりするようになれば、アプリケーション・スタック全体は必要なくなるのですから。私は SaaS（）での仕事に全キャリアを費やしてきました。会社を設立し、そこで働いてきましたが、正直なところ、今すぐに新しい SaaS ビジネスを立ち上げることはないでしょう。正直なところ、私は今すぐに SaaS の新規事業を立ち上げることはないでしょうし、SaaS 企業に投資することもないでしょう。状況があまりにも不透明だからです。将来的にソフトウェア企業が存在しないとは言いませんが、その姿は大きく変わるでしょう。将来のシステムは、ビジネス・ロジックを[AI]エージェントに取り込んだデータベースとなるだろう。これらのエージェントは、単一のデータベースに限定されることなく、複数のデータ・リポジトリを同時に扱うことになる。すべてのロジックは AI レイヤー [46]に移行する。
>
> - マシュー・バーマン、フォワード・フューチャーCEO

新しいパラダイムの重要な違いは、技術的バラストの最小化である。巨大で複雑なクローズド・ソフトウェア・システムの代わりに、柔軟性があり、オープンで素早くカスタマイズ可能な、文字通りデータフローの中に "生きる "モジュールが登場する（図 7.4-1 - Apache Airflow, NiFi）。将来のプロセス管理のアーキテクチャは、マイクロアプリケーションの使用を想定している。コンパクトで、目的に合わせて構築されたツールは、大規模でクローズドな ERP、PMIS 、CDE 、CAFM システム とは根本的に異なる。新しいエージェントは、可能な限り適応性があり、統合され、ビジネスに特化したものとなる（例：ローコード/ノーコード 図 7.4-6）。

> すべてのビジネスロジックは、これらの[AI]エージェントに送られ、これらのエージェントは複数のリポジトリに対して CRUD [作成、読み込み、更新、削除]操作を実行します。エージェントは複数のデータベースを更新し、すべてのロジックはいわゆる AI -レベルに集約される。AI レイヤーにすべてのロジックが集約されれば、人々はバックエンドを交換し始めるだろう。私たちはすでに、Dynamics のバックエンドとエージェントの使用において、かなり高い割合で市場獲得を見ています。カスタマーサービスであれ、他の分野であれ、例えば CRM（）だけでなく、財務やオペレーション・ソリューションもそうです。ロジックレイヤーが AI や AI エージェントによって駆動されるような、AI 主導のビジネスアプリケーションが求められているからです。[...].私にとって最もエキサイティングなことの 1 つは、Excel with Python で、これは GitHub with Copilot に匹敵します。つまり、Excel があれば、それを開いて Copilot を実行し、遊び始めることができるのです。もはや利用可能な数字を理解するだけではなく、勝手に計画を立ててくれる。GitHub Copilot Workspace がプランを作成し、それを実行するように、それはデータアナリストのようなものだ 、Excel を分析のための行や列を視覚化するツールとして使用する。Copilot は、データを生成でき、Python インタプリタを備えているため、あらゆる機能を備えたツールとして Excel を使用するわけです。
>
> - Satya Nadella, CEO, Microsoft, interview with BG2 channel December 2024.[28]

モジュール式のクローズド・システムから、オープン・データを直接扱う AI エージェントへの移行という、オフィス・アプリケーションのロジックに見られる変革は、より大きなプロセスの一部に過ぎない。この変化は、単にインターフェースやソフトウェア・アーキテクチャの変更にとどまらず、仕事の組織化、意思決定、ビジネス・マネジメントの基本原則に影響を与えるだろう。建設業界では、データが設計から資源管理、建設モニタリングに至るプロセスの中心となる、データ主導のロジックにつながるだろう。

次世代デジタルオフィス：AI がワークスペースをどう変えるか

ほぼ 1 世紀前、人類はすでに同様の技術革命を経験していた。蒸気機関から電気モーターへの移行には 40 年以上かかったが、最終的には、主にエネルギー容量の分散化と新しいソリューションの柔軟性によって、前例の

ない生産性の伸びをもたらした。この転換は歴史の流れを変え、人口の大部分を農村部から都市部へと移動させただけでなく、近代経済の基礎を築いた。テクノロジーの歴史は、肉体労働からオートメーションとインテリジェント・システムへの旅である。トラクターが土を耕す何台もの耕運機に取って代わったように、現代のデジタル技術は、伝統的な事務所ベースの建設管理手法に取って代わりつつある（図 10.1-3）。20 世紀初頭、1930 年代に機械やトラクターによる労働の機械化が始まるまで、世界人口のほとんどが手作業で土地を耕していた。

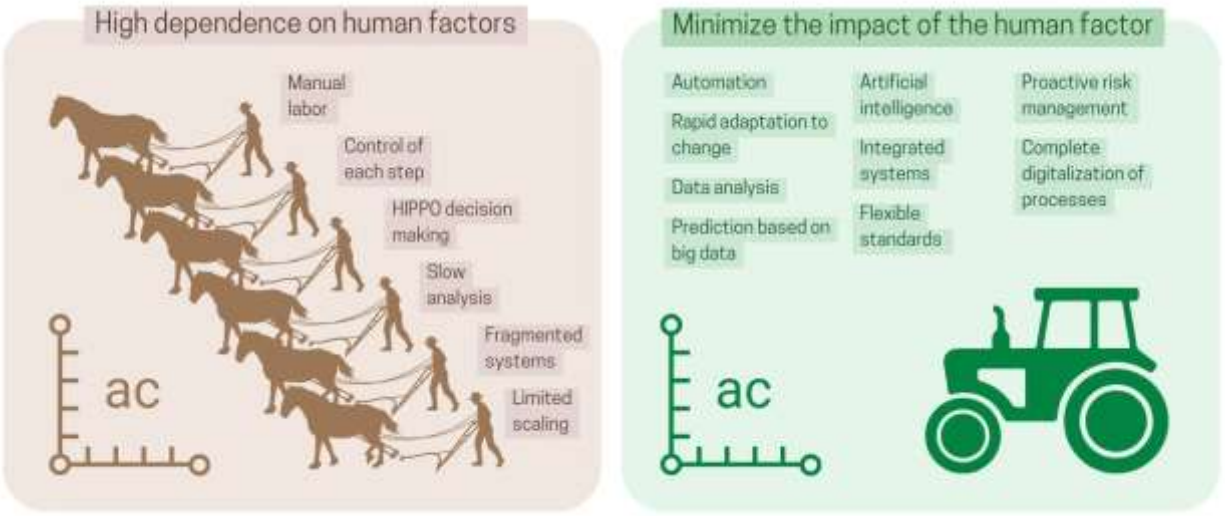


図 10.1-3 20 世紀初頭にトラクターが何十人もの人間に取って代わったように、21 世紀には機械学習が従来のビジネスやプロジェクト管理の手法に取って代わるだろう。

100 年前の人類が、原始的な道具を使った個々の土地の耕作から、機械を使った大規模農業へと移行したように、今日、私たちは、情報のバラバラの「サイロ」の処理から、ETL パイプラインと人工知能アルゴリズムという強力な「トラクター」の助けを借りて、データのアレイを扱うことへと移行しつつある。

従来の手作業による経営管理からデータ主導のモデルへ。

本格的なデータ・ドリブン・アーキテクチャーへの道には、時間と投資、組織的な努力が必要である。しかし、この道は徐々に改善されるだけでなく、建設プロセスの効率性、透明性、管理性の向上へと質的に飛躍する道を開く。これらすべては、デジタルツールの体系的な導入と、時代遅れのビジネス慣行の放棄が前提となる。

> タスクのパラメータ化、ETL、LLM 、IoT コンポーネント 、RFID 、トークン化、ビッグデータ、機械学習は、従来の建設をデータ駆動型建設に変え、プロジェクトと建設事業のあらゆる細部がデータによって制御され、最適化される。

以前は情報を分析するのに何千時間もの工数がかかっていた。今では、こうした作業はアルゴリズムや LLM によって実行され、プロンプトを使って異種のデータセットを戦略的な情報源に変える。技術の世界では、農業と同じことが起こっている。鍬から自動化された農業ビジネスへと移行しているのだ。エクセルファイルと手作業による要約から、データが収集され、クリーニングされ、構造化され、洞察に変わるインテリジェントなシステムに移行しつつあるのだ。

すでに今日、企業は質の高いデータ収集と情報構造化によって情報フィールドを「耕し」、クリーニングと正規化ツールで「肥やし」にし、予測分析と自動化ソリューションの形で「収穫」する必要がある。現代の農家が機械によって 100 人の土を耕す人の代わりになれるとすれば、インテリジェントなアルゴリズムは従業員からルーティンを取り除き、情報の流れの戦略的管理者の役割に移すことができるだろう。

しかし、真にデータ主導の -組織を作ることは、すぐにできるプロセスではないことを理解することが重要である。それは、システムの新しい森（図 1.2-5）を育てるための新しい場所を作るのと同じような、長期的な戦略的方向性である。この生態系におけるそれぞれの「木」は、成長し発展するのに時間を必要とする個別のプロセス、能力、またはツールである。実際の森と同じように、成功するかどうかは、植える材料（技術）の質だけでなく、土壌（企業文化）、気候（ビジネス環境）、手入れ（システム・アプローチ）にも左右される。

企業はもはや、クローズドで既成概念にとらわれないソリューションだけに頼ることはできないだろう。これまでの技術開発の段階とは異なり、オープン・データ・アクセス、人工知能、オープン・ソースといった現在の移行は、大手ベンダーがサポートする可能性は低い。

第 4 および第 5 の技術革命の章ですでに述べたハーバード・ビジネス・スクールの研究[40]が示すように、最も利用されているオープン・ソース・ソリューションをすべての企業がゼロから作成した場合のコストは、約 41 億 5,000 万ドルになる。しかし、各企業が既存のオープン・ソース・ツールにアクセスすることなく、独自の代替案を開発すると想像すると、過去数十年間起こってきたことだが、ビジネス・コストの総額は 8 兆 8000 億ドルという途方もない額に達する可能性がある。これは、 、ソフトウェア市場が評価できる非合理的な需要の価格である。

技術の進歩は、必然的に既存のビジネスモデルの見直しにつながる。かつて企業は、複雑で不透明なプロセスやクローズドなデータから収益を上げることができたが、AI（ ）やアナリティクス（ ）の発展により、このア

プローチはますます通用しなくなってきている。

> データやツールへのアクセスが民主化された結果、従来のソフトウェア販売市場は大幅に縮小する可能性がある。しかし同時に、デジタル専門知識、カスタマイズ、統合、ソリューション・デザインの市場という新たな市場も成長するだろう。ここでは、ライセンス販売ではなく、柔軟でオープンで適応可能なデジタル・プロセスを構築する能力から価値が生まれるだろう。電化やトラクターの登場が新しい産業を生み出したように、ビッグデータ、AI（）、LLM（）の応用は、建設業界のビジネスにまったく新しい地平を切り開くだろう。そのためには、技術投資だけでなく、考え方、プロセス、組織構造の大幅な変革が必要になる。そして、このことに気づき、今日行動を開始する企業や専門家が、明日のリーダーとなるだろう。

オープンデータが主要な資産となりつつある世界では、情報の入手可能性はゲームチェンジャーとなるだろう。投資家、顧客、規制当局はますます透明性を求めるようになり、機械学習アルゴリズム、見積もり、スケジュール、コストの食い違いを自動的に特定できるようになるだろう。これにより、デジタル変革の新たな段階への条件が整い、建設業界の「ユーバー化」が徐々に進んでいる。

オープンデータとウーバライゼーションは既存の建設ビジネスにとって脅威である

建設は情報管理プロセスになりつつある。データがより正確で、質が高く、完全であればあるほど、建築物の設計、計算、積算、建設、運営はより効率的になる。将来、重要な資源は、クレーン、コンクリート、鉄筋ではなく、情報を収集し、分析し、利用する能力になるだろう。

> 建設会社の顧客（投資家や建設資金を調達する顧客）は今後、オープンデータやアナリティクス（）の価値を活用することは必至である。これによって、建設会社を原価計算の問題に巻き込むことなく、プロジェクトのタイムラインとコストの計算を自動化する機会が開かれるだろう。、コストを管理し、余分なコストをより迅速に特定するのに役立つ。

レーザースキャナー、クアドロコプター、写真測量システムが、使用されたコンクリート量に関する正確なリアルタイムデータを収集する建設現場を想像してみてほしい。この情報は、複雑な幾何学カーネル（）、ERP（）、PMIS（）に依存することなく、煩雑な CAD（）システム（BIM（））をバイパスして、メタデータ付

きのシンプルなフラット MESH（ ）モデルに自動的に変換される。建設現場から収集されたこのデータは、独立した分析のために、クライアントが利用できる単一の構造化されたレポジトリに集中的に転送され、さまざまな建設店からの実際の価格がアップロードされ、信用融資の割合から、気象条件、建設資材の証券取引所の相場、物流関税、労働価格の統計的な季節変動などの動的に変化する要因に至るまで、さまざまなパラメータが表示されます。このような条件下では、設計と実際の資材量の食い違いが即座に明らかになるため、設計段階でもプロジェクト引き渡し時でも見積もりを操作することが不可能になる。その結果、建設プロセスの透明性は、監督者や管理者の軍隊によってではなく、客観的なデジタルデータによって達成される。

将来、このようなデータ管理業務は、むしろ顧客側のデータ管理者が行うようになるだろう（図 1.2-4 CQMS manager）。これは特に、 、プロジェクトの見積もりに当てはまる。以前は見積もり担当の部署があったのが、明日はすでに機械学習 、建設会社が適合する価格の上限を設定する予測ツールが存在するだろう。

> システムやサブシステムのほとんどが中小企業によって供給されている建設業界の細分化された性質を考えると、デジタル戦略は顧客から生まれなければならない。顧客は、サプライチェーンのデジタル能力を引き出すための条件と仕組みを作らなければならない[20]。
>
> - アンドリュー・デイヴィス、ジュリアーノ・デニコル（アクセンチュア）"資本プロジェクトを通じてさらなる価値を創造する"

このようなデータのオープン性と透明性は、不透明なプロセスや紛らわしい報告書から利益を得ることに慣れている建設会社にとって脅威となる。建設会社では、複雑で閉鎖的なフォーマットやモジュール式の専有データプラットフォームの背後に、思惑や隠れたコストが隠されている可能性がある。そのため、建設会社は、オープンソース・ソリューションを推進するベンダーと同様、オープンデータを自社のビジネスプロセスに完全に導入することに関心を持ちにくい。データが利用可能で、顧客にとって処理しやすいものであれば、自動的にチェックすることができるため、数量を過大評価したり、見積もりを操作したりする可能性がなくなる。

> 世界経済フォーラムの報告書「Shaping the Future of Construction」（2016 年）[5]によると、建設業界にとって重要な課題のひとつは、依然として顧客の受動的な役割である。とはいえ、早期計画から持続可能な相互作用モデルの選択、パフォーマンスのモニタリングに至るまで、プロジェクトの結果に対してより大きな責任を負うべきなのは顧客である。プロジェクト・オーナーの積極的な参加なくして、建設業界のシステマティックな変革は不可能である。

量とコスト計算のコントロールができなくなったことで、過去 20 年間、他の業界はすでに変貌を遂げ、顧客は仲介者を介さずに直接、、目標を阻止できるようになった。デジタル化とデータの透明性は、Uber の登場によるタクシー運転手（図 10.1-4）、Airbnb の登場によるホテル経営者、アマゾンの台頭による小売業者や商店、ネオバンクや分散型フィンテック・エコシステムの台頭による銀行など、多くの伝統的なビジネスモデルを変革し、情報への直接アクセスや時間・コスト計算の自動化によって、仲介者の役割が大幅に減少している。

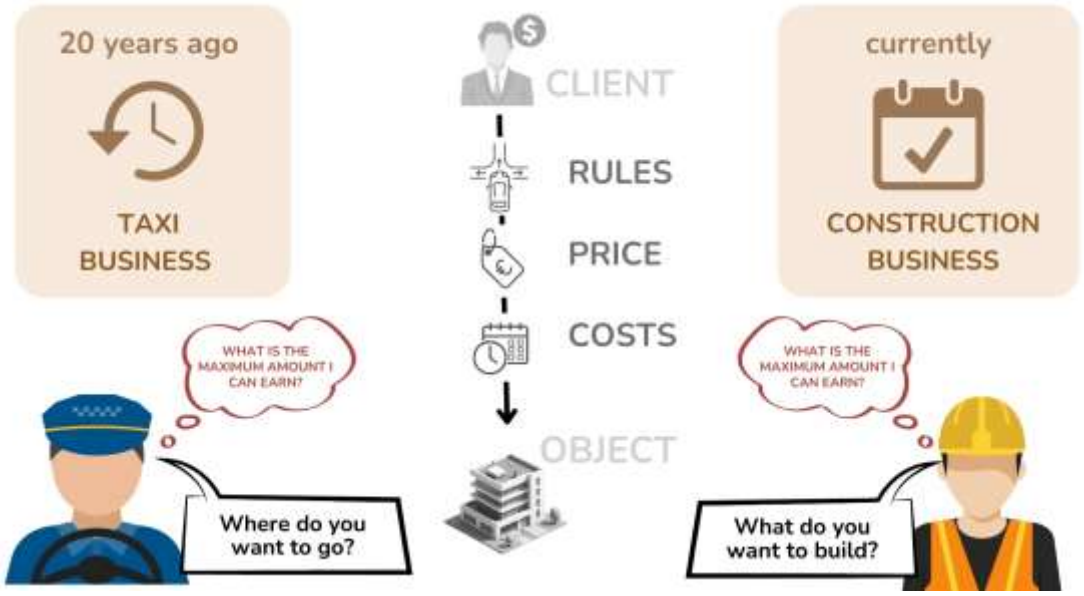


図 10.1-4 建設事業は、10 年前にタクシー運転手、ホテル経営者、小売業者が直面しなければならなかったユーバーライゼーションに直面するだろう。

データとその処理ツールへのアクセス民主化のプロセスは不可避であり、時間の経過とともに、すべてのプロジェクト構成要素に関するオープンデータ、顧客の要件となり、新たな標準となるだろう。したがって、オープン・フォーマットと透明性のある計算の導入という問題は、投資家（）、顧客、銀行、プライベート・エクイティ・ファンド（）、つまり、建設された対象物のエンド・ユーザーであり、その後数十年にわたって対象物を運用する人々によって推進されるだろう。

> 主要な投資家、顧客、銀行は、建設業界においてすでに透明性を求めている。アクセンチュアの調査「資本プロジェクトを通じてより多くの価値を創造する」（2020 年）[20]によると、透明で信頼性の高いデータは、建設業における投資判断の決め手となりつつある。専門家が指摘するように、信頼できる効果的なプロジェクト管理は、透明性なしには不可能であり、特に危機的状況下ではなおさらである。さらに、投資家 、銀行、規制当局から説明責任と透明性に対する要求 が高まっていることを反映し、アセットオーナーや請負業者は、データ共有や共同分析にインセンティブを与える契約への移行を強めている。

> 投資家、つまり顧客のアイデアから完成した建物までの移動は、将来的には自動操縦で移動するようなもので、建設会社という運転手がいなければ、投機や不確実性から独立することが約束される。

オープンデータとオートメーションの時代は、銀行、商業、農業、物流がすでにそうであったように、建設ビジネスも必然的に変えていくだろう。これらの業界では、仲介業者の役割や伝統的なビジネスのやり方が自動化やロボット化に取って代わられ、不当な値上げや投機の余地がなくなっている。

人間のあらゆる経済活動におけるデータとプロセスは、建設業界の専門家が対処しなければならないことと何ら変わりはない。長期的には、今日、価格とサービスの品質基準を設定することで市場を支配している建設会社は、顧客と建設プロジェクトの間の重要な仲介者としての役割を失うかもしれない。

ユーバー化の未解決問題 変革のために時間を使う最後のチャンスとして

しかし、建設業界の現実に話を戻そう。自動運転車、分散型金融システム、人工知能ベースのソリューションが経済の一部分野で台頭している一方で、 、建設会社のかなりの割合は依然として紙ベースの組織であり、重要な意思決定は個々の専門家の直感や経験に基づいて行われている。

このパラダイムでは、現代の建設会社は、リソース、ルート、納期を管理し、プロジェクトのアイデア（ロジスティクスと設置プロセス）からプロジェクトの納品までの「旅」のタイミングとコストに責任を持つ 20 年前のタクシー会社に例えることができる。かつて GPS（建設 IoT 、RFID ）や時間／コスト計算の機械学習アルゴリズムが輸送分野を変革したように、データ、アルゴリズム、AI -エージェントは、直感的な評価から予測的なガイド付きモデルへと、建設管理を変革する可能性を秘めている。過去 20 年間、金融、農業、小売、物流など多くの業界では、データの不透明性を利用した投機能力が徐々に失われてきた。価格、配送コスト、あるいは

金融取引は、デジタル・プラットフォーム上で、わずか数秒のうちに、自動的に、統計的に正しく計算される。

将来を見据えて、建設会社は、データへのアクセスとそれを分析するツールの民主化が、プロジェクトのコストと期間を見積もる従来のアプローチを破壊し、不透明な数量と価格のデータから推測する機会をなくすことを認識する必要がある。

> 自動化された時間とコストの見積もり、透明性の高いタスクのルーティング、人的要因への最小限の依存。これは、アイデアから実現までの「旅」の本質を変え、より予測可能で、管理しやすく、データ駆動型にする。

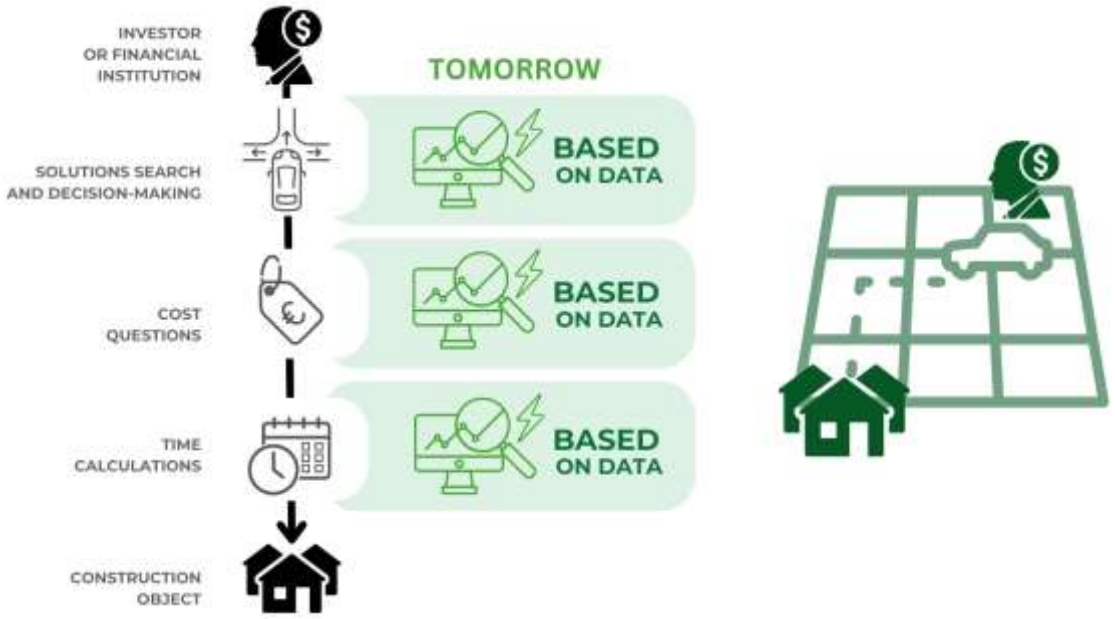


図 10.1-5 建設期間中のコストと「移動」時間は、機械学習 と統計ツールを使って決定される。

世界のほぼすべての国で、建設プロジェクトに融資する顧客や銀行に CAD- (BIM)-モデルの転送を義務付ける新しい規制や要件が徐々に導入されているため、顧客と発注者は、コストとスコープ計算の透明性を独自に確保する機会を得ている。これは、数量を迅速に分析し、市場価格を監視する十分な能力とツールを有する大口顧客や投資家（ ）にとって特に重要である。店舗、オフィスビル、集合住宅など、大規模で標準的なプロジェクトを実施する企業にとって、このような慣行は標準になりつつある。

モデルの情報内容がより完全で標準化されるにつれ、操作や推測の可能性は事実上なくなっていく。デジタルトランスフォーメーションは、建設業界のゲームのルールを徐々に変えつつあり、こうした変化に適応しない

企業は、深刻な課題に直面する可能性がある。

> 競争の激化、テクノロジーの破壊、利幅の縮小は、ビジネスの持続可能性に影響を与える可能性がある。流動性が限られる中、効率性とプロセスの透明性を向上させる方法として、自動化、分析、データ・テクノロジーに注目する業界関係者が増えている。これらのツールは、変化する経済環境の中で競争力を維持するための重要なリソースになりつつある。

デジタル・コンピテンシーを強化し、最新のソリューションを導入し、データ中心の企業文化を構築することによって、今日から準備を始める方がはるかに効果的なのだ。

建設業界の大規模なデジタル変革に対する最後の重要な技術的障壁のひとつであり、今後数年間にあらゆる企業に影響を及ぼすであろう、建設プロジェクトの 要素の自動分類の問題である。

> 信頼性が高く、正確でスケーラブルな分類ができなければ、AI や予測モデルを使った本格的な分析、プロセスの自動化、ライフサイクル管理の基盤を構築することは不可能だ。対象物の分類が、現場監督、設計者、積算担当者など、経験豊富な専門家による手作業による解釈に依存している限り、建設業界にはまだチャンスの窓がある。透明性への要求の高まり、ツールやデータの民主化、ゲームのルールを根本的に変える自動分類システムの出現といった避けられない変化に備えるために、この時間を使うことができる。

建設世界の自動的に分類するタスクは、その複雑さにおいて、無人運転システムにおける物体認識に匹敵し、これは主要な課題の 1 つである。A 地点から B 地点へ移動する無人自動車を想像してみよう（図 10.1-5）。現在の自動運転システムは、ライダーやカメラで認識された物体を分類する問題に頭を悩ませている。車が障害物やランドマークを「見る」だけでは不十分で、それが歩行者なのか、道路標識なのか、ゴミ箱なのかを認識できなければならない。

同様の基本的な課題は、建設業界全体が直面している。窓、ドア、柱などのプロジェクト要素は、文書に記録したり、CAD モデルで表現したり、建設現場で撮影したり、レーザースキャンによる点群データで認識したりすることができる。しかし、真に自動化されたプロジェクト管理システムを構築するには、視覚的あるいは大まかな幾何学的認識だけでは不十分です。各要素を正確かつ一貫して、見積もりや仕様書からロジスティクス、在庫管理、そして最も重要なオペレーションに至るまで、その後のすべてのプロセスで一意に識別できるタイプに分類することが必要である（図 4.2-6）。

認識から意味のある分類への移行というこの段階で、重要な障害のひとつが発生する。デジタル・システムが技術的にモデルや建設現場のオブジェクトを区別・識別できるとしても、主な困難は、異なるソフトウェア環境に対して、要素タイプを正しく、文脈的に安定して定義することにある。。例えば、設計者が CAD モデルで "ドア "というカテゴリの要素としてラベル付けしたドアが、ERP や PMIS システムに転送されると、設計者のミスやシステム間の不整合により、誤入力されることがあります。さらに、データのエクスポートやインポートの際に、その要素が重要な属性を失ったり、システム会計から消えてしまったりすることもよくあります。これはデータの流れにギャップをもたらし、建設プロセスのエンド・ツー・エンド・デジタル化の原則を損なう。これは、「目に見える」意味と「理解できる」意味との間に決定的なギャップを生み、データの完全性を損ない、建設プロジェクトのライフサイクル全体を通じてプロセスの自動化を著しく複雑にしている。

ビッグデータと機械学習技術 （図 10.1-6）を用いた、 建築要素の普遍的な分類という課題に対する解決策は、業界全体の変革のきっかけとなり、おそらく多くの建設会社にとって予期せぬ発見となるだろう。統一された学習可能な分類システム は、スケーラブルな分析 、デジタル管理、建設組織の日常業務への AI の導入の基盤となる。

エヌビディアや他のテクノロジー・リーダーは、膨大な量のテキスト情報や視覚情報を自動的に分類・構造化するソリューションを、他の産業ですでに提供している。

> 例えば NVIDIA の NeMo Curator モデル[161]は、データを定義済みのカテゴリーに自動的に分類・分類することに特化しており、生成 AI モデル の微調整と事前学習のための情報処理パイプラインの最適化において重要な役割を果たす。Cosmos プラットフォームは、実世界のビデオや 3D [162]のシーンで学習され、NVIDIA エコシステムですでに構築されている自律システムやデジタルツインの基盤を提供する。NVIDIA Omniverse は、2025 年までに、設計情報伝達プロセスにおいて最終的に IFC フォーマットに取って代わる可能性のあるユニバーサルシーン記述である USD フォーマットで作業するための主要ツールとなった。NeMo Curator、Cosmos、Omniverse などのソリューションは、ロボット・プロセス・シミュレータである Isaac Sim [163]とともに、データのクリーニングやフィルタリングから、トレーニング・セットの生成、オブジェクト・プロパティのモデリング、建設現場でのロボット・トレーニングに至るまで、自動化の新たなレベルを示している。また、これらのツールはすべて無料でオープンソースであるため、エンジニアリングや建設現場での導入障壁が大幅に軽減される。

構造化された表レベルでのデータの自動分類は、一見したところそれほど難しくはない。前章で示したように（図 9.1-10）、過去のデータが蓄積されていれば、他の要素の類似したパラメー タに基づいて、欠落している、あるいは誤ったクラス値を補うことが可能です。類似した特性を持つ要素が、すでに完成した複数のプロジ

ェクトで正しく分類されている場合、シス テムは高い確率で、新規または未完成の要素に適した値を提案することができます（図 10.1-6）。平均値と文脈分析に基づくこのようなロジックは、見積書、仕様書、CAD モデルから得られる表データを大量処理する場合に特に効果的です。

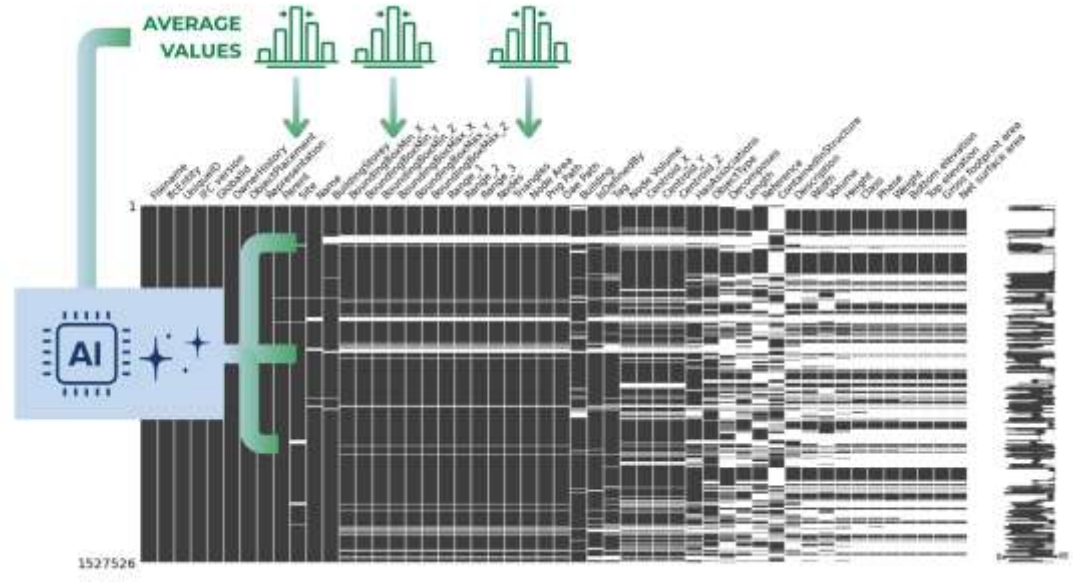


図 10.1-6 機械学習 は、過去のプロジェクトに基づいて、未記入（白枠）の表パラメータの平均値を自動的に見つけるのに役立ちます。

> このような機械学習の急速な進歩を背景にすれば、2025 年において、建築要素を自動的に分類する問題がいつまでも未解決のままであると考えるのは甘いということは明らかである。確かに、現在のアルゴリズムはまだ完全な成熟には至っておらず、特にデータが不完全であったり、異種データであったりする場合はそうであるが、適応の機会の窓は急速に閉ざされつつある。

すでにデータの収集、クリーニング、整理に投資し、ETL 自動化ツールを導入している企業が有利になるだろう。それ以外の企業は遅れをとるリスクがある。かつて運輸や金融業界で企業がデジタルトランスフォーメーションの課題に対応できなかったように

手作業によるデータ管理や伝統的なコストと時間の見積もり方法に依存し続ける企業は、2020 年代初頭までにモバイルアプリや自動ルート計算の時代に適応できず、2000 年代のタクシー会社のような立場に陥る危険性がある。

章 10.2.

データ主導のアプローチを導入するための実践的ガイド

理論から実践へ：建設業におけるデジタルトランスフォーメーションのロードマップ

建設業界は徐々に新たな発展段階を迎えており、慣れ親しんだプロセスはますますデジタル・プラットフォームや透明性の高いインタラクション・モデルによって補完され、時には取って代わられつつある。これは、企業にとって課題であると同時に、大きなチャンスでもある。現在、すでに長期的なデジタル戦略を構築している企業は、市場での地位を維持できるだけでなく、顧客に最新のアプローチと信頼性の高い技術的裏付けのあるソリューションを提供することで、その地位を拡大することもできるだろう。

> 概念や技術の知識は出発点に過ぎないことを認識することが重要である。経営者や専門家は、どこから実行に移せばいいのか、理論的なアイデアを実際の価値に変えるにはどうすればいいのかという現実的な問題に直面する。さらに、従来の原価計算やタイミングが顧客によっていつでも変更される可能性があるのであれば、そのビジネスは何を基礎とするのかという疑問もますます生じてくる。

その答えはおそらく、技術にあるのではなく、データを扱うことが日常業務の不可欠な一部として認識されるような、新しい職業文化の形成にある。デジタル技術とイノベーションへの関心の欠如が、建設業界を過去数十年にわたって観察されてきた深刻な後進性に陥れているのだ[43]。

> マッキンゼーによると、建設業における研究開発費は売上高の 1%未満であるのに対し、自動車産業や航空宇宙産業では 3.5〜4.5%に達する。同様に、建設業における IT コストも総収入の 1%未満にとどまっている[107]。

その結果、建設業では自動化のレベルだけでなく労働生産性も低下しており、2020 年には建設労働者の生産量は半世紀前よりも減少している（図 10.2-1）

このような建設部門の生産性の問題は、ほとんどの先進国・発展途上国に共通しており（OECD 加盟 29 カ国中 16 カ国で建設生産性が低下している（図 2.2-1））、技術不足だけでなく、マネジメント、トレーニング、イノベーションへのアプローチそのものを体系的に変える必要性を指摘している。

デジタルトランスフォーメーションの成功は、ツールの数や利用可能性よりも、組織のプロセスを見直し、変化を受け入れる文化を発展させる能力にかかっている。重要なのはテクノロジーそのものではなく、その効果的な利用を保証し、継続的な学習を支援し、新しいアイデアの受け入れを促す人材とプロセスなのだ。

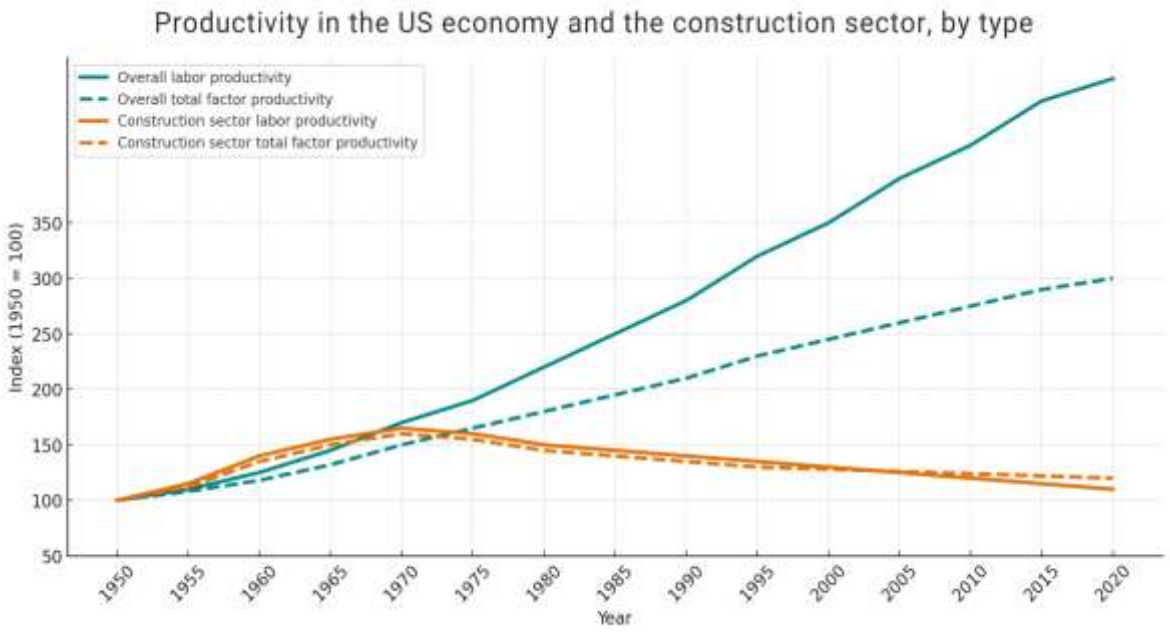


図 10.2-1 米国経済と建設部門における労働生産性と総資源生産性のパラドックス（1950-2020年）（[43]に基づく）。

本書の序盤では、ビジネス環境モデルを森林生態系に例えた（図 2.1-2、図 1.2-4、図 1.3-2）。健全な森林では、定期的に行われる火災が、その破壊力の大きさにもかかわらず、長期的な更新において重要な役割を果たしている。火災は、古い植生を土壌から除去し、蓄積された養分を戻し、新たな生命を育む空間を作り出す。植物種の中には、火災による高温にさらされたときにのみ種子が開くように進化したものさえある。

ビジネスにおいても同様に、危機は「コントロールされた燃え尽き症候群」として機能し、時代遅れのシステムに縛られない新しいアプローチや企業の出現を促す。そのような時期は、非効率的な慣行を放棄させ、イノベーションのためのリソースを解放する。火事の後の森が先駆的な植物から始まるように、危機の後のビジネスも、成熟した情報環境の基礎となる新しい柔軟なプロセスを形成する。

こうした「シグナル・ファイア」を正しく解釈し、破壊的なエネルギーを建設的な変化へと転換させることに

成功した企業は、新たなレベルの業績を達成することができる。

> 人工知能）と機械学習がビジネス環境に与える影響の拡大は、もはや疑う余地がない。これは単なる一時的なトレンドではなく、戦略的に必要なことなのだ。AI を無視する企業は、イノベーションと柔軟性がますます奨励される市場で競争力を失うリスクがある。

未来は、AI を単なるツールとしてではなく、プロセスの最適化から経営判断に至るまで、ビジネスのあらゆる側面を見直す機会として捉える人たちのものだ。

デジタル基盤の構築：デジタル成熟への 1-5 ステップ

本章では、デジタルトランスフォーメーションのロードマップに注目し、企業文化と企業の情報エコシステムの両方を変革するのに役立つデータ主導のアプローチを導入するために必要な主要ステップを明らかにする。

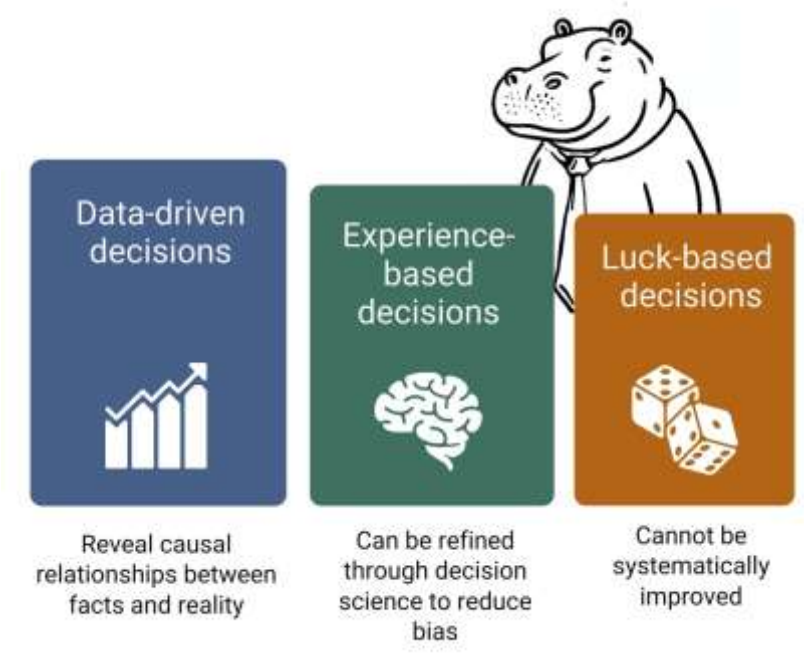


図 10.2-2 管理された更新と戦略の選択：事例、経験、またはデータ。

マッキンゼーの調査「デジタル戦略が失敗する理由」（2018 年）によると、企業がデジタルトランスフォーメーションの目標を達成できない理由は、少なくとも次の 5 つである[164]。

- 定義の曖昧さ：経営幹部と管理職は「デジタル」の意味について異なる理解を持っており、誤解や矛

盾が生じている。

- デジタル経済の誤解：多くの企業は、デジタル化がビジネスモデルや産業力学にもたらす変化の規模を過小評価している（図 10.1-6）。
- エコシステムの無視：企業は個々のテクノロジー・ソリューション（データ・サイロ）に集中し、より広範なデジタル・エコシステムへの統合の必要性を見落としている（図 2.2-2、図 4.1-12）。
- 競合他社によるデジタル化の過小評価：経営者は、競合他社もデジタル技術を積極的に導入していることを考慮していない。
- デジタル化の二面性を見逃している：最高経営責任者（CEO）はデジタル変革の責任を他の経営幹部に委ねるが、これは統制を官僚化し、変革プロセスを遅らせる。

こうした課題に対処するには、組織のあらゆるレベルでデジタル戦略を明確に理解し、整合させる必要がある。デジタル戦略を構築する前に、出発点を理解することが重要である。多くの組織は、現状を完全に把握しないまま、新しいツールやプラットフォームを導入する傾向がある。

ステップ 1：現在のシステムとデータの監査を行う。

プロセスを変更する前に、すでに何が実施されているかを理解することが重要である。監査を実施することで、データ管理の弱点を特定し、どのようなリソースを活用できるかを理解することができる。監査は、ビジネスプロセスの「X 線検査」のようなものです。これにより、リスクのある領域を特定し、プロジェクトやビジネスにとってどのデータが重要で、どのデータが二次的なものかを判断することができる。

主なアクション

- IT 環境のマップを作成する（Draw.io、Lucidchart、Miro、Visio、Canva を使用）。近代建設における技術と管理システム」の章（図 1.2-4）で説明した、プロセスに関与する使用システム（ERP 、CAD 、CAFM 、CPM 、SCM など）をリストアップ
- 各システムにおける重複の頻度、欠損値の可能性、フォーマットの不整合について、各システムのデータ品質の問題を評価する。
- インポート、エクスポート、追加の検証プロセスなど、プロセスが破綻したり、しばしば手作業が必要になったりする「ペインポイント」を特定する。

> チームにレポートを信頼してもらいたいなら、最初からデータが正しいことを確認する必要がある。

質の高いデータ監査は、どのようなデータがあるかを示してくれる：

- さらなる開発が必要（自動洗浄プロセスまたは追加の変換を設定する必要がある）
- それらはシステムを詰まらせるだけの "ゴミ "であり、もうプロセスで使用しないことで取り除くことができる。

このような監査は、自社で行うことも可能である。建設業界の「特殊性」から独立した新鮮な視点は、現状を冷静に評価し、特定のソリューションや技術に偏るという典型的な落とし穴を避けるのに役立つ。

ステップ 2：データ調和のための主要基準を特定する。

監査後、データを扱うための共通ルールを作る必要がある。標準：ランダムファイルからインテリジェントなデータモデルへ」の章で述べたように、これはサイロ化したデータフローを排除するのに役立つ。

> 単一の標準がなければ、各チームは「それぞれのやり方」で作業を続け、変換のたびにデータが失われる統合の「動物園」を維持することになる。

主なアクション

- システム間で情報を交換するためのデータ標準を選択する：
  - 表形式データの場合、CSV や XLSX のような構造化フォーマットや、Parquet のようなより効率的なフォーマットを使用することができます。
  - 緩く構造化されたデータやドキュメントの交換： JSON または XML
- データモデルを使いこなす：
  - データモデリング：概念モデル、論理モデル、物理モデル」（図 4.3-2）で説明されているように、概念データモデルのレベルでタスクをパラメータ化することから始める。
  - ビジネスプロセスロジックを深く検討するにつれて、論理モデルと物理モデルのパラメータを使用して、要件を形式化することに移ります（図 4.3-6 参照）。

- 主要なエンティティ、その属性、およびプロセス内の関係を特定し、エンティティ間およびパラメー タ間の関係を可視化する（図 4.3-7）。

- 正規表現（RegEx ）を使用してデータを検証し標準化する（図 4.4-7）「構造化要件 と RegEx 正規表現」の章で説明したとおりである。RegEx は複雑なものではありませんが、物理データモデルのレベルで要件を作成する作業では非常に重要なトピックです。

データレベルの標準化とプロセスの可視化なくして、一貫性と拡張性のあるデジタル環境を提供することは不可能である。覚えておいてほしいのは、「悪いデータは高くつく」ということだ。プロジェクトや組織が複雑になればなるほど、エラーのコストは増大する。フォーマットを統一し、命名、構造、検証ルールを定義することは、将来のソリューションの安定性と拡張性への投資である。

ステップ 3. DataOps を導入し、プロセスを自動化する。

明確に定義されたアーキテクチャがなければ、企業は必然的に、サイロ化した情報システムに含まれる異種データに直面することになる。データは統合されず、複数の場所に重複し、維持にコストがかかる。

> データは水であり、データアーキテクチャはその水を貯蔵源から使用地点まで輸送するパイプラインの複雑なシステムであると想像してほしい。情報がどのように収集され、保存され、変換され、分析され、エンドユーザーやアプリケーションに提供されるかを決定するのがデータアーキテクチャである。

DataOps (Data Operations)は、本書のパート 8 で詳しく説明したように、データの収集、クレンジング、検証、利用を単一の自動化されたプロセスフローに統合する方法論である。

主なアクション

- プロセスを自動化するために ETL -コンベアを作成し、カスタマイズする：

  - Extract: PDF ドキュメント（Fig. 4.1-2、Fig. 4.1-5、Fig. 4.1-7）、Excel スプレッドシート、CAD -モデル（Fig. 7.2-4）、ERP -システム、その他お客様が扱うソースからの自動データ収集。

  - Transform: データを単一の構造化されたフォーマットに変換する自動プロセスを設定し、閉じたアプリ ケーションの外部で行われる計算を自動化する（図 7.2-8）。

- ロード: 要約表、文書、または一元管理されたリポジトリへの自動データアップロードを作成してみる（図 7.2-9、図 7.2-13、図 7.2-16）

- QTO Quantity Take-Off: Grouping Project Data by Attributes」の章で説明したように、計算と QTO (Quantity Take-Off)のプロセスを自動化する：
  - API、プラグイン、またはリバースエンジニアリングツールを使用して、CAD -モデルから体積の自動抽出を設定する（図 5.2-5）。
  - 異なるクラスの要素を属性でグループ化するルールを表の形で作成する（図 5.2-12）
  - モジュール式クローズドシステム以外では、頻繁に繰り返される数量計算とコスト計算の自動化を試みる（図 5.2-15）
- Python Pandas: an indispensable tool for working with data " の章で説明したように、Python と Pandas を使ってデータ処理を開始する：
  - DataFrame を適用して XLSX ファイルを操作し、表形式データの処理を自動化する (図 3.4-6)
  - 様々な Python ライブラリを使って、情報の集約と変換を自動化する。
  - LLM を使用すると、既製のコードブロックやパイプライン全体の記述が簡単になります (図 7.2-18)
  - Pipeline を Python で作ってみてください。このパイプラインは、エラーを発見し たり、異常を発見したりすると、責任者（プロジェクトマネージャーなど）に通知を送ります（図 7.4-2）。

DataOps の原則に基づく自動化により、手作業による断片的なデータ処理から、持続可能で反復可能なプロセスへの移行が可能になります。これにより、毎日同じ変換を行う従業員の負担が軽減されるだけでなく、情報システム全体の信頼性、拡張性、透明性が飛躍的に向上します。

ステップ 4：オープンデータガバナンスのエコシステムを構築する。

クローズド・モジュラー・システムの開発と新しいツールとの統合にもかかわらず、企業は深刻な問題に直面している。すべてのビジネス・プロセスをカバーする単一の独自プラットフォームを構築するという当初のアイデアは、過度の集中化を招き、いかなる変更にも対応するために多大なリソースと時間が必要となっている。

企業の菌糸体：データがビジネスプロセスをつなぐ仕組み」の章で述べたように、効果的なデータ管理には、あらゆる情報源をつなぐオープンで統一されたエコシステムが必要である。

エコシステムの主要要素：

- 適切なデータストアを選択する：
  - テーブルや計算にはデータベースを使用する。例えば PostgreSQL や MySQL などである（図 3.1-7）。
  - 文書やレポートについては、クラウドストレージ（Google Drive、OneDrive）や JSON 形式をサポートするシステムが適している場合があります。
  - データウェアハウスやデータレイクなど、大量の情報を一元的に保管・分析するツールの機能をチェックする（図 8.1-8）。
- 専有データにアクセスするためのソリューションを導入する：
  - 独自のシステムを使用している場合は、API または SDK を介してアクセスできるように設定し、外部処理用のデータを取得する（図 4.1-2）。
  - リバースエンジニアリングツールの可能性をよく理解する。 CAD フォーマットに対応（図 4.1-13）
  - アプリケーションやサーバーから定期的にデータを収集し、オープンな構造化フォーマットに変換し、 リポジトリに保存する ETL パイプラインを設定する（図 7.2-3）。
  - プロプライエタリなソフトウェアを必要とせずにデータへのアクセスを提供する方法をチーム内で話し合う。
  - 覚えておいてほしいのは、データはインターフェイスよりも重要だということだ。長期的な価値を提供するのは、情報の構造と可用性であり、特定のユーザー・インターフェース・ツールではない。
- データモデリングのための卓越したセンター（CoE）」の章で述べたように、データのための卓越したセンター （CoE ）の設立、あるいは他の方法でデータの専門知識を提供する方法を考える（図 4.3-9）。

データマネジメント・エコシステムは、すべてのプロジェクト参加者が一貫性のある最新かつ検証された情報を使って作業できるよう、統一された情報空間を作り出します。これは、スケーラブルで柔軟かつ信頼性の高

いデジタル・プロセスの基盤です

データの可能性を解き放つ：デジタル成熟への 5〜10 ステップ

技術的な統合に加え、デジタルソリューションの導入を成功させる重要な要因は、エンドユーザーによる採用である。顧客やユーザーをパフォーマンス測定に参加させることは、ユーザーエクスペリエンスを向上させる課題であると同時に、企業内の変化を管理する課題でもある。ソリューションが使い慣れたワークフローに適合しなかったり、実際のユーザーや顧客の問題を解決できなかったりすれば、そのソリューションは利用されない。

トランスフォーメーションは、新しいプロセスにおけるユーザー・インタラクション・データの分析に基づく反復的なプロセスであり、頻繁なテスト・サイクル、絶え間ないフィードバックと改良が行われる。

ステップ 5：データ文化の構築、スタッフのトレーニング、フィードバックの収集

どんなに先進的なシステムでも、従業員のエンゲージメントがなければ機能しない。データが日常的に使用され、チームがその価値を理解する環境を作る必要がある。

> 英国政府が発表した 2024 年の報告書「Data Analytics and AI in Government Project Delivery」[83] は、 データ分析と AI の導入を成功させるには、データ処理と解釈に必要な能力を備えた専門家の育成が不可欠であると指摘している。

データ分析の専門知識の欠如は、デジタルトランスフォーメーションを制限する重要な問題の一つである。経営幹部は、四半期ごとのサイクル、優先順位をつけた取り組み、プロジェクトを進める伝統的な方法など、確立されたルーチンに慣れている。変革には、影響力を持てるほど地位は高いが、長期的な変革プロジェクトをリードする時間とモチベーションを持てるほど地位は高くない、個性的なリーダーが必要だ。

主なアクション

- ■ 高給取りの従業員の意見に基づく主観的な意思決定（HiPPO ）から、「HiPPO あるいは意思決定における意見の危険性」（図 2.1-9）の章で述べたように、事実とデータに基づく意思決定文化に移行する必要性を認識する。
- ■ 体系的なトレーニングを実施する：
    - ▢ 構造化データの使用方法に関する研修会を開催し、現在の建設業界で普及している製品やコ

ンセプトに偏らない他業界の専門家を招く。

- ☐ データ分析のアプローチやツールについて同僚と議論し、Python、pandas、LLM などのツールを使った実践的な作業を自主的に学ぶ（図 4.1-3、図 4.1-6）。
- ☐ データの構造化（図 3.2-15）とデータモデルの作成（図 4.3-6、図 4.3-7）に関するチュートリアル（できれば短いビデオ付き）のライブラリを作成する。

■ 最新の学習テクノロジーを活用する：

- ☐ 言語モデル（LLM）を使用して、コード生成、リファクタリング、解析、および表形式情報の処理と解釈を含む、コードとデータ操作をサポートする（図 3.4-1）
- ☐ オフラインの開発環境（IDE）で作業する場合、LLM が生成したコードをどのように適合させ、完全な Pipeline ソリューションに統合できるかを探る（図 4.4-14、図 5.2-13）。

> マネジャーが "昔ながらの方法 "で意思決定を続けている場合、いくらトレーニングを受けても、アナリティクスを真剣に受け止めるように人々を説得することはできない。

データ文化の構築は、継続的なフィードバックなしには不可能である。フィードバックによって、内部レポートや正式な KPI メトリクスでは発見できないプロセス、ツール、戦略のギャップを特定することができる。ソリューションのユーザーからの褒め言葉では、実用的な価値は得られない。価値があるのは批判的なフィードバックであり、それが具体的な観察や事実に基づくものであればなおさらである。しかし、そのような情報を得るには努力が必要である。社内外の参加者が、意見を歪めることなく、またその意見が自分の仕事に影響することを恐れることなく、コメントを共有できる（匿名でも意味があるかもしれない）プロセスを構築する必要がある。参加者が、歪めることなく、また自分自身への悪影響を恐れることなく、そうすることが重要である。

> どんな学習も最終的には自己学習である[165]。
>
> - ミルトン・フリードマン、アメリカの経済学者、統計学者

分析ツールの導入にあたっては、その有効性を定期的に検証する必要がある（ROI、KPI）。これにより、企業は失敗を繰り返さないだけでなく、環境の変化により迅速に適応することができる。フィードバックを収集・

分析する仕組みを持つことは、その時々のデジタル・イニシアティブから、継続的改善の持続可能なモデルへと移行する組織の成熟の兆候のひとつである（図 2.2-5）。

ステップ 6：パイロット・プロジェクトからスケールアップへ

> 重要な戦いは大きく、勝つための戦いは小さく。
>
> - ジョナサン・コゾール

デジタルトランスフォーメーションを「一度に、どこでも」開始することは、非常にリスクが高い。より効果的なアプローチは、パイロット・プロジェクトから始め、徐々に成功事例を拡大していくことである。

主なアクション

- ■ 試験的に適切なプロジェクトを選択する：
  - ☐ 測定可能な結果（KPI 、ROI ）を伴う特定のビジネス目的またはプロセスを定義する（図 7.1-5 ）。
  - ☐ Python と Pandas を使用した自動データ検証やワークロード計算 (QTO) などの ETL 自動化プロセスを選択(図 5.2-10)
  - ☐ 成功のための明確な指標を確立する（例えば、検査仕様書やデータ検証報告書の作成時間を 1 週間から 1 日に短縮する）。
- ■ 反復的なアプローチをとる：
  - ☐ 簡単なデータ変換処理から始め、マルチフォーマットデータを処理に必要なフォーマットにストリーミ ング変換する（図 4.1-2、図 4.1-5）。
  - ☐ 文書化されたコードブロックに基づき IDE で完全なパイプラインを形成することで、タスクの複雑さを徐々に増し、プロセスの自動化を拡大する（図 4.1-7、図 7.2-18）。
  - ☐ 成功した解決策を文書化し（できれば短いビデオで）記録し、同僚やプロのコミュニティで

共有する。

- 同僚（またはソーシャルメディアユーザーを含むプロフェッショナルコミュニティのメンバー）が効果的に使用できるように、そのようなソリューションを複製するためのテンプレートと付随する文書を作成する。

ステップ・バイ・ステップの「ロールアップ」によって、変更の質を高く維持し、並行実装の混乱に陥らない。小さな失敗から大きな失敗へ」戦略により、リスクを最小限に抑え、小さな失敗から学ぶことができます。

従業員が部分的にしか関与しないプロジェクト・アプローチから、常設チーム（例えば、専門家センター＝CoE）の形成に切り替えることは、最初のバージョンがリリースされた後も、製品の持続可能な開発を保証するのに役立つ。このようなチームは、既存のソリューションをサポートするだけでなく、その改善も継続する。

これにより、長い承認に依存することが少なくなり、チームメンバーは自分の責任範囲内で意思決定する権限を与えられる。その結果、マネジャーは 、マイクロマネジメントの必要性から解放され、チームは真の価値創造に集中できる。

> 新しいソリューションの開発はスプリントではなく、マラソンである。それに成功するのは、最初に長期的で一貫性のある仕事を目指した人である。

技術には絶え間ない発展が必要であることを認識することが重要である。技術的なソリューションの長期的な開発に投資することは、事業を成功させるための基本である。

ステップ 7：オープンデータ形式とソリューションの使用

モジュラープラットフォーム（ERP、PMIS、CAFM、CDE など）の章で述べたように、ベンダーのソリューションからの独立性を確保し、すべての利害関係者にとって情報の可用性を高めるオープンで普遍的なデータフォーマットに焦点を当てることが重要である。

主なアクション

- クローズド・フォーマットからオープン・フォーマットへ：
  - ☐ プロプライエタリなフォーマットの代わりにオープンフォーマットを使うか、クローズドフォーマットの自動アップロードやオープンフォーマットへの変換を設定する方法を見つける

（図 3.2-15）。

- ☐ Parquet, CSV, JSON, XLSX を扱うツールを実装する。これらは、最新のシステム間の交換標準である（図 8.1-2）。
- ☐ 3D ジオメトリを扱うことがプロセスで重要な役割を果たす場合は、USD、glTF、DAE、OBJ などのオープンフォーマットの使用を検討してください（図 3.1-14）。

■ ベクトルデータベース データベースを利用して、効率的に情報を分析・検索する：

- ☐ バウンディングボックス やその他の方法を用いて、3D -geometry を単純化する（図 8.2-1）。
- ☐ データのベクトル化（テキスト、オブジェクト、ドキュメントを数値表現に変換すること）を実装できる場所を考える（図 8.2-2）。

■ ビッグデータ分析ツールを適用する：

- ☐ 蓄積された履歴データ（例：PDF, XLSX, CAD ）を、分析に適したフォーマット（Apache Parquet, CSV, ORC）に整理して保存する（図 8.1-2）。
- ☐ 基本的な統計手法の適用を開始し、代表サンプルを扱う - あるいは、最低限、統計の基本原理に精通する (図 9.2-5)
- ☐ データおよびデータ間の関係を視覚化するツールを導入し、学習して、分析結果を視覚化する。優れた視覚化なしには、データそのものやそれに基づくプロセスを完全に理解することは不可能である（図 7.1-4）。

オープンデータ形式への移行と、情報を分析・保存・視覚化するツールの導入は、持続可能で独立したデジタル・ガバナンスの基礎を築く。これはベンダーへの依存を減らすだけでなく、すべての利害関係者のデータへの平等なアクセスを保証する。

ステップ 8.予測のための機械学習の導入開始

多くの企業には、膨大な量のデータ、つまり「情報の間欠泉」のようなものが蓄積されているが、それらはまだ使われていない。これらのデータは、何百、何千ものプロジェクトで収集されたものだが、一度しか使用されていないか、その後のプロセスでまったく使用されていないことが多い。クローズドなフォーマットやシステムに保存された文書やモデルは、時代遅れで役に立たないバラストとして認識されがちだ。しかし実際には

、これらは貴重な資源であり、間違いを分析し、ルーチン作業を自動化し、将来のプロジェクトで自動分類や特徴認識のための革新的なソリューションを開発するための基盤なのです。

重要な課題は、このデータを抽出し、実用的な洞察に変換する方法を学ぶことである。機械学習と予測」の章で述べたように、機械学習技術は、建設関連の様々なプロセスにおいて、見積もりや予測の精度を大幅に向上させる可能性を秘めている。蓄積されたデータを十分に活用することで、効率性の向上、リスクの低減、持続可能なデジタルプロセスの構築への道が開ける。

主なアクション

- 単純なアルゴリズムから始める：
  - LLM からのヒントを使用して、線形回帰を適用して、多数の要因の依存関係が存在しないか最小のデータセットにおけるリカレント・パフォーマンスを予測してみよう（図 9.3-4）。
  - 例えば、分類タスク、オブジェクトの類似性評価、または歴史的類似に基づく予測などです（図 9.3-5）。
- モデルを訓練するためのデータを収集し、構造化する：
  - 過去のプロジェクトデータを一箇所に集め、一つのフォーマットで管理する（図 9.1-10）
  - 自動化された ETL （図 9.2-8）を通して、トレーニングサンプルの質と代表性に取り組む。
  - タイタニック・データセットの例で行ったように、データをトレーニング・セットとテスト・セットに分けることを学ぶ（図 9.2-6、図 9.2-7）
- プロジェクトのスケジュール予測から、ロジスティクスの最適化、リソース管理、潜在的な問題の早期発見まで、幅広いタスクへの機械学習技術の適用拡大を検討する。

機械学習は、アーカイブされたデータを、予測、最適化、情報に基づいた意思決定のための貴重な資産 に変えるツールである。小さなデータセット（図 9.2-5）と単純なモデルから始め、徐々に複雑さを増していく。

ステップ 9. IoT と最新のデータ収集技術の統合

建設業界は急速にデジタル化している。すべての工事写真やチームへの投稿は、すでに現実をパラメータ化し、トークン化する大きなプロセスの一部となっている。かつて GPS が物流を変えたように、IoT、RFID、自動データ収集技術が建設業界を変えようとしている。IoT モノのインターネットとスマート・コントラクト」の章で述べたように、センサーと自動モニタリングによるデジタル建設現場は、業界の未来である。

主なアクション

- IoT -デバイス、RFID -タグを導入し、それらに関連するプロセスを詳述する：
  - 例えば、温度、振動、湿度、動きのモニタリングなど、プロジェクトのどの分野やプロジェクトフェーズで最大の投資利益率（ROI）が得られるかを評価します。
  - サプライチェーン全体を通して、材料、工具、機器を追跡するために RFID を使用することを検討する。
  - 自動化されたリアルタイム処理と分析のために、収集されたデータを Apache NiFi のような単一の情報システムに統合する方法を検討する (図 7.4-5)
- リアルタイムの監視システムを確立する：
  - Streamlit、Flask、Power BI などの可視化ツールを使用して、主要なプロセスやプロジェクトの指標を追跡するためのダッシュボードを開発する)
  - 計画や規範からの重大な逸脱を知らせる自動通知を設定する（図 7.4-2）
  - 収集されたデータと特定されたパターンに基づき、機器の予知保全の可能性を評価する（図 9.3-6）
- 異なるソースからのデータを組み合わせる：
  - CAD システム、IoT デバイス、ERP プラットフォームからの情報フローと主要パラメータの構造を反映する（図 4.3-1）。
  - データ分析と経営意思決定支援のための統一プラットフォームの大まかなアウトラインを作成することから始める。主要機能、データソース、ユーザー、想定されるアプリケーションシナリオを把握する（図 4.3-7）。

実際のプロセスをデジタルの世界に接続し始めれば、効率的で透明性の高い、リアルタイムでのデータ管理が可能になる。

ステップ 10.将来の業界の変化に備える

経済危機、技術の飛躍、規制の変更など、建設会社は常に外部環境からのプレッシャーにさらされている。雨、雪、干ばつ、灼熱の太陽に耐えなければならない森のように、企業は絶え間ない適応の条件下で生きている

。木々が深い根系によって霜や干ばつに強くなるように、自動化されたプロセスの強固な基盤、変化を予測し、柔軟に戦略を適応させる能力を持つ組織だけが、存続し、競争力を維持することができるのだ。

生存戦略：競争優位の構築」の章で述べたように、建設業界は激変の局面を迎えている。発注者と請負業者の相互作用は、透明性、予測可能性、デジタルツールが従来のアプローチに取って代わる、説得のモデル（ ）へと移行しつつある。この新しい現実において、勝つのは最大手ではなく、最も柔軟で技術的に成熟した企業である。

主なアクション

- オープンデータの文脈でビジネスの脆弱性を分析する：
  - ☐ ウベリゼーションの一環としてデータへのアクセスを民主化することが、自社の競争優位性とビジネスにどのような壊滅的な影響を与えるかを評価する（図 10.1-5）
  - ☐ 不透明でサイロ化されたプロセスから、オープンなソリューション、システムの相互運用性、データの透明性に基づくビジネスモデルに移行する戦略を考える（図 2.2-5）。
- 長期的なデジタル戦略を策定する：
  - ☐ イノベーション・リーダーを目指すのか、それともリソースを節約する「キャッチアップ」シナリオを好むのかを判断する。
  - ☐ 短期（プロセス自動化、データ構造化）、中期（LLM と ETL の導入）、長期（デジタルエコシステム、一元化されたリポジトリ）。
- サービスのポートフォリオを拡大することを考える：
  - ☐ 新たなサービス（エネルギー効率、ESG、データ・サービスに特化）の提供を検討する。新しいビジネスモデルについては、次の章で詳しく説明する。
  - ☐ 設計から運営に至るまで、施設のライフサイクル全体をサポートする信頼できる技術パートナーとしての地位を確立すること。貴社への信頼は、体系的なアプローチ、透明性の高いプロセス、持続可能な技術ソリューションを提供する能力に基づいているべきである。

変革的な環境では、単に変化に対応する者が勝つのではなく、積極的に行動する者が勝つのだ。柔軟性、開放性、デジタル成熟度が、明日を築く持続可能性の基礎となる。

変革のロードマップ：混沌からデータ駆動型へ 企業

以下のプランは、自社のデータ主導型デジタル変革戦略を策定するための最初のベンチマーク、つまり出発点として役立つ：

- 監査と基準：現状分析、データの統一
- データの構造化と分類：非構造化データと疎構造化データの変換を自動化する。
- グルーピング、計算、算出の自動化: 自動化のためのオープンソースツールやライブラリを使用する。
- エコシステムと COE：社内で統一されたデータ・エコシステムを形成するための社内チームを構築する。
- 文化と学習：HiPPO - ソリューションからデータ主導型ソリューションへの移行
- パイロット、フィードバック、スケーリング：反復的に行動する：新しい方法を限定的な規模でテストし、有効なフィードバックを収集し、解決策を徐々に拡大する。
- オープンフォーマット：ソフトウェアベンダーから独立するために、ユニバーサルでオープンなフォーマットを使用する。
- 機械学習: 予測と最適化のためのプロセスに ML アルゴリズムを組み込む
- IoT とデジタル建設現場：最新のデータ収集技術をプロセスに統合する
- 戦略的適応：将来の業界の変化に備える

> 最も重要なことは、「データだけでは会社は変わらない。企業文化、透明なプロセス、継続的改善へのコミットメントを重視する。

システム・アプローチは、サイロ化されたデジタル・イニシアティブから、意思決定が直感や仮定ではなく、データ、事実、数学的に計算された確率に基づく本格的なデータ駆動型管理モデルへの移行を可能にする。建設業界のデジタルトランスフォーメーションは、単にテクノロジーを採用することではなく、プロジェクト情報が異なるシステム間でシームレスかつ反復的に転送されるビジネス・エコシステムを形成することである。その際、機械学習アルゴリズムは、自動化された継続的な分析、予測、プロセスの最適化を提供する。このような環境では、推測や隠されたデータは無関係となり、実績のあるモデル、透明性のある計算、予測可能な結果だけが残る。

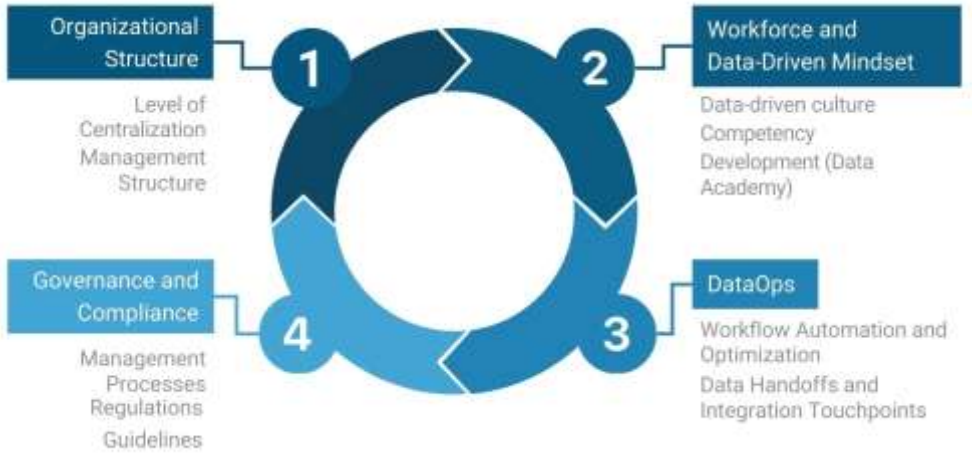


図 10.2-3 企業レベルのデータ管理を成功させるための主要要素。

本書の各パートは、建設プロジェクトにおけるデータ処理と分析の特定の段階に対応している（図 2.2-5）。もし、以前に取り上げたトピックのひとつに戻り、データ活用の流れを全体的な理解から見たい場合は、図 10.2-4 の各パートのタイトルを参照することができます。

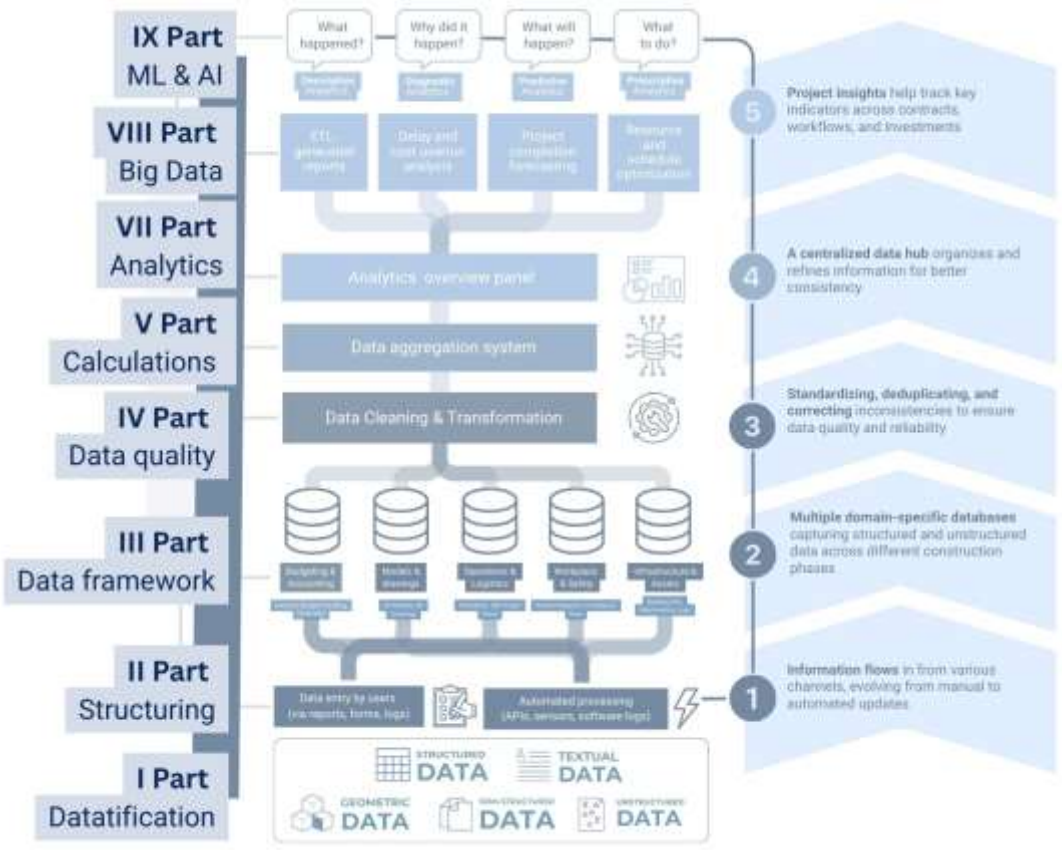


図 10.2-4 データ処理パイプライン（図 2.2-5）の文脈における本の一部：情報のデジタル化から
アナリティクス、人工知能まで。

組織の規模、テクノロジーの成熟度、予算にかかわらず、データ主導の 。正しい方向への小さな一歩でも、時間の経過とともに成果を生むだろう。

> データ主導の変革は、1 回限りのプロジェクトではなく、新しいツールの導入、プロセスの再設計、データ主導の意思決定文化の育成を含む、継続的かつ反復的な改善プロセスである。

インダストリー5.0 の構築：隠れることができなくなったときにお金を稼ぐ方法

長い間、建設会社はプロセスの不透明性で儲けてきた。主なビジネスモデルは投機であり、クローズドな ERP（）や PMIS（）、外部監査にアクセスできないシステムで、資材費、工事範囲、歩掛りを過大に見積もることだった。顧客とその信頼できる関係者が元のプロジェクト・データにアクセスすることが制限されていたため、計算の信頼性を検証することがほとんど不可能なスキームが生まれた。

しかし、このモデルは急速に妥当性を失いつつある。データアクセスの民主化、LLM の出現、オープンデータの登場、ETL 自動化ツールの登場により、業界は新たな作業標準へと移行しつつある。

> その結果、不透明性はもはや競争上の優位性ではなく、やがて手放すのが難しい重荷となる。透明性はオプションから、市場にとどまるための必須条件へと変わる。

銀行、投資家、現物顧客、プライベート・エクイティ 、政府系顧客など、顧客は新しいデジタルの現実の中で誰と仕事をするのだろうか？答えは明白で、結果だけでなく、そこに至るまでの各ステップの正当性を提供できる企業である。オープンデータの量が増えるにつれ、パートナーや顧客は、結果の透明性、正確性、予測可能性を保証する企業を選ぶようになるだろう。

このような背景から、投機ではなくデータ管理と信頼に基づく新しいビジネスモデルが生まれつつある：

- 平方メートルの代わりにプロセスを売る：重要な資産は、値引きされた具体的な契約ではなく、信頼と効率になる。主な価値は、信頼できる検証済みのデータに基づく結果の予測可能性である。現代の企業は、建設物そのものを売るのではなく、そのプロセスを売るのである：
    - 正確な納期と透明性のある作業スケジュール；
    - 妥当な見積もり 、計算によって裏付けられている；
    - プロジェクトの全段階における完全なデジタル・トレーサビリティと管理。
- サービスとしてのエンジニアリングとアナリティクス：「Data-as-a-Service 」モデル（既製のデータをサービスとしてインターネット経由でユーザーに提供する方法）。各プロジェクトはデジタルデータチェーンの一部となり、ビジネス価値はこのチェーンを管理する能力にある。企業は、自動化とアナリティクスに基づくソリューションを提供するインテリジェント・プラットフォームへと変貌しつつある ：
    - 自動化された透明性の高い見積もりと計画の作成；

- 機械学習アルゴリズムに基づくリスクとタイミングの評価；
- 環境指標（ESG、CO_2、エネルギー効率）の計算；
- 監査済みのオープンソースからレポートを作成する。

■ エンジニアリング・ノウハウの製品化：企業が開発したものは、社内で繰り返し使用され、別個の製品として流通させることができる。新しい環境では、企業はプロジェクトだけでなくデジタル資産も創造する：

- コンポーネントと見積もりテンプレートのライブラリ；
- 自動検証モジュール；
- データを扱うためのオープンソースのプラグインとスクリプト（コンサルティング販売）。

■ 新しいタイプの企業：データ主導型インテグレーター：特定のソフトウェア・ベンダーやモジュラー・システムに依存せず、単一のソフトウェア・インターフェースに「ロック」されない市場プレーヤー。特定のソフトウェア・ベンダーやモジュラー・システムに依存することなく、単一のソフトウェア・インターフェースに「ロック」されることもない。データとともに自由に活動し、データによって競争力を構築する。未来の建設会社は、単なる請負業者ではなく、顧客のために次のような機能を実行できる情報インテグレーターである：

- 異種ソースからのデータを結合し、分析を実行する；
- プロセスの透明性と信頼性を確保する；
- ビジネスプロセスの最適化に関するアドバイス
- オープンデータエコシステム、LLM 、ETL、パイプライン で動作するツールを開発する。

インダストリー5.0（図 2.1-12）は、「手作業による平均比率の時代」と、CEO と見積・経理部門との夜の会議の終わりを告げる。これまでは隠されていた計算、見積もり（ ）、数量などすべてがオープンになり、検証可能で、専門家でなくても理解できるようになる。真っ先に方向転換した者が勝者となる。それ以外の者は、建設部門の新しいデジタル経済から取り残されることになる。

結論

建設業界は、根本的な変化の時代を迎えている。粘土板に記された最初の記録から、プロジェクト・サーバーや建設現場から流れ出る大量のデジタル・データまで、建設における情報管理の歴史は、常にその時代の技術の成熟度を反映してきた。今日、自動化、オープンフォーマット、インテリジェントな分析システムの出現により、業界は緩やかな進化ではなく、急速なデジタル変革に直面している。

経済の他の分野と同様、建設業も、道具だけでなく仕事の原則も見直さなければならないだろう。かつては市場を左右し、発注者とプロジェクトの主な仲介者としての役割を果たしていた企業も、独自の地位を失いつつある。信頼と、データの収集と構造化から分析、予測、意思決定の自動化まで、データを扱う能力が前面に出てきている。

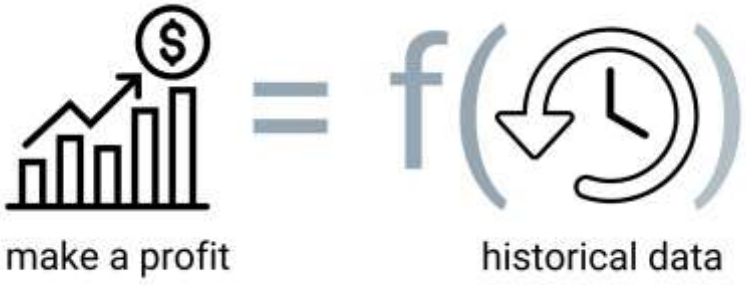


図 10.2-1 構造化された履歴データは、効率的で管理しやすいビジネスの燃料である。

本書では、監査や標準化からプロセスの自動化、可視化ツールの活用、インテリジェント・アルゴリズムの実装に至るまで、建設業界におけるデータ管理の主要原則について詳述する。限られたリソースでも、実用的なデータ・アーキテクチャを構築し、直感ではなく検証可能な事実に基づいた意思決定を開始する方法を考察した。データ活用はもはや IT 部門だけの仕事ではなく、企業の柔軟性、適応性、長期的な持続可能性を左右する経営文化の基盤となる。

機械学習技術（）、自動処理システム、デジタル・ツイン、オープン・フォーマットの応用により、以前は不可欠だった人的要素を排除することが、今日すでに可能になっている。建設は自律性と制御性へと向かっており、アイデアからプロジェクト実現までの動きは、自動操縦モードでのナビゲーションに例えることができる。主観的な判断に依存することなく、すべての段階で手作業による介入を必要とせず、完全なデジタル・トレ

ーサビリティと制御が可能である（図 10.2-2）。

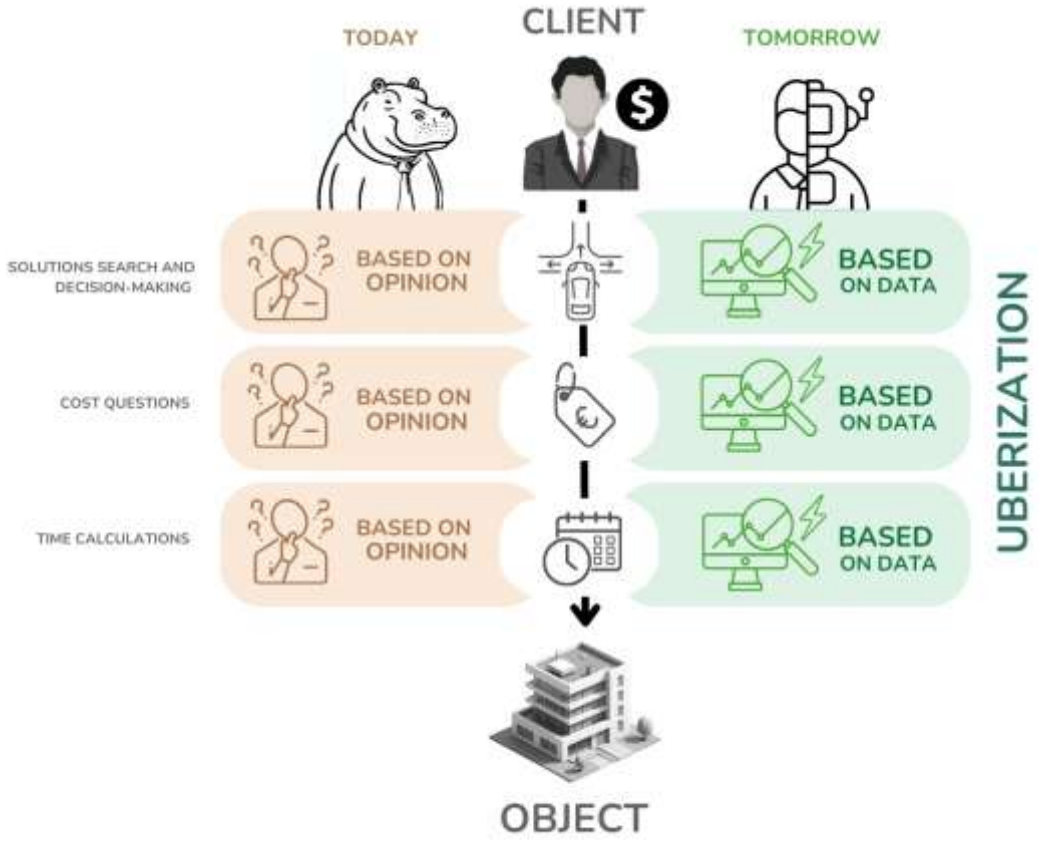


図 10.2-2 重要な専門家の意見に基づく意思決定（HiPPO ）からデータ分析へのシフトは、主に顧客によって推進される。

本書で紹介する方法、原則、ツールを学ぶことで、直感的な意思決定ではなく、データドリブンな意思決定を社内で行えるようになる。また、LLM でモジュールチェーンを実行し、すぐに使える ETL パイプラインを開発環境（IDE）にコピーし、データを自動的に処理して、必要な情報を必要な形で得られるようになる。ビッグデータと機械学習に関する本書の章を基に、過去のデータから新たな洞察を抽出し、機械学習アルゴリズムを適用してプロセスを予測・最適化するなど、より複雑なシナリオを実装できるようになります。

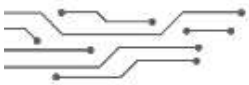
オープンなデータとプロセスは、プロジェクト・コストとスケジュールをより正確に見積もるための基礎となり、建設会社が不透明なデータに基づいて推測することを防ぐ。これは、業界が自らの役割を再考し、透明性と効率性が重要な成功要因となる新しい環境に適応するための課題であると同時にチャンスでもある。

> デジタル・トランスフォーメーションの時代における成功の鍵は、知識を積極的に取り入れ、実践することである。

このことにいち早く気づいた企業が、新たなデジタル競争で優位に立てるだろう。しかし、データだけでは何も変わらないことを認識することが重要である。多くの人が考え方を変える必要があり、それにはインセンティブが必要だ。あなたの会社は、データを共有する方法を考え直さなければならない

会社を変えるのは、このデータの扱い方を知り、それを解釈し、それを使って最適化し、それに基づいて新しいプロセス・アーキテクチャを作り上げる人たちだ。

この行を読んでいるということは、あなたは変化への準備ができていて、すでに一歩先を進んでいるということだ。この道を選んでくれてありがとう。デジタルトランスフォーメーションの時代へようこそ！



著者について



私の名前はアルチョム・ボイコ。私の建設現場での旅は 2007 年に始まった。故郷のオイルシェール鉱山で鉱夫として働きながら、サンクトペテルブルク鉱山大学で鉱山と地下建設を専攻した。本書の裏表紙には、私たちが何百もの立方体のオイルシェールを採掘し、発破をかけた切羽での爆発作業員が描かれている。私のキャリアは、鉱山や地下の建設作業員から、産業用クライマー、屋根葺き職人、リフト機器設置職人まで、さまざまな方向に進んできた。世界のさまざまな地域で、個人住宅の建設から大規模な工業施設の建設まで、さまざまな規模のプロジェクトに携わることができたのは光栄だった。

時が経つにつれ、私の仕事は物理的な建設から情報管理やデジタルプロセスへとシフトしてきました。2013 年以来、ドイツのいくつかの地域の中小・大手の建設会社で、設計者からデータ管理マネージャーまで、さまざまなポジションで仕事をしてきました。データ管理に関しては、様々な ERP システム 、CAD (BIM)、MEP 、FEM、CMS のデータを扱った経験があります。工業、住宅、インフラ、公共事業などの建設会社で、建設工事の計画、計算、実行段階における最適化、プロセスの自動化、分析、機械学習、データ処理に携わってきました。

2003 年以来、オープンソースソフトウェアとオープンデータに携わってきました。この間、ウェブサイトやオンラインショップから本格的なウェブアプリケーションまで、オープンソース・ソリューションとオープン CMS を使って、多くのウェブプロジェクトを実装してきました。これらのプラットフォームは、多くの点で最新のビルディング ERP に似ており、モジュラーアーキテクチャ、高い適応性とアクセシビリティを備えています。この経験は、オープンソース技術と共同開発の文化に焦点を当てた私の専門的アプローチの基礎を築きました。オープンソースと知識の自由な交換を尊重することは、私が建設業界で推進しようと努めていることです。建設業界におけるデータ・アクセシビリティを向上させるための私の活動は、データのオープン性と建設におけるオープンソースの使用について議論するいくつかのソーシャルメディア・コミュニティの創設や、様々なクローズド・システムやプラットフォームからのデータへのアクセスを提供するソリューションを開発するいくつかの新興企業の立ち上げへとつながっています。

私の専門家コミュニティへの貢献は、CAD の相互運用性 (BIM) 、ERP 、4D-5D、LLM 機械学習 、人工知能 を扱った会議へのスピーカーとしての参加や、欧州の建設業界の出版物に掲載された記事を通してである。私の特筆すべき業績のひとつは、"BIM の歴史"[111]の作成です。[111]は、建設業界におけるデータ管理のための重要なソフトウェアソリューションの包括的なマップです。数カ国語に翻訳された私の 7 部構成の記事シリーズ「BIM の開発とロビー活動ゲーム」は、デジタル標準の隠れた力学を明らかにする試みとして広く認知されている。

こうして私は、岩石の採掘から建設データの採掘と体系化を行うようになった。私は、プロフェッショナルな対話、新しいアイデア、共同プロジェクトに対して常にオープンです。どんなフィードバックでもありがたくお受けしますし、メッセージや私のソーシャルメディアのフォロワーの中でお会いできるのを楽しみにしています。本書を最後までお読みいただき、本当にありがとうございました！本書が、建設業界におけるデータの話題について理解を深める一助となれば幸いです。

さらなる関係

読者の意見は、出版物をさらに発展させ、優先的なトピックを選択する上で重要な役割を果たす。どの考え方が有用で、どの考え方に疑問が生じ、さらなる説明や出典の引用が必要であるかについてのコメントは、特に貴重なものである。本書にはさまざまな資料や分析が含まれており、そのなかには議論を呼ぶようなものや主観的なものも含まれている。もし、本書をお読みになる中で、不正確な記述、誤った出典の引用、論理的な矛盾、誤字脱字などを発見された場合は、コメント、ご感想、ご批判を boikoartem@gmail.com までお寄せいただければ幸いです。または、LinkedIn の linkedin.メッセージを通じて:

ソーシャルメディア上で『Data-Driven Construction』について言及していただけると非常にありがたい。読書体験を共有することは、オープンデータやツールについての情報を広めることにつながり、私の仕事をサポートしてくれる。

翻訳解説

本書は人工知能技術 を用いて翻訳された。これによって、翻訳作業は格段に速くなりました。しかし、どのような技術的な操作でもそうであるように、誤りや不正確さが発生する可能性があります。誤訳や不正確と思われる箇所にお気づきの際は、メールにてご連絡ください。あなたのコメントは、翻訳の品質を向上させるのに役立ちます。

コミュニティ

自由に質問し、問題や解決策を共有できる場所です：

DataDrivenConstruction.io: https://datadrivenconstruction.io

LinkedIn: https://www.linkedin.com/company/datadrivenconstruction/

Twitter: https://twitter.com/datadrivenconst

Telegram: https://t.me/datadrivenconstruction

YouTube: https://www.youtube.com/@datadrivenconstruction

その他のスキルとコンセプト

DataDrivenConstruction では、建設業界におけるデータ活用の主要原則に加え、データ活用のプロフェッショナルに不可欠な概念、プログラム、スキルを幅広く取り上げている。これらの中には、概要のみを紹介しているものもあるが、実務には欠かせないものである。

興味のある読者は、DataDrivenConstruction.io のウェブサイトを訪れ、主要スキルに関する追加資料へのリンクを参照することができる。これらの資料には、Python と Pandas を使った作業、ETL -プロセスの構築、建設 CAD プロジェクトにおけるデータ処理の例、ビッグデータシステム、建設データの可視化と分析への最新のアプローチなどが含まれる。

本書 "DataDrivenConstruction "およびすべてのケース・スタディの作成にあたり、多くのオープン・ソース・ツールおよびソフトウェア を使用した。著者は、以下のソリューションの開発者と共著者に感謝したい：

- Python と Pandas - データ操作と自動化のバックボーン
- Scipy、NumPy、Matplotlib および Scikit-Learn - データ分析および機械学習用ライブラリ
- SQL と Apache Parquet - 大容量の建設データの保存と処理のためのツール。
- オープンソース CAD (BIM) オープンフォーマットのオープンデータツール
- N8n, Apache Airflow, Apache NiFi - オーケストレーションおよびワークフロー自動化システム
- DeepSeek, LlaMa, Mistral - オープンソース LLM

本書の内容と構成を改善するために、批判、コメント、アイデアを提供してくれた、専門的なコミュニティやソーシャルネットワークにおけるオープンデータとツールというトピックに関する議論に参加したすべての人々に感謝する。

DataDrivenConstruction.io のウェブサイトでは書籍の更新や修正だけでなく、新しい章やチュートリアル、説明されているテクニックを適用するための実践例も公開されています。

印刷版で最大限の利便性を

データ駆動型建設』の無料デジタル版をご覧いただけます。より便利な作業や資料への迅速なアクセスのためには、印刷版のご利用をお勧めします：



■いつでも手元に：印刷された本書は信頼できる作業ツールとなり、どのような作業状況でも適切なビジュアライゼーションやダイアグラムを素早く見つけて使用することができます。

■高品質のイラスト：印刷版のすべての画像とグラフィックは最高の品質で表示されます。

■情報への素早いアクセス：簡単なナビゲーション、メモやブックマークを作成する機能、どこでも本で作業する機能。

印刷された完全版を購入することで、快適で効率的な情報作業のための便利なツールを手に入れることができます。日々の作業でビジュアル資料を素早く利用し、必要なスキームを素早く見つけ、メモを取ることができます。さらに、ご購入はオープンな知識の普及をサポートします。

印刷版のご注文はこちらから：datadrivenconstruction.io/books



またとない機会

戦略的ポジショニング

DataDrivenConstruction の無料版に広告を掲載することができます。有料版は、発行後 1 年間で、ラテンアメリカからアジア太平洋地域まで、世界 50 カ国以上の専門家の注目を集めました。個別の協力条件や掲載機会に関する詳細情報をご希望の方は、公式ポータル datadrivenconstruction.io のフィードバック・フォームにご記入いただくか、巻末に記載されている連絡先までご連絡ください。



本書の各章は datadrivenconstruction.io で入手可能。

Data-Driven Construction のウェブサイトでは、Data-Driven Construction の各章を読むことができます。このウェブサイトでは、必要な情報をすぐに見つけ、仕事に役立てることができるよう、書籍の各セクションが順次公開されています。また、同様のトピックに関する他の出版物や、応用例やソリューションも多数掲載されており、スキルアップやデータの建設への応用に役立ちます。



最新版

公式サイトからダウンロード

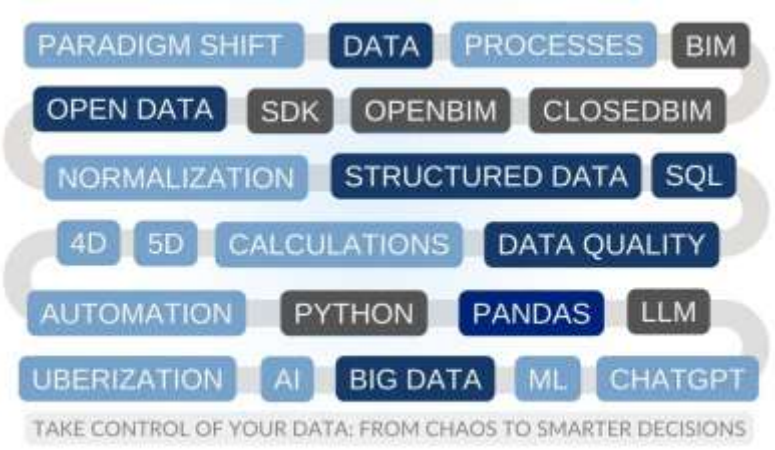
DataDrivenConstruction の最新版は、datadrivenconstruction.io からダウンロードできます。新しい書籍の章、実践的なヒント、新しいアプリケーションのレビューなどの最新情報をご希望の方は、ニュースレターをご購読ください：

- 本書の新セクションをいち早くご紹介します。
- 建設業界におけるアナリティクスとオートメーションに関する実践的なケーススタディとアドバイスを受ける。
- トレンド、出版物、応用例の最新情報を入手する。

申し込みは datadrivenconstruction.io へ！

データドリブン・コンストラクション
コンサルティング、ワークショップ、トレーニング

DataDrivenConstruction のトレーニングプログラムとコンサルティングは、世界中の大手建設会社数十社の効率化、コスト削減、ソリューションの品質向上を支援してきました。DataDrivenConstruction の顧客には、建設会社、コンサルティング会社、IT 企業など、10 億ユーロ規模の市場で最大手の企業が名を連ねています。



私たちを選ぶ理由

- 関連性：業界の主なトレンドや見識について語る
- 実践的：PoC を通じて、プロフェッショナルが日常業務を効率的に解決できるよう支援する。
- 個人的なアプローチ：お客様のビジネスに合わせて、トレーニングとアドバイスのメリットを最大限に引き出します。

DataDrivenConstruction チームの主な重点分野：

- データ品質管理：タスクのパラメータ化、要件の収集 、自動処理のためのデータの検証および準備を支援する。
- データマイニング - データの抽出と構造化：ETL プロセスを設定し、電子メール、PDF 、Excel 、画像 、その他のソースからデータを抽出します。
- BIM and CAD analytics: RVT ファイル, IFC, DWG and other CAD (BIM) formats から情報を収集し、構造化し、分析する。
- アナリティクスとデータ変換：異種情報を構造化データ、アナリティクス、洞察、意思決定に変える 。
- データ統合とプロセスの自動化：ドキュメントの自動作成から、社内システムや外部データベースとの統合まで。

DataDrivenConstruction.io にお問い合わせいただき、自動化を活用することで御社がどのように具体的なビジネス成果を達成できるかをご確認ください。

用語集

AI (Artificial Intelligence) - 人工知能。パターン認識、学習、意思決定など、通常は人間の知性を必要とするタスクをコンピュータ・システムが実行する能力。

Apache Airflow はオープンソースのワークフロー・オーケストレーション・プラットフォームで、DAG（有向非循環グラフ）を使ってワークフローや ETL をプログラムで作成、スケジュール、追跡することができる。

Apache NiFi は、システム間のデータフローを自動化するツールで、データのルーティングと変換に特化している。

Apache Parquet は、ビッグデータ解析システム用に最適化された、カラム型データストレージ用の効率的なファイルフォーマットです。大幅な圧縮と高速処理が可能です。

API (Application Programming Interface) - あるプログラムがソースコードにアクセスすることなく他のプログラムと相互作用することを可能にする公式化されたインターフェースで、標準化されたリクエストとレスポンスを通じてデータと機能を交換する。

属性 - 物体の特徴（面積、体積、コスト、材質など）を表す特性や性質。

データベースは、情報を保存、管理、アクセスするための組織化された構造であり、効率的なデータ検索と処理のために使用される。

BEP（BIM 実行計画） - プロジェクトで BIM を実施するための目的、方法、プロセスを定義するビルディング・インフォメーション・モデリングの実施計画。

ビッグデータ - かなりの量、種類、更新速度を持つ情報の配列で、処理と分析に特別な技術を必要とする。

BI（Business Intelligence） - ビジネスインテリジェンス。データを意思決定のための有意義な情報に変換するプロセス、テクノロジー、ツール。

BIM（ビルディング・インフォメーション・モデリング） - ビルディング・インフォメーション・モデリング。3D モデルだけでなく、性能、材料、時期、コストなどの情報も含め、建設プロジェクトの物理的・機能的特性をデジタルで表現し、作成・管理するプロセス。

BlackBox/WhiteBox-システムを理解するためのアプローチ：最初のケースでは、内部ロジックは隠され、入力と出力だけが見える。

バウンディングボックスは、3 次元空間におけるオブジェクトの境界を、X、Y、Z 座標の最小値と最大値によって記述する幾何学的な構成要素で、オブジェクトの周囲に「ボックス」を作成します。

BREP（Boundary Representation）とは、サーフェスの境界を通してオブジェクトを定義する幾何学的表現である。

CAD（Computer-Aided Design）は、建築、建設、エンジニアリング、その他の産業において、正確な図面や 3D モデルを作成、編集、分析するために使用されるコンピュータ支援設計システムです。

CAFM（Computer-Aided Facility Management）は、スペース計画、資産管理、メンテナンス、コスト監視を含む、不動産およびインフラ管理ソフトウェアである。

CDE（共通データ環境） - 施設のライフサイクルのあらゆる段階で、プロジェクト情報を管理、保存、共有、コラボレーションするための一元化されたデジタルスペース。

センター・オブ・エクセレンス（CoE）とは、特定の知識分野の開発、標準やベストプラクティスの開発、スタッフのトレーニング、イノベーションの導入支援などを担当する組織内の専門組織である。

CoClass は、近代的な第 3 世代の建築要素分類システムです。

概念データモデルとは、基本的なエンティティとその関係を、属性の詳細を伴わずに高レベルで表現したもので、データベース設計の初期段階で使用される。

CRM（Customer Relationship Management）とは、販売やサービスのプロセスを自動化するための顧客関係管理システムである。

DAG（Directed Acyclic Graph）は、データ・オーケストレーション・システム（Airflow、NiFi）でタスクの順序と依存関係を決定するために使用される有向非循環グラフである。

Dash は、インタラクティブな Web ベースのデータビジュアライゼーションを作成するための Python フレームワークです。

ダッシュボード - 主要なパフォーマンス指標と測定基準をリアルタイムで視覚的に表示するダッシュボード。

データ中心アプローチとは、アプリケーションやソフトウェアコードよりもデータを優先し、データを組織の中心的資産とする方法論である。

データガバナンス - アクセス、品質、セキュリティ管理など、組織内でのデータの適切かつ効果的な利用を確保するための一連の慣行、プロセス、方針。

データレイクは、大量の生データを使用するまで元の形式で保存するために設計されたストレージ施設である。

データレイクハウスは、データレイク（Data Lake）の柔軟性と拡張性と、データウェアハウス（DWH）の管理性とパフォーマンスを組み合わせたアーキテクチャアプローチである。

データ・ドリブン・コンストラクションとは、設計からオペレーションまで、施設のライフサイクルのあらゆる段階を、自動化された相互接続システムによってサポートする戦略的アプローチである。このアプローチは、継続的で事実に基づいた学習を提供し、不確実性を低減し、企業が持続可能な業界リーダーシップを達成することを可能にする。

データ・ドリブン・インテグレーター（Data-Driven Integrator） - 異種ソースからのデータを組み合わせて分析し、経営上の意思決定を行うことを専門とする企業。

データ主導型アプローチ - データを戦略的資産とみなし、主観的な意見ではなく、情報の客観的分析に基づいて意思決定を行う方法論。

データミニマリズム - データを最も価値があり意味のあるものに絞り込み、情報の処理と分析を簡素化するアプローチ。

データ沼 - 情報が適切に整理されずに収集され、無秩序に保存された場合に発生する、構造化されていないデータの散在。

DataOps は、DevOps の原則、データ、アナリティクスを組み合わせた方法論であり、データフローのコラボレーション、統合、自動化の改善に焦点を当てている。

情報のデジタル化とは、建設活動のあらゆる側面を、分析、解釈、自動化に適したデジタル形式に変換するプロセスである。

DataFrame - Pandas ライブラリの 2 次元表データ構造で、行は個々のレコードやオブジェクトを表し、列はその特徴や属性を表す。

記述分析 - 過去のデータを分析し、過去に何が起こったかを理解する。

診断分析 - データを分析し、何かが起こった理由を特定する。

ガントチャートとは、タスクをタイムライン上の横棒で表したプロジェクト計画ツールで、作業の順序と期間を視覚化することができます。

DWH（データウェアハウス）とは、複数のソースからの情報を集約し、構造化し、分析やレポーティングに利用できるようにする一元化されたデータウェアハウスシステムである。

ESG（環境、社会、ガバナンス） - 企業やプロジェクトの環境、社会、ガバナンスへの影響を評価するための一連の基準。

ELT（Extract, Load, Transform）とは、まずデータをソースから抽出し、リポジトリにロードし、分析目的のために変換するプロセスである。

ETL（Extract, Transform, Load）とは、様々なソースからデータを抽出し、目的のフォーマットに変換し、分析のためにターゲットストレージにロードするプロセスである。

ER 図（Entity-Relationship） - データモデリングで使用される、エンティティ、その属性、およびエンティティ間の関係を示す視覚的な図。

ERP（Enterprise Resource Planning）は、建設プロセスのさまざまな側面を管理し、最適化するために使用される包括的なモジュール式の企業資源計画システムである。

特徴 - 機械学習では、モデルの入力として使用される独立変数または属性。

物理データモデル - 特定の DBMS に最適化された、テーブル、カラム、データ型、キー、インデックスを含むデータベース構造の詳細な表現。

FPDF は PDF 文書を作成するための Python ライブラリです。

ジオメトリックコアは、CAD、BIM、その他のエンジニアリングアプリケーションで幾何学オブジェクトを作成、編集、分析するための基本的なアルゴリズムを提供するソフトウェアコンポーネントです。

HiPPO（Highest Paid Person's Opinion）：客観的なデータではなく、組織内で最も高給をもらっている人の意見に基づいて意思決定を行うアプローチ。

IDE (統合開発環境) - 統合開発環境。コードを書いたり、テストしたり、デバッグしたりするための包括的なツール（PyCharm、VS Code、Jupyter Notebook など）。

IDS（Information Delivery Specification）は、プロジェクトのさまざまな段階におけるデータ要件を定義した情報提供仕様書である。

IFC（Industry Foundation Classes）は、異なるソフトウェア・ソリューション間の相互運用性を保証する BIM データ交換フォーマットです。

インダストリー5.0 は、デジタル化、自動化、人工知能の能力と、人間の可能性、環境の持続可能性を組み合わせた産業開発コンセプトである。

データ統合とは、異なるソースからのデータを単一の首尾一貫したシステムに統合し、情報の統一されたビューを提供するプロセスである。

情報サイロとは、他のシステムと情報を共有しない孤立したデータ保管システムのことで、データの効率的な利用を妨げる障害となっている。

IoT（Internet of Things）とは、物理的なモノをインターネットに接続し、データを収集、処理、送信するという概念である。

k-NN (k-Nearest Neighbors)は、学習サンプルの最近傍のオブジェクトとの類似性に基づいてオブジェクトを分類する機械学習アルゴリズムです。

Kaggle は、データ分析と機械学習コンテストのためのプラットフォームです。

計算 - ある単位（例：石膏ボードの壁 $1m^2$、コンクリート $1m^3$）の建設工事または工程の費用の計算。

KPI（Key Performance Indicators）とは、重要業績評価指標のことで、企業や特定のプロジェクトの成功を評価するために使用される定量化可能な指標である。

ラベル - 機械学習において、モデルが予測すべき対象の変数または属性。

学習アルゴリズム - 学習データのセットを使用して、ターゲット関数に対応するモデルの最良の仮説を見つけるプロセス。

線形回帰 - 従属変数と 1 つ以上の独立変数の関係をモデル化する統計手法。

LLM (Large Language Model) - 大規模言語モデル。膨大な量のデータからテキストを理解・生成するために訓練された人工知能で、文脈を分析し、プログラミングコードを記述することができる。

LOD (Level of Detail/Development) - 幾何学的精度と情報量の程度を決定するモデルの詳細レベル。

論理データモデルは、ビジネス情報とルールを反映したエンティティ、属性、キー、関係の詳細な記述であり、概念モデルと物理モデルの中間段階である。

機械学習 - コンピュータシステムが明示的なプログラミングなしにデータから学習し、予測を行うことを可能にする人工知能技術の一種。

マスターフォーマットは、建設仕様書をセクションや分野に分類するための第一世代の分類システムである。

MEP（Mechanical, Electrical, Plumbing） - 機械、電気、配管コンポーネントを含む建築工学システム。

メッシュは、頂点、辺、面で構成される 3D オブジェクトのメッシュ表現である。

モデル - 機械学習では、予測または近似されるターゲット関数を近似する、異なる仮説のセット。

データモデリングとは、概念的、論理的、物理的レベルを含む情報システムに実装するために、データとその関係を構造的に表現するプロセスである。

n8n は、ワークフローを自動化し、ローコードアプローチでアプリケーションを統合するためのオープンソースツールです。

正規化 - 機械学習において、処理と分析を容易にするために、異なる数値データを共通の尺度にするプロセス。

リバースエンジニアリング - 対象物の構造、機能、動作を分析することによって、その装置、機能、製造技術を研究するプロセス。データの文脈では - オープンシステムで使用するために、独自のフォーマットから情報を抽出すること。

OCR（Optical Character Recognition）とは、テキスト画像（スキャンした文書や写真）を機械が読み取り可能なテキスト形式に変換する光学式文字認識技術である。

OmniClass は、建設情報管理のための第 2 世代の国際分類基準です。

オントロジー - 特定の知識分野を形式化した概念の相互関係の体系。

オープンソース - 自由な使用、研究、変更が可能なオープンソースソフトウェアを開発、配布するモデル。

オープン BIM とは、異なるソフトウェア・ソリューション間のデータ交換にオープンな標準とフォーマットを使用するという概念である。

オープンスタンダード - 特定のタスクを達成するための仕様で、異なるシステムが相互作用し、データを交換できるようにするもの。

Pandas は、データ処理と分析のためのオープンソースの Python ライブラリで、表形式の情報を効率的に扱うための DataFrame と Series データ構造を提供しています。

オープンデータパラダイムとは、情報を誰でも自由に利用、再利用、普及できるようにするデータ処理のアプローチである。

パラメトリック法とは、統計モデルを用いてプロジェクトのパラメータに基づく価値を見積もる建設プロジェクトの見積もり方法である。

PIMS（プロジェクト情報モデル）は、すべてのプロジェクト情報を整理、保存、共有するために設計されたデジタルシステムである。

パイプライン - 抽出、変換から分析、視覚化までの一連のデータ処理プロセス。

PMIS（プロジェクト情報管理システム）は、個々の建設プロジェクトのレベルで詳細なタスク管理を行うために設計されたプロジェクト管理システムである。

予測分析は、統計的手法と機械学習を用いて、過去のデータに基づいて将来の結果を予測する分析の一分野である。

処方的アナリティクス - 将来の結果を予測するだけでなく、望ましい結果を達成するための最適な行動を提案するアナリティクスのセクション。

プロプライエタリ・フォーマットとは、特定の企業が管理する閉じたデータフォーマットのことで、情報を共有する能力を制限し、特定のソフトウェアへの依存度を高める。

QTO（Quantity Take-Off）とは、プロジェクトの実現に必要な資材の数量を計算するために、設計書から要素の数量を抽出するプロセスのことである。

品質マネジメントシステム - プロセスと結果が確立された要求事項を満たすことを保証する品質マネジメントシステム。

RAG（Retrieval-Augmented Generation）は、言語モデルの生成能力と企業データベースからの関連情報の抽出を組み合わせた手法で、回答の精度と関連性を向上させる。

RDBMS（Relational Database Management System）は、相互に関連するテーブルの形で情報を整理するリレーショナル・データベース管理システムである。

RegEx（正規表現）は、文字列を検索・処理するための形式化された言語で、テキストデータが特定の基準に適合しているかどうかをチェックするためのテンプレートを指定することができます。

回帰は、変数間の関係を分析する統計的手法である。

CO_2計算とは、建設資材の生産・使用や工程に伴う二酸化炭素排出量を推定する方法である。

資源法は、建設工事を行うために必要なすべての資源（資材、労働力、設備）の詳細な分析に基づいて見積もりを行う方法である。

RFID（Radio Frequency Identification）とは、無線信号を使って自動的に物体を識別する技術で、材料や機械、人員の追跡に使われる。

ROI（投資利益率）とは、利益と投資資金の比率を表す指標で、投資の効果を評価するために用いられる。

SaaS（Software as a Service）とは、アプリケーションをプロバイダーがホストし、インターネット経由でユーザーが利用できるようにする、サービスとしてのソフトウェアのモデルである。

SCM（Supply Chain Management） - サプライチェーンマネジメントのことで、原材料の調達から完成品の納品までのすべてのプロセスの調整と最適化が含まれる。

データ・サイロとは、組織内の孤立した情報保管庫のことで、他のシステムと統合されていないため、データの共有が難しく、非効率的である。

SQL（Structured Query Language）は、リレーショナル・データベースを操作するために使用される構造化問い合わせ言語である。

SQLite は軽量で組み込み可能なクロスプラットフォーム DBMS で、別のサーバーを必要とせず、基本的な SQL 機能をサポートしており、モバイルアプリケーションや組み込みシステムで広く使用されています。

構造化データ - リレーショナル・データベースやテーブルなど、明確な構造を持つ特定のフォーマットで整理された情報。

緩やかに構造化されたデータ - JSON や XML のような、部分的に整理された柔軟な構造を持つ情報。

エンティティとは、現実世界に存在する具体的または抽象的な物体で、一意に識別され、記述され、データの形で表現できるもの。

教師あり学習 - 機械学習の一種で、分割されたデータでアルゴリズムを学習する。

分類法は、共通の特徴に基づいて要素を体系的に分類するために使用される階層的な分類システムである。

タイタニック・データセットは、機械学習モデルのトレーニングとテストによく使われるデータセットである。

トレーニング - 機械学習アルゴリズムがデータを分析してパターンを特定し、モデルを形成するプロセス。

転移学習とは、あるタスクのために訓練されたモデルを、別のタスクの出発点として使用する機械学習手法である。

変換 - データの形式、構造、または内容を後で使用するために変更するプロセス。

データ要件 - ビジネスプロセスをサポートするために必要な情報の構造、形式、完全性、品質を定義する公式化された基準。

建設業界のユーバー化とは、仲介業者を介さずに顧客と請負業者が直接やりとりできるデジタル・プラットフォームの影響下、建設業界における従来のビジネスモデルが変容するプロセスのことである。

ユニクラスは、英国で広く使われている第 2 世代および第 3 世代の建築要素分類システムである。

USD（Universal Scene Description）は、コンピュータグラフィックス用に開発されたデータフォーマットだが、そのシンプルな構造と幾何学的コアからの独立性により、エンジニアリングシステムにも応用されている。

データの検証とは、データの正確性、完全性、一貫性を確保するために、確立された基準や要件に照らし合わせて情報をチェックするプロセスである。

ベクトルデータベース - データを多次元ベクトルとして保存し、効率的なセマンティック検索やオブジェクト比較を行う特殊なデータベース。

ベクトル表現（埋め込み）とは、データを多次元の数値ベクトルに変換する手法で、機械アルゴリズムが効率的に情報を処理・分析できるようにする。

VectorOps は、多次元ベクトルデータの処理、保存、分析に焦点を当てた方法論であり、特にデジタルツインやセマンティック検索などの分野に関連する。

視覚化 - 情報のより良い知覚と分析のためのデータのグラフィカルな表現。

用語のアルファベットによる分類は、英語名で行った。

参考文献とオンライン資料のリスト

[1] ガートナー「IT Key Metrics Data 2017: Index of Published Documents and Metrics」2016 年 12 月 12 日。[オンライン]。利用可能: https://www.gartner.com/en/documents/3530919.[アクセス日：2025 年 3 月 1 日]。

[2] KPMG「おなじみの課題-新しいアプローチ。2023 Global Construction Survey」、2023 年 1 月 1 日。[オンライン]。利用可能: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf.[Accessed: 5 March 2025].

[3] F.R.バーナード「百聞は一見にしかず」10 Mapi 1927.[オンライン]。利用可能: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words.[参照日：2025 年 3 月 15 日]。

[4] M.Bastian, "Microsoft CEO Satya Nadella says self-claiming AGI is 'nonsensical benchmark hacking'," 21 Feb 2025.[オンライン]。Available: https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/.[アクセス日：2025 年 3 月 15 日]。

[5] W.E.フォーラム、「建設業の未来を形作るフォーラム - 変貌する風景：」、2016 年 1 月 1 日。[オンライン]。Available: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf.[アクセス日：2025 年 3 月 2 日]。

[6] C.D. Gillespie, "Clay: Earth's entanglement in the age of clay," 2024.[オンライン]。利用可能: https://ufl.pb.unizin.org/imos/chapter/clay/.

[7] 「紀元前 3 世紀のパピルス、言語はギリシア語」、2024 年。[オンライン]。Available: https://www.facebook.com/429710190886668/posts/595698270954525.

[8] 「モニタリング：利用可能なツールの活用」1980 年。[オンライン]。利用可能: https://pubmed.ncbi.nlm.nih.gov/10246720/.[講演日：2025 年 3 月 15 日）。

[9] PWC, "Data driven What students need to succeed in a rapidly changing business world," 2015 年 2 月 15 日.[オンライン]。利用可能: https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-

driven-paper-Feb2015.pdf.[アクセス日：2025 年 3 月 15 日]。

[10] スカンスカ USA「秋の建設市場動向」2023 年 11 月 2 日。[オンライン]。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.[アクセス日：2025 年 3 月 1 日].

[12] "引用：ソンダーガード、データ分析について"[オンライン]。利用可能: https://www.causeweb.org/cause/resources/library/r2493.[アクセス日：2025 年 3 月 15 日]。

[13] 「世界的な AI への関心がデータ管理市場をどのように押し上げるか」2024 年 5 月 28 日。[オンライン]。利用可能: https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/.[アクセス日：2025 年 3 月 15 日]。

[14] И.McCue, "ERP History", 2024.[オンライン]。利用可能: https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml.

[15] erpscout、"ERP の価格：ERP システムのコストはいくらですか？","[オンライン]。利用可能: https://erpscout.de/en/erp-costs/.[アクセス日：2025 年 3 月 15 日].

[16] softwarepath、「1,384 の ERP プロジェクトが語る ERP の選び方（2022 年 ERP レポート）」、2022 年 1 月 18 日。[オンライン]。利用可能: https://softwarepath.com/guides/erp-report.[アクセス日：2025 年 3 月 15 日]。

[17] デロイト「デジタル資本プロジェクトにおけるデータ駆動型マネジメント」2016 年 12 月 16 日。[オンライン]。利用可能: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf.[アドレスの日付：2025 年 3 月 1 日]。

[18] マッキンゼー、「2025 年のデータ駆動型企業」、2022 年 1 月 28 日。[オンライン]。利用可能: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-

2025.[参照日：2024 年 5 月 22 日]。

[19] ウィキペディア『ムーアの法則』。[オンライン]。利用可能: https://en.wikipedia.org/wiki/Moore%27s_law.[アクセス日：2025 年 3 月 15 日].

[20] アクセンチュア、「資本プロジェクトでさらなる価値を構築する」、2020 年 1 月 1 日。[Online].利用可能: https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf.[アクセス日：2024 年 3 月 3 日]。

[21] Б.Marr, "How much data do we create every day?誰もが読むべき驚くべき統計」、2018 年。[オンライン]。利用可能: 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] 「毎日どれだけのデータが作られているか」、2024 年。[オンライン]。利用可能: https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/.

[23] T.サリバン、「AI と世界の『データスフィア』：2025 年までに人類はどれだけの情報を手に入れるのか」、2024 年。[オンライン]。Available: https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By- 2025.html.

[24] Statista, "Total number of printed books produced in various regions of Western Europe in each half century between 1454 and 1800", [Internet].[オンライン]。利用可能: https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/.[参照日：2025 年 3 月 1 日]。

[25] 「価格設定の例」2024 年[オンライン]。https://cloud.google.com/storage/pricing-examples.

[26] M.Ashare, "Enterprises outsource data storage as complexity rises," 2024 年 5 月 10 日.[オンライン]。利用可能: https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/.[アクセス日：2025 年 3 月 15 日]。

[27] JETSOFTTPRO「SaaS は死んだ？マイクロソフト CEO の衝撃的な予測を解説」、2025 年 1 月 13 日。[オンライン]。利用可能: https://jetsoftpro.com/blog/saas-is-dead/.

[28] BG2 Pod、「サティア・ナデラ | BG2 w/ ビル・ガーリー & ブラッド・ガースナー」、2024 年 12 月 12 日。[オンライン]。利用可能: https://www.youtube.com/watch?v=9NtsnzRFJ_o.[アクセス日：2025 年 3 月 15 日]。

[29] GoodReads, "Tim Berners-Lee," [Online].利用可能: https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than.[アクセス日：2025 年 3 月 15 日]。

[30] KPMG「Cue Construction 4.0: Make-or-Break Time」2023 年 1 月 1 日。[オンライン]。利用可能: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html.[Accessed: 5 March 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.[オンライン]。Available: 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] マッキンゼー「REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY」2017 年 2 月 1 日。[オンライン]。利用可能: 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] 建設タスクフォースから副首相へ、「建設再考」、2014 年 10 月 1 日。[オンライン].利用可能: 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.[Online].Available: https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/.[Date of access: 15 March 2025].

[35] Wikiquote、「チャールズ・バベッジ」、[オンライン]。利用可能:

https://en.wikiquote.org/wiki/Charles_Babbage.[アクセス日：2025 年 3 月 15 日]。

[36] SAP, "New Research Finds Nearly Half of Executives Trust AI Over Themselves,", 12 Mar 2025.[オンライン]。Available: https://news.sap.com/2025/03/new-research-executive-trust-ai/.[アクセス日：2025 年 3 月 15 日]。

[37] カナダ建設協会とカナダ KPMG、2021 年、「デジタル世界における建設」、2021 年 5 月 1 日。[オンライン]。利用可能: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf.[アクセス日：2025 年 3 月 5 日]。

[38] ZCS、"第 5 次産業革命の解読"[オンライン]。利用可能：https://www.pwc.in/decoding-the-fifth-industrial-revolution.html.[アクセス日：2025 年 3 月 15 日]。

[39] M.K, Private Rights and Public Problems: The Global Economics of, Peterson Institute for Intern.Economics, 2012.

[40] F.N. a. Y. Z.ハーバード・ビジネス・スクール：マニュエル・ホフマン、「オープンソース・ソフトウェアの価値」、2024 年 1 月 24 日。[オンライン]。利用可能: https://www.hbs.edu/faculty/Pages/item.aspx?num=65230.[アクセス日：2025 年 3 月 15 日]。

[41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1 September 2008.[オンライン]。利用可能: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf.

[42] マッキンゼー、「建設業の生産性向上」、[オンライン]。Available: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity.[アドレスの日付：2025 年 3 月 15 日]。

[43] A.G．Syverson、「米国建設セクターにおける生産性の奇妙で恐ろしい道のり」、2023 年 1 月 19 日。[オンライン]。Available: https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/.[Date of access: 1 March 2025].

[44] McKinsey, "Delivering on construction productivity is no longer optional," 2024 年 8 月 9 日.[オンライ

ン]。利用可能: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional.[アクセス日：2025 年 3 月 5 日]。

[45] ING Group, "Lagging productivity in construction is driving up construction costs," 12 December 2022.[オンライン]。Available: https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/.[アドレスの日付：2025 年 3 月 15 日]。

[46] M.バーマン、「マイクロソフト CEO の衝撃的予測：『エージェントがすべてのソフトウェアに取って代わる』」、2024 年 12 月 19 日。[オンライン]。利用可能: https://www.youtube.com/watch?v=uGOLYz2pgr8.[アクセス日：2025 年 3 月 15 日]。

[47] Business Insider, "Anthropic's CEO says that 3 to 6 months, AI will be written 90% of the code were charge of software developers," 15 Mar 2025.[オンライン]。Available: https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3.[Date of access: 30 Mar 2025].

[48] Statista、「2024 年 6 月時点における世界のデータベース管理システム（DBMS）のカテゴリー別普及率比較」。[オンライン]。利用可能: https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/.[アドレスの日付：2025 年 3 月 15 日]。

[49] DB-Engines, "DB-Engines Ranking," [Online].利用可能: https://db-engines.com/en/ranking.[参照日：2025 年 3 月 15 日]。

[50] 「Stack Overflow Developer Survey 2023」、2024 年。[オンライン]。利用可能: https://survey.stackoverflow.co/2023/.

[51] 「SQL」2024 年[オンライン]。利用可能: https://en.wikipedia.org/wiki/SQL.

[52] 「構造化データと非構造化データ：その違いは？[オンライン]。利用可能: https://www.ibm.com/blog/structured-vs-unstructured-data/.

[53] DataDrivenConstruction、「PDF COMPARISON OF DATA FORMATS FOR CONSTRUCTION PROJECTS」、2024 年 4 月 23 日。[オンライン]。利用可能: https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf.

[54] 「ビルディング・インフォメーション・モデリング白書サイト", 2003.[オンライン]。利用可能: https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112.

[55] A.ボイコ「ロビー合戦と BIM 開発第 5 回：ブラックロックはあらゆる技術の支配者。企業がオープンソース・コードをいかにコントロールするか」、2024 年。[オンライン]。利用可能: 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.[オンライン]。利用可能: https://isicad.net/articles.php?article_num=14805.[アドレスの日付：02 2025].

[57] C.Eastman, Charles; And Cthers, "Eastman, Charles; And Cthers," September 1974.[オンライン]。利用可能: https://files.eric.ed.gov/fulltext/ED113833.pdf.[アクセス日：2025 年 3 月 15 日]。

[58] D.Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies," 2011 年 11 月 11 日.[オンライン]。利用可能: https://isicad.net/articles.php?article_num=14805.[参照日：2025 年 3 月 15 日]。

[59] D.ワイズバーグ「CAD の歴史」2022 年 12 月 12 日。[オンライン]。利用可能: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet.[参照日：2025 年 3 月 15 日]。

[60] ADSK, "ビルディング・インフォメーション・モデリング白書", 2002.[オンライン]。利用可能: https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand.[参照日：2025 年 3 月 15 日]。

[61] ADSK, "White Paper Building Information Modeling in Practice," [Online].利用可能: https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf.[アクセス日：2025 年 3 月 15 日]。

[62] A.ボイコ「ロビー合戦と BIM 開発。第 2 部：オープン BIM 対クローズド BIM。ヨーロッパ VS 世界の他の地域」、2024 年。[オンライン]。利用可能: 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.Boyko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs," 2024.[オンライン]。Available: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.

[64] ADSK、「ホワイトペーパー BIM」、2002 年。[オンライン]。Available: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand.[アドレスの日付：2025 年 3 月 15 日]。

[65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale", [Online].利用可能: https://web.archive.org/web/20010615093351/http:/www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf.[アクセス日：2025 年 3 月 15 日]。

[66] M.シャクレット「構造化データと非構造化データ：主な違い」、2024 年。[オンライン]。利用可能: https://www.datamation.com/big-data/structured-vs-unstructured-data/.

[67] K.Woolard, "Making sense of the growth of unstructured data," 2024.[オンライン]。利用可能: 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 the," 2004.[オンライン]。Available: https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf.[Date of access: 02 2025].

[69] CrowdFlower, "Data Science Report 2016," 2016.[オンライン]。利用可能: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf.[アクセス日：2025 年 3 月 15 日].

[70] Analyticsindiamag, "6 Most Time Consuming Tasks For Data Scientists," 15 May 2019.[オンライン]。利用可能: https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/.

[71] BizReport, "Report: Data scientists spend bulk of time clean up," 2015 年 7 月 6 日.[オンライン]。Available: https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-

data-scientists-spend-bulk-of-time-cleaning-up.html.[アクセス日：2025 年 3 月 5 日]。

[72] S.Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27 July 2015.[オンライン]。利用可能: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/.[アクセス日：2025 年 3 月 15 日].

[73] Б.Cyphers and K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability," 2024.[オンライン]。利用可能: https://www.eff.org/wp/interoperability-and-privacy.

[74] マッキンゼー・グローバル・インスティテュート、「オープンデータ：流動的な情報でイノベーションとパフォーマンスを解き放つ」、2013 年 10 月 1 日。[オンライン]。利用可能: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information.[アクセス日：2025 年 3 月 15 日]。

[75] A.Boyko, "The struggle for open data in the construction industry.AUTOLISP, intelliCAD, openDWG, ODA and openCASCADE," 15 05 2024.[オンライン]。利用可能: https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393.[アドレスの日付：16 02 2025].

[76] ウィキペディア、「マイクロソフトとオープンソース」、[オンライン]。利用可能: https://en.wikipedia.org/wiki/Microsoft_and_open_source.[アクセス日: 2025 年 3 月 15 日].

[77] TIME, "The Gap Between Open and Closed AI Models Might Be Shrinking.これが重要な理由だ」、2024 年 11 月 5 日。[オンライン]。Available: https://time.com/7171962/open-closed-ai-models-epoch/.[アクセス日：2025 年 3 月 15 日]。

[78] The Verge、「Google の新しいコードの 4 分の 1 以上が AI によって生成されている」、2024 年 10 月 29 日。[オンライン]。Available: https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024.[アクセス日：2025 年 3 月 15 日]。

[79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing," 2020 年 12 月 15 日.[オンライン]。利用可能: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics- 処理.[Accessed: 15

March 2025].

[80] PWC, "PwC Open Source Monitor 2019," 2019.[オンライン]。利用可能: https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf.[2025 年 3 月 15 日アクセス]。

[81] トラヴァース・スミス「オープン・シークレット：オープンソース・ソフトウェア」2024 年。[オンライン]。利用可能: https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/.[アクセス日：2025 年 3 月 15 日]。

[82] デロイト、「企業変革におけるデータ移転プロセス」、2021 年。[オンライン]。利用可能: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf.[参照日：2025 年 3 月 15 日]。

[83] gov.uk、「政府プロジェクトにおけるデータ分析と AI」、2024 年 3 月 20 日。[オンライン]。Available: https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery.[アクセス日：2025 年 3 月 5 日]。

[84] "引用元: すべては可能な限りシンプルになるべきだが、よりシンプルにはならない", 13 May 2011.[オンライン]。利用可能: https://quoteinvestigator.com/2011/05/13/einstein-simple/.[アクセス日：2025 年 3 月 15 日]。

[85] "トランスフォーマー（深層学習アーキテクチャ）"[オンライン]。利用可能: https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture).[アクセス日：2025 年 3 月 15 日].

[86] 「Python Packages Download Stats", 2024.[オンライン].利用可能: https://www.pepy.tech/projects/pandas.

[87] インタビュー・ビット、「Python ライブラリ・トップ 10」、2023 年。[オンライン]。利用可能: https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more.[アクセス日: 2025 年 3 月 30 日].

[88] 「エヌビディアと HP がワークステーションでデータサイエンスとジェネレーティブ AI を強化」、2025 年 3 月 7 日。[オンライン]。Available: https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations.[Date of access: 15 March 2025].

[89] P.Orac, "How to process a DataFrame with millions of rows in seconds", 2024.[オンライン]。利用可能: https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds.

[90] Ç.Uslu, "Kaggle とは？", 2024.[オンライン]。利用可能: https://www.datacamp.com/blog/what-is-kaggle.

[91] "NVIDIA CEO Jensen Huang Keynote at COMPUTEX 2024", 2 June 2024.[オンライン]。利用可能: https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w.[Accessed 15 March 2025].

[92] 2024 年「会員：創立者と法人会員」。[オンライン]。https://www.opendesign.com/member-showcase.

[93] A.Boyko, "The Age of Change: IFC is a thing of past or why ADSK and other CAD vendors are willing to give IFC for USD in 14 key facts", 24 November 2024.[オンライン]。利用可能: 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.[Accessed: 23 February 2025].

[94] A.Boyko「ポスト BIM の世界。データとプロセスへの移行と、建設業界にセマンティクス、フォーマット、相互運用性が必要かどうか」、2024 年 12 月 20 日。[オンライン]。Available: https://boikoartem.medium.com/the-post-bim-world-7e35b7271119.[アドレスの日付：2025 年 2 月 23 日]。

[95] N.I.o.Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE,", 2016.[オンライン]。Available: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf.[アクセス日：2025 年 2 月 23 日）。

[96] Harvard Business Review, "Bad Data Costs the U.S. $3 Trillion Per Year," 22 September 2016.[オンライン]。利用可能：https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

[97] Delpha、「データ品質の影響」、2025 年 1 月 1 日。[オンライン]。利用可能: https://delpha.io/impacts-of-data-quality/.

[98] W.B.D.ガイド、"Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects,".[オンライン]。利用可能: https://www.wbdg.org/resources/design-for-maintainability.[アクセス日：2025 年 3 月 15 日）。

[99] O.o.D.C.P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment," April 2014.[オンライン]。利用可能: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf.[アクセス日：2025 年 3 月 15 日]。

[100] ガートナー、「データ品質：正確な洞察のためのベストプラクティス」、2025 年 1 月 1 日。[オンライン]。利用可能: https://www.gartner.com/en/data-analytics/topics/data-quality.

[101] "釘が欲しくて"[オンライン]。利用可能: https://en.wikipedia.org/wiki/For_Want_of_a_Nail.[アクセス日：2025 年 3 月 15 日]。

[102] マッキンゼー・グローバル・インスティテュート「オープンデータ：流動的な情報でイノベーションとパフォーマンスを解き放つ」2013 年 10 月。[オンライン]。Available: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20i nformation/mgi_open_data_fullreport_oct2013.pdf.[Date of access: 15 March 2025].

[103] EY、「カーボンニュートラルへの道」、2023 年 3 月 10 日。[オンライン]。利用可能: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality.[アクセス日：2025 年 3 月 15 日]。

[104] PWC「ESG の認識」2024 年 7 月 1 日。[オンライン]。利用可能: https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf.[アクセス日：2025 年 3 月 15 日]。

[105] G.Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)," 2024.[オンライン]。利

用可能: https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf.

[106] "CO2_体積炭素の計算", 2024.[オンライン]。利用可能: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.

[107] McKinsey, "Imagining Construction's Digital Future", 24 June 2016.[オンライン]。利用可能: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future.[アクセス日：2025 年 2 月 25 日]。

[108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10 October 2024.[オンライン]。Available: https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/.[アドレスの日付：15 3 月 2025]。

[109] SAS, "Data lake and data warehouse - know the difference," [Online].利用可能: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html.[参照日：2025 年 3 月 15 日]。

[110] ADSK, "ビルディング・インフォメーション・モデリング", 2002.[オンライン]。利用可能: https://www.laiserin.com/features/bim/autodesk_bim.pdf.[アドレスの日付：15 3 月 2025]。

[111] A.Boyko, "BIM History Map", 2024.[オンライン]。利用可能: https://bigdataconstruction.com/history-of-bim/.

[112] A.S.Borkowski、「組織と規格による BIM の定義」、2023 年 12 月 27 日。[オンライン]。利用可能: https://encyclopedia.pub/entry/53149.[アドレスの日付：2025 年 3 月 5 日]。

[113] CAD Vendor, "OPEN BIM Programme," 2012.[オンライン]。利用可能: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/.[参照日：2025 年 3 月 30 日]。

[114] ウィキペディア、"業界基礎クラス"[オンライン]。利用可能: https://en.wikipedia.org/wiki/Industry_Foundation_Classes.[アクセス日：2025 年 3 月 15 日]。

[115] ウィキペディア『IGES』。[オンライン]。利用可能: https://en.wikipedia.org/wiki/IGES.[アクセス日：2025 年 3 月 30 日].

[116] A.Boyko, "History of CAD (BIM)," 15 December 2021.[オンライン]。利用可能: https://miro.com/app/board/o9J_laML2cs=/.[アドレスの日付：2025 年 2 月 24 日]。

[117] T.K.K.A.O.F.B.C.E.L. H. H. E. L. P. N. S. H. T. J. v. L. H. G. D. H. T. K. C.L.A.W.J. S. Francesca Noardo, "Reference study of IFC software support: GeoBIM benchmark 2019 - Part I, "8 Jan.[オンライン]。Available: https://arxiv.org/pdf/2007.10951.[アクセス日：2025 年 3 月 5 日]。

[118] И.Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM," 13 April 2021.[オンライン]。利用可能: https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be.[講演日：2025 年 3 月 5 日]。

[119] D.Ares, "Real Estate: Why It's Crucial for Efficiency & Growth", 17 December 2024.[オンライン]。利用可能: https://www.realpha.com/blog/rets-importance-in-real-estate-explained.[アクセス日：2025 年 3 月 5 日]。

[120] 「フレックス・トークンのコスト」2024 年[オンライン]。https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex.

[121] A.Boyko, "Forget BIM and democratise access to data (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024.[オンライン]。Available: 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, and F. Plaschke, "Solving the construction industry productivity puzzle," 2015.[Online].Available: https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle.

[123] "SCOPE - Projektdatenumgebung und Modelerung multifunktionaler Bauprodukte mit Fokus a Focus a die Gebäudehülle," 2018 年 1 月 1 日。[Online].Available: https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html.[アクセス日：2025 年 3 月 2 日].

[124] Apple.com, "Pixar, Adobe, Apple and NVIDIA form Alliance for OpenUSD for 3D content", 1 August 2023.[オンライン]。利用可能: https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/.[アクセス：2025 年 3 月 2 日]。

[125] AECmag、「ADSK のきめ細かなデータ戦略」、2024 年 7 月 25 日。[オンライン]。利用可能: https://aecmag.com/technology/autodesks-granular-data-strategy/.[アクセス日：2025 年 3 月 15 日]。

[126] A.Boyko, "The Age of Change: IFC is a thing of past or why ADSK and other CAD vendors are willing to give IFC for USD in 14 key facts," 24 11 2024.[オンライン]。利用可能: 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.[アドレスの日付：23 fevryall 2025]。

[127] A.Boyko, "ENG BIM Cluster 2024 | 建設における LLM と ChatGPT のデータと応用の戦い", 7 August 2024.[オンライン]。Available: ENG BIM Cluster 2024 | 建設における LLM と ChatGPT のデータと応用の戦い.[Accessed: 15 March 2025].

[128] 「ジェフリー・ゼルドマンプレゼンツ」2008 年 5 月 6 日。[オンライン]。利用可能: https://zeldman.com/2008/05/06/content-precedes-design/.[アクセス日：2025 年 3 月 15 日]。

[129] A.Boyko、「ChatGPT による DWG 解析 | DataDrivenConstruction」、2024 年 3 月 5 日。[オンライン]。利用可能: https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction.[参照日：2025 年 3 月 15 日]。

[130] McKinsey, "The McKinsey guide to outcompeting in the age of digital and AI," 2023.[オンライン]。Available: https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired.[アドレスの日付：2025 年 3 月 30 日].

[131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs," 31 March 2016.[オンライン]。利用可能: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/.[アクセス日：2025 年 3 月 15 日].

[132] J.Bertin, "Graphics and Graphic Information Processing," 2011 年 9 月 8 日.[オンライン]。利用可能:

https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y.[アクセス日：2025 年 3 月 15 日]。

[133] CauseWeb, "Wells/Wilks on Statistical Thinking," [Online].利用可能: https://www.causeweb.org/cause/resources/library/r1266.[アクセス日：2025 年 3 月 15 日]。

[134] Ministrymagazine「科学はいかにして創造を発見したか」1986 年 1 月号。[オンライン]。利用可能: https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation.[参照日：2025 年 3 月 15 日）。

[135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23 May 2017.[オンライン]。利用可能: https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation.[2024 年 5 月 15 日アクセス]。

[136] "イノベーションを推進するデータアーキテクチャの構築方法-今日と明日", 3 June 2020.[オンライン]。Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow.[講演日：2025 年 3 月 15 日]。

[137] オックスフォード、"Woodrow Wilson 1856-1924"[Online].Available: https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630.[講演日：2025 年 3 月 15 日］。

[138] 「コンバーター」2024 年[オンライン]。https://datadrivenconstruction.io/index.php/convertors/.

[139] PWC, "Sizing the prize AI's real value for your business and how can you capitalise?", 1 January 2017.[オンライン]。利用可能: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf.[アクセス日：2025 年 2 月 18 日]。

[140] 「建設中のパイプライン」、2024 年。[オンライン]。https://datadrivenconstruction.io/index.php/pipeline-in-construction/.

[141] ウィキペディア、「Apache NiFi」、2025 年 1 月 1 日。[オンライン]。利用可能: https://de.wikipedia.org/wiki/Apache_NiFi.[アクセス日: 2025 年 3 月 5 日].

[142] n8n、「Gmail の AI 自動応答：受信メールに下書き返信を作成」、2024 年 5 月 1 日。[オンライン]。利用可能: https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/.[アクセス日：2025 年 3 月 15 日]。

[143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail," 2025 年 3 月 1 日.[オンライン]。利用可能: https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/.[アクセス日: 15 Mar 2025].

[144] B.T.O'Neill、「アナリティクス、AI、ビッグデータ・プロジェクトの失敗率＝85% - ヤバい！」、2025 年 1 月 1 日。[オンライン]。利用可能: https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/.

[145] J.Neyman, On the Two Different Aspect of Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.

[146] T.J. S. Jesse Perla, "A Problem that Stumped Milton Friedman," Quantitative Economics with Julia, 1 Jan 2025.[オンライン]。Available: https://julia.quantecon.org/dynamic_programming/wald_friedman.html.[Date of access: 1 May 2024].

[147] T.ランズオール・ウェルフェアー、『国民の気分を予測する』、シグニフィカンス、2012 年。

[148] A.Boyko「サンフランシスコ。建設部門 1980-2019", 2024.[オンライン]。Available: https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019.

[149] A.Boyko, "Kaggle: RVT IFC Files 5000 Projects," 2024.[オンライン]。利用可能: https://www.kaggle.com/datasets/artemboiko/rvtifc-projects.

[150] CFMA、"コネクテッド・コンストラクションで未来に備える"[オンライン]。利用可能: https://cfma.org/articles/preparing-for-the-future-with-connected-construction.[2025 年 3 月 15 日アクセス]。

[151] Cisco, "Cisco Survey Reveals Close to Three-Fourths Are Failing of IoT Projects," 2017 年 5 月 22 日。[Online].利用可能: 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] 「植物化石保存に必要な条件」、2024 年。[オンライン]。利用可能: https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html.

[153] "ブラックロックのフィンク、債券、M&A、米景気後退、選挙について：フルインタビュー", 2023.[オンライン]。利用可能: https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video.

[154] cio、「12 の有名な AI 災害」、2024 年 10 月 02 日。[オンライン]。利用可能: https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html.[アクセス日：2025 年 3 月 15 日]。

[155] G.カスパロフ『ディープ・シンキング』PublicAffairs、2017 年。

[156] ウィキペディア、「Kaggle」、2025 年 1 月 1 日。[オンライン]。利用可能: https://en.wikipedia.org/wiki/Kaggle.[アクセス日: 2025 年 3 月 15 日].

[157] Kaggle, "Titanic - Machine Learning from Disaster," 2025 年 1 月 1 日.[オンライン]。利用可能: https://www.kaggle.com/competitions/titanic/overview.[2025 年 3 月 10 日アクセス]。

[158] Ш.ジョフリ『ChatGPT を創る：データから対話へ』2024 年。[オンライン]。利用可能: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/.

[159] П.Domingos, "A few useful things to know about machine learning," 2024.[オンライン]。利用可能: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

[160] J.Saramago, "Quotable Quote," [Online].利用可能: https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered.[アクセス日：2025 年 3 月 17 日]。

[161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models", 19 December 2024.[オンライン]。利用可能: https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/.[2025 年 3 月 25 日アクセス]。

[162] "NVIDIA、Cosmos World Foundation モデルと物理 AI データツールのメジャーリリースを発表", 18 Mar 2025.[Online].利用可能: https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools.[Date of access: 25 March 2025].

[163] NVIDIA, "NVIDIA Isaac Sim,", [On Internet].利用可能: https://developer.nvidia.com/isaac/sim.[2025 年 3 月 25 日アクセス]。

[164] M.クォータリー「デジタル戦略はなぜ失敗するのか」2018 年 1 月 25 日。[オンライン]。利用可能: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail.[アクセス日：2025 年 3 月 15 日].

[165] M.J.ペリー、「私の好きなミルトン・フリードマンの名言」、2006 年 11 月 17 日。[オンライン]。利用可能: https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/.[アクセス日：2025 年 3 月 1 日]。

[166] J.A.ウィーラー、「情報、物理、量子：リンクの探求」、1990 年。

[169] A.ボイコ「ロビー合戦と BIM 開発第 5 回：ブラックロックはあらゆる技術の支配者。企業がオープンソース・コードをいかにコントロールするか」、2024 年。[オンライン]。利用可能: https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d.

[170] T.Krijnen and J.Beetz, "A SPARQL query engine for binary-formatted IFC building models," Advanced Engineering Informatics, 2024.

[171] 「2021 年の英国の建設部門企業数（事業規模別）」2024 年。[オンライン].利用可能: https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/.

[172] 「5,000 の IFC&RVT プロジェクト」、2024 年。[オンライン]。https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io.

[173] M.ポポヴァ, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe," 2008.[オンライン]。利用可能: https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/.[参照日：2025 年 2 月]。

[174] 建設におけるデータをめぐるロビー合戦｜テクノ封建主義と BIM の隠された歴史。[映画]。ドイツ：アルチョム・ボイコ、2023 年。

[175] A.Boyko、「CHATGPT WITH REVIT AND IFC｜プロジェクトからのドキュメントとデータの自動検索」、2023 年 11 月 16 日。[オンライン]。利用可能: https://www.youtube.com/watch?v=ASXolti_YPs&t.[アクセス日：2025 年 3 月 2 日]。

[176] M.&.Company, "Three new mandates for capturing a digital transformation's full value," 2022 年 1 月 22 日。[オンライン]。Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value.[Date of access: 15 February 2025].

[177] KPMG、「デジタル世界における建設」、2021 年 5 月 1 日。[オンライン]。Available: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-ja.pdf.[アクセス日：2024 年 4 月 5 日]。

[178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time," 2023 年 3 月 17 日.[オンライン]。利用可能: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html.[アクセス日：2025 年 2 月 15 日]。

[179] O.Business, "Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models'," 10 Jan 2025.[オンライン]。Available: https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models.[アクセス日：2025 年 3 月 15 日]。

[180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," 23 Mar 2016.[Online].利用可能: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/.[Date of access: 15 March 2025].

[181] 英国外務・英連邦・開発省「デジタル開発戦略 2024〜2030」、2024 年 3 月 18 日。[オンライン]。Available: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030.[アクセス日：2025 年 3 月 15 日]。

[182] 「建築設計業界におけるビジョンと戦略」、2003 年 11 月 7 日。[オンライン]。https://web.archive.org/web/20030711125527/http:/usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112.[講演日：2025 年 3 月 5 日].

[183] M.Bocharov、「情報モデリング」、2025 年 3 月。[オンライン]。Available: https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5.[講演日：2025 年 3 月 15 日]。

[184] 「統合デザイン・スルー・マニュファクチャリング：その利点と合理性」、2000 年。[オンライン]。利用可能: https://web.archive.org/web/20010615093351/http:/www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf.[参照日: 2025 年 3 月 25 日].

[185] CAD Vendor, "Open BIM Programme is a marketing campaign," 2012 年 3 月 12 日.[Online].Available: https://web.archive.org/web/20120827193840/http:/www.graphisoft.com/openbim/.[アドレスの日付：2025 年 3 月 30 日]。

サブジェクト・インデックス

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
アパッチ・エアフロー、311、361、362、363、364、366、367、369、371、399、494
アパッチ NiFi、116、311、361、367、368、371、399、482、494
アパッチ ORC、62、330、378
アパッチ・パーケット、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
ビッグデータ, 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
ボケ, 320, 337
bom, 76, 77, 79, 257, 263
バウンディングボックス、234、373、392、393、394、414、415、480
ブレップ、142、234、263、264、276、283、284

C

カド、6、14、18、24、56、57、58、63、70、71、72、73、74、75、76、77、78、79、80、84、85、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
CFM、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
クロード、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
副操縦士、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
ダグ, 362, 363, 365, 366
ダッシュ、320、336、337。
データガバナンス、373, 395, 396, 398, 400, 401
データレイク、214、373、376、384、385、386、387、388、389、390、400
データレイクハウス, 67, 373, 386, 387, 388
データミニマリズム, 373, 395, 396, 400, 401
データ沼、373、395、397、401
データウェアハウス、382, 383, 400
データ・アズ・ア・サービス, 487
データ駆動型, 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
エクセル、57、61、62、65、66、85、88、111、120、123、125、132、154、167、187、210、223、224、226、227、228、278、291、329、333、342、351、365、376、378、459、475、497
抜粋、81、128、134、193、311、323、324、326、328、330、345、361、362、363、365、383、384、450、475

F

フェザー、62、123、330、378
fpdf, 339, 340, 341, 342, 343

G

GDPR, 109
GIS, 58
GLTF, 143, 278
グーグル・シート, 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

ID、289、290、291。
IFC、8、73、138、142、186、227、261、262、263、264、265、266、267、268、272、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、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
ジュピター・ノートブック、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

リード、235、236、238。
LlaMa、103、104、120、124、129、133、166、219、300、329、 343、349、381、421、429、443、494
LM、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、187、219、220、221、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
負荷、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
ローコード, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
マカド、77、78、257、284
MEP, 14, 175, 492
メッシュ、234、283、284、285、296、462
Microsoft SQL, 65
ミストラル、103、104、107、110、129、133、166、219、221、300、329、343、349、381、421、429、443、494
MRP, 11, 12
MS Project, 70
MySQL, 63, 64, 65, 291, 329, 330

N

N8N, 116, 311, 361, 368, 369, 370, 371, 372
NLP、69、394
ノーコード、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
オムニクラス、154、155、156。

オープン BIM, 142, 216, 219, 256, 261, 278, 291
オープンソース、45、55、97、98、107、108、273、275、494
フクロウ、267、268、269。

P

パンダ、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
寄木細工、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、344、345、346、347、354、355、356、357、365、376、475、480、497
PDM, 16
PHP, 63, 339
パイプライン、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
プライベート・エクイティ, 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
RDBM、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
rpm、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
スコープ、273、277
sdk, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
シーボーン、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
ステップ、261、262、263、266、272、276、277、292
ストリームリット、336、337
svf、142、276、285

T

変形、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

ユニクラス、154、155、156。
米ドル、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、291、299、329、378、414、474

データが建設業界に与える変革を探求する

中身
- AECにおける100を超える主要なデータ関連トピック
- 300を超えるユニークな視覚化と図
- 50件以上の実際のビジネスケース
- 実用的なLLMおよびAIアプリケーション
- コードサンプルと即利用可能なワークフロー

トピック
- データ駆動型建設業
- AECにおけるデジタル変革
- データ解析と自動化
- データの質とガバナンス
- CAD、BIM、データの相互運用性
- 建設業界における法学修士、人工知能、機械学習
- プロジェクトのコストおよび時間の見積もり

観客
- 建設プロジェクト管理者
- 建築家および土木技術者
- BIMコーディネーターおよびデータ管理者
- デジタルトランスフォーメーションのリーダー
- AEC分野におけるITおよびAIの専門家
- 都市計画専門家および持続可能性アドバイザー
- 建築学および工学の学生
- 教育者と研究者

- **ウェブサイト**
- • www.datadrivenconstruction.io

初版に関するコメント:

「ボイコはIT界のジェームズ・カーヴィルである。彼のよく引用される『経済が問題なんだ、バカめ』という言葉を、この名著に置き換えるなら、たった一言で済む。『データが問題なんだ、バカめ』（ソフトウェアではない）。そして、データの世界で進むべき道を見つけるには、古代ローマ人がギリシャにまで遡って言い伝えた「ナビゲートは必要だ」という言葉が、今日でも通用する。著者は、確かな手腕と揺るぎない羅針盤、包括的な歴史的アプローチ、そして独創的なグラフィックによって、読者をデータの海の深淵へと導いてくれる…」

- — ブルクハルト・タレヴィタリ博士

「アルチョム・ボイコ氏の著作は、建設業界におけるデジタル化の民主化に向けた画期的な出来事であり、中小企業にとって真のゲームチェンジャーとなるでしょう。本書はまさに行動への呼びかけです。建設業界におけるデジタル変革を理解したいだけでなく、実践的かつ効率的、そして将来を見据えた方法で積極的に変革を推進したいと考えるすべての人にとって、貴重なガイドとなるでしょう。今こそ、この知識を共有し、建設業界の生産性を持続的に向上させるために、共に協力すべき時です。」

- — マイケル・マックス・ビュラー博士

この実践的なガイドは、専門家と初心者の両方が急速に進化するデータ駆動型建設の領域を効果的にナビゲートできるよう支援します。データ管理の基礎から高度なデジタルワークフロー、AIツール、さらには実社会での応用に至るまで、本書はよりスマートで迅速かつ効率的な建設プロセスへの道筋を示します。

ISBN: 978-3-98262-558-4
9 783982 625584