2. UDGAVE OPDATERET
OG UDVIDET

DATA DRIVEN CONSTRUCTION

NAVIGERING I DATAALDEREN I BYGGEBRANCHEN

MED AI- OG LLM-
BRUGSSCENARIER



Artem Boiko


100+
KEY DATA-RELATED TOPICS


DLC GUIDE


DLC GUIDE


50+
PRACTICAL PROBLEMS SOLVED


300+
UNIQUE ILLUSTRATIONS AND GRAPHS

DATA-DRIVEN CONTRUCTION

Navigering i data-alderen i byggebranchen

Anden udgave, revideret og suppleret

ARTEM BOIKO

"Boiko er it's James Carville - i sidstnævntes meget citerede "It's the economy, stupid" er der kun ét ord, der skal udskiftes med denne berømte bog. "Det er dataene, dumme." (Og for at finde vej i datauniverset gælder et ordsprog fra de gamle romere, der går tilbage til græsk, stadig i dag: "Navigare necesse est". Forfatteren navigerer sine læsere gennem alle datahavets dybder og lavvandede områder med sikker hånd og et urokkeligt kompas, for ikke at nævne en omfattende historisk tilgang og sidst, men ikke mindst, meget original grafik og en god sans for humor, som ikke kun er tydelig ved andet øjekast. Den internationale respons på Boikos bog spænder fra euforisk bifald til temmelig galgenhumoristisk skepsis, hvilket har gjort den anden tyske udgave af bogen godt. Boiko er en original og udogmatisk datatænker. Han præsenterer læseren for spændende indsigter og altid modige, endda provokerende teser, som inspirerer til videre overvejelser. Fremragende medicin mod den tyske sygdom latent konsensualisme. I øvrigt har ovenstående latinske ordsprog et supplement: "vivere non est necesse". Det gælder ikke for Boikos tilgang til dataverdenen - data lever, og deres liv er nødvendigt, for ikke at sige afgørende."

- Dr. Burkhard Talebitari, freelance-redaktør - bl.a. for tidsskriftet: BIM, som Ernst & Sohn har udgivet hvert år siden 2013.

"Artem Boikos bog er en milepæl for demokratiseringen af digitaliseringen i byggebranchen - og en ægte game changer for små og mellemstore virksomheder (SMV'er). Særligt banebrydende: Ved at bruge moderne open source low-code- og no-code-værktøjer kan virksomheder allerede nu effektivt integrere data i deres forretningsprocesser og analysere dem rentabelt - uden nogen dybtgående viden om programmering. Det gør den dyre brug af besværlige kommercielle softwarepakker overflødig. Denne bog er en opfordring til handling! Den er en værdifuld guide for alle, der ikke kun ønsker at forstå den digitale transformation i byggebranchen, men også ønsker at forme den aktivt - pragmatisk, effektivt og på en fremadrettet måde. Nu er det tid til at arbejde sammen om at dele denne viden og øge produktiviteten i byggebranchen på en bæredygtig måde."

-Dr. Michael Max Buehler, professor i byggeledelse ved HTWG Konstanz, medejer af GemeinWerk Ventures og uafhængig direktør i DevvStream**.**

"DataDrivenConstruction" er et af de første skridt ud over grænserne for bygherrernes sædvanlige verden med deres komplekse design- og styringssystemer, hvor det ser ud til, at kompleksiteten og mætningen af data ikke engang giver en chance for radikal forenkling og øget gennemsigtighed i arbejdet med konstruktionsdata. I sin bog viser Artem i et enkelt sprog, hvilke muligheder moderne teknologier til at arbejde med data åbner for os, og giver bogstaveligt talt konkrete trin, som du straks kan anvende i dit arbejde. Jeg opfordrer alle, der ønsker at forstå, hvor automatiseringssystemer vil bevæge sig hen i byggebranchen, til at læse denne bog grundigt for at indse, at datarevolutionen i byggeriet allerede banker på vores dør. Det er kun interessant for nørder nu, men om få år vil sådanne tilgange og software være allestedsnærværende, ligesom BIM!"

- Ihor Rogachew, leder af IMT-kompetencecentret, BIM & Digital Transformation hos RGD og grundlægger af InfraBIM.Pro.

"Jeg kan varmt anbefale DataDrivenConstruction-bogen, der, som titlen siger, omhandler en datadrevet informationsstyringstilgang for AECO. Jeg bruger den i øjeblikket til at sætte gang i en række diskussioner med forskellige grupper. Jeg synes, det er en meget tilgængelig reference. Ud over en grundig oversigt over den historiske kontekst for værktøjer i AECO, data og introduktion af flere nøgleteknologier indeholder bogen en række meget nyttige diagrammer, der skitserer omfanget af datakilder og slutbrugerartefakter med eksempler på workflows. Det slår mig, at det er den type diagrammer, vi har brug for flere af, når vi udvikler og overvåger informationsstrategier og bidrager til BEP'er - definerer den overordnede virksomhedsdatamodel, som grænsen for en PIM og AIM kan overlejres på."

- Paul Ransley, chefkonsulent hos Acmena og systemintegrationsingeniør hos Transport for London.

"Hvis "data er den nye olie", skal vi lære at definere den, finde den, udvinde den, forfine den for at gøre den værdifuld. Jeg har fundet bogen DataDrivenConstruction meget informativ og indsigtsfuld. Bogen giver en nyttig historisk baggrund og forklarer arbejdet med data i et almindeligt sprog. For dem, der er interesserede i digital transformation, giver den en god forståelse af data - hvordan de fungerer, hvordan de er struktureret, og hvordan de kan bruges."

- Ralph Montague, direktør for ArcDox, direktør for BIM Coordinators Summit og formand for BIM National Mirror Committee hos National Standards Authority of Ireland.

"Som det blev understreget i bogen, er information et afgørende aktiv for byggesektoren, og at have det i tilgængelige formater letter i høj grad nøjagtig beslutningstagning og fremskynder projektets tidslinjer. Bogen tilbyder en neutral og effektiv tilgang til at få adgang til og drage fordel af denne kilde i beslutningsprocessen. Den metode, der præsenteres i bogen, udnytter en moderne tilgang, der kombinerer kunstig intelligens-drevet programmering med tilgængelige open source-værktøjer. Ved at udnytte kraften i AI og bruge open source-software har metoden til formål at forbedre automatiseringen, optimere processerne og fremme tilgængeligheden og samarbejdet inden for området. Sproget i bogen er klart og let at følge."

- Dr. Salih Ofluoğlu, dekan for fakultetet for kunst og arkitektur ved Antalya Bilim University og arrangør af Eurasian BIM Forum.

"Alt, hvad jeg kan sige, er: WOW! Den måde, du har indarbejdet historien, LLM, grafikken og den generelle lethed til at forstå dine pointer, er virkelig bemærkelsesværdig. Bogens flow er fantastisk. Der er så mange geniale aspekter ved denne bog; den er virkelig en game-changer. Det er en fantastisk kilde til information, og jeg roser dig for den indsats og passion, du har lagt i den. Tillykke med at have skabt et så bemærkelsesværdigt værk. Jeg kunne blive ved, men det er nok at sige, at jeg er utroligt imponeret!"

- Natasha Prinsloo, Digital Practice Lead hos energylab_.

"For alle i byggebranchen, fra nybegyndere til garvede professionelle, er denne bog en game-changer! Det er ikke en typisk støvet bog - den er fyldt med indsigt, strategier og et strejf af humor, som holder dig beskæftiget. Den dækker udviklingen af dataanvendelse i byggeriet, lige fra gamle dataregistreringsmetoder til banebrydende digitale teknologier. Det er som at tage en tidsmaskine gennem udviklingen af byggedata. Uanset om du er arkitekt, ingeniør, projektleder eller dataanalytiker, vil denne omfattende guide revolutionere den måde, du griber projekter an på. Gør dig klar til at optimere processer, forbedre beslutningstagningen og styre projekter som aldrig før!"

- Pierpaolo Vergati, lektor ved Sapienza-universitetet i Rom og Senior Construction Project Manager hos Fintecna.

"Jeg læste bogen i ét åndedrag, på mindre end 6 timer. Bogens fremstillingskvalitet er fremragende, tæt blankt papir, farveskemaer, en behagelig skrifttype. Det store antal praktiske eksempler på, hvordan man arbejder med LLM, der er specifikke for byggebranchen, vil spare dig for måneder, hvis ikke år, med selvstudier. Arbejdseksemplerne er meget forskellige og spænder fra enkle til komplekse, uden at du behøver at købe kompleks og dyr software. Bogen vil give ejere af enhver virksomhed i byggebranchen mulighed for at se på deres forretningsstrategi, digitalisering og udviklingsmuligheder med friske øjne. Og for mindre virksomheder at øge effektiviteten med overkommelige og gratis værktøjer."

- Mikhail Kosarev, underviser og konsulent i digital transformation i byggebranchen hos TIM-ASG.

Bogen "DATA DRIVEN CONSTRUCTION" er en game-changer for alle, der er nysgerrige efter at vide, hvor byggebranchen er på vej hen i dataenes tidsalder. Artem skraber ikke bare i overfladen; han dykker dybt ned i den aktuelle udvikling, udfordringer og lovende muligheder inden for byggeriet. Det, der adskiller denne bog fra andre, er dens tilgængelighed - Artem forklarer komplekse ideer ved hjælp af relaterbare analogier, der gør indholdet let at forstå. Jeg synes, at bogen er utrolig informativ og samtidig engagerende. Kort sagt har Artem skabt en værdifuld ressource, som ikke kun informerer, men også inspirerer. Uanset om du er en erfaren professionel eller en nybegynder inden for byggeriet, vil denne bog udvide dit perspektiv og uddybe din forståelse af, hvor branchen er på vej hen. Den kan varmt anbefales!"

- Moayad Saleh, arkitekt og BIM Implementation Manager hos TMM GROUP Gesamtplanungs GmbH.

"Jeg vil sige, at Data-Driven Construction er værdig til at blive undervist som lærebog på universiteterne og er en bog, der vil give værdifulde bidrag til det udviklende BIM felt. Data-Driven Construction indeholder en teknisk ordliste, der forklarer begreberne meget godt. Emner, der er ekstremt vanskelige at forklare, gøres enkle og forståelige med et meget smukt visuelt sprog. Jeg synes, at det, der skal forklares i billederne, skal udtrykkes for læseren, selv om det er kortfattet. Forståeligheden af nogle billeder, med andre ord, at læse det visuelle kræver separat information. Jeg vil også gerne sige, at jeg er glad for at introducere Artem Boikos værdifulde arbejde i mine forelæsninger og seminarer på universiteterne."

- Dr. Ediz Yazicioglu, ejer af ArchCube og underviser i byggeprojektledelse på afdelingen for arkitektur på Istanbuls tekniske universitet og på Medipol-universitetet.

"Data Driven Construction formidler på en levende måde det grundlæggende i informationsbaseret arbejde med bygningsdata. En bog, der beskæftiger sig med informationsstrømme og grundlæggende økonomiske begreber og dermed adskiller sig fra andre BIM bøger, fordi den ikke kun repræsenterer en softwareproducents perspektiv, men også forsøger at formidle grundlæggende begreber. En bog, der er værd at læse og se."

- Jakob Hirn, CEO og medstifter af Build Informed GmbH og initiativtager til innovationsforummet "On Top With BIM ".

"Data er den nye olie", som man siger, så guldgraverne skal have de rette værktøjer og den rette tankegang til at udvinde værdi af denne ressource fra det 21. århundrede. Byggebranchen har alt for længe befundet sig på en glidebane med processer baseret på "3D-information", hvor projektleverancen er baseret på en andens færdige information (f.eks. de har allerede tegnet cirkel- eller søjlediagrammet), mens de underliggende "data" (f.eks. det rå regneark) er i stand til at levere meget mere, især fordi multidatafusion og AI giver ubegrænsede potentialer. Hvis du leverer (eller underviser/forsker i) byggeri, er denne bog din bedste - og indtil videre eneste - ressource til at navigere i den datadrevne verden, vi har fundet os selv i."

- Dr. Zulfikar Adamu, lektor i strategisk it i byggeriet ved LSBU, Storbritannien.

"Data-Driven Construction" af Artem Boiko er et imponerende værk, der giver et solidt grundlag for byggebranchen i en tid med konstant voksende teknologier og informationsmuligheder. Boiko formår at præsentere komplekse emner på en forståelig måde, samtidig med at han introducerer visionære ideer. Bogen er et gennemtænkt kompendium, der ikke kun fremhæver den aktuelle udvikling, men også giver et udsyn til fremtidige innovationer. Den kan varmt anbefales til alle, der ønsker at sætte sig ind i datadrevet byggeplanlægning og -udførelse."

- Markus Eiberger, lektor ved Stuttgart University of Applied Sciences, Senior Project Manager og Deputy Branch Manager hos Konstruktionsgruppe Bauen, bestyrelsesmedlem i BIM Cluster Baden-Württemberg Association.

Anden udgave, marts 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-01-0



Artem Boiko Ophavsret

boikoartem@gmail.com
info@datadrivenconstruction.io

Ingen del af denne bog må gengives eller overføres i nogen form eller på nogen måde, elektronisk eller mekanisk, herunder fotokopiering, optagelse eller noget informationslagrings- og genfindingssystem, uden forfatterens skriftlige tilladelse - undtagen til ikke-kommerciel distribution i uændret form. Bogen distribueres gratis og kan frit deles med andre brugere til personlige, uddannelsesmæssige eller forskningsmæssige formål, forudsat at forfatterskab og referencer til originalen bevares. Forfatteren bevarer alle ikke-proprietære rettigheder til teksten og giver ingen udtrykkelige eller underforståede garantier. Virksomheder, produkter og navne, der nævnes i bogen, kan være fiktive eller bruges som eksempler. Forfatteren er ikke ansvarlig for eventuelle konsekvenser af brugen af de givne oplysninger. Oplysningerne i bogen leveres "som de er" uden garanti for fuldstændighed eller relevans. Forfatteren er ikke ansvarlig for tilfældige skader eller følgeskader, der opstår som følge af brugen af oplysningerne, koden eller programmerne i denne bog. De kodeeksempler, der præsenteres i denne bog, er kun til uddannelsesmæssige formål. Læsere bruger dem på egen risiko. Forfatteren anbefaler, at alle softwareløsninger testes, før de bruges i et produktionsmiljø. Alle varemærker og produktnavne, der er nævnt i teksten, er varemærker, registrerede varemærker eller servicemærker tilhørende deres respektive virksomheder og tilhører deres respektive ejere. Brugen af disse navne i bogen indebærer ikke noget forhold til eller godkendelse af deres ejere. Omtale af tredjeparts produkter eller tjenester udgør ikke en anbefaling eller indebærer en godkendelse. Firma- og produktnavne, der bruges i eksemplerne, kan være varemærker tilhørende deres ejere. Links til tredjepartswebsteder gives for nemheds skyld og betyder ikke, at forfatteren støtter oplysningerne på disse websteder. Alle statistikker, citater og undersøgelser, der citeres, var aktuelle på tidspunktet for skrivningen. Data kan ændre sig over tid.

Denne bog er distribueret under en Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International-licens (CC BY-NC-ND 4.0). Du må kopiere og distribuere den til ikke-kommercielle formål, forudsat at du bevarer forfatterskabet og ikke foretager nogen ændringer.

CC BY-NC-ND 4.0

© 2024 Artyom Boiko. Første udgave.
© 2025 Artyom Boiko. Anden udgave, revideret og suppleret.
Alle rettigheder forbeholdes.

FORORD TIL ANDEN UDGAVE

Denne bog er resultatet af en livlig dialog med det professionelle samfund. Den er baseret på talrige faglige diskussioner om datahåndtering i byggebranchen, som fandt sted på forskellige faglige platforme og sociale medier. Disse diskussioner blev grundlaget for artikler, publikationer og visuelt materiale, som har skabt en bred respons i det professionelle samfund. Forfatterens indhold tiltrækker millioner af visninger hvert år på forskellige platforme og sprog og bringer fagfolk inden for digitalisering af byggeriet sammen.

Inden for et år efter udgivelsen af den første udgave blev bogen bestilt af eksperter fra mere end 50 lande, fra Brasilien og Peru til Mauritius og Japan. Den anden udgave af bogen, som du nu holder i hånden, er blevet revideret og udvidet på baggrund af ekspertfeedback, kritik af den første udgave og diskussioner i fagkredse. Takket være feedbacken er anden udgave blevet udvidet betydeligt: Der er tilføjet nye kapitler om CAD (BIM) teknologier og om at skabe effektive ETL -processer. Antallet af praktiske eksempler og casestudier er også blevet øget betydeligt. Af særlig værdi er tilbagemeldingerne fra ledere i byggebranchen, konsulentfirmaer og store it-virksomheder, som henvendte sig til forfatteren med spørgsmål om digitalisering og interoperabilitet både før og efter udgivelsen af den første version af bogen. Mange af dem har allerede anvendt de tilgange, der er beskrevet i bogen, eller planlægger at gøre det i den nærmeste fremtid

Du holder en bog i hånden, som er blevet til gennem diskussion og aktiv udveksling af synspunkter. Fremskridt fødes i dialog, i sammenstødet mellem synspunkter og åbenhed over for nye tilgange. Tak, fordi du er en del af denne dialog. Din konstruktive kritik er grundlaget for fremtidige forbedringer. Hvis du finder fejl i teksten eller gerne vil dele ideer og forslag, er al feedback velkommen. Kontaktoplysninger findes i slutningen af bogen

HVORFOR ER BOGEN GRATIS?

Denne bog er tænkt som en åben uddannelsesressource, der har til formål at udbrede moderne tilgange til datahåndtering i byggebranchen. Den første version af bogen tjente som grundlag for at indsamle kommentarer og forslag fra det professionelle samfund, hvilket gjorde det muligt at forbedre materialets struktur og indhold. Alle kommentarer, forslag og ideer er blevet omhyggeligt analyseret og indarbejdet i denne reviderede version. Målet med bogen er at hjælpe fagfolk i byggeriet med at forstå vigtigheden af at arbejde med data: systematisk, bevidst og med øje for informationens langsigtede værdi. Forfatteren har samlet eksempler, illustrationer og praktiske observationer fra mere end 10 års arbejde med digitalisering af byggeriet. Det meste af dette materiale stammer fra virkelige projekter, diskussioner med ingeniører og bygherrer, deltagelse i internationale initiativer og uddannelsesseminarer. Bogen er et forsøg på at strukturere den akkumulerede erfaring og dele den på en tilgængelig måde. Hvis du vil støtte yderligere udbredelse af bogens ideer og få et praktisk format til at læse, arbejde med eksempler og visuelle materialer - kan du købe en trykt version.

RETTIGHEDER TIL BRUG

Alle materialer, illustrationer og fragmenter i denne bog må gengives, citeres eller bruges i ethvert format og på ethvert medie, forudsat at kilden angives: forfatterskab af Artem Boiko og bogens titel "Data-Driven Construction". Tak for respekten for arbejde og formidling af viden.

Det er med oprigtig taknemmelighed, at jeg dedikerer denne bog til min familie, som fra en tidlig alder indpodede mig en dyb kærlighed til byggeri, til min hjemlige mineby for lektioner i modstandsdygtighed og til min landmålerhustru, hvis usvigelige støtte har været min konstante inspiration.

HVEM DENNE BOG ER TIL

Denne bog er skrevet i et tilgængeligt sprog og henvender sig til en bred vifte af læsere i byggebranchen - fra studerende og nybegyndere, der ønsker at forstå det grundlæggende i moderne byggeprocesser, til fagfolk, der har brug for en opdateret metode til at styre data i byggeriet. Uanset om du er arkitekt, ingeniør, formand, byggeleder eller dataanalytiker, giver denne omfattende guide med mange unikke illustrationer og grafikker værdifuld indsigt i, hvordan man bruger data i erhvervslivet til at optimere og automatisere processer, forbedre beslutningstagningen og styre byggeprojekter på forskellige niveauer hjælp af moderne værktøjer.

Bogen er en omfattende guide, der kombinerer teoretisk grundlag og praktiske anbefalinger til integration af datastyringsteknikker i byggeprocesser. Bogen fokuserer på strategisk brug af information til at optimere driften, automatisere processer, forbedre beslutningstagningen og effektivt styre projekter ved hjælp af moderne digitale værktøjer.

Denne bog dækker de teoretiske og praktiske aspekter af at arbejde med information i byggebranchen. Gennem detaljerede eksempler udforskes metodologien for opgaveparameterisering, indsamling af krav, behandling af ustrukturerede data i flere formater og omdannelse af dem til effektive løsninger for byggevirksomheder.

Læseren går successivt fra opstilling af krav og udvikling af grundlæggende datamodeller til mere komplekse processer med integration af heterogene informationskilder, oprettelse af ETL -processer, opbygning af informationspipelines og maskinlæringsmodeller. Den sekventielle tilgang gør det muligt klart at demonstrere mekanismerne for organisering og automatisering af forretningsprocesser og beslutningsstøttesystemer i byggebranchen. Hver del af bogen afsluttes med et praktisk kapitel, der indeholder trinvise instruktioner, så du straks kan anvende den erhvervede viden i virkelige projekter.

SYNOPSIS AF DELE AF BOGEN

Denne bog er struktureret omkring begrebet datatransformation i værdikæden: fra dataindsamling og kvalitetssikring til analytisk behandling og udtrækning af værdifulde praktiske løsninger ved hjælp af moderne værktøjer og metoder.

Del 1: Digital udvikling i byggeriet - sporer den historiske transformation af datahåndtering fra lertavler til moderne digitale systemer, analyserer fremkomsten af modulære systemer og den voksende betydning af informationsdigitalisering i forbindelse med industrielle revolutioner.

Del 2: Informationsudfordringer for byggebranchen - udforsker problemerne med datafragmentering, "informationssiloer", HiPPO-tilgangens indvirkning beslutningstagningen og begrænsningerne ved proprietære formater og foreslår, at man overvejer overgangen til AI- og LLM-økosystemer.

Del 3: Systematisering af data i byggeriet - opstiller en typologi for byggedata, beskriver metoder til organisering og integration med virksomhedssystemer og diskuterer oprettelsen af kompetencecentre til standardisering af informationsprocesser.

Del 4: Kvalitetssikring af data - afslører metoder til at omdanne forskelligartet information til strukturerede kvalitetsdata, herunder dataudtræk fra forskellige kilder, validering og modellering ved hjælp af LLM.

Del 5: Omkostnings- og tidsberegninger - omhandler digitalisering af omkostnings- og planlægningsberegninger, automatisering af indhentning af mængder fra CAD (BIM)-modeller, 4D-8D-modelleringsteknologier og ESG-beregning af byggeprojekter.

Del 6: CAD og BIM - analyserer kritisk udviklingen af designteknologier, problemer med systemers interoperabilitet, tendenser til åbne dataformater og udsigterne til at anvende kunstig intelligens til design.

Del 7: Dataanalyse og automatisering - ser på principperne for informationsvisualisering, key performance indicators, ETL-processer, workflow-orkestreringsværktøjer og anvendelse af sprogmodeller til at automatisere rutineopgaver.

Del 8: Datalagring og -styring - udforsker datalagringsformater, data warehouse- og data lake-koncepter, datastyringsprincipper og nye tilgange, herunder vektordatabaser og DataOps- og VectorOps-metodologierne.

Del 9: Big Data og maskinlæring - fokuserer på overgangen til objektiv analyse baseret på historiske data, tingenes internet på byggepladser og anvendelse af maskinlæringsalgoritmer til at forudsige projektomkostninger og tidslinjer.

Del 10: Byggebranchen i de digitale datas tidsalder - præsenterer et kig på byggebranchens fremtid og analyserer skiftet fra kausalanalyse til arbejde med korrelationer, begrebet "Uberisering" af byggeriet og strategier for digital transformation.

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

Led by CEO and project managers

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

Focuses on real-time analytics

It's a mindset shift!

Streamline operations & enhance collaboration

The ultimate goal



INDLEDNING

Hvor længe kan din virksomhed forblive konkurrencedygtig i en verden, hvor teknologien udvikler sig hurtigt, og alle aspekter af forretningen, fra timing og omkostningsberegning til risikoanalyse, bliver automatiseret af maskinlæringsmodeller?

Byggebranchen, som har eksisteret lige så længe som menneskeheden selv, står på tærsklen til revolutionerende forandringer, som lover at ændre den måde, vi tænker på traditionelt byggeri, fuldstændigt. Allerede i andre økonomiske sektorer ændrer digitaliseringen ikke bare reglerne, men driver hensynsløst de virksomheder ud af markedet, som ikke har formået at tilpasse sig det nye databehandlingsmiljø og ikke er i stand til at forbedre beslutningshastigheden (fig. 1).



Fig. 1 Beslutningshastigheden i byggebranchen afhænger oftere af den menneskelige faktor end i andre brancher.

Bankvæsen, detailhandel, logistik og landbrugsindustri bevæger sig hurtigt mod fuld digitalisering, hvor unøjagtigheder og subjektive meninger ikke længere har nogen plads. Moderne algoritmer er i stand til at analysere enorme mængder data og give kunderne præcise forudsigelser - uanset om det drejer sig om sandsynligheden for tilbagebetaling af lån, optimale leveringsruter eller risikoprognoser.

Byggeriet er en af de sidste brancher, der foretager den uundgåelige overgang fra løsninger baseret på højt betalte specialisters meninger til datadrevne løsninger. Denne overgang er ikke kun drevet af nye teknologiske muligheder, men også af markedets og kundernes øgede krav om gennemsigtighed, nøjagtighed og hastighed.

Robotisering, procesautomatisering, åbne data og prognoser baseret på dem er ikke længere bare muligheder, men uundgåelige. De fleste virksomheder i byggebranchen, som for nylig var ansvarlige over for bygherren for at beregne mængden, omkostningerne, projekttiden og kvalitetskontrollen, risikerer nu at blive til rene udførere af ordrer uden at træffe vigtige beslutninger (fig. 2).

Med fremskridt inden for computerkraft, maskinlæringsalgoritmer og demokratiseret adgang til data er det nu muligt automatisk at kombinere data fra forskellige kilder, hvilket muliggør dybere procesanalyse, risikoforudsigelse og omkostningsoptimering i de indledende faser af et byggeprojekt. Disse teknologier har potentiale til radikalt at forbedre effektiviteten og reducere omkostningerne i hele sektoren.

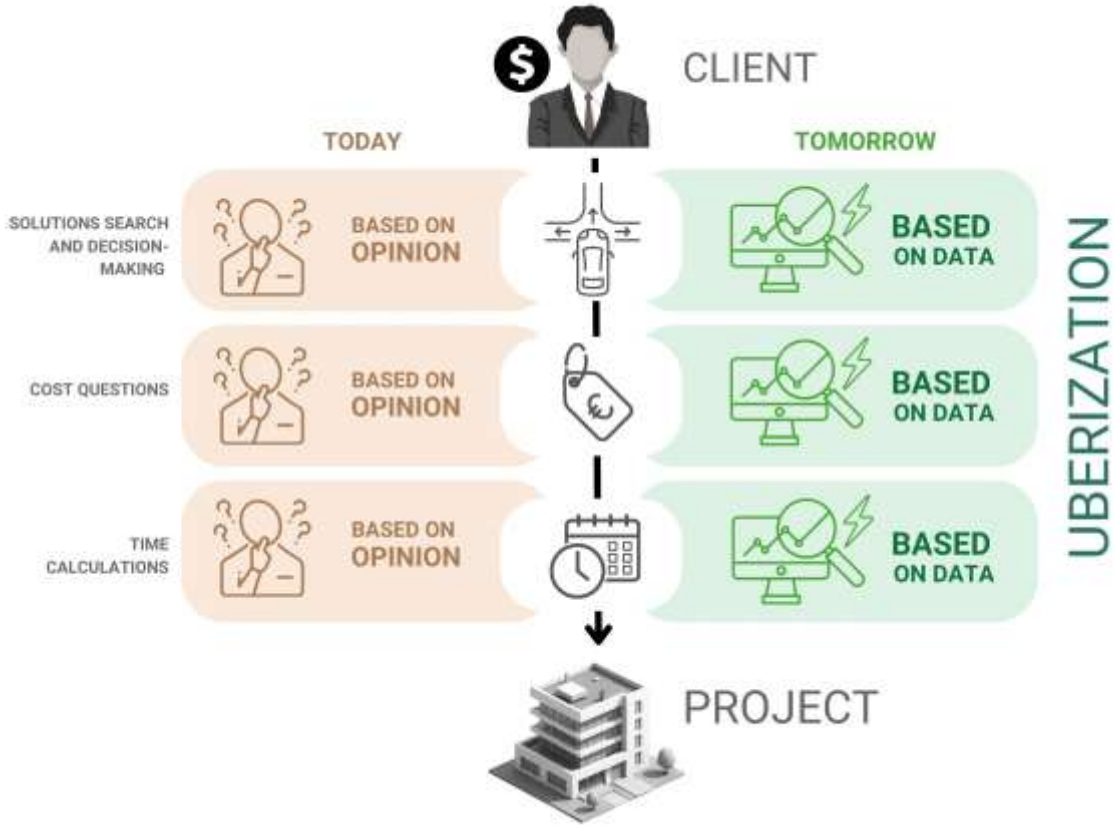


Fig. 2 Bygherren er ikke interesseret i for mange menneskelige faktorer i forbindelse med realiseringen af sit projekt.

På trods af alle fordelene ved nye værktøjer og koncepter halter byggebranchen langt bagefter andre sektorer i økonomien, når det gælder om at tage nye teknologier i brug.

> Ifølge rapporten IT Metrics Key Data 2017 ligger byggebranchen sidst i it-udgifter blandt 19 andre brancher [1].

Den hurtige vækst i datamængden og kompleksiteten i processerne er ved at blive en hovedpine for virksomhedsledelsen, og det største problem ved at bruge nye teknologier er, at data på trods af deres overflod forbliver fragmenterede, ustrukturerede og ofte inkompatible mellem forskellige systemer og softwareprodukter. Derfor er mange virksomheder i byggesektoren nu primært bekymrede over problemer med datakvalitet, som kun kan løses med implementering af effektive, automatiserede styrings- og analysesystemer.

> Ifølge en undersøgelse foretaget af KPMG® blandt byggeledere i 2023 [2] har projektstyringsinformationssystemer (PMIS), avanceret og grundlæggende dataanalyse og bygningsinformationsmodellering (BIM) det største potentiale til at forbedre projekternes ROI (figur 3).

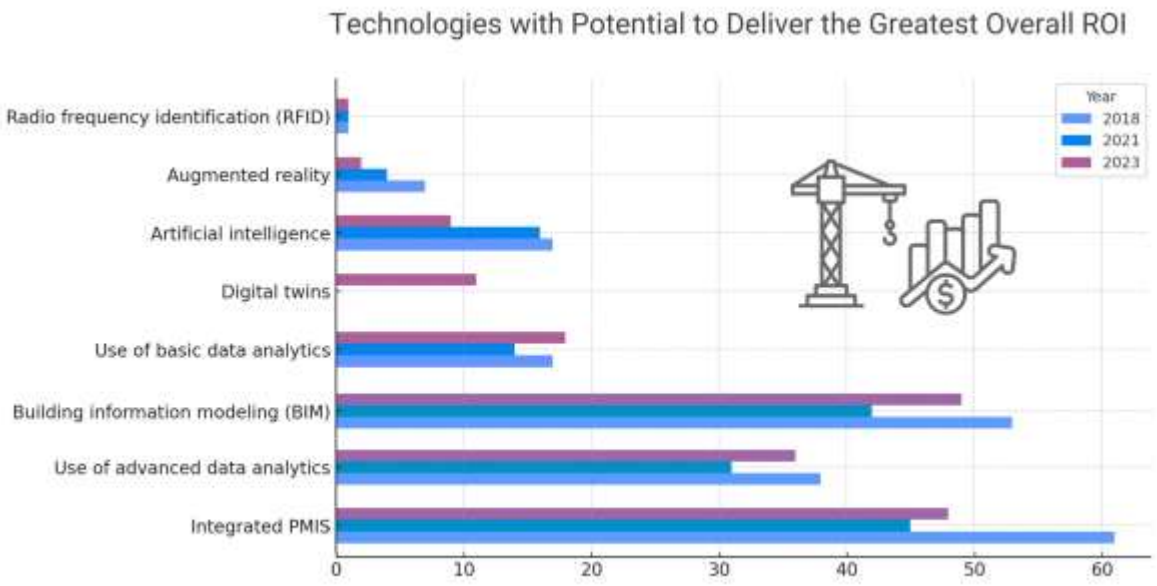


Fig. 3 Undersøgelse blandt ledere af byggevirksomheder: Hvilke teknologier vil give det højeste investeringsafkast (ROI) i kapitalprojekter? (baseret på materialer [2]).

Løsningen på de udfordringer, der er forbundet med at integrere data i forretningsprocesser, er at sikre information af høj kvalitet, bruge passende dataformater og anvende effektive metoder til at skabe, lagre, analysere og behandle data.

Erkendelsen af værdien af data tvinger forskellige brancher til at bevæge sig væk fra siloopdelte applikationer og komplekse bureaukratiske ledelsesstrukturer. I stedet skifter fokus til at skabe nye tilgange til informationsarkitektur og omdanne virksomheder til moderne datadrevne virksomheder. Før eller siden vil byggebranchen selv tage dette skridt og bevæge sig fra en gradvis digital udvikling til en ægte digital revolution, der påvirker alle virksomheder.

> Overgangen til datadrevne forretningsprocesser bliver ikke let. Mange virksomheder vil møde udfordringer, fordi lederne ikke altid forstår, hvordan de skal bruge kaotiske datasæt til at forbedre effektiviteten og væksten i virksomheden.

Denne bog dykker ned i en verden af data, hvor information er ved at blive en vigtig strategisk ressource, der bestemmer effektiviteten og bæredygtigheden af forretningsprocesser. Med den hurtige vækst i information står virksomheder over for nye udfordringer. Digital transformation er ikke længere bare et buzzword - det er ved at blive en nødvendighed.



Fig. 4 Data og processer er fundamentet for byggeriet.

At forstå transformation betyder at kunne forklare det komplekse med enkle ord. Derfor er bogen skrevet i et tilgængeligt sprog og er ledsaget af forfatterens illustrationer, der er skabt specielt til at forklare nøglebegreber tydeligt. Disse diagrammer, tabeller og visualiseringer er designet til at fjerne barrierer for opfattelsen og gøre materialet forståeligt selv for dem, der tidligere anså sådanne emner for at være for komplekse. Alle illustrationer, diagrammer og grafikker i denne bog er skabt af forfatteren og designet specifikt til at visualisere de nøglebegreber, der beskrives i teksten.

> *Et billede siger mere end tusind ord [3].*
>
> - Fred R. Barnard, engelsk illustrator, 1927.

For at forbinde teori med praksis bruger vi værktøjer til kunstig intelligens (især sprogmodeller), som giver dig mulighed for at udvikle løsninger uden behov for dyb viden om programmering. Hvis du er orienteret mod praktisk materiale og er mere interesseret i praktisk arbejde med data, kan du springe den første introduktionsdel over og gå direkte til anden del af bogen, hvor beskrivelsen af konkrete eksempler og cases starter.

Men man skal ikke have for store forventninger til AI (kunstig intelligens), maskinlæring og LLM-værktøjer (Large Language Models) i almindelighed. Uden inputdata af høj kvalitet og en dyb forståelse af emnet kan selv de mest avancerede algoritmer ikke give pålidelige og meningsfulde resultater.

> Microsofts CEO Satya Nadella advarer om risikoen for en boble inden for kunstig intelligens i begyndelsen af 2025 [4] og sammenligner den nuværende hype med dot-com-boblen. Han understreger, at påstande om at nå AGI-milepæle (Artificial General Intelligence) uden ordentlig begrundelse er "meningsløs manipulation af metrikker". Nadella mener, at den virkelige succes med AI skal måles på dens bidrag til den globale BNP-vækst i stedet for at lægge for stor vægt på buzzwords.

Bag alle buzzwords om nye teknologier og koncepter ligger et komplekst og omhyggeligt arbejde med at

sikre datakvalitet, parametrisering af forretningsprocesser og tilpasning af værktøjer til reelle opgaver.

> En datadrevet tilgang er ikke et produkt, man bare kan downloade eller købe. Det er en strategi, der skal opbygges. Den starter med et nyt blik på eksisterende processer og problemer og kræver derefter en disciplineret bevægelse i den valgte retning.

Førende softwareudviklere og applikationsleverandører vil ikke være motoren for forandring i byggebranchen, for mange af dem er en datadrevet tilgang en trussel mod deres etablerede forretningsmodel.

> *Andre brancher [i modsætning til byggeriet], som f.eks. bilindustrien, har allerede gennemgået radikale og forstyrrende forandringer, og deres digitale transformation er godt i gang. Byggevirksomhederne skal handle hurtigt og beslutsomt: De kvikke virksomheder vil høste store gevinster, mens de, der tøver, løber en stor risiko. Tænk på den omvæltning, som digital fotografering har forårsaget i denne branche [5].*
>
> - World Economic Forums rapport Shaping the Future of Construction, 2016

De virksomheder, der i tide erkender mulighederne og fordelene ved den nye tilgang, vil få en bæredygtig konkurrencefordel og vil være i stand til at udvikle sig og vokse uden at være afhængige af løsninger fra store leverandører.

Dette er din chance for ikke bare at klare den kommende storm med digitalisering af information, men for at tage kontrol over den. I denne bog finder du ikke bare en analyse af branchens nuværende tilstand, men også konkrete anbefalinger til at gentænke og omstrukturere dine processer og din virksomhed for at blive førende i den nye æra inden for byggeriet og forbedre din professionelle erfaring.

Byggeriets digitale fremtid handler ikke kun om at bruge nye teknologier og programmer, men også om grundlæggende at gentænke datahåndtering og forretningsmodeller.

Er din virksomhed klar til denne strategiske forandring?

INDHOLDSFORTEGNELSE

INDLEDNING **1**

INDHOLDSFORTEGNELSE ..... **I**

II DEL FRA LERTAVLER TIL DEN DIGITALE REVOLUTION: HVORDAN INFORMATION HAR UDVIKLET SIG I BYGGERIET **2**

KAPITEL 2.1. UDVIKLINGEN I BRUGEN AF DATA I BYGGEBRANCHEN 3

Fødslen af dataæraen i byggeriet 3

Fra ler og papyrus til digital teknologi 4

Proces som værktøj til datadrevet oplevelse 5

Digitalisering af information om byggeprocessen 7

KAPITEL 2.2. TEKNOLOGIER OG LEDELSESSYSTEMER I MODERNE BYGGERI 11

Den digitale revolution og fremkomsten af modulære MRP/ERP -systemer 11

Datastyringssystemer: fra datamining til forretningsudfordringer 13

Virksomhedens mycelium: hvordan data forbindes med forretningsprocesser 16

KAPITEL 2.3. DEN DIGITALE REVOLUTION OG EKSPLOSIONEN AF DATA 20

Begyndelsen på data-boomet som en evolutionær bølge 20

Mængden af data, der genereres i en moderne virksomhed 22

Omkostningerne ved datalagring: det økonomiske aspekt 23

Grænser for dataakkumulering: fra masse til mening 25

Næste skridt: fra datateori til praktisk forandring 27

III DEL HVORDAN BYGGEBRANCHEN DRUKNER I DATA-KAOS **28**

KAPITEL 3.1. DATAFRAGMENTERING OG SILOER 29

Jo flere værktøjer, jo mere effektiv virksomhed? 29

Datasiloer og deres indvirkning på virksomhedens resultater 31

Dobbeltarbejde og manglende datakvalitet som følge af uenighed 34

HiPPO eller faren ved meninger i beslutningsprocessen 36

Kontinuerlig stigning i forretningsprocessernes kompleksitet og dynamik 39

Den fjerde industrielle revolution (Industri 4.0) og den femte industrielle revolution (Industri 5.0) i byggeriet 42

KAPITEL 3.2. AT GØRE KAOS TIL ORDEN OG REDUCERE KOMPLEKSITETEN 45

Overflødig kode og lukkede systemer som en barriere for produktivitetsforbedringer 45

Fra siloer til et enkelt data warehouse 47
Integrerede lagersystemer muliggør overgangen til AI agenter 48
Fra dataindsamling til beslutningstagning: vejen til automatisering 51
Næste skridt: Gør kaos til et håndterbart system 53
IV DEL DATARAMME I FORRETNINGSPROCESSER I BYGGERIET 55
KAPITEL 4.1. DATATYPER I BYGGERIET 56
De vigtigste datatyper i byggebranchen 56
Strukturerede data 60
Relationelle databaser RDBMS og SQL-forespørgselssprog 61
SQL - forespørgsler i databaser og nye tendenser 64
Ustrukturerede data 66
Tekstdata: mellem ustruktureret kaos og struktureret kaos y 67
Semistrukturerede og løst strukturerede data 68
Geometriske data og deres anvendelse 69
CAD data: fra design til datalagring 72
Fremkomsten af BIM-konceptet (BOM) og brugen af CAD i processer 75
KAPITEL 4.2. DATAHARMONISERING OG -STRUKTURERING 81
Fyldning af systemer med data i byggebranchen 81
Datatransformation: det kritiske fundament for moderne forretningsanalyse 83
Datamodeller: relationer i data og relationer mellem elementer 87
Proprietære formater og deres indvirkning på digitale processer 91
Åbne formater ændrer tilgangen til digitalisering 95
Paradigmeskift: Open Source som afslutningen på æraen med softwareleverandørernes dominans 96
Strukturerede åbne data: fundamentet for digital transformation 98
KAPITEL 4.3. LLM OG DERES ROLLE I DATABEHANDLING OG FORRETNINGSPROCESSER 102
LLM chatrum: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok for at automatisere dataprocesser. 102
Store LLM-sprogmodeller: Sådan fungerer det 103
Brug af lokale LLM'er til følsomme virksomhedsdata 106
Fuld kontrol over AI i virksomheden, og hvordan du implementerer din egen LLM 108
RAG: Intelligent LLM -assistenter med adgang til virksomhedsdata 110
KAPITEL 4.4. IDE MED LLM-SUPPORT OG FREMTIDIGE PROGRAMÆNDRINGER 112
Valg af IDE: fra LLM eksperimenter til forretningsløsninger 112

IDE med LLM-understøttelse og fremtidige programmeringsændringer 114
Python Pandas: et uundværligt værktøj til at arbejde med data 115
DataFrame: universelt tabeldataformat 119
Næste skridt: opbygning af en bæredygtig dataramme 122
V DEL DATAKVALITET: ORGANISERING, STRUKTURERING, MODELLERING 124
KAPITEL 5.1. KONVERTERING AF DATA TIL EN STRUKTUREret FORM 125
At lære at omdanne dokumenter, PDF, billeder og tekster til strukturerede formater 125
Eksempel på konvertering af et PDF -dokument til en tabel 126
Konverter JPEG- og PNG-billeder til struktureret form 130
Konvertering af tekstdata til en struktureret form 132
Konvertering af CAD-data (BIM) til en struktureret form 135
Leverandører af CAD -løsninger bevæger sig mod strukturerede data 140
KAPITEL 5.2. KLASSIFIKATION OG INTEGRATION: ET FÆLLES SPROG FOR BYGGEDATA 143
Hastigheden af beslutningstagningen afhænger af kvaliteten af data 143
Datastandardisering og -integration 144
Digital interoperabilitet starter med krav 146
Et fælles konstruktionssprog: Klassifikatorernes rolle i den digitale transformation 149
Masterformat, OmniClass, Uniclass og CoClass: udviklingen af klassifikationssystemer 151
KAPITEL 5.3. DATAMODELLERING OG EKSPERTISECENTER 156
Datamodellering: konceptuel, logisk og fysisk model 156
Praktisk datamodellering i forbindelse med byggeri 160
Oprettelse af en database ved hjælp af LLM 162
Kompetencecenter (CoE) for datamodellering 164
KAPITEL 5.4. SYSTEMATISERING AF KRAV OG VALIDERING AF INFORMATION 167
Indsamling og analyse af krav: omdanner kommunikation til strukturerede data 167
Procesflowdiagrammer og effektiviteten af konceptuelle rammer 171
Strukturerede krav og RegEx regulære udtryk 173
Dataindsamling til verifikationsprocessen 178
Verifikation af data og resultater af verifikation 180
Visualisering af verifikationsresultater 185
Sammenligning af datakvalitetstjek med menneskelige livsbehov 187
Næste skridt: Omdan data til præcise beregninger og planer 189
VI DEL OMKOSTNINGS- OG TIDSBEREGNINGER: INDARBEJDELSE AF DATA I

BYGGEPROCESSER 191

KAPITEL 6.1. OMKOSTNINGSBEREGNINGER OG -OVERSLAG FOR BYGGEPROJEKTER 192

Grundlæggende om byggeri: estimering af mængde, pris og tid 192

Metoder til beregning af de anslåede projektomkostninger 193

Ressourcebaseret metode omkostningsberegning og estimering i byggeriet 194

Database over byggeressourcer: katalog over byggematerialer og -arbejder 194

Udarbejdelse af beregninger og omkostningsberegning af arbejder baseret på ressourcegrundlag. 196

Endelig projektberegning: fra estimater til budgetter 201

KAPITEL 6.2. MÆNGDEBEREGNING OG AUTOMATISK OPRETTELSE AF ESTIMATER OG TIDSPLANER 205

At gå fra 3D til 4D og 5D: ved hjælp af volumetriske og kvantitative parametre 205

5D-attributter og hentning af attributmængder fra CAD 205

QTO Quantity Take-Off: gruppering af projektdata efter attributter 209

QTO-automatisering ved hjælp af LLM og strukturerede data 214

QTO beregning af hele projektet ved hjælp af grupperegler fra et Excel-regneark 218

KAPITEL 6.3. 4D, 6D -8D OG BEREGNING AF KULDIOXIDEMISSIONER CO_2 224

4D-model: Integration af tid i byggeoverslag 224

Byggeplan og dens automatisering baseret på omkostningsdata 225

Udvidede attributlag 6D -8D: fra energieffektivitet til sikkerhedssikring 227

Estimering af CO_2 og beregning af kuldioxidudledning fra byggeprojekter 230

KAPITEL 6.4. KONSTRUKTION ERP OG PMIS SYSTEMER 234

Construction ERP -systemer på eksemplet med beregninger og estimater 234

PMIS: Mellemled mellem ERP og byggepladsen 239

Spekulation, profit, isolerethed og mangel på gennemsigtighed i ERP og PMIS 240

Slut på æraen med lukkede ERP /PMIS: Byggebranchen har brug for nye tilgange 242

Næste skridt: effektiv brug af projektdata 244

VII DEL CAD OG BIM: MARKEDSFØRING, VIRKELIGHED OG FREMTIDEN FOR DESIGNDATA I BYGGERIET 247

KAPITEL 7.1. FREMKOMSTEN AF BIM- KONCEPTER I BYGGEBRANCHEN 248

Historien om fremkomsten af BIM og open BIM som markedsføringskoncepter for CAD- leverandører 248

BIM's virkelighed: I stedet for integrerede databaser - lukkede modulære systemer 251

Fremkomsten af det åbne format IFC i byggebranchen 253

IFC formatproblem afhængigt af den geometriske kerne 255

Optræden i konstruktionen af emnet semantik og ontologi 258
Hvorfor semantiske teknologier ikke lever op til forventningerne i byggebranchen 260
KAPITEL 7.2. LUKKEDE PROJEKTFORMATER OG PROBLEMER MED INTEROPERABILITET 263
Lukkede data og faldende produktivitet: CAD-industriens blindgyde (BIM) 263
Myten om interoperabilitet mellem CAD-systemer 265
Gå til USD og granulære data m 269
KAPITEL 7.3. GEOMETRI I BYGGERIET: FRA LINJER TIL KUBIKMETER 273
Når linjer bliver til penge, eller hvorfor bygherrer har brug for geometri 273
Fra linjer til volumener: Hvordan areal og volumen bliver til data 273
Flytning til MESH, USD og polygoner: brug af tessellation til geometri 275
LOD, LOI, LOMD - unik klassifikation af detaljer i CAD (BIM) 277
Nye CAD-standarder (BIM) - AIA, BEP, IDS, LOD, COBie 280
KAPITEL 7.4. DESIGNPARAMETERISERING OG BRUG AF LLM TIL CAD-DRIFT 284
Illusionen om unikke CAD-data (BIM): vejen til analyse og åbne formater 284
Design gennem parametre: fremtiden for CAD og BIM 287
Fremkomsten af LLM i design CAD databehandlingsprocesser 290
Automatiseret analyse af DWG -filer med LLM og Pandas 293
Næste skridt: fra lukkede formater til åbne data 299
VIII DEL DATADREVET BESLUTNINGSTAGNING, ANALYSE, AUTOMATISERING OG MASKINLÆRING 301
KAPITEL 8.1. DATAANALYSE OG DATADREVET BESLUTNINGSTAGNING 302
Data som en ressource i beslutningsprocessen 302
Visualisering af data: nøglen til forståelse og beslutningstagning 305
KPIog ROI 307
Dashboards og dashboards: Visualisering af målinger til effektiv ledelse 309
Dataanalyse og kunsten at stille spørgsmål 311
KAPITEL 8.2. DATAFLOW UDEN MANUEL INDSATS: HVORFOR DER ER BRUG FOR ETL 314
ETL-automatisering: reducerer omkostninger og fremskynder datahåndtering 314
ETL Udtræk: dataindsamling 317
ETL Transform: anvendelse af validerings- og transformationsregler 320
ETL Load: Visualiser resultater i diagrammer og grafer 323
ETL Load: Automatisk oprettelse af PDF-dokumenter 328
ETL Load: automatisk generering af dokumenter fra FPDF 329

ETL Load: Rapportering og indlæsning til andre systemer 333
ETL med LLM: Visualisering af data fra PDF -dokumenter 334
KAPITEL 8.3. AUTOMATISK ETL TRANSPORTØR (RØRLEDNING) 339
Pipeline: Automatisk ETL transportør data 339
Rørledning -ETL datavalideringsproces med LLM 343
Rørledning -ETL: verifikation af data og oplysninger om projektelementer i CAD (BIM) 345
KAPITEL 8.4. ORKESTRERING AF ETL OG WORKFLOWS: PRAKTISKE LØSNINGER 351
DAG og Apache Airflow: automatisering og orkestrering af arbejdsgange 351
Apache Airflow: praktisk anvendelse ETL-automatisering 352
Apache NiFi til routing og datakonvertering 356
n8n Low-Code, No-Code procesorkestrering 357
Næste skridt: at gå fra manuel drift til analysebaserede løsninger 360
IX DEL DATALAGRING OG -STYRING I BYGGERIET 362
KAPITEL 9.1. DATAINFRASTRUKTUR: FRA LAGRINGSFORMATER TIL DIGITALE ARKIVER 363
Dataatomer: grundlaget for effektiv informationsstyring 363
Lagring af information: filer eller data 364
Big Data Storage: Analyse af populære formater og deres effektivitet 366
Optimer storage med Apache Parquet 369
DWH: Data Warehouse data warehouses 371
Data Lake - udvikling af ETL til ELT: fra traditionel rensning til fleksibel behandling 372
Data Lakehouse-arkitektur: synergi mellem lagre og datasøer 374
CDE, PMIS, ERP eller DWH og Data Lake 376
KAPITEL 9.2. STYRING AF DATALAGER OG FOREBYGGELSE AF KAOS 380
Vektordatabaser og den afgrænsende boks 380
Data Governance), Data Minimalism) og Data Swamp) 383
DataOps og VectorOps: nye datastandarder 386
Næste skridt: fra kaotisk opbevaring til struktureret opbevaring 388
X DEL BIG DATA, MASKINLÆRING OG FORUDSIGELSER 390
KAPITEL 10.1. BIG DATA OG ANALYSE HERAF 391
Big data i byggeriet: fra intuition til forudsigelighed 391
Spørgsmål om gennemførligheden af big data: korrelation, statistik og datasampling 392
Big data: analyse af data fra San Franciscos datasæt med en million byggetilladelser 395
Eksempel på big data baseret på CAD-data (BIM) 400

IoT Internet of Things og smarte kontrakter 404
KAPITEL 10.2. MASKINLÆRING OG FORUDSIGELSER 408
Maskinlæring og kunstig intelligens vil ændre den måde, vi bygger på 408
Fra subjektiv vurdering til statistisk prognose 410
Titanic-datasæt: Hello World i analyseverdenen data og big data 412
Machine learning i aktion: fra Titanic-passagerer til projektledelse 417
Forudsigelser og prognoser baseret på historiske data 421
Nøglebegreber inden for maskinlæring 423
KAPITEL 10.3. FORUDSIGELSE AF OMKOSTNINGER OG DEADLINES VED HJÆLP AF MASKINLÆRING 426
Et eksempel på brug af maskinlæring til at finde projektomkostninger og tidslinjer 426
Forudsigelse af projektomkostninger og -tid ved hjælp af lineær regression 428
Forudsigelse af projektomkostninger og -tid ved hjælp af K-nearest neighbour-algoritmen (k-NN) .. 431
Næste skridt: fra opbevaring til analyse og prognoser 435
XI DEL BYGGEBRANCHEN I DEN DIGITALE TIDSALDER. MULIGHEDER OG UDFORDRINGER..... 438
KAPITEL 11.1. OVERLEVELSESSTRATEGIER: OPBYGNING AF KONKURRENCEFORDELE..... 439
Korrelationer i stedet for beregninger: fremtiden for byggeanalyser 439
Datadrevet tilgang i byggeriet: et nyt niveau af infrastruktur..... 442
Den næste generation af det digitale kontor: Hvordan AI ændrer arbejdsområdet 444
Åbne data og uberisering er en trussel mod eksisterende byggevirksomheder 446
Uløste problemer med uberisering som sidste chance for at bruge tiden til transformation 449
KAPITEL 11.2. EN PRAKTISK GUIDE TIL IMPLEMENTERING AF EN DATADREVET TILGANG..... 454
Fra teori til praksis: En køreplan for digital transformation i byggeriet..... 454
At lægge det digitale fundament: 1-5 trin til digital modenhed 456
Frigør potentialet i data: 5-10 trin til digital modenhed..... 460
Køreplan for transformation: fra kaos til datadrevet virksomhed 466
Byggeri i Industri 5.0: Sådan tjener du penge, når du ikke længere kan gemme dig..... 469
KONKLUSION..... 471
OM FORFATTEREN 474
YDERLIGERE RELATION..... 475
KOMMENTAR TIL OVERSÆTTELSE 475
ANDRE FÆRDIGHEDER OG KONCEPTER..... 476
GLOSSAR 480

LISTE OVER REFERENCER OG ONLINE-MATERIALER 486

EMNEINDEKS 502

MAKSIMAL BEKVEMMELIGHED MED DEN TRYKTE VERSION

Du sidder med en gratis digital version af **Datadrevet byggeri**. For mere praktisk arbejde og hurtig adgang til materialerne anbefaler vi, at du er opmærksom på den **trykte udgave**:



- **Altid ved hånden**: Bogen i trykt format bliver et pålideligt arbejdsredskab, så du hurtigt kan finde og bruge de rigtige visualiseringer og diagrammer i enhver arbejdssituation.
- **Illustrationer i høj kvalitet**: alle billeder og grafikker i den trykte udgave præsenteres i maksimal kvalitet
- **Hurtig adgang til information**: nem navigation, mulighed for at lave noter, bogmærker og arbejde med bogen overalt.

Ved at købe den fulde trykte version af bogen får du et praktisk værktøj til komfortabelt og effektivt arbejde med information: muligheden for hurtigt at bruge visuelle materialer i hverdagens opgaver, hurtigt finde de nødvendige skemaer og tage noter. Desuden støtter dit køb udbredelsen af åben viden.

Bestil en trykt version af bogen på: datadrivenconstruction.io/books



I DEL
FRA LERTAVLER TIL DEN DIGITALE REVOLUTION: HVORDAN INFORMATION HAR UDVIKLET SIG I BYGGERIET

Den første del af bogen undersøger den historiske udvikling af datastyring i byggebranchen, fra primitive optegnelser på fysiske medier til moderne digitale økosystemer. Den analyserer transformationen af informationsstyringsteknologier, fremkomsten af ERP -systemer og indvirkningen af datafragmentering på effektiviteten af forretningsprocesser. Der lægges særlig vægt på informationsdigitaliseringsprocessen og den voksende betydning af objektiv analyse i stedet for subjektive ekspertvurderinger. Den eksponentielle vækst i informationsmængder, som den moderne byggebranche står over for, og de tilhørende udfordringer for virksomhedssystemer undersøges i detaljer. Byggebranchens positionering i forbindelse med den fjerde og femte industrielle revolution udforskes, samt potentialet i at bruge kunstig intelligens og datacentrerede tilgange til at skabe bæredygtige konkurrencefordele.

KAPITEL 1.1.
UDVIKLINGEN I BRUGEN AF DATA I BYGGEBRANCHEN

Fødslen af dataæraen i byggeriet

For ca. 10.000 år siden, i den neolitiske periode, foretog menneskeheden en revolutionerende ændring i sin udvikling og opgav den nomadiske livsstil til fordel for et fastboende liv, hvilket førte til fremkomsten af de første primitive bygninger lavet af ler, træ og sten [6]. Fra dette øjeblik begynder byggeindustriens historie.

Efterhånden som civilisationerne udviklede sig, blev arkitekturen mere og mere kompleks, hvilket førte til de første rituelle templer og offentlige bygninger. Den stigende kompleksitet i de arkitektoniske designs krævede, at antikkens ingeniører og ledere skabte de første optegnelser og beregninger. De første optegnelser på lertavler og papyri indeholdt ofte en beskrivelse af logikken bag beregningen af mængden af nødvendige byggematerialer, deres omkostninger og beregningen af betalingen for det udførte arbejde [7]. Dermed begyndte æraen med brug af data i byggeriet - længe før de moderne digitale teknologier kom til (figur 1.1-1).

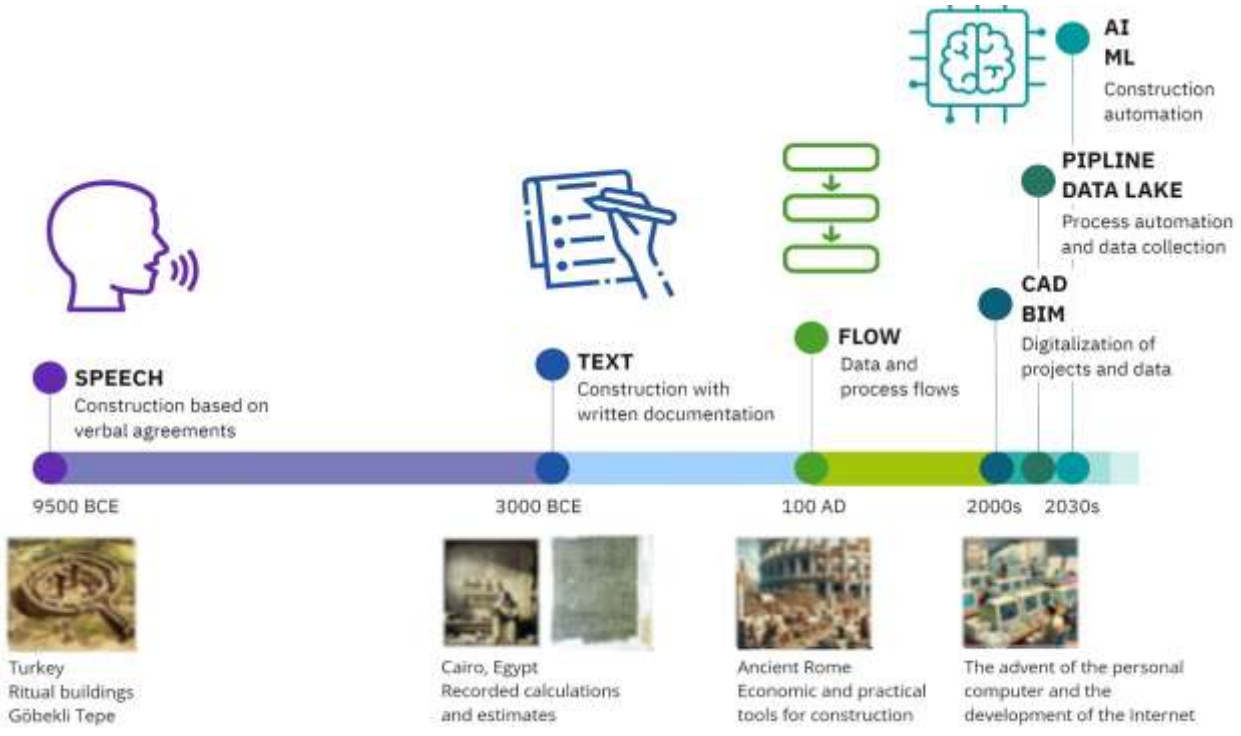


Fig. 1.1-1 Kronologi over informationsteknologiens udvikling i byggeriet: fra mundtlig information til kunstig intelligens.

Fra ler og papyrus til digital teknologi

Den første dokumentation om byggeri stammer fra pyramidebyggeriet omkring 3000-4000 f.Kr.[7]. Siden da har det at føre skriftlige optegnelser lettet og ledsaget udviklingen i byggebranchen og gjort det muligt at akkumulere og systematisere viden, som i løbet af de næste 10.000 år førte til betydelige innovationer inden for byggemetoder og arkitektur.

Brugen af de første fysiske medier i byggeriet, såsom lertavler, papyrus fra tusinder af år siden (figur 1.1-2) eller "A0"-papir i 1980'erne, til at registrere data var oprindeligt ikke beregnet til at anvende disse oplysninger på nye projekter. Hovedformålet med sådanne optegnelser var at beskrive den aktuelle status for projektet, herunder beregninger af de nødvendige materialer og omkostningerne ved arbejdet. På samme måde garanterer tilgængeligheden af digitale designdata og modeller i dag ikke altid, at de kan anvendes i fremtidige projekter, og de tjener ofte primært som information til aktuelle beregninger af nødvendige materialer og byggeomkostninger.

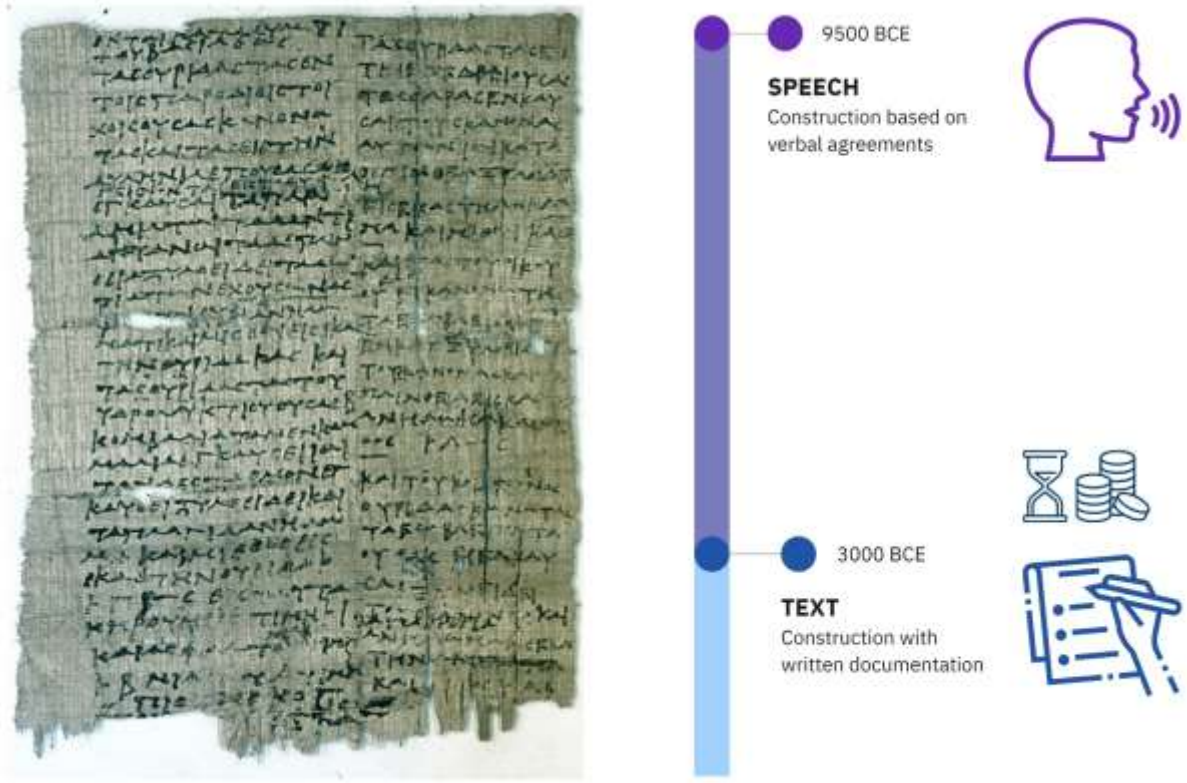


Fig. 1.1-2 En papyrus fra det 3. århundrede f.Kr., der beskriver omkostningerne ved at male forskellige typer vinduer i et kongeligt palads ved hjælp af encaustic-teknikken.

Det tog menneskeheden ca. 5.000 år at gå fra mundtlige samtaler til skriftlige dokumenter i byggeprojektledelse, og lige så lang tid at gå fra papir til digitale data som den primære ressource for planlægning og kontrol.

Ligesom udviklingen af handel og monetære relationer stimulerede fremkomsten af skrift og de første advokater til at løse tvister, så førte de første registreringer af materialeomkostninger og arbejdsomfang i byggeriet til fremkomsten af de første ledere i byggebranchen, hvis opgaver omfattede dokumentation, overvågning og ansvar for vigtige oplysninger om projektets tidslinjer og omkostninger.

I dag spiller data en meget vigtigere rolle: De registrerer ikke kun de beslutninger, der er truffet, men bliver også et værktøj til at forudsige og modellere fremtiden. Det er det fundament, som den moderne procestilgang i projektledelse bygger på - at omdanne akkumuleret erfaring til et beslutningssystem baseret på strukturerede og verificerbare data.

Proces som værktøj til datadrevet oplevelse

Kernen i enhver proces er omdannelsen af tidligere erfaringer til et værktøj til planlægning af fremtiden. Erfaring i moderne forstand er et struktureret sæt data, der kan analyseres for at komme med kvalificerede forudsigelser.

> Det er historiske data, der er grundlaget for prognoser, da de tydeligt viser resultaterne af det udførte arbejde og giver indsigt i de faktorer, der påvirker disse resultater.

Lad os tage et konkret eksempel fra monolitisk byggeri: Når man planlægger tidspunktet for arbejdet, tager man normalt højde for betonmængden, konstruktionens kompleksitet og vejrforholdene. Lad os antage, at en bestemt arbejdsleder eller virksomhedens historiske data for de sidste tre år (2023-2025) viser, at det tog mellem 4,5 og 6 dage at støbe en 200 m² stor monolitisk struktur i regnvejr (fig. 1.1-3). Det er disse akkumulerede statistikker, der bliver grundlaget for at forudsige gennemløbstider og beregne ressourcer, når man planlægger lignende arbejde i fremtidige projekter. Baseret på disse historiske data kan formanden eller beregneren lave en informeret prognose, baseret på erfaring, af den tid, der kræves for at færdiggøre fremtidige lignende arbejder i 2026 under lignende forhold.

I dette tilfælde med tidsestimater fungerer den analytiske proces som en mekanisme til at omdanne forskelligartede data til strukturerede oplevelser og derefter til et præcist planlægningsværktøj. Data og processer er et enkelt økosystem, hvor den ene ikke kan eksistere uden den anden.

> *Tæl det, der kan tælles, mål det, der kan måles, og gør det, der ikke kan måles, målbart [8].*
>
> - Galileo Galilei

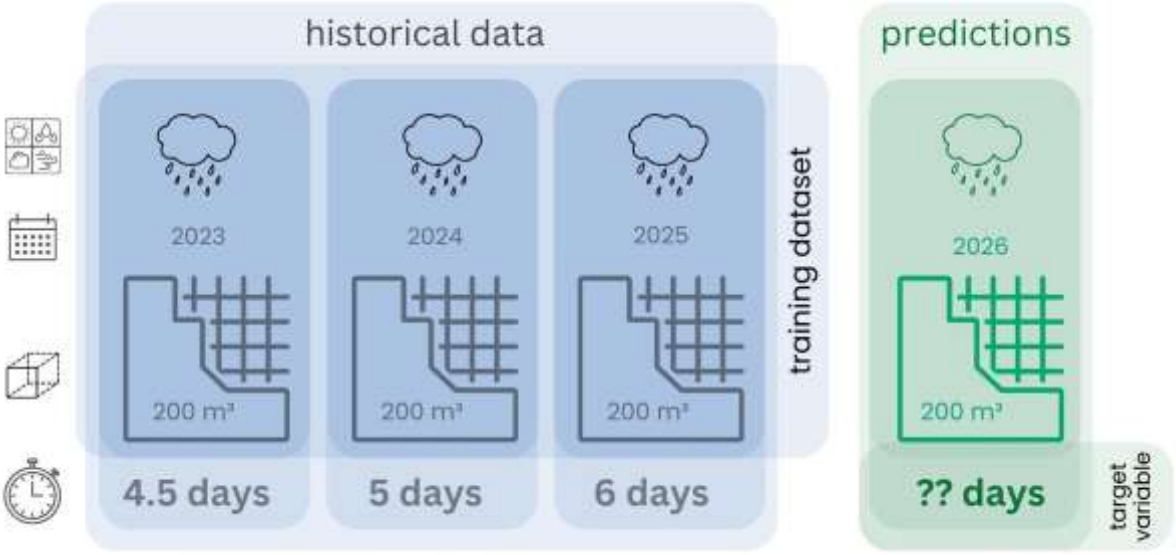


Fig. 1.1-3 Historiske data fungerer som et træningsdatasæt til forudsigelse af en af værdierne i fremtiden.

I dagens forretningslandskab er dataanalyse ved at blive en kritisk komponent i effektiv projektstyring, procesoptimering og strategisk beslutningstagning. Byggebranchen er gradvist ved at mestre fire vigtige niveauer af analyse, som hver især besvarer et specifikt spørgsmål og giver unikke fordele (fig. 1.1-4):

- **Deskriptiv analyse** - besvarer spørgsmålet "*hvad skete der*?" og leverer historiske data og rapporter om tidligere begivenheder og resultater: I løbet af de sidste tre år (2023-2025) tog det mellem 4,5 og 6 dage at støbe en 200 m² stor monolitisk struktur i regnvejr.
- **Diagnostisk analyse** - besvarer spørgsmålet "*hvorfor skete dette?" ved at* identificere årsagerne til problemerne: analysen viser, at hældningstiden for den monolitiske struktur steg på grund af regnvejr, som bremsede betonens hærdningsproces.
- **Predictive analytics** - fremadrettet, forudsiger mulige risici og leveringstider ved at besvare spørgsmålet "*hvad vil der ske*?": Baseret på historiske data forudsiges det, at det vil tage ca. 5,5 dage at støbe en lignende 200 m² monolitisk struktur i regnvejr i 2026 under hensyntagen til alle kendte faktorer og tendenser.
- **Prescriptive analytics** - giver automatiserede anbefalinger og besvarer spørgsmålet "*hvad skal jeg gøre*?", så virksomhederne kan vælge de optimale handlinger: For at optimere arbejdet anbefales det f.eks. at: bruge særlige tilsætningsstoffer til at fremskynde betonens hærdning under forhold med høj luftfugtighed; planlægge støbning i perioder med den laveste sandsynlighed for nedbør; arrangere midlertidige læskure til konstruktionen, hvilket vil reducere arbejdstiden til 4-4,5 dage, selv under ugunstige vejrforhold.

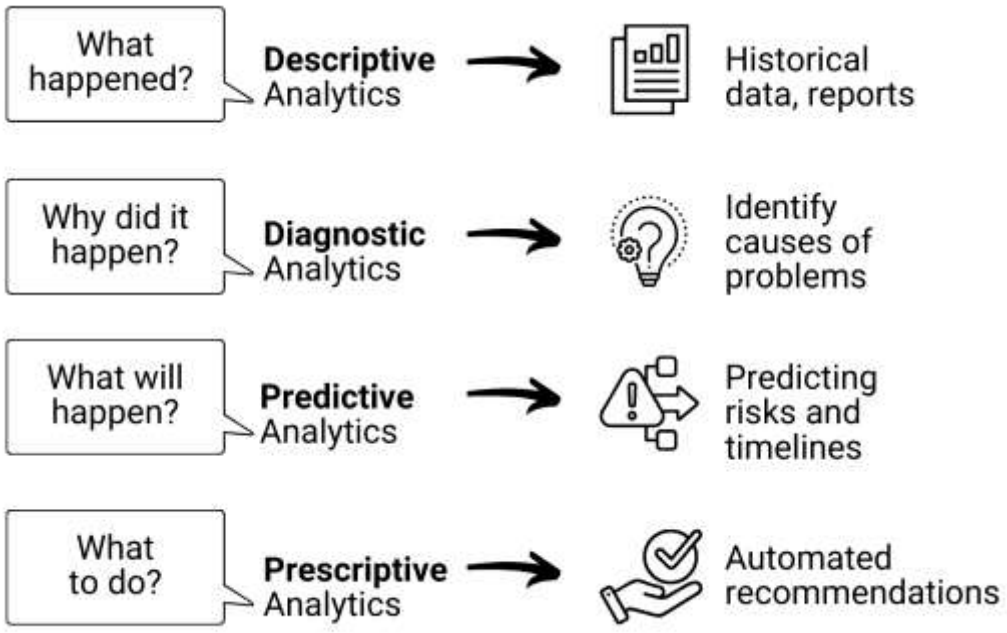


Fig. 1.1-4 De vigtigste typer af analyser: fra tidligere beskrivelse til automatiseret beslutningstagning.

En fuldgyldig digital transformation, som indebærer en overgang til systemanalyse og datadrevet ledelse, kræver ikke kun outsourcing, men også dannelsen af et kompetent internt team. Nøglemedlemmerne i et sådant team bør være produktchefer, dataingeniører, analytikere og udviklere, som skal arbejde tæt sammen med forretningsenhederne (fig. 4.3-9). Dette samarbejde er nødvendigt for at stille intelligente analytiske spørgsmål og effektivt parametrisere forretningens beslutningsopgaver. I et informationssamfund bliver data ikke bare et hjælpeværktøj, men grundlaget for prognoser og optimering.

I byggeriet ændrer den digitale transformation fundamentalt den måde, hvorpå faciliteter designes, administreres og drives. Denne proces kaldes digitalisering af information - hvor alle aspekter af byggeprocessen digitaliseres til en digital form, der er egnet til analyse.

Digitalisering af information om byggeprocessen

I årtusinder har mængden af information, der er registreret i byggeriet, næsten ikke ændret sig, men den er vokset hurtigt i de seneste årtier (figur 1.1-5).

> Ifølge PwC-undersøgelsen® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] er 90% af alle data i verden blevet skabt inden for de seneste to år (fra 2015). Men de fleste virksomheder udnytter ikke disse data fuldt ud, da de enten forbliver i silosystemer eller blot arkiveres uden egentlig analyse.

Stigningen i datamængden er kun accelereret i de seneste år og er fordoblet fra 15 zettabytes i 2015 til 181 zettabytes i 2025 [10]. Hver dag fyldes serverne hos bygge- og designvirksomheder med projektdokumen-

tation, arbejdsplaner, beregninger og kalkulationer, finansielle rapporter. Til 2D/3D -tegninger bruges formaterne DWG, DXF og DGN, og til 3D-modeller - RVT, NWC, PLN og IFC™. Tekstdokumenter, tabeller og præsentationer gemmes i DOC, XLSX og PPT. Ud over video og billeder fra byggepladsen - i MPG og JPEG, realtidsdata fra IoT-komponenter, RFID® tags (identifikation og sporing) og BMS bygningsstyringssystemer (overvågning og kontrol)



Fig. 1.1-5 Parabolsk datavækst 2010-2025 (baseret på [10]).

Med den hurtige vækst i information står byggebranchen over for behovet for ikke kun at indsamle og lagre data, men også at sikre deres verificering, validering, målbarhed og analytiske behandling. I dag gennemgår branchen en aktiv fase af informationsdigitalisering - den systematiske omdannelse af alle aspekter af byggeaktiviteter til en digital form, der er egnet til analyse, fortolkning og automatisering.

> Digitalisering af information betyder, at man tager information om alle enheder og elementer i et byggeprojekt og selve byggeprocessen - også dem, som vi tidligere slet ikke betragtede som information - og konverterer dem til et dataformat, der gør informationen kvantificerbar og let at analysere.

I forbindelse med byggeri betyder det at indsamle og digitalisere oplysninger om alle elementer i projekter og alle processer - fra maskiners og menneskers bevægelser på byggepladsen til vejr- og klimaforhold på byggepladsen, aktuelle materialepriser og centralbankrenter - for at kunne generere analytiske modeller.

> *Hvis du kan måle det, du taler om, og udtrykke det i tal, så ved du noget om emnet. Men hvis du ikke kan udtrykke det kvantitativt, er din viden ekstremt begrænset og utilfredsstillende. Det kan være et udgangspunkt, men det er ikke niveauet for ægte videnskabelig viden. [11].*
>
> - W. Thomson (Lord Kelvin), 1824-1907, britisk videnskabsmand

Digitaliseringen af information går langt ud over den traditionelle tilgang til informationsindsamling, hvor man kun registrerede grundlæggende målinger som f.eks. arbejdstimer eller faktiske materialeomkostninger. I dag kan stort set enhver begivenhed omdannes til en datastrøm, der egner sig til dybdegående analyse ved hjælp af avancerede analyseværktøjer og maskinlæringsteknikker. Byggebranchen har gennemgået et grundlæggende skift fra papirtegninger, Excel-regneark og mundtlige instruktioner til digitale systemer (figur 1.2-4), hvor hvert element i et projekt bliver en datakilde. Selv medarbejderne - fra ingeniører til bygningsarbejdere på byggepladsen - betragtes nu som en samling af digitale variabler og datasæt.

> Ifølge KPMG's "Familiar Challenges - New Approaches: Global Construction Survey 2023" er digitale tvillinger, AI (kunstig intelligens) og Big Data ved at blive de vigtigste drivkræfter for at forbedre projekternes rentabilitet [2].

Moderne teknologier forenkler ikke kun informationsindsamlingen og gør den stort set automatisk, men reducerer også omkostningerne til datalagring radikalt. Som følge heraf bevæger virksomhederne sig væk fra en selektiv tilgang og foretrækker at gemme hele spektret af information til senere analyse (fig. 2.1-5), hvilket åbner op for potentielle muligheder for at optimere processer i fremtiden.

> Digitalisering af information og digitalisering gør det muligt at afdække den skjulte, tidligere uudnyttede værdi af information. Hvis data organiseres korrekt, kan de genbruges, genfortolkes og integreres i nye tjenester og løsninger.

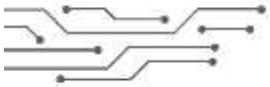
I fremtiden vil digitaliseringen af information sandsynligvis føre til fuld automatisering af dokumenthåndtering, indførelse af selvstyrende byggeprocesser og fremkomsten af nye professioner - byggedataanalytikere, AI-projektledelseseksperter og digitale ingeniører. Byggeprojekter vil blive dynamiske informationskilder, og beslutningstagning vil ikke være baseret på intuition eller subjektiv erfaring, men på pålidelige og reproducerbare digitale fakta

> *Information er det 21. århundredes olie, og analytics er den interne forbrændingsmotor [12].*
>
> - Peter Søndergaard, Senior Vice President, Gartner®

> Ifølge IoT Analytics 2024 [13] forventes de globale udgifter til datahåndtering og -analyse at vokse dramatisk fra 185,5 milliarder dollars i 2023 til 513,3 milliarder dollars i 2030 med en samlet årlig vækstrate på 16%. Det er dog ikke alle komponenter, der vokser i samme takt: Analytics vokser hurtigt, mens væksten i storage er aftagende. Analyse vil give den hurtigste vækst i økosystemet for datastyring: Det forventes at vokse fra 60,6 mia. dollars i 2023 til 227,9 mia. dollars i 2030, en samlet årlig vækstrate på 27%.

Med den accelererende digitalisering af information og den hurtige vækst i informationsmængderne står ledelsen af byggeprojekter og -virksomheder over for behovet for systematisk at lagre, analysere og behandle forskellige, ofte heterogene data. Som svar på denne udfordring påbegyndte branchen fra midten af 1990'erne et massivt skift til elektronisk oprettelse, lagring og styring af dokumentation - fra regneark og designberegninger til tegninger og kontrakter.

Traditionelle papirdokumenter, som kræver underskrifter, fysisk opbevaring, regelmæssig revision og arkivering i skabe, bliver gradvist erstattet af digitale systemer, som lagrer data på en struktureret måde - i databaser med specialiserede applikationer.



KAPITEL 1.2.
TEKNOLOGIER OG LEDELSESSYSTEMER I MODERNE BYGGERI

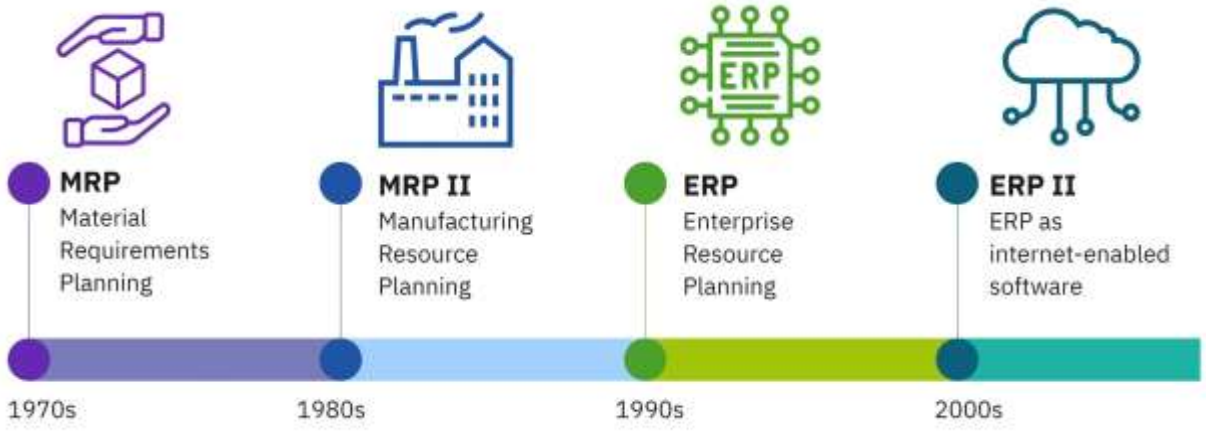
Den digitale revolution og fremkomsten af modulære MRP/ERP -systemer

Æraen med moderne digital datalagring og -behandling begyndte med fremkomsten af magnetbånd i 1950'erne, som åbnede op for muligheden for at lagre og bruge store mængder information. Det næste gennembrud var fremkomsten af diskdrev, som radikalt ændrede tilgangen til datahåndtering i byggebranchen.

Med udviklingen af data warehousing er et stort antal virksomheder kommet ind på markedet for løsninger og er begyndt at udvikle modulær software til at skabe, lagre, behandle data og automatisere rutineopgaver

> Den eksponentielle vækst i information og værktøjer har ført til et behov for integrerede, modulære løsninger, der ikke arbejder med individuelle filer, men hjælper med at styre og kontrollere datastrømmen på tværs af processer og projekter.

De første omfattende platformsværktøjer skulle ikke kun lagre dokumenter, men også dokumentere alle ændringsanmodninger og operationer i processer: hvem der tog initiativ til dem, hvad omfanget af anmodningen var, og hvad der til sidst blev registreret som en værdi eller attribut. Til disse formål var der brug for et system, der kunne spore nøjagtige beregninger og beslutninger (fig. 1.2-1). Sådanne platforme var de første MRP (Material Requirements Planning) og ERP (Enterprise Resource Planning) systemer, der blev populære fra begyndelsen af 1990'erne [14]



Figur 1.2-1 Fremskridt inden for datalagringsteknologi førte til fremkomsten af ERP -systemer i 1980'erne.

De første MRP - og ERP - systemer lagde grunden til digitaliseringsæraen inden for styring af forretningsprocesser og byggeprojekter. Modulære systemer, der oprindeligt var designet til at automatisere vigtige forretningsprocesser, blev efterhånden integreret med yderligere, mere fleksible og tilpasningsdygtige softwareløsninger.

Disse ekstra løsninger var designet til databehandling og styring af projektindhold (fig. 1.2-2), og de erstattede enten visse moduler i store systemer eller supplerede dem effektivt ved at udvide funktionaliteten i hele systemet.

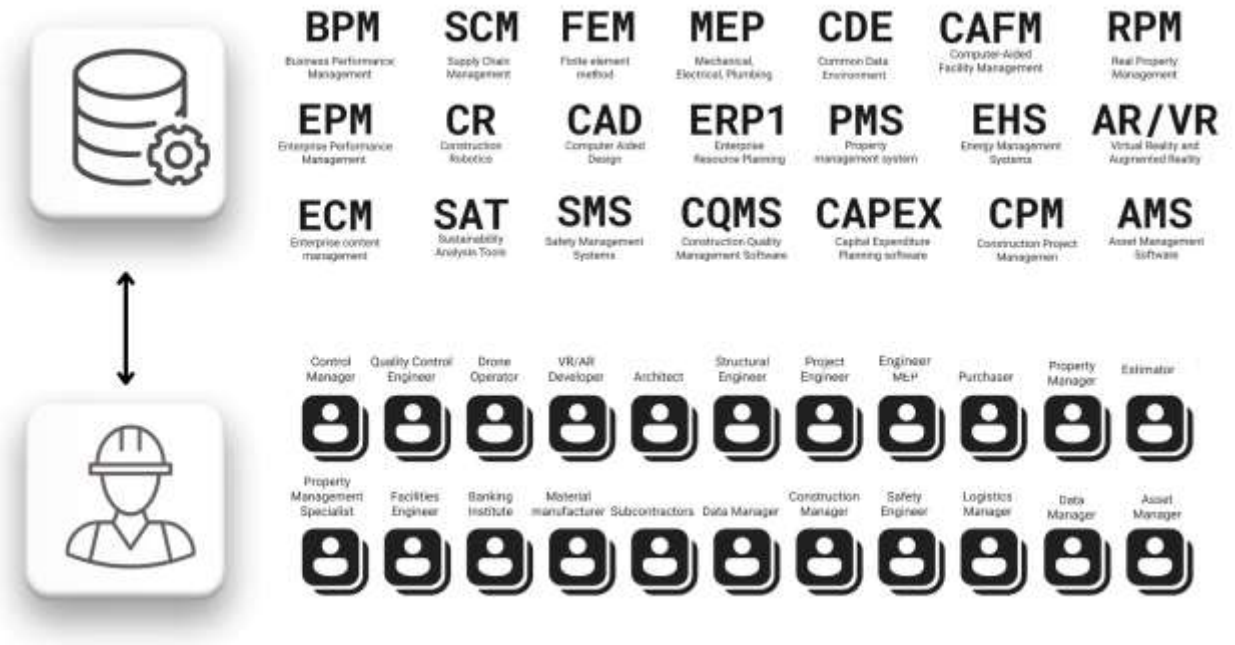


Fig. 1.2-2 Nye softwareløsninger har trukket en hær af ledere ind i virksomheden for at styre datastrømmene.

I løbet af de sidste årtier har virksomheder investeret kraftigt i modulære systemer [15], idet de opfatter dem som langsigtede integrerede løsninger.

> Ifølge Software Path-rapporten for 2022 [16] er det gennemsnitlige budget pr. bruger af et ERP - system 9.000 dollars. I gennemsnit bruger ca. 26% af virksomhedens medarbejdere sådanne systemer. For en organisation med 100 brugere når de samlede omkostninger til ERP-implementering således op på ca. 900.000 dollars.

Investeringer i proprietære, lukkede, modulære løsninger bliver mindre og mindre berettigede på baggrund af den hurtige udvikling af moderne, fleksible og åbne teknologier. Hvis sådanne investeringer allerede er foretaget, er det vigtigt objektivt at revurdere de eksisterende systemers rolle: om de fortsat er vigtige på lang sigt, eller om deres funktioner kan revideres og implementeres mere effektivt og gennemsigtigt.

Et af hovedproblemerne med nutidens modulære dataplatforme er, at de centraliserer datahåndteringen i lukkede applikationer. Resultatet er, at data - en virksomheds kerneaktiv - bliver afhængig af specifikke softwareløsninger i stedet for omvendt. Det begrænser genbrug af information, komplicerer migration og reducerer virksomhedens smidighed i et hurtigt skiftende digitalt landskab.

Hvis det er sandsynligt, at relevansen eller betydningen af lukket modulær arkitektur vil aftage i fremtiden, giver det mening at anerkende de omkostninger, der er afholdt i dag, som sunk costs og fokusere på et strategisk skift til et mere åbent, skalerbart og adaptivt digitalt økosystem.

> Proprietær software er kendetegnet ved, at udvikleren har eksklusiv kontrol over kildekoden og de brugerdata, der skabes som led i brugen af sådanne løsninger. I modsætning til open source-software har brugerne ikke adgang til applikationens interne struktur og kan ikke selv gennemgå, ændre eller tilpasse den til deres behov. I stedet skal de købe licenser, der giver ret til at bruge softwaren inden for de grænser, som leverandøren har fastsat.

En moderne datacentreret tilgang tilbyder et andet paradigme: Data skal ses som et vigtigt strategisk aktiv - uafhængigt, holdbart og adskilt fra specifikke softwareløsninger. Applikationer bliver til gengæld blot dataværktøjer, som frit kan udskiftes uden risiko for at miste kritisk information.

Udviklingen af ERP- og MRP-systemer i 1990'erne (figur 1.2-1) gav virksomhederne stærke værktøjer til processtyring, men havde også den utilsigtede konsekvens, at antallet af mennesker, der var involveret i at vedligeholde informationsstrømme, steg betydeligt. I stedet for at automatisere og forenkle driftsopgaverne skabte disse systemer ofte nye niveauer af kompleksitet, bureaukrati og afhængighed af interne IT-ressourcer.

Datastyringssystemer: fra datamining til forretningsudfordringer

Nutidens virksomheder står over for behovet for at integrere flere datastyringssystemer. At vælge datastyringssystemer, styre disse systemer med omtanke og integrere forskellige datakilder er ved at blive afgørende for virksomhedens resultater.

I midten af 2020'erne kan du finde hundredvis (og tusindvis i store byggefirmaer) af forskellige systemer (figur 1.2-3), som skal fungere i harmoni for at få alle aspekter af byggeprocessen til at køre gnidningsløst og sammenhængende.

> Ifølge Deloittes undersøgelse fra 2016® "Data-Driven Management in Digital Capital Projects" - bruger den gennemsnitlige byggefagmand 3,3 softwareprogrammer dagligt, men kun 1,7 af dem er integreret med hinanden [17].

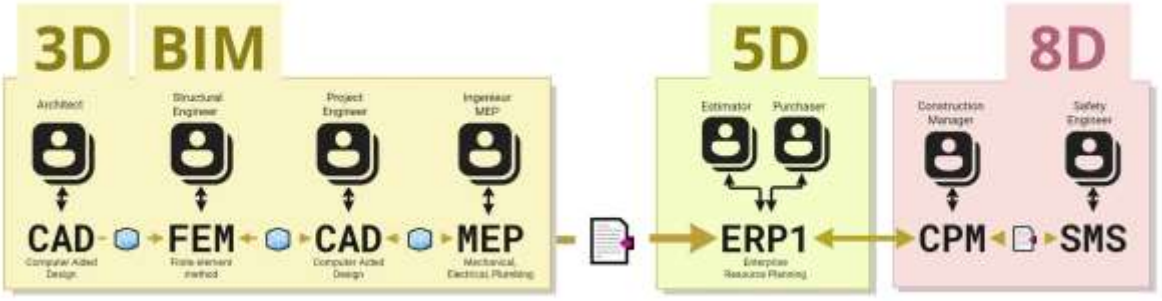


Fig. 1.2-3 Ethvert forretningssystem kræver et professionelt team og en ansvarlig leder til kvalitetsstyring af data.

Her er en liste over populære systemer til mellemstore og store virksomheder i byggebranchen, som bruges til effektiv styring af byggeprojekter:

- **ERP (Enterprise Resource Planning)** - giver integration af forretningsprocesser, herunder regnskab, indkøb og projektstyring.
- **CAPEX (Capital Expenditure Planning Software)** - bruges til budgettering og styring af finansielle investeringer i byggeprojekter, hjælper med at bestemme omkostningerne ved anlægsaktiver og investeringer i langsigtede aktiver.
- **CAD (Computer-Aided Design)** og **BIM (Building Information Modeling)** - bruges til at skabe detaljerede og præcise tekniske tegninger og 3D -modeller af projekter. Fokus i disse systemer er på at arbejde med geometrisk information.
- **MEP (Mechanical, Electrical, Plumbing**) - Tekniske systemer, der omfatter mekaniske, elektriske og sanitære komponenter og beskriver et projekts interne "kredsløbssystem".
- **GIS (Geografiske Informationssystemer)** - bruges til terrænanalyse og planlægning, herunder kartografi og rumlig analyse.
- **CQMS (software til kvalitetsstyring i byggeriet**) - sikrer, at byggeprocesser overholder etablerede standarder og regler, og hjælper med at eliminere fejl.
- **CPM (Construction Project Management)** - omfatter planlægning, koordinering og styring af byggeprocesser.
- **CAFM (Computer-Aided Facility Management) -** bygningsstyrings- og vedligeholdelsessystemer.
- **SCM (Supply Chain Management) er** nødvendig for at optimere flowet af materialer og information mellem leverandører og byggepladsen.
- **EPM (Enterprise Performance Management) - har til** formål at forbedre forretningsprocesser og performance.
- **AMS (Asset Management Software)** - bruges til at optimere brugen, styringen og vedligeholdelsen af udstyr og infrastruktur i hele aktivets livscyklus.
- **RPM (Real Property Management)** - omfatter opgaver og processer i forbindelse med forvaltning og drift af bygninger og jord samt tilhørende ressourcer og aktiver.

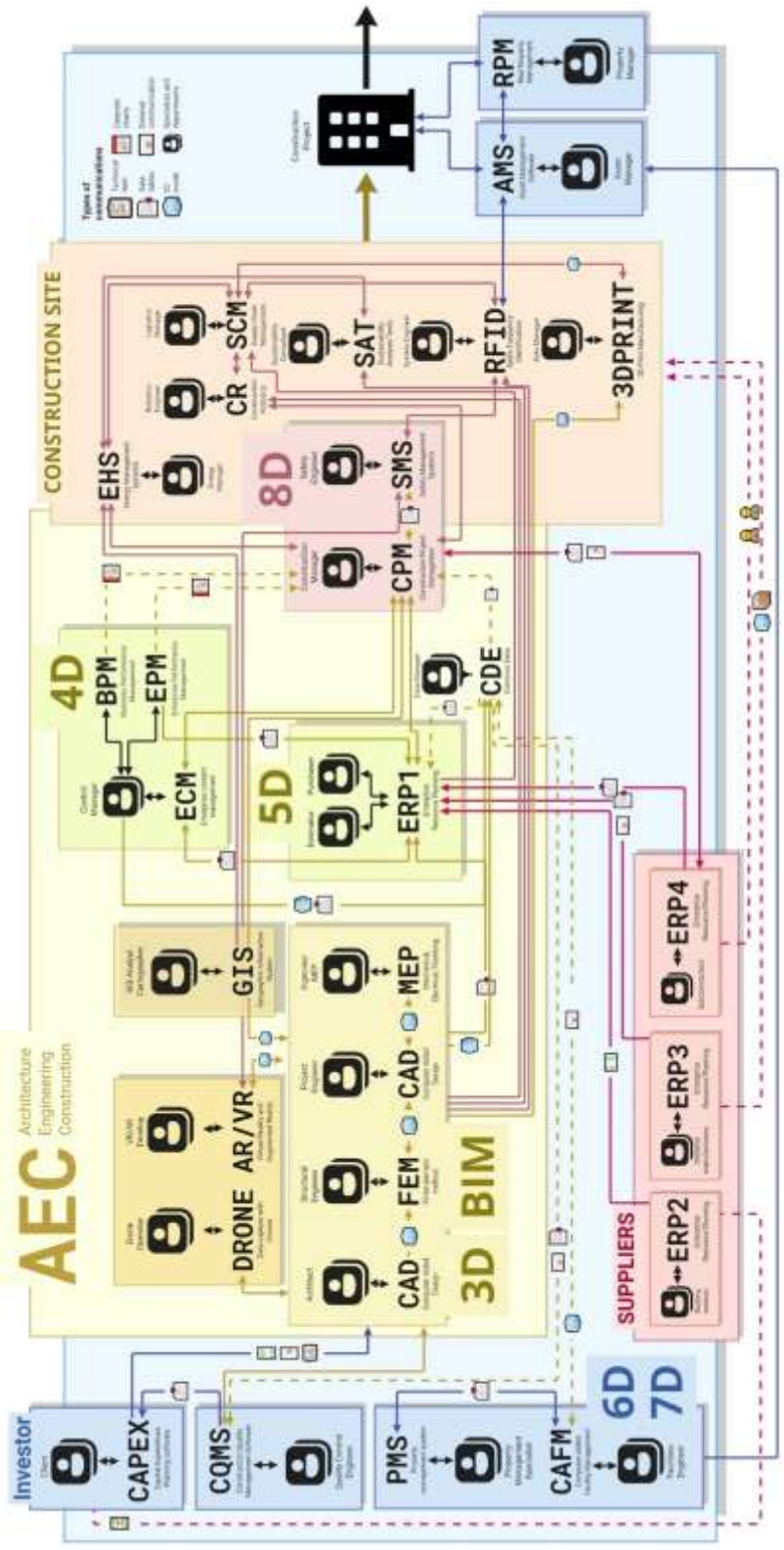


Fig. 1.2-4 Sammenkobling af systemer, der forbinder virksomhedens processer med informationsstrømmen mellem forskellige afdelinger.

- **CAE (Computer-Aided** Engineering) - Computerstøttet ingeniørarbejde omfatter beregnings- og simuleringssystemer som finite element-analyse (FEA) og computational fluid dynamics (CFD).
- **CFD (**Computational **Fluid** Dynamics**)** - Computational fluid dynamics, modellering af væske- og gasstrømme. CAE-underkategori.
- **CAPP (Computer-Aided Process Planning)** - Computerstøttet procesplanlægning. Det bruges til at skabe rute- og proceskort.
- **CAM (Computer-Aided Manufacturing)** - computerstøttet fremstilling, generering af kontrolprogrammer til CNC-maskiner.
- **PDM (Product Data Management)** - Product Data Management, et system til lagring og styring af teknisk dokumentation.
- **MES (Manufacturing Execution System)** er et system til styring af produktionsprocesser i realtid.
- **PLM (Product Lifecycle Management)** - livscyklusstyring af et projektelement, integrerer PDM, CAPP, CAM og andre systemer til komplet produktkontrol fra udvikling til bortskaffelse.

Disse og mange andre systemer, der indeholder en række forskellige softwareløsninger, er blevet en integreret del af den moderne byggeindustri (fig. 1.2-4). Kernen i sådanne systemer er specialiserede databaser med intuitive grænseflader, der giver effektiv indtastning, behandling og analyse af information i alle faser af design og konstruktion. Integrationen af digitale værktøjer hjælper ikke kun med at optimere arbejdsprocesserne, men forbedrer også beslutningernes nøjagtighed betydeligt, hvilket har en positiv indvirkning på timingen og kvaliteten af projektimplementeringen.

Men i halvdelen af tilfældene er der ingen integration. Ifølge statistikkerne er kun hver anden applikation eller hvert andet system integreret med andre løsninger [17]. Dette indikerer den fortsatte fragmentering af det digitale miljø og understreger behovet for at udvikle åbne standarder og ensartede grænseflader for at sikre end-to-end informationsudveksling inden for et byggeprojekt.

En af de største udfordringer i integrationen for moderne virksomheder er fortsat de digitale systemers høje kompleksitet og kravene til brugerkompetencer, der er nødvendige for effektiv informationssøgning og -fortolkning. Et team af specialister, ledet af en nøgleperson, dannes for at understøtte hvert system, der implementeres i virksomheden (fig. 1.2-2).

> Den centrale systemadministrator spiller en afgørende rolle i den rigtige retning for datastrømmen og er ansvarlig for kvaliteten af den endelige information, ligesom de første administratorer for tusinder af år siden var ansvarlige for de tal, der blev skrevet på papyrus eller lertavler.

For at gøre forskellige informationsstrømme til et ledelsesværktøj er evnen til systematisk at integrere og styre data afgørende. I denne arkitektur skal ledere fungere som elementer i et enkelt netværk - som et mycelium, der forbinder de enkelte dele af virksomheden til en sammenhængende levende organisme, der er i stand til at tilpasse sig og udvikle sig.

Virksomhedens mycelium: hvordan data forbindes med forretningsprocesser

Processen med at integrere data i applikationer og databaser er baseret på aggregering af information fra en række forskellige kilder, herunder forskellige afdelinger og specialister (figur 1.2-4). Specialister søger efter relevante data, behandler dem og overfører dem til deres systemer og applikationer til videre brug.

Hvert virksomhedssystem, der består af et sæt værktøjer, teknologier og databaser, er et vidensträe med

rødder i jorden af historiske data, der vokser og bærer nye frugter i form af færdige løsninger: dokumenter, beregninger, tabeller, grafer og dashboards (Fig. 1.2-5). Systemerne i en virksomhed interagerer og kommunikerer med hinanden som træer i et bestemt skovområde og udgør et komplekst og velstruktureret system, der understøttes og styres af ekspertledere.

> En virksomheds informationssøgnings- og overførselssystem fungerer som et komplekst skovnetværk bestående af træer (systemer) og mycelium-svampe (ledere), der fungerer som ledere og genbrugere og sikrer, at information overføres og flyder til de rigtige systemer. Det er med til at opretholde et sundt og effektivt flow og distribution af data i virksomheden.

Eksperter absorberer ligesom rødder rådata i de indledende faser af et projekt og omdanner dem til næringsstoffer for virksomhedens økosystem. Data- og indholdsstyringssystemer (fig. 1.2-4 - ERP, CPM, BIM, osv.) fungerer som stærke informationsmotorveje, hvorigennem denne viden cirkulerer gennem alle niveauer i virksomheden.

Ligesom i naturen, hvor hvert element i økosystemet spiller sin egen rolle, bidrager hver procesdeltager i en virksomheds forretningslandskab - fra ingeniør til analytiker - til informationsmiljøets vækst og frugtbarhed. Disse systemiske "datatræer" (fig. 1.2-5) er ikke bare mekanismer til at indsamle information, men en konkurrencefordel, der sikrer en virksomheds bæredygtighed.

Skovøkosystemer er en overraskende præcis afspejling af, hvordan digitale virksomhedsstrukturer er organiseret. Ligesom den lagdelte struktur i en skov - fra underskoven til trætoppene - tildeler virksomhedsledelse opgaver til ansvarsniveauer og funktionelle afdelinger.

> Dybe og forgrenede trærødder giver modstandskraft og adgang til næringsstoffer. På samme måde understøtter en solid organisationsstruktur og stabile processer til at arbejde med kvalitetsdata hele informationsøkosystemet i en virksomhed og bidrager til dens bæredygtige vækst og udvikling, selv i perioder med ustabilitet og kriser på markedet (med meget vind).

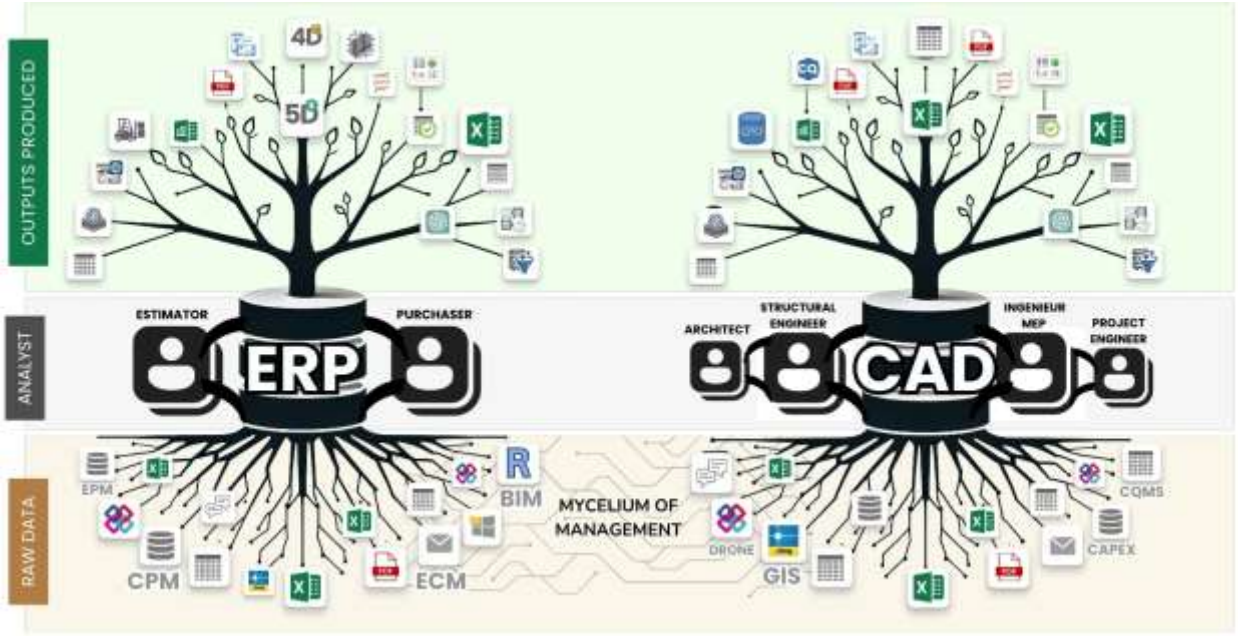


Fig. 1.2-5 Integration af data gennem forskellige systemer er som et mycelium, der forbinder ledere og specialister i et enkelt informationsnetværk.

Den moderne forståelse af skala i erhvervslivet har udviklet sig. I dag bestemmes en virksomheds værdi ikke kun af dens synlige del - "kronerne" i form af slutdokumenter og rapporter - men også af dybden af "rodsystemet" af kvalitativt indsamlede og systematisk behandlede data. Jo mere information, der kan indsamles og behandles, jo højere bliver forretningsværdien. Virksomheder, der metodisk akkumulerer en "kompost" af allerede behandlede data og er i stand til at uddrage nyttige indsigter fra dem, får strategisk fordel.

> Historisk information er ved at blive en ny form for kapital, der muliggør vækst, procesoptimering og konkurrencemæssige fordele. I en datadrevet verden er det ikke den, der har mest, men den, der ved mest, der vinder.

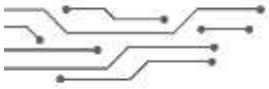
For byggebranchen betyder det, at man går over til projektstyring i realtid, hvor alle processer - fra design og indkøb til koordinering af entreprenører - vil være baseret på relevante, dagligt opdaterede data. Integrationen af information fra forskellige kilder (ERP -systemer, CAD -modeller, sensorer IoT på byggepladser, RFID) vil gøre det muligt at lave mere præcise prognoser, reagere hurtigt på ændringer og undgå forsinkelser, der skyldes mangel på opdaterede data.

> Ifølge McKinsey & Company's Data-Driven Enterprise 2025 (McKinsey & Company®, 2022 [18]) vil fremtidens succesfulde virksomheder være afhængige af data i alle vigtige aspekter af deres drift, fra strategiske beslutninger til operationelle interaktioner.

Data vil ophøre med blot at være et analyseværktøj og vil blive en integreret del af alle forretningsprocesser, der giver gennemsigtighed, kontrol og automatisering af ledelsen. Et datadrevet skridt vil gøre det muligt for organisationer at minimere indvirkningen af den menneskelige faktor, reducere driftsrisici og øge gennemsigtigheden og effektiviteten i beslutningstagningen.

Det 21. århundrede vender op og ned på det økonomiske paradigme: Hvor olie tidligere blev kaldt "sort guld"

for sin evne til at drive maskiner og transport, er historiske data i dag, komprimeret under tidspres, ved at blive en ny strategisk ressource, der ikke driver maskiner, men beslutningsalgoritmer, som vil drive forretningen.



KAPITEL 1.3.
DEN DIGITALE REVOLUTION OG EKSPLOSIONEN AF DATA

Begyndelsen på data-boomet som en evolutionær bølge

Byggebranchen oplever en hidtil uset informationseksplosion. Hvis vi tænker på erhvervslivet som et videnstræ (fig. 1.2-5), der fodres med data, kan det nuværende digitaliseringsstadie sammenlignes med den hurtige vækst i vegetationen i kultiden, en æra, hvor jordens biosfære blev forvandlet af den hurtige ophobning af biomasse (fig. 1.3-1).

Med den globale digitalisering fordobles mængden af information i byggebranchen hvert år. Moderne teknologi gør det muligt at indsamle data i baggrunden, analysere dem i realtid og bruge dem i et omfang, der virkede umuligt for bare kort tid siden.

> Ifølge Moores lov, formuleret af Gordon Moore (medstifter af Intel®), fordobles tætheden og kompleksiteten af integrerede kredsløb og mængden af data, der behandles og lagres, cirka hvert andet år [19].

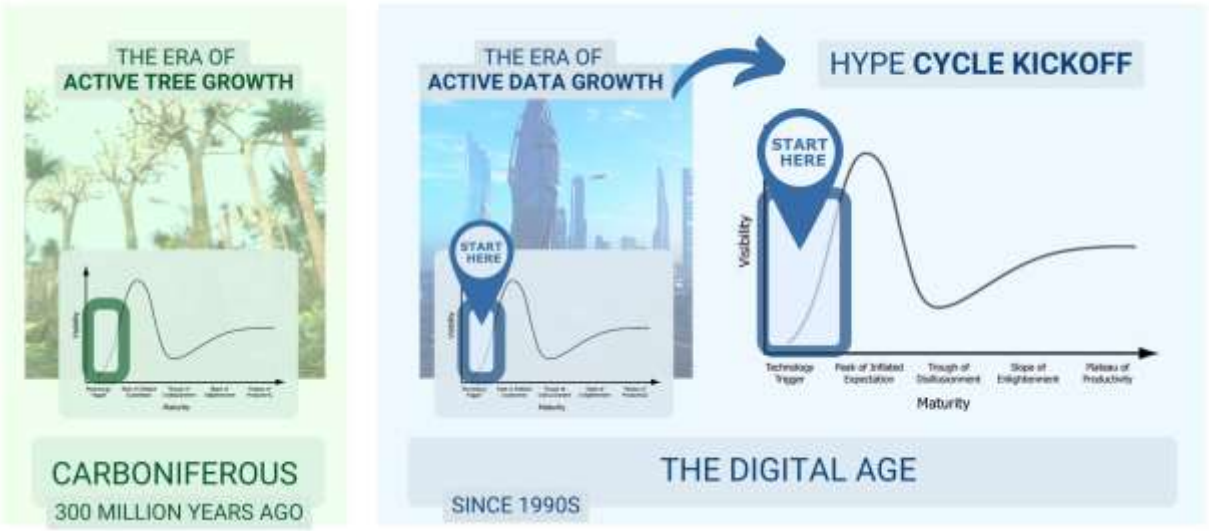


Fig. 1.3-1 Den begyndende digitalisering har ført til en eksponentiel vækst af data, ligesom den kraftige vækst af vegetation i kulalderen.

Mens gamle megalitiske strukturer som Göbekli Tepe (Tyrkiet) ikke efterlod sig dokumenteret viden, der kunne genbruges, gør digitale teknologier det i dag muligt at akkumulere og genbruge information. Dette kan sammenlignes med den evolutionære overgang fra sporeplanter til frøplanter (angiospermer): Fremkomsten af frøet gav anledning til den omfattende spredning af liv på planeten. (Fig. 1.3-2).

På samme måde bliver data fra tidligere projekter en slags "digitale frø" - DNA-videnbærere, der kan skaleres og bruges i nye projekter og pro dukter. Fremkomsten af moderne værktøjer til kunstig intelligens - maskinlæring og store sprogmodeller (LLM'er) som ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN,

Grok - gør det muligt automatisk at udtrække, fortolke og anvende data nye sammenhænge.

> Ligesom frø revolutionerede spredningen af liv på en oprindeligt livløs planet, bliver "datafrø" grundlaget for den automatiske fremkomst af nye informationsstrukturer og viden, så digitale økosystemer kan udvikle sig uafhængigt og tilpasse sig skiftende brugerkrav.

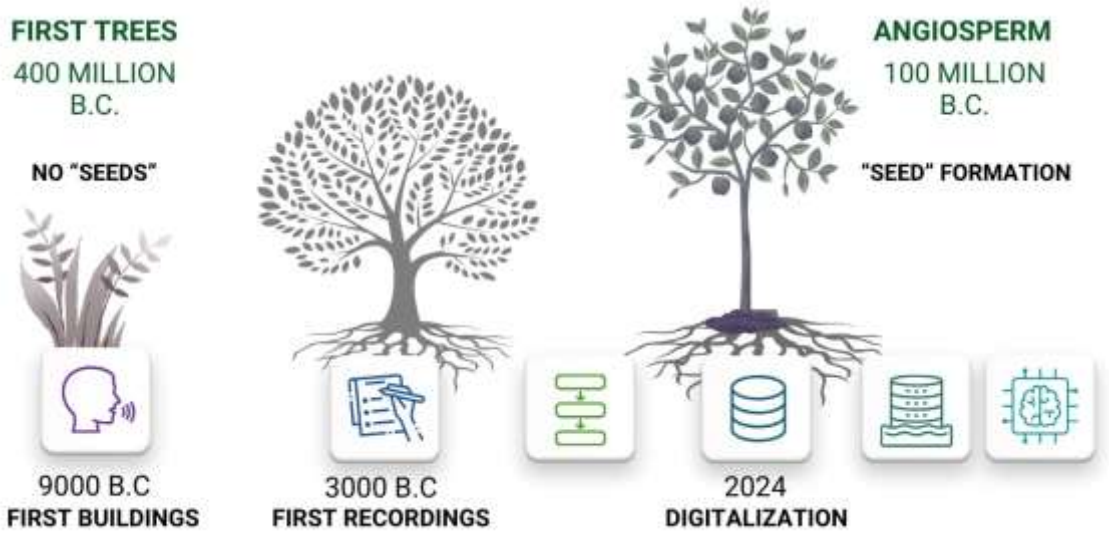


Fig. 1.3-2 Digitale "datafrø" spiller den samme evolutionære rolle som angiospermer, de blomstrende planter, der forvandlede jordens økosystem.

Vi står på tærsklen til en ny æra i byggeriet, hvor eksplosionen af data og den aktive formidling af "data seeds" - struktureret information fra tidligere og igangværende projekter - danner grundlaget for branchens digitale fremtid. Deres "bestøvning" gennem big data-sprogmodeller (LLM'er) giver os mulighed for ikke bare at observere digital forandring, men for aktivt at deltage i skabelsen af selvlærende, adaptive økosystemer. Det er ikke evolution - det er en digital revolution, hvor data bliver den vigtigste byggesten i ny virkelighed.

> *Mængden af data i byggebranchen stiger dramatisk på grund af oplysninger fra forskellige discipliner i hele byggeprojektets livscyklus. Denne enorme ophobning af data har skubbet byggebranchen i retning af Big Data-æraen [20].*
>
> - Prof. Hang Yang, Institut for Byggeri og Arkitektur, Wuhan University of Technology, Wuhan,

Væksten i data i informationsalderen minder om evolutionære processer i naturen: Ligesom udviklingen af skove ændrede det gamle landskab på planeten, ændrer den nuværende informationseksplosion landskabet i hele byggebranchen.

Mængden af data, der genereres i en moderne virksomhed

I løbet af de sidste to år er 90% af alle eksisterende data i verden blevet skabt [21]. Fra 2023 genererer hver person, inklusive fagfolk i byggebranchen, omkring 1,7 megabyte data i sekundet [22], og den samlede mængde data i verden vil nå 64 zettabyte i 2023 og forventes at overstige 180 zettabyte, eller 180*10^15 megabyte, i 2025 [23].

> Denne informationseksplosion har et historisk fortilfælde - Johannes Gutenbergs opfindelse af trykpressen i det 15. århundrede. Bare 50 år efter opfindelsen blev antallet af bøger i Europa fordoblet: I løbet af få årtier blev der trykt lige så mange bøger, som der var blevet skabt i hånden i de foregående 1.200 år [24]. I dag ser vi en endnu hurtigere vækst: Mængden af data i verden fordobles hvert tredje år.

Med den nuværende datavækst har byggebranchen potentiale til at generere lige så meget information i løbet af de næste par årtier, som den har akkumuleret i hele sin tidligere historie

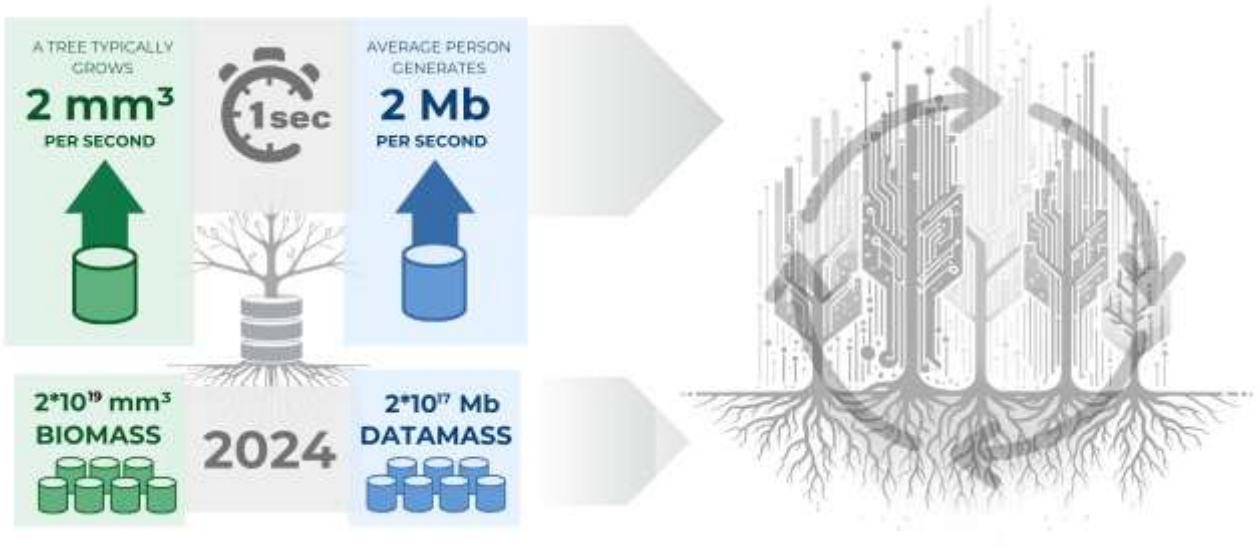
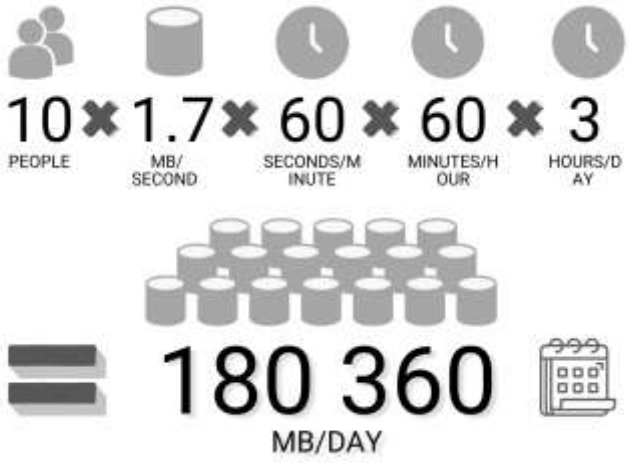


Fig. 1.3-3 Den enkelte medarbejders daglige lagring af data på virksomhedens servere bidrager til den konstante vækst af data.

I dagens byggebranche genererer selv små virksomheder dagligt en enorm mængde information i flere formater, og selv et lille byggefirmas digitale fodaftryk kan nå op på titusindvis af gigabytes om dagen - fra modeller og tegninger til fotografiske optagelser og sensorer på byggepladsen. Hvis vi antager, at hver tekniker i gennemsnit genererer ca. 1,7 MB data pr. sekund, svarer det til ca. 146 GB pr. dag eller 53 TB pr. år (fig. 1.3-3).

Når et team på 10 personer arbejder aktivt i kun 3 timer dagligt, når den samlede mængde information, der genereres pr. dag, op på 180 gigabyte (Fig. 1.3-4).



Figur 1.3-4 En virksomhed med 10 ansatte genererer ca. 50-200 gigabyte data om dagen.

Hvis man antager, at 30% af arbejdsdataene er nye (resten overskrives eller slettes), kan et firma med 10 ansatte skabe flere hundrede gigabyte nye data om måneden (det faktiske tal afhænger af, hvilken type virksomhed der er om).

Så det er klart: Vi genererer ikke bare flere og flere data - vi står over for et voksende behov for effektiv styring, lagring og langsigtet tilgængelighed. Og mens data tidligere kunne "ligge" på lokale servere uden omkostninger, er flere og flere virksomheder i forbindelse med den digitale transformation begyndt at bruge cloud-løsninger som grundlag for deres informationsinfrastruktur.

Omkostningerne ved datalagring: det økonomiske aspekt

I de senere år har flere og flere virksomheder outsourcet datalagring til cloud-tjenester. Hvis en virksomhed f.eks. lægger halvdelen af sine data i skyen til en gennemsnitspris på 0,015 dollar pr. gigabyte pr. måned, kan lageromkostningerne stige med 10-50 dollar [25] hver måned.

For en lille virksomhed med typiske datagenereringsmønstre kan omkostningerne til lagring i skyen variere fra hundreder til potentielt over tusind dollars om måneden (figur 1.3-5) på et par år, hvilket skaber en potentielt betydelig økonomisk byrde.

> Ifølge Forresters undersøgelse "Enterprises Outsource Data Storage as Complexity Grows" [26], som omfattede 214 beslutningstagere inden for teknologiinfrastruktur. [26], som omfattede 214 beslutningstagere inden for teknologisk infrastruktur, outsourcer mere end en tredjedel af organisationerne lagring for at klare den voksende mængde og kompleksitet i datadriften, og næsten to tredjedele af virksomhederne foretrækker en abonnementsbaseret model.

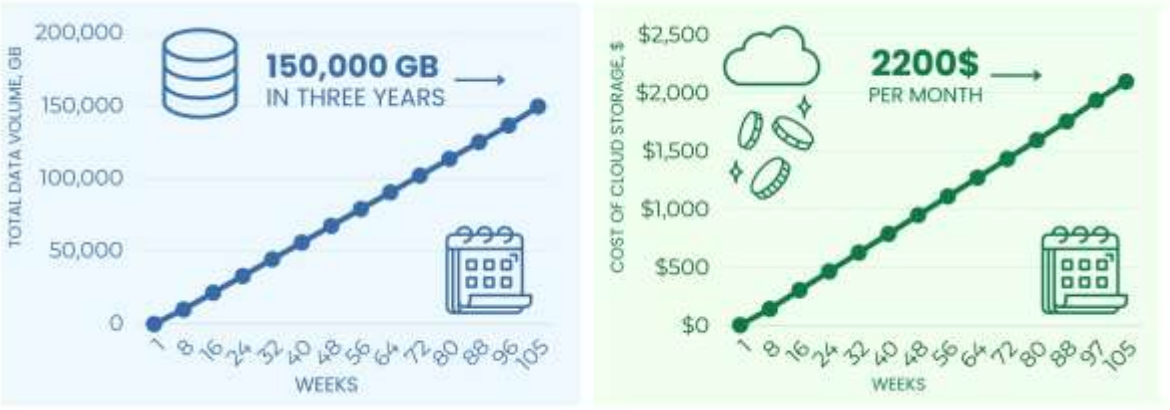


Fig. 1.3-5 Flytning af data til skyen kan øge de månedlige lageromkostninger med op til 2.000 dollars, selv for en virksomhed med kun 10 ansatte.

Situationen kompliceres yderligere af den accelererende anvendelse af cloud-baserede teknologier som CAD (BIM), CAFM, PMIS og ERP - systemer, der yderligere øger omkostningerne til datalagring og -behandling. Som følge heraf er virksomhederne tvunget til at finde måder at optimere omkostningerne på og reducere deres afhængighed af cloud-udbydere.

Siden 2023, med den aktive udvikling af store sprogmodeller (LLM), er tilgangen til datalagring begyndt at ændre sig. Flere og flere virksomheder overvejer at tage kontrollen over deres data tilbage, da det bliver sikrere og mere rentabelt at behandle oplysninger på deres egne servere.

I denne sammenhæng kommer tendensen væk fra cloudbaseret lagring og behandling af kun de nødvendige data til fordel for lokal implementering af enterprise LLM og AI -løsninger til syne. Som Microsofts CEO påpegede i et af sine interviews [27], vil AI-agenter styre processer i databaser og automatisere forskellige systemers funktioner i stedet for at være afhængige af flere separate applikationer eller cloudbaserede SaaS-løsningerat udføre forskellige opgaver.

> *[...] den gamle tilgang til dette [databehandling] problem var: Hvis du tænker tilbage på, hvordan forskellige forretningsapplikationer håndterede integration, så brugte de connectors. Virksomhederne solgte licenser til disse konnektorer, og forretningsmodellen var bygget op omkring det. SAP [ERP] er et af de klassiske eksempler: Man kunne kun få adgang til SAP-data, hvis man havde den rigtige connector. Så det forekommer mig, at noget lignende vil opstå i forbindelse med [AI] agentinteraktion [...]. Vores tilgang er i hvert fald: Jeg tror, at konceptet med eksistensen af forretningsapplikationer sandsynligvis vil kollapse i [AI]-agenternes æra. For hvis man tænker over det, er de i bund og grund databaser med en masse forretningslogik*
>
> - Satya Nadella, Microsofts CEO, interview med BG2 channel, 2024. [28]

I dette paradigme går den datadrevne LLM-tilgang ud over klassiske systemer. Kunstig intelligens bliver et mellemled mellem brugeren og data (fig. 2.2-3, fig. 2.2-4), hvilket eliminerer behovet for flere mellemliggende grænseflader og øger effektiviteten af forretningsprocesser. Vi vil tale mere om denne tilgang til at

arbejde med data i kapitlet "Gør kaos til orden og reducer kompleksiteten".

Mens fremtidens arkitektur stadig er ved at tage form, står virksomheder allerede over for konsekvenserne af tidligere beslutninger. Den massive digitalisering i de seneste årtier, ledsaget af indførelsen af forskellige systemer og ukontrolleret ophobning af data, har ført til et nyt problem - informationsoverbelastning.

Grænser for dataakkumulering: fra masse til mening

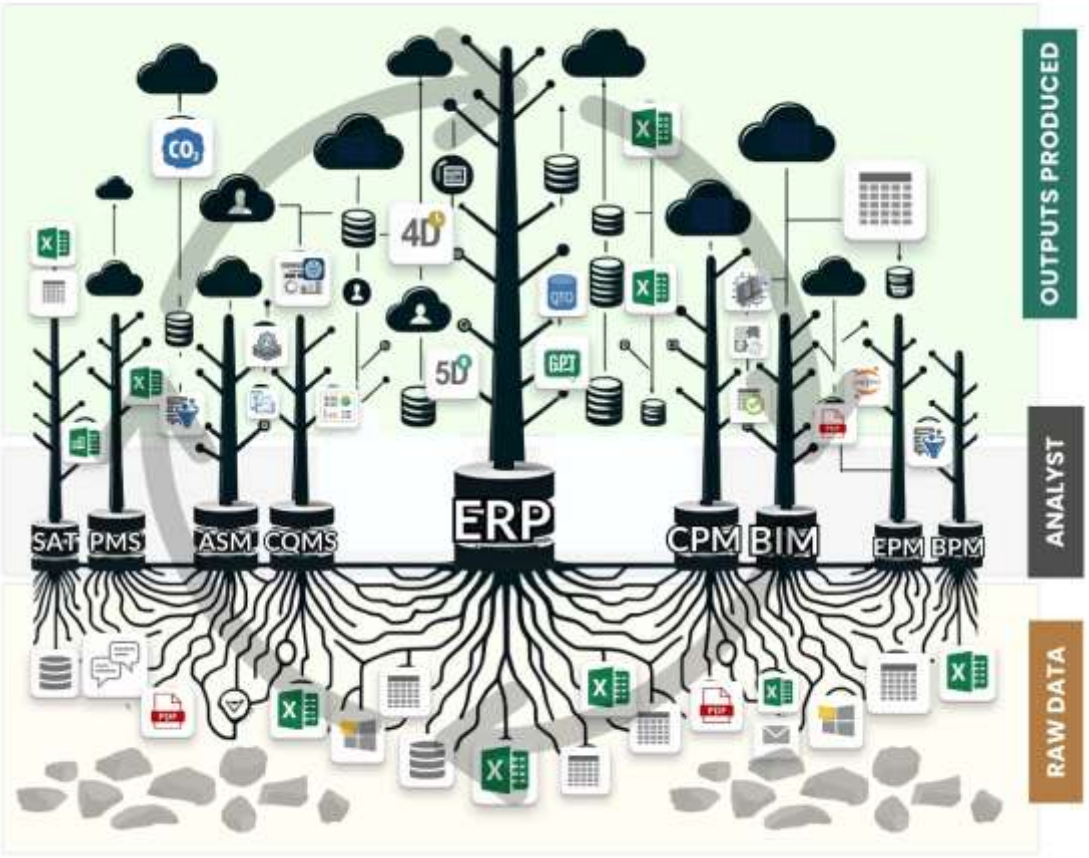
Moderne virksomhedssystemer udvikler sig og fungerer med succes under kontrolleret vækst, når datamængden og antallet af applikationer er i balance med it-afdelingernes og ledernes evner. Men i de seneste årtier har digitaliseringen ført til en ukontrolleret stigning i datamængden og -kompleksiteten, hvilket har skabt en overmætningseffekt i virksomhedernes informationsøkosystem.

I dag udsættes servere og lagerfaciliteter for en hidtil uset tilstrømning af ubehandlede og multiformaterede oplysninger, som ikke har tid til at blive til kompost og hurtigt bliver irrelevante. Virksomhedens begrænsede ressourcer kan ikke klare denne syndflod, og data ophobes i isolerede siloer (såkaldte "siloer "), der kræver manuel behandling for at udtrække nyttig information.

> Som følge heraf lider moderne virksomhedsledelsessystemer ofte af informationsoverbelastning som en skov, der er overgroet med efeu og dækket af skimmelsvamp. I stedet for at give næring til informationshumus i hjertet af virksomhedens økosystem dannes der isolerede områder med data i forskellige formater, hvilket uundgåeligt fører til et fald i forretningsprocessernes samlede effektivitet.

Den lange periode med eksponentiel datavækst i løbet af de sidste 40 år vil uundgåeligt blive efterfulgt af en mætningsfase og efterfølgende afkøling. Når lagring når sine grænser, vil der ske et kvalitativt skift: Data vil ikke længere bare være et lagringsobjekt, men en strategisk ressource.

Med udviklingen af kunstig intelligens og maskinlæring har virksomhederne mulighed for at reducere omkostningerne til informationsbehandling og gå fra kvantitativ vækst til kvalitativ brug af data. I løbet af det næste årti bliver byggebranchen nødt til at skifte fokus fra at skabe flere og flere data til at sikre deres struktur, integritet og analytiske værdi.



Figur 1.3-6 Isolerede datakilder forhindrer informationsdeling mellem datasystemer.

> Hovedværdien ligger ikke længere i mængden af information, men i evnen til automatisk at fortolke den og omdanne den til anvendt viden, der kan bruges til at træffe ledelsesmæssige beslutninger. For at data kan blive virkelig nyttige, skal de håndteres korrekt: indsamles, verificeres, struktureres, lagres og analyseres i forbindelse med specifikke forretningsopgaver.

Dataanalyseprocessen i en virksomhed svarer til cyklussen af liv og forfald af træer i en skov og fremkomsten af nye unge og stærke træer: modne træer dør, nedbrydes og bliver et nærende miljø for ny vækst. Færdige og afsluttede processer bliver, når de er færdige, en del af virksomhedens informationsøkosystem og bliver til sidst den informationshumus, der giver brændstof til den fremtidige vækst af nye systemer og data.

I praksis bliver denne cyklus dog ofte brudt. I stedet for organisk fornyelse dannes der et lagdelt kaos, der ligner geologiske lag, hvor nye systemer lægges oven på gamle uden dyb integration og strukturering. Resultatet er, at der opstår forskellige informations-"siloer", som hindrer cirkulation af viden og komplicerer datahåndtering.

Næste skridt: fra datateori til praktisk forandring

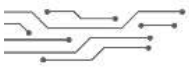
Udviklingen af data i byggeriet er en rejse fra lertavler til moderne modulære platforme. Udfordringen i dag er ikke at indsamle information, men at skabe en ramme, der gør forskellige og forskelligartede data til en strategisk ressource. Uanset om du er virksomhedsleder eller ingeniør, vil det være en nøglefærdighed i fremtiden at forstå værdien af data, og hvordan man arbejder med dem.

For at opsummere denne del er det værd at fremhæve de vigtigste praktiske trin, der vil hjælpe dig med at anvende de diskuterede tilgange i dine daglige opgaver:

- Gennemfør en personlig revision af informationsstrømme
  - [ ] Lav en liste over alle de systemer og applikationer, du arbejder med til daglig.
  - [ ] Marker, hvor du bruger mest tid på at søge eller dobbelttjekke data
  - [ ] Identificer dine vigtigste informationskilder
  - [ ] Analyser dit nuværende applikationslandskab for redundans og duplikering af funktioner
- Stræb efter at bevæge dig gennem processerne efter niveauer af analytisk modenhed
  - [ ] Start dine opgaver med beskrivende analyser (hvad skete der?)
  - [ ] Introducer gradvist en diagnose (hvorfor skete dette?)
  - [ ] Tænk over, hvordan du i processer kan gå over til prædiktive (hvad vil der ske?) og præskriptive (hvad skal der gøres?) analyser.
- Begynd at strukturere dine arbejdsdata
  - [ ] Implementer et fælles system til navngivning af filer og mapper, som du ofte bruger i dit arbejde.
  - [ ] Opret skabeloner til ofte anvendte dokumenter og rapporter
  - [ ] Arkiver regelmæssigt afsluttede projekter med en klar struktur

Selv om du ikke kan ændre hele informationsinfrastrukturen i dit team eller din virksomhed, så start med dine egne processer og små forbedringer i dit daglige arbejde. Husk, at den virkelige værdi af data ikke ligger i mængden, men i evnen til at uddrage brugbare indsigter fra dem. Selv små, men korrekt strukturerede og analyserede datasæt kan have en betydelig indvirkning, når de integreres i beslutningsprocesser.

I de næste dele af bogen går vi videre til specifikke metoder og værktøjer til at arbejde med data, ser på måder at omdanne ustruktureret information til strukturerede sæt, udforsker teknologier til automatisering af analyser og beskriver, hvordan man opbygger et effektivt økosystem for analyser i en byggevirksomhed.





II DEL
HVORDAN BYGGEBRANCHEN DRUKNER I DATA-KAOS

Den anden del er helliget en kritisk analyse af de udfordringer, som byggevirksomhederne står over for, når de skal håndtere de stigende datamængder. Konsekvenserne af informationsfragmentering og fænomenet "data i siloer", som hindrer effektiv beslutningstagning, undersøges i detaljer. Problemerne med HiPPO -tilgangen (Highest Paid Person's Opinion) og dens indvirkning på kvaliteten af ledelsesbeslutninger i byggeprojekter undersøges. Indvirkningen af dynamiske forretningsprocesser og deres voksende kompleksitet på informationsstrømme og driftseffektivitet vurderes. Der gives specifikke eksempler på, hvordan overdreven systemkompleksitet øger omkostningerne og reducerer organisationernes fleksibilitet. Der lægges særlig vægt på de begrænsninger, der skabes af proprietære formater, og udsigterne til brug af åbne standarder i byggebranchen. Konceptet med at bevæge sig mod AI- og LLM-baserede softwareøkosystemer præsenteres, som minimerer overdreven kompleksitet og tekniske barrierer.

KAPITEL 2.1.
DATAFRAGMENTERING OG SILOER

Jo flere værktøjer, jo mere effektiv virksomhed?

Ved første øjekast kan det se ud til, at flere digitale værktøjer fører til større effektivitet. I praksis er det dog ikke tilfældet. Med hver ny løsning, hvad enten det er en cloud-tjeneste, et ældre system eller endnu en Excel-rapport, tilføjer en virksomhed endnu et lag til sit digitale landskab - et lag, der ofte ikke er integreret med resten (figur 2.1-1).

Data kan sammenlignes med kul eller olie: Det tager år at opbygge dem, komprimeret under lag af kaos, fejl, ustrukturerede processer og glemte formater. For at få virkelig brugbar information ud af dem skal virksomhederne bogstaveligt talt vade gennem lag af forældede løsninger og digital støj.

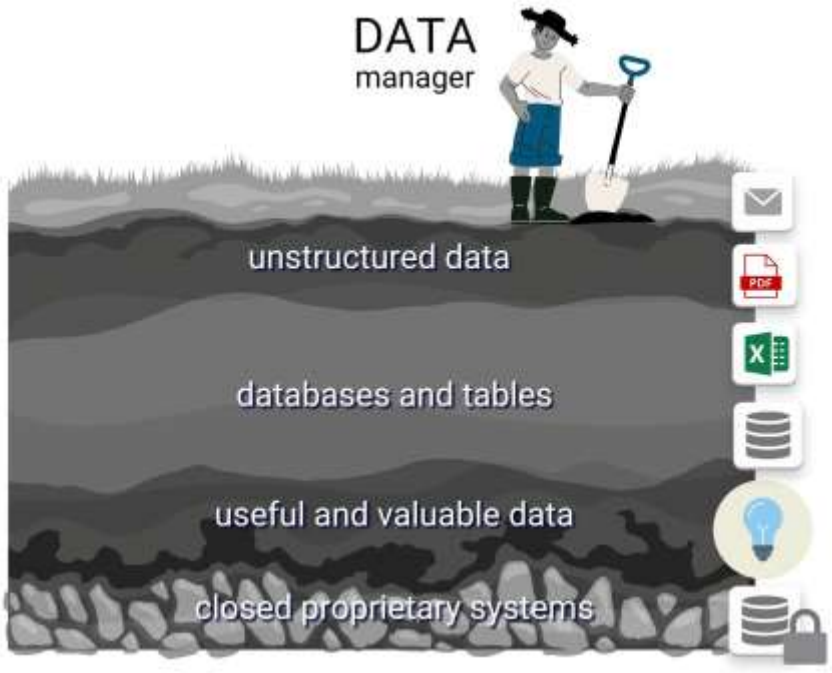


Fig. 2.1-1 Forskellige data danner afgrænsede lag - selv "gyldne" indsigter går tabt i geologiske klipper af systemisk kompleksitet.

Hver ny applikation efterlader et spor: en fil, en tabel eller en hel isoleret "silo" på serveren. Et lag er ler (forældede og glemte data), et andet er sand (forskellige tabeller og rapporter), og det tredje er granit (lukkede proprietære formater, der ikke kan integreres). Over tid ligner en virksomheds digitale miljø i stigende grad et reservoir med ukontrolleret ophobning af information, hvor værdien går tabt dybt inde i virksomhedens servere.

Med hvert nyt projekt og hvert nyt system bliver ikke kun infrastrukturen mere kompleks, men også vejen til brugbare kvalitetsdata. At nå frem til den værdifulde "klippe" kræver dyb udrensning, strukturering af information, "chunking", gruppering i meningsfulde bidder og udtrækning af strategisk vigtige indsigter

gennem analyse og datamodellering.

> *Data er en værdifuld ting, og de vil holde længere end selve systemerne [der behandler data] [29].*
>
> - Tim Berners-Lee, far til World Wide Web og skaber af den første hjemmeside

Før data kan blive en "værdifuld ting" og et pålideligt grundlag for beslutningstagning, skal de gennemgå en omhyggelig forberedelse. Det er den rette forbehandling, der forvandler uensartede data til struktureret erfaring, nyttig informationshumus, som derefter bliver et værktøj til prognoser og optimering.

> Det er en misforståelse, at man skal have helt rene data for at begynde at analysere, men i praksis er det en vigtig del af processen at kunne arbejde med beskidte data.

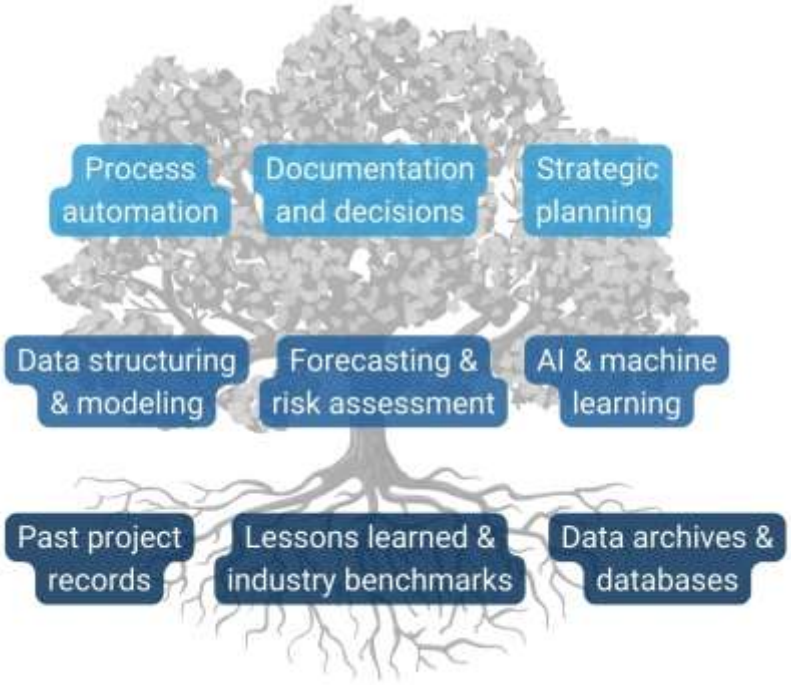


Fig. 2.1-2 Data er rodsystemet og fundamentet for forretningen, som igen er baseret på beslutningsprocesser.

I takt med at teknologien fortsætter med at udvikle sig, skal din virksomhed også bevæge sig fremad og lære at skabe værdi ud af data. Ligesom olie- og kulselskaber opbygger infrastrukturen til at udvinde mineraler, skal virksomheder også lære at styre strømmen af nye oplysninger på deres egne servere og udtrække værdifulde indsigter fra ubrugte, uformaterede og forældede data og gøre dem til en strategisk ressource.

Oprettelse af felter (datavarehuse) er det første skridt. Selv de mest kraftfulde værktøjer løser ikke problemet med dataisolering og data i flere formater, hvis virksomhederne fortsætter med at arbejde i silosystemer. Når data eksisterer adskilt fra hinanden uden at krydse hinanden og dele information, står virksomheder over for en "datasilo"-effekt. I stedet for en enkelt, konsistent infrastruktur er virksomhederne tvunget til at bruge ressourcer på at flette og synkronisere data.

Datasiloer og deres indvirkning på virksomhedens resultater

Forestil dig, at du er ved at bygge et boligområde, men at hvert hold har sit eget projekt. Nogle bygger mure, andre lægger kommunikation, og andre igen lægger veje uden at tjekke med hinanden. Resultatet er, at rørene ikke passer til åbningerne i væggene, at elevatorskakterne ikke passer til etagerne, og at vejene skal rives ned og lægges om.

Denne situation er ikke bare et hypotetisk scenarie, men en realitet i mange moderne byggeprojekter. På grund af det store antal hoved- og underentreprenører, der arbejder med forskellige systemer og uden et enkelt koordinerende center, bliver processen til en række endeløse godkendelser, omarbejdninger og konflikter. Alt dette fører til betydelige forsinkelser og flere projektomkostninger.

En klassisk situation på en byggeplads er enkel: Forskallingen er klar, men leveringen af armering er ikke kommet til tiden. Når man tjekker oplysningerne i forskellige systemer, ser kommunikationen nogenlunde sådan ud:

- **Formanden** på byggepladsen den 20. skriver til projektlederen*: "Vi er færdige med at sætte forskallingen, hvor er armeringsjernet?*"
- **Projektleder** (PMIS) til indkøbsafdelingen**:** - "*Forskallingen er klar. I mit system [PMIS] står der, at armeringsjernet skulle være kommet den 18. Hvor er armeringsjernet?"*
- **Specialist i forsyningskæden** (ERP): - "Vores ERP siger, at leveringen vil ske den 25.".
- **Dataingeniør** eller IT-afdeling (ansvarlig for integrationer): - I PMIS er datoen den 18., i ERP er det den 25. Der er ikke noget OrderID-link mellem ERP og PMIS, så dataene er ikke synkroniseret. Dette er et typisk eksempel på et informationsgab.
- **Projektleder** til General **Manager** - "*Leveringen af inventar er forsinket, byggepladsen står stille, og det er uklart, hvem der har ansvaret*".

Årsagen til hændelsen var isoleringen af data i forskellige systemer. Ved at integrere og forene datakilder, skabe et enkelt informationslager og automatisere ved hjælp af ETL -værktøjer (Apache NiFi, Airflow eller n8n) kan siloerne mellem systemerne elimineres. Disse og andre metoder og værktøjer vil blive diskuteret i detaljer i senere afsnit af bogen.

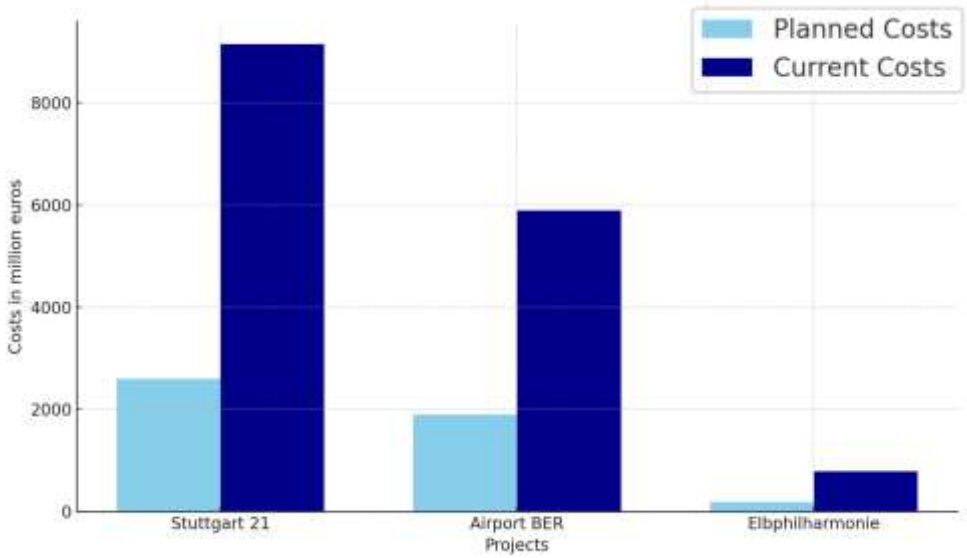


Fig. 2.1-3 Sammenligning af planlagte og faktiske omkostninger ved store infrastrukturprojekter i Tyskland.

Det samme sker med virksomhedssystemer: Først skabes isolerede løsninger, og så skal der bruges enorme budgetter på at integrere og harmonisere dem. Hvis data- og kommunikationsmodellerne var blevet gennemtænkt fra starten, ville der slet ikke være behov for integration. Silodata skaber kaos i den digitale verden, ligesom en ukoordineret byggeproces.

> Ifølge KPMG's 2023-undersøgelse "Cue construction 4.0: Time to make or break" deler kun 36% af virksomhederne data effektivt på tværs af afdelinger, mens 61% står over for alvorlige problemer på grund af isolerede data-"siloer" [30].

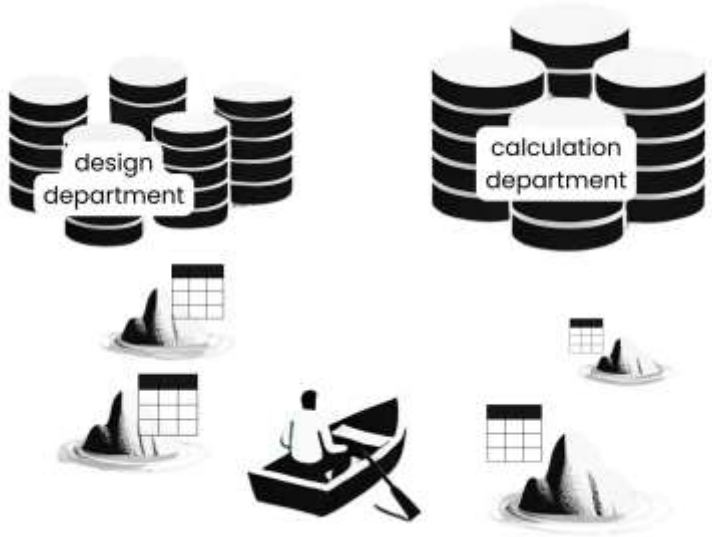


Fig. 2.1-4 Årelange data, der er svære at indsamle, ophobes i isolerede "siloer" med risiko for aldrig at blive brugt.

Virksomhedsdata lagres i isolerede systemer, som individuelle træer spredt ud over landskabet. De

indeholder hver især værdifulde oplysninger, men manglen på forbindelser mellem dem forhindrer, at der skabes et enkelt, sammenkoblet økosystem. Denne siloopdeling hindrer dataflowet og begrænser organisationens evne til at se det fulde billede. At forbinde disse siloer er en ekstremt lang og kompleks proces, hvor man dyrker svampemycelium på ledelsesniveau for at lære at overføre individuelle informationer mellem systemerne.

Ifølge en WEF-undersøgelse fra 2016 er en af de største barrierer for digital transformation manglen på fælles datastandarder og fragmentering.

> *Byggebranchen er en af de mest fragmenterede i verden og er afhængig af et gnidningsløst samspil mellem alle deltagere i værdikæden [5].*
>
> - World Economic Forum 2016: Formning af fremtidens

Designere, ledere, koordinatorer og udviklere foretrækker ofte at arbejde selvstændigt for at undgå kompleksiteten ved koordinering. Denne naturlige tilbøjelighed fører til oprettelse af informations-"siloer", hvor data er isoleret i separate systemer. Jo flere sådanne isolerede systemer der er, jo sværere er det at få dem til at arbejde sammen. Med tiden får hvert system sin egen database og en specialiseret supportafdeling med ledere (figur 1.2-4), hvilket komplicerer integrationen yderligere.

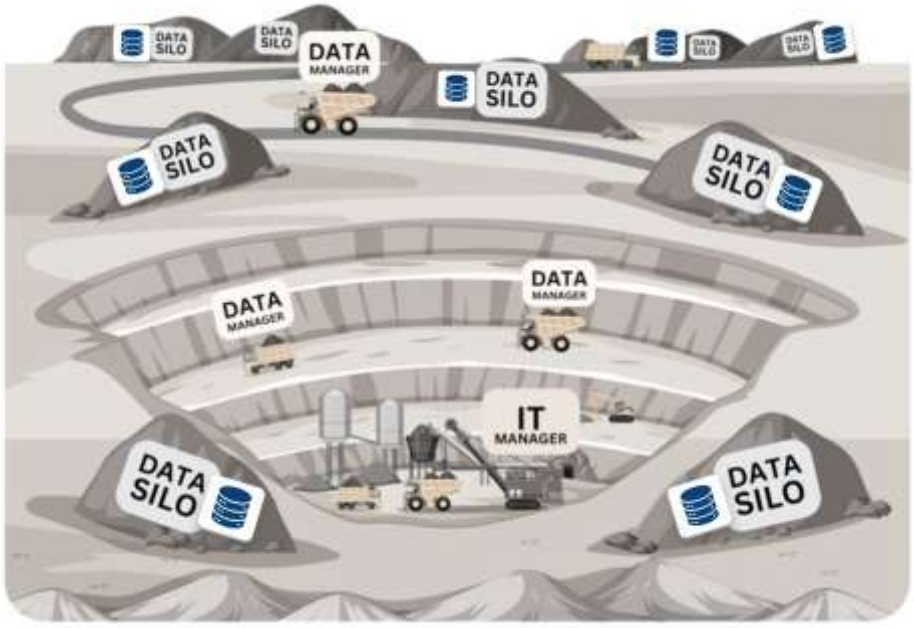


Fig. 2.1-5 Hvert system har en tendens til at skabe sin egen unikke silo af data, som skal behandles af passende værktøjer [31].

Den onde cirkel i virksomhedssystemer ser således ud: Virksomhederne investerer i komplekse iso leerede løsninger, står derefter over for høje omkostninger til deres integration, og udviklere, der indser kompleksiteten i at kombinere systemer, foretrækker at arbejde i deres lukkede økosystemer. Alt dette øger fragmenteringen af IT-landskabet og gør det sværere at migrere til nye løsninger (figur 2.1-5). Ledere ender med at kritisere datasiloer, men analyserer sjældent årsagerne til dem, og hvordan de kan forhindres. Ledere

klager over forældede IT-systemer, men udskiftning af dem kræver betydelige investeringer og giver sjældent de forventede resultater. Resultatet er, at selv forsøg på at bekæmpe problemet ofte gør tingene værre.

Hovedårsagen til den manglende forbindelse er prioriteringen af applikationer frem for data. Virksomheder udvikler først separate systemer eller køber standardløsninger fra leverandører og forsøger derefter at forene dem ved at oprette dobbelt og inkompatibel lagring og databaser.

At overvinde problemet med fragmentering kræver en radikal ny tilgang - at prioritere data frem for applikationer. Virksomheder skal først udvikle datastyringsstrategier og datamodeller og derefter bygge systemer eller købe løsninger, der arbejder med et enkelt sæt oplysninger i stedet for at skabe nye barrierer.

> *Vi er på vej ind i en ny verden, hvor data måske er vigtigere end software.*
>
> - Tim O'Reilly, administrerende direktør for O'Reilly Media, Inc.

> McKinsey Global Institutes undersøgelse "Rethinking Construction: the path to improved productivity" (2016) viser, at byggebranchen halter bagefter andre sektorer, når det gælder digital transformation [32]. Ifølge rapporten kan indførelsen af automatiseret datastyring og digitale platforme forbedre produktiviteten betydeligt og reducere tab i forbindelse med inkonsekvente processer. Dette behov for digital transformation understreges også af Egan-rapporten (UK, 1998) [33], som fremhæver den centrale rolle, som integrerede processer og en samarbejdsorienteret tilgang spiller i byggeriet.

> Resultatet er, at mens det største problem for dataadministratorer i de sidste 10.000 år har været mangel på data, står brugere og administratorer med en lavine af data og dataadministrationssystemer over for et problem - en overflod af data, der gør det vanskeligt at finde juridisk korrekt information af høj kvalitet.

De forskellige datasiloer fører uundgåeligt til det alvorlige problem med reduceret datakvalitet. Med flere uafhængige systemer kan de samme data findes i forskellige versioner, ofte med modstridende værdier, hvilket skaber yderligere vanskeligheder for brugere, der skal afgøre, hvilke oplysninger der er relevante og pålidelige.

Dobbeltarbejde og manglende datakvalitet som følge af uenighed

På grund af problemet med datasiloer skal ledere bruge meget tid på at søge efter og afstemme data. For at sikre sig mod kvalitetsproblemer skaber virksomheder komplekse informationsstyringsstrukturer, hvor en vertikal af ledere er ansvarlig for at søge, verificere og afstemme data. Men denne tilgang øger kun bureaukratiet og gør beslutningstagningen langsommere. Jo flere data der er, jo sværere er det at analysere og fortolke dem, især hvis der ikke er nogen ensartet standard for opbevaring og behandling af dem.

Med den overflod af softwareapplikationer og systemer, der er vokset frem som paddehatte i det sidste årti, er problemet med siloer og uhensigtsmæssig datakvalitet blevet stadig vigtigere for slutbrugerne. De samme data, men med forskellige værdier, kan nu findes i forskellige systemer og applikationer (figur 2.1-

6). Det fører til vanskeligheder for slutbrugerne, når de skal finde ud af, hvilken version af data der er relevant og korrekt blandt de mange tilgængelige. Det fører til fejl i analyser og i sidste ende i beslutningstagningen.

For at sikre sig mod problemer med at finde de rigtige data opretter virksomhedslederne et bureaukrati af verifikationsansvarlige på flere niveauer. Deres opgave er hurtigt at kunne finde, kontrollere og sende de nødvendige data i form af tabeller og rapporter og navigere i labyrinten af forskellige systemer.

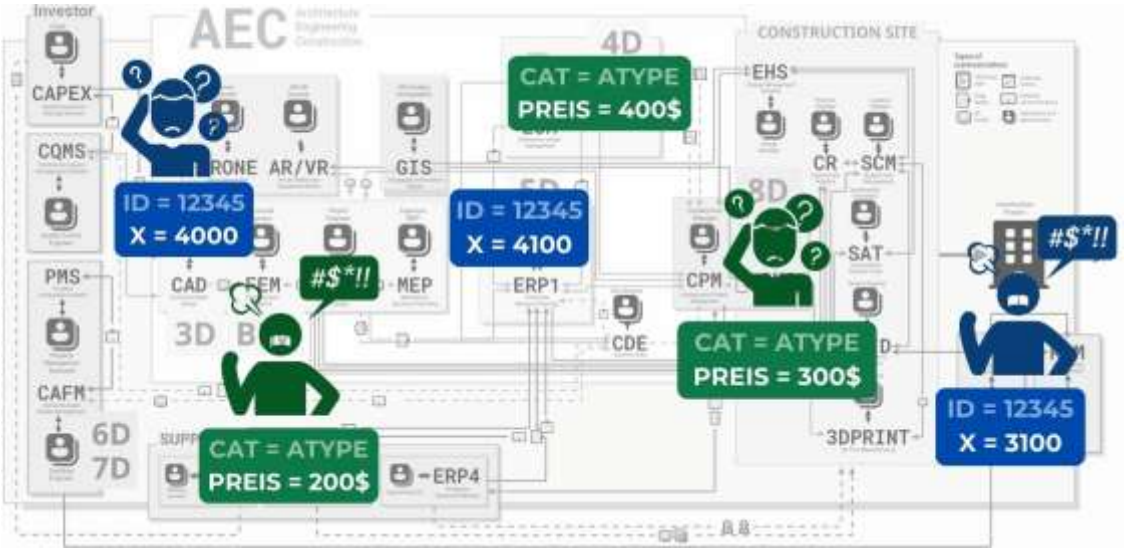


Fig. 2.1-6 I forsøget på at finde de rigtige data skal ledere sikre kvaliteten og den juridiske pålidelighed af data mellem forskellige systemer.

I praksis skaber denne model dog nye kompleksiteter. Når data håndteres manuelt, og information er spredt over mange ikke-relaterede beslutninger, bliver ethvert forsøg på at få nøjagtig og opdateret information gennem en pyramide af beslutningstagere (figur 2.1-7) en flaskehals - tidskrævende og fejlbehæftet.

Situationen forværres af en lavine af digitale løsninger. Softwaremarkedet oversvømmes fortsat af nye værktøjer, der virker lovende. Men uden en klar datastyringsstrategi kan disse løsninger ikke integreres i et samlet system, men skaber i stedet yderligere lag af kompleksitet og overlapning. Resultatet er, at virksomhederne i stedet for at forenkle processerne befinder sig i et endnu mere fragmenteret og kaotisk informationsmiljø.

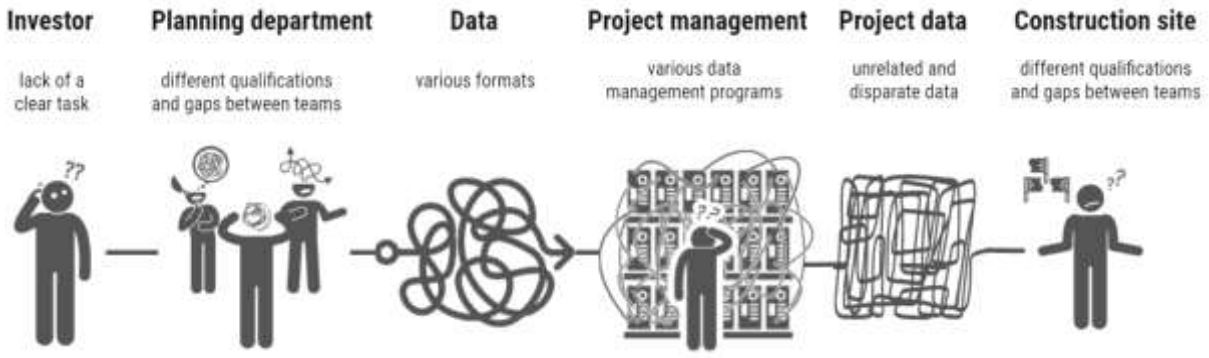


Fig. 2.1-7 Systemernes kompleksitet og de mange forskellige dataformater fører til et tab af konsistens i byggeprocessen.

Alle disse problemer, der er forbundet med at administrere et væld af forskellige løsninger, får før eller siden virksomhedsledelsen til at indse noget vigtigt: Det er ikke datamængden eller jagten på det næste "one-size-fits-all"-databehandlingsværktøj, der er problemet. Den egentlige årsag ligger i kvaliteten af data, og hvordan organisationen skaber, modtager, lagrer og bruger dem.

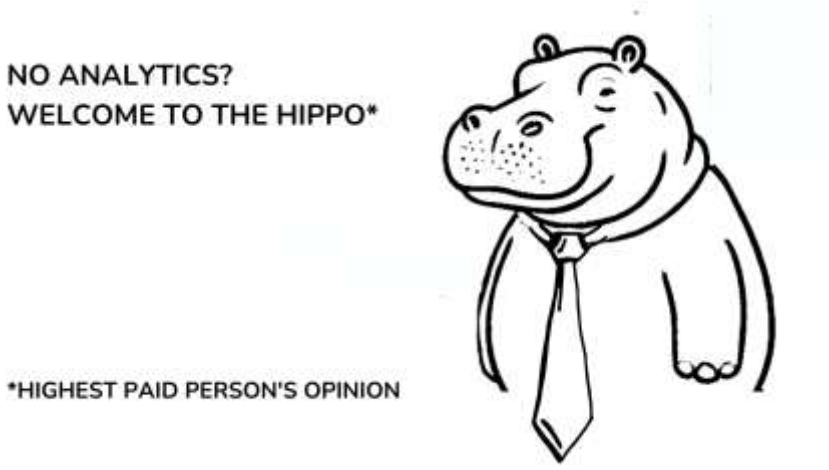
> Nøglen til bæredygtig succes ligger ikke i at jagte nye "magiske" applikationer, men i at opbygge en datakultur i virksomheden. Det betyder, at data skal behandles som et strategisk aktiv, og at datakvalitet, -integritet og -relevans skal prioriteres på alle niveauer i organisationen.

Løsningen på dilemmaet mellem kvalitet og kvantitet ligger i at skabe en samlet datastruktur, der eliminerer dobbeltarbejde, fjerner uoverensstemmelser og ensretter informationsstrømme. Denne arkitektur giver en enkelt, pålidelig datakilde, som kan bruges til at træffe informerede, nøjagtige og rettidige beslutninger.

Ellers fortsætter virksomhederne med at stole på HiPPO-eksperternes subjektive meninger og intuitive vurderinger snarere end på pålidelige fakta, hvilket stadig ofte er tilfældet. I byggebranchen, hvor ekspertise traditionelt spiller en væsentlig rolle, er dette særligt mærkbart.

HiPPO eller faren ved meninger i beslutningsprocessen

Traditionelt træffes vigtige beslutninger i byggebranchen på baggrund af erfaring og subjektiv vurdering. Uden rettidige og pålidelige data er virksomhedsledere nødt til at handle i blinde og stole på de højest betalte medarbejderes intuition (HiPPO - Highest Paid Person's Opinion) snarere end på objektive fakta (Fig. 2.1-8).



Figur 2.1-8 I mangel af analyser afhænger forretningen af erfarne fagfolks subjektive mening.

Denne tilgang kan være berettiget i et stabilt og langsomt skiftende miljø, men i en tid med digital transformation bliver det en alvorlig risiko. Beslutninger baseret på intuition og gætværk er tilbøjelige til at blive forvrænget, er ofte baseret på uunderbyggede hypoteser og tager ikke højde for det komplekse billede der afspejles i dataene.

> Det, der går for at være en intelligent debat på beslutningsniveau i en virksomhed, er ofte ikke baseret på noget konkret. En virksomheds succes bør ikke afhænge af eksperters autoritet og løn, men af evnen til at arbejde effektivt med data, identificere mønstre og træffe informerede beslutninger.

Det er vigtigt at opgive forestillingen om, at autoritet eller erfaring automatisk betyder, at en beslutning er rigtig. Den datadrevne tilgang er en game changer: data og analyser, ikke stilling og løn, er nu grundlaget for beslutningstagning. Big data, maskinlæring og visuel analyse giver os mulighed for at identificere mønstre og stole på fakta i stedet for gætværk (fig. 1.1-4).

> *Uden data er du bare endnu en person med en mening [34].*
>
> - W. Edwards Deming, forsker og ledelseskonsulent

Moderne datastyringsmetoder sikrer også kontinuitet i viden inden for virksomheden. Klart beskrevne processer, automatisering og en systematisk tilgang gør det muligt at overføre selv nøgleroller uden at miste effektivitet.

Men blind tillid til data kan også føre til alvorlige fejl. Data i sig selv er bare en samling tal. Uden ordentlig analyse, kontekst og evnen til at identificere mønstre har de ingen værdi og kan ikke drive processer. Nøglen til succes ligger ikke i at vælge mellem HiPPO-intuition og analyse, men i at opbygge intelligente værktøjer, der omdanner forskelligartet information til håndterbare, informerede beslutninger.

> I et digitalt byggemiljø er det ikke anciennitet og plads i hierarkiet, der bliver afgørende succesfaktorer, men lydhørhed, beslutningsnøjagtighed og ressourceeffektivitet

Data er værktøjer, ikke absolutte sandheder. De skal supplere menneskelig tænkning, ikke erstatte den. På trods af fordelene ved analyse kan data ikke helt erstatte menneskelig intuition og erfaring. Deres rolle er at hjælpe med at træffe mere præcise og informerede beslutninger.

Konkurrencefordele opnås ikke kun ved at opfylde standarder, men ved at være i stand til at overgå konkurrenterne i effektiv brug af ressourcer, der er ens for alle. I fremtiden vil datafærdigheder blive lige så vigtige som læse- og skrivefærdigheder eller matematiske færdigheder engang var. Fagfolk, der kan analysere og fortolke data, vil være i stand til at træffe mere præcise beslutninger og fortrænge dem, der kun baserer sig på personlig erfaring (figur 2.1-9).

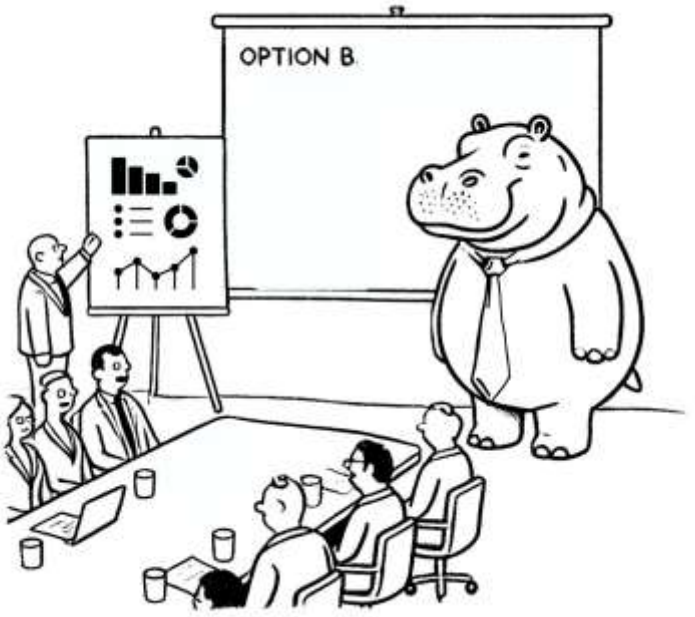


Fig. 2.1-9 Beslutninger skal baseres på en objektiv analyse, ikke på den højest lønnede medarbejders mening.

Ledere, specialister og ingeniører vil fungere som dataanalytikere, der studerer projekternes struktur, dynamik og nøgleindikatorer. Menneskelige ressourcer vil blive elementer i systemet, som kræver fleksibel datadrevet tilpasning for at maksimere effektiviteten.

> *Fejl ved brug af utilstrækkelige data er meget mindre end ved brug af ingen data [35].*
>
> - Charles Babbage, opfinder af den første analytiske regnemaskine

Fremkomsten af big data og introduktionen af LLM (Large Language Models) har radikalt ændret ikke kun den måde, vi analyserer på, men også selve beslutningstagningens natur. Hvor der tidligere var fokus på kausalitet (hvorfor noget skete - diagnostisk analyse) (Fig. 1.1-4), er det i dag evnen til at forudsige fremtiden (prædiktiv analyse) og i fremtiden præskriptiv analyse, hvor maskinlæring og AI foreslår det bedste valg i beslutningsprocessen, der kommer i forgrunden.

> Ifølge den nye SAP™-undersøgelse, "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36], ville 44% af topledere være villige til at ændre deres tidligere beslutning baseret på AI-rådgivning, og 38% ville stole på, at AI træffer forretningsbeslutninger på deres vegne. I mellemtiden sagde 74% af lederne, at de stoler mere på AI-rådgivning end på deres venner og familie, og 55% arbejder i virksomheder, hvor AI-afledte indsigter erstatter eller ofte omgår traditionelle beslutningsmetoder - især i organisationer med en årlig omsætning på over 5 milliarder dollars. Derudover bruger 48% af de adspurgte generative AI-værktøjer dagligt, herunder 15%, som bruger dem flere gange om dagen.

Med udviklingen af LLM og automatiserede datastyringssystemer opstår der en ny udfordring: hvordan man

bruger information effektivt uden at miste dens værdi i kaosset af inkompatible formater og heterogene kilder, som suppleres af den voksende kompleksitet og dynamik i forretningsprocesserne.

Kontinuerlig stigning i forretningsprocessernes kompleksitet og dynamik

Byggebranchen står i dag over for alvorlige udfordringer inden for data- og processtyring. De største udfordringer er siloopdelte informationssystemer, overdrevent bureaukrati og manglende integration mellem digitale værktøjer. Disse udfordringer intensiveres, efterhånden som forretningsprocesserne selv bliver mere komplekse - drevet af teknologi, ændrede kundekrav og nye regler.

At byggeprojekter er unikke, skyldes ikke kun deres tekniske særpræg, men også forskelle i nationale standarder og lovkrav i forskellige lande (fig. 4.2-10, fig. 5.1-7). Det kræver en fleksibel, individuel tilgang til hvert projekt, som er vanskelig at implementere i traditionelle modulære kontrolsystemer. På grund af processernes kompleksitet og den store mængde data henvender mange virksomheder sig til leverandører, der tilbyder specialiserede løsninger. Men markedet er overbelastet - mange nystartede virksomheder tilbyder lignende produkter med fokus på snævre opgaver. Resultatet er, at en holistisk tilgang til datahåndtering ofte går tabt.

> Tilpasning til den kontinuerlige strøm af nye teknologier og markedskrav er ved at blive en kritisk faktor for konkurrenceevnen. Men eksisterende proprietære applikationer og modulsystemer har lav tilpasningsevne - enhver ændring kræver ofte langvarige og dyre revisioner af udviklere, som ikke altid forstår detaljerne i byggeprocesserne.

Virksomhederne bliver gidsler i den teknologiske forsinkelse og venter på nye opdateringer i stedet for straks at implementere innovative, integrerede tilgange. Som følge heraf er byggeorganisationernes interne struktur ofte et komplekst økosystem af indbyrdes forbundne hierarkiske og ofte lukkede systemer, der koordineres gennem et netværk af ledere på flere niveauer (figur 2.1-10).

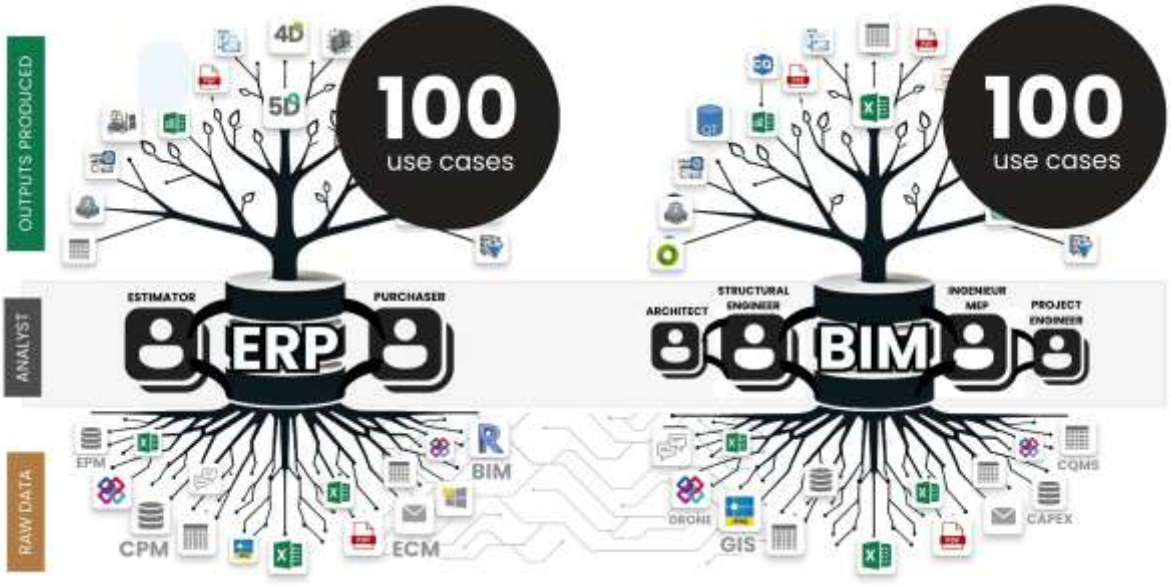


Fig. 2.1-10 Virksomheder består af sammenkoblede systemer, hvis sammenkobling danner processer, der kræver automatisering.

> Ifølge en undersøgelse foretaget af Canadian Construction Association og KPMG Canada i 2021 [37] mener kun 25% af virksomhederne, at de befinder sig i en væsentlig eller anderledes position sammenlignet med konkurrenterne, når det gælder indførelse af teknologi eller digitale løsninger. Kun 23% af de adspurgte rapporterede, at deres løsninger i høj grad eller i meget høj grad er datadrevne. Samtidig karakteriserede størstedelen af de adspurgte deres brug af en række andre teknologier som rent eksperimentelle eller indrømmede, at de slet ikke brugte dem.

Denne modvilje mod at deltage i teknologiske eksperimenter er særlig tydelig i store infrastrukturprojekter, hvor fejl kan koste millioner af dollars. Selv de mest avancerede teknologier - digitale tvillinger, prædiktive analyser - møder ofte modstand, ikke på grund af deres effektivitet, men på grund af manglen på dokumenteret pålidelighed i virkelige projekter.

> Ifølge World Economic Forums (WEF) rapport "Shaping the Future of Construction" [5] står indførelsen af nye teknologier i byggeriet ikke kun over for tekniske vanskeligheder, men også psykologiske barrierer fra kundernes side [5] står indførelsen af nye teknologier i byggeriet ikke kun over for tekniske vanskeligheder, men også psykologiske barrierer fra kundernes side. Mange kunder frygter, at brugen af avancerede løsninger vil gøre deres projekter til et eksperimentarium og gøre dem til "forsøgskaniner", og uforudsigelige konsekvenser kan føre til yderligere omkostninger og risici.

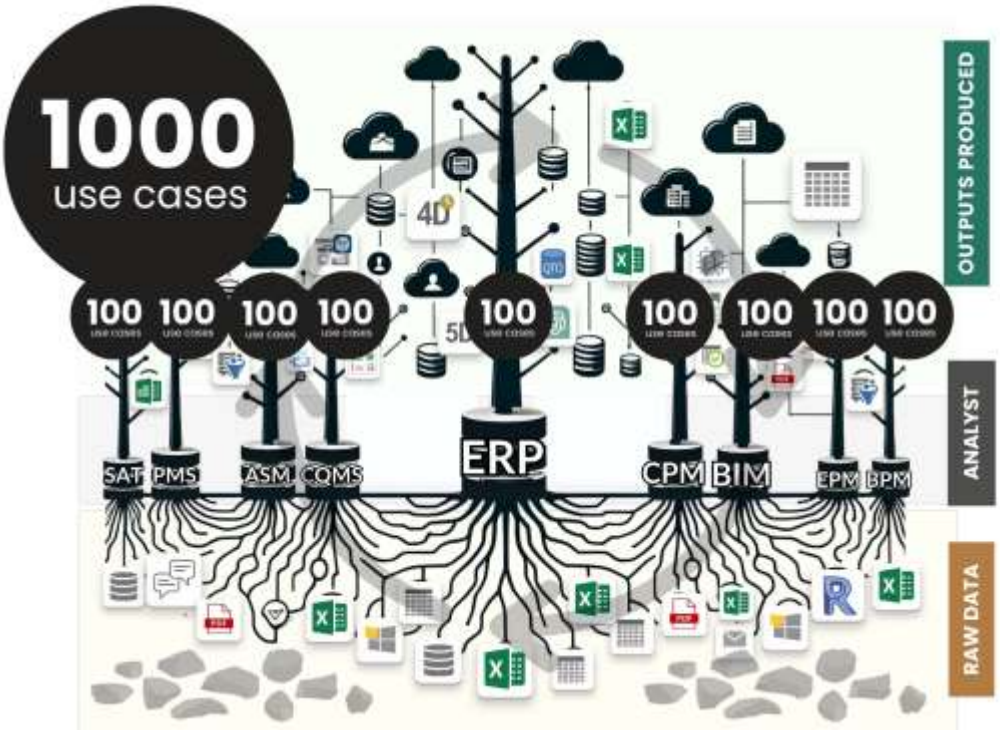


Fig. 2.1-11 For hver data use case tilbyder markedet for løsninger applikationer til optimering og automatisering af processer.

> Byggebranchen er meget forskelligartet: forskellige projekter har forskellige krav, regionale særpræg, lovbestemte klassifikationsregler (fig. 4.2-10), beregningsstandarder (fig. 5.1-7) osv. Derfor er det praktisk talt umuligt at skabe en proprietær universel applikation eller et system, der passer perfekt til alle disse krav og projektspecifikationer.

I et forsøg på at håndtere den voksende kompleksitet i systemerne og afhængigheden af softwareleverandører indser man i stigende grad, at nøglen til effektiv datahåndtering ikke kun er åbenhed og standardisering, men også forenkling af selve procesarkitekturen. Den voksende kompleksitet og dynamik i forretningsprocesserne kræver nye tilgange, hvor prioriteten skifter fra at akkumulere data til at strukturere og organisere dem. Det er dette skift, der vil være det næste skridt i udviklingen af byggebranchen og markere afslutningen på æraen med softwareleverandørernes dominans og begyndelsen på æraen med meningsfuld informationsorganisering.

Erkendelsen af begrænsningerne i one-size-fits-all-løsninger og sårbarheden over for stigende kompleksitet fører til et skift i prioriteterne fra lukkede platforme og datahopning til gennemsigtighed, tilpasningsevne og struktureret informationshåndtering. Dette skift i tankegang afspejler bredere ændringer i den globale økonomi og teknologi, som beskrives gennem de såkaldte "industrielle revolutioner". For at forstå, hvor byggeriet er på vej hen, og hvad dets fremtidige retning er, er det nødvendigt at overveje branchens plads i forbindelse med den fjerde og femte industrielle revolution - fra automatisering og digitalisering til personalisering, åbne standarder og den servicebaserede datamodel.

Den fjerde industrielle revolution (Industri 4.0) og den femte industrielle revolution (Industri 5.0) i byggeriet

Teknologiske og økonomiske stadier er teoretiske begreber, der bruges til at beskrive og analysere samfundets og økonomiens udvikling på forskellige udviklingstrin. De kan fortolkes forskelligt af forskellige forskere og eksperter.

- **Den fjerde industrielle revolution** (4IR eller Industry 4.0) er relateret til informationsteknologi, automatisering, digitalisering og globalisering. Et af nøgleelementerne er skabelsen af proprietære softwareløsninger, dvs. specialiserede digitale produkter, der er designet til specifikke opgaver og virksomheder. Disse løsninger bliver ofte en vigtig del af it-infrastrukturen, men er dårligt skalerbare uden yderligere ændringer.
- **Den femte industrielle revolution (5IR)** er nu på et tidligere stadie af konceptualisering og udvikling end 4IR. Dens kerneprincipper omfatter øget personalisering af produkter og tjenester. 5IR er en bevægelse mod mere tilpasningsdygtige, fleksible og personaliserede økonomiske aktiviteter med fokus på personalisering, rådgivning og serviceorienterede modeller. Et centralt aspekt af den femte økonomiske modus er brugen af data til beslutningstagning, hvilket stort set er umuligt uden brug af åbne data og åbne værktøjer (figur 2.1-12).

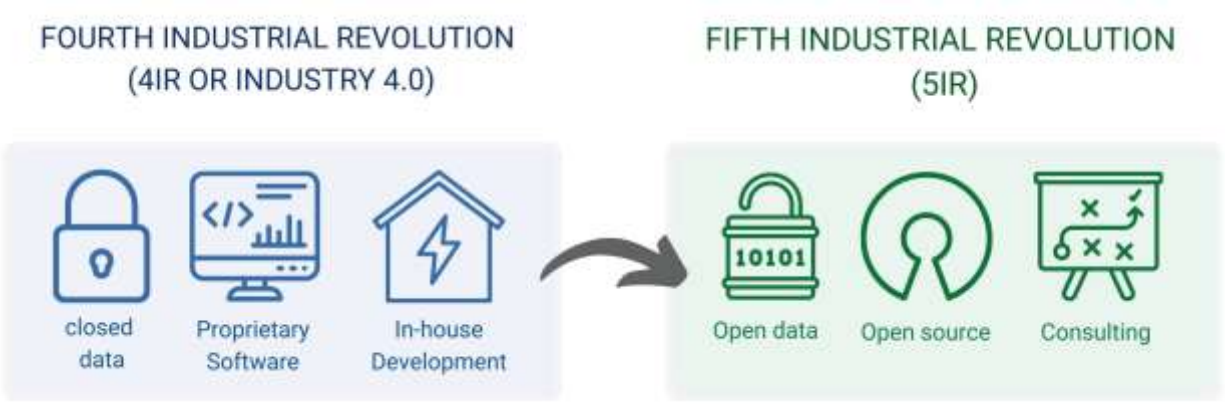


Fig. 2.1-12 Det fjerde mønster fokuserer på løsninger, mens det femte mønster fokuserer på personalisering og data.

> At skabe en applikation til virksomheder i byggebranchen til brug i ti eller hundrede organisationer garanterer ikke, at den med succes kan skaleres til andre virksomheder, regioner eller lande uden væsentlige ændringer og forbedringer. Sandsynligheden for en vellykket skalering af sådanne løsninger er fortsat lav, da hver organisation har unikke processer, krav og betingelser, som kan kræve personlige tilpasninger.

Det er vigtigt at forstå, at en vellykket integration af teknologiske løsninger allerede i dag forudsætter en dybt personlig tilgang til hver enkelt proces, projekt og virksomhed. Det betyder, at selv efter at der er udviklet en universel ramme, et værktøj eller et program, vil det kræve detaljeret tilpasning og tilpasning for at opfylde de unikke krav og betingelser for hver enkelt virksomhed og projekt.

> Ifølge PwC-rapporten "Decoding the Fifth Industrial Revolution" [38] [38] er omkring 50% af de øverste ledere i forskellige brancher i år afhængige af integrationen af avanceret teknologi og menneskelig ekspertise. Denne tilgang giver dem mulighed for hurtigt at tilpasse sig ændringer i produktdesign eller kundekrav og skabe personlig produktion.

Hver proces kræver udvikling af en unik funktion eller applikation, hvilket i betragtning af den globale byggeindustris størrelse og projekternes mangfoldighed fører til eksistensen af et stort antal business cases, der hver gang repræsenterer en unik pipeline logik (Fig. 2.1-13). Hver af disse sager har sine egne særegenheder og kræver en skræddersyet tilgang. Vi vil se nærmere på de mange forskellige mulige løsninger på det samme analytiske problem i forbindelse med forskellige tilgange i kapitlet om maskinlæring og analyse af Titanic-datasættet (fig. 9.2-9).

Pipeline er i forbindelse med digitale processer en sekvens af aktiviteter, processer og værktøjer, der muliggør et automatiseret eller struktureret flow af data og arbejde gennem de forskellige faser af projektets livscyklus.

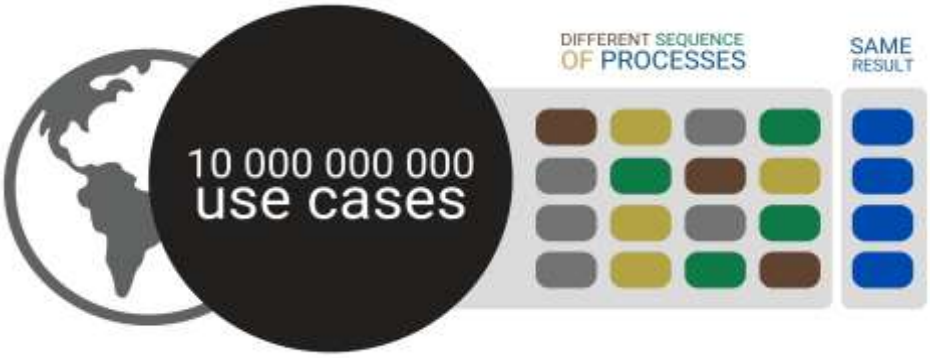


Fig. 2.1-13 Forretningssagernes individualitet og foranderlighed gør det umuligt at skabe skalerbare, lukkede platforme og værktøjer.

Vores liv har allerede ændret sig på mange måder under indflydelse af den digitale transformation, og i dag kan vi tale om indgangen til en ny fase i den økonomiske udvikling af byggebranchen. I denne "nye økonomi" vil konkurrencen blive organiseret efter andre regler: Den, der er i stand til effektivt at omdanne offentlig viden og åbne data til efterspurgte produkter og tjenester, får en afgørende fordel under den femte industrielle revolution.

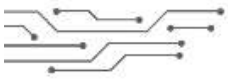
> Som økonomen Kate Maskus bemærker i bogen "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39] 2012 [39], "lever vi i en global vidensøkonomi, og fremtiden tilhører dem, der forstår at gøre videnskabelige opdagelser til varer".

Overgangen til den femte økonomiske tilstand indebærer et skift i fokus fra lukkede it-løsninger til åbne standarder og platforme. Virksomheder vil begynde at bevæge sig væk fra traditionelle softwareprodukter til fordel for serviceorienterede modeller, hvor data snarere end proprietære teknologier bliver det vigtigste aktiv.

> Harvard Business School 2024-undersøgelsen [40] viser den enorme økonomiske værdi af open source software (Open Source Software, OSS). Ifølge undersøgelsen er OSS til stede i 96% af alle softwarekoder, og noget kommercielt software består af 99,9% OSS-komponenter. Uden OSS ville virksomheder bruge 3,5 gange mere på software.

> Opbygning af virksomhedsøkosystemer, der følger de globale tendenser, vil gradvist skifte til et femte økonomisk paradigme, hvor datacentrerede analyser og konsulenttjenester vil blive prioriteret højere end isolerede, lukkede løsninger med fast definerede brugsscenarier.

Digitaliseringens æra vil ændre magtbalancen i branchen: I stedet for at være afhængige af leverandørløsninger vil virksomhederne basere deres konkurrenceevne på deres evne til at bruge data effektivt. Som følge heraf vil byggebranchen bevæge sig fra gamle, stive systemer til fleksible, adaptive økosystemer, hvor åbne standarder og interoperable værktøjer vil være grundlaget for projektstyring. Afslutningen på æraen med applikationsleverandørernes dominans vil skabe et nyt miljø, hvor værdi ikke defineres af besiddelse af lukket kildekode og specialiserede connectorer, men af evnen til at gøre data til en strategisk fordel.



KAPITEL 2.2.
AT GØRE KAOS TIL ORDEN OG REDUCERE KOMPLEKSITETEN

Overflødig kode og lukkede systemer som en barriere for produktivitetsforbedringer

I de seneste årtier har de teknologiske ændringer på it-området primært været drevet af softwareleverandører. De har sat kursen for udviklingen og bestemt, hvilke teknologier virksomhederne skulle tage til sig, og hvilke der skulle lades i stikken. I en tid, hvor man gik fra siloløsninger til centraliserede databaser og integrerede systemer, promoverede leverandørerne licensprodukter, der gav kontrol over adgang og skalerbarhed. Senere, med fremkomsten af cloud-teknologier og SaaS-modeller (Software as a Service), udviklede denne kontrol sig til en abonnementsmodel, der cementerede brugerne som loyale kunder af digitale tjenester.

Denne tilgang har givet anledning til et paradoks: På trods af de hidtil usete mængder af skabt programkode er det kun en lille del af den, der rent faktisk bliver brugt. Måske er der hundreder eller tusinder af gange mere kode end nødvendigt, fordi de samme forretningsprocesser beskrives og duplikeres i snesevis eller hundreder af programmer på forskellige måder, selv inden for den samme virksomhed. Samtidig er der allerede betalt for udviklingsomkostningerne, og de kan ikke hentes hjem igen. Ikke desto mindre fortsætter industrien med at reproducere denne cyklus og skabe nye produkter med minimal merværdi for slutbrugeren, oftere under pres fra markedets forventninger end reelle behov.

> Ifølge Defence Acquisition University (DAU) Software Development Cost Estimating Guide [41] kan omkostningerne til softwareudvikling variere betydeligt afhængigt af flere faktorer, herunder systemets kompleksitet og den valgte teknologi. Historisk set har udviklingsomkostningerne i 2008 ligget på omkring 100 dollars pr. kildekode-linje (SLOC), mens vedligeholdelsesomkostningerne kan stige til 4.000 dollars pr. SLOC.

Bare en af komponenterne i CAD-applikationer - den geometriske kerne - kan have titusindvis af kodelinjer (fig. 6.1-5). En lignende situation ses i ERP-systemer (fig. 5.4-4), hvis kompleksitet vi vender tilbage til i bogens femte del. Et nærmere kig afslører dog, at meget af denne kode ikke tilfører værdi, men blot fungerer som et "postbud" - der mekanisk flytter data mellem databasen, API'en, brugergrænsefladen og andre tabeller i systemet. På trods af den populære myte om den kritiske betydning af såkaldt forretningslogik er den barske virkelighed meget mere prosaisk: Moderne kodebaser er fulde af forældede skabelonblokke (legacy-kode), hvis eneste formål er at sikre dataoverførsel mellem tabeller og komponenter uden at påvirke beslutningstagning eller forretningseffektivitet.

Resultatet er, at lukkede løsninger, der behandler data fra forskellige kilder, uundgåeligt bliver til forvirrende "spaghetti-økosystemer". Disse komplekse, sammenflettede systemer kan kun håndteres af en hær af ledere, der arbejder i en semi-rutinemæssig tilstand. Denne organisering af datahåndtering er ikke kun ineffektiv med hensyn til ressourcer, men skaber også kritiske sårbarheder i forretningsprocesserne og gør virksomheden afhængig af en snæver kreds af specialister, der forstår, hvordan denne teknologiske labyrint fungerer.

Den kontinuerlige stigning i mængden af kode, antallet af applikationer og den stigende kompleksitet i de koncepter, som leverandørerne tilbyder, har ført til et naturligt resultat - en stigning i kompleksiteten af IT-økosystemet i byggeriet. Det har gjort den praktiske implementering af digitalisering ved at øge antallet af

applikationer i branchen ineffektiv. Softwareprodukter, der er skabt uden behørig opmærksomhed på brugernes behov, kræver ofte betydelige ressourcer til implementering og support, men giver ikke det forventede afkast.

> Ifølge McKinseys undersøgelse "Increasing Construction Productivity" [42] har den globale vækst i arbejdsproduktiviteten i byggeriet i løbet af de sidste to årtier i gennemsnit kun været på 1% om året sammenlignet med en vækst på 2,8% for verdensøkonomien som helhed og 3,6% for fremstillingsindustrien. I USA er arbejdsproduktiviteten pr. arbejder i byggeriet halveret siden 1960'erne [43].

Stigende kompleksitet i systemer, isolation og lukkede data har forringet kommunikationen mellem fagfolk, hvilket gør byggebranchen til en af de mindst effektive (figur 2.2-1). til 22 billioner dollars i 2040, hvilket vil kræve betydelige effektivitetsforbedringer.

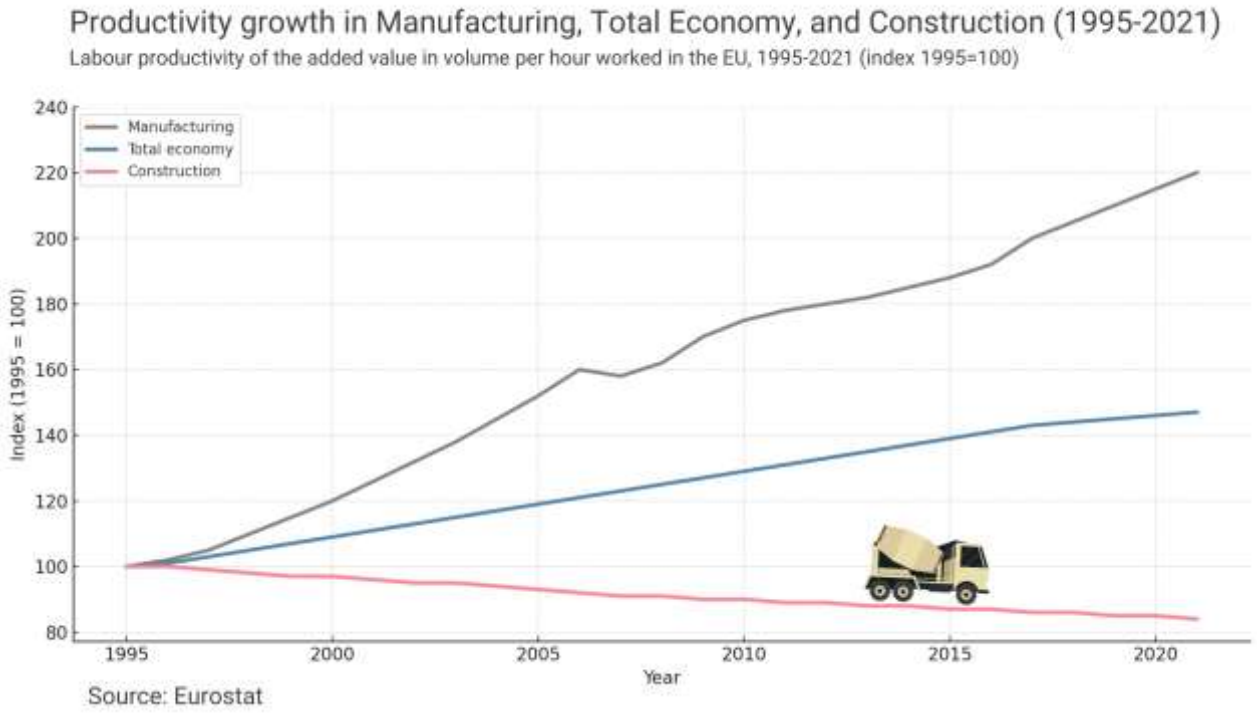


Fig. 2.2-1 Lukkede og komplekse data og som følge heraf dårlig kommunikation mellem specialister gjorde byggebranchen til en af de mindst effektive sektorer i økonomien (baseret på [44], [45]).

> Som det understreges i McKinsey-undersøgelsen (2024) "Ensuring construction productivity is no longer optional", har byggeriet med den stigende ressourceknaphed og branchens ønske om at fordoble sin vækstrate ikke længere råd til at forblive på det nuværende produktivitetsniveau [44]. De globale byggeomkostninger forventes at stige fra 13 billioner dollars i 2023 til langt højere niveauer ved udgangen af årtiet, hvilket gør spørgsmålet om effektivitet ikke bare relevant, men afgørende.

En af de vigtigste måder at forbedre effektiviteten på vil være den uundgåelige forening og forenkling af applikationsstrukturer og dataøkosystemarkitekturer. Denne tilgang til rationalisering vil eliminere overflødige abstraktionslag og unødvendig kompleksitet, som er blevet akkumuleret gennem årene i

virksomhedssystemer.

Fra siloer til et enkelt data warehouse

Jo flere data en organisation akkumulerer, jo sværere bliver det at få reel værdi ud af dem. På grund af den fragmenterede lagring af information i isolerede siloer er moderne virksomheders forretningsprocesser som bygherrer, der forsøger at bygge en skyskraber ud af materialer, der er lagret i tusindvis af forskellige lagre. Overskuddet af information gør det ikke kun vanskeligt at få adgang til juridisk relevant information, men gør også beslutningstagningen langsommere: Hvert trin skal kontrolleres og bekræftes igen og igen.

> Hver opgave eller proces er koblet til en separat tabel eller database, og dataudveksling mellem systemer kræver komplekse integrationer. Fejl og uoverensstemmelser i ét system kan forårsage kædefejl i andre. Forkerte værdier, sene opdateringer og duplikerede oplysninger tvinger medarbejderne til at bruge meget tid på manuelt at afstemme og samkøre data. Resultatet er, at organisationen bruger mere tid på at håndtere konsekvenserne af fragmentering end på at udvikle og optimere processer

Problemet er universelt: Nogle virksomheder kæmper fortsat med kaos, mens andre finder en løsning i integration - at flytte informationsstrømme ind i et centralt lagringssystem. Tænk på det som én stor tabel, hvor du kan gemme alle enheder, der er relateret til opgaver, projekter og objekter. I stedet for dusinvis af forskellige tabeller og formater opstår der et enkelt sammenhængende arkiv (figur 2.2-2), som gør det muligt:

- minimere tab af data;
- eliminere behovet for konstant harmonisering af information;
- forbedre datatilgængelighed og -kvalitet;
- forenkle analytisk behandling og maskinlæring

At bringe data til en fælles standard betyder, at information, uanset kilden, konverteres til et ensartet og maskinlæsbart format. En sådan organisering af data gør det muligt at kontrollere deres integritet, analysere dem i realtid og straks bruge dem til at træffe ledelsesmæssige beslutninger.

Konceptet med integrerede lagringssystemer og deres anvendelse inden for analyse og maskinlæring vil blive diskuteret mere detaljeret i kapitlet "Big Data Storage and Machine Learning". Emnerne datamodellering og -strukturering vil blive dækket i detaljer i kapitlerne "Omdannelse af data til en

struktureret form" og "Hvordan standarder ændrer spillet: fra tilfældige filer til en detaljeret datamodel".

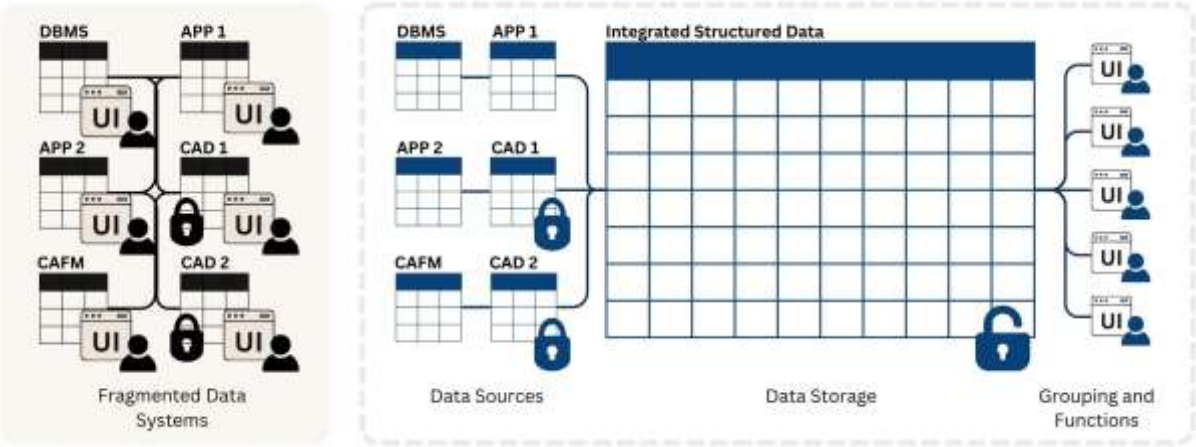


Fig. 2.2-2 Dataintegration eliminerer siloer, forbedrer informationstilgængeligheden og optimerer forretningsprocesserne.

Når dataene er blevet struktureret og flettet, er det næste logiske skridt at validere dem. Med et enkelt integreret datalager er denne proces meget forenklet: Ikke flere forskellige inkonsekvente skemaer, duplikerede strukturer og komplekse relationer mellem tabeller. Alle oplysninger er tilpasset en enkelt datamodel, hvilket eliminerer interne uoverensstemmelser og fremskynder valideringsprocessen. Validering af og sikring af datakvalitet er hjørnesten i alle forretningsprocesser, og vi vil se nærmere på dem i de relevante kapitler i bogen.

I den sidste fase grupperes, filtreres og analyseres dataene. Der anvendes forskellige funktioner på dem: aggregering (addition, multiplikation), beregninger mellem tabeller, kolonner eller rækker (fig. 2.2-4). Arbejdet med data bliver en række trin: indsamling, strukturering, validering, transformation, analytisk behandling og aflæsning til endelige applikationer, hvor oplysningerne bruges til at løse praktiske problemer. Vi vil diskutere mere om opbygning af sådanne scenarier, automatisering af trin og opbygning af behandlingsstrømme i kapitlerne om ETL -processer og datapipeline-tilgang.

Digital transformation handler således ikke kun om at forenkle håndteringen af information. Det handler om at eliminere overdreven kompleksitet i datahåndteringen, at gå fra kaos til forudsigelighed, fra flere systemer til en håndterbar proces. Jo mindre kompleks arkitekturen er, jo mindre kode kræves der for at understøtte den. Og i fremtiden kan kode som sådan forsvinde helt og give plads til intelligente agenter, der uafhængigt analyserer, systematiserer og transformerer data.

Integrerede lagersystemer muliggør overgangen til AI agenter

Jo mindre komplekse data og systemer er, jo mindre kode skal du skrive og vedligeholde. Og den nemmeste måde at spare udvikling på er at slippe helt af med koden og erstatte den med data. Når udviklingen af applikationskoder flyttes fra kode til datamodeller, sker der uundgåeligt et skift til en datacentreret (datadrevet) tilgang, fordi der ligger en helt anden måde at tænke på bag disse koncepter.

> Når man vælger at arbejde med data i centrum, begynder man at se deres rolle på en anden måde. Data er ikke længere bare "råmateriale" til applikationer - det er nu det fundament, som arkitektur, logik og interaktion er bygget op omkring.

Den traditionelle tilgang til datahåndtering starter normalt på applikationsniveau og ligner i konstruktionen et besværligt bureaukratisk system: godkendelser på flere niveauer, manuelle kontroller, endeløse versioner af dokumenter gennem relevante softwareprodukter. Med udviklingen af digitale teknologier vil flere og flere virksomheder blive tvunget til at skifte til princippet om minimalisme - kun at gemme og bruge det, der virkelig er nødvendigt og vil blive brugt.

Logikken om minimering er blevet taget op af leverandørerne. For at forenkle datalagring og -behandling flyttes brugernes arbejde fra offline-applikationer og -værktøjer til cloud-tjenester og såkaldte SaaS -løsninger.

> SaaS-konceptet (Software as a Service, eller "software som en tjeneste") er en af de vigtigste tendenser i moderne it-infrastrukturer, som giver brugerne adgang til applikationer via internettet uden at skulle installere og vedligeholde software på deres egne computere.

På den ene side har SaaSskalering, versionskontrol og reduceret support- og vedligeholdelsesomkostningerne, men på den anden side har det ud over afhængigheden af logikken i en bestemt applikation også gjort brugeren helt afhængig af udbyderens cloud-infrastruktur. Hvis en tjeneste går ned, kan adgangen til data og forretningsprocesser blive midlertidigt eller endda permanent blokeret. Desuden gemmes alle brugerdata, når man arbejder med SaaS-applikationer, på udbyderens servere, hvilket skaber risici for sikkerhed og overholdelse af lovgivningen. Ændringer i takster eller brugsbetingelser kan også resultere i øgede omkostninger eller behov for hurtig migrering.

Udviklingen af AI, LLM -agenter og den datacentrerede tilgang har sat spørgsmålstegn ved fremtiden for applikationer i deres traditionelle form og SaaS -udførelse. Mens applikationer og tjenester tidligere skulle håndtere forretningslogik og behandle data, kan disse funktioner med fremkomsten af AI-agenter skifte til intelligente systemer, der arbejder direkte med data.

Derfor diskuteres hybridarkitekturer i stigende grad i it-afdelinger og på ledelsesniveau, hvor AI -agenter og lokale løsninger supplerer cloud-tjenester og reducerer afhængigheden af SaaS -platforme.

> *Vores tilgang anerkender, at traditionelle forretningsapplikationer eller SaaS -applikationer kan ændre sig dramatisk i agenternes tidsalder. Disse applikationer er i bund og grund CRUD [create, read, update and delete] databaser med forretningslogik. Men i fremtiden vil denne logik blive overtaget af AI-agenter [46].*
>
> - Satya Nadella, administrerende direktør for Microsoft, 2024.

En datacentreret tilgang og brug af AI/LLM-agenter kan reducere overflødige processer og dermed reducere medarbejdernes arbejdsbyrde. Når data er organiseret korrekt, bliver det lettere at analysere, visualisere og

anvende dem til at træffe beslutninger. I stedet for endeløse rapporter og kontroller får specialisterne adgang til opdaterede oplysninger med nogle få klik eller ved hjælp af LLM agenter automatisk i form af færdige dokumenter og dashboards.

Vi vil få hjælp til datamanipulation af værktøjer til kunstig intelligens (AI) og LLM chats. I de senere år har der været en tendens væk fra traditionelle CRUD-operationer (create, read, update, delete) til brug af store sprogmodeller (LLM'er) til datahåndtering. LLM'er er i stand til at fortolke naturligt sprog og automatisk generere passende databaseforespørgsler, hvilket forenkler interaktionen med datastyringssystemer (fig. 2.2-3).

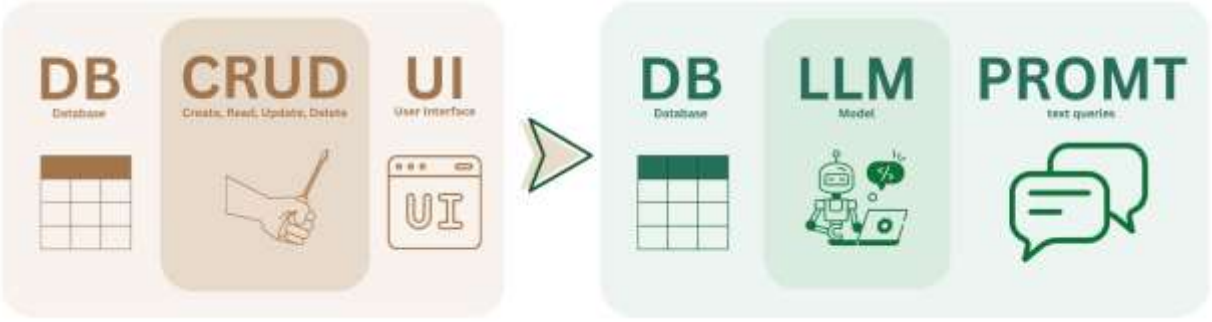


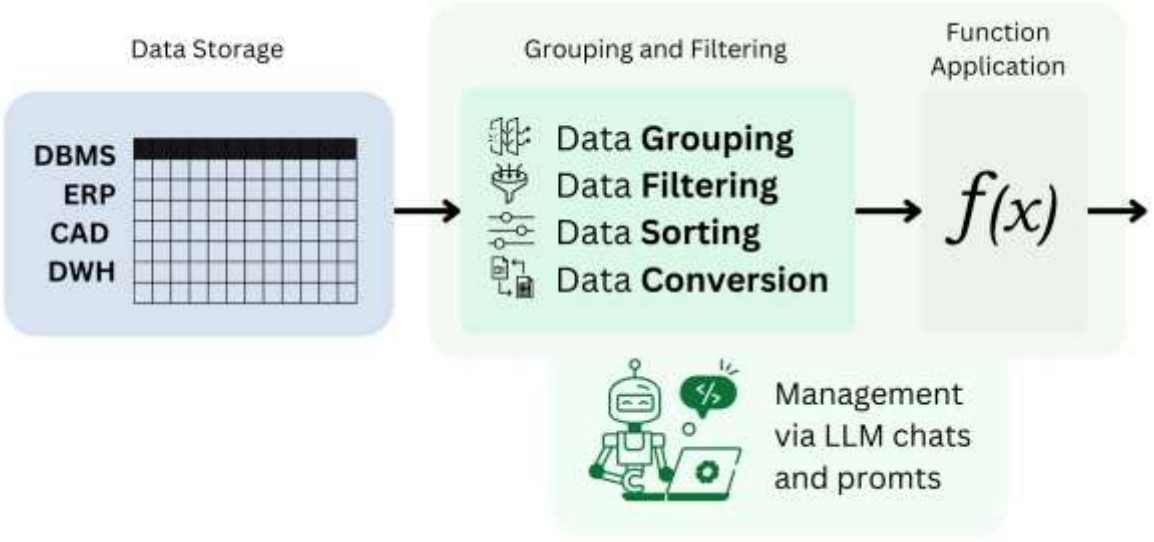
Fig. 2.2-3 AI vil erstatte og integrere storage- og databaseløsninger og gradvist fortrænge traditionelle applikationer og CRUD -operationer.

> *I løbet af de næste 3-6 måneder vil AI skrive 90% af koden, og om 12 måneder kan næsten al koden være genereret af AI [47].*
>
> - Dario Amodei, CEO for LLM Anthropic, marts 2025.

På trods af den hurtige udvikling af AI-udviklingsværktøjer (f.eks. GitHub Copilot) spiller udviklere i 2025 stadig en nøglerolle i denne proces. AI-agenter bliver mere og mere nyttige assistenter: De fortolker automatisk brugerens -forespørgsler, genererer SQL- og Pandas-forespørgsler (mere om dette i de følgende kapitler) eller skriver kode til at analysere data. På den måde erstatter kunstig intelligens gradvist de traditionelle brugergrænseflader i applikationer.

Udbredelsen af modeller for kunstig intelligens, som f.eks. sprogmodeller, vil drive udviklingen af hybridarkitekturer. I stedet for helt at opgive cloud-løsninger og SaaS-produkter vil vi måske se en integration af cloud-tjenester med lokale datahåndteringssystemer. F.eks. muliggør fødereret læring kraftfulde AI modeller uden at skulle flytte følsomme data til skyen. På den måde kan virksomheder bevare kontrollen over

deres data og samtidig få adgang til avancerede teknologier.



Figur 2.2-4 De grundlæggende operationer med gruppering, filtrering og sortering efterfulgt af anvendelse af funktioner vil blive håndteret af LLM chats.

Byggebranchens fremtid vil være baseret på en kombination af lokale løsninger, cloud power og intelligente modeller, der arbejder sammen om at skabe effektive og sikre datahåndteringssystemer. LLM vil gøre det muligt for brugere uden dyb teknisk viden at interagere med databaser og datalagre ved at formulere deres forespørgsler på et naturligt sprog. Vi vil tale mere om LLM og AI agenter, og hvordan de fungerer, i kapitlet "LLM-agenter og strukturerede dataformater".

Korrekt organiserede data og enkle, brugervenlige LLM-analyseværktøjer vil ikke kun gøre det lettere at arbejde med information, men også hjælpe med at minimere fejl, øge effektiviteten og automatisere processer.

Fra dataindsamling til beslutningstagning: vejen til automatisering

I senere dele af bogen vil vi se nærmere på, hvordan specialisterne interagerer med hinanden, og hvordan data bliver grundlaget for beslutningstagning, automatisering og driftseffektivitet. Figur 2.2-5 er et eksempel på et diagram, der viser rækkefølgen af databehandlingstrin i en datacentreret tilgang. Dette diagram illustrerer Continuous Improvement Pipeline), hvis dele vil blive diskuteret i detaljer senere i bogen.

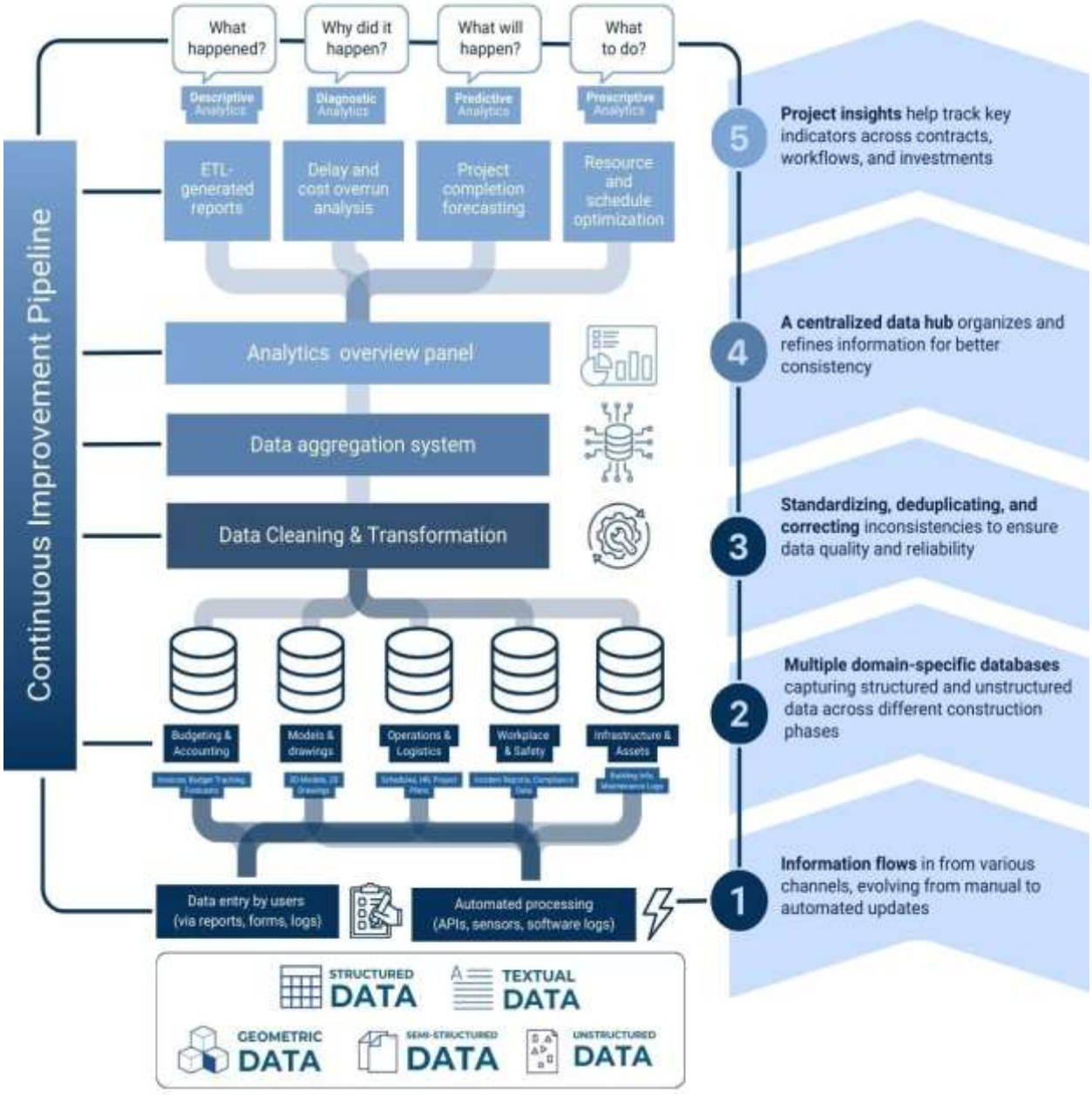


Fig. 2.2-5 Et eksempel på en kontinuerlig dataforbedringspipeline: databehandlings- og analyseflowet i byggeprojekter.

Systemet, der beskriver forretningsprocesserne i en mellemstor virksomhed, er bygget på et princip med flere niveauer. Det omfatter: dataindsamling, rensning, aggregering, analytisk behandling og beslutningstagning baseret på resultaterne. Vi vil studere alle disse faser senere i bogen - både i en teoretisk sammenhæng og gennem praktiske eksempler:

- På det første niveau finder **dataindtastningen** sted (fig. 3.1-1**)**. Information modtages både manuelt (gennem rapporter, formularer, logfiler) og i automatiseret form (fra API, sensorer, softwaresystemer). Data kan have forskellig struktur: geometrisk, tekst, ustruktureret. På dette stadie er der behov for standardisering, strukturering og ensretning af informationsstrømme.

- Det næste niveau er **databehandling og -transformation**. Det omfatter processerne med at rense, fjerne dubletter, rette fejl og forberede information til yderligere analyse (figur 4.2-5). Denne fase er kritisk, fordi kvaliteten af analyserne afhænger direkte af, hvor rene og nøjagtige dataene er.
- **Dataene flyder** derefter **ind i specialiserede tabeller, datarammer eller databaser**, der er opdelt efter funktionsområder: budgettering og regnskab, modeller og tegninger, logistik, sikkerhed og infrastruktur. Denne opdeling gør det nemt at få adgang til og analysere oplysninger på tværs.
- Dataene bliver derefter **samlet og vist i et analytisk dashboard** (showcase). Her anvendes deskriptive, diagnostiske, prædiktive og præskriptive analysemetoder. Det hjælper med at besvare nøglespørgsmål (fig. 1.1-4): Hvad skete der, hvorfor skete det, hvad vil der ske i fremtiden, og hvad skal der gøres? Systemet kan f.eks. identificere forsinkelser, forudsige færdiggørelsen af projekter eller optimere ressourcer.
- Endelig genererer det sidste niveau **analytiske konklusioner og nøgleindikatorer**, der hjælper med at overvåge kontraktopfyldelse, styre investeringer og forbedre forretningsprocesser (fig. 7.4-2). Denne information bliver grundlaget for beslutningstagning og virksomhedens udviklingsstrategi.

På samme måde går data fra at blive indsamlet til at blive brugt i strategisk ledelse. I de følgende dele af bogen vil vi se nærmere på hvert trin med fokus på datatyper, databehandlingsteknikker, analyseværktøjer og eksempler fra det virkelige liv på, hvordan disse tilgange bruges i byggebranchen.

Næste skridt: Gør kaos til et håndterbart system

I denne del undersøgte vi udfordringerne ved informationssiloer og så på, hvordan overdreven systemkompleksitet påvirker virksomhedernes resultater, og vi analyserede overgangen fra den fjerde industrielle revolution til den femte, hvor data snarere end applikationer er i centrum. Vi så, hvordan siloopdelte informationssystemer skaber barrierer for videndeling, og den fortsatte kompleksitet i IT-landskabet reducerer produktiviteten og hæmmer innovation i byggebranchen.

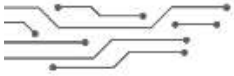
For at opsummere denne del er det værd at fremhæve de vigtigste praktiske trin, der vil hjælpe dig med at anvende de diskuterede tilgange i dine daglige opgaver:

- Visualiser dit informationslandskab
  - [ ] Lav et visuelt kort over de datakilder (Miro, Figma, Canva), du arbejder med regelmæssigt.
  - [ ] Tilføj de systemer og applikationer, du bruger i dit arbejde, til dette kort
  - [ ] Identificer potentielt dobbelt funktionalitet og overflødige løsninger
  - [ ] Identificer kritiske punkter, hvor datatab eller -korruption kan opstå under overførsel mellem systemer.
- Implementer personlig datahåndteringspraksis
  - [ ] Skift fokus fra applikationer til data som et centralt aktiv i processer
  - [ ] Dokumentér datakilder og behandlingsmetoder for at sikre gennemsigtighed
  - [ ] Udvikle mekanismer til at vurdere og forbedre datakvaliteten
  - [ ] Stræb efter at sikre, at data indtastes én gang og bruges igen og igen - det er grundlaget for en effektiv procesorganisation.
- Fremme en datacentreret (datadrevet) tilgang i dit team

- [ ] Foreslå brug af standardiserede og ensartede formater til peer-to-peer-dataudveksling
- [ ] Tag regelmæssigt spørgsmål om datakvalitet og -tilgængelighed op på teammøder
- [ ] Lær Open Source at kende alternativer til de værktøjer, du bruger til at løse dine problemer

Start i det små - vælg en specifik proces eller et datasæt, der er afgørende for dit arbejde, og anvend en datacentreret tilgang til det, så du skifter fokus fra værktøjer til data. Ved at opnå succes med et enkelt pilotprojekt får du ikke kun praktisk erfaring, men også en klar demonstration af fordelene ved den nye metode for dit team. Hvis du har spørgsmål, når du gennemfører de fleste af disse trin, kan du søge afklaring og hjælp hos en opdateret LLM.

I de følgende dele af bogen går vi videre til et mere detaljeret kig på datastrukturerings- og harmoniseringsteknikker og udforsker praktiske tilgange til at integrere heterogen information. Der vil blive lagt særlig vægt på overgangen fra forskellige siloer til samlede dataøkosystemer, som spiller en nøglerolle i den digitale transformation af byggebranchen.





III DEL
DATARAMME I FORRETNINGSPROCESSER I BYGGERIET

I tredje del skabes der en omfattende forståelse af typologien af data i byggeriet og metoder til effektiv organisering af dem. Egenskaberne og detaljerne ved at arbejde med strukturerede, ustrukturerede, semistrukturerede, tekstuelle og geometriske data i forbindelse med byggeprojekter analyseres. Moderne lagringsformater og protokoller til udveksling af information mellem forskellige systemer, der bruges i branchen, gennemgås. Praktiske værktøjer og teknikker til konvertering af data i flere formater til et enkelt struktureret miljø beskrives, herunder hvordan man integrerer CAD (BIM) data. Der foreslås metoder til at sikre kvaliteten af data gennem standardisering og validering, som er afgørende for nøjagtigheden af konstruktionsberegninger. Praktiske aspekter af brugen af moderne teknologier (Python Pandas, LLM -modeller) med kodeeksempler til at løse typiske problemer i byggebranchen analyseres i detaljer. Værdien af at oprette et kompetencecenter (CoE) som en organisatorisk struktur til koordinering og standardisering af tilgange til informationsstyring underbygges.

KAPITEL 3.1.
DATATYPER I BYGGERIET

De vigtigste datatyper i byggebranchen

I den moderne byggebranche fyldes systemer, applikationer og datalagre i virksomheder aktivt med information og data af forskellige typer og formater (Fig. 3.1-1). Lad os se nærmere på de vigtigste datatyper, der udgør informationslandskabet i en moderne virksomhed i byggebranchen:

- **Strukturerede** data: Disse data har en klar organisatorisk struktur, f.eks. Excel-regneark og relationsdatabaser.
- **Ustrukturerede** data: Dette er information, der ikke er organiseret efter strenge regler. Eksempler på sådanne data er tekst, video, fotos og lydoptagelser.
- **Løst strukturerede** data: Disse data indtager en mellemposition mellem strukturerede og ustrukturerede data. De indeholder elementer af struktur, men denne struktur er ikke altid klar eller ofte beskrevet gennem forskellige skemaer. Eksempler på semistrukturerede data i byggeriet er: tekniske specifikationer, projektdokumentation eller statusrapporter.
- **Tekstdata**: omfatter alt, der stammer fra mundtlig og skriftlig kommunikation, f.eks. e-mails, udskrifter af møder og aftaler.
- **Geometriske** data: disse data kommer fra CAD-programmer, hvor specialister skaber geometriske data for projektelementer til visualisering, bekræftelse af volumenværdier eller kollisionskontrol.

Det er vigtigt at bemærke, at geometriske data og tekstdata (alfanumeriske) ikke er en separat kategori, men kan være til stede i alle tre typer data. Geometriske data kan f.eks. være en del af både strukturerede data (parametriske CAD -formater) og ustrukturerede data (indscannede tegninger). Tekstdata kan på samme måde både være organiseret i databaser (strukturerede data) og eksistere som dokumenter uden en klar struktur.

Hver type data i en byggevirksomhed er et unikt element i mosaikken af virksomhedens informationsaktiver. Fra ustrukturerede data som f.eks. billeder fra byggepladser og lydoptagelser af møder til strukturerede optegnelser, herunder tabeller og databaser, spiller hvert element en vigtig rolle i udformningen af virksomhedens informationslandskab.

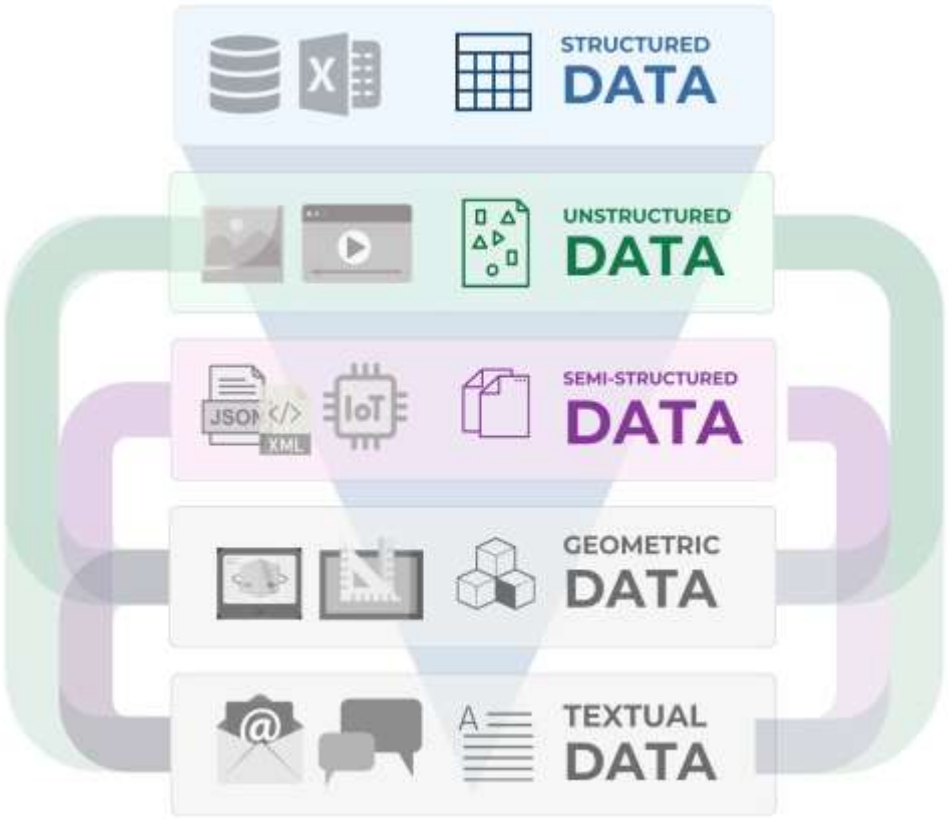


Fig. 3.1-1 Ingeniører og dataansvarlige skal lære at arbejde med alle typer data, der bruges i byggebranchen.

Her er et eksempel på nogle af de systemer og tilhørende datatyper (figur 3.1-2), der bruges i byggeriet:

- **ERP** (Enterprise Resource Planning) - håndterer generelt strukturerede data for at hjælpe med at styre virksomhedens ressourcer og integrere forskellige forretningsprocesser.
- **CAD** (Computer-Aided Design) kombineret med **BIM** (Building Information Modeling) - bruger geometriske og semistrukturerede data til at designe og modellere byggeprojekter, hvilket sikrer nøjagtighed og konsistens i informationerne i designfasen.
- **GIS** (Geographic Information Systems) - arbejder med geometriske og strukturerede data for at skabe og analysere kartografiske data og rumlige relationer.
- **RFID** (Radio-Frequency Identification) - bruger semistrukturerede data til effektivt at spore materialer og udstyr på en byggeplads ved hjælp af radiofrekvensidentifikation.
- **ECM** (Engineering Content Management) er et system til håndtering af tekniske data og dokumentation, herunder semistrukturerede og ustrukturerede data som f.eks. tekniske tegninger og designdokumenter.

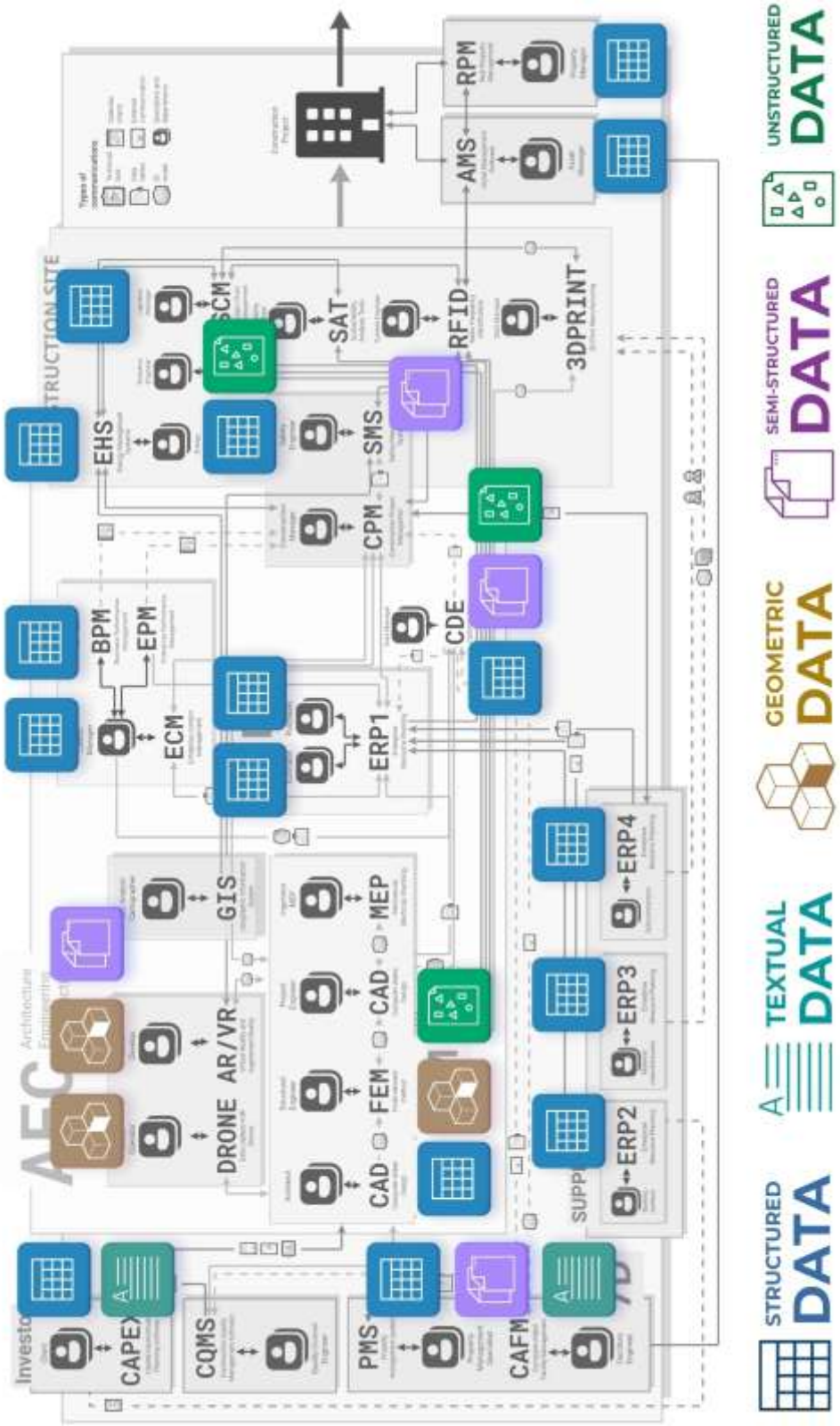


Fig. 3.1-2 Forskellige formater og data fylder forskellige systemer, hvilket kræver oversættelse til en form, der er egnet til kompleks integration.

Disse og mange af virksomhedens andre systemer håndterer en bred vifte af data, fra strukturerede tabeldata til komplekse geometriske modeller, hvilket giver et integreret samarbejde i design-, planlægnings- og byggestyringsprocesserne.

I eksemplet på en forenklet dialog (fig. 3.1-3) udveksles forskellige typer data mellem byggeprojektets specialister:

- **Arkitekt**: "*Under hensyntagen til kundens ønsker har jeg tilføjet et siddeområde på taget. Se venligst det nye design*" (geometriske data - model).
- **Bygningsingeniør**: "*Projektet er modtaget. Jeg er ved at beregne tagets bæreevne for det nye rekreative område*" (strukturerede og semistrukturerede data - beregningstabeller).
- **Indkøbschef**: "*Har brug for specifikationer og mængder af materialer til det rekreative område for at kunne organisere indkøbet*" (tekstuelle og semistrukturerede data - lister og specifikationer).
- **Arbejdsmiljøingeniør**: "*Modtaget data om nyt område. Jeg er ved at vurdere risiciene og opdatere sikkerhedsplanen*" semistrukturerede data - dokumenter og planer).
- **Specialist i BIM** -modellering: "*At foretage ændringer i den overordnede projektmodel for at tilpasse arbejdsdokumentationen*" (geometriske data og semistrukturerede data).
- **Projektleder**: "*Jeg indarbejder den nye rasteplads i arbejdsplanen. Jeg opdaterer tidsplaner og ressourcer i projektstyringssystemet*" (strukturerede og semistrukturerede data - tidsplaner og planer).
- **Specialist i vedligeholdelse af faciliteter (FM)**: "*Jeg forbereder data til fremtidig vedligeholdelse af det rekreative område og lægger dem ind i ejendomsadministrationssystemet*" (strukturerede og semistrukturerede data - instruktioner og vedligeholdelsesplaner).

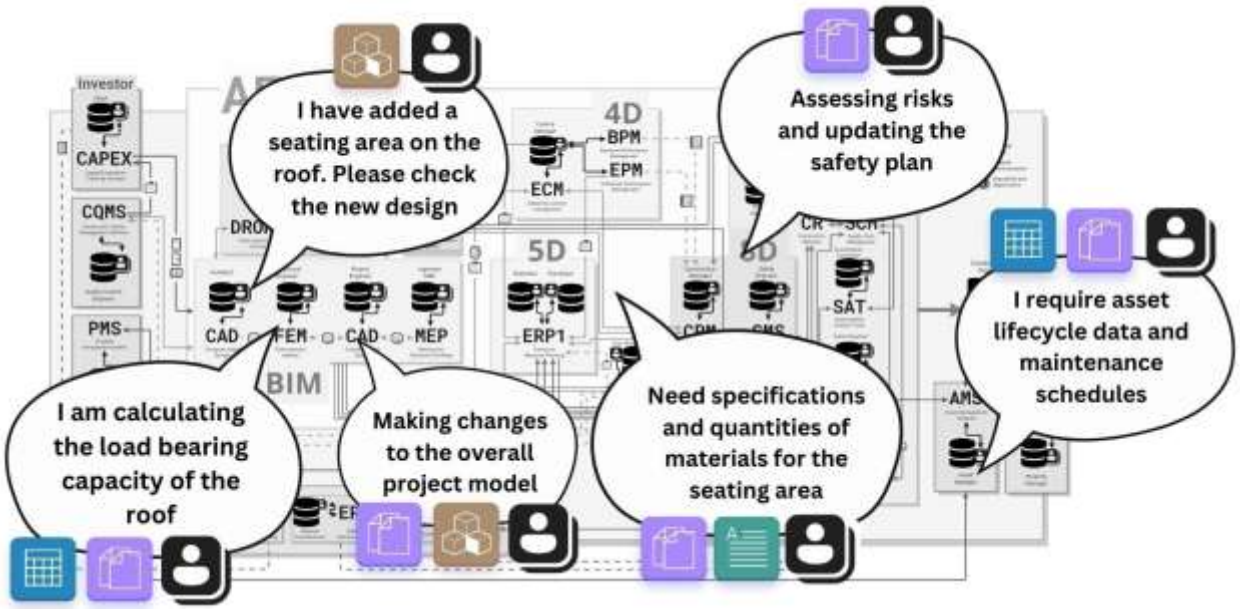


Fig. 3.1-3 Kommunikation mellem specialister foregår både på tekst- og dataniveau.

Alle professionelle arbejder med forskellige typer data for at sikre effektivt teamsamarbejde og projektsucces. Ved at forstå forskellene mellem strukturerede, semistrukturerede og ustrukturerede data kan du genkende den unikke rolle, som hver type spiller i digitale forretningsprocesser. Det er ikke kun vigtigt at vide, at der findes forskellige former for data, men også at forstå, hvordan, hvor og hvorfor de bruges.

> For ikke så længe siden virkede ideen om at kombinere så forskellige data ambitiøs, men svær at realisere. I dag er det allerede en del af den daglige praksis. Integration af data fra forskellige skemaer og strukturer er blevet en integreret del af moderne informationssystemers arkitektur.

I de følgende kapitler vil vi se nærmere på de vigtigste standarder og tilgange, der gør det muligt at kombinere strukturerede, semistrukturerede og ustrukturerede data til en enkelt sammenhængende visning. Der vil blive lagt særlig vægt på strukturerede data og relationsdatabaser som de vigtigste mekanismer til lagring, behandling og analyse af information i byggebranchen.

Strukturerede data

I byggebranchen kommer information fra mange kilder - tegninger, specifikationer, tidsplaner og rapporter. For effektivt at kunne håndtere denne informationsstrøm skal den struktureres. Strukturerede data giver dig mulighed for at organisere information i en praktisk, læsbar og tilgængelig form.

> Ifølge JB Knowledge's 5th Annual Construction Technology Report [17] sporer og evaluerer 67% af de professionelle i byggeprojektledelse arbejdsindsatsen manuelt eller ved hjælp af regneark.

Nogle af de mest almindelige strukturerede dataformater er XLSX og CSV. De bruges i vid udstrækning til at lagre, behandle og analysere oplysninger i regneark. I sådanne regneark præsenteres data i form af rækker og kolonner, hvilket gør dem nemme at læse, redigere og analysere.

XLSX, et format skabt af Microsoft, er baseret på brugen af XML -strukturer og arkiveres ved hjælp af ZIP-algoritmen. De vigtigste funktioner i formatet:

- Understøttelse af komplekse formler, diagrammer og makroer.
- Mulighed for gemme data i forskellige ark samt formatere information.
- Optimeret til Microsoft Excel, men kompatibel med andre kontorpakker.

CSV-format er en ren tekstfil, hvor værdierne er adskilt af kommaer, semikoloner eller andre afgrænsningstegn. Vigtigste fordele:

- Universel kompatibilitet med forskellige programmer og operativsystemer.
- Nem import/eksport til databaser og analysesystemer.
- Nem behandling, selv i tekstredigeringsprogrammer.

CSV understøtter dog ikke formler og formatering, så den vigtigste anvendelse er dataudveksling mellem systemer og masseopdatering af information. På grund af sin alsidighed og platformsuafhængighed er CSV blevet et populært værktøj til dataoverførsel i heterogene IT-miljøer.

De to formater XLSX og CSV fungerer som bindeled mellem forskellige systemer, der håndterer strukturerede data (figur 3.1-4). De er især nyttige til opgaver, hvor læsbarhed manuel redigering og grundlæggende kompatibilitet er vigtig.

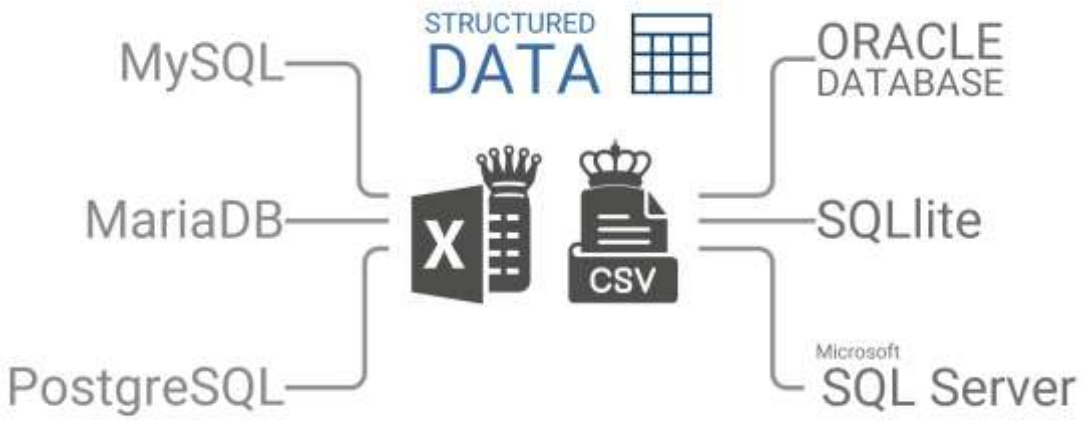


Fig. 3.1-4 Formaterne XLSX og CSV er bindeleddet mellem forskellige systemer, der arbejder med strukturerede data.

> Platformsuafhængighed gør CSV til det mest populære format til dataoverførsel i heterogene IT-miljøer og -systemer.

XLSX og CSV er dog ikke designet til højtydende databehandling eller langtidsopbevaring af store datamængder. Mere moderne strukturerede formater som Apache Parquet, Apache ORC, Feather, HDF5 bruges til sådanne formål. Disse formater vil blive diskuteret mere detaljeret i kapitlet "Storing Big Data: Analysing Popular Formats and Their Effectiveness" del 9 af denne bog.

I praksis bruges Excel med XLSX-format oftere til små opgaver og automatisering af rutineprocesser. Mere komplekse scenarier kræver brug af datastyringssystemer som ERP, PMIS CAFM, CPM, SCM og andre (fig. 3.2-1). Disse systemer lagrer strukturerede data, som organisationen og styringen af en virksomheds informationsstrømme er baseret på.

Moderne informationssystemer til datahåndtering i byggebranchen bygger på strukturerede data, der er organiseret i form af tabeller. Til pålidelig, skalerbar og holistisk styring af store informationsmængder vender applikations- og systemudviklere sig mod relationelle databasestyringssystemer (RDBMS).

Relationelle databaser RDBMS og SQL-forespørgselssprog

For effektivt at kunne lagre, behandle og analysere data er **relationsdatabaser (RDBMS**) datalagringssystemer, der organiserer information i tabeller med definerede relationer mellem dem.

> Data organiseret i databaser (RDBMS) er ikke bare digital information; de er grundlaget for transaktioner og interaktioner mellem forskellige systemer.

Her er nogle af de mest almindelige relationelle databasestyringssystemer (RDBMS) (figur 3.1-5):

- **MySQL** (Open Source) er en af de mest populære RDBMS'er, som er en del af LAMP-stakken

(Linux, Apache, MySQL, PHP /Perl/Python). Det bruges i vid udstrækning til webudvikling på grund af dets enkelhed og høje ydeevne.

- **PostgreSQL** (Open Source) er et kraftfuldt objektrelationelt system, der er kendt for sin pålidelighed og avancerede funktioner. Det er velegnet til komplekse virksomhedsløsninger.
- **Microsoft SQL Server** er et kommercielt system fra Microsoft, som er meget udbredt i virksomhedsmiljøer på grund af dets integration med andre virksomhedsprodukter og høje sikkerhedsniveau.
- **Oracle Database** er en af de mest kraftfulde og pålidelige DBMS'er, der bruges i store virksomheder og missionskritiske applikationer.
- **IBM DB2** - rettet mod store virksomheder og giver høj ydeevne og fejltolerance.
- **SQLite** (Open Source) er en let, indlejret database, der er ideel til mobile applikationer og selvstændige systemer som CAD-designsoftware (BIM).

Populære databasestyringssystemer i byggebranchen - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 og SQLite - arbejder med strukturerede data. Alle disse DBMS'er er kraftfulde og fleksible løsninger til styring af en lang række forretningsprocesser og applikationer, fra små hjemmesider til store virksomhedssystemer (fig. 3.2-1).

> Ifølge Statista [48] udgør relationelle databasestyringssystemer (RDBMS) omkring 72% af de samlede DBMS'er, der er i brug i 2022.



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

Fig. 3.1-5 Populariteten af at bruge strukturerede databaser (markeret med blåt) i DBMS-rangordningen (baseret på [49]).

Det er ret nemt at installere open source-databaser - selv uden omfattende teknisk viden. Open source-systemer, såsom PostgreSQL, MySQL eller SQLite, er gratis tilgængelige og fungerer på de fleste operativsystemer: Windows, macOS og Linux. Alt, hvad du behøver, er at gå til projektets officielle hjemmeside, downloade installationsprogrammet og følge instruktionerne. I de fleste tilfælde tager installationen ikke mere end 10-15 minutter. Vi vil modellere og oprette en sådan database i bogens fjerde del (Fig. 4.3-8).

Hvis din virksomhed bruger cloud-tjenester (f.eks. Amazon Web Services, Google Cloud eller Microsoft Azure), kan du implementere databasen med et par klik - platformen tilbyder dig færdige skabeloner til installation. Takket være kodens åbenhed er sådanne databaser nemme at tilpasse til dine opgaver, og et stort fællesskab af brugere vil altid hjælpe dig med at finde en løsning på ethvert problem.

RDBMS er fortsat grundlaget for en lang række forretningsapplikationer og analyseplatforme (figur 3.1-6), som gør det muligt for virksomheder at lagre, behandle og analysere data effektivt - og dermed træffe informerede og rettidige beslutninger.

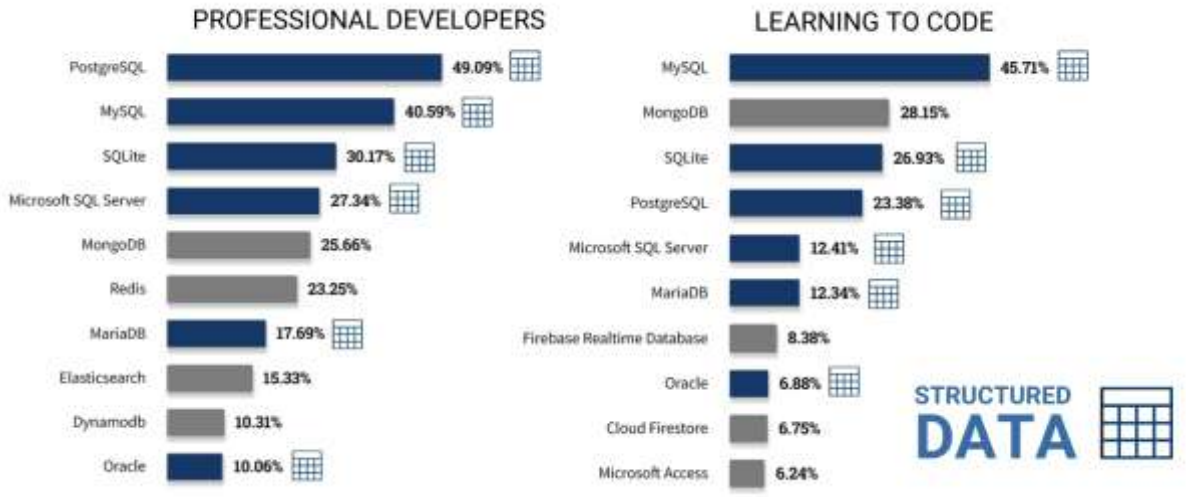


Fig. 3.1-6 Undersøgelse af udviklere på StackOverFlow (det største IT-forum) om, hvilke databaser de brugte sidste år, og hvilke de vil bruge næste år (RDBMS'er er markeret med blåt) (baseret på [50]).

RDBMS giver pålidelighed, datakonsistens, transaktionsstøtte og bruger et stærkt forespørgselssprog - SQL (Structured Query Language), som ofte bruges i analyser og giver dig mulighed for nemt at indhente, ændre og analysere oplysninger, der er gemt i databaser. SQL er det vigtigste værktøj til at arbejde med data i relationelle systemer.

SQL - forespørgsler i databaser og nye tendenser

Den største fordel ved SQL-sproget, der ofte bruges i relationsdatabaser, i forhold til andre typer informationsstyring (f.eks. ved hjælp af klassiske Excel-regneark) er understøttelsen af meget store mængder databaser med høj hastighed i behandlingen af forespørgsler.

> Structured Query Language (SQL) er et specialiseret programmeringssprog, der er designet til at lagre, behandle og analysere information i relationsdatabaser. SQL bruges til at oprette, administrere og få adgang til data, så man effektivt kan finde, filtrere, kombinere og aggregere information. Det er et vigtigt værktøj til at få adgang til data og giver en praktisk og formaliseret måde at interagere med informationslagre på.

Udviklingen af SEQUEL-SQL systemer går gennem vigtige produkter og virksomheder som Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL og MySQL, og kulminerer i fremkomsten af SQLite og MariaDB [51]. SQL giver regnearksfunktioner, der ikke findes i Excel, hvilket gør datamanipulation mere skalerbar, sikker og nem at automatisere:

- **Oprettelse og styring af datastrukturer (DDL):** I SQL kan man oprette, ændre og slette tabeller i en database, etablere links mellem dem og definere datalagringsstrukturer. Excel arbejder derimod med faste ark og celler uden klart definerede relationer mellem ark og datasæt.
- **Datamanipulation (DML):** SQL gør det muligt at tilføje, ændre, slette og hente data i stort omfang

og med høj hastighed ved at udføre komplekse forespørgsler med filtrering, sortering og sammenkædning af tabeller (figur 3.1-7). I Excel kræver behandling af store mængder information manuelle handlinger eller særlige makroer, hvilket gør processen langsommere og øger sandsynligheden for fejl.

- **Adgangskontrol (DCL):** SQL giver dig mulighed for at differentiere adgangsrettigheder til data for forskellige brugere, hvilket begrænser muligheden for at redigere eller se oplysninger. I Excel er adgangen derimod enten delt (når man overfører en fil) eller kræver komplekse indstillinger med deling af tilladelser via cloud-tjenester.

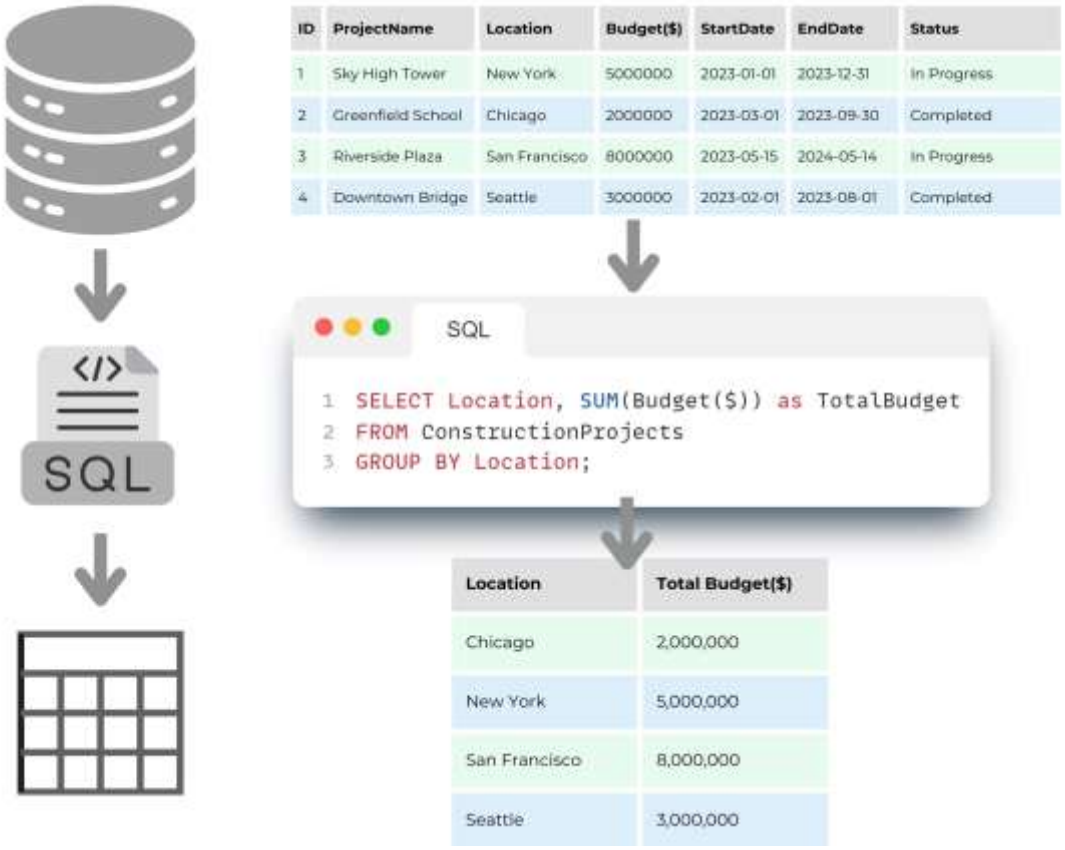


Fig. 3.1-7 Eksempel på DML i SQL: hurtig behandling, gruppering og aggregering med få kodelinjer til automatisk databehandling.

Excel gør det nemt at arbejde med data på grund af sin visuelle og intuitive struktur. Men når mængden af data stiger, falder Excels ydeevne. Excel har også grænser for, hvor mange data det kan gemme - maksimalt en million rækker - og ydeevnen forringes, længe før denne grænse nås. Så mens Excel ser ud til at være at foretrække til at visualisere og manipulere små mængder data, er SQL bedre egnet til at håndtere store datasæt.

Det næste trin i udviklingen af strukturerede data var fremkomsten af kolonnedatabaser (Columnar Databases), som er et alternativ til traditionelle relationsdatabaser, især når det drejer sig om betydeligt større datamængder og analytiske beregninger. I modsætning til rækkedatabaser, hvor data gemmes linje for linje, registrerer kolonnedatabaser information efter kolonne. Sammenlignet med klassiske databaser giver dette mulighed for:

- Reducer lagerpladsen ved effektivt at komprimere ensartede data i kolonner.
- Hurtigere analytiske forespørgsler, da det kun er de nødvendige kolonner, der læses, og ikke hele tabellen.
- Optimer Big Data og data warehousing, f.eks. Data Lakehouse Architecture.

Vi vil tale mere om kolonnedatabaser, Pandas DataFrame, Apache Parquet, HDF5, samt om oprettelse af Big Data -lagre baseret på dem til dataanalyse og -behandling i de følgende kapitler i denne bog - "Data-Frame: et universelt tabulært dataformat" og "Datalagringsformater og arbejde med Apache Parquet: DWH -datalagre og Data Lakehouse-arkitektur ".

Ustrukturerede data

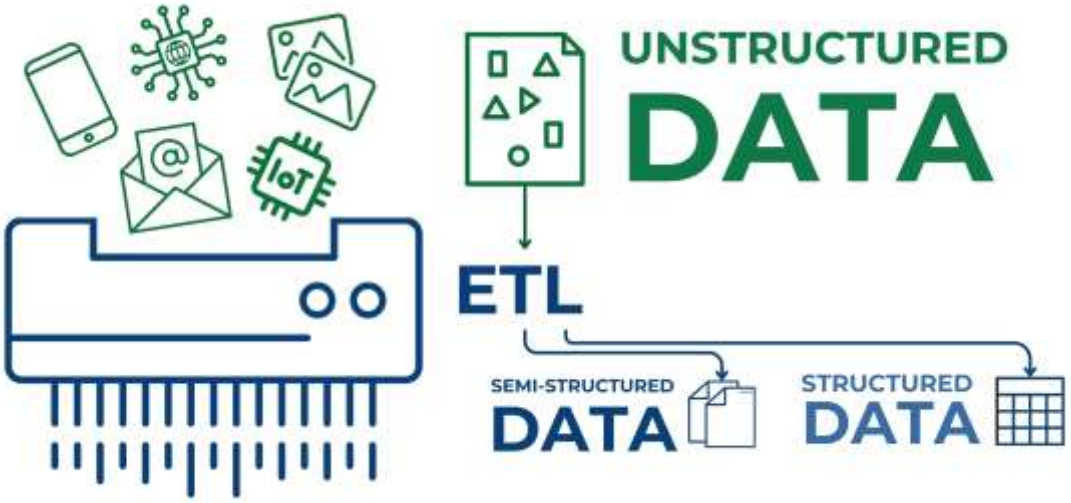
Selvom de fleste data, der bruges i applikationer og informationssystemer, er i struktureret form, er de fleste oplysninger, der genereres i byggeriet, i form af ustrukturerede data - billeder, videoer, tekstdokumenter, lydoptagelser og andre former for indhold. Det gælder især i bygge- og anlægsfasen, i driften og i det tekniske tilsyn, hvor visuel og tekstlig information dominerer.

> **Ustrukturerede data** er information, der ikke har nogen foruddefineret model eller struktur, og som ikke er organiseret i traditionelle rækker og kolonner som i databaser eller tabeller.

Generelt kan ustrukturerede data klassificeres i to kategorier:

- Menneskeskabte ustrukturerede data, som omfatter forskellige typer menneskeskabt indhold: tekstdokumenter, e-mails, billeder, videoer og så videre.
- Maskingenererede ustrukturerede data skabes af enheder og sensorer: de omfatter logfiler, GPS-data, Internet of Things-resultater (IoT) og andre telemetrioplysninger fra f.eks. en byggeplads.

I modsætning til strukturerede data, som nemt kan organiseres i tabeller og databaser, kræver ustrukturerede data yderligere behandlingstrin, før de kan integreres i informationssystemer (fig. 3.1-8). Brugen af teknologier til automatiseret indsamling, analyse og omdannelse af sådanne data åbner nye muligheder for at forbedre byggeriets effektivitet, reducere fejl og minimere den menneskelige faktors indflydelse.



Figur 3.1-8 Behandlingen af ustrukturerede data begynder med, at de omdannes til halvstrukturerede og strukturerede data.

Ustrukturerede data udgør op til 80% af al information [52], som fagfolk i virksomheder støder på, så vi vil diskutere deres typer og behandling i detaljer med eksempler i de følgende kapitler i bogen.

For at lette diskussionen er tekstdata kategoriseret separat. Selvom de er en type ret ustrukturerede data, kræver deres betydning og udbredelse i byggebranchen særlig opmærksomhed.

Tekstdata: mellem ustruktureret kaos og struktureret kaos y

Tekstdata i byggebranchen dækker over en bred vifte af formater og informationstyper, fra papirdokumenter til uformelle kommunikationsmetoder som breve, samtaler, arbejdskorrespondance og mundtlige møder på byggepladsen. Alle disse tekstdata indeholder vigtige oplysninger til styring af byggeprojekter, fra detaljer om designbeslutninger og ændringer i planer til diskussioner om sikkerhedsspørgsmål og forhandlinger med entreprenører og kunder (figur 3.1-9).



Fig. 3.1-9 Tekstdata, en af de mest populære informationstyper, der bruges i kommunikationen mellem projektdeltagere.

Tekstinformation kan være både formaliseret og ustruktureret. Formaliserede data omfatter Word-doku-

menter (.doc,.docx), PDF samt tekstfiler med mødereferater (.txt). Uformaliserede data omfatter messenger- og e-mailkorrespondance, mødeudskrifter (Teams, Zoom, Google Meet) og lydoptagelser af diskussioner (.mp3,.wav), der skal konverteres til tekst.

Men mens skriftlige dokumenter som formelle anmodninger, kontraktbetingelser og e-mails normalt allerede har en vis struktur, forbliver mundtlig kommunikation og arbejdskorrespondance ofte ustruktureret, hvilket gør det vanskeligt at analysere og integrere dem i projektstyringssystemer.

Nøglen til effektiv håndtering af tekstdata er at konvertere dem til et struktureret format. På den måde kan de behandlede oplysninger automatisk integreres i eksisterende systemer, der allerede arbejder med strukturerede data.

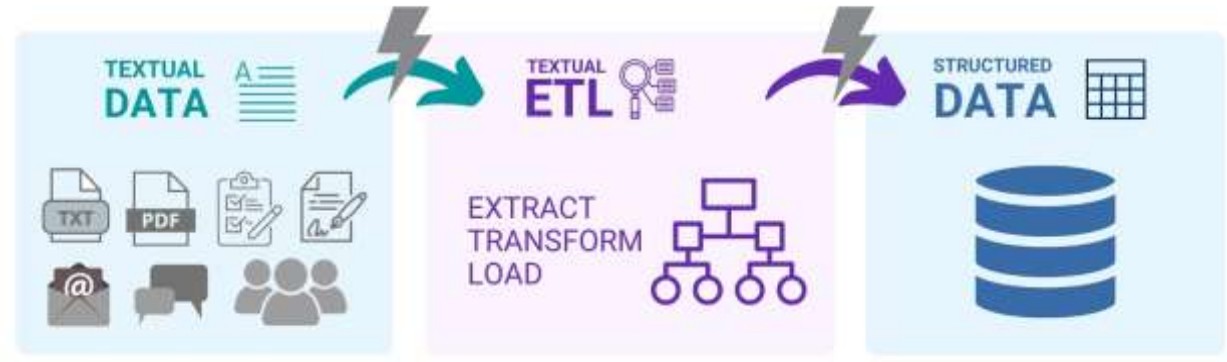


Fig. 3.1-10 Konvertering af tekstindhold til strukturerede data.

For at gøre effektiv brug af tekstinformation skal den automatisk konverteres til en struktureret form (figur 3.1-10). Denne proces involverer normalt flere trin:

- **Tekstgenkendelse (OCR**) - konvertering af billeder af dokumenter og tegninger til et maskinlæsbart format.
- **Tekstanalyse (NLP**) - automatisk identifikation af nøgleparametre (datoer, beløb og tal, der er relevante for projektet).
- **Dataklassificering** - kategorisering af information (økonomi, logistik, risikostyring).

Efter genkendelse og klassificering kan de allerede strukturerede data integreres i databaser og bruges i automatiserede rapporterings- og styringssystemer.

Semistrukturerede og løst strukturerede data

Semistrukturerede data indeholder en vis grad af organisering, men har ikke et strengt skema eller en streng struktur. Selvom sådanne oplysninger indeholder strukturerede elementer (f.eks. datoer, medarbejdernavne og lister over udførte opgaver), kan præsentationsformatet variere betydeligt fra projekt til projekt eller endda fra en medarbejder til en anden. Eksempler på sådanne data er tidslogge, statusrapporter og tidsplaner, som kan præsenteres i mange forskellige formater.

Semistrukturerede data er lettere at analysere end ustrukturerede data, men kræver yderligere behandling for at kunne integreres i standardiserede projektstyringssystemer.

At arbejde med semistrukturerede data, der er kendetegnet ved en konstant skiftende struktur, giver betydelige udfordringer. Det skyldes, at variationen i datastrukturen kræver separate, individuelle tilgange til behandling og analyse af hver enkelt kilde til semistrukturerede data.

Men mens det kræver en stor indsats at håndtere ustrukturerede data, kan behandling af semistrukturerede data gøres med relativt enkle metoder og værktøjer.

Svagt strukturerede data er et mere generelt begreb, der beskriver data med minimal eller ufuldstændig struktur. Oftest er det tekstdokumenter, chats, e-mails, hvor der findes nogle metadata (f.eks. dato, afsender), men de fleste oplysninger er præsenteret på en kaotisk måde.

I byggeriet findes løst strukturerede data i en række forskellige processer. De kan for eksempel omfatte:

- Overslag og tilbud - tabeller med data om materialer, mængder og omkostninger, men uden et ensartet format.
- Tegninger og tekniske diagrammer - filer i PDF eller DWG, der indeholder tekstkommentarer og metadata, men uden en strengt fast struktur.
- Arbejdsplaner - data fra MS Project, Primavera P6 eller andre systemer, som kan have en anden eksportstruktur.
- CAD (BIM -modeller) - indeholder elementer af strukturen, men datarepræsentationen afhænger af softwaren og projektstandarden.

Geometriske data, der produceres af CAD-systemer, kan kategoriseres på samme måde som semistrukturerede data. Vi vil dog klassificere geometriske CAD-data (BIM) som en separat datatype, fordi de, ligesom tekstdata, ofte kan behandles som en separat datatype i virksomhedens processer.

Geometriske data og deres anvendelse

Mens metadata om projektelementer næsten altid lagres i form af tabeller, strukturerede eller løst strukturerede formater, oprettes geometriske data om projektelementer i de fleste tilfælde ved hjælp af særlige CAD værktøjer (Fig. 3.1-11), der gør det muligt at visualisere projektelementer i detaljer som et sæt linjer (2D) eller geometriske legemer (3D).



Fig. 3.1-11 CAD værktøjer har hjulpet med at flytte geometrisk information fra fysiske medier til databaseform.

Når man arbejder med geometriske data inden for byggeri og arkitektur, kan man identificere tre hovedanvendelser af geometriske data (figur 3.1-12):

- **Bekræftelse af volumener**: Geometriske data, der genereres i CAD-programmer (BIM) ved hjælp af særlige geometriske kerner, er nødvendige for automatisk og nøjagtigt at bestemme volumener og dimensioner af projektelementer. Disse data omfatter automatisk beregnede arealer, volumener, længder og andre vigtige egenskaber, der er nødvendige for planlægning, budgettering og bestilling af ressourcer og materialer
- **Visualisering af projektet**: I tilfælde af ændringer i projektet giver visualiseringen af elementer mulighed for automatisk at generere opdaterede tegninger i forskellige planer. Visualisering af projektet i de indledende faser giver mulighed for hurtigere gensidig forståelse mellem alle deltagere for at spare tid og ressourcer under byggeprocessen.
- **Kontrol af kollisioner**: I komplekse bygge- og anlægsprojekter, hvor samspillet mellem flere kategorier af elementer (f.eks. rør og vægge) uden "geometriske konflikter" er kritisk, spiller kollisionskontrol en vigtig rolle. Ved at bruge collision detection software kan du proaktivt identificere potentielle geometriske konflikter mellem projektelementer og forhindre dyre fejl under byggeprocessen.

Helt fra begyndelsen af ingeniørkontorer, fra tidspunktet for opførelsen af de første komplekse strukturer, leverede bygningsingeniører geometriske oplysninger i form af tegninger, linjer og flade geometriske elementer (på papyrus, "A0"-papir eller i DWG, PDF, PLT-formater), på grundlag af hvilke formænd og estimatorer (Fig. 3.1-11), i de sidste årtusinder, ved hjælp af linealer og vinkelmålere indsamlede attributive volumener eller antal elementer og grupper af elementer. 3.1-11), i de sidste årtusinder ved hjælp af linealer og transportmidler indsamlede attributive volumener eller mængder af elementer og grupper af elementer.

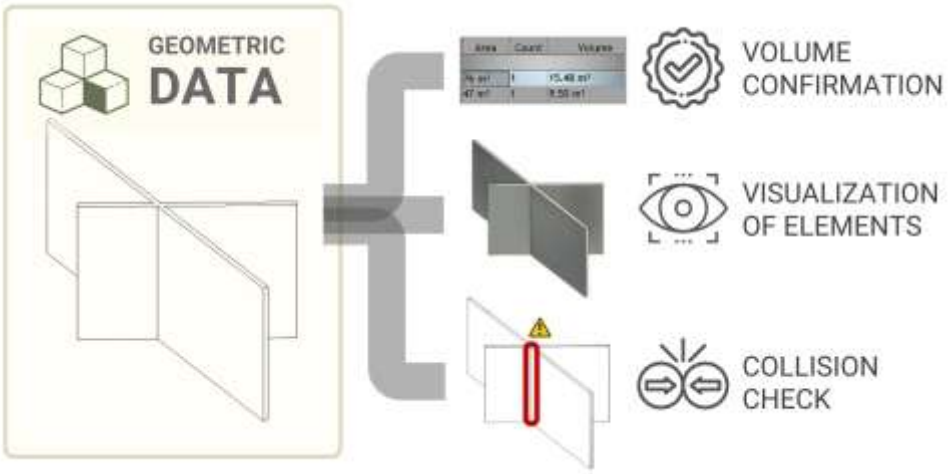


Fig. 3.1-12 Geometri er grundlaget for at opnå volumetriske parametre for elementer, som derefter bruges til at beregne projektets omkostninger og timing.

I dag løses denne manuelle og tidskrævende opgave ved fuld automatisering takket være fremkomsten af volumetrisk modellering i moderne CAD-værktøjer (BIM), som gør det muligt automatisk ved hjælp af en særlig geometrisk kerne at opnå volumetriske attributter for ethvert element uden behov for at beregne volumetriske parametre manuelt.

Moderne CAD-værktøjer giver dig også mulighed for at klassificere og kategorisere projektelementer, så du kan uploade specifikationstabeller fra projektdatabasen til brug i forskellige systemer, f.eks. omkostningsestimering, planlægning eller $CO_{(2)}$-beregning (fig. 3.1-13). Vi diskuterer Indhentning af specifikationer, QTO-tabeller og mængder samt praktiske eksempler i kapitlet "Indhentning af mængder og kvantificering".

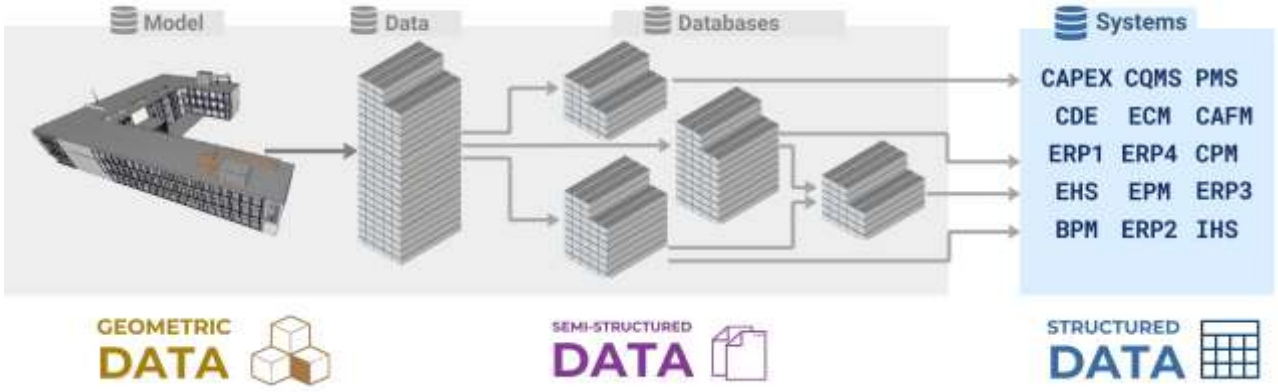


Fig. 3.1-13 CAD-værktøjer (BIM) gemmer data i databaser, der er designet til at integrere og interagere med andre systemer.

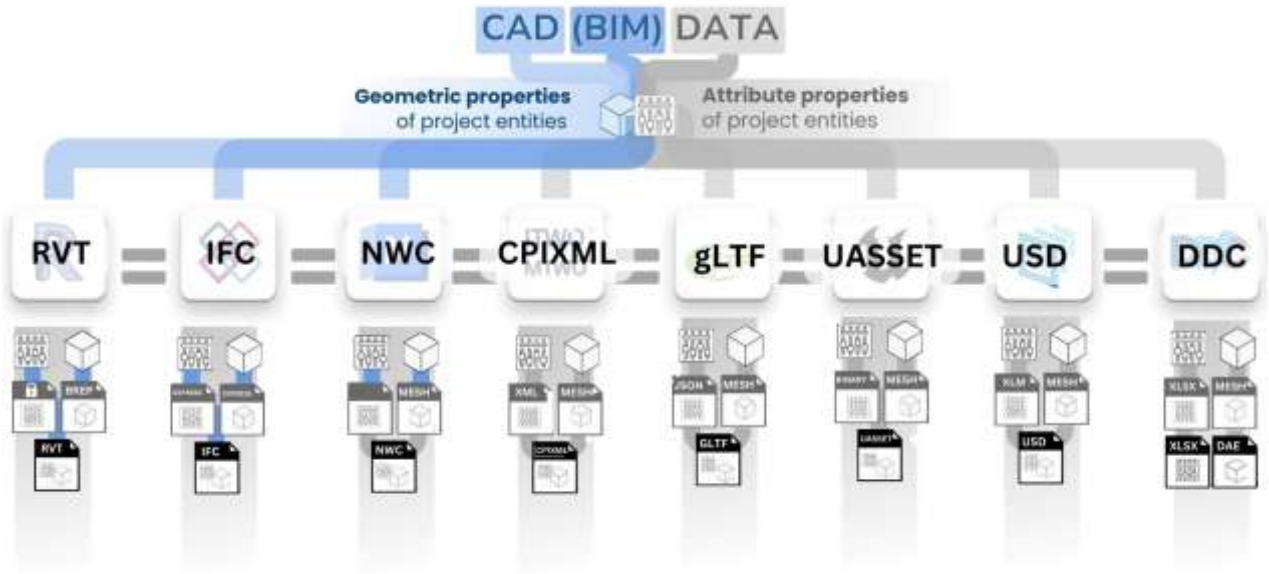
På grund af den lukkede karakter af databaser og formater, der bruges i CAD-miljøet, er geometriske data, der skabes i CAD-løsninger, faktisk blevet en separat type information. De kombinerer både elementernes geometri og metainformation (struktureret eller semistruktureret), der er indkapslet i specialiserede filer og formater.

CAD data: fra design til datalagring

Moderne CAD- og BIM-systemer gemmer data i deres egne, ofte proprietære formater: DWG, DXF, RVT, DGN, PLN og andre. Disse formater understøtter både 2D- og 3D-repræsentationer af objekter og bevarer ikke kun geometrien, men også de attributter, der er forbundet med objekterne. Her er de mest almindelige:

- **DWG** er et binært filformat, der bruges til at gemme todimensionelle (og mindre almindeligt tredimensionelle) designdata og metadata.
- **DXF** er et tekstformat til udveksling af 2D- og 3D -tegninger mellem CAD -systemer. Det indeholder geometri, lag og attributdata og understøtter både ASCII og binær repræsentation.
- **RVT** er et binært format til lagring af CAD modeller, herunder 3D -geometri, elementattributter, relationer og designparametre.
- **IFC** er et åbent tekstformat til udveksling af konstruktionsdata mellem CAD (BIM) systemer. Det omfatter geometri, objektegenskaber og oplysninger om deres relationer.

Ud over disse anvendes andre formater: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Selv om de har forskellige formål og er forskellige i graden af åbenhed (figur 3.1-14), kan de alle repræsentere den samme projektinformationsmodel i forskellige former. I komplekse projekter bruges disse formater ofte parallelt, fra tegning til koordinering af projektmodeller.



Figur 3.1-14 Populære CAD-lagringsformater beskriver geometri gennem BREP- eller MESH-parametre, suppleret med attributdata.

Alle ovenstående formater gør det muligt at gemme data om hvert enkelt element i et byggeprojekt, og alle ovenstående formater indeholder to vigtige typer data:

- **Geometriske parametre** - beskriver et objekts form, placering og dimensioner. Geometri og brugen af den vil blive diskuteret i detaljer i den sjette del af bogen, der er dedikeret til CAD-løsninger (BIM);
- **Attributegenskaber** - indeholder forskellige oplysninger: materialer, elementtyper, tekniske egenskaber, unikke identifikatorer og andre egenskaber, som projektelementer kan have.

Attributdata er særligt vigtige i moderne projekter, da de definerer objekternes driftsegenskaber, giver

mulighed for ingeniør- og omkostningsberegninger og giver mulighed for end-to-end-interaktion mellem deltagere i design, konstruktion og drift. For eksempel:

- For vinduer og døre: konstruktionstype, rudetype, åbningsretning (fig. 3.2-1).
- For vægge registreres oplysninger om materialer, varmeisolering og akustisk ydeevne.
- For tekniske systemer gemmes parametre for rørledninger, kanaler, kabelruter og deres forbindelser.

Disse parametre kan lagres både i selve CAD-(BIM -)filerne og i eksterne databaser - som et resultat af eksport, konvertering eller direkte adgang til interne CAD-strukturer via reverse engineering-værktøjer. Denne tilgang gør det lettere at integrere designinformation med andre virksomhedssystemer og -platforme

> **Reverse engineering i** forbindelse med CAD (BIM) er processen med at udtrække og analysere den interne struktur i en digital model for at genskabe dens logik, datastruktur og afhængigheder uden adgang til de oprindelige algoritmer eller dokumentation.

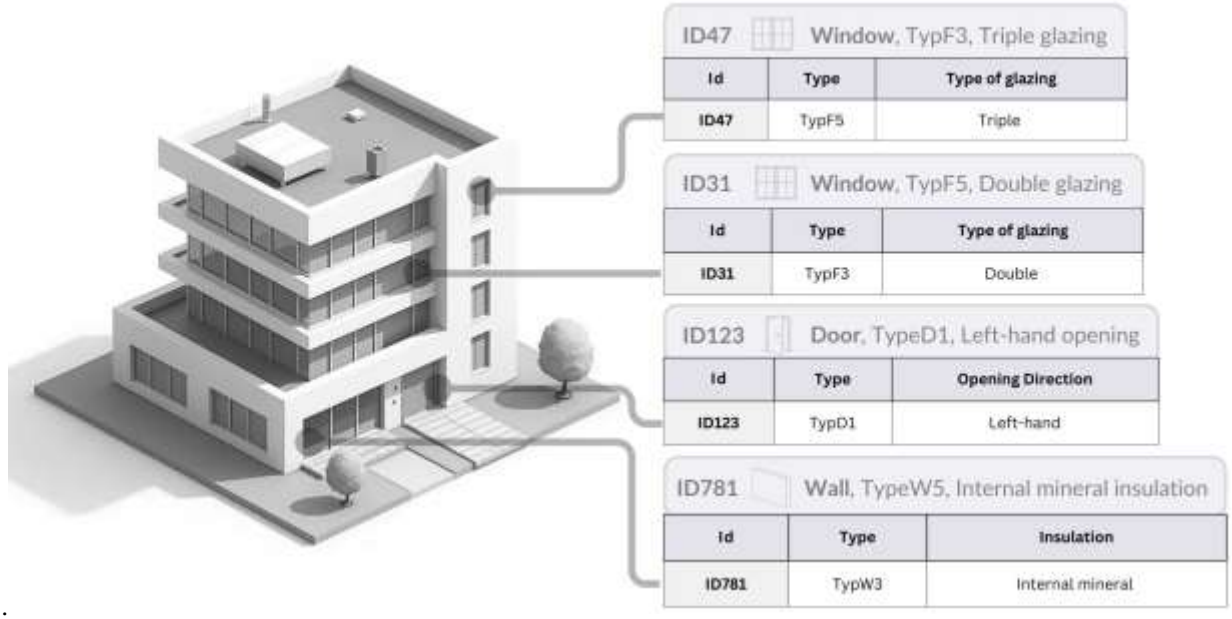


Fig. 3.1-15 Et projektelement indeholder, ud over at beskrive parametrisk eller polygonal geometri, information om parametre og egenskaber for elementer.

Resultatet er, at der dannes et unikt sæt parametre og egenskaber omkring hvert element, herunder både unikke karakteristika for hvert objekt (f.eks. identifikator og dimensioner) og fælles attributter for grupper af elementer. Det gør det muligt ikke blot at analysere de enkelte elementer og enheder i projektet, men også at kombinere dem i logiske grupper, som derefter kan bruges af andre specialister til deres opgaver og beregninger i systemer og databaser.

> En entitet er et konkret eller abstrakt objekt i den virkelige verden, som kan identificeres, beskrives og repræsenteres entydigt i form af data.

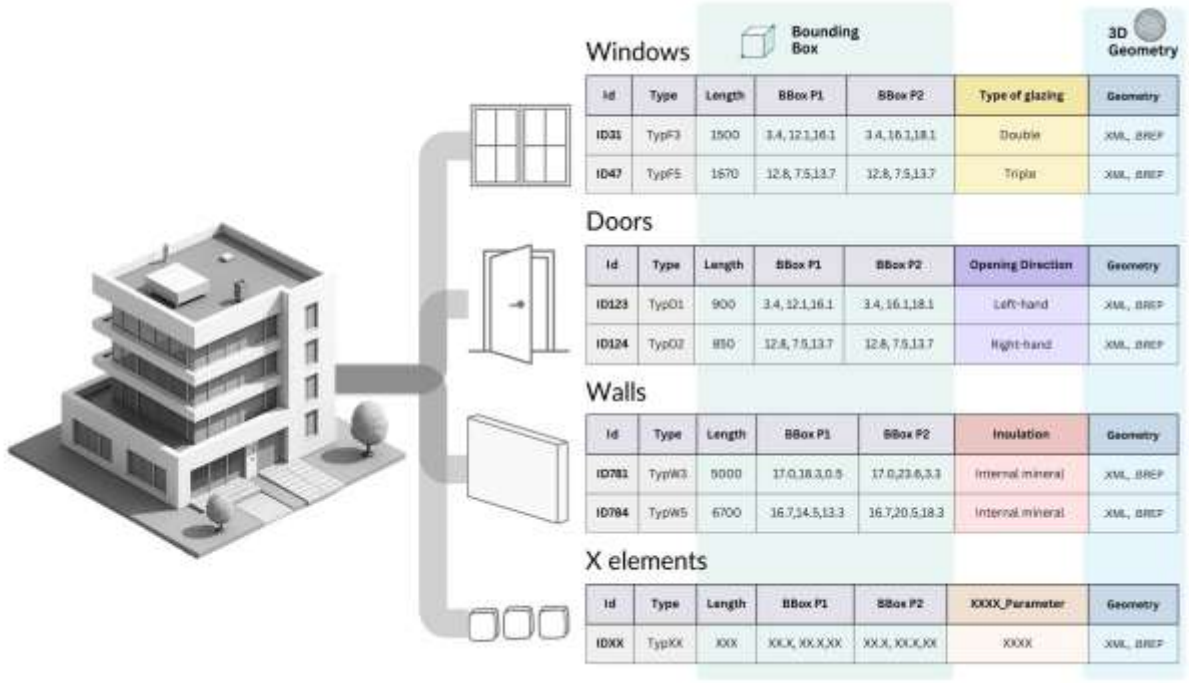
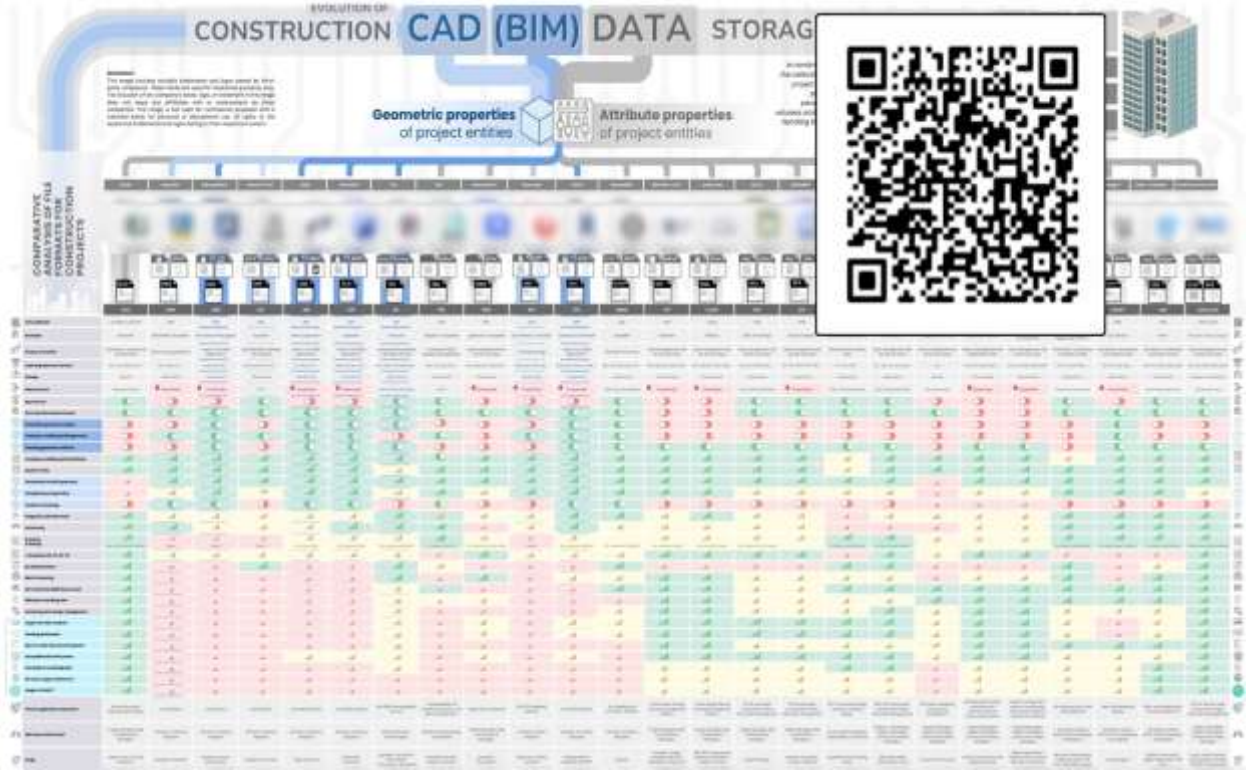


Fig. 3.1-16 Hvert projektelement indeholder attributter, som enten indtastes af designeren eller beregnes i CAD-programmet.

I løbet af de sidste årtier har byggebranchen udviklet mange nye CAD-formater (BIM), der forenkler oprettelse, lagring og overførsel af data. Disse formater kan være lukkede eller åbne, tabelformede, parametriske eller grafiske. Men deres mangfoldighed og fragmentering komplicerer i høj grad datahåndteringen i alle faser af projektets livscyklus. En sammenligningstabel over de vigtigste formater, der bruges til informationsudveksling i byggeriet, er vist i figur 3.1-17 (fuld version tilgængelig via QR-kode).

For at løse problemerne med interoperabilitet og adgang til CAD-data inddrages BIM-ledere (BIM) og koordinatorer, hvis opgave er at kontrollere eksporten, kontrollere datakvaliteten og integrere dele af CAD-data (BIM) i andre systemer.

Men på grund af formaternes lukkede natur og kompleksitet er det vanskeligt at automatisere denne proces, hvilket tvinger specialister til at udføre mange operationer manuelt uden mulighed for at opbygge fuldgyldige in-line databehandlingsprocesser (pipeline).



Figur 3.1-17 Tabel, der sammenligner de vigtigste dataformater, som oplysninger om projektelementer lagres i [53].

For at forstå, hvorfor der er så mange forskellige dataformater, og hvorfor de fleste af dem er lukkede, er det vigtigt at dykke ned i de processer, der finder sted inden for CAD (BIM) programmer, som vil blive udforsket i detaljer i bogens sjette del.

Et ekstra informationslag til geometrien blev introduceret af CAD-systemudviklere i form af BIM-konceptet (Building Information Modeling), et markedsføringsudtryk, der er blevet aktivt promoveret i byggebranchen siden 2002 [54].

Fremkomsten af BIM-konceptet (BOM) og brugen af CAD i processer

Begrebet bygningsinformationsmodellering (BIM), der første gang blev skitseret i fra 2002 [54], stammer fra CAD-softwareproducenternes markedsføringsinitiativer. Det udsprang af CAD-softwareudviklernes markedsføringsinitiativer og var et forsøg på at tilpasse principper, der allerede var veletablerede inden for maskinteknik, til byggebranchens behov.

Inspirationen til BIM kom fra konceptet BOM (Bill of Materials) - stykliste - som har været brugt i stor udstrækning i industrien siden slutningen af 1980'erne. Inden for maskinteknik gjorde BOM det muligt at forbinde data fra CAD-systemer med PDM (Product Data Management), PLM (Product Lifecycle Management) og ERP-systemer, hvilket gav en holistisk styring af teknisk information gennem hele produktets livscyklus (fig. 3.1-8).

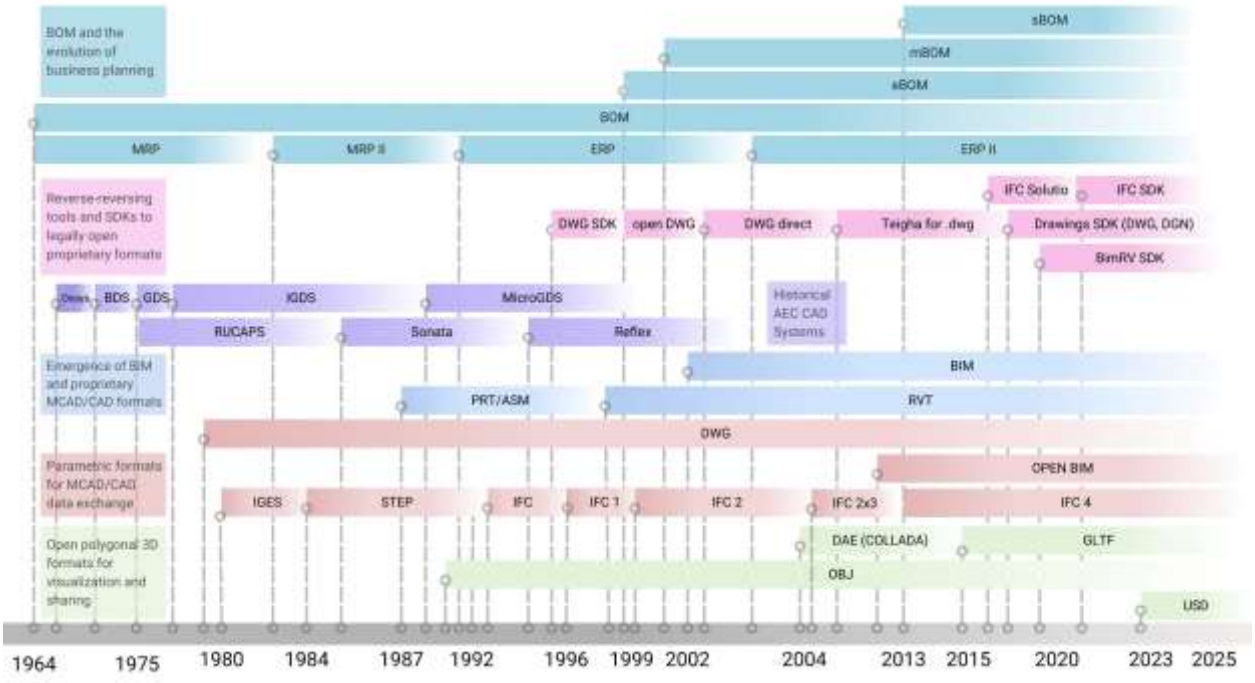


Fig. 3.1-18 Udvikling af specifikationer (BOM), informationsmodellering (BIMdigitale formater i bygge- og anlægsbranchen.

Den moderne udvikling af BOM-konceptet har ført til fremkomsten af en udvidet ramme - XBOM (Extended BOM), som ikke kun omfatter produktsammensætning, men også adfærdsscenarier, driftskrav, bæredygtighedsparametre og data til prædiktiv analyse. XBOM udfylder i bund og grund samme rolle som BIM i byggeriet: Begge tilgange søger at gøre den digitale model til en enkelt kilde til sandhed for alle projektdeltagere i hele projektets livscyklus.

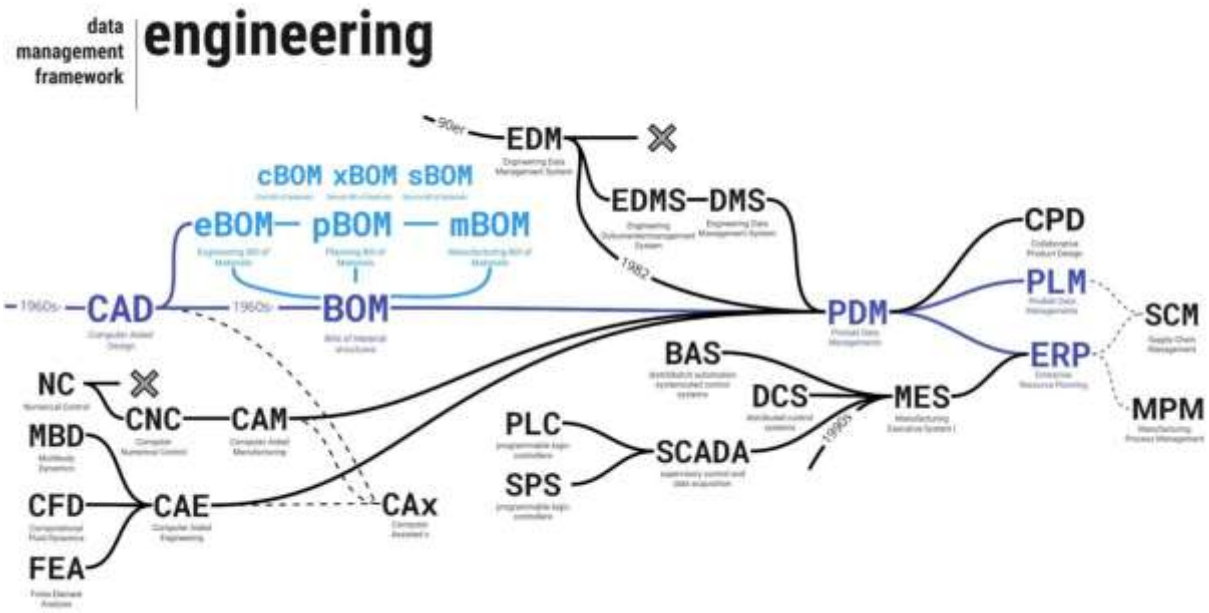
En vigtig milepæl i fremkomsten af styklister i byggeriet var introduktionen af det første parametriske CAD-system (MCAD), der var specielt tilpasset byggebranchen i 2002. Det blev udviklet af det team, der tidligere havde skabt Pro-E®, et revolutionerende MCAD-system til maskinteknik, der var kommet frem i slutningen af 1980'erne og var blevet en industristandard [55].

Allerede i slutningen af 1980'erne var målet at fjerne begrænsningerne [56] i de dengang eksisterende CAD -programmer. Hovedformålet var at reducere det arbejde, der kræves for at foretage ændringer i designelementernes parametre, og at gøre det muligt at opdatere modellen på basis af data uden for CAD-programmerne via en database [57]. Den vigtigste rolle i dette skulle spilles af parametrisering: automatisk hentning af egenskaber fra databasen og brug af dem til at opdatere modellen i CAD-systemer.

> Pro-E og konceptet med elementær parametrisk modellering c BOM, der ligger til grund for det, har haft en betydelig indflydelse på udviklingen af CAD - og MCAD - markedet [58]. I 25 år har denne model været i branchen, og mange moderne systemer er blevet dens konceptuelle efterfølgere.

> *Målet er at skabe et system, der er fleksibelt nok til at opmuntre ingeniøren til nemt at overveje forskellige designs. Og omkostningerne ved at foretage ændringer i designet skal være så tæt på nul som muligt. Traditionel CAD / CAM software begrænser på urealistisk vis muligheden for at foretage billige ændringer helt i begyndelsen af designprocessen [59].*
>
> - Samuel Heisenberg, grundlægger af Parametric Technology Corporation®, udvikler af MCAD -produktet Pro-E og lærer af skaberen af et CAD-produkt, der bruger RVT-formatet

Inden for maskinteknik er PDM-, PLM-, MRP- og ERP-systemer blevet nøgleplatforme. De spiller en central rolle i data- og processtyringen, idet de indsamler oplysninger fra CAx-systemer (CAD, CAM, CAE) og organiserer designaktiviteter baseret på produktstrukturen (BOM: eBOM, pBOM, mBOM) (fig. 3.1-18). Denne integration reducerer fejl, undgår dataduplikation og sikrer end-to-end sporbarhed fra design til produktion.



Figur 3.1-19 Historisk set opstod BOM i 1960'erne som en måde at strukturere data fra CAx-systemer på og sende dem videre til kontrolsystemer.

En af de førende leverandørers køb af en CAD-løsning udviklet af det tidligere Pro-E-team og baseret på BOM-tilgangen blev næsten øjeblikkeligt markeret med udgivelsen af BIM Whitepaper-serien (2002-2003)[60][61]. Allerede fra midten af 2000'erne begyndte BIM-konceptet at blive aktivt fremmet i byggebranchen, hvilket markant øgede interessen for parametrisk software. Populariteten voksede så hurtigt, at konstruktionsdelen af maskinteknik Pro-E - parametrisk CAD promoveret af denne leverandør - faktisk har fortrængt konkurrenterne i segmentet for arkitektonisk og strukturelt design (Fig. 3.1-20). I begyndelsen af 2020'erne har den de facto konsolideret den globale dominans på BIM (CAD)-markedet [62].



Fig. 3.1-20 Populariteten af Google-søgninger (RVT versus IFC): parametrisk CAD skabt af det tidligere Pro-E-team med BOM-understøttelse -BIM er blevet populær i næsten de fleste lande i verden.

I løbet af de sidste 20 år har forkortelsen BIM fået et væld af fortolkninger, hvis mange betydninger har rødder i de oprindelige markedsføringskoncepter, der opstod i begyndelsen af 2000'erne. ISO 19650-standarden, som spillede en vigtig rolle i populariseringen af begrebet, sikrede faktisk BIM's status som en "videnskabsbaseret" tilgang til informationsstyring. Men i selve teksten til standarden, som er dedikeret til datastyring i hele objektets livscyklus ved hjælp af BIM, nævnes forkortelsen BIM, men den defineres aldrig klart

Leverandørens oprindelige hjemmeside, som udgav en række whitepapers om BIM i 2002[60] og 2003[61], gengav faktisk markedsføringsmateriale om BOM (Bills of Materials) og PLM (Product Lifecycle Management)-koncepter, som tidligere blev brugt i Pro-E-softwaren til maskinteknik tilbage i 1990'erne [63].

> *Building Information Modelling, en innovativ ny tilgang til bygningsdesign, -konstruktion og -forvaltning, som blev introduceret af..... [CAD-leverandørens firmanavn] i 2002, har ændret den måde, hvorpå branchefolk over hele verden tænker på, hvordan teknologi kan anvendes til design, konstruktion og forvaltning af bygninger.*
>
> - BIM Whitepaper, 2003 [61]

Disse tidlige publikationer forbandt BIM direkte med konceptet om en centraliseret, integreret database. Som det fremgår af hvidbogen fra 2003, er BIM bygningsinformationsstyring, hvor alle opdateringer sker i et enkelt arkiv, hvilket sikrer, at alle tegninger, udskæringer og specifikationer (BOM - Bills of Materials) er synkroniserede.

> *BIM beskrives som bygningsinformationsstyring, hvor alle opdateringer og alle ændringer finder sted i en database. Så uanset om det drejer sig om skemaer, snit eller tegninger, er alt altid koordineret, konsistent og opdateret.*
>
> - CAD-firmaets hjemmeside leverandør med BIM Whitepaper, 2003 [54]

Ideen om at styre design gennem en enkelt integreret database er blevet diskuteret bredt så tidligt som i forskningen i 1980'erne. For eksempel indeholdt Charles Eastmans BDS-koncept [57] 43 referencer til begrebet "database" (figur 6.1-2). I 2004 var dette antal næsten halveret til 23 i hvidbogen om BIM fra 2002 [64]. Og i midten af 2000'erne var emnet databaser stort set forsvundet fra leverandørernes markedsføringsmateriale og fra digitaliseringsdagsordenen i almindelighed.

Selv om det var databasen og adgangen til den, der oprindeligt var tænkt som kernen i BIM -systemet, er vægten med tiden skiftet til geometri, visualisering og 3D. Den selvsamme registrator af IFC-standarden i 1994, som udgav BIM-hvidbogen i 2002 - den samme leverandør - påpegede i hvidbogen fra begyndelsen af 2000'erne udtrykkeligt begrænsningerne i neutrale formater som IGES, STEP og IFC og behovet for direkte adgang til CAD-databaser:

> *Forskellige applikationer kan være inkompatible, og genindtastede data kan være unøjagtige [...]. Resultatet af traditionelt computerstøttet design [CAD]: højere omkostninger, længere time-to-market og lavere produktkvalitet. I dag bruger alle større applikationer industristandardgrænseflader til dataudveksling på lavt niveau. Ved at bruge de gamle IGES-standarder eller det nye STEP [IFC er en de facto og de jure kopi af STEP/IGES-formatet] til at udveksle data mellem programmer fra forskellige leverandører, kan brugerne opnå en vis datakompatibilitet mellem best-of-breed-produkter. Men IGES og STEP fungerer kun på lave niveauer, og de kan ikke udveksle data, der er lige så rige som de oplysninger, der genereres af nutidens førende programmer [...]. Og selv om disse og andre standarder forbedres næsten dagligt, vil de altid halte bagefter nutidens leverandørprodukter med hensyn til datarigdom. [...] programmer inden for en applikation skal kunne udveksle og bevare datarigdommen uden at ty til neutrale oversættere som IGES, STEP [IFC] eller PATRAN. I stedet skal rammeprogrammerne kunne få direkte adgang til den underliggende CAD-database, så informationens detaljer og nøjagtighed ikke går tabt.*
>
> - CAD-leverandørens hvidbog (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale", 2000 [65].

Allerede i 1980'erne og begyndelsen af 2000'erne blev nøgleelementet i digitalt design i CAD-miljøet således anset for at være databasen snarere end formatfilen eller det neutrale IFC-format. Det blev foreslået, at oversættere skulle droppes, og at applikationer skulle have direkte adgang til dataene. Men i virkeligheden begyndte konceptet BIM i midten af 2020'erne at ligne en "del og hersk"-strategi, hvor softwareleveran-

dørernes interesser ved hjælp af lukkede geometriske kerner prioriteres frem for udviklingen af åben informationsudveksling.

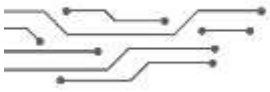
I dag opfattes BIM som en integreret del af byggebranchen. Men i løbet af de sidste to årtier er løfterne om forenklet samarbejde og dataintegration stort set ikke blevet indfriet. De fleste løsninger er stadig bundet til lukkede eller neutrale formater og specialiserede værktøjer. I del 6 af bogen "CAD and BIM: Marketing, Reality and the Future of Design Data in Construction" vil vi se nærmere på BIM's historie, open BIM og IFC samt spørgsmålene om interoperabilitet og geometriske kerner.

> I dag står industrien over for den store udfordring at bevæge sig fra den traditionelle forståelse af CAD (BIM) som et modelleringsværktøj til at bruge det som en fuldgyldig database. Det kræver nye tilgange til at arbejde med information, hvor man forlader afhængigheden af lukkede økosystemer og implementerer åbne løsninger.

Med udviklingen af reverse engineering-værktøjer, der giver adgang til CAD-databaser, samt udbredelsen af Open Source- og LLM-teknologier bevæger brugere og udviklere i byggebranchen sig i stigende grad væk fra softwareleverandørernes vage termer. I stedet skifter fokus til det, der virkelig betyder noget: data (databaser) og processer.

> Bag de trendy akronymer og visualiseringer ligger standardpraksisser for datahåndtering: lagring, overførsel og transformation - dvs. den klassiske ETL-proces (Extract, Transform, Load). Som i andre brancher kræver digitaliseringen af byggeriet ikke kun udvekslingsstandarder, men også en klart struktureret håndtering af heterogen information.

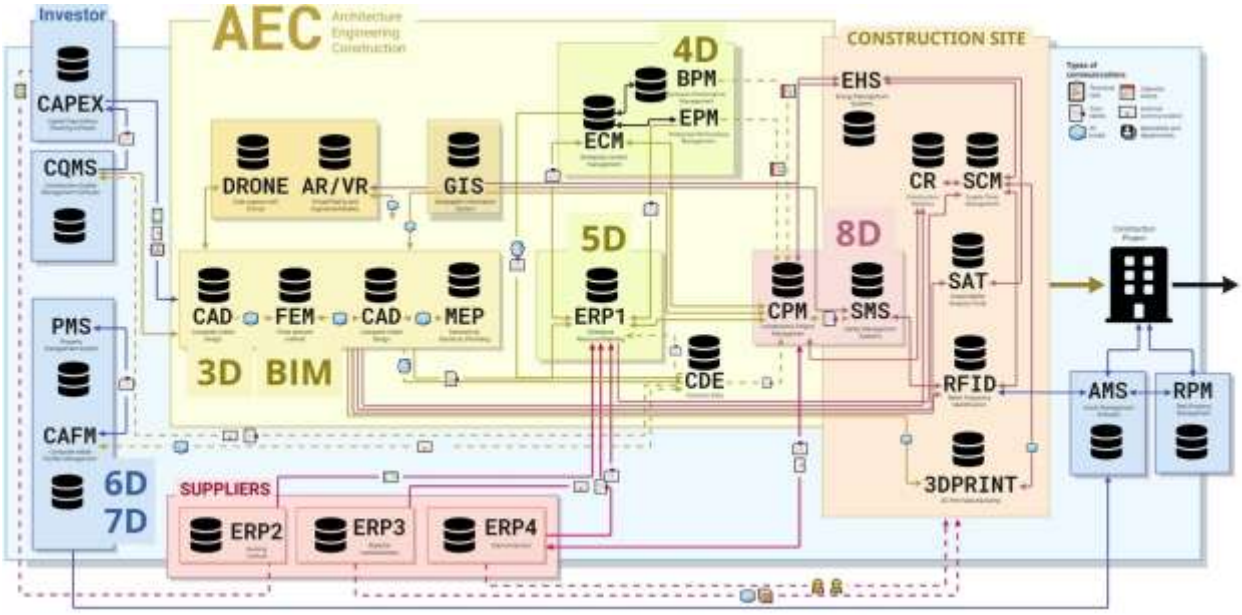
For at kunne udnytte potentialet i CAD-data (BIM) fuldt ud er virksomhederne nødt til at gentænke deres tilgang til informationsstyring. Det vil uundgåeligt føre til et nøgleelement i den digitale transformation - ensretning, standardisering og meningsfuld strukturering af de data, som fagfolk i byggeriet arbejder med hver dag.



KAPITEL 3.2.
DATAHARMONISERING OG -STRUKTURERING

Fyldning af systemer med data i byggebranchen

Uanset om det er store koncerner eller mellemstore virksomheder, er specialister dagligt beskæftiget med at fylde softwaresystemer og databaser med forskellige grænseflader med information i flere formater (fig. 3.2-1), som med hjælp fra ledere skal samarbejde med hinanden. Det er dette kompleks af interagerende systemer og processer, der i sidste ende skaber omsætning og profit for virksomheden.

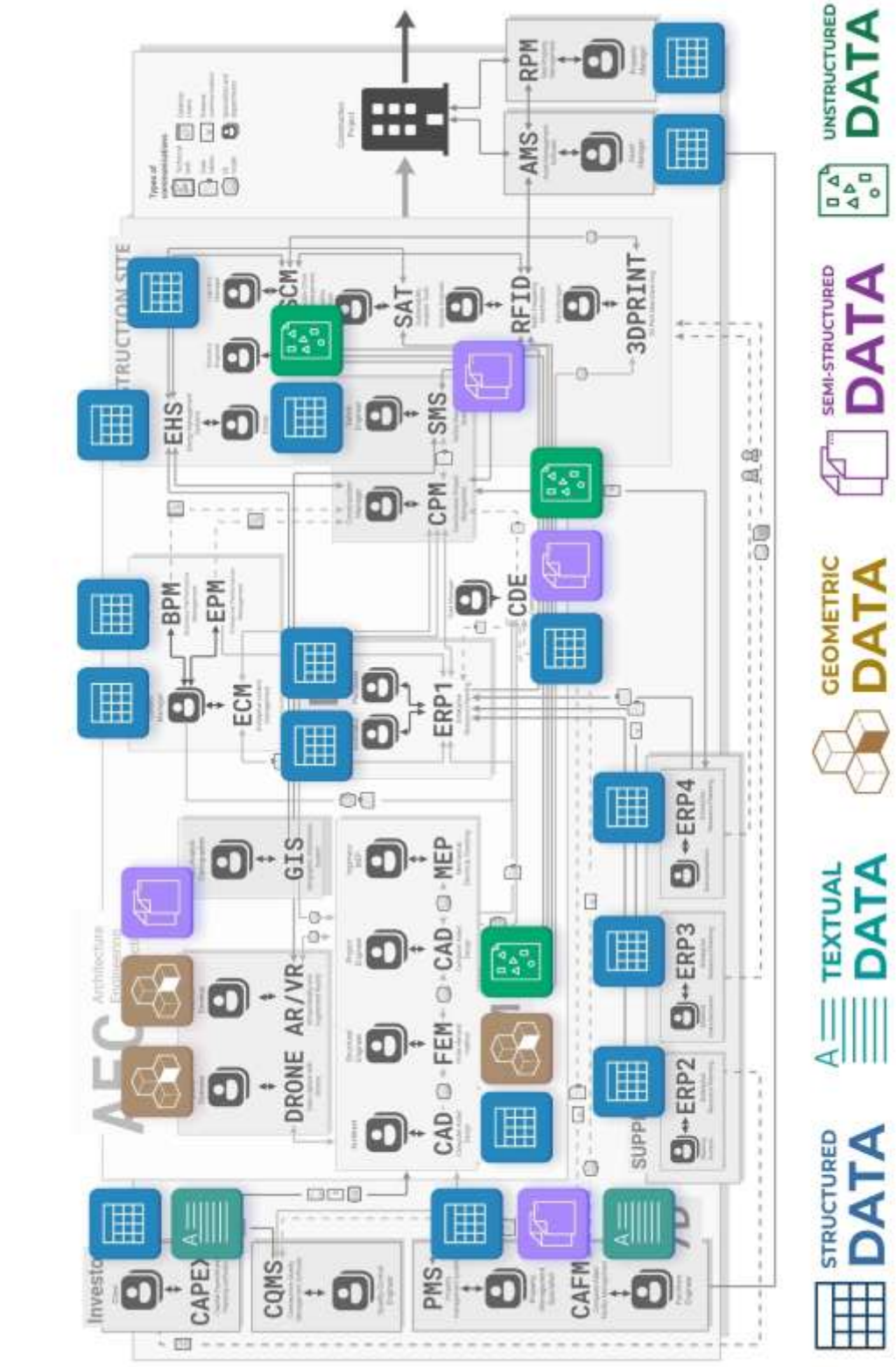


Figur 3.2-1 Stort set alle systemer eller applikationer, der bruges i byggebranchen, har en af de populære RDBMS databaser som kerne.

Hver af de kategorier af systemer, der er nævnt tidligere, og som anvendes i byggebranchen, arbejder med sine egne datatyper, der svarer til disse systemers funktionelle rolle. For at gå fra det abstrakte niveau til det konkrete, går vi fra datatyper til deres repræsentation som formater og dokumenter.

Til den tidligere liste over systemer (figur 1.2-4) tilføjer vi nu de specifikke typer af formater og dokumenter, som de ofte arbejder med:

- **Investor (CAPEX)**
  - Finansielle data: budgetter, udgiftsprognoser (strukturerede data).
  - Data om markedstendenser: markedsanalyser (strukturerede og ustrukturerede data).
  - Juridiske og kontraktlige data: kontrakter (tekstdata).



Figur 3.2-2 Byggebranchen bruger mange systemer med forskellige grænseflader, der håndterer forskellige typer data.

- **Ledelsessystemer (PMS, CAFM, CQMS)**
  - Projektdata: grafer, opgaver (strukturerede data).
  - Vedligeholdelsesdata for anlæg: vedligeholdelsesplaner (tekst og semistrukturerede data).
  - Kvalitetskontroldata: standarder, inspektionsrapporter (tekstuelle og ustrukturerede data).
- **CAD, FEM og BIM**
  - Tekniske tegninger: arkitektoniske, strukturelle planer (geometriske data, ustrukturerede data).
  - Bygningsmodeller: 3D -modeller, materialedata (geometriske og semistrukturerede data).
  - Tekniske beregninger: belastningsanalyse (strukturerede data).
- **Ledelsessystemer for byggepladser (EHS, SCM)**
  - Sikkerheds- og sundhedsdata: sikkerhedsprotokoller (tekstuelle og strukturerede data).
  - Data om forsyningskæden: lagerbeholdninger, ordrer (strukturerede data).
  - Daglige rapporter: arbejdstid, produktivitet (strukturerede data).
- **Droner, AR/VR, GIS, 3D -printing**
  - Geodata: topografiske kort (geometriske og strukturerede data).
  - Realtidsdata: video og fotos (ustrukturerede data).
  - Modeller til 3D -printing: digitale tegninger (geometriske data).
- **Yderligere ledelsessystemer (4D BPM, 5D ERP1)**
  - Tids- og omkostningsdata: tidsplaner, estimater (strukturerede data).
  - Ændringshåndtering: projektændringsregistreringer (tekst og strukturerede data).
  - Præstationsrapportering: indikatorer for succes (strukturerede data).
- **Dataintegration og kommunikation (CDE, RFID, AMS, RPM)**
  - Dataudveksling: dokumentudveksling, datamodeller (strukturerede og tekstuelle data).
  - RFID og sporingsdata: logistik, asset management (strukturerede data).
  - Overvågning og kontrol: sensorer på steder (strukturerede og ustrukturerede data).

Hvert system i byggebranchen - fra byggepladsstyringssystemer til driftsdatabaser - opererer således med sin egen type information: struktureret, tekstuel, geometrisk og andet. Det "datalandskab", som fagfolk skal arbejde med på daglig basis, er ekstremt forskelligt. Men en simpel opremsning af formater afslører ikke kompleksiteten i det virkelige arbejde med information.

I praksis står virksomheder over for det faktum, at data, selv når de hentes fra systemer, ikke er klar til at blive brugt "som de er". Det gælder især for tekster, billeder, PDF'er, CAD-filer og andre formater, som er vanskelige at analysere med standardværktøjer. Derfor er det næste vigtige skridt datatransformation - en proces, uden hvilken behandling, analyse, visualisering og beslutningstagning ikke kan automatiseres effektivt.

Datatransformation: det kritiske fundament for moderne forretningsanalyse

I dag står de fleste virksomheder over for et paradoks: Omkring 80% af deres daglige processer er stadig afhængige af klassisk strukturerede data - velkendte Excel-regneark og relationsdatabaser (RDBMS) [66]. Men samtidig er 80% af den nye information, der kommer ind i virksomhedernes digitale økosystem, ustruktureret eller løst struktureret (fig. 3.2-3) [52]. Dette omfatter tekst, grafik, geometri, billeder, CAD -modeller, dokumentation i PDF, lyd- og videooptagelser, elektronisk korrespondance og meget mere.

Desuden fortsætter mængden af ustrukturerede data med at vokse hurtigt - den årlige vækstrate anslås til 55-65% [67]. En sådan dynamik skaber alvorlige vanskeligheder med at integrere ny information i eksisterende forretningsprocesser. Hvis man ignorerer denne strøm af data i flere formater, opstår der informationshuller, og hele virksomhedens digitale miljø bliver mindre håndterbart.



Fig. 3.2-3 Den årlige vækst i ustrukturerede data skaber udfordringer med at integrere strømmende information i forretningsprocesser.

> At ignorere komplekse ustrukturerede og forvirrende løst strukturerede data i automatiseringsprocesser kan føre til betydelige huller i en virksomheds informationslandskab. I dagens verden med ukontrollerbar og lavineagtig informationsbevægelse er virksomheder nødt til at anvende en hybrid tilgang til datahåndtering, der omfatter effektive metoder til at håndtere alle typer data.

Nøglen til effektiv datastyring ligger i at organisere, strukturere og klassificere forskellige typer data "Babel" (herunder ustrukturerede, tekstuelle og geometriske formater, til strukturerede eller løst strukturerede data). Denne proces omdanner kaotiske datasæt til organiserede strukturer, der kan integreres i systemer og dermed muliggøre beslutningstagning baseret på dem (figur 3.2-4).

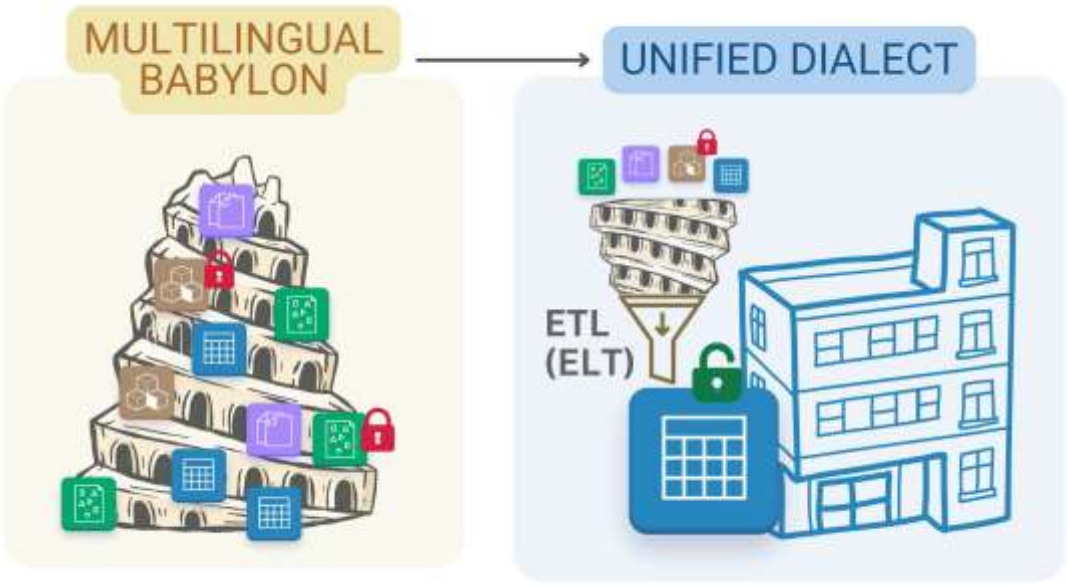


Fig. 3.2-4 Hovedopgaven for datastyringsafdelinger er at oversætte "Babylon" af forskelligartede data i flere formater til et struktureret og kategoriseret system.

En af de største forhindringer for en sådan harmonisering er fortsat den lave grad af interoperabilitet mellem de forskellige digitale platforme - de "siloer", vi diskuterede i de foregående kapitler.

> Ifølge rapporten understreger National Institute of Standards and Technology (NIST, USA) [68], at dårlig datakompatibilitet mellem forskellige bygningsplatforme fører til tab af information og betydelige ekstraomkostninger. Alene i 2002 forårsagede problemer med softwareinteroperabilitet tab på i alt 15,8 milliarder dollars om året i USA, hvor to tredjedele af disse tab bæres af bygningsejere og -operatører, især under drift og vedligeholdelse [68]. Undersøgelsen bemærker også, at standardisering af dataformater kan reducere disse tab og forbedre effektiviteten i hele anlæggets livscyklus.

Ifølge CrowdFlower-undersøgelsen fra 2016 [69], som omfattede 16.000 dataforskere over hele verden, er hovedproblemet fortsat "beskidte" og multiformaterede data. Ifølge denne undersøgelse er den mest værdifulde ressource ikke de endelige databaser eller maskinlæringsmodeller, men den tid, der bruges på at forberede information.

Rengøring, formatering og organisering tager op til 60 procent af en analytikers og datamanagers tid. Næsten en femtedel bruges på at søge efter og indsamle de rigtige datasæt, som ofte er skjult i siloer og utilgængelige for analyse. Og kun ca. 9 procent af tiden bruges direkte på modellering, analyse, forudsigelser og test af hypoteser. Resten bruges på at kommunikere, visualisere, rapportere og undersøge understøttende informationskilder

I gennemsnit er en leders dataarbejde fordelt på følgende måde (figur 3.2-5):

- **Rensning og organisering af data (60%):** Rene og strukturerede data kan reducere analytikerens arbejdstid betydeligt og fremskynde processen med at udføre opgaver.
- **Dataindsamling (19%):** En stor udfordring for data science-medarbejdere er at finde relevante datasæt. Ofte er virksomhedsdata stablet i kaotisk organiserede "siloer", hvilket gør det vanskeligt at

få adgang til de oplysninger, de har brug for.

- **Modellering/maskinlæring (9%):** Hæmmes ofte af manglende klarhed over kundernes forretningsmål. Manglen på en klar mission statement kan ødelægge potentialet i selv den bedste model.
- **Andre opgaver (5%):** Ud over at behandle data skal analytikere beskæftige sig med forskning, udforske data fra forskellige vinkler, kommunikere resultater gennem visualiseringer og rapporter og anbefale optimering af processer og strategier.

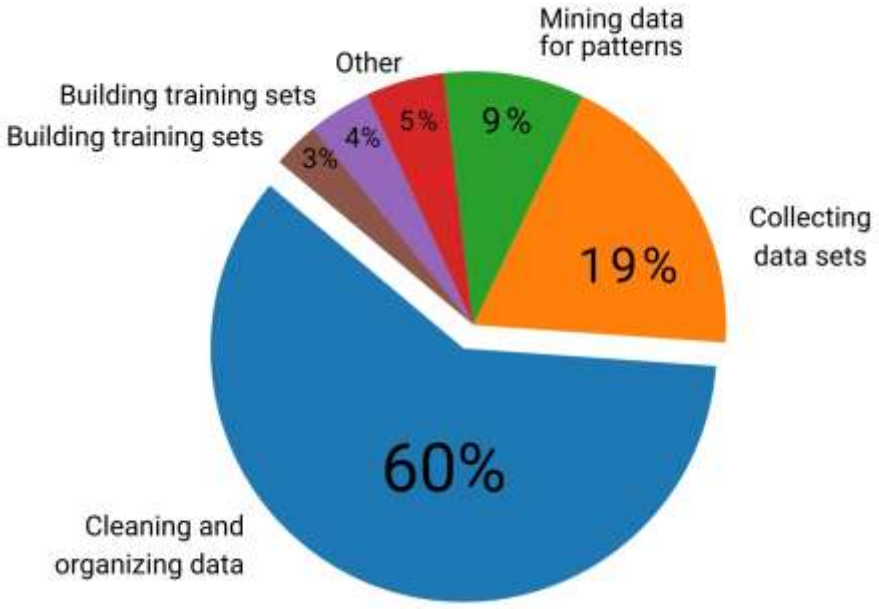


Fig. 3.2-5 Hvad data managers, der arbejder med data, bruger mest tid på (baseret på [70]).

> Disse skøn understøttes af andre undersøgelser. Ifølge Xplenty-undersøgelsen, der blev offentliggjort i BizReport i 2015 [71], bruges mellem 50 og 90% af business intelligence (BI)-medarbejdernes tid (BI) på at forberede data til analyse.

Rensning, validering og organisering af data udgør et kritisk fundament for alle efterfølgende data- og analyseprocesser og tager op til 90% af dataforskernes tid.

Dette omhyggelige arbejde, som er usynligt for slutbrugeren, er afgørende. Fejl i rådata forvrænger uundgåeligt analyser, er vildledende og kan føre til dyre ledelsesfejl. Derfor er datarensnings- og standardiseringsprocesser - fra eliminering af dubletter og udfyldning af mangler til harmonisering af måleenheder og tilpasning til en fælles model - ved at blive en hjørnesten i den moderne digitale strategi.

Grundig transformation, rensning og standardisering af data optager således ikke kun størstedelen af specialisternes tid (op til 80% af arbejdet med data), men afgør også muligheden for at bruge dem effektivt inden for rammerne af moderne forretningsprocesser. Organisering og rensning af data alene er dog ikke nok til at sikre en optimal styring af virksomhedens informationsstrømme. I organiserings- og struktureringsfasen bliver det valget af en passende datamodel, der direkte påvirker bekvemmeligheden og effektiviteten ved at arbejde med information i de efterfølgende faser af behandlingen.

Da data og forretningsmål er forskellige, er det vigtigt at forstå datamodellernes egenskaber og være i stand til at vælge eller skabe den rigtige struktur. Afhængigt af graden af strukturering og den måde, relationerne

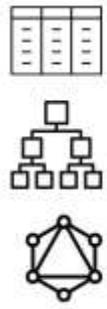
mellem elementerne beskrives på, er der tre hovedmodeller: struktureret, løst struktureret og grafisk. Hver af dem egner sig til forskellige opgaver og har sine egne styrker og svagheder.

Datamodeller: relationer i data og relationer mellem elementer

Data i informationssystemer organiseres på forskellige måder - afhængigt af opgaverne og kravene til lagring, behandling og overførsel af information. Den vigtigste forskel mellem typerne af datamodeller, den form, som information lagres i, er graden af strukturering og den måde, hvorpå relationerne mellem elementer beskrives.

> Strukturerede data har et klart og repeterbart skema: de er organiseret som tabeller med faste kolonner. Dette format giver forudsigelighed, nem behandling og effektivitet ved udførelse af SQL forespørgsler, filtrering og aggregering. Eksempler - databaser (RDBMS), Excel, CSV.

Løst strukturerede data giver mulighed for fleksibel struktur: forskellige elementer kan indeholde forskellige sæt af attributter og lagres som hierarkier. Eksempler er JSON, XML eller andre dokumentformater. Disse data er praktiske, når det er nødvendigt at modellere indlejrede objekter og relationer mellem dem, men på den anden side komplicerer det dataanalyse og standardisering (Fig. 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

Figur 3.2-6 En datamodel er en logisk struktur, der beskriver, hvordan data organiseres, lagres og behandles i et system.

Valget af det rette format afhænger af målene:

- Hvis hastigheden af filtrering og analyse er vigtig - er relationelle tabeller (SQL, CSV, RDBMS, kolonnedatabaser) nok.
- Hvis der er behov for en fleksibel struktur, er det bedre at bruge JSON eller XML.
- Hvis dataene har komplekse relationer - grafdatabaser giver synlighed og skalerbarhed.

I klassiske relationsdatabaser (RDBMS) er hver enhed (f.eks. en dør) repræsenteret af en række og dens egenskaber af tabelkolonner. For eksempel kan en tabel med elementer fra kategorien "Døre" indeholde felterne ID, højde, bredde, brandmodstand og rum-ID, der angiver rummet (Fig. 3.2-7).

I klassiske relationsdatabaser (RDBMS) dannes relationer i form af tabeller, hvor hver post repræsenterer

et objekt, og kolonnerne repræsenterer dets parametre. I tabelformatet ser dataene om døre i projektet sådan ud, hvor hver række repræsenterer et separat element - en dør med dens unikke identifikator og attributter, og forbindelsen med rummet udføres gennem parameteren "Rum-ID".



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

Fig. 3.2-7 Oplysninger om de tre elementer i projektets kategori "Døre" i struktureret tabelform.

I løst strukturerede formater som JSON eller XML lagres data i en hierarkisk eller indlejret form, hvor elementer kan indeholde andre objekter, og deres struktur kan variere. Det gør det muligt at modellere komplekse relationer mellem elementer. Lignende oplysninger om døre i projektet, som blev registreret i struktureret form (figur 3.2-7), er repræsenteret i et løst struktureret format (JSON) på en sådan måde (figur 3.2-8), at de bliver indlejrede objekter i Rooms (Rooms - ID), hvilket logisk set afspejler hierarkiet.

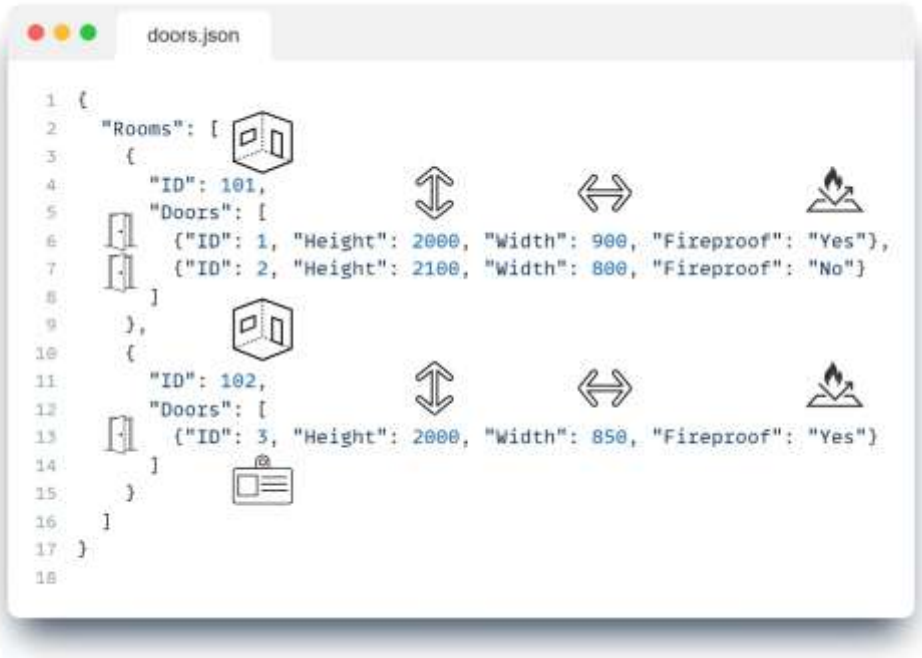


Fig. 3.2-8 Information om elementerne i projektets kategori "Døre" i JSON-format.

I en grafmodel er data repræsenteret som knudepunkter (vertices) og links (edges) mellem dem. Det gør det muligt at visualisere de komplekse relationer mellem objekter og deres attributter. Når det gælder dør- og rumdata i projektet, er grafrepræsentationen af som følger:

- **Noder (knudepunkter)** repræsenterer de vigtigste enheder: rum (rum 101, rum 102) og døre (ID1001, ID1002, ID1003).
- **Ribber (links)** viser relationerne mellem disse enheder, f.eks. at en dør hører til et bestemt rum.
- **Attributter** mappes til noder og indeholder enhedsegenskaber (højde, bredde, brandmodstandsdygtighed for døre).



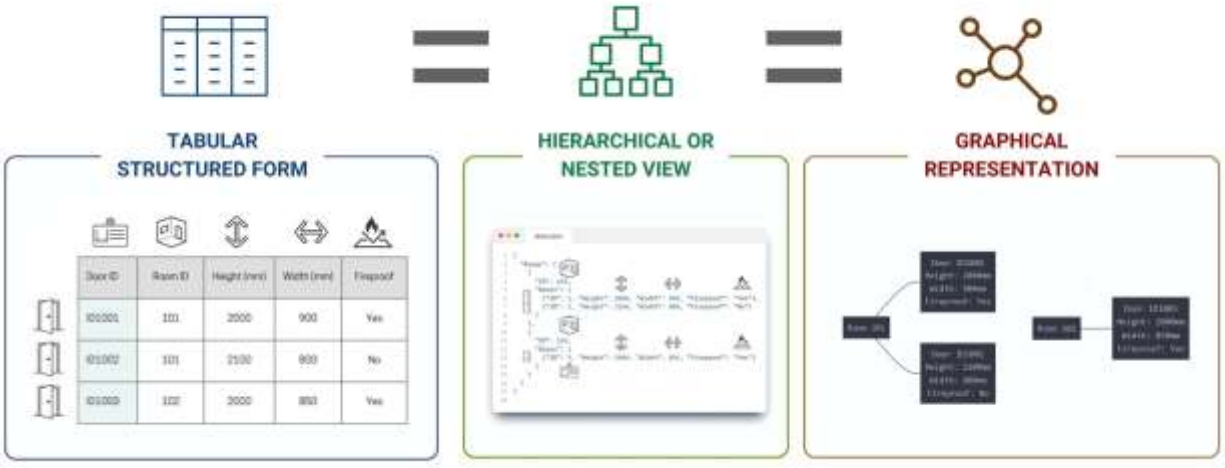
Fig. 3.2-9 Oplysninger om projektdørens enhed i grafvisning.

I den grafiske datamodel for dørbeskrivelser er hvert rum og hver dør separate noder. Dørene er knyttet til rummene via kanter, som angiver, at døren hører til et bestemt rum. Dørenes attributter (højde, bredde,

brandmodstand) gemmes som egenskaber for de tilsvarende noder. Flere detaljer om grafformater, og hvordan grafsemantik opstod i byggebranchen, vil blive diskuteret i kapitlet "Fremkomsten af semantik og ontologi i byggeriet".

Grafdatabaser er effektive, når det ikke så meget er selve dataene, der er vigtige, men relationerne mellem dem, som f.eks. i anbefalingssystemer, routing-systemer eller ved modellering af komplekse relationer i facility management-projekter. Grafformatet forenkler oprettelsen af nye relationer ved at gøre det muligt at tilføje nye datatyper til grafen uden at ændre lagringsstrukturen. Men sammenlignet med relationstabeller og strukturerede formater er der ingen yderligere dataforbindelse i en graf - at overføre todimensionelle databasedata til en graf øger ikke antallet af relationer og giver ikke mulighed for at indhente nye oplysninger.

> Dataenes form og skema skal være skræddersyet til den specifikke brugssag og de opgaver, der skal udføres. For at arbejde effektivt i forretningsprocesser er det vigtigt at bruge de værktøjer og de datamodeller, der hjælper med at få resultater så hurtigt og nemt som muligt.



Figur 3.2-10 De samme oplysninger om projektelementer kan gemmes i forskellige formater ved hjælp af forskellige datamodeller.

> I dag står de fleste store virksomheder over for problemet med overdreven datakompleksitet. Hver af de hundredvis eller tusindvis af applikationer bruger sin egen datamodel, hvilket skaber overdreven kompleksitet - en individuel model er ofte snesevis af gange mere kompleks end nødvendigt, og summen af alle modeller er tusindvis af gange mere kompleks. Denne overdrevne kompleksitet hæmmer både udviklernes og slutbrugernes arbejde betydeligt.

En sådan kompleksitet sætter alvorlige begrænsninger for udvikling og vedligeholdelse af virksomhedens systemer. Hvert nyt element i modellen kræver ekstra kode, implementering af ny logik, grundig testning og tilpasning til eksisterende løsninger. Alt dette øger omkostningerne og forsinker arbejdet i virksomhedens automatiseringsteam, hvilket gør selv enkle opgaver til dyre og tidskrævende processer.

Kompleksitet påvirker alle niveauer af dataarkitekturen. I relationsdatabaser kommer den til udtryk i det

voksende antal tabeller og kolonner, som ofte er overflødige. I objektorienterede systemer øges kompleksiteten af de mange klasser og indbyrdes forbundne egenskaber. I formater som XML eller JSON kommer kompleksiteten til udtryk gennem forvirrende indlejrede strukturer, unikke nøgler og inkonsekvente skemaer.

> Datamodellernes overdrevne kompleksitet gør ikke kun systemerne mindre effektive, men også vanskelige at forstå for slutbrugerne, og i fremtiden vil store sprogmodeller og LLM-agenter. Det er problemet med forståelse og kompleksitet i datamodeller og databehandling, der rejser spørgsmålet: Hvordan gør man data så nemme at bruge, at de faktisk hurtigt bliver nyttige?

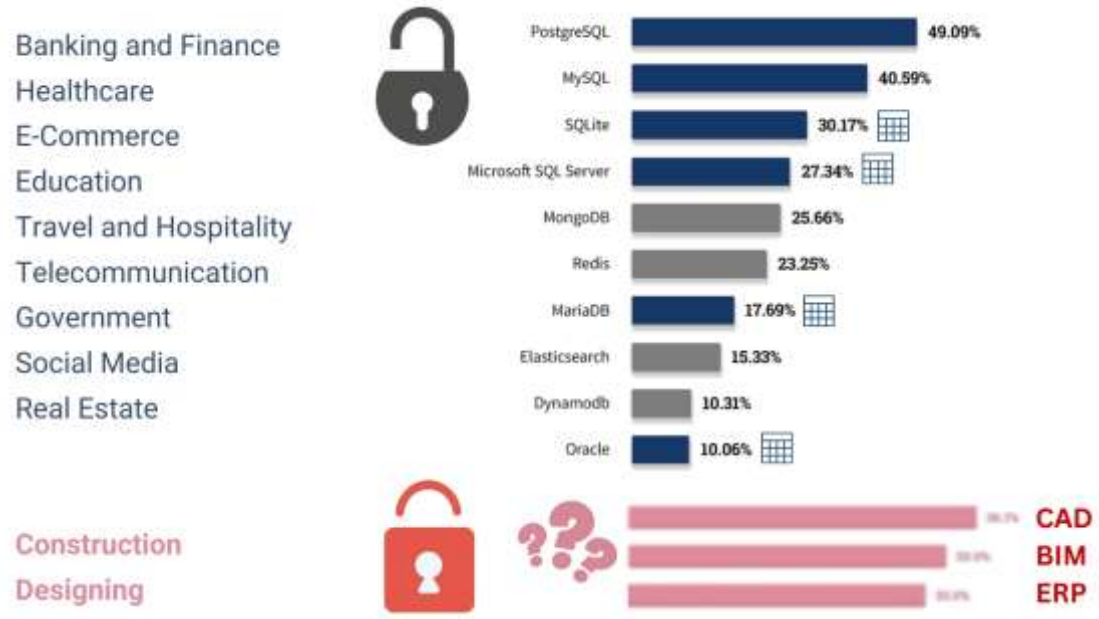
Selv når datamodeller er valgt med omhu, reduceres deres anvendelighed dramatisk, hvis adgangen til data er begrænset. Proprietære formater og lukkede platforme hindrer integration, komplicerer automatisering og fjerner kontrollen med ejendomsretligt beskyttede oplysninger, hvilket ikke bare skaber en silo af nye data, men en låst silo, som kun kan tilgås med leverandørens tilladelse. For at forstå problemets omfang er det vigtigt at overveje, præcis hvordan lukkede systemer påvirker digitale processer i byggeriet.

Proprietære formater og deres indvirkning på digitale processer

En af de største udfordringer for byggevirksomheder i forbindelse med digitalisering er begrænset adgang til data. Det gør det svært at integrere systemer, reducerer informationskvaliteten og gør det vanskeligt at organisere effektive processer. Brugen af proprietære formater og lukkede softwareløsninger er ofte årsagen til disse vanskeligheder.

Desværre har mange af de programmer, der bruges i byggebranchen, indtil nu kun givet brugeren mulighed for at gemme data i proprietære formater eller cloud storage, som kun kan tilgås via strengt begrænsede grænseflader. Og det er ikke ualmindeligt, at disse løsninger er bygget i afhængighed af endnu mere lukkede systemer fra større leverandører. Resultatet er, at selv de udviklere, der gerne vil tilbyde mere åbne arkitekturer, er tvunget til at overholde de regler, som de store leverandører har dikteret.

> Mens moderne byggedatastyringssystemer i stigende grad understøtter åbne formater og standarder (fig. 3.1-5), er CAD- (BIM)-baserede databaser og relaterede ERP- og CAFM-systemer fortsat isolerede proprietære "øer" i branchens digitale landskab (fig. 3.2-11).



Figur 3.2-11 Datas lukkede og proprietære karakter skaber barrierer for dataintegration og -adgang.

Lukkede og monopoliserede formater og protokoller er ikke kun et problem for byggebranchen. I mange økonomiske sektorer er kampen mod lukkede standarder og begrænset adgang til data startet med langsommere innovation (figur 3.2-12), tilstedeværelsen af kunstige adgangsbarrierer for nye aktører og større afhængighed af store leverandører. Med den hurtige vækst i datas betydning har konkurrencemyndighederne simpelthen ikke tid til at reagere på udfordringerne fra nye digitale markeder, og som følge heraf bliver lukkede formater og begrænset adgang til data i bund og grund digitale "grænser", der strømmen af information og vækst [63].

> *Hvis maskiner producerer alt, hvad vi har brug for, så afhænger vores situation af, hvordan disse varer fordeles. Alle vil kun kunne nyde et liv i velstand, hvis den rigdom, der produceres af maskiner, deles. Eller også vil de fleste mennesker ende med at leve i dyb fattigdom, hvis bilejere med held kan lobbye mod omfordeling af velstand. Indtil videre ser det ud til, at det går den anden vej, hvor teknologien fører til stadig større ulighed [72].*
>
> - Stephen Hawking, astrofysiker, 2015

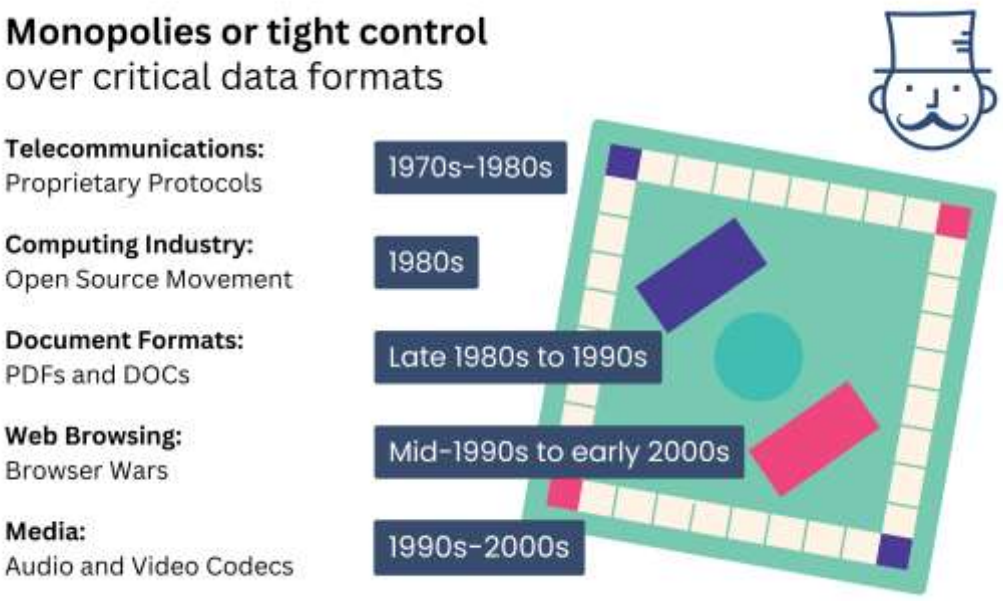


Fig. 3.2-12 Monopolejerskab over vigtige dataformater og protokoller er ikke et problem, der kun findes i byggebranchen.

Som følge af den lukkede adgang til databaser står programmer, dataadministratorer, dataanalytikere, it-specialister og udviklere, der skaber applikationer til dataadgang, -behandling og -automatisering i byggebranchen, i dag over for mange afhængigheder af softwareleverandører (fig. 3.2-13). Disse afhængigheder i form af yderligere adgangslag kræver, at der skabes løsninger med specialiserede API -forbindelser og særlige værktøjer og software.

> En API (Application Programming Interface) er en formaliseret grænseflade, hvorigennem et program kan interagere med et andet og udveksle data og funktionalitet uden at skulle have adgang til kildekoden. En API beskriver, hvilke anmodninger et eksternt system kan fremsætte, hvilket format de skal have, og hvilke svar det vil modtage. Det er en standardiseret "kontrakt" mellem softwaremoduler.

Det store antal afhængigheder af lukkede løsninger får hele kodearkitekturen og forretningsproceslogikken i en virksomhed til at blive en "spaghettiarkitektur" af værktøjer, der er afhængige af softwareleverandørens politik for at give kvalitetsadgang til data.

Afhængighed af lukkede løsninger og platforme fører ikke kun til tab af fleksibilitet, men også til reelle forretningsrisici. Ændring af licensbetingelser, lukning af adgang til data, ændring af formater eller API-struktur - alt dette kan blokere kritiske processer. Pludselig viser det sig, at opdatering af en tabel kræver omarbejdning af en hel blok af integrationer og connectorer (fig. 3.2-13), og enhver storstilet opdatering af software eller dens API-leverandør bliver en potentiel trussel mod stabiliteten i hele virksomhedens system.

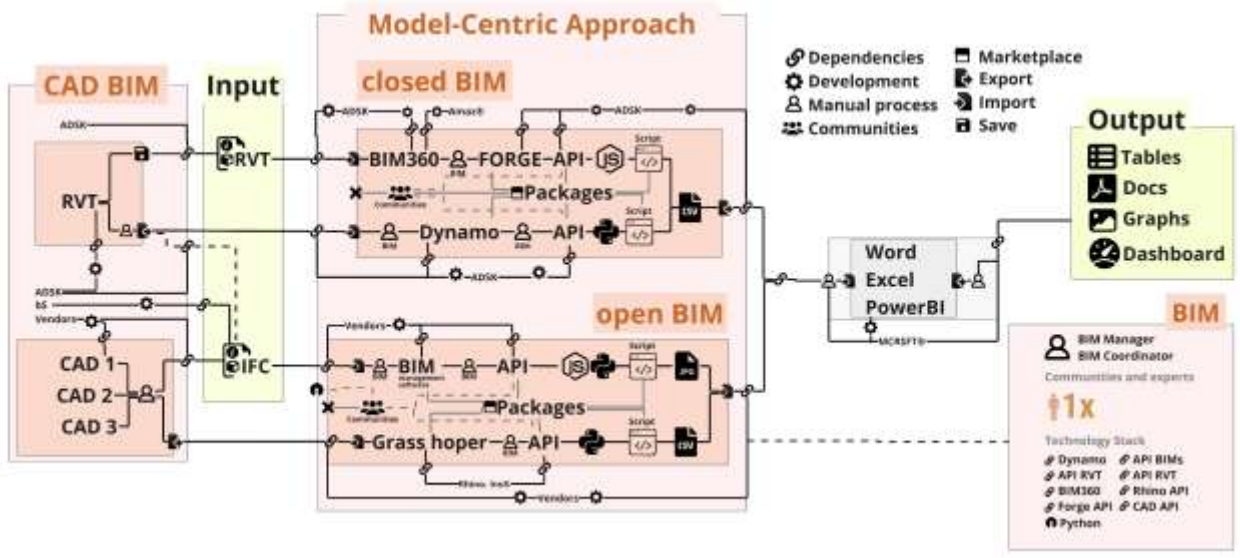


Fig. 3.2-13 Et eksempel på det store antal afhængigheder i CAD-behandling -data skaber barrierer for dataintegration i byggevirksomhedernes økosystem.

> Udviklere og systemarkitekter er under sådanne forhold tvunget til ikke at arbejde for at forudse, men for at overleve. I stedet for at implementere nye løsninger tilpasser de sig. I stedet for at udvikle forsøger de at opretholde kompatibilitet. I stedet for at automatisere og fremskynde processer bruger de deres tid på at studere de næste lukkede grænseflader, API dokumentation og endeløs genopbygning af kode.

At arbejde med lukkede formater og systemer er ikke bare en teknisk udfordring - det er en strategisk begrænsning. På trods af de åbenlyse muligheder, der ligger i moderne automatisering, AI, LLM og predictive analytics, er der mange virksomheder, der ikke udnytter deres fulde potentiale. Og de barrierer, der opstilles af proprietære formater (figur 3.2-13), nægter virksomhederne adgang til deres egne data. Det er måske det ironiske ved den digitale transformation i byggeriet.

> Datatransparens og åbne systemer er ikke en luksus, men en forudsætning for hastighed og effektivitet. Uden åbenhed er forretningsprocesserne fyldt med unødvendigt bureaukrati, godkendelseskæder i flere lag og en voksende afhængighed af HiPPO-princippet - at træffe beslutninger baseret på den bedst betalte persons mening.

Ikke desto mindre er der et paradigmeskift på vej i horisonten. På trods af de proprietære løsningers dominans er flere og flere virksomheder ved at indse begrænsningerne i de arkitekturer, der er inspireret af den fjerde industrielle revolution. I dag skifter vektoren til principperne for den femte revolution, hvor data som et strategisk aktiv, åbne grænseflader (API'er) og ægte interoperabilitet mellem systemer er i centrum.

Denne overgang markerer et skift væk fra lukkede økosystemer til fleksible, modulære digitale arkitekturer, hvor åbne formater, standarder og gennemsigtig dataudveksling er nøglen.

Åbne formater ændrer tilgangen til digitalisering

Byggebranchen var en af de sidste, der tog fat på problemet med lukkede og proprietære data. I modsætning til andre økonomiske sektorer har digitaliseringen været langsom til at udvikle sig her. Årsagerne er bl.a. branchens traditionelle konservative natur, udbredelsen af forskellige lokale løsninger og den dybt forankrede karakter af papirbaseret dokumenthåndtering. I årtier har vigtige byggeprocesser været baseret på fysiske tegninger, telefonopkald og usynkroniserede databaser. I denne sammenhæng har lukkede formater længe været opfattet som normen snarere end en hindring.

Erfaringer fra andre brancher viser, at fjernelse af barrierer for lukkede data fører til en kraftig stigning i innovation, hurtigere udvikling og øget konkurrence [73]. Inden for videnskaben giver udveksling af åbne data mulighed for at fremskynde opdagelser og udvikle internationalt samarbejde. Inden for medicin kan det forbedre effektiviteten af diagnose og behandling. Inden for softwareteknik - for at skabe økosystemer for samskabelse og hurtig produktforbedring.

> Ifølge McKinsey-rapporten "Open Data: Unlock Innovation and Productivity with Information Flow" fra 2013. [74] har åbne data potentiale til at frigøre 3 til 5 billioner dollars årligt på tværs af syv nøgleindustrier, herunder byggeri, transport, sundhedspleje og energi. Ifølge samme undersøgelse gør decentraliserede dataøkosystemer det muligt for store byggefirmaer og entreprenører at reducere omkostningerne til softwareudvikling og -vedligeholdelse, hvilket fremskynder den digitale indførelse.

Overgangen til åbne arkitekturer, som for længst er begyndt i andre sektorer af økonomien, breder sig gradvist til byggebranchen. Store virksomheder og offentlige bygherrer, og især de finansielle organisationer, der kontrollerer investeringer i byggeprojekter, kræver i stigende grad brug af åbne data og adgang til kildekoden til beregninger, kalkulationer og applikationer. Udviklere forventes ikke længere bare at skabe digitale løsninger og vise de endelige tal for et projekt - de forventes at være gennemsigtige, reproducerbare og uafhængige af tredjepartsleverandører af applikationer.

> Brug af open source-løsninger giver kunden tillid til, at selv om eksterne udviklere stopper samarbejdet eller forlader projektet, vil det ikke påvirke muligheden for at videreudvikle værktøjer og systemer. En af de største fordele ved åbne data er deres evne til at eliminere applikationsudviklernes afhængighed af specifikke platforme for at få adgang til data.

Hvis en virksomhed ikke helt kan opgive proprietære løsninger, er et muligt kompromis brugen af reverse engineering-teknikker. Disse juridiske og teknisk sunde metoder gør det muligt at konvertere lukkede formater til mere tilgængelige, strukturerede og egnede til integration. Det er især vigtigt, når der er behov for at oprette forbindelse til ældre systemer eller migrere oplysninger fra et softwarelandskab til et andet.

> Et af de klareste eksempler i historien om overgang til åbne formater og anvendelse af reverse engineering (lovlig hacking af proprietære systemer) i byggeriet er historien om kampen for åbning af DWG-formatet, der i vid udstrækning anvendes i computerstøttede designsystemer (CAD). Som reaktion på en softwareleverandørs monopol dannede de øvrige 15 CAD-leverandører i 1998 en ny alliance kaldet "Open DWG" for at give udviklere gratis og uafhængige værktøjer til at arbejde med DWG-formatet (de facto-standarden for tegningsoverførsel) uden behov for proprietær software eller lukkede API'er. Denne begivenhed var et vendepunkt, der gjorde det muligt for titusindvis af virksomheder at få fri adgang til det lukkede format i en populær CAD-løsning fra slutningen af 1980'erne og frem til i dag og skabe kompatible løsninger, der fremmede konkurrencen på CAD-markedet [75]. I dag bruges "Open DWG"-SDK'et, som blev skabt tilbage i 1996, i næsten alle løsninger, hvor det er muligt at importere, redigere og eksportere DWG-formatet uden for DWG-formatudviklerens officielle applikation.

Andre tech-giganter gennemfører lignende forandringer. Microsoft, der engang var et symbol på en proprietær tilgang, åbnede kildekoden til.NET Framework, begyndte at bruge Linux i Azure cloud service-infrastrukturen og opkøbte GitHub for at styrke sin position i Open Source-samfundet. [76]. Meta (tidligere Facebook) udgav open source AI-modeller , såsom Llama-serien, for at fremme innovation og samarbejde inden for udvikling af AI-agenter. CEO Mark Zuckerberg forventer, at open source platforme vil føre an i de teknologiske fremskridt i løbet af det næste årti [77].

> Open Source er en softwareudviklings- og distributionsmodel, hvor kildekoden er åben for fri brug, undersøgelse, ændring og distribution.

Åbne data og open source-løsninger er ikke bare en trend, men grundlaget for digital bæredygtighed. De giver virksomhederne fleksibilitet, modstandsdygtighed, kontrol over deres egne beslutninger og mulighed for at skalere digitale processer uden at være afhængig af leverandørernes politikker. Og lige så vigtigt er det, at de giver virksomhederne kontrollen tilbage over det 21. århundredes mest værdifulde ressource - deres data.

Paradigmeskift: Open Source som afslutningen på æraen med softwareleverandørernes dominans

Byggebranchen gennemgår et skift, som man ikke kan tjene penge på på den sædvanlige måde. Konceptet med datadrevet, datacentreret tilgang og brugen af Open Source værktøjer fører til en nytænkning af de spilleregler, som softwaregiganterne på markedet står på.

> I modsætning til tidligere teknologitransformationer vil denne overgang ikke blive aktivt fremmet af leverandørerne. Paradigmeskiftet truer deres traditionelle forretningsmodeller baseret på licenser, abonnementer og rådgivning. Den nye virkelighed involverer ikke et færdigt produkt eller et betalt abonnement - det kræver en omlægning af processer og tankegang.

For at kunne administrere og udvikle datacenterløsninger baseret på åbne teknologier bliver virksomhederne nødt til at gentænke de interne processer. Specialister fra forskellige afdelinger skal ikke kun samarbejde, men også gentænke, hvordan de arbejder sammen.

Det nye paradigme indebærer brug af åbne data og Open Source-løsninger, hvor værktøjer baseret på kunstig intelligens og store sprogmodeller (LLM) snarere end programmører vil spille en særlig rolle i skabelsen af softwarekode. Allerede i midten af 2024 er mere end 25 procent af den nye kode hos Google skabt med AI [78]. I fremtiden vil kodning med LLM'er udføre 80% af arbejdet på blot 20% af tiden (figur 3.2-14).

> Ifølge McKinseys 2020-undersøgelse [79] erstatter GPU'er i stigende grad CPU'er i analyser på grund af deres høje ydeevne og understøttelse af moderne Open Source-værktøjer. Det gør det muligt for virksomheder at fremskynde databehandling uden betydelige investeringer i dyr software eller ansættelse af få specialister.

Førende konsulentfirmaer som McKinsey, PwC og Deloitte understreger den voksende betydning af åbne standarder, Open Source applikationer på tværs af brancher.

> Ifølge rapporten PwC Open Source Monitor 2019 [80] bruger 69% af virksomheder med 100 eller flere ansatte bevidst open source-løsninger. OSS bruges især aktivt i store virksomheder: 71% af virksomheder med 200-499 ansatte, 78% i kategorien 500-1999 ansatte og op til 86% blandt virksomheder med mere end 2000 ansatte. Ifølge Synopsys' OSSRA 2023-rapport indeholdt 96% af de analyserede kodebaser open source-komponenter [81].

Fremtidens udviklerrolle er ikke at skrive kode manuelt, men at designe datamodeller, flow-arkitekturer og styre AI-agenter, der skaber de rigtige beregninger efter behov. Brugergrænseflader vil blive minimalistiske, og interaktion vil blive dialogbaseret. Klassisk programmering vil vige for design på højt niveau og orkestrering af digitale løsninger (figur 3.2-14). Nuværende tendenser - såsom low-code-platforme (fig. 7.4-6) og LLM-aktiverede økosystemer (fig. 7.4-4) - vil reducere omkostningerne til udvikling og vedligeholdelse af IT-systemer betydeligt.

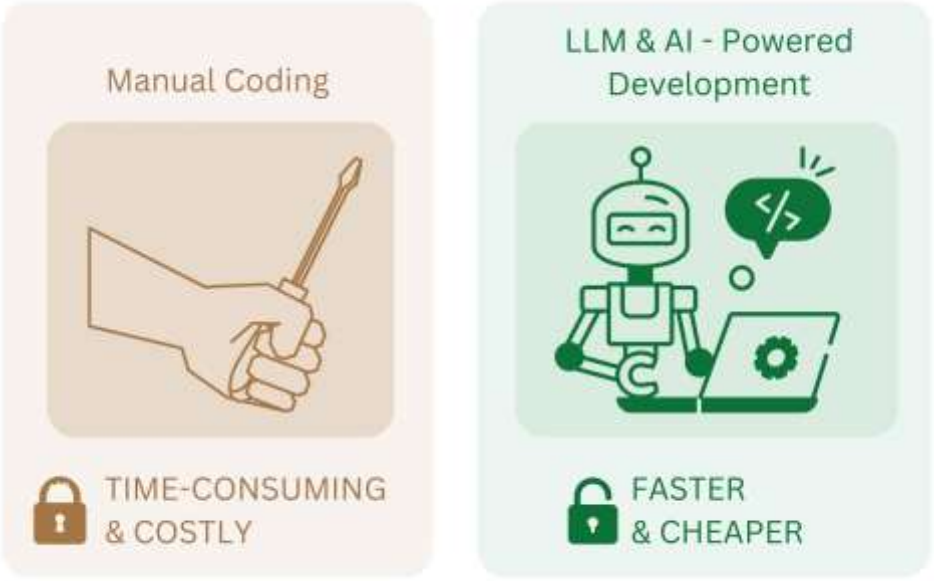


Fig. 3.2-14 Mens applikationer i dag skabes manuelt af programmører, vil en betydelig del af koden i fremtiden blive genereret af AI- og LLM-baserede løsninger.

Denne overgang vil være anderledes end tidligere, og det er usandsynligt, at de store softwareleverandører vil katalysere den.

> I Harvard Business School-undersøgelsen "The Value of Open Source Software" 2024 [40] anslås den samlede værdi af open source-software ud fra to synsvinkler. Hvis vi på den ene side beregner, hvor meget det ville koste at bygge alle eksisterende open source-løsninger fra bunden, ville beløbet være ca. 4,15 milliarder dollars. Hvis vi på den anden side forestiller os, at hver virksomhed selv udvikler sine egne analoger til Open Source-løsninger (hvilket sker overalt) uden at have adgang til eksisterende værktøjer, så vil de samlede forretningsomkostninger nå op på kolossale 8,8 billioner dollars - dette er omkostningerne ved efterspørgsel.

> Det er ikke svært at gætte, at ingen større softwareleverandør er interesseret i at formindske et softwaremarked med en potentiel værdi på 8,8 billioner dollars til kun 4,15 milliarder dollars. Det ville betyde en reduktion af efterspørgslen med mere end 2.000 gange. En sådan transformation er simpelthen urentabel for leverandører, hvis forretningsmodeller er bygget på mange års opretholdelse af kundernes afhængighed af lukkede løsninger. Så virksomheder, der forventer, at nogen vil tilbyde dem en praktisk og åben nøglefærdig løsning, kan blive skuffede - de leverandører dukker simpelthen ikke op.

Skiftet til en åben digital arkitektur betyder ikke tab af arbejdspladser eller indtægter. Tværtimod skaber det forudsætningerne for fleksible og tilpasningsdygtige forretningsmodeller, som i sidste ende kan fortrænge det traditionelle licens- og boks-softwaremarked.

I stedet for at sælge licenser - tjenester, i stedet for lukkede formater - åbne platforme, i stedet for afhængighed af en leverandør - uafhængighed og mulighed for at bygge løsninger til reelle behov. De, der før blot brugte værktøjer, vil kunne blive deres medforfattere. Og de, der kan arbejde med data, modeller, scenarier og logik, vil befinde sig i centrum af branchens nye digitale økonomi. Vi vil tale mere om disse forandringer og om, hvilke nye roller, forretningsmodeller og samarbejdsformer der opstår omkring åbne data i bogens sidste og tiende del.

> Løsninger baseret på åbne data og åben kode vil gøre det muligt for virksomheder at fokusere på effektiviteten af forretningsprocesser i stedet for at kæmpe med forældede API'er og integrere lukkede systemer. En bevidst overgang til åben arkitektur kan forbedre produktiviteten betydeligt og reducere afhængigheden af leverandører.

Overgangen til en ny virkelighed er ikke kun en ændring i tilgangen til softwareudvikling, men også en gentænkning af selve princippet om at arbejde med data. I centrum for denne transformation står ikke kode, men information: dens struktur, tilgængelighed og fortolkningsmuligheder. Det er her, åbne og strukturerede data kommer i forgrunden og bliver en integreret del af den nye digitale arkitektur.

Strukturerede åbne data: fundamentet for digital transformation

Mens virksomheders bæredygtighed i tidligere årtier i høj grad blev bestemt af valget af softwareløsninger og afhængighed af specifikke leverandører, er nøglefaktoren i dagens digitale økonomi datakvalitet og evnen til at arbejde effektivt med den. Open source-kode er en vigtig del af det nye teknologiske paradigme, men dets potentiale frigøres først for alvor, når data er forståelige, organiserede og maskinlæsbare. Blandt alle typer datamodeller er strukturerede åbne data ved at blive hjørnestenen i en bæredygtig digital transformation.

Den største fordel ved strukturerede åbne data er den entydige fortolkning og muligheden for automatiseret behandling. Det giver mulighed for betydelige effektivitetsgevinster både i de enkelte operationer og på tværs af organisationen.

> Ifølge Deloittes rapport "The Data Transfer Process in Enterprise Transformation" [82] er det afgørende at arbejde med IT for at styre overførslen af strukturerede data. Ifølge den britiske regerings rapport "Data Analytics and AI in Government Project Delivery" (2024) [83] er det vigtigt at fjerne barrierer for datadeling mellem forskellige projekter og organisationer for at forbedre effektiviteten i projektledelsen. Dokumentet understreger, at standardisering af dataformater og indførelsen af åbne dataprincipper kan undgå dobbeltarbejde, minimere tidstab og forbedre nøjagtigheden af prognoser.

I byggebranchen, hvor der traditionelt hersker en høj grad af fragmentering og mangfoldighed af formater, spiller den strukturerede foreningsproces og strukturerede åbne data en afgørende rolle i udformningen af sammenhængende og håndterbare processer (fig. 4.1-14). De giver projektdeltagerne mulighed for at fokusere på at forbedre produktiviteten i stedet for at løse tekniske problemer i forbindelse med inkompatibilitet mellem lukkede platforme, datamodeller og formater.

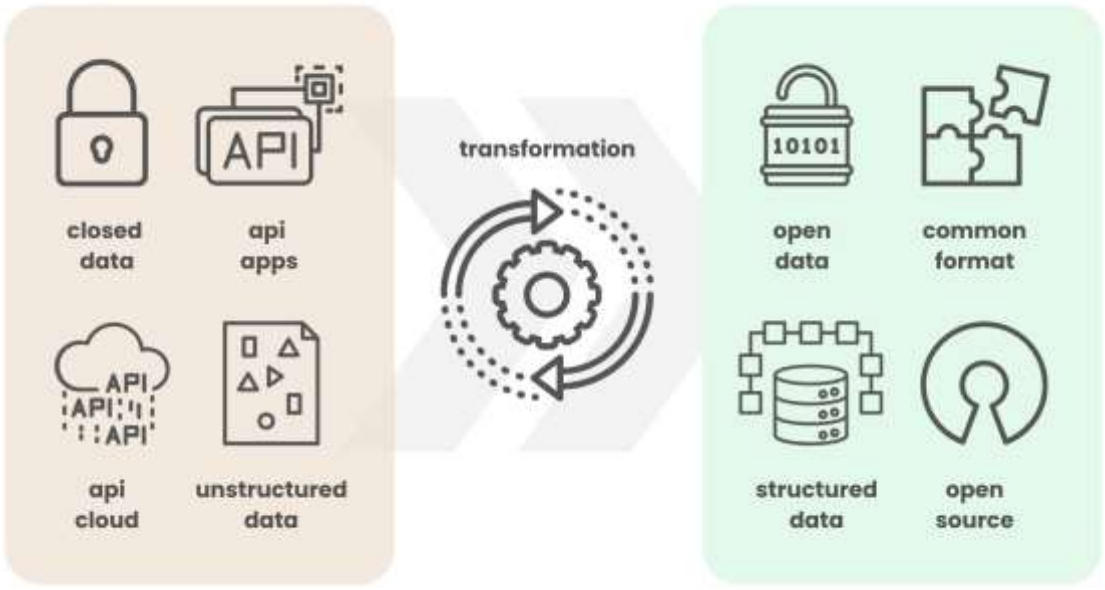


Fig. 3.2-15 Åbne strukturerede data reducerer afhængigheden af softwareløsninger og -platforme og fremskynder innovation.

Moderne teknologiske værktøjer, som vi vil diskutere i detaljer senere i bogen, gør det muligt ikke bare at indsamle information, men også at rense den automatisk: fjerne duplikationer, rette fejl og normalisere værdier. Det betyder, at analytikere og ingeniører ikke arbejder med forskellige dokumenter, men med en organiseret vidensbase, der er egnet til analyse, automatisering og beslutningstagning.

> *Gør det så enkelt som muligt, men ikke enklere.*
>
> - Albert Einstein, teoretisk fysiker (forfatterskabet til citatet er omstridt [84])

I dag kan de fleste brugergrænseflader til at arbejde med data skabes automatisk - uden at det er nødvendigt manuelt at skrive kode til hver enkelt forretningssag. Det kræver et infrastrukturlag, der forstår datastrukturen, -modellen og -logikken uden yderligere instruktioner (figur 4.1-15). Det er strukturerede data, der gør denne tilgang mulig: formularer, tabeller, filtre og visninger kan genereres automatisk med en minimal programmeringsindsats.

De vigtigste brugerkritiske grænseflader kan stadig kræve manuel finpudsning. Men i de fleste tilfælde - og det er mellem 50 og 90 procent af arbejdsscenarierne - er det tilstrækkeligt med automatisk generering af applikationer og beregninger uden brug af særlige applikationer til dette formål (Fig. 3.2-16), hvilket reducerer udviklings- og vedligeholdelsesomkostningerne betydeligt, reducerer fejl og fremskynder implementeringen af digitale løsninger.

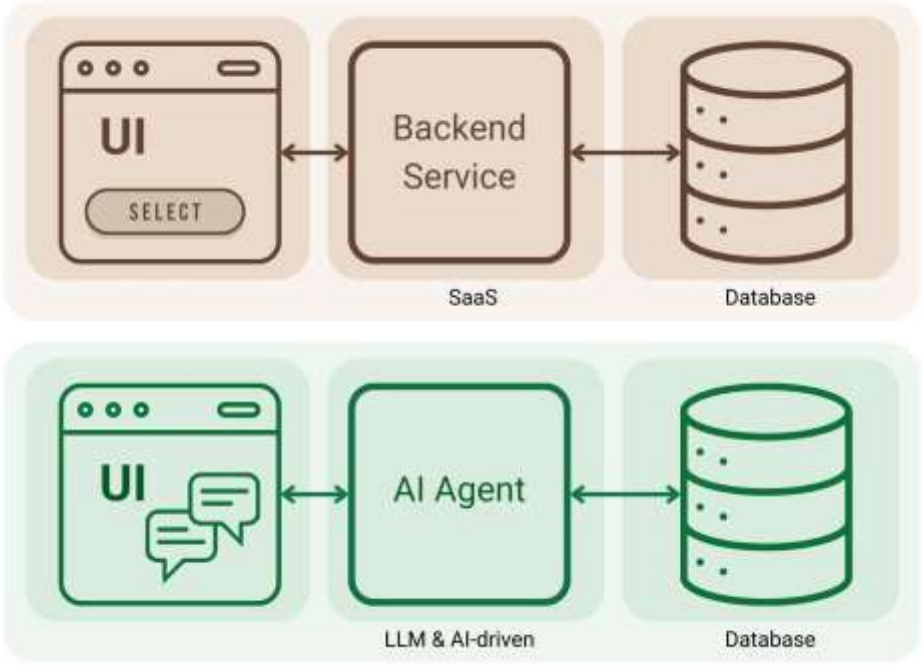
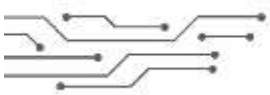


Fig. 3.2-16 Arkitekturmodeller til at arbejde med data: traditionel applikationsarkitektur og AI - orienteret model med LLM.

Skiftet fra arkitekturer bygget på individuelle applikationer til intelligent styrede systemer baseret på sprogmodeller (LLM'er) er det næste skridt i den digitale udvikling. I en sådan arkitektur bliver strukturerede data ikke kun et opbevaringsobjekt, men også grundlaget for interaktion med AI-værktøjer, der er i stand til at analysere, fortolke og anbefale handlinger baseret på kontekst.

I de følgende kapitler vil vi se på eksempler fra det virkelige liv på implementering af en arkitektur baseret på åbne strukturerede data og vise, hvordan sprogmodeller anvendes til automatisk at fortolke, validere og behandle data. Disse praktiske cases vil hjælpe dig til bedre at forstå, hvordan den nye digitale logik fungerer i praksis - og hvilke fordele den giver til virksomheder, der er klar til transformation.



KAPITEL 3.3.
LLM OG DERES ROLLE I DATABEHANDLING OG FORRETNINGSPROCESSER

LLM chatrum: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok for at automatisere dataprocesser.

Fremkomsten af Large Language Models (LLM'er) var en naturlig forlængelse af bevægelsen mod strukturerede åbne data og Open Source-filosofien. Når data bliver organiseret, tilgængelige og maskinlæsbare, er det næste skridt et værktøj, der kan interagere med disse oplysninger uden at skulle skrive kompleks kode eller have specialiseret teknisk viden.

> LLM'er er et direkte produkt af åbenhed: store åbne datasæt, publikationer og Open Source-bevægelsen. Uden åbne videnskabelige artikler, offentligt tilgængelige tekstdata og en kultur med fælles udvikling ville der ikke være nogen ChatGPT eller andre LLM'er. LLM er på en måde et "destillat" af menneskehedens akkumulerede digitale viden, indsamlet og uddannet gennem principperne om åbenhed.

Moderne store sprogmodeller (LLM - Large Language Models) som ChatGPT ® (OpenAI), LlaMa ™ (Meta AI), Mistral DeepSeek™, Grok ™ (xAI), Claude ™ (Anthropic), QWEN™ giver brugerne mulighed for at formulere forespørgsler til data på naturligt sprog. Det gør arbejdet med information tilgængeligt ikke kun for udviklere, men også for analytikere, ingeniører, planlæggere, ledere og andre fagfolk, der tidligere var langt fra programmering

> **LLM** (Large Language Model) er en kunstig intelligens, der er trænet til at forstå og generere tekst baseret på store mængder data indsamlet fra hele internettet. Den er i stand til at analysere kontekst, besvare spørgsmål, indgå i dialog, skrive tekster og generere softwarekode.

Hvis visualisering, behandling eller analyse af data tidligere krævede kendskab til et særligt programmeringssprog: Python, SQL, R eller Scala, samt evnen til at arbejde med biblioteker som Pandas, Polars eller DuckDB og mange andre, har situationen fra 2023 ændret sig radikalt. Nu kan brugeren blot beskrive, hvad han ønsker at opnå - og modellen genererer selv koden, udfører den, viser en tabel eller graf og forklarer resultatet. For første gang i årtier er den teknologiske udvikling ikke gået i retning af komplikationer, men i retning af radikal forenkling og tilgængelighed.

Dette princip - "behandle data med ord (prompts)". - markerede en ny fase i udviklingen af arbejdet med information og førte effektivt skabelsen af løsninger op på et endnu højere abstraktionsniveau. Ligesom det engang ikke længere var nødvendigt for brugerne at forstå det tekniske grundlag for internettet for at drive onlinebutikker eller oprette hjemmesider ved hjælp af WordPress, Joomla og andre open source-modulsystemer (av tor bogen har arbejdet med sådanne systemer siden 2005, herunder onlineplatforme til uddannelse og teknik). - Dette har igen ført til et boom i digitalt indhold og online-forretning - i dag kan ingeniører, analytikere og ledere automatisere arbejdsgange uden kendskab til programmeringssprog. Dette gøres lettere af kraftfulde LLM'er - både gratis og open source, såsom LLaMA, Mistral, Qwen, DeepSeek og andre -

der gør avancerede teknologier tilgængelige for det bredest mulige publikum.

Store LLM-sprogmodeller: Sådan fungerer det

Store sprogmodeller (ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok) er neurale netværk, der er trænet på enorme mængder tekstdata fra internettet, bøger, artikler og andre kilder. Deres hovedopgave er at forstå konteksten for menneskelig tale og generere meningsfulde svar.

Moderne LLM er baseret på Transformer-arkitekturen, der blev foreslået af Google-forskere i 2017 [85]. Nøglekomponenten i denne arkitektur er opmærksomhedsmekanismen, som gør det muligt for modellen at tage hensyn til relationer mellem ord uanset deres placering i teksten.

> LLM-læringsprocessen ligner til forveksling den måde, mennesker lærer et sprog på, bare millioner af gange større. Modellen analyserer milliarder af eksempler på ord og udtryk og identificerer mønstre i sprogets struktur og i logikken i semantiske overgange. Hele teksten opdeles i tokens - minimale semantiske enheder (ord eller dele af dem), som derefter omdannes til vektorer i et flerdimensionelt rum (fig. 8.2-2). Disse vektorrepræsentationer gør det muligt for maskinen at "forstå" de skjulte relationer mellem begreber i stedet for blot at arbejde med teksten som en række symboler.

Store sprogmodeller er ikke bare værktøjer til at generere tekst. De er i stand til at genkende mening, finde forbindelser mellem begreber og arbejde med data, selv om de præsenteres i forskellige formater. Det vigtigste er, at informationerne nedbrydes til forståelige modeller og repræsenteres som tokens, som LLM'en kan arbejde med.

Den samme tilgang kan anvendes på byggeprojekter. Hvis vi tænker på et projekt som en slags tekst, hvor hver bygning, element eller konstruktion er et symbol, kan vi begynde at behandle sådanne oplysninger på en lignende måde. Byggeprojekter kan sammenlignes med bøger, der er organiseret i kategorier, kapitler og grupper af afsnit, der består af minimale symboler - elementer i et byggeprojekt (figur 3.3-1). Ved at oversætte datamodeller til et struktureret format kan vi også oversætte strukturerede data til vektorbaser (fig. 8.2-2), som er en ideel kilde til maskinlæring og teknologier som LLM.

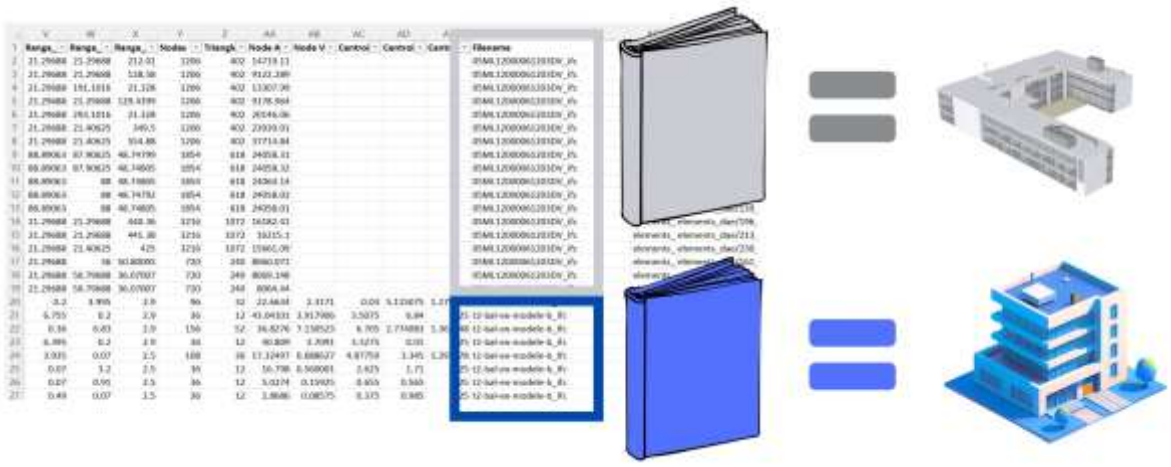


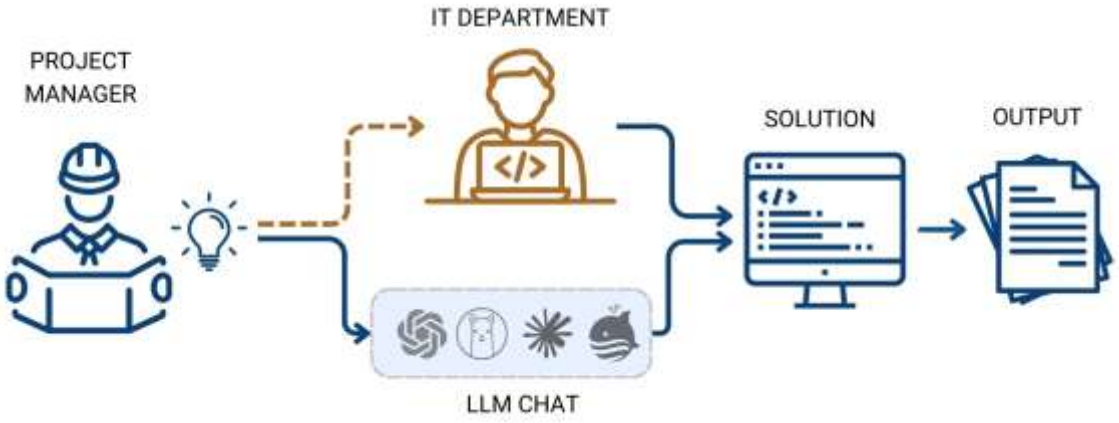
Fig. 3.3-1 Et byggeprojektelement er som et symbol i en tekst: en minimumsenhed, som grupper (afsnit) af sektioner (kategorier) af hele projektet dannes ud fra.

Hvis et byggeprojekt er digitaliseret, og dets elementer er repræsenteret som tokens eller vektorer, bliver det muligt at få adgang til dem i naturligt sprog i stedet for gennem stive formelle forespørgsler. Det er her, en af de vigtigste fordele ved LLM kommer i spil - evnen til at forstå betydningen af en forespørgsel og linke den til de relevante data.

Ingeniøren behøver ikke længere at skrive SQL -forespørgsel eller Python-kode for at få de nødvendige data - han kan blot, med forståelse for LLM og datastrukturen, formulere opgaven på den sædvanlige måde: "*Find alle armerede betonkonstruktioner med en betonklasse højere end B30 og beregn deres samlede volumen*". Modellen vil genkende betydningen af forespørgslen, omdanne den til en maskinlæsbar form, finde dataene (gruppere og transformere) og returnere det endelige resultat.

Dokumenter, tabeller og projektmodeller konverteres til vektorrepræsentationer (indlejring) og gemmes i databasen. Når en bruger stiller et spørgsmål, konverteres forespørgslen også til en vektor, og systemet finder de mest relevante data. Dette gør det muligt for LLM ikke kun at stole på sin egen trænede viden, men også på faktiske virksomhedsdata, selv om de allerede er dukket op efter afslutningen af modeltræningen.

> En af de vigtigste fordele ved LLM i byggeriet er muligheden for at generere programkode. I stedet for at give den tekniske opgave til en programmør kan specialister beskrive opgaven i et naturligt sprog, og modellen vil skabe den nødvendige kode, som kan bruges (ved at kopiere den fra chatten) til at skabe kode til procesautomatisering. LLM -modeller giver specialister uden dyb viden om programmering mulighed for at bidrage til automatisering og forbedring af virksomhedens forretningsprocesser.



Figur 3.3-2 LLM'er giver brugerne mulighed for at skrive kode og få resultater uden behov for programmeringsfærdigheder.

> Ifølge en undersøgelse foretaget af Wakefield Research og sponsoreret af SAP i 2024 [36], som omfattede 300 topledere i virksomheder med en årlig omsætning på mindst 1 milliard dollars i USA: 52% af topledere stoler på AI til at analysere data og komme med anbefalinger til beslutningstagning. Yderligere 48% bruger AI til at identificere risici, der ikke tidligere er taget højde for, og 47% bruger AI til at foreslå alternative planer. Derudover bruger 40% AI til udvikling af nye produkter, budgetplanlægning og markedsundersøgelser. Undersøgelsen viste også den positive indvirkning af AI på privatlivet, hvor 39% af respondenterne rapporterede om forbedret balance mellem arbejde og privatliv, 38% om forbedret mental sundhed og 31% om lavere stressniveauer.

Men på trods af deres styrke er LLM'er stadig et værktøj, som det er vigtigt at bruge bevidst. Som enhver anden teknologi har de begrænsninger. Et af de mest kendte problemer er såkaldte "hallucinationer" - tilfælde, hvor modellen med sikkerhed producerer et plausibelt, men faktuelt forkert svar. Det er derfor vigtigt at forstå, hvordan modellen fungerer: Hvilke data og datamodeller den kan fortolke uden fejl, hvordan den fortolker forespørgsler, og hvor den får sine oplysninger fra. Det er også værd at huske, at LLM'ens viden er begrænset til datoen for dens uddannelse, og uden forbindelse til eksterne data tager modellen måske ikke højde for aktuelle normer, standarder, priser eller teknologier.

Løsningen på disse problemer er regelmæssigt at opdatere vektordatabaser, oprette forbindelse til relevante kilder og udvikle autonome AI - agenter, der ikke bare besvarer spørgsmål, men proaktivt bruger data til træning, styrer opgaver, identificerer risici, foreslår optimeringsmuligheder og overvåger projektets resultater.

> Overgangen til LLM -grænseflader i byggeriet er ikke bare en teknologisk nyhed. Det er et paradigmeskift, der fjerner barrierer mellem mennesker og data. Det er evnen til at arbejde med information lige så let, som vi taler med hinanden - og stadig få nøjagtige, verificerede og brugbare resultater.

De virksomheder, der begynder at bruge sådanne værktøjer tidligere end andre, vil få en betydelig konkurrencefordel. Det omfatter hurtigere arbejde, lavere omkostninger og bedre kvalitet af designløsninger gennem hurtig adgang til dataanalyse og mulighed for hurtigt at finde svar på komplekse spørgsmål. Men der

er også sikkerhedsspørgsmål, der skal overvejes. Brugen af cloud-baserede LLM -tjenester kan være forbundet med risiko for datalækage. Derfor er organisationer i stigende grad på udkig efter alternative løsninger, der giver dem mulighed for at implementere LLM-værktøjer i deres egen infrastruktur - lokalt, med fuld beskyttelse og kontrol over oplysningerne.

Brug af lokale LLM'er til følsomme virksomhedsdata

Fremkomsten af de første chat-LLM'er i 2022 markerede en ny fase i udviklingen af kunstig intelligens. Men umiddelbart efter den udbredte anvendelse af disse modeller opstod der et legitimt spørgsmål: Hvor sikkert er det at overføre virksomhedsrelaterede data og forespørgsler til skyen? De fleste cloud-baserede sprogmodeller gemte kommunikationshistorik og uploadede dokumenter på deres servere, og for virksomheder, der håndterer følsomme oplysninger, var dette en alvorlig barriere for indførelse af AI.

> En af de mest bæredygtige og logiske løsninger på dette problem har været udrulningen af Open Source LLM lokalt i virksomhedens it-infrastruktur. I modsætning til cloud-tjenester fungerer lokale modeller uden internetforbindelse, overfører ikke data til eksterne servere og giver virksomhederne fuld kontrol over informationerne

> *Den bedste åbne model [Open Source LLM] kan i øjeblikket sammenlignes med lukkede modeller [såsom ChatGPT, Claude], men med en forsinkelse på omkring et år 77].*
>
> - Ben Cottier, ledende forsker ved Epoch AI, en non-profit forskningsorganisation, 2024

Store teknologivirksomheder er begyndt at stille deres LLM'er til rådighed for lokal brug. Metas open source LLaMA-serie og det hurtigt voksende DeepSeek -projekt fra Kina var eksempler på overgangen til åben arkitektur. Sideløbende med dem har Mistral og Falcon også frigivet kraftfulde modeller uden begrænsninger fra proprietære platforme. Disse initiativer har ikke bare fremskyndet udviklingen af global AI, men har også givet privatlivsbevidste virksomheder reelle alternativer til uafhængighed, fleksibilitet og sikkerhedsoverholdelse.

I et virksomhedsmiljø, især i byggebranchen, er databeskyttelse ikke bare et spørgsmål om bekvemmelighed, men om overholdelse af lovgivningen. Arbejdet med udbudsdokumenter, overslag, tegninger og fortrolig korrespondance kræver streng kontrol. Og det er her, at det lokale LLM giver den nødvendige sikkerhed for, at data forbliver inden for virksomhedens grænser.

	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

Fig. 3.3-3 Lokale modeller giver fuld kontrol og sikkerhed, mens cloud-baserede løsninger giver nem integration og automatiske opdateringer.

Vigtige fordele ved lokal Open Source LLM:

- Fuldstændig kontrol over data. Alle oplysninger forbliver inden for virksomheden, hvilket eliminerer uautoriseret adgang og datalækage.
- Offline drift. Ingen afhængighed af internetforbindelse, hvilket er særligt vigtigt for isolerede it-infrastrukturer. Det sikrer også uafbrudt drift i tilfælde af sanktioner eller blokerede cloud-tjenester.
- Fleksibilitet i anvendelsen. Modellen kan bruges til tekstgenerering, dataanalyse, skrivning af programkode, designstøtte og styring af forretningsprocesser.
- Tilpasning til virksomhedens mål. LLM kan trænes på interne dokumenter, hvilket gør det muligt at tage hensyn til de særlige forhold, der gør sig gældende for virksomhedens arbejde og dens branche. Den lokale LLM kan forbindes med CRM, ERP eller BI-platforme, så du kan automatisere

analysen af kundeforespørgsler, oprettelsen af rapporter eller endda trendprognoser.

> Implementering af DeepSeeks gratis og open source-model -R1-7B på en server med adgang for et helt team af brugere til en pris på 1000 dollars om måneden kan potentielt koste mindre end årlige gebyrer for cloud-API'er, såsom ChatGPT eller Claude og giver virksomheder mulighed for at tage fuld kontrol over deres data, eliminerer overførslen til internettet og hjælper med at overholde lovkrav såsom GDPR

I andre brancher er lokale LLM'er allerede i gang med at ændre deres tilgang til automatisering. I supportafdelinger svarer de på hyppige kundeforespørgsler og reducerer operatørernes arbejdsbyrde. I HR-afdelinger analyserer de CV'er og udvælger relevante kandidater. I e-handel genererer de personlige tilbud uden at afsløre brugerdata.

En lignende effekt forventes i byggesektoren. Takket være integrationen af LLM med projektdata og standarder er det muligt at fremskynde udarbejdelsen af dokumentation, automatisere udarbejdelsen af estimater og forudsigelige omkostningsanalyser. Brugen af LLM i forbindelse med strukturerede tabeller og datarammer er ved at blive et særligt lovende område.

Fuld kontrol over AI i virksomheden, og hvordan du implementerer din egen LLM

Moderne værktøjer gør det muligt for virksomheder at implementere en stor sprogmodel (LLM) lokalt på bare et par timer. Det giver fuld kontrol over data og infrastruktur, eliminerer afhængigheden af eksterne cloud-tjenester og minimerer risikoen for informationslækage. Denne løsning er især relevant for organisationer, der arbejder med følsom projektdokumentation eller kommercielt følsomme data.

Afhængigt af opgaverne og ressourcerne er der forskellige implementeringsscenarier til rådighed, fra færdige løsninger til mere fleksible og skalerbare arkitekturer. Et af de nemmeste værktøjer er Ollama, som giver dig mulighed for at køre sprogmodeller bogstaveligt talt med et enkelt klik uden behov for dyb teknisk viden. En hurtig start med Ollama:

1. Download distributionen til dit operativsystem (Windows / Linux / macOS) fra den officielle hjemmeside: ollama.com
2. Installer modellen via kommandolinjen. For eksempel for *Mistral-modellen*:

```
ollama kører mistral
```

3. Når du har kørt modellen, er den klar til at arbejde - du kan sende tekstforespørgsler via terminalen eller integrere den i andre værktøjer. Kør modellen, og udfør en forespørgsel:

```
ollama run mistral "Hvordan opretter man en beregning med alle ressourcer til arbejdet med at installere en 100 mm bred gipspladevæg?".
```

For dem der foretrækker at arbejde i et velkendt visuelt miljø, er der LM Studio, et gratis program med en grænseflade, der minder om ChatGPT:

- Installer LM Studio ved at downloade distributionskittet fra den officielle hjemmeside - lmstudio.ai
- Vælg en model (f.eks. Falcon eller GPT-Neo-X) via det indbyggede katalog, og download den.
- Arbejd med modellen via en intuitiv grænseflade, der minder om ChatGPT, men som er helt lokal.

	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

Figur 3.3-4 Sammenligning af populære lokale open source LLM -modeller.

Valget af model afhænger af kravene til hastighed, nøjagtighed og tilgængelige hardwarekapaciteter (figur 3.3-4). Små modeller som Mistral 7B og Baichuan 7B er velegnede til lette opgaver og mobile enheder, mens kraftfulde modeller som DeepSeek -V3 kræver betydelige beregningsressourcer, men tilbyder høj ydeevne og understøttelse af flere sprog. I de kommende år vil LLM-markedet vokse hurtigt - vi vil se flere og flere lette og specialiserede modeller. I stedet for generelle LLM'er, der dækker alt menneskeligt indhold, vil der dukke modeller op, der er trænet i snæver domæneekspertise. For eksempel kan vi forvente at se fremkomsten af modeller, der udelukkende er designet til at håndtere ingeniørberegninger, byggeoverslag eller data i CAD-formater. Sådanne specialiserede modeller vil være hurtigere, mere præcise og sikrere at bruge - især i professionelle miljøer, hvor høj pålidelighed og faglig dybde er vigtig.

Når den lokale LLM er lanceret, kan den tilpasses til virksomhedens specifikke opgaver. Til dette formål bruges finjusteringsteknikken, hvor modellen trænes yderligere på interne dokumenter, tekniske instruktioner, kontraktskabeloner eller projektdokumentation.

RAG: Intelligent LLM -assistenter med adgang til virksomhedsdata

Den næste fase i udviklingen af LLM-anvendelse i erhvervslivet er integrationen af modeller med faktiske virksomhedsdata i realtid. Denne tilgang kaldes RAG (Retrieval-Augmented Generation) - Retrieval-Augmented Generation. I denne arkitektur bliver sprogmodellen ikke bare en dialoggrænseflade, men en fuldgyldig intelligent assistent, der er i stand til at navigere i dokumenter, tegninger og databaser og give præcise, kontekstuelle svar.

> Den største fordel ved RAG er muligheden for at bruge interne virksomhedsdata uden behov for at træne modellen på forhånd, samtidig med at man bevarer en høj nøjagtighed og fleksibilitet i informationshåndteringen.

RAG -teknologien kombinerer to hovedkomponenter:

- Hentning: Modellen opretter forbindelse til datalagre - dokumenter, tabeller, PDF - filer, tegninger - og henter relevante oplysninger som ønsket af brugeren.
- Augmented Generation: Baseret på de udtrukne data genererer modellen et præcist, informeret svar, der tager højde for konteksten og forespørgslens specificitet.

For at køre LLM med RAG-support er der et par trin, der skal følges:

- **Dataforberedelse**: indsaml de nødvendige dokumenter, tegninger, specifikationer, tabeller. De kan være i forskellige formater og strukturer, fra PDF til Excel.
- **Indeksering og vektorisering**: Ved hjælp af værktøjer som LlamaIndex eller LangChain konverteres data til vektorrepræsentationer, der gør det muligt at finde semantiske links mellem tekstfragmenter (mere om vektordatabaser og oversættelse af store arrays til vektorrepræsentation, herunder CAD projekter, i del 8).
- **Spørg assistenten**: Når dataene er blevet uploadet, kan du stille modellen spørgsmål, og den vil søge efter svar inden for virksomhedens rammer i stedet for i generel viden indsamlet fra internettet.

Antag, at en virksomhed har en mappe constructionsite_docs, hvor kontrakter, instruktioner, overslag og tabeller er gemt. Ved hjælp af et Python-script (fig. 3.3-5) kan vi scanne denne mappe og opbygge en vektorindeksering: Hvert dokument konverteres til et sæt vektorer, der afspejler tekstens semantiske indhold. Dette gør dokumenterne til en slags "kort over betydninger", som modellen effektivt kan navigere på og finde forbindelser mellem termer og sætninger.

For eksempel "husker" modellen, at ordene "returnering" og "reklamation" ofte findes i den del af kontrakten, der handler om forsendelse af materialer til byggepladsen. Hvis der så stilles et spørgsmål - for eksempel "Hvad er vores returperiode?" (Fig. 3.3-5 - linje 11 i koden) - vil LLM'en analysere interne dokumenter og finde de nøjagtige oplysninger, idet den fungerer som en intelligent assistent, der er i stand til at læse

og forstå indholdet af alle virksomhedens filer.

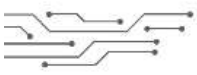


Fig. 3.3-5 LM læser en mappe med filer - på samme måde som en person åbner den og søger efter ønskede dokument.

Koden kan køres på enhver computer med Python installeret. Vi vil tale mere om at bruge Python og IDE'er til at køre koden i næste kapitel.

Lokal implementering af LLM er ikke bare en trend, men en strategisk løsning for virksomheder, der sætter pris på sikkerhed og fleksibilitet. Implementering af LLM, uanset om det er på lokale virksomhedscomputere eller ved hjælp af onlineløsninger, er dog kun det første skridt. For at kunne anvende LLM-funktioner til opgaver i den virkelige verden skal virksomheder bruge værktøjer, der ikke kun giver dem mulighed for at modtage chatsvar, men også gemme den logik, der er skabt i form af kode, som kan køres uden for den kontekst, hvor LLM bruges. Det er vigtigt for at kunne skalere løsningerne - korrekt organiserede processer gør det muligt at anvende AI-udvikling på flere projekter på én gang eller endda på tværs af hele virksomheden.

I denne sammenhæng spiller valget af et passende udviklingsmiljø (IDE) en vigtig rolle. Moderne programmeringsværktøjer gør det ikke kun muligt at udvikle LLM-baserede løsninger, men også at integrere dem i eksisterende forretningsprocesser og gøre dem til automatiserede ETL -Pipeline



KAPITEL 3.4.
IDE MED LLM-SUPPORT OG FREMTIDIGE PROGRAMÆNDRINGER

Valg af IDE: fra LLM eksperimenter til forretningsløsninger

Når man dykker ned i en verden af automatisering, dataanalyse og kunstig intelligens - især når man arbejder med store sprogmodeller (LLM'er) - er det vigtigt at vælge det rigtige integrerede udviklingsmiljø (IDE). Dette IDE bliver dit vigtigste arbejdsredskab: det sted, hvor den kode, der genereres af LLM'en, køres, både på en lokal computer og i virksomhedens netværk. Valget af IDE afgør ikke kun, hvor nemt det er at arbejde, men også hvor hurtigt du kan gå fra eksperimentelle LLM-forespørgsler til fuldgyldige løsninger, der er indlejret i virkelige forretningsprocesser.

> **IDE** (Integrated Development Environment) er en alsidig byggesten på din computer til procesautomatisering og databehandling. I stedet for at have en sav, en hammer, en boremaskine og andre værktøjer hver for sig, har du én enhed, der kan det hele - skære, fastgøre, bore og endda kontrollere materialernes kvalitet. IDE for programmører er et enkelt sted, hvor du kan skrive kode (i analogi med byggeri - lave tegninger), teste dets arbejde (samling af bygningsmodeller), finde fejl (som at kontrollere styrken af strukturer i byggeriet) og køre det færdige projekt (ibrugtagning af huset).

En oversigt over populære IDE'er:

- **PyCharm**® (JetBrains) er et stærkt professionelt IDE til Python. Det er velegnet til seriøse projekter på grund af det store antal indbyggede funktioner. Grundlæggende understøttelse af interaktive Jupyter-filer (IPYNB) er dog kun tilgængelig i betalingsversionen, og begyndere kan finde grænsefladen overvældende.

> En fil med IPYNB-udvidelsen (Interactive Python Notebook) er et format til interaktive Jupyter®-notesbøger (figur 3.4-1), hvor kode, visualiseringer og forklaringer er kombineret i et enkelt dokument. Dette format er ideelt til opbygning af rapporter, analyser og træningsscenarier.

- **VS Code**® (Microsoft) er et hurtigt, fleksibelt og brugertilpasset værktøj med gratis IPYNB-understøttelse og mange plug-ins. Velegnet til både begyndere og professionelle. Tillader integration af GitHub Copilot og sprogmodel-plugins, hvilket gør det til et godt valg til AI og datavidenskabelige projekter.
- **Jupyter Notebook** - Et klassisk og populært valg til eksperimenter og læring. Det giver dig mulighed for at skrive kode, tilføje forklaringer og visualisere resultater i en enkelt grænseflade (fig. 3.4-1). Ideel til hurtigt at teste hypoteser, arbejde med LLM og skabe reproducerbare vilde dataanalysetrin. For at styre afhængigheder og biblioteker anbefaler vi at bruge Anaconda Navigator, en visuel grænseflade til styring af Python -miljøet.

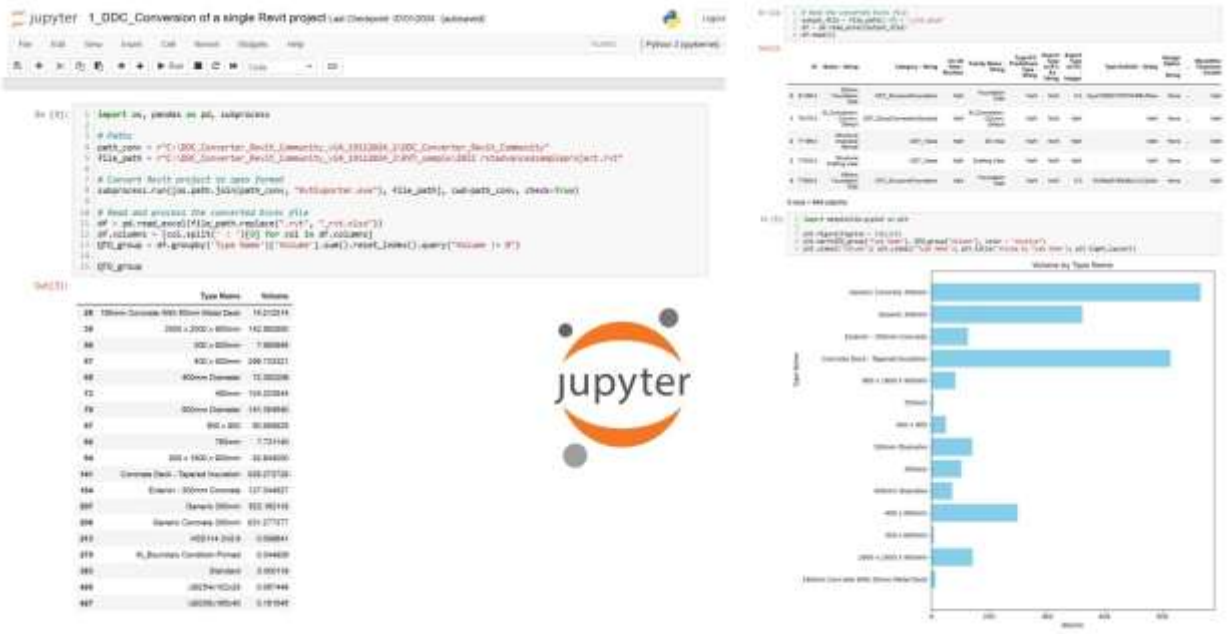


Fig. 3.4-1 Jupyter Notebook et af de mest praktiske og populære værktøjer til at skabe Pipeline processer.

- **Google Collab**™ (og Kaggle-platformen (figur 9.2-5)) er et skybaseret alternativ til Jupyter, der giver gratis GPU/TPU-adgang. Det er en god løsning til at komme i gang - ingen lokal softwareinstallation og mulighed for at arbejde direkte fra en browser. Det understøtter integration med Google Drive og for nylig med Gemini (Googles LLM).

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

Figur 3.4-2 IDE-sammenligning: Jupyter Notebook et af de mest praktiske og nemme værktøjer til at skabe Pipeline processer.

Valget af IDE afhænger af dine opgaver. Hvis du hurtigt vil i gang med at arbejde med AI, kan du prøve Jupyter Notebook eller Google Collab. Til seriøse projekter er det bedre at bruge PyCharm eller VS Code. Det vigtigste er at komme i gang. Med moderne værktøjer kan du hurtigt forvandle dine eksperimenter til fungerende løsninger.

Alle de IDE'er, der er beskrevet ovenfor, giver dig mulighed for at oprette databehandlingspipelines - det vil sige kæder af kodeblokmoduler (som kan genereres af LLM), som hver især er ansvarlige for f.eks. et forskelligt trin:

- analytiske scenarier,
- kæder af informationsudtræk fra dokumenter,
- automatiske reaktioner baseret på RAG,
- generering af rapporter og visualiseringer.

Takket være den modulære struktur kan hvert trin repræsenteres som en separat blok: dataindlæsning→ filtrering→ analyse→ visualisering→ eksport af resultater. Disse blokke kan genbruges, - tilpasses og samles i nye kæder som en konstruktør, bare til data.

For ingeniører, ledere og analytikere åbner det op for muligheden for at dokumentere beslutningslogik i form af kode, der kan genereres med LLM. Denne tilgang hjælper med at fremskynde rutineopgaver, automatisere typiske operationer og skabe gentagelige processer, hvor hvert trin er klart dokumenteret og gennemsigtigt for alle teammedlemmer.

De automatiserede ETL-værktøjer Pipelines (fig. 7.2-3), Apache Airflow (fig. 7.4-4), Apache NiFi (fig. 7.4-5) og n8n (fig. 7.4-6) til opbygning af logikblokke til procesautomatisering vil blive diskuteret mere detaljeret i del 7 og del 8 af bogen.

IDE med LLM-understøttelse og fremtidige programmeringsændringer

Integrationen af kunstig intelligens i udviklingsprocesser er ved at ændre programmeringslandskabet. Moderne miljøer er ikke længere bare teksteditorer med syntaksfremhævning - de bliver til intelligente assistenter, der er i stand til at forstå projektlogik, færdiggøre kode og endda forklare, hvordan et bestemt kodefragment fungerer. Der kommer produkter på markedet, som bruger AI til at flytte grænserne for konventionel udvikling:

- **GitHub Copilot** (integreres med VS Code, PyCharm): AI -assistent, der genererer kode baseret på kommentarer eller delvise beskrivelser og omdanner tekstlige hints til færdige løsninger.
- **Cursor** (en fork af VS Code med AI -kernel): giver ikke kun mulighed for at skrive kode færdig, men også for at stille spørgsmål til projektet, lede efter afhængigheder og lære af kodebasen.
- **JetBrains AI Assistant**: et plugin til JetBrains IDE (inklusive PyCharm) med den funktion at forklare kompleks kode, optimering og testoprettelse.
- **Amazon CodeWhisperer**: en analog til Copilot med fokus på sikkerhed og support til Amazons AWS-tjenester.

Programmering vil undergå en dramatisk forandring i de kommende år. Hovedfokus vil skifte fra rutinemæssig kodeskrivning til design af modeller og dataarkitektur - udviklere vil være mere involveret i systemdesign, mens AI vil overtage skabelonopgaver: kodegenerering, test, dokumentation og grundlæggende funktioner. Fremtidens programmering er et samarbejde mellem mennesker og AI, hvor maskiner overtager den tekniske rutine, og mennesker fokuserer på kreativitet.

Programmering på naturligt sprog bliver hverdagskost. IDE-personalisering vil nå et nyt niveau udviklings-

miljøer vil lære at tilpasse sig brugerens arbejdsstil og virksomheder ved at forudse mønstre, tilbyde kontekstuelle løsninger og lære af tidligere projekter.

> Det afskaffer ikke udviklerens rolle, men det vil ændre den radikalt: fra at skrive kode til viden-, kvalitets- og processtyring. Denne udvikling vil også påvirke business intelligence, hvor udarbejdelsen af rapporter, visualiseringer og beslutningsstøtteapplikationer i stigende grad vil ske gennem generering af kode og logik ved hjælp af AI og LLM, chat- og agentgrænseflader.

Når en virksomhed har oprettet LLM-chats og valgt et passende udviklingsmiljø, er det næste vigtige skridt at organisere dataene. Denne proces indebærer at udtrække oplysninger fra forskellige kilder, rense dem, omdanne dem til en struktureret form og integrere dem i virksomhedens systemer.

I en moderne datacentreret tilgang til datahåndtering er et vigtigt mål at bringe data ind i en enkelt universel form, der er kompatibel med et stort antal værktøjer og applikationer. Der er brug for specialiserede biblioteker til at håndtere struktureringsprocesser og strukturerede data. Et af de mest kraftfulde, fleksible og populære er Pandas-biblioteket for Python. Det giver dig mulighed for nemt at behandle tabeldata: filtrere, gruppere, rense, tilføje, aggregere og rapportere.

Python Pandas: et uundværligt værktøj til at arbejde med data

Pandas indtager en særlig plads i verdenen af dataanalyse og automatisering. Det er et af de mest populære og udbredte biblioteker i programmeringssproget Python [86], designet til at arbejde med strukturerede data.

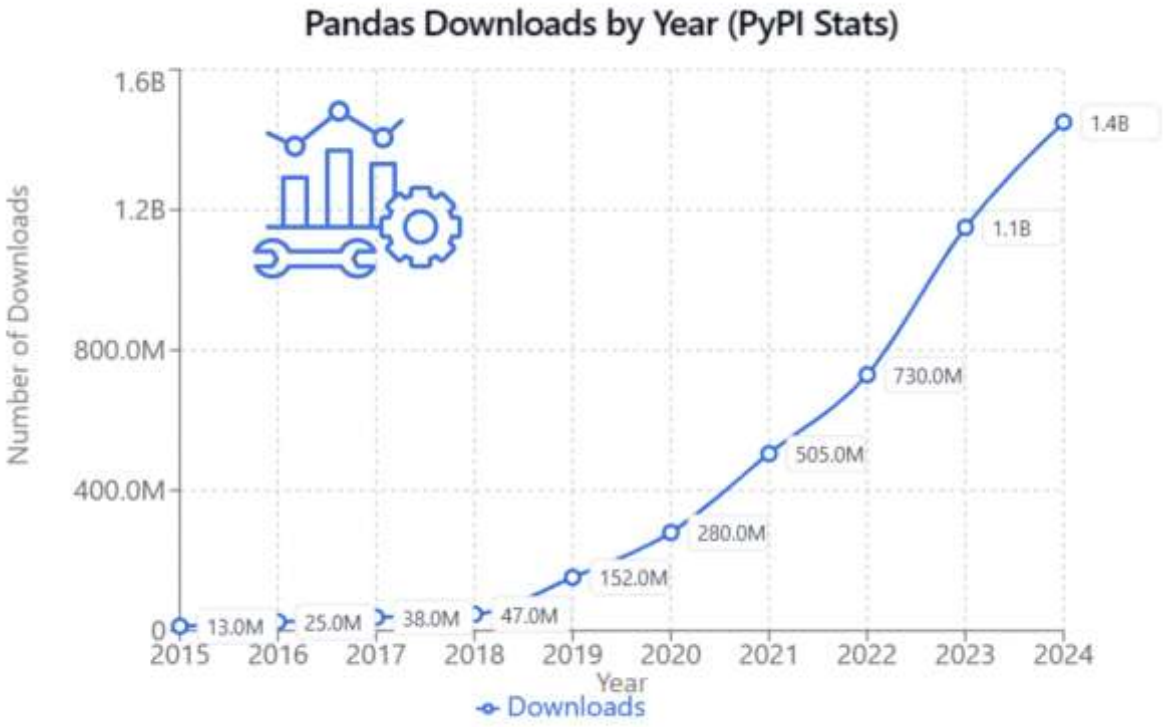
Et bibliotek er som et sæt færdige værktøjer: funktioner, moduler, klasser. Ligesom man på en byggeplads ikke behøver at opfinde en hammer eller et vaterpas hver gang, giver biblioteker inden for programmering mulighed for hurtigt at løse problemer uden at skulle genopfinde grundlæggende funktioner og løsninger.

> Pandas er et open source Python-bibliotek , der leverer højtydende og intuitive datastrukturer, især DataFrame, et universelt format til at arbejde med tabeller. Pandas er en schweizerkniv for analytikere, ingeniører og udviklere, der arbejder med data.

Python er et programmeringssprog på højt niveau med en enkel syntaks, der bruges aktivt inden for analyse, automatisering, maskinlæring og webudvikling. Dets popularitet skyldes, at det er let at læse, kan bruges på flere platforme og har et rigt økosystem af biblioteker. Til dato er der oprettet mere end 137.000 open source-pakker til Python [87], og dette antal fortsætter med at vokse næsten dagligt. Hvert af disse biblioteker er en slags lager af færdige funktioner: fra enkle matematiske operationer til komplekse værktøjer til billedbehandling, big data-analyse, neurale netværk og integration med eksterne tjenester.

> Forestil dig med andre ord, at du har fri og åben adgang til hundredtusindvis af færdige softwareløsninger - biblioteker og værktøjer, som du kan integrere direkte i dine forretningsprocesser. Det er som et stort katalog af applikationer til automatisering, analyse, visualisering, integration og meget mere - og alt sammen tilgængeligt umiddelbart efter installation af Python.

Pandas er en af de mest populære pakker i Python-økosystemet. I 2022 nåede det gennemsnitlige antal downloads af Pandas-biblioteket op på 4 millioner om dagen (figur 3.4-3), mens dette tal i begyndelsen af 2025 var steget til 12 millioner downloads om dagen, hvilket afspejler dets voksende popularitet og udbredte brug inden for dataanalyse og LLM-chat [86]



Figur 3.4-3 Pandas er et af de mest downloadede biblioteker. I 2024 oversteg det årlige antal downloads 1,4 milliarder.

Forespørgselssproget i Pandas-biblioteket har samme funktionalitet som SQL-forespørgselssproget, som vi diskuterede i kapitlet "Relationelle databaser og SQL-forespørgselssprog".

> I en verden af analyser og struktureret datahåndtering skiller Pandas sig ud ved sin enkelhed, hastighed og kraft, og giver brugerne en bred vifte af værktøjer til effektivt at analysere og behandle information.

Begge værktøjer - SQL og Pandas - giver kraftige datamanipulationsfunktioner, især sammenlignet med traditionel Excel. De understøtter operationer som udvælgelse og filtrering (figur 3.4-4), og den eneste forskel er, at SQL er optimeret til at arbejde med relationsdatabaser, mens Pandas behandler data i RAM, hvilket gør, at det kan køre på enhver computer uden behov for at oprette databaser og implementere en separat infrastruktur.



Fig. 3.4-4 Pandas har i modsætning til SQL fleksibiliteten til at arbejde med en række forskellige dataformater, der ikke er begrænset til databaser.

Pandas foretrækkes ofte til videnskabelig forskning, procesautomatisering, oprettelse af pipelines (herunder ETL) og datamanipulation i Python, mens SQL er en standard for databasestyring og ofte bruges i virksomhedsmiljøer til at håndtere store mængder data.

> Pandas-biblioteket i programmeringssproget Python giver dig mulighed for ikke kun at udføre grundlæggende operationer som at læse og skrive tabeller, men også mere komplekse opgaver, herunder fletning af data, gruppering af data og udførelse af komplekse analytiske beregninger.

I dag bruges biblioteket Pandas ikke kun i akademisk forskning og forretningsanalyse, men også i forbindelse med LLM -modeller. For eksempel var Meta®-divisionen (Facebook™), da den udgav en ny open source-model LlaMa 3.1 i 2024, særlig opmærksom på at arbejde med strukturerede data, hvilket gjorde en af de vigtigste og første cases i udgivelsen til netop behandling af strukturerede datarammer (fig. 3.4-5) i CSV-format og integration med Pandas-biblioteket direkte i chatten.

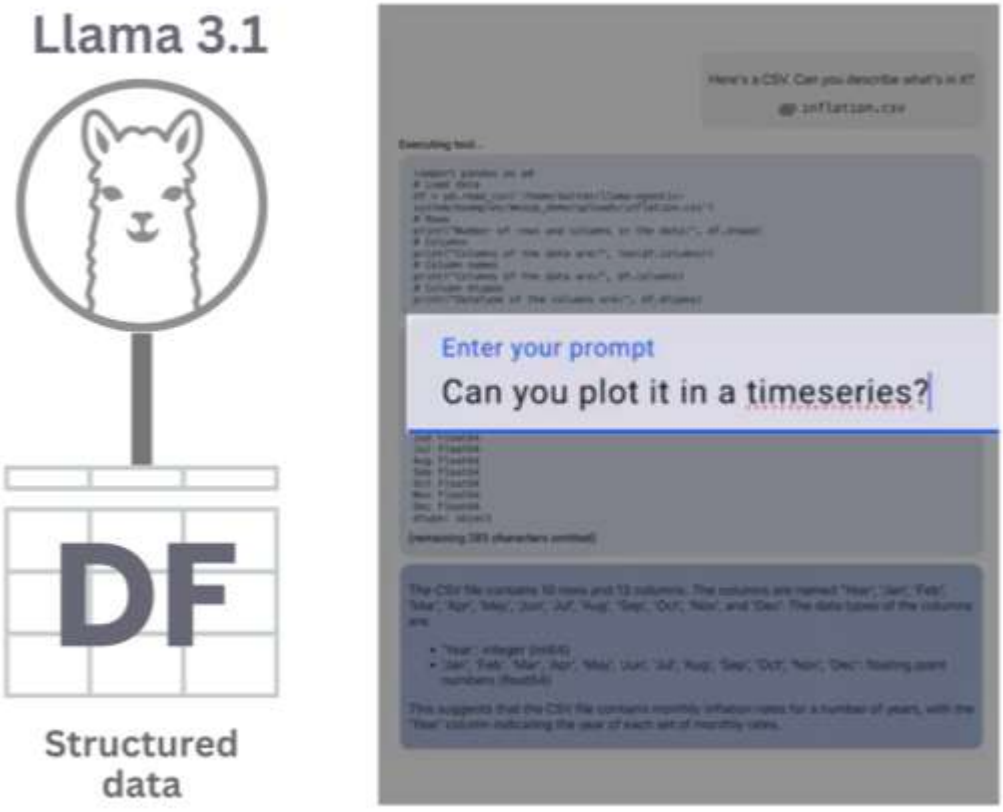


Fig. 3.4-5 En af Meta-teamets første og vigtigste cases, der blev præsenteret i LlaMa 3.1 i 2024, var at bygge applikationer ved hjælp af Pandas.

> *Pandas er et vigtigt værktøj for millioner af dataforskere, der behandler og forbereder data til generativ AI. Acceleration af Pandas med nul kodeændringer vil være et stort skridt fremad. Dataforskere vil kunne behandle data på få minutter i stedet for timer og få mange flere data til at træne generative AI-modeller [88].*
>
> - Jensen Huang, grundlægger og administrerende direktør for NVIDIA

Ved hjælp af Pandas er det muligt at håndtere og analysere datasæt, der ligger langt ud over mulighederne i Excel. Mens Excel typisk er i stand til at håndtere op til 1 million datarækker, kan Pandas nemt håndtere datasæt (figur 9.1-2, figur 9.1-10), der indeholder snesevis af millioner rækker [89]. Denne evne giver brugerne mulighed for at udføre sofistikeret dataanalyse og visualisering på store datasæt, hvilket giver dyb indsigt og letter datadrevet beslutningstagning. Derudover har Pandas en stærk samfundsstøtte [90]: hundredvis af millioner af udviklere og analytikere verden over (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) bruger det online eller offline hver dag, hvilket giver et stort antal out-of-the-box-løsninger til ethvert forretningsproblem.

Kernen i de fleste Python-analyseprocesser er en struktureret form for data kaldet DataFrame, som leveres af Pandas-biblioteket. Det er et kraftfuldt og fleksibelt værktøj til at organisere, analysere og visualisere

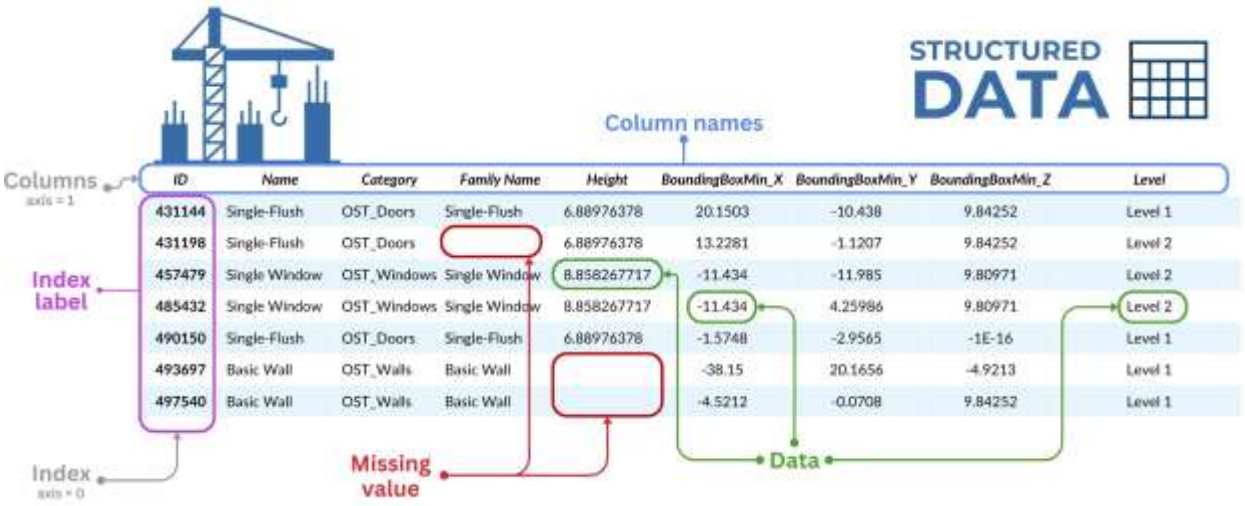
tabeldata.

DataFrame: universelt tabeldataformat

DataFrame er den centrale struktur i Pandas-biblioteket, som er en todimensionel tabel (figur 3.4-6), hvor rækkerne svarer til individuelle objekter eller poster, og kolonnerne svarer til deres egenskaber, parametre eller kategorier. Denne struktur ligner visuelt Excel-regneark, men er langt bedre, hvad angår fleksibilitet, skalerbarhed og funktionalitet.

> En **DataFrame** er en måde at repræsentere og behandle tabulære data, der er gemt i computerens RAM.

DataFrame er en måde at repræsentere og behandle tabeldata, der er gemt i computerens RAM. I en tabel kan rækkerne f.eks. afspejle elementer i et byggeprojekt, og kolonnerne deres egenskaber: kategorier, dimensioner, koordinater, omkostninger, vilkår og så videre. Desuden kan en sådan tabel indeholde både oplysninger om ét projekt (fig. 4.1-13) og data om millioner af objekter fra tusindvis af forskellige projekter (fig. 9.1-10). Takket være vektoriserede Pandas-operationer er det nemt at filtrere, gruppere og aggregere sådanne informationsmængder med høj hastighed.



Figur 3.4-6 Byggeprojekt som dataramme er en todimensionel tabel med elementer i rækker og attributter i kolonner.

Nvidia vurderer, at op til 30 procent af alle computerressourcer allerede i dag bruges til at behandle strukturerede data - dataframes - og denne andel fortsætter med at vokse.

> *Databehandling er, hvad sandsynligvis en tredjedel af verdens databehandling udføres i hver eneste virksomhed. De fleste virksomheders databehandling og data ligger i DataFrame, i tabelformat*
>
> - Jensen Huang, administrerende direktør for Nvidia [91].

Lad os liste nogle af de vigtigste funktioner i DataFrame i Pandas:

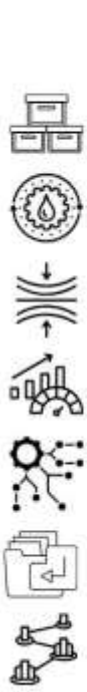
- Kolonner: I DataFrame er data organiseret i kolonner, som hver har et unikt navn. Attributkolonner kan indeholde data af forskellige typer, svarende til kolonner i databaser eller kolonner i tabeller.
- **Pandas Series** er en endimensionel datastruktur i Pandas, svarende til en liste eller kolonne i en tabel, hvor hver værdi svarer til et forskelligt indeks

> Pandas Series har over 400 attributter og metoder, som gør arbejdet med data utroligt fleksibelt. Du kan anvende en af de 400 tilgængelige funktioner direkte på en kolonne, udføre matematiske operationer, filtrere data, erstatte værdier, arbejde med datoer, strenge og meget mere. Desuden understøtter Series vektoriserede operationer, som i høj grad fremskynder behandlingen af store datasæt sammenlignet med cykliske beregninger. Du kan f.eks. nemt gange alle værdier med et tal, erstatte manglende data eller anvende komplekse transformationer uden at skrive komplicerede loops.

- **Rækker**: I DataFrame kan indekseres med unikke værdier. Dette indeks gør det muligt hurtigt at ændre og justere data i specifikke rækker.
- **Indeks**: Når du opretter en DataFrame Pandas, tildeles hver række som standard et indeks fra 0 til N-1 (hvor N er antallet af alle rækker i DataFrame). Indekset kan dog ændres til at omfatte særlige betegnelser som f.eks. datoer eller unikke egenskaber.
- **Indeksering af** rækker i en DataFrame betyder, at hver række tildeles et unikt navn eller en etiket, som kaldes DataFrame-indekset.
- Datatyper**:** DataFrame understøtter en række forskellige datatyper, herunder: `int`, `float`, `bool`, `datetime64` og `obect` til tekstdata. Hver DataFrame-kolonne har sin egen datatype, der bestemmer, hvilke operationer der kan udføres på dens indhold.
- **Dataoperationer**: DataFrame understøtter en bred vifte af operationer til databehandling, herunder aggregering (`groupby`), fletning (`merge` og `join`), sammenkædning (`concat`), split-apply-combine og mange andre datatransformationsteknikker.
- **Manipulation af størrelse**: DataFrame giver dig mulighed for at tilføje og fjerne kolonner og rækker, hvilket gør det til en dynamisk struktur, der kan ændres i henhold til dine behov for dataanalyse.
- **Visualisering af data: Ved hjælp af** indbyggede visualiseringsteknikker eller interaktion med populære datavisualiseringsbiblioteker som Matplotlib eller Seaborn kan DataFrame nemt konverteres til grafer og diagrammer for at præsentere data grafisk.
- **Datainput og -output:** Pandas har funktioner til at læse, importere og eksportere data til forskellige filformater som CSV, Excel, JSON, HTML og SQL, hvilket potentielt kan gøre DataFrame til et centralt knudepunkt for dataindsamling og -distribution.

I modsætning til CSV og XLSX giver Pandas DataFrame større fleksibilitet og ydeevne, når man arbejder med data: Den kan håndtere store mængder information i RAM, understøtter udvidede datatyper (herunder datoer, logiske værdier og tidsserier) og giver omfattende muligheder for filtrering, aggregering, sammenlægning og visualisering af data. Mens CSV ikke gemmer oplysninger om datatyper og -struktur, og XLSX ofte er overbelastet med formatering og har lav skalerbarhed, er DataFrame stadig det optimale valg til hurtig analyse, procesautomatisering og integration med AI -modeller (figur 3.4-7). De følgende kapitler vil udforske hvert af disse aspekter af data i detaljer, og i del 8 af bogen vil lignende formater som Parquet,

Apache Orc, JSON, Feather, HDF5 og datalagre blive diskuteret i detaljer (fig. 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)

Fig. 3.4-7 DataFrame er det optimale valg til datamanipulation med høj ydeevne og avanceret understøttelse af datatyper.

Takket være deres fleksibilitet, styrke og brugervenlighed er Pandas-biblioteket og DataFrame-formatet blevet de facto-standarden inden for Python-dataanalyse. De er ideelle til både at skabe enkle rapporter og opbygge komplekse analytiske pipelines, især i forbindelse med LLM-modeller.

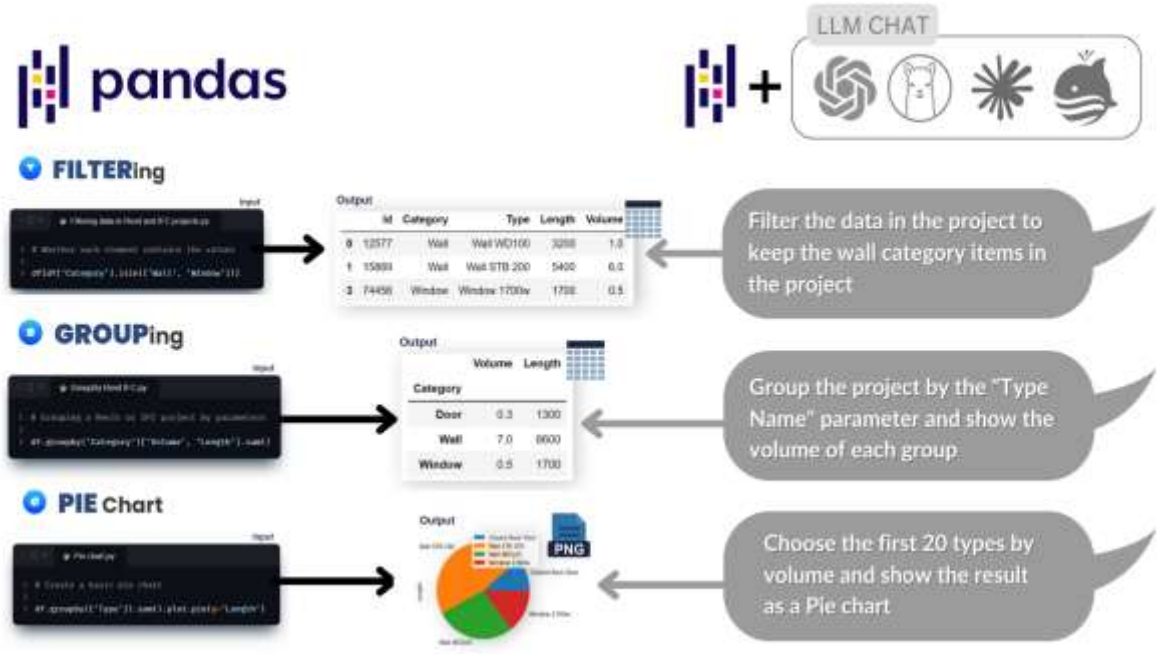


Fig. 3.4-8 LLM'er forenkler interaktionen med Pandas: I stedet for kode er det tilstrækkeligt med en tekstforespørgsel.

I dag bruges Pandas aktivt i LLM-baserede chatrum som ChatGPT, LlaMa, DeepSeek, QWEN og andre. I mange tilfælde, når en model modtager en forespørgsel relateret til tabelbehandling, datavalidering eller analyse, genererer den kode nøjagtigt ved hjælp af Pandas-biblioteket. Det gør DataFrame til et naturligt "sprog" til at repræsentere data i dialoger med AI (figur 3.4-8).

Moderne datateknologier som Pandas gør det lettere at analysere, automatisere og integrere data i forretningsprocesser. De leverer resultater hurtigt, reducerer arbejdsbyrden for specialister og sikrer gentagelige operationer.

Næste skridt: opbygning af en bæredygtig dataramme

I denne del så vi på de vigtigste typer data, der bruges i byggebranchen, lærte om de forskellige formater til lagring af dem og analyserede den rolle, som moderne værktøjer, herunder LLM og IDE'er, spiller i behandlingen af information. Vi lærte, at effektiv datastyring er grundlaget for informeret beslutningstagning og automatisering af forretningsprocesser. Organisationer, der er i stand til at strukturere og organisere deres data, får en betydelig konkurrencefordel i faserne med databehandling og -transformation.

For at opsummere denne del er det værd at fremhæve de vigtigste praktiske trin, der vil hjælpe dig med at anvende de diskuterede tilgange i dine daglige opgaver:

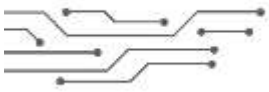
- Gennemfør en datarevision af dine processer
  - [ ] Lav en oversigt over alle de datatyper, du bruger i dine projekter
  - [ ] Find ud af, hvilke datatyper og modeller der er mest kritiske for dine forretningsprocesser

- [ ] Identificer problemområder, hvor information ofte forbliver ustruktureret, dårligt struktureret eller utilgængelig

- Begynd at formulere en strategi for datahåndtering
  - [ ] Rejser politiske spørgsmål og standarder for håndtering af forskellige datatyper
  - [ ] Analyser, hvilke af dine arbejdsgange der kan forbedres ved at konvertere ustrukturerede data til strukturerede data.
  - [ ] Opret en politik for datalagring og -adgang, der tager højde for sikkerhed og fortrolighed
- Installere og mestre grundlæggende værktøjer til at arbejde med data
  - [ ] Vælg et passende IDE, der passer til dine opgaver (installer f.eks. VS Code eller Jupyter Notebook)
  - [ ] Prøv at oprette en lokal LLM til at håndtere dine personlige data fortroligt
  - [ ] Begynd at eksperimentere med Pandas-biblioteket for at behandle XLSX-tabeldata
  - [ ] Beskriv for LLM de typiske opgaver, du håndterer i regnearksværktøjer eller databaser, og bed LLM om at automatisere arbejdet ved hjælp af Pandas.

Ved at anvende disse trin kan du gradvist ændre din tilgang til at arbejde med data og gå fra forskellige, ustrukturerede datasæt til et samlet økosystem, hvor data bliver et tilgængeligt og forståeligt aktiv. Start i det små - opret din første DataFrame i Pandas, kør en lokal LLM, automatiser din første rutineopgave ved hjælp af Python (f.eks. Excel-regneark).

Den fjerde del af bogen fokuserer på datakvalitet, dataorganisation, strukturering og modellering. Vi vil fokusere på metoder, der omdanner forskellige datakilder - fra PDF'er og tekster til billeder og CAD-modeller - til strukturerede datasæt, der egner sig til analyse og automatisering. Vi vil også undersøge, hvordan datakrav formaliseres, hvordan konceptuelle og logiske modeller opbygges i byggeprojekter, og hvordan moderne sprogmodeller (LLM'er) kan hjælpe i denne proces.





IV DEL
DATAKVALITET: ORGANISERING, STRUKTURERING, MODELLERING

Den fjerde del fokuserer på metoder og teknologier, der sikrer, at forskelligartet information omdannes til strukturerede datasæt af høj kvalitet. Processerne med at danne og dokumentere datakrav som grundlag for effektiv informationsarkitektur i byggeprojekter diskuteres i detaljer. Praktiske metoder til at udtrække struktureret information fra forskellige kilder (PDF -dokumenter, billeder, tekstfiler, CAD -modeller) med eksempler på implementering præsenteres. Brugen af regulære udtryk (RegEx) og andre værktøjer til automatisk validering og verificering af data analyseres. Processen med datamodellering på konceptuelt, logisk og fysisk niveau beskrives trin for trin under hensyntagen til de særlige forhold, der gør sig gældende i byggebranchen. Specifikke eksempler på brug af sprogmodeller (LLM) til at automatisere processerne med strukturering og validering af information demonstreres. Der foreslås effektive tilgange til visualisering af analyseresultater, hvilket øger tilgængeligheden af analytisk information på alle niveauer af byggeprojektledelsen

KAPITEL 4.1.
KONVERTERING AF DATA TIL EN STRUKTURERET FORM

I den datadrevne økonomis æra er data ved at blive grundlaget for beslutningstagning snarere end en hindring. I stedet for konstant at tilpasse information til hvert nyt system og dets formater forsøger virksomheder i stigende grad at skabe en enkelt struktureret datamodel, der fungerer som en universel sandhedskilde for alle processer. Moderne informationssystemer er ikke designet omkring formater og grænseflader, men omkring betydningen af data - fordi strukturen kan ændre sig, men betydningen af information forbliver den samme i meget længere tid.

> Nøglen til at arbejde effektivt med data er ikke at konvertere og transformere dem i en uendelighed, men at organisere dem korrekt fra starten: at skabe en universel struktur, der kan give gennemsigtighed, automatisering og integration i alle faser af projektets livscyklus.

Den traditionelle tilgang fremtvinger manuelle justeringer ved hver ny platformsimplementering: migrering af data, ændring af attributnavne, justering af formater. Disse trin forbedrer ikke kvaliteten af selve dataene, men maskerer kun problemer og skaber en ond cirkel af endeløse transformationer. Som følge heraf bliver virksomheder afhængige af specifikke softwareløsninger, og den digitale transformation går langsommere.

I de følgende kapitler vil vi se på, hvordan man strukturerer data korrekt og derefter skaber universelle modeller, minimerer platformsafhængighed og fokuserer på det, der betyder mest - data som en strategisk ressource, som bæredygtige processer bygges op omkring.

At lære at omdanne dokumenter, PDF, billeder og tekster til strukturerede formater

I byggeprojekter findes langt de fleste oplysninger i ustruktureret form: tekniske dokumenter, arbejdsbeskrivelser, tegninger, specifikationer, tidsplaner og protokoller. Deres forskellighed - både i format og indhold - gør det vanskeligt at integrere og automatisere dem.

> Processen med at konvertere til strukturerede eller semistrukturerede formater kan variere afhængigt af typen af inputdata og de ønskede behandlingsresultater.

At omdanne data fra ustruktureret til struktureret form er både en kunst og en videnskab. Denne proces varierer afhængigt af typen af inputdata og formålet med analysen og optager ofte en betydelig del af dataingeniørens (figur 3.2-5) og analytikerens arbejde, med det formål at producere et rent, organiseret datasæt.



Fig. 4.1-1 Konvertering af et ustruktureret indscannet dokument til et struktureret tabelformat.

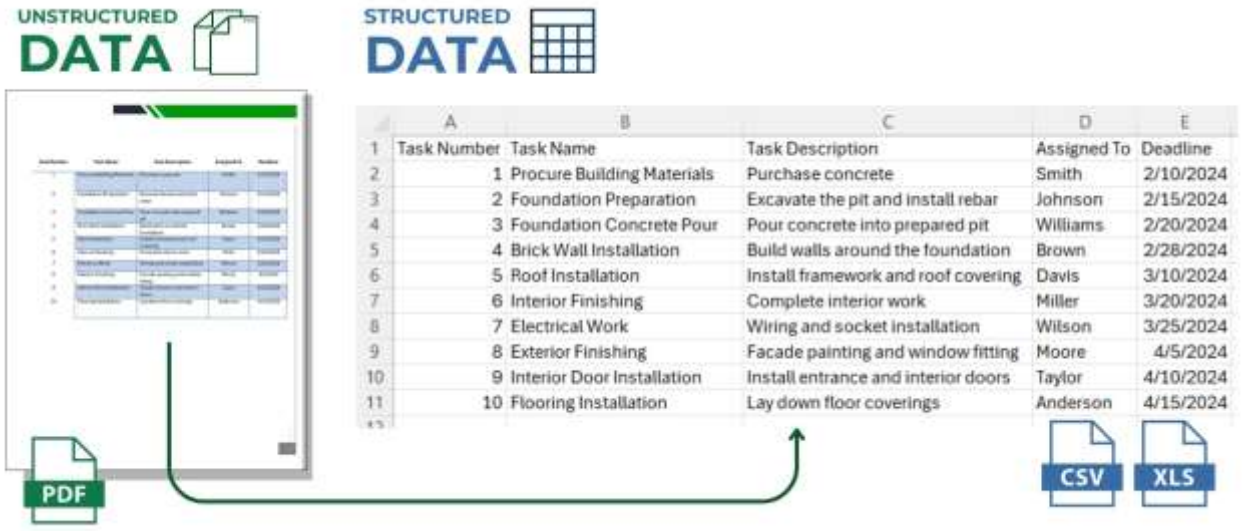
At omdanne dokumenter, PDF, billeder og tekster til et struktureret format (figur 4.1-1) er en trinvis proces, der omfatter følgende trin:

- Udtræk): I dette trin indlæses et kildedokument eller et billede, der indeholder ustrukturerede data. Det kan f.eks. være et PDF -dokument, et fotografi, en tegning eller et skema.
- **Datatransformation** (Transform): Herefter følger det trin, hvor ustrukturerede data omdannes til et struktureret format. Det kan f.eks. indebære genkendelse og fortolkning af tekst fra billeder ved hjælp af optisk tegngenkendelse (OCR) eller andre behandlingsmetoder.
- **Indlæsning og lagring af data** (Load): Det sidste trin omfatter lagring af de behandlede data i forskellige formater såsom CSV, XLSX, XML, JSON, til videre arbejde, hvor valget af format afhænger af specifikke krav og præferencer.

Denne proces, kendt som ETL (Extract, Transform, Load), spiller en nøglerolle i automatiseret databehandling og vil blive diskuteret mere detaljeret i kapitlet "ETL og Pipeline: Extract, Transform, Load". Dernæst vil vi se på eksempler på, hvordan dokumenter i forskellige formater omdannes til strukturerede data.

Eksempel på konvertering af et PDF -dokument til en tabel

En af de mest almindelige opgaver i byggeprojekter er at behandle specifikationer i PDF-format. For at demonstrere overgangen fra ustrukturerede data til et struktureret format, lad os se på et praktisk eksempel: at udtrække en tabel fra et PDF-dokument og konvertere den til CSV- eller Excel-format (Fig. 4.1-2).



	A	B	C	D	E
1	Task Number	Task Name	Task Description	Assigned To	Deadline
2	1	Procure Building Materials	Purchase concrete	Smith	2/10/2024
3	2	Foundation Preparation	Excavate the pit and install rebar	Johnson	2/15/2024
4	3	Foundation Concrete Pour	Pour concrete into prepared pit	Williams	2/20/2024
5	4	Brick Wall Installation	Build walls around the foundation	Brown	2/28/2024
6	5	Roof Installation	Install framework and roof covering	Davis	3/10/2024
7	6	Interior Finishing	Complete interior work	Miller	3/20/2024
8	7	Electrical Work	Wiring and socket installation	Wilson	3/25/2024
9	8	Exterior Finishing	Facade painting and window fitting	Moore	4/5/2024
10	9	Interior Door Installation	Install entrance and interior doors	Taylor	4/10/2024
11	10	Flooring Installation	Lay down floor coverings	Anderson	4/15/2024

Fig. 4.1-2 I modsætning til PDF er formaterne CSV og XLSX meget udbredte og kan nemt integreres i forskellige datahåndteringssystemer.

LLM-sprogmodeller, såsom ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN forenkler i høj grad den måde, dataforskere arbejder med data på, idet de reducerer behovet for dyb læring af programmeringssprog og gør det muligt at løse mange opgaver med tekstforespørgsler.

I stedet for at bruge tid på at søge efter løsninger på internettet (normalt StackOverFlow-webstedet eller tematiske fora og chats) eller kontakte databehandlingsspecialister, kan vi bruge funktionerne i moderne online eller lokale LLM'er. Det er nok at stille en forespørgsel, og modellen vil give klar kode til at konvertere PDF -dokumentet til et tabelformat.

- Send følgende tekstanmodning til enhver LLM -model (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller enhver anden):

> Skriv venligst en kode til at udtrække tekst fra en PDF -fil, der indeholder en tabel. Koden skal tage filstien som et argument og returnere den udpakkede tabel som en Data-Frame⏎

- LLM -modelsvaret vil i de fleste tilfælde være i form af kode i Python , da dette sprog er meget brugt til databehandling, automatisering og arbejde med forskellige filformater:

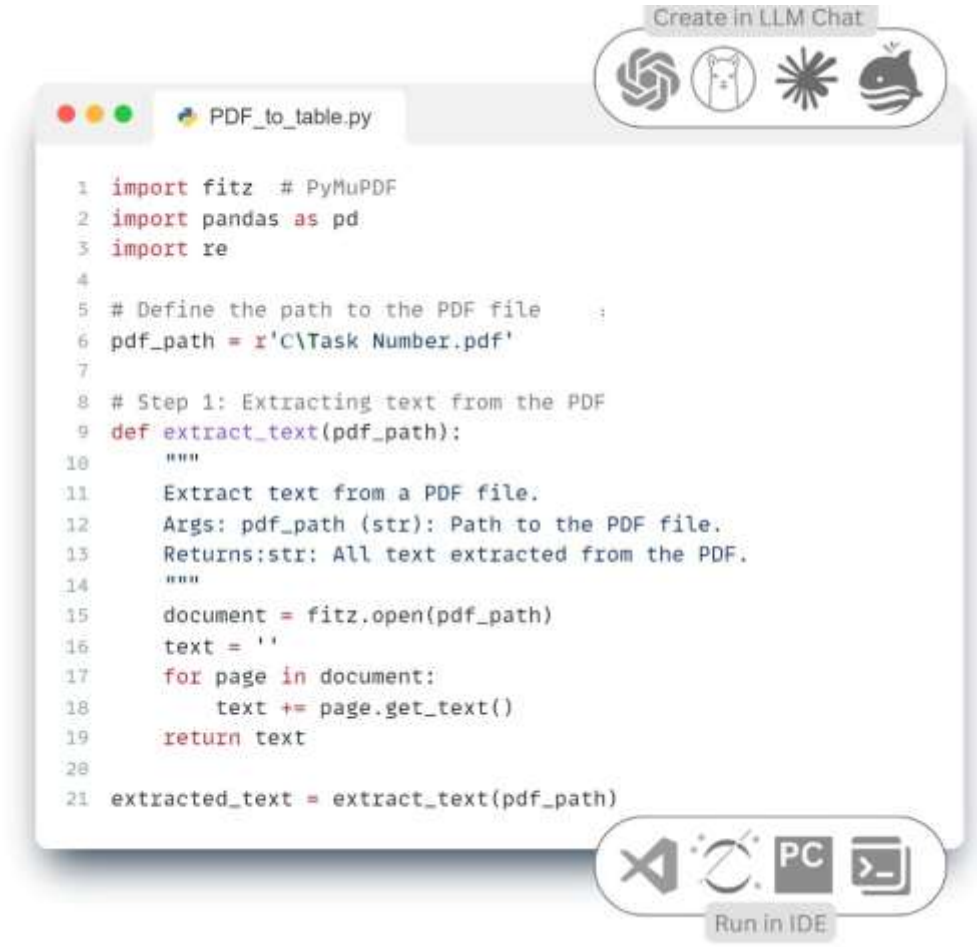


Fig. 4.1-3 LLM-svaret i form af Python-koden og de tilhørende biblioteker og pakker (Pandas, Fitz) udtrækker tekst fra en PDF -fil.

Denne kode (figur 4.1-3) kan køres offline i en af de populære IDE'er, vi nævnte ovenfor: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære onlineværktøjer: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- I trinnet "Konverter" bruger vi det populære Pandas-bibliotek (som vi diskuterede i detaljer i kapitlet "Python Pandas: et uundværligt værktøj til at arbejde med data") til at læse den udtrukne tekst ind i DataFrame og gemme DataFrame i en CSV-tabel-fil eller XLXS:

> Jeg har brug for kode, der konverterer den resulterende tabel fra en PDF -fil til en Data-Frame. Tilføj også kode til at gemme DataFrame til en CSV -fil ⏎

LLM's svar :

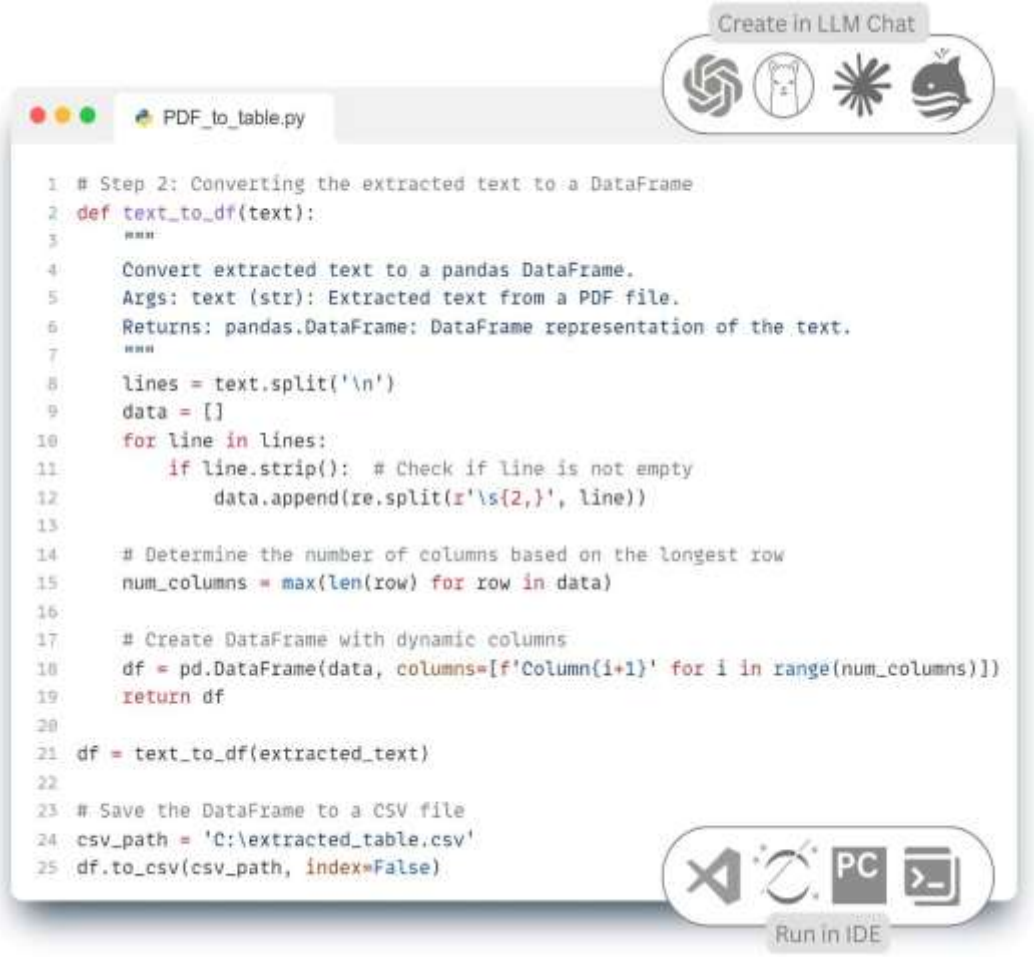


Fig. 4.1-4 Konvertering af den udtrukne tabel fra PDF til DataFrame og lagring af tabellen i en CSV -fil.

Hvis der opstår en fejl under udførelsen af koden (fig. 4.1-3, fig. 4.1-4) - f.eks. på grund af manglende biblioteker eller forkert filsti - kan fejlteksten blot kopieres sammen med kildekoden og indsendes til LLM - modellen. Modellen vil analysere fejlmeddelelsen, forklare, hvad problemet er, og foreslå rettelser eller yderligere trin.

> På den måde bliver interaktionen med AI LLM en komplet c→→test→ feedback→ korrektion - uden behov for dyb teknisk viden.

Ved hjælp af en almindelig tekstforespørgsel i LLM chat og et dusin linjer Python, som vi kan køre lokalt i ethvert IDE, konverterede vi et PDF -dokument til et tabelformet CSV-format, som i modsætning til et PDF-dokument er let maskinlæsbart og hurtigt kan integreres i ethvert datastyringssystem.

Vi kan anvende denne kode (figur 4.1-3, figur 4.1-4) ved at kopiere den fra et hvilket som helst LLM-chatrum til titusindvis af nye PDF-dokumenter på serveren og derved automatisere processen med at konvertere en strøm af ustrukturerede dokumenter til et struktureret CSV-tabelformat.

Men PDF-dokumenter indeholder ikke altid tekst, oftest er det scannede dokumenter, der skal behandles som billeder. Selv om billeder i sagens natur er ustrukturerede, giver udviklingen og anvendelsen af genkendelsesbiblioteker os mulighed for at udtrække, behandle og analysere deres indhold, så vi kan udnytte disse data fuldt ud i vores forretningsprocesser.

Konverter JPEG- og PNG-billeder til struktureret form

Billeder er en af de mest almindelige former for ustrukturerede data. I byggebranchen og mange andre brancher lagres en enorm mængde information i form af indscannede dokumenter, skemaer, fotografier og tegninger. Sådanne data indeholder værdifuld information, men kan ikke behandles direkte, som f.eks. et Excel-regneark eller en database. Billeder indeholder mange komplekse oplysninger, fordi deres indhold, farver og teksturer er varierede, og der kræves særlig behandling for at udtrække nyttige oplysninger.

Vanskeligheden ved at bruge billeder som datakilde er manglen på struktur. Billeder formidler ikke mening på en direkte, let kvantificerbar måde, som en computer umiddelbart kan forstå eller behandle, som et Excel-regneark eller en databasetabel gør. For at konvertere ustrukturerede billeddata til en struktureret form skal der bruges særlige biblioteker, der er i stand til at fortolke den visuelle information, de indeholder (figur 4.1-5).

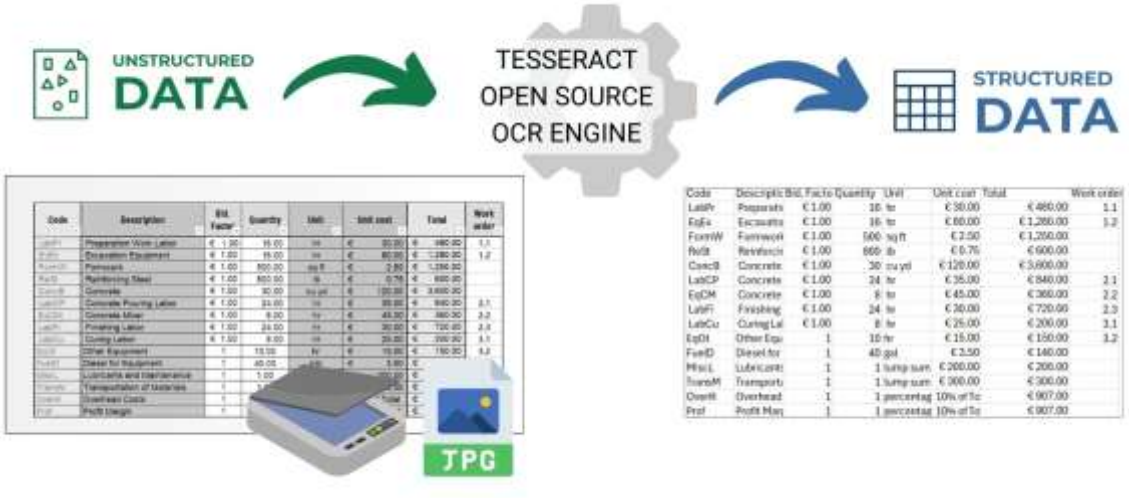


Fig. 4.1-5 Det er muligt at konvertere indscannede dokumenter og billeder til strukturerede formater med særlige OCR-værktøjer.

OCR (Optical Character Recognition) er en teknologi, der bruges til at udtrække tekst fra billeder. Den gør det muligt at genkende bogstaver og tal i dokumentscanninger, fotos og PDF -filer og omdanne dem til redigerbar og maskinlæsbar tekst. OCR-teknologier har længe været brugt til automatisering af dokumenter,

og i dag kan de nemt integreres i alle forretningsprocesser og Python -applikationer. Et af de mest populære OCR-værktøjer er Tesseract, en open source, som oprindeligt blev udviklet af HP™ og nu understøttes af Google™. Det understøtter over 100 sprog og har en høj genkendelsesnøjagtighed.

Lad os bede LLM chat om at skrive et kodeeksempel for at hente data fra en scannet eller fotograferet tabel på en struktureret måde.

- Send en tekstanmodning til LLM chat (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller andre):

> Skriv kode til at konvertere et JPEG-billede, der indeholder en tabel, til en DataFrame-tabel ⏎

- LLM's svar vil i de fleste tilfælde foreslå at bruge Pytesseract-biblioteket til at genkende tekst i billeder :

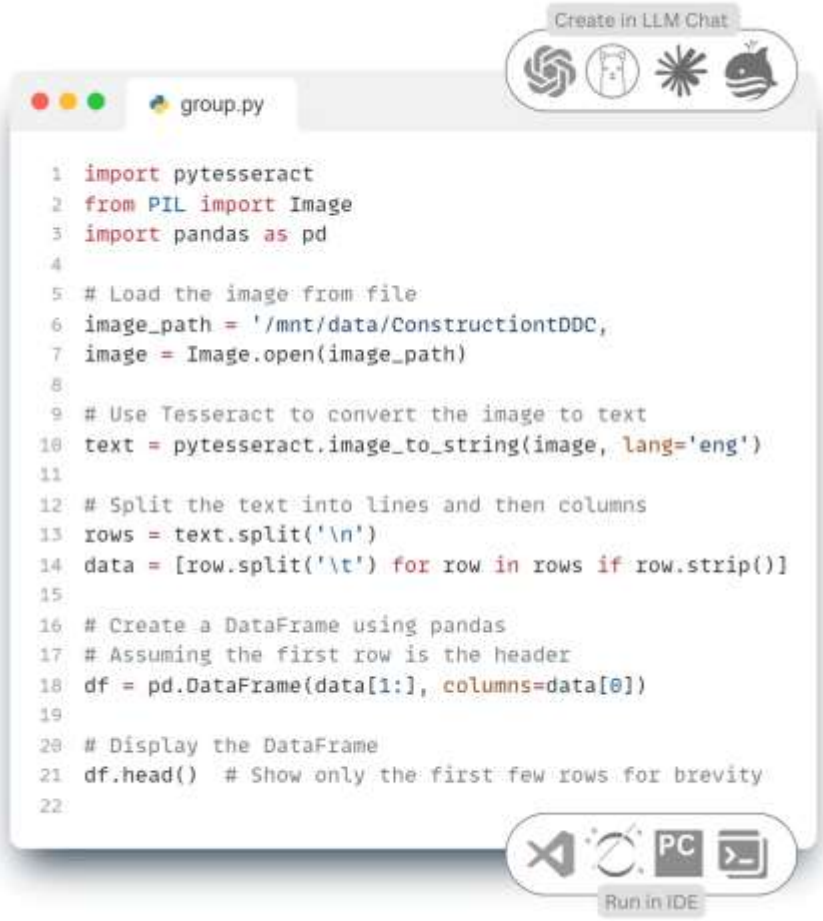


Fig. 4.1-6 Konverter tekst fra en tabel med billeder eller fotos til en struktureret tabelrepræsentation.

I dette eksempel bruger koden (figur 4.1-6) fra LLM pytesseract-biblioteket (Tesseract for Python) til at konvertere et billede til tekst ved hjælp af OCR (optisk tegngenkendelse) og Pandas-biblioteket til at konvertere denne tekst til en struktureret form, dvs. en DataFrame.

> Konverteringsprocessen involverer normalt forbehandling for at forbedre billedkvaliteten, hvorefter der anvendes forskellige algoritmer til mønsterregistrering, udtrækning af funktioner eller genkendelse af objekter. Resultatet er, at ustruktureret visuel information konverteres til strukturerede data.

Selv om PDF og billeder er vigtige kilder til ustruktureret information, er den virkelige mester, hvad angår mængde, tekst genereret i e-mails, chats, møder og messengers. Disse data er ikke bare mange - de er også spredte, uformelle og ekstremt dårligt strukturerede.

Konvertering af tekstdata til en struktureret form

Ud over PDF-dokumenter med tabeller (fig. 4.1-2) og indscannede versioner af tabelformularer (fig. 4.1-5) er en betydelig del af oplysningerne i projektdokumentationen præsenteret i tekstform. Det kan både være sammenhængende sætninger i tekstdokumenter og fragmentariske optegnelser spredt over tegninger og skemaer. Under moderne databehandlingsforhold er en af de mest almindelige opgaver at konvertere en sådan tekst til et struktureret format, der er egnet til analyse, visualisering og beslutningstagning.

Centralt i denne proces er taksonomi, et klassifikationssystem, der gør det muligt at organisere information i kategorier og underkategorier baseret på fælles træk.

> **En taksonomi** er en hierarkisk klassifikationsstruktur, der bruges til at gruppere og organisere objekter. I forbindelse med tekstbehandling tjener den som grundlag for systematisk at kategorisere emner i semantiske kategorier og dermed forenkle analysen og forbedre kvaliteten af databehandlingen.

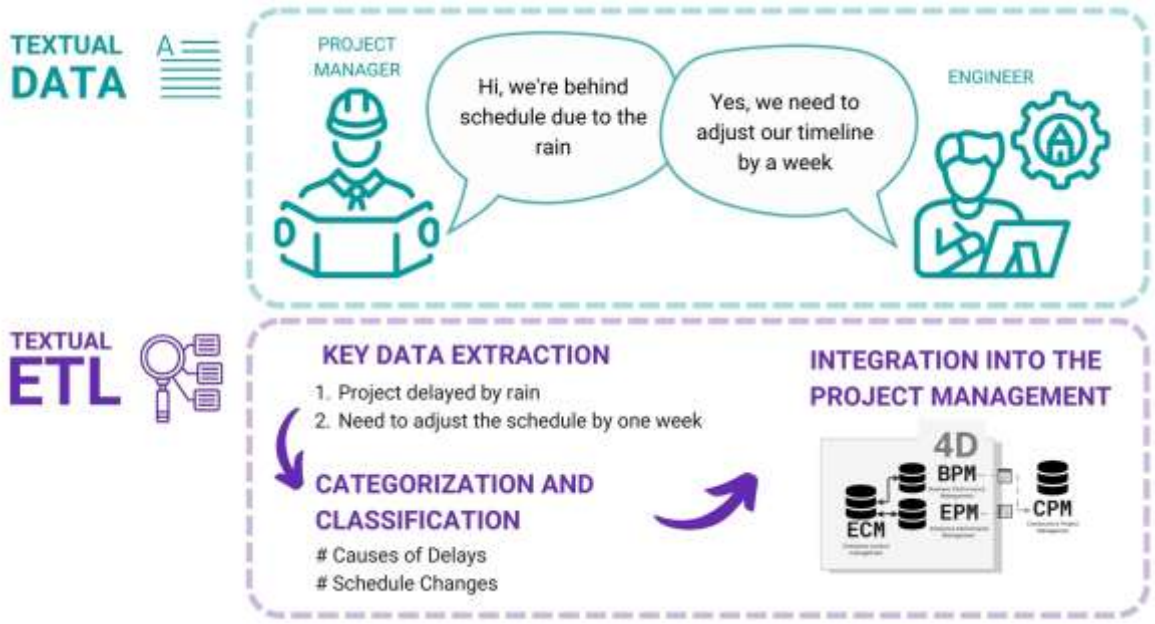
Oprettelse af en taksonomi ledsages af trinene udtrækning af enheder, kategorisering og kontekstualisering. For at modellere processen med at udtrække information fra tekstdata skal vi følge følgende trin, som svarer til dem, vi allerede har anvendt til at strukturere data fra PDF-dokumenter:

- Udtræk): Du skal analysere tekstdata for at udtrække oplysninger om forsinkelser og ændringer i projektplanen.
- **Kategorisering og klassificering** (Transofrm): Kategoriser den modtagne information, f.eks. årsager til forsinkelser og ændringer i tidsplanen.
- **Integration** (Load): Til sidst forbereder vi strukturerede data til integration i eksterne datastyringssystemer.

Overvej en situation: Vi har en dialog mellem en projektleder og en ingeniør, der diskuterer problemer med tidsplanen. Vores mål er at uddrage de vigtigste elementer (årsager til forsinkelser, justeringer af tidsplanen) og præsentere dem på en struktureret måde (figur 4.1-7).

Lad os udføre udtrækket baseret på de forventede nøgleord, oprette en DataFrame for at simulere

dataudtræk og efter transformation en ny DataFrame-tabel, der vil indeholde kolonner for dato, begivenhed (f.eks. årsag til forsinkelse) og handling (f.eks. ændring af tidsplan).



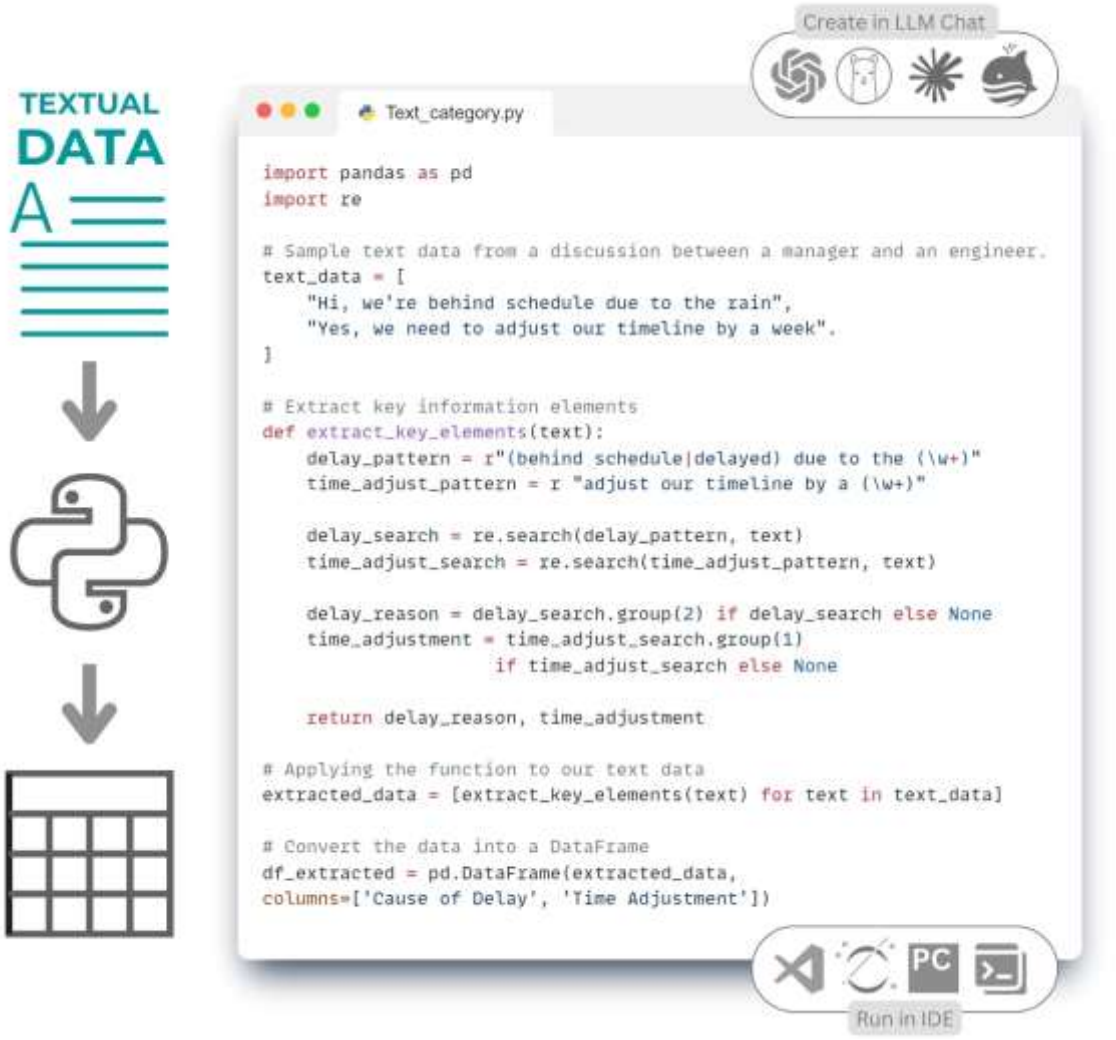
Figur 4.1-7 Fremhæver nøgleinformation fra teksten om behovet for at justere tidslinjer og integrere ændringer i projektledelsessystemet.

Her er koden til at løse problemet ved hjælp af en tekstforespørgsel i en af sprogmodellerne som i de tidligere eksempler.

- Send en tekstanmodning til et hvilket som helst LLM chatrum:

> Jeg har en samtale mellem en leder: "Hej, vi er bagud med tidsplanen på grund af regn" og en ingeniør: "Ja, vi er nødt til at justere deadline med en uge". Jeg har brug for et script, der kan analysere fremtidige lignende tekstdialoger, uddrage årsagerne til forsinkelser og nødvendige justeringer af deadlines og derefter generere en DataFrame ud fra disse data. DataFrame skal derefter gemmes i en CSV -fil ⏎

Svaret fra LLM vil typisk indeholde Python -kode, der bruger regulære udtryk (re - Regex) og Pandas (pd)-biblioteket :

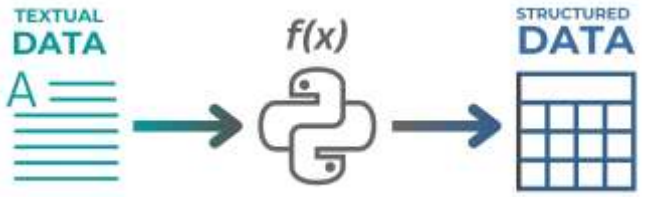


Figur 4.1-8 Fremhævelse af nøgleoplysninger fra teksten om behovet for at justere tidsfrister i en tabel.

I dette eksempel (figur 4.1-7) analyseres tekstdata, der indeholder korrespondance mellem en projektleder og en ingeniør, for at identificere og udtrække specifikke oplysninger, der kan påvirke styringen af fremtidige projekter med lignende dialoger. Ved hjælp regulære udtryk (vi vil diskutere mere om regulære udtryk i kapitlet "Strukturerede krav og regulære udtryk RegEx ") identificeres årsagerne til projektforsinkelser og nødvendige justeringer af tidsplanen ved hjælp af mønstre. Funktionen i dette eksempel udtrækker enten årsagen til forsinkelsen eller tidsjusteringen fra strengene baseret på mønstrene: vælger ordet efter "because of" som årsagen til forsinkelsen eller ordet efter "by" som tidsjusteringen.

Hvis en række nævner en forsinkelse på grund af vejret, identificeres "regn" som årsagen; hvis en række nævner en justering af tidsplanen for en bestemt periode, udtrækkes denne periode som en tidsjustering

(figur 4.1-9). Hvis der ikke er nogen af disse ord i en række, får den tilsvarende attribut-kolonne værdien "None".



Text string	Cause of Delay	Time Adjustment
0	rain	None
1	None	week

Fig. 4.1-9 Den oversigtstabel, der fås som DataFrame, efter udførelse af koden, indeholder oplysninger om forekomsten af forsinkelser og nødvendige tidsjusteringer.

Strukturering og parametrisering af betingelser fra teksten (dialog, brev, dokument) gør det muligt hurtigt at eliminere forsinkelser i byggeriet: For eksempel kan mangel på arbejdere påvirke arbejdstempoet i dårligt vejr, så virksomheder, der kender forsinkelsesparametrene fra dialogerne (fig. 4.1-9) mellem formanden på byggepladsen og projektlederen - på forhånd kan styrke mandskabet i tilfælde af en ugunstig prognose.

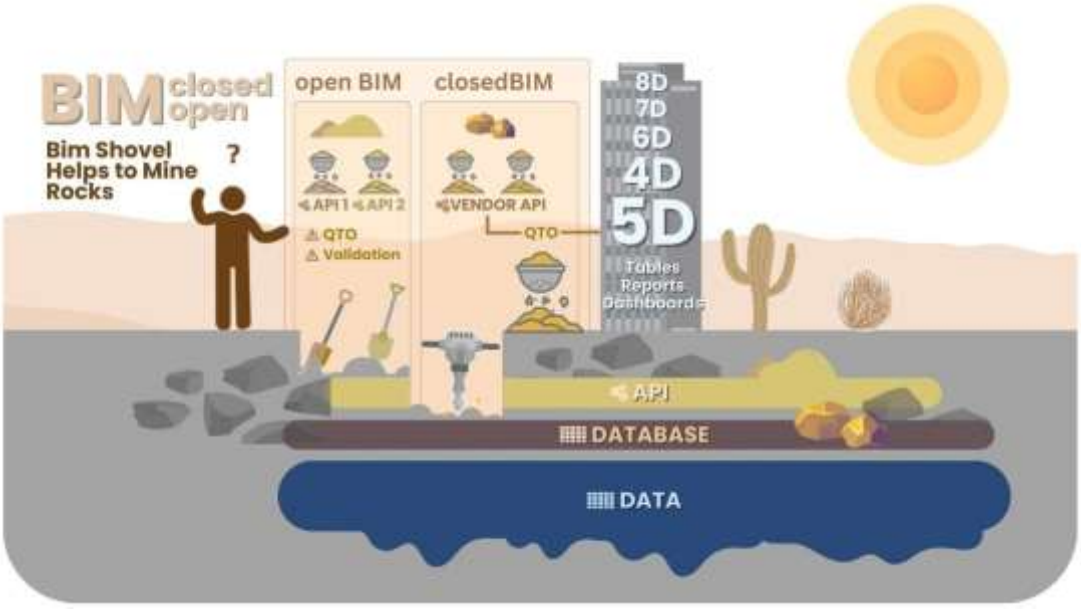
Konvertering af dokumenter og billeder til et struktureret format kan ske ved hjælp af relativt enkle, åbne og gratis kategoriseringsbaserede værktøjer.

Kategorisering af elementer er også en vigtig del af arbejdet med projektdata, især i forbindelse med brug af CAD-software (BIM).

Konvertering af CAD-data (BIM) til en struktureret form

Strukturering og kategorisering af CAD-data (BIM) er en mere kompleks opgave, fordi de data, der er lagret i CAD-databaser (BIM), næsten altid er i lukkede eller komplekse parametriske formater, der ofte kombinerer geometriske dataelementer (semistrukturerede) og metainformationselementer (semistrukturerede eller strukturerede data) på samme tid.

Native dataformater i CAD (BIM) systemer er normalt beskyttede og utilgængelige for direkte brug, medmindre der er tale om specialiseret software eller API - udviklerens egne grænseflader (Fig. 4.1-10). Denne isolering af data danner lukkede lagringssiloer, der begrænser den frie udveksling af information og hæmmer skabelsen af end-to-end digitale processer i virksomheden.



Figur 4.1-10 CAD-professionelle (BIM) kan få adgang til oprindelige data via API -forbindelser eller leverandørværktøjer.

I særlige CAD (BIM) formater samles oplysninger om projektelementernes egenskaber og attributter i et hierarkisk klassifikationssystem, hvor enheder med tilsvarende egenskaber er placeret, som frugten på et frugttræ, i de nyeste knuder på dataklassifikationsgrene (Fig. 4.1-11).

Dataudtræk fra sådanne hierarkier er muligt på to måder: enten manuelt ved at klikke på hver node, som om man bearbejder et træ og skærer udvalgte grene af kategorier og typer ned med en økse. Alternativt giver brugen af programmeringsgrænseflader (API'er) en mere effektiv, automatiseret tilgang til at hente og gruppere data og i sidste ende omdanne dem til en struktureret tabel til brug i andre systemer.

Forskellige værktøjer som Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC eller open source-løsninger, som IfcOpSh eller IFCjs til IFC-format, kan bruges til at udtrække strukturerede datatabeller fra CAD (BIM) -projekter.

Moderne dataeksport- og konverteringsværktøjer gør det muligt at forenkle databehandling og -forberedelse ved at opdele indholdet af CAD-modeller i to hovedkomponenter: geometriinformation og attributdata (fig. 4.1-13) - metainformation, der beskriver designelementernes egenskaber (fig. 3.1-16). Disse to datalag forbliver forbundet via unikke identifikatorer, som gør det muligt præcist at mappe hvert element med geometribeskrivelse (via parametre eller polygoner) til dets attributter: navn, materiale, færdiggørelsestrin, omkostninger osv. Denne tilgang sikrer modellens integritet og giver mulighed for fleksibel brug af data både til visualisering (geometriske modeldata) og til analyse- eller ledelsesopgaver (struktureret eller løst struktureret), hvor man kan arbejde med de to typer data separat eller parallelt.

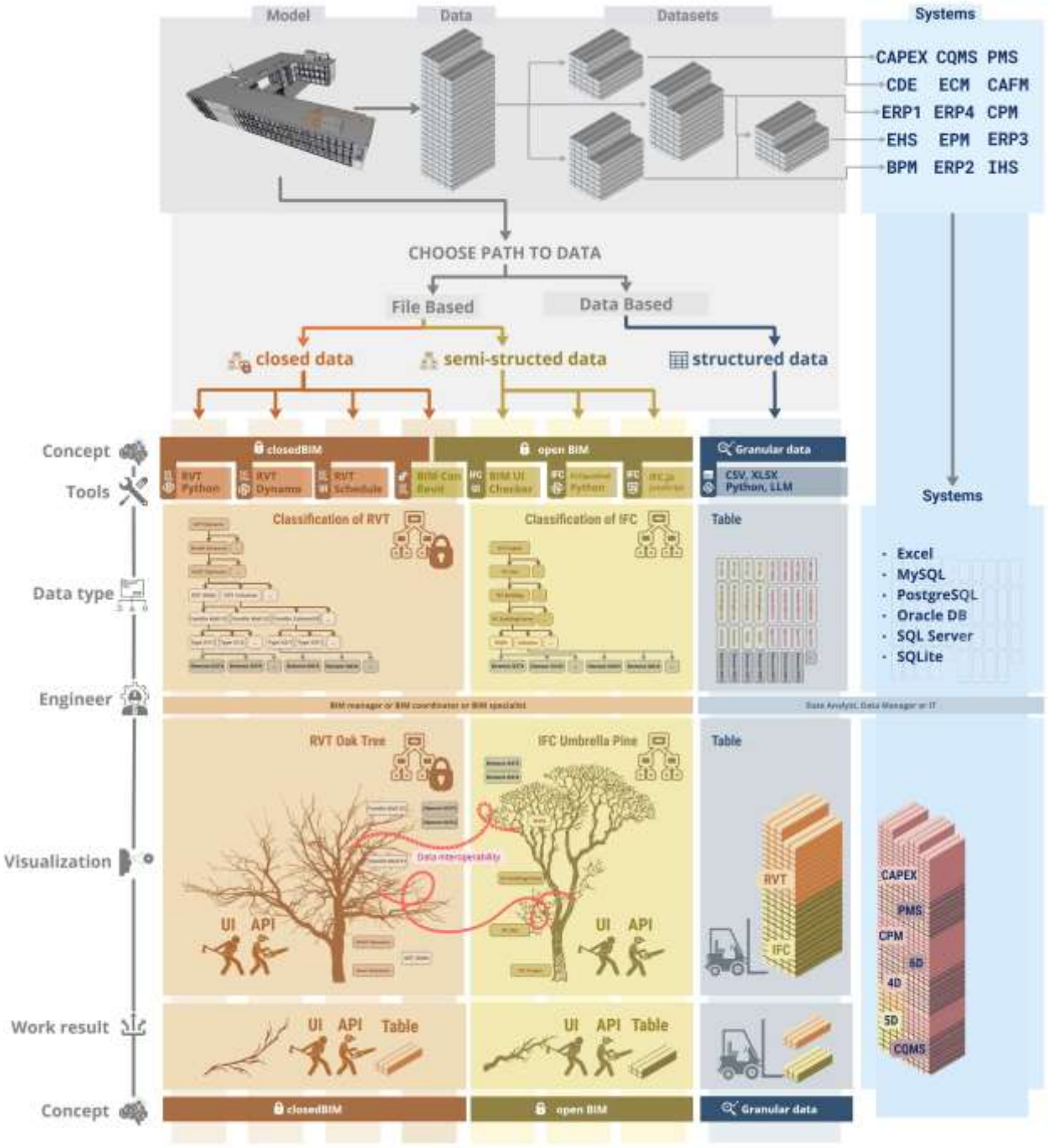


Fig. 4.1-11 Visningen af information fra CAD-databaser (BIM) præsenteres for brugeren i form af klassifikationstræer.

Med udviklingen af reverse engineering-teknologier og fremkomsten af SDK'er (Software Development Kit) til CAD-datakonvertering - er tilgængeligheden og konverteringen af data fra lukkede CAD-softwareformater (BIM) blevet meget lettere. Det er nu muligt lovligt og sikkert at konvertere data fra lukkede formater til universelle formater, der er egnede til analyse og brug i andre systemer. Historien om de første reverse

engineering-værktøjer ("Open DWG") og kampen om dominans over CAD-leverandørernes formater blev diskuteret i kapitlet "Strukturerede data: grundlaget for digital transformation".

> Reverse engineering-værktøjer gør det muligt at hente data fra lukkede, proprietære formater og nedbryde oplysninger fra det blandede CAD-format (BIM) til de datatyper og -formater, som brugeren har brug for, hvilket gør det lettere at behandle og analysere dem.

Ved hjælp af reverse engineering og direkte adgang til information fra CAD-databaser gør information tilgængelig, hvilket giver mulighed for åbne data og åbne værktøjer samt dataanalyse ved hjælp af standardværktøjer, opbygning af rapporter, visualiseringer og integration med andre digitale systemer (Fig. 4.1-12).

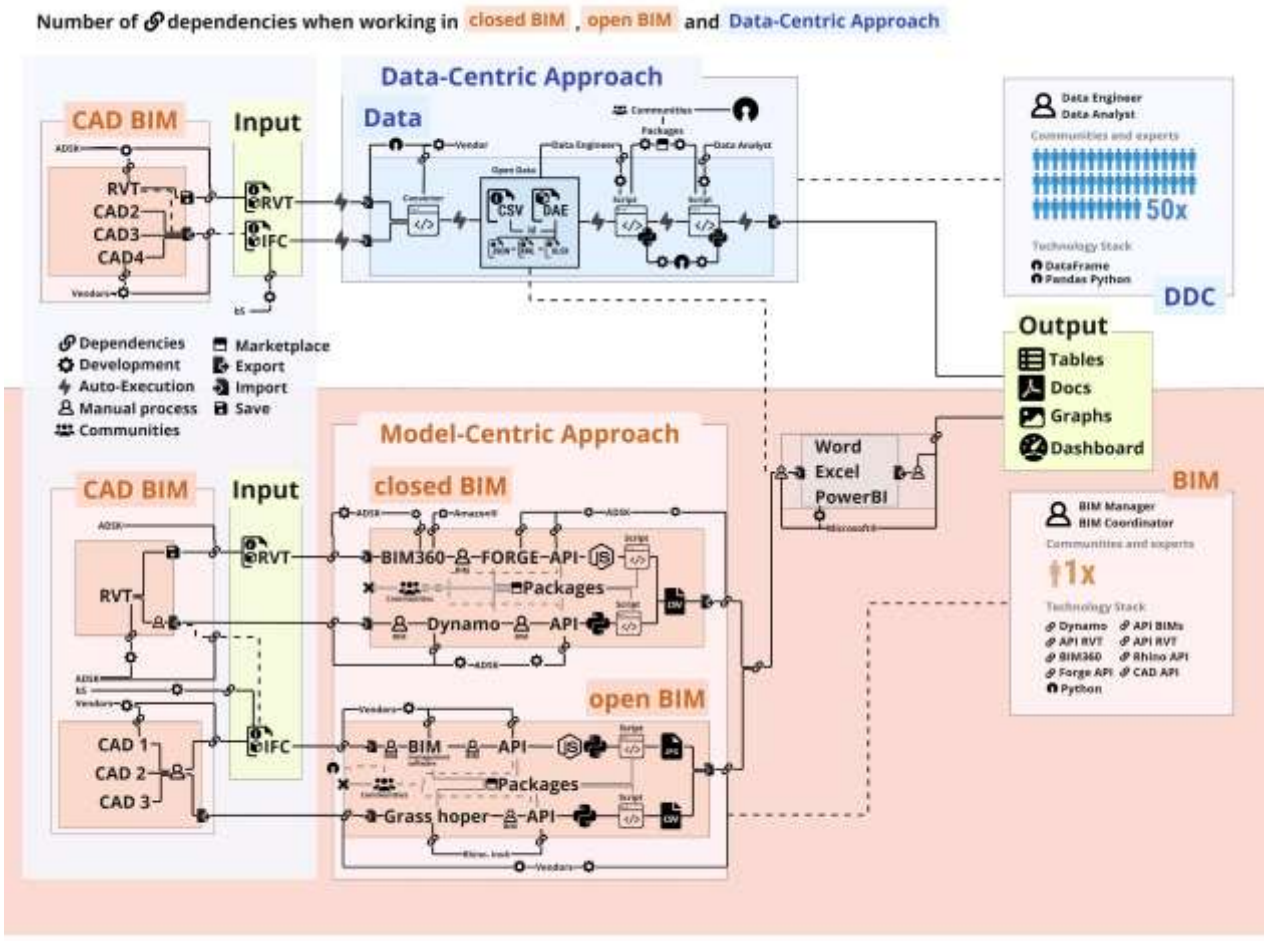


Fig. 4.1-12 Direkte adgang til CAD-data minimerer afhængigheden af softwareplatforme og bevæger sig i retning af en datacentreret tilgang.

Siden 1996 for DWG-formatet, siden 2008 for DGN-formatet og siden 2018 for RVT har det været muligt at konvertere oprindeligt lukkede CAD-dataformater til alle andre formater, herunder strukturerede formater, nemt og effektivt ved hjælp af reverse engineering-værktøjer (Fig. 4.1-13). I dag bruger næsten alle større CAD (BIM) og store ingeniørfirmaer i verden SDK'er - reverse engineering-værktøjer til at udtrække data fra lukkede CAD (BIM)-leverandørformater [92].



Fig. 4.1-13 Brugen af reverse engineering-værktøjer gør det muligt at konvertere CAD (BIM) programdatabaser til en hvilken som helst praktisk datamodel.

Konvertering af data fra lukkede, proprietære formater til åbne formater og adskillelse af blandede CAD (BIM) formater i geometriske og metainformationsattributdata forenkler processen med at arbejde med dem og gør dem tilgængelige for analyse, manipulation og integration med andre systemer (Fig. 4.1-14).

> I dagens arbejde med CAD-data (BIM) er vi nået til et punkt, hvor vi ikke behøver at søge tilladelse fra CAD-(BIM)-leverandører for at få adgang til oplysninger fra CAD-formater.

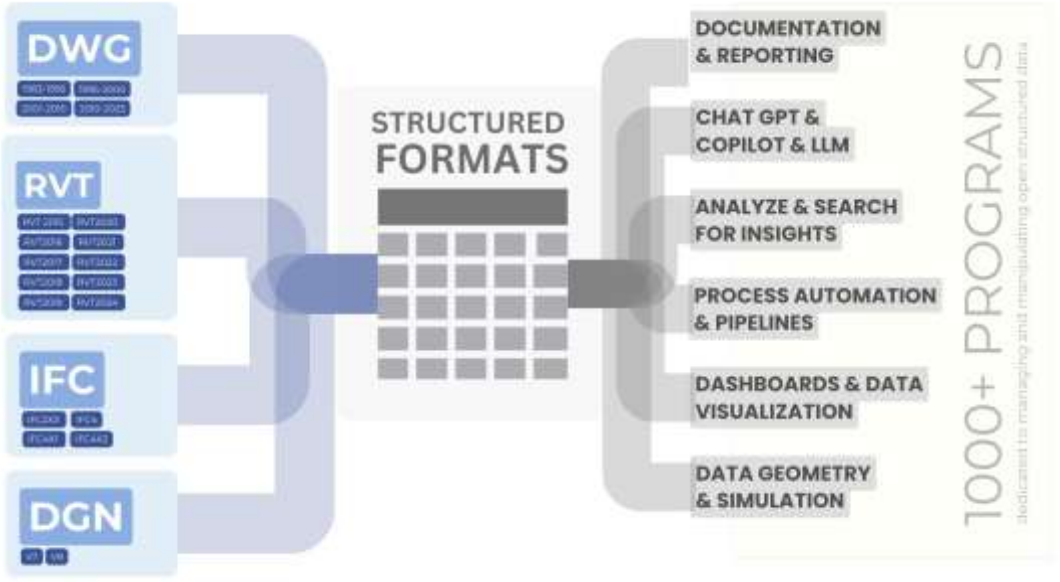


Fig. 4.1-14 Moderne SDK-værktøjer tillader lovlig konvertering af data fra proprietære CAD-databaseformater (BIM).

De aktuelle tendenser inden for behandling af CAD-designdata formes fortsat af de vigtigste markedsaktører - CAD - leverandører, som arbejder på at styrke deres position i dataverdenen og skabe nye formater og koncepter.

Leverandører af CAD -løsninger bevæger sig mod strukturerede data

Fra 2024 og frem gennemgår design- og byggebranchen et betydeligt teknologisk skift i brugen og behandlingen af data. I stedet for fri adgang til designdata fokuserer CAD -systemleverandører på at fremme de næste nye koncepter. Tilgange som BIM (oprettet i 2002) og åben BIM (oprettet i 2012) viger gradvist for moderne teknologiske løsninger, som CAD-leverandører begynder at promovere [93]:

- Overgang til brug af "granulære" data, der giver mulighed for effektiv informationsstyring og en overgang til dataanalyse
- Fremkomsten af USD format og implementering af Entity-component-system approach (ECS) for fleksibel dataorganisation.
- Aktiv brug af kunstig intelligens i databehandling, procesautomatisering og dataanalyse
- Udvikling af interoperabilitet - forbedret interaktion mellem forskellige programmer, systemer og databaser

Hvert af disse aspekter vil blive diskuteret mere detaljeret i sjette del af bogen "CAD and BIM: marketing, reality and the future of design data in construction". I dette kapitel vil vi kun kort skitsere den generelle forandringsvektor: De store CAD-leverandører forsøger nu at gentænke den måde, designinformation struktureres på. Et af de vigtigste skift er opgivelsen af den klassiske filbaserede lagringsmodel til fordel for en granulær, analyseorienteret dataarkitektur, der giver kontinuerlig adgang til individuelle modelkomponenter [93].

> Essensen af det, der sker, er, at industrien udfaser besværlige, specialiserede og parametriske formater, der kræver geometriske kerner, til fordel for mere universelle, maskinlæsbare og fleksible løsninger.

En sådan drivkraft for forandring er USD-formatet (Universal Scene Description), som oprindeligt blev udviklet i computergrafikindustrien, men som allerede er anerkendt i tekniske anvendelser takket være udviklingen af NVIDIA Omniverse (og Isaac Sim)-platformen til simuleringer og visualiseringer [93]. I modsætning til den parametriske IFC har USD en enklere struktur og gør det muligt at beskrive geometri og objektegenskaber i JSON-format (figur 4.1-15), hvilket letter informationsbehandlingen og fremskynder integrationen i digitale processer. Det nye format gør det muligt at lagre geometri (ud over BREP -NURBS - flere detaljer i del 6 af bogen) i form af MESH-polygoner og objektegenskaber i JSON, hvilket gør det mere praktisk for automatiserede processer og arbejde i cloud-økosystemer [94].

Nogle CAD- og ERP-leverandører bruger allerede lignende formater (f.eks. NWD, SVF, CP2, CPIXML), men de fleste af dem er fortsat lukkede og utilgængelige for ekstern brug, hvilket begrænser mulighederne for dataintegration og genbrug. I denne sammenhæng kan USD spille samme rolle, som DXF gjorde i sin tid - et åbent alternativ til proprietære formater som DWG.

General Information			Comparison / Notes
Year of format creation	1991	2016	IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar	IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF	IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK	ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK	ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD	Different alliances for each format
Year of Alliance Formation	1994	2023	The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co	ADSK and Co actively promotes both formats in bS (IAI) since the introduction

Purpose and Usage			Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification	IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing	IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction	USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera	USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

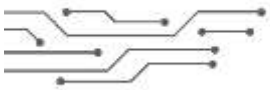
Fig. 4.1-15 USD format som et forsøg fra CAD-leverandører på at imødekomme kravet om interoperabilitet og uafhængighed af designdata fra geometriske kerner.

De store udvikleres overgang til åbne og forenklede USD, GLTF, OBJ, XML (lukket NWD, CP2, SVF, SVF2, CPIXML) og lignende formater (Fig. 3.1-17) afspejler den globale tendens og industriens efterspørgsel efter dataforenkling og øget tilgængelighed. I de kommende år kan vi forvente en gradvis bevægelse væk fra komplekse parametriske standarder og formater med afhængighed af geometriske kerner til fordel for lettere og mere strukturerede løsninger. Denne overgang vil fremskynde digitaliseringen af byggebranchen, lette procesautomatiseringen og forenkle dataudvekslingen.

På trods af CAD -leverandørernes strategiske planer om at fremme nye åbne formater, kan fagfolk i byggebranchen også få fuld adgang til data fra lukkede CAD-systemer uden behov for CAD-værktøjer (BIM) ved at bruge reverse engineering-værktøjer.

> Alle disse tendenser fører uundgåeligt til et skift fra store, monolitiske 3D-modeller til universelle, strukturerede data og til brug af formater, der længe har været afprøvet i andre brancher. Når projektteams begynder at se CAD-modeller ikke bare som visuelle objekter eller et sæt filer, men som databaser, der indeholder viden og information, ændrer tilgangen til design og ledelse sig dramatisk.

Når teams har lært at udtrække strukturerede data fra dokumenter, tekster, tegninger og CAD-modeller og har adgang til databaser, er det næste vigtige skridt datamodellering og kvalitetssikring. Det er dette trin, der i høj grad bestemmer, hvor hurtigt man kan behandle og omdanne de oplysninger, der i sidste ende skal bruges til at træffe beslutninger i specifikke opgaver.



KAPITEL 4.2.
KLASSIFIKATION OG INTEGRATION: ET FÆLLES SPROG FOR BYGGEDATA

Hastigheden af beslutningstagningen afhænger af kvaliteten af data

Nutidens designdataarkitektur undergår grundlæggende forandringer. Branchen bevæger sig væk fra omfangsrige, isolerede modeller og lukkede formater og hen imod mere fleksible, maskinlæsbare strukturer med fokus på analyse, integration og procesautomatisering. Overgangen til nye formater alene garanterer dog ikke effektivitet - selve datakvaliteten er uundgåeligt i centrum.

> På siderne i denne bog taler vi meget om formater, systemer og processer. Men alle disse bestræbelser er meningsløse uden et nøgleelement: data, som man kan stole på. Datakvalitet er hjørnestenen i digitaliseringen, som vi vil vende tilbage til i de følgende dele.

Moderne byggevirksomheder - især de store - bruger snesevis og nogle gange tusindvis af forskellige systemer og databaser (fig. 4.2-1). Disse systemer skal ikke kun fyldes med ny information regelmæssigt, men også interagere effektivt med hinanden. Alle nye data, der genereres som følge af behandling af indgående information, integreres i disse miljøer og tjener til at løse specifikke forretningsopgaver.

Og mens beslutninger om specifikke forretningsopgaver tidligere blev truffet af topledere - såkaldte HiPPO-'er (Fig. 2.1-9) - på grundlag af erfaring og intuition, er denne tilgang i dag, med den kraftige stigning i informationsmængden, ved at blive kontroversiel. Automatiseret analyse, som arbejder med data i realtid, er ved at erstatte den.

> "Traditionelle manuelle" diskussioner om forretningsprocesser på ledelsesniveau vil skifte til operationelle analyser, som kræver hurtige svar på forretningsforespørgsler.

Den tid, hvor revisorer, formænd og beregnere manuelt genererede rapporter og oversigtsskemaer og projektdatashows over dage og uger, hører fortiden til. I dag er hastighed og rettidighed i beslutningstagningen ved at blive en nøglefaktor i konkurrencefordelen.



Fig. 4.2-1 I byggebranchen tager det dage at beregne og træffe beslutninger i modsætning til andre brancher, hvor det sker på timer eller minutter.

Den største forskel mellem byggebranchen og mere digitalt avancerede brancher (figur 4.2-1) er det lave niveau af datakvalitet og standardisering. Forældede tilgange til generering, transmission og behandling af information bremser processer og skaber kaos. Manglen på ensartede standarder for datakvalitet hindrer implementeringen af end-to-end-automatisering

> En af de største udfordringer er fortsat den dårlige kvalitet af inputdata samt manglen på formaliserede processer til forberedelse og validering af dem. Uden pålidelige og konsistente data kan der ikke være nogen effektiv integration mellem systemerne. Det fører til forsinkelser, fejl og øgede omkostninger i alle faser af projektets livscyklus.

I de følgende afsnit af bogen ser vi nærmere på, hvordan du kan forbedre datakvaliteten, standardisere processer og forkorte rejsen fra information til validerede og konsistente data af høj kvalitet.

Datastandardisering og -integration

Effektiv datahåndtering kræver en klar standardiseringsstrategi. Kun med klare krav til datastruktur og -kvalitet kan datavalidering automatiseres, manuelle operationer reduceres, og informeret beslutningstagning fremskyndes i alle faser af et projekt.

I den daglige praksis skal en byggevirksomhed behandle hundredvis af filer hver dag: e-mails, PDF -dokumenter, CAD-designfiler, data fra IOT-sensorer, som skal integreres i virksomhedens forretningsprocesser.

Skoven i en virksomheds økosystem af databaser og værktøjer (figur 4.2-2) skal lære at udlede næringsstoffer fra indkommende data i flere formater for at opnå de resultater, virksomheden ønsker.

For at håndtere datastrømmen effektivt behøver du ikke nødvendigvis at ansætte en hær af ledere, du skal først udvikle strenge krav og standarder for dataene og bruge passende værktøjer til automatisk at validere, harmonisere og behandle dem.

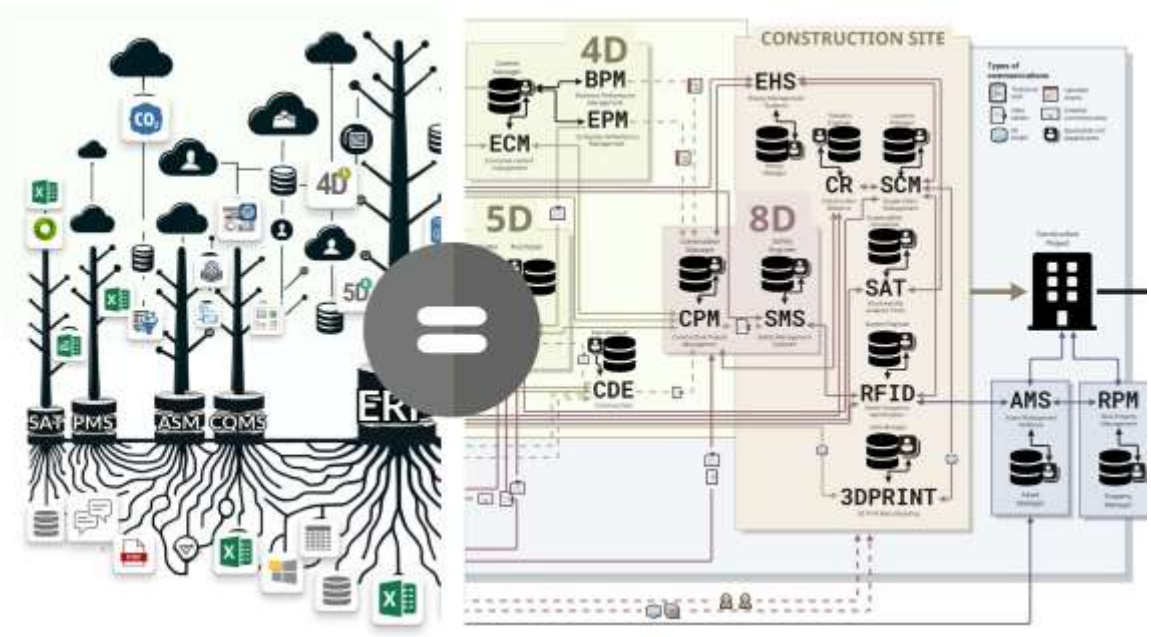


Fig. 4.2-2 For at sikre et sundt virksomhedsøkosystem er det nødvendigt med kvalitet og rettidige ressourcer til systemerne.

For at kunne automatisere datavaliderings- og harmoniseringsprocessen (med henblik på efterfølgende automatisk integration) bør du starte med at beskrive de nødvendige minimumskrav til data for hvert enkelt system. Disse krav definerer:

- Hvad har du helt præcist brug for at få?
- I hvilken form (struktur, format)?
- Hvilke egenskaber er obligatoriske?
- Hvilke tolerancer i nøjagtighed og fuldstændighed er acceptable?

Datakrav beskriver kriterier for kvaliteten, strukturen og fuldstændigheden af den information, der modtages og behandles. For eksempel er det vigtigt for tekster i PDF -dokumenter, at de er nøjagtigt formateret i henhold til industristandarder (fig. 7.2-14 - fig. 7.2-16). Objekter i CAD -modeller skal have korrekte attributter (dimensioner, koder, links til klassifikatorer) (Fig. 7.3-9, Fig. 7.3-10). Og for kontraktscanninger er det vigtigt med klare datoer og mulighed for automatisk at udtrække beløb og nøgleord (Fig. 4.1-7 - Fig. 4.1-10).

At formulere datakrav og automatisk kontrollere, at de overholdes, er et af de mest tidskrævende, men kritiske trin. Det er det mest tidskrævende trin i forretningsprocesser.

Som nævnt i del 3 af denne bog bruges mellem 50 og 90% af business intelligence (BI)-medarbejdernes tid på dataforberedelse frem for analyse (figur 3.2-5). Denne proces omfatter dataindsamling, verificering, validering, harmonisering og strukturering.

Ifølge en undersøgelse fra 2016 [95] sagde dataforskere inden for en lang række bredspektrede områder, at de bruger det meste af deres arbejdstid (ca. 80%) på at gøre det, de mindst kan lide at gøre (figur 4.2-3): at indsamle eksisterende datasæt og organisere (forene, strukturere) dem. Dermed er der mindre end 20 procent af deres tid tilbage til kreative opgaver, som f.eks. at finde mønstre og regelmæssigheder, der kan føre til nye indsigter og opdagelser.

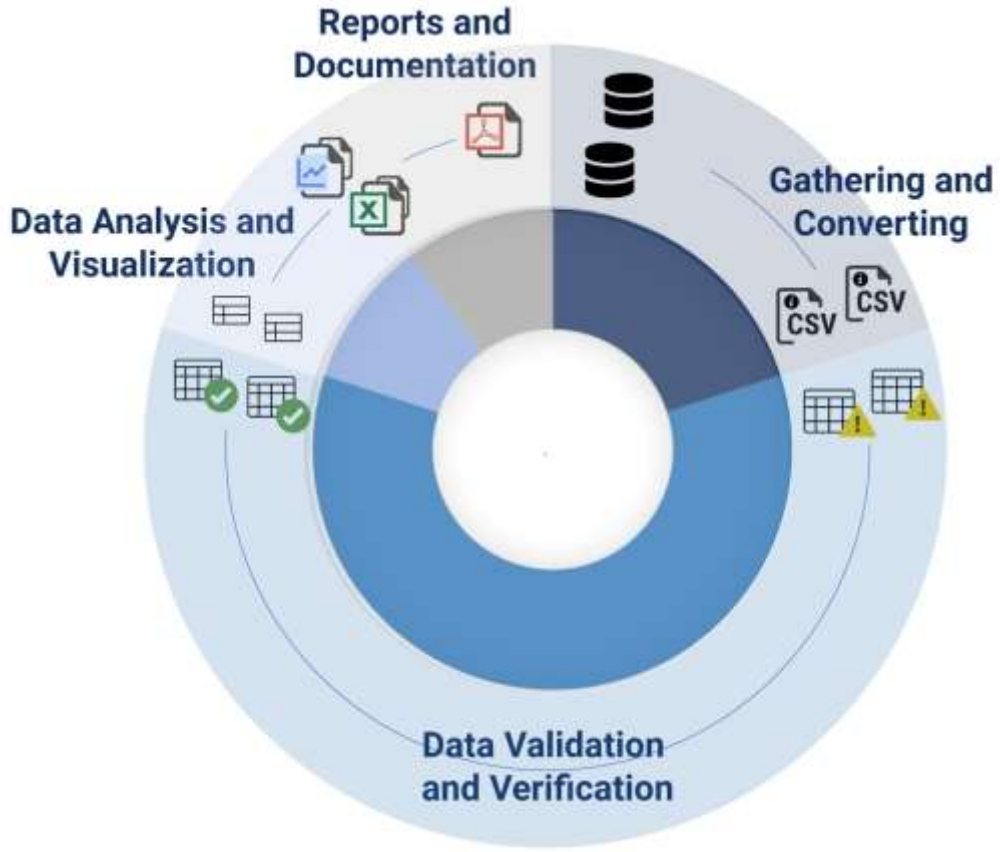


Fig. 4.2-3 Verificering af og sikring af datakvalitet er det mest kostbare, tidskrævende og komplekse trin i forberedelsen af data til integration i andre systemer.

Vellykket datahåndtering i en byggevirksomhed kræver en omfattende tilgang, der omfatter parametrisering af opgaver, formulering af krav til datakvalitet og brug af egnede værktøjer til automatiseret validering.

Digital interoperabilitet starter med krav

I takt med at antallet af digitale systemer i virksomhederne vokser, vokser også behovet for datakonsistens mellem dem. Ledere med ansvar for forskellige IT-systemer er ofte ude af stand til at holde trit med den stigende informationsmængde og de mange forskellige formater. Under sådanne omstændigheder er de tvunget til at bede specialister om at skabe data i en form, der er egnet til brug i andre applikationer og

platforme.

Det kræver til gengæld, at ingeniører og datagenereringsmedarbejdere tilpasser sig en lang række krav, ofte uden gennemsigtighed og en klar forståelse af, hvor og hvordan dataene vil blive brugt i fremtiden. Manglen på standardiserede tilgange til håndtering af information fører til ineffektivitet og øgede omkostninger i verifikationsfasen, som ofte er manuel på grund af dataenes kompleksitet og ikke-standardiserede karakter.

> Spørgsmålet om datastandardisering er ikke bare et spørgsmål om bekvemmelighed eller automatisering. Det er et direkte økonomisk tab. Ifølge en IBM-rapport fra 2016 er det årlige tab som følge af dårlig datakvalitet i USA på 3,1 billioner dollars [96]. Derudover viser undersøgelser fra MIT og andre analytiske konsulentfirmaer, at omkostningerne ved dårlig datakvalitet kan være så høje som 15-25% af en virksomheds omsætning [97].

Under disse forhold bliver det afgørende at have klart definerede datakrav og beskrivelser af, hvilke parametre, i hvilket format og med hvilken detaljeringsgrad, der skal indgå i de oprettede objekter. Uden formalisering af disse krav er det umuligt at garantere datakvalitet og kompatibilitet mellem systemer og projektfaser (Fig. 4.2-4).

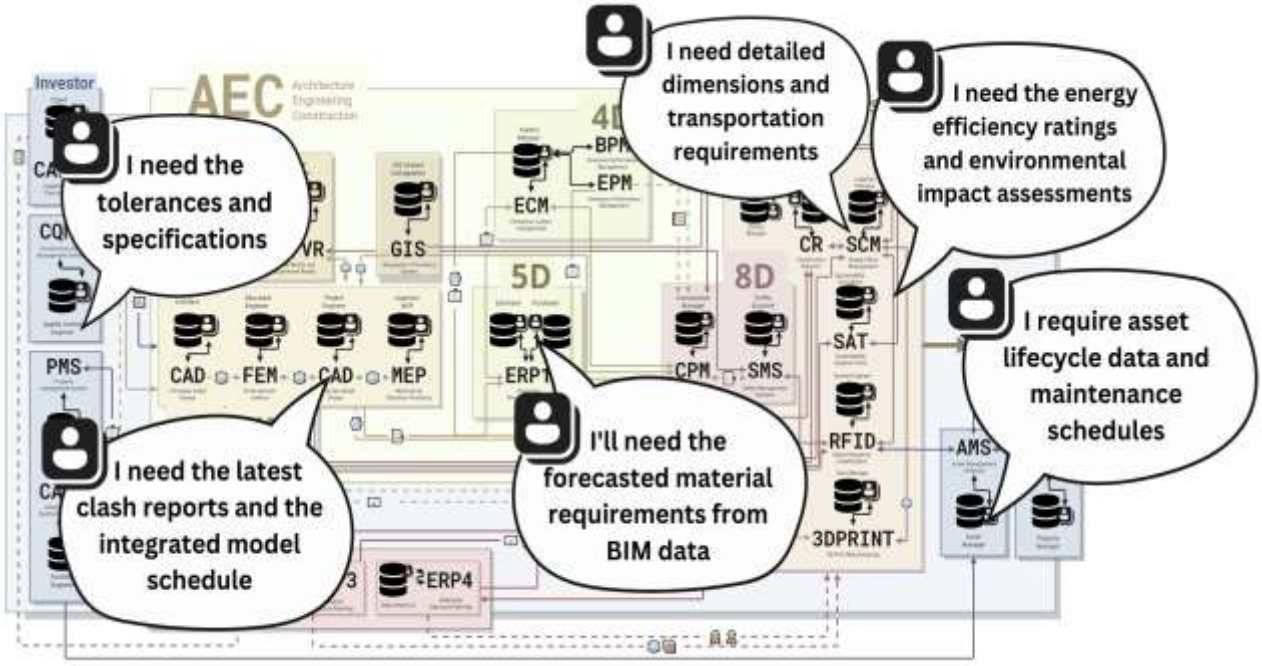
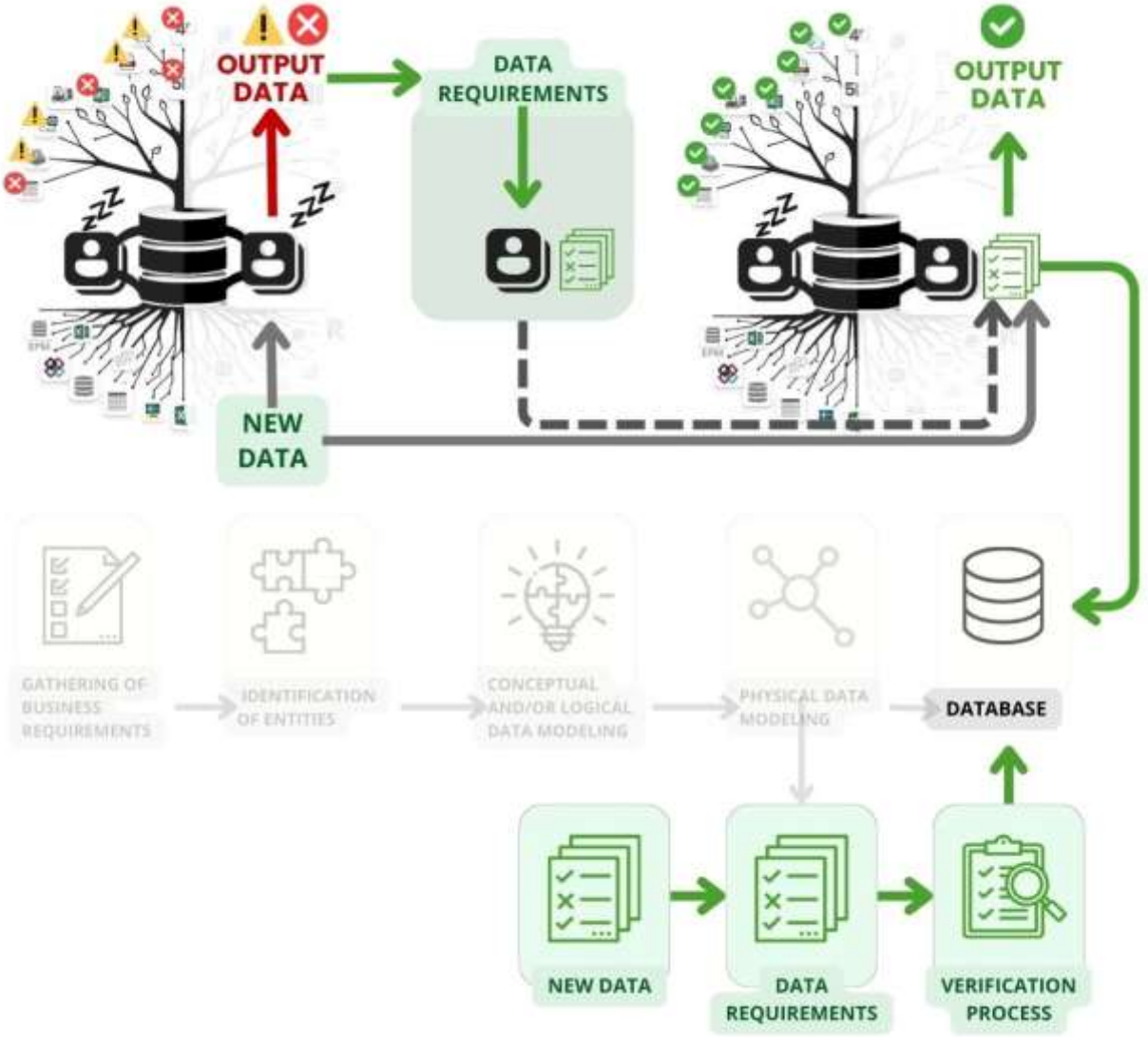


Fig. 4.2-4 En virksomhed er baseret på samspillet mellem forskellige roller, som hver især kræver bestemte parametre og værdier, der er afgørende for at opfylde virksomhedens mål.

For at kunne formulere de korrekte datakrav skal man forstå forretningsprocesserne på dataniveau. Byggeprojekter varierer i type, omfang og antal deltagere, og hvert system - uanset om det er modellering (CAD (BIM)), planlægning (ERP 4D), omkostningsberegning (ERP 5D) eller logistik (SCM) - kræver sine egne unikke parametre for input (input entity-elements).

Afhængigt af disse behov skal virksomhedslederne enten designe nye datastrukturer for at opfylde kravene eller tilpasse eksisterende tabeller og databaser. Kvaliteten af de data, der skabes, afhænger direkte af, hvor præcist og korrekt kravene er formuleret (Fig. 4.2-5).



Figur 4.2-5 Datakvaliteten afhænger af kvaliteten af de krav, der stilles til specifikke data use cases.

Da hvert system har sine egne specifikke datakrav, er det første skridt i formuleringen af generelle krav at kategorisere alle elementer, der er involveret i forretningsprocesser. Det betyder, at det er nødvendigt at opdele objekter i klasser og grupper af klasser, der svarer til specifikke systemer eller anvendelsesopgaver. For hver sådan gruppe udvikles der separate krav til datastruktur, attributter og kvalitet.

I praksis står implementeringen af denne tilgang dog over for en stor udfordring: manglen på et fælles sprog til gruppering af data. Forskellige klassifikationer, dobbelte identifikatorer og inkompatible formater resulterer i, at hver virksomhed, hver software og endda hvert projekt danner sine egne, isolerede datamodeller og -klasser. Resultatet er et digitalt "Babelstårn", hvor overførsel af information mellem systemer kræver flere konverteringer til de rigtige datamodeller og -klasser, som ofte udføres manuelt. Denne barriere kan kun overvindes ved at bevæge sig i retning af universelle klassifikatorer og standardiserede sæt af krav.

Et fælles konstruktionssprog: Klassifikatorernes rolle i den digitale transformation

I forbindelse med digitalisering og automatisering af inspektions- og forarbejdningsprocesser spiller klassifikationssystemerne en særlig rolle - en slags "digitale ordbøger", der sikrer ensartethed i beskrivelsen og parametriseringen af objekter. Klassifikatorer udgør det "fælles sprog", der gør det muligt at gruppere data efter betydning og integrere dem mellem forskellige systemer, ledelsesniveauer og faser i projektets livscyklus.

Den mest håndgribelige effekt af klassifikatorer er økonomien i bygningens livscyklus, hvor det vigtigste aspekt er optimeringen af de langsigtede driftsomkostninger. Undersøgelser viser, at driftsomkostningerne udgør op til 80% af de samlede ejeromkostninger for en bygning, hvilket er tre gange højere end de oprindelige byggeomkostninger (fig. 4.2-6) [98]. Det betyder, at beslutningen om de fremtidige omkostninger i høj grad træffes i designfasen

> Derfor bør krav fra driftsingeniører (CAFM, AMS, PMS, RPM) blive udgangspunktet for generering af datakrav i designfasen (fig. 1.2-4). Disse systemer skal ikke ses som den sidste fase af projektet, men som en integreret del af hele projektets digitale økosystem, fra koncept til afvikling

En moderne klassifikator er ikke bare et system af koder til gruppering. Det er en mekanisme til gensidig forståelse mellem arkitekter, ingeniører, beregnere, logistikere, vedligeholdelses- og IT-systemer. Ligesom en bils autopilot skal genkende vejobjekter entydigt og præcist, skal digitale byggesystemer og deres brugere fortolke det samme projektelement entydigt for forskellige systemer via elementklassen.

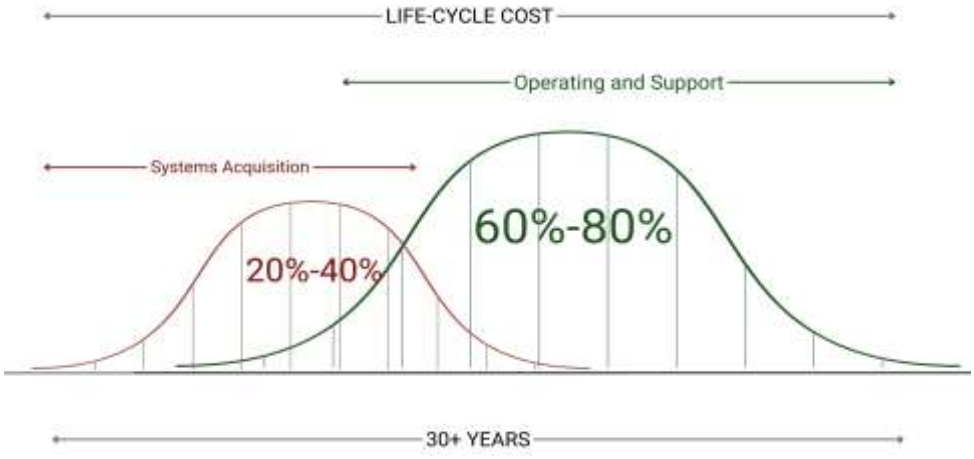


Fig. 4.2-6 Drifts- og vedligeholdelsesomkostningerne overstiger byggeomkostningerne tre gange og udgør 60-80% af alle omkostninger i en bygnings livscyklus (ifølge [99]).

Niveauet for udvikling af klassifikatorer hænger direkte sammen med dybden af en virksomheds digitalisering og dens digitale modenhed. Organisationer med et lavt niveau af digital modenhed står over for fragmenterede data, inkompatible informationssystemer og som følge heraf inkompatible og ineffektive klassifikatorer. I sådanne virksomheder kan det samme element ofte have forskellige

grupperingsidentifikatorer i forskellige systemer, hvilket på kritisk vis hæmmer den endelige integration og gør procesautomatisering umulig.

For eksempel kan det samme vindue i et projekt være mærket forskelligt i CAD model, estimerings- og vedligeholdelsessystem (Fig. 4.2-7) på grund af den flerdimensionelle opfattelse af elementer hos forskellige deltagere i processen. For beregneren i vindueskategorien element er volumen og omkostninger vigtige, for vedligeholdelsestjenesten - tilgængelighed og vedligeholdelsesevne, for arkitekten - æstetiske og funktionelle egenskaber. Som følge heraf kan det samme element kræve forskellige parametre.

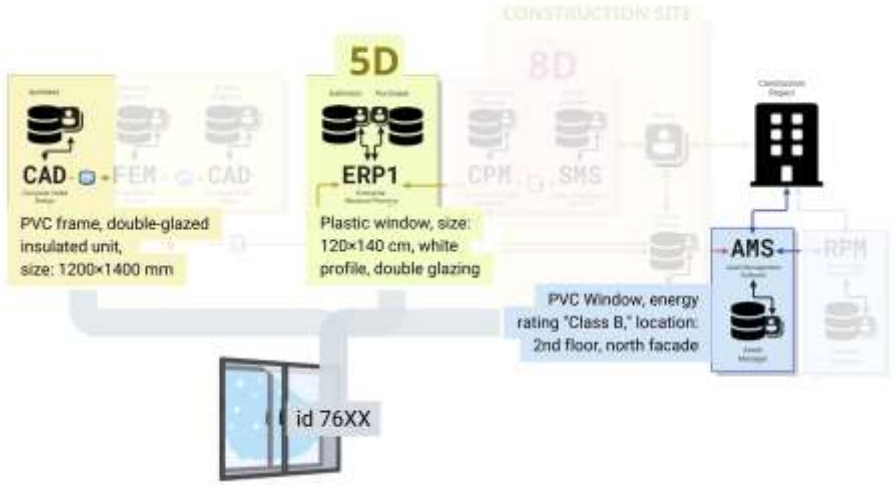


Fig. 4.2-7 Med inkonsekvent klassifikation mellem systemer vil et element miste nogle attributoplysninger ved hvert trin i overgangen til et andet system.

På grund af vanskelighederne ved entydigt at definere klassifikationen af bygningselementer tildeler specialister fra forskellige områder ofte uforenelige klasser til det samme element. Dette resulterer i et tab af et samlet billede af objektet, hvilket kræver efterfølgende manuel indgriben for at harmonisere de forskellige klassifikationssystemer og for at etablere konsistens mellem de typer og klasser, der er defineret af forskellige specialister.

Som følge af denne uoverensstemmelse kan den operationelle dokumentation, som indkøbsafdelingen (ERP) modtager, når en byggevare købes fra en producent, ofte ikke knyttes korrekt til klassifikationen af den pågældende vare på byggepladsen (PMIS, SCM). Som følge heraf er det ikke sandsynligt, at kritiske oplysninger integreres i infrastruktur- og asset management-systemer (CAFM, AMS), hvilket skaber alvorlige problemer under idriftsættelse samt under efterfølgende vedligeholdelse (AMS, RPM) eller udskiftning af elementet.

> I virksomheder med høj digital modenhed spiller klassifikatorer rollen som et nervesystem, der forener alle informationsstrømme. Det samme element får en unik identifikator, så det kan overføres mellem CAD, ERP, AMS og CAFM -systemer og deres klassifikatorer uden forvrængning eller tab.

For at opbygge effektive klassifikatorer skal man forstå, hvordan dataene bruges. Den samme ingeniør kan navngive og klassificere et element forskelligt i forskellige projekter. Kun ved at indsamle brugsstatistikker gennem årene kan man udvikle et stabilt klassifikationssystem. Maskinlæring hjælper med dette: Algoritmer analyserer tusindvis af projekter (fig. 9.1-10) og identificerer sandsynlige klasser og parametre ved

hjælp af maskinlæring (fig. 10.1-6). Automatisk klassificering er især værdifuld i miljøer, hvor manuel klassificering ikke er mulig på grund af datamængden. Automatiske klassifikationssystemer vil være i stand til at skelne mellem grundlæggende kategorier baseret på minimalt udfyldte vareparametre (flere detaljer i bogens niende og tiende del).

Udviklede klassificeringssystemer bliver katalysatorer for yderligere digitalisering og skaber grundlag for..:

- Automatiseret estimering af projektomkostninger og tidslinjer.
- Forudsigende analyse af potentielle risici og konflikter
- Optimering af indkøbsprocesser og logistikkæder
- Skabe digitale tvillinger af bygninger og strukturer
- Integrationer med smart city- og Internet of Things-systemer

> Tiden til transformation er begrænset - med udviklingen af maskinlæring og computersynsteknologier vil problemet med automatisk klassificering, som har været uløseligt i årtier, blive løst i de kommende år, og bygge- og designvirksomheder, der ikke tilpasser sig i tide, risikerer at gentage skæbnen for taxaflåder, der er blevet fortrængt af digitale platforme.

Automatiseringen af omkostnings- og tidsplanberegninger samt big data og maskinlæring vil blive diskuteret mere detaljeret i bogens femte og niende del. Risikoen for en gentagelse af taxaflådernes skæbne og Uberiseringen af byggebranchen diskuteres i detaljer i bogens tiende del.

For at forstå klassifikatorernes nøglerolle i den digitale transformation af byggebranchen er det nødvendigt at se på deres udviklingshistorie. Det er den historiske kontekst, der gør det muligt for os at forstå, hvordan tilgangen til klassifikation har udviklet sig, og hvilke tendenser der definerer deres nuværende tilstand.

Masterformat, OmniClass, Uniclass og CoClass: udviklingen af klassifikationssystemer

Historisk set har byggeelement- og arbejdsklassifikatorer udviklet sig i tre generationer, der hver især afspejler niveauet af tilgængelig teknologi og branchens aktuelle behov i en bestemt tidsperiode (figur 4.2-8):

- **Første generation** (begyndelsen af 1950'erne til slutningen af 1980'erne) - papirkataloger, hierarkiske klassifikatorer brugt lokalt (f.eks. Masterformat, SfB).
- **Den anden generation** fra slutningen af 1990'erne til midten af 2010'erne) er regneark og strukturerede databaser implementeret i Excel og Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Tredje generation** (2010'erne til i dag- digitale tjenester og API'er - grænseflader, integration med CAD (BIM), automatisering (Uniclass 2015, CoClass).

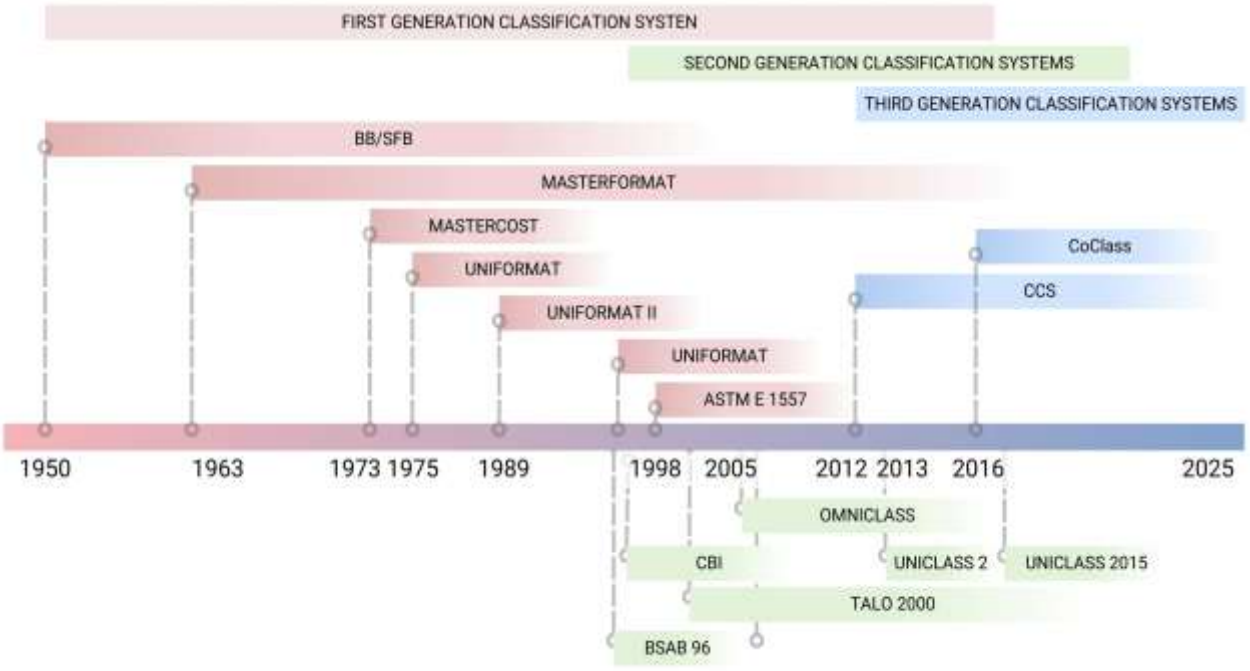


Fig. 4.2-8 Tre generationer af klassifikatorer til byggebranchen.

I løbet af de sidste årtier er der sket en reduktion i den hierarkiske kompleksitet (Fig. 4.2-9) af klassifikatorer: Mens tidlige systemer som OmniClass brugte op til 7 indlejringsniveauer til at beskrive 6887 klasser, er moderne løsninger som CoClass begrænset til 3 niveauer med 750 klasser. Det gør det lettere at arbejde med data, samtidig med at den nødvendige granularitet bevares. Uniclass 2015, der ofte bruges som standard i Storbritannien, kombinerer 7210 klasser på kun 4 niveauer, hvilket gør det praktisk for CAD projekter og offentlige indkøb.

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

Fig. 4.2-9 For hver ny generation af klassifikatorer falder kompleksiteten i kategoriseringen med stormskridt.

I de forskellige landes systemer til vurdering af byggeri kan selv et typisk element som en betonfundament-væg beskrives på meget forskellige måder på grund af forskellige klassifikationer (figur 4.2-10). Disse forskelle afspejler national byggepraksis, anvendte målesystemer, tilgange til materialeklassifikation og de lovmæssige og tekniske krav, der gælder i hvert land.

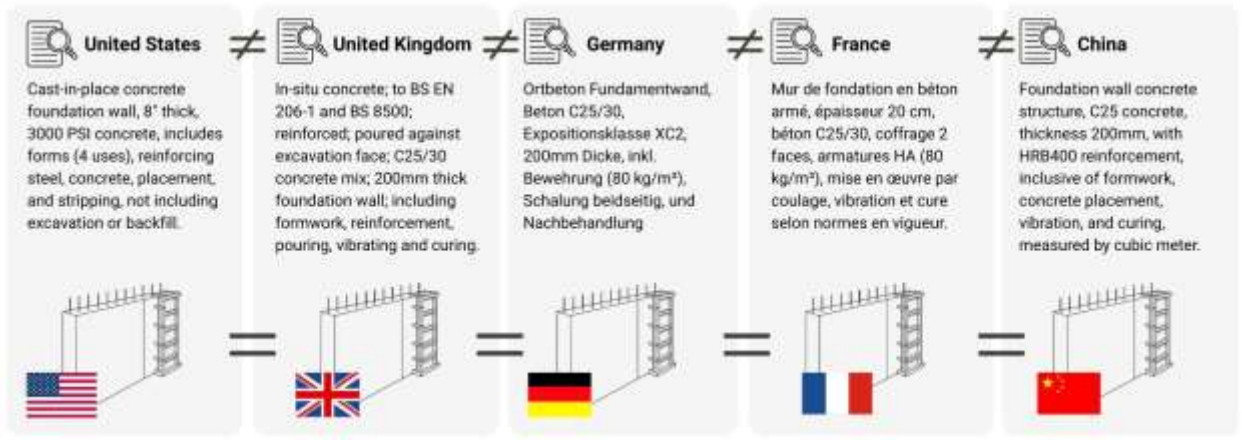


Fig. 4.2-10 Det samme element bruges i projekter i forskellige lande med forskellige beskrivelser og klassifikationer.

De mange forskellige klassifikationer af de samme elementer komplicerer det internationale samarbejde og gør sammenligningen af omkostninger og arbejdets omfang i internationale projekter tidskrævende og nogle gange næsten umulig. I øjeblikket findes der ikke én universel klassifikation på globalt plan - hvert land eller hver region udvikler sine egne systemer baseret på lokale normer, sprog og forretningskultur:

- **CCS** (Danmark): Cost Classification System - et system til at klassificere omkostninger i hele anlæggets livscyklus (design, konstruktion, drift). Vægten ligger på drifts- og vedligeholdelseslogik, men omfatter også budget- og ressourcestyring.
- **NS 3451** (Norge): Kategoriserer faciliteter efter funktion, designelementer og livscyklusstadier. Bruges til projektledelse, omkostningsestimering og langtidsplanlægning.
- **MasterFormat** (USA): Et system til strukturering af byggespecifikationer i sektioner (f.eks. beton, el, efterbehandling). Fokus på discipliner og arbejdstyper snarere end funktionelle elementer (i modsætning til UniFormat).
- **Uniclass 2** (UK): en af de mest detaljerede klassifikatorer, der bruges i offentlige indkøb og BIM projekter. Samler data om objekter, arbejder, materialer og rum i et enkelt system.
- **OmniClass**: en international standard (udviklet af CSI i USA) til håndtering af objektinformation fra komponentbiblioteker til elektroniske specifikationer. Velegnet til langtidsopbevaring af data, kompatibel med CAD (BIM) og andre digitale værktøjer.
- **COBie**: Construction-Operation Building information exchange er en international standard for udveksling af data mellem design-, konstruktions- og driftsfaserne. Indgår i BS 1192-4:2014 som en del af konceptet "BIM - model klar til brug". Fokuserer på informationsoverførsel (f.eks. udstyrsspecifikationer, garantier, entreprenørkontakter).

Globaliseringen af byggebranchen vil sandsynligvis føre til en gradvis ensretning af klassifikationssystemerne for bygningselementer, hvilket vil reducere afhængigheden af lokale nationale standarder betydeligt. Denne proces kan udvikle sig på linje med udviklingen af internetkommunikation, hvor universelle dataoverførselsprotokoller i sidste ende erstattede forskellige lokale formater og sikrede global interoperabilitet mellem systemer.

En alternativ udviklingsvej kan være en direkte overgang til automatiske klassifikationssystemer baseret på maskinlæringsteknologier. Disse teknologier, som i dag hovedsageligt udvikles inden for autonom transport, har et betydeligt potentiale for anvendelse på store CAD-designdatasæt (Fig. 10.1-6).

> I dag er situationen ikke kun begrænset til den nationale klynge af klassifikatorer. På grund af de mange særegenheder, der ikke er taget højde for på nationalt plan, er hver virksomhed nødt til at forene og standardisere de kategorier af elementer og ressourcer, som den arbejder med.

Som regel starter denne proces i det små - med lokale tabeller over objekter eller interne mærkningssystemer. Det strategiske mål er imidlertid at gå over til et fælles sprog til beskrivelse af alle elementer, som ikke kun er forståeligt inden for virksomheden, men også uden for den - ideelt set harmoniseret med internationale klassifikatorer eller brancheklassifikatorer (fig. 4.2-8). Denne tilgang letter integrationen med eksterne partnere, digitale systemer og fremmer dannelsen af ensartede end-to-end-processer inden for objekternes livscyklus.

Før man går over til automatisering og skalerbare IT-systemer, er det nødvendigt enten at bruge klassifikatorer på nationalt niveau eller at opbygge sin egen logiske og entydige struktur til identifikation af elementer. Hvert objekt - det være sig et vindue (figur 4.2-11), en dør eller et teknisk system - skal beskrives på en sådan måde, at det umiskendeligt kan genkendes i enhver virksomheds digitale system. Dette er afgørende i overgangen fra flade tegninger til digitale modeller, der dækker både designfasen og bygningens drift.

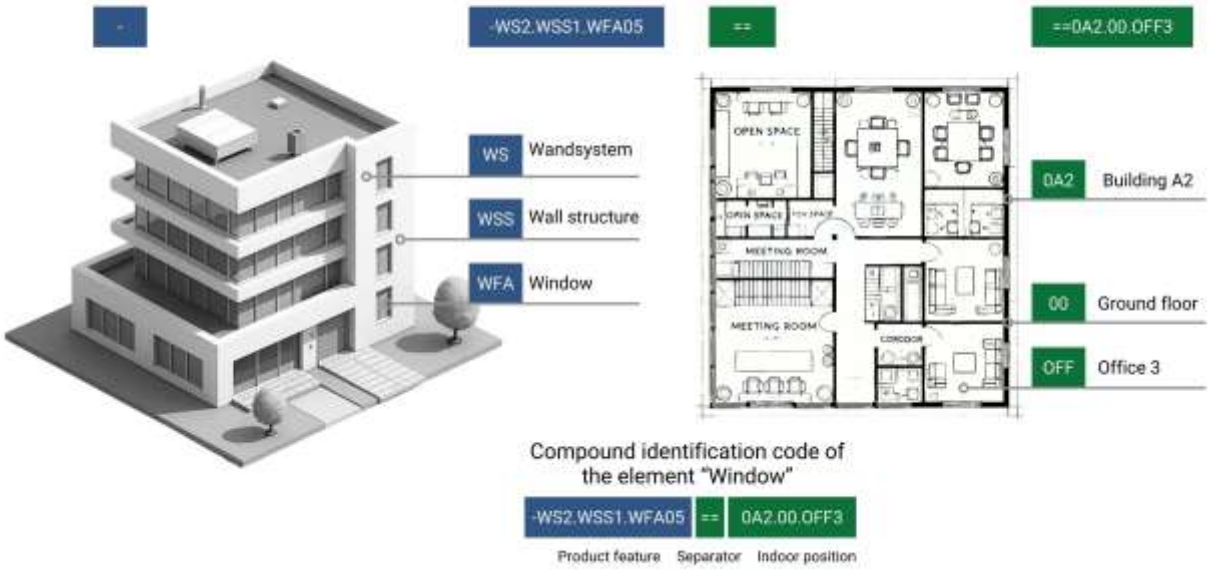


Fig. 4.2-11 Eksempel på identifikation af sammensatte vinduesbygningselementer baseret på klassifikation og placering i bygningen.

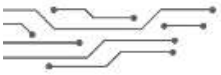
Et eksempel på interne klassifikatorer kunne være udviklingen af en sammensat identifikationskode (figur 4.2-11). En sådan kode kombinerer flere informationsniveauer: elementets funktionelle formål (f.eks. "vindue i væggen"), dets type og den nøjagtige rumlige reference - bygning A2, etage 0, rum 3. En sådan struktur på flere niveauer gør det muligt at skabe et samlet system til navigation gennem digitale modeller og dokumentation, især i stadierne med dataverifikation og -transformation, hvor entydig gruppering af elementer er påkrævet. Entydig elementgenkendelse sikrer konsistens mellem afdelinger og reducerer risikoen for dobbeltarbejde, fejl og tab af information.

En velbygget klassifikator er ikke bare et teknisk dokument, det er grundlaget for en virksomheds digitale økosystem:

- sikrer datakompatibilitet mellem systemer;
- reducerer omkostningerne ved at søge og behandle information;
- øger gennemsigtigheden og håndterbarheden;
- skaber et fundament for skalering og automatisering.

Den standardiserede beskrivelse af objekter gennem brug af nationale klassifikatorer eller proprietære sammensatte identifikationskoder bliver grundlaget for konsistente data, pålidelig informationsudveksling og den efterfølgende implementering af intelligente tjenester - fra automatiserede indkøb til digitale tvillinger.

Når man har gennemført struktureringsfasen for data i flere formater og valgt den klassifikator, der skal bruges til at genkende og gruppere elementerne, er næste skridt at modellere dataene korrekt. Denne proces omfatter identifikation af nøgleparametre, opbygning af en logisk datastruktur og beskrivelse af forholdet



mellem elementer.

KAPITEL 4.3.
DATAMODELLERING OG EKSPERTISECENTER

Datamodellering: konceptuel, logisk og fysisk model

Effektiv håndtering af data (som vi tidligere har struktureret og kategoriseret) er umulig uden en gennemtænkt opbevarings- og behandlingsstruktur. For at sikre adgang til og konsistens af information i lagrings- og behandlingsfasen bruger virksomheder datamodellering, en metode, der giver dem mulighed for at designe tabeller, databaser og links mellem dem i overensstemmelse med forretningskravene.

> Datamodellering er det fundament, som ethvert digitalt økosystem er bygget på. Uden en beskrivelse af systemer, krav og datamodellering ved og forstår ingeniører og fagfolk, der skaber data, ikke, hvad de data, de skaber, skal bruges til.

Ligesom når man bygger en bygning, hvor man ikke kan begynde at lægge mursten uden en plan, kræver oprettelsen af et data warehouse-system en klar forståelse af, hvilke data der skal bruges, hvordan de skal forbindes, og hvem der skal arbejde med dem. Uden en beskrivelse af processer og krav mister de ingeniører og fagfolk, der skaber dataene, overblikket over, hvor og hvordan dataene vil blive brugt i fremtiden.

Datamodellen fungerer som en bro mellem forretning og IT. Den gør det muligt at formalisere krav, strukturere information og lette kommunikationen mellem interessenter. I den forstand svarer datamodellering til en arkitekts arbejde, som i henhold til kundens plan udvikler en bygningsplan og derefter overdrager den til håndværkerne - databaseadministratorer og udviklere - til implementering (databaseoprettelse).

Derfor skal enhver byggevirksomhed, ud over at strukturere og kategorisere elementer og ressourcer (fig. 4.2-11), mestre kunsten at "bygge" databaser (tabeller) og lære at skabe forbindelser mellem dem, som om man forbinder murstenene til en pålidelig og stærk mur af viden fra virksomhedens data. Nøglebegreberne i datamodellering (figur 4.3-1) omfatter:

- **Enheder** er objekter, som der skal indsamles data om. I den tidlige designfase kan en enhed være et enkelt element (f.eks. "dør"), og i beregningsmodellen kan det være en gruppe af elementer, der er kategoriseret (f.eks. "indvendige døre").
- **Attributter** er egenskaber ved enheder, der beskriver vigtige detaljer: dimensioner, egenskaber, monteringsomkostninger, logistik og andre parametre.
- **Relationer (links)** - viser, hvordan enheder interagerer med hinanden. De kan være af en af typerne: "en til en", "mange til en", "mange til mange".
- ER-diagrammer (Entity-Relationship-diagrammer) er visuelle diagrammer, der viser entiteter, attributter og relationerne mellem dem. ER-diagrammer kan være konceptuelle, logiske og fysiske - hver især afspejler de et forskelligt detaljeringsniveau.

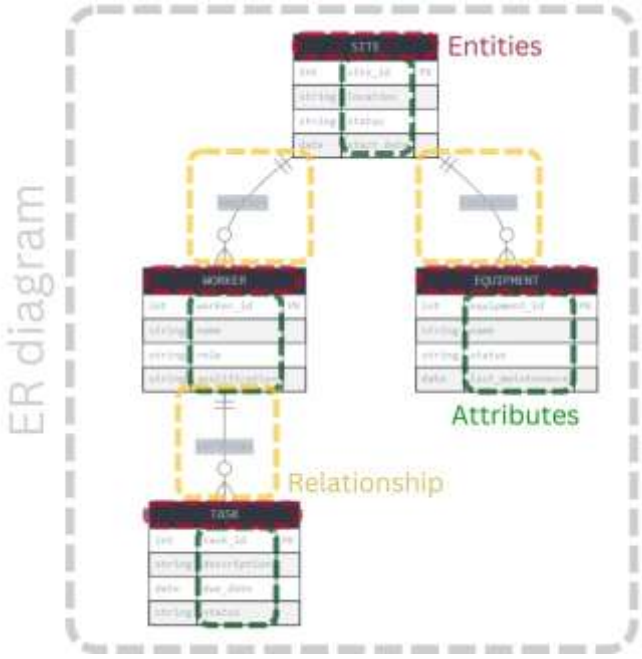


Fig. 4.3-1 ER-diagram over en konceptuel databasestruktur med enheder, attributter og relationer.

Processen med at designe data og definere relationer mellem dem er traditionelt opdelt i tre hovedmodeller. Hver af dem opfylder visse funktioner og varierer i detaljeringsgrad og abstraktionsniveau i repræsentationen af datastrukturen:

- **Konceptuel datamodel**: Denne model beskriver de vigtigste enheder og deres relationer uden at gå i attributdetaljer. Den bruges normalt i de indledende faser af planlægningen. På dette stadie kan vi skitsere fra databaser og systemer for at vise forholdet mellem forskellige afdelinger og specialister.

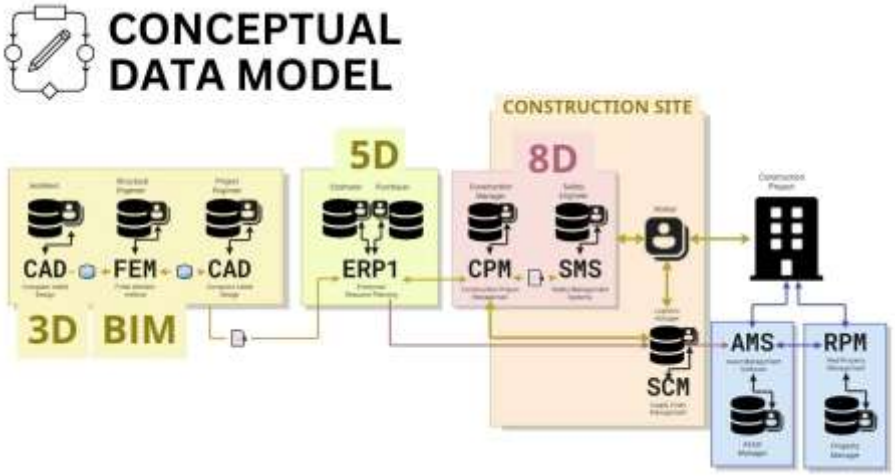
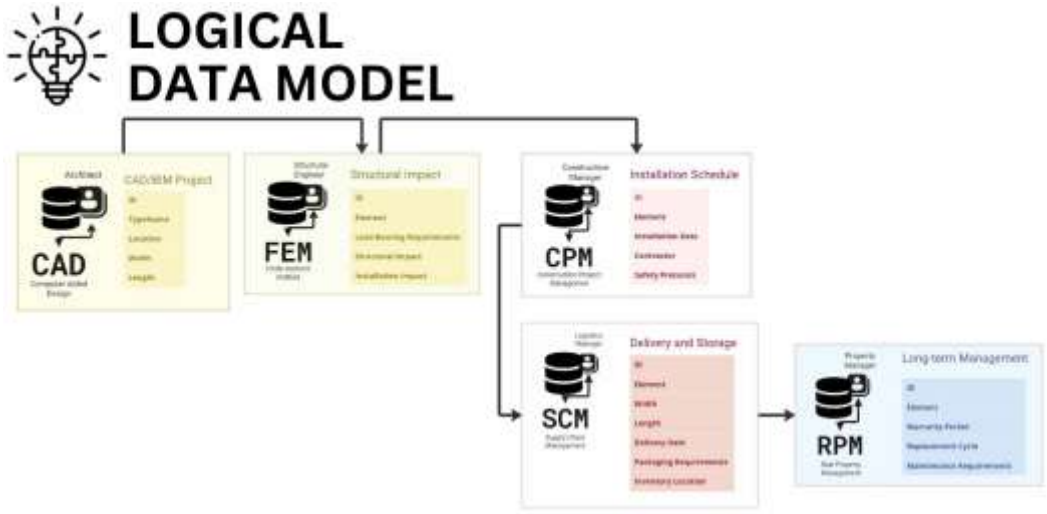


Fig. 4.3-2 Konceptdiagrammet beskriver systemets indhold: en repræsentation på højt niveau af relationerne uden tekniske detaljer.

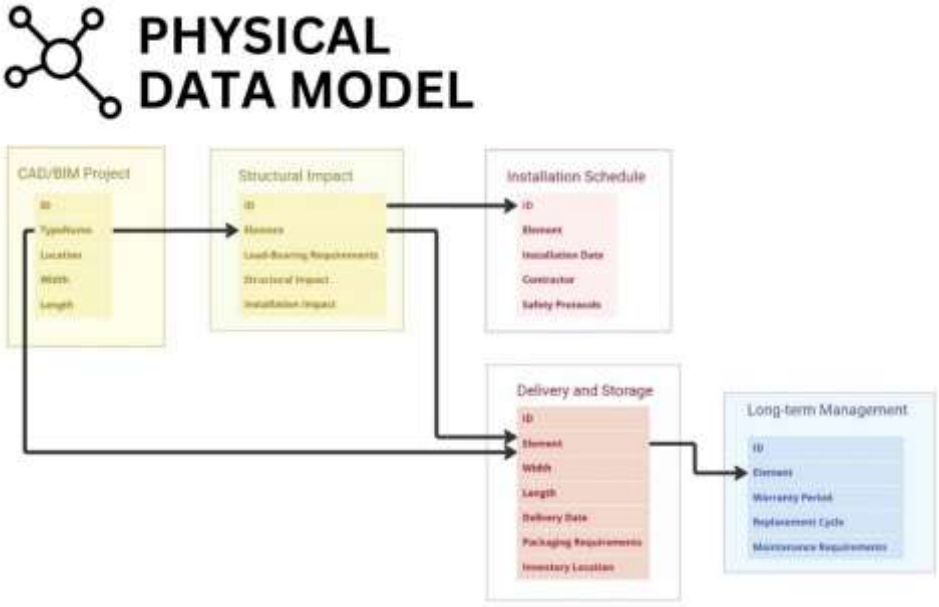
- **Logisk** datamodel: Baseret på den konceptuelle model indeholder den logiske datamodel detaljerede beskrivelser af enheder, attributter, nøgler og relationer, der kortlægger forretningsinformation og

regler.



Figur 4.3-3 Logisk datamodel beskriver datatyper, relationer og nøgler i detaljer, men uden systemimplementering.

- **Fysisk datamodel**: Denne model beskriver de nødvendige strukturer til implementering af en database, herunder tabeller, kolonner og relationer. Den fokuserer på databasens ydeevne, indekseringsstrategier og fysisk lagring for at optimere den fysiske implementering af databaser.



Figur 4.3-4 Den fysiske datamodel definerer, hvordan systemet skal implementeres, herunder tabellerne og de specifikke detaljer i databasen.

> Når man designer databaser og tabelrelationer, spiller forståelsen af abstraktionsniveauerne en vigtig rolle i opbygningen af en effektiv systemarkitektur.

En effektiv datamodelleringsmetode giver dig mulighed for at kombinere forretningsmæssige mål med teknisk implementering, hvilket gør hele proceskæden mere gennemsigtig og håndterbar. Datamodellering er ikke en engangsopgave, men en proces med sekventielle trin (figur 4.3-5):

- **Indsamling af forretningskrav**: Nøgleopgaver, mål og informationsstrømme defineres. Dette er fasen med aktiv interaktion med eksperter og brugere.
- **Entitetsidentifikation: De** vigtigste objekter, kategorier og datatyper, der er vigtige at tage højde for i det fremtidige system, fremhæves.
- **Udvikling af en konceptuel og logisk model:** Først indfanges de vigtigste enheder og deres relationer, derefter attributter, regler og detaljeret struktur.
- **Fysisk modellering: Den** tekniske implementering af modellen designes: tabeller, felter, relationer, begrænsninger, indekser.
- **Oprettelse af database:** Det sidste trin er at implementere den fysiske model i det valgte DBMS, udføre test og gøre klar til drift.

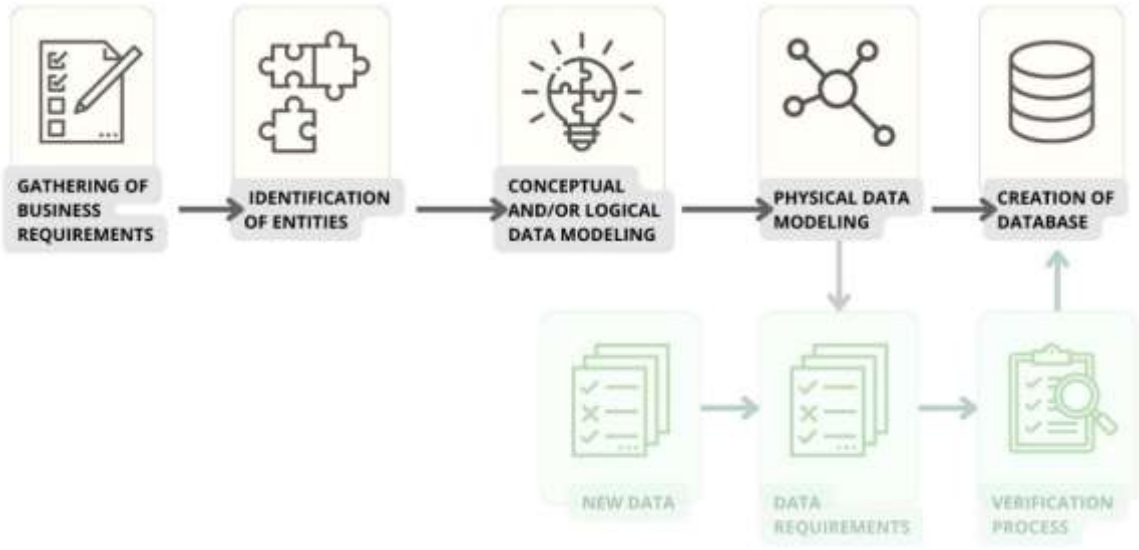


Fig. 4.3-5 Oprettelse af databaser og datastyringssystemer til forretningsprocesser begynder med generering af krav og datamodellering.

Korrekt designede datamodelleringsprocesser giver mulighed for gennemsigtige informationsstrømme, hvilket er særligt vigtigt i komplekse projekter som f.eks. byggeprojekt- eller byggepladsledelse. Lad os se på, hvordan det at gå fra en konceptuel model til en logisk model og derefter til en fysisk model kan hjælpe med at strømline processerne.

Praktisk datamodellering i forbindelse med byggeri

Lad os tage en byggepladsledelsesopgave som et eksempel på datamodellering og konvertere kravene fra til en struktureret logisk model. Baseret på de grundlæggende behov for byggepladsledelse definerer vi nøgleenheder for: byggeplads (SITE), medarbejdere (WORKER), udstyr (EQUIPMENT), opgaver (TASK) og brug af udstyr (EQUIPMENT_USAGE). Hver enhed indeholder et sæt attributter, der afspejler vigtige egenskaber. For TASK kan det f.eks. være en beskrivelse af opgaven, forfaldsdato, status, prioritet; for WORKER kan det være navn, rolle på stedet, nuværende ansættelse osv.

Den logiske model etablerer relationer mellem disse enheder og viser, hvordan de interagerer med hinanden i virkelige arbejdsprocesser (figur 4.3-6). For eksempel indikerer forholdet mellem sted og medarbejdere, at mange medarbejdere kan arbejde på et sted, mens forholdet mellem medarbejdere og opgaver afspejler, at en medarbejder kan udføre flere opgaver.

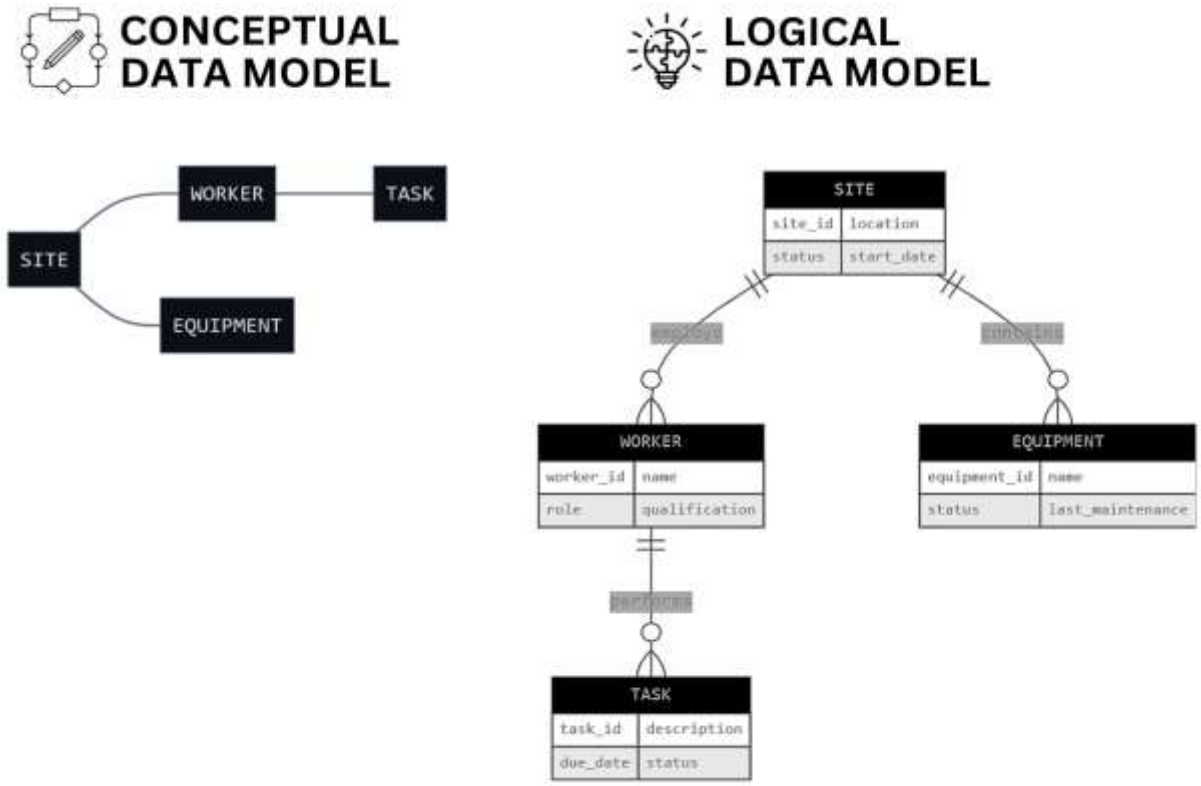
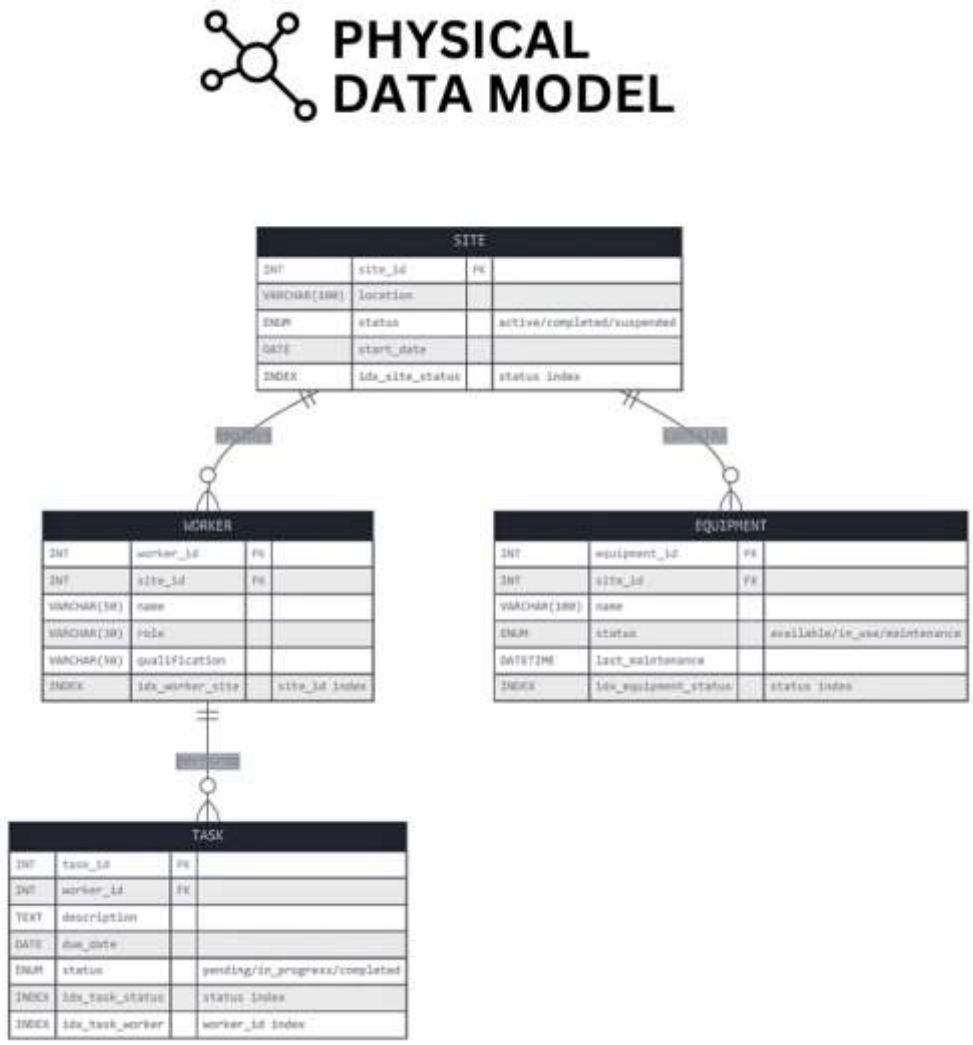


Fig. 4.3-6 Konceptuel og logisk datamodel genereret af formandens krav til beskrivelse af byggepladsprocesser.

Når man går over til den fysiske model, tilføjes tekniske implementeringsdetaljer: specifikke datatyper (VARCHAR, INT, DATE), primær- og fremmednøgler til relationer mellem tabeller og indekser for at optimere databasens ydeevne (figur 4.3-7).

For eksempel skal der defineres specifikke typer med mulige værdier for statusser, og der skal tilføjes indekser på nøglefelter som status og worker_id for at forbedre søgeydelsen. På den måde bliver en logisk beskrivelse af systemet til en konkret databaseimplementeringsplan, som er klar til at blive oprettet og implementeret.



Figur 4.3-7 Den fysiske datamodel beskriver enhederne på en byggeplads ved hjælp af de nødvendige minimumsparametre.

> Den fysiske model er ofte forskellig fra den logiske model. I gennemsnit er fordelingen af modelleringstiden som følger: ca. 50% bruges på den konceptuelle model (indsamling af krav, diskussion af processer, identifikation af enheder), 10% på den logiske model (specificering af attributter og relationer) og 40% på den fysiske model (implementering, testning, tilpasning til DBMS).

Denne balance skyldes, at den konceptuelle fase lægger fundamentet for datastrukturen, mens den logiske model kun specificerer relationer og attributter. Den fysiske model kræver flest ressourcer, da det er i denne fase, at data implementeres i specifikke platforme og værktøjer

Oprettelse af en database ved hjælp af LLM

Når vi har en datamodel og en beskrivelse af enheder gennem parametre, er vi klar til at oprette databaser - lagre, hvor vi vil gemme oplysninger, der kommer efter struktureringsfasen om specifikke processer.

Lad os prøve at skabe et eksempel på en enkel, men funktionel database med et minimum af kode ved hjælp af SQLite ved hjælp af eksemplet med programmeringssproget Python. Relationsdatabaser blev diskuteret i detaljer i kapitlet "Strukturerede relationsdatabaser og SQL-forespørgselssprog ".

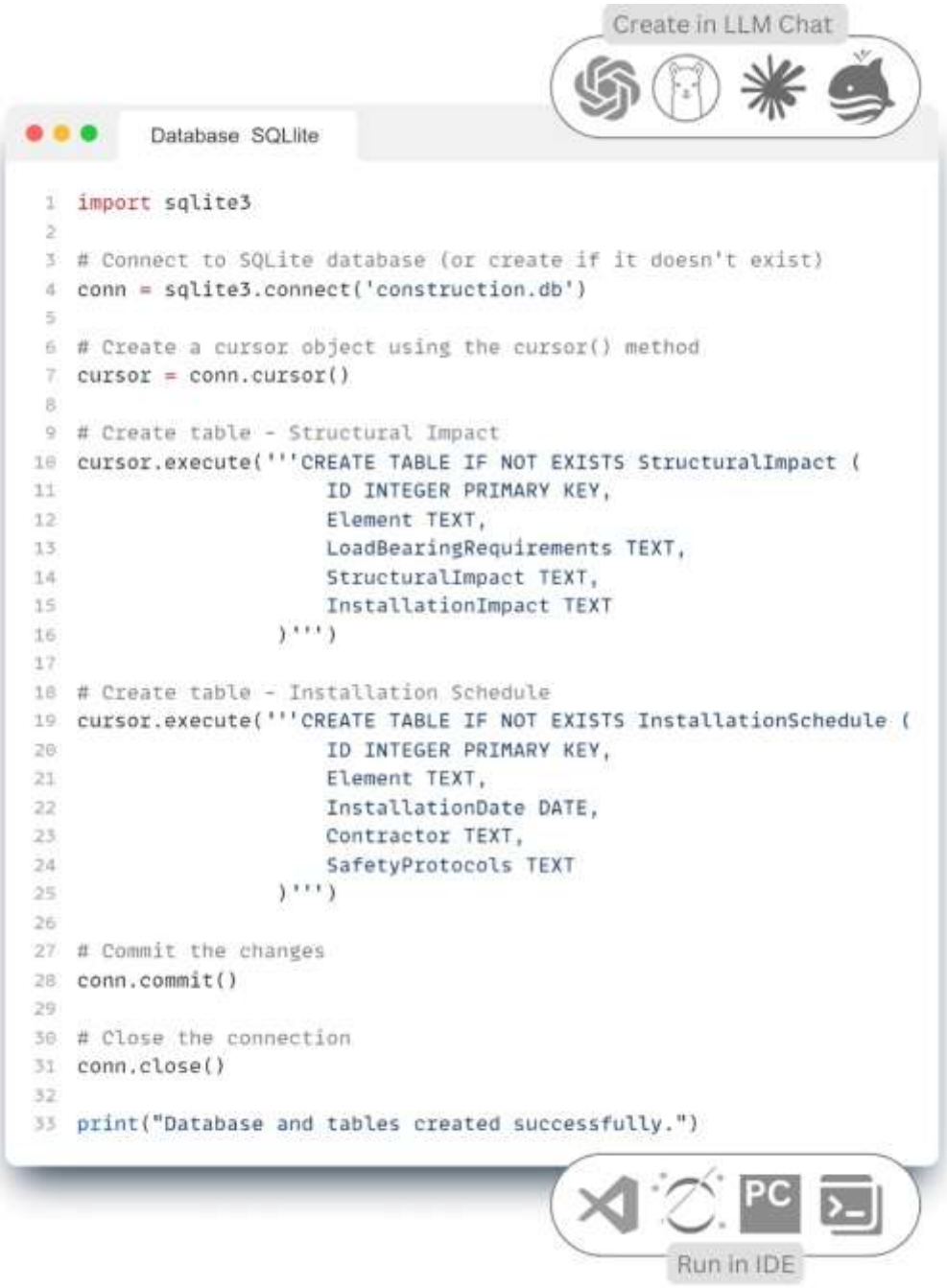
> SQLite er en let, open source, integrerbar database på tværs af platforme, som understøtter centrale SQL-funktioner. SQLite bruges i vid udstrækning i CAD (BIM), ERP, EPM og findes i næsten alle smartphones.

Lad os bede LLM -modellen om at oprette en lille SQLlite-database til os baseret på den fysiske model, der er oprettet ovenfor (figur 4.3-4) for FEM- (strukturel påvirkning) og CPM (montageplan) systemerne.

- Send en tekstanmodning til LLM chat (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller andre):

> Opret SQLlite-databaser for systemerne "Constructive Impact" og "Installation Schedule" ud fra figuren i bilaget (indsæt skemaet i chatten som et billede (Fig. 4.3-4))⏎

LLM's svar :



Figur 4.3-8 Den genererede kode opretter databaser og tabeller ud fra det uploadede billede, som vi nu kan indlæse data i.

Denne databasekode kan køres i Python-miljøet med SQLite-understøttelse, som normalt er inkluderet i standard-Python-miljøet. Den kan køres på enhver lokal Python-maskine, der er i stand til at køre scripts og oprette filer. Når den er kørt og udført, opretter koden en fil på din harddisk - en SQLite-database med navnet construction.db svarende til en Excel-fil med regneark), som indeholder tabellerne "Structural Impact" og "Installation Schedule".

Når en ny **construction.db-database** er oprettet i disse tabeller, kan der tilføjes data til disse tabeller via SQL _forespørgsler eller import, hvilket giver dig mulighed for at begynde at oprette automatisk databehandling senere. Data kan importeres til SQLite-databasen fra CSV-filer, Excel-regneark eller eksporteres fra andre databaser og arkiver via API'en.

For at etablere bæredygtig datamodellering og effektive databasestyringsprocesser har en virksomhed brug for en klart defineret strategi samt koordinering mellem tekniske og forretningsmæssige teams. Med forskellige projekter og flere datakilder er det ofte svært at sikre konsistens, standardisering og kvalitetskontrol på alle niveauer. En vigtig løsning kan være at oprette et dedikeret Center of Excellence (COE) for datamodellering i virksomheden.

Kompetencecenter (CoE) for datamodellering

Da data er ved at blive et af de vigtigste strategiske aktiver, skal virksomhederne gøre mere end blot at indsamle og opbevare information korrekt - det er vigtigt at lære at håndtere data systematisk. Kompetencecentret for klassifikation og datamodellering (CoE) er en strukturel enhed, der sikrer konsistens, kvalitet og effektivitet i al datahåndtering i organisationen.

> Centre of Excellence (CoE) er kernen i ekspertsupport og det metodiske fundament for digital transformation i en virksomhed. Det opbygger en datadrevet kultur og gør det muligt for organisationer at opbygge processer, der træffer beslutninger baseret på strukturerede, validerede og repræsentative data i stedet for på intuition eller lokal information.

Et datacenter of excellence består normalt af tværfunktionelle teams, der arbejder efter "to pizzaer"-princippet. Dette princip, som er foreslået af Jeff Bezos, betyder, at teamet skal være så stort, at det kan brødfødes med to pizzaer, dvs. ikke mere end 6-10 personer. Denne tilgang hjælper med at undgå overdrevent bureaukrati og øger arbejdets fleksibilitet. CoE-teamet bør omfatte medarbejdere med mange forskellige tekniske færdigheder, fra dataanalyse og maskinlæring til ekspertise inden for specifikke forretningsområder. Med deres dybe tekniske viden bør dataingeniører ikke kun optimere processer og modellere data, men også støtte kolleger ved at reducere tiden til rutineopgaver (fig. 4.3-9).

> Ligesom økosystemets modstandsdygtighed i naturen sikres af biodiversitet, opnås fleksibilitet og tilpasningsevne i den digitale verden gennem en mangfoldighed af tilgange til håndtering af data. Men denne mangfoldighed skal understøttes af fælles regler og koncepter.

Et ekspertisecenter (CoE) kan sammenlignes med "klimaforholdene" i et skovøkosystem, som bestemmer, hvilke typer data der vil blomstre, og hvilke der automatisk vil blive kasseret. Ved at skabe et gunstigt "klima" for kvalitetsdata letter CoE det naturlige valg af bedste praksis og metoder, som derefter bliver standarder for organisationen.

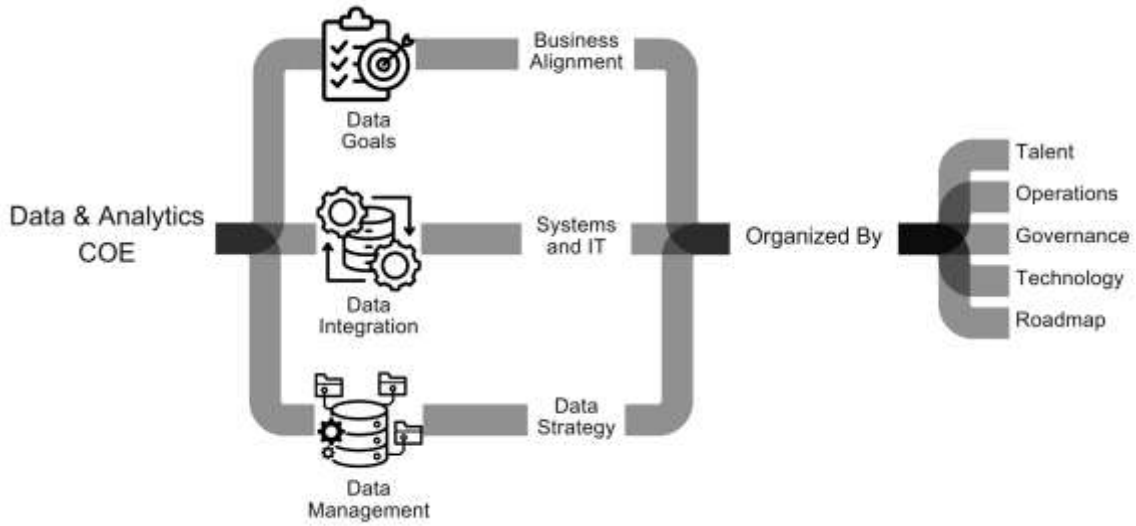


Fig. 4.3-9 Kompetencecentret (CoE) for data og analyse samler ekspertise inden for centrale aspekter af datastyring, -integration og -strategi.

For at fremskynde integrationscyklusser og opnå bedre resultater bør CoE give sine medlemmer en tilstrækkelig grad af selvstændighed i beslutningstagningen. Det er især vigtigt i et dynamisk miljø, hvor forsøg og fejl, konstant feedback og hyppige udgivelser kan give betydelige fordele. Denne autonomi er dog kun effektiv, hvis der er klar kommunikation og støtte fra den øverste ledelse. Uden strategisk vision og koordinering på højeste niveau kan selv det mest kompetente team møde barrierer i implementeringen af deres initiativer.

> Det er COE eller virksomhedens øverste ledelse, der er ansvarlig for at sikre, at datamodelleringstilgangen ikke er begrænset til et eller to projekter, men er indlejret i det overordnede informationsstyrings- og forretningsprocesstyringssystem.

Ud over sine opgaver i forbindelse med datamodellering og datastyring er ekspertisecentret (CoE) ansvarligt for at udvikle fælles standarder og tilgange til udrulning og drift af datainfrastrukturen. Derudover fremmer det en kultur med løbende forbedringer, procesoptimering og effektiv brug af data i organisationen (figur 4.3-10).

Den systematiske tilgang til data- og modelstyring inden for CoE kan groft sagt opdeles i flere nøgleblokke:

- **Standardisering af processer og styring af modellens livscyklus:** CoE udvikler og implementerer metoder til at ensrette oprettelsen og styringen af datamodeller. Dette omfatter: etablering af strukturelle skabeloner, kvalitetskontrolmetoder og versionskontrolsystemer for at sikre datakontinuitet på tværs af alle arbejdsfaser.
- **Rollestyring og ansvarstildeling:** COE definerer nøgleroller i datamodelleringsprocessen. Hver projektdeltager tildeles klart definerede roller og ansvarsområder, hvilket letter teamarbejdet og reducerer risikoen for uoverensstemmelser i data.
- **Kvalitetskontrol og revision:** Effektiv styring af byggedata kræver løbende overvågning af kvaliteten. Automatiserede mekanismer til kontrol af data, identifikation af fejl og manglende attributter

er ved at blive implementeret.

- **Ledelse af metadata og informationsarkitektur:** CoE er ansvarlig for at skabe et samlet system af klassifikation og identifikatorer, navngivning og standarder for enhedsbeskrivelser, som er afgørende for integration mellem systemer.

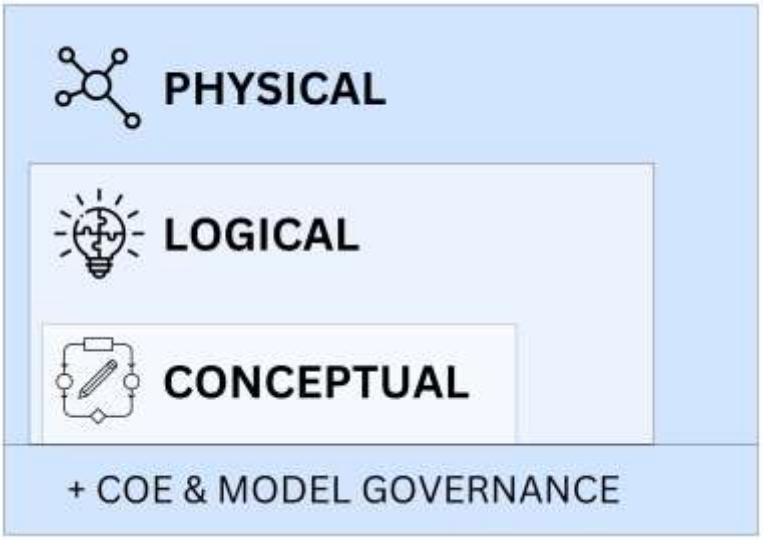


Fig. 4.3-10 Datamodellering og datakvalitetsstyring er en af de største udfordringer for CoE

Kompetencecentret (CoE) for data er ikke bare en gruppe eksperter, men en systemisk mekanisme, der skaber en ny datadrevet kultur og sikrer en ensartet tilgang til arbejdet med data i hele virksomheden. Gennem kompetent integration af modelleringsprocesser i det overordnede informationsstyringssystem, standardisering, klassificering og datakvalitetskontrol hjælper CoE virksomheder med løbende at forbedre deres produkter og forretningsprocesser, reagere hurtigere på markedsændringer og træffe informerede beslutninger baseret på pålidelige analyser.

Sådanne centre er særligt effektive, når de kombineres med moderne DataOps-principper - under et tiltag, der sikrer kontinuerlig levering, automatisering og kvalitetskontrol af data. Vi vil tale mere om DataOps i del 8 i kapitlet "Moderne datateknologier i byggebranchen".

I de følgende kapitler bevæger vi os fra strategi til praksis - lad os betinget "transformere" til et datacenter: Vi ser på flere eksempler på, hvordan opgaveparameterisering, kravindsamling og den automatiske valideringsproces finder sted.

KAPITEL 4.4.
SYSTEMATISERING AF KRAV OG VALIDERING AF INFORMATION

Indsamling og analyse af krav: omdanner kommunikation til strukturerede data

Indsamling og håndtering af krav er det første skridt til at sikre datakvalitet. På trods af udviklingen af digitale værktøjer formuleres de fleste krav stadig på en ustruktureret måde: gennem breve, mødereferater, telefonopkald og mundtlige diskussioner. Denne form for kommunikation gør det vanskeligt at automatisere, validere og genbruge information. I dette kapitel ser vi på, hvordan man oversætter tekstuelle krav til formelle strukturer og sikrer, at forretningskravene er gennemsigtige og systematiske.

> Gartners forskning, "Data Quality: Best Practices for Accurate Insights", fremhæver den kritiske betydning af datakvalitet for vellykkede data- og analyseinitiativer [100]. De bemærker, at dårlig datakvalitet i gennemsnit koster organisationer mindst 12,9 millioner dollars om året, og at pålidelige data af høj kvalitet er afgørende for at skabe en datadrevet virksomhed.

Manglen på strukturerede krav fører til, at det samme element (enhed) og dets parametre kan være lagret i forskellige systemer i forskellige variationer. Det reducerer ikke kun processernes effektivitet, men resulterer også i spildtid, duplikering af information og behov for at genvalidere data, før de kan bruges. Som følge heraf kan selv en enkelt udeladelse - en mistet parameter eller et enkelt forkert beskrevet element - forsinke beslutningsprocessen og medføre ineffektiv brug af ressourcer.

> *I mangel af et søm gik hesteskoen tabt.*
> *I mangel af en hestesko gik hesten tabt.*
> *I mangel af en hest gik rytteren tabt.*
> *I mangel af en rytter gik et budskab tabt.*
> *I mangel af et budskab gik slaget tabt.*
> *I mangel af et slag gik et kongerige tabt.*
> *Alt sammen på grund af et manglende søm i hesteskoen.*
>
> - Ordsprog [101]

Analyse og indsamling af krav til processen med at udfylde og lagre data starter med at identificere alle interessenter. Ligesom det ordsprogede tab af et enkelt søm fører til en kæde af kritiske konsekvenser, kan tabet af en enkelt interessent, et overset krav eller tabet af selv en enkelt parameter have betydelig indflydelse på ikke bare en individuel forretningsproces, men hele økosystemet i et projekt og organisationen som helhed. Derfor er det afgørende at identificere selv de elementer, parametre og roller, som ved første øjekast virker ubetydelige, men som senere kan vise sig at være kritiske for virksomhedens bæredygtighed.

Lad os forestille os, at en virksomhed har et projekt, hvor *bygherren* fremsætter en ny anmodning - "*tilføj et ekstra vindue på nordsiden af bygningen*". Den lille proces "*bygherrens anmodning om at tilføje et nyt vindue til det nuværende projekt*" involverer arkitekt, bygherre, CAD-specialist (BIM), byggeleder, logistikchef, ERP - analytiker, kvalitetskontrolingeniør, sikkerhedsingeniør, kontrolchef og ejendomsadministrator.

> Selv en lille proces kan involvere dusinvis af forskellige specialister. Hver procesdeltager skal forstå kravene fra de specialister, som de er forbundet med på dataniveau.

På tekstniveau (fig. 4.4-1) foregår kommunikationen mellem kunden og specialisterne i proceskæden på følgende måde:

- **Kunde:** "*Vi har besluttet at tilføje et ekstra vindue på nordsiden for at få bedre lys. Kan det lade sig gøre*?"
- **Arkitekt**: "Ja*, jeg vil revidere projektet, så det omfatter det nye vindue, og sende opdaterede CAD-planer til (BIM*)".
- **CAD-specialist (BIM):** "*Modtager et nyt projekt. Jeg opdaterer CAD (BIM) modellen med det ekstra vindue og efter koordinering med FEM-ingeniøren leverer jeg den nøjagtige placering og dimensioner af det nye vindue*."
- **Byggeleder**: "*Vi har modtaget et nyt projekt. Vi justerer installationsdatoerne for 4D og informerer alle relevante underleverandører*."
- **Anlægsingeniør (CAFM**): "*Jeg vil indtaste data fra 6D på det nye vindue i CAFM-systemet til fremtidig facilitetsstyring og vedligeholdelsesplanlægning*."
- **Logistikchef**: "*Jeg skal bruge målene og vægten på det nye vindue for at kunne organisere leveringen af vinduet til byggepladsen*".
- **ERP** -analytiker: "*Jeg har brug for omfangstabellerne og den nøjagtige vinduestype til 5D budgetopdatering i vores ERP-system for at afspejle omkostningerne til det nye vindue i det samlede projektestimat*."
- **Kvalitetskontrolingeniør**: "*Når vinduesspecifikationerne er klar, sørger jeg for, at de lever op til vores kvalitets- og materialestandarder.*"
- **Sikkerhedsingeniør**: "*Jeg vil vurdere de sikkerhedsmæssige aspekter af det nye vindue med særligt fokus på overholdelse og evakuering i henhold til 8D-ordningen ".*
- **Kontrolchef**: "*Baseret på det nøjagtige arbejdsomfang fra ERP opdaterer vi vores 4D-tidslinje, så den afspejler installationen af det nye vindue, og gemmer de nye data i projektets content management-system*."
- **Arbejder (installatør)**: "*Du har brug for instruktioner om installation, montering og timing af arbejdet. Er der desuden indført særlige sikkerhedsregler, som jeg skal følge?*"
- **Ejendomsadministrator**: "*Når den er installeret, vil jeg dokumentere garanti- og vedligeholdelsesoplysninger til brug for den langsigtede administration*".
- **Asset Manager**: "*Udstyrsingeniør, send venligst endelige data til sporing af aktiver og livscyklusstyring*."
- **Kunden**: "*Vent, måske har jeg travlt, og der bliver ikke brug for vinduet. Måske skulle jeg lave en balkon*".

I sådanne scenarier, som sker ofte, forårsager selv en lille ændring en kædereaktion mellem flere systemer og roller. I dette tilfælde er næsten al kommunikation i den indledende fase i tekstform: e-mails, chats, mødereferater (fig. 4.4-1).

I et sådant tekstbaseret kommunikationssystem til et byggeprojekt er det meget vigtigt med et system til juridisk bekræftelse og registrering af alle dataudvekslingsoperationer og alle beslutninger, der træffes. er for at sikre, at enhver beslutning, instruktion eller ændring er juridisk gyldig og kan spores, hvilket reducerer risikoen for fremtidige "misforståelser".

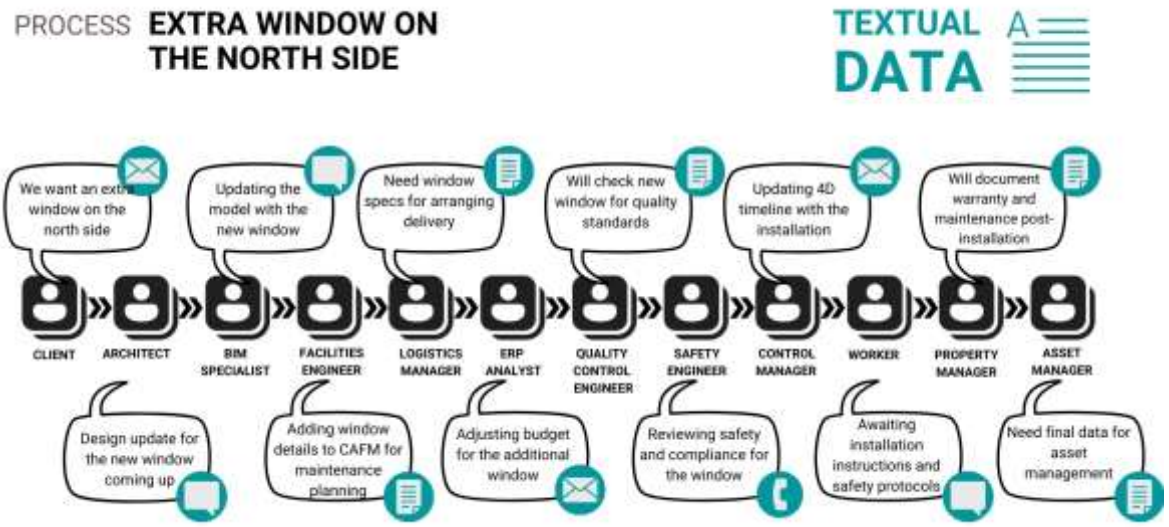


Fig. 4.4-1 Kommunikation mellem bygherre og entreprenør i de indledende faser af et projekt indeholder ofte tekstdata i flere formater.

> Manglen på juridisk kontrol og validering af beslutninger i de relevante systemer i et byggeprojekt kan føre til alvorlige problemer for alle involverede. Enhver beslutning, ordre eller ændring, der foretages uden ordentlig dokumentation og validering, kan føre til tvister (og retssager).

Juridisk konsolidering af alle beslutninger i tekstkommunikation kan kun sikres ved et stort antal underskrevne dokumenter, som vil falde på skuldrene af ledelsen, der er forpligtet til at registrere alle transaktioner. Hvis alle deltagere skal underskrive dokumenter for hver eneste handling, mister systemet sin fleksibilitet og bliver til en bureaukratisk labyrint. Manglende transaktionsbekræftelser vil ikke kun forsinke projektimplementeringen, men kan også føre til økonomiske tab og forringede relationer mellem deltagerne, helt op til og med juridiske problemer.

En sådan transaktionsgodkendelsesproces, som normalt starter med tekstbaserede diskussioner, udvikler sig gradvist til en dokumentudveksling i flere formater i de følgende faser (Fig. 4.4-2), hvilket komplicerer den kommunikation, der tidligere kun foregik via tekst, betydeligt. Uden klart definerede krav bliver det næsten umuligt at automatisere sådanne processer, som er fyldt med data i flere formater og et stort antal tekstkrav.

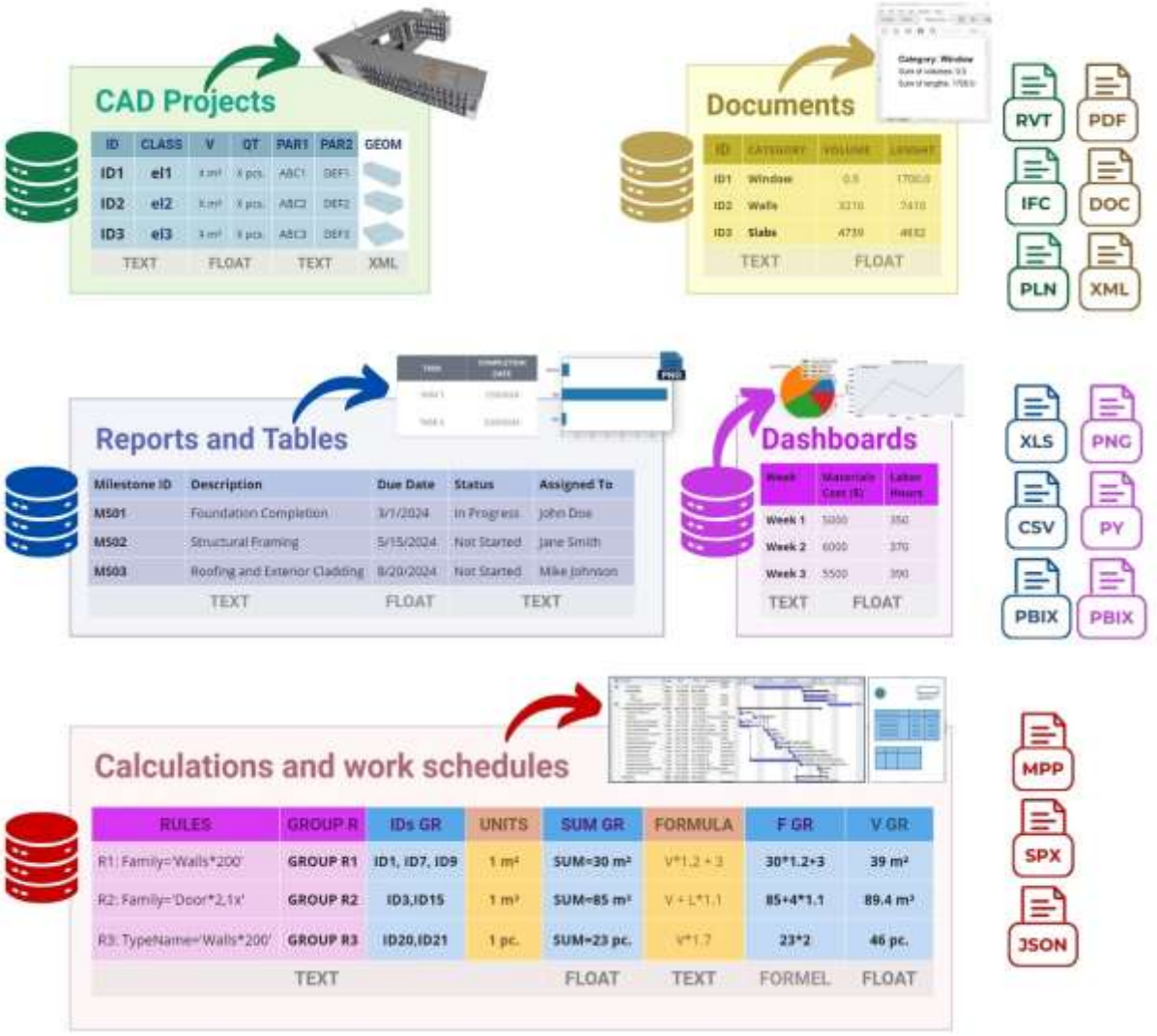


Fig. 4.4-2 Hvert system i byggefirmaets landskab fungerer som kilde til juridisk relevante dokumenter i en række forskellige formater.

Tekstkommunikation kræver, at hver fagperson enten sætter sig ind i hele korrespondancen eller deltager i alle møder regelmæssigt for at forstå projektets aktuelle status.

For at overvinde denne begrænsning er det nødvendigt med en overgang fra tekstlig kommunikation til en struktureret kravmodel. Det er kun muligt gennem systematisk analyse, procesvisualisering og beskrivelse af interaktioner i form af flowdiagrammer og datamodeller (fig. 4.4-3). Ligesom i datamodelleringen (Fig. 4.3-7) bevægede vi os fra det kontekstuelle idéniveau til det konceptuelle niveau ved at tilføje de systemer og værktøjer, som deltagerne bruger, og forbindelserne mellem dem.

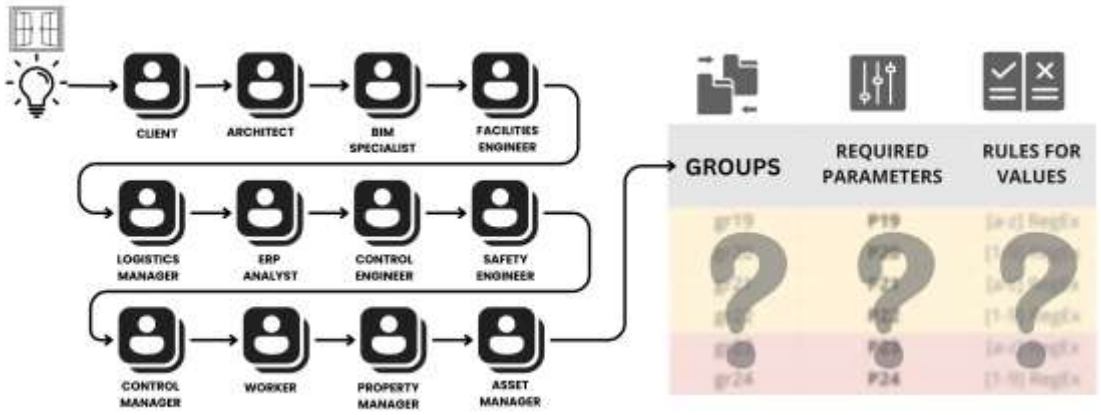


Fig. 4.4-3 For at lære, hvordan man styrer og automatiserer valideringsprocessen, er det nødvendigt at visualisere processerne og strukturere kravene.

Det første skridt i systematiseringen af krav og relationer er at visualisere alle links og relationer ved hjælp af konceptuelle flowdiagrammer. Det konceptuelle niveau vil ikke kun gøre det lettere for alle procesdeltagere at forstå hele proceskæden, men vil også tydeligt vise, hvorfor og for hvem data (og krav) er nødvendige på hvert procestrin.

Procesflowdiagrammer og effektiviteten af konceptuelle rammer

For at bygge bro over kløften mellem traditionelle og moderne tilgange til datahåndtering skal virksomheder foretage et bevidst skift fra fragmenterede tekstbeskrivelser til strukturerede procesrepræsentationer. Udviklingen af data - fra lertavler til digitale økosystemer - kræver nye tænkeværktøjer. Og et sådant værktøj er konceptuel modellering ved hjælp af flowcharts. Ved at skabe visuelle diagrammer - flowcharts, procesdiagrammer, interaktionsdiagrammer - kan projektdeltagerne se, hvordan deres handlinger og beslutninger påvirker hele beslutningssystemet.

> Hvis processerne ikke bare kræver lagring af data, men også analyse eller automatisering af dem, skal man begynde at skabe et konceptuelt og visuelt kravlag.

I vores eksempel (fig. 4.4-1) kan hver specialist være en del af et lille team, men også af en større afdeling med op til et dusin eksperter under kontrol af en daglig leder. Hver afdeling bruger en specialiseret applikationsdatabase (fig. 1.2-4 f.eks. ERP, CAD, MEP, CDE, ECM, CPM osv.), som regelmæssigt opdateres med indgående oplysninger, der er nødvendige for at oprette dokumenter, registrere den juridiske status for beslutninger og styre processer.

> Transaktionsprocessen ligner de gamle lederes arbejde for 4.000 år siden, hvor lertavler og papyrus blev brugt til at bekræfte beslutninger på lovlig vis. Forskellen mellem moderne systemer og deres forgængere af ler og papir er, at moderne metoder desuden omfatter processen med at konvertere tekstinformation til digital form til videre automatisk behandling i andre systemer og værktøjer.

At skabe en visualisering af processen i form af konceptuelle flowcharts hjælper med at beskrive hvert trin og samspillet mellem forskellige roller, hvilket gør et komplekst workflow klart og enkelt.

> Visualisering af processer sikrer, at proceslogikken er gennemsigtig og tilgængelig for alle teammedlemmer.

Den samme kommunikative proces for tilføjelse af et vindue til et projekt, som blev beskrevet i form af tekst, meddelelser (figur 4.4-1) og blokdiagram, ligner den konceptuelle model, vi diskuterede i kapitlet om datamodellering (figur 4.4-4).

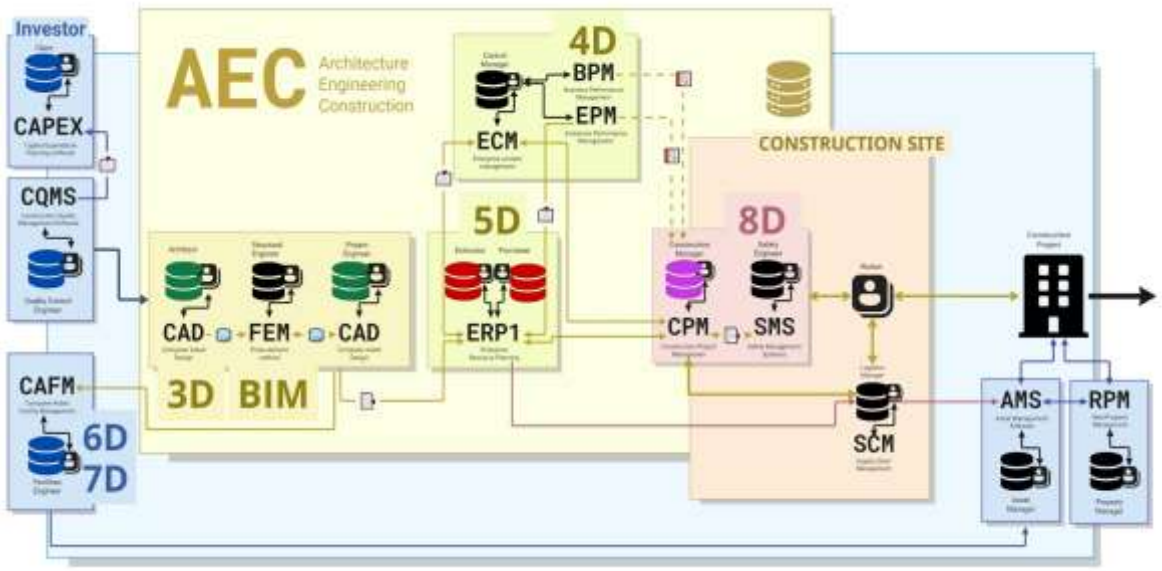


Fig. 4.4-4 Konceptdiagrammet viser projektdeltagerne som brugere af en database, hvor deres forespørgsler forbinder forskellige systemer.

Selv om konceptuelle diagrammer er et vigtigt skridt, begrænser mange virksomheder sig til dette niveau og tror, at et visuelt diagram er tilstrækkeligt til at forstå processerne. Det skaber en illusion af håndterbarhed: Ledere kan lettere se det store billede i et sådant flowdiagram og se forbindelserne mellem deltagere og faser. Men sådanne skemaer giver ikke en klar idé om, hvilke data der er nødvendige for hver deltager, i hvilket format de skal overføres, og hvilke parametre og attributter der er obligatoriske for automatisering. Et konceptuelt flowchart er mere som et rutekort: Det angiver, hvem der interagerer med hvem, men afslører ikke, hvad der overføres i disse interaktioner.

> Selv om en proces er beskrevet i detaljer på konceptuelt niveau ved hjælp af flowcharts, er det ikke nogen garanti for, at den er effektiv. Visualisering forenkler ofte ledernes arbejde, så de lettere kan spore processen med trinvis rapportering. Men for databaseingeniører er den konceptuelle repræsentation måske ikke klar nok og giver måske ikke en klar forståelse af, hvordan processen skal implementeres på parameter- og kravniveau.

Når vi bevæger os mod mere komplekse dataøkosystemer, bliver den indledende implementering af konceptuelle og visuelle værktøjer afgørende for at sikre, at dataprocesserne ikke kun er effektive, men også

i overensstemmelse med organisationens strategiske mål. For fuldt ud at oversætte denne vinduestilføjelsesproces (figur 4.4-1) til datakravniveauet skal vi gå et niveau dybere og oversætte den konceptuelle visualisering af processen til det logiske og fysiske niveau af data, nødvendige attributter og deres grænseværdier.

Strukturerede krav og RegEx regulære udtryk

Op til 80% af de data, der skabes i virksomheder, er ustrukturerede eller semi-strukturerede formater [52] - tekst, dokumenter, breve, PDF -filer, samtaler. Sådanne data (fig. 4.4-1) er vanskelige at analysere, verificere, overføre mellem systemer og bruge i automatisering.

For at sikre håndterbarhed, gennemsigtighed og automatisk validering er det nødvendigt at oversætte tekstuelle og halvstrukturerede krav til veldefinerede, strukturerede formater. Struktureringsprocessen vedrører ikke kun dataene (som vi har diskuteret i detaljer i de første kapitler i denne del af bogen), men også selve kravene, som projektdeltagerne normalt formulerer i fritekstform gennem hele projektets livscyklus, ofte uden at tænke på, at disse processer kan automatiseres.

> Ligesom vi allerede har konverteret data fra en ustruktureret tekstform til en struktureret form, vil vi i krav-workflowet konvertere tekstkrav til et struktureret "logisk og fysisk lag"-format.

Som en del af eksemplet med at tilføje et vindue (figur 4.4-1) er det næste skridt at beskrive datakravene i tabelform. Vi vil strukturere oplysningerne for hvert system, der bruges af projektdeltagerne, ved at angive nøgleattributter og deres grænseværdier

Tag f.eks. et sådant system (fig. 4.4-5) - Construction Quality Management System (CQMS), som bruges af kvalitetskontrolingeniøren på bygherrens side. Ved hjælp af det kontrollerer han, om et nyt element i projektet - i dette tilfælde "det nye vindue" - overholder de fastlagte standarder og krav.

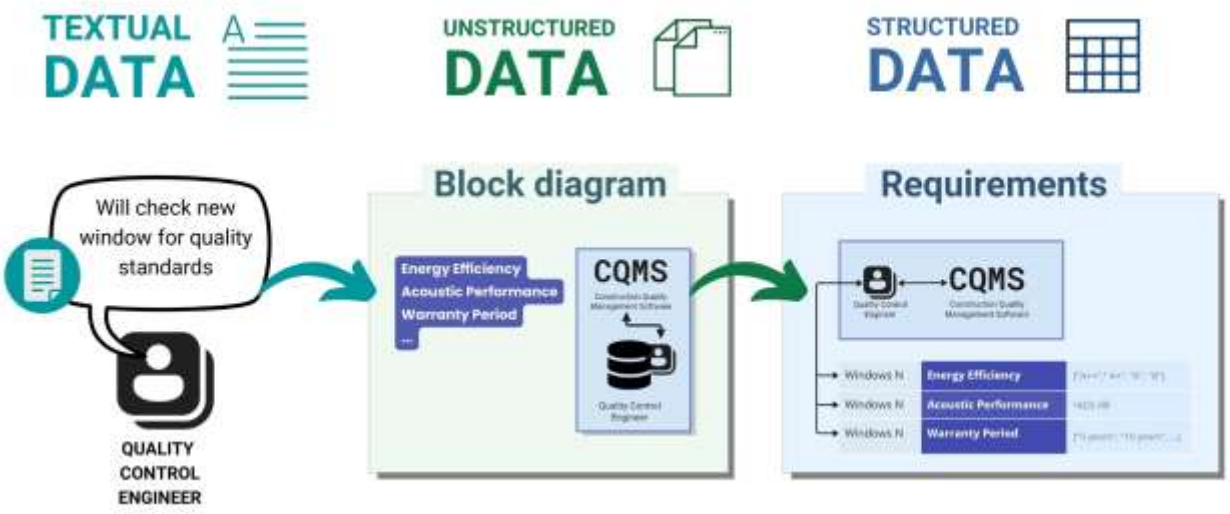


Fig. 4.4-5 Konvertering af tekstuelle krav til et tabelformat med beskrivelser af entitetens attributter gør det lettere for andre specialister at forstå.

Som eksempel kan nævnes nogle vigtige krav til attributter for enheder af typen "vinduessystemer" i CQMS -system (Fig. 4.4-6): energieffektivitet, akustisk ydeevne og garantiperiode. Hver kategori omfatter visse standarder og specifikationer, som skal tages i betragtning ved design og installation af vinduessystemer.



Fig. 4.4-6 Kvalitetskontrolingeniøren skal inspicere nye vinduestypeelementer for energieffektivitet, lydisolering og garantistandarder.

De datakrav, som en kvalitetskontrolingeniør specificerer i form af en tabel, har f.eks. følgende grænseværdier:

- **Vinduernes energieffektivitetsklasse** spænder fra "A++", der angiver den højeste effektivitet, til "B", der betragtes som det mindst acceptable niveau, og disse klasser er repræsenteret ved en liste over acceptable værdier **["A++", "A+", "A", "A", "A", "B"].**
- **Vinduers akustiske isolering**, målt i decibel og som viser deres evne til at reducere gadestøj, er defineret ved **det** almindelige udtryk \d{2}dB.

- **Attributten "Garantiperiode**" for entiteten "Vinduestype" starter ved fem år og indstiller denne periode som det minimum, der er tilladt, når man vælger et produkt; garantiperiodeværdier **som ["5 år", "10 år" osv.]** eller den logiske betingelse **">5 (år)"** er også angivet**.**

I henhold til de indsamlede krav inden for de etablerede attributter vil nye vindueskategorier eller klasseelementer med karakterer under "B", såsom "C" eller "D", ikke bestå energieffektivitetstesten. Vinduesakustisk isolering i data eller dokumenter til QA Engineer skal mærkes med et tocifret tal efterfulgt af postfixet "dB", såsom "35 dB" eller "40 dB", og værdier uden for dette format, såsom "9 D B" eller "100 decibel", vil ikke blive accepteret (da de ikke vil bestå mønsteret for RegEx-strenge). Garantiperioden skal begynde med mindst "5 år", og vinduer med kortere garantiperioder som "3 år" eller "4 år" vil ikke opfylde de krav, som kvalitetsingeniøren har beskrevet i tabelformatet.

For at kontrollere sådanne attributparameterværdier mod grænseværdier fra krav i valideringsprocessen bruger vi enten en liste over tilladte værdier (["A", "B", "C"]), ordbøger (["A": "H1", "H2"; "B": W1", "W2"]), logiske operationer (f.eks. ">", "<", "<=", ">=" "==") for numeriske værdier) og regulære udtryk (for streng- og tekstværdier som f.eks. i attributten "Acoustic Performance"). Regulære udtryk er et ekstremt vigtigt værktøj, når man arbejder med strengværdier.

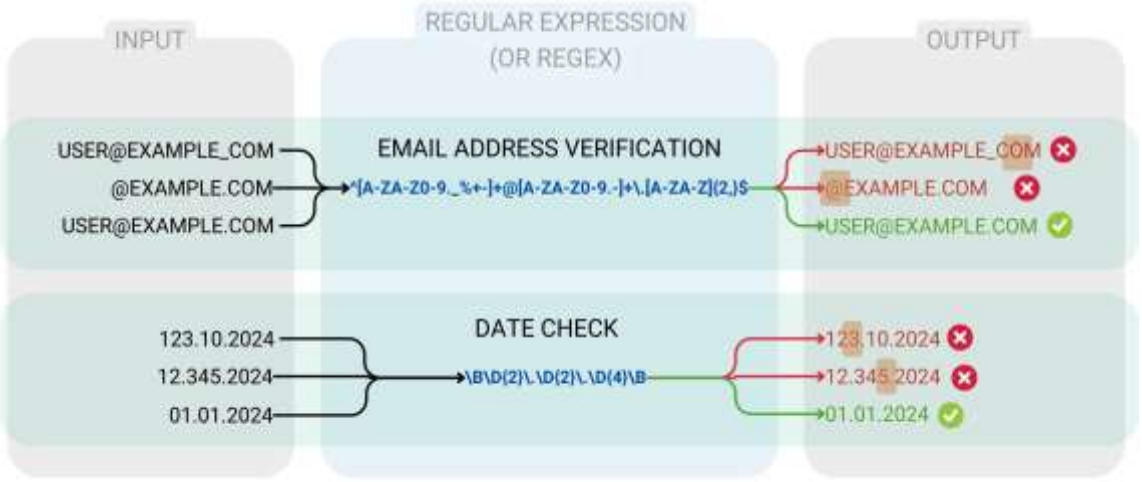
> **Regulære udtryk** (RegEx) bruges i programmeringssprog, herunder Python (Re-biblioteket), til at finde og ændre strenge. Regex er som en detektiv i strengverdenen, der er i stand til at identificere tekstmønstre i tekst med præcision.

I regulære udtryk beskrives bogstaver direkte ved hjælp af de tilsvarende tegn i alfabetet, mens tal kan repræsenteres ved hjælp af specialtegnet \d, som svarer til et hvilket som helst ciffer fra 0 til 9. Firkantede parenteser bruges til at angive en række bogstaver eller cifre, f.eks. [a-z] for ethvert lille bogstav i det latinske alfabet eller [0-9], som svarer til \d. For ikke-numeriske tegn og tegn, der ikke er bogstaver, bruges henholdsvis \D og \W.

Populære RegEx-brugsscenarier (figur 4.4-7):

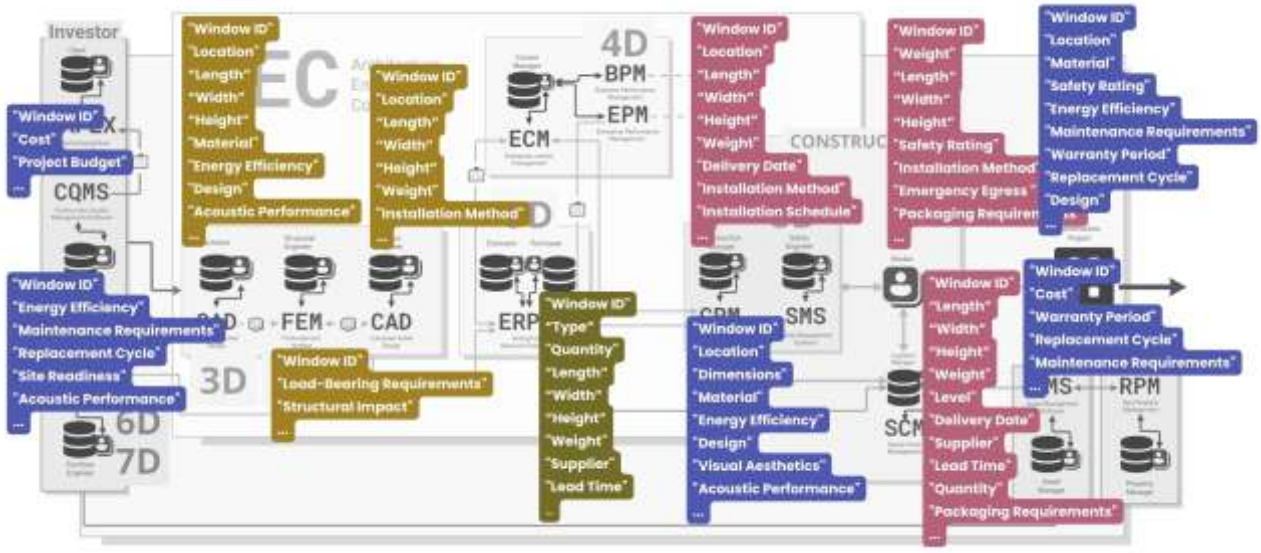
- **Verificering af e-mailadresse**: For at kontrollere, om en streng er en gyldig e-mailadresse, kan du bruge skabelonen "^ **[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$".**
- **Datoudtræk**: "\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}**.\d{4}\b"**-skabelonen kan bruges til at udtrække dato fra tekst i DD.MM.YYYYY-format.
- Verificering af telefonnumre: For at verificere telefonnumre i formatet +49(000)000-0000 vil mønsteret se ud som **"\+\d{2}\(\d{3}\)\d{3}-\d{4}".**

Ved at oversætte kravene fra en QA-ingeniør til formatet for attributter og deres grænseværdier (Fig. 4.4-6) har vi omdannet dem fra deres oprindelige tekstformat (samtaler, breve og forskriftsdokumenter) til en organiseret og struktureret tabel, hvilket gør det muligt automatisk at kontrollere og analysere alle indgående data (f.eks. nye elementer i Window-kategorien). Tilstedeværelsen af krav gør det muligt automatisk at kassere data, der ikke er blevet kontrolleret, og automatisk at overføre de kontrollerede data til systemerne til videre behandling.



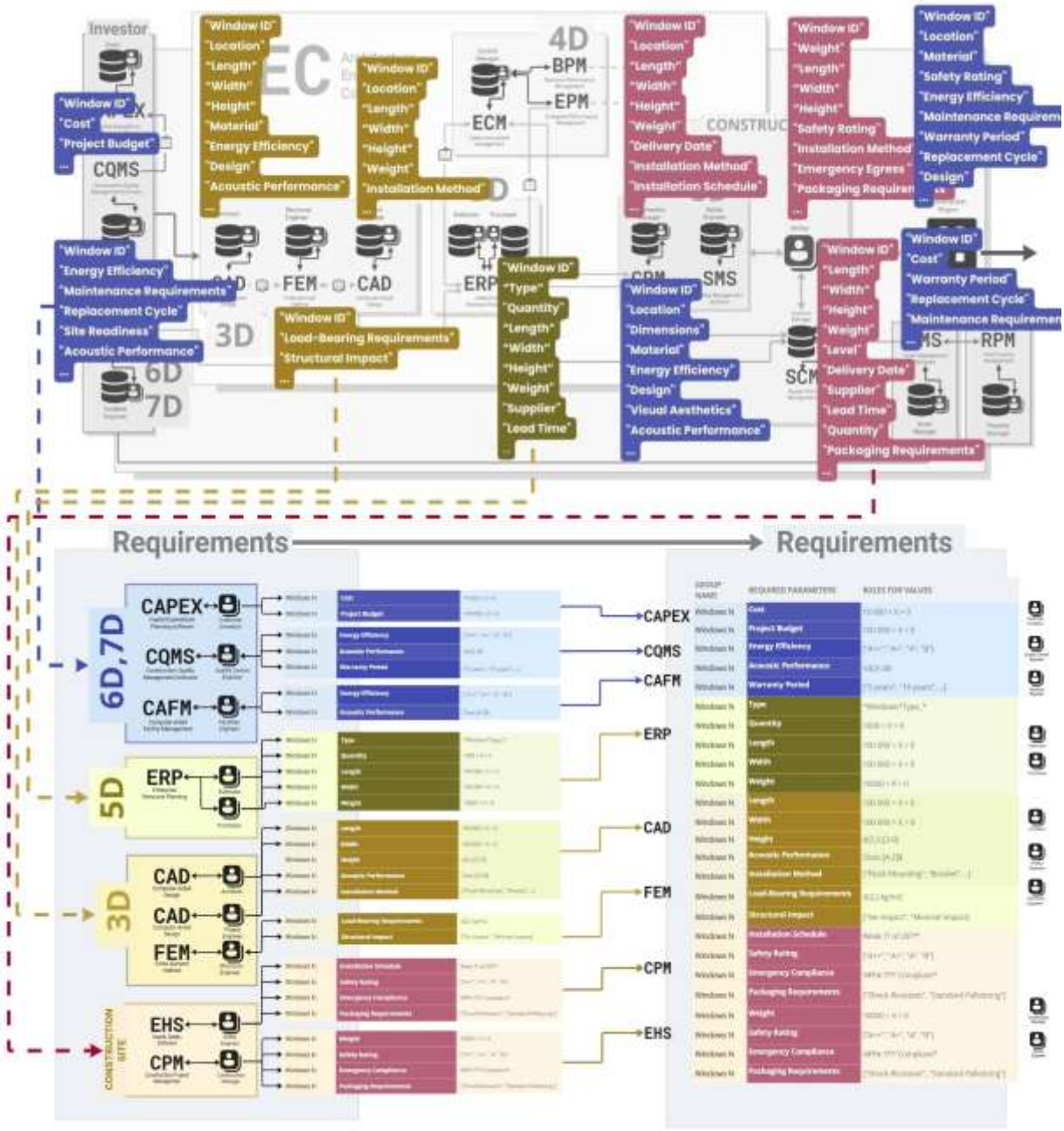
Figur 4.4-7 Brugen af regulære udtryk er et ekstremt vigtigt værktøj i valideringsprocessen for tekstdata.

Når vi nu bevæger os fra det konceptuelle til det logiske niveau i arbejdet med krav, vil vi konvertere alle krav fra alle specialister i vores proces med at installere et nyt vindue (fig. 4.4-4) til en ordnet liste i attributformat og tilføje disse lister med de nødvendige attributter til vores flowdiagram for hver specialist (fig. 4.4-8).



Figur 4.4-8 På det logiske procesniveau tilføjes de attributter, der håndteres af hver specialist, til deres respektive systemer.

Ved at tilføje alle attributter til en fælles procestabel omdanner vi den information, der tidligere blev præsenteret som tekst og dialog på det konceptuelle niveau (fig. 4.4-1), til den strukturerede og systematiske form af tabeller på det fysiske niveau (fig. 4.4-9).



Figur 4.4-9 Konvertering af ustruktureret specialistdialog til strukturerede tabeller hjælper med at forstå kravene til på det fysiske niveau.

Nu skal datakravene kommunikeres til de specialister, der skaber information til specifikke systemer. Hvis du f.eks. arbejder i en CAD-database, skal du, før du begynder at modellere elementer, indsamle alle de nødvendige parametre baseret på slutbrugsscenarierne for dataene. Dette starter normalt med driftsfasen, efterfulgt af byggepladsen, logistikafdelingen, beregningsafdelingen, afdelingen for strukturelle beregninger og så videre. Først når du har taget højde for kravene fra alle disse områder, kan du begynde at skabe data - baseret på de indsamlede parametre. Det giver dig mulighed for at automatisere verificeringen og overførslen af data i kæden i fremtiden.

Når nye data opfylder kravene, integreres de automatisk i virksomhedens dataøkosystem og sendes direkte til de brugere og systemer, som de er beregnet til. Verificering af data i forhold til attributter og deres værdier sikrer, at oplysningerne opfylder de nødvendige kvalitetsstandarder og er klar til brug i virksomhedens scenarier.

Datakravene er defineret, og før verifikationen kan begynde, skal de data, der skal verificeres, oprettes, skaffes eller indsamles, eller den aktuelle status for information i databaser skal registreres for at blive brugt i verifikationsprocessen.

Dataindsamling til verifikationsprocessen

Før valideringen kan begynde, er det vigtigt at sikre, at dataene er tilgængelige i en form, der er egnet til valideringsprocessen. Det betyder ikke bare, at oplysningerne skal være tilgængelige, men at de skal forberedes: Dataene skal indsamles og omdannes fra ustrukturerede, løst strukturerede, tekstuelle og geometriske formater til en struktureret form. Denne proces er beskrevet i detaljer i de foregående kapitler, hvor metoder til at transformere forskellige typer data blev diskuteret. Som et resultat af alle transformationer har de indgående data form af åbne strukturerede tabeller (fig. 4.1-2, fig. 4.1-9, fig. 4.1-13).

Med kravene og de strukturerede tabeller med de nødvendige parametre og grænseværdier (fig. 4.4-9) kan vi begynde at validere dataene - enten som en enkelt automatiseret proces (Pipeline) eller som en trinvis validering af hvert enkelt indgående dokument.

For at starte kontrollen er det nødvendigt enten at modtage en ny fil som input eller at fastsætte dataenes aktuelle tilstand - at oprette et øjebliksbillede eller eksportere aktuelle og indgående data eller at oprette en forbindelse til en ekstern eller intern database. I det aktuelle eksempel skabes et sådant øjebliksbillede ved automatisk at konvertere CAD-data fra til et struktureret format, der registreres f.eks. kl. 23:00:00 fredag den 29. marts 2024, efter at alle designerne er gået hjem.



Figur 4.4-10 CAD-databasesnapshot (BIM), der viser de aktuelle attributoplysninger for en ny enhed af klassen "Vindue" i den aktuelle version af projektmodellen.

Takket være de reverse engineering-værktøjer, der omtales i kapitlet "Oversættelse af CAD-data (BIM) til en struktureret form", kan disse oplysninger fra forskellige CAD (BIM)-værktøjer og editorer organiseres i separate tabeller (fig. 4.4-11) eller kombineres i en fælles tabel, der kombinerer forskellige dele af projektet

(fig. 9.1-10).

En sådan tabel - database viser unikke identifikatorer for vinduer og døre (ID-attribut), typenavne (TypeName), dimensioner (Bredde, Længde), materialer (Materiale) samt indikatorer for energimæssig og akustisk ydeevne og andre egenskaber. En sådan tabel udfyldt i CAD-programmet (BIM) indsamles af en designingeniør fra forskellige afdelinger og dokumenter og danner en informationsmodel for projektet.

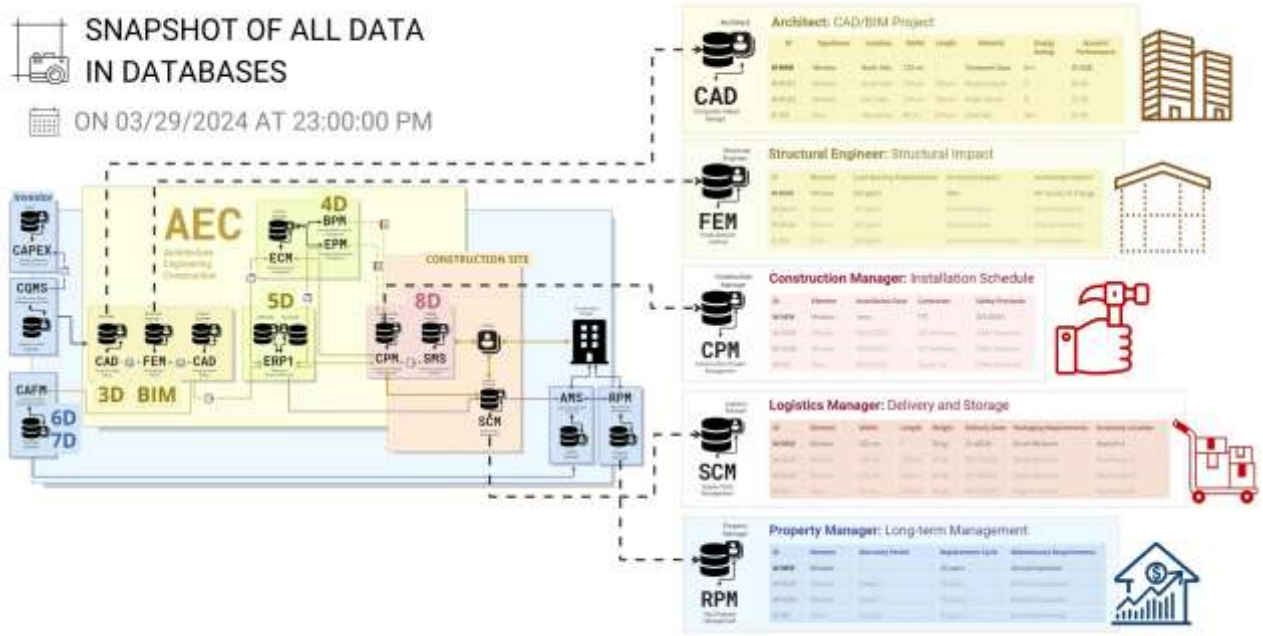


Figur 4.4-11 Strukturerede data fra CAD systemer kan være en todimensionel tabel med kolonner, der angiver elementernes attributter.

Rigtige CAD (BIM) projekter omfatter titusinder eller hundredtusinder af elementer (fig. 9.1-10). Elementer i CAD-formater kategoriseres automatisk efter type og kategori - fra vinduer og døre til plader og vægge. Unikke identifikatorer (f.eks. native ID, som indstilles automatisk af CAD-løsningen) eller typeattributter (typenavn, type, familie) gør det muligt at spore det samme objekt i forskellige systemer. For eksempel kan et nyt vindue på nordvæggen i en bygning identificeres entydigt med en enkelt identifikator "W-NEW" i alle relevante systemer i organisationen.

> Mens entitetsnavne og -identifikatorer bør være ensartede på tværs af alle systemer, kan de attributter og værdier, der er forbundet med disse entiteter, variere betydeligt afhængigt af brugskonteksten. Arkitekter, bygningsingeniører, fagfolk inden for byggeri, logistik og ejendomsvedligeholdelse opfatter alle de samme elementer på forskellige måder. De har hver især deres egne klassifikatorer, standarder og mål: Nogle ser på et vindue ud fra et rent æstetisk synspunkt og vurderer dets form og proportioner, mens andre ser på det ud fra et teknisk eller driftsmæssigt synspunkt og analyserer varmeledningsevne, installationsmetode, vægt eller vedligeholdelseskrav. Når man modellerer data og beskriver elementer, er det derfor vigtigt at tage højde for deres alsidige brug og sikre konsistens i data, samtidig med at man tager højde for branchespecifikke forhold.

For hver rolle i virksomhedens processer er der specialiserede databaser med deres egen brugergrænseflade - fra design og beregninger til logistik, installation og bygningsdrift (fig. 4.4-12). Hvert af disse systemer styres af et professionelt team af specialister via en særlig brugergrænseflade eller via databaseforespørgsler, og bag summen af alle beslutninger, der træffes om de indtastede værdier i slutningen af kæden, står systemlederen eller afdelingslederen, som er ansvarlig for de indtastede datas juridiske gyldighed og kvalitet over for sine modparter, der betjener andre systemer.



Figur 4.4-12 Den samme enhed har den samme identifikator i forskellige systemer, men forskellige attributter, som kun er vigtige i det pågældende system.

Når vi har organiseret indsamlingen af strukturerede krav og data på det logiske og fysiske niveau, er det op til os at etablere en proces, der automatisk kontrollerer data fra forskellige indgående dokumenter og forskellige systemer i forhold til de tidligere indsamlede krav.

Verifikation af data og resultater af verifikation

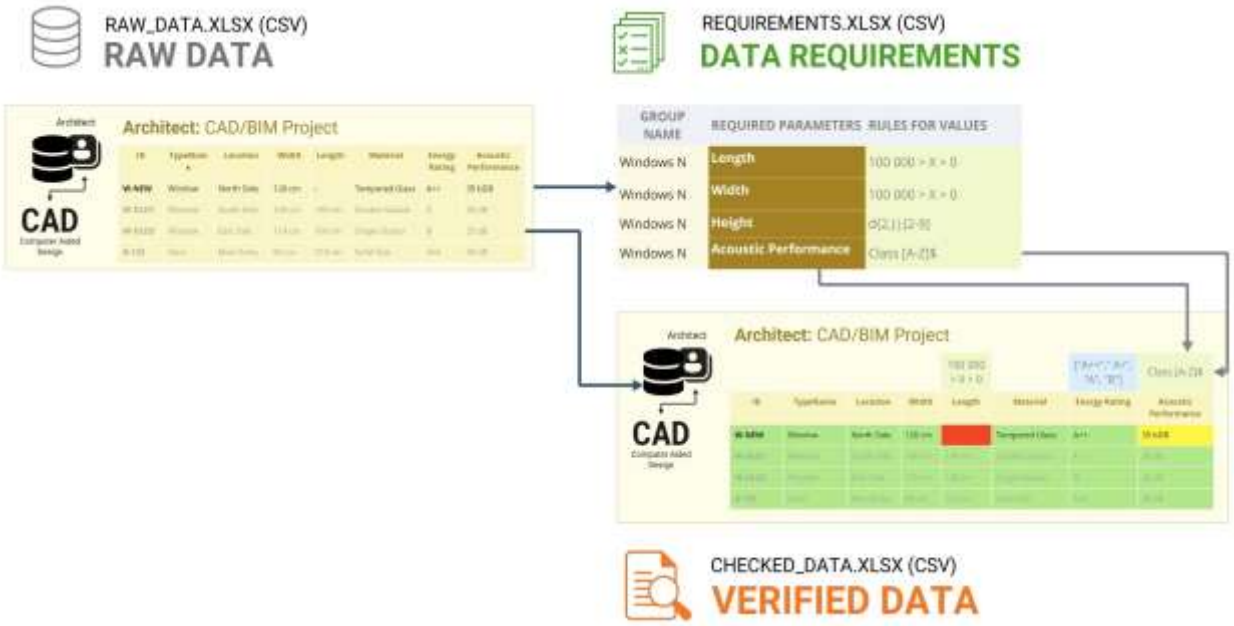
Alle nye data, der kommer ind i systemet - hvad enten det er dokumenter, tabeller eller databaseposter fra bygherren, arkitekten, ingeniøren, formanden, logistikeren eller ejendomsadministratoren - skal valideres i forhold til de krav, der er formuleret tidligere (fig. 4.4-9). Valideringsprocessen er kritisk: Enhver fejl i dataene kan føre til forkerte beregninger, forsinkelser i tidsplanen og endda økonomiske tab. For at minimere sådanne risici bør der tilrettelægges en systematisk og gentagelig, iterativ datavalideringsprocedure.

> For at validere nye data, der kommer ind i systemet - ustrukturerede, tekstuelle eller geometriske - skal de konverteres til et løst struktureret eller struktureret format. Valideringsprocessen skal derefter kontrollere dataene i forhold til en komplet liste over påkrævede attributter og deres tilladte værdier.

Konvertering af forskellige typer data: tekst, billeder, PDF -dokumenter og blandede CAD (BIM) data til en struktureret form blev diskuteret i detaljer i kapitlet "Konvertering af data til en struktureret form".

Et eksempel er en tabel hentet fra et CAD (BIM) projekt (Fig. 4.4-11). Den indeholder semistrukturerede geometriske data og strukturerede attributoplysninger om projektenheder (fig. 3.1-14) - f.eks. et element fra klassen "Windows".

For at udføre valideringen sammenligner vi attributværdierne (fig. 4.4-11) med de referencegrænseværdier, der er defineret af eksperter i form af et krav (fig. 4.4-9). Den endelige sammenligningstabel (fig. 4.4-13) vil gøre det klart, hvilke værdier der er acceptable, og hvilke der skal korrigeres, før dataene kan bruges uden for CAD-applikationer (BIM).



Figur 4.4-13 Den endelige valideringstabel fremhæver de attributværdier for den nye entitet af klassen "Windows", som du skal være opmærksom på.

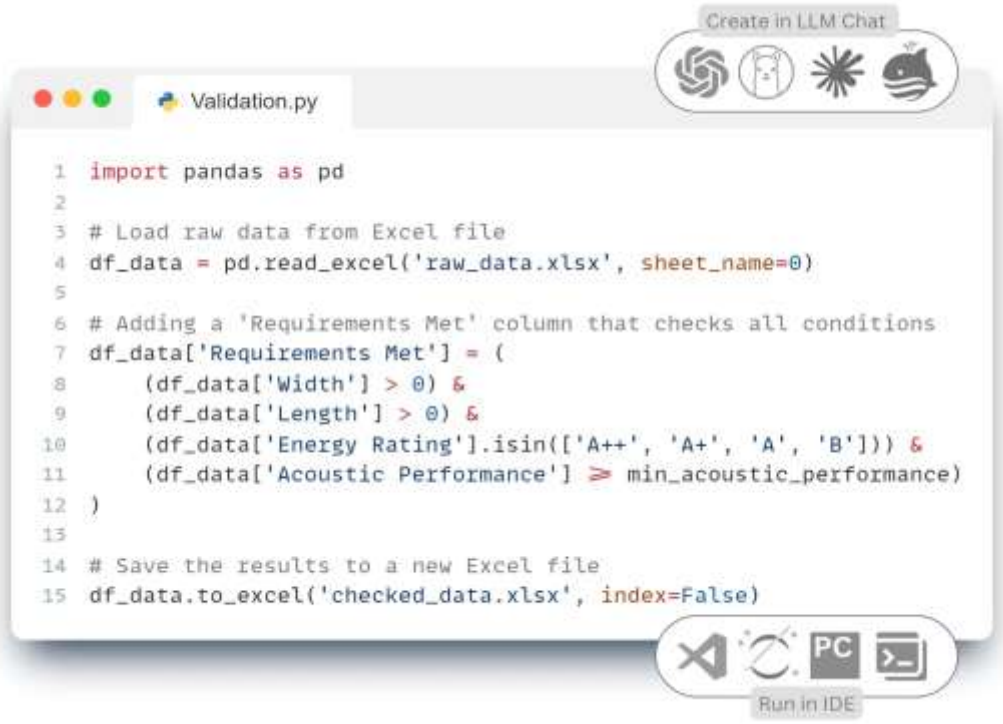
Ved at implementere en lignende løsning ved hjælp af Pandas-biblioteket, som vi beskrev tidligere i kapitlet "Pandas: Et uundværligt værktøj til dataanalyse", vil vi validere data fra en tabelfil, der er hentet fra en CAD-fil (BIM) (RVT, IFC, DWG, NWS, DGN) (figur 4.4-11) ved hjælp af kravene fra en anden tabelfil med krav (figur 4.4-9).

For at få koden skal vi i prompten til LLM beskrive, at vi skal indlæse dataene fra filen **raw_data.xlsx** (et komplet sæt data fra CAD-databasen (BIM))**, kontrollere** dem og gemme resultatet i en ny fil **checked_data.xlsx (**fig. 4.4-13).

- Lad os få koden ved hjælp af LLM uden at nævne Pandas-biblioteket:

> Skriv kode til at kontrollere tabellen fra filen raw_data.xlsx og validere dem ved hjælp af følgende valideringsregler: værdierne i kolonnerne 'Width' og 'Length' er større end nul, 'Energy Rating' er inkluderet i listen ['A++', 'A+', 'A', 'B'] og 'Acoustic Performance' som en variabel, som vi vil specificere senere - med den sidste valideringskolonne tilføjet, og gem den endelige tabel i en ny Excel-fil checked_data.xlsx ⏎

- LLM's svar vil beskrive et kort eksempel på Python-kode, som kan forfines og udvides ved hjælp af efterfølgende forespørgsler :



Figur 4.4-14 Den kode, der genereres af LLM- modellen, kontrollerer det konverterede CAD (BIM) design i forhold til attributkravene i form af grænseværdier.

Den kode, der genereres af LLM-sprogmodellen, kan bruges i ethvert populært IDE eller onlineværktøj: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære onlineværktøjer Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Udførelse af koden (fig. 4.4-14) vil vise, at "enhedselementerne" W-OLD1, W-OLD2, D-122 (og andre - elementer) fra CAD-databasen (BIM) opfylder attributkravene: bredde og længde er større end nul, og energieffektivitetsklassen er en af listeværdierne "A++", "A", "B", "C" (fig. 4.4-15).

W-NEW-elementet, som vi har brug for og for nylig har tilføjet, og som er ansvarlig for den nye elementklasse 'Vindue' på nordsiden, er ikke i overensstemmelse (attributten 'Krav opfyldt'), fordi dets længde er nul (en værdi på '0,0' betragtes som uacceptabel af vores 'Bredde'>0-regel), og det angiver ikke en energieffektivitetsklasse.



	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

Figur 4.4-15 Checking identificerer enheder, der ikke har gennemgået verifikationsprocessen, og tilføjer en ny attribut med værdierne 'False' eller 'True' til resultaterne.

På samme måde kontrollerer vi konsistensen af alle projektelementer (entiteter) og de nødvendige attributter for hvert af systemerne, tabellerne eller databaserne i alle de data, vi modtager fra forskellige specialister (figur 4.4-1) i løbet af processen med at tilføje et vindue til projektet.

I den endelige tabel er det praktisk at fremhæve resultaterne af kontrollen med farver til visualisering: grøn bruges til at markere de attributter, der har bestået kontrollen, gul - værdier med ikke-kritiske afvigelser, og rød - kritiske uoverensstemmelser (Fig. 4.4-16).

Som et resultat af valideringen (figur 4.4-16) får vi en liste over pålidelige og validerede elementer med deres identifikatorer, som er blevet verificeret til at opfylde attributkravene. Validerede elementer giver sikkerhed for, at disse elementer opfylder de angivne standarder og specifikationer for alle systemer, der er involveret i processen med at tilføje elementer i Window-klassen eller en hvilken som helst anden klasse (vi vil diskutere mere om at automatisere datavalidering og skabe en automatiseret ETL-proces i kapitlet "Automatisering af ETL og datavalidering").

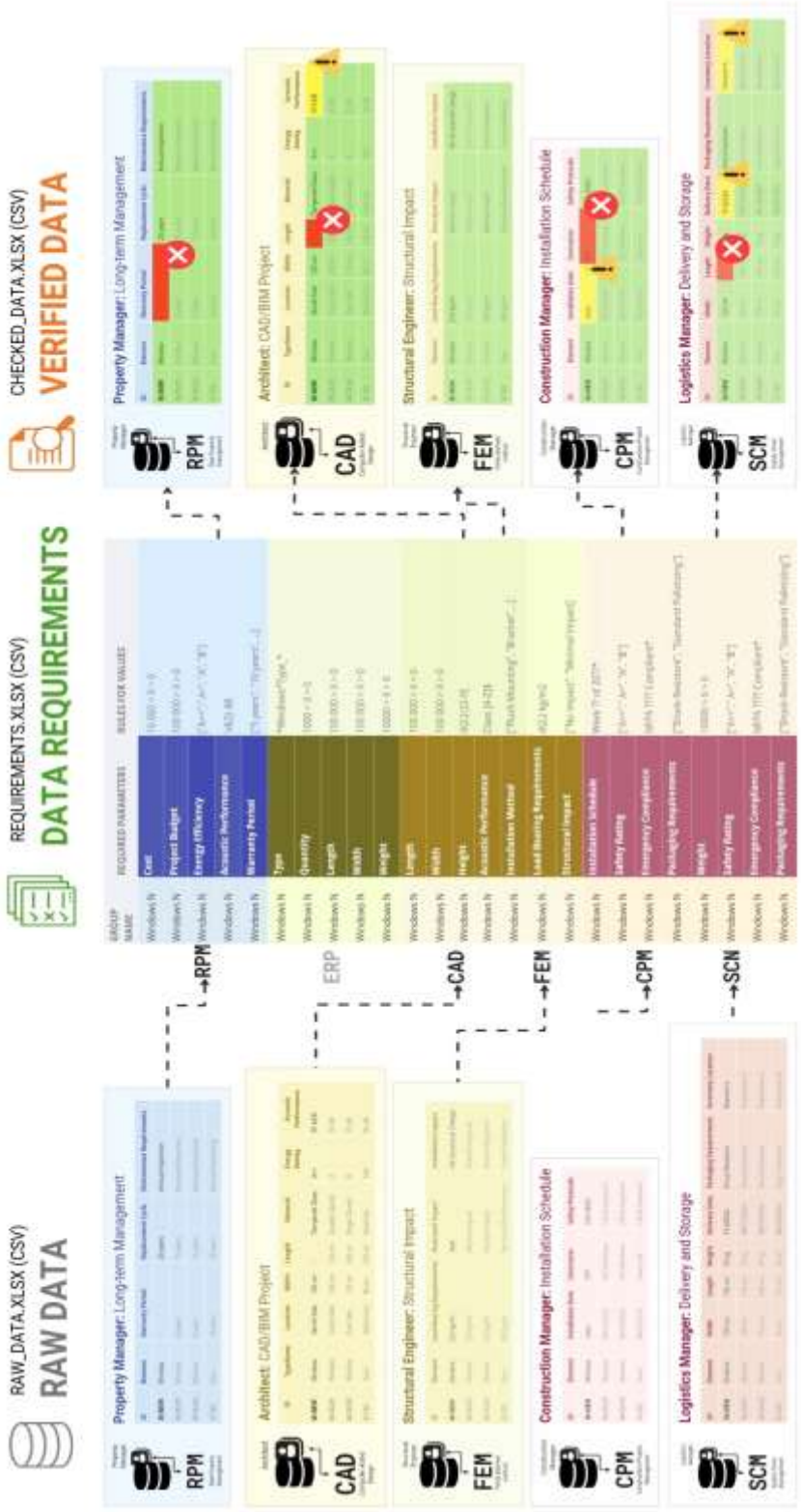


Fig. 4.4-16 Resultatet af den kontrol, der udføres for alle systemer, gør det muligt at afgøre, hvilke data der ikke opfylder virksomhedens krav.

Enheder, der er blevet valideret med succes, kræver normalt ikke megen opmærksomhed. De går videre til de næste stadier af behandling og integration i andre systemer uden forhindringer. I modsætning til "kvalitets"-elementer er det de elementer, der ikke bliver valideret, der er af størst interesse. Oplysninger om sådanne afvigelser er afgørende: De bør ikke kun formidles i tabelrapporter, men også ved hjælp af forskellige visualiseringsværktøjer. En grafisk fremstilling af verifikationsresultaterne hjælper med hurtigt at vurdere den overordnede tilstand af datakvaliteten, identificere problemområder og straks træffe korrigerende eller afhjælpende foranstaltninger.

Visualisering af verifikationsresultater

Visualisering er et vigtigt værktøj til fortolkning af inspektionsresultater. Ud over de sædvanlige oversigtstabeller kan det omfatte informationspaneler, diagrammer og automatisk genererede PDF dokumenter, der grupperer projektelementer efter deres inspektionsstatus. Her kan farvekodning spille en støttende rolle: Grøn kan angive elementer, der er blevet valideret med succes, gul kan angive elementer, der kræver yderligere opmærksomhed, og rød kan angive elementer, der har kritiske fejl eller mangler nøgledata.

I vores eksempel (fig. 4.4-1) analyserer vi data fra hvert system trin for trin: fra CAD (BIM) og ejendomsadministration til logistik og installationsplaner (fig. 4.4-16). Efter revisionen genereres der automatisk individuelle advarsler eller rapportdokumenter til hver specialist, f.eks. i PDF-format (fig. 4.4-17). Hvis dataene er korrekte, modtager specialisten en kort besked: "Tak for samarbejdet". Hvis der findes uoverensstemmelser, sendes en detaljeret rapport med følgende ordlyd: "Dette dokument viser elementer, deres identifikatorer, attributter og værdier, som ikke er blevet kontrolleret for overensstemmelse".



Figur 4.4-17 Validering og automatisk rapportgenerering fremskynder processen med at finde og forstå datamangler for den professionelle, der skaber dataene.

Takket være den automatiserede valideringsproces - så snart der opdages en fejl eller et datahul, sendes der en øjeblikkelig meddelelse i form af en chatbesked, e-mail eller PDF -dokument til den person, der er ansvarlig for at oprette eller behandle de relevante enheder og deres attributter (fig. 4.4-18), med en liste over elementer og attributbeskrivelser, der ikke er blevet valideret.



Fig. 4.4-18 Automatiske inspektionsrapporter gør det lettere at forstå fejl og fremskynder arbejdet med at udfylde projektdata.

Hvis et dokument f.eks. ankommer til ejendomsadministrationssystemet (efter strukturering) med attributten "Garantiperiode" forkert udfyldt, modtager ejendomsadministratoren en advarsel med en liste over attributter, der skal kontrolleres og rettes.

På samme måde vil eventuelle mangler i installationsplanen eller logistikdata resultere i, at der automatisk genereres en rapport, og at der f.eks. sendes en chatbesked eller en e-mail med resultaterne af inspektionen til den relevante specialist.

Ud over PDF -dokumenter og grafer med resultater er det muligt at oprette dashboards og interaktive 3D -modeller (figur 7.1-6, figur 7.2-12), der fremhæver elementer med manglende attributter, så brugerne visuelt kan bruge 3D-geometrier af elementer til at filtrere og evaluere kvaliteten og fuldstændigheden af disse elementer i projektet.

> Visualisering af inspektionsresultater i form af automatisk genererede dokumenter, grafer eller dashboards forenkler i høj grad fortolkningen af data og letter effektiv kommunikation mellem projektdeltagerne.

Processen med automatisk at kontrollere data fra forskellige systemer og informationskilder kan sammenlignes med informeret beslutningstagning i hverdagen. Ligesom virksomheder i byggebranchen overvejer mange variabler - fra pålideligheden af inputdata til deres indvirkning på timingen, omkostningerne og kvaliteten af projektets gennemførelse - afvejer en person også en lang række faktorer, når han træffer vigtige beslutninger, f.eks. når han vælger et sted at bo: transporttilgængelighed, infrastruktur, omkostninger, sikkerhed, livskvalitet. Alle disse overvejelser udgør et system af kriterier, som danner grundlag for de endelige beslutninger, der udgør vores liv.

Sammenligning af datakvalitetstjek med menneskelige livsbehov

På trods af den konstante udvikling af metoder og værktøjer til datakvalitetskontrol forbliver det grundlæggende princip om informationsoverensstemmelse uændret. Dette princip er indbygget i fundamentet for et modent ledelsessystem, uanset om det er i erhvervslivet eller i hverdagen.

> Processen med iterativ datavalidering minder meget om den beslutningsproces, som alle står over for på daglig basis. I begge tilfælde er vi afhængige af erfaring, data og nye oplysninger, efterhånden som de bliver tilgængelige. Og flere og flere livs- og professionelle beslutninger - fra strategiske til dagligdags beslutninger - træffes på baggrund af data.

Når vi f.eks. vælger et sted at bo eller en livspartner, danner vi intuitivt en tabel med kriterier og egenskaber, som vi sammenligner alternativerne med (Fig. 4.4-19). Disse egenskaber - hvad enten det er personlige kvaliteter hos en person eller parametre for en ejendom - repræsenterer attributter, der påvirker den endelige beslutning.

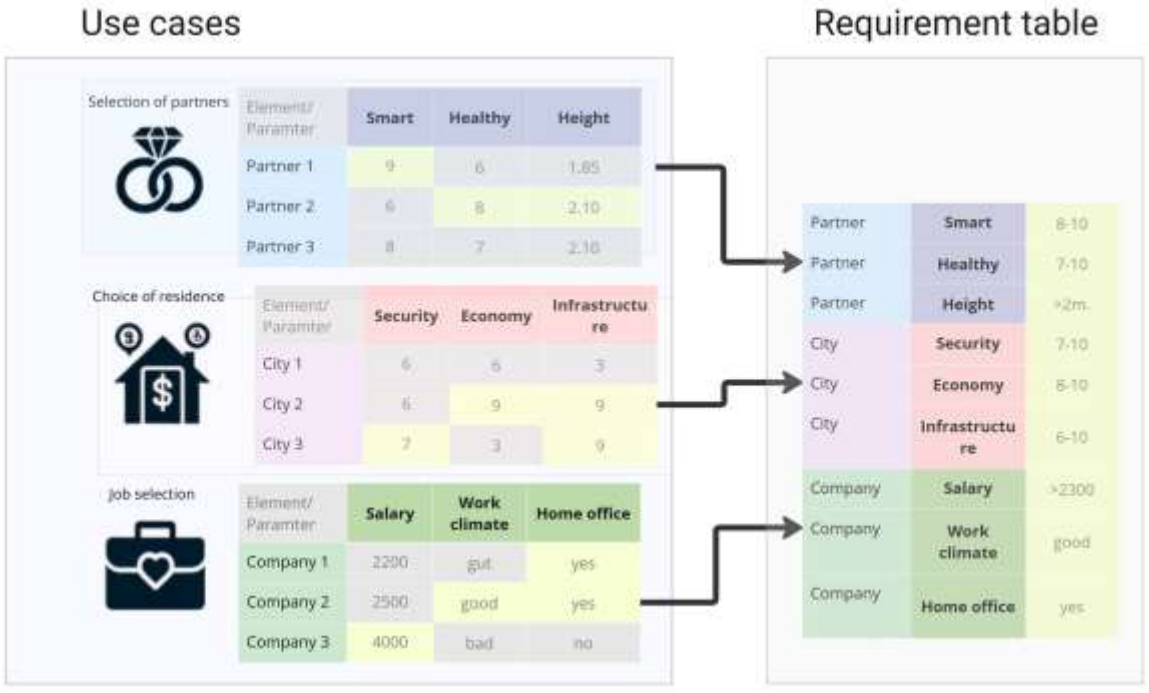


Fig. 4.4-19 Valget af bopæl, arbejde eller partnerskab er baseret på individuelle krav til egenskaber.

Brugen af strukturerede data og en formaliseret tilgang til at beskrive krav (figur 4.4-20) bidrager til mere informerede og velbegrundede valg i både arbejdslivet og privatlivet.

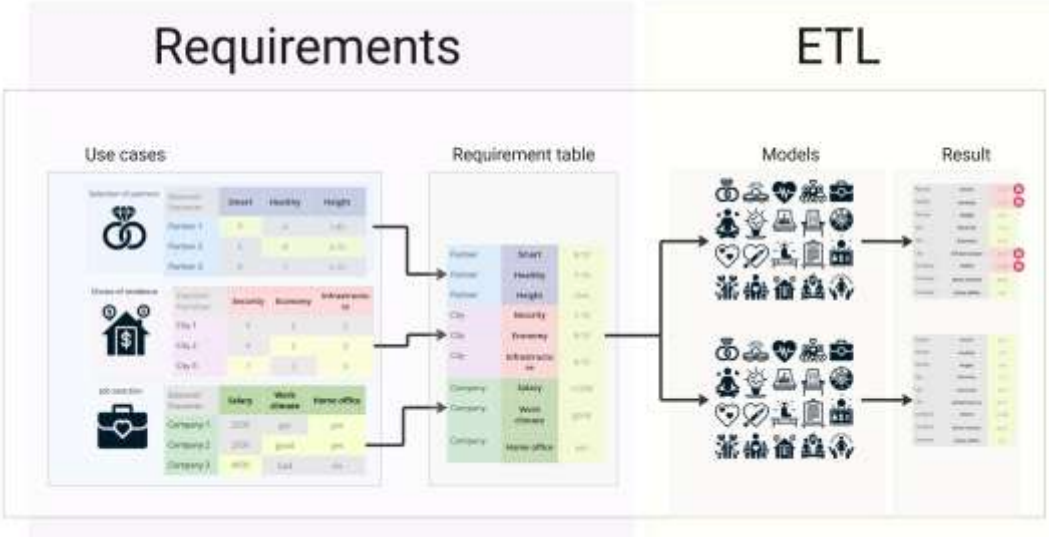


Fig. 4.4-20 Formaliseringen af krav gør det muligt at systematisere opfattelsen af livs- og forretningsbeslutninger.

Den datadrevne tilgang til beslutningstagning er ikke udelukkende et forretningsværktøj. Den er også problemfrit integreret i hverdagen og følger almindelige databehandlingstrin (fig. 4.4-21), der ligner ETL-processen (Extract, Transform, Load), som vi allerede har diskuteret i begyndelsen af denne del i forbindelse med strukturering af data, og som vi vil undersøge i detaljer i forbindelse med automatisering af opgaver i bogens syvende del:

- **Data som fundament (uddrag**): Inden for alle områder - uanset om det er arbejde eller privatliv - indsamler vi information. I erhvervslivet kan det være rapporter, tal, markedsdata; i privatlivet kan det være personlige erfaringer, råd fra ens nærmeste, feedback, observationer.
- **Evalueringskriterier (Transform): De** indsamlede oplysninger fortolkes på grundlag af foruddefinerede kriterier. På arbejdspladsen er det præstationsindikatorer (KPI'er), budgetbegrænsninger og normer; i privatlivet er det parametre som pris, bekvemmelighed, pålidelighed, karisma osv.
- **Forudsigelse og risikoanalyse (Load):** Den sidste fase involverer beslutningstagning baseret på analyse af de transformerede data og sammenligning af mulige konsekvenser. Dette svarer til forretningsprocesser, hvor data går gennem et forretningslogik- og risikofilter.

> De beslutninger, vi træffer - fra trivielle præferencer som hvad vi skal spise til morgenmad til store livsbegivenheder som valg af karriere eller livspartner - er i sagens natur et resultat af behandling og evaluering af data.

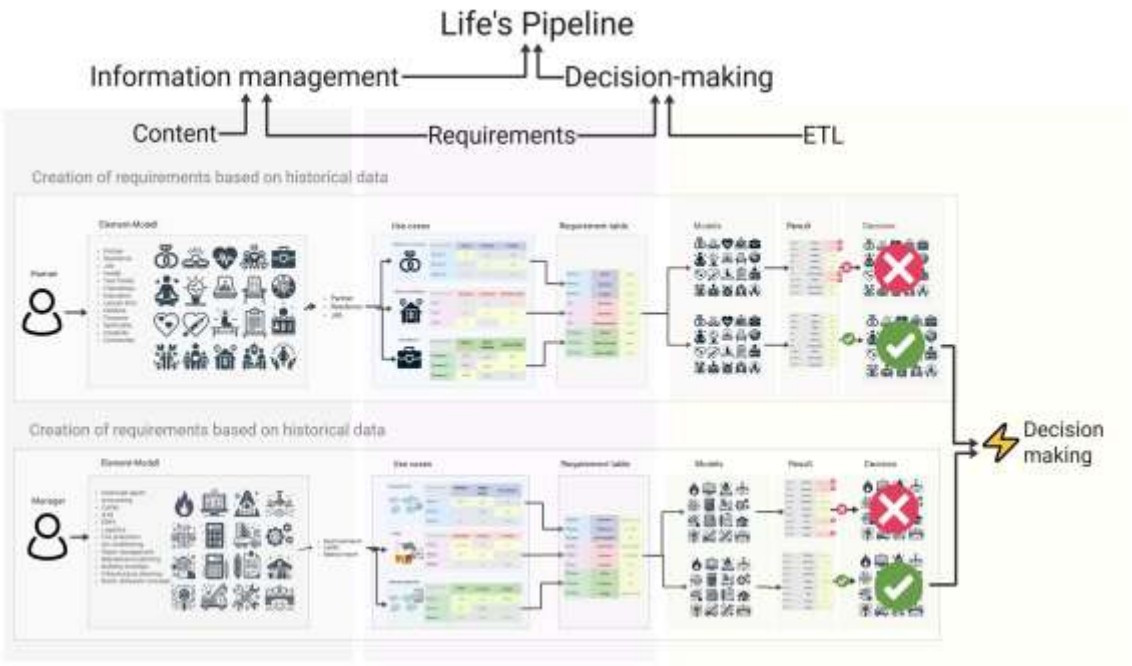


Fig. 4.4-21 Erhvervslivet og livet generelt er en række datadrevne beslutninger, hvor kvaliteten af de data, der bruges til at træffe beslutninger, er en nøglefaktor.

Alt i vores liv hænger sammen, og ligesom levende organismer, herunder mennesker, følger naturens love, udvikler sig og tilpasser sig skiftende forhold, så afspejler menneskelige processer, herunder hvordan vi indsamler og analyserer data, disse naturlige principper. Det tætte forhold mellem natur og menneskelig aktivitet bekræfter ikke kun vores afhængighed af naturen, men også vores ønske om at anvende love, der er finpudset gennem millioner af års evolution, til at skabe dataarkitekturer, processer og systemer til beslutningstagning.

Nye teknologier, især inden for byggeri, er et godt eksempel på, hvordan menneskeheden gang på gang lader sig inspirere af naturen til at skabe bedre, mere bæredygtige og effektive løsninger.

Næste skridt: Omdan data til præcise beregninger og planer

I denne del så vi på, hvordan man konverterer ustrukturerede data til et struktureret format, udvikler datamodeller og organiserer processer til kontrol af informationskvaliteten i byggeprojekter. Datastyring, standardisering og klassificering er en grundlæggende proces, som kræver en systematisk tilgang og en klar forståelse af forretningskravene. De teknikker og værktøjer, der diskuteres i denne del, giver mulighed for pålidelig integration mellem forskellige systemer gennem hele et objekts livscyklus.

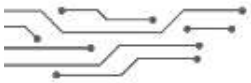
For at opsummere denne del, lad os fremhæve de vigtigste praktiske trin, der vil hjælpe dig med at anvende de diskuterede tilgange i dine daglige opgaver:

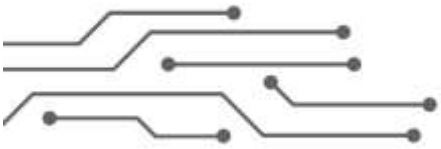
- Start med at systematisere kravene
  - Opret et register med attributter og parametre for nøgleelementer i dine projekter og processer
  - Dokumentér grænseværdierne for hver attribut

- [ ] Visualisere processer og relationer mellem klasser, systemer og attributter ved hjælp af flowdiagrammer (f.eks. i Miro, Canva, Visio).

- Automatiser datakonvertering
  - [ ] Tjek, hvilke af dine dokumenter, der ofte bruges i processer, der kan digitaliseres ved hjælp af OCR-biblioteker og konverteres til tabelform.
  - [ ] Tjek reverse engineering-værktøjer til at udtrække data fra CAD (BIM)
  - [ ] Prøv at opsætte automatisk datahentning fra dokumenter eller formater, du ofte bruger i dit arbejde, til en tabelformular
  - [ ] Opsæt automatiske konverteringer mellem forskellige dataformater
- Opret en vidensbase til kategorisering
  - [ ] Udvikl en intern eller brug en eksisterende elementklassifikator, der er tilpasset industristandarder.
  - [ ] Dokumentér sammenhængen mellem forskellige klassifikationssystemer
  - [ ] Diskuter med dit team emnet om at bruge et samlet system til identifikation og entydig klassificering af elementer.
  - [ ] Begynd at opbygge en proces til automatisk validering af data - både dem, du arbejder med i teamet, og dem, der sendes videre til eksterne systemer.

Ved at bruge disse tilgange kan du forbedre kvaliteten af dine data betydeligt og forenkle den efterfølgende behandling og transformation. I de følgende dele af denne bog vil vi se på, hvordan man kan anvende allerede strukturerede og forberedte data til automatiserede beregninger, omkostningsestimering, planlægning og styring af byggeprojekter.





V DEL

OMKOSTNINGS- OG TIDSBEREGNINGER: INDARBEJDELSE AF DATA I BYGGEPROCESSER

Den femte del er helliget praktiske aspekter af at bruge data til at optimere omkostningsberegning og planlægning af byggeprojekter. Den ressourcebaserede metode til omkostningsestimering og automatisering af estimeringsprocesser analyseres i detaljer. Metoder til automatiseret indhentning af Quantity Take-Off) fra CAD (BIM)-modeller og deres integration med beregningssystemer overvejes. 4D og 5D modelleringsteknologier til tidsplanlægning og styring af byggeomkostninger udforskes med specifikke eksempler på deres anvendelse. Der præsenteres en analyse af de udvidede informationslag 6D -8D, som giver en integreret tilgang til vurdering af bæredygtighed, drift og sikkerhed i forbindelse med fast ejendom. Metoderne til beregning af carbon footprint og ESG -indikatorer for byggeprojekter i forbindelse med moderne miljøkrav og -standarder diskuteres i detaljer. Mulighederne og begrænsningerne i de traditionelle ERP og PMIS systemer til styring af byggeprocessen vurderes kritisk med en analyse af deres indvirkning på prisgennemsigtigheden. Der forudsiges udsigter for overgangen fra lukkede løsninger til åbne standarder og fleksible dataanalyseværktøjer, der kan sikre større effektivitet i byggeprocesserne.

KAPITEL 5.1.
OMKOSTNINGSBEREGNINGER OG -OVERSLAG FOR BYGGEPROJEKTER

Grundlæggende om byggeri: estimering af mængde, pris og tid

Blandt de mange forretningsprocesser, der er afgørende for en virksomheds bæredygtighed i byggebranchen, er processerne til nøjagtig estimering af antallet af elementer, projektomkostninger og leveringstid særligt vigtige - ligesom de var det for tusinder af år siden (fig. 5.1-1).

Udviklingen af skriften var resultatet af et kompleks af faktorer, herunder behovet for at registrere økonomiske transaktioner, handel og ressourceforvaltning i tidlige samfund. De første juridisk betydningsfulde dokumenter, lertavler med beregninger af materialeomkostninger og arbejdsløn, blev brugt i forbindelse med handel og byggeri. Disse tavler registrerede parternes forpligtelser i forbindelse med opførelsen af bygninger og blev opbevaret som bevis på aftaler og penge- og handelsrelationer.

I årtusinder har tilgangen til estimering været stort set uændret: Beregninger blev udført manuelt og var afhængige af den estimerende ingeniørs erfaring og intuition. Men med fremkomsten af modulære ERP-systemer og CAD-værktøjer begyndte den traditionelle tilgang til estimering af mængder, omkostninger og tid hurtigt at ændre sig. Nutidens digitale teknologier gør det muligt at fuldautomatisere de vigtigste beregninger af tid og omkostninger, hvilket giver større nøjagtighed, hastighed og gennemsigtighed i ressourceplanlægningen af byggeprojekter.

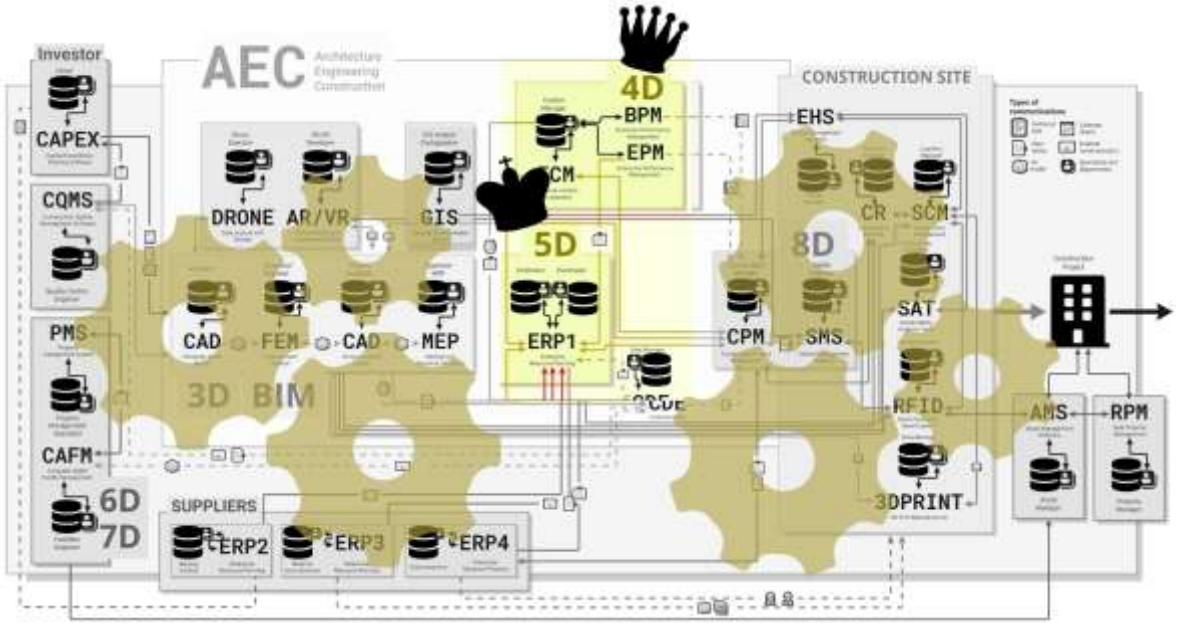


Fig. 5.1-1 Af de mange forskellige systemer er de værktøjer, der er ansvarlige for volumen-, omkostnings- og tidsmålinger, de vigtigste i erhvervslivet.

Byggevirksomheder har fokus på nøjagtige tids- og omkostningsdata. Disse afhænger igen af mængden af anvendte materialer og arbejdskraft, og deres gennemsigtighed påvirker rentabiliteten. Men kompleksiteten i beregningsprocesserne og deres manglende gennemsigtighed fører ofte til højere projektomkostninger, overskredne deadlines og endda konkurs.

> Ifølge KPMG-rapporten "Familiar Problems - New Approaches" (2023) bliver kun 50% af byggeprojekterne afsluttet til tiden, og 87% af virksomhederne rapporterer om øget kontrol med økonomien i kapitalprojekter. De største problemer er relateret til manglen på kvalificeret personale og vanskeligheden ved at forudsige risici [2].

Historiske omkostnings- og procestidsdata indsamles under opførelsen af tidligere projekter i hele byggevirksomhedens levetid og indtastes i databaser i forskellige systemer (ERP, PMIS BPM, EPM osv.).

> At have historiske omkostningsdata af god kvalitet er en stor konkurrencemæssig fordel for en byggeorganisation, som direkte påvirker dens overlevelsesevne.

Estimerings- og omkostningsafdelinger i bygge- og anlægsvirksomheder er oprettet for at indsamle, lagre og opdatere historiske data om projektberegninger. Deres vigtigste funktion er at akkumulere og systematisere virksomhedens erfaringer, hvilket gør det muligt at forbedre nøjagtigheden af estimeringen af omfang, timing og omkostninger for nye projekter over tid. Denne tilgang hjælper med at minimere fejl i fremtidige beregninger baseret på praksis og resultater fra allerede gennemførte projekter.

Metoder til beregning af de anslåede projektomkostninger

Omkostningsspecialister bruger en række forskellige estimeringsmetoder, som hver især fokuserer på en bestemt type data, tilgængelighed af information og projektets detaljeringsgrad. De mest almindelige omfatter:

- **Ressourcebaseret metode**: estimering af de anslåede omkostninger ved et projekt baseret på en detaljeret analyse af alle de nødvendige ressourcer, f.eks. materialer, udstyr og arbejdskraft. Denne metode kræver en detaljeret liste over alle opgaver og ressourcer, der kræves for at udføre hver opgave, efterfulgt af en beregning af deres omkostninger. Denne metode er meget præcis og bruges i vid udstrækning til omkostningsestimering.
- **Parametrisk metode**: bruger statistiske modeller til at estimere omkostninger baseret på projektparametre. Det kan indebære en analyse af omkostningerne pr. måleenhed, f.eks. bygningsareal eller arbejdsomfang, og en tilpasning af disse værdier til specifikke projektforhold. Metoden er især effektiv i de tidlige faser, hvor der endnu ikke foreligger detaljerede oplysninger.
- Enhedsmetode (enhedsomkostningsmetode): Beregner de anslåede omkostninger for et projekt baseret på en omkostning pr. måleenhed (f.eks. pr. kvadratmeter eller kubikmeter). Det giver en hurtig og nem måde at sammenligne og analysere omkostningerne ved forskellige projekter eller dele af projekter.
- **Ekspertvurdering** (Delphi-metoden): baseret på udtalelser fra eksperter, der bruger deres erfaring og viden til at estimere værdien af et projekt. Metoden er nyttig, når der ikke findes nøjagtige baseline-data, eller når projektet er unikt.

Det er værd at bemærke, at den parametriske metode og ekspertvurderinger kan tilpasses til maskinlæringsmodeller. Det giver mulighed for automatisk at generere prognoser for projektomkostninger

og tidsplaner baseret på træningseksempler. Eksempler på anvendelse af sådanne modeller diskuteres mere detaljeret i kapitlet "Et eksempel på brug af maskinlæring til at finde projektomkostninger og tidslinjer" (Fig. 9.3-5).

Ikke desto mindre er den ressourcebaserede metode stadig den mest populære og udbredte i verdenspraksis. Den giver ikke kun en nøjagtig vurdering af de anslåede omkostninger, men gør det også muligt at beregne varigheden af de enkelte processer på byggepladsen og hele projektet som helhed (flere detaljer i kapitlet "Byggeplaner og 4D - projektdata").

Ressourcebaseret metode omkostningsberegning og estimering i byggeriet

Ressourcebaseret omkostningsberegning er en metode til ledelsesregnskab, hvor omkostningerne ved et projekt er baseret på en direkte beregning af alle involverede ressourcer. I byggeriet indebærer denne tilgang en detaljeret analyse og evaluering af alle materialer, arbejdskraft og tekniske ressourcer, der kræves for at udføre arbejdet.

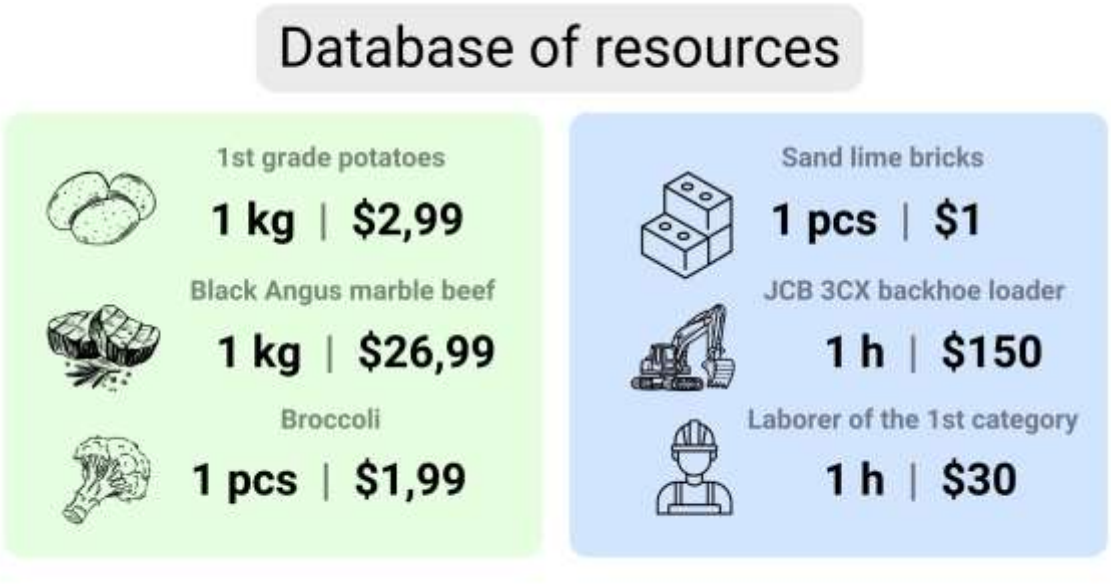
> Den ressourcebaserede metode giver en høj grad af gennemsigtighed og nøjagtighed i budgetplanlægningen, da den fokuserer på de faktiske priser på ressourcer på estimeringstidspunktet. Dette er især vigtigt i et ustabilt økonomisk miljø, hvor prisudsving kan påvirke de samlede omkostninger ved et projekt betydeligt.

I de følgende kapitler vil vi se nærmere på den ressourcebaserede omkostningsproces. For bedre at forstå dens principper i konstruktionen vil vi drage en analogi til omkostningsberegningen af en middag på en restaurant. Når restaurantchefen planlægger aftenen, laver han en liste over de nødvendige produkter, tager højde for tilberedningstiden for hver ret og ganger derefter omkostningerne med antallet af gæster. I byggeriet er processen den samme: For hver kategori af projektelementer (objekter) genereres specificerede estimater Opskrifter, og projektets samlede omkostninger bestemmes ved at summere alle omkostninger i en samlet regning - det endelige estimat pr. kategori.

Den centrale og indledende fase i den ressourcebaserede tilgang er oprettelsen af virksomhedens oprindelige database. I den første fase af omkostningsberegningen udarbejdes en struktureret liste over alle varer, materialer, arbejdstyper og ressourcer, som virksomheden har til rådighed i sine byggeprojekter - fra et søm på lageret til en beskrivelse af mennesker med deres kvalifikationer og timepris. Disse oplysninger systematiseres i en enkelt "Construction Resources and Materials Database" - et tabelregister, der indeholder data om navne, egenskaber, måleenheder og aktuelle priser. Det er denne database, der bliver den vigtigste og primære informationskilde til alle efterfølgende ressourceberegninger - både omkostninger og tidsplan for arbejdet.

Database over byggeressourcer: katalog over byggematerialer og -arbejder

En database eller tabel over byggeressourcer og -materialer - indeholder detaljerede oplysninger om hvert element, der kan bruges i et byggeprojekt - et produkt, en genstand, et materiale eller en tjeneste, herunder navn, beskrivelse, måleenhed og enhedsomkostninger, registreret i en struktureret form. I denne tabel kan du finde alt fra forskellige typer brændstof og materialer, der bruges i projekter, til detaljerede lister over specialister i form af forskellige kategorier med beskrivelser af timepriser (fig. 5.1-2).



Figur 5.1-2 En ressourcetabel er en ingrediensliste, der beskriver et materiale og en tjeneste med en enhedsomkostning.

"Ressourcedatabasen" svarer til produktkataloget i en onlinebutik, hvor hver vare har en detaljeret beskrivelse af sine egenskaber. Det gør det lettere for kalkulatører at vælge de rigtige ressourcer (som at vælge produkter, når man lægger dem i indkøbskurven), der er nødvendige for at beregne specifikke byggeprocesser i form af beregninger (endelig ordre i onlinebutikken).

En ressourcedatabase kan også betragtes som en liste over alle ingredienserne i en restaurantkogebog. Hvert byggemateriale, udstyr og service svarer til de ingredienser, der bruges i opskrifterne. "Ressourcedatabasen" er en detaljeret liste over alle ingredienser - byggematerialer og serviceydelser, inklusive deres pris pr. enhed: styk, meter, time, liter osv.

Nye enhedselementer kan tilføjes til tabellen "Construction Resource Databases" på to måder - manuelt (fig. 5.1-3) eller automatisk ved at integrere med virksomhedens lagerstyringssystemer eller leverandørdatabaser.

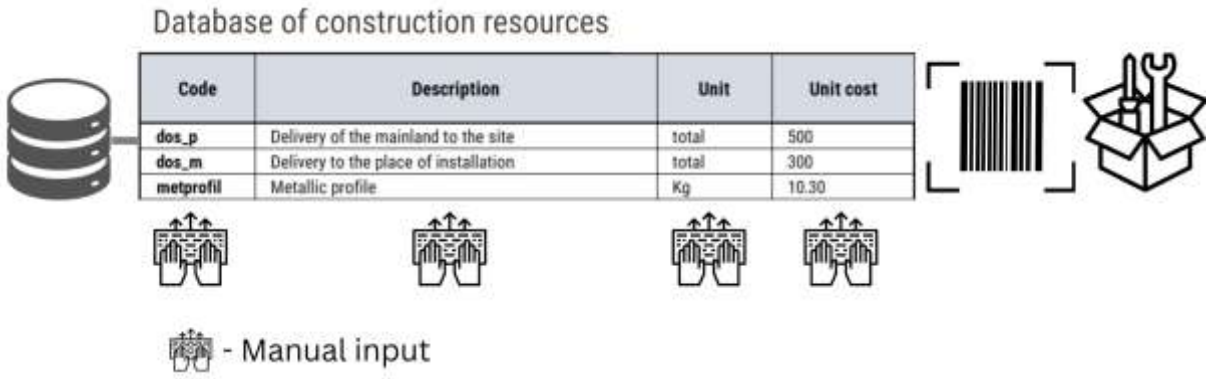


Fig. 5.1-3 Ressourcedatabasen udfyldes manuelt eller overtager automatisk data fra andre databaser.

En typisk mellemstor byggevirksomhed bruger en database, der indeholder tusindvis og nogle gange titusindvis af elementer med detaljerede beskrivelser, som kan bruges i byggeprojekter. Disse data bruges derefter automatisk i kontrakter og projektdokumentation til nøjagtigt at beskrive arbejdets sammensætning og processer

> For at holde trit med ændrede markedsforhold som f.eks. inflation opdateres attributten "enhedsomkostninger" for hvert produkt (vare eller tjenesteydelse) i ressourcedatabasen (figur 5.1-3) regelmæssigt manuelt eller ved automatisk at downloade aktuelle priser fra andre systemer eller onlineplatforme.

Opdatering af enhedsomkostningerne for en ressource kan ske månedligt, kvartalsvist eller årligt - afhængigt af ressourcens art, inflationen og det eksterne økonomiske klima. Sådanne opdateringer er nødvendige for at opretholde nøjagtigheden af beregninger og estimater, da disse er de grundlæggende elementer, der tjener som udgangspunkt for omkostningsberegnernes arbejde. Opdaterede data bruges til at generere estimater, budgetter og tidsplaner, der afspejler de reelle markedsforhold og reducerer risikoen for fejl i efterfølgende projektberegninger.

Udarbejdelse af beregninger og omkostningsberegning af arbejder baseret på ressourcegrundlag

Når du har fyldt "Construction Resource Database" (Fig. 5.1-3) med minimumsenheder, kan du begynde at oprette beregninger, som beregnes for hver proces eller arbejde på byggepladsen for bestemte måleenheder: for eksempel for en kubikmeter beton, en kvadratmeter gipspladevæg, pr. meter kantsten eller pr. vinduesinstallation.

For eksempel, for at bygge en mur på 1 m² (figur 5.1-4), baseret på erfaring fra tidligere projekter, kræves der ca. 65 mursten (enhed "Silikatsten") til en pris på 1 $ pr. stk (attribut "Pris pr. stk."), i alt 65 $. Det er også min erfaring, at det er nødvendigt at bruge entreprenørmaskiner (enheden "JCB 3CX Loader") i 10 minutter til at placere mursten i nærheden af arbejdsområdet. Da det koster $150 pr. time at leje udstyret, vil 6 minutters brug koste ca. $15. Derudover er der brug for en murer i 2 timer med en timepris på 30 dollars og i alt 60 dollars.

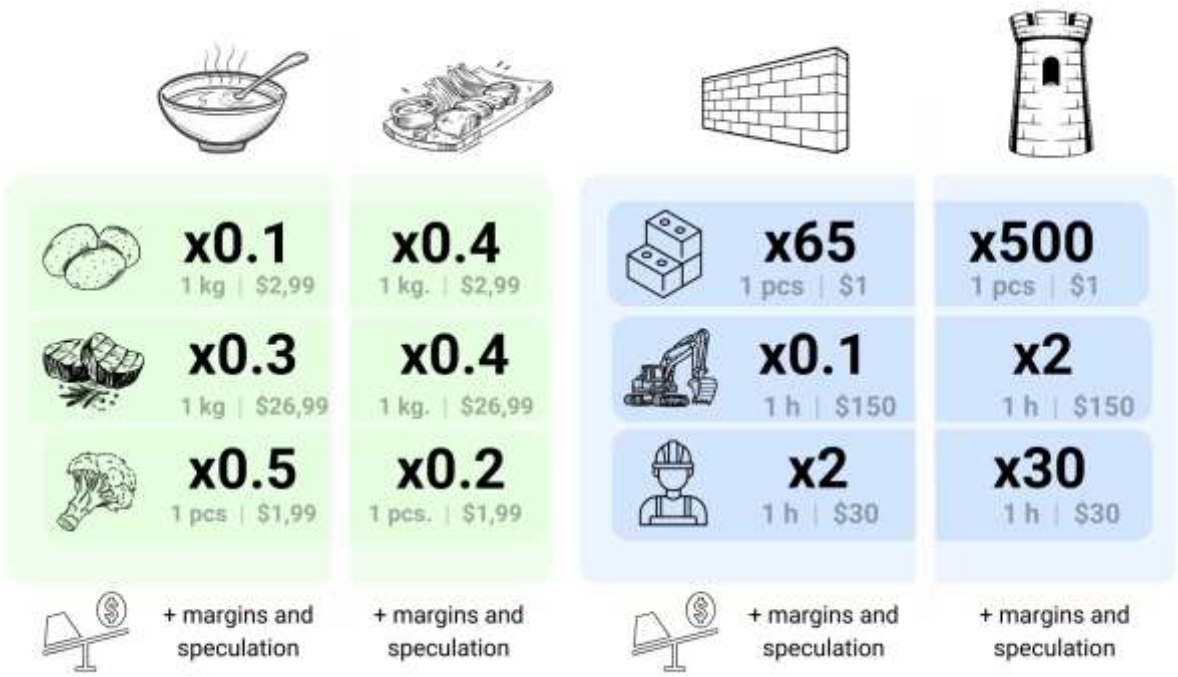


Fig. 5.1-4 Omkostningsberegningerne indeholder en detaljeret liste over byggematerialer og tjenesteydelser, der er nødvendige for at udføre arbejdet og processerne.

Sammensætningen af beregninger (såkaldte "opskrifter") er dannet på baggrund af historiske erfaringer, som virksomheden har akkumuleret i forbindelse med udførelsen af en stor mængde lignende arbejde. Denne praktiske erfaring akkumuleres normalt gennem feedback fra byggepladsen. Især indsamler formanden oplysninger direkte på byggepladsen og registrerer de faktiske lønomkostninger, materialeforbrug og nuancer i den teknologiske drift. I samarbejde med kalkulationsafdelingen forfines disse oplysninger derefter iterativt: Procesbeskrivelser forfines, ressourcemix justeres, og omkostningsberegninger opdateres for at afspejle faktiske data fra de seneste projekter.

> Ligesom en opskrift beskriver de ingredienser og mængder, der er nødvendige for at tilberede en ret, indeholder et overslag en detaljeret liste over alle byggematerialer, ressourcer og tjenester, der er nødvendige for at gennemføre et bestemt job eller en bestemt proces.

Regelmæssigt udført arbejde gør det muligt for arbejdere, formænd og kalkulatorer at orientere sig i den nødvendige mængde ressourcer: materialer, brændstof, arbejdstid og andre parametre, der kræves for at udføre en arbejdsenhed (fig. 5.1-5). Disse data indtastes i kalkulationssystemerne i form af tabeller, hvor hver opgave og operation beskrives ved hjælp af minimumselementerne i ressourcebasen (med konstant opdaterede priser), hvilket sikrer nøjagtigheden af beregningerne.

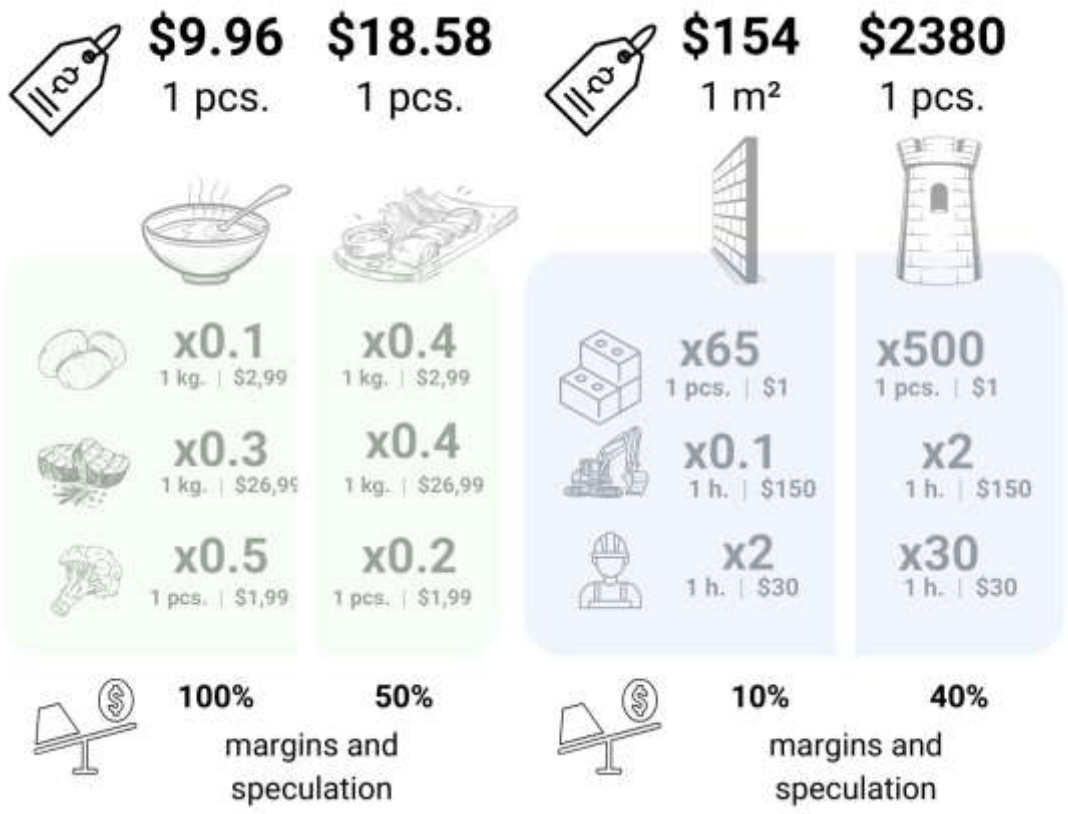


Fig. 5.1-5 Enhedspriser indsamles for hvert job, hvor en enheds volumenattribut ganges med dens mængde, og der tilføjes en profitprocent.

For at få de samlede omkostninger for hver proces eller aktivitet (kalkulationsobjekt) ganges omkostningsattributten med dens antal og faktorer. Koefficienterne kan tage højde for forskellige faktorer, f.eks. arbejdets kompleksitet, regionale karakteristika, inflationsrate, potentielle risici (forventet overheadprocent) eller spekulation (ekstra profitfaktor).

> Som analytiker omdanner beregneren formandens erfaringer og anbefalinger til standardiserede beregninger, der beskriver byggeprocesser gennem ressourceenheder i tabelform. I bund og grund er beregnerens opgave at indsamle og strukturere de oplysninger, der kommer fra byggepladsen, ved hjælp af parametre og koefficienter.

De endelige omkostninger pr. arbejdsenhed (f.eks. kvadratmeter eller kubikmeter eller en installation af en enkelt enhed) omfatter således ikke kun de direkte omkostninger til materialer og arbejdskraft, men også virksomhedstillæg, generalomkostninger, forsikring og andre faktorer (figur 5.1-6)

Samtidig behøver vi ikke længere at bekymre os om de faktiske priser i (opskrifts-)beregninger, da de reelle priser altid afspejles i "ressourcebasen" (ingrediens-tabellen). På beregningsniveauet indlæses data fra ressourcebasen automatisk i tabellen (f.eks. efter varekode eller dens unikke identifikator), som indlæser beskrivelsen og den faktiske pris pr. enhed, som igen automatisk kan indlæses fra onlineplatforme eller onlinebutikker med byggematerialer. Estimatoren på beregningsniveau behøver kun at beskrive arbejdet

eller processen gennem attributten "ressourcemængde" og yderligere faktorer.

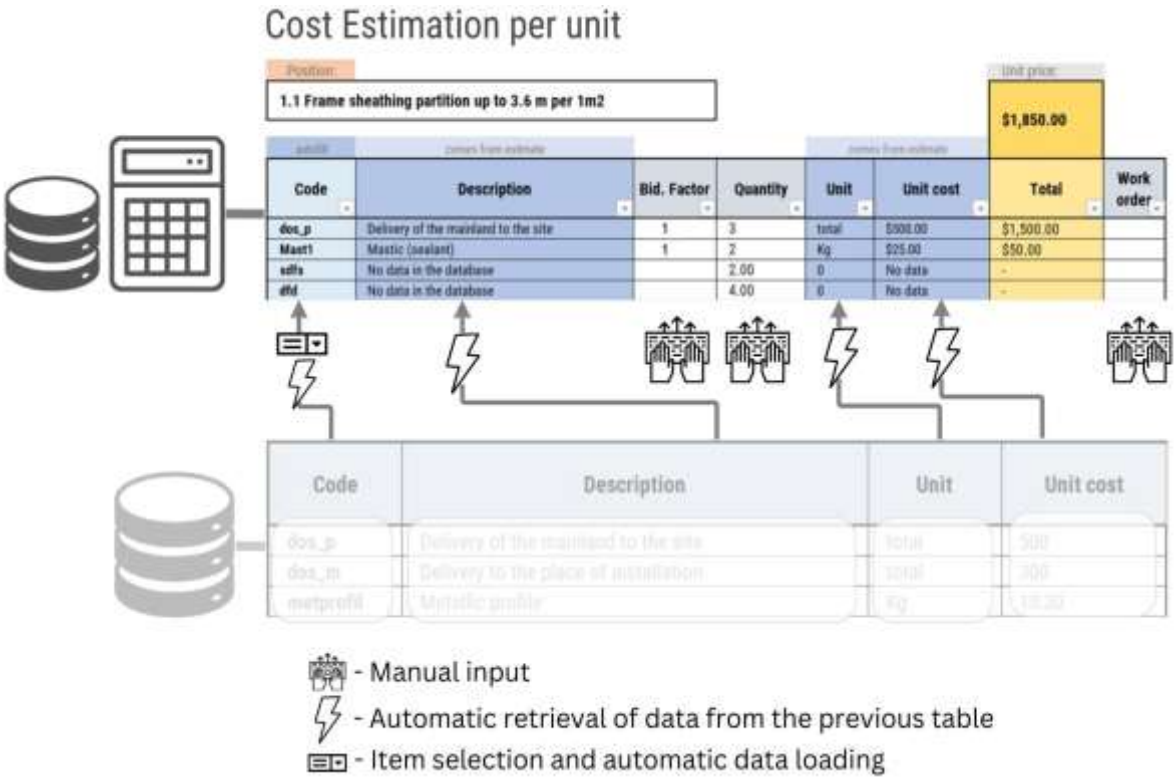


Fig. 5.1-6 Når man beregner arbejdets enhedsomkostninger, er det kun attributterne for den nødvendige mængde ressourcer, der udfyldes, alt andet indlæses automatisk fra ressourcedatabasen.

De oprettede jobkalkulationer gemmes i form af skabelontabeller med typiske projekter, som er direkte forbundet med databasen over byggeressourcer og materialer. Disse skabeloner repræsenterer standardiserede opskrifter på gentagne arbejdstyper til fremtidige projekter, hvilket sikrer ensartethed i beregningerne på tværs af virksomheden.

Når omkostningerne for en ressource ændres i databasen (fig. 5.1-3) - enten manuelt eller automatisk via download af aktuelle markedspriser (f.eks. under inflationsforhold) - afspejles opdateringerne straks i alle tilknyttede omkostningskalkulationer (fig. 5.1-6). Det betyder, at det kun er ressourcegrundlaget, der skal ændres, mens omkostningsskemaerne og estimaterne forbliver uændrede over tid. Denne tilgang sikrer, at beregningerne er stabile og reproducerbare i forhold til eventuelle prisudsving, som der kun tages højde for i en relativt enkel ressourcetabel (figur 5.1-3).

For hvert nyt projekt oprettes en kopi af standardomkostningsskabelonen, som giver dig mulighed for at foretage ændringer og justere aktiviteter for at opfylde specifikke krav uden at ændre den oprindelige skabelon, som virksomheden har vedtaget. Denne tilgang giver fleksibilitet i tilpasningen af beregningerne: Man kan tage højde for byggepladsens særlige forhold, kundens ønsker, indføre risiko- eller rentabilitetskoefficienter (spekulation) - alt sammen uden at bryde med virksomhedens standarder. Det hjælper virksomheden med at finde en balance mellem profitmaksimering, kundetilfredshed og opretholdelse af konkurrenceevnen.

I nogle lande er sådanne omkostningsskabeloner, der er akkumuleret gennem årtier, standardiseret på nationalt niveau og bliver en del af de nationale standarder for byggeomkostningssystemer (figur 5.1-7).

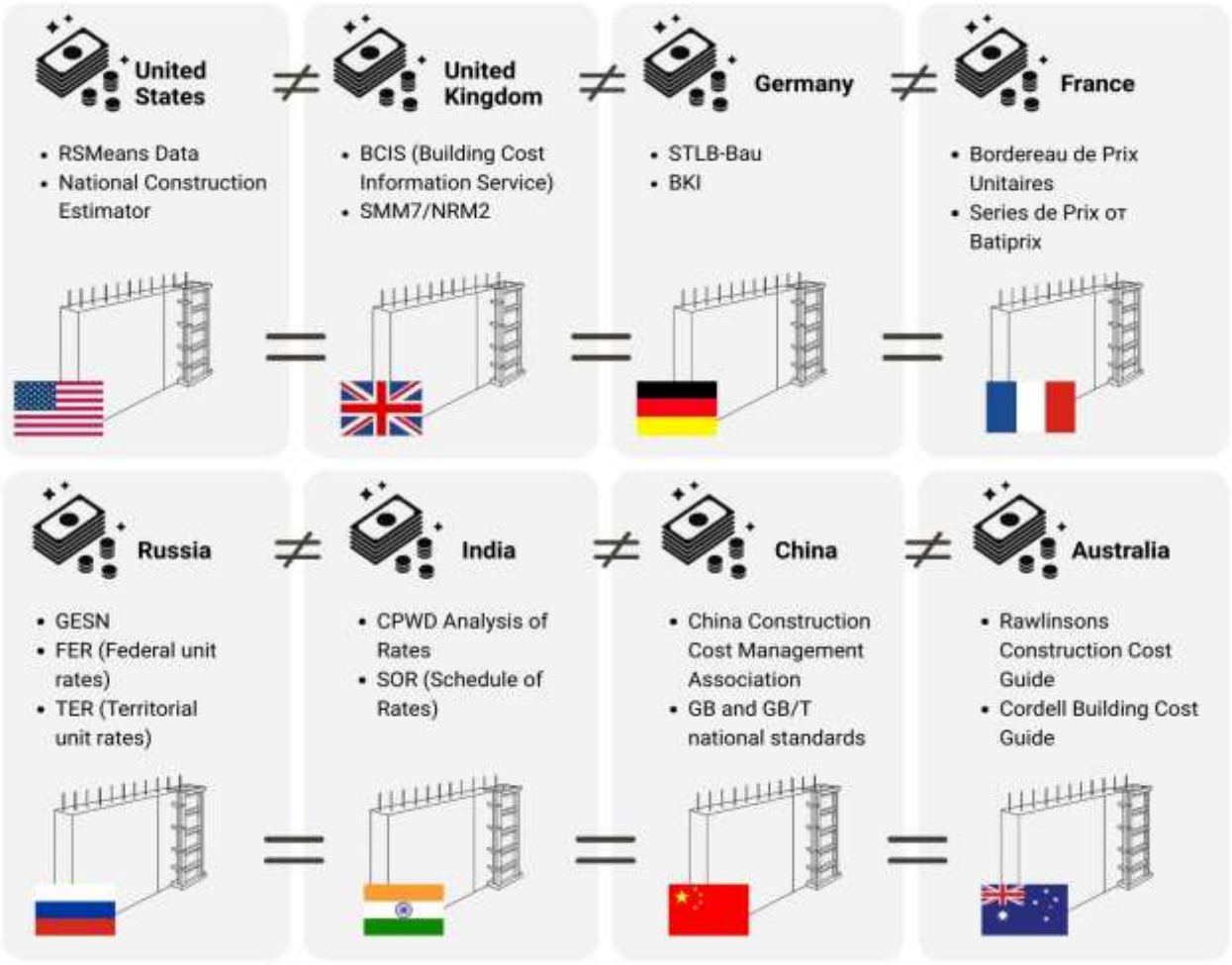


Fig. 5.1-7 Forskellige lande rundt om i verden har deres egne regler for omkostningsberegning med deres egne (præskriptive) kompendier og standarder for byggearbejde til omkostningsberegning af det samme element.

Sådanne standardiserede ressourcebaser med estimater (figur 5.1-7) er obligatoriske alle

markedsdeltagere, især for offentligt finansierede projekter. En sådan standardisering sikrer gennemsigtighed, sammenlignelighed og retfærdighed i prisdannelsen og kontraktlige forpligtelser for kunden

Endelig projektberegning: fra estimater til budgetter

Statslige og branchespecifikke beregningsstandarder spiller forskellige roller i byggepraksis i forskellige lande. Mens nogle lande kræver streng overholdelse af en enkelt standard, anvender de fleste udviklede økonomier en mere fleksibel tilgang. I markedsøkonomier fungerer statslige byggestandarder normalt kun som en basislinje. Byggefirmaer tilpasser disse standarder til deres driftsmodeller eller reviderer dem helt og supplerer dem med deres egne tilpassede faktorer. Disse justeringer afspejler virksomhedens erfaring, ressourceforvaltningseffektivitet og ofte faktorer, hvor f.eks. en virksomheds spekulationsgevinster kan indregnes.

> Som følge heraf kan konkurrenceniveauet, markedsefterspørgslen, målmarginer og endda forholdet til specifikke kunder føre til betydelige afvigelser fra standardiserede normer. Denne praksis giver markedsfleksibilitet, men gør det også vanskeligt at sammenligne bud fra forskellige entreprenører på en gennemsigtig måde, hvilket introducerer et element af spekulativ prissætning i byggebranchen på dette stadie af beregningsprocessen.

Når beregningsskabelonerne for de enkelte aktiviteter og processer er udarbejdet - eller oftere blot kopieret fra regeringens standardoverslag (fig. 5.1-7) med koefficienter tilføjet for at afspejle en bestemt virksomheds "særegenheder" - er det sidste trin at gange omkostningerne for hver post med den tilsvarende egenskab for arbejdsomfanget eller processerne i det nye projekt.

> Når man beregner de samlede omkostninger for et nyt byggeprojekt, er det vigtigste skridt at sammenfatte omkostningerne for alle kalkulationsposter, ganget med mængden af disse arbejdsposter i projektet.

For at skabe projektets samlede omkostninger starter vi i vores forenklede eksempel med at beregne omkostningerne ved at bygge en kvadratmeter væg og ganger omkostningerne ved beregningen (f.eks. arbejdet "1m² standardinstallation af vægelementer") med det samlede antal kvadratmeter væg i projektet (f.eks. attributten "Areal" eller "Mængde" (fig. 5.1-8) for en enhed af typen "Vægelementer" fra projektets CAD eller værkførerens beregninger).

Vi beregner omkostningerne for alle projektelementer på samme måde (fig. 5.1-8): Vi tager omkostningerne pr. arbejdsenhed og ganger dem med mængden af et bestemt element eller en gruppe af elementer i projektet. Kalkulatøren skal kun indtaste antallet af disse elementer, aktiviteter eller processer i projektet på i form af volumen eller mængde. Dette gør det muligt automatisk at generere et komplet byggeoverslag.

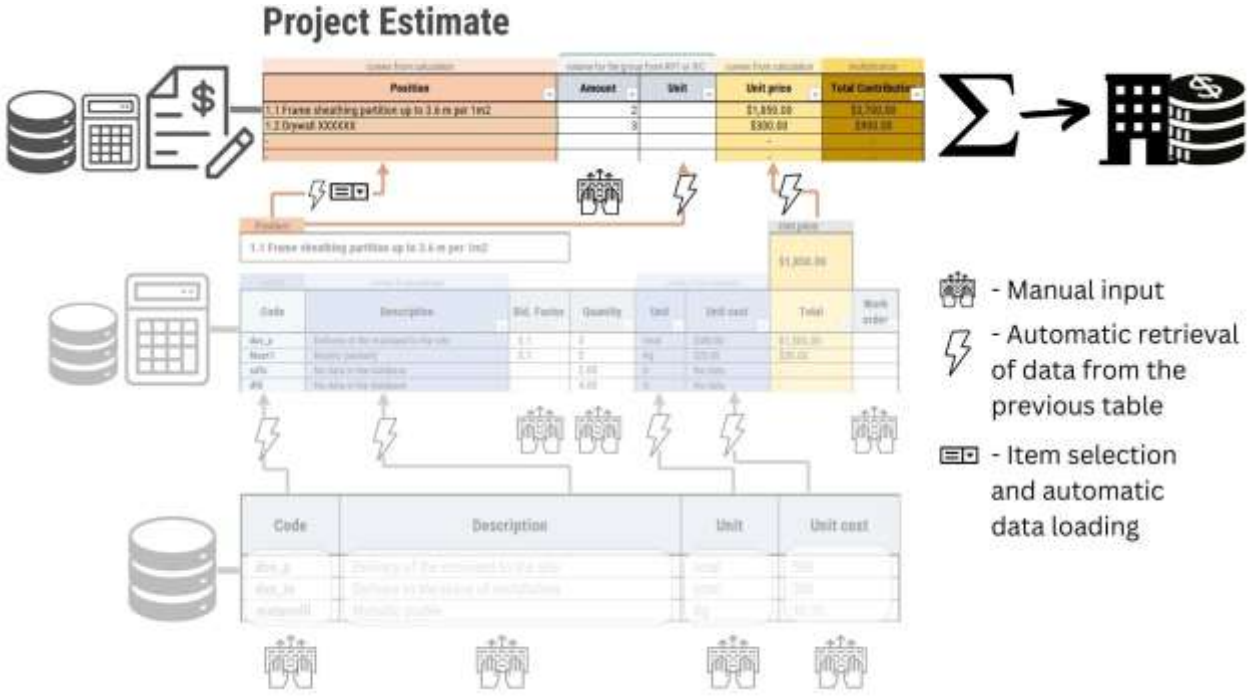


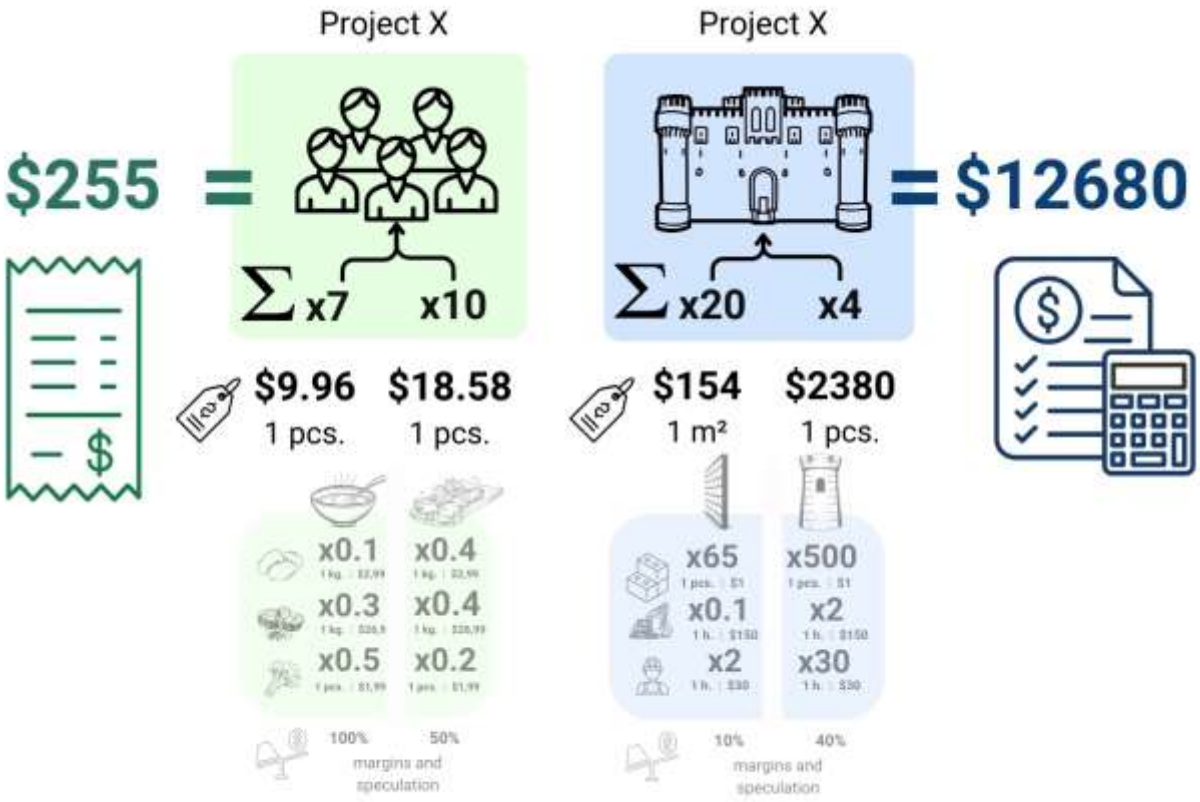
Fig. 5.1-8 I den fase, hvor vi opretter et overslag, indtaster vi kun arbejdets omfang.

Som i tilfældet med beregninger uploader vi på dette niveau automatisk færdige beregnede poster (fra beregningsskabelonen eller nye, der er kopieret fra skabelonen og redigeret), som automatisk medbringer de aktuelle omkostninger pr. arbejdsenhed (som opdateres automatisk fra ressourcedatabasen (Fig. 5.1-8 nederste tabel)). I tilfælde af ændringer af data i ressourcedatabasen eller kalkulationstabellerne vil dataene i estimatet derfor automatisk blive opdateret for den aktuelle dag, uden at det er nødvendigt at ændre selve kalkulationen eller estimatet.

I en restaurantsammenhæng beregnes de endelige omkostninger for et arrangement på samme måde og svarer til de endelige omkostninger for hele middagen, hvor omkostningerne for hver ret ganget med antallet af gæster giver de samlede omkostninger for checken (figur 5.1-9). Og ligesom i byggebranchen er det ikke sikkert, at opskrifterne på en restaurant ændrer sig i årtier. I modsætning til priser, hvor prisen på ingredienser kan ændre sig hver time.

> Ligesom en restaurantejer ganger prisen på hvert måltid med antallet af portioner og personer for at finde frem til de samlede omkostninger ved arrangementet, lægger lederen af omkostningsberegningen prisen på alle projektets komponenter sammen for at nå frem til et komplet anlægsoverslag.

Således bestemmes de endelige omkostninger for hver aktivitet i projektet (fig. 5.1-9), som ganget med attributmængden for den enhed, der svarer til denne aktivitet, giver omkostningerne for grupper af aktiviteter, hvorfra de endelige omkostninger for hele projektet opnås.



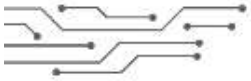
Figur 5.1-9 Det endelige estimat beregnes ved at summere hvert elements arbejdsomkostningsattribut med dets scope-attribut.

De samlede projektomkostninger (figur 5.1-8) giver et økonomisk billede af projektet, så kunder, investorer eller finansielle organisationer kan forstå det samlede budget og de finansielle ressourcer, der kræves for at levere projektet på en given dag, under hensyntagen til de aktuelle priser.

Og hvis processerne med at udarbejde ressourcebaser, beregninger og estimater (procesopskrifter) allerede er blevet udarbejdet, halvautomatiseret og finpudset i titusindvis af år og registreret på statsniveau, så er automatisk indhentning af kvalitativ information om volumen og mængde af elementer til den sidste fase af det endelige estimat - i dag stadig en flaskehals i processerne med alle beregninger af projektets omkostnings- og tidsattributter og generelt projektets samlede budget.

> I tusindvis af år har den traditionelle metode til at beregne volumen været manuelle metoder til at måle volumen og mængde ved hjælp af flade tegninger. Med fremkomsten af den digitale tidsalder har virksomheder opdaget, at volumen- og mængdeinformation nu automatisk kan udtrækkes fra geometriske data i CAD-modeller, hvilket revolutionerer årtusinders gamle måder at indhente kvantitative data på.

Moderne tilgange til procesestimering og estimering involverer automatisk udtrækning af volumetriske og kvantitative attributter fra CAD-databaser, som kan uploades og forbindes med omkostningsprocessen for at få opdaterede projektgruppemængder på ethvert trin fra design til drift.



KAPITEL 5.2.
MÆNGDEBEREGNING OG AUTOMATISK OPRETTELSE AF ESTIMATER OG TIDSPLANER

At gå fra 3D til 4D og 5D: ved hjælp af volumetriske og kvantitative parametre

Med kalkulationstabellerne med de beskrevne processer gennem ressourcer (fig. 5.1-8) i hånden er det næste skridt automatisk at indhente volumen- eller mængdeparametre for en gruppe elementer, der er nødvendige for beregningerne og for det endelige estimat.

Volumetriske egenskaber for projektelementer - f.eks. vægge eller plader - kan automatisk udtrækkes fra CAD-databaser. Parametriske objekter, der er oprettet i CAD-programmer, konverteres ved hjælp af en geometrisk kerne til numeriske værdier for længde, bredde, areal, volumen og andre parametre. Processen med at opnå volumener baseret på 3D-geometri vil blive diskuteret mere detaljeret i den næste, sjette del (Fig. 6.3-3), der er dedikeret til at arbejde med CAD (BIM). Ud over volumener kan antallet af lignende elementer også fås fra CAD-modeldatabasen ved at filtrere og gruppere objekter efter kategorier og egenskaber. Disse parametre, som tillader gruppering, bliver grundlaget for at forbinde projektelementer gennem ressourceberegninger med beregninger, endelige estimater og budget for hele projektet.

Datamodellen fra 3D-modellen (CAD) udvides således med nye parameterlag, der betegnes som 4D og 5D. I de nye entitetsattributlag, 4D (tid) og 5D (omkostninger), bruges 3D-geometriske data som kilde til entitetsvolumenattributværdier.

- **4D** er et parameterinformationslag, der tilføjer oplysninger om byggeoperationernes varighed til elementernes 3D-parametre. Disse data er vigtige for planlægning af tidsplaner og styring af projektets tidslinjer
- **5D** er det næste niveau i udvidelsen af datamodellen, hvor elementerne suppleres med omkostningskarakteristika. Dette tilføjer et økonomisk aspekt til den geometriske information: omkostningerne til materialer, arbejde og udstyr, hvilket muliggør budgetberegninger, rentabilitetsanalyser og omkostningsstyring under byggeprocessen.

Omkostninger og 3D, 4D og 5D attributdata for grupper af projektenheder beskrives på samme måde som beregninger i modulære ERP, PIMS -systemer (eller Excel -lignende værktøjer) og bruges til automatisk omkostningskalkulation og budgetplanlægning af både individuelle grupper og det komplette projektbudget.

5D-attributter og hentning af attributmængder fra CAD

Når man udarbejder det endelige overslag for et byggeprojekt, hvis udarbejdelse vi har diskuteret i tidligere kapitler (figur 5.1-8), indsamles volumenattributterne for hver kategori af projektelementer enten manuelt eller udtrækkes fra volumenattributspecifikationerne i CAD-softwaren.

Den traditionelle manuelle metode til beregning af mængder indebærer, at formanden og beregneren analyserer tegninger, der i tusinder af år er blevet præsenteret som linjer på papir og i de sidste 30 år i digitale formater som PDF (PLT) eller DWG. De trækker på deres professionelle erfaring og måler de nødvendige mængder arbejde og materialer, ofte med lineal og vinkelmåler. Denne metode kræver en betydelig indsats og tid samt en særlig opmærksomhed på detaljer.

Det kan tage alt fra et par dage til flere måneder at bestemme scope-attributter på denne måde, afhængigt af projektets omfang. Da alle målinger og beregninger udføres manuelt, er der desuden risiko for menneskelige fejl, der kan føre til unøjagtige data, som efterfølgende påvirker fejl i estimeringen af projektets tid og omkostninger, hvilket hele virksomheden vil blive holdt ansvarlig for.

Moderne metoder baseret på brug af CAD-databaser forenkler i høj grad beregningen af volumener. I CAD-modeller indeholder elementernes geometri allerede volumenattributter, som kan beregnes automatisk (via geometrikernen (fig. 6.3-3)) og præsenteres eller eksporteres i tabelform.

I et sådant scenarie beder beregningsafdelingen CAD-designeren om data om projektelementernes mængde- og volumenegenskaber. Disse data eksporteres som regneark eller integreres direkte i kalkulationsdatabaser - hvad enten det er Excel-, ERP- eller PMIS - systemer. Denne proces starter ofte ikke med en formel anmodning, men med en kort dialog mellem bygherren (initiativtageren) og arkitekten beregneren fra bygge- eller designfirmaet. Nedenfor ses et forenklet eksempel på, hvordan en struktureret tabel til automatiske beregninger (QTO) dannes ud fra den daglige kommunikation:

- Kunde - "*Jeg vil gerne tilføje endnu en etage til bygningen, i samme konfiguration som den første etage*"
- Arkitekt (CAD) - "*Tilføjelse af en tredje etage, konfigurationen er den samme som på anden etage*". Og efter denne besked sendes en ny CAD-version af projektet til beregneren.
- *Estimator* udfører automatisk gruppering og beregning (ERP, PMIS, Excel) - "*Jeg vil køre projektet gennem et Excel-regneark med QTO-regler (ERP, PMIS), få mængderne efter kategori for den nye etage og generere et estimat*"

Resultatet er, at tekstdialogen omdannes til en tabelstruktur med grupperingsregler:

Element	Kategori	Gulv
Overlapning	OST_Floors	3

Element	Kategori	Gulv
Søjle	OST_StrukturelleKolonner	3
trappe	OST_trapper	3

Efter processen med automatisk gruppering af CAD-modellen fra designeren i henhold til estimatorens QTO-regler og automatisk multiplikation af mængder med ressourceberegninger (fig. 5.1-8) får vi følgende resultater, som sendes til kunden:

Element	Volumen	Gulv	Pris pr. enhed.	Samlede omkostninger
Overlapning	420 м²	3	150 €/м²	**63 000 €**
Søjle	4 stk.	3	2450 €/stk.	**9 800 €**
trappe	2 stk.	3	4.300 €/stk.	**8 600 €**
I ALT:	-	-	-	**81 400 €**

- Kunden - "*Tak, det er ret meget, vi er nødt til at skære et par værelser ned*". Og cyklussen gentager sig mange gange.

Dette scenarie kan gentages mange gange, især i godkendelsesfasen, hvor kunden forventer øjeblikkelig feedback. I praksis kan sådanne processer dog trække ud i dage eller endda uger. Takket være indførelsen af automatiske grupperings- og beregningsregler kan aktiviteter, der tidligere tog lang tid, i dag gennemføres på få minutter. Automatiseret indhentning af mængder gennem grupperingsregler fremskynder ikke kun beregninger og estimater, men ved at minimere den menneskelige faktor reduceres sandsynligheden for fejl, hvilket giver en gennemsigtig og nøjagtig vurdering af projektomkostningerne.

> Hvis der oprindeligt blev taget højde for beregningsafdelingens krav, da 3D-modellen blev oprettet i CAD-systemet (hvilket stadig er sjældent i praksis), og navnene, identifikatorerne for elementgrupper og deres klassifikationsattributter er indstillet i form af parametre, der falder sammen med beregningsgruppernes og -klassernes strukturer, kan volumetriske attributter automatisk overføres til beregningssystemerne uden yderligere omformninger.

Automatisk udtrækning af volumenattributter fra CAD i form af specifikationstabeller gør det muligt hurtigt at få opdaterede data om omkostningerne ved de enkelte arbejder og projektet som helhed (fig. 5.2-1). Ved kun at opdatere CAD-filen med projektmængder i beregningsprocessen eller beregningssystemet kan virksomheden hurtigt genberegne estimatet under hensyntagen til de seneste ændringer, hvilket sikrer høj nøjagtighed og konsistens i alle efterfølgende beregninger.

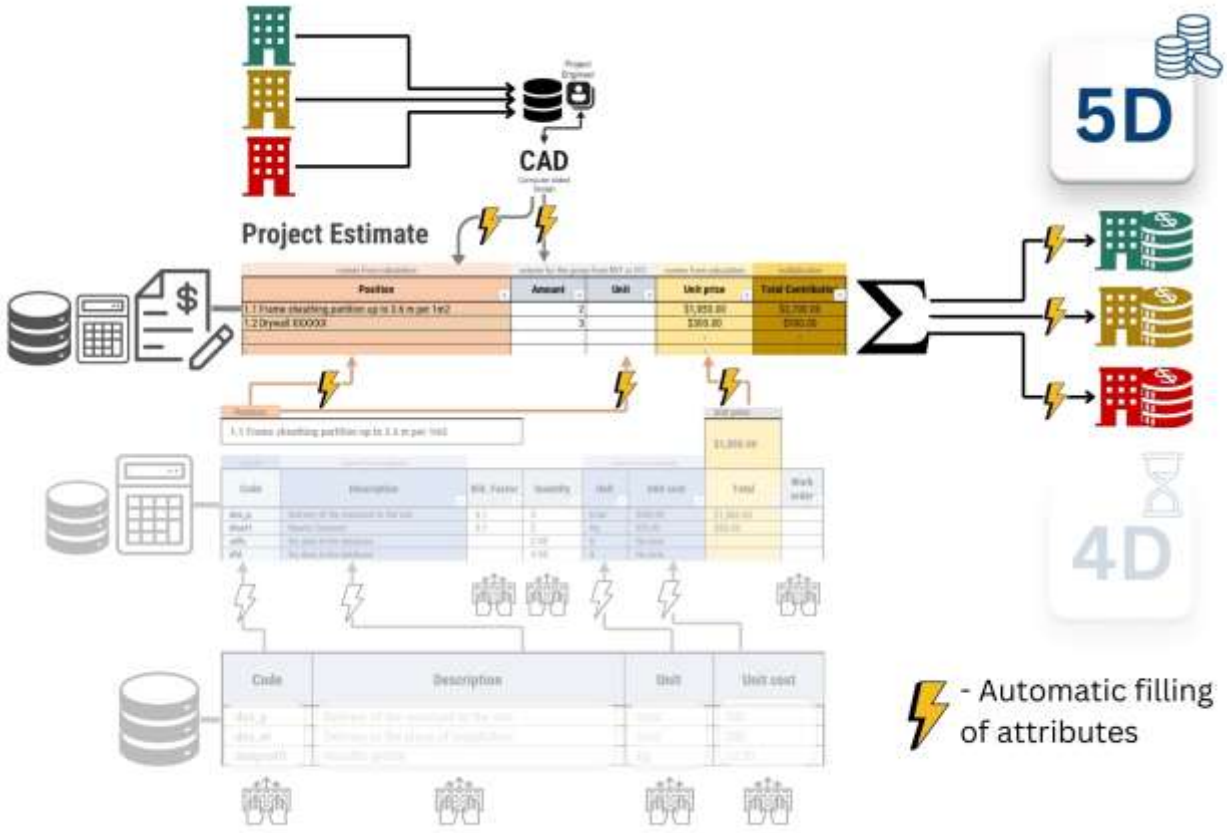


Fig. 5.2-1 Volumenattributter fra CAD-tabeller eller databaser indtastes automatisk i overslaget, så du straks kan beregne de samlede projektomkostninger.

Med den stigende kompleksitet i kapitalprojekter bliver beregning af det fulde budget og analyse af de samlede projektomkostninger under et sådant scenarie (figur 5.2-1) - et vigtigt redskab til informeret beslutningstagning.

> Ifølge Accentures undersøgelse Creating More Value through Capital Projects (2024) [20] integrerer førende virksomheder aktivt dataanalyse i digitale initiativer ved at bruge historiske oplysninger til at forudsige og optimere resultater. Forskning viser, at flere ejere og operatører anvender big data-analyser til at forudsige markedstendenser og vurdere den kommercielle levedygtighed, før designet påbegyndes. Dette opnås ved at analysere datalagre fra en eksisterende portefølje af projekter. Desuden implementerer 79 procent af bygherrerne "robuste" forudsigelige analyser til at vurdere projektresultater og understøtte operationelle beslutninger i realtid.

Moderne effektiv styring af byggeprojekter er uløseligt forbundet med behandling og analyse af store mængder information på alle stadier af designet og de processer, der går forud for designet. Brugen af data warehouses, ressourceberegninger, prædiktive modeller og maskinlæring gør det ikke kun muligt at minimere risici i beregninger, men også at træffe strategiske beslutninger om projektfinansiering i de tidlige stadier af designet. Vi vil tale mere om datalagre og prædiktive modeller, der supplerer beregningerne, i bogens niende del.

Automatisk indhentning af volumetriske parametre for elementer fra CAD-projekter, som er nødvendige for

udarbejdelse af estimater, udføres ved hjælp af grupperingsværktøjer QTO (Quantity Take-Off). QTO-værktøjer fungerer ved at gruppere alle projektobjekter efter særlige elementidentifikatorer eller elementattributparametre ved hjælp af specifikationer og tabeller, der er oprettet i CAD-databasen.

QTO Quantity Take-Off: gruppering af projektdata efter attributter

QTO (Quantity Take-Off) i byggeriet er processen med at udtrække de mængder af elementer, der er nødvendige for at realisere et projekt. I praksis er QTO ofte en semi-manuel proces, der involverer indsamling af data fra forskellige kilder: PDF-dokumenter, DWG-tegninger og digitale CAD-modeller.

Når man arbejder med data hentet fra CAD-databaser, gennemføres QTO-processen som en sekvens af filtrerings-, sorterings-, grupperings- og aggregeringsoperationer. Modelelementer udvælges i henhold til parametrene for klasser, kategorier og typer, hvorefter deres kvantitative attributter - såsom volumen, areal, længde eller mængde - opsummeres i henhold til beregningslogikken (fig. 5.2-2).

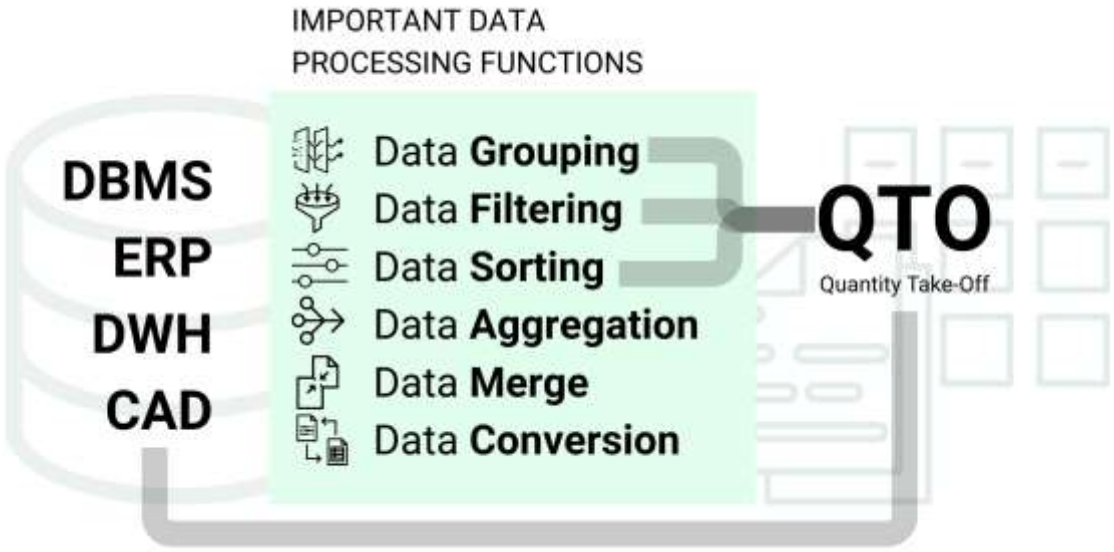


Fig. 5.2-2 Gruppering og filtrering af data er de mest populære funktioner, der anvendes i databaser og datalagre.

QTO-processen (filtrering og gruppering) gør det muligt at systematisere data, danne specifikationer og forberede inputinformation til beregning af estimater, indkøb og arbejdsplaner. Grundlaget for QTO er klassificering af elementer i henhold til den type attributter, der skal måles. For hvert element eller gruppe af elementer vælges den tilsvarende kvantitative måleparameter. Det kan f.eks. være

- **Længdeattribut** (kantsten - i meter)
- **Arealattribut** (gipspladearbejde - i kvadratmeter)
- **Volumenattribut** (betonarbejde - i kubikmeter)
- **Mængdeattribut** (vinduer - pr. stk.)

Ud over de volumetriske egenskaber, der genereres matematisk på grundlag af geometrien, anvendes der ofte overskridelsesfaktorer (fig. 5.2-12 f.eks. 1,1 for at tage højde for 10 procent til logistik og installation) i

beregninger efter QTO-grupperingen - korrektionsværdier, der tager højde for tab, installation, opbevaring eller transport. Det gør det muligt at forudsige det faktiske forbrug af materialer mere præcist og at undgå både mangel på og for store lagre på byggepladsen.

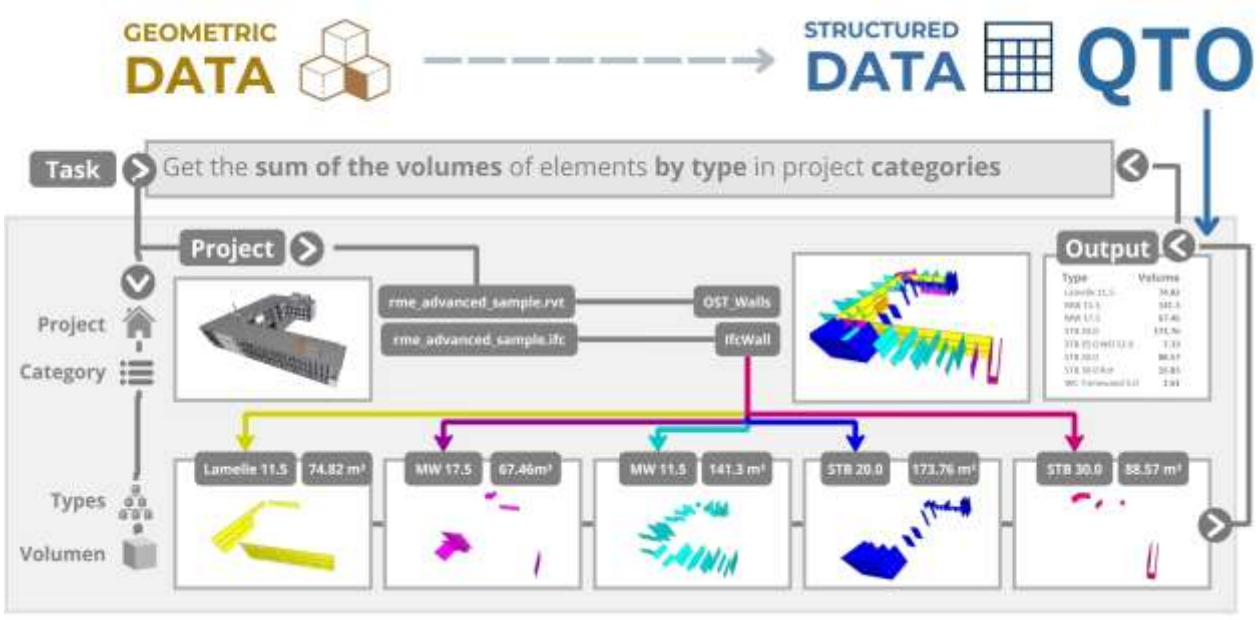
> En automatiseret quantity take-off-proces (QTO) er afgørende for at producere nøjagtige beregninger og estimater, reducere menneskelige fejl i mængdespecifikationsprocesser og forhindre over- eller underbestilling af materialer.

Lad os som et eksempel på QTO-processen betragte et almindeligt tilfælde, hvor det er nødvendigt fra CAD-databasen at vise en tabel-specifikation af volumener efter elementtyper for en bestemt kategori, klasser af elementer. Lad os gruppere alle projektelementer efter type fra CAD-projektets vægkategori og opsummere volumenattributterne for hver type for at præsentere resultatet som en QTO-volumentabel (Fig. 5.2-3).

I eksemplet med et typisk CAD-projekt (fig. 5.2-3) er alle vægkategorielementer i CAD-databasen grupperet efter vægtype, f.eks. "Lamelle 11.5", "MW 11.5" og "STB 20.0", og har veldefinerede volumenattributter repræsenteret i metriske kuber.

Målet for manageren, som befinder sig i grænsefladen mellem designere og beregningsspecialister, er at få en automatiseret tabel over mængder efter elementtype i den valgte kategori. Ikke kun for et specifikt projekt, men også i en universel form, der kan anvendes på andre projekter med en lignende modelstruktur. Det gør tilgangen skalerbar og gør det muligt at genbruge data uden dobbeltarbejde.

> De dage er forbi, hvor erfarne designere og beregnere bevæbnede sig med en lineal og omhyggeligt opmålte hver linje på papir eller PDF -planer - en tradition, der ikke har ændret sig i de sidste årtusinder. Med udviklingen af 3D -modellering, hvor hvert elements geometri nu er direkte forbundet med automatisk beregnede volumetriske attributter, er processen med at bestemme volumener og QTO -mængder blevet automatiseret.



Figur 5.2-3 Indhentning af QTO volumen- og mængdeattributter fra et projekt indebærer gruppering og filtrering af projektelementer.

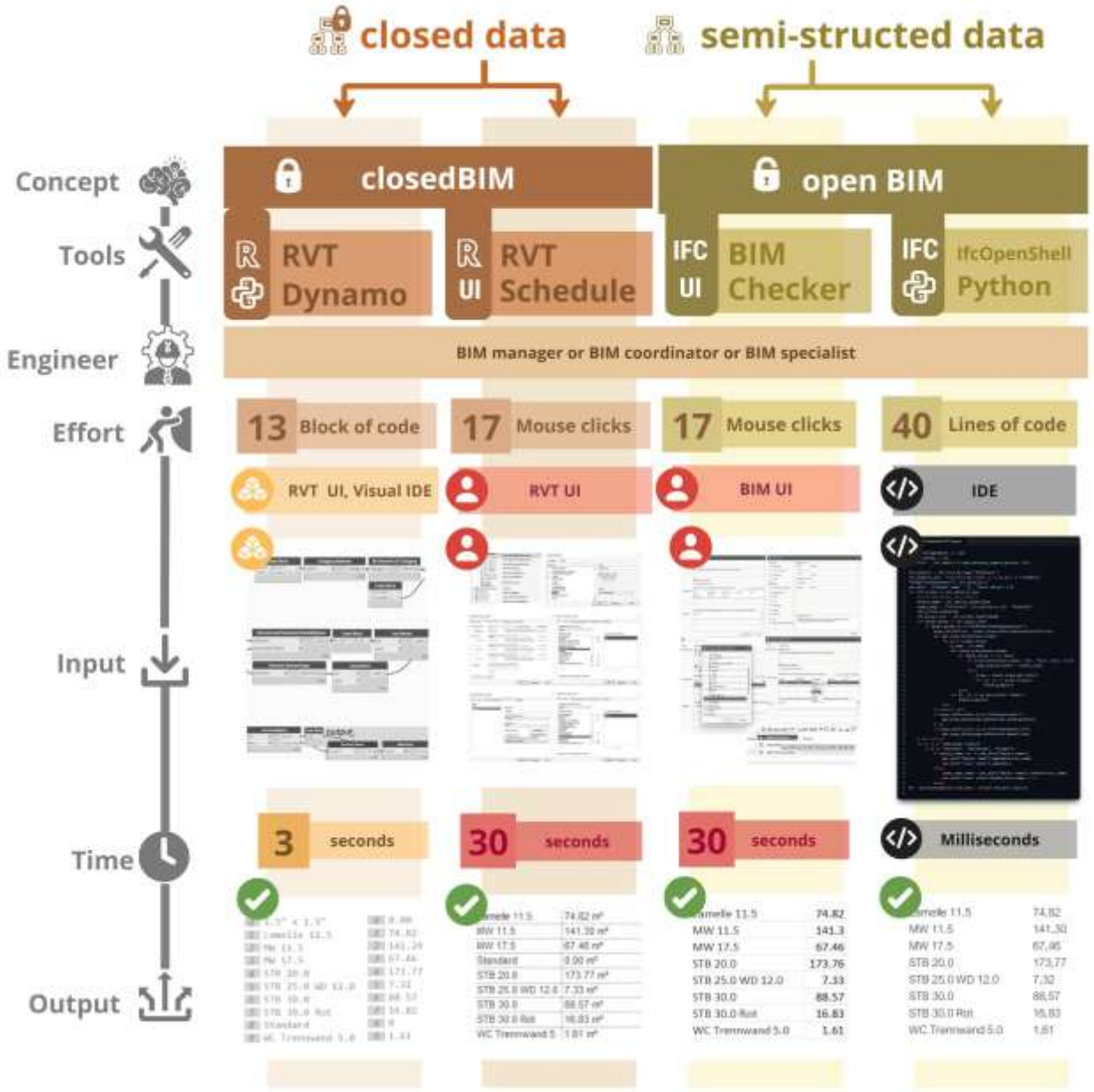
I vores eksempel er opgaven at "vælge en kategori af vægge i et projekt, gruppere alle elementer efter type og præsentere oplysninger om volumenattributter i et struktureret tabelformat", så denne tabel kan bruges af dusinvis af andre specialister til beregninger af omkostninger, logistik, arbejdsplaner og andre business cases (fig. 6.1-3).

På grund af CAD-dataenes lukkede karakter kan ikke alle specialister i dag bruge direkte adgang til CAD-databasen (årsagerne til og løsningerne på adgangsproblemet er beskrevet i bogens sjette del). Derfor er mange nødt til at ty til specialiserede BIM-værktøjer baseret på begreberne åben BIM og lukket BIM [63]. Når man arbejder med specialiserede BIM-værktøjer eller direkte i CAD-softwaremiljøet, kan tabellen med QTO-resultater (Quantity Take-Off) genereres på forskellige måder - afhængigt af, om der anvendes en manuel grænseflade eller softwareautomatisering.

Hvis man f.eks. bruger brugergrænsefladen til CAD (BIM)-software, er det nok at udføre ca. 17 handlinger (knapklik) for at få en klar tabel over volumener (fig. 5.2-4). Brugeren skal dog have en god forståelse af modelstrukturen og funktionerne i CAD (BIM)-softwaren.

Hvis automatisering anvendes via programkode eller via plug-ins og API-værktøjer i CAD-programmer, reduceres antallet af manuelle trin for at få volumentabellerne, men der skal skrives 40 til 150 kodelinjer, afhængigt af det anvendte bibliotek eller værktøj:

- **IfcOpSh (åben BIM**) eller **Dynamo IronPython (lukket BIM)** - giver dig mulighed for at få en QTO-tabel fra et CAD-format eller CAD-program på bare ~40 linjer kode.
- **IFC_js (åben BIM**) - kræver ca. 150 linjer kode for at udtrække omfattende attributter fra IFC-modellen.
- **Interface CAD-værktøjer (BIM**) - giver dig mulighed for at få det samme resultat manuelt med 17 museklik.



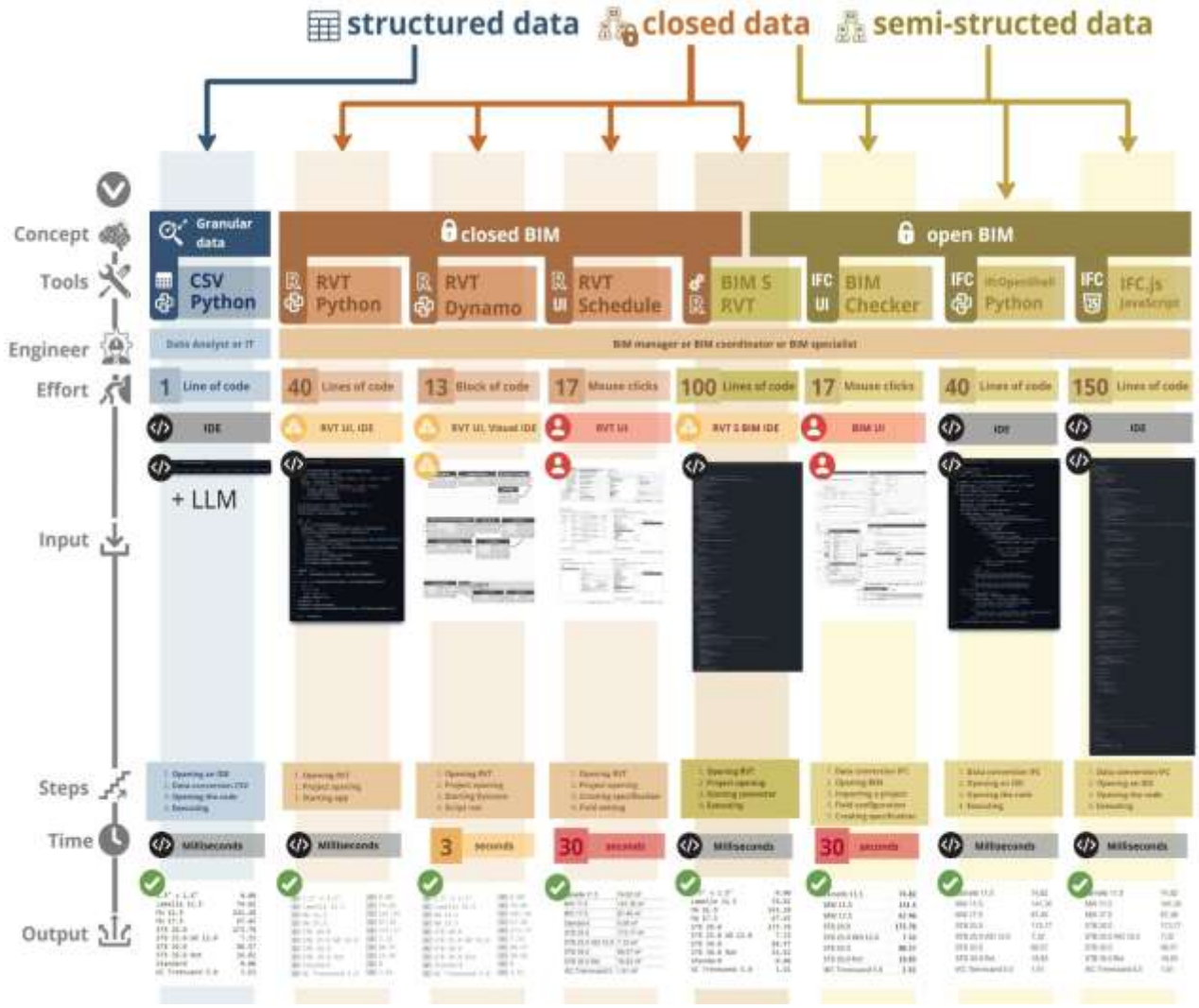
Figur 5.2-4 CAD (BIM) designere ledere bruger 40 til 150 linjer kode eller et dusin tastetryk til at oprette QTO tabeller.

Resultatet er det samme - en struktureret tabel med volumenattributter for en gruppe af elementer. Den eneste forskel er lønomkostningerne og det niveau af teknisk ekspertise, som brugeren skal have (fig. 5.2-4). Moderne værktøjer i forhold til manuel indsamling af mængder fremskynder QTO-processen betydeligt og reducerer sandsynligheden for fejl. De gør det muligt at udtrække data direkte fra projektmodellen, hvilket eliminerer behovet for manuelt at genberegne mængder fra tegninger, som man gjorde før i tiden.

Uanset hvilken metode der anvendes - uanset om det er åben BIM eller lukket BIM - er det muligt at få en identisk QTO - tabel med projektelementmængder (fig. 5.2-4). Men når man arbejder med projektdata i CAD - (BIM-) koncepter, er brugerne afhængige af specialiserede værktøjer og API'er, der leveres af leverandører

(fig. 3.2-13). Dette skaber yderligere lag af afhængighed og kræver, at man lærer unikke dataskemaer, samtidig med at den direkte adgang til dataene begrænses.

På grund af CAD-datas lukkede karakter vanskeliggør indhentning af QTO-tabeller og andre parametre automatisering af beregninger og integration med eksterne systemer. Ved at bruge værktøjer til direkte adgang til databaser og overføre CAD -data fra projektet ved hjælp af reverse engineering-værktøjer til et åbent struktureret dataframe-format (fig. 4.1-13), kan en identisk QTO -tabel opnås med blot én kodelinje (fig. 5.2-5 - variant med granulære data).



Figur 5.2-5 Forskellige værktøjer producerer de samme resultater i form af attributtabeller for projektenheder, men med forskellige arbejdsomkostninger.

Når man bruger åbne strukturerede data fra CAD projekter, som nævnt i kapitlet "Konvertering af CAD (BIM) data til struktureret form", er grupperingsprocessen, QTO, meget forenklet.

Tilgange baseret på brugen af åbne strukturerede data eller direkte adgang til CAD-modeldatabaser er fri for de markedsføringsmæssige begrænsninger, der er forbundet med akronymet BIM. De bygger på gennemprøvede værktøjer, der længe har været brugt i andre brancher (fig. 7.3-10 ETL proces).

> Ifølge McKinsey-undersøgelsen "Open Data: Unleash Innovation and Productivity with Streaming Information" [102], der blev gennemført i 2013, kan brugen af åbne data skabe muligheder for besparelser på 30 til 50 milliarder dollars om året inden for design, teknik, indkøb og opførelse af elproduktionsanlæg. Det svarer til en besparelse på 15 procent i anlægsomkostningerne.

Arbejdet med åbne strukturerede (granulære) data forenkler informationssøgning og -behandling, reducerer afhængigheden af specialiserede BIM-platforme og baner vejen for automatisering uden behov for at bruge proprietære systemer eller parametriske og komplekse datamodeller fra CAD-formater.

QTO-automatisering ved hjælp af LLM og strukturerede data

Oversættelse af ustrukturerede data til en struktureret form forbedrer effektiviteten af forskellige processer betydeligt: Det forenkler databehandlingen (Fig. 4.1-1, Fig. 4.1-2) og fremskynder valideringsprocessen ved at gøre kravene klare og gennemsigtige, som vi allerede har diskuteret i de foregående kapitler. På samme måde letter oversættelsen af CAD-data (BIM) til en struktureret åben form (fig. 4.1-12, fig. 4.1-13) attributgrupperingsprocessen og QTO-processen.

QTO attributtabellen har en struktureret form, så når vi bruger strukturerede CAD data, arbejder vi med en enkelt datamodel (figur 5.2-5), hvilket eliminerer behovet for at konvertere og oversætte projektdatamodeller og grupperingsregler til en fællesnævner. Det giver os mulighed for at gruppere data efter en eller flere attributter med blot en enkelt kodelinje. I modsætning hertil kræver behandling i åben BIM og lukket BIM, hvor data lagres i semistrukturerede, parametriske eller lukkede formater, snesevis eller endda hundredvis af kodelinjer og brug af API til at interagere med geometri- og attributoplysninger.

- Eksempel på gruppering af et QTO-struktureret projekt efter en attribut. Tekstforespørgsel i et hvilket som helst LLM chatrum (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller et andet):

  > Jeg har CAD -projekt som DataFrame - filtrer venligst projektdataene for at få elementer med "Type"-parameteren, der kun indeholder "Type 1"⏎

- LLM's svar vil højst sandsynligt være i form af Python-kode ved hjælp af Pandas :



Fig. 5.2-6 En kodelinje skrevet med LLM gør det muligt at gruppere et helt CAD -projekt efter "Type"-attributten og få den ønskede gruppe af elementer.

Takket være den enkle struktur i den todimensionelle DataFrame behøver vi ikke at forklare LLM -skemaet og datamodellen, hvilket forkorter fortolkningstrinnene og fremskynder oprettelsen af de endelige løsninger. Tidligere krævede det at lære programmeringssprog at skrive selv simpel kode, men nu giver moderne sprogmodeller (LLM'er) os mulighed for automatisk at oversætte proceslogik til kode, når vi arbejder med strukturerede data ved hjælp af tekstforespørgsler.

> LLM automatisering og sprogmodeller kan helt fjerne behovet for, at fagfolk, der arbejder med CAD (BIM) datagruppering og -behandling, skal lære programmeringssprog eller BIM-værktøjer ved at give mulighed for at løse problemer ved hjælp af tekstforespørgsler.

Den samme forespørgsel - gruppering af alle projektelementer fra kategorien "vægge" og beregning af mængder for hver type (fig. 5.2-5) - som i et CAD-miljø (BIM) kræver 17 klik i grænsefladen eller skrivning af 40 linjer kode, ser i åbne databehandlingsværktøjer (f.eks. SQL eller Pandas) ud som en enkel og intuitiv forespørgsel:

- Med en enkelt linje i Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

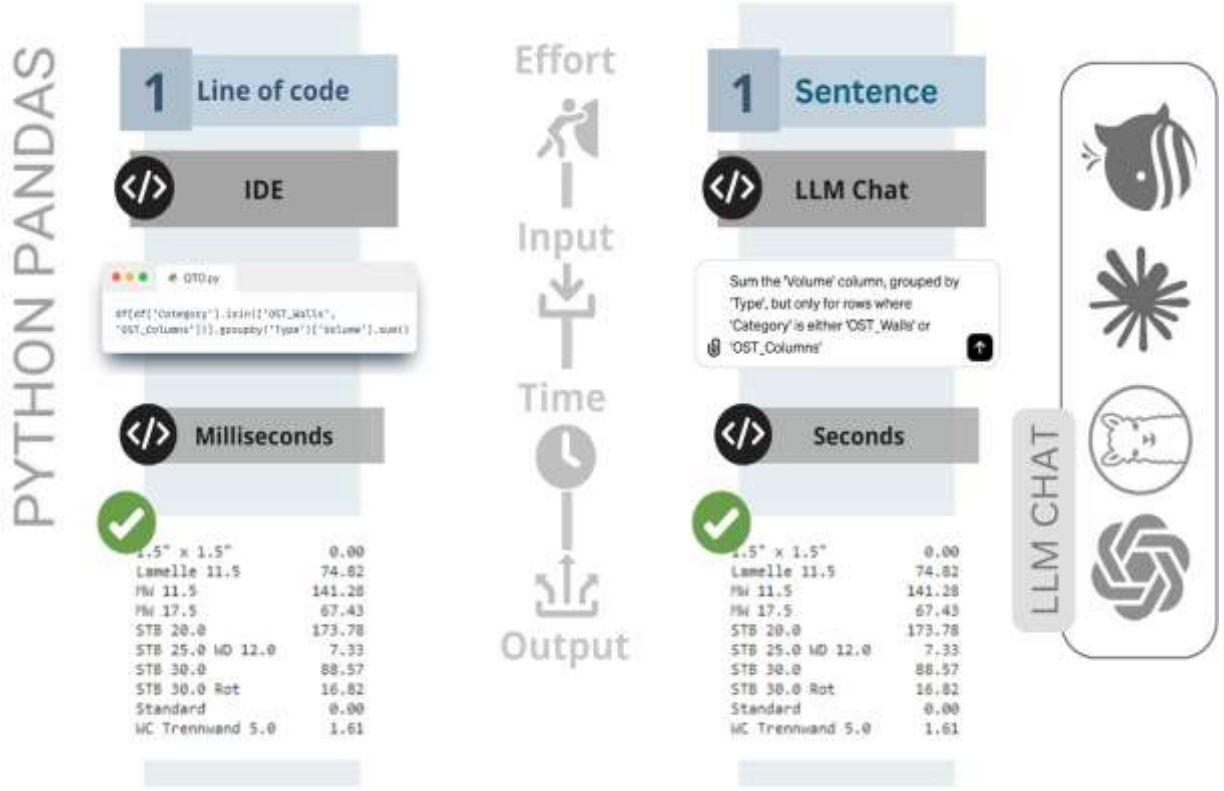
Kodeafkodning: tag fra df (DataFrame) de elementer, hvis attribut-kolonne "Category" har værdien "OST_Walls", gruppér alle de opnåede elementer efter attribut-kolonnen "Type", og opsummer for den opnåede gruppe af elementer attributten "Volume".

- Gruppering af et struktureret projekt hentet fra CAD ved hjælp af SQL:

```
SELECT Type, SUM(Volume) AS TotalVolume
FRA elementer
WHERE Category = 'OST_Walls'
GROUP BY Type;
```

- Ved hjælp af LLM kan vi skrive en grupperingsanmodning til projektdatabasen som en simpel tekstreference - en prompt (fig. 5.2-7):

> For projektdatarammen skal du gruppere elementerne efter parameteren "Type", men kun for elementer med parameteren "Kategori", der er lig med "OST_Walls" eller "OST_Columns", og opsummere kolonneparameteren "Volume" for den resulterende ⏎



Figur 5.2-7 Ved hjælp af SQL, Pandas og LLM er det nu muligt at automatisere databehandling med nogle få linjer kode og tekstforespørgsler.

Indhentning af QTO fra CAD-data ved hjælp af LLM-værktøjer (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) ændrer drastisk de traditionelle metoder til at udtrække attributinformation, kvantitative og volumetriske data for individuelle objekter og grupper af objekter.

Nu kan selv projektledere, omkostnings- eller logistikspecialister, der ikke har et dybt kendskab til design og ikke har specialiseret CAD-software - (BIM-) leverandører, der har adgang til CAD-databasen, få den samlede mængde af elementer i kategorien vægge eller andre objekter i løbet af få sekunder, blot ved at skrive eller diktere en forespørgsel.

I tekstforespørgsler (Fig. 5.2-8) behandler LLM-agenten i modellen brugerens anmodning om at anvende en bestemt funktion på en eller flere parametre - kolonner i tabellen. Som følge heraf modtager brugeren i kommunikation med LLM enten en ny kolonneparameter med nye værdier eller en specifik værdi efter gruppering

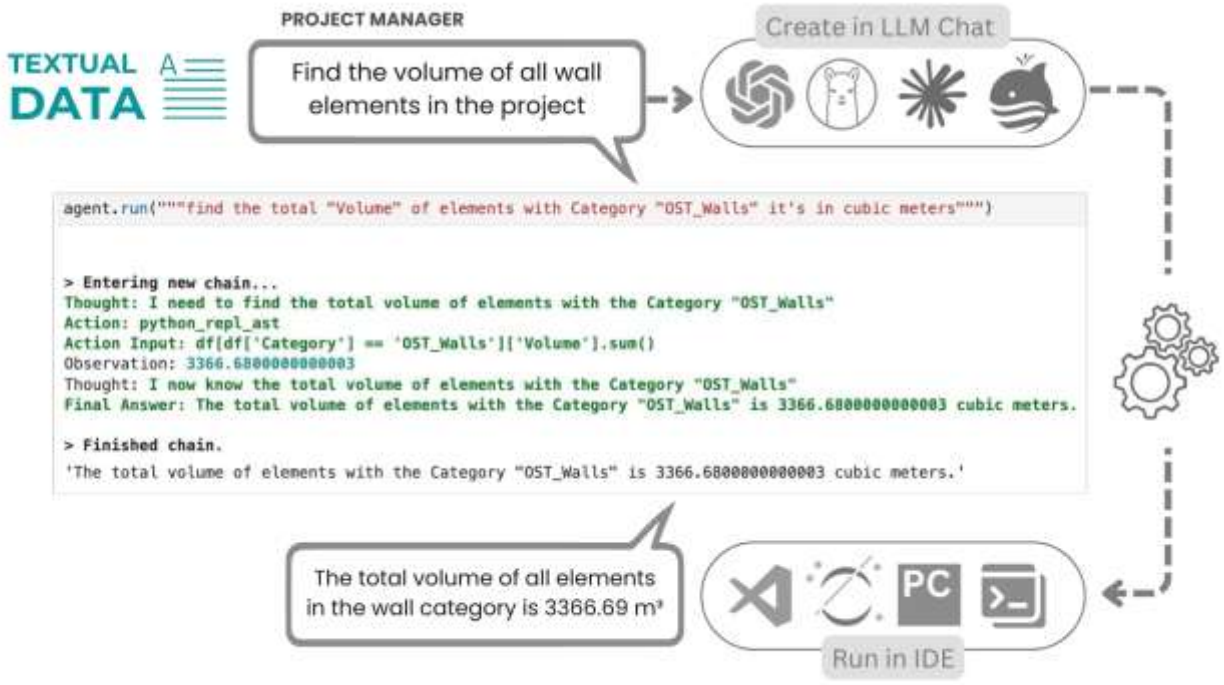


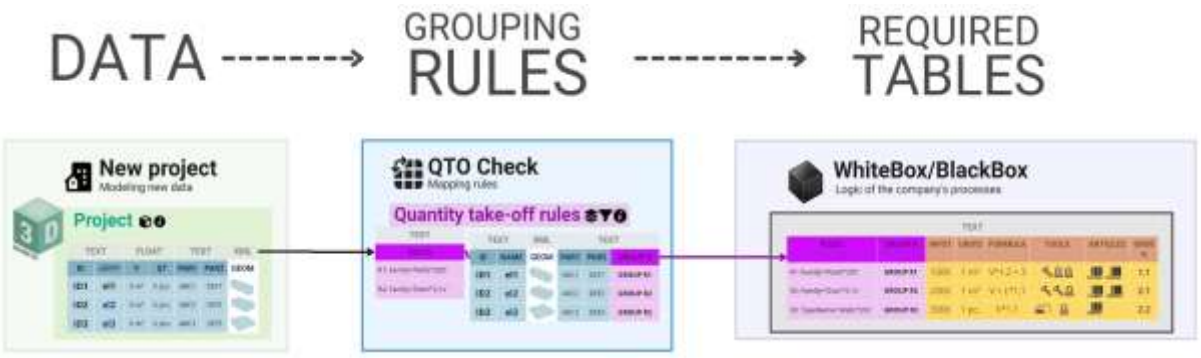
Fig. 5.2-8 LLM model, der arbejder med strukturerede data, forstår ud fra konteksten i en tekstforespørgsel, hvilken gruppering og hvilke attributter brugeren spørger om.

Hvis det kun er nødvendigt at få mængder for én gruppe af elementer, er det tilstrækkeligt at udføre en simpel QTO-forespørgsel (figur 5.2-7) på CAD-modeldataene. Men når man beregner et budget eller et overslag for et helt projekt, der består af mange grupper af elementer, er det ofte nødvendigt at udtrække mængder for alle typer af elementer (klasser0 , hvor hver kategori af elementer behandles separat - med gruppering efter relevante attributter.

I praksis bruger estimatorer og vurderingsfolk individuelle grupperings- og beregningsregler for forskellige typer af objekter. For eksempel grupperes vinduer normalt efter etager eller zoner (grupperingsparameter - attribut Level, Rooms), og vægge - efter materiale eller konstruktionstype (parameter Material, Type). For at automatisere grupperingsprocessen er sådanne regler beskrevet på forhånd i i form af grupperingsregeltabeller. Disse tabeller fungerer som konfigurationsskabeloner, der definerer, hvilke attributter der skal bruges i beregninger for hver gruppe af elementer i projektet.

QTO beregning af hele projektet ved hjælp af grupperegler fra et Excel-regneark

I virkelige byggeprojekter er det ofte nødvendigt at udføre aggregering af flere attributter samtidig inden for en gruppe af elementer. Når man f.eks. arbejder med kategorien "Vinduer" (hvor attributten Kategori indeholder værdier som OST_Windows eller IfcWindows), kan elementer ikke kun grupperes efter type - f.eks. efter værdien i feltet Typenavn eller Type - men også efter yderligere egenskaber, som f.eks. det termiske konduktivitetsniveau, der er angivet i den tilsvarende attribut. Denne flerdimensionelle gruppering giver mulighed for mere præcise resultater for en bestemt gruppe. På samme måde kan man ved beregning af væg- eller gulvkategorier bruge vilkårlige kombinationer af attributter - såsom materiale, niveau, gulv, brandmodstand og andre parametre - som filtre eller grupperingskriterier (figur 5.2-9).



Figur 5.2-9 For hver gruppe eller kategori af enheder i et projekt er der en anden grupperingsformel, som består af et eller flere kriterier.

Processen med at definere sådanne grupperingsregler svarer til processen med at oprette datakrav, der er beskrevet i kapitlet "Oprettelse af krav og kvalitetskontrol data " (Figur 4.4-5), hvor vi diskuterede arbejdet med datamodeller i detaljer. Sådanne grupperings- og beregningsregler sikrer nøjagtigheden og relevansen af resultaterne for automatisk at beregne de samlede attributter for mængden eller volumen af en kategori af enheder under hensyntagen til alle de nødvendige betingelser, der skal tages højde for i beregninger og udregninger.

- Følgende kodeeksempel filtrerer projekttabellen, så det resulterende datasæt kun indeholder enheder, hvor attributkolonnen "Category" indeholder værdierne "OST_Windows" eller "IfcWindows", og hvor attributkolonnen "Type" samtidig indeholder værdien "Type 1":

> Jeg har et DataFrame -projekt - filtrer dataene, så kun elementer, der har attributten "Category", der indeholder værdierne "OST_Windows" eller "IfcWindows", og samtidig indeholder Type-attributten værdien "Type 1, forbliver i datasættet. ⏎

LLM's svar :



Figur 5.2-10 Med en enkelt kodelinje, der ligner Excel-formlen , kan du gruppere alle projektenheder efter flere attributter.

Den resulterende kode (fig. 5.2-10) efter oversættelse af CAD-data i strukturerede åbne formater (fig. 4.1-13) kan køres i et af de populære IDE'er (integrerede udviklingsmiljøer), vi nævnte ovenfor, i offline-tilstand: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære online-værktøjer: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

For at hente projektenhederne i QTO DataFrame-formularen under kategorien "Windows" kun med en specifik varmeledningsevneværdi, kan vi bruge følgende forespørgsel til LLM:

> Jeg har et DataFrame -projekt - filtrer dataene, så kun poster med "Kategori", der indeholder "OST_Windows" eller "IfcWindows"-værdier, forbliver i datasættet, og samtidig skal kolonnen ThermalConductivity have en værdi på 0,⏎

LLM's svar :



Figur 5.2-11 Det ekstremt enkle Pandas-forespørgselssprog Python giver dig mulighed for at køre QTO'er for et vilkårligt antal projekter samtidigt.

I svaret fra LLM (Fig. 5.2-11) bruges den logiske betingelse "&" til at kombinere to kriterier:

varmeledningsevne og tilhørsforhold til en af de to kategorier. Metoden "isin" kontrollerer, om værdien af attribut-kolonnen "Category" er indeholdt i den angivne liste.

I projekter med et stort antal elementgrupper med forskellig grupperingslogik - for hver kategori af projektenheder (f.eks. vinduer, døre, plader) skal der fastlægges individuelle grupperingsregler, som kan omfatte yderligere koefficienter eller samlede attributberegningsformler. Disse formler (figur 5.2-12 attribut "formel", f.eks. x-værdi af mængde og y-volumen af gruppe) og koefficienter tager f.eks. højde for hver gruppes unikke egenskaber:

- % tillæg til materialemængden for at tage højde for overskridelser
- fast ekstra mængde materiale
- justeringer i forbindelse med mulige risici og beregningsfejl i form af formler

Når filtrerings- og grupperingsreglerne er formuleret i form af parameterformler for hver emnekategori, kan de gemmes som en linje-for-linje-tabel - for eksempel i Excel-format (figur 5.2-12). Ved at gemme disse regler i en struktureret form kan processen med at udtrække, filtrere og gruppere projektdata automatiseres fuldt ud. I stedet for manuelt at skrive mange separate forespørgsler læser systemet blot parametertabellen og anvender de relevante regler på modellen (den overordnede projektdataramme (fig. 4.1-13)) og genererer de endelige QTO-tabeller for hver kategori af projektelementer.



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

Figur 5.2-12 QTO Attribute Grouping Table opstiller regler for gruppering af projektelementer, så man får et nøjagtigt samlet antal og omfang for hver kategori.

De indsamlede regler giver dig mulighed for at gruppere hele projektet og udføre alle nødvendige beregninger, herunder justeringer af volumenattributter. Som følge heraf bringes volumenerne til den "virkelige volumen", som bruges til beregninger og udregninger, og ikke dem, der oprindeligt var på designstadiet i CAD-modellen.

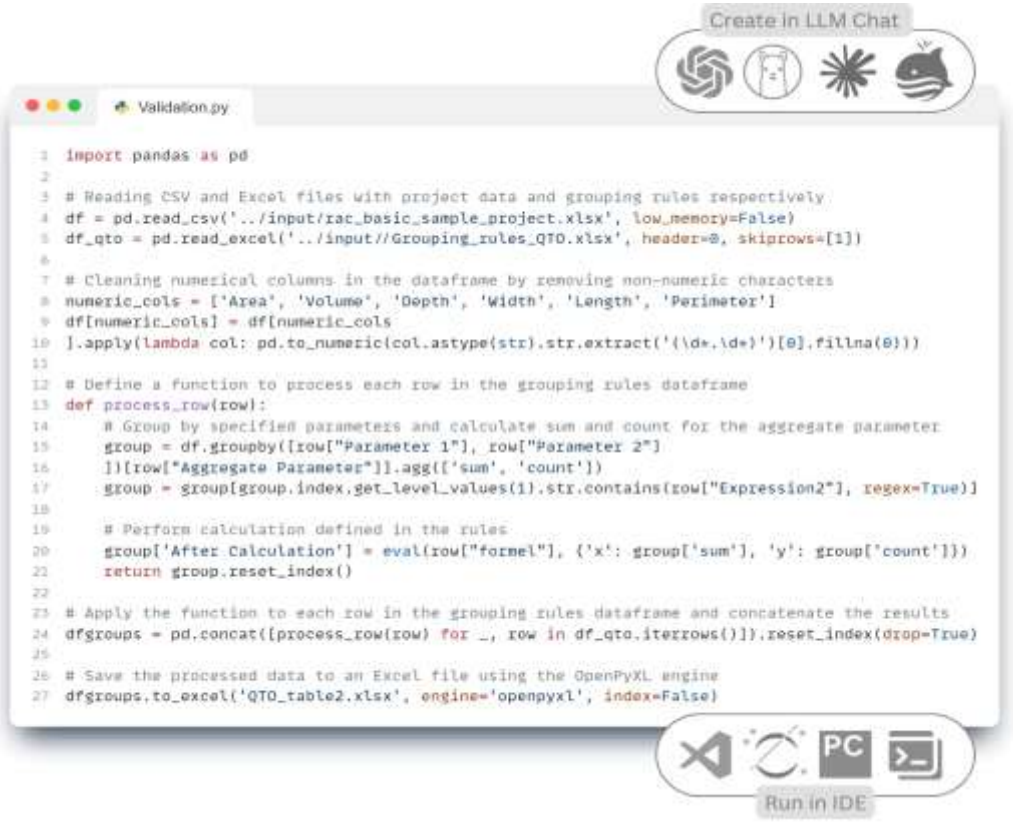
> I processen med automatisk at oprette QTO volumentabeller for hele projektet skal programmet gennemgå alle kategorierne i tabellen med grupperingsregler, tage grupperingsattributterne, gruppere alle projektelementerne efter dem og aggregere volumenattributten for den gruppe og desuden gange den med en raffineringsfaktor eller -koefficient.

Lad os bede LLM om at skrive kode til os til en sådan løsning, hvor koden skal indlæse to tabeller - en tabel med grupperingsregler (figur 5.2-12) og en tabel med data fra selve projektet (figur 4.1-13), og derefter anvende grupperingsreglerne, gruppere elementerne i henhold til de givne regler, beregne de aggregerede værdier og gemme resultaterne i en ny Excel-fil.

- Send en tekstanmodning til chatten på LLM:

> Jeg har brug for kode til at læse projektdata fra filen 'basic_sample_project.xlsx' og derefter regler fra 'Grouping_rules_QTO.xlsx' for at gruppere alle data efter 'Parameter 1' og 'Parameter 2', aggregere 'Aggregate Parameter', filtrere efter 'Expression2', udføre beregninger fra 'Formel1' og gemme QTO-tabellen i 'QTQ_table2.xlsx. ⏎

LLM's svar :



Figur 5.2-13 Python-kode Pandas behandler designdata fra et CAD-projekt og grupperer dem i henhold til regler, der er defineret i Excel .

Det endelige resultat af kodekørslen (fig. 5.2-13) vil være en entitetsgruppetabel, der ikke kun indeholder de generaliserede volumenattributter fra den oprindelige CAD- (BIM-) model, men også en ny reel volumenattribut, der tager højde for alle krav til korrekt oprettelse af beregninger og estimater (eksempel fig. 5.2-14).

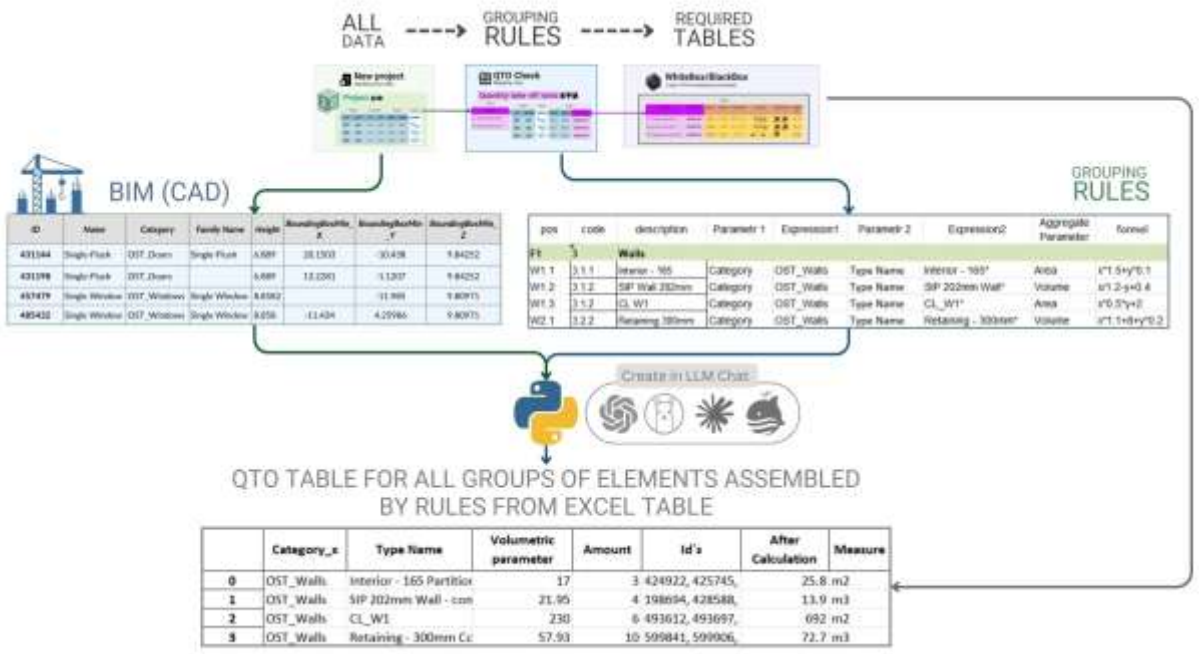
QTO TABLE FOR ALL GROUPS OF ELEMENTS
ASSEMBLED BY RULES FROM EXCEL TABLE

	Category_x	Type Name	Volumetric parameter	Amount	Id´s	After Calculation	Measure
0	OST_Walls	Interior - 165 Partitior	17	3	424922, 425745,	25.8	m2
1	OST_Walls	SIP 202mm Wall - con	21.95	4	198694, 428588,	13.9	m3
2	OST_Walls	CL_W1	230	6	493612, 493697,	692	m2
3	OST_Walls	Retaining - 300mm Cc	57.93	10	599841, 599906,	72.7	m3

Figur 5.2-14 Attributten "After Calculation" tilføjes til oversigtstabellen, efter at der er udført kode, som automatisk beregner den faktiske mængde.

Den resulterende kode (fig. 5.2-13) kan køres i et af de populære IDE'er (som vi nævnte ovenfor) og

anvendes på et hvilket som helst antal eksisterende eller nye indgående projekter (RVT, IFC, DWG, NWS, DGN osv.), hvad enten det drejer sig om nogle få projekter eller måske hundredvis af projekter i forskellige formater i en struktureret form (fig. 5.2-15).



Figur 5.2-15 Den automatiske grupperingsproces for byggedata forbinder BIM-data (CAD) med QTO-tabeller via regler fra et Excel-regneark.

Den tilpassede og parametriserede proces til indsamling af volumetriske data (figur 5.2-15) muliggør fuldautomatisk indsamling af data om kvantitative egenskaber og mængder af projektelementer til videre arbejde med dem, herunder omkostningsestimering, logistik, arbejdsplaner og beregning af CO2-fodaftryk og andre analytiske opgaver.

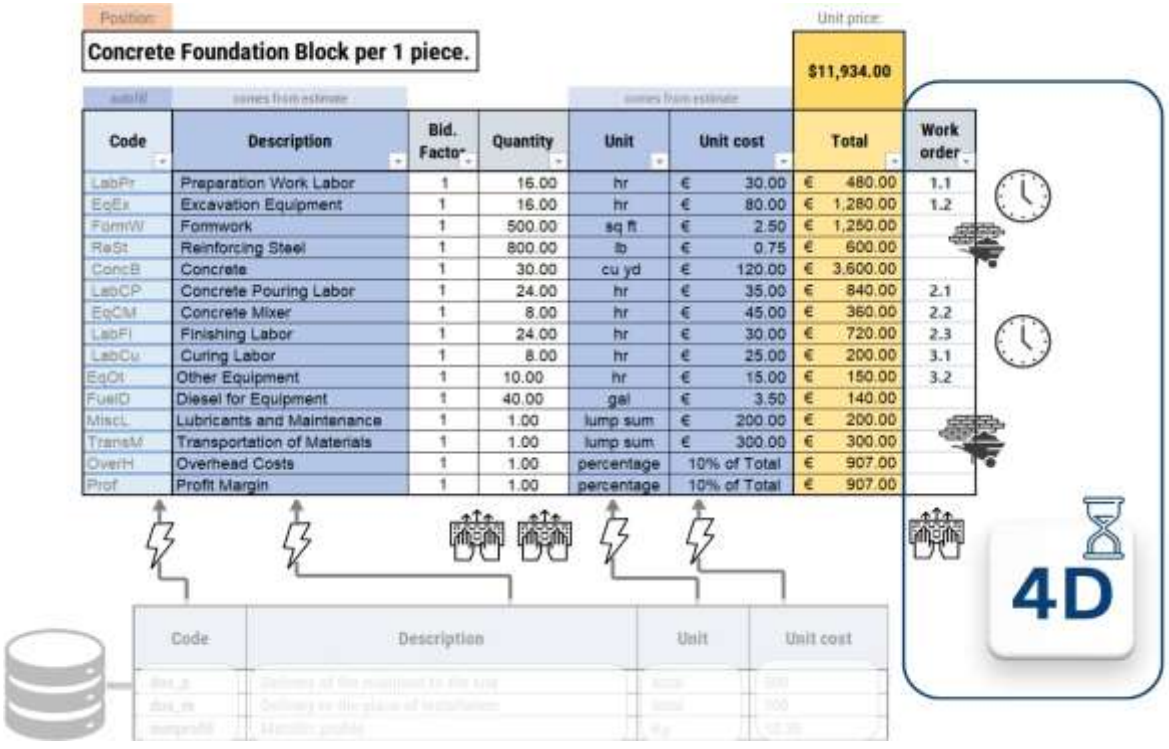
Efter at have lært værktøjer, der giver os mulighed for nemt at organisere og gruppere grupper af projektelementer i henhold til bestemte attributter, er vi nu klar til at integrere grupperede og filtrerede projekter med forskellige virksomhedsberegninger og forretningsscenarier.

KAPITEL 5.3.
4D, 6D -8D OG BEREGNING AF KULDIOXIDEMISSIONER CO_2

4D-model: Integration af tid i byggeoverslag

Ud over omkostningsberegning er en af de vigtigste anvendelser af designdata i byggeriet fastlæggelsen af tidsparametre - både for individuelle byggeoperationer og for hele projektet. Den ressourcebaserede estimeringsmetode og den tilhørende beregningsdatabase, som blev diskuteret i detaljer i det foregående kapitel "Beregninger og estimater for byggeprojekter", bruges ofte som grundlag for automatiseret beregning af tid og oprettelse af en tidsplan for arbejdets udførelse.

Den ressourcebaserede tilgang tager ikke kun hensyn til materialeomkostninger, men også til tidsressourcer. I omkostningsberegningen kan hver proces tildeles en arbejdsordreattribut (figur 5.3-1 - Arbejdsordreparameter) og den mængde tid og omkostninger, der er forbundet med udførelsen af den pågældende proces. Disse parametre er særligt vigtige for at beskrive operationer, der ikke har en fast markedspris og ikke kan købes direkte - som f.eks. brug af byggematerialer, ansættelse af arbejdere eller logistiske processer (som normalt udtrykkes i timer). I sådanne tilfælde bestemmes omkostningerne ikke af indkøbsafdelingen, men direkte af den udførende virksomhed baseret på interne normer eller produktionsrater (figur 5.3-1).



Code	Description	Bid. Factor	Quantity	Unit	Unit cost	Total	Work order
LabPr	Preparation Work Labor	1	16.00	hr	€ 30.00	€ 480.00	1.1
EqEx	Excavation Equipment	1	16.00	hr	€ 80.00	€ 1,280.00	1.2
FormW	Formwork	1	500.00	sq ft	€ 2.50	€ 1,250.00	
ReSt	Reinforcing Steel	1	800.00	lb	€ 0.75	€ 600.00	
ConcB	Concrete	1	30.00	cu yd	€ 120.00	€ 3,600.00	
LabCP	Concrete Pouring Labor	1	24.00	hr	€ 35.00	€ 840.00	2.1
EqCM	Concrete Mixer	1	8.00	hr	€ 45.00	€ 360.00	2.2
LabFi	Finishing Labor	1	24.00	hr	€ 30.00	€ 720.00	2.3
LabCu	Curing Labor	1	8.00	hr	€ 25.00	€ 200.00	3.1
EqOt	Other Equipment	1	10.00	hr	€ 15.00	€ 150.00	3.2
FuelD	Diesel for Equipment	1	40.00	gal	€ 3.50	€ 140.00	
MiscL	Lubricants and Maintenance	1	1.00	lump sum	€ 200.00	€ 200.00	
TransM	Transportation of Materials	1	1.00	lump sum	€ 300.00	€ 300.00	
OverH	Overhead Costs	1	1.00	percentage	10% of Total	€ 907.00	
Prof	Profit Margin	1	1.00	percentage	10% of Total	€ 907.00	

Fig. 5.3-1 Beregninger af arbejde i den ressourcebaserede estimeringsmetode omfatter timing af arbejdstimer.

Beregninger på omkostningsniveau omfatter således ikke kun brændstof- og materialeomkostninger (indkøbsomkostninger), men også den tid, som chauffører, teknikere og hjælpearbejdere bruger på stedet.

I det givne eksempel (fig. 5.3-1) er omkostningstabellen en beregning af omkostningerne ved at installere en fundamentblok, herunder arbejdets bestanddele, såsom forberedelse, rammeinstallation og betonstøbning, samt de materialer og den arbejdskraft, der kræves. Men enkelte operationer, som f.eks. forberedelsesarbejdet, har måske ingen materialeomkostninger, men kan indeholde betydelige midlertidige arbejdsomkostninger udtrykt i mandetimer.

For at planlægge arbejdssekvensen (for arbejdsplanen) på byggepladsen tilføjes attributten "Arbejdsordre" manuelt til beregningstabellen (fig. 5.3-1). Denne attribut er kun specificeret i en ekstra kolonne for emner, hvis måleenhed er udtrykt i tid (time, dag). Denne attribut er et supplement til arbejdskode, beskrivelse, mængde, måleenhed (parameter "Unit") og omkostninger. Den numeriske rækkefølge (parameter "Arbejdsordre") for aktiviteter gør det muligt at fastlægge den rækkefølge, som opgaverne udføres i på byggepladsen, og bruge den til planlægning.

Byggeplan og dens automatisering baseret på omkostningsdata

Byggeplanen er en visuel fremstilling af den arbejdsplan og de processer, der skal udføres som en del af projektgennemførelsen. Den oprettes på grundlag af detaljerede ressourceberegninger (figur 5.3-1), hvor hver opgave/job ud over ressourceomkostningerne er planlagt efter tid og rækkefølge.

I modsætning til gennemsnitsmetoder, hvor tidsestimater er baseret på typiske timer til installation af materialer eller udstyr, er planlægningen i den ressourcebaserede metode baseret på faktiske data i omkostningsberegningen. Hver post i estimatet relateret til lønomkostninger er baseret på den anvendte kalender, som tager højde for de faktiske betingelser for ressourceanvendelse i arbejdsperioden. Justering af produktive timer gennem koefficienter på omkostningsniveau (Fig. 5.3-1 parameter "Bid. Factor") gør det muligt at tage højde for forskelle i produktivitet og sæsonbestemte særegenheder, der påvirker arbejdets timing.

For at bestemme processens start- og slutdatoer for byggeplanen i Gantt-diagrammet tager vi værdierne for tidsattributten for hvert element fra fundamentblokkalkulationen og ganger dem med antallet af blokke (i dette tilfælde antallet af betonfundamentblokke). Denne beregning giver varigheden af hver opgave. Vi plotter derefter disse varigheder ind på en tidslinje fra projektets startdato for at skabe en graf, og resultatet er en visuel repræsentation, der viser, hvornår hver opgave skal starte og slutte. Parameteren "Arbejdsordre" for processer giver os desuden mulighed for at forstå, om arbejdsprocessen kører parallelt ("Arbejdsordre", f.eks. 1.1-1.1) eller sekventielt (1.1-1.2).

> **Gantt Chart** er et grafisk værktøj til planlægning og projektstyring, der repræsenterer opgaver som vandrette søjler på en tidslinje. Hver bjælke repræsenterer opgavens varighed, dens start og slutning.

En tidsplan eller et Gantt-diagram hjælper projektledere og medarbejdere med klart at forstå, hvornår og i hvilken rækkefølge de forskellige faser af byggeriet skal udføres, hvilket sikrer effektiv brug af ressourcer og overholdelse af deadlines.

Lad os præsentere en tidsplan for installationen af tre betonfundamentblokke ved hjælp af beregningerne fra tabellen ovenfor. Brug omkostningstabellen (fig. 5.3-1) fra eksemplet ovenfor og bed LLM om at planlægge installationen af de 3 fundamentblokke, f.eks. til den første maj 2024.

For at sende en kalkulation til LLM kan vi uploade kalkulationstabellen i XLSX-format eller blot indsætte et

skærmbillede af et JPEG-billede af kalkulationen direkte i LLM-chatten (fig. 5.3-1). LLM vil selv finde et bibliotek til at visualisere tabelbilledet, og ved at gange tidsattributterne for aktiviteterne fra tabellen med deres volumen, vil det tilføje alle data til en graf.

Send en tekstanmodning til LLM:

> Opret en arbejdsplan i form af et Gantt-diagram ved hjælp af værdierne fra den tidsrelaterede tabel (fig. 5.3-1 som JPEG i bilaget) til installation af 3 fundamentblokke. Arbejdet for hver blok skal udføres sekventielt. Arbejdet skal påbegyndes den 01/05/2024. ⏎

LLM's svar

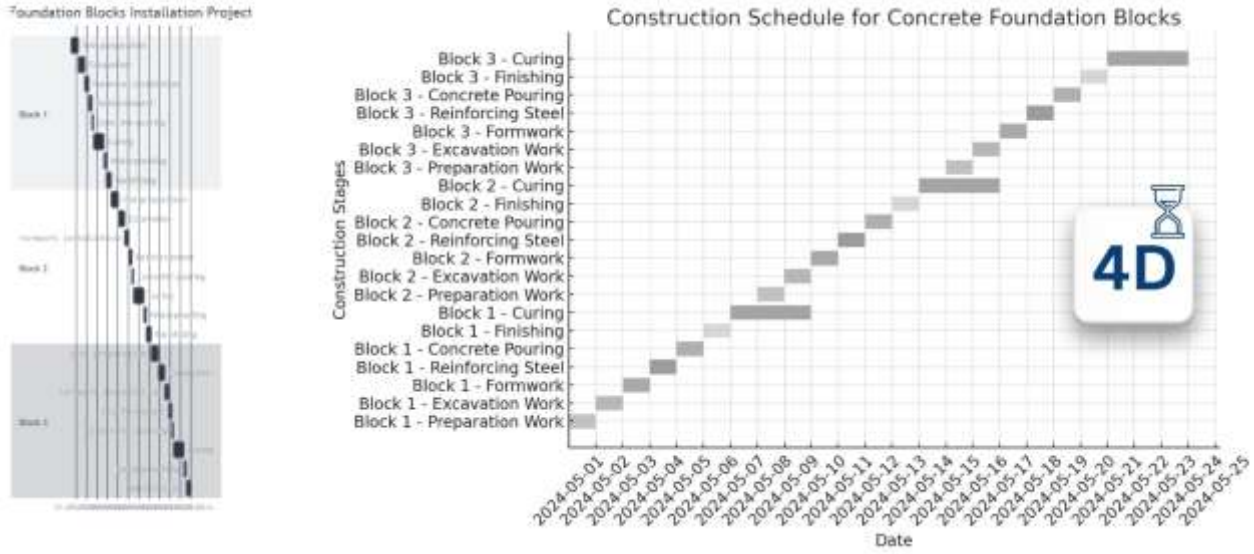


Fig. 5.3-2 Automatisk genereret af flere LLM'Gantt-diagrammet viser faserne i konstruktionen af tre betonblokke i henhold til betingelserne fra prompten.

Den resulterende graf (fig. 5.3-2) er et tidsdiagram, hvor hver vandret bjælke svarer til en bestemt fase af arbejdet på fundamentblokken og viser rækkefølgen af operationer (parameter "Arbejdsordre"), såsom forberedelse, udgravning, forskallingsinstallation, armering, betonstøbning og efterbehandling„ dvs. de processer, der har udfyldt tidsparametre og rækkefølge i beregningerne.

En sådan tidsplan (fig. 5.3-2) tager ikke hensyn til begrænsninger i forbindelse med arbejdsdage, skift eller arbejdstidsstandarder, men er udelukkende beregnet til konceptuel visualisering af processen. En nøjagtig tidsplan, der afspejler arbejdets parallelitet, kan suppleres med passende promts eller yderligere instruktioner i chatrummet

Ved hjælp af en enkelt kalkulation (fig. 5.3-1), takket være volumenattributter fra 3D -geometri, er det muligt automatisk at estimere både projektomkostningerne via automatiserede -estimater og samtidig at beregne gruppernes tidskarakteristika i form af tabeller eller grafer for forskellige projektvarianter (fig. 5.3-3).

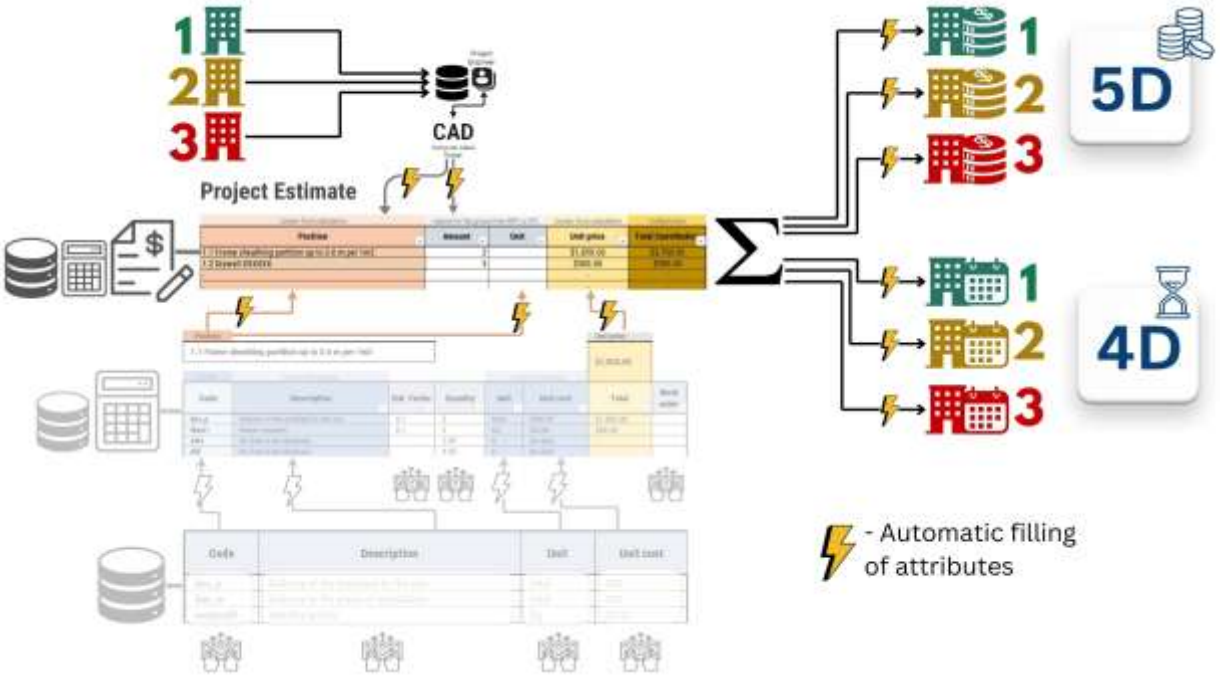


Fig. 5.3-3 Automatisk beregning, giver mulighed for øjeblikkelig og automatisk omkostnings- og tidsprognose for forskellige projektmuligheder.

Moderne modulære ERP -systemer (fig. 5.4-4), der indlæser data fra CAD -modeller, bruger lignende automatiserede tidsberegningsmetoder, der reducerer beslutningsprocessen betydeligt. Det giver dig mulighed for øjeblikkeligt og præcist at planlægge arbejdsplaner og beregne den samlede tid, der kræves for at fuldføre alle opgaver i projektgennemførelsen, under hensyntagen til reelle priser.

Udvidede attributlag 6D -8D: fra energieffektivitet til sikkerhedssikring

6D, 7D og 8D er udvidede niveauer af informationsmodellering, som hver især bidrager med yderligere lag af attributter til den omfattende projektinformationsmodel, hvis grundlag er attributterne i 3D -modellen med deres antal og omfang. Hvert ekstra lag introducerer specifikke parametre, der er nødvendige for yderligere gruppering eller yderligere identifikation i andre systemer, som f.eks. ejendomsadministrationssystemer (PMS), computerstøttet facilitetsstyring (CAFM), byggeprojektstyring (CPM) og sikkerhedsstyringssystemer (SMS).



Figur 5.3-4 Attributterne 6D, 7D og 8D i datainformationsmodellen udvider overvejelserne om forskellige aspekter af projektet, fra energieffektivitet til sikkerhed.

- I **6D** tilføjes ud over projektdatabasen (eller datarammen (figur 4.1-13)) med geometriske og volumetriske attributter for elementer også oplysninger (attributkolonner) om miljømæssig bæredygtighed. Dette omfatter oplysninger om energieffektivitet, CO2-fodaftryk, genanvendelighed af materialer og brug af miljøvenlige teknologier. Disse data gør det muligt at vurdere projektets miljøpåvirkning, optimere projektbeslutninger og nå målene for bæredygtig udvikling (ESG).
- **7D** -attributter supplerer de attributter, der kræves for at styre bygningsvedligeholdelsen. Det drejer sig om data om vedligeholdelsesplaner, komponenternes livscyklus, teknisk dokumentation og reparationshistorik. Dette sæt af oplysninger sikrer, at modellen kan integreres med vedligeholdelsessystemer (CAFM, AMS), giver mulighed for effektiv planlægning af vedligeholdelse, udskiftning af udstyr og giver støtte gennem hele anlæggets livscyklus.
- **8D** ekstra attributlag, - indeholder oplysninger om sikkerhed - både i byggefasen og under den efterfølgende drift. Modellen omfatter personalets sikkerhedsforanstaltninger, nødinstruktioner, evakueringssystemer og brandbeskyttelseskrav. Integrationen af disse data i den digitale model hjælper med at tage højde for risici på forhånd og udvikle arkitektoniske, tekniske og organisatoriske løsninger, der tager højde for sundheds- og sikkerhedskrav.

I struktureret tabelform repræsenterer lag 4D til 8D yderligere attributter i form af kolonner med udfyldte værdier (figur 5.3-5), der føjes til de allerede udfyldte 3D -modelattributter såsom navn, kategori, type og volumetriske egenskaber. Værdierne i attributlag 6D, 7D og 8D indeholder yderligere tekstlige og numeriske data som f.eks. genbrugsprocent, CO2-fodaftryk, garantiperiode, udskiftningscyklus, installationsdato, sikkerhedsprotokoller osv.



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

Figur 5.3-5 6D -8D tilføjer attributlag til datainformationsmodellen, som allerede indeholder geometriske og volumetriske attributter fra 3D -modellen.

For vores nye vindue (figur 4.4-1) kan elementet med identifikatoren W-NEW (figur 5.3-5) have følgende 3D -8D attributter:

3D -attributter - geometrisk information hentet fra CAD systemer:

- "Typenavn" - element "Vindue"
- "Bredde" - 120 cm
- Derudover kan du tilføje "Bounding Box "-punkterne for et element eller dets "geometri BREP / MESH " som en separat attribut

Egenskaber ved 6D - miljømæssig bæredygtighed:

- Genbrugsprocent på 90 procent
- "Carbon footprint - 1.622 kg CO_2

Attributter 7D - objektstyringsdata:

- "Garantiperiode" - 8 år
- "Replacement Cycle" er 20 år gammel
- "Vedligeholdelse" - kræves hvert år

Egenskaber ved 8Dat sørge for sikker brug og drift af bygninger:

- Vinduet "Installeret" - af firmaet "XYZ Windows
- "Sikkerhedsstandard" - i overensstemmelse med ISO 45001

Alle parametre, der er registreret i en database eller et datasæt (fig. 5.3-5), er nødvendige for specialister i forskellige afdelinger til gruppering, søgning eller beregninger. En sådan flerdimensionel attributbaseret beskrivelse af projektobjekter giver et komplet billede af deres livscyklus, driftskrav og mange andre aspekter, der er nødvendige for projektdesign, konstruktion og drift.

Estimering af CO_2 og beregning af kuldioxidudledning fra byggeprojekter

Sideløbende med emnet bæredygtighed i byggeprojekter i fase 6D (figur 5.3-5) fokuserer moderne byggeri på projekters miljømæssige bæredygtighed, hvor et af de vigtigste aspekter bliver vurdering og minimering af kuldioxid CO_2 emissioner, der opstår i løbet af et projekts livscyklusfaser (f.eks. fremstilling og installation).

> Estimering og beregning af kulstofemissioner fra byggematerialer er en proces, hvor de samlede kulstofemissioner bestemmes ved at gange de volumetriske egenskaber for et element eller en gruppe af elementer, der anvendes i et projekt, med en passende kulstofemissionsfaktor for kategorien.

At tage højde for kulstofemissioner i vurderingen af byggeprojekter som en del af de bredere ESG kriterier (miljø, sociale forhold og ledelse) tilføjer et nyt niveau af kompleksitet til analysen. Det er især vigtigt for bygherren og investoren, når de skal opnå en relevant certificering som LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) eller DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Opnåelse af en af disse certificeringer kan øge en ejendoms salgbarhed betydeligt, forenkle ibrugtagning og sikre overholdelse af kravene fra bæredygtighedsorienterede lejere (ESG). Afhængigt af projektkravene kan HQE (Haute Qualité Environnementale, den franske standard for grønt byggeri), WELL (WELL Building Standard, med fokus på brugernes sundhed og komfort) GRESB (Global Real Estate Sustainability Benchmark) også anvendes

> Miljø, sociale forhold og god selskabsledelse **ESG** (miljø, sociale forhold og god selskabsledelse) er et bredt sæt af principper, der kan bruges til at vurdere en virksomheds selskabsledelse, sociale og miljømæssige påvirkning både internt og eksternt.

ESG, der oprindeligt blev udviklet i begyndelsen af 2000'erne af finansielle fonde for at give investorer oplysninger om brede miljømæssige, sociale og ledelsesmæssige kriterier, har udviklet sig til en nøgleindikator for evaluering af både virksomheder og projekter, herunder byggeprojekter. Ifølge undersøgelser foretaget af store konsulentfirmaer er miljømæssige, sociale og ledelsesmæssige overvejelser (ESG) ved at blive en integreret del af byggebranchen.

> Ifølge EY (2023) "The Path to Carbon Neutrality" reducerer virksomheder, der aktivt implementerer ESG -principper, ikke kun de langsigtede risici, men øger også effektiviteten af deres forretningsmodeller, hvilket er særligt vigtigt i den globale transformation af markederne [103]. PwC's ESG Awareness-rapport bemærker, at virksomhedernes bevidsthed om betydningen af ESG -faktorer spænder fra 67% til 97%, og at de fleste organisationer ser disse tendenser som nøglen til fremtidig bæredygtighed [104], og at virksomhederne for det meste oplever et betydeligt pres fra interessenter for at integrere ESG-principper.

> Integrationen af ESG -principper i byggeprojekter bidrager således ikke kun til at opnå internationale bæredygtighedscertifikater som LEED, BREEAM, DGNB, men sikrer også den langsigtede bæredygtighed og konkurrenceevne for virksomheder i branchen.

En af de mest betydningsfulde faktorer, der påvirker et byggeprojekts samlede CO2-aftryk, er produktions-

og logistikfaserne for byggematerialer og komponenter. De materialer, der bruges på byggepladsen, har ofte en afgørende indflydelse på den samlede CO_2-udledning, især i de tidlige stadier af projektets livscyklus - fra udvinding af råmaterialer til levering på byggepladsen.

Beregning af emissioner pr. kategori eller type bygningselement kræver brug af reference-kulstofemissionsfaktorer, der afspejler mængden af CO_2, der genereres fra produktionen af forskellige materialer. Sådanne materialer omfatter beton, mursten, genbrugsstål, aluminium og andre. Disse værdier er generelt hentet fra anerkendte kilder og internationale databaser som UK ICE 2015 (Inventory of Carbon and Energy) og US EPA 2006 (U.S. Environmental Protection Agency) [105]. Den følgende tabel (figur 5.3-6) opsummerer baseline-emissionsfaktorerne for en række almindelige byggematerialer. To nøgleparametre er angivet for hvert materiale: specifikke CO_2-emissioner (i kg pr. kg materiale) og volumen-til-vægt-omregningsfaktorer (i kg pr. kubikmeter), som er nødvendige for at integrere beregningerne i designmodellen og linke til QTO-datagrupperingen.



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

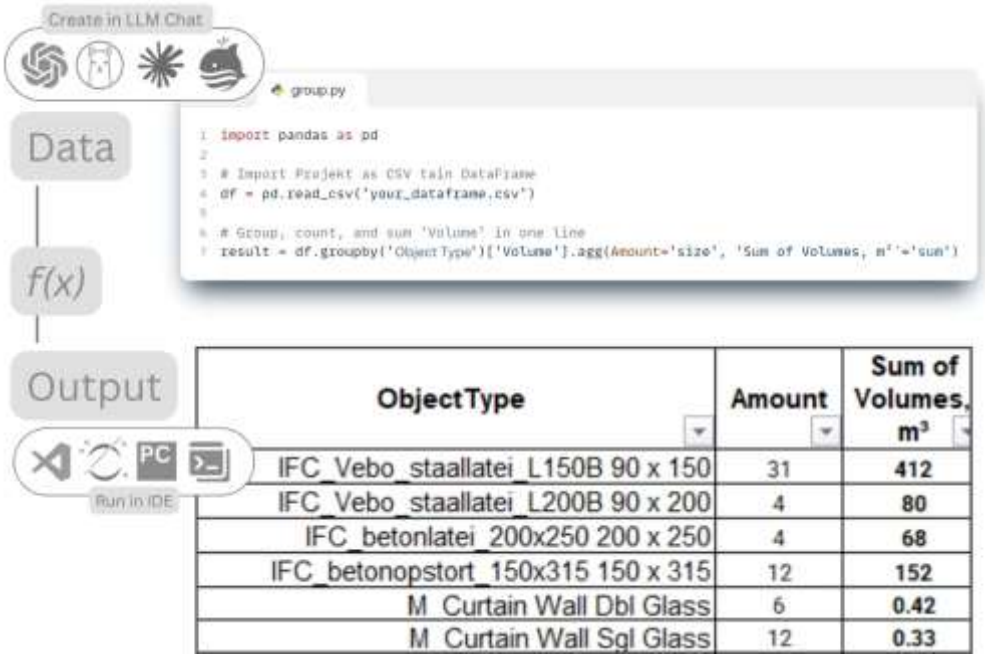
Figur 5.3-6 Mængden af kulstof, der udledes under produktionen af forskellige byggematerialer, ifølge den britiske ICE- og US EPA-database.

For at beregne de samlede CO_2 emissioner for et projekt, som med 4D og 5D beregningerne, er det nødvendigt at bestemme attributmængderne for hver gruppe af faciliteter. Dette kan gøres ved hjælp af kvantitative analyseværktøjer (QTO) ved at indhente attributmængderne i kubikmeter, som beskrevet i detaljer i afsnittet om Quantity take-off. Disse mængder multipliceres derefter med de relevante koefficienter for attributten "CO_2-procesemissioner" for hver materialegruppe.

- Lad os automatisk udtrække tabellen med mængder efter elementtype fra CAD (BIM) -projektet ved at gruppere alle projektdata som allerede gjort i de foregående kapitler. For at udføre denne opgave henviser vi til LLM.

> Gruppér venligst DataFrame-tabellen fra CAD (BIM) efter kolonneparameteren "Object Name" (eller "Type"), og vis antallet af elementer i hver gruppe, og opsummer "Volume"-parameteren for alle elementer i typen. ⏎

LLM's svar :



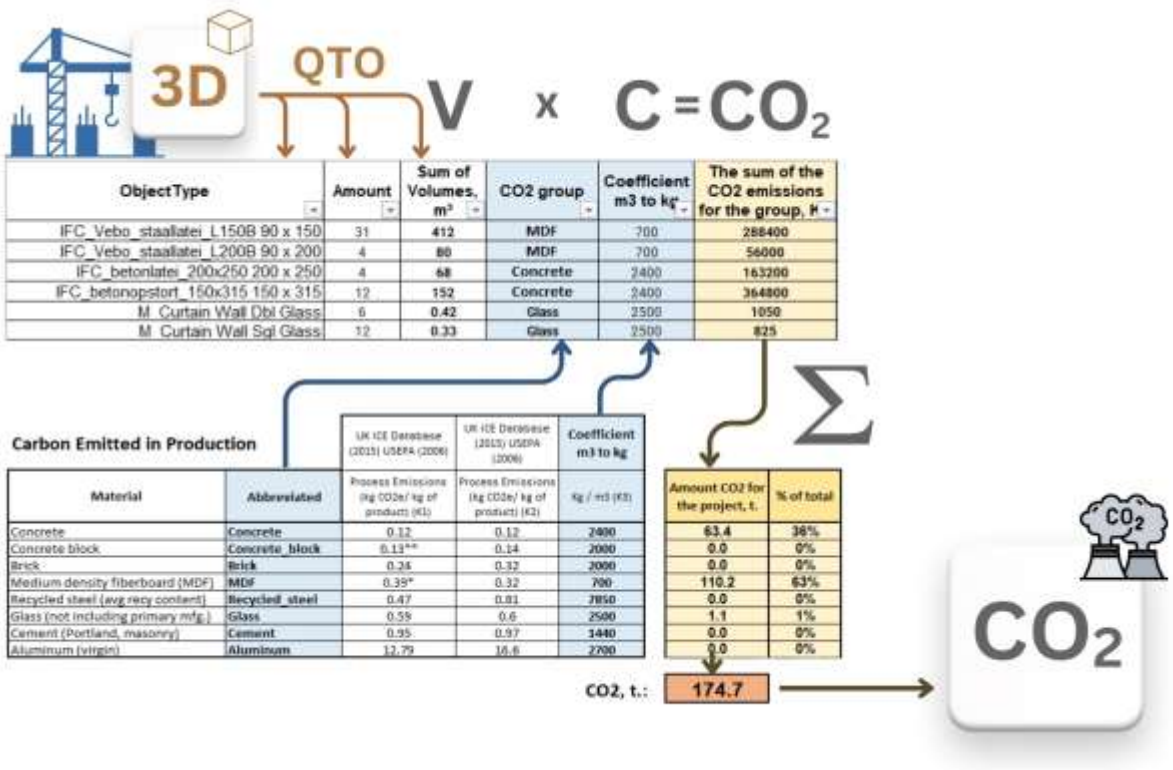
ObjectType	Amount	Sum of Volumes, m^3
IFC_Vebo_staallatei_L150B 90 x 150	31	412
IFC_Vebo_staallatei_L200B 90 x 200	4	80
IFC_betonlatei_200x250 200 x 250	4	68
IFC_betonopstort_150x315 150 x 315	12	152
M_Curtain Wall Dbl Glass	6	0.42
M_Curtain Wall Sgl Glass	12	0.33

Figur 5.3-7 Den genererede kode i LLM har grupperet projektenhederne for os efter type (ObjectType) med den sammenfattede attribut "Volume".

For at automatisere beregningen af de samlede CO_2-emissioner for hele projektet skal man blot opsætte automatisk datamapping i tabellen eller manuelt knytte elementtyperne (figur 5.3-7) til de tilsvarende materialetyper (figur 5.3-6) fra emissionsfaktortabellen. Den færdige tabel med emissionsfaktorer og formler samt kode til at hente volumener fra CAD-formater (BIM) og automatisere CO_2-bestemmelse kan findes på GitHub ved at søge efter "CO_2_calculating-the-embodied-carbon. DataDrivenConstruction." [106].

Således giver dataintegration efter gruppering af QTO-elementer fra CAD-databasen mulighed for automatisk beregning af kuldioxidemissioner (fig. 5.3-8) for forskellige designmuligheder. Det gør det muligt at analysere indvirkningen af forskellige materialer i forskellige varianter og kun vælge de løsninger, der opfylder krav til CO_2-udledning for at opnå et bestemt certifikatnår bygningen tages i brug.

> Estimering af CO_2-emissioner ved at gange faktorer med mængden af grupperede projektelementer er et typisk eksempel på en opgave i forbindelse med, at en byggevirksomhed opnår en ESG-klassificering (f.eks. LEED-certificering) for et anlæg.



Figur 5.3-8 Integration af QTO-grupper fra CAD-databaser giver nøjagtighed og automatisering ved beregning af den endelige CO_2-udledning.

> På samme måde kan vi ved at definere elementgruppemængder udføre beregninger for materialekontrol og logistik, kvalitetsovervågning og -styring, energimodellering og -analyse og en lang række andre opgaver for at opnå en ny attributstatus (parameter i tabellen) for både individuelle elementgrupper og hele projektet.

Hvis antallet af sådanne beregningsprocesser i virksomheden begynder at vokse, opstår spørgsmålet om behovet for at automatisere sådanne beregninger og implementere resultaterne af beregningerne i virksomhedens processer og datahåndteringssystemer.

På grund af kompleksiteten i en komplet løsning outsourcer mellemstore og store virksomheder i byggebranchen en sådan automatisering til ERP (eller PMIS) systemudviklingsfirmaer. Udviklingsfirmaerne opretter et enkelt omfattende modulært system til store kunder til at håndtere mange forskellige informationslag, herunder materiale- og ressourceberegninger.

KAPITEL 5.4.
KONSTRUKTION ERP OG PMIS SYSTEMER

Construction ERP -systemer på eksemplet med beregninger og estimater

Modulære ERP systemer integrerer forskellige attributlag (information) og datastrømme i et enkelt omfattende system, så projektledere kan styre ressourcer, økonomi, logistik og andre aspekter af et projekt på en synkroniseret måde inden for en enkelt platform. Et ERP-system til byggeri fungerer som "hjernen" i byggeprojekter og forenkler gentagne processer gennem automatisering, hvilket giver gennemsigtighed og kontrol i hele byggeprocessen.

> **Construction ERP** -systemer (Enterprise Resource Planning) er omfattende softwareløsninger, der er designet til at styre og optimere forskellige aspekter af byggeprocessen. Kernen i ERP-systemer til byggeriet er moduler til styring af omkostningsberegning og planlægning, hvilket gør dem til et vigtigt værktøj til effektiv ressourceplanlægning.

ERP-moduler -systemer giver brugerne mulighed for at indtaste, behandle og analysere data på en struktureret måde, der dækker forskellige aspekter af et projekt, som kan omfatte regnskab for materiale- og lønomkostninger, udnyttelse af udstyr, logistikstyring, menneskelige ressourcer, kontakter og andre byggeaktiviteter.

En af systemets funktionelle blokke er modulet til automatisering af forretningslogik - BlackBox/WhiteBox, som spiller rollen som proceskontrolcenter.

BlackBox /WhiteBox giver specialister, der bruger et ERP -system, mulighed for fleksibelt at styre forskellige aspekter af virksomheden via adgangsrettigheder, der allerede er blevet forudkonfigureret af andre brugere eller administratorer. I forbindelse med ERP-systemer henviser begreberne *BlackBox* og *WhiteBox* til niveauet af gennemsigtighed og kontrollerbarhed af systemets interne logik:

- **BlackBox** ("sort boks") - brugeren interagerer med systemet via grænsefladen uden adgang til den interne logik i procesudførelsen. Systemet udfører beregninger på egen hånd, baseret på foruddefinerede regler, der er skjult for slutbrugeren. Han indtaster data og får resultatet uden at vide, hvilke attributter eller koefficienter der blev brugt indeni.
- **WhiteBox** ("hvid boks") - proceslogikken er tilgængelig for visning, tilpasning og ændring. Avancerede brugere, administratorer eller integratorer kan manuelt definere databehandlingsalgoritmer, beregningsregler og interaktionsscenarier mellem projektenheder.

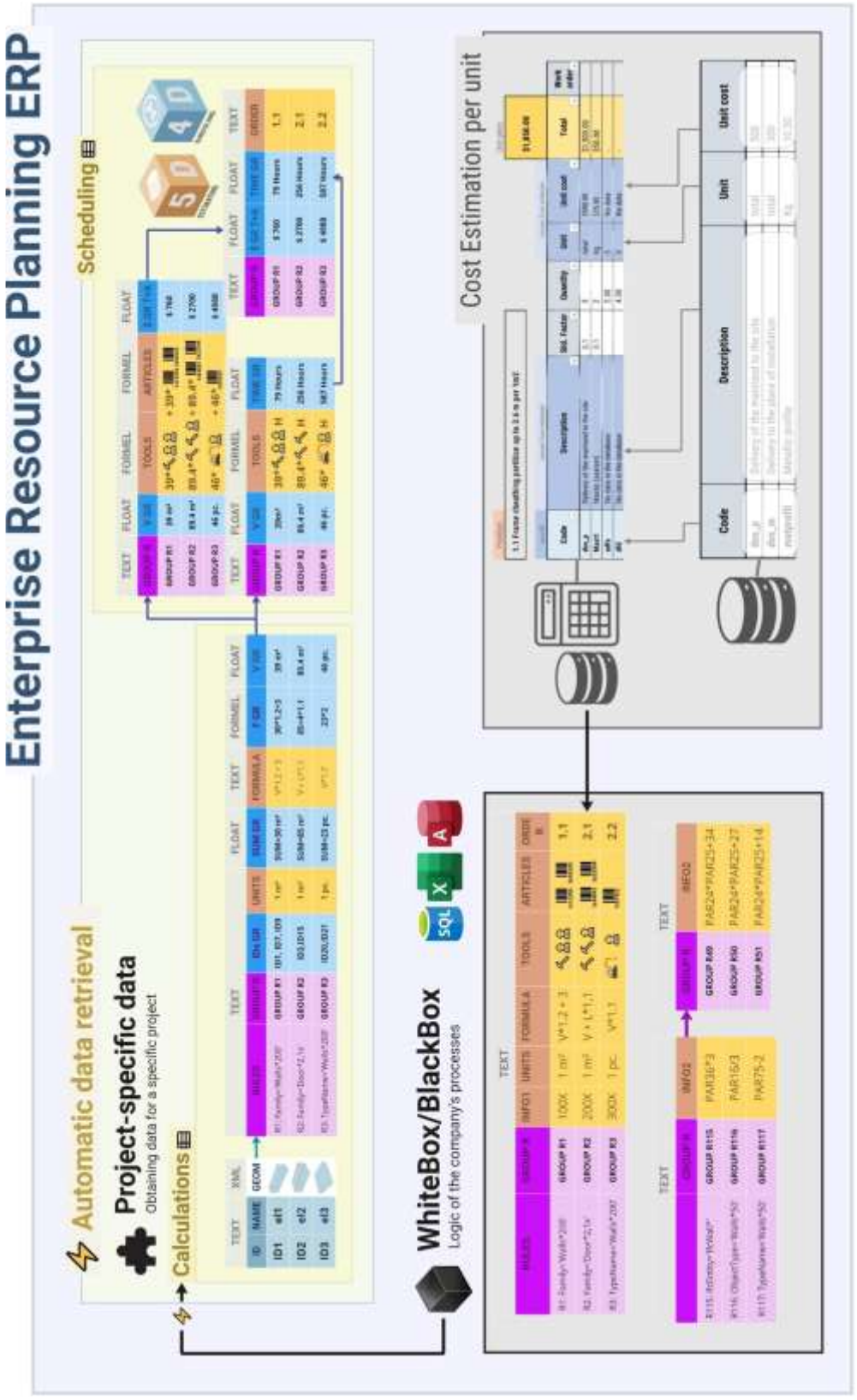


Fig. 5.4-1 Architecture of construction ERP -system, for at få estimater og arbejdsplaner ved manuelt at udfylde volumenattributterne.

Et eksempel er, når en erfaren bruger eller administrator indstiller en regel: hvilke attributter i et estimat, der skal ganges med hinanden eller grupperes efter en bestemt egenskab, og hvor det endelige resultat skal registreres. Efterfølgende kan mindre uddannede fagfolk, som f.eks. beregningsingeniører, blot indlæse nye data i ERP-systemet via brugergrænsefladen - og få færdige beregninger, tidsplaner eller specifikationer uden at skulle skrive kode eller forstå de tekniske detaljer i logikken.

> I de foregående kapitler blev beregnings- og logikmodulerne diskuteret i forbindelse med LLM-interaktion. I et ERP-miljø foregår sådanne beregninger og transformationer i moduler, der er skjult bag en grænseflade med knapper og formularer.

I det følgende eksempel (fig. 5.4-1) har administratoren af ERP -systemet i BlackBox /WhiteBox -modulet defineret regler for matchning af attributter for enheder fra estimater med attributter til gruppering af QTO. Takket være dette BlackBox/WhiteBox-modul, der er konfigureret (af lederen eller administratoren), modtager brugeren (estimator eller ingeniør) ved manuelt at tilføje en mængde- eller volumenattribut via ERP-brugergrænsefladen automatisk de færdige estimater og arbejdsplaner. På den måde bliver de beregnings- og estimatgenereringsprocesser, der blev diskuteret i de foregående kapitler, ved hjælp af kode inde i ERP-systemet til et halvautomatisk transportbånd.

Ved at forbinde denne halvautomatiske proces med de volumetriske attributter fra CAD (BIM) modeller (fig. 4.1-13), f.eks. ved at indlæse CAD-projektet i ERP-modulet, der er forudkonfigureret til dette formål, bliver datastrømmen til en synkroniseret mekanisme, der er i stand til autonomt og øjeblikkeligt at opdatere værdien af individuelle grupper af elementer eller hele projektet som reaktion på eventuelle ændringer i det i designfasen, når CAD-modellen indlæses i ERP-systemet.

For at skabe et automatiseret dataflow (fig. 5.4-2) mellem CAD (BIM) og ERP systemer skal de grundlæggende processer og krav til data fra CAD (BIM) modeldatabaser defineres på en struktureret måde, som vi allerede har diskuteret i kapitlet ovenfor "Krav og kvalitetssikring af data ". Denne proces i ERP er opdelt i lignende trin:

- **Oprettelse af valideringsregler (1),** som spiller en vigtig rolle i at sikre nøjagtigheden af de data, der kommer ind i ERP -systemet. Valideringsregler fungerer som filtre, der validerer enheder og deres attributter, så kun de elementer, der opfylder kravene, kan komme ind i systemet. Du kan lære mere om verifikation og validering i kapitlet "Oprettelse af krav og validering af datakvalitet ".
- Derefter finder **en verificeringsproces (2)** sted i ERP, som bekræfter, at alle projektenhedselementer med deres attributter og værdier er blevet oprettet korrekt og er klar til de næste behandlingstrin.
- Hvis der er problemer med ufuldstændige attributdata**, genereres der en rapport (3),** og projektet sendes sammen med instruktioner til korrektion til revision, indtil det er klar til næste iteration.
- Når projektdataene er blevet valideret og verificeret, bruges de i et andet ERP-modul **(4) til at oprette Quantity Take-Off-tabeller (QTO**), der opretter mængdeattributter for enhedsgrupper, materialer og ressourcer i henhold til tidligere genererede regler (WhiteBox/BlackBox).
- Grupperede data efter matchningsregler eller QTO **integreres** automatisk **med beregninger (f.eks. omkostninger og tid) (5).**
- I det sidste trin i ERP -systemet **genererer** brugeren **automatisk beregningsresultater (6) (**f.eks. omkostningsestimater, arbejdsplaner eller CO_2-emissioner) for hver enhedsgruppe og for

projektet som helhed ved at gange scope-attributterne fra QTO-tabellen med attributterne i procestabellerne (f.eks. estimerede elementer).

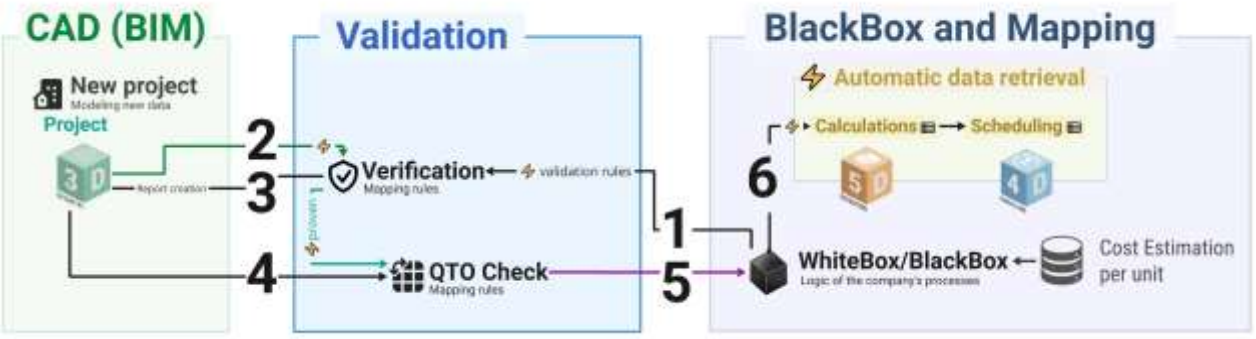


Fig. 5.4-2 Arkitektur af et ERP -system til byggeriet med CAD (BIM), fra oprettelse af valideringsregler (1) til automatisk beregning af omkostninger og arbejdsplaner (5-6).

I et modulopbygget ERP -system integreres processer ved hjælp af software, der omfatter en brugergrænseflade. Bag grænsefladen ligger back-end, hvor strukturerede tabeller behandler data ved at udføre forskellige operationer, som er blevet forudkonfigureret af lederen eller administratoren. Resultatet er, at brugeren takket være den foruddefinerede og tilpassede automatiseringslogik (i modulerne BlackBox /WhiteBox) modtager halvautomatisk forberedte dokumenter, der opfylder hans opgaver

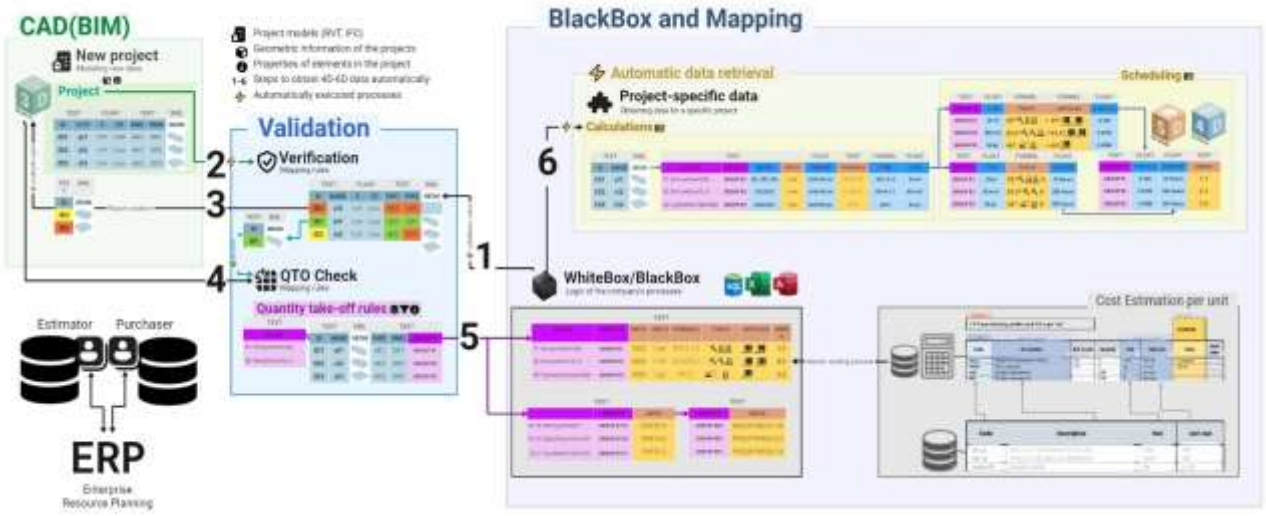


Fig. 5.4-3 ERP -systemet hjælper ledere og brugere med at bevæge sig mellem specialtabeller for at generere nye data.

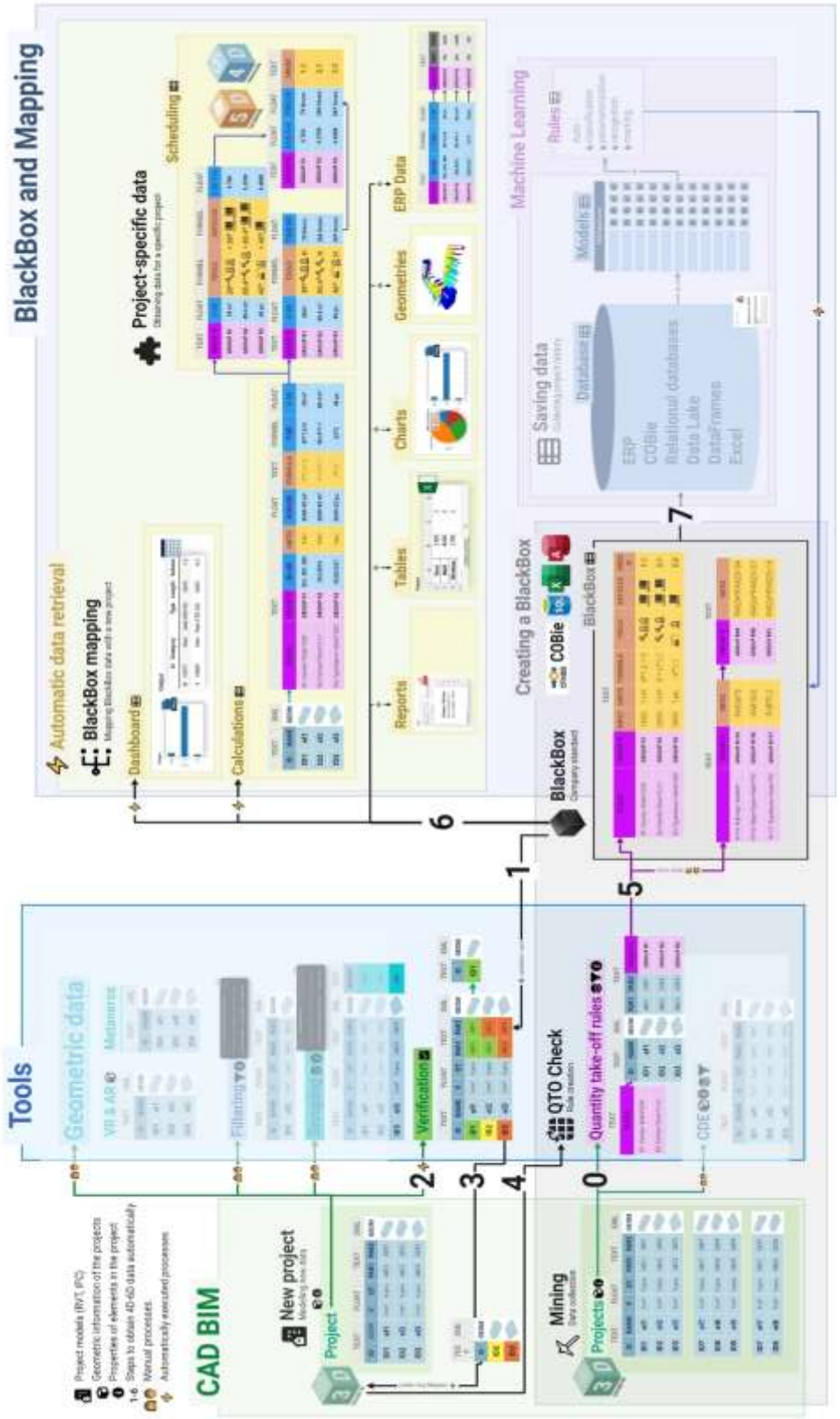


Fig. 5.4-4 ERP -systemet er integreret med analytiske værktøjer og automatiserer virksomhedens beslutningsproces.

På samme måde er processerne i ERP -systemer, fra begyndelse til endelig beregning (trin 1-6 Figur 5.4-3), en kæde af indbyrdes forbundne trin, der i sidste ende giver gennemsigtighed, effektivitet og nøjagtighed i planlægningen.

Moderne ERP-systemer til byggeriet omfatter ikke kun moduler til beregning af omkostninger og tidsplaner, men også dusinvis af andre forudkonfigurerede moduler, der normalt dækker dokumenthåndtering, sporing af projektforløb, kontraktstyring, forsyningskæde og logistik samt integration med andre forretningssystemer og platforme. De integrerede analyseværktøjer i ERP giver brugerne mulighed for at automatisere oprettelsen af dashboards til overvågning af projekt-KPI'er (KPI - key performance indicators). Dette giver en centraliseret og konsekvent styring af alle aspekter af et byggeprojekt, hvor man forsøger at kombinere et stort antal applikationer og systemer på en enkelt platform.

I fremtiden vil ERP -analytics blive brugt i kombination med maskinlæring for at øge nøjagtigheden og optimere processen med at beregne fremtidige projektattributter. De data og attributter, der analyseres og indsamles fra ERP-systemer i Big Data (fig. 5.4-4), vil i fremtiden danne grundlag for at skabe forudsigelsesmodeller, der præcist kan forudse potentielle forsinkelser, risici eller f.eks. mulige ændringer i materialeomkostninger.

Som et alternativ til ERP bruger byggebranchen ofte PMIS (Project Management Information System), et projektstyringssystem, der er designet til detaljeret styring af opgaver på det enkelte byggeprojekts niveau.

PMIS: Mellemled mellem ERP og byggepladsen

I modsætning til ERP, som dækker hele kæden af en virksomheds forretningsprocesser, fokuserer PMIS på styring af et specifikt projekt, overvågning af tidslinjer, budgetter, ressourcer og dokumentation.

> PMIS (Project Management Information System) er software til styring af byggeprojekter, der er designet til at planlægge, spore, analysere og rapportere om alle aspekter af et projekt.

PMIS giver dig mulighed for at styre dokumenter, tidsplaner, budgetter og ved første øjekast kan PMIS se ud til at være en kopi af ERP, men den vigtigste forskel er styringsniveauet:

- **ERP** er fokuseret på virksomhedens forretningsprocesser som helhed: omkostninger, kontrakter, indkøb, personale og ressourcestyring på virksomhedsniveau.
- **PMIS** fokuserer på styring af individuelle projekter og sørger for detaljeret planlægning, ændringskontrol, rapportering og koordinering af deltagere.

I mange tilfælde er det ERP -systemerne, der allerede har tilstrækkelig funktionalitet, og implementeringen af PMIS bliver mere et spørgsmål om bekvemmelighed og virksomhedens præferencer. Mange entreprenører og kunder bruger ikke PMIS, fordi det er nødvendigt, men fordi det er pålagt af leverandøren eller en stor kunde, der ønsker at samle data på en bestemt platform.

Det skal nævnes, at der i den internationale terminologi for byggeprojektledelse er andre separate populære begreber som PLM (Product Lifecycle Management) og EPC og EPC-M (Engineering, Procurement and Construction Management) - entreprisemetoder i byggebranchen.

Hvis en virksomhed allerede bruger ERP med projektstyringsmoduler, kan indførelsen af PMIS være et unødvendigt link, der duplikerer funktionalitet. Men hvis processerne ikke er automatiserede, og data er fragmenterede, kan PMIS være et mere praktisk og let at vedligeholde værktøj.

Spekulation, profit, isolerethed og mangel på gennemsigtighed i ERP og PMIS

På trods af den ydre enkelhed i grænseflader og procedurer er ERP- og PMIS-systemer til byggeriet i de fleste tilfælde lukkede og ufleksible løsninger. Sådanne systemer leveres normalt som en forudkonfigureret softwarepakke fra en enkelt leverandør med begrænset adgang til interne databaser og proceslogik.

CAD-(BIM-) leverandører overtager i stigende grad udviklingen og styringen af sådanne systemer, da deres databaser indeholder de oplysninger, der kræves af ERP-systemer: kvantitative og volumetriske attributter for projektelementer. Men i stedet for at give adgang til disse data i et åbent eller maskinlæsbart format, tilbyder leverandørerne kun begrænsede brugerscenarier og lukket behandlingslogik - foruddefineret i BlackBox-moduler. Det reducerer systemets fleksibilitet og forhindrer, at det kan tilpasses specifikke projektforhold.

> Begrænset datatransparens er fortsat en af de største udfordringer ved digitale processer i byggeriet. Lukket databasearkitektur, manglende adgang til komplette sæt af bygningselementers attributter, fokus på BlackBox-automatiseringsmoduler og mangel på åbne grænseflader øger risikoen for dokumentbureaukrati betydeligt. Sådanne begrænsninger skaber flaskehalse i beslutningsprocessen, gør det vanskeligt at verificere oplysninger og åbner døren for data, der skjules eller spekuleres i ERP/PMIS-systemer. Brugerne får typisk kun begrænset adgang - det være sig en nedbarberet grænseflade eller en delvis API - uden mulighed for at interagere direkte med de primære datakilder. Dette er især kritisk, når det drejer sig om parametre, der automatisk genereres fra CAD-projekter, f.eks. volumener, arealer og mængder, der bruges til QTO-beregninger.

I stedet for at søge effektivitet gennem procesautomatisering, åbne data, reducere transaktionsomkostninger og skabe nye forretningsmodeller, fokuserer mange byggevirksomheder på at styre eksterne parametre - manipulere faktorer, justeringsfaktorer og beregningsmetoder, der påvirker projektomkostningerne i lukkede ERP/PMIS-platforme. Det skaber plads til spekulation, fordrejer de reelle produktionsomkostninger og reducerer tilliden mellem alle deltagere i byggeprocessen

> I byggeriet er overskuddet forskellen mellem indtægterne fra et afsluttet projekt og de variable omkostninger, som omfatter design, materialer, arbejdskraft og andre direkte omkostninger, der er direkte relateret til gennemførelsen af projektet. Men den vigtigste faktor, der påvirker værdien af disse omkostninger, er ikke kun teknologi eller logistik, men også hastigheden og nøjagtigheden af beregningerne og kvaliteten af ledelsesbeslutningerne i virksomheden.

Problemet forværres af, at omkostningsberegningsprocesserne i de fleste byggevirksomheder forbliver uigennemsigtige, ikke kun for kunderne, men også for medarbejderne selv, som ikke er en del af beregnings- eller økonomiafdelingerne. En sådan lukkethed fremmer dannelsen af en privilegeret gruppe af specialister i virksomheden - bærere af "økonomisk ekspertise", som har eneret til at redigere attributter og korrektionsfaktorer i ERP/PMIS -systemer. Disse medarbejdere kan sammen med virksomhedslederne faktisk kontrollere projektets finansielle logik.

> Under sådanne forhold bliver beregnere til "økonomiske jonglører", der balancerer mellem maksimering af virksomhedens overskud og behovet for at opretholde en konkurrencedygtig pris for kunden. Samtidig skal de undgå åbenlyse og grove manipulationer for ikke at underminere virksomhedens omdømme. Det er i denne fase, at der fastsættes koefficienter for at skjule overvurderede mængder eller omkostninger til materialer og arbejde.

Som følge heraf er den vigtigste metode til at øge effektiviteten og rentabiliteten hos virksomheder i byggebranchen ikke automatisering og fremskyndelse af beslutningsprocesser, men spekulation i priserne på materialer og arbejder (fig. 5.4-5). Overvurdering af omkostningerne til arbejder og materialer udføres ved "grå" bogføring i lukkede ERP /PMIS - systemer ved at oppuste procenterne over de gennemsnitlige markedspriser for materialer eller arbejdsmængder ved hjælp af koefficienter (fig. 5.1-6), som blev diskuteret i kapitlet "Udarbejdelse af beregninger og beregning af omkostningerne til arbejder på grundlag af ressourcegrundlaget".

> Resultatet er, at kunden modtager en beregning, der ikke afspejler de reelle omkostninger eller arbejdets omfang, men er en afledning af mange skjulte interne koefficienter. Samtidig tvinges underleverandører ofte til at købe billigere materialer af lav kvalitet i et forsøg på at leve op til de undervurderede priser, som hovedentreprenøren har fastsat, hvilket forringer den endelige kvalitet af byggeriet.

Den spekulative proces med at finde overskud ud af den blå luft ender med at skade både kunder, der modtager upålidelige data, og ledere, der er tvunget til at finde flere og flere spekulationsmodeller

Resultatet er, at jo større projektet er, jo mere bureaukratisk er data- og processtyringen. Hvert trin og hvert modul skjuler ofte uigennemsigtige koefficienter og tillæg, der er indlejret i beregningsalgoritmer og interne procedurer. Det gør ikke kun revisionen vanskelig, men forvrænger også det økonomiske billede af projektet betydeligt. I store byggeprojekter fører en sådan praksis ofte til en mangedobling (nogle gange op til en tidobling) af de endelige omkostninger, mens de reelle mængder og omkostninger forbliver uden for bygherrens effektive kontrol (fig. 2.1-3 Sammenligning af planlagte og faktiske omkostninger ved store infrastrukturprojekter i Tyskland.

> Ifølge McKinsey & Companys rapport Imagining the Digital Future of Construction (2016) bliver store byggeprojekter i gennemsnit afsluttet 20% senere end planlagt og op til 80% over budgettet [107].

Estimerings- og budgetteringsafdelinger er ved at blive det mest beskyttede led i en virksomhed. Adgangen til dem er strengt begrænset, selv for interne specialister, og på grund af den lukkede logik og databasestrukturerne er det umuligt objektivt at vurdere effektiviteten af projektbeslutninger uden forvrængninger. Den manglende gennemsigtighed fører til, at virksomheder ikke er tvunget til at optimere processer, men til at kæmpe for overlevelse ved "kreativ" styring af tal og faktorer (Fig. 5.3-1, Fig. 5.1-6 - for eksempel parameteren "Bid. Factor").

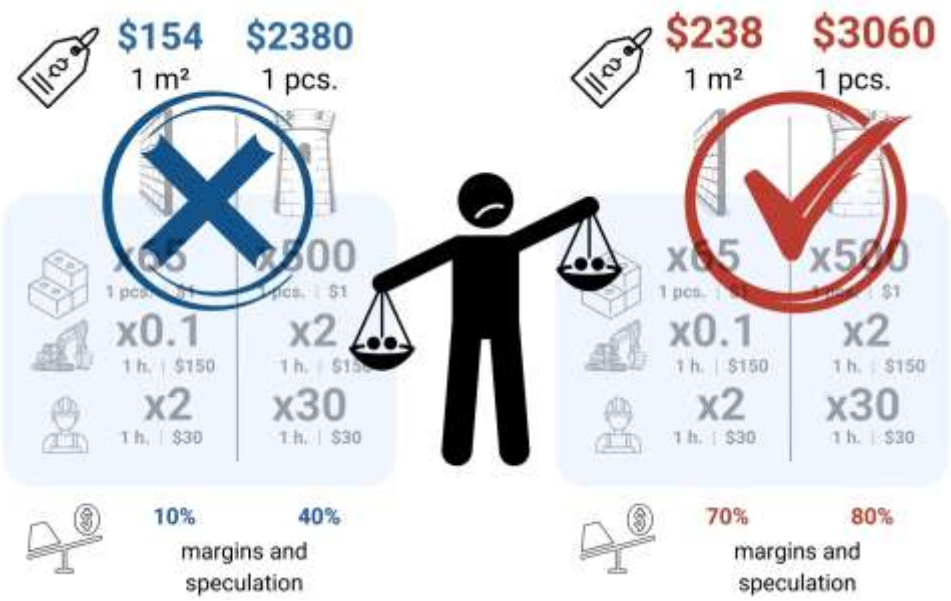


Fig. 5.4-5 Spekulationsrater på afregningsniveau er virksomhedernes største fortjeneste og kunsten at jonglere mellem arbejdskvalitet og omdømme.

Alt dette sår tvivl om den fortsatte brug af lukkede ERP/PMIS -systemer i byggeriet. I forbindelse med den digitale transformation og stigende krav om gennemsigtighed fra kundernes side (fig. 10.2-3) er det usandsynligt, at projektrealisering på lang sigt vil forblive afhængig af proprietære løsninger, der begrænser fleksibiliteten, hindrer integration og hæmmer forretningsudviklingen.

Og uanset hvor fordelagtigt det måtte være for byggevirksomheder at arbejde med datasiloer og uigennemsigtige data i lukkede databaser - så vil fremtiden for byggebranchen uundgåeligt indebære en overgang til åbne platforme, maskinlæsbare og gennemsigtige datastrukturer og tillidsbaseret automatisering. Denne transformation vil blive drevet fra toppen - under pres fra kunder, myndigheder og samfundet, som i stigende grad kræver ansvarlighed, bæredygtighed, gennemsigtighed og økonomisk gennemførlighed.

Slut på æraen med lukkede ERP /PMIS: Byggebranchen har brug for nye tilgange

Brugen af store modulopbyggede ERP/PMIS -systemer, der består af titusindvis af kodelinjer, gør det ekstremt vanskeligt at foretage ændringer i dem. I dette tilfælde bliver overgangen til en ny platform med moduler, der allerede er forudkonfigureret til virksomheden, titusindvis af artikler i ressourcedatabaser (fig. 5.1-3) og tusindvis af færdige beregninger (fig. 5.1-6) til en kostbar og langvarig proces. Jo mere kode og ældre arkitektur - jo højere er niveauet af intern ineffektivitet, og hvert nyt projekt vil kun gøre tingene værre. I mange virksomheder bliver datamigrering og integration af nye løsninger et flerårigt epos ledsaget af konstant omarbejde og endeløs søgen efter kompromiser. Resultatet er ofte en tilbagevenden til gamle, velkendte platforme på trods af deres begrænsninger.

Som fremhævet i den tyske Black Book-rapport [108] om systemiske fejl i byggedatastyringen er fragmenteringen af information og manglen på en centraliseret tilgang til styringen af den en vigtig årsag til ineffektivitet. Uden standardisering og integration mister data deres værdi og bliver et arkiv snarere end

et ledelsesværktøj.

> En væsentlig årsag til tab af datakvalitet er utilstrækkelig planlægning og kontrol af byggeprojekter, hvilket ofte fører til betydelige omkostningsstigninger. Black Books afsnit "Fokus: Omkostningseksplosionen" analyserer de vigtigste faktorer, der bidrager til disse uønskede konsekvenser. De omfatter utilstrækkelige behovsanalyser, manglende gennemførlighedsundersøgelser og ukoordineret planlægning, der fører til ekstraomkostninger, som kunne have været undgået.

> I en virksomheds modne IT-økosystem kan udskiftning af et forældet system sammenlignes med udskiftning af en bærende søjle i en allerede konstrueret bygning. Det er ikke nok bare at fjerne den gamle og installere en ny - det er vigtigt at gøre det på en sådan måde, at bygningen forbliver stabil, lofterne ikke kollapser, og al kommunikation fortsat fungerer. Det er her, vanskelighederne ligger: Enhver fejl kan få alvorlige konsekvenser for hele virksomhedens system.

Alligevel fortsætter udviklere af store ERP-produkter til byggebranchen med at bruge mængden af skrevet kode som et argument for deres platform. På specialiserede konferencer kan man stadig høre sætninger som: "Det ville tage 150 mandeår at genskabe sådan et system", på trods af at det meste af funktionaliteten i sådanne systemer gemmer på databaser og ret enkle funktioner til at arbejde med tabeller, pakket ind i en særlig fast brugergrænseflade. I praksis bliver kodemængden på "150 mandeår" til en byrde snarere end en konkurrencefordel. Jo mere kode - jo højere omkostninger til support, jo sværere er det at tilpasse sig nye forhold, og jo højere er indgangstærsklen for nye udviklere og kunder.

> Mange modulære byggesystemer ligner i dag besværlige og forældede "Frankenstein-konstruktioner", hvor enhver skødesløs ændring kan føre til fejl. Hvert nyt modul øger kompleksiteten i et allerede overbelastet system og gør det til en labyrint, som kun nogle få specialister kan forstå, hvilket gør det endnu sværere at vedligeholde og modernisere.

Kompleksiteten erkendes også af udviklerne selv, som med jævne mellemrum holder pause for at refaktorere - revidere arkitekturen for at tage højde for fremkomsten af nye teknologier. Men selv om refaktorering udføres regelmæssigt, vokser kompleksiteten uundgåeligt. Arkitekterne af sådanne systemer vænner sig til den voksende kompleksitet, men for nye brugere og specialister bliver det en uoverstigelig barriere. Resultatet er, at al ekspertise koncentreres hos nogle få udviklere, og systemet ophører med at være skalerbart. På kort sigt er sådanne eksperter nyttige, men på lang sigt bliver de en del af problemet.

> *Organisationer vil fortsætte med at integrere "små" data med deres "store" data, og det er tåbeligt at tro, at én applikation - uanset hvor dyr eller robust den er - kan håndtere alt [109].*
>
> - Phil Simon, vært for podcasten Conversations About Collaboration

Der opstår et legitimt spørgsmål: Har vi virkelig brug for så besværlige og lukkede systemer til at beregne omkostninger og timing af arbejde i form af tabeller, hvis andre brancher længe har været i stand til at klare

lignende opgaver ved hjælp af analytiske værktøjer med åbne data og gennemsigtig logik?

I øjeblikket er lukkede modulære platforme stadig efterspurgte i byggebranchen, primært på grund af de særlige forhold, der gør sig gældende for omkostningsregnskaber (fig. 5.1-7). Sådanne systemer bruges ofte til at køre grå eller uigennemsigtige ordninger, der gør det muligt at skjule de reelle omkostninger for kunden. Men efterhånden som branchen modnes digitalt, primært kunderne, og bevæger sig ind i den såkaldte "Uberised Era", vil mellemleddene, nemlig byggevirksomhederne med deres ERP-systemer, miste deres betydning for tids- og omkostningsberegninger. Det vil ændre byggebranchens ansigt for altid. Læs mere i den sidste del af bogen og i kapitlet "Byggeri 5.0: Hvordan man tjener penge, når man ikke længere kan gemme sig".

> Tusindvis af ældre løsninger, der er akkumuleret i løbet af de sidste 30 år med tusindvis af mandeår investeret i udvikling, vil begynde at forsvinde hurtigt. Overgangen til åben, gennemsigtig og fleksibel datahåndtering er uundgåelig. Det eneste spørgsmål er, hvilke virksomheder der vil være i stand til at tilpasse sig disse ændringer, og hvilke der vil forblive gidsler i den gamle model.

En lignende situation ses inden for CAD-værktøjer (BIM -), hvis data i dag udfylder de volumetriske parametre for konstruktionsenheder i ERP/PMIS -systemer. Oprindeligt var ideen om BIM (udviklet tilbage i 2002 [110]) baseret på konceptet om en enkelt integreret database, men i praksis kræver arbejdet med BIM i dag et helt sæt specialiseret software og formater. Det, der skulle forenkle design- og byggestyring, er blevet til endnu et lag af proprietære løsninger, der komplicerer integrationen og reducerer forretningsfleksibiliteten.

Næste skridt: effektiv brug af projektdata

I denne del har vi vist, hvordan strukturerede data bliver grundlaget for nøjagtige omkostnings- og tidsplanberegninger for byggeprojekter. Automatisering af QTO-, planlægnings- og estimeringsprocesserne reducerer arbejdsomkostningerne og forbedrer resultaternes nøjagtighed betydeligt.

For at opsummere denne del er det værd at fremhæve de vigtigste praktiske trin, der vil hjælpe dig med at anvende de diskuterede tilgange til dine daglige opgaver. Disse tilgange er universelle - de er nyttige både til den digitale transformation af en virksomhed og til det daglige arbejde for fagfolk, der er involveret i beregninger:

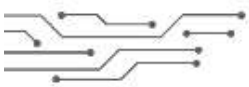
- Automatiser rutineberegninger
  - Prøv at finde standardomkostninger, som du kan relatere til i dit arbejde.
  - Analyser, hvilke metoder der bruges til at beregne omkostninger eller processer på byggepladsen i dit land (figur 5.1-7).
  - Hvis du arbejder med et CAD-system - lær funktionerne til automatisk udtræk af specifikationer og QTO-data i din CAD-software (BIM-).
  - Brug LLM til at skrive udkast til kode for at automatisere beregninger
- Udvikl dine egne værktøjer til QTO

- [ ] Opret scripts eller tabeller til at automatisere volumenoptælling
- [ ] Standardiser kategorier og grupper af elementer for en konsekvent tilgang til evaluering
- [ ] Dokumenter beregningsmetoden for at sikre reproducerbarhed af resultater i nye projekter

- Integrer forskellige aspekter af projektet i dit arbejde
  - [ ] Hvis du arbejder med modulopbyggede systemer, så prøv at visualisere dine processer ikke kun som diagrammer eller diagrammer, men også på dataniveau - især i form af tabeller.
  - [ ] Behersk automatisk sammenlægning af data fra CAD-databaser med beregninger - med Python-kode ved hjælp af gruppering, filtrering og aggregering.
  - [ ] Skab klare visualiseringer af QTO-grupper for at præsentere komplekse oplysninger for kolleger og kunder

Disse skridt vil hjælpe med at opbygge et bæredygtigt beregningssystem baseret på automatisering og standardisering af data. Denne tilgang vil forbedre nøjagtigheden og reducere rutinen i de daglige beregningsproblemer.

De følgende kapitler fokuserer på de tekniske aspekter af CAD - (BIM-) produkter og årsagerne til, at CAD-databaser stadig er vanskelige at integrere i virksomhedernes forretningsprocesser. Hvis du ikke er interesseret i historien om BIM-implementering i byggeriet, udviklingen af CAD-værktøjer og de tekniske aspekter ved at arbejde med disse teknologier, kan du gå direkte til syvende del af bogen "Datadrevet beslutningstagning".



MAKSIMAL BEKVEMMELIGHED MED DEN TRYKTE VERSION

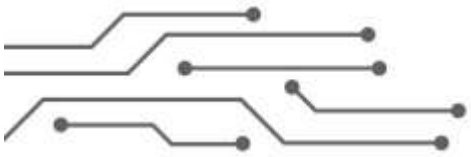
Du sidder med en gratis digital version af **Datadrevet byggeri**. For mere praktisk arbejde og hurtig adgang til materialerne anbefaler vi, at du er opmærksom på den **trykte udgave**:



- **Altid ved hånden**: Bogen i trykt format bliver et pålideligt arbejdsredskab, så du hurtigt kan finde og bruge de rigtige visualiseringer og diagrammer i enhver arbejdssituation.
- **Illustrationer i høj kvalitet**: alle billeder og grafikker i den trykte udgave præsenteres i maksimal kvalitet
- **Hurtig adgang til information**: nem navigation, mulighed for at lave noter, bogmærker og arbejde med bogen overalt.

Ved at købe den fulde trykte version af bogen får du et praktisk værktøj til komfortabelt og effektivt arbejde med information: muligheden for hurtigt at bruge visuelle materialer i hverdagens opgaver, hurtigt finde de nødvendige skemaer og tage noter. Desuden støtter dit køb udbredelsen af åben viden.

Bestil en trykt version af bogen på: datadrivenconstruction.io/books



VI DEL

CAD OG BIM: MARKEDSFØRING, VIRKELIGHED OG FREMTIDEN FOR DESIGNDATA I BYGGERIET

Den sjette del af bogen præsenterer en kritisk analyse af udviklingen af CAD og BIM -teknologier og deres indvirkning på datastyringsprocesser i byggeriet. Den sporer den historiske transformation af BIM-konceptet fra den oprindelige idé om en integreret database til de nuværende markedsføringskonstruktioner, der promoveres af softwareleverandører. Indvirkningen af proprietære formater og lukkede systemer på effektiviteten af projektdatahåndtering og byggebranchens samlede præstation vurderes. Problemerne med kompatibilitet mellem forskellige CAD-systemer og vanskelighederne med at integrere dem med byggevirksomhedernes forretningsprocesser analyseres i detaljer. Aktuelle tendenser i retning af forenklede åbne dataformater, såsom USD, og deres potentielle indvirkning på branchen diskuteres. Alternative tilgange til at udtrække information fra lukkede systemer præsenteres, herunder reverse engineering-teknikker. Udsigterne for anvendelse af kunstig intelligens og maskinlæring til at automatisere design- og dataanalyseprocesser i byggeriet analyseres. Der formuleres prognoser for udviklingen af designteknologier, der er orienteret mod brugernes reelle behov snarere end mod softwareleverandørernes interesser.

KAPITEL 6.1.
FREMKOMSTEN AF BIM- KONCEPTER I BYGGEBRANCHEN

Oprindeligt var denne sjette del, der er dedikeret til CAD (BIM), ikke med i den første version af bogen. Emnerne proprietære formater, geometriske kerner og lukkede systemer er alt for tekniske, overlæssede med detaljer og tilsyneladende ubrugelige for dem, der bare gerne vil forstå, hvordan man arbejder med data. Men feedback og anmodninger om at tilføje afklaring til den første version af bogen viste, at uden at forstå kompleksiteten i CAD-systemernes indre arbejde, geometriske kerner, de forskellige formater og inkompatible lagringsordninger for de samme data, er det umuligt virkelig at forstå, hvorfor de koncepter, der fremmes af leverandørerne, ofte gør det vanskeligt at arbejde med information og hindrer overgangen til åbent parametrisk design. Derfor har denne del fået sin egen plads i bogens struktur. Hvis CAD (BIM) ikke er en prioritet for dig, kan du springe direkte til næste del - "DEL VII: Datadrevet beslutningstagning, analyse, automatisering og maskinlæring".

Historien om fremkomsten af BIM og open BIM som markedsføringskoncepter for CAD- leverandører

Med fremkomsten af digitale data i 1990'erne blev computerteknologi ikke kun introduceret i forretningsprocesser, men også i designprocesser, hvilket førte til begreber som CAD (computerstøttet designsystem) og senere BIM (bygningsinformationsmodellering)

Men som enhver anden innovation er de ikke udviklingens slutpunkt. Koncepter som BIM er blevet en vigtig milepæl i byggebranchens historie, men før eller siden må de vige pladsen for bedre værktøjer og tilgange, der bedre kan imødekomme fremtidens udfordringer.

> Overvældet af indflydelsen fra CAD-leverandørerne og forvirret af kompleksiteten i sin egen implementering kan det godt være, at konceptet BIM, som dukkede op i 2002, ikke vil opleve sit 30-års jubilæum, som en rockstjerne, der blussede kraftigt op, men hurtigt forsvandt. Årsagen er enkel: Dataforskernes krav ændrer sig hurtigere, end CAD-leverandørerne kan tilpasse sig dem.

Stillet over for en mangel på kvalitetsdata kræver fagfolk i byggebranchen i dag interoperabilitet på tværs af platforme og adgang til åbne data fra CAD- projekter for at forenkle deres analyse og behandling. Kompleksiteten af CAD-data og den forvirrende behandling af CAD-data har en negativ indvirkning på alle, der er involveret i byggeprocessen: designere, projektledere, bygningsarbejdere på byggepladsen og i sidste ende bygherren.

I stedet for et fuldgyldigt datasæt til drift modtager kunden og investoren i dag containere i CAD- formater, der kræver komplekse geometriske kerner, forståelse af dataskemaer, årligt opdateret API -dokumentation og specialiseret CAD-software (BIM) for at arbejde med dataene. Samtidig forbliver en stor del af designdataene ubrugte.

I dagens design- og byggeverden fører kompleksiteten i adgangen til CAD-data til overengineering af projektledelsen. Mellemstore og store virksomheder, der arbejder med CAD-data eller udvikler BIM -

løsninger, er enten tvunget til at opretholde tætte relationer med CAD-leverandører -løsninger for at få adgang til data via API'er, eller omgå CAD-leverandørernes begrænsninger ved at bruge dyre SDK-konvertere til reverse-engineering, for at få adgang til åbne data [75].

> Den proprietære datatilgang er forældet og lever ikke længere op til kravene i dagens digitale miljø. Fremtiden vil dele virksomhederne op i to typer: dem, der bruger åbne data effektivt, og dem, der forlader markedet.

Begrebet BIM (Building Information Modeling) dukkede op i byggebranchen med udgivelsen af en af de store CAD-leverandørers Whitepaper BIM [54] i 2002, og suppleret med det maskintekniske begreb BOM (Bills of Materials) stammer det fra den parametriske tilgang til oprettelse og behandling af projektdata (Fig. 6.1-1). Den parametriske tilgang til oprettelse og behandling af designdata var en af de første, der blev implementeret i Pro-E-systemet til maskinteknisk design (MCAD). Dette system blev en prototype [111] for mange moderne CAD løsninger, herunder dem, der bruges i byggebranchen i dag.

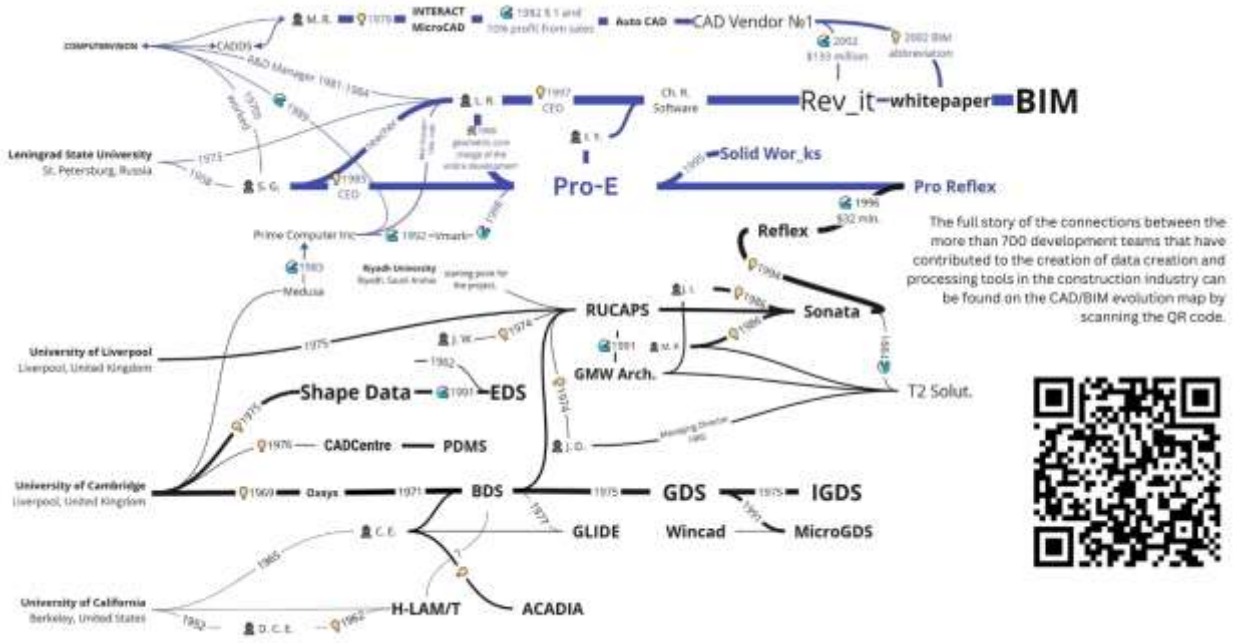


Fig. 6.1-1 Kort over BIM-konceptets historie og lignende koncepter.

Journalister og AEC-konsulenter, som promoverede CAD-værktøjer -leverandører indtil begyndelsen af 2000'erne, skiftede deres opmærksomhed til Whitepaper BIM fra 2002. Det var hvidbogen BIM 2002-2004 og de artikler, der blev offentliggjort i 2002, 2003, 2005 og 2007, som spillede en central rolle i populariseringen af BIM-konceptet i bygge- og anlægsbranchen [112].

> *Bygningsinformationsmodellering er en strategi..... [CAD-leverandørens firmanavn] for at anvende informationsteknologi i byggebranchen.*
>
> - BIM Whitepaper, 2002 [60].

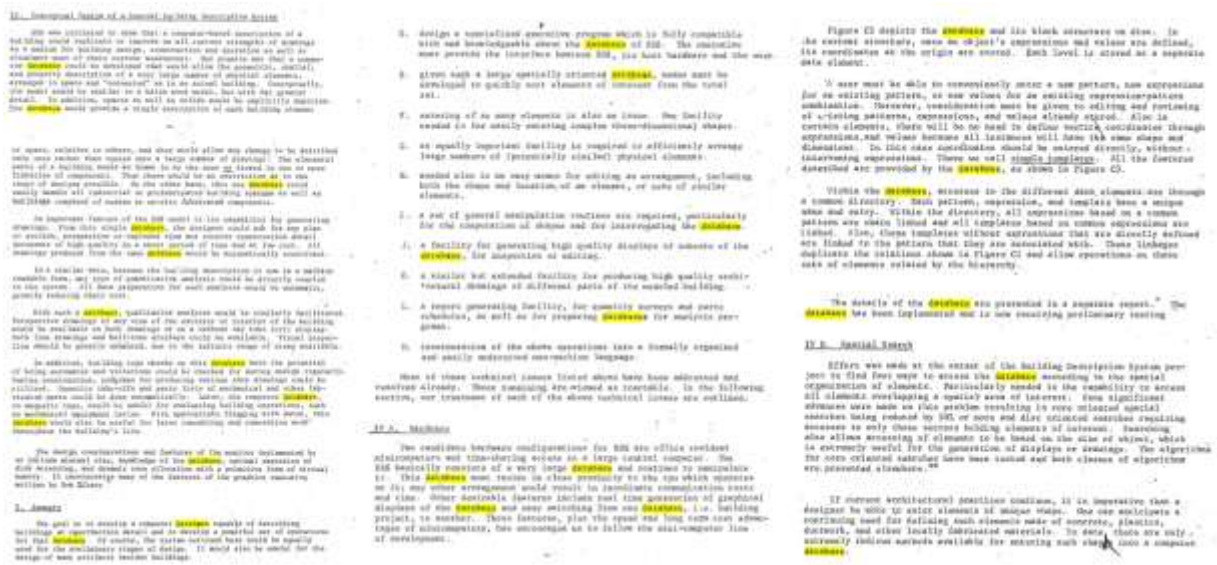
I midten af 2000'erne begyndte "forskere" at forbinde BIM- -konceptet, der blev offentliggjort af CAD- -leverandøren i 2002, med tidligere videnskabelige værker, såsom Charles Eastmans BDS, som blev grundlaget for systemer som GLIDE, GBM, BPM og RUCAPS. I sit pionerarbejde Building Description System (1974) lagde Charles Eastman det teoretiske grundlag for moderne informationsmodellering. Udtrykket "database " optræder 43 gange i hans arbejde (Fig. 6.1-2) - oftere end noget andet, bortset fra ordet "bygning".

> Eastmans hovedidé var, at alle oplysninger om en bygning - fra geometri til elementernes egenskaber og deres indbyrdes forhold - skulle lagres i en enkelt struktureret database. Det er ud fra denne database, at tegninger, specifikationer, beregninger og overholdelse af regler automatisk kan genereres og analyseres. Eastman kritiserede udtrykkeligt tegninger som en forældet og overflødig kommunikationsmetode og pegede på dobbeltarbejde, problemer med opdatering og behovet for manuelle opdateringer, når der foretages ændringer. I stedet foreslog han en enkelt digital model i en database, hvor enhver ændring foretages én gang og automatisk afspejles i alle visninger.

Det er bemærkelsesværdigt, at Eastman i sit koncept ikke satte visualisering i centrum. Information var centralt i hans system: parametre, relationer, attributter, analyse- og automatiseringsfunktioner. Tegninger var i hans forståelse kun en af formerne for visning af data fra databasen, ikke den primære kilde til designinformation.

I det første whitepaper om BIM fra den førende CAD-leverandør blev udtrykket "database " brugt lige så ofte som i Charles Eastmans BDS - 23 gange [60] over syv sider og var et af de mest populære ord i dokumentet efter "bygning", "information", "modellering" og "design". Men i 2003 optrådte termen "database" kun to gange i lignende dokumenter [61], og i slutningen af 2000'erne var emnet databaser stort set forsvundet fra diskussionen om designdata. Derfor blev konceptet med "en enkelt integreret database til visuel og kvantitativ analyse" aldrig helt realiseret.

Byggebranchen er således gået fra Charles Eastmans progressive BDS-koncept med dets vægt på databaser og Samuel Geisbergs ideer om automatisk opdatering af designdata fra databaser i maskinteknikproduktet Pro-E (forgængeren for de populære CAD -løsninger, der bruges i byggeriet i dag) til den nuværende markedsførte BIM, hvor datahåndtering gennem databaser næppe nævnes, på trods af at dette var konceptet bag den oprindelige teoretiske



Figur 6.1-2 I BDS-konceptet, der blev beskrevet af Charles Eastman i 1974, blev udtrykket "Database" (fremhævet med gult) brugt 43 gange.

BDS og lignende koncepter frem til 2000'erne blev udviklet som en digital database over bygninger snarere end som et visualiseringsværktøj. BIM blev i 2002 et designværktøj, hvor databasen trådte i baggrunden. Hvad har vi mistet i overgangen fra BDS og lignende koncepter i 1990'erne til BIM i midten af 2010'erne?

- Åbne databaser: BDS og andre lignende koncepter lagde vægt på analyse, BIM lagde vægt på design.
- Fleksibilitet til at arbejde med data: BDS lagde vægt på dataanalyse, BIM lagde vægt på processer, der skulle baseres på uklare data.
- Gennemsigtighed: BDS var tænkt som en åben integreret database, mens CAD-leverandører i BIM har gjort deres databaser helt lukkede og har kæmpet uden held i 20 år mod reverse engineering-værktøjer, der åbner proprietære formater.

I løbet af de sidste 30 år har designere aldrig haft adgang til en "integreret database", og efter tyve års markedsføringseufori omkring BIM -værktøjer er byggebranchen begyndt at indse konsekvenserne af denne dille.

BIM's virkelighed: I stedet for integrerede databaser - lukkede modulære systemer

I stedet for at fokusere på data, strukturere dem og integrere dem i ensartede processer, er brugere af CAD - (BIM-) systemer tvunget til at arbejde med et fragmenteret sæt af proprietære løsninger, der hver især dikterer deres egne spilleregler:

- **Den fælles database**, som blev diskuteret i den første BIM-hvidbog**, er fortsat en myte.** På trods af højlydte påstande er adgangen til data stadig begrænset og fordelt på lukkede systemer.
- **BIM -modeller er blevet et lukket økosystem** snarere end et værktøj. I stedet for gennemsigtig informationsudveksling er brugerne tvunget til at betale for abonnementer og bruge proprietære API'er.

- **Data tilhører leverandørerne, ikke brugerne.** Projektinformation er låst fast i proprietære formater eller cloud-tjenester i stedet for at være tilgængelig i åbne og uafhængige formater.

Konstruktører og projektledere har ofte ikke adgang til CAD-databasen -systemerne eller til det format, som deres egne projektdata er lagret i. Det gør det umuligt hurtigt at verificere oplysninger eller formulere krav til datastruktur og -kvalitet (fig. 6.1-3). Adgang til sådanne data kræver et helt sæt specialiserede programmer, der er forbundet via API'er og plug-ins, hvilket fører til overdreven bureaukratisering af processerne i byggebranchen. I mellemtiden bruges disse data samtidig af dusinvis af informationssystemer og hundredvis af specialister.

> *Vi er nødt til at kunne administrere alle disse data [CAD (BIM)], gemme dem digitalt og sælge livscyklus- og processtyringssoftware,* ***for for hver ingeniør*** *[designer], der skaber noget [i CAD-software]****, er der ti personer, der*** *arbejder med disse data" [41].*
>
> - CEO for CAD - den leverandør, der skabte BIM-konceptet, 2005.

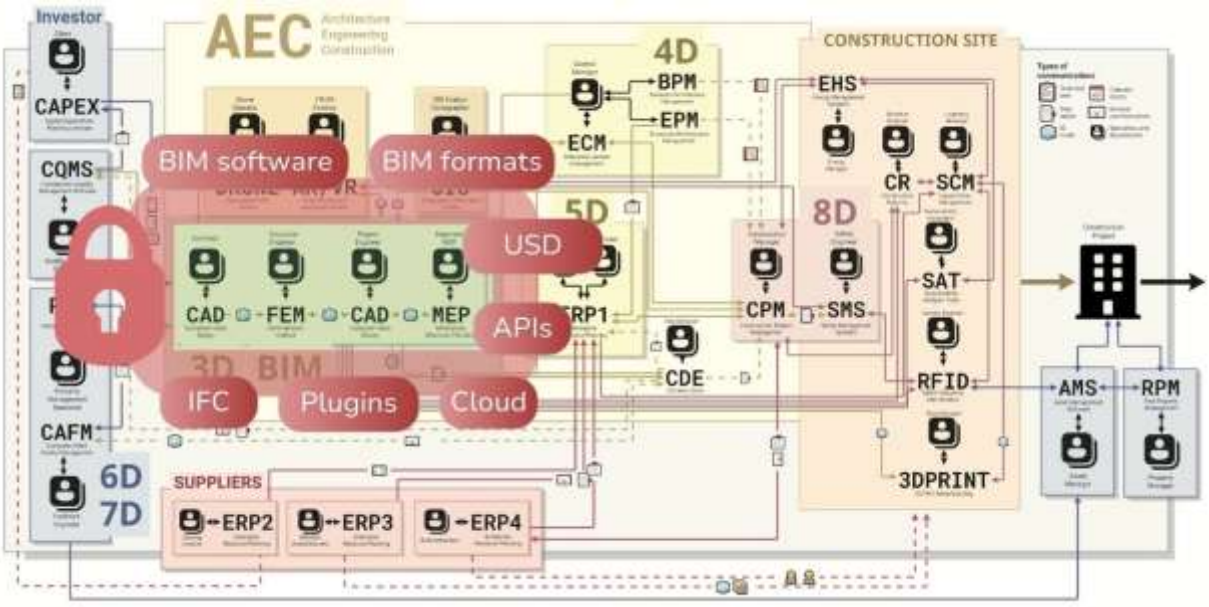


Fig. 6.1-3 CAD- (BIM-) databaser er stadig et af de sidste lukkede systemer for IT-afdelinger og dataadministratorer i byggeriets økosystem.

Når det bliver tydeligt, at BIM mere er et middel til at kommercialisere databaser end et fuldgyldigt værktøj til databasestyring, opstår der et logisk spørgsmål: Hvordan kan vi genvinde kontrollen over dataene? Svaret er at bruge åbne datastrukturer, hvor brugeren, ikke softwareleverandøren, bliver ejer af informationerne.

Brugere og løsningsudviklere i byggebranchen vil, ligesom deres kolleger i andre brancher, uundgåeligt bevæge sig væk fra den vage softwareleverandørterminologi, der har domineret de sidste 30 år, og fokusere på de vigtigste aspekter af digitalisering - "data" og "processer".

Tilbage i slutningen af 1980'erne blev nøgleområdet for den digitale udvikling i byggeriet præsenteret som et spørgsmål om dataadgang og projektinformationsstyring. Med tiden har fokus dog flyttet sig. I stedet for at udvikle gennemsigtige og tilgængelige tilgange til at arbejde med data blev IFC-formatet og det åbne BIM-koncept aktivt fremmet som forsøg på at aflede specialisternes opmærksomhed fra emnerne projektdatabasehåndtering.

Fremkomsten af det åbne format IFC i byggebranchen

Det såkaldte åbne format IFC (Industry Foundation Classes) er placeret som en standard, der skal sikre interoperabilitet mellem forskellige CAD-systemer (BIM -). Det blev udviklet inden for rammerne af organisationer, der blev oprettet og kontrolleret af store CAD-leverandører. Baseret på IFC-formatet udviklede to CAD- virksomheder i 2012 markedsføringskonceptet OPEN BIM[63].

> **IFC** (Industry Foundation Classes) er en åben standard for dataudveksling i byggebranchen, der er designet til at sikre interoperabilitet mellem forskellige CAD - (BIM-) systemer.

Open BIM - konceptet indebærer, at man arbejder med oplysninger fra CAD-databaser og udveksler oplysninger mellem systemer via et åbent format til udveksling af CAD-data - IFC.

> *Open BIM-programmet er en markedsføringskampagne iværksat af... [1 CAD-leverandør],... [2 CAD-leverandør] og andre virksomheder for at tilskynde til og lette global koordineret fremme af OPEN BIM-konceptet i hele AEC-industrien med konsekvent kommunikation og fælles branding til rådighed for programmets deltagere.*
>
> - Fra CAD-leverandørens hjemmeside, OPEN BIM-programmet, 2012 [113].

IFC blev tilpasset af det tekniske universitet i München fra det maskintekniske format STEP i slutningen af 1980'erne, og blev senere registreret af en stor designvirksomhed og en stor CAD- leverandør for at danne IAI (Industry Alliance for Interoperability) i 1994 [114] (Fig. 6.1-4). IFC -formatet blev udviklet for at sikre interoperabilitet mellem forskellige CAD -systemer og var baseret på principperne i det maskintekniske format STEP, som igen udsprang af IGES -formatet, der blev skabt i 1979 af en gruppe CAD-brugere og -leverandører med støtte fra NIST (The National Institute of Standards and Technology) og det amerikanske forsvarsministerium [115].

Men den komplekse struktur i IFC, den tætte afhængighed af den geometriske kerne samt uoverensstemmelser i implementeringen af formatet i forskellige softwareløsninger har ført til mange problemer i den praktiske anvendelse. Lignende vanskeligheder - tab af detaljer, begrænsning af nøjagtighed og behovet for at bruge mellemformater - blev tidligere mødt af maskintekniske specialister, når de arbejdede med IGES- og STEP-formater, som IFC opstod fra.



Fig. 6.1-4 Kort over forbindelser mellem udviklingsteams og CAD-produkter (BIM) [116].

> I 2000 udgiver den samme CAD -leverandør, som registrerede IFC -formatet og skabte IAI-organisationen (senere bS), hvidbogen "Integrated Design and Manufacturing: Benefits and Rationale". [65]. Papiret understregede vigtigheden af at opretholde fuld datagranularitet ved udveksling mellem programmer inden for det samme system uden at bruge neutrale formater som IGES, STEP [identisk med IFC]. I stedet blev det foreslået, at applikationer skulle have direkte adgang til den underliggende CAD-database for at forhindre tab af informationens nøjagtighed.

I 2002 køber den samme CAD-leverandør produktet parametrisk stykliste (fig. 3.1-18, flere detaljer i i tredje del) og danner på grundlag heraf BIM-konceptet. Som følge heraf anvendes kun lukkede CAD-formater eller IFC-formatet (STEP) til udveksling af byggeprojektdata, hvis begrænsninger CAD-leverandøren selv skrev om i 2000, da han bragte dette format til byggebranchen.

En detaljeret historie om samspillet mellem mere end 700 udviklingsteams, der er involveret i skabelsen af værktøjer til at skabe og behandle konstruktionsdata, præsenteres på kortet "The Evolution of CAD (BIM)" [116]. [116].

Den åbne form IFC består af en geometrisk beskrivelse af designelementerne og en beskrivelse af metainformationen. Der bruges forskellige metoder til at repræsentere geometri i IFC-format, såsom CSG og Swept Solids: Den parametriske repræsentation BREP er dog blevet den førende standard for overførsel af elementgeometri i IFC-format, da dette format understøttes ved eksport fra CAD- (BIM-) programmer og

giver mulighed for potentiel redigering af elementer ved import af IFC tilbage til CAD-programmer.

IFC formatproblem afhængigt af den geometriske kerne

I de fleste tilfælde, når geometrien i IFC er defineret parametrisk (BREP), bliver det umuligt at visualisere eller hente geometriske egenskaber såsom volumen eller areal af projektenheder med kun en IFC-fil, fordi det i dette tilfælde er nødvendigt med en geometrikerne (figur 6.1-5) for at arbejde med og visualisere geometrien, hvilket i første omgang mangler.

> **Geometrikernen** er en softwarekomponent, der leverer grundlæggende algoritmer til at skabe, redigere og analysere geometriske objekter i CAD (CAD), BIM og andre tekniske applikationer. Den er ansvarlig for at opbygge 2D- og 3D -geometri og for operationer på den, såsom: Boolske operationer, udjævning, skæringspunkter, transformationer og visualisering.

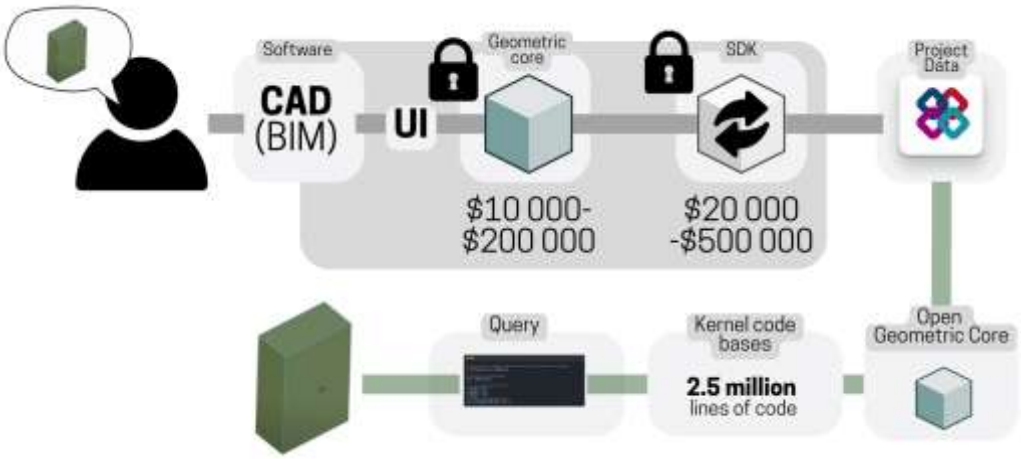


Fig. 6.1-5 Oprettelse af geometri via CAD- software foregår i dag via proprietære geometrikerner og SDK'er, som ofte ikke ejes af CAD-leverandørerne.

Hvert CAD-program og alle programmer, der arbejder med parametriske eller IFC-formater, har deres egen eller købte geometriske kerne. Og hvis der med primitive elementer i IFC -BREP format ikke er nogen problemer, og i programmer med forskellige geometriske kerner kan disse elementer vises på samme måde, men ud over problemer med forskellige motorer til geometriske kerner, er der nok elementer, der har deres egne særegenheder til korrekt visning. Dette problem diskuteres i detaljer i den internationale undersøgelse "A reference study of IFC software support" udgivet i 2019 [117].

> *De samme standardiserede datasæt giver modstridende resultater med få fælles mønstre, og der er konstateret alvorlige problemer med at understøtte standarden [IFC], sandsynligvis på grund af den meget høje kompleksitet i standarddatamodellen. Standarderne selv har en del af skylden her, da de ofte efterlader nogle detaljer udefineret, med høje frihedsgrader og forskellige mulige fortolkninger. De tillader høj kompleksitet i organiseringen og lagringen af objekter, hvilket ikke er befordrende for en effektiv universel forståelse, unikke implementeringer og konsistent datamodellering [117].*
>
> - Referenceundersøgelse af IFC-softwaresupport, 2021

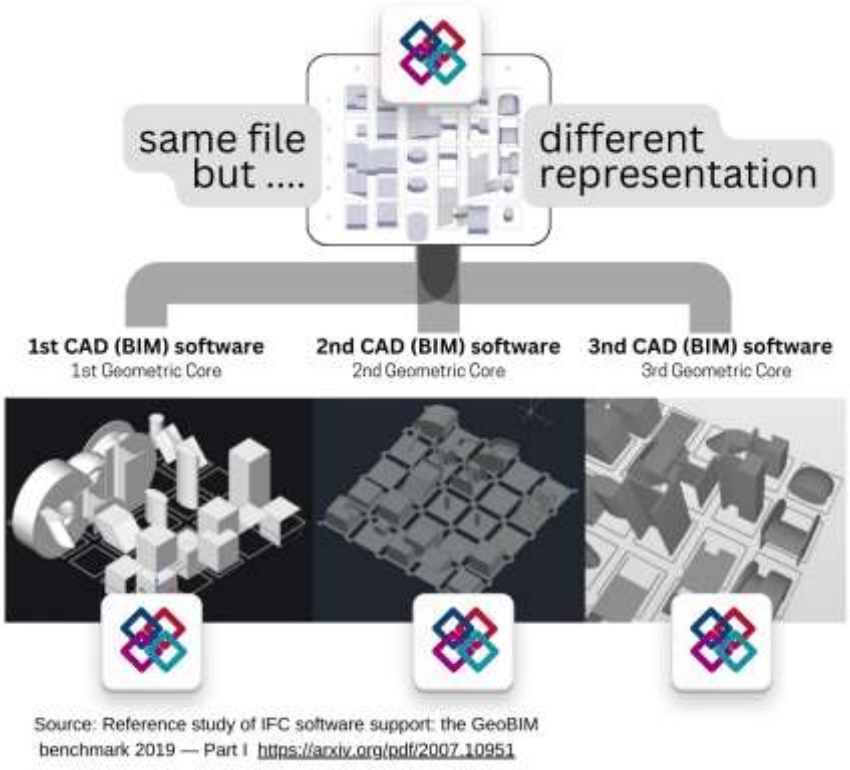


Fig. 6.1-6 Forskellige geometriske kerner giver forskellige repræsentationer af den samme geometri beskrevet parametrisk (baseret på [117]).

Den korrekte forståelse af "visse bestemmelser" er tilgængelig for betalende medlemmer af særlige organisationer, der udvikler IFC. Som følge heraf vil den, der ønsker at få adgang til vigtig viden om visse funktioner i IFC, forsøge at samarbejde med store CAD- leverandører eller at nå frem til en kvalitativ vurdering funktionerne ved egen forskning.

> *Du falder over et spørgsmål om import og eksport af data via IFC-formatet og spørger andre leverandører: "Hvorfor er der i IFC-filen oplysninger om parametrisk overførsel af lokaler? Den åbne specifikation siger ikke noget om det". Svar fra "mere vidende" europæiske leverandører: "Ja, det står der ikke, men det er tilladt".*
>
> - Fra interviewet med CAD 2021-udvikleren [118]

IFC beskriver geometrien gennem parametriske primitiver, men indeholder ikke en indlejret kerne - dens rolle udfyldes af CAD-programmet, som kompilerer geometrien gennem geometrikernen. Geometrikernen udfører de matematiske beregninger og definerer skæringspunkterne, mens IFC'en kun leverer data til fortolkning. Hvis IFC'en indeholder forkerte flader, kan forskellige programmer med forskellige geometrikerne enten ignorere dem eller producere fejl, afhængigt af kernen.

For at arbejde med IFC-formatet er det derfor nødvendigt at besvare hovedspørgsmålet, som det er svært at finde et entydigt svar på - hvilket værktøj, med hvilken geometrisk kerne, skal man bruge for at få den datakvalitet, som projektet oprindeligt havde i CAD-programmet, hvorfra IFC blev hentet?

Problemer med datakvalitet og kompleksiteten i IFC -formatet tillader ikke direkte brug af designdata til procesautomatisering, analyse og databehandling, hvilket ofte fører udviklere til det uundgåelige behov for at bruge lukkede CAD -løsninger med "kvalitets"-adgang til data[63], hvilket blev skrevet om af leverandøren selv, der registrerede IFC i 1994 [65].

Alle særegenheder ved kortlægning og generering af IFC parametre i geometrikernen kan kun realiseres af store teams af udviklere, der har erfaring med at arbejde med geometrikerner. Derfor er den nuværende praksis med IFC-formatets særegenheder og kompleksitet primært til gavn for CAD- leverandører og har meget til fælles med store softwareleverandørers strategi "adopt, extend, destroy", når standardens voksende kompleksitet faktisk skaber barrierer for små markedsaktører [94].

De store leverandørers strategi kan være tilpasse åbne standarder, tilføje proprietære udvidelser og funktioner for at skabe brugerafhængighed af deres produkter og derefter presse konkurrenterne ud.

IFC-formatet, der er tænkt som en universel bro mellem forskellige CAD- (BIM-) systemer, fungerer i virkeligheden som en indikator for kompatibilitetsproblemer mellem de geometriske kerner i forskellige CAD-platforme, ligesom STEP -formatet, som det oprindeligt udsprang af.

Som følge heraf er en fuld implementering af IFC ontologien i dag mulig for store CAD-leverandører, som kan investere betydelige ressourcer i at understøtte alle enheder og deres mapping til deres egen interne geometrikerne, som ikke findes for IFC som standard. Store leverandører har også mulighed for indbyrdes at koordinere tekniske detaljer om funktioner, som måske ikke er tilgængelige for selv den mest aktive deltager i IFC-formatets udviklingsorganisationer.

> For små uafhængige teams og open source-projekter, der stræber efter at støtte udviklingen af interoperable formater, bliver manglen på en intern geometrikerne et alvorligt problem. Uden den er det stort set umuligt at tage højde for alle de forskellige finesser og nuancer, der er forbundet med dataudveksling på tværs af platforme.

Med udviklingen af det parametriske IFC-format og det åbne BIM-koncept er diskussionerne i byggebranchen om ontologiens og semantikkens rolle i data- og processtyring blevet intensiveret.

Optræden i konstruktionen af emnet semantik og ontologi

Takket være ideerne om det semantiske internet slutningen af 1990'erne og indsatsen fra organisationer, der er involveret i udviklingen af IFC-formatet, er semantik og ontologier blevet et af de vigtigste elementer i den standardisering, der diskuteres i byggebranchen i midten af 2020'erne.

> Semantiske teknologier er ensretning, standardisering og ændring af store mængder heterogene data og implementering af komplekse søgninger.

OWL (Web Ontology Language), repræsenteret som RDF-grafer -triplets (Resource Description Framework) (Fig. 6.1-7), bruges til at lagre semantiske data. OWL henviser til grafdatamodeller, hvis typer vi diskuterede mere detaljeret i kapitlet "Datamodeller: datarelationer og relationer mellem elementer".

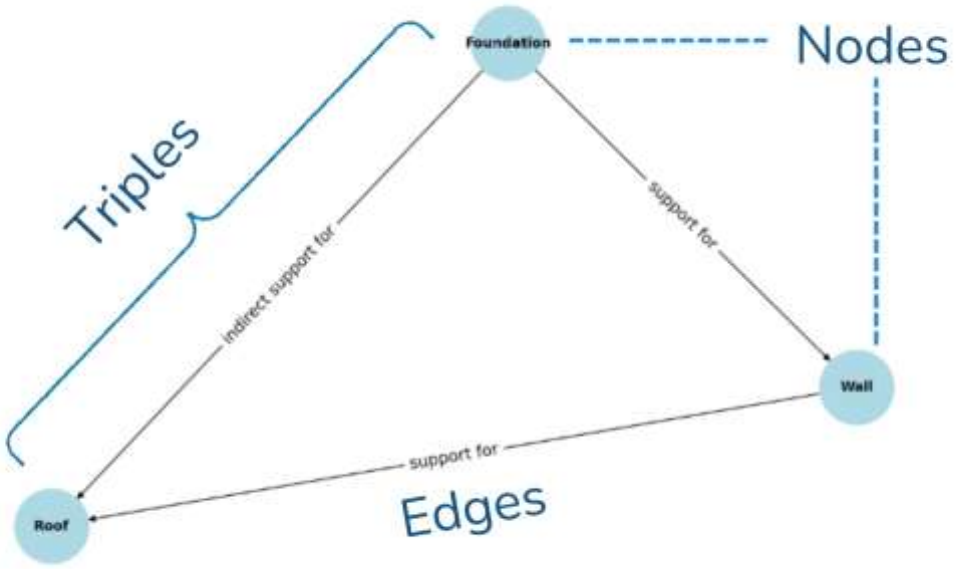


Fig. 6.1-7 RDF-datamodel: Nodes, Edges og Triples illustrerer relationerne mellem byggestenene.

Teoretisk set gør den logiske slutning af risoner (automatisk logisk slutningssoftware) det muligt at udlede nye udsagn fra ontologier. Hvis bygningsontologien f.eks. registrerer, at "et fundament er en støtte for en væg" og "en væg er en støtte for et tag" (fig. 6.1-7), kan risoneren automatisk udlede, at "et fundament er en støtte for et tag".

En sådan mekanisme er nyttig til at optimere dataanalyse, fordi den undgår eksplicit at foreskrive alle afhængigheder. Den skaber dog ikke ny viden, men identificerer og strukturerer kun allerede kendte fakta.

> Semantik skaber ikke ny mening eller viden i sig selv og er ikke bedre end andre teknologier til datalagring og -behandling i dette aspekt. At repræsentere data fra relationsdatabaser som tripletter gør dem ikke mere meningsfulde. At erstatte tabeller med grafstrukturer kan være nyttigt for at forene datamodeller, gøre det nemt at hente data og sikre redigering, men det gør ikke dataene "klogere" - computeren begynder ikke at forstå deres indhold bedre.

Logiske relationer i data kan organiseres uden komplekse semantiske teknologier (Fig. 6.1-8). Traditionelle relationsdatabaser (SQL) såvel som CSV eller XLSX formater gør det muligt at opbygge lignende afhængigheder. I en søjledatabase kan man f.eks. tilføje et "tagstøtte"-felt og automatisk forbinde taget med fundamentet, når man opretter en væg. Denne tilgang er implementeret uden brug af RDF, OWL, grafer eller risonere, og forbliver en enkel og effektiv løsning til lagring og analyse af data.

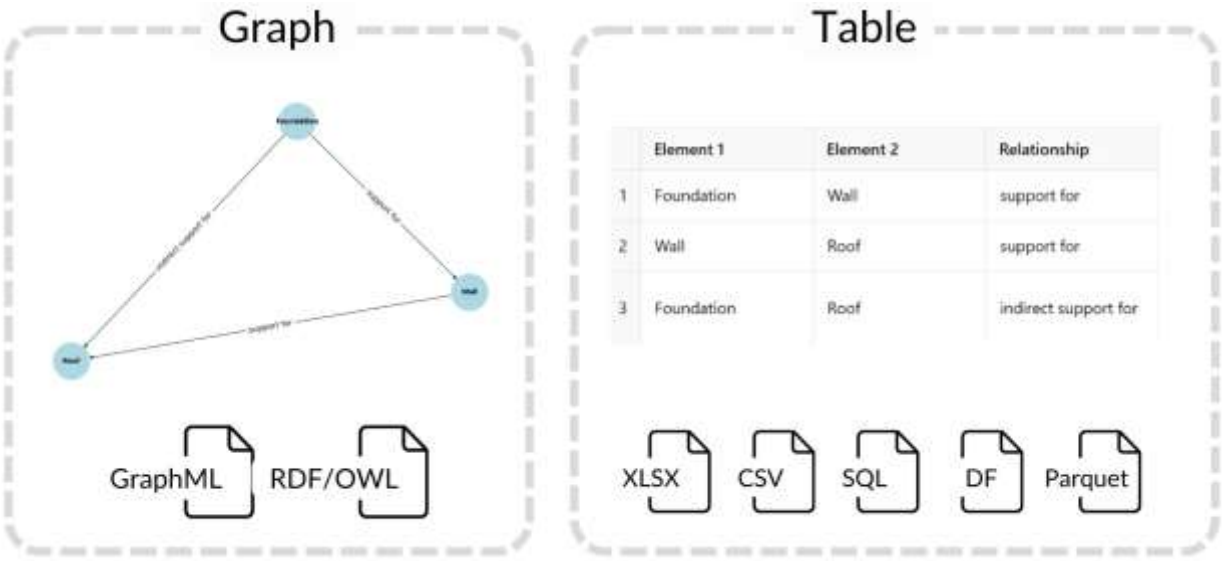


Fig. 6.1-8 Sammenligning af graf- og tabeldatamodeller til repræsentation af de samme logiske relationer.

Beslutningen fra en række store byggevirksomheder og IFC-formatudviklingsorganisationen [94] om at følge det semantiske webkoncept, som virkede lovende i slutningen af 1990'erne, har haft en betydelig indflydelse på udviklingen af standarder i byggebranchen.

Det paradoksale er imidlertid, at selve konceptet med det semantiske web, der oprindeligt var beregnet til internettet, ikke er blevet udbredt selv i sit oprindelige miljø. På trods af udviklingen af RDF og OWL er det fuldgyldige semantiske web ikke dukket op i sin oprindelige udformning, og det er allerede usandsynligt, at det vil blive skabt.

Hvorfor semantiske teknologier ikke lever op til forventningerne i byggebranchen

Andre brancher har oplevet begrænsninger i teknologierne til at udnytte semantik. I spilbranchen har forsøg på at beskrive spilobjekter og deres interaktioner gennem ontologier vist sig at være ineffektive på grund af den høje forandringsdynamik. Derfor har man foretrukket enklere dataformater som XML og JSON sammen med algoritmiske løsninger. Situationen var den samme i ejendomssektoren: På grund af regionale forskelle i terminologi og hyppige markedsændringer viste brugen af ontologier sig at være alt for kompleks, mens enkle databaser og standarder som RETS [119] var bedre i stand til at klare udfordringerne med dataudveksling.

> Tekniske vanskeligheder som f.eks. kompleksiteten i opmærkningen, arbejdskrævende support og lav motivation hos udviklerne bremsede udbredelsen af det semantiske web og i andre økonomiske sektorer. RDF (Resource Description Framework) blev ikke en massestandard, og ontologier viste sig at være for komplekse og økonomisk ubegrundede.

Derfor blev den ambitiøse idé om at skabe et globalt semantisk web ikke til noget. Selvom nogle elementer af teknologien, såsom ontologier og SPARQL, har fundet vej til virksomhedsløsninger, er det oprindelige mål om at skabe en enkelt omfattende datastruktur ikke blevet nået.

Konceptet med et internet, hvor computere er i stand til at forstå indholdet, har vist sig at være teknisk vanskeligt og kommercielt urentabelt. Det er grunden til, at de virksomheder, der støttede idéen, til sidst reducerede brugen til enkelte nyttige værktøjer og overlod RDF og OWL til højt specialiserede virksomhedsbehov snarere end til internettet som helhed. En analyse af Google Trends (figur 6.1-9) i løbet af de sidste 20 år tyder på, at der måske ikke er flere udsigter for det semantiske web.

> *Der er ingen grund til at multiplicere enheder unødigt. Hvis der er flere logisk konsistente forklaringer på et fænomen, som forklarer det lige godt, bør man alt andet lige foretrække den simpleste af dem.*
>
> ragekniv

Her opstår et logisk spørgsmål: Hvorfor overhovedet bruge tripletter, risers og SPARQL i konstruktionen, når man kan behandle data ved hjælp af populære strukturerede forespørgsler (SQL, Pandas, Apache®)? I virksomhedsapplikationer er SQL standarden for at arbejde med databaser. SPARQL kræver derimod komplekse grafstrukturer og specialiseret software og tiltrækker ifølge tendenser i Google ikke udviklernes interesse.

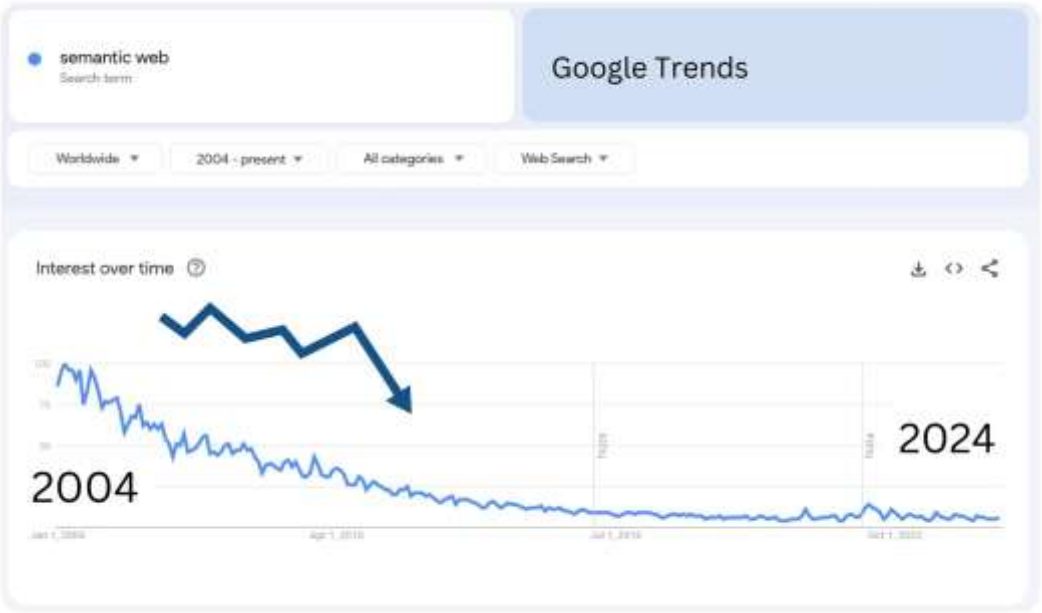


Fig. 6.1-9 Interesse i "semantisk internet"-forespørgsler ifølge Google-statistikker.

Grafdatabaser og klassifikationstræer kan være nyttige i nogle tilfælde, men deres anvendelse er ikke altid berettiget til de fleste hverdagsopgaver. Derfor giver oprettelse af vidensgrafer og brug af semantiske webteknologier kun mening, når det er nødvendigt at forene data fra forskellige kilder eller gennemføre komplekse logiske konklusioner.

> At gå fra tabeller til grafiske datamodeller forbedrer søgningen og ensretter informationsstrømmen, men gør ikke dataene mere meningsfulde for maskiner. Spørgsmålet er ikke, om semantiske teknologier skal bruges, men hvor de virkelig gør en forskel. Før du implementerer ontologi, semantik og grafdatabaser i din virksomhed, skal du finde ud af, hvilke virksomheder der allerede bruger disse teknologier med succes, og hvor de har fejlet.

På trods af ambitiøse forventninger blev semantiske teknologier aldrig en universel løsning til strukturering af data i byggebranchen. I praksis har disse teknologier ikke ført til en universel løsning, men har kun tilføjet ny kompleksitet, og disse bestræbelser er et ekko af de urealiserede ambitioner for det semantiske internetkoncept, hvor forventningerne langt oversteg virkeligheden.

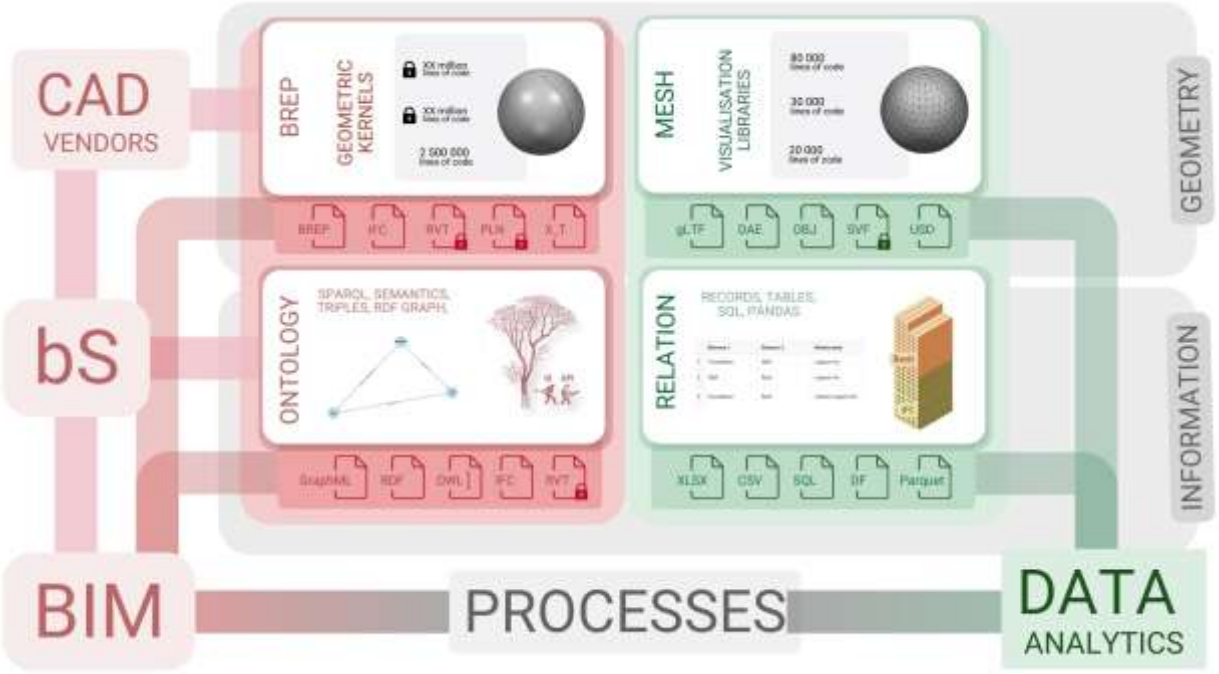


Fig. 6.1-10 Geometri og information i byggeprocesser: fra komplekse CAD og BIM- systemer til forenklede data til analyse.

Mens det semantiske webs fiaskoer inden for IT er blevet opvejet af fremkomsten af nye teknologier (big data, IoT, machine learning, AR/VR), har byggebranchen ikke sådanne muligheder.

Ud over udfordringerne med at bruge koncepter til at kommunikere datarelationer mellem projektelementer er der stadig et grundlæggende problem - selve tilgængeligheden af disse data. Byggebranchen er stadig domineret af lukkede systemer, hvilket gør det vanskeligt at arbejde med data, dele information og forbedre proceseffektiviteten.

Det er dataenes lukkede karakter, der er ved at blive en af de vigtigste barrierer for udviklingen af digitale løsninger i byggeriet. I modsætning til IT-industrien, hvor åbne og harmoniserede dataformater er blevet standarden, bruger hver software sit eget format i CAD-sektoren (BIM), hvilket skaber lukkede økosystemer og kunstigt begrænser brugerne.

KAPITEL 6.2.
LUKKEDE PROJEKTFORMATER OG PROBLEMER MED INTEROPERABILITET

Lukkede data og faldende produktivitet: CAD-industriens blindgyde (BIM)

CAD -systemernes proprietære karakter har ført til, at hvert program har sit eget unikke dataformat, som enten er lukket og utilgængeligt udefra - RVT, PLN, DWG, NDW, NWD, SKP, eller er tilgængeligt i semistruktureret form gennem en ret kompleks konverteringsproces - JSON, XML (CPIXML), IFC, STEP og ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, osv..

De forskellige dataformater, som de samme data om de samme projekter kan gemmes i, har ikke kun forskellig struktur, men indeholder også forskellige versioner af den interne markup, som udviklere skal tage højde for for at sikre applikationskompatibilitet. For eksempel vil et CAD-format fra 2025 kunne åbnes i et CAD-program fra 2026, men det samme projekt vil aldrig kunne åbnes i alle de versioner af CAD-programmet, der måtte have været tilgængelige før 2025.

> Ved ikke at give direkte adgang til databaser skaber en softwareleverandør i byggebranchen ofte sit eget unikke format og værktøjer, som en fagperson (designingeniør eller dataansvarlig) skal bruge for at få adgang til, importere og eksportere data.

Som følge heraf hæver leverandørerne af grundlæggende CAD (BIM) og relaterede løsninger (f.eks. ERP/PMIS)) konstant priserne for at bruge produkterne, og almindelige brugere tvinges til at betale en "provision" på hvert trin i dataoverførslen ved hjælp af formater [63]: for at forbinde, importere, eksportere og arbejde med data, som brugerne selv har skabt.

Omkostningerne ved at få adgang til data i cloud storage fra populære CAD - (BIM-) produkter vil nå op på 1 dollar pr. transaktion i 2025 [120], og abonnementer på ERP -produkter til byggeri for mellemstore virksomheder når op på fem- og sekscifrede beløb om året [121].

> Essensen af moderne byggesoftware er, at det ikke er automatisering eller øget effektivitet, men ingeniørernes evne til at forstå en bestemt højt specialiseret software, der påvirker kvaliteten og omkostningerne ved databehandling af byggeprojekter samt overskuddet og den langsigtede overlevelse for virksomheder, der udfører byggeprojekter.

Den manglende adgang til databaser CAD - systemer, der bruges i dusinvis af andre systemer og hundredvis af processer [63], og den deraf følgende mangel på kvalitetskommunikation mellem individuelle specialister har ført til, at byggebranchen er en af de mest ineffektive sektorer i økonomien, hvad angår produktivitet [44].

I løbet af de sidste 20 år med CAD- (BIM-) designapplikationer, fremkomsten af nye systemer (ERP), nye byggeteknologier og materialer er produktiviteten i hele byggeindustrien faldet med 20% (figur 2.2-1), mens den samlede produktivitet i alle sektorer af økonomien, der ikke har store problemer med at få adgang til databaser og markedsføringslignende BIM-koncepter, er steget med 70% (96% i fremstillingsindustrien)

[122].



Fig. 6.2-1 På grund af isolationen og kompleksiteten af projektdata, som dusinvis af afdelinger og hundredvis af processer er afhængige af i byggebranchen, er beslutningshastigheden flere gange langsommere end i andre brancher.

Der er dog også enkelte eksempler på alternative tilgange til at skabe interoperabilitet mellem CAD-løsninger. Europas største byggefirma demonstrerer med SCOPE-projektet [123], der blev startet tilbage i 2018, hvordan det er muligt at gå ud over den klassiske logik i CAD- (BIM-) systemer. I stedet for at forsøge at underlægge sig IFC eller stole på proprietære geometrikerner, bruger SCOPE-udviklere API'er og SDKreverse engineering til at udtrække data fra forskellige CAD-programmer, konvertere dem til neutrale formater som OBJ eller CPIXML baseret på den eneste Open Source geometrikerne OCCT, og yderligere anvende dem til hundredvis af forretningsprocesser i bygge- og designvirksomheder. Men på trods af idéens progressivitet står sådanne projekter over for begrænsningerne og kompleksiteten i gratis geometriske kerner, og de forbliver stadig en del af lukkede økosystemer i en virksomhed, der reproducerer logikken i monovirksomhedsløsninger.

På grund af begrænsningerne i lukkede systemer og forskelle i dataformater samt manglen på effektive værktøjer til at forene dem, står virksomheder, der skal arbejde med CAD-formater, over for en ophobning af betydelige mængder data med varierende grader af struktur og lukkethed. Disse data bliver ikke brugt ordentligt og forsvinder i arkiver, hvor de for evigt forbliver glemte og ubrugte.

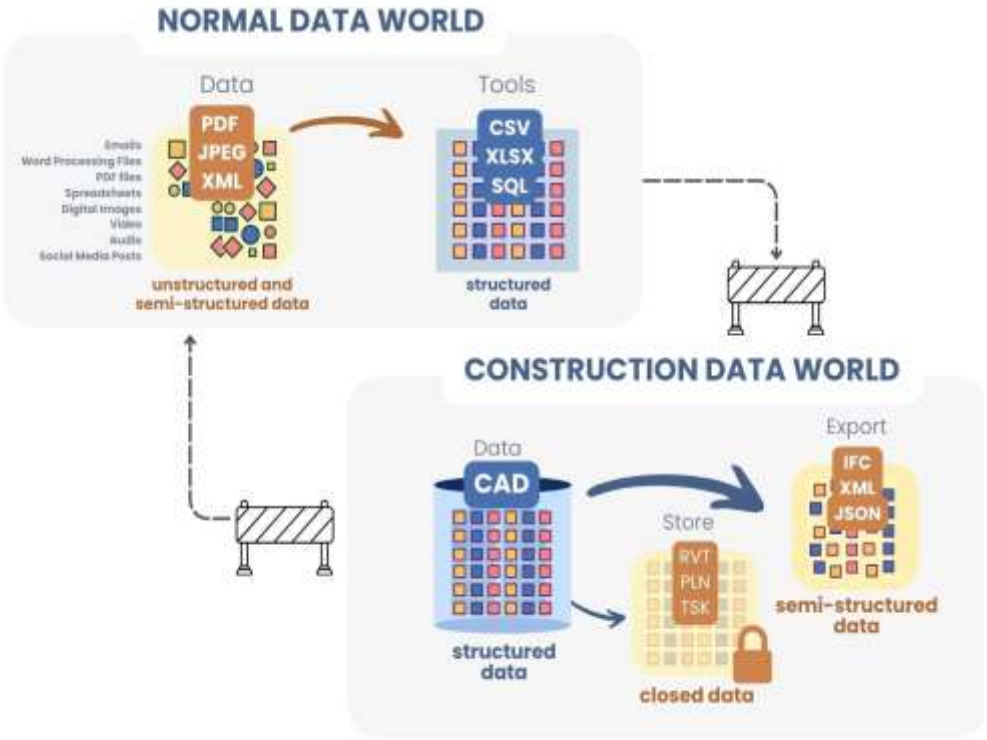
> Data, der er indhentet gennem en betydelig indsats i designfasen, bliver utilgængelige for videre brug på grund af deres kompleksitet og lukkede natur.

I løbet af de sidste 30 år har udviklere i byggebranchen derfor været tvunget til at stå over for det samme problem igen og igen: Hvert nyt lukket format eller hver ny proprietær løsning skaber et behov for at integrere med eksisterende åbne og lukkede CAD-systemer. Disse konstante forsøg på at sikre interoperabilitet

mellem forskellige CAD- og BIM-løsninger er kun med til at komplicere dataøkosystemet i stedet for at bidrage til dets forenkling og standardisering.

Myten om interoperabilitet mellem CAD-systemer

Hvis den vigtigste retning for udviklingen af interoperabilitet i CAD-miljøet i midten af 1990'erne var at bryde med det proprietære DWG-format - hvilket kulminerede med sejren for Open DWG-alliancen [75] og den faktiske åbning af det mest populære tegneformat for hele byggebranchen - så er fokus i midten af 2020'erne skiftet. En ny tendens vinder frem i byggebranchen: Talrige udviklingsteams fokuserer på at skabe såkaldte "broer" mellem lukkede CAD-systemer (closed BIM), IFC-formatet og åbne løsninger (open BIM). De fleste af disse initiativer er baseret på brugen af IFC-formatet og OCCT-geometrikernen, som skaber en teknisk bro mellem forskellige platforme. Denne tilgang ses som en lovende retning, der kan forbedre dataudveksling og interoperabilitet mellem softwareværktøjer betydeligt.

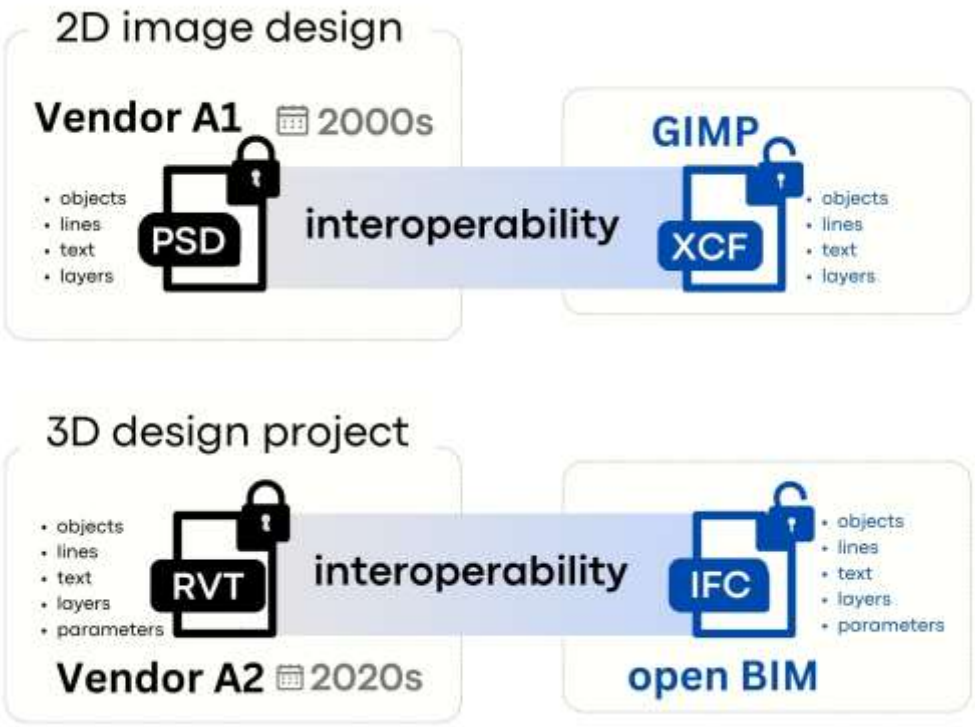


Figur 6.2-2 Mens andre brancher arbejder med åbne data, er byggebranchen nødt til at arbejde med lukkede eller løst strukturerede CAD-formater (BIM).

Denne tilgang har historiske paralleller. I 2000'erne forsøgte udviklere at overvinde dominansen fra den største leverandør af grafiske redigeringsprogrammer (2D world) ved at skabe en problemfri integration mellem deres proprietære løsning og den gratis Open Source - et alternativ til GIMP (Fig. 6.2-3). Dengang, som i dag inden for byggeri, handlede det om at forsøge at bygge bro mellem lukkede og åbne systemer, samtidig med at man bevarede komplekse parametre, lag og intern softwarelogik.

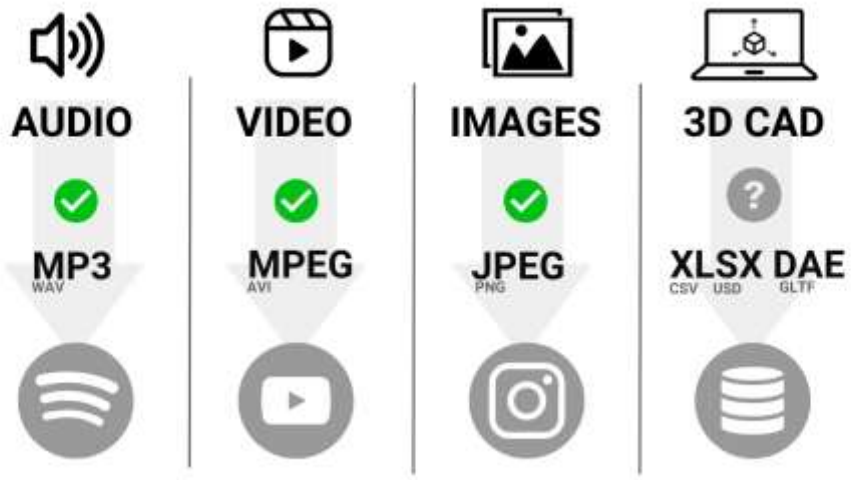
Men brugerne var faktisk på udkig efter enkle løsninger - flade, åbne data uden overdreven kompleksitet med lag og programparametre (svarende til den geometriske kerne i CAD). Brugerne søgte enkle og åbne

dataformater uden overdreven logik. JPEG, PNG og GIF blev sådanne formater inden for grafik. I dag bruges de i sociale netværk, på hjemmesider og i applikationer - de er nemme at behandle og fortolke, uanset platform eller softwareleverandør.



Figur 6.2-3 Interoperabiliteten mellem dataformater i byggeriet svarer til vejen fra forsøg på at fusionere en populær leverandørs proprietære produkt og Open Source GIMP i 2000'erne.

Resultatet er, at næsten ingen i billedindustrien i dag bruger lukkede formater som PSD eller åbne XCF til applikationer, sociale netværk som Facebook og Instagram eller som indhold på hjemmesider. I stedet bruger de fleste flade og åbne JPEG-, PNG- og GIF-formater, fordi de er nemme at bruge og har bred kompatibilitet. Åbne formater som JPEG og PNG er blevet standarden for billeddeling på grund af deres alsidighed og brede understøttelse, hvilket gør dem nemme at bruge på en række forskellige platforme. En lignende overgang kan observeres i andre udvekslingsformater, f.eks. video og lyd, hvor universelle formater som MPEG og MP3 fremhæves for deres komprimeringseffektivitet og brede kompatibilitet. En sådan bevægelse mod standardisering har forenklet deling og afspilning af indhold og information og gjort dem tilgængelige for alle brugere på tværs af flere platforme (figur 6.2-4).



Figur 6.2-4 Forenklede formater uden komplekse redigeringsfunktioner er blevet populære til deling og brug af data.

Lignende processer finder sted inden for 3D modellering. Enkle og åbne formater som USD, OBJ, glTF, DAE, DXF, SQL og XLSX bruges i stigende grad i projekter til dataudveksling uden for CAD-miljøet (BIM). Disse formater gemmer alle de nødvendige oplysninger, herunder geometri og metadata, uden at det er nødvendigt at anvende en kompleks BREP-struktur, geometrikerner eller leverandørspecifikke interne klassifikatorer. Proprietære formater som NWC, SVF, SVF2, CPIXML og CP2, der leveres af førende softwareleverandører, udfører også lignende funktioner, men forbliver lukkede i modsætning til åbne standarder.

> Det er bemærkelsesværdigt (og værd at minde om igen, som allerede nævnt i det foregående kapitel), at denne idé - afvisningen af neutrale og parametriske mellemformater som IGES, STEP og IFC - blev støttet tilbage i 2000 af den store CAD-leverandør, der skabte BIM-hvidbogen og registrerede IFC-formatet i 1994. I Whitepaper 2000 "Integrated Design and Manufacturing" [65] understreger CAD-leverandøren vigtigheden af naturlig adgang til CAD-databasen i softwaremiljøet uden behov for at bruge mellemliggende oversættere og parametriske formater for at bevare informationens fuldstændighed og nøjagtighed.

Byggebranchen er endnu ikke blevet enig om hverken værktøjer til at få adgang til CAD-databaser eller deres tvungne reverse engineering, eller om vedtagelsen af et fælles forenklet dataformat til brug uden for CAD-platforme (BIM). For eksempel bruger mange store virksomheder i Centraleuropa og tysktalende regioner, der opererer i byggesektoren, CPIXML -formatet i deres ERP -systemer [121]. Dette proprietære format, som er en slags XML, kombinerer CAD (BIM)-projektdata, herunder geometriske data og metadata, i en enkelt organiseret forenklet struktur. Store byggevirksomheder skaber også deres egne nye formater og systemer, som i SCOPE-projektet, som vi diskuterede i det foregående kapitel

Den lukkede logik i parametriske CAD-formater eller komplekse parametriske filer IFC (STEP) er overflødig i de fleste forretningsprocesser. Brugerne er på udkig efter forenklede og flade formater som USD, CPIXML, XML &OBJ, DXF, glTF, SQLlite, DAE &XLSX, som indeholder alle de nødvendige elementoplysninger, men som ikke er belastet af overflødig BREP-geometrilogik, afhængighed af geometrikerne og interne klassifikationer af specifikke CAD og BIM -produkter (Fig. 6.2-5).

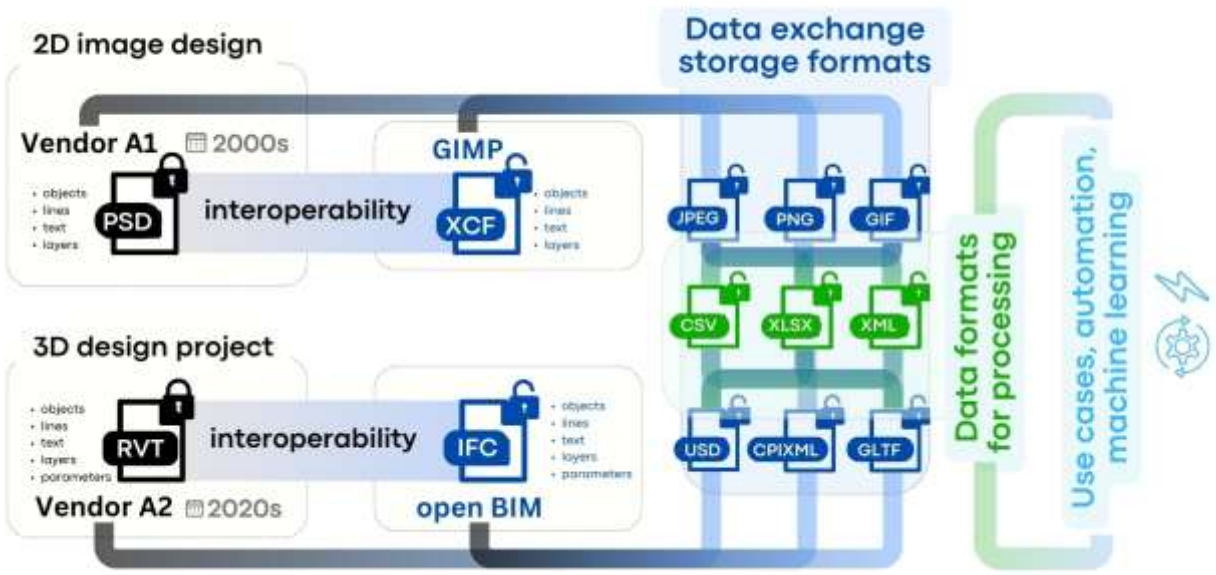


Fig. 6.2-5 I de fleste tilfælde vælger brugerne de simplest mulige formater, som er uafhængige af leverandørens programmer.

Fremkomsten af flade billedformater som JPEG, PNG og GIF, der er fri for den overflødige logik i leverandørernes interne motorer, har sat gang i udviklingen af tusindvis af interoperable løsninger til behandling og udnyttelse af grafik. Det har ført til fremkomsten af applikationer lige fra retoucherings- og filtreringsværktøjer til sociale netværk som Instagram, Snapchat og Canva, hvor disse forenklede data kan udnyttes uden at være bundet til en bestemt softwareudvikler.

> Standardisering og forenkling af design-CAD -formater vil stimulere fremkomsten af mange nye brugervenlige og uafhængige værktøjer til at arbejde med byggeprojekter.

At bevæge sig væk fra den komplekse logik i leverandørapplikationer, der er bundet til lukkede geometriske kerner, til universelle åbne formater baseret på biblioteker med forenklede elementer, skaber forudsætningerne for en mere fleksibel, gennemsigtig og effektiv datahåndtering. Det åbner også op for adgang til information for alle parter, der er involveret i byggeprocessen - fra designere til kunder og vedligeholdelsestjenester.

Ikke desto mindre er det meget sandsynligt, at CAD leverandører i de kommende år vil forsøge at flytte debatten om interoperabilitet og adgang til CAD-databaser igen. Det vil allerede handle om "nye" koncepter - såsom granulære data, intelligente grafer, "fødererede modeller", digitale tvillinger i cloud-lagre - samt oprettelsen af industrialliancer og standarder, der fortsætter vejen til BIM og åben BIM. På trods af den attraktive terminologi kan sådanne initiativer igen blive værktøjer til at fastholde brugere i proprietære økosystemer. Et eksempel er den aktive promovering af USD (Universal Scene Description)-formatet som den "nye standard" for CAD (BIM)-samarbejde på tværs af platforme fra 2023.

Gå til USD og granulære data m

Fremkomsten af AOUSD-alliancen [124] i 2023 markerer en vigtig vending i byggebranchen. Vi er vidne til begyndelsen på en ny virkelighed, formet af CAD-leverandører, i håndteringen af byggedata gennem flere væsentlige ændringer. Den første store ændring vedrører opfattelsen af CAD -data. Fagfolk, der er involveret i de tidlige stadier af konceptuelt design, indser i stigende grad, at det at skabe et design i et CAD-miljø kun er et startpunkt. De data, der genereres under designprocessen, bliver i sidste ende grundlaget for at analysere, betjene og administrere objekter. Det betyder, at de skal være tilgængelige og anvendelige i systemer, der rækker ud over traditionelle CAD-værktøjer.

Sideløbende sker der en revolution i de førende udvikleres tilgang. Branchens førende CAD- leverandør, skaberen af BIM-konceptet og IFC-formatet, tager en uventet drejning i sin strategi. Fra 2023 bevæger virksomheden sig væk fra den traditionelle lagring af data i separate filer, fokuserer på at arbejde med granulære (normaliserede og strukturerede) data og går over til en datacentreret tilgang [125].

> Leverandørerne følger de historiske tendenser i andre brancher: De fleste brugere har ikke brug for lukkede CAD-formater (svarende til PSD) eller komplekse parametriske IFC-filer (svarende til GIMP med laglogik). De har brug for enkle objektbilleder, der kan bruges i CAFM (Construction Instagram), ERP (Facebook) og tusindvis af andre processer, der er fyldt med Excel regneark og PDF dokumenter.

De nuværende tendenser i byggebranchen lægger muligvis op til et gradvist skift væk fra parametriske og komplekse formater til fordel for de mere universelle og uafhængige formater USD, GLTF, DAE, OBJ (med metainformation både inden for hybriden og i separate strukturerede eller løst strukturerede formater). Historiske ledere, herunder store designvirksomheder, som engang aktivt promoverede IFC i midten af 1990'erne, promoverer nu åbent det nye USD-format [93] og fremhæver dets enkelhed og alsidighed (Fig. 6.2-6). Den store udbredelse af USD i produkter, GLTF-kompatibiliteten og den aktive integration i værktøjer som Blender, Unreal Engine og Omniverse viser potentialet i begyndelsen af et nyt paradigme for arbejde med data. Sammen med populariteten af lokaliserede løsninger som det europæiske flade USD-format - CPIXML, der bruges i populære europæiske ERP'er kan det potentielt styrke USD's position i Centraleuropa. Organisationer, der er involveret i udviklingen af IFC-formatet, er allerede i gang med at tilpasse deres strategi til USD [126], hvilket kun bekræfter, at skiftet er uundgåeligt.

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

Figur 6.2-6 Sammenligning af tekniske specifikationer for IFC og USD-formater.

> På denne baggrund har USD potentiale til at blive de facto-standarden og lover at overvinde mange nuværende begrænsninger, primært relateret til kompleksiteten af eksisterende CAD - (BIM-) formater og afhængigheden af deres fortolkning af geometriske kerner.

I stedet for parametriske og komplekse CAD -formater og IFC - vil forenklede dataformater USD, gLTF, DAE, OBJ med elementmetainformation i CSV, XLSX, JSON, XML få en plads i byggebranchen på grund af deres enkelhed og fleksibilitet.

De aktuelle ændringer i byggebranchen ligner ved første øjekast et teknologisk gennembrud i forbindelse med overgangen fra den forældede IFC til den mere moderne USD. Det er dog værd at overveje, at den samme CAD-leverandør, som udviklede IFC, allerede i 2000 skrev om problemerne med IFC og behovet for

adgang til databasen [65], og nu aktivt fremmer overgangen til en ny standard - USD.

> Bag endnu en facade af "åbne data" USD og "nye" koncepter for granulær datastyring gennem cloud-baserede applikationer, som CAD -leverandører begynder at promovere, kan der gemme sig leverandørernes hensigt om at monopolisere projektdatastyring, hvor brugerne befinder sig i en position, hvor valget af format er mere relateret til virksomhedsinteresser end til reelle behov.

En analyse af nøglefakta [93] viser, at hovedformålet med disse ændringer snarere handler om at bevare kontrollen over økosystemer og datastrømme til fordel for leverandører, som i 40 år aldrig har været i stand til at give adgang til CAD-databaser.

Måske er tiden inde til, at virksomhederne holder op med at vente på nye koncepter fra softwareleverandører og i stedet fokuserer på selvudvikling i datacentreret retning. Efter at have frigjort sig fra problemer med dataadgang ved hjælp af reverse engineering-værktøjer vil branchen være i stand til selvstændigt at gå over til moderne, gratis og praktiske værktøjer til at arbejde med og analysere data uden at påtvinge nye koncepter.

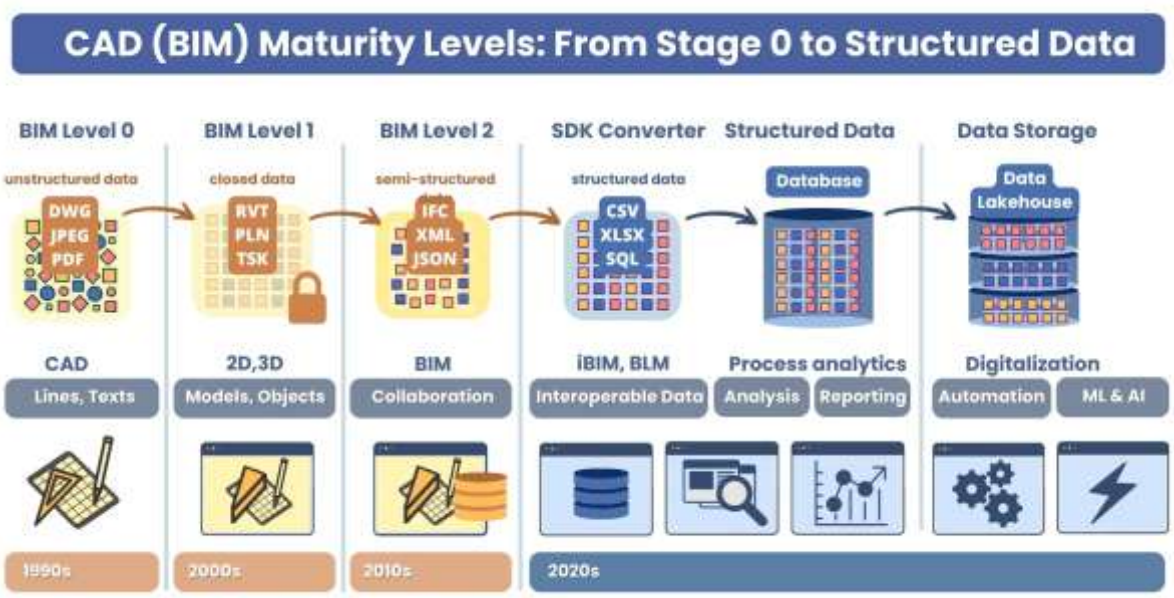


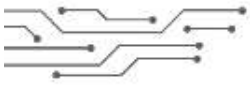
Fig. 6.2-7 CAD-modenhedsniveau (BIM): fra ustrukturerede data til strukturerede data og arkiver.

Adgang til databaser, åbne data og formater vil uundgåeligt blive standard i byggebranchen, uanset leverandørernes forsøg på at forsinke processen - det er kun et spørgsmål om tid (fig. 6.2-7). Tempoet i denne overgang kan øges betydeligt, hvis flere og flere fagfolk bliver fortrolige med åbne formater, databaseværktøjer og tilgængelige reverse engineering SDK'er, der giver direkte adgang til CAD-data - systemer [92].

> Fremtiden ligger i åbne, ensartede og analytisk tilgængelige data. For at undgå afhængighed af leverandørløsninger og for at undgå at blive taget som gidsel i lukkede økosystemer bliver bygge- og anlægsvirksomheder før eller siden nødt til at satse på åbenhed og uafhængighed og vælge formater og løsninger, der giver fuld kontrol over data.

De data, der skabes i byggebranchen i dag, vil være en vigtig ressource for forretningsbeslutninger i fremtiden. De vil fungere som det strategiske "brændstof", der driver byggevirksomhedernes udvikling og effektivitet. Byggebranchens fremtid ligger i evnen til at arbejde med data, ikke i valget af dataformater eller modeller.

For at forstå forskellen mellem de åbne formater USD, glTF, DAE, OBJ og de proprietære parametriske CAD-formater er det vigtigt at se på et af de mest komplekse og centrale dataelementer i visualisering og designberegninger - geometrien og dens genereringsprocesser. Og for at forstå, hvordan geometriske data bliver grundlaget for analyser og beregninger i byggeriet, er det nødvendigt at se nærmere på mekanismerne for generering, transformation og lagring af geometri.



KAPITEL 6.3.
GEOMETRI I BYGGERIET: FRA LINJER TIL KUBIKMETER

Når linjer bliver til penge, eller hvorfor bygherrer har brug for geometri

Geometri i byggeriet er ikke kun en visualisering, men også grundlaget for nøjagtige kvantitative beregninger. I projektmodellen supplerer geometrien listerne over elementparametre (fig. 3.1-16) med vigtige volumetriske egenskaber som længde, areal og volumen. Disse volumetriske parameterværdier beregnes automatisk ved hjælp af geometrikernerne og er udgangspunktet for estimater, tidsplaner og ressourcemodeller. Som vi allerede har diskuteret i del 5 af denne bog og i kapitlet "Costing and Estimating construction projects", er det de volumetriske parametre for grupper af objekter fra CAD -modeller, der danner grundlaget for moderne ERP, PMIS -systemer/ Geometri spiller en grundlæggende rolle ikke kun i designfasen, men også i projektimplementeringsstyring, tidsplanlægning, budgettering og drift. Ligesom nøjagtigheden af et projekt for tusinder af år siden, da man byggede egyptiske pyramider, afhang af længdemål som alen og alen, har nøjagtigheden af geometrifortolkningen i CAD -programmer i dag direkte indflydelse på resultatet fra budget og tidsfrister til valg af entreprenør og leveringslogistik.

> I et meget konkurrencepræget og budgetbegrænset miljø bliver nøjagtigheden af volumenberegninger, som er direkte afhængige af geometrien, en overlevelsesfaktor. Moderne ERP -systemer er direkte afhængige af korrekte volumetriske egenskaber fra CAD - og BIM - modeller. Derfor er en nøjagtig geometrisk beskrivelse af elementer ikke bare en visualisering, men et vigtigt værktøj til at styre omkostningerne og timingen af byggeriet.

Historisk set har geometri været det primære sprog for ingeniørkommunikation. Fra linjer på papyrus til digitale modeller har tegninger og geometriske repræsentationer fungeret som et middel til at udveksle information mellem designere, formænd og beregnere. Før computerens indtog blev beregninger udført manuelt ved hjælp af linealer og vinkelmålere. I dag er denne opgave automatiseret takket være volumetrisk modellering: De geometriske kerner i CAD-software konverterer linjer og punkter til tredimensionelle kroppe, hvorfra alle de nødvendige egenskaber automatisk udtrækkes.

Når man arbejder i CAD -programmer, oprettes geometriske elementer til beregninger via brugergrænsefladen i CAD- (BIM-) -programmer. For at omdanne punkter og linjer til volumetriske legemer bruges den geometriske kerne, som opfylder nøgleopgaven - omdannelse af geometri til volumetriske modeller, hvorfra elementets volumetriske egenskaber automatisk beregnes efter tilnærmelse.

Fra linjer til volumener: Hvordan areal og volumen bliver til data

I ingeniørpraksis beregnes volumener og arealer ud fra geometriske overflader beskrevet analytisk eller gennem parametriske modeller som NURBS (nonuniform rational B-splines) inden for rammerne af BREP (boundary element representation).

NURBS (Non-Uniform Rational B-Splines) er en matematisk måde at beskrive kurver og overflader på, mens BREP er en ramme til at beskrive den komplette tredimensionelle geometri af et objekt, inklusive dets grænser, som kan defineres ved hjælp af NURBS.

På trods af BREP's og NURBS' nøjagtighed kræver de kraftige beregningsressourcer og komplekse algoritmer. Men direkte beregninger ud fra sådanne matematisk nøjagtige beskrivelser er ofte beregningsmæssigt vanskelige, så i praksis bruges næsten altid tessellering - omdannelsen af overflader til et gitter af trekanter - hvilket forenkler de efterfølgende beregninger. Tessellation er opdelingen af en kompleks overflade i trekanter eller polygoner. I CAD /CAE miljøer bruges denne metode til visualisering, volumenberegninger, kollisionssøgning, eksport til formater som MESH og kollisionsanalyse. Et eksempel fra naturen er en bikage, hvor en kompleks form opdeles i et regelmæssigt gitter (fig. 6.3-1).

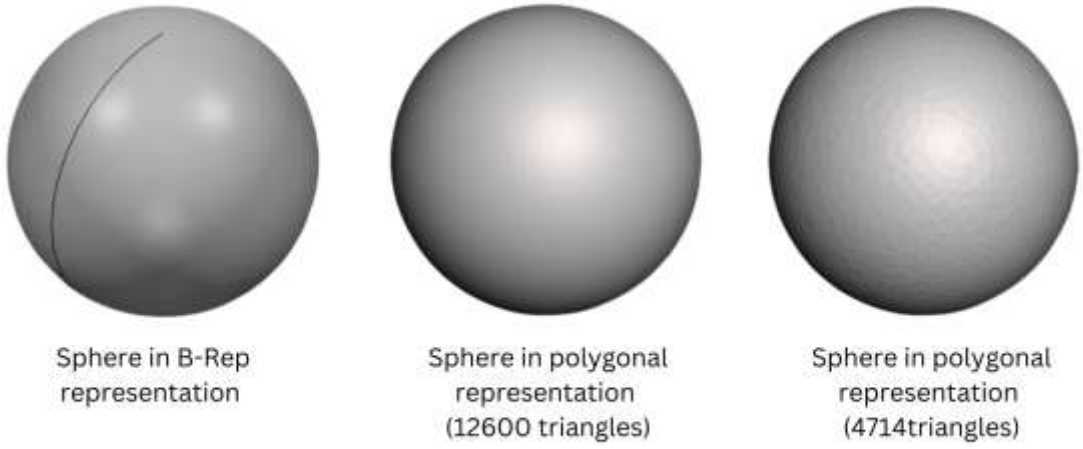


Fig. 6.3-1 Den samme kugle i parametrisk beskrivelse BREP og polygonal repræsentation med forskelligt antal trekanter.

BREP (NURBS), som bruges i CAD, er ikke en grundlæggende model for geometri. Den blev skabt som et praktisk værktøj til at repræsentere cirkler og rationelle splines og til at minimere lagringen af geometridata. Den har dog begrænsninger - for eksempel den manglende evne til nøjagtigt at beskrive den sinusform, der ligger til grund for spiralformede linjer og overflader, og behovet for at bruge komplekse geometriske kerner.

I modsætning hertil er trekantede net og tesselering af parametriske former kendetegnet ved enkelhed, effektiv brug af hukommelse og evnen til at behandle store mængder data (Fig. 6.3-2). Disse fordele gør det muligt at undvære komplekse og dyre geometriske kerner, og de titusindvis af kodelinjer, der er indlejret i dem, når man beregner geometriske former.

I de fleste bygningssager er det ligegyldigt, hvordan de volumetriske egenskaber defineres - gennem parametriske modeller (BREP, IFC) eller gennem polygoner (USD, glTF, DAE, OBJ). Geometrien forbliver en form for tilnærmelse: uanset om det er gennem NURBS eller MESH, er det altid en omtrentlig beskrivelse af formen.

Geometri defineret som polygoner eller BREP (NURBS) er til en vis grad kun en måde at tilnærme sig med

en omtrentlig beskrivelse af en kontinuerlig form. Ligesom Fresnel-integraler ikke har noget eksakt analytisk udtryk, er diskretisering af geometri gennem polygoner eller NURBS altid en tilnærmelse, ligesom trekantede MESH.

Parametrisk geometri i BREP -format er hovedsageligt nødvendig, hvor minimal datastørrelse er vigtig, og det er muligt at bruge ressourcekrævende og dyre geometrikerner til behandling og visning. Oftest er det karakteristisk for udviklere af CAD -programmer, som til dette formål anvender geometriske kerner fra MCAD -leverandører i deres produkter. I dette tilfælde, selv i disse programmer, konverteres BREP-modeller i tessellationsprocessen til visualisering og beregninger ofte til trekanter (på samme måde som PSD-filer forenkles til JPEG).

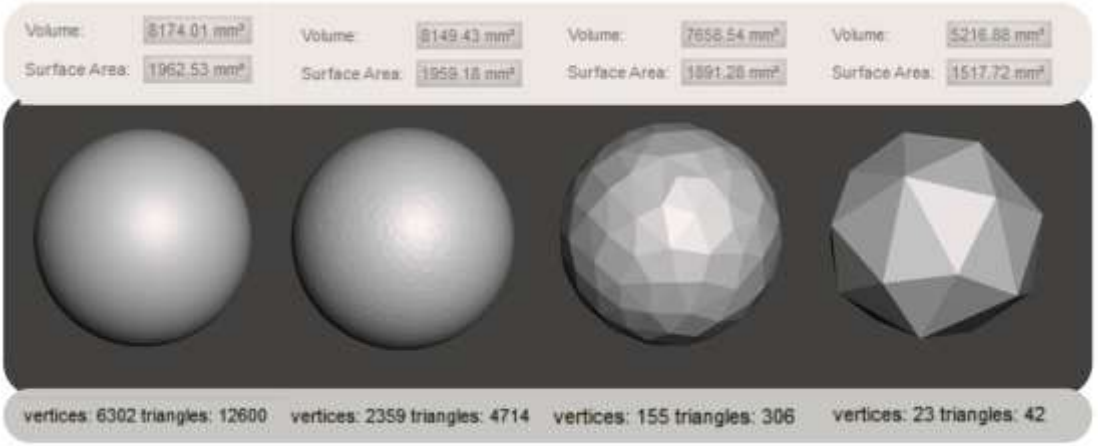


Fig. 6.3-2 Forskel på volumetriske egenskaber i figurer med forskelligt antal polygoner.

Polygonal MESH, såvel som parametrisk BREP, har deres egne fordele og begrænsninger, men målet er det samme - at beskrive geometrien under hensyntagen til brugerens opgaver. I sidste ende afhænger nøjagtigheden af en geometrisk model ikke kun af metoden til dens repræsentation, men også af kravene til en bestemt opgave.

> I de fleste konstruktionsproblemer kan behovet for parametrisk geometri og komplekse geometriske kerner være overflødigt.

I hver enkelt automatiseret beregningsopgave er det værd at overveje, om betydningen af parametrisk geometri overdrives af CAD-udviklere, der er interesserede i at promovere og sælge deres egne softwareprodukter.

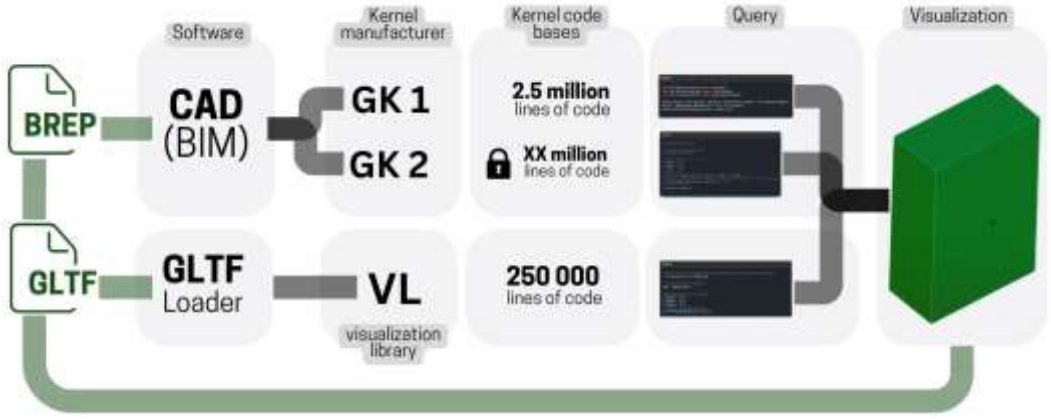
Flytning til MESH, USD og polygoner: brug af tessellation til geometri

Når man i byggebranchen streamer, udvikler systemer, databaser eller automatiserer processer, der arbejder med designinformation og featuregeometri, er det vigtigt at stræbe efter uafhængighed af specifikke CAD-editorer og geometrikerner.

> Det udvekslingsformat, der skal bruges både i beregningsafdelingerne og på byggepladsen, bør ikke være baseret på et specifikt CAD- (BIM-) program. Geometrisk information skal repræsenteres i formatet direkte gennem tessellering uden henvisning til den geometriske kerne eller CAD-arkitektur.

Parametrisk geometri fra CAD kan betragtes som en mellemliggende kilde, men ikke som grundlaget for et universelt format. De fleste parametriske beskrivelser (herunder BREP og NURBS) konverteres under alle omstændigheder til polygonal MESH til videre behandling. Hvis resultatet er det samme (tesselering og polygoner), og processen er enklere, er valget indlysende. Dette svarer til valget mellem grafontologier og strukturerede tabeller (som vi diskuterede i del fire): overdreven kompleksitet er sjældent berettiget (fig. 3.2-10, fig. 6.1-8).

Åbne formater som: OBJ, STL, glTF, SVF, CPIXML, USD og DAE, bruger en universel trekantnetstruktur, hvilket giver dem betydelige fordele. Disse formater har fremragende interoperabilitet - de er nemme at læse og visualisere ved hjælp af tilgængelige open source-biblioteker uden behov for komplekse specialiserede geometrikerner, der indeholder millioner af kodelinjer (figur 6.3-3). Disse alsidige geometriformater bruges i applikationer, der spænder fra relativt enkle køkkendesignværktøjer i IKEA™ til komplekse objektvisualiseringssystemer i biografer og VR -applikationer. En vigtig fordel er, at der findes et stort antal gratis og open source-biblioteker til at arbejde med disse formater, som er tilgængelige for de fleste platforme og programmeringssprog.



Figur 6.3-3 Den samme geometriske repræsentation opnås ved brug af parametriske formater og geometriske kerner, eller ved brug af trianguleredе formater og open source-visualiseringsbiblioteker.

Ud over brugerne selv har CAD -leverandørerne problemer med at fortolke udenlandske parametriske CAD-formater eller åbne IFC på grund af forskellige geometriske kerner. I praksis bruger alle CAD -leverandører uden undtagelse reverse engineering SDK til at overføre data mellem systemer, og ingen af dem er afhængige formater som IFC eller USD [93] til interoperabilitetsformål.

> I stedet for at bruge koncepter, der fremmes af alliancer af CAD- leverandører, som de ikke selv bruger, er det mere produktivt for udviklere og brugere af CAD-løsninger at fokusere på at forstå fordelene ved hver tilgang i en specifik sammenhæng og at vælge den ene eller den anden type geometri afhængigt af brugssituationen. At vælge mellem forskellige geometriske repræsentationer er en afvejning mellem nøjagtighed, beregningseffektivitet og de praktiske behov i forbindelse med en bestemt opgave.

Den kompleksitet, der er forbundet med brugen af geometriske kerner, og som traditionelt pålægges byggebranchen af store leverandører ved behandling af designdata, viser sig ofte at være overflødig. USD-formatet baseret på MESH-geometri kan blive en slags "Pandoras æske" for branchen og åbne nye muligheder for udviklere til at organisere dataudveksling - uden for rammerne af IFC og parametriske BREP-strukturer, der er typiske for CAD-leverandører.

Når man ser nærmere på strukturen i USD, DAE, gLTF, OBJ osv., bliver det tydeligt, at der findes enklere, åbne formater, som gør det muligt effektivt at organisere overførslen og brugen af geometrisk information uden at være afhængig af kompleks parametrik og lukkede geometriske kerner. Denne tilgang sænker ikke kun den tekniske indgangstærskel for udviklere, men fremmer også udviklingen af fleksible, skalerbare og virkelig åbne løsninger til digitalt byggeri.

LOD, LOI, LOMD - unik klassifikation af detaljer i CAD (BIM)

Ud over de geometriske repræsentationsformater tilbyder CAD - (BIM-) metoder deres egne unikke klassifikationssystemer, som strukturerer tilgangen til at informere bygningsmodeller i en verden, hvor forskellige industrier bruger forskellige niveauer af detaljer og datadybde.

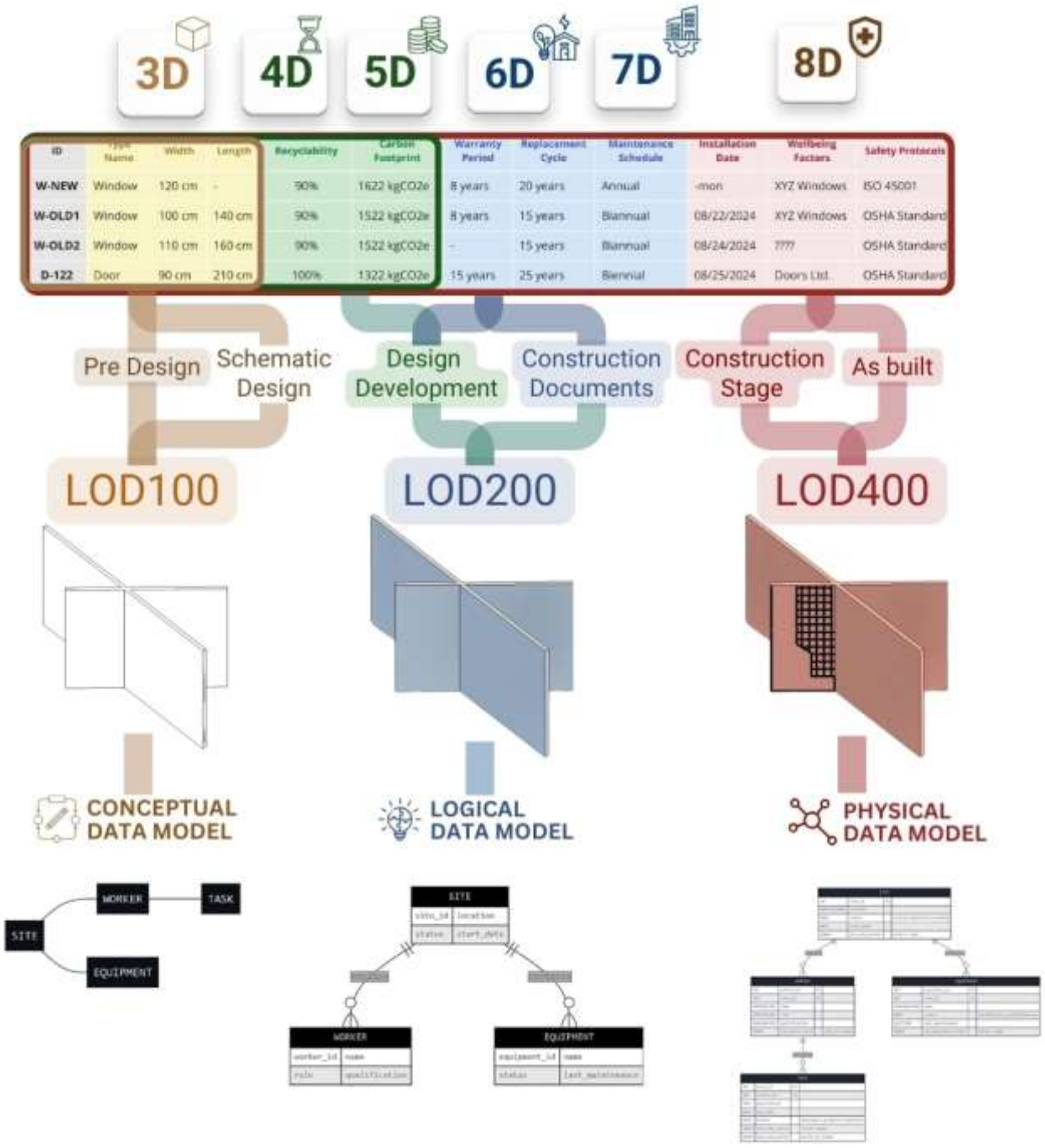
Et af eksemplerne på nye tilgange til standardisering er indførelsen af niveauer for modeludvikling, der afspejler graden af parathed og pålidelighed af både grafiske og informationsmæssige komponenter. Til differentiering af informationsindholdet i arbejdet med CAD - (BIM-) data opstod LOD (Level Of Detail) - detaljeringsniveauet for den grafiske del af modellen, og LOI (Level Of Information) - niveauet for dataudarbejdelse. Derudover blev begrebet LOA (Level of Accuracy) introduceret for den integrerede tilgang - nøjagtigheden af de præsenterede elementer og LOG (Level of Geometry) for at bestemme nøjagtigheden af den grafiske repræsentation.

Detaljeringsniveauer (LOD) er angivet med tal fra 100 til 500, hvilket afspejler graden af modeludvikling. LOD 100 er en konceptuel model med generelle former og dimensioner. LOD 200 omfatter mere præcise dimensioner og former, men med betingede detaljer. LOD 300 er en detaljeret model med præcise dimensioner, former og elementplaceringer. LOD 400 indeholder detaljerede oplysninger, der er nødvendige for fremstilling og installation af elementer. LOD 500 afspejler anlæggets faktiske tilstand efter opførelsen og bruges til drift og vedligeholdelse. Disse niveauer beskriver strukturen i CAD (BIM)-modellens informationsmætning i forskellige faser af livscyklussen, herunder 3D, 4D, 5D og videre.

I virkelige projekter er det høje detaljeringsniveau (LOD400) ofte for højt, og det er tilstrækkeligt at bruge LOD100-geometri eller endda flade tegninger, mens resten af dataene kan indhentes enten beregningsmæssigt eller fra relaterede elementer, der måske ikke har en tydelig geometri. For eksempel har rum og rumelementer (rumelementkategorier) måske ingen visuel geometri, men indeholder betydelige mængder information og databaser, som mange forretningsprocesser er bygget op omkring.

> Det er derfor vigtigt klart at definere den nødvendige detaljeringsgrad, før man begynder at designe. For 4D -7D use cases er selv DWG-tegninger og minimal LOD100-geometri ofte tilstrækkeligt. Den vigtigste opgave i kravprocessen er at finde en balance mellem modellens rigdom og praktiske anvendelighed.

Hvis vi betragter CAD (BIM)-data som en database (hvilket de er), er beskrivelsen af modelmætning gennem nye akronymer i bund og grund ikke andet end en trinvis datamodellering for informationssystemer, der starter med det konceptuelle niveau og slutter med det fysiske (fig. 6.3-4), som blev diskuteret i detaljer i bogens tredje og fjerde del. Hver forøgelse af LOD og LOI betyder tilføjelse af information, der er nødvendig for nye opgaver: beregninger, byggestyring, drift og er kendetegnet ved successiv berigelse af modellen med yderligere informationslag (3D -8D) i form af forskellige parametre, som vi diskuterede i bogens femte del.



Figur 6.3-4 Processen med at konkretisere detaljerne i et projekt er identisk med datamodellering fra en konceptuel til en fysisk datamodel.

Geometri er kun en del af designdataene, som der ikke altid er behov for i byggeprojekter, og det centrale spørgsmål i arbejdet med CAD -data er ikke så meget, hvordan modellerne visualiseres, men mere hvordan dataene fra disse modeller kan bruges uden for CAD- (BIM -) programmer.

I midten af 2000'erne stod byggebranchen over for en hidtil uset udfordring med den hurtige stigning i mængden af data i ledelses- og databehandlingssystemer, især dem, der kom fra CAD-afdelinger (BIM).

Denne dramatiske stigning i datamængden kom bag på virksomhedsledere, og de var uforberedte på de voksende krav til datakvalitet og -styring.

Nye CAD-standarder (BIM) - AIA, BEP, IDS, LOD, COBie

Ved at udnytte den manglende åbne adgang til CAD-databaser og den begrænsede konkurrence på markedet for databehandling og ved at bruge markedsføringskampagnerne i forbindelse med det nye akronym BIM er organisationer, der er involveret i at udvikle tilgange til håndtering af CAD-data, begyndt at skabe nye standarder og koncepter, der de jure skulle have til formål at forbedre dataforvaltningspraksis.

> Selvom næsten alle initiativer, der direkte eller indirekte støttes af CAD-leverandører og udviklere (BIM), har haft til formål at optimere arbejdsgange, har de resulteret i et væld af standarder, som forskellige interessenter har presset på for at få indført, hvilket har ført til en vis tvetydighed og forvirring i byggebranchen omkring dataprocesser.

Lad os liste nogle af de nye datastandarder, ud over LOD, LOI, LOA, LOG, som er dukket op i de senere år i byggebranchen:

- **BEP** (BIM Execution Plan) - beskriver, hvordan man integrerer og bruger CAD (BIM) i et projekt, og definerer datahåndteringsmetoder og -processer.
- **EIR-dokument /AIA** (bygherrens informationskrav) - udarbejdes af bygherren forud for udbuddet og indeholder kravene til entreprenørens udarbejdelse og tilvejebringelse af information. Det fungerer som grundlag for BEP i det pågældende projekt.
- **AIM** (Asset Information Model) er en del af BIM-processen. Når projektet er leveret og afsluttet, kaldes datamodellen for Asset Information Model eller AIM. Formålet med AIM er at styre, vedligeholde og drive det realiserede aktiv.
- **IDS** (Information Delivery Specification) - definerer kravene til og hvilke data og i hvilket format, der kræves i forskellige faser af et byggeprojekt.
- **iLOD** er detaljeringsniveauet LOD, som informationerne er repræsenteret med **i** BIM -modellen. Den definerer, hvor detaljerede og komplette oplysningerne i modellen er, fra grundlæggende geometriske repræsentationer til detaljerede specifikationer og data.
- **eLOD** - LOD detaljeringsgraden af de enkelte elementer i en CAD-model (BIM). Det definerer, i hvilket omfang hvert element er modelleret, og de tilknyttede oplysninger såsom dimensioner, materialer, ydeevneegenskaber og andre relevante attributter.
- **APS** (Platform Services) og andre produkter fra store CAD-leverandører (BIM) - beskriver de værktøjer og den infrastruktur, der er nødvendig for at skabe forbundne og åbne datamodeller.

Selv om det erklærede formål med at implementere CAD (BIM)-standarder - såsom LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - er at forbedre kvaliteten af datastyringen og udvide automatiseringsmulighederne, fører brugen af dem i praksis ofte til overdreven kompleksitet og fragmentering af processerne. Hvis vi betragter CAD (BIM)-modellen som en slags database, bliver det tydeligt, at mange af disse standarder duplikerer veletablerede og effektive tilgange, der bruges i andre brancher til at arbejde med informationssystemer. I stedet for at forenkle og ensrette skaber sådanne

initiativer ofte en yderligere terminologisk byrde og hindrer implementeringen af virkelig åbne og fleksible løsninger.

Det er bemærkelsesværdigt, at mange af disse nye koncepter faktisk erstatter de modellerings- og datavalideringsprocesser, der blev diskuteret i detaljer i de første dele af bogen, og som længe har været brugt i andre økonomiske sektorer. I byggeriet går standardiseringsprocessen derimod ofte i den modsatte retning - der skabes nye databeskrivelsesformater, nye standarder og nye koncepter for datavalidering, som ikke altid fører til reel ensartethed og praktisk anvendelighed. I stedet for at forenkle og automatisere behandlingen står branchen derfor over for flere niveauer af regulering og bureaukrati (fig. 6.3-1), hvilket ikke altid er befordrende for øget effektivitet.

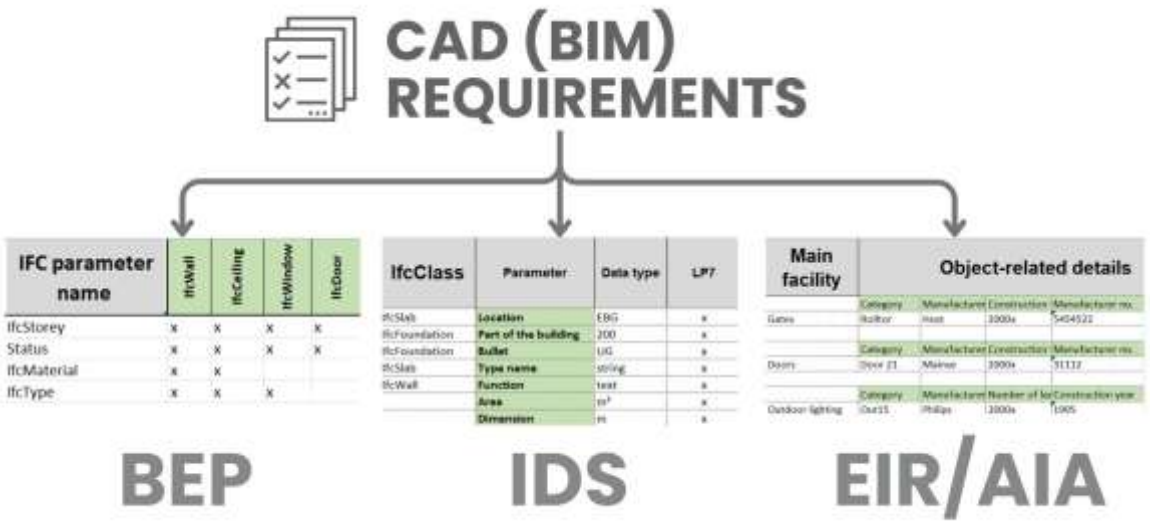


Fig. 6.3-1 Krav til data- og informationsindhold reduceres til beskrivelsen af attributter og deres grænseværdier, der beskrives ved hjælp af tabeller.

I stedet for at forenkle databehandlingen skaber nye koncepter i forbindelse med CAD (BIM)-data ofte yderligere kompleksitet og tvister allerede på fortolkningsstadiet og i forbindelse med grundlæggende definitioner.

Et af de seneste eksempler på nye koncepter er IDS-formatet (indført i 2020), som gør det muligt at beskrive kravene til attributsammensætningen af en informationsmodel i det åbne BIM-koncept. IDS-krav beskriver oplysninger om attributter og deres grænseværdier i form af en struktureret tabel (Excel eller MySQL), som derefter oversættes til markeringen af et semi-struktureret XML-format, omdøbt fra XML til den særlige forkortelse IDS.

I modsætning til det synspunkt, der fremmes af leverandører og støttes af BIM og åben BIM, at datahåndtering i byggeriet er unik på grund af brugen af specialiserede værktøjer som CAD og BIM, er dataformaterne og datastyringspraksisserne i denne branche ikke forskellige fra dem andre brancher.

Antallet af krav til projekter og CAD (BIM) formater kan forenkles ved at bruge en enkelt kravtabel med attribut-kolonner, som beskrevet i kapitlet "Oversættelse af krav til struktureret form", uden at skulle oversætte de oprindeligt strukturerede krav til ikke-tabelformater (IDS er oprindeligt beskrevet via en tabel).

Den forenklede tilgang (figur 6.3-2), som omfatter kolonner til entitetsidentifikatorer, egenskaber og grænseværdier, der blev diskuteret i detaljer i tidligere kapitler (figur 4.4-9, figur 4.4-16, figur 7.3-10), eliminerer behovet for at konvertere krav til IDS-XML-format. Denne metode giver en direkte, mindre besværlig og mere gennemsigtig mekanisme til datakvalitetskontrol. Den bygger på udbredte værktøjer, fra regulære udtryk (RegEx) til dataframes, Pandas og standard ETL -payplanes - præcis som dem, der bruges af fagfolk i andre økonomiske sektorer til at arbejde med data.

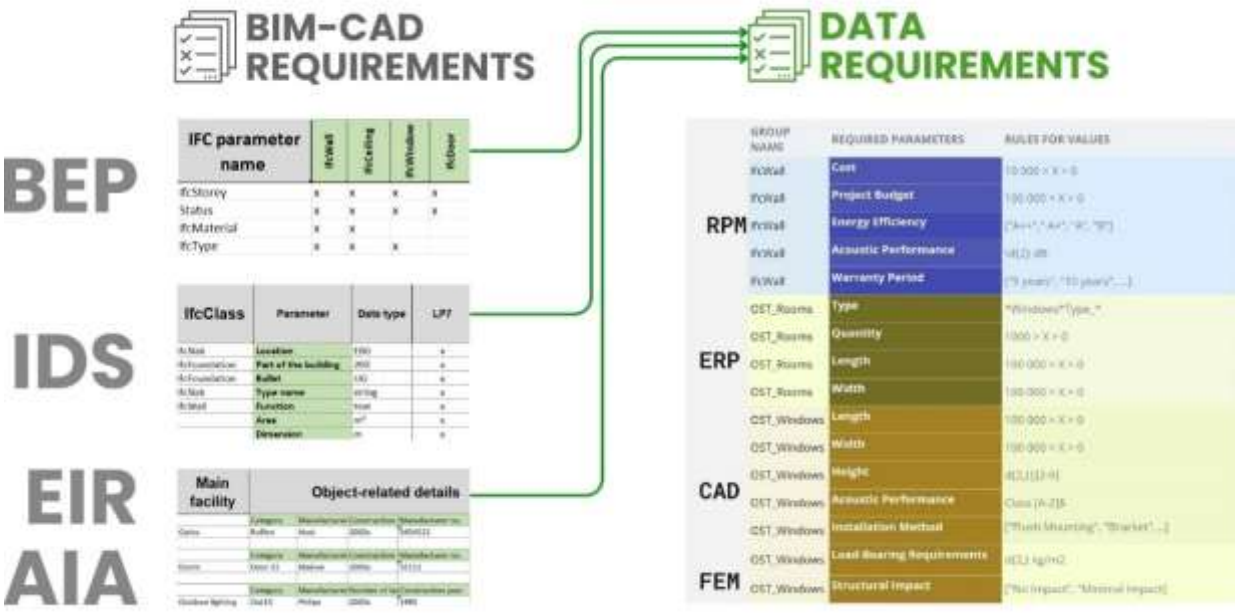


Fig. 6.3-2 Datakrav i andre brancher er forenklet til en struktureret beskrivelse af attributter og deres grænseværdier.

På grund af dataenes lukkede karakter er der med tiden dukket flere og flere nye tilgange og teknikker op i byggebranchen til at kontrollere og styre disse forskelligartede data, selv om dataene i byggeprojekter stort set er de samme som på andre områder. Mens andre brancher med succes har klaret sig med standardiserede tilgange til databehandling, fortsætter byggeriet med at udvikle nye og unikke dataformater, krav og valideringskoncepter.

> De metoder og værktøjer, der bruges til at indsamle, forberede og analysere data i byggeriet, bør ikke være fundamentalt forskellige fra dem, der bruges af specialister i andre økonomiske sektorer.

Branchen har udviklet et særskilt terminologisk økosystem, som kræver kritisk refleksion og revurdering:

- STEP-formatet er placeret under det nye navn IFC, suppleret med konstruktionskategorisering, uden at tage hensyn til begrænsningerne i selve STEP-formatet.
- Det parametriske format IFC bruges i datakommunikationsprocesser på trods af manglen på en

samlet geometrisk kerne, der er nødvendig for visualisering og beregning.
- Adgang til databaser CAD -systemer promoveres under betegnelsen "BIM ", uden at man diskuterer funktionerne i disse databaser og adgangen til dem.
- Leverandører fremmer interoperabilitet gennem formaterne IFC og USD, ofte uden at omsætte dem til praksis, ved hjælp af kostbar reverse engineering, som de selv har kæmpet med.
- Udtrykkene LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD bruges universelt til at beskrive de samme enhedsparametre uden henvisning til modellerings- og verifikationsværktøjer, der længe har været brugt i andre brancher.

Byggebranchen viser, at alt det ovenstående, selv om det lyder mærkeligt, er muligt i byggebranchen - især hvis hovedformålet er at tjene penge på hvert trin i databehandlingen gennem salg af specialiserede tjenester og software. Fra et forretningsmæssigt synspunkt er der ikke noget galt med dette. Men det er stadig et åbent spørgsmål, om sådanne akronymer og tilgange i forbindelse med CAD (BIM) virkelig tilfører værdi og forenkler de professionelle processer.

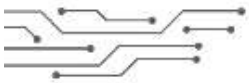
> I byggebranchen fungerer sådan et system, fordi branchen selv tjener det meste af sin spekulationsgevinst i denne labyrint af systemer og akronymer. Virksomheder, der er interesserede i gennemsigtige processer og åbne data, er sjældne. Denne komplekse situation vil sandsynligvis fortsætte på ubestemt tid - indtil kunder, klienter, investorer, banker og private equity begynder at kræve klarere og mere informerede tilgange til informationsstyring.

Branchen har samlet et stort antal akronymer, men de beskriver alle de samme processer og datakrav i varierende grad. Deres reelle anvendelighed til at forenkle arbejdsgange er stadig tvivlsom.

Mens koncepter og markedsføringsakronymer kommer og går, vil selve valideringsprocesserne for datakrav altid forblive en integreret del af forretningsprocesserne. I stedet for at skaber flere og flere specialiserede formater og regler, bør byggebranchen se på værktøjer, der allerede har vist sig at være effektive på andre områder som finans, industri og IT.

Overfloden af termer, akronymer og formater skaber en illusion om dybt gennemarbejdede digitale byggeprocesser. Men markedsføringskoncepter og kompleks terminologi skjuler ofte en simpel, men ubekvem sandhed: Data er stadig svært tilgængelige, dårligt dokumenterede og stift bundet til specifikke softwareløsninger.

For at komme ud af denne onde cirkel af akronymer og formater for formaternes skyld er det nødvendigt at se på CAD (BIM) -systemer ikke som magiske informationsstyringsværktøjer, men som det, de virkelig er - specialiserede databaser. Og det er gennem dette prisme, at man kan forstå, hvor marketing slutter, og det virkelige arbejde med information begynder.



KAPITEL 6.4.
DESIGNPARAMETERISERING OG BRUG AF LLM TIL CAD-DRIFT

Illusionen om unikke CAD-data (BIM): vejen til analyse og åbne formater

Moderne CAD (BIM) -platforme har i høj grad ændret tilgangen til design- og byggeinformationsstyring. Hvor disse værktøjer tidligere primært blev brugt til at skabe tegninger og 3D-modeller, fungerer de i dag som fuldgyldige projektdatalagre. I henhold til Single Source of Truth-konceptet bliver den parametriske model i stigende grad den vigtigste og ofte den eneste kilde til projektinformation, hvilket sikrer dens integritet og relevans gennem hele projektets livscyklus.

> Den vigtigste forskel mellem CAD - (BIM -) platforme og andre systemer til styring af byggedata er behovet for at bruge specialiserede værktøjer og API'er for at få adgang til oplysningerne (den eneste kilde til sandhed). Disse databaser er ikke universelle i traditionel forstand: I stedet for en åben struktur og fleksibel integration er de et lukket miljø, der er bundet til en bestemt platform og et bestemt format.

På trods af kompleksiteten i arbejdet med CAD -data er der et vigtigere spørgsmål, som går ud over den tekniske erkendelse: Hvad er CAD-databaser (BIM) egentlig? For at besvare dette spørgsmål er det nødvendigt at gå ud over de sædvanlige akronymer og begreber, som softwareudviklere har indført. I stedet er det værd at fokusere på essensen af at arbejde med projektinformation: data og deres behandling.

> Forretningsprocessen i byggeriet starter ikke med arbejde i CAD - eller BIM - værktøjer, men med udformning af projektkrav og datamodellering. Først defineres opgaveparametrene: listen over enheder, deres oprindelige egenskaber og grænseværdier, der skal tages i betragtning, når en bestemt opgave skal løses. Først derefter oprettes modeller og elementer i CAD (BIM)-systemer på grundlag af de angivne parametre

Den proces, der går forud for oprettelsen af oplysninger i CAD - (BIM-) databaser, er fuldstændig den samme som datamodelleringsprocessen, der blev diskuteret i detaljer i bogens fjerde del og kapitlet "Datamodellering: konceptuel, logisk og fysisk model" (Fig. 4.3-1).

Ligesom vi i datamodellering opstiller krav til de data, vi senere vil behandle i databasen, opstiller ledere af CAD-databaser designkrav i form af flere tabelkolonner eller lister med nøgleværdipar (fig. 6.4-1, trin 1-2). Og kun på grundlag af disse indledende parametre ved hjælp af API'en automatisk eller manuelt opretter (eller rettere forfiner) designeren objekter i CAD- (BIM) databaser (trin 3-4), hvorefter de igen kontrolleres for overholdelse af de oprindelige krav (trin 5-6). Denne proces - definition→ oprettelse→ validering→ justering (trin 2-6) - gentages iterativt, indtil datakvaliteten, i datamodellering, når det ønskede niveau for målsystemet - dokumenter, tabeller eller dashboards (trin 7).

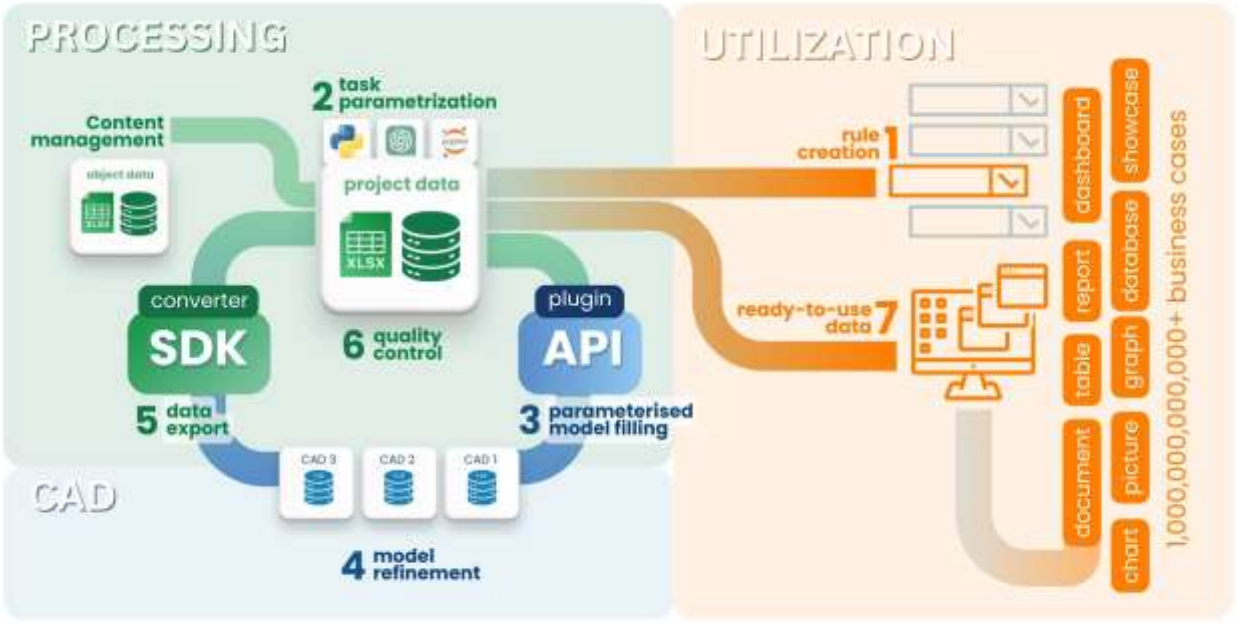


Fig. 6.4-1 Cyklussen for informationsmætning af databaser til forretningsprocesser i gennemførelsen af byggeprojekter.

Hvis vi betragter CAD (BIM) som en mekanisme til parameteroverførsel i form af et sæt nøgleværdipar, der genereres på grundlag af krav, der er defineret uden for designmiljøet (fig. 6.4-1, trin 1-2), skifter diskussionens fokus fra specifikke softwareløsninger og deres begrænsninger til mere grundlæggende aspekter - datastruktur, datamodeller og datakrav. I bund og grund taler vi om parametermætning af databasen og den klassiske datamodelleringsproces (trin 2-3 og 5-6). Den eneste forskel er, at på grund af CAD-databasernes lukkede karakter og de anvendte formaters særegenheder ledsages denne proces af brugen af specialiserede BIM-værktøjer. Spørgsmålet er: Hvad er det unikke ved BIM, hvis der ikke findes lignende tilgange i andre brancher?

I de sidste 20 år er BIM blevet positioneret som mere end blot en enkelt datakilde. CAD -BIM-pakken markedsføres ofte som et parametrisk værktøj med en iboende integreret database [64], der er i stand til at automatisere processerne for design, modellering og livscyklusstyring af byggeprojekter. Men i virkeligheden er BIM mere blevet et værktøj til at holde brugerne på leverandørernes platform end en praktisk metode til data- og processtyring.

Som følge heraf er CAD- (BIM-) data isoleret på deres platforme og gemmer projektinformation bag proprietære API'er og geometrikerner. Det har frataget brugerne muligheden for at få uafhængig adgang til databaser og udtrække, analysere, automatisere og overføre data til andre systemer uden om leverandørernes økosystemer.

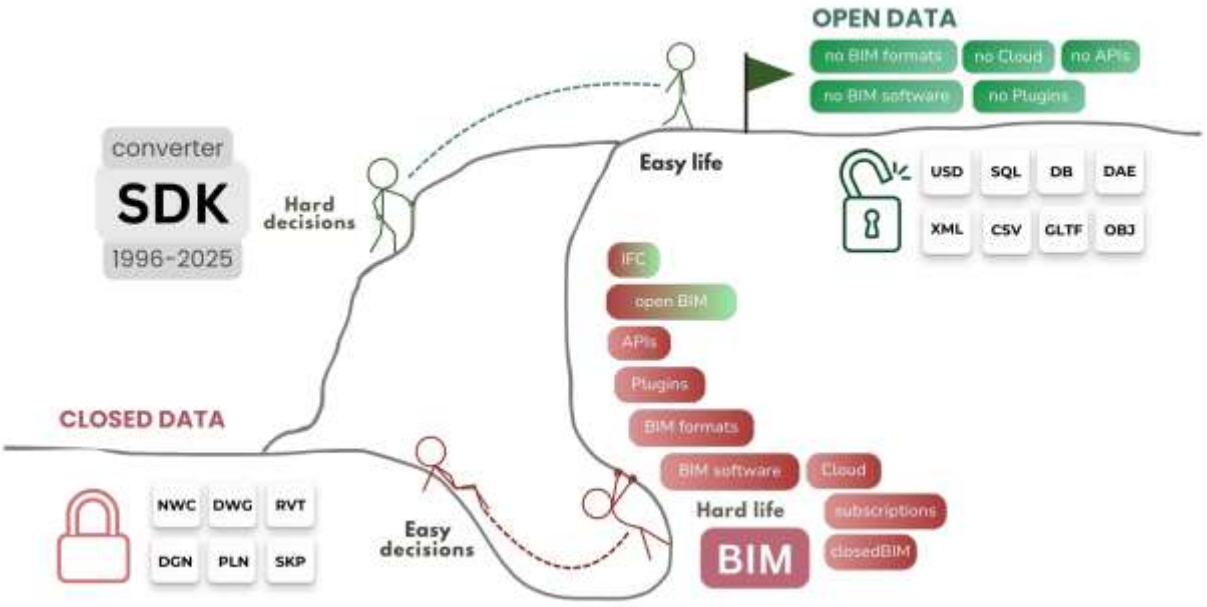


Fig. 6.4-2 I byggeriet kræver moderne formater sofistikerede geometrikerner, en årligt opdateret API og særlige licenser til CAD -(BIM -) software.

Virksomheder, der arbejder med moderne CAD-værktøjer, bør bruge den samme tilgang til at arbejde med data, som alle CAD-leverandører selv uden undtagelse bruger i praksis: datatransformation ved hjælp af SDK - reverse engineering-værktøjer, som CAD-leverandører har kæmpet imod siden 1995 [75]. Med fuld adgang til CAD-databasen og ved hjælp af reverse engineering-værktøjer kan vi få [127] et fladt sæt enheder med attributter og eksportere dem til ethvert praktisk åbent format (fig. 6.4-2), herunder både geometri og parametre for designelementer. Denne tilgang ændrer fundamentalt paradigmet for arbejdet med information - fra filorienteret til datacentreret arkitektur:

- Dataformater som RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET og andre indeholder identiske oplysninger om elementer i det samme projekt. Det betyder, at kendskab til et bestemt format og dets skema ikke bør være en barriere for at arbejde med selve dataene.
- Data fra alle formater kan kombineres til en enkelt åben struktureret og granulær struktur (figur 9.1-10), der indeholder MESH's trekantede geometri og egenskaberne for alle objektenheder, uden begrænsninger fra geometriske kerner.
- Dataanalyse stræber efter universalitet: Ved hjælp af åbne data kan du arbejde med projektdata uanset det anvendte format.
- Minimering og afhængighed af API'er og leverandør-plugins: Arbejdet med data er ikke længere afhængigt af API-færdigheder.

Når og CAD -datakrav omdannes til strukturerede repræsentationsformater, er nemme at analysere - er udviklere ikke længere afhængige af specifikke dataskemaer og lukkede økosystemer.

Design gennem parametre: fremtiden for CAD og BIM

Intet byggeprojekt i verden er nogensinde startet i et CAD-program. Før en tegning eller model tager form i CAD, gennemgår den konceptualiseringsfasen (fig. 6.4-1, trin 1-2), hvor der fokuseres på de parametre, der definerer den grundlæggende idé og logik i det fremtidige objekt. Denne fase svarer til det konceptuelle niveau i datamodellering (fig. 4.3-6). Parametrene kan eksistere udelukkende i designerens hoved, men ideelt set er de arrangeret i form af strukturerede lister, tabeller eller lagret i databaser (fig. 6.4-3), hvilket giver mulighed for gennemsigtighed, reproducerbarhed og yderligere automatisering af designprocessen.

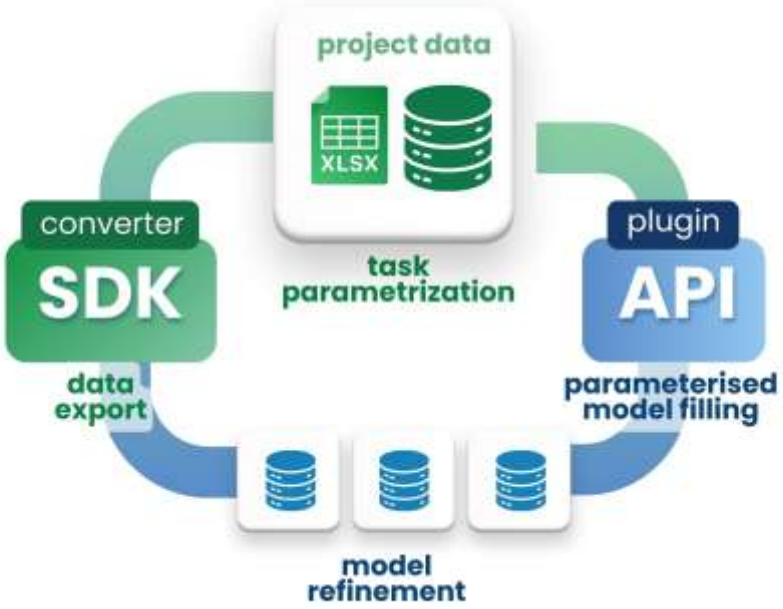


Fig. 6.4-3 Designprocessen er en iterativ proces, hvor CAD-databasen fyldes med information udefra ved hjælp af krav i værdikæden.

Før selve CAD-modelleringen påbegyndes (datamodelleringens logiske og fysiske fase (fig. 4.3-7)), er det vigtigt at definere de grænseparametre, der ligger til grund for projektet. Disse attributter indsamles, som med andre krav, helt fra slutningen af dataanvendelseskæden (f.eks. systemer), og gennem dem er begrænsninger, mål og nøgleegenskaber for fremtidige objekter i projektet allerede defineret.

Selve modelleringen kan automatiseres med 60-100% ved hjælp af parametriske modelleringsværktøjer (fig. 6.4-3), hvis kravene er veldefinerede. Så snart projektet er beskrevet i form af parametre, bliver dets udformning teknisk mulig, f.eks. ved hjælp af visuelle programmeringssprog som Grasshopper Dynamo, der er indlejret i moderne CAD -miljøer eller gratis løsninger i Blender, UE, Omniverse.

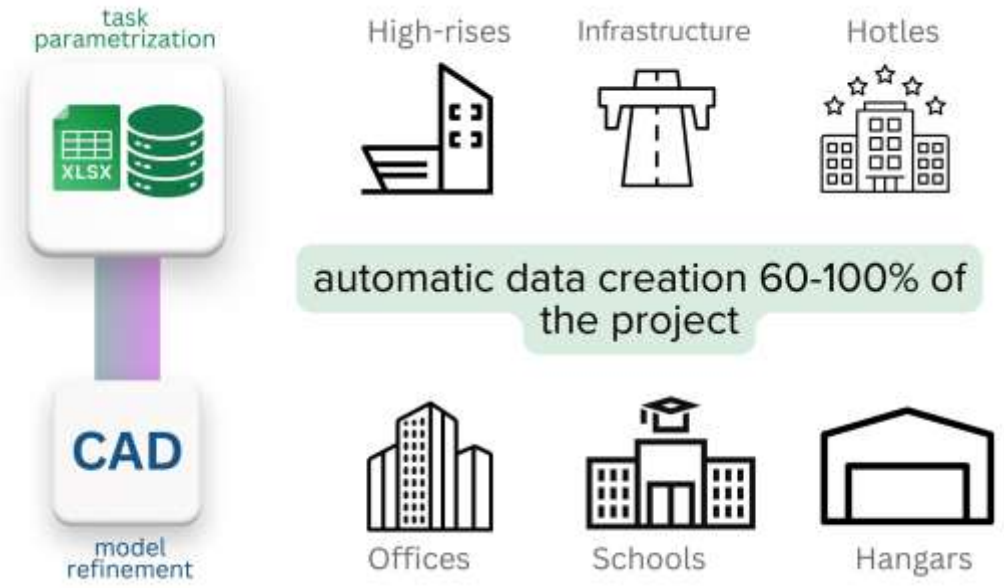


Fig. 6.4-4 Størstedelen af de typificerede projekter skabes allerede i dag helt automatisk takket være parametriske programmeringsværktøjer.

Allerede i dag skabes store industrielle og typificerede projekter ikke af designafdelingens hænder, men ved hjælp af parametriske værktøjer og visuel programmering. Det gør det muligt at bygge en model baseret på data i stedet for på en bestemt designers eller leders subjektive beslutninger.

> *Indhold går forud for design. Design uden indhold er ikke design, men dekoration [128].*
>
> - Jeffrey Zeldman, webdesigner og iværksætter

Processen starter ikke med tegning eller 3D -modellering, men med opstilling af krav. Det er kravene, der bestemmer, hvilke elementer der skal bruges i projektet, og hvilke data der skal overføres til andre afdelinger og systemer. Kun eksistensen af strukturerede krav gør det muligt automatisk at kontrollere modeller regelmæssigt (f.eks. hvert 10. minut uden at distrahere designeren fra hans arbejde).

> Måske vil CAD- (BIM-) systemet i fremtiden blot blive en grænseflade til at udfylde databasen, og det vil være ligegyldigt, i hvilket CAD-værktøj modelleringen udført (fysisk niveau).

På samme måde bruges 3D-modellering ofte inden for maskinteknik, men det er ikke et nødvendigt eller obligatorisk element i projektet. I de fleste tilfælde er den klassiske 2D-dokumentation tilstrækkelig, og den nødvendige informationsmodel oprettes på grundlag af den. Denne model er sammensat af komponenter, der er struktureret i henhold til industristandarder, og indeholder alle de oplysninger, der er nødvendige for at forstå design- og produktionsorganisationen. Fabriksinformationsmodellen bruges derefter til at skabe en fabriksinformationsmodel, hvortil der tilføjes specifikke produkter og flowdiagrammer, som allerede er orienteret mod teknologernes behov. Hele processen kan organiseres uden unødig kompleksitet og uden

at overbelaste systemet med 3D-grafik, hvor det ikke giver reelle fordele.

> Det er vigtigt at forstå, at selve 3D-modellen og CAD -systemet ikke bør spille hovedrollen - det er kun et værktøj til kvantitativ og geometrisk analyse. Alle andre parametre, undtagen geometri, som beskriver enheden, bør lagres og behandles uden for CAD-miljøet, hvis det er muligt (BIM).

Design gennem parametre er ikke bare en trend, men den uundgåelige fremtid for byggebranchen. I stedet for at skabe komplekse 3D -modeller manuelt vil designere arbejde med data, validere dem og automatisere processer, hvilket bringer byggeriet tættere på programmeringsverdenen. Med tiden vil designprocesser blive bygget på principperne for softwareudvikling:

- Opret krav → Opret model→ Upload til server→ Valider ændringer→ Pull request
- Pull-anmodningen kører automatisk modeltjek i forhold til krav, der blev oprettet før eller under designprocessen.
- Efter datakvalitetstjek og godkendelse implementeres ændringerne i projektet, den fælles database eller overføres automatisk til andre systemer.

Inden for maskinteknik starter sådanne designændringer allerede nu med dannelsen af en ændringsmeddelelse. En lignende ordning venter på byggebranchen: Design vil være en iterativ proces, hvor hvert trin understøttes af parametriske krav. Et sådant system vil gøre det muligt for designere at oprette automatiserede kontroller og en automatiseret pull-anmodning til specifikke krav.

> Fremtidens designer er først og fremmest en dataoperatør, ikke en manuel modellerer. Hans opgave er at fylde projektet med parametriske enheder, hvor geometri kun er en af attributterne.

Det er forståelsen af vigtigheden af datamodellering, klassifikation og standardisering, som er blevet diskuteret i detaljer i de foregående kapitler i bogen, der vil spille en vigtig rolle i transformationen. Fremtidens designregler vil blive formaliseret som nøgle-værdi-parameterpar i form af XLSX eller XML -skemaer.

> Byggebranchens fremtid handler om at indsamle data, analysere dem, validere dem og automatisere processer ved hjælp af analyseværktøjer. BIM (eller CAD) er ikke det endelige mål, men kun et udviklingsstadie. Når fagfolk indser, at de kan arbejde direkte med data uden om de traditionelle CAD-værktøjer, vil selve udtrykket "BIM" gradvist vige for begreberne om brug af strukturerede og detaljerede byggeprojektdata.

En af de vigtigste faktorer, der har fremskyndet transformationen, har været fremkomsten af store sprogmodeller (LLM) og de værktøjer, der er baseret på dem. Disse teknologier ændrer den måde, designdata håndteres på, og giver adgang til information uden behov for indgående kendskab til API'er eller leverandørløsninger. Med LLM'er bliver processen med at skabe et krav og interagere med CAD data intuitiv og tilgængelig.

Fremkomsten af LLM i design CAD databehandlingsprocesser

Ud over udviklingen af værktøjer til adgang til CAD-databaser og åbne og forenklede CAD -formater er fremkomsten af LLM -værktøjer (Large Language Models) ved at revolutionere behandlingen af designdata. Hvor adgangen til information tidligere hovedsageligt foregik via komplekse grænseflader og krævede programmeringsfærdigheder og viden om API, er det nu muligt at interagere med data ved hjælp af naturligt sprog.

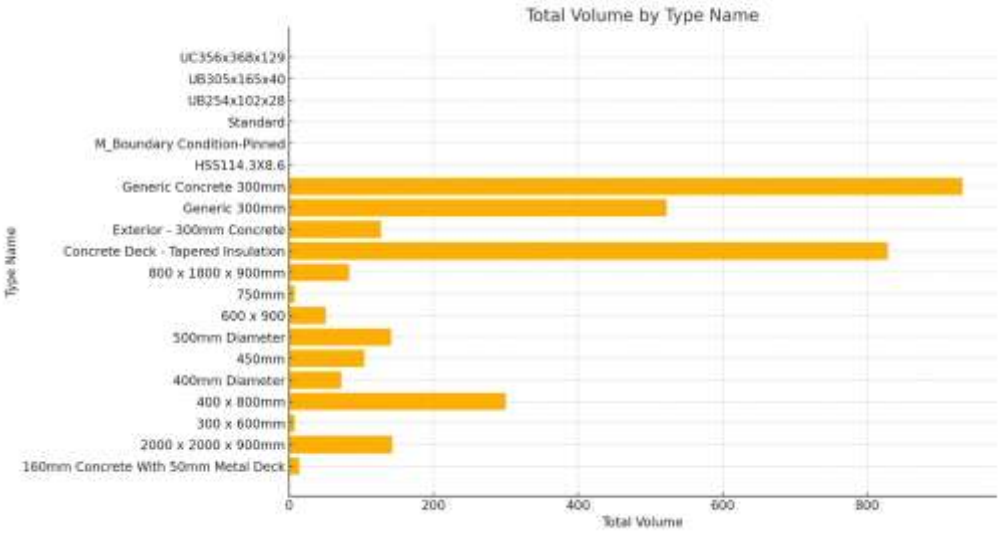
Ingeniører, ledere og planlæggere uden teknisk baggrund kan få de nødvendige oplysninger fra projektdata ved at formulere forespørgsler i almindeligt sprog. Forudsat at dataene er strukturerede og tilgængelige (fig. 4.1-13), er det nok at stille et spørgsmål i LLM chatrummet som: "*Vis i en tabel med gruppering efter type alle vægge med et volumen på mere end 10 kubikmeter*" - og modellen vil automatisk konvertere denne forespørgsel til SQL eller kode i Pandas, og generere en oversigtstabel, graf eller endda et færdigt dokument.

Nedenfor er nogle eksempler fra det virkelige liv på, hvordan LLM -modeller interagerer med designdata repræsenteret i forskellige CAD- (BIM-) formater.

- Eksempel på en forespørgsel LLM chat til et CAD-projekt i RVT format efter konvertering (Fig. 4.1-13) til en tabulær dataramme (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller en anden):

> Gruppér dataene i datarammen fra RVT-filen efter "Typenavn", når du summerer parameteren "Volumen", og vis antallet af elementer i gruppen. Og vis venligst alt dette som et vandret histogram uden nulværdier.

LLM-respons som et vandret søjlediagram (PNG-format):



Figur 6.4-5 I stedet for 17 museklik eller 40 linjer kode ved brug af plug-ins henter vi i LLM øjeblikkeligt tabellen QTO ved hjælp af en tekstforespørgsel.

For at generere en QTO -tabel med vægtyper med samlet areal og mængde fra kategorien "Vægge" skal vi formulere en tekstforespørgsel til LLM -chat:

> Tag kun de elementer fra projektets dataramme, der har "OST_Walls" i "Category"-parameteren, gruppér dem efter "Type Name", summer værdien af "Area"-kolonnen, tilføj antallet, og vis dem i tabellen ved at fjerne null-værdier.

LLM-svar i form af en færdig QTO-tabel:

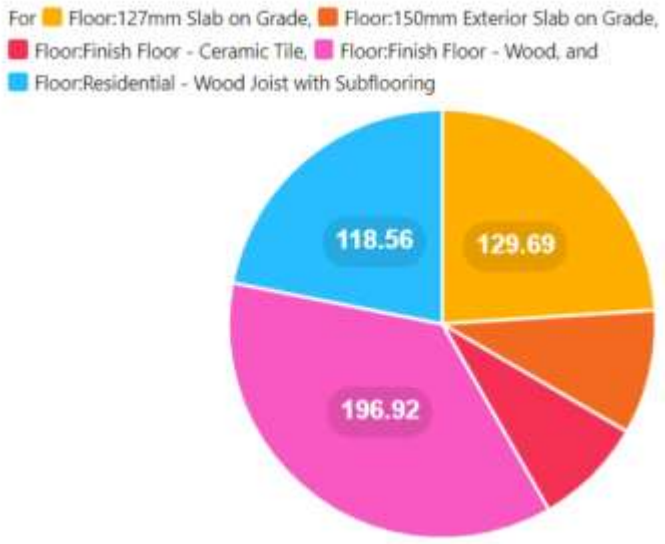
Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

Fig. 6.4-6 Oprettelse af en QTO tabel i naturligt sprog giver samme kvalitetsresultat som ved brug af CAD - (BIM-) værktøjer.

Lad os spørge ind til projektet i IFC -format efter at have konverteret det til en tabulær dataramme og indtaste en lignende tekstforespørgsel i en LLM -chat:

> Tag kun de elementer fra projektet, der har Level 1- og Level 2-værdier i parameteren "Parent", og tag de elementer, der har IfcSlab-værdier i parameteren "Category", gruppér derefter disse elementer efter parameteren "ObjectType", summer værdierne i parameteren "PSet_RVT _Dimensions Area", og vis dem som et cirkeldiagram.

- LLM-svar som et færdigt cirkeldiagram over elementgrupper fra IFC-data :



Figur 6.4-7 Resultatet af en forespørgsel på IFC -data i et struktureret format kan være en hvilken som helst type graf, der er praktisk for at forstå dataene.

Bag hver af de resulterende færdige løsninger (Fig. 6.4-5 - Fig. 6.4-7) er der et dusin linjer Python-kode, der bruger Pandas-biblioteket. Den resulterende kode kan kopieres fra LLM-chatrummet og bruges i et hvilket som helst lokalt eller online IDE for at få identiske resultater uden for LLM-chatrummet.

I den samme LLM-chat kan vi ikke kun arbejde med projekter hentet fra 3D CAD (BIM) formater, men også med flade tegninger i DWG-format, som vi kan spørge LLM chat om at vise, for eksempel data om elementgrupper i form af linjer eller 3D-geometrier efter konvertering til en struktureret form.

Automatiseret analyse af DWG -filer med LLM og Pandas

Processen med databehandling fra DWG -filer har på grund af informationens ustrukturerede karakter altid været en kompleks opgave, der kræver specialiseret software og ofte manuel analyse. Men med udviklingen af kunstig intelligens og LLM-værktøjer er det blevet muligt at automatisere mange trin i denne, i dag, for det meste manuelle proces. Lad os se på en rigtig pipeline med anmodninger til LLM (i dette eksempel ChatGPT) om at arbejde med DWG-tegninger, som giver dig mulighed for at arbejde med projektet:

- Filtrer DWG-data efter lag, ID og koordinater
- Visualiser elementernes geometri
- Annoter automatisk tegninger baseret på parametre
- Udvid vægpolylinjer til det vandrette plan
- Skab interaktive 3D -visualiseringer af plane data

- Strukturer og analyser konstruktionsdata uden komplekse CAD -værktøjer

I vores tilfælde starter processen med at opbygge Pipelinesekventiel kodegenerering via LLM. Først genereres en forespørgsel, der beskriver opgaven. ChatGPT genererer Python -kode, som udføres og analyseres og viser resultatet inde i chatrummet. Hvis resultatet ikke er som forventet, rettes forespørgslen, og processen gentages

> **Pipeline** er en sekvens af automatiserede trin, der udføres for at behandle og analysere data. I en sådan proces tager hvert trin data som input, udfører transformationer og sender resultatet videre til det næste trin.

Når det ønskede resultat er opnået, kopieres koden fra LLM og indsættes i koden i form af blokke i et af de praktiske IDE'er, i vores tilfælde på Kaggle-platformen.com. De resulterende kodefragmenter kombineres til en enkelt pipeline, som automatiserer hele processen - fra dataindlæsning til den endelige analyse. Denne tilgang giver mulighed for hurtig udvikling og skalering af analytiske processer uden dyb programmeringsekspertise. Den fulde kode for alle nedenstående fragmenter kan sammen med eksempler på forespørgsler findes på Kaggle.com-platformen ved at søge på "DWG Analyse with ChatGPT | DataDrivenConstruction" [129].

Lad os starte processen med at arbejde med DWG-data efter konvertering til struktureret form (Fig. 4.1-13), med et klassisk trin - gruppering og filtrering af alle tegningsdata, der er nødvendige for vores opgave, vægelementer, specifikt polyliner (parameter 'ParentID' gør det muligt at gruppere linjer i grupper), som i parameteren (dataframe-kolonne) "Layer" har en strengværdi, der indeholder følgende bogstavkombination (RegEx) - "wall".

- For at få koden til en lignende opgave og resultatet i form af et billede skal du skrive følgende forespørgsel i LLM:

> Tjek først, om den dataramme, der er hentet fra DWG, indeholder de definerede kolonner: 'Layer', 'ID', 'ParentID' og 'Point'. Filtrer derefter de ID'er fra kolonnen 'Layer', der indeholder strengen 'wall'. Find de elementer i kolonnen 'ParentID', der matcher disse identifikatorer. Definer en funktion til at rense og opdele dataene i kolonnen 'Point'. Dette inkluderer at fjerne parenteser og opdele værdierne i 'x', 'y' og 'z'-koordinater. Plot dataene ved hjælp af matplotlib. Tegn en separat polylinje, der forbinder 'Point'-koordinaterne, for hvert unikt 'ParentID'. Sørg for, at det første og sidste punkt er forbundet, hvis det er muligt. Indstil de relevante etiketter og titler, og sørg for, at x- og y-akserne er skaleret ens.

- Svaret LLM vil give dig et færdigt billede, bag hvilket Python-koden, der genererede det, er skjult:

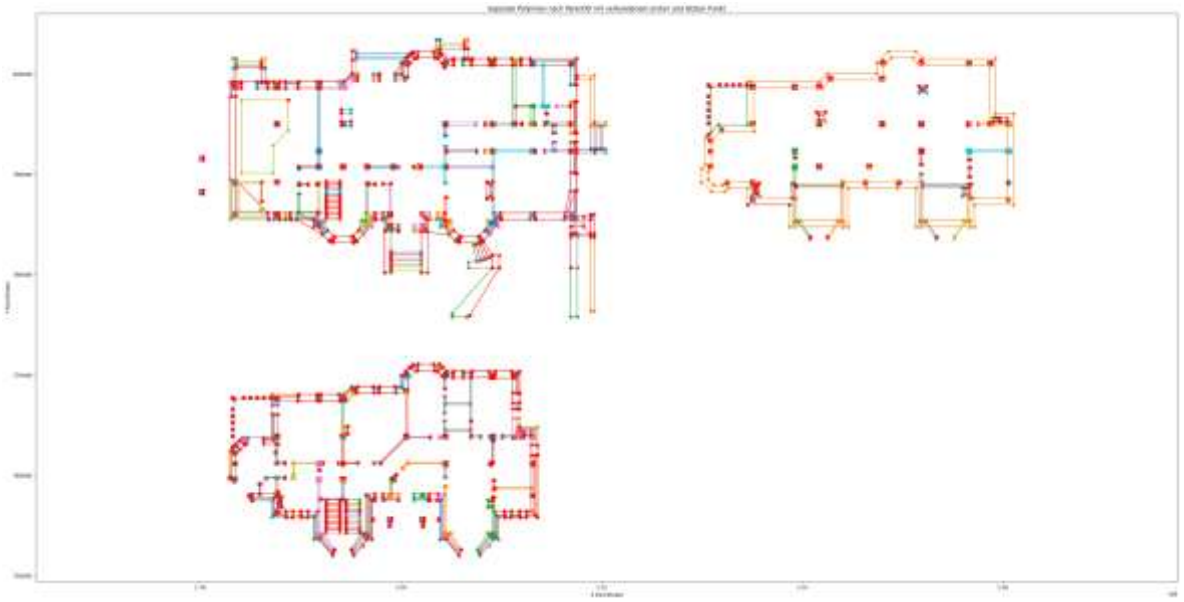


Fig. 6.4-8 LLM -koden udtrak alle linjerne i "væg"-laget fra DWG -filen, rensede deres koordinater og konstruerede polylinjerne ved hjælp af et af Python-bibliotekerne.

Lad os nu tilføje den arealparameter til linjerne, som hver polylinje har i sine egenskaber (i en af dataramme-kolonnerne):

> Find nu kun et "ParentID" fra hver polylinje - find dette ID i "ID"-kolonnen, tag "Area"-værdien, divider med 1.000., og tilføj denne værdi til grafen.

LLM-svaret viser en ny graf, hvor hver polylinje har en overskrift med sit areal:

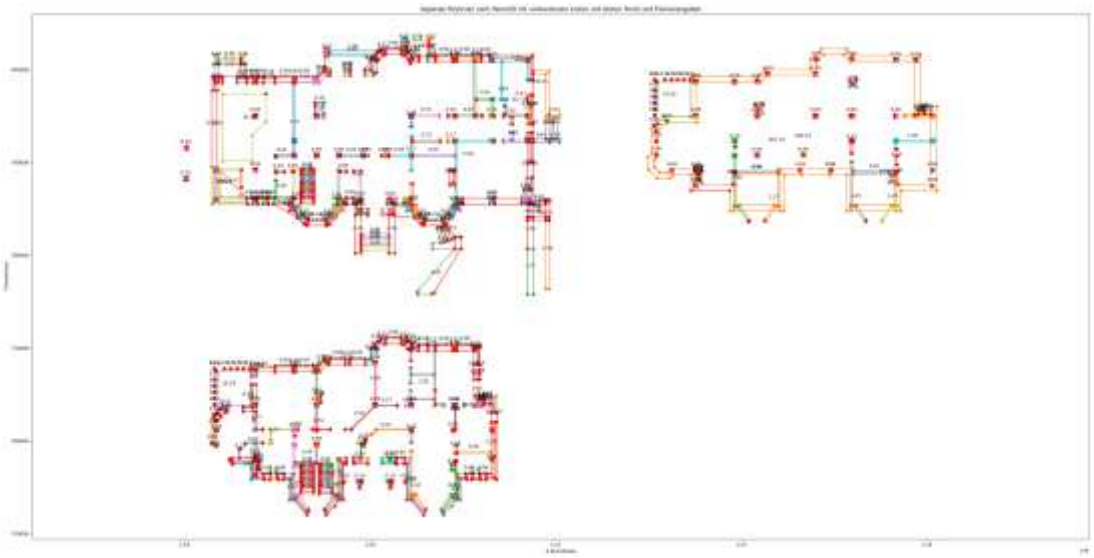


Fig. 6.4-9 LLM har tilføjet kode, der tager arealværdierne for hver polylinje og føjer dem til billedet med linjevisualisering.

Derefter konverterer vi hver polylinje til en vandret linje, tilføjer en parallel linje i en højde af 3000 mm og forbinder dem til et enkelt plan for på denne måde at vise layoutet af vægelementernes overflader:

> Du skal tage alle elementer fra kolonnen "Layer" med værdien "wall". Tag disse ID'er som en liste fra kolonnen "ID", og find disse ID'er fra hele dataarammen i kolonnen "ParentID". Alle elementer er linjer, der er kombineret til en enkelt polylinje. Hver linje har en forskellig x, y-geometri for det første punkt i kolonnen "Point". Du skal tage hver polylinje efter tur og fra punktet 0,0 horisontalt tegne længden af hvert segment fra polylinjen. længden af hvert segment af polylinjen til en linje. Tegn derefter nøjagtig de samme linjer, bare 3000 højere, og forbind alle punkter til et plan.

LLM-svaret udsender kode, der giver dig mulighed for at plotte vægtegninger i planet:

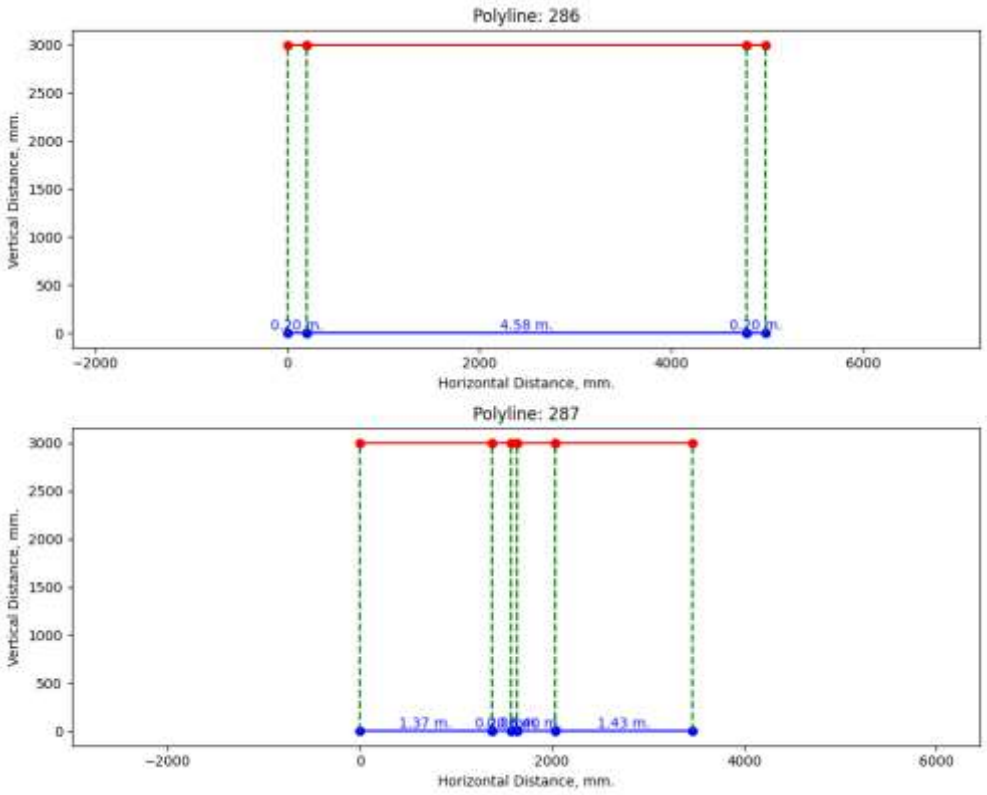


Fig. 6.4-10 Vi omdanner hver polylinje ved hjælp af prompts til et layout, der visualiserer vægplanerne direkte i LLM-chatten.

Lad os nu gå fra 2D-projektion til 3D - modellering af vægge ud fra flade linjer ved at forbinde de øverste og nederste lag af polylinjer:

> Visualiser vægelementer i 3D ved at forbinde polylinjer i højderne z = 0 og z = 3000 mm. For at skabe en lukket geometri, der repræsenterer bygningens vægge. Brug Matplotlib 3D-grafværktøj.

- LLM vil generere en interaktiv 3D -graf, hvor hver polylinje vil blive repræsenteret som et sæt planer. Brugeren vil kunne bevæge sig frit mellem elementer med en computermus og udforske modellen i 3D-tilstand ved at kopiere koden fra chatten til IDE'en:

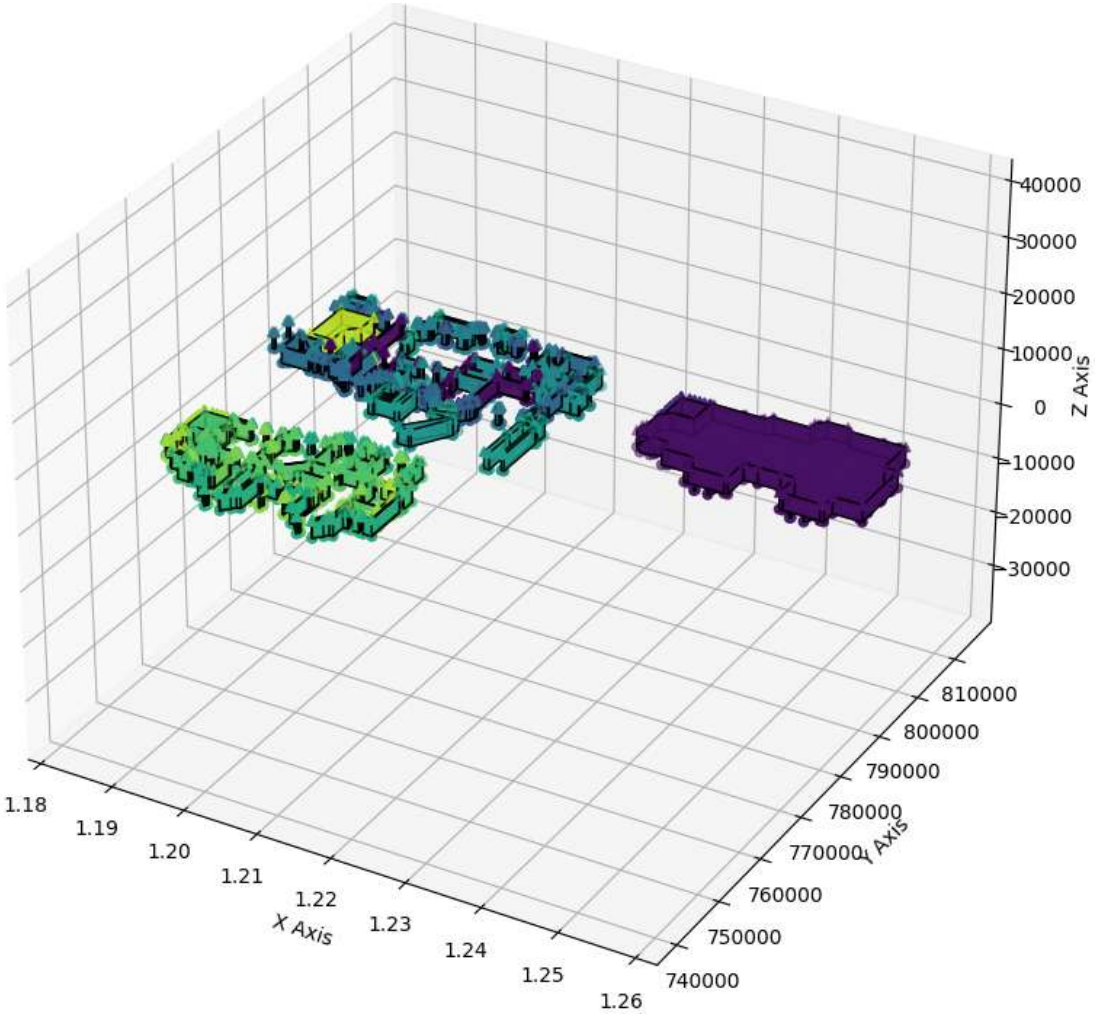


Fig. 6.4-11 LLM hjalp med at bygge kode [129] til at visualisere flade tegningslinjer i en 3D visning, der kan udforskes i 3D-vieweren i IDE'et.

For at opbygge en logisk og reproducerbar pipeline - fra den første konvertering og indlæsning af DWG - filen til det endelige resultat - anbefales det at kopiere den genererede LLM -kodeblok til IDE'en efter hvert trin. På den måde kan du ikke kun tjekke resultatet i chatten, men også køre det i dit udviklingsmiljø med det samme. Det giver dig mulighed for at opbygge processen sekventielt, fejlfinde og tilpasse den efter behov.

Du kan finde den komplette Pipeline kode for alle fragmenter (figurerne 6.4-8 til figurerne 6.4-11) sammen med eksempler på forespørgsler på Kaggle-platformen.com ved at søge på "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. På Kaggle kan du ikke kun se koden og de anvendte forespørgsler, men også kopiere og teste hele pipelinen med de originale DWG-datafelter i skyen gratis uden at skulle

installere yderligere software eller selve IDE'et.

Den tilgang, der præsenteres i dette kapitel, giver dig mulighed for fuldt ud at automatisere kontrol, behandling og generering af dokumenter baseret på DWG -projekter. Den udviklede pipeline er velegnet både til behandling af individuelle tegninger og til batchbehandling af ti, hundrede og tusindvis af DWG-filer med automatisk generering af de nødvendige rapporter og visualiseringer for hvert projekt.

Processen kan organiseres på en sekventiel og gennemsigtig måde: Først konverteres dataene fra CAD -filen automatisk til XLSX-format, derefter indlæses de i en dataramme, efterfulgt af gruppering, kontrol og resultatgenerering - alt dette implementeres i en enkelt Jupyter-notebook eller Python -skript, i enhver populær IDE. Hvis det er nødvendigt, kan processen nemt udvides gennem integration med systemer til styring af projektdokumentation: CAD-filer kan hentes automatisk efter bestemte kriterier, resultaterne kan sendes tilbage til lagringssystemet, og brugerne kan få besked, når resultaterne er klar - via e-mail eller messengers.

> Brug af LLM chats og agenter til at arbejde med designdata reducerer afhængigheden af specialiserede CAD -programmer og giver dig mulighed for at udføre analyse og visualisering af arkitektoniske designs uden behov for manuel interaktion med grænsefladen - uden museklik og uden at skulle huske kompleks menunavigation.

For hver dag der går, vil byggebranchen høre mere og mere om LLM, granulære strukturerede data, DataFrames og kolonnedatabaser. Ensartede todimensionelle DataFrames dannet af forskellige databaser og CAD-formater, vil være det ideelle brændstof til moderne analyseværktøjer, der håndteres aktivt af specialister i andre brancher.

Selve automatiseringsprocessen vil blive væsentligt forenklet - i stedet for at studere API for lukkede nicheprodukter og skrive komplekse scripts for at analysere eller transformere parametre, vil det nu være nok at formulere en opgave i form af et sæt individuelle tekstkommandoer, som vil blive foldet ind i den nødvendige Pipeline eller Workflow-proces for det nødvendige programmeringssprog, som kører gratis på næsten enhver enhed. Det er slut med at vente på nye produkter, formater, plug-ins eller opdateringer fra CAD- (BIM-) værktøjsleverandører. Ingeniører og bygherrer vil kunne arbejde selvstændigt med data ved hjælp af enkle, gratis og letforståelige værktøjer, assisteret af LLM chats og agenter.

Næste skridt: fra lukkede formater til åbne data

Når man arbejder med fremtidens designdata, er det usandsynligt, at nogen virkelig har brug for at forstå de geometriske kerner i proprietære værktøjer eller lære hundredvis af inkompatible formater, der indeholder de samme oplysninger. Men uden at forstå, hvorfor overgangen til åbne strukturerede data er vigtig, er det svært at argumentere for brugen af nye gratis værktøjer, åbne data og tilgange, der sandsynligvis ikke vil blive fremmet af softwareleverandører.

I dette kapitel har vi diskuteret de vigtigste egenskaber ved CAD (BIM) data, deres begrænsninger og muligheder, og at ingeniører og designere på trods af leverandørernes markedsføringsløfter står over for vanskeligheder med at udtrække, overføre og analysere designinformation hver dag. At forstå arkitekturen i disse systemer og lære om alternative tilgange - baseret på åbne formater og automatisering gennem LLM - kan gøre livet meget lettere for selv en enkelt fagperson, for slet ikke at tale om virksomheder. For at opsummere denne del er det værd at fremhæve de vigtigste praktiske trin, der vil hjælpe dig med at anvende

de tilgange, der diskuteres i dette afsnit, på dine daglige opgaver:

- Udvid din værktøjskasse til at arbejde med projektdata
  - [ ] Udforsk tilgængelige plug-ins og værktøjer til at udtrække data fra de CAD - (BIM-) systemer, du bruger.
  - [ ] Udforsk tilgængelige SDK'er og API'er, der giver dig mulighed for at automatisere dataudtræk fra lukkede formater uden manuelt at skulle åbne specialiseret software.
  - [ ] Beherske grundlæggende færdigheder i at arbejde med åbne ikke-parametriske geometriformater (OBJ, glTF, USD, DAE) og tilsvarende open source-biblioteker.
  - [ ] Prøv at tænke på et system til lagring af projektmetadata adskilt fra geometri uden for CAD (BIM) løsninger for at forenkle analyse og integration med andre systemer.
  - [ ] Brug LLM til at automatisere problemer med datakonvertering mellem formater
- Skab dine egne processer til håndtering af projektinformation
  - [ ] Begynde at beskrive opgaver og modelleringskrav ved hjælp af parametre og deres værdier i enkle og strukturerede formater.
  - [ ] Opret et personligt bibliotek med scripts eller kodeblokke til ofte udførte operationer
- Fremme brugen af åbne standarder i dit arbejde
  - [ ] Inviter kolleger og partnere til at dele data i åbne formater, der ikke er begrænset af softwareleverandørens økosystem.
  - [ ] Demonstrer fordelene ved at bruge strukturerede data med specifikke eksempler
  - [ ] Indled diskussioner om problemer med lukkede formater og mulige løsninger

Selv om du ikke kan ændre din virksomheds politik vedrørende CAD - (BIM-) platforme, vil en personlig forståelse af principperne for at arbejde med projektdata i åbne formater give dig mulighed for at øge effektiviteten i dit arbejde betydeligt. Ved at skabe dine egne værktøjer og metoder til at udtrække og omdanne data fra forskellige formater optimerer du ikke kun dine arbejdsgange, men får også fleksibiliteten til at omgå begrænsningerne i standardsoftwareløsninger.



VII DEL

DATADREVET BESLUTNINGSTAGNING, ANALYSE, AUTOMATISERING OG MASKINLÆRING

Den syvende del fokuserer på dataanalyse og procesautomatisering i byggebranchen. Den diskuterer, hvordan data bliver grundlaget for beslutningstagning, og forklarer principperne for visualisering af information til effektiv analyse. Key performance indicators (KPI), metoder til evaluering af investeringsafkast (ROI) og oprettelse af dashboards til projektovervågning beskrives i detaljer. Der lægges særlig vægt på ETL-processer (Extract, Transform, Load) og deres automatisering ved hjælp af pipelines (Pipeline) for at omdanne forskelligartede data til struktureret information til analyse. Workflow-orkestreringsværktøjer som Apache Airflow, Apache NiFi og n8n, som gør det muligt at opbygge automatiserede datapipelines uden dyb viden om programmering, diskuteres. Store sprogmodeller (LLM'er) og deres brug til at forenkle dataanalyse og automatisere rutineopgaver spiller en vigtig rolle

KAPITEL 7.1.
DATAANALYSE OG DATADREVET BESLUTNINGSTAGNING

Efter at have indsamlet, struktureret, renset og verificeret informationerne er der opstået et sammenhængende og analyserbart datasæt. De tidligere dele af bogen dækkede systematisering og strukturering af heterogene kilder - fra PDF-dokumenter og tekstoptegnelser over møder til CAD-modeller og geometriske data. Processen med at kontrollere og tilpasse oplysninger til kravene i forskellige systemer og klassifikatorer og fjerne dubletter og uoverensstemmelser er beskrevet i detaljer.

Alle de beregninger, der udføres på disse data (tredje og fjerde del af bogen) - fra simple transformationer til beregninger af tid, omkostninger og ESG-indikatorer (femte del) - er samlede analyseopgaver. De danner grundlag for at forstå et projekts aktuelle tilstand, vurdere dets parametre og derefter træffe beslutninger. Som et resultat af beregningerne forvandles dataene fra et sæt forskelligartede poster til en håndterbar ressource, der kan besvare vigtige forretningsspørgsmål.

De foregående kapitler har beskrevet dataindsamlings- og kvalitetskontrolprocesserne til brug i typiske forretningssager og processer, der er specifikke for byggebranchen. Analyse i denne sammenhæng ligner på mange måder applikationer i andre brancher, men har en række specifikke funktioner.

De følgende kapitler beskriver dataanalyseprocessen i detaljer, herunder automatiseringstrinnene fra den indledende indsamling af information og dens transformation til den efterfølgende overførsel til målsystemer og dokumenter. Først præsenteres en teoretisk del med fokus på udvalgte aspekter af datadrevet beslutningstagning. Derefter, i de følgende kapitler, den praktiske del i forbindelse med automatisering og konstruktion af ETL -Pipeline.

Data som en ressource i beslutningsprocessen

Datadrevet beslutningstagning er ofte en iterativ proces og begynder med systematisk indsamling af information fra en række forskellige informationskilder. Som en naturlig cyklus falder individuelle dataelementer og hele informationssystemer gradvist ned i jorden - og akkumuleres i virksomhedernes informationslagre (figur 1.3-2). Med tiden bliver disse data, ligesom nedfaldne blade og grene, omdannet til værdifuldt materiale. Myceliet af dataingeniører og analytikere organiserer og forbereder information til fremtidig brug og forvandler nedfaldne data og systemer til værdifuld kompost, så der kan vokse nye skud og nye systemer (fig. 1.2-5).

Tendenser til udbredt brug af analyser i forskellige brancher markerer begyndelsen på en ny æra, hvor det at arbejde med data bliver grundlaget for professionel aktivitet (fig. 7.1-1). Det er vigtigt, at fagfolk i byggebranchen tilpasser sig disse ændringer og er klar til at bevæge sig ind i en ny æra - æraen med data og analyser

> At flytte data manuelt mellem tabeller og udføre beregninger manuelt hører gradvist fortiden til og giver plads til automatisering, dataflowanalyse, analyse og maskinlæring. Disse værktøjer er ved at blive nøgleelementer i moderne beslutningsstøttesystemer.

I McKinseys bog "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] citeres en undersøgelse foretaget i 2022 med 1.330 topledere fra forskellige regioner, brancher og funktionsområder. I en undersøgelse foretaget i 2022 med 1.330 topledere fra forskellige regioner, brancher og funktionsområder citeres [130]. Ifølge resultaterne bruger 70% af lederne avancerede analyser til at generere deres egne ideer, og 50% implementerer kunstig intelligens til at forbedre og automatisere beslutningsprocesser.



Fig. 7.1-1 Dataanalyse og analyse er det vigtigste værktøj til at øge hastigheden af beslutningstagningen i en virksomhed.

Dataanalyse trænger ligesom mycelium ned i humus fra tidligere beslutninger, hjælper med at forbinde individuelle systemer og guider ledere til værdifuld indsigt. Denne viden, ligesom næringsstoffer fra rådne datasystemtræer, giver næring til nye beslutninger i virksomheden, hvilket fører til effektiv forandring og kvalitetsinformationsvækst, ligesom nye skud og spirer, der kommer ud af rig og sund jord (figur 1.2-5).

> *Tal har en vigtig historie at fortælle. De regner med, at du giver dem en klar og overbevisende stemme* [131].
>
> - Stephen Few, ekspert i datavisualisering

I mellemstore og små virksomheder er arbejdet med at udtrække og forberede information til videre analyse i dag en ekstremt arbejdskrævende proces (figur 7.1-2), der kan sammenlignes med kulminedrift i det 18. århundrede. Indtil for nylig var arbejdet med datamining og -forberedelse snarere forbeholdt eventyrere, der arbejdede i en højt specialiseret niche med et lille og begrænset sæt værktøjer til at arbejde med forskellige typer data fra ustrukturerede, løst strukturerede, blandede og lukkede kilder.

Beslutningstagere og ledere er ofte uerfarne med heterogene data og systemer, men er alligevel nødt til at træffe beslutninger baseret på dem. Som følge heraf har datadrevet beslutningstagning i den moderne byggebranche i de seneste årtier været mindre som en automatiseret proces og mere som en minearbejders manuelle flerdagesarbejde i de tidlige kulminer.

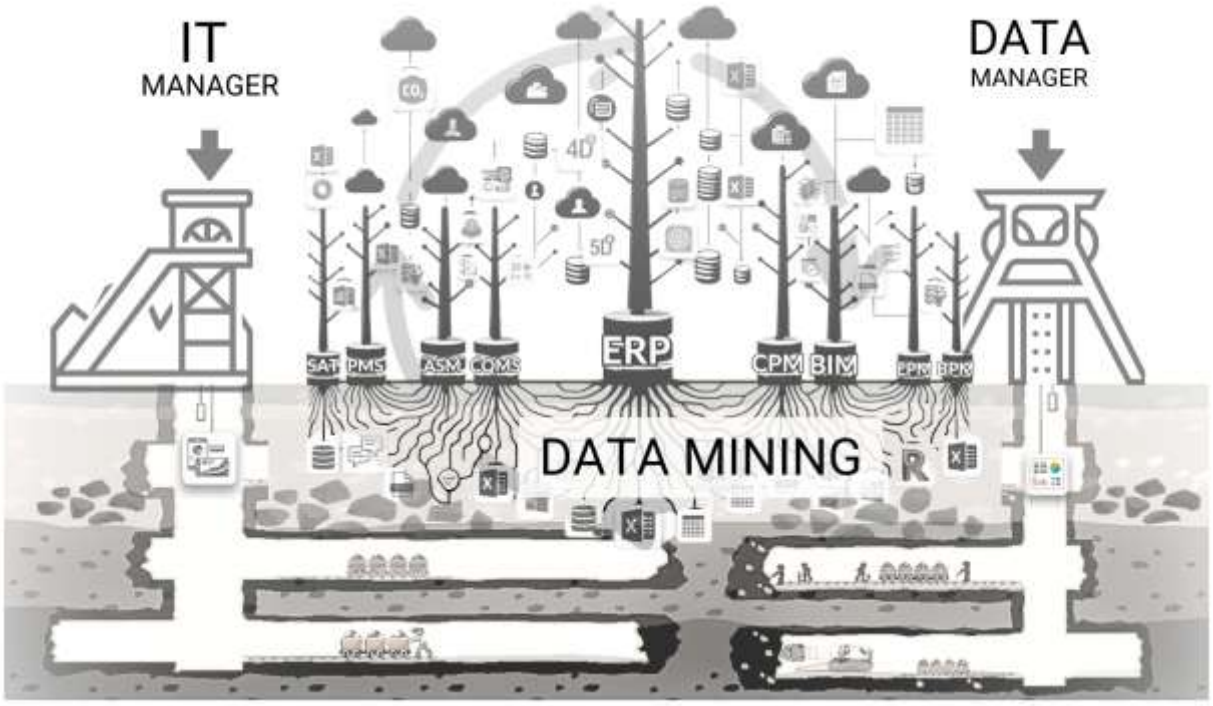


Fig. 7.1-2 I dataminingprocessen gennemgår eksperterne en kompleks proces med forberedelse af data - fra rensning til strukturering med henblik på efterfølgende analyse.

Selv om moderne metoder til dataudtræk i byggebranchen bestemt er mere avancerede end de primitive teknikker, som minearbejderne brugte i det 12. århundrede, er det stadig en kompleks og risikofyldt opgave, der kræver betydelige ressourcer og ekspertise, som kun store virksomheder har råd til. Processerne med at udtrække og analysere data fra den akkumulerede arv fra tidligere projekter er indtil for nylig overvejende blevet udført af store, teknologisk avancerede virksomheder, der har indsamlet og lagret data konsekvent i årtier

> Tidligere blev den førende rolle inden for analyse spillet af teknologisk modne virksomheder, der havde akkumuleret data i årtier. I dag er situationen ved at ændre sig: Adgangen til data og databehandlingsværktøjer er ved at blive demokratiseret - tidligere komplekse løsninger er nu gratis tilgængelige for alle.

Ved at anvende analyser kan virksomheder træffe mere præcise og informerede beslutninger i realtid. Følgende casestudie illustrerer, hvordan historiske data kan hjælpe med at træffe økonomisk sunde beslutninger:

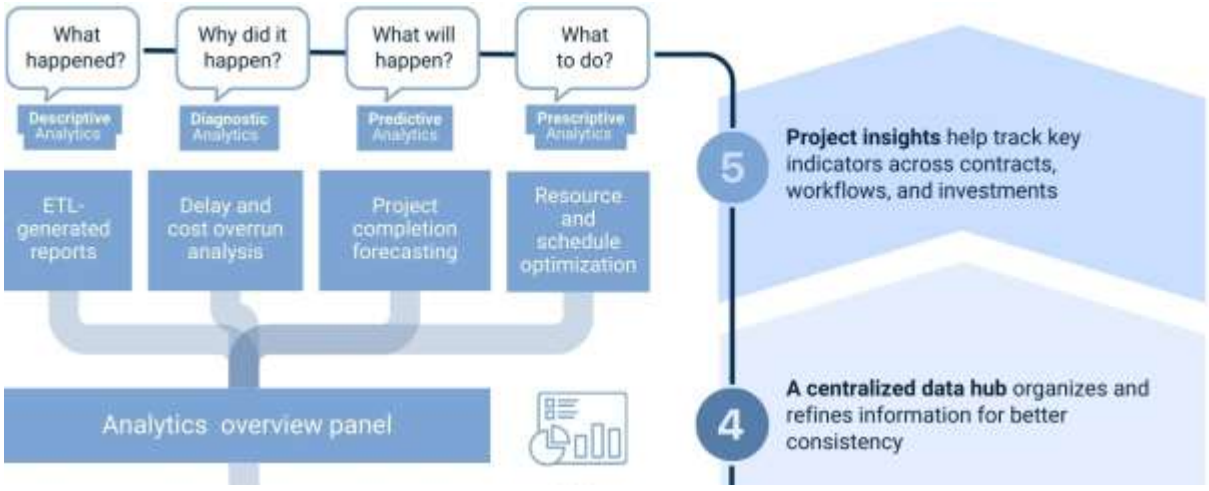
- **Projektleder** - "*Nu er den gennemsnitlige pris på beton i byen 82€ /m³, vi har 95 €/m³ i overslaget."*
- **Estimator** - "*På tidligere projekter var overskridelsen ca. 15%, så jeg trak mig tilbage*."
- **Datamanager eller kontrolingeniør på kundesiden** - "*Lad os se på analyserne for de sidste tre udbud*."

Efter at have analyseret DataFrame fra tidligere projekter får vi:

- **Gennemsnitlig faktisk købspris:** 84,80 /m³€

- **Gennemsnitlig overskridelsesprocent:** +4,7%.
- **Anbefalet hastighed i estimatet:** ~ 85 /m³€

En sådan beslutning vil ikke længere være baseret på subjektive følelser, men på specifikke historiske statistikker, som hjælper med at reducere risici og øge udbuddets gyldighed. Analyse af data fra tidligere projekter bliver en slags "organisk gødning", hvorfra nye, mere præcise løsninger spirer.



Figur 7.1-3 Dataanalyse besvarer tre centrale spørgsmål: Hvad skete der, hvorfor skete det, og hvad skal der gøres nu?

Beslutningstagere og ledere står ofte over for behovet for at arbejde med heterogene data og systemer uden tilstrækkelig teknisk ekspertise. I sådanne situationer er visualisering, som er et af de første og vigtigste trin i den analytiske proces, en vigtig hjælp til at forstå data. Det gør det muligt at præsentere information i en visuel og forståelig form.

Visualisering af data: nøglen til forståelse og beslutningstagning

I dagens byggebranche, hvor projektdata er kendetegnet ved kompleksitet og struktur på flere niveauer, spiller visualisering en nøglerolle. Visualisering af data giver projektledere og ingeniører mulighed for at visualisere komplekse mønstre og tendenser, der er skjult i store, heterogene datamængder.

> Visualisering af data gør det lettere at forstå et projekts status: ressourceallokering, omkostningstendenser eller materialeforbrug. Grafer og diagrammer gør kompleks og tør information tilgængelig og forståelig, så du hurtigt kan identificere nøgleområder, der kræver opmærksomhed, og få øje på potentielle problemer.

Visualisering af data letter ikke kun fortolkningen af information, det er et afgørende skridt i den analytiske proces og informeret ledelsesbeslutningstagning, der hjælper med *at besvare spø*rgsmå*lene "hvad ske*te *d*er?" og "hvordan skete det?" (Figur 2.2-5). (Figur 2.2-5).

> *Grafik er visuelle værktøjer til at løse logiske problemer [132].*
>
> - Jacques Bertin, "Grafik og grafisk informationsbehandling

Før de træffer vigtige beslutninger, er projektledere mere tilbøjelige til at bruge visuelle repræsentationer af data i stedet for tørre og svært fortolkelige tal fra regneark eller tekstbeskeder.

> Data uden visualisering er som byggematerialer, der er strøet tilfældigt rundt på en byggeplads: Deres potentiale er uklart. Det er først, når de visualiseres tydeligt, som et hus lavet af mursten og beton, at deres værdi bliver klar. Før huset er bygget, er det umuligt at sige, om bunken af materialer bliver til en lille hytte, en luksusvilla eller en skyskraber.

Virksomheder har data fra forskellige systemer (Figur 1.2-4 til Figur 2.1-10), finansielle transaktioner og omfattende tekstdata. Men det er ofte en udfordring at udnytte disse data til gavn for forretningen. I sådanne situationer bliver visualisering et vigtigt redskab til at kommunikere betydningen af dataene og hjælpe med at præsentere oplysningerne i formater, som enhver ekspert kan forstå, f.eks. dashboards, grafer og diagrammer.

> PwC's undersøgelse "What Students Need to Succeed in a Rapidly Changing Business World" (2015) fremhæver [9], at succesfulde virksomheder ikke begrænser sig til dataanalyse, men aktivt bruger interaktive visualiseringsværktøjer som grafer, infografik og analytiske dashboards til at understøtte beslutningstagningen. Ifølge rapporten hjælper datavisualisering kunderne med at forstå den historie, som data fortæller gennem diagrammer, grafer, diagrammer, dashboards og interaktive datamodeller.

Processen med at konvertere information til visuelle grafiske former som diagrammer, grafer og skemaer forbedrer den menneskelige hjernes forståelse og fortolkning af data (figur 7.1-4). Det giver projektledere og analytikere mulighed for hurtigere at vurdere komplekse scenarier og træffe informerede beslutninger baseret på visuelt genkendelige tendenser og mønstre i stedet for intuition.

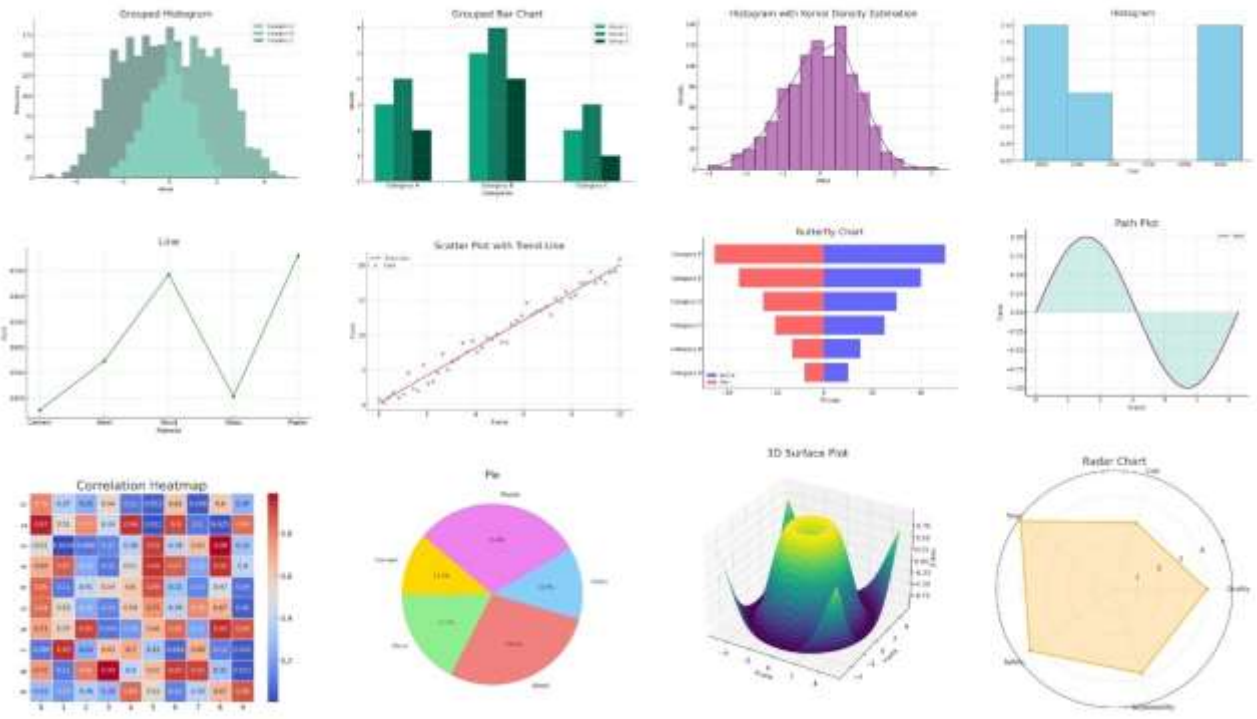


Fig. 7.1-4 Forskellige former for visualisering er designet til at hjælpe den menneskelige hjerne med bedre at forstå og give mening til tallenes tørre information.

Oprettelsen af visualiseringer fra data og brugen af forskellige gratis visualiseringsbiblioteker vil blive diskuteret mere detaljeret i det næste kapitel om ETL -processer.

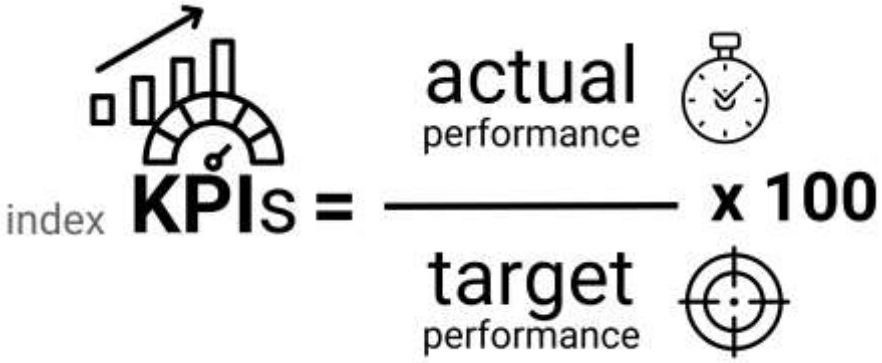
Visualisering er ved at blive et integreret element i arbejdet med data i byggebranchen - det hjælper ikke bare med at "se" data, men også med at forstå deres betydning i forbindelse med ledelsesopgaver. Men for at visualisering virkelig skal være nyttig, er det nødvendigt på forhånd at afgøre, hvad der præcist skal visualiseres, og hvilke målinger der virkelig er vigtige for at vurdere projektets resultater. Det er her, præstationsmålinger som KPI'er og ROI kommer ind i billedet. Uden dem risikerer selv de smukkeste dashboards bare at være "informationsstøj".

KPIog ROI

I dagens byggebranche spiller styringen af præstationsindikatorer (KPI og ROI) og visualiseringen af dem gennem rapporter og dashboards en central rolle i forbedringen af produktiviteten og effektiviteten i projektstyringen.

> Som i enhver anden virksomhed er det også i byggeriet nødvendigt klart at definere de parametre, som succes, investeringsafkast og præstationer måles på. Når man indhenter data om forskellige processer, skal en datadrevet organisation først lære at identificere **de vigtigste KPI'er (Key Performance Indicators)** - kvantitative mål, der afspejler, i hvor høj grad strategiske og operationelle mål nås.

For at beregne en KPI bruges der normalt en formel (fig. 7.1-5), som omfatter faktiske og planlagte indikatorer. For at beregne en individuel KPI for et projekt, en medarbejder eller en proces skal man f.eks. dividere den faktiske præstation med den planlagte præstation og gange resultatet med 100%.



Figur 7.1-5 KPI'er bruges til at måle et projekts eller en proces' succes med at nå de vigtigste mål.

På lokalitetsniveau kan man bruge mere detaljerede KPI'er metrics:

- **Timing af vigtige milepæle** (fundament, installation, færdiggørelse) - gør det muligt at kontrollere overholdelse af arbejdsplaner.
- **Procentdel af materialeoverskridelser** - hjælper med at styre indkøb og minimere spild.
- **Antal uplanlagte maskinnedetider** - påvirker produktivitet og omkostninger.

Hvis man vælger de forkerte parametre, kan det føre til fejlagtige *"hvad skal jeg gøre?*"-beslutninger (figur 2.2-5). Hvis en virksomhed f.eks. kun fokuserer på omkostninger pr. kvadratmeter, men ikke tager højde for omkostningerne ved ændringer, kan besparelser på materialer føre til dårligere kvalitet og højere omkostninger i fremtidige projekter.

Når man opstiller mål, er det vigtigt at være klar over, hvad der skal måles. Vage formuleringer fører til forkerte konklusioner og komplicerer kontrollen. Lad os se på eksempler på vellykkede og mislykkede KPI'er i byggeriet.

Gode KPI'er:

- *"Inden årets udgang skal du reducere andelen af renoveringsarbejde med* 10 *procent.*"
- "*Øg hastigheden på facademontage med 15 procent uden at gå på kompromis med kvaliteten i næste kvartal*"
- "*Reducer maskinernes nedetid med 20% ved at optimere arbejdsplanerne inden årets udgang*"

Disse målinger er klart målbare, med specifikke værdier og tidsrammer.

Dårlige KPI'er:

- "*Vi vil bygge hurtigere*" (Hvor meget hurtigere? Hvad betyder "hurtigere"?).
- "*Vi vil forbedre kvaliteten af det konkrete arbejde*" (Hvordan måles kvalitet helt præcist?)
- "*Vi vil forbedre samspillet med entreprenøren på byggepladsen*" (Hvilke kriterier vil vise

forbedring?)

> En god KPI er en, der kan måles og vurderes objektivt. I byggeriet er det særligt vigtigt, for uden klare indikatorer er det umuligt at overvåge præstationerne og opnå stabile resultater.

Ud over KPI'en er der en yderligere måleenhed til at vurdere investeringernes effektivitet: **ROI (Return on Investment)** - en indikator for investeringsafkast, der afspejler forholdet mellem overskud og investerede midler. ROI hjælper med at vurdere, om indførelsen af nye metoder, teknologier eller værktøjer er berettiget: fra digitale løsninger og automatisering (se f.eks. figur 7.3-2til brugen af nye byggematerialer. Denne indikator hjælper med at træffe informerede beslutninger om yderligere investeringer baseret på deres reelle indvirkning på virksomhedens rentabilitet

> I forbindelse med byggeprojektledelse kan ROI (return on investment) bruges som en af de vigtigste præstationsindikatorer (KPI), hvis en virksomheds mål er at måle investeringsafkastet af et projekt, en teknologi eller en procesforbedring. Hvis man f.eks. er ved at implementere en ny byggestyringsteknik, kan ROI vise, hvor meget den har forbedret rentabiliteten.

Regelmæssig måling af KPI'er og ROI'er baseret på data indsamlet fra forskellige kilder, f.eks. materialeforbrug, arbejdstimer og omkostninger, gør det muligt for projektledelsen at styre ressourcerne effektivt og træffe hurtige beslutninger. Ved at gemme disse data på lang sigt kan man analysere fremtidige tendenser og optimere processerne.

Forskellige diagrammer og grafer bruges til at visualisere KPI'er, ROI'er og andre målinger, som normalt kombineres i dashboards.

Dashboards og dashboards: Visualisering af målinger til effektiv ledelse

Der bruges en række diagrammer og grafer til at visualisere indikatorer og målinger, som normalt kombineres til dataudstillingsvinduer og dashboards. Disse dashboards giver et centraliseret overblik over status for et projekt eller dele af et projekt og viser nøgleindikatorer (ideelt set i realtid). Up-to-date og løbende opdaterede dashboards gør det muligt for teamet at reagere hurtigt på ændringer.

> **Dashboards** er værktøjer, der visualiserer kvantitative vurderinger og gør dem let tilgængelige og forståelige for alle projektdeltagere.

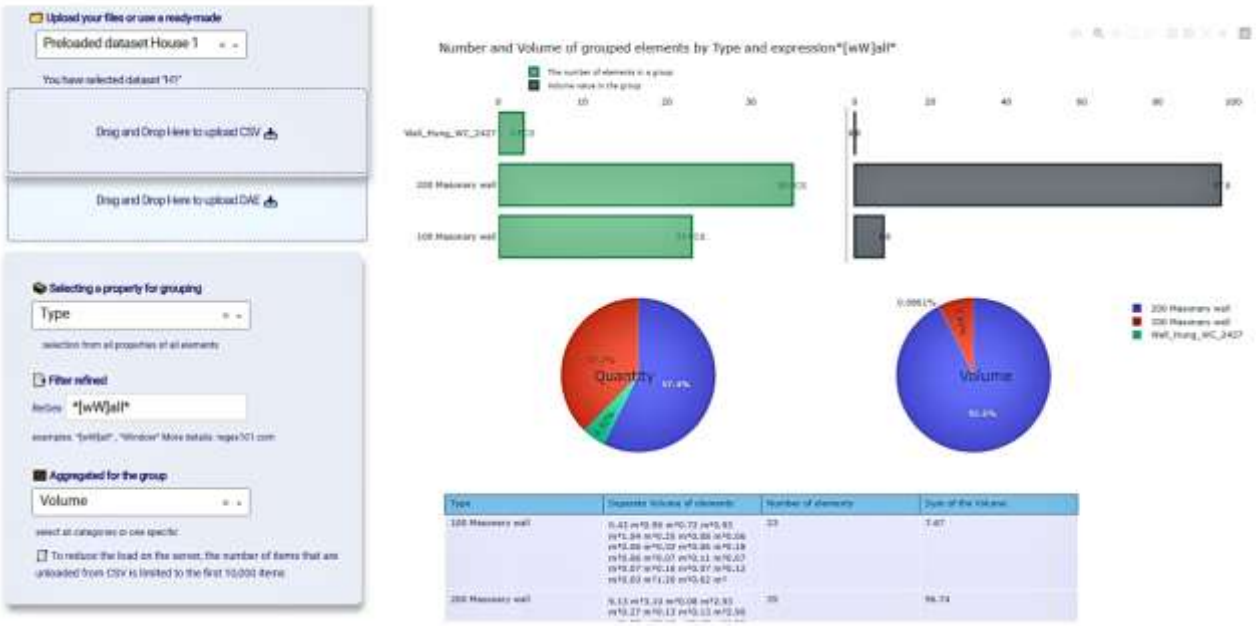


Fig. 7.1-6 Styring af KPI'er og visualisering af dem via dashboards er nøglen til at forbedre projektets produktivitet og effektivitet.

Her er nogle eksempler på populære værktøjer, som du kan bruge til at lave dashboards:

- **Power BI** er et værktøj fra Microsoft til at skabe interaktive rapporter og dashboards.
- **Tableau og Google Data Studio** er stærke værktøjer til at visualisere data og skabe dashboards uden at skulle skrive kode.
- **Plotly** (figur 7.1-6, figur 7.2-12) er et bibliotek til at skabe interaktive grafer, og Dash er en ramme til at skabe webapplikationer til analyse af data. De kan bruges i kombination til at skabe interaktive dashboards.
- **Talrige Python-biblioteker** (Figur 7.2-9 - Figur 7.2-11) - Python har mange open source og gratis biblioteker til datavisualisering, f.eks. Matplotlib, Seaborn, Plotly, Bokeh og andre. Disse kan bruges til at skabe grafer og integrere dem i en webapplikation ved hjælp af frameworks som Flask eller Django.
- **JavaScript-biblioteker**: giver dig mulighed for at oprette interaktive dashboards ved hjælp af Open Source JavaScript-biblioteker som D3.js eller Chart.js og integrere dem i en webapplikation.

> For at kunne vurdere KPI'er og oprette dashboards har du brug for opdaterede data og en klar tidsplan for indsamling og analyse af oplysninger.

Generelt danner KPI, ROI og dashboards i byggebranchen grundlaget for en analytisk tilgang til projektledelse. De hjælper ikke kun med at overvåge og evaluere den aktuelle status, men giver også værdifuld indsigt i fremtidige planlægnings- og optimeringsprocesser - processer, der afhænger direkte af at fortolke data og stille de rigtige og rettidige spørgsmål.

Dataanalyse og kunsten at stille spørgsmål

Datafortolkning er den sidste fase af analysen, hvor informationerne giver mening og begynder at "tale". Det er her, svarene på nøglespørgsmålene formuleres: *"Hvad skal jeg gøre?"* og *"Hvordan skal jeg gøre?"* (Figur 2.2-5). Denne fase gør det muligt at opsummere resultater, identificere mønstre, etablere årsags- og virkningsforhold og drage konklusioner baseret på visualisering og statistisk analyse.

> *Måske er den tid ikke langt væk, hvor man vil indse, at for at blive en effektiv borger i en af de store nye komplekse verdensstater, der nu er ved at udvikle sig, er det lige så nødvendigt at kunne regne, at tænke i gennemsnit, maksima og minima, som det nu er nødvendigt at kunne læse og skrive* [133].
>
> - Samuel S. Wilkes, citeret i en formandstale til American Statistical Association i 1951

> Ifølge rapporten "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) udgivet af den britiske regering [83] kan implementeringen af analyser data og kunstig intelligens (AI) forbedre projektstyringsprocesserne betydeligt, øge nøjagtigheden af tids- og omkostningsprognoser samt reducere risiko og usikkerhed. Artiklen fremhæver, at offentlige organisationer, der bruger avancerede analytiske værktøjer, opnår bedre resultater med infrastrukturinitiativer.

Moderne byggevirksomheder, der opererer i den fjerde industrielle revolutions stærkt konkurrenceprægede miljø med lave marginer, kan sammenlignes med militære operationer. Her afhænger virksomhedens overlevelse og succes af, hvor hurtigt den kan få fat i ressourcer og kvalitetsinformation - og dermed af rettidig og informeret beslutningstagning (Fig. 7.1-7).

Hvis datavisualisering er den "intelligens", der giver overblik, så er dataanalyse den "ammunition", der skal til for at handle. Den besvarer spørgsmålene*: Hvad skal man gøre,* og *hvordan skal man gøre det,* og danner grundlaget for at opnå en konkurrencefordel på markedet.

> Analytics omdanner forskelligartede data til struktureret og meningsfuld information, som kan danne grundlag for beslutninger.

Analytikernes og ledernes opgave er ikke bare at fortolke information, men at træffe informerede beslutninger, identificere tendenser, bestemme relationer mellem forskellige typer data og kategorisere dem i overensstemmelse med projektets mål og specifikationer. Ved hjælp af visualiseringsværktøjer og statistiske analysemetoder gør de data til et strategisk aktiv for virksomheden.

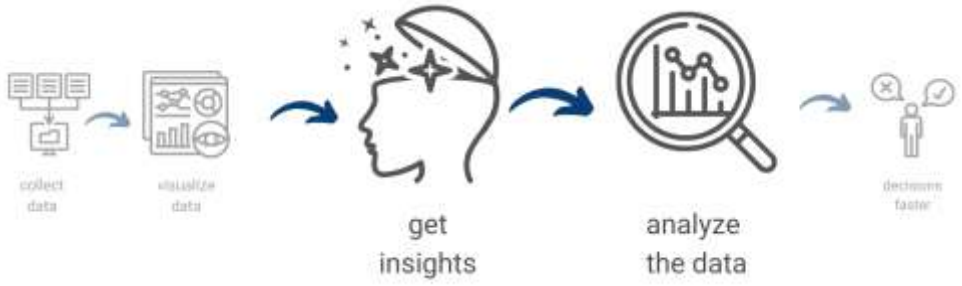


Fig. 7.1-7 Det er dataanalysen der i sidste ende gør de indsamlede oplysninger til en kilde til beslutningstagning.

For at kunne træffe virkelig informerede beslutninger i analyseprocessen er det nødvendigt at lære at formulere de spørgsmål, der stilles til dataene, korrekt. Kvaliteten af disse spørgsmål påvirker direkte dybden af den opnåede indsigt og som følge heraf kvaliteten af ledelsens beslutninger.

> *Fortiden eksisterer kun i det omfang, den er til stede i nutidens optegnelser. Og hvad disse optegnelser repræsenterer, bestemmes af de spørgsmål, vi stiller. Der findes ingen anden historie end denne [134].*
>
> - John Archibald Wheeler, fysiker 1982

Kunsten at stille dybe spørgsmål og tænke kritisk er en vigtig færdighed, når man arbejder med data. De fleste mennesker har en tendens til at stille enkle, overfladiske spørgsmål, som det ikke kræver den store indsats at besvare. Men ægte analyse begynder med meningsfulde og tankevækkende spørgsmål, der kan afdække skjulte forhold og årsagssammenhænge i information, som kan være skjult bag flere lag af ræsonnementer.

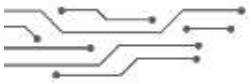
> Ifølge undersøgelsen "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135] kræver en vellykket digital transformation investering i analysekapacitet, programmer for forandringsledelse og tilpasning af forretningsmål til it-initiativer. Virksomheder, der skaber en datadrevet kultur, bør investere i dataanalysefunktioner og iværksætte programmer for forandringsledelse for at indgyde ny tænkning, adfærd og arbejdsmetoder.

Uden investeringer i udvikling af en analytisk kultur, forbedring af dataværktøjer og uddannelse af specialister vil virksomheder fortsat risikere at træffe beslutninger baseret på forældede eller ufuldstændige oplysninger - eller stole på HiPPO-chefernes subjektive meninger (figur 2.1-9).

> Erkendelsen af relevansen og behovet for konstant at opdatere analyser og dashboards får uundgåeligt ledelsen til at forstå vigtigheden af at automatisere analytiske processer. Automatisering øger beslutningshastigheden, reducerer den menneskelige faktors indflydelse og sikrer datarelevans. Med den eksponentielle vækst i informationsmængderne bliver hastighed ikke bare en konkurrencefordel, men en nøglefaktor for bæredygtig succes.

Automatisering af dataanalyse- og behandlingsprocesser generelt er uløseligt forbundet med emnet ETL

(Extract, Transform, Load). Ligesom vi i automatiseringsprocessen er nødt til at transformere data, udtrækkes data i ETL-processen fra forskellige kilder, transformeres i henhold til de nødvendige krav og indlæses i målsystemer til videre brug.



KAPITEL 7.2.
DATAFLOW UDEN MANUEL INDSATS: HVORFOR DER ER BRUG FOR ETL

ETL-automatisering: reducerer omkostninger og fremskynder datahåndtering

Når key performance indicators (KPI'er) holder op med at vokse på trods af stigende datamængder og teamstørrelse, indser virksomhedsledelsen uundgåeligt behovet for at automatisere processer. Før eller siden bliver denne erkendelse et incitament til at iværksætte kompleks automatisering, hvis hovedmål er at reducere processernes kompleksitet, fremskynde behandlingen og reducere afhængigheden af den menneskelige faktor.

> Ifølge McKinseys undersøgelse "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136] får virksomheder, der bruger streaming-dataarkitekturer, en betydelig fordel, fordi de kan analysere information i realtid. Streamingteknologier muliggør direkte analyse af realtidsbeskeder og anvendelse af prædiktiv vedligeholdelse i produktionen ved at analysere sensordata i realtid.

> Procesforenkling er automatisering, hvor traditionelle manuelle funktioner erstattes af algoritmer og systemer.

Spørgsmålet om automatisering, eller rettere "minimering af menneskets rolle i databehandling", er en irreversibel og meget følsom proces for enhver virksomhed. Specialister inden for et hvilket som helst fagområde tøver ofte med at afsløre deres arbejdsmetoder og finesser for andre optimeringsfolk, fordi de er klar over risikoen for at miste deres job i et teknologisk miljø i hastig udvikling.

> *Hvis du vil have fjender, skal du prøve at ændre tingene [137].*
>
> - Woodrow Wilson, tale til en kongres for sælgere, Detroit, 1916

På trods af de åbenlyse fordele ved automatisering har mange virksomheder stadig en høj andel af manuelt arbejde i deres daglige praksis, især inden for tekniske data. Lad os se på et typisk eksempel på sekventiel databehandling inden for sådanne processer for at illustrere den nuværende situation.

Manuel datahåndtering kan illustreres med et eksempel på interaktion med oplysninger fra CAD-databaser. Traditionel databehandling ("manuel" ETL -proces) i CAD (BIM)-afdelinger til oprettelse af attributtabeller eller oprettelse af dokumentation baseret på designdata finder sted i følgende rækkefølge (Fig. 7.2-1):

1. Manuel **udtrækning (Extract**): Brugeren åbner projektet manuelt - ved at starte CAD-programmet (BIM) (Fig. 7.2-1 trin 1).
2. **Verifikation**: Det næste trin involverer normalt manuel kørsel af flere plug-ins eller hjælpeprogrammer for at forberede dataene og vurdere deres kvalitet (Fig. 7.2-1 trin 2-3).
3. Manuel **transformation (Transform):** Efter forberedelsen begynder databehandlingen, som kræver manuel betjening af forskellige softwareværktøjer, hvor dataene forberedes til upload (Fig. 7.2-1 trin 4).

4. Manuel upload **(Load):** manuel upload af konverterede data til eksterne systemer, dataformater og dokumenter (Fig. 7.2-1 trin 5).

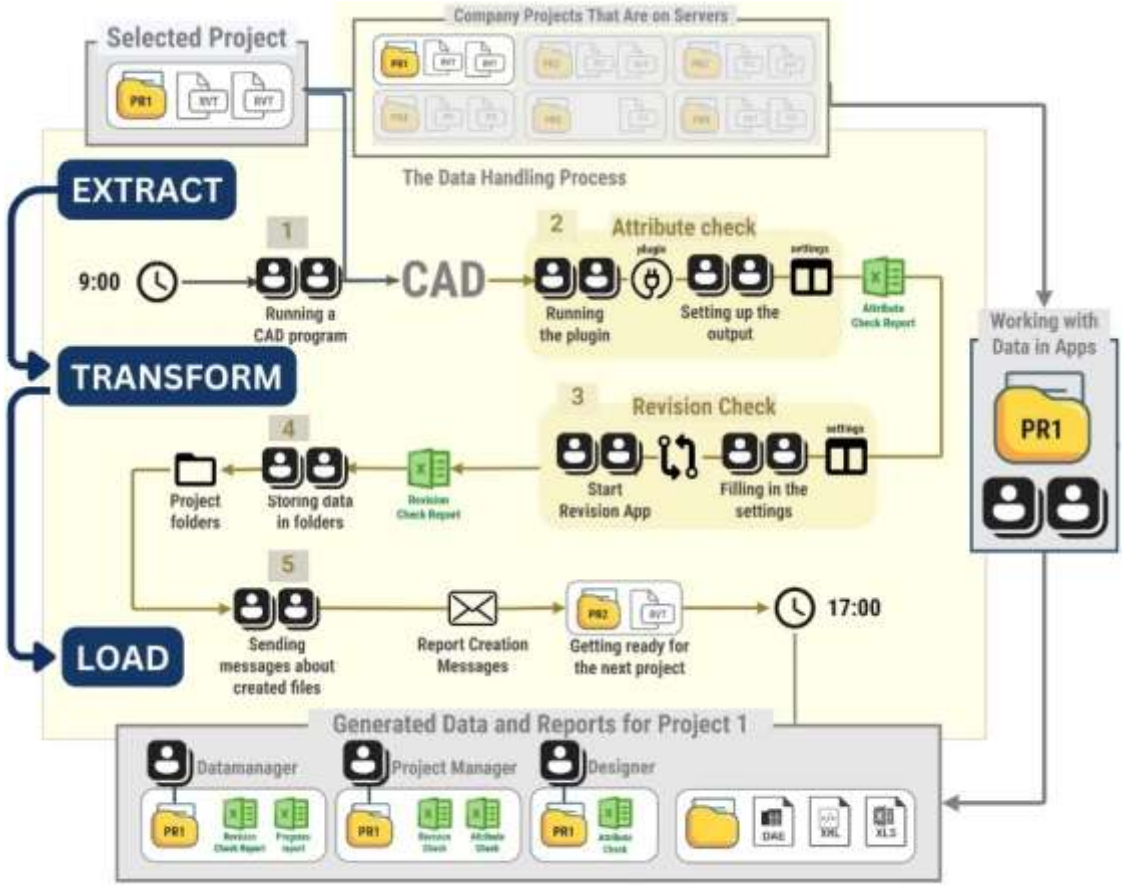


Fig. 7.2-1 Traditionel manuel ETL behandling er begrænset af den enkelte teknikers ønsker og fysiske formåen.

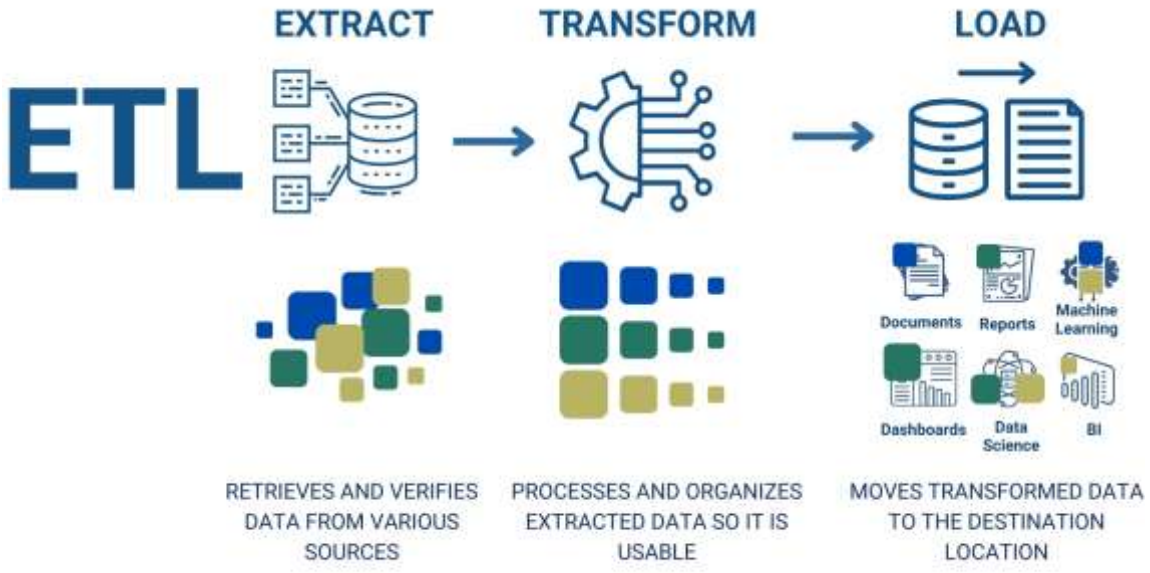
En sådan arbejdsgang er et eksempel på en klassisk ETL -proces - extraction, transformation and loading (ETL). I modsætning til andre brancher, hvor automatiske ETL-pipelines længe har været standarden, er byggebranchen stadig domineret af manuelt arbejde, hvilket forsinker processerne og øger omkostningerne

> **ETL (Extract, Transform, Load**) er processen med at udtrække data fra forskellige kilder, omdanne dem til det ønskede format og indlæse dem i målsystemet til videre analyse og brug.

ETL er en proces, der betegner tre nøglekomponenter i databehandling: Extract, Transform og Load (figur 7.2-2):

- Udtræk - udtræk data fra forskellige kilder (filer, databaser, API).
- **Transform** - datarensning, aggregering, normalisering og logisk behandling.
- **Indlæs** - indlæs struktureret information i et data warehouse, en rapport eller et BI-system.

Tidligere i bogen blev begrebet ETL kun berørt lejlighedsvis: i konverteringen af et ustruktureret indscannet dokument til et struktureret tabelformat (fig. 4.1-1), i forbindelse med formalisering af krav for at systematisere opfattelsen af både livs- og forretningsprocesser (fig. 4.4-20) og i automatiseringen af datavalidering og databehandling fra CAD-løsninger. Lad os nu se nærmere på ETL i forbindelse med typiske arbejdsgange.



Figur 7.2-2 ETL automatiserer gentagne databehandlingsopgaver.

Manuel eller halvautomatisk ETL -proces indebærer en leder eller tekniker, der håndterer alle trin manuelt - fra dataindsamling til rapportgenerering. En sådan proces tager lang tid, især når arbejdstiden er begrænset (f.eks. 9:00 til 17:00).

Virksomheder forsøger ofte at løse problemet med lav effektivitet og langsom hastighed ved at købe modulære integrerede løsninger (ERP, PMIS, CPM, CAFM, etc.), som derefter videreudvikles af eksterne leverandører og konsulenter. Men disse leverandører og tredjepartsudviklere bliver ofte et kritisk afhængighedspunkt: deres tekniske begrænsninger påvirker direkte hele systemets ydeevne og virksomheden som helhed, som beskrevet i de foregående kapitler om proprietære systemer og formater. De problemer, der skabes af fragmentering og afhængighed, blev diskuteret i detaljer i kapitlet "Hvordan byggevirksomheder drukner i data-kaos".

> Hvis en virksomhed ikke er klar til at implementere en stor modulær platform fra en af leverandørerne, begynder den at lede efter alternative måder at automatisere på. En af dem er at udvikle deres egen modulære åbne ETL -transportører, hvor hvert trin (udtrækning, transformation, validering, indlæsning) er implementeret som scripts, der udføres efter en tidsplan.

I den automatiserede version af det samme ETL-workflow (fig. 7.2-1) ser workflowet ud som en modulær kode, der starter med at behandle data og oversætte dem til en åben struktureret form. Når de strukturerede

data er modtaget, køres forskellige scripts eller moduler automatisk, på en planlagt basis, for at kontrollere ændringer, transformere og sende beskeder (Fig. 7.2-3).

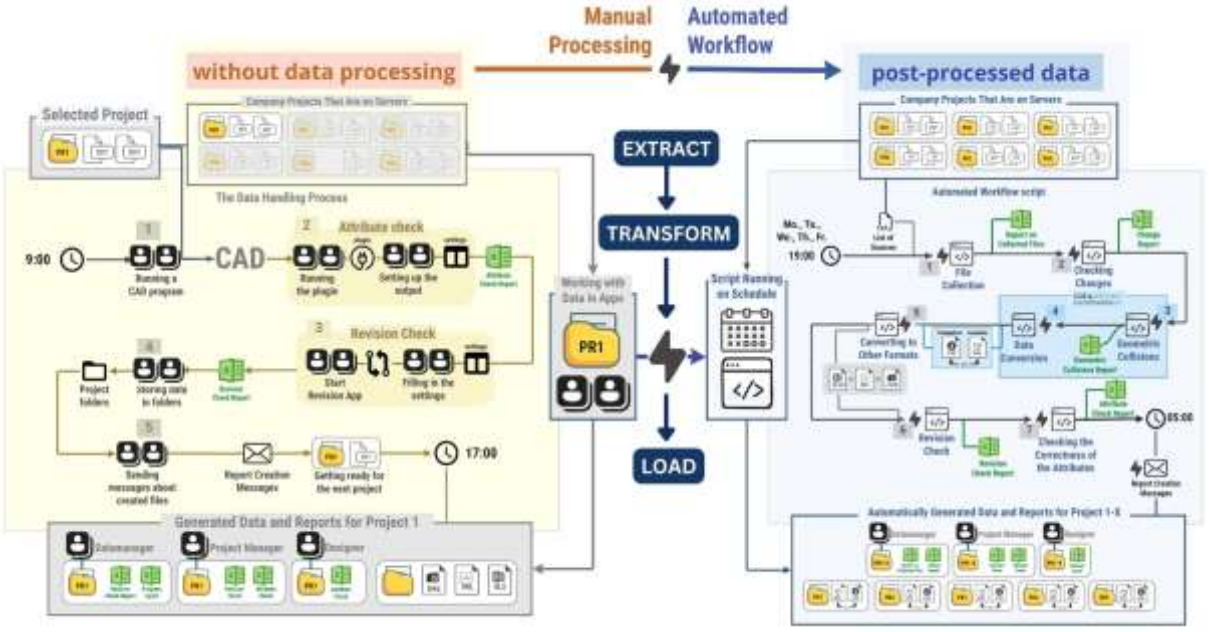


Fig. 7.2-3 Til venstre ses manuel bearbejdning, til højre ses en automatisk proces, der i modsætning til traditionel manuel bearbejdning ikke er begrænset af brugerens evner.

I en automatiseret arbejdsgang forenkles databehandlingen ved hjælp af ET(L)-forbehandling af data: strukturering og forening.

> I traditionelle behandlingsmetoder arbejder specialister med data, "som de er" - som de hentes fra systemer eller software. I automatiserede processer passerer data derimod ofte først gennem en ETL -payplane, hvor de bringes til en ensartet struktur og et format, der er egnet til videre brug og analyse.

Lad os tage et praktisk ETL-eksempel, der demonstrerer valideringsprocessen for datatabeller, som er beskrevet i kapitlet "Validering af data og valideringsresultater" (figur 4.4-13). For at gøre dette bruger vi Pandas-biblioteket sammen med LLM til automatiserede dataanalyse- og behandlingsprocesser.

ETL Udtræk: dataindsamling

Den første fase af ETL -processen - Extract) - starter med at skrive kode til indsamling af datasæt, der skal kontrolleres og behandles yderligere. For at gøre dette vil vi scanne alle mapperne på produktionsserveren, indsamle dokumenter med et bestemt format og indhold og derefter konvertere dem til en struktureret form. Denne proces beskrives i detaljer i kapitlerne "Konvertering af ustrukturerede og tekstlige data til struktureret form" og "Konvertering af CAD-data (BIM) til struktureret form" (Fig. 4.1-1 - Fig. 4.1-12).

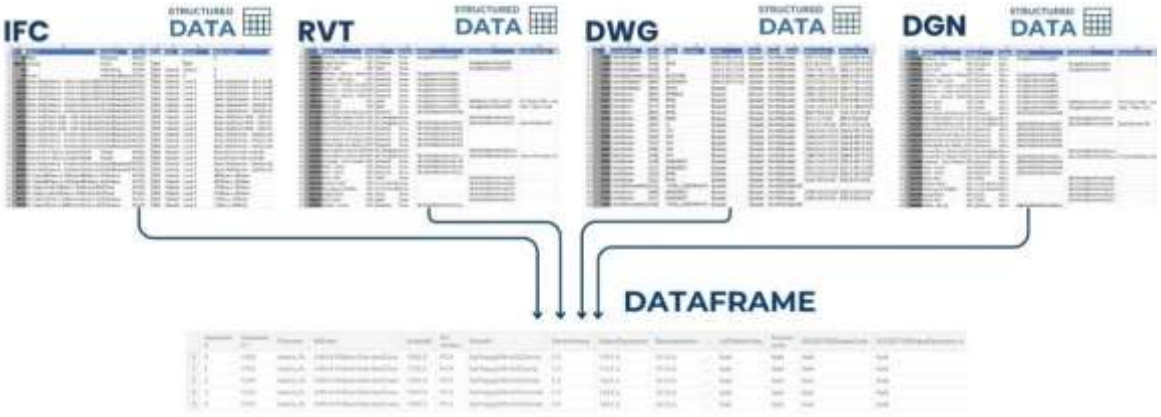


Fig. 7.2-4 Konverter CAD-data (BIM) til en stor dataramme, der indeholder alle projektafsnit.

Som et illustrativt eksempel bruger vi Extract dataindlæsningstrinnet og får en tabel over alle CAD- (BIM-) projekter (Fig. 7.2-4) bruger reverse engineering-aktiverede konvertere [138] til RVT- og IFC-formater til at få strukturerede tabeller fra alle projekter og kombinere dem til en stor DataFrame-tabel.



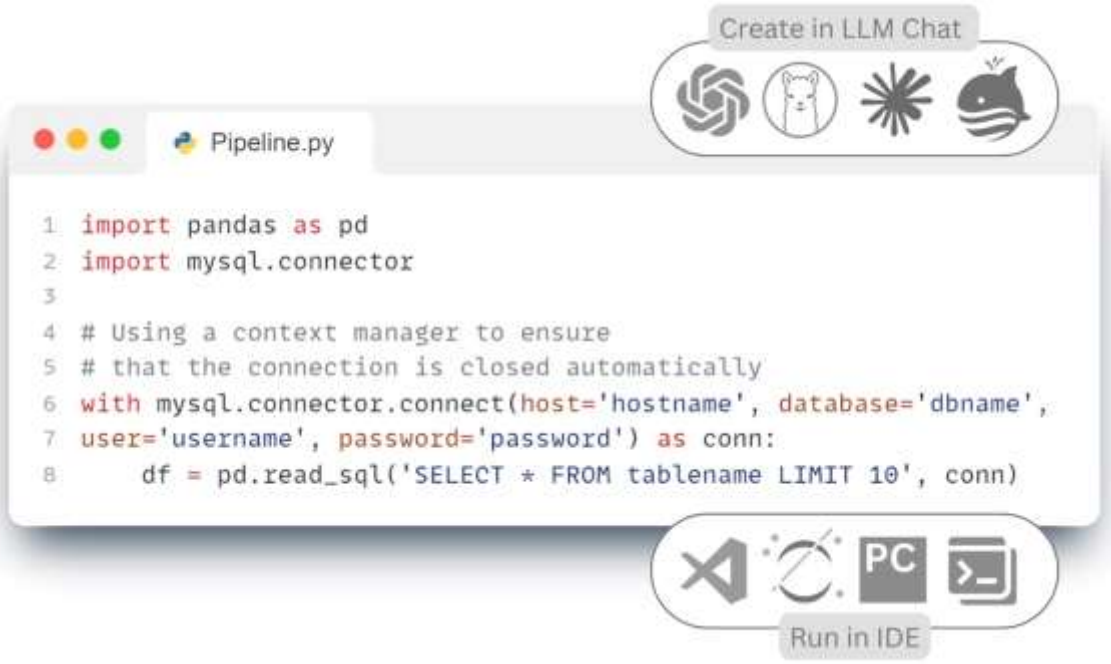
Fig. 7.2-5 Konvertering ved hjælp af Python kode og SDK reverse engineering-værktøj til RVT og IFC filer til en stor struktureret (df) DataFrame.

Pandas DataFrame kan indlæse data fra en række forskellige kilder, herunder CSV-tekstfiler, Excel-regneark, JSON - og XML - filer, big data-lagringsformater som Parquet og HDF5, og fra MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle og andre databaser. Derudover understøtter Pandas indlæsning af data fra API'er, websider, skytjenester og lagringssystemer som Google BigQuery, Amazon Redshift og Snowflake.

- Hvis du vil skrive kode til at oprette forbindelse og indsamle oplysninger fra databaser, skal du sende en lignende tekstanmodning til LLM chat (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller en anden):

> Skriv venligst et eksempel på at oprette forbindelse til MySQL og konvertere data til. ⏎

LLM's svar :



Figur 7.2-6 Eksempel på forbindelse via Python til en MySQL-database og import af data fra MySQL-databasen til en DataFrame .

Den resulterende kode (fig. 7.2-5, fig. 7.2-6) kan køres i et af de populære IDE'er (integrerede udviklingsmiljøer), vi nævnte ovenfor, i offline-tilstand: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære online-værktøjer: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Ved at indlæse multiformatdataene i variablen "df" (Figur 7.2-5 - række 25; Figur 7.2-6 - række 8) konverterede vi dataene til Pandas DataFrame -formatet, en af de mest populære strukturer til databehandling, som er en todimensionel tabel med rækker og kolonner. Vi vil tale mere om andre lagringsformater, der bruges i ETL -Pipelines som Parquet, Apache ORC, JSON, Feather, HDF5, samt moderne data warehouses i kapitlet "Datalagring og -styring i byggebranchen" (figur 8.1-2).

Efter dataudtræk og -strukturering (Udtræk) er der dannet en enkelt informationsrække (Fig. 7.2-5, Fig. 7.2-6), som er klar til videre behandling. Men før disse data indlæses i målsystemer eller bruges til analyse, er det nødvendigt at sikre deres kvalitet, integritet og overholdelse af de specificerede krav. Det er i denne fase, at datatransformationen (Transform) finder sted - et vigtigt skridt for at sikre pålideligheden af de efterfølgende konklusioner og beslutninger.

ETL Transform: anvendelse af validerings- og transformationsregler

Trinnet Transform er det trin, hvor dataene behandles og transformeres. Denne proces kan omfatte kontrol af korrekthed, normalisering, udfyldning af manglende værdier og validering hjælp af automatiserede

værktøjer.

> Ifølge PwC-undersøgelsen "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] bevæger moderne revisionsfirmaer sig væk fra tilfældig dataverifikation og hen imod analyse af massive mængder information ved hjælp af automatiserede værktøjer. Denne tilgang gør det ikke kun muligt at identificere uoverensstemmelser i rapporteringen, men også at komme med anbefalinger til optimering af forretningsprocesser.

I byggeriet kan lignende metoder f.eks. bruges til automatisk validering af designdata, kvalitetskontrol af byggeriet og evaluering af entreprenørernes præstationer. Et værktøj til at automatisere og fremskynde databehandling er brugen af regulære udtryk (RegEx) i datatransformationsfasen (Transform) i ETL-processen. RegEx giver dig mulighed for effektivt at validere datastrenge, opdage uoverensstemmelser og sikre informationens integritet med minimalt ressourceforbrug. Flere detaljer om RegEx (fig. 4.4-7) blev diskuteret i kapitlet "Oversættelse af krav til struktureret form".

Lad os se på et praktisk eksempel: I systemet til forvaltning af ejendomsobjekter (RPM) opstiller lederen kravene til nøgleattributter for objekter (fig. 7.2-7). I transformationsfasen skal følgende parametre valideres:

- Verifikation af formater for objektidentifikatorer (attribut "ID")
- Kontrol af værdier for erstatningsgarantiperiode (attribut "Garantiperiode")
- verifikation af elementets udskiftningscyklus (attribut "Vedligeholdelseskrav")



Figur 7.2-7 Verifikation af kvalitet begynder med at opstille krav til attributter og deres grænseværdier.

For at fastsætte grænseværdier for parametervalidering kan vi f.eks. antage, at vi fra ved, at gyldige værdier for "ID"-attributten kun kan omfatte strengværdierne "W-NEW", "W-OLD1" eller "D-122" eller lignende værdier, hvor det første tegn er et bogstav efterfulgt af en bindestreg og derefter tre alfabetiske tegn 'NEW', 'OLD' eller et trecifret tal (figur 7.2-7). Følgende regulære udtryk (RegEx) kan bruges til at validere disse identifikatorer:

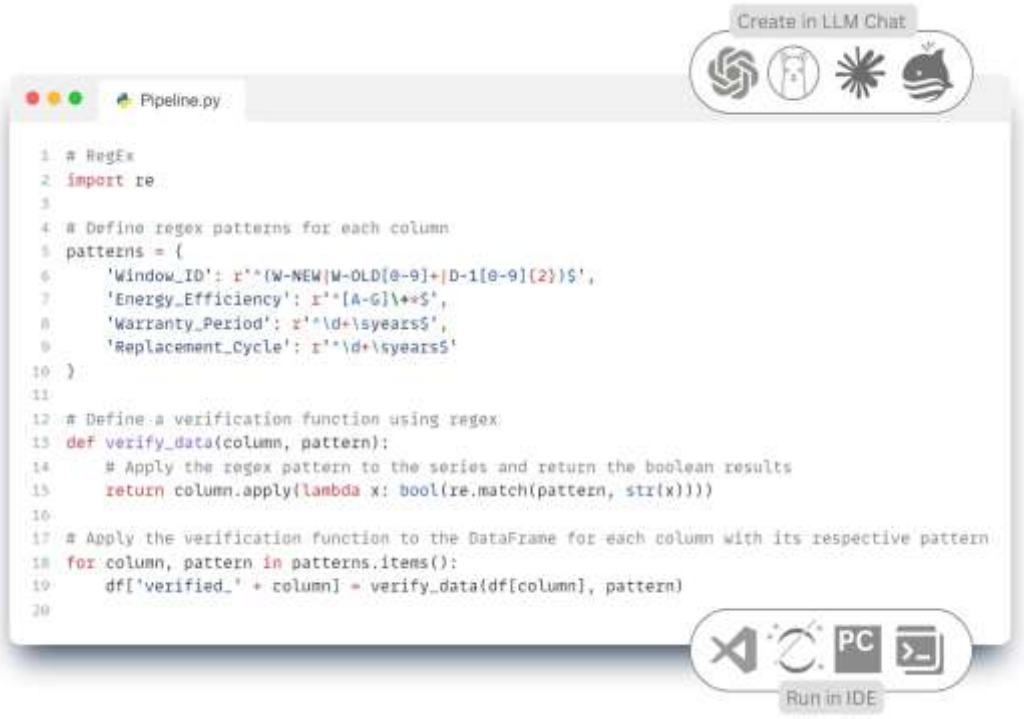
```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

Denne skabelon sørger for, at alle identifikatorer i dataene matcher de angivne kriterier. Hvis en værdi ikke består kontrollen, logger systemet en fejl. For at oprette Python -kode til at transformere dataene og bruge de resulterende data til at oprette en resultattabel, skal du blot formulere en forespørgsel i LLM -chatten.

- Sms-anmodning til LLM:

> Skriv kode til at validere kolonnerne i DataFrame ved hjælp af regulære udtryk, som kontrollerer identifikatorer i formatet 'W-NEW' eller 'W-OLD' via RegEx, energieffektivitet med bogstaverne 'A' til 'G', garantiperiode og udskiftningscyklus med numeriske værdier i år. ⏎

- LLM's svar :



Figur 7.2-8 Koden automatiserer valideringsprocessen ved at anvende RegEx-skabeloner på dataframens parameterkolonner .

Ovenstående automatisk genererede Python -kode (figur 7.2-8) bruger biblioteket "re" (RegEx regular expressions) til at definere en funktion, der kontrollerer hver attribut for et dataelement i en DataFrame. For hver kolonne (attribut), der er angivet, anvender funktionen et RegEx-mønster til at kontrollere, at hver post er i overensstemmelse med det forventede format, og tilføjer resultaterne som nye værdier (False/True) til den nye attribut-kolonne i DataFrame.

> En sådan automatiseret validering sikrer formel dataoverensstemmelse og kan bruges som en del af et kvalitetskontrolsystem under transformationsfasen.

Når trinnet Transform og kvalitetskontrollen er gennemført, er dataene klar til at blive uploadet til målsystemerne. Transformerede og validerede data kan uploades til CSV, JSON, Excel, databaser og andre formater til videre brug. Afhængigt af opgaven kan resultaterne også præsenteres i rapporter, grafer eller analytiske dashboards.

ETL Load: Visualiser resultater i diagrammer og grafer

Når Transform -fasen er afsluttet, og dataene er blevet struktureret og verificeret, er den sidste fase Load, hvor dataene både kan indlæses i -målsystemet og visualiseres til analyse. Visuel præsentation af data gør det muligt hurtigt at identificere afvigelser, analysere fordelinger og kommunikere vigtige konklusioner til alle projektdeltagere, også dem uden teknisk baggrund.

I stedet for at præsentere information som tabeller og tal kan vi bruge infografik, grafer og dashboards (instrumentbrætter). Et af de mest almindelige og fleksible værktøjer til at visualisere strukturerede data i Python er Matplotlib-biblioteket (figur 7.2-9, figur 7.2-10). Det giver dig mulighed for at skabe statiske, animerede og interaktive diagrammer og understøtter en bred vifte af diagramtyper.

- For at visualisere resultaterne af attributkontrol fra RPM-systemet (figur 7.2-7) kan du bruge følgende forespørgsel til sprogmodellen:

> Skriv kode til at visualisere DataFrame-dataene, ovenfor (figur 7.2-7), med et histogram for resultaterne for at vise hyppigheden af fejl i attributten ⏎

- LLM-svar i kodeform og klar visualisering direkte i LLM-chatrummet af resultaterne af kodeafviklingen :

Create in LLM Chat

Pipeline.py

```python
# Re-importing necessary libraries for visualization
import matplotlib.pyplot as plt
import seaborn as sns

# Visualization 1: Bar Chart
plt.figure(figsize=(10, 6))
df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
plt.title('Data Verification Summary - Bar Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Run in IDE



Fig. 7.2-9 Visualisering af resultaterne af trinnet Transform med kontrol af attributværdier fra RPM-systemet (Fig. 7.2-7) som et histogram i trinnet Load .

Der findes mange open source og gratis visualiseringsbiblioteker, som giver dig mulighed for at præsentere strukturerede data i forskellige formater. Lad os fortsætte med at visualisere resultaterne med en anden type graf med følgende prompt i chatten:

> Plot de samme data som en graf med linjer ⏎

LLM's svar :

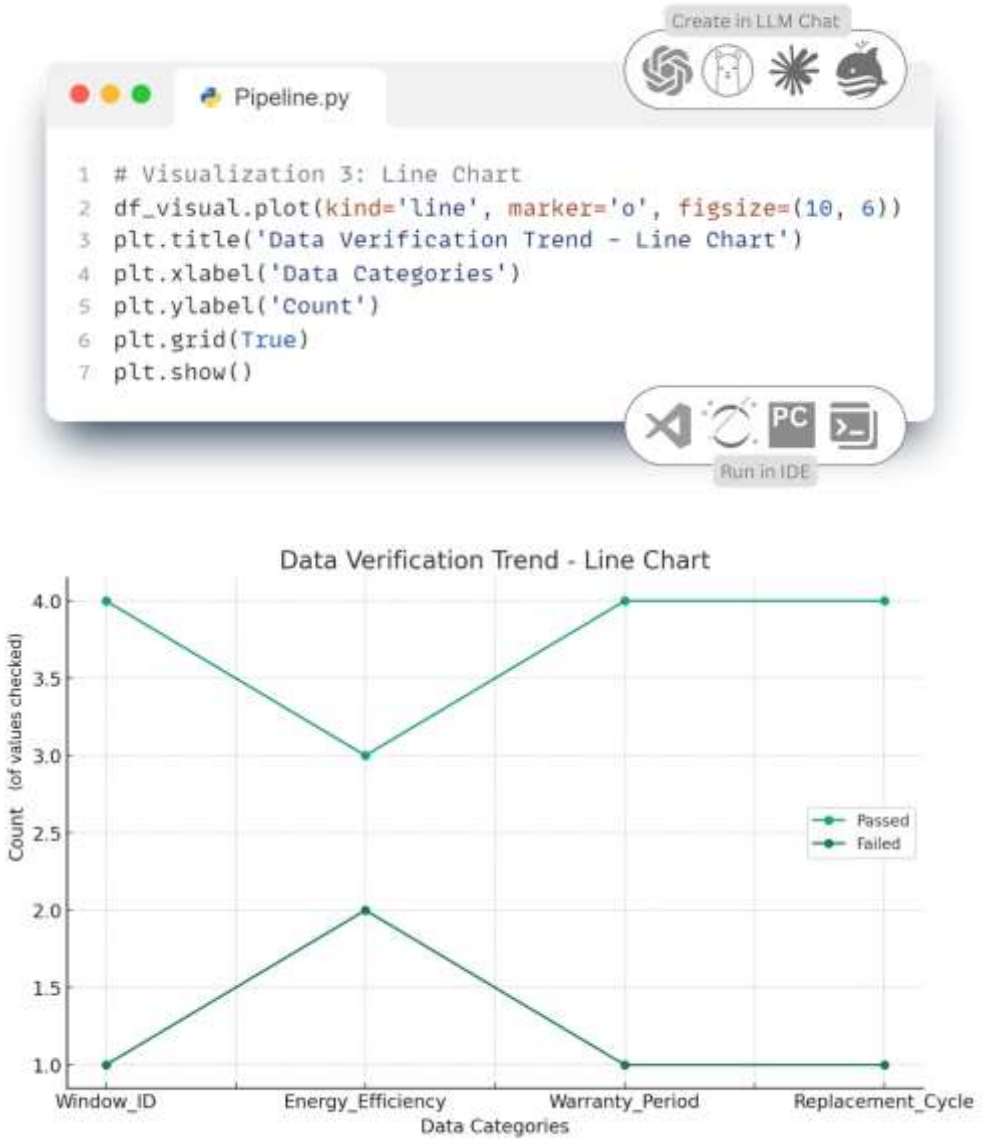


Fig. 7.2-10 Visualisering af valideringsdataene (Fig. 7.2-8) som et linjediagram opnået ved hjælp af Matplotlib-biblioteket .

Der findes mange open source og gratis visualiseringsbiblioteker som f.eks:

- Seaborn - til statistiske grafer (Fig. 7.2-11)
- Plotly - til interaktive webvisualiseringer (Fig. 7.2-12, Fig. 7.1-6)
- Altair - til deklarativ visualisering
- Dash eller Streamlit - for at skabe komplette dashboards

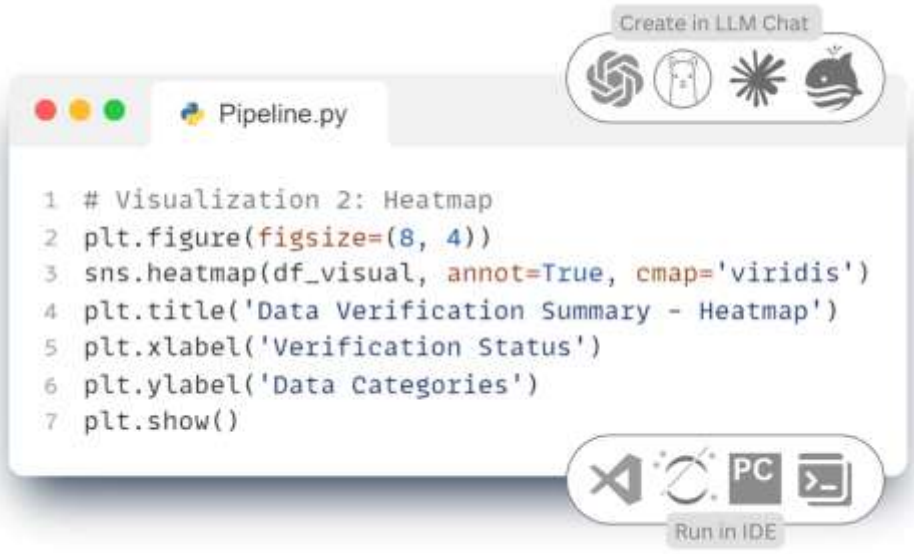
Kendskab til specifikke biblioteker til visualisering er ikke nødvendigt - moderne værktøjer, herunder LLM, giver dig mulighed for automatisk at generere kode til grafer og hele applikationer baseret på opgavebeskrivelsen.

Valget af værktøj afhænger af projektets mål: om det er en rapport, en præsentation eller et online dashboard. For eksempel er open source-biblioteket Seaborn særligt godt til at arbejde med kategoriske data og hjælpe med at identificere mønstre og tendenser.

- For at se Seaborn-biblioteket i drift kan du enten bede LLM om at bruge det ønskede bibliotek direkte eller sende en lignende tekstforespørgsel til LLM på:

> Vis heatmap for ⏎

- LLM's svar i form af kode og en færdig graf, hvis plottekode nu kan kopieres ind i IDE, og selve grafen kan kopieres eller gemmes for at blive indsat i et dokument:





Figur 7.2-11 Visualisering af resultaterne af validering (Figur 7.2-8) af data ved hjælp af Seaborn-biblioteket.

For dem, der foretrækker en interaktiv tilgang, findes der værktøjer, som gør det muligt at skabe dynamiske diagrammer og paneler med mulighed for interaktion. Plotly-biblioteket (figur 7.1-6, figur 7.2-12) giver mulighed for at skabe meget interaktive diagrammer og paneler, der kan indlejres i websider og give brugeren mulighed for at interagere med dataene i realtid.

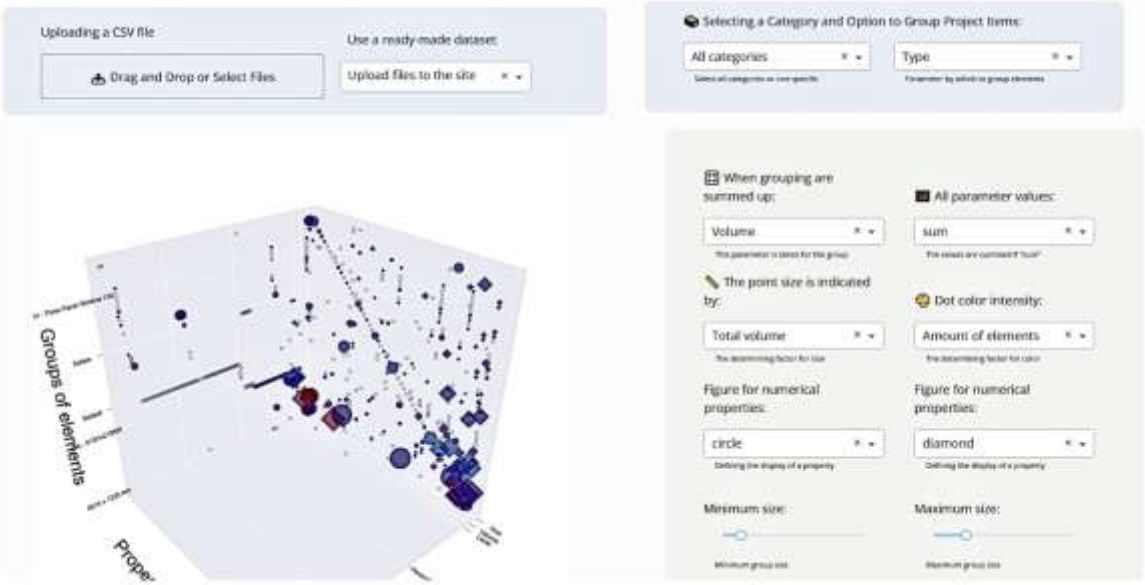


Fig. 7.2-12 Interaktiv 3D -visualisering af elementattributter fra et CAD- (BIM-) projekt ved hjælp af Plotly-biblioteket.

De specialiserede open source-biblioteker Bokeh, Dash og Streamlit giver en praktisk måde at præsentere data på uden at kræve dyb viden om webudvikling. Bokeh er velegnet til komplekse interaktive grafer, Dash bruges til at bygge fuldgyldige analytiske dashboards, og Streamlit giver dig mulighed for hurtigt at oprette webapplikationer til dataanalyse.

Med sådanne visualiseringsværktøjer kan udviklere og analytikere effektivt formidle resultater til kolleger og interessenter, hvilket muliggør intuitiv interaktion med data og forenkler beslutningstagningen.

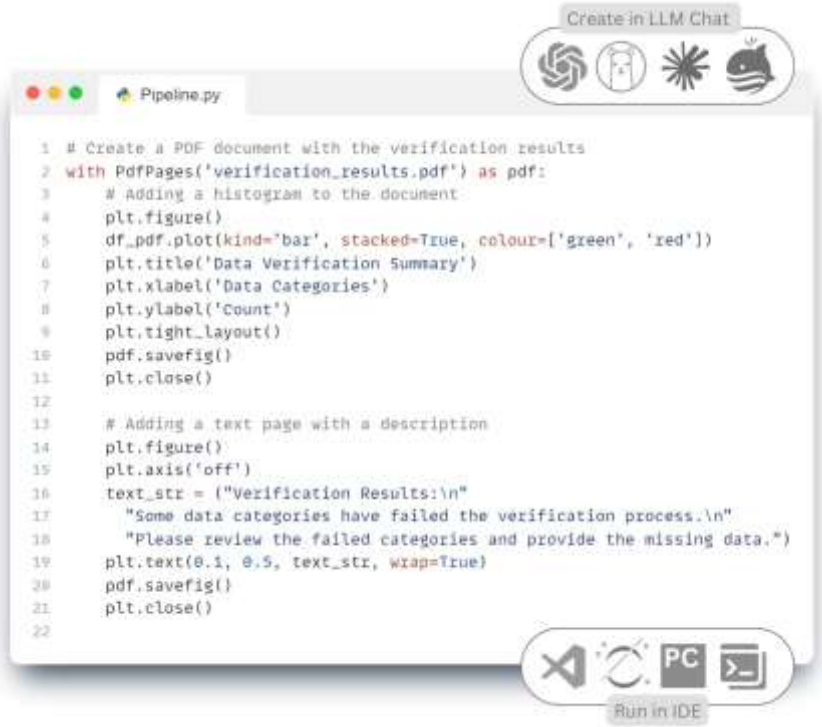
ETL Load: Automatisk oprettelse af PDF-dokumenter

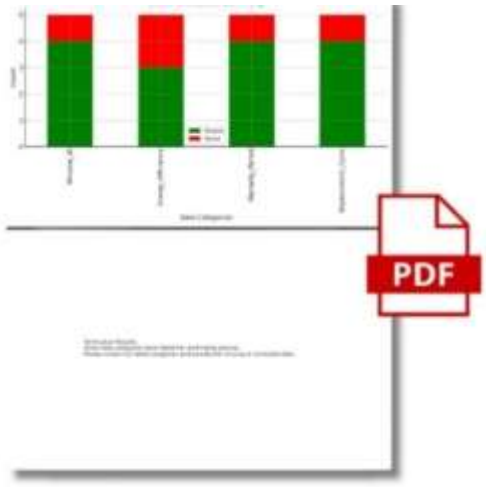
I dataindlæsningsfasen kan du ikke kun visualisere data, uploade dem til tabeller eller databaser, men også automatisk generere rapporter, herunder de nødvendige grafer, diagrammer og analytiske nøgleindikatorer, som skal modtages af den leder eller specialist, der afventer resultaterne af revisionen. Automatiserede rapporter kan indeholde både kommentarer og tekstfortolkning af data og visualiseringselementer - tabeller og grafer.

- For at oprette en PDF -rapport med et histogram (Fig. 7.2-9) og en beskrivelse af analysen baseret på den validering, vi udførte i de foregående kapitler, er det tilstrækkeligt at formulere en anmodning i en dialogfortsættelse med LLM, f.eks.:

> Skriv kode til at oprette en PDF -fil med et histogram og en beskrivelse af resultaterne af datavalideringen ovenfor (i chatten), og skriv en tekstadvarsel om, at nogle kategorier ikke er blevet valideret, og at manglende skal udfyldes. ⏎

- LLM's svar som kode og færdig PDF med resultater:





Figur 7.2-13 Den automatiserede kode opretter et PDF -dokument, der indeholder et histogram med testdata og en tekst med testresultater.

En automatisk skrevet løsning på kun 20 kodelinjer ved hjælp af LLM skaber øjeblikkeligt det ønskede PDF (eller DOC) dokument med visualisering i form af et attributhistogram (fig. 7.2-13), der viser antallet af data, der bestod og ikke bestod valideringen, og med tilføjelse af en tekstblok, der opsummerer resultaterne og anbefalinger yderligere handling.

Automatiseret dokumentgenerering er et nøgleelement i Load stage, især i et projektmiljø, hvor hurtig rapportering og nøjagtighed er afgørende.

ETL Load: automatisk generering af dokumenter fra FPDF

Automatisering af rapportering på ETL-stadiet Load er et vigtigt skridt i databehandlingen, især når resultaterne af analyserne skal præsenteres i et format, der er let at kommunikere og forstå. I byggebranchen er dette ofte relevant for statusrapporter, projektdatastatistikker, kvalitetssikringsrapporter eller finansiel dokumentation.

Et af de mest praktiske værktøjer til sådanne opgaver er open source-biblioteket, FPDF, der findes til både Python og PHP.

> Open source **FPDF-biblioteket** giver en fleksibel måde at generere dokumenter på ved hjælp af kode, så man kan tilføje overskrifter, tekst, tabeller og billeder. Brug af kode i stedet for manuel redigering reducerer fejl og fremskynder processen med at udarbejde rapporter i PDF format.

Et af de vigtigste trin i oprettelsen af et PDF -dokument er at tilføje overskrifter og hovedtekst i form af kommentarer eller beskrivelser. Men når man opretter en rapport, er det vigtigt ikke kun at tilføje tekst, men også at strukturere den korrekt. Overskrifter, indrykninger, linjeafstand - alt dette påvirker dokumentets læsbarhed. Med FPDF kan du indstille formateringsparametre, styre placeringen af elementer og tilpasse dokumentets stil.

FPDF er i princippet meget lig HTML. De, der allerede er fortrolige med HTML, kan nemt generere PDF dokumenter af enhver kompleksitet ved hjælp af FPDF, da kodestrukturen er meget lig HTML-markup:

overskrifter, tekst, billeder og tabeller tilføjes på samme måde. De, der ikke er fortrolige med HTML, skal ikke bekymre sig - de kan bruge LLM, som straks hjælper dem med at komponere koden til at generere det ønskede dokumentlayout.

Følgende eksempel viser, hvordan man genererer en rapport med en overskrift og brødtekst. Ved at udføre denne kode i en IDE med Python-understøttelse oprettes en PDF -fil, der indeholder den ønskede overskrift og tekst:

```
from fpdf import FPDF     # Importer FPDF-biblioteket
pdf = FPDF()   # Opret PDF -document
pdf.add_page()     # Tilføj en side

pdf.set_font("Arial", style='B', size=16) # Indstil skrifttype: Arial, fed, størrelse 16
pdf.cell(200, 10, "Projektrapport", ln=True, align='C')  # Opret en overskrift og centrer den
pdf.set_font("Arial", size=12) # Skift skrifttype til almindelig Arial, størrelse 12
pdf.multi_cell(0, 10, "Dette dokument indeholder data om resultaterne af verifikation af projektfiler
...")  # Tilføj tekst på flere linjer
pdf.output(r "C:\reports\report.pdf") # Gem PDF -file
```



Figur 7.2-14 Med et par linjer Python-kode kan vi automatisk generere det PDF tekstdokument, vi har brug for.

Når man udarbejder rapporter, er det vigtigt at tage højde for, at de data, som dokumentet dannes ud fra, sjældent er statiske. Overskrifter, tekstblokke (fig. 7.2-14) dannes ofte dynamisk og modtager værdier i transformationsfasen i ETL-processen.

> Ved hjælp af koden kan du oprette dokumenter, der indeholder opdaterede oplysninger: projektnavn, dato for rapportgenerering samt oplysninger om deltagere eller aktuel status. Brugen af variabler i koden gør det muligt automatisk at indsætte disse data på de ønskede steder i rapporten, hvilket helt eliminerer behovet for manuel redigering inden afsendelse.

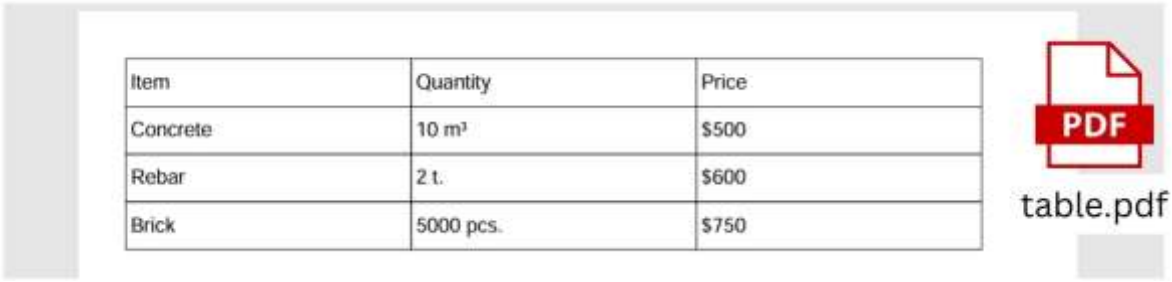
Ud over simpel tekst og overskrifter indtager tabeller en særlig plads i projektdokumentation. Næsten alle dokumenter indeholder strukturerede data: fra objektbeskrivelser til inspektionsresultater. Automatisk generering af tabeller baseret på data fra Transformgør det ikke kun muligt at fremskynde processen med at udarbejde dokumenter, men også at minimere fejl ved overførsel af information. FPDF gør det muligt at indsætte tabeller i PDF -filer (som tekst eller billeder) og indstille cellekanter, kolonnestørrelser og skrifttyper (fig. 7.2-15). Det er især praktisk, når man arbejder med dynamiske data, hvor antallet af rækker og kolonner kan variere afhængigt af dokumentets opgaver.

- Følgende eksempel viser, hvordan man kan automatisere oprettelsen af tabeller, f.eks. med en liste over materialer, estimater eller parametertestresultater:

```
data = [
    ["Vare", "Mængde", "Pris"],   # Kolonneoverskrifter
    ["Beton", "10 m³", "$ 500."],   # Data for første række
    ["Armeringsjern", "2 tons", "$ 600"],   # Data for anden række.
    ["Mursten", "5.000 stykker", "$ 750."],   # Data for linje 3.
]

pdf = FPDF () # Opret PDF -document
pdf.add_page() # Tilføj en side
pdf.set_font("Arial", size=12)   # Indstil skrifttypen

for row in data:   # Søg i tabellens rækker
    for item in row: # Gå gennem cellerne i rækken
        pdf.cell(60, 10, item, border=1)   # Opret en celle med en kant, bredde 60 og højde 10
    pdf.ln()   # Gå til næste linje
pdf.output(r "C:\reports\table.pdf")   # Gem PDF -file
```



Figur 7.2-15 Du kan automatisk generere ikke kun tekst, men også alle tabeloplysninger fra transformeringstrinnet i PDF.

I virkelige rapporteringsscenarier er tabeller normalt dynamisk genererede oplysninger, der opnås i datatransformationsfasen. I det givne eksempel (Fig. 7.2-15) er tabellen indsat i PDF -dokumentet i en statisk form: dataene til eksemplet blev placeret i dataordbogen (første linje i koden), under virkelige forhold udfyldes en sådan datavariabel automatisk efter f.eks. dataramme-gruppering.

I praksis opbygges sådanne tabeller ofte på grundlag af strukturerede data, der kommer fra forskellige dynamiske kilder: databaser, Excel -filer, API -grænseflader eller resultater af analytiske beregninger. Oftest i Transform (ETL)-fasen aggregeres, grupperes eller filtreres data - og først derefter omdannes de til totaler i form af grafer eller todimensionelle tabeller, der vises i rapporter. Det betyder, at tabelindholdet kan ændre sig afhængigt af de valgte parametre, analyseperiode, projektfiltre eller brugerindstillinger.

> Brugen af dynamiske datarammer og datasæt i Transform -fasen gør rapporteringsprocessen i Load -fasen så fleksibel, skalerbar og let at gentage som muligt uden behov for manuel indgriben.

Ud over tabeller og tekst understøtter FPDF også tilføjelse af grafer til tabeldata, hvilket giver dig mulighed for at integrere billeder genereret med Matplotlib eller andre visualiseringsbiblioteker, som vi har diskuteret ovenfor, i rapporten. Alle grafer, diagrammer og skemaer kan føjes til dokumentet ved hjælp af koden.

- Lad os ved hjælp af Python-biblioteket FPDF tilføje en graf, der er genereret på forhånd med Matplotlib, til PDF-dokumentet:

```
import matplotlib.pyplot as plt  # Importer Matplotlibat skabe grafer

fig, ax =  plt.subplots()  # Opret grafens figur og akser
categories = ["Beton", "Armeringsjern", "Mursten"]  # Kategorinavne
values = [50000, 60000, 75000]  # Kategoriens værdier
ax.bar(categories, values)  # Opret et søjlediagram
plt.ylabel("Værdi,$.")  # Tegn Y-aksen
plt.title("Omkostningsfordeling")  # Tilføj en titel
plt.savefig(r "C:\reports\chart\chart\chart.png") # Gem diagrammet som et billede

pdf = FPDF () # Opret PDF -document
pdf.add_page()  # Tilføj en side
pdf.set_font("Arial", size=12)  # Indstil skrifttypen
pdf.cell(200, 10, "Cost Chart", ln=True, align='C')  # Tilføj en overskrift

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100)  # Indsæt billedet
i PDF'en (x, y - koordinater, w - bredde)
pdf.output(r "C:\reports\chart_report.pdf")  # Gem PDF-filen
```

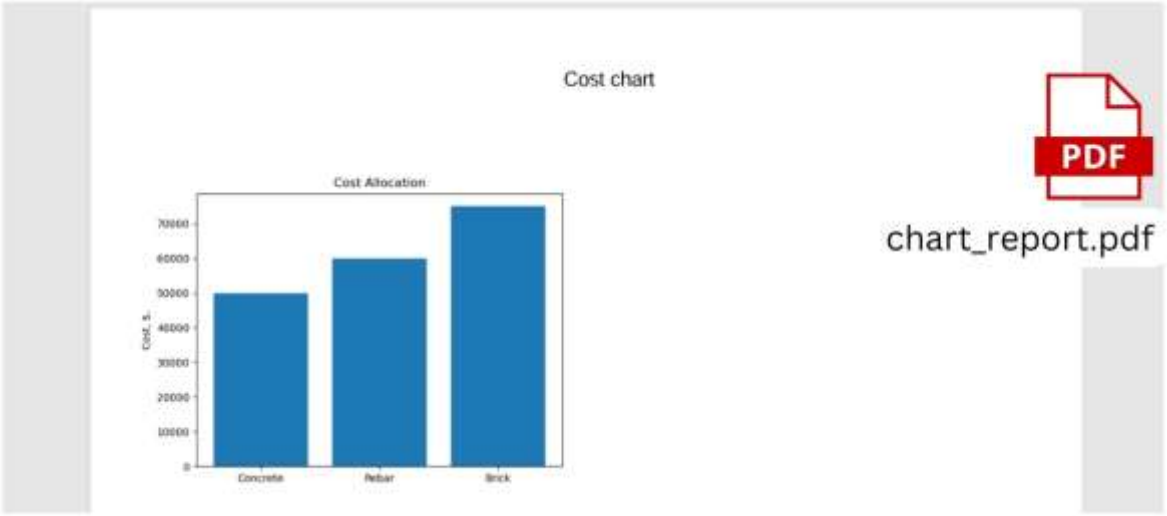


Fig. 7.2-16et dusin kodelinjer kan du generere en graf, gemme den og derefter indsætte den i et PDF-dokument.

Med FPDF bliver processen med dokumentforberedelse og -logik gennemsigtig, hurtig og bekvem. Skabeloner indbygget i koden gør det muligt at generere dokumenter med opdaterede data, hvilket eliminerer behovet for manuel udfyldning.

> Brug af ETL-automatisering - i stedet for tidskrævende manuel rapportering kan fagfolk fokusere på at analysere data og træffe beslutninger i stedet for at vælge det rigtige værktøj til at arbejde med en bestemt datasilo med en klar brugergrænseflade.

FPDF-biblioteket er således et fleksibelt værktøj til automatisk oprettelse af dokumenter af enhver kompleksitet - fra korte tekniske rapporter til komplekse analytiske resuméer med tabeller og diagrammer, hvilket ikke kun gør det muligt at fremskynde dokumentflowet, men også at reducere sandsynligheden for fejl i forbindelse med manuel dataindtastning og formatering betydeligt.

ETL Load: Rapportering og indlæsning til andre systemer

I Load-fasen blev resultaterne genereret i form af tabeller, grafer og endelige PDF-rapporter udarbejdet i overensstemmelse med de fastlagte krav. Det er desuden muligt at eksportere disse data til maskinlæsbare formater (f.eks. CSV), hvilket er nødvendigt for integration med eksterne systemer som ERP, CAFM, CPM, BI-platforme og andre virksomheds- eller brancheløsninger. Ud over CSV kan der uploades til XLSX, JSON, XML eller direkte til databaser, der understøtter automatisk informationsudveksling.

- For at generere den passende kode til at automatisere indlæsningstrinnet skal du blot forespørge LLM -grænsefladen, f.eks: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude eller QWEN:

> Skriv kode til at generere en rapport over datavalideringsresultater i DataFrame, hvor kolonner med præfikset 'verified_' tælles, omdøbes til 'Passed' og 'Failed', manglende værdier erstattes med 0, og derefter eksporteres kun de rækker, der består alle valideringer, til en CSV -fil.

- LLM's svar:



Figur 7.2-17 Validerede data opnået i Transform-trinnet fra den endelige dataramme eksporteres til en CSV -fil til integration med andre systemer.

Ovenstående kode (fig. 7.2-17) implementerer det sidste trin i ETL -processen - Load, hvor de verificerede data gemmes i CSV-format, der er kompatibelt med de fleste eksterne systemer og databaser. Dermed har vi gennemført hele cyklussen i ETL -processen, herunder udtræk, transformation, visualisering, dokumentation og eksport af data til de systemer og formater, vi har brug for, hvilket sikrer reproducerbarhed, gennemsigtighed og automatisering af arbejdet med information.

ETL pipeline kan bruges både til behandling af enkeltprojekter og til store applikationer - når man analyserer hundredvis eller tusindvis af indgående data i form af dokumenter, billeder, scanninger, CAD projekter, punktskyer, PDF filer eller andre kilder, der kommer fra distribuerede systemer. Muligheden for at automatisere processen fuldt ud gør ETL til ikke bare et teknisk behandlingsværktøj, men grundlaget for en digital infrastruktur for byggeinformation.

ETL med LLM: Visualisering af data fra PDF -dokumenter

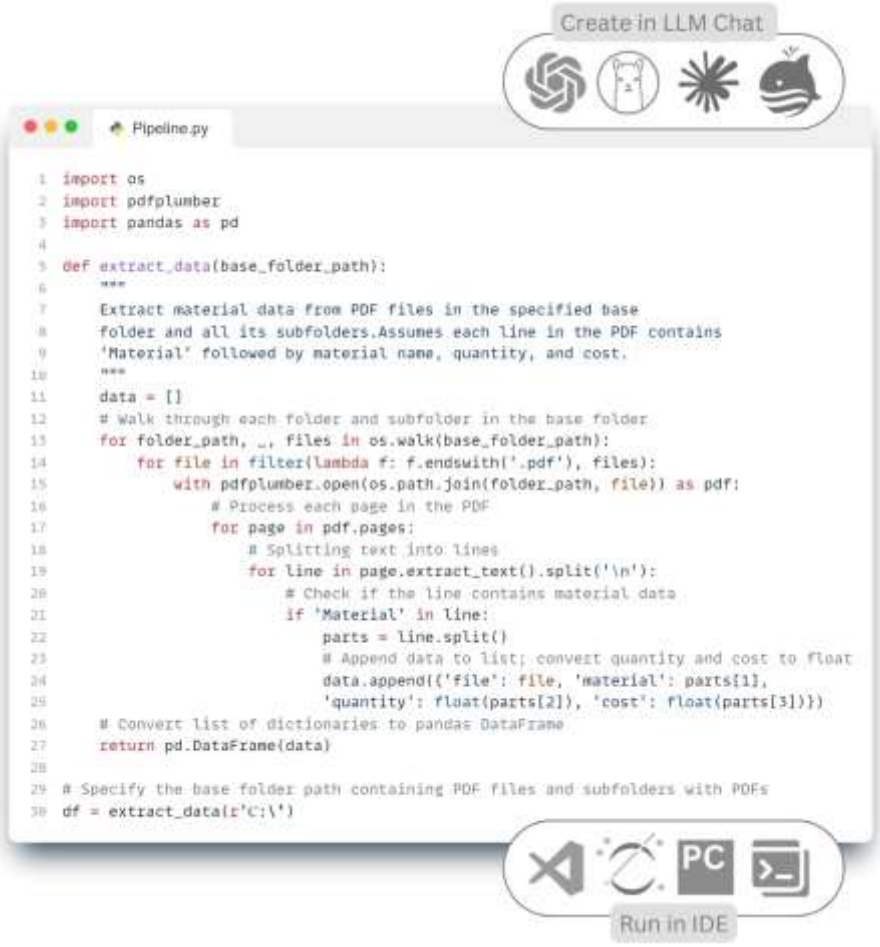
Det er på tide at gå videre til at opbygge en fuldgyldig ETL-proces, der dækker alle de vigtigste faser af datahåndtering i et enkelt scenarie - udtræk, transformation og indlæsning. Lad os opbygge en automatiseret ETL-pipeline, der giver dig mulighed for at behandle PDF-dokumenter uden manuelt arbejde - udtrække data fra dokumenter, visualisere, analysere og overføre til andre systemer.

ETL-processen i vores eksempel vil blive beskrevet gennem prompts, som skal forklare sprogmodellen (LLM) alle ETL-processerne med en beskrivelse af det slutresultat, der skal opnås. I dette tilfælde er opgaven at finde alle PDF-filer i den angivne mappe og dens undermapper, udtrække relevante oplysninger fra dem - f.eks. materialenavne, mængder og omkostninger - og præsentere resultatet som en struktureret tabel (DataFrame) til yderligere analyse

- Første tekstforespørgsel i LLM for automatisk at udtrække data fra flere PDF-dokumenter og oprette en dataramme med data til udtrækstrinnet:

> Skriv kode til at udtrække materialeoplysninger fra PDF -filer i en given mappe og dens undermapper. Dataene i PDF'en omfatter materialenavn, mængde og pris. Resultatet skal gemmes i en DataFrame ⏎

LLM's svar :

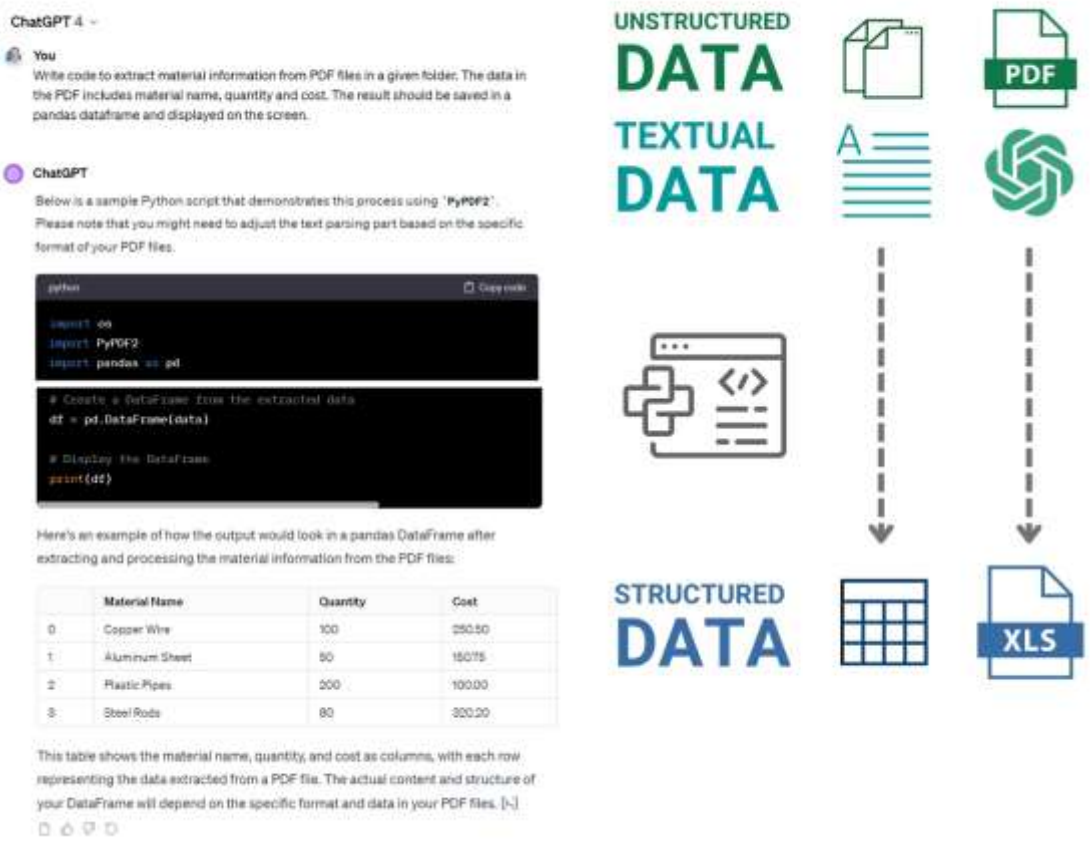


Figur 7.2-18 LLM skaber Python -kode til at udtrække data fra PDF -filer i en bestemt mappe og alle dens undermapper.

LLM-svaret (figur 7.2-18) er et færdigt Python -script, der automatisk gennemsøger alle mapper, åbner de fundne PDF -filer, udtrækker tekstinformation fra dem og konverterer den til en tabel. Den resulterende chatkode kan køres i en af de populære IDE'er PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære onlineværktøjer Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

For at forenkle indsamlingsprocessen kan vi i stedet for at kopiere koden fra LLM og bruge koden i IDE'en også uploade et dusin PDF -filer direkte til LLM-chatten (figur 7.2-19) og få en tabel som output uden at skulle se koden eller køre den. Resultatet af at udføre denne kode vil være en tabel med de attributter, vi har

valgt.

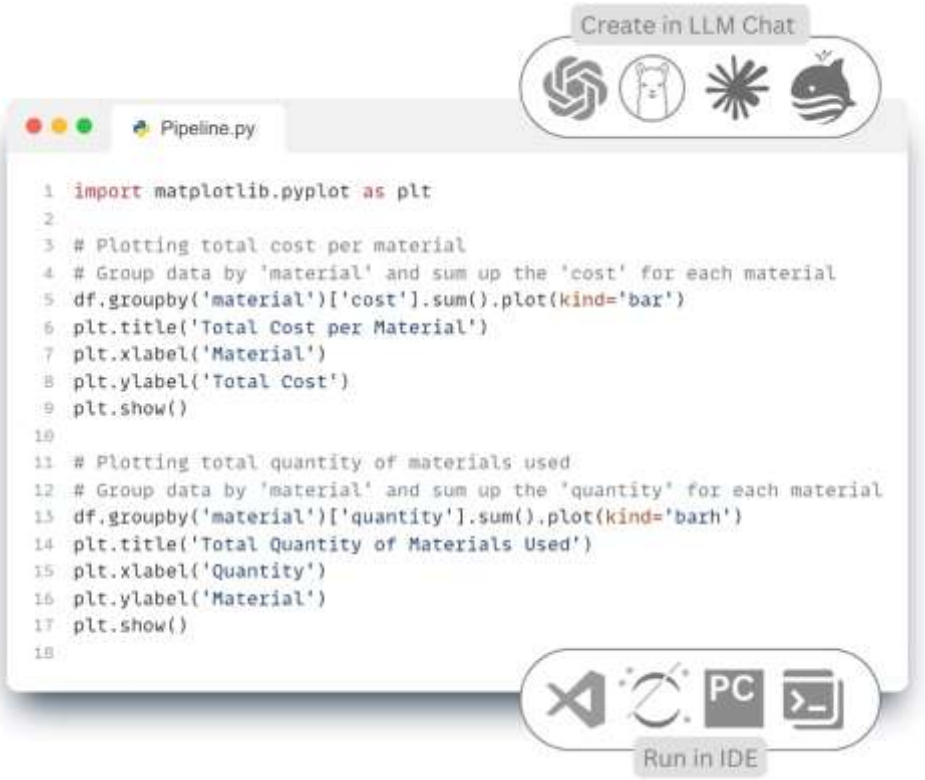


Figur 7.2-19 Resultatet af at udføre kode i LLM, som udtrækker data fra PDF -filer i en struktureret dataframevisning med udvalgte attributter.

I næste trin beder vi om en sprogmodel på de indhentede data - for eksempel for at sammenligne omkostninger og mængde af materialeforbrug og skabe nogle eksempler på visualiseringer, der kan tjene som grundlag for yderligere analyse.

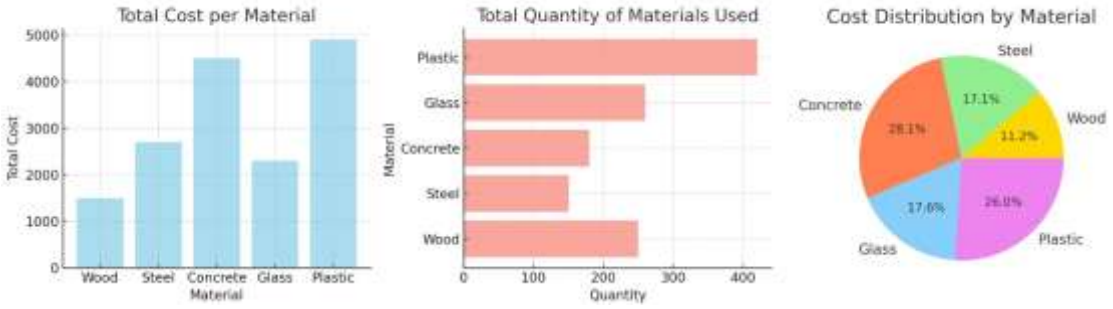
- Bed i en fortsat chat med LLM om at plotte nogle grafer fra de tabeller, der blev produceret i Transform-trinnet (figur 7.2-18):

> Visualiser de samlede omkostninger og mængden af hvert materiale fra DataFrame (figur 7.2-18)⏎



Figur 7.2-20 LLM-respons -modeller som Python-kode for at visualisere data fra en dataramme ved hjælp af matplotlib-biblioteket.

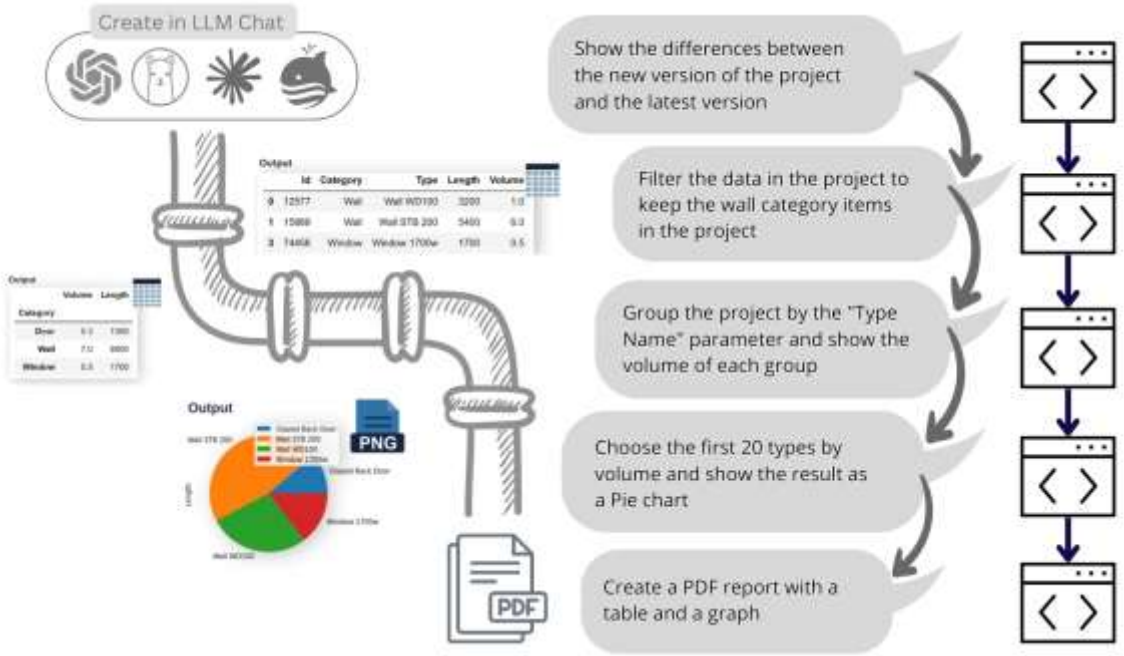
LLM genererer og udfører automatisk Python -kode (fig. 7.2-20) ved hjælp af matplotlib-biblioteket. Efter udførelse af denne kode får vi grafer over omkostninger og materialeforbrug i byggeprojekter direkte i chat (fig. 7.2-21), hvilket i høj grad forenkler det analytiske arbejde.



Figur 7.2-21 Visualisering af LLM-svaret som grafer baseret på data indsamlet i DataFrame.

Støtte til at udvikle ideer til at skrive ETL kode, analysere og udføre kode og visualisere resultater er tilgængelig gennem enkle tekstforespørgsler i LLM, uden at det er nødvendigt at lære det grundlæggende i programmering. Fremkomsten af AI -værktøjer som LLM ændrer helt klart tilgangen til programmering og automatisering af databehandling (figur 7.2-22).

> Ifølge PwC-rapporten "Hvad er den reelle værdi af kunstig intelligens for din virksomhed, og hvordan kan du udnytte den?" (2017) [139] vil procesautomatisering og produktivitetsforbedringer være de vigtigste drivkræfter for økonomisk vækst. Og produktivitetsforbedringer forventes at stå for mere end 55% af al AI-drevet BNP-vækst mellem 2017 og 2030."



Figur 7.2-22 AI LLM hjælper med at generere udkast til kode, der kan anvendes i fremtidige projekter uden behov for en LLM.

> Ved hjælp af værktøjer som ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, samt åbne data og open source-software kan vi automatisere processer, der tidligere kun blev udført med specialiserede, dyre og vanskeligt vedligeholdelige modulopbyggede proprietære systemer.

I byggesammenhæng betyder det, at virksomheder, der er de første til at implementere automatiserede Pipeline -dataprocesser, vil høste betydelige fordele, fra forbedret projektstyringseffektivitet til reducerede økonomiske tab og eliminering af fragmenterede applikationer og siloopdelte datalagre.

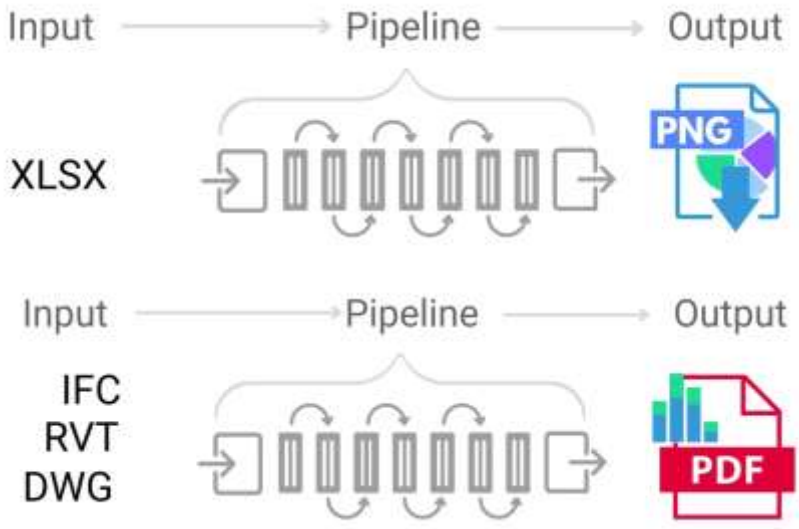
Den beskrevne logik for udførelse af forretningsopgaver i ETL-processen er en afgørende del af automatiseringen af analytics og databehandlingsprocesser, som er en specifik variation af et bredere koncept - pipelines (rørledninger).

KAPITEL 7.3.
AUTOMATISK ETL TRANSPORTØR (RØRLEDNING)

Pipeline: Automatisk ETL transportør data

ETL-processen er traditionelt blevet brugt til at behandle data i analytiske systemer, der dækker både strukturerede og ustrukturerede kilder. Men i dagens digitale miljø bruges der i stigende grad et bredere begreb - Pipeline (transportbånd), som beskriver enhver sekventiel behandlingskæde, hvor output fra et trin bliver input til det næste.

Denne tilgang gælder ikke kun for data, men også for andre typer af automatisering: opgavebehandling, bygningsrapportering, integration med software og digital arbejdsgang (Fig. 7.3-1).



Figur 7.3-1 Pipeline er en behandlingssekvens, hvor outputtet fra et trin bliver input til det næste trin.

Brugen af Pipeline er et af hovedelementerne i automatisering, især når man arbejder med store mængder heterogene data. Pipeline-arkitekturen gør det muligt at organisere komplekse behandlingstrin i et modulært, konsistent og håndterbart format, hvilket øger læsbarheden, forenkler vedligeholdelse af koden og muliggør trinvis fejlsøgning og skalerbar testning.

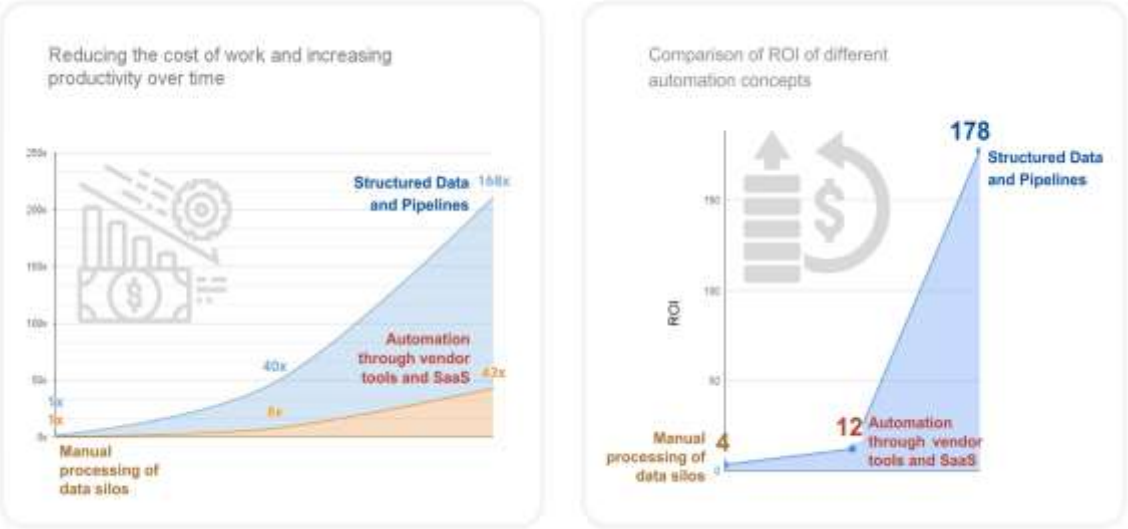


Fig. 7.3-2 ROI Pipeline datavalideringsprocessen reducerer udførelsestiden med titusinder og hundredvis af gange sammenlignet med behandling med klassiske værktøjer [74].

I modsætning til manuelt arbejde i proprietære systemer (ERP, PMIS, CAD osv.) giver pipelining dig mulighed for at øge hastigheden på opgaverne betydeligt (fig. 7.3-2), undgå gentaget arbejde og automatisere starten af processer på det rigtige tidspunkt (fig. 7.3-3).

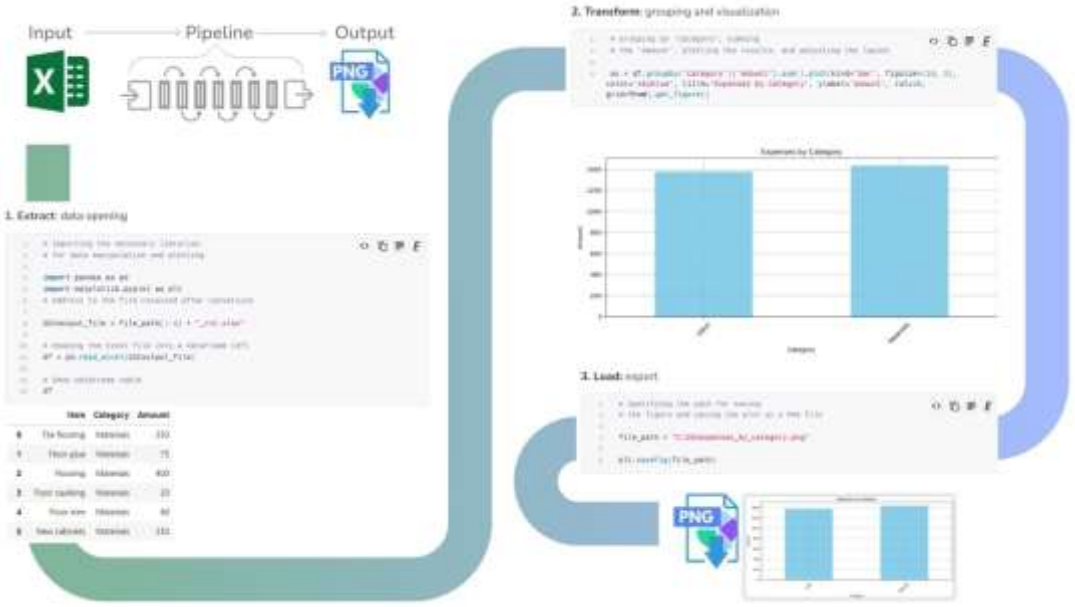


Fig. 7.3-3 ETL-eksempel Pipeline på automatisk hentning af et diagram fra tabeldata i en XLSX-fil uden at åbne Excel.

For at behandle streamingdata og opbygge en automatiseret pipeline, der ligner ETL-processen, skal du identificere datakilderne på forhånd samt tidsrammen for indsamling af dem - enten for en bestemt forretningsproces eller på tværs af hele virksomheden.

I byggeprojekter kommer data fra mange heterogene kilder med forskellige opdateringsintervaller. For at skabe et pålideligt dataudstillingsvindue er det vigtigt at registrere, hvornår oplysninger hentes og opdateres. Det muliggør rettidig beslutningstagning og forbedrer effektiviteten i projektledelsen.

En mulighed er at starte samlingsprocessen på et fast tidspunkt - f.eks. kl. 19.00, når arbejdsdagen er slut. På dette tidspunkt aktiveres det første script, der er ansvarligt for at samle data fra forskellige systemer og lagre (fig. 7.3-4 trin 1). Derefter følger automatisk behandling og omdannelse af data til et struktureret format, der er egnet til analyse (fig. 7.3-4 trin 2-4). I den sidste fase genereres rapporter, dashboards og andre produkter, der er beskrevet i de foregående kapitler, automatisk ved hjælp af de forberedte data (fig. 7.3-4 trin 6-7). Resultatet er, at lederne allerede kl. 05.00 om morgenen har opdaterede projektstatusrapporter i det ønskede format (fig. 7.3-5).

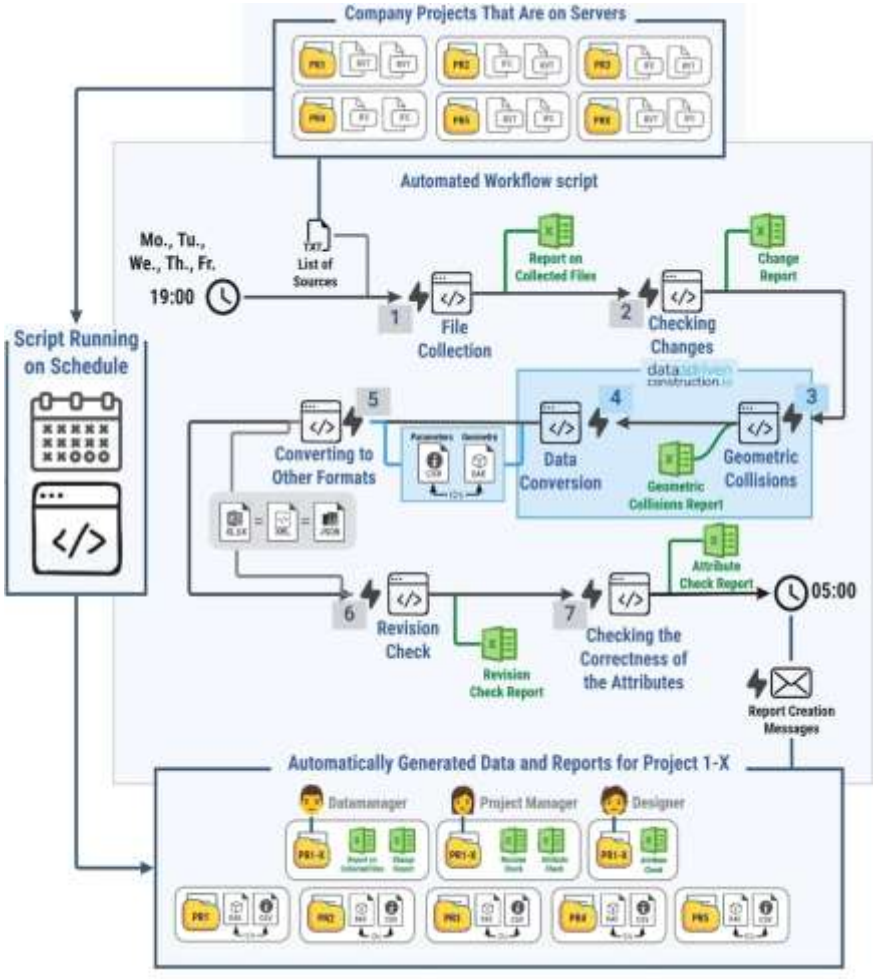


Fig. 7.3-4 Data i Pipeline, der automatisk indsamles om aftenen, behandles i løbet af natten, så lederne har opdaterede rapporter og friske rapporter om morgenen.

> Rettidig dataindsamling, KPI-definition, automatisering af transformationsprocesser og visualisering gennem dashboards er nøgleelementer i en vellykket datadrevet beslutningstagning.

Sådanne automatiserede processer (fig. 7.3-4) kan udføres helt uafhængigt: De kører efter en tidsplan, behandler data uden opsyn og kan implementeres enten i skyen eller på virksomhedens egen server (fig. 7.3-5). Det gør det muligt at integrere sådanne ETL-pipelines i den eksisterende IT-infrastruktur, bevare kontrollen over data og give fleksibilitet i skaleringen.

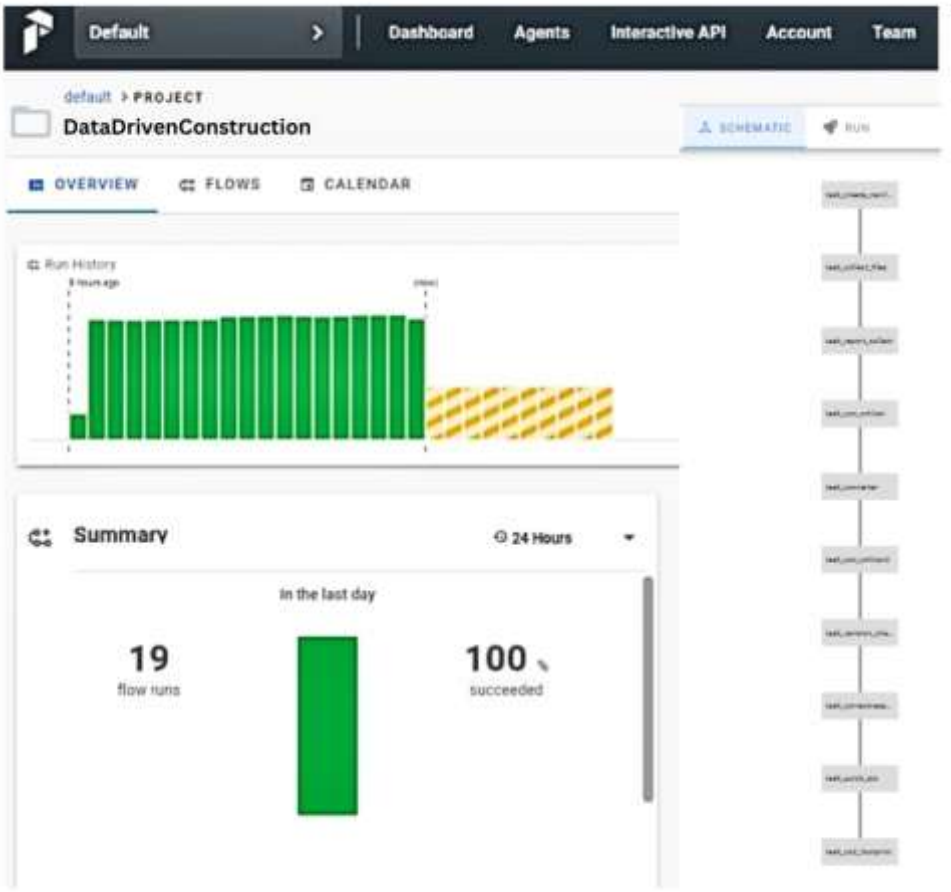


Fig. 7.3-5 Automatiske ETL -conveyor processer (Fig. 7.3-4) på Prefect-platformen, hvor 10 python-scripts køres skiftevis efter kl. 19.00 hver arbejdsdag.

Automatisering af arbejdsgange øger ikke kun teamets produktivitet ved at frigøre tid til mere meningsfulde og mindre rutineprægede opgaver, men fungerer også som et vigtigt første skridt i retning af at indarbejde teknologier med kunstig intelligens (AI) i forretningsprocesser, som vi vil diskutere mere detaljeret i kapitlet om Predictive Analytics og Machine Learning.

Rørledning -ETL datavalideringsproces med LLM

I de tidligere kapitler om oprettelse af datakrav og automatisering af ETL opdelte vi trin for trin processen med dataforberedelse, -transformation, -validering og -visualisering. Disse aktiviteter blev implementeret som separate kodeblokke (Figur 7.2-18 - Figur 7.2-20), der hver især udførte en specifik opgave.

Nu har vi det næste mål - at kombinere disse elementer til en enkelt, sammenhængende og automatiseret pipeline for databehandling - en pipeline, ETL -Pipeline - hvor alle faser (indlæsning, validering, visualisering, eksport) udføres sekventielt i et enkelt automatisk eksekverbart script.

I det følgende eksempel implementeres en komplet databehandlingscyklus: fra indlæsning af CSV-kildefilen →til kontrol af strukturen og værdierne ved hjælp af regulære udtryk →beregning af resultaterne→ generering af en visuel rapport i PDF-format.

- Du kan bruge følgende tekstforespørgsel til LLM. til at hente den relevante kode:

> Skriv et kodeeksempel, der indlæser data fra CSV, validerer DataFrame ved hjælp af regulære udtryk, kontrollerer identifikatorer i formatet 'W-NEW' eller 'W-OLD', energieffektivitet med bogstaverne 'A' til 'G', garantiperiode og udskiftningscyklus med numeriske værdier i år og til sidst opretter en rapport med en optælling af godkendte og mislykkede værdier, genererer en PDF med et histogram af resultater og tilføjer en tekstbeskrivelse. ⏎

LLM's svar :



Fig. 7.3-6 Pipeline (ETL) automatiserer hele databehandlingscyklussen: fra download og validering til oprettelse af en struktureret rapport i PDF-format .

Den automatiserede kode (fig. 7.3-6) i LLM-chatrummet eller i DIE vil, efter at have kopieret koden, validere dataene fra CSV -filen ved hjælp af de angivne regulære udtryk, oprette en rapport om antallet af godkendte og mislykkede poster og derefter gemme valideringsresultaterne som en PDF -fil.

Denne ETL -transportørstruktur, hvor hvert trin - fra dataindlæsning til rapportgenerering - er implementeret som et separat modul, sikrer gennemsigtighed, skalerbarhed og reproducerbarhed. Præsentation af valideringslogikken som en letlæselig Python-kode gør processen gennemsigtig og forståelig, ikke kun for udviklere, men også for specialister inden for datastyring, kvalitet og analyse.

> Pipelines -tilgang til automatisering af databehandling giver dig mulighed for at standardisere processer, øge deres repeterbarhed og forenkle tilpasningen til nye projekter. Det skaber en ensartet metode til at analysere data, uanset kilden eller opgavetypen - uanset om det er compliance-test, rapportering eller dataoverførsel til eksterne systemer.

En sådan automatisering reducerer menneskelige fejl, reducerer afhængigheden af proprietære løsninger og øger nøjagtigheden og pålideligheden af resultaterne, hvilket gør dem velegnede til både operationelle analyser på projektniveau og strategiske analyser på virksomhedsniveau.

Rørledning -ETL: verifikation af data og oplysninger om projektelementer i CAD (BIM)

Data fra CAD-systemer og databaser (BIM) er nogle af de mest sofistikerede og dynamisk opdaterede datakilder i byggevirksomheder. Disse applikationer beskriver ikke kun projektet ved hjælp af geometri, men supplerer det også med tekstinformation i flere lag: mængder, materialeegenskaber, rumtildelinger, energieffektivitetsniveauer, tolerancer, forventet levetid og andre attributter.

> Attributter, der tildeles enheder i CAD -modeller, dannes i designfasen og bliver grundlaget for yderligere forretningsprocesser, herunder omkostningsberegning, planlægning, livscyklusvurdering og integration med ERP og CAFM -systemer, hvor processernes effektivitet i høj grad afhænger af kvaliteten af de data, der kommer fra designafdelingerne.

Den traditionelle tilgang til attributvalidering i CAD- (BIM-) modeller involverer manuel validering (Fig. 7.2-1), hvilket bliver en lang og kostbar proces, når mængden af modeller er stor. I betragtning af mængden og antallet af moderne byggeprojekter og deres regelmæssige opdateringer bliver processen med datavalidering og -transformation uholdbar og uoverkommelig.

> Hovedentreprenører og projektledere står over for behovet for at behandle store mængder projektdata, herunder flere versioner og fragmenter af de samme modeller. Dataene kommer fra designorganisationer i RVT-, DWG-, DGN-, IFC-, NWD- og andre formater (figur 3.1-14) og kræver regelmæssig kontrol af, om de er i overensstemmelse med industri- og virksomhedsstandarder

Afhængigheden af manuelle handlinger og specialiseret software gør datavalideringsprocessen til en flaskehals i arbejdsgange, der er relateret til data fra virksomhedsdækkende modeller. Automatisering og

brug af strukturerede krav kan eliminere denne afhængighed og øge hastigheden og pålideligheden af datavalidering dramatisk (figur 7.3-7).

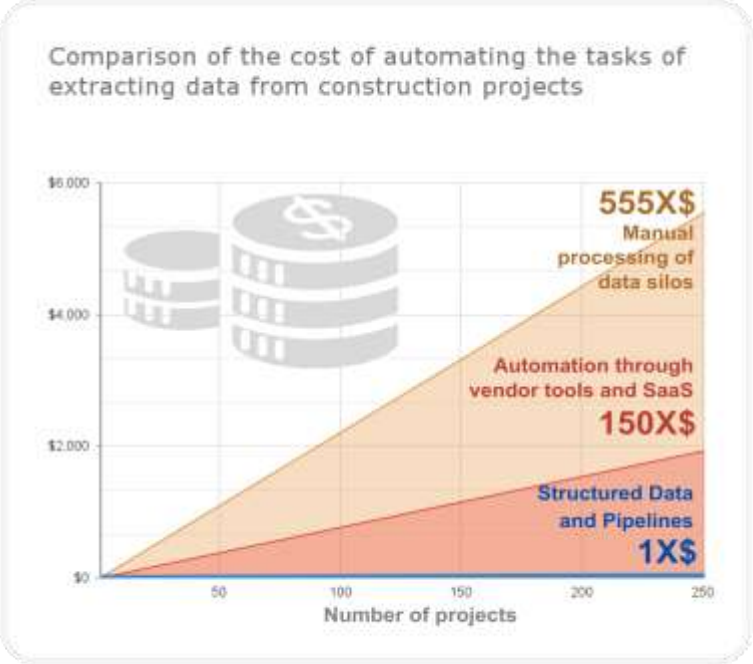
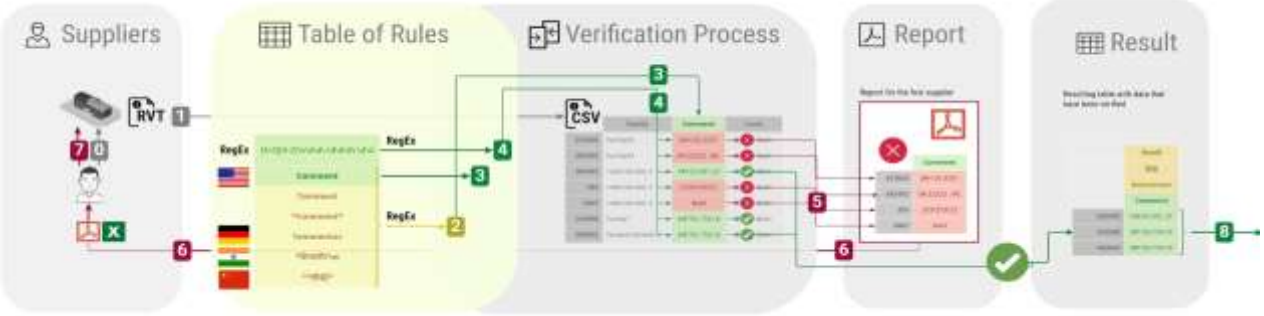


Fig. 7.3-7 Automatisering øger hastigheden af dataverifikation og -behandling, hvilket reducerer arbejdsomkostningerne med mange gange [140].

CAD-datavalideringsprocessen omfatter dataudtræk (ETL-fase Udtræk) fra forskellige lukkede (RVT, DWG, DGN, NWS osv.) eller åbne semistrukturerede og parametriske formater (IFC, CPXML, USD).) eller åbne semistrukturerede og parametriske formater (IFC, CPXML, USD), hvor regeltabeller kan anvendes på hver attribut og dens værdier (Transform stage) ved hjælp af regulære udtryk RegEx (Fig. 7.3-8), en proces, som vi diskuterede i detaljer i bogens fjerde del.

Oprettelsen af en PDF-fejlrapport og succesfuldt validerede poster bør afsluttes med output (Load step) i strukturerede formater, der kun tager højde for validerede enheder, der kan bruges til yderligere processer.



Figur 7.3-8 Datavalideringsprocessen fra projektets dataleverandører til den endelige rapport valideres ved hjælp af regulære udtryk.

Automatisering af valideringen af data fra CAD-systemer (BIM) med strukturerede krav og streaming af nye data, der behandles gennem ETL-Pipelines (figur 7.3-9), reducerer behovet for manuel involvering i valideringsprocessen (hver af validerings- og datakravsprocesserne er blevet diskuteret i tidligere kapitler).

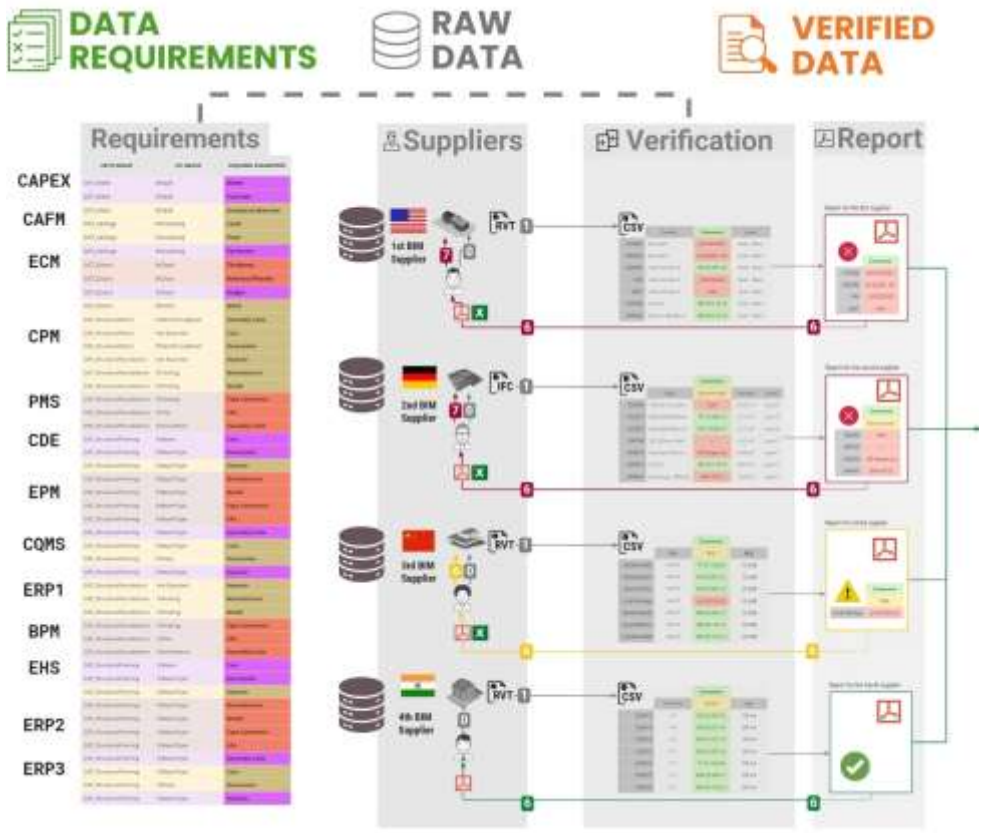


Fig. 7.3-9 Automatisering af datavalidering gennem ETL forenkler ledelsen af byggeprojekter ved at fremskynde processerne.

Traditionelt kan validering af modeller leveret af entreprenører og CAD (BIM)-specialister tage dage til uger. Men med indførelsen af automatiserede ETL-processer kan dette reduceres til nogle få minutter. I en typisk situation siger entreprenøren: *"Modellen er valideret og kompatibel."* Dette udsagn starter en kæde af verifikationer af entreprenørens påstand om datakvalitet:

- Projektleder - "*Entreprenøren siger: 'Modellen er blevet testet, alt er i orden'."*
- Data Manager - Indlæsningsvalidering:
  - Et simpelt script i Pandas opdager et brud på få sekunder. Automatisering eliminerer tvister:
    - Kategori: OST_StructuralColumns, Parameter: FireRating IS NULL.
    - Generer en liste over overtrædelses-ID'er→ eksporter til Excel/PDF.

  Et simpelt script i Pandas opdager bruddet på få sekunder:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"].  # Filtrering
issues = df[df["FireRating"].isnull()]  # Tomme værdier
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Eksportér ID'er
```

- Data Manager til Project Manager - *"En kontrol af viser, at 18 kolonner ikke har FireRating-parameteren udfyldt."*
- Projektleder til entreprenør - "*Modellen returneres til revision: FireRating-parameteren er obligatorisk, uden den er accept umulig"*

Resultatet er, at CAD-modellen ikke skal valideres, automatisering eliminerer tvister, og entreprenøren modtager næsten øjeblikkeligt en struktureret rapport med en liste over ID'er på problematiske elementer. På denne måde bliver valideringsprocessen gennemsigtig, gentagelig og beskyttet mod menneskelige fejl (fig. 7.3-10).

> Denne tilgang gør datavalideringsprocessen til en teknisk funktion i stedet for manuel kvalitetskontrol. Det øger ikke kun produktiviteten, men gør det også muligt at anvende den samme logik på alle virksomhedens projekter, hvilket muliggør en ende-til-ende digital transformation af processer, fra design til drift.

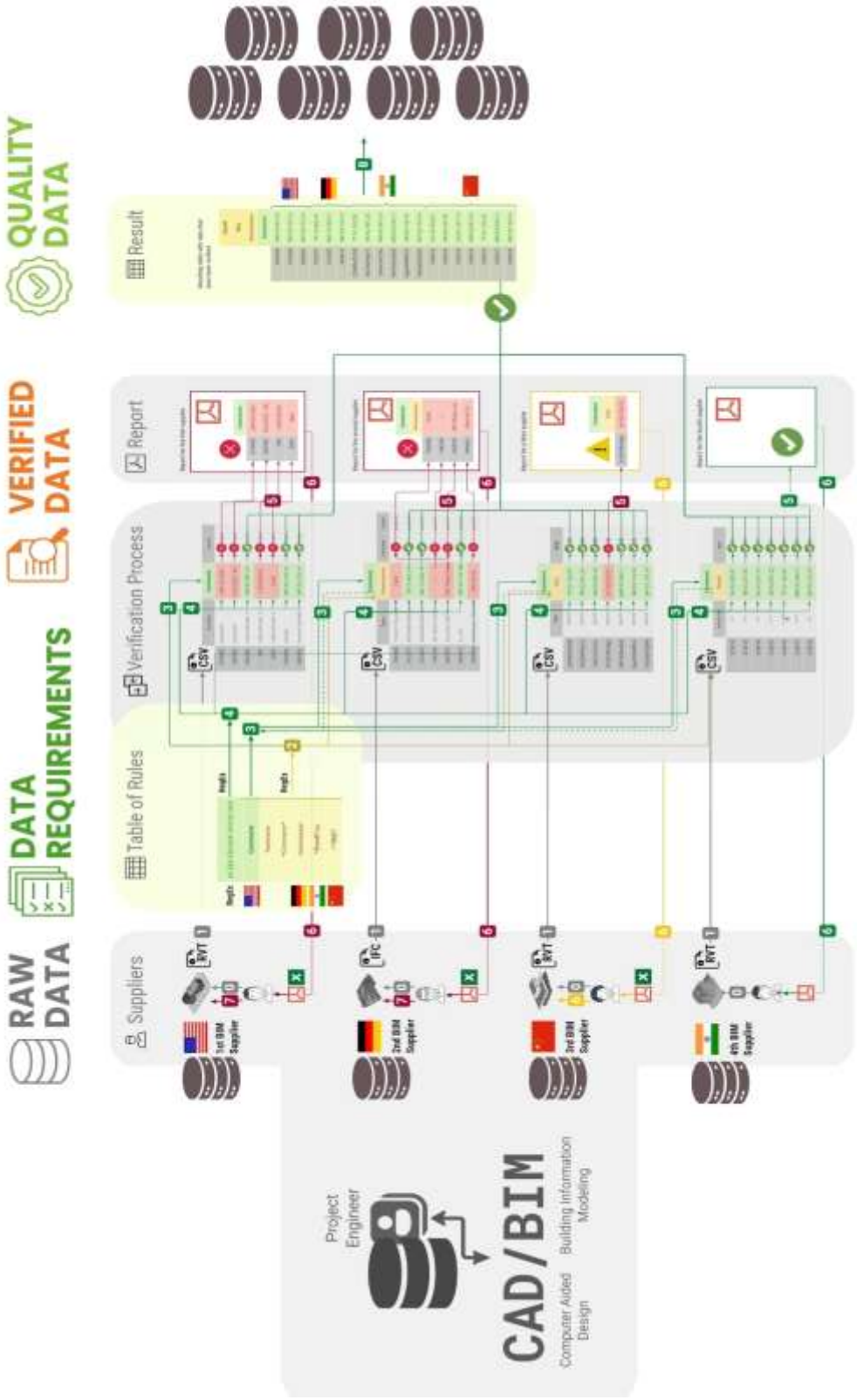


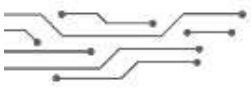
Fig. 7.3-10 Ved at automatisere kontrollen af elementattributter eliminerer man menneskelige fejl og reducerer sandsynligheden for fejl.

Ved hjælp af automatiserede pipelines (fig. 7.3-10) kan systembrugere, der forventer kvalitetsdata fra CAD-(BIM-) systemer, øjeblikkeligt få de outputdata, de har brug for - tabeller, dokumenter, billeder - og hurtigt integrere dem i deres arbejdsopgaver.

> Automatiseringen af kontrol, behandling og analyse medfører en ændring i den måde, byggeprojektledelse gribes an på, især interoperabiliteten mellem forskellige systemer, uden brug af komplekse og dyre modulære proprietære systemer eller lukkede leverandørløsninger.

Mens koncepter og markedsføringsakronymer kommer og går, vil selve valideringsprocesserne for datakrav altid forblive en integreret del af forretningsprocesserne. I stedet for at skabe flere og flere specialiserede formater og standarder bør byggebranchen se på værktøjer, der allerede har bevist deres effektivitet i andre brancher. I dag findes der stærke platforme til automatisering af databehandling og procesintegration, som gør det muligt for virksomheder at reducere den tid, de bruger på rutinearbejde, betydeligt og minimere fejl i Extract, Transform og Load.

Et af de populære eksempler på løsninger til automatisering og orkestrering af ETL processer er Apache Airflow, som gør det muligt at organisere komplekse beregningsprocesser og styre ETL-pipelines. Sammen med Airflow bruges andre lignende løsninger som Apache NiFi til datarouting og -streaming og n8n til automatisering af forretningsprocesser også aktivt.



KAPITEL 7.4.
ORKESTRERING AF ETL OG WORKFLOWS: PRAKTISKE LØSNINGER

DAG og Apache Airflow: automatisering og orkestrering af arbejdsgange

Apache Airflow er en gratis open source-platform, der er designet til at automatisere, orkestrere og overvåge workflows (ETL -conveyors).

Det er nødvendigt at arbejde med store mængder data hver dag:

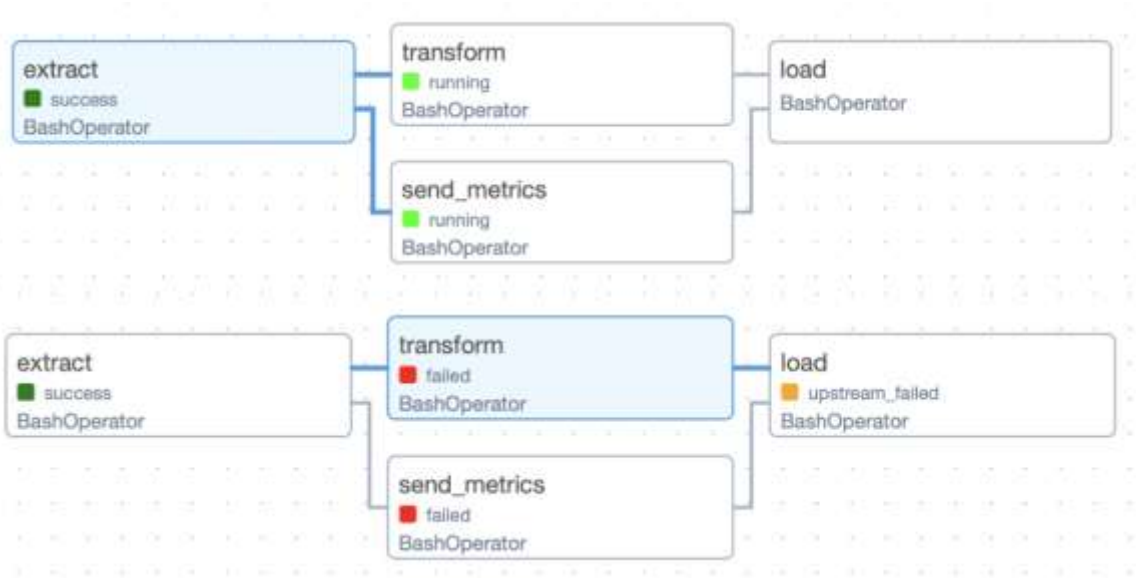
- Download filer fra forskellige kilder - Udtræk.eks. fra leverandører eller kunder).
- Transformér disse data til det ønskede format - Transformér (strukturér, rens og valider)
- Send resultater til verifikation og opret rapporter - Indlæs (upload til nødvendige systemer, dokumenter, databaser eller dashboards).

Manuel udførelse af sådanne ETL-processer tager lang tid og medfører risiko for menneskelige fejl. En ændring i datakilden eller en fejl i et af trinnene kan forårsage forsinkelser og forkerte resultater.

Automatiseringsværktøjer som Apache Airflow giver dig mulighed for at opbygge en pålidelig ETL - transportør, minimere fejl, reducere behandlingstiden og sikre, at data er korrekte i alle faser. Kernen i Apache Airflow er begrebet DAG (Directed Acyclic Graph) - en rettet acyklisk graf, hvor hver opgave (operatør) er forbundet med andre afhængigheder og udføres strengt i en bestemt rækkefølge. DAG eliminerer cyklusser, hvilket giver en logisk og forudsigelig struktur for opgaveudførelse.

Airflow tager sig af orkestrering - styring af afhængigheder mellem opgaver, kontrol af udførelsesplaner, sporing af status og automatisk reaktion på fejl. Denne tilgang minimerer manuel indgriben og sikrer pålideligheden af hele processen.

> **Task Orchestrator** er et værktøj eller system, der er designet til at styre og kontrollere udførelsen af opgaver i komplekse computer- og informationsmiljøer. Det letter processen med at implementere, automatisere og styre opgaveudførelse for at forbedre ydeevnen og optimere ressourcerne.



Figur 7.4-1 Apache Airflow har en brugervenlig grænseflade, hvor du kan visualisere DAG -ETL, se udførelseslogs, opstartsstatus for opgaver og meget mere.

Airflow bruges i vid udstrækning til orkestrering og automatisering af distribueret computing, databehandling, ETL (Extract, Transform, Load) processtyring, opgaveplanlægning og andre datascenarier. Som standard bruger Apache Airflow SQLite som database.

Et eksempel på en simpel DAG, der ligner ETL, består af opgaver - Extract, Transform og Load. I grafen, som styres via brugergrænsefladen (fig. 7.4-1), defineres rækkefølgen for udførelse af opgaver (kodefragmenter): For eksempel udføres extract først, derefter transform (og sending_metrics), og load-opgaven afslutter arbejdet. Når alle opgaver er udført, betragtes dataindlæsningsprocessen som vellykket.

Apache Airflow: praktisk anvendelse ETL-automatisering

Apache Airflow bruges i vid udstrækning til at organisere komplekse databehandlingsprocesser, hvilket gør det muligt at opbygge fleksible ETL -transportører. Apache Airflow kan køres enten via en webgrænseflade eller programmatisk via Python-kode (fig. 7.4-2). I webinterfacet (fig. 7.4-3) kan administratorer og udviklere visuelt overvåge DAG'er, køre opgaver og analysere udførelsesresultater.

Ved hjælp af DAG kan du indstille en klar rækkefølge af opgaver, styre afhængighederne mellem dem og automatisk reagere på ændringer i kildedataene. Lad os se på et eksempel, hvor man bruger Airflow til at automatisere rapporteringsbehandling (fig. 7.4-2).

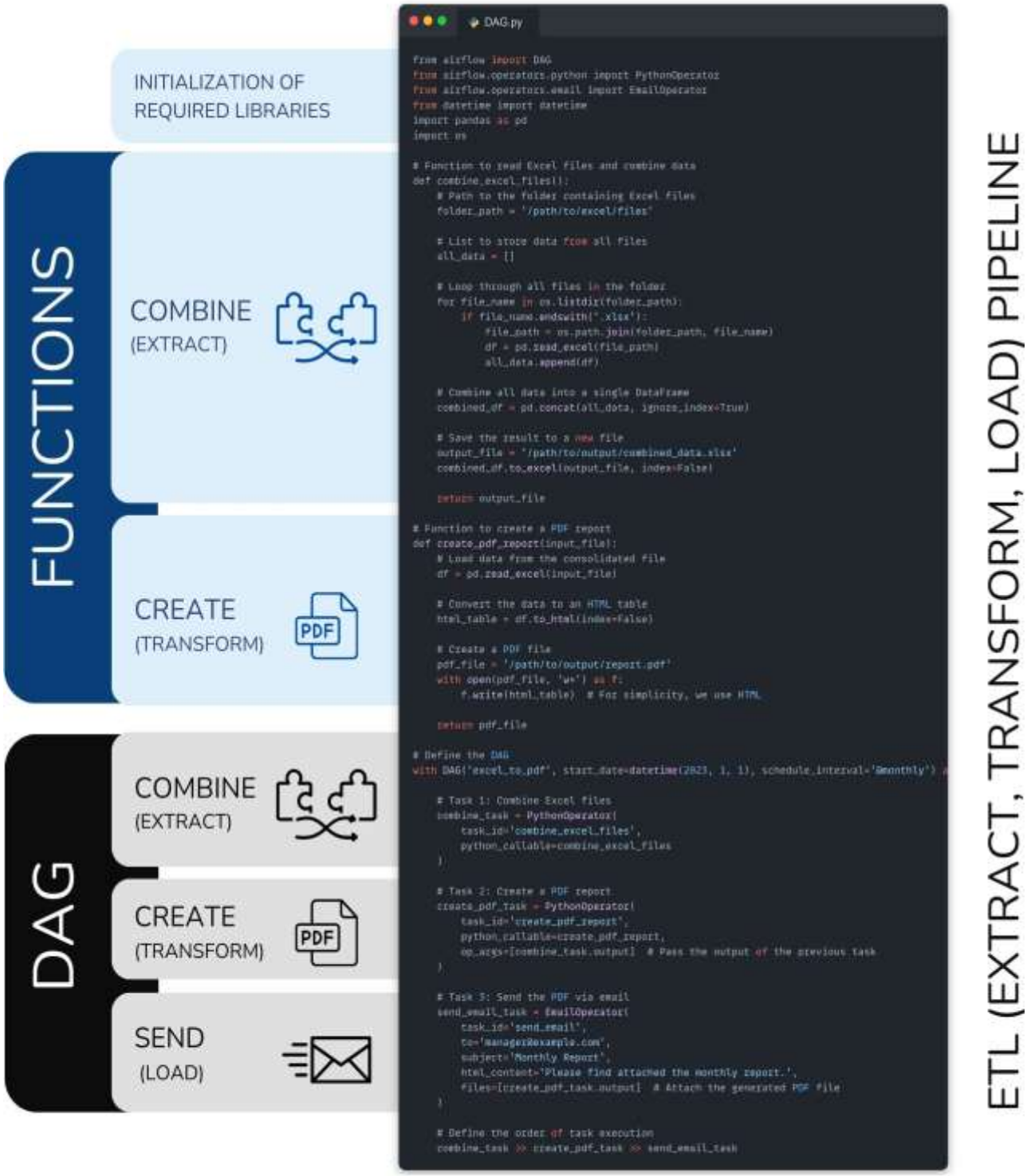


Fig. 7.4-2 ETL-koncept -transportør til databehandling ved hjælp af Apache Airflow.

Dette eksempel (figur 7.4-2) omhandler DAG, som udfører nøgleopgaver inden for ETL -transportøren:

- **Læs Excel -filer (Udtræk):**
  - Sekventiel gennemgang af alle filer i en given mappe.
  - Læs data fra hver fil ved hjælp af pandas-biblioteket.
  - Flet alle data til en enkelt DataFrame.
- **Opret PDF -dokument (Transform):**
  - Transformér den flettede DataFrame til en HTML -tabel.
  - Gem tabellen som PDF (i demoversionen - via HTML).
- **Afsendelse af en rapport via e-mail (Indlæs):**
  - Anvend EmailOperator til at sende PDF -dokumentet via e-mail.
- **Konfiguration af DAG:**
  - Definere rækkefølgen af opgaver: udtrække data→ generere rapport→ sende.
  - Tildeling af en startplan (@monthly - første dag i hver måned).

Det automatiserede ETL -eksempel (figur 7.4-2) viser, hvordan man indsamler data fra Excel -filer, opretter et PDF -dokument og sender det med e-mail. Dette er blot en af mange mulige brugsscenarier for Airflow. Dette eksempel kan tilpasses enhver specifik opgave for at forenkle og automatisere databehandlingen.

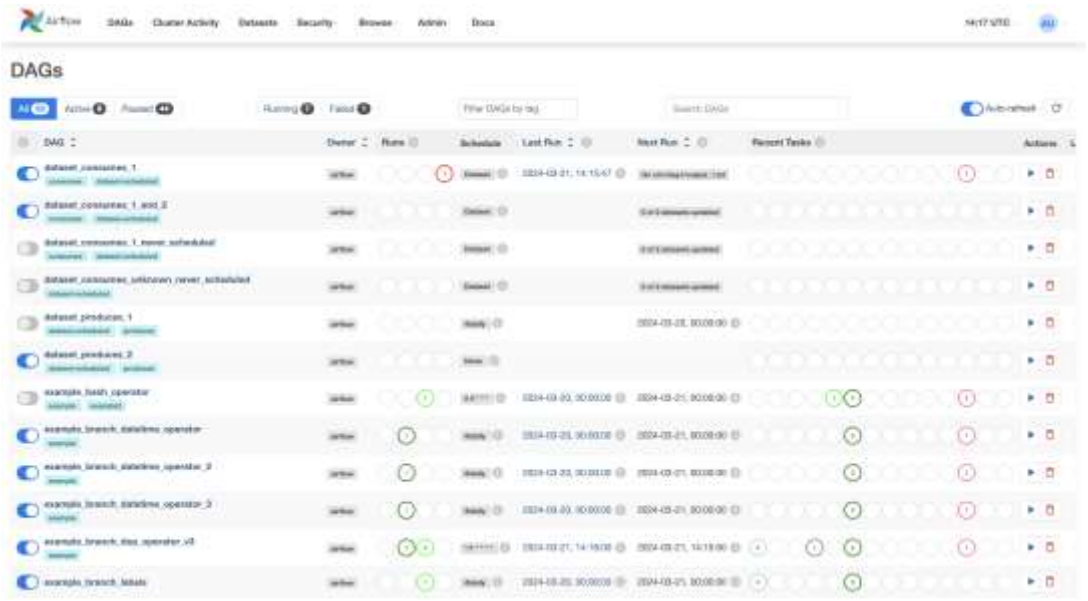


Fig. 7.4-3 Oversigt over alle DAG'er i miljøet med oplysninger om de seneste kørsler.

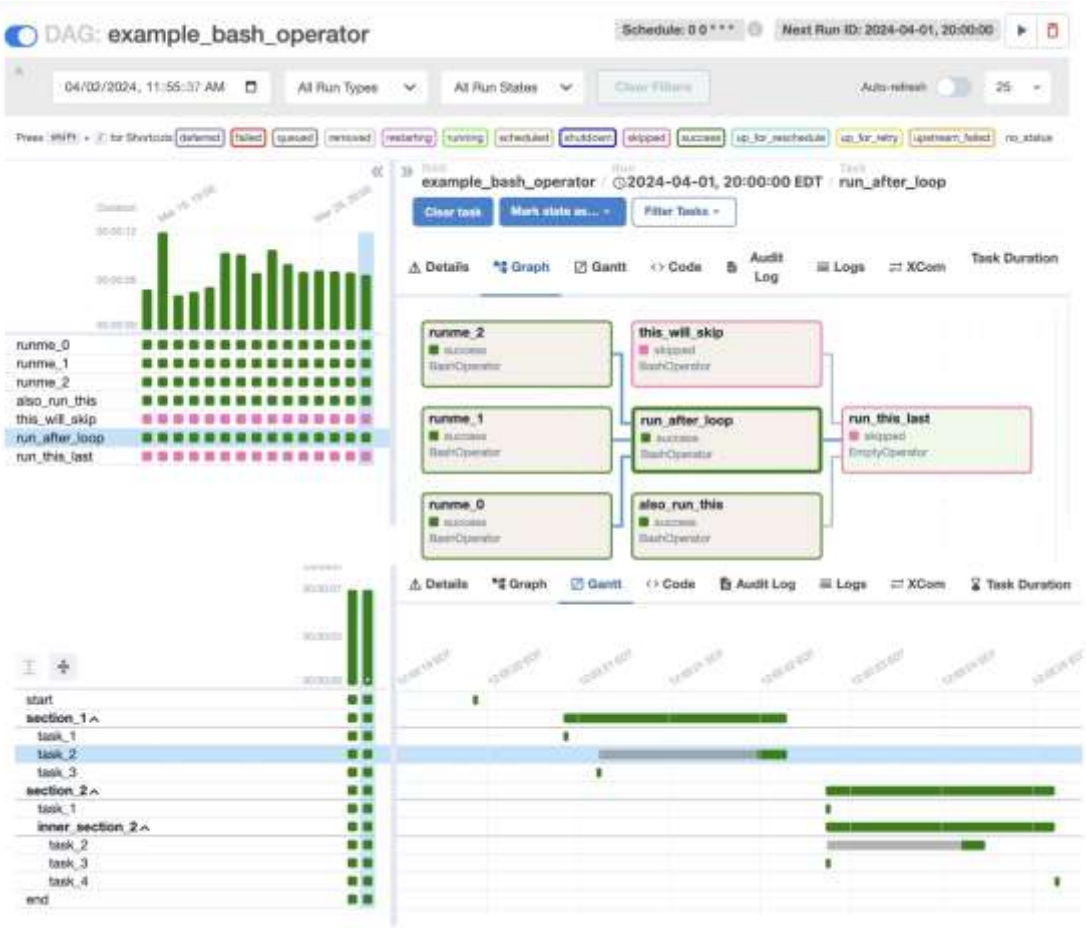
Apache Airflow-webgrænsefladen (figur 7.4-3) giver et omfattende visuelt miljø til styring af dataarbejdsgange. Det viser DAG'er som interaktive grafer, hvor noder repræsenterer opgaver, og kanter repræsenterer afhængigheder mellem dem, hvilket gør det nemt at spore komplekse data-workflows. Interfacet indeholder et dashboard med oplysninger om status for opgaveudførelse, kørselshistorik, detaljerede logfiler og præstationsmålinger. Administratorer kan manuelt starte opgaver, genstarte mislykkede operationer, suspendere DAG'er og konfigurere miljøvariabler, alt sammen via en intuitiv brugergrænseflade.

En sådan arkitektur kan suppleres med datavalidering, meddelelser om udførelsesstatus, integration med

eksterne API'er eller databaser. Airflow giver dig mulighed for fleksibelt at tilpasse DAG: tilføje nye opgaver, ændre deres rækkefølge, kombinere kæder - hvilket gør det til et effektivt værktøj til automatisering af komplekse databehandlingsprocesser. Når du kører DAG i Airflows webinterface (fig. 7.4-3, fig. 7.4-4), kan du overvåge opgavernes status. Systemet bruger farveindikation:

- Grøn - opgaven er gennemført med succes.
- Gul - processen er i gang.
- Rød - en fejl under udførelsen af opgaven.

I tilfælde af fejl (f.eks. manglende fil eller ødelagt datastruktur) begynder systemet automatisk at sende en meddelelse.



Figur 7.4-4 Apache Airflow forenkler i høj grad problemdiagnosticering, procesoptimering og teamsamarbejde om komplekse databehandlingspipelines.

Apache Airflow er praktisk, fordi den automatiserer rutineopgaver og eliminerer behovet for at udføre dem manuelt. Det sikrer pålidelighed ved at overvåge procesudførelsen og give øjeblikkelig fejlmeddelelse. Systemets fleksibilitet gør det nemt at tilføje nye opgaver eller ændre eksisterende og tilpasse arbejdsgange, så de opfylder skiftende krav.

Ud over Apache Airflow findes der lignende værktøjer til orkestrering af workflows. For eksempel tilbyder open source og gratis Prefect (figur 7.3-5) en enklere syntaks og integreres bedre med Python, Luigi, der er udviklet af Spotify, giver lignende funktionalitet og fungerer godt med big data. Det er også værd at bemærke Kronos og Dagster, som tilbyder moderne tilgange til opbygning af pipeline med fokus på modularitet og skalerbarhed. Valget af opgaveorkestreringsværktøj afhænger af projektets specifikke behov, men de hjælper alle med at automatisere komplekse ETL dataprocesser

Særligt interessant er Apache NiFi, en open source-platform, der er designet til streaming og routing af data. I modsætning til Airflow, som fokuserer på batchbehandling og afhængighedsstyring, fokuserer NiFi på datatransformation i realtid, on-the-fly og fleksibel routing mellem systemer.

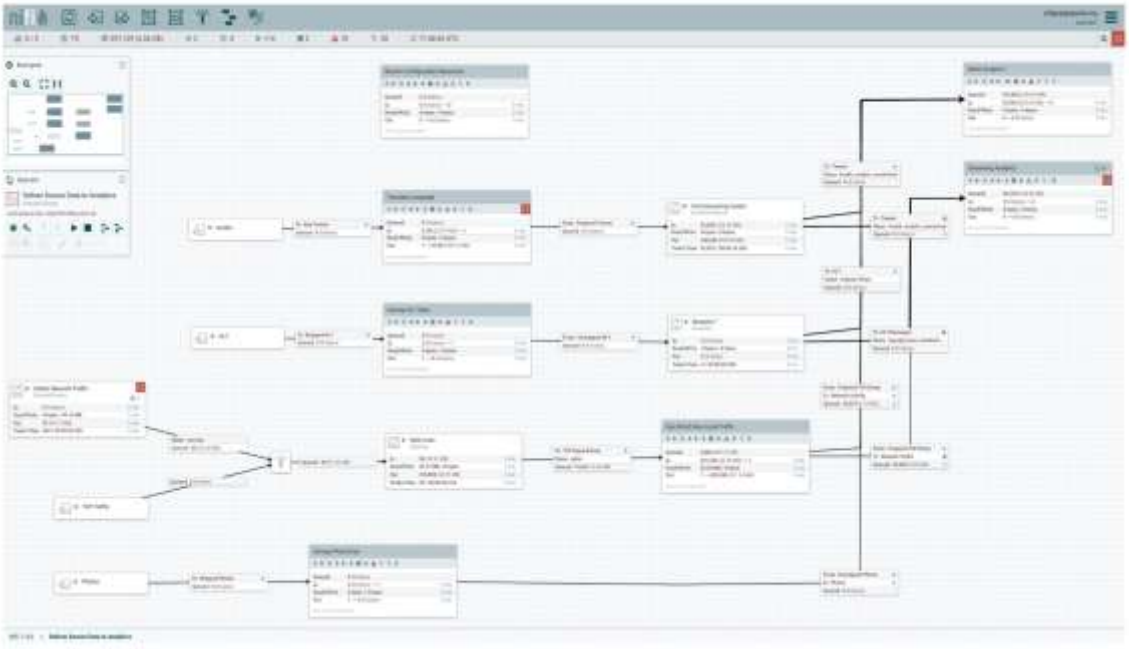
Apache NiFi til routing og datakonvertering

Apache NiFi er en kraftfuld open source-platform, der er designet til at automatisere datastrømme mellem forskellige systemer. Den blev oprindeligt udviklet i 2006 af det amerikanske National Security Agency (NSA) under navnet "Niagara Files" til internt brug. I 2014 blev projektet open source og overført til Apache Software Foundation, hvor det blev en del af deres teknologioverførselsinitiativer [141].

Apache NiFi er designet til at indsamle, behandle og overføre data i realtid. I modsætning til Airflow, som arbejder med batch-opgaver og kræver veldefinerede tidsplaner, arbejder NiFi i en strømbehandlingstilstand, så data løbende kan overføres mellem forskellige tjenester.

> Apache NiFi er ideel til integration med IoT -enheder, byggesensorer, overvågningssystemer og f.eks. streaming-validering af CAD -formater på en server, hvor der kan være behov for øjeblikkelig reaktion på ændringer i data.

Med indbyggede værktøjer til filtrering, transformation og routing gør NiFi det muligt at standardisere data (Transform), før de overføres (Load) til lager- eller analysesystemer. En af de største fordele er den indbyggede sikkerhedsunderstøttelse og adgangskontrol, som gør det til en pålidelig løsning til håndtering af følsomme oplysninger.



Figur 7.4-5 Grafisk fremstilling af dataflow i Apache NiFi-grænsefladen.

Apache NiFi håndterer effektivt datastreaming, filtrering og routing i realtid. Det er ideelt til teknisk intensive scenarier, hvor stabil informationsoverførsel mellem systemer og høj gennemstrømning er vigtig.

Men når hovedformålet er at integrere forskellige tjenester, automatisere rutineoperationer og hurtigt oprette arbejdsgange uden dyb viden om programmering, er der behov for løsninger med en lav indgangstærskel og maksimal fleksibilitet. Et sådant værktøj er n8n - en Low-Code /No-Code klasseplatform med fokus på forretningsautomatisering og visuel orkestrering af processer.

n8n Low-Code, No-Code procesorkestrering

n8n er en Open Source Low-Code / No-Code platform til opbygning af automatiserede arbejdsgange, der er kendetegnet ved brugervenlighed, fleksibilitet og evnen til hurtigt at integrere med en lang række eksterne tjenester.

> **No-Code** er en metode til at skabe digitale produkter uden at skrive kode. Alle elementer i processen - fra logik til grænseflade - realiseres udelukkende ved hjælp af visuelle værktøjer. No-Code-platforme henvender sig til brugere uden teknisk baggrund og gør det muligt hurtigt at skabe automatiseringer, formularer, integrationer og webapplikationer. Eksempel: En bruger opretter automatisk afsendelse af notifikationer eller Google Sheets-integration via en træk-og-slip-grænseflade uden programmeringsviden.

Med open source og lokale implementeringsmuligheder giver n8n i automatiserings- og ETL Pipeline-oprettelsesprocesser virksomheder fuld kontrol over deres data, samtidig med at der sikres sikkerhed og uafhængighed af cloud-udbydere.

I modsætning til Apache Airflow, som er orienteret mod beregningsopgaver med rigid orkestrering og kræver kendskab til Python, giver n8n en visuel editor, der tillader scripting uden at kræve kendskab til programmeringssprog (figur 7.4-6). Selv om grænsefladen gør det muligt at oprette automatiserede processer uden at skrive kode (No-Code), kan brugerne i mere komplekse scenarier tilføje deres egne JavaScript og Python -funktioner for at udvide funktionerne (Low-Code).

> **Low-Code** er en tilgang til softwareudvikling, hvor den grundlæggende logik i en applikation eller proces skabes ved hjælp af en grafisk grænseflade og visuelle elementer, og hvor programkoden kun bruges til at tilpasse eller udvide funktionaliteten. Low-Code-platforme gør det muligt at fremskynde udviklingen af løsninger betydeligt ved ikke kun at involvere programmører, men også forretningsbrugere med grundlæggende tekniske færdigheder. Eksempel: En bruger kan opbygge en forretningsproces ud fra færdige blokke og om nødvendigt tilføje sit eget script i JavaScript eller Python.

Selvom n8n er positioneret som en platform med en lav indgangstærskel, kræver det grundlæggende viden om programmering, forståelse af webteknologier og færdigheder i at arbejde med API. Systemets fleksibilitet gør det muligt at tilpasse det til en bred vifte af opgaver - fra automatiseret databehandling til integration med messengers, IoT -enheder og cloudtjenester.

Vigtige funktioner og fordele ved at bruge n8n:

- **Open source** og lokale implementeringsmuligheder sikrer fuld datakontrol, overholdelse af sikkerhedskrav og uafhængighed af cloud-udbydere.
- **Integration med over 330 tjenester,** herunder CRM, ERP, e-handel, cloud-platforme, messengers og databaser.
- **Scenariefleksibilitet**: fra enkle meddelelser til komplekse kæder med API-behandling af -forespørgsler, beslutningslogik og tilslutning af AI -tjenester.
- **Understøttelse af JavaScript og Python**: Brugere kan indlejre brugerdefineret kode efter behov og udvide automatiseringsmulighederne.
- **Intuitiv visuel grænseflade:** Giver dig mulighed for hurtigt at konfigurere og visualisere alle procestrin.

Low-Code-klasseplatforme giver værktøjer til at skabe digitale løsninger med minimal kode, hvilket gør dem ideelle til teams, der ikke har dyb teknisk ekspertise, men som har brug for at automatisere processer.

I byggeriet kan n8n bruges til at automatisere en lang række processer som f.eks. integration med projektstyringssystemer, stream checking, skrive out-of-the-box rapporter og breve, automatisk opdatering af materialelagerdata, sende meddelelser om opgavestatus til teams og meget mere. En skræddersyet pipeline i n8n kan reducere manuelle operationer med flere gange, reducere sandsynligheden for fejl og fremskynde beslutningstagningen i forbindelse med projektudførelse.

> Du kan vælge mellem næsten to tusind færdige, gratis og open source n8n Pipeline, som findes på: n8n.io/workflows, for at automatisere både arbejdsgange i byggeriet og personlige opgaver og reducere rutinemæssige operationer.

Tag en af de færdige Pipeline-skabeloner, der er gratis tilgængelige på n8n.io [142], som automatisk opretter udkast til svar i Gmail (fig. 7.4-6), hvilket hjælper brugere, der modtager en stor mængde e-mails eller har svært ved at skrive svar.

Denne n8n-skabelon "Gmail AI Auto-Responder: Create draft responses to incoming emails" (figur 7.4-6) analyserer indgående e-mails ved hjælp af LLM fra ChatGPT, afgør, om der er behov for et svar, genererer en kladde fra ChatGPT og konverterer teksten til HTML og føjer den til meddelelseskæden i Gmail. Dette sender ikke automatisk e-mailen, så du kan manuelt redigere og godkende svaret. Opsætningen tager ca. 10 minutter og omfatter OAuth-konfiguration af Gmail API og OpenAI API-integration. Resultatet er en praktisk og gratis løsning til at automatisere rutinemæssig e-mailkommunikation uden at miste kontrollen over indholdet af e-mails.

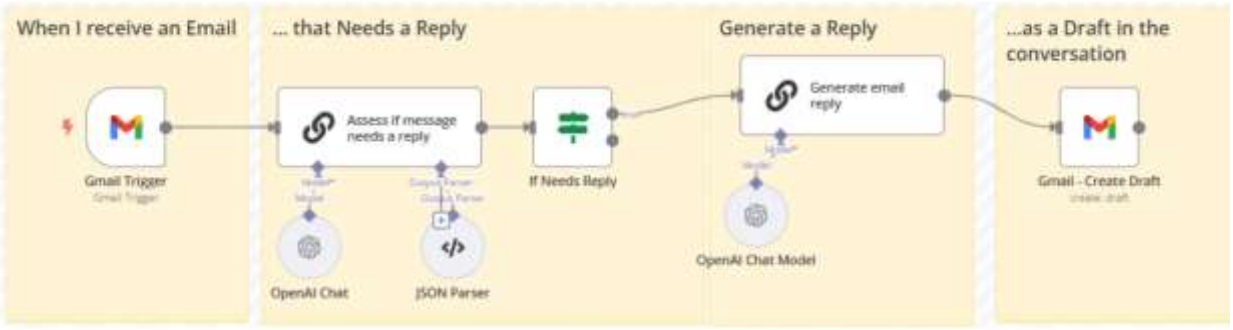


Fig. 7.4-6 Automatisk generering af e-mail-svar ved hjælp af n8n.

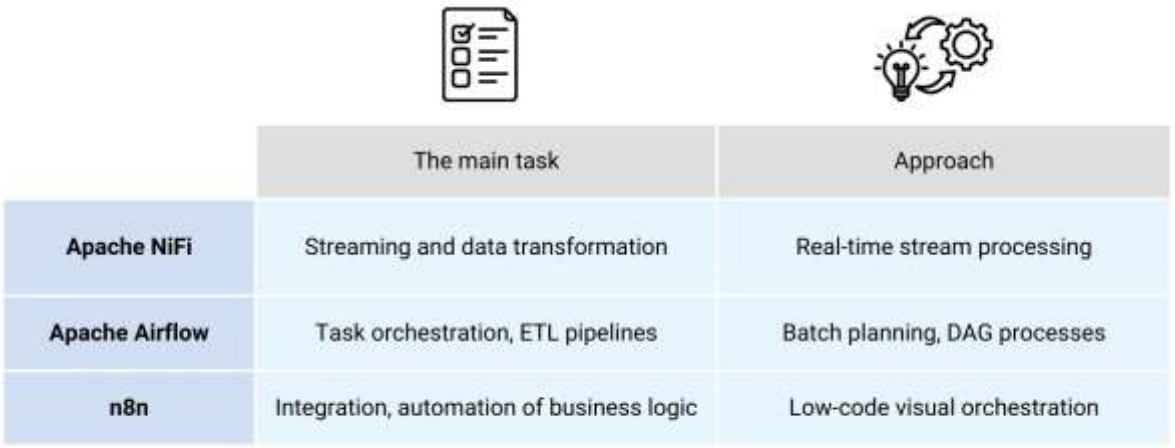
Et andet eksempel på automatisering med n8n er at finde gode tilbud på ejendomsmarkedet [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", indsamler dagligt relevante tilbud, der matcher de givne kriterier, ved hjælp af Zillow API. Den beregner automatisk vigtige investeringsmålinger (Cash on Cash ROI, Monthly Cash Flow, Down Payment), opdaterer Google Sheets og sender en sammenfattende rapport til e-mail (figur 7.4-7), så investorer kan spare tid og reagere hurtigt på de bedste tilbud.



Fig. 7.4-7 Automatiseret proces til vurdering af ejendommes investeringsattraktivitet.

Dens fleksibilitet og udvidelsesmuligheder gør n8n til et værdifuldt værktøj for virksomheder, der ønsker at transformere sig digitalt og blive mere konkurrencedygtige på markedet med relativt enkle og gratis open source-værktøjer.

Værktøjer som Apache NiFi, Airflow og n8n kan betragtes som tre lag af databehandling (figur 7.4-8). NiFi styrer datastrømmen og sikrer levering og transformation, Airflow orkestrerer opgaveudførelsen ved at samle data i behandlingspipelines, og n8n automatiserer integrationen med eksterne tjenester og styrer forretningslogikken.



	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

Fig. 7.4-8 Apache Airflow, Apache NiFi og n8n kan ses som tre komplementære lag i en moderne datastyringsarkitektur.

Tilsammen udgør disse gratis og open source-værktøjer potentielt et eksempel på et effektivt økosystem til data- og processtyring i byggebranchen, som gør det muligt for virksomheder at udnytte information til beslutningstagning og procesautomatisering.

Næste skridt: at gå fra manuel drift til analysebaserede løsninger

Nutidens byggevirksomheder opererer i et miljø med stor usikkerhed: skiftende materialepriser, forsinkede leverancer, mangel på arbejdskraft og stramme projektdeadlines. Brugen af analytiske dashboards, ETL -conveyors og BI-systemer hjælper virksomheder med hurtigt at identificere problemområder, vurdere ressourceeffektivitet og forudsige ændringer, før de fører til økonomiske tab.

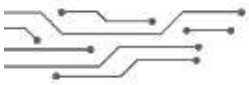
For at opsummere denne del er det værd at fremhæve de vigtigste praktiske trin, der vil hjælpe dig med at anvende de diskuterede teknologier i dine daglige opgaver:

- Implementere datavisualiseringer og analytiske dashboards
  - Beherske processen med at oprette dashboards til overvågning af nøglepræstationsindikatorer (KPI'er)
  - Brug visualiseringsværktøjer til dine data (Power BI, Tableau, Matplotlib, Plotly)
- Automatiser databehandling gennem ETL -processer
  - Opsæt automatisk dataindsamling fra forskellige kilder (dokumentation, tabeller, CAD) via ETL-processer.
  - Organiser datatransformation (f.eks. kontrol af regulære udtryk eller beregning) ved hjælp af Python-scripts
  - Prøv at opsætte automatisk PDF- (eller DOC-) rapportering med FPDF-biblioteket, ved at bruge data fra Excel-filer eller udtrække information fra andre PDF-dokumenter.
- Brug sprogmodeller (LLM) til automatisering

- [ ] Brug store sprogmodeller (LLM'er) til at generere kode, der hjælper med at udtrække og analysere data fra ustrukturerede dokumenter.
- [ ] Gør dig bekendt med automatiseringsværktøjet n8n, og udforsk de færdige skabeloner og casestudier på deres hjemmeside. Find ud af, hvilke processer i dit arbejde der kan automatiseres fuldt ud ved hjælp af No-Code/Low-Code-tilgangen.

En analytisk tilgang til data- og procesautomatisering reducerer ikke kun den tid, der bruges på rutinearbejde, men forbedrer også kvaliteten af beslutningstagningen. Virksomheder, der implementerer visuelle analyseværktøjer og ETL -conveyors, får mulighed for at reagere hurtigt på ændringer

Automatisering af forretningsprocesser ved hjælp af værktøjer som n8n, Airflow og NiFi er kun det første skridt mod digital modenhed. Det næste skridt er kvalitetsopbevaring og -styring af de data, der ligger til grund for automatiseringen. I del 8 ser vi nærmere på, hvordan byggevirksomheder kan opbygge en bæredygtig datalagringsarkitektur og gå fra et kaos af dokumenter og filer i flere formater til centraliserede lagrings- og analyseplatforme.





VIII DEL
DATALAGRING OG -STYRING I BYGGERIET

Del 8 udforsker moderne datalagrings- og styringsteknologier til byggebranchen. Den analyserer effektive formater til håndtering af store informationsmængder - fra simple CSV og XLSX til de mere effektive Apache Parquet og ORC med en detaljeret sammenligning af deres muligheder og begrænsninger. Begreberne datalager (DWH), datasøer) og deres hybridløsninger (Data Lakehouse) samt principperne for Data Governance) og Dataminimalisme) diskuteres. Problemerne med Data Swamp) og strategier til at forhindre kaos i informationssystemer dækkes i detaljer. Nye tilgange til at arbejde med data præsenteres, herunder vektordatabaser og deres anvendelse i konstruktion gennem begrebet Bounding Box. Denne del berører også metoderne DataOps og VectorOps som nye standarder for organisering af dataarbejdsgange.

KAPITEL 8.1.
DATAINFRASTRUKTUR: FRA LAGRINGSFORMATER TIL DIGITALE ARKIVER

Dataatomer: grundlaget for effektiv informationsstyring

Alt i universet består af de mindste byggesten - atomer og molekyler, og med tiden vender alle levende og ikke-levende ting uundgåeligt tilbage til denne oprindelige tilstand. I naturen sker denne proces med en forbløffende hastighed, som vi forsøger at overføre til de processer, der styres af mennesker.

> I skoven bliver alle levende organismer i sidste ende omdannet til et næringsrigt stof, der danner grundlag for nye planter. Disse planter bliver igen føde for nye levende ting, der består af de samme atomer, som skabte universet for millioner af år siden.

I forretningsverdenen er det også vigtigt at nedbryde komplekse, flerlagede strukturer til deres mest grundlæggende, minimalt forarbejdede enheder ligesom atomer og molekyler i naturen. Det gør det muligt at lagre og administrere atomer af data effektivt og forvandle dem til et rigt, frugtbart fundament, der bliver en nøgleressource for væksten i analyse og beslutningskvalitet.



Fig. 8.1-1 Analyse og beslutningstagning er baseret på genbrugte data, som engang er blevet behandlet og lagret.

Musikalske kompositioner består af toner, der kombineres for at skabe komplekse musikstykker, mens ord skabes ud fra en primitiv enhed, bogstavlyden. Uanset om det drejer sig om natur, videnskab, økonomi, kunst eller teknologi, udviser verden en bemærkelsesværdig enhed og harmoni i sin stræben efter

ødelæggelse, struktur, cykling og skabelse. På samme måde nedbrydes processer i omkostningssystemer til små strukturerede enheder - ressourceelementer - på omkostningsniveau og tidsplaner. Disse enheder bruges så, ligesom sedler, til at danne mere komplekse beregninger og skemaer. Det samme princip anvendes af computerstøttede designsystemer, hvor komplekse arkitektoniske og tekniske projekter opbygges af grundelementer - individuelle elementer og bibliotekskomponenter, hvorfra der skabes en komplet 3D -model af projektet af en kompleks bygning eller struktur.

> Konceptet med cykliskhed og struktur i naturen og videnskaben afspejles også i den moderne dataverden. Ligesom alle levende væsener i naturen vender tilbage til atomer og molekyler, så har information en tendens til at vende tilbage til sin mest primitive form i en verden med moderne databehandlingsværktøjer.

De mindste elementer, med deres endelige udelelighed, er de grundlæggende byggesten i forretningsprocesser. Det er vigtigt, at man fra starten nøje overvejer, hvordan man indsamler, strukturerer (nedbryder til atomer) og lagrer disse små byggesten fra forskellige kilder. Organisering og lagring af data er ikke kun et spørgsmål om at bryde dem ned i deres bestanddele. Det er lige så vigtigt at sikre, at de integreres og lagres på en struktureret måde, så data nemt kan hentes, analyseres og bruges til beslutningstagning, når der er brug for det.

For at behandle information effektivt skal formatet og metoderne til datalagring vælges omhyggeligt - ligesom jorden skal forberedes til træernes vækst. Datalagre skal organiseres for at sikre opdateret information af høj kvalitet og eliminere overflødige eller irrelevante data. Jo bedre denne "informationsjord" er struktureret, jo hurtigere og mere præcist vil brugerne kunne finde de rigtige data og løse analytiske problemer.

Lagring af information: filer eller data

Datavarehuse giver virksomheder mulighed for at indsamle og kombinere oplysninger fra forskellige systemer og skabe et enkelt center for yderligere analyser. De indsamlede historiske data muliggør ikke kun en dybere analyse af processer, men også identifikation af mønstre, der kan påvirke virksomhedens resultater.

Lad os sige, at en virksomhed arbejder på flere projekter på samme tid. En ingeniør vil gerne vide, hvor meget beton der er blevet støbt, og hvor meget der stadig mangler at blive købt. Med en traditionel tilgang ville han være nødt til manuelt at søge på serveren og åbne flere estimattabeller, sammenligne dem med certifikaterne for udført arbejde og kontrollere de aktuelle lagerbeholdninger. Det tager timer eller endda dage. Selv med ETL-processer og automatiske scripts forbliver opgaven semi-manuel: Ingeniøren skal stadig manuelt angive stien til mapper eller specifikke filer på serveren. Det reducerer den samlede effekt af automatisering, da det fortsat tager værdifuld arbejdstid.

Når man skifter til datastyring, får teknikeren i stedet for at arbejde med serverens filsystem adgang til en samlet lagerstruktur, hvor oplysningerne opdateres i realtid. En enkelt forespørgsel - i form af kode, SQL - forespørgsel eller endda et LLM -agentopkald - kan øjeblikkeligt give nøjagtige data om aktuelle saldi, udført arbejde og kommende leverancer, hvis dataene er blevet forberedt på forhånd og kombineret i et datalager, hvor der ikke er behov for at vandre gennem mapper, åbne dusinvis af filer og manuelt sammenligne værdier.

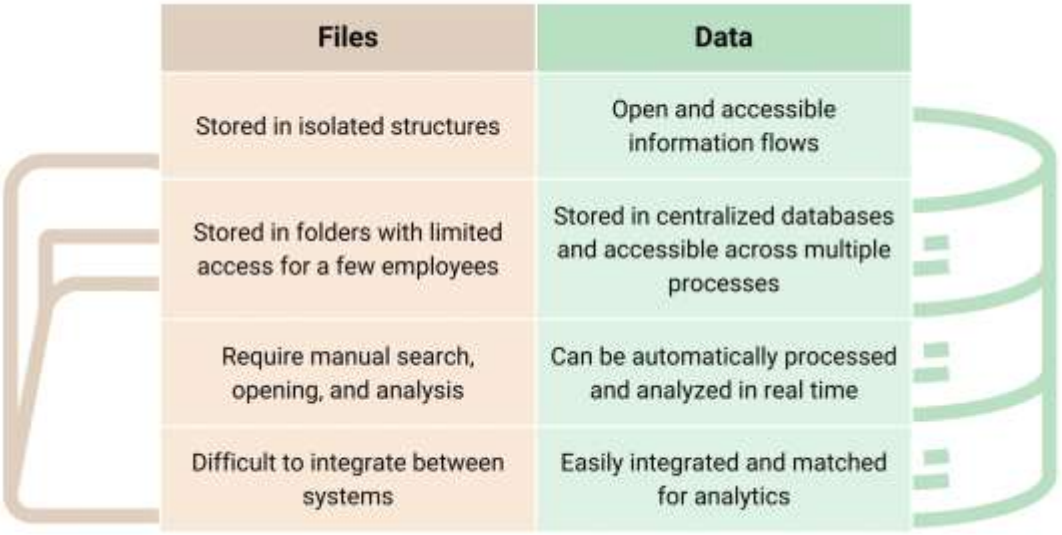
I lang tid har byggevirksomheder brugt PDF -dokumenter, DWG -tegninger, RVT -modeller og hundredvis og tusindvis af Excel -tabeller og andre forskellige formater, som er gemt i specifikke mapper på virksomhedens servere, hvilket gør det vanskeligt at søge efter oplysninger, kontrollere og analysere dem. Som følge heraf flyttes de filer, der er tilbage efter afslutningen af projekter, oftest tilbage til serveren i arkivmapper, som praktisk talt ikke bruges i fremtiden. En sådan traditionel filbaseret datalagring mister sin relevans, efterhånden som datastrømmen øges, fordi den er sårbar over for menneskelige fejl.

> **En fil** er bare en isoleret beholder, hvor data er gemt. Filer er skabt til mennesker, ikke til systemer, så de kræver manuel åbning, læsning og fortolkning. Eksempler er en Excel -tabel, et PDF -dokument eller en CAD -tegning, der skal åbnes specifikt i et bestemt værktøj for at få adgang til de ønskede oplysninger. Uden struktureret hentning og behandling forbliver informationen i det ubrugt.
>
> **Data** er til gengæld maskinlæsbare oplysninger, der sammenkædes, opdateres og analyseres automatisk. I et enkelt data warehouse (f.eks. database, DWH eller Data Lake) er information repræsenteret i form af tabeller, poster og relationer. Det giver mulighed for ensartet lagring, automatiserede forespørgsler, værdianalyser og rapportering i realtid.

Brug af data i stedet for filer (fig. 8.1-1) gør det muligt at eliminere den manuelle søgeproces og ensrette behandlingen. Virksomheder, der allerede har implementeret en sådan tilgang, får en konkurrencemæssig fordel på grund af den hurtige adgang til information og muligheden for hurtigt at integrere den i forretningsprocesser.

> Skiftet fra at bruge filer til data er en uundgåelig forandring, som vil forme byggebranchens fremtid.

Alle virksomheder i byggebranchen vil stå over for et vigtigt valg: Fortsæt med at gemme information i forskellige filer og siloer, som skal læses af mennesker ved hjælp af særlige programmer, eller omdan dem i de første faser af behandlingen til strukturerede data og skab et enkelt integreret digitalt fundament for automatiseret projektstyring.



Figur 8.1-1 Udvikling af informationsflow: fra isolerede filer til integrerede data.

Med eksplosionen af information bliver de traditionelle metoder til lagring og behandling af filer mindre og mindre effektive. I byggebranchen, som i andre sektorer, er det ikke længere tilstrækkeligt at forlade sig på forskellige filmapper med forskellige filformater eller databaser uden forbindelse.

Virksomheder, der ønsker at forblive konkurrencedygtige i den digitale tidsalder, vil uundgåeligt bevæge sig i retning af integrerede digitale platforme, bruge big data-teknologier og automatiserede analysesystemer.

At gå fra filbaseret lagring til datadrevne arbejdsgange vil kræve en nytænkning af tilgangen til informationsstyring og et bevidst valg af formater, der egner sig til yderligere integration i centraliserede arkiver. Dette valg vil afgøre, hvor effektivt data kan behandles, hvor hurtigt man kan få adgang til dem, og hvor let de kan integreres i virksomhedens digitale processer.

Big Data Storage: Analyse af populære formater og deres effektivitet

Lagerformater spiller en vigtig rolle for analyseinfrastrukturens skalerbarhed, pålidelighed og ydeevne. Til at analysere og behandle data - f.eks. filtrering, gruppering og aggregering - brugte vores eksempler Pandas DataFrame - en populær struktur til at arbejde med data i RAM.

Pandas DataFrame har dog ikke sit eget lagringsformat, så når behandlingen er færdig, eksporteres dataene til et af de eksterne formater - oftest CSV eller XLSX. Disse tabelformater er nemme at udveksle og kompatible med de fleste eksterne systemer, men har en række begrænsninger: lav lagringseffektivitet, manglende komprimering og dårlig understøttelse af versionering:

- **CSV** (Comma-Separated Values): et simpelt tekstformat, der understøttes bredt af forskellige platforme og værktøjer. Det er nemt at bruge, men understøtter ikke komplekse datatyper og komprimering.
- **XLSX** (Excel Open XML Spreadsheet): et Microsoft Excel-filformat, der understøtter sofistikerede funktioner som formler, diagrammer og styling. Det er nyttigt til manuel dataanalyse og visualisering, men det er ikke optimeret til databehandling i stor skala.

Ud over de populære tabelformater XLSX og CSV findes der flere populære formater til effektiv lagring af strukturerede data (figur 8.1-2), som hver især har unikke fordele afhængigt af specifikke krav til datalagring og analyse:

- **Apache Parquet**: et filformat til kolonnelagring af data, der er optimeret til brug i dataanalysesystemer. Det tilbyder effektiv datakomprimering og kodningsordninger, hvilket gør det ideelt til komplekse datastrukturer og behandling af store datamængder.
- **Apache ORC** (Optimised Row Columnar): I lighed med Parquet giver ORC høj komprimering og effektiv datalagring. Det er optimeret til tunge læseoperationer og egner sig godt til lagring af datasøer.
- **JSON** (JavaScript Object Notation): Selv om JSON ikke er så effektivt med hensyn til datalagring sammenlignet med binære formater som Parquet eller ORC, er det meget tilgængeligt og nemt at arbejde med, hvilket gør det ideelt til scripts, hvor læsbarhed og webkompatibilitet er vigtig.
- **Feather**: et hurtigt, let og brugervenligt analytisk orienteret binært søjleformet datalagringsformat. Det er designet til effektivt at overføre data mellem Python (Pandas) og R, hvilket gør det til et fremragende valg til projekter, der involverer disse programmeringsmiljøer.
- **HDF5**Hierarchical Data Format version 5): designet til at lagre og organisere store mængder data. Det understøtter en lang række datatyper og er velegnet til at arbejde med komplekse datasamlinger. HDF5 er især populært inden for videnskabelig databehandling på grund af dets evne til effektivt at lagre og få adgang til store datasæt.

	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)

Fig. 8.1-2 Sammenligning af dataformater, der viser de største forskelle i lagrings- og behandlingsaspekter.

For at udføre en sammenlignende analyse af de formater, der bruges i ETL-processens indlæsningsfase, blev der oprettet en tabel, der viser filstørrelser og læsetider (fig. 8.1-3). Filer med identiske data blev brugt i undersøgelsen: Tabellen indeholdt 10.000 rækker og 10 kolonner fyldt med tilfældige værdier.

Følgende lagringsformater er inkluderet i undersøgelsen: CSV, Parquet, XLSX og HDF5 samt deres komprimerede versioner i ZIP-arkiver. De rå data blev genereret ved hjælp af NumPy-biblioteket og repræsenteret som en Pandas DataFrame-struktur. Testprocessen bestod af følgende trin:

- Filopbevaring: Datarammen gemmes i fire forskellige formater: CSV, Parquet, XLSX og HDF5. Hvert format har unikke funktioner i den måde, det lagrer data på, hvilket påvirker filstørrelsen og læsehastigheden.
- ZIP-filkomprimering: For at analysere effektiviteten af standardkomprimering blev hver fil yderligere komprimeret til et ZIP-arkiv.
- Læsning af filer (ETL - Load): Læsetiden blev målt for hver fil efter udpakning fra ZIP. Dette gør det muligt at estimere dataadgangshastigheden efter udpakning fra arkivet.

Det er vigtigt at bemærke, at Pandas DataFrame ikke blev brugt direkte i analysen af størrelse eller læsetid, da den ikke repræsenterer et lagringsformat i sig selv. Den fungerede kun som en mellemstruktur til generering og efterfølgende lagring af data i forskellige formater.



	Original Size (MB)	ZIP Size (MB)	Read Time (s)
CSV	18.82	8.73	0.48
Parquet	9.67	9.22	0.17
XLSX	12.59	12.59	41.73
HDF5	8.50	7.62	0.13

Figur 8.1-3 Sammenligning af lagringsformater efter størrelse og læsehastighed.

CSV og HDF5 filer udviser (figur 8.1-3) høj komprimeringseffektivitet og reducerer deres størrelse betydeligt, når de pakkes i ZIP, hvilket kan være særligt nyttigt i scenarier, der kræver lageroptimering. XLSX -filer kan derimod stort set ikke komprimeres, og deres størrelse i ZIP forbliver sammenlignelig med den originale, hvilket gør dem mindre velegnede til brug i store datamængder eller i miljøer, hvor hurtig dataadgang er vigtig. Desuden er læsetiden for XLSX betydeligt højere sammenlignet med andre formater, hvilket gør det

mindre velegnet til hurtige datalæsninger. Apache Parquet har demonstreret høj ydeevne til analytiske opgaver og store datamængder på grund af sin søjlestruktur.

Optimer storage med Apache Parquet

Et af de populære formater til lagring og behandling af big data er Apache Parquet. Dette format er specielt designet til kolonnelagring (i lighed med Pandas), hvilket gør det muligt at reducere hukommelsesaftrykket betydeligt og øge hastigheden af analytiske forespørgsler. I modsætning til traditionelle formater som CSV og XLSX understøtter Parquet indbygget komprimering og er optimeret til big data-systemer, herunder Spark, Hadoop og cloud storage.

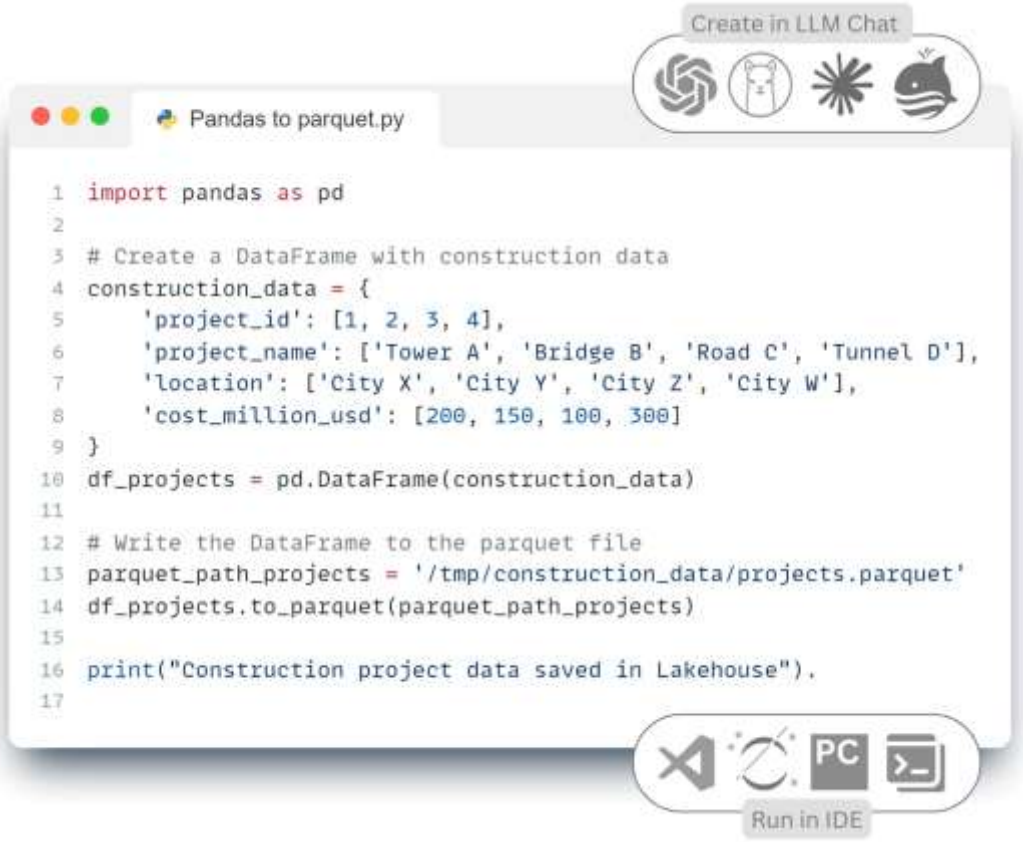
> Nøglefunktionerne i Parquet omfatter understøttelse af datakomprimering og -kodning, hvilket reducerer lagerstørrelsen betydeligt og fremskynder datalæsning ved at arbejde direkte på de ønskede kolonner i stedet for alle rækker af data.

Lad os bruge LLM som et illustrativt eksempel på, hvor nemt det er at få den nødvendige kode til at konvertere data til Apache Parquet.

- Send en tekstanmodning til LLM chat (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

> Skriv kode til at gemme data fra Pandas DataFrame til Apache Parquet. ⏎

LLM's svar :



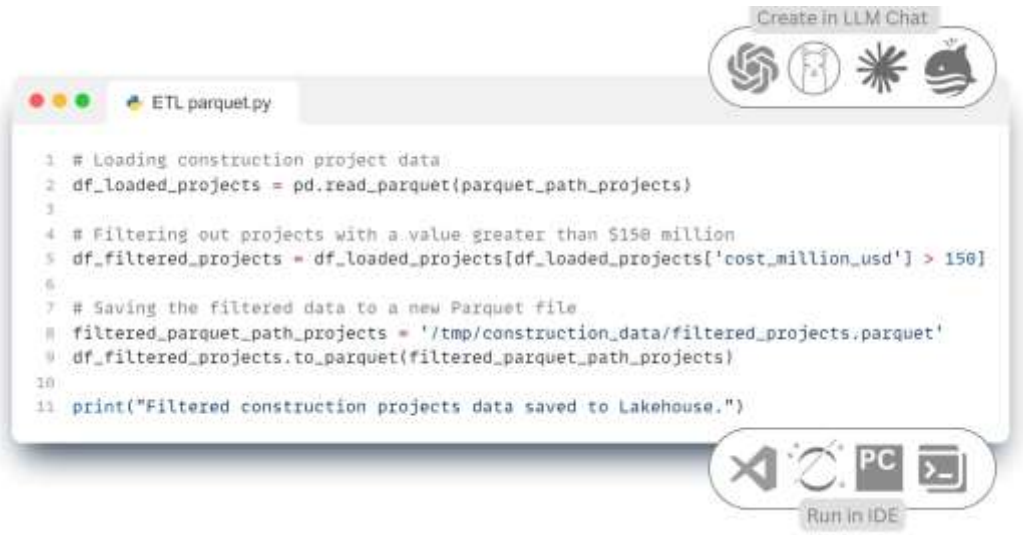
Figur 8.1-4 Overførsel af dataframe-data fra RAM til et lagringseffektivt format Apache Parquet ved hjælp af nogle få linjer Python .

Næste eksempel: Lad os simulere ETL-processen med data gemt i Parquetfor at filtrere projekter efter en bestemt værdi af en af attributterne "cost_million_usd" (fig. 8.1-4).

Mens chatten fortsætter, skal du sende en sms-anmodning til LLM:

> Skriv en kode, hvor vi vil filtrere dataene i en tabel og kun gemme de projekter (tabelrækker) fra Apache Parquet-dataene, hvis omkostninger (parameter cost_million_usd) overstiger 150 millioner dollars. ⏎

LLM's svar :



Figur 8.1-5 ETL-processen , når man arbejder med data i formatet Apache Parquet , ser ud på samme måde som med andre strukturerede formater.

Ved at bruge Parquet-formatet (i forhold til XLSX, CSV osv.) reducerer mængden af lagret information betydeligt og gør søgninger hurtigere. Det gør det fremragende til både lagring og analyse af data. Parquet kan integreres med forskellige behandlingssystemer og giver effektiv adgang i hybride arkitekturer.

Men et effektivt lagringsformat er kun ét element i en komplet dataoplevelse. For at skabe et bæredygtigt og skalerbart miljø er det nødvendigt med en veldesignet datastyringsarkitektur. Systemer i DWH-klassen (Data Warehouse) opfylder denne funktion. De giver mulighed for aggregering af data fra heterogene kilder, gennemsigtighed i forretningsprocesser og mulighed for komplekse analyser ved hjælp af BI-værktøjer og maskinlæringsalgoritmer.

DWH: Data Warehouse data warehouses

Ligesom formatet Parquet er optimeret til effektiv lagring af store mængder information, er Data Warehouse optimeret til at integrere og strukturere data for at understøtte analyser, prognoser og ledelsens beslutningstagning.

I nutidens virksomheder kommer data fra mange forskellige kilder: ERP-, CAFM-, CPM-, CRM-systemer, regnskabs- og lagerstyring, digitale CAD-modeller af bygninger, IoT-sensorer og andre løsninger. For at få et holistisk billede er det ikke nok bare at indsamle data - de skal organiseres, standardiseres og centraliseres i et enkelt lager. Det er præcis, hvad DWH gør - et centralt lagringssystem, der gør det muligt at samle oplysninger fra forskellige kilder, strukturere dem og gøre dem tilgængelige for analyser og strategisk ledelse.

> **DWH (Data Warehouse**) er et centraliseret data warehouse-system, der samler oplysninger fra flere kilder, strukturerer dem og gør dem tilgængelige for analyse og rapportering.

I mange virksomheder er data spredt over forskellige systemer, som vi diskuterede i de første dele af bogen (figur 1.2-4). DWH integrerer disse kilder og sikrer fuldstændig gennemsigtighed og pålidelighed af information. Et DWH-datalager er en specialiseret database (en stor database), der indsamler, behandler og lagrer data fra flere kilder. De vigtigste egenskaber ved et DWH er:

- **Bruge ETL -processer** Extract, Transform, Load) - udtrække data fra kilder, rense dem, transformere dem, indlæse dem i repositoriet og automatisere disse processer, som blev diskuteret i bogens syvende del.
- **Datagranularitet** - data i DWH kan gemmes både i aggregeret form (oversigtsrapporter) og i granulær form (rådata). Fra 2024 og fremefter er det CAD- leverandører, der er begyndt at tale om granulære data [125], hvilket måske indikerer, at branchen forbereder sig på overgangen til specialiseret cloud-lagring til håndtering af digitale bygningsmodeldata.
- **Understøttelse af analyser og forudsigelser** - datalagre danner grundlag for BI-værktøjer, Big Data -analyser og maskinlæring.

DWH er grundlaget for business intelligence, der muliggør analyse af nøgleindikatorer, prognoser for salg, indkøb og omkostninger samt automatiseret rapportering og datavisualisering (figur 8.1-6).

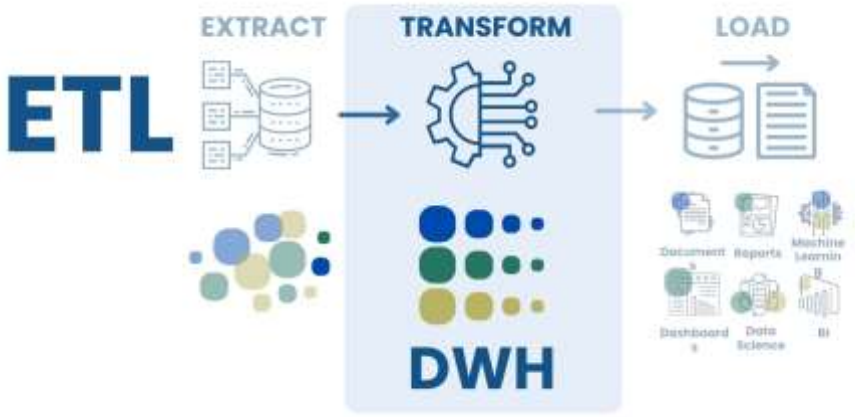


Fig. 8.1-6 I en ETL -proces kan DWH fungere som et centralt opbevaringssted, hvor data, der er hentet fra forskellige systemer, gennemgår transformations- og aflæsningsfaser.

DWH spiller en nøglerolle i integration, rensning og strukturering af information, hvilket giver et solidt grundlag for business intelligence og beslutningsprocesser. Men i dagens miljø, hvor datamængderne vokser hurtigt, og datakilderne bliver mere forskelligartede, kræver den traditionelle DWH-tilgang til informationslagring ofte en udvidelse i form af ELT og Data Lake

Data Lake - udvikling af ETL til ELT: fra traditionel rensning til fleksibel behandling

Klassiske DWH - data warehouses, der er designet til at lagre strukturerede data i et format, der er optimeret til analytiske forespørgsler, har haft begrænsninger med hensyn til håndtering af ustrukturerede data og skalerbarhed. Som svar på disse udfordringer er der opstået Data Lakes), som tilbyder fleksibel lagring af

store mængder heterogene data.

> **Data Lake** tilbyder en alternativ DWH -tilgang, der gør det muligt at arbejde med ustrukturerede, halvstrukturerede og rå data uden et forudgående stift skema. Denne lagringsmetode er ofte relevant for databehandling i realtid, maskinlæring og avanceret analyse. I modsætning til DWH, som strukturerer og aggregerer data før indlæsning, gør Data Lake det muligt at lagre information i sin rå form, hvilket giver fleksibilitet og skalerbarhed

Det var frustrationen over traditionelle datalagre (RDBMS, DWH) og interessen for "big data", der førte til fremkomsten af datasøer, hvor data i stedet for kompleks ETL nu blot uploades til et løst struktureret depot, hvor behandlingen finder sted i analysestadiet:

- I traditionelle datalagre bliver data typisk forbehandlet, transformeret og renset (ETL - Extract, Transform, Load), før de indlæses i lageret (figur 8.1-6). Det betyder, at data struktureres og optimeres til specifikke fremtidige analyser og rapporteringsopgaver. Der lægges vægt på at opretholde en høj forespørgselsydelse og dataintegritet. Denne tilgang kan dog være dyr og mindre fleksibel med hensyn til at integrere nye datatyper og hurtigt skiftende dataskemaer.
- Datasøer er derimod designet til at lagre store mængder rådata i deres oprindelige format (figur 8.1-7). ETL-processen (Extract, Transform, Load), er ved at blive erstattet af ELT (Extract, Load, Transform), hvor data først indlæses i lageret "som de er" og først derefter kan transformeres og analyseres efter behov. Det giver større fleksibilitet og mulighed for at lagre heterogene data, herunder ustrukturerede data som tekst, billeder og logfiler.

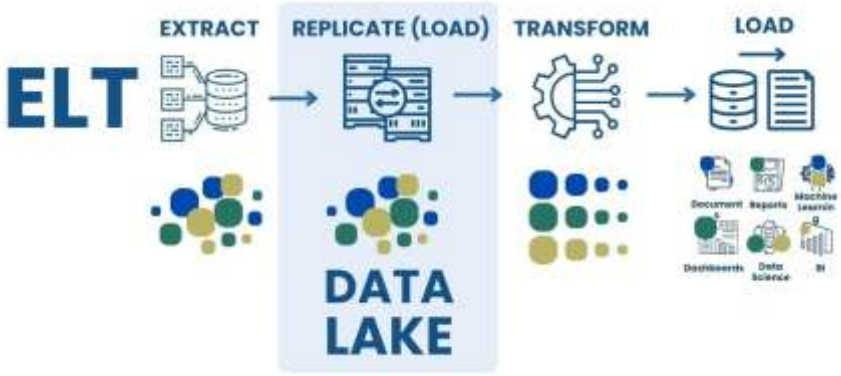


Fig. 8.1-7 I modsætning til ETL bruger Data Lake ELT, hvor information først uploades i "rå" form, og transformationen udføres i uploadfasen.

Traditionelle data warehouses fokuserer på forbehandling af data for at sikre en høj forespørgselsydelse, hvorimod data lakes prioriterer fleksibilitet: de lagrer rådata og transformerer dem efter behov (figur 8.1-8).

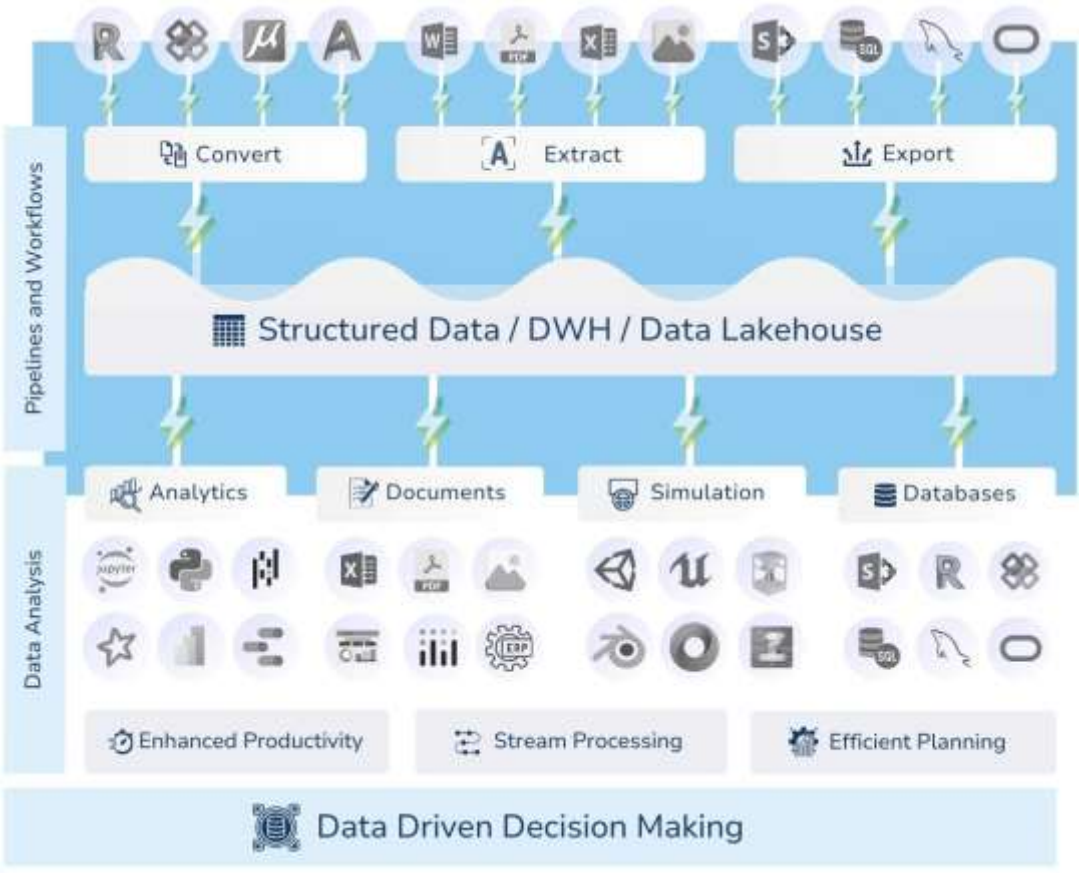


Fig. 8.1-8 Moderne lagringskoncepter sigter mod at lagre og behandle alle typer data med henblik på beslutningstagning.

Men på trods af alle fordelene er datasøer ikke uden ulemper. Manglen på en streng struktur og kompleksiteten i informationsstyringen kan føre til kaos, hvor data duplikeres, modsiger hinanden eller bliver irrelevante. Desuden kræver søgning og analyse af data i et sådant arkiv en betydelig indsats, især når man har at gøre med heterogene oplysninger. For at overvinde disse begrænsninger og kombinere de bedste funktioner i traditionelle datalagre og datasøer blev Data Lakehouse-arkitekturen udviklet.

Data Lakehouse-arkitektur: synergi mellem lagre og datasøer

For at kombinere de bedste egenskaber ved DWH (struktureret, håndterbar, højtydende analyse) og Data Lake (skalerbarhed, håndtering af heterogene data) blev Data Lakehouse-tilgangen udviklet. Denne arkitektur kombinerer fleksibiliteten i datasøer med de kraftfulde behandlings- og administrationsværktøjer, der er typiske for traditionelle lagre, og skaber balance mellem lagring, analyse og maskinlæring. Data Lakehouse er en syntese af datasøer og datalagre, der kombinerer førstnævntes fleksibilitet og skalerbarhed med sidstnævntes håndterbarhed og optimering af forespørgsler.

Data Lakehouse er en arkitektonisk tilgang, der forsøger at kombinere datasøernes fleksibilitet og skalerbarhed med datalagrenes håndterbarhed og forespørgselsydelse (figur 8.1-9).

De vigtigste funktioner i Data Lakehouse omfatter:

- **Åbent datalagringsformat:** Brug af åbne formater til datalagring, f.eks. Apache Parquet, giver effektivitet og optimerede forespørgsler.
- Skrivebeskyttet **skema: I** modsætning til den traditionelle tilgang med et skrivebeskyttet skema i DWH understøtter Lakehouse et skrivebeskyttet skema, som giver mere fleksibilitet i håndteringen af datastrukturen.
- **Fleksibel og skalerbar**: understøtter lagring og analyse af strukturerede og ustrukturerede data og giver høj forespørgselsydelse gennem optimering på lagerniveau.

Data Lakehouse tilbyder en kompromisløsning, der kombinerer fordelene ved begge tilgange, hvilket gør den ideel til moderne analysearbejde, der kræver fleksibilitet i databehandlingen.

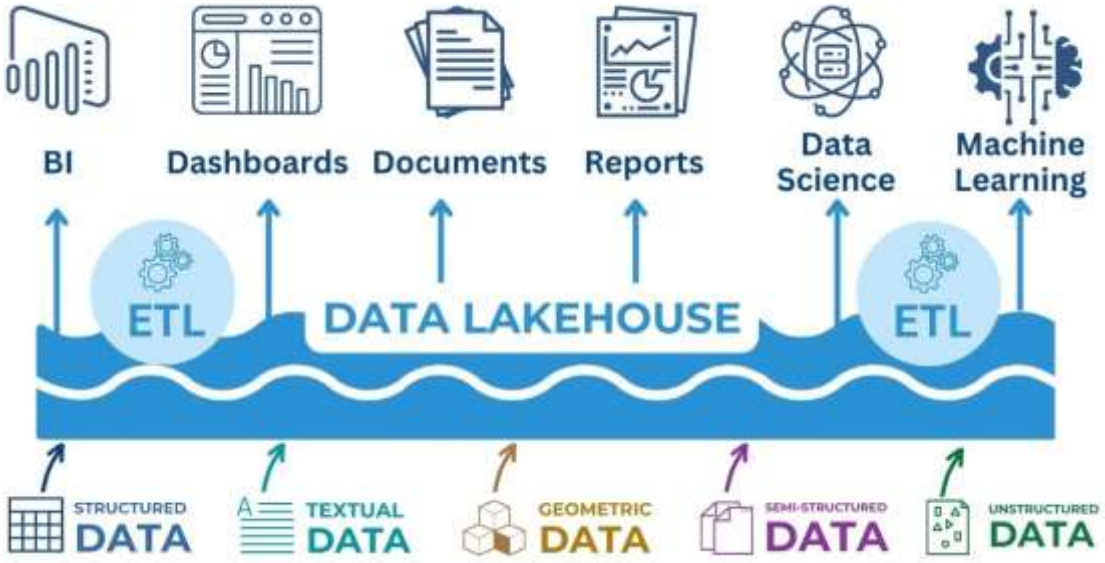


Fig. 8.1-9 Data Lakehouse er den næste generation af lagersystemer, der er designet til at opfylde komplekse og stadigt skiftende krav.

Ideen bag moderne datavarehuse virker enkel: Hvis alle data er samlet ét sted, er de lettere at analysere. Men i praksis er alt ikke så glat. Forestil dig, at en virksomhed beslutter sig for helt at opgive de sædvanlige regnskabs- og ledelsessystemer (ERP, PMIS, CAFM eller andre) og erstatte dem med én stor datasø, som alle har adgang til. Hvad vil der så ske? Højst sandsynligt vil der opstå kaos: data vil blive duplikeret, modstridende, og vigtige oplysninger vil gå tabt eller blive ødelagt. Selv hvis datasøen kun bruges til analyser, vil den blive alvorligt kompromitteret uden ordentlig styring:

- Data er svære at forstå: I konventionelle systemer har data en klar struktur, men i en sø er det bare en enorm ophobning af filer og tabeller. For at finde noget skal specialist finde ud af, hvad hver række og kolonne er ansvarlig for.
- Data kan være unøjagtige: Hvis mange versioner af de samme oplysninger er gemt ét sted, er det svært at vide, hvilken version der er opdateret. Resultatet er, at der træffes beslutninger baseret på forældede eller fejlagtige data.
- Det er svært at forberede data til arbejdet: Dataene skal ikke kun gemmes, men også præsenteres

i en praktisk form - i form af rapporter, grafer og tabeller. I traditionelle systemer sker dette automatisk, men i datasøer kræver det yderligere behandling.

Som følge heraf har hvert datalagerkoncept sine egne karakteristika, behandlingsmetoder og forretningsapplikationer. Traditionelle databaser fokuserer på transaktionelle operationer, datalagre (DWH) giver en struktur til analyse, datasøer (Data Lake) lagrer information i rå form, og hybridlagre (Data Lakehouse) kombinerer fordelene ved DWH og Data Lake (figur 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

Fig. 8.1-10 DWH, Data Lake og Data Lakehouse: vigtige forskelle i datatyper, brugsscenarier, behandlingsmetoder og lagringsmetoder.

At vælge en lagringsarkitektur er en kompleks proces, der afhænger af forretningsbehov, informationsmængde og analysekrav. Hver løsning har sine fordele og ulemper: DWH giver struktur, Data Lake giver fleksibilitet, og Lakehouse giver en balance mellem de to. Organisationer er sjældent begrænset til en enkelt dataarkitektur.

> Uanset den valgte arkitektur er automatiserede datahåndteringssystemer betydeligt bedre end manuelle metoder. De minimerer menneskelige fejl, fremskynder informationsbehandlingen og sikrer gennemsigtighed og sporbarhed af data i alle faser af forretningsprocesserne.

Og mens centraliserede datavarehuse allerede er blevet en industristandard i mange dele af økonomien, er situationen i byggeriet stadig fragmenteret. Data her er fordelt på forskellige platforme (CDE, PMIS, ERP osv.), hvilket gør det vanskeligt at skabe et samlet billede af, hvad der foregår, og kræver arkitekturer, der er i stand til at kombinere disse kilder til et holistisk, analytisk anvendeligt digitalt miljø.

CDE, PMIS, ERP eller DWH og Data Lake

Nogle bygge- og anlægsvirksomheder bruger allerede begrebet Common Data Environment (CDE) i henhold til ISO 19650. I bund og grund udfører CDE de samme funktioner som et data warehouse (DWH) i andre brancher: centralisering af information, versionskontrol og adgang til valideret information.

> **Et Common Data Environment (CDE**) er et centraliseret digitalt rum, der bruges til at administrere, lagre, dele og samarbejde om projektinformation i alle faser af et anlægs livscyklus. CDE implementeres ofte ved hjælp af cloud-baserede teknologier og integreres med CAD (BIM) -systemer.

Finans-, detail-, logistik- og industrisektoren har i årtier brugt centraliserede datastyringssystemer, der kombinerer oplysninger fra forskellige kilder, kontrollerer deres relevans og leverer analyser. CDE tager disse principper videre ved at tilpasse dem til udfordringerne ved bygningsdesign og livscyklusstyring.

Ligesom DWH strukturerer CDE data, registrerer ændringer og giver et enkelt adgangspunkt til verificerede oplysninger. Med flytningen til skyen og integrationen med analyseværktøjer bliver forskellene mellem de to mindre og mindre tydelige. Hvis man tilføjer granulære data til CDE, hvis koncept har været diskuteret af CAD -leverandører siden 2023[93, 125], kan man se endnu flere paralleller til klassisk DWH.

Tidligere i kapitlet "Byggeriets ERP og PMIS systemer" har vi allerede set på PMIS (Project Management Information System) og ERP (Enterprise Resource Planning). I byggeprojekter arbejder CDE og PMIS sammen: CDE fungerer som lager for data, herunder tegninger, modeller og projektdokumentation, mens PMIS styrer processer som f.eks. styring af deadlines, opgaver, ressourcer og budgetter.

ERP, der er ansvarlig for at styre virksomheden som helhed (økonomi, indkøb, personale, produktion), kan integreres med PMIS, der giver omkostnings- og budgetkontrol på virksomhedsniveau. Til analyse og rapportering kan DWH bruges til at indsamle, strukturere og aggregere data fra CDE, PMIS og ERP for at evaluere finansielle KPI'er (ROI) og identificere mønstre. Til gengæld kan Data Lake (DL) supplere DWH ved at lagre rå og ustrukturerede data (f.eks. logfiler, sensordata, billeder). Disse data kan behandles og indlæses i DWH til yderligere analyse.

> Således fokuserer CDE og PMIS på projektledelse, ERP fokuserer på forretningsprocesser, og DWH og Data Lake fokuserer på analyse og datadrift.

Når man sammenligner CDE, PMIS og ERP systemer med DWH og Data Lake, kan man se betydelige forskelle med hensyn til leverandøruafhængighed, omkostninger, integrationsfleksibilitet, dataafhængighed, hastighed for tilpasning til ændringer og analytiske muligheder (figur 8.1-11). Traditionelle systemer som CDE, PMIS og ERP er ofte bundet til specifikke leverandørløsninger og standarder, hvilket gør dem mindre fleksible og øger deres omkostninger på grund af licenser og support. Desuden er data i sådanne systemer ofte indkapslet i proprietære, lukkede formater, hvilket begrænser brugen og analysen af dem.

		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)

Fig. 8.1-11 DWH og Data Lake tilbyder større fleksibilitet og datauafhængighed end systemer som CDE, PMIS og ERP.

I modsætning hertil giver DWH og Data Lake større fleksibilitet i integrationen med forskellige datakilder, og deres brug af åbne teknologier og platforme er med til at reducere de samlede ejeromkostninger. Desuden understøtter DWH og Data Lake en bred vifte af analyseværktøjer, hvilket forbedrer analyse- og administrationsmulighederne.

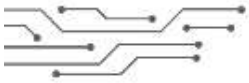
> Med udviklingen af reverse-engineering-værktøjer til CAD-formater og adgang til CAD-applikationsdatabaser bliver spørgsmålet mere og mere akut: Hvor berettiget er det at fortsætte med at bruge lukkede, isolerede platforme, hvis designdata skal være tilgængelige for en lang række specialister, der arbejder i dusinvis af entreprenører og designorganisationer?

Denne afhængighed af leverandørspecifik teknologi kan i høj grad begrænse fleksibiliteten i datahåndteringen, gøre det langsommere at reagere på projektændringer og hæmme et effektivt samarbejde mellem deltagerne.

Traditionelle tilgange til datahåndtering - herunder DWH, Data Lake, CDE og PMIS - har primært fokuseret på lagring, strukturering og behandling af information. Men med udviklingen af kunstig intelligens og maskinlæring er der et voksende behov for nye måder at organisere data på, som ikke bare samler, men også afslører komplekse relationer, finder skjulte mønstre og giver øjeblikkelig adgang til de mest relevante

oplysninger.

Vektordatabaser - en ny type lagring, der er optimeret til højdimensionelle indlejringer - er begyndt at spille en særlig rolle i denne retning.



KAPITEL 8.2.
STYRING AF DATALAGER OG FOREBYGGELSE AF KAOS

Vektordatabaser og den afgrænsende boks

Vektordatabaser er en ny klasse af arkiver, der ikke bare gemmer data, men gør det muligt at søge efter betydning, sammenligne objekter efter semantisk nærhed og skabe intelligente systemer: fra anbefalinger til automatisk analyse og generering af kontekst. I modsætning til traditionelle databaser, der fokuserer på nøjagtige matches, finder vektordatabaser lignende objekter baseret på attributter - også selvom der ikke er noget nøjagtigt match

> **En** vektordatabase er en specialiseret type database, der lagrer data som flerdimensionelle vektorer, der hver især repræsenterer bestemte egenskaber eller kvaliteter. Disse vektorer kan have forskellige antal dimensioner, afhængigt af dataenes kompleksitet (i ét tilfælde kan det være nogle få dimensioner og i et andet — tusinder).

Den største fordel ved vektordatabaser er søgning efter semantisk relevans snarere end efter nøjagtig matchning af værdier. I stedet for SQL- og Pandas -forespørgsler med "equals"- eller "contains"-filtre bruges søgning efter nærmeste naboer (k-NN) (vi vil tale mere om k-NN i næste del af bogen) i funktionsområdet.

Med udviklingen af LLM (Large Language Models) og generative modeller er interaktionen med databaser begyndt at ændre sig. Det er nu muligt at forespørge data på naturligt sprog, få semantiske søgninger på dokumenter, automatisk udtrække nøglebegreber og opbygge kontekstuelle relationer mellem objekter - alt sammen uden behov for færdigheder i SQL eller viden om tabelstruktur. Dette blev diskuteret mere detaljeret i afsnittet "LLM'er og deres rolle i databehandling og forretningsprocesser".

Det er dog vigtigt at indse, at LLM'er ikke automatisk strukturerer og ordner information. Modellen flyder bare gennem dataene og finder det mest relevante stykke data baseret på konteksten for forespørgslen. Hvis dataene ikke er blevet renset eller omdannet på forhånd, vil dybdesøgning være som at forsøge at finde et svar i digitalt "affald" - det kan virke, men kvaliteten af resultaterne vil være lavere. Det ideelle er, hvis dataene kan struktureres (f.eks. ved at oversætte dokumenter til Markdown) og indlæses i en vektordatabase. Det øger nøjagtigheden og relevansen af resultatet betydeligt.

> Oprindeligt blev vektordatabaser brugt i maskinlæring, men i dag finder de flere og flere anvendelser uden for det - i søgemaskiner, personalisering af indhold og intelligent analyse.

Et af de mest indlysende eksempler på vektortilgangen i konstruktionen er Bounding Box (afgrænsende parallelepipedium). Det er en geometrisk konstruktion, der beskriver grænserne for et objekt i det tredimensionelle rum. Bounding Box defineres af de minimale og maksimale X-, Y- og Z-koordinater, der danner en "kasse" omkring objektet. Denne metode giver dig mulighed for at estimere størrelsen og placeringen af et element uden at skulle analysere hele geometrien.

Hver Bounding Box kan repræsenteres som en vektor i et flerdimensionelt rum: for eksempel [x, y, z, bredde, højde, dybde] - allerede 6 dimensioner (Fig. 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

Fig. 8.2-1 Bounding Box -elementernes koordinatoplysninger og deres placering i projektmodellen svarer til en vektordatabase.

Denne datarepræsentation letter mange opgaver, herunder kontrol af skæringspunkter mellem objekter, planlægning af den rumlige fordeling af bygningselementer og udførelse af automatiserede beregninger. Bounding Box kan fungere som en bro mellem komplekse 3D-modeller og traditionelle vektordatabaser, så du effektivt kan bruge fordelene ved begge tilgange i arkitektonisk og teknisk modellering

> Bounding Box er "vektorisering af geometri", og indlejringen måde at transformere noget abstrakt på) er "vektorisering af betydning". Begge tilgange gør det muligt at gå fra manuel søgning til intelligent søgning, hvad enten det drejer sig om 3D -objekter i en projektmodel eller begreber i en tekst.

Søgning efter objekter i projektet (f.eks. "find alle vinduer med bredde > 1,5 m") svarer til søgning efter nærmeste naboer (k-NN) i en vektordatabase, hvor kriterierne definerer en "zone" i funktionsrummet. (Vi vil tale mere om k-NN-søgning efter nærmeste naboer i den næste del om maskinlæring) (Fig. 8.2-2). Hvis vi tilføjer yderligere parametre (materiale, vægt, produktionstid) til bounding box-attributterne, bliver tabellen til en højdimensionel vektor, hvor hver attribut er en ny dimension. Dette er tættere på moderne vektorbaser, hvor dimensionerne tælles i hundreder eller tusinder (f.eks. indlejring fra neurale netværk).

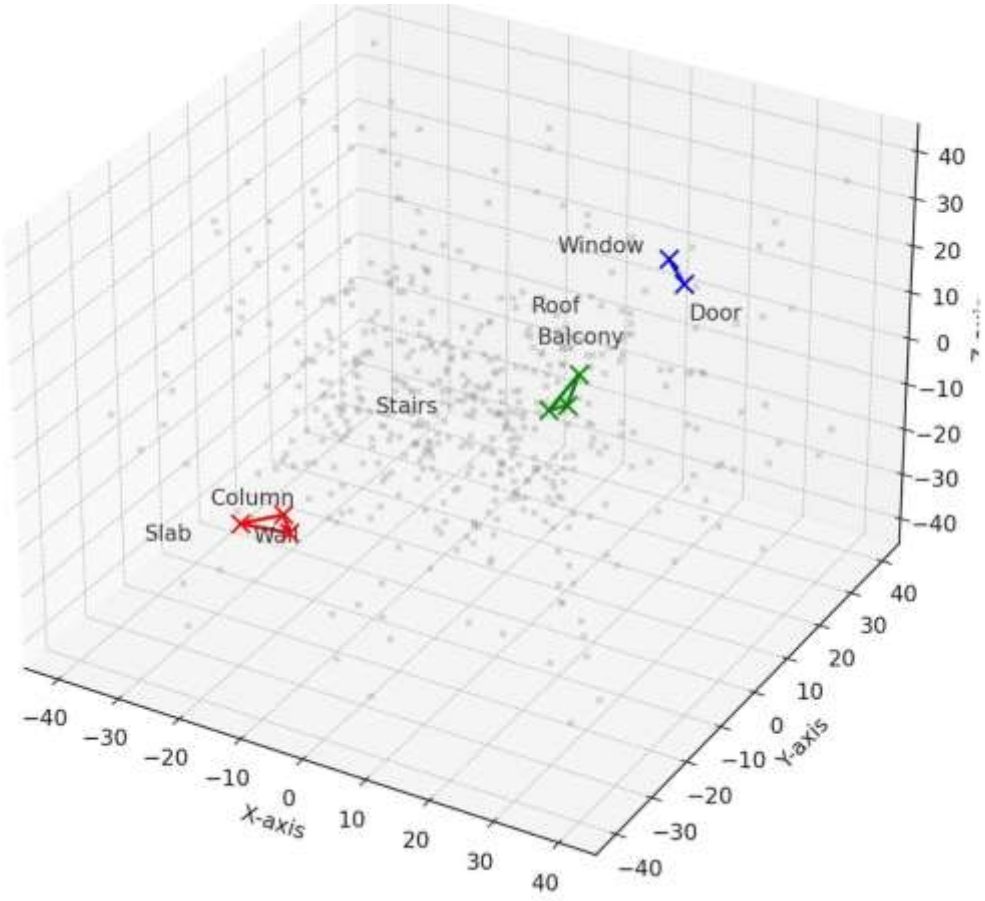


Fig. 8.2-2 Søg efter objekter i projektet ved hjælp af vektordatabaser.

Den tilgang, der anvendes i Bounding Box, kan ikke kun anvendes til geometriske objekter, men også til tekst- og sproganalyse. Vektorrepræsentationer af data bruges allerede aktivt i naturlig sprogbehandling (NLP). Ligesom objekter i et byggeprojekt kan grupperes i henhold til deres rumlige nærhed (figur 8.2-2), kan ord i en tekst analyseres i henhold til deres semantiske og kontekstuelle nærhed.

For eksempel vil ordene "arkitekt", "konstruktion" og "design" være ved siden af hinanden i vektorrummet, fordi de har en lignende betydning. I LLM giver denne mekanisme mulighed for automatisk, ikke manuel kategorisering:

- Identificere emnet for en tekst
- Udfør semantiske søgninger på indholdet af dokumenter
- Generer automatiske anmærkninger og tekstresuméer
- Find synonymer og relaterede termer

Vektordatabaser giver dig mulighed for at analysere tekst og finde relaterede termer i den på samme måde, som Bounding Box hjælper dig med at analysere rumlige objekter i 3D -modeller. Bounding Box-eksemplet med projektelementer hjælper med at forstå, at vektorrepræsentation ikke er et rent "kunstigt" koncept fra ML, men en naturlig måde at strukturere data på til løsning af anvendte problemer, uanset om det er søgning efter kolonner i et CAD-projekt eller semantisk tætte billeder i en database.

> Specialister, der arbejder med databaser, bør være opmærksomme på vektorlagre. Deres udbredelse indikerer en ny fase i databaseudviklingen, hvor klassiske relationelle systemer og AI -orienterede teknologier begynder at flette sig ind i hinanden og danne fremtidens hybridløsninger.

Brugere, der udvikler komplekse og store AI -applikationer, vil bruge specialiserede databaser til vektorsøgning. Samtidig er det mere sandsynligt, at de, der kun har brug for separate AI-funktioner til integration i eksisterende applikationer, vælger indbyggede vektorsøgningsfunktioner i de databaser, de allerede bruger (PostgreSQL, Redis).

Selvom systemer som DWH, Data Lake, CDE, PMIS, vektor-databaser og andre tilbyder forskellige tilgange til datalagring og -styring, bestemmes deres effektivitet ikke kun af deres arkitektur, men også af, hvor godt selve dataene er organiseret og styret. Selv når man bruger moderne løsninger - det være sig vektordatabaser, klassiske relationelle DBMS'er eller warehouses af Data Lake-typen - kan manglen på klare regler for håndtering, strukturering og opdatering af data føre til de samme vanskeligheder, som brugere, der arbejder med forskellige filer og data i flere formater, står over for.

Uden Data Governance) kan selv de mest kraftfulde løsninger blive kaotiske og ustrukturerede og forvandle datasøer til datasumpe). For at undgå dette skal virksomheder ikke kun vælge den rigtige lagringsarkitektur, men også implementere dataminimalisme), adgangsstyring og kvalitetskontrolstrategier for at gøre data til et effektivt værktøj til beslutningstagning.

Data Governance), Data Minimalism) og Data Swamp)

Forståelse og implementering af begreberne Data Governance), Dataminimalisme) og forebyggelse af Datasump) er nøglen til succesfuld styring af datalagre og levering af forretningsværdi (figur 8.2-3).

> Ifølge en undersøgelse foretaget af Gartner (2017) mislykkes 85% af big data-projekter, og en af hovedårsagerne er utilstrækkelig datakvalitet og datastyring [144].

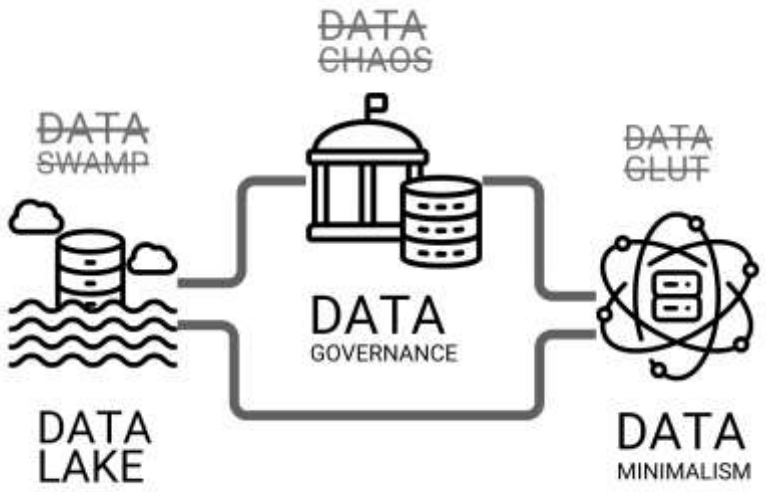


Fig. 8.2-3 Nogle af de vigtigste aspekter af datastyring er Data Governance og Dataminimalisme.

Data Governance (Data Governance) er en grundlæggende komponent i datastyring, der sikrer, at data bruges hensigtsmæssigt og effektivt i alle forretningsprocesser. Det handler ikke kun om at etablere regler og procedurer, men også om at sikre tilgængeligheden, pålideligheden og sikkerheden af data:

- Definere og klassificere data: En klar definition og klassificering af enheder gør det muligt for organisationer at forstå, hvilke enheder der er brug for i virksomheden, og bestemme, hvordan de skal bruges.
- Adgangsrettigheder og -styring: Udvikling af politikker og procedurer for adgang til og styring af data sikrer, at kun autoriserede brugere kan få adgang til bestemte data.
- Beskyttelse af data mod eksterne trusler: Beskyttelse af data mod eksterne trusler er et centralt aspekt af datahåndtering. Det omfatter ikke kun tekniske foranstaltninger, men også uddannelse af medarbejdere i de grundlæggende principper for informationssikkerhed.

Dataminimalisme (Data Minimalism) er en tilgang til at reducere data til de mest værdifulde og meningsfulde attributter og enheder i formationen (figur 8.2-4) og derved reducere omkostningerne og forbedre dataudnyttelsen:

- Forenkling af beslutningstagning: Ved at reducere antallet af objekter og deres egenskaber til de mest relevante forenkles beslutningstagningen ved at reducere den tid og de ressourcer, der kræves for at analysere og behandle data.
- Fokus på det vigtige: Ved at udvælge de mest relevante enheder og attributter kan du fokusere på de oplysninger, der virkelig betyder noget for virksomheden, og eliminere støj og unødvendige data.
- Effektiv ressourceallokering: Dataminimering muliggør en mere effektiv ressourceallokering, reducerer omkostningerne til datalagring og -behandling og forbedrer datakvaliteten og -sikkerheden.

Logikken i arbejdet med data bør ikke starte med oprettelsen af dem som sådan (Fig. 8.2-4), men med en forståelse af fremtidige scenarier for brug af disse data, selv før genereringsprocessen starter. Denne tilgang gør det muligt på forhånd at definere de nødvendige minimumskrav til attributter, deres typer og grænseværdier. Disse krav danner grundlag for at skabe korrekte og stabile enheder i

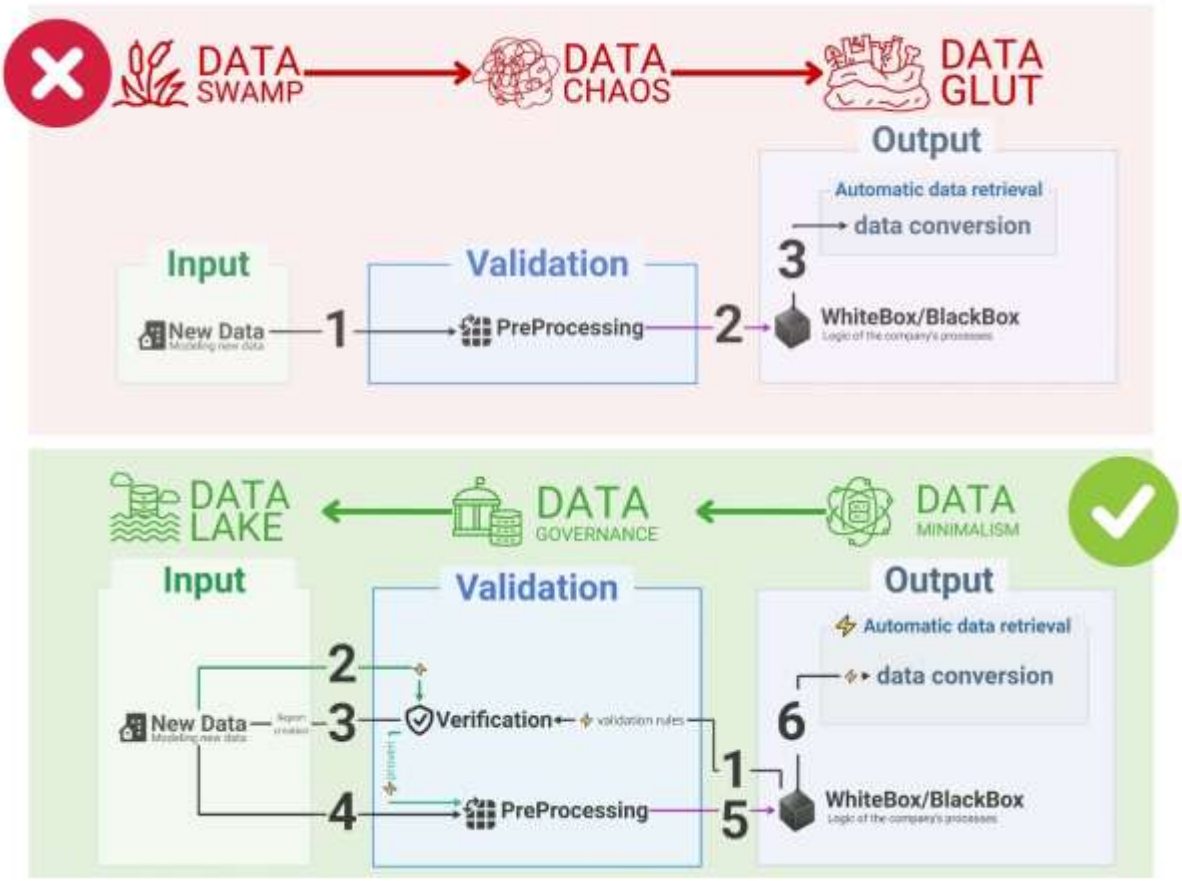
informationsmodellen. En foreløbig forståelse af dataenes formål og anvendelse bidrager til dannelsen af en struktur, der er egnet til analyse. Flere detaljer om tilgange til datamodellering på det konceptuelle, logiske og fysiske niveau blev diskuteret i kapitlet "Datamodellering: konceptuel, logisk og fysisk model".

> I traditionelle forretningsprocesser i byggefirmaer ligner databehandling oftere at dumpe data i en sump, hvor data først skabes, og derefter forsøger specialister at integrere dem i andre systemer og værktøjer.

Datasump (Data Swamp) er resultatet af ukontrolleret indsamling og lagring af data uden ordentlig organisering, strukturering og styring, hvilket resulterer i data, der er ustrukturerede, vanskelige at bruge og af ringe værdi.

Hvordan man forhindrer informationsstrømmen i at blive et morads:

- **Styring af datastruktur**: Ved at sikre, at data er struktureret og kategoriseret, kan man forhindre, at data drukner, ved at gøre dem overskuelige og let tilgængelige.
- **Forståelse og fortolkning af** data: En klar beskrivelse af dataenes oprindelse, ændringer og betydning sikrer, at data forstås og fortolkes korrekt.
- **Opretholdelse af datakvalitet**: Regelmæssig vedligeholdelse og rensning af data hjælper med at opretholde datakvalitet, relevans og værdi for analyser og forretningsprocesser.



Figur 8.2-4 For at undgå rod i datalageret bør du starte dataoprettelsesprocessen med at indsamle krav til attributter.

> Ved at integrere principperne om datastyring og dataminimalisme i datahåndteringsprocesserne og aktivt forhindre, at datalagre bliver til datasumpe, kan organisationer maksimere potentialet i deres data.

Det næste trin i udviklingen af arbejdet med data, efter at man har løst problemerne med styring og minimalisme, er standardisering af automatisk behandling, kvalitetssikring og implementering af metoder, der gør data anvendelige til analyse, transformation og beslutningstagning. Det er, hvad DataOps og VectorOps gør, og de er ved at blive vigtige værktøjer for virksomheder, der arbejder med big data og maskinlæring.

DataOps og VectorOps: nye datastandarder

Mens Data Governance er ansvarlig for at kontrollere og organisere data, hjælper DataOps med at sikre deres nøjagtighed, konsistens og smidige flow i virksomheden. Dette er især kritisk for en række forretningsområder i byggeriet, hvor data genereres løbende og kræver rettidig behandling. For eksempel i situationer, hvor bygningsinformationsmodeller, projektkrav og analytiske rapporter skal synkroniseres mellem forskellige systemer inden for en enkelt arbejdsdag, kan DataOps' rolle være afgørende. Det giver

dig mulighed for at opbygge stabile og gentagelige databehandlingsprocesser, hvilket reducerer risikoen for forsinkelser og tab af informationens relevans.

Data Governance alene er ikke nok - det er vigtigt, at data ikke bare opbevares, men bruges aktivt i den daglige drift. Det er her, DataOps - en metode med fokus på automatisering, integration og kontinuerligt dataflow - kommer ind i billedet.

> **DataOps** fokuserer på at forbedre samarbejde, integration og automatisering af datastrømme i organisationer. Vedtagelse af DataOps-praksis fremmer datanøjagtighed, -konsistens og -tilgængelighed, hvilket er afgørende for datacentrerede applikationer.

Nøgleværktøjer i DataOps økosystem er Apache Airflow (figur 7.4-4) til orkestrering af arbejdsgange og Apache NiFi (figur 7.4-5) til routing og transformering af datastrømme. Sammen muliggør disse teknologier fleksible, pålidelige og skalerbare datapipelines til automatisk behandling, kontrol og integration af information mellem systemer flere detaljer i kapitlet "Automatisk ETL -conveyor "). Når man implementerer DataOps-tilgangen i byggeprocesser, er det vigtigt at overveje fire grundlæggende aspekter:

1. **Mennesker og værktøjer er vigtigere** end data: siloopdelte datalagre kan ses som et stort problem, men virkeligheden er mere kompleks. Ud over datafragmentering spiller isolationen af teams og de forskellige værktøjer, de bruger, en væsentlig rolle. I byggeriet arbejder specialister fra forskellige discipliner med data: dataingeniører og analytikere, BI- og visualiseringsteams samt projektledelses- og kvalitetseksperter. Hver af dem har forskellige arbejdsmetoder, så det er vigtigt at skabe et økosystem, hvor data flyder frit mellem deltagerne og giver en enkelt, ensartet version af informationerne.
2. **Automatiser test og fejlsøgning**: Byggedata indeholder altid fejl, hvad enten det er unøjagtigheder i modeller, beregningsfejl eller forældede specifikationer. Regelmæssig testning af data og eliminering af tilbagevendende fejl kan forbedre datakvaliteten betydeligt. Som en del af DataOps skal du implementere automatiserede kontrol- og valideringsmekanismer, der overvåger dataenes korrekthed, analyserer fejl og identificerer mønstre samt fanger og håndterer systemfejl i alle arbejdsgange. Jo højere grad af automatiseret validering, jo højere er den samlede datakvalitet, og jo lavere er sandsynligheden for fejl i de sidste faser.
3. **Data bør testes på samme måde som programkode**: De fleste bygningsapplikationer er baseret på databehandling, men kontrollen af dem overlades ofte til sekundære roller. Hvis maskinlæringsmodeller trænes på unøjagtige data, fører det til forkerte forudsigelser og økonomiske tab. Inden for DataOps bør data underkastes samme strenge kontrol som softwarekode: logiske kontroller, stresstest, evaluering af modellernes adfærd, når inputværdierne ændres. Kun validerede og pålidelige data kan bruges som grundlag for ledelsesbeslutninger.
4. **Dataobservabilitet uden at gå på kompromis med ydeevnen**: Dataovervågning er ikke bare en samling af målinger, men et strategisk kvalitetsstyringsværktøj. For at DataOps kan fungere effektivt, skal der indbygges observerbarhed i alle faser af datahåndteringen, fra design til drift. Samtidig er det vigtigt, at overvågningen ikke gør systemet langsommere. I byggeprojekter er det afgørende ikke kun at indsamle data, men også at gøre det på en sådan måde, at arbejdet for de fagfolk (f.eks. designere), der skaber dataene, ikke forstyrres på nogen måde. Denne balance giver dig mulighed for at kontrollere datakvaliteten uden at gå på kompromis med produktiviteten.

DataOps er ikke en ekstra byrde for dataforskere, men rygraden i deres arbejde. Ved at implementere DataOps kan byggevirksomheder gå fra kaotisk datahåndtering til et effektivt økosystem, hvor data arbejder

for virksomheden.

Til gengæld repræsenterer VectorOpsnæste fase i udviklingen af DataOps, der fokuserer på behandling, lagring og analyse af multidimensionelle vektordata (som blev diskuteret i det foregående kapitel). Det er især relevant inden for områder som digitale tvillinger, neurale netværksmodeller og semantisk søgning, som er begyndt at komme ind i byggebranchen. VectorOps er afhængig af vektordatabaser for effektivt at kunne lagre, indeksere og søge i flerdimensionelle repræsentationer af objekter.

> **VectorOps** er det næste skridt efter DataOps, der fokuserer på behandling, analyse og brug af vektordata i byggeriet. I modsætning til DataOps, som fokuserer på dataflow, konsistens og kvalitet, fokuserer VectorOps på at håndtere de multidimensionelle objektrepræsentationer, der kræves til maskinlæring.

I modsætning til traditionelle tilgange giver VectorOps dig mulighed for at opnå mere nøjagtige objektbeskrivelser, hvilket er afgørende for digitale tvillinger, generative designsystemer og automatisk fejlregistrering i CAD-data, der er konverteret til vektorformat. Den kombinerede implementering af DataOps og VectorOps danner et solidt fundament for skalerbart, automatiseret arbejde med store mængder information - fra klassiske tabeller til semantisk rige rumlige modeller

Næste skridt: fra kaotisk opbevaring til struktureret opbevaring

Traditionelle tilgange til opbygning af datalagring resulterer ofte i oprettelse af forskellige "informationssiloer", hvor vigtige indsigter er utilgængelige for analyse og beslutningstagning. Moderne lagringskoncepter, såsom Data Warehouse, Data Lake og deres hybrider, gør det muligt at forene forskellig information og gøre den tilgængelig på en centraliseret måde til datastreaming og business intelligence. Det er ikke kun vigtigt at vælge den rigtige storage-arkitektur, men også at implementere Data Governance) og Data Minimalism) for at forhindre, at storage-faciliteter bliver ukontrollerbare Data Swamps).

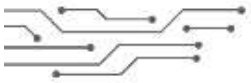
For at opsummere denne del er det værd at fremhæve de vigtigste praktiske trin, der vil hjælpe dig med at anvende de diskuterede koncepter på dine daglige opgaver:

- [ ] Vælg effektive datalagringsformater
  - [ ] Gå væk fra CSV og XLSX til mere effektive formater (Apache Parquet, ORC) til lagring af store mængder data.
  - [ ] Implementer et dataversioneringssystem til at spore ændringer
  - [ ] Brug metadata til at beskrive informationens struktur og oprindelse
- [ ] Skab en samlet dataarkitektur for virksomheden
  - [ ] Sammenlign forskellige storage-arkitekturer: RDBMS, DWH og Data Lake. Vælg den, der bedst opfylder dine behov for skalerbarhed, kildeintegration og analytisk behandling.
  - [ ] Design et proceskort til at udtrække, indlæse og transformere data (ETL) fra forskellige kilder til dine opgaver. Brug visualiseringsværktøjer som Miro, Lucidchart eller Draw.io til at visualisere vigtige trin og integrationspunkter.
- [ ] Implementer Data Governance-praksisser og Dataminimalisme

- [ ] Følg den dataminimalistiske tilgang - gem og behandl kun det, der virkelig er værdifuldt
- [ ] Implementer Data Governance-principper - definer ansvar for data, sørg for kvalitet og gennemsigtighed
- [ ] Få mere at vide om datastyringspolitikker og DataOps-koncepter, VectorOps
- [ ] Definer datakvalitetskriterier og procedurer for datavalidering inden for DataOps

Velorganiseret datalagring skaber grundlaget for centralisering af en virksomheds analytiske processer. Overgangen fra kaotisk ophobning af filer til struktureret lagring gør det muligt at gøre information til et strategisk aktiv, der hjælper med at træffe informerede beslutninger og forbedre effektiviteten af forretningsprocesser.

Når processerne for dataindsamling, -transformation, -analyse og struktureret lagring er blevet automatiseret og standardiseret, er den næste fase af den digitale transformation den fulde håndtering af Big Data.





IX DEL
BIG DATA, MASKINLÆRING OG FORUDSIGELSER

Den niende del fokuserer på big data, maskinlæring og prædiktiv analyse i byggebranchen. Den udforsker overgangen fra intuitiv beslutningstagning til objektive analyser baseret på historiske data. Praktiske eksempler bruges til at demonstrere big data-analyse i byggeriet - fra analyse af San Franciscos byggetilladelsesdatasæt til behandling af CAD - projekter med millioner af elementer. Der lægges særlig vægt på maskinlæringsmetoder til forudsigelse af byggeprojekters omkostninger og tidsplan med en detaljeret diskussion af lineær regression og k-nearest neighbours-algoritmer. Det vises, hvordan strukturerede data bliver grundlaget for forudsigelige modeller til vurdering af risici, optimering af ressourcer og forbedring af projektledelsens effektivitet. Delen giver også anbefalinger til udvælgelse af repræsentative dataprøver og forklarer, hvorfor store datasæt ikke altid er nødvendige for effektive analyser.

KAPITEL 9.1.
BIG DATA OG ANALYSE HERAF

Big data i byggeriet: fra intuition til forudsigelighed

Udtrykket "big data" har ikke en fast definition. Begrebet dukkede oprindeligt op, da informationsmængden begyndte at overstige kapaciteten i traditionelle metoder til behandling af den. I dag er mængden og kompleksiteten af data i mange brancher, herunder byggeriet, steget så meget, at de ikke passer ind i computernes lokale hukommelse og kræver brug af nye teknologier til at behandle dem.

Essensen af at arbejde med Big Data er ikke kun lagring og behandling, men også forudsigelsesmuligheder. I byggebranchen baner Big Data vejen fra intuitive beslutninger baseret på subjektiv fortolkning af tabeller og visualiseringer (som diskuteret tidligere) til informerede prognoser understøttet af reelle observationer og statistikker.

I modsætning til hvad mange tror, er målet med at arbejde med big data ikke at "få en maskine til at tænke som et menneske", men at anvende matematiske modeller og algoritmer til at analysere enorme datamængder for at identificere mønstre, forudsige begivenheder og optimere processer.

> **Big Data** er ikke en kold verden af algoritmer uden menneskelig indflydelse. Tværtimod arbejder big data sammen med vores instinkter, fejltagelser og kreativitet. Det er den menneskelige tænknings ufuldkommenhed, der giver os mulighed for at finde ikke-standardiserede løsninger og skabe gennembrud.

Med udviklingen af digital teknologi er byggebranchen begyndt aktivt at bruge databehandlingsteknikker, der kommer fra IT-sektoren. Takket være værktøjer som Pandas og Apache Parquet kan strukturerede og ustrukturerede data kombineres, hvilket forenkler adgangen til information og reducerer tabet ved analyse, mens store datasæt fra dokumenter eller CAD-projekter (fig. 9.2-10 - fig. 9.2-12) gør det muligt at indsamle, analysere og forudsige data i alle faser af projektets livscyklus.

Big Data har en transformerende indvirkning på byggebranchen og påvirker den potentielt på en række forskellige måder. Anvendelsen af Big Data-teknologier giver resultater på en række nøgleområder, herunder f.eks. følgende:

- **Analyse af investeringspotentiale** - forudsigelse af rentabilitet og tilbagebetalingsperioder for projekter baseret på data fra tidligere anlæg.
- **Forudseende vedligeholdelse** - identificering af sandsynlige udstyrsfejl, før de rent faktisk opstår, hvilket reducerer nedetid.
- **Optimering af forsyningskæden** - forudsigelse af afbrydelser og forbedring af logistikeffektiviteten.
- **Energieffektivitetsanalyser** - hjælper med at designe lavenergibygninger.
- **Sikkerhedsovervågning** - brug af sensorer og bærbare enheder til at overvåge forholdene på stedet.
- **Kvalitetskontrol** - realtidsovervågning af overholdelse af processtandarder.
- **Personalestyring** - resultatanalyse og prognoser for personalebehov.

Det er svært at finde et område inden for byggeriet, hvor dataanalyse og forudsigelser ikke er efterspurgt. Den største fordel ved forudsigelsesalgoritmer er deres evne til selv at lære og konstant forbedre sig, efterhånden som data akkumuleres.

> I den nærmeste fremtid vil kunstig intelligens ikke bare hjælpe bygherrer, men træffe vigtige beslutninger - fra designprocesser til bygningsdrift.

Mere om, hvordan forudsigelser genereres og læringsmodeller bruges, vil blive diskuteret i næste del af bogen, "Maskinlæring og forudsigelser".

Overgangen til fuldgyldigt arbejde med big data kræver en ændring i selve tilgangen til analyse. Mens de klassiske systemer, vi har set på indtil nu, fokuserede på årsagssammenhænge, skifter big data-analysen fokus til at søge efter statistiske mønstre og sammenhænge, som gør det muligt at identificere skjulte sammenhænge og forudsige objekters adfærd, selv uden at have en fuld forståelse af alle faktorer.

Spørgsmål om gennemførligheden af big data: korrelation, statistik og datasampling

Traditionelt var byggeri baseret på subjektive hypoteser og personlige erfaringer. Ingeniører antog - med en vis grad af sandsynlighed - hvordan materialet ville opføre sig, hvilke belastninger konstruktionen ville modstå, og hvor længe projektet ville vare. Disse antagelser blev testet i praksis, ofte på bekostning af tid, ressourcer og fremtidige risici.

> Med fremkomsten af big data ændrer tilgangen sig dramatisk: Beslutninger træffes ikke længere på baggrund af intuitive fornemmelser, men som et resultat af analyse af store datasæt. Byggeri holder gradvist op med at være en intuitiv kunst og bliver en præcis videnskab om forudsigelser.

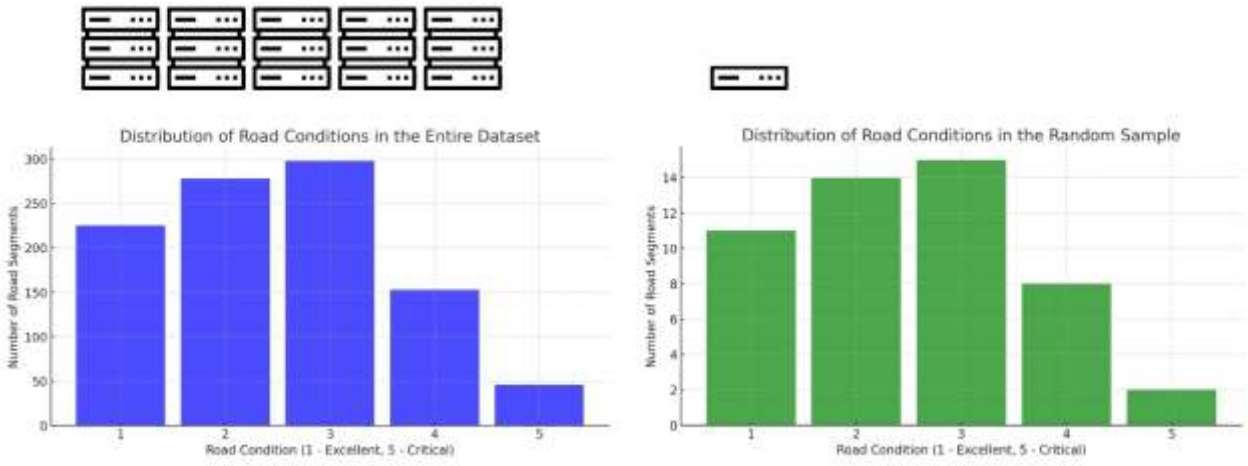
Overgangen til ideen om at bruge big data rejser uundgåeligt et vigtigt spørgsmål: Hvor kritisk er mængden af data, og hvor meget information er der egentlig brug for til pålidelig forudsigende analyse? Den udbredte tro på, at "jo flere data, jo større nøjagtighed", viser sig ikke altid at være statistisk gyldig i praksis.

> Tilbage i 1934 beviste statistikeren Jerzy Neumann [145], at nøglen til nøjagtigheden af statistiske slutninger ikke så meget ligger i mængden af data som i deres repræsentativitet og tilfældighed i prøveudtagningen.

Det gælder især i byggebranchen, hvor store datamængder indsamles ved hjælp af IoT -sensorer, scannere, overvågningskameraer, droner og endda CAD -modeller i flere formater, hvilket øger risikoen for blinde vinkler, outliers og dataforvrængninger.

Lad os se på et eksempel med overvågning af vejoverfladens tilstand. Et komplet datasæt med alle vejstrækninger kan fylde X GB og tage omkring en dag at behandle. Samtidig vil en tilfældig stikprøve, der kun omfatter hver 50. vejstrækning, kun tage X/50 GB og blive behandlet på en halv time, samtidig med at

den giver samme nøjagtighed i estimaterne for visse beregninger (figur 9.1-1).



Figur 9.1-1 Histogrammer over belægningens tilstand: det fulde datasæt og tilfældige stikprøver viser identiske resultater.

Derfor er nøglen til en vellykket dataanalyse ofte ikke mængden af data, men stikprøvens repræsentativitet og kvaliteten af de anvendte behandlingsmetoder. Overgangen til tilfældig prøveudtagning og en mere selektiv tilgang kræver et skift i tankegangen i byggebranchen. Historisk set har virksomheder fulgt logikken "jo flere data, jo bedre" i den tro, at det ville maksimere nøjagtigheden at dække alle mulige indikatorer.

Denne tilgang minder om en populær misforståelse inden for projektledelse: "Jo flere specialister jeg tiltrækker, jo mere effektivt bliver arbejdet". Men som med menneskelige ressourcer er det kvalitet og værktøjer, der er vigtigere end kvantitet. Hvis man ikke tager højde for de indbyrdes relationer (korrelationer) mellem data eller projektdeltagere, kan øget volumen kun føre til støj, forvrængning, dobbeltarbejde og unødvendigt spild.

I sidste ende viser det sig ofte, at det er meget mere produktivt at have et mindre, men kvalitativt forberedt datasæt, der kan producere stabile og rimelige prognoser, end at stole på massiv, men kaotisk information, der indeholder mange modstridende signaler.

> For stor datamængde garanterer ikke kun større nøjagtighed, men kan også føre til forvrængede konklusioner på grund af tilstedeværelsen af støj, overflødige funktioner, skjulte sammenhænge og irrelevante oplysninger. Under sådanne omstændigheder øges risikoen for overtilpasning af modeller, og pålideligheden af analytiske resultater mindskes.

I byggeriet er en stor udfordring i forbindelse med big data at bestemme den optimale mængde og kvalitet af data. Når man f.eks. overvåger betonkonstruktioners tilstand, kan brugen af tusindvis af sensorer og indsamling af oplysninger hvert minut overvælde lagrings- og analysesystemet. Men hvis du udfører korrelationsanalyse og vælger de 10% mest informative sensorer, kan du få næsten samme nøjagtighed i forudsigelserne og bruge mange gange, nogle gange titusinder og hundreder af gange, færre ressourcer.

Ved at bruge en mindre delmængde af data reduceres både den nødvendige lagerplads og behandlingstiden, hvilket reducerer omkostningerne til lagring og analyse af data betydeligt og ofte gør stikprøver til en ideel løsning til prædiktiv analyse, især i store infrastrukturprojekter eller når der arbejdes i

realtid. I sidste ende bestemmes effektiviteten af byggeprocesser ikke af mængden af indsamlede data, men af kvaliteten af deres analyse. Uden en kritisk tilgang og omhyggelig analyse kan data føre til forkerte konklusioner.

> Efter en vis mængde data giver hver ny informationsenhed mindre og mindre brugbare resultater. I stedet for at indsamle oplysninger i en uendelighed er det vigtigt at fokusere på deres repræsentativitet og analysemetoder (figur 9.2-2).

Dette fænomen er godt beskrevet af Allen Wallis [146], som illustrerer brugen af statistiske metoder ved hjælp af et eksempel på test af to alternative projektildesigns i den amerikanske flåde.

> *Flåden testede to alternative projektildesigns (A og B) ved at gennemføre en række parvise runder. I hver runde får A 1 eller 0, afhængigt af om dets præstation er bedre eller dårligere end B's, og omvendt. Den statistiske standardtilgang indebærer, at man gennemfører et fast antal forsøg (f.eks. 1000) og bestemmer vinderen ud fra en procentfordeling (hvis A f.eks. får 1 i mere end 53% af tilfældene, betragtes den som den bedste). Da Allen Wallis diskuterede et sådant problem med (flåde)kaptajn Garrett L. Schuyler, indvendte kaptajnen, at en sådan test, for at citere Allens beretning, kunne være ubrugelig. Hvis en klog og erfaren artillerioffícer som Schuyler havde været på stedet, ville han efter de første par hundrede [skud] have set, at eksperimentet ikke behøvede at blive afbrudt, enten fordi den nye metode er klart ringere, eller fordi den er klart bedre end det, man havde håbet på [146].*
>
> - Den amerikanske regerings statistiske forskningsgruppe ved Columbia University, perioden under Anden Verdenskrig

Dette princip er meget udbredt i forskellige brancher. Inden for medicin udføres kliniske forsøg med nye lægemidler f.eks. på tilfældige stikprøver af patienter, hvilket gør det muligt at opnå statistisk signifikante resultater, uden at det er nødvendigt at teste lægemidlet på hele jordens befolkning. Inden for økonomi og sociologi gennemføres repræsentative undersøgelser for at afspejle samfundets holdning, uden at det er nødvendigt at interviewe alle i landet.

Ligesom regeringer og forskningsorganisationer undersøger små populationer for at forstå generelle sociale tendenser, kan virksomheder i byggebranchen bruge tilfældige datastikprøver til effektivt at overvåge og skabe prognoser for projektstyring (figur 9.1-1).

> *Big data kan ændre tilgangen til samfundsvidenskab, men det vil ikke erstatte statistisk sund fornuft [147].*
>
> - Thomas Landsall-Welfair, "Forecasting the nation's current mood", Significance v. 9(4), 2012 .

Ud fra et ressourcebesparende perspektiv er det ved indsamling af data til fremtidige forudsigelser og beslutningstagning vigtigt at besvare spørgsmålet: Giver det mening at bruge betydelige ressourcer på at

indsamle og behandle enorme datasæt, når der kan bruges et meget mindre og billigere testdatasæt, der kan opskaleres trinvist? Effektiviteten af tilfældig prøveudtagning viser, at virksomheder kan reducere omkostningerne til indsamling og træning af modeller med ti eller endda tusinder af gange ved at vælge dataindsamlingsmetoder, der ikke kræver omfattende dækning, men stadig giver tilstrækkelig nøjagtighed og repræsentativitet. Denne tilgang gør det muligt for selv små virksomheder at opnå resultater på niveau med store virksomheder ved hjælp af betydeligt færre ressourcer og datamængder, hvilket er vigtigt for virksomheder, der ønsker at optimere omkostningerne og fremskynde informeret beslutningstagning ved hjælp af små ressourcer. I de følgende kapitler udforskes eksempler på analytics og predictive analytics baseret på offentlige datasæt ved hjælp af big data-værktøjer.

Big data: analyse af data fra San Franciscos datasæt med en million byggetilladelser

Arbejdet med åbne datasæt giver en unik mulighed for at omsætte de principper, der er blevet diskuteret i de foregående kapitler, til praksis: fornuftig udvælgelse af funktioner, repræsentativ prøveudtagning, visualisering og kritisk analyse. I dette kapitel vil vi undersøge, hvordan komplekse fænomener som byggeaktivitet i en storby kan undersøges ved hjælp af åbne data - i særdeleshed over en million optegnelser over byggetilladelser i San Francisco.

Offentligt tilgængelige data om over en million byggetilladelser (figur 9.1-2) (registreringer i to datasæt i CSV-format) fra "San Francisco Department of Buildings" [148] giver os mulighed for at bruge den rå CSV - tabel til ikke kun at analysere byggeaktiviteten i byen,, men også til kritisk at analysere de seneste tendenser og historien om San Franciscos byggeindustri i løbet af de sidste 40 år, fra 1980 til 2019.

De kodeeksempler, der blev brugt til at skabe datasætvisualiseringerne (figurerne 9.1-3- figurerne 9.1-8), samt visuelle grafer med kode, forklaringer og kommentarer, kan findes på Kaggle-platformen ved at søge efter "San Francisco. Construction Sector 1980-2019." [149].

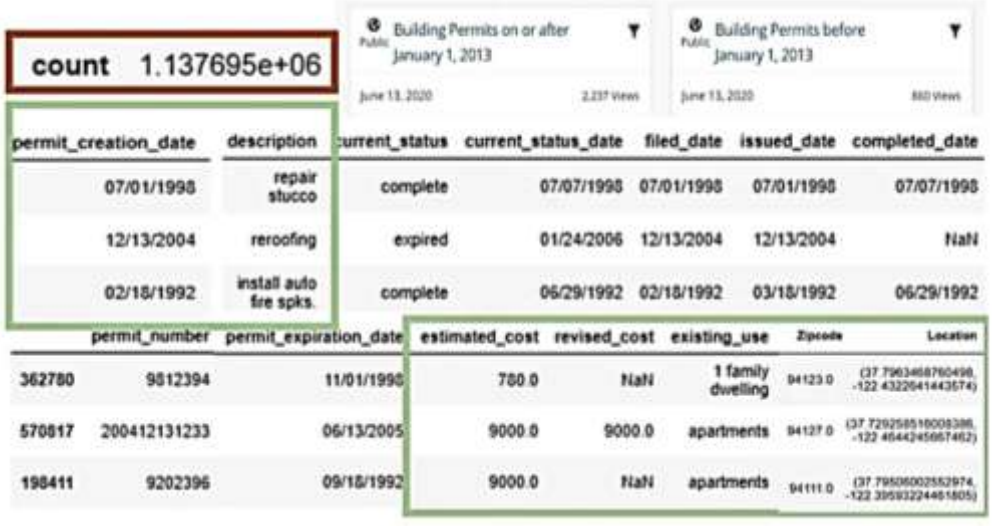


Fig. 9.1-2 Datasættene indeholder oplysninger om udstedte byggetilladelser med forskellige objektattributter.

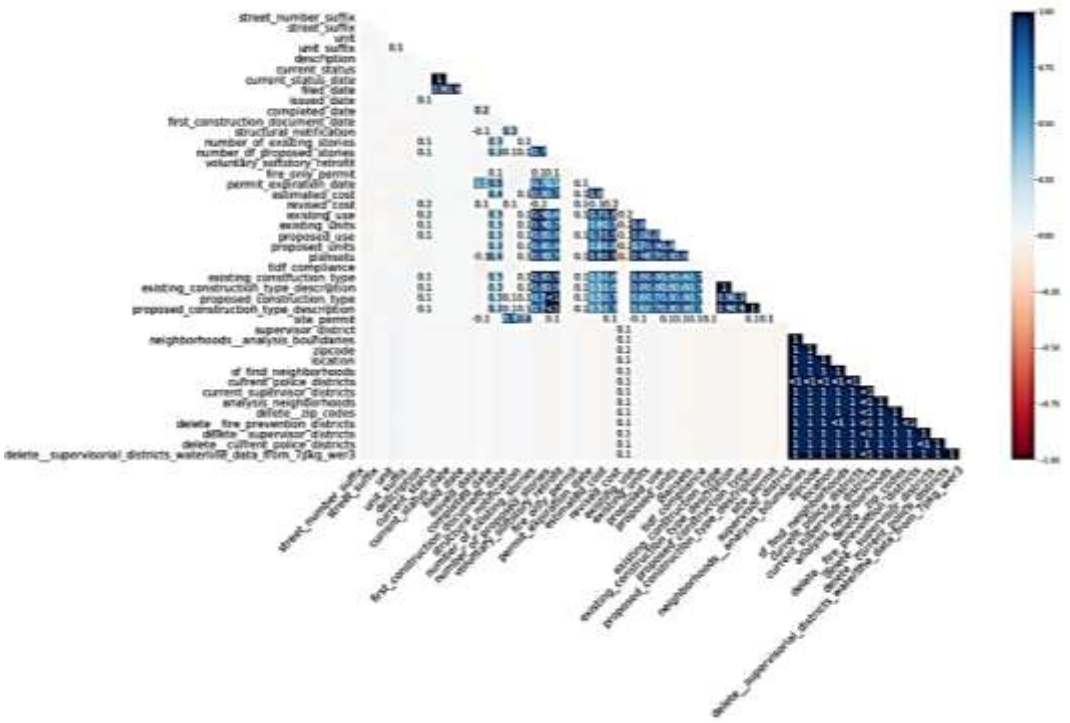


Fig. 9.1-3 Et varmekort (Pandas og Seaborn), der visualiserer alle attributter i et datasæt og hjælper med at identificere relationer mellem par af attributter.

Der fremgår ingen tendenser eller konklusioner af tabellen fra San Francisco Department of Buildings (figur 9.1-2). Tørre tal i tabelform er ikke et grundlag for beslutningstagning. For at gøre dataene visuelt forståelige skal de, som beskrevet i detaljer i kapitlerne om datavisualisering, visualiseres ved hjælp af de forskellige biblioteker, der omtales i bogens syvende del om emnet "ETL og visualisering af resultater som grafer".

Ved at analysere data ved hjælp af Pandas DataFrame og Python-visualiseringsbibliotekerneværdien af 1.137.695 tilladelser [148] kan vi konkludere, at byggeaktiviteten i San Francisco er tæt forbundet med økonomiske cyklusser, især i den blomstrende Silicon Valley-teknologiindustri (figur 9.1-4).

Økonomiske op- og nedture har en betydelig indvirkning på antallet og værdien af byggeprojekter. For eksempel faldt det første højdepunkt i byggeaktiviteten sammen med elektronikboomet i midten af 1980'erne (brugt Pandas og Matplotlib), og efterfølgende højdepunkter og fald var forbundet med dot-com-boblen og de seneste års teknologiboom.

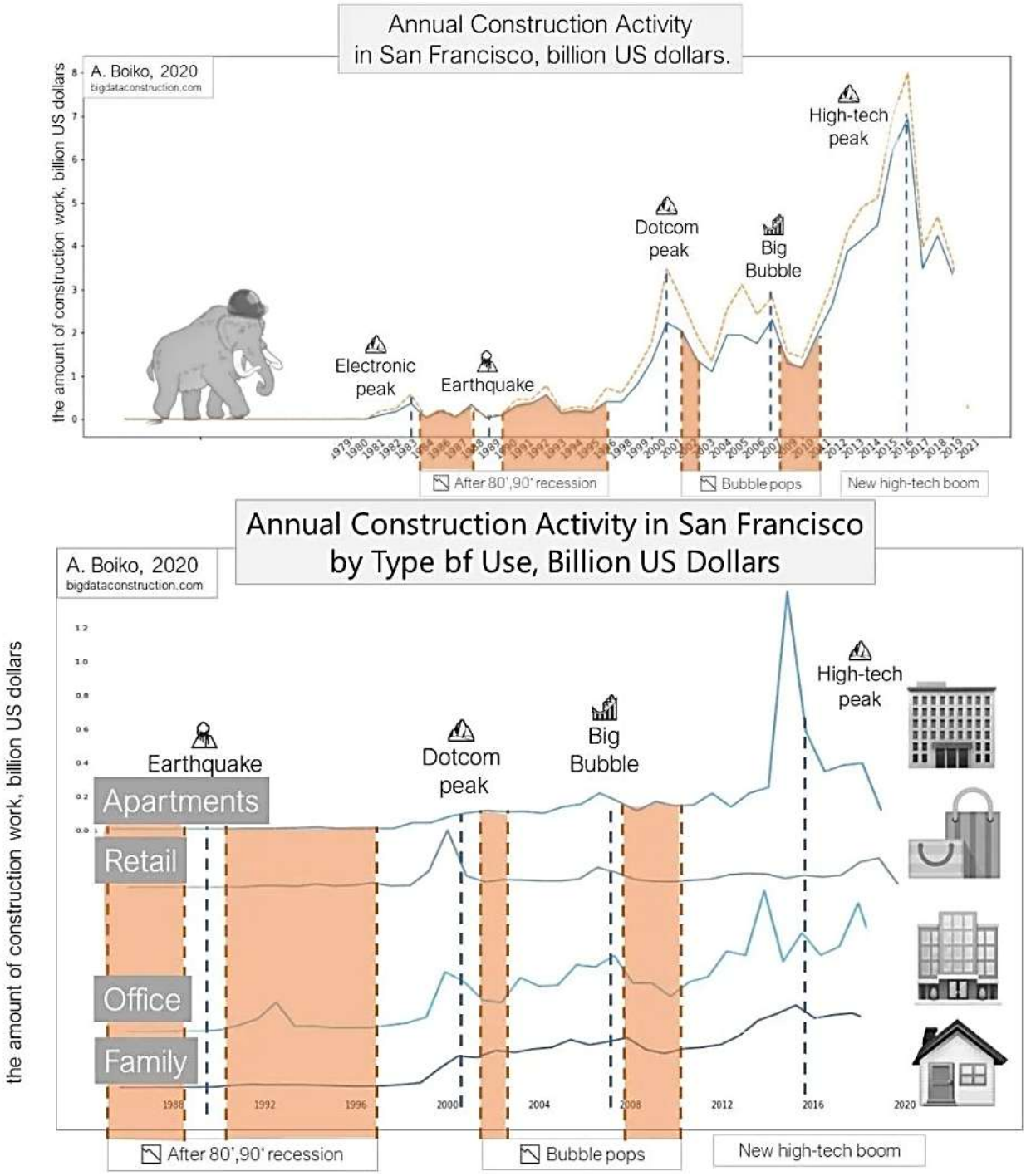
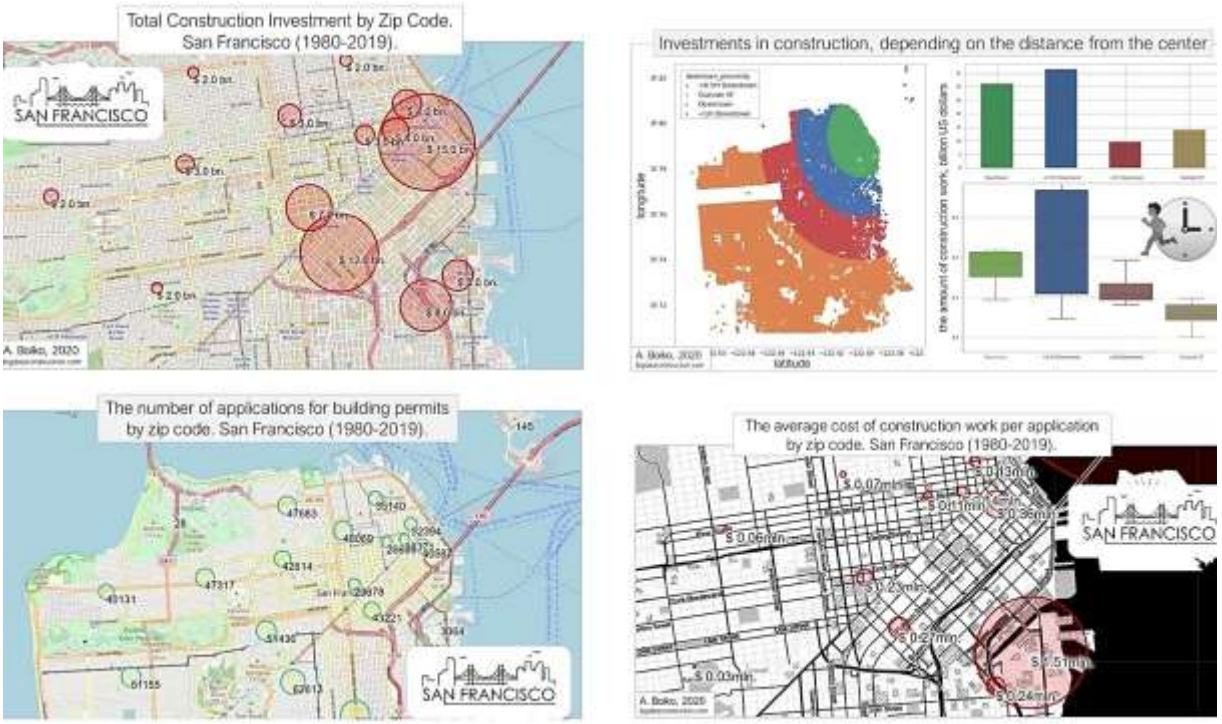


Fig. 9.1-4 I San Franciscos ejendomssektor korrelerer investeringerne med Silicon Valleys teknologiske udvikling.

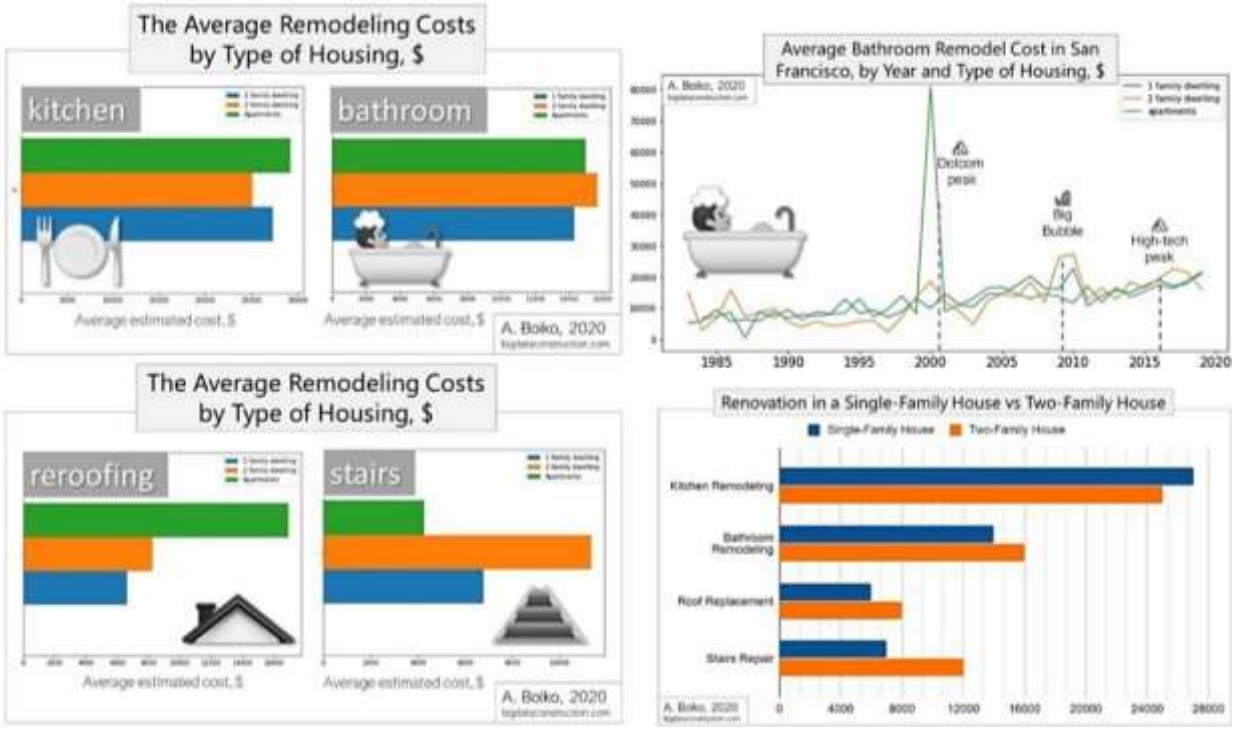
Dataanalyser viser, at i San Francisco er størstedelen af de 91,5 milliarder dollars, der er investeret i byggeri og ombygning i løbet af det sidste årti - næsten 75% koncentreret i byens centrum (figur 9.1-5 - brugt Pandas og Folium visualiseringsbibliotek) og inden for en radius af 2 km fra byens centrum, hvilket afspejler den højere investeringstæthed i disse centrale zoner.

De gennemsnitlige omkostninger til byggetilladelser varierer betydeligt fra bydel til bydel, og ansøgninger i byens centrum koster tre gange så meget som uden for centrum på grund af højere omkostninger til jord, arbejdskraft og materialer samt strenge bygningsregler, der kræver dyrere materialer for at forbedre energieffektiviteten.



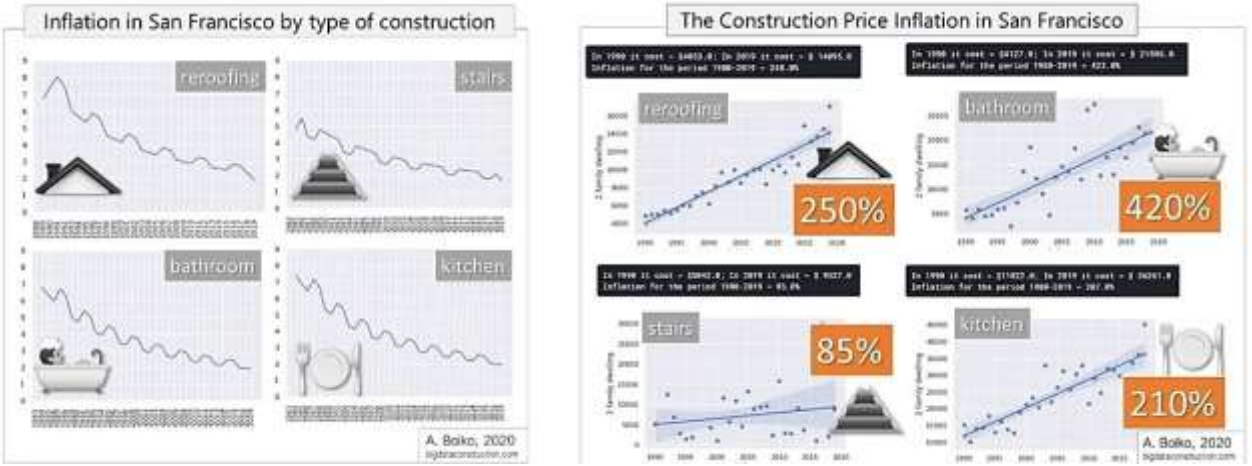
Figur 9.1-5 I San Francisco er 75 procent af byggeinvesteringerne (91,5 milliarder dollars) koncentreret i centrum.

Datasættet giver også mulighed for at beregne gennemsnitlige reparationspriser ikke kun efter hustype, men også efter bydele og individuelle adresser (postnumre). I San Francisco viser dynamikken i omkostningerne til boligrenovering forskellige tendenser for forskellige typer af renoveringer og boliger (figur 9.1-6 brugt Pandas og Matplotlib). Køkkenrenoveringer er mærkbart dyrere end badeværelsesrenoveringer: Den gennemsnitlige køkkenrenovering i et enfamiliehus koster omkring 28.000 dollars sammenlignet med 25.000 dollars i et tofamiliehus.



Figur 9.1-6 I SF koster køkkenrenoveringer næsten dobbelt så meget som badeværelsesrenoveringer, og husejere skal lægge$ 350 til side hver måned i 15 år for at dække udgifterne til større boligreparationer.

Byggeomkostningsinflationen i San Francisco gennem årene kan spores ved at analysere data grupperet efter boligtype og år (figur 9.1-7 - brugt af Pandas og Seaborn), som viser en jævn stigning i de gennemsnitlige reparationsomkostninger siden 1990 og afslører kortvarige treårige cyklusser i omkostningerne til reparationer af flerfamiliehuse.



Figur 9.1-7 Fra 1980 til 2019 er udgifterne til badeværelsesrenoveringer i SF femdoblet, mens tag- og køkkenrenoveringer er tredoblet i pris, og trapperenoveringer kun er steget med 85%.

En undersøgelse af offentlige data fra San Francisco Building Department (figur 9.1-3) viser, at

byggeomkostningerne i byen er ekstremt varierende og ofte uforudsigelige, påvirket af en række faktorer. Disse faktorer omfatter økonomisk vækst, teknologisk innovation og de unikke krav til forskellige boligtyper.

Tidligere krævede sådanne analyser indgående kendskab til programmering og analyse. Men med fremkomsten af LLM- værktøjer er processen blevet tilgængelig og forståelig for en bred vifte af fagfolk i byggebranchen, fra ingeniører i designafdelinger til den øverste ledelse.

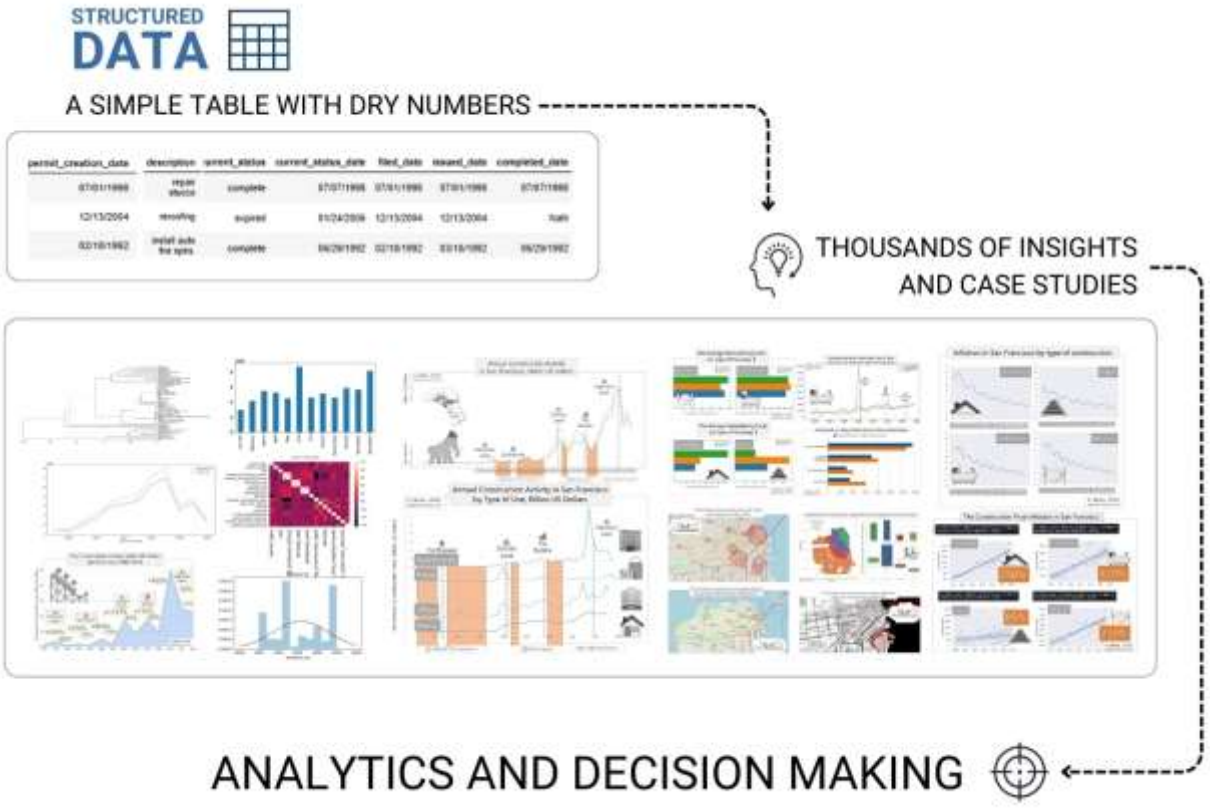


Fig. 9.1-8 Overgangen til visuelt forståelige data muliggør automatiseret beslutningstagning ved at genkende skjulte mønstre.

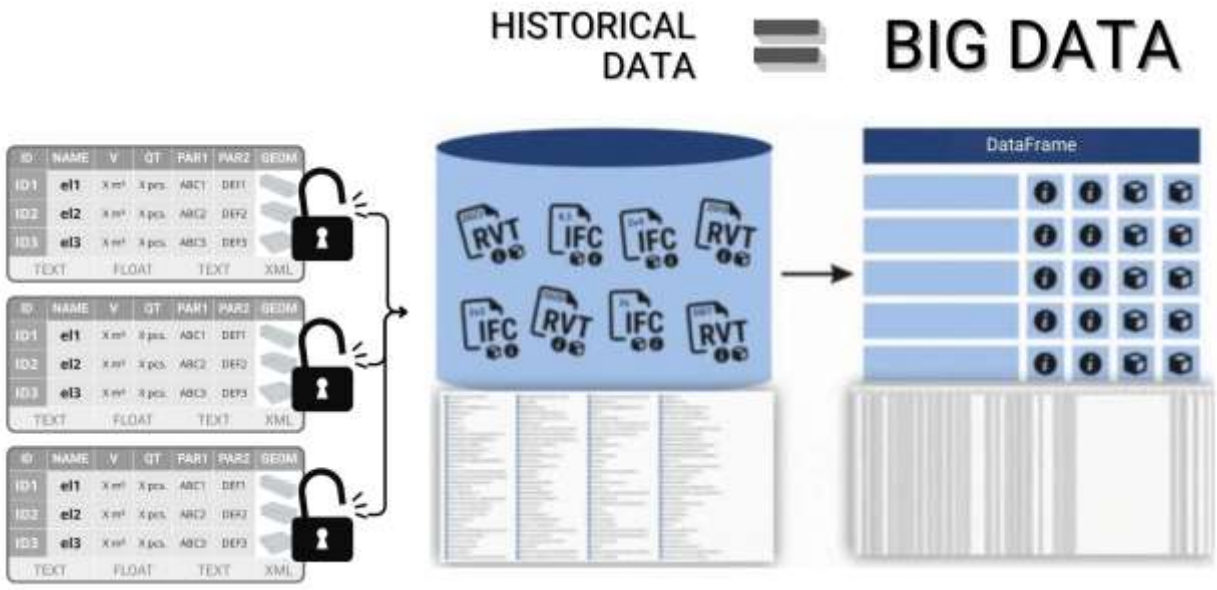
Ligesom vi analyserede data fra "San Francisco Building Authority"-datasættet i tabelform, kan vi visualisere og analysere ethvert datasæt - fra billeder og dokumenter til IoT-data eller data fra afledte CAD-databaser.

Eksempel på big data baseret på CAD-data (BIM)

I det følgende eksempel vil vi analysere et stort datasæt med data fra forskellige CAD-værktøjer (BIM). Til at indsamle og oprette det store datasæt blev der brugt en specialiseret automatiseret webcrawler (script), der var konfigureret til automatisk at søge og indsamle designfiler fra websteder, der tilbyder gratis arkitektoniske modeller i formaterne RVT og IFC. I løbet af et par dage fandt og downloadede crawleren 4.596 IFC-filer, 6.471 RVT-filer og 156.024 DWG-filer[149].

Efter at have indsamlet projekter i RVT og IFC formater i forskellige versioner og konverteret dem til et struktureret CSV-format ved hjælp af de gratis reverse engineering SDK'er, blev næsten 10.000 RVT- og IFC-projekter samlet i en stor Apache Parquet-tabel-fil og uploadet til Pandas DataFrame til analyse (figur 9.1-

9).



Figur 9.1-9 Strukturerede data projektdata giver dig mulighed for at kombinere et vilkårligt antal projekter i en enkelt todimensionel tabel.

Dataene fra denne store samling indeholder følgende oplysninger: IFC-filsættet indeholder ca. 4 millioner enheder (rækker) og 24.962 attributter (kolonner), og RVT-filsættet, der består af ca. 6 millioner enheder (rækker), indeholder 27.025 forskellige attributter (kolonner).

Disse informationssæt (figur 9.1-10) dækker millioner af elementer, for hvilke koordinaterne for Bounding Box -geometrien (et rektangel, der definerer grænserne for et objekt i projektet) desuden blev indhentet og tilføjet til en fælles tabel - koordinaterne for Bounding Box -geometrien (et rektangel, der definerer grænserne for et objekt i projektet) og billeder af hvert element i PNG-format og geometrien i det åbne XML-format - DAE (Collada) blev oprettet.

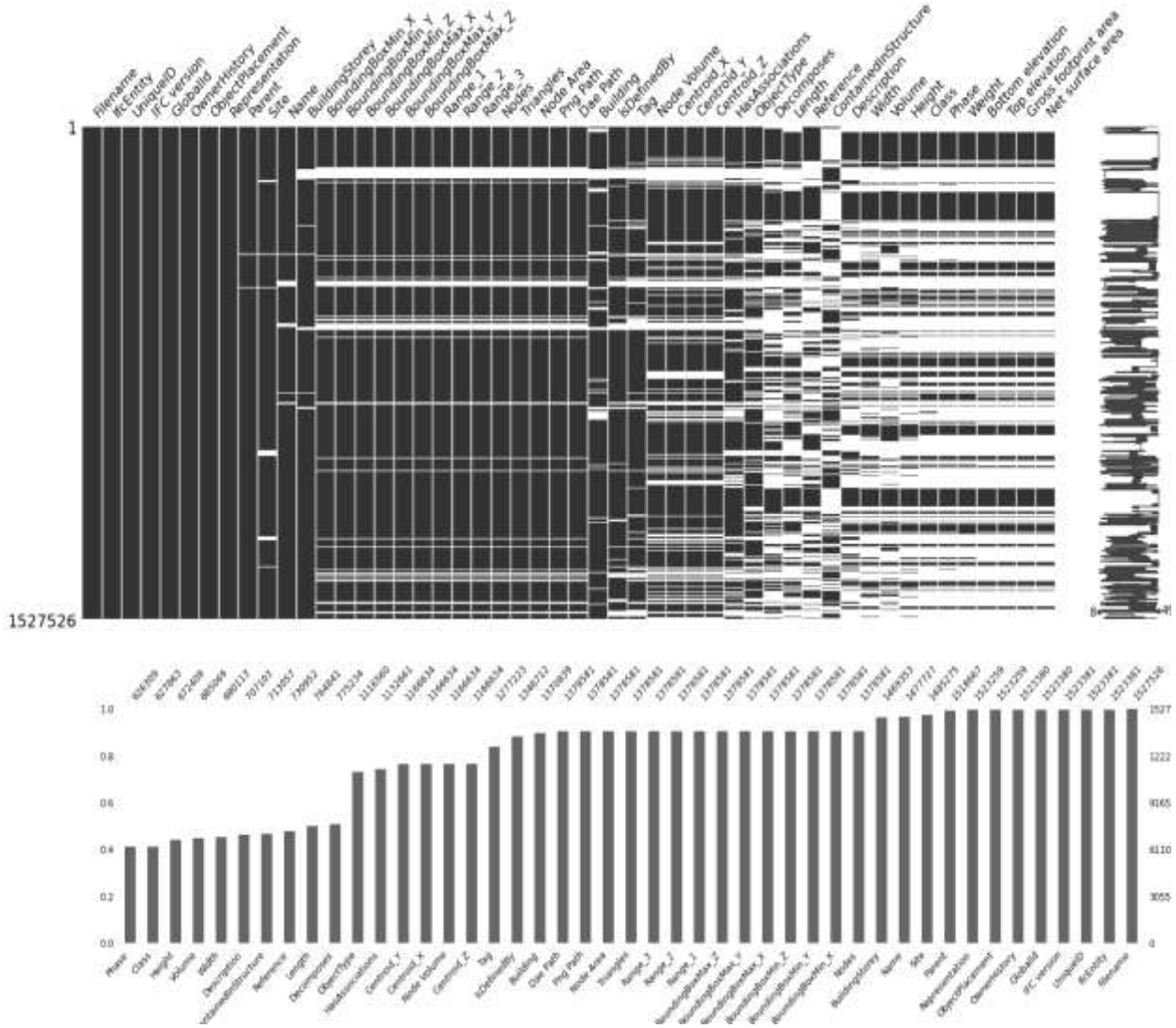


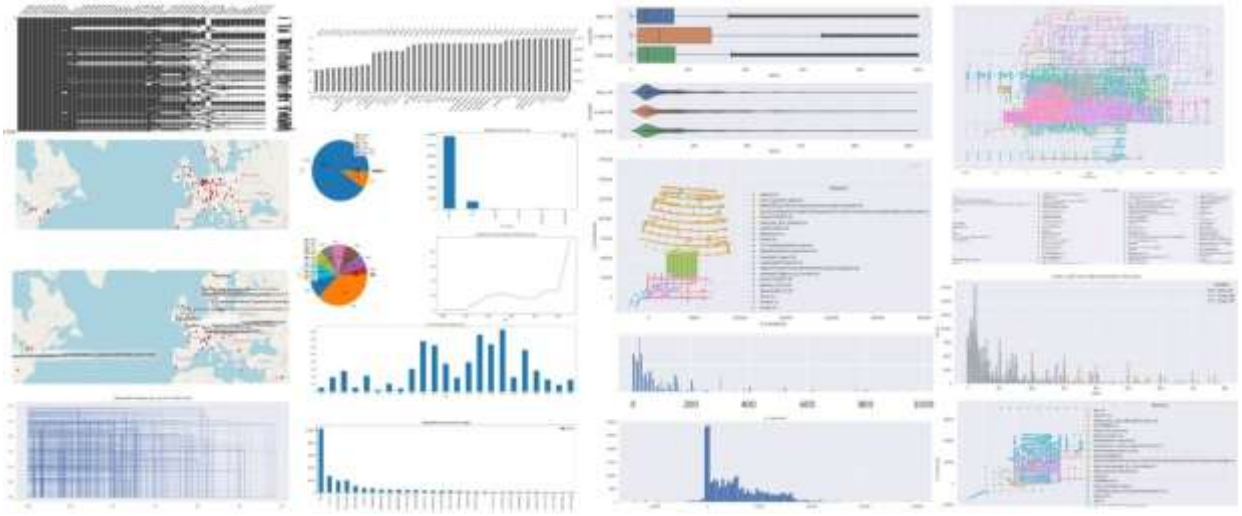
Fig. 9.1-10 Sabset med 1,5 millioner elementer og visualisering (missingno library) af de første 100 attributters belægning som et histogram.

Således fik vi alle oplysninger om titusindvis af elementer fra 4.596 IFC-projekter og 6.471 RVT-projekter, hvor alle attributter og egenskaber for alle enhedselementer og deres geometri (Bounding Box) blev oversat til en struktureret form af en enkelt tabel (DataFrame) (Fig. 9.1-10 - data om dataframepopulationer vises som histogrammer).

> Histogrammerne (fig. 9.1-10, fig. 9.2-6, fig. 9.2-7), der tegnes under analyseprocessen, giver mulighed for en hurtig vurdering af datatætheden og hyppigheden af forekomsten af værdier i kolonnerne. Dette giver et første indblik i fordelingen af funktioner, tilstedeværelsen af outliers og den potentielle anvendelighed af individuelle attributter til analyse og opbygning af maskinlæringsmodeller.

Et eksempel på en praktisk anvendelse af dette datasæt (fig. 9.1-10) er projektet "5000 IFC and RVT ". [149], der er tilgængeligt på Kaggle-platformen. Det præsenterer Jupyter Notebook med en komplet pipeline-

løsning: fra forbehandling af data og analyse til visualisering af resultater ved hjælp af Python-biblioteker - pandas, matplotlib, seaborn, folium og andre (fig. 9.1-11).



Figur 9.1-11 Eksempler på analyse af data fra CAD-formater (BIM) ved hjælp af Python-visualiseringsbiblioteker og pandas-biblioteket.

Baseret på metainformation er det muligt at afgøre, i hvilke byer bestemte projekter er blevet udviklet, og at vise dette på et kort (f.eks. ved hjælp af folium-biblioteket). Desuden giver tidsstempler i dataene mulighed for at udforske mønstre i, hvornår filer blev gemt eller redigeret: efter ugedag, tidspunkt på dagen og måned.

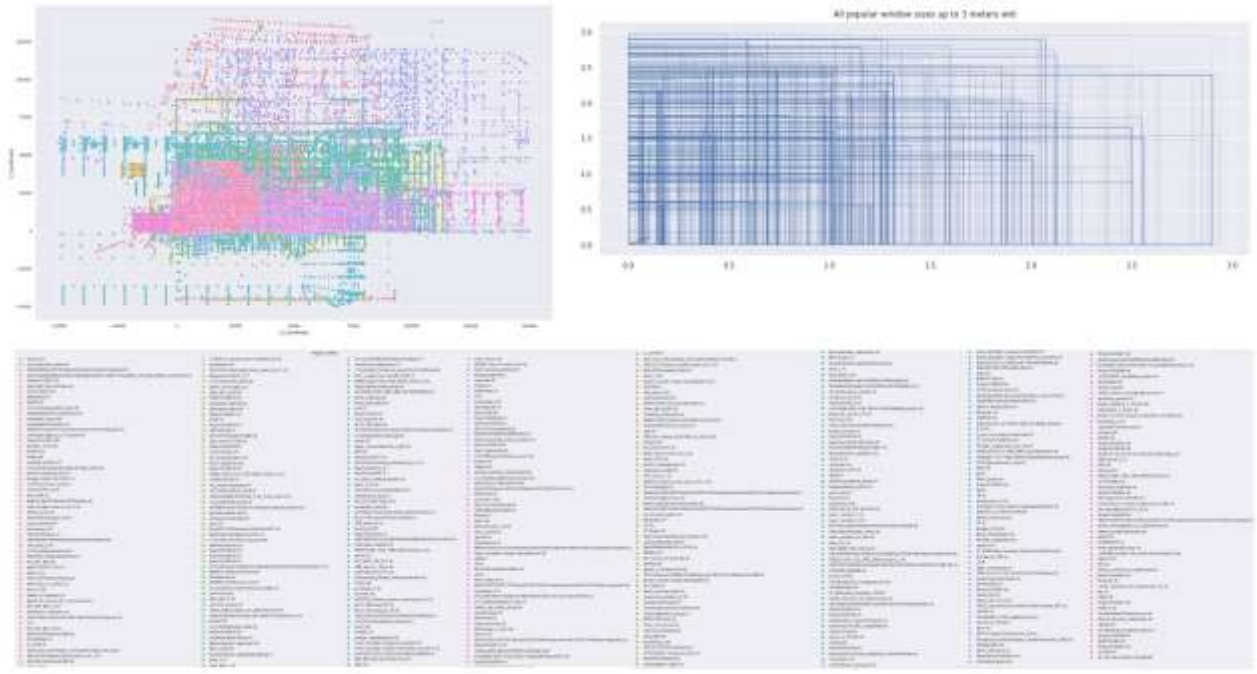


Fig. 9.1-12 Visualisering af den geometriske placering af alle søjler og dimensionerne af alle vinduer op til 3 meter i projekter fra listen nederst i diagrammet.

Geometriske parametre i form af Bounding Box uddraget fra modellerne egner sig også til aggregerede analyser. For eksempel viser figur 9.1-12 to grafer: Den venstre viser fordelingen af afstande mellem søjler

for alle projekter i forhold til nulpunktet, og den højre viser dimensionerne af alle vinduer op til 3 meters højde i en prøve på titusindvis af vindueselementer (efter gruppering af hele datasættet efter parameteren "Kategori" med værdien "OST_Windows", "IfcWindows").

Pipeline-analysekoden for dette eksempel og selve datasættet er tilgængelige på Kaggle-webstedet under titlen "5000 IFC and RVT | DataDrivenCo-nstruction.io projects" [149]. Denne færdige pipeline sammen med datasættet kan kopieres og køres gratis online på Kaggle eller offline i et af de populære IDE'er: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære online-værktøjer Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

> Den analytiske indsigt, der opnås ved at behandle og studere enorme mængder strukturerede data, vil spille en afgørende rolle i beslutningsprocesserne i byggebranchen.

Med denne form for informationsanalyse baseret på tidligere projekter kan specialisterne effektivt forudsige f.eks. materiale- og arbejdskraftbehov og optimere designløsninger, før byggeriet begynder

Men mens designdata eller byggetilladelser er relativt statiske oplysninger, der ændrer sig relativt langsomt, bliver selve byggeprocessen hurtigt mættet med en række sensorer og IoT -enheder: kameraer, automatiserede overvågningssystemer, der overfører data i realtid - alt dette gør byggepladsen til et dynamisk digitalt miljø, hvor data skal analyseres i realtid.

IoT Internet of Things og smarte kontrakter

IoT Internet of Things repræsenterer en ny bølge af digital transformation, hvor hver enhed får sin egen IP-adresse og bliver en del af et globalt netværk. IoT er et koncept, der involverer tilslutning af fysiske objekter til internettet for at indsamle, behandle og overføre data. I byggeriet betyder det muligheden for at styre byggeprocesser i realtid, minimere materialespild, forudsige slitage på udstyr og automatisere beslutningstagning.

> Ifølge CFMA-artiklen "Preparing for the Future with Connected Construction" [150] vil byggebranchen gennemgå en stor digital transformation i det næste årti, som vil kulminere i konceptet Connected Construction - en fuldt integreret og automatiseret byggeplads.

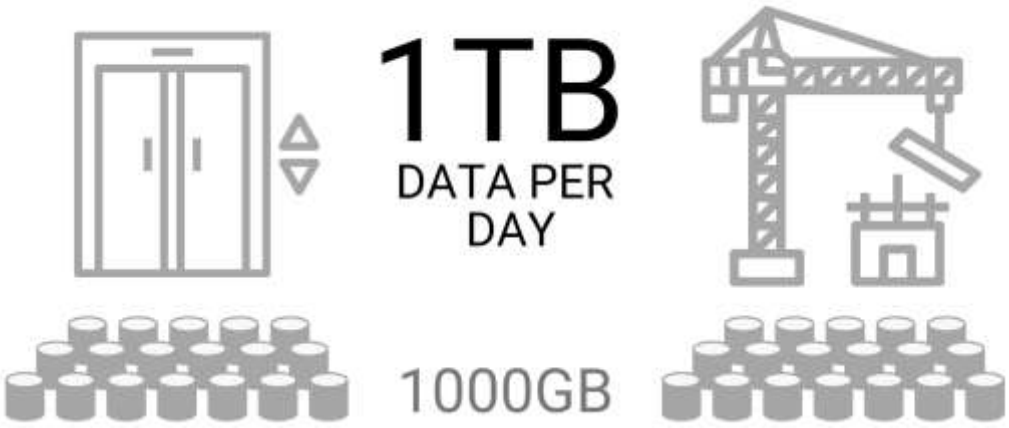


Fig. 9.1-13 IOT- eller byggepladsdataenheder kan producere og overføre terabytes af data om dagen.

En digital byggeplads indebærer, at alle elementer i byggeriet - fra planlægning og logistik til arbejdets udførelse og kvalitetskontrol på byggepladsen ved hjælp af faste kameraer og quadrocoptere - integreres i et enkelt dynamisk digitalt økosystem. Tidligere, i del 7 af denne bog, har vi allerede set på mulighederne i Apache NiFi (figur 7.4-5), et gratis open source-værktøj, der muliggør datastreaming i realtid - fra indsamling fra forskellige kilder til overførsel til lager- eller analyseplatforme.

Data om byggeriets fremskridt, materialeforbrug, udstyrsstatus og sikkerhed overføres i realtid til analysesystemer (fig. 9.1-13). Det gør det muligt at forudsige potentielle risici, reagere hurtigt på afvigelser og optimere byggepladsprocesserne. Nøglekomponenterne i en digital byggeplads omfatter:

- IoT -sensorer - sporing af miljøparametre, overvågning af byggeudstyr og kontrol af arbejdsforhold.
- Digitale tvillinger - virtuelle modeller af bygninger og infrastruktur til at forudsige mulige afvigelser og forhindre fejl.
- Automatiserede logistiksystemer - styring af forsyningskæden i realtid for at reducere nedetid og omkostninger.
- Robotbyggesystemer - brugen af autonome maskiner til at udføre rutinemæssige og farlige opgaver.

Robotisering brug af IoT og det digitale byggepladskoncept Connected Site (Construction) vil ikke bare øge effektiviteten og reducere omkostningerne, men også indlede en ny æra med sikkerhed, bæredygtigt byggeri og forudsigelig projektstyring.

> RFID (Radio Frequency Identification) tags er også en af de vigtigste komponenter i IoT. De bruges til at identificere og spore materialer, maskiner og endda personale på en byggeplads, hvilket øger gennemsigtigheden og kontrollen med projektets ressourcer.

RFID -teknologi bruges til automatisk at genkende genstande ved hjælp af radiosignaler. Den består af tre nøgleelementer:

- RFID -tags (passive eller aktive) - indeholder en unik identifikator og er fastgjort til materialer, værktøj eller maskiner.

- Scannere er enheder, der aflæser oplysninger fra tags og sender dem til systemet.
- Centraliseret database - gemmer oplysninger om objekternes placering, status og bevægelse.

Anvendelse af RFID i byggeriet:

- Automatisk materialeregnskab - tags på færdigblandede betonprodukter, armering eller sandwichpanelpakker giver mulighed for lagerstyring og forhindrer tyveri.
- Kontrol af personalets arbejde - RFID - medarbejderbadges registrerer skiftets start- og sluttidspunkt og giver en oversigt over arbejdstiden.
- Overvågning af udstyr - RFID - systemet sporer udstyrets bevægelser, forhindrer nedetid og forbedrer logistikkens effektivitet.

Denne teknologipakke suppleres af blockchain-baserede smarte kontrakter, der automatiserer betalinger, leveringskontrol og kontraktoverholdelse uden behov for mellemmænd, hvilket reducerer risikoen for svindel og forsinkelser.

> I dag er smarte kontrakter i mangel af en fælles datamodel blot kode, som deltagerne bliver enige om. Men med en datacentreret tilgang er det muligt at skabe en fælles model for kontraktparametre, kode den i en blockchain og automatisere opfyldelsen af vilkårene.

I et system til styring af forsyningskæden kan en smart kontrakt f.eks. spore leveringen af en forsendelse ud fra IoT -sensorer og RFID -tags og automatisk overføre betalingen, når den ankommer. På samme måde kan en smart kontrakt på en byggeplads registrere afslutningen af en arbejdsfase - f.eks. installation af armeringsjern eller støbning af et fundament - baseret på data fra droner eller byggesensorer og automatisk igangsætte den næste betaling til entreprenøren uden behov for manuelle kontroller og papircertifikater.

Men på trods af nye teknologier og internationale standardiseringsorganisationers indsats komplicerer et væld af konkurrerende standarder IoT-landskabet.

> Ifølge en Cisco-undersøgelse, der blev offentliggjort i 2017 [151], stopper næsten 60% af Internet of Things-initiativerne (IoT) på proof-of-concept-stadiet, og kun 26% af virksomhederne anser deres IoT-projekter for at være fuldt ud vellykkede. Desuden når en tredjedel af de afsluttede projekter ikke deres erklærede mål og anerkendes ikke som vellykkede, selv efter implementeringen.

En af hovedårsagerne er den manglende interoperabilitet mellem platforme, der behandler data fra forskellige sensorer. Resultatet er, at data forbliver isolerede i separate løsninger. Et alternativ til denne tilgang, som i andre lignende tilfælde (som vi har dækket i denne bog), er en arkitektur, der er bygget op omkring selve dataene som det primære aktiv.

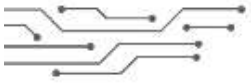
> IoT-sensorer spiller ikke kun en vigtig rolle i overvågningen af udstyrets tekniske tilstand, men også i forudsigende analyser for at reducere risici på byggepladsen og forbedre den samlede procesydelse ved at forudsige fejl og afvigelser.

De data, der indsamles af IoT-sensorer og RFID-tags, kan behandles i realtid af maskinlæringsalgoritmer, der kan opdage uregelmæssigheder og advare ingeniører om potentielle funktionsfejl på forhånd. Det kan være alt fra mikrorevner i betonkonstruktioner til ukarakteristiske pauser i tårnkranens drift, som indikerer

tekniske fejl eller overtrædelser af lovgivningen. Desuden kan avancerede adfærdsanalysealgoritmer opfange adfærdsmønstre, der f.eks. kan indikere fysisk træthed hos personalet, hvilket forbedrer den proaktive styring af sikkerheden og medarbejdernes trivsel på byggepladsen.

I byggebranchen sker ulykker og funktionsfejl - hvad enten det drejer sig om maskiner eller mennesker - sjældent pludseligt. De indledes som regel med mindre afvigelser, der går ubemærket hen. Predictive analytics og maskinlæring gør det muligt at opdage disse signaler på et tidligt tidspunkt, selv før der opstår kritiske konsekvenser.

Mens dokumenter, projektfiler og data fra IoT-enheder og RFID-tags udgør byggeprojekternes digitale fodaftryk, kan maskinlæring hjælpe med at uddrage nyttige indsigter fra dem. Med væksten i data og demokratiseringen af dataadgang får byggebranchen nye muligheder inden for analyse, forudsigende analyse og kunstig intelligens.



KAPITEL 9.2.
MASKINLÆRING OG FORUDSIGELSER

Maskinlæring og kunstig intelligens vil ændre den måde, vi bygger på

Databaserne i de forskellige systemer i byggebranchen - med deres uundgåeligt forfaldne og stadig mere komplekse infrastruktur - er ved at blive en yngleplads for fremtidige løsninger. Virksomhedens servere er som en skov rig på en biomasse af vigtig information, som ofte er skjult under jorden i mapper og servere. Masserne af data fra de forskellige systemer, der skabes i dag - efter brug, efter at være faldet til bunds i serveren og efter mange års fossilisering - vil give næring til maskinlæring og sprogmodeller i fremtiden. Virksomhedens interne chatrum (f.eks. en separat instans af lokalt konfigurerede ChatGPT, LlaMa, Mistral, DeepSeek) vil blive bygget på disse interne modeller ved hjælp af centraliseret lagring for hurtigt og bekvemt at hente oplysninger og generere de nødvendige grafer, dashboards og dokumenter.

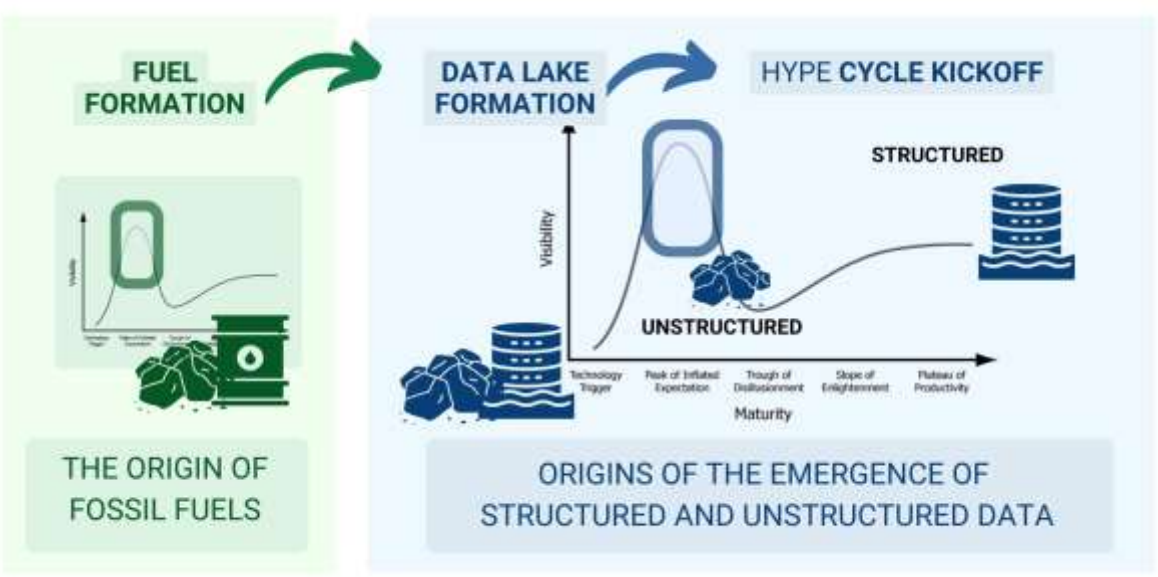


Fig. 9.2-1 Ligesom træer bliver til kul, bliver information også til værdifuld forretningsenergi med tiden under pres fra tid og analyse.

Fossilisering af plantemasse i kombination med tryk og temperatur skaber en homogen og unikt struktureret homogen masse af træer af forskellige arter, der levede på forskellige tidspunkter - trækul [152]. På samme måde danner information, der er optaget på harddiske i forskellige formater og på forskellige tidspunkter under pres fra analyseafdelinger og temperatur fra kvalitetsstyring, i sidste ende en homogen struktureret masse af værdifuld information (Fig. 9.2-1).

Disse lag (eller oftere isolerede nuggets) af information skabes gennem omhyggelig dataorganisering af erfarne analytikere, som gradvist begynder at udtrække værdifuld information fra tilsyneladende lange irrelevante data.

I det øjeblik disse modne datalag ikke længere bare "brændes" i rapporter, men begynder at cirkulere i forretningsprocesser, berige beslutninger og forbedre processer, bliver virksomheden klar til det næste skridt - overgangen til maskinlæring og kunstig intelligens (Fig. 9.2-2).

Maskinlæring (ML - Machine learning) er en klasse af metoder til at løse problemer med kunstig intelligens. Maskinlæringsalgoritmer genkender mønstre i store datasæt og bruger dem til at lære sig selv. Hvert nyt datasæt giver de matematiske algoritmer mulighed for at forbedre og tilpasse sig i henhold til de indhentede oplysninger, hvilket gør det muligt konstant at forbedre nøjagtigheden af anbefalinger og forudsigelser.

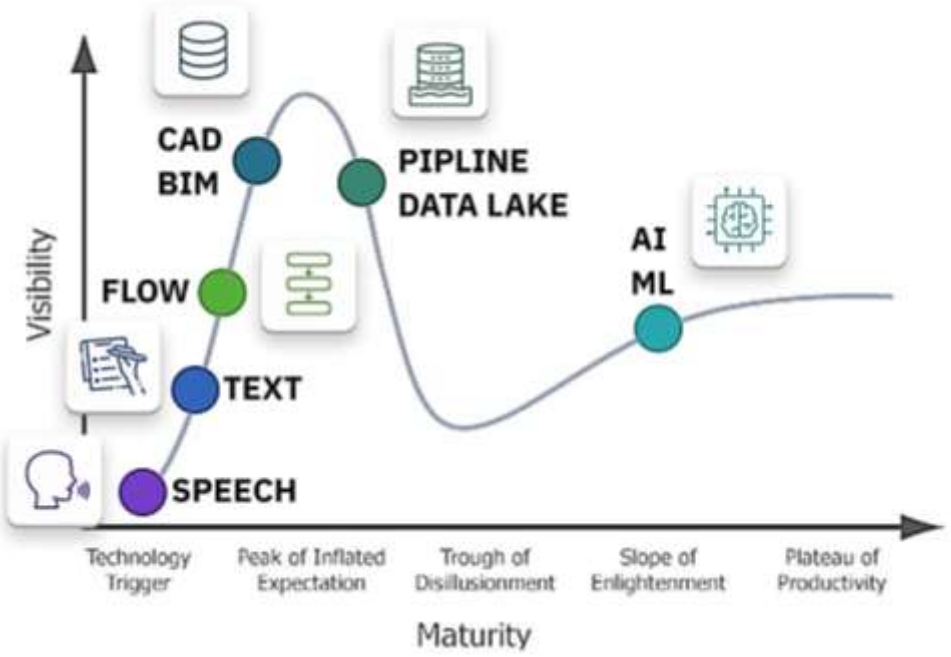


Fig. 9.2-2 Udfasningen af teknologier til oprettelse af data og anvendelsen af analyseværktøjer åbner døren til emnet maskinlæring.

Som den indflydelsesrige CEO for verdens største investeringsfond (som ejer vigtige andele i næsten alle de største byggesoftwarevirksomheder samt de virksomheder, der ejer den største mængde fast ejendom i verden [55]) sagde i et interview i 2023 - maskinlæring vil ændre byggeriets verden.

> *AI har et enormt potentiale. Det vil ændre den måde, vi arbejder på, og den måde, vi lever på. AI og robotteknologi vil ændre den måde, vi arbejder på, og den måde, vi bygger på, og vi vil kunne bruge AI og robotteknologi som et middel til at skabe meget større produktivitet [153].*
>
> - CEO for verdens største investeringsfond, interview, september 2023.

Machine Learning (ML) fungerer ved at behandle store mængder data og bruge statistiske teknikker til at efterligne aspekter af menneskelig tænkning. Men de fleste virksomheder har ikke sådanne datasæt, og hvis de har, er de ofte ikke tilstrækkeligt mærkede. Det er her, semantiske teknologier og transfer learning, en teknik, der gør det muligt for ML at være mere effektiv, når man håndterer små datamængder, og hvis gennemførlighed er blevet diskuteret i tidligere kapitler i denne del, kan hjælpe.

Essensen af transfer learning er, at man i stedet for at lære hver opgave fra bunden kan bruge viden fra beslægtede områder. Det er nødvendigt at indse, at mønstre og opdagelser fra andre brancher kan tilpasses og anvendes i byggebranchen. For eksempel kan metoder til optimering af logistikprocesser, der er udviklet i detailhandlen, hjælpe med at forbedre effektiviteten af forsyningskæden i byggeriet. Big data-analyse, som bruges aktivt i finansverdenen, kan anvendes til omkostningsprognoser og risikostyring i byggeprojekter. Og computersyn og robotteknologi, som udvikles i industrien, finder allerede anvendelse i automatiseret kvalitetskontrol, sikkerhedsovervågning og styring af byggepladsens faciliteter.

> Transfer learning gør det ikke kun muligt at fremskynde indførelsen af innovationer, men også at reducere omkostningerne ved deres udvikling ved at bruge den allerede akkumulerede erfaring fra andre brancher.



Fig. 9.2-3 Kunstig intelligens teknologier og robotteknologi vil være den vigtigste drivkraft i fremtiden for at øge produktiviteten i byggebranchen.

Menneskelig tænkning er organiseret efter et lignende princip: Vi bygger på tidligere erhvervet viden for at løse nye problemer (fig. 4.4-19, fig. 4.4-20, fig. 4.4-21). Inden for maskinlæring fungerer denne tilgang også - ved at forenkle datamodellen og gøre den mere elegant kan vi reducere problemets kompleksitet for ML-algoritmer. Dette reducerer igen behovet for store mængder data og reducerer beregningsomkostningerne.

Fra subjektiv vurdering til statistisk prognose

Den tid, hvor strategiske beslutninger afhang af den enkelte leders intuition (figur 9.2-4), hører fortiden til. I et stadig mere konkurrencepræget og udfordrende økonomisk miljø er en subjektiv tilgang ved at blive for risikabel og ineffektiv. Virksomheder, der fortsætter med at stole på personlige meninger i stedet for objektivt at analysere data, mister evnen til at reagere hurtigt på forandringer.

> Konkurrencemiljøet kræver nøjagtighed og repeterbarhed baseret på data, statistiske mønstre og beregnelig sandsynlighed. Beslutninger kan ikke længere baseres på følelser, de skal baseres på sammenhænge, tendenser og forudsigelige modeller, der stammer fra analyser og maskinlæring. Det er ikke bare en ændring i værktøjer - det er en ændring i tænkningens logik: fra antagelser til beviser, fra subjektive sandsynligheder til statistisk beregnede afvigelser, fra følelser til fakta.



Fig. 9.2-4 Tiden, hvor beslutninger blev truffet af HiPPO (den bedst betalte medarbejders mening), vil høre fortiden til med fremkomsten af big data og maskinlæring.

Ledere, der tidligere udelukkende stolede på deres egne følelser, vil uundgåeligt stå over for en ny virkelighed: Autoritet bestemmer ikke længere valg. I centrum for ledelse står nu systemer, der analyserer millioner af parametre og vektorer, identificerer skjulte mønstre og foreslår optimale strategier.

Hovedårsagen til, at virksomheder i dag stadig undgår at implementere ML, er den manglende gennemsigtighed. De fleste modeller fungerer som "sorte bokse" for ledere uden at forklare, hvordan de præcist kommer frem til deres konklusioner. Det fører til problemer: Algoritmer kan forstærke stereotyper og endda skabe humoristiske situationer, som i tilfældet med Microsofts chatbot, der hurtigt blev til et giftigt kommunikationsværktøj [154].

> I Deep Thinking reflekterer Garry Kasparov, tidligere verdensmester i skak, over sit nederlag til IBM's Big Blue-computer [155]. Han argumenterer for, at den sande værdi af AI ikke ligger i at kopiere menneskelig intelligens, men i at supplere vores evner. AI skal udføre opgaver, hvor mennesker er svage, mens mennesker bidrager med kreativitet. Computere har ændret den traditionelle tilgang til at analysere skak. I stedet for at skabe fascinerende historier om partier vurderer computerskakprogrammer hvert træk upartisk, kun baseret på dets faktiske styrke eller svaghed. Kasparov bemærker, at den menneskelige tendens til at se begivenheder som sammenhængende historier snarere end individuelle handlinger ofte fører til forkerte konklusioner - ikke kun i skak, men i livet generelt.

Hvis du planlægger at bruge maskinlæring til forudsigelse og analyse, er det derfor vigtigt at forstå de grundlæggende principper - hvordan algoritmer fungerer, og hvordan data behandles - før du begynder at bruge maskinlæringsværktøjer og AI i dit arbejde. Den bedste måde at komme i gang på er gennem praktisk erfaring.

Et af de mest praktiske værktøjer til en første introduktion til emnet maskinlæring og forudsigelse er Jupyter

Notebook og det populære klassiske Titanic-datasæt, som vil give en visuel introduktion til de vigtigste metoder til at analysere data og opbygge ML-modeller.

Titanic-datasæt: Hello World i analyseverdenen data og big data

Et af de mest berømte eksempler på brug af ML i dataanalyse er analysen af Titanic-datasættet, som ofte bruges til at undersøge sandsynligheden for, at passagererne overlever. At lære denne tabel svarer til "Hello World"-programmet, når man lærer programmeringssprog.

RMS Titanics forlis i 1912 resulterede i, at 1502 ud af 2224 mennesker døde. Titanic-datasættet indeholder ikke kun oplysninger om, hvorvidt en passager overlevede, men også attributter som: alder, køn, billetklasse og andre parametre. Dette datasæt er gratis tilgængeligt og kan åbnes og analyseres på forskellige offline- og onlineplatforme.

```
Link til Titanic-datasæt:
https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv
```

Tidligere i kapitlet "LLM-aktiverede IDE'er og fremtidige programmeringsændringer" diskuterede vi allerede Jupyter Notebook - et af de mest populære udviklingsmiljøer til dataanalyse og maskinlæring. Gratis cloud-analoger til Jupyter Notebook er Kaggle og Google Collab-platformene, som giver dig mulighed for at køre Python-kode uden at installere software og giver gratis adgang til computerressourcer.

> Kaggle er den største konkurrenceplatform for dataanalyse og maskinlæring med et integreret miljø til udførelse af kode. I oktober 2023 havde Kaggle over 15 millioner brugere [156] fra 194 lande.

Download og brug Titanic-datasættet på Kaggle platformen (figur 9.2-5) for at gemme datasættet (en kopi af det) og køre Python kode med forudinstallerede biblioteker direkte i en browser uden at skulle installere et dedikeret IDE.

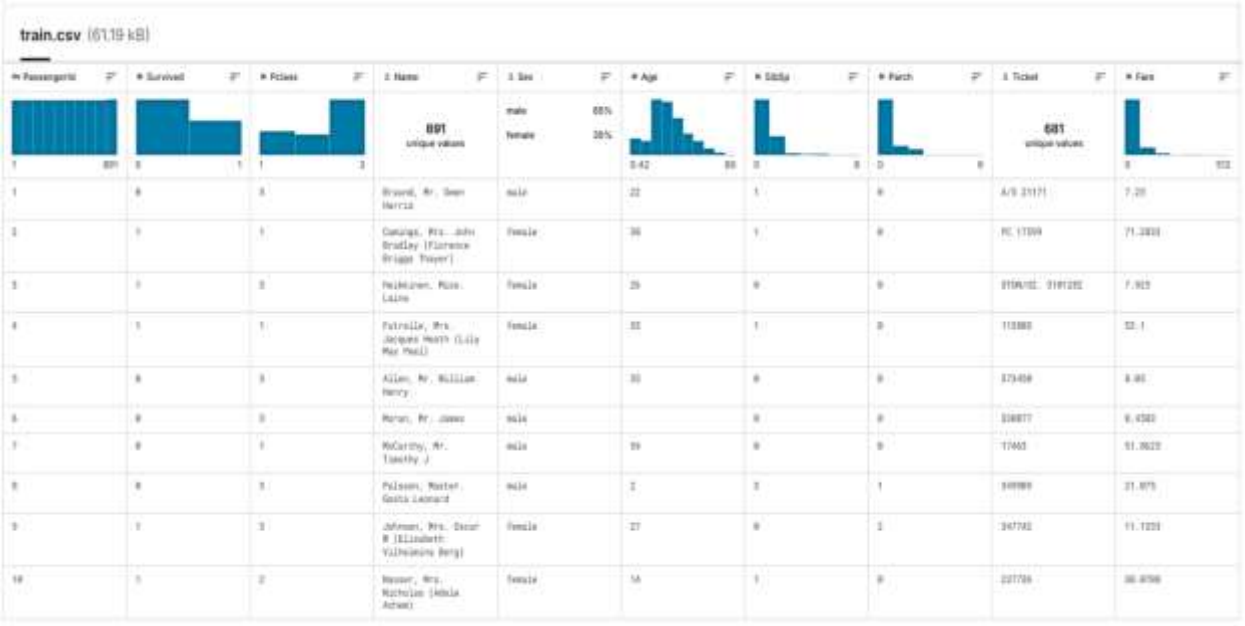


Fig. 9.2-5 Titanic table statistics - det mest populære træningsdatasæt til læring af dataanalyse og maskinlæring.

Titanic-datasættet indeholder data om de 2.224 passagerer, der var om bord på *RMS Titanic* på tidspunktet for dets forlis i 1912. Datasættet præsenteres som to separate tabeller, en trænings- (train.csv) og en testprøve (test.csv), så det både kan bruges til at træne modeller og til at evaluere deres nøjagtighed på nye data.

Træningsdatasættet indeholder både passagerattributter (alder, køn, billetklasse og andet) og oplysninger om, hvem der overlevede (kolonne med binære værdier "Overlevede"). Træningsdatasættet (figur 9.2-6 - fil train.csv) bruges til at træne modellen. Testdatasættet (Fig. 9.2-7 - fil test.csv) indeholder kun passagerattributter uden oplysninger om overlevende (uden en enkelt kolonne "Survivor"). Testdatasættet er beregnet til at teste modellen på nye data og til at evaluere dens nøjagtighed.

Vi har således næsten identiske attributter for passagerer i trænings- og testdatasættene. Den eneste vigtige forskel er, at vi i testdatasættet har en liste over passagerer, der ikke har kolonnen "Survivor" - målvariablen, som vi vil lære at forudsige ved hjælp af forskellige matematiske algoritmer. Og når vi har opbygget modellen, vil vi kunne sammenligne resultatet af vores model med den virkelige parameter "Survivor" fra testdatasættet, som vi vil tage i betragtning for at evaluere resultaterne.

Tabellens hovedkolonner, passagerparametre i trænings- og testdatasættet:

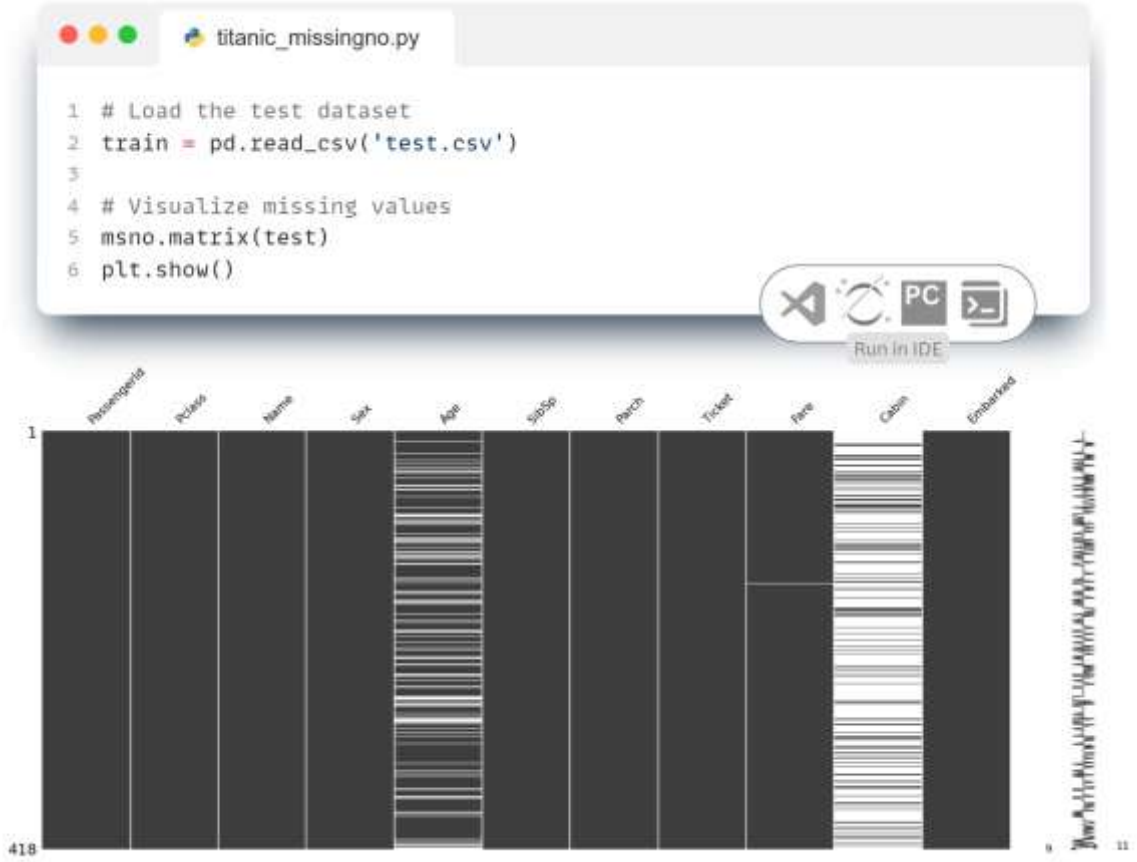
- **PassengerId** - unik passageridentifikation
- **Overlevede** - 1 hvis passageren overlevede, 0 hvis død (ikke tilgængelig i testsættet)
- **Pclass** - billetklasse (1, 2 eller 3)
- **Navn** - passagerens navn
- Køn - passagerens køn (mand/kvinde)
- Alder
- **SibSp** - antal brødre/søstre eller ægtefæller om bord

- **Parch** - antal forældre eller børn om bord
- **Billet** - billetnummer
- **Fare** - billetpris
- Kahyt - kahyttens nummer (mange data mangler)
- **Embarked** er indskibningshavnen (C = Cherbourg, Q = Queenstown, S = Southampton).

For at visualisere manglende data i begge tabeller kan du bruge missingno-biblioteket (fig. 9.2-6, fig. 9.2-7), som viser manglende værdier i form af et histogram, hvor hvide felter viser manglende data. Denne visualisering giver mulighed for en hurtig vurdering af datakvaliteten før behandling.



Fig. 9.2-6 Et par kodelinjer bruges til at visualisere manglende data i Titanic-træningsdatasættet, hvor nøgleparameteren for træning er parameteren "Overlevet".



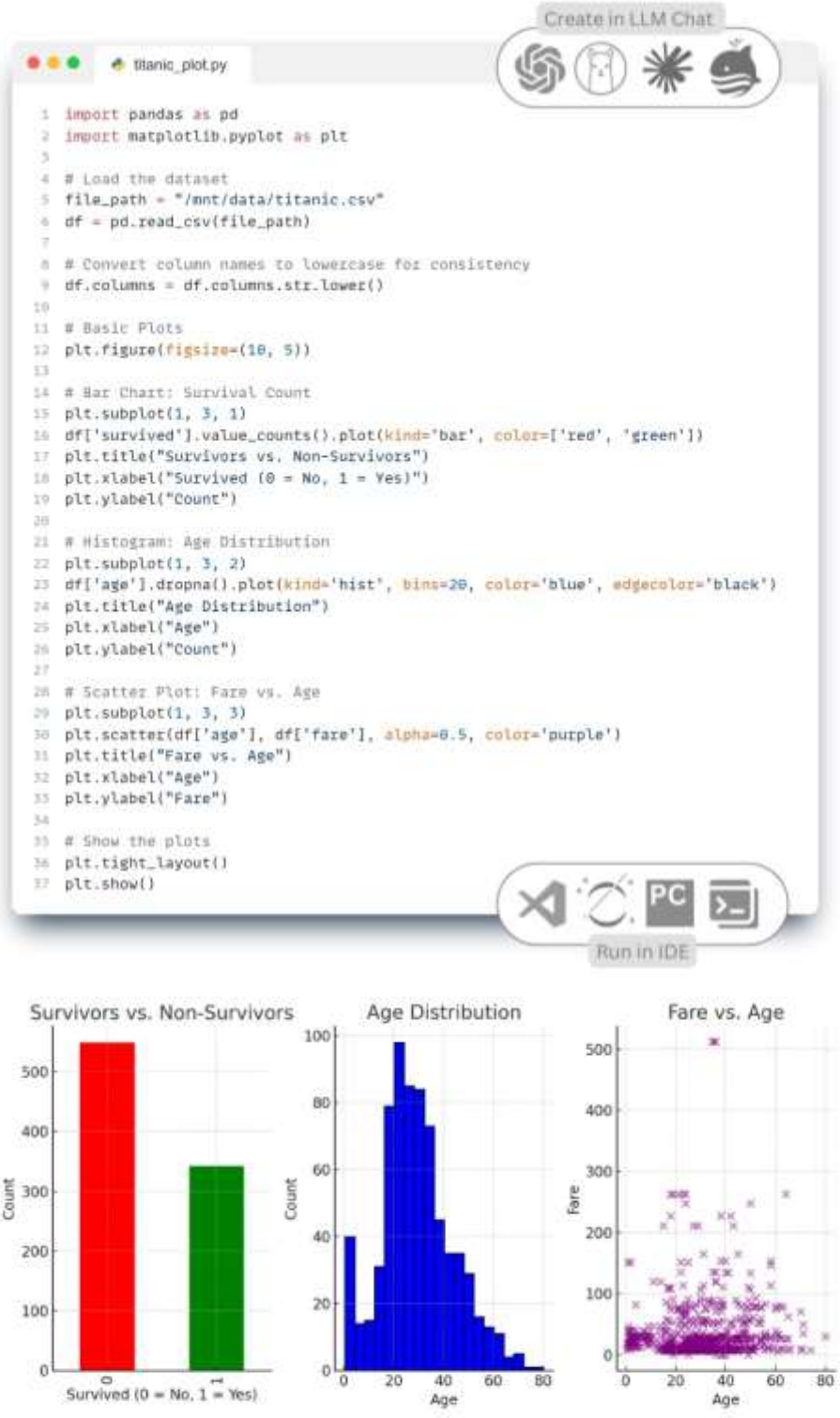
Figur 9.2-7 Visualisering af manglende data i Titanic-testdatasættet, som kun indeholder passagerkarakteristika uden oplysninger.

Før man formulerer hypoteser og kommer med forudsigelser baseret på datasættet, hjælper visuelle analyser med at identificere vigtige mønstre i dataene, vurdere deres kvalitet og identificere mulige afhængigheder. Der er mange visualiseringsteknikker, som kan hjælpe dig med at forstå Titanic-datasættet bedre. Du kan bruge fordelingsdiagrammer til at analysere passagerernes aldersgrupper, overlevelsesdiagrammer efter køn og klasse og matricer med manglende data til at vurdere kvaliteten af oplysningerne og forstå dataene.

- Lad os bede LLM om at hjælpe os med at visualisere dataene fra Titanic-datasættet ved at sende følgende tekstanmodning til en hvilken som helst LLM-model (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller en anden):

> Vis nogle enkle grafer for Titanic-datasættet. Download selv datasættet og vis ⏎

  - LLM-svar i form af færdiglavet kode og grafer, visualiserer datasætparametrene:



Figur 9.2-8 LLM hjælper dig med at få en øjeblikkelig visualisering af dine datasætdata.

Datavisualisering er et vigtigt skridt for at forberede datasættet til den efterfølgende konstruktion af en

maskinlæringsmodel, som man kun kan få adgang til ved at forstå dataene.

Machine learning i aktion: fra Titanic-passagerer til projektledelse

Hovedhypotesen, der blev brugt til at udforske maskinlæringsrammen baseret på Titanic-datasættet, er, at visse grupper af passagerer havde en større chance for at overleve.

> Den lille tabel med Titanic-passagerer er blevet populær over hele verden, og millioner af mennesker bruger den til træning, eksperimenter og modeltest for at finde ud af, hvilke algoritmer og hypoteser der vil opbygge den mest nøjagtige model til forudsigelse af overlevelse baseret på træningsdatasættet for Titanic-passagerer.

Titanic-datasættets tiltrækningskraft skyldes dets kompakthed: med flere hundrede rækker og tolv kolonner (fig. 9.2-6) giver det rig mulighed for analyse. Datasættet er relativt enkelt et klassisk eksempel på en binær klassifikationsløsning, hvor målet med problemet - overlevelse - udtrykkes i det praktiske format 0 eller 1.

> John Wheeler i "It from Bit" [7], at universet er baseret på binære valg. På samme måde er en virksomhed, der drives af mennesker, som består af molekyler, faktisk bygget på en række binære valg.

Desuden er dataene baseret på en virkelig historisk begivenhed, hvilket gør dem værdifulde for forskningen i modsætning til kunstigt skabte eksempler. Alene på Kaggle-platformen, en af de største Data Pipeline og ETL, deltog 1.355.998 personer i de Titanic-datasætbaserede udfordringer og udviklede 53.963 unikke Data Pipeline-løsninger [157] (Figur 9.2-9).

> Det virker utroligt, men bare 1000 linjer data om Titanic-passagererne med 12 parametre er blevet til et felt for millioner af hypoteser, logiske kæder og unikke datapipelines. Fra et lille datasæt fødes endeløse indsigter, hypoteser og fortolkninger - fra simple overlevelsesmodeller til komplekse ensembler, der tager højde for skjulte mønstre og komplekse labyrinter af ræsonnementer.

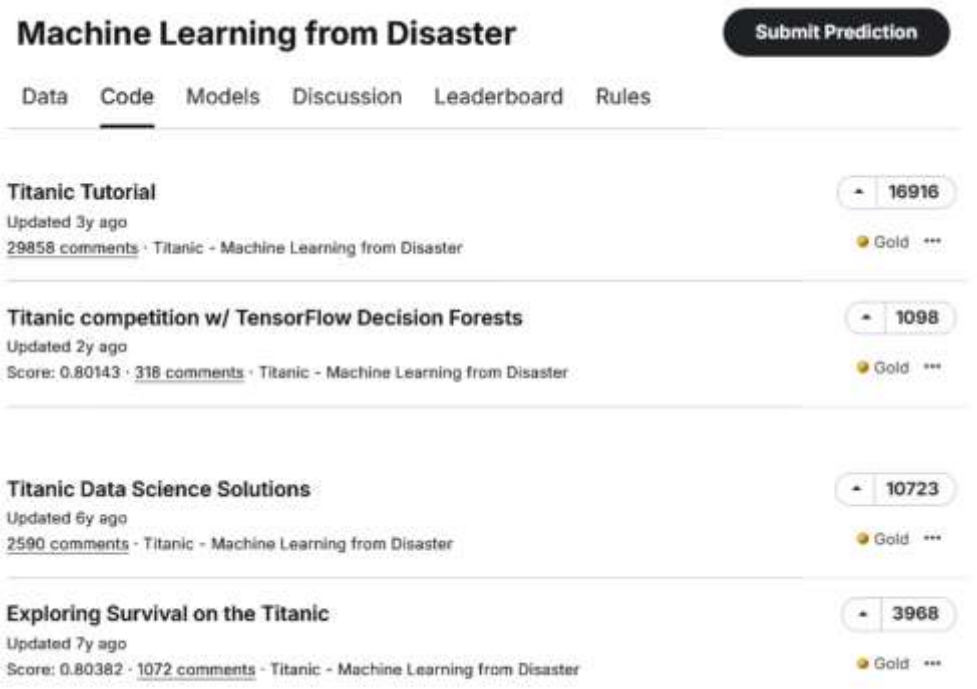


Fig. 9.2-9 De første fem løsninger ud af i alt 53.963 færdige og open source Pipeline-løsninger. Næsten 1,5 millioner mennesker har allerede forsøgt at løse dette problem alene på Kaggle [157].

Hvis selv sådan en lille tabel kan generere millioner af unikke løsninger (Fig. 9.2-9), hvad kan vi så sige om virkelige industrielle konstruktionsdatasæt, hvor parametrene måles i titusindvis?

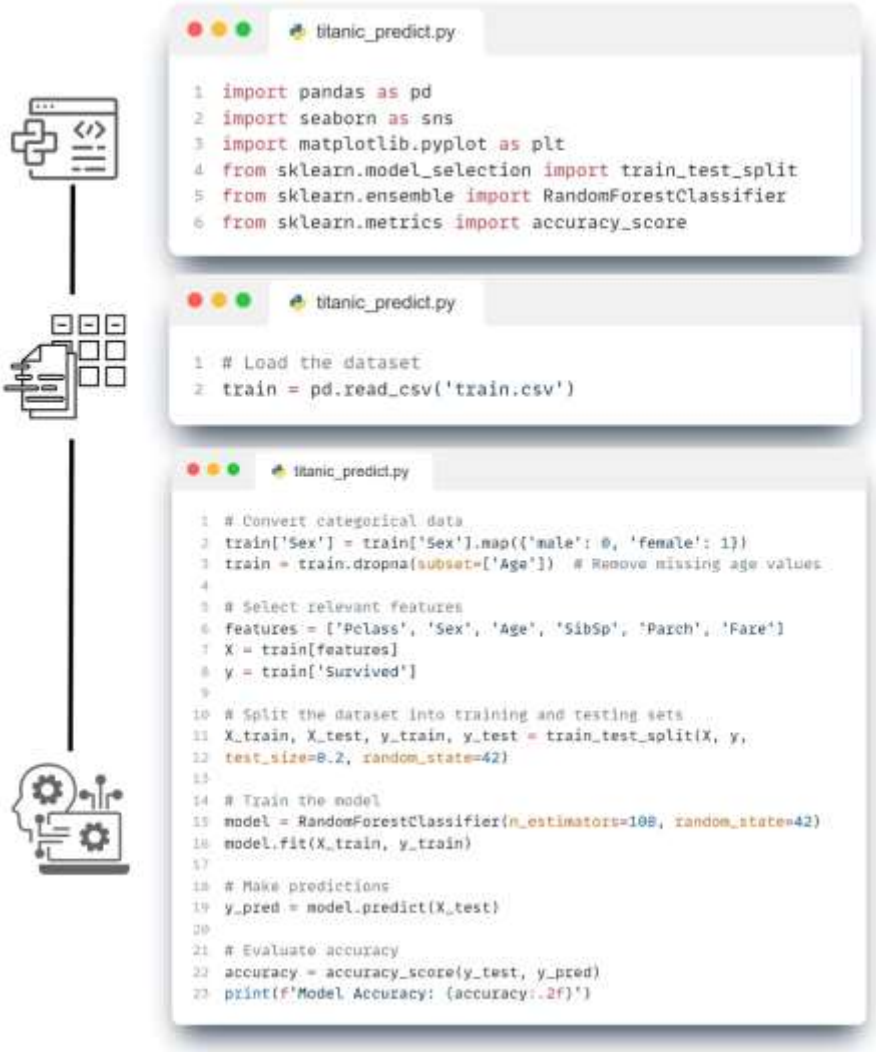
> Et standard CAD -design af en relativt lille bygning indeholder titusindvis af enheder med tusindvis af parametre - fra geometriske egenskaber til omkostnings- og tidsattributter. Forestil dig, hvor mange potentielle indsigter, relationer, forudsigelser og ledelseshypoteser der ligger gemt i data fra alle din virksomheds projekter indsamlet gennem årene. Historiske projektdata er ikke bare et arkiv - det er en organisations levende hukommelse, dens digitale fodaftryk, som kan analyseres for at opbygge et stort antal unikke hypoteser.

Det vigtigste er, at du ikke behøver at vente på, at Kaggle interesserer sig for din virksomhed eller dine data. Du kan begynde at arbejde med det, du har i dag: Kør analyser på dine egne data, træn modeller på dine egne data, identificer gentagelser, anomalier og mønstre. Hvor det tidligere krævede årelange eksperimenter og dyr rådgivning, er det nu nok med initiativ, en LLM, en åben tilgang til data og en vilje til at lære.

- For at opbygge en maskinlæringsalgoritme, som kan forudsige passagerernes overlevelsesrate baseret på train.csv-passagerernes træningsdatasæt, beder vi LLM om at løse dette problem for os:

> Baseret på Titanic-passagerernes træningsdatasæt skal du opbygge en maskinlæringsmodel til at forudsige overlevelsesrater ⏎

LLM's svar :



titanic_predict.py

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
```

titanic_predict.py

```python
# Load the dataset
train = pd.read_csv('train.csv')
```

titanic_predict.py

```python
# Convert categorical data
train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
train = train.dropna(subset=['Age'])  # Remove missing age values

# Select relevant features
features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']
X = train[features]
y = train['Survived']

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Train the model
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f'Model Accuracy: {accuracy:.2f}')
```

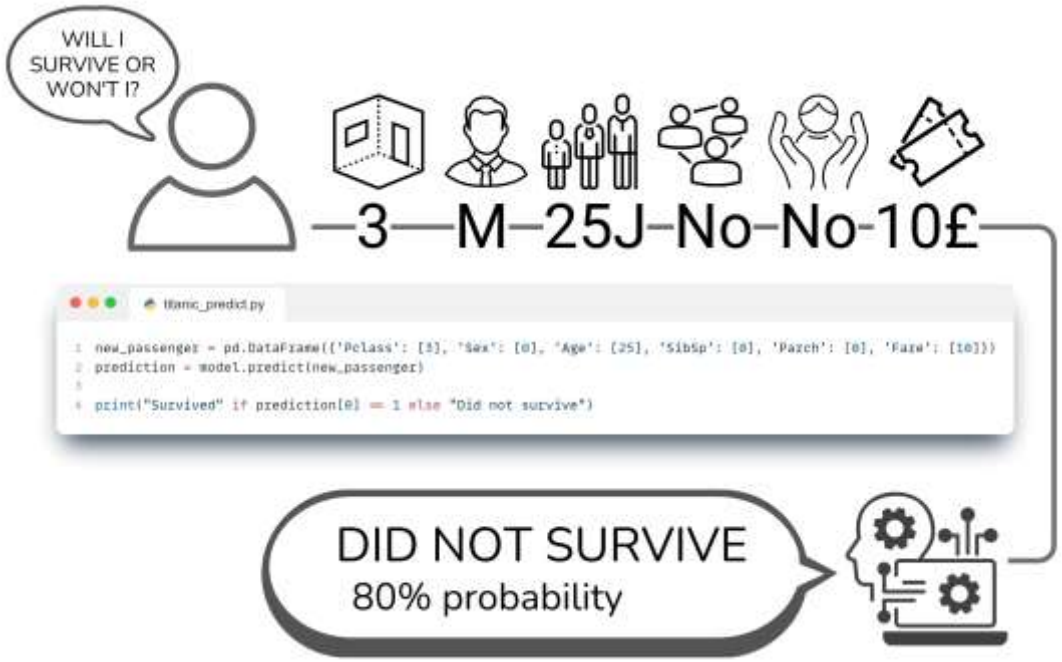
Fig. 9.2-10 LLM byggede en forudsigelse af Titanic-overlevende ved hjælp af maskinlæringsalgoritmen Random Forest.

Den resulterende kode fra LLMFig. 9.2-10) indlæser Titanic-passagerdata, renser dem, konverterer kategoriske variabler (f.eks. køn til numerisk format) og træner modellen gennem RandomForestClassifier-algoritmen for at forudsige, om en passager overlevede eller ej (vi vil tale mere om populære algoritmer i de følgende kapitler).

Koden opdeler træningsdataene i trænings- og testsæt (Kaggles hjemmeside har allerede oprettet færdige test.csv (fig. 9.2-7) og train.csv (fig. 9.2-6) til træning, hvorefter modellen trænes på træningsdataene og testes på testdataene for at se, hvor god en bestemt forudsigelsesmodel er. Efter træningen føres testdataene fra test.csv (med rigtige data om dem, der overlevede eller ikke overlevede) ind i modellen, og den forudsiger, hvem der overlevede, og hvem der ikke gjorde. I vores tilfælde er nøjagtigheden af den maskinlæringsmodel, vi fik, ca. 80%, hvilket viser, at den fanger mønstrene ret godt.

> Machine learning kan sammenlignes med et barn, der forsøger at få en rektangulær klods til at passe i et rundt hul. I de indledende faser prøver algoritmen mange tilgange og støder på fejl og uoverensstemmelser. Denne proces kan virke ineffektiv, men den giver vigtig læring: Ved at analysere hver fejl forbedrer modellen sine forudsigelser og træffer stadig mere præcise beslutninger.

Nu kan denne model (fig. 9.2-10) bruges til at forudsige overlevelsesraten for nye passagerer, og hvis du f.eks. fodrer den med passageroplysninger ved hjælp af funktionen model.predict parametrene: "mand", "3. klasse", "25 år gammel", "ingen slægtninge om bord", vil modellen give en forudsigelse - at passageren med 80% sandsynlighed ikke vil overleve katastrofen, hvis han var på Titanic i 1912 (Fig. 9.2-11).



Figur 9.2-11 Den model, vi skabte ovenfor, kan nu med 80% sandsynlighed forudsige, om en ny Titanic-passager vil overleve eller ej.

Modellen til forudsigelse af Titanic-passagerernes overlevelse illustrerer et meget bredere koncept: Hver dag træffer tusindvis af fagfolk i byggebranchen lignende "dobbelte" beslutninger - liv eller død for en beslutning, et projekt, et overslag, et værktøj, fortjeneste eller tab, sikkerhed eller risiko. Som i Titanic-eksemplet, hvor udfaldet afhang af faktorer (køn, alder, klasse), er hvert aspekt af beslutningen i byggeriet påvirket af mange af sine egne faktorer og variabler (kolonner i tabeller): materialeomkostninger, medarbejdernes kvalifikationer, tidsfrister, vejr, logistik, tekniske risici, kommentarer og hundredtusindvis

af andre parametre.

> I byggebranchen følger maskinlæring de samme principper som inden for andre områder: Modeller trænes på historiske data - fra projekter, kontrakter og estimater - for at teste forskellige hypoteser og finde de mest effektive løsninger. Denne proces minder meget om at undervise et barn gennem forsøg og fejl: For hver cyklus tilpasser modellerne sig og bliver mere præcise.

Brugen af akkumulerede data åbner nye horisonter for byggeriet. I stedet for tidskrævende manuelle beregninger kan der oplæres modeller, som med stor nøjagtighed kan forudsige de vigtigste egenskaber ved fremtidige projekter. På den måde forvandler predictive analytics byggebranchen til et sted, hvor man ikke bare kan planlægge, men også med sikkerhed forudsige udviklingen.

Forudsigelser og prognoser baseret på historiske data

De data, der indsamles om virksomhedens projekter, åbner mulighed for at opbygge modeller, der kan forudsige omkostnings- og tidskarakteristika for fremtidige, endnu ikke realiserede objekter - uden tidskrævende manuelle beregninger og sammenligninger. Det gør det muligt at fremskynde og forenkle værdiansættelsesprocesser betydeligt, idet man ikke baserer sig på subjektive antagelser, men på solide matematiske prognoser.

Tidligere, i bogens fjerde del, har vi set nærmere på traditionelle metoder til estimering af projektomkostninger, herunder den ressourcebaserede metode, og også nævnt parametriske tilgange og eksperttilgange. Disse metoder er stadig relevante, men i moderne praksis er de begyndt at blive beriget med værktøjer til statistisk analyse og maskinlæring, som kan forbedre nøjagtigheden og reproducerbarheden af estimater betydeligt.

Processerne med manuel og halvautomatisk beregning af priser og tidsattributter vil i fremtiden blive suppleret med udtalelser og forudsigelser fra ML-modeller, der er i stand til at analysere historiske data, finde skjulte mønstre og foreslå informerede beslutninger.
Nye data og scenarier vil blive genereret automatisk ud fra allerede tilgængelige oplysninger - på samme måde som sprogmodeller (LLM) skaber tekster, billeder og kode baseret på data, der er indsamlet i årenes løb fra åbne kilder [158].

> Ligesom mennesker i dag er afhængige af erfaring, intuition og interne statistikker for at vurdere fremtidige begivenheder, vil fremtiden for byggeprojekter i de kommende år i stigende grad blive bestemt af en kombination af akkumuleret viden og matematiske maskinlæringsmodeller.

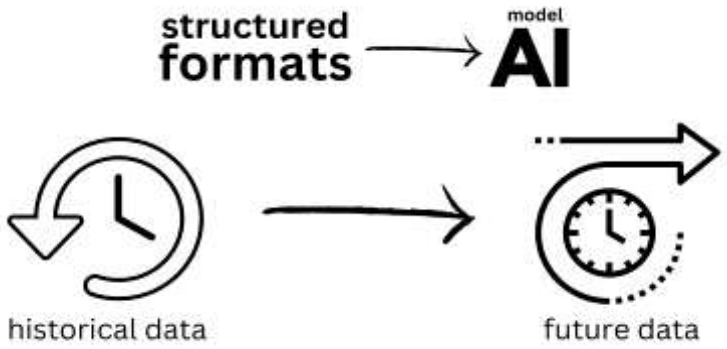


Fig. 9.2-12 Kvalitative og strukturerede historiske virksomhedsdata er det materiale, som maskinlæringsmodeller og forudsigelser bygger på.

Tag et enkelt eksempel: Forudsigelse af prisen på et hus baseret på dets areal, grundstørrelse, antal værelser og geografiske placering. En tilgang er at opbygge en klassisk model, der analyserer disse parametre og beregner den forventede pris (figur 9.2-13). Denne tilgang kræver en præcis og kendt formel på forhånd, hvilket er praktisk talt umuligt i praksis.

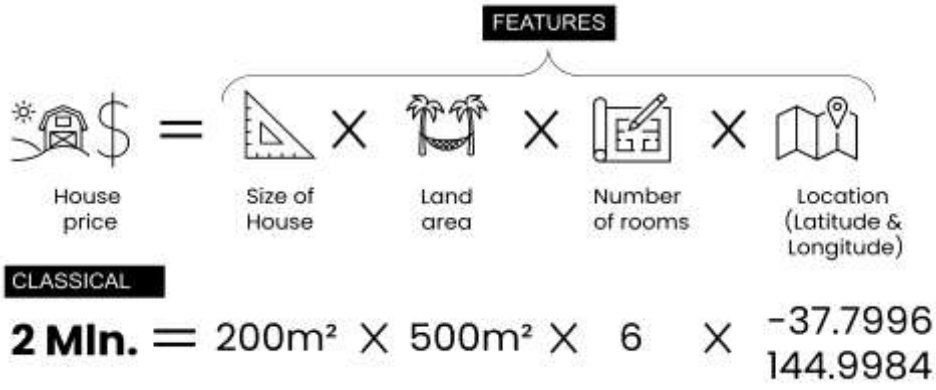


Fig. 9.2-13 En klassisk algoritme kan bruges til at estimere værdien af et hus med en fast formel, der skal findes.

Maskinlæring eliminerer den manuelle søgning efter formler og erstatter dem med trænede algoritmer, der uafhængigt identificerer afhængigheder, som er mange gange mere nøjagtige end nogen forudbestemte ligninger. Lad os alternativt oprette en maskinlæringsalgoritme, som genererer en model baseret på en forudgående forståelse af problemet og historiske data, der kan være ufuldstændige (figur 9.2-14).

Med prisfastsættelse som eksempel gør maskinlæring det muligt at skabe forskellige typer af matematiske modeller, som ikke kræver kendskab til den nøjagtige mekanisme for omkostningsdannelse. Modellen "lærer" af data fra tidligere projekter og tilpasser sig reelle mønstre mellem bygningsparametre, deres omkostninger og deadlines.

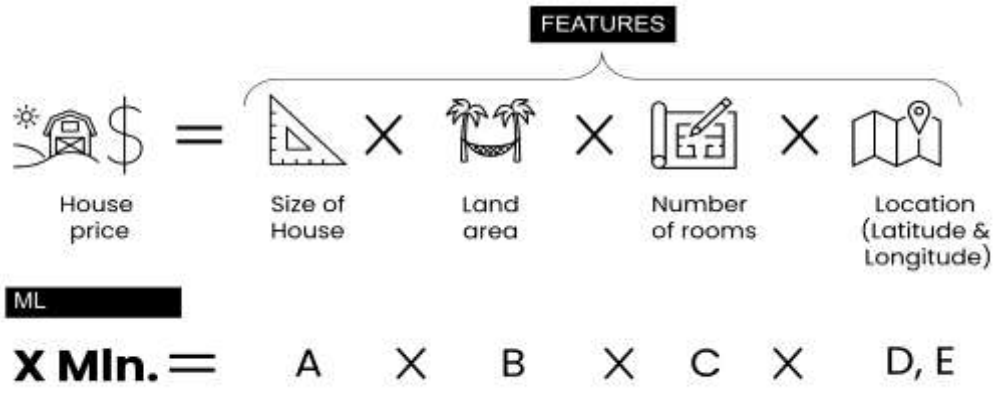
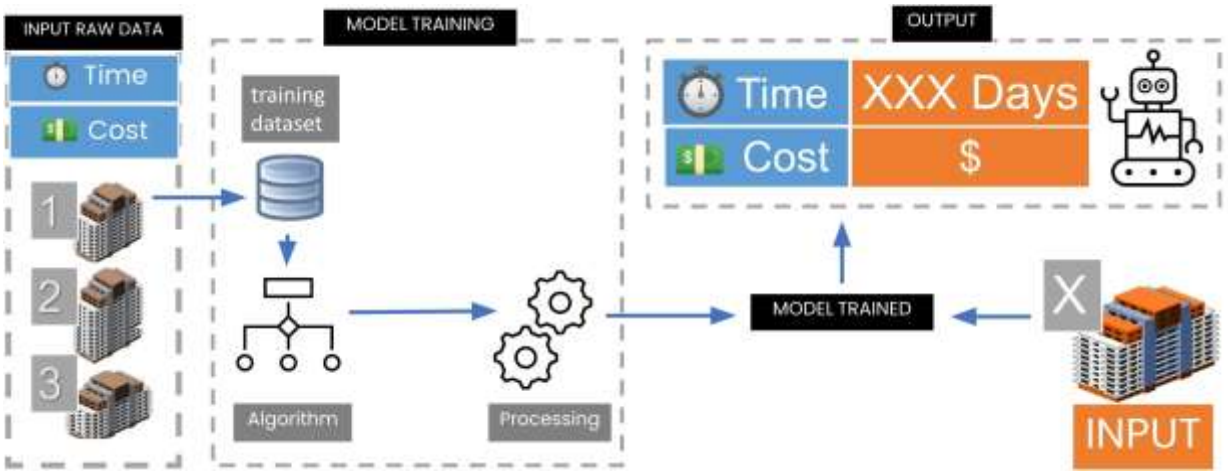


Fig. 9.2-14 I modsætning til klassisk formelbaseret estimering trænes maskinlæringsalgoritmen på historiske data.

I forbindelse med overvåget maskinlæring indeholder hvert projekt i træningsdatasættet både inputattributter (f.eks. omkostninger og tid for lignende bygninger) og forventede outputværdier (f.eks. omkostninger eller tid). Et lignende datasæt bruges til at oprette og tilpasse en maskinlæringsmodel (figur 9.2-15). Jo større datasættet er, og jo højere kvaliteten af dataene er, jo mere nøjagtig bliver modellen, og jo mere nøjagtige bliver forudsigelsesresultaterne.



Figur 9.2-15 En ML-model, der er trænet på omkostnings- og tidsdata fra tidligere projekter, vil med en vis sandsynlighed bestemme omkostningerne og tidsplanen for et nyt projekt.

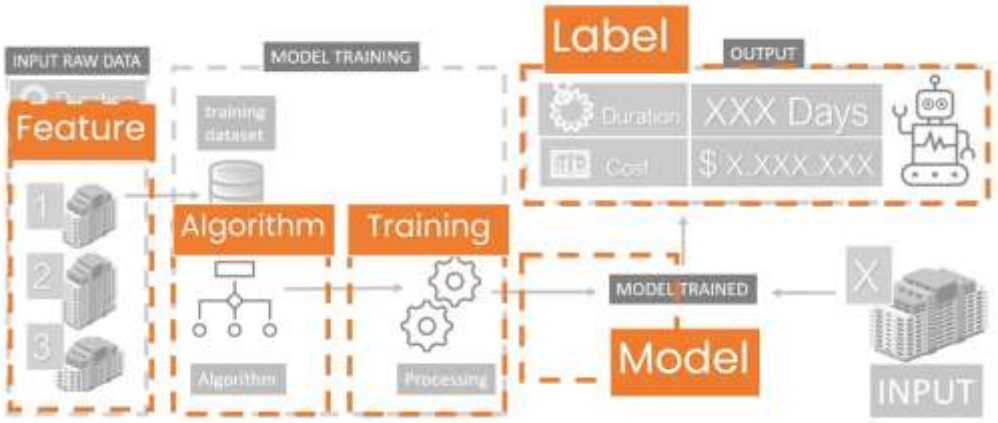
Når modellen er oprettet og trænet til at estimere opførelsen af et nyt projekt, skal man blot give modellen nye attributter for det nye projekt, og modellen vil med en vis sandsynlighed give estimerede resultater baseret på tidligere indlærte mønstre.

Nøglebegreber inden for maskinlæring

Maskinlæring er ikke magi, det er bare matematik, data og at finde mønstre. Den har ingen egentlig intelligens, men er et program, der er trænet i data til at genkende mønstre og træffe beslutninger uden konstant menneskelig indblanding.

Maskinlæring bruger en række nøglebegreber til at beskrive sin struktur (figur 9.2-16):

- **Labels** er målvariabler eller -attributter (parameteren "Survivor" i Titanic-datasættet), som modellen skal forudsige. Eksempel: byggeomkostninger (f.eks. i dollars), byggearbejdets varighed (f.eks. i måneder).
- **Funktioner** er uafhængige variabler eller attributter, der fungerer som input til modellen. I en prognosemodel bruges de til at forudsige etiketter. Eksempler: grundstørrelse (i kvadratmeter), antal etager i en bygning, bygningens samlede etageareal (i kvadratmeter), geografisk placering (bredde- og længdegrad), type af materialer, der bruges i byggeriet. Antallet af karakteristika bestemmer også dataenes dimensionalitet.
- **En model** er **et** sæt af forskellige hypoteser, hvoraf den ene tilnærmer sig den målfunktion, der skal forudsiges eller tilnærmes. Eksempel: maskinlæringsmodel, som bruger regressionsanalyseteknikker til at forudsige omkostninger og timing af byggeri.
- **Læringsalgoritme Læringsalgoritme** er processen med at finde den bedste hypotese i en model, der nøjagtigt matcher målfunktionen ved hjælp af et sæt træningsdata. Eksempel: En lineær regressions-, KNN- eller random forest-algoritme, der analyserer data om omkostninger og byggetid for at identificere relationer og mønstre.
- **Træning** - Under træningsprocessen analyserer algoritmen træningsdataene for at finde mønstre, der svarer til forholdet mellem inputattributterne og målmærkerne. Resultatet af denne proces er en trænet maskinlæringsmodel, der er klar til forudsigelse. Eksempel: En proces, hvor en algoritme analyserer historiske byggedata (omkostninger, tid, anlægskarakteristika) for at skabe en forudsigelsesmodel.



Figur 9.2-16 ML bruger etiketter og attributter til at skabe modeller, der trænes på data ved hjælp af algoritmer til at forudsige resultater.

Maskinlæring eksisterer ikke isoleret, men er en del af et bredere økosystem af analytiske discipliner, herunder statistik, databaser, datamining, mønstergenkendelse, big data-analyse og kunstig intelligens. Figur 9.2-17 viser, hvordan disse områder overlapper og supplerer hinanden, hvilket giver en omfattende ramme for moderne beslutnings- og automatiseringssystemer.

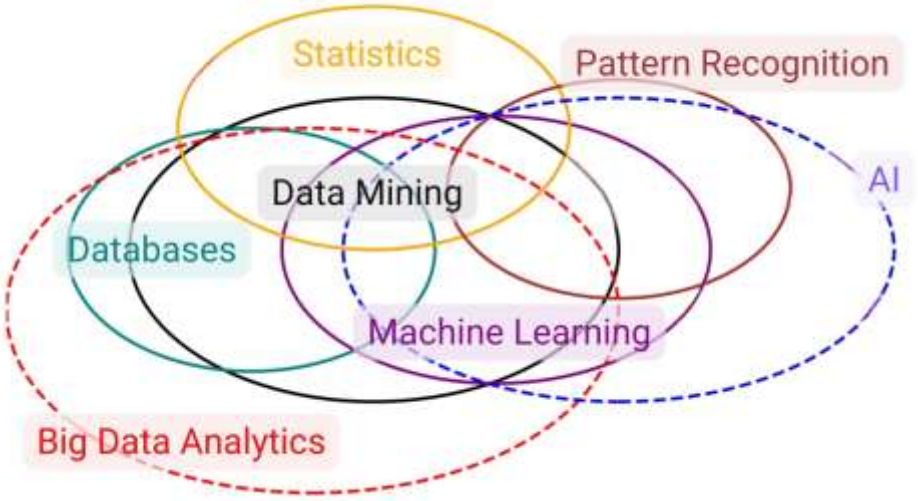


Fig. 9.2-17 Forholdet mellem de forskellige områder inden for dataanalyse: statistik, maskinlæring, kunstig intelligens, big data, mønstergenkendelse og datamining.

> Hovedformålet med maskinlæring er at udstyre computere med evnen til automatisk at lære viden uden menneskelig indgriben eller hjælp og justere deres handlinger i overensstemmelse hermed [159].

Så i fremtiden vil menneskets rolle kun være at give maskinen kognitive evner - de vil indstille betingelserne, vægtene og parametrene, og maskinlæringsmodellen vil gøre resten.

I det næste kapitel vil vi se på konkrete eksempler på anvendelse af algoritmen. Reelle tabeller og forenklede modeller vil blive brugt til at vise, hvordan prognosen opbygges trin for trin.

KAPITEL 9.3.
FORUDSIGELSE AF OMKOSTNINGER OG DEADLINES VED HJÆLP AF MASKINLÆRING

Et eksempel på brug af maskinlæring til at finde projektomkostninger og tidslinjer

Estimering af byggetid og -omkostninger er en af de vigtigste processer i en byggevirksomheds aktiviteter. Traditionelt er sådanne estimater blevet foretaget af eksperter baseret på erfaring, referencebøger og lovpligtige databaser. Men med den digitale transformation og den stigende datatilgængelighed er det nu muligt at bruge maskinlæringsmodeller (ML) til at forbedre nøjagtigheden og automatiseringen af sådanne estimater.

Indførelsen af maskinlæring i processen med at beregne omkostninger og tidsplan for byggeri giver ikke kun mulighed for mere effektiv planlægning, men bliver også et udgangspunkt for at integrere intelligente modeller i andre forretningsprocesser - fra risikostyring til optimering af logistik og indkøb.

> Det er vigtigt hurtigt at kunne afgøre, hvor lang tid det vil tage at bygge et projekt, og hvad de samlede omkostninger vil være. Disse spørgsmål om projekttid og -omkostninger har traditionelt været i højsædet hos både bygherrer og byggefirmaer siden byggeindustriens fødsel.

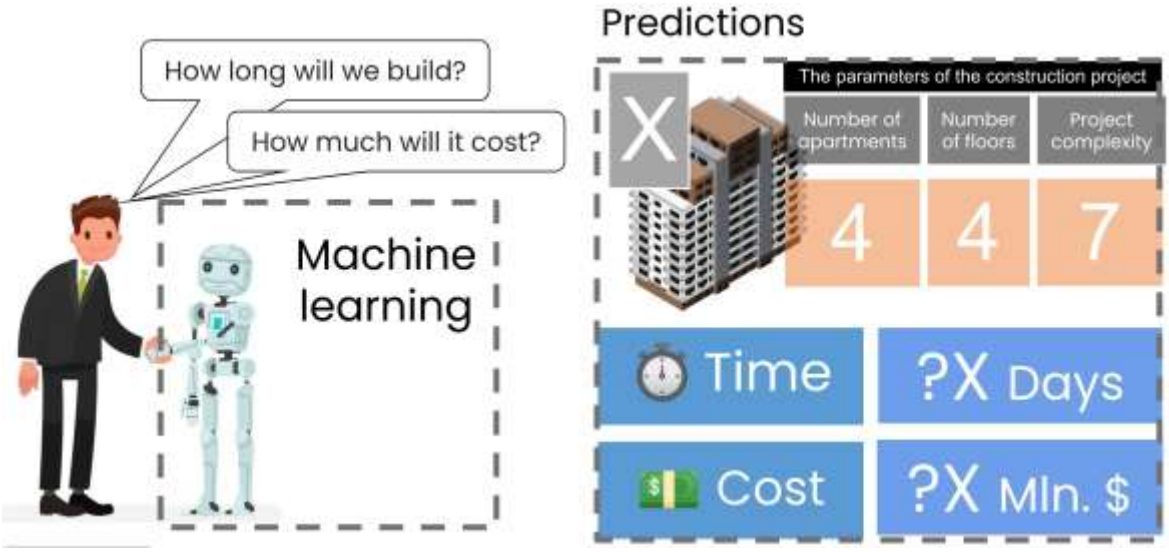


Fig. 9.3-1 I byggeprojekter er hastigheden og kvaliteten af estimeringen af byggetid og -omkostninger vigtige succesfaktorer.

I det følgende eksempel udtrækkes nøgledata fra tidligere projekter og bruges til at udvikle en maskinlæringsmodel, som giver os mulighed for at bruge modellen til at estimere omkostningerne og timingen for nye byggeprojekter med nye parametre (figur 9.3-1).

Overvej tre projekter med tre nøgleattributter: antallet af lejligheder (hvor 100 lejligheder svarer til tallet 10 for at lette visualiseringen), antallet af etager og et betinget mål for byggeriets kompleksitet på en skala fra 1 til 10, hvor 10 er den højeste kompleksitetsscore. Inden for maskinlæring kaldes processen med at konvertere og forenkle værdier som 100 til 10 eller 50 til 5 for "normalisering".

> **Normalisering** i maskinlæring er processen med at bringe forskellige numeriske data til en fælles skala for at lette behandling og analyse. Denne proces er især vigtig, når dataene har forskellige skalaer og måleenheder.

Antag, at det første projekt (figur 9.3-2) havde 50 lejligheder (efter normalisering, 5), 7 etager og en kompleksitetsscore på 2, hvilket betød en relativt enkel konstruktion. Det andet projekt havde allerede 80 lejligheder, 9 etager og et relativt komplekst design. Under disse forhold tog det 270 og 330 dage at bygge den første og den anden boligblok, og de samlede projektomkostninger var henholdsvis 4,5 og 5,8 millioner dollars.



Figur 9.3-2 Et eksempel på et sæt tidligere projekter, der skal bruges til at estimere tid og omkostninger for det fremtidige projekt X.

Når man bygger en maskinlæringsmodel til sådanne data, er hovedopgaven at identificere kritiske attributter (eller labels) til forudsigelse, i dette tilfælde byggetid og -omkostninger. Med et lille datasæt vil vi bruge oplysninger om tidligere byggeprojekter til at planlægge nye: Ved hjælp af maskinlæringsalgoritmer skal vi forudsige omkostningerne og byggetiden for et nyt projekt X baseret på givne attributter for det nye projekt, f.eks. 40 lejligheder, 4 etager og en relativ høj projektkompleksitet på 7 (fig. 9.3-2). I den virkelige verden kan antallet af inputparametre være meget større og variere fra flere dusin til hundredvis af faktorer. Disse kan omfatte: typen af byggematerialer, klimazone, entreprenørernes kvalifikationsniveau, tilgængelighed af forsyningsselskaber, fundamenttype, årstid for påbegyndelse af arbejdet, formændenes kommentarer osv.

> For at skabe en prædiktiv maskinlæringsmodel skal vi vælge en algoritme til at skabe den. En algoritme i maskinlæring er som en matematisk opskrift, der lærer computeren, hvordan den skal komme med forudsigelser (blande parametre i den rigtige rækkefølge) eller træffe beslutninger baseret på data.

For at analysere data om tidligere byggeprojekter og forudsige tid og omkostninger for fremtidige projekter (figur 9.3-2) kan en populær maskinlæringsalgoritme bruges:

- **Lineær regression (Lineær regression):** Denne algoritme forsøger at finde et direkte forhold mellem attributter, f.eks. mellem antallet af etager og byggeomkostningerne. Målet med algoritmen er at finde en lineær ligning, der bedst beskriver dette forhold, og som gør det muligt at lave forudsigelser.
- **Algoritme k-nærmeste naboer (k-NN):** Denne algoritme sammenligner et nyt projekt med tidligere projekter, der har samme størrelse eller kompleksitet. k-NN klassificerer data ud fra, hvilke af de k (antal) træningseksempler der er tættest på. I forbindelse med regression er resultatet gennemsnittet eller medianen af de k nærmeste naboer.
- **Beslutningstræer**: er en prædiktiv modelleringsmodel, der opdeler data i delmængder baseret på forskellige forhold ved hjælp af en træstruktur. Hver node i træet repræsenterer en betingelse eller et spørgsmål, der fører til yderligere opdeling af data, og hvert blad repræsenterer den endelige forudsigelse eller det endelige resultat. Algoritmen opdeler data i mindre grupper baseret på forskellige karakteristika, f.eks. først efter antal historier, derefter efter kompleksitet osv. for at komme med en forudsigelse.

Lad os se på maskinlæringsalgoritmer til at estimere omkostningerne ved et nyt projekt ved hjælp af to populære algoritmer som eksempler: lineær regression og K-nearest neighbours-algoritmen.

Forudsigelse af projektomkostninger og -tid ved hjælp af lineær regression

Lineær regression er en grundlæggende dataanalysealgoritme, der forudsiger værdien af en variabel baseret på et lineært forhold til en eller flere andre variabler. Denne model antager, at der er et direkte lineært forhold mellem den afhængige variabel og en eller flere uafhængige variabler, og målet med algoritmen er at finde dette forhold.

> Den lineære regressions enkelhed og klarhed har gjort den til et populært værktøj inden for mange forskellige områder. Når man har med en enkelt variabel at gøre, handler lineær regression om at finde den bedst passende linje gennem datapunkterne.

Lineær regression finder den bedste lige linje (rød linje), der tilnærmer sig afhængigheden mellem inputvariablen X og outputvariablen Y. Denne linje gør det muligt at forudsige Y-værdier for nye X-værdier baseret på det identificerede lineære forhold (figur 9.3-3).

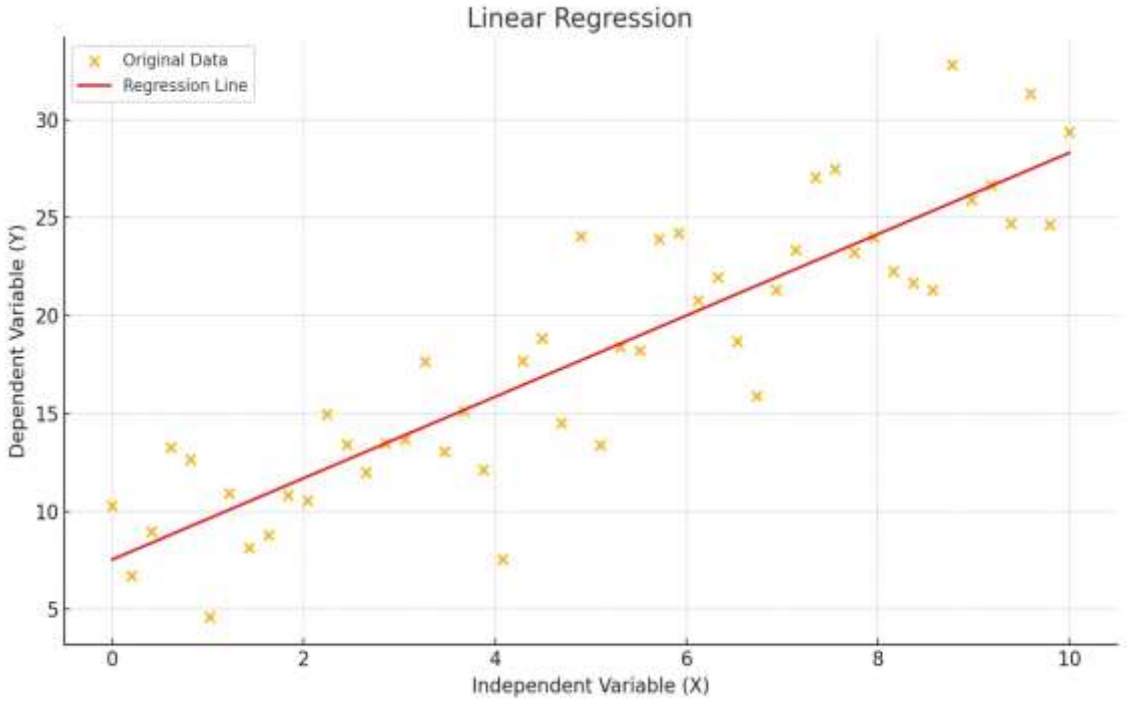


Fig. 9.3-3 Princippet i lineær regression er at finde den bedste lige linje gennem træningsværdierne.

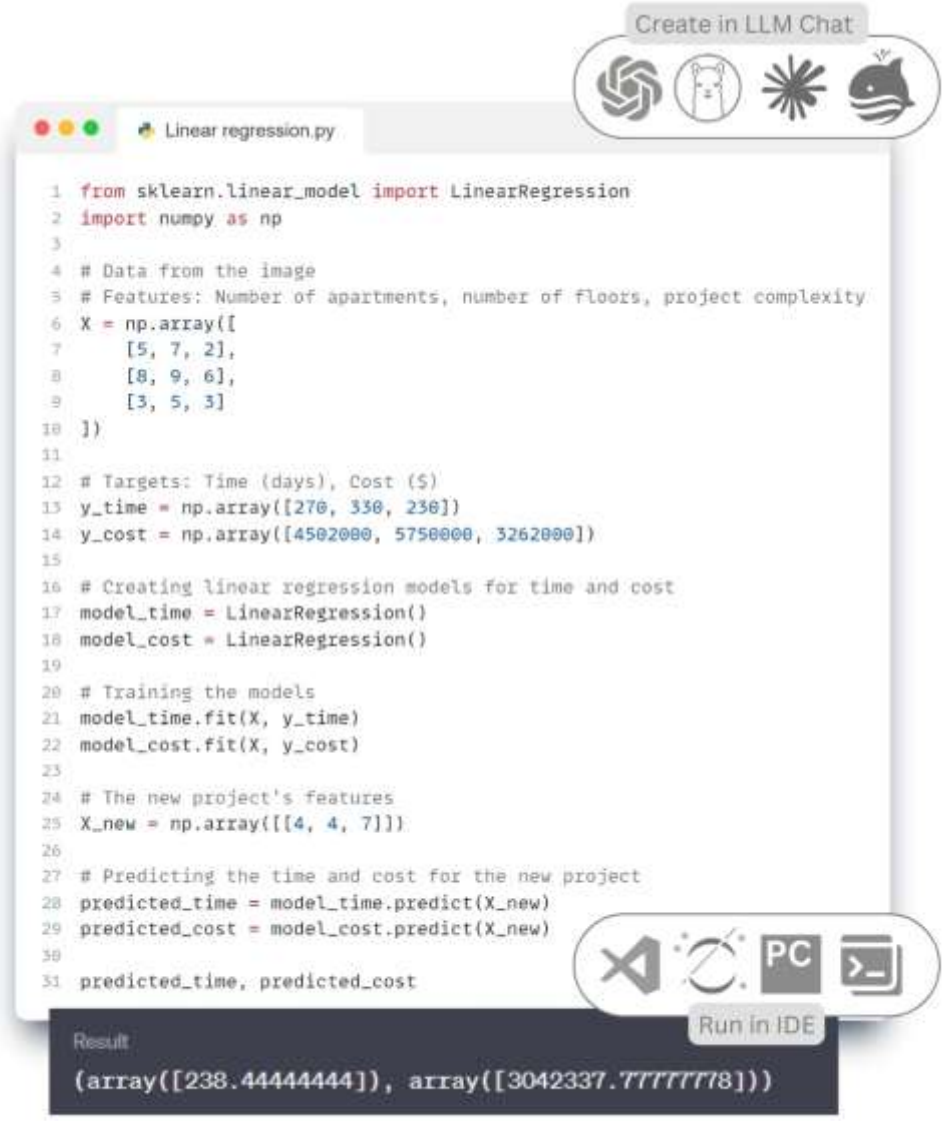
Denne linje er repræsenteret ved en ligning, hvor indtastning af en værdi af den uafhængige variabel (X) giver en forudsagt værdi af den afhængige variabel (Y). Denne proces gør det muligt effektivt at forudsige Y ud fra kendte værdier af X ved hjælp af et lineært forhold mellem dem. Et eksempel på at finde en sådan statistisk gennemsnitlig linje kan ses i San Francisco Building Permit Data Assessment (figur 9.1-7), hvor inflationen ved hjælp af lineær regression blev beregnet for forskellige typer af faciliteter.

Lad os indlæse projektdatatabellen (figur 9.3-2 fra forrige kapitel) direkte i LLM og bede den om at opbygge en simpel maskinlæringsmodel for os.

- Send en tekstanmodning til LLM chat (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

> Vi skal vise konstruktionen af en simpel maskinlæringsmodel til at forudsige omkostningerne og tiden til at realisere et nyt projekt X (Fig. 9.3-2 som vedhæftet billede). ⏎

- LLM's svar :



Figur 9.3-4 LLM valgte lineær regression til at skabe en maskinlæringsmodel til at forudsige projektomkostninger og -tid.

LLM genkendte automatisk tabellen fra det vedhæftede billede og konverterede dataene fra et visuelt format til et tabelarray (fig. 9.3-4 - række 6). Dette array blev brugt som grundlag for at skabe funktioner og etiketter, hvorfra der blev skabt en maskinlæringsmodel (Fig. 9.3-4 - 17.-22. række), som brugte lineær regression.

Ved hjælp af en grundlæggende lineær regressionsmodel, der blev trænet på et "ekstremt lille" datasæt, blev der lavet forudsigelser for et nyt hypotetisk byggeprojekt mærket Projekt X. I vores problem er dette projekt kendetegnet ved 40 lejligheder, 4 etager og et kompleksitetsniveau på 7 (figur 9.3-2).

Som forudsagt af en lineær regressionsmodel baseret på et begrænset og lille datasæt for det nye projekt X (figur 9.3-4 - linje 24-29):

- **Byggeperioden** vil være ca. 238 dage (238.4444444).
- **De samlede udgifter** vil være ca.$ 3,042,338 (3042337.777)

For yderligere at udforske hypotesen om projektomkostninger er det nyttigt at eksperimentere med forskellige maskinlæringsalgoritmer og -metoder. Lad os derfor forudsige de samme omkostnings- og tidsværdier for et nyt projekt X baseret på et lille sæt historiske data ved hjælp af K-Nearest Neighbours-algoritmen (k-NN).

Forudsigelse af projektomkostninger og -tid ved hjælp af K-nearest neighbour-algoritmen (k-NN)

Vi bruger k-Nearest Neighbours (k-NN) algoritmen som en ekstra prædiktor til at estimere omkostningerne og varigheden af et nyt projekt. K-Nearest Neighbours (k-NN)-algoritmen er en overvåget maskinlæringsmetode (supervised machine learning) til både klassificering og regression. Vi har også tidligere diskuteret k-NN-algoritmen i forbindelse med søgning i vektordatabaser (figur 8.2-2), hvor den bruges til at finde de tætteste vektorer (f.eks. tekster, billeder eller tekniske beskrivelser). I denne tilgang er hvert projekt repræsenteret som et punkt i et flerdimensionelt rum, hvor hver dimension svarer til en specifik attribut for projektet.

I vores tilfælde repræsenterer vi de tre attributter for hvert projekt som punkter i et tredimensionelt rum (figur 9.3-5). Vores kommende projekt X vil således blive lokaliseret i dette rum med koordinaterne (x=4, y=4, z=7). Det skal bemærkes, at i den virkelige verden kan antallet af punkter og rummets dimensionalitet være mange gange større.

> K-NN (k-nearest neighbours)-algoritmen fungerer ved at måle afstanden mellem det ønskede projekt X og projekterne i træningsdatabasen. Ved at sammenligne disse afstande bestemmer algoritmen de projekter, der er tættest på punktet for det nye projekt X.

Hvis det andet projekt (x=8, y=9, z=6) fra vores oprindelige datasæt f.eks. er meget længere væk fra X (figur 9.3-5) end de andre projekter, kan det udelukkes fra den videre analyse. Som følge heraf kan kun de to (k=2) tætteste projekter bruges til beregninger, på grundlag af hvilke gennemsnitsværdien vil blive bestemt.

En sådan metode gør det muligt at vurdere lighederne mellem projekter ved hjælp af en nabosøgning, hvilket igen hjælper med at drage konklusioner om de sandsynlige omkostninger og tidsplanen for et nyt projekt baseret på lignende projekter, der er blevet gennemført tidligere.

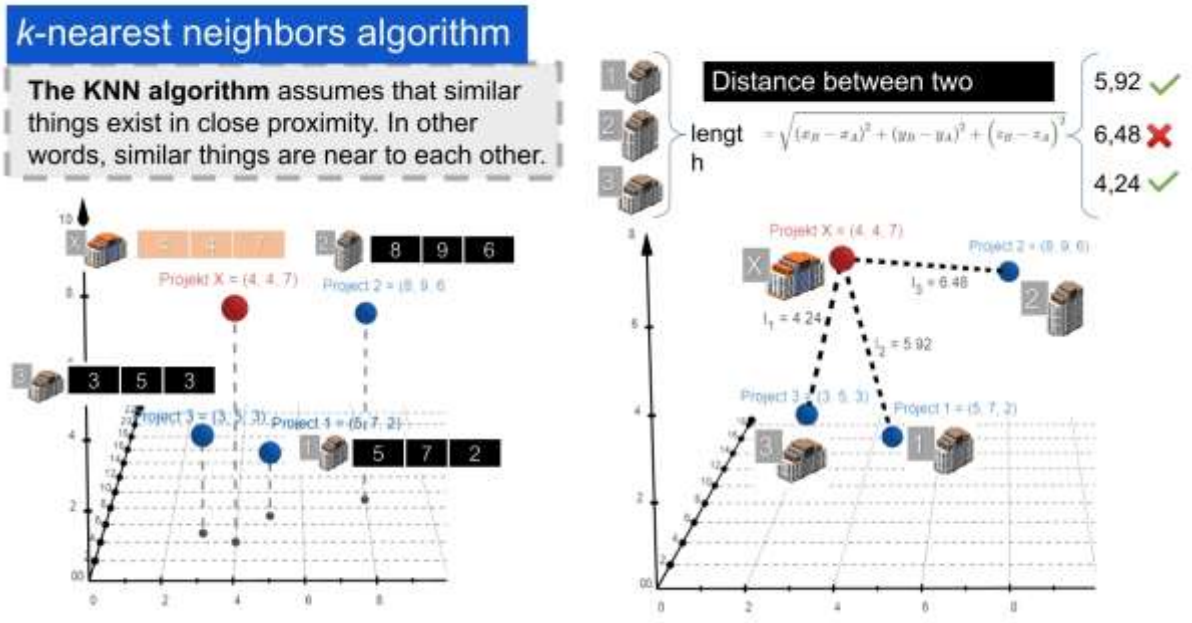


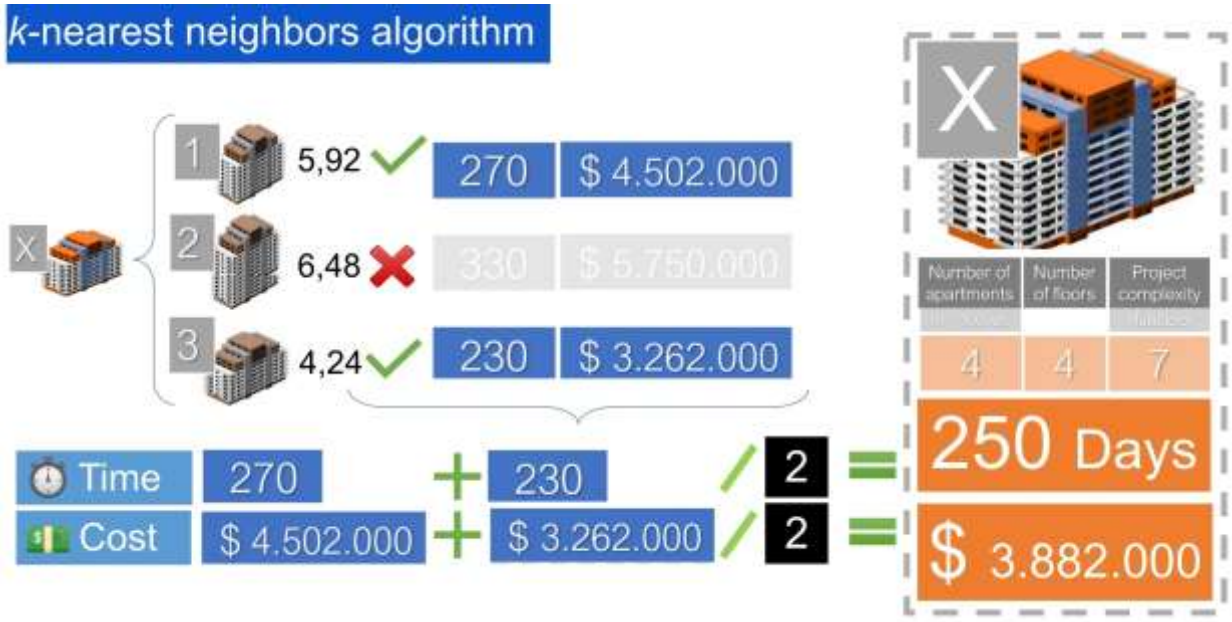
Fig. 9.3-5 I K-NN-algoritmen repræsenteres projekter som punkter i et flerdimensionelt rum, og projekter i nærheden udvælges på baggrund af afstande for at evaluere ligheden og lave forudsigelser.

Arbejdet med k-NN involverer flere vigtige trin:

- **Dataforberedelse:** trænings- og testdatasæt indlæses først. Træningsdata bruges til at "træne" algoritmen, og testdata bruges til at kontrollere dens effektivitet.
- **Valg af parameter K: Der** vælges et tal K, som angiver, hvor mange nærmeste naboer (datapunkter) der skal tages i betragtning i algoritmen. Værdien af "K" er meget vigtig, fordi den påvirker resultatet.
- Klassifikationsproces og regression for testdata:
  - **Beregning af afstande:** For hvert element fra testdataene beregnes afstanden til hvert element fra træningsdataene (figur 9.3-5). Der kan bruges forskellige afstandsmålemetoder til dette, f.eks. euklidisk afstand (den mest almindelige metode), Manhattan-afstand eller Hamming-afstand.
  - **Sortering og udvælgelse af K nærmeste naboer:** Efter beregning af afstandene sorteres de, og der udvælges K punkter, der er tættest på testpunktet.
  - **Bestemmelse af et testpunkts klasse eller værdi:** Hvis det er en klassificeringsopgave, bestemmes testpunktets klasse ud fra den hyppigste klasse blandt K udvalgte naboer. Hvis det er en regressionsopgave, beregnes gennemsnittet (eller et andet mål for den centrale tendens) af de K naboers værdier.
- **Afslutning af processen**: Når alle testdata er blevet klassificeret, eller der er lavet forudsigelser for dem, er processen afsluttet.

Algoritmen k-nearest neighbours (k-NN) er effektiv i mange praktiske anvendelser og er et af de vigtigste værktøjer i maskinlæringsspecialisternes arsenal. Denne algoritme er populær på grund af sin enkelhed og effektivitet, især i opgaver, hvor forholdet mellem data er let at fortolke.

I vores eksempel blev de to projekter (fra vores lille stikprøve) med den korteste afstand til projekt X identificeret efter anvendelse af K-nearest neighbours-algoritmen (figur 9.3-5). Baseret på disse projekter bestemmer algoritmen gennemsnittet af deres pris og byggevarighed. Efter analysen (fig. 9.3-6) konkluderer algoritmen ved at beregne gennemsnittet af de nærmeste naboer, at projekt X vil koste ca.$ $3.800.000 og tage ca. 250 dage at gennemføre.



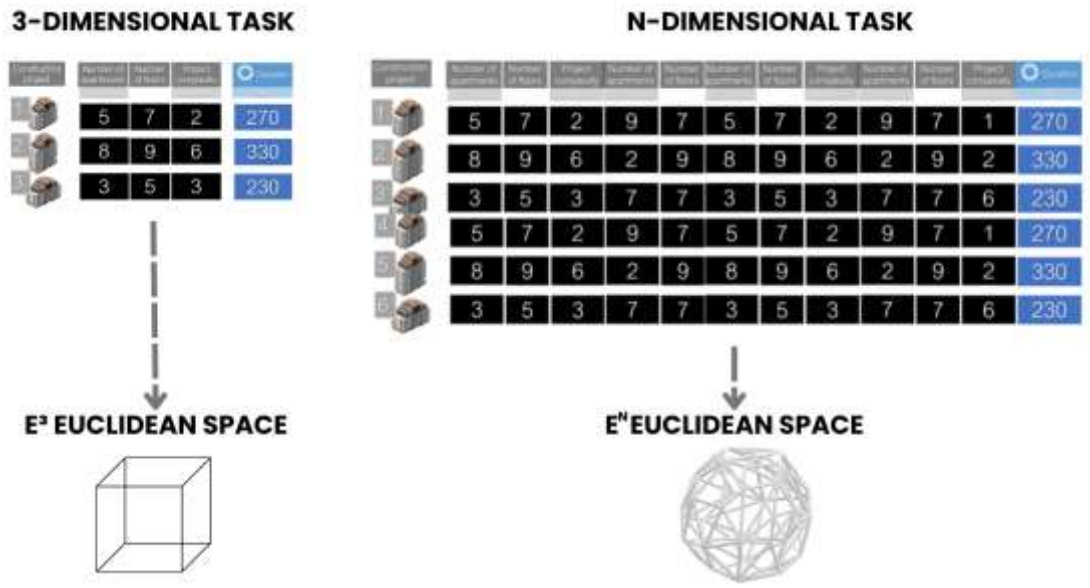
Figur 9.3-6 K-nearest neighbours-algoritmen bestemmer omkostningerne og tidsplanen for projekt X ved at analysere de to nærmeste projekter i stikprøven.

k-Nearest Neighbors (k-NN) algoritmen er især populær i klassificerings- og regressionsopgaver, såsom anbefalingssystemer, hvor den bruges til at foreslå produkter eller indhold baseret på præferencer, der svarer til en bestemt brugers interesser. Derudover bruges k-NN i vid udstrækning i medicinsk diagnostik til at klassificere typer af sygdomme baseret på patientsymptomer, i mønstergenkendelse og i den finansielle sektor til at vurdere kunders kreditværdighed.

Selv med begrænsede data kan maskinlæringsmodeller give nyttige forudsigelser og forbedre den analytiske del af byggeprojektledelsen betydeligt. Efterhånden som historiske data udvides og ryddes op, er det muligt at gå over til mere sofistikerede modeller - for eksempel ved at tage hensyn til konstruktionstype, placering, sæson for byggestart og andre faktorer.

Vores forenklede opgave brugte tre attributter til visualisering i 3D-rummet, men virkelige projekter indeholder i gennemsnit hundredvis eller tusindvis af attributter (se datasættet fra kapitlet "Et eksempel på

CAD-baserede big data (BIM)"), hvilket øger rummets dimensionalitet og kompleksiteten af at repræsentere projekter som vektorer betydeligt (figur 9.3-7).



Figur 9.3-7 I det forenklede eksempel blev der brugt tre attributter til 3D -visualisering, mens virkelige projekter har flere.

Anvendelse af forskellige algoritmer på det samme datasæt for projekt X, som har 40 lejligheder, 4 etager og kompleksitetsniveau 7, gav forskellige forudsagte værdier. Den lineære regressionsalgoritme forudsagde en færdiggørelsestid på 238 dage og en omkostning på$ 3.042.338 (figur 9.3-4), mens k-NN-algoritmen forudsagde 250 dage og$ 3.882.000 (figur 9.3-6).

Nøjagtigheden af forudsigelser, der opnås ved hjælp af maskinlæringsmodeller, afhænger direkte af mængden og kvaliteten af inputdataene. Jo flere projekter der er involveret i træningen, og jo mere fuldstændigt og præcist deres egenskaber (attributter) og resultater (etiketter) er repræsenteret, jo større er sandsynligheden for at opnå pålidelige forudsigelser med minimale fejlværdier.

Teknikker til forbehandling af data spiller en vigtig rolle i denne proces, herunder:

- Normalisering for at bringe funktioner til en fælles skala;
- Detektering og eliminering af afvigelser, hvilket eliminerer modelforvrængning;
- Kodning af kategoriske attributter for at muliggøre manipulation af tekstdata;
- Udfyldning af manglende værdier øger modellens robusthed.

Derudover bruges krydsvalideringsmetoder til at vurdere modellens generaliserbarhed og dens robusthed over for nye datasæt for at opdage overtilpasning og forbedre forudsigelsens pålidelighed.

> *Kaos er en orden, der skal afkodes [160].*
>
> - José Saramago, "The Double"

Selv om du synes, at kaoset i dine opgaver ikke kan beskrives formelt, skal du vide, at enhver begivenhed i verden og især byggeprocesser er underlagt matematiske love, som kan have brug for støtte til at beregne værdier, ikke gennem strenge formler, men ved hjælp af statistikker og historiske data.

Både traditionelle estimater fra omkostningsafdelinger og maskinlæringsmodeller er uundgåeligt behæftet med usikkerhed og potentielle fejlkilder. Men hvis der er tilstrækkelige kvalitetsdata til rådighed, kan maskinlæringsmodeller vise sammenlignelig og nogle gange endda højere forudsigelsesnøjagtighed end ekspertvurderinger.

> Maskinlæring vil sandsynligvis blive et pålideligt supplerende værktøj til analyser, der kan: forfine beregninger, foreslå alternative scenarier og identificere skjulte afhængigheder mellem projektparametre. Sådanne modeller vil ikke gøre krav på at være universelle, men de vil snart indtage en vigtig plads i beregninger og beslutningsprocesser. Maskinlæringsteknologier vil ikke udelukke deltagelse af ingeniører, estimatorer og analytikere, men vil tværtimod udvide deres muligheder ved at tilbyde et ekstra synspunkt baseret på historiske data.

Hvis maskinlæring integreres korrekt i byggefirmaernes forretningsprocesser, har den potentiale til at blive et vigtigt element i ledelsens beslutningsstøttesystem - ikke som en erstatning for mennesker, men som en udvidelse af deres professionelle intuition og ingeniørlogik.

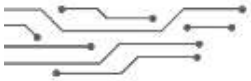
Næste skridt: fra opbevaring til analyse og prognoser

Moderne tilgange til at arbejde med data er begyndt at ændre den måde, man træffer beslutninger på i byggebranchen. At gå fra intuitive vurderinger til objektiv dataanalyse forbedrer ikke kun nøjagtigheden, men åbner også op for nye muligheder for at optimere processer. For at opsummere denne del er det værd at fremhæve de vigtigste praktiske trin, der vil hjælpe dig med at anvende de diskuterede metoder i dine daglige opgaver:

- Etablering af en bæredygtig lagerinfrastruktur
  - [ ] Prøv at kombinere forskellige dokumenter og projektdata i en enkelt tabelmodel, der samler vigtige oplysninger i en enkelt dataramme til videre analyse.
  - [ ] Brug effektive datalagringsformater - for eksempel kolonneformater som Apache Parquet i stedet for CSV eller XLSX - især til de sæt, der potentielt kan bruges til at træne maskinlæringsmodeller i fremtiden.
  - [ ] Etabler et dataversioneringssystem til at spore ændringer gennem hele projektet
- Implementering af analyse- og automatiseringsværktøjer
  - [ ] Begynd at analysere historiske projektdata - via dokumentation, modeller og estimater - for at identificere mønstre, tendenser og afvigelser.
  - [ ] Master ETL-processer (Extract, Transform, Load) til automatisk at indlæse og forberede data
  - [ ] Lær at visualisere nøgletal ved hjælp af forskellige gratis Python-visualiseringsbiblioteker
  - [ ] Begynde at anvende statistiske metoder og tilfældig prøveudtagning til at producere repræsentative og reproducerbare analytiske resultater.

- Øget modenhed i arbejdet med data
  - [ ] Lær et par grundlæggende maskinlæringsalgoritmer med enkle og ligefremme eksempler som Titanic-datasættet
  - [ ] Analyser de nuværende processer, og find ud af, hvor du kan gå fra rigid årsags- og virkningslogik til statistiske metoder til forudsigelse og estimering.
  - [ ] Begynd at behandle data som et strategisk aktiv i stedet for et biprodukt: Byg beslutningsprocesser op omkring datamodeller i stedet for omkring specifikke softwareløsninger.

Byggevirksomheder, der har indset værdien af data, er på vej ind i en ny udviklingsfase, hvor konkurrencefordele ikke afgøres af mængden af ressourcer, men af hastigheden af beslutningstagning baseret på analyser.



MAKSIMAL BEKVEMMELIGHED MED DEN TRYKTE VERSION

Du sidder med en gratis digital version af **Datadrevet byggeri**. For mere praktisk arbejde og hurtig adgang til materialerne anbefaler vi, at du er opmærksom på den **trykte udgave**:



- **Altid ved hånden**: Bogen i trykt format bliver et pålideligt arbejdsredskab, så du hurtigt kan finde og bruge de rigtige visualiseringer og diagrammer i enhver arbejdssituation.
- **Illustrationer i høj kvalitet**: alle billeder og grafikker i den trykte udgave præsenteres i maksimal kvalitet
- **Hurtig adgang til information**: nem navigation, mulighed for at lave noter, bogmærker og arbejde med bogen overalt.

Ved at købe den fulde trykte version af bogen får du et praktisk værktøj til komfortabelt og effektivt arbejde med information: muligheden for hurtigt at bruge visuelle materialer i hverdagens opgaver, hurtigt finde de nødvendige skemaer og tage noter. Desuden støtter dit køb udbredelsen af åben viden.

Bestil en trykt version af bogen på: datadrivenconstruction.io/books



X DEL

BYGGEBRANCHEN I DEN DIGITALE TIDSALDER. MULIGHEDER OG UDFORDRINGER

Den sidste tiende del er et omfattende kig på fremtiden for byggebranchen i en tid med digital transformation. Den analyserer skiftet fra kausalanalyse til arbejde med big data-korrelationer. Der drages paralleller mellem udviklingen af billedkunst og udviklingen af dataarbejde i byggebranchen, som viser, hvordan branchen bevæger sig fra detaljeret kontrol til en holistisk forståelse af processer. Begrebet "uberisering " af byggebranchen udforskes, hvor datatransparens og automatiserede beregninger radikalt kan ændre traditionelle forretningsmodeller, fjerne behovet for mellemmænd og reducere mulighederne for spekulation. Uafklarede spørgsmål, som f.eks. den universelle klassificering af elementer, diskuteres i detaljer, så byggevirksomhederne får tid til at tilpasse sig det nye miljø. Delen afsluttes med specifikke anbefalinger til udformning af en digital transformationsstrategi, der omfatter analyse af sårbarheder og udvidelse af tjenester for at forblive konkurrencedygtig i en branche i forandring.

KAPITEL 10.1.
OVERLEVELSESSTRATEGIER: OPBYGNING AF KONKURRENCEFORDELE

Korrelationer i stedet for beregninger: fremtiden for byggeanalyser

På grund af den hurtige digitalisering af information (figur 1.1-5) gennemgår moderne byggeri en grundlæggende transformation, hvor data ikke bare bliver et værktøj, men et strategisk aktiv, der fundamentalt kan ændre traditionelle tilgange til projekt- og forretningsstyring.

I tusindvis af år har byggeaktiviteter været baseret på deterministiske metoder - præcise beregninger, detaljering og streng kontrol af parametre. I de første århundreder e.Kr. anvendte romerske ingeniører matematiske principper til at bygge akvædukter og broer. I middelalderen stræbte arkitekter efter ideelle proportioner i gotiske katedraler, og med industrialiseringen i det 20. århundrede blev der dannet systemer med standardiserede normer og regler som blev grundlaget for massebyggeri.

I dag skifter udviklingsvektoren fra søgen efter rene årsags- og virkningsforhold til probabilistisk analyse, søgen efter korrelationer og skjulte mønstre. Branchen er på vej ind i en ny fase - data er ved at blive en nøgleressource, og analyser baseret på dem erstatter intuitive og lokalt optimerede tilgange.



Fig. 10.1-1 Byggedataenes skjulte potentiale: De eksisterende beregninger i virksomheden er kun toppen af isbjerget, som ledelsen kan analysere.

> En virksomheds informationssystem er som et isbjerg (fig. 10.1-1): Kun en lille del af datapotentialet er synligt for virksomhedens ledelse, mens den største værdi er skjult i dybet. Det er vigtigt ikke kun at evaluere data ud fra deres nuværende anvendelse, men også ud fra de muligheder, de vil åbne for i fremtiden. er de virksomheder, der lærer at uddrage skjulte mønstre og skabe ny indsigt fra data, der vil skabe en bæredygtig konkurrencefordel.

At finde skjulte mønstre og forstå data handler ikke kun om at arbejde med tal, men er en kreativ proces, der kræver abstrakt tænkning og evnen til at se hele billedet bag forskellige elementer. I den forstand kan udviklingen af arbejdet med data sammenlignes med udviklingen af billedkunsten (figur 10.1-2).

Udviklingen inden for byggeri minder meget om udviklingen inden for billedkunst. I begge tilfælde har menneskeheden udviklet sig fra primitive metoder til sofistikerede visualiserings- og analyseteknologier. I forhistorisk tid brugte mennesker huletegninger og primitive redskaber til at løse hverdagens opgaver. I løbet af middelalderen og renæssancen steg niveauet af sofistikeret arkitektur og kunst dramatisk. I den tidlige middelalder havde byggeværktøjerne udviklet sig fra en simpel økse til omfattende værktøjssæt, der symboliserede væksten i teknisk viden.

Realismens tidsalder var den første revolution inden for billedkunsten: Kunstnere lærte at gengive de mindste detaljer og opnå maksimal troværdighed. I byggeriet var modstykket til denne periode præcise ingeniørteknikker, detaljerede tegninger og strengt regulerede beregninger, som blev grundlaget for designpraksis i århundreder.

> Senere ændrede impressionismen selve opfattelsen af den kunstneriske virkelighed: I stedet for bogstaveligt at gengive formen begyndte kunstnerne at indfange stemningen, lyset og dynamikken med det formål at afspejle et helhedsindtryk snarere end absolut nøjagtighed. På samme måde bevæger maskinlæring inden for bygningsanalyse sig væk fra stive logiske modeller til mønstergenkendelse og probabilistiske mønstre, der gør det muligt at "se" skjulte afhængigheder i data, som er utilgængelige i klassisk analyse. Denne tilgang er et ekko af Bauhaus' ideer om minimalisme og funktionalitet, hvor mening (funktion) er vigtigere end form. Bauhaus forsøgte at fjerne det overflødige, at opgive ornamentering til fordel for klarhed, anvendelighed og masseappel. Tingene skulle være forståelige og nyttige uden overflødighed - æstetik blev født ud af logikken i design og formål.

Med fremkomsten af fotografiet i slutningen af det 19. århundrede fik kunsten et nyt værktøj til at indfange virkeligheden med en hidtil uset nøjagtighed og vendte op og ned på holdningen til billedkunst. På samme måde fører den industrielle revolution i det 21. århundrede til brug af robotteknologier, lasere, IoT, RFID og koncepter som Connected Construction, hvor indsamlingen af individuelle parametre har udviklet sig til en skalerbar intelligent indfangning af byggepladsens fulde virkelighed.

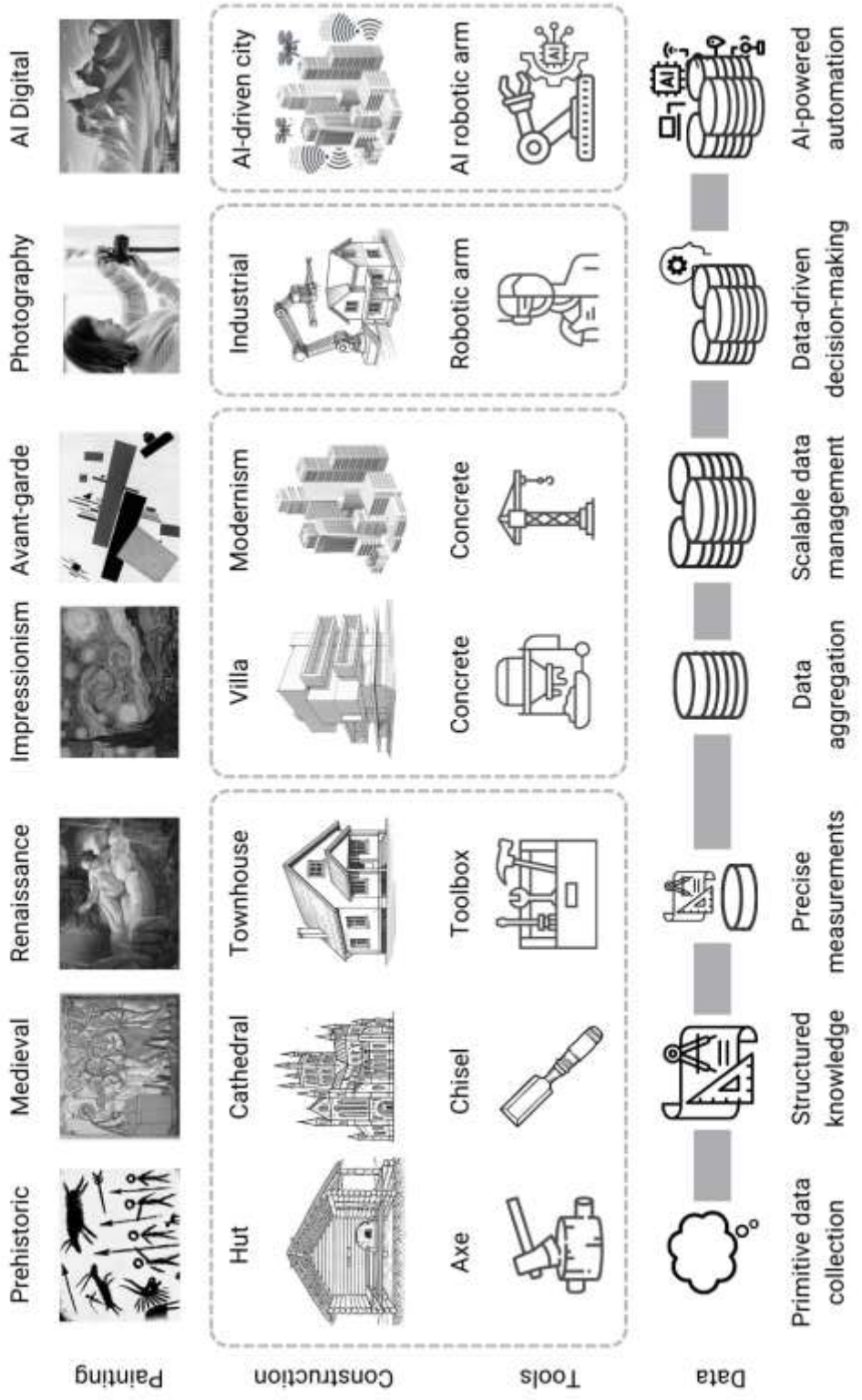


Fig. 10.1-2 Era of Fine Art Evolutions er i overensstemmelse med udviklingen i tilgangen til at arbejde med data i byggebranchen.

Ligesom billedkunsten oplever en nytænkning med ankomsten af AI-værktøjerne og LLM, oplever byggebranchen i dag endnu et kvantespring: Intelligente systemer drevet af kunstig intelligens (AI), LLM-chats gør det muligt at forudsige, optimere og generere løsninger med minimal menneskelig indgriben.

Datas rolle i design og ledelse har ændret sig radikalt. Hvor viden tidligere blev overført verbalt og empirisk - ligesom virkeligheden blev indfanget af håndmalede billeder indtil det 19. århundrede - er der i dag fokus på den komplette digitale indfangning af byggeriets "billede". Ved hjælp af maskinlæringsalgoritmer omdannes dette digitale billede til en impressionistisk repræsentation af byggeriets virkelighed - ikke en nøjagtig kopi, men en generaliseret, probabilistisk forståelse af processerne.

> Vi nærmer os hastigt en æra, hvor processerne med at designe, konstruere og drive bygninger ikke bare vil blive udvidet, men i høj grad drevet af systemer med kunstig intelligens. Ligesom moderne digital kunst skabes uden en pensel - ved hjælp af tekstbeskeder og generative modeller - vil fremtidens arkitektoniske og tekniske løsninger blive formet af nøgleforespørgsler og parametre, der indstilles af brugeren.

I det 21. århundrede er adgang til data, fortolkningen af dem og kvaliteten af analyserne ved at blive uundværlige for projektsucces. Og værdien af data bestemmes ikke af mængden, men af specialisternes evne til at analysere, verificere og omsætte dem til handling.

Datadrevet tilgang i byggeriet: et nyt niveau af infrastruktur

I menneskehedens historie har hvert af disse teknologiske spring medført grundlæggende ændringer i økonomien og samfundet. I dag er vi vidne til en ny bølge af forandringer, der i omfang kan sammenlignes med den industrielle revolution i det 19. århundrede. Men hvor det for 100 år siden var mekaniske kræfter og energiteknologi, der var den primære drivkraft bag forandringerne, er det nu data og kunstig intelligens.

Machine learning, LLM og AI agenter ændrer selve essensen af applikationer og gør traditionelle softwarestakke (diskuteret i anden del af bogen) overflødige (fig. 2.2-3). AI datalogik er centreret i AI-agenter snarere end i hårdt kodede forretningsregler (fig. 2.2-4).

> I dataalderen er det traditionelle syn på applikationer ved at blive fundamentalt forandret. Vi bevæger os mod en model, hvor store, modulopbyggede virksomhedssystemer uundgåeligt vil vige for åbne, lette, skræddersyede løsninger.

> *I fremtiden vil kun den underliggende datastruktur være tilbage, og al interaktion med den vil ske gennem agenter, der arbejder direkte med databasen. Jeg tror virkelig, at hele applikationsstakken vil forsvinde, fordi der simpelthen ikke er brug for den, når kunstig intelligens interagerer direkte med den underliggende database. Jeg har brugt hele min karriere på at arbejde i SaaS - opbygge virksomheder, arbejde i dem, og for at være ærlig ville jeg nok ikke starte en ny SaaS-virksomhed lige nu. Og jeg ville nok heller ikke investere i SaaS-virksomheder lige nu. Situationen er for usikker. Det betyder ikke, at der ikke vil være softwarevirksomheder i fremtiden, men de vil se meget anderledes ud. Fremtidens systemer vil være databaser med forretningslogik, der bringes ind i [AI]-agenter. Disse agenter vil arbejde med flere datalagre på samme tid og ikke være begrænset til en enkelt database. AI logik vil flytte ind i AI-laget [46].*
>
> - Matthew Berman, administrerende direktør Forward Future

Den vigtigste forskel i det nye paradigme er minimeringen af den teknologiske ballast. I stedet for monumentale, komplekse og lukkede softwaresystemer får vi fleksible, åbne og hurtigt tilpasselige moduler, der bogstaveligt talt "lever" i datastrømmen (fig. 7.4-1 - Apache Airflow, NiFi). Arkitekturen for fremtidens processtyring forudser brugen af mikroapplikationer - kompakte, specialbyggede værktøjer, der er fundamentalt forskellige fra massive og lukkede ERP-, PMIS-, CDE-, CAFM-systemer. De nye agenter vil være så tilpasningsdygtige, integrerede og forretningsspecifikke som muligt (f.eks. Low-Code/No-Code Fig. 7.4-6).

> *AI forretningslogik vil gå til disse [AI] agenter, og disse agenter vil udføre CRUD [Create, Read, Update, and Delete]-operationer på flere repositorier, hvilket betyder, at de ikke skelner mellem, hvilken backend der bruges. De vil opdatere flere databaser, og al logikken vil ende i det såkaldte AI -niveau. Og når AI-laget er der, hvor al logikken er, vil folk begynde at udskifte backends. Vi ser allerede en ret høj procentdel af markedsgevinster inden for Dynamics-backends og agentbrug, og vi vil bevæge os aggressivt i den retning og forsøge at bringe det hele sammen. Uanset om det er i kundeservice eller på andre områder, som ikke bare CRM, men vores finans- og driftsløsninger. Fordi folk vil have flere AI-drevne forretningsapplikationer, hvor det logiske lag kan drives af AI og AI-agenter. [...]. En af de mest spændende ting for mig er Excel med Python, som kan sammenlignes med GitHub med Copilot. Det vil sige, hvad vi har gjort: Når du har Excel, skal du bare åbne det, køre Copilot og begynde at lege med det. Det handler ikke længere bare om at forstå de tilgængelige tal - det laver selv en plan. Ligesom GitHub Copilot Workspace opretter en plan og derefter udfører den, er det som en dataanalytiker, der bruger Excel som et værktøj til at visualisere rækker og kolonner til analyse. Så Copilot bruger Excel som et værktøj med alle dets muligheder, fordi det kan generere data og har en Python-fortolker.*
>
> - Satya Nadella, CEO, Microsoft, interview med BG2-kanal december 2024. [28]

Den transformation, vi er vidne til i kontorapplikationernes logik - fra modulære, lukkede systemer til AI-agenter, der arbejder direkte med åbne data - er kun en del af en meget større proces. Det handler ikke kun om at ændre grænseflader eller softwarearkitektur: Ændringerne vil påvirke de grundlæggende principper

for arbejdstilrettelæggelse, beslutningstagning og virksomhedsledelse. I byggeriet vil det føre til en datadrevet logik, hvor data bliver det centrale element i processer fra design til ressourcestyring og byggeovervågning.

Den næste generation af det digitale kontor: Hvordan AI ændrer arbejdsområdet

For næsten et århundrede siden oplevede menneskeheden allerede en lignende teknologisk revolution. Overgangen fra dampmaskiner til elektriske motorer tog mere end fire årtier, men i sidste ende katalyserede den en hidtil uset produktivitetsvækst - primært gennem decentralisering af energikapaciteten og nye løsningers fleksibilitet. Dette skift ændrede ikke kun historiens gang og flyttede størstedelen af befolkningen fra land til by, men lagde også grunden til den moderne økonomi. Teknologiens historie er en rejse fra fysisk arbejde til automatisering og intelligente systemer. Ligesom traktoren erstattede dusinvis af jordbearbejdere, erstatter moderne digital teknologi traditionelle kontorbaserede byggestyringsmetoder (figur 10.1-3). Så sent som i begyndelsen af det 20. århundrede arbejdede størstedelen af verdens befolkning med håndkraft, indtil mekaniseringen af arbejdet med maskiner og traktorer begyndte i 1930'erne.

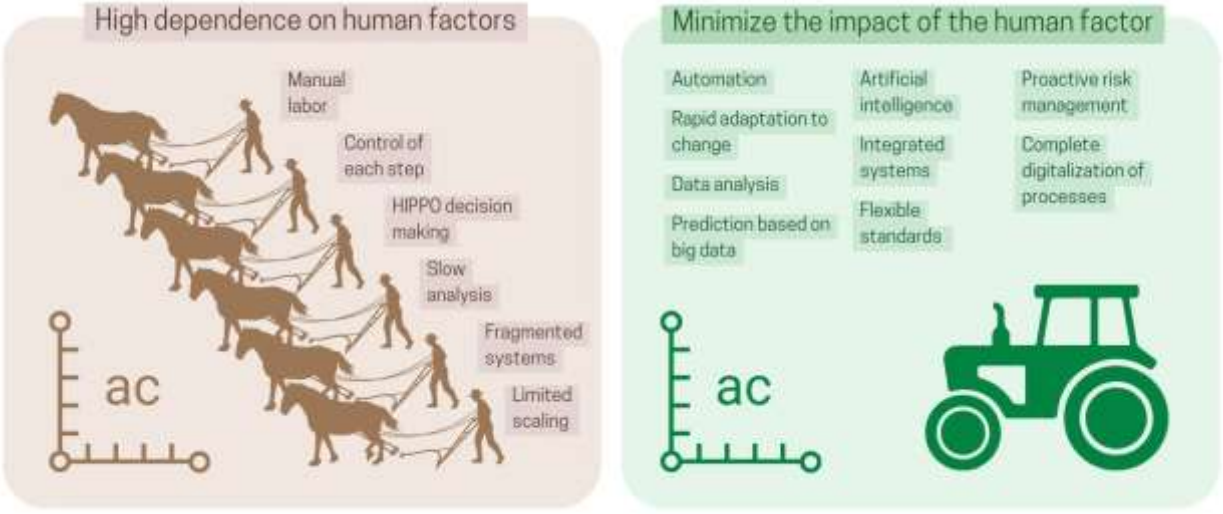


Fig. 10.1-3 Ligesom traktoren erstattede dusinvis af mennesker i begyndelsen af det 20. århundrede, vil maskinlæring erstatte traditionelle forretnings- og projektstyringsmetoder i det 21. århundrede.

Ligesom menneskeheden for hundrede år siden gik fra at dyrke individuelle jordlodder med primitive redskaber til storstilet landbrug med brug af maskiner, går vi i dag fra at behandle forskellige "siloer" af information til at arbejde med datamængder ved hjælp af kraftige "traktorer" - ETL-pipeline og algoritmer med kunstig intelligens.

Vi står på tærsklen til et lignende spring - men på det digitale plan: fra traditionel, manuel forretningsstyring til datadrevne modeller.

Vejen til en fuldt udbygget datadrevet arkitektur vil kræve tid, investeringer og en organisatorisk indsats. Men denne vej baner vejen for ikke bare en gradvis forbedring, men et kvalitativt spring mod større effektivitet, gennemsigtighed og håndterbarhed i byggeprocesserne. Alt dette forudsætter en systematisk

indførelse af digitale værktøjer og opgivelse af forældede forretningsmetoder.

> Opgaveparametrisering, ETL, LLM, IoT-komponenter, RFID, tokenisering, big **data** og maskinlæring vil omdanne traditionelt byggeri **til datadrevet byggeri**, hvor hver eneste detalje i projektet og byggevirksomheden vil blive kontrolleret og optimeret af data.

Det plejede at tage tusindvis af arbejdstimer at analysere information. Nu udføres disse opgaver af algoritmer og LLM'er, der forvandler forskellige datasæt til strategiske kilder ved hjælp af prompts. I teknologiverdenen sker det samme som i landbruget: Vi går fra hakken til automatiseret agribusiness. På samme måde er kontorarbejdet i byggeriet - fra Excel-filer og manuel opsummering - på vej til et intelligent system, hvor data indsamles, renses, struktureres og omdannes til indsigt.

Allerede i dag bør virksomheder begynde at "dyrke" informationsfelter ved hjælp af kvalitetsdataindsamling og informationsstrukturering og "gøde" dem med rengørings- og normaliseringsværktøjer og derefter "høste" dem i form af prædiktive analyser og automatiserede løsninger. Hvis en moderne landmand med en maskine er i stand til at erstatte hundrede jordbearbejdere, vil intelligente algoritmer være i stand til at fjerne rutinen fra medarbejderne og overføre dem til rollen som strategiske ledere af informationsstrømme.

Det er dog vigtigt at forstå, at det ikke er en hurtig proces at skabe en ægte datadrevet -organisation. Det er en langsigtet strategisk retning, der svarer til at skabe et nyt sted til dyrkning af en ny skov (fig. 1.2-5) af systemer, hvor hvert "træ" i dette økosystem er en separat proces, kompetence eller værktøj, der tager tid at vokse og udvikle. Som i tilfældet med en rigtig skov afhænger succesen ikke kun af plantematerialets kvalitet (teknologi), men også af jordbunden (virksomhedskultur), klimaet (forretningsmiljø) og pleje (systemtilgang).

Virksomheder vil ikke længere kunne stole udelukkende på lukkede, færdige løsninger. I modsætning til tidligere faser af den teknologiske udvikling vil den nuværende overgang - til åben dataadgang, kunstig intelligens og Open Source - sandsynligvis ikke blive støttet af store leverandører, da den direkte truer deres etablerede forretningsmodeller og centrale indtægtsstrømme.

Som det fremgår af Harvard Business School-undersøgelsen [40], som allerede er blevet diskuteret i kapitlet om den fjerde og femte teknologiske revolution, ville omkostningerne ved at skabe de mest anvendte Open Source-løsninger fra bunden for alle virksomheder være omkring 4,15 milliarder dollars. Men hvis vi forestiller os, at hver virksomhed ville udvikle sine egne alternativer uden adgang til eksisterende Open Source-værktøjer, hvilket er, hvad der er sket i de sidste årtier, kunne de samlede forretningsomkostninger nå op på kolossale 8,8 billioner dollars - dette er prisen på irrationel efterspørgsel, som softwaremarkedet kan værdisættes til.

Teknologiske fremskridt vil uundgåeligt føre til en gentænkning af etablerede forretningsmodeller. Tidligere kunne virksomheder tjene penge på komplekse, uigennemsigtige processer og lukkede data, men med udviklingen af AI og analytics bliver denne tilgang mindre og mindre levedygtig.

> Som følge af demokratiseringen af adgangen til data og værktøjer kan det traditionelle marked for softwaresalg skrumpe betydeligt. Men samtidig vil et nyt marked vokse - markedet for digital ekspertise, tilpasning, integration og løsningsdesign. Her vil værdien ikke komme fra licenssalg, men fra evnen til at opbygge fleksible, åbne og tilpasningsdygtige digitale processer. Ligesom elektrificeringen og fremkomsten af traktorer har skabt nye industrier, vil anvendelsen af big data, AI og LLM også åbne op for helt nye horisonter i byggebranchen, hvilket ikke kun vil kræve teknologiske investeringer, men også en dybtgående transformation af tankegange, processer og organisationsstrukturer. Og de virksomheder og fagfolk, der indser dette og begynder at handle i dag, vil være morgendagens ledere.

I en verden, hvor åbne data er ved at blive et stort aktiv, vil tilgængeligheden af information være en game changer. Investorer, kunder og lovgivere vil i stigende grad kræve gennemsigtighed, og maskinlæringsalgoritmer vil automatisk kunne identificere uoverensstemmelser i estimater, tidslinjer og omkostninger. Dette skaber forudsætningerne for en ny fase af digital transformation, som gradvist fører os til "uberisering " af byggebranchen.

Åbne data og uberisering er en trussel mod eksisterende byggevirksomheder

Byggeri er ved at blive en informationsstyringsproces. Jo mere nøjagtige, komplette og af høj kvalitet dataene er, desto mere effektivt bliver design, beregninger, omkostningsoverslag, opførelse og drift af bygninger. I fremtiden vil den vigtigste ressource ikke være en kran, beton og armeringsjern, men evnen til at indsamle, analysere og bruge information.

> I fremtiden vil byggefirmaernes kunder - investorer og bygherrer, der finansierer byggeri - uundgåeligt udnytte værdien af åbne data og analyser af historiske data. Det vil give mulighed for at automatisere beregningen af projektets tidslinjer og omkostninger uden at involvere byggevirksomhederne i omkostningsspørgsmål, hvilket vil hjælpe med at kontrollere omkostningerne og identificere overflødige omkostninger hurtigere.

Forestil dig en byggeplads, hvor laserscannere, quadrokoptere og fotogrammetrisystemer indsamler nøjagtige data i realtid om de anvendte betonmængder. Disse oplysninger konverteres automatisk til enkle flade MESH -modeller med metadata, uden om besværlige CAD (BIM) -systemer, uden afhængighed af komplekse geometriske kerner, ERP eller PMIS. Disse data, der indsamles fra byggepladsen, overføres centralt til et enkelt struktureret lager, der er tilgængeligt for kunden til uafhængig analyse, hvor reelle priser fra forskellige byggeforretninger uploades med parametre, der spænder fra kreditfinansieringssatsen til dynamisk skiftende faktorer som vejrforhold, børsnoteringer for byggematerialer, logistiktariffer og statistiske sæsonudsving i arbejdskraftpriser. Under sådanne forhold bliver eventuelle uoverensstemmelser mellem design og faktiske materialemængder øjeblikkeligt indlysende, hvilket gør det umuligt at manipulere med estimater både i designfasen og når projektet overdrages. Resultatet er, at gennemsigtigheden i byggeprocessen ikke opnås gennem en hær af tilsynsførende og ledere, men gennem objektive digitale data, som minimerer den menneskelige faktor og muligheden for spekulation.

I fremtiden vil denne form for datakontrolarbejde snarere blive udført af dataadministratorer på kundesiden (fig. 1.2-4 CQMS manager). Det gælder især for beregninger og projektestimater: Hvor der tidligere var en

hel afdeling med estimatorer, vil der i morgen allerede være maskinlæring og prognoseværktøjer, der sætter prisgrænser, som byggefirmaer skal passe ind i.

> *I betragtning af [bygge]industriens fragmenterede natur, hvor de fleste systemer og undersystemer leveres af SMV'er, skal den digitale strategi komme fra kunden. Kunderne skal skabe betingelserne og mekanismerne for at frigøre de digitale muligheder i forsyningskæden [20].*
>
> - Andrew Davis og Giuliano Denicol, Accenture "Skab mere værdi gennem kapitalprojekter"

En sådan åbenhed og gennemsigtighed i data udgør en trussel mod byggevirksomheder, som er vant til at tjene penge på uigennemsigtige processer og forvirrende rapporter, hvor spekulationer og skjulte omkostninger kan gemme sig bag komplekse og lukkede formater og modulære, proprietære dataplatforme. Derfor er det usandsynligt, at byggefirmaer, ligesom leverandører, der promoverer Open Source-løsninger, er interesserede i at implementere åbne data fuldt ud i deres forretningsprocesser. Hvis dataene er tilgængelige og nemme at behandle for kunden, kan de kontrolleres automatisk, hvilket vil eliminere muligheden for at overestimere mængder og manipulere estimater.

> Ifølge World Economic Forums rapport "Shaping the Future of Construction" (2016) [5] er en af de vigtigste udfordringer for branchen fortsat bygherrens passive rolle. Ikke desto mindre er det kunderne, der bør tage et større ansvar for resultatet af projekterne - fra tidlig planlægning til valg af bæredygtige interaktionsmodeller og overvågning af resultaterne. Uden projektejernes aktive deltagelse er en systemisk transformation af byggebranchen umulig.

Tabet af kontrol over mængde- og omkostningsberegninger har allerede transformeret andre brancher i løbet af de sidste 20 år, så kunderne kan direkte, uden mellemmænd, modvirke deres mål. Digitalisering og datatransparens har ændret mange traditionelle forretningsmodeller, f.eks. taxachauffører med fremkomsten af Uber (figur 10.1-4), hotelejere med fremkomsten af Airbnb og detailhandlere og butikker med fremkomsten af Amazon, og banker med fremkomsten af neo-banker og decentraliserede fintech-økosystemer, hvor direkte adgang information og automatisering af tids- og omkostningsberegninger har reduceret mellemmændenes rolle betydeligt.

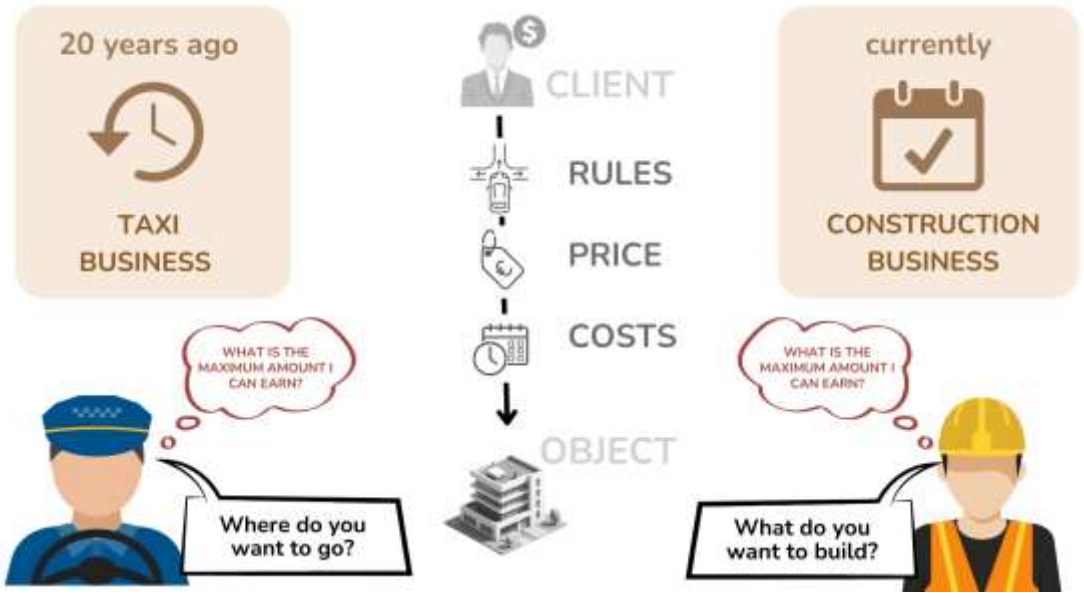


Fig. 10.1-4 Byggebranchen vil blive udsat for den uberisering, som taxachauffører, hotelejere og detailhandlere blev udsat for for 10 år siden.

Processen med at demokratisere adgangen til data og værktøjer til at behandle dem er uundgåelig, og med tiden vil åbne data om alle projektkomponenter blive et kundekrav og en ny standard. Derfor vil spørgsmålet om at indføre åbne formater og gennemsigtige beregninger blive fremmet af investorer, kunder, banker og private equity-fonde (private equity) - dem, der er slutbrugere af de opførte objekter og derefter driver objektet i årtier.

> Store investorer, kunder og banker kræver allerede gennemsigtighed i byggebranchen. Ifølge Accentures undersøgelse "Creating more value through capital projects" (2020) [20] er gennemsigtige og pålidelige data ved at blive en afgørende faktor for investeringsbeslutninger i byggeriet. Som eksperter påpeger, er troværdig og effektiv projektstyring umulig uden gennemsigtighed, især i krisetider. Derudover bevæger ejere af aktiver og entreprenører sig i stigende grad mod kontrakter, der tilskynder til datadeling og samarbejdsanalyser, hvilket afspejler de voksende krav fra investorer, banker og tilsynsmyndigheder om ansvarlighed og gennemsigtighed.

> Investorens, kundens bevægelse fra idé til færdigt byggeri vil i fremtiden være som at køre på autopilot - uden en chauffør i form af et byggefirma, lover at blive uafhængig af spekulation og usikkerhed.

Æraen med åbne data og automatisering vil uundgåeligt ændre byggebranchen, ligesom den allerede har gjort det inden for bankvæsen, handel, landbrug og logistik. I disse brancher er mellemmændenes rolle og de traditionelle måder at drive forretning på ved at blive afløst af automatisering og robotisering, hvilket ikke giver plads til uberettigede avancer og spekulation.

Data og processer i alle menneskelige økonomiske aktiviteter adskiller sig ikke fra det, som fagfolk i byggebranchen skal forholde sig til. På lang sigt kan byggefirmaer, som i dag dominerer markedet ved at sætte standarder for pris og servicekvalitet, miste deres rolle som et vigtigt mellemled mellem kunden og

deres byggeprojekt.

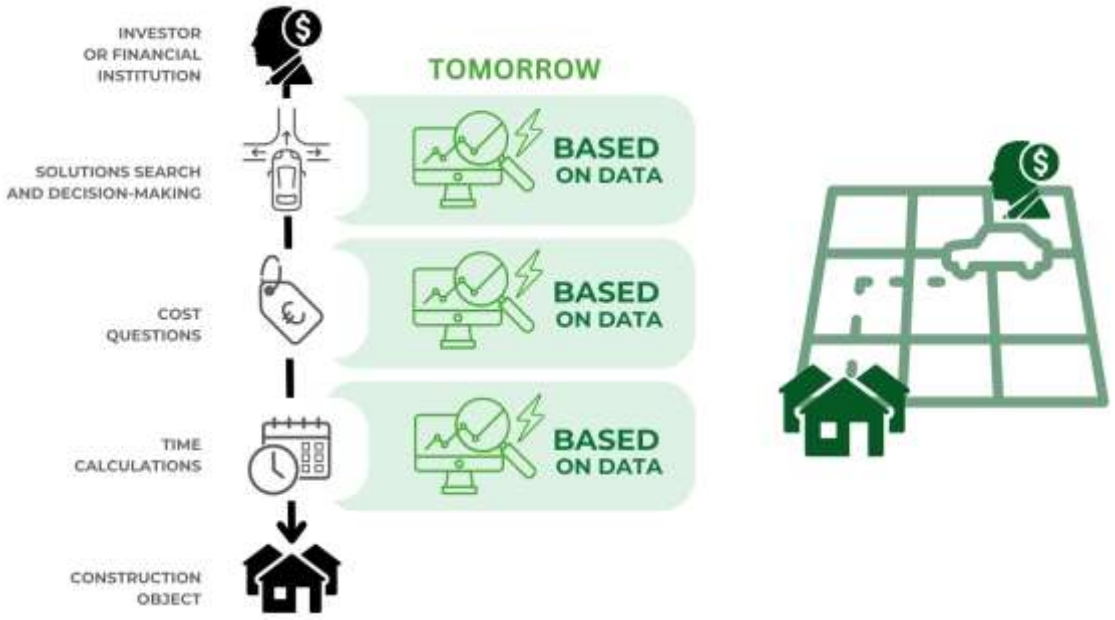
Uløste problemer med uberisering som sidste chance for at bruge tiden til transformation

Men lad os vende tilbage til realiteterne i byggebranchen. Mens selvkørende biler, decentraliserede finansielle systemer og løsninger baseret på kunstig intelligens vinder frem i nogle sektorer af økonomien, er en betydelig del af byggevirksomhederne stadig papirbaserede organisationer, hvor vigtige beslutninger i højere grad træffes på grundlag af de enkelte specialisters intuition og erfaring.

I dette paradigme kan et moderne byggefirma sammenlignes med et 20 år gammelt taxaselskab, der kontrollerer ressourcer, ruter og leveringstider og er ansvarlig for timingen og omkostningerne ved "rejsen" - fra projektideen (logistik- og installationsprocessen) til leveringen af projektet. Ligesom GPS (i byggeriets IoT, RFID) og maskinlæringsalgoritmer i tids-/omkostningsberegninger engang transformerede transportområdet, har data, algoritmer og AI -agenter potentiale til at transformere byggeledelse - fra intuitive vurderinger til forudsigelige, guidede modeller. I løbet af de sidste 20 år har mange brancher - finans, landbrug, detailhandel og logistik - gradvist elimineret muligheden for at spekulere på grund af uigennemsigtige data. Priser, leveringsomkostninger eller finansielle transaktioner beregnes automatisk og statistisk forsvarligt - på få sekunder på digitale platforme.

Når man ser på fremtiden, må byggevirksomhederne erkende, at en demokratisering af adgangen til data og værktøjerne til at analysere dem vil forstyrre den traditionelle tilgang til estimering af projektomkostninger og tidsplaner og fjerne muligheden for at spekulere i uigennemsigtige mængde- og prisdata.

> Ligesom når man kører på en reguleret vej uden indblanding fra føreren, vil fremtidens byggeprocesser i stigende grad ligne et "Uberiseret" system - med automatiseret tids- og omkostningsestimering, gennemsigtig routing af opgaver og minimal afhængighed af menneskelige faktorer. Det vil ændre selve karakteren af "rejsen" fra idé til realisering - og gøre den mere forudsigelig, håndterbar og datadrevet.



Figur 10.1-5 Omkostninger og "rejsetid" under byggeriet bestemmes ved hjælp af maskinlæring og statistiske værktøjer.

Med den gradvise indførelse af nye regler og krav i næsten alle lande i verden, som forpligter til at overføre CAD- (BIM)-modeller til bygherrer eller banker, der finansierer byggeprojekter, har bygherren og kunden mulighed for selvstændigt at sikre gennemsigtighed i omkostnings- og omfangskalkulationer. Dette er især vigtigt for store kunder og investorer, som har tilstrækkelige kompetencer og værktøjer til at analysere mængder hurtigt og overvåge markedspriser. For virksomheder, der gennemfører store standardprojekter - butikker, kontorbygninger, boligkomplekser - er en sådan praksis ved at blive standard.

Efterhånden som informationsindholdet i modellerne bliver mere komplet og standardiseret, forsvinder muligheden for manipulation og spekulation stort set. Den digitale transformation ændrer gradvist spillereglerne i byggebranchen, og virksomheder, der ikke tilpasser sig disse ændringer, kan komme til at stå over for alvorlige udfordringer.

> Øget konkurrence, teknologiske forstyrrelser og faldende marginer har potentiale til at påvirke virksomhedernes bæredygtighed. Med begrænset likviditet vender flere og flere aktører i branchen sig mod automatisering, analyse og datateknologi som en måde at forbedre effektiviteten og procesgennemsigtigheden på. Disse værktøjer bliver en vigtig ressource for at forblive konkurrencedygtig i et skiftende økonomisk miljø.

Måske skal vi ikke vente på, at ydre omstændigheder tvinger os til at handle hurtigt - det er meget mere effektivt at begynde at forberede sig i dag ved at styrke de digitale kompetencer, implementere moderne løsninger og opbygge en datacentreret kultur.

En af de sidste vigtige teknologiske barrierer for den store digitale transformation af byggebranchen, som vil påvirke alle virksomheder i de kommende år, er problemet med automatisk klassificering af elementer i byggeprojekter.

> Uden pålidelig, nøjagtig og skalerbar klassificering er det umuligt at skabe grundlag for fuldgyldige analyser, procesautomatisering og livscyklusstyring ved hjælp af AI og forudsigelige modeller. Så længe objektklassificering stadig afhænger af manuel fortolkning af erfarne specialister - formænd, designere, estimatorer - har byggebranchen stadig et vindue af muligheder. Denne tid kan bruges til at forberede sig på de uundgåelige ændringer: stigende krav om gennemsigtighed, demokratisering af værktøjer og data og fremkomsten af automatiske klassifikationssystemer, der vil ændre spillereglerne radikalt.

Opgaven med automatisk klassificering af elementer i byggeriets verden kan i sin kompleksitet sammenlignes med objektgenkendelse i ubemandede køresystemer, som er en af de største udfordringer. Lad os forestille os en ubemandet bil, der kører fra punkt A til punkt B (fig. 10.1-5). Nuværende automatiske køresystemer er kørt fast i problemet med at klassificere objekter, der genkendes af lidarer og kameraer. Det er ikke nok for en bil blot at "se" en forhindring eller et landemærke, men den skal kunne genkende, om det er en fodgænger, et vejskilt eller en skraldespand.

Hele byggebranchen står over for en lignende grundlæggende udfordring. Projektelementer - som f.eks. vinduer, døre eller søjler - kan registreres i dokumentation, repræsenteres i CAD-modeller, fotograferes på byggepladsen eller genkendes i punktskyer fra laserscanning. Men deres visuelle eller grove geometriske genkendelse er ikke nok til at opbygge et ægte automatiseret projektstyringssystem. Det er nødvendigt at sikre, at hvert element er nøjagtigt og konsekvent kategoriseret i en type, der kan identificeres entydigt i alle efterfølgende processer - fra overslag og specifikationer til logistik, lagerstyring og vigtigst af alt, drift (fig. 4.2-6).

Det er på dette stadie - overgangen fra genkendelse til meningsfuld klassificering - at en af de vigtigste forhindringer opstår. Selv om digitale systemer teknisk set er i stand til at skelne mellem og identificere objekter i modeller og på byggepladsen, ligger den største vanskelighed i den korrekte og kontekstuelt stabile definition af elementtypen for forskellige softwaremiljøer.. For eksempel kan en dør være mærket af designeren i en CAD-model som et element i kategorien "dør", men når den overføres til et ERP- eller PMIS-system, kan den være skrevet forkert, enten på grund af en fejl fra designerens side eller på grund af uoverensstemmelser mellem systemerne. Desuden mister elementet ofte nogle vigtige attributter eller forsvinder fra systemregnskabet under dataeksport og -import. Det fører til et hul i dataflowet og underminerer princippet om end-to-end-digitalisering af byggeprocesser. Det skaber en kritisk kløft mellem "synlig" og "forståelig" semantisk betydning, som underminerer dataintegriteten og i høj grad komplicerer automatiseringen af processer gennem hele byggeprojektets livscyklus.

Løsningen på udfordringen med universel klassificering af bygningselementer ved hjælp af big data og maskinlæringsteknologier (fig. 10.1-6) vil være en katalysator for transformation i hele branchen - og måske en uventet opdagelse for mange byggevirksomheder. Et samlet, lærbart klassifikationssystem vil være grundlaget for skalerbar analyse, digital ledelse og indførelse af AI i byggeorganisationernes daglige praksis.

NVIDIA og andre teknologiledere tilbyder allerede løsninger i andre brancher, som automatisk kan kategorisere og strukturere store mængder tekst og visuel information.

> NVIDIA's NeMo Curator-model [161] er f.eks. specialiseret i automatisk klassificering og kategorisering af data i foruddefinerede kategorier, hvilket spiller en vigtig rolle i optimeringen af informationsbehandlingspipelines til finjustering og prætræning af generative AI-modeller. Cosmos-platformen er trænet på virkelige video- og 3D-scener [162], hvilket giver et fundament for autonome systemer og digitale tvillinger, som allerede er ved at blive bygget i NVIDIA-økosystemet. NVIDIA Omniverse, som i 2025 er blevet det førende værktøj til at arbejde med USD-formatet, en universel scenebeskrivelse, der i sidste ende kan erstatte IFC-formatet i overførselsprocesser for designinformation. Sammen med Isaac Sim, en robotprocessimulator [163], repræsenterer løsninger som NeMo Curator, Cosmos og Omniverse et nyt niveau af automatisering: fra datarensning og -filtrering til generering af træningssæt, modellering af objektegenskaber og robottræning på byggepladsen. Og alle disse værktøjer er gratis og open source, hvilket reducerer barriererne for indførelse i ingeniør- og byggepraksis betydeligt.

Automatisk klassificering af data på niveau med strukturerede tabeller er ikke så vanskelig, som det kan se ud ved første øjekast. Som vi har vist i det foregående kapitel (fig. 9.1-10), er det muligt at kompensere for manglende eller forkerte klasseværdier på basis af lignende parametre for andre elementer, hvis der er akkumuleret historiske data. Hvis elementer med lignende egenskaber allerede er blevet klassificeret korrekt i flere afsluttede projekter, kan systemet med stor sandsynlighed foreslå en passende værdi for et nyt eller ufuldstændigt element (fig. 10.1-6). En sådan logik, der er baseret på gennemsnitsværdier og kontekstanalyse, kan være særlig effektiv ved massebehandling af tabeldata, der kommer fra estimater, specifikationer eller CAD-modeller.

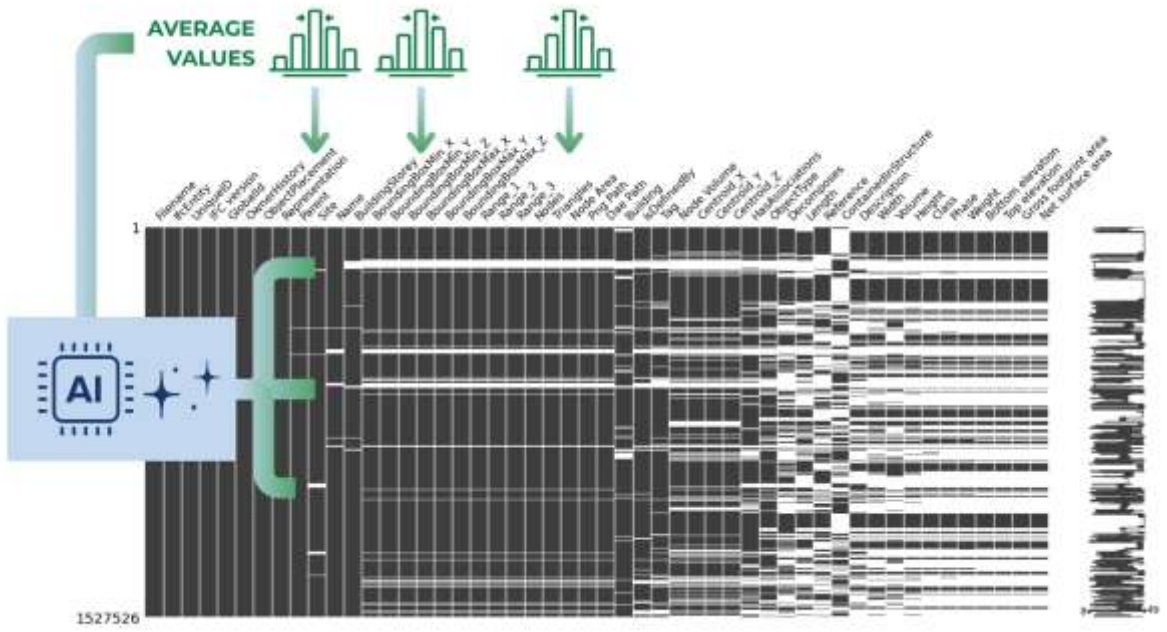
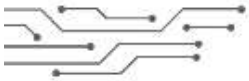


Fig. 10.1-6 Machine learning kan hjælpe dig med automatisk at finde gennemsnitsværdier for uudfyldte (hvide felter) tabelparametre baseret på tidligere projekter.

> På baggrund af disse hurtige fremskridt inden for maskinlæring står det klart: I 2025 er det naivt at tro, at problemet med automatisk klassificering af bygningselementer vil forblive uløst i lang tid. Ja, de nuværende algoritmer har endnu ikke nået fuld modenhed, især med ufuldstændige eller heterogene data, men vinduet med muligheder for tilpasning lukker hurtigt.

Virksomheder, der allerede investerer i at indsamle, rense og organisere deres data samt anvender ETL-automatiseringsværktøjer, vil have en fordel. Resten risikerer at sakke bagud - ligesom det engang mislykkedes for virksomheder at klare udfordringerne ved den digitale transformation i transport- og finanssektoren

De, der fortsætter med at stole på manuel datahåndtering og traditionelle metoder til omkostnings- og tidsestimering, risikerer at befinde sig i samme situation som taxaflåderne i 2000'erne, der ikke var i stand til at tilpasse sig æraen med mobilapps og automatiserede ruteberegninger i begyndelsen af 2020'erne.



KAPITEL 10.2.
EN PRAKTISK GUIDE TIL IMPLEMENTERING AF EN DATADREVET TILGANG

Fra teori til praksis: En køreplan for digital transformation i byggeriet

Byggebranchen er gradvist på vej ind i en ny udviklingsfase, hvor velkendte processer i stigende grad suppleres - og nogle gange endda erstattes - af digitale platforme og transparente interaktionsmodeller. Det giver ikke kun virksomhederne udfordringer, men også store muligheder. De organisationer, der allerede i dag er i gang med at opbygge en langsigtet digital strategi, vil ikke kun kunne fastholde deres position på markedet, men også udvide den ved at tilbyde kunderne moderne tilgange og pålidelige, teknologisk understøttede løsninger.

> Det er vigtigt at indse, at viden om koncepter og teknologier kun er et udgangspunkt. Ledere og specialister står over for et praktisk spørgsmål: Hvor skal man starte implementeringen, og hvordan omsætter man teoretiske ideer til reel værdi. Derudover opstår spørgsmålet i stigende grad: Hvad skal forretningen baseres på, hvis traditionelle metoder til omkostningsberegning og timing til enhver tid kan revideres af kunden.

Svaret ligger nok ikke så meget i teknologien, men i dannelsen af en ny professionel kultur, hvor arbejdet med data opfattes som en integreret del af den daglige praksis. Det er den manglende opmærksomhed på digital teknologi og innovation, der har ført byggebranchen ind i den alvorlige tilbagestående, der er blevet observeret i løbet af de sidste årtier [43].

> Ifølge McKinsey er F&U-udgifterne i byggebranchen mindre end 1% af omsætningen, mens tallet er 3,5-4,5% i bil- og luftfartsindustrien. På samme måde ligger IT-omkostningerne i byggebranchen på under 1% af den samlede omsætning [107].

Som følge heraf falder ikke kun automatiseringsniveauet, men også arbejdsproduktiviteten i byggeriet, og i 2020 vil bygningsarbejderen producere mindre end for et halvt århundrede siden (Fig. 10.2-1)

Sådanne produktivitetsproblemer i byggesektoren er fælles for de fleste udviklede lande og udviklingslande (produktiviteten i byggeriet er faldet i 16 ud af 29 OECD-lande (figur 2.2-1)) og peger ikke kun på manglende teknologi, men også på behovet for systemiske ændringer i selve tilgangen til ledelse, uddannelse og innovation.

> Den digitale transformations succes afhænger ikke så meget af antallet og tilgængeligheden af værktøjer, men af organisationernes evne til at gennemgå deres processer og udvikle en kultur, der er åben for forandring. Det er ikke teknologien i sig selv, der er nøglen, men de mennesker og processer, der sikrer en effektiv brug af den, understøtter løbende læring og tilskynder til accept af nye ideer.

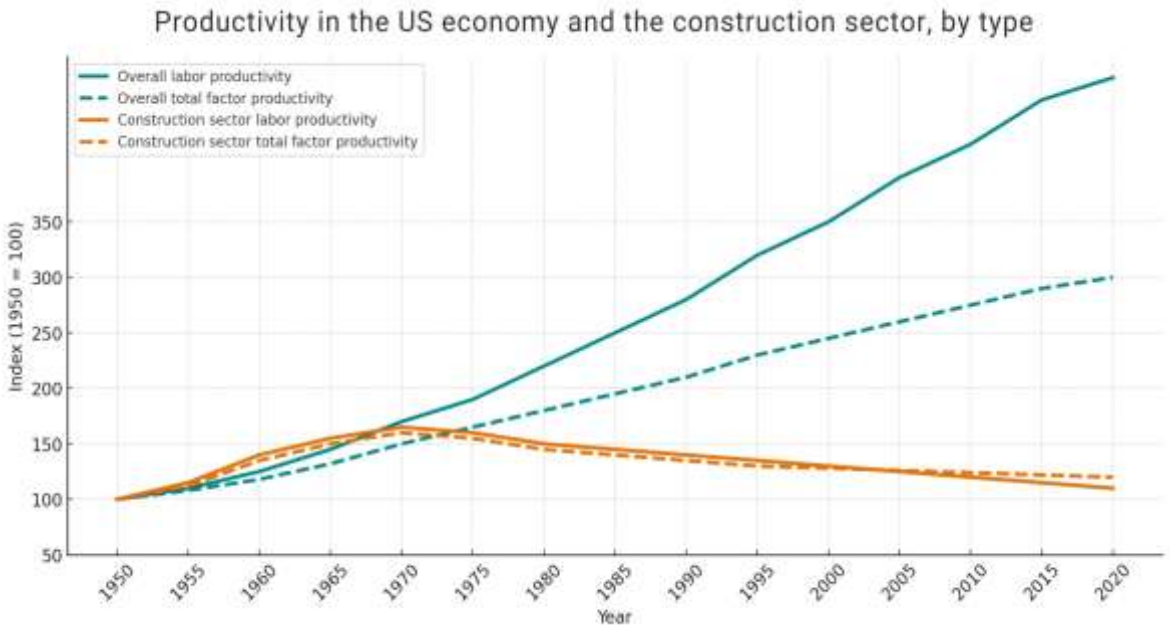


Fig. 10.2-1 Paradokset med arbejdsproduktivitet og total ressourceproduktivitet i den amerikanske økonomi og byggesektor (1950-2020) (baseret på [43]).

I de første dele af bogen blev forretningsmiljømodellen sammenlignet med et skovøkosystem (figur 2.1-2, figur 1.2-4, figur 1.3-2). I en sund skov spiller periodiske brande, på trods af deres destruktive kraft, en vigtig rolle i den langsigtede fornyelse. De rydder jorden for gammel vegetation, giver oplagrede næringsstoffer tilbage og skaber plads til nyt liv. Nogle plantearter har endda udviklet sig, så deres frø kun åbner sig, når de udsættes for høje brandtemperaturer - en naturlig mekanisme, der giver det ideelle tidspunkt for spiring.

På samme måde kan kriser i erhvervslivet fungere som en "kontrolleret udbrændthed", der fremmer fremkomsten af nye tilgange og virksomheder, som ikke er bundet til forældede systemer. Sådanne perioder tvinger til at opgive ineffektiv praksis og frigør ressourcer til innovation. Ligesom en skov efter en brand starter med pionerplanter, så danner en virksomhed efter en krise nye, fleksible processer, som bliver grundlaget for et modent informationsmiljø.

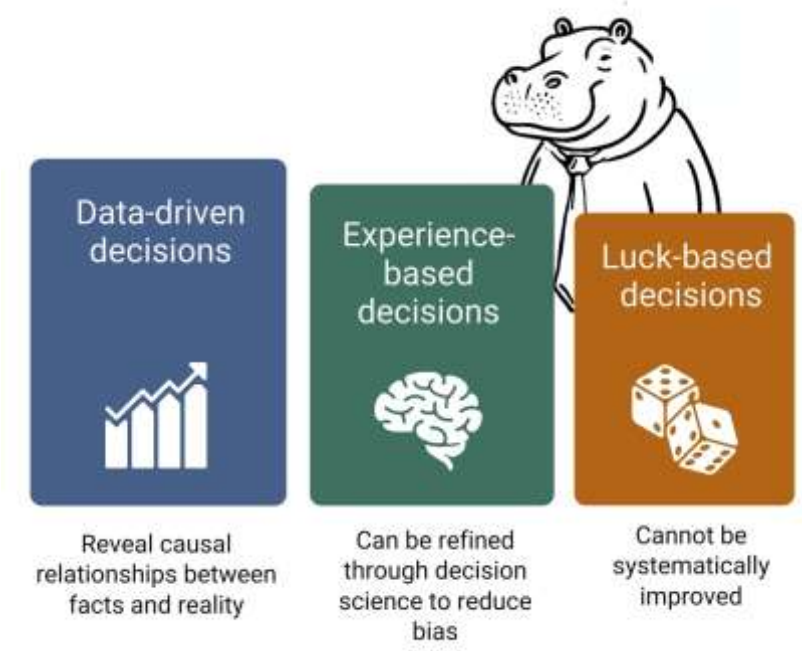
Virksomheder, der formår at fortolke disse "signalbrande" korrekt og omdanne deres forstyrrende energi til konstruktive forandringer, vil nå et nyt præstationsniveau - med mere gennemsigtige, adaptive dataprocesser, der forbedrer en organisations naturlige evne til at forny sig og vokse.

> Der er ikke længere tvivl om den voksende indflydelse, som kunstig intelligens og maskinlæring har på forretningsmiljøet. Det er ikke bare en midlertidig tendens, men en strategisk nødvendighed. Virksomheder, der ignorerer AI, risikerer at miste konkurrenceevne på et marked, der i stigende grad opfordrer til innovation og fleksibilitet.

Fremtiden tilhører dem, der ikke bare ser AI som et værktøj, men som en mulighed for at gentænke alle aspekter af deres virksomhed - fra optimering af processer til ledelsesbeslutninger.

At lægge det digitale fundament: 1-5 trin til digital modenhed

I dette kapitel ser vi på køreplanen for digital transformation og identificerer de vigtigste trin, der kræves for at implementere en datadrevet tilgang, der kan hjælpe med at transformere både virksomhedskulturen og virksomhedens informationsøkosystem.



Figur 10.2-2 Kontrolleret opdatering og valg af strategi: case, erfaring eller data.

Ifølge McKinseys undersøgelse "Why Digital Strategies Fail" (2018) er der mindst fem grunde [164] til, at virksomheder ikke når deres mål for digital transformation

- **Uklare definitioner**: Chefer og ledere har forskellige opfattelser af, hvad "digital" betyder, hvilket fører til misforståelser og uoverensstemmelser.
- **Misforståelse af den digitale økonomi**: Mange virksomheder undervurderer omfanget af de ændringer, som digitaliseringen medfører for forretningsmodeller og branchedynamik (figur 10.1-6).
- **Ignorerer økosystemer**: Virksomheder fokuserer på individuelle teknologiløsninger (datasiloer) og overser behovet for at integrere i bredere digitale økosystemer (figur 2.2-2, figur 4.1-12).
- **Undervurdering af konkurrenternes digitalisering**: Lederne tager ikke højde for, at konkurrenterne også aktivt indfører digitale teknologier, hvilket kan føre til tab af konkurrencefordele.
- **Mangler digitaliseringens dualitet**: CEO'er uddelegerer ansvaret for digital transformation til andre ledere, hvilket bureaukratiserer kontrollen og forsinker forandringsprocessen.

Det kræver en klar forståelse og tilpasning af de digitale strategier på alle niveauer i organisationen at håndtere disse udfordringer. Før man opbygger en digital strategi, er det vigtigt at forstå udgangspunktet. Mange organisationer har en tendens til at tage nye værktøjer og platforme i brug uden at have et fuldstændigt billede af den nuværende situation.

Trin 1: Gennemfør en revision af dine nuværende systemer og data.

Før man ændrer processer, er det vigtigt at forstå, hvad der allerede er på plads. Ved at gennemføre en audit kan du identificere svagheder i datahåndteringen og forstå, hvilke ressourcer der kan bruges. En audit er en slags "røntgenbillede" af dine forretningsprocesser. Den giver dig mulighed for at identificere risikoområder og afgøre, hvilke data der er kritiske for dit projekt eller din virksomhed, og hvilke der er sekundære.

De vigtigste handlinger:

- Kortlæg dit it-miljø (i Draw.io, Lucidchart, Miro, Visio eller Canva). Lav en liste over de anvendte systemer (ERP, CAD, CAFM, CPM, SCM og andre), der er involveret i dine processer, og som vi diskuterede i kapitlet "Teknologier og ledelsessystemer i moderne byggeri" (fig. 1.2-4)
- Vurder datakvalitetsproblemer for hvert system for hyppigheden af dubletter, mulige manglende værdier og formatuoverensstemmelser i hvert system.
- Identificer "smertepunkter" - steder, hvor processer kan bryde sammen eller ofte kræver manuel indgriben - import, eksport og yderligere valideringsprocesser.

> Hvis du vil have teamet til at stole på rapporterne, skal du sørge for, at dataene er korrekte fra starten.

En kvalitetsdatarevision vil vise dig, hvilke data:

- Behov for yderligere udvikling (automatiske rengøringsprocesser eller yderligere transformationer skal sættes op)
- De er "affald", som kun tilstopper systemer og kan fjernes ved ikke længere at bruge dem i processer.

Det er muligt at gennemføre en sådan revision på egen hånd. Men nogle gange er det nyttigt at engagere en ekstern konsulent - især fra andre brancher: Et nyt perspektiv og uafhængighed af byggeriets "særegenheder" vil hjælpe med at vurdere status quo nøgternt og undgå de typiske faldgruber med forudindtagethed over for bestemte løsninger og teknologier.

Trin 2: Identificer de vigtigste standarder for dataharmonisering.

Efter revisionen er det nødvendigt at skabe fælles regler for arbejdet med data. Som vi diskuterede i kapitlet "Standarder: Fra tilfældige filer til en intelligent datamodel", vil dette hjælpe med at eliminere siloopdelte datastrømme.

> Uden en enkelt standard vil hvert team fortsætte med at arbejde på "deres egen måde", og du vil opretholde en "zoologisk have" af integrationer, hvor data går tabt ved hver konvertering.

De vigtigste handlinger:

- Vælg datastandarder til at udveksle information mellem systemer:
  - For tabeldata kan dette være strukturerede formater som CSV, XLSX eller mere effektive

formater som Parquet

- Til udveksling af løst strukturerede data og dokumenter: JSON eller XML

- Mester i at arbejde med datamodeller:
  - Start med at parametrisere opgaverne på niveauet for den konceptuelle datamodel - som beskrevet i kapitlet "Datamodellering: konceptuel, logisk og fysisk model" (Fig. 4.3-2).
  - Når du dykker dybere ned i forretningsprocessens logik, skal du formalisere kravene ved hjælp af parametre i de logiske og fysiske modeller (figur 4.3-6).
  - Identificer centrale enheder, deres attributter og relationer inden for processer, og visualiser disse relationer - både mellem enheder og mellem parametre (figur 4.3-7).
- Brug regulære udtryk (RegEx) til at validere og standardisere data (figur 4.4-7), som vi diskuterede i kapitlet "Strukturerede krav og regulære udtryk RegEx". RegEx er ikke et komplekst, men et ekstremt vigtigt emne i arbejdet med at skabe krav på niveau med fysiske datamodeller.

Uden standarder på dataniveau og procesvisualisering er det umuligt at levere et ensartet og skalerbart digitalt miljø. Husk, at "dårlige data er dyre". Og omkostningerne ved fejl stiger i takt med, at et projekt eller en organisation bliver mere kompleks. At forene formater og definere regler for navngivning, struktur og validering er en investering i fremtidige løsningers stabilitet og skalerbarhed.

Trin 3. Implementer DataOps og automatiser processer.

Uden en veldefineret arkitektur vil virksomheder uundgåeligt stå over for forskellige data i siloopdelte informationssystemer. Data vil være uintegrerede, duplikerede flere steder og dyre at vedligeholde.

> Forestil dig, at data er vand, og at dataarkitekturen er det komplekse system af rørledninger, der transporterer vandet fra dets kilde til dets anvendelsessted. Det er dataarkitekturen, der bestemmer, hvordan information indsamles, lagres, omdannes, analyseres og leveres til slutbrugere eller applikationer.

DataOps (Data Operations) er en metode, der integrerer indsamling, rensning, validering og brug af data i et enkelt automatiseret procesflow, som vi diskuterede i detaljer i del 8 af bogen.

De vigtigste handlinger:

- Opret og tilpas ETL -conveyors for at automatisere processer:
  - Udtræk: organiser automatisk dataindsamling fra PDF dokumenter (Fig. 4.1-2, Fig. 4.1-5, Fig. 4.1-7), Excel regneark, CAD -modeller (Fig. 7.2-4), ERP -systemer og andre kilder, du arbejder med.
  - Transform: opsæt automatiske processer til at transformere data til et enkelt struktureret format og automatisere beregninger, der vil finde sted uden for lukkede applikationer (Fig. 7.2-8).
  - Indlæs: prøv at oprette en automatisk dataoverførsel til oversigtstabeller, dokumenter eller centrale arkiver (fig. 7.2-9, fig. 7.2-13, fig. 7.2-16).
- Automatiser beregnings- og QTO (Quantity Take-Off)-processerne, som vi diskuterede i kapitlet

"QTO Quantity Take-Off: Gruppering af projektdata efter attributter":

- [ ] Opsæt automatisk udtrækning af volumener fra CAD -modeller ved hjælp af API'er, plug-ins eller reverse engineering-værktøjer (Fig. 5.2-5).
- [ ] Opret regler for gruppering af elementer til forskellige klasser efter attributter i form af tabeller (figur 5.2-12).
- [ ] Prøv at automatisere hyppigt gentagne mængde- og omkostningsberegninger uden for modulære lukkede systemer (Fig. 5.2-15)

- Begynd at bruge Python og Pandas til at behandle data, som vi diskuterede i kapitlet "Python Pandas: et uundværligt værktøj til at arbejde med data":
  - Anvend DataFrame til at arbejde med XLSX-filer og automatisere behandlingen af tabeldata (Fig. 3.4-6).
  - Automatiser aggregering og transformation af information gennem forskellige Python-biblioteker
  - Brug LLM til at gøre det nemmere at skrive færdige kodeblokke og hele pipelines (figur 7.2-18).
  - Prøv at bygge en Pipeline i Python, som finder fejl eller ser afvigelser og sender en meddelelse til den ansvarlige person (f.eks. projektlederen) (figur 7.4-2).

Automatisering baseret på DataOps-principper giver dig mulighed for at gå fra manuel og fragmenteret datahåndtering til bæredygtige og gentagelige processer. Det reducerer ikke kun byrden for de medarbejdere, der arbejder med de samme transformationer hver dag, men øger også pålideligheden, skalerbarheden og gennemsigtigheden i hele informationssystemet dramatisk.

Trin 4: Skab et åbent økosystem for datastyring.

På trods af udviklingen af lukkede modulære systemer og deres integration med nye værktøjer står virksomheder over for et alvorligt problem - den voksende kompleksitet i sådanne systemer overgår deres anvendelighed. Den oprindelige idé om at skabe en enkelt proprietær platform, der dækker alle forretningsprocesser, har ført til overdreven centralisering, hvor enhver ændring kræver betydelige ressourcer og tid til at blive tilpasset.

Som vi diskuterede i kapitlet "Virksomhedens mycelium: Hvordan data forbinder forretningsprocesser", kræver effektiv datastyring et åbent og samlet økosystem, der forbinder alle informationskilder.

Nøgleelementer i økosystemet:

- Vælg et passende datalager:
  - [ ] Brug databaser til tabeller og beregninger - for eksempel PostgreSQL eller MySQL (Fig. 3.1-7).
  - [ ] Til dokumenter og rapporter kan cloud storage (Google Drive, OneDrive) eller systemer, der understøtter JSON-formatet, være velegnet.
  - [ ] Tjek mulighederne i Data Warehouse, Data Lakes og andre værktøjer til centraliseret lagring og analyse af store mængder information (figur 8.1-8).

- Implementer løsninger til at få adgang til proprietære data:
  - Hvis du bruger proprietære systemer, skal du konfigurere adgang til dem via API eller SDK for at få data til ekstern behandling (fig. 4.1-2).
  - Gør dig bekendt med potentialet i reverse engineering-værktøjer for CAD formater (figur 4.1-13)
  - Opsæt ETL-pipelines, der med jævne mellemrum indsamler data fra applikationer eller servere, konverterer dem til åbne strukturerede formater og gemmer dem i arkiver (figur 7.2-3).
  - Diskuter i teamet, hvordan man kan give adgang til data uden behov for proprietær software.
  - Husk: Data er vigtigere end grænseflader. Det er strukturen og tilgængeligheden af information, ikke specifikke brugergrænsefladeværktøjer, der giver langsigtet værdi.
- Overvej at oprette et kompetencecenter (CoE) for data, som vi diskuterede i kapitlet "Kompetencecenter (CoE) for datamodellering", eller hvordan du kan levere dataekspertise på andre måder (figur 4.3-9).

Datastyringsøkosystemet skaber et samlet informationsrum, hvor alle projektdeltagere arbejder med ensartede, opdaterede og verificerede oplysninger. Det er grundlaget for skalerbare, fleksible og pålidelige digitale processer

Frigør potentialet i data: 5-10 trin til digital modenhed

Ud over den tekniske integration er en vigtig faktor for en vellykket implementering af digitale løsninger, at slutbrugerne tager dem til sig. At engagere kunder eller brugere i præstationsmåling er både en udfordring i forhold til at forbedre brugeroplevelsen og håndtere forandringer i virksomheden. Hvis en løsning ikke passer ind i en velkendt arbejdsgang eller ikke løser reelle bruger- eller kundeproblemer, vil den ikke blive brugt, og det kan ingen yderligere foranstaltninger og incitamenter rette op på.

Transformation er en iterativ proces baseret på analyse af data om brugerinteraktion med nye processer, med hyppige testcyklusser, konstant feedback og forbedringer.

Trin 5: Opbyg en datakultur, træn personalet og indsaml feedback

Selv det mest avancerede system fungerer ikke uden medarbejdernes engagement. Du er nødt til at skabe et miljø, hvor data bruges dagligt, og hvor teamet forstår deres værdi.

> Den britiske regerings offentliggjorte 2024-rapport "Data Analytics and AI in Government Project Delivery" bemærker [83], at uddannelse af fagfolk med de nødvendige kompetencer inden for databehandling og -fortolkning er afgørende for en vellykket implementering af data analytics and AI.

Mangel på ekspertise i dataanalyse er et af de vigtigste problemer, der begrænser den digitale transformation. Ledere er vant til etablerede rutiner: kvartalsvise cyklusser, prioriterede initiativer og traditionelle måder at drive projekter fremad på. Forandring kræver en markant leder - høj nok i rang til at have indflydelse, men ikke så høj, at han eller hun har tid og motivation til at lede et langsigtet transformationsprojekt.

De vigtigste handlinger:

- Anerkendelse af behovet for at gå fra subjektive beslutninger baseret på en højtlønnet medarbejders mening (HiPPO) til en beslutningskultur baseret på fakta og data, som diskuteret i kapitlet "HiPPO eller faren ved meninger i beslutningstagningen" (figur 2.1-9).
- Organiser systematisk træning:
  - Hold træningssessioner om, hvordan man bruger strukturerede data, og inviter eksperter fra andre brancher, som ikke har en forudindtagethed over for produkter og koncepter, der er populære i byggebranchen i dag.
  - Diskutere dataanalysetilgange og -værktøjer med kolleger og selvstændigt lære praktisk arbejde med værktøjer som Python, pandas og LLM (Fig. 4.1-3, Fig. 4.1-6).
  - Opret et bibliotek med vejledninger (helst med korte videoer) om emnet strukturering af data (fig. 3.2-15) og oprettelse af datamodeller (fig. 4.3-6, fig. 4.3-7).
- Brug moderne læringsteknologier:
  - Brug sprogmodeller (LLM'er) til at understøtte kode- og datamanipulation, herunder kodegenerering, refaktorering og analyse samt behandling og fortolkning af tabeloplysninger (figur 3.4-1).
  - Udforsk, hvordan LLM-genereret kode kan tilpasses og integreres i en komplet Pipeline-løsning, når du arbejder i et offline udviklingsmiljø (IDE) (Fig. 4.4-14, Fig. 5.2-13).

> Når en leder fortsætter med at træffe beslutninger på den "gammeldags måde", kan ingen uddannelse overbevise folk om, at de skal tage analyser alvorligt.

Det er umuligt at opbygge en datakultur uden løbende feedback. Feedback giver dig mulighed for at identificere huller i processer, værktøjer og strategier, som ikke kan opdages gennem interne rapporter eller formelle KPImetrics. Komplimenterende kommentarer fra brugere af dine løsninger vil ikke give praktisk værdi. Det er kritisk feedback, der er værdifuld, især hvis den er baseret på konkrete observationer og fakta. Men det kræver en indsats at indhente sådanne oplysninger: Du skal opbygge processer, hvor deltagerne - både interne og eksterne - kan dele kommentarer (det kan give mening at gøre det anonymt) uden forvrængning og uden frygt for, at deres meninger kan påvirke deres eget arbejde. Det er vigtigt, at de gør det uden forvrængning og uden frygt for negative konsekvenser for dem selv.

> *Enhver læring er i sidste ende selvlærende [165].*
>
> - Milton Friedman, amerikansk økonom og statistiker

Implementering af analyseværktøjer bør ledsages af regelmæssig verificering af deres effektivitet i praksis (ROI, KPI'er), hvilket kun kan opnås gennem struktureret feedback fra medarbejdere, kunder og partnere. Det gør det muligt for virksomheder ikke kun at undgå at gentage fejl, men også at tilpasse sig hurtigere til ændringer i miljøet. At have en mekanisme til at indsamle og analysere feedback er et af tegnene på modenhed i en organisation, der bevæger sig fra lejlighedsvise digitale initiativer til en bæredygtig model for løbende forbedringer (figur 2.2-5).

Trin 6: Fra pilotprojekter til opskalering

> *Vælg kampe, der er store nok til at betyde noget og små nok til at vinde.*
>
> - Jonathan Kozol

Det er ekstremt risikabelt at lancere digital transformation "på én gang og overalt". En mere effektiv tilgang er at starte med pilotprojekter og gradvist opskalere succesfuld praksis.

De vigtigste handlinger:

- Vælg det rigtige projekt til pilotprojektet:
  - Definer et specifikt forretningsmæssigt mål eller en proces med målbare resultater (KPI, ROI) (Fig. 7.1-5).
  - Vælg en ETL-automatiseringsproces, f.eks. automatisk datavalidering eller beregning af arbejdsbyrde (QTO) ved hjælp af Python og Pandas (figur 5.2-10)
  - Fastlæg klare mål for succes (f.eks. reducer tiden til at producere inspektionsspecifikationer eller datavalideringsrapporter fra en uge til en dag).
- Brug iterative tilgange:
  - Start med enkle datakonverteringsprocesser, og opret streamingkonverteringer af data i flere formater til de formater, du har brug for til dine processer (fig. 4.1-2, fig. 4.1-5).
  - Øg gradvist opgavernes kompleksitet og udvid automatiseringen af processer ved at danne en komplet pipeline i IDE'et baseret på dokumenterede kodeblokke (fig. 4.1-7, fig. 7.2-18).
  - Dokumentér og optag (helst med korte videoer) vellykkede løsninger, og del dem med kolleger eller i faglige fællesskaber.
- Udvikle skabeloner og ledsagende dokumentation til at kopiere sådanne løsninger, så de kan bruges effektivt af dine kolleger (eller medlemmer af det professionelle fællesskab, herunder brugere af sociale medier).

En trinvis "roll-up" giver dig mulighed for at opretholde den høje kvalitet af ændringer og ikke falde i kaos med parallelle implementeringer. Strategien "fra lille til stor" minimerer risici og giver dig mulighed for at lære af små fejl uden at lade dem vokse til kritiske problemer.

At skifte fra en projekttilgang, hvor medarbejderne kun er delvist involveret, til dannelsen af permanente teams (f.eks. ekspertisecentre - CoE'er) hjælper med at sikre en bæredygtig udvikling af produktet, selv efter at den første version er udgivet. Sådanne teams understøtter ikke kun eksisterende løsninger, men fortsætter også med at forbedre dem.

Det reducerer afhængigheden af langvarige godkendelser: Teammedlemmerne får mulighed for at træffe

beslutninger inden for deres ansvarsområde. Resultatet er, at lederne slipper for behovet for mikromanagement, og teamene kan fokusere på at skabe reel værdi.

> At udvikle nye løsninger er ikke en sprint, men et maraton. De, der får succes med det, er dem, der i første omgang satser på et langsigtet og konsekvent arbejde.

Det er vigtigt at indse, at teknologi kræver konstant udvikling. Investering i den langsigtede udvikling af teknologiske løsninger er grundlaget for en vellykket drift.

Trin 7: Brug åbne dataformater og -løsninger

Som vi diskuterede i kapitlerne om modulære platforme (ERP, PMIS, CAFM, CDE osv.), er det vigtigt at fokusere på åbne og universelle dataformater, der sikrer uafhængighed af leverandørløsninger og øger tilgængeligheden af information for alle interessenter.

De vigtigste handlinger:

- Gå fra lukkede formater til åbne formater:
  - Brug åbne formater i stedet for proprietære formater, eller find en måde at sætte automatisk upload eller konvertering af lukkede formater til åbne formater op på (figur 3.2-15).
  - Implementere værktøjer til at arbejde med Parquet, CSV, JSON, XLSX, som er udvekslingsstandarder mellem de fleste moderne systemer (Fig. 8.1-2).
  - Hvis arbejdet med 3D-geometri spiller en vigtig rolle i dine processer, bør du overveje at bruge åbne formater som USD, glTF, DAE eller OBJ (figur 3.1-14).
- Brug vektor-databaser databaser til effektivt at analysere og søge efter information:
  - Brug Bounding Box og andre metoder til at forenkle 3D -geometrien (figur 8.2-1).
  - Tænk over, hvor du kan implementere datavektorisering - konvertere tekster, objekter eller dokumenter til numeriske repræsentationer (figur 8.2-2).
- Anvend værktøjer til big data-analyse:
  - Organiser lagring af akkumulerede historiske data (f.eks. PDF, XLSX, CAD) i passende formater til analyse (Apache Parquet, CSV, ORC) (Fig. 8.1-2).
  - Begynde at anvende grundlæggende statistiske metoder og arbejde med repræsentative stikprøver - eller som minimum gøre dig bekendt med de grundlæggende principper for statistik (figur 9.2-5).
  - Implementer og lær værktøjer til at visualisere data og relationerne mellem data for at visualisere resultaterne af analysen. Uden god visualisering er det umuligt fuldt ud at forstå selve dataene eller de processer, der er baseret på dem (figur 7.1-4).

Overgangen til åbne dataformater og indførelsen af værktøjer til analyse, lagring og visualisering af information lægger grunden til en bæredygtig og uafhængig digital forvaltning. Det reducerer ikke kun afhængigheden af leverandører, men sikrer også lige adgang til data for alle interessenter.

Trin 8. Begynd at implementere maskinlæring til forudsigelse

Mange virksomheder har akkumuleret store mængder data - en slags "informationsgejsere", som stadig er ubrugte. Disse data er blevet indsamlet i hundredvis eller tusindvis af projekter, men er ofte kun blevet brugt én gang eller slet ikke i de videre processer. Dokumenter og modeller, der er gemt i lukkede formater og systemer, opfattes ofte som forældet og ubrugelig ballast. Men i virkeligheden er de en værdifuld ressource - grundlaget for at analysere begåede fejl, automatisere rutineoperationer og udvikle innovative løsninger til automatisk klassifikation og funktionsgenkendelse i fremtidige projekter.

Hovedudfordringen er at lære at udtrække disse data og omdanne dem til brugbar indsigt. Som beskrevet i kapitlet om maskinlæring og forudsigelser har maskinlæringsteknikker potentialet til at forbedre nøjagtigheden af estimater og forudsigelser i en række forskellige byggerelaterede processer. Fuld udnyttelse af de akkumulerede data baner vejen for at forbedre effektiviteten, reducere risikoen og opbygge bæredygtige digitale processer.

De vigtigste handlinger:

- Start med enkle algoritmer:
  - Prøv at anvende lineær regression - ved hjælp af tips fra LLM - til at forudsige tilbagevendende præstationer i datasæt, hvor afhængigheden af et stort antal faktorer er fraværende eller minimal (figur 9.3-4).
  - Overvej, i hvilke faser af dine processer k-NN-algoritmen (k-nearest neighbours) teoretisk kunne anvendes - f.eks. til klassificeringsopgaver, vurdering af objektlighed eller forudsigelse baseret på historiske analogier (Fig. 9.3-5).
- Indsaml og strukturér data for at træne modeller:
  - Saml historiske projektdata på ét sted og i ét format (figur 9.1-10)
  - Arbejde med kvaliteten og repræsentativiteten af træningsprøver gennem automatiseret ETLFigur 9.2-8)
  - Lær at opdele data i trænings- og testsæt, som vi gjorde i eksemplet med Titanic-datasættet (figur 9.2-6, figur 9.2-7).
- Overvej at udvide anvendelsen af maskinlæringsteknikker til en bred vifte af opgaver, fra forudsigelse af projekttidslinjer til optimering af logistik, ressourcestyring og tidlig identifikation af potentielle problemer.

Maskinlæring er et værktøj til at gøre arkiverede data til et værdifuldt aktiv til forudsigelse, optimering og informeret beslutningstagning. Start med små datasæt (figur 9.2-5) og enkle modeller, og øg gradvist kompleksiteten.

Trin 9. Integrer IoT og moderne dataindsamlingsteknologier

Byggeverdenen bliver hurtigt digital: Hvert eneste byggefoto, hvert eneste Teams-indlæg er allerede en del af en større proces med at parametrisere og tokenisere virkeligheden. Ligesom GPS i sin tid forandrede logistikken, er IoT, RFID og teknologier til automatisk dataindsamling i færd med at forandre byggebranchen. Som beskrevet i kapitlet "IoT Internet of Things and Smart Contracts" er den digitale byggeplads

med sensorer og automatiseret overvågning branchens fremtid.

De vigtigste handlinger:

- Implementer IoT -enheder, RFID -tags og beskriv de processer, der er forbundet med dem:
  - Evaluer, hvilke områder eller projektfaser i et projekt der kan få det største investeringsafkast (ROI) - f.eks. til overvågning af temperatur, vibrationer, fugtighed eller.
  - Overvej at bruge RFID til at spore materialer, værktøj og udstyr i hele forsyningskæden.
  - Overvej, hvordan de indsamlede data kan integreres i et enkelt informationssystem, f.eks. Apache NiFi, til automatiseret behandling og analyse i realtid (figur 7.4-5)
- Etabler et overvågningssystem i realtid:
  - Udvikle dashboards til at spore vigtige proces- eller projektmålinger ved hjælp af visualiseringsværktøjer som Streamlit, Flask eller Power BI)
  - Opsæt automatiske meddelelser for at signalere kritiske afvigelser fra planen eller normerne (Fig. 7.4-2)
  - Vurder potentialet for forebyggende vedligeholdelse af udstyr baseret på de indsamlede data og identificerede mønstre (figur 9.3-6).
- Kombiner data fra forskellige kilder:
  - Start med at visualisere datamodellen på det fysiske niveau - afspejl strukturen af informationsstrømme og nøgleparametre, der kommer fra CAD-systemer, IoT-enheder og ERP-platforme (fig. 4.3-1).
  - Start med at lave en grov skitse af en samlet platform til dataanalyse og beslutningsstøtte til ledelsen. Fastlæg nøglefunktioner, datakilder, brugere og forventede anvendelsesscenarier (figur 4.3-7).

Jo hurtigere du begynder at forbinde virkelige processer med den digitale verden, jo hurtigere kan du styre dem med data - effektivt, gennemsigtigt og i realtid.

Trin 10. Forbered dig på fremtidens ændringer i branchen

Byggevirksomheder er konstant under pres fra det ydre miljø: økonomiske kriser, teknologiske spring, lovændringer. Ligesom en skov, der skal modstå regn, sne, tørke og brændende sol, lever virksomheder under forhold med kontinuerlig tilpasning. Og ligesom træer bliver modstandsdygtige over for frost og tørke gennem dybe rodsystemer, er det kun de organisationer, der har et solidt fundament af automatiserede processer, evnen til at forudse forandringer og fleksibelt tilpasse strategier, der forbliver levedygtige og konkurrencedygtige.

Som nævnt i kapitlet "Overlevelsesstrategier: Opbygning af konkurrencefordele" er byggebranchen på vej ind i en fase med radikale forandringer. Samspillet mellem bygherre og entreprenør bevæger sig i retning af en overtalelsesmodel, hvor gennemsigtighed, forudsigelighed og digitale værktøjer erstatter traditionelle tilgange. I denne nye virkelighed er det ikke de største, men de mest fleksible og teknologisk modne, der vinder.

De vigtigste handlinger:

- Analyser virksomhedens sårbarheder i forbindelse med åbne data:
  - Vurder, hvordan demokratisering af adgang til data som en del af Uberisering kan have en ødelæggende indvirkning på din konkurrencefordel og din virksomhed (figur 10.1-5).
  - Overvej en strategi for at gå fra uigennemsigtige og siloopdelte processer til forretningsmodeller baseret på åbne løsninger, systeminteroperabilitet og datatransparens (Fig. 2.2-5).
- Udvikl en langsigtet digital strategi:
  - Find ud af, om du ønsker at være innovationsleder eller foretrækker et "indhentningsscenarie", hvor du sparer på dine ressourcer.
  - Beskriv stadierne: på kort sigt (procesautomatisering, datastrukturering), på mellemlang sigt (implementering af LLM og ETL), på lang sigt (digitale økosystemer, centraliserede repositorier).
- Overvej at udvide din portefølje af tjenester:
  - Overvej at tilbyde nye tjenester (med fokus på energieffektivitet, ESG, datatjenester). Vi vil tale mere om nye forretningsmodeller i næste kapitel.
  - Forsøg at positionere dig som en pålidelig teknologipartner, der understøtter hele livscyklussen for et anlæg, fra design til drift. Tilliden til dig skal være baseret på en systematisk tilgang, gennemsigtige processer og evnen til at levere bæredygtige teknologiske løsninger.

I et transformationsmiljø er det ikke dem, der blot reagerer på forandringer, der vinder, men dem, der er proaktive. Fleksibilitet, åbenhed og digital modenhed er grundlaget for bæredygtighed i opbygningen af morgendagen.

Køreplan for transformation: fra kaos til datadrevet virksomhed

Følgende plan kan tjene som et første benchmark - et udgangspunkt for at forme din egen datadrevne digitale transformationsstrategi:

- **Revision og standarder**: analyser den nuværende tilstand, saml data
- **Datastrukturering og -klassificering**: automatiser transformationen af ustrukturerede og løst strukturerede data
- **Automatiser grupperinger**, beregninger og udregninger: brug open source-værktøjer og biblioteker til automatisering
- **Økosystem og COE**: opbyg et internt team til at danne et samlet dataøkosystem i virksomheden
- **Kultur og læring**: gå væk fra HiPPO -løsninger til datadrevne løsninger
- **Pilotprojekter, feedback og opskalering**: Handl iterativt: Test nye metoder i begrænset omfang, indsaml valid feedback, og opskalér løsningerne gradvist.
- **Åbne formater**: Brug universelle og åbne formater for at være uafhængig af softwareleverandører
- **Maskinlæring**: integrer ML-algoritmer i processer til forudsigelse og optimering

- **IoT og den digitale byggeplads**: integrer moderne dataindsamlingsteknologier i processer
- **Strategisk tilpasning**: forbered dig på fremtidige ændringer i branchen

> Vigtigst af alt skal du huske, at "data alene ændrer ikke en virksomhed: Det er de mennesker, der ved, hvordan de skal arbejde med disse data, der ændrer den". Læg vægt på kultur, gennemsigtige processer og en forpligtelse til løbende forbedringer

En systemtilgang muliggør et skift fra siloopdelte digitale initiativer til en fuldgyldig datadrevet ledelsesmodel, hvor beslutninger ikke er baseret på intuition eller antagelser, men på data, fakta og matematisk beregnede sandsynligheder. Den digitale transformation af byggebranchen handler ikke kun om at indføre teknologi, men om at forme et forretningsøkosystem, hvor projektoplysninger overføres problemfrit og iterativt mellem forskellige systemer. På den måde sørger maskinlæringsalgoritmer for automatisk, kontinuerlig analyse, prognoser og procesoptimering. I et sådant miljø bliver spekulation og skjulte data irrelevante - der er kun gennemprøvede modeller, gennemsigtige beregninger og forudsigelige resultater tilbage.

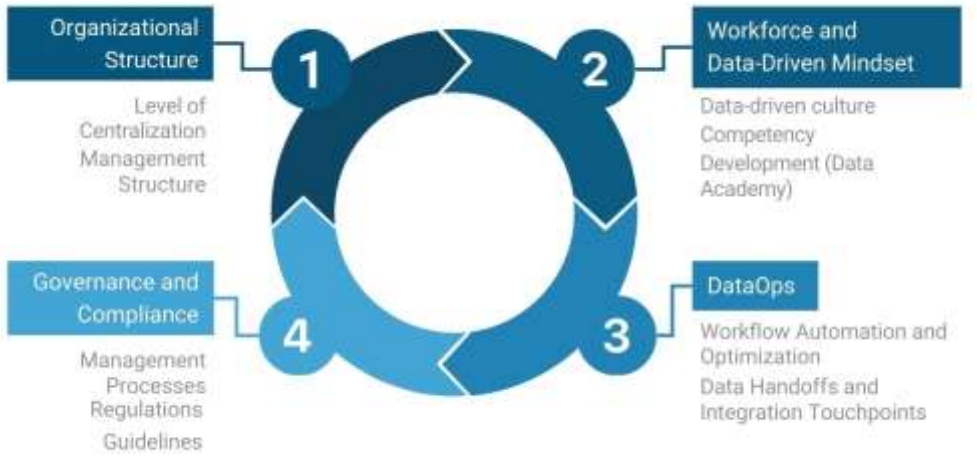


Fig. 10.2-3 Nøgleelementer i en vellykket datastyring på virksomhedsniveau.

Hver del af bogen svarer til en bestemt fase af databehandling og analyse i byggeprojekter (fig. 2.2-5). Hvis du vil vende tilbage til et af de emner, der er blevet diskuteret tidligere, og se på det ud fra en holistisk forståelse af dataudnyttelsesflowet, kan du henvise til deltitlerne i fig. 10.2-4.

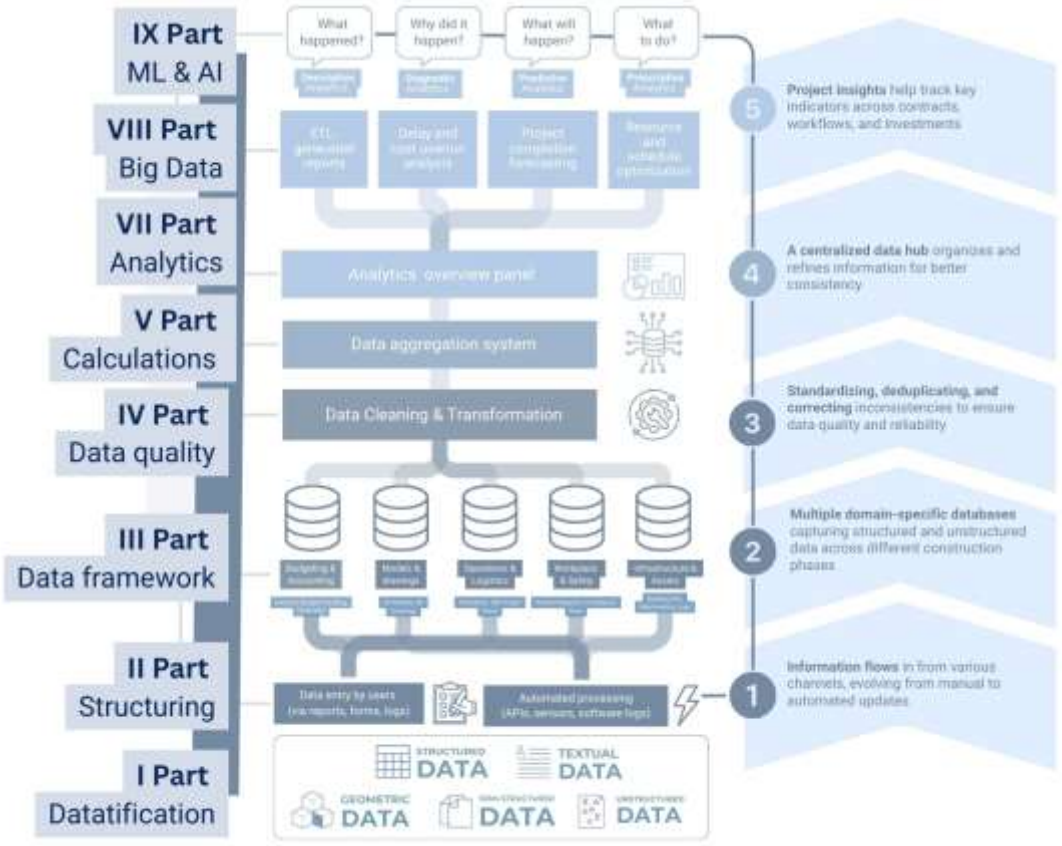


Fig. 10.2-4 Dele af bogen i sammenhæng med databehandlingspipelinen (Fig. 2.2-5): fra digitalisering af information til analyse og kunstig intelligens.

Uanset din organisations størrelse, teknologiske modenhed eller budget kan du begynde at bevæge dig i retning af en datadrevet tilgang i dag. Selv små skridt i den rigtige retning vil give resultater over tid.

> Datadrevet transformation er ikke et engangsprojekt, men en kontinuerlig, iterativ forbedringsproces, der omfatter introduktion af nye værktøjer, redesign af processer og udvikling af en kultur med datadrevet beslutningstagning.

Byggeri i Industri 5.0: Sådan tjener du penge, når du ikke længere kan gemme dig

I lang tid har byggefirmaer tjent penge på uigennemsigtige processer. Den vigtigste forretningsmodel var spekulation - overvurdering af materialeomkostninger, arbejdsomfang og procentvise tillæg i lukkede ERP - og PMIS - systemer, der er utilgængelige for eksterne revisioner. Kundernes og deres betroede personers begrænsede adgang til de oprindelige projektdata skabte grobund for ordninger, hvor det blev næsten umuligt at verificere beregningernes pålidelighed.

Men denne model er hurtigt ved at miste sin relevans. Med demokratiseringen af dataadgang, fremkomsten af LLM'er, ankomsten af åbne data og ETL-automatiseringsværktøjer bevæger branchen sig mod en ny arbejdsstandard.

> Resultatet er, at uigennemsigtighed ikke længere er en konkurrencefordel - snart bliver det en byrde, som det bliver svært at skille sig af med. Gennemsigtighed går fra at være en mulighed til at være en forudsætning for at blive på markedet.

Hvem vil kunderne - banker, investorer, fysiske kunder, private equity, offentlige kunder - arbejde sammen med i den nye digitale virkelighed? Svaret er indlysende: med dem, der er i stand til at levere ikke bare resultatet, men også begrundelsen for hvert skridt på vejen til det. Efterhånden som mængden af åbne data vokser, vil partnere og kunder vælge virksomheder, der garanterer gennemsigtighed, nøjagtighed og forudsigelighed i resultaterne.

På den baggrund opstår der nye forretningsmodeller, som er baseret på datahåndtering og tillid i stedet for spekulation:

- **Salg af processer i stedet for kvadratmeter: Det** vigtigste aktiv bliver tillid og effektivitet snarere end konkrete aftaler. Hovedværdien vil være forudsigelighed af resultatet baseret på pålidelige og verificerede data. Moderne virksomheder vil ikke sælge byggeobjektet som sådan, men:
  - præcise deadlines og gennemsigtige arbejdsplaner;
  - rimelige skøn, understøttet af beregninger;
  - fuld digital sporbarhed og kontrol i alle faser af projektet.
- **Engineering og analyse som en tjeneste**: "Data-as-a-Service "-modellen (en måde at levere færdige data til brugere via internettet, som en tjeneste), hvor hvert projekt bliver en del af en digital datakæde, og forretningsværdien ligger i evnen til at styre denne kæde. Virksomheder omdannes til intelligente platforme, der tilbyder løsninger baseret på automatisering og analyse:
  - Automatiseret og gennemsigtig udarbejdelse af estimater og planer;
  - risiko- og tidsvurdering baseret på maskinlæringsalgoritmer;
  - beregning af miljøindikatorer (ESG, CO_2, energieffektivitet);
  - generering af rapporter fra reviderede åbne kilder.
- **Produktivisering af teknisk ekspertise**: Virksomhedens udvikling kan bruges gentagne gange i virksomheden og distribueres som et separat produkt - og dermed udgøre en ekstra indtægtskilde gennem digitale tjenester. I det nye miljø skaber virksomheder ikke kun projekter, men også digitale

aktiver:

- biblioteker med komponenter og estimatskabeloner;
- automatiserede verifikationsmoduler;
- Open source-plugins og scripts (salg af rådgivning) til at arbejde med data.

- **En ny type virksomhed: den datadrevne integrator**: en markedsaktør, der ikke er afhængig af specifikke softwareleverandører eller modulære systemer og ikke er "låst" til en enkelt softwaregrænseflade. Den arbejder frit med data - og bygger sin konkurrenceevne på dem. Fremtidens byggefirma er ikke bare en entreprenør, men en informationsintegrator, der er i stand til at udføre følgende funktioner for kunden:
  - Kombiner data fra forskellige kilder og udfør analyser;
  - Sikre gennemsigtighed og troværdighed i processerne;
  - rådgive om optimering af forretningsprocesser;
  - udvikle værktøjer, der fungerer i økosystemet for åbne data, LLM, ETL og Pipelines.

Industri 5.0 (fig. 2.1-12) markerer afslutningen på "æraen med manuelle gennemsnitsberegninger" og aftenmøder mellem administrerende direktører og kalkulations- og regnskabsafdelingen. Alt, hvad der tidligere var skjult - beregninger, estimater, mængder - bliver åbent, verificerbart og forståeligt, selv for ikke-eksperter. De, der er de første til at omstille sig, bliver vinderne. Alle andre vil blive udelukket fra den nye digitale økonomi i byggesektoren.

KONKLUSION

Byggebranchen er på vej ind i en æra med grundlæggende forandringer. Fra de første optegnelser på lertavler til de enorme mængder af digitale data, der flyder fra projektservere og byggepladser, har historien om informationsstyring i byggeriet altid afspejlet den daværende teknologis modenhed. I dag, med fremkomsten af automatisering, åbne formater og intelligente analysesystemer, står branchen ikke over for en gradvis udvikling, men en hurtig digital transformation.

Som i andre økonomiske sektorer bliver byggeriet nødt til at gentænke ikke bare værktøjerne, men også arbejdsprincipperne. Virksomheder, der tidligere dikterede markedet og fungerede som det vigtigste mellemled mellem kunden og projektet, er ved at miste deres unikke position. Tillid og evnen til at arbejde med data - fra indsamling og strukturering af dem til analyse, prognoser og automatisering af beslutninger - kommer i forgrunden.

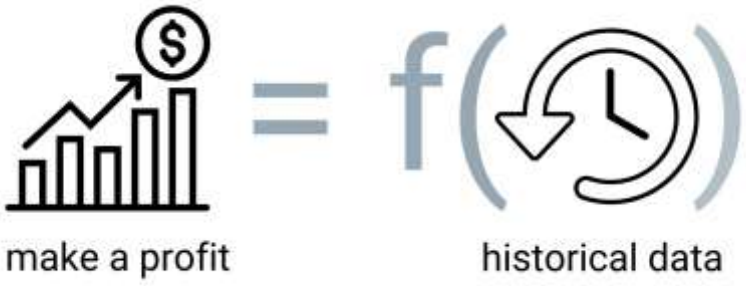


Fig. 10.2-1 Strukturerede historiske data er brændstoffet til en effektiv og håndterbar virksomhed.

Denne bog beskriver de vigtigste principper for datahåndtering i byggebranchen - fra revision og standardisering til procesautomatisering, brug af visualiseringsværktøjer og implementering af intelligente algoritmer. Vi ser på, hvordan man selv med begrænsede ressourcer kan opbygge en fungerende dataarkitektur og begynde at træffe beslutninger baseret på verificerbare fakta i stedet for intuition. At arbejde med data er ikke længere kun en opgave for IT-afdelingen - det bliver grundlaget for ledelseskulturen, som bestemmer virksomhedens fleksibilitet, tilpasningsevne og langsigtede bæredygtighed.

Anvendelsen af maskinlæringsteknologier, automatiske behandlingssystemer, digitale tvillinger og åbne formater gør det allerede i dag muligt at eliminere den menneskelige faktor, hvor den før var afgørende. Byggeriet bevæger sig i retning af autonomi og kontrollerbarhed, hvor bevægelsen fra idé til projektrealisering kan sammenlignes med navigation i autopilottilstand: uden afhængighed af subjektive beslutninger, uden behov for manuel indgriben i alle faser, men med fuld digital sporbarhed og kontrol (fig. 10.2-2).

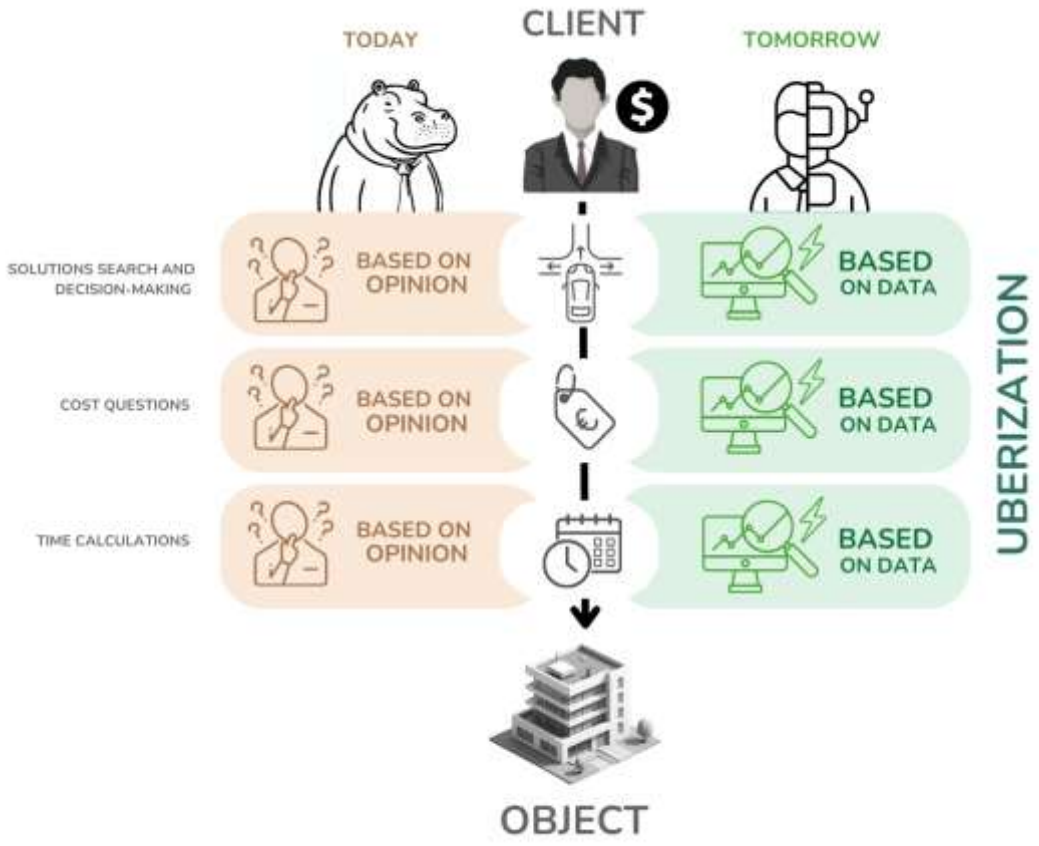


Fig. 10.2-2 Skiftet fra beslutningstagning baseret på udtalelser fra vigtige eksperter (HiPPO) til dataanalyse vil primært blive fremmet af kunden.

Ved at lære de metoder, principper og værktøjer, der præsenteres i denne bog, vil du være i stand til at begynde at træffe datadrevne beslutninger i din virksomhed i stedet for intuitive. Du bliver også i stand til at køre modulkæder i LLM'er, kopiere færdige ETL-pipelines ind i dit udviklingsmiljø (IDE) og automatisk behandle data for at få de oplysninger, du har brug for, i den form, du ønsker. Ved at bygge videre på bogens kapitler om big data og maskinlæring bliver du i stand til at implementere mere komplekse scenarier - udtrække ny indsigt fra historiske data og anvende maskinlæringsalgoritmer til at forudsige og optimere dine processer.

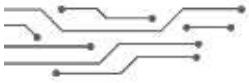
Åbne data og processer vil danne grundlag for mere præcise estimater af projektomkostninger og tidsfrister og forhindre byggevirksomheder i at spekulere på uigennemsigtige data. Dette er både en udfordring og en mulighed for branchen til at gentænke sin rolle og tilpasse sig et nyt miljø, hvor gennemsigtighed og effektivitet bliver vigtige succesfaktorer.

> Viljen til at tage viden til sig og omsætte den i praksis er nøglen til succes i den digitale transformations tidsalder.

Virksomheder, der først indser dette, vil have en fordel i den nye digitale konkurrence. Men det er vigtigt at indse, at data alene ikke vil ændre noget. Mange mennesker bliver nødt til at ændre deres måde at tænke på, og det kræver et incitament. Din virksomhed skal gentænke den måde, den deler data.

De mennesker, der forandrer virksomheden, er dem, der ved, hvordan man arbejder med disse data, fortolker dem, bruger dem til at optimere dem og skaber en ny procesarkitektur baseret på dem.

Hvis du læser disse linjer, er du klar til forandring, og du er allerede et skridt foran. Tak, fordi du har valgt denne vej. Velkommen til den digitale transformations æra!



OM FORFATTEREN



Mit navn er Artyom Boiko. Min rejse på byggepladsen startede i 2007 - med et job som minearbejder i en olieskifermine i min hjemby, mens jeg studerede på Skt. Petersborgs mineuniversitet med speciale i mine- og underjordisk byggeri. På bagsiden af denne bog kan du se en eksplosionist i den mine, hvor vi udvandt og sprængte hundredvis af kuber af olieskifer. Min karriere har taget mig i mange forskellige retninger, fra at arbejde som mine- og underjordisk bygningsarbejder til industriel klatrer, tagdækker og installatør af liftudstyr. Jeg har haft æren af at være involveret i projekter i forskellige skalaer, fra opførelse af private huse til store industrianlæg i forskellige regioner i verden.

Med tiden har mit arbejde flyttet sig fra fysisk byggeri til informationsstyring og digitale processer. Siden 2013 har jeg arbejdet i forskellige stillinger i små, mellemstore og store byggefirmaer i flere regioner i Tyskland, fra designer til data management manager. Hvad angår datastyring, består min erfaring i at arbejde med data i forskellige ERP-systemer, CAD (BIM), MEP, FEM, CMS. Jeg har været involveret i optimering, procesautomatisering samt analyse, maskinlæring og databehandling i planlægnings-, beregnings- og udførelsesfasen af byggearbejder i industri-, bolig-, infrastruktur- og forsyningsvirksomheder.

Jeg har arbejdet med open source-software og åbne data siden 2003. I løbet af denne tid har jeg implementeret mange webprojekter - fra hjemmesider og onlinebutikker til fuldgyldige webapplikationer ved hjælp af open source-løsninger og åbent CMS. Disse platforme, der på mange måder ligner moderne ERP-systemer, har modulær arkitektur, høj tilpasningsevne og tilgængelighed. Denne erfaring lagde grunden til min professionelle tilgang - et fokus på open source-teknologier og en kultur med fælles udvikling. Respekt for open source og fri udveksling af viden er noget, jeg bestræber mig på at fremme i byggebranchen. Mit arbejde med at forbedre datatilgængeligheden i byggebranchen har udmøntet sig i oprettelsen af flere fællesskaber på de sociale medier, hvor man diskuterer dataåbenhed og brugen af open source i byggeriet, samt i lanceringen af flere startups, der udvikler løsninger til at give adgang til data fra forskellige lukkede systemer og platforme.

Mit bidrag til det professionelle samfund er gennem deltagelse som taler på konferencer, der dækker CAD-interoperabilitet (BIM), ERP, 4D-5D, LLM Machine Learning og Artificial Intelligence, samt artikler udgivet i europæiske publikationer for byggebranchen. Et af mine bemærkelsesværdige resultater er oprettelsen af "BIM History" [111], et omfattende kort over vigtige softwareløsninger til datastyring i byggebranchen. Min artikelserie i 7 dele "BIM Development and Lobbying Games", der er oversat til flere sprog, er blevet bredt anerkendt som et forsøg på at belyse den skjulte dynamik i digitale standarder.

Det var sådan, jeg gik fra at udvinde sten - til at udvinde og systematisere byggedata. Jeg er altid åben for faglig dialog, nye ideer og fælles projekter. Jeg tager gerne imod feedback og ser frem til dine beskeder eller til at se dig blandt mine følgere på de sociale medier. Tusind tak, fordi du læste denne bog til ende! Det vil glæde mig, hvis denne bog hjælper dig til bedre at forstå emnet data i byggebranchen.

YDERLIGERE RELATION

Læsernes meninger spiller en vigtig rolle i den videre udvikling af publikationer og udvælgelsen af prioriterede emner. Kommentarer om, hvilke ideer der har vist sig nyttige, og hvilke der har givet anledning til tvivl og kræver yderligere afklaring eller kildehenvisning, er særligt værdifulde. Bogen indeholder en bred vifte af materiale og analyser, hvoraf nogle kan virke kontroversielle eller subjektive. Hvis du under læsningen finder unøjagtigheder, forkert citerede kilder, logiske uoverensstemmelser eller slåfejl, vil jeg være taknemmelig for dine kommentarer, tanker eller kritik, som du kan sende til:boikoartem@gmail.com. Eller via beskeder på LinkedIn: linkedin.com/in/boikoartem

Jeg vil sætte pris på enhver omtale af bogen Data-Driven Construction på de sociale medier - deling af læseoplevelser hjælper med at sprede budskabet om åbne data og værktøjer og støtter mit arbejde.

KOMMENTAR TIL OVERSÆTTELSE

Denne bog er oversat ved hjælp af kunstig intelligens-teknologi. Det har gjort oversættelsesprocessen meget hurtigere. Men som ved enhver teknologisk operation kan der forekomme fejl eller unøjagtigheder. Hvis du opdager noget, der ser ud til at være ukorrekt eller forkert oversat, så send mig en e-mail. Dine kommentarer vil hjælpe med at forbedre kvaliteten af oversættelsen.

DATADREVNE BYGGEFÆLLESSKABER

Det er et sted, hvor du frit kan stille spørgsmål og dele dine problemer og løsninger:

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

ANDRE FÆRDIGHEDER OG KONCEPTER

Ud over de vigtigste principper for at arbejde med data i byggebranchen behandler DataDrivenConstruction en lang række yderligere koncepter, programmer og færdigheder, som er afgørende for den datadrevne professionelle. Nogle af disse præsenteres kun i en oversigt, men er afgørende for praksis.

Den interesserede læser kan besøge webstedet DataDrivenConstruction.io for at finde links til yderligere materialer om nøglefærdigheder. Disse materialer omfatter arbejde med Python og Pandas, opbygning af ETL -processer, eksempler på databehandling i CAD-projekter inden for byggeri, big data-systemer og moderne tilgange til visualisering og analyse af byggedata.

Mange open source-værktøjer og software. blev brugt i forbindelse med udarbejdelsen af bogen "DataDrivenConstruction" og alle casestudierne. Forfatteren vil gerne takke udviklerne og medforfatterne af følgende løsninger:

- Python og Pandas - rygraden i datamanipulation og automatisering
- Scipy, NumPy, Matplotlib og Scikit-Learn - biblioteker til dataanalyse og maskinlæring
- SQL og Apache Parquet - værktøjer til lagring og behandling af store mængder byggedata
- Open Source CAD (BIM) åbne dataværktøjer i åbne formater
- N8n, Apache Airflow, Apache NiFi - systemer til orkestrering og automatisering af arbejdsgange
- DeepSeek, LlaMa, Mistral - Open Source LLM

En særlig tak til alle deltagere i diskussioner om emnet åbne data og værktøjer i faglige fællesskaber og sociale netværk, hvis kritik, kommentarer og idéer har været med til at forbedre bogens indhold og struktur.

Følg udviklingen af projektet på webstedet DataDrivenConstruction.io, hvor der ikke kun offentliggøres bogopdateringer og rettelser, men også nye kapitler, vejledninger og praktiske eksempler på anvendelse af de beskrevne teknikker.

MAKSIMAL BEKVEMMELIGHED MED DEN TRYKTE VERSION

Du sidder med en gratis digital version af **Datadrevet byggeri**. For mere praktisk arbejde og hurtig adgang til materialerne anbefaler vi, at du er opmærksom på den **trykte udgave**:



- **Altid ved hånden**: Bogen i trykt format bliver et pålideligt arbejdsredskab, så du hurtigt kan finde og bruge de rigtige visualiseringer og diagrammer i enhver arbejdssituation.
- **Illustrationer i høj kvalitet**: alle billeder og grafikker i den trykte udgave præsenteres i maksimal kvalitet
- **Hurtig adgang til information**: nem navigation, mulighed for at lave noter, bogmærker og arbejde med bogen overalt.

Ved at købe den fulde trykte version af bogen får du et praktisk værktøj til komfortabelt og effektivt arbejde med information: muligheden for hurtigt at bruge visuelle materialer i hverdagens opgaver, hurtigt finde de nødvendige skemaer og tage noter. Desuden støtter dit køb udbredelsen af åben viden.

Bestil en trykt version af bogen på: datadrivenconstruction.io/books



UNIK MULIGHED FOR STRATEGISK POSITIONERING

Vi tilbyder dig at placere reklamemateriale i den gratis version af DataDrivenConstruction. Den betalte version af publikationen har tiltrukket sig opmærksomhed fra specialister fra mere end 50 lande i verden - fra Latinamerika til Asien og Stillehavsområdet - i det første år efter udgivelsen. For at diskutere individuelle samarbejdsvilkår og få detaljerede oplysninger om placeringsmuligheder, bedes du udfylde feedbackformularen på den officielle portal datadrivenconstruction.io er i eller skrive til de kontaktpersoner, der anført slutningen af bogen.



BOGENS KAPITLER ER TILGÆNGELIGE PÅ DATADRIVENCONSTRUCTION.IO

Du kan læse kapitler af Datadrevet byggeri på hjemmesiden for Datadrevet byggeri, hvor dele af bogen løbende bliver offentliggjort, så du hurtigt kan finde de oplysninger, du har brug for, og bruge dem i dit arbejde. Du kan også finde mange andre publikationer om lignende emner samt eksempler på anvendelser og løsninger, der kan hjælpe dig med at udvikle dine færdigheder og anvende data i byggeriet.



SENESTE VERSIONER AF BOGEN DOWNLOAD FRA DEN OFFICIELLE HJEMMESIDE

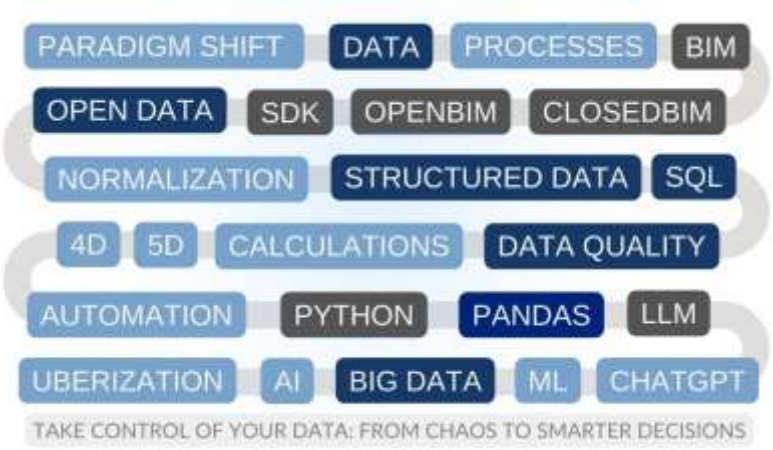
Den aktuelle og seneste version af DataDrivenConstruction-bogen kan downloades på datadrivenconstruction.io. Hvis du gerne vil modtage opdateringer med nye bogkapitler, praktiske tips eller anmeldelser af nye applikationer, kan du abonnere på nyhedsbrevet:

- Du vil være den første til at blive introduceret til de nye afsnit i bogen
- Modtag praktiske casestudier og råd om analyse og automatisering i byggebranchen
- Hold dig opdateret med trends, publikationer og anvendelseseksempler

Gå til datadrivenconstruction.io for at abonnere!

DATADREVET KONSTRUKTION:
RÅDGIVNING, WORKSHOPS OG UDDANNELSE

DataDrivenConstructions træningsprogrammer og rådgivning har hjulpet snesevis af førende byggevirksomheder over hele verden med at øge effektiviteten, reducere omkostningerne og forbedre kvaliteten af løsningerne. DataDrivenConstructions kunder omfatter nogle af de største aktører på milliardmarkedet, herunder bygge-, konsulent- og IT-virksomheder.



Hvorfor vælge os?

- **Relevans**: at tale om de vigtigste tendenser og indsigter i branchen
- **Praktisk:** hjælper fagfolk med effektivt at løse hverdagsopgaver gennem PoC.
- **Personlig tilgang:** skræddersyet til din virksomhed, maksimering af fordelene ved træning og rådgivning

DataDrivenConstruction-teamets vigtigste fokusområder:

- **Styring af datakvalitet**: Hjælp med at parametrisere opgaver, indsamle krav, validere og forberede data til automatiseret behandling.
- **Data Mining - udtræk og strukturering af data:** opsæt ETL-processer og udtræk data fra e-mails, PDF, Excel, billeder og andre kilder.
- **BIM og CAD analytics**: indsamle, strukturere og analysere oplysninger fra RVT-filer, IFC, DWG og andre CAD (BIM)-formater.
- **Analyse og datatransformation:** Omdannelse af forskellig information til strukturerede data, analyser, indsigt og beslutninger.
- **Dataintegration og procesautomatisering**: fra automatisk oprettelse af dokumenter til integration med interne systemer og eksterne databaser.

Kontakt DataDrivenConstruction.io for at høre, hvordan automatisering kan hjælpe din virksomhed med at opnå håndgribelige forretningsresultater.

GLOSSAR

AI (Artificial Intelligence) - Kunstig intelligens; computersystemernes evne til at udføre opgaver, der normalt kræver menneskelig intelligens, som f.eks. mønstergenkendelse, indlæring og beslutningstagning.

Apache Airflow er en open source workflow-orkestreringsplatform, der giver dig mulighed for programmatisk at oprette, planlægge og spore workflows og ETL'er ved hjælp af DAG'er (directed acyclic graphs).

Apache NiFi er et værktøj til automatisering af datastrømme mellem systemer med speciale i datarouting og -transformation.

Apache Parquet er et effektivt filformat til kolonnelagring af data, optimeret til brug i store dataanalysesystemer. Det giver betydelig komprimering og hurtig behandling.

API (Application Programming Interface) - en formaliseret grænseflade, der gør det muligt for et program at interagere med et andet uden adgang til kildekoden og udveksle data og funktionalitet via standardiserede anmodninger og svar.

Attribut - et kendetegn eller en egenskab ved et objekt, der beskriver dets egenskaber (f.eks. areal, volumen, pris, materiale).

Databaser er organiserede strukturer til lagring, styring og adgang til information, som bruges til effektiv datahentning og -behandling.

BEP (BIM Execution Plan) - En implementeringsplan for bygningsinformationsmodellering, der definerer mål, metoder og processer for implementering af BIM i et projekt.

Big Data - informationsmængder af betydelig størrelse, variation og opdateringshastighed, som kræver særlige teknologier til behandling og analyse.

BI (Business Intelligence) - Business **Intelligence**; processer, teknologier og værktøjer til at omdanne data til meningsfuld information til brug for beslutningstagning.

BIM (Building Information Modeling**)** - Bygningsinformationsmodellering; processen med at skabe og styre digitale repræsentationer af byggeprojekters fysiske og funktionelle egenskaber, herunder ikke kun 3D-modeller, men også oplysninger om ydeevne, materialer, timing og omkostninger.

BlackBox/WhiteBox - tilgange til at forstå systemet: I det første tilfælde er den interne logik skjult, kun input og output er synlige; i det andet tilfælde er behandlingen gennemsigtig og tilgængelig for analyse.

Bounding Box er en geometrisk konstruktion, der beskriver grænserne for et objekt i det tredimensionelle rum ved hjælp af de minimale og maksimale X-, Y- og Z-koordinater, hvilket skaber en "kasse" omkring objektet.

BREP (Boundary Representation) er en geometrisk repræsentation af objekter, der definerer dem gennem overfladernes grænser.

CAD (Computer-Aided Design) er et computerstøttet designsystem, der bruges til at skabe, redigere og analysere nøjagtige tegninger og 3D-modeller inden for arkitektur, byggeri, teknik og andre brancher.

CAFM (Computer-Aided Facility Management) er en software til ejendoms- og infrastrukturforvaltning, der omfatter planlægning af rum, forvaltning af aktiver, vedligeholdelse og omkostningsovervågning.

CDE (Common Data Environment) - et centraliseret digitalt rum til styring, lagring, deling og samarbejde med projektinformation i alle faser af anlæggets livscyklus.

Et **Centre of Excellence (CoE)** er en specialiseret struktur i en organisation, der er ansvarlig for at udvikle et specifikt vidensområde, udvikle standarder og bedste praksis, uddanne personale og støtte indførelsen af innovationer.

CoClass er et moderne tredjegenerations klassifikationssystem for bygningselementer.

En konceptuel datamodel er en repræsentation på højt niveau af grundlæggende enheder og deres relationer uden attributdetaljer, som bruges i de indledende faser af databasedesignet.

CRM (Customer Relationship Management) er et system til styring af kunderelationer, som bruges til at automatisere salgs- og serviceprocesser.

DAG (Directed Acyclic Graph) er en rettet acyklisk graf, der bruges i dataorkestreringssystemer (Airflow, NiFi) til at bestemme opgavesekvenser og afhængigheder.

Dash er en Python-ramme til at skabe interaktive webbaserede datavisualiseringer.

Dashboard - Et **dashboard**, der visuelt præsenterer vigtige præstationsindikatorer og metrikker i realtid.

Den datacentrerede tilgang er en metode, der prioriterer data frem for applikationer eller softwarekode og gør data til organisationens centrale aktiv.

Datastyring - et sæt praksisser, processer og politikker, der sikrer en passende og effektiv brug af data i en organisation, herunder adgangs-, kvalitets- og sikkerhedskontrol.

Data Lake er en lagerfacilitet, der er designet til at opbevare store mængder rådata i deres oprindelige format, indtil de skal bruges.

Data Lakehouse er en arkitektonisk tilgang, der kombinerer fleksibiliteten og skalerbarheden i datasøer (Data **Lake**) med håndterbarheden og ydeevnen i datalagre (DWH).

Datadrevet byggeri er en strategisk tilgang, hvor alle faser af byggeriets livscyklus - fra design til drift - understøttes af automatiserede, indbyrdes forbundne systemer. Denne tilgang giver kontinuerlig, faktabaseret læring, reducerer usikkerhed og gør det muligt for virksomheder at opnå bæredygtig brancheledelse.

Data-Driven Integrator - en virksomhed, der har specialiseret sig i at kombinere data fra forskellige kilder og analysere dem for at træffe ledelsesbeslutninger.

Datadrevet tilgang - en metode, hvor data ses som et strategisk aktiv, og hvor beslutninger træffes på baggrund af en objektiv analyse af information frem for subjektive meninger.

Dataminimalisme - en tilgang til at reducere data til det mest værdifulde og meningsfulde, hvilket giver mulighed for forenklet behandling og analyse af information.

Datasump - En spredt masse af ustrukturerede data, der opstår, når information indsamles og lagres på en ukontrolleret måde uden ordentlig organisering.

DataOps er en metode, der kombinerer DevOps-principper, data og analyse med fokus på at forbedre samarbejde, integration og automatisering af datastrømme.

Digitalisering **af information** er processen med at konvertere alle aspekter af byggeaktiviteter til en digital form, der er egnet til analyse, fortolkning og automatisering.

DataFrame - En todimensionel tabeldatastruktur i Pandas-biblioteket, hvor rækkerne repræsenterer individuelle poster eller objekter, og kolonnerne repræsenterer deres egenskaber eller attributter.

Deskriptiv analyse - Analyse af historiske data for at forstå, hvad der skete i fortiden.

Diagnostisk analyse - Analyse af data for at finde ud af, hvorfor noget skete.

Et Gantt-diagram er et projektplanlægningsværktøj, der repræsenterer opgaver som vandrette søjler på en tidslinje, så du kan visualisere arbejdets rækkefølge og varighed.

DWH (Data Warehouse) er et centraliseret data warehouse-system, der samler information fra flere kilder, strukturerer den og gør den tilgængelig for analyse og rapportering.

ESG (Environmental, Social, Governance) - et sæt kriterier til vurdering af en virksomheds eller et projekts miljømæssige, sociale og ledelsesmæssige konsekvenser.

ELT (Extract, Load, Transform) er en proces, hvor data først udtrækkes fra kilder og indlæses i et repository og derefter transformeres til analytiske formål.

ETL (Extract, Transform, Load) er processen med at udtrække data fra forskellige kilder, omdanne dem til det ønskede format og indlæse dem i mållageret til analyse.

ER-diagram (Entity-Relationship) - et visuelt diagram, der viser enheder, deres attributter og relationerne mellem dem, og som bruges i datamodellering.

ERP (Enterprise Resource Planning) er et omfattende modulopbygget ressourceplanlægningssystem, der bruges til at styre og optimere forskellige aspekter af byggeprocessen.

Funktioner - I maskinlæring er det uafhængige variabler eller attributter, der bruges som input til en model.

Fysisk datamodel - en detaljeret repræsentation af databasestrukturen, herunder tabeller, kolonner, datatyper, nøgler og indekser, optimeret til et bestemt DBMS.

FPDF er et Python-bibliotek til oprettelse af PDF-dokumenter.

Geometric Core er en softwarekomponent, der indeholder grundlæggende algoritmer til at skabe, redigere og analysere geometriske objekter i CAD, BIM og andre tekniske applikationer.

HiPPO (Highest Paid Person's Opinion) - en tilgang til beslutningstagning baseret på den højest betalte persons mening i organisationen i stedet for objektive data.

IDE (Integrated Development Environment) - integreret udviklingsmiljø, et omfattende værktøj til at skrive, teste og fejlfinde kode (f.eks. PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) er en specifikation for informationslevering, der definerer datakravene i forskellige faser af et projekt.

IFC (Industry Foundation Classes) er et format til udveksling af BIM-data, som sikrer interoperabilitet mellem forskellige softwareløsninger.

Industri 5.0 er et industrielt udviklingskoncept, der kombinerer mulighederne i digitalisering, automatisering og kunstig intelligens med menneskeligt potentiale og miljømæssig bæredygtighed.

Dataintegration er processen med at kombinere data fra forskellige kilder til et enkelt, sammenhængende system for at give et samlet overblik over information.

Informationssiloer er isolerede datalagringssystemer, som ikke deler information med andre systemer, hvilket skaber barrierer for effektiv brug af data.

IoT (Internet of Things) er konceptet med at forbinde fysiske objekter til internettet for at indsamle, behandle og overføre data.

k-NN (k-Nearest Neighbors) er en maskinlæringsalgoritme, der klassificerer objekter baseret på lighed med de nærmeste naboer i træningsprøven.

Kaggle er en platform for dataanalyse og maskinlæringskonkurrencer.

Beregning - beregning af omkostningerne ved byggearbejder eller -processer for en bestemt måleenhed (f.eks. 1 m² gipspladevæg, 1 m³ beton).

KPI'er (Key Performance Indicators) er centrale præstationsindikatorer, kvantificerbare målinger, der bruges til at evaluere en virksomheds eller et specifikt projekts succes.

Etiketter - I maskinlæring de målvariabler eller attributter, som modellen skal forudsige.

Læringsalgoritme - Processen med at finde den bedste hypotese i en model, der svarer til en målfunktion, ved hjælp af et sæt træningsdata.

Lineær regression - En statistisk metode til modellering af forholdet mellem en afhængig variabel og en eller flere uafhængige variabler.

LLM (Large Language Model) - Large Language Model, en kunstig intelligens, der er trænet til at forstå og generere tekst fra store mængder data, og som er i stand til at analysere kontekst og skrive programmeringskode.

LOD (Level of Detail/Development) - modellens detaljeringsgrad, der bestemmer graden af geometrisk nøjagtighed og informationsindhold.

En logisk datamodel er en detaljeret beskrivelse af enheder, attributter, nøgler og relationer, der afspejler forretningsinformation og -regler, et mellemstadie mellem den konceptuelle og den fysiske model.

Machine Learning - En klasse af teknikker til kunstig intelligens, der gør det muligt for computersystemer at lære og komme med forudsigelser ud fra data uden eksplicit programmering.

Masterformat er et førstegenerations klassifikationssystem, der bruges til at strukturere byggespecifikationer i sektioner og discipliner.

MEP (Mechanical, Electrical, Plumbing) - Bygningstekniske systemer, der omfatter mekaniske, elektriske og sanitære komponenter.

Mesh er en maskerepræsentation af 3D-objekter, der består af hjørner, kanter og flader.

Model - I maskinlæring et sæt af forskellige hypoteser, hvoraf den ene tilnærmer sig den målfunktion, der skal forudsiges eller tilnærmes.

Datamodellering er processen med at skabe en struktureret repræsentation af data og deres relationer med henblik på implementering i informationssystemer, herunder konceptuelle, logiske og fysiske niveauer.

n8n er et open source-værktøj til automatisering af arbejdsgange og integration af applikationer via en low-code-tilgang.

Normalisering - i maskinlæring, processen med at bringe forskellige numeriske data til en fælles skala for at lette behandling og analyse.

Reverse engineering - processen med at studere et objekts enhed, funktion og fremstillingsteknologi ved at analysere dets struktur, funktioner og drift. I forbindelse med data - udtrækning af information fra proprietære formater til brug i åbne systemer.

OCR (Optical Character Recognition) er en optisk tegngenkendelsesteknologi, der konverterer tekstbilleder (scannede dokumenter, fotos) til et maskinlæsbart tekstformat.

OmniClass er anden generation af en international klassifikationsstandard til styring af byggeinformation.

Ontologi - Et system af indbyrdes relationer mellem begreber, der formaliserer et bestemt vidensområde.

Open Source - en model for udvikling og distribution af open source-software, der er tilgængelig til fri brug, undersøgelse og ændring.

Open BIM er begrebet åben BIM, som indebærer brug af åbne standarder og formater til dataudveksling mellem forskellige softwareløsninger.

Åbne standarder - offentligt tilgængelige specifikationer til at udføre en bestemt opgave, som gør det muligt for forskellige systemer at interagere og udveksle data.

Pandas er et open source Python-bibliotek til databehandling og -analyse, der leverer DataFrame- og Series-datastrukturer til effektiv håndtering af tabeloplysninger.

Det åbne dataparadigme er en tilgang til databehandling, hvor information gøres frit tilgængelig til brug, genbrug og formidling af alle.

Parametrisk metode er en metode til estimering af byggeprojekter, der bruger statistiske modeller til at estimere værdien baseret på projektparametre.

PIMS (Project Information Model) er et digitalt system, der er designet til at organisere, lagre og dele al projektinformation.

Pipeline - En sekvens af databehandlingsprocesser, fra udtræk og transformation til analyse og visualisering.

PMIS (Project Information Management System) er et projektstyringssystem, der er designet til detaljeret opgavestyring på niveauet for et individuelt byggeprojekt.

Predictive Analytics er en del af analytics, der bruger statistiske metoder og maskinlæring til at forudsige fremtidige resultater baseret på historiske data.

Præskriptiv analyse - En del af analysen, der ikke kun forudsiger fremtidige resultater, men også foreslår optimale handlinger for at opnå de ønskede resultater.

Proprietære formater er lukkede dataformater, der kontrolleres af en bestemt virksomhed, og som begrænser muligheden for at dele information og øger afhængigheden af specifik software.

QTO (Quantity Take-Off) er processen med at udtrække mængderne af elementer fra designdokumenterne for at beregne de mængder af materialer, der kræves for at realisere projektet.

Kvalitetsstyringssystem - et kvalitetsstyringssystem, der sikrer, at processer og resultater opfylder fastlagte krav.

RAG (Retrieval-Augmented Generation) er en metode, der kombinerer sprogmodellers generative evner med udtræk af relevant information fra virksomhedsdatabaser, hvilket forbedrer svarenes nøjagtighed og relevans.

RDBMS (Relational Database Management System) er et relationelt databasestyringssystem, der organiserer information i form af indbyrdes forbundne tabeller.

RegEx (Regular Expressions) er et formaliseret sprog til søgning og behandling af strenge, som gør det muligt at specificere skabeloner til kontrol af tekstdata for overholdelse af bestemte kriterier.

Regression er en statistisk metode til at analysere forholdet mellem variabler.

CO_2-beregninger er en metode til at estimere kuldioxidemissioner i forbindelse med produktion og brug af byggematerialer og -processer.

Ressourcemetoden er en metode til at lave overslag baseret på en detaljeret analyse af alle de ressourcer (materialer, arbejdskraft, udstyr), der er nødvendige for at udføre byggearbejdet.

RFID (Radio Frequency Identification) er en teknologi til automatisk identifikation af objekter ved hjælp af radiosignaler, som bruges til at spore materialer, maskiner og personale.

ROI (Return on Investment) er en indikator, der afspejler forholdet mellem profit og investerede midler, og som bruges til at vurdere effektiviteten af investeringer.

SaaS (Software as a Service) er en model for software som en tjeneste, hvor applikationer hostes af en udbyder og stilles til rådighed for brugerne via internettet.

SCM (Supply Chain Management) - styring af forsyningskæden, som omfatter koordinering og optimering af alle processer fra indkøb af materialer til levering af færdige produkter.

Datasiloer er isolerede informationslagre i en organisation, som ikke er integreret med andre systemer, hvilket gør det svært at dele data og ineffektivt.

SQL (Structured Query **Language)** er et struktureret forespørgselssprog, der bruges til at arbejde med relationsdatabaser.

SQLite er et let, integrerbart DBMS på tværs af platforme, som ikke kræver en separat server, og som understøtter grundlæggende SQL-funktioner, der er meget udbredt i mobilapplikationer og integrerede systemer.

Strukturerede data - information, der er organiseret i et bestemt format med en klar struktur, f.eks. i relationsdatabaser eller tabeller.

Løst strukturerede data - information med delvis organisering og fleksibel struktur, såsom JSON eller XML, hvor forskellige elementer kan indeholde forskellige sæt af attributter.

En entitet er et konkret eller abstrakt objekt i den virkelige verden, som kan identificeres, beskrives og repræsenteres entydigt i form af data.

Supervised Learning - En type maskinlæring, hvor en algoritme trænes på opdelte data, hvor det ønskede resultat er kendt for hvert eksempel.

Taksonomi er et hierarkisk klassifikationssystem, der bruges til systematisk at kategorisere elementer baseret på fælles træk.

Titanic Dataset er et populært datasæt til træning og test af maskinlæringsmodeller.

Træning - Den proces, hvor en maskinlæringsalgoritme analyserer data for at identificere mønstre og danne en model.

Transfer learning er en maskinlæringsteknik, hvor en model, der er trænet til én opgave, bruges som udgangspunkt for en anden opgave.

Transformation - Processen med at ændre formatet, strukturen eller indholdet af data til senere brug.

Datakrav - formaliserede kriterier, der definerer strukturen, formatet, fuldstændigheden og kvaliteten af de oplysninger, der kræves for at understøtte forretningsprocesserne.

Uberisering af byggebranchen er en proces, hvor traditionelle forretningsmodeller i byggebranchen omdannes under indflydelse af digitale platforme, der giver direkte interaktion mellem kunder og entreprenører uden mellemled.

Uniclass er et anden- og tredjegenerations klassifikationssystem for bygningselementer, som er meget udbredt i Storbritannien.

USD (Universal Scene Description) er et dataformat, der er udviklet til computergrafik, men som har vundet indpas i tekniske systemer på grund af dets enkle struktur og uafhængighed af geometriske kerner.

Datavalidering er den proces, hvor man kontrollerer oplysninger i forhold til fastlagte kriterier og krav for at sikre, at data er nøjagtige, fuldstændige og konsistente.

Vektordatabase - En specialiseret type database, der gemmer data som flerdimensionelle vektorer til effektiv semantisk søgning og sammenligning af objekter.

Vektorrepræsentation (indlejring) er en metode til at omdanne data til flerdimensionelle numeriske vektorer, som gør det muligt for maskinalgoritmer at behandle og analysere information effektivt.

VectorOps er en metode, der fokuserer på behandling, lagring og analyse af multidimensionelle vektordata, som er særligt relevante inden for områder som digitale tvillinger og semantisk søgning.

Visualisering - Grafisk repræsentation af data for bedre opfattelse og analyse af information.

Den alfabetiske kategorisering af termerne blev foretaget efter deres engelske navne.

LISTE OVER REFERENCER OG ONLINE-MATERIALER

[1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12. december 2016. [Online]. Tilgængelig: https://www.gartner.com/en/documents/3530919. [Dato for adgang: 1. marts 2025].

[2] KPMG, "Velkendte udfordringer - nye tilgange. 2023 Global Construction Survey", 1. januar 2023. [Online]. Available: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf. [Tilgået: 5. marts 2025].

[3] F. R. Barnard, "Et billede siger mere end tusind ord", 10 Mapi 1927. [Online]. Tilgængelig: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Referencedato: 15. marts 2025].

[4] M. Bastian, "Microsofts CEO Satya Nadella siger, at selvudnævnelse af AGI er 'meningsløs benchmark-hacking'," 21. februar 2025. [Online]. Tilgængelig: https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/. [Dato for adgang: 15. marts 2025].

[5] W. E. Forum, "Forum Shaping the Future of Construction - A Landscape in Transformation:", 1. januar 2016. [Online]. Tilgængelig: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Dato for adgang: 2. marts 2025].

[6] C. D. Gillespie, "Ler: jordens sammenfiltring i lerets tidsalder", 2024. [Online]. Tilgængelig: https://ufl.pb.unizin.org/imos/chapter/clay/.

[7] "Papyrus fra det 3. århundrede f.v.t. Sproget er græsk", 2024. [Online]. Available: https://www.facebook.com/429710190886668/posts/595698270954525.

[8] "Overvågning: brug af de værktøjer, der er til rådighed", 1980. [Online]. Tilgængelig: https://pubmed.ncbi.nlm.nih.gov/10246720/. [Dato for adresse: 15. marts 2025].

[9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15. februar 2015. [Online]. Tilgængelig: https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf. [Dato for adgang: 15. marts 2025].

[10] Skanska USA, "Fall Construction Market Trends," 2. november 2023. [Online]. Available: https://x.com/SkanskaUSA/status/1720167220817588714.

[11] "Oxford Essential Quotations (4. udgave)", Oxford University Press, 2016. [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236. [Dato for adgang: 1. marts 2025].

[12] "Citat: Søndergaard om dataanalyse," [Online]. Available: https://www.causeweb.org/cause/resources/library/r2493. [Dato for adgang: 15. marts 2025].

[13] "Hvordan den globale AI-interesse øger markedet for datahåndtering", 28. maj 2024. [Online]. Tilgængelig: https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/. [Dato for adgang: 15. marts 2025].

[14] И. McCue, "ERP-historie", 2024. [Online]. Tilgængelig: https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml.

[15] erpscout, "ERP-pris: Hvor meget koster et ERP-system?"," [Online]. Available: https://erpscout.de/en/erp-costs/. [Dato for adgang: 15. marts 2025].

[16] softwarepath, "What 1,384 ERP projects tell us about selecting ERP (2022 ERP report)," 18 Jan. 2022. [Online]. Available: https://softwarepath.com/guides/erp-report. [Dato for adgang: 15. marts 2025].

[17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16. december 2016. [Online]. Tilgængelig: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf. [Dato for adresse: 1. marts 2025].

[18] Mckinsey, "Den datadrevne virksomhed i 2025", 28. januar 2022. [Online]. Tilgængelig: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025. [Referencedato: 22. maj 2024].

[19] Wikipedia, "Moores lov". [Online]. Tilgængelig: https://en.wikipedia.org/wiki/Moore%27s_law. [Dato for adgang: 15. marts 2025].

[20] Accenture, "Building More Value With Capital Projects", 1. januar 2020. [Online]. Tilgængelig: https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf. [Dato for adgang: 3. marts 2024].

[21] Б. Marr, "Hvor meget data skaber vi hver dag? De forbløffende statistikker, som alle bør læse", 2018. [Online]. Tilgængelig: 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] "Hvor meget data produceres der hver dag?", 2024. [Online]. Tilgængelig: https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/.

[23] T. Sullivan, "AI og den globale 'datasfære': hvor meget information vil menneskeheden have i 2025?", 2024. [Online]. Tilgængelig: https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By- 2025.html.

[24] Statista, "Samlet antal trykte bøger produceret i forskellige regioner i Vesteuropa i hvert halve århundrede mellem 1454 og 1800," [Internet]. [Online]. Available: https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/. [Referencedato: 1. marts 2025].

[25] "Eksempler på prissætning", 2024. [Online]. Tilgængelig: https://cloud.google.com/storage/pricing-examples.

[26] M. Ashare, "Enterprises outsource data storage as complexity rises", 10. maj 2024. [Online]. Tilgængelig: https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/. [Dato for adgang: 15. marts 2025].

[27] JETSOFTPRO, "SaaS er død? Microsoft CEO's chokerende forudsigelse forklaret", 13. januar 2025. [Online]. Tilgængelig: https://jetsoftpro.com/blog/saas-is-dead/.

[28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12. december 2024. [Online]. Tilgængelig: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Dato for adgang: 15. marts 2025].

[29] GoodReads, "Tim Berners-Lee," [Online]. Tilgængelig: https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than. [Dato for adgang: 15. marts 2025].

[30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1. januar 2023. [Online]. Tilgængelig: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Tilgået: 5. marts 2025].

[31] И. Deininger, B. Koch, R. Bauknecht og M. Langhans, "Using digital models for decarbonising a production site: An example of connecting a building model, a production model and an energy model", 2024. [Online]. Tilgængelig: https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Study_of_Connecting_the_Building_Model_Production_Model_and_Energy_Model.

[32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY", 1. februar 2017. [Online]. Tilgængelig: 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] Construction Task Force to the Deputy Prime Minister, "Rethinking Construction", 1. oktober 2014. [Online]. Tilgængelig: 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. marts 2016. [Online]. Available: https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/. [Dato for adgang: 15. marts 2025].

[35] Wikiquote, "Charles Babbage", [Online]. Available: https://en.wikiquote.org/wiki/Charles_Babbage. [Dato for adgang: 15. marts 2025].

[36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves", 12. marts 2025. [Online]. Tilgængelig: https://news.sap.com/2025/03/new-research-executive-trust-ai/. [Dato for adgang: 15. marts 2025].

[37] The Canadian Construction Association og KPMG i Canada, 2021, "Construction in a digital world", 1. maj 2021. [Online]. Available: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Dato for adgang: 5. marts 2025].

[38] ZCS, "Afkodning af den femte industrielle revolution". [Online]. Tilgængelig: https://www.pwc.in/decoding-the-fifth-industrial-revolution.html. [Dato for adgang: 15. marts 2025].

[39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.

[40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "The Value of Open Source Software", 24. januar 2024. [Online]. Tilgængelig: https://www.hbs.edu/faculty/Pages/item.aspx?num=65230. [Dato for adgang: 15. marts 2025].

[41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1. september 2008. [Online]. Tilgængelig: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf.

[42] McKinsey, "Forbedring af produktiviteten i byggeriet", [Online]. Tilgængelig: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity. [Dato for adresse: 15. marts 2025].

[43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector", 19. januar 2023. [Online]. Available: https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/. [Dato for adgang: 1. marts 2025].

[44] McKinsey, "At levere produktivitet i byggeriet er ikke længere valgfrit", 9. august 2024. [Online]. Tilgængelig: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional. [Dato for adgang: 5. marts 2025].

[45] ING Group, "Lagging productivity in construction is driving up construction costs," 12. december 2022. [Online]. Tilgængelig: https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/. [Dato for adresse: 15. marts 2025].

[46] M. Berman, "Microsoft CEO's Shocking Prediction: 'Agents Will Replace ALL Software'", 19. december 2024. [Online]. Tilgængelig: https://www.youtube.com/watch?v=uGOLYz2pgr8. [Dato for

adgang: 15. marts 2025].

[47] Business Insider, "Anthropics CEO siger, at om 3 til 6 måneder vil AI skrive 90 % af den kode, som softwareudviklere havde ansvaret for", 15. marts 2025. [Online]. Tilgængelig: https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3. [Dato for adgang: 30. marts 2025].

[48] Statista, "Sammenligning af populariteten af databasestyringssystemer (DBMS'er) på verdensplan fra juni 2024, efter kategori", juni 2024. [Online]. Tilgængelig: https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/. [Dato for adresse: 15. marts 2025].

[49] DB-Engines, "DB-Engines Ranking", [Online]. Available: https://db-engines.com/en/ranking. [Referencedato: 15. marts 2025].

[50] "Stack Overflow Developer Survey 2023", 2024. [Online]. Tilgængelig: https://survey.stackoverflow.co/2023/.

[51] "SQL", 2024. [Online]. Tilgængelig: https://en.wikipedia.org/wiki/SQL.

[52] "Strukturerede og ustrukturerede data: Hvad er forskellen?", 2024. [Online]. Tilgængelig: https://www.ibm.com/blog/structured-vs-unstructured-data/.

[53] DataDrivenConstruction, "PDF COMPARISON OF DATA FORMATS FOR CONSTRUCTION PROJECTS," 23 Apr 2024. [Online]. Tilgængelig: https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf.

[54] "Building Information Modeling Whitepaper site," 2003. [Online]. Tilgængelig: https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112.

[55] A. Boyko, "Lobbykrig og BIM-udvikling. Del 5: BlackRock er herre over alle teknologier. Hvordan virksomheder kontrollerer open source-kode," 2024. [Online]. Tilgængelig: https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/.

[56] D. Ushakov, "Direkte modellering - hvem og hvorfor har brug for det? En gennemgang af konkurrencedygtige teknologier", 14.11.2011. [Online]. Tilgængelig: https://isicad.net/articles.php?article_num=14805. [Dato for adresse: 02 2025].

[57] C. Eastman og A. Cthers, "Eastman, Charles; og Cthers", september 1974. [Online]. Available: https://files.eric.ed.gov/fulltext/ED113833.pdf. [Dato for adgang: 15. marts 2025].

[58] D. Ushakov, "Direkte modellering - hvem og hvorfor har brug for det? En gennemgang af konkurrencedygtige teknologier", 11. november 2011. [Online]. Tilgængelig: https://isicad.net/articles.php?article_num=14805. [Referencedato: 15. marts 2025].

[59] D. Weisberg, "CAD's historie", 12. december 2022. [Online]. Tilgængelig: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Referencedato: 15. marts 2025].

[60] ADSK, "Hvidbog om bygningsinformationsmodellering", 2002. [Online]. Available: https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Referencedato: 15. marts 2025].

[61] ADSK, "Hvidbog om bygningsinformationsmodellering i praksis", [Online]. Available: https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Dato for adgang: 15. marts 2025].

[62] A. Boyko, "Lobbykrige og BIM-udvikling. Del 2: åben BIM VS lukket BIM. Europa VS resten af verden," 2024. [Online]. Tilgængelig: 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. [Online]. Tilgængelig: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.

[64] ADSK, "Hvidbog BIM", 2002. [Online]. Tilgængelig: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Dato for adresse: 15. marts 2025].

[65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [Online]. Tilgængelig: https://web.archive.org/web/20010615093351/http:/www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Dato for adgang: 15. marts 2025].

[66] M. Shacklett, "Strukturerede og ustrukturerede data: Vigtige forskelle", 2024. [Online]. Tilgængelig: https://www.datamation.com/big-data/structured-vs-unstructured-data/.

[67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Tilgængelig: https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/.

[68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the", 2004. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf . [Dato for adgang: 02 2025].

[69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Tilgængelig: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Dato for adgang: 15. marts 2025].

[70] Analyticsindiamag, "6 mest tidskrævende opgaver for dataforskere", 15. maj 2019. [Online]. Tilgængelig: https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/.

[71] BizReport, "Report: Data scientists spend bulk of time cleaning up," 06. juli 2015. [Online]. Available: https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html. [Dato for adgang: 5. marts 2025].

[72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27. juli 2015. [Online]. Tilgængelig: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Dato for adgang: 15. marts 2025].

[73] Б. Cyphers og K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [Online]. Tilgængelig: https://www.eff.org/wp/interoperability-and-privacy.

[74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 1. oktober 2013. [Online]. Tilgængelig: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information. [Dato for adgang: 15. marts 2025].

[75] A. Boyko, "Kampen for åbne data i byggebranchen. Historien om AUTOLISP, intelliCAD, openDWG, ODA og openCASCADE," 15 05 2024. [Online]. Tilgængelig: https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393. [Dato for adresse: 16 02 2025].

[76] Wikipedia, "Microsoft og open source", [Online]. Tilgængelig: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Dato for adgang: 15. marts 2025].

[77] TIME, "Forskellen mellem åbne og lukkede AI-modeller bliver måske mindre. Her er, hvorfor det betyder noget", 5. november 2024. [Online]. Tilgængelig: https://time.com/7171962/open-closed-ai-models-epoch/. [Dato for adgang: 15. marts 2025].

[78] The Verge, "Mere end en fjerdedel af den nye kode hos Google genereres af AI", 29. oktober 2024. [Online]. Available: https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024. [Dato for adgang: 15. marts 2025].

[79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15. december 2020. [Online]. Tilgængelig: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-processing. [Tilgået: 15. marts 2025].

[80] PWC, "PwC Open Source Monitor 2019", 2019. [Online]. Available: https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf. [Tilgået 15. marts 2025].

[81] Travers Smith, "Den åbne hemmelighed: Open Source-software", 2024. [Online]. Tilgængelig: https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/. [Dato for adgang: 15. marts 2025].

[82] Deloitte, "Dataoverførselsprocessen i virksomhedstransformationer", 2021. [Online]. Tilgængelig: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-

process-in-corporate-transformations.pdf. [Referencedato: 15. marts 2025].

[83] gov.uk, "Data Analytics and AI in Government Project Delivery," 20. marts 2024. [Online]. Available: https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery. [Dato for adgang: 5. marts 2025].

[84] "Citat Origin: Alt skal gøres så enkelt som muligt, men ikke enklere," 13. maj 2011. [Online]. Tilgængelig: https://quoteinvestigator.com/2011/05/13/einstein-simple/. [Dato for adgang: 15. marts 2025].

[85] "Transformer (arkitektur for dyb læring)". [Online]. Tilgængelig: https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture). [Dato for adgang: 15. marts 2025].

[86] "Python Packages Download Stats," 2024. [Online]. Tilgængelig: https://www.pepy.tech/projects/pandas.

[87] Interview Bit, "Top 10 Python-biblioteker", 2023. [Online]. Tilgængelig: https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more. [Dato for adgang: 30. marts 2025].

[88] "NVIDIA and HP Supercharge Data Science and Generative AI on Workstations", 7. marts 2025. [Online]. Tilgængelig: https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations. [Dato for adgang: 15. marts 2025].

[89] P. Orac, "Sådan behandler du en DataFrame med millioner af rækker på få sekunder", 2024. [Online]. Tilgængelig: https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds.

[90] Ç. Uslu, "Hvad er Kaggle?", 2024. [Online]. Tilgængelig: https://www.datacamp.com/blog/what-is-kaggle.

[91] "NVIDIA CEO Jensen Huang Keynote på COMPUTEX 2024," 2. juni 2024. [Online]. Tilgængelig: https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w. [Tilgået 15. marts 2025].

[92] "Medlemmer: Stiftere og virksomhedsmedlemmer," 2024. [Online]. Tilgængelig: https://www.opendesign.com/member-showcase.

[93] A. Boyko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts," 24. november 2024. [Online]. Available: 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. [Tilgået: 23. februar 2025].

[94] A. Boyko, "Verden efter BIM. Overgang til data og processer, og om byggebranchen har brug for semantik, formater og interoperabilitet," 20. dec. 2024. [Online]. Tilgængelig:

https://boikoartem.medium.com/the-post-bim-world-7e35b7271119. [Dato for adresse: 23. februar 2025].

[95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE," 2016. [Online]. Available: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Dato for adgang: 23. februar 2025].

[96] Harvard Business Review, "Bad Data Costs the U.S. $3 Trillion Per Year", 22. september 2016. [Online]. Tilgængelig: https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

[97] Delpha, "Impacts of Data Quality", 1. januar 2025. [Online]. Tilgængelig: https://delpha.io/impacts-of-data-quality/.

[98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects," [Online]. Available: https://www.wbdg.org/resources/design-for-maintainability. [Dato for adgang: 15. marts 2025].

[99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", april 2014. [Online]. Tilgængelig: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf. [Dato for adgang: 15. marts 2025].

[100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1. januar 2025. [Online]. Tilgængelig: https://www.gartner.com/en/data-analytics/topics/data-quality.

[101] "I mangel af et søm" [Online]. Tilgængelig: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Dato for adgang: 15. marts 2025].

[102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", oktober 2013. [Online]. Tilgængelig: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [Dato for adgang: 15. marts 2025].

[103] EY, "Vejen til kulstofneutralitet", 10. marts 2023. [Online]. Tilgængelig: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Dato for adgang: 15. marts 2025].

[104] PWC, "ESG Awareness", 1. juli 2024. [Online]. Tilgængelig: https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf. [Dato for adgang: 15. marts 2025].

[105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)," 2024. [Online]. Tilgængelig: https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf.

[106] "CO2_beregning af det indbyggede kulstof", 2024. [Online]. Tilgængelig:

https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.

[107] McKinsey, "Imagining Construction's Digital Future", 24. juni 2016. [Online]. Tilgængelig: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future. [Dato for adgang: 25. februar 2025].

[108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch," 10. oktober 2024. [Online]. Tilgængelig: https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/. [Referencedato: 15. marts 2025].

[109] SAS, "Data lake and data warehouse - know the difference," [Online]. Available: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Referencedato: 15. marts 2025].

[110] ADSK, "Bygningsinformationsmodellering", 2002. [Online]. Available: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Dato for adresse: 15. marts 2025].

[111] A. Boyko, "BIM History Map", 2024. [Online]. Tilgængelig: https://bigdataconstruction.com/history-of-bim/.

[112] A. S. Borkowski, "Definitioner af BIM efter organisationer og standarder", 27. december 2023. [Online]. Tilgængelig: https://encyclopedia.pub/entry/53149. [Dato for adresse: 5. marts 2025].

[113] CAD Vendor, "OPEN BIM Programme," 2012. [Online]. Tilgængelig: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Referencedato: 30. marts 2025].

[114] Wikipedia, "Industry Foundation Classes". [Online]. Available: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Dato for adgang: 15. marts 2025].

[115] Wikipedia, "IGES," [Online]. Tilgængelig: https://en.wikipedia.org/wiki/IGES. [Referencedato: 30. marts 2025].

[116] A. Boyko, "Historien om CAD (BIM)," 15. december 2021. [Online]. Tilgængelig: https://miro.com/app/board/o9J_laML2cs=/. [Dato for adresse: 24. februar 2025].

[117] T. K. K. A. O. F. B. C. E. L. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Reference study of IFC software support: the GeoBIM benchmark 2019 - Part I," 8 Jan. 2021. [Online]. Tilgængelig: https://arxiv.org/pdf/2007.10951. [Dato for adgang: 5. marts 2025].

[118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM," 13. april 2021. [Online]. Tilgængelig: https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be. [Dato for adresse: 5. marts 2025].

[119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth", 17. december 2024. [Online]. Tilgængelig: https://www.realpha.com/blog/rets-importance-in-real-estate-explained. [Dato for adgang: 5. marts 2025].

[120] "Flex token cost," 2024. [Online]. Tilgængelig: https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex.

[121] A. Boyko, "Glem BIM og demokratiser adgangen til data (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024. [Online]. Tilgængelig: 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 og F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [Online]. Tilgængelig: https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle.

[123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle," 1. januar 2018. [Online]. Tilgængelig: https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html. [Dato for adgang: 2. marts 2025].

[124] Apple.com, "Pixar, Adobe, Apple og NVIDIA danner Alliance for OpenUSD for at fremme åbne standarder for 3D-indhold," 1. august 2023. [Online]. Tilgængelig: https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/. [Tilgået: 2. marts 2025].

[125] AECmag, "ADSK's granulære datastrategi", 25. juli 2024. [Online]. Available: https://aecmag.com/technology/autodesks-granular-data-strategy/. [Dato for adgang: 15. marts 2025].

[126] A. Boyko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts", 24.11.2024. [Online]. Tilgængelig: 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. [Dato for adresse: 23. februar 2025].

[127] A. Boyko, "ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction," 7. august 2024. [Online]. Tilgængelig: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [Tilgået: 15. marts 2025].

[128] "Jeffrey Zeldman præsenterer", 6. maj 2008. [Online]. Tilgængelig: https://zeldman.com/2008/05/06/content-precedes-design/. [Dato for adgang: 15. marts 2025].

[129] A. Boyko, "DWG Analyse with ChatGPT | DataDrivenConstruction," 5. marts 2024. [Online]. Tilgængelig: https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction. [Referencedato: 15. marts 2025].

[130] McKinsey , "The McKinsey guide to outcompeting in the age of digital and AI", 2023. [Online]. Tilgængelig: https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired. [Dato for adresse: 30. marts 2025].

[131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31. marts 2016. [Online]. Tilgængelig: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/. [Dato for adgang: 15. marts 2025].

[132] J. Bertin, "Grafik og grafisk informationsbehandling", 8. september 2011. [Online]. Tilgængelig: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Dato for adgang: 15. marts 2025].

[133] CauseWeb, "Wells/Wilks om statistisk tænkning", [Online]. Tilgængelig: https://www.causeweb.org/cause/resources/library/r1266. [Dato for adgang: 15. marts 2025].

[134] Ministrymagazine, "Hvordan videnskaben opdagede skabelsen", januar 1986. [Online]. Tilgængelig: https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation. [Referencedato: 15. marts 2025].

[135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23. maj 2017. [Online]. Tilgængelig: https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation. [Tilgået 15. maj 2024].

[136] "Sådan opbygger du en dataarkitektur, der driver innovation - i dag og i morgen", 3. juni 2020. [Online]. Tilgængelig: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow. [Dato for adresse: 15. marts 2025].

[137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630. [Dato for adresse: 15. marts 2025].

[138] "Convertors", 2024. [Online]. Tilgængelig: https://datadrivenconstruction.io/index.php/convertors/.

[139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1. januar 2017. [Online]. Tilgængelig: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf. [Dato for adgang: 18. februar 2025].

[140] "Rørledning i konstruktion", 2024. [Online]. Tilgængelig: https://datadrivenconstruction.io/index.php/pipeline-in-construction/.

[141] Wikipedia, "Apache NiFi," 1. januar 2025. [Online]. Tilgængelig: https://de.wikipedia.org/wiki/Apache_NiFi. [Dato for adgang: 5. marts 2025].

[142] n8n, "Gmail AI Auto-Responder: Create Draft Replies to incoming emails," 1. maj 2024. [Online]. Tilgængelig: https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-

incoming-emails/. [Tilgået 15. marts 2025].

[143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail," 1. marts 2025. [Online]. Tilgængelig: https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/. [Dato for adgang: 15. marts 2025].

[144] B. T. O'Neill, "Failure rates for analytics, AI, and big data projects = 85% - yikes!", 1. januar 2025. [Online]. Tilgængelig: https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/.

[145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.

[146] T. J. S. a. J. S. Jesse Perla, "A Problem that Stumped Milton Friedman", Quantitative Economics with Julia, 1. januar 2025. [Online]. Tilgængelig: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Dato for adgang: 1. maj 2024].

[147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.

[148] A. Boyko, "San Francisco. Bygge- og anlægssektoren 1980-2019," 2024. [Online]. Tilgængelig: https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019.

[149] A. Boyko, "Kaggle: RVT IFC Files 5000 Projects," 2024. [Online]. Tilgængelig: https://www.kaggle.com/datasets/artemboiko/rvtifc-projects.

[150] CFMA, "Forberedelse til fremtiden med Connected Construction," [Online]. Tilgængelig: https://cfma.org/articles/preparing-for-the-future-with-connected-construction. [Tilgået 15. marts 2025].

[151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing", 22. maj 2017. [Online]. Tilgængelig: 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] "Betingelser, der kræves for at bevare plantefossiler", 2024. [Online]. Tilgængelig: https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html.

[153] "BlackRocks Fink om obligationer, fusioner og opkøb, den amerikanske recession og valget: Fuldt interview," 2023. [Online]. Tilgængelig: https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video.

[154] cio, "12 berømte AI-katastrofer", 02. oktober 2024. [Online]. Tilgængelig: https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html. [Dato for adgang: 15. marts 2025].

[155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.

[156] Wikipedia, "Kaggle," 1. januar 2025. [Online]. Tilgængelig: https://en.wikipedia.org/wiki/Kaggle. [Dato for adgang: 15. marts 2025].

[157] Kaggle, "Titanic - Machine Learning from Disaster", 1. januar 2025. [Online]. Tilgængelig: https://www.kaggle.com/competitions/titanic/overview. [Tilgået 10. marts 2025].

[158] Ш. Johri, "Creating ChatGPT: From Data to Dialogue", 2024. [Online]. Tilgængelig: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/.

[159] П. Domingos, "Et par nyttige ting at vide om maskinlæring", 2024. [Online]. Tilgængelig: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

[160] J. Saramago, "Citerbart citat", [Online]. Available: https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered. [Dato for adgang: 17. marts 2025].

[161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models", 19. december 2024. [Online]. Tilgængelig: https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/. [Tilgået 25. marts 2025].

[162] "NVIDIA annoncerer en større udgivelse af Cosmos World Foundation Models og Physical AI Data Tools", 18. marts 2025. [Online]. Tilgængelig: https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools. [Dato for adgang: 25. marts 2025].

[163] NVIDIA, "NVIDIA Isaac Sim," [På internettet]. Tilgængelig: https://developer.nvidia.com/isaac/sim. [Tilgået 25. marts 2025].

[164] M. Quarterly, "Why digital strategies fail", 25. januar 2018. [Online]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail. [Dato for adgang: 15. marts 2025].

[165] M. J. Perry, "Mine foretrukne Milton Friedman-citater", 17. november 2006. [Online]. Tilgængelig: https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/. [Dato for adgang: 1. marts 2025].

[166] J. A. Wheeler, "Information, physics, quantum: the search for links", 1990.

[169] A. Boyko, "Lobbykrig og BIM-udvikling. Del 5: BlackRock er herre over alle teknologier. Hvordan virksomheder kontrollerer open source-kode," 2024. [Online]. Tilgængelig: https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d.

[170] T. Krijnen og J. Beetz, "A SPARQL query engine for binary-formatted IFC building models", *Advanced Engineering Informatics*, 2024.

[171] "Antal virksomheder i byggesektoren i Storbritannien i 2021, efter virksomhedsstørrelse," 2024. [Online]. Tilgængelig: https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/.

[172] "5.000 IFC&RVT-projekter," 2024. [Online]. Tilgængelig: https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io.

[173] M. Popova, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe", 2008. [Online]. Tilgængelig: https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/. [Referencedato: februar 2025].

[174] *Lobbykrig om data i byggeriet | Teknofeudalisme og historien om BIM's skjulte fortid.* [Film]. Tyskland: Artyom Boiko, 2023.

[175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Automatic retrieval of documents and data from projects," 16 nov. 2023. [Online]. Tilgængelig: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Dato for adgang: 2. marts 2025].

[176] M. &. Company, "Three new mandates for capturing a digital transformation's full value," 22 Jan. 2022. [Online]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value. [Dato for adgang: 15. februar 2025].

[177] KPMG, "Byggeri i en digital verden", 1. maj 2021. [Online]. Tilgængelig: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Dato for adgang: 5. april 2024].

[178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time," 17. marts 2023. [Online]. Tilgængelig: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Dato for adgang: 15. februar 2025].

[179] O. Business, "Satya Nadella afslører 'hvordan AI-agenter vil forstyrre SaaS-modeller'", 10. januar 2025. [Online]. Available: https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models. [Dato for adgang: 15. marts 2025].

[180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," 23. marts 2016. [Online]. Available: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/. [Dato for adgang: 15. marts 2025].

[181] UK Foreign, Commonwealth and Development Office, "Digital development strategy 2024 to 2030", 18. marts 2024. [Online]. Available: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030. [Dato for adgang: 15. marts 2025].

[182] "Vision og strategi i bygningsdesignbranchen", 7. november 2003. [Online]. Tilgængelig: https://web.archive.org/web/20030711125527/http:/usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112. [Dato for adresse: 5. marts 2025].

[183] M. Bocharov, "Informationsmodellering", marts 2025. [Online]. Tilgængelig: https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5. [Dato for adresse: 15. marts 2025].

[184] "Integrated Design-Through-Manufacturing: Benefits and Rationale", 2000. [Online]. Tilgængelig: https://web.archive.org/web/20010615093351/http:/www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Referencedato: 25. marts 2025].

[185] CAD Vendor, "Open BIM Programme is a marketing campaign," 12. marts 2012. [Online]. Tilgængelig: https://web.archive.org/web/20120827193840/http:/www.graphisoft.com/openbim/. [Dato for adresse: 30. marts 2025].

EMNEINDEKS

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477
AIA, 289
AIM, 3, 289
AMS, 14, 84, 151, 153
Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494
Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494
Apache ORC, 62, 330, 378
Apache Parket, 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
Big Data, 9, 67, 245, 383
BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 389, 413, 416, 448, 462, 466, 475, 476, 492, 494, 497
BlackBox, 240, 242, 243.
BMS, 8
Bokeh, 320, 337
BOM, 76, 77, 79, 257, 263
Afgrænsningsboks, 234, 373, 392, 393, 394, 414, 415, 480
BREP, 142, 234, 263, 264, 276, 283, 284

C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497
CAE, 16, 283
CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473
CAM, 16, 78
CAPEX, 14, 82
CDE, 84, 175, 388, 389, 390
ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421
Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443
CO_2, 72
CO_2, 229, 234, 235, 236, 237, 238, 239, 243
COBie, 156, 289, 292.
CoE, 56, 168, 169, 170, 477
Andenpilot, 114, 116, 459
CPIXML, 143, 272, 273, 276, 277, 279, 285, 296
CPM, 14, 17, 62, 166, 175, 233, 326, 473
CQMS, 14, 84, 177, 178, 462
CRM, 109, 369, 459.
CRUD, 51, 459
CSG, 263
CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 354, 356, 373, 377, 378, 379, 380, 401, 407, 414, 474, 480

D

DAE, 276, 277, 278, 280, 281, 284, 285, 414
DAG, 362, 363, 365, 366
Dash, 320, 336, 337.
Datastyring, 373, 395, 396, 398, 400, 401
Data Lake, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
Data Lakehouse, 67, 373, 386, 387, 388
Dataminimalisme, 373, 395, 396, 400, 401
Data Sump, 373, 395, 397, 401
Datavarehus, 382, 383, 400
Data-som-en-tjeneste, 487
datadrevet, 50, 170, 460, 461, 484, 486
DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475
DataOps, 170, 373, 398, 399, 400, 401, 475
DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
DGN, 8, 140, 186, 227, 357
DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390
DXF, 8, 73, 277

E

ECM, 58, 175
ECS, 142
EIR, 289
eLOD, 289
ELT, 384, 385
EPM, 14, 166, 198
ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
ESG, 196, 235, 236, 238
ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497
Uddrag, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Fjer, 62, 123, 330, 378
FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
GIS, 58
GLTF, 143, 278
Google Sheets, 368, 370
Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380
HiPPO, 29, 37, 95, 424, 477, 484, 490
HTML, 123, 340, 365, 370

I

IDS, 289, 290, 291.
IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497
IGES, 262, 263, 276.
iLOD, 289
IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484
ISO 19650, 388

J

JavaScript, 320, 369, 378
JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480
Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
k-NN, 392, 393, 442, 445, 446, 447, 448
KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494
LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 187, 219, 220, 220, 221, 222, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494
Belastning, 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
Low-Code, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
MCAD, 77, 78, 257, 284
MEP, 14, 175, 492
NET, 234, 283, 284, 285, 296, 462
Microsoft SQL, 65
Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
MRP, 11, 12
MS Project, 70
MySQL, 63, 64, 65, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372
NLP, 69, 394
Ingen kode, 368, 369
NURBS, 142, 282, 283, 284, 285
NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285
OCCT, 273
OCR, 69, 128, 132, 134
OmniClass, 154, 155, 156.
åben BIM, 142, 216, 219, 256, 261, 278, 291
Open Source, 45, 55, 97, 98, 107, 108, 273, 275, 494
UGLE, 267, 268, 269.

P

Pandaer, 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
Parket, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480
PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497
PDM, 16
PHP, 63, 339
Rørledning, 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
privat egenkapital, 464, 487
Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479
QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 381, 429, 443

R

RAG, 111, 116
RDBMS, 63, 64, 65, 82, 89.
RDF, 267, 268, 269
RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474
RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482
ROI, 311, 317, 319, 321, 351, 370, 389, 479
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
OMFANG, 273, 277
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
Seaborn, 123, 320, 336, 337, 408, 412
SPARQL, 269
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494
SQLite, 63, 64, 65, 166, 167, 296, 329, 363
TRIN, 261, 262, 263, 266, 272, 276, 277, 292
Streamlit, 336, 337
SVF, 142, 276, 285

T

Transform, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

U

Uniclass, 154, 155, 156.
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

V

VectorOps, 373, 398, 400, 401
VR, 84, 271, 285

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474