

2. IZDAJA  
POSODOBLJENA IN  
RAZŠIRJENA

# DATA DRIVEN CONSTRUCTION

KRMARJENJE PO PODATKOVNI DOBI  
V GRADBENI INDUSTRIJI

Z UPORABO PRIMEROV  
UMETNE INTELIGENCE IN LLM

Artem Boiko



# **DATA-DRIVEN CONSTRUCTION**

**Navigacija v podatkovni dobi  
v gradbeništvu**

Druga, popravljena in dopolnjena izdaja

ARTEM BOIKO

“

*"Boiko je James Carville na področju informacijske tehnologije - v njegovi večkrat citirani izjavi "It's the economy, stupid" je treba zamenjati le eno besedo za to slavno knjigo. "Gre za podatke, neu-mno." (Za iskanje poti v podatkovnem vesolju še danes velja rek starih Rimljjanov, ki izvira iz grščine: "Navigare necesse est". Avtor svoje bralce vodi skozi vse globine in plitvine podatkovnega oceana z zanesljivo roko in neomajnjim kompasom, pri čemer ne smemo pozabiti na izčrpen zgodovinski pristop in nenazadnje na zelo izvirne grafike ter dober smisel za humor, ki ni viden le na drugi pogled. Mednarodni odziv na Bojkovo knjigo sega od evforičnega odobravanja do precej žolčne skepse, kar je drugi nemški izdaji knjige še kako koristilo. Boiko je izviren in nedogmatičen mislec podatkov. Bralcu predstavi vznemirljiva spoznanja in vedno po-gumne, celo provokativne teze, ki spodbujajo k nadaljnemu raz-mišljanju. Odlično zdravilo za nemško bolezen latentnega konsen-zualizma. Mimogrede, zgornji latinski pregovor ima dopolnilo: "vi-vere non est necesse". Za Bojkov pristop k svetu podatkov ne velja - podatki živijo in njihovo življenje je nujno, da ne rečem ključno."*

- Dr. Burkhard Talebitari, samostojni urednik - tudi za revijo:  
BIM, ki jo od leta 2013 vsako leto izdaja družba Ernst & Sohn.

**"Knjiga Artema Bojka je mejnik za demokratizacijo digitalizacije v gradbeništvu - in pravi preobrat za mala in srednje velika podjetja (MSP). Še posebej prelomna: z uporabo sodobnih odprt-kodnih nizkokodnih in nekodnih orodij lahko podjetja že zdaj učinkovito vključujejo podatke v svoje poslovne procese in jih dobičkonosno analizirajo - brez poglobljenega programerskega znanja. S tem postane draga uporaba okornih komercialnih programskih paketov nepotrebna. Ta knjiga je poziv k ukrepanju! Je dragocen vodnik za vse, ki ne želijo le razumeti digitalne preobrazbe v gradbeništvu, temveč jo želijo tudi dejavno oblikovati - pragmatično, učinkovito in v prihodnost usmerjeno. Zdaj je čas, da si skupaj izmenjamo to znanje in trajnostno povečamo produktivnost gradbene industrije."**

**- Dr. Michael Max Buehler**, profesor gradbenega menedžmenta na HTWG Konstanz, solastnik podjetja Gemein-Werk Ventures in neodvisni direktor podjetja DevStream.

**"Knjiga DataDrivenConstruction je eden prvih korakov čez meje običajnega sveta gradbenikov z njihovimi zapletenimi sistemi načrtovanja in upravljanja, ko, kot se zdi, zaplenost in zasičenost s podatki niti ne daje možnosti za korenito poenostavitev in večjo preglednost dela z gradbenimi podatki. Artem v svoji knjigi v preprostem jeziku pokaže, kakšne možnosti pred nami odpirajo sodobne tehnologije dela s podatki, in dobesedno poda konkretno korake, ki jih lahko takoj uporabite pri svojem delu. Pozivam vse, ki želijo razumeti, kam bodo šli sistemi avtomatizacije v gradbeništvu, naj natančno preučijo to knjigo, da bi spoznali, da podatkovna revolucija v gradbeništvu že trka na naša vrata. Zdaj je zanimiva le za geeke, čez nekaj let pa bodo, tako kot BIM, takšni pristopi in programska oprema vseprisotni!"**

**- Ihor Rogachew**, vodja kompetenčnega centra IMT, BIM in digitalne preobrazbe pri podjetju RGD ter ustanovitelj podjetja InfraBIM.Pro.

**"Toplo priporočam knjigo DataDrivenConstruction, ki obravnava, kot pravi naslov, pristop k upravljanju informacij na podlagi podatkov za podjetje AECO. Trenutno jo uporabljam za pomoč pri začetku številnih razprav z različnimi skupinami. Ugotovil sem, da je zelo dostopna referenca. Poleg temeljitega pregleda zgodovinskega konteksta orodij v skupini AECO, podatkov in predstavitev več ključnih tehnologij knjiga vsebuje številne zelo uporabne diagrame, ki opisujejo obseg podatkovnih virov in artefaktov končnega uporabnika z vzorčnimi delovnimi tokovi. Zdi se mi, da so to tiste vrste diagramov, ki jih potrebujemo več pri razvoju in spremljanju informacijskih strategij in prispevajo k BEP-om - opredeljujejo splošni podatkovni model podjetja, na katerega se lahko prekrije meja za PIM in AIM."**

**- Paul Ransley**, glavni svetovalec pri podjetju Acmena in inženir za integracijo sistemov pri podjetju Transport for London.

**"Če so "podatki nova nafta", se jih moramo naučiti opredeliti, poiskati, izkopati in izboljšati, da bodo dragoceni. Knjiga Data-DrivenConstruction se mi je zdela zelo informativna in pronicljiva. Knjiga vsebuje koristno zgodovinsko ozadje in v preprostem jeziku pojasnjuje delo s podatki. Tistem, ki jih zanima digitalna preobrazba, omogoča dobro razumevanje podatkov - kako delujejo, kako so strukturirani in kako jih je mogoče uporabiti."**

**- Ralph Montague**, direktor podjetja ArcDox, direktor vrha koordinatorjev BIM in predsednik nacionalnega zrcalnega odbora BIM pri irskem nacionalnem organu za standarde.

**"Kot je bilo poudarjeno v knjigi, so informacije ključna dobrina za gradbeni sektor, njihova dostopnost v dostopnih oblikah pa močno olajša sprejemanje natančnih odločitev in pospešuje časovni potek projektov. Knjiga ponuja neutralen in učinkovit pristop za dostop do tega vira in njegovo izkorisčanje pri odločjanju. Metodologija, predstavljena v knjigi, izkorišča sodoben pristop, ki združuje programiranje na podlagi umetne inteligenčne z dostopnimi odprt-kodnimi orodji. Z izkorisčanjem moči umetne inteligence in uporabo odprt-kodne programske opreme je cilj metodologije povečati avtomatizacijo, optimizirati procese ter spodbujati dostopnost in sodelovanje na tem področju. Jezik knjige je jasen in enostaven za razumevanje."**

**- Dr. Salih Ofluoğlu**, dekan Fakultete za likovno umetnost in arhitekturo na Univerzi Antalya Bilim in organizator Evrazijskega foruma BIM.

**"Vse, kar lahko rečem, je: WOW! Način, kako ste vključili zgodovino, LLM, grafike in splošno enostavnost razumevanja vaših točk, je resnično izjemен. Potek knjige je neverjeten. Ta knjiga ima toliko briljantnih vidikov; resnično spreminja pravila igre. Je odličen vir informacij in pohvalim vas za trud in strast, ki ste ju vložili vanjo. Čestitam vam za tako izjemno delo. Lahko bi še nadaljeval, a dovolj je, če rečem, da sem neverjetno navdušen!"**

**- Natasha Prinsloo**, vodja digitalne prakse v podjetju energylab\_

**"Ta knjiga je za vse, ki se ukvarjajo z gradbeništvom, od novincev do izkušenih profesionalcev, prava popotnica! Ni tipično zaprašeno branje - polna je vpogledov, strategij in kančka humorja, ki vas bo držal v napetosti. Knjiga zajema razvoj uporabe podatkov v gradbeništvu, od starodavnih metod zapisovanja podatkov do najsodobnejših digitalnih tehnologij. To je kot vožnja s časovnim strojem skozi razvoj gradbenih podatkov. Ne glede na to, ali ste arhitekt, inženir, vodja projekta ali podatkovni analitik, bo ta izčrpen vodnik korenito spremenil vaš način pristopa k projektom. Pripravite se na optimizacijo procesov, izboljšanje odločanja in upravljanje projektov kot še nikoli prej!"**

**- Pierpaolo Vergati**, predavatelj na Univerzi Sapienza v Rimu in višji vodja gradbenih projektov pri podjetju Fintecna.

*"Knjigo sem prebral v enem dihu, v manj kot šestih urah. Kakovost izdelave knjige je odlična, gost sijajni papir, barvne sheme, prijetna pisava. Veliko število praktičnih primerov o delu z LLM, značilnih za gradbeništvo, vam bo prihranilo mesece, če ne celo leta samostojnega učenja. Primeri dela so zelo raznoliki, od preprostih do zapletenih, ne da bi vam bilo treba kupiti zapleteno in drago programsko opremo. Knjiga bo lastnikom vseh podjetij v gradbeništvu omogočila nov pogled na njihovo poslovno strategijo, digitalizacijo in razvojne možnosti. Manjšim podjetjem pa, da s cenovno dostopnimi in brezplačnimi orodji prevečajo učinkovitost."*

- **Mikhail Kosarev**, predavatelj in svetovalec za digitalno preobrazbo v gradbeništvu pri TIM-ASG.

*"Knjiga "DATA DRIVEN CONSTRUCTION" spreminja pravila igre za vse, ki jih zanima, kam gre gradbena industrija v dobi podatkov. Artem ne opiše le površja, temveč se poglobi v trenutni razvoj, izzive in obetavne priložnosti v gradbeništvu. To knjigo odlikuje njena dostopnost - Artem razлага zapletene zamisli s primerljivimi analogijami, zaradi katerih je vsebina enostavno razumljiva. Knjiga se mi je zdela izjemno informativna in hkrati zanimiva. Skratka, Artem je ustvaril dragocen vir, ki ne le obvešča, temveč tudi navdihuje. Ne glede na to, ali ste izkušen strokovnjak ali novinec v gradbeništvu, bo ta knjiga razširila vaš pogled in poglobila vaše razumevanje tega, kam gre ta panoga. Zelo priporočljivo!"*

- **Moayad Saleh**, arhitekt in vodja izvajanja BIM pri podjetju TMM GROUP Gesamtplanungs GmbH.

*"Naj povem, da je knjiga Data-Driven Construction vredna poučevanja kot učbenik na univerzah in je knjiga, ki bo dragoceno prispevala k razvijajočemu se področju BIM. Knjiga Data-Driven Construction vsebuje tehnični slovarček, ki zelo dobro pojasnjuje koncepte. Teme, ki jih je izjemno težko razložiti, so poenostavljene in razumljive z zelo lepim vizualnim jezikom. Menim, da je treba bralcu, četudi na kratko, izraziti tisto, kar naj bi bilo razloženo z vizualnimi prikazi. Razumljivost nekaterih vizualizacij, z drugimi besedami, branje vizualizacije zahteva ločene informacije. Rad bi še povedal, da dragoceno delo Artema Bojka z veseljem predstavljam na svojih predavanjih in seminarjih na univerzah."*

- **Dr. Ediz Yazicioglu**, lastnik podjetja ArchCube in predavatelj vodenja gradbenih projektov na oddelku za arhitekturo na Tehnični univerzi v Istanbulu in na univerzi Medipol.

*"Podatkovno vodena gradnja nazorno predstavi osnove informacijsko podprtega dela z gradbenimi podatki. Knjiga, ki se ukvarja z informacijskimi tokovi in temeljnimi ekonomskimi koncepti, se tako razlikuje od drugih knjig o BIM, saj ne predstavlja le pogleda proizvajalca programske opreme, temveč poskuša posredovati tudi temeljne koncepte. Knjiga, ki jo je vredno prebrati in videti."*

- **Jakob Hirn**, generalni direktor in soustanovitelj podjetja Build Informed GmbH ter pobudnik foruma inovacij "On Top With BIM".

*"Podatki so nova nafta", kot pravijo, zato morajo imeti iskalci ali rudarji prava orodja in način razmišljanja za pridobivanje vrednosti iz tega vira 21. stoletja. Gradbena industrija je predolgo na spolzkem terenu procesov, ki temeljijo na "3D informacijah", pri čemer izvedba projekta temelji na pečenih informacijah nekoga drugega (npr. že so narisali krožni ali stolpčni diagram), medtem ko lahko osnovni "podatki" (npr. neobdelana preglednica) prinesejo veliko več, zlasti ker zlitje več podatkov in umetna inteligence prinašata neomejene potenciale. Če izvajate (ali poučujete/raziskujete) gradbeništvo, je ta knjiga vaš najboljši - in zaenkrat edini - vir za krmarjenje po svetu, ki ga poganjajo podatki in v katerem smo se znašli."*

- **Dr. Zulfikar Adamu**, izredni profesor za strateško informatiko v gradbeništvu na LSBU, Velika Britanija.

*"Data-Driven Construction" avtorja Artyoma Boika je impresivno delo, ki ponuja trdne temelje za gradbeno industrijo v času ne-nehno rastočih tehnologij in informacijskih možnosti. Boiku je uspelo kompleksne teme predstaviti na razumljiv način, hkrati pa predstaviti vizionarske zamisli. Knjiga je dobro premišljen zbornik, ki ne izpostavlja le sedanjega razvoja, temveč ponuja tudi pogled na prihodnje inovacije. Priporočamo jo vsem, ki se želijo spoprijeti z načrtovanjem in izvajanjem gradnje na podlagi podatkov."*

- **Markus Eiberger**, predavatelj na Univerzi uporabnih znanosti v Stuttgartu, višji vodja projektov in namestnik vodje podružnice pri podjetju Konstruktionsgruppe Bauen, član odборa združenja BIM Cluster Baden-Württemberg.



Druga izdaja, marec 2025.  
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-18-8



Artem Boiko Avtorske pravice

[boikoartem@gmail.com](mailto:boikoartem@gmail.com)  
[info@datadrivenconstruction.io](mailto:info@datadrivenconstruction.io)

Nobenega dela te knjige ni dovoljeno razmnoževati ali prenašati v kakršni koli obliki ali na kakršen koli način, elektronski ali mehanski, vključno s fotokopiranjem, snemanjem ali sistemom za shranjevanje in iskanje informacij, brez pisnega dovoljenja avtorja - razen za nekomercialno distribucijo v nespremenjeni obliki. Knjiga se distribuira brezplačno in se lahko prosto deli z drugimi uporabniki v osebne, izobraževalne ali raziskovalne namene, pod pogojem, da se ohrani avtorstvo in sklicevanje na izvirnik. Avtor obdrži vse nelastniške pravice do besedila in ne daje nobenih izrecnih ali implicitnih jamstev. Podjetja, izdelki in imena, omenjena v knjigi, so lahko izmišljena ali uporabljenata kot primeri. Avtor ni odgovoren za kakršne koli posledice uporabe navedenih informacij. Informacije v knjigi so na voljo "takšne, kot so", brez jamstva za popolnost ali ustreznost. Avtor ni odgovoren za naključno ali posledično škodo, ki bi nastala zaradi uporabe informacij, kode ali programov iz te knjige. Vzorci kode, ki so predstavljeni v tej knjigi, so namenjeni le za izobraževalne namene. Bralci jih uporabljajo na lastno odgovornost. Avtor priporoča, da se vse programske rešitve pred uporabo v produkcijskem okolju preizkusijo. Vse blagovne znamke in imena izdelkov, omenjena v besedilu, so blagovne znamke, registrirane blagovne znamke ali storitvene znamke zadevnih podjetij in so last njihovih lastnikov. Uporaba teh imen v knjigi ne pomeni kakršne koli povezave z njihovimi lastniki ali njihove podpore. Omembu izdelkov ali storitev tretjih oseb ne pomeni priporočila ali podpore. Imena podjetij in izdelkov, uporabljeni v primerih, so lahko blagovne znamke njihovih lastnikov. Povezave na spletna mesta tretjih oseb so na voljo zaradi priročnosti in ne pomenijo, da avtor podpira informacije na teh spletnih mestih. Vsi navedeni statistični podatki, citati in študije so bili aktualni v času pisanja. Podatki se lahko sčasoma spremeni.

Ta knjiga je distribuirana pod licenco Creative Commons Priznanje avtorstva - nekomercialno - brez avtorskih pravic 4.0 International (CC BY-NC-ND 4.0). Lahko jo kopirate in razširjate v nekomercialne namene, če ohranite avtorstvo in ne vnašate sprememb.



CC BY-NC-ND 4.0

© 2024 Artyom Boiko. Prva izdaja.  
© 2025 Artyom Boiko. Druga, popravljena in dopolnjena izdaja.  
Vse pravice pridržane.

## PREDGOVOR K DRUGI IZDAJI

Ta knjiga je rezultat živahnega dialoga s strokovno skupnostjo. Temelji na številnih strokovnih razpravah o upravljanju podatkov v gradbeništvu, ki so potekale na različnih strokovnih platformah in družbenih medijih. Te razprave so postale podlaga za članke, publikacije in vizualno gradivo, ki so v strokovni skupnosti sprožili širok odziv. Avtorjeve vsebine vsako leto na različnih platformah in v različnih jezikih pritegnejo na milijone ogledov ter združujejo strokovnjake s področja digitalizacije gradbeništva.

V enem letu po objavi prve izdaje so knjigo naročili strokovnjaki iz več kot 50 držav, od Brazilije in Peruja do Mauritiusa in Japonske. Druga izdaja knjige, ki jo zdaj držite v rokah, je bila na podlagi povratnih informacij strokovnjakov, kritik prve izdaje in razprav v strokovnih krogih pregledana in razširjena. Zaradi povratnih informacij je bila druga izdaja bistveno razširjena: dodani sta bili novi poglavji o tehnologijah CAD (BIM) in oblikovanju učinkovitih ETL -procesov. Bistveno se je povečalo tudi število praktičnih primerov in študij primerov. Posebno dragocene so povratne informacije vodilnih v gradbeništvu, svetovalnih podjetjih in večjih IT-podjetjih, ki so se na avtorja obrnili z vprašanji o digitalizaciji in interoperabilnosti tako pred objavo prve različice knjige kot po njej. Mnogi med njimi so že uporabili pristope, opisane v knjigi, ali pa to nameravajo storiti v bližnji prihodnosti.

V rokah držite knjigo, ki je nastala z razpravo in aktivno izmenjavo mnenj. Napredek se rojeva v dialogu, spopadu mnenj in odprtosti za nove pristope. Hvala, ker ste del tega dialoga. Vaša konstruktivna kritika je osnova za prihodnje izboljšave. Če v besedilu najdete napake ali želite deliti ideje in predloge, so dobrodoše vse povratne informacije. Kontaktni podatki so navedeni na koncu knjige.

## ZAKAJ JE KNJIGA BREZPLAČNA?

Ta knjiga je bila zasnovana kot odprt izobraževalni vir, namenjen širjenju sodobnih pristopov k upravljanju podatkov v gradbeništvu. Prva različica knjige je služila kot podlaga za zbiranje pripomb in predlogov strokovne skupnosti, ki so omogočili izboljšanje strukture in vsebine gradiva. Vse pripombe, predloge in zamisli smo skrbno analizirali in jih vključili v to revidirano različico. Cilj knjige je pomagati gradbenim strokovnjakom razumeti pomen dela s podatki: sistematično, zavestno in z misljivo na dolgoročno vrednost informacij. Avtor je zbral primere, ilustracije in praktična opažanja iz več kot desetletnega dela na področju digitalizacije gradbeništva. Večina tega gradiva je nastala na podlagi resničnih projektov, pogоворov z inženirji in gradbeniki, sodelovanja v mednarodnih pobudah in izobraževalnih seminarjih. Knjiga je poskus strukturiranja zbranih izkušenj in njihove delitve na dostopen način. Če želite podpreti nadaljnje širjenje idej iz knjige in pridobiti priročno obliko za branje, delo s primeri in vizualnim gradivom - lahko kupite [tiskano različico](#).

## PRAVICE UPORABE

Vsa gradiva, ilustracije in deli te knjige se lahko razmnožujejo, citirajo ali uporabljajo v kateri koli obliki in na katerem koli mediju, če je naveden vir: avtorstvo Artema Bojka in naslov knjige "Data-Dri-ven Construction". Zahvaljujemo se vam za spoštovanje dela in širjenje znanja.

To knjigo z iskreno hvaležnostjo posvečam svoji družini, ki mi je že od malih nog vcepila globoko ljubezen do gradbeništva, domačemu rudarskemu mestu za lekcije odpornosti in svoji ženi geodetki, ki mi je bila v nenehno oporo in navdih.

## **ZA KOGA JE TA KNJIGA NAMENJENA**

Knjiga je napisana v dostopnem jeziku in je namenjena širokemu krogu bralcev v gradbeništvu - od študentov in začetnikov, ki želijo razumeti osnove sodobnih gradbenih procesov, do strokovnjakov, ki potrebujejo sodobno metodologijo za upravljanje podatkov v gradbeništvu. Ne glede na to, ali ste arhitekt, inženir, mojster, vodja gradnje ali podatkovni analistik, vam ta obsežen priročnik s številnimi edinstvenimi ilustracijami in grafikami ponuja dragocen vpogled v uporabo podatkov v poslovanju za optimizacijo in avtomatizacijo procesov, izboljšanje odločanja in upravljanje gradbenih projektov na različnih ravneh z uporabo sodobnih orodij.

Knjiga je celovit vodnik, ki združuje teoretične osnove in praktična pripomočka za vključevanje tehnik upravljanja podatkov v gradbene procese. Knjiga se osredotoča na strateško uporabo informacij za optimizacijo poslovanja, avtomatizacijo procesov, izboljšanje odločanja in učinkovito upravljanje projektov z uporabo sodobnih digitalnih orodij.

Knjiga obravnava teoretične in praktične vidike dela z informacijami v gradbeništvu. S podrobnnimi primeri raziskuje metodologijo parametriziranja nalog, zbiranja zahtev, obdelave nestrukturiranih in večplastnih podatkov ter njihovega preoblikovanja v učinkovite rešitve za gradbena podjetja.

Bralec postopoma prehodi pot od oblikovanja zahtev in razvoja osnovnih podatkovnih modelov do kompleksnejših procesov integracije heterogenih informacijskih virov, oblikovanja procesov ETL, gradnje informacijskih cevovodov in modelov strojnega učenja. Zaporedni pristop omogoča jasen prikaz mehanizmov organizacije in avtomatizacije poslovnih procesov ter sistemov za podporo odločanju v gradbeništvu. Vsak del knjige se zaključi s praktičnim poglavjem, ki vsebuje navodila korak za korakom, ki omogočajo takojšnjo uporabo pridobljenega znanja v resničnih projektih.

## **POVZETEK DELOV KNJIGE**

Knjiga je zasnovana na konceptu preoblikovanja podatkov v vrednostni verigi: od zbiranja in zagotavljanja kakovosti podatkov do analitične obdelave in pridobivanja dragocenih praktičnih rešitev z uporabo sodobnih orodij in metodologij.

**1. del: Digitalna evolucija v gradbeništvu** - spremišča zgodovinsko preobrazbo upravljanja podatkov od glinenih ploščic do sodobnih digitalnih sistemov, analizira pojav modularnih sistemov in vse večji pomen informacijske digitalizacije v okviru industrijskih revolucij.

**2. del: Informacijski izzivi v gradbeništvu** - obravnava težave razdrobljenosti podatkov, "informacijske silose", vpliv pristopa HiPPO sprejemanje odločitev in omejitve lastniških formatov ter predlaga razmislek o prehodu na ekosisteme umetne inteligence in LLM.

**Del 3: Sistematisacija podatkov v gradbeništvu** - oblikuje tipologijo gradbenih podatkov, opisuje metode njihove organizacije, integracije s korporativnimi sistemi in obravnava vzpostavitev kompetenčnih centrov za standardizacijo informacijskih procesov.

**4. del: Zagotavljanje kakovosti podatkov** - razkriva metodologije za spremiščanje različnih informacij v kakovostne, strukturirane podatke, vključno s pridobivanjem podatkov iz različnih virov, potrjevanjem in modeliranjem z uporabo LLM.

**5. del: Izračuni stroškov in časa** - obravnava digitalizacijo izračunov stroškov in načrtovanja, avtomatizacijo pridobivanja količin iz modelov CAD (BIM), tehnologije modeliranja 4D-8D in izračun ESG gradbenih projektov.

**6. del: CAD in BIM** - kritično analizira razvoj tehnologij za projektiranje, vprašanja interoperabilnosti sistemov, trende v smeri odprtih podatkovnih formatov in možnosti za uporabo umetne inteligence pri projektiranju.

**7. del: Analiza podatkov in avtomatizacija** - obravnava načela vizualizacije informacij, ključne kazalnike uspešnosti, postopke ETL, orodja za orkestracijo delovnih tokov in uporabo jezikovnih modelov za avtomatizacijo rutinskih opravil.

**8. del: Shranjevanje in upravljanje podatkov** - obravnava formate za shranjevanje podatkov, koncepte podatkovnega skladišča in podatkovnega jezera, načela upravljanja podatkov in nove pristope, vključno z vektorskimi podatkovnimi zbirkami ter metodologijama DataOps in VectorOps.

**9. del: Veliki podatki in strojno učenje** - osredotoča se na prehod na objektivno analizo na podlagi preteklih podatkov, internet stvari na gradbiščih in uporabo algoritmov strojnega učenja za napovedovanje stroškov in časovnih rokov projektov.

**10. del: Gradbeništvo v dobi digitalnih podatkov** - predstavlja pogled na prihodnost gradbeništva, analizira prehod z vzročno-posledične analize na delo s korelacijami, koncept "uberizacije" gradbeništva in strategije za digitalno preobrazbo.

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



## UVOD

Kako dolgo lahko vaše podjetje ostane konkurenčno v svetu, v katerem se tehnologija hitro razvija in kjer se z modeli strojnega učenja avtomatizirajo vsi vidiki poslovanja, od časovnega načrtovanja in izračunavanja stroškov do analize tveganja?

Gradbena industrija, ki obstaja tako dolgo kot človeštvo samo, je na pragu revolucionarne spremembe, ki bo popolnoma spremenila naš pogled na tradicionalno gradbeništvo. Že v drugih gospodarskih panogah digitalizacija ne spreminja le pravil, temveč s trga neusmiljeno izriva podjetja, ki se niso uspela prilagoditi novemu okolju obdelave podatkov in niso sposobna izboljšati hitrosti sprejemanja odločitev (slika 1).



**Slika 1** Hitrost odločanja v gradbeništvu je pogosteje kot v drugih panogah odvisna od človeškega dejavnika.

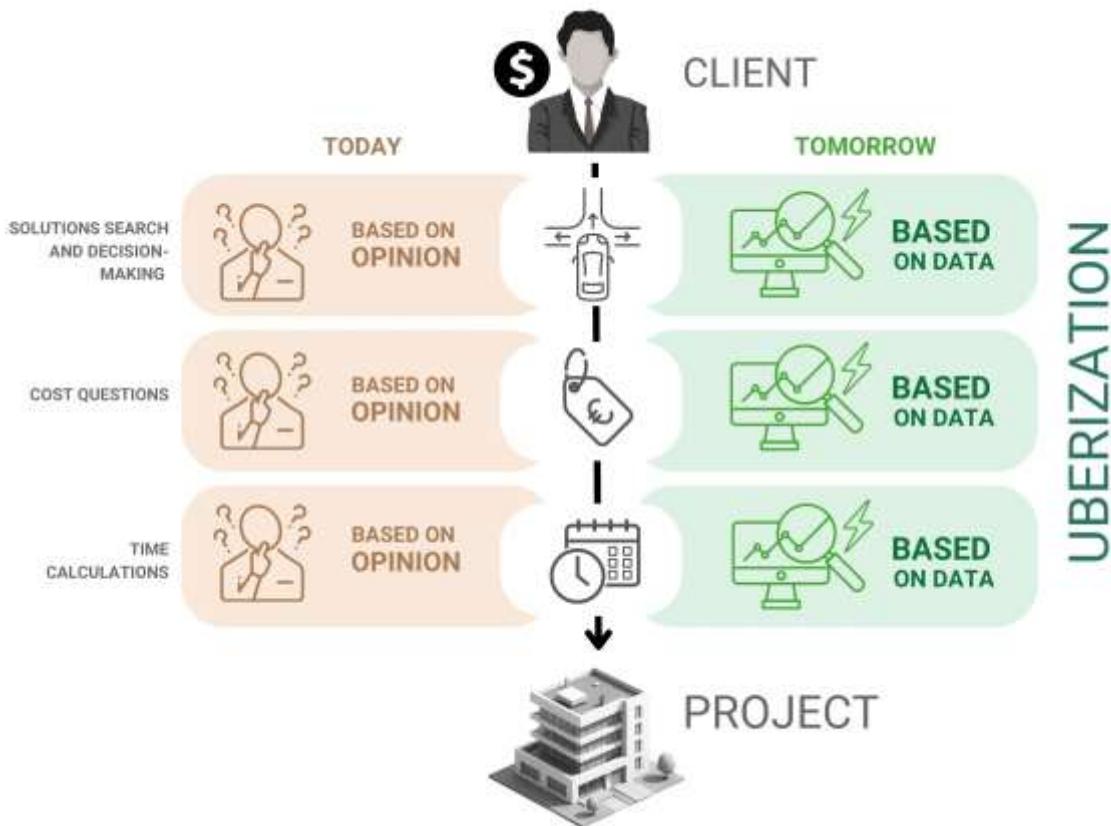
Bančni, maloprodajni, logistični in agroživilski sektor se hitro približujejo popolni digitalizaciji, kjer netočnosti in subjektivna mnenja nimajo več mesta. Sodobni algoritmi so sposobni analizirati ogromne količine podatkov in strankam zagotoviti natančne napovedi - pa naj gre za verjetnost vračila posojila, optimalne dobavne poti ali napovedovanje tveganj.

Gradbeništvo je ena od zadnjih panog, ki neizogibno prehaja od rešitev, ki temeljijo na mnenjih visoko plačanih strokovnjakov, k rešitvam, ki temeljijo na podatkih. Tega prehoda ne poganjajo le nove tehnološke zmogljivosti, temveč tudi večje zahteve trga in strank po preglednosti, natančnosti in hitrosti.

Robotizacija, avtomatizacija procesov, odprti podatki in na njih temelječe napovedi niso več le možnosti, temveč neizogibnost. Večina podjetij v gradbeništvu, ki so bila pred kratkim odgovorna naročniku za izračun obsega, stroškov, časa projektov in nadzor kakovosti, zdaj tvega, da se bodo

spremenila v zgolj izvajalce naročil in ne bodo sprejemala ključnih odločitev (slika 2).

Z razvojem računalniške moči, algoritmov strojnega učenja in demokratizacijo dostopa do podatkov je zdaj mogoče samodejno združevati podatke iz različnih virov, kar omogoča poglobljeno analizo procesov, napovedovanje tveganj in optimizacijo stroškov v razpravnih fazah gradbenega projekta. Te tehnologije lahko korenito izboljšajo učinkovitost in zmanjšajo stroške v celotnem sektorju.



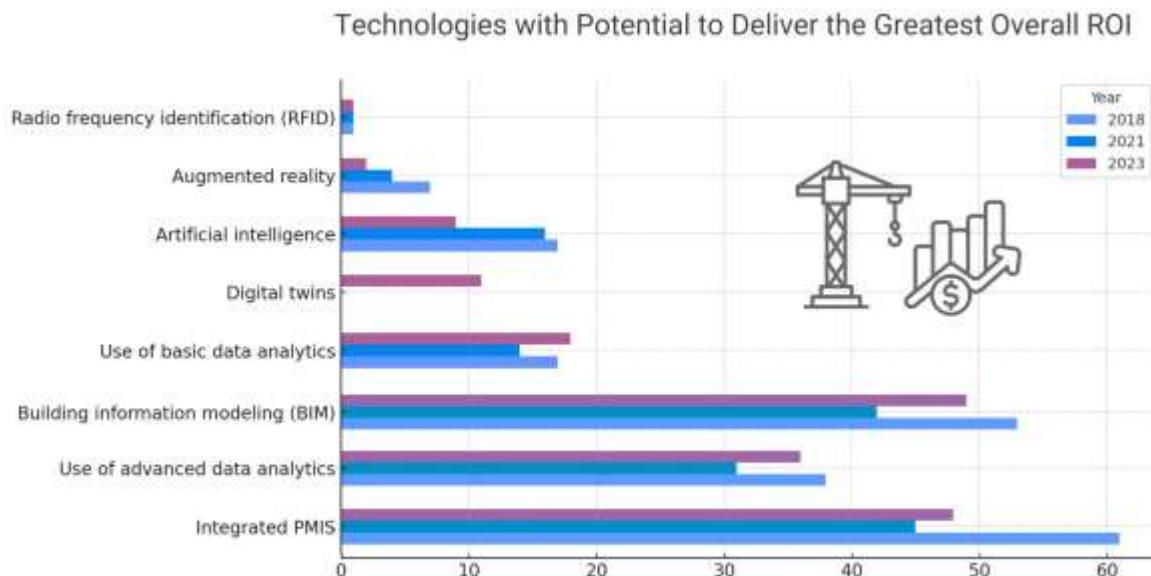
**Slika 2 Naročnika na poti do uresničitve njegovega projekta ne zanima pretiran človeški dejavnik.**

Klub vsem prednostim, ki jih prinašajo nova orodja in koncepti, gradbeništvo pri uvajanju novih tehnologij močno zaostaja za drugimi gospodarskimi panogami.

Po podatkih poročila IT Metrics Key Data 2017 je gradbeništvo med 19 drugimi panogami na zadnjem mestu po izdatkih za IT [1].

Hitra rast količine podatkov in zapletenosti procesov postaja glavobol za vodstvo podjetja, glavna težava pri uporabi novih tehnologij pa je, da so podatki kljub svojemu obilju razdrobljeni, nestrukturirani in pogosto nezdružljivi med različnimi sistemi in programskimi izdelki. Zato se številna podjetja v gradbenem sektorju zdaj ukvarjajo predvsem s težavami s kakovostjo podatkov, ki jih je mogoče rešiti le z uvedbo učinkovitih, avtomatiziranih sistemov za upravljanje in analitiko.

Glede na raziskavo, ki jo je leta 2023 med gradbenimi menedžerji izvedla družba KPMG® [2], imajo informacijski sistemi za vodenje projektov (PMIS), napredna in osnovna podatkovna analitika ter informacijsko modeliranje stavb (BIM) največji potencial za izboljšanje ROI projektov (slika 3).



Slika 3 Anketa med vodstvenimi delavci gradbenih podjetij: katere tehnologije bodo zagotovile najvišjo donosnost naložb (ROI) v investicijske projekte? (na podlagi gradiva [2]).

Rešitev izzivov, povezanih z vključevanjem podatkov v poslovne procese, je zagotavljanje visokokakovostnih informacij, uporaba ustreznih oblik podatkov ter uporaba učinkovitih metod za ustvarjanje, shranjevanje, analiziranje in obdelavo podatkov.

Zavedanje vrednosti podatkov sili različne panoge, da opustijo ločene aplikacije in zapletene birokratske strukture upravljanja. Namesto tega se osredotočajo na oblikovanje novih pristopov k informacijski arhitekturi, s čimer se podjetja spreminjajo v sodobna podjetja, ki temeljijo na podatkih. Prej ali slej bo ta korak naredila tudi gradbena industrija, ki bo iz postopne digitalne evolucije prešla v pravo digitalno revolucijo, ki bo prizadela vsa podjetja.

Prehod na poslovne procese, ki temeljijo na podatkih, ne bo enostaven. Mnoga podjetja se bodo soočala z izzivi, saj vodstveni delavci ne razumejo vedno, kako uporabiti kaotične nabore podatkov za izboljšanje učinkovitosti in poslovne rasti.

Ta knjiga se poglavlja v svet podatkov, kjer informacije postajajo ključni strateški vir, ki določa učinkovitost in trajnost poslovnih procesov. S hitro rastjo informacij se podjetja soočajo z novimi izzivi. Digitalna preobrazba ni več le modna reč - postaja nujnost.



**Slika 4** Podatki in procesi so temelj gradnje.

Razumeti transformacijo pomeni znati razložiti zapleteno s preprostimi besedami. Zato je knjiga napisana v dostopnem jeziku in opremljena z avtorjevimi ilustracijami, ustvarjenimi posebej za jasno razlago ključnih pojmov. Ti diagrami, sheme in vizualizacije so zasnovani tako, da odpravljajo ovire pri dojemanju in naredijo gradivo razumljivo tudi tistim, ki so se jih prej takšne teme zdele preveč zapletene. Vse ilustracije, diagrame in grafikone v tej knjigi je ustvaril avtor in so oblikovani posebej za vizualizacijo ključnih pojmov, opisanih v besedilu.

*Ena slika je vredna tisoč besed [3].*

- Fred R. Barnard, angleški ilustrator, 1927.

Da bi teorijo povezali s prakso, bomo uporabili orodja umetne inteligenčne (zlasti jezikovne) modele, ki vam omogočajo razvoj rešitev brez poglobljenega znanja programiranja. Če ste usmerjeni v praktično gradivo in vas bolj zanima praktično delo s podatki, lahko preskočite prvi uvodni del in preidete na ravnost na drugi del knjige, kjer se začne opisovanje konkretnih primerov in primerov.

Vendar ne pričakujte preveč od orodij umetne inteligenčne, strojnega učenja in LLM (Large Language Models) na splošno. Brez kakovostnih vhodnih podatkov in poglobljenega razumevanja predmeta tudi najnaprednejši algoritmi ne morejo zagotoviti zanesljivih in smiselnih rezultatov.

Microsoftov izvršni direktor Satya Nadella je v začetku leta 2025 opozoril na nevarnost balona na področju umetne inteligenčne [4] in primerjal trenutni razmah z balonom dot-com. Poudarja, da so trditve o doseganju mejnikov AGI (Artificial General Intelligence) brez ustrezne utemeljitve "nesmiselno manipuliranje z metrikami". Nadella meni, da je treba resnični uspeh umetne inteligenčne meriti po njenem prispevku k rasti svetovnega BDP in ne po pretiranem poudarjanju modnih besed.

Za vsemi besedami o novih tehnologijah in konceptih se skriva zapleteno in naporno delo za zagotavljanje kakovosti podatkov, parametrizacijo poslovnih procesov in prilaganje orodij dejanskim nalogam.

Pristop, ki temelji na podatkih, ni izdelek, ki ga lahko samo prenesete ali kupite. Gre za strategijo, ki jo je treba zgraditi. Začne se s svežim pogledom na obstoječe procese in težave, nato pa je treba disciplinirano delovati v izbrani smeri.

Vodilni razvijalci programske opreme in prodajalci aplikacij ne bodo gonilo sprememb v gradbeništvu, saj za mnoge od njih pristop, ki temelji na podatkih, pomeni grožnjo njihovemu ustaljenemu poslovemu modelu.

*V drugih panogah [za razliko od gradbeništva], na primer v avtomobilski industriji, so se že zgodile korenite in prelomne spremembe, njihova digitalna preobrazba pa je v polnem teku. Gradbena podjetja morajo ukrepati hitro in odločno: vitka podjetja bodo požela velike koristi, medtem ko bodo tveganja za tiste, ki bodo oklevali, velika. Pомislite na preobrat, ki ga je v tej panogi povzročila digitalna fotografija [5].*

- Poročilo Svetovnega gospodarskega foruma Oblikovanje prihodnosti gradbeništva, 2016.

Podjetja, ki bodo pravočasno prepoznala priložnosti in prednosti novega pristopa, bodo pridobila trajnostno konkurenčno prednost ter se lahko razvijala in rasla brez odvisnosti od rešitev velikih ponudnikov.

To je vaša priložnost, da ne le preživite prihajajočo nevihto digitalizacije informacij, temveč da jo tudi prevzamete v svoje roke. V tej knjigi ne boste našli le analize trenutnega stanja v panogi, temveč tudi konkretna priporočila za premislek in prestrukturiranje vaših procesov in poslovanja, da bi postali vodilni v novi dobi gradbeništva in izboljšali svoje poklicne izkušnje.

Digitalna prihodnost gradbeništva ne pomeni le uporabe novih tehnologij in programov, temveč tudi temeljiti premislek o ravnanju s podatki in poslovnih modelih.

Ali je vaše podjetje pripravljeno na to strateško spremembo?



## KAZALO VSEBINE

<b>UVOD .....</b>	<b>1</b>
<b>KAZALO VSEBINE .....</b>	<b>I</b>
<b>II DEL OD GLINENIH TABLIC DO DIGITALNE REVOLUCIJE: KAKO SO SE RAZVIJALE INFORMACIJE V GRADBENIŠTVU.....</b>	<b>2</b>
POGLAVJE 2.1.RAZVOJ UPORABE PODATKOV V GRADBENIŠTVU.....	3
Rojstvo podatkovne dobe v gradbeništvu .....	3
Od gline in papirusa do digitalne tehnologije .....	4
Proces kot orodje za izkušnje, ki temeljijo na podatkih .....	5
Digitalizacija informacij o gradbenem procesu .....	7
POGLAVJE 2.2.TEHNOLOGIJE IN SISTEMI UPRAVLJANJA V SODOBNEM GRADBENIŠTVU ..	11
Digitalna revolucija in pojav modularnih MRP/ERP -sistemov .....	11
Sistemi za upravljanje podatkov: od podatkovnega rudarjenja do poslovnih izzivov .....	13
Korporativni micelij: kako so podatki povezani s poslovnimi procesi.....	17
POGLAVJE 2.3.DIGITALNA REVOLUCIJA IN EKSPLOZIJA PODATKOV .....	20
Začetek podatkovnega razcveta kot evolucijski val .....	20
Količina podatkov, ki jih ustvari sodobno podjetje .....	22
Stroški shranjevanja podatkov: ekonomski vidik.....	23
Meje kopiranja podatkov: od mase do pomena .....	25
Naslednji koraki: od teorije podatkov do praktičnih sprememb .....	27
<b>III DEL KAKO SE GRADBENIŠTVO UTAPLJA V PODATKOVNEM KAOSU .....</b>	<b>28</b>
POGLAVJE 3.1.RAZDROBLJENOST PODATKOV IN SILOSI.....	29
Več kot je orodij, bolj učinkovito je podjetje? .....	29
Podatkovni silosi in njihov vpliv na uspešnost podjetja .....	31
Podvajanje in pomanjkanje kakovosti podatkov kot posledica neenotnosti.....	34
HiPPO ali nevarnost mnenj pri odločanju.....	36
nenehno povečevanje kompleksnosti in dinamičnosti poslovnih procesov .....	39
Četrti industrijska revolucija (industrija 4.0) in peta industrijska revolucija (industrija 5.0) v gradbeništvu.....	42
POGLAVJE 3.2.SPREMINJANJE KAOSA V RED IN ZMANJŠEVANJE KOMPLEKSNOTI .....	45
Odvečna koda in zaprti sistemi kot ovira za izboljšanje produktivnosti .....	45
Od silosov do enotnega podatkovnega skladišča .....	47

Integrirani sistemi za shranjevanje omogočajo prehod na agente z umetno inteligenco .....	48
Od zbiranja podatkov do odločanja: pot do avtomatizacije.....	51
Naslednji koraki: spreminjanje kaosa v obvladljiv sistem.....	53
<b>IV DEL PODATKOVNI OKVIR V POSLOVNIH PROCESIH V GRADBENIŠTVU .....</b>	<b>55</b>
POGLAVJE 4.1.PODATKOVNE VRSTE V GRADBENIŠTVU .....	56
Najpomembnejše vrste podatkov v gradbeništvu .....	56
Strukturirani podatki .....	60
Relacijske zbirke podatkov RDBMS in poizvedovalni jezik SQL .....	61
SQL - poizvedbe v zbirkah podatkov in novi trendi .....	64
Nestrukturirani podatki .....	66
Besedilni podatki: med nestrukturiranim kaosom in strukturiranim kaosom y.....	67
Polstrukturirani in ohlapno strukturirani podatki .....	68
Geometrijski podatki in njihova uporaba .....	69
Podatki CAD: od načrtovanja do shranjevanja podatkov .....	72
Pojav koncepta BIM (BOM) in uporaba CAD v procesih .....	75
POGLAVJE 4.2.USKLAJEVANJE IN STRUKTURIRANJE PODATKOV.....	81
Polnjenje sistemov s podatki v gradbeništvu.....	81
Preoblikovanje podatkov: ključni temelj sodobne poslovne analize .....	83
podatkovni modeli: razmerja v podatkih in razmerja med elementi .....	87
Lastniški formati in njihov vpliv na digitalne procese .....	91
Odprti formati spreminjajo pristop k digitalizaciji.....	95
Sprememba paradigm: odprta koda kot konec obdobja prevlade proizvajalcev programske opreme.....	96
Strukturirani odprti podatki: temelj digitalne preobrazbe.....	99
POGLAVJE 4.3.LLM IN NJIHOVA VLOGA PRI OBDELDI PODATKOV IN POSLOVNIH PROCESIH .....	102
Klepetalnice LLM: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok za avtomatizacijo podatkovnih procesov.....	102
Veliki jezikovni modeli LLM: kako deluje .....	103
Uporaba lokalnih LLM za občutljive podatke podjetja.....	106
Popoln nadzor nad umetno inteligenco v podjetju in kako namestiti lasten LLM .....	108
RAG: Inteligentni LLM - pomočniki z dostopom do podatkov podjetja.....	110
POGLAVJE 4.4.IDE S PODPORO LLM IN PRIHODNJIMI PROGRAMSKIMI SPREMSEMBAMI ....	112
Izbira IDE: od eksperimentov LLM do poslovnih rešitev .....	112
IDE s podporo LLM in prihodnje spremembe programiranja.....	114

Python Pandas: nepogrešljivo orodje za delo s podatki .....	115
DataFrame: univerzalni format tabelaričnih podatkov .....	119
Naslednji koraki: vzpostavitev trajnostnega podatkovnega okvira .....	122
<b>V DEL KAKOVOST PODATKOV: ORGANIZACIJA, STRUKTURIRANJE, MODELIRANJE ..... 124</b>	
POGLAVJE 5.1. PRETVORBA PODATKOV V STRUKTURIRANO OBLIKO .....	125
Naučite se, kako pretvoriti dokumente, PDF, slike in besedila v strukturirane formate.....	125
Primer pretvorbe dokumenta PDF v tabelo .....	126
Pretvarjanje slik JPEG, PNG v strukturirano obliko.....	130
Pretvarjanje besedilnih podatkov v strukturirano obliko .....	133
Pretvorba podatkov CAD (BIM) v strukturirano obliko .....	136
Ponudniki rešitev CAD se usmerjajo k strukturiranim podatkom.....	141
POGLAVJE 5.2. KLASIFIKACIJA IN INTEGRACIJA: SKUPNI JEZIK ZA GRADBENE PODATKE .144	
Hitrost odločanja je odvisna od kakovosti podatkov .....	144
Standardizacija in integracija podatkov .....	145
Digitalna interoperabilnost se začne z zahtevami.....	148
Skupni jezik gradnje: vloga klasifikatorjev pri digitalni preobrazbi.....	150
Masterformat, OmniClass, Uniclass in CoClass: razvoj klasifikacijskih sistemov .....	153
POGLAVJE 5.3. MODELIRANJE PODATKOV IN CENTER ODLIČNOSTI .....	158
Modeliranje podatkov: konceptualni, logični in fizični model .....	158
Praktično modeliranje podatkov v gradbeništvu .....	162
Ustvarjanje podatkovne zbirke s programom LLM.....	164
Center odličnosti (CoE) za modeliranje podatkov .....	166
POGLAVJE 5.4. SISTEMATIZACIJA ZAHTEV IN POTRJEVANJE INFORMACIJ.....	169
Zbiranje in analiziranje zahtev: preoblikovanje komunikacij v strukturirane podatke.....	169
Procesne sheme in učinkovitost konceptualnih okvirov .....	173
Strukturirane zahteve in regularni izrazi RegEx.....	175
Zbiranje podatkov za postopek preverjanja .....	180
Preverjanje podatkov na spletni strani in rezultati preverjanja .....	182
Vizualizacija rezultatov preverjanja .....	187
Primerjava preverjanja kakovosti podatkov s potrebami človeškega življenja .....	189
Naslednji koraki: pretvorba podatkov v natančne izračune in načrte .....	191
<b>VI DEL IZRAČUNI STROŠKOV IN ČASA: VKLJUČEVANJE PODATKOV V GRADBENE PROCESE..... 193</b>	
POGLAVJE 6.1. IZRAČUNI IN OCENE STROŠKOV ZA GRADBENE PROJEKTE.....	194

Osnove gradbeništva: ocenjevanje količine, stroškov in časa.....	194
Metode za izračun ocenjenih stroškov projektov .....	195
Metoda, ki temelji na virih obračunavanje stroškov in ocenjevanje v gradbeništvu .....	196
Podatkovna baza gradbenih virov: katalog gradbenih materialov in del .....	196
Sestavljanje izračunov in obračunavanje stroškov del na podlagi baze virov.....	198
Stroški končnega projekta: od ocen do proračunov.....	203
<b>POGLAVJE 6.2. KOLIČINSKI IZPIS TER SAMODEJNO OBLIKOVANJE PREDRAČUNOV IN RAZPOREDOV .....</b>	<b>207</b>
Prehod iz 3D v 4D in 5D: uporaba volumetričnih in kvantitativnih parametrov .....	207
5D atributi in pridobivanje volumnov atributov iz CAD .....	207
QTO Količinski prevzem: združevanje podatkov o projektu po atributih .....	211
Avtomatizacija QTO z uporabo LLM in strukturiranih podatkov.....	216
QTO izračun celotnega projekta z uporabo skupinskih pravil iz Excelove preglednice .....	220
<b>POGLAVJE 6.3. 4D, 6D -8D IN IZRAČUN EMISIJ OGLJIKOVEGA DIOKSIDA CO<sub>2</sub> .....</b>	<b>226</b>
Model 4D: vključitev časa v gradbene ocene .....	226
Časovni načrt gradnje in njegova avtomatizacija na podlagi podatkov o stroških.....	227
Razširjeni atributni sloji 6D -8D: od energetske učinkovitosti do zagotavljanja varnosti .....	229
Ocena emisij CO <sub>2</sub> in izračun emisij ogljikovega dioksida pri gradbenih projektih.....	232
<b>POGLAVJE 6.4. GRADBENI SISTEMI ERP IN PMIS.....</b>	<b>236</b>
Gradbeni ERP -sistemi na primeru izračunov in ocen .....	236
PMIS: vmesni člen med ERP in gradbiščem .....	241
Špekulacije, dobiček, izoliranost in pomanjkanje preglednosti v ERP in PMIS.....	242
Konec obdobja zaprtih sistemov ERP /PMIS: gradbena industrija potrebuje nove pristope	245
Naslednji koraki: učinkovita uporaba podatkov o projektih .....	246
<b>VII DEL CAD IN BIM: TRŽENJE, RESNIČNOST IN PRIHODNOST PROJEKTNIH PODATKOV V GRADBENIŠTVU .....</b>	<b>249</b>
<b>POGLAVJE 7.1. POJAV KONCEPTOV BIM- V GRADBENIŠTVU.....</b>	<b>250</b>
Zgodovina pojava BIM in odprtega BIM kot tržnih konceptov prodajalcev CAD-.....	250
Realnost BIM: namesto integriranih podatkovnih baz - zaprti modularni sistemi .....	253
Pojav odprtega formata IFC v gradbeništvu .....	255
Težava s formatom IFC glede na geometrijsko jedro .....	257
Pojavljanje v konstrukciji teme semantike in ontologije .....	260
Zakaj semantične tehnologije v gradbeništvu ne izpolnjujejo pričakovanj.....	262
<b>POGLAVJE 7.2. ZAPRTI FORMATI PROJEKTOV IN VPRAŠANJA INTEROPERABILNOSTI. ....</b>	<b>266</b>
Zaprti podatki in padajoča produktivnost: slepa ulica industrije CAD (BIM) .....	266

Mit o interoperabilnosti med sistemi CAD .....	268
Prehod na USD in granularni podatki m.....	272
<b>POGLAVJE 7.3. GEOMETRIJA V GRADBENIŠTVU: OD ČRT DO KUBIČNIH METROV.....</b>	<b>276</b>
Ko se črte spremenijo v denar ali zakaj gradbeniki potrebujejo geometrijo .....	276
Od črt do volumnov: kako površina in volumen postaneta podatki .....	276
Prehod na MESH, USD in poligone: uporaba teselacije za geometrijo.....	278
LOD, LOI, LOMD - enotna klasifikacija podrobnosti v CAD (BIM) .....	280
Novi standardi CAD (BIM) - AIA, BEP, IDS, LOD, COBie .....	283
<b>POGLAVJE 7.4. PARAMETRIZACIJA ZASNOVE IN UPORABA LLM ZA DELOVANJE S KADROM .....</b>	<b>288</b>
Iluzija edinstvenosti podatkov CAD (BIM): pot do analitike in odprtih formatov.....	288
Oblikovanje prek parametrov: prihodnost CAD in BIM .....	291
Pojav LLM v procesih obdelave podatkov CAD za oblikovanje .....	294
Avtomatisirana analiza datotek DWG z LLM in Pandas .....	297
Naslednji koraki: prehod z zaprtih oblik na odprte podatke.....	303
<b>VIII DEL ODLOČANJE NA PODLAGI PODATKOV, ANALITIKA, AVTOMATIZACIJA IN STROJNO UČENJE. .....</b>	<b>305</b>
<b>POGLAVJE 8.1. PODATKOVNA ANALITIKA IN ODLOČANJE NA PODLAGI PODATKOV. ....</b>	<b>306</b>
Podatki kot vir pri sprejemanju odločitev .....	306
Vizualizacija podatkov: ključ do razumevanja in odločanja.....	309
KPI in ROI .....	311
Nadzorne plošče in nadzorne plošče: vizualizacija metrik za učinkovito upravljanje .....	313
Analiza podatkov in umetnost postavljanja vprašanj.....	315
<b>POGLAVJE 8.2. PRETOK PODATKOV BREZ ROČNEGA DELA: ZAKAJ JE POTREBEN ETL .....</b>	<b>318</b>
Avtomatisacija ETL: zmanjšanje stroškov in pospešitev obdelave podatkov.....	318
ETL Izvleček: zbiranje podatkov.....	322
ETL Preoblikovanje: uporaba pravil potrjevanja in preoblikovanja .....	325
ETL Load: vizualizacija rezultatov v preglednicah in grafih .....	327
ETL Load: samodejno ustvarjanje dokumentov PDF.....	332
ETL Load: samodejno ustvarjanje dokumentov iz FPDF.....	333
ETL Load: poročanje in nalaganje v druge sisteme.....	337
ETL z LLM: vizualizacija podatkov iz dokumentov PDF .....	338
<b>POGLAVJE 8.3. AVTOMATSKI TRANSPORTER ETL (CEVOVOD) .....</b>	<b>343</b>
Cevovod: Samodejno ETL transporter podatkov.....	343
Cevovod - Postopek potrjevanja podatkov v skladu s standardom LLM .....	347

Cevovod -ETL: preverjanje podatkov in informacij o elementih projekta v CAD (BIM) .....	349
POGLAVJE 8.4. ORKESTRACIJA ETL IN DELOVNIH TOKOV: PRAKTIČNE REŠITVE .....	355
DAG in Apache Airflow: avtomatizacija in orkestracija poteka dela .....	355
Apache Airflow: praktična uporaba pri avtomatizaciji ETL .....	356
Apache NiFi za usmerjanje in pretvorbo podatkov .....	360
n8n Low-Code, No-Code orkestracija procesov .....	361
Naslednji koraki: prehod z ročnih postopkov na rešitve, ki temeljijo na analitiki .....	364
<b>IX DEL SHRANJEVANJE IN UPRAVLJANJE PODATKOV V GRADBENIŠTVU .....</b>	<b>366</b>
POGLAVJE 9.1. PODATKOVNA INFRASTRUKTURA: OD OBLIK SHRANJEVANJA DO DIGITALNIH REPOZITORIJEV .....	367
Podatkovni atomi: temelj učinkovitega upravljanja informacij .....	367
Shranjevanje informacij: datoteke ali podatki .....	368
Shranjevanje velikih količin podatkov: analiza priljubljenih formatov in njihove učinkovitosti .....	370
Optimizacija shranjevanja z Apache Parquet .....	373
DWH: podatkovno skladišče podatkovna skladišča .....	375
Podatkovno jezero - razvoj ETL do ELT: od tradicionalnega čiščenja do prilagodljive obdelave .....	377
Arhitektura podatkovnega skladišča: sinergija skladišč in podatkovnih jezer .....	379
CDE, PMIS, ERP ali DWH in podatkovno jezero .....	381
POGLAVJE 9.2. UPRAVLJANJE PODATKOVNEGA SKLADIŠČA IN PREPREČEVANJE KAOSA .....	385
Vektorske zbirke podatkov in omejitveno polje .....	385
Upravljanje podatkov), podatkovni minimalizem) in podatkovno močvirje) .....	388
DataOps in VectorOps: novi podatkovni standardi .....	391
Naslednji koraki: od kaotičnega do strukturiranega shranjevanja .....	393
<b>X DEL VELIKI PODATKI, STROJNO UČENJE IN NAPOVEDI .....</b>	<b>395</b>
POGLAVJE 10.1. VELIKI PODATKI IN NJIHOVA ANALIZA .....	396
Veliki podatki v gradbeništvu: od intuicije do predvidljivosti .....	396
Vprašanje izvedljivosti velikih podatkov: korelacija, statistika in vzorčenje podatkov .....	397
Veliki podatki: analiza podatkov iz zbirke podatkov o milijonih gradbenih dovoljenj v San Franciscu .....	400
Primer velikih podatkov na podlagi podatkov CAD (BIM) .....	406
Internet stvari Internet stvari in pametne pogodbe .....	410
POGLAVJE 10.2. STROJNO UČENJE IN NAPOVEDI .....	414
Strojno učenje in umetna inteligenca bosta spremenila način gradnje .....	414

Od subjektivne ocene do statistične napovedi .....	416
Zbirka podatkov Titanik: Pozdravljen svet v svetu analitike podatkov in velikih podatkov ..	418
Strojno učenje v akciji: od potnikov na Titaniku do upravljanja projektov .....	423
Napovedi in napovedi na podlagi preteklih podatkov.....	427
Ključni koncepti strojnega učenja.....	430
<b>POGLAVJE 10.3.NAPOVEDOVANJE STROŠKOV IN ROKOV Z UPORABO STROJNEGA UČENJA .....</b>	<b>433</b>
Primer uporabe strojnega učenja za iskanje stroškov in časovnice projekta .....	433
Napovedovanje stroškov in časa projekta z uporabo linearne regresije .....	435
Napovedi stroškov in časa projekta z uporabo algoritma K-najbližji sosed (k-NN).....	438
Naslednji koraki: od shranjevanja do analize in napovedovanja .....	442
<b>XI DEL GRADBENIŠTVO V DIGITALNI DOBI. PRILOŽNOSTI IN IZZIVI .....</b>	<b>445</b>
<b>POGLAVJE 11.1.STRATEGIJE PREŽIVETJA: USTVARJANJE KONKURENČNIH PREDNOSTI..</b>	<b>446</b>
Korelacije namesto izračunov: prihodnost gradbene analitike.....	446
Na podatkih temelječ pristop v gradbeništvu: nova raven infrastrukture .....	449
Digitalna pisarna naslednje generacije: kako umetna inteligenca spreminja delovni prostor .....	451
Odprti podatki in uberizacija ogrožajo obstoječa gradbena podjetja.....	453
Nerešeni problemi uberizacije kot zadnja priložnost, da izkoristimo čas za preobrazbo ....	456
<b>POGLAVJE 11.2.PRAKTIČNI VODNIK ZA IZVAJANJE PRISTOPA, KI TEMELJI NA PODATKIH</b>	<b>461</b>
Od teorije k praksi: načrt za digitalno preobrazbo v gradbeništvu .....	461
Postavitev digitalnih temeljev: 1-5 korakov do digitalne zrelosti.....	463
Izkoriščanje potenciala podatkov: 5-10 korakov do digitalne zrelosti .....	467
Načrt preobrazbe: od kaosa do podatkovno usmerjenega podjetja .....	474
Gradnja v industriji 5.0: Kako zaslužiti, ko se ne morete več skrivati.....	477
<b>ZAKLJUČEK .....</b>	<b>479</b>
<b>O AVTORJU .....</b>	<b>482</b>
<b>NADALJNJI ODNOS .....</b>	<b>483</b>
<b>KOMENTAR PREVODA .....</b>	<b>483</b>
<b>DRUGA ZNANJA IN VEŠČINE TER KONCEPTI.....</b>	<b>484</b>
<b>GLOSARIJ .....</b>	<b>488</b>
<b>SEZNAM LITERATURE IN SPLETNEGA GRADIVA .....</b>	<b>495</b>
<b>INDEKS SUBJEKTOV .....</b>	<b>512</b>



## NAJVEČJE UDOBJE S TISKANO RAZLIČICO

V rokah imate brezplačno digitalno različico knjige **Data-Driven Construction**. Za udobnejše delo in hiter dostop do gradiva vam priporočamo, da se posvetite **tiskani izdaji**:



■ **Vedno pri roki:** knjiga v tiskani obliki bo postala zanesljivo delovno orodje, s katerim boste lahko hitro našli in uporabili prave vizualizacije in diagrame v vsaki delovni situaciji.

■ **Visoka kakovost ilustracij:** vse slike in grafike v tiskani izdaji so predstavljene v najvišji kakovosti.

■ **Hiter dostop do informacij:** enostavna navigacija, možnost beleženja opomb, zaznamkov in dela s knjigo kjer koli.



Z nakupom polne tiskane različice knjige dobite priročno orodje za udobno in učinkovito delo z informacijami: možnost hitre uporabe vizualnega gradiva pri vsakodnevnih opravilih, hitrega iskanja potrebnih shem in zapisovanja. Poleg tega s svojim nakupom podpirate širjenje odprtega znanja.

Tiskano različico knjige naročite na: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## I DEL

# **OD GLINENIH TABLIC DO DIGITALNE REVOLUCIJE: KAKO SO SE RAZVIJALE INFORMACIJE V GRADBENIŠTVU**

Prvi del knjige obravnava zgodovinski razvoj upravljanja podatkov v gradbeništvu, od primitivnih zapisov na fizičnih medijih do sodobnih digitalnih ekosistemov. Analizira preoblikovanje tehnologij za upravljanje informacij, pojav sistemov ERP in vpliv razdrobljenosti podatkov na učinkovitost poslovnih procesov. Posebna pozornost je namenjena procesu digitalizacije informacij in vse večjemu pomenu objektivne analize namesto subjektivne strokovne presoje. Podrobno so obravnavani eksponentna rast obsega informacij, s katero se sooča sodobna gradbena industrija, in z njo povezani izzivi za sisteme podjetij. Raziskana je umestitev gradbeništva v kontekst četrte in pete industrijske revolucije ter možnosti uporabe umetne inteligence in podatkovno usmerjenih pristopov za ustvarjanje trajnostne konkurenčne prednosti.

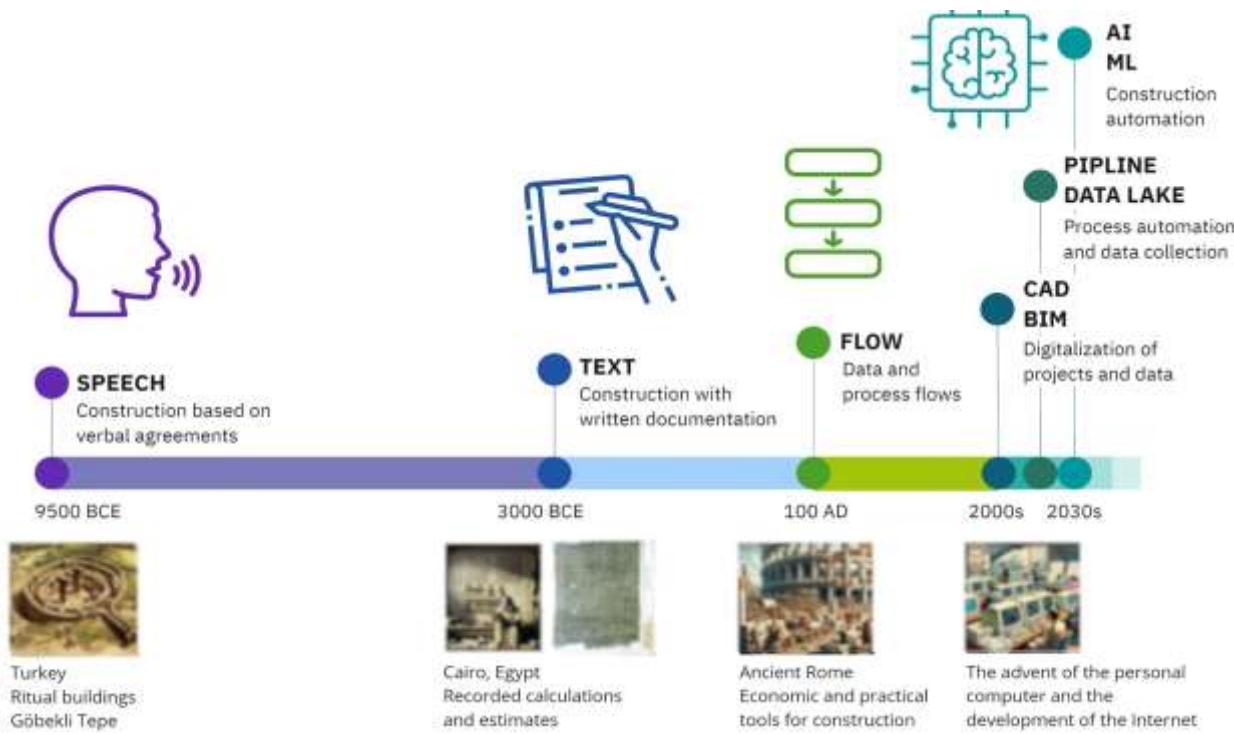
## POGLAVJE 1.1.

# RAZVOJ UPORABE PODATKOV V GRADBENIŠTVU

### Rojstvo podatkovne dobe v gradbeništvu

Pred približno 10.000 leti, v neolitiku, je človeštvo naredilo revolucionarno spremembo v svojem razvoju, ko je opustilo nomadski način življenja in se preusmerilo v sedentarno življenje, kar je privedlo do pojava prvih primitivnih stavb iz gline, lesa in kamna [6]. Od tega trenutka se začenja zgodovina gradbeništva.

Z razvojem civilizacij je arhitektura postajala vse bolj zapletena, kar je pripeljalo do prvih obrednih templjev in javnih zgradb. Vedno bolj zapleteni arhitekturni načrti so od inženirjev in menedžerjev v antiki zahtevali prve zapise in izračune. Prvi zapisi na glinenih tablicah in papirusih so pogosto vsebovali opis logike izračuna količine potrebnih gradbenih materialov, njihove cene in izračuna plačila za opravljeno delo [7]. Tako se je začelo obdobje uporabe podatkov v gradbeništvu - dolgo pred pojavom sodobnih digitalnih tehnologij (slika 1.1-1).



Slika 1.1-1 Kronologija razvoja informacijske tehnologije v gradbeništvu: od verbalnih informacij do umetne inteligence.

## Od gline in papirusa do digitalne tehnologije

Prvi dokumenti o gradnji segajo v obdobje gradnje piramid, približno 3000-4000 let pred našim štetjem[7]. Od takrat je vodenje pisnih zapisov olajšalo in spremljalo napredek v gradbeništvu ter omogočilo zbiranje in sistematizacijo znanja, ki je v naslednjih 10.000 letih privedlo do pomembnih inovacij na področju gradbenih metod in arhitekture.

Uporaba prvih fizičnih medijev v gradbeništvu, kot so glinene tablice, papirus izpred tisočletja (slika 1.1-2) ali papir "A0" v osemdesetih letih prejšnjega stoletja, za zapisovanje podatkov prvotno ni bila namenjena uporabi teh informacij pri novih projektih. Glavni namen takšnih zapisov je bil podrobno opisati trenutno stanje projekta, vključno z izračuni potrebnih materialov in stroškov dela. Podobno tudi v današnjem svetu razpoložljivost digitalnih projektnih podatkov in modelov ne zagotavlja vedno njihove uporabe v prihodnjih projektih in pogosto služi predvsem kot informacija za trenutne izračune potrebnih materialov in stroškov gradnje.



Slika 1.1-2 Papirus iz 3. stoletja pred našim štetjem, ki opisuje stroške poslikave različnih vrst oken v kraljevi palači v tehniki enkaustike.

Človeštvo je potrebovalo približno 5.000 let, da je prešlo od ustnih pogоворов k pisnim dokumentom pri upravljanju gradbenih projektov, in prav toliko časa, da je prešlo od papirja k digitalnim podatkom kot glavnemu viru za načrtovanje in nadzor.

Tako kot je razvoj trgovinskih in denarnih odnosov spodbudil nastanek pisave in prvih odvetnikov za reševanje sporov, so prvi zapisi materialnih stroškov in obsega del v gradbeništvu priveli do nastanka prvih vodij v gradbeništvu, katerih naloge so vključevale dokumentiranje, spremljanje in odgovornost za ključne informacije o časovnem načrtu projekta in stroških.

Danes imajo podatki veliko pomembnejšo vlogo: ne le da beležijo sprejete odločitve, temveč postajajo tudi orodje za napovedovanje in modeliranje prihodnosti. To je temelj, na katerem temelji sodobni procesni pristop pri vodenju projektov - zbrane izkušnje se spremenijo v sistem odločanja, ki temelji na strukturiranih in preverljivih podatkih.

## Proces kot orodje za izkušnje, ki temeljijo na podatkih

Bistvo vsakega procesa je preoblikovanje preteklih izkušenj v orodje za načrtovanje prihodnosti. Izkušnje v sodobnem smislu so strukturiran niz podatkov, ki jih je mogoče analizirati in na njihovi podlagi pripraviti utemeljene napovedi.

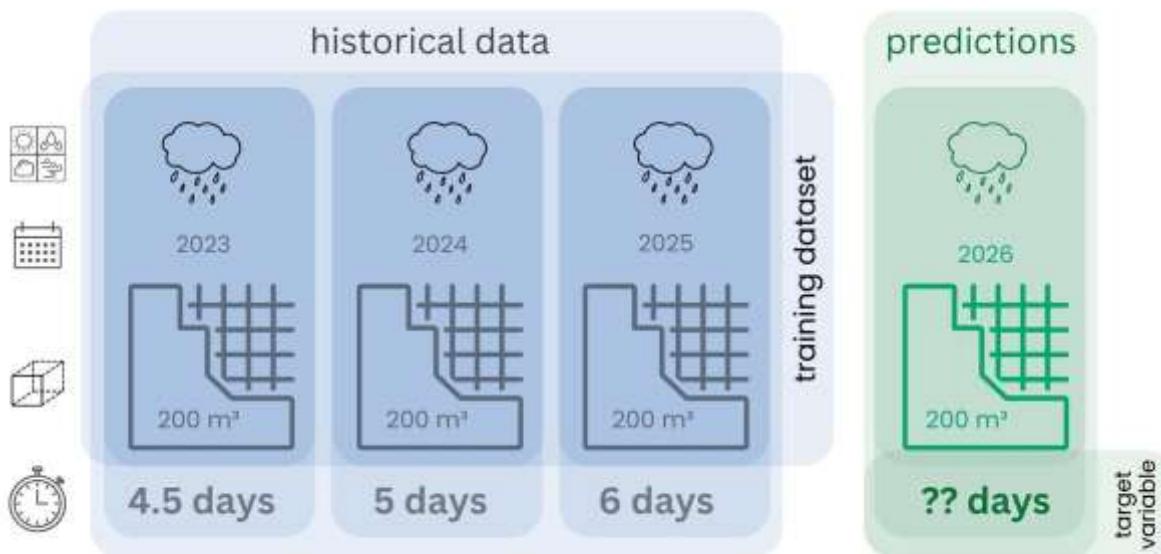
Zgodovinski podatki so temelj napovedovanja, saj jasno prikazujejo rezultate opravljenega dela in omogočajo vpogled v dejavnike, ki vplivajo na te rezultate.

Vzemimo konkreten primer iz monolitne gradnje: običajno se pri načrtovanju časovnega razporeda del upoštevajo količina betona, zapletenost konstrukcije in vremenske razmere. Predpostavimo, da določen mojster na gradbišču ali zgodovinski podatki podjetja za zadnja tri leta (2023-2025) kažejo, da je betoniranje  $200 \text{ m}^2$  velike monolitne konstrukcije v deževnem vremenu trajalo od 4,5 do 6 dni (slika 1.1-3). Prav ti zbrani statistični podatki postanejo osnova za napovedovanje časa izvedbe in stroškov virov pri načrtovanju podobnih del v prihodnjih projektih. Na podlagi teh preteklih podatkov lahko mojster ali ocenjevalec na podlagi izkušenj utemeljeno napove čas, ki bo potreben za dokončanje podobnih del v prihodnosti leta 2026 v podobnih pogojih.

V primeru časovne analitike analitični proces deluje kot mehanizem za preoblikovanje raznovrstnih podatkov v strukturirane izkušnje in nato v natančno orodje za načrtovanje. Podatki in procesi so enoten ekosistem, kjer eden brez drugega ne more obstajati.

*Preštejte, kar je mogoče prešteti, izmerite, kar je mogoče izmeriti, in poskrbite, da bo mogoče izmeriti, česar ni mogoče izmeriti [8].*

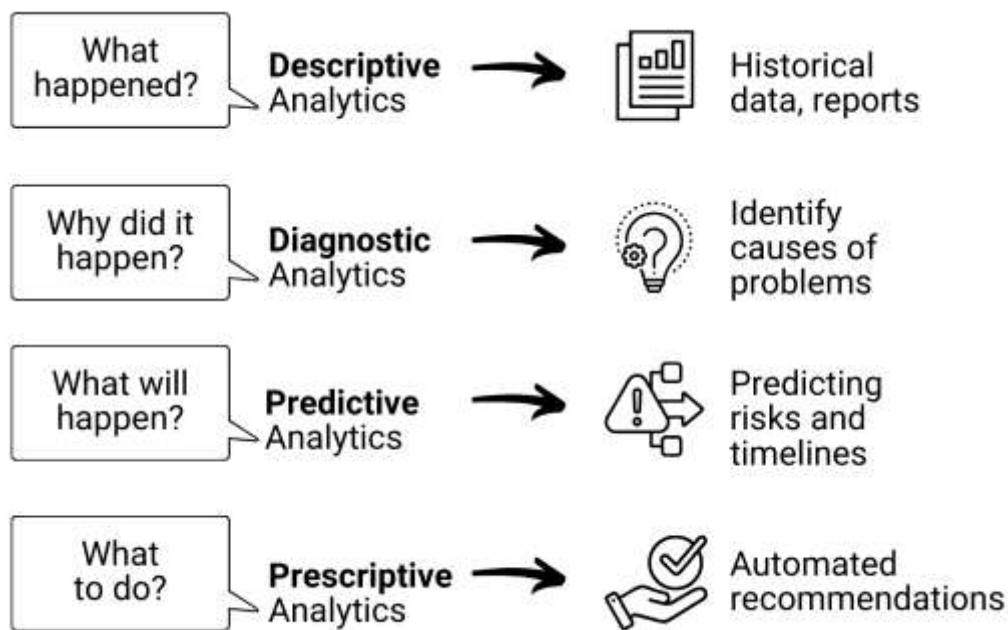
- Galileo Galilei



Slika 1.1-3 Zgodovinski podatki služijo kot nabor učnih podatkov za napovedovanje ene od vrednosti v prihodnosti.

V današnjem poslovnem okolju postaja podatkovna analitika ključna sestavina učinkovitega vodenja projektov, optimizacije procesov in strateškega odločanja. Gradbeništvo postopoma obvladuje štiri ključne ravni analitike, od katerih vsaka odgovarja na določeno vprašanje in zagotavlja edinstvene koristi (slika 1.1-4):

- **Opisna analitika** - odgovarja na vprašanje "kaj se je zgodilo?" ter zagotavlja zgodovinske podatke in poročila o preteklih dogodkih in rezultatih: v zadnjih treh letih (2023-2025) je v deževnem vremenu trajalo od 4,5 do 6 dni, da se je zlila monolitna konstrukcija velikosti 200 m<sup>2</sup>.
- **Diagnostična analitika** - odgovori na vprašanje "zakaj se je to zgodilo?" z ugotavljanjem vzrokov za težave: analiza pokaže, da se je čas betoniranja monolitne konstrukcije podaljšal zaradi deževnega vremena, kar je upočasnilo proces strjevanja betona.
- **Prediktivna analitika** - usmerjena v prihodnost, napovedovanje možnih tveganj in časov izvedbe z odgovorom na vprašanje "kaj se bo zgodilo?": na podlagi preteklih podatkov je ob upoštevanju vseh znanih dejavnikov in trendov napovedano, da bo zlivanje podobne monolitne konstrukcije s površino 200 m<sup>2</sup> v deževnem vremenu leta 2026 trajalo približno 5,5 dni.
- **Preskriptivna analitika** - zagotavlja avtomatizirana priporočila in odgovore na vprašanje "kaj storiti?", kar podjetjem omogoča, da izberejo optimalne ukrepe: Za optimizacijo dela je na primer priporočljivo: uporabiti posebne dodatke za pospešitev strjevanja betona v pogojih visoke vlažnosti; načrtovati betoniranje za obdobja z najmanjšo verjetnostjo padavin; urediti začasne nadstreške za konstrukcijo, ki bodo skrajšali čas dela na 4-4,5 dneva tudi v neugodnih vremenskih razmerah.



Slika 1.1-4 Glavne vrste analitike: od preteklega opisa do avtomatiziranega odločanja.

Polnopravna digitalna preobrazba, ki pomeni prehod na sistemsko analitiko in upravljanje na podlagi podatkov, ne zahteva le zunanjega izvajanja, temveč tudi oblikovanje kompetentne notranje ekipe. Ključni člani takšne ekipe naj bodo produktivi vodje, podatkovni inženirji, analitiki in razvijalci, ki bodo tesno sodelovali s poslovnimi enotami (slika 4.3-9). To sodelovanje je potrebno za postavljanje inteligentnih analitičnih vprašanj in učinkovito parametriranje nalog poslovnega odločanja. V informacijski družbi podatki ne postanejo le pomožno orodje, temveč osnova za napovedovanje in optimizacijo.

V gradbeništvu digitalna preobrazba korenito spreminja način načrtovanja, upravljanja in obratovanja objektov. Ta proces se imenuje digitalizacija informacij - pri čemer se vsi vidiki gradbenega procesa digitalizirajo v digitalno obliko, primerno za analizo.

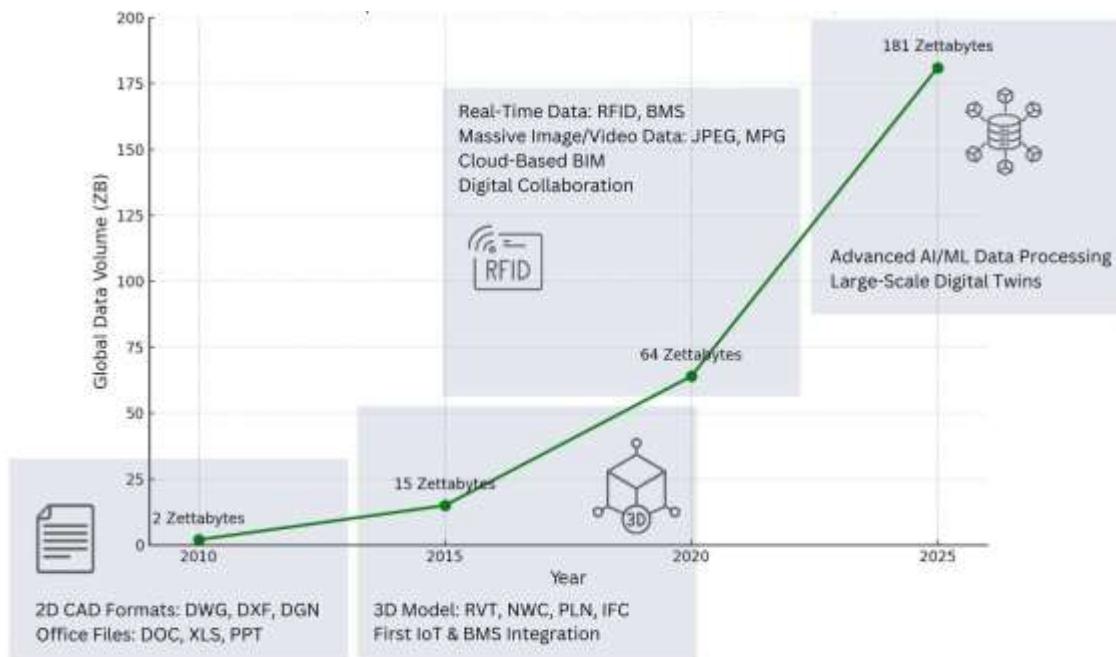
## Digitalizacija informacij o gradbenem procesu

V tisočletjih se količina informacij, zapisanih v gradbeništvu, skoraj ni spremenjala, vendar se je v zadnjih desetletjih hitro povečala (slika 1.1-5).

Glede na študijo družbe PwC® "Upravljeni podatki. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] je bilo 90% vseh podatkov na svetu ustvarjenih v zadnjih dveh letih (od leta 2015). Vendar večina podjetij teh podatkov ne izkorišča v celoti, saj ostajajo v ločenih sistemih ali pa se preprosto arhivirajo brez prave analize.

Obseg podatkov se je v zadnjih letih le še povečal in se podvojil s 15 zetabajtov leta 2015 na 181 zetabajtov leta 2025 [10]. Vsak dan se strežniki gradbenih in projektantskih podjetij polnijo s projektno

dokumentacijo, urniki dela, izračuni in kalkulacijami, finančnimi poročili. Za 2D/3D -črte se uporabljajo formati DWG, DXF in DGN, za 3D-modele pa RVT, NWC, PLN in IFC™. Besedilni dokumenti, tabele in predstavitev so shranjeni v formatih DOC, XLSX in PPT. Poleg videoposnetkov in slik z gradbišča - v formatih MPG in JPEG, se v realnem času shranjujejo podatki iz komponent interneta stvari, oznak RFID® (identifikacija in sledenje) in sistemov za upravljanje stavb BMS (spremljanje in nadzor)



Slika 1.1-5 Parabolična rast podatkov 2010-2025 (na podlagi [10]).

Zaradi hitre rasti informacij se gradbena industrija ne sooča le z zbiranjem in shranjevanjem podatkov, temveč tudi z zagotavljanjem njihovega preverjanja, potrjevanja, merljivosti in analitične obdelave. Danes je panoga v aktivni fazi digitalizacije informacij - sistematičnega preoblikovanja vseh vidikov gradbene dejavnosti v digitalno obliko, primerno za analizo, interpretacijo in avtomatizacijo.

Digitalizacija informacij pomeni, da je treba informacije o vseh subjektih in elementih gradbenega projekta in samega gradbenega procesa - vključno s tistimi, ki jih prej sploh nismo obravnavali kot informacije - prenesti v podatkovno obliko, da jih je mogoče količinsko opredeliti in enostavno analizirati.

Na področju gradbeništva to pomeni zajemanje in digitalizacijo informacij vseh elementih projektov in vseh procesov - od gibanja strojev in ljudi na gradbišču do vremenskih in podnebnih razmer na gradbišču, trenutnih cen materialov in obrestnih mer centralnih bank - za izdelavo analitičnih modelov.

*Če lahko to, o čemer govorite, izmerite in izrazite v številkah, potem o tem nekaj veste.  
 Če pa tega ne morete količinsko izraziti, je vaše znanje zelo omejeno in nezadovoljivo.  
 To je lahko izhodišče, vendar ni raven pravega znanstvenega znanja. [11].*

- W. Thomson (Lord Kelvin), 1824-1907, britanski znanstvenik

Digitalizacija informacij daleč presega tradicionalni pristop k zbiranju informacij, pri katerem so se beležili le osnovni kazalniki, kot so število delovnih ur ali dejanski stroški materiala. Danes je mogoče praktično vsak dogodek pretvoriti v podatkovni tok, primeren za poglobljeno analizo z uporabo naprednih analitičnih orodij in tehnik strojnega učenja. V gradbeništvu je prišlo do temeljnega premika od papirnatih risb, Excelovih preglednic in ustnih navodil k digitalnim sistemom (slika 1.2-4), v katerih vsak element projekta postane vir podatkov. Celo zaposleni - od inženirjev do gradbenih delavcev na gradbišču - so zdaj obravnavani kot zbirka digitalnih spremenljivk in podatkovnih nizov.

V skladu z raziskavo družbe KPMG "Familiar Challenges - New Approaches: Global Construction Survey 2023" se digitalni dvojčki, umetna inteligenca (AI) in veliki podatki, pojavljajo kot ključni dejavniki za izboljšanje donosnosti projektov [2].

Sodobne tehnologije ne le poenostavljajo zbiranje informacij in ga v veliki meri avtomatizirajo, temveč tudi bistveno zmanjšujejo stroške shranjevanja podatkov. Podjetja se zato odmikajo od selektivnega pristopa in raje shranjujejo celoten nabor informacij za poznejšo analizo (slika 2.1-5), kar odpira potencialne možnosti za optimizacijo procesov v prihodnosti.

Digitalizacija informacij in digitalizacija omogočata odkrivanje skrite, prej neizkoriščene vrednosti informacij. Če so podatki ustrezno organizirani, jih je mogoče ponovno uporabiti, interpretirati in vključiti v nove storitve in rešitve.

V prihodnosti bo digitalizacija informacij verjetno privedla do popolne avtomatizacije upravljanja dokumentov, uvedbe samoupravnih gradbenih procesov in pojava novih poklicev - gradbenih podatkovnih analitikov, strokovnjakov za upravljanje projektov z umetno inteligenco in digitalnih inženirjev. Gradbeni projekti bodo postali dinamični viri informacij, odločanje pa ne bo temeljilo na intuiciji ali subjektivnih izkušnjah, temveč na zanesljivih in ponovljivih digitalnih dejstvih

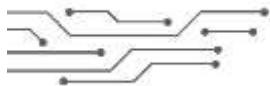
*Informacije so nafta 21. stoletja, analitika pa motor z notranjim izgorevanjem [12].*

- Peter Sondergaard, višji podpredsednik, Gartner®

V skladu z raziskavo IoT Analytics 2024 [13] naj bi se globalna poraba za upravljanje in analizo podatkov močno povečala, in sicer s 185,5 milijarde USD leta 2023 na 513,3 milijarde USD do leta 2030, kar pomeni 16-odstotno letno stopnjo rasti. Vendar pa vse komponente ne rastejo enako hitro: analitika hitro raste, medtem ko se rast shranjevanja upočasnuje. Najhitrejšo rast v ekosistemu za upravljanje podatkov bo zagotovljala analitika: po napovedih naj bi se povečala s 60,6 milijarde dolarjev leta 2023 na 227,9 milijarde dolarjev do leta 2030, kar pomeni 27-odstotno sestavljeni letno stopnjo rasti.

Zaradi vse hitrejše digitalizacije informacij in hitrega povečevanja obsega informacij se vodstvo gradbenih projektov in podjetij sooča s potrebo po sistematičnem shranjevanju, analizi in obdelavi različnih, pogosto heterogenih podatkov. Kot odgovor na ta izziv se je od sredine devetdesetih let prejšnjega stoletja dalje v panogi začel množičen prehod na elektronsko ustvarjanje, shranjevanje in upravljanje dokumentacije - od preglednic in projektnih izračunov do risb in pogodb.

Tradicionalne papirnate dokumente, ki zahtevajo podpise, fizično hrambo, redno pregledovanje in arhiviranje v omarah, postopoma nadomeščajo digitalni sistemi, ki podatke hranijo na strukturiran način - v podatkovnih zbirkah specializiranih aplikacij.



## POGLAVJE 1.2.

# TEHNOLOGIJE IN SISTEMI UPRAVLJANJA V SODOBNEM GRADBENIŠTVU.

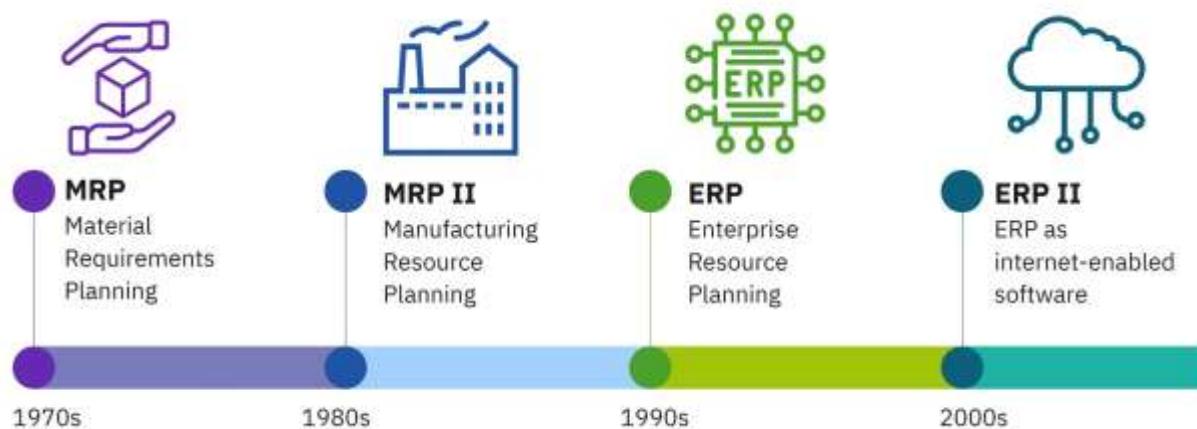
### Digitalna revolucija in pojav modularnih MRP/ERP -sistemov

Doba sodobnega digitalnega shranjevanja in obdelave podatkov se je začela s pojavom magnetnega traku v petdesetih letih prejšnjega stoletja, ki je odprl možnost shranjevanja in uporabe velikih količin informacij. Naslednji preboj je bil pojav diskovnih pogonov, ki so korenito spremenili pristop k upravljanju podatkov v gradbeništvu.

Z razvojem podatkovnih skladišč je na trg rešitev vstopilo veliko število podjetij, ki so začela razvijati modularno programsko opremo za ustvarjanje, shranjevanje in obdelavo podatkov ter avtomatizacijo rutinskih nalog.

Eksponentna rast informacij in orodij je povzročila potrebo po integriranih, modularnih rešitvah, ki ne delajo s posameznimi datotekami, temveč pomagajo upravljati in nadzorovati pretok podatkov med procesi in projekti.

Prva celovita platformna orodja so morala ne le shranjevati dokumente, temveč tudi dokumentirati vse zahteve za spremembe in operacije v procesih: kdo jih je sprožil, kakšen je bil obseg zahteve in kaj je bilo na koncu zapisano kot vrednost ali atribut. Za te namene je bil potreben sistem, ki je lahko spremljal natančne izračune in sprejete odločitve (slika 1.2-1). Takšne platforme so bili prvi sistemi MRP (Material Requirements Planning) in ERP (Enterprise Resource Planning), ki so postali priljubljeni od začetka devetdesetih let prejšnjega stoletja [14].

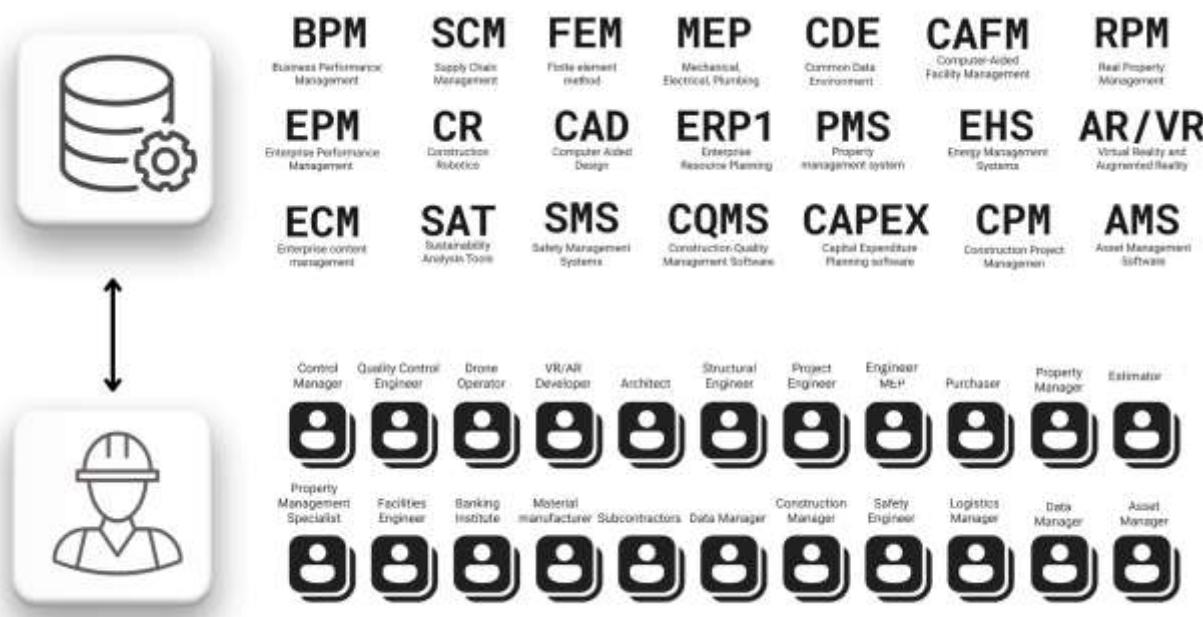


Slika 1.2-1 Napredok v tehnologiji shranjevanja podatkov je v osemdesetih letih prejšnjega stoletja priveljal do pojava sistemov ERP.

Prvi sistemi MRP - in ERP - so postavili temelje za obdobje digitalizacije poslovnih procesov in

upravljanja gradbenih projektov. Modularni sistemi, prvotno zasnovani za avtomatizacijo ključnih poslovnih procesov, so bili sčasoma povezani z dodatnimi, bolj prilagodljivimi in prilagodljivimi programskimi rešitvami.

Te dodatne rešitve so bile namenjene obdelavi podatkov in upravljanju projektnih vsebin (slika 1.2-2), nadomestile so nekatere module velikih sistemov ali pa so jih učinkovito dopolnile in razširile funkcionalnost celotnega sistema.



Slika 1.2-2 Nove programske rešitve so v podjetja privabile vojsko menedžerjev za upravljanje podatkovnih tokov.

V zadnjih desetletjih so podjetja veliko vlagala v modularne sisteme [15] in jih dojemala kot dolgoročne integrirane rešitve.

Glede na poročilo Software Path za leto 2022 [16] je povprečni proračun na uporabnika sistema ERP 9 000 USD. V povprečju takšne sisteme uporablja približno 26% zaposlenih v podjetju. Tako za organizacijo s 100 uporabniki skupni stroški uvedbe ERP dosežejo približno 900.000 USD.

Naložbe v lastniške, zaprte in modularne rešitve so zaradi hitrega razvoja sodobnih, prilagodljivih in odprtih tehnologij vse manj upravičene. Če so bile takšne naložbe že izvedene, je pomembno, da objektivno ponovno ocenimo vlogo obstoječih sistemov: ali so dolgoročno še vedno bistveni ali pa je mogoče njihove funkcije spremeniti in izvajati učinkoviteje in preglednejše.

Ena od ključnih težav današnjih modularnih podatkovnih platform je, da upravljanje podatkov centralizirajo v zaprtih aplikacijah. Zaradi tega postanejo podatki - glavno premoženje podjetja - odvisni od posebnih programskih rešitev in ne obratno. To omejuje ponovno uporabo informacij, otežuje migracijo na in zmanjšuje poslovno agilnost v hitro spreminjačem se digitalnem okolju.

Če je verjetno, da se bo pomen ali pomembnost zaprte modularne arhitekture v prihodnosti zmanjšal,

je smiselno, da danes nastale stroške prepoznamo kot potopljene stroške in se osredotočimo na strateški prehod na bolj odprt, razširljiv in prilagodljiv digitalni ekosistem.

Za lastniško programsko opremo je značilno, da ima razvijalec izključni nadzor nad izvorno kodo in uporabniškimi podatki, ki nastanejo pri uporabi takih rešitev. V nasprotju z odprtokodno programsko opremo uporabniki nimajo dostopa do notranje strukture aplikacije in je ne morejo neodvisno pregledovati, spremenjati ali prilagajati svojim potrebam. Namesto tega morajo kupiti licence, ki jim dajejo pravico do uporabe programske opreme v okviru omejitev, ki jih določi prodajalec.

Sodoben pristop, osredotočen na podatke, ponuja drugačno paradigma: podatke je treba obravnavati kot glavno strateško sredstvo - neodvisno, trajno in ločeno od posebnih programskih rešitev. Aplikacije pa postanejo zgolj podatkovna orodja, ki jih je mogoče prosto zamenjati brez tveganja izgube ključnih informacij.

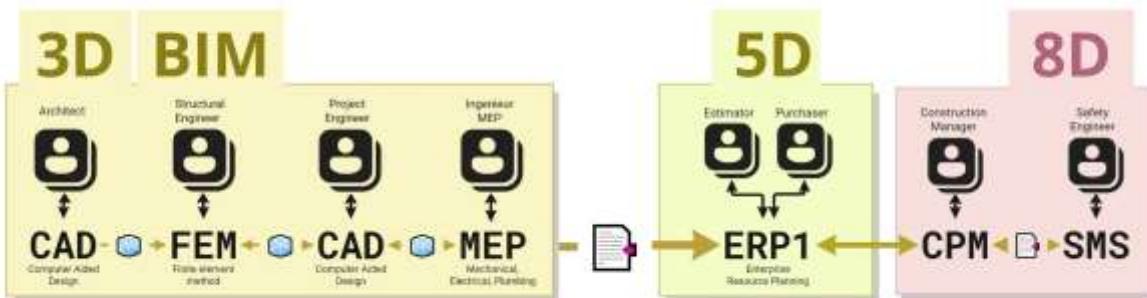
Razvoj sistemov ERP in MRP v devetdesetih letih prejšnjega stoletja (slika 1.2-1) je podjetjem zagotovil zmogljiva orodja za upravljanje procesov, vendar je imel tudi nenamerno posledico, da se je znatno povečalo število ljudi, vključenih v vzdrževanje informacijskih tokov. Namesto da bi ti sistemi avtomatizirali in poenostavili operativne naloge, so pogosto ustvarili nove ravni zapletenosti, birokracije in odvisnosti od notranjih virov IT.

## **Sistemi za upravljanje podatkov: od podatkovnega rudarjenja do poslovnih izzivov**

Današnja podjetja se soočajo s potrebo po integraciji več sistemov za upravljanje podatkov. Izbira sistemov za upravljanje podatkov, pametno upravljanje teh sistemov in povezovanje različnih virov podatkov postajajo ključnega pomena za poslovno uspešnost.

Sredi leta 2020 lahko najdete na stotine (v velikih gradbenih podjetjih pa na tisoče) različnih sistemov (slika 1.2-3), ki morajo delovati usklajeno, da vsi vidiki gradbenega procesa potekajo gladko in povezano.

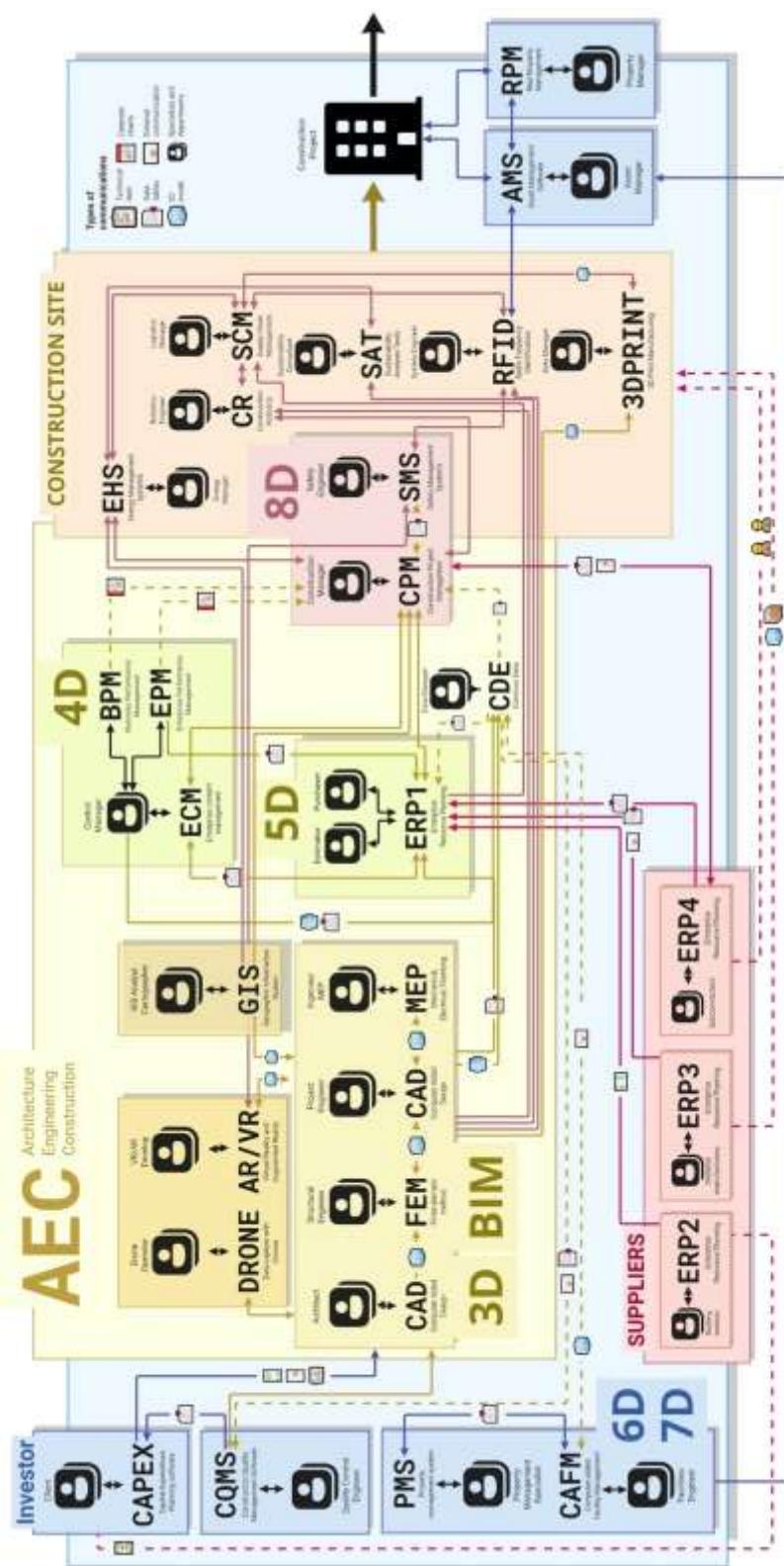
Po podatkih študije družbe Deloitte iz leta 2016® "Data-Driven Management in Digital Capital Projects" povprečen gradbeni strokovnjak dnevno uporablja 3,3 programske aplikacije, vendar je le 1,7 od njih medsebojno integriranih [17].



Slika 1.2-3 Vsak poslovni sistem potrebuje strokovno ekipo in odgovornega vodjo za kakovostno upravljanje podatkov.

V nadaljevanju je seznam priljubljenih sistemov za srednje velika in velika podjetja v gradbeništvu, ki se uporabljajo za učinkovito vodenje gradbenih projektov:

- **ERP (Enterprise Resource Planning)** - omogoča integracijo poslovnih procesov, vključno z računovodstvom, nabavo in upravljanjem projektov.
- **CAPEX (programska oprema za načrtovanje kapitalskih izdatkov)** - uporablja se za načrtovanje in upravljanje finančnih naložb v gradbene projekte, pomaga določiti stroške osnovnih sredstev in naložb v dolgoročna sredstva.
- **CAD (Computer-Aided Design) in BIM (Building Information Modeling)** - se uporabljata za izdelavo podrobnih in natančnih tehničnih risb in 3D - modelov projektov. Ti sistemi se osredotočajo na delo z geometrijskimi informacijami.
- **MEP (Mechanical, Electrical, Plumbing)** - inženirski sistemi, ki vključujejo mehanske, električne in vodovodne komponente ter podrobno opisujejo notranji "obtočni" sistem projekta.
- **GIS (geografski informacijski sistemi)** - uporabljajo se za analizo in načrtovanje terena, vključno s kartografijo in prostorsko analizo.
- **CQMS (programska oprema za upravljanje kakovosti v gradbeništvu)** - zagotavlja skladnost gradbenih postopkov z uveljavljenimi standardi in predpisi ter pomaga odpravljati napake.
- **CPM (vodenje gradbenih projektov)** - vključuje načrtovanje, usklajevanje in nadzor gradbenih procesov.
- **CAFM (Computer-Aided Facility Management)** - sistemi za upravljanje in vzdrževanje stavb.
- **SCM (Supply Chain Management)** je potreben za optimizacijo pretoka materialov in informacij med dobavitelji in gradbiščem.
- **EPM (Enterprise Performance Management)** - namenjen izboljšanju poslovnih procesov in učinkovitosti.
- **AMS (programska oprema za upravljanje sredstev)** - uporablja se za optimizacijo uporabe, upravljanja in vzdrževanja opreme in infrastrukture v celotnem življenjskem ciklu sredstev.
- **RPM (Upravljanje nepremičnin)** - vključuje naloge in procese, povezane z upravljanjem in delovanjem stavb in zemljišč ter z njimi povezanih virov in sredstev.



Slika 1.2-4 Medsebojna povezanost sistemov, ki povezuje procese podjetja s pretokom informacij med različnimi oddelki.

- **CAE (Computer-Aided Engineering)** - Računalniško podprto inženirstvo vključuje računalniške in simulacijske sisteme, kot sta analiza končnih elementov (FEA) in računalniška dinamika tekočin (CFD).
- **CFD (Computational Fluid Dynamics)** - Računalniška dinamika tekočin, modeliranje tokov tekočin in plinov. Podkategorija CAE.
- **CAPP (računalniško podprto načrtovanje procesov)** - Računalniško podprto načrtovanje procesov. Uporablja se za izdelavo zemljevidov poti in procesov.
- **CAM (Computer-Aided Manufacturing)** - računalniško podprta proizvodnja, izdelava krmilnih programov za stroje CNC.
- **PDM (Product Data Management)** - Upravljanje podatkov o izdelku, sistem za shranjevanje in upravljanje tehnične dokumentacije.
- **MES (Manufacturing Execution System)** je sistem za nadzor proizvodnih procesov v realnem času.
- **PLM (Product Lifecycle Management)** - upravljanje življenjskega **cikla** elementa projekta, ki združuje PDM, CAPP, CAM in druge sisteme za popoln nadzor izdelka od razvoja do odstranitve.

Ti in številni drugi sistemi, ki vključujejo različne programske rešitve, so postali sestavni del sodobne gradbene industrije (slika 1.2-4). Bistvo takšnih sistemov so specializirane podatkovne zbirke z intuitivnimi vmesniki, ki zagotavljajo učinkovit vnos, obdelavo in analizo informacij v vseh fazah projektiranja in gradnje. Medsebojno povezovanje digitalnih orodij ne priomore le k optimizaciji delovnih procesov, temveč tudi bistveno izboljša natančnost odločitev, kar pozitivno vpliva na čas in kakovost izvedbe projekta.

Vendar v polovici primerov ni integracije. Po statističnih podatkih je le vsaka druga aplikacija ali sistem integriran z drugimi rešitvami [17]. To kaže na nenehno razdrobljenost digitalnega okolja in poudarja potrebo po razvoju odprtih standardov in enotnih vmesnikov, da se zagotovi celovita izmenjava informacij znotraj gradbenega projekta.

Eden od glavnih izzivov pri integraciji za sodobna podjetja ostaja visoka kompleksnost digitalnih sistemov in zahteve po usposobljenosti uporabnikov, ki so potrebne za učinkovito iskanje in interpretacijo informacij. Za podporo vsakemu sistemu, uvedenemu v podjetju, se oblikuje skupina strokovnjakov, ki jo vodi ključni vodja (slika 1.2-2).

Ključni upravljavec sistema ima odločilno vlogo pri pravilnem usmerjanju pretoka podatkov in je odgovoren za kakovost končnih informacij, tako kot so bili prvi upravljavci pred tisočletji odgovorni za številke, zapisane na papirusu ali glinenih tablicah.

Da bi raznovrstne informacijske tokove spremenili v orodje za upravljanje, je bistvena sposobnost sistematičnega povezovanja in upravljanja podatkov. V tej arhitekturi morajo menedžerji delovati kot elementi enotnega omrežja - kot micelij, ki povezuje posamezne dele podjetja v skladen živ organizem, ki se lahko prilagaja in razvija.

## Korporativni micelij: kako so podatki povezani s poslovnimi procesi

Postopek vključevanja podatkov v aplikacije in zbirke podatkov temelji na združevanju informacij iz različnih virov, vključno z različnimi oddelki in strokovnjaki (slika 1.2-4). Strokovnjaki poiščejo ustreerne podatke, jih obdelajo in prenesejo v svoje sisteme in aplikacije za nadaljnjo uporabo.

Vsek sistem podjetja, sestavljen iz niza orodij, tehnologij in podatkovnih zbirk, je drevo znanja, ki korenini v tleh zgodovinskih podatkov in raste, da bi obrodilo nove sadove v obliki končnih rešitev: dokumentov, izračunov, tabel, grafov in nadzornih plošč (slika 1.2-5). Sistemi v podjetju kot drevesa v dolženem delu gozda medsebojno delujejo in komunicirajo ter tvorijo zapleten in dobro strukturiran sistem, ki ga podpirajo in upravljajo strokovni vodje.

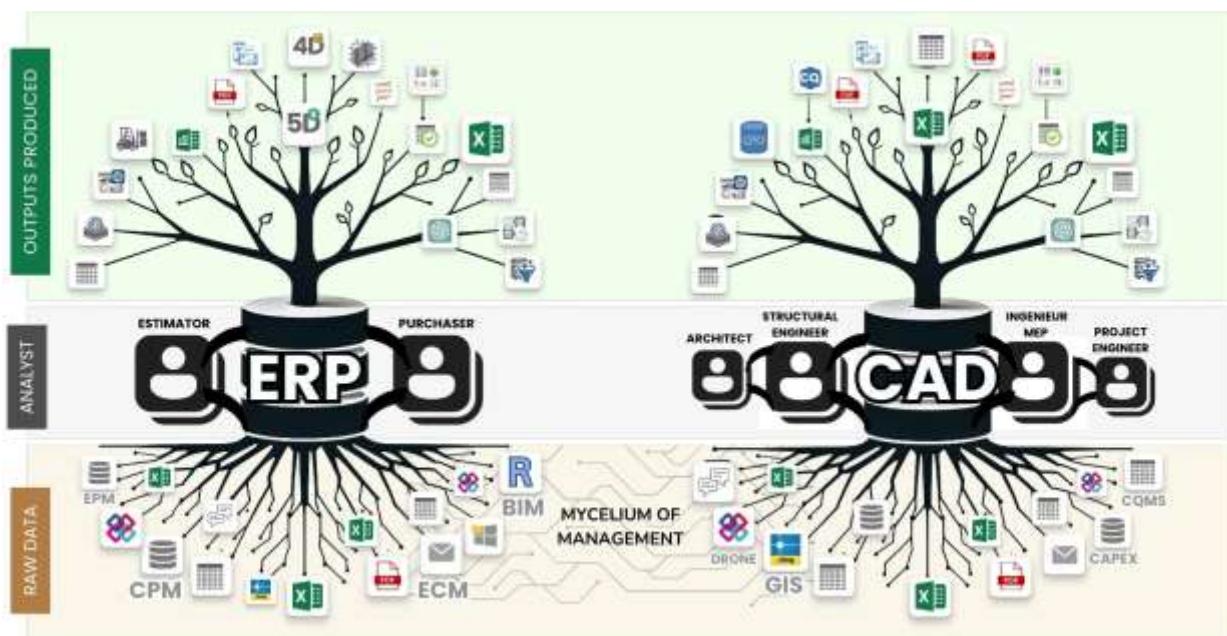
Sistem za pridobivanje in prenos informacij v podjetju deluje kot zapleteno gozdno omrežje, sestavljeno iz dreves (sistemov) in gob (upraviteljev), ki delujejo kot prevodniki in predelovalci ter skrbijo za prenos in pretok informacij v prave sisteme. To pomaga ohranjati zdrav in učinkovit pretok ter distribucijo podatkov v podjetju.

Strokovnjaki kot korenine absorbirajo surove podatke v začetnih fazah projekta in jih spremenijo v hranilne snovi za ekosistem podjetja. Sistemi za upravljanje podatkov in vsebin (slika 1.2-4 - ERP, CPM, BIM itd.) delujejo kot zmogljive informacijske avtoceste, po katerih to znanje kroži po vseh ravneh podjetja.

Tako kot v naravi, kjer ima vsak element ekosistema svojo vlogo, tudi v poslovniem okolju podjetja vsak udeleženec procesa - od inženirja do analitika - prispeva k rasti in plodnosti informacijskega okolja. Ta sistemski "podatkovna drevesa" (slika 1.2-5) niso le mehanizmi za zbiranje informacij, temveč konkurenčna prednost, ki zagotavlja trajnost podjetja.

Gozdni ekosistemi so presenetljivo natančen odraz organiziranosti digitalnih podjetniških struktur. Podobno kot večstopenjska struktura gozda - od podrasti do krošenj dreves - tudi upravljanje podjetij dodeljuje naloge ravnom odgovornosti in funkcionalnim oddelkom.

Globoke in razvezjane korenine dreves zagotavljajo odpornost in dostop do hranil. Podobno trdna organizacijska struktura in stabilni procesi za delo s kakovostnimi podatki podpirajo celoten informacijski ekosistem podjetja ter prispevajo k njegovi trajnostni rasti in razvoju tudi v obdobjih nestabilnosti in kriz na trgu (z močnim vetrom).



Slika 1.2-5 Povezovanje podatkov prek različnih sistemov je kot micelij, ki povezuje vodje in strokovnjake v enotno informacijsko omrežje.

Sodobno razumevanje obsega v poslovanju se je razvilo. Danes vrednosti podjetja ne določa le njegov vidni del - "krone" v obliki končnih dokumentov in poročil - temveč tudi globina "koreninskega sistema" kakovostno zbranih in sistematicno obdelanih podatkov. Več kot je mogoče zbrati in obdelati podatkov, večja je vrednost podjetja. Podjetja, ki metodično kopijo "kompost" že obdelanih podatkov in so iz njih sposobna izluščiti uporabna spoznanja, pridobijo strateško prednost.

Zgodovinske informacije postajajo nova vrsta kapitala, ki omogoča rast, optimizacijo procesov in konkurenčno prednost. V svetu, ki temelji na podatkih, ne zmaga tisti, ki ima več, temveč tisti, ki več ve.

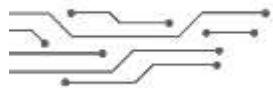
Za gradbeno industrijo to pomeni prehod na upravljanje projektov v realnem času, kjer bodo vsi procesi - od načrtovanja in naročanja do usklajevanja izvajalcev - temeljili na ustreznih, dnevno posodobljenih podatkih. Integracija informacij iz različnih virov (sistemi ERP, modeli CAD, senzorji IoT na gradbiščih, RFID) bo omogočila natančnejše napovedi, hitro odzivanje na spremembe in prečevanje zamud zaradi pomanjkanja posodobljenih podatkov.

V skladu z dokumentom Data-Driven Enterprise 2025 družbe McKinsey & Company (McKinsey & Company®, 2022 [18]) se bodo uspešna podjetja prihodnosti zanašala na podatke v vseh ključnih vidikih svojega delovanja, od strateških odločitev do operativnih interakcij.

Podatki ne bodo več le analitično orodje in bodo postali sestavni del vseh poslovnih procesov, saj bodo zagotavljali preglednost, nadzor in avtomatizacijo upravljanja. Na podatkih temelječa poteza pod bo organizacijam omogočila, da zmanjšajo vpliv človeškega dejavnika, zmanjšajo operativna tveganja ter povečajo preglednost in učinkovitost odločanja.

21. stoletje obrača gospodarsko paradigmo na glavo: medtem ko se je nafta včasih imenovala "črno

zlato" zaradi svoje sposobnosti poganjanja strojev in prevoza, danes zgodovinski podatki, stisnjeni pod časovnim pritiskom, postajajo nov strateški vir, ki ne poganja strojev, temveč algoritme za odločanje, ki bodo vodili poslovanje.



## POGLAVJE 1.3.

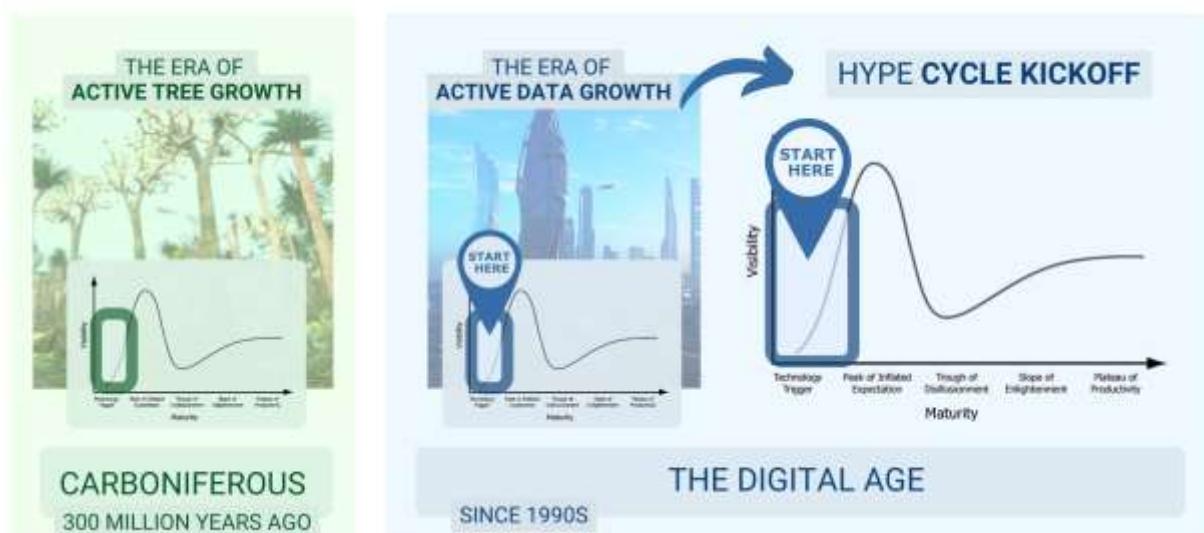
# DIGITALNA REVOLUCIJA IN EKSPLOZIJA PODATKOV

### Začetek podatkovnega razcveta kot evolucijski val

Gradbena industrija doživlja informacijsko eksplozijo brez primere. Če si poslovanje predstavljamo kot drevo znanja (slika 1.2-5), ki se napaja s podatki, lahko sedanjo stopnjo digitalizacije primerjamo s hitro rastjo vegetacije v karbonskem obdobju, v katerem se je zemeljska biosfera preoblikovala s hitrim kopičenjem biomase (slika 1.3-1).

Zaradi globalne digitalizacije se količina informacij v gradbeništvu vsako leto podvoji. Sodobna tehnologija omogoča zbiranje podatkov v ozadju, njihovo analizo v realnem času in uporabo v obsegu, ki se je še pred kratkim zdel nemogoč.

V skladu z Moorovim zakonom, ki ga je oblikoval Gordon Moore (soustanovitelj podjetja Intel®), se gostota in zapletenost integriranih vezij ter količina obdelanih in shranjenih podatkov podvojita približno vsaki dve leti [19].



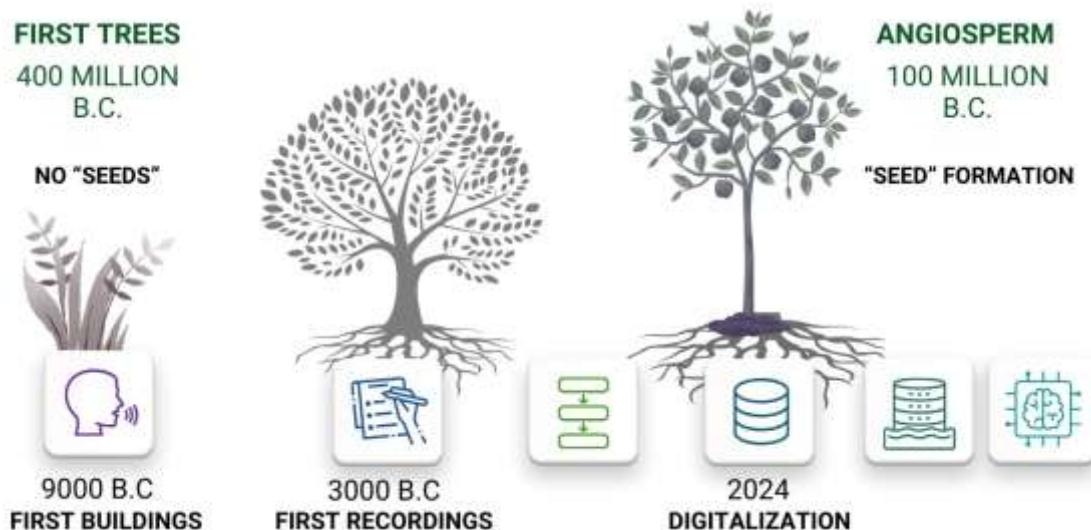
Slika 1.3-1 Z začetkom digitalizacije je prišlo do eksponentne rasti podatkov, podobno kot se je v dobi premoga razmahnila vegetacija.

Medtem ko starodavne megalitske strukture, kot je Göbekli Tepe (Turčija), niso pustile dokumentiranega znanja, primerenega za ponovno uporabo, danes digitalne tehnologije omogočajo zbiranje in ponovno uporabo informacij. To lahko primerjamo z evolucijskim prehodom od spor v semenske rastline (angiosperme): pojav semena je omogočil široko razširjenost življenja na planetu. (Slika 1.3-2).

Podobno podatki iz preteklih projektov postanejo nekakšna "digitalna semena" - nosilci znanja DNK, ki jih je mogoče razširiti in uporabiti pri novih projektih in producēti. Pojav sodobnih orodij umetne inteligence - strojnega učenja in velikih jezikovnih modelov (LLM), kot so ChatGPT, LLaMa, Mistral,

Claude, DeepSeek, QWEN, Grok - omogoča samodejno pridobivanje, razlago in uporabo podatkov v novih kontekstih

Tako kot so semena revolucionarno pripomogla k širjenju življenja na sprva neživem planetu, postajajo "podatkovna semena" osnova za samodejno nastajanje novih informacijskih struktur in znanja, kar digitalnim ekosistemom omogoča, da se neodvisno razvijajo in prilagajajo spremenljajočim se zahtevam uporabnikov.



Slika 1.3-2 Digitalna "podatkovna semena" imajo enako evolucijsko vlogo kot angiosperme, cvetoče rastline, ki so preoblikovale zemeljski ekosistem.

Stojimo na pragu novega obdobja v gradbeništvu, kjer eksplozija podatkov in aktivno širjenje "podatkovnih semen" - strukturiranih informacij iz preteklih in tekočih projektov - oblikujeta temelje digitalne prihodnosti panoge. Njihovo "opraševanje" s pomočjo velikih podatkovnih jezikovnih modelov (LLM) nam omogoča, da digitalnih sprememb ne le opazujemo, temveč aktivno sodelujemo pri ustvarjanju samoučečih se, prilagodljivih ekosistemov. To ni evolucija - to je digitalna revolucija, v kateri podatki postajajo glavni gradnik nove resničnosti

*Količina podatkov v gradbeništvu se močno povečuje zaradi informacij iz različnih področij v celotnem življenjskem ciklu gradbenih projektov. To ogromno kopiranje podatkov je gradbeno industrijo potisnilo v dobo velikih podatkov [20].*

- Prof. Hang Yang, Oddelek za gradbeništvo in arhitekturo, Tehnološka univerza Wuhan, Wuhan,

Rast podatkov v informacijski dobi spominja na evolucijske procese v naravi: tako kot je razvoj gozdov spremenil starodavno pokrajino planeta, sedanja informacijska eksplozija spreminja pokrajino celotne gradbene industrije.

## Količina podatkov, ki jih ustvari sodobno podjetje

V zadnjih dveh letih je bilo ustvarjenih 90% vseh obstoječih podatkov na svetu [21]. Od leta 2023 vsak človek, vključno s strokovnjaki v gradbeništvu, ustvari približno 1,7 megabajta podatkov na sekundo [22], skupna količina podatkov na svetu pa bo leta 2023 dosegla 64 zetabajtov, do leta 2025 pa naj bi presegla 180 zetabajtov ali  $180 \times 10^{15}$  megabajtov [23].

Ta informacijska eksplozija ima zgodovinski precedens - izum tiskarskega stroja Johanna Gutenberga v 15. stoletju. Le petdeset let po njegovi uvedbi se je število knjig v Evropi podvojilo: v nekaj desetletjih je bilo natisnjenih toliko knjig, kot jih je bilo ročno ustvarjenih v prejšnjih 1 200 letih [24]. Danes smo priča še hitrejši rasti: količina podatkov na svetu se podvoji vsaka tri leta.

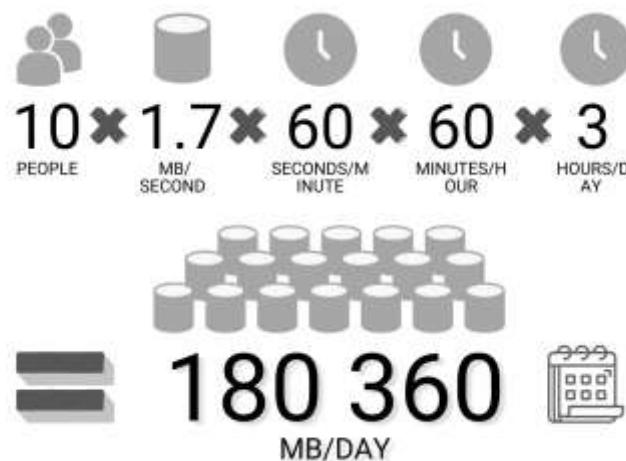
Glede na sedanj hitrost rasti podatkov lahko gradbena industrija v naslednjih nekaj desetletjih ustvari toliko informacij, kot jih je zbrala v vsej svoji dosedanji zgodovini



Slika 1.3-3 Vsakodnevno shranjevanje podatkov s strani vsakega zaposlenega na strežnikih podjetja prispeva k nenehnemu povečevanju količine podatkov.

V današnjem svetu gradbeništva tudi majhna podjetja dnevno ustvarjajo ogromne količine večpredstavnostnih informacij, digitalni odtis celo majhnega gradbenega podjetja pa lahko doseže več deset gigabajtov na dan - od modelov in risb do fotografskih posnetkov in senzorjev na gradbišču. Če predpostavimo, da vsak tehnik v povprečju ustvari približno 1,7 MB podatkov na sekundo, to ustreza približno 146 GB na dan ali 53 TB na leto (slika 1.3-3).

Če ekipa desetih ljudi dnevno aktivno dela le tri ure, skupna količina informacij, ustvarjenih na dan, doseže 180 gigabajtov (slika 1.3-4).



**Slika 1.3-4 Podjetje z 10 zaposlenimi ustvari približno 50-200 gigabajtov podatkov na dan.**

Ob predpostavki, da je 30% delovnih podatkov novih (ostali so prepisani ali izbrisani), lahko podjetje z 10 zaposlenimi ustvari več sto gigabajtov novih podatkov na mesec (dejanske številke so odvisne od vrste poslovanja podjetja)

Jasno je torej, da ne ustvarjamo le vedno več podatkov, temveč se soočamo z vse večjo potrebo po njihovem učinkovitem upravljanju, shranjevanju in dolgoročni razpoložljivosti. In medtem ko so prej podatki lahko brezplačno "ležali" na lokalnih strežnikih, v okviru digitalne preobrazbe vse več podjetij kot osnovo svoje informacijske infrastrukture začenja uporabljati rešitve v oblaku.

### Stroški shranjevanja podatkov: ekonomski vidik

V zadnjih letih vse več podjetij prenaša shranjevanje podatkov v storitve v oblaku. Če na primer podjetje polovico svojih podatkov gosti v oblaku po povprečni ceni 0,015 USD na gigabajt na mesec, se lahko njegovi stroški shranjevanja vsak mesec povečajo za 10-50 USD [25].

Za majhno podjetje s tipičnimi vzorci ustvarjanja podatkov lahko stroški shranjevanja v oblaku v nekaj letih znašajo od nekaj sto do več kot tisoč dolarjev na mesec (slika 1.3-5), kar lahko predstavlja znatno finančno breme.

Glede na Forresterjevo študijo "Enterprises Outsource Data Storage as Complexity Grows" [26], v kateri je sodelovalo 214 odločevalcev na področju tehnološke infrastrukture. [26], ki je zajela 214 odločevalcev na področju tehnološke infrastrukture, več kot tretjina organizacij za obvladovanje vse večje količine in kompleksnosti podatkovnih operacij oddaja hrambo v zunanje izvajanje, pri čemer se skoraj dve tretjini podjetij raje odloča za naročniški model.



Slika 1.3-5 Prenos podatkov v oblak lahko poveča mesečne stroške shranjevanja za do 2.000 USD, tudi v podjetju z le 10 zaposlenimi.

Razmere dodatno otežuje pospešeno uvajanje tehnologij v oblaku, kot so sistemi CAD (BIM), CAFM, PMIS in ERP, ki dodatno povečujejo stroške shranjevanja in obdelave podatkov. Zato so podjetja prisiljena iskati načine za optimizacijo stroškov in zmanjšanje odvisnosti od ponudnikov storitev v oblaku.

Od leta 2023 se z aktivnim razvojem velikih jezikovnih modelov (LLM) pristopi k shranjevanju podatkov začenjajo spremenjati. Vedno več podjetij razmišlja o tem, da bi ponovno prevzela nadzor nad svojimi podatki, saj postaja obdelava informacij na lastnih strežnikih varnejša in donosnejša.

V tem okviru je v ospredju trend opuščanja shranjevanja in obdelave le potrebnih podatkov v oblaku v korist lokalne uporabe rešitev LLM in AI v podjetjih. Kot je enem od intervjujev [27] poudaril izvršni direktor družbe Microsoft, bodo agenti umetne inteligence namesto na več ločenih aplikacij ali rešitev SaaS v oblaku za izvajanje različnih nalog upravljalci procese v podatkovnih zbirkah in avtomatizirali funkcije različnih sistemov.

[...] stari pristop k temu vprašanju [obdelave podatkov] je bil: če se spomnите, kako so različne poslovne aplikacije reševale integracijo, so uporabljale konektorje. Podjetja so prodajala licence za te konektorje in poslovni model se je oblikoval okoli tega. SAP [ERP] je eden od klasičnih primerov: do podatkov SAP ste lahko dostopali le, če ste imeli pravi priključek. Zato se mi zdi, da se bo nekaj podobnega pojavilo tudi v primeru interakcije agentov [AI] [...]. Pristop, vsaj ta, ki ga uporabljammo mi, je: mislim, da se bo koncept obstoja poslovnih aplikacij v dobi agentov [umetne inteligence] verjetno zrušil. Če namreč pomislite, so to v bistvu podatkovne zbirke s kopico poslovne logike

- Satya Nadella, izvršni direktor Microsofta, intervju s kanalom BG2, 2024. [28]

V tej paradigmi pristop LLM, ki temelji na podatkih, presega klasične sisteme. Umetna inteligenco postane posrednik med uporabnikom in podatki (slika 2.2-3, slika 2.2-4), kar odpravi potrebo po več vmesnikih in poveča učinkovitost poslovnih procesov. Več o tem pristopu k delu s podatki bomo povedali

v poglavju "Spreminjanje kaosa v red in zmanjševanje kompleksnosti".

Medtem ko se arhitektura prihodnosti še oblikuje, se podjetja že soočajo s posledicami preteklih odločitev. Množična digitalizacija v zadnjih desetletjih, ki jo spremljata uvedba različnih sistemov in nena-dzorovano kopiranje podatkov, je privedla do nove težave - preobremenjenosti z informacijami.

## Meje kopiranja podatkov: od mase do pomena

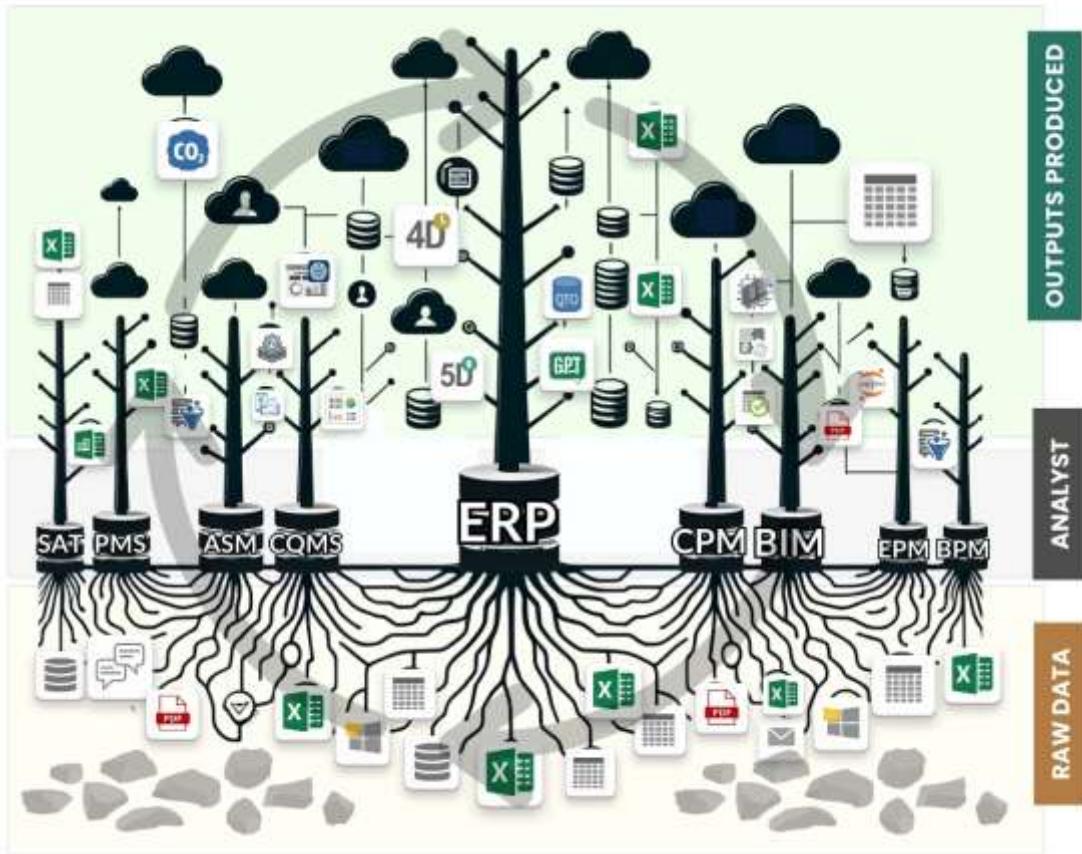
Sodobni sistemi v podjetjih se uspešno razvijajo in delujejo v okviru nadzorovane rasti, ko sta količina podatkov in število aplikacij v ravnovesju z zmožnostmi oddelkov IT in vodij. Vendar pa je digitalizacija v zadnjih desetletjih povzročila neobvladljivo povečanje količine in kompleksnosti podatkov, kar je povzročilo učinek prenasičenosti v informacijskem ekosistemu podjetij.

Danes so strežniki in shrambe izpostavljeni neverjetnemu navalu neobdelanih in večformatnih informacij, ki nimajo časa, da bi se spremenile v kompost, in hitro postanejo nepomembne. Omejeni viri podjetja ne morejo obvladati te poplave, zato se podatki kopijo v izoliranih silosih (tako imenovani "silosi"), ki zahtevajo ročno obdelavo, da bi iz njih pridobili uporabne informacije.

Zato so sodobni sistemi upravljanja podjetij pogosto preobremenjeni z informacijami, tako kot gozd, porasel z bršljanom in prekrit s plesnijo. Namesto da bi v osrčju podjetniškega ekosistema hranili informacijski humus, nastajajo izolirana območja podatkov različnih formatov, kar neizogibno vodi k zmanjšanju splošne učinkovitosti poslovnih procesov.

Dolgemu obdobju eksponentne rasti podatkov v zadnjih 40 letih bo neizogibno sledila faza nasičenja in nato ohlajanja. Ko bo pomnilnik dosegel svoje meje, bo prišlo do kvalitativnega premika: podatki ne bodo več le objekt za shranjevanje, temveč strateški vir.

Z razvojem umetne inteligence in strojnega učenja imajo podjetja priložnost, da zmanjšajo stroške obdelave informacij in preidejo od kvantitativne rasti h kvalitativni uporabi podatkov. V naslednjem desetletju se bo morala gradbena industrija preusmeriti od ustvarjanja vedno več podatkov k zagotavljanju njihove strukture, celovitosti in analitične vrednosti.



Slika 1.3-6 Izolirani viri podatkov preprečujejo izmenjavo informacij med podatkovnimi sistemmi.

Glavna vrednost ni več v količini informacij, temveč v zmožnosti, da jih samodejno interpretiramo in spremenimo v uporabno znanje, uporabno za sprejemanje upravljalnih odločitev. Da bi podatki postali resnično uporabni, jih je treba ustrezno upravljati: zbirati, preverjati, strukturirati, shranjevati in analizirati v okviru posebnih poslovnih nalog.

Proces podatkovne analitike v podjetju je podoben ciklu življenja in propadanja dreves v gozdu ter nastajanju novih mladih in močnih dreves: zrela drevesa odmrejo, propadejo in postanejo gojišče za novo rast. Zaključeni in dokončani procesi, ko so končani, postanejo del informacijskega ekosistema podjetja in sčasoma postanejo informacijski humus, ki poganja prihodnjo rast novih sistemov in podatkov.

V praksi pa se ta krog pogosto prekine. Namesto organske prenove se oblikuje večplastni kaos, podoben geološkim plastem, kjer se novi sistemi nalagajo na stare brez globljega povezovanja in strukturiranja. Posledično se pojavijo različni informacijski "silosi", ki ovirajo kroženje znanja in otežujejo upravljanje podatkov.

## Naslednji koraki: od teorije podatkov do praktičnih sprememb

Razvoj podatkov v gradbeništvu je pot od glinenih tablic do sodobnih modularnih platform. Današnji izziv ni zbiranje informacij, temveč ustvarjanje okvira, ki razpršene in raznolike podatke spreminja v strateški vir. Ne glede na to, ali je vaša vloga vodja podjetja ali inženir, bo razumevanje vrednosti podatkov in načina dela z njimi v prihodnosti ključna veščina.

Za povzetek tega dela velja izpostaviti glavne praktične korake, ki vam bodo pomagali pri uporabi obravnavanih pristopov pri vsakodnevnih nalogah:

■ Izvedite osebno revizijo informacijskih tokov.

- Sestavite seznam vseh sistemov in aplikacij, s katerimi vsakodnevno delate.
- označite, kje porabite največ časa za iskanje ali ponovno preverjanje podatkov.
- Določite ključne vire informacij.
- Analizirajte trenutno okolje aplikacij glede odvečnih in podvojenih funkcij.

■ Prizadevajte si za prehod skozi procese po stopnjah analitične zrelosti.

- Naloge začnite z opisno analitiko (kaj se je zgodilo?)
- Postopoma uvedite diagnostiko (zakaj se je to zgodilo?)
- Razmislite o tem, kako lahko v procesih preidete na napovedno (kaj se bo zgodilo?) in predpisno (kaj storiti?) analitiko.

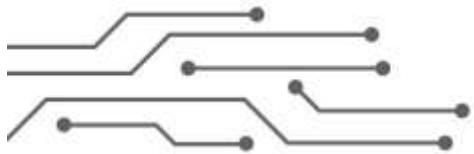
■ Začnite strukturirati svoje delovne podatke

- Uvedite enoten sistem poimenovanja datotek in map, ki jih pogosto uporabljate pri svojem delu.
- Ustvarjanje predlog za pogosto uporabljene dokumente in poročila
- redno arhiviranje zaključenih projektov z jasno strukturo

Tudi če ne morete spremeniti celotne informacijske infrastrukture v svoji ekipi ali podjetju, začnite z lastnimi procesi in majhnimi izboljšavami pri vsakdanjem delu. Ne pozabite, da prava vrednost podatkov ni v njihovi količini, temveč v zmožnosti, da iz njih pridobimo uporabne vpoglede. Tudi majhni, a ustrezno strukturirani in analizirani podatkovni nizi imajo lahko pomemben vpliv, če so vključeni v procese odločanja.

V naslednjih delih knjige se bomo posvetili posebnim metodam in orodjem za delo s podatki, preučili načine za preoblikovanje nestrukturiranih informacij v strukturirane nize, raziskali tehnologije za avtomatizacijo analitike in podrobno opisali, kako vzpostaviti učinkovit analitični ekosistem v gradbenem podjetju.





## II DEL

### **KAKO SE GRADBENIŠTVO UTAPLJA V PODATKOVNEM KAOSU**

Drugi del je posvečen kritični analizi izzivov, s katerimi se soočajo gradbena podjetja pri ravnjanju z vse večjimi količinami podatkov. Podrobno so obravnavane posledice razdrobljenosti informacij in pojav "podatkov v silosih", ki ovirajo učinkovito odločanje. Preučeni so problemi pristopa HiPPO (Highest Paid Person's Opinion) in njegov vpliv na kakovost vodstvenih odločitev v gradbenih projektih. Ocenjen je vpliv dinamičnih poslovnih procesov in njihove vse večje kompleksnosti na pretok informacij in operativno učinkovitost. Navedeni so konkretni primeri, kako prevelika kompleksnost sistemov povečuje stroške in zmanjšuje prilagodljivost organizacij. Posebna pozornost je namenjena omejitvam, ki jih povzročajo lastniški formati, in možnostim uporabe odprtih standardov v gradbeništvu. Predstavljen je koncept prehoda na programske ekosisteme, ki temeljijo na umetni inteligenci in LLM, ki zmanjšujejo pretirano zapletenost in tehnične ovire.

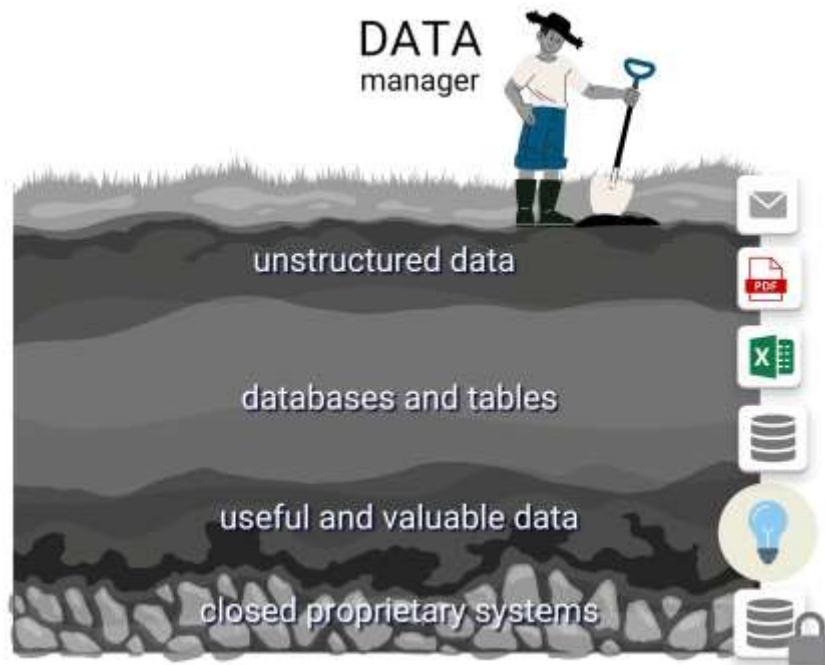
## POGLAVJE 2.1.

# RAZDROBLJENOST PODATKOV IN SILOSI.

### Več kot je orodij, bolj učinkovito je podjetje?

Na prvi pogled se zdi, da več digitalnih orodij vodi k večji učinkovitosti. Vendar v praksi ni tako. Z vsako novo rešitvijo, pa naj bo to storitev v oblaku, starejši sistem ali še eno poročilo v Excelu, podjetje doda še eno plast v svoj digitalni prostor - plast, ki pogosto ni integrirana s preostalimi (slika 2.1-1).

Podatke lahko primerjamo s premogom ali nafto: zbirajo se več let, saj so zgoščeni pod plastmi kaosa, napak, nestrukturiranih procesov in pozabljenih formatov. Da bi iz njih pridobila resnično uporabne informacije, se morajo podjetja dobesedno prebijati skozi plasti zastarelih rešitev in digitalnega hrupa.



Slika 2.1-1 Različni podatki tvorijo razdrobljene plasti - tudi "zlata" spoznanja se izgubijo v geoloških kamninah sistemskih kompleksnosti.

Vsaka nova aplikacija za seboj pusti sled: datoteko, tabelo ali celoten izoliran "silos" v strežniku. Ena plast je glina (zastareli in pozabljeni podatki), druga pesek (različne tabele in poročila), tretja pa granit (zaprti lastniški formati, ki jih ni mogoče integrirati). Sčasoma digitalno okolje podjetja vse bolj spominja na rezervoar nenadzorovanega kopiranja informacij, kjer se vrednost izgublja globoko v strežnikih podjetja.

Z vsakim novim projektom in vsakim novim sistemom postaja kompleksnejša ne le infrastruktura, temveč tudi pot do uporabnih kakovostnih podatkov. Da pridemo do dragocene "skale", je treba podatke temeljito očistiti, jih strukturirati, "razdrobiti", združiti v smiselne dele ter z analitiko in

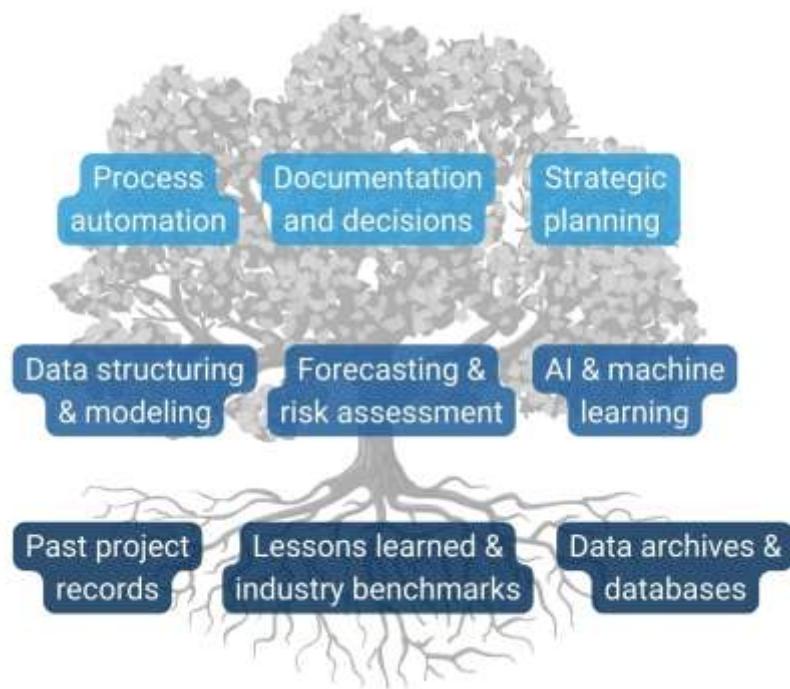
modeliranjem podatkov pridobiti strateško pomembna spoznanja.

*Podatki so dragocena stvar, ki bo trajala dlje kot sami sistemi [ki obdelujejo podatke] 29].*

- Tim Berners-Lee, oče svetovnega spleta in avtor prve spletne strani

Preden lahko podatki postanejo "dragocena stvar" in zanesljiva podlaga za odločanje, jih je treba skrbno pripraviti. Prav ustrezna predobdelava je tista, ki razpršene podatke spremeni v strukturirano izkušnjo, uporaben informacijski humus, ki nato postane orodje za napovedovanje in optimizacijo.

Obstaja napačno prepričanje, da za začetek analiziranja potrebujete popolnoma čiste podatke, vendar je v praksi sposobnost dela z umazanimi podatki bistven del procesa.



Slika 2.1-2 Podatki so koreninski sistem in temelj poslovanja, ki temelji na procesih odločanja.

Ker tehnologija še naprej napreduje, mora tudi vaše podjetje napredovati in se naučiti, kako iz podatkov ustvariti vrednost. Tako kot naftna in premogovniška podjetja gradijo infrastrukturo za pridobivanje mineralov, se morajo tudi podjetja naučiti, kako upravljati tok novih informacij na svojih strežnikih in iz neuporabljenih, neformatiranih in zastarelih podatkov pridobiti dragocene informacije ter jih spremeniti v strateški vir.

Ustvarjanje polj (podatkovna skladišča) je prvi korak. Tudi najzmožljivejša orodja ne rešijo problema izolacije podatkov in podatkov različnih formatov, če podjetja še naprej delujejo v ločenih sistemih. Če podatki obstajajo ločeno drug od drugega, ne da bi se križali in izmenjevali informacije, se podjetja soočajo z učinkom "podatkovnih silosov". Namesto enotne, dosledne infrastrukture so podjetja prisiljena porabljati sredstva za združevanje in sinhronizacijo podatkov.

## Podatkovni silosi in njihov vpliv na uspešnost podjetja

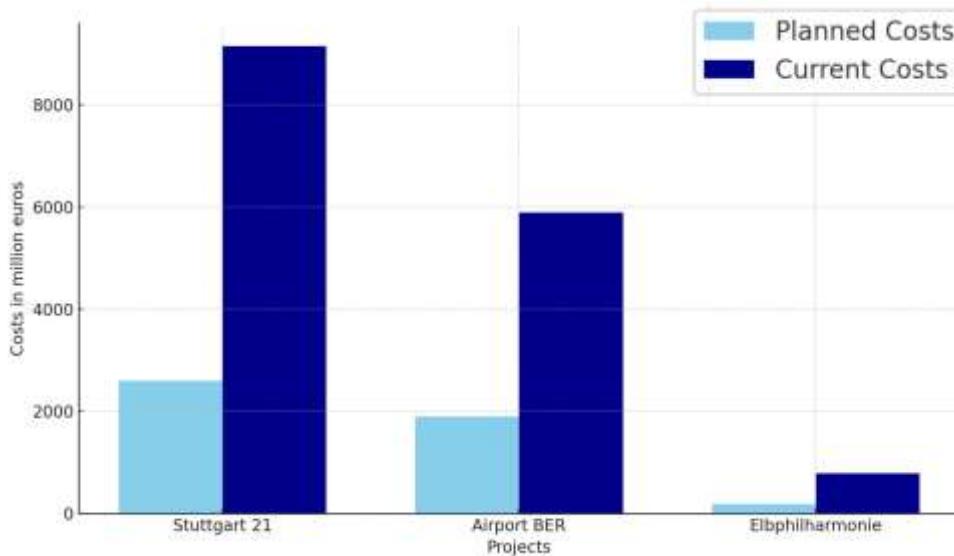
Predstavljajte si, da gradite stanovanjsko sosesko, vendar ima vsaka ekipa svoj projekt. Nekatere gradijo zidove, druge polagajo komunikacije, tretje polagajo ceste, ne da bi se med seboj posvetovale. Zaradi tega se cevi ne ujemajo z odprtinami v zidovih, dvigalni jaški ne ustrezajo nadstropjem, ceste pa je treba razstaviti in na novo položiti.

Ta situacija ni le hipotetičen scenarij, temveč realnost številnih sodobnih gradbenih projektov. Zaradi velikega števila generalnih in podizvajalskih izvajalcev, ki delajo z različnimi sistemi in nimajo enotnega koordinacijskega centra, se postopek spremeni v niz neskončnih odobritev, predelav in konfliktov. Vse to vodi v velike zamude in večkratne stroške projekta.

Klasičen primer na gradbišču je preprost: opaž je pripravljen, vendar dobava armature ni prispela pravočasno. Pri preverjanju informacij v različnih sistemih je komunikacija v grobem naslednja:

- ❷ **Mojster** na gradbišču 20. dne piše vodji projekta: "*Končali smo z montažo opaža, kje je armatura?*"
- ❷ **Vodja projekta** (PMIS) oddelku za nabavo: - "*Opaži so pripravljeni. V mojem sistemu [PMIS] piše, da naj bi armature prispele 18. septembra. Kje so armaturne palice?*"
- ❷ **Strokovnjak za oskrbovalno verigo** (ERP): - "*Naš sistem ERP pravi, da bo dostava 25.*".
- ❷ **podatkovni inženir** ali oddelek IT (odgovoren za integracije): - V PMIS je datum 18., v ERP pa 25. Med ERP in PMIS ni povezave OrderID, zato podatki niso sinhronizirani. To je tipičen primer informacijske vrzeli.
- ❷ **Vodja projekta** generalnemu **direktorju** - "*Dobava napeljav zamuja, gradbišče stoji in ni jasno, kdo je odgovoren*".

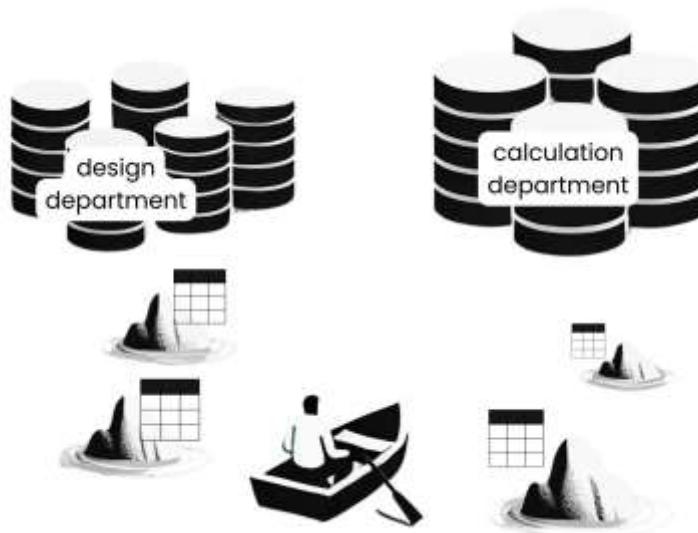
Vzrok incidenta je bila izolacija podatkov v različnih sistemih. S povezovanjem in poenotenjem podatkovnih virov, vzpostavljivjo enotnega skladišča informacij in avtomatizacijo s pomočjo orodij ETL (Apache NiFi, Airflow ali n8n) je mogoče odpraviti silose med sistemi. Te in druge metode in orodja bodo podrobno obravnavane v nadaljnjih delih knjige.



**Slika 2.1-3 Primerjava načrtovanih in dejanskih stroškov večjih infrastrukturnih projektov v Nemčiji.**

Enako se dogaja s podjetniškimi sistemi: najprej se oblikujejo izolirane rešitve, nato pa je treba za njihovo integracijo in uskladitev porabiti ogromna proračunska sredstva. Če bi bili podatkovni in komunikacijski modeli premišljeni od začetka, povezovanje sploh ne bi bilo potrebno. Posamezni podatki v digitalnem svetu ustvarjajo kaos, podobno kot neusklajen gradbeni proces.

Glede na študijo družbe KPMG "Cue construction 4.0: Time to make or break" iz leta 2023 je 36% podjetij učinkovito deli podatke med oddelki, 61% pa se jih sooča z resnimi težavami zaradi izoliranih podatkovnih "silosov" [30].



**Slika 2.1-4 Leta težko zbranih podatkov se kopijo v izoliranih skladniščnih "silosih" in tvegajo, da ne bodo nikoli uporabljeni.**

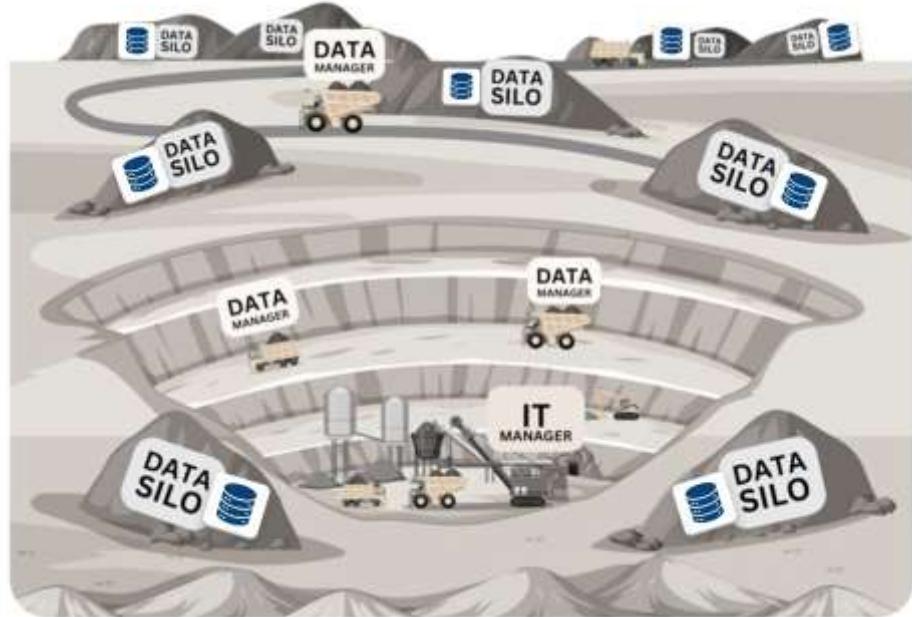
Podatki podjetja so shranjeni v izoliranih sistemih, kot posamezna drevesa, raztresena po pokrajini. Vsak od njih vsebuje dragocene informacije, vendar pomanjkanje povezav med njimi preprečuje vzpostavitev enotnega, medsebojno povezanega ekosistema. Ta ločenost ovira pretok podatkov in omejuje sposobnost organizacije, da vidi celotno sliko. Povezovanje teh silosov je izjemno dolg in zapleten proces gojenja gobjega micelija na ravni upravljanja, da bi se naučili, kako prenesti posamezne informacije med sistemi.

Po študiji WEF iz leta 2016 je ena glavnih ovir za digitalno preobrazbo pomanjkanje skupnih podatkovnih standardov in razdrobljenost.

*Gradbena industrija je ena najbolj razdrobljenih na svetu in je odvisna od nemotenega sodelovanja vseh udeležencev v vrednostni verigi [5].*

- Svetovni gospodarski forum 2016: Oblikovanje prihodnosti

Oblikovalci, vodje, koordinatorji in razvijalci pogosto raje delajo samostojno in se izogibajo zapletom usklajevanja. To naravno nagnjenje vodi v oblikovanje informacijskih "silosov", v katerih so podatki izolirani v ločenih sistemih. Več kot je takih izoliranih sistemov, teže jih je povezati v celoto. Sčasoma vsak sistem dobi svojo podatkovno zbirko in specializiran podporni oddelek menedžerjev (slika 1.2-4), kar še dodatno otežuje povezovanje.



**Slika 2.1-5** Vsak sistem ponavadi ustvari svoj edinstven silos podatkov, ki jih je treba obdelati z ustreznimi orodji [31].

Začarani krog pri korporativnih sistemih je videti takole: podjetja vlagajo v zapletene rešitve, povezane s sistemom ISO, nato se soočajo z visokimi stroški za njihovo integracijo, razvijalci pa se zavedajo zapletenosti združevanja sistemov in raje delajo v svojih zaprtih ekosistemih. Vse to povečuje

razdrobljenost informacijskega okolja in otežuje prehod na nove rešitve (slika 2.1-5). Vodje na koncu kritizirajo podatkovne silose, vendar le redko analizirajo njihove vzroke in načine, kako jih preprečiti. Vodje se pritožujejo nad zastarelimi sistemi IT, vendar njihova zamenjava zahteva velike naložbe in le redko prinese pričakovane rezultate. Posledično se tudi pri poskusih odpravljanja težav razmere pogosto še poslabšajo.

Glavni razlog za to je dajanje prednosti aplikacijam pred podatki. Podjetja najprej razvijejo ločene sisteme ali kupijo gotove rešitve od prodajalcev, nato pa jih poskušajo poenotiti z ustvarjanjem podvojenih in nezdružljivih shramb in podatkovnih zbirk.

Za rešitev problema razdrobljenosti je potreben radikalno nov pristop - prednost podatkov pred aplikacijami. Podjetja morajo najprej razviti strategije upravljanja podatkov in podatkovne modele, nato pa zgraditi sisteme ali kupiti rešitve, ki delujejo z enotnim naborom informacij, namesto da bi ustvarjala nove ovire.

*Vstopamo v nov svet, v katerem so podatki morda pomembnejši od programske opreme.*

- Tim O'Reilly, izvršni direktor družbe O'Reilly Media, Inc.

Študija McKinsey Global Institute "Rethinking Construction: the path to improved productivity" (2016) kaže, da gradbeništvo zaostaja za drugimi sektorji pri digitalni preobrazbi [32]. V skladu s poročilom lahko uvedba avtomatiziranega upravljanja podatkov in digitalnih platform bistveno izboljša produktivnost in zmanjša izgube, povezane z nedoslednostjo procesov. To potrebo po digitalni preobrazbi poudarja tudi poročilo Egan (Združeno kraljestvo, 1998) [33], ki poudarja ključno vlogo integriranih procesov in sodelovalnega pristopa v gradbeništvu.

Medtem ko je bilo v zadnjih 10.000 letih glavna težava upravljavcev podatkov pomanjkanje podatkov, se s plazom podatkov in sistemov za upravljanje podatkov uporabniki in upravljavci soočajo s problemom preobilice podatkov, zaradi česar je težko najti pravno pravilne in kakovostne informacije.

Različni silosi podatkov neizogibno vodijo do resne težave z zmanjšano kakovostjo podatkov. Pri več neodvisnih sistemih lahko isti podatki obstajajo v različnih različicah, pogosto z nasprotujočimi si vrednostmi, kar povzroča dodatne težave uporabnikom, ki morajo ugotoviti, katere informacije so pomembne in zanesljive.

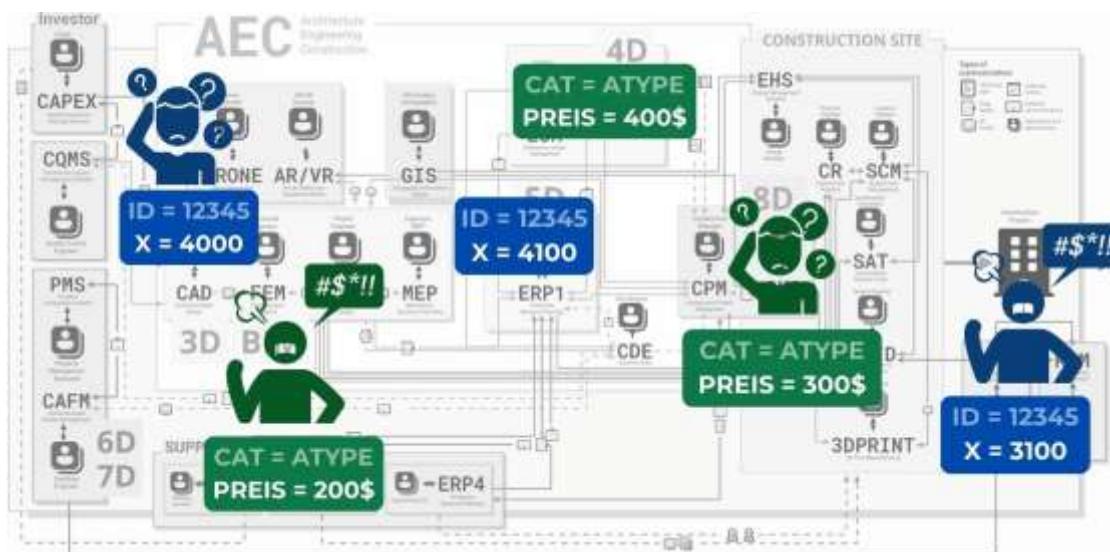
## **Podvajanje in pomanjkanje kakovosti podatkov kot posledica neenotnosti.**

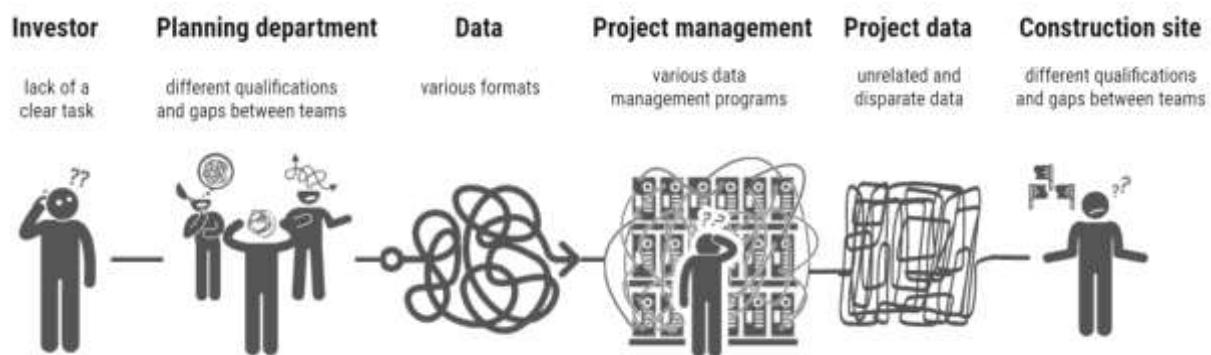
Zaradi problema podatkovnih silosov morajo vodje porabiti veliko časa za iskanje in usklajevanje podatkov. Da bi se zavarovala pred težavami s kakovostjo, podjetja oblikujejo zapletene strukture upravljanja informacij, v katerih je za iskanje, preverjanje in usklajevanje podatkov odgovorna vertikala

vodij. Vendar ta pristop le povečuje birokracijo in upočasnuje sprejemanje odločitev. Več kot je podatkov, težje jih je analizirati in razlagati, zlasti če ni enotnega standarda za njihovo shranjevanje in obdelavo.

Zaradi množice programskih aplikacij in sistemov, ki v zadnjem desetletju rastejo kot gobe po dežju, je problem silosov in neustrezne kakovosti podatkov za končne uporabnike vse pomembnejši. Iste podatke, vendar z različnimi vrednostmi, je zdaj mogoče najti v različnih sistemih in aplikacijah (slika 2.1-6). To povzroča težave končnim uporabnikom, ko poskušajo ugotoviti, katera različica podatkov je ustrezna in pravilna med številnimi, ki so na voljo. To vodi do napak pri analizah in nenazadnje pri sprejemanju odločitev.

Da bi se zavarovali pred težavami pri iskanju pravih podatkov, vodje podjetij ustvarijo večnivojsko birokracijo vodij preverjanja. Njihova naloga je, da lahko hitro poiščejo, preverijo in pošljejo zahtevane podatke v obliki preglednic in poročil ter se pri tem znajdejo v labirintu različnih sistemov.





**Slika 2.1-7 Kompleksnost sistemov in raznolikost podatkovnih formatov povzročata izgubo doslednosti v procesu gradnje.**

Vse te težave, povezane z upravljanjem številnih različnih rešitev, vodstvo podjetja prej ali slej pripeljejo do pomembnega spoznanja: ne gre za količino podatkov ali iskanje naslednjega "univerzalnega" orodja za obdelavo podatkov. Pravi razlog se skriva v kakovosti podatkov ter v tem, kako jih organizacija ustvarja, prejema, shranjuje in uporablja.

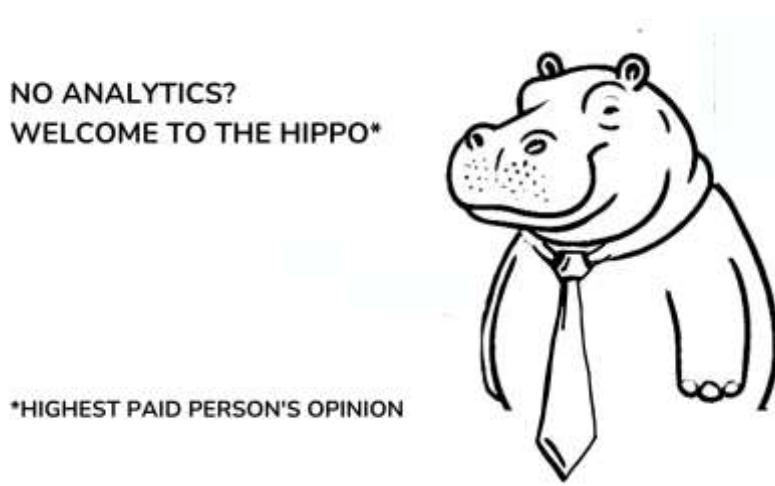
Ključ do trajnostnega uspeha ni v iskanju novih "čarobnih" aplikacij, temveč v vzpostaviti podatkovne kulture v podjetju. To pomeni, da je treba podatke obravnavati kot strateško sredstvo ter da so kakovost, celovitost in ustreznost podatkov prednostna naloga na vseh ravneh organizacije.

Rešitev dileme "kakovost proti količini" je v oblikovanju enotne podatkovne strukture, ki odpravlja podvajanje, nedoslednosti in poenoti pretok informacij. Ta struktura zagotavlja en sam zanesljiv vir podatkov, na podlagi katerega je mogoče sprejemati utemeljene, natančne in pravočasne odločitve.

V nasprotnem primeru se podjetja še naprej zanašajo na subjektivna mnenja in intuitivne ocene strokovnjakov HiPPO in ne na zanesljiva dejstva, kar se pogosto dogaja. To je še posebej opazno v gradbeništvu, kjer ima strokovno znanje tradicionalno pomembno vlogo.

### HiPPO ali nevarnost mnenj pri odločanju

V gradbeništvu se ključne odločitve tradicionalno sprejemajo na podlagi izkušenj in subjektivne presoje. Brez pravočasnih in zanesljivih podatkov morajo vodje podjetij delovati slepo in se zanašati na intuicijo najbolje plačanih zaposlenih (HiPPO - Highest Paid Person's Opinion) in ne na objektivna dejstva (slika 2.1-8).



**Slika 2.1-8 V odsotnosti analitike je poslovanje odvisno od subjektivnega mnenja izkušenih strokovnjakov.**

Ta pristop je lahko upravičen v stabilnem in počasi spreminjajočem se okolju, vendar v dobi digitalne preobrazbe predstavlja resno tveganje. Odločitve, ki temeljijo na intuiciji in ugibanju, so nagnjene k izkriviljanju, pogosto temeljijo na nepodprtih hipotezah in ne upoštevajo kompleksne slike, ki se odraža v podatkih

To, kar se v podjetju na ravni odločanja obravnava kot inteligentna razprava, pogosto ne temelji na ničemer konkremem. Uspeh podjetja ne bi smel biti odvisen od avtoritete in plače strokovnjakov, temveč od sposobnosti učinkovitega dela s podatki, prepoznavanja vzorcev in sprejemanja premišljenih odločitev.

Pomembno je opustiti prepričanje, da avtoriteta ali izkušnje samodejno pomenijo, da je odločitev pravilna. Pристоп, ki temelji na podatkih, spreminja pravila igre: podatki in analitika, ne pa položaj in plača, so zdaj osnova za odločanje. Veliki podatki, strojno učenje in vizualna analitika nam omogočajo prepoznavanje vzorcev in opiranje na dejstva namesto na ugibanja (slika 1.1-4).

*Brez podatkov ste le še ena oseba z lastnim mnenjem [34].*

- W. Edwards Deming, znanstvenik in svetovalec za upravljanje

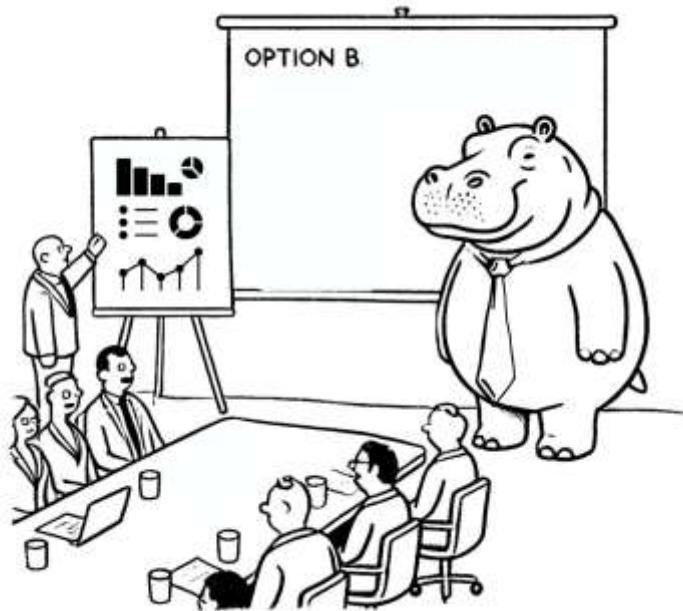
Sodobne metode upravljanja podatkov zagotavljajo tudi kontinuiteto znanja v podjetju. Jasno opisani procesi, avtomatizacija in sistematičen pristop omogočajo prenos celo ključnih vlog brez izgube učinkovitosti.

Vendar pa lahko slepo zaupanje v podatke privede tudi do resnih napak. Podatki so le zbirka številk. Brez ustrezne analize, konteksta in sposobnosti prepoznavanja vzorcev nimajo vrednosti in ne morejo voditi procesov. Ključ do uspeha se ne skriva v izbiri med intuicijo HiPPO in analitiko, temveč v izgradnji inteligentnih orodij, ki raznovrstne podatke pretvorijo v obvladljive in informirane odločitve.

V digitalnem gradbenem okolju odločilna dejavnika uspeha nista višina in mesto v hierarhiji, temveč odzivnost, natančnost odločitev in učinkovitost virov

Podatki so orodje in ne absolutna resnica. Dopolnjujejo človeško razmišljjanje in ga ne nadomeščajo. Kljub prednostim analitike podatki ne morejo popolnoma nadomestiti človeške intuicije in izkušenj. Njihova vloga je, da pomagajo pri sprejemanju natančnejših in bolj informiranih odločitev.

Konkurenčne prednosti ne bomo dosegli le z izpolnjevanjem standardov, temveč tudi s tem, da bomo lahko pri učinkoviti rabi virov, ki so enaki za vse, presegli konkurenco. V prihodnosti bo znanje o podatkih postalo tako pomembno, kot je bilo nekoč znanje pismenosti ali matematike. Strokovnjaki, ki bodo znali analizirati in razlagati podatke, bodo lahko sprejemali natančnejše odločitve in izpodrinili tiste, ki se zanašajo le na osebne izkušnje (slika 2.1-9).



**Slika 2.1-9 Odločitve morajo temeljiti na objektivni analizi in ne na mnenju najbolje plačanega zaposlenega.**

Vodje, strokovnjaki in inženirji bodo delovali kot podatkovni analitiki, ki bodo preučevali strukturo, dinamiko in ključne kazalnike projektov. Človeški viri bodo postali elementi sistema, ki bodo za čim večjo učinkovitost zahtevali prilagodljivo prilaganje na podlagi podatkov.

*Napake pri uporabi neustreznih podatkov so veliko manjše kot pri uporabi brez podatkov [35].*

- Charles Babbage, izumitelj prvega analitičnega računskega stroja

Pojav velikih količin podatkov in uvedba modelov LLM (Large Language Models) sta korenito spremenila ne le način analiziranja, temveč tudi samo naravo odločanja. Če je bila prej v ospredju vzročnost (zakaj se je nekaj zgodilo - diagnostična analitika) (slika 1.1-4), danes prihaja v ospredje sposobnost napovedovanja prihodnosti (napovedna analitika), v prihodnosti pa tudi preskriptivna analitika, kjer strojno učenje in umetna inteligenca predlagata najboljšo izbiro v procesu odločanja.

Glede na novo študijo SAP™ "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36] bi bilo 44% vodstvenih delavcev pripravljenih spremeniti svojo prejšnjo odločitev na podlagi nasveta umetne inteligence, 38% pa bi jih umetni inteligenčni zaupalo sprejemanje poslovnih odločitev v njihovem imenu. Medtem je 74% vodstvenih delavcev izjavilo, da nasvetom UI zaupajo bolj kot svojim prijateljem in družini, 55% pa jih dela v podjetjih, kjer spoznanja, pridobljena z UI, nadomeščajo ali pogosto zaobidejo tradicionalne metode odločanja - zlasti v organizacijah z letnimi prihodki nad 5 milijard dolarjev. Poleg tega 48% vprašanih vsakodnevno uporablja generativna orodja umetne inteligence, od tega jih 15% ta orodja uporablja večkrat na dan.

Z razvojem sistema vseživljenskega učenja in avtomatiziranih sistemov za upravljanje podatkov se pojavlja nov izliv: kako učinkovito uporabljati informacije, ne da bi izgubili njihovo vrednost v kaosu nezdružljivih formatov in heterogenih virov, ki ga dopolnjujeta vse večja zapletenost in dinamika poslovnih procesov.

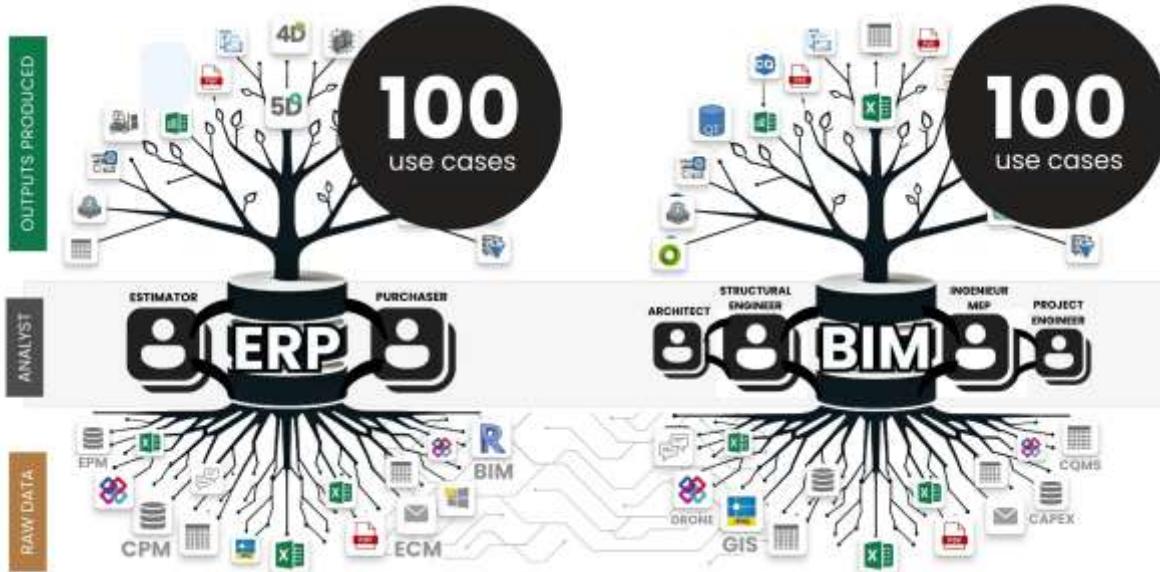
## **nenehno povečevanje kompleksnosti in dinamičnosti poslovnih procesov**

Gradbena industrija se danes sooča z resnimi izzivi na področju upravljanja podatkov in procesov. Glavni izzivi so razdrobljeni informacijski sistemi, pretirana birokracija in pomanjkanje integracije med digitalnimi orodji. Ti izzivi se stopnjujejo, saj so tudi sami poslovni procesi vse bolj zapleteni - zaradi tehnologije, spremenjajočih se zahtev strank in razvijajočih se predpisov.

Edinstvenost gradbenih projektov ni le posledica njihovih tehničnih posebnosti, temveč tudi razlik v nacionalnih standardih in regulativnih zahtevah v različnih državah (slika 4.2-10, slika 5.1-7). To zahteva prilagodljiv, individualen pristop k vsakemu projektu, ki ga je v tradicionalnih modularnih nadzornih sistemih težko izvajati. Zaradi zapletenosti procesov in velike količine podatkov se številna podjetja obračajo na prodajalce, ki ponujajo specializirane rešitve. Toda trg je preobremenjen - številna zagonska podjetja ponujajo podobne izdelke in se osredotočajo na ozke naloge. Zaradi tega se pogosto izgubi celovit pristop k upravljanju podatkov.

Prilaganje stalnemu toku novih tehnologij in zahtev trga postaja ključni dejavnik konkurenčnosti. Vendar so obstoječe lastniške aplikacije in modularni sistemi slabo prilagodljivi - vse spremembe pogosto zahtevajo dolgotrajne in drage popravke s strani razvijalcev, ki ne razumejo vedno posebnosti gradbenih procesov.

Podjetja so talci tehnološkega zaostanka in čakajo na nove posodobitve, namesto da bi takoj uvela inovativne celostne pristope. Zato je notranja struktura gradbenih organizacij pogosto zapleten ekosistem medsebojno povezanih hierarhičnih in pogosto zaprtih sistemov, ki jih usklajuje večnivojska mreža vodij (slika 2.1-10).

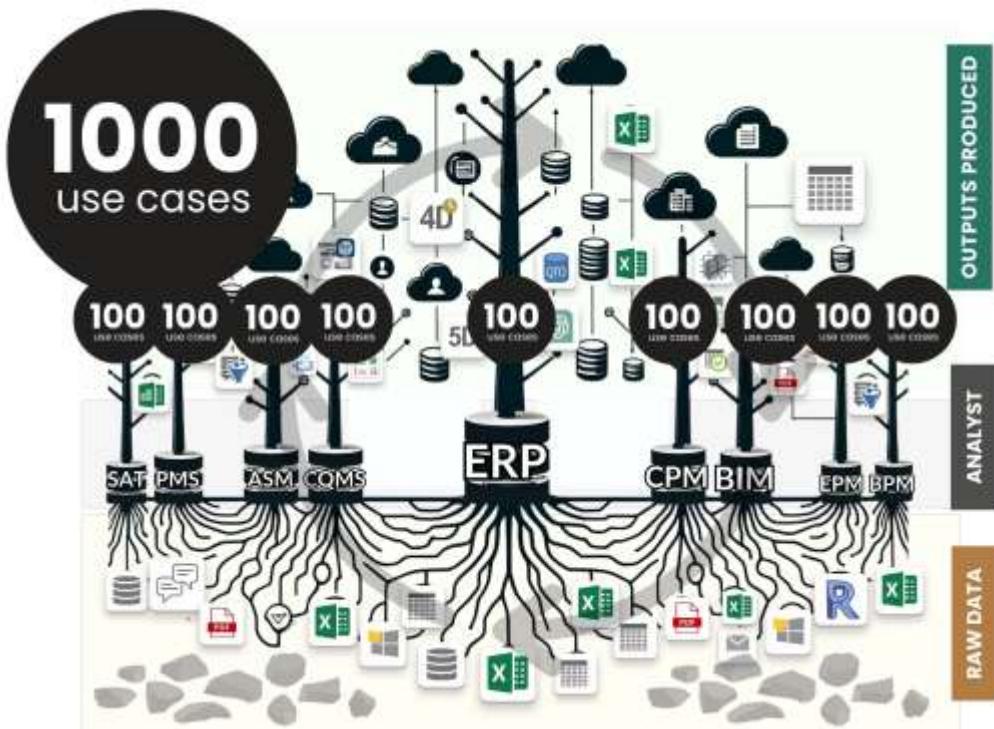


Slika 2.1-10 Podjetja so sestavljena iz medsebojno povezanih sistemov, katerih medsebojna povezanost tvori procese, ki zahtevajo avtomatizacijo.

Glede na študijo, ki sta jo izvedla Kanadsko gradbeno združenje in KPMG Canada 2021 [37], le 25% podjetij meni, da so v primerjavi s konkurenčnimi ali drugačnimi položaji, kar zadeva uvajanje tehnologije ali digitalnih rešitev. Le 23% anketirancev je poročalo, da so njihove rešitve bistveno ali močno usmerjene v podatke. Hkrati je večina anketirancev svojo uporabo vrste drugih tehnologij opredelila kot povsem eksperimentalno ali priznala, da jih sploh ne uporabljajo.

Ta nepripravljenost za sodelovanje pri tehničkih poskusih je še posebej očitna pri velikih infrastrukturnih projektih, kjer lahko napake stanejo milijone dolarjev. Tudi najnaprednejše tehnologije - digitalni dvojčki, napovedna analitika - pogosto naletijo na odpor ne zaradi svoje učinkovitosti, temveč zaradi pomanjkanja dokazane zanesljivosti v resničnih projektih.

V skladu s poročilom Svetovnega gospodarskega foruma (WEF) "Oblikovanje prihodnosti gradbeništva" [5] se pri uvajanju novih tehnologij v gradbeništvo poleg tehničnih težav srečujejo tudi s psihološkimi ovirami s strani strank. [5] se uvajanje novih tehnologij v gradbeništvo ne sooča le s tehničnimi težavami, temveč tudi s psihološkimi ovirami na strani strank. Mnogi naročniki se bojijo, da bodo zaradi uporabe naprednih rešitev njihovi projekti postali poskusno gradbišče in da bodo postali "poskusni zajčki", nepredvidljive posledice pa lahko povzročijo dodatne stroške in tveganja.



Slika 2.1-11 Za vsak primer uporabe podatkov trg rešitev ponuja aplikacije za optimizacijo in avtomatizacijo procesov.

Gradbena industrija je zelo raznolika: različni projekti imajo različne zahteve, regionalne posebnosti, zakonske predpise o klasifikaciji (slika 4.2-10), računske standarde (slika 5.1-7) itd. Zato je praktično nemogoče ustvariti lastno univerzalno aplikacijo ali sistem, ki bi popolnoma ustrezal vsem tem zahtevam in posebnostim projekta.

Pri poskusu obvladovanja vse večje kompleksnosti sistemov in odvisnosti od ponudnikov programske opreme se vse bolj zavedamo, da ključ do učinkovitega upravljanja podatkov nista le odprtost in standardizacija, temveč tudi poenostavitev same arhitekture procesov. Vse večja kompleksnost in dinamičnost poslovnih procesov zahteva nove pristope, pri katerih se prednostna naloga preusmeri od zbiranja podatkov k njihovemu strukturiranju in organiziraju. Prav ta premik bo naslednji korak v razvoju gradbene industrije, ki bo pomenil konec obdobja prevlade proizvajalcev programske opreme in začetek obdobja smiselne organizacije informacij.

Zaradi spoznanja o omejitvah univerzalnih rešitev in občutljivosti na vse večjo kompleksnost se prednostne naloge spreminjajo od zaprtih platform in kopiranja podatkov k preglednosti, prilagodljivosti in strukturiranemu ravnanju z informacijami. Ta premik v razmišljanju odraža širše spremembe v svetovnem gospodarstvu in tehnologiji, opisane skozi prizmo tako imenovanih "industrijskih revolucij". Da bi razumeli, kam gre gradbeništvo in kakšna bo njegova prihodnja usmeritev, je treba upoštevati mesto panoge v kontekstu četrte in pete industrijske revolucije - od avtomatizacije in digitalizacije do personalizacije, odprtih standardov in podatkovnega modela, ki temelji na storitvah.

## Četrta industrijska revolucija (industrija 4.0) in peta industrijska revolucija (industrija 5.0) v gradbeništvu

Tehnološke in gospodarske stopnje so teoretični koncepti, ki se uporabljajo za opis in analizo razvoja družbe in gospodarstva na različnih stopnjah razvoja. Različni raziskovalci in strokovnjaki jih lahko različno razlagajo.

- **Četrta industrijska revolucija** (4IR ali Industrija 4.0) je povezana z informacijsko tehnologijo, avtomatizacijo, digitalizacijo in globalizacijo. Eden njenih ključnih elementov je oblikovanje lastnih programskih rešitev, tj. specializiranih digitalnih izdelkov, ki so namenjeni določenim nalogam in podjetjem. Te rešitve pogosto postanejo pomemben del informacijske infrastrukture, vendar so brez dodatnih sprememb slabo razširljive.
- **Peta industrijska revolucija** (5IR) je zdaj v zgodnejši fazi zasnove in razvoja kot 4IR. Njena temeljna načela vključujejo večjo personalizacijo izdelkov in storitev. 5IR je gibanje v smeri bolj prilagodljivih, prožnih in personaliziranih gospodarskih dejavnosti s poudarkom na personalizaciji, svetovanju in storitveno usmerjenih modelih. Ključni vidik petega gospodarskega načina je uporaba podatkov za sprejemanje odločitev, ki je praktično nemogoča brez uporabe odprtih podatkov in odprtih orodij (slika 2.1-12).



Slika 2.1-12 Četrti vzorec se osredotoča na rešitve, peti vzorec pa na personalizacijo in podatke.

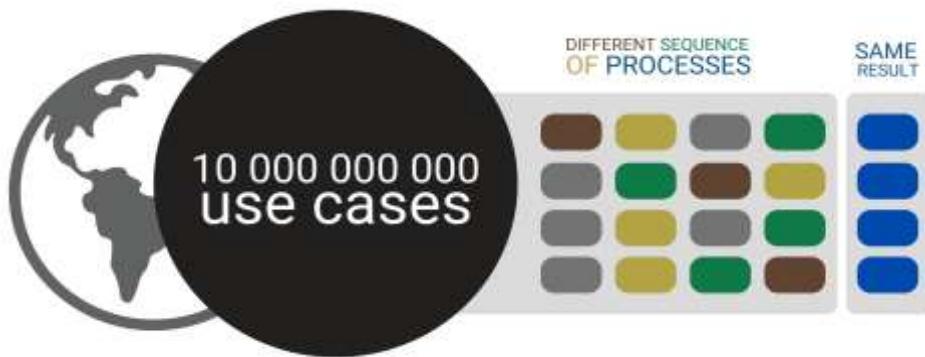
Izdelava aplikacije za podjetja v gradbeništvu za uporabo v desetih ali stotih organizacijah ne zagotavlja njene uspešne razširitve na druga podjetja, regije ali države brez bistvenih sprememb in izboljšav. Verjetnost uspešnega razširjanja takšnih rešitev ostaja majhna, saj ima vsaka organizacija edinstvene procese, zahteve in pogoje, ki lahko zahtevajo prilagojene prilagoditve.

Pomembno je razumeti, da že danes uspešno vključevanje tehnoloških rešitev pomeni zelo prilagojen pristop k vsakemu procesu, projektu in podjetju. To pomeni, da bo tudi po razvoju univerzalnega okvira, orodja ali programa potrebna njegova podrobna prilagoditev in prilagoditev, da bo ustrezal edinstvenim zahtevam in pogojem vsakega posameznega podjetja in projekta.

Po poročilu družbe PwC "Decoding the Fifth Industrial Revolution" (Dekodiranje pete industrijske revolucije) [38] [38] se letos približno 50% vodilnih kadrov v različnih panogah zanaša na povezovanje napredne tehnologije in človeškega strokovnega znanja. Ta pristop jim omogoča hitro prilagajanje spremembam v zasnovi izdelka ali zahtevam strank ter ustvarjanje personalizirane proizvodnje.

Vsek proces zahteva razvoj edinstvene funkcije ali aplikacije, kar glede na velikost svetovne gradbene industrije in raznolikost projektov vodi v obstoj velikega števila poslovnih primerov, ki vsakič predstavljajo edinstveno logiko cevovoda (slika 2.1-13). Vsak tak primer ima svoje posebnosti in zahteva prilagojen pristop. Raznolikost možnih rešitev istega analitičnega problema v okviru različnih pristopov si bomo podrobnejše ogledali v poglavju, posvečenem strojnemu učenju in analizi nabora podatkov o Titaniku (slika 9.2-9).

**Pipeline** v kontekstu digitalnih procesov je zaporedje dejavnosti, procesov in orodij, ki omogočajo avtomatiziran ali strukturiran pretok podatkov in dela skozi različne faze življenjskega cikla projekta.



Slika 2.1-13 Individualnost in spremenljivost poslovnih primerov onemogočata poskuse ustvarjanja skalabilnih zaprtih platform in orodij.

Naše življenje se je pod vplivom digitalne preobrazbe že v marsičem spremenilo, danes pa lahko govorimo o nastopu nove faze v gospodarskem razvoju gradbene industrije. V tem "novem gospodarstvu" bo konkurenca organizirana po drugačnih pravilih: kdor bo zнал javno znanje in odprte podatke učinkovito pretvoriti v iskane izdelke in storitve, bo v pogojih pete industrijske revolucije pridobil ključno prednost.

Kot ugotavlja ekonomistka Kate Maskus v knjigi "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39], 2012 [39], "živimo v globalnem gospodarstvu znanja in prihodnost pripada tistim, ki znajo znanstvena odkritja spremeniti v blago".

Prehod na peti gospodarski način pomeni premik od zaprtih informacijskih rešitev k odprtим standardom in platformam. Podjetja se bodo začela umikati od tradicionalnih programskih izdelkov v korist storitveno usmerjenih modelov, kjer bodo glavno sredstvo podatki in ne lastniške tehnologije.

Študija Harvard Business School 2024 [40] kaže na veliko gospodarsko vrednost odprtokodne programske opreme (Open Source Software, OSS). Glede na študijo je OSS prisoten v 96% vseh programskih kod, nekatera komercialna programska oprema pa je v 99,9% sestavljena iz komponent OSS. Brez OSS bi podjetja za programsko opremo porabila 3,5-krat več sredstev.

Oblikovanje ekosistemov podjetij bo v skladu s svetovnimi trendi postopoma prešlo v peto gospodarsko paradigma, v kateri bodo podatkovno usmerjene analitične in svetovalne storitve postale pomembnejše od izoliranih, zaprtih rešitev s togo določenimi scenariji uporabe.

Doba digitalizacije bo spremenila razmerje moči v industriji: namesto da bi se zanašala na rešitve prodajalcev, bodo podjetja svojo konkurenčnost utemeljila na sposobnosti učinkovite uporabe podatkov. Posledično se bo gradbena industrija premaknila od podedovanih togih sistemov k prožnim, prilagodljivim ekosistemom, v katerih bodo odprti standardi in interoperabilna orodja temelj upravljanja projektov. Konec obdobja prevlade ponudnikov aplikacij bo ustvaril novo okolje, v katerem vrednost ne bo opredeljena z lastništvom zaprte kode in specializiranih povezovalnikov, temveč s sposobnostjo spremnjanja podatkov v strateško prednost.



## POGLAVJE 2.2.

### SPREMINJANJE KAOSA V RED IN ZMANJŠEVANJE KOMPLEKSNOTI

#### Odvečna koda in zaprti sistemi kot ovira za izboljšanje produktivnosti

V zadnjih desetletjih so tehnološke spremembe na področju IT spodbujali predvsem prodajalci programske opreme. Ti so določali smer razvoja in določali, katere tehnologije naj podjetja sprejmejo in katere naj pustijo za seboj. V obdobju prehoda od ločenih rešitev k centraliziranim zbirkam podatkov in integriranim sistemom so prodajalci spodbujali licenčne izdelke, ki so zagotavljali nadzor nad dostopom in razširljivostjo. Kasneje, s prihodom tehnologij v oblaku in modelov programske opreme kot storitve (SaaS), se je ta nadzor razvil v naročniški model, kar je uporabnike utrdilo kot zveste stranke digitalnih storitev.

Ta pristop je privedel do paradoksa: kljub izjemni količini ustvarjene programske kode se dejansko uporablja le majhen del te kode. Morda je kode stokrat ali tisočkrat več, kot je potrebno, saj so isti poslovni procesi opisani in podvojeni v desetinah ali stotinah programov na različne načine, celo znotraj istega podjetja. Hkrati so bili stroški razvoja že plačani in teh stroškov ni mogoče povrniti. Kljub temu industrija še naprej reproducira ta krog in ustvarja nove izdelke z minimalno dodano vrednostjo za končnega uporabnika, pogosteje pod pritiskom tržnih pričakovanj kot dejanskih potreb.

V skladu z Vodnikom za ocenjevanje stroškov razvoja programske opreme (Defence Acquisition University, DAU) [41] se lahko stroški razvoja programske opreme zelo razlikujejo glede na več dejavnikov, vključno s kompleksnostjo sistema in izbrano tehnologijo. V preteklosti so stroški razvoja za leto 2008 znašali približno 100 USD na vrstico izvorne kode (SLOC), medtem ko se lahko stroški vzdrževanja povzpnejo do 4 000 USD na SLOC.

Samo ena od komponent aplikacij CAD - geometrijsko jedro - ima lahko več deset milijonov vrstic kode (slika 6.1-5). Podobno velja za sisteme ERP (slika 5.4-4), k razpravi o kompleksnosti katerih se bomo vrnili v petem delu knjige. Vendar podrobnejši pogled razkrije, da večina te kode ne ustvarja dodane vrednosti, temveč deluje le kot "poštar" - mehanično prenaša podatke med zbirko podatkov, vmesnikom API, uporabniškim vmesnikom in drugimi tabelami v sistemu. Kljub priljubljenemu mitu o ključnem pomenu tako imenovane poslovne logike je kruta realnost veliko bolj prozaična: sodobne baze kode so polne zastarelih blokov predlog (legacy code), katerih edini namen je zagotoviti prenos podatkov med tabelami in komponentami, ne da bi to vplivalo na sprejemanje odločitev ali poslovno učinkovitost.

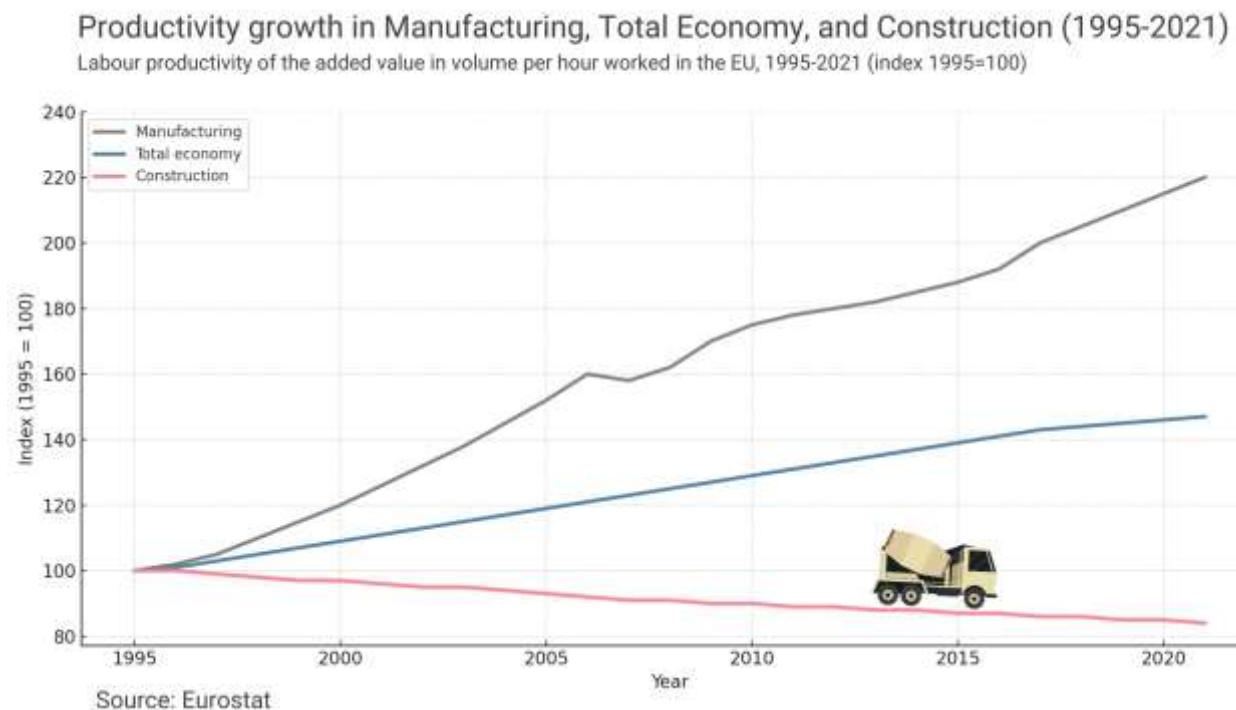
Zato se zaprte rešitve, ki obdelujejo podatke iz različnih virov, neizogibno spremenijo v zmedene "ekosisteme špagetov". Te zapletene in prepletene sisteme lahko upravlja le vojska menedžerjev, ki delajo v polurutinskem načinu. Takšna organizacija upravljanja podatkov ni le neučinkovita z vidika virov, temveč ustvarja tudi kritične ranljivosti v poslovnih procesih, zaradi česar je podjetje odvisno od ozkega kroga strokovnjakov, ki razumejo delovanje tega tehnološkega labirinta.

Nenehno povečevanje količine kode, števila aplikacij in vse večje zapletenosti konceptov, ki jih ponujajo prodajalci, je privedlo do naravne posledice - povečanja zapletenosti ekosistema IT v

gradbeništvu. Zaradi tega je praktično izvajanje digitalizacije s povečanjem števila aplikacij v panogi postalo neučinkovito. Programski izdelki, ustvarjeni brez ustreznega upoštevanja potreb uporabnikov, pogosto zahtevajo veliko sredstev za izvajanje in podporo, vendar ne prinašajo pričakovanih koristi.

V McKinseyjevi študiji "Povečanje produktivnosti v gradbeništvu" je navedeno [42] je v zadnjih dveh desetletjih svetovna rast produktivnosti dela v gradbeništvu v povprečju znašala le 1% na leto v primerjavi z 2,8-odstotno rastjo za celotno svetovno gospodarstvo in 3,6-odstotno rastjo v predelovalnih dejavnostih. V Združenih državah Amerike se je produktivnost dela v gradbeništvu na delavca od šestdesetih let prejšnjega stoletja prepolovila [43].

Vse večja kompleksnost sistemov, izoliranost in zaprtost podatkov so poslabšali komunikacijo med strokovnjaki, zaradi česar je gradbeništvo med najmanj učinkovitimi (slika 2.2-1). do leta 2040 na 22 trilijonov dolarjev, kar bo zahtevalo znatno povečanje učinkovitosti.



Slika 2.2-1 Zaprti in kompleksni podatki ter posledično slaba komunikacija med strokovnjaki so gradbeništvo pripeljali med najmanj učinkovite gospodarske panoge (na podlagi [44], [45]).

Kot je poudarjeno v študiji McKinsey (2024) "Zagotavljanje produktivnosti v gradbeništvu ni več neobvezno", si gradbeništvo zaradi vse večjega pomanjkanja virov in prizadevanj panoge, da podvoji svojo stopnjo rasti, ne more več privoščiti, da bi ostalo na sedanji ravni produktivnosti [44]. Svetovni stroški gradbeništva naj bi se s 13 bilijonov dolarjev leta 2023 povečali na precej višjo raven do konca desetletja, zaradi česar vprašanje učinkovitosti ni le pomembno, temveč ključno.

Eden od ključnih načinov za izboljšanje učinkovitosti bo neizogibno poenotenje in poenostavitev

struktur aplikacij in arhitekture podatkovnih ekosistemov. Ta pristop k racionalizaciji bo odpravil odvečne plasti abstrakcije in nepotreбno zapletenost, ki so se z leti nakopičile v podjetniških sistemih.

## Od silosov do enotnega podatkovnega skladišča

Več podatkov kot jih organizacija zbere, težje je iz njih izluščiti pravo vrednost. Zaradi razdrobljene narave shranjevanja informacij v izoliranih silosih so poslovni procesi sodobnih podjetij podobni gradbenikom, ki poskušajo zgraditi nebotičnik iz materialov, shranjenih v tisočih različnih skladiščih. Zaradi preobilice informacij ni le težko dostopati do pravno relevantnih informacij, temveč se upočasni tudi odločanje: vsak korak je treba večkrat preveriti in potrditi.

Vsako opravilo ali postopek je trdno povezan v ločeno tabelo ali zbirko podatkov, za izmenjavo podatkov med sistemi pa so potrebne zapletene integracije. Napake in nedoslednosti v enem sistemu lahko povzročijo verižne napake v drugih. Nepravilne vrednosti, pozne posodobitve in podvojene informacije prisilijo zaposlene, da porabijo veliko časa za ročno usklajevanje in usklajevanje podatkov. Posledično organizacija porabi več časa za odpravljanje posledic razdrobljenosti kot za razvoj in optimizacijo procesov

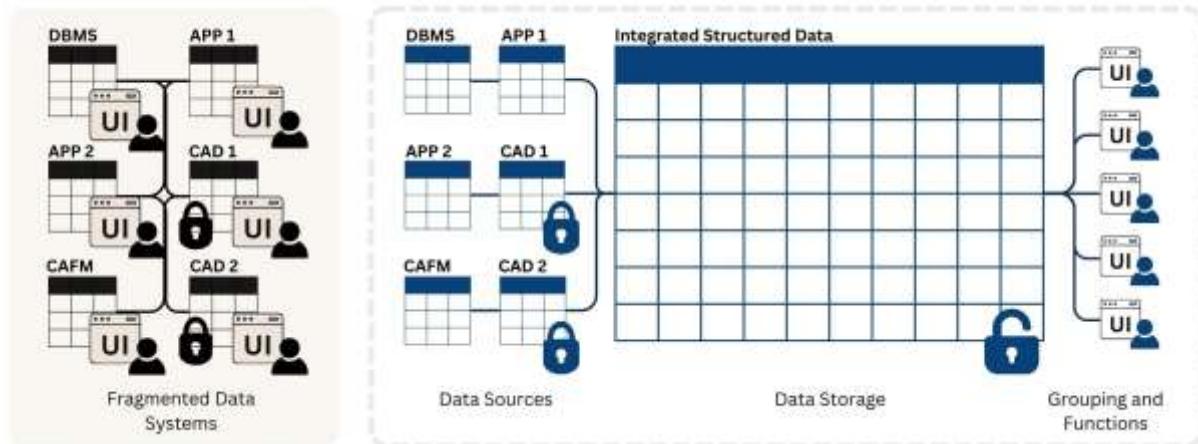
Ta težava je univerzalna: nekatera podjetja se še naprej borijo s kaosom, druga pa najdejo rešitev v integraciji - prenosu informacijskih tokov v centraliziran sistem shranjevanja. Predstavljajte si ga kot eno veliko tabelo, v katero lahko shranite vse entitete, povezane z nalogami, projekti in predmeti. Namesto več deset različnih tabel in formatov nastane eno samo povezano skladišče (slika 2.2-2), ki omogoča

- zmanjšanje izgube podatkov;
- odpraviti potrebo po stalnem usklajevanju informacij;
- izboljšanje razpoložljivosti in kakovosti podatkov;
- poenostavitev analitične obdelave in strojnega učenja.

Približevanje podatkov skupnemu standardu pomeni, da se informacije ne glede na vir pretvorijo v enotno in strojno berljivo obliko. Takšna organizacija podatkov omogoča preverjanje njihove celovitosti, analiziranje v realnem času in takojšnjo uporabo pri sprejemanju vodstvenih odločitev.

Koncept integriranih sistemov za shranjevanje ter njihova uporaba pri analitiki in strojnem učenju bosta podrobnejše obravnavana v poglavju "Shranjevanje velikih količin podatkov in strojno učenje". Teme modeliranja in strukturiranja podatkov bodo podrobno obravnavane v poglavjih "Preoblikovanje podatkov v strukturirano obliko" in "Kako standardi spremenijo igro: od naključnih datotek do

izdelanega podatkovnega modela".



Slika 2.2-2 Integracija podatkov odpravlja silose, izboljšuje razpoložljivost informacij in optimizira poslovne procese.

Ko so podatki strukturirani in združeni, je naslednji logični korak njihovo preverjanje. Z enotnim integriranim skladisčem je ta postopek zelo poenostavljen: nič več več večjih neskladnih shem, podvojenih struktur in zapletenih razmerij med tabelami. Vse informacije so usklajene z enotnim podatkovnim modelom, kar odpravlja notranje nedoslednosti in pospešuje postopek potrjevanja. Potrjevanje in zagotavljanje kakovosti podatkov sta temeljna vidika vseh poslovnih procesov, zato ju bomo podrobnejše obravnavali v ustreznih poglavijih knjige.

V zadnji fazi se podatki razvrstijo, filtrirajo in analizirajo. Uporabljajo se različne funkcije: združevanje (seštevanje, množenje), izračuni med tabelami, stolpci ali vrsticami (slika 2.2-4). Delo s podatki postane zaporedje korakov: zbiranje, strukturiranje, potrjevanje, preoblikovanje, analitična obdelava in prenos v končne aplikacije, kjer se informacije uporabijo za reševanje praktičnih problemov. Več o oblikovanju takšnih scenarijev, avtomatizaciji korakov in oblikovanju tokov obdelave bomo obravnavali v poglavjih o ETL -procesih in pristopu podatkovnih cevovodov.

Digitalna preobrazba torej ne pomeni le poenostavitev ravnjanja z informacijami. Gre za odpravljanje pretirane zapletenosti pri upravljanju podatkov, prehod od kaosa k predvidljivosti, od številnih sistemov k obvladljivemu procesu. Manjša kot je kompleksnost arhitekture, manj kode je potrebne za njeni podporo. V prihodnosti bo morda koda kot tako povsem izginila, saj bo prepustila mesto inteligentnim agentom, ki bodo samostojno analizirali, sistematizirali in preoblikovali podatke.

## Integrirani sistemi za shranjevanje omogočajo prehod na agente z umetno inteligenco

Manj zapleteni kot so podatki in sistemi, manj kode morate napisati in vzdrževati. Najlažji način za prihranek pri razvoju pa je, da se kode popolnoma znebite in jo nadomestite s podatki. Ko se razvoj aplikacijske kode premakne od kode k podatkovnim modelom, neizogibno pride do premika k podatkovno usmerjenemu pristopu (data-driven), saj se za temi koncepti skriva povsem drugačen

način razmišljanja.

Ko se odločimo za delo s podatki v središču, začnemo na njihovo vlogo gledati drugače. Podatki niso več le "surovina" za aplikacije - zdaj so temelj, na katerem se gradijo arhitektura, logika in interakcija.

Tradicionalni pristop k upravljanju podatkov se običajno začne na ravni aplikacij in pri gradnji spominja na okoren birokratski sistem: odobritve na več ravneh, ročni pregledi, neskončne različice dokumentov prek ustreznih programskih izdelkov. Z razvojem digitalnih tehnologij bo vse več podjetij prisiljenih preiti na načelo minimalizma - hraniti in uporabljati le tisto, kar je resnično potrebno in bo uporabljeno.

Prodajalci so prevzeli logiko zmanjševanja. Da bi poenostavili shranjevanje in obdelavo podatkov, se delo uporabnikov z aplikacij in orodij, ki niso povezani z internetom, seli v storitve v oblaku in tako imenovane rešitve SaaS.

Koncept SaaS (Software as a Service ali "programska oprema kot storitev") je eden ključnih trendov v sodobnih infrastrukturah IT, ki uporabnikom omogoča dostop do aplikacij prek interneta, ne da bi jim bilo treba namestiti in vzdrževati programsko opremo v lastnih računalnikih.

Po eni strani je storitev SaaS omogočila lažje skaliranje, nadzor različic ter zmanjšala stroške podpore in vzdrževanja, po drugi strani pa je poleg odvisnosti od logike določene aplikacije uporabnika naredila tudi popolnoma odvisnega od infrastrukture v oblaku ponudnika. Če storitev odpove, je lahko dostop do podatkov in poslovnih procesov začasno ali celo trajno onemogočen. Poleg tega so vsi uporabniški podatki pri delu z aplikacijami SaaS shranjeni na strežnikih ponudnika, kar povzroča varnostna tveganja in tveganja glede skladnosti z zakonodajo. Spremembe tarif ali pogojev uporabe lahko povzročijo tudi višje stroške ali potrebo po nujni migraciji.

Razvoj umetne inteligence, LLM -agentov in podatkovno usmerjenega pristopa je postavil pod vprašaj prihodnost aplikacij v njihovi tradicionalni obliki in izvedbe SaaS. Medtem ko so aplikacije in storitve prej morale upravljati poslovno logiko in obdelovati podatke, se lahko s pojavom agentov umetne inteligence te funkcije prenesejo na inteligentne sisteme, ki delajo neposredno s podatki.

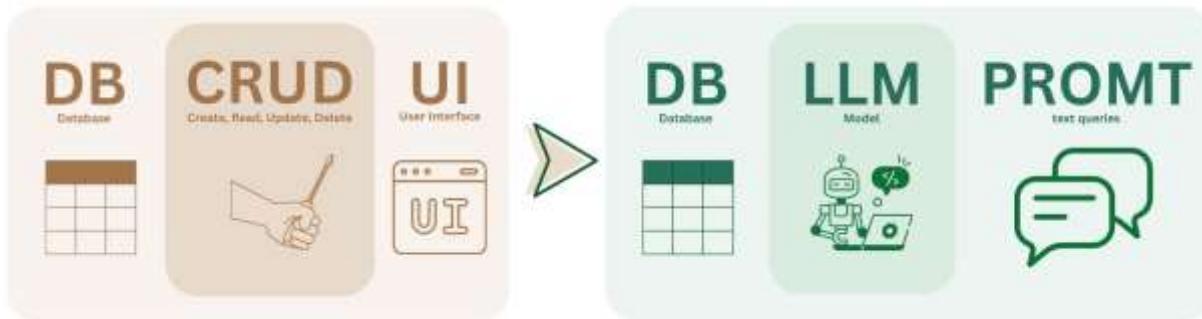
Zato se v oddelkih IT in na ravni vodstva vse pogosteje razpravlja o hibridnih arhitekturah, kjer storitve v oblaku dopolnjujejo agenti za umetno inteligenco in lokalne rešitve, kar zmanjšuje odvisnost od platform SaaS.

*Pristop, ki ga uporabljam, priznava, da se lahko tradicionalne poslovne aplikacije ali aplikacije SaaS v dobi agentov močno spremenijo. Te aplikacije so v bistvu CRUD [ustvarjanje, branje, posodabljanje in brisanje] podatkovne zbirke s poslovno logiko. V prihodnosti pa bodo to logiko prevzeli agenti umetne inteligence [46].*

- Satya Nadella, izvršni direktor Microsofta, 2024.

Pristop, osredotočen na podatke, in uporaba agentov AI/LLM lahko zmanjšata odvečne procese in s tem zmanjšata delovno obremenitev zaposlenih. Če so podatki ustrezno organizirani, jih je lažje analizirati, vizualizirati in uporabiti pri sprejemanju odločitev. Namesto neskončnih poročil in preverjanj strokovnjaki z nekaj kliki ali s pomočjo agentov LLM samodejno dobijo dostop do najnovejših informacij v obliki pripravljenih dokumentov in nadzornih plošč.

Pri obdelavi podatkov nam bodo pomagala orodja umetne inteligence (AI) in klepetalnice LLM. V zadnjih letih je opazen trend opuščanja tradicionalnih operacij CRUD (ustvarjanje, branje, posodabljanje, brisanje) v smeri uporabe velikih jezikovnih modelov (LLM) za upravljanje podatkov. LLM lahko interpretirajo naravni jezik in samodejno generirajo ustrezne poizvedbe po zbirkah podatkov, kar poenostavlja interakcijo s sistemi za upravljanje podatkov (slika 2.2-3).



Slika 2.2-3 Umetna inteligenco bo nadomestila in integrirala rešitve za shranjevanje in podatkovne zbirke ter postopoma izpodrinila tradicionalne aplikacije in operacije CRUD.

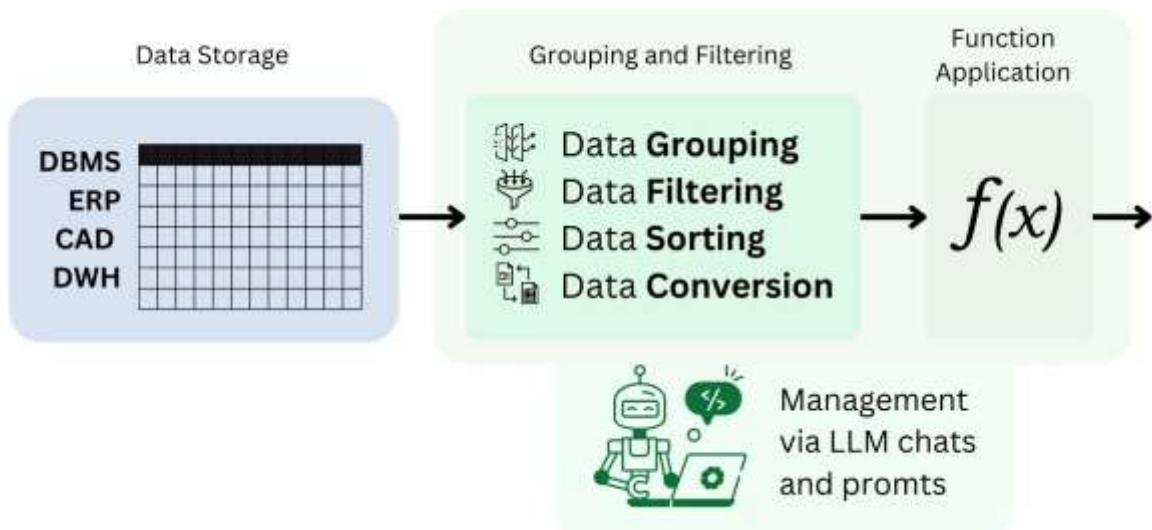
*V naslednjih 3-6 mesecih bo umetna inteligenco napisala 90% kode, v 12 mesecih pa bi lahko skoraj vso kodo ustvarila umetna inteligenco [47].*

- Dario Amodei, izvršni direktor LLM Anthropic, marec 2025.

Kljud hitremu razvoju orodij za razvoj umetne inteligence (npr. GitHub Copilot) bodo leta 2025 razvijalci še vedno imeli ključno vlogo v tem procesu. Agenti umetne inteligence postajajo vse bolj uporabni pomočniki: samodejno interpretirajo uporabnikove poizvedbe, generirajo poizvedbe SQL in Pandas (več o tem v naslednjih poglavjih) ali pišejo kodo za analizo podatkov. Na ta način umetna inteligenco postopoma nadomešča tradicionalne uporabniške vmesnike aplikacij.

Širjenje modelov umetne inteligence, kot so jezikovni modeli, bo spodbudilo razvoj hibridnih arhitektur. Namesto da bi popolnoma opustili rešitve v oblaku in izdelke SaaS, bomo morda priča integraciji storitev v oblaku z lokalnimi sistemmi za upravljanje podatkov. Federativno učenje na primer omogoča zmogljive modele umetne inteligence, ne da bi bilo treba občutljive podatke prenesti v oblak. Na ta način lahko podjetja ohranijo nadzor nad svojimi podatki in hkrati pridobijo dostop do naprednih

tehnologij.



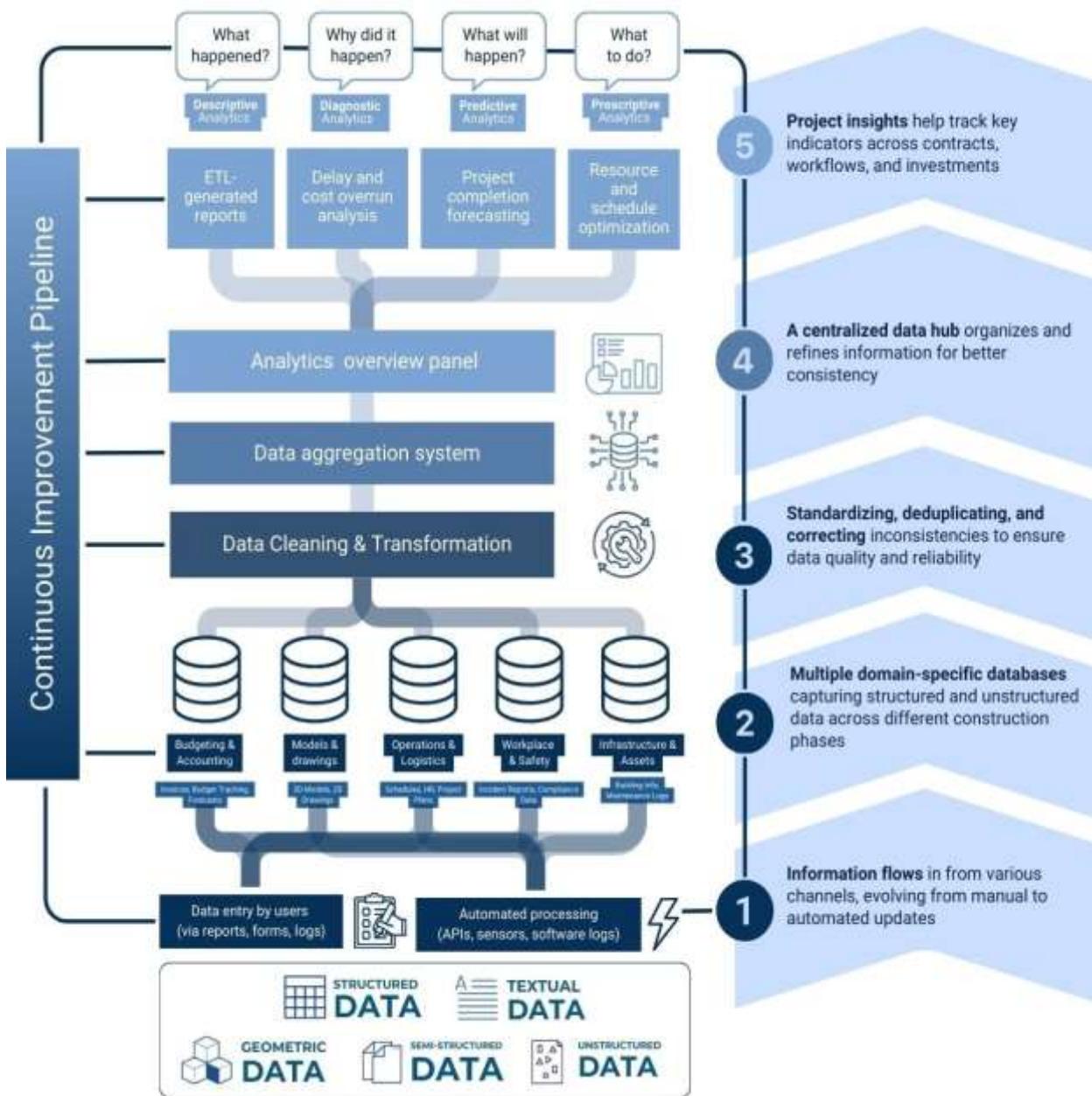
Slika 2.2-4 Osnovne operacije razvrščanja v skupine, filtriranja in razvrščanja, ki jim sledi uporaba funkcij, bodo obdelane v klepetalnicah LLM.

Prihodnost gradbene industrije bo temeljila na kombinaciji lokalnih rešitev, zmogljivosti v oblaku in inteligentnih modelov, ki skupaj ustvarjajo učinkovite in varne sisteme za upravljanje podatkov. LLM bo uporabnikom brez poglobljenega tehničnega znanja omogočil interakcijo s podatkovnimi zbirkami in podatkovnimi skladišči z oblikovanjem poizvedb v naravnem jeziku. Več o agentih LLM in umetni inteligenci ter njihovem delovanju bomo povedali v poglavju "Agenti LLM in strukturirani podatkovni formati".

Ustrezno organizirani podatki in preprosta analitična orodja, ki jih je mogoče uporabljati v okviru LLM, ne bodo le olajšali dela z informacijami, temveč bodo tudi pomagali zmanjšati število napak, povečati učinkovitost in avtomatizirati procese.

### Od zbiranja podatkov do odločanja: pot do avtomatizacije

V nadalnjih delih knjige bomo podrobno preučili, kako strokovnjaki medsebojno sodelujejo in kako podatki postanejo podlaga za sprejemanje odločitev, avtomatizacijo in operativno učinkovitost. Na sliki 2.2-5 je primer diagrama, ki prikazuje zaporedje korakov obdelave podatkov v podatkovno usmerjenem pristopu. Ta diagram ponazarja cevovod za nenehno izboljševanje), katerega deli bodo podrobno obravnavani v nadaljevanju knjige.



Slika 2.2-5 Primer stalnega cevovoda za izboljšanje podatkov: tok obdelave in analize podatkov v gradbenih projektih.

Sistem, ki opisuje poslovne procese srednje velikega podjetja, je zgrajen po večnivojskem načelu. Vključuje: zbiranje podatkov, čiščenje, združevanje, analitično obdelavo in odločanje na podlagi rezultatov. Vse te stopnje bomo preučili v nadaljevanju knjige - tako v teoretičnem kontekstu kot na praktičnih primerih:

- Na prvi ravni poteka **vnos podatkov** (slika 3.1-1). Informacije se prejemajo ročno (prek poročil, obrazcev, dnevnikov) in v avtomatizirani obliki (iz API, senzorjev, programskih sistemov). Podatki so lahko različne strukture: geometrijski, besedilni, nestrukturirani. Na tej stopnji

- obstaja potreba po standardizaciji, strukturiranju in poenotenju informacijskih tokov.
- Naslednja stopnja je **obdelava in preoblikovanje podatkov**. Vključuje postopke čiščenja, odstranjevanja podvajanj, popravljanja napak in priprave informacij za nadaljnjo analizo (slika 4.2-5). Ta stopnja je ključnega pomena, saj je kakovost analitike neposredno odvisna od čistosti in natančnosti podatkov.
  - **Podatki se nato prenesejo v specializirane tabele, podatkovne okvire ali zbirke podatkov**, razdeljene po funkcionalnih področjih: proračun in računovodstvo, modeli in risbe, logistika, varnost in infrastruktura. Ta delitev omogoča enostaven dostop do informacij in njihovo navzkrižno analizo.
  - Podatki se nato **zdržijo in prikažejo v analitični nadzorni plošči** (vitrina). Pri tem se uporabljajo opisne, diagnostične, napovedne in preskriptivne analitične metode. To pomaga odgovoriti na ključna vprašanja (slika 1.1-4): kaj se je zgodilo, zakaj se je zgodilo, kaj se bo zgodilo v prihodnosti in katere ukrepe je treba sprejeti. Sistem lahko na primer ugotovi zamude, predvidi dokončanje projektov ali optimizira vire.
  - Na zadnji ravni se oblikujejo **analitične ugotovitve in ključni kazalniki**, ki pomagajo spremeljati izpolnjevanje pogodb, upravljati naložbe in izboljšati poslovne procese (slika 7.4-2). Te informacije postanejo podlaga za sprejemanje odločitev in razvojno strategijo podjetja.

Podobno se podatki od zbiranja preusmerijo k uporabi v strateškem upravljanju. V naslednjih delih knjige si bomo podrobno ogledali vsako stopnjo, pri čemer se bomo osredotočili na vrste podatkov, tehnike obdelave podatkov, analitična orodja in resnične primere uporabe teh pristopov v gradbeništvu.

## Naslednji koraki: spremjanje kaosa v obvladljiv sistem

V tem delu smo raziskali izzive informacijskih silosov in preučili vpliv prevelike kompleksnosti sistema na poslovno uspešnost ter analizirali prehod iz četrte industrijske revolucije v peto, kjer so v ospredju podatki in ne aplikacije. Videli smo, kako ločeni informacijski sistemi ustvarjajo ovire za izmenjavo znanja, nadaljnja zapletenost informacijskega okolja pa zmanjšuje produktivnost in zavira inovacije v gradbeništvu.

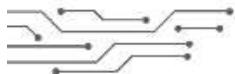
Za povzetek tega dela velja izpostaviti glavne praktične korake, ki vam bodo pomagali pri uporabi obravnavanih pristopov pri vsakodnevnih nalogah:

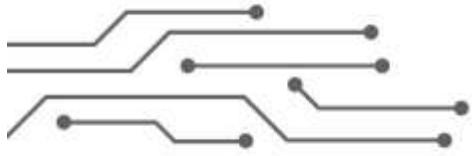
- Vizualizirajte svojo informacijsko pokrajino
  - Ustvarite vizualni zemljevid virov podatkov (Miro, Figma, Canva), s katerimi redno delate.
  - Na ta zemljevid dodajte sisteme in aplikacije, ki jih uporabljate pri svojem delu.
  - Ugotavljanje morebitnih podvojenih funkcij in odvečnih rešitev
  - Opredelitev kritičnih točk, kjer lahko pride do izgube ali poškodbe podatkov med prenosom med sistemi.
- Izvajanje praks upravljanja prilagojenih podatkov.
  - Preusmeritev pozornosti z aplikacij na podatke kot ključno sredstvo v procesih

- dokumentiranje virov podatkov in metodologije obdelave za zagotavljanje preglednosti.
  - Razvoj mehanizmov za ocenjevanje in izboljšanje kakovosti podatkov
  - Prizadevajte si, da se podatki vnesejo enkrat in uporabijo večkrat - to je osnova za učinkovito organizacijo procesov.
- Spodbujajte v svoji ekipi pristop, osredotočen na podatke (podatkovno usmerjen).
- predlagajo uporabo standardiziranih in enotnih formatov za medsebojno izmenjavo podatkov.
  - na sestankih skupine redno izpostavljajte vprašanja, povezana s kakovostjo in razpoložljivostjo podatkov.
  - Spoznajte odprtakodne alternative orodjem, ki jih uporabljate za reševanje svojih težav.

Začnite z majhnimi koraki - izberite en specifičen proces ali niz podatkov, ki je ključnega pomena za vaše delo, in pri njem uporabite pristop, osredotočen na podatke, pri čemer preusmerite pozornost z orodij na podatke. Z uspehom v enem samem pilotnem projektu ne boste pridobili le praktičnih izkušenj, temveč tudi jasno predstavili prednosti nove metodologije svoji ekipi. Pri izvedbi večine teh korakov lahko v primeru vprašanj poiščete pojasnila in pomoč pri katerem koli aktualnem programu LLM.

V naslednjih delih knjige si bomo podrobnejše ogledali tehnike strukturiranja in usklajevanja podatkov ter raziskali praktične pristope k povezovanju heterogenih informacij. Posebno pozornost bomo namenili prehodu od raznovrstnih silosov k poenotenim podatkovnim ekosistemom, ki imajo ključno vlogo pri digitalni preobrazbi gradbeništva.





### III DEL

## PODATKOVNI OKVIR V POSLOVNIH PROCESIH V GRADBENIŠTVU

V tretjem delu je oblikovano celovito razumevanje tipologije podatkov v gradbeništvu in metod njihove učinkovite organizacije. Analizirane so značilnosti in posebnosti dela s strukturiranimi, nestrukturiranimi, polstrukturiranimi, besedilnimi in geometrijskimi podatki v okviru gradbenih projektov. Pregledani so sodobni formati shranjevanja in protokoli za izmenjavo informacij med različnimi sistemmi, ki se uporabljajo v industriji. Opisana so praktična orodja in tehnike za pretvorbo podatkov različnih formatov v enotno strukturirano okolje, vključno s tem, kako vključiti podatke CAD (BIM). Predlagani so pristopi za zagotavljanje kakovosti podatkov s standardizacijo in potrjevanjem, ki so ključnega pomena za natančnost gradbenih izračunov. Podrobno so analizirani praktični vidiki uporabe sodobnih tehnologij (Python Pandas, LLM -modeli) s primeri kode za reševanje tipičnih problemov v gradbeništvu. Utemeljena je vrednost ustanovitve kompetenčnega centra (CoE) kot organizacijske strukture za usklajevanje in standardizacijo pristopov k upravljanju informacij.

## POGLAVJE 3.1.

# PODATKOVNE VRSTE V GRADBENIŠTVU

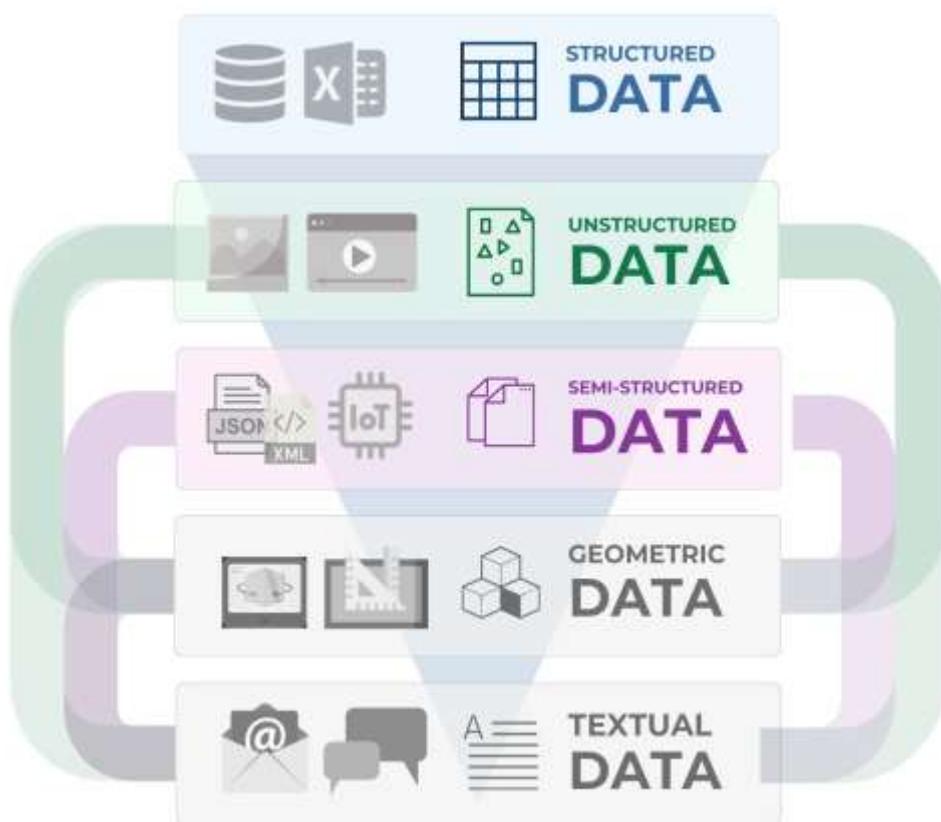
### Najpomembnejše vrste podatkov v gradbeništvu

V sodobni gradbeni industriji se sistemi, aplikacije in podatkovna skladišča podjetij aktivno polnijo z informacijami in podatki različnih vrst in oblik (slika 3.1-1). Podrobneje si oglejmo glavne vrste podatkov, ki tvorijo informacijsko krajino sodobnega podjetja, ki deluje v gradbeništvu:

- **Strukturirani** podatki: ti podatki imajo jasno organizacijsko strukturo, npr. Excelove preglednice in relacijske podatkovne zbirke.
- **Nestukturirani** podatki: to so informacije, ki niso urejene v skladu s strogimi pravili. Primeri takih podatkov so besedilo, video, fotografije in zvočni posnetki.
- **Ohlapno strukturirani** podatki: ti podatki zavzemajo vmesni položaj med strukturiranimi in nestukturiranimi podatki. Vsebujejo elemente strukture, vendar ta struktura ni vedno jasna ali pa je pogosto opisana z različnimi shemami. Primeri polstrukturiranih podatkov v gradbeništvu so: tehnične specifikacije, projektna dokumentacija ali poročila o napredku.
- **Besedilni podatki**: vključujejo vse, kar izhaja iz ustne in pisne komunikacije, kot so elektronska sporočila, prepisi sestankov in srečanj.
- **Geometrijski** podatki: ti podatki izvirajo iz programov CAD, v katerih strokovnjaki ustvarjajo geometrijske podatke elementov projekta za vizualizacijo, potrditev prostorninskih vrednosti ali preverjanje trkov.

Pomembno je opozoriti, da geometrijski in besedilni (alfanumerični) podatki niso ločena kategorija, temveč so lahko prisotni v vseh treh vrstah podatkov. Geometrijski podatki so na primer lahko del strukturiranih podatkov (parametrični formati CAD) in nestukturiranih podatkov (skenirane risbe). Besedilni podatki so lahko podobno organizirani v podatkovnih zbirkah (strukturirani podatki) in obstajajo kot dokumenti brez jasne strukture.

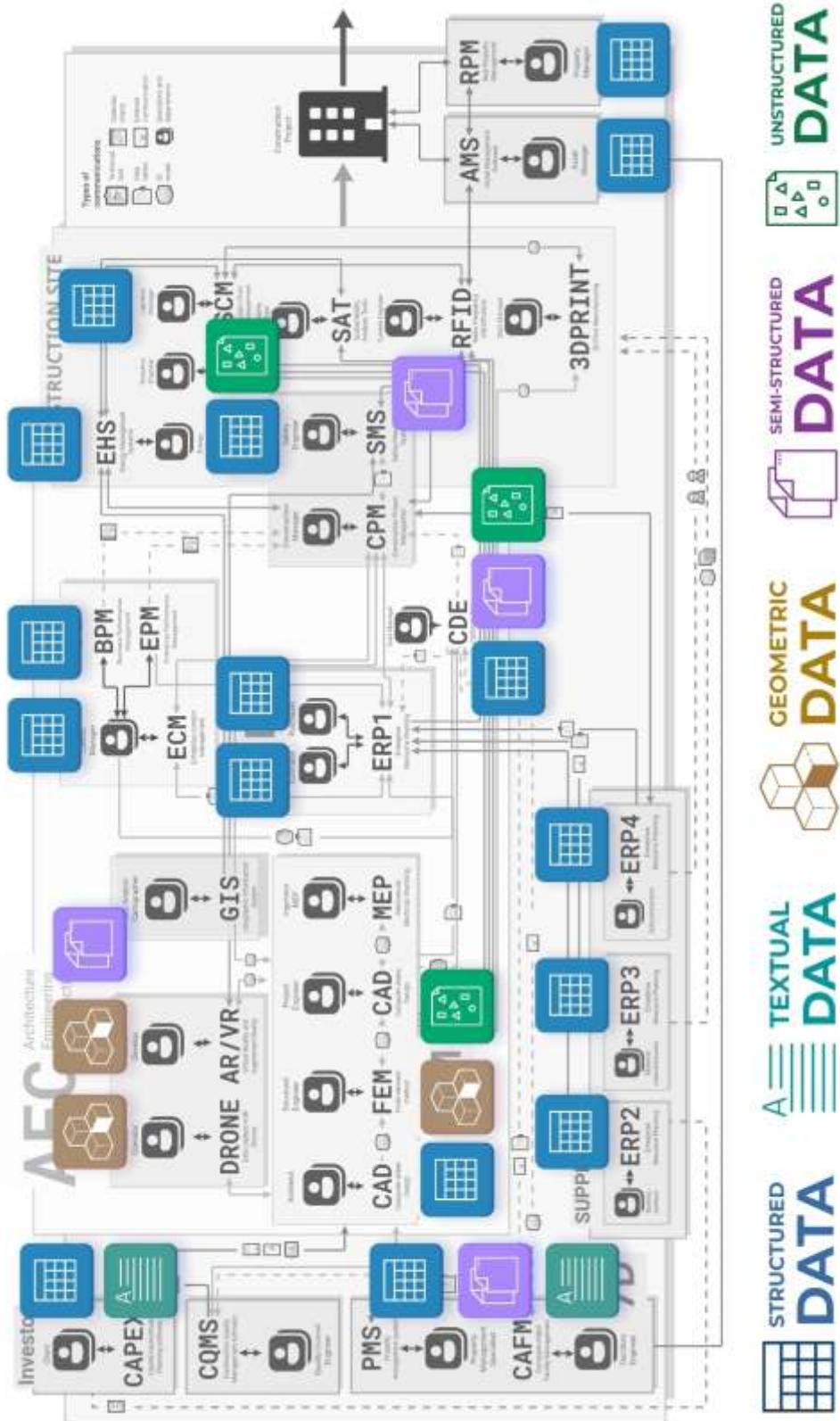
Vsaka vrsta podatkov v gradbenem podjetju je edinstven element v mozaiku informacijskih sredstev podjetja. Od nestukturiranih podatkov, kot so slike z gradbišč in zvočni posnetki sestankov, do strukturiranih zapisov, vključno s tabelami in podatkovnimi zbirkami, ima vsak element pomembno vlogo pri oblikovanju informacijske krajine podjetja.



Slika 3.1-1 Inženirji in upravljavci podatkov se morajo naučiti delati z vsemi vrstami podatkov, ki se uporabljajo v gradbeništvu.

Tukaj je vzorčni seznam le nekaterih sistemov in z njimi povezanih vrst podatkov (slika 3.1-2), ki se uporabljajo v gradbeništvu:

- **ERP** (Enterprise Resource Planning) - obdeluje splošno strukturirane podatke za pomoč pri upravljanju virov podjetja in povezovanju različnih poslovnih procesov.
- **CAD** (Computer-Aided Design) v kombinaciji z **BIM** (Building Information Modeling) - uporablja geometrijske in polstrukturirane podatke za načrtovanje in modeliranje gradbenih projektov ter zagotavlja natančnost in doslednost informacij v fazi načrtovanja.
- **GIS** (geografski informacijski sistemi) - dela z geometrijskimi in strukturiranimi podatki za ustvarjanje in analizo kartografskih podatkov in prostorskih odnosov.
- **RFID** (Radio-Frequency Identification) - uporablja polstrukturirane podatke za učinkovito sledenje materialu in opremi na gradbišču z uporabo radiofrekvenčne identifikacije.
- **ECM** (Engineering Content Management) je sistem za upravljanje inženirskega podatkov in dokumentacije, vključno s polstrukturiranimi in nestrukturiranimi podatki, kot so tehnične risbe in projektna dokumentacija.

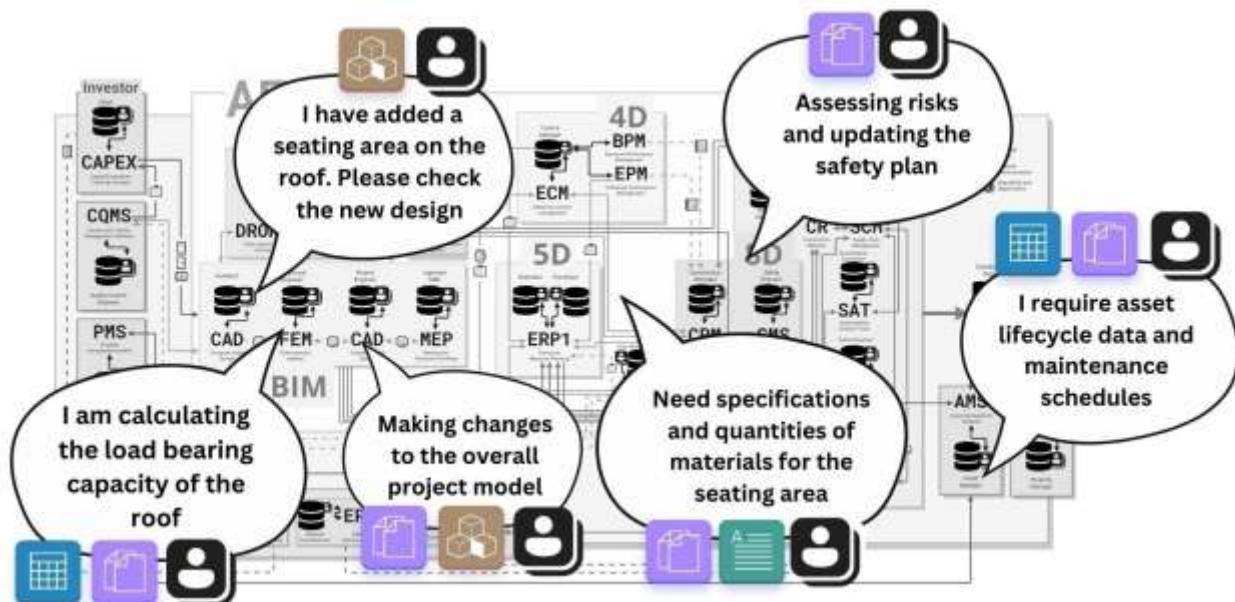


Slika 3.1-2 Različni formati in podatki polnijo različne sisteme, zato jih je treba prevesti v obliko, primerno za kompleksno integracijo.

Ti in številni drugi sistemi podjetja upravljajo širok nabor podatkov, od strukturiranih tabeličnih podatkov do zapletenih geometrijskih modelov, ter zagotavljajo celostno sodelovanje v procesih načrtovanja, načrtovanja in upravljanja gradnje.

V primeru poenostavljenega dialoga (slika 3.1-3) si strokovnjaki za gradbeni projekt izmenjujejo različne vrste podatkov:

- ⦿ **Arhitekt:** "Ob upoštevanju naročnikovih želja sem na streho dodal prostor za sedenje. Prosim, oglejte si novo zasnova" (geometrijski podatki - model).
- ⦿ **Gradbeni inženir:** "Projekt je bil sprejet. Izračunavam nosilnost strehe za novo rekreacijsko območje" (strukturirani in polstrukturirani podatki - računske tabele).
- ⦿ **Vodja nabave:** "Potrebujem specifikacije in količine materialov za rekreacijsko območje za organizacijo nakupa" (besedilni in polstrukturirani podatki - sezname in specifikacije).
- ⦿ **Inženir za zdravje in varnost:** "Prejeli smo podatke o novem območju. Ocenjujem tveganja in posodabljam varnostni načrt" (polstrukturirani podatki - dokumenti in načrti).
- ⦿ **Strokovnjak za BIM - modeliranje:** "Izvajanje sprememb celotnega modela projekta za prilagoditev delovne dokumentacije" (geometrijski podatki in polstrukturirani podatki).
- ⦿ **Vodja projekta:** "Novo počivališče sem vključil v delovni načrt. Posodabljam urnike in vire v sistemuh za upravljanje projekta" (strukturirani in polstrukturirani podatki - urniki in načrti).
- ⦿ **Strokovnjak za vzdrževanje objektov (FM):** "(strukturirani in polstrukturirani podatki - navodila in načrti vzdrževanja).



Slika 3.1-3 Komunikacija med strokovnjaki poteka na ravni besedila in podatkov.

Vsek strokovnjak dela z različnimi vrstami podatkov, da bi zagotovil učinkovito skupinsko sodelovanje in uspeh projekta. Razumevanje razlik med strukturiranimi, polstrukturiranimi in nestrukturiranimi podatki vam omogoča, da prepoznate edinstveno vlogo, ki jo ima vsaka vrsta podatkov v digitalnih poslovnih procesih. Pomembno je ne le vedeti, da obstajajo različne oblike podatkov, temveč tudi razumeti, kako in zakaj se uporabljajo.

Še pred kratkim se je zdela zamisel o združevanju tako raznolikih podatkov ambiciozna, a težko uresničljiva. Danes je to že del vsakdanje prakse. Integracija podatkov različnih schem in struktur je postala sestavni del arhitekture sodobnih informacijskih sistemov.

V naslednjih poglavjih si bomo podrobno ogledali ključne standarde in pristope, ki omogočajo združevanje strukturiranih, polstrukturiranih in nestukturiranih podatkov v enoten pregled. Posebno pozornost bomo namenili strukturiranim podatkom in relacijskim podatkovnim bazam kot glavnim mehanizmom za shranjevanje, obdelavo in analizo informacij v gradbeništvu.

## Strukturirani podatki

V gradbeništvu informacije prihajajo iz številnih virov - iz risb, specifikacij, načrtov in poročil. Za učinkovito upravljanje tega toka informacij jih je treba strukturirati. Strukturirani podatki omogočajo organizacijo informacij v priročni, berljivi in dostopni obliki.

Po podatkih 5. letnega poročila o tehnologiji v gradbeništvu, ki ga je pripravila družba JB Knowledge [17], 67% strokovnjakov za vodenje gradbenih projektov spremja in ocenjuje delovno uspešnost ročno ali z uporabo preglednic.

Med najpogostejšimi oblikami strukturiranih podatkov sta XLSX in CSV. Ti se pogosto uporabljajo za shranjevanje, obdelavo in analizo informacij v preglednicah. V takšnih preglednicah so podatki predstavljeni v obliki vrstic in stolpcev, kar omogoča enostavno branje, urejanje in analiziranje.

**Format XLSX**, ki ga je ustvaril Microsoft, temelji na uporabi struktur XML in se arhivira z uporabo algoritma ZIP. Glavne značilnosti tega formata:

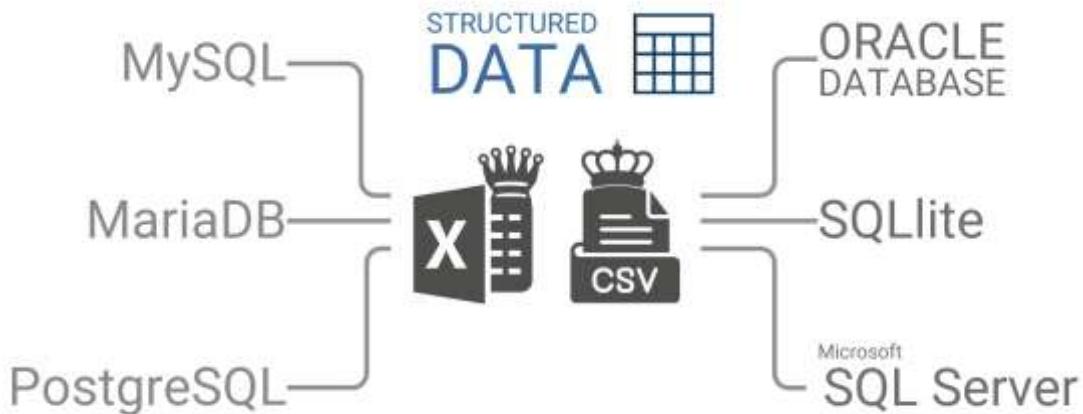
- Podpora za zapletene formule, grafe in makra.
- Sposobnost shranjevanja podatkov v različnih listih in oblikovanja informacij.
- Optimizirano za Microsoft Excel, vendar je združljivo z drugimi pisarniškimi paketi.

**Oblika CSV** je navadna besedilna datoteka, v kateri so vrednosti ločene z vejicami, podpičji ali drugimi ločilnimi znaki. Glavne prednosti:

- Univerzalna združljivost z različnimi programi in operacijskimi sistemi.
- Enostaven uvoz/izvoz v podatkovne zbirke in analitične sisteme.
- Enostavna obdelava tudi v urejevalnikih besedil.

Vendar CSV ne podpira formul in oblikovanja, zato se uporablja predvsem za izmenjavo podatkov med sistemi in množično posodabljanje informacij. Zaradi svoje vsestranskoosti in neodvisnosti od platforme je CSV postal priljubljeno orodje za prenos podatkov v heterogenih okoljih IT.

Formata XLSX in CSV delujeta kot povezava med različnimi sistemi, ki se ukvarjajo s strukturiranimi podatki (slika 3.1-4). Še posebej sta uporabna pri nalogah, kjer so pomembni berljivost, ročno urejanje in osnovna združljivost



Slika 3.1-4 Formata XLSX in CSV predstavljata povezavo med različnimi sistemi, ki delajo s strukturiranimi podatki.

Zaradi neodvisnosti od platforme je CSV najbolj priljubljen format za prenos podatkov v heterogenih okoljih in sistemih IT.

Vendar XLSX in CSV nista zasnovana za visoko zmogljivo računalništvo ali dolgoročno shranjevanje velikih količin podatkov. Za take namene se uporabljajo sodobnejši strukturirani formati, kot so Apache Parquet, Apache ORC, Feather, HDF5. Ti formati bodo podrobneje obravnavani v poglavju "Shranjevanje velikih podatkov: analiza priljubljenih formatov in njihova učinkovitost" v 9. delu te knjige.

V praksi se Excel z obliko XLSX pogosteje uporablja za manjša opravila in avtomatizacijo rutinskih procesov. Za bolj zapletene scenarije je potrebna uporaba sistemov za upravljanje podatkov, kot so ERP, PMIS CAFM, CPM, SCM in drugi (slika 3.2-1). Ti sistemi hranijo strukturirane podatke, na katerih temeljita organizacija in upravljanje informacijskih tokov podjetja.

Sodobni informacijski sistemi za upravljanje podatkov, ki se uporabljajo v gradbeništvu, temeljijo na strukturiranih podatkih, organiziranih v obliki tabel. Razvijalci aplikacij in sistemov se za zanesljivo, zaznavno in celostno upravljanje velikih količin informacij obračajo k sistemom za upravljanje relacijskih podatkovnih zbirk (RDBMS).

## Relacijske zbirke podatkov RDBMS in poizvedovalni jezik SQL

Za učinkovito shranjevanje, obdelavo in analizo podatkov so **relacijske zbirke podatkov (RDBMS)** sistemi za shranjevanje podatkov, ki organizirajo informacije v tabele z opredeljenimi odnosi med njimi.

Podatki, organizirani v zbirkah podatkov (RDBMS), niso le digitalne informacije; so osnova za transakcije in interakcije med različnimi sistemi.

Tukaj je nekaj najpogostejših sistemov za upravljanje relacijskih podatkovnih zbirk (RDBMS) (slika 3.1-5):

- **MySQL** (odprta koda) je ena izmed najbolj priljubljenih RDBMS, ki je del sklada LAMP (Linux, Apache, MySQL, PHP /Perl/Python). Zaradi svoje preprostosti in visoke zmogljivosti se pogosto uporablja pri razvoju spletnih strani.
- **PostgreSQL** (odprta kodaje zmogljiv objektno-relacijski sistem, znan po svoji zanesljivosti in naprednih funkcijah. Primeren je za kompleksne rešitve za podjetja.
- **Strežnik Microsoft SQL** je Microsoftov komercialni sistem, ki se zaradi integracije z drugimi izdelki podjetja in visoke stopnje varnosti pogosto uporablja v poslovnih okoljih.
- **Podatkovna zbirka Oracle** je eden od najzmogljivejših in najzanesljivejših sistemov DBMS, ki se uporabljajo v velikih podjetjih in kritičnih aplikacijah.
- **IBM DB2** - namenjen velikim korporacijam, zagotavlja visoko zmogljivost in odpornost na napake.
- **SQLite** (odprta kodaje lahka vgrajena podatkovna zbirka, idealna za mobilne aplikacije in samostojne sisteme, kot je programska oprema za načrtovanje CAD (BIM).

Priljubljeni sistemi za upravljanje podatkovnih zbirk v gradbeništvu - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 in SQLite - delujejo s strukturiranimi podatki. Vsi ti sistemi DBMS so zmogljive in prilagodljive rešitve za upravljanje številnih poslovnih procesov in aplikacij, od majhnih spletnih strani do velikih poslovnih sistemov (slika 3.2-1).

Po podatkih Statista [48] bodo sistemi za upravljanje relacijskih podatkovnih zbirk (RDBMS) leta 2022 predstavljali približno 72% vseh uporabljenih DBMS.

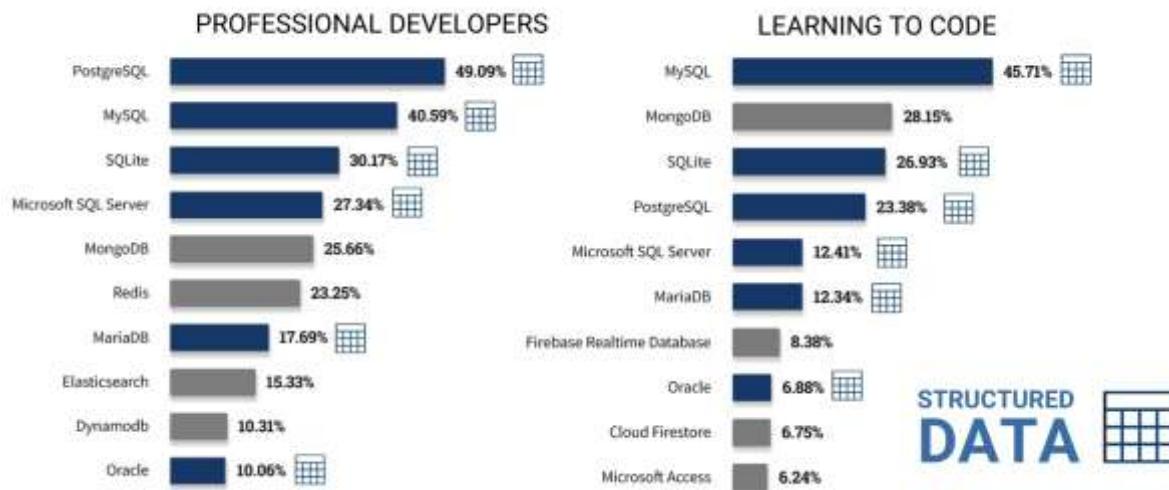
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open 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

Slika 3.1-5 Priljubljenost uporabe strukturiranih podatkovnih zbirk (označeno z modro) pri razvrščanju DBMS (na podlagi [49]).

Namestitev odprtokodnih podatkovnih zbirk je precej preprosta - tudi brez obsežnega tehničnega znanja. Odprtokodni sistemi, kot so PostgreSQL, MySQL ali SQLite, so na voljo brezplačno in delujejo v večini operacijskih sistemov: Windows, macOS in Linux. Vse, kar potrebujete, je, da obiščete uradno spletno stran projekta, prenesete namestitveni program in sledite navodilom. V večini primerov namestitev ne traja več kot 10 do 15 minut. V četrtem delu knjige bomo modelirali in ustvarili eno takšno zbirko podatkov (slika 4.3-8).

Če vaše podjetje uporablja storitve v oblaku (na primer Amazon Web Services, Google Cloud ali Microsoft Azure), lahko podatkovno zbirko namestite z nekaj kliki - platforma vam bo ponudila pripravljene predloge za namestitev. Zaradi odprtosti kode je takšne zbirke podatkov enostavno prilagoditi vašim nalogam, velika skupnost uporabnikov pa vam bo vedno pomagala najti rešitev za vsako težavo.

RDBMS ostajajo temelj za številne poslovne aplikacije in analitične platforme (slika 3.1-6), ki podjetjem omogočajo učinkovito shranjevanje, obdelavo in analizo podatkov ter s tem sprejemanje informiranih in pravočasnih odločitev.



Slika 3.1-6 Raziskava med razvijalci na največjem forumu IT StackOverFlow o tem, katere zbirke podatkov so uporabljali lani in katere želijo uporabljati naslednje leto (RDBMS so označene z modro) (na podlagi [50]).

Sistemi RDBMS zagotavljajo zanesljivost, doslednost podatkov, podporo transakcijam in uporabljajo zmogljiv jezik za poizvedbe - SQL (strukturirani jezik za poizvedbe), ki se pogosto uporablja v analitiki in omogoča enostavno pridobivanje, spreminjanje in analiziranje informacij, shranjenih v zbirkah podatkov. SQL je glavno orodje za delo s podatki v relacijskih sistemih.

## SQL - poizvedbe v zbirkah podatkov in novi trendi

Glavna prednost jezika SQL, ki se pogosto uporablja v relacijskih zbirkah podatkov, pred drugimi vrstami upravljanja informacij (na primer s pomočjo klasičnih preglednic Excel) je podpora zelo velikim količinam podatkovnih zbirk pri visoki hitrosti obdelave poizvedb.

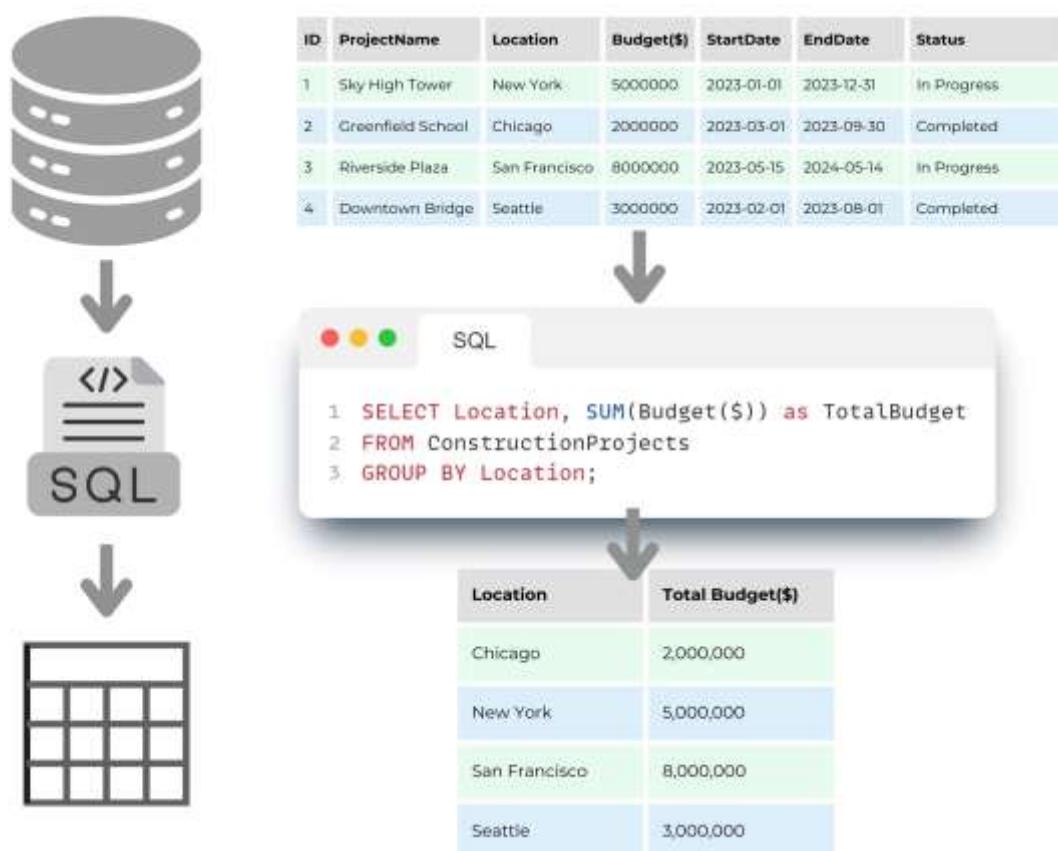
Strukturirani poizvedovalni jezik (SQL) je specializiran programski jezik, namenjen shranjevanju, obdelavi in analizi informacij v relacijskih podatkovnih zbirkah. SQL se uporablja za ustvarjanje, upravljanje in dostop do podatkov ter omogoča učinkovito iskanje, filtriranje, združevanje in združevanje informacij. Služi kot ključno orodje za dostop do podatkov ter zagotavlja priročen in formaliziran način interakcije s skladišči informacij.

Razvoj sistemov SEQUEL-SQL poteka skozi pomembne izdelke in podjetja, kot so Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL in MySQL, vrhunc pa predstavlja pojav SQLite in MariaDB [51]. SQL zagotavlja možnosti preglednic, ki jih ni mogoče najti v Excelu, zaradi česar je ravnanje s podatki bolj skalabilno, varno in enostavno za avtomatizacijo:

- **Ustvarjanje in upravljanje podatkovnih struktur (DDL):** V jeziku SQL lahko ustvarjate, spreminjate in brišete tabele v zbirki podatkov, vzpostavljate povezave med njimi in določate strukture za shranjevanje podatkov. Excel po drugi strani deluje s fiksнимi listi in celicami,

brez jasno opredeljenih povezav med listi in podatkovnimi nizi.

- **Manipulacija s podatki (DML):** SQL omogoča množično dodajanje, spremjanje, brisanje in priklic podatkov z veliko hitrostjo z izvajanjem zapletenih poizvedb s filtriranjem, razvrščanjem in združevanjem tabel (slika 3.1-7). V programu Excel je za obdelavo velikih količin podatkov potrebno ročno ukrepanje ali posebni makri, kar upočasni postopek in poveča verjetnost napak.
- **Nadzor dostopa (DCL):** SQL omogoča razlikovanje pravic dostopa do podatkov za različne uporabnike, s čimer omejite možnost urejanja ali pregledovanja informacij. V Excelu pa je dostop bodisi deljen (pri prenosu datoteke) bodisi zahteva zapletene nastavitev z deljenjem pravic prek storitev v oblaku.



Slika 3.1-7 Primer DML v SQL: hitra obdelava, grupiranje in agregiranje z nekaj vrsticami kode za samodejno obdelavo podatkov.

Excel zaradi svoje vizualne in intuitivne strukture omogoča enostavno delo s podatki. Vendar se z večanjem količine podatkov Excelova zmogljivost zmanjuje. Excel ima tudi omejitve glede količine podatkov, ki jih lahko shrani - največ en milijon vrstic -, zmogljivost pa se poslabša že dolgo pred dosegom te meje. Medtem ko je Excel boljši za vizualizacijo in upravljanje majhnih količin podatkov, je SQL primernejši za obdelavo velikih podatkovnih nizov.

Naslednja stopnja v razvoju strukturiranih podatkov je bil pojav stolpčnih podatkovnih zbirk (Columnar Databases), ki so alternativa tradicionalnim relacijskim podatkovnim zbirkam, zlasti ko gre za bistveno

večje količine podatkov in analitične izračune. V nasprotju z vrstilnimi podatkovnimi zbirkami, kjer so podatki shranjeni vrstico za vrstico, stolpcne podatkovne zbirke zapisujejo podatke po stolpcih. V primerjavi s klasičnimi podatkovnimi zbirkami to omogoča:

- Zmanjšajte prostor za shranjevanje z učinkovitim stiskanjem enotnih podatkov v stolpcih.
- Pospešite analitične poizvedbe, saj se berejo samo zahtevani stolpci in ne celotna tabela.
- Optimizacija velikih podatkov in podatkovnega skladiščenja, npr. Arhitektura podatkovnega skladišča.

Več o stolpčnih podatkovnih zbirkah, Pandas DataFrame, Apache Parquet, HDF5 studi o ustvarjanju velikih podatkovnih -skladišč na njihovi podlagi za namene analize in obdelave podatkov bomo povedali v naslednjih poglavjih te knjige - "DataFrame: univerzalni tabelarni podatkovni format" in "Formati za shranjevanje podatkov in delo z Apache Parquet: DWH - podatkovna skladišča in arhitektura Data Lakehouse".

## Nestrukturirani podatki

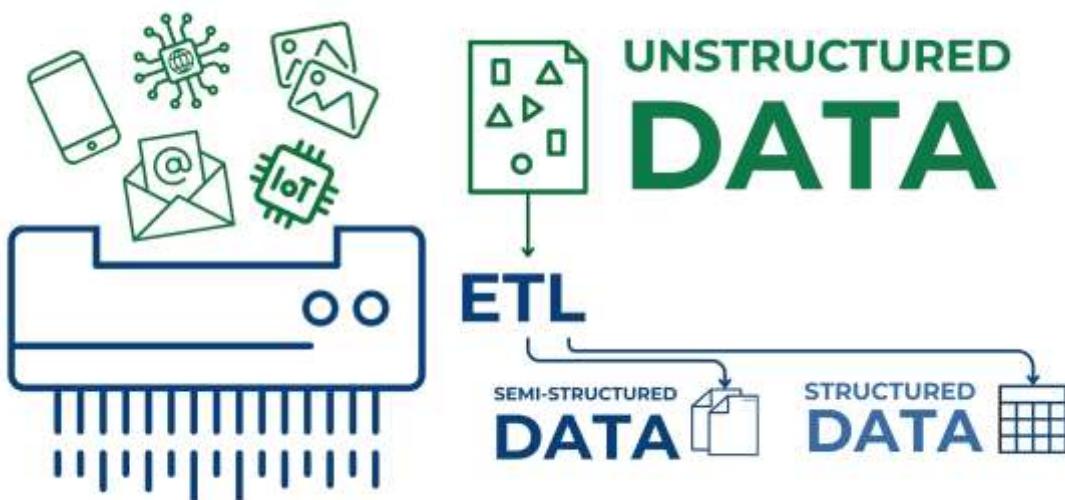
Čeprav je večina podatkov, ki se uporabljajo v aplikacijah in informacijskih sistemih, v strukturirani obliki, je večina informacij, ki nastajajo v gradbeništvu, v obliki nestrukturiranih podatkov - slik, videoposnetkov, besedilnih dokumentov, zvočnih posnetkov in drugih oblik vsebine. To še posebej velja za faze gradnje, obratovanja in tehničnega nadzora, kjer prevladujejo vizualne in besedilne informacije.

**Nestrukturirani podatki** so informacije, ki nimajo vnaprej določenega modela ali strukture in niso urejeni v tradicionalne vrstice in stolpce kot v podatkovnih zbirkah ali tabelah.

Na splošno lahko nestrukturirane podatke razvrstimo v dve kategoriji:

- nestrukturirani podatki, ki jih je ustvaril človek in vključujejo različne vrste vsebine, ki jo je ustvaril človek: besedilne dokumente, e-poštna sporočila, slike, videoposnetke itd.
- Strojno generirane nestrukturirane podatke ustvarjajo naprave in senzorji: to so na primer dnevniške datoteke, podatki GPS, rezultati interneta stvari (IoT) in druge telemetrične informacije z gradbišča.

Za razliko od strukturiranih podatkov, ki jih je priročno urediti v tabele in podatkovne zbirke, nestrukturirani podatki pred vključitvijo v informacijske sisteme zahtevajo dodatne korake obdelave (slika 3.1-8). Uporaba tehnologij za avtomatizirano zbiranje, analizo in preoblikovanje takšnih podatkov odpira nove možnosti za izboljšanje učinkovitosti gradnje, zmanjšanje napak in zmanjšanje vpliva človeškega dejavnika.



Slika 3.1-8 Obdelava nestrukturiranih podatkov se začne s preoblikovanjem v polstrukturirane in strukturirane podatke.

Nestrukturirani podatki predstavljajo do 80% vseh informacij [52], s katerimi se srečujejo strokovnjaki v podjetjih, zato bomo v naslednjih poglavjih knjige podrobno obravnavali njihove vrste in obdelavo s primeri.

Zaradi lažje razprave so besedilni podatki razvrščeni ločeno. Čeprav gre za vrsto precej nestrukturiranih podatkov, je treba njihovemu pomenu in razširjenosti v gradbeništvu nameniti posebno pozornost.

### Besedilni podatki: med nestrukturiranim kaosom in strukturiranim kaosom y

Besedilni podatki v gradbeništvu zajemajo široko paletu formatov in vrst informacij, od papirnih dokumentov do neformalnih načinov komunikacije, kot so pisma, pogovori, delovna korespondenca in ustni sestanki na gradbišču. Vsi ti besedilni podatki vsebujejo pomembne informacije za upravljanje gradbenih projektov, od podrobnosti o projektnih odločitvah in spremembah načrtov do razprav o varnostnih vprašanjih in pogajanj z izvajalci in naročniki (slika 3.1-9).



Slika 3.1-9 Besedilni podatki, ena od najbolj priljubljenih vrst informacij, ki se uporablja pri komunikaciji med udeleženci projekta.

Besedilne informacije so lahko formalizirane in nestrukturirane. Med formalizirane podatke spadajo Wordovi dokumenti (.doc,.docx), PDF in besedilne datoteke zapisnikov sestankov (.txt). Neformalizirani podatki vključujejo korespondenco prek messengerja in elektronske pošte, zapise sestankov (Teams, Zoom, Google Meet) in zvočne posnetke razprav (.mp3,.wav), ki jih je treba pretvoriti v besedilo.

Medtem ko imajo pisni dokumenti, kot so uradne zahteve, pogodbeni pogoji in elektronska sporočila, običajno že določeno strukturo, pa so ustna sporočila in delovna korespondenca pogosto nestrukturirani, zato jih je težko analizirati in vključiti v sisteme za upravljanje projektov.

Ključ do učinkovitega upravljanja besedilnih podatkov je njihova pretvorba v strukturirano obliko. To omogoča samodejno vključevanje obdelanih informacij v obstoječe sisteme, ki že delajo s strukturiranimi podatki.



Slika 3.1-10 Pretvarjanje besedilne vsebine v strukturirane podatke.

Za učinkovito uporabo besedilnih informacij jih je treba samodejno pretvoriti v strukturirano obliko (slika 3.1-10). Ta postopek običajno vključuje več korakov:

- **Prepoznavanje besedila (OCR)** - pretvorba slik dokumentov in risb v strojno berljivo obliko.
- **Analiza besedila (NLP)** - samodejna identifikacija ključnih parametrov (datumi, zneski in številke, pomembne za projekt).
- **Razvrščanje podatkov** - razvrščanje informacij (finance, logistika, upravljanje tveganj).

Po prepoznavanju in razvrščanju je mogoče že strukturirane podatke vključiti v podatkovne zbirke in jih uporabiti v avtomatiziranih sistemih poročanja in upravljanja.

## Polstrukturirani in ohlapno strukturirani podatki

Delno strukturirani podatki vsebujejo določeno raven organizacije, vendar nimajo stroge sheme ali strukture. Čeprav takšne informacije vključujejo strukturirane elemente (npr. datume, imena zaposlenih in sezname opravljenih nalog), se lahko oblika predstavitev od projekta do projekta ali celo od zaposlenega do zaposlenega precej razlikuje. Primeri takih podatkov so dnevnički delovnega časa, poročila o napredku in urniki, ki so lahko predstavljeni v različnih oblikah.

Polstrukturirane podatke je lažje analizirati kot nestrukturirane podatke, vendar zahtevajo dodatno

obdelavo za vključitev v standardizirane sisteme za upravljanje projektov.

Delo s polstrukturiranimi podatki, za katere je značilna stalno spremenljajoča se struktura, predstavlja velik izziv. Razlog za to je, da spremenljivost strukture podatkov zahteva ločene individualne pristope k obdelavi in analizi vsakega vira polstrukturiranih podatkov.

Medtem ko obravnava nestrukturiranih podatkov zahteva veliko truda, pa je obdelavo polstrukturiranih podatkov mogoče opraviti z razmeroma preprostimi metodami in orodji.

**Slabo strukturirani** podatki je splošnejši izraz, ki opisuje podatke z minimalno ali nepopolno strukturo. Najpogosteje gre za besedilne dokumente, klepete, elektronska sporočila, v katerih najdemo nekaj metapodatkov (npr. datum, pošiljatelj), vendar je večina informacij predstavljena kaotično.

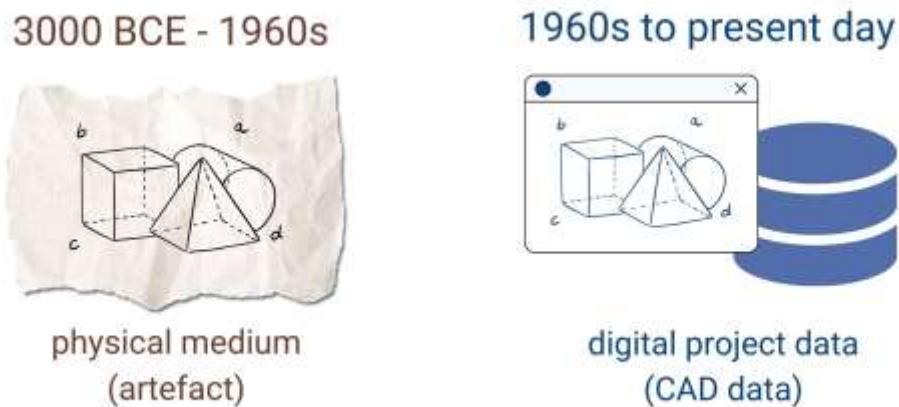
V gradbeništvu se ohlapno strukturirani podatki pojavljajo v različnih procesih. Med njimi so na primer:

- Ocene in ponudbe - tabele s podatki o materialih, količinah in stroških, vendar brez enotne oblike.
- Risbe in inženirski diagrami - datoteke v formatu PDF ali DWG, ki vsebujejo besedilne opombe in metapodatke, vendar nimajo strogog določene strukture.
- Urniki dela - podatki iz MS Project, Primavera P6 ali drugih sistemov, ki imajo lahko drugačno strukturo izvoza.
- CAD (BIM -modeli) - vsebujejo elemente konstrukcije, vendar je predstavitev podatkov odvisna od programske opreme in projektnega standarda.

Geometrične podatke, ki jih ustvarjajo sistemi CAD, je mogoče razvrstiti na enak način kot polstrukturirane podatke. Vendar bomo geometrijske podatke CAD (BIM) razvrstili kot ločeno vrsto podatkov, ker jih je mogoče, tako kot besedilne podatke, v procesih podjetja pogosto obravnavati kot ločeno vrsto podatkov.

## Geometrijski podatki in njihova uporaba

Medtem ko so metapodatki o elementih projekta skoraj vedno shranjeni v obliki tabel, strukturiranih ali ohlapno strukturiranih formatov, so geometrijski podatki elementov projekta v večini primerov ustvarjeni s posebnimi orodji CAD (slika 3.1-11), ki omogočajo podrobno vizualizacijo elementov projekta kot množice črt (2D) ali geometrijskih teles (3D).

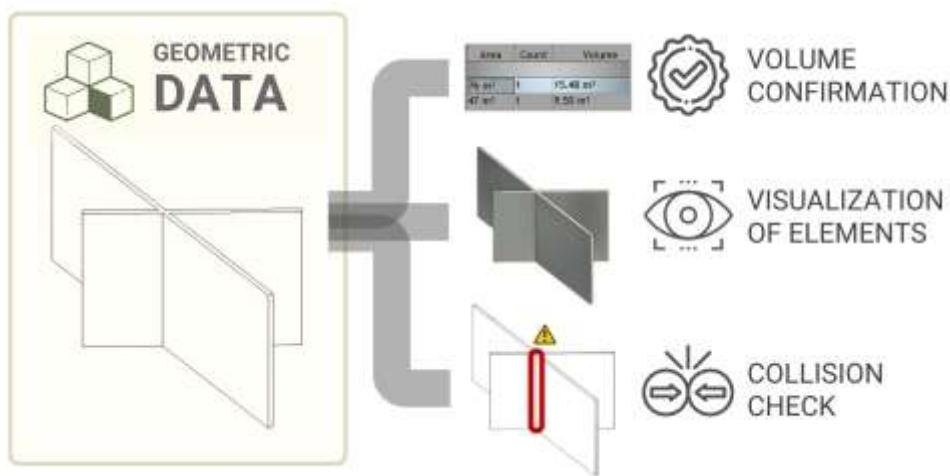


**Slika 3.1-11 Orodja CAD so pripomogla k prenosu geometrijskih informacij iz fizičnega medija v obliko podatkovne zbirke.**

Pri delu z geometrijskimi podatki v gradbeništvu in arhitekturi ločimo tri glavne načine uporabe geometrijskih podatkov (slika 3.1-12):

- **Potrditev volumnov:** geometrijski podatki, ustvarjeni v programih CAD (BIM) z uporabo posebnih geometrijskih jeder, so potrebni za samodejno in natančno določitev volumnov in dimenziј elementov projekta. Ti podatki vključujejo samodejno izračunane površine, prostornine, dolžine in druge pomembne attribute, ki so potrebni za načrtovanje, pripravo proračuna ter naročanje virov in materialov
- **Vizualizacija projekta:** v primeru kakršnih koli sprememb v projektu vizualizacija elementov omogoča samodejno izdelavo posodobljenih risb v različnih ravninah. Vizualizacija projekta v začetnih fazah pomaga pospešiti razumevanje med vsemi udeleženci in tako prihraniti čas in sredstva med postopkom gradnje.
- **Preverjanje trkov:** Pri zapletenih gradbenih in inženirskih projektih, kjer je ključnega pomena medsebojno delovanje več kategorij elementov (npr. cevi in sten) brez "geometrijskih konfliktov", ima preverjanje trkov ključno vlogo. Z uporabo programske opreme za odkrivanje trkov lahko proaktivno prepozname morebitne geometrijske konflikte med elementi projekta in tako preprečite drage napake med gradnjo.

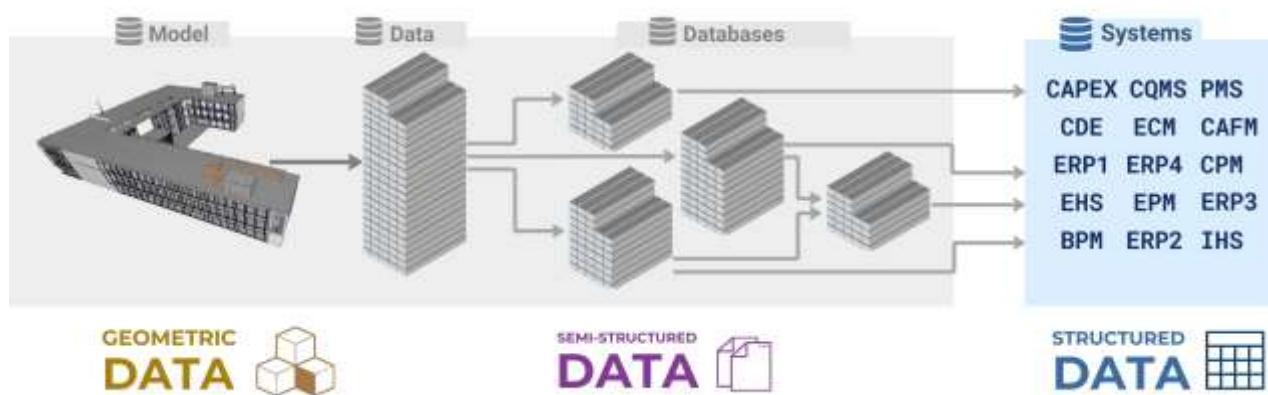
Od samega začetka delovanja inženirskih projektantskih pisarn, od časa gradnje prvih kompleksnih objektov, so gradbeni inženirji zagotavljali geometrijske podatke v obliki risb, črt in ravnih geometrijskih elementov (na papirusu, papirju "A0" ali v formatih DWG, PDF, PLT), na podlagi katerih so mojstri in ocenjevalci (slika 3.1-11), v zadnjih tisočletjih s pomočjo ravnila in kotnikov zbirali atributne količine ali število elementov in skupin elementov. 3.1-11), za zadnja tisočletja s pomočjo ravnila in transporterjev zbirali atributne volumne ali količine elementov in skupin elementov.



**Slika 3.1-12 Geometrija je osnova za pridobitev volumskih parametrov elementov, ki se nato uporabijo za izračun stroškov in časovnega načrta projekta.**

Danes je to ročno in zamudno opravilo popolnoma avtomatizirano zaradi pojava volumetričnega modeliranja v sodobnih orodjih CAD (BIM), ki s pomočjo posebnega geometrijskega jedra omogoča samodejno pridobivanje volumetričnih atributov katerega koli elementa brez potrebe po ročnem izračunavanju volumetričnih parametrov.

Sodobna orodja CAD omogočajo tudi razvrščanje in kategoriziranje elementov projekta, tako da lahko iz podatkovne baze projekta prenesete tabele specifikacij za uporabo v različnih sistemih, kot so ocenjevanje stroškov, načrtovanje ali izračun CO<sub>2</sub> (slika 3.1-13). O pridobivanju specifikacij, tabel QTO in količin ter praktičnih primerih bomo govorili v poglavju "Pridobivanje količin in količinski izračun".



**Slika 3.1-13 Orodja CAD (BIM) shranjujejo podatke v podatkovnih zbirkah, ki so namenjene povezovanju in interakciji z drugimi sistemi.**

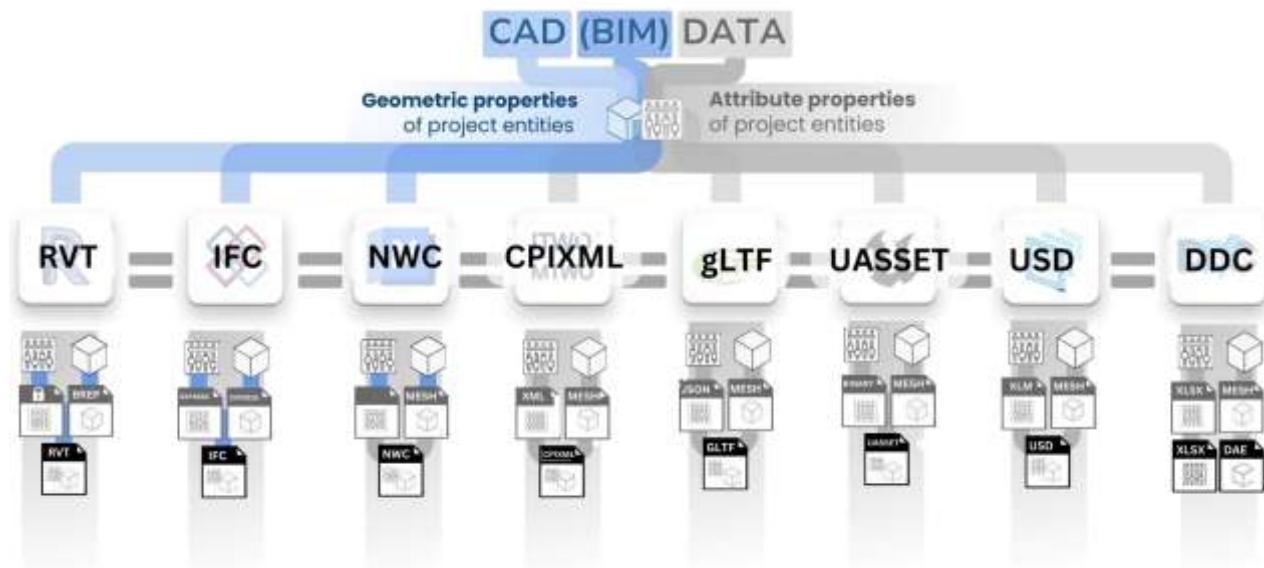
Zaradi zaprte narave podatkovnih zbirk in formatov, ki se uporabljajo v okolju CAD, so geometrijski podatki, ustvarjeni v rešitvah CAD, dejansko postali ločena vrsta informacij. Združujejo geometrijo elementov in metainformacije (strukturirane ali polstrukturirane), ki so zaprte v specializiranih datotekah in formatih.

## Podatki CAD: od načrtovanja do shranjevanja podatkov

Sodobni sistemi CAD in BIM shranjujejo podatke v lastnih, pogosto lastniških formatih: DWG, DXF, RVT, DGN, PLN in drugih. Ti formati podpirajo 2D- in 3D-predstavitev objektov, pri čemer ohranjajo ne le geometrijo, temveč tudi atribute, povezane z objekti. Tu so najpogostejši:

- **DWG** je binarni datotečni format, ki se uporablja za shranjevanje dvodimenzionalnih (in redkeje tridimenzionalnih) podatkov o načrtovanju in metapodatkov.
- **DXF** je besedilni format za izmenjavo 2D in 3D -risb med sistemi CAD. Vsebuje podatke o geometriji, slojih in atributih ter podpira tako ASCII kot binarno predstavitev.
- **RVT** je binarni format za shranjevanje modelov CAD, vključno s 3D -geometrijo, atributi elementov, odnosi in parametri načrtovanja.
- **IFC** je odprt besedilni format za izmenjavo gradbenih podatkov med sistemi CAD (BIM). Vključuje geometrijo, lastnosti objektov in informacije o njihovih odnosih.

Poleg teh se uporabljajo tudi drugi formati: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Čeprav se razlikujejo po namenu in stopnji odprtosti (slika 3.1-14), lahko vsi predstavljajo isti informacijski model projekta v različnih oblikah. Pri kompleksnih projektih se ti formati pogosto uporabljajo vzporedno, od risanja do usklajevanja projektnih modelov.



Slika 3.1-14 Priljubljeni formati za shranjevanje CAD opisujejo geometrijo s parametri BREP ali MESH, ki jih dopolnjujejo atributni podatki.

Vsi zgornji formati omogočajo shranjevanje podatkov o vsakem elementu gradbenega projekta, vsi zgornji formati pa vsebujejo dve ključni vrsti podatkov:

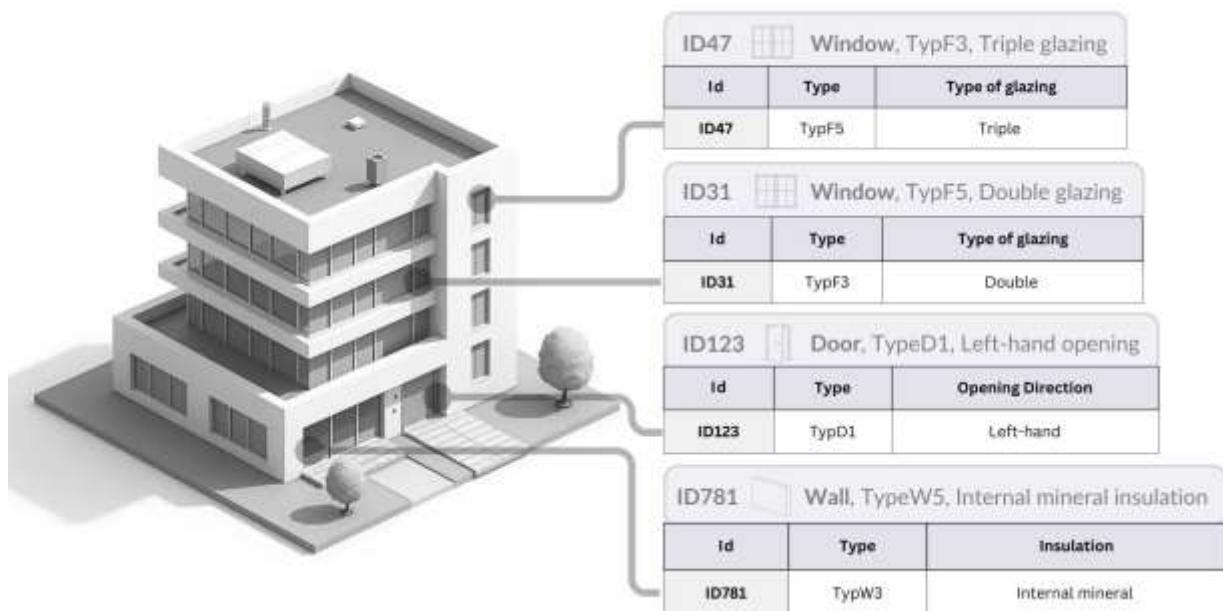
- **Geometrijski parametri** - opisujejo obliko, lokacijo in dimenzijske predmetne. Geometrijo in njeni uporabo bomo podrobno obravnavali v šestem delu knjige, ki je namenjen rešitvam CAD (BIM);
- **Lastnosti atributov** - vsebujejo različne informacije: materiale, vrste elementov, tehnične značilnosti, edinstvene identifikatorje in druge lastnosti, ki jih lahko imajo elementi projekta.

Podatki o atributih so v sodobnih projektih še posebej pomembni, saj določajo operativne značilnosti objektov, omogočajo inženirske in stroškovne izračune ter zagotavljajo celovito interakcijo med udeleženci pri načrtovanju, gradnji in obratovanju. Na primer:

- Za okna in vrata: vrsta konstrukcije, vrsta zasteklitve, smer odpiranja (slika 3.2-1).
- Pri stenah se zabeležijo podatki o materialih, topotni izolaciji in akustičnih lastnostih.
- Za inženirske sisteme so shranjeni parametri cevovodov, kanalov, kabelskih tras in njihovih priključkov.

Ti parametri so lahko shranjeni v samih datotekah CAD-(BIM) in v zunanjih podatkovnih zbirkah - kot rezultat izvoza, pretvorbe ali neposrednega dostopa do notranjih struktur CAD prek orodij za povratni inženiring. Ta pristop olajša integracijo informacij o projektiranju z drugimi podjetniškimi sistemmi in platformami

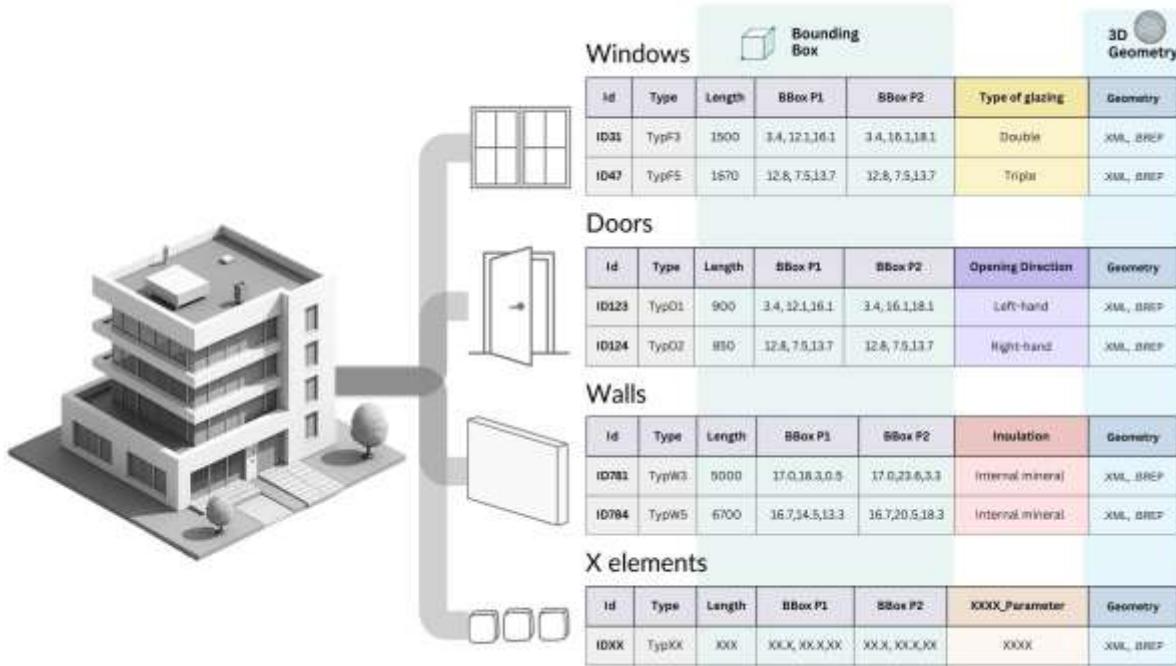
**Povratni inženiring** v okviru CAD (BIM) je postopek pridobivanja in analiziranja notranje strukture digitalnega modela, da bi ponovno ustvarili njegovo logiko, podatkovno strukturo in odvisnosti brez dostopa do izvirnih algoritmov ali dokumentacije.



Slika 3.1-15 Projektni element poleg opisa parametrične ali poligonalne geometrije vsebuje informacije o parametrih in lastnostih elementov.

Tako se okoli vsakega elementa oblikuje edinstven nabor parametrov in lastnosti, vključno z edinstvenimi značilnostmi vsakega predmeta (npr. identifikator in dimenzije) ter skupnimi atributi za skupine elementov. To omogoča ne le analizo posameznih elementov-subjektov projekta, temveč tudi njihovo združevanje v logične skupine, ki jih lahko nato uporabljajo drugi strokovnjaki za svoje naloge in izračune v sistemih in podatkovnih zbirkah.

Entiteta je konkreten ali abstrakten predmet realnega sveta, ki ga je mogoče enolično prepoznati, opisati in predstaviti v obliki podatkov.

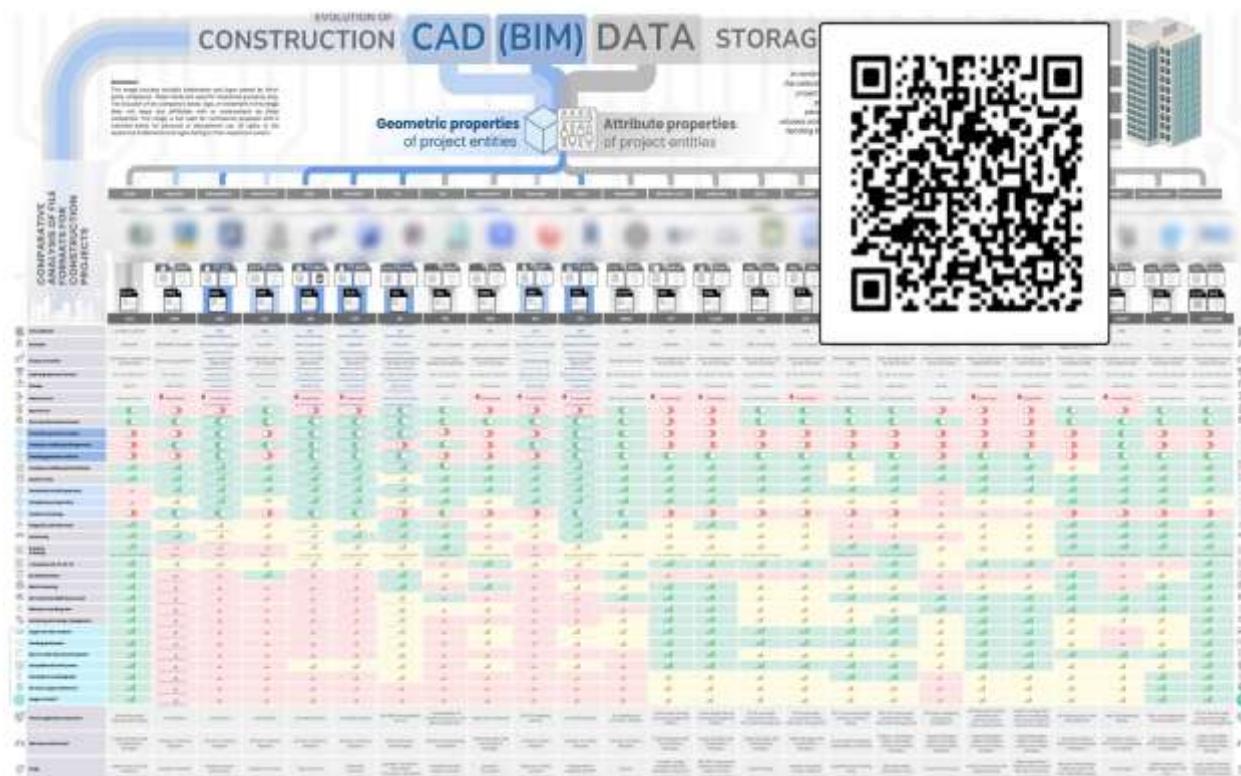


Slika 3.1-16 Vsak element projekta vsebuje atribute, ki jih vnese projektant ali pa se izračunajo v programu CAD.

V zadnjih desetletjih je gradbena industrija razvila številne nove formate CAD (BIM), ki poenostavljajo ustvarjanje, shranjevanje in prenos podatkov. Ti formati so lahko zaprti ali odprtji, tabelarni, parametrični ali grafični. Vendar njihova raznolikost in razdrobljenost znatno otežuje upravljanje podatkov v vseh fazah življenjskega cikla projekta. Primerjalna tabela glavnih formatov, ki se uporabljajo za izmenjavo informacij v gradbeništvu, je predstavljena na sliki 3.1-17 (celotna različica je na voljo s kodo QR).

Za reševanje težav z interoperabilnostjo in dostopom do podatkov CAD so vključeni upravljavci BIM (BIM) in koordinatorji, katerih naloga je nadzor izvoza, preverjanje kakovosti podatkov in vključevanje delov podatkov CAD (BIM) v druge sisteme.

Vendar je zaradi zaprte narave in zapletenosti formatov ta postopek težko avtomatizirati, zaradi česar morajo strokovnjaki številne operacije izvajati ročno, ne da bi lahko vzpostavili celovite postopke obdelave podatkov v liniji (pipeline).



Slika 3.1-17 Preglednica s primerjavo glavnih podatkovnih formatov, v katerih so shranjene informacije o elementih projekta [53].

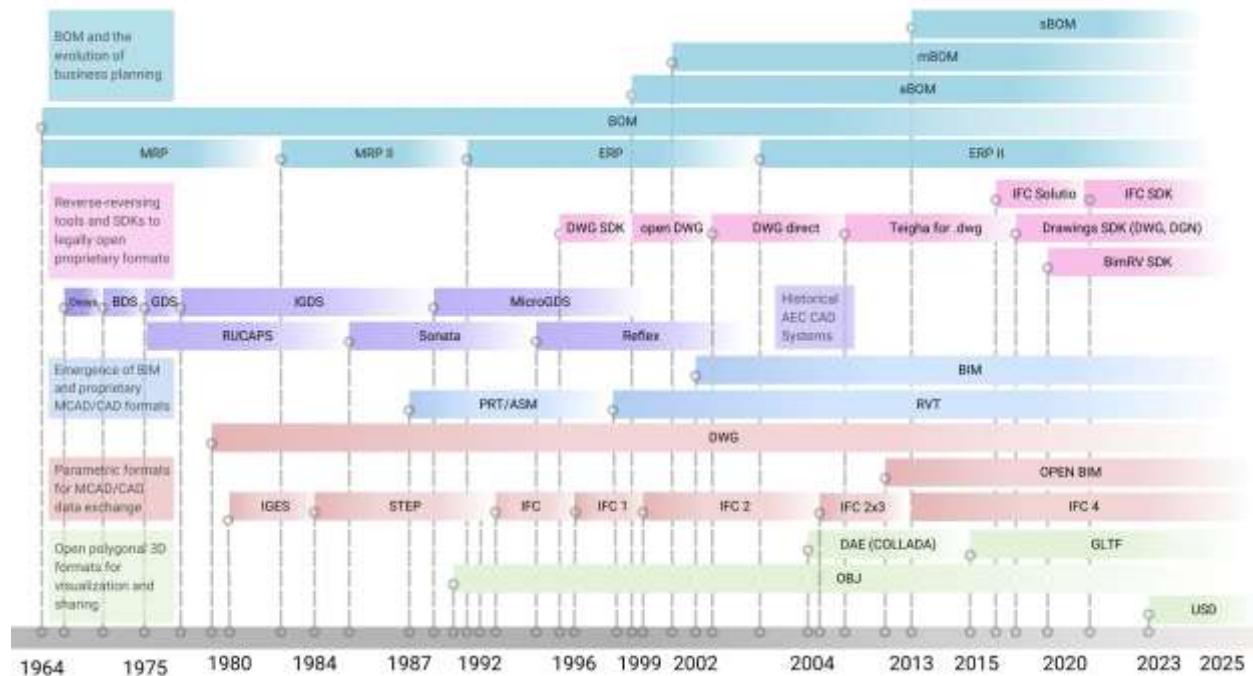
Da bi razumeli, zakaj obstaja toliko različnih podatkovnih formatov in zakaj je večina od njih zaprtih, se je treba poglobiti v procese, ki potekajo znotraj programov CAD (BIM), ki jih bomo podrobno raziskali v šestem delu knjige.

Razvijalci sistemov CAD so geometriji dodali dodatno informacijsko plast v obliki koncepta BIM (Building Information Modeling), marketinškega izraza, ki se v gradbeni industriji aktivno promovira od leta 2002 [54].

## Pojav koncepta BIM (BOM) in uporaba CAD v procesih

Koncept informacijskega modeliranja stavb (BIM), ki je bil prvič predstavljen v beli knjigi BIM iz leta 2002 [54], izhaja iz tržnih pobud proizvajalcev programske opreme CAD. Nastal je iz trženjskih pobud razvijalcev programske opreme CAD in je bil poskus prilagoditve načel, ki so se že dobro uveljavila v strojništву, potrebam gradbene industrije.

Navdih za BIM je prišel iz koncepta BOM (Bill of Materials) - seznama materialov, ki se v industriji pogosto uporablja že od konca osemdesetih let prejšnjega stoletja. V strojništvu je BOM omogočal povezovanje podatkov iz sistemov CAD s sistemi PDM (Product Data Management), PLM (Product Lifecycle Management) in ERP, kar je zagotovljalo celovito upravljanje inženirskih informacij v celotnem življenjskem ciklu izdelka (slika 3.1-8).



Slika 3.1-18 Razvoj specifikacij (BOM), informacijskega modeliranja (BIM) in digitalnih formatov v inženirski gradbeni industriji.

Sodobni razvoj koncepta BOM je privdel do pojava razširjenega okvira - XBOM (Extended BOM), ki poleg sestave izdelka vključuje tudi vedenjske scenarije, operativne zahteve, trajnostne parametre in podatke za napovedno analitiko. XBOM ima v bistvu enako vlogo kot BIM v gradbeništvu: oba pristopa si prizadevata digitalni model spremeniti v enotni vir resnice za vse udeležence projekta v celotnem življenjskem ciklu projekta.

Ključni mejnik pri pojavu BOM v gradbeništvu je bila uvedba prvega parametričnega CAD (MCAD), posebej prilagojenega za gradbeno industrijo, leta 2002. Razvila ga je ekipa, ki je pred tem ustvarila Pro-E®, revolucionarni sistem MCAD za strojništvo, ki je nastal konec osemdesetih let prejšnjega stoletja in je postal industrijski standard [55].

Že konec osemdesetih let prejšnjega stoletja je bil cilj odpraviti omejitve [56] takrat obstoječih programov CAD. Glavni cilj je bil zmanjšati delo, potrebno za spremembe parametrov konstrukcijskih elementov, in omogočiti posodabljanje modela na podlagi podatkov zunaj programov CAD prek podatkovne baze [57]. Najpomembnejšo vlogo pri tem naj bi imela parametrizacija: samodejno pridobivanje lastnosti iz podatkovne baze in njihova uporaba za posodobitev modela znotraj sistemov CAD.

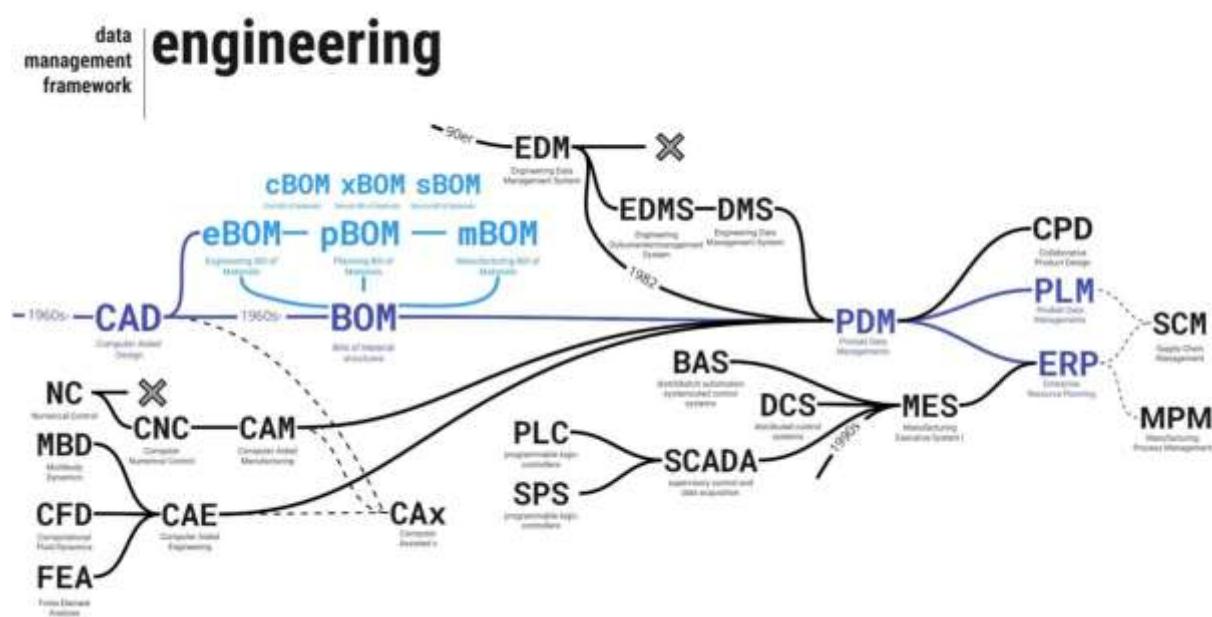
Pro-E in koncept elementarnega parametričnega modeliranja c BOM, na katerem temelji, sta pomembno vplivala na razvoj trga CAD - in MCAD - [58]. Ta model je v industriji prisoten že 25 let, številni sodobni sistemi pa so postali njegovi konceptualni nasledniki.

Cilj je ustvariti sistem, ki bo dovolj prilagodljiv, da bo inženirja spodbudil, da bo zlahka preučil različne zasnove. Stroški sprememb zasnove pa morajo biti čim bliže ničli.

Tradicionalna programska oprema CAD / CAM nerealistično omejuje izvajanje poceni sprememb le na samem začetku procesa načrtovanja [59].

- Samuel Heisenberg, ustanovitelj podjetja Parametric Technology Corporation®, razvijalec MCAD -izdelka Pro-E in učitelj ustvarjalec izdelka CAD, ki uporablja format RVT

V strojništvu so sistemi PDM, PLM, MRP in ERP postali ključne platforme. Imajo osrednjo vlogo pri upravljanju podatkov in procesov, zbiranju informacij iz sistemov CAx (CAD, CAM, CAE) in organiziranju dejavnosti načrtovanja na podlagi strukture izdelka (BOM: eBOM, pBOM, mBOM) (slika 3.1-18). Ta integracija zmanjšuje število napak, preprečuje podvajanje podatkov in zagotavlja celovito sledljivost od načrtovanja do proizvodnje.



Slika 3.1-19 Zgodovinsko gledano se je sistem BOM pojavil v šestdesetih letih prejšnjega stoletja kot način za strukturiranje podatkov iz sistemov CAx in njihovo posredovanje nadzornim sistemom.

Nakup rešitve CAD, ki jo je razvila nekdanja ekipa Pro-E in je temeljila na pristopu BOM, s strani enega od vodilnih prodajalcev je skoraj takoj zaznamovala objava serije belih knjig BIM (2002-2003)[60][61]. Že od sredine leta 2000 se je koncept BIM začel aktivno spodbujati v gradbeništvu, kar je občutno povečalo zanimanje za parametrično programsko opremo. Priljubljenost je rasla tako hitro, da je gradbena vilica za strojništvo Pro-E - parametrični CAD, ki ga spodbuja ta proizvajalec - dejansko izpodrinila konkurenco v segmentu arhitekturnega in konstrukcijskega projektiranja (slika 3.1-20). Do začetka leta 2020 je de facto utrdil globalno prevlado na trgu BIM (CAD) [62].



Slika 3.1-20 Priljubljenost iskalnih poizvedb na Googlu (RVT v primerjavi z IFC): parametrični CAD, ki ga je ustvarila nekdanja ekipa Pro-E s podporo BOM -BIM, je postal priljubljen v skoraj večini držav sveta.

V zadnjih 20 letih je kratica BIM dobila številne razlage, katerih več pomenov izhaja iz začetnih marketinških konceptov, ki so se pojavili v začetku leta 2000. Standard ISO 19650, ki je imel pomembno vlogo pri popularizaciji izraza, je pravzaprav zagotovil status BIM kot "znanstveno utemeljenega" pristopa k upravljanju informacij. Vendar je v samem besedilu standarda, ki je posvečeno upravljanju podatkov v celotnem življenjskem ciklu objektov z uporabo BIM, kratica BIM omenjena, vendar nikoli jasno opredeljena

Prvotna spletna stran prodajalca, na kateri je bila leta 2002 [60] in 2003 [61] objavljena serija belih knjig o BIM, je dejansko povzemala marketinško gradivo o konceptih BOM (Bills of Materials) in PLM (Product Lifecycle Management), ki so bili v 90. letih prejšnjega stoletja uporabljeni v strojniški programske opremi Pro-E [63].

*Informacijsko modeliranje stavb, inovativen nov pristop k načrtovanju, gradnji in upravljanju stavb, ki ga je leta 2002 predstavil..... [ime podjetja, ki izdeluje CAD], je spremenil način razmišljanja strokovnjakov po vsem svetu o tem, kako je mogoče tehnologijo uporabiti pri načrtovanju, gradnji in upravljanju stavb.*

- Bela knjiga BIM, 2003 [61]

Te zgodne publikacije so BIM neposredno povezovale s konceptom centralizirane integrirane podatkovne zbirke. Kot je navedeno v Beli knjigi iz leta 2003, je BIM upravljanje informacij o stavbi, pri katerem se vse posodobitve izvajajo v enem samem skladišču, kar zagotavlja sinhronizacijo vseh risb, kosov in specifikacij (BOM - Bills of Materials).

*BIM je opisan kot upravljanje informacij o stavbi, pri katerem se vse posodobitve in spremembe izvajajo v podatkovni zbirki. Ne glede na to, ali imate opravka s shemami, prerezi ali risbami listov, je vse vedno usklajeno, dosledno in posodobljeno.*

- Spletna stran podjetja CAD prodajalec z BIM Whitepaper, 2003 [54]

Zamisel o upravljanju oblikovanja z enotno integrirano podatkovno zbirko je bila pogosto obravnavana že v raziskavah v 80. letih prejšnjega stoletja. Na primer, koncept BDS Charlesa Eastmana [57] je vseboval 43 omemb izraza "podatkovna baza" (slika 6.1-2). Do leta 2004 se je to število skoraj prepolovilo na 23 v Beli knjigi o BIM iz leta 2002 [64]. Do sredine leta 2000 pa je tema podatkovnih zbirk praktično izginila iz marketinškega gradiva prodajalcev in agende digitalizacije na splošno.

Čeprav sta bila podatkovna zbirka in dostop do nje prvotno zasnovana kot jedro sistema BIM, se je sčasoma poudarek preusmeril na geometrijo, vizualizacijo in 3D. Sam registrar standarda IFC iz leta 1994, ki je leta 2002 objavil BIM Whitepaper - isti prodajalec -, je v Whitepaperju z začetka leta 2000 izrecno opozoril na omejitve nevtralnih formatov, kot so IGES, STEP in IFC, ter na potrebo po neposrednem dostopu do podatkovnih zbirk CAD:

*Različne aplikacije so lahko nezdružljive, ponovno vneseni podatki pa netočni [...]. Rezultat tradicionalnega računalniško podprtga načrtovanja [CAD]: višji stroški, daljši čas uvajanja na trg in slabša kakovost izdelkov. Danes vse glavne aplikacije uporabljajo standardne vmesnike za izmenjavo podatkov na nizki ravni. Z uporabo starih standardov IGES ali novih STEP [IFC je dejanska in pravna kopija formata STEP/IGES] za izmenjavo podatkov med aplikacijami različnih proizvajalcev lahko uporabniki dosežejo določeno združljivost podatkov med najboljšimi izdelki. Vendar IGES in STEP delujeta le na nizkih ravneh in ne moreta izmenjevati tako bogatih podatkov, kot so podatki, ki jih ustvarjajo današnje vodilne aplikacije [...]. In čeprav se ti in drugi standardi skoraj vsak dan izboljšujejo, bodo glede bogastva podatkov vedno zaostajali za današnjimi izdelki proizvajalcev. [...] Programi znotraj aplikacije morajo biti sposobni izmenjevati in ohranjati bogastvo podatkov, ne da bi se zatekali k nevtralnim prevajalnikom, kot so IGES, STEP [IFC] ali PATRAN. Namesto tega morajo imeti okvirne aplikacije možnost neposrednega dostopa do osnovne podatkovne zbirke CAD, da se podrobnosti in natančnost informacij ne izgubijo.*

- Bela knjiga proizvajalca CAD (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale", 2000 [65].

Tako je že v osemdesetih letih prejšnjega stoletja in na začetku novega tisočletja veljalo, da je ključni element digitalnega oblikovanja v okolju CAD podatkovna zbirka in ne datoteka formata ali nevtralni format IFC. Predlagano je bilo, da se opustijo prevajalniki in da imajo aplikacije neposreden dostop do podatkov. V resnici pa je do sredine 20. let prejšnjega stoletja koncept BIM začel spominjati na strategijo "deli in vladaj", kjer so interesi proizvajalcev programske opreme, ki uporabljajo zaprta

geometrijska jedra, pomembnejši od razvoja odprte izmenjave informacij.

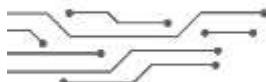
BIM danes velja za sestavni del gradbene industrije. Toda v zadnjih dveh desetletjih so se obljube o poenostavljenem sodelovanju in povezovanju podatkov večinoma ne uresničile. Večina rešitev je še vedno vezana na zaprte ali nevtralne formate in specializirana orodja. Zgodovino BIM, odprtga BIM in IFC ter vprašanja interoperabilnosti in geometrijskih jader si bomo podrobno ogledali v 6. delu knjige "CAD in BIM: trženje, realnost in prihodnost projektnih podatkov v gradbeništvu".

Industrija se danes sooča s ključnim izzivom, kako preiti od tradicionalnega razumevanja CAD (BIM) kot orodja za modeliranje k njegovi uporabi kot polnopravne podatkovne zbirke. To zahteva nove pristope k delu z informacijami, opustitev odvisnosti od zaprtih ekosistemov in uvajanje odprtih rešitev.

Z razvojem orodij za povratni inženiring, ki omogočajo dostop do podatkovnih zbirk CAD, ter širjenjem odprtih tehnologij in tehnologij LLM se uporabniki in razvijalci v gradbeništvu vse bolj oddaljujejo od nejasnih pogojev prodajalcev programske opreme. Namesto tega se osredotočajo na tisto, kar je resnično pomembno: na podatke (podatkovne zbirke) in procese.

Za modnimi kraticami in vizualizacijami se skrivajo standardne prakse upravljanja podatkov: shranjevanje, prenos in preoblikovanje - tj. klasični postopek ETL (Extract, Transform, Load). Tako kot v drugih panogah tudi digitalizacija gradbeništva ne zahteva le standardov izmenjave, temveč tudi jasno strukturirano ravnanje s heterogenimi informacijami.

Da bi v celoti izkoristili potencial podatkov CAD (BIM), morajo podjetja ponovno premisliti o svojem pristopu k upravljanju informacij. To bo neizogibno privedlo do ključnega elementa digitalne preobrazbe - poenotenja, standardizacije in smiselnega strukturiranja podatkov, s katerimi gradbeni strokovnjaki vsakodnevno delajo.

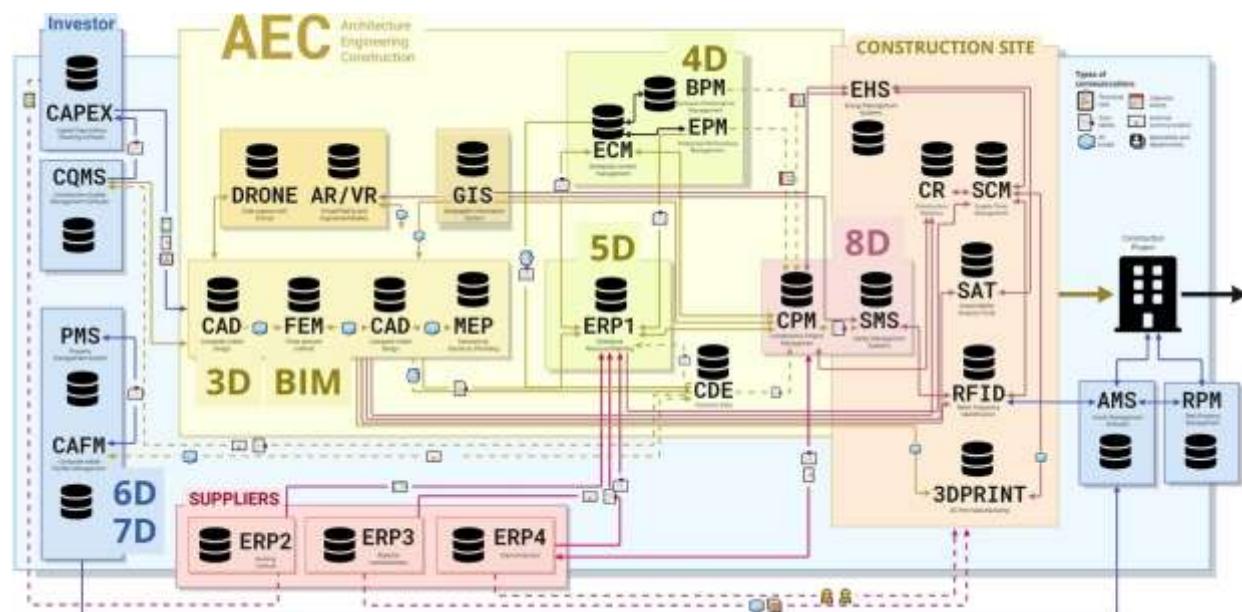


## POGLAVJE 3.2.

# USKLAJEVANJE IN STRUKTURIRANJE PODATKOV

### Polnjenje sistemov s podatki v gradbeništvu

Ne glede na to, ali gre za velike korporacije ali srednje velika podjetja, se strokovnjaki dnevno ukvarjajo s polnjenjem programskih sistemov in podatkovnih zbirk z različnimi vmesniki z večplastnimi informacijami (slika 3.2-1), ki morajo s pomočjo vodij medsebojno sodelovati. Prav ta kompleks medsebojno delujočih sistemov in procesov na koncu ustvarja prihodek in dobiček podjetja.



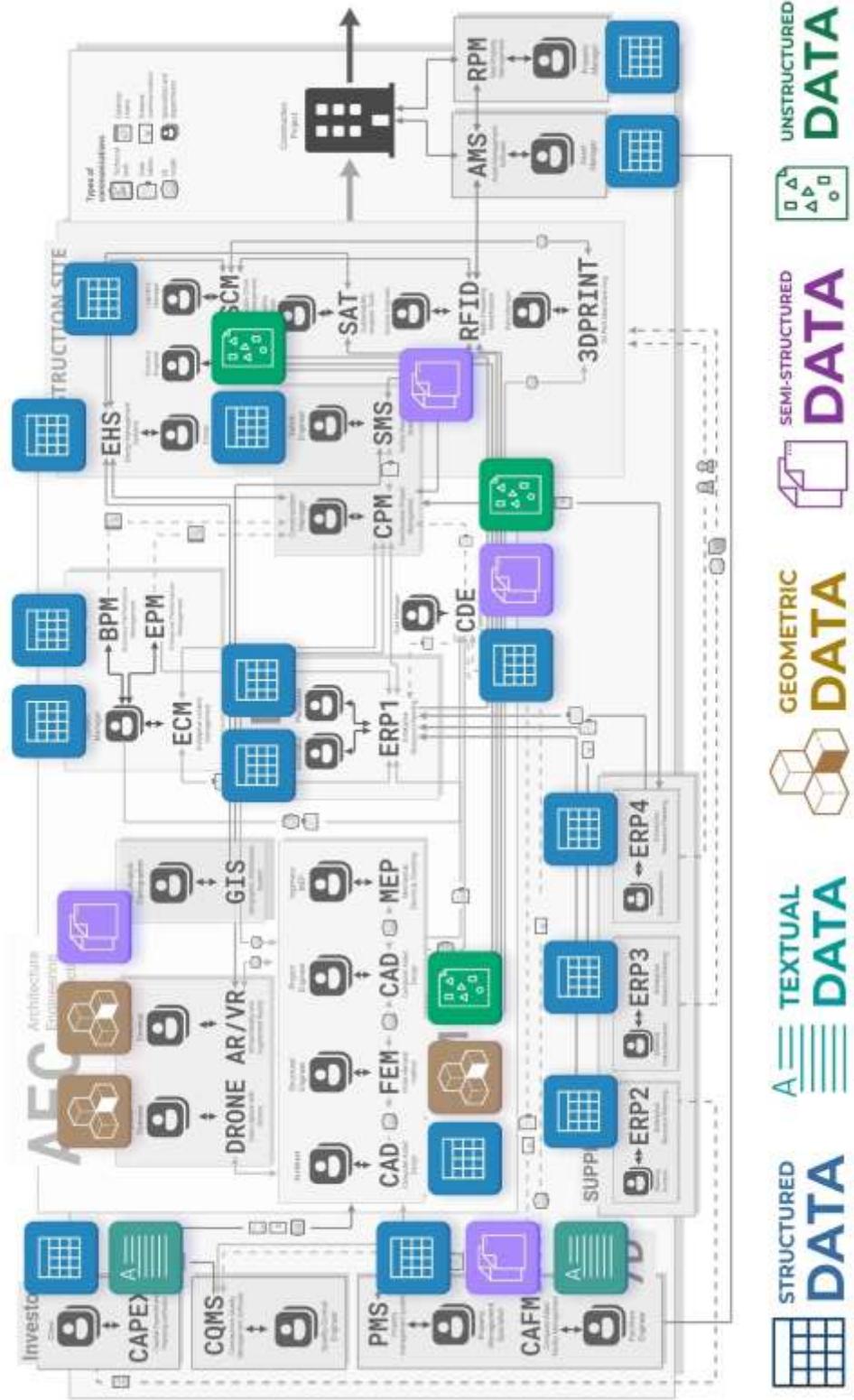
Slika 3.2-1 Skoraj vsak sistem ali aplikacija, ki se uporablja v gradbeništvu, ima v svojem jedru eno od priljubljenih podatkovnih zbirk RDBMS.

Vsaka od prej omenjenih kategorij sistemov, ki se uporabljajo v gradbeništvu, deluje s svojimi podatkovnimi tipi, ki ustrezajo funkcionalni vlogi teh sistemov. Da bi prešli z abstraktne ravni na konkretno, preidemo od podatkovnih tipov k njihovi predstavitvi v obliki formatov in dokumentov.

Prej navedenemu seznamu sistemov (slika 1.2-4) zdaj dodajamo posebne vrste formatov in dokumentov, s katerimi pogosto delajo:

#### **Vlagatelj (CAPEX)**

- Finančni podatki: proračuni, napovedi odhodkov (strukturirani podatki).
- Podatki o tržnih trendih: tržne analize (strukturirani in nestrukturirani podatki).
- Pravni in pogodbeni podatki: pogodbe (besedilni podatki).



Slika 3.2-2 Gradbena industrija uporablja številne sisteme z različnimi vmesniki, ki obravnavajo različne vrste podatkov.

### **Sistemi upravljanja (PMS, CAFM, CQMS)**

- Podatki o projektu: grafikoni, naloge (strukturirani podatki).
- Podatki o vzdrževanju objektov: načrti vzdrževanja (besedilni in polstrukturirani podatki).
- Podatki o nadzoru kakovosti: standardi, poročila o pregledih (besedilni in nestrukturirani podatki).

### **CAD, FEM in BIM**

- Tehnične risbe: arhitekturni in strukturni načrti (geometrijski podatki, nestrukturirani podatki).
- Modeli za gradnjo: 3D - modeli, podatki o materialih (geometrijski in polstrukturirani podatki).
- Inženirski izračuni: analiza obremenitve (strukturirani podatki).

### **Sistemi upravljanja gradbišča (EHS, SCM)**

- Podatki o varnosti in zdravju: varnostni protokoli (besedilni in strukturirani podatki).
- Podatki o dobavni verigi: zaloge, naročila (strukturirani podatki).
- Dnevna poročila: delovni čas, produktivnost (strukturirani podatki).

### **Droni, AR/VR, GIS, 3D - tiskanje**

- Geopodatki: topografski zemljevidi (geometrijski in strukturirani podatki).
- Podatki v realnem času: videoposnetki in fotografije (nestrukturirani podatki).
- Modeli za 3D -tiskanje: digitalne risbe (geometrijski podatki).

### **Dodatni sistemi upravljanja (4D BPM, 5D ERP1)**

- Podatki o času in stroških: urniki, ocene (strukturirani podatki).
- Upravljanje sprememb: zapisi o spremembah projekta (besedilo in strukturirani podatki).
- Poročanje o uspešnosti: kazalniki uspeha (strukturirani podatki).

### **Integracija in komunikacija podatkov (CDE, RFID, AMS, RPM)**

- Izmenjava podatkov: izmenjava dokumentov, podatkovni modeli (strukturirani in besedilni podatki).
- RFID in podatki za sledenje: logistika, upravljanje sredstev (strukturirani podatki).
- Spremljanje in nadzor: senzorji na lokacijah (strukturirani in nestrukturirani podatki).

Tako vsak sistem v gradbeništvu - od sistemov za upravljanje gradbišč do operativnih podatkovnih zbirk - deluje s svojo vrsto informacij: strukturiranimi, besedilnimi, geometrijskimi in drugimi. "Podatkovna pokrajina", s katero morajo strokovnjaki vsakodnevno delati, je zelo raznolika. Vendar pa preprosto naštevanje formatov ne razkrije kompleksnosti resničnega dela z informacijami.

V praksi se podjetja soočajo z dejstvom, da podatki, tudi ko so pridobljeni iz sistemov, niso pripravljeni za uporabo "takšni, kot so". To še posebej velja za besedila, slike, datoteke PDF, datoteke CAD in druge formate, ki jih je težko analizirati s standardnimi orodji. Zato je naslednji ključni korak preoblikovanje podatkov - proces, brez katerega ni mogoče učinkovito avtomatizirati obdelave, analize, vizualizacije in odločanja.

## **Preoblikovanje podatkov: ključni temelj sodobne poslovne analize**

Danes se večina podjetij sooča s paradoksom: približno 80% njihovih vsakodnevnih procesov še vedno

temelji na klasičnih strukturiranih podatkih - znanih Excelovih preglednicah in relacijskih podatkovnih bazah (RDBMS) [66]. Hkrati pa je 80% novih informacij, ki vstopajo v digitalni ekosistem podjetij, nestrukturiranih ali ohlapno strukturiranih (slika 3.2-3) [52]. To vključuje besedilo, grafiko, geometrijo, slike, modele CAD, dokumentacijo v formatu PDF, zvočne in video posnetke, elektronsko korespondenco in še veliko več.

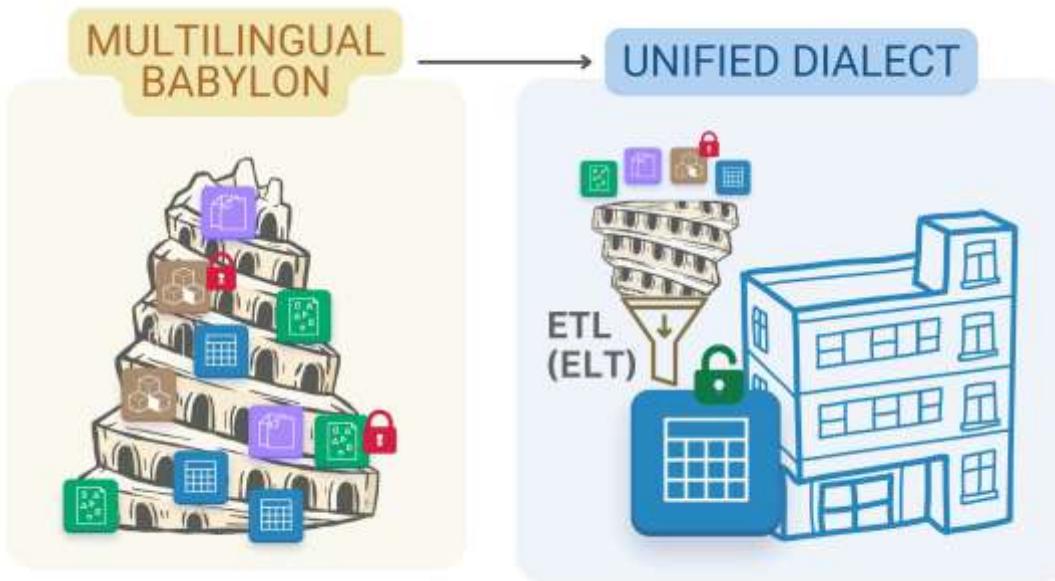
Poleg tega količina nestrukturiranih podatkov še naprej hitro narašča - letna stopnja rasti je ocenjena na 55-65% [67]. Takšna dinamika povzroča resne težave pri vključevanju novih informacij v obstoječe poslovne procese. Neupoštevanje tega toka večplastnih podatkov vodi v nastanek informacijskih vrzeli in zmanjšuje obvladljivost celotnega digitalnega okolja podjetja.



Slika 3.2-3 Vsakiletna rast nestrukturiranih podatkov povzroča izzive pri vključevanju pretočnih informacij v poslovne procese.

Neupoštevanje zapletenih nestrukturiranih in zmedenih ohlapno strukturiranih podatkov v procesih avtomatizacije lahko privede do velikih vrzeli v informacijski krajini podjetja. V današnjem svetu nenadzorovanega in lavinskega gibanja informacij morajo podjetja sprejeti hibridni pristop k upravljanju podatkov, ki vključuje učinkovite metode za ravnanje z vsemi vrstami podatkov.

Ključ do učinkovitega upravljanja podatkov je v organiziranju, strukturiraju in razvrščanju različnih vrst podatkov "Babel" (vključno z nestrukturiranimi, besedilnimi in geometrijskimi oblikami, v strukturirane ali ohlapno strukturirane podatke). Ta proces preoblikuje kaotične podatkovne nize v organizirane strukture za vključevanje v sisteme in s tem omogoča sprejemanje odločitev na njihovi podlagi (slika 3.2-4).



Slika 3.2-4 Glavna naloga oddelkov za upravljanje podatkov je, da "babilon" različnih in večplastnih podatkov prevedejo v strukturiran in kategoriziran sistem.

Ena od ključnih ovir za takšno usklajevanje ostaja nizka raven interoperabilnosti med različnimi digitalnimi platformami - "silosi", o katerih smo govorili v prejšnjih poglavjih.

Nacionalni inštitut za standarde in tehnologijo (NIST, ZDA) v poročilu poudarja [68], da slaba združljivost podatkov med različnimi gradbenimi platformami povzroča izgubo informacij in velike dodatne stroške. Samo leta 2002 so težave z interoperabilnostjo programske opreme povzročile izgube v investicijski gradnji v ZDA v skupni vrednosti 15,8 milijarde dolarjev na leto, pri čemer dve tretjini teh izgub nosijo lastniki in upravljavci stavb, zlasti med obratovanjem in vzdrževanjem [68]. Študija tudi ugotavlja, da lahko standardizacija podatkovnih formatov zmanjša te izgube in izboljša učinkovitost v celotnem življenjskem ciklu objekta.

Glede na študijo CrowdFlower iz leta 2016 [69], ki je zajela 16 000 podatkovnih znanstvenikov po vsem svetu, glavna težava ostajajo "umazani" in večformatni podatki. Po tej študiji najdragocenjeji vir niso končne zbirke podatkov ali modeli strojnega učenja, temveč čas, porabljen za pripravo podatkov.

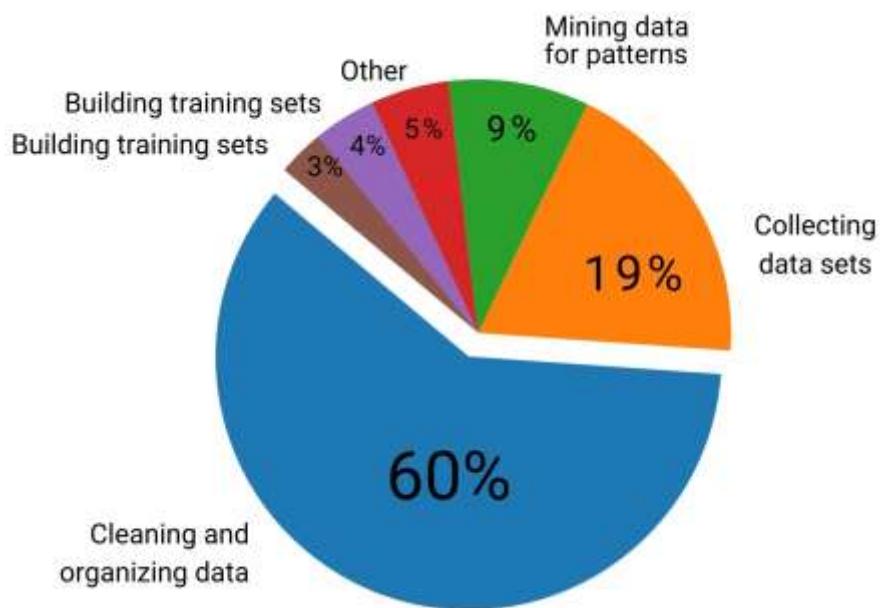
Čiščenje, oblikovanje in urejanje vzame do 60 odstotkov časa analitika in upravljavca podatkov. Skoraj petino porabi za iskanje in zbiranje ustreznih podatkovnih nizov, ki so pogosto skriti v silosih in nedostopni za analizo. In le približno 9 odstotkov časa je porabljenega neposredno za modeliranje, analitiko, napovedovanje in testiranje hipotez. Preostali del je namenjen komuniciranju, vizualizaciji, poročanju in iskanju podpornih virov informacij.

V povprečju je delo s podatki, ki ga opravlja vodja, razporejeno na naslednji način (slika 3.2-5):

- **Čiščenje in urejanje podatkov (60%):** čisti in strukturirani podatki lahko bistveno skrajšajo delovni čas analitika in pospešijo postopek opravljanja nalog.
- **Zbiranje podatkov (19%):** Velik izziv za strokovnjake s področja podatkovnih znanosti je

iskanje ustreznih zbirk podatkov. Pogosto so podatki podjetij zbrani v kaotično urejenih "silosih", kar otežuje dostop do informacij, ki jih potrebujejo.

- **Modeliranje/strojno učenje (9%):** Pogosto ovira pomanjkanje jasnosti poslovnih ciljev s strani strank. Pomanjkanje jasnega poslanstva lahko izniki potencial še takoj dobrega modela.
- **Druge naloge (5%):** poleg obdelave podatkov se analitiki ukvarjajo z raziskavami, raziskovanjem podatkov z različnih vidikov, sporočanjem rezultatov z vizualizacijami in poročili ter priporočanjem optimizacije procesov in strategij.



Slika 3.2-5 Za kaj upravljavci podatkov, ki se ukvarjajo s podatki, porabijo največ časa (na podlagi [70]).

Te ocene potrjujejo tudi druge študije. Glede na študijo Xplenty, objavljeno v reviji BizReport leta 2015 [71], med 50% in 90% časa strokovnjakov za poslovno obveščanje (BI) (BI) porabijo za pripravo podatkov za analizo.

Čiščenje, potrjevanje in urejanje podatkov na splettem mestu je ključna podlaga za vse nadaljnje podatkovne in analitične postopke, saj podatkovnim znanstvenikom vzame do 90% časa.

To skrbno delo, ki je končnemu uporabniku nevidno, je ključnega pomena. Napake v neobdelanih podatkih neizogibno izkrivljajo analize, so zavajajoče in lahko vodijo do dragih napak pri upravljanju. Zato postopki čiščenja in standardizacije podatkov - od odpravljanja podvajanj in dopolnjevanja izpustov do usklajevanja merskih enot in prilaganja skupnemu modelu - postajajo temelj sodobne digitalne strategije.

Temeljito preoblikovanje, čiščenje in standardizacija podatkov tako ne zavzemajo le večine časa strokovnjakov (do 80% dela s podatki), temveč določajo tudi možnost njihove učinkovite uporabe v okviru sodobnih poslovnih procesov. Vendar pa samo organizacija in čiščenje podatkov ne izčrpata

naloge optimalnega upravljanja informacijskih tokov podjetja. V fazi organiziranja in strukturiranja postane pomembna izbira ustreznega podatkovnega modela, ki neposredno vpliva na priročnost in učinkovitost dela z informacijami v naslednjih fazah obdelave.

Ker so podatki in poslovni cilji različni, je pomembno razumeti značilnosti podatkovnih modelov in znati izbrati ali ustvariti pravo strukturo. Glede na stopnjo strukturiranosti in način opisovanja odnosov med elementi obstajajo trije glavni modeli: strukturirani, ohlapno strukturirani in grafični. Vsak je primeren za različne naloge ter ima svoje prednosti in slabosti.

## **podatkovni modeli: razmerja v podatkih in razmerja med elementi**

Podatki v informacijskih sistemih so organizirani na različne načine - odvisno od nalog in zahtev za shranjevanje, obdelavo in prenos informacij. Ključna razlika med vrstami podatkovnih modelov, torej obliko, v kateri so informacije shranjene, je v stopnji strukturiranosti in načinu opisa odnosov med elementi.

Strukturirani podatki imajo jasno in ponovljivo shemo: organizirani so kot tabele z določenimi stolpcji. Ta oblika zagotavlja predvidljivost, enostavnost obdelave in učinkovitost pri izvajanju poizvedb SQL, filtriranju in agregirjanju. Primeri: podatkovne zbirke (RDBMS), Excel, CSV.

Ohlapno strukturirani podatki omogočajo prilagodljivo strukturo: različni elementi lahko vsebujejo različen nabor atributov in so shranjeni kot hierarhije. Primeri so JSON, XML ali drugi formati dokumentov. Ti podatki so priročni, kadar je treba modelirati vgnezdené objekte in odnose med njimi, po drugi strani pa otežujejo analizo in standardizacijo podatkov (slika 3.2-6).

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

**Slika 3.2-6 Podatkovni model je logična struktura, ki opisuje, kako so podatki organizirani, shranjeni in obdelani v sistemu.**

Izbira ustrezne oblike je odvisna od ciljev:

- Če je pomembna hitrost filtriranja in analitike - zadostujejo relacijske tabele (SQL, CSV, RDBMS, stolpčne podatkovne zbirke).

- Če je potrebna prilagodljivost strukture, je bolje uporabiti JSON ali XML.
- Če imajo podatki zapletene odnose, grafne podatkovne zbirke zagotavljajo preglednost in skalabilnost.

V klasičnih relacijskih podatkovnih zbirkah (RDBMS) je vsaka entiteta (npr. vrata) predstavljena z vrstico, njene lastnosti pa s stolci tabele. Na primer, tabela elementov iz kategorije "Vrata" lahko vsebuje polja ID, Višina, Širina, Požarna odpornost in ID prostora, ki označuje prostor (slika 3.2-7).

V klasičnih relacijskih podatkovnih zbirkah (RDBMS) so relacije oblikovane v oblikih tabel, kjer vsak zapis predstavlja objekt, stolpci pa njegove parametre. V tabelarični obliku so podatki o vratih v projektu videti takole, kjer vsaka vrstica predstavlja ločen element - vrata z edinstvenim identifikatorjem in atributi, povezava s sobo pa je izvedena prek parametra "ID sobe".



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

Slika 3.2-7 Informacije o treh elementih kategorije "Vrata" projekta v tabelarični strukturirani obliku.

V ohlapno strukturiranih formatih, kot sta JSON ali XML, so podatki shranjeni v hierarhični ali gnezdeni oblik, pri čemer lahko elementi vsebujejo druge objekte, njihova struktura pa se lahko spreminja. To omogoča modeliranje zapletenih razmerij med elementi. Podobne informacije o vratih v projektu, ki so bile zapisane v strukturirani obliku (slika 3.2-7), so predstavljene v ohlapno strukturirani obliku (JSON) tako (slika 3.2-8), da postanejo ugnezdeni objekti znotraj sob (Rooms - ID), kar logično odraža hierarhijo.

```

1 {
2   "Rooms": [
3     {
4       "ID": 101,
5       "Doors": [
6         {"ID": 1, "Height": 2000, "Width": 900, "Fireproof": "Yes"},
7         {"ID": 2, "Height": 2100, "Width": 800, "Fireproof": "No"}
8       ]
9     },
10    {
11      "ID": 102,
12      "Doors": [
13        {"ID": 3, "Height": 2000, "Width": 850, "Fireproof": "Yes"}
14      ]
15    }
16  ]
17}
18

```

Slika 3.2-8 Informacije o elementih kategorije "Vrata" projekta v obliki JSON.

V modelu grafa so podatki predstavljeni kot vozlišča (vrhovi) in povezave (robovi) med njimi. To omogoča vizualizacijo zapletenih razmerij med predmeti in njihovimi atributi. V primeru podatkov o vratih in sobah v projektu je predstavitev grafa naslednja:

- **Vozlišča (nodes)** predstavljajo glavne entitete: sobe (Room 101, Room 102) in vrata (ID1001, ID1002, ID1003).
- **Rebra (povezave)** prikazujejo odnose med temi entitetami, npr. pripadnost vrat določenemu prostoru.
- **Atributi** so preslikani na vozlišča in vsebujejo lastnosti entitete (višina, širina, požarna odpornost za vrata).

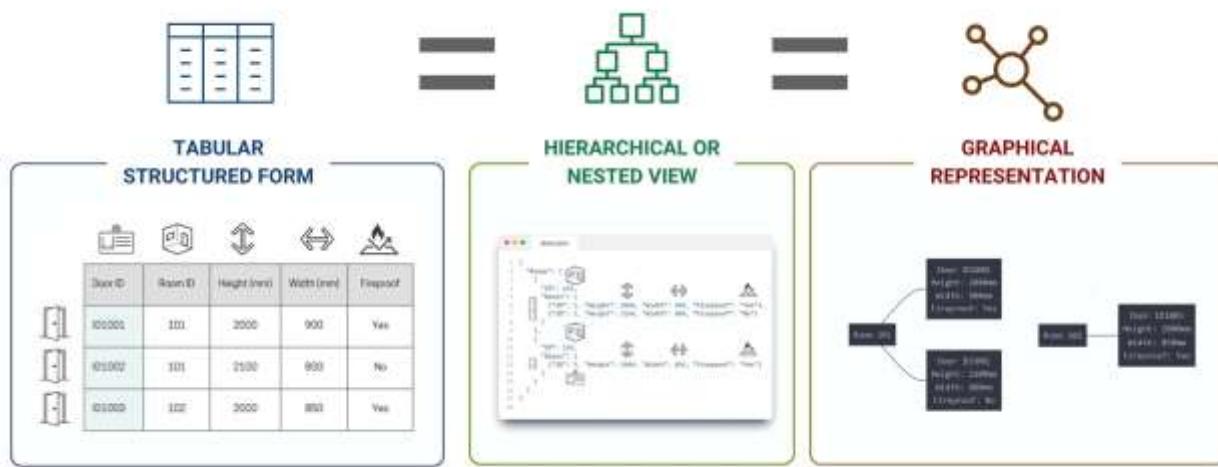


Slika 3.2-9 Informacije o entiteti projektnih vrat v grafičnem pogledu.

V grafnem podatkovnem modelu opisa vrat so vsaka soba in vsaka vrata ločena vozlišča. Vrata so s sobami povezana z robovi, ki označujejo, da vrata pripadajo določeni sobi. Atributi vrat (višina, širina, požarna odpornost) so shranjeni kot lastnosti ustreznih vozlišč. Več podrobnosti o grafnih formatih in o tem, kako se je grafna semantika pojavila v gradbeništvu, bo obravnavano v poglavju "Pojav semantike in ontologije v gradbeništvu".

Grafne podatkovne zbirke so učinkovite, kadar so pomembni odnosi med podatki in ne podatki sami, na primer v priporočilnih sistemih, sistemih usmerjanja ali pri modeliranju zapletenih odnosov v projektih upravljanja objektov. Oblika grafa poenostavlja ustvarjanje novih razmerij, saj omogoča dodajanje novih vrst podatkov v graf brez spremnjanja strukture shranjevanja. Vendar v primerjavi z relacijskimi tabelami in strukturiranimi formati v grafu ni dodatne povezljivosti podatkov - prenos dvodimenzionalnih podatkov iz zbirke podatkov v graf ne poveča števila povezav in ne zagotavlja novih informacij.

Oblika in shema podatkov morata biti prilagojeni posebnemu primeru uporabe in nalogam, ki jih je treba opraviti. Za učinkovito delo v poslovnih procesih je pomembno uporabljati tista orodja in tiste podatkovne modele, ki pripomorejo k čim hitrejšemu in enostavnnejšemu doseganju rezultatov.



Slika 3.2-10 Iste informacije o elementih projekta je mogoče shraniti v različnih oblikah z uporabo različnih podatkovnih modelov.

Večina velikih podjetij se danes sooča s problemom prevelike kompleksnosti podatkov. Vsaka od več sto ali več tisoč aplikacij uporablja svoj podatkovni model, kar povzroča pretirano zapletenost - posamezen model je pogosto desetkrat bolj zaplenjen, kot je potrebno, skupek vseh modelov pa je tisočkrat bolj zaplenjen. Ta pretirana zapletenost znatno otežuje delo razvijalcev in končnih uporabnikov.

Takšna zapletenost povzroča resne omejitve pri razvoju in vzdrževanju sistemov podjetja. Vsak nov element v modelu zahteva dodatno kodo, implementacijo nove logike, temeljito testiranje in

prilagajanje obstoječim rešitvam. Vse to povečuje stroške in upočasnuje delo ekipe za avtomatizacijo v podjetju, zaradi česar se tudi preproste naloge spremenijo v drage in dolgotrajne procese.

Kompleksnost vpliva na vse ravni podatkovne arhitekture. V relacijskih podatkovnih zbirkah se kaže v naraščajočem številu tabel in stolpcev, ki so pogosto nepotrebni. V objektno usmerjenih sistemih se zapletenost povečuje z množico razredov in medsebojno povezanih lastnosti. V formatih, kot sta XML ali JSON, se zapletenost kaže v zmedenih gnezdenih strukturah, edinstvenih ključih in nedoslednih shemah.

Zaradi prevelike zapletenosti podatkovnih modelov sistemi niso le manj učinkoviti, temveč jih končni uporabniki in v prihodnosti tudi veliki jezikovni modeli ter agenti LLM težko razumejo. Prav problem razumevanja in zapletenosti podatkovnih modelov in obdelave podatkov sproža vprašanje: kako narediti podatke dovolj enostavne za uporabo, da bodo dejansko hitro začeli biti uporabni.

Tudi če so podatkovni modeli izbrani pametno, se njihova uporabnost močno zmanjša, če je dostop do podatkov omejen. Lastniški formati in zaprte platforme ovirajo integracijo, otežujejo avtomatizacijo in jemljejo nadzor nad lastniškimi informacijami, kar ne ustvarja le silosa novih podatkov, temveč zaklenjen silos, do katerega je mogoče dostopati le z dovoljenjem prodajalca. Da bi razumeli razsežnost problema, je treba natančno preučiti, kako zaprti sistemi vplivajo na digitalne procese v gradbeništvu.

## Lastniški formati in njihov vpliv na digitalne procese

Eden ključnih izzivov, s katerimi se gradbena podjetja soočajo med digitalizacijo, je omejen dostop do podatkov. To otežuje povezovanje sistemov, zmanjšuje kakovost informacij in otežuje organizacijo učinkovitih procesov. Te težave so pogosto posledica uporabe lastniških formatov in zaprtih programskih rešitev.

Žal so številni programi, ki se uporabljajo v gradbeništvu, do zdaj uporabniku omogočali shranjevanje podatkov le v lastniških formatih ali v oblaku, do katerih je mogoče dostopati le prek strogo omejenih vmesnikov. Neredko pa so te rešitve zgrajene v odvisnosti od še bolj zaprtih sistemov večjih ponudnikov. Zato so tudi tisti razvijalci, ki bi radi ponudili bolj odprte arhitekture, prisiljeni upoštevati pravila, ki jih narekujejo veliki ponudniki.

Medtem ko sodobni sistemi za upravljanje gradbenih podatkov vse bolj podpirajo odprte formate in standarde (slika 3.1-5), podatkovne zbirke, ki temeljijo na CAD (BIM), ter povezani sistemi ERP in CAFM ostajajo izolirani lastniški "otoki" v digitalnem okolju industrije (slika 3.2-11).

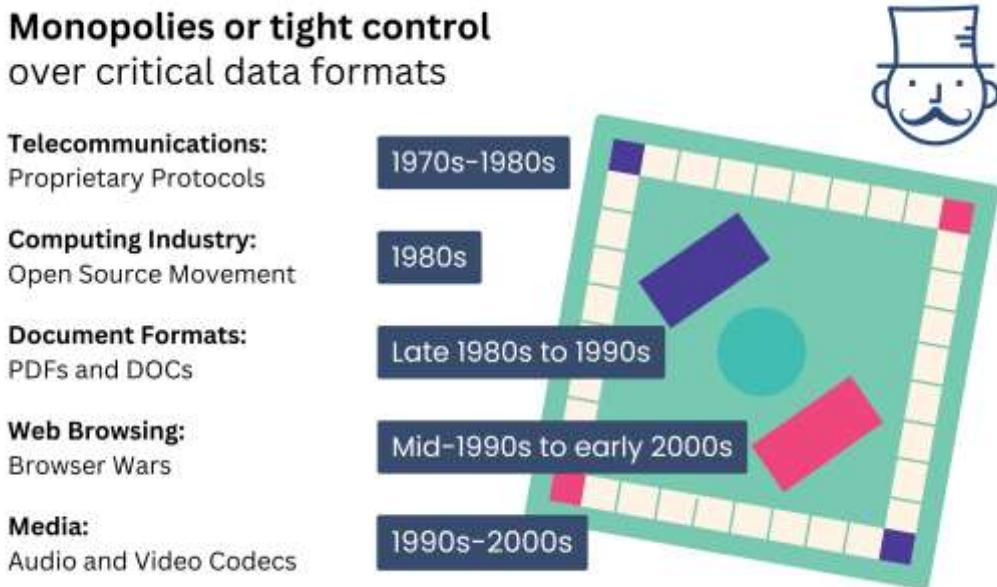


Slika 3.2-11 Zaprta in lastniška narava podatkov ustvarja ovire za povezovanje in dostop do podatkov.

Zaprti in monopolizirani formati in protokoli niso težava le v gradbeništvu. V številnih gospodarskih sektorjih se je boj proti zaprtim standardom in omejenemu dostopu do podatkov začel z upočasnitvijo inovacij (slika 3.2-12), obstojem umetnih ovir za vstop novih akterjev in poglabljanjem odvisnosti od velikih dobaviteljev. Zaradi hitre rasti pomena podatkov organi, pristojni za konkurenco, preprosto nimajo časa, da bi se odzvali na izzive, ki jih prinašajo novi digitalni trgi, zato zaprti formati in omejen dostop do podatkov v bistvu postajajo digitalne "meje", ki omejujejo pretok informacij in rast [63].

*Če stroji proizvedejo vse, kar potrebujemo, bo naš položaj odvisen od tega, kako se te dobrine porazdelijo. Vsi bodo lahko uživali v blaginji le, če bo bogastvo, ki ga proizvedejo stroji, razdeljeno. Ali pa bo večina ljudi na koncu živila v skrajni revščini, če bodo lastniki avtomobilov lahko uspešno lobirali proti prerazdelitvi bogastva. Zaenkrat se zdi, da se stvari odvijajo v drugo smer, saj tehnologija vodi v vse večjo neenakost [72].*

- Stephen Hawking, astrofizik, 2015



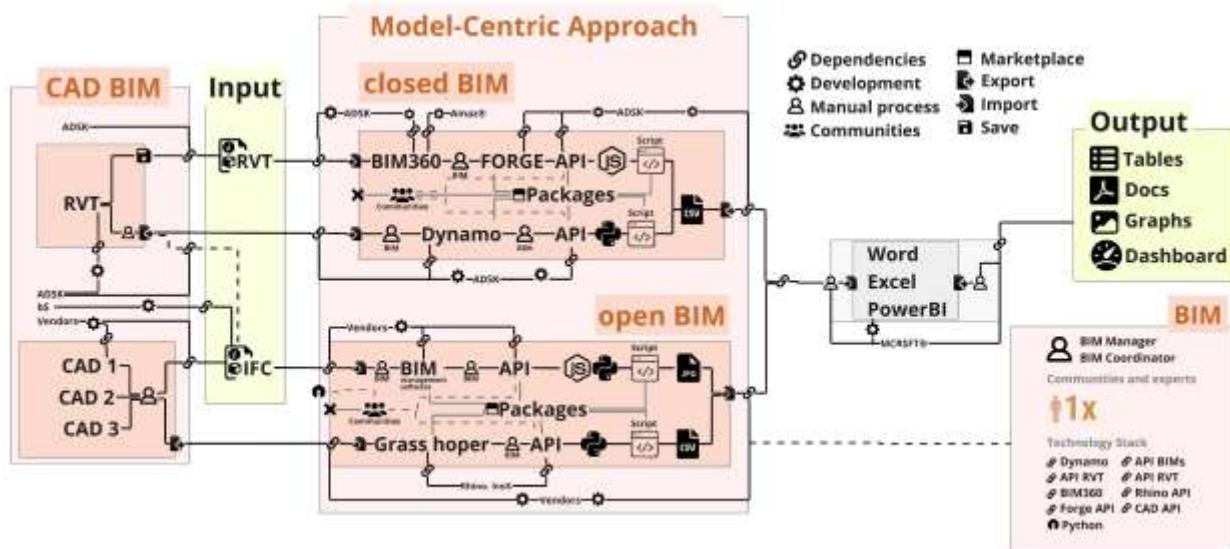
Slika 3.2-12 Monopolno lastništvo nad ključnimi podatkovnimi formati in protokoli ni izključni problem gradbene industrije.

Zaradi zaprtega dostopa do podatkovnih zbirk programi, upravljavci podatkov, podatkovni analitiki, strokovnjaki za IT in razvijalci, ki ustvarjajo aplikacije za dostop do podatkov, njihovo obdelavo in avtomatizacijo v gradbeništvu, se danes soočajo s številnimi odvisnostmi od ponudnikov programske opreme (slika 3.2-13). Te odvisnosti v obliki dodatnih ravni dostopa zahtevajo oblikovanje rešitev s specializiranimi API -povezavami ter posebnimi orodji in programsko opremo.

API (vmesnik za programiranje aplikacij) je formaliziran vmesnik, prek katerega lahko en program komunicira z drugim ter izmenjuje podatke in funkcije, ne da bi mu bilo treba dostopati do izvirne kode. API opisuje, kakšne zahteve lahko pošlje zunanji sistem, v kakšni obliki morajo biti in kakšne odgovore bo prejel. Je standardizirana "pogodba" med programskimi moduli.

Veliko število odvisnosti od zaprtih rešitev povzroči, da celotna arhitektura kode in logika poslovnih procesov v podjetju postane "špagetna arhitektura" orodij, ki so odvisna od politike prodajalca programske opreme glede zagotavljanja kakovostnega dostopa do podatkov.

Odvisnost od zaprtih rešitev in platform ne pomeni le izgube prilagodljivosti, temveč tudi resnična poslovna tveganja. Sprememba licenčnih pogojev, zaprtje dostopa do podatkov, spremembformatov ali strukture API - vse to lahko blokira kritične procese. Nenadoma se izkaže, da posodobitev ene tabele zahteva preoblikovanje celotnega bloka integracij in priključkov (slika 3.2-13), in vsaka obsežna posodobitev programske opreme ali njenega prodajalca API postane potencialna grožnja stabilnosti celotnega sistema podjetja.



Slika 3.2-13 Primer velikega števila odvisnosti pri obdelavi CAD -podatkov ustvarja ovire za povezovanje podatkov v ekosistemu gradbenih podjetij.

Razvijalci in sistemski arhitekti so v takšnih razmerah prisiljeni delati ne za pričakovanje, temveč za preživetje. Namesto da bi uvajali nove rešitve, se prilagajajo. Namesto da bi razvijali, poskušajo ohraniti združljivost. Namesto da bi avtomatizirali in pospešili procese, svoj čas porabijo za preučevanje naslednjih zaprtih vmesnikov, dokumentacije API in neskončno obnovo kode.

Delo z zaprtimi formati in sistemi ni le tehnični izviv, temveč tudi strateška omejitve. Kljub očitnim priložnostim, ki jih ponujajo sodobna avtomatizacija, umetna inteligenca, LLM in napovedna analitika, številna podjetja ne izkoristijo vseh njihovih možnosti. Ovire, ki jih postavljajo lastniški formati (slika 3.2-13), pa podjetjem onemogočajo dostop do lastnih podatkov. V tem je morda ironija digitalne preobrazbe v gradbeništву.

Preglednost podatkov in odprti sistemi niso razkošje, temveč predpogoj za hitrost in učinkovitost. Brez odprtosti so poslovni procesi polni nepotrebne birokracije, večplastnih verig odobritve in vse večje odvisnosti od načela HiPPO - odločanje na podlagi mnenja najbolje plačane osebe.

Kljub temu se na obzorju pojavlja sprememb paradigm. Kljub prevladi lastniških rešitev se vse več podjetij zaveda omejitev arhitektur, ki jih navdihuje četrta industrijska revolucija. Danes se vektor preusmerja k načelom pete revolucije, kjer so v središču podatki kot strateško sredstvo, odpri vmesniki (API) in resnična interoperabilnost med sistemi.

Ta prehod pomeni premik od zaprtih ekosistemov k prilagodljivim, modularnim digitalnim arhitekturam, kjer so ključni odprti formati, standardi in pregledna izmenjava podatkov.

## Odprti formati spreminjajo pristop k digitalizaciji

Gradbena industrija je bila ena zadnjih, ki se je lotila problema zaprtih in lastniških podatkov. Za razliko od drugih gospodarskih sektorjev se je digitalizacija tu razvijala počasi. Razlogi za to so med drugim tradicionalna konservativna narava panoge, razširjenost različnih lokalnih rešitev in globoko zakoreninjena narava upravljanja dokumentov v papirni obliki. Desetletja so se ključni gradbeni postopki opirali na fizične risbe, telefonske klice in nesinhronizirane zbirke podatkov. V tem kontekstu so bili zaprti formati dolgo časa dojeti kot norma in ne kot ovira.

Izkušnje iz drugih panog kažejo, da odprava ovir za dostop do zaprtih podatkov vodi v porast inovacij, hitrejši razvoj in večjo konkurenco [73]. V znanosti izmenjava odprtih podatkov omogoča hitrejša odkritja in razvoj mednarodnega sodelovanja. V medicini lahko izboljša učinkovitost diagnosticiranja in zdravljenja. V programskem inženirstvu - za ustvarjanje ekosistemov za soustvarjanje in hitro izboljševanje izdelkov.

V skladu z McKinseyjevim poročilom "Odprti podatki: 2013. [74] lahko odprti podatki letno sprostijo od 3 do 5 bilijonov dolarjev v sedmih ključnih panogah, vključno z gradbeništvom, prometom, zdravstvom in energetiko. Po isti študiji decentralizirani podatkovni ekosistemi omogočajo velikim gradbenim podjetjem in izvajalcem, da zmanjšajo stroške razvoja in vzdrževanja programske opreme, kar pospešuje uvajanje digitalnih tehnologij.

Prehod na odprte arhitekture, ki se je že zdavnaj začel v drugih gospodarskih sektorjih, se postopoma širi tudi v gradbeništvo. Velika podjetja in javni naročniki, zlasti pa finančne organizacije, ki nadzorujejo naložbe v gradbene projekte, vse bolj zahtevajo uporabo odprtih podatkov in dostop do izvirne kode izračunov, kalkulacij in aplikacij. Od razvijalcev se ne pričakuje več le ustvarjanje digitalnih rešitev in prikaz končnih številk projekta - od njih se pričakuje, da bodo pregledni, ponovljivi in neodvisni od tretjih ponudnikov aplikacij.

Z uporabo odprtokodnih rešitev je stranka prepričana, da tudi če zunanji razvijalci prenehajo sodelovati ali zapustijo projekt, to ne bo vplivalo na možnost nadaljnjega razvoja orodij in sistemov. Ena od glavnih prednosti odprtih podatkov je, da lahko odpravijo odvisnost razvijalcev aplikacij od posebnih platform za dostop do podatkov.

Če se podjetje ne more popolnoma odpovedati lastniškim rešitvam, je možen kompromis uporaba tehnik povratnega inženiringa. Te pravno in tehnično zanesljive metode omogočajo pretvorbo zaprtih formatov v bolj dostopne, strukturirane in primerne za integracijo. To je še posebej pomembno, kadar se je treba povezati s starejšimi sistemi ali prenesti informacije iz enega programskega okolja v drugega.

Eden najsvetlejših primerov v zgodovini prehoda na odprte formate in uporabe povratnega inženiringa (zakonito vdiranje v lastniške sisteme) v gradbeništvu je zgodovina boja za odprtje formata DWG, ki se pogosto uporablja v sistemih računalniško podprtga načrtovanja (CAD). Leta 1998 je kot odgovor na monopol enega proizvajalca programske opreme ostalih 15 proizvajalcev CAD oblikovalo novo zavezništvo, imenovano "Open DWG", da bi razvijalcem zagotovili brezplačna in neodvisna orodja za delo z DWG formatom (dejanski standard za prenos risb) brez potrebe po lastniški programske opremi ali zaprtih API-jih. Ta dogodek je bil prelomnica, ki je deset tisočim podjetjem omogočila brezplačen dostop do zaprtega formata priljubljene rešitve CAD od konca osemdesetih let prejšnjega stoletja do danes in ustvarjanje združljivih rešitev, kar je spodbudilo konkurenco na trgu CAD [75]. Danes se SDK "Open DWG", ki je bil prvič ustvarjen že leta 1996, uporablja v skoraj vseh rešitvah, v katerih je mogoče uvoziti, urejati in izvoziti format DWG zunaj uradne aplikacije razvijalca formata DWG.

Tudi drugi tehnološki velikani so prisiljeni v podobno preoblikovanje. Microsoft, nekoč simbol lastniškega pristopa, je odprl izvorno kodo ogrodja.NET Framework, začel uporabljati Linux v infrastrukturi storitev v oblaku Azure in kupil GitHub, da bi okrepil svoj položaj v odprtakodni skupnosti. [76]. Meta (nekdanji Facebook) je objavila odprtakodne modele umetne inteligence, kot je serija Llama, da bi spodbudila inovacije in sodelovanje pri razvoju agentov umetne inteligence. Izvršni direktor Mark Zuckerberg predvideva, da bodo odprtakodne platforme v naslednjem desetletju vodilne pri tehnološkem napredku [77].

Odprta koda je model razvoja in distribucije programske opreme, pri katerem je izvorna koda odprta za prosto uporabo, preučevanje, spremištanje in distribucijo.

Odprti podatki in odprtakodne rešitve ne postajajo le trend, temveč temelj digitalne trajnosti. Podjetjem omogočajo prilagodljivost, odpornost, nadzor nad lastnimi odločitvami in možnost razširjanja digitalnih procesov brez odvisnosti od politik prodajalcev. In kar je prav tako pomembno, podjetjem vračajo nadzor nad najdragocenejšim virom 21. stoletja - njihovimi podatki.

## **Sprememba paradigme: odprta koda kot konec obdobja prevlade proizvajalcev programske opreme**

Gradbeništvo doživlja spremembe, ki jih ni mogoče ovrednotiti na običajen način. Koncept podatkovno usmerjenega pristopa, osredotočenega na podatke, in uporaba odprtakodnih orodij vodita v premislek o pravilih igre, na katerih stojijo programski velikani na trgu.

Za razliko od prejšnjih tehnoloških preobrazb tega prehoda prodajalci ne bodo aktivno spodbujali. Sprememba paradigme ogroža njihove tradicionalne poslovne modele, ki temeljijo na licenciranju, naročninah in svetovanju. Nova realnost ne vključuje izdelka iz škatle ali plačljive naročnine - zahteva preusmeritev procesov in razmišljanja.

Za upravljanje in razvoj rešitev podatkovnih centrov, ki temeljijo na odprtih tehnologijah, bodo morala

podjetja ponovno razmisliti o notranjih procesih. Strokovnjaki iz različnih oddelkov bodo morali ne le sodelovati, temveč tudi na novo razmisliti o tem, kako sodelujejo.

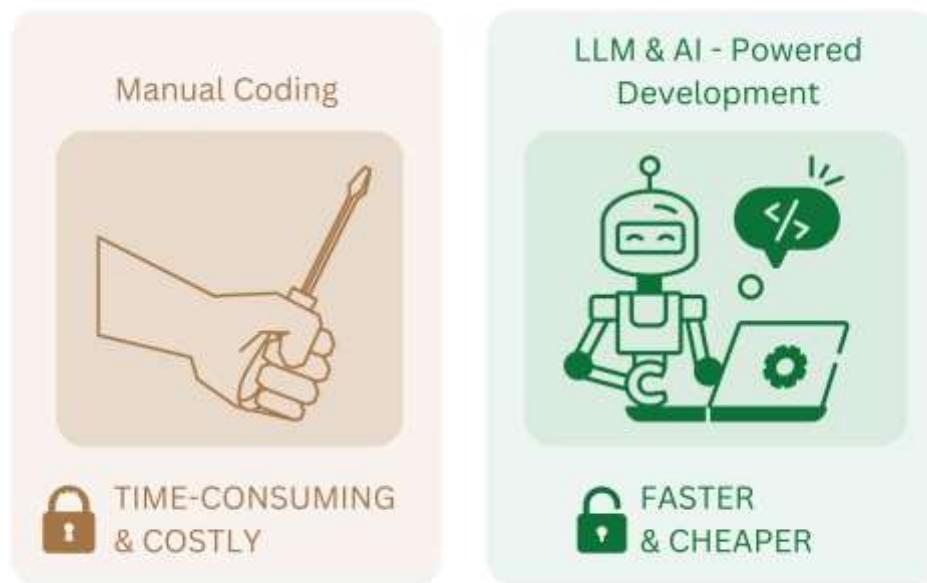
Nova paradigma predvideva uporabo odprtih podatkov in odprtakodnih rešitev, pri čemer bodo imela pri ustvarjanju programske kode posebno vlogo orodja, ki temeljijo na umetni inteligenci in velikih jezikovnih modelih (LLM), in ne programerji. že sredi leta 2024 bo pri Googlu več kot 25 odstotkov nove kode ustvarjene z umetno inteligenco [78]. V prihodnosti bo kodiranje z LLM opravilo 80% dela v samo 20% časa (slika 3.2-14).

Po podatkih študije McKinsey 2020 [79] grafični procesorji zaradi visoke zmogljivosti in podpore sodobnih odprtakodnih orodij v analitiki vse bolj nadomeščajo centralne procesorje. To podjetjem omogoča, da pospešijo obdelavo podatkov brez večjih naložb v drago programsko opremo ali zaposlovanja redkih strokovnjakov.

Vodilna svetovalna podjetja, kot so McKinsey, PwC in Deloitte, poudarjajo vse večji pomen odprtih standardov in odprtakodnih aplikacij v različnih panogah.

Po podatkih poročila PwC Open Source Monitor 2019 [80] 69% podjetij s 100 ali več zaposlenimi zavestno uporablja odprtakodne rešitve. OSS se še posebej aktivno uporablja v velikih podjetjih: OSS uporablja 71% podjetij z 200-499 zaposlenimi, 78% v kategoriji 500-1999 zaposlenih in do 86% med podjetji z več kot 2000 zaposlenimi. Po podatkih poročila Synopsys OSSRA 2023 je 96% analiziranih kodnih zbirk vsebovalo odprtakodne komponente [81].

Vloga razvijalca v prihodnosti ne bo ročno pisanje kode, temveč oblikovanje podatkovnih modelov, arhitekture tokov in upravljanje agentov umetne inteligence, ki na zahtevo ustvarjajo prave izračune. Uporabniški vmesniki bodo postali minimalistični, interakcija pa bo temeljila na dialogu. Klasično programiranje se bo umaknilo oblikovanju na visoki ravni in orkestraciji digitalnih rešitev (slika 3.2-14). Trenutni trendi - kot so platforme z nizko kodo (slika 7.4-6) in ekosistemi, podprtih z LLM (slika 7.4-4) - bodo znatno zmanjšali stroške razvoja in vzdrževanja IT-sistemov.



**Slika 3.2-14** Medtem ko danes aplikacije ročno ustvarjajo programerji, bodo v prihodnosti znaten del kode generirale rešitve na osnovi umetne inteligence in LLM.

Ta prehod bo drugačen od prejšnjih in veliki prodajalci programske opreme verjetno ne bodo katalizatorji.

V študiji Harvard Business School "The Value of Open Source Software" 2024 [40] je skupna vrednost odprtakodne programske opreme ocenjena z dveh vidikov. Po eni strani, če izračunamo, koliko bi potrebovali, da bi vse obstoječe odprtakodne rešitve zgradili od začetka, bi znesek znašal približno 4,15 milijarde dolarjev. Po drugi strani pa, če si predstavljamo, da vsako podjetje samostojno razvija svoje analogije odprtakodnih rešitev (kar se dogaja povsod), ne da bi imelo dostop do obstoječih orodij, bi skupni stroški poslovanja dosegli kolosalnih 8,8 bilijona dolarjev - to je strošek povpraševanja.

Ni težko uganiti, da nobenega velikega prodajalca programske opreme ne zanima krčenje trga programske opreme s potencialno vrednostjo 8,8 bilijona dolarjev na samo 4,15 milijarde dolarjev. To bi pomenilo zmanjšanje obsega povpraševanja za več kot 2000-krat. Takšno preoblikovanje je za prodajalce, katerih poslovni modeli temeljijo na dolgoletnem ohranjanju odvisnosti strank od zaprtih rešitev, preprosto nerentabilno. Podjetja, ki pričakujejo, da jim bo nekdo ponudil priročno in odprto rešitev na ključ, bodo morda razočarana - ti prodajalci se preprosto ne bodo pojavili.

Prehod na odprto digitalno arhitekturo ne pomeni izgube delovnih mest ali prihodkov. Nasprotno, ustvarja pogoje za prožne in prilagodljive poslovne modele, ki lahko sčasoma izpodrinejo tradicionalni trg licenčne in škatlaste programske opreme.

Namesto prodaje licenc - storitve, namesto zaprtih formatov - odprte platforme, namesto odvisnosti od prodajalca - neodvisnost in možnost oblikovanja rešitev za resnične potrebe. Tisti, ki so prej

orodja samo uporabljali, bodo lahko postali njihovi soavtorji. In tisti, ki znajo delati s podatki, modeli, scenariji in logiko, se bodo znašli v središču novega digitalnega gospodarstva v industriji. Več o teh spremembah in o tem, kakšne nove vloge, poslovni modeli in oblike sodelovanja se pojavljajo v zvezi z odprtimi podatki, bomo govorili v zadnjem, desetem delu knjige.

Rešitve, ki temeljijo na odprtih podatkih in odprti kodi, bodo podjetjem omogočile, da se osredotočijo na učinkovitost poslovnih procesov, ne pa na težave z zastarelimi vmesniki API in integracijo zaprtih sistemov. Zavesten prehod na odprto arhitekturo lahko bistveno izboljša produktivnost in zmanjša odvisnost od dobaviteljev.

Prehod v novo realnost ne pomeni le spremembe pristopov k razvoju programske opreme, temveč tudi premislek o samem načelu dela s podatki. V središču te preobrazbe ni koda, temveč informacije: njihova struktura, dostopnost in razlagalnost. In tu v ospredje stopajo odpri in strukturirani podatki, ki postajajo sestavni del nove digitalne arhitekture.

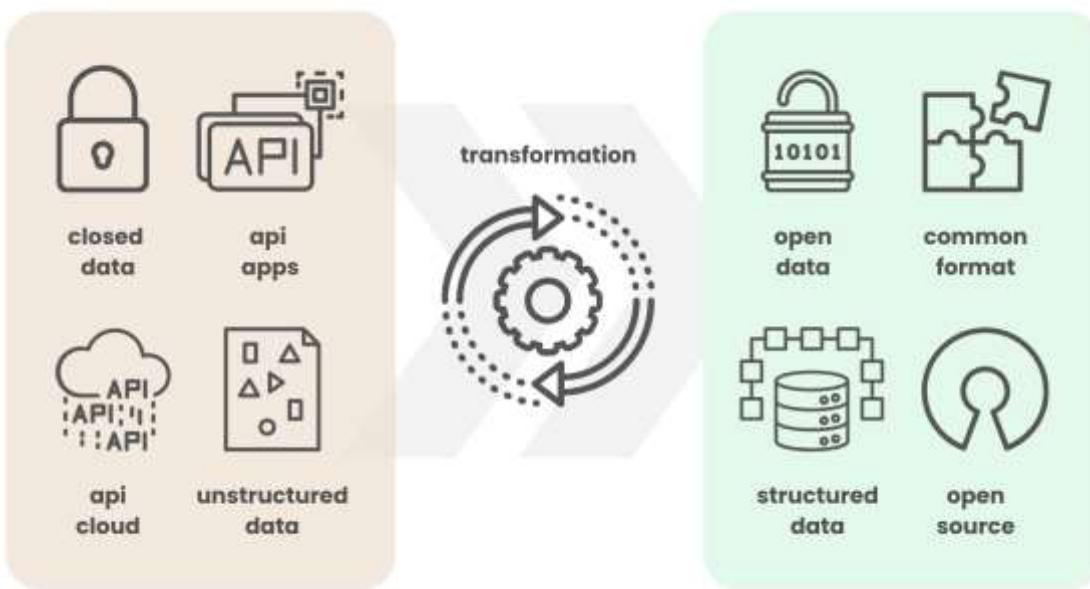
## Strukturirani odprti podatki: temelj digitalne preobrazbe

Medtem ko je bila v preteklih desetletjih trajnost poslovanja odvisna predvsem od izbire programskih rešitev in odvisnosti od določenih ponudnikov, je v današnjem digitalnem gospodarstvu ključni dejavnik kakovost podatkov in sposobnost učinkovitega dela z njimi. Odprta koda je pomemben del nove tehnološke paradigme, vendar je njen potencial resnično izkoriščen šele, ko so podatki razumljivi, organizirani in strojno berljivi. Med vsemi vrstami podatkovnih modelov postajajo strukturirani odprti podatki temeljni kamen trajnostne digitalne preobrazbe.

Glavna prednost strukturiranih odprtih podatkov je nedvoumna razlaga in možnost samodejne obdelave. To omogoča znatno povečanje učinkovitosti tako na ravni posameznih operacij kot v celotni organizaciji.

V poročilu družbe Deloitte "Proces prenosa podatkov pri preoblikovanju podjetij" je navedeno, da [82] je sodelovanje z IT pri upravljanju prenosa strukturiranih podatkov ključnega pomena. V skladu s poročilom vlade Združenega kraljestva Data Analytics and AI in Government Project Delivery (2024) [83] je odprava ovir pri izmenjavi podatkov med različnimi projektmi in organizacijami ključna za izboljšanje učinkovitosti pri upravljanju projektov. V dokumentu je poudarjeno, da lahko s standardizacijo podatkovnih formatov in uvedbo načel odprtih podatkov preprečimo podvajanje informacij, zmanjšamo izgubo časa in izboljšamo natančnost napovedi.

V gradbeništvu, kjer tradicionalno prevladujeta velika razdrobljenost in raznolikost formatov, imata strukturiran proces združevanja in strukturirani odprti podatki ključno vlogo pri oblikovanju skladnih in obvladljivih procesov (slika 4.1-14). Udeležencem projekta omogočata, da se osredotočijo na izboljšanje produktivnosti in ne na reševanje tehničnih težav, povezanih z nezdružljivostjo med zaprtimi platformami, podatkovnimi modeli in formati.



Slika 3.2-15 Odprti strukturirani podatki zmanjšujejo odvisnost od programskega rešitev in platform ter pospešujejo inovacije.

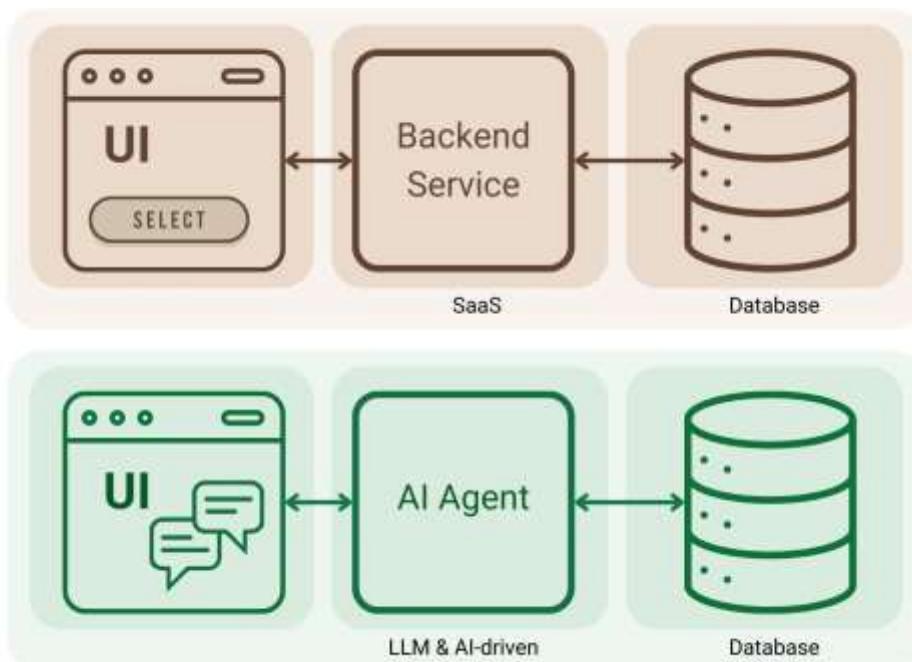
Sodobna tehnološka orodja, ki jih bomo podrobno obravnavali v nadaljevanju knjige, omogočajo ne le zbiranje informacij, temveč tudi njihovo samodejno čiščenje: odpravljanje podvajanj, popravljanje napak in normalizacijo vrednosti. To pomeni, da analitiki in inženirji ne delajo z raznovrstnimi dokumenti, temveč z organizirano bazo znanja, primerno za analizo, avtomatizacijo in odločanje.

*Naj bo čim bolj preprosta, vendar ne enostavnejša.*

- Albert Einstein, teoretični fizik (avtorstvo citata je sporno [84])

Danes je mogoče večino uporabniških vmesnikov za delo s podatki ustvariti samodejno - brez potrebe po ročnem pisanju kode za vsak poslovni primer. Za to je potrebna infrastruktura plast, ki brez dodatnih navodil razume strukturo, model in logiko podatkov (slika 4.1-15). Ta pristop omogočajo prav strukturirani podatki: obrazce, tabele, filtre in poglede je mogoče samodejno ustvariti z minimalnim programerskim naporom.

Najpomembnejše vmesnike, ki so za uporabnika ključni, bo morda še vedno treba ročno izpopolniti. V večini primerov - in to je od 50 do 90 odstotkov delovnih scenarijev - pa zadostuje samodejno generiranje aplikacij in izračunov brez uporabe posebnih aplikacij v ta namen (slika 3.2-16), kar bistveno zmanjša stroške razvoja in vzdrževanja, zmanjša število napak in pospeši izvajanje digitalnih rešitev.



Slika 3.2-16 Arhitekturna modela za delo s podatki: tradicionalna arhitektura aplikacije in model, usmerjen v umetno inteligenco z LLM.

Naslednji korak v digitalni evoluciji je prehod od arhitektur, ki temeljijo na posameznih aplikacijah, k intelligentno upravljanim sistemom, ki temeljijo na jezikovnih modelih (LLM). V takšni arhitekturi strukturirani podatki ne postanejo le predmet shranjevanja, temveč tudi podlaga za interakcijo z orodji umetne inteligence, ki so sposobna analizirati, razlagati in priporočati ukrepe na podlagi konteksta.

V naslednjih poglavjih si bomo ogledali resnične primere izvajanja arhitekture, ki temelji na odprtih strukturiranih podatkih, in pokazali, kako se jezikovni modeli uporabljajo za samodejno interpretacijo, potrjevanje in obdelavo podatkov. Ti praktični primeri vam bodo pomagali bolje razumeti, kako nova digitalna logika deluje v praksi - in kakšne koristi prinaša podjetjem, ki so pripravljena na preoblikovanje.



## POGLAVJE 3.3.

### **LLM IN NJIHOVA VLOGA PRI OBDELAVI PODATKOV IN POSLOVNIH PROCESIH**

**klepetalnice LLM: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok za avtomatizacijo podatkovnih procesov.**

Pojav velikih jezikovnih modelov (LLM) je bil naravna posledica gibanja za strukturirane odprte podatke in filozofije odprte kode. Ko podatki postanejo organizirani, dostopni in strojno berljivi, je naslednji korak orodje, ki lahko sodeluje s temi podatki, ne da bi bilo treba pisati zapleteno kodo ali imeti specializirano tehnično znanje.

Programi vseživljenskega učenja so neposreden rezultat odprtosti: velikih odprtih zbirk podatkov, publikacij in gibanja za odprto kodo. Brez odprtih znanstvenih člankov, javno dostopnih besedilnih podatkov in kulture sodelovalnega razvoja ne bi bilo programa ChatGPT ali drugih programov LLM. Program LLM je na neki način "destilat" nakopičenega digitalnega znanja človeštva, zbranega in izobraženega po načelih odprtosti.

Sodobni veliki jezikovni modeli (LLM - Large Language Models), kot so ChatGPT ® (OpenAI), LlaMa ™ (Meta AI), Mistral DeepSeek™, Grok ™ (xAI), Claude ™ (Anthropic), QWEN™, uporabnikom omogočajo oblikovanje poizvedb po podatkih v naravnem jeziku. S tem delo s podatki postane dostopno ne le razvijalcem, temveč tudi analitikom, inženirjem, načrtovalcem, menedžerjem in drugim strokovnjakom, ki jim je bilo programiranje prej oddaljeno

**LLM** (Large Language Model) je umetna inteligenca, ki je usposobljena za razumevanje in ustvarjanje besedila na podlagi velikih količin podatkov, zbranih na internetu. Sposoben je analizirati kontekst, odgovarjati na vprašanja, sodelovati v dialogu, pisati besedila in ustvarjati programsko kodo.

Če je bilo prej za vizualizacijo, obdelavo ali analizo podatkov potrebno znanje posebnega programskega jezika: Python, SQL, R ali Scala ter sposobnost dela s knjižnicami, kot so Pandas, Polars ali DuckDB in številne druge, se bodo od leta 2023 razmere korenito spremenile. Zdaj lahko uporabnik preprosto opiše, kaj želi dobiti - in model bo sam ustvaril kodo, jo izvedel, prikazal tabelo ali graf ter pojasnil rezultat. Prvič po desetletjih razvoj tehnologije ni šel po poti zapletanja, temveč po poti radikalne poenostavitve in dostopnosti.

To načelo je "obdelovanje podatkov z besedami (pozivi)". - je pomenilo novo stopnjo v razvoju dela z informacijami, saj je ustvarjanje rešitev preneslo na še višjo raven abstrakcije. Tako kot nekoč ni bilo več potrebno, da bi uporabniki razumeli tehnične temelje interneta, da bi lahko upravliali spletnе trgovine ali ustvarjali spletna mesta z uporabo WordPress, Joomla in drugih odprtokodnih modularnih

sistemov (avtor knjige že od leta 2005 dela s takšnimi sistemmi, vključno z izobraževalnimi in inženirskimi spletnimi platformami). - To pa je povzročilo razcvet digitalnih vsebin in spletnega poslovanja - danes lahko inženirji, analitiki in menedžerji avtomatizirajo delovne procese brez znanja programskih jezikov. K temu pripomorejo zmogljivi programi LLM - brezplačni in odprtakodni, kot so LLaMA, Mistral, Qwen, DeepSeek in drugi - ki omogočajo dostop do naprednih tehnologij najširšemu občinstvu.

## **Veliki jezikovni modeli LLM: kako deluje**

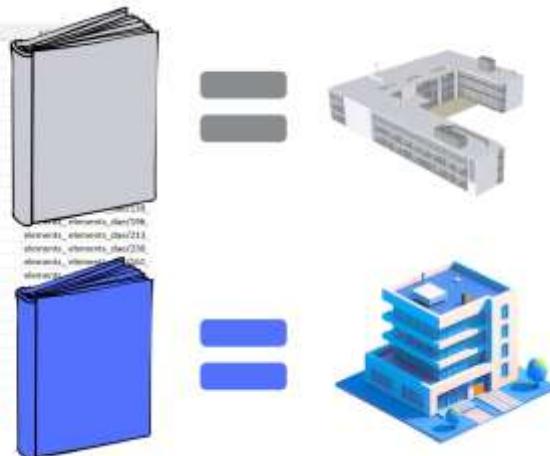
Veliki jezikovni modeli (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) so nevronske mreže, usposobljene na velikih količinah besedilnih podatkov iz interneta, knjig, člankov in drugih virov. Njihova glavna naloga je razumevanje konteksta človeškega govora in ustvarjanje smiselnih odzivov.

Sodobni LLM temelji na arhitekturi Transformer, ki so jo leta 2017 predlagali Googlovi raziskovalci [85]. Ključni sestavni del te arhitekture je mehanizem pozornosti, ki modelu omogoča, da upošteva odnose med besedami ne glede na njihov položaj v besedilu.

Učni proces LLM je zelo podoben načinu, na katerega se ljudje učimo jezika, le da je milijonkrat večji. Model analizira milijarde primerov besed in izrazov ter ugotavlja vzorce v strukturi jezika in logiki pomenskih prehodov. Celotno besedilo se razdeli na žetone - minimalne pomenske enote (besede ali njihove dele), ki se nato pretvorijo v vektorje v večdimenzionalnem prostoru (slika 8.2-2). Te vektorske predstavitve omogočajo stroju, da "razume" skrite odnose med pojmi, namesto da bi z besedilom preprosto deloval kot z zaporedjem simbolov.

Veliki jezikovni modeli niso le orodja za generiranje besedila. Sposobni so prepoznati pomen, najti povezave med pojmi in delati s podatki, tudi če so ti predstavljeni v različnih oblikah. Najpomembnejše je, da so informacije razčlenjene na razumljive modele in predstavljene v obliki žetonov, s katerimi lahko deluje program LLM.

Enak pristop lahko uporabimo tudi pri gradbenih projektih. Če si projekt predstavljamo kot nekakšno besedilo, v katerem je vsaka stavba, element ali konstrukcija simbol, lahko začnemo takšne informacije obdelovati na podoben način. Gradbene projekte lahko primerjamo s knjigami, ki so organizirane v kategorije, poglavja in skupine odstavkov, sestavljene iz minimalnih žetonov - elementov gradbenega projekta (slika 3.3-1). S prevajanjem podatkovnih modelov v strukturirano obliko lahko strukturirane podatke prevedemo tudi v vektorske baze (slika 8.2-2), ki so idealen vir za strojno učenje in tehnologije, kot je LLM.



**Slika 3.3-1 Element gradbenega projekta je kot žeton v besedilu: minimalna enota, iz katere se oblikujejo skupine (odstavki) delov (kategorije) celotnega projekta.**

Če je gradbeni projekt digitaliziran in so njegovi elementi predstavljeni kot žetoni ali vektorji, je mogoče do njih dostopati v naravnem jeziku in ne s togimi formalnimi poizvedbami. Tu pride do izraza ena od ključnih prednosti LLM - zmožnost razumevanja pomena poizvedbe in njene povezave z ustreznimi podatki.

Inženirju ni več treba pisati poizvedb SQL ali kode Python, da bi pridobil zahtevane podatke - preprosto lahko ob razumevanju LLM in podatkovne strukture oblikuje nalogu na običajen način: "*Poisci vse armiranobetonske konstrukcije z razredom betona, višjim od B30, in izračunaj njihovo skupno prostornino*". Model bo prepoznal pomen poizvedbe, jo pretvoril v strojno berljivo obliko, poiskal podatke (združil in preoblikoval) ter vrnil končni rezultat.

Dokumenti, tabele, modeli projektov se pretvorijo v vektorske predstavitev (vstavljanje) in shranijo v podatkovno zbirko. Ko uporabnik postavi vprašanje, se poizvedba prav tako pretvori v vektor in sistem poišče najustreznejše podatke. To omogoča, da se sistem LLM ne zanaša le na lastno usposobljeno znanje, temveč tudi na dejanske podatke podjetij, tudi če so se pojavili že po koncu usposabljanja modela.

Ena najpomembnejših prednosti programa LLM v gradbeništvu je možnost ustvarjanja programske kode. Namesto da bi tehnično nalogo predali programerju, lahko strokovnjaki nalogu opišejo v naravnem jeziku, model pa bo ustvaril potrebno kodo, ki jo je mogoče uporabiti (s kopiranjem iz klepeta) pri ustvarjanju kode za avtomatizacijo procesa. LLM - modeli omogočajo strokovnjakom brez poglobljenega programerskega znanja, da prispevajo k avtomatizaciji in izboljšanju poslovnih procesov podjetja.



**Slika 3.3-2 LLM uporabnikom omogoča pisanje kode in pridobivanje rezultatov, ne da bi za to potrebovali znanje programiranja.**

Po podatkih študije, ki jo je leta 2024 izvedla družba Wakefield Research, sponzoriral pa SAP [36] in v kateri je sodelovalo 300 vodilnih delavcev v podjetjih z letnimi prihodki najmanj 1 milijardo USD v ZDA: 52% vodilnih delavcev zaupa umetni inteligenici pri analizi podatkov in pripravi priporočil za odločanje. Drugih 48% jih AI uporablja za prepoznavanje prej neupoštevanih tveganj, 47% pa jih AI uporablja za predlaganje alternativnih načrtov. Poleg tega jih 40 odstotkov uporablja UI za razvoj novih izdelkov, načrtovanje proračuna in tržne raziskave. Študija je pokazala tudi pozitiven vpliv UI na osebno življenje, saj 39% vprašanih poroča o boljšem ravnovesju med delom in zasebnim življnjem, 38% o boljšem duševnem zdravju, 31% pa o nižji ravni stresa.

Vendar pa je vseeno pomembno, da se pri uporabi LLM zavestno zavedamo njihove moči. Kot vsaka tehnologija imajo tudi te omejitve. Ena od najbolj znanih težav so tako imenovane "halucinacije" - primeri, ko model zanesljivo poda verjeten, vendar dejansko napačen odgovor. Zato je ključnega pomena razumeti, kako model deluje: katere podatke in podatkovne modele lahko interpretira brez napak, kako interpretira poizvedbe in od kod pridobiva informacije. Prav tako se je treba zavedati, da je znanje modela LLM omejeno na datum njegovega usposabljanja, brez povezave z zunanjimi podatki pa model morda ne bo upošteval trenutnih normativov, standardov, cen ali tehnologij.

Rešitev za te težave je redno posodabljanje podatkovnih zbirk vektorjev, povezovanje z ustreznimi viri in razvoj avtonomnih agentov z umetno inteligenco, ki ne odgovarjajo le na vprašanja, temveč proaktivno uporabljajo podatke za usposabljanje, upravljajo naloge, prepoznavajo tveganja, predlagajo možnosti optimizacije in spremljajo uspešnost projekta.

Prehod na LLM -vmesnike v gradbeništvu ni le tehnološka novost. Gre za spremembo paradigme, ki odpravlja ovire med ljudmi in podatki. Gre za možnost, da z informacijami delamo tako preprosto, kot se pogovarjamo med seboj - in še vedno dobimo natančne, preverjene in uporabne rezultate.

Podjetja, ki bodo začela uporabljati takšna orodja prej kot druga, bodo pridobila pomembno konkurenčno prednost. To vključuje pospešitev dela, zmanjšanje stroškov in izboljšanje kakovosti

projektnih rešitev hitrim dostopom do analize podatkov in zmožnostjo hitrega iskanja odgovorov na zapletena vprašanja. Vendar je treba upoštevati tudi varnostna vprašanja. Uporaba storitev LLM v oblaku je lahko povezana s tveganji uhajanja podatkov. Zato organizacije vse pogosteje iščejo alternativne rešitve, ki jim omogočajo uporabo orodij LLM v lastni infrastrukturi - lokalno, s popolno zaščito in nadzorom nad informacijami.

## Uporaba lokalnih LLM za občutljive podatke podjetja

Pojav prvih klepetalnic-LLM leta 2022 je pomenil novo stopnjo v razvoju umetne inteligence. Vendar pa se je takoj po široki uveljavitvi teh modelov pojavilo upravičeno vprašanje: kako varen je prenos podatkov in poizvedb, povezanih s podjetjem, v oblak? Večina jezikovnih modelov v oblaku je na svojih strežnikih hranila zgodovino komunikacije in naložene dokumente, kar je za podjetja, ki se ukvarjajo z občutljivimi informacijami, predstavljalo resno oviro za sprejetje umetne inteligence.

Ena od najbolj trajnostnih in logičnih rešitev tega problema je uvedba odprtakodnega LLM lokalno, znotraj infrastrukture IT podjetja. V nasprotju s storitvami v oblaku lokalni modeli delujejo brez internetne povezave, ne prenašajo podatkov na zunanje strežnike in podjetjem omogočajo popoln nadzor nad informacijami

*Najboljši odprti model [Open Source LLM] je trenutno po učinkovitosti primerljiv z zaprtimi modeli [kot sta ChatGPT, Claude], vendar s približno enoletnim zamikom [77].*

- Ben Cottier, vodilni raziskovalec pri neprofitni raziskovalni organizaciji Epoch AI, 2024

Večja tehnološka podjetja so začela dajati svoje programe LLM na voljo za lokalno uporabo. Primera prehoda na odprto arhitekturo sta bila odprtakodna serija LLaMA podjetja Meta in hitro rastoči kitajski projekt DeepSeek. Poleg njiju sta tudi podjetji Mistral in Falcon izdala zmogljive modele, ki so osvobojeni omejitev lastniških platform. Te pobude niso le pospešile razvoja globalnega sistema umetne inteligence, temveč so podjetjem, ki se zavedajo zasebnosti, zagotovile tudi prave alternative za neodvisnost, prilagodljivost in skladnost z varnostjo.

V poslovnom okolju, zlasti v gradbeništvu, varovanje podatkov ni le vprašanje udobja, temveč tudi skladnosti s predpisi. Delo z razpisno dokumentacijo, predračuni, risbami in zaupno korespondenco zahteva strog nadzor. In tu lokalni LLM zagotavlja potrebno zagotovilo, da podatki ostanejo znotraj meja podjetja.

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

**Slika 3.3-3 Lokalni modeli zagotavljajo popoln nadzor in varnost, rešitve v oblaku pa enostavno integracijo in samodejne posodobitve.**

Glavne prednosti lokalne odprte kode LLM:

- Popoln nadzor nad podatki. Vse informacije ostanejo v podjetju, kar preprečuje nepooblaščen dostop in uhajanje podatkov.
- Delovanje brez povezave. Ni odvisen od internetne povezave, kar je še posebej pomembno za izolirane IT infrastrukture. To zagotavlja neprekinjeno delovanje tudi v primeru sankcij ali blokiranih storitev v oblaku.
- Prilagodljivost uporabe. Model se lahko uporablja za generiranje besedila, analizo podatkov, pisanje programske kode, podporo načrtovanju in upravljanje poslovnih procesov.
- Prilagoditev ciljem podjetja. LLM se lahko usposablja na podlagi internih dokumentov, kar vam omogoča, da upoštevate posebnosti dela podjetja in značilnosti njegove panoge. Lokalni

LLM je mogoče povezati s platformami CRM, ERP ali BI, kar vam omogoča avtomatizacijo analize poizvedb strank, izdelavo poročil ali celo napovedovanje trendov.

Namestitev brezplačnega in odprtakodnega modela DeepSeek -R1-7B na strežnik, do katerega lahko dostopa celotna ekipa uporabnikov, po ceni 1000 USD na mesec je lahko potencialno cenejša od letnih pristojbin za API v oblaku, kot sta ChatGPT ali Claude, in podjetjem omogoča, da prevzamejo popoln nadzor nad svojimi podatki, odpravijo njihov prenos v internet ter pomagajo izpolnjevati regulativne zahteve, kot je GDPR

V drugih panogah lokalni študenti LLM že spreminja svoj pristop k avtomatizaciji. V podpornih storitvah se odzivajo na pogosta vprašanja strank in tako zmanjšujejo delovno obremenitev operaterjev. V kadrovskih oddelkih analizirajo življenjepise in izbirajo ustrezne kandidate. V e-trgovini ustvarjajo prilagojene ponudbe, ne da bi razkrili podatke o uporabniku.

Podoben učinek se pričakuje tudi v gradbenem sektorju. Z integracijo LLM s projektnimi podatki in standardi je mogoče pospešiti pripravo dokumentacije, avtomatizirati pripravo ocen in napovednih stroškovnih analiz. Uporaba LLM v povezavi s strukturiranimi tabelami in podatkovnimi okvirji postaja še posebej obetavno področje.

## **Popoln nadzor nad umetno inteligenco v podjetju in kako namestiti lasten LLM**

Sodobna orodja podjetjem omogočajo lokalno namestitev velikega jezikovnega modela (LLM) v samo nekaj urah. To omogoča popoln nadzor nad podatki in infrastrukturo, odpravlja odvisnost od zunanjih storitev v oblaku in zmanjšuje tveganje uhajanja informacij. Ta rešitev je še posebej pomembna za organizacije, ki delajo z občutljivo projektno dokumentacijo ali komercialno občutljivimi podatki.

Glede na naloge in vire so na voljo različni scenariji uvajanja, od rešitev, ki so na voljo že v osnovi, do bolj prilagodljivih in razširljivih arhitektur. Eno najlažjih orodij je Ollama, ki omogoča zagon jezikovnih modelov dobesedno z enim klikom, ne da bi za to potrebovali poglobljeno tehnično znanje. Hiter začetek z orodjem Ollama:

1. Prenesite distribucijo za svoj operacijski sistem (Windows / Linux / macOS) z uradne spletnne strani: [ollama.com](https://ollama.com)
2. Model namestite z ukazno vrstico. Na primer za model *Mistral*:

```
ollama run mistral
```

3. Ko model zaženete, je pripravljen za delo - prek terminala lahko pošiljate besedilne poizvedbe ali ga vključite v druga orodja. Zaženite model in izvedite poizvedbo:

```
ollama run mistral "Kako izdelati izračun z vsemi viri za delo za namestitev 100 mm široke predelne stene iz mavčnih plošč?".
```

Za tiste, ki raje delajo v znanem vizualnem okolju, je na voljo LM Studio, brezplačna aplikacija z vmesnikom, ki spominja na ChatGPT

- LM Studio namestite tako, da prenesete distribucijski komplet z uradne spletne strani - lmstudio.ai
- V vgrajenem katalogu izberite model (npr. Falcon ali GPT-Neo-X) in ga prenesite.
- Delo z modelom prek intuitivnega vmesnika, ki spominja na ChatGPT, vendar je popolnoma lokaliziran.

	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

Slika 3.3-4 Primerjava priljubljenih lokalnih odprtakodnih modelov LLM.

Izbira modela je odvisna od zahtev glede hitrosti, natančnosti in razpoložljivih zmogljivosti strojne opreme (slika 3.3-4). Majhni modeli, kot sta Mistral 7B in Baichuan 7B, so primerni za lahke naloge in mobilne naprave, medtem ko zmogljivi modeli, kot je DeepSeek -V3, zahtevajo precej računalniških virov, vendar ponujajo visoko zmogljivost in podporo za več jezikov. V prihodnjih letih bo trg LLM hitro rasel - videli bomo vse več lahkih in specializiranih modelov. Namesto splošnih LLM, ki pokrivajo vse človeške vsebine, se bodo pojavili modeli, usposobljeni na podlagi ozkega domenskega znanja. Pričakujemo lahko na primer pojav modelov, zasnovanih izključno za obdelavo inženirskega izračunov, gradbenih ocen ali podatkov v formatih CAD. Takšni specializirani modeli bodo hitrejši, natančnejši in varnejši za uporabo - zlasti v strokovnih okoljih, kjer sta pomembna visoka zanesljivost in vsebinska poglobljenost.

Ko je lokalni LLM vzpostavljen, ga je mogoče prilagoditi posebnim nalogam podjetja. V ta namen se uporablja tehnika finega prilagajanja, pri čemer se model dodatno usposobi na internih dokumentih, tehničnih navodilih, pogodbenih predlogah ali projektni dokumentaciji.

## RAG: Inteligentni LLM - pomočniki z dostopom do podatkov podjetja

Naslednja stopnja v razvoju uporabe LLM v podjetjih je povezovanje modelov z dejanskimi podatki podjetij v realnem času. Ta pristop se imenuje RAG (Retrieval-Augmented Generation) - pridobivanje in razširjeno generiranje. V tej arhitekturi jezikovni model ne postane le dialogni vmesnik, temveč polnopravni inteligentni pomočnik, ki je sposoben krmariti po dokumentih, risbah, zbirkah podatkov in zagotavljati natančne, kontekstualne odgovore.

Glavna prednost sistema RAG je možnost uporabe notranjih podatkov podjetja, ne da bi bilo treba model predhodno usposobiti, hkrati pa ohranja visoko natančnost in prilagodljivost pri ravnjanju z informacijami.

Tehnologija RAG združuje dve glavni komponenti:

- Pridobivanje: model se poveže s podatkovnimi skladišči - dokumenti, tabelami, datotekami PDF - datotekami, risbami - in pridobi ustrezne informacije, kot jih zahteva uporabnik.
- Razširjeno generiranje: model na podlagi pridobljenih podatkov ustvari natančen in utemeljen odgovor, pri čemer upošteva kontekst in specifičnost poizvedbe.

Za zagon LLM s podporo RAG, je treba opraviti nekaj korakov:

- **Priprava podatkov:** zbiranje potrebnih dokumentov, risb, specifikacij, tabel. Lahko so v različnih formatih in strukturah, od PDF do Excel.
- **Indeksiranje in vektorizacija:** z orodji, kot sta LlamaIndex ali LangChain, se podatki pretvorijo v vektorske predstavitve, ki omogočajo iskanje semantičnih povezav med fragmenti besedila (več o vektorskih podatkovnih zbirkah in prevajanju velikih polj v vektorske predstavitve, vključno s projekti CAD, v delu 8).
- **Poizvedovanje po pomočniku:** ko so podatki naloženi, lahko modelu zastavite vprašanja in model bo poiskal odgovore v okviru podjetja in ne v splošnem znanju, zbranem na internetu.

Predpostavimo, da ima podjetje mapo constructionsite\_docs, v kateri so shranjene pogodbe, navodila, ocene in preglednice. Z uporabo skripte Python (slika 3.3-5) lahko preiščemo to mapo in zgradimo vektorsko indeksiranje: vsak dokument bo pretvorjen v niz vektorjev, ki odražajo semantično vsebino besedila. S tem se dokumenti spremenijo v nekakšen "zemljevid pomenov", po katerem se lahko model učinkovito pomika in išče povezave med izrazi in besednimi zvezami.

Model na primer "pomni", da se besedi "vračilo" in "reklamacija" pogosto nahajata v delu pogodbe, ki se nanaša na pošiljanje materiala na gradbišče. Če se postavi vprašanje, na primer: "Kolikšen je rok za vračilo?" (slika 3.3-5 - 11. vrstica kode) - bo program LLM analiziral interne dokumente in poiskal

točno določene informacije ter deloval kot inteligentni pomočnik, ki je sposoben prebrati in razumeti vsebino vseh datotek podjetja.



```

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

```

**Slika 3.3-5 LM prebere mapo z datotekami - podobno kot jo človek odpre in poišče želeni dokument**

Kodo lahko zaženete na katerem koli računalniku z nameščenim programom Python. Več o uporabi programa Python in IDE za izvajanje kode bomo povedali v naslednjem poglavju.

Lokalna namestitev LLM ni le trend, temveč strateška rešitev za podjetja, ki cenijo varnost in prilagodljivost. Vendar je uvedba LLM, bodisi na lokalnih računalnikih podjetja bodisi z uporabo spletnih rešitev, le prvi korak. Za uporabo zmožnosti LLM pri dejanskih nalogah morajo podjetja uporabljati orodja, ki jim omogočajo ne le prejemanje odgovorov klepeta, temveč tudi shranjevanje ustvarjene logike v obliki kode, ki jo je mogoče zagnati zunaj konteksta uporabe LLM. To je pomembno za razširjanje rešitev - ustrezno organizirani procesi omogočajo, da se razvoj umetne inteligence uporablja za več projektov hkrati ali celo v celotnem podjetju.

Pri tem ima pomembno vlogo izbira ustreznega razvojnega okolja (IDE). Sodobna programska orodja omogočajo ne le razvoj rešitev, ki temeljijo na LLM, temveč tudi njihovo integracijo v obstoječe poslovne procese, tako da jih spremenijo v avtomatizirane ETL -Pipeline



## POGLAVJE 3.4.

### IDE S PODPORO LLM IN PRIHODNJIMI PROGRAMSKIMI SPREMEMBAMI

#### Izbira IDE: od eksperimentov LLM do poslovnih rešitev

Ko se podajate v svet avtomatizacije, analize podatkov in umetne inteligence - zlasti pri delu z velikimi jezikovnimi modeli (LLM) - je ključnega pomena izbrati pravo integrirano razvojno okolje (IDE). To IDE bo vaše glavno delovno orodje: kraj, kjer se bo izvajala koda, ustvarjena z LLM, tako na lokalnem računalniku kot v omrežju podjetja. Od izbire IDE ni odvisno le udobje vašega dela, temveč tudi to, kako hitro boste lahko prešli od eksperimentalnih poizvedb LLM do polnopravnih rešitev, vgrajenih v resnične poslovne procese.

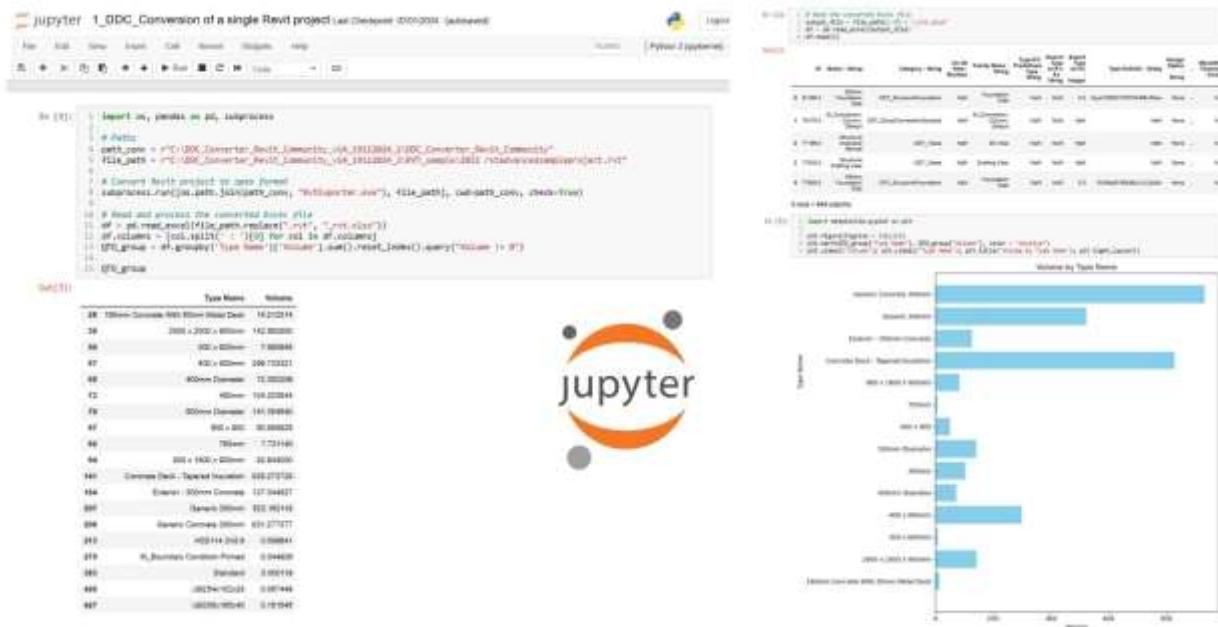
**IDE** (integrirano razvojno okolje) je vsestranski gradnik v računalniku za avtomatizacijo procesov in obdelavo podatkov. Namesto da bi žago, kladivo, vrtalnik in druga orodja hranili ločeno, imate na voljo eno napravo, ki lahko počne vse - reže, pritrjuje, vrta in celo preverja kakovost materialov. IDE za programerje je enoten prostor, kjer lahko pišete kodo (po analogiji z gradbeništvom - ustvarjate risbe), preizkušate njeno delo (sestavljanje modelov stavb), iščete napake (kot pri preverjanju trdnosti konstrukcij pri gradnji) in zaženete dokončan projekt (predaja hiše v uporabo).

Pregled priljubljenih IDE:

- **PyCharm®** (JetBrains) je zmogljiv profesionalni IDE za Python. Zaradi velikega števila vgrajenih funkcij je primeren za resne projekte. Vendar je osnovna podpora za interaktivne datoteke Jupyter (IPYNB) na voljo le v plačljivi različici, vmesnik pa je za začetnike morda preobsežen.

Datoteka s končnico IPYNB (Interactive Python Notebook) je format za interaktivne beležnice Jupyter® (slika 3.4-1), kjer so koda, vizualizacije in razlage združene v enem dokumentu. Ta format je idealen za izdelavo poročil, analitike in scenarijev usposabljanja.

- **VS Code®** (Microsoft) je hitro, prilagodljivo in prilagodljivo orodje z brezplačno podporo IPYNB in številnimi vtičniki. Primerno je tako za začetnike kot za strokovnjake. Omogoča integracijo vtičnikov GitHub Copilot in jezikovnega modela, zato je odlična izbira za projekte umetne inteligence in podatkovne znanosti.
- **Beležnica Jupyter** - Klasična in priljubljena izbira za eksperimentiranje in učenje. Omogoča pisanje kode, dodajanje razlag in vizualizacijo rezultatov v enem samem vmesniku (slika 3.4-1). Idealen je za hitro preverjanje hipotez, delo z LLM in ustvarjanje ponovljivih divjih korakov analize podatkov. Za upravljanje odvisnosti in knjižnic priporočamo uporabo programa Anaconda Navigator, vizualnega vmesnika za upravljanje okolja Python.



Slika 3.4-1 Beležnica Jupyter eno najpriročnejših in najbolj priljubljenih orodij za ustvarjanje procesov Pipeline.

- **Google Collab™** (in platforma Kaggle (slika 9.2-5)) je alternativa Jupyterju v oblaku, ki omogoča brezplačen dostop do GPU/TPU. To je odlična rešitev za začetek - brez lokalne namestitve programske opreme in z možnostjo dela neposredno iz brskalnika. Podpira integracijo z Googlovim diskom in pred kratkim tudi z Geminijem (Googlov 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

Slika 3.4-2 Primerjava IDE: beležnica Jupyter eno najpriročnejših in najlažjih orodij za ustvarjanje procesov Pipeline.

Izbira IDE je odvisna od vaših nalog. Če želite hitro začeti delati z umetno inteligenco, poskusite Jupyter Notebook ali Google Collab. Za resne projekte je bolje uporabiti PyCharm ali VS Code. Glavna stvar je, da začnete. Sodobna orodja vam omogočajo, da svoje poskuse hitro spremenite v delujoče rešitve.

Vsi zgoraj opisani IDE omogočajo ustvarjanje cevovodov za obdelavo podatkov - to so verige modulov blokov kode (ki jih lahko ustvari program LLM), od katerih je vsak na primer odgovoren za drugo fazo:

- analitične scenarije,
- verige pridobivanja informacij iz dokumentov,
- samodejni odzivi na podlagi RAG,
- izdelavo poročil in vizualizacij.

Zaradi modularne strukture je vsak korak mogoče predstaviti kot ločen blok: nalaganje podatkov→filtriranje→ analiza→ vizualizacija→ izvoz rezultatov. Te bloke je mogoče ponovno uporabiti, - prilagoditi in sestaviti v nove verige kot konstruktor, samo za podatke.

Inženirjem, vodjem in analitikom to odpira možnost dokumentiranja logike odločanja v obliki kode, ki jo je mogoče ustvariti z LLM. Ta pristop pomaga pospešiti rutinska opravila, avtomatizirati tipične operacije in ustvariti ponovljive procese, pri katerih je vsak korak jasno dokumentiran in pregleden za vse člane ekipe.

Avtomatizirana orodja ETL Pipelines (slika 7.2-3), Apache Airflow (slika 7.4-4), Apache NiFi (slika 7.4-5) in n8n (slika 7.4-6) za gradnjo logičnih blokov za avtomatizacijo procesov bodo podrobnejše obravnavana v 7. in 8. delu knjige.

## **IDE s podporo LLM in prihodnje spremembe programiranja**

Vključevanje umetne inteligence v razvojne procese spreminja pokrajino programiranja. Sodobna okolja niso več le urejevalniki besedil z označevanjem sintakse - spreminjajo se v intelligentne pomočnike, ki so sposobni razumeti logiko projekta, dopolnjevati kodo in celo razložiti, kako določen del kode deluje. Na trgu se pojavljajo izdelki, ki z uporabo umetne inteligence premikajo meje običajnega razvoja:

- **GitHub Copilot** (povezuje se z VS Code, PyCharm): AI je pomočnik, ki na podlagi komentarjev ali delnih opisov ustvarja kodo in besedilne namige spreminja v dokončane rešitve.
- **Cursor** (vilica VS Code z AI -kernel): omogoča ne le dokončanje pisanja kode, temveč tudi postavljanje vprašanj projektu, iskanje odvisnosti in učenje iz baze kode.
- **JetBrains AI Assistant**: vtičnik za JetBrains IDE (vključno s PyCharmom) s funkcijo razlage zapletene kode, optimizacije in ustvarjanja testov.
- **Amazon CodeWhisperer**: analogen program kot Copilot s poudarkom na varnosti in podpori za Amazonove storitve AWS.

Programiranje se bo v prihodnjih letih močno spremenilo. Glavni poudarek se bo z rutinskega pisanja kode premaknil na oblikovanje modelov in podatkovne arhitekture - razvijalci bodo bolj vključeni v oblikovanje sistema, medtem ko bo umetna inteligenco prevzela naloge, povezane s predlogami: ustvarjanje kode, testi, dokumentacija in osnovne funkcije. Prihodnost programiranja je sodelovanje med ljudmi in umetno inteligenco, kjer stroji prevzemajo tehnično rutino, ljudje pa se osredotočajo na ustvarjalnost.

Programiranje v naravnem jeziku bo postalo vsakdanji pojav. Personalizacija IDE bo dosegla novo raven - razvojna okolja se bodo s predvidevanjem vzorcev, ponujanjem kontekstualnih rešitev in učenjem iz prejšnjih projektov naučila prilagajati slogu dela in podjetjem uporabnika

To ne odpravlja vloge razvijalca, vendar jo bo korenito spremenilo: od pisanja kode k upravljanju znanja, kakovosti in procesov. Ta razvoj bo vplival tudi na poslovno obveščanje, kjer bo ustvarjanje poročil, vizualizacij in aplikacij za podporo odločanju vse bolj potekalo z generiranjem kode in logike s pomočjo umetne inteligence in LLM, klepeta in agentskih vmesnikov.

Ko podjetje vzpostavi klepete LLM in izbere ustrezeno razvojno okolje, je naslednji pomemben korak organizacija podatkov. Ta postopek vključuje pridobivanje podatkov iz različnih virov, njihovo čiščenje, preoblikovanje v strukturirano obliko in vključitev v sisteme podjetja.

Pri sodobnem podatkovnem usmerjenem pristopu k upravljanju podatkov je ključni cilj združiti podatke v enotno univerzalno obliko, ki je združljiva s številnimi orodji in aplikacijami. Za obdelavo postopkov strukturiranja in strukturiranih podatkov so potrebne specializirane knjižnice. Ena od najzmogljivejših, najbolj prilagodljivih in priljubljenih je knjižnica Pandas za Python. Omogoča priročno obdelavo tabelaričnih podatkov: filtriranje, grupiranje, čiščenje, dodajanje, agregiranje in poročanje.

## **Python Pandas: nepogrešljivo orodje za delo s podatki**

Pandas ima posebno mesto v svetu analize in avtomatizacije podatkov. Gre za eno najbolj priljubljenih in pogosto uporabljenih knjižnic programskega jezika Python [86], ki je namenjena delu s strukturiranimi podatki.

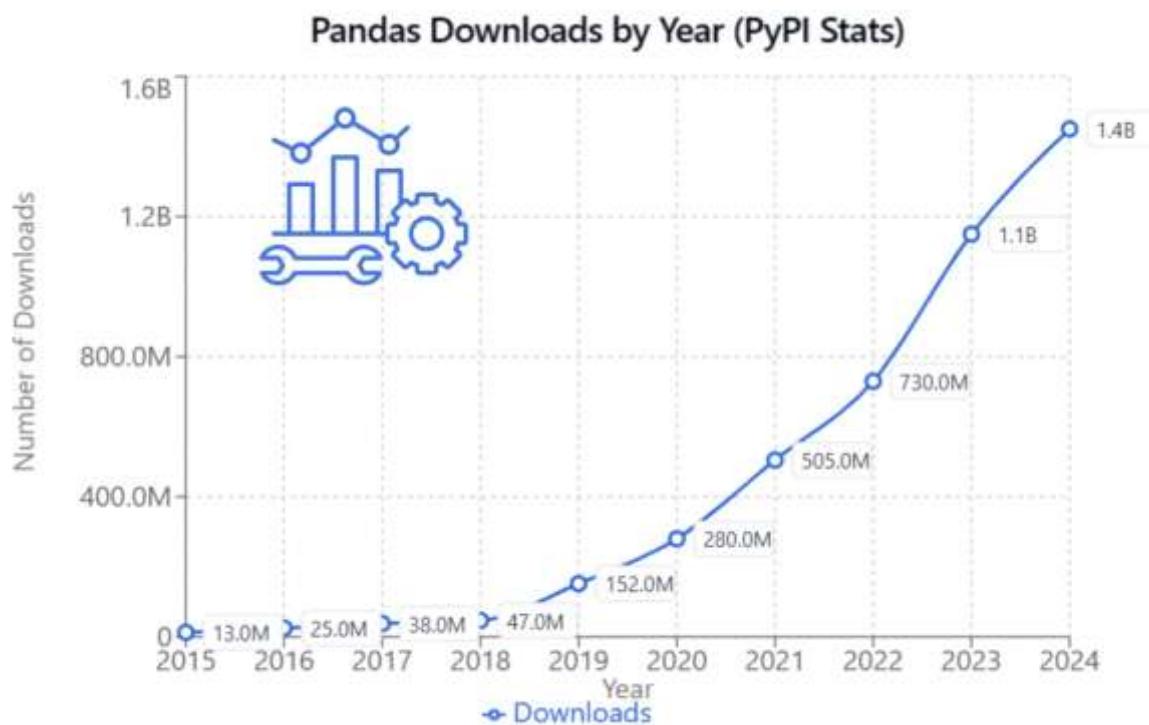
Knjižnica je kot niz pripravljenih orodij: funkcij, modulov, razredov. Tako kot na gradbišču ni treba vsakič znova izumljati kladiva ali ravnila, tako tudi pri programiranju knjižnice omogočajo hitro reševanje težav brez ponovnega izumljanja osnovnih funkcij in rešitev.

Pandas je odprtokodna knjižnica za Python, ki zagotavlja visoko zmogljive in intuitivne podatkovne strukture, zlasti DataFrame, univerzalni format za delo s tabelami. Pandas je švicarski nož za analitike, inženirje in razvijalce, ki delajo s podatki.

Python je programski jezik visoke ravni z enostavno sintakso, ki se aktivno uporablja v analitiki, avtomatizaciji, strojnem učenju in spletnem razvoju. Njegova priljubljenost je posledica berljivosti kode, medplatformnosti in bogatega ekosistema knjižnic. Do danes je bilo za Python ustvarjenih več kot 137.000 odprtakodnih paketov [87], to število pa se skoraj vsak dan povečuje. Vsaka takšna knjižnica je nekakšno skladišče pripravljenih funkcij: od preprostih matematičnih operacij do kompleksnih orodij za obdelavo slik, analizo velikih količin podatkov, nevronske mreže in integracijo z zunanjimi storitvami.

Z drugimi besedami, predstavljajte si, da imate brezplačen in odprt dostop do več sto tisoč programskih rešitev - knjižnic in orodij, ki jih lahko neposredno vključite v svoje poslovne procese. To je kot ogromen katalog aplikacij za avtomatizacijo, analizo, vizualizacijo, integracijo in še več - in vse so na voljo takoj po namestitvi Pythona.

Pandas je eden najbolj priljubljenih paketov v ekosistemu Python. Leta 2022 je povprečno število prenosov knjižnice Pandas doseglo 4 milijone na dan (slika 3.4-3), do začetka leta 2025 pa se je to število povečalo na 12 milijonov prenosov na dan, kar odraža njeno vse večjo priljubljenost in razširjeno uporabo v podatkovni analitiki in klepetu LLM [86]



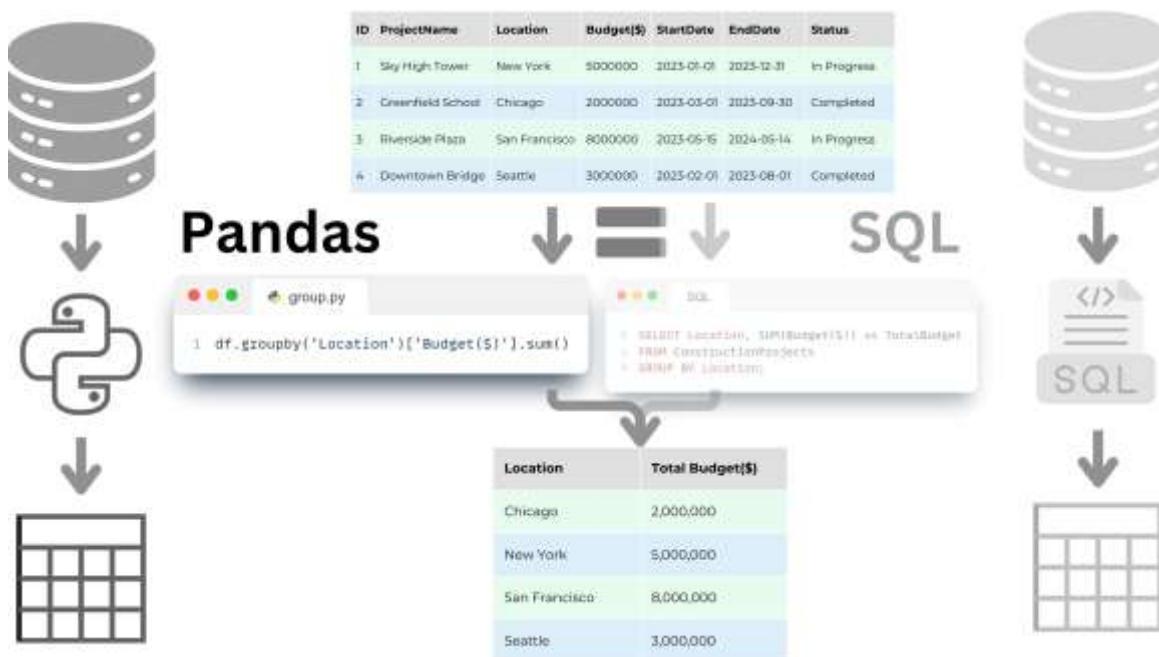
Slika 3.4-3 Pandas je ena izmed najbolj prenesenih knjižnic. Leta 2024 je njeno letno število prenosov preseglo 1,4 milijarde.

Jezik poizvedb v knjižnici Pandas je po funkcionalnosti podoben jeziku poizvedb SQL, ki smo ga obravnavali v poglavju "Relacijske podatkovne zbirke in jezik poizvedb SQL".

V svetu analitike in upravljanja strukturiranih podatkov Pandas odlikujejo preprostost, hitrost in zmogljivost, saj uporabnikom zagotavlja širok nabor orodij za učinkovito analizo in obdelavo informacij.

Obe orodji - SQL in Pandas - zagotavljata zmogljive možnosti za manipulacijo s podatki, zlasti v

primerjavi s tradicionalnim Excelom. Podpirata operacije, kot so izbiranje, filtriranje (slika 3.4-4), razlika je le v tem, da je SQL optimiziran za delo z relacijskimi podatkovnimi zbirkami, medtem ko Pandas obdeluje podatke v pomnilniku RAM, kar mu omogoča delovanje na katerem koli računalniku, ne da bi bilo treba ustvarjati podatkovne zbirke in uvajati ločeno infrastrukturo.

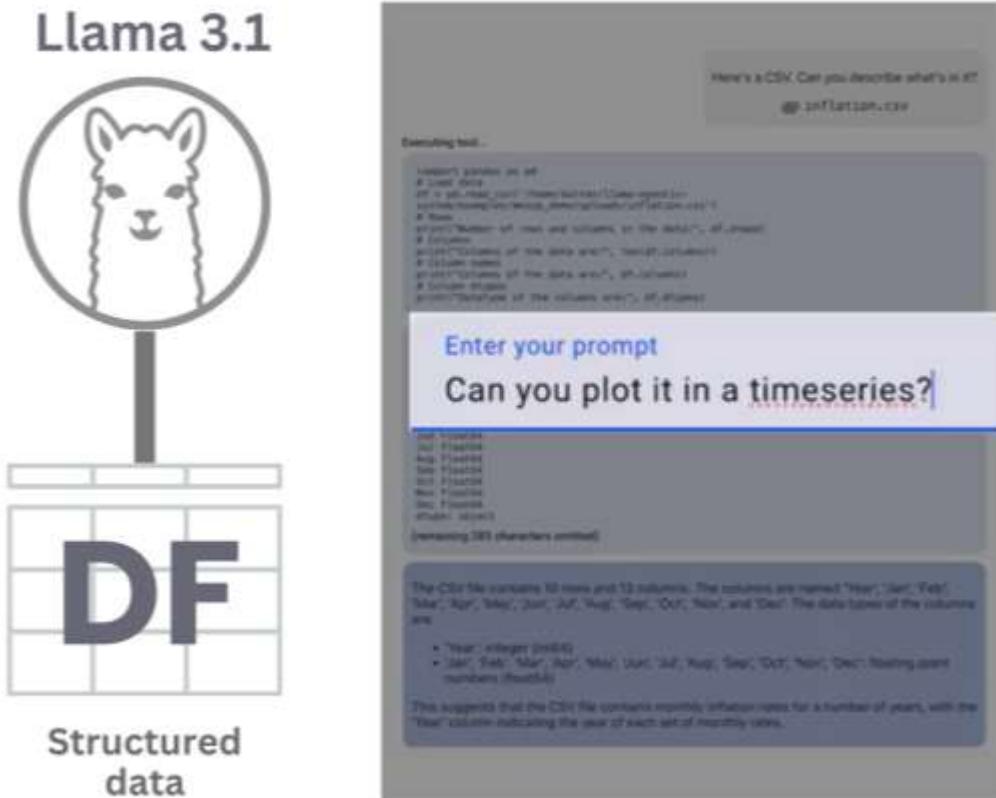


**Slika 3.4-4 Pandas, za razliko od SQL, je prilagodljiv pri delu z različnimi podatkovnimi oblikami, ki niso omejene na podatkovne zbirke.**

Pandas se pogosto raje uporablja za znanstvene raziskave, avtomatizacijo procesov, ustvarjanje cevovodov (vključno z ETL) in manipulacijo podatkov v Pythonu, medtem ko je SQL standard za upravljanje podatkovnih zbirk in se pogosto uporablja v podjetniških okoljih za obdelavo velikih količin podatkov.

Knjižnica Pandas programskega jezika Python omogoča izvajanje ne le osnovnih operacij, kot sta branje in pisanje tabel, temveč tudi bolj zapletena opravila, vključno z združevanjem podatkov, grupiranjem podatkov in izvajanjem zapletenih analitičnih izračunov.

Danes se knjižnica Pandas ne uporablja le v akademskih raziskavah in poslovni analitiki, temveč tudi v povezavi z modeli LLM. Na primer, oddelek Meta® (Facebook™) je ob objavi novega odprtakodnega modela LLaMa 3.1 leta 2024 posebno pozornost namenil delu s strukturiranimi podatki, pri čemer je bila ena ključnih in prvih zadev v njegovi izdaji prav obdelava strukturiranih podatkovnih okvirjev (slika 3.4-5) v formatu CSV in integracija s knjižnico Pandas neposredno v klepetu.



Slika 3.4-5 Eden od prvih in glavnih primerov ekipe Meta, predstavljenih v LlaMa 3.1 leta 2024, je bila gradnja aplikacij z uporabo Pandas.

Pandas je bistveno orodje za milijone podatkovnih znanstvenikov, ki obdelujejo in pripravljajo podatke za generativno umetno inteligenco. Pospešitev programa Pandas z ničelnimi spremembami kode bo velik korak naprej. Podatkovni znanstveniki bodo lahko podatke obdelali v nekaj minutah namesto v urah in pridobili za red velikosti več podatkov za usposabljanje generativnih modelov umetne inteligence [88].

- Jensen Huang, ustanovitelj in izvršni direktor družbe NVIDIA

S programom Pandas je mogoče upravljati in analizirati nabore podatkov, ki daleč presegajo zmožnosti programa Excel. Medtem ko je Excel običajno sposoben obdelati do 1 milijon vrstic podatkov, lahko Pandas brez težav obdela podatkovne nize (slika 9.1-2, slika 9.1-10), ki vsebujejo več deset milijonov vrstic [89]. Ta zmožnost uporabnikom omogoča izvajanje zahtevnih analiz podatkov in vizualizacijo velikih podatkovnih nizov, kar omogoča poglobljen vpogled in lažje sprejemanje odločitev na podlagi podatkov. Poleg tega ima Pandas močno podporo skupnosti [90]: na spletu ali brez njega ga vsak dan uporablja več sto milijonov razvijalcev in analitikov po vsem svetu (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker), kar zagotavlja veliko število izhodnih rešitev za vsak poslovni problem.

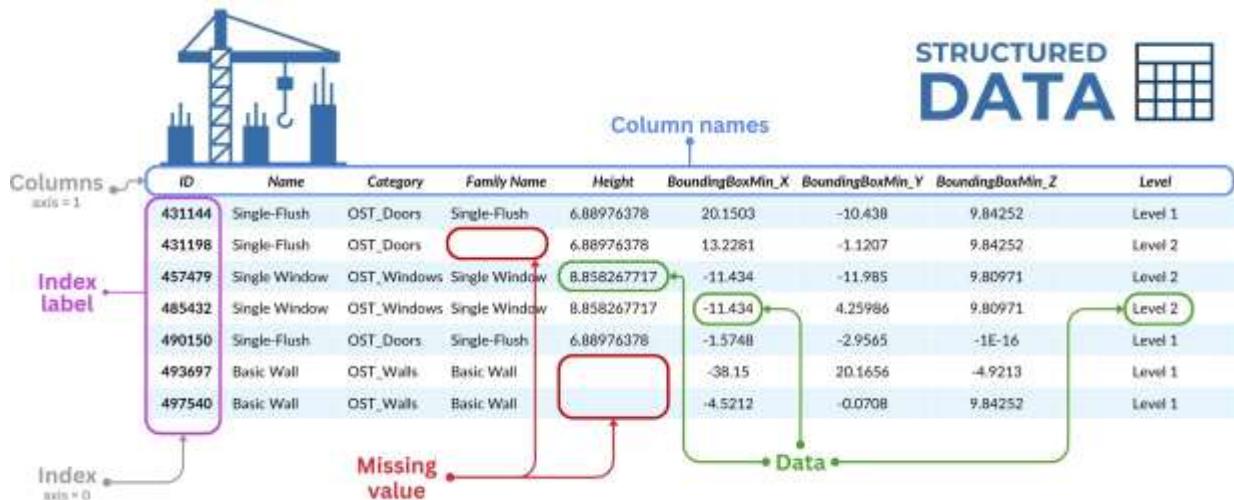
Srce večine analitičnih postopkov v Pythonu je strukturirana oblika podatkov, imenovana DataFrame, ki jo zagotavlja knjižnica Pandas. Gre za zmogljivo in prilagodljivo orodje za organiziranje, analizo in vizualizacijo tabeličnih podatkov.

## DataFrame: univerzalni format tabeličnih podatkov

DataFrame je osrednja struktura v knjižnici Pandas, ki je dvodimenzionalna tabela (slika 3.4-6), kjer vrstice ustreza posameznim predmetom ali zapisom, stolpci pa njihovim značilnostim, parametrom ali kategorijam. Ta struktura vizualno spominja na Excelove preglednice, vendar je po prilagodljivosti, razširljivosti in funkcionalnosti veliko boljša.

**DataFrame** je način predstavitev in obdelave tabeličnih podatkov, shranjenih v pomnilniku RAM računalnika.

DataFrame je način predstavitev in obdelave tabeličnih podatkov, shranjenih v pomnilniku RAM računalnika. V tabeli lahko vrstice prikazujejo na primer elemente gradbenega projekta, stolpci pa njihove lastnosti: kategorije, dimenzijske, koordinate, stroške, pogoje itd. Poleg tega lahko takšna tabela vsebuje tako podatke o enem projektu (slika 4.1-13) kot tudi podatke o milijonih objektov iz tisočih različnih projektov (slika 9.1-10). Zahvaljujoč vektoriziranim operacijam Pandas je tako količino informacij enostavno filtrirati, grupirati in združevati z veliko hitrostjo.



Slika 3.4-6 Gradbeni projekt kot podatkovni okvir je dvodimenzionalna tabela z elementi v vrsticah in atributi v stolpcih.

Nvidia ocenjuje, da se že danes do 30 odstotkov vseh računalniških virov uporablja za obdelavo strukturiranih podatkov - podatkovnih sličic - in ta delež se še povečuje.

*Obdelava podatkov je tisto, kar se v vsakem podjetju opravi verjetno tretjina svetovnega računalništva. Obdelava podatkov in podatki večine podjetij so v DataFrame, v obliki tabele*

- Jensen Huang, izvršni direktor družbe Nvidia [91]

Naštejmo nekaj ključnih funkcij DataFrame v Pandas:

- **Stolpci:** v DataFrame so podatki razvrščeni v stolpce, vsak z edinstvenim imenom. Stolpci atributov lahko vsebujejo podatke različnih vrst, podobno kot stolpci v podatkovnih bazah ali stolpci v tabelah.
- **Pandas Serija** je enodimensionalna podatkovna struktura v programu Pandas, podobna seznamu ali stolpcu v tabeli, kjer vsaka vrednost ustreza drugemu indeksu.  
Serija Pandas ima več kot 400 atributov in metod, zato je delo s podatki izjemno prilagodljivo. Za stolpec lahko neposredno uporabite eno od štiristo razpoložljivih funkcij, izvajate matematične operacije, filtrirate podatke, zamenjate vrednosti, delate z datumimi, nizi in še veliko več. Poleg tega Series podpira vektorske operacije, kar v primerjavi s cikličnimi izračuni močno pospeši obdelavo velikih naborov podatkov. Tako lahko na primer enostavno pomnožite vse vrednosti s številom, zamenjate manjkajoče podatke ali uporabite zapletene transformacije, ne da bi pisali zapletene zanke.
- **Vrstice:** v podatkovnem okviru se lahko indeksirajo z edinstvenimi vrednostmi. Ta indeks vam omogoča hitro spremjanje in prilaganje podatkov v določenih vrsticah.
- **Indeks:** Ko ustvarite podatkovni okvir, Pandas vsaki vrstici privzeto dodeli indeks od 0 do N-1 (kjer je N število vseh vrstic v podatkovnem okvirju). Vendar pa lahko indeks spremenite, tako da vključuje posebne oznake, kot so datumii ali edinstvene značilnosti.
- **Indeksiranje vrstic** v podatkovnem okviru pomeni, da je vsaki vrstici dodeljeno edinstveno ime ali oznaka, ki se imenuje indeks podatkovnega okvira.
- Podatkovne **vrste:** DataFrame podpira različne podatkovne vrste, vključno z: `int`, `float`, `bool`, `datetime64` in `object` za besedilne podatke. Vsak stolpec DataFrame ima svoj podatkovni tip, ki določa, katere operacije je mogoče izvajati z njegovo vsebino.
- **Operacije s podatki:** DataFrame podpira široko paleto operacij za obdelavo podatkov, vključno z združevanjem (`groupby`), združevanjem (`merge` in `join`), povezovanjem (`concat`), razdeljevanjem in združevanjem ter številnimi drugimi tehnikami preoblikovanja podatkov.
- **Manipulacija velikosti:** DataFrame omogoča dodajanje in odstranjevanje stolpcev ter vrstic, zato je dinamična struktura, ki jo lahko spreminjate glede na potrebe analize podatkov.
- **Vizualizacija podatkov:** z uporabo vgrajenih tehnik vizualizacije ali povezovanja s priljubljenimi knjižnicami za vizualizacijo podatkov, kot sta Matplotlib ali Seaborn, lahko DataFrame enostavno pretvorite v grafe in diagrame ter podatke grafično predstavite.

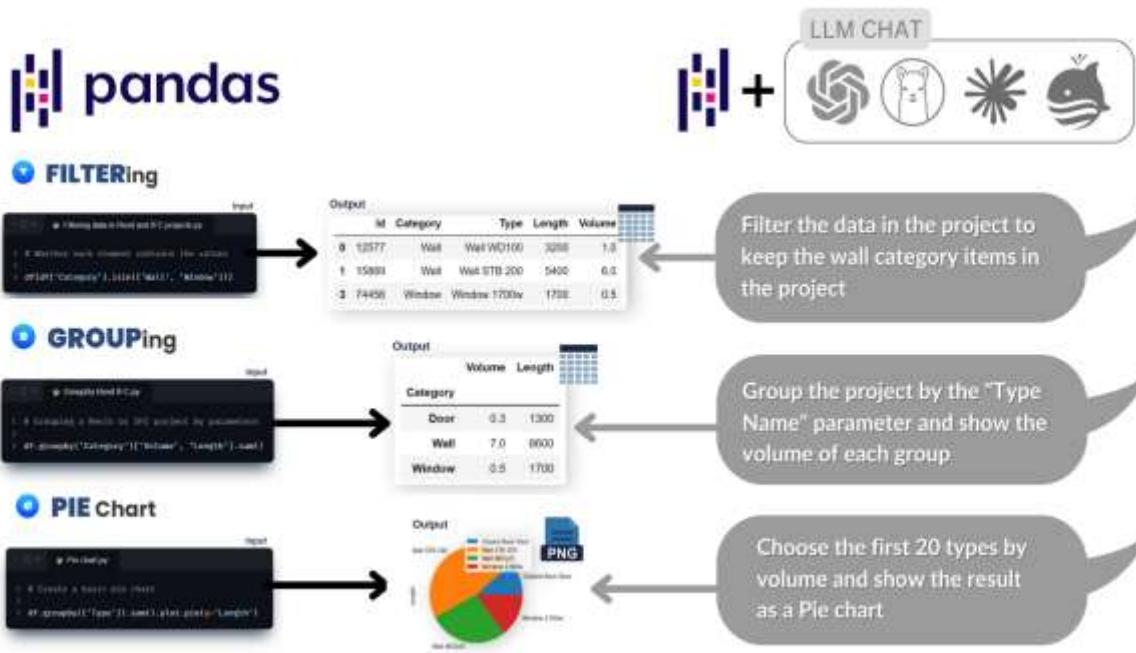
**Vnos in iznos podatkov:** Pandas zagotavlja funkcije za branje, uvoz in izvoz podatkov v različne formate datotek, kot so CSV, Excel, JSON, HTML in SQL, s čimer lahko DataFrame postane osrednje središče za zbiranje in distribucijo podatkov.

Za razliko od CSV in XLSX omogoča Pandas DataFrame večjo prilagodljivost in zmogljivost pri delu s podatki: v pomnilniku RAM lahko obdeluje velike količine informacij, podpira razširjene vrste podatkov (vključno z datumimi, logičnimi vrednostmi in časovnimi vrstami) ter zagotavlja obsežne možnosti za filtriranje, združevanje, združevanje in vizualizacijo podatkov. Medtem ko CSV ne shranjuje informacij o vrstah in strukturi podatkov, XLSX pa je pogosto preobremenjen s formatiranjem in ima nizko skalabilnost, DataFrame ostaja optimalna izbira za hitro analitiko, avtomatizacijo procesov in integracijo z umetno inteligenco -modelov (slika 3.4-7). V naslednjih poglavjih bo podrobno obravnavan vsak od teh vidikov podatkov, prav tako bodo v osmem delu knjige podrobno obravnavani podobni formati, kot so Parquet, Apache Orc, JSON, Feather, HDF5 in podatkovna skladišča (slika 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)

Slika 3.4-7 DataFrame je optimalna izbira za manipulacijo s podatki z visoko zmogljivostjo in napredno podporo podatkovnim tipom.

Knjižnica Pandas in format DataFrame sta zaradi svoje prilagodljivosti, zmogljivosti in enostavne uporabe postala dejanski standard za analizo podatkov v jeziku Python. Sta idealna tako za ustvarjanje preprostih poročil kot za gradnjo zapletenih analitičnih cevovodov, zlasti v povezavi z modeli LLM.



Slika 3.4-8 LLM poenostavi interakcijo s programom Pandas: namesto kode zadostuje besedilna poizvedba.

Danes se Pandas aktivno uporablja v klepetalnicah, ki temeljijo na LLM, kot so ChatGPT, LLaMa, DeepSeek, QWEN in druge. V številnih primerih, ko model prejme poizvedbo, povezano z obdelavo tabel, potrjevanjem podatkov ali analitiko, ustvari kodo prav z uporabo knjižnice Pandas. Zaradi tega je DataFrame naravni "jezik" za predstavitev podatkov v dialogih z umetno inteligenco (slika 3.4-8).

Sodobne podatkovne tehnologije, kot je Pandas, omogočajo lažje analiziranje, avtomatizacijo in vključevanje podatkov v poslovne procese. Hitro zagotavljajo rezultate, zmanjšujejo delovno obremenitev strokovnjakov in zagotavljajo ponovljivo delovanje.

### Naslednji koraki: vzpostavitev trajnostnega podatkovnega okvira

V tem delu smo si ogledali ključne vrste podatkov, ki se uporabljajo v gradbeništvu, spoznali različne formate za njihovo shranjevanje in analizirali vlogo sodobnih orodij, vključno z LLM in IDE, pri obdelavi informacij. Spoznali smo, da je učinkovito upravljanje podatkov temelj za informirano odločanje in avtomatizacijo poslovnih procesov. Organizacije, ki znajo strukturirati in organizirati svoje podatke, pridobijo pomembno konkurenčno prednost v fazah obdelave in preoblikovanja podatkov.

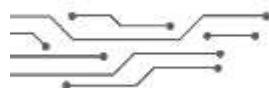
Za povzetek tega dela velja izpostaviti glavne praktične korake, ki vam bodo pomagali pri uporabi obravnavanih pristopov pri vsakodnevnih nalogah:

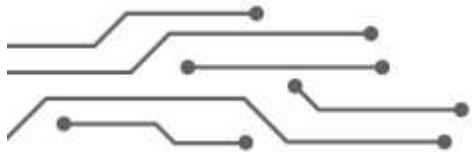
- Izvedite revizijo podatkov o svojih postopkih
- Pripravite seznam vseh podatkovnih vrst, ki jih uporabljate v svojih projektih.

- Določite, katere vrste podatkov in modeli so najbolj pomembni za vaše poslovne procese.
  - opredelitev problematičnih področij, na katerih so informacije pogosto nestrukturirane, slabo strukturirane ali nedostopne
- Začnite oblikovati strategijo upravljanja podatkov
- izpostavljanje vprašanj politike in standardov za ravnanje z različnimi vrstami podatkov.
  - analizirajte, katere delovne postopke lahko izboljšate s pretvorbo nestrukturiranih podatkov v strukturirane.
  - oblikovanje pravilnika o shranjevanju podatkov in dostopu do njih, ki upošteva varnost in zaupnost.
- namestitev in obvladovanje osnovnih orodij za delo s podatki
- Izberite ustrezen IDE, ki ustreza vašim nalogam (npr. namestite VS Code ali Jupyter Notebook)
  - Poskusite ustanoviti lokalno spletno stran LLM, ki bo zaupno obravnavala vaše osebne podatke.
  - Začnite eksperimentirati s knjižnico Pandas za obdelavo tabelaričnih podatkov XLSX
  - opišite mu tipična opravila, ki jih opravljate v orodjih za preglednice ali podatkovnih zbirkah, in ga prosite, naj delo avtomatizira z uporabo programa Pandas.

Z uporabo teh korakov boste lahko postopoma preoblikovali svoj pristop k delu s podatki in prešli od različnih, nestrukturiranih podatkovnih nizov k enotnemu ekosistemu, v katerem podatki postanejo dostopno in razumljivo sredstvo. Začnite z majhnimi koraki - ustvarite svoj prvi okvir DataFrame v programu Pandas, zaženite lokalni LLM, avtomatizirajte svoje prvo rutinsko opravilo z uporabo Pythona (npr. preglednice Excel).

Četrti del knjige se osredotoča na vprašanja kakovosti podatkov, organizacijo, strukturiranje in modeliranje podatkov. Osredotočili se bomo na metodologije, ki različne vire podatkov - od PDF-jev in besedil do slik in modelov CAD - preoblikujejo v strukturirane podatkovne nize, primerne za analizo in avtomatizacijo. Raziskali bomo tudi, kako se formalizirajo podatkovne zahteve, kako se gradijo konceptualni in logični modeli v gradbenih projektih in kako lahko pri tem pomagajo sodobni jezikovni modeli (LLM).





## IV DEL

### **KAKOVOST PODATKOV: ORGANIZACIJA, STRUKTURIRANJE, MODELIRANJE.**

Četrti del se osredotoča na metodologije in tehnologije, ki zagotavljajo pretvorbo raznovrstnih informacij v strukturirane podatkovne nize visoke kakovosti. Podrobno so obravnavani procesi oblikovanja in dokumentiranja podatkovnih zahtev kot podlage za učinkovito informacijsko arhitekturo v gradbenih projektih. Predstavljeni so praktične metode pridobivanja strukturiranih informacij iz različnih virov (PDF -dokumenti, slike, besedilne datoteke, CAD -modeli) s primeri izvedbe. Analizirana je uporaba regularnih izrazov (RegEx) in drugih orodij za samodejno potrjevanje in preverjanje podatkov. Postopek modeliranja podatkov na konceptualni, logični in fizični ravni je opisan korak za korakom, pri čemer so upoštevane posebnosti gradbene industrije. Prikazani so posebni primeri uporabe jezikovnih modelov (LLM) za avtomatizacijo postopkov strukturiranja in potrjevanja podatkov. Predlagani so učinkoviti pristopi k vizualizaciji rezultatov analize, ki povečujejo razpoložljivost analitičnih informacij za vse ravni upravljanja gradbenih projektov.

## POGLAVJE 4.1.

### PRETVORBA PODATKOV V STRUKTURIRANO OBLIKO.

V dobi podatkovno usmerjenega gospodarstva podatki postajajo osnova za sprejemanje odločitev in ne ovira. Namesto nenehnega prilagajanja informacij vsakemu novemu sistemu in njegovim oblikam si podjetja vse bolj prizadevajo oblikovati enoten strukturiran podatkovni model, ki služi kot univerzalni vir resnice za vse procese. Sodobni informacijski sistemi niso zasnovani na podlagi formatov in vmesnikov, temveč na podlagi pomena podatkov - struktura se namreč lahko spreminja, pomen informacij pa ostaja enak veliko dlje časa.

Ključ do učinkovitega dela s podatki ni v njihovem neskončnem pretvarjanju in preoblikovanju, temveč v tem, da jih že od začetka pravilno organiziramo: ustvarimo univerzalno strukturo, ki lahko zagotovi preglednost, avtomatizacijo in integracijo v vseh fazah življenjskega cikla projekta.

Tradicionalni pristop zahteva ročne prilagoditve ob vsaki uvedbi nove platforme: selitev podatkov, spreminjanje imen atributov, prilagajanje formatov. Ti koraki ne izboljšajo kakovosti samih podatkov, temveč le zakrijejo težave in ustvarijo začaran krog neskončnih preoblikovanj. Posledično podjetja postanejo odvisna od določenih programskih rešitev, digitalna preobrazba pa se upočasni.

V naslednjih poglavjih si bomo ogledali, kako pravilno strukturirati podatke, nato pa ustvariti univerzalne modele, zmanjšati odvisnost od platform in se osredotočiti na najpomembnejše - podatke kot strateški vir, okoli katerega se gradijo trajnostni procesi.

#### **Naučite se, kako pretvoriti dokumente, PDF, slike in besedila v strukturirane formate.**

Pri gradbenih projektih je velika večina informacij v nestrukturirani obliki: tehnična dokumentacija, popisi del, risbe, specifikacije, urniki, protokoli. Njihova raznolikost - tako po obliku kot vsebini - otežuje integracijo in avtomatizacijo.

Postopek pretvorbe v strukturirane ali polstrukturirane oblike se lahko razlikuje glede na vrsto vhodnih podatkov in želene rezultate obdelave.

Pretvarjanje podatkov iz nestrukturirane v strukturirano obliko je umetnost in znanost. Ta postopek se razlikuje glede na vrsto vhodnih podatkov in namen analize ter pogosto zavzema pomemben del dela podatkovnega inženirja (slika 3.2-5) in analitika, njegov cilj pa je pripraviti čist, urejen nabor podatkov.



Slika 4.1-1 Pretvarjanje nestrukturiranega optično prebranega dokumenta v strukturirano obliko tabele.

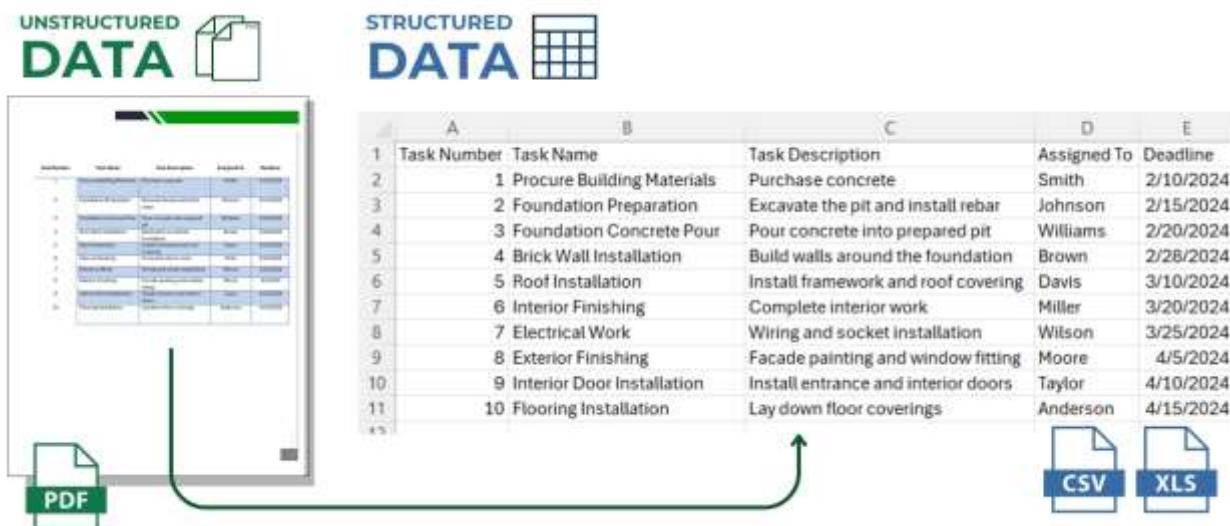
Pretvarjanje dokumentov, PDF, slik in besedil v strukturirano obliko (slika 4.1-1) je postopen postopek, ki vključuje naslednje korake:

- **Izvleček**: V tem koraku se naloži izvorni dokument ali slika, ki vsebuje nestrukturirane podatke. To je lahko na primer dokument PDF, fotografija, risba ali shema.
- **Preoblikovanje podatkov** (Transform): Sledi korak preoblikovanja nestrukturiranih podatkov v strukturirano obliko. To lahko na primer vključuje prepoznavanje in razlago besedila iz slik z uporabo optičnega prepoznavanja znakov (OCR) ali drugih metod obdelave.
- **Nalaganje in shranjevanje podatkov** (Load): zadnji korak vključuje shranjevanje obdelanih podatkov v različnih formatih, kot so CSV, XLSX, XML, JSON, za nadaljnje delo, pri čemer je izbira formata odvisna od posebnih zahtev in želja.

Ta postopek, znan kot ETL (Extract, Transform, Load), ima ključno vlogo pri avtomatizirani obdelavi podatkov in bo podrobneje obravnavan v poglavju "ETL in Pipeline: Extract, Transform, Load". V nadaljevanju si bomo ogledali primere preoblikovanja dokumentov različnih formatov v strukturirane podatke.

### Primer pretvorbe dokumenta PDF v tabelo

Ena najpogostejših nalog pri gradbenih projektih je obdelava specifikacij v formatu PDF. Za prikaz prehoda iz nestrukturiranih podatkov v strukturirano obliko si oglejmo praktični primer: izločanje tabele iz dokumenta PDF in pretvorba v obliko CSV ali Excel (slika 4.1-2).



Slika 4.1-2 Za razliko od PDF so formati CSV in XLSX zelo razširjeni in se zlahka vključujejo v različne sisteme za upravljanje podatkov.

Jezikovni modeli LLM, kot so ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN, močno poenostavljajo način dela podatkovnih znanstvenikov s podatki, saj zmanjšujejo potrebo po poglobljenem učenju programskih jezikov in omogočajo reševanje številnih nalog z besedilnimi poizvedbami.

Zato lahko namesto da bi porabili čas za iskanje rešitev na spletu (običajno na spletišču StackOverflow ali tematskih forumih in klepetalnicah) ali se obrnili na strokovnjake za obdelavo podatkov, uporabimo zmogljivosti sodobnih spletnih ali lokalnih LLM. Dovolj je, da zastavimo poizvedbo, in model nam bo ponudil pripravljeno kodo za pretvorbo dokumenta PDF -dokumenta v tabelarično obliko.

- ❷ Pošljite naslednjo besedilno zahtevo kateremu koli modelu LLM (CHATGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN ali kateremu koli drugemu):

Napišite kodo za izpis besedila iz datoteke PDF, ki vsebuje tabelo. Koda naj kot argument vzame pot do datoteke in vrne izvlečeno tabelo kot podatkovni okvir ↵

- ❷ Odziv modela LLM bo v večini primerov v obliki kode v jeziku Python, saj se ta jezik pogosto uporablja za obdelavo podatkov, avtomatizacijo in delo z različnimi oblikami datotek:

```

1 import fitz # PyMuPDF
2 import pandas as pd
3 import re
4
5 # Define the path to the PDF file
6 pdf_path = r'C\Task Number.pdf'
7
8 # Step 1: Extracting text from the PDF
9 def extract_text(pdf_path):
10     """
11     Extract text from a PDF file.
12     Args: pdf_path (str): Path to the PDF file.
13     Returns: str: All text extracted from the PDF.
14     """
15     document = fitz.open(pdf_path)
16     text = ''
17     for page in document:
18         text += page.get_text()
19     return text
20
21 extracted_text = extract_text(pdf_path)

```

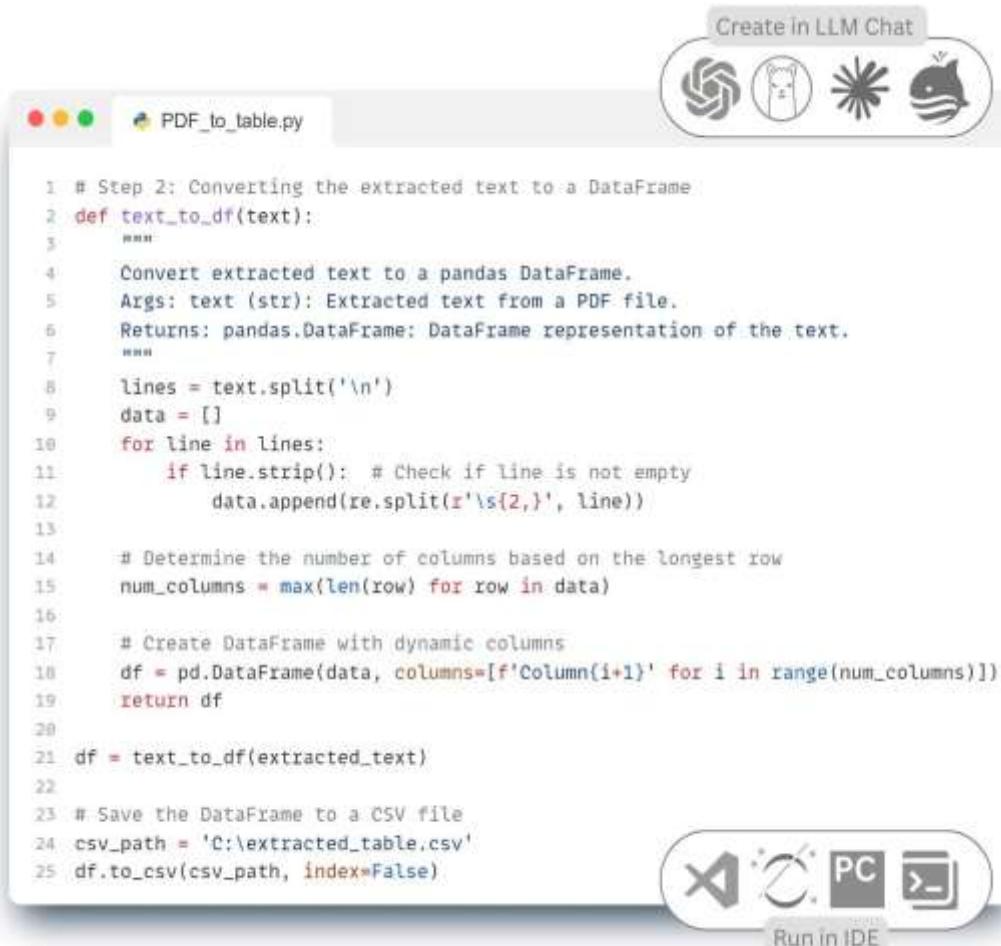
**Slika 4.1-3 Odziv LLM v obliki kode Python ter njenih knjižnic in paketov (Pandas, Fitz) izvleče besedilo iz datoteke PDF.**

To kodo (slika 4.1-3) lahko brez povezave zaženete v enem od priljubljenih IDE, ki smo jih omenili zgoraj: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z vtičnikom PyDev, Thonny, Wing IDE, IntelliJ IDEA z vtičnikom Python, JupyterLab ali priljubljena spletna orodja: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- ❸ V koraku "Convert" uporabimo priljubljeno knjižnico Pandas (ki smo jo podrobno obravnavali v poglavju "Python Pandas: nepogrešljivo orodje za delo s podatki") za branje izpisanega besedila v DataFrame in shranjevanje DataFrame v datoteko tabele CSV ali XLXS:

Potrebujem kodo, ki bo pretvorila tabelo iz datoteke PDF v datoteko DataFrame.  
Dodajte tudi kodo za shranjevanje DataFrame v datoteko CSV. ↵

💡 Odgovor LLM :



The screenshot shows a code editor window titled "PDF\_to\_table.py". The code is as follows:

```

1 # Step 2: Converting the extracted text to a DataFrame
2 def text_to_df(text):
3     """
4         Convert extracted text to a pandas DataFrame.
5         Args: text (str): Extracted text from a PDF file.
6         Returns: pandas.DataFrame: DataFrame representation of the text.
7     """
8     lines = text.split('\n')
9     data = []
10    for line in lines:
11        if line.strip(): # Check if line is not empty
12            data.append(re.split(r'\s{2,}', line))
13
14    # Determine the number of columns based on the longest row
15    num_columns = max(len(row) for row in data)
16
17    # Create DataFrame with dynamic columns
18    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
19    return df
20
21 df = text_to_df(extracted_text)
22
23 # Save the DataFrame to a CSV file
24 csv_path = 'C:\extracted_table.csv'
25 df.to_csv(csv_path, index=False)

```

At the top right of the code editor are icons for "Create in LLM Chat" and four AI models. At the bottom right are icons for "Run in IDE" and "Run in Jupyter".

Slika 4.1-4 Pretvarjanje izpisane tabele iz PDF v DataFrame in shranjevanje tabele v datoteko CSV .

Če se med izvajanjem kode pojavi napaka (slika 4.1-3, slika 4.1-4) - npr. zaradi manjkajočih knjižnic ali napačne poti do datoteke - lahko besedilo napake preprosto kopirate skupaj z izvorno kodo in ga ponovno pošljete v LLM -model. Model bo analiziral sporočilo o napaki, pojasnil, v čem je težava, in predlagal popravke ali dodatne korake.

Na ta način interakcija z umetno inteligenco LLM postane popoln c→→test→ povratne informacije→ popravek - brez potrebe po poglobljenem tehničnem znanju.

Z uporabo poizvedbe navadnega besedila v LLM chat in ducata vrstic programa Python, ki ga lahko zaženemo lokalno v katerem koli IDE, smo pretvorili dokument PDF v tabelarno obliko CSV, ki je za razliko od dokumenta PDF enostavno strojno berljiva in hitro vključena v kateri koli sistem za upravljanje podatkov.

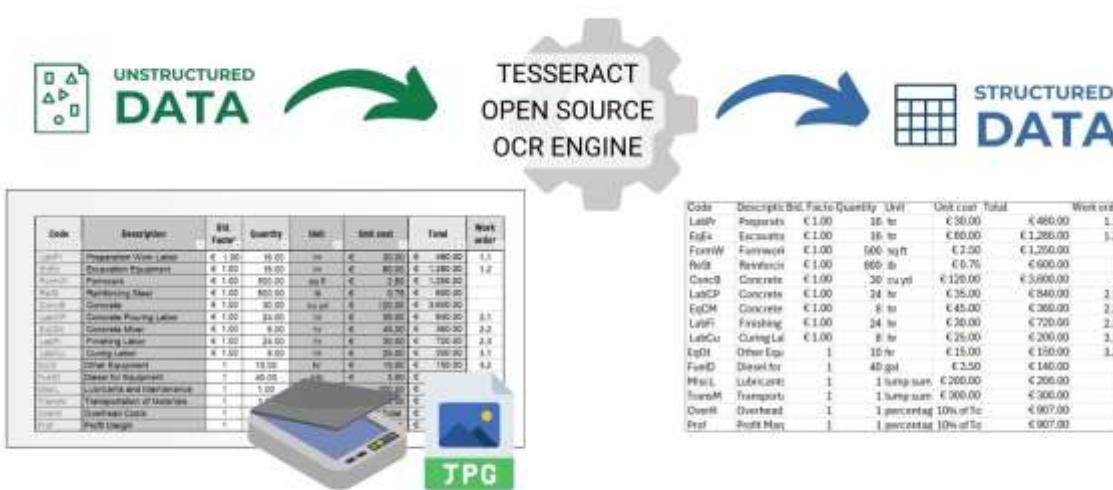
To kodo (slika 4.1-3, slika 4.1-4) lahko s kopiranjem iz katere koli klepetalnice LLM uporabimo desetine ali tisoče novih dokumentov PDF v strežniku in tako avtomatiziramo postopek pretvorbe toka nestrukturiranih dokumentov v strukturirano obliko tabele CSV.

Vendar dokumenti PDF ne vsebujejo vedno besedila, saj gre največkrat za skenirane dokumente, ki jih je treba obdelati kot slike. Čeprav so slike že po naravi nestrukturirane, nam razvoj in uporaba knjižnic za prepoznavanje omogoča pridobivanje, obdelavo in analizo njihove vsebine, kar nam omogoča, da te podatke v celoti izkoristimo v svojih poslovnih procesih.

## Pretvarjanje slik JPEG, PNG v strukturirano obliko

Slike so ena najpogostejših oblik nestrukturiranih podatkov. V gradbeništvu in številnih drugih panogah je ogromno informacij shranjenih v obliki skeniranih dokumentov, shem, fotografij in risb. Takšni podatki vsebujejo dragocene informacije, vendar jih ni mogoče neposredno obdelati, na primer v preglednici Excel ali podatkovni zbirki. Slike vsebujejo veliko kompleksnih informacij, saj so njihova vsebina, barve in tekture raznolike, zato je za pridobivanje uporabnih informacij potrebna posebna obdelava.

Težava pri uporabi slik kot vira podatkov je pomanjkanje strukture. Slike ne izražajo pomena na neposreden, lahko merljiv način, ki bi ga računalnik lahko takoj razumel ali obdelal, kot to počne Excelova preglednica ali tabela v zbirki podatkov. Za pretvorbo nestrukturiranih slikovnih podatkov v strukturirano obliko je treba uporabiti posebne knjižnice, ki so sposobne interpretirati vizualne informacije, ki jih vsebujejo (slika 4.1-5).



Slika 4.1-5 Pretvarjanje skeniranih dokumentov in slik v strukturirane oblike je mogoče s posebnimi orodji OCR.

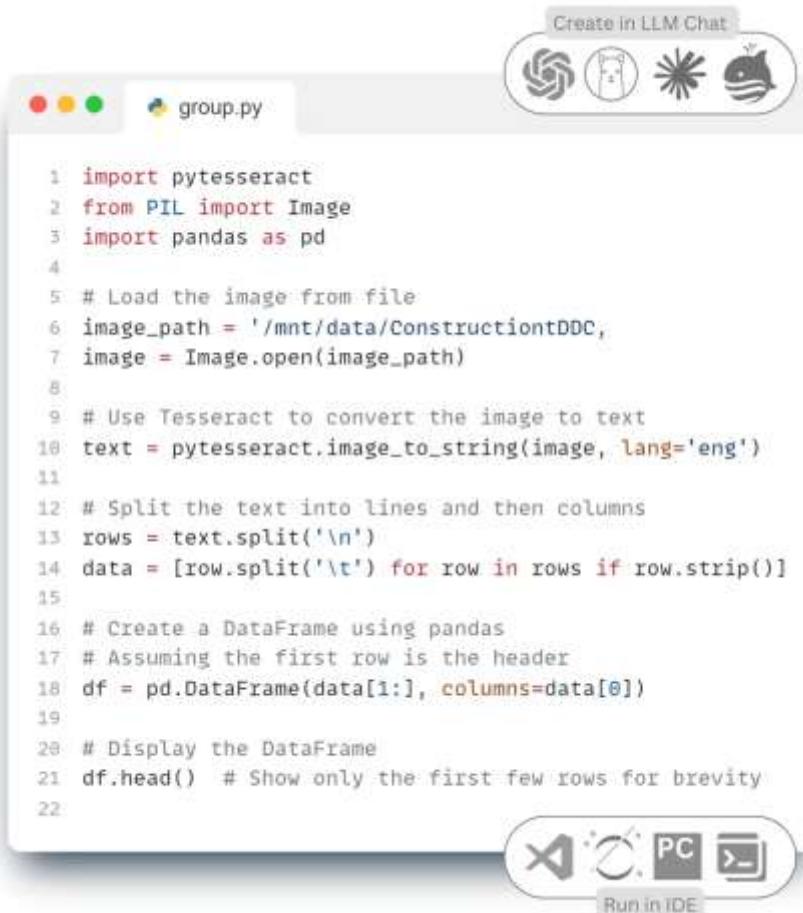
OCR (optično prepoznavanje znakov) se uporablja za pridobivanje besedila iz slik. Omogoča prepoznavanje črk in številk v skeniranih dokumentih, fotografijah in datotekah PDF ter jih spremeni v besedilo, ki ga je mogoče urejati in strojno brati. Tehnologije OCR se že dolgo uporabljajo pri avtomatizaciji dokumentov, danes pa jih zlahka vključite v vse poslovne procese in aplikacije Python - aplikacije. Eno najbolj priljubljenih orodij OCR je Tesseract, odprtokodno orodje, ki ga je prvotno razvil HP™, zdaj pa ga podpira Google™. Podpira več kot 100 jezikov in ima visoko natančnost prepoznavanja.

Prosimo LLM chat, da napiše vzorčno kodo za strukturirano pridobivanje podatkov iz skenirane ali fotografirane tabele.

- ❷ Pošljite besedilno zahtevo v klepet LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ali katero koli drugo):

Napišite kodo za pretvorbo slike JPEG, ki vsebuje tabelo, v tabelo DataFrame ↴

- 2 V odgovoru LLM je v večini primerov predlagana uporaba knjižnice Pytesseract za prepoznavanje besedila na slikah :



```

1 import pytesseract
2 from PIL import Image
3 import pandas as pd
4
5 # Load the image from file
6 image_path = '/mnt/data/ConstructionDDC'
7 image = Image.open(image_path)
8
9 # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head() # Show only the first few rows for brevity
22

```

Slika 4.1-6 Pretvarjanje besedila, pridobljenega iz tabele slik ali fotografij, v strukturirano tabelarno predstavitev.

V tem primeru - koda (slika 4.1-6), ki izhaja iz LLM, uporablja knjižnico pytesseract (Tesseract za Python) za pretvorbo slike v besedilo z uporabo OCR (optično prepoznavanje znakov) in knjižnico Pandas za pretvorbo tega besedila v strukturirano obliko, tj. podatkovni okvir.

Postopek pretvorbe običajno vključuje predobdelavo za izboljšanje kakovosti slike, nato pa se uporabijo različni algoritmi za zaznavanje vzorcev, ekstrakcijo značilnosti ali prepoznavanje predmetov. Tako se nestrukturirane vizualne informacije pretvorijo v strukturirane podatke.

Čeprav so PDF in slike ključni viri nestrukturiranih informacij, je pravi prvak po obsegu besedilo, ustvarjeno v e-poštnih sporočilih, klepetalnicah, sestankih in sporočilnikih. Teh podatkov ni le veliko, temveč so razpršeni, neformalizirani in izredno slabo strukturirani.

## Pretvarjanje besedilnih podatkov v strukturirano obliko

Poleg dokumentov PDF s preglednicami (slika 4.1-2) in skeniranih različic preglednic (slika 4.1-5) je pomemben del informacij v projektni dokumentaciji predstavljen v obliki besedila. To so lahko tako koherentni stavki v besedilnih dokumentih kot tudi fragmentarni zapisi, raztreseni po risbah in shemah. V sodobnih pogojih obdelave podatkov je ena najpogostejših nalog pretvorba takšnega besedila v strukturirano obliko, primerno za analizo, vizualizacijo in odločanje.

Osrednji del tega procesa je taksonomija, sistem klasifikacije, ki omogoča razvrščanje informacij v kategorije in podkategorije na podlagi skupnih značilnosti.

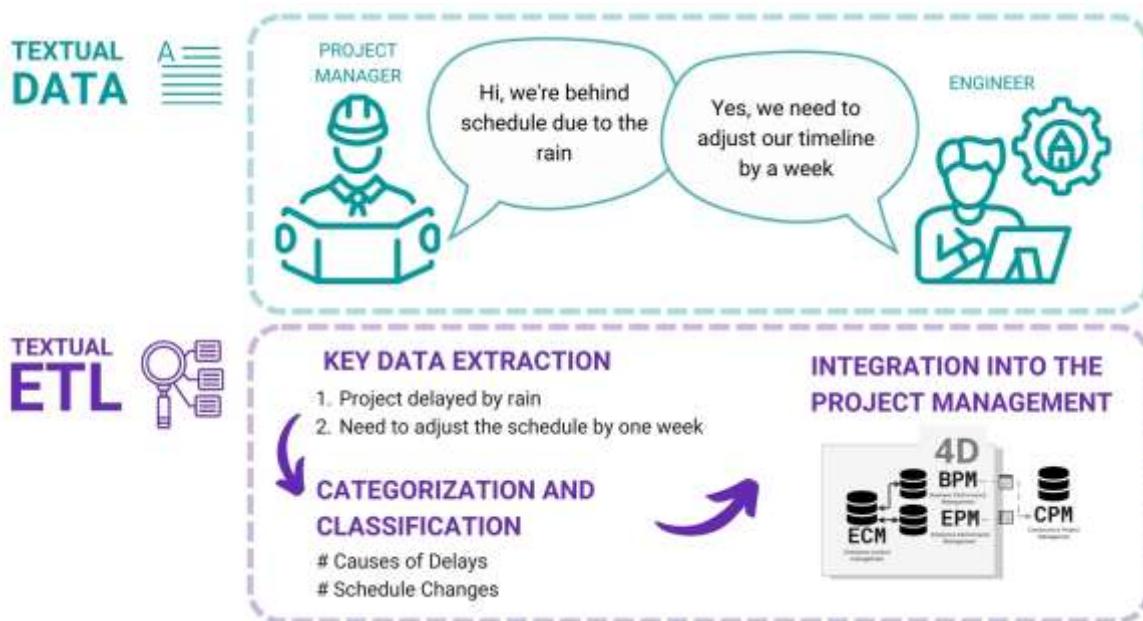
**Taksonomija** je hierarhična klasifikacijska struktura, ki se uporablja za razvrščanje in organiziranje predmetov. V okviru obdelave besedila služi kot osnova za sistematično razvrščanje elementov v semantične kategorije, s čimer se poenostavi analiza in izboljša kakovost obdelave podatkov.

Oblikovanje taksonomije spremljajo koraki ekstrakcije entitet, kategorizacije in kontekstualizacije. Za modeliranje postopka pridobivanja informacij iz besedilnih podatkov so naslednji koraki podobni tistim, ki smo jih že uporabili za strukturiranje podatkov iz dokumentov PDF:

- **Izvleček**): analizirati morate besedilne podatke, da bi izluščili informacije o zamudah in spremembah v časovnem načrtu projekta.
- **Kategorizacija in razvrščanje** (Transofrm): kategorizacija prejetih informacij, npr. razlogov za zamude in spremembe voznega reda.
- **Integracija** (Load): na koncu pripravimo strukturirane podatke za integracijo v zunanje sisteme za upravljanje podatkov.

Razmislite o situaciji: vodja projekta in inženir razpravlja o težavah s časovnim načrtom. Naš cilj je izluščiti ključne elemente (razlogi za zamudo, prilagoditve časovnega načrta) in jih predstaviti na strukturiran način (slika 4.1-7).

Izvedimo ekstrakcijo na podlagi pričakovanih ključnih besed, ustvarimo DataFrame za simulacijo ekstrakcije podatkov in po preoblikovanju novo tabelo DataFrame, ki bo vsebovala stolpce za datum, dogodek (npr. razlog za zamudo) in dejanje (npr. sprememba urnika).



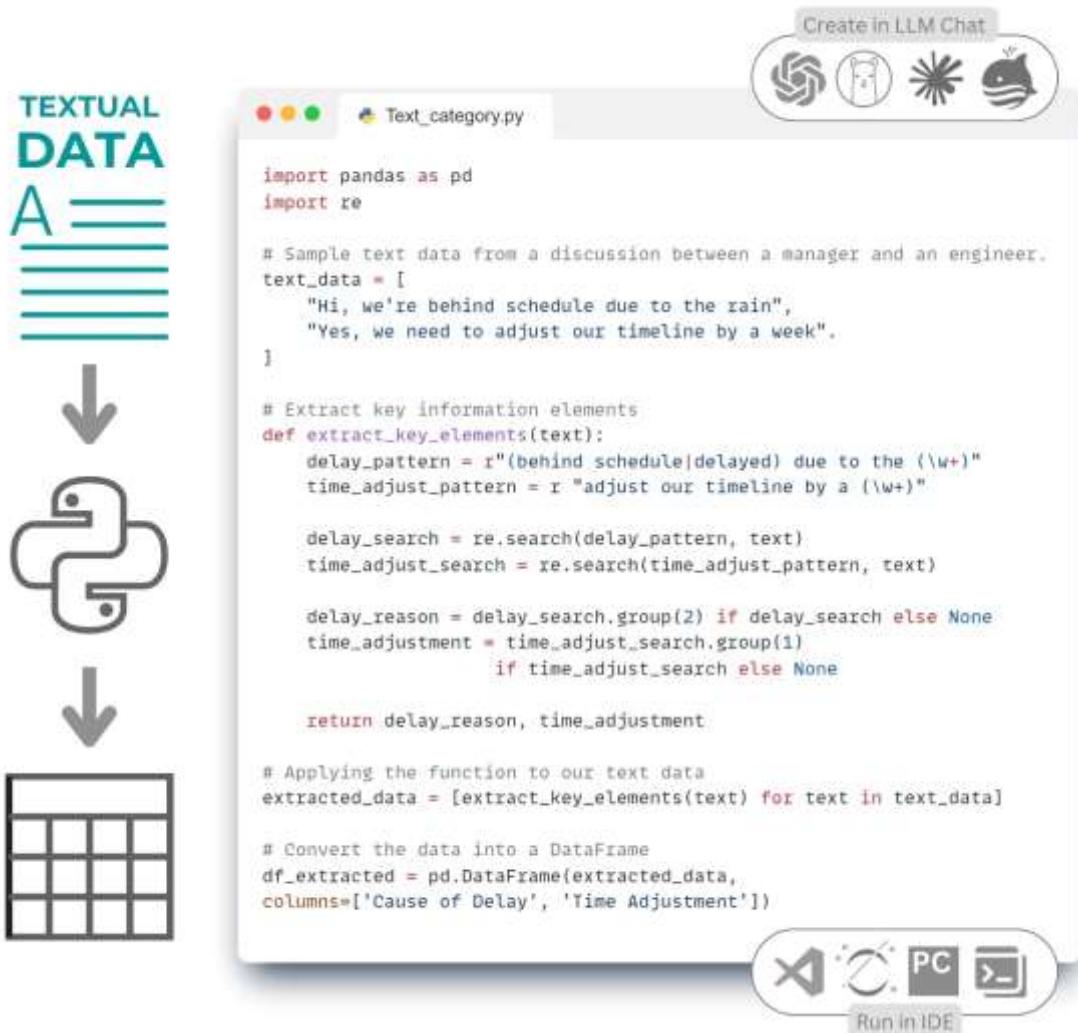
Slika 4.1-7 Poudarjanje ključnih informacij iz besedila o potrebi po prilagoditvi časovnih načrtov in vključitvi sprememb v sistem vodenja projekta.

Tukaj je koda za reševanje problema z uporabo besedilne poizvedbe v enem od jezikovnih modelov kot v prejšnjih primerih.

❷ Pošljite zahtevo za besedilo v katero koli klepetalnico LLM:

Pogovor med vodo: "Pozdravljeni, zaradi dežja zaostajamo za načrtom" in inženirjem: "Da, rok moramo prilagoditi za en teden". Potrebujem skripto, ki bo analizirala prihodnje podobne besedilne dialoge, iz njih pridobila razloge za zamude in potrebne prilagoditve rokov, nato pa iz teh podatkov ustvarila okvir DataFrame. DataFrame je treba nato shraniti v datoteko CSV. ↵

- Odgovor programa LLM bo običajno vključeval kodo v jeziku Python z uporabo regularnih izrazov (re - Regex) in knjižnice Pandas (pd) :

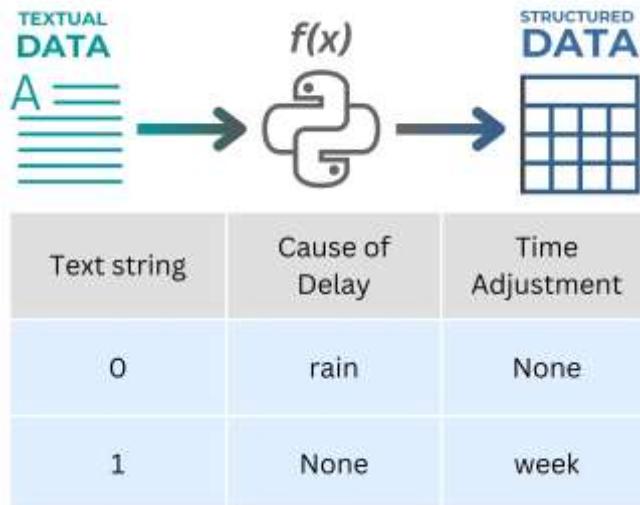


Slika 4.1-8 Poudarjanje ključnih informacij iz besedila o potrebi po prilaganju rokov v tabeli.

V tem primeru (slika 4.1-7) so analizirani besedilni podatki, ki vsebujejo korespondenco med vodjo projekta in inženirjem, da bi ugotovili in izluščili posebne informacije, ki lahko vplivajo na upravljanje prihodnjih projektov s podobnimi dialogi. Z uporabo regularnih izrazov (več o regularnih izrazih bomo obravnavali v poglavju "Strukturirane zahteve in regularni izrazi RegEx") se prek vzorcev ugotovijo vzroki za zamude pri projektu in potrebne prilagoditve časovnega načrta. Funkcija, zapisana v tem primeru, iz nizov na podlagi vzorcev izlušči bodisi vzrok zamude bodisi časovno prilagoditev: kot vzrok zamude izbere besedo za "zaradi", kot časovno prilagoditev pa besedo za "s".

Če je v vrstici omenjena zamuda zaradi vremena, je kot vzrok opredeljen "dež"; če je v vrstici omenjena prilagoditev urnika za določeno obdobje, je to obdobje izločeno kot časovna prilagoditev (slika 4.1-9).

Če v vrstici ni nobene od teh besed, je vrednost "Ni" za ustrezen stolpec z atributom.



Slika 4.1-9 Zbirna tabela, ki se po izvedbi kode pridobi kot podatkovni okvir, vsebuje informacije o obstoju zamud in potrebnih časovnih prilagoditvah.

Strukturiranje in parametrirjanje pogojev iz besedila (dialog, dopis, dokument) omogoča hitro odpravo zamud pri gradnji: na primer, pomanjkanje delavcev lahko vpliva na hitrost dela v slabem vremenu, zato lahko podjetja, ki poznajo parametre zamud iz dialogov (slika 4.1-9) med mojstrom na gradbišču in vodjo projekta - vnaprej, okrepijo ekipo v primeru neugodne napovedi.

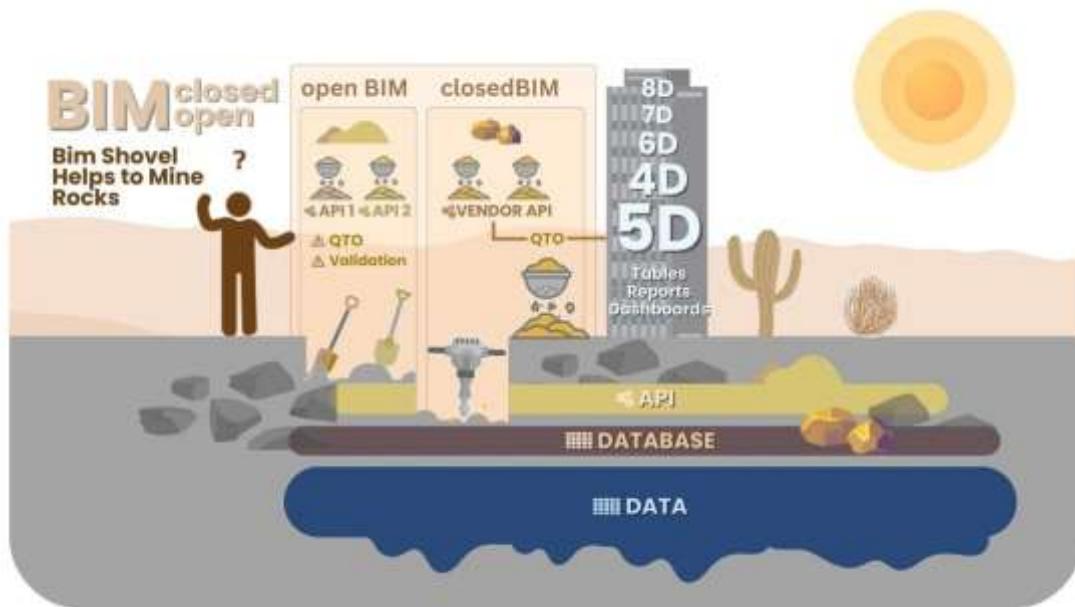
Pretvarjanje dokumentov in slik v strukturirano obliko je mogoče doseči z razmeroma preprostimi, odprtimi in brezplačnimi orodji, ki temeljijo na kategorizaciji.

Kategorizacija elementov je prav tako ključni del dela s projektnimi podatki, zlasti pri uporabi programske opreme CAD (BIM).

## Pretvorba podatkov CAD (BIM) v strukturirano obliko

Strukturiranje in kategoriziranje podatkov CAD (BIM) je bolj zapletena naloga, saj so podatki, shranjeni v podatkovnih zbirkah CAD (BIM), skoraj vedno v zaprtih ali zaplenenih parametričnih oblikah, ki pogosto hkrati združujejo geometrijske podatkovne elemente (polstrukturirane) in metainformacijske elemente (polstrukturirane ali strukturirane podatke).

Nativni podatkovni formati v sistemih CAD (BIM) so običajno zaščiteni in nedostopni za neposredno uporabo, razen če gre za specializirano programsko opremo ali API - vmesnike samega razvijalca (slika 4.1-10). Takšna izolacija podatkov oblikuje zaprte silose za shranjevanje, ki omejujejo prosto izmenjavo informacij in zavirajo oblikovanje celovitih digitalnih procesov v podjetju.



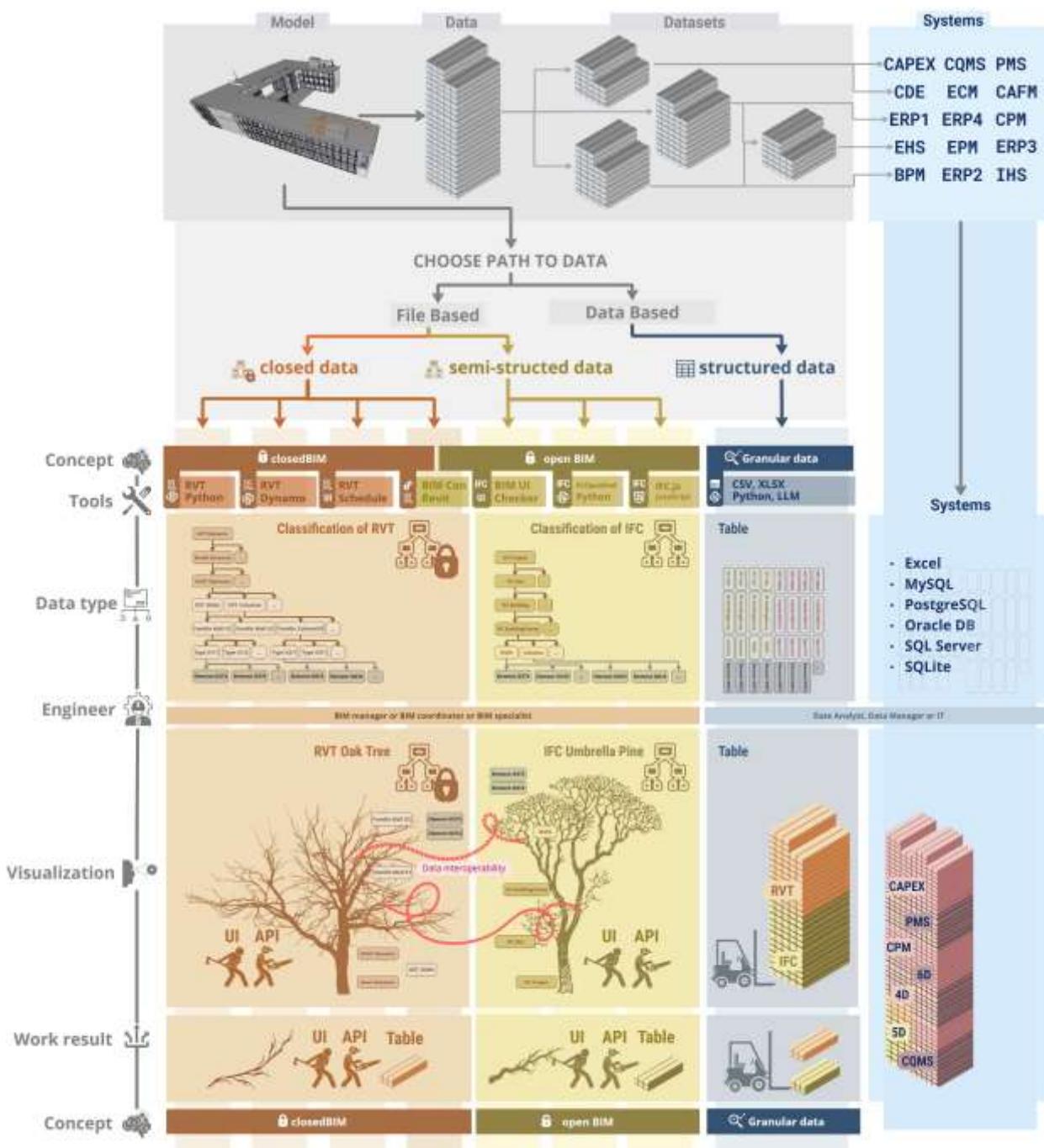
Slika 4.1-10 Strokovnjaki CAD (BIM) lahko do izvornih podatkov dostopajo prek povezav API ali orodij prodajalcev.

V posebnih formatih CAD (BIM) so informacije o značilnostih in atributih elementov projekta zbrane v hierarhičnem klasifikacijskem sistemu, v katerem se entitete z ustreznimi lastnostmi nahajajo kot plodovi sadnega drevesa v najnovejših vozliščih vej klasifikacije podatkov (slika 4.1-11).

Pridobivanje podatkov iz takih hierarhij je mogoče na dva načina: ročno, s klikom na vsako vozlišče, kot da bi obdelovali drevo in s sekiro sekali izbrane veje kategorij in tipov. Druga možnost je uporaba vmesnikov za programiranje aplikacij (API), ki omogoča učinkovitejši, avtomatiziran pristop k pridobivanju in združevanju podatkov ter njihovo pretvorbo v strukturirano tabelo za uporabo v drugih sistemih.

Za pridobivanje strukturiranih podatkovnih tabel iz projektov CAD (BIM) se lahko uporablja različna orodja, kot so Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC ali odprtakodne rešitve, na primer IfcOpSh ali IFCjs za format IFC.

Sodobna orodja za izvoz in pretvorbo podatkov omogočajo poenostavitev obdelave in priprave podatkov z razdelitvijo vsebine modelov CAD na dve ključni komponenti: geometrijske informacije in atributne podatke (slika 4.1-13) - metainformacije, ki opisujejo lastnosti konstrukcijskih elementov (slika 3.1-16). Ti dve podatkovni plasti ostajata povezani prek edinstvenih identifikatorjev, zaradi katerih je mogoče vsak element z opisom geometrije (prek parametrov ali poligonov) natančno povezati z njegovimi atributi: imenom, materialom, stopnjo dokončanosti, stroški itd. Ta pristop zagotavlja celovitost modela in omogoča prožno uporabo podatkov tako za vizualizacijo (podatki o geometrijskem modelu) kot za analitične ali upravljaške naloge (strukturirani ali ohlapno strukturirani), pri čemer se z obema vrstama podatkov dela ločeno ali vzporedno.



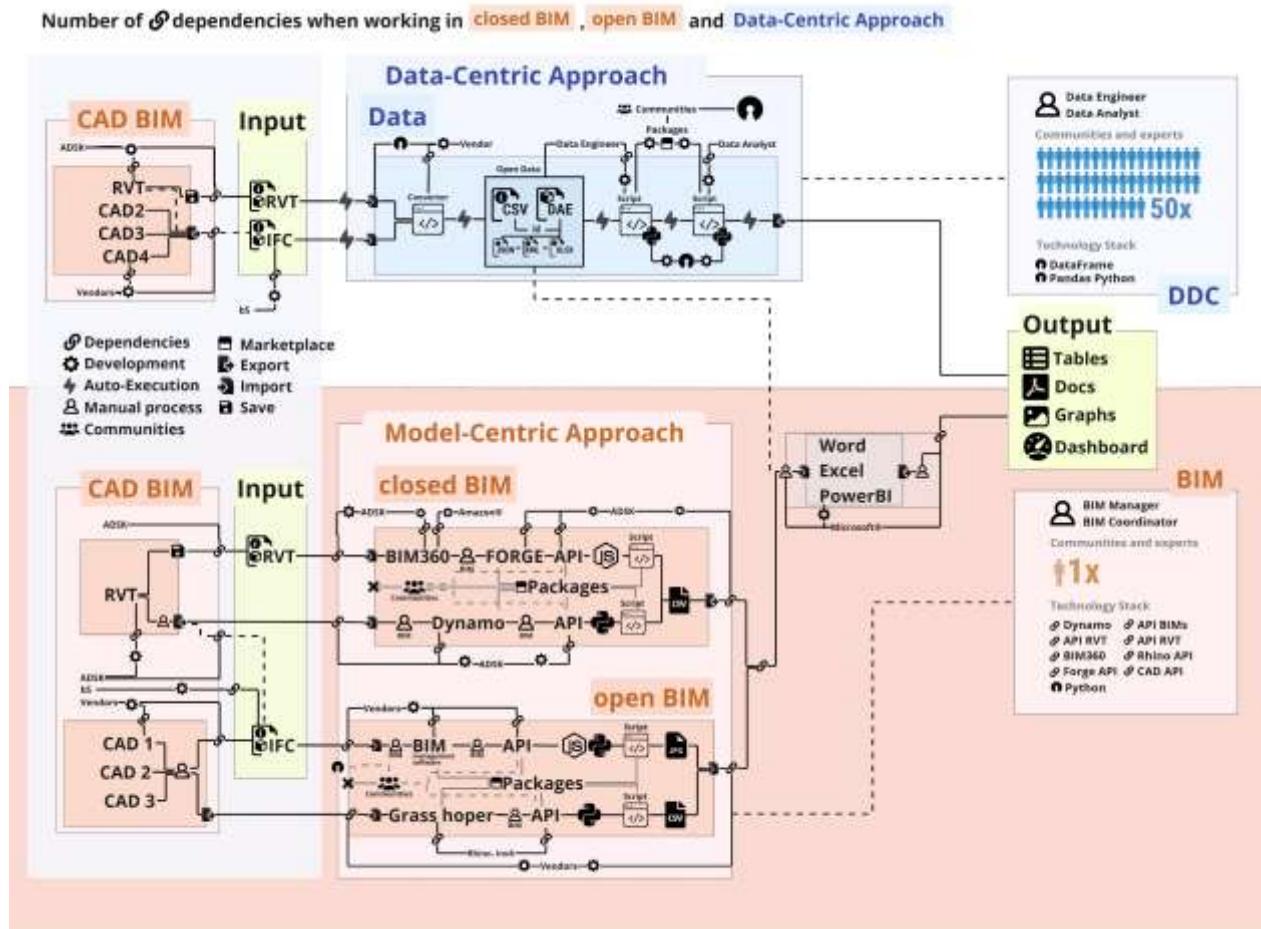
Slika 4.1-11 Prikaz informacij iz podatkovnih zbirk CAD (BIM) je uporabniku predstavljen v obliki klasifikacijskih dreves.

Z razvojem tehnologij povratnega inženiringa in pojavom SDK (Software Development Kit) za pretvorbo podatkov CAD sta postala razpoložljivost in pretvorba podatkov iz zaprtih formatov programske opreme CAD (BIM) veliko lažja. Zdaj je mogoče zakonito in varno pretvoriti podatke iz zaprtih formatov v univerzalne formate, primerne za analizo in uporabo v drugih sistemih. Zgodovina prvih orodij za povratni inženiring ("Open DWG") in boj za prevlado nad formati proizvajalcev CAD sta bila

obravnavana v poglavju "Strukturirani podatki: temelj digitalne preobrazbe".

Orodja za povratni inženiring omogočajo zakonito pridobivanje podatkov iz zaprtih lastniških formatov, pri čemer informacije iz mešanega formata CAD (BIM) razčlenijo v vrste podatkov in formate, ki jih zahteva uporabnik, kar olajša njihovo obdelavo in analizo.

Z uporabo povratnega inženiringa in neposrednega dostopa do informacij iz podatkovnih zbirk CAD omogoča dostop do informacij, odprte podatke in odprta orodja, pa tudi analizo podatkov z uporabo standardnih orodij, izdelavo poročil, vizualizacij in povezovanje z drugimi digitalnimi sistemi (slika 4.1-12).



Slika 4.1-12 Neposredni dostop do podatkov CAD zmanjšuje odvisnost od programskih platform in se približuje podatkovno usmerjenemu pristopu.

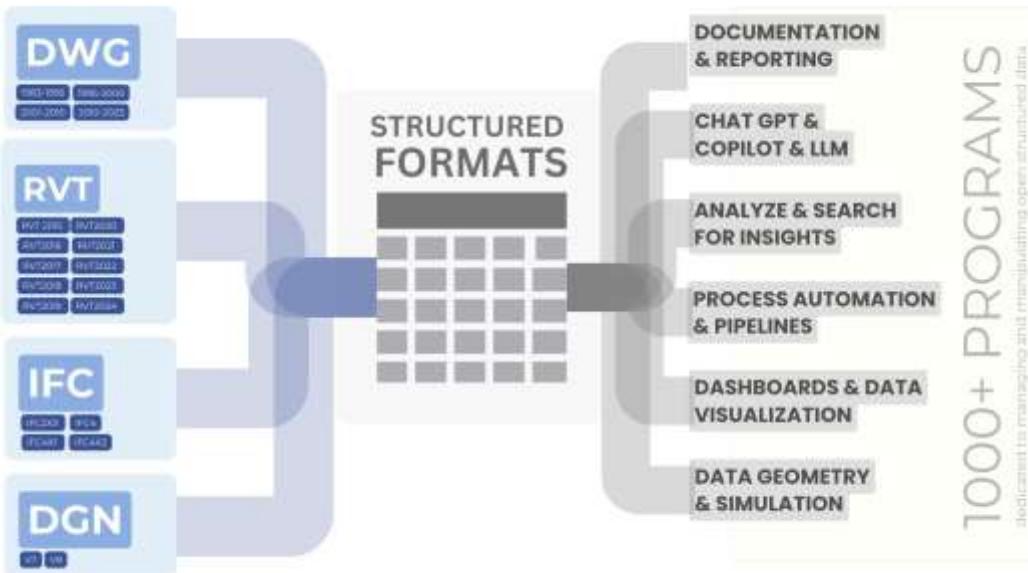
Od leta 1996 za format DWG, od leta 2008 za format DGN in od leta 2018 za RVT je mogoče s pomočjo orodij za obratno inženirstvo (slika 4.1-13) priročno in učinkovito pretvoriti prvotno zaprte formate podatkov CAD v katere koli druge formate, tudi strukturirane formate. Danes skoraj vsa večja podjetja CAD (BIM) in velika inženirska podjetja na svetu uporabljajo SDK - orodja za obratno inženirstvo za pridobivanje podatkov iz zaprtih formatov proizvajalcev CAD (BIM) [92].



**Slika 4.1-13** Uporaba orodij za obratno inženirstvo omogoča pretvorbo podatkovnih baz programov CAD (BIM) v kateri koli primeren podatkovni model.

Pretvarjanje podatkov iz zaprtih, lastniških formatov v odprte formate in ločevanje mešanih formatov CAD (BIM) v geometrijske in metainformacijske atributne podatke poenostavi postopek dela z njimi ter jih naredi dostopne za analizo, manipulacijo in integracijo z drugimi sistemmi (slika 4.1-14).

Pri današnjem delu s podatki CAD (BIM) smo dosegli točko, ko nam za dostop do informacij iz formatov CAD ni treba zahtevati dovoljenja prodajalcev CAD (BIM).



**Slika 4.1-14** Sodobna orodja SDK omogočajo zakonito pretvorbo podatkov iz lastniških formatov podatkovnih baz CAD (BIM).

Trenutne tende na področju obdelave podatkov o oblikovanju CAD še naprej oblikujejo ključni akterji na trgu - prodajalci CAD -, ki si prizadevajo okrepliti svoj položaj v svetu podatkov ter ustvarjajo nove formate in koncepte.

## Ponudniki rešitev CAD se usmerjajo k strukturiranim podatkom

Od leta 2024 dalje se bo v panogi projektiranja in gradbeništva zgodil pomemben tehnološki premik pri uporabi in obdelavi podatkov. Namesto prostega dostopa do podatkov o projektiranju se ponudniki sistemov CAD -sistema osredotočajo na spodbujanje naslednjih novih konceptov. Pristopi, kot sta BIM (ustvarjen leta 2002) in odprti BIM (ustvarjen leta 2012), se postopoma umikajo sodobnim tehnološkim rešitvam, ki jih začenjajo spodbujati prodajalci CAD-sistemov [93]:

- Prehod na uporabo "granularnih" podatkov, ki omogočajo učinkovito upravljanje informacij in prehod na podatkovno analitiko.
- Oblikovanje formata USD in izvajanje pristopa Entity-component-system (ECS) za prožno organizacijo podatkov.
- Aktivna uporaba umetne inteligenčne pri obdelavi podatkov, avtomatizaciji procesov in analizi podatkov.
- Razvoj interoperabilnosti - izboljšana interakcija med različnimi programi, sistemi in zbirkami podatkov.

Vsek od teh vidikov bo podrobneje obravnavan v šestem delu knjige "CAD in BIM: trženje, resničnost in prihodnost projektnih podatkov v gradbeništvu". V tem poglavju bomo le na kratko orisali splošni vektor sprememb: glavni ponudniki CAD si zdaj prizadevajo na novo razmisljiti o načinu strukturiranja podatkov o projektiranju. Eden ključnih premikov je opustitev klasičnega modela shranjevanja na podlagi datotek v korist granularne, analitično usmerjene podatkovne arhitekture, ki zagotavlja stalen dostop do posameznih komponent modela [93].

Bistvo dogajanja je, da industrija opušča okorne, specializirane in parametrične formate, ki zahtevajo geometrijska jedra, v korist bolj univerzalnih, strojno berljivih in prilagodljivih rešitev.

Eden od takšnih dejavnikov sprememb je format USD (Universal Scene Description), ki je bil prvotno razvit v industriji računalniške grafike, vendar je zaradi razvoja platforme NVIDIA Omniverse (in Isaac Sim) za simulacije in vizualizacije že priznan v inženirske aplikacijah [93]. Za razliko od parametričnega IFC, USD ponuja preprostejšo strukturo in omogoča opis geometrije in lastnosti objektov v formatu JSON (slika 4.1-15), kar olajša obdelavo informacij in pospeši njihovo vključevanje v digitalne procese. Novi format omogoča shranjevanje geometrije (poleg BREP -NURBS - več podrobnosti v 6. delu knjige) v obliki poligonov MESH, lastnosti objektov pa v JSON, kar je priročnejše za avtomatizirane procese in delo v ekosistemih v oblaku [94].

Nekateri ponudniki CAD in ERP že uporabljajo podobne formate (npr. NWD, SVF, CP2, CPIXML), vendar jih večina ostaja zaprtih in nedostopnih za zunanjou uporabo, kar omejuje možnosti povezovanja in ponovne uporabe podatkov. V tem kontekstu ima lahko USD enako vlogo, kot jo je imel svoj čas DXF - odprta alternativa lastniškim formatom, kot je 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

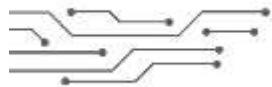
Slika 4.1-15 USD format kot poskus prodajalcev CAD, da bi izpolnili zahtevo po interoperabilnosti in neodvisnosti podatkov o načrtovanju od geometrijskih jedor.

Prehod glavnih razvijalcev na odprte in poenostavljene formate USD, GLTF, OBJ, XML (zaprti NWD, CP2, SVF, SVF2, CPIXML) in podobne formate (slika 3.1-17) odraža svetovni trend in zahteve industrije po poenostavitevi podatkov in večji dostopnosti. V prihodnjih letih lahko pričakujemo postopen odmak od zapletenih parametričnih standardov in formatov z odvisnostjo od geometrijskih jedor v korist lažjih in bolj strukturiranih rešitev. Ta prehod bo pospešil digitalizacijo gradbene industrije, olajšal avtomatizacijo procesov in poenostavil izmenjavo podatkov.

Kljud strateškim načrtom ponudnikov CAD za spodbujanje novih odprtih formatov lahko strokovnjaki v gradbeništvu pridobijo popoln dostop do podatkov iz zaprtih sistemov CAD tudi brez orodij CAD (BIM) z uporabo orodij za obratni inženiring.

Vsi ti trendi neizogibno vodijo k prehodu od obsežnih, monolitnih 3D modelov k univerzalnim, strukturiranim podatkom in k uporabi formatov, ki so se že zdavnaj izkazali v drugih panogah. Ko projektne skupine začnejo modele CAD obravnavati ne le kot vizualne objekte ali nabor datotek, temveč kot podatkovne zbirke, ki vsebujejo znanje in informacije, se pristop k načrtovanju in upravljanju bistveno spremeni.

Ko se ekipe naučijo pridobivati strukturirane podatke iz dokumentov, besedil, risb in modelov CAD ter imajo dostop do podatkovnih zbirk, je naslednji ključni korak modeliranje podatkov in zagotavljanje kakovosti. Prav ta korak v veliki meri določa hitrost obdelave in preoblikovanja informacij, ki se bodo na koncu uporabljale za sprejemanje odločitev pri določenih aplikacijskih nalogah.



## POGLAVJE 4.2.

### KLASIFIKACIJA IN INTEGRACIJA: SKUPNI JEZIK ZA GRADBENE PODATKE

#### Hitrost odločanja je odvisna od kakovosti podatkov

Današnja arhitektura podatkov o oblikovanju se korenito spreminja. Industrija se odmika od obsežnih, izoliranih modelov in zaprtih formatov k bolj prilagodljivim, strojno berljivim strukturam, osredotočenim na analitiko, integracijo in avtomatizacijo procesov. Vendar pa sam prehod na nove formate ne zagotavlja učinkovitosti - v središču pozornosti je neizogibno kakovost samih podatkov.

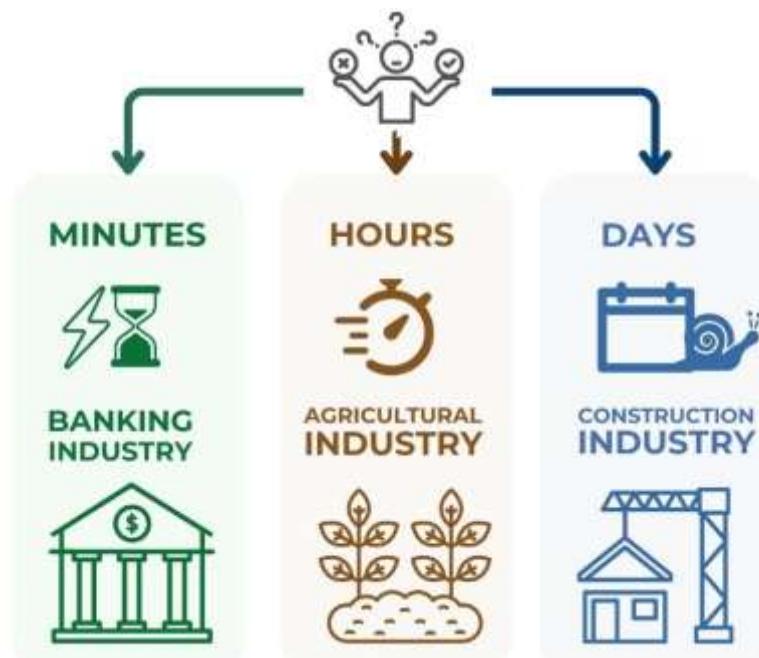
Na straneh te knjige veliko govorimo o formatih, sistemih in procesih. Vendar so vsa ta prizadevanja brez pomena brez enega ključnega elementa: podatkov, ki jim je mogoče zaupati. Kakovost podatkov je temeljni kamen digitalizacije in k njej se bomo vračali v vseh naslednjih delih.

Sodobna gradbena podjetja - zlasti velika - uporabljajo na desetine in včasih na tisoče različnih sistemov in zbirk podatkov (slika 4.2-1). Teh sistemov je treba ne le redno polniti z novimi informacijami, temveč tudi učinkovito medsebojno sodelovati. Vsi novi podatki, ki nastanejo kot rezultat obdelave vhodnih informacij, se vključijo v ta okolja in služijo reševanju določenih poslovnih nalog.

Medtem ko so v preteklosti odločitve o posameznih poslovnih nalogah sprejemali najvišji vodje - tako imenovani HiPPO (slika 2.1-9) - na podlagi izkušenj in intuicije, pa danes, ko se količina informacij močno povečuje, ta pristop postaja sporen. Nadomešča ga avtomatizirana analitika, ki dela s podatki v realnem času.

"Tradicionalno-ročne" razprave o poslovnih procesih na vodstveni ravni se bodo preusmerile v operativno analitiko, ki zahteva hitre odgovore na poslovne poizvedbe.

Obdobje, ko so računovodje, mojstri in ocenjevalci več dni in tednov ročno ustvarjali poročila, zbirne tabele in vitrine s podatki o projektu, je preteklost. Danes hitrost in pravočasnost odločanja postajata ključna dejavnika konkurenčne prednosti.



Slika 4.2-1 V gradbeništvu so za izračune in odločitve potrebni dnevi, za razliko od drugih panog, kjer se to zgodi v urah ali minutah.

Glavna razlika med gradbeništvom in bolj digitalno naprednimi panogami (slika 4.2-1) je nizka raven kakovosti in standardizacije podatkov. Zastareli pristopi k ustvarjanju, prenosu in obdelavi informacij upočasnujejo procese in ustvarjajo kaos. Pomanjkanje enotnih standardov kakovosti podatkov ovira izvajanje celovite avtomatizacije.

Eden glavnih izzivov ostaja slaba kakovost vhodnih podatkov ter pomanjkanje formaliziranih postopkov za njihovo pripravo in potrjevanje. Brez zanesljivih in doslednih podatkov ni učinkovitega povezovanja med sistemi. To vodi do zamud, napak in povečanih stroškov na vseh stopnjah življenjskega cikla projekta.

V naslednjih poglavjih knjige podrobno obravnavamo, kako lahko izboljšate kakovost podatkov, standardizirate procese in skrajšate pot od informacij do kakovostnih, potrjenih in doslednih podatkov.

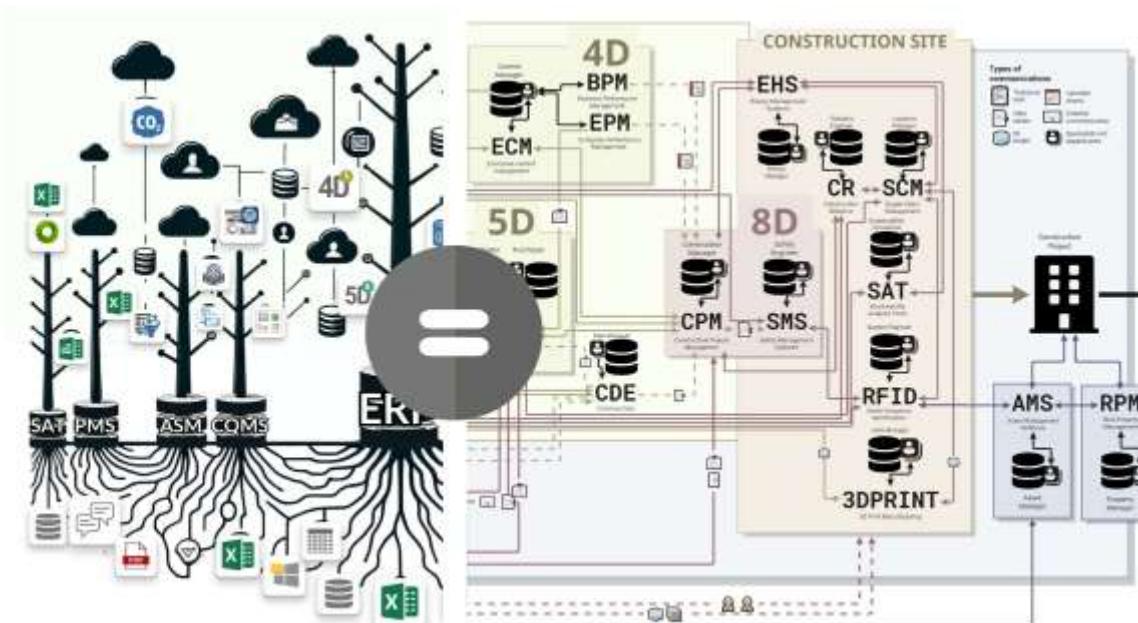
## Standardizacija in integracija podatkov

Za učinkovito upravljanje podatkov je potrebna jasna strategija standardizacije. Samo z jasnimi zahtevami glede strukture in kakovosti podatkov je mogoče avtomatizirati potrjevanje podatkov, zmanjšati število ročnih postopkov in pospešiti sprejemanje odločitev na podlagi informacij v vseh fazah projekta.

V vsakodnevni praksi mora gradbeno podjetje vsak dan obdelati na stotine datotek: e-pošto, dokumente PDF, projektne datoteke CAD, podatke iz senzorjev IOT, ki jih je treba vključiti v poslovne procese podjetja.

Gozd ekosistema podatkovnih zbirk in orodij podjetja (slika 4.2-2) se mora naučiti pridobivati hranila iz vhodnih večplastnih podatkov, da bi dosegel rezultate, ki jih želi podjetje.

Za učinkovito obvladovanje pretoka podatkov ni nujno, da boste zaposlili vojsko vodij, temveč morate najprej razviti stroge zahteve in standarde za podatke ter uporabiti ustrezna orodja za njihovo samodejno potrjevanje, usklajevanje in obdelavo.



Slika 4.2-2 Za zagotavljanje zdravega ekosistema podjetja so potrebni kakovostni in pravočasni viri za njegove sisteme.

Za avtomatizacijo postopka potrjevanja in usklajevanja podatkov (za poznejše samodejno vključevanje) morate najprej opisati minimalne potrebne zahteve za podatke za vsak posamezen sistem. Te zahteve opredeljujejo:

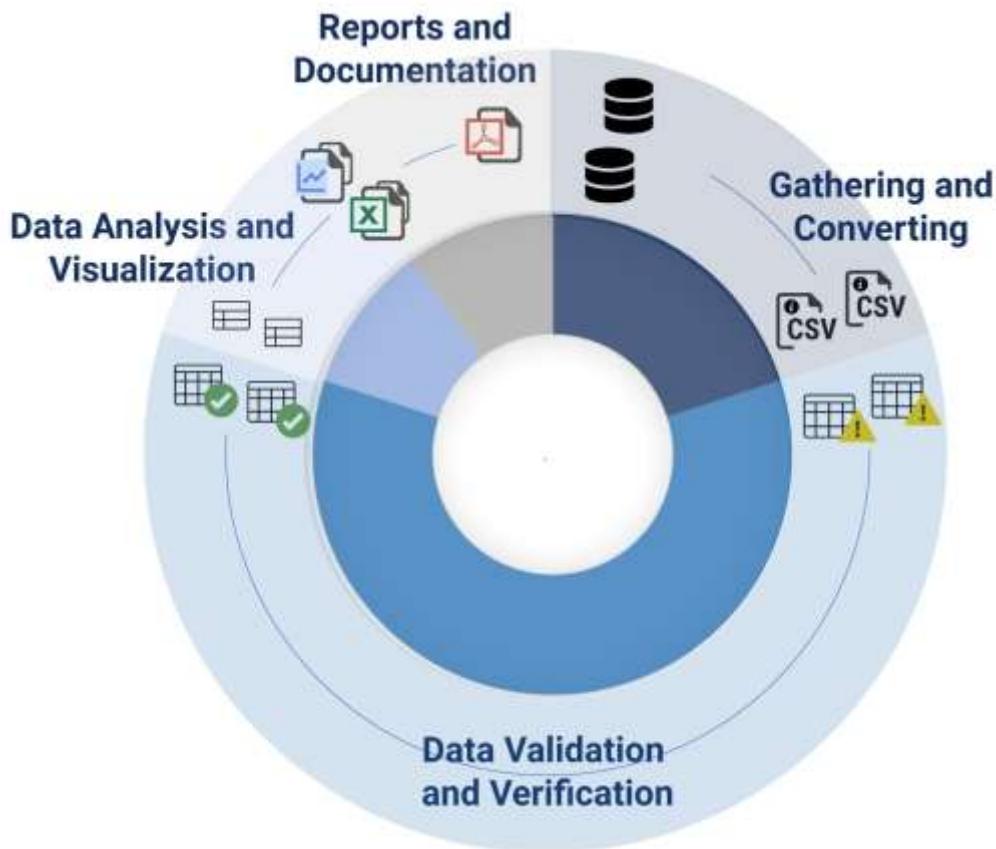
- Kaj točno potrebuje?
- V kakšni obliki (strukturi, formatu)?
- Kateri atributi so obvezni?
- Kakšne tolerance natančnosti in popolnosti so sprejemljive?

Zahteve za podatke opisujejo merila za kakovost, strukturo in popolnost prejetih in obdelanih informacij. Na primer, za besedila v dokumentih PDF je pomembno, da so natančno oblikovana v skladu z industrijskimi standardi (Slika 7.2-14 - Slika 7.2-16). Objekti v modelih CAD - morajo imeti pravilne atribute (dimenzijske, kode, povezave s klasifikatorji) (Slika 7.3-9, Slika 7.3-10). Za skeniranje pogodb pa so pomembni jasni datumi ter možnost samodejnega izpisa zneska in ključnih izrazov (Slika 4.1-7 - Slika 4.1-10).

Oblikovanje podatkovnih zahtev in samodejno preverjanje njihove skladnosti je eden najbolj zamudnih, vendar ključnih korakov. Gre za časovno najzahtevnejši korak v poslovnih procesih.

Kot je omenjeno v tretjem delu te knjige, od 50% do 90% časa strokovnjaki za poslovno obveščanje (BI) porabijo za pripravo podatkov in ne za analizo (slika 3.2-5). Ta postopek vključuje zbiranje, preverjanje, potrjevanje, usklajevanje in strukturiranje podatkov.

Po raziskavi iz leta 2016 [95] so podatkovni znanstveniki z različnih področij širokega spektra navedli, da večino svojega delovnega časa (približno 80%) porabijo za to, kar najmanj radi počnejo (slika 4.2-3): zbiranje obstoječih zbirk podatkov in njihovo organiziranje (poenotenje, strukturiranje). Tako jim manj kot 20 odstotkov časa ostane za ustvarjalne naloge, kot je iskanje vzorcev in zakonitosti, ki vodijo do novih spoznanj in odkritij.



Slika 4.2-3 Preverjanje in zagotavljanje kakovosti podatkov je najdražji, najbolj zamuden in zapleten korak pri pripravi podatkov za vključitev v druge sisteme.

Uspešno upravljanje podatkov v gradbenem podjetju zahteva celovit pristop, ki vključuje parametriranje nalog, oblikovanje zahtev glede kakovosti podatkov in uporabo ustreznih orodij za njihovo samodejno potrjevanje.

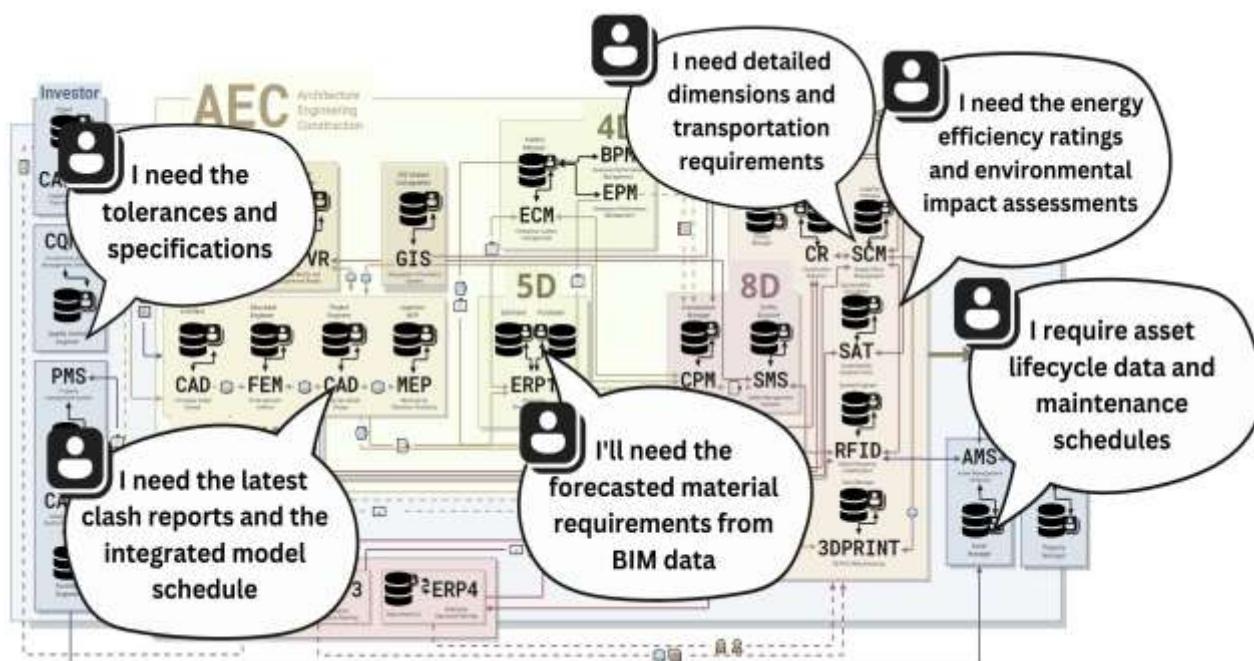
## Digitalna interoperabilnost se začne z zahtevami

Z naraščanjem števila digitalnih sistemov v podjetjih narašča tudi potreba po skladnosti podatkov med njimi. Vodje, ki so odgovorni za različne informacijske sisteme, pogosto ne morejo slediti naraščajoči količini informacij in različnim oblikam. V takšnih okoliščinah so prisiljeni prositi strokovnjake, da ustvarijo podatke v obliki, primerni za uporabo v drugih aplikacijah in platformah.

To pa od inženirjev in osebja, ki sodelujejo pri ustvarjanju podatkov, zahteva, da se prilagodijo številnim zahtevam, pogosto brez preglednosti in jasnega razumevanja, kje in kako se bodo podatki uporabljali v prihodnosti. Pomanjkanje standardiziranih pristopov k ravnanju z informacijami povzroča neučinkovitost in večje stroške v fazi preverjanja, ki je zaradi zapletenosti in nestandardizirane narave podatkov pogosto ročna.

Vprašanje standardizacije podatkov ni le vprašanje udobja ali avtomatizacije. Gre za neposredno finančno izgubo. Glede na poročilo IBM iz leta 2016 znaša letna izguba zaradi slabe kakovosti podatkov v ZDA 3,1 bilijona dolarjev [96]. Poleg tega študije MIT in drugih podjetij za analitično svetovanje kažejo, da lahko stroški slabe kakovosti podatkov znašajo tudi 15-25% prihodkov podjetja [97].

V takšnih razmerah postane ključnega pomena, da so jasno opredeljene zahteve po podatkih in opisi, katere parametre, v kakšni obliki in s kakšno stopnjo podrobnosti je treba vključiti v ustvarjene predmete. Brez formalizacije teh zahtev je nemogoče zagotoviti kakovost podatkov in združljivost med sistemi in fazami projekta (slika 4.2-4).

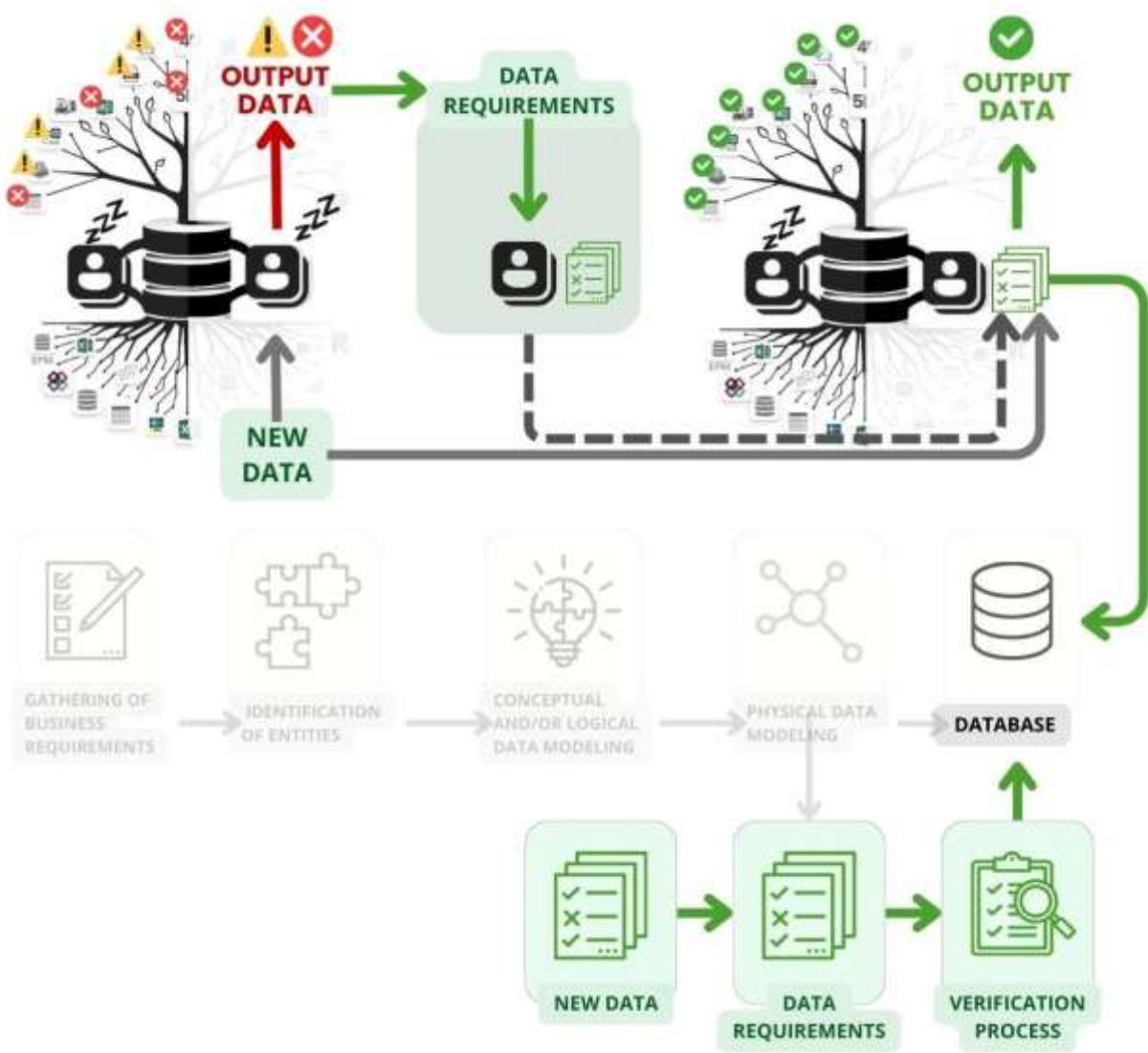


**Slika 4.2-4 Poslovanje temelji na interakciji različnih vlog, od katerih vsaka zahteva določene parametre in vrednosti, ki so ključni za izpolnjevanje poslovnih ciljev.**

Za oblikovanje pravilnih podatkovnih zahtev, morate razumeti poslovne procese na ravni podatkov.

Gradbeni projekti se razlikujejo po vrsti, obsegu in številu udeležencev, vsak sistem - naj bo to modeliranje (CAD (BIM)), načrtovanje (ERP 4D), obračunavanje stroškov (ERP 5D) ali logistika (SCM) - pa zahteva svoje edinstvene parametre za vhodne podatke (vhodne entitete-elementi).

Glede na te potrebe morajo poslovni vodje oblikovati nove podatkovne strukture, da izpolnijo zahteve, ali pa prilagoditi obstoječe tabele in zbirke podatkov. Kakovost ustvarjenih podatkov bo neposredno odvisna od tega, kako natančno in pravilno so oblikovane zahteve (slika 4.2-5).



Slika 4.2-5 Kakovost podatkov je odvisna od kakovosti zahtev, ki so ustvarjene za določene primere uporabe podatkov.

Ker ima vsak sistem svoje posebne zahteve glede podatkov, je prvi korak pri oblikovanju splošnih zahtev razvrstitev vseh elementov, vključenih v poslovne procese. To pomeni, da je treba objekte razdeliti v razrede in skupine razredov, ki ustrezajo določenim sistemom ali aplikacijskim nalogam. Za

vsako takšno skupino se oblikujejo ločene zahteve za podatkovno strukturo, atribute in kakovost.

V praksi pa se izvajanje tega pristopa sooča z velikim izzivom: pomanjkanjem skupnega jezika za združevanje podatkov v skupine. Zaradi različnih klasifikacij, podvojenih identifikatorjev in nezdružljivih formatov vsako podjetje, vsaka programska oprema in celo vsak projekt oblikuje svoje lastne, izolirane podatkovne modele in razrede. Rezultat je digitalni "babilonski stolp", kjer je za prenos informacij med sistemi potrebno večkratno pretvarjanje v prave podatkovne modele in razrede, ki se pogosto izvaja ročno. To oviro je mogoče premagati le s prehodom na univerzalne klasifikatorje in standardizirane nabore zahtev.

## **Skupni jezik gradnje: vloga klasifikatorjev pri digitalni preobrazbi**

V okviru digitalizacije in avtomatizacije postopkov pregleda in obdelave imajo klasifikacijski sistemi posebno vlogo - so nekakšni "digitalni slovarji", ki zagotavljajo enotnost pri opisu in parametrizaciji predmetov. Klasifikatorji tvorijo "skupni jezik", ki omogoča združevanje podatkov po pomenu in povezovanje med različnimi sistemami, ravnni upravljanja in fazami življenjskega cikla projekta.

Najbolj oprijemljiv vpliv sevalnikov je v ekonomiki življenjskega cikla stavbe, kjer je najpomembnejši vidik optimizacija dolgoročnih obratovalnih stroškov. Študije kažejo, da stroški obratovanja predstavljajo do 80% skupnih stroškov lastništva stavbe, kar je trikrat več kot začetni stroški gradnje (slika 4.2-6) [98]. To pomeni, da se odločitev o prihodnjih stroških v veliki meri oblikuje v fazi projektiranja

Zato morajo zahteve operativnih inženirjev (CAFM, AMS, PMS, RPM) postati izhodišče za oblikovanje podatkovnih zahtev v fazi načrtovanja (slika 1.2-4). Teh sistemov ne bi smeli obravnavati kot zadnjo fazo projekta, temveč kot sestavni del celotnega digitalnega ekosistema projekta, od zasnove do razgradnje

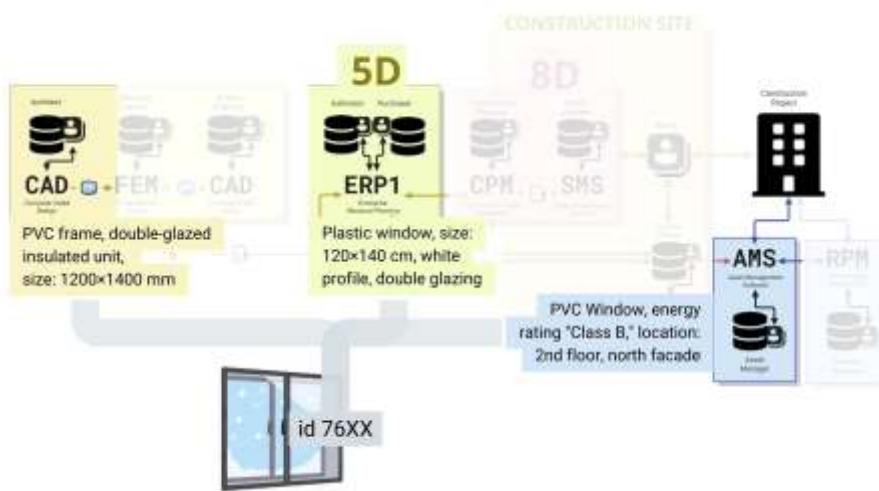
Sodobni klasifikator ni le sistem kod za razvrščanje v skupine. Je mehanizem za medsebojno razumevanje med arhitekti, inženirji, ocenjevalci, logisti, vzdrževalci in informacijskimi sistemi. Tako kot mora avtopilot v avtomobilu nedvoumno in natančno prepoznati cestne objekte, morajo digitalni gradbeni sistemi in njihovi uporabniki prek razreda elementov nedvoumno razlagati isti element projekta za različne sisteme.



Slika 4.2-6 Stroški obratovanja in vzdrževanja trikrat presegajo stroške gradnje in predstavljajo 60-80% skupnih stroškov življenjskega cikla stavbe (po [99]).

Stopnja razvoja klasifikatorja je neposredno povezana z globino digitalizacije podjetja in njegovo digitalno zrelostjo. Organizacije z nizko stopnjo digitalne zrelosti se soočajo z razdrobljenimi podatki, nezdružljivimi informacijskimi sistemi ter posledično z nezdružljivimi in neučinkovitimi klasifikatorji. V takih podjetjih ima lahko isti element v različnih sistemih pogosto različne identifikatorje razvrščanja, kar kritično ovira končno integracijo in onemogoča avtomatizacijo procesov.

Na primer, isto okno v projektu je lahko različno označeno v sistemu CAD model, ocenjevanje in vzdrževanje (slika 4.2-7) zaradi večdimenzionalnega dojemanja elementov s strani različnih udeležencev v procesu. Za ocenjevalca v kategoriji oken so pomembni element, volumen in stroški, za vzdrževalno službo razpoložljivost in možnost vzdrževanja, za arhitekta pa estetske in funkcionalne značilnosti. Posledično lahko isti element zahteva različne parametre.



Slika 4.2-7 Pri nedoslednem razvrščanju med sistemi bo element na vsakem koraku prehoda v drug sistem izgubil nekaj informacij o atributih.

Ker je klasifikacijo gradbenih elementov težko nedvoumno opredeliti, strokovnjaki z različnih področij

istem elementu pogosto dodelijo nezdružljive razrede. Zaradi tega se izgubi enoten pogled na objekt, kar zahteva naknadno ročno posredovanje, da se uskladijo različni sistemi razvrščanja ter vzpostavi skladnost med tipi in razredi, ki so jih opredelili različni strokovnjaki.

Zaradi te nedoslednosti operativne dokumentacije, ki jo prejme oddelek za nabavo (ERP) ob nakupu gradbenega elementa od proizvajalca, pogosto ni mogoče pravilno povezati z razvrstitvijo tega elementa na gradbišču (PMIS, SCM). Posledično je malo verjetno, da bodo kritične informacije vključene v sisteme upravljanja infrastrukture in sredstev (CAFM, AMS), kar povzroča resne težave med zagonom, pa tudi med poznejšim vzdrževanjem (AMS, RPM) ali zamenjavo elementa.

V podjetjih z visoko digitalno zrelostjo imajo klasifikatorji vlogo živčnega sistema, ki združuje vse informacijske tokove. Isti element dobi edinstven identifikator, ki omogoča, da se prenaša med sistemi CAD, ERP, AMS in CAFM ter njihovimi klasifikatorji brez popačenja ali izgube.

Če želite izdelati učinkovite klasifikatorje, morate razumeti, kako se podatki uporabljajo. Isti inženir lahko v različnih projektih različno poimenuje in razvrsti element. Le z večletnim zbiranjem statističnih podatkov o uporabi je mogoče razviti stabilen sistem razvrščanja. Pri tem pomaga strojno učenje: algoritmi analizirajo na tisoče projektov (slika 9.1-10) in s pomočjo strojnega učenja določijo verjetne razrede in parametre (slika 10.1-6). Samodejno razvrščanje je še posebej dragoceno v okoljih, kjer ročno razvrščanje zaradi količine podatkov ni mogoče. Sistemi samodejnega razvrščanja bodo lahko razlikovali osnovne kategorije na podlagi minimalno izpolnjenih parametrov elementov (več podrobnosti v devetem in desetem delu knjige).

Razviti sistemi za razvrščanje postanejo katalizatorji za nadaljnjo digitalizacijo in ustvarijo podlago za:

- Avtomatizirano ocenjevanje stroškov in časovnih okvirov projekta.
- Prediktivna analiza morebitnih tveganj in konfliktov
- Optimizacija postopkov nabave in logističnih verig
- Ustvarjanje digitalnih dvojčkov stavb in struktur
- Integracija s sistemi pametnih mest in interneta stvari

Čas za preoblikovanje je omejen - z razvojem tehnologij strojnega učenja in računalniškega vida bo problem samodejnega razvrščanja, ki je bil desetletja nerešljiv, rešen v prihodnjih letih, gradbena in projektantska podjetja, ki se ne bodo pravočasno prilagodila, pa tvegajo ponovitev usode taksi flot, ki so jih izpodrinile digitalne platforme.

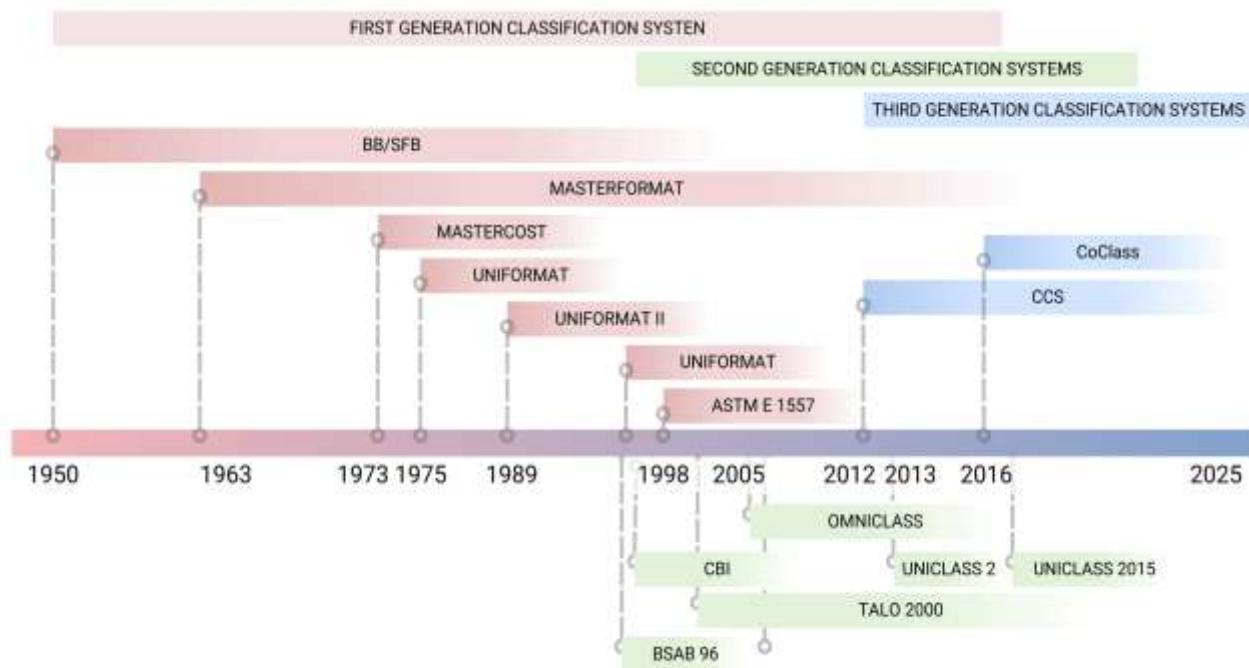
Avtomatizacija izračunov stroškov in časovnega načrta ter veliki podatki in strojno učenje bodo podrobneje obravnavani v petem in devetem delu knjige. Nevarnost ponovitve usode voznih parkov taksijev in uberizacije gradbeništva sta podrobno obravnavana v desetem delu knjige.

Za razumevanje ključne vloge klasifikatorjev pri digitalnem preoblikovanju gradbene industrije se je treba posvetiti zgodovini njihovega razvoja. Prav zgodovinski kontekst nam omogoča, da se zavemo, kako so se razvijali pristopi h klasifikaciji in kateri trendi določajo njihovo sedanje stanje.

## Masterformat, OmniClass, Uniclass in CoClass: razvoj klasifikacijskih sistemov

Zgodovinsko gledano so se gradbeni elementi in klasifikatorji del razvijali v treh generacijah, od katerih je vsaka odražala raven razpoložljive tehnologije in trenutne potrebe industrije v določenem časovnem obdobju (slika 4.2-8):

- **Prva generacija** (od začetka 50. do konca 80. let) - papirnati imeniki, hierarhični klasifikatorji se uporabljajo lokalno (npr. Masterformat, SfB).
- **Druga generacija** (od poznih devetdesetih do sredine leta 2010) so preglednice in strukturirane zbirke podatkov, ki se izvajajo v programih Excel in Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Tretja generacija** (od leta 2010 do danes) - digitalne storitve in vmesniki API - vmesniki, integracija s CAD (BIM), automatizacija (Uniclass 2015, CoClass).



Slika 4.2-8 Tri generacije klasifikatorjev za gradbeništvo.

V zadnjih desetletjih se je hierarhična zapletenost (slika 4.2-9) klasifikatorjev zmanjšala: medtem ko so zgodnji sistemi, kot je OmniClass, za opis 6887 razredov uporabljali do 7 ravni gnezdenja, so sodobne rešitve, kot je CoClass, omejene na 3 ravni s 750 razredi. To olajša delo s podatki in hkrati ohranja potrebno granularnost. Uniclass 2015, ki se pogosto uporablja kot standard v Združenem kraljestvu, združuje 7210 razredov na samo 4 ravneh, zaradi česar je primeren za projekte CAD in javna naročila.

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

Slika 4.2-9 Z vsako novo generacijo klasifikatorjev se zahtevnost kategorizacije skokovito zmanjšuje.

V sistemih za ocenjevanje gradnje v različnih državah je zaradi razlik v klasifikacijah celo tipičen element, kot je betonski temeljni zid, lahko opisan na zelo različne načine (slika 4.2-10). Te razlike odražajo nacionalne gradbene prakse, uporabljene meritve sisteme, pristope h klasifikaciji materialov ter regulativne in tehnične zahteve, ki veljajo v posamezni državi.



Slika 4.2-10 Isti element se v projektih v različnih državah uporablja z različnimi opisi in klasifikacijami.

Različne klasifikacije istih elementov otežujejo mednarodno sodelovanje in povzročajo, da je primerjava stroškov in obsega dela v okviru mednarodnih projektov dolgotrajna in včasih skoraj nemogoča. Trenutno na svetovni ravni ni univerzalnega klasifikatorja - vsaka država ali regija razvija svoje sisteme, ki temeljijo na lokalnih normah, jeziku in poslovni kulturi:

- **CCS** (Danska): Sistem razvrščanja stroškov - sistem za razvrščanje stroškov v celotnem življenjskem ciklu objekta (načrtovanje, gradnja, obratovanje). Poudarek je na logiki obratovanja in vzdrževanja, vključuje pa tudi upravljanje proračuna in virov.
- **NS 3451** (Norveška): kategorizira objekte po funkcijah, oblikovnih elementih in fazah življenjskega cikla. Uporablja se za vodenje projektov, oceno stroškov in dolgoročno načrtovanje.
- **MasterFormat** (ZDA): sistem za strukturiranje gradbenih specifikacij v razdelke (npr. beton, elektrika, zaključna dela). Osredotoča se na discipline in vrste del namesto na funkcionalne

elemente (za razliko od UniFormata).

- **Uniclass 2** (Združeno kraljestvo): eden od najbolj podrobnih klasifikatorjev, ki se uporablja pri javnih naročilih in projektih BIM. Združuje podatke o objektih, delih, materialih in prostorih v enoten sistem.
- **OmniClass**: mednarodni standard (razvil ga je CSI v ZDA) za upravljanje informacij o predmetih od knjižnic komponent do elektronskih specifikacij. Primeren za dolgoročno shranjevanje podatkov, združljiv s CAD (BIM) in drugimi digitalnimi orodji.
- **COBie**: Izmenjava informacij o gradnji in obratovanju stavb je mednarodni standard za izmenjavo podatkov med fazami projektiranja, gradnje in obratovanja. Vključen v BS 1192-4:2014 kot del koncepta "BIM -model, pripravljen za uporabo". Osredotoča se na prenos informacij (npr. specifikacije opreme, garancije, stiki z izvajalci).

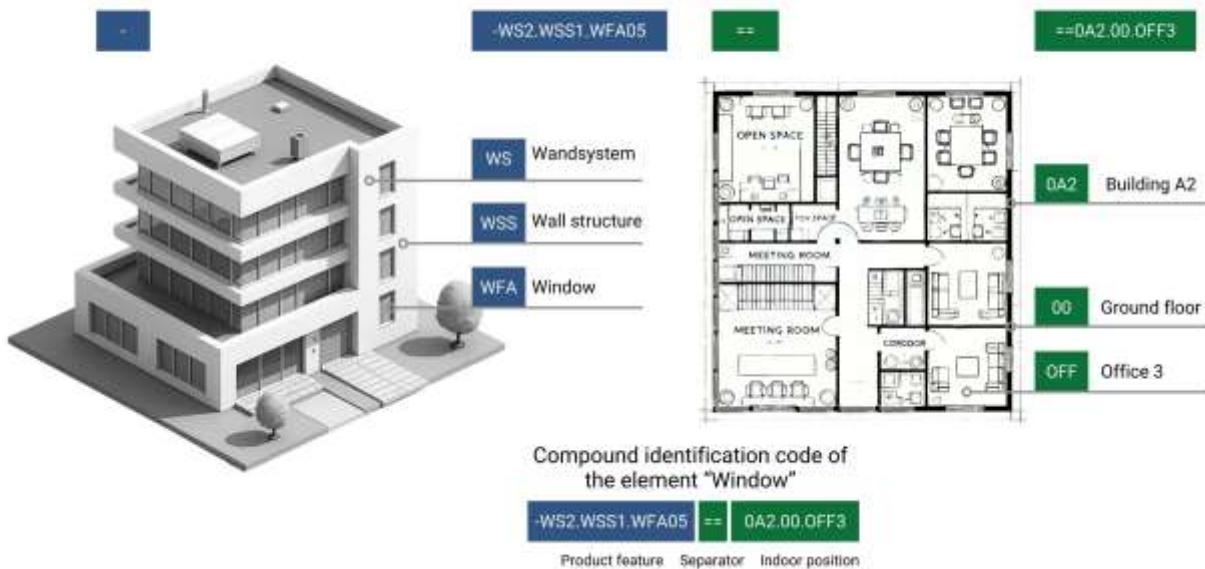
Globalizacija gradbene industrije bo verjetno privedla do postopnega poenotenja sistemov klasifikacije gradbenih elementov, kar bo znatno zmanjšalo odvisnost od lokalnih nacionalnih standardov. Ta proces se lahko razvije po vzoru razvoja internetnih komunikacij, kjer so univerzalni protokoli za prenos podatkov sčasoma nadomestili različne lokalne formate in zagotovili globalno interoperabilnost sistemov.

Alternativna razvojna pot je lahko neposreden prehod na sisteme samodejnega razvrščanja, ki temeljijo na tehnologijah strojnega učenja. Te tehnologije, ki se danes razvijajo predvsem na področju avtonomnega prometa, imajo velik potencial za uporabo v velikih podatkovnih nizih CAD načrtovanja (slika 10.1-6).

Danes razmere niso omejene le na nacionalno združevanje klasifikatorjev. Zaradi številnih posebnosti, ki se na nacionalni ravni ne upoštevajo, mora vsako podjetje poenotiti in standardizirati kategorije elementov in virov, s katerimi dela.

Praviloma se ta proces začne z majhnim obsegom - z lokalnimi tabelami predmetov ali notranjimi sistemi označevanja. Vendar je strateški cilj prehod na skupni jezik za opisovanje vseh elementov, ki bi bil razumljiv ne le v podjetju, temveč tudi zunaj njega - v idealnem primeru usklajen z mednarodnimi ali industrijskimi klasifikatorji (slika 4.2-8). Ta pristop olajša povezovanje z zunanjimi partnerji, digitalnimi sistemi in spodbuja oblikovanje enotnih celovitih procesov v življenjskem ciklu predmetov.

Pred prehodom na avtomatizacijo in razširljive sisteme IT je treba uporabiti klasifikatorje na nacionalni ravni ali zgraditi lastno, logično in nedvoumno strukturo identifikacije elementov. Vsak objekt - naj bo to okno (slika 4.2-11), vrata ali inženirski sistem - mora biti opisan tako, da ga je mogoče nedvoumno prepoznati v katerem koli digitalnem sistemu podjetja. To je ključnega pomena pri prehodu s ploščatih risb na digitalne modele, ki zajemajo tako fazo projektiranja kot tudi obratovanje stavbe.



**Slika 4.2-11 Primer identifikatorja sestavljenega okenskega elementa stavbe na podlagi klasifikacije in položaja v stavbi.**

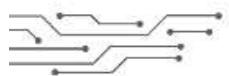
Eden od primerov notranjih klasifikatorjev je lahko razvoj sestavljene identifikacijske kode (slika 4.2-11). Takšna koda združuje več ravni informacij: funkcionalni namen elementa (npr. "okno v steni"), njegovo vrsto in natančno prostorsko referenco - stavba A2, nadstropje 0, soba 3. Takšna večnivojska struktura omogoča vzpostavitev enotnega sistema navigacije po digitalnih modelih in dokumentaciji, zlasti v fazah preverjanja in preoblikovanja podatkov, kjer je potrebno nedvoumno združevanje elementov v skupine. Nedvoumno prepoznavanje elementov zagotavlja skladnost med oddelki ter zmanjšuje tveganje podvajanja, napak in izgube informacij.

Dobro izdelan klasifikator ni le tehnični dokument, temveč temelj digitalnega ekosistema podjetja:

- zagotavlja združljivost podatkov med sistemi;
- zmanjša stroške iskanja in obdelave informacij;
- poveča preglednost in upravljivost;
- ustvarja temelje za razširjanje in avtomatizacijo.

Standardiziran opis predmetov z uporabo nacionalnih klasifikatorjev ali lastnih sestavljenih identifikacijskih kod postane podlaga za dosledne podatke, zanesljivo izmenjavo informacij in poznejše izvajanje inteligentnih storitev - od avtomatiziranega javnega naročanja do digitalnih dvojčkov.

Po končani fazi strukturiranja podatkov več oblik in izbiri klasifikatorja, ki se bo uporabil za prepoznavanje značilnosti in razvrščanje v skupine, je naslednji korak pravilno modeliranje podatkov. Ta postopek vključuje določitev ključnih parametrov, oblikovanje logične strukture podatkov in opis



odnosov med elementi.

## POGLAVJE 4.3.

### MODELIRANJE PODATKOV IN CENTER ODLIČNOSTI

#### Modeliranje podatkov: konceptualni, logični in fizični model

Učinkovito upravljanje podatkov (ki smo jih prej strukturirali in kategorizirali) ni mogoče brez dobro premišljene strukture za shranjevanje in obdelavo. Da bi zagotovila dostop in doslednost podatkov v fazi shranjevanja in obdelave, podjetja uporabljajo modeliranje podatkov, metodologijo, ki jim omogoča oblikovanje tabel, podatkovnih zbirk in povezav med njimi v skladu s poslovnimi zahtevami.

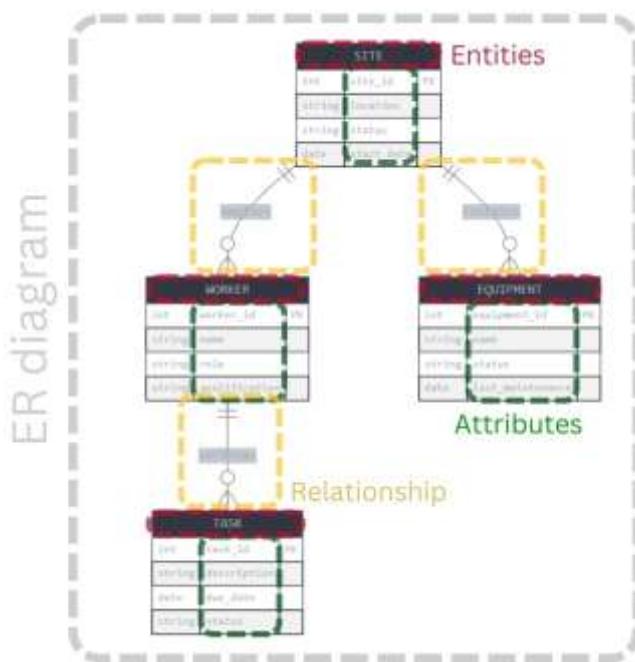
Modeliranje podatkov je temelj, na katerem je zgrajen vsak digitalni ekosistem. Brez opisa sistemov, zahtev in podatkovnega modeliranja inženirji in strokovnjaki, ki ustvarjajo podatke, ne vedo ali ne razumejo, kje bodo uporabljeni podatki, ki jih ustvarjajo.

Podobno kot pri gradnji stavbe, kjer ne morete začeti zlagati opeke brez načrta, je tudi pri ustvarjanju sistema podatkovnega skladišča treba jasno razumeti, kateri podatki se bodo uporabljali, kako bodo povezani in kdo bo z njimi delal. Brez opisa procesov in zahtev inženirji in strokovnjaki, ki ustvarjajo podatke, izgubijo pregled nad tem, kje in kako bodo podatki uporabljeni v prihodnosti.

Podatkovni model služi kot most med poslovanjem in IT. Omogoča formalizacijo zahtev, strukturiranje informacij in lažjo komunikacijo med zainteresiranimi stranmi. V tem smislu je podatkovno modeliranje podobno delu arhitekta, ki v skladu z naročnikovim načrtom razvije načrt stavbe in ga nato preda gradbenikom - skrbnikom in razvijalcem podatkovnih zbirk - v izvedbo (vzpostavitev podatkovne zbirke).

Tako mora vsako gradbeno podjetje poleg strukturiranja in kategoriziranja elementov in virov (slika 4.2-11) obvladati tudi umetnost "gradnje" podatkovnih zbirk (tabel) in se naučiti ustvarjati povezave med njimi, kot da bi opeke povezali v zanesljiv in močan zid znanja iz podatkov podjetja. Ključni koncepti pri modeliranju podatkov (slika 4.3-1) vključujejo:

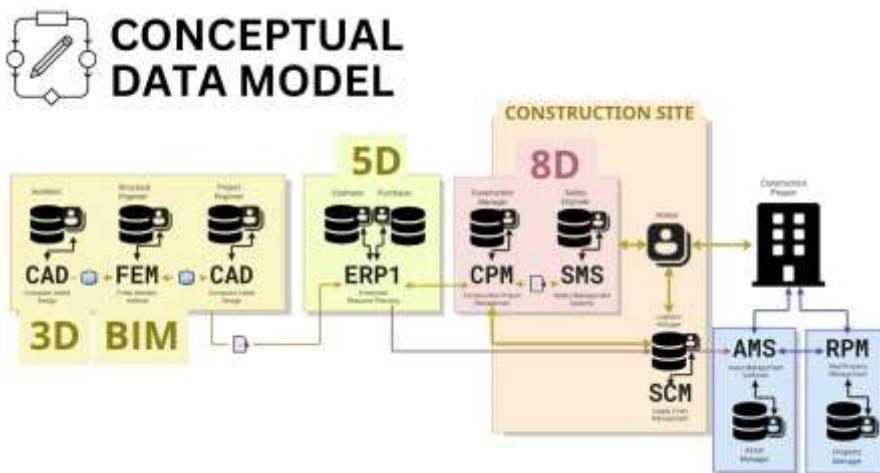
- **Entitete** so predmeti, o katerih je treba zbirati podatke. V zgodnji fazi načrtovanja je entiteta lahko posamezen element (npr. "vrata"), v modelu ocenjevanja pa je lahko skupina kategoriziranih elementov (npr. "notranja vrata").
- **Atributi** so značilnosti entitet, ki opisujejo pomembne podrobnosti: dimenzijske, lastnosti, stroške sestavljanja, logistiko in druge parametre.
- **Odnosi (povezave)** - prikazujejo, kako entitete vplivajo druga na drugo. Lahko so ene od vrst: "ena z eno", "veliko z eno", "veliko z veliko".
- Diagrami **ER** (Entity-Relationship diagrams) so vizualni diagrami, ki prikazujejo entitete, attribute in odnose med njimi. Diagrami ER so lahko konceptualni, logični in fizični - vsak od njih odraža različno raven podrobnosti.



Slika 4.3-1 ER diagram konceptualne strukture podatkovne zbirke z entitetami, atributmi in odnosi.

Postopek oblikovanja podatkov in določanja odnosov med njimi se tradicionalno deli na tri glavne modele. Vsak od njih izpoljuje določene funkcije, razlikujejo pa se po stopnji podrobnosti in stopnji abstrakcije pri predstavitvi podatkovne strukture:

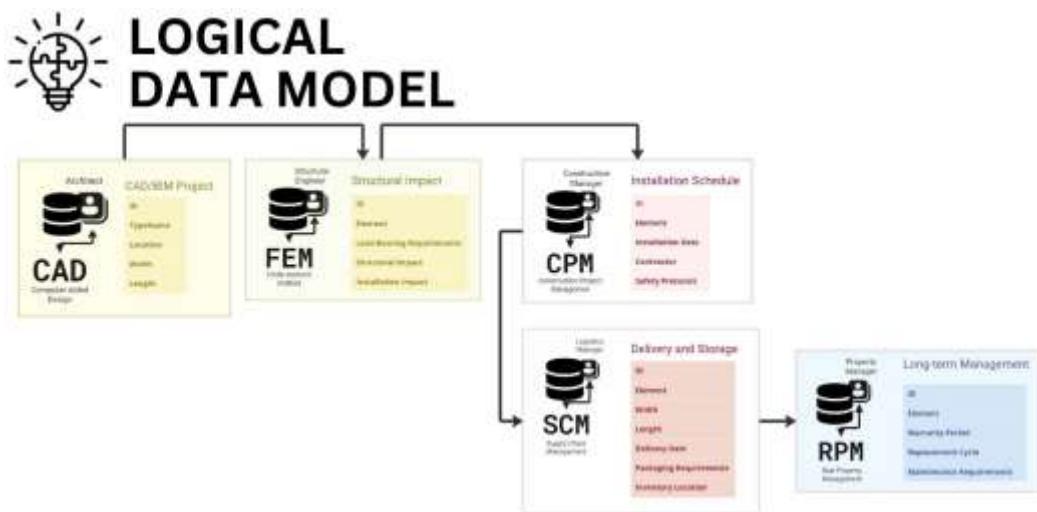
- **Konceptualni podatkovni model:** ta model opisuje glavne entitete in njihova razmerja, ne da bi se spuščal v podrobnosti atributov. Običajno se uporablja v začetnih fazah načrtovanja. V tej fazi lahko iz podatkovnih zbirk in sistemov skiciramo, da prikažemo odnose med različnimi oddelki in strokovnjaki.



Slika 4.3-2 Konceptualni diagram opisuje vsebino sistema: predstavitev odnosov na visoki ravni, brez tehničnih podrobnosti.

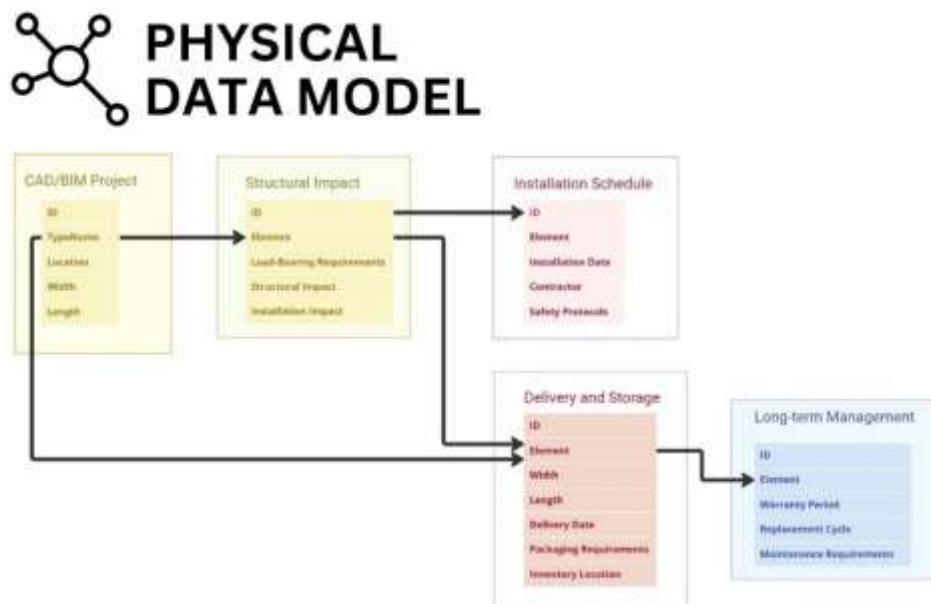
- **Logični podatkovni model:** Logični podatkovni model na podlagi konceptualnega modela

vključuje podrobne opise entitet, atributov, ključev in povezav ter preslikavo poslovnih informacij in pravil.



Slika 4.3-3 Logični podatkovni model podrobno opisuje podatkovne vrste, razmerja in ključe, vendar brez sistemskih izvedb.

- **Fizični podatkovni model:** Ta model opisuje potrebne strukture za izvajanje podatkovne zbirke, vključno s tabelami, stolpci in razmerji. Osredotoča se na zmogljivost podatkovne zbirke, strategije indeksiranja in fizično shranjevanje za optimizacijo fizične namestitve podatkovnih zbirk.

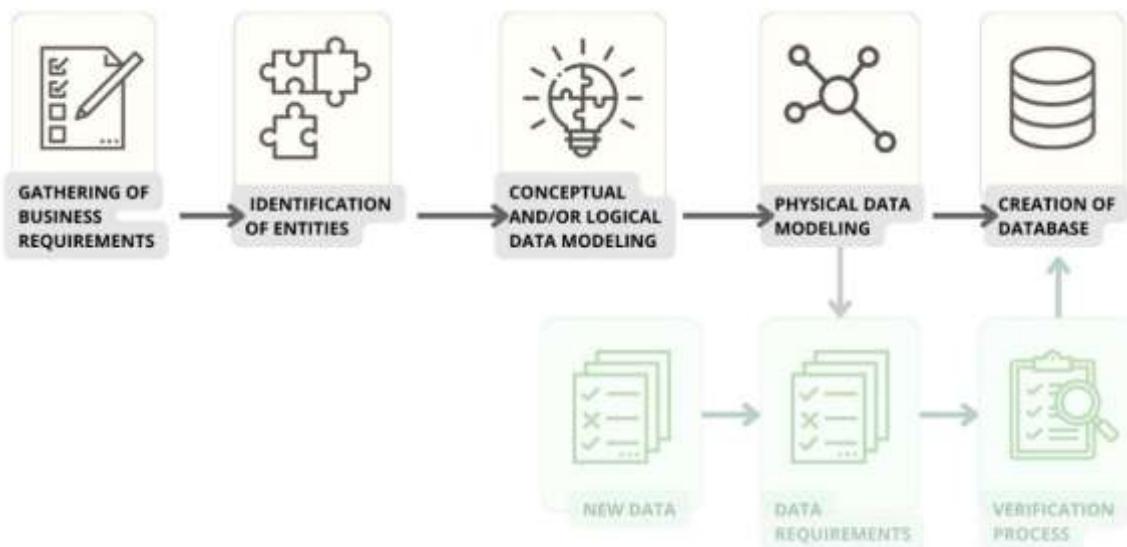


Slika 4.3-4 Fizični podatkovni model določa, kako bo sistem izveden, vključno s tabelami in posebnimi podrobnostmi podatkovne zbirke.

Pri načrtovanju podatkovnih zbirk in tabelaričnih povezav ima razumevanje ravni abstrakcije ključno vlogo pri oblikovanju učinkovite arhitekture sistema.

Učinkovita metodologija podatkovnega modeliranja vam omogoča, da poslovne cilje združite s tehnično izvedbo, zaradi česar je celotna veriga postopkov preglednejša in lažje obvladljiva. Modeliranje podatkov ni enkratna naloga, temveč proces, ki vključuje zaporedne korake (slika 4.3-5):

- **Zbiranje poslovnih zahtev:** opredelijo se ključne naloge, cilji in informacijski tokovi. To je faza aktivnega sodelovanja s strokovnjaki in uporabniki.
- **Identifikacija entitet:** poudarjeni so glavni predmeti, kategorije in vrste podatkov, ki jih je treba upoštevati v prihodnjem sistemu.
- **Razvoj konceptualnega in logičnega modela:** najprej se zajamejo ključne entitete in njihovi odnosi, nato atributi, pravila in podrobna struktura.
- **Fizično modeliranje:** zasnovana je tehnična izvedba modela: tabele, polja, povezave, omejitve, indeksi.
- **Ustvarjanje podatkovne zbirke:** zadnji korak je implementacija fizičnega modela v izbrani DBMS, testiranje in priprava na delovanje.



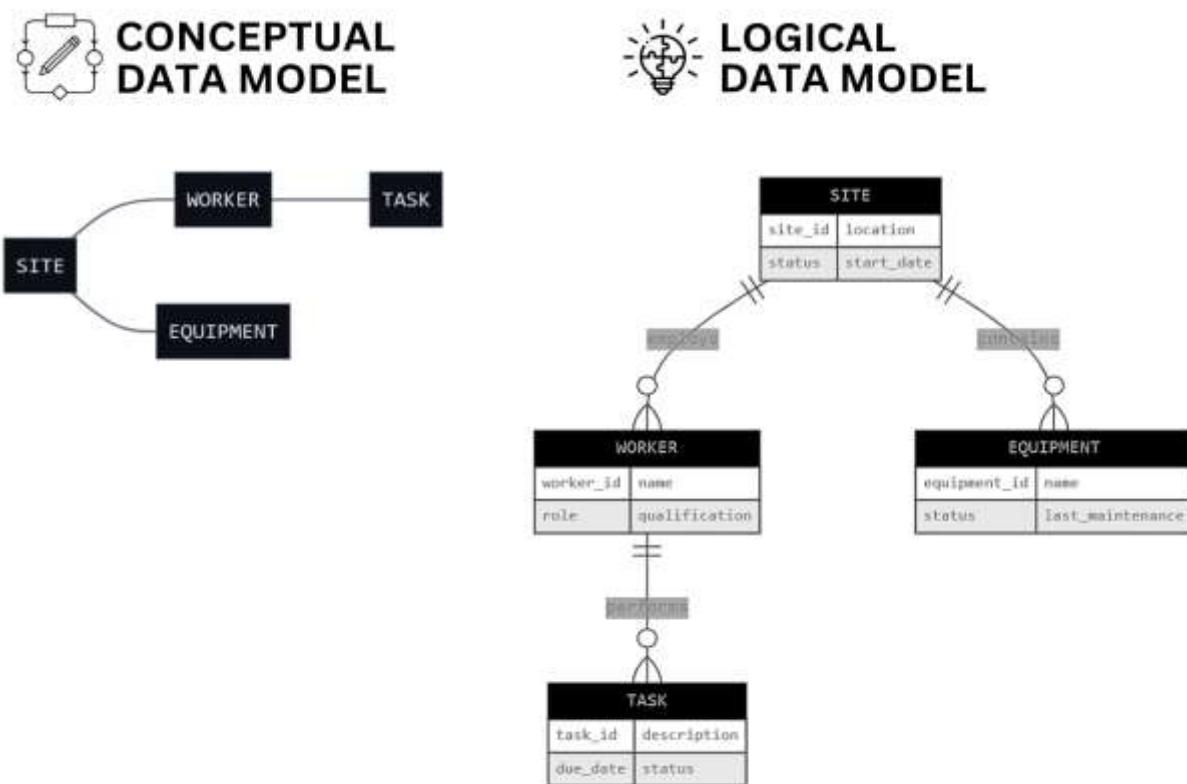
Slika 4.3-5 Ustvarjanje podatkovnih zbirk in sistemov za upravljanje podatkov za poslovne procese se začne z oblikovanjem zahtev in modeliranjem podatkov.

Ustrezno zasnovani postopki modeliranja podatkov omogočajo pregleden pretok informacij, kar je še posebej pomembno pri kompleksnih projektih, kot je upravljanje gradbenih projektov ali gradbišč. Poglejmo, kako lahko prehod od konceptualnega modela k logičnemu modelu in nato k fizičnemu modelu pripomore k racionalizaciji procesov.

## Praktično modeliranje podatkov v gradbeništvu

Kot primer modeliranja podatkov vzemimo nalogu upravljanja gradbišča in pretvorimo zahteve mojstra v strukturiran logični model. Na podlagi osnovnih potreb upravljanja gradbišča opredelimo ključne entitete za: gradbišče (SITE), delavce (WORKER), opremo (EQUIPMENT), naloge (TASK) in uporabo opreme (EQUIPMENT\_USAGE). Vsaka entiteta vsebuje niz atributov, ki odražajo pomembne značilnosti. Na primer, pri nalogi TASK so to lahko opis naloge, datum izvedbe, status, prednostna naloga; pri delavcu WORKER (Delavec) so to lahko ime, njegova vloga na lokaciji, trenutna zaposlitev itd.

Logični model vzpostavlja odnose med temi entitetami in prikazuje, kako medsebojno delujejo v dejanskih delovnih procesih (slika 4.3-6). Na primer, razmerje med lokacijo in delavci kaže, da lahko na eni lokaciji dela več delavcev, medtem ko razmerje med delavci in nalogami odraža, da lahko en delavec opravlja več nalog.

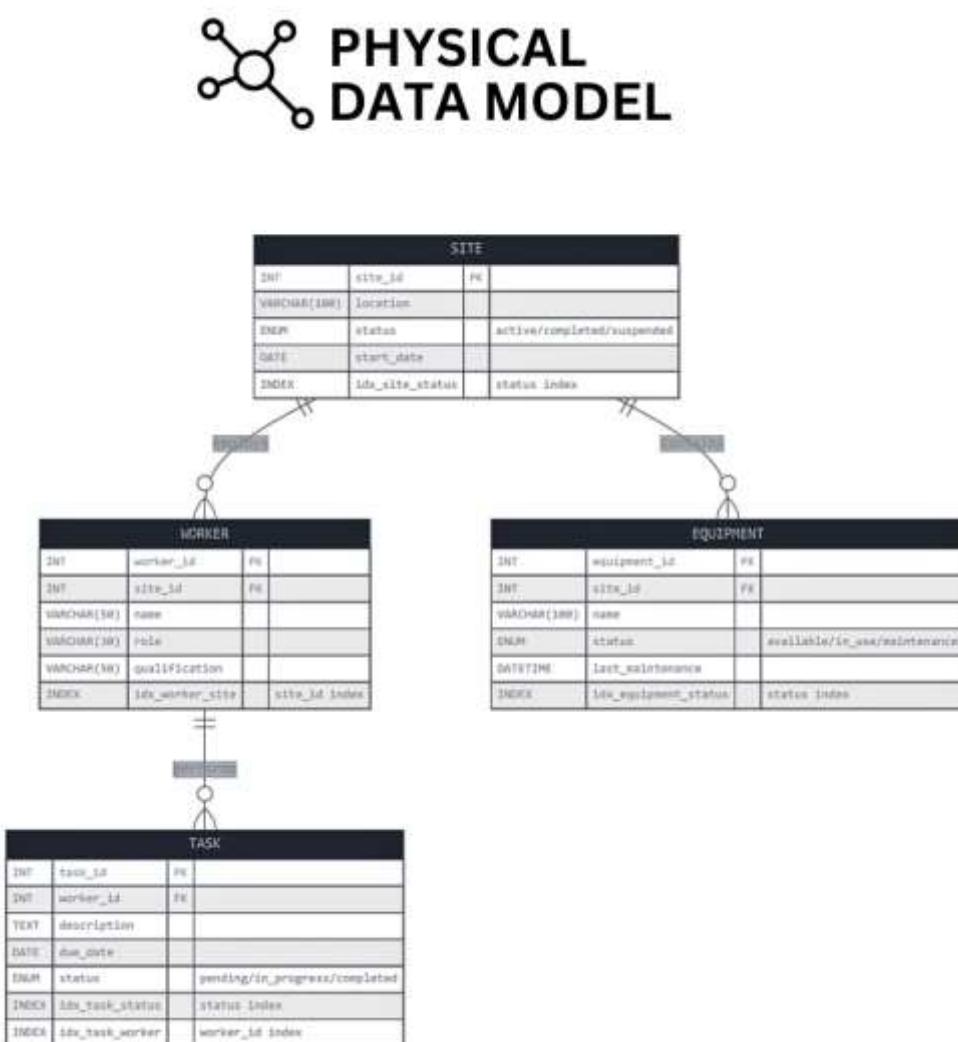


Slika 4.3-6 Konceptualni in logični podatkovni model, ustvarjen na podlagi zahtev mojstra za opis procesov na gradbišču.

Pri prehodu na fizični model se dodajo tehnične izvedbene podrobnosti: posebne vrste podatkov (VARCHAR, INT, DATE), primarni in tuji ključi za razmerja med tabelami ter indeksi za optimizacijo delovanja podatkovne zbirke (slika 4.3-7).

Na primer, za statuse je treba opределiti posebne vrste z možnimi vrednostmi, za izboljšanje učinkovitosti iskanja pa je treba dodati indekse za ključna polja, kot sta status in worker\_id. S tem se logični opis sistema spremeni v konkreten izvedbeni načrt podatkovne zbirke, ki je pripravljen za

ustvarjanje in izvajanje.



Slika 4.3-7 Fizični podatkovni model opisuje entitete gradbišča z minimalnimi zahtevanimi parametri.

Fizični model se pogosto razlikuje od logičnega modela. V povprečju je porazdelitev časa modeliranja naslednja: približno 50% se porabi za konceptualni model (zbiranje zahtev, razprava o procesih, določanje entitet), 10% za logični model (določanje atributov in odnosov) in 40% za fizični model (izvajanje, testiranje, prilagajanje DBMS).

To ravnoesje je posledica dejstva, da konceptualna faza postavlja temelje podatkovne strukture, medtem ko logični model določa le razmerja in atribute. Fizični model zahteva največ virov, saj se v tej fazi podatki implementirajo v posebne platforme in orodja

## Ustvarjanje podatkovne zbirke s programom LLM

Ko imamo podatkovni model in opis entitet s parametri, smo pripravljeni na ustvarjanje podatkovnih baz - shramb, v katerih bomo shranjevali informacije, ki bodo prišle po fazi strukturiranja o določenih procesih.

Poskusimo ustvariti primer preproste, a funkcionalne podatkovne zbirke z minimalno količino kode z uporabo SQLite na primeru programskega jezika Python. Relacijske podatkovne zbirke so bile podrobno obravnavane v poglavju "Strukturirane relacijske podatkovne zbirke in poizvedovalni jezik SQL".

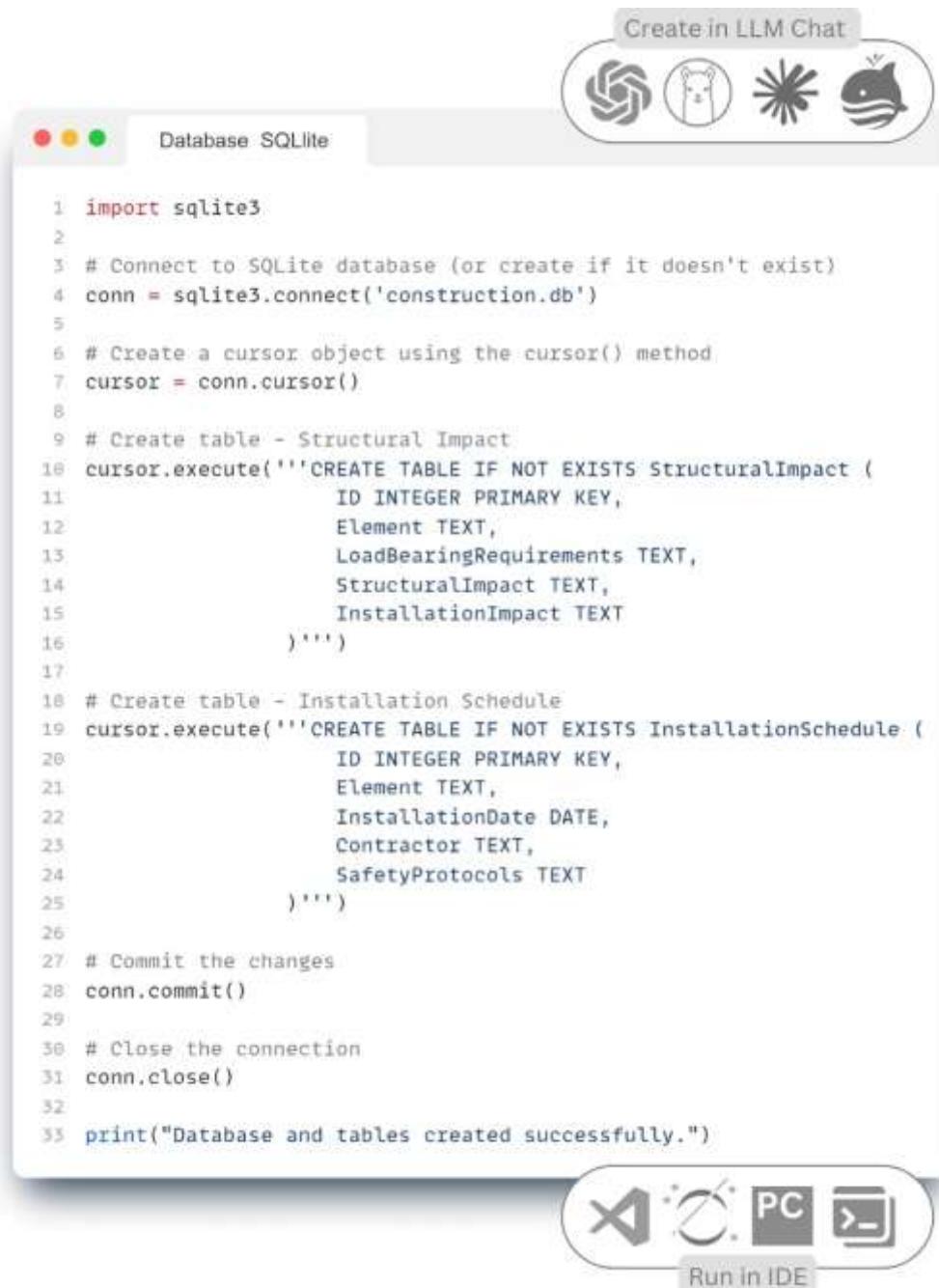
SQLite je lahka, odprtokodna, vgradljiva, medplatformna podatkovna zbirka, ki podpira osnovne funkcije SQL. SQLite se pogosto uporablja v CAD (BIM), ERP, EPM in je prisoten v skoraj vsakem pametnem telefonu.

Prosimo model LLM, da nam na podlagi zgoraj ustvarjenega fizičnega modela (slika 4.3-4) za sistema FEM (strukturni vpliv) in CPM (urnik sestavljanja) ustvari majhno podatkovno zbirko SQLlite.

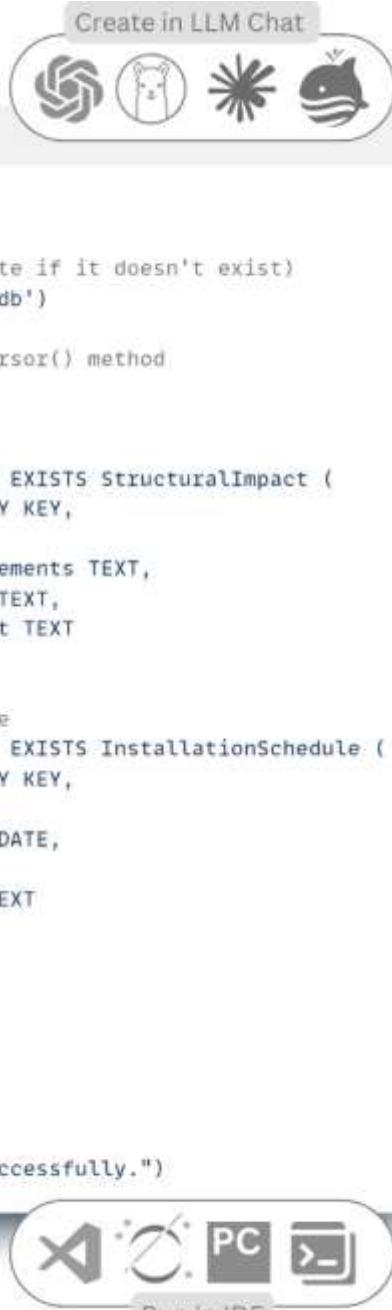
- ❷ Pošljite besedilno zahtevo v klepet LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ali katero koli drugo):

Ustvarite podatkovni zbirki SQLlite za sistema "Constructive Impact" in "Installation Schedule" s slike v priponki (prilepite shemo v klepet kot sliko (slika 4.3-4) ↴

2 Odgovor LLM :



```
1 import sqlite3
2
3 # Connect to SQLite database (or create if it doesn't exist)
4 conn = sqlite3.connect('construction.db')
5
6 # Create a cursor object using the cursor() method
7 cursor = conn.cursor()
8
9 # Create table - Structural Impact
10 cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
11                 ID INTEGER PRIMARY KEY,
12                 Element TEXT,
13                 LoadBearingRequirements TEXT,
14                 StructuralImpact TEXT,
15                 InstallationImpact TEXT
16             )''' )
17
18 # Create table - Installation Schedule
19 cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
20                 ID INTEGER PRIMARY KEY,
21                 Element TEXT,
22                 InstallationDate DATE,
23                 Contractor TEXT,
24                 SafetyProtocols TEXT
25             )''' )
26
27 # Commit the changes
28 conn.commit()
29
30 # Close the connection
31 conn.close()
32
33 print("Database and tables created successfully.")
```



Slika 4.3-8 Ustvarjena koda iz naložene slike ustvari podatkovne zbirke in tabele, v katere lahko zdaj naložimo podatke.

To kodo zbirke podatkov lahko zaženete v okolju Python s podporo za SQLite, ki je običajno vključena v standardno okolje Python. Zagnati jo je mogoče na katerem koli lokalnem računalniku s sistemom Python, ki lahko zažene skripte in ustvari datoteke. Po zagonu in izvedbi bo ta koda na trdem disku ustvarila datoteko - zbirko podatkov SQLite z imenom construction.db (podobno kot datoteka Excel z delovnimi listi), ki bo vsebovala preglednici "Structural Impact" (Strukturni vpliv) in "Installation Schedule" (Načrt namestitve).

Ko je v teh tabelah ustvarjena nova podatkovna zbirka **construction.db**, lahko podatke v te tabele dodajate prek poizvedb SQL \_queries ali uvoza, kar vam bo omogočilo, da kasneje začnete ustvarjati samodejno obdelavo podatkov. Podatke lahko v podatkovno zbirko SQLite uvozite iz datotek CSV, preglednic Excel ali izvozite iz drugih podatkovnih zbirk in skladišč prek vmesnika API.

Za vzpostavitev trajnostnega modeliranja podatkov in učinkovitih postopkov upravljanja podatkovnih zbirk potrebuje podjetje jasno opredeljeno strategijo ter usklajevanje med tehničnimi in poslovnimi ekipami. Pri različnih projektih in več virih podatkov je pogosto težko zagotoviti skladnost, standardizacijo in nadzor kakovosti na vseh ravneh. Ena od ključnih rešitev je lahko ustanovitev posebnega centra odličnosti za modeliranje podatkov (CoE) v podjetju.

## Center odličnosti (CoE) za modeliranje podatkov

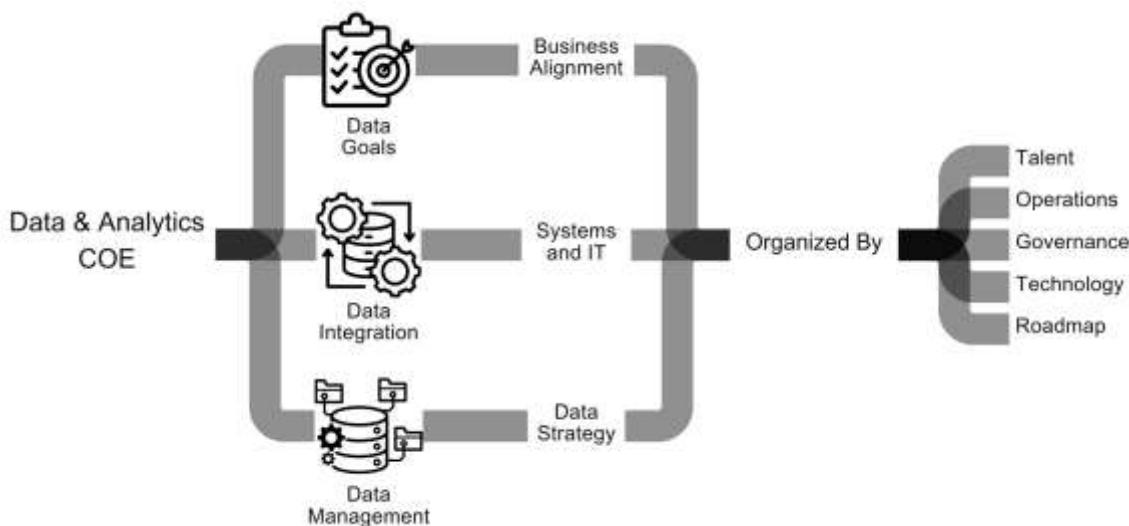
Ker podatki postajajo eno ključnih strateških sredstev, morajo podjetja storiti več kot le pravilno zbirati in shranjevati informacije - pomembno je, da se naučijo sistematicno upravljati podatke. Center odličnosti za klasifikacijo in modeliranje podatkov (CoE) je strukturna enota, ki zagotavlja doslednost, kakovost in učinkovitost vseh postopkov ravnjanja s podatki v organizaciji.

Center odličnosti (CoE) je jedro strokovne podpore in metodološka podlaga za digitalno preobrazbo v podjetju. Gradi kulturo, ki temelji na podatkih, in organizacijam omogoča, da vzpostavijo procese, ki sprejemajo odločitve na podlagi strukturiranih, potrjenih in reprezentativnih podatkov in ne na podlagi intuicije ali lokalnih informacij.

Podatkovni center odličnosti je običajno sestavljen iz medfunkcijskih skupin, ki delujejo po načelu "dveh pic". To načelo, ki ga je predlagal Jeff Bezos, pomeni, da mora biti ekipa tako velika, da se lahko nasiti z dvema picama, tj. ne več kot 6-10 ljudi. Ta pristop pomaga preprečevati pretirano birokracijo in povečuje prožnost dela. Ekipa CoE bi morala vključevati zaposlene z različnimi tehničnimi znanji, od podatkovne analitike in strojnega učenja do strokovnega znanja na posameznih poslovnih področjih. Podatkovni inženirji naj s svojim poglobljenim tehničnim znanjem ne bi le optimizirali procesov in modelirali podatkov, temveč tudi podpirali sodelavce s skrajševanjem časa za rutinske naloge (slika 4.3-9).

Tako kot v naravi odpornost ekosistemov zagotavlja biotska raznovrstnost, se v digitalnem svetu prožnost in prilagodljivost dosegata z različnimi pristopi k ravnanju s podatki. Vendar morajo to raznolikost podpirati skupna pravila in koncepti.

Center odličnosti (CoE) lahko primerjamo s "podnebnimi razmerami" gozdnega ekosistema, ki določajo, katere vrste podatkov bodo uspevale in katere bodo samodejno zavrhene. Z ustvarjanjem ugodne "klime" za kakovostne podatke center odličnosti omogoča naravno izbiro najboljših praks in metodologij, ki nato postanejo standardi za organizacijo.



**Slika 4.3-9 Center odličnosti (CoE) za podatke in analitiko združuje strokovno znanje o ključnih vidikih upravljanja, integracije in strategije podatkov.**

Da bi pospešili cikle vključevanja in dosegli boljše rezultate, bi moral CoE svojim članom zagotoviti zadostno stopnjo avtonomije pri sprejemanju odločitev. To je še posebej pomembno v dinamičnem okolju, kjer lahko poskusi in napake, stalne povratne informacije in pogoste izdaje prinesajo pomembne koristi. Vendar je ta avtonomija učinkovita le, če je zagotovljena jasna komunikacija in podpora višjega vodstva. Brez strateške vizije in usklajevanja na najvišji ravni se lahko tudi najbolj kompetentna ekipa sreča z ovirami pri izvajanju svojih pobud.

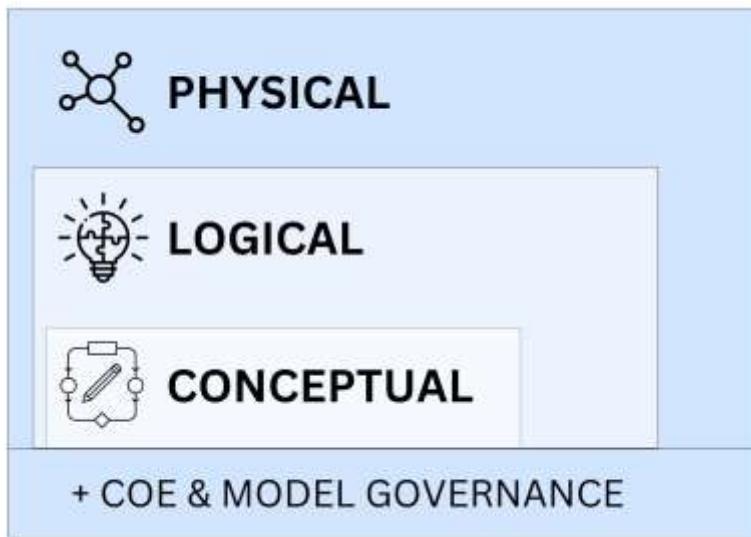
COE ali višje vodstvo podjetja je odgovorno za to, da pristop k modeliranju podatkov ni omejen na enega ali dva projekta, temveč je vključen v celoten sistem upravljanja informacij in poslovnih procesov.

Strokovni center (CoE) je poleg nalog, povezanih z modeliranjem in upravljanjem podatkov, odgovoren za razvoj skupnih standardov in pristopov k uvajanju in delovanju podatkovne infrastrukture. Poleg tega spodbuja kulturo stalnih izboljšav, optimizacije procesov in učinkovite uporabe podatkov v organizaciji (slika 4.3-10).

Sistematični pristop k upravljanju podatkov in modelov v okviru CoE lahko v grobem razdelimo na več ključnih sklopov:

- **Standardizacija procesov in upravljanje življenjskega cikla modela:** CoE razvija in izvaja metodologije za poenotenje ustvarjanja in upravljanja podatkovnih modelov. To vključuje: vzpostavitev strukturnih predlog, metod za nadzor kakovosti in sistemov za nadzor različic, da se zagotovi neprekinjenost podatkov v vseh fazah dela.

- **Upravljanje vlog in dodeljevanje odgovornosti:** CoE opredeli ključne vloge v postopku modeliranja podatkov. Vsakemu udeležencu projekta so dodeljene jasno opredeljene vloge in področja odgovornosti, kar olajša timsko delo in zmanjša tveganje za neskladnost podatkov.
- **Nadzor kakovosti in revizija:** učinkovito upravljanje podatkov o gradnji zahteva stalno spremljanje njihove kakovosti. Izvajajo se avtomatizirani mehanizmi za preverjanje podatkov, ugotavljanje napak in manjkajočih atributov.
- **Upravljanje metapodatkov in informacijske arhitekture:** CoE je odgovoren za oblikovanje enotnega sistema klasifikacije in identifikatorjev ter standardov za poimenovanje in opis entitet, ki je ključnega pomena za integracijo med sistemi.



Slika 4.3-10 Modeliranje podatkov in upravljanje kakovosti podatkov je eden glavnih izzivov CoE

Center odličnosti (CoE) za podatke ni le skupina strokovnjakov, temveč sistemski mehanizem, ki ustvarja novo kulturo, ki temelji na podatkih, in zagotavlja enoten pristop k delu s podatki v celotnem podjetju. S kompetentnim vključevanjem procesov modeliranja v celoten sistem upravljanja informacij, standardizacijo, klasifikacijo in nadzorom kakovosti podatkov, CoE pomaga podjetjem nenehno izboljševati njihove izdelke in poslovne procese, se hitreje odzivati na spremembe na trgu in sprejemati utemeljene odločitve na podlagi zanesljive analitike.

Takšni centri so še posebej učinkoviti v kombinaciji s sodobnimi načeli DataOps - v okviru, ki zagotavljajo neprekinjeno dostavo, avtomatizacijo in nadzor kakovosti podatkov. Več o DataOps bomo govorili v 8. delu, v poglavju "Sodobne podatkovne tehnologije v gradbeništvu".

V naslednjih poglavjih bomo prešli od strategije k praksi - pogojno se "preoblikujmo" v podatkovni center: ogledali si bomo več primerov, kako potekajo parametrizacija nalog, zbiranje zahtev in postopek samodejnega potrjevanja.



## POGLAVJE 4.4.

### SISTEMATIZACIJA ZAHTEV IN POTRJEVANJE INFORMACIJ.

#### Zbiranje in analiziranje zahtev: preoblikovanje komunikacij v strukturirane podatke.

Zbiranje in upravljanje zahtev je prvi korak k zagotavljanju kakovosti podatkov. Kljub razvoju digitalnih orodij se večina zahtev še vedno oblikuje v nestrukturirani obliki: s pismi, zapisniki sestankov, telefonskimi klici in ustnimi razpravami. Takšna oblika komunikacije otežuje avtomatizacijo, potrjevanje in ponovno uporabo informacij. V tem poglavju si ogledamo, kako besedilne zahteve prevesti v formalne strukture in tako zagotoviti preglednost in sistematičnost poslovnih zahtev.

Gartnerjeva raziskava "Data Quality: Best Practices for Accurate Insights" poudarja ključni pomen kakovosti podatkov za uspešne podatkovne in analitične pobude [100]. Ugotavljajo, da slaba kakovost podatkov organizacije v povprečju stane vsaj 12,9 milijona dolarjev letno in da so zanesljivi, visokokakovostni podatki bistveni za oblikovanje podatkovno usmerjenega podjetja.

Pomanjkanje strukturiranih zahtev povzroča, da so lahko isti element (entiteta) in njegovi parametri shranjeni v različnih sistemih v različnih različicah. To ne le zmanjšuje učinkovitost procesov, temveč povzroča tudi izgubo časa, podvajanje informacij in potrebo po ponovnem potrjevanju podatkov, preden jih je mogoče uporabiti. Posledično lahko že ena sama opustitev - izgubljen parameter ali en sam nepravilno opisan element - upočasni sprejemanje odločitev in povzroči neučinkovito rabo virov.

*Zaradi pomanjkanja žebbla je bila podkev izgubljena.*

*Zaradi pomanjkanja podkev je bil konj izgubljen.*

*Ker ni bilo konja, se je izgubil jezdec.*

*Zaradi pomanjkanja jezdeca je bilo sporočilo izgubljeno.*

*Zaradi pomanjkanja sporočila je bila bitka izgubljena.*

*Zaradi pomanjkanja bitke je bilo izgubljeno kraljestvo.*

*Vse to zaradi pomanjkanja žebbla v podkvi.*

- Pregovor [101]

Analiza in zbiranje zahtev za postopek izpolnjevanja in shranjevanja podatkov se začneta z opredelitvijo vseh zainteresiranih strani. Tako kot pregovorna izguba enega samega žebbla vodi v verigo kritičnih posledic, lahko v poslu izguba ene same zainteresirane strani, spregledana zahteva ali izguba celo enega samega parametra bistveno vpliva ne le na posamezen poslovni proces, temveč na celoten ekosistem projekta in organizacijo kot celoto. Zato je ključnega pomena, da prepoznamo tiste elemente, parametre in vloge, ki se na prvi pogled zdijo nepomembni, vendar se lahko pozneje izkažejo za ključne za trajnost poslovanja.

Predstavljajmo si, da ima podjetje projekt, pri katerem stranka postavi novo zahtevo: "Dodajte dodatno

okno na severni strani stavbe". V manjši proces "zahteva stranke za dodajanje novega okna v trenutni projekt" so vključeni arhitekt, stranka, strokovnjak CAD (BIM), vodja gradnje, vodja logistike, analistik ERP, inženir za nadzor kakovosti, varnostni inženir, vodja nadzora in vodja nepremičnin.

Tudi majhen postopek lahko vključuje več deset različnih strokovnjakov. Vsak udeleženec procesa mora razumeti zahteve strokovnjakov, s katerimi je povezan na ravni podatkov.

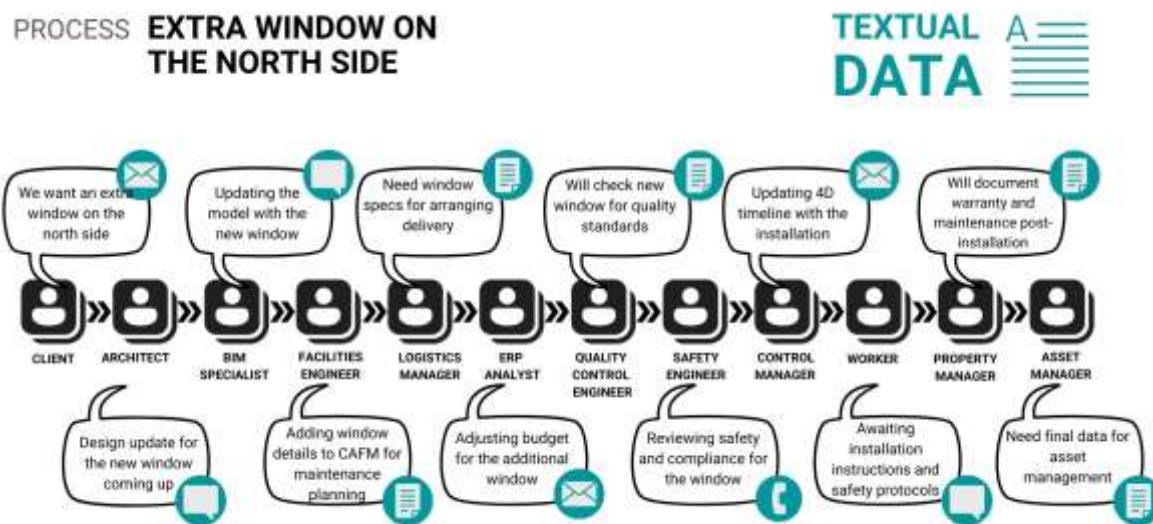
Na ravni besedila (slika 4.4-1) poteka komunikacija med stranko in strokovnjaki v procesni verigi na naslednji način:

- ⦿ **Stranka:** "Odločili smo se, da na severni strani dodamo dodatno okno za boljšo osvetlitev. Ali je to mogoče izvesti?"
- ⦿ **Arhitekt:** "Seveda, spremenil bom projekt, da bo vključeval novo okno, in poslal posodobljene načrte CAD (BIM)".
- ⦿ **Strokovnjak za CAD (BIM):** "Prejel nov projekt. Posodobim model CAD (BIM) z dodatnim oknom in po uskladitvi z inženirjem FEM zagotovim natančno lokacijo in dimenzije novega okna."
- ⦿ **Vodja gradnje:** "Prejeli smo nov projekt. Prilagajamo datume vgradnje 4D in o tem obveščamo vse ustrezne podizvajalce."
- ⦿ **Inženir objektov (CAFM):** "Podatke 6D o novem oknu bom vnesel v sistem CAFM za prihodnje upravljanje objektov in načrtovanje vzdrževanja."
- ⦿ **Vodja logistike:** "Potrebujem dimenzije in težo novega okna, da lahko organiziram dostavo okna na lokacijo."
- ⦿ **ERP -analitik:** "Potrebujem preglednice obsega in natančno vrsto okna za posodobitev proračuna 5D v našem sistemu ERP, da se stroški novega okna odražajo v skupni oceni projekta."
- ⦿ **Inženir za nadzor kakovosti:** "Ko bodo specifikacije oken pripravljene, bom poskrbel, da bodo izpolnjevale naše standarde kakovosti in materialov."
- ⦿ **Varnostni inženir:** "Ocenil bom varnostne vidike novega okna, s posebnim poudarkom na skladnosti in evakuaciji v okviru sheme 8D".
- ⦿ **Vodja nadzora:** "Na podlagi natančnega obsega dela iz ERP bomo posodobili našo časovnico 4D, da bo odražala namestitev novega okna, in shranili nove podatke v sistem za upravljanje vsebine projekta."
- ⦿ **Delavec (monter):** "Potrebujete navodila za namestitev, montažo in časovni razpored dela. Ali so poleg tega uvedena kakšna posebna varnostna pravila, ki jih moram upoštevati?"
- ⦿ **Upravitelj nepremičnine:** "Po namestitvi bom dokumentiral informacije o garanciji in vzdrževanju za dolgoročno upravljanje".
- ⦿ **Upravitelj premoženja:** "Inženir opreme, pošljite končne podatke za sledenje in upravljanje življenjskega cikla sredstev."
- ⦿ **Stranka:** "Počakajte, morda se mi mudi in okno ne bo potrebno. Morda bi moral narediti balkon."

V takšnih scenarijih, ki se pogosto dogajajo, že majhna sprememba povzroči verižno reakcijo med več sistemi in vlogami. V tem primeru je skoraj vsa komunikacija v začetni fazi v besedilni obliki: elektronska sporočila, klepetalnice, zapisniki sestankov (slika 4.4-1).

V takšnem besedilnem komunikacijskem sistemu za gradbeni projekt je zelo pomemben sistem pravne potrditve in evidentiranja vseh postopkov izmenjave podatkov in vseh sprejetih odločitev. S tem

se zagotovi, da je vsaka odločitev, navodilo ali sprememba pravno veljavna in sledljiva, s čimer se zmanjša tveganje prihodnjih "nesporazumov"

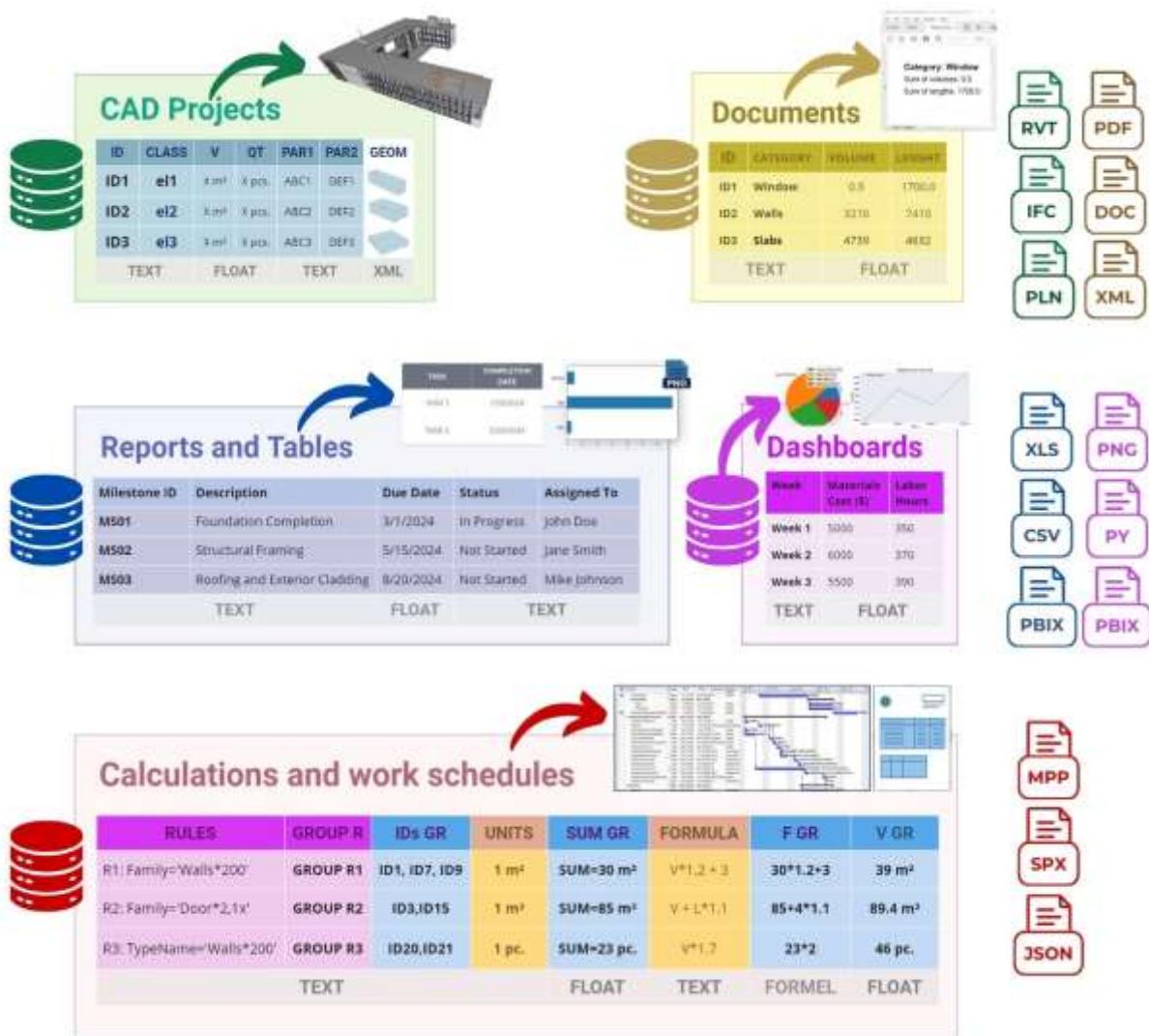


**Slika 4.4-1 Komunikacija med naročnikom in izvajalcem v začetnih fazah projekta pogosto vsebuje besedilne podatke več oblik.**

Pomanjkanje pravnega nadzora in potrjevanja odločitev v ustreznih sistemih gradbenega projekta lahko privede do resnih težav za vse vpletene. Vsaka odločitev, naročilo ali sprememba, sprejeta brez ustrezne dokumentacije in potrditve, lahko privede do sporov (in sodnih postopkov).

Pravno utrditev vseh odločitev v besedilni komunikaciji je mogoče zagotoviti le z velikim številom podpisanih dokumentov, kar bo padlo na ramena uprave, ki mora evidentirati vse transakcije. Če mora vsak udeleženec podpisati dokumente za vsako dejanje, sistem izgubi prožnost in postane birokratski labirint. Pomanjkanje potrdil o transakcijah ne bo samo zavleklo izvajanja projekta, temveč lahko povzroči tudi finančne izgube in poslabšanje odnosov med udeleženci, vključno s pravnimi težavami.

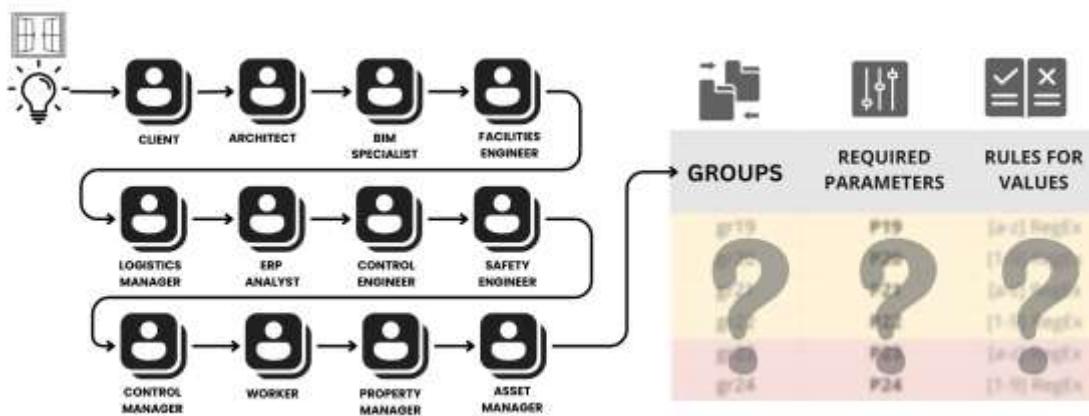
Takšen postopek odobritve transakcije, ki se običajno začne z besedilnimi razpravami, se v naslednjih fazah postopoma razvije v izmenjavo dokumentov več oblik (slika 4.4-2), kar znatno oteži komunikacijo, ki je prej potekala le prek besedila. Brez jasno opredeljenih zahtev je avtomatizacija takšnih procesov, ki so polni večpredstavnostnih podatkov in velikega števila besedilnih zahtev, skoraj nemogoča.



Slika 4.4-2 Vsak sistem v gradbenem podjetju služi kot vir pravno pomembnih dokumentov v različnih oblikah.

Zaradi besedilnih sporočil se mora vsak strokovnjak seznaniti s celotno korespondenco ali pa se redno udeleževati vseh sestankov, da bi razumel trenutno stanje projekta.

Za odpravo te omejitve je potreben prehod od besedilne komunikacije k strukturiranemu modelu zahtev. To je mogoče le s sistematično analizo, vizualizacijo procesov in opisom interakcij v obliki diagramov poteka in podatkovnih modelov (slika 4.4-3). Tako kot pri podatkovnem modeliranju (slika 4.3-7) smo se s kontekstualno-idejne ravni premaknili na konceptualno raven z dodajanjem sistemov in orodij, ki jih uporabljajo udeleženci, ter povezav med njimi.



Slika 4.4-3 Da bi se naučili, kako upravljati in avtomatizirati postopek potrjevanja, je treba vizualizirati procese in strukturirati zahteve.

Prvi korak pri sistematizaciji zahtev in odnosov je vizualizacija vseh povezav in odnosov z uporabo konceptualnih diagramov poteka. Konceptualna raven ne bo le olajšala razumevanja celotne procesne verige vsem udeležencem procesa, temveč bo tudi jasno pokazala, zakaj in za koga so podatki (in zahteve) potrebni v vsakem koraku procesa.

### Procesne sheme in učinkovitost konceptualnih okvirov

Da bi premostili vrzel med tradicionalnimi in sodobnimi pristopi k upravljanju podatkov, morajo podjetja zavestno preiti od razdrobljenih besedilnih opisov k strukturiranim predstavitvam procesov. Razvoj podatkov - od glinenih tablic do digitalnih ekosistemov - zahteva nova orodja za razmišljanje. Eno od teh orodij je konceptualno modeliranje z uporabo diagramov poteka. Ustvarjanje vizualnih diagramov - diagramov poteka, diagramov procesov, diagramov interakcij - omogoča udeležencem projekta, da se zavedajo, kako njihova dejanja in odločitve vplivajo na celoten sistem odločanja.

Če procesi ne zahtevajo le shranjevanja podatkov, temveč tudi njihovo analizo ali avtomatizacijo, se morate začeti ukvarjati z ustvarjanjem konceptualnega in vizualnega sloja zahtev.

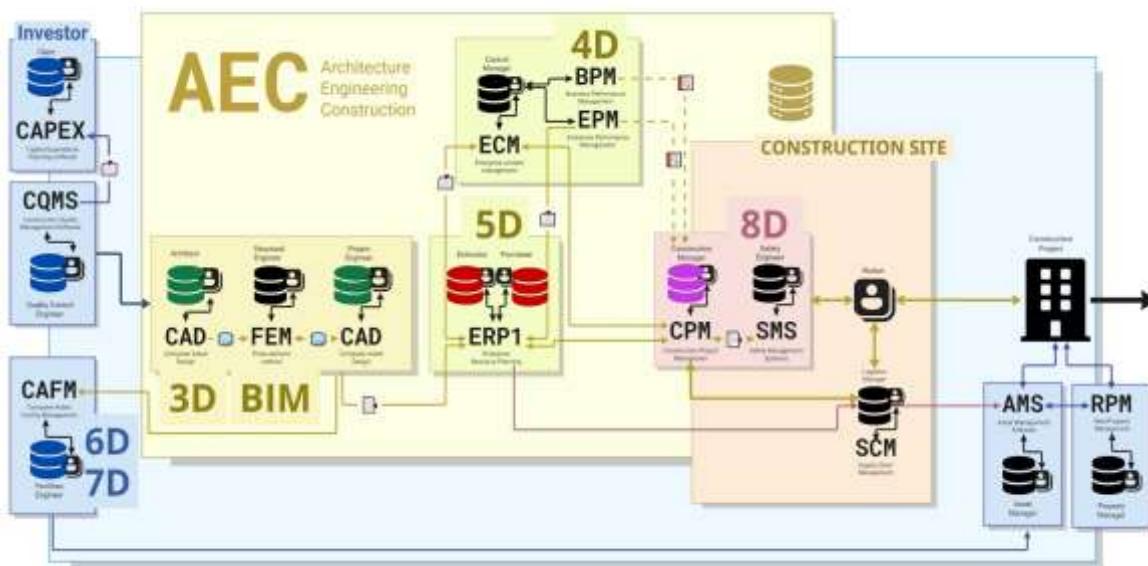
V našem primeru (slika 4.4-1) je lahko vsak strokovnjak del majhne skupine, lahko pa tudi del večjega oddelka z do ducat strokovnjaki pod nadzorom glavnega vodje. Vsak oddelek uporablja specializirano aplikacijsko podatkovno zbirko (slika 1.2-4, npr. ERP, CAD, MEP, CDE, ECM, CPM itd.), ki se redno posodablja z vhodnimi informacijami, potrebnimi za ustvarjanje dokumentov, beleženje pravnega statusa odločitev in upravljanje procesov.

Postopek transakcije je podoben delu antičnih menedžerjev pred 4000 leti, ko so za pravno potrditev odločitev uporabljali glinene tablice in papirus. Razlika med sodobnimi sistemmi in njihovimi glinenimi in papirnatimi predhodniki je v tem, da sodobne metode dodatno vključujejo postopek pretvorbe besedilnih informacij v digitalno obliko za nadaljnjo samodejno obdelavo v drugih sistemih in orodjih.

Oblikovanje vizualizacije procesa v obliki konceptualnih diagramov poteka bo pomagalo opisati vsak korak in interakcije med različnimi vlogami, s čimer bo zapleten potez dela postal jasen in preprost.

Vizualizacija procesov zagotavlja, da je logika procesa pregledna in dostopna vsem članom ekipe.

Isti komunikacijski postopek dodajanja okna v projekt, ki je bil opisan v obliki besedila, sporočil (slika 4.4-1) in blokovnega diagrama, je podoben konceptualnemu modelu, ki smo ga obravnavali v poglavju o modeliranju podatkov (slika 4.4-4).



Slika 4.4-4 Konceptualni diagram prikazuje udeležence projekta kot uporabnike podatkovne zbirke, kjer njihove poizvedbe povezujejo različne sisteme.

Čeprav so konceptualni diagrami pomemben korak, se mnoga podjetja omejijo na to raven, saj menijo, da vizualni diagram zadostuje za razumevanje procesov. To ustvarja iluzijo obvladljivosti: vodje na takšnem diagramu laže zaznajo celotno sliko in vidijo povezave med udeleženci in stopnjami. Vendar takšne sheme ne dajejo jasne predstave o tem, kateri podatki so potrebni za vsakega udeleženca, v kakšni obliki jih je treba posredovati ter kateri parametri in atributi so obvezni za izvedbo avtomatizacije. Konceptualna shema poteka je bolj podobna zemljevidu poti: nakazuje, kdo je v interakciji s kom, vendar ne razkriva, kaj se v teh interakcijah prenaša.

Tudi če je proces podrobno opisan na konceptualni ravni s pomočjo diagramov poteka, to še ne zagotavlja njegove učinkovitosti. Vizualizacija pogosto poenostavi delo vodij, saj jim omogoča lažje spremljanje procesa s poročanjem po korakih. Vendar pa za inženirje podatkovnih zbirk konceptualna predstavitev morda ni dovolj jasna in morda ne omogoča jasnega razumevanja, kako izvesti proces na ravni parametrov in zahtev.

S prehodom na bolj zapletene podatkovne ekosisteme postaja začetno uvajanje konceptualnih in vizualnih orodij ključnega pomena za zagotavljanje, da so podatkovni procesi ne le učinkoviti, temveč tudi usklajeni s strateškimi cilji organizacije. Da bi ta proces dodajanja oken (slika 4.4-1) v celoti prenesli na raven podatkovnih zahtev, moramo iti še globlje in konceptualno vizualizacijo procesa prenesti na logično in fizično raven podatkov, zahtevanih atributov in njihovih mejnih vrednosti.

## **Strukturirane zahteve in regularni izrazi RegEx**

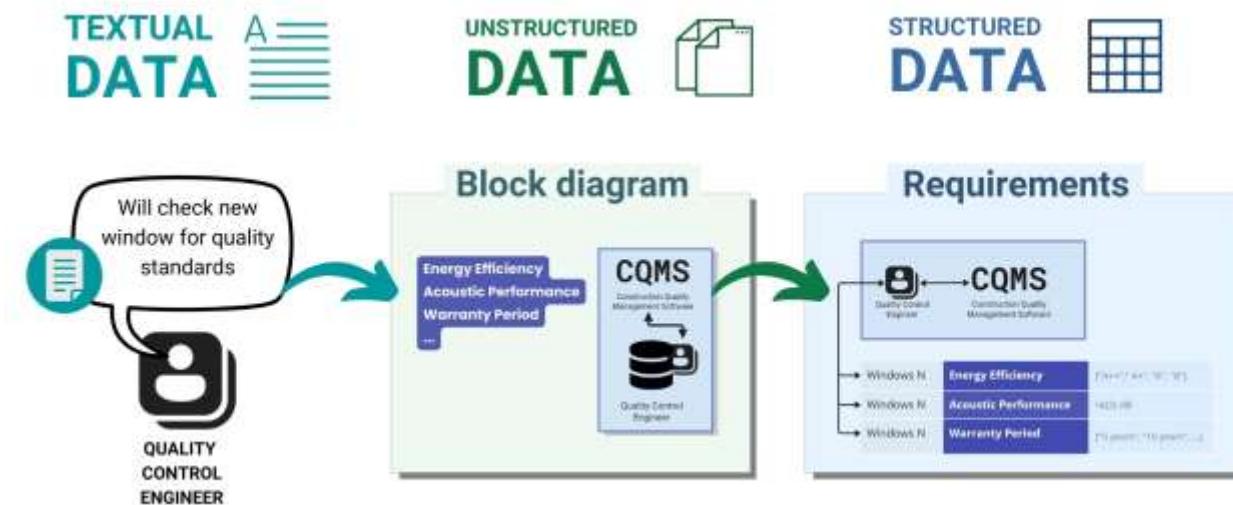
Do 80% podatkov, ustvarjenih v podjetjih, je v nestrukturiranih ali polstrukturiranih oblikah [52] - besedilo, dokumenti, pisma, datoteke PDF, pogовори. Takšne podatke (slika 4.4-1) je težko analizirati, potrjevati, prenašati med sistemi in uporabljati pri avtomatizaciji.

Da bi zagotovili obvladljivost, preglednost in samodejno potrjevanje, je treba besedilne in polstrukturirane zahteve prevesti v dobro opredeljene, strukturirane oblike. Proses strukturiranja ne zadeva le podatkov (ki smo jih podrobno obravnavali v prvih poglavjih tega dela knjige), temveč tudi same zahteve, ki jih udeleženci projekta običajno oblikujejo v obliki prostega besedila v celotnem življenjskem ciklu projekta, pri čemer pogosto ne pomislico, da je te procese mogoče avtomatizirati.

Tako kot smo že pretvorili podatke iz nestrukturirane besedilne oblike v strukturirano obliko, bomo v delovnem procesu zahtev pretvorili besedilne zahteve v strukturirano obliko "logične in fizične plasti".

V okviru primera dodajanja okna (slika 4.4-1) je naslednji korak opis podatkovnih zahtev v obliki preglednice. Podatke za vsak sistem, ki ga uporabljajo udeleženci projekta, bomo strukturirali z določitvijo ključnih atributov in njihovih mejnih vrednosti.

Poglejmo na primer en tak sistem (slika 4.4-5) - sistem vodenja kakovosti v gradbeništvu (CQMS), ki ga uporablja inženir za nadzor kakovosti na strani naročnika. Z njegovo pomočjo preveri, ali je nov element projekta - v tem primeru "novo okno" - skladen z uveljavljenimi standardi in zahtevami.



Slika 4.4-5 Pretvarjanje besedilnih zahtev v obliko tabele z opisi atributov entitete poenostavi razumevanje za druge strokovnjake.

Kot primer si oglejte nekaj pomembnih zahtev za atribute entitet tipa "okenski sistemi" v sistemu CQMS (slika 4.4-6): energetska učinkovitost, akustične lastnosti in garancijski rok. Vsaka kategorija vključuje določene standarde in specifikacije, ki jih je treba upoštevati pri načrtovanju in vgradnji okenskih sistemov.



Slika 4.4-6 Inženir za nadzor kakovosti mora pregledati nove elemente tipa okno glede na energijsko učinkovitost, zvočno izolacijo in garancijske standarde.

Zahteve po podatkih, ki jih inženir za nadzor kakovosti določi v obliki preglednice, imajo na primer naslednje mejne vrednosti:

- **Razred energetske učinkovitosti oken** sega od "A++", ki označuje najvišjo učinkovitost, do "B", ki velja za najnižjo sprejemljivo raven, ti razredi pa so predstavljeni s seznamom sprejemljivih vrednosti **["A++", "A+", "A", "B"]**.

- **Akustična izolacija oken**, merjena v decibelih, ki kaže njihovo sposobnost zmanjševanja hrupa z ulice, je opredeljena z običajnim izrazom \d{2}dB.
- **Atribut "Garancijsko obdobje"** za entiteto "Vrsta okna" se začne pri petih letih, s čimer se to obdobje določi kot najmanjše dovoljeno pri izbiri izdelka; določene so tudi vrednosti garancijskega obdobja, kot so ["5 let", "10 let" itd.] ali logični pogoj ">5 (let)".

V skladu z zbranimi zahtevami v okviru določenih atributov novi elementi kategorije ali razreda oken z ocenami pod "B", na primer "C" ali "D", ne bodo opravili preskusa energetske učinkovitosti. Akustična izolacija oken v podatkih ali dokumentih za inženirja za zagotavljanje kakovosti mora biti označena z dvomestno številko, ki ji sledi predpona "dB", na primer "35 dB" ali "40 dB", vrednosti zunaj te oblike, kot sta "9 D B" ali "100 decibelov", pa ne bodo sprejete (ker ne bodo prestale vzorca za nize RegEx). Garancijski rok se mora začeti z najmanj "5 let", okna s krajsimi garancijskimi roki, kot sta "3 leta" ali "4 leta", pa ne bodo izpolnjevala zahteve, ki jih je inženir kakovosti opisal v obliki tabele.

Za preverjanje takšnih vrednosti atributnih parametrov glede na mejne vrednosti iz zahtev v postopku potrjevanja uporabimo seznam dovoljenih vrednosti ([A", "B", "C"]), slovarje ([A": "H1", "H2"; "B": "W1", "W2"]), logične operacije (npr. ">", "<", "<=", ">=" "=="") za numerične vrednosti) in regularne izraze (za verige in besedilne vrednosti, kot je atribut "Akustična zmogljivost"). Regularni izrazi so izjemno pomembno orodje pri delu z vrednostmi nizov.

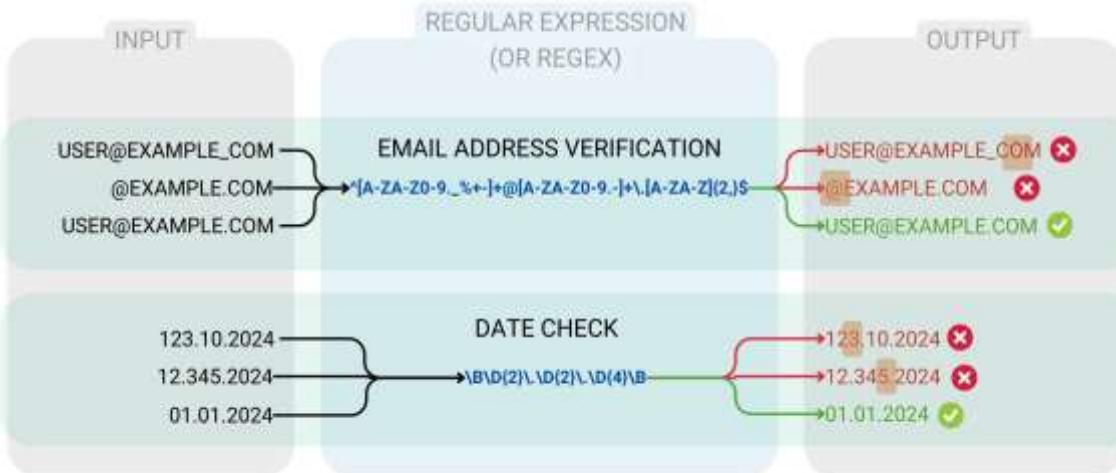
**Regularni izrazi** (RegEx) se v programskih jezikih, vključno s Pythonom (knjižnica Re), uporabljo za iskanje in spremicanje nizov. Regex je kot detektiv v svetu nizov, ki lahko natančno prepozna besedilne vzorce v besedilu.

V regularnih izrazih so črke opisane neposredno z ustrezнимi znaki abecede, številke pa lahko predstavimo s posebnim znakom \d, ki ustreza katerikoli številki od 0 do 9. Kvadratni oklepaji se uporabljajo za označevanje območja črk ali številk, npr. [a-z] za katero koli malo črko latinske abecede ali [0-9], ki ustreza \d. Za neštevilčne in nečrkovne znake se uporablja \D oziroma \W.

Priljubljeni primeri uporabe RegEx (slika 4.4-7):

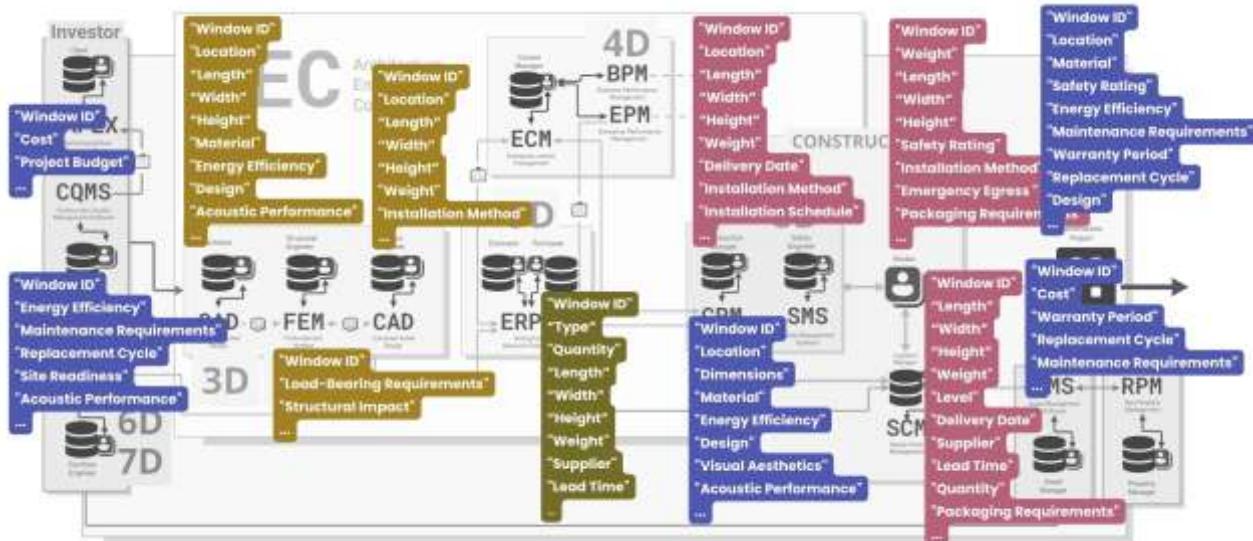
- **Preverjanje e-poštnega naslova** : če želite preveriti, ali je niz veljaven e-poštni naslov, lahko uporabite predlogo "^ [a-zA-Z0-9.\_%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}\$".
- **Izvleček datuma**: predlogo "\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b" lahko uporabite za izvleček datuma iz besedila v obliki DD.MM.YYYY YYYY.
- Preverjanje telefonskih številk : če želite preveriti telefonske številke v obliki +49(000)000-0000, bo vzorec videti kot "+\d{2}\(\d{3}\)\d{3}-\d{4}".

S prevodom zahtev inženirja za zagotavljanje kakovosti v obliko atributov in njihovih mejnih vrednosti (slika 4.4-6) smo jih iz prvotne besedilne oblike (pogovori, dopisi in regulativni dokumenti) pretvorili v organizirano in strukturirano preglednico, kar omogoča samodejno preverjanje in analizo vseh prejetih podatkov (npr. novih elementov kategorije Okno). Prisotnost zahtev omogoča, da se podatki, ki niso bili preverjeni, samodejno zavrnejo, preverjeni podatki pa se samodejno prenesejo v sisteme za nadaljnjo obdelavo.



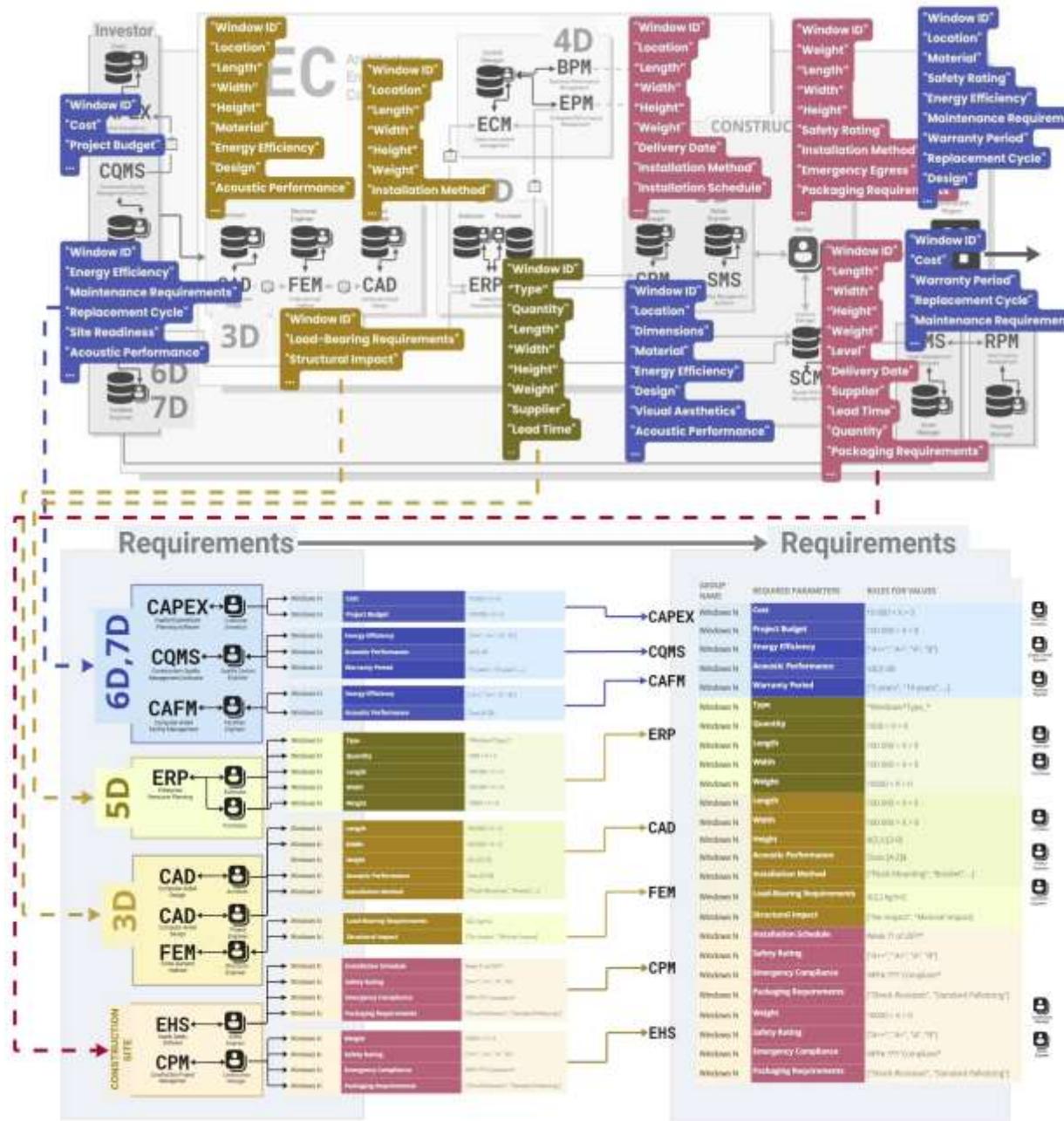
Slika 4.4-7 Uporaba regularnih izrazov je izredno pomembno orodje v postopku potrjevanja besedilnih podatkov.

Sedaj, ko preidemo s konceptualne na logično raven dela z zahtevami, bomo vse zahteve vseh strokovnjakov v našem postopku namestitve novega okna (slika 4.4-4) pretvorili v urejen seznam v obliki atributov in te sezname s potrebnimi atributi dodali v našo shemo poteka za vsakega strokovnjaka (slika 4.4-8).



Slika 4.4-8 Na ravni logičnega procesa se atributi, ki jih obdelujejo posamezni strokovnjaki, dodajo njihovim sistemom.

Z dodajanjem vseh atributov v eno skupno procesno tabelo informacije, ki so bile prej predstavljene kot besedilo in dialog na konceptualni ravni (slika 4.4-1), spremenimo v strukturirano in sistematično obliko tabel na fizični ravni (slika 4.4-9).



**Slika 4.4-9 Pretvarjanje nestrukturiranih strokovnih dialogov v strukturirane tabele pomaga razumeti zahteve na fizični ravni.**

Zdaj je treba zahteve po podatkih sporočiti strokovnjakom, ki pripravljajo informacije za določene sisteme. Če na primer delate v podatkovni zbirki CAD, morate, preden začnete modelirati elemente, zbrati vse potrebne parametre na podlagi scenarijev končne uporabe podatkov. Običajno se to začne v operativni fazi, sledijo gradbišče, oddelek za logistiko, oddelek za ocenjevanje, oddelek za konstrukcijske izračune itd. Šele ko upoštevate zahteve vseh teh področij, lahko začnete ustvarjati podatke - na podlagi zbranih parametrov. To vam bo v prihodnosti omogočilo avtomatizacijo preverjanja in prenosa podatkov vzdolž verige.

Ko novi podatki izpolnjujejo zahteve, se samodejno vključijo v podatkovni ekosistem podjetja in so neposredno namenjeni uporabnikom in sistemom, za katere so namenjeni. Preverjanje podatkov glede na attribute in njihove vrednosti zagotavlja, da podatki izpolnjujejo zahtevane standarde kakovosti in so pripravljeni za uporabo v scenarijih podjetja.

Zahteve po podatkih so bile opredeljene, zdaj pa je treba pred začetkom preverjanja ustvariti, pridobiti ali zbrati podatke, ki jih je treba preveriti, ali pa zabeležiti trenutno stanje informacij v podatkovnih zbirkah, da se lahko uporabijo v postopku preverjanja.

## Zbiranje podatkov za postopek preverjanja

Pred začetkom potrjevanja je treba zagotoviti, da so podatki na voljo v obliki, primerni za postopek potrjevanja. To ne pomeni, da so podatki preprosto na voljo, temveč da jih je treba pripraviti: podatke je treba zbrati in pretvoriti iz nestrukturiranih, ohlapno strukturiranih, besedilnih in geometrijskih oblik v strukturirano obliko. Ta postopek je podrobno opisan v prejšnjih poglavjih, kjer so bile obravnavane metode za preoblikovanje različnih vrst podatkov. Kot rezultat vseh pretvorb so vhodni podatki v obliki odprtih strukturiranih tabel (slika 4.1-2, slika 4.1-9, slika 4.1-13).

Z zahtevami in strukturiranimi preglednicami s potrebnimi parametri in mejnimi vrednostmi (slika 4.4-9) lahko začnemo potrjevati podatke - bodisi kot en sam samodejni postopek (Pipeline) bodisi kot postopno potrjevanje vsakega vhodnega dokumenta.

Za začetek preverjanja je treba prejeti novo datoteko kot vhod ali določiti trenutno stanje podatkov - ustvariti posnetek ali izvoziti trenutne in vhodne podatke ali vzpostaviti povezavo z zunanjim ali notranjim podatkovnim sistemom. V obravnavanem primeru se takšen posnetek ustvari s samodejnim pretvarjanjem podatkov CAD s strukturirano obliko, zabeleženo na primer ob 23:00:00 v petek, 29. marca 2024, potem ko so vsi projektanti odšli domov.



Slika 4.4-10 Posnetek baze podatkov CAD (BIM), ki prikazuje trenutne informacije o atributih za novo entiteto razreda "Okno" v trenutni različici modela projekta.

Zahvaljujoč orodjem za povratni inženiring, ki so obravnavana v poglavju "Prevajanje podatkov CAD (BIM) v strukturirano obliko", lahko te informacije iz različnih orodij CAD (BIM) in urejevalnikov organizirate v ločene tabele (slika 4.4-11) ali združite v eno skupno tabelo, ki združuje različne dele

projekta (slika 9.1-10).

Takšna tabela - zbirka podatkov prikazuje edinstvene identifikatorje oken in vrat (atribut ID), imena tipov (TypeName), dimenzijs (Width, Length), materiale (Material) ter kazalnike energijske in zvočne učinkovitosti in druge značilnosti. Takšno tabelo, izpolnjeno v programu CAD (BIM), projektant zbere iz različnih oddelkov in dokumentov ter tako oblikuje informacijski model projekta.



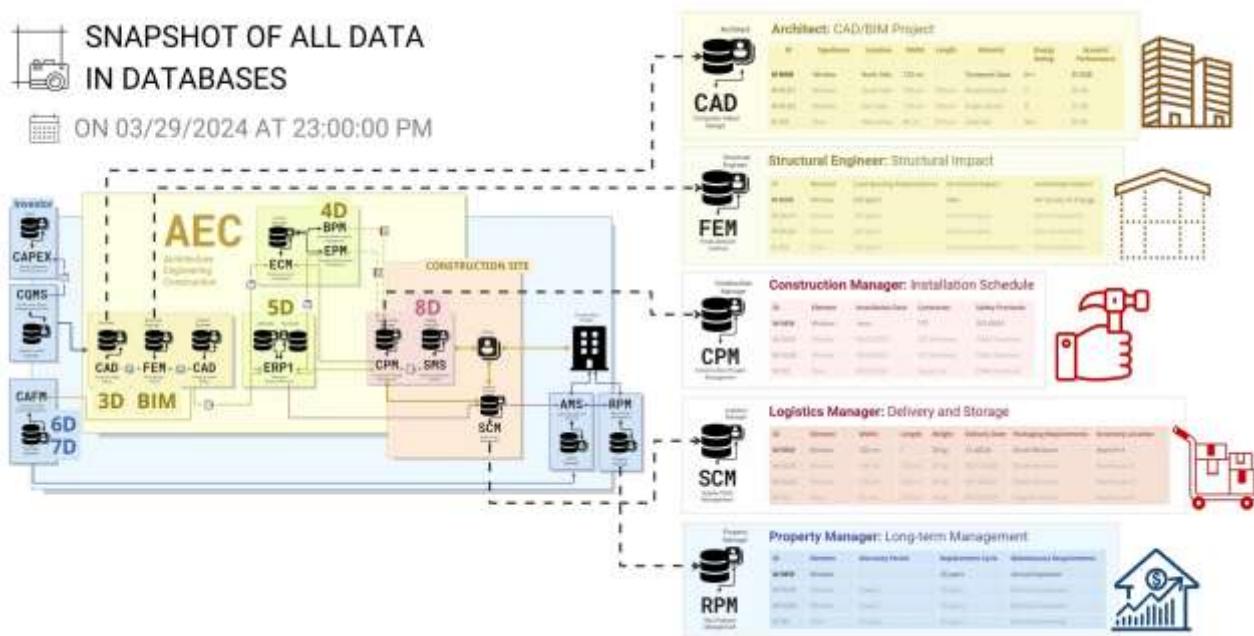
**Slika 4.4-11 Strukturirani podatki iz sistemov CAD so lahko dvodimensionalna tabela s stolpcji, ki označujejo atributi elementov.**

Pravi projekti CAD (BIM) vključujejo več deset ali več sto tisoč elementov (slika 9.1-10). Elementi v formatih CAD so samodejno razvrščeni po vrstah in kategorijah - od oken in vrat do plošč, tal in sten. Enolični identifikatorji (npr. izvorni ID, ki ga samodejno določi rešitev CAD) ali atributi tipa (ime tipa, tip, družina) omogočajo sledenje istega objekta v različnih sistemih. Na primer, novo okno na severni steni stavbe je lahko v vseh ustreznih sistemih organizacije enolično označeno z enotnim identifikatorjem "W-NEW".

Medtem ko morajo biti imena in identifikatorji entitet v vseh sistemih dosledni, se lahko nabor atributov in vrednosti, povezanih s temi entitetami, bistveno razlikuje glede na kontekst uporabe. Arhitekti, gradbeni inženirji, strokovnjaki za gradbeništvo, logistiko in vzdrževanje nepremičnin dojemajo iste elemente na različne načine. Vsak od njih se zanaša na lastne klasifikatorje, standarde in cilje: nekateri gledajo na okno zgolj z estetskega vidika in ocenjujejo njegovo obliko in proporce, drugi pa ga obravnavajo z inženirskega ali operativnega vidika in analizirajo toplotno prevodnost, način vgradnje, težo ali zahteve glede vzdrževanja. Zato je pri modeliranju podatkov in opisovanju elementov pomembno upoštevati vsestransko njihove uporabe in zagotoviti doslednost podatkov ob upoštevanju panožnih posebnosti.

Za vsako vlogo v procesih podjetja obstajajo specializirane podatkovne zbirke s svojim uporabniškim vmesnikom - od načrtovanja in izračunov do logistike, montaže in obratovanja stavbe (slika 4.4-12). Vsak tak sistem upravlja strokovna ekipa strokovnjakov prek posebnega uporabniškega vmesnika ali prek poizvedb v zbirki podatkov, pri čemer za vsoto vseh odločitev o vnesenih vrednostih na koncu verige stoji vodja sistema ali oddelka, ki je odgovoren za pravno veljavnost in kakovost vnesenih

podatkov svojim sogovornikom, ki služijo drugim sistemom.



Slika 4.4-12 Ista entiteta ima v različnih sistemih enak identifikator, vendar različne atribute, ki so pomembni samo v tem sistemu.

Ko smo organizirali zbiranje strukturiranih zahtev in podatkov na logični in fizični ravni, moramo vzpostaviti postopek za samodejno preverjanje podatkov iz različnih vhodnih dokumentov in različnih sistemov glede na predhodno zbrane zahteve.

### Preverjanje podatkov na spletni strani in rezultati preverjanja

Vse nove podatke, ki vstopajo v sistem - naj bodo to dokumenti, tabele ali vnosi v zbirkovo podatkov od naročnika, arhitekta, inženirja, mojstra, logista ali upravljalca nepremičnin - je treba potrditi glede na predhodno oblikovane zahteve (slika 4.4-9). Postopek potrjevanja je ključnega pomena: kakršne koli napake v podatkih lahko privedejo do napačnih izračunov, zamud pri terminskem načrtu in celo finančnih izgub. Za zmanjšanje takšnih tveganj je treba organizirati sistematičen in ponavljajoč se iterativni postopek potrjevanja podatkov.

Za potrditev novih podatkov, ki vstopajo v sistem - nestrukturiranih, besedilnih ali geometrijskih -, jih je treba pretvoriti v ohlapno strukturirano ali strukturirano obliko. Postopek potrjevanja mora nato preveriti podatke glede na popoln seznam zahtevanih atributov in njihovih dovoljenih vrednosti.

Pretvarjanje različnih vrst podatkov: besedila, slike, dokumentov PDF in mešanih podatkov CAD (BIM) v strukturirano obliko je bilo podrobno obravnavano v poglavju "Pretvarjanje podatkov v strukturirano

obliko".

Primer je tabela, pridobljena iz projekta CAD (BIM) (slika 4.4-11). Vključuje polstrukturirane geometrijske podatke in strukturirane atributne informacije o entitetah projekta (slika 3.1-14) - npr. element iz razreda "Windows".

Za izvedbo validacije primerjamo vrednosti atributov (slika 4.4-11) z referenčnimi mejnimi vrednostmi, ki so jih opredelili strokovnjaki v obliki zahteve (slika 4.4-9). Iz končne primerjalne tabele (slika 4.4-13) bo razvidno, katere vrednosti so sprejemljive in katere je treba popraviti, preden lahko podatke uporabimo zunaj aplikacij CAD (BIM).



Slika 4.4-13 V končni preglednici potrjevanja so izpostavljene tiste vrednosti atributov za novo entiteto razreda "Windows", na katere morate biti pozorni.

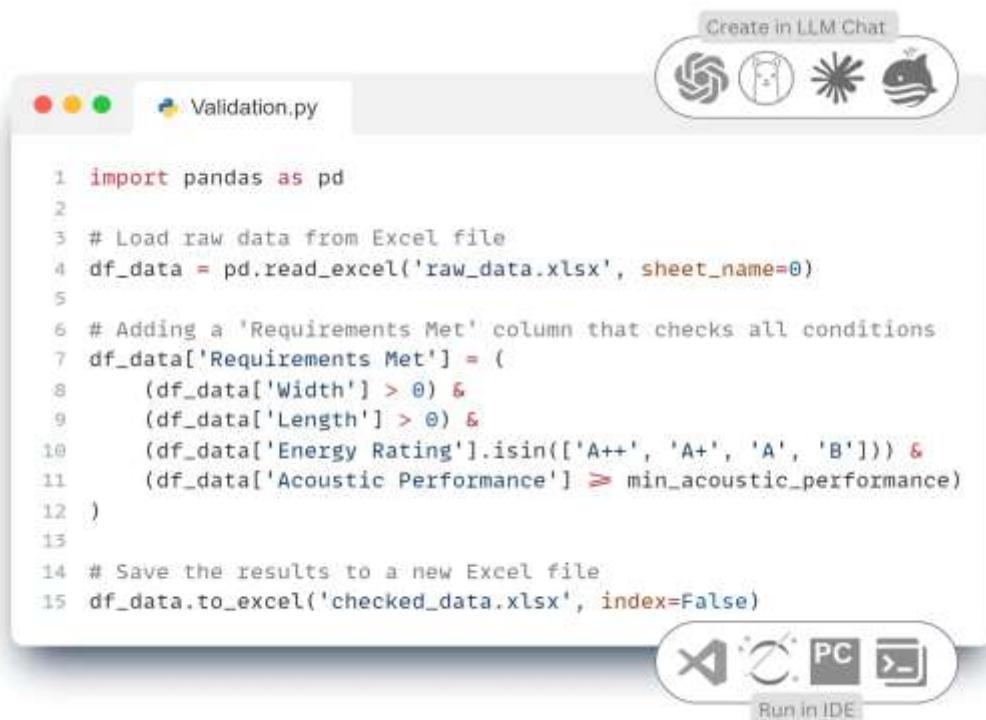
Z izvedbo podobne rešitve z uporabo knjižnice Pandas, ki smo jo opisali prej v poglavju "Pandas: Nepogrešljivo orodje za analizo podatkov", bomo potrdili podatke iz tabelarne datoteke, pridobljene iz datoteke CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (slika 4.4-11), z uporabo zahtev iz druge tabelarne datoteke zahtev (slika 4.4-9).

Za pridobitev kode moramo v pozivu za LLM opisati, da moramo naložiti podatke iz datoteke **raw\_data.xlsx** (celoten niz podatkov iz podatkovne zbirke CAD (BIM)), jih **preveriti** in shraniti rezultat v novo datoteko **checked\_data.xlsx** (slika 4.4-13).

💡 Pridobimo kodo z uporabo LLM brez omembe knjižnice Pandas:

Napišite kodo za preverjanje tabele iz datoteke raw\_data.xlsx in jih potrdite z uporabo naslednjih pravil potrjevanja: vrednosti stolpcev "Width" in "Length" sta večji od nič, "Energy Rating" je vključen v seznam ["A++", "A+", "A", "B"], "Acoustic Performance" pa kot spremenljivka, ki jo bomo določili pozneje - z dodanim zadnjim stolpcem potrjevanja, in končno tabelo shranite v novo datoteko Excel checked\_data.xlsx ↵

- V odgovoru LLM bo opisan kratek primer kode Pythona, ki jo je mogoče z nadaljnimi pozivi izpopolniti in razširiti :



The screenshot shows a code editor window with the file name 'Validation.py' at the top. The code itself is as follows:

```

1 import pandas as pd
2
3 # Load raw data from Excel file
4 df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)
5
6 # Adding a 'Requirements Met' column that checks all conditions
7 df_data['Requirements Met'] = (
8     (df_data['Width'] > 0) &
9     (df_data['Length'] > 0) &
10    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
11    (df_data['Acoustic Performance'] >= min_acoustic_performance)
12 )
13
14 # Save the results to a new Excel file
15 df_data.to_excel('checked_data.xlsx', index=False)

```

At the top of the editor, there are several icons: three colored dots (red, yellow, green), a blue gear icon labeled 'Validation.py', and a 'Create in LLM Chat' button. Below the code, there are four circular icons representing different AI models: a spiral brain, a brain with a gear, a starburst, and a whale. At the bottom right, there is a 'Run in IDE' button with icons for VS Code, PyCharm, and Eclipse.

Slika 4.4-14 Koda, ki jo ustvari model LLM- , preveri pretvorjeno zasnovo CAD (BIM ) glede na zahteve atributov v obliki mejnih vrednosti.

Koda, ki jo ustvari jezikovni model LLM, se lahko uporablja v katerem koli priljubljenem IDE ali spletnem orodju: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z vtičnikom PyDev, Thonny, Wing IDE, IntelliJ IDEA z vtičnikom Python, JupyterLab ali priljubljena spletna orodja Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Izvedba kode (slika 4.4-14) pokaže, da "elementi entitete" W-OLD1, W-OLD2, D-122 (in drugi elementi) iz baze podatkov CAD (BIM) izpolnjujejo zahteve za attribute: širina in dolžina sta večji od nič, razred energetske učinkovitosti pa je ena od vrednosti s seznama "A++", "A", "B", "C" (slika 4.4-15).

Element W-NEW, ki ga potrebujemo in smo ga nedavno dodali ter je odgovoren za nov razred elementa "Window" na severni strani, ni skladen (atribut "Requirements Met"), ker je njegova dolžina enaka nič

(vrednost "0,0" se po našem pravilu "Width">0 šteje za nesprejemljivo) in ne določa razreda energetske učinkovitosti.



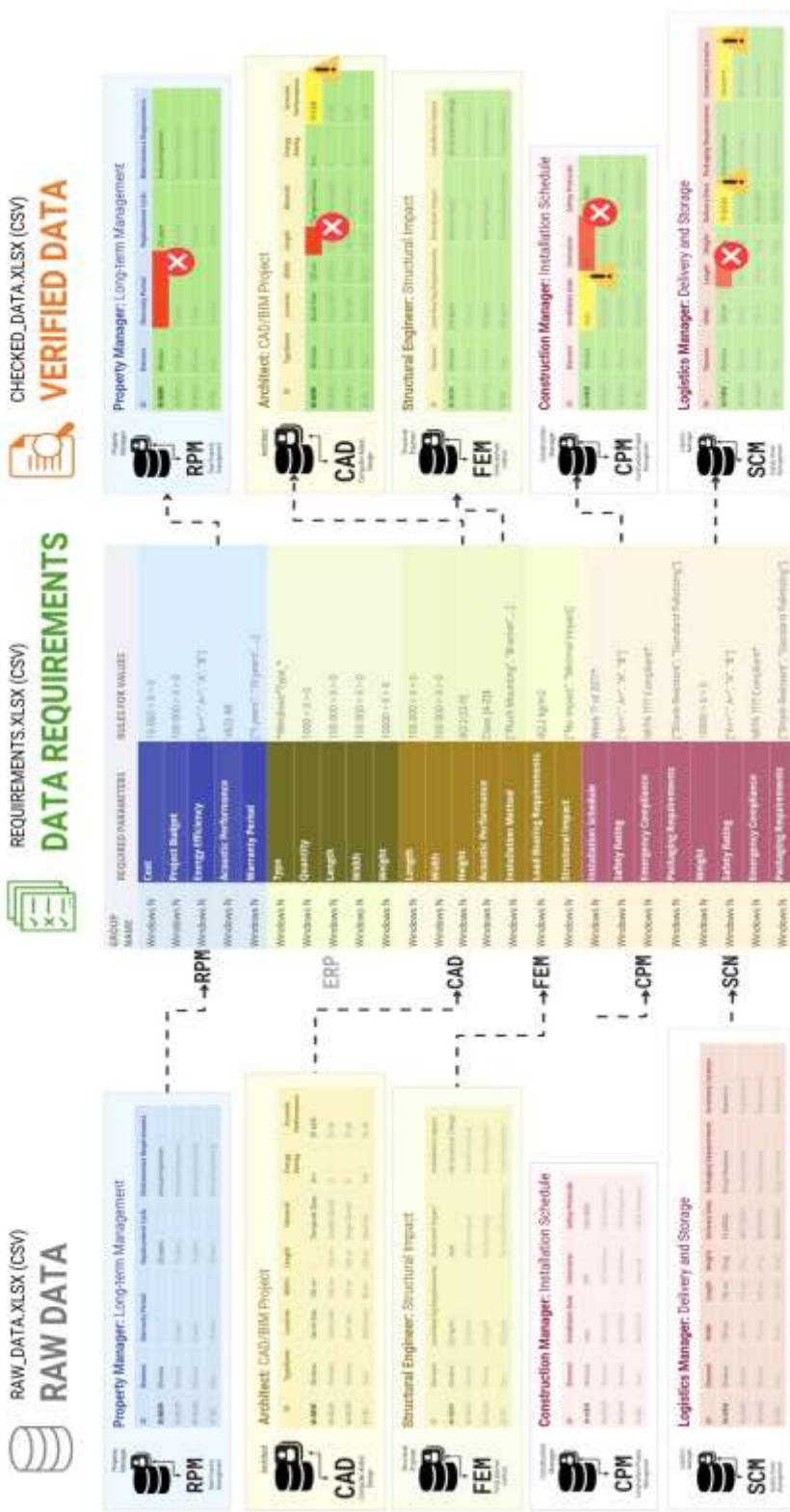
	ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

**Slika 4.4-15** Preverjanje identificira entitete, ki niso šle skozi postopek preverjanja, in rezultatom doda nov atribut z vrednostmi "False" ali "True".

Podobno preverimo skladnost vseh elementov projekta (entitet) in zahtevanih atributov za vsak sistem, tabelo ali podatkovno zbirko v vseh podatkih, ki jih prejmemo od različnih strokovnjakov (slika 4.4-1) med postopkom dodajanja okna v projekt.

V končni preglednici je zaradi vizualizacije primerno barvno poudariti rezultate preverjanja: z zeleno barvo so označeni atributi, ki so uspešno prestali preverjanje, z rumeno vrednostmi z nekritičnimi odstopanj, z rdečo pa kritična odstopanja (slika 4.4-16).

Kot rezultat potrjevanja (slika 4.4-16) dobimo seznam zaupanja vrednih in potrjenih elementov z njihovimi identifikatorji, za katere je bilo preverjeno, da izpolnjujejo zahteve za attribute. Potrjeni elementi zagotavljajo, da ti elementi izpolnjujejo navedene standarde in specifikacije za vse sisteme, vključene v postopek dodajanja elementov razreda Okno ali katerega koli drugega razreda (več o samodejnem potrjevanju podatkov in ustvarjanju samodejnega postopka ETL bomo obravnavali v poglavju "Avtomatizacija ETL in potrjevanje podatkov").



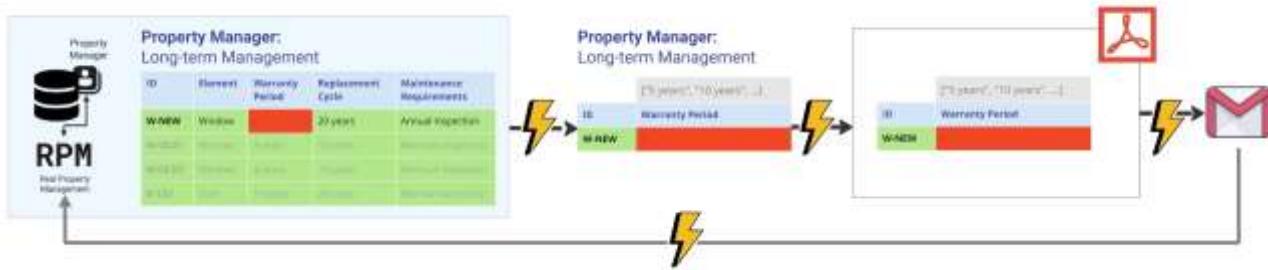
Slika 4.4-16 Rezultat preverjanja, opravljenega za vse sisteme, omogoča ugotoviti, kateri podatki ne ustrezajo zahtevam podjetja.

Entitetam, ki so bile uspešno potrjene, običajno ni treba posvečati veliko pozornosti. Brez ovir preidejo v naslednje faze obdelave in vključevanja v druge sisteme. V nasprotju s "kakovostnimi" postavkami so najbolj zanimive tiste, ki so pri potrjevanju neuspešne. Informacije o takšnih odstopanjih so ključnega pomena: posredovati jih je treba ne le v tabeličnih poročilih, temveč tudi z uporabo različnih orodij za vizualizacijo. Grafični prikaz rezultatov preverjanja pomaga hitro oceniti splošno stanje kakovosti podatkov, prepoznati problematična področja in nemudoma sprejeti korektivne ali popravljalne ukrepe.

## Vizualizacija rezultatov preverjanja

Vizualizacija je bistveno orodje za interpretacijo rezultatov pregleda. Poleg običajnih zbirnih tabel lahko vključuje nadzorne plošče, diagrame in samodejno ustvarjene dokumente PDF, ki združujejo elemente projekta glede na njihov status pregleda. Barvno označevanje ima lahko pri tem podporno vlogo: zelena barva lahko označuje elemente, ki so bili uspešno potrjeni, rumena barva lahko označuje elemente, ki zahtevajo dodatno pozornost, rdeča barva pa elemente, ki imajo kritične napake ali jim manjkajo ključni podatki.

V našem primeru (slika 4.4-1) analiziramo podatke iz vsakega sistema po korakih: od CAD (BIM) in upravljanja nepremičnin do logistike in urnikov montaže (slika 4.4-16). Po reviziji se za vsakega strokovnjaka samodejno ustvarijo posamezna opozorila ali dokumenti s poročili, npr. v formatu PDF (slika 4.4-17). Če so podatki pravilni, strokovnjak prejme kratko sporočilo: "Hvala za sodelovanje". Če so ugotovljena neskladja, se pošlje podrobno poročilo z naslednjim besedilom: "V tem dokumentu so navedeni elementi, njihovi identifikatorji, atributi in vrednosti, katerih skladnost ni bila preverjena".



**Slika 4.4-17 Potrjevanje in samodejno ustvarjanje poročil strokovnjaku, ki ustvarja podatke, pospešita postopek iskanja in razumevanja pomanjkljivosti podatkov.**

Zahvaljujoč samodejnemu postopku potrjevanja - takoj ko je odkrita napaka ali podatkovna vrzel, se osebi, odgovorni za ustvarjanje ali obdelavo ustreznih entitet in njihovih atributov, pošlje takojšnje obvestilo v obliki sporočila v klepetu, e-pošte ali dokumenta PDF (slika 4.4-18) s seznamom elementov in opisov atributov, ki niso bili potrjeni.



Slika 4.4-18 Samodejna poročila o pregledih olajšajo razumevanje napak in pospešijo delo pri izpolnjevanju podatkov o projektu.

Če na primer v sistem za upravljanje nepremičnin prispe dokument (po strukturiranju) z nepravilno izpolnjenim atributom "Garancijski rok", upravitelj nepremičnine prejme opozorilo s seznamom atributov, ki jih je treba preveriti in popraviti.

Podobno se ob morebitnih pomanjkljivostih v urniku namestitve ali logističnih podatkih samodejno ustvari poročilo, ustreznemu strokovnjaku pa se na primer pošlje obvestilo v klepetu ali elektronsko sporočilo z rezultati pregleda.

Poleg dokumentov PDF -dokumentov in grafov z rezultati je mogoče ustvariti nadzorne plošče in interaktivne 3D -modele (slika 7.1-6, slika 7.2-12), ki poudarjajo elemente z manjkajočimi atributi, kar uporabnikom omogoča vizualno uporabo 3D geometrije elementov za filtriranje in ocenjevanje kakovosti in popolnosti teh elementov v projektu.

Vizualizacija rezultatov pregledov na spletni strani v obliki samodejno ustvarjenih dokumentov, grafov ali nadzornih plošč močno poenostavi razlago podatkov in olajša učinkovito komunikacijo med udeleženci projekta.

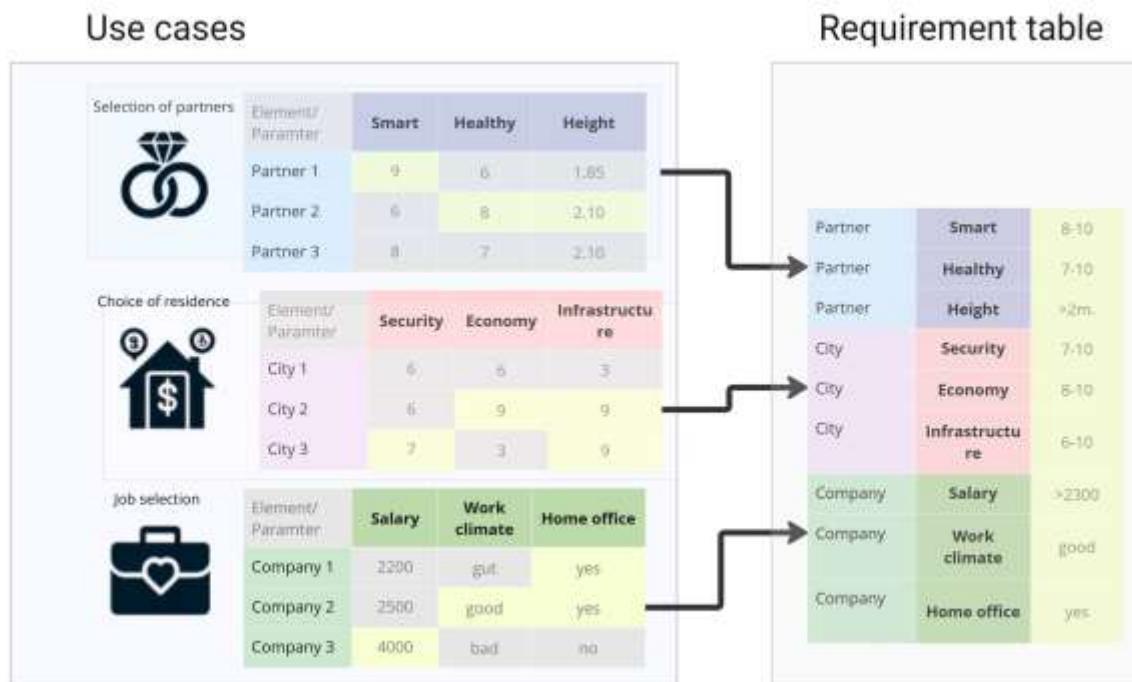
Postopek samodejnega preverjanja podatkov iz različnih sistemov in informacijskih virov lahko primerjamo z informiranim odločanjem v vsakdanjem življenju. Tako kot podjetja v gradbeništvu upoštevajo številne spremenljivke - od zanesljivosti vhodnih podatkov do njihovega vpliva na čas, stroške in kakovost izvedbe projekta - tako tudi človek pri sprejemanju pomembnih odločitev, na primer pri izbiri kraja za življenje, pretehta celo vrsto dejavnikov: prometno dostopnost, infrastrukturo, stroške, varnost, kakovost življenja. Vsi ti premisleki tvorijo sistem merit, ki so osnova za končne odločitve, na katerih temelji naše življenje.

## Primerjava preverjanja kakovosti podatkov s potrebami človeškega življenja

Kljud nenehnemu razvoju metod in orodij za nadzor kakovosti podatkov ostaja temeljno načelo skladnosti informacij nespremenjeno. To načelo je vgrajeno v temelje zrelega sistema upravljanja, tako v poslovnem kot v vsakdanjem življenju.

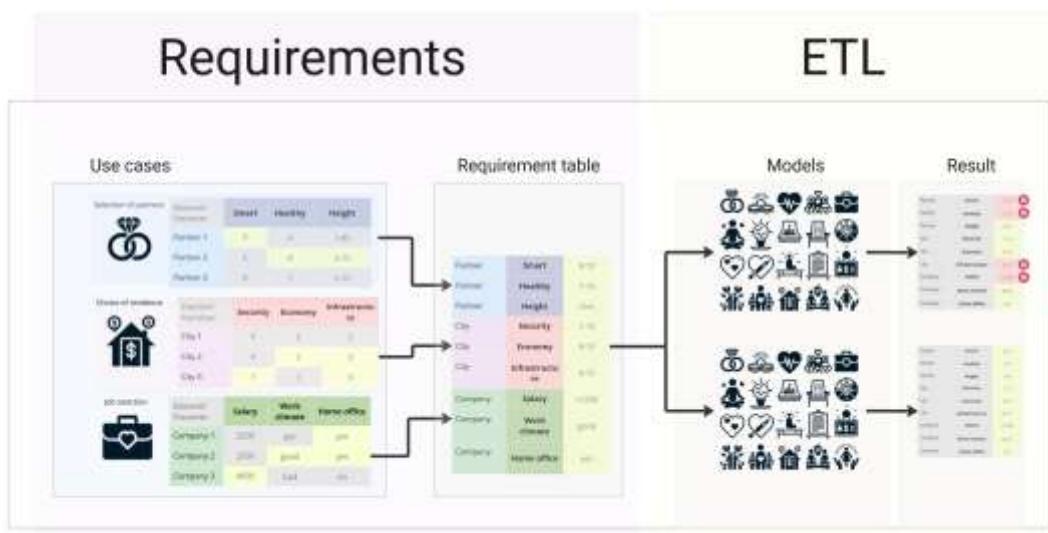
Postopek ponavljajočega se potrjevanja podatkov je zelo podoben procesu odločanja, s katerim se vsakdo vsakodnevno srečuje. V obeh primerih se zanašamo na izkušnje, podatke in nove informacije, ko so na voljo. Vse več življenjskih in poklicnih odločitev - od strateških do vsakodnevnih - pa se sprejema na podlagi podatkov.

Ko na primer izbiramo kraj za življenje ali življenjskega partnerja, si v mislih intuitivno oblikujemo tabelo merit in značilnosti, po katerih primerjamo alternative (slika 4.4-19). Te značilnosti - pa naj gre za osebne lastnosti osebe ali parametre nepremičnine - predstavljajo attribute, ki vplivajo na končno odločitev.



Slika 4.4-19 Izbira prebivališča, zaposlitve ali partnerstva temelji na zahtevah posameznih atributov.

Uporaba strukturiranih podatkov in formaliziranega pristopa k opisovanju zahtev (slika 4.4-20) prispeva k bolj premišljenim in informiranim odločitvam v poklicnem in zasebnem življenju.

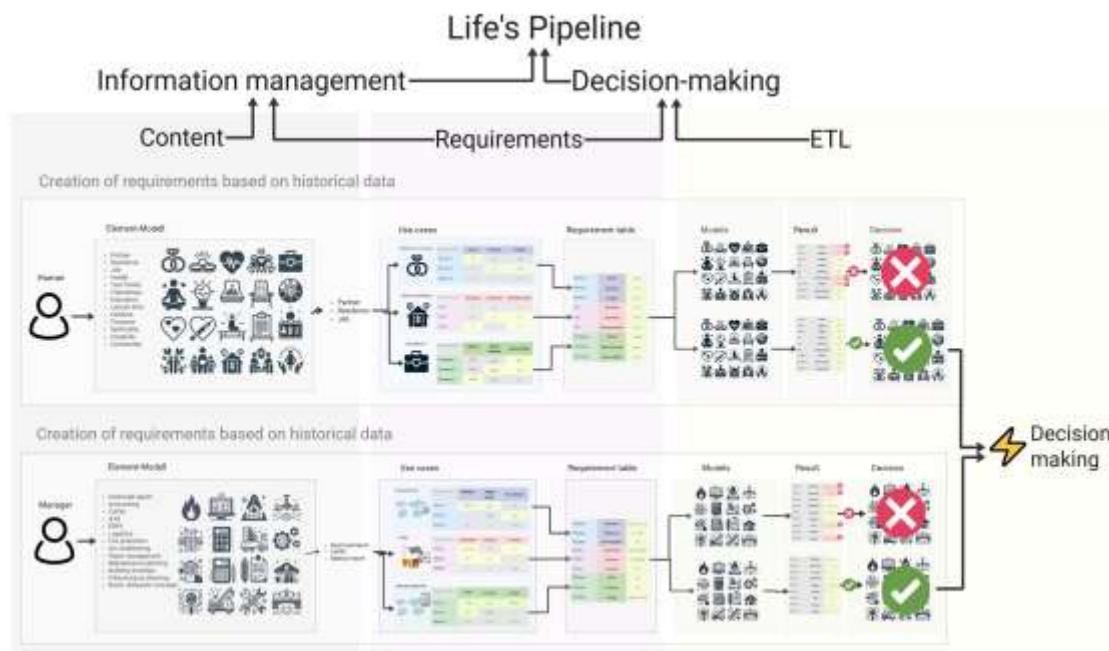


Slika 4.4-20 Formalizacija zahtev pomaga sistematizirati dojemanje življenjskih in poslovnih odločitev.

Pristop k odločanju na podlagi podatkov ni izključno poslovno orodje. Neovirano se vključuje tudi v vsakdanje življenje, in sicer po skupnih korakih obdelave podatkov (slika 4.4-21), podobnih procesu ETL (Extract, Transform, Load), ki smo ga obravnavali že na začetku tega dela pri strukturiranju podatkov in ga bomo podrobno preučili v kontekstu avtomatizacije nalog v sedmem delu knjige:

- **Podatki kot temelj (izvleček):** Na vseh področjih - tako na delovnem kot v zasebnem življenju - zbiramo informacije. V poslu so to lahko poročila, številke, tržni podatki, v osebnem življenju pa osebne izkušnje, nasveti bližnjih, povratne informacije, opažanja.
- **Merila vrednotenja (preoblikovanje):** zbrane informacije se razlagajo na podlagi vnaprej določenih meril. Pri delu so to kazalniki uspešnosti (KPI), proračunske omejitve in norme; v osebnem življenju pa parametri, kot so cena, udobje, zanesljivost, karizma itd.
- **Napovedovanje in analiza tveganja (obremenitev):** zadnji korak vključuje sprejetje odločitve na podlagi analize preoblikovanih podatkov in primerjave možnih posledic. To je podobno poslovnim procesom, pri katerih gredo podatki skozi filter poslovne logike in tveganja.

Odločitve, ki jih sprejemamo - od trivialnih želja, kot je, kaj jesti za zajtrk, do pomembnih življenjskih dogodkov, kot je izbira poklicne poti ali življenjskega partnerja - so po naravi rezultat obdelave in vrednotenja podatkov.



**Slika 4.4-21 Poslovanje in življenje na splošno je niz odločitev, ki temeljijo na podatkih, pri čemer je ključni dejavnik kakovost podatkov, ki se uporablja za sprejemanje odločitev.**

Vse v našem življenju je medsebojno povezano in tako kot živi organizmi, vključno z ljudmi, sledijo zakonom narave, se razvijajo in prilagajajo spreminjačom se razmeram, tudi človeški procesi, vključno s tem, kako zbiramo in analiziramo podatke, odražajo ta naravna načela. Tesna povezanost med naravo in človeškimi dejavnostmi potrjuje ne le našo odvisnost od narave, temveč tudi našo željo po uporabi zakonov, izpopolnjenih v milijonih let evolucije, za ustvarjanje podatkovnih arhitektur, procesov in sistemov za odločanje.

Nove tehnologije, zlasti v gradbeništvu, so odličen primer, kako narava vedno znova navdihuje človeštvo pri ustvarjanju boljših, trajnostnih in učinkovitih rešitev.

### Naslednji koraki: pretvorba podatkov v natančne izračune in načrte

V tem delu smo si ogledali, kako pretvoriti nestrukturirane podatke v strukturirano obliko, razviti podatkovne modele in organizirati postopke za preverjanje kakovosti informacij v gradbenih projektih. Upravljanje, standardizacija in razvrščanje podatkov je temeljni proces, ki zahteva sistematičen pristop in jasno razumevanje poslovnih zahtev. Tehnike in orodja, obravnavana v tem delu, omogočajo zanesljivo povezovanje med različnimi sistemi v celotnem življenjskem ciklu objekta.

Za povzetek tega dela poudarimo glavne praktične korake, ki vam bodo pomagali pri uporabi obravnavanih pristopov pri vsakodnevnih nalogah:

- Začnite s sistematizacijo zahtev
- Ustvarite register atributov in parametrov za ključne elemente vaših projektov in

procesov.

- Dokumentiranje mejnih vrednosti za vsak atribut
- vizualizacija procesov in odnosov med razredi, sistemi in atributi z uporabo diagramov poteka (npr. v programih Miro, Canva, Visio).

#### ■ Avtomatiziranje pretvorbe podatkov

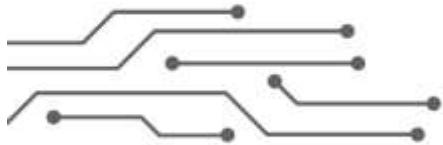
- Preverite, katere dokumente, ki se pogosto uporabljajo v procesih, lahko digitalizirate s pomočjo knjižnic OCR in jih pretvorite v tabelarno obliko.
- Oglejte si orodja za povratni inženiring za pridobivanje podatkov iz CAD (BIM).
- Poskusite nastaviti samodejno pridobivanje podatkov iz dokumentov ali formatov, ki jih pogosto uporabljate pri svojem delu, v obliki tabele.
- Nastavite samodejno pretvorbo med različnimi podatkovnimi formati.

#### ■ Ustvarite bazo znanja za kategorizacijo

- Razvoj notranjega ali uporaba obstoječega klasifikatorja elementov, usklajenega z industrijskimi standardi.
- dokumentiranje medsebojnih povezav med različnimi klasifikacijskimi sistemi
- Z ekipo razpravljamte o uporabi enotnega sistema identifikacije in nedvoumne klasifikacije elementov.
- Začnite vzpostavljati postopek za samodejno potrjevanje podatkov - tako tistih, s katerimi delate znotraj ekipe, kot tistih, ki se posredujejo v zunanje sisteme.

Z uporabo teh pristopov lahko znatno izboljšate kakovost podatkov ter poenostavite njihovo nadaljnjo obdelavo in preoblikovanje. V naslednjih delih te knjige si bomo ogledali, kako uporabiti že strukturirane in pripravljene podatke za avtomatizirane izračune, ocenjevanje stroškov, načrtovanje in vodenje gradbenih projektov.





## V DEL

### **IZRAČUNI STROŠKOV IN ČASA: VKLJUČEVANJE PODATKOV V GRADBENE PROCESE.**

Peti del je posvečen praktičnim vidikom uporabe podatkov za optimizacijo obračunavanja stroškov in načrtovanja gradbenih projektov. Podrobno sta analizirana metoda ocenjevanja stroškov, ki temelji na virih, in avtomatizacija postopkov ocenjevanja. Obravnavane so metode avtomatiziranega pridobivanja količinskega prevzema) iz modelov CAD (BIM) in njihova integracija s sistemi za izračun. Raziskane so tehnologije 4D in 5D modeliranja za časovno načrtovanje in upravljanje gradbenih stroškov s konkretnimi primeri njihove uporabe. Predstavljena je analiza razširjenih informacijskih slojev 6D -8D, ki zagotavljajo celovit pristop k ocenjevanju trajnosti, delovanja in varnosti nepremičninskih objektov. Podrobno so obravnavane metode izračuna ogljičnega odtisa in kazalnikov ESG gradbenih projektov v okviru sodobnih okoljskih zahtev in standardov. Kritično so ocenjene možnosti in omejitve tradicionalnih sistemov ERP in PMIS pri upravljanju gradbenih procesov z analizo njihovega vpliva na preglednost oblikovanja cen. Napovedane so možnosti za prehod od zaprtih rešitev k odprtим standardom in prilagodljivim orodjem za analizo podatkov, ki lahko zagotovijo večjo učinkovitost gradbenih procesov.

## POGLAVJE 5.1.

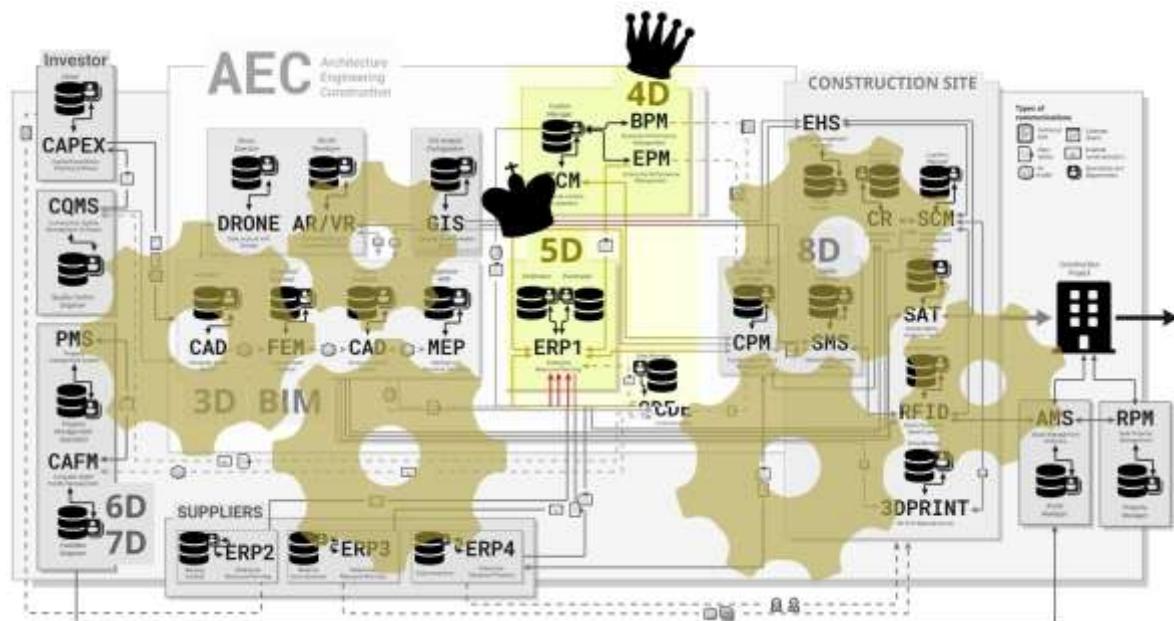
### IZRAČUNI IN OCENE STROŠKOV ZA GRADBENE PROJEKTE.

#### Osnove gradbeništva: ocenjevanje količine, stroškov in časa

Med številnimi poslovnimi procesi, ki določajo trajnost podjetja v gradbeništvu, so - tako kot pred tisočletji - še posebej pomembni procesi za natančno ocenjevanje števila elementov, stroškov projekta in dobavnega roka (slika 5.1-1).

Razvoj pisave je bil posledica številnih dejavnikov, med drugim potrebe po zapisovanju gospodarskih transakcij, trgovine in upravljanja virov v zgodnjih družbah. Prvi pravno pomembni dokumenti, glinene ploščice z izračuni stroškov materiala in dela, so se uporabljali v okviru trgovine in gradbeništva. Te tablice so beležile obveznosti strank pri gradnji objektov in so se hranile kot dokazilo o sporazumih ter denarnih in trgovinskih odnosih.

Tisočletja je pristop k ocenjevanju ostal v glavnem nespremenjen: izračuni so se izvajali ročno, zanašali pa so se na izkušnje in intuicijo ocenjevalca. S prihodom modularnih sistemov ERP in orodij CAD pa se je tradicionalni pristop k ocenjevanju količine, stroškov in časa začel hitro spremenjati. Današnje digitalne tehnologije omogočajo popolno avtomatizacijo ključnih izračunov časa in stroškov, kar omogoča večjo natančnost, hitrost in preglednost pri načrtovanju virov gradbenih projektov.



Slika 5.1-1 Med številnimi različnimi sistemi so v poslovanju najpomembnejša orodja, ki so odgovorna za merjenje obsega, stroškov in časa.

Gradbena podjetja se osredotočajo na natančne podatke o času in stroških. Ti pa so odvisni od količine porabljenega materiala in dela, njihova preglednost pa vpliva na dobičkonosnost. Vendar pa zapletenost računskih postopkov in njihova nepreglednost pogosto privедeta do višjih stroškov projekta, zamujanja rokov in celo stečaja.

Po podatkih poročila KPMG "Znani problemi - novi pristopi" (2023) je le 50% gradbenih projektov končanih pravočasno, 87% podjetij pa poroča o povečanem nadzoru nad ekonomičnostjo investicijskih projektov. Glavne težave so povezane s pomanjkanjem usposobljenega osebja in težavami pri napovedovanju tveganj [2].

Zgodovinski podatki o stroških in času procesa se zbirajo med gradnjo preteklih projektov v celotnem obdobju delovanja gradbenega podjetja in se vnašajo v zbirke podatkov različnih sistemov (ERP, PMIS BPM, EPM itd.).

Kakovostni podatki o preteklih stroških so pomembna konkurenčna prednost gradbene organizacije, saj neposredno vplivajo na njeno sposobnost preživetja.

V gradbenih in inženirskih podjetjih so oddelki za ocenjevanje in kalkulacijo stroškov ustanovljeni za zbiranje, shranjevanje in posodabljanje preteklih podatkov o projektnih izračunih. Njihova glavna naloga je zbirati in sistematizirati izkušnje podjetja, kar omogoča, da se natančnost ocenjevanja obsega, časovnega okvira in stroškov novih projektov sčasoma izboljša. Ta pristop pomaga zmanjšati napake v prihodnjih izračunih, ki temeljijo na praksi in rezultatih že izvedenih projektov.

## Metode za izračun ocenjenih stroškov projektov

Strokovnjaki za ocenjevanje stroškov uporabljajo različne metode ocenjevanja, pri čemer se vsaka osredotoča na določeno vrsto podatkov, razpoložljivost informacij in raven podrobnosti projekta. Najpogostejše so:

- **Metoda, ki temelji na virih:** ocena ocenjenih stroškov projekta na podlagi podrobne analize vseh potrebnih virov, kot so materiali, oprema in delo. Ta metoda zahteva podroben seznam vseh nalog in virov, potrebnih za izvedbo vsake naloge, ki mu sledi izračun njihovih stroškov. Ta metoda je zelo natančna in se pogosto uporablja pri ocenjevanju stroškov.
- **Parametrična metoda:** uporablja statistične modele za oceno stroškov na podlagi parametrov projekta. To lahko vključuje analizo stroškov na mersko enoto, kot je površina stavbe ali obseg del, in prilagoditev teh vrednosti posebnim pogojem projekta. Metoda je še posebej učinkovita v zgodnjih fazah, ko podrobne informacije še niso na voljo.
- Metoda na enoto (**metoda stroškov na enoto**): izračuna ocenjene stroške projekta na podlagi stroškov na mersko enoto (npr. na kvadratni meter ali kubični meter). To omogoča hiter in enostaven način za primerjavo in analizo stroškov različnih projektov ali delov projektov.
- **Strokovna presoja** (metoda Delphi): temelji na mnenjih strokovnjakov, ki za oceno vrednosti projekta uporabijo svoje izkušnje in znanje. Ta pristop je uporaben, kadar natančni izhodiščni podatki niso na voljo ali kadar je projekt edinstven.

Omeniti velja, da je mogoče parametrično metodo in strokovne ocene prilagoditi modelom strojnega

učenja. To omogoča samodejno izdelavo napovedi stroškov in časovnega načrta projekta na podlagi učnih vzorcev. Primeri uporabe takšnih modelov so podrobneje obravnavani v poglavju "Primer uporabe strojnega učenja za iskanje stroškov in časovnic projekta" (slika 9.3-5).

Kljud temu je metoda, ki temelji na virih, še vedno najbolj priljubljena in se pogosto uporablja v svetovni praksi. Ne zagotavlja le natančne ocene ocenjenih stroškov, temveč omogoča tudi izračun trajanja posameznih procesov na gradbišču in celotnega projekta kot celote (več podrobnosti v poglavju "Gradbeni urniki in 4D -podatki o projektu").

## **Metoda, ki temelji na virih obračunavanje stroškov in ocenjevanje v gradbeništvu**

Obračunavanje stroškov na podlagi virov je metoda poslovodnega računovodstva, pri kateri stroški projekta temeljijo na neposrednem obračunavanju vseh vključenih virov. V gradbeništvu ta pristop vključuje podrobno analizo in ovrednotenje vseh materialnih, delovnih in tehničnih virov, potrebnih za izvedbo dela.

Metoda, ki temelji na virih, zagotavlja visoko stopnjo preglednosti in natančnosti pri načrtovanju proračuna, saj se osredotoča na dejanske cene virov v času ocenjevanja. To je še posebej pomembno v nestabilnem gospodarskem okolju, kjer lahko nihanja cen bistveno vplivajo na skupne stroške projekta.

V naslednjih poglavjih si bomo podrobno ogledali postopek izračunavanja stroškov na podlagi virov. Za boljše razumevanje njegovih načel v gradbeništvu bomo naredili analogijo s kalkulacijo stroškov večerje v restavraciji. Vodja restavracije pri načrtovanju večera sestavi seznam potrebnih izdelkov, upošteva čas priprave posamezne jedi in nato stroške pomnoži s številom gostov. Pri gradnji je postopek podoben: za vsako kategorijo elementov projekta (objektov) se izdelajo predračuni po postavkah Recepti, skupni stroški projekta pa se določijo tako, da se vsi stroški seštejejo v skupni račun - končni predračun po kategorijah.

Ključna in začetna faza pristopa, ki temelji na virih, je oblikovanje začetne podatkovne zbirke podjetja. V prvi fazi izračuna stroškov se sestavi strukturiran seznam vseh predmetov, materialov, vrst dela in virov, ki jih ima podjetje na voljo v okviru svojih gradbenih projektov - od žebbla v skladišču do opisa ljudi z njihovo usposobljenostjo in urno postavko. Te informacije so sistematizirane v enotni "zbirki podatkov o gradbenih sredstvih in materialih" - tabelaričnem registru, ki vsebuje podatke o imenih, značilnostih, merskih enotah in trenutnih cenah. Ta podatkovna zbirka postane glavni in primarni vir informacij za vse nadaljnje izračune virov - tako stroškov kot časovnega razporeda del.

## **Podatkovna baza gradbenih virov: katalog gradbenih materialov in del**

Podatkovna zbirka ali tabela gradbenih virov in materialov - vključuje podrobne informacije o vsakem elementu, ki se lahko uporabi v gradbenem projektu - izdelek, predmet, material ali storitev, vključno z njegovim imenom, opisom, mersko enoto in stroški na enoto, zapisane v strukturirani obliki. V tej

preglednici lahko najdete vse od različnih vrst goriv in materialov, ki se uporabljajo v projektih, do podrobnih seznamov strokovnjakov v obliki različnih kategorij z opisi urnih postavk (slika 5.1-2).

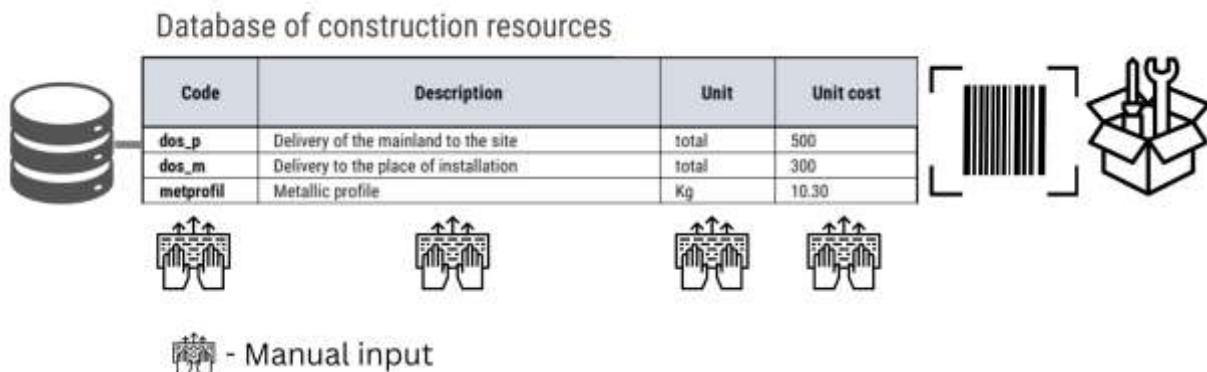
Database of resources	
 1st grade potatoes <b>1 kg   \$2,99</b>	 Sand lime bricks <b>1 pcs   \$1</b>
 Black Angus marble beef <b>1 kg   \$26,99</b>	 JCB 3CX backhoe loader <b>1 h   \$150</b>
 Broccoli <b>1 pcs   \$1,99</b>	 Laborer of the 1st category <b>1 h   \$30</b>

Slika 5.1-2 Preglednica virov je seznam sestavin, ki opisuje material in storitev s stroški na enoto.

"Podatkovna baza virov" je podobna katalogu izdelkov v spletni trgovini, kjer ima vsak izdelek podrobni opis svojih lastnosti. Tako ocenjevalci stroškov lažje izberejo prave vire (podobno kot izbirajo izdelke pri dodajanju v nakupovalno košarico), ki so potrebni za izračun določenih gradbenih postopkov v obliki kalkulacij (končno naročilo v spletni trgovini).

Podatkovno zbirkovo virov si lahko predstavljamo tudi kot seznam vseh sestavin v kuharski knjigi restavracije. Vsak gradbeni material, oprema in storitev so podobni sestavinam, ki se uporabljajo v receptih. "Baza podatkov o virih" je podrobni seznam vseh sestavin - gradbenih materialov in storitev, vključno z njihovimi stroški na enoto: kos, meter, uro, liter itd.

V preglednico "Podatkovne zbirke gradbenih virov" lahko nove elemente entitet dodate na dva načina - ročno (slika 5.1-3) ali samodejno z integracijo s sistemi za upravljanje zalog podjetja ali podatkovnimi zbirkami dobaviteljev.



**Slika 5.1-3 Podatkovna zbirka virov se izpolni ročno ali samodejno prevzame podatke iz drugih podatkovnih zbirk.**

Tipično srednje veliko gradbeno podjetje uporablja zbirko podatkov, ki vsebuje na tisoče, včasih pa tudi več deset tisoč elementov s podrobnnimi opisi, ki se lahko uporabljajo pri gradbenih projektih. Ti podatki se nato samodejno uporabijo v pogodbah in projektni dokumentaciji za natančen opis sestave del in postopkov

Da bi sledili spreminjačim se tržnim razmeram, kot je inflacija, se atribut "stroški na enoto" za vsak izdelek (blago ali storitev) v podatkovni zbirki virov (slika 5.1-3) redno posodablja ročno ali s samodejnim prenosom trenutnih cen iz drugih sistemov ali spletnih platform.

Posodabljanje stroškov na enoto vira se lahko izvaja mesečno, četrletno ali letno - odvisno od narave vira, inflacije in zunanjih gospodarskih razmer. Takšne posodobitve so potrebne za ohranjanje natančnosti izračunov in ocen, saj so to osnovni elementi, ki služijo kot izhodišče za delo ocenjevalcev stroškov. S posodobljenimi podatki se izdelajo ocene, proračuni in časovni načrti, ki odražajo dejanske tržne razmere in zmanjšujejo tveganje napak pri poznejših projektnih izračunih.

## Sestavljanje izračunov in obračunavanje stroškov del na podlagi baze virov

Potem ko ste napolnili "Podatkovno zbirko gradbenih virov" (slika 5.1-3) z minimalnimi enotami, lahko začnete ustvarjati izračune, ki so izračunani za vsak postopek ali delo na gradbišču za določene merske enote: na primer za en kubični meter betona, en kvadratni meter stene iz mavčnih plošč, na meter robnika ali na vgradnjo okna.

Na primer, za gradnjo opečnega zidu s površino 1 m<sup>2</sup> (slika 5.1-4) je na podlagi izkušenj iz prejšnjih projektov potrebnih približno 65 opek (entiteta "Silikatna opeka") po ceni 1 USD na kos (atribut "Cena na kos"), kar skupaj znaša 65 USD. Prav tako je po mojih izkušnjah treba za 10 minut uporabit gradbeno opremo (entiteta "JCB 3CX Loader"), ki bo opeke položila v bližino delovnega območja. Ker najem opreme stane 150 USD na uro, bi 6 minut uporabe stalo približno 15 USD. Poleg tega bo za 2 uri potreben izvajalec za polaganje opeke, katerega urna postavka znaša 30 USD, skupaj pa 60 USD.

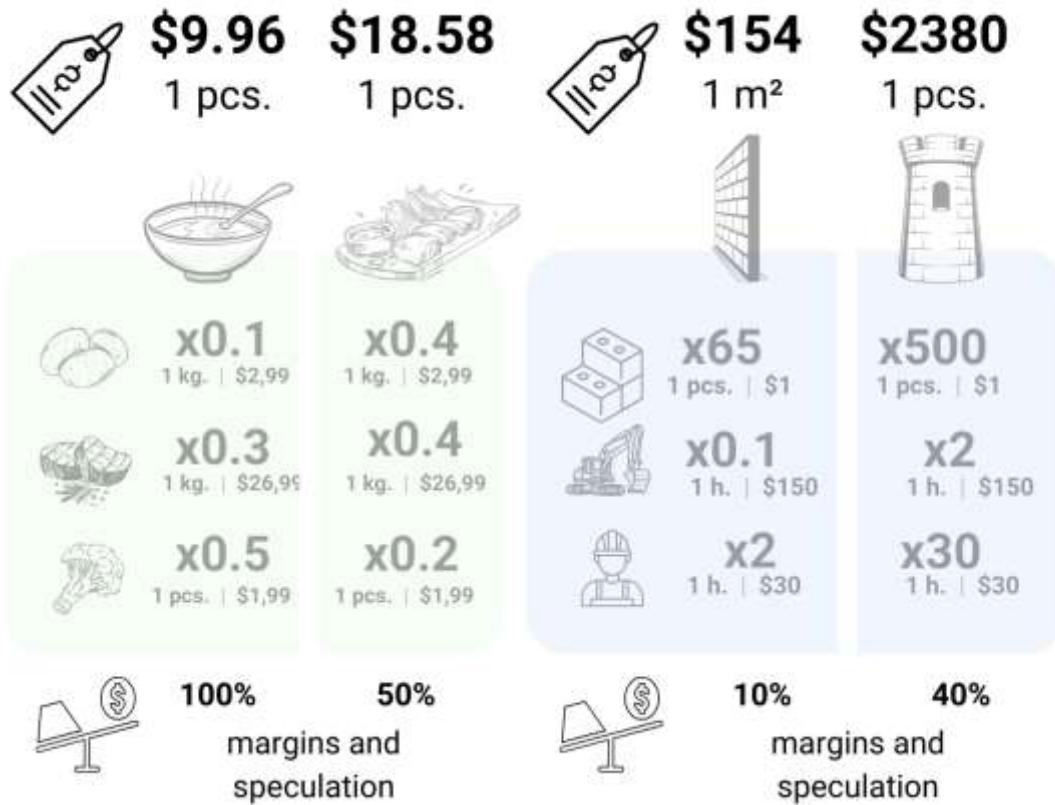
<b>x0.1</b> 1 kg   \$2,99	<b>x0.4</b> 1 kg.   \$2,99	<b>x65</b> 1 pcs   \$1	<b>x500</b> 1 pcs   \$1
<b>x0.3</b> 1 kg   \$26,99	<b>x0.4</b> 1 kg.   \$26,99	<b>x0.1</b> 1 h   \$150	<b>x2</b> 1 h   \$150
<b>x0.5</b> 1 pcs   \$1,99	<b>x0.2</b> 1 pcs.   \$1,99	<b>x2</b> 1 h   \$30	<b>x30</b> 1 h   \$30
+ margins and speculation	+ margins and speculation	+ margins and speculation	+ margins and speculation

Slika: 5.1-4 Izračuni stroškov vsebujejo podroben seznam gradbenega materiala in storitev, potrebnih za izvedbo del in postopkov.

Sestava izračunov (tako imenovani "recepti") se oblikuje na podlagi preteklih izkušenj, ki jih je podjetje pridobilo pri opravljanju velikega obsega podobnega dela. Te praktične izkušnje se običajno nabirajo s povratnimi informacijami z gradbišča. Zlasti mojster zbira informacije neposredno na gradbišču, kjer beleži dejanske stroške dela, porabo materiala in nianse tehnoloških postopkov. V sodelovanju z oddelkom za ocenjevanje se te informacije nato iterativno izboljšujejo: izboljšujejo se opisi postopkov, prilagaja se sestava virov in posodablja se kalkulacije stroškov, da odražajo dejanske podatke iz nedavnih projektov.

Tako kot so v receptu opisane sestavine in količine, potrebne za pripravo jedi, je v predračunske listu podroben seznam vseh gradbenih materialov, sredstev in storitev, potrebnih za dokončanje določenega dela ali postopka.

Redno opravljeno delo omogoča delavcem, mojstrom in ocenjevalcem, da se orientirajo v potrebni količini virov: materiala, goriva, delovnega časa in drugih parametrov, potrebnih za izvedbo enote dela (slika 5.1-5). Ti podatki se v ocenjevalne sisteme vnašajo v obliki tabel, v katerih je vsaka naloga in operacija opisana z minimalnimi elementi baze virov (s stalno posodobljenimi cenami), kar zagotavlja natančnost izračunov.



Slika 5.1-5 Cene na enoto se zbirajo za vsako delo, pri čemer se atribut obsega entitete pomnoži z njeno količino in doda odstotek dobička.

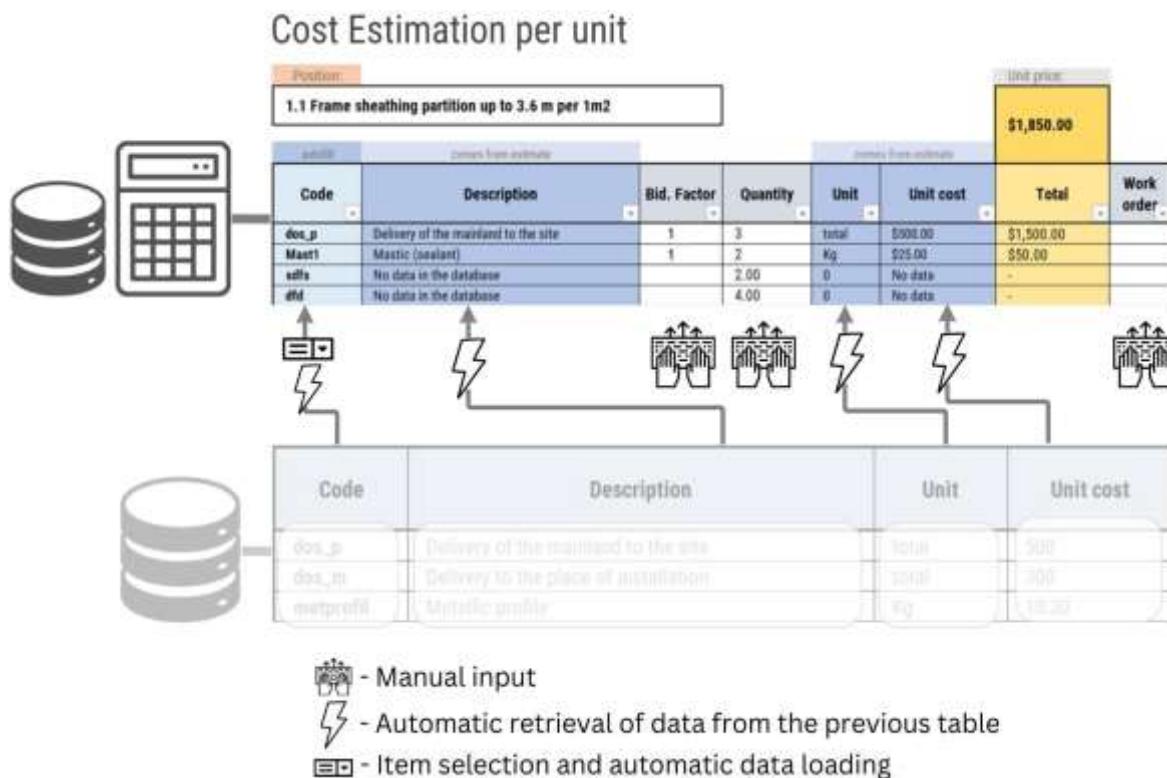
Da bi dobili celotne stroške vsakega procesa ali dejavnosti (predmeta kalkulacije stroškov), se atribut stroškov pomnoži z njihovim številom in faktorji. Koeficienti lahko upoštevajo različne dejavnike, kot so zapletenost dela, regionalne značilnosti, stopnja inflacije, morebitna tveganja (pričakovani odstotek splošnih stroškov) ali špekulacije (faktor dodatnega dobička).

Ocenjevalec kot analistik pretvori izkušnje in priporočila mojstra v standardizirane ocene, ki opisujejo gradbene procese prek enot virov v obliki preglednic. V bistvu je naloga ocenjevalca, da s pomočjo parametrov in koeficientov zbira in strukturira informacije, ki prihajajo z gradbišča.

Tako končni strošek na enoto dela (npr. kvadratni ali kubični meter ali ena vgradnja posamezne enote) poleg neposrednih stroškov materiala in dela vključuje tudi pribitke podjetja, splošne stroške, zavarovanje in druge dejavnike (slika 5.1-6)

Hkrati nam ni treba več skrbeti za dejanske cene pri izračunu (recepta), saj se dejanske cene vedno odražajo v "bazi virov" (preglednici sestavin). Na ravni izračunov se v tabelo samodejno naložijo podatki iz baze virov (npr. po kodi predmeta ali njegovem edinstvenem identifikatorju), ki naložijo opis in dejansko ceno na enoto, ta pa se lahko samodejno naloži iz spletnih platform ali spletne trgovine

gradbenih materialov. Ocenjevalec na ravni kalkulacij dela mora delo ali postopek opisati le z atributom "količina virov" in dodatnimi dejavniki.



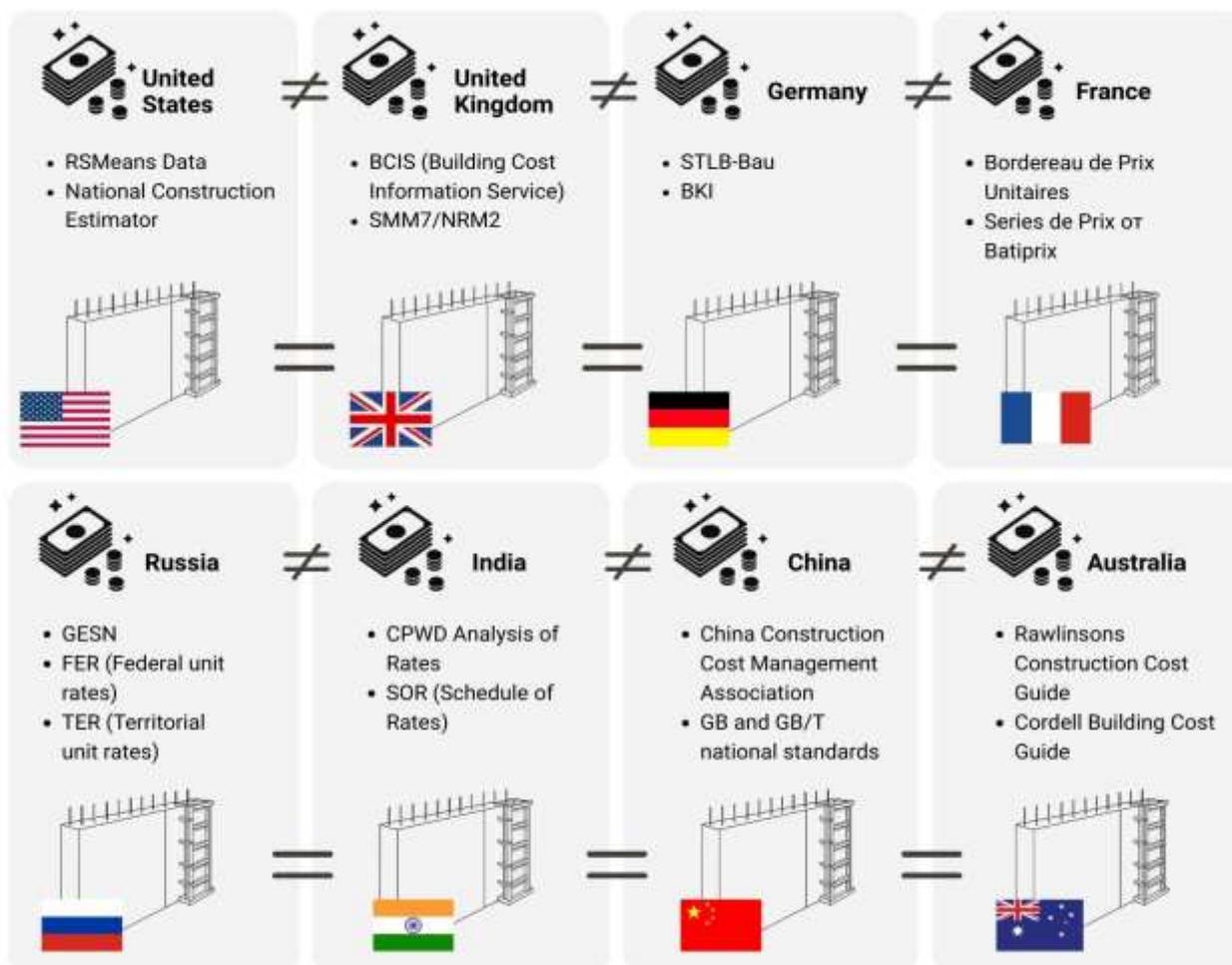
Slika 5.1-6 V fazi izračuna stroškov dela na enoto se izpolnijo le atributi števila potrebnih virov, vse drugo se samodejno naloži iz baze podatkov o virih.

Ustvarjene kalkulacije stroškov dela so shranjene v obliki predlog tabel tipičnih projektov, ki so neposredno povezane s podatkovno bazo gradbenih virov in materialov. Te predloge predstavljajo standardizirane recepte za ponavljajoče se vrste del za prihodnje projekte, kar zagotavlja enotnost izračunov v podjetju.

Ko se stroški katerega koli vira v podatkovni zbirki spremenijo (slika 5.1-3) - ročno ali samodejno s prenosom tekočih tržnih cen (npr. v inflacijskih razmerah) - se posodobitve takoj odražajo v vseh povezanih kalkulacijah stroškov (slika 5.1-6). To pomeni, da je treba spremeniti le osnovno viro, medtem ko predloge za izračun stroškov in ocene ostanejo sčasoma nespremenjene. Ta pristop zagotavlja stabilnost in ponovljivost izračunov za vsa nihanja cen, ki so upoštevana le v razmeroma preprosti tabeli virov (slika 5.1-3).

Za vsak nov projekt se ustvari kopija standardne predloge za izračun stroškov, kar omogoča, da lahko izvedete spremembe in prilagodite dejavnosti za izpolnjevanje posebnih zahtev, ne da bi spremenili prvotno predlogo, ki jo je sprejelo podjetje. Ta pristop zagotavlja prilagodljivost pri prilagajanju izračunov: upoštevate lahko posebnosti gradbišča, želje stranke, uvedete koeficiente tveganja ali donosnosti (špekulacije) - vse to brez kršenja standardov podjetja. To podjetju pomaga najti ravnotesje med maksimiranjem dobička, zadovoljstvom strank in ohranjanjem konkurenčnosti.

V nekaterih državah so takšne predloge za izračun stroškov, ki so se kopičile več desetletij, standardizirane na nacionalni ravni in so postale del standardov nacionalnega sistema za izračun stroškov gradnje (slika 5.1-7).



Slika 5.1-7 Različne države po svetu imajo svoja pravila za obračun stroškov z lastnimi (predpisanimi) zbirkami in standardi za gradbena dela za obračun stroškov istega elementa.

Takšne standardizirane baze ocen virov (slika 5.1-7) morajo obvezno uporabljati vsi udeleženci na trgu zlasti za projekte, ki se financirajo iz javnih sredstev. Takšna standardizacija zagotavlja

preglednost, primerljivost in pravičnost pri oblikovanju cen in pogodbenih obveznosti za naročnika

## **Stroški končnega projekta: od ocen do proračunov**

Državni in panožni standardi ocenjevanja imajo v gradbeni praksi v različnih državah različno vlogo. Medtem ko nekatere države zahtevajo strogo upoštevanje enotnega standarda, večina razvitih gospodarstev uporablja bolj prožen pristop. V tržnih gospodarstvih državni gradbeni standardi običajno služijo le kot izhodišče. Gradbena podjetja te standarde prilagodijo svojim modelom delovanja ali jih popolnoma spremenijo in dopolnijo z lastnimi prilagojenimi dejavniki. Te prilagoditve odražajo izkušnje podjetij, učinkovitost upravljanja virov in pogosto tudi dejavnike, pri katerih se lahko na primer upoštevajo špekulativni dobički podjetja.

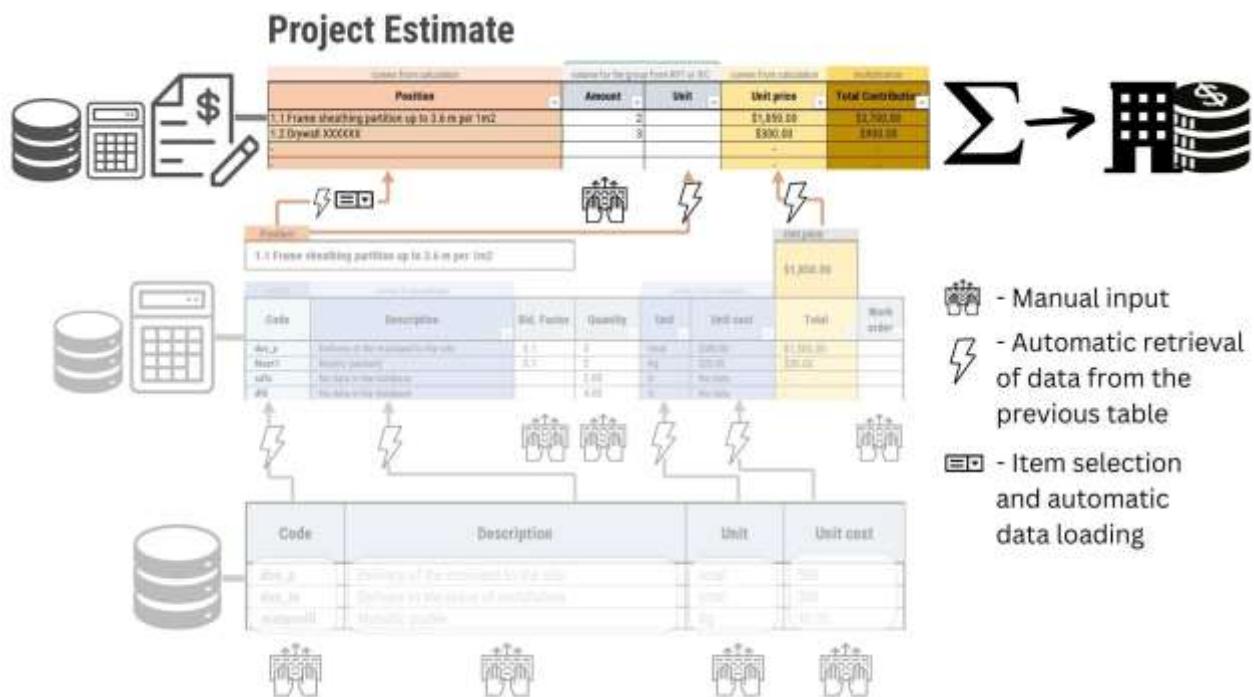
Zato lahko raven konkurence, povpraševanje na trgu, ciljne marže in celo odnosi z določenimi strankami povzročijo znatna odstopanja od standardiziranih norm. Ta praksa zagotavlja prožnost trga, vendar tudi otežuje pregledno primerjavo ponudb različnih izvajalcev in v gradbeništvo na tej stopnji postopka izračuna vnaša element špekulativnega oblikovanja cen.

Ko so pripravljene predloge izračunov za posamezne dejavnosti in procese - ali pa so pogosteje preprosto prekopirane iz standardnih vladnih ocen (slika 5.1-7) z dodanimi koeficienti, ki odražajo "posebnosti" določenega podjetja - je zadnji korak pomnožiti stroške vsake postavke z ustreznim atributom obsega dela ali procesov v novem projektu.

Pri izračunu skupnih stroškov novega gradbenega projekta je ključni korak povzetek stroškov vseh stroškovnih postavk, pomnoženih z obsegom teh postavk dela v projektu.

Za oblikovanje skupnih stroškov projekta v našem poenostavljenem primeru začnemo z izračunom stroškov gradnje enega kvadratnega metra stene in pomnožimo stroške njenega izračuna (npr. delo "1m<sup>2</sup> standardna namestitev stenskih elementov") s skupnim številom kvadratnih metrov sten v projektu (npr. atribut "površina" ali "količina" (slika 5.1-8) entitete tipa "stenski elementi" iz CAD projekta ali izračuni mojstra).

Stroške za vse elemente projekta izračunamo na enak način (slika 5.1-8): vzamemo stroške na enoto dela in jih pomnožimo z obsegom določenega elementa ali skupine elementov v projektu. Ocjenjevalec mora na spletni strani v obliki obsega ali količine vnesti le število teh elementov, dejavnosti ali procesov v projektu. To omogoča samodejno izdelavo popolnega gradbenega predračuna.



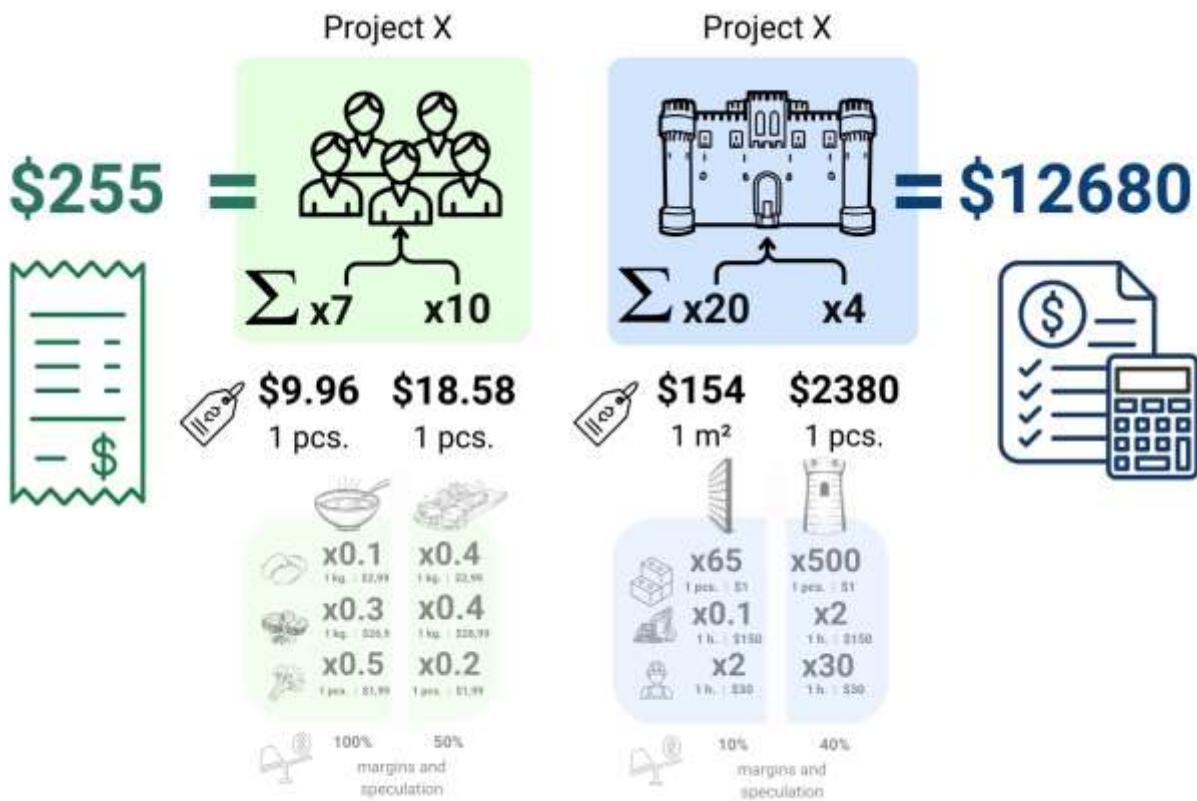
Slika: 5.1-8 V fazi izdelave predračuna vnesemo le obseg dela.

Tako kot pri kalkulacijah tudi na tej ravni samodejno naložimo pripravljene kalkulativne elemente (iz predloge za kalkulacijo ali nove, kopirane iz predloge in urejene), ki s seboj samodejno prinesejo trenutne stroške na enoto dela (ki se samodejno posodabljajo iz podatkovne zbirke virov (slika 5.1-8 spodnja tabela)). V skladu s tem se v primeru kakršne koli spremembe podatkov v zbirki podatkov o virih ali tabelah za izračun stroškov - podatki v predračunu samodejno posodobijo za tekoči dan, ne da bi bilo treba spremenjati kalkulacijo stroškov ali sam predračun.

V restavraciji se končni strošek dogodka izračuna na podoben način in je enak končnemu strošku celotne večerje, pri čemer se strošek posamezne jedi, pomnožen s številom gostov, sešteje v skupni strošek čeka (slika 5.1-9). Tako kot v gradbeništvu se tudi recepti za kuhanje v restavracji morda ne bodo spremenili več desetletij. Za razliko od cen, kjer se lahko stroški sestavin spreminja vsako uro.

Podobno kot lastnik restavracije pomnoži stroške posameznega obroka s številom obrokov in ljudi, da bi določil skupne stroške dogodka, tudi vodja ocenjevanja stroškov sešteje stroške vseh sestavnih delov projekta, da dobi popolno gradbeno oceno.

Tako se za vsako dejavnost v projektu določi njen končni strošek (slika 5.1-9), ki se pomnoži z obsegom atributa entitete, ki ustreza tej dejavnosti - dobimo strošek skupin dejavnosti, iz katerega dobimo končni strošek celotnega projekta.



Slika 5.1-9 Končna ocena se izračuna tako, da se atribut stroškov dela vsakega elementa sešteje z njegovim atributom obsega.

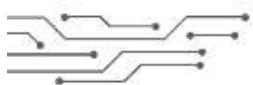
Skupni stroški projekta (slika 5.1-8) zagotavljajo finančno sliko projekta, ki strankam, vlagateljem ali finančnim organizacijam omogoča razumevanje celotnega proračuna in finančnih virov, potrebnih za izvedbo projekta na kateri koli dan, ob upoštevanju trenutnih cen.

In če so procesi sestavljanja baz virov, izračunov in ocen (procesni recepti) že izdelani, polavtomatizirani in izpopolnjeni v desettisočih letih ter zabeleženi na državni ravni, potem samodejno pridobivanje kakovostnih informacij o obsegu in količini elementov za zadnjo fazo končne ocene - danes ostaja ozko grlo v procesih vseh izračunov stroškovnih in časovnih atributov projekta ter na splošno celotnega proračuna projekta.

Tisočletja je bila tradicionalna metoda izračuna prostornine ročna metoda merjenja prostornine in količine z uporabo ravnih risb. S prihodom digitalne dobe so podjetja odkrila, da je zdaj mogoče informacije o prostornini in količini samodejno pridobiti iz geometrijskih podatkov, ki jih vsebujejo modeli CAD, kar je revolucionarno spremenilo tisočletja stare načine pridobivanja količinskih podatkov.

Sodobni pristopi k ocenjevanju in vrednotenju procesov vključujejo samodejno pridobivanje volumetričnih in kvantitativnih atributov iz podatkovnih zbirk CAD, ki jih je mogoče naložiti in povezati s procesom obračunavanja stroškov, da se pridobijo posodobljeni volumni projektne skupine v kateri

koli faz od načrtovanja do delovanja.



## POGLAVJE 5.2.

# KOLIČINSKI IZPIS TER SAMODEJNO OBLIKOVANJE PREDRAČUNOV IN RAZPOREDOV

### Prehod iz 3D in 4D in 5D: uporaba volumetričnih in kvantitativnih parametrov

Ko imate v rokah preglednice za izračun stroškov z opisanimi procesi prek virov (slika 5.1-8), je naslednji korak samodejna pridobitev parametrov obsega ali količine za skupino elementov, ki so potrebni za izračune in končno oceno.

Prostorninske značilnosti elementov projekta - npr. sten ali plošč - je mogoče samodejno pridobiti iz podatkovnih zbirk CAD. Parametrični objekti, ustvarjeni v programih CAD, se s pomočjo geometrijskega jedra pretvorijo v numerične vrednosti dolžine, širine, površine, prostornine in drugih parametrov. Postopek pridobivanja volumnov na podlagi 3D geometrije bo podrobnejše obravnavan v naslednjem, šestem delu (slika 6.3-3), ki je namenjen delu s CAD (BIM). Poleg volumnov je mogoče iz podatkovne zbirke CAD-modelov pridobiti tudi število podobnih elementov, in sicer s filtriranjem in grupiranjem objektov po kategorijah in lastnostih. Ti parametri, ki omogočajo združevanje v skupine, postanejo osnova za povezovanje elementov projekta prek izračunov virov z izračuni, končnimi ocenami in proračunom celotnega projekta.

Tako je podatkovni model, pridobljen iz 3D modela (CAD), dopolnjen z novimi plastmi parametrov, označenimi kot 4D in 5D. V novih slojih atributov entitet, 4D (čas) in 5D (stroški), se 3D geometrijski podatki uporabljajo kot vir vrednosti atributov obsega entitet.

- **4D** je sloj informacij o parametrih, ki 3D parametrom elementov dodaja informacije o trajanju gradbenih operacij. Ti podatki so bistveni za načrtovanje časovnih razporedov in upravljanje časovnice projekta
- **5D** je naslednja stopnja razširitve podatkovnega modela, pri kateri so elementi dopolnjeni s stroškovnimi značilnostmi. S tem se geometrijskim podatkom doda finančni vidik: stroški materiala, del in opreme, kar omogoča proračunske izračune, analize donosnosti in upravljanje stroškov med postopkom gradnje.

Stroški in atributni podatki 3D, 4D in 5D skupin projektnih entitet so opisani na podoben način kot izračuni v modularnih sistemih ERP, PIMS (ali Excelu podobnih orodjih) in se uporabljajo za samodejno izračunavanje stroškov in načrtovanje proračuna posameznih skupin in celotnega proračuna projekta.

### 5D atributi in pridobivanje volumnov atributov iz CAD

Pri pripravi končnega predračuna gradbenega projekta, katerega pripravo smo obravnavali v prejšnjih poglavjih (slika 5.1-8), se atributi prostornine za vsako kategorijo elementov projekta zbirajo ročno ali pa se pridobijo iz specifikacij atributov prostornine, ki jih zagotavlja programska oprema CAD.

Tradicionalna ročna metoda izračunavanja količin vključuje analizo risb, ki so bile tisočletja predstavljene kot črte na papirju, zadnjih 30 let pa v digitalnih formatih, kot so PDF (PLT) ali DWG. Na podlagi strokovnih izkušenj izmerita količine potrebnih del in materialov, pogosto z ravnalom in kotomerom. Ta metoda zahteva veliko truda in časa ter posebno pozornost do podrobnosti.

Določanje atributov obsega na ta način lahko traja od nekaj dni do več mesecev, odvisno od obsega projekta. Poleg tega, ker se vse meritve in izračuni izvajajo ročno, obstaja tveganje človeške napake, ki lahko privede do netočnih podatkov, kar posledično vpliva na napake pri ocenjevanju časa in stroškov projekta, za katere bo odgovorno celotno podjetje.

Sodobne metode, ki temeljijo na uporabi podatkovnih baz CAD, močno poenostavijo izračun volumnov. V modelih CAD geometrija elementov že vključuje atribute prostornine, ki jih je mogoče samodejno izračunati (prek geometrijskega jedra (slika 6.3-3)) in predstaviti ali izvoziti v tabelarični obliki.

V takem primeru oddelek za ocenjevanje od projektanta CAD zahteva podatke o količinskih in prostorninskih značilnostih elementov projekta. Ti podatki se izvozijo v obliki preglednic ali neposredno vključijo v podatkovne zbirke za izračun stroškov - bodisi v sisteme Excel, ERP ali PMIS. Ta postopek se pogosto ne začne z uradno zahtevo, temveč s kratkim dialogom med naročnikom (pobudnikom) ter arhitektom ocenjevalcem iz gradbenega ali projektantskega podjetja. Spodaj je poenostavljen primer, ki prikazuje, kako iz vsakdanje komunikacije nastane strukturirana tabela za samodejne izračune (QTO):

- ☛ Stranka - "Stavbi želim dodati še eno nadstropje v enaki konfiguraciji kot prvo nadstropje"
- ☛ Arhitekt (CAD) - "Če dodamo tretje nadstropje, je konfiguracija enaka kot v drugem nadstropju". In po tem sporočilu pošlje novo različico projekta CAD ocenjevalcu.
- ☛ Estimator samodejno izvede združevanje v skupine in izračune (ERP, PMIS, Excel) - "Projekt poženem skozi Excelovo preglednico s pravili QTO (ERP, PMIS), pridobim količine po kategorijah za novo nadstropje in ustvarim predračun"

Tako se besedilni dialog pretvori v tabelarično strukturo s pravili za razvrščanje v skupine:

Element	Kategorija	Tla
Prekrivanje	OST_Floors	3
Stolpec	OST_StructuralColumns	3
stopnišče	OST_stropi	3

Po postopku samodejnega združevanja modela CAD od oblikovalca v skupine v skladu s pravili QTO ocenjevalca in samodejnega množenja količin z izračuni virov (slika 5.1-8) dobimo naslednje rezultate, ki jih pošljemo stranki:

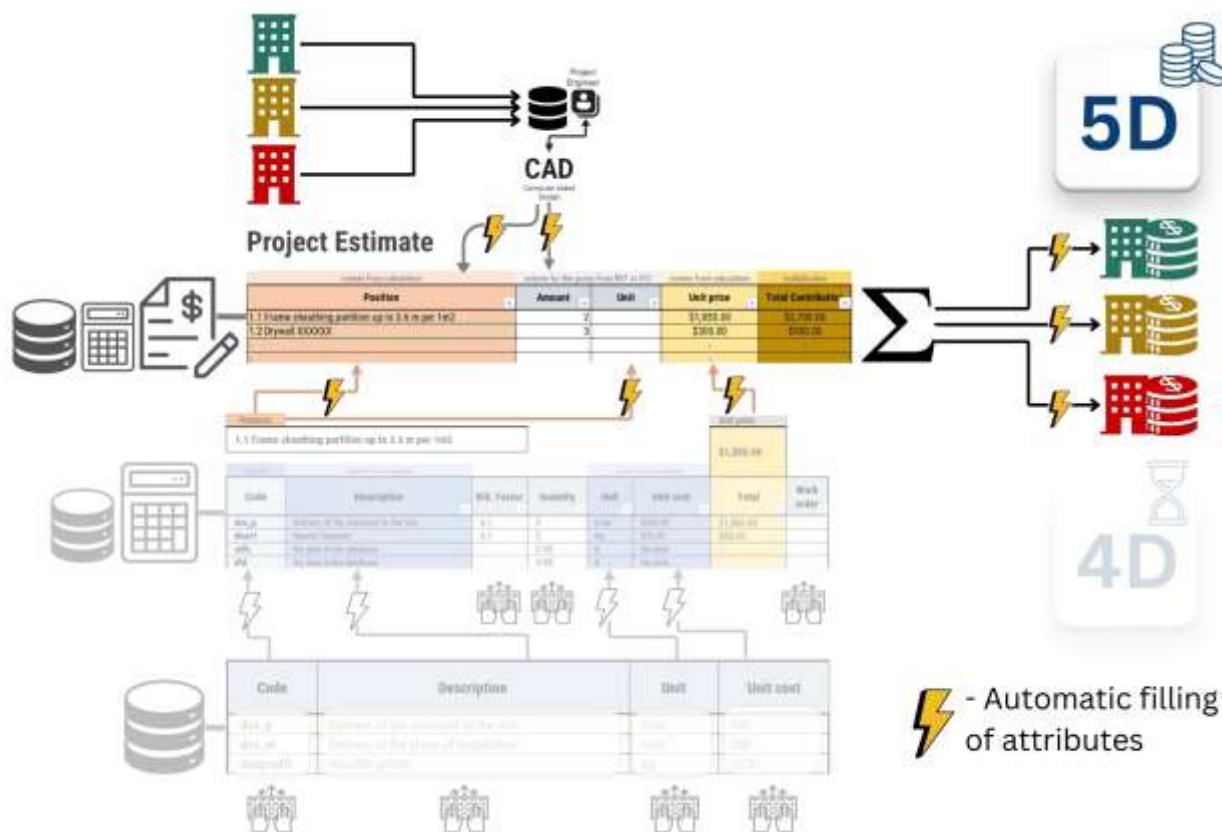
Element	Zvezek	Tla	Cena na enoto.	Skupni stroški
Prekrivanje	420 m <sup>2</sup>	3	150 €/m <sup>2</sup>	<b>63 000 €</b>
Stolpec	4 kosi.	3	2450 €/kos.	<b>9 800 €</b>
stopnišče	2 kosa.	3	4.300 €/kos.	<b>8 600 €</b>
<b>SKUPAJ:</b>	-	-	-	<b>81 400 €</b>

❸ Stranka - "Hvala, to je precej veliko, nekaj sob moramo zmanjšati". In cikel se ponovi.

Ta scenarij se lahko večkrat ponovi, zlasti v fazi odobritve, ko stranka pričakuje takojšnje povratne informacije. V praksi pa se takšni postopki lahko zavlečajo za več dni ali celo tednov. Danes je treba zaradi uvedbe samodejnega združevanja v skupine in pravil za izračun dejavnosti, ki so včasih zahtevale veliko časa, opraviti v nekaj minutah. Avtomatizirano pridobivanje količin s pomočjo pravil združevanja ne le pospešuje izračune in ocene, temveč z zmanjšanjem človeškega dejavnika zmanjuje verjetnost napak ter zagotavlja pregledno in natančno oceno stroškov projekta.

Če so bile zahteve oddelka za ocenjevanje prvotno upoštevane pri ustvarjanju 3D modela v sistemu CAD (kar je v praksi še vedno redko) in so imena, identifikatorji skupin elementov in njihovi klasifikacijski atributi določeni v obliki parametrov, ki se ujemajo s strukturami skupin in razredov za ocenjevanje, se lahko volumetrični atributi samodejno prenesejo v sisteme za ocenjevanje brez dodatnih pretvorb.

Samodejno pridobivanje volumetričnih atributov iz CAD v obliki specifikacijskih tabel omogoča hitro pridobivanje posodobljenih podatkov o stroških posameznih del in projekta kot celote (slika 5.2-1). S posodobitvijo samo datoteke CAD z volumni projekta v postopku izračuna ali v kalkulacijskem sistemu lahko podjetje hitro preračuna predračun ob upoštevanju najnovejših sprememb, kar zagotavlja visoko natančnost in doslednost vseh nadaljnjih izračunov.



Slika 5.2-1 Atributi obsega iz tabel CAD ali podatkovnih zbirk se samodejno vnesejo v oceno, kar omogoča takojšen izračun skupnih stroškov projekta.

Zaradi vse večje kompleksnosti investicijskih projektov postaja izračun celotnega proračuna in analiza skupnih stroškov projektov po takem scenariju (slika 5.2-1) - ključno orodje za informirano sprejemanje odločitev.

Po podatkih študije Creating More Value through Capital Projects (2024) [20], ki jo je izvedla družba Accenture, vodilna podjetja aktivno vključujejo podatkovno analitiko v digitalne pobude ter uporabljajo pretekle informacije za napovedovanje in optimizacijo rezultatov. Raziskave kažejo, da vse več lastnikov-upravljavcev uporablja analitiko velikih količin podatkov za napovedovanje tržnih trendov in ocenjevanje komercialne upravičenosti pred začetkom projektiranja. To dosežejo z analizo podatkovnih skladišč iz obstoječega portfelja projektov. Poleg tega 79 odstotkov lastnikov-upravljavcev izvaja "robustno" napovedno analitiko za oceno uspešnosti projektov in podporo operativnemu odločanju v realnem času.

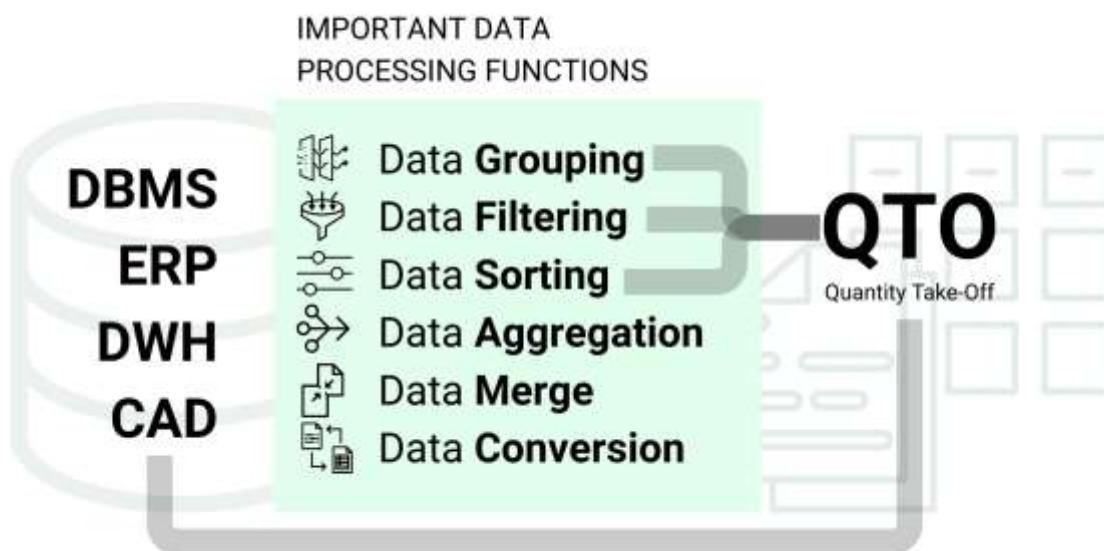
Sodobno učinkovito upravljanje gradbenih projektov je neločljivo povezano z obdelavo in analizo velikih količin informacij v vseh fazah projektiranja in tistih procesov, ki so pred projektiranjem. Uporaba podatkovnih skladišč, izračunov virov, napovednih modelov in strojnega učenja omogoča ne le zmanjšanje tveganj pri izračunih, temveč tudi sprejemanje strateških odločitev o financiranju projekta v zgodnjih fazah projektiranja. Več o podatkovnih skladiščih in napovednih modelih, ki dopolnjujejo izračune, bomo govorili v devetem delu knjige.

Samodejno pridobivanje volumenskih parametrov elementov iz projektov CAD, ki so potrebni za pripravo ocen, poteka s pomočjo orodij za združevanje QTO (Quantity Take-Off). Orodja QTO delujejo tako, da združujejo vse objekte projekta po posebnih identifikatorjih elementov ali parametrih atributov elementov s pomočjo specifikacij in tabel, ustvarjenih v podatkovni zbirki CAD.

## **QTO Količinski prevzem: združevanje podatkov o projektu po atributih**

QTO (Quantity Take-Off) v gradbeništvu je postopek določanja količin elementov, potrebnih za izvedbo projekta. V praksi je QTO pogosto delno ročni postopek, ki vključuje zbiranje podatkov iz različnih virov: dokumentov PDF, risb DWG in digitalnih modelov CAD.

Pri delu s podatki, pridobljenimi iz podatkovnih zbirk CAD, se postopek QTO izvaja kot zaporedje operacij filtriranja, razvrščanja, združevanja in združevanja. Elementi modela so izbrani glede na parametre razredov, kategorij in tipov, nato pa so njihovi kvantitativni atributi - kot so prostornina, površina, dolžina ali količina - povzeti v skladu z logiko izračuna (slika 5.2-2).



Slika 5.2-2 Združevanje in filtriranje podatkov sta najbolj priljubljeni funkciji, ki se uporabljalata v podatkovnih zbirkah in podatkovnih skladiščih.

Postopek QTO (filtriranje in združevanje) omogoča sistematizacijo podatkov, oblikovanje specifikacij in pripravo vhodnih informacij za izračun ocen, nakupov in razporedov dela. Osnova QTO je razvrščanje elementov glede na vrsto merjenih lastnosti. Za vsak element ali skupino elementov se izbere ustrezni kvantitativni merski parameter. Na primer:

- **Atribut dolžine** (robnik - v metrih)
- **Pripis površine** (delo z mavčnimi ploščami - v kvadratnih metrih)
- **Pripis količine** (betonska dela - v kubičnih metrih)
- **Atribut količine** (okna - na kos)

Poleg prostorninskih karakteristik, ki se matematično oblikujejo na podlagi geometrije, se pri izračunih po grupiranju QTO pogosto uporablajo faktorji prekoračitve (slika 5.2-12, npr. 1,1, da se upošteva 10 odstotkov za logistiko in namestitev) - koreksijske vrednosti, ki upoštevajo izgube, namestitev, skladiščenje ali prevoz. Tako je mogoče natančneje predvideti dejansko porabo materialov in se izogniti tako pomanjkanju kot prevelikim zalogam na gradbišču.

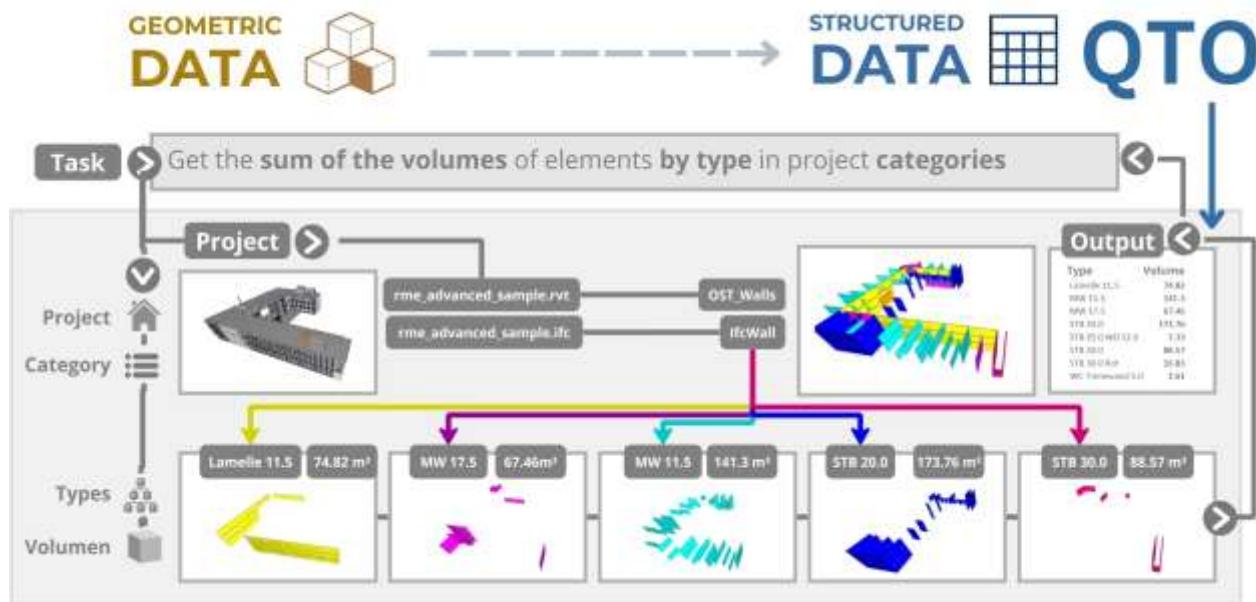
Avtomatiziran proces količinskega prevzema (QTO) je bistvenega pomena za izdelavo natančnih izračunov in ocen, zmanjšanje človeških napak v procesih specifikacij količin in preprečevanje prevelikega ali premajhnega naročanja materialov.

Kot primer postopka QTO si oglejmo pogost primer, ko je treba iz podatkovne baze CAD prikazati tabelo s specifikacijo prostornin po vrstah elementov za določeno kategorijo, razrede elementov. Združimo vse elemente projekta po tipih iz kategorije stene projekta CAD in povzamemo atributе volumnov za vsak tip ter rezultat predstavimo kot tabelo volumnov QTO (slika 5.2-3).

V primeru tipičnega projekta CAD (slika 5.2-3) so vsi elementi kategorije sten v podatkovni zbirki CAD združeni po tipu sten, npr. "Lamelle 11.5", "MW 11.5" in "STB 20.0", in imajo natančno opredeljene atributе prostornine, predstavljene v metričnih kockah.

Cilj upravitelja, ki je vmesnik med projektanti in strokovnjaki za izračune, je pridobiti avtomatizirano preglednico količin po vrstah elementov v izbrani kategoriji. Ne le za določen projekt, temveč tudi v univerzalni obliki, ki se uporablja za druge projekte s podobno strukturo modela. To omogoča razširljivost pristopa in ponovno uporabo podatkov brez podvajanja naporov.

Minili so časi, ko so se izkušeni projektanti in ocenjevalci oborožili z ravnilom in skrbno izmerili vsako črto na papirju ali načrtih PDF - tradicija, ki se v zadnjih tisočletjih ni spremenila. Z razvojem 3D -modeliranja, kjer je geometrija vsakega elementa zdaj neposredno povezana s samodejno izračunanimi prostorninskimi atributi, je postopek določanja prostornin in količin QTO postal avtomatiziran.



Slika 5.2-3 Pridobivanje atributov obsega in količine QTO iz projekta vključuje združevanje in filtriranje elementov projekta.

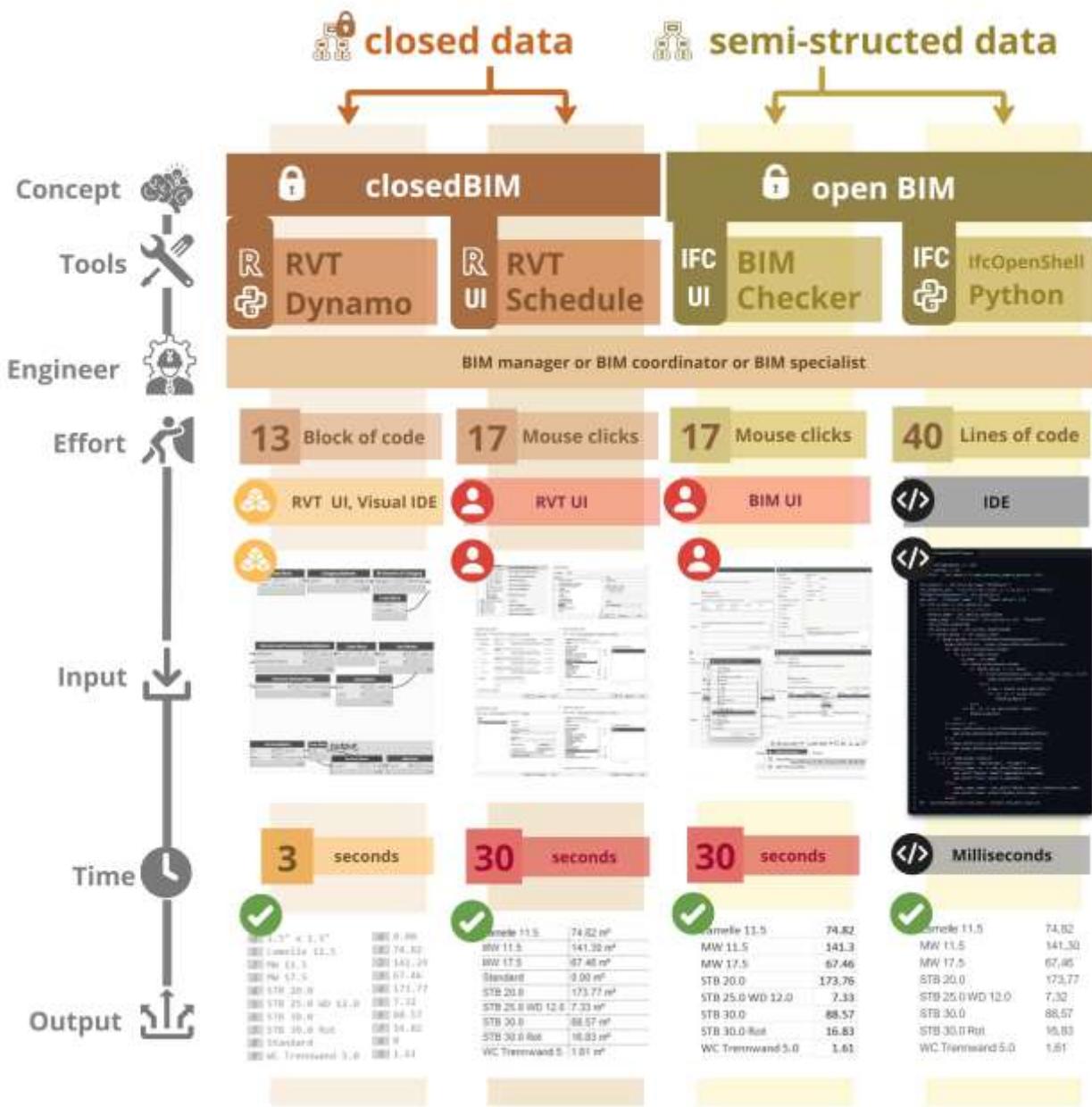
V našem primeru je naloga "izbrati kategorijo zidov v projektu, združiti vse elemente po vrsti in predstaviti informacije o atributih obsega v strukturirani, tabelarični obliki", tako da lahko to tabelo uporabi na desetine drugih strokovnjakov za izračune stroškov, logistiko, urnike dela in druge poslovne primere (slika 6.1-3).

Zaradi zaprte narave podatkov CAD danes ne more vsak strokovnjak uporabljati neposrednega dostopa do baze podatkov CAD (razlogi in rešitve problema dostopa so podrobno opisani v šestem delu knjige). Zato se morajo mnogi obrniti na specializirana orodja BIM, ki temeljijo na konceptih odprtrega BIM in zaprtega BIM [63]. Pri delu s specializiranimi orodji BIM ali neposredno v programskem okolju CAD se lahko tabela z rezultati QTO (Quantity Take-Off) ustvari na različne načine - odvisno od tega, ali se uporablja ročni vmesnik ali avtomatizacija programske opreme.

Na primer, z uporabniškim vmesnikom programske opreme CAD (BIM) je dovolj opraviti približno 17 dejanj (klikov na gume), da dobimo pripravljeno tabelo prostornin (slika 5.2-4). Vendar mora uporabnik dobro poznati strukturo modela in funkcije programske opreme CAD (BIM).

Če se avtomatizacija uporablja prek programske kode ali vtičnikov in orodij API v programih CAD, se zmanjša število ročnih korakov za pridobitev prostorninskih tabel, vendar je treba napisati od 40 do 150 vrstic kode, odvisno od uporabljeni knjižnice ali orodja:

- **IfcOpSh (odpri ti BIM)** ali **Dynamo IronPython (zapri ti BIM)** - omogočata pridobitev tabele QTO iz formata CAD ali programa CAD v samo ~40 vrsticah kode.
- **IFC\_js (odpri ti BIM)** - za pridobivanje obsežnih atributov iz modela IFC je potrebnih približno 150 vrstic kode.
- **Orodja vmesnika CAD (BIM)** - omogočajo, da enak rezultat dosežete ročno, s 17 kliki z miško.



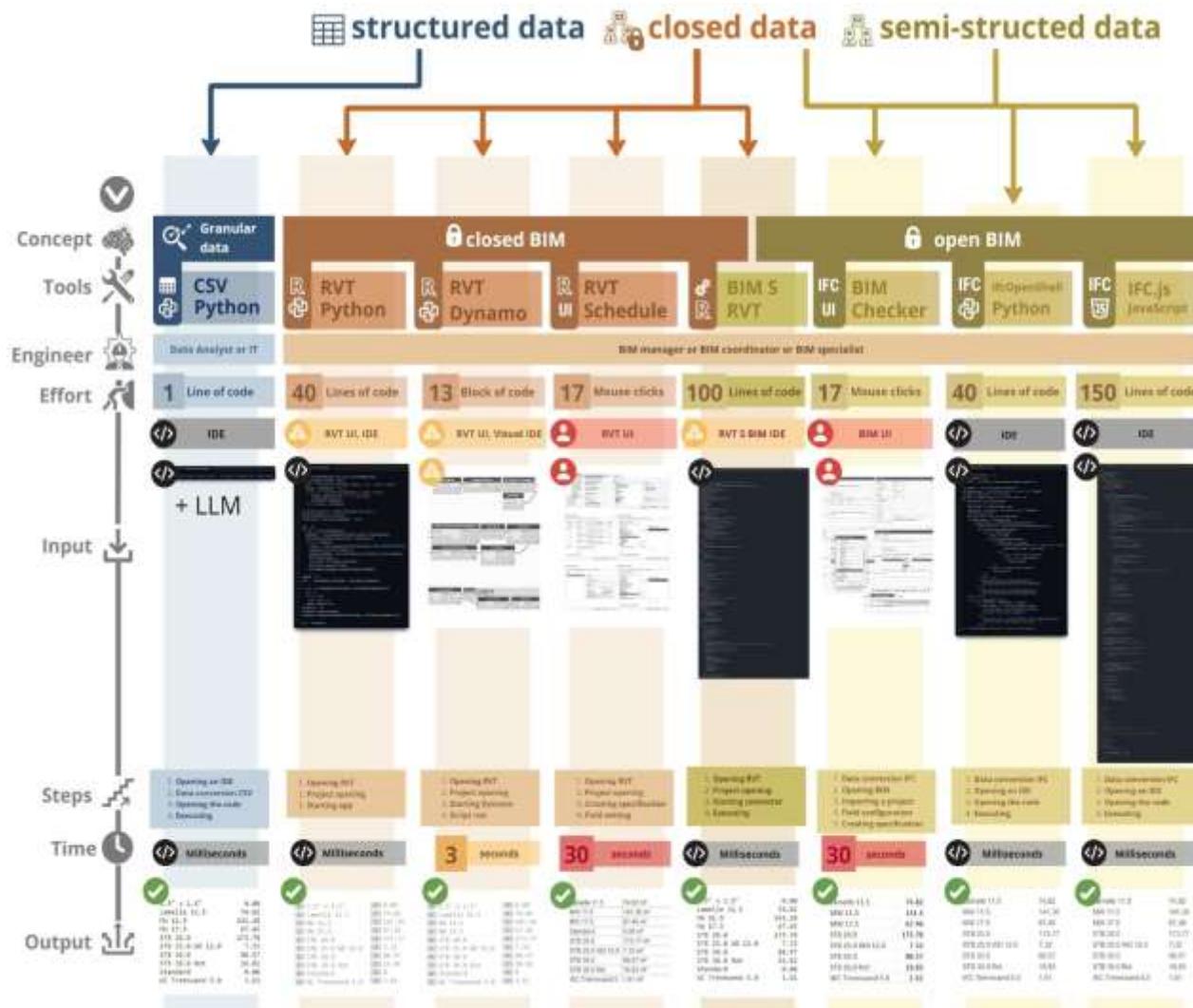
Slika 5.2-4 Oblikovalci in upravljevalci CAD (BIM) za izdelavo tabel QTO uporabijo od 40 do 150 vrstic kode ali ducat pritiskov tipk.

Rezultat je enak - strukturirana tabela z atributi obsega za skupino elementov. Edina razlika je v stroških dela in stopnji tehničnega znanja, ki ga potrebuje uporabnik (slika 5.2-4). Sodobna orodja v primerjavi z ročnim zbiranjem količin znatno pospešijo postopek QTO in zmanjšajo verjetnost napak. Omogočajo, da se podatki pridobijo neposredno iz modela projekta, kar odpravlja potrebo po ročnem preračunavanju volumnov iz risb, kot se je to počelo v preteklosti.

Ne glede na uporabljen metodo - bodisi odprt BIM bodisi zaprt BIM - je mogoče dobiti identično QTO - tabelo z obsegom elementov projekta (slika 5.2-4). Pri delu s projektnimi podatki v konceptnih CAD - (BIM-

) pa so uporabniki odvisni od specializiranih orodij in vmesnikov API, ki jih zagotavljajo ponudniki (slika 3.2-13). To ustvarja dodatne plasti odvisnosti in zahteva učenje edinstvenih podatkovnih shem, hkrati pa omejuje neposreden dostop do podatkov.

Zaradi zaprte narave podatkov CAD pridobivanje tabel QTO in drugih parametrov otežuje avtomatizacijo izračunov in integracijo z zunanjimi sistemmi. Z uporabo orodij za neposreden dostop do podatkovnih zbirk in prenosom podatkov CAD projekta z orodji za povratno inženirstvo v odprto obliko strukturiranega podatkovnega okvira (slika 4.1-13) lahko identično tabelo QTO pridobimo z eno samo vrstico kode (slika 5.2-5 - različica z granularnimi podatki).



Slika 5.2-5 Različna orodja dajejo enake rezultate v obliki atributnih tabel projektnih entitet, vendar z različnimi stroški dela.

Pri uporabi odprtih strukturiranih podatkov iz projektov CAD, kot je navedeno v poglavju "Prevarjanje podatkov CAD (BIM) v strukturirano obliko", je postopek razvrščanja v skupine, QTO, zelo poenostavljen.

Pristopi, ki temeljijo na uporabi odprtih strukturiranih podatkov ali neposrednem dostopu do

podatkovnih zbirk modelov CAD, so prosti tržnih omejitev, povezanih z akronimom BIM. Zanašajo se na preizkušena orodja, ki se že dolgo uporabljajo v drugih panogah (slika 7.3-10 Postopek ETL).

Glede na McKinseyjevo študijo "Odprti podatki: Sprostitev inovacij in produktivnosti s pretočnimi informacijami". [102], ki je bila izvedena leta 2013, bi lahko uporaba odprtih podatkov ustvarila priložnosti za prihranke v višini od 30 do 50 milijard dolarjev na leto pri načrtovanju, inženiringu, naročanju in gradnji objektov za proizvodnjo električne energije. To pomeni 15-odstotne prihranke pri investicijskih stroških gradnje.

Delo z odprtimi strukturiranimi (granularnimi) podatki poenostavlja iskanje in obdelavo informacij, zmanjšuje odvisnost od specializiranih platform BIM in odpira pot k avtomatizaciji brez potrebe po uporabi lastniških sistemov ali parametričnih in kompleksnih podatkovnih modelov iz formatov CAD.

## Avtomatizacija QTO z uporabo LLM in strukturiranih podatkov

Prevajanje nestrukturiranih podatkov v strukturirano obliko bistveno izboljša učinkovitost različnih procesov: poenostavi obdelavo podatkov (slika 4.1-1, slika 4.1-2) in pospeši postopek potrjevanja, saj zahteve postanejo jasne in pregledne, kot smo že obravnavali v prejšnjih poglavjih. Podobno prevod podatkov CAD (BIM) v strukturirano odprto obliko (slika 4.1-12, slika 4.1-13) olajša postopek razvrščanja atributov v skupine in postopek QTO.

Tabela atributov QTO ima strukturirano obliko, zato pri uporabi strukturiranih podatkov CAD delamo z enotnim podatkovnim modelom (slika 5.2-5), zaradi česar ni treba pretvarjati in prevajati podatkovnih modelov projektov in pravil za razvrščanje v skupine na skupni imenovalec. To nam omogoča združevanje podatkov po enem ali več atributih z eno samo vrstico kode. V nasprotju s tem je pri odprttem BIM in zaprtem BIM, kjer so podatki shranjeni v polstrukturiranih, parametričnih ali zaprtih formatih, za obdelavo potrebnih na desetine ali celo stotine vrstic kode ter uporaba vmesnika API za interakcijo z informacijami o geometriji in atributih.

- ❷ Primer združevanja strukturiranega projekta QTO po enem atributu. Besedilna poizvedba v kateri koli klepetalnici LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ali kateri koli drugi):

Imam CAD -projekt kot podatkovni okvir - filtrirajte podatke o projektu, da dobite elemente s parametrom "Type", ki vsebuje samo vrednost "Type 1" ↴

- Odgovor LLM bo zelo verjetno v obliki kode Python z uporabo Pandas :

```
1 df[df['Type'].str.contains("Type 1")]
```

**Slika 5.2-6** Ena vrstica kode, napisana z LLM , omogoča, da celoten projekt CAD združite v skupine po atributu "Type" in dobite želeno skupino elementov.

Zaradi preproste strukture dvodimenzionalnega podatkovnega okvira nam ni treba razlagati sheme LLM in podatkovnega modela, kar skrajša korake razlage in pospeši oblikovanje končnih rešitev. Prej je bilo za pisanje celo preproste kode potrebno učenje programskih jezikov, zdaj pa nam sodobni jezikovni modeli (LLM) omogočajo, da pri delu s strukturiranimi podatki s pomočjo besedilnih poizvedb samodejno prevedemo procesno logiko v kodo.

Avtomatizacija LLM in jezikovni modeli lahko strokovnjakom, ki se ukvarjajo z združevanjem in obdelavo podatkov CAD (BIM), popolnoma odpravijo potrebo po učenju programskih jezikov ali orodij BIM, saj omogočajo reševanje težav z uporabo besedilnih poizvedb.

Ista poizvedba - združevanje vseh elementov projekta iz kategorije "stene" in izračun volumnov za vsako vrsto (slika 5.2-5) - ki v okolju CAD (BIM) zahteva 17 klikov v vmesniku ali pisanje 40 vrstic kode, je v odprtih orodjih za obdelavo podatkov (npr. SQL ali Pandas) videti kot preprosta in intuitivna poizvedba:

- Z eno samo vrstico v programu Pandas:

```
df[df['Kategorija'].isin(['OST_Walls'])].groupby('Vrsta')['Obseg'].sum()
```

Dekodiranje kode: iz df (DataFrame) vzemite elemente, katerih atribut-stolpec "Category" ima vrednosti "OST\_Walls", združite vse dobljene elemente po atributu-stolpcu "Type" in za dobljeno skupino elementov seštejte atribut "Volume".

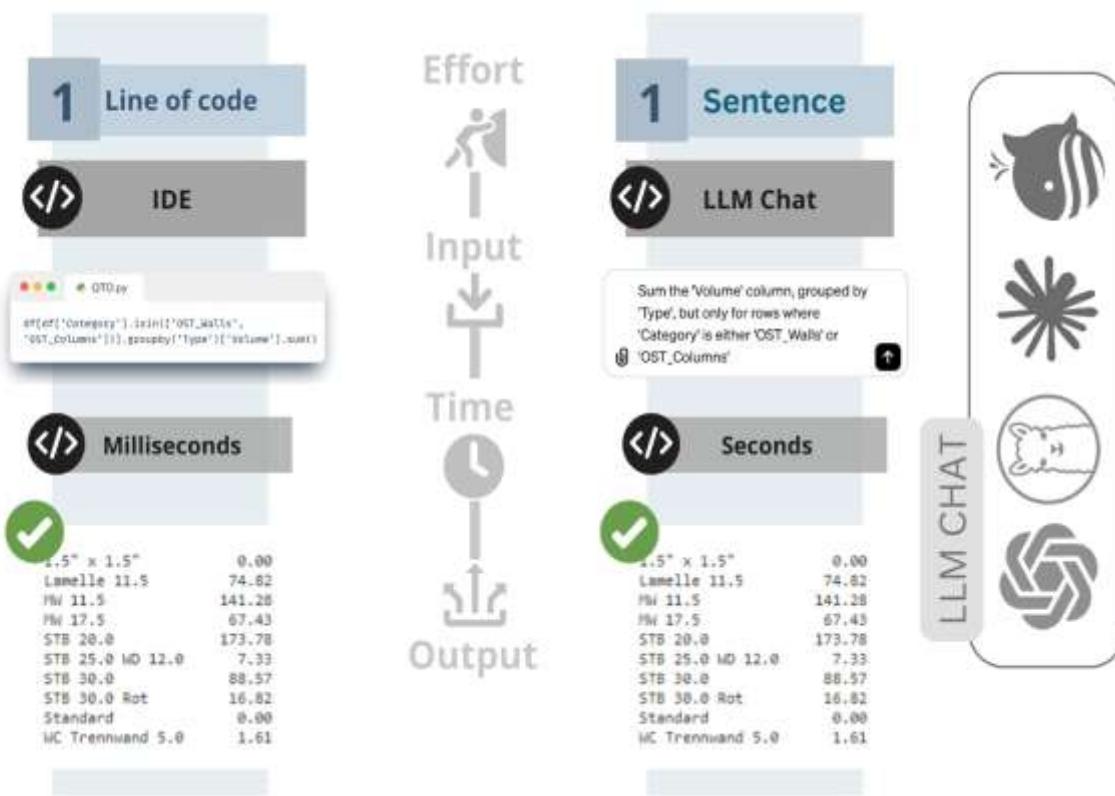
- Združevanje strukturiranega projekta, pridobljenega iz CAD, z uporabo SQL:

```
SELECT Vrsta, SUM(Obseg) AS TotalVolume
IZ elementov
Kjer je kategorija = 'OST_Walls'
GROUP BY Type;
```

- S pomočjo programa LLM lahko zahtevo za združevanje v podatkovno zbirko projekta zapišemo kot preprost besedilni sklic - poziv (slika 5.2-7):

Za podatkovni okvir projekta združite elemente po parametru "Type" (Vrsta), vendar samo za elemente s parametrom "Category" (Kategorija), ki je enak "OST\_Walls" (Stene) ali "OST\_Columns" (Stolpci), in povzemite parameter stolpca "Volume" (Obseg) za nastalo. ↵

## PYTHON PANDAS

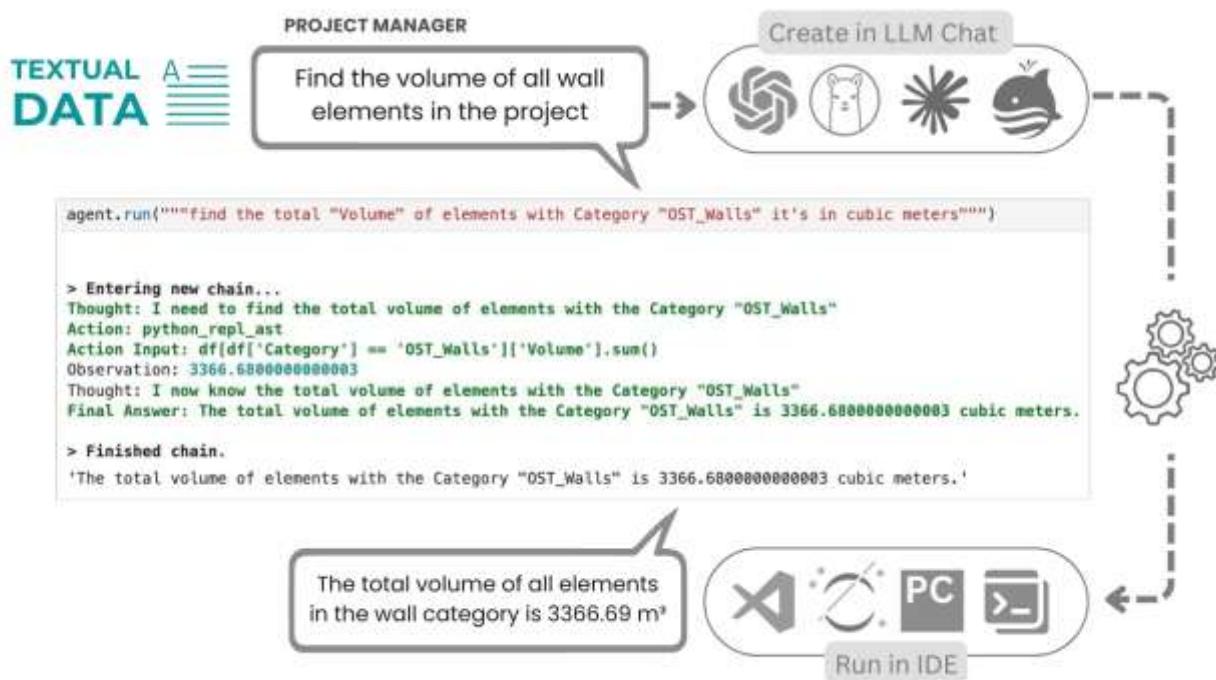


Slika 5.2-7 Z uporabo SQL, Pandas in LLM je zdaj mogoče avtomatizirati obdelavo podatkov z nekaj vrsticami kode in besedilnimi poizvedvabmi.

Pridobivanje QTO iz podatkov CAD z orodji LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) dramatično spreminja tradicionalne metode pridobivanja atributnih informacij, kvantitativnih in volumetričnih podatkov za posamezne objekte in skupine objektov.

Zdaj lahko tudi vodje projektov, strokovnjaki za izračun stroškov ali logistiko, ki nimajo poglobljenega znanja o projektiranju in nimajo specializirane programske opreme CAD - (BIM-), ki imajo dostop do podatkovne baze CAD, v nekaj sekundah dobijo skupno količino elementov kategorije sten ali drugih objektov, tako da preprosto napišejo ali narekujejo poizvedbo.

Pri besedilnih poizvedbah (slika 5.2-8) agent LLM modela obdela uporabnikovo zahtevo uporabi določene funkcije za enega ali več parametrov - stolpcov tabele. Posledično uporabnik v komunikaciji z LLM prejme bodisi nov stolpec-parameter z novimi vrednostmi bodisi eno določeno vrednost po grupiranju



Slika 5.2-8 LLM model, ki dela s strukturiranimi podatki, iz konteksta besedilne poizvedbe razbere, po katerih skupinah in atributih sprašuje uporabnik.

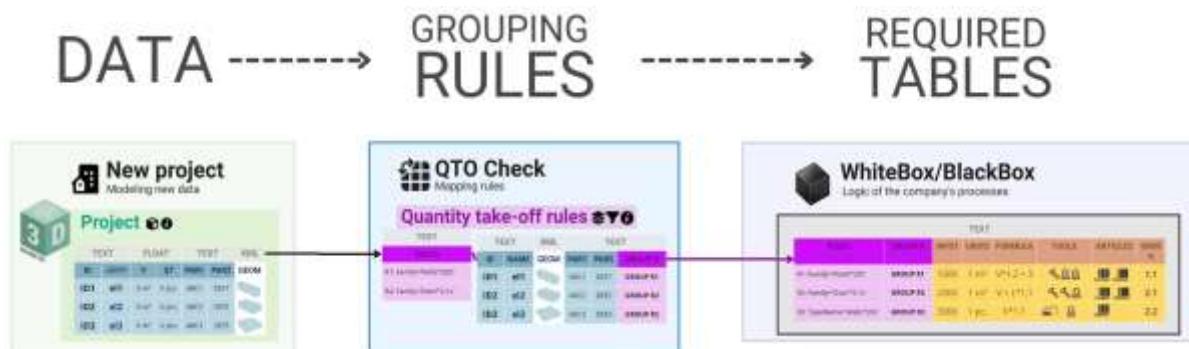
Če je treba pridobiti količine samo za eno skupino elementov, je dovolj, da izvedete preprosto poizvedbo QTO (slika 5.2-7) na podatkih modela CAD. Pri izračunu proračuna ali predračuna za celoten projekt, sestavljen iz več skupin elementov, pa je pogosto treba pridobiti količine za vse vrste elementov (razrede0, pri čemer se vsaka kategorija elementov obdela posebej - z razvrščanjem v skupine po ustreznih atributih).

V praksi ocenjevalcev in cenilcev se za različne vrste predmetov uporablajo posamezna pravila za razvrščanje v skupine in izračunavanje. Na primer, okna se običajno razvrščajo po nadstropijih ali območjih (parameter razvrščanja - atribut Level, Rooms), stene pa po materialu ali vrsti konstrukcije (parameter Material, Type). Za avtomatizacijo postopka združevanja v skupine so takšna pravila

vnaprej opisana v spletni strani v obliki tabel pravil združevanja v skupine. Te tabele delujejo kot konfiguracijske predloge, ki določajo, katere atribute je treba uporabiti pri izračunih za vsako skupino elementov v projektu.

## **QTO izračun celotnega projekta z uporabo skupinskih pravil iz Excelove preglednice**

V resničnih gradbenih projektih je pogosto treba v eni skupini elementov hkrati izvesti združevanje po več atributih. Na primer, pri delu s kategorijo "Okna" (kjer atribut Kategorija vsebuje vrednosti, kot sta OST\_Windows ali IfcWindows) je mogoče elemente grupirati ne le po vrsti - na primer po vrednosti v polju Ime vrste ali Vrsta - temveč tudi po dodatnih lastnostih, kot je stopnja topotne prevodnosti, določena v ustreznom atributu. To večdimenzionalno grupiranje omogoča natančnejše rezultate za določeno skupino. Podobno lahko pri izračunu kategorij sten ali tal kot filtre ali merila za razvrščanje v skupine uporabimo poljubne kombinacije atributov - kot so material, raven, tla, požarna odpornost in drugi parametri (slika 5.2-9).



**Slika 5.2-9 Za vsako skupino ali kategorijo entitet v projektu obstaja drugačna formula za razvrščanje v skupine, sestavljena iz enega ali več meril.**

Postopek opredelitev takšnih pravil združevanja je podoben postopku oblikovanja zahtev za podatke, opisanemu v poglavju "Oblikovanje zahtev in preverjanje kakovosti podatkov" (slika 4.4-5), kjer smo podrobno obravnavali delo s podatkovnimi modeli. Takšna pravila združevanja v skupine in izračuna zagotavljajo natančnost in ustreznost rezultatov za samodejni izračun skupnih atributov količine ali obsega kategorije entitete, pri čemer se upoštevajo vsi potrebni pogoji, ki jih je treba upoštevati pri izračunih in izračunih.

- Naslednji vzorec kode filtrira tabelo projektov tako, da dobljeni niz podatkov vsebuje samo entitete, pri katerih stolpec atributa "Kategorija" vsebuje vrednosti "OST\_Windows" ali "IfcWindows" in hkrati stolpec atributa "Vrsta" vsebuje vrednost "Vrsta 1":

Imam projekt DataFrame - filtriranje podatkov tako, da v podatkovnem nizu ostanejo samo elementi, ki imajo atribut "Kategorija" z vrednostmi "OST\_Windows" ali "IfcWindows" in hkrati atribut Tip vsebuje vrednost "Tip 1" ↴

💡 Odgovor LLM :

```
group.py
df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

Slika 5.2-10 Ena sama vrstica kode, podobna Excelovi formuli , omogoča združevanje vseh entitet projekta po več atributih.

Nastala koda (slika 5.2-10) po prevodu podatkov CAD v strukturirane odprte formate (slika 4.1-13) lahko zaženemo v enem od priljubljenih IDE (integriranih razvojnih okolij), ki smo jih omenili zgoraj, v načinu brez povezave: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z vtičnikom PyDev, Thonny, Wing IDE, IntelliJ IDEA z vtičnikom Python, JupyterLab ali priljubljena spletna orodja: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

💡 Za pridobitev projektnih entitet v obrazcu QTO DataFrame v kategoriji "Windows" samo z določeno vrednostjo toplotne prevodnosti lahko uporabimo naslednjo poizvedbo v LLM:

Imam projekt DataFrame - podatke filtriram tako, da v podatkovnem nizu ostanejo samo zapisi z vrednostmi "Category", ki vsebujejo "OST\_Windows" ali "IfcWindows", hkrati pa mora imeti stolpec ThermalConductivity vrednost 0.. ↴

💡 Odgovor LLM :

```
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

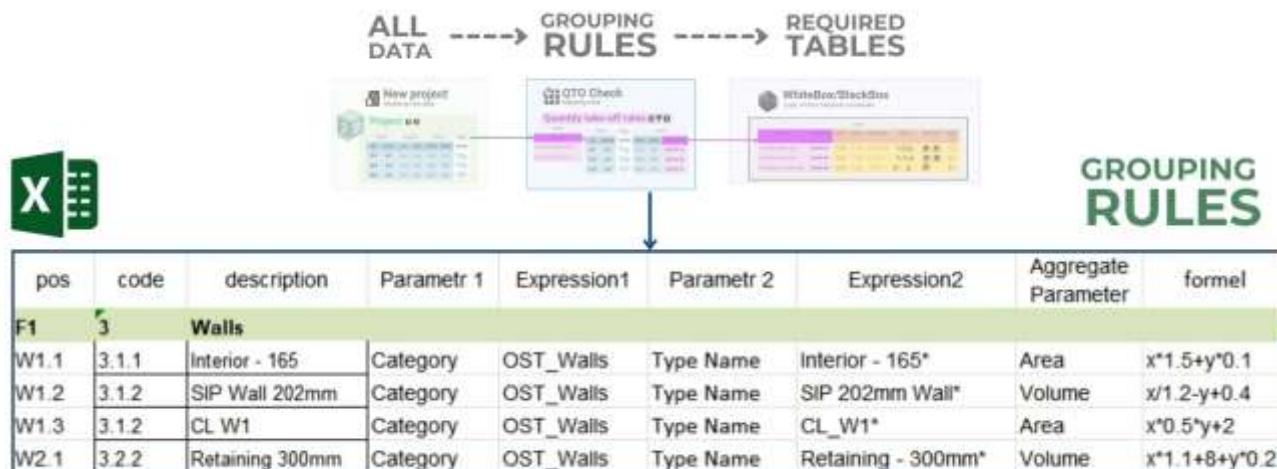
Slika 5.2-11 Izjemno preprost poizvedovalni jezik Pandas Python omogoča hkratno izvajanje QTO za poljubno število projektov.

V odgovoru, prejetem od LLM (slika 5.2-11), je logični pogoj "&" uporabljen za združitev dveh meril: vrednosti toplotne prevodnosti in pripadnosti eni od dveh kategorij. Z metodo "isin" se preveri, ali je vrednost atributa-stolpca "Kategorija" na predloženem seznamu.

Pri projektih z velikim številom skupin elementov z različno logiko združevanja - za vsako kategorijo enot projekta (npr.: okna, vrata, plošče) je treba določiti posamezna pravila združevanja, ki lahko vključujejo dodatne koeficiente ali formule za izračun skupnega atributa. Te formule (slika 5.2-12 atributni "formel", npr. x-vrednost količine in y-obseg skupine) in koeficienti upoštevajo edinstvene značilnosti vsake skupine, npr.:

- % dodatkov k količini materiala zaradi prekoračitve.
- določena dodatna količina materiala
- prilagoditve, povezane z morebitnimi tveganji in napakami v izračunih v obliki formul

Ko so pravila filtriranja in razvrščanja v skupine oblikovana v obliki formul parametrov za vsako kategorijo elementov, jih lahko shranite kot tabelo po vrsticah - na primer v formatu Excel (slika 5.2-12). S shranjevanjem teh pravil v strukturirani obliki je mogoče postopek pridobivanja, filtriranja in združevanja podatkov o projektu popolnoma avtomatizirati. Namesto ročnega pisanja številnih ločenih poizvedb sistem preprosto prebere tabelo parametrov in uporabi ustrezna pravila za model (celotni podatkovni okvir projekta (slika 4.1-13)) ter ustvari končne tabele QTO za vsako kategorijo elementov projekta.



**Slika 5.2-12 Tabela za razvrščanje atributov QTO v skupine določa pravila za razvrščanje elementov projekta v skupine ter zagotavlja natančno skupno število in obseg za vsako kategorijo.**

Zbrana pravila vam bodo omogočila združevanje celotnega projekta in izvedbo vseh potrebnih izračunov, vključno s prilagoditvami atributov obsega. Posledično se volumni pripeljejo do "pravega volumna", ki se uporablja za izračune in kalkulacije, in ne tistih, ki so bili prvotno v fazi projektiranja v modelu CAD.

Pri samodejnem ustvarjanju preglednic obsega QTO za celoten projekt mora aplikacija iti skozi vse kategorije preglednice pravil združevanja, vzeti attribute združevanja, po njih združiti vse elemente projekta in združiti atribut obsega za to skupino ter ga dodatno pomnožiti z izboljševalnim faktorjem ali koeficientom.

Prosimo LLM, da nam napiše kodo za takšno rešitev, pri čemer bi morala koda naložiti dve preglednici - preglednico s pravili grupiranja (slika 5.2-12) in preglednico s podatki iz samega projekta (slika 4.1-13), nato pa uporabiti pravila grupiranja, grupirati elemente po danih pravilih, izračunati agregirane vrednosti in shraniti rezultate v novo datoteko Excel.

- Pošljite besedilno zahtevo v klepetalnico LLM:

Potrebujem kodo za branje podatkov o projektu iz datoteke 'basic\_sample\_project.xlsx' in nato pravila iz datoteke 'Grouping\_rules\_QTO.xlsx', s katerimi se vsi podatki združijo po 'Parametru 1' in 'Parametru 2', agregirajo 'Aggregate Parameter', filtrirajo po 'Expression2', izvedejo izračuni iz 'Formel1' in shranijo tabela QTO v 'QTQ\_table2.xlsx'.

 Odgovor LLM :



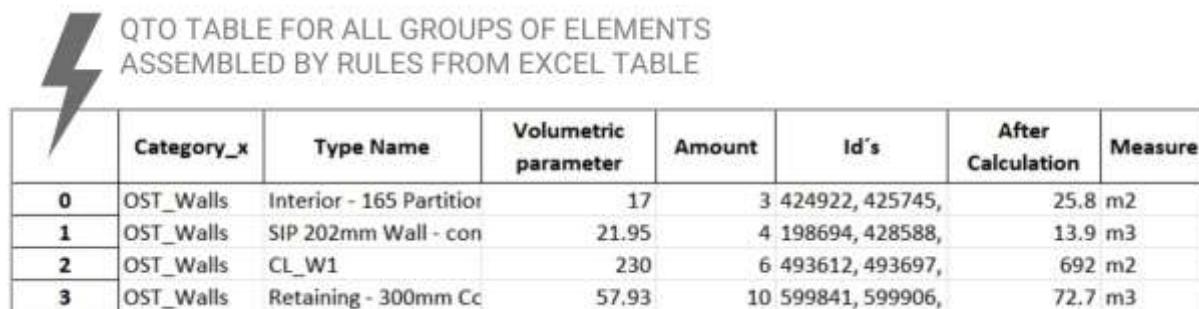
```

1 import pandas as pd
2
3 # Reading CSV and Excel files with project data and grouping rules respectively
4 df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
5 df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])
6
7 # Cleaning numerical columns in the dataframe by removing non-numeric characters
8 numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
9 df[numeric_cols] = df[numeric_cols]
10 df.apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+\.\d*)')|0], fillna(0)))
11
12 # Define a function to process each row in the grouping rules dataframe
13 def process_row(row):
14     # Group by specified parameters and calculate sum and count for the aggregate parameter
15     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
16     [][row["Aggregate Parameter"]].agg(['sum', 'count'])
17     group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]
18
19     # Perform calculation defined in the rules
20     group['After_Calculation'] = eval(row["formula"], {'x': group['sum'], 'y': group['count']})
21     return group.reset_index()
22
23 # Apply the function to each row in the grouping rules dataframe and concatenate the results
24 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)
25
26 # Save the processed data to an Excel file using the OpenPyXL engine
27 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

Slika 5.2-13 Pythonova koda Pandas obdeluje podatke o načrtovanju iz projekta CAD in jih razvršča v skupine v skladu s pravili, določenimi v Excelu .

Končni rezultat izvajanja kode (slika 5.2-13) bo tabela skupine entitet, ki poleg pospoljenih atributov prostornine iz prvotnega modela CAD- (BIM-) vsebuje tudi nov atribut dejanske prostornine, ki upošteva vse zahteve za pravilno izdelavo izračunov in ocen (primer slika 5.2-14).

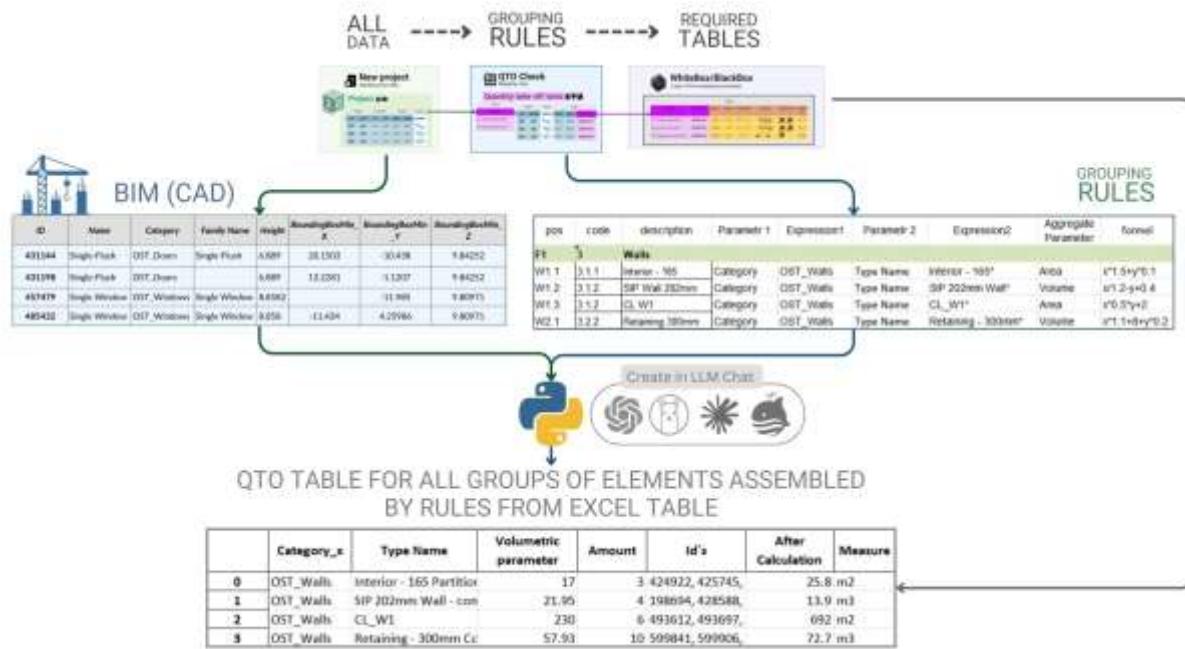


	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition		17	3 424922, 425745,	25.8 m2	
1	OST_Walls	SIP 202mm Wall - con		21.95	4 198694, 428588,	13.9 m3	
2	OST_Walls	CL_W1		230	6 493612, 493697,	692 m2	
3	OST_Walls	Retaining - 300mm Cc		57.93	10 599841, 599906,	72.7 m3	

Slika 5.2-14 Atribut "Po izračunu" se doda v zbirno tabelo po izvedbi kode, ki samodejno izračuna dejansko količino.

Dobljeno kodo (slika 5.2-13) je mogoče zagnati v enem od priljubljenih IDE (ki smo jih omenili zgoraj) in jo uporabiti za poljubno število obstoječih ali novih vhodnih projektov (RVT, IFC, DWG, NWS, DGN

itd.), bodisi nekaj projektov ali morda več sto projektov v različnih formatih v strukturirani obliki (slika 5.2-15).



**Slika 5.2-15 Postopek samodejnega združevanja gradbenih podatkov povezuje podatke BIM (CAD) s tabelami QTO prek pravil iz Excelove preglednice.**

Prilagojeni in parametrizirani postopek zbiranja volumetričnih podatkov (slika 5.2-15) omogoča popolnoma avtomatizirano zbiranje podatkov o kvantitativnih lastnostih in volumnih elementov projekta za nadaljnje delo z njimi, vključno z oceno stroškov, logistiko, urniki dela in izračunom ogljičnega odtisa ter drugimi analitičnimi nalogami.

Po spoznavanju orodij, ki nam omogočajo enostavno organiziranje in združevanje skupin elementov projekta glede na določene atribute, smo zdaj pripravljeni, da združene in filtrirane projekte vključimo v različne izračune podjetja in poslovne scenarije.



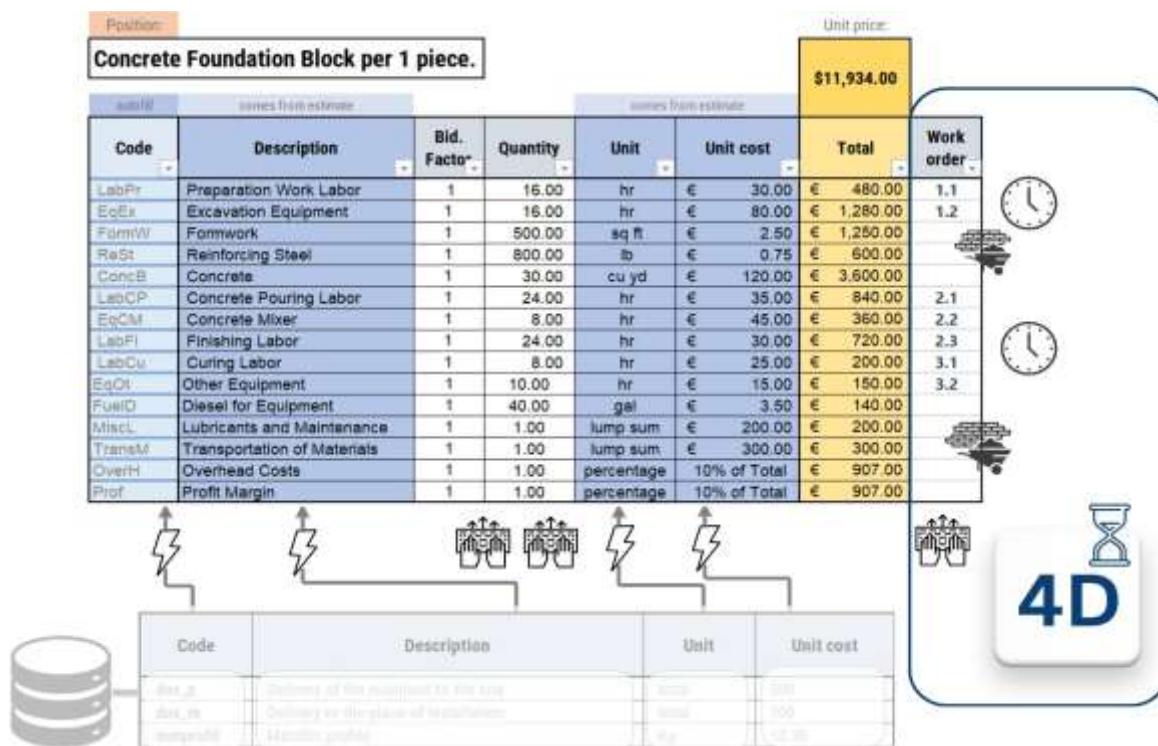
## POGLAVJE 5.3.

### 4D, 6D -8D IN IZRAČUN EMISIJ OGLJIKOVEGA DIOKSIDA CO<sub>2</sub>

#### Model 4D: vključitev časa v gradbene ocene

Ena od ključnih aplikacij projektnih podatkov v gradbeništvu je poleg obračunavanja stroškov tudi določanje časovnih parametrov - tako za posamezne gradbene operacije kot za celoten projekt. Metoda ocenjevanja na podlagi virov in z njo povezana zbirka kalkulativnih podatkov, ki je bila podrobno obravnavana v prejšnjem poglavju "Izračuni in ocene za gradbene projekte", se pogosto uporablja kot podlaga za avtomatiziran izračun časa in izdelavo časovnega načrta za izvedbo del.

Pristop, ki temelji na virih, ne upošteva le materialnih stroškov, temveč tudi časovne vire. Pri obračunavanju stroškov lahko vsakemu procesu pripisemo atribut delovnega naloga (slika 5.3-1 - Parameter delovnega naloga) ter količino časa in stroškov, povezanih z izvedbo tega procesa. Ti parametri so še posebej pomembni za opis postopkov, ki nimajo fiksne tržne cene in jih ni mogoče neposredno kupiti - na primer uporaba gradbene opreme, zaposlitev delavcev ali logistični postopki (ki so običajno običajno izraženi v urah). V takšnih primerih stroškov ne določi oddelek za nabavo, temveč jih neposredno določi izvajalsko podjetje na podlagi notranjih normativov ali proizvodnih stopenj (slika 5.3-1).



Slika 5.3-1 Izračuni dela pri metodi ocenjevanja, ki temelji na virih, vključujejo časovni razpored delovnih ur.

Tako izračuni na ravni stroškov ne vključujejo le stroškov goriva in materiala (stroški nakupa), temveč

tudi čas, ki ga vozniki, tehnični in pomožni delavci porabijo na kraju samem. V navedenem primeru (slika 5.3-1) je stroškovna tabela izračun stroškov vgradnje temeljnega bloka, ki vključuje sestavne faze dela, kot so priprava, vgradnja okvirja in betoniranje, ter potreben material in delo. Vendar posamezne operacije, kot so pripravljalna dela, morda nimajo stroškov materiala, lahko pa vsebujejo znatne začasne stroške dela, izražene v delovnih urah.

Za načrtovanje zaporedja del (za delovni načrt) na gradbišču se v tabelo za izračun ročno doda atribut "Delovni nalog" (slika 5.3-1). Ta atribut je naveden v dodatnem stolpcu samo za elemente, katerih merska enota je izražena v času (ura, dan). Ta atribut je dodatek k šifri dela, opisu, količini, merski enoti (parameter "Enota") in stroškom. Številčno zaporedje (parameter "Zaporedje dela") dejavnosti omogoča določitev vrstnega reda izvajanja nalog na gradbišču in njegovo uporabo pri načrtovanju.

## Časovni načrt gradnje in njegova avtomatizacija na podlagi podatkov o stroških

Gradbeni načrt je vizualni prikaz delovnega načrta in postopkov, ki jih je treba izvesti v okviru izvajanja projekta. Izdelan je na podlagi podrobnih izračunov virov (slika 5.3-1), kjer je vsaka naloga-delo poleg stroškov virov načrtovana po času in zaporedju.

V nasprotju s pristopi povprečenja, pri katerih ocene časa temeljijo na tipičnih urah za vgradnjo materialov ali opreme, pri metodi, ki temelji na virih, načrtovanje temelji na dejanskih podatkih v kalkulaciji stroškov. Vsaka postavka ocene, povezana s stroški dela, temelji na uporabljenem koledarju, ki upošteva dejanske pogoje uporabe virov v delovnem obdobju. Prilagoditev produktivnih ur s pomočjo koeficientov na ravni stroškov (slika 5.3-1 parameter "Bid. faktor"), omogoča upoštevanje razlik v produktivnosti in sezonskih posebnosti, ki vplivajo na časovni potek dela.

Za določitev datumov začetka in konca procesa za časovni načrt gradnje v Ganttovem diagramu vzamemo vrednosti atributa količina časa za vsako postavko iz kalkulacije temeljnih blokov in jih pomnožimo s številom blokov (v tem primeru s številom betonskih temeljnih blokov). S tem izračunom dobimo trajanje vsake naloge. Ta trajanja nato narišemo na časovnico, začenši z datumom začetka projekta, in tako ustvarimo graf, rezultat pa je vizualni prikaz, ki prikazuje, kdaj naj bi se posamezna naloga začela in končala. Parameter "Zaporedje dela" za procese nam dodatno omogoča razumeti, ali delovni proces poteka vzporedno ("Zaporedje dela", npr. 1.1-1.1) ali zaporedno (1.1-1.2).

**Gantsov diagram** je grafično orodje za načrtovanje in upravljanje projektov, ki naloge predstavlja kot vodoravne črte na časovnici. Vsaka vrstica predstavlja trajanje naloge, njen začetek in konec.

Časovni načrt ali Gantsov diagram pomaga vodjem projekta in delavcem jasno razumeti, kdaj in v kakšnem zaporedju je treba izvesti različne faze gradnje, kar zagotavlja učinkovito uporabo virov in spoštovanje rokov.

Predstavimo časovni razpored vgradnje treh betonskih temeljnih blokov z uporabo izračunov iz zgornje preglednice. S pomočjo preglednice stroškov (slika 5.3-1) iz zgornjega primera prosimo LLM, da naredi urnik vgradnje treh elementov temeljnih blokov, npr. za prvi maj 2024.

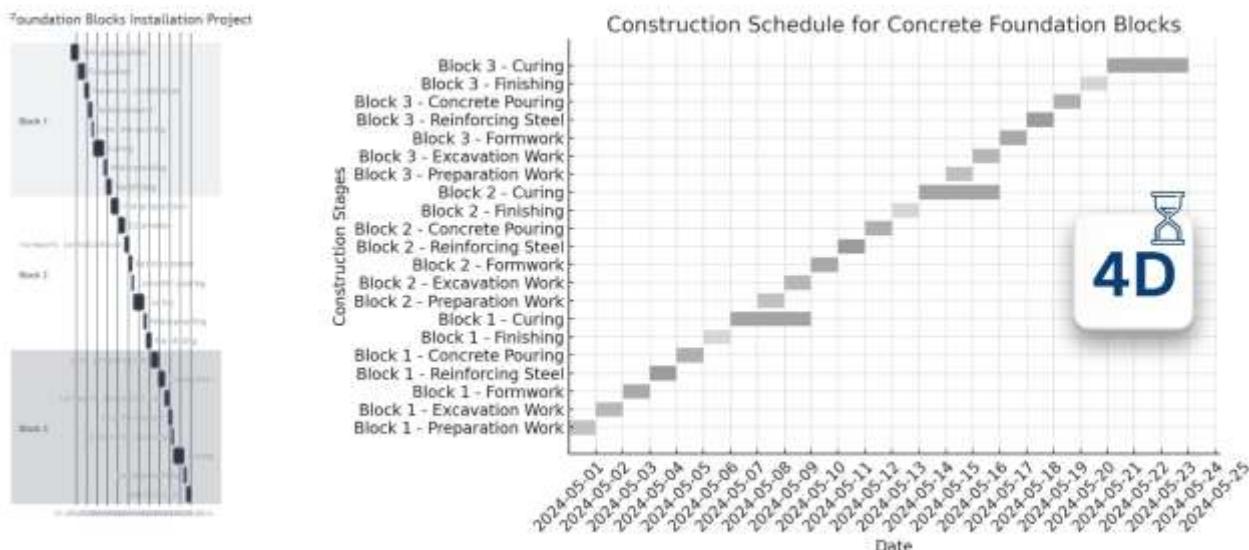
Za pošiljanje kalkulacije stroškov v program LLM, lahko naložimo preglednico stroškov v obliki XLSX

ali preprosto vstavimo zaslonsko sliko JPEG kalkulacije stroškov neposredno v klepetalnik LLM (slika 5.3-1). LLM bo samostojno poiškal knjižnico za vizualizacijo slike tabele in z množenjem časovnih atributov dejavnosti iz tabele z njihovim obsegom vse podatke dodal v graf.

- ❷ Pošljite zahtevo za besedilo na LLM:

Ustvarite razpored dela v obliki Ganttovega diagrama z uporabo vrednosti iz časovno povezane preglednice (slika 5.3-1 kot JPEG v prilogi) za vgradnjo 3 temeljnih blokov. Dela za vsak blok se izvajajo zaporedno. Začetek del se določi od 01.05.2024 ↴

- ❸ Odgovor LLM

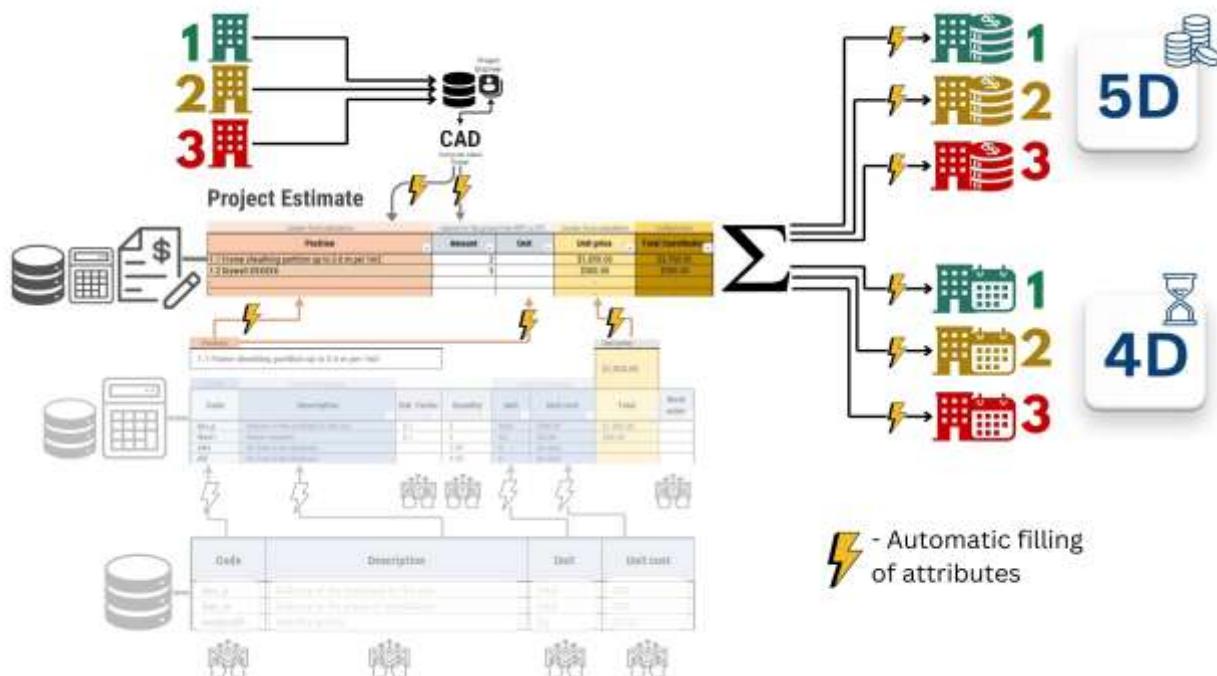


Slika 5.3-2 Samodejno ustvarjeno z več LLM-ji Ganttov diagram prikazuje faze gradnje treh betonskih blokov glede na pogoje iz poziva.

Dobljeni graf (slika 5.3-2) je časovni diagram, v katerem vsaka vodoravna vrstica ustreza določeni fazi del na temeljnem bloku in prikazuje zaporedje operacij (parameter "Vrstni red del"), kot so priprava, izkop, vgradnja opaža, armiranje, betoniranje in zaključna obdelava,, tj. tistih procesov, ki imajo v izračunih izpolnjene časovne parametre in zaporedje.

Tak urnik (slika 5.3-2) ne upošteva omejitev, povezanih z delovnimi dnevi, izmenami ali standardi delovnega časa, temveč je namenjen zgolj konceptualni vizualizaciji procesa. Natančen razpored, ki bo odražal vzporednost dela, je mogoče dopolniti z ustrezнимi pozivi ali dodatnimi navodili v klepetalnici

Z uporabo enotne kalkulacije stroškov (slika 5.3-1) je mogoče zahvaljujoč atributom volumna iz 3D - geometrije samodejno oceniti tako stroške projekta z avtomatiziranimi ocenami kot hkrati izračunati časovne značilnosti skupin v obliki tabel ali grafov za različne variante projekta (slika 5.3-3).



Slika 5.3-3 Samodejni izračun, omogoča takojšnje in samodejno napovedovanje stroškov in časa za različne možnosti projekta.

Sodobni modularni sistemi ERP (slika 5.4-4), ki nalagajo podatke iz modelov CAD, uporabljajo podobne samodejne metode za izračun časa, ki bistveno skrajšajo postopek odločanja. To omogoča takojšnje in natančno načrtovanje razporedov dela ter izračun skupnega časa, potrebnega za izvedbo vseh nalog pri realizaciji projekta, ob upoštevanju realnih cen.

### Razširjeni atributni sloji 6D -8D: od energetske učinkovitosti do zagotavljanja varnosti

6D, 7D in 8D so razširjene ravni informacijskega modeliranja, od katerih vsaka prispeva dodatne plasti atributov k celovitemu informacijskemu modelu projekta, katerih osnova so atributi 3D -modela z njihovim številom in obsegom. Vsaka dodatna raven uvaja posebne parametre, ki so potrebni za nadaljnje razvrščanje v skupine ali nadaljnjo identifikacijo v drugih sistemih, kot so na primer sistemi za upravljanje nepremičnin (PMS), računalniško podprto upravljanje objektov (CAFM), upravljanje gradbenih projektov (CPM) in sistemi za upravljanje varnosti (SMS).



Slika 5.3-4 Atributi 6D, 7D in 8D v podatkovnem informacijskem modelu razširjajo obravnavo različnih vidikov projekta, od energetske učinkovitosti do varnosti.

- V **6D** so poleg podatkovne zbirke projekta (ali podatkovnega okvira (slika 4.1-13)) z geometrijskimi in volumetričnimi atributi elementov dodane informacije (atributni stolpci) o okoljski trajnosti. To vključuje informacije, povezane z energetsko učinkovitostjo, ogljičnim odtisom, možnostjo recikliranja materialov in uporabo okolju prijaznih tehnologij. Ti podatki omogočajo oceno vpliva projekta na okolje, optimizacijo projektnih odločitev in doseganje ciljev trajnostnega razvoja (ESG).
- **7D** lastnosti dopolnjujejo lastnosti, ki so potrebne za upravljanje vzdrževanja stavb. To so podatki o urnikih vzdrževanja, življenjskih ciklih sestavnih delov, tehnični dokumentaciji in zgodovini popravil. Ta nabor informacij zagotavlja, da je model mogoče povezati s sistemi za vzdrževanje (CAFM, AMS), omogoča učinkovito načrtovanje vzdrževanja, zamenjavo opreme in zagotavlja podporo v celotnem življenjskem ciklu objekta.
- **8D** dodatna plast atributov, - vključuje informacije, povezane z varnostjo - tako v fazi gradnje kot med nadaljnjjim delovanjem. Model vključuje ukrepe za varnost osebja, navodila za ravnanje v sili, sisteme za evakuacijo in zahteve za požarno zaščito. Vključitev teh podatkov v digitalni model pomaga vnaprej upoštevati tveganja ter razviti arhitekturne, inženirske in organizacijske rešitve, ki upoštevajo zdravstvene in varnostne zahteve.

V strukturirani tabelarni obliki plasti 4D do 8D predstavljajo dodatne attribute v obliki stolpcev z izpolnjenimi vrednostmi (slika 5.3-5), ki so dodani že izpolnjenim atributom modela 3D, kot so ime, kategorija, tip in volumetrične značilnosti. Vrednosti v slojih atributov 6D, 7D in 8D vsebujejo dodatne besedilne in številčne podatke, kot so odstotek recikliranja, ogljični odtis, garancijski rok, cikel zamenjave, datum namestitve, varnostni protokoli itd.

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 kgCO <sub>2</sub> e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO <sub>2</sub> e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO <sub>2</sub> e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO <sub>2</sub> e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

**Slika 5.3-5 6D -8D dodaje atributne sloje podatkovnemu informacijskemu modelu, ki že vsebuje geometrijske in volumetrične attribute iz 3D -modela.**

Za naše novo okno (slika 4.4-1) ima lahko element z identifikatorjem W-NEW (slika 5.3-5) naslednje attribute 3D -8D:

**3D -atributi** - geometrijske informacije, pridobljene iz sistemov CAD:

- "Ime tipa" - element "Okno"
- "Širina" - 120 cm
- Poleg tega lahko dodate točke "Bounding Box" elementa ali njegove "geometrije BREP / MESH" kot ločen atribut.

**Lastnosti 6D** - okoljska trajnost:

- 90-odstotna stopnja recikliranja
- "Ogljični odtis" - 1.622 kg CO<sub>2</sub>

**Atributi 7D** - podatki o upravljanju predmetov:

- "Garancijski rok" - 8 let
- "Nadomestni cikel" je star 20 let
- "Vzdrževanje" - zahteva se vsako leto

**Atributi 8D** - zagotavljanje varne uporabe in obratovanja stavb:

- Okno "Nameščeno" - podjetje "XYZ Windows"
- "Varnostni standard" - skladen s standardom ISO 45001

Vse parametre, zapisane v podatkovni zbirkki ali naboru podatkov (slika 5.3-5), potrebujejo strokovnjaki v različnih oddelkih za združevanje, iskanje ali izračune. Takšen večdimensionalni atributni opis projektnih objektov zagotavlja popolno sliko njihovega življenjskega cikla, operativnih zahtev in številnih drugih vidikov, potrebnih za načrtovanje, gradnjo in delovanje projekta.

## Ocena emisij CO<sub>2</sub> in izračun emisij ogljikovega dioksida pri gradbenih projektih

Poleg teme trajnosti gradbenih projektov na stopnji 6D (slika 5.3-5) se sodobno gradbeništvo osredotoča na okoljsko trajnost projektov, pri čemer eden od ključnih vidikov postaja ocena in zmanjšanje emisij ogljikovega dioksida CO<sub>2</sub>, ki nastanejo v fazah življenjskega cikla projekta (npr. proizvodnja in namestitev).

Ocenjevanje in izračun emisij ogljika gradbenih materialov je postopek, pri katerem se skupne emisije ogljika določijo z množenjem prostorninskih lastnosti elementa ali skupine elementov, uporabljenih v projektu, z ustreznim faktorjem emisij ogljika za to kategorijo.

Upoštevanje emisij ogljika pri ocenjevanju gradbenih projektov v okviru širših meril ESG (okolje, družba in upravljanje) dodaja analizi novo raven zapletenosti. To je še posebej pomembno za naročnika-investitorja pri pridobivanju ustreznih certifikatov, kot so LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) ali DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Pridobitev enega od teh certifikatov lahko bistveno poveča tržno zanimivost nepremičnine, poenostavi začetek obratovanja in zagotovi skladnost z zahtevami trajnostno usmerjenih najemnikov (ESG). Glede na zahteve projekta se lahko uporabijo tudi HQE (Haute Qualité Environnementale, francoski standard za zelene stavbe), WELL (WELL Building Standard, ki se osredotoča na zdravje in udobje uporabnikov) in GRESB (Global Real Estate Sustainability Benchmark).

Okoljsko, družbeno in upravljavsko **ESG** (environmental, social and governance) je širok nabor načel, ki se lahko uporablajo za ocenjevanje upravljanja podjetja ter družbenega in okoljskega vpliva podjetja tako interna kot eksterno.

Spletna stran ESG, ki so jo v začetku leta 2000 prvotno razvili finančni skladi, da bi vlagateljem zagotovili informacije o širših okoljskih, socialnih in upravljavskih merilih, se je razvila v ključni kazalnik za ocenjevanje podjetij in projektov, vključno z gradbenimi projekti. Glede na raziskave večjih svetovalnih podjetij postajajo okoljski, socialni in upravljavski vidiki (ESG) sestavni del gradbene industrije.

Po podatkih EY (2023) "The Path to Carbon Neutrality" podjetja, ki aktivno izvajajo načela ESG, ne zmanjšujejo le dolgoročnih tveganj, temveč tudi povečujejo učinkovitost svojih poslovnih modelov, kar je še posebej pomembno pri globalnem preoblikovanju trgov [103]. Poročilo družbe PwC o ozaveščenosti o ESG navaja, da se podjetja zavedajo pomena dejavnikov ESG od 67% do 97%, pri čemer večina organizacij meni, da so ti trendi ključni za prihodnjo trajnost [104], in da podjetja večinoma opažajo velik pritisk deležnikov, da vključijo načela ESG.

Tako vključevanje načel ESG v gradbene projekte ne prispeva le k pridobivanju mednarodnih trajnostnih certifikatov, kot so LEED, BREEAM, DGNB, temveč zagotavlja tudi dolgoročno trajnost in konkurenčnost podjetij v panogi.

Eden najpomembnejših dejavnikov, ki vplivajo na skupni ogljični odtis gradbenega projekta, so proizvodne in logistične faze gradbenih materialov in komponent. Materiali, ki se uporabljajo na gradbišču, pogosto odločilno vplivajo na skupne emisije CO<sub>2</sub>, zlasti v zgodnjih fazah življenjskega cikla projekta - od pridobivanja surovin do dostave na gradbišče.

Za izračun emisij po kategorijah ali vrstah gradbenih elementov je treba uporabiti referenčne faktorje emisij ogljika, ki odražajo količino CO<sub>2</sub>, ki nastane pri proizvodnji različnih materialov. Takšni materiali vključujejo beton, opeko, reciklirano jeklo, aluminij in druge. Te vrednosti se običajno pridobijo iz uglednih virov in mednarodnih podatkovnih zbirk, kot sta UK ICE 2015 (Inventory of Carbon and Energy) in US EPA 2006 (U.S. Environmental Protection Agency) [105]. V naslednji preglednici (slika 5.3-6) so povzeti izhodiščni emisijski faktorji za vrsto običajnih gradbenih materialov. Za vsak material sta navedena dva ključna parametra: specifične emisije CO<sub>2</sub> (v kilogramih na kilogram materiala) in pretvorbeni faktorji prostornine v maso (v kilogramih na kubični meter), ki so potrebni za vključitev izračunov v model projektiranja in povezavo s skupino podatkov QTO.



Carbon Emitted in Production		UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m³ to kg
Material	Abbreviated	Process Emissions (kg CO <sub>2</sub> / kg of product) (X1)	Process Emissions (kg CO <sub>2</sub> / kg of product) (X2)	
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

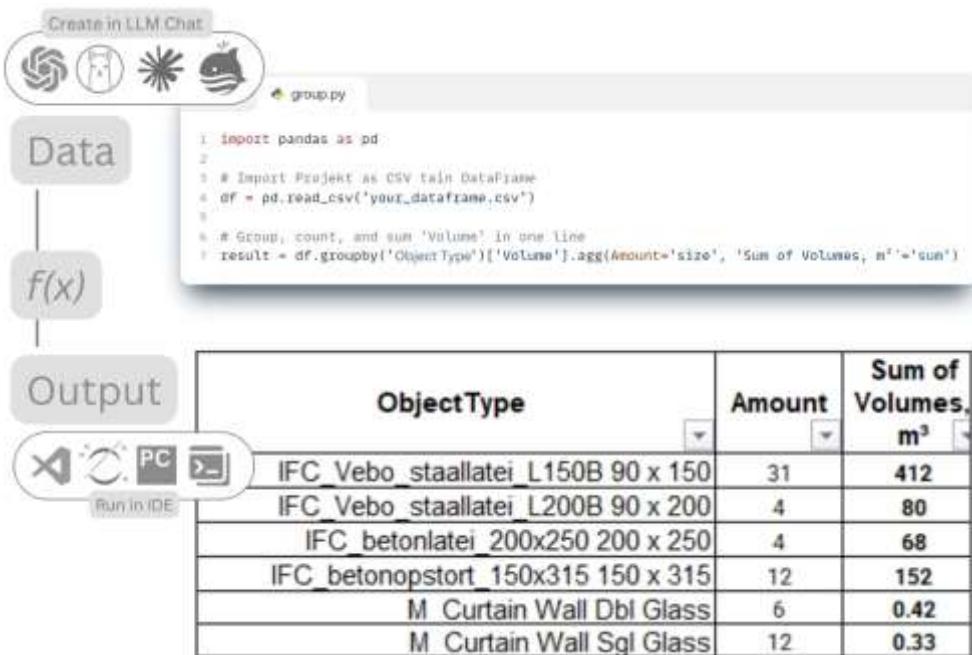
Slika 5.3-6 Količina izpustov ogljika pri proizvodnji različnih gradbenih materialov po podatkih podatkovne zbirke UK ICE in US EPA.

Za izračun skupnih emisij CO<sub>2</sub> za projekt je treba tako kot pri izračunih 4D in 5D določiti količine atributov vsake skupine objektov. To je mogoče storiti z uporabo orodij za kvantitativno analizo (QTO), tako da se pridobijo atributne količine v kubičnih metrih, kot je podrobno obravnavano v razdelku o količinski odbitki. Te količine se nato pomnožijo z ustreznimi koeficienti za atribut "emisije CO<sub>2</sub> iz procesov" vsake skupine materialov.

- Iz projekta CAD (BIM) samodejno izluščimo preglednico volumnov po vrsti elementa tako, da združimo vse podatke projekta, kot smo to že storili v prejšnjih poglavjih. Za izvedbo te naloge se sklicujmo na program LLM.

Preglednico DataFrame iz projekta CAD (BIM) razvrstite v skupine po parametru stolpca "Ime objekta" (ali "Vrsta") in prikažite število elementov v vsaki skupini ter povzemite parameter "Obseg" za vse elemente v vrsti. ↴

💡 Odgovor LLM :



```

Create in LLM Chat
Data
f(x)
Output
Run in IDE
group.py

1 import pandas as pd
2
3 # Import Projekt as CSV tain DataFrame
4 df = pd.read_csv('your_dataframe.csv')
5
6 # Group, count, and sum 'Volume' in one line
7 result = df.groupby('ObjectType')['Volume'].agg(Count='size', Sum_of_Volumes='sum')

```

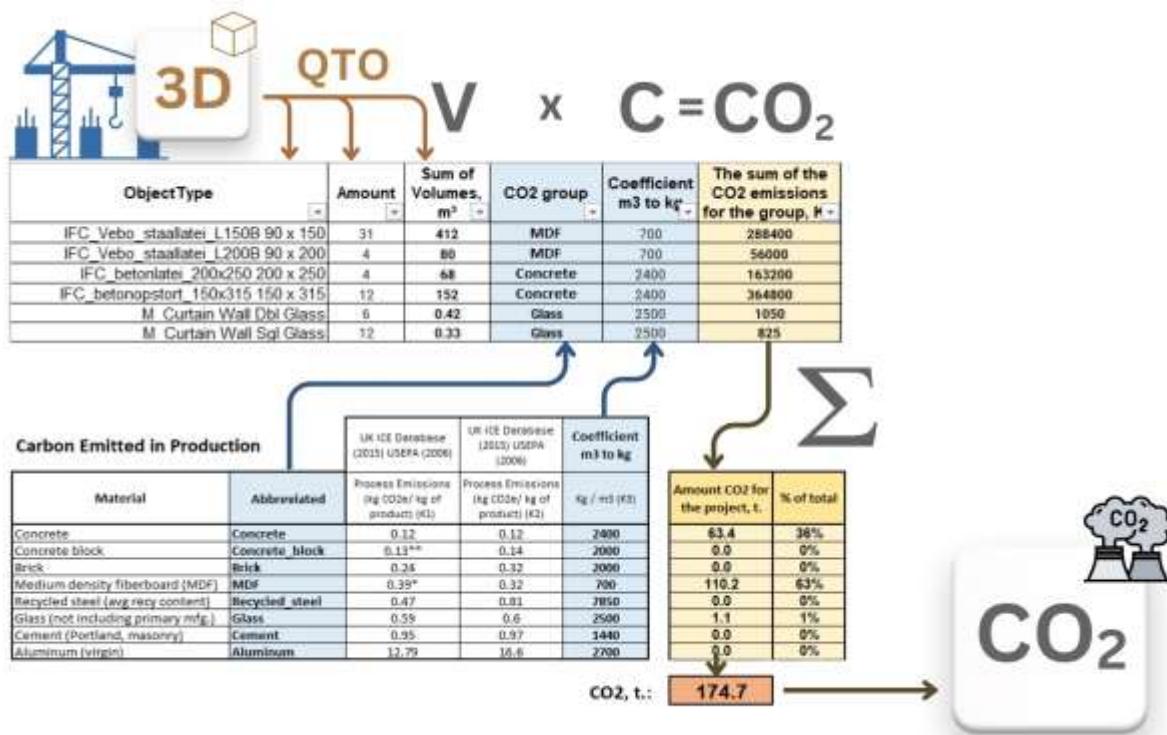
ObjectType	Amount	Sum of Volumes, m <sup>3</sup>
IFC_Vebo_staalatei_L150B_90_x_150	31	412
IFC_Vebo_staalatei_L200B_90_x_200	4	80
IFC_betonlatei_200x250_200_x_250	4	68
IFC_betonopstort_150x315_150_x_315	12	152
M_Curtain_Wall_Dbl_Glass	6	0.42
M_Curtain_Wall_Sgl_Glass	12	0.33

Slika 5.3-7 Ustvarjena koda v LLM je za nas združila projektne entitete po vrsti (ObjectType) s povzetkom atributa "Volume".

Če želite avtomatizirati izračun skupnih emisij CO<sub>2</sub> za celoten projekt, preprosto nastavite samodejno preslikavo podatkov v preglednici ali ročno povežite vrste elementov (slika 5.3-7) z ustreznimi vrstami materialov (slika 5.3-6) iz preglednice faktorjev emisij. Končno tabelo z emisijskimi faktorji in formulami ter kodo za pridobivanje volumnov iz formatov CAD (BIM) in avtomatizacijo določanja CO<sub>2</sub> lahko najdete na GitHubu z iskanjem "CO<sub>2</sub>\_calculating-the-embodyed-carbon. DataDrivenConstruction." [106].

Tako integracija podatkov po združevanju elementov QTO iz podatkovne zbirke CAD omogoča samodejni izračun emisij ogljikovega dioksida (slika 5.3-8) za različne možnosti načrtovanja. Tako je mogoče analizirati vpliv različnih materialov v različnih variantah in izbrati le tiste rešitve, ki izpolnjujejo zahteve glede emisij CO<sub>2</sub> za pridobitev določenega certifikata ob predaji stavbe v uporabo

Ocena emisij CO<sub>2</sub> z množenjem faktorjev s količinami združenih elementov projekta je tipičen primer naloge v postopku gradbenega podjetja, ki za objekt pridobi oceno ESG (npr. certifikat LEED).



Slika 5.3-8 Integracija skupin QTO iz podatkovnih zbirk CAD zagotavlja natančnost in avtomatizacijo pri izpeljavi ocen končnih emisij CO<sub>2</sub>.

Podobno lahko z opredelitvijo obsega skupin elementov izvedemo izračune za nadzor in logistiko materiala, spremljanje in upravljanje kakovosti, energetsko modeliranje in analizo ter številne druge naloge za pridobitev novega statusa atributa (parametra v preglednici) tako za posamezne skupine elementov kot za celoten projekt.

Če se število takšnih računskih postopkov v podjetju začne povečevati, se pojavi vprašanje o potrebi po avtomatizaciji takšnih izračunov in implementaciji rezultatov izračunov v procese in sisteme za upravljanje podatkov v podjetju.

Zaradi kompleksnosti celovite rešitve srednje velika in velika podjetja v gradbeništvu takšno avtomatizacijo naročajo pri podjetjih za razvoj sistemov ERP (ali PMIS). Razvojna podjetja za velike stranke ustvarijo en sam celovit modularni sistem za upravljanje številnih različnih informacijskih plasti, vključno z izračuni materiala in virov.



## POGLAVJE 5.4.

### GRADBENI SISTEMI ERP IN PMIS

#### Gradbeni ERP -sistemi na primeru izračunov in ocen

Modularni sistemi ERP združujejo različne atributne (informacijske) plasti in podatkovne tokove v en sam celovit sistem, ki vodjem projektov omogoča sinhronizirano upravljanje virov, financ, logistike in drugih vidikov projekta znotraj ene platforme. Gradbeni sistem ERP deluje kot "možgani" gradbenih projektov, saj z avtomatizacijo poenostavlja ponavljajoče se procese ter zagotavlja preglednost in nadzor v celotnem procesu gradnje.

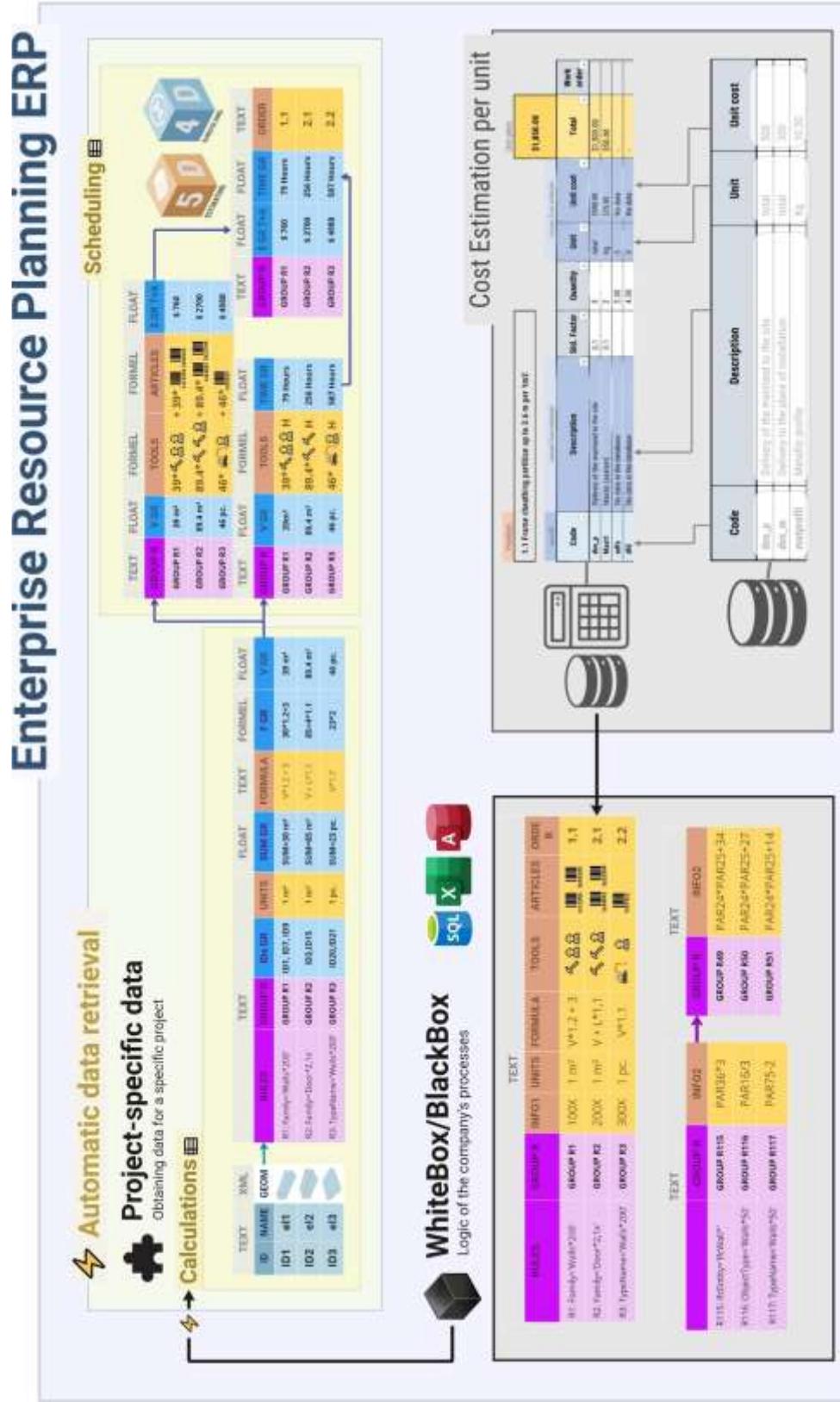
**Gradbeni ERP -sistemi** (Enterprise Resource Planning) so celovite programske rešitve, namenjene upravljanju in optimizaciji različnih vidikov gradbenega procesa. Jedro gradbenih sistemov ERP so moduli za upravljanje stroškov in načrtovanje, zato so pomembno orodje za učinkovito načrtovanje virov.

Moduli ERP -sistemi uporabnikom omogočajo strukturiran vnos, obdelavo in analizo podatkov, ki zajemajo različne vidike projekta, ki lahko vključujejo obračunavanje stroškov materiala in dela, uporabo opreme, upravljanje logistike, človeške vire, stike in druge gradbene dejavnosti.

Eden od funkcionalnih blokov sistema je modul za avtomatizacijo poslovne logike - BlackBox/WhiteBox, ki ima vlogo nadzornega centra procesa.

BlackBox /WhiteBox omogoča strokovnjakom, ki uporablajo sistem ERP, da s pravicami dostopa prilagodljivo upravljajo različne vidike poslovanja, ki so jih že predhodno konfigurirali drugi uporabniki ali skrbniki. V okviru sistemov ERP se izraza *BlackBox* in *WhiteBox* nanašata na ravni preglednosti in obvladljivosti notranje logike sistema:

- **BlackBox** ("črna skrinjica") - uporabnik komunicira s sistemom prek vmesnika brez dostopa do notranje logike izvajanja procesa. Sistem sam izvaja izračune na podlagi vnaprej določenih pravil, ki so končnemu uporabniku skrita. Ta vnese podatke in prejme rezultat, ne da bi vedel, kateri atributi ali koeficienti so bili uporabljeni znotraj.
- **WhiteBox** ("bela škatla") - procesna logika je na voljo za ogled, prilagajanje in spreminjanje. Napredni uporabniki, skrbniki ali integratorji lahko ročno opredelijo algoritme obdelave podatkov, pravila za izračun in scenarije interakcije med subjekti projekta.



Slika 5.4-1 Arhitektura gradbenega ERP -sistema za pridobivanje ocen in razporedov del z ročnim izpolnjevanjem atributov obsega.

Primer je, ko izkušen uporabnik ali skrbnik določi pravilo: katere atributi v oceni je treba pomnožiti med seboj ali jih združiti glede na določeno značilnost in kje je treba zabeležiti končni rezultat. Nato manj usposobljeni strokovnjaki, kot so inženirji za ocenjevanje, preprosto naložijo nove podatke v ERP prek uporabniškega vmesnika - in dobijo pripravljene ocene, razporede ali specifikacije, ne da bi morali pisati kodo ali razumeti tehnične podrobnosti logike.

V prejšnjih poglavjih sta bila računski in logični modul obravnavana v okviru interakcije LLM. V okolju ERP takšni izračuni in pretvorbe potekajo v modulih, skritih za vmesnikom z gumbi in obrazci.

V naslednjem primeru (slika 5.4-1) je skrbnik sistema ERP -sistema v modulu BlackBox /WhiteBox določil pravila za ujemanje atributov entitet iz ocen z atributi za grupiranje QTO. Zahvaljujoč tej konfiguraciji modula BlackBox/WhiteBox (s strani upravitelja ali skrbnika) uporabnik (ocenjevalec ali inženir) z ročnim dodajanjem atributa količine ali obsega prek uporabniškega vmesnika ERP samodejno prejme dokončane oceane in razporede dela. Na ta način postanejo postopki izračuna in izdelave ocen, obravnavani v prejšnjih poglavjih, s pomočjo kode znotraj sistema ERP polautomatski tekoči.

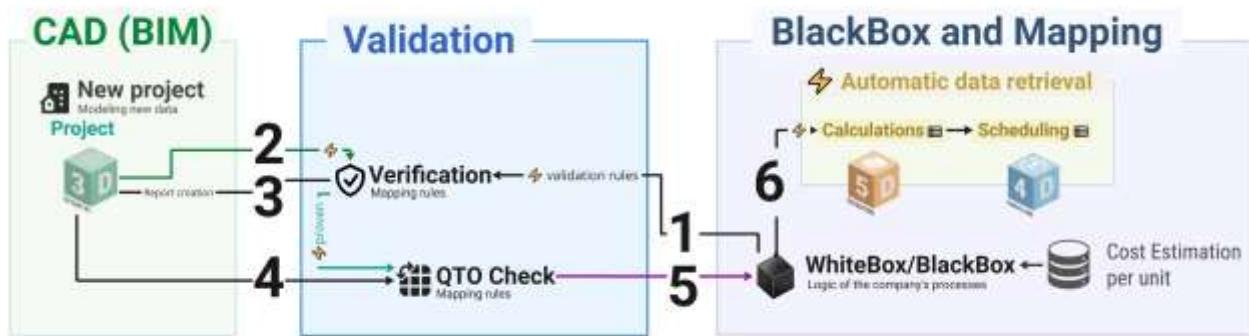
Povezava tega polautomatskega postopka z volumetričnimi atributi iz modelov CAD (BIM) (slika 4.1-13), na primer z nalaganjem projekta CAD v modul ERP, ki je v ta namen vnaprej konfiguriran, spremeni pretok podatkov v sinhroniziran mehanizem, ki lahko samostojno in takoj posodobi vrednost posameznih skupin elementov ali celotnega projekta glede na morebitne spremembe v njem v fazi projektiranja, ko se model CAD naloži v ERP.

Za vzpostavitev avtomatiziranega pretoka podatkov (slika 5.4-2) med sistemi CAD (BIM) in ERP je treba strukturirano opredeliti osnovne procese in zahteve za podatke iz podatkovnih zbirk modelov CAD (BIM), kot smo že obravnavali v zgornjem poglavju "Zahteve in zagotavljanje kakovosti podatkov ". Ta proces v sistemu ERP je razdeljen na podobne korake:

- **Oblikovanje pravil potrjevanja (1)**, ki imajo pomembno vlogo pri zagotavljanju točnosti podatkov, ki vstopajo v sistem ERP. Validacijska pravila služijo kot filtri, ki potrjujejo entitete in njihove attribute ter omogočajo vstop v sistem samo tistim elementom, ki izpolnjujejo zahteve. Več o preverjanju in potrjevanju lahko izveste v poglavju "Ustvarjanje zahtev in potrjevanje kakovosti podatkov ".
- Nato se v ERP izvede **postopek preverjanja (2)**, ki potrdi, da so bili vsi elementi projektne entitete s svojimi atributi in vrednostmi pravilno ustvarjeni in pripravljeni za naslednje korake obdelave.
- Če se pojavi težave zaradi nepopolnih podatkov o attributih, **se pripravi poročilo (3)** in projekt se skupaj z navodili za popravke pošlje v revizijo, dokler ni pripravljen za naslednjo iteracijo.
- Ko so podatki o projektu potrjeni in preverjeni, se uporabijo v drugem modulu ERP **(4) za izdelavo tabel za prevzem količine (QTO)**, ki ustvarijo attribute količine za skupine entitet, materiale in vire v skladu s predhodno oblikovanimi pravili (WhiteBox/BlackBox).
- Podatki, združeni po pravilih ujemanja ali QTO, se samodejno **vključijo v izračune (npr.**

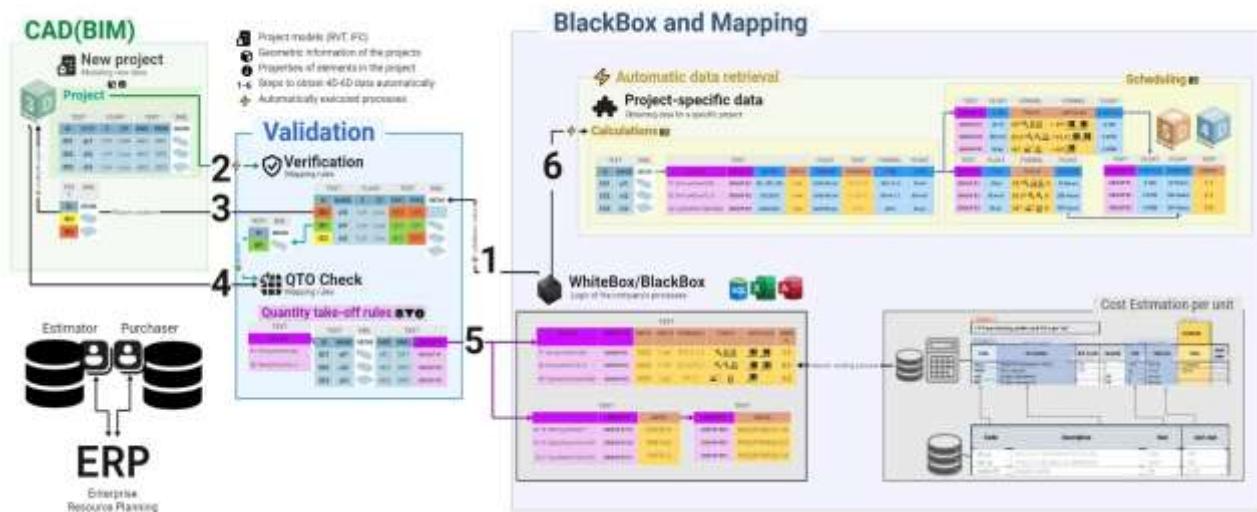
### stroški in čas) (5).

- V zadnjem koraku sistema ERP -sistema uporabnik z množenjem atributov obsega iz preglednice QTO z atributi procesnih preglednic (npr. ocenjene postavke) **samodejno ustvari rezultate izračunov (6)** npr. ocene stroškov , urnike dela ali emisije CO<sub>2</sub> za vsako skupino subjektov in za projekt kot celoto.

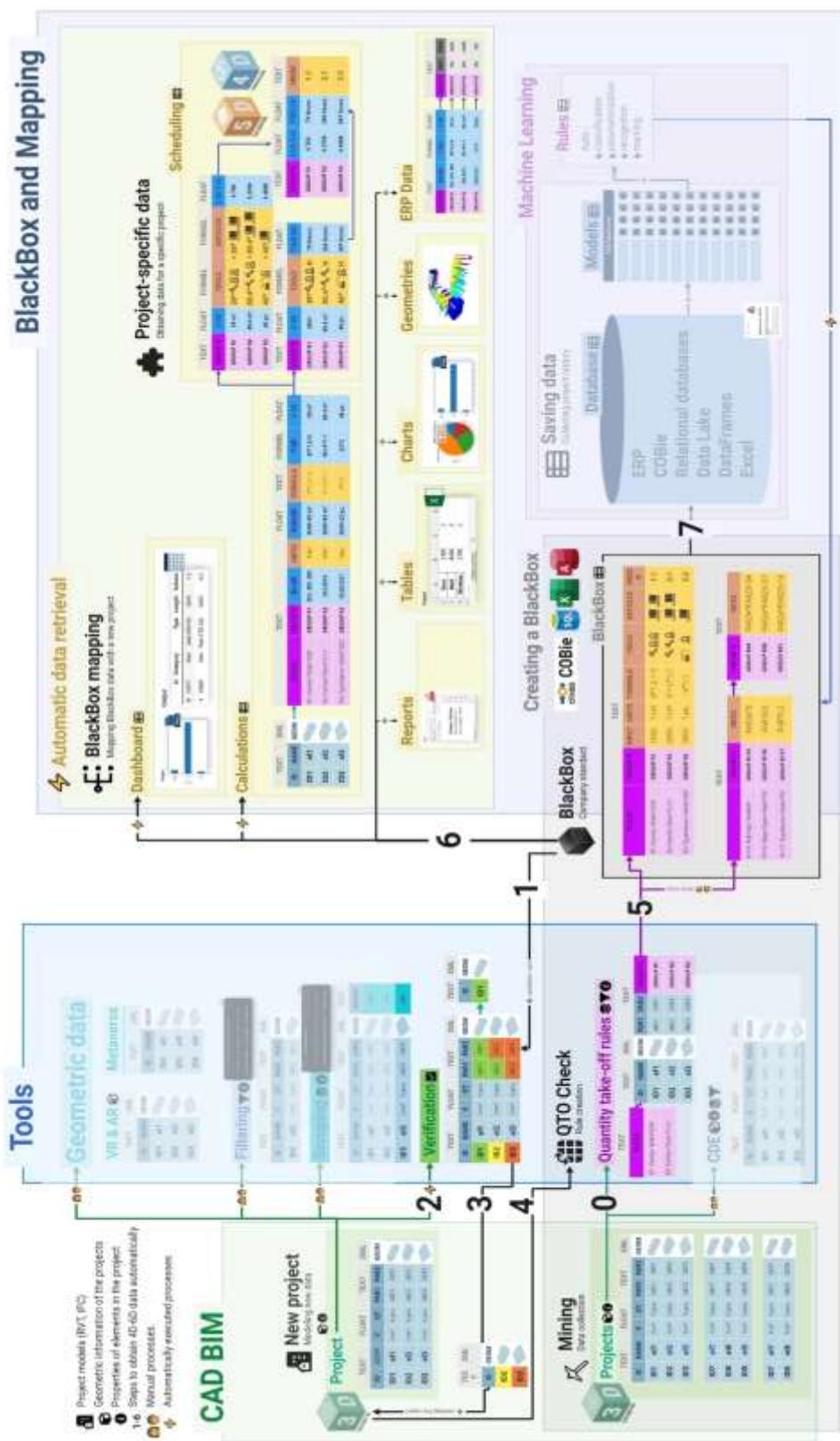


Slika 5.4-2 Arhitektura gradbenega ERP -sistema s CAD (BIM), od oblikovanja pravil potrjevanja (1) do samodejnega izračuna stroškov in urnikov dela (5-6).

V modularnem sistemu ERP so procesi integrirani z uporabo programske opreme, ki vključuje uporabniški vmesnik. Za vmesnikom je zaledni del, kjer strukturirane tabele obdelujejo podatke z izvajanjem različnih operacij, ki jih je predhodno konfiguriral vodja ali skrbnik. Tako uporabnik zaradi vnaprej določene in prilagojene logike avtomatizacije (v modulih BlackBox /WhiteBox) prejme polavtomatsko pripravljeni dokumente, ki izpolnjujejo njegove naloge



Slika 5.4-3 ERP -sistem pomaga upravljavcem in uporabnikom, da se premikajo med strokovnimi tabelami in ustvarjajo nove podatke.



Slika 5.4-4 ERP -sistem je integriran z analitičnimi orodji in avtomatizira proces odločanja v podjetju.

Podobno so tudi procesi v sistemih ERP -od začetka do končnega izračuna (koraki 1-6 Slika 5.4-3) veriga medsebojno povezanih korakov, ki na koncu zagotavljajo preglednost, učinkovitost in natančnost pri načrtovanju.

Sodobni gradbeni ERP -sistemi poleg modulov za izračun stroškov in časovnega načrta vključujejo tudi več deset drugih vnaprej konfiguriranih modulov, ki običajno zajemajo upravljanje dokumentov, sledenje napredku projekta, upravljanje pogodb, dobavno verigo in logistiko ter integracijo z drugimi poslovnimi sistemi in platformami. Integrirana analitična orodja ERP uporabnikom omogočajo avtomatizirano izdelavo nadzornih plošč za spremeljanje ključnih kazalnikov uspešnosti projekta (KPI - key performance indicators). To zagotavlja centralizirano in dosledno upravljanje vseh vidikov gradbenega projekta, pri čemer se poskuša združiti veliko število aplikacij in sistemov na eni platformi.

V prihodnosti se bo ERP -analitika uporabljal v kombinaciji s strojnim učenjem za povečanje natančnosti in optimizacijo postopka izračuna prihodnjih atributov projekta. Podatki in atributi, ki se analizirajo in zbirajo iz sistemov ERP v sistemu Big Data (slika 5.4-4), bodo v prihodnosti podlaga za oblikovanje napovednih modelov, ki bodo lahko natančno predvideli morebitne zamude, tveganja ali na primer morebitne spremembe stroškov materiala.

Kot alternativa ERP, se v gradbeništvu pogosto uporablja PMIS (Project Management Information System), sistem za vodenje projektov, ki je namenjen podrobному nadzoru nalog na ravni posameznega gradbenega projekta.

## **PMIS: vmesni člen med ERP in gradbiščem**

Za razliko od ERP, ki zajema celotno verigo poslovnih procesov podjetja, se PMIS osredotoča na upravljanje določenega projekta, spremeljanje časovnih rokov, proračunov, virov in dokumentacije.

PMIS (Project Management Information System) je programska oprema za vodenje gradbenih projektov, namenjena načrtovanju, spremeljanju, analiziranju in poročanju o vseh vidikih projekta.

PMIS omogoča upravljanje dokumentov, časovnih razporedov, proračunov in na prvi pogled se morda zdi, da je PMIS podvojena rešitev ERP, vendar je ključna razlika v ravni upravljanja:

- **ERP** se osredotoča na poslovne procese podjetja kot celote: stroške, pogodbe, nabavo, človeške vire in upravljanje virov na ravni podjetja.
- **PMIS** se osredotoča na vodenje posameznih projektov, zagotavlja podrobno načrtovanje, nadzor sprememb, poročanje in usklajevanje udeležencev.

V mnogih primerih so to sistemi ERP, ki že imajo zadostno funkcionalnost, in uvedba PMIS postane bolj stvar udobja in želja podjetja. Mnogi izvajalci in naročniki ne uporabljajo PMIS zato, ker bi bilo to potrebno, ampak ker jim to vsili prodajalec ali velika stranka, ki želi združevati podatke na določeni platformi.

Omeniti je treba, da v mednarodni terminologiji za upravljanje gradbenih projektov obstajajo še drugi ločeni priljubljeni pojmi, kot so PLM (Product Lifecycle Management) ter EPC in EPC-M (Engineering,

Procurement and Construction Management) - metode sklepanja pogodb v gradbeništvu.

Če podjetje že uporablja ERP z moduli za vodenje projektov, je lahko uvedba PMIS nepotrebna povezava, ki podvaja funkcionalnost. Če pa procesi niso avtomatizirani in so podatki razdrobljeni, je lahko PMIS priročnejše in enostavnejše orodje za vzdrževanje.

## **Špekulacije, dobiček, izoliranost in pomanjkanje preglednosti v ERP in PMIS**

Kljud zunanjih preprostosti vmesnikov in postopkov so gradbeni sistemi ERP in PMIS v večini primerov zaprte in neprilagodljive rešitve. Takšni sistemi so običajno dobavljeni kot vnaprej konfiguriran programski paket enega prodajalca, z omejenim dostopom do notranjih podatkovnih zbirk in procesne logike.

Ponudniki sistemov CAD-(BIM-) vse bolj prevzemajo razvoj in nadzor takih sistemov, saj njihove zbirke podatkov vsebujejo informacije, ki jih zahtevajo sistemi ERP: količinske in prostorninske atribute elementov projekta. Vendar pa prodajalci namesto dostopa do teh podatkov v odprtih ali strojno berljivih oblikih ponujajo le omejene uporabniške scenarije in zaprto logiko obdelave - vnaprej določeno v modulih BlackBox. To zmanjšuje prilagodljivost sistema in preprečuje, da bi ga prilagodili posebnim projektnim pogojem.

Omejena preglednost podatkov ostaja eden od ključnih izzivov digitalnih procesov v gradbeništvu. Zaprti arhitektura podatkovnih zbirk, pomanjkanje dostopa do popolnih sklopov atributov gradbenih elementov, osredotočenost na *module* za avtomatizacijo *BlackBox* in pomanjkanje odprtih vmesnikov znatno povečujejo tveganja birokracije dokumentov. Takšne omejitve ustvarjajo ozka grla v procesu odločanja, otežujejo preverjanje informacij in odpirajo vrata skrivanju podatkov ali špekulacijam v sistemih ERP/PMIS. Uporabniki navadno dobijo le omejen dostop - naj bo to okrnjen vmesnik ali delni API - brez možnosti neposredne interakcije s primarnimi viri podatkov. To je še posebej kritično, ko gre za parametre, samodejno ustvarjene iz projektov CAD, kot so volumni, površine in količine, ki se uporabljajo za izračune QTO.

Posledično se številna gradbena podjetja namesto iskanja učinkovitosti z avtomatizacijo procesov, odprtimi podatki, zmanjševanjem transakcijskih stroškov in oblikovanjem novih poslovnih modelov osredotočajo na upravljanje zunanjih parametrov - manipulativnih dejavnikov, faktorjev prilagajanja in metod izračuna, ki vplivajo na stroške projekta v zaprtih platformah ERP/PMIS. To ustvarja prostor za špekulacije, izkrivilja dejanske proizvodne stroške in zmanjšuje zaupanje med vsemi udeleženci v gradbenem procesu.

V gradbeništvu se dobiček oblikuje kot razlika med prihodki iz dokončanega projekta in spremenljivimi stroški, ki vključujejo stroške projektiranja, materiala, dela in druge neposredne stroške, neposredno povezane z izvedbo projekta. Vendar ključni dejavnik, ki vpliva na vrednost teh stroškov, ni le tehnologija ali logistika, temveč tudi hitrost in natančnost izračunov ter kakovost vodstvenih odločitev v podjetju.

Težava je še večja, ker so v večini gradbenih podjetij postopki izračuna stroškov nepregledni ne le za stranke, temveč tudi za zaposlene, ki niso del ocenjevalnega ali finančnega oddelka. Takšna zaprtost spodbuja oblikovanje privilegirane skupine strokovnjakov - nosilcev "finančnega znanja", ki imajo izključno pravico urejati atributte in korekcijske faktorje v sistemih ERP/PMIS. Ti zaposleni lahko skupaj z vodji podjetij dejansko nadzorujejo finančno logiko projekta.

V takšnih razmerah se ocenjevalci spremenijo v "finančne žonglerje", ki balansirajo med maksimiranjem dobička podjetja in potrebo po ohranjanju konkurenčne cene za naročnika. Hkrati se morajo izogibati očitnim in grobim manipulacijam, da ne bi ogrozili ugleda podjetja. Na tej stopnji se določijo koeficienti, ki prikrijejo precenjene količine ali stroške materiala in del.

Zato glavna shema za povečanje učinkovitosti in dobičkonosnosti podjetij, ki delujejo v gradbeništvu, ni avtomatizacija in pospešitev procesov odločanja, temveč špekulacije s cenami materialov in del (slika 5.4-5). Precenjevanje stroškov del in materiala se izvaja s sivim računovodstvom v zaprtih sistemih ERP /PMIS - tako, da se odstotki nad povprečnimi tržnimi cenami materialov ali količinami del napihnejo s pomočjo koeficientov (slika 5.1-6), ki so bili obravnavani v poglavju "Sestava kalkulacij in izračun stroškov del na podlagi baze virov".

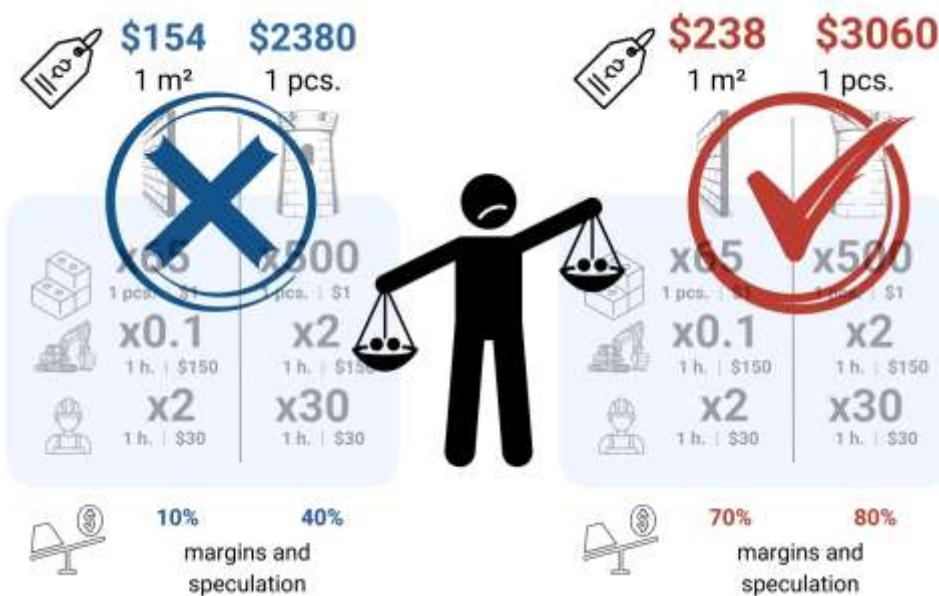
Zato stranka prejme izračun, ki ne odraža dejanskih stroškov ali obsega dela, temveč je izpeljanka številnih skritih notranjih koeficientov. Hkrati so podizvajalci, da bi zadostili podcenjenim cenam, ki jih je določil generalni izvajalec, pogosto prisiljeni kupovati cenejše in manj kakovostne materiale, kar poslabša končno kakovost gradnje.

Špekulativni postopek iskanja dobička iz zraka na koncu škoduje tako strankam, ki prejemajo nezanesljive podatke, kot tudi izvršnim direktorjem, ki so prisiljeni iskati vedno več modelov za špekulacije

Večji kot je projekt, večja je stopnja birokracije pri upravljanju podatkov in procesov. Vsak korak in vsak modul pogosto skriva nepregledne koeficiente in pribitke, vgrajene v algoritme izračunov in notranje postopke. To ne le otežuje revizijo, temveč tudi znatno izkrivilja finančno sliko projekta. Pri velikih gradbenih projektih takšne prakse pogosto vodijo do večkratnega (včasih tudi do desetkratnega) povečanja končnih stroškov, medtem ko dejanski obseg in stroški ostajajo zunaj učinkovitega nadzora naročnika (slika 2.1-3 Primerjava načrtovanih in dejanskih stroškov velikih infrastrukturnih projektov v Nemčiji).

Po podatkih poročila družbe McKinsey & Company *Imagining the Digital Future of Construction* (2016) so veliki gradbeni projekti v povprečju zaključeni 20% pozneje od načrtovanega in do 80% prekoračijo proračun [107].

Oddelki za ocenjevanje in pripravo proračuna postajajo najbolj varovani člen v podjetju. Dostop do njih je strogo omejen celo za notranje strokovnjake, zaradi zaprte logike in struktur podatkovnih baz pa je nemogoče objektivno oceniti učinkovitost projektnih odločitev brez izkrivljanja. Pomanjkanje preglednosti vodi do tega, da podjetja niso prisiljena optimizirati procesov, temveč se za preživetje borijo s "kreativnim" upravljanjem številk in faktorjev (slika 5.3-1, slika 5.1-6 - na primer parameter "Bid. faktor").



Slika 5.4-5 Razmerja špekulacij na ravni poravnave so glavni dobiček podjetij in umetnost žongliranja med kakovostjo dela in ugledom.

Vse to postavlja pod vprašaj nadaljnjo uporabo zaprtih sistemov ERP/PMIS v gradbeništvu. V kontekstu digitalne preobrazbe in vse večjih zahtev strank po preglednosti (slika 10.2-3) je malo verjetno, da bo izvajanje projektov dolgoročno ostalo odvisno od zaprtih rešitev, ki omejujejo prilagodljivost, ovirajo integracijo in zavirajo razvoj poslovanja.

Ne glede na to, kako ugodno je za gradbena podjetja delo s podatkovnimi silosi in nepreglednimi podatki v zaprtih zbirkah podatkov - prihodnost gradbene industrije bo neizogibno vključevala prehod na odprte platforme, strojno berljive in pregledne podatkovne strukture ter avtomatizacijo, ki temelji na zaupanju. Ta preobrazba bo potekala od zgoraj - pod pritiskom strank, regulatorjev in družbe, ki vse bolj zahtevajo odgovornost, trajnost, preglednost in ekonomsko upravičenost.

## Konec obdobja zaprtih sistemov ERP /PMIS: gradbena industrija potrebuje nove pristope

Uporaba obsežnih modularnih sistemov ERP/PMIS, ki so sestavljeni iz več deset milijonov vrstic kode, izjemno oteže kakršne koli spremembe v njih. V tem primeru se prehod na novo platformo ob prisotnosti modulov, ki so že vnaprej konfigurirani za podjetje, več deset tisoč člankov v podatkovnih bazah virov (slika 5.1-3) in več tisoč pripravljenih izračunov (slika 5.1-6) spremeni v drag in dolgotrajen proces. Več kot je kode in starejše arhitekture - višja je stopnja notranje neučinkovitosti, vsak nov projekt pa stvari le še poslabša. V številnih podjetjih postaneta migracija podatkov in integracija novih rešitev večletni epopeji, ki ju spremljata nenehno popravljanje in neskončno iskanje kompromisov. Posledica tega je pogosto vrnitev k starim, zanim platformam, kljub njihovim omejitvam.

Kot je poudarjeno v nemškem poročilu Black Book [108] o sistemskih napakah pri upravljanju podatkov v gradbeništvu, sta razdrobljenost informacij in pomanjkanje centraliziranega pristopa k njihovemu upravljanju ključna vzroka neučinkovitosti. Brez standardizacije in integracije podatki izgubijo svojo vrednost in postanejo arhiv in ne orodje za upravljanje.

Glavni vzrok za izgubo kakovosti podatkov je neustrezno načrtovanje in nadzor gradbenih projektov, kar pogosto povzroči znatno povečanje stroškov. V poglavju črne knjige z naslovom "Osrednji poudarek: Eksplozija stroškov" so analizirani ključni dejavniki, ki prispevajo k tem nezaželenim posledicam. Ti vključujejo neustrezne analize potreb, pomanjkanje študij izvedljivosti in neuskajeno načrtovanje, kar vodi do dodatnih stroškov, ki bi se jim lahko izognili.

V zrelem ekosistemu IT podjetja je zamenjava zastarelega sistema primerljiva z zamenjavo nosilnega stebra v že zgrajeni stavbi. Ni dovolj le odstraniti starega in namestiti novega - pomembno je, da to storimo tako, da stavba ostane stabilna, da se stropi ne zrušijo in da vse komunikacije še naprej delujejo. Prav v tem je težava: vsaka napaka ima lahko resne posledice za celoten sistem podjetja.

Kljub temu razvijalci velikih izdelkov ERP za gradbeno industrijo še vedno uporabljajo količino napisane kode kot argument v prid svoji platformi. Na specializiranih konferencah je še vedno mogoče slišati stavke, kot so: "Za poustvaritev takšnega sistema bi potrebovali 150 človeških let", čeprav se za večino funkcionalnosti takšnih sistemov skrivajo zbirke podatkov in precej preproste funkcije za delo s tabelami, zapakirane v poseben fiksni uporabniški vmesnik. V praksi se obseg kode "150 človeških let" spremeni v breme in ne v konkurenčno prednost. Več kot je kode - višji so stroški podpore, teže se je prilagajati novim razmeram in višji je vstopni prag za nove razvijalce in stranke.

Številni modularni gradbeni sistemi danes spominjajo na okorne in zastarele "Frankensteinove konstrukcije", pri katerih lahko vsaka neprevidna sprememba privede do okvare. Vsak nov modul poveča kompleksnost že tako preobremenjenega sistema in ga spremeni v labirint, ki ga lahko razume le nekaj strokovnjakov, kar še dodatno oteže vzdrževanje in posodobitev.

Kompleksnosti se zavedajo tudi razvijalci sami, ki se občasno ustavijo pri refaktorizaciji - pregledu arhitekture, da bi upoštevali pojav novih tehnologij. Vendar se zapletenost neizogibno povečuje, tudi če se refaktorizacija izvaja redno. Arhitekti takšnih sistemov se na naraščajočo zapletenost navadijo, za nove uporabnike in strokovnjake pa postane nepremostljiva ovira. Posledično se vse strokovno znanje in izkušnje skoncentrirajo v rokah nekaj razvijalcev, sistem pa preneha biti skalabilen. Kratkoročno so takšni strokovnjaki koristni, dolgoročno pa postanejo del problema.

*Organizacije bodo še naprej združevale "majhne" podatke z njihovimi velikimi podatki, zato je neumno verjeti, da lahko ena sama aplikacija - pa naj bo še tako draga ali zanesljiva - obvlada vse [109].*

- Phil Simon, gostitelj podkasta Conversations About Collaboration

Postavlja se upravičeno vprašanje: ali res potrebujemo tako okorne in zaprte sisteme za izračunavanje stroškov in časovnega okvira dela v obliki tabel, če so druge industrije že dolgo sposobne opravljati podobne naloge z analitičnimi orodji z odprtimi podatki in pregledno logiko?

Trenutno so zaprte modularne platforme v gradbeništvu še vedno iskane, predvsem zaradi posebnosti stroškovnega računovodstva (slika 5.1-7). Takšni sistemi se pogosto uporabljajo za vodenje sivih ali nepreglednih shem, kar omogoča, da so dejanski stroški skriti pred naročnikom. Ker pa industrija digitalno dozoreva, predvsem stranke, in prehaja v tako imenovano "uberizirano dobo", bodo posredniki, torej gradbena podjetja s svojimi ERP, izgubili svoj pomen pri izračunavanju časa in stroškov. To bo za vedno spremenilo podobo gradbene industrije. Več si preberite v zadnjem delu knjige in v poglavju "Gradbeništvo 5.0: Kako zaslužiti, ko se ne morete več skriti".

Na tisoče starejših rešitev, ki so se nabrale v zadnjih 30 letih in v razvoj katerih je bilo vloženih na tisoče človeških let, bo začelo hitro izginjati. Prehod na odprto, pregledno in prilagodljivo upravljanje podatkov je neizogiven. Edino vprašanje je, katera podjetja se bodo sposobna prilagoditi tem spremembam in katera bodo ostala talci starega modela.

Podobno velja za orodja CAD (BIM), katerih podatki danes zapolnjujejo volumetrične parametre projektantov v sistemih ERP/PMIS. Sprva je ideja BIM (razvita že leta 2002 [110]) temeljila na konceptu enotne integrirane podatkovne zbirke, v praksi pa danes delo z BIM zahteva celo vrsto specializirane programske opreme in formatov. Kar naj bi poenostavilo upravljanje projektiranja in gradnje, se je spremenilo v še en sloj lastniških rešitev, ki otežujejo integracijo in zmanjšujejo prilagodljivost poslovanja.

## Naslednji koraki: učinkovita uporaba podatkov o projektih

V tem delu smo pokazali, kako strukturirani podatki postanejo osnova za natančne izračune stroškov in časovnega načrta gradbenih projektov. Avtomatizacija postopkov QTO, načrtovanja in ocenjevanja zmanjša stroške dela in znatno izboljša natančnost rezultatov.

Za povzetek tega dela velja izpostaviti glavne praktične korake, ki vam bodo pomagali pri uporabi obravnavanih pristopov pri vsakodnevnih nalogah. Ti pristopi so univerzalni - uporabni so tako za digitalno preobrazbo podjetja kot za vsakodnevno delo strokovnjakov, vključenih v izračune:

■ **Avtomatizacija rutinskih izračunov**

- Poskusite najti standardne stroške dela, ki jih lahko uporabite pri svojem delu.
- Analizirajte, katere metode se v vaši državi uporablajo za izračun stroškov ali stroškov del ali postopkov na gradbišču (slika 5.1-7).
- Če delate s sistemom CAD - spoznajte funkcije samodejnega pridobivanja specifikacij in podatkov QTO v programski opremi CAD (BIM-).
- Uporabite program LLM za pisanje osnutka kode za avtomatizacijo izračunov.

■ **Razvoj lastnih orodij za QTO**

- Ustvarjanje skript ali tabel za avtomatizacijo štetja volumnov
- standardizacija kategorij in skupin elementov za dosleden pristop k ocenjevanju
- dokumentiranje metodologije izračunov za zagotovitev ponovljivosti rezultatov v novih projektih.

■ **Vključite različne vidike projekta v svoje delo.**

- Če delate z modularnimi sistemi, poskušajte svoje procese vizualizirati ne le v oblikih diagramov ali grafov, temveč tudi na ravni podatkov - zlasti v oblikih tabel.
- Obvladajte samodejno združevanje podatkov, pridobljenih iz podatkovnih zbirk CAD, z izračuni - s kodo Python z uporabo združevanja v skupine, filtriranja in agregiranja
- Ustvarjanje jasnih vizualizacij skupin QTO za predstavitev zapletenih informacij sodelavcem in strankam.

Ti koraki bodo pomagali vzpostaviti trajnostni sistem izračuna, ki bo temeljil na avtomatizaciji in standardizaciji podatkov. Ta pristop bo izboljšal natančnost in zmanjšal rutino vsakodnevnih vprašanj v zvezi z izračuni.

Naslednja poglavja se osredotočajo na tehnične vidike izdelkov CAD - (BIM-) in razloge, zakaj je podatkovne zbirke CAD še vedno težko vključiti v poslovne procese podjetij. Če vas zdaj ne zanimajo zgodovina uvajanja BIM v gradbeništvu, razvoj orodij CAD in tehnične podrobnosti dela s temi tehnologijami, lahko preidete naravnost na sedmi del knjige "Odločanje na podlagi podatkov".



## NAJVEČJE UDOBJE S TISKANO RAZLIČICO

V rokah imate brezplačno digitalno različico knjige **Data-Driven Construction**. Za udobnejše delo in hiter dostop do gradiva vam priporočamo, da se posvetite **tiskani izdaji**:



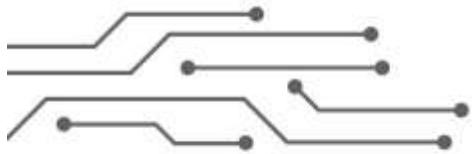
**Vedno pri roki:** knjiga v tiskani obliki bo postala zanesljivo delovno orodje, s katerim boste lahko hitro našli in uporabili prave vizualizacije in diagrame v vsaki delovni situaciji.

**Visoka kakovost ilustracij:** vse slike in grafike v tiskani izdaji so predstavljene v najvišji kakovosti.

**Hiter dostop do informacij:** enostavna navigacija, možnost beleženja opomb, zaznamkov in dela s knjigo kjer koli.

orodje za udobno in učinkovito delo z informacijami: možnost hitre uporabe vizualnega gradiva pri vsakodnevnih opravilih, hitrega iskanja potrebnih shem in zapisovanja. Poleg tega s svojim nakupom podpirate širjenje odprtega znanja.

Tiskano različico knjige naročite na: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## VI DEL

# **CAD IN BIM: TRŽENJE, RESNIČNOST IN PRIHODNOST PROJEKTNIH PODATKOV V GRADBENIŠTVU**

V šestem delu knjige je predstavljena kritična analiza razvoja tehnologij CAD in BIM ter njihovega vpliva na procese upravljanja podatkov v gradbeništvu. Sledi zgodovinski preobrazbi koncepta BIM od prvotne zamisli o integrirani podatkovni zbirki do sedanjih tržnih konstrukcij, ki jih spodbujajo prodajalci programske opreme. Ocenjen je vpliv lastniških formatov in zaprtih sistemov na učinkovitost ravnanja s projektnimi podatki in splošno uspešnost gradbene industrije. Podrobno so analizirani problemi združljivosti različnih sistemov CAD in težave pri njihovem povezovanju s poslovnimi procesi gradbenih podjetij. Obravnavani so sedanji trendi v smeri poenostavljenih odprtih podatkovnih formatov, kot je USD, in njihov potencialni vpliv na industrijo. Predstavljeni so alternativni pristopi k pridobivanju informacij iz zaprtih sistemov, vključno s tehnikami obratnega inženiringa. Analizirane so možnosti uporabe umetne inteligence in strojnega učenja za avtomatizacijo postopkov načrtovanja in analize podatkov v gradbeništvu. Oblikovane so napovedi za razvoj tehnologij projektiranja, ki so usmerjene k dejanskim potrebam uporabnikov in ne k interesom prodajalcev programske opreme.

## POGLAVJE 6.1.

### POJAV KONCEPTOV BIM- V GRADBENIŠTVU

Prvotno ta šesti del, namenjen CAD (BIM), ni bil vključen v prvo različico knjige. Teme o lastniških formatih, geometrijskih jedrih in zaprtih sistemih so preveč tehnične, preobremenjene s podrobnostmi in navidezno neuporabne za tiste, ki želijo le razumeti, kako delati s podatki. Vendar so povratne informacije in prošnje za dodana pojasnila k prvi različici knjige pokazale, da brez razumevanja zapletenosti notranjega delovanja sistemov CAD, geometrijskih jeder, različnih formatov in nezdružljivih shem za shranjevanje istih podatkov ni mogoče resnično razumeti, zakaj koncepti, ki jih spodbujajo prodajalci, pogosto otežujejo delo s podatki in ovirajo prehod na odprto parametrično načrtovanje. Zato je ta del zavzel samostojno mesto v strukturi knjige. Če CAD (BIM) za vas ni prednostna naloga, lahko preskočite naravnost na naslednji del - "DEL VII: Odločanje na podlagi podatkov, analitika, avtomatizacija in strojno učenje".

#### Zgodovina pojava BIM in odprtega BIM kot tržnih konceptov prodajalcev CAD-

S pojavom digitalnih podatkov v 90. letih prejšnjega stoletja je bila računalniška tehnologija uvedena ne le v poslovne procese, temveč tudi v procese projektiranja, kar je privedlo do konceptov, kot sta CAD (računalniško podprtji sistemi projektiranja) in pozneje BIM (informacijsko modeliranje stavb)

Vendar kot vsaka inovacija tudi te niso končna točka razvoja. Koncepti, kot je BIM, so postali pomemben mejnik v zgodovini gradbene industrije, vendar bodo prej ali slej morda odstopili mesto boljšim orodjem in pristopom, ki bodo bolje kos izzivom prihodnosti.

Koncept BIM, ki se je pojavil leta 2002, zaradi vpliva prodajalcev CAD in zapletenosti lastnega izvajanja morda ne bo dočakal svoje tridesete obletnice, kot rock zvezda, ki je močno zasijala, a hitro ugasnila. Razlog je preprost: zahteve podatkovnih znanstvenikov se spreminjajo hitreje, kot se jim lahko prilagodijo ponudniki CAD.

Današnji strokovnjaki v gradbeništvu, ki se soočajo s pomanjkanjem kakovostnih podatkov, zahtevajo interoperabilnost med platformami in dostop do odprtih podatkov iz projektov CAD-, da bi poenostavili njihovo analizo in obdelavo. Zapletenost podatkov CAD in zmedena obdelava podatkov CAD negativno vplivata na vse, ki sodelujejo v gradbenem procesu: projektante, vodje projektov, gradbene delavce na gradbišču in navsezadnje na naročnika.

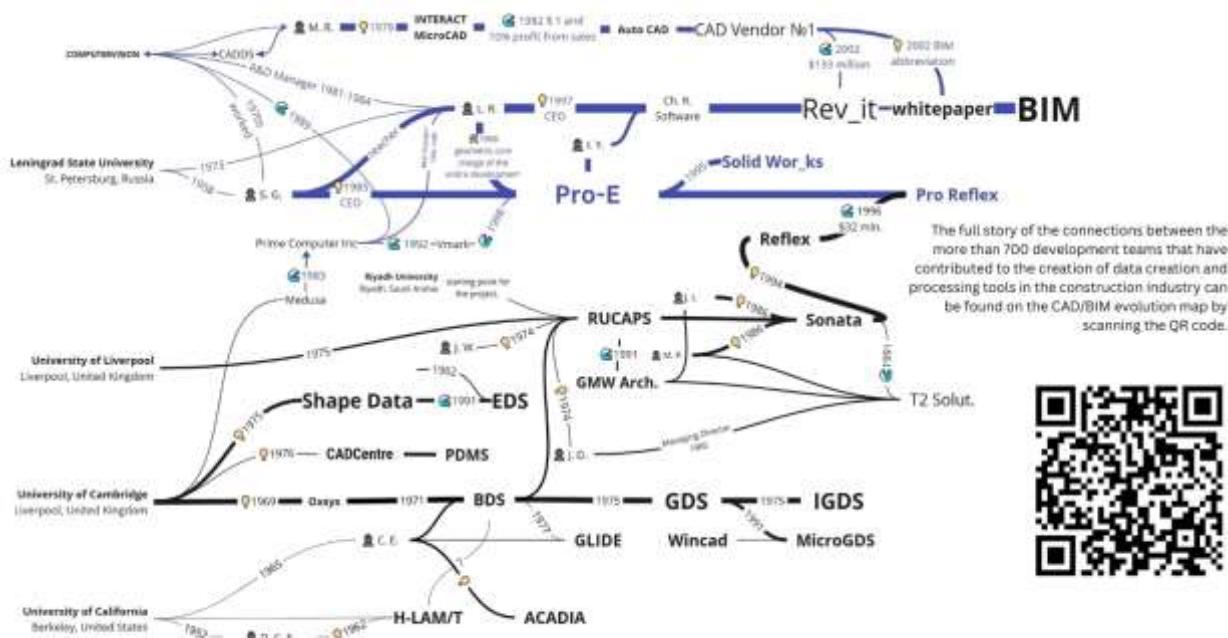
Namesto popolnega nabora podatkov za delovanje danes stranka in investitor prejmeta zabojnike v formatih CAD-, ki za delo s podatki zahtevajo kompleksna geometrijska jedra, razumevanje podatkovnih shem, letno posodobljeno dokumentacijo API in specializirano programsko opremo CAD (BIM). Hkrati pa velik del podatkov o projektiranju ostaja neizkoriščen.

V današnjem svetu projektiranja in gradnje zapletenost dostopa do podatkov CAD vodi do pretiranega inženiringa pri upravljanju projektov. Srednja in velika podjetja, ki delajo s podatki CAD ali razvijajo rešitve BIM, so za dostop do podatkov prek vmesnikov API prisiljena vzdrževati tesne odnose s

ponudniki rešitev CAD ali pa obiti omejitve ponudnikov CAD z uporabo dragih pretvornikov SDK za povratno inženirstvo, da bi pridobila odprte podatke [75].

Pristop z lastniškimi podatki je zastarel in ne ustreza več zahtevam današnjega digitalnega okolja. Prihodnost bo podjetja razdelila na dve vrsti: tista, ki bodo učinkovito uporabljala odprte podatke, in tista, ki bodo zapustila trg.

Koncept BIM (Building Information Modeling) se je v gradbeništvu pojavil z objavo Whitepaper BIM [54] enega od glavnih proizvajalcev CAD leta 2002 in je, dopolnjen s strojniškim konceptom BOM (Bills of Materials), izhajal iz parametričnega pristopa k ustvarjanju in obdelavi projektnih podatkov (slika 6.1-1). Parametrični pristop k oblikovanju in obdelavi projektnih podatkov je bil med prvimi implementirani v sistemu Pro-E za strojniško projektiranje (MCAD). Ta sistem je postal prototip [111] številne sodobne rešitve CAD, vključno s tistimi, ki se danes uporabljajo v gradbeništvu.



Slika 6.1-1 Zemljevid zgodovine koncepta BIM in podobnih konceptov.

Novinarji in svetovalci za AEC, ki so do zgodnjih let 2000 promovirali orodja CAD -prodajalci, so od leta 2002 preusmerili svojo pozornost na bele knjige BIM. Prav Bela knjiga BIM 2002-2004 in članki, objavljeni v letih 2002, 2003, 2005 in 2007, so imeli ključno vlogo pri popularizaciji koncepta BIM v gradbeništvu [112].

*Informacijsko modeliranje stavb je strategija..... [ime podjetja, ki prodaja CAD] za uporabo informacijske tehnologije v gradbeništvu.*

- Bela knjiga BIM, 2002 [60]

Sredi leta 2000 so "raziskovalci" začeli povezovati koncept BIM-, ki ga je leta 2002 objavil prodajalec CAD-, s prejšnjimi znanstvenimi deli, kot je BDS Charlesa Eastmana , ki so postala podlaga za sisteme, kot so GLIDE, GBM, BPM, RUCAPS. Charles Eastman je v svojem pionirskem delu Building Description System (1974) postavil teoretične temelje sodobnega informacijskega modeliranja. Izraz "podatkovna zbirka " se v njegovem delu pojavi 43-krat (slika 6.1-2) - pogosteje kot katerikoli drug, razen besede "stavba".

Eastmanova ključna zamisel je bila, da je treba vse informacije o stavbi - od geometrije do lastnosti elementov in njihovih medsebojnih povezav - shraniti v enotni strukturirani zbirki podatkov. Iz te zbirke podatkov je mogoče samodejno ustvariti in analizirati risbe, specifikacije, izračune in skladnost s predpisi. Eastman je izrecno kritiziral risbe kot zastarello in nepotrebno metodo komunikacije, pri čemer je opozoril na podvajanje informacij, težave s posodabljanjem in potrebo po ročnih posodobitvah ob spremembah. Namesto tega je predlagal enoten digitalni model v podatkovni zbirki, kjer se vsaka sprememba izvede enkrat in se samodejno odrazi v vseh pogledih.

Omeniti velja, da Eastman v svojem konceptu vizualizacije ni postavil v središče. V njegovem sistemu so bile osrednjega pomena informacije: parametri, razmerja, atributi, možnosti analize in avtomatizacije. Risbe so bile v njegovem pojmovanju le ena od oblik prikaza podatkov iz podatkovne zbirke, ne pa primarni vir informacij o načrtovanju.

V prvi beli knjigi o BIM vodilnega ponudnika CAD je bila besedna zveza "podatkovna baza " uporabljena tako pogosto kot v BDS Charlesa Eastmana - 23-krat [60] na sedmih straneh in je bila ena izmed najbolj priljubljenih besed v dokumentu za besedami "zgradba", "informacije", "modeliranje" in "oblikovanje". Vendar se do leta 2003 izraz "podatkovna baza" v podobnih dokumentih pojavi le dvakrat [61], do konca leta 2000 pa je tema podatkovnih baz praktično izginila iz razprav o podatkih o projektiranju. Zato koncept "enotne integrirane podatkovne zbirke za vizualno in kvantitativno analizo" ni bil nikoli v celoti uresničen.

Tako je gradbena industrija prešla od naprednega koncepta BDS Charlesa Eastmana s poudarkom na podatkovnih bazah in zamisli Samuela Geisberga o samodejnem posodabljanju projektnih podatkov iz podatkovnih baz v strojniškem izdelku Pro-E (predhodnik priljubljenih rešitev CAD, ki se danes uporablajo v gradbeništvu) do trenutno tržnega BIM, kjer je upravljanje podatkov prek podatkovnih baz komaj omenjeno, čeprav je bil to koncept prvotne teoretične

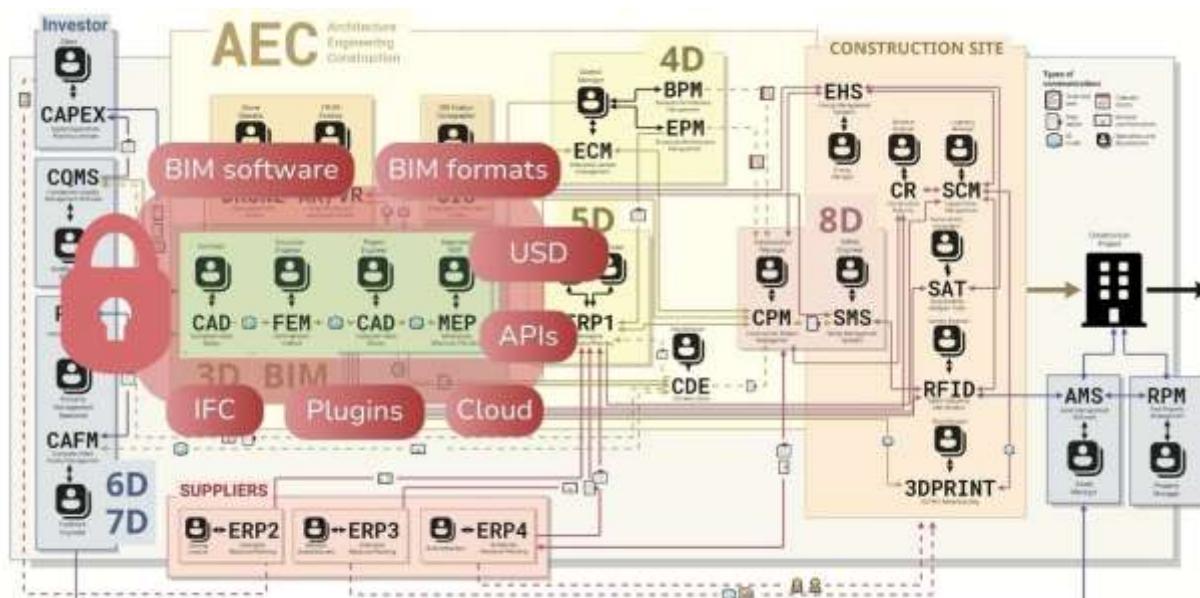


■ **Podatki pripadajo prodajalcem in ne uporabnikom.** Podatki o projektih so zaklenjeni v lastniških formatih ali storitvah v oblaku, namesto da bi bili na voljo v odprtih in neodvisnih formatih.

Projektanti in vodje projektov pogosto nimajo dostopa do podatkovnih zbirk CAD -sistemov niti do formata, v katerem so shranjeni njihovi lastni projektni podatki. Zaradi tega je nemogoče hitro preveriti informacije ali oblikovati zahteve glede strukture in kakovosti podatkov (slika 6.1-3). Za dostop do takšnih podatkov je potreben cel niz specializiranih programov, povezanih prek vmesnikov API in vtičnikov, kar vodi v pretirano birokratizacijo procesov v gradbeništvu. Medtem te podatke hkrati uporablja na desetine informacijskih sistemov in na stotine strokovnjakov.

*Vse te podatke [CAD (BIM)] moramo znati upravljati, jih shranjevati v digitalni obliki ter prodajati programsko opremo za upravljanje življenjskega cikla in procesov, saj **na vsakega inženirja** [projektanta], ki nekaj ustvari [v programski opremi CAD], **pride deset ljudi, ki delajo s temi podatki**" [41].*

- Generalni direktor podjetja CAD - ponudnika, ki je ustvaril koncept BIM, 2005.



Slika 6.1-3 Podatkovne baze CAD- (BIM-) ostajajo eden zadnjih zaprtih sistemov za oddelke IT in upravljalce podatkov v ekosistemu gradbenih podjetij.

Ko postane jasno, da je BIM bolj sredstvo za komercializacijo podatkovnih zbirk kot pa polnopravno orodje za upravljanje podatkovnih zbirk, se pojavi logično vprašanje: kako lahko ponovno pridobimo nadzor nad podatki? Odgovor je uporaba odprtih podatkovnih struktur, kjer lastnik podatkov postane uporabnik in ne prodajalec programske opreme.

Uporabniki in razvijalci rešitev v gradbeništvu se bodo, tako kot njihovi kolegi v drugih panogah, neizogibno odmaknili od nejasne terminologije proizvajalcev programske opreme, ki je prevladovala v zadnjih 30 letih, in se osredotočili na ključne vidike digitalizacije - "podatke" in "procese".

V poznih 80. letih prejšnjega stoletja je bilo ključno področje digitalnega razvoja v gradbeništvu predstavljeno kot vprašanje dostopa do podatkov in upravljanja projektnih informacij. Sčasoma pa se je težišče spremenilo. Namesto razvoja preglednih in dostopnih pristopov k delu s podatki sta se aktivno promovirala format IFC in odprt koncept BIM kot poskusa odvračanja pozornosti strokovnjakov od teme upravljanja projektnih podatkovnih baz.

## Pojav odprtega formata IFC v gradbeništvu

Tako imenovani odprt format IFC (Industry Foundation Classes) je postavljen kot standard za zagotavljanje interoperabilnosti med različnimi sistemi CAD (BIM -). Njegov razvoj je potekal v okviru organizacij, ki so jih ustanovili in nadzorovali glavni ponudniki CAD. Na podlagi formata IFC sta dve podjetji CAD- leta 2012 razvili tržni koncept OPEN BIM[63].

**IFC** (Industry Foundation Classes) je odprt standard za izmenjavo podatkov v gradbeništvu, namenjen zagotavljanju interoperabilnosti med različnimi sistemi CAD - (BIM-).

Koncept odprtega BIM - vključuje delo z informacijami iz podatkovnih zbirk CAD in izmenjavo informacij med sistemi prek odprtega formata za izmenjavo podatkov CAD - IFC.

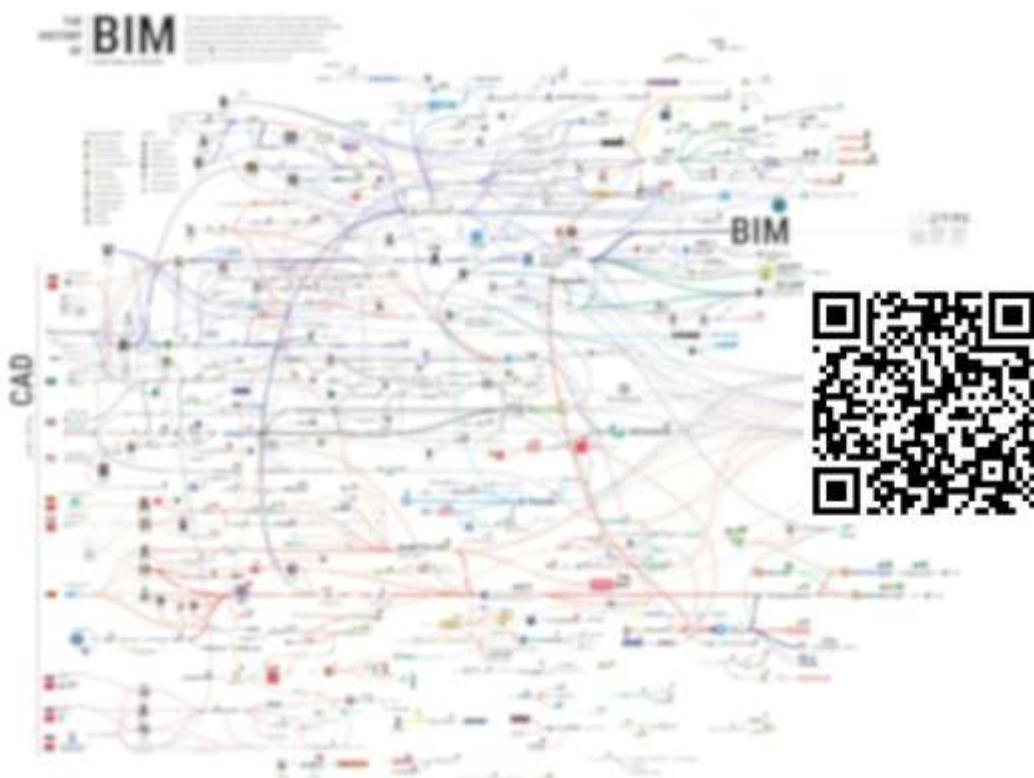
*Program Open BIM je marketinška kampanja, ki so jo začeli... [1 prodajalec CAD]... [2 prodajalec CAD] in druga podjetja, da bi spodbudili in olajšali globalno usklajeno promocijo koncepta OPEN BIM v celotni industriji AEC z dosledno komunikacijo in skupno blagovno znamko, ki je na voljo udeležencem programa.*

- S spletne strani prodajalca CAD, program OPEN BIM, 2012 [113]

IFC je Tehnična univerza v Münchnu konec osemdesetih let prejšnjega stoletja prilagodila iz strojniškega formata STEP, pozneje pa sta ga registrirala večje oblikovalsko podjetje in večji prodajalec CAD- ter leta 1994 ustanovila združenje IAI (Industry Alliance for Interoperability) [114] (slika 6.1-4). Format IFC je bil razvit za zagotavljanje interoperabilnosti med različnimi sistemi CAD in je temeljil na načelih, določenih v strojniškem formatu STEP, ki pa je nastal iz formata IGES, ki ga je leta 1979 ustvarila skupina uporabnikov in prodajalcev CAD s podporo NIST (The National Institute of Standards and Technology) in Ministrstva za obrambo ZDA [115].

Vendar so zapletena struktura IFC, njegova tesna odvisnost od geometrijskega jedra ter razlike pri izvajanju formata v različnih programskih rešitvah povzročili številne težave pri njegovi praktični uporabi. S podobnimi težavami - izgubo podrobnosti, omejitvijo natančnosti in potrebo po uporabi vmesnih formatov - so se strokovnjaki s področja strojništva že prej srečevali pri delu s formati IGES

in STEP, iz katerih je izšel IFC.



Slika 6.1-4 Zemljevid povezav med razvojnimi skupinami in izdelki CAD (BIM) [116].

Leta 2000 isti ponudnik CAD, ki je registriral format IFC in ustanovil organizacijo IAI (pozneje bS), objavi belo knjigo "Integrated Design and Manufacturing: Benefits and Rationale". [65]. V dokumentu je poudarjen pomen ohranjanja popolne granularnosti podatkov pri izmenjavi med programi znotraj istega sistema brez uporabe nevtralnih formatov, kot sta IGES, STEP [enaka IFC]. Namesto tega je bilo predlagano, da bi morali imeti programi neposreden dostop do osnovne podatkovne zbirke CAD, da se prepreči izguba natančnosti informacij.

Leta 2002 isti prodajalec CAD kupi parametrični izdelek BOM (slika 3.1-18, več podrobnosti v tretjega dela) in na njegovi podlagi oblikuje koncept BIM. Posledično se pri izmenjavi podatkov o gradbenih projektih uporablajo le zaprti formati CAD ali format IFC (STEP), o omejitvah katerega je leta 2000 pisal že sam prodajalec CAD, ki je ta format uvedel v gradbeništvo.

Podrobna zdobovina sodelovanja več kot 700 razvojnih skupin, ki so sodelovale pri ustvarjanju orodij za ustvarjanje in obdelavo gradbenih podatkov, je predstavljena na zemljevidu "Razvoj CAD (BIM)" [116]. [116].

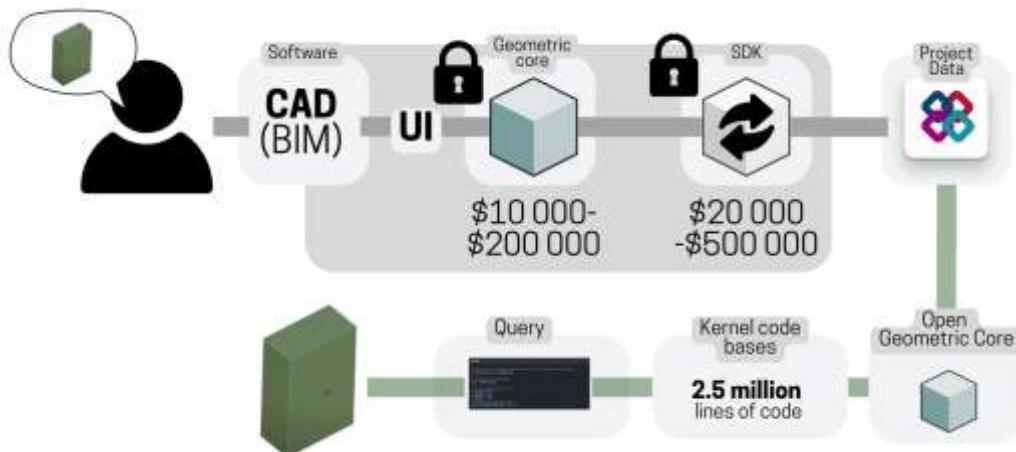
Odperta oblika IFC je sestavljena iz geometrijskega opisa konstrukcijskih elementov in opisa metainformacij. Za predstavitev geometrije v obliki IFC se uporabljajo različne metode, na primer CSG in Swept Solids: vendar je parametrična predstavitev BREP postala vodilni standard za prenos

geometrije elementov v obliki IFC, saj je ta oblika podprta pri izvozu iz programov CAD- (BIM-) in omogoča morebitno urejanje elementov pri uvozu IFC nazaj v programe CAD.

## Težava s formatom IFC glede na geometrijsko jedro

V večini primerov, ko je geometrija v datoteki IFC opredeljena parametrično (BREP), je nemogoče vizualizirati ali pridobiti geometrijske lastnosti, kot sta prostornina ali površina enot projekta, samo z datoteko IFC, saj je za delo z geometrijo in njenou vizualizacijo v tem primeru potrebno geometrijsko jedro (slika 6.1-5), ki ga na začetku ni.

**Jedro geometrije** je programska komponenta, ki zagotavlja osnovne algoritme za ustvarjanje, urejanje in analizo geometrijskih objektov v programih CAD (CAD), BIM in drugih inženirskeih aplikacijah. Odgovorno je za izdelavo 2D- in 3D- geometrije ter za operacije z njo, kot so: logične operacije, glajenje, presečišča, transformacije in vizualizacija.

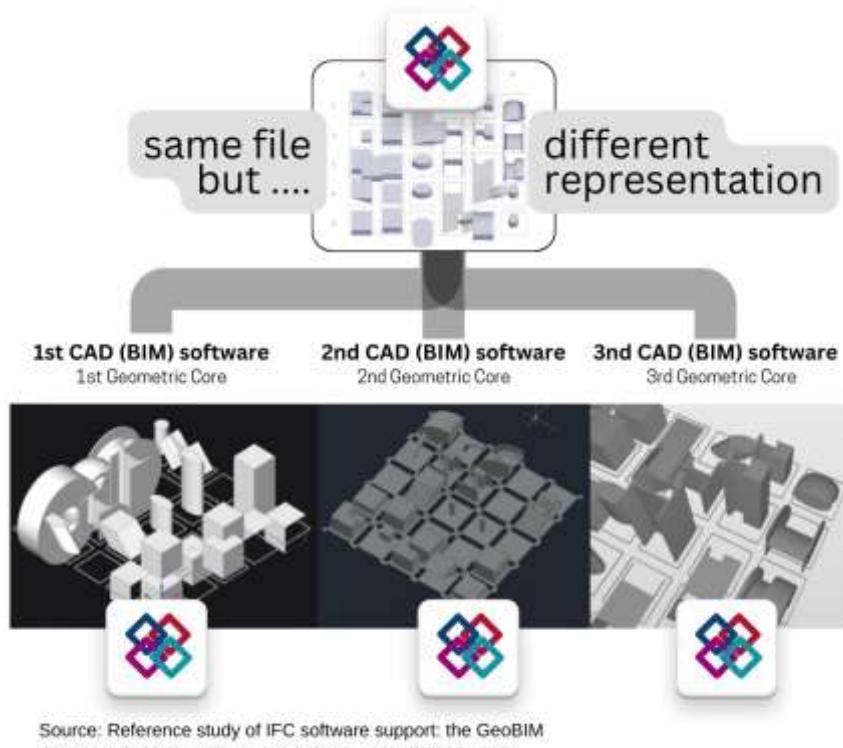


Slika 6.1-5 Ustvarjanje geometrije v programski opremi CAD- danes poteka prek lastniških geometrijskih jeder in SDK-jev, ki pogosto niso v lasti ponudnikov CAD.

Vsak program CAD in vsi programi, ki delajo s parametričnimi formati ali formati IFC, imajo lastno ali kupljeno geometrijsko jedro. In če s primitivnimi elementi v formatu IFC -BREP ne more biti težav in je v programih z različnimi geometrijskimi jedri te elemente mogoče prikazati podobno, pa je poleg težav z različnimi motorji geometrijskih jeder, dovolj elementov, ki imajo svoje posebnosti za pravilno prikazovanje. Ta problem je podrobno obravnavan v mednarodni študiji " A reference study of IFC software support", objavljeni leta 2019 [117].

*Isti standardizirani podatkovni nizi dajejo nasprotujoče si rezultate, saj bilo ugotovljenih le malo skupnih vzorcev, pri podpori standarda [IFC] pa so bile ugotovljene resne težave, verjetno zaradi zelo velike zapletenosti standardnega podatkovnega modela. Pri tem so delno krivi tudi sami standardi, saj pogosto puščajo nekatere podrobnosti nedefinirane, z veliko stopnjo svobode in različnimi možnimi razlagami. Omogočajo veliko kompleksnost pri organizaciji in shranjevanju objektov, kar ne prispeva k učinkovitemu univerzalnemu razumevanju, edinstvenim izvedbam in doslednemu modeliranju podatkov [117].*

- Referenčna študija o podpori programske opreme IFC, 2021



**Slika 6.1-6 Različna geometrijska jedra dajejo različne predstavitev iste geometrije, opisane parametrično (na podlagi [117]).**

Pravilno razumevanje "nekaterih določb" je na voljo plačanim članom posebnih organizacij, ki razvijajo IFC. Posledično bo tisti, ki želi priti do pomembnega znanja o določenih lastnostih IFC, poskušal sodelovati z velikimi ponudniki CAD- ali pa bo do kakovostne obravnave lastnosti prišel z lastnimi raziskavami

*Naletite na vprašanje o uvozu in izvozu podatkov prek formata IFC in ga zastavite drugim prodajalcem: "Zakaj je v datoteki IFC informacija o parametričnem prenosu prostorov? Odprta specifikacija o tem ne pove ničesar". Odgovor "bolj ozaveščenih" evropskih prodajalcev: "Da, to ni povedano, vendar je dovoljeno".*

- Iz intervjuja razvijalca CAD 2021 [118]

IFC opisuje geometrijo s parametričnimi primitivi, vendar ne vsebuje vgrajenega jedra - njegovo vlogo opravi program CAD, ki sestavi geometrijo s pomočjo geometrijskega jedra. Jedro geometrije izvaja matematične izračune in določa presečišča, medtem ko IFC zagotavlja le podatke za njegovo interpretacijo. Če IFC vsebuje napačne površine, jih lahko različni programi z različnimi geometrijskimi jedri prezrejo ali pa glede na jedro povzročijo napake.

Zato je za delo z IFC formatom treba odgovoriti na glavno vprašanje, na katerega je težko najti nedvoumen odgovor - katero orodje, s katerim geometrijskim jedrom je treba uporabiti, da dobimo kakovost podatkov, ki jo je projekt prvotno imel v programu CAD, iz katerega je bil pridobljen IFC?

Vprašanja kakovosti podatkov in zapletenost formata IFC ne omogočajo neposredne uporabe projektnih podatkov za avtomatizacijo procesov, analizo in obdelavo podatkov, kar razvijalce pogosto pripelje do neizogibne potrebe po uporabi zaprtih rešitev CAD-rešitev s "kakovostnim" dostopom do podatkov [63], o čemer je pisal sam prodajalec, ki je IFC registriral leta 1994 [65].

Vse posebnosti preslikave in generiranja parametrov IFC v geometrijskem jedru lahko uresničijo le velike skupine razvijalcev, ki imajo izkušnje z delom z geometrijskimi jedri. Zato je trenutna praksa posebnosti in zapletenosti formata IFC koristna predvsem za prodajalce CAD- in ima veliko skupnega s strategijo velikih prodajalcev programske opreme "posvoji, razširi, uniči", ko vse večja zapletenost standarda dejansko ustvarja ovire za male tržne udeležence [94].

Strategija velikih prodajalcev pri takšni strategiji je lahko prilagoditev odprtih standardov, dodajanje lastniških razširitev in funkcij, da bi ustvarili odvisnost uporabnikov od svojih izdelkov in tako izrinili konkurente.

Format IFC, ki naj bi bil univerzalni most med različnimi sistemi CAD- (BIM-), ima v resnici vlogo kazalnika težav z združljivostjo med geometrijskimi jedri različnih platform CAD, podobno kot format STEP, iz katerega je prvotno izšel.

Zato je danes popolna in kakovostna izvedba ontologije IFC izvedljiva za velike ponudnike CAD, ki lahko vložijo veliko sredstev za podporo vseh entitet in njihovo preslikavo v lastno notranje geometrijsko jedro, ki za standard IFC ne obstaja. Veliki prodajalci imajo tudi možnost, da med seboj usklajujejo tehnične podrobnosti funkcij, ki morda niso na voljo niti najbolj aktivnemu udeležencu v organizacijah za razvoj formata IFC.

Za majhne neodvisne skupine in odprtakodne projekte, ki si prizadevajo podpirati razvoj interoperabilnih formatov, postane pomanjkanje lastnega geometrijskega jedra resna težava. Brez njega je praktično nemogoče upoštevati vse številne subtilnosti in nianse, povezane z izmenjavo podatkov med platformami.

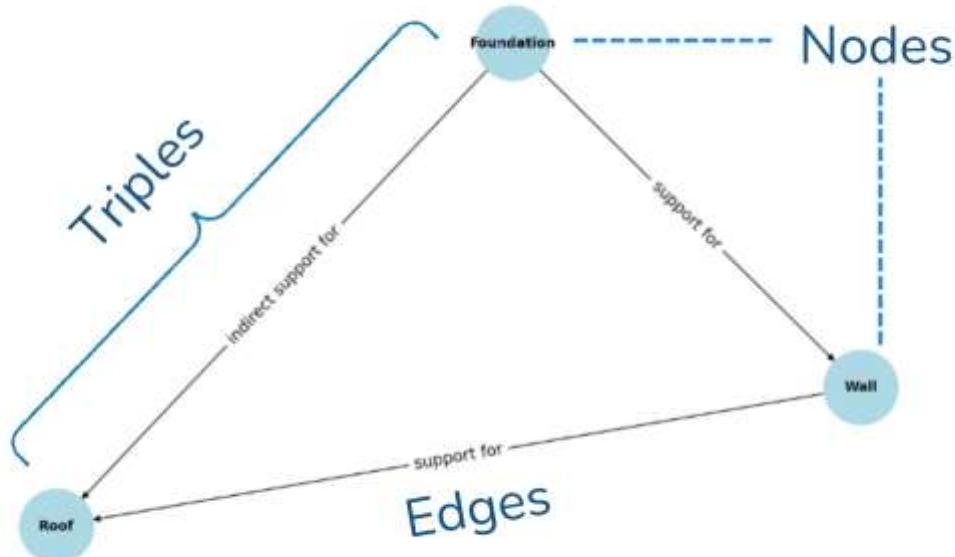
Z razvojem parametričnega formata IFC in odprtega koncepta BIM so se v gradbeništvu okrepile razprave o vlogi ontologije in semantike pri upravljanju podatkov in procesov.

## Pojavljanje v konstrukciji teme semantike in ontologije

Zahvaljujoč zamislom o semantičnem internetu konec devetdesetih let prejšnjega stoletja in prizadevanjem organizacij, vključenih v razvoj formata IFC, so semantika in ontologije postali eni od ključnih elementov standardizacije, o kateri se bo v gradbeni industriji razpravljalo do sredine 20. stoletja.

Semantične tehnologije so poenotenje, standardizacija in spreminjanje velikih nizov heterogenih podatkov ter izvajanje zapletenih iskanj.

OWL (Web Ontology Language), ki je predstavljen kot graf RDF-triplets (Resource Description Framework) (slika 6.1-7), se uporablja za shranjevanje semantičnih podatkov. OWL se nanaša na grafne podatkovne modele, katerih vrste smo podrobnejše obravnavali v poglavju "Podatkovni modeli: podatkovna razmerja in razmerja med elementi".



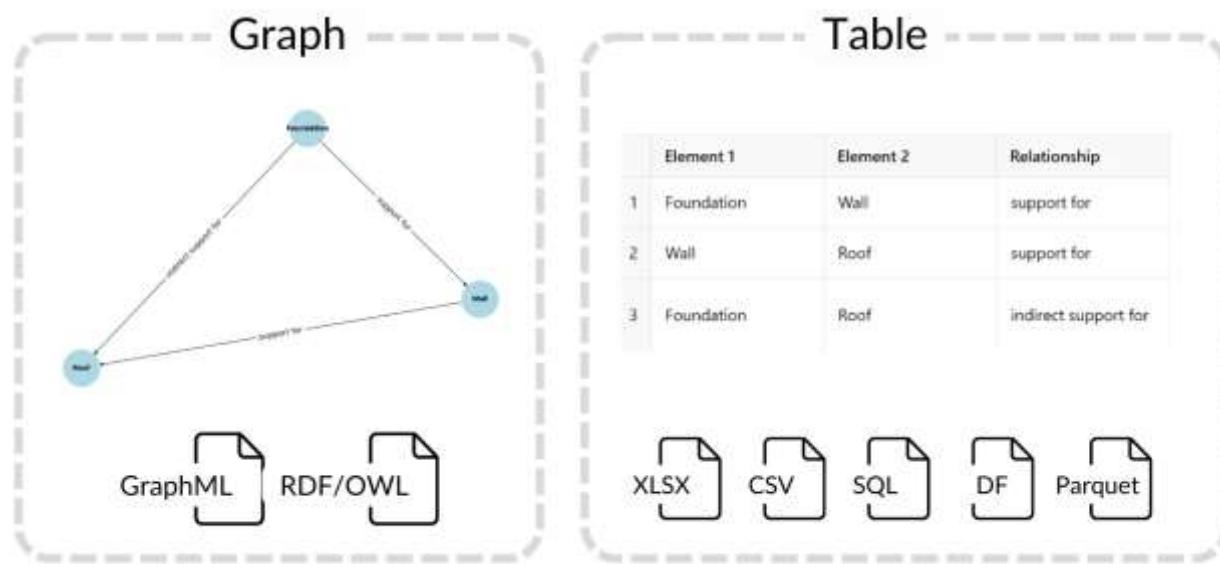
**Slika 6.1-7 Podatkovni model RDF: vozlišča, robovi in trojčki, ki ponazarjajo odnose med gradniki.**

Teoretično lahko z logičnim sklepanjem risonerjev (programska oprema za samodejno logično sklepanje) iz ontologij izpeljemo nove izjave. Če je na primer v ontologiji stavb zapisano, da je "temelj opora za steno" in "stena opora za streho" (slika 6.1-7), lahko risoner samodejno sklepa, da je "temelj opora za streho".

Tak mehanizem je uporaben za optimizacijo analize podatkov, saj se izogne izrecnemu predpisovanju vseh odvisnosti. Vendar pa ne ustvarja novega znanja, temveč le identificira in strukturira že znana dejstva.

Semantika sama po sebi ne ustvarja novega pomena ali znanja in v tem pogledu ni boljša od drugih tehnologij za shranjevanje in obdelavo podatkov. Če podatke iz relacijskih podatkovnih zbirk predstavimo kot trojčke, jih ne naredimo bolj smiselne. Zamenjava tabel z grafnimi strukturami je morda koristna za poenotenje podatkovnih modelov, enostavno iskanje in varno urejanje, vendar zaradi tega podatki ne postanejo "pametnejši" - računalnik ne začne bolje razumeti njihove vsebine.

Logične povezave v podatkih je mogoče organizirati brez zapletenih semantičnih tehnologij (slika 6.1-8). Tradicionalne relacijske podatkovne zbirke (SQL) ter formati CSV ali XLSX omogočajo gradnjo podobnih odvisnosti. V stolčni podatkovni zbirki lahko na primer dodate polje "podpora strehi" in pri ustvarjanju zidu samodejno povežete streho s temelji. Ta pristop se izvaja brez uporabe RDF, OWL, grafov ali rizonerjev in ostaja preprosta in učinkovita rešitev za shranjevanje in analizo podatkov.



Slika 6.1-8 Primerjava podatkovnih modelov grafov in tabel za predstavitev istih logičnih povezav.

Odločitev številnih velikih gradbenih podjetij in organizacije za razvoj formata IFC [94], da sledijo konceptu semantičnega spletja, ki se je konec 90. let prejšnjega stoletja zdel obetaven, je pomembno vplivala na razvoj standardov v gradbeništvu.

Paradoksalno pa je, da se sam koncept semantičnega spletja, ki je bil prvotno namenjen internetu, ni široko uveljavil niti v svojem domačem okolju. Kljub razvoju RDF in OWL se polnopravni semantični splet v svoji prvotni zasnovi še ni pojavit, njegova vzpostavitev pa je že malo verjetna.

### Zakaj semantične tehnologije v gradbeništvu ne izpolnjujejo pričakovanj

V drugih panogah se soočajo z omejitvami tehnologij za uporabo semantike. V industriji iger na srečo so se poskusi opisovanja predmetov iger in njihovih interakcij z ontologijami zaradi velike dinamike sprememb izkazali za neučinkovite. Posledično so bili v ospredju preprostejši podatkovni formati, kot sta XML in JSON, skupaj z algoritemskimi rešitvami. Podobno je bilo v nepremičninskem sektorju: zaradi regionalnih razlik v terminologiji in pogostih sprememb na trgu se je uporaba ontologij izkazala za preveč zapleteno, medtem ko so se preproste zbirke podatkov in standardi, kot je RETS [119], bolje spopadali z izzivi izmenjave podatkov.

Tehnične težave, kot so zapletenost označevanja, visoka delovna intenzivnost podpore in nizka motivacija razvijalcev, so upočasnilo sprejetje semantičnega spletja in v drugih gospodarskih sektorjih. RDF (Resource Description Framework) ni postal množični standard, ontologije pa so se izkazale za preveč zapletene in ekonomsko neupravičene.

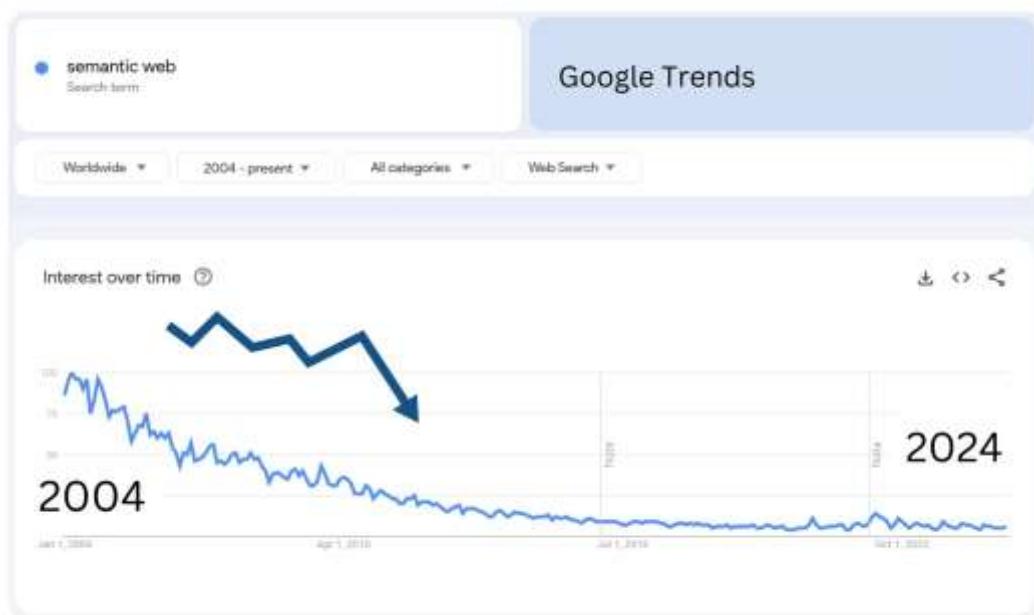
Zato se ambiciozna zamisel o vzpostaviti globalnega semantičnega spletja ni uresničila. Čeprav so nekateri elementi tehnologije, kot sta ontologije in SPARQL, našli pot v rešitve za podjetja, prvotni cilj oblikovanja enotne celovite podatkovne strukture ni bil dosežen.

Koncept interneta, v katerem lahko računalniki razumejo vsebino, se je izkazal za tehnično zahtevnega in komercialno nerentabilnega. Zato so podjetja, ki so podpirala to zamisel, sčasoma omejila njen uporabo na posamezna uporabna orodja, tako da sta RDF in OWL ostala za zelo specializirane potrebe podjetij in ne za internet kot celoto. Analiza Googlovin trendov (slika 6.1-9) v zadnjih 20 letih kaže, da za semantični splet morda ni več obetov.

*Po nepotrebнем ni treba množiti entitet. Če obstaja več logično skladnih razlag nekega pojava, ki ga enako dobro pojasnjujejo, je treba ob enakih pogojih dati prednost najpreprostejši izmed njih.*

britev

Pri tem se pojavi logično vprašanje: zakaj pri gradnji sploh uporabljati trojčke, riserje in SPARQL, če lahko podatke obdelujemo s priljubljenimi strukturiranimi poizvedbami (SQL, Pandas, Apache®)? V podjetniških aplikacijah je SQL standard za delo s podatkovnimi zbirkami. SPARQL pa nasprotno zahteva zapletene grafne strukture in specializirano programsko opremo ter glede na trende v Googlu ne pritegne zanimaanja razvijalcev.



Slika 6.1-9 Zanimanje za poizvedbe po "semantičnem internetu" glede na Googlovo statistiko.

Podatkovne baze grafov in klasifikacijska drevesa so lahko v nekaterih primerih koristni, vendar njihova uporaba ni vedno upravičena za večino vsakodnevnih nalog. Zato je oblikovanje grafov znanja in uporaba tehnologij semantičnega spleta smiselna le, kadar je treba poenotiti podatke iz različnih virov ali uresničiti zapletene logične skele.

Prehod s tabel na grafične podatkovne modele izboljša iskanje in poenoti pretok informacij, vendar pa podatki ne postanejo bolj smiseln za stroje. Vprašanje ni, ali je treba uporabljati semantične tehnologije, temveč kje so resnično pomembne. Preden v svojem podjetju uvedete ontologijo, semantiko in grafne podatkovne baze, ugotovite, katera podjetja te tehnologije že uspešno uporablajo in kje jim ni uspelo.

Klub ambicioznim pričakovanjem semantične tehnologije nikoli niso postale univerzalna rešitev za strukturiranje podatkov v gradbeništvu. V praksi te tehnologije niso pripeljale do univerzalne rešitve, temveč so le dodale nove zaplete, ta prizadevanja pa odražajo neuresničene ambicije koncepta semantičnega interneta, kjer so pričakovanja močno presegla realnost.

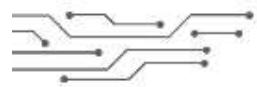


Slika 6.1-10 Geometrija in informacijske procese: od kompleksnih sistemov CAD in BIM- do poenostavljenih podatkov za analitiko.

Medtem ko so v IT neuspehe semantičnega spletja nadomestili s pojavom novih tehnologij (veliki podatki, internet stvari, strojno učenje, AR/VR), gradbeništvo nima takih priložnosti.

Poleg izzivov uporabe konceptov za sporočanje podatkovnih povezav med elementi projekta ostaja temeljna težava - sama razpoložljivost teh podatkov. V gradbeništvu še vedno prevladujejo zaprti sistemi, kar otežuje delo s podatki, izmenjavo informacij in izboljšanje učinkovitosti procesov.

Prav zaprta narava podatkov postaja ena od ključnih ovir za razvoj digitalnih rešitev v gradbeništvu. V nasprotju z industrijo IT, kjer so odprti in usklajeni podatkovni formati postali standard, v sektorju CAD (BIM) vsaka programska oprema uporablja svoj format, kar ustvarja zaprte ekosisteme in umetno omejuje uporabnike.



## POGLAVJE 6.2.

### ZAPRTI FORMATI PROJEKTOV IN VPRAŠANJA INTEROPERABILNOSTI.

#### Zaprti podatki in padajoča produktivnost: slepa ulica industrije CAD (BIM)

Zaradi lastniške narave sistemov CAD -sistemov ima vsak program svoj edinstveni podatkovni format, ki je bodisi zaprt in od zunaj nedostopen - RVT, PLN, DWG, NDW, NWD, SKP, ali pa je na voljo v polstrukturirani obliki s precej zapletenim postopkom pretvorbe JSON, XML (CPIXML), IFC, STEP in ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV itd..

Različni podatkovni formati, v katerih so lahko shranjeni isti podatki o istih projektih, se ne razlikujejo le po strukturi, temveč vključujejo tudi različne različice notranjega označevanja, kar morajo razvijalci upoštevati, da zagotovijo združljivost aplikacij. Na primer, format CAD iz leta 2025 se bo odprl v programu CAD iz leta 2026, vendar se isti projekt nikoli ne bo odprl v vseh različicah programa CAD, ki so bile morda na voljo pred letom 2025.

Ker ponudnik programske opreme v gradbeništvu ne zagotavlja neposrednega dostopa do podatkovnih zbirk, pogosto ustvari svoj lasten edinstven format in orodja, ki jih mora strokovnjak (projektant ali upravljavec podatkov) uporabiti za dostop, uvoz in izvoz podatkov.

Posledično prodajalci osnovnih sistemov CAD (BIM) in povezanih rešitev (npr. ERP/PMIS)) nenehno zvišujejo cene za uporabo izdelkov, navadni uporabniki pa so prisiljeni plačati "provizijo" na vsaki stopnji prenosa podatkov po formatih [63]: za povezovanje, uvoz, izvoz in delo s podatki, ki so jih uporabniki ustvarili sami.

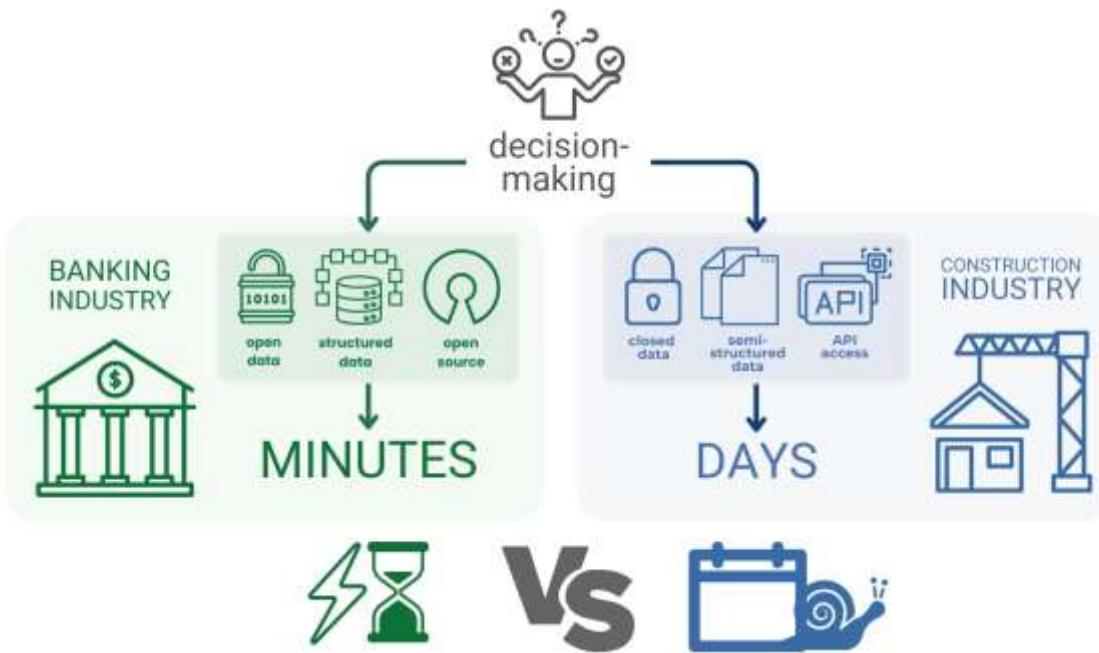
Stroški dostopa do podatkov v oblačni shrambi iz priljubljenih izdelkov CAD (BIM-) bodo leta 2025 dosegli 1 dolar na transakcijo [120], naročnine na gradbene ERP izdelke za srednje velika podjetja pa dosegajo pet- in šestmestne zneske na leto [121].

Bistvo sodobne programske opreme za gradbeništvo je, da na kakovost in stroške obdelave podatkov o gradbenih projektih ter na dobiček in dolgoročno preživetje podjetij, ki izvajajo gradbene projekte, ne vpliva avtomatizacija ali večja učinkovitost, temveč sposobnost inženirjev, da razumejo določeno visoko specializirano programsko opremo.

Pomanjkanje dostopa do baz podatkov CAD -sistemov, ki se uporablajo v več deset drugih sistemih in več sto procesih [63], in posledično pomanjkanje kakovostne komunikacije med posameznimi strokovnjaki sta gradbeništvo pripeljala do statusa enega najbolj neučinkovitih sektorjev gospodarstva v smislu produktivnosti [44].

V zadnjih 20 letih uporabe aplikacij za projektiranje CAD- (BIM-), pojava novih sistemov (ERP), novih gradbenih tehnologij in materialov se je produktivnost celotne gradbene industrije zmanjšala za 20% (slika 2.2-1), medtem ko se je skupna produktivnost vseh gospodarskih sektorjev, ki nimajo večjih

težav z dostopom do podatkovnih baz in trženjem podobnih konceptov BIM, povečala za 70% (96% v predelovalni industriji) [122].



Slika 6.2-1 Zaradi izoliranosti in kompleksnosti projektnih podatkov, od katerih je v gradbeništvu odvisnih več deset oddelkov in več sto procesov, je hitrost odločanja nekajkrat počasnejša kot v drugih panogah.

Vendar pa obstajajo tudi posamezni primeri alternativnih pristopov k ustvarjanju interoperabilnosti med rešitvami CAD. Največje evropsko gradbeno podjetje s projektom SCOPE [123], ki se je začel že leta 2018, dokazuje, kako je mogoče preseči klasično logiko sistemov CAD- (BIM-). Namesto da bi si poskušali podrediti IFC ali se zanašali na lastniška geometrijska jedra, razvijalci SCOPE z uporabo API-jev in SDK-jev povratnega inženiringa pridobivajo podatke iz različnih programov CAD, jih pretvarjajo v neutralne formate, kot sta OBJ ali CPIXML, ki temeljijo na edinem odprtakodnem geometrijskem jedru OCCT, in jih nadalje uporabljajo v več sto poslovnih procesih gradbenih in projektantskih podjetij. Kljub naprednosti zamisli pa se takšni projekti soočajo z omejitvami in zapletenostjo brezplačnih geometrijskih jeder in še vedno ostajajo del zaprtih ekosistemov enega podjetja, ki reproducirajo logiko enoprodajnih rešitev.

Zaradi omejitev zaprtih sistemov in razlik v podatkovnih formatih ter pomanjkanja učinkovitih orodij za njihovo poenotenje se podjetja, ki morajo delati z oblikami CAD, soočajo s kopiranjem velikih količin podatkov z različno stopnjo strukture in zaprtosti. Ti podatki se ne uporabljajo pravilno in izginejo v arhivih, kjer ostanejo za vedno pozabljeni in neuporabljeni.

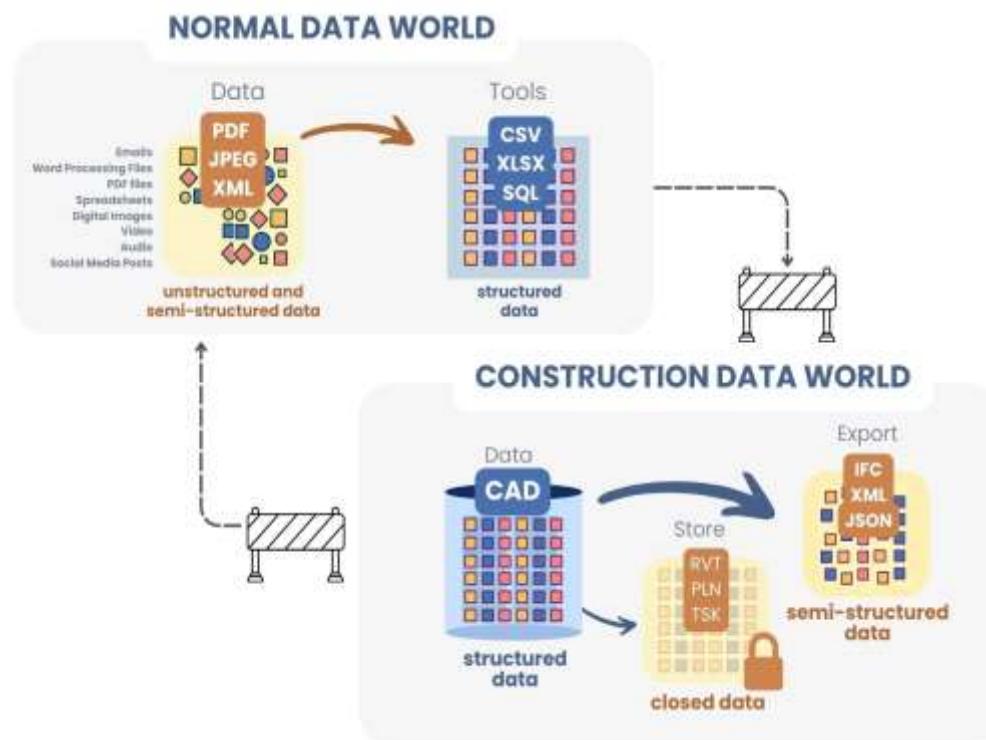
Podatki, pridobljeni z veliko truda v fazi načrtovanja, postanejo zaradi svoje kompleksnosti in zaprtosti nedostopni za nadaljnjo uporabo.

V zadnjih 30 letih so se zato razvijalci v gradbeništvu vedno znova soočali z isto težavo: vsak nov zaprt

format ali lastniška rešitev je povzročila potrebo po integraciji z obstoječimi odprtimi in zaprtimi sistemami CAD. Ti nenehni poskusi zagotavljanja interoperabilnosti med različnimi rešitvami CAD in BIM le zapletajo podatkovni ekosistem, namesto da bi prispevali k njegovi poenostavitvi in standardizaciji.

## Mit o interoperabilnosti med sistemi CAD

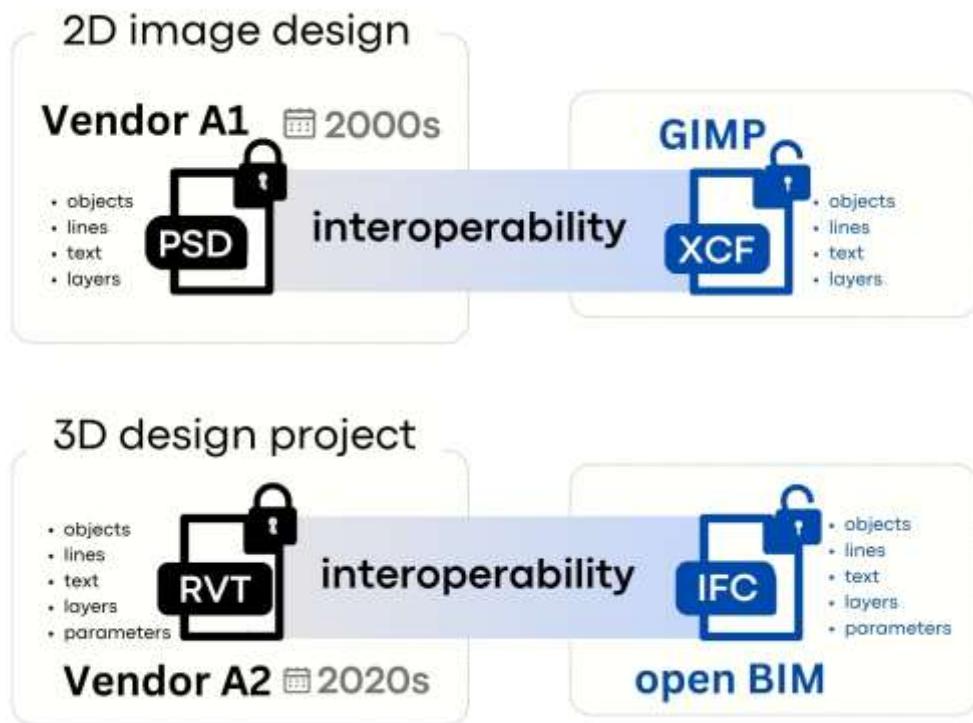
Medtem ko je bila sredi devetdesetih let prejšnjega stoletja ključna usmeritev razvoja interoperabilnosti v okolju CAD razbijanje lastniškega formata DWG - kar je doseglo vrhunec z zmago zavezništva Open DWG [75] in dejanskim odprtjem najbolj priljubljenega formata risb za celotno gradbeno industrijo -, se je sredi dvajsetih let poudarek preusmeril. V gradbeništvu se uveljavlja nov trend: številne razvojne skupine se osredotočajo na ustvarjanje tako imenovanih "mostov" med zaprtimi sistemi CAD (zaprti BIM), formatom IFC in odprtimi rešitvami (odprti BIM). Večina teh pobud temelji na uporabi formata IFC in geometrijskega jedra OCCT, kar zagotavlja tehnični most med različnimi platformami. Ta pristop velja za obetavno smer, ki lahko bistveno izboljša izmenjavo podatkov in interoperabilnost programskih orodij.



Slika 6.2-2 Medtem ko druge panoge delajo z odprtimi podatki, mora gradbena industrija delati z zaprtimi ali ohlapno strukturiranimi formati CAD (BIM).

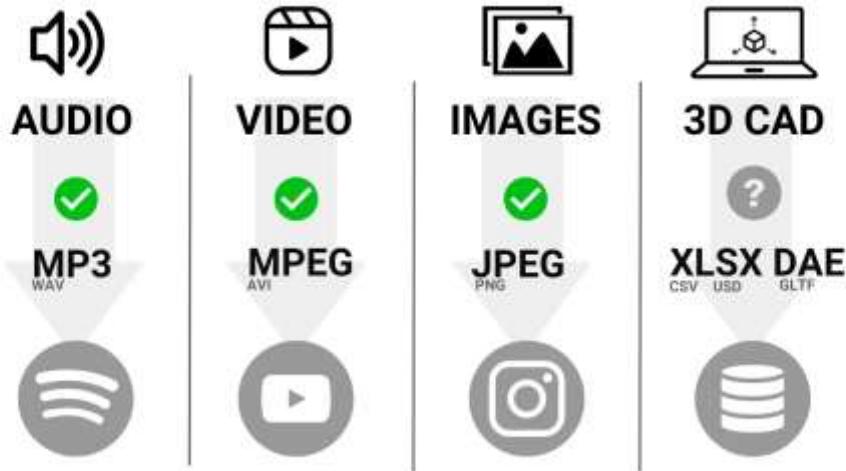
Takšen pristop ima zgodovinske vzporednice. Leta 2000 so razvijalci, ki so skušali premagati prevlado največjega prodajalca grafičnih urejevalnikov (2D world), poskušali ustvariti nemoteno integracijo med svojo lastniško rešitvijo in brezplačno odprto kodo - alternativo GIMP-u (slika 6.2-3). Takrat, tako kot danes pri gradnji, je šlo za poskus povezovanja zaprtih in odprtih sistemov ob ohranjanju zapletenih parametrov, slojev in notranje programske logike.

Vendar so uporabniki dejansko iskali preproste rešitve - ravne, odprte podatke brez pretirane zapletenosti slojev in programskih parametrov (analogije geometrijskega jedra v CAD). Uporabniki so iskali preproste in odprte podatkovne formate brez pretirane logike. Takšni formati so postali JPEG, PNG in GIF v grafiki. Danes se uporablajo v družabnih omrežjih, na spletnih straneh, v aplikacijah - enostavno jih je obdelati in interpretirati ne glede na platformo ali proizvajalca programske opreme.



Slika 6.2-3 Interoperabilnost podatkovnih formatov v gradbeništvu je podobna poti od poskusov združitve lastniškega izdelka priljubljenega prodajalca in odprtakodnega programa GIMP v letu 2000.

Zato danes skoraj nihče v industriji upodabljanja ne uporablja zaprtih formatov, kot sta PSD ali odprti XCF, za aplikacije, družbena omrežja, kot sta Facebook in Instagram, ali za vsebino na spletnih straneh. Namesto tega večina nalog zaradi enostavne uporabe in široke združljivosti uporablja ploščate in odprte formate JPEG, PNG in GIF. Odprti formati, kot sta JPEG in PNG, so zaradi svoje vsestransnosti in široke podpore postali standard za izmenjavo slik, saj jih je enostavno uporabljati na različnih platformah. Podoben prehod je mogoče opaziti pri drugih formatih za izmenjavo, kot sta video in zvok, kjer so univerzalni formati, kot sta MPEG in MP3, poudarjeni zaradi svoje učinkovitosti stiskanja in široke združljivosti. Takšen prehod k standardizaciji je poenostavil izmenjavo in predvajanje vsebin in informacij ter jih naredil dostopne vsem uporabnikom na različnih platformah (slika 6.2-4).



Slika 6.2-4 Poenostavljeni formati brez zapletenih funkcij urejanja so postali priljubljeni za izmenjavo in uporabo podatkov.

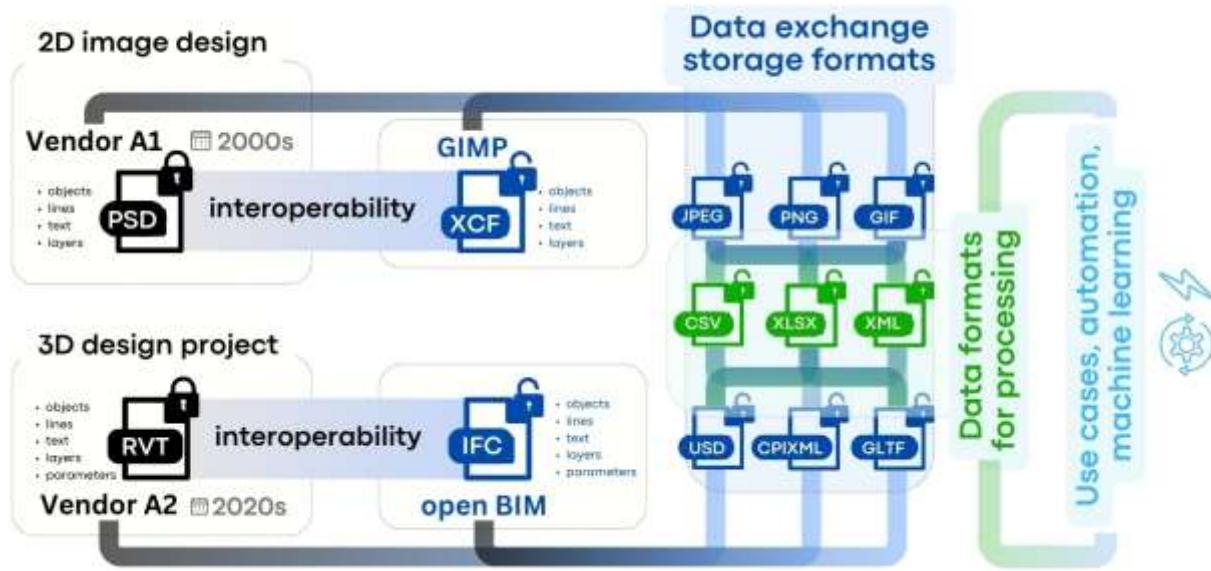
Podobni procesi se pojavljajo tudi pri 3D modeliranju. Enostavni in odprti formati, kot so USD, OBJ, glTF, DAE, DXF, SQL in XLSX, se v projektih vse pogosteje uporabljajo za izmenjavo podatkov zunaj okolja CAD (BIM). V teh formatih so shranjene vse potrebne informacije, vključno z geometrijo in metapodatki, brez potrebe po uporabi zapletene strukture BREP, geometrijskih jeder ali notranjih klasifikatorjev, specifičnih za posameznega dobavitelja. Lastniški formati, kot so NWC, SVF, SVF2, CPIXML in CP2, ki jih zagotavljajo vodilni proizvajalci programske opreme, prav tako opravljajo podobne funkcije, vendar za razliko od odprtih standardov ostajajo zaprti.

Omeniti velja (in ponovno opozoriti, kot je bilo že omenjeno v prejšnjem poglavju), da je to zamisel - zavrnitev vmesnih nevtralnih in parametričnih formatov, kot so IGES, STEP in IFC - že leta 2000 podprt glavni prodajalec CAD, ki je pripravil BIM Whitepaper in leta 1994 registriral format IFC. V beli knjigi iz leta 2000 "Integrated Design and Manufacturing" (Integrirano načrtovanje in proizvodnja) [65] prodajalec CAD poudarja pomen izvornega dostopa do podatkovne zbirke CAD v programskem okolju, brez potrebe po uporabi vmesnih prevajalnikov in parametričnih formatov, da se ohranita popolnost in točnost informacij.

Gradbena industrija se še ni dogovorila niti o orodjih za dostop do podatkovnih zbirk CAD ali njihovem prisilnem povratnem inženiringu, niti o sprejetju skupnega poenostavljenega podatkovnega formata za uporabo zunaj platform CAD (BIM). Številna velika podjetja v Srednji Evropi in nemško govorečih regijah, ki delujejo v gradbenem sektorju, na primer uporabljajo format CPIXML v svojih sistemih ERP [121]. Ta lastniški format, ki je neke vrste XML, združuje projektne podatke CAD (BIM), vključno z geometrijskimi in metapodatki, v enoto organizirano poenostavljenou strukturo. Tudi velika gradbena podjetja ustvarjajo nove lastne formate in sisteme, kot v projektu SCOPE, ki smo ga obravnavali v prejšnjem poglavju.

Zaprta logika parametričnih formatov CAD ali kompleksnih parametričnih datotek IFC (STEP) je v

večini poslovnih procesov odveč. Uporabniki iščejo poenostavljene in ravne formate, kot so USD, CPIXML, XML &OBJ, DXF, glTF, SQLite, DAE &XLSX, ki vsebujejo vse potrebne informacije o elementih, vendar niso obremenjeni z odvečno geometrijsko logiko BREP, odvisnostjo od geometrijskih jeder in notranjimi klasifikacijami posebnih izdelkov CAD in BIM (slika 1). 6.2-5).



Slika 6.2-5 Za večino primerov uporabniki izberejo najpreprostejše možne formate, ki so neodvisni od programov prodajalcev.

Pojav enotnih slikovnih formatov, kot so JPEG, PNG in GIF, brez odvečne logike notranjih mehanizmov proizvajalcev, je spodbudil razvoj na tisoče interoperabilnih rešitev za obdelavo in uporabo grafike. To je privedlo do pojava različnih aplikacij, od orodij za retuširanje in filtriranje do družabnih omrežij, kot so Instagram, Snapchat in Canva, kjer je mogoče te poenostavljene podatke uporabljati brez vezanosti na določenega razvijalca programske opreme.

Standardizacija in poenostavitev oblikovanja CAD -formatov bo spodbudila nastanek številnih novih uporabniku prijaznih in neodvisnih orodij za delo z gradbenimi projektmi.

Prehod od zapletene logike aplikacij proizvajalcev, vezanih na zaprta geometrijska jedra, k univerzalnim odprtim formatom, ki temeljijo na knjižnicah poenostavljenih elementov, ustvarja predpogoje za bolj prilagodljivo, pregledno in učinkovito obdelavo podatkov. S tem se odpira tudi dostop do informacij za vse udeležence v procesu gradnje - od projektantov do strank in vzdrževalnih služb.

Kljub temu je zelo verjetno, da bodo v prihodnjih letih ponudniki sistemov CAD poskušali ponovno premakniti razpravo o interoperabilnosti in dostopu do podatkovnih zbirk CAD. Šlo bo že za "nove" koncepte - kot so granularni podatki, inteligentni grafi, "združeni modeli", digitalni dvojčki v repositorijih v oblaku - ter za oblikovanje industrijskih zavezništev in standardov, ki nadaljujejo pot BIM in odprtega BIM. Kljub privlačni terminologiji lahko takšne pobude ponovno postanejo orodja za

zadrževanje uporabnikov znotraj lastniških ekosistemov. Eden od primerov je dejavno spodbujanje formata USD (Universal Scene Description) kot "novega standarda" za medplatformno sodelovanje CAD (BIM) od leta 2023 naprej.

## Prehod na USD in granularni podatki m

Nastanek zveze AOUSD [124] leta 2023 pomeni pomemben preobrat v gradbeništvu. Priča smo začetku nove realnosti, ki jo oblikujejo ponudniki CAD, pri obravnavi gradbenih podatkov z več pomembnimi spremembami. Prva večja sprememba se nanaša na dojemanje podatkov CAD - podatkov. Strokovnjaki, ki sodelujejo v zgodnjih fazah konceptualnega projektiranja, se vse bolj zavedajo, da je izdelava projekta v okolju CAD le izhodišče. Podatki, ki nastanejo med procesom projektiranja, sčasoma postanejo osnova za analizo, delovanje in upravljanje objektov. To pomeni, da morajo biti dostopni in uporabni v sistemih, ki presegajo tradicionalna orodja CAD.

Vzporedno s tem poteka revolucija v pristopu vodilnih razvijalcev. Vodilni ponudnik CAD- v panogi, ustvarjalec koncepta BIM in formata IFC, je v svoji strategiji naredil nepričakovani preobrat. Od leta 2023 dalje se podjetje odmika od tradicionalnega shranjevanja podatkov v posameznih datotekah in se osredotoča na delo z granularnimi (normaliziranimi in strukturiranimi) podatki ter prehaja na podatkovno usmerjen pristop [125].

Prodajalci sledijo zgodovinskim trendom drugih panog: večina uporabnikov ne potrebuje zaprtih formatov CAD (podobno kot PSD) ali zapletenih parametričnih datotek IFC (podobno kot GIMP z logiko plasti). Potrebujejo preproste objektne slike, ki jih je mogoče uporabiti v CAFM (gradbeni Instagram), ERP (Facebook) in tisočih drugih procesov, polnih preglednic Excel in dokumentov PDF.

Trenutni trendi v gradbeništvu morda postavljajo temelje za postopno opuščanje parametričnih in kompleksnih formatov v korist bolj univerzalnih in neodvisnih formatov USD, GLTF, DAE, OBJ (z metainformacijami v hibridnem formatu in v ločenih strukturiranih ali ohlapno strukturiranih formatih). Zgodovinski voditelji, vključno z velikimi projektantskimi podjetji, ki so nekoč sredi devetdesetih let prejšnjega stoletja dejavno spodbujali IFC, zdaj odkrito promovirajo novi format USD [93] ter poudarjajo njegovo preprostost in vsestranskoščnost (slika 6.2-6). Množična uporaba USD v izdelkih, združljivost z GLTF in aktivna integracija v orodja, kot so Blender, Unreal Engine in Omniverse, kažejo na možnost začetka nove paradigme dela s podatki. Poleg priljubljenosti lokaliziranih rešitev, kot je evropski ploščati format USD - CPIXML, ki se uporablja v priljubljenih evropskih ERP, bi se lahko položaj USD v Srednji Evropi potencialno okreplil. Organizacije, ki sodelujejo pri razvoju formata IFC, že prilagajajo svojo strategijo USD [126], kar samo potrjuje neizogibnost premika.

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	

Slika 6.2-6 Primerjava tehničnih specifikacij formatov IFC in USD.

Glede na to lahko USD postane dejanski standard, saj obljudlja, da bo premagal številne sedanje omejitve, ki so predvsem povezane z zapletenostjo obstoječih formatov CAD - (BIM-) in odvisnostjo njihove razlage od geometrijskih jeder.

Namesto parametričnih in zaplenenih formatov CAD in IFC bodo v gradbeništvu zaradi svoje preprostosti in prilagodljivosti pridobili mesto poenostavljeni podatkovni formati USD, gLTF, DAE, OBJ z metainformacijami elementov v CSV, XLSX, JSON, XML.

Sedanje spremembe v gradbeništvu so na prvi pogled videti kot tehnološki preboj, povezan s prehodom z zastarelega sistema IFC na sodobnejši sistem USD. Vendar pa je treba upoštevati, da je

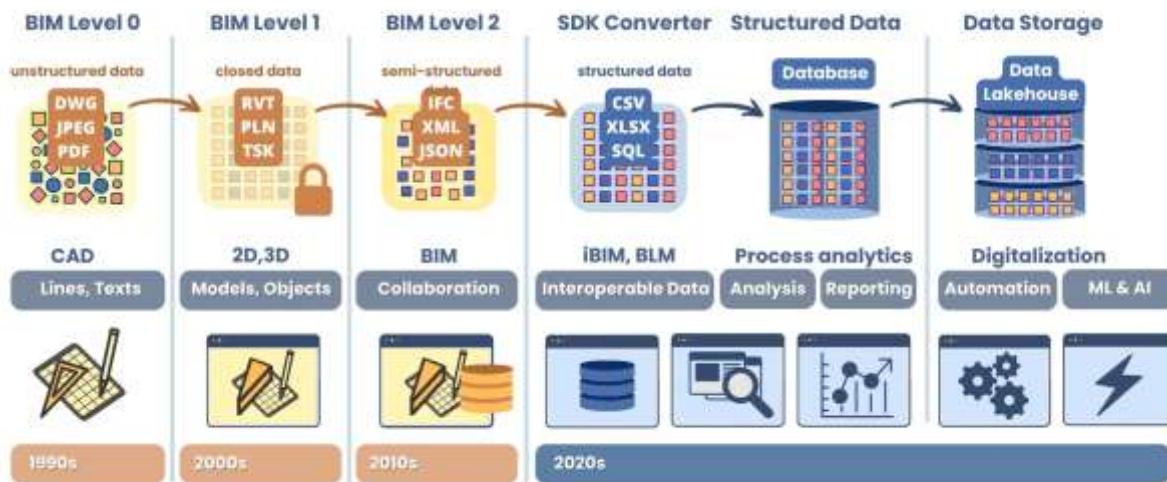
že leta 2000 isti ponudnik CAD, ki je razvil IFC, pisal o njegovih težavah in potrebi po dostopu do podatkovne zbirke [65], zdaj pa aktivno spodbuja prehod na nov standard - USD.

Za še eno fazado "odprtih podatkov" USD in "novimi" koncepti za granularno upravljanje podatkov se lahko prek aplikacij v oblaku, ki jih začenjajo promovirati prodajalci CAD, skriva namen prodajalcev monopolizirati upravljanje projektnih podatkov, pri čemer se uporabniki znajdejo v položaju, ko je izbira formata bolj povezana z interesu podjetij kot z dejanskimi potrebami.

Analiza ključnih dejstev [93] kaže, da glavni namen teh sprememb ni toliko v udobju uporabnikov kot predvsem v ohranjanju nadzora nad ekosistemi in pretokom podatkov v korist prodajalcev, ki v 40 letih nikoli niso mogli zagotoviti dostopa do podatkovnih zbirk CAD.

Morda je zdaj čas, da podjetja prenehajo čakati na nove koncepte prodajalcev programske opreme in se osredotočijo na lasten razvoj v smeri podatkovne usmerjenosti. Ko se bo industrija osvobodila težav z dostopom do podatkov s pomočjo orodij povratnega inženiringa, bo lahko samostojno prešla na sodobna, brezplačna in priročna orodja za delo in analizo podatkov brez vsiljevanja novih konceptov.

### CAD (BIM) Maturity Levels: From Stage 0 to Structured Data



Slika 6.2-7 Raven zrelosti CAD (BIM): od nestrukturiranih podatkov do strukturiranih podatkov in repozitorijev.

Dostop do podatkovnih zbirk, odprti podatki in formati bodo v gradbeništvu neizogibno postali standard, ne glede na poskuse prodajalcev, da bi ta proces ustavili - to je le vprašanje časa (slika 6.2-7). Hitrost tega prehoda se lahko bistveno poveča, če se bo vse več strokovnjakov seznanilo z odptimi formati, orodji za podatkovne zbirke in razpoložljivimi SDK-ji za obratno inženirstvo, ki omogočajo neposreden dostop do podatkov CAD -sistemov [92].

Prihodnost je v odprtih, poenotenih in analitično dostopnih podatkih. Da bi se izognila odvisnosti od rešitev prodajalcev in da ne bi postala talec zaprtih ekosistemov, se bodo morala gradbena in inženirska podjetja prej ali slej opreti na odprtost in neodvisnost ter izbrati formate in rešitve, ki zagotavljajo popoln nadzor nad podatki.

Podatki, ki danes nastajajo v gradbeništvu, bodo v prihodnosti ključni vir za poslovne odločitve. Delovali bodo kot strateško "gorivo", ki bo spodbujalo razvoj in učinkovitost gradbenih podjetij. Prihodnost gradbeništva je v sposobnosti dela s podatki in ne v izbiri podatkovnih formatov ali modelov.

Za razumevanje razlike med odprtimi formati USD, glTF, DAE, OBJ in lastniškimi parametričnimi formati CAD je pomembno upoštevati enega najbolj zapletenih in ključnih podatkovnih elementov pri vizualizaciji in oblikovalskih izračunih - geometrijo in postopke njenega ustvarjanja. In da bi razumeli, kako geometrijski podatki postanejo osnova za analitiko in izračune v gradbeništvu, se je treba poglobiti v mehanizme generiranja, preoblikovanja in shranjevanja geometrije.



## POGLAVJE 6.3.

### GEOMETRIJA V GRADBENIŠTVU: OD ČRT DO KUBIČNIH METROV

#### Ko se črte spremenijo v denar ali zakaj gradbeniki potrebujejo geometrijo

Geometrija v gradbeništvu ni le vizualizacija, temveč tudi osnova za natančne kvantitativne izračune. V modelu projekta geometrija dopolnjuje sezname parametrov elementov (slika 3.1-16) s pomembnimi prostorninskimi značilnostmi, kot so dolžina, površina in prostornina. Vrednosti teh volumetričnih parametrov se samodejno izračunajo s pomočjo geometrijskih jeder in so izhodišče za ocene, razporede in modele virov. Kot smo že obravnavali v 5. delu te knjige in v poglavju "Obračunavanje stroškov in ocenjevanje gradbenih projektov", so prav volumetrični parametri skupin objektov iz CAD - modelov osnova sodobnih ERP, PMIS - sistemov/ Geometrija ima temeljno vlogo ne le v fazi projektiranja, temveč tudi pri upravljanju izvajanja projekta, nadzoru časovnega načrta, načrtovanju proračuna in obratovanju. Tako kot je bila pred tisočletji pri gradnji egiptanskih piramid natančnost projekta odvisna od dolžinskih mer, kot so komolci in loki, danes natančnost interpretacije geometrije v CAD - programih neposredno vpliva na rezultat: od proračuna in rokov do izbire izvajalca in logistike dostave

V zelo konkurenčnem okolju z omejenim proračunom postane natančnost volumetričnih izračunov, ki je neposredno odvisna od geometrije, dejavnik preživetja. Sodobni sistemi ERP so neposredno odvisni od pravilnih prostorninskih lastnosti, pridobljenih iz modelov CAD in BIM. Zato natančen geometrijski opis elementov ni le vizualizacija, temveč ključno orodje za obvladovanje stroškov in časa gradnje.

V preteklosti je bila geometrija glavni jezik inženirske komunikacije. Od črt na papirusu do digitalnih modelov so risbe in geometrijski prikazi služili kot sredstvo za izmenjavo informacij med projektanti, mojstri in ocenjevalci. Pred pojavom računalnikov so se izračuni izvajali ročno, z uporabo ravnila in kotomerov. Danes je ta naloga avtomatizirana zaradi volumetričnega modeliranja: geometrijska jedra programske opreme CAD pretvorijo črte in točke v tridimenzionalna telesa, iz katerih se samodejno pridobijo vse potrebne lastnosti.

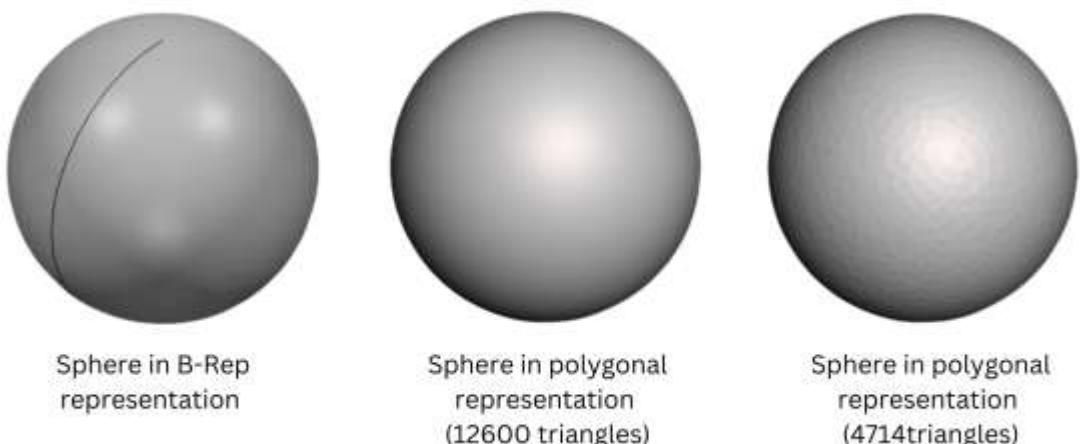
Pri delu v programih CAD se ustvarjanje geometrijskih elementov za izračune izvaja prek uporabniškega vmesnika programov CAD- (BIM-). Za pretvorbo točk in črt v volumenska telesa se uporablja geometrijsko jedro, ki opravlja ključno nalogu - pretvorbo geometrije v volumenske modele, iz katerih se po aproksimaciji samodejno izračunajo volumenske značilnosti elementa.

#### Od črt do volumnov: kako površina in volumen postaneta podatki

V inženirski praksi se prostornine in površine izračunavajo iz geometrijskih površin, opisanih analitično ali s parametričnimi modeli, kot so NURBS (neenakomerni racionalni B-splini) v okviru BREP (predstavitev mejnih elementov).

NURBS (Non-Uniform Rational B-Splines) je matematični način za opisovanje krivulj in površin, medtem ko je BREP okvir za opisovanje celotne tridimenzionalne geometrije predmeta, vključno z njegovimi mejami, ki jih je mogoče opredeliti z NURBS.

Kljub natančnosti BREP in NURBS zahtevata zmogljive računalniške vire in zapletene algoritme. Vendar je neposredno računanje iz tako matematično natančnih opisov pogosto računsko zahtevno, zato se v praksi skoraj vedno uporablja teselacija - pretvorba površin v mrežo trikotnikov -, ki poenostavi nadaljnje izračune. Tessellation je razdelitev kompleksne površine na trikotnike ali poligone. V okoljih CAD /CAE se ta metoda uporablja za vizualizacijo, izračune volumnov, iskanje trkov, izvoz v formate, kot je MESH, in analizo trkov. Primer iz narave je čebelji sat, pri katerem je kompleksna oblika razčlenjena na pravilno mrežo (slika 6.3-1).



**Slika 6.3-1** Ista krogla v parametričnem opisu BREP in poligonalni predstavitev z različnim številom trikotnikov.

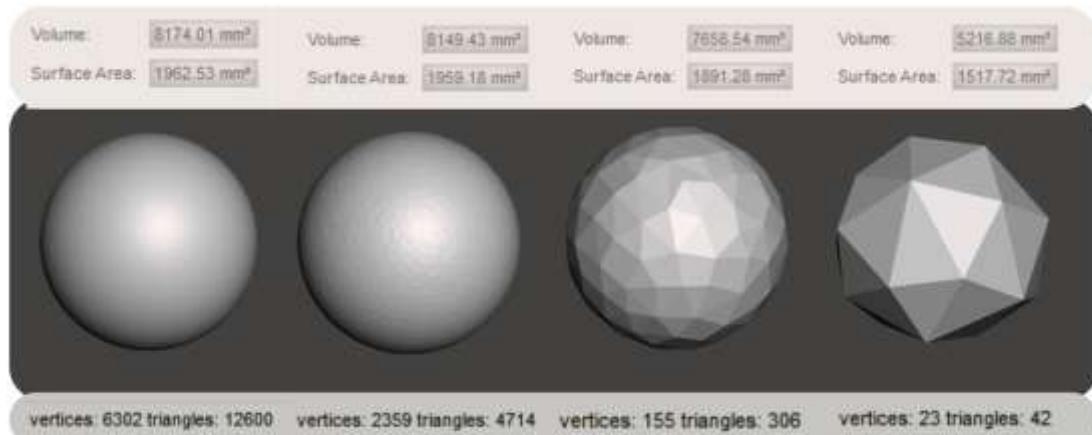
BREP (NURBS), ki se uporablja v CAD, ni temeljni model geometrije. Ustvarjen je bil kot priročno orodje za predstavitev krogov in racionalnih ploskev ter za čim manjše shranjevanje geometrijskih podatkov. Vendar ima omejitve - na primer nezmožnost natančnega opisa sinusoide, ki je osnova spiralnih črt in površin, ter potrebo po uporabi zapletenih geometrijskih jeder.

V nasprotju s tem so za trikotne mreže in teselacijo parametričnih oblik značilne preprostost, učinkovita uporaba pomnilnika in zmožnost obdelave velikih količin podatkov (slika 6.3-2). Te prednosti omogočajo, da se pri izračunavanju geometrijskih oblik lahko izognemo zapletenim in dragim geometrijskim jedrom in vanje vgrajenim desetinam milijonov vrstic kode.

V večini primerov stavb ni pomembno, kako natančno so prostorninske značilnosti opredeljene - s parametričnimi modeli (BREP, IFC) ali s poligoni (USD, glTF, DAE, OBJ). Geometrija ostaja oblika približka: bodisi prek NURBS bodisi prek MESH, vedno gre za približen opis oblike.

Geometrija, opredeljena kot poligoni ali BREP (NURBS), ostaja do neke mere le način približnega opisa zvezne oblike. Tako kot Fresnelovi integrali nimajo natančnega analitičnega izraza, je tudi diskretizacija geometrije s poligoni ali NURBS vedno približek, tako kot trikotniški MESH.

Parametrična geometrija v formatu BREP je potrebna predvsem tam, kjer je pomembna minimalna velikost podatkov in je za njeno obdelavo in prikaz mogoče uporabiti draga geometrijska jedra, ki zahtevajo veliko virov. Najpogosteje je to značilno za razvijalce programov CAD, ki v ta namen v svojih izdelkih uporabljajo geometrijska jedra ponudnikov MCAD. V tem primeru se tudi v teh programih modeli BREP v postopku teselacije za vizualizacijo in izračune pogosto pretvorijo v trikotnike (podobno kot se datoteke PSD poenostavijo v JPEG).



**Slika 6.3-2 Razlika volumenskih karakteristik na slikah z različnim številom poligonov.**

Polygonalni MESH, kot tudi parametrični BREP, imajo svoje prednosti in omejitve, vendar je cilj enak - opisati geometrijo z upoštevanjem uporabnikove naloge. Navsezadnje natančnost geometrijskega modela ni odvisna le od metode njegove predstavitve, temveč tudi od zahtev določene naloge.

Pri večini konstrukcijskih problemov so parametrična geometrija in zapletena geometrijska jedra lahko odveč.

Pri vsaki posamezni nalogi avtomatizacije izračuna je treba razmisiliti, ali razvijalci CAD, ki jih zanima promocija in prodaja lastnih programskih izdelkov, pretiravajo s pomenom parametrične geometrije.

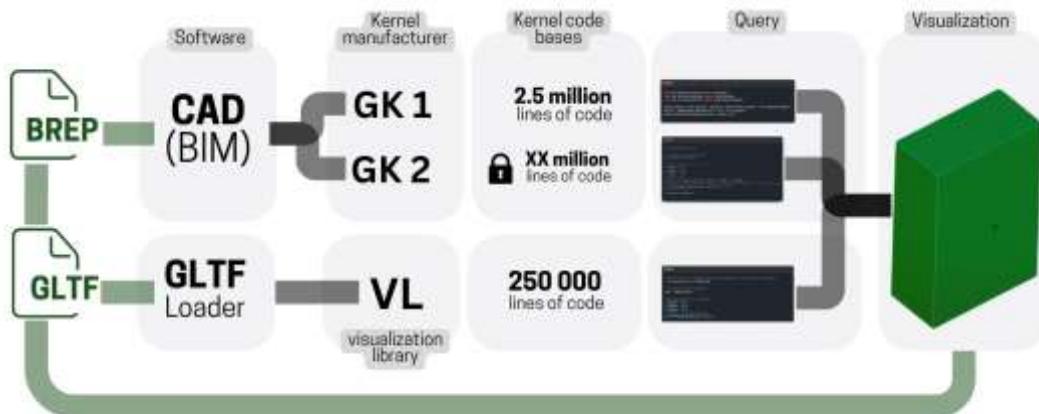
## Prehod na MESH, USD in poligone: uporaba teselacije za geometrijo

V gradbeništvu si je pri pretakanju, razvoju sistemov, podatkovnih zbirk ali avtomatizaciji procesov, ki delajo z informacijami o projektiranju in geometrijo elementov, pomembno prizadevati za neodvisnost od posebnih urejevalnikov CAD in geometrijskih jeder.

Oblika izmenjave, ki se uporablja v oddelkih za izračun in na gradbišču, ne sme temeljiti na posebnem programu CAD- (BIM-). Geometrične informacije bi morale biti v formatu predstavljene neposredno s teselacijo, brez sklicevanja na geometrijsko jedro ali arhitekturo CAD.

Parametrično geometrijo iz sistema CAD lahko obravnavamo kot vmesni vir, vendar ne kot osnovo za univerzalni format. Večina parametričnih opisov (vključno z BREP in NURBS) se za nadaljnjo obdelavo v vsakem primeru pretvori v poligonski MESH. Če je rezultat enak (teselacija in poligoni) in je postopek enostavnejši, je izbira očitna. To je analogno izbiri med grafnimi ontologijami in strukturiranimi tabelami (ki smo jo obravnavali v četrtem delu): pretirana zapletenost je redko upravičena (slika 3.2-10, slika 6.1-8).

Odprtji formati, kot so: OBJ, STL, glTF, SVF, CPIXML, USD in DAE, uporabljajo univerzalno trikotno mrežno strukturo, kar jim daje pomembne prednosti. Ti formati imajo odlično interoperabilnost - zlahka jih je mogoče brati in vizualizirati z uporabo razpoložljivih odprtokodnih knjižnic brez potrebe po zapletenih specializiranih geometrijskih jedrih, ki vsebujejo na milijone vrstic kode (slika 6.3-3). Ti vsestranski geometrijski formati se uporabljajo v različnih aplikacijah, od razmeroma preprostih orodij za oblikovanje kuhinj v IKEA™ do kompleksnih sistemov za vizualizacijo objektov v kinematografih in VR-aplikacijah. Pomembna prednost je razpoložljivost velikega števila brezplačnih in odprtokodnih knjižnic za delo s temi formati, ki so na voljo za večino platform in programskih jezikov.



Slika 6.3-3 Enako predstavitev geometrije je mogoče doseči z uporabo parametričnih formatov in geometrijskih jeder ali z uporabo trikotniških formatov in odprtokodnih knjižnic za vizualizacijo.

Poleg uporabnikov samih imajo tudi ponudniki CAD težave s tolmačenjem tujih parametričnih formatov CAD ali odprtih formatov IFC zaradi različnih geometrijskih jeder. V praksi vsi -prodajalci CAD brez izjeme za prenos podatkov med sistemi uporabljajo SDK za povratni inženiring in nihče od njih se za namene interoperabilnosti ne zanaša na formate, kot sta IFC ali USD [93]

Namesto da bi uporabljali koncepte, ki jih spodbujajo združenja ponudnikov CAD- in jih sami ne uporabljajo, je za razvijalce in uporabnike rešitev CAD bolj produktivno, da se osredotočijo na razumevanje prednosti vsakega pristopa v določenem kontekstu in izberejo eno ali drugo vrsto geometrije glede na primer uporabe. Izbera med različnimi geometrijskimi predstavitevami je kompromis med natančnostjo, računsko učinkovitostjo in praktičnimi potrebami določene naloge.

Zapletenost, povezana z uporabo geometrijskih jeder, ki jih veliki prodajalci tradicionalno vsiljujejo gradbeni industriji pri obdelavi projektnih podatkov, se pogosto izkaže za nepotrebno. Format USD, ki temelji na geometriji MESH, lahko postane nekakšna "Pandorina skrinjica" za industrijo, saj razvijalcem odpira nove možnosti za organizacijo izmenjave podatkov - zunaj okvira struktur IFC in parametričnih struktur BREP, značilnih za prodajalce CAD.

Po podrobnejšem pregledu strukture USD, DAE, glTF, OBJ itd. postane jasno, da obstajajo preprostejši, odprtii formati, ki omogočajo učinkovito organizacijo prenosa in uporabe geometrijskih informacij, ne da bi se bilo treba zanašati na zapletene parametre in zaprta geometrijska jedra. Ta pristop ne le znižuje tehnični prag za vstop razvijalcev, temveč tudi spodbuja razvoj prilagodljivih, razširljivih in resnično odprtih rešitev za digitalno gradnjo.

## **LOD, LOI, LOMD - enotna klasifikacija podrobnosti v CAD (BIM)**

V svetu, kjer različne panoge uporabljajo različne ravni podrobnosti in globine podatkov, metodologije CAD - (BIM-) poleg formatov geometrijske predstavitev ponujajo tudi svoje edinstvene sisteme klasifikacije, ki strukturirajo pristop k informiranju modelov stavb.

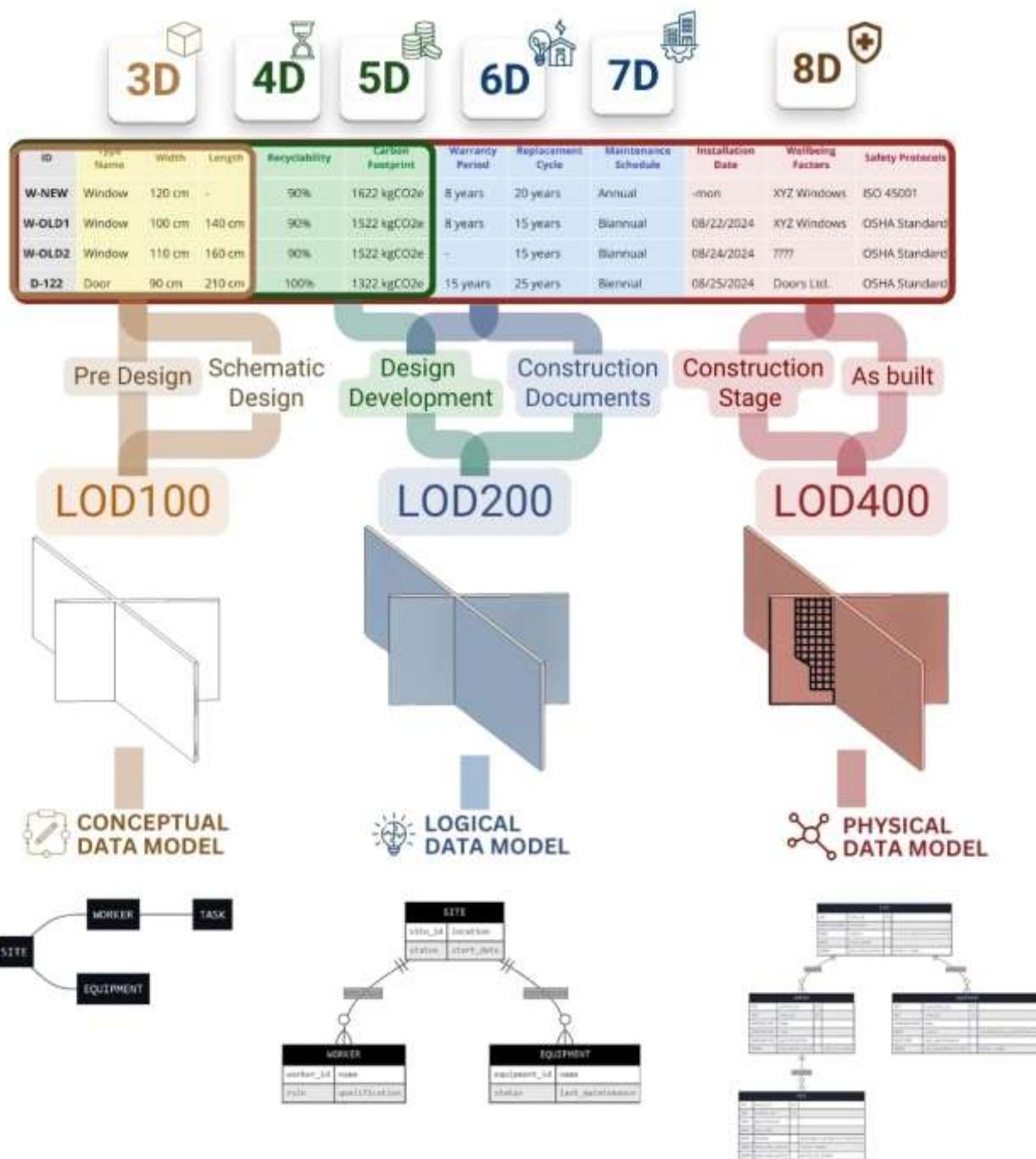
Eden od primerov novih pristopov k standardizaciji je uvedba ravni razvoja modela, ki odraža stopnjo pripravljenosti in zanesljivosti grafičnih in informacijskih komponent. Za razlikovanje informacijske vsebine pri delu s podatki CAD - (BIM-) sta se pojavila LOD (Level Of Detail) - raven podrobnosti grafičnega dela modela, in LOI (Level Of Information) - raven izdelave podatkov. Poleg tega je bil za celostni pristop uveden koncept LOA (Level of Accuracy) - natančnost predstavljenih elementov in LOG (Level of Geometry) za določitev natančnosti grafične predstavitev.

Stopnje podrobnosti (LOD) so označene s številkami od 100 do 500, kar odraža stopnjo razvoja modela. LOD 100 je konceptualni model s splošnimi oblikami in dimenzijami. LOD 200 vključuje natančnejše dimenzijske in oblike, vendar s pogojnimi podrobnostmi. LOD 300 je podrobni model z natančnimi dimenzijskimi, oblikami in lokacijami elementov. LOD 400 vsebuje podrobne informacije, potrebne za izdelavo in namestitev elementov. LOD 500 odraža dejansko stanje objekta po izgradnji in se uporablja za obravnavanje in vzdrževanje. Te ravni opisujejo strukturo zasičenosti modela CAD (BIM) z informacijami v različnih fazah življenjskega cikla, vključno s 3D, 4D, 5D in naprej.

V resničnih projektih je visoka raven podrobnosti (LOD400) pogosto pretirana in zadostuje uporaba geometrije LOD100 ali celo ravnihs risb, medtem ko se preostali podatki lahko pridobijo z računanjem ali iz povezanih elementov, ki morda nimajo ločene geometrije. Na primer, prostori in elementi sob (kategorije elementov sob) morda nimajo vizualne geometrije, vendar vsebujejo velike količine informacij in podatkovnih zbirk, okoli katerih so zgrajeni številni poslovni procesi.

Zato je pomembno, da pred začetkom načrtovanja jasno opredelite zahtevano raven podrobnosti. Za primere uporabe 4D -7D pogosto zadostujejo že risbe DWG in minimalna geometrija LOD100. Ključna naloga v procesu zahtev je najti ravnovesje med bogastvom in praktičnostjo modela.

Če podatke CAD (BIM) obravnavamo kot podatkovno zbirko (kar tudi je), opis nasičenosti modela z novimi kraticami v bistvu ni nič drugega kot postopno modeliranje podatkov za informacijske sisteme, ki se začne na konceptualni ravni in konča na fizični (slika 6.3-4), kar smo podrobno obravnavali v tretjem in četrtem delu knjige. Vsako povečanje LOD in LOI pomeni dodajanje informacij, potrebnih za nove naloge: izračune, vodenje gradnje, obratovanje, zanj pa je značilno zaporedno bogatenje modela z dodatnimi informacijskimi plastmi (3D -8D) v obliki različnih parametrov, ki smo jih obravnavali v petem delu knjige.



Slika 6.3-4 Postopek izpopolnjevanja podrobnosti projekta je enak podatkovnemu modeliranju od konceptualnega do fizičnega podatkovnega modela.

Geometrija je le del projektnih podatkov, katerih potreba po njih v gradbenih projektih ni vedno upravičena, zato ključno vprašanje pri delu s podatki CAD ni toliko v tem, kako se modeli vizualizirajo, temveč bolj v tem, kako se lahko podatki iz teh modelov uporabljajo zunaj programov CAD (BIM -).

Sredi leta 2000 se je gradbena industrija soočila z izvodom brez primere zaradi hitrega povečanja količine podatkov v sistemih za upravljanje in obdelavo podatkov, zlasti tistih, ki prihajajo iz oddelkov

CAD (BIM). To dramatično povečanje količine podatkov je presenetilo vodje podjetij, ki niso bili pripravljeni na naraščajoče zahteve glede kakovosti in upravljanja podatkov.

## **Novi standardi CAD (BIM) - AIA, BEP, IDS, LOD, COBie**

Organizacije, ki se ukvarjajo z razvojem pristopov k ravnjanju s podatki CAD, so začele oblikovati nove standarde in koncepte, ki naj bi bili de iure namenjeni izboljšanju praks upravljanja podatkov, pri čemer so izkoristile pomanjkanje odprtrega dostopa do podatkovnih baz CAD in omejeno konkurenco na trgu obdelave podatkov ter uporabile tržne kampanje, povezane z novo kratico BIM.

Čeprav so bile skoraj vse pobude, ki so jih neposredno ali posredno podpirali prodajalci CAD in razvijalci (BIM), namenjene optimizaciji delovnih postopkov, so privedle do množice standardov, za katere so lobirali različni deležniki, kar je v gradbeništvu povzročilo nekaj nejasnosti in zmede glede podatkovnih postopkov.

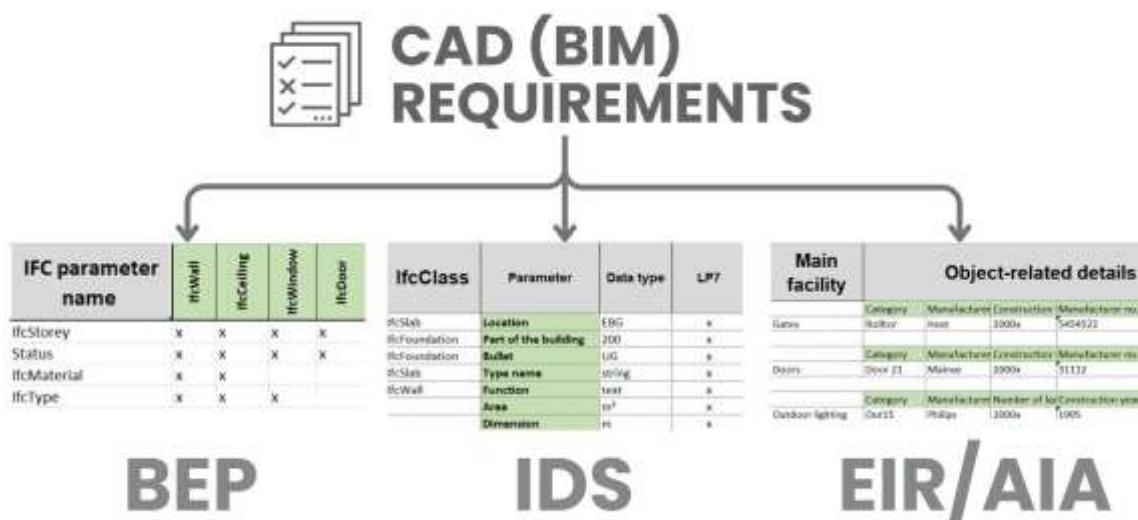
Naštejmo nekaj novih podatkovnih standardov, ki so se poleg LOD, LOI, LOA, LOG, v zadnjih letih pojavili v gradbeništvu:

- **BEP** (BIM Execution Plan) - opisuje, kako vključiti in uporabiti CAD (BIM) v projektu, ter opredeljuje metode in postopke ravnanja s podatki.
- **Dokument EIR /AIA** (Informacijske zahteve naročnika) - pripravi ga naročnik pred razpisom in vsebuje zahteve za izvajalca, da pripravi in zagotovi informacije. Služi kot podlaga za BEP v zadevnem projektu.
- **AIM** (Asset Information Model) je del procesa BIM. Ko je projekt dobavljen in dokončan, se podatkovni model imenuje informacijski model sredstev ali AIM. Namen AIM je upravljanje, vzdrževanje in upravljanje realiziranega sredstva.
- **IDS** (Information Delivery Specification) - opredeljuje zahteve ter katere podatke in v kakšni obliki so potrebni v različnih fazah gradbenega projekta.
- **iLOD** je raven podrobnosti LOD, s katero so informacije predstavljene v modelu BIM. Določa, kako podrobne in popolne so informacije v modelu, od osnovnih geometrijskih predstavitev do podrobnih specifikacij in podatkov.
- **eLOD** - LOD raven podrobnosti posameznih elementov v modelu CAD (BIM). Opredeljuje obseg, v katerem je vsak element modeliran, in z njim povezane informacije, kot so dimenzije, materiali, značilnosti delovanja in drugi pomembni atributi.
- **APS** (Platform Services) in drugi izdelki glavnih ponudnikov CAD (BIM) - na spletni strani so opisana orodja in infrastruktura, potrebna za ustvarjanje povezanih in odprtih podatkovnih modelov.

Čeprav je deklarirani namen izvajanja standardov CAD (BIM), kot so LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD, izboljšati kakovost upravljanja podatkov in razširiti možnosti avtomatizacije, njihova uporaba v praksi pogosto vodi v pretirano zapletenost in razdrobljenost procesov. Če model

CAD (BIM) obravnavamo kot neke vrste podatkovno zbirko, postane očitno, da številni od teh standardov podvajajo že dolgo uveljavljene in učinkovite pristope, ki se uporabljajo v drugih panogah pri delu z informacijskimi sistemmi. Namesto poenostavitev in poenotenja takšne pobude pogosto ustvarjajo dodatno terminološko breme in ovirajo izvajanje resnično odprtih in prilagodljivih rešitev.

Pomembno je, da številni od teh novih konceptov dejansko nadomeščajo postopke modeliranja in potrjevanja podatkov, ki so bili podrobno obravnavani v prvih delih knjige in ki se že dolgo uporabljajo v drugih gospodarskih sektorjih. Po drugi strani pa v gradbeništvu proces standardizacije pogosto poteka v nasprotni smeri - nastajajo novi formati za opis podatkov, novi standardi in novi koncepti za potrjevanje podatkov, ki pa ne vodijo vedno do dejanske enotnosti in praktične uporabnosti. Posledično se panoga namesto s poenostavitvijo in avtomatizacijo obdelave srečuje z dodatnimi ravnimi predpisov in birokracije (slika 6.3-1), kar ne prispeva vedno k večji učinkovitosti.



Slika 6.3-1 Zahteve glede vsebine podatkov in informacij so omejene na opis atributov in njihovih mejnih vrednostih, opisanih s pomočjo tabel.

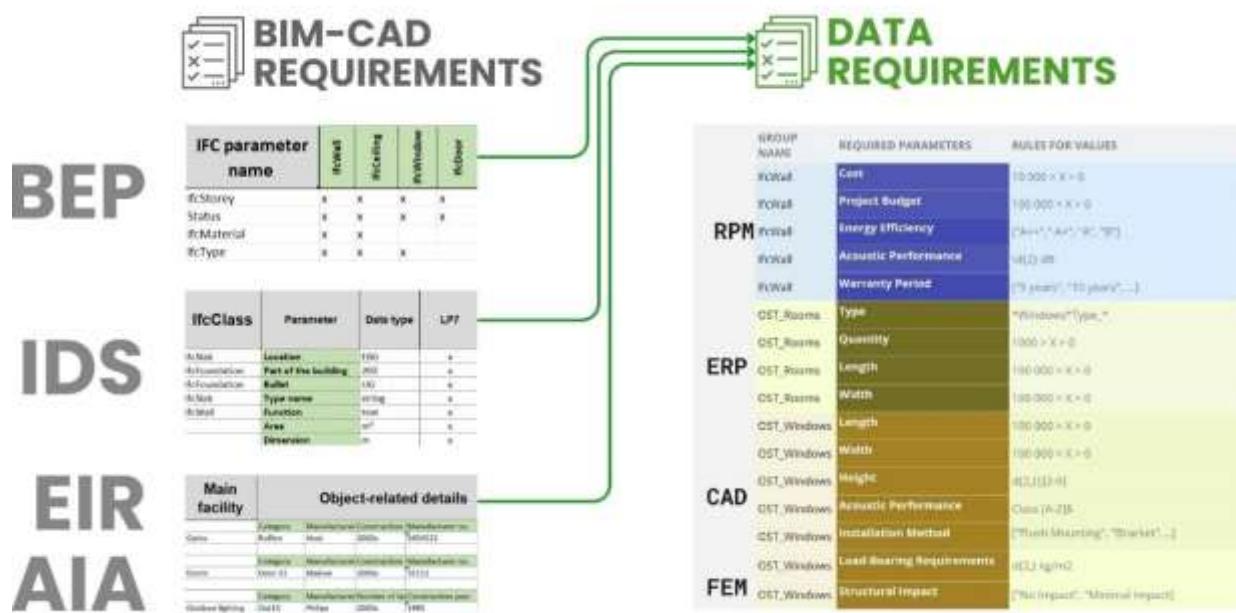
Namesto da bi poenostavili obdelavo podatkov, novi koncepti, povezani s podatki CAD (BIM), najpogosteje povzročajo dodatne zaplete in spore že v fazi razlage in osnovnih opredelitev.

Eden najnovejših primerov novih konceptov je format IDS (uveden leta 2020), ki omogoča opis zahtev za sestavo atributov informacijskega modela v odprttem konceptu BIM. Zahteve IDS opisujejo informacije o atributih in njihovih mejnih vrednostih v obliki strukturirane tabele (Excel ali MySQL), ki se nato prevede v označevanje polstrukturiranega formata XML, preimenovanega iz XML v posebno kratico IDS.

V nasprotju s stališčem, ki ga spodbujajo prodajalci in podpirajo BIM in odprti BIM, da je ravnanje s podatki v gradbeništvu edinstveno zaradi uporabe specializiranih orodij, kot sta CAD in BIM, se formati podatkov in prakse upravljanja podatkov v tej panogi ne razlikujejo od tistih drugih panogah.

Število zahtev za projekte in formate CAD (BIM) je mogoče poenostaviti z uporabo ene same preglednice zahtev s stolpci atributov, podrobno opisane v poglavju "Prevajanje zahtev v strukturirano obliko", ne da bi bilo treba prvotno strukturirane zahteve prevesti v formate, ki niso tabele (IDS je na začetku opisan s preglednico).

Poenostavljeni pristop (slika 6.3-2), ki vključuje stolpce za identifikatorje entitet, lastnosti in mejne vrednosti, ki so bili podrobno obravnavani v prejšnjih poglavjih (slika 4.4-9, slika 4.4-16, slika 7.3-10), odpravlja potrebo po pretvorbi zahtev v obliko IDS-XML. Ta metoda zagotavlja neposreden, manj okoren in preglednejši mehanizem za nadzor kakovosti podatkov. Zanaša se na široko uporabljenih orodja, od regularnih izrazov (RegEx) do podatkovnih okvirjev, Pandas in standardnih ETL -payplanes - prav takšna, kot jih za delo s podatki uporablja strokovnjaki v drugih gospodarskih sektorjih.



Slika 6.3-2 Podatkovne zahteve v drugih panogah so poenostavljene na strukturiran opis atributov in njihovih mejnih vrednosti.

V gradbeništvu se zaradi zaprte narave podatkov sčasoma pojavlja vse več novih pristopov in tehnik za nadzor in upravljanje teh različnih podatkov, čeprav so podatki v gradbenih projektih v bistvu enaki kot na drugih področjih. Medtem ko so druge panoge uspešno shajale s standardiziranimi pristopi k obdelavi podatkov, gradbeništvo še naprej razvija nove in edinstvene podatkovne formate, zahteve in koncepte potrjevanja.

Metode in orodja, ki se uporabljam za zbiranje, pripravo in analizo podatkov v gradbeništvu, se ne smejo bistveno razlikovati od tistih, ki jih uporabljam strokovnjaki v drugih gospodarskih sektorjih.

Industrija je razvila poseben terminološki ekosistem, ki zahteva kritičen razmislek in ponovno ovrednotenje:

- Format STEP je umeščen pod novim imenom IFC, ki ga dopoljuje kategorizacija konstrukcij, pri čemer se ne upoštevajo omejitve samega formata STEP.
- Parametrični format IFC se uporablja v procesih sporočanja podatkov kljub pomanjkanju enotnega geometrijskega jedra, potrebnega za vizualizacijo in izračunavanje.
- Dostop do podatkovnih zbirk CAD -sistemov se spodbuja pod izrazom "BIM", brez razprave o posebnostih teh podatkovnih zbirk in dostopa do njih.
- Prodajalci spodbujajo interoperabilnost prek formatov IFC in USD, ki jih pogosto ne izvajajo v praksi, temveč uporabljajo drag povratni inženiring, s katerim imajo sami težave.
- Izrazi LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD se uporablja univerzalno za opis istih parametrov entitete brez sklicevanja na orodja za modeliranje in preverjanje, ki se že dolgo uporabljajo v drugih industrijah.

Gradbena industrija dokazuje, da je vse zgoraj navedeno, čeprav se sliši nenavadno, v gradbeništvu mogoče - zlasti če je glavni cilj zaslužiti na vsaki stopnji obdelave podatkov s prodajo specializiranih storitev in programske opreme. S poslovnega vidika s tem ni nič narobe. Vendar pa ostaja odprto vprašanje, ali takšne kratice in pristopi, povezani s CAD (BIM), resnično prinašajo dodano vrednost in poenostavljajo strokovne postopke.

V gradbeništvu takšen sistem deluje, saj industrija sama v tem labirintu sistemov in kratic ustvari večino svojih špekulativnih dobičkov. Podjetja, ki jih zanimajo pregledni procesi in odprti podatki, so redka. Ta zapleten položaj se bo verjetno nadaljeval v nedogled - dokler stranke, naročniki, investitorji, banke in zasebni kapital ne bodo začeli zahtevati jasnejših in bolj informiranih pristopov k upravljanju informacij.

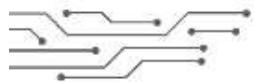
V panogi se je nabralo preveliko število kratic, vendar vse različnem obsegu opisujejo iste procese in zahteve po podatkih. Njihova dejanska uporabnost pri poenostavljavi delovnih postopkov ostaja vprašljiva.

Medtem ko koncepti in marketinške kratice prihajajo in odhajajo, bodo sami postopki potrjevanja podatkovnih zahtev za vedno ostali sestavni del poslovnih procesov. Namesto da ustvarja vse bolj specializirane formate in predpise, bi morala gradbena industrija poiskati orodja, ki so se že izkazala za učinkovita na drugih področjih, kot so finance, industrija in IT.

Obilica izrazov, kratic in formatov ustvarja iluzijo, da so digitalni gradbeni procesi zelo izpopolnjeni. Vendar pa marketinški koncepti in zapletena terminologija pogosto skrivajo preprosto, a neprijetno resnico: podatki ostajajo težko dostopni, slabo dokumentirani in togo vezani na določene programske rešitve.

Da bi se rešili iz tega začaranega kroga kratic in formatov zaradi formatov, je treba na sisteme CAD

(BIM) gledati ne kot na čarobna orodja za upravljanje informacij, temveč kot na to, kar v resnici so - specializirane podatkovne zbirke. Skozi to prizmo lahko razumemo, kje se konča trženje in začne resnično delo z informacijami.



## POGLAVJE 6.4.

### PARAMETRIZACIJA ZASNOVE IN UPORABA LLM ZA DELOVANJE S KADROM

#### Iluzija edinstvenosti podatkov CAD (BIM): pot do analitike in odprtih formatov

Sodobne platforme CAD (BIM) so bistveno spremenile pristop k upravljanju informacij o projektiranju in gradnji. Medtem ko so se prej ta orodja uporabljala predvsem za izdelavo risb in 3D-modelov, danes delujejo kot polnovredna skladišča projektnih podatkov. V okviru koncepta enotnega vira resnice postaja parametrični model vse bolj glavni in pogosto edini vir projektnih informacij, kar zagotavlja njegovo celovitost in ustreznost v celotnem življenjskem ciklu projekta.

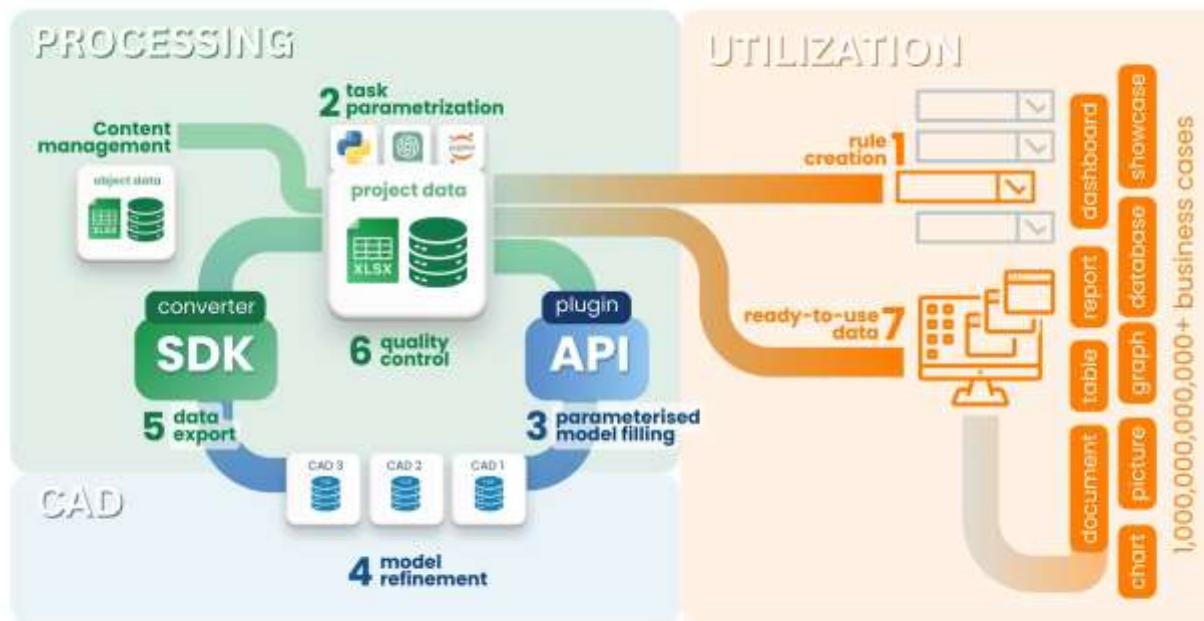
Ključna razlika med platformami CAD - (BIM-) in drugimi sistemmi za upravljanje gradbenih podatkov je, da so za dostop do informacij (edini vir resnice) potrebna specializirana orodja in vmesniki API. Te zbirke podatkov niso univerzalne v tradicionalnem smislu: namesto odprte strukture in prilagodljivega povezovanja so zaprto okolje, trdno vezano na določeno platformo in format.

Kljud zapletenosti dela s podatki CAD je pomembnejše vprašanje, ki presega tehnično realizacijo: kaj pravzaprav so podatkovne zbirke CAD (BIM)? Za odgovor na to vprašanje je treba preseči običajne kratice in pojme, ki jih vsiljujejo razvijalcji programske opreme. Namesto tega se je treba osredotočiti na bistvo dela s projektnimi informacijami: podatke in njihovo obdelavo.

Poslovni proces v gradbeništvu se ne začne z delom v orodjih CAD - ali BIM -, temveč z oblikovanjem projektnih zahtev in modeliranjem podatkov. Najprej se opredelijo parametri naloge: seznam entitet, njihovih začetnih značilnosti in mejnih vrednosti, ki jih je treba upoštevati pri reševanju določene naloge. Šele nato se v sistemih CAD (BIM) na podlagi določenih parametrov ustvarijo modeli in elementi

Postopek, ki predhodi ustvarjanju informacij v podatkovnih zbirkah CAD - (BIM-), je popolnoma enak postopku modeliranja podatkov, ki je bil podrobno obravnavan v četrtem delu knjige in poglavju "Modeliranje podatkov: konceptualni, logični in fizični model" (slika 4.3-1).

Tako kot pri modeliranju podatkov ustvarimo zahteve za podatke, ki jih želimo pozneje obdelati v podatkovni zbirki, za podatkovne zbirke CAD upravljavci ustvarijo projektne zahteve v obliki več stolpcev tabel ali seznamov parov ključ-vrednost (slika 6.4-1, korak 1-2). In šele na podlagi teh začetnih parametrov z uporabo API samodejno ali ročno projektant ustvari (ali bolje rečeno izpopolni) objekte v podatkovnih zbirkah CAD- (BIM) (koraki 3-4), nakar ponovno preveri njihovo skladnost z začetnimi zahtevami (koraki 5-6). Ta postopek - opredelitev → ustvarjanje → preverjanje → prilaganje (koraki 2-6) - se ponavlja iterativno, dokler kakovost podatkov, tako kot pri modeliranju podatkov, ne doseže želene ravni za ciljni sistem - dokumente, tabele ali nadzorne plošče (korak 7).

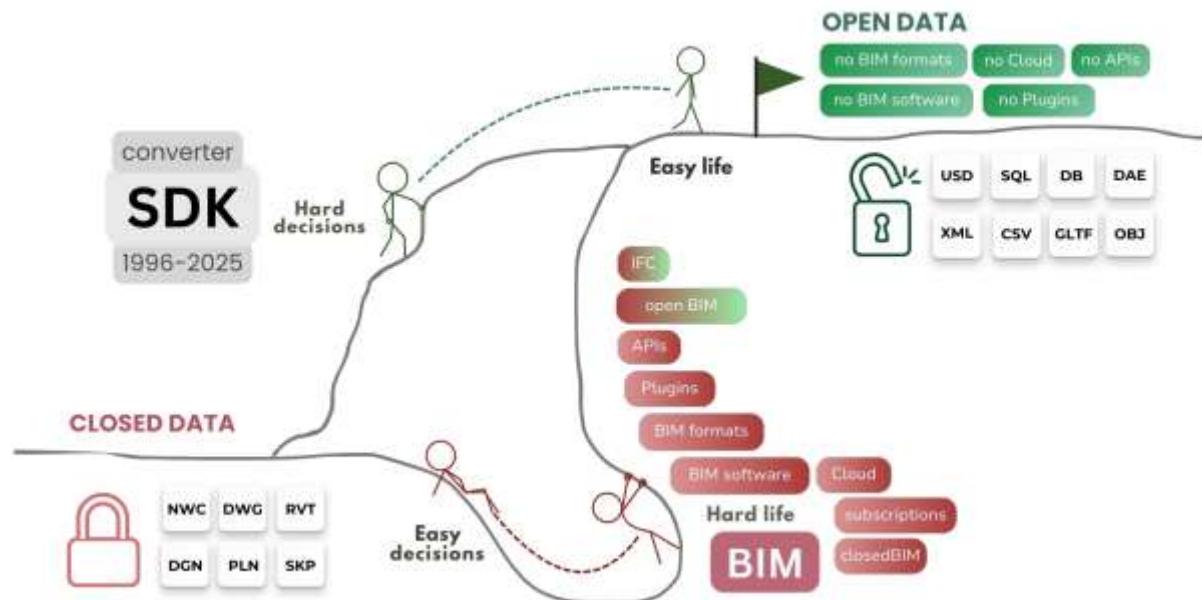


Slika 6.4-1 Cikel informacijske zasičenosti podatkovnih baz za poslovne procese pri izvajjanju gradbenih projektov.

Če CAD (BIM) obravnavamo kot mehanizem za prenos parametrov v obliki niza parov ključ-vrednost, ustvarjenih na podlagi zahtev, opredeljenih zunaj projektnega okolja (slika 6.4-1, korak 1-2), se težišče razprave premakne s specifičnih programskih rešitev in njihovih omejitev na temeljnješ vidike - podatkovno strukturo, podatkovne modele in podatkovne zahteve. V bistvu govorimo o parametrični zasičenosti podatkovne zbirke in klasičnem postopku podatkovnega modeliranja (koraki 2-3 in 5-6). Edina razlika je, da zaradi zaprte narave podatkovnih zbirk CAD in posebnosti uporabljenih formatov ta proces spremišča uporabo specializiranih orodij BIM. Postavlja se vprašanje: v čem je edinstvenost BIM, če v drugih panogah ni podobnih pristopov?

V zadnjih 20 letih je bil BIM predstavljen kot več kot le en vir podatkov. Paket CAD-BIM se pogosto trži kot parametrično orodje z vgrajeno zbirko podatkov [64], ki lahko avtomatizira postopke načrtovanja, modeliranja in upravljanja življenjskega cikla gradbenih projektov. Vendar je v resnici BIM postal bolj orodje za ohranjanje uporabnikov na platformi prodajalcev kot pa priročna metoda upravljanja podatkov in procesov.

Zato so podatki CAD- (BIM-) na njihovih platformah izolirani in skrivajo projektne informacije za lastniškimi API-ji in geometrijskimi jedri. To je uporabnike prikrajšalo za možnost neodvisnega dostopa do podatkovnih zbirk ter za pridobivanje, analizo, avtomatizacijo in prenos podatkov v druge sisteme mimo ekosistemov prodajalcev.



Slika 6.4-2 V gradbeništvu sodobni formati zahtevajo zapletena geometrijska jedra, vsako leto posodobljen API in posebne licence za programsko opremo CAD -(BIM -).

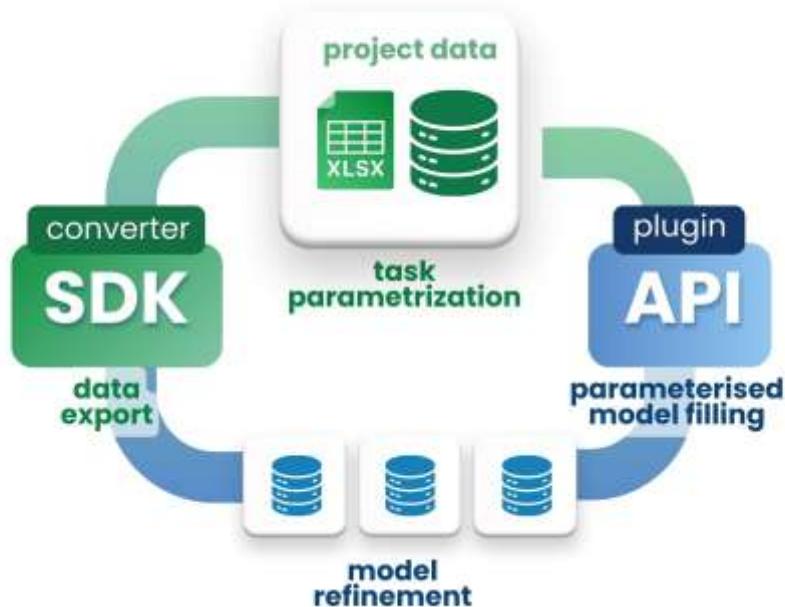
Podjetja, ki delajo s sodobnimi orodji CAD, bi morala pri delu s podatki uporabljati enak pristop, kot ga v praksi uporabljajo vsi prodajalci CAD brez izjeme sami: preoblikovanje podatkov z uporabo SDK - orodja za povratni inženiring, proti kateremu se prodajalci CAD borijo že od leta 1995 [75]. Če imamo popoln dostop do podatkovne zbirke CAD in uporabljamo orodja za obratno inženirstvo, lahko pridobimo [127] ploski nabor entitet z atributi in jih izvozimo v katero koli priročno odprto obliko (slika 6.4-2), vključno z geometrijo in parametri konstrukcijskih elementov. Ta pristop temeljito spremeni paradigmo dela z informacijami - od datotečno usmerjene k podatkovno usmerjeni arhitekturi:

- Podatkovni formati, kot so RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET in drugi, vsebujejo enake informacije o elementih istega projekta. To pomeni, da poznavanje določenega formata in njegove sheme ne sme biti ovira za delo s samimi podatki.
- Podatke iz vseh formatov je mogoče združiti v enotno odprto strukturirano in zrnato strukturo (slika 9.1-10), ki vsebuje trikotniško geometrijo MESH in lastnosti vseh entitet objektov brez omejitev geometrijskih jader.
- Analitika podatkov si prizadeva za univerzalnost: z uporabo odprtih podatkov lahko delate s projektnimi podatki ne glede na uporabljeno obliko.
- Zmanjšanje odvisnosti od vmesnikov API in vtičnikov prodajalcev: delo s podatki ni več odvisno od znanja API.

Ko se in CAD -podatkovne zahteve preoblikujejo v enostavno analizirljive strukturirane oblike predstavitev - razvijalci niso več odvisni od posebnih podatkovnih shem in zaprtih ekosistemov.

## Oblikovanje prek parametrov: prihodnost CAD in BIM

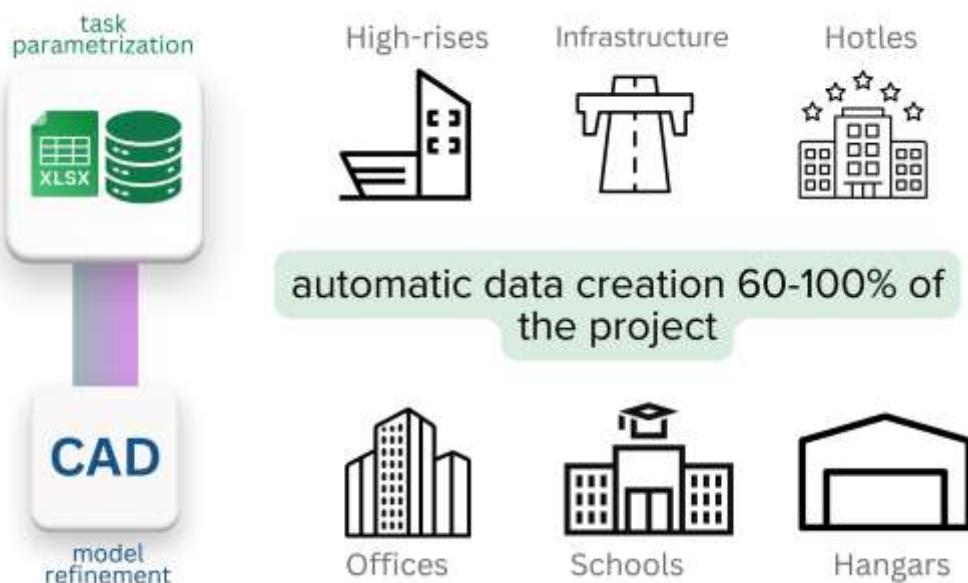
Še noben gradbeni projekt na svetu se ni začel v programu CAD. Preden risba ali model dobi obliko v programu CAD, gre skozi fazo konceptualizacije (slika 6.4-1, faze 1-2), kjer je poudarek na parametrih, ki določajo osnovno idejo in logiko prihodnjega objekta. Ta stopnja ustreza konceptualni ravni pri podatkovnem modeliraju (slika 4.3-6). Parametri lahko obstajajo zgolj v mislih oblikovalca, v idealnem primeru pa so urejeni v obliki strukturiranih seznamov, tabel ali shranjeni v podatkovnih bazah (slika 6.4-3), kar omogoča preglednost, ponovljivost in nadaljnjo avtomatizacijo postopka oblikovanja.



Slika 6.4-3 Postopek načrtovanja je ponavljajoči se proces polnjenja podatkovne zbirke CAD z informacijami od zunaj z uporabo zahtev v vrednostni verigi.

Pred začetkom samega modeliranja CAD (logična in fizična faza modeliranja podatkov (slika 4.3-7)) je treba opredeliti mejne parametre, ki so osnova projekta. Te lastnosti se, tako kot druge zahteve, zbirajo na samem koncu verige uporabe podatkov (npr. sistemi) in prek njih so že opredeljene omejitve, cilji in ključne značilnosti prihodnjih objektov v projektu.

Samo modeliranje je mogoče s pomočjo orodij za parametrično modeliranje (slika 6.4-3) popolnoma avtomatizirati za 60-100%, če so zahteve dobro opredeljene. Takoj ko je projekt opisan v obliki parametrov, postane njegovo oblikovanje tehnično izvedljivo, na primer s pomočjo vizualnih programskih jezikov, kot je Grasshopper Dynamo, vgrajenih v sodobne CAD -okolja ali brezplačne rešitve v programih Blender, UE, Omniverse.



**Slika 6.4-4 Večina tipiziranih projektov je danes zaradi orodij za parametrično programiranje ustvarjena že popolnoma samodejno.**

Velikih industrijskih in tipiziranih projektov že danes ne ustvarjajo roke oblikovalskega oddelka, temveč parametrična orodja in vizualno programiranje. To omogoča izdelavo modela na podlagi podatkov in ne na podlagi subjektivnih odločitev posameznega projektanta ali vodje.

*Vsebina ima prednost pred oblikovanjem. Oblikovanje brez vsebine ni oblikovanje, temveč dekoracija [128].*

- Jeffrey Zeldman, spletni oblikovalec in podjetnik

Postopek se ne začne z risanjem ali 3D -modeliranjem, temveč z oblikovanjem zahtev. Prav zahteve določajo, kateri elementi bodo uporabljeni v projektu, katere podatke je treba prenesti v druge oddelke in sisteme. Le obstoj strukturiranih zahtev omogoča samodejno preverjanje modelov (na primer tudi vsakih 10 minut, ne da bi pri tem odvrnili projektanta od dela).

Morda bo v prihodnosti sistem CAD- (BIM-) postal le vmesnik za polnjenje podatkovne baze, pri čemer ne bo pomembno, v katerem orodju CAD modeliranje izvedeno (fizična raven).

Podobno se pri strojništvu pogosto uporablja 3D modeliranje, ki pa ni nujen ali obvezen element projekta. V večini primerov zadostuje klasična 2D-dokumentacija, na njeni podlagi pa se ustvari potreben informacijski model. Ta model je sestavljen iz sestavnih delov, strukturiranih v skladu z industrijskimi standardi, in vsebuje vse potrebne informacije za razumevanje načrtovanja in organizacije proizvodnje. Na podlagi tovarniškega informacijskega modela se nato ustvari tovarniški informacijski model, ki se mu dodajo specifični izdelki in diagrami poteka, ki so že usmerjeni k

potrebam tehnologov. Celoten proces je mogoče organizirati brez nepotrebne zapletenosti, ne da bi sistem preobremenili s 3D grafikami, kjer te ne prinašajo pravih prednosti.

Pomembno je razumeti, da sam 3D model in sistem CAD ne smeta imeti glavne vloge - sta le orodje za kvantitativno in geometrijsko analizo. Vse druge parametre, razen geometrije, ki opisujejo entiteto, je treba shraniti in obdelati zunaj okolja CAD, če je to mogoče (BIM).

Oblikovanje s pomočjo parametrov ni le trend, temveč neizogibna prihodnost gradbene industrije. Namesto ročnega ustvarjanja kompleksnih 3D -modelov bodo projektanti delali s podatki, jih potrjevali in avtomatizirali procese, s čimer se bo gradbeništvo približalo svetu programiranja. Sčasoma bodo postopki projektiranja temeljili na načelih razvoja programske opreme:

- Ustvarjanje zahtev → Ustvarjanje modela → Prenos na strežnik → Potrjevanje sprememb → Zahteva za prenos
- V zahtevkih za prenos se samodejno izvajajo preverjanja modela glede na zahteve, ki so bile ustvarjene pred oblikovanjem ali med njim kot del zahtevka za prenos.
- Po preverjanju kakovosti podatkov in odobritvi se spremembe prenesejo v projekt, skupno podatkovno zbirko ali samodejno prenesejo v druge sisteme.

Že zdaj se v strojništvu takšne spremembe zaslove začnejo z oblikovanjem obvestila o spremembah. Podobna shema čaka tudi gradbeno industrijo: projektiranje bo iterativni proces, pri katerem bo vsak korak podprt s parametričnimi zahtevami. Takšen sistem bo projektantom omogočil oblikovanje samodejnih preverjanj in samodejnih zahtevkov za povišanje za posebne zahteve.

Oblikovalec prihodnosti je predvsem upravljač podatkov in ne ročni modelar. Njegova naloga je, da projekt napolni s parametričnimi entitetami, pri katerih je geometrija le eden od atributov.

Pomembno vlogo pri preoblikovanju bo imelo razumevanje pomena modeliranja podatkov, klasifikacije in standardizacije, ki so bili podrobno obravnavani v prejšnjih poglavjih knjige. Predpisi za oblikovanje v prihodnosti bodo formalizirani kot pari parametrov ključ-vrednost v obliki schem XLSX ali XML.

Prihodnost gradbeništva je v zbiranju podatkov, njihovi analizi, potrjevanju in avtomatizaciji procesov s pomočjo analitičnih orodij. BIM (ali CAD) ni končni cilj, temveč le stopnja razvoja. Ko bodo strokovnjaki spoznali, da lahko delajo neposredno s podatki in obidejo tradicionalna orodja CAD, bo sam izraz "BIM" postopek odstopil mesto konceptom uporabe strukturiranih in granularnih podatkov o gradbenih projektih.

Eden od ključnih dejavnikov, ki so pospešili preoblikovanje, je pojav velikih jezikovnih modelov (LLM) in orodij, ki temeljijo na njih. Te tehnologije spreminjajo način ravnanja s podatki o projektiranju, saj omogočajo dostop do informacij, ne da bi za to potrebovali poglobljeno poznavanje vmesnikov API ali rešitev prodajalcev. Z modeli LLM postane postopek oblikovanja zahteve in interakcije s podatki CAD

intuitiven in dostopen.

## Pojav LLM v procesih obdelave podatkov CAD za oblikovanje

Poleg razvoja orodij za dostop do podatkovnih zbirk CAD ter odprtih in poenostavljenih formatov CAD je revolucijo pri obdelavi podatkov o načrtovanju povzročil tudi pojav orodij LLM (Large Language Models). Medtem ko je bil v preteklosti dostop do informacij večinoma prek zapletenih vmesnikov in je zahteval znanje programiranja ter poznavanje API, je zdaj mogoče s podatki komunicirati z uporabo naravnega jezika.

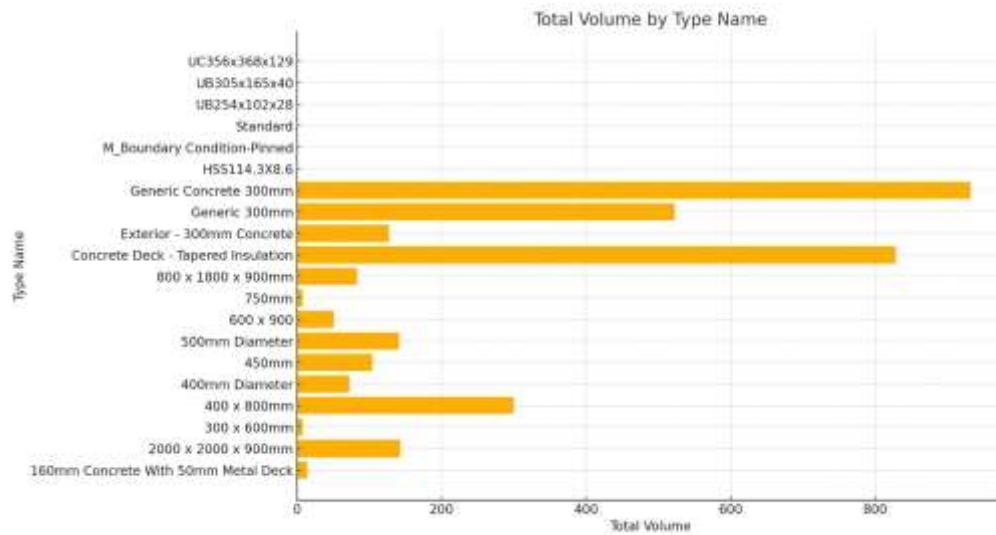
Inženirji, vodje in načrtovalci brez tehničnega znanja lahko pridobijo potrebne informacije iz projektnih podatkov z oblikovanjem poizvedb v običajnem jeziku. Če so podatki strukturirani in dostopni (slika 4.1-13), je dovolj, da v LLM klepetu zastavite vprašanje, kot je na primer: "Prikaži v tabeli z razvrščanjem po tipu vse stene s prostornino nad 10 kubičnih metrov" - in model bo to poizvedbo samodejno pretvoril v SQL ali kodo v Pandas, pri čemer bo ustvaril zbirno tabelo, graf ali celo končni dokument.

V nadaljevanju je nekaj primerov iz resničnega življenja, kako modeli LLM sodelujejo s projektnimi podatki, predstavljenimi v različnih formatih CAD- (BIM-).

- ❷ Primer poizvedbe v LLM chat na projekt CAD v formatu RVT po pretvorbi (slika 4.1-13) v tabelarni podatkovni okvir (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ali kateri koli drug):

Podatke v podatkovnem okviru, pridobljene iz datoteke RVT, pri seštevanju parametra "Obseg" združite v skupine po "imenu vrste" in prikažite število elementov v skupini. Vse to prikažite kot vodoravni histogram brez ničelnih vrednosti.

- Odziv LLM v obliki vodoravnega stolpčnega grafa (format PNG):



Slika 6.4-5 Namesto 17 klikov z miško ali 40 vrstic kode z uporabo vtičnikov v LLM takoj pridobimo tabelo QTO z besedilno poizvedbo.

- Za izdelavo preglednice QTO vrst sten s skupno površino in količino iz kategorije "Stene" oblikujmo besedilno poizvedbo za LLM -chat:

Iz podatkovnega okvira projekta vzemite samo tiste elemente, ki imajo v parametru "Kategorija" navedeno "OST\_Stene", jih združite po "Ime vrste", seštejte vrednost stolpca "Površina", dodajte količino in jih prikažite v preglednici tako, da odstranite ničelne vrednosti.

- Odgovor LLM v obliki dokončane tabele QTO:

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

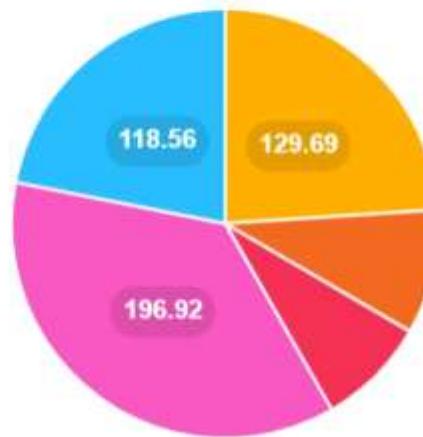
**Slika 6.4-6 Izdelava preglednice QTO v naravnem jeziku zagotavlja enako kakovosten rezultat kot pri uporabi orodij CAD - (BIM- ).**

- Poizvedujmo po projektu v formatu IFC po pretvorbi v tabelarični podatkovni okvir in vnesimo podobno besedilno poizvedbo v kateri koli klepetalnik LLM:

Iz projekta vzemite samo tiste elemente, ki imajo v parametru "Starš" vrednosti 1. in 2. ravni, in elemente, ki imajo v parametru "Kategorija" vrednosti IfcSlab, nato te elemente združite po parametru "Vrsta objekta", seštejte vrednosti v parametru "PSet\_RVT \_Dimensions Area" in jih prikažite kot krožni diagram.

■ Odziv LLM kot končni krožni diagram skupin elementov iz podatkov IFC :

For Floor:127mm Slab on Grade, Floor:150mm Exterior Slab on Grade,  
 Floor:Finish Floor - Ceramic Tile, Floor:Finish Floor - Wood, and  
 Floor:Residential - Wood Joist with Subflooring



**Slika 6.4-7 Rezultat poizvedbe po podatkih IFC v strukturirani obliki je lahko katera koli vrsta grafa, ki je primerna za razumevanje podatkov.**

Za vsako od pripravljenih rešitev (slika 6.4-5 - slika 6.4-7) se skriva ducat vrstic kode Python z uporabo knjižnice Pandas. Nastalo kodo lahko kopirate iz klepetalnice LLM in jo uporabite v katerem koli lokalnem ali spletnem IDE, da dobite enake rezultate zunaj klepetalnice LLM.

V istem klepetu LLM lahko delamo ne le s projekti, pridobljenimi iz formatov 3D CAD (BIM), temveč tudi s ploščatimi risbami v formatu DWG, za katere lahko v klepetu LLM poizvedujemo in prikažemo na primer podatke o skupinah elementov v obliki črt ali 3D geometrije po pretvorbi v strukturirano obliko.

## Avtomatizirana analiza datotek DWG z LLM in Pandas

Postopek obdelave podatkov iz datotek DWG je bil zaradi nestrukturirane narave informacij vedno zapletena naloga, ki je zahtevala specializirano programsko opremo in pogosto ročno analizo. Vendar je z razvojem umetne inteligenčne in orodij LLM postalo mogoče avtomatizirati številne korake tega, danes večinoma ročnega postopka. Oglejmo si pravi Pipeline zahtevkov za LLM (v tem primeru ChatGPT) za delo z risbami DWG, ki omogočajo delo s projektom:

- Filtriranje podatkov DWG po sloju, ID in koordinatah
- Vizualizacija geometrije elementov
- Samodejno komentiranje risb na podlagi parametrov
- Razširite poliline sten na vodoravno ravnino

- Ustvarjanje interaktivnih 3D - vizualizacije ravninskih podatkov
- Strukturirajte in analizirajte gradbene podatke brez zapletenih orodij CAD

V našem primeru se postopek gradnje cevovoda začne z zaporednim generiranjem kode prek LLM. Najprej se ustvari poizvedba, ki opisuje nalogu. ChatGPT ustvari kodo Python, ki se izvede in analizira ter prikaže rezultat znotraj klepetalnice. Če rezultat ni v skladu s pričakovanji, se zahteva popravi in postopek se ponovi.

**Pipeline** je zaporedje samodejnih korakov, ki se izvajajo za obdelavo in analizo podatkov. V takem postopku vsak korak prevzame podatke kot vhodne podatke, izvede pretvorbe in rezultat posreduje naslednjemu koraku.

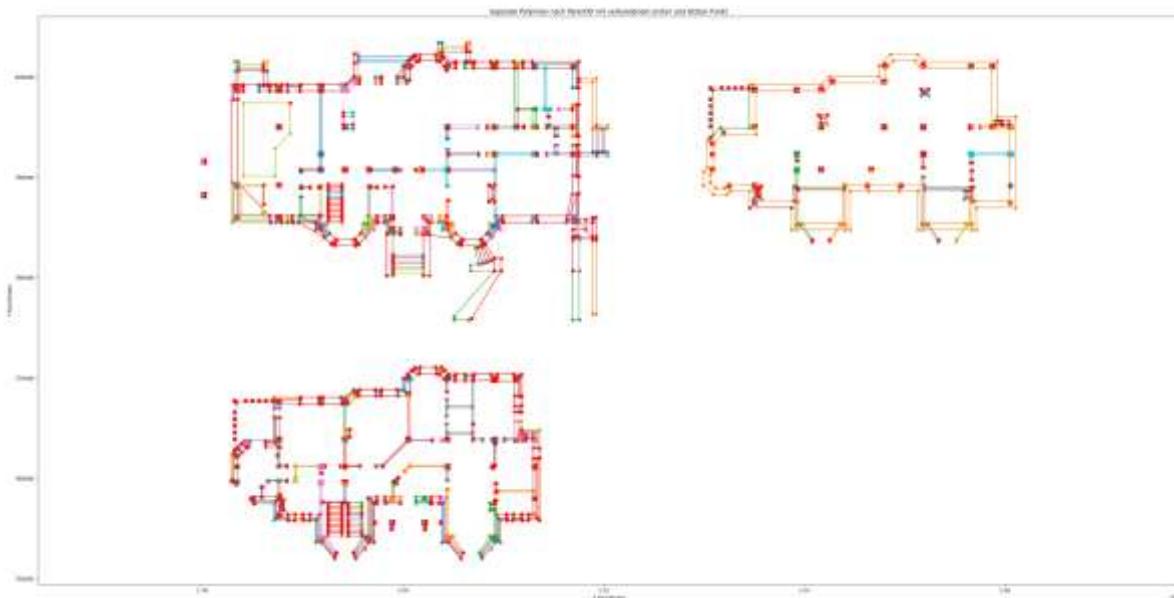
Po pridobitvi želenega rezultata kodo kopiramo iz LLM in jo v obliki blokov prilepimo v katero od priročnih IDE, v našem primeru na platformi Kaggle.com. Dobljeni deli kode so združeni v enotno cevovodno linijo, ki avtomatizira celoten postopek - od nalaganja podatkov do njihove končne analize. Ta pristop omogoča hiter razvoj in razširjanje analitičnih procesov brez poglobljenega programerskega znanja. Celotno kodo vseh spodnjih delov skupaj z vzorčnimi poizvedbami lahko najdete na platformi Kaggle.com z iskanjem "DWG Analyse with ChatGPT | DataDrivenConstruction". [129].

Postopek dela s podatki DWG po pretvorbi v strukturirano obliko (slika 4.1-13), začimo s klasičnim korakom - združevanjem in filtriranjem vseh podatkov risbe, ki so za našo nalogu potrebni elementi stene, natančneje polilinij (parameter "ParentID" omogoča združevanje linij v skupine), ki imajo v parametru (stolpec podatkovnega okvira) "Layer" vrednost niza, ki vsebuje naslednjo kombinacijo črk (RegEx) - "wall".

- ➲ Če želite dobiti kodo za podobno nalogu in rezultat v obliki slike, morate v LLM napisati naslednjo poizvedbo:

Najprej preverite, ali podatkovni okvir, pridobljen iz DWG, vsebuje opredeljene stolpce: "Layer", "ID", "ParentID" in "Point". Nato iz stolpca "Layer" filtrirajte ID-je, ki vsebujejo niz "wall". V stolpcu "ParentID" poiščite elemente, ki ustrezajo tem identifikatorjem. Določite funkcijo za čiščenje in razdelitev podatkov v stolpcu "Point". To vključuje odstranitev oklepajev in razdelitev vrednosti na koordinate "x", "y" in "z". Podatke prikažite z uporabo matplotlib. Za vsako edinstveno "ParentID" narišite ločeno poljsko črto, ki povezuje koordinate "Point". Poskrbite, da bosta prva in zadnja točka povezani, če je to mogoče. Nastavite ustrezne oznake in naslove ter poskrbite, da sta osi x in y enako pomanjšani.

- ➲ Odgovor LLM vam bo dal pripravljeno sliko, za katero se skriva koda Pythona, ki jo je ustvarila:



Slika 6.4-8 Koda LLM je iz datoteke DWG izbral vse črte sloja "zid", očistila njihove koordinate in konstruirala poligone s pomočjo ene od knjižnic Python.

- Zdaj dodajmo linijam parameter površine, ki ga ima vsaka polilinija v svojih lastnostih (v enem od stolpcev podatkovnega okvira):

Zdaj iz vsake poljine izberite samo en "ParentID" - poiščite ta ID v stolpcu "ID", vzemite vrednost "Area", jo delite z 1.000.000 in dodajte to vrednost v graf.

- Odziv LLM bo prikazal nov graf, v katerem bo imela vsaka polilinija napis s svojo površino:

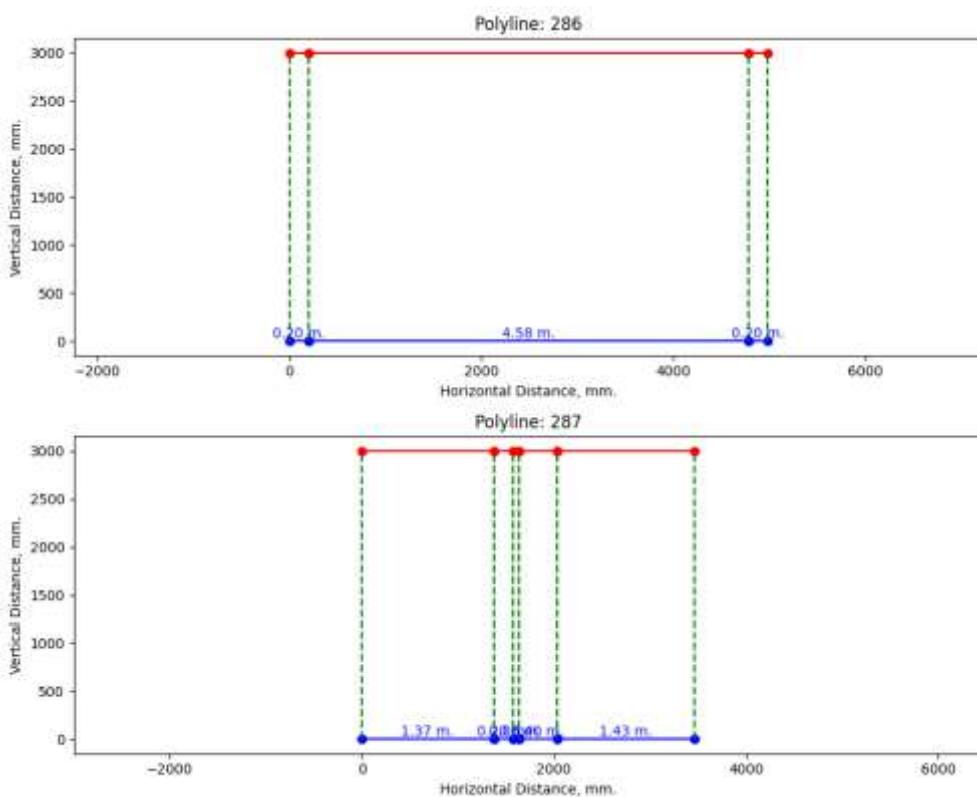


**Slika 6.4-9 LLM je dodal kodo, ki za vsako poljsko črto vzame vrednosti površin in jih doda na sliko z vizualizacijo črt.**

- Nato bomo vsako polilinijo pretvorili v vodoravno črto, dodali vzporedno črto na višini 3000 mm in jih povezali v enotno ravnilo, da bi na ta način prikazali razporeditev površin stenskih elementov:

Iz stolpca "Layer" morate vzeti vse elemente z vrednostjo "wall". Te ID-je vzemite kot seznam iz stolpca "ID" in jih poiščite v celotnem podatkovnem okviru v stolpcu "ParentID". Vsi elementi so črte, ki so združene v eno samo poljsko črto. Vsaka črta ima drugačno geometrijo x, y prve točke v stolpcu "Point" (Točka). Vzeti morate vsako polilinijo po vrsti in iz točke 0,0 vodoravno narisati dolžino vsakega segmenta iz poliline. dolžino vsakega segmenta poliline v eno črto. Nato narišite popolnoma enake črte, le 3000 višje, in vse točke povežite v eno ravnilo.

- ❷ Odziv LLM bo prikazal kodo, ki omogoča izrisovanje risb sten v ravnini:

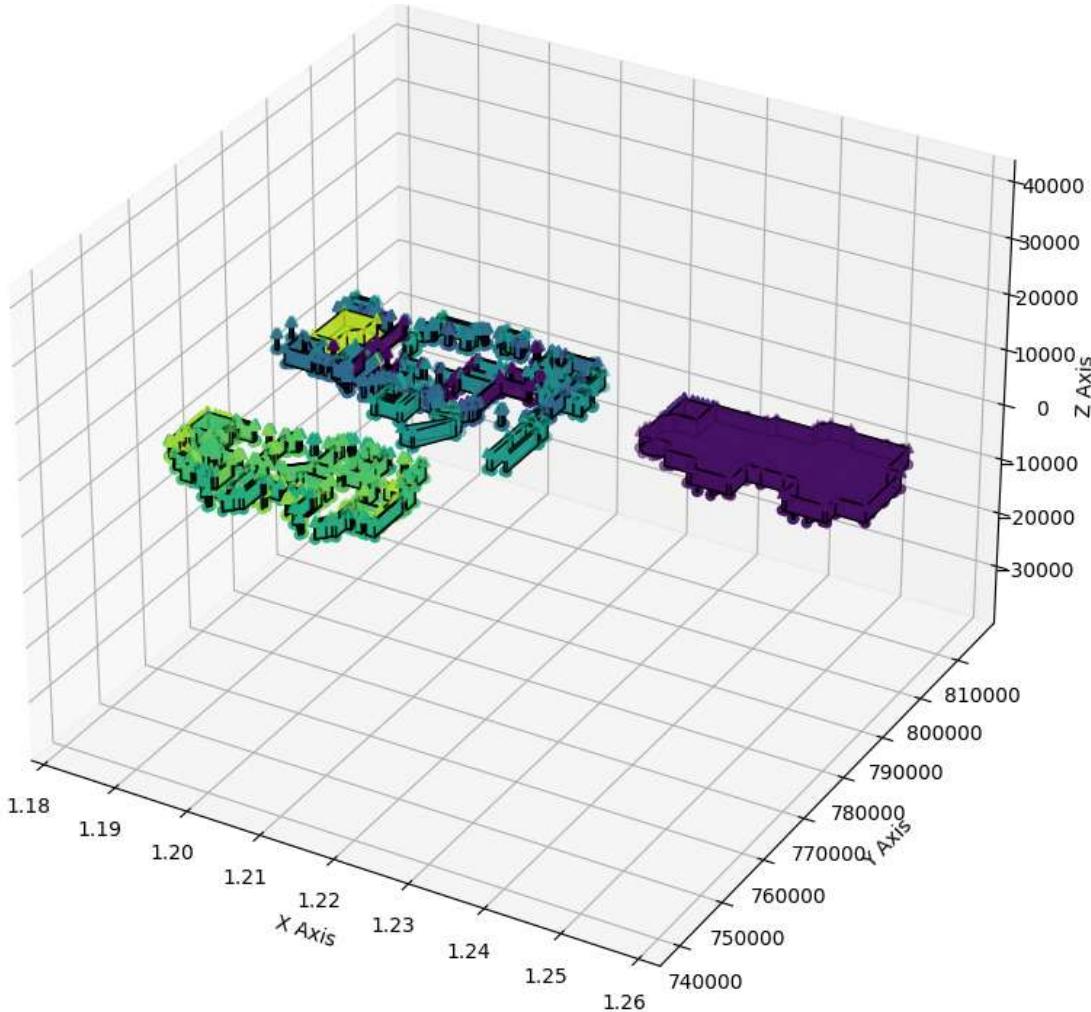


Slika 6.4-10 Vsako poligonsko črto s pomočjo napotkov spremenimo v postavitev, ki vizualizira stenske ploskve neposredno v klepetu LLM.

- ❸ Sedaj preidimo od 2D projekcije k 3D - modeliranje sten iz ravnih črt s povezovanjem zgornjega in spodnjega sloja polilinij:

Vizualizirajte stenske elemente v 3D, pri čemer povežite poligone na višinah  $z = 0$  in  $z = 3000 \text{ mm}$ . Ustvarite zaproto geometrijo, ki predstavlja stene stavbe. Uporabite Matplotlibovo orodje za izdelavo 3D grafik.

- LLM bo ustvaril interaktivni 3D -graf, v katerem bo vsaka polilinija predstavljena kot niz ravnin. Uporabnik se bo lahko z računalniško miško prosto premikal med elementi in raziskoval model v 3D načinu s kopiranjem kode iz klepeta v IDE:



Slika 6.4-11 LLM je pomagal zgraditi kodo [129] za vizualizacijo ravnih risalnih črt v 3D prikaz, ki ga je mogoče raziskati v 3D pregledovalniku v IDE.

Za vzpostavitev logičnega in ponovljivega cevovoda - od začetne pretvorbe in nalaganja datoteke DWG do končnega rezultata - je priporočljivo, da po vsakem koraku kopirate ustvarjeni blok kode LLM v IDE. Na ta način ne le preverite rezultat v klepetu, temveč ga tudi takoj zaženete v svojem razvojnem okolju. To vam omogoča zaporedno izgradnjo postopka, odpravljanje napak in prilagajanje po potrebi.

Celotno kodo cevovoda vseh fragmentov (slike 6.4-8 do slike 6.4-11) skupaj z vzorčnimi poizvedbami lahko najdete na platformi Kaggle.com z iskanjem "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. Na portalu Kaggle si lahko ne le ogledate kodo in uporabljeni

poizvedbe, temveč tudi brezplačno kopirate in preizkusite celoten Pipeline z izvirnimi podatkovnimi okvirji DWG v oblaku, ne da bi morali namestiti kakršno koli dodatno programsko opremo ali sam IDE.

Pristop, predstavljen v tem poglavju, vam omogoča popolno avtomatizacijo preverjanja, obdelave in generiranja dokumentov na podlagi DWG -projektov. Razvita cevovodna linija je primerna tako za obdelavo posameznih risb kot za paketno obdelavo deset, sto in tisoč datotek DWG s samodejnim generiranjem potrebnih poročil in vizualizacij za vsak projekt.

Postopek je mogoče organizirati zaporedno in pregledno: najprej se podatki iz datoteke CAD samodejno pretvorijo v obliko XLSX, nato se naložijo v podatkovni okvir, sledijo razvrščanje v skupine, preverjanje in ustvarjanje rezultatov - vse to se izvede v enem samem beležnici Jupyter ali Python - skriptu, v katerem koli priljubljenem IDE. Po potrebi je mogoče postopek preprosto razširiti z integracijo s sistemi za upravljanje projektne dokumentacije: datoteke CAD je mogoče samodejno priklicati v skladu z določenimi merili, rezultate je mogoče vrniti nazaj v sistem za shranjevanje, uporabnike pa je mogoče obvestiti, ko so rezultati pripravljeni - prek e-pošte ali sporočilnikov.

Uporaba klepetalnic in agentov LLM za delo s podatki o projektiranju zmanjšuje odvisnost od specializiranih programov CAD ter omogoča izvajanje analize in vizualizacije arhitekturnih načrtov brez potrebe po ročni interakciji z vmesnikom - brez klikov z miško in pomnjenja zapletenih menujev.

Z vsakim dnem bo gradbena industrija slišala vse več o LLM, zrnatih strukturiranih podatkih, DataFrames in stolpčnih podatkovnih bazah. Enotni dvodimensionalni DataFrame, oblikovani iz različnih podatkovnih zbirk in formatov CAD, bodo idealno gorivo za sodobna analitična orodja, s katerimi se aktivno ukvarjajo strokovnjaki v drugih panogah.

Sam postopek avtomatizacije bo bistveno poenostavljen - namesto preučevanja API zaprtih nišnih izdelkov in pisanja zapletenih skript za analizo ali preoblikovanje parametrov bo zdaj dovolj, da se naloga oblikuje v obliku niza posameznih besedilnih ukazov, ki se bodo zložili v zahtevani cevovod ali proces poteka dela za zahtevani programski jezik, ki brezplačno deluje na skoraj vsaki napravi. Nič več čakanja na nove izdelke, formate, vtičnike ali posodobitve proizvajalcev orodij CAD- (BIM-). Inženirji in gradbeniki bodo lahko samostojno delali s podatki z uporabo preprostih, brezplačnih in razumljivih orodij, pri čemer jim bodo pomagali klepetalnice in agenti LLM.

## Naslednji koraki: prehod z zaprtih oblik na odprte podatke

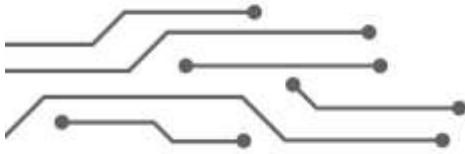
Pri delu z oblikovalskimi podatki prihodnosti verjetno ne bo treba razumeti geometrijskih jeder lastniških orodij ali se učiti več sto nezdružljivih formatov, ki vsebujejo iste informacije. Vendar je brez razumevanja, zakaj je prehod na odprte strukturirane podatke pomemben, težko zagovarjati uporabo novih brezplačnih orodij, odprtih podatkov in pristopov, ki jih prodajalci programske opreme verjetno ne bodo spodbujali.

V tem poglavju smo obravnavali ključne značilnosti podatkov CAD (BIM), njihove omejitve in priložnosti ter dejstvo, da se inženirji in projektanti kljub marketinškim obljudbam prodajalcev vsak dan srečujejo s

težavami pri pridobivanju, prenašanju in analiziranju projektnih informacij. Razumevanje arhitekture teh sistemov in spoznavanje alternativnih pristopov - ki temeljijo na odprtih formatih in avtomatizaciji z LLM - lahko olajša življenje že posameznemu strokovnjaku, kaj šele podjetjem. Za povzetek tega dela velja izpostaviti glavne praktične korake, ki vam bodo pomagali pri uporabi pristopov, obravnavanih v tem poglavju, pri vaših vsakodnevnih nalogah:

- Razširite svoj nabor orodij za delo s podatki o projektih
  - Raziščite razpoložljive vtičnike in pripomočke za pridobivanje podatkov iz sistemov CAD - (BIM-), ki jih uporabljate.
  - Raziščite razpoložljive komplete SDK in vmesnike API, ki vam omogočajo samodejno pridobivanje podatkov iz zaprtih formatov, ne da bi vam bilo treba ročno odpirati specializirano programsko opremo.
  - Osvojite osnovna znanja za delo z odptimi neparametričnimi geometrijskimi formati (OBJ, glTF, USD, DAE) in ustrezni odprtokodnimi knjižnicami.
  - Poskusite si zamisliti sistem za shranjevanje metapodatkov projekta ločeno od geometrije zunaj rešitev CAD (BIM), da bi poenostavili analizo in integracijo z drugimi sistemi.
  - Uporabite LLM za avtomatizacijo pretvorbe podatkov med formati.
- Ustvarite lastne postopke za ravnanje s projektnimi informacijami.
  - Začetek opisovanja nalog in zahtev modeliranja s parametri in njihovimi vrednostmi v preprostih in strukturiranih oblikah.
  - Ustvarite osebno knjižnico skript ali blokov kode za pogosto izvajane operacije.
- Spodbujanje uporabe odprtih standardov pri svojem delu
  - Povabite sodelavce in partnerje, da delijo podatke v odprtih formatih, ki jih ne omejuje ekosistem proizvajalcev programske opreme.
  - predstavitev prednosti uporabe strukturiranih podatkov s konkretnimi primeri
  - sprožite razpravo o težavah z zaprtimi formati in možnih rešitvah.

Tudi če ne morete spremeniti politike svojega podjetja glede platform CAD - (BIM-), vam bo osebno razumevanje načel dela s projektnimi podatki v odprtih formatih omogočilo bistveno povečati učinkovitost vašega dela. Z ustvarjanjem lastnih orodij in metod za pridobivanje in preoblikovanje podatkov iz različnih formatov ne boste le optimizirali svojih delovnih postopkov, temveč boste pridobili tudi prožnost, s katero boste lahko zaobšli omejitve standardnih programskih rešitev.



## VII DEL

# **ODLOČANJE NA PODLAGI PODATKOV, ANALITIKA, AVTOMATIZACIJA IN STROJNO UČENJE.**

Sedmi del se osredotoča na podatkovno analitiko in avtomatizacijo procesov v gradbeništvu. Obravnava, kako podatki postanejo podlaga za sprejemanje odločitev, in pojasnjuje načela vizualizacije informacij za učinkovito analizo. Podrobno so opisani ključni kazalniki uspešnosti (KPI), metode za ocenjevanje donosnosti naložb (ROI) in oblikovanje nadzornih plošč za spremljanje projektov. Posebna pozornost je namenjena postopkom ETL (Extract, Transform, Load) in njihovi avtomatizaciji z uporabo cevovodov (Pipeline), da se raznovrstni podatki spremeniijo v strukturirane informacije za analizo. Obravnavana so orodja za orkestracijo delovnih tokov, kot so Apache Airflow, Apache NiFi in n8n, ki omogočajo gradnjo avtomatiziranih podatkovnih cevovodov brez poglobljenega znanja programiranja. Pomembno vlogo imajo veliki jezikovni modeli (LLM) in njihova uporaba za poenostavitev analize podatkov in avtomatizacijo rutinskih opravil.

## POGLAVJE 7.1.

### **PODATKOVNA ANALITIKA IN ODLOČANJE NA PODLAGI PODATKOV.**

Po korakih zbiranja, strukturiranja, čiščenja in preverjanja informacij je nastal skladen in analizabilen nabor podatkov. Prejšnji deli knjige so obravnavali sistematizacijo in strukturiranje heterogenih virov - od dokumentov PDF in besedilnih zapisov sestankov do modelov CAD in geometrijskih podatkov. Podrobno je opisan postopek preverjanja in usklajevanja informacij z zahtevami različnih sistemov in klasifikatorjev ter odpravljanja podvajanj in nedoslednosti.

Vsi izračuni, ki se izvajajo na teh podatkih (tretji in četrti del knjige) - od preprostih pretvorb do izračunov časa, stroškov in kazalnikov ESG (peti del) - so združene analitične naloge. So podlaga za razumevanje trenutnega stanja projekta, ocenjevanje njegovih parametrov in nato sprejemanje odločitev. Tako se podatki kot rezultat izračunov iz množice različnih zapisov spremenijo v obvladljiv vir, ki lahko odgovori na ključna poslovna vprašanja.

V prejšnjih poglavjih so bili podrobno opisani postopki zbiranja podatkov in nadzora kakovosti za uporabo v tipičnih poslovnih primerih in procesih, značilnih za gradbeništvo. Analitika v tem kontekstu je v marsičem podobna aplikacijam v drugih panogah, vendar ima številne posebnosti.

V naslednjih poglavjih bo podrobno opisan postopek analize podatkov, vključno s koraki avtomatizacije od začetnega pridobivanja informacij in njihovega preoblikovanja do poznejšega prenosa v ciljne sisteme in dokumente. Najprej bo predstavljen teoretični del, ki se osredotoča na izbrane vidike odločanja na podlagi podatkov. Nato bo v naslednjih poglavjih predstavljen praktični del, povezan z avtomatizacijo in izgradnjo ETL -Pipeline.

#### **Podatki kot vir pri sprejemanju odločitev**

Odločanje na podlagi podatkov je pogosto ponavljajoč se proces, ki se začne s sistematičnim zbiranjem informacij iz različnih virov informacij. Podobno kot naravni cikel posamezni podatkovni elementi in celotni informacijski sistemi postopoma padajo v zemljo - kopijo se v informacijskih skladiščih podjetij (slika 1.3-2). Sčasoma se ti podatki, podobno kot odpadlo listje in veje, spremenijo v dragoceno gradivo. Michel podatkovnih inženirjev in analitikov organizira in pripravlja informacije za prihodnjo uporabo ter spreminja odpadle podatke in sisteme v dragocen kompost, da bi zrasli novi poganjenki in novi sistemi (slika 1.2-5).

Trendi razširjene uporabe analitike v različnih panogah pomenijo začetek novega obdobja, v katerem delo s podatki postaja temelj poklicne dejavnosti (slika 7.1-1). Za strokovnjake v gradbeništvu je pomembno, da se prilagodijo tem spremembam in so pripravljeni na prehod v novo dobo - dobo podatkov in analitike

Ročno premikanje podatkov med tabelami in ročno izvajanje izračunov postopoma postajata stvar preteklosti, saj se umikajo avtomatizaciji, analizi pretoka podatkov, analitiki in strojnemu učenju. Ta orodja postajajo ključni elementi sodobnih sistemov za podporo odločjanju.

V McKinseyjevi knjigi "Rebooting. vodnik za premagovanje konkurenčne v dobi digitalne tehnologije in umetne inteligence" [130] je navedena raziskava, ki je bila leta 2022 izvedena med 1 330 vodstvenimi delavci iz različnih regij, panog in funkcionalnih področij [130]. navaja raziskavo, ki je bila izvedena leta 2022 s 1 330 vodilnimi delavci iz različnih regij, panog in funkcionalnih področij. Glede na njene rezultate 70% vodilnih uporablja napredno analitiko za ustvarjanje lastnih zamisli, 50% pa izvaja umetno inteligenco za izboljšanje in avtomatizacijo procesov odločanja.



Slika 7.1-1 Analiza podatkov in analitika sta glavni orodji za povečanje hitrosti odločanja v podjetju.

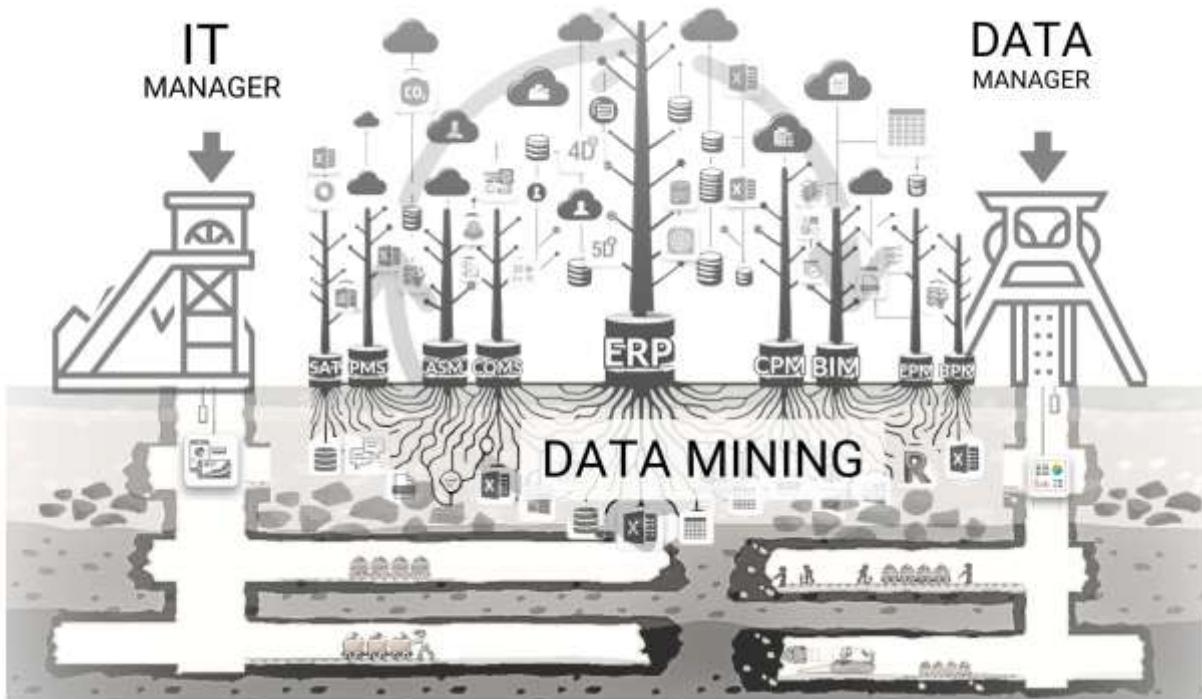
Analiza podatkov, podobno kot širjenje micelija, prodira v humus preteklih odločitev, pomaga povezati posamezne sisteme in vodje vodi do dragocenih spoznanj. To znanje, podobno kot hrana iz propadlih dreves podatkovnih sistemov, napaja nove odločitve v podjetju, kar vodi k učinkovitim spremembam in rasti kakovostnih informacij, podobno kot novi poganjki in kalčki, ki vznikajo iz bogate in zdrave zemlje (slika 1.2-5).

*Številke so pomembna zgodba. Od vas pričakujejo, da jim boste dali jasen in prepričljiv glas [131].*

- Stephen Few, strokovnjak za vizualizacijo podatkov

V srednje velikih in malih podjetjih je danes pridobivanje in priprava informacij za nadaljnjo analizo izredno delovno intenziven proces (slika 7.1-2), primerljiv s pridobivanjem premoga v osemnajstem stoletju. Do nedavnega je bilo delo pridobivanja in priprave podatkov bolj rezervirano za avanturiste, ki so delali v visoko specializirani niši z majhnim in omejenim naborom orodij za delo z različnimi vrstami podatkov iz nestrukturiranih, ohlapno strukturiranih, mešanih in zaprtih virov.

Odločevalci in vodje pogosto nimajo izkušenj s heterogenimi podatki in sistemmi, vendar morajo na njihovi podlagi sprejemati odločitve. Zato je bilo v zadnjih desetletjih odločanje na podlagi podatkov v sodobni gradbeni industriji manj podobno avtomatiziranemu procesu in bolj večdnevnu ročnemu delu rudarja v prvih premogovnikih.



Slika 7.1-2 V procesu podatkovnega rudarjenja strokovnjaki prehodijo zapleteno pot priprave podatkov - od čiščenja do strukturiranja za nadaljnjo analitiko.

Čeprav so sodobne metode pridobivanja podatkov v gradbeništvu vsekakor naprednejše od primitivnih tehnik rudarjev iz 12. stoletja, je to še vedno zapletena in tvegana naloga, ki zahteva veliko sredstev in strokovnega znanja, kar si lahko privoščijo le velika podjetja. Postopke pridobivanja in analize podatkov iz nakopičene zapaščine preteklih projektov so do nedavnega izvajala predvsem velika, tehnološko napredna podjetja, ki podatke dosledno zbirajo in shranjujejo že desetletja.

Prej so imela vodilno vlogo na področju analitike tehnološko zrela podjetja, ki so podatke zbirala že desetletja. Danes se razmere spreminjajo: dostop do podatkov in orodij za obdelavo podatkov se demokratizira - prej zapletene rešitve so zdaj brezplačno na voljo vsem.

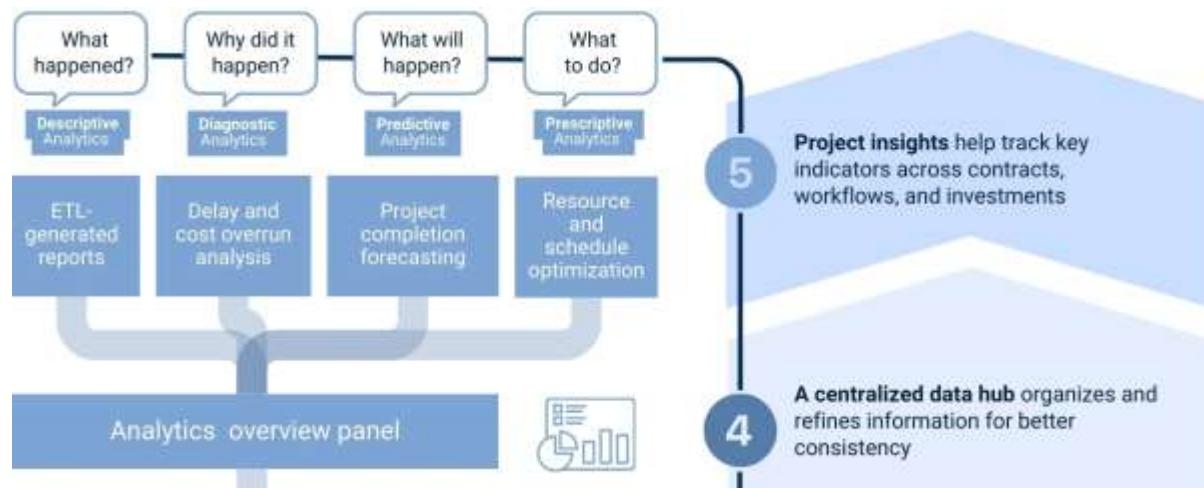
Uporaba analitike podjetjem omogoča sprejemanje natančnejših in bolj informiranih odločitev v realnem času. Naslednja študija primera ponazarja, kako lahko zgodovinski podatki pomagajo pri sprejemanju finančno utemeljenih odločitev:

- ☛ **Vodja projekta** - "Zdaj je povprečna cena betona v mestu 82€ /m<sup>3</sup>, v oceni imamo 95 €/m<sup>3</sup>."
- ☛ **Ocenjevalec** - "Pri prejšnjih projektih je bila prekoračitev približno 15-odstotna, zato sem se umaknil."
- ☛ **Vodja podatkov ali nadzorni inženir na strani stranke**: "Poglejmo analitiko zadnjih treh razpisov."

Po analizi podatkovnega okvira DataFrame iz preteklih projektov dobimo:

- **Povprečna dejanska nakupna cena:** 84,80 /m<sup>3</sup>€
- **Povprečni delež presežene porabe:** +4,7%.
- **Priporočena stopnja v oceni:** ~ 85 /m<sup>3</sup>€

Takšna odločitev ne bo več temeljila na subjektivnih občutkih, temveč na konkretnih zgodovinskih statističnih podatkih, kar pomaga zmanjšati tveganja in povečati veljavnost razpisne ponudbe. Analiza podatkov iz preteklih projektov postane neke vrste "organsko gnojilo", iz katerega vzklijejo nove, natančnejše rešitve.



Slika 7.1-3 Analitika podatkov odgovarja na tri ključna vprašanja: kaj se je zgodilo, zakaj se je zgodilo in kaj je treba storiti naprej.

Odločevalci in vodstveni delavci se pogosto soočajo s potrebo po delu s heterogenimi podatki in sistemi brez zadostnega tehničnega znanja in izkušenj. V takih primerih je vizualizacija, ki je eden prvih in najpomembnejših korakov v analitičnem procesu, ključna pomoč pri razumevanju podatkov. Omogoča predstavitev informacij v vizualni in razumljivi obliki.

## Vizualizacija podatkov: ključ do razumevanja in odločanja

V današnji gradbeni industriji, za katero sta značilni kompleksnost in večnivojska struktura projektnih podatkov, ima vizualizacija ključno vlogo. Vizualizacija podatkov omogoča vodjem projektov in inženirjem, da vizualizirajo zapletene vzorce in trende, skrite v velikih, heterogenih količinah podatkov.

Z vizualizacijo podatkov na spletni strani je lažje razumeti stanje projekta: dodeljevanje virov, stroškovne tende ali porabo materiala. Grafi in diagrami naredijo kompleksne in suhoperne informacije dostopne in razumljive, kar vam omogoča, da hitro prepoznate ključna področja, ki potrebujejo pozornost, in odkrijete morebitne težave.

Vizualizacija podatkov ne le olajša razlago informacij, temveč je tudi ključni korak v analitičnem

postopku in pri sprejemanju odločitev o upravljanju na podlagi informacij, saj pomaga *odgovoriti na vprašanji "kaj se je zgodilo?" in "kako se je to zgodilo?"* (slika 2.2-5). (Slika 2.2-5).

*Grafike so vizualna orodja za reševanje logičnih problemov [132].*

- Jacques Bertin, "Grafika in grafična obdelava informacij

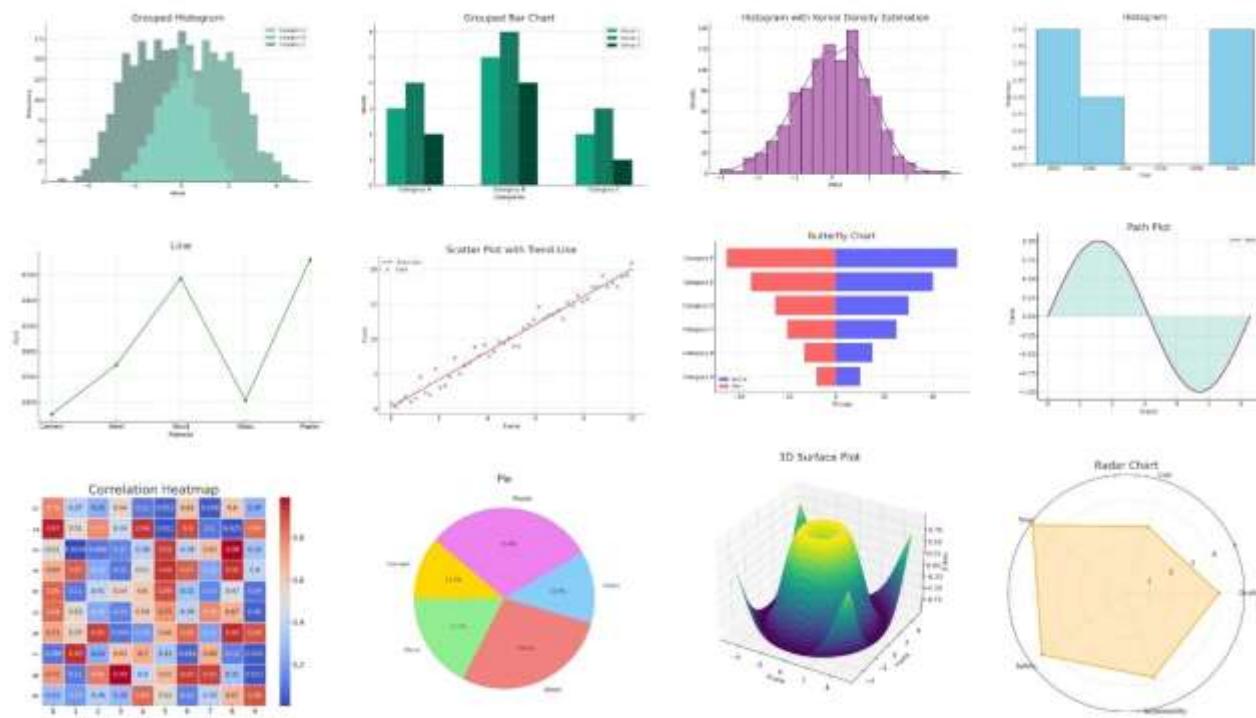
Pred sprejemanjem ključnih odločitev vodje projektov pogosteje uporabljajo vizualne predstavitev podatkov kot suhe in težko razumljive številke iz preglednic ali besedilnih sporočil.

Podatki brez vizualizacije so kot gradbeni material, naključno razmetan po gradbišču: njihov potencial je nejasen. Njihova vrednost postane jasna šele, ko so jasno vizualizirani, kot hiša iz opeke in betona. Dokler hiša ni zgrajena, ni mogoče reči, ali bo kup materialov postal majhna koča, luksuzna vila ali nebotičnik.

Podjetja imajo podatke iz različnih sistemov (Slika 1.2-4 do Slika 2.1-10), finančne transakcije in obsežne besedilne podatke. Vendar je uporaba teh podatkov za poslovne koristi pogosto zahtevna. V takšnih primerih postane vizualizacija pomembno orodje za sporočanje pomena podatkov, saj pomaga predstaviti informacije v oblikah, ki jih lahko razume vsak strokovnjak, kot so nadzorne plošče, grafikoni in diagrami.

Študija PwC "Kaj potrebujejo študenti za uspeh v hitro spreminjajočem se poslovnom svetu" (2015) poudarja [9], da uspešna podjetja presegajo analizo podatkov in aktivno uporabljajo orodja za interaktivno vizualizacijo, kot so grafikoni, infografike in analitične nadzorne plošče za podporo odločanju. V skladu s poročilom - vizualizacija podatkov pomaga strankam razumeti zgodbo, ki jo pripovedujejo podatki, s pomočjo grafov, diagramov, nadzornih plošč in interaktivnih podatkovnih modelov.

Proces pretvorbe informacij v vizualne grafične oblike, kot so preglednice, grafi in diagrami, izboljša razumevanje in razlago podatkov v človeških možganih (slika 7.1-4). Tako lahko vodje projektov in analitiki hitreje ocenijo zapletene scenarije in sprejemajo utemeljene odločitve na podlagi vizualno prepoznavnih trendov in vzorcev namesto intuicije.



**Slika 7.1-4 Različne vrste vizualizacije so zasnovane tako, da človeškim možganom pomagajo bolje razumeti in osmislieti suhoperne informacije številk.**

Izdelava vizualizacij iz podatkov in uporaba različnih brezplačnih knjižnic za vizualizacijo bosta podrobneje obravnavani v naslednjem poglavju o postopkih ETL.

Vizualizacija postaja sestavni del dela s podatki v gradbeništву - pomaga ne le "videti" podatke, temveč tudi razumeti njihov pomen v okviru nalog upravljanja. Vendar pa je za resnično uporabno vizualizacijo treba vnaprej določiti, kaj natančno je treba vizualizirati in katere metrike so resnično pomembne za ocenjevanje uspešnosti projekta. Tu pridejo v poštev metrike uspešnosti, kot so ključni kazalniki uspešnosti in donosnost naložbe. Brez njih tudi najlepše nadzorne plošče tvegajo, da bodo le "informacijski šum".

## KPI in ROI

V današnji gradbeni industriji imajo upravljanje kazalnikov uspešnosti (KPI in ROI) ter njihova vizualizacija prek poročil in nadzornih plošč ključno vlogo pri izboljšanju produktivnosti in učinkovitosti upravljanja projektov.

Kot v vsakem poslu je tudi v gradbeništvu treba jasno opredeliti kazalnike, s katerimi se merijo uspeh, donosnost naložb in uspešnost. Pri pridobivanju podatkov o različnih procesih se mora podatkovno usmerjena organizacija najprej naučiti opredeliti **ključne KPI (Key Performance Indicators)** - kvantitativna merila, ki odražajo, v kolikšni meri se dosegajo strateški in operativni cilji.

Za izračun KPI se običajno uporablja formula (slika 7.1-5), ki vključuje dejanske in načrtovane kazalnike. Na primer, za izračun posameznega KPI za projekt, zaposlenega ali proces delimo dejansko uspešnost z načrtovano uspešnostjo in rezultat pomnožimo s 100%.

$$\text{index KPIs} = \frac{\text{actual performance}}{\text{target performance}} \times 100$$

Slika 7.1-5 KPI se uporabljajo za merjenje uspešnosti projekta ali procesa pri doseganju ključnih ciljev.

Na ravni lokacije se lahko uporabijo podrobnejši ključni kazalniki uspešnosti metrike:

- **Časovni razpored ključnih mejnikov** (temeljenje, vgradnja, zaključna dela) - omogoča nadzor skladnosti z delovnimi načrti.
- **Odstotek prekoračitev materiala** - pomaga pri upravljanju nabave in zmanjševanju izgub.
- **Število nenačrtovanih izpadov strojev** - vpliva na produktivnost in stroške.

Izbira napačnih metrik lahko privede do napačnih odločitev "kaj storiti?" (slika 2.2-5). Če se na primer podjetje osredotoča le na stroške na kvadratni meter, ne upošteva pa stroškov predelav, lahko prihranki pri materialih privedejo do slabše kakovosti in višjih stroškov pri prihodnjih projektih.

Pri določanju ciljev je treba jasno opredeliti, kaj se meri. Nejasna formulacija vodi do napačnih sklepov in otežuje nadzor. Oglejmo si primere uspešnih in neuspešnih KPI v gradbeništvu.

#### Dobri ključni kazalniki uspešnosti:

- ❑ "Do konca leta zmanjšajte odstotek obnovitvenih del za 10 odstotkov."
- ❑ "Do naslednjega četrtletja povečajte hitrost vgradnje fasad za 15 odstotkov brez poslabšanja kakovosti"
- ❑ "Z optimizacijo delovnih urnikov do konca leta za 20% zmanjšajte izpade strojev."

Te metrike so jasno merljive, imajo določene vrednosti in časovne okvire.

#### **Slabi ključni kazalniki uspešnosti:**

- ☒ "Gradili bomo hitreje" (Koliko hitreje? Kaj pomeni "hitreje"?).
- ☒ "Izboljšali bomo kakovost betonskega dela" (Kako natančno se meri kakovost?)
- ☒ "Izboljšali bomo interakcijo med izvajalci na gradbišču" (Katera merila bodo pokazala izboljšanje?)

Dober KPI je tisti, ki ga je mogoče meriti in objektivno oceniti. V gradbeništvu je to še posebej pomembno, saj je brez jasnih kazalnikov nemogoče spremljati uspešnost in dosegati stabilne rezultate.

Poleg KPI obstaja dodatna metrika za ocenjevanje učinkovitosti naložb: **ROI (Return on Investment)** - kazalnik donosnosti naložb, ki odraža razmerje med dobičkom in vloženimi sredstvi. ROI pomaga oceniti, ali je uvedba novih metod, tehnologij ali orodij upravičena: od digitalnih rešitev in avtomatizacije (npr. slika 7.3-2) do uporabe novih gradbenih materialov. Ta kazalnik pomaga pri sprejemanju premišljenih odločitev o nadaljnjih naložbah na podlagi njihovega dejanskega vpliva na dobičkonosnost podjetja

V okviru vodenja gradbenih projektov se lahko ROI (donosnost naložbe) uporablja kot eden od ključnih kazalnikov uspešnosti (KPI), če je cilj podjetja izmeriti donosnost naložbe v projekt, tehnologijo ali izboljšanje procesa. Na primer, če se izvaja nova tehnika upravljanja gradnje, lahko ROI pokaže, koliko je izboljšala dobičkonosnost.

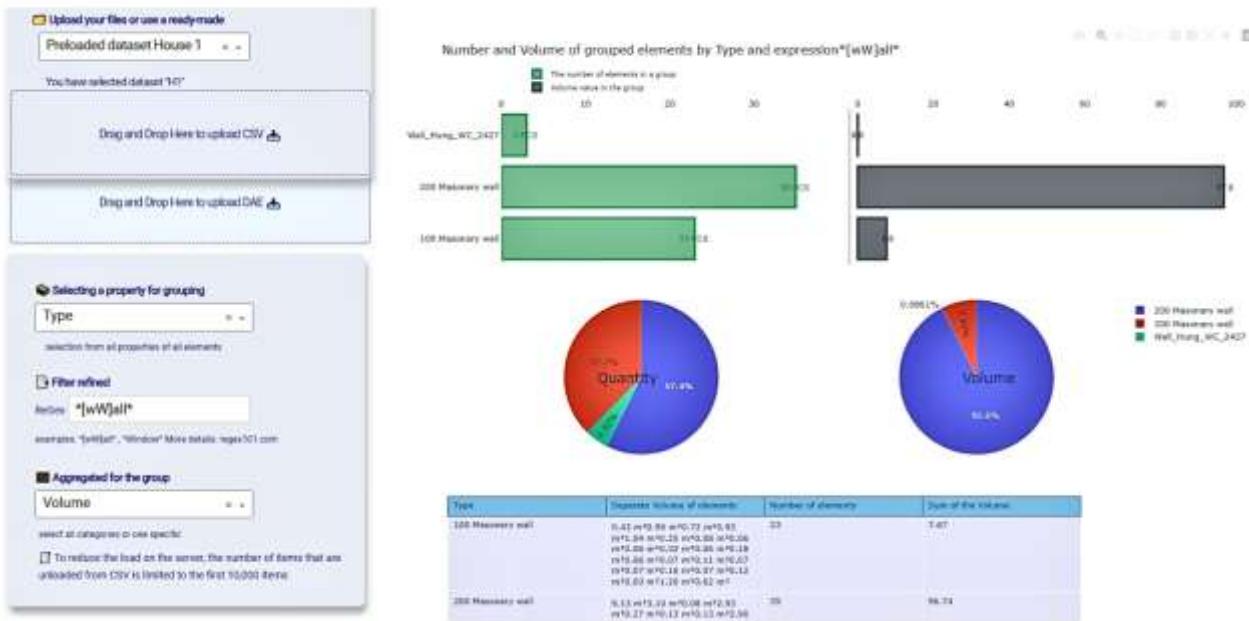
Redno merjenje ključnih kazalnikov uspešnosti in donosnosti naložb na podlagi podatkov, zbranih iz različnih virov, kot so poraba materiala, delovne ure in stroški, omogoča vodstvu projekta učinkovito upravljanje virov in sprejemanje hitrih odločitev. Dolgoročno shranjevanje teh podatkov omogoča analizo prihodnjih trendov in optimizacijo procesov.

Za vizualizacijo ključnih kazalnikov uspešnosti, donosnosti naložb in drugih kazalnikov se uporabljajo različni diagrami in grafikoni, ki so običajno združeni v nadzorne plošče.

#### **Nadzorne plošče in nadzorne plošče: vizualizacija metrik za učinkovito upravljanje**

Za vizualizacijo kazalnikov in kazalnikov se uporabljajo različni diagrami in grafi, ki so običajno združeni v podatkovne vitrine in nadzorne plošče. Te nadzorne plošče zagotavljajo centraliziran pregled nad stanjem projekta ali delov projekta in prikazujejo ključne kazalnike (po možnosti v realnem času). Ažurne in stalno posodobljene nadzorne plošče omogočajo ekipi, da se hitro odzove na spremembe.

**Nadzorne plošče** so orodja za vizualizacijo kvantitativnih ocen, ki so lahko dostopne in razumljive vsem udeležencem projekta.



Slika 7.1-6 Upravljanje ključnih kazalnikov uspešnosti in njihova vizualizacija s pomočjo nadzornih plošč je ključnega pomena za izboljšanje produktivnosti in učinkovitosti projekta.

Tukaj je nekaj primerov priljubljenih orodij, v katerih lahko ustvarite nadzorne plošče:

- **Power BI** je Microsoftovo orodje za ustvarjanje interaktivnih poročil in nadzornih plošč.
- **Tableau in Google Data Studio** sta zmogljivi orodji za vizualizacijo podatkov in ustvarjanje nadzornih plošč, ne da bi morali pisati kodo.
- **Plotly** (Slika 7.1-6, Slika 7.2-12) je knjižnica za ustvarjanje interaktivnih grafov, Dash pa je ogrodje za ustvarjanje spletnih aplikacij za analizo podatkov. V kombinaciji ju lahko uporabite za ustvarjanje interaktivnih nadzornih plošč.
- **Številne knjižnice za Python** (Slika 7.2-9 - Slika 7.2-11) - Python ima številne odprtakodne in brezplačne knjižnice za vizualizacijo podatkov, kot so Matplotlib, Seaborn, Plotly, Bokeh in druge. Te je mogoče uporabiti za ustvarjanje grafov in jih vključiti v spletno aplikacijo z uporabo ogrodij, kot sta Flask ali Django.
- **Knjižnice JavaScript**: omogoča ustvarjanje interaktivnih nadzornih plošč z uporabo odprtakodnih knjižnic JavaScript, kot sta D3.js ali Chart.js, in njihovo vključitev v spletno aplikacijo.

Za ocenjevanje ključnih kazalnikov uspešnosti in izdelavo nadzornih plošč potrebuje potrebujete ažurne podatke ter jasen časovni razpored zbiranja in analiziranja informacij.

Na splošno so KPI, ROI in nadzorne plošče v gradbeništvu podlaga za analitični pristop k upravljanju projektov. Ne pomagajo le pri spremeljanju in ocenjevanju trenutnega stanja, temveč zagotavljajo tudi dragocene vpoglede za prihodnje procese načrtovanja in optimizacije - procese, ki so neposredno odvisni od interpretacije podatkov ter postavljanja pravih in pravočasnih vprašanj.

## Analiza podatkov in umetnost postavljanja vprašanj

Interpretacija podatkov je zadnja faza analize, ko informacije dobijo smisel in začnejo "govoriti". Tu se oblikujejo odgovori na ključna vprašanja: "Kaj storiti?" in "Kako storiti?" (slika 2.2-5). Ta faza omogoča povzemanje rezultatov, prepoznavanje vzorcev, ugotavljanje vzročno-posledičnih povezav ter oblikovanje zaključkov na podlagi vizualizacije in statistične analize.

*Morda ni daleč čas, ko bo spoznala, da je za polnopravno državljanstvo v eni od velikih novih kompleksnih svetovnih držav, ki se zdaj razvijajo, treba znati računati, razmišljati v smislu povprečja, maksimuma in minimuma, kot je zdaj treba znati brati in pisati [133].*

- Samuel S. Wilkes, citirano v predsedniškem nagovoru Ameriškemu statističnemu združenju leta 1951

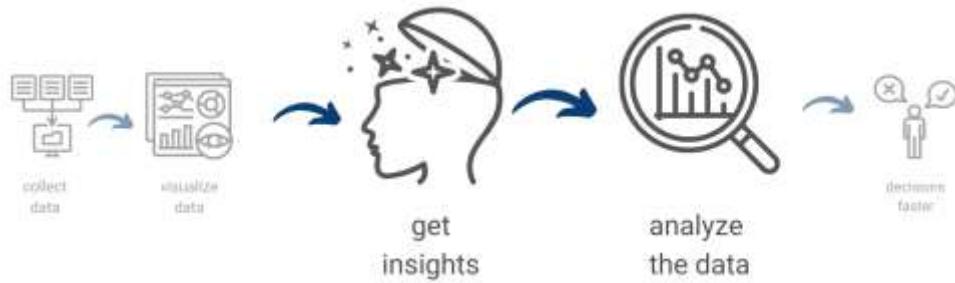
V skladu s poročilom "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024), ki ga je objavila vlada Združenega kraljestva [83], lahko izvajanje analitike podatkov in umetne inteligenčne (AI) bistveno izboljša procese upravljanja projektov, poveča natančnost napovedovanja časa in stroškov ter zmanjša tveganje in negotovost. Dokument poudarja, da javne organizacije, ki uporabljajo napredna analitična orodja, dosegajo večjo uspešnost pri infrastrukturnih pobudah.

Sodobno gradbeno dejavnost, ki deluje v zelo konkurenčnem okolju četrte industrijske revolucije z nizkimi maržami, lahko primerjamo z vojaškimi operacijami. Tu sta preživetje in uspeh podjetja odvisna od hitrosti pridobivanja virov in kakovostnih informacij - ter s tem od pravočasnega in informiranega odločanja (slika 7.1-7).

Če je vizualizacija podatkov "obveščevalna informacija", ki omogoča pregled, je podatkovna analitika "strelivo", potrebno za ukrepanje. Odgovarja na vprašanja: *kaj storiti* in *kako to storiti*, kar je osnova za pridobitev konkurenčne prednosti na trgu.

Analitika spreminja različne podatke v strukturirane in smiselne informacije, na katerih temeljijo odločitve.

Naloga analitikov in vodij ni le razlaganje informacij, temveč tudi sprejemanje utemeljenih odločitev, prepoznavanje trendov, določanje razmerij med različnimi vrstami podatkov in njihovo razvrščanje v skladu s cilji in posebnostmi projekta. Z uporabo orodij za vizualizacijo in metod statistične analize podatke spremenijo v strateško premoženje podjetja.



**Slika 7.1-7 Analiza podatkov je tista, ki na koncu zbrane informacije spremeni v vir za sprejemanje odločitev.**

Za sprejemanje resnično utemeljenih odločitev v procesu analitike, se je treba naučiti, kako pravilno oblikovati vprašanja, ki se zastavlajo podatkom. Kakovost teh vprašanj neposredno vpliva na globino pridobljenih vpogledov in posledično na kakovost upravljaških odločitev.

*Preteklost obstaja le, če je prisotna v današnjih zapisih. Kaj ti zapisi predstavljajo, pa je odvisno od vprašanj, ki si jih zastavljamo. Druge zgodovine kot te ni [134].*

- John Archibald Wheeler, fizik 1982

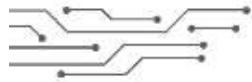
Umetnost postavljanja poglobljenih vprašanj in kritičnega razmišljanja je bistvena spremnost pri delu s podatki. Večina ljudi običajno postavlja preprosta, površinska vprašanja, na katera ni treba odgovoriti z veliko truda. Vendar se prava analiza začne s smiselnimi in premišljenimi vprašanji, ki lahko v podatkih odkrijejo skrite odnose ter vzročno-posledične povezave, ki so morda skrite za več plastmi razmišljanja.

V skladu s študijo "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135] so za uspešno digitalno preobrazbo potrebne naložbe v analitične zmogljivosti, programi za upravljanje sprememb in uskladitev poslovnih ciljev s pobudami IT. Podjetja, ki ustvarjajo podatkovno usmerjeno kulturo, morajo vlagati v zmogljivosti za analitiko podatkov in uvajati programe za upravljanje sprememb, da bi uveljavila novo razmišljanje, vedenje in načine dela.

Brez naložb v razvoj analitične kulture, izboljšanje podatkovnih orodij in usposabljanje strokovnjakov bodo podjetja še naprej tvegala sprejemanje odločitev na podlagi zastarelih ali nepopolnih informacij - ali pa se bodo zanašala na subjektivna mnenja vodij HiPPO (slika 2.1-9).

Zavedanje pomembnosti in potrebe po nenehnem posodabljanju analitičnih in nadzornih plošč neizogibno pripelje vodstvo do razumevanja pomena avtomatizacije analitičnih procesov. Avtomatizacija poveča hitrost odločanja, zmanjša vpliv človeškega dejavnika in zagotovi ustreznost podatkov. Z eksponentno rastjo obsega informacij hitrost ne postaja le konkurenčna prednost, temveč ključni dejavnik trajnostnega uspeha.

Avtomatizacija postopkov analize in obdelave podatkov na splošno je neločljivo povezana s temo ETL (Extract, Transform, Load). Tako kot moramo v procesu avtomatizacije preoblikovati podatke, se v procesu ETL podatki pridobivajo iz različnih virov, preoblikujejo v skladu s potrebnimi zahtevami in nalagajo v ciljne sisteme za nadaljnjo uporabo.



## POGLAVJE 7.2.

### PRETOK PODATKOV BREZ ROČNEGA DELA: ZAKAJ JE POTREBEN ETL

#### Avtomatizacija ETL: zmanjšanje stroškov in pospešitev obdelave podatkov

Ko ključni kazalniki uspešnosti (KPI) prenehajo rasti kljub povečanju količine podatkov in velikosti ekipe, se vodstvo podjetja neizogibno zave, da je treba avtomatizirati procese. To spoznanje prej ali slej postane spodbuda za uvedbo kompleksne avtomatizacije, katere glavni cilj je zmanjšati zapletenost procesov, pospešiti obdelavo in zmanjšati odvisnost od človeškega dejavnika.

V skladu z McKinseyjevo študijo "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136] imajo podjetja, ki uporabljajo arhitekture pretočnih podatkov, veliko prednost, saj lahko informacije analizirajo v realnem času. Tehnologije pretočnih podatkov omogočajo neposredno analizo sporočil v realnem času in uporabo napovednega vzdrževanja v proizvodnji z analizo podatkov senzorjev v realnem času.

Poenostavitev procesov je avtomatizacija, pri kateri tradicionalne ročne funkcije nadomestijo algoritmi in sistemi.

Vprašanje avtomatizacije oziroma "zmanjševanja vloge ljudi pri obdelavi podatkov" je za vsako podjetje nepovraten in zelo občutljiv proces. Strokovnjaki na katerem koli strokovnem področju se pogosto obotavljam, da bi svoje metode in subtilnosti dela v celoti razkrili kolegom optimistom, saj se zavedajo tveganja, da bodo v hitro razvijajočem se tehnološkem okolju izgubili službo.

*Če si želite pridobiti sovražnike, poskušajte spremeniti stvari [137].*

- Woodrow Wilson, govor na kongresu prodajalcev, Detroit, 1916

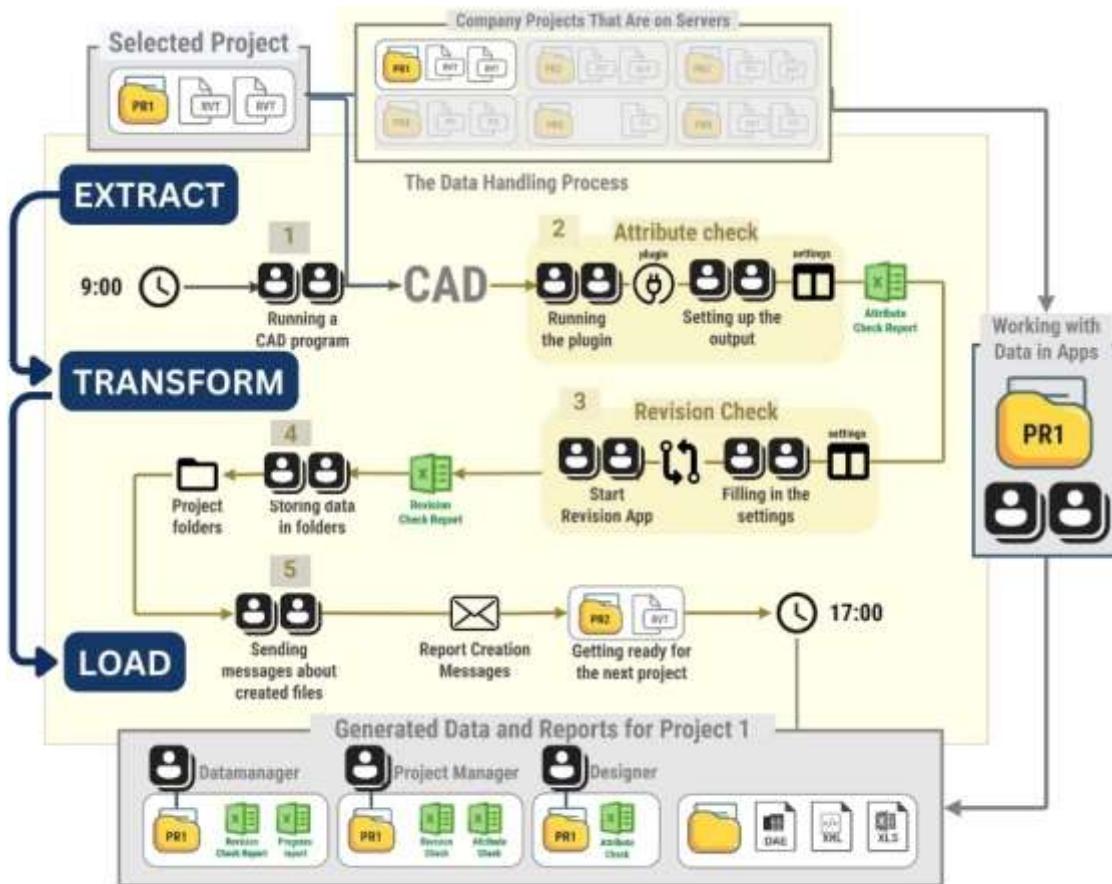
Kljud očitnim prednostim avtomatizacije številna podjetja v svojih vsakodnevnih praksah še vedno uporabljajo velik delež ročnega dela, zlasti na področju inženirskeih podatkov. Za ponazoritev trenutnega stanja si oglejmo tipičen primer zaporedne obdelave podatkov znotraj takšnih procesov.

Ročno obdelavo podatkov lahko ponazorimo s primerom interakcije z informacijami, pridobljenimi iz podatkovnih zbirk CAD. Tradicionalna obdelava podatkov ("ročni" ETL -proces) v oddelkih CAD (BIM) za izdelavo atributnih tabel ali izdelavo dokumentacije na podlagi projektnih podatkov poteka v naslednjem vrstnem redu (slika 7.2-1):

1. Ročno **pridobivanje (Extract)**: uporabnik ročno odpre projekt - z zagonom aplikacije CAD (BIM) (slika 7.2-1 korak 1).
2. **Preverjanje**: naslednji korak običajno vključuje ročno izvajanje več vtičnikov ali pomožnih aplikacij za pripravo podatkov in oceno njihove kakovosti (slika 7.2-1, korak 2-3).
3. Ročno **preoblikovanje (Transform)**: po pripravi se začne obdelava podatkov, ki zahteva ročno

upravljanje različnih programskih orodij, v katerih se podatki pripravijo za prenos (slika 7.2-1 korak 4).

4. Ročno **nalaganje (Load)**: ročno nalaganje pretvorjenih podatkov v zunanje sisteme, podatkovne formate in dokumente (slika 7.2-1 korak 5).



Slika 7.2-1 Tradicionalna ročna obdelava ETL je omejena z željami in fizičnimi zmogljivostmi posameznega tehnika.

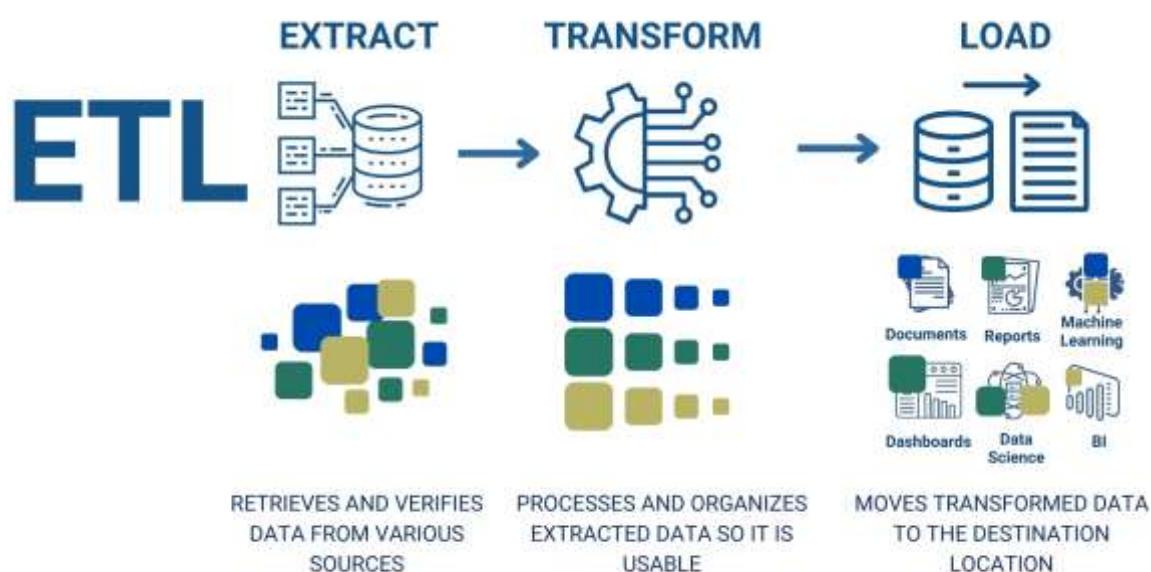
Takšen potek dela je primer klasičnega postopka ETL - ekstrakcija, transformacija in nalaganje (ETL). V nasprotju z drugimi panogami, kjer so samodejni cevovodi ETL že dolgo standard, v gradbeništvu še vedno prevladuje ročno delo, ki upočasnuje procese in povečuje stroške

**ETL (Extract, Transform, Load)** je postopek pridobivanja podatkov iz različnih virov, njihovega preoblikovanja v želeno obliko in nalaganja v ciljni sistem za nadaljnjo analizo in uporabo.

**ETL** je postopek, ki označuje tri ključne sestavine obdelave podatkov: Izvleček, preoblikovanje in nalaganje (slika 7.2-2):

- Izvleček - izvleček podatkov iz različnih virov (datotek, podatkovnih zbirk, API).
- **Preoblikovanje** - čiščenje, združevanje, normalizacija in logična obdelava podatkov.
- **Load** - nalaganje strukturiranih informacij v podatkovno skladišče, poročilo ali sistem BI.

Na začetku knjige smo se koncepta ETL dotknili le občasno: pri pretvorbi nestrukturiranega skeniranega dokumenta v strukturirano tabelarno obliko (slika 4.1-1), v okviru formalizacije zahtev za sistematizacijo dojemanja življenjskih in poslovnih procesov (slika 4.4-20) ter pri avtomatizaciji potrjevanja in obdelave podatkov iz rešitev CAD. Zdaj si podrobneje oglejmo ETL v kontekstu tipičnih delovnih procesov.



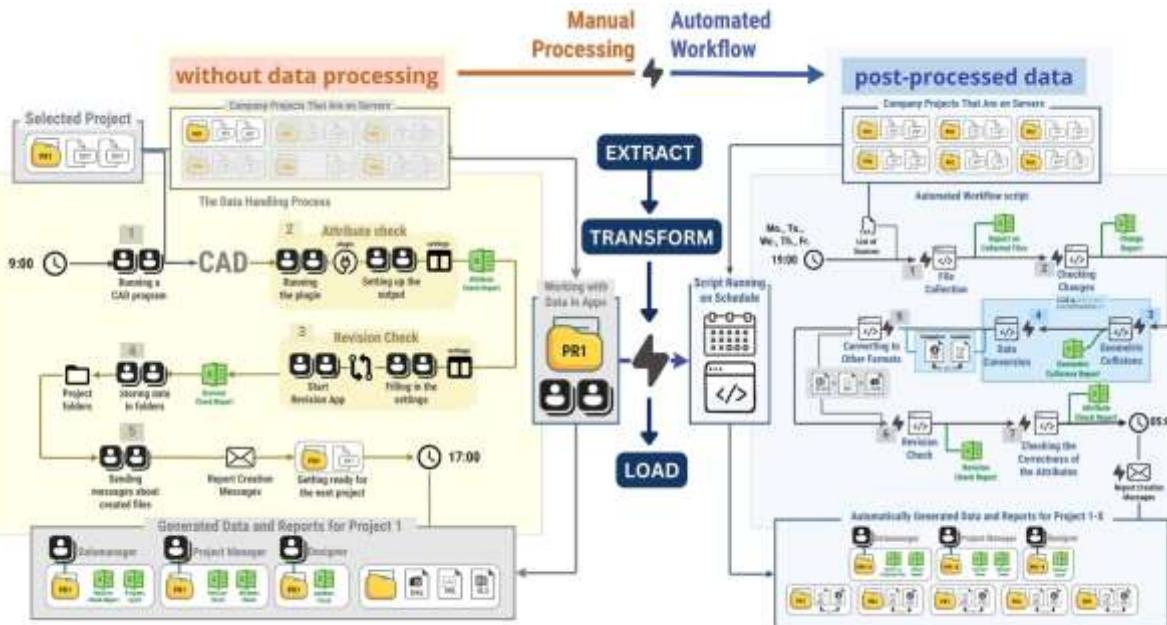
Slika 7.2-2 ETL avtomatizira ponavljajoča se opravila obdelave podatkov.

Ročni ali polavtomatski postopek ETL - proces vključuje vodjo ali tehnika, ki ročno upravlja vse korake - od zbiranja podatkov do izdelave poročila. Takšen postopek zahteva precej časa, zlasti če je delovni čas omejen (npr. od 9.00 do 17.00).

Podjetja pogosto poskušajo rešiti problem nizke učinkovitosti in počasne hitrosti z nakupom modularnih integriranih rešitev (ERP, PMIS, CPM, CAFM itd.), ki jih nato dodatno razvijajo zunanji prodajalci in svetovalci. Toda ti prodajalci in zunanji razvijalci pogosto postanejo kritična točka odvisnosti: njihove tehnične omejitve neposredno vplivajo na učinkovitost celotnega sistema in poslovanja kot celote, kot je podrobno opisano v prejšnjih poglavjih o lastniških sistemih in formatih. Težave, ki jih povzročata razdrobljenost in odvisnost, so bile podrobno obravnavane v poglavju "Kako se gradbena podjetja utapljamajo v podatkovnem kaosu".

Če podjetje ni pripravljeno uvesti velike modularne platforme enega od ponudnikov, začne iskati druge načine avtomatizacije. Eden od njih je razvoj lastnega modularnega odprtga ETL -prevoznikov, kjer se vsaka faza (ekstrakcija, transformacija, validacija, nalaganje) izvaja kot skripte, ki se izvajajo po urniku.

V avtomatizirani različici istega postopa dela ETL (slika 7.2-1) je potek dela videti kot modularna koda, ki se začne z obdelavo podatkov in njihovim prevajanjem v odprto strukturirano obliko. Ko so strukturirani podatki prejeti, se samodejno, po urniku, zaženejo različne skripte ali moduli za preverjanje sprememb, preoblikovanje in pošiljanje sporočil (slika 7.2-3).



Slika 7.2-3 Na levi strani je ročna obdelava, na desni pa samodejni postopek, ki za razliko od tradicionalne ročne obdelave ni omejen z zmožnostmi uporabnika.

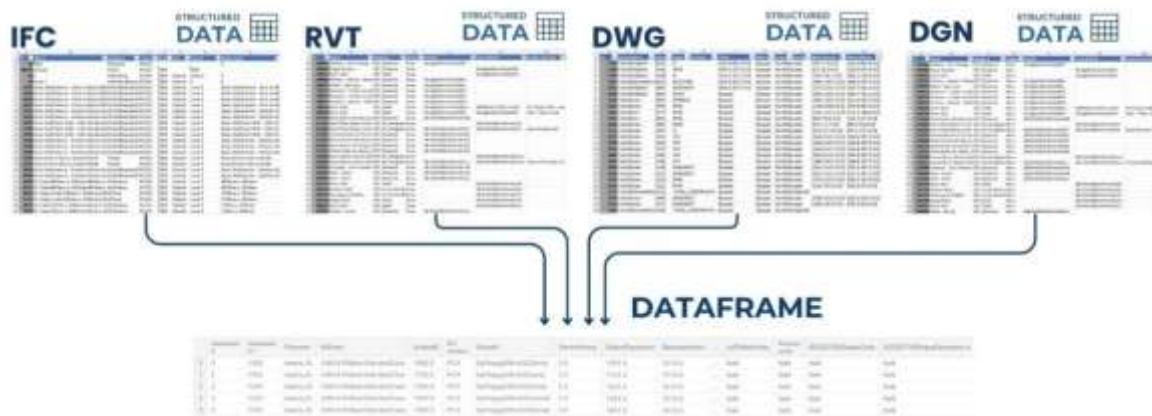
V avtomatiziranem delovnem postopku je obdelava podatkov poenostavljena s predobdelavo podatkov ET(L): strukturiranjem in poenotenjem.

Pri tradicionalnih metodah obdelave strokovnjaki delajo s podatki, "kakršni so" - kot so pridobljeni iz sistemov ali programske opreme. Pri avtomatiziranih procesih pa podatki pogosto najprej preidejo skozi postopek ETL -payplane, kjer se uskladijo v dosledno strukturo in obliko, primerno za nadaljnjo uporabo in analizo.

Vzemimo praktični primer ETL, ki prikazuje postopek potrjevanja podatkovne tabele, opisan v poglavju "Potrjevanje podatkov in rezultati potrjevanja" (slika 4.4-13). V ta namen uporabimo knjižnico Pandas v povezavi z LLM za avtomatizirane postopke analize in obdelave podatkov.

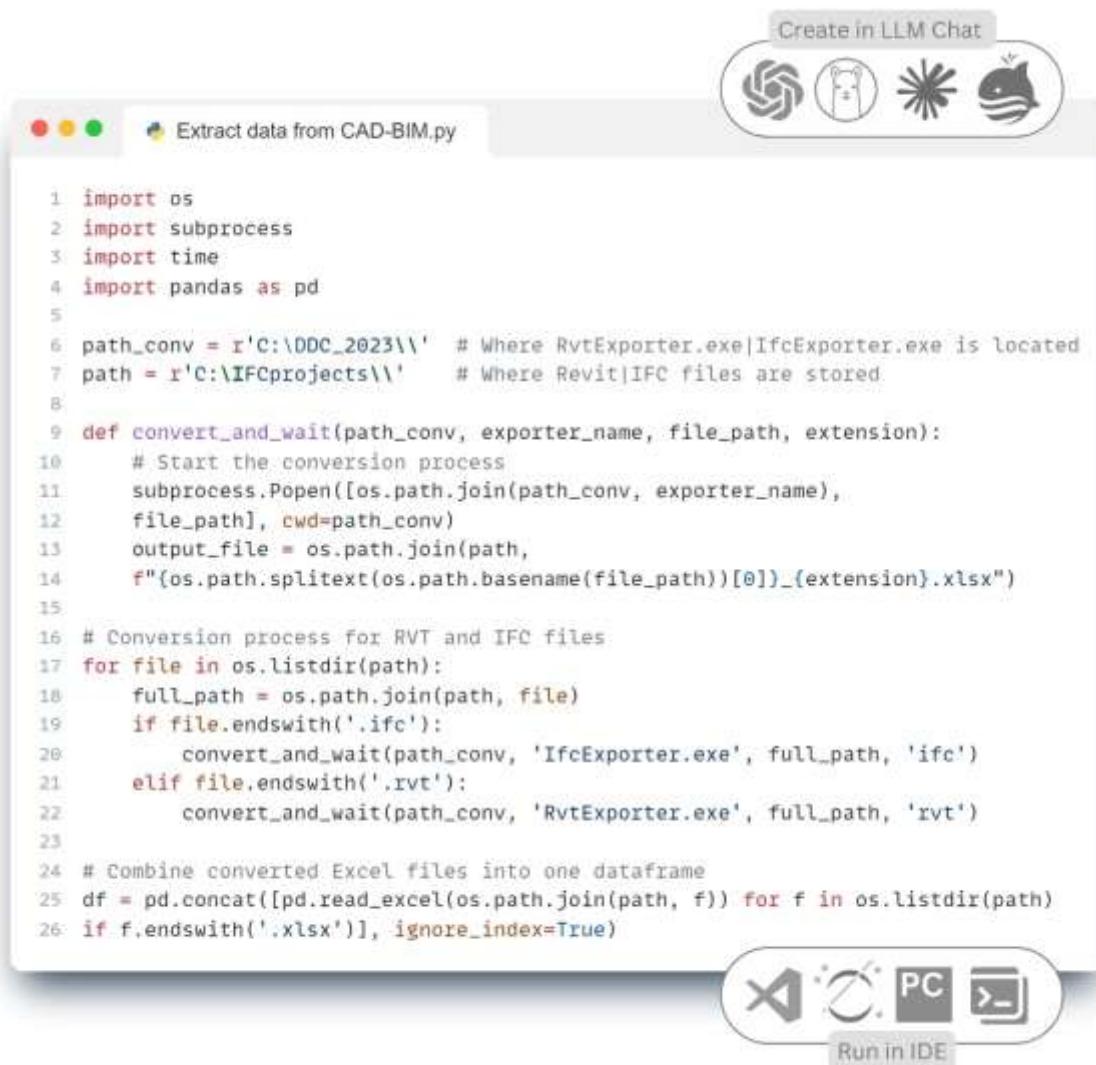
## ETL Izvleček: zbiranje podatkov

Prva faza postopka ETL - Izvleček) - se začne s pisanjem kode za zbiranje podatkovnih nizov, ki jih je treba nadalje preveriti in obdelati. V ta namen bomo pregledali vse mape produkcijskega strežnika, zbrali dokumente določene oblike in vsebine ter jih nato pretvorili v strukturirano obliko. Ta postopek je podrobno obravnavan v poglavjih "Pretvarjanje nestrukturiranih in besedilnih podatkov v strukturirano obliko" in "Pretvarjanje podatkov CAD (BIM) v strukturirano obliko" (Slika 4.1-1 - Slika 4.1-12).



Slika 7.2-4 Pretvorite podatke CAD (BIM) v en velik podatkovni okvir, ki bo vseboval vse dele projekta.

Kot nazoren primer uporabimo korak za nalaganje podatkov Extract in pridobimo preglednico vseh projektov CAD- (BIM-) (slika 7.2-4), ki uporablja pretvornike z možnostjo povratnega inženiringa [138] formate RVT in IFC, da pridobi strukturirane preglednice iz vseh projektov in jih združi v eno veliko preglednico DataFrame.



```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\' # Where RvtExporter.exe|IfcExporter.exe is located
7 path = r'C:\IFCprojects\' # Where Revit|IFC files are stored
8
9 def convert_and_wait(path_conv, exporter_name, file_path, extension):
10    # Start the conversion process
11    subprocess.Popen([os.path.join(path_conv, exporter_name),
12                      file_path], cwd=path_conv)
13    output_file = os.path.join(path,
14                               f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")
15
16    # Conversion process for RVT and IFC files
17    for file in os.listdir(path):
18        full_path = os.path.join(path, file)
19        if file.endswith('.ifc'):
20            convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
21        elif file.endswith('.rvt'):
22            convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')
23
24    # Combine converted Excel files into one dataframe
25 df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
26 if f.endswith('.xlsx')], ignore_index=True)

```

Run in IDE

**Slika 7.2-5 Pretvarjanje z uporabo kode Python in orodja SDK za obratno inženirstvo za datoteke RVT in IFC v en velik strukturiran (df) DataFrame.**

Pandas DataFrame lahko naloži podatke iz različnih virov, vključno z besedilnimi datotekami CSV, preglednicami Excel, datotekami JSON - in XML -, datotekami za shranjevanje velikih količin podatkov, kot sta Parquet in HDF5, ter iz zbirk podatkov MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle in drugih zbirk podatkov. Poleg tega Pandas podpira nalaganje podatkov iz API, spletnih strani, storitev v oblaku in sistemov za shranjevanje, kot so Google BigQuery, Amazon Redshift in Snowflake.

- ☞ Če želite napisati kodo za povezovanje in zbiranje informacij iz podatkovnih zbirk, pošljite podobno besedilno zahtevo v klepet LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ali katero koli drugo):

Prosimo, napišite primer povezovanja s strežnikom MySQL in pretvorbe podatkov v .



Odgovor LLM :

```

1 import pandas as pd
2 import mysql.connector
3
4 # Using a context manager to ensure
5 # that the connection is closed automatically
6 with mysql.connector.connect(host='hostname', database='dbname',
7 user='username', password='password') as conn:
8     df = pd.read_sql('SELECT * FROM tablename LIMIT 10', conn)

```

Slika 7.2-6 Primer povezovanja prek Pythona s podatkovno zbirko MySQL in uvoza podatkov iz podatkovne zbirke MySQL v podatkovni okvir.

Dobljena koda (Slika 7.2-5, Slika 7.2-6) lahko zaženete v enem od priljubljenih IDE (integriranih razvojnih okolij), ki smo jih omenili zgoraj, v načinu brez povezave: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z vtičnikom PyDev, Thonny, Wing IDE, IntelliJ IDEA z vtičnikom Python, JupyterLab ali priljubljena spletna orodja: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Z nalaganjem večpredstavnih podatkov v spremenljivko "df" (slika 7.2-5 - vrstica 25; slika 7.2-6 - vrstica 8) smo podatke pretvorili v format Pandas DataFrame, eno najbolj priljubljenih struktur za obdelavo podatkov, ki je dvodimenzionalna tabela z vrsticami in stolpcji. O drugih formatih shranjevanja, ki se uporabljajo v ETL -plinijah, kot so Parquet, Apache ORC, JSON, Feather, HDF5, ter sodobnih podatkovnih skladiščih bomo več povedali v poglavju "Shranjevanje in upravljanje podatkov v gradbeništvu" (slika 8.1-2).

Po fazi pridobivanja in strukturiranja podatkov (Izvleček) nastane enotno polje informacij (slika 7.2-5, slika 7.2-6), ki je pripravljeno za nadaljnjo obdelavo. Vendar je treba pred nalaganjem teh podatkov v ciljne sisteme ali njihovo uporabo za analizo zagotoviti njihovo kakovost, celovitost in skladnost z določenimi zahtevami. Na tej stopnji se izvede preoblikovanje podatkov (Transform) - ključni korak za zagotovitev zanesljivosti poznejših sklepov in odločitev.

## ETL Preoblikovanje: uporaba pravil potrjevanja in preoblikovanja

V koraku Transform se podatki obdelajo in preoblikujejo. Ta postopek lahko vključuje preverjanje pravilnosti, normalizacijo, dopolnjevanje manjkajočih vrednosti in potrjevanje z uporabo samodejnih orodij

Glede na študijo družbe PwC "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] se sodobna revizijska podjetja oddaljujejo od naključnega preverjanja podatkov in se usmerjajo v analiziranje velikih količin informacij z uporabo avtomatiziranih orodij. Ta pristop ne omogoča le ugotavljanja neskladij v poročanju, temveč tudi ponujanje priporočil za optimizacijo poslovnih procesov.

V gradbeništvu se lahko podobne metode uporabljajo na primer za samodejno potrjevanje projektnih podatkov, nadzor kakovosti gradnje in ocenjevanje uspešnosti izvajalcev. Eno od orodij za avtomatizacijo in pospešitev obdelave podatkov je uporaba regularnih izrazov (RegEx) v fazi preoblikovanja podatkov (Transform) procesa ETL. RegEx omogoča učinkovito preverjanje podatkovnih nizov, odkrivanje nedoslednosti in zagotavljanje celovitosti informacij z minimalno uporabo virov. Več podrobnosti o RegEx (slika 4.4-7) je bilo obravnavanih v poglavju "Prevajanje zahtev v strukturirano obliko".

Poglejmo praktični primer: v sistemu za upravljanje nepremičinskih objektov (RPM) upravljavec določi zahteve za ključne atribute objektov (slika 7.2-7). V fazi preoblikovanja je treba potrditi naslednje parametre:

- preverjanje oblik identifikatorjev objektov (atribut "ID")
- nadzor nad vrednostmi nadomestnih garancijskih rokov (atribut "Garancijski rok").
- preverjanje cikla zamenjave elementa (atribut "Zahteve za vzdrževanje").



ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Maintenance

Slika 7.2-7 Preverjanje kakovosti se začne z določanjem zahtev za atribute in njihove mejne vrednosti.

Za določitev mejnih vrednosti za potrjevanje parametrov na primer predpostavimo, da iz naših izkušenj vemo, da lahko veljavne vrednosti za atribut "ID" vključujejo samo vrednosti nizov "W-NEW", "W-OLD1"

ali "D-122" ali podobne vrednosti, kjer je prvi znak črka, ki ji sledi pomicljaj in nato trije abecedni znaki "NEW", "OLD" ali katero koli trimestno število (slika 7.2-7). Za preverjanje teh identifikatorjev se lahko uporabi naslednji regularni izraz (RegEx):

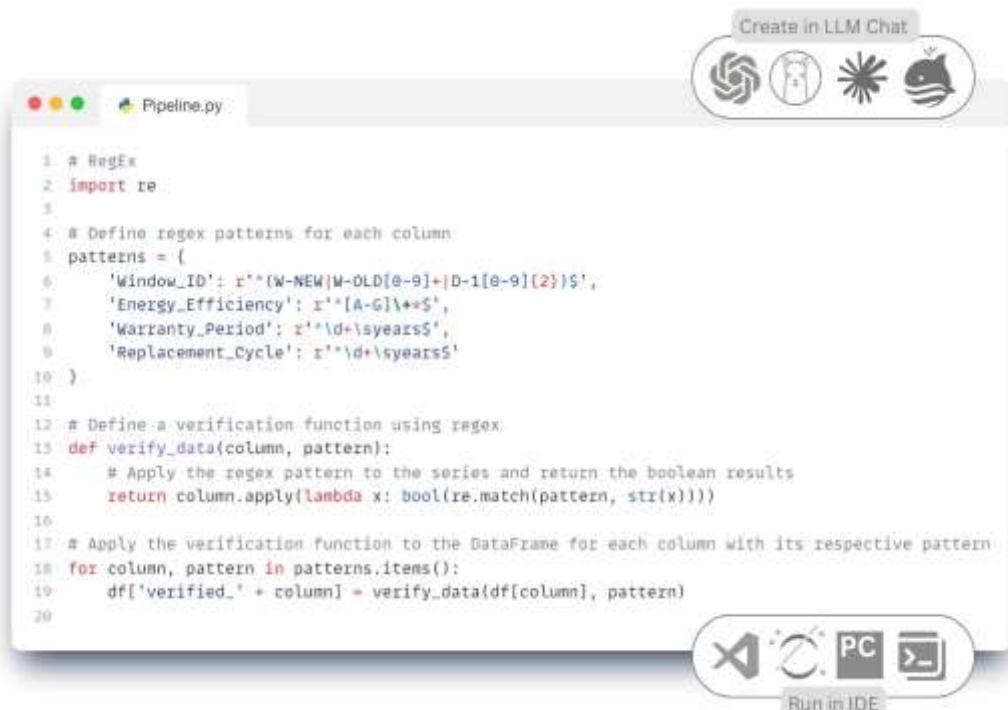
```
^W-NEW$ | ^W-OLD[0-9]+$ | ^D-1[0-9]{2}$
```

Ta predloga poskrbi, da se vsi identifikatorji v podatkih ujemajo z določenimi merili. Če katera koli vrednost ne prestane preverjanja, sistem zabeleži napako. Če želite ustvariti Python -kodo za preoblikovanje podatkov in uporabiti dobljene podatke za ustvarjanje preglednice z rezultati, preprosto oblikujte poizvedbo v klepetu LLM.

💡 Besedilo zahtevka za LLM:

Napišite kodo za preverjanje stolpcev podatkovnega okvira z uporabo regularnih izrazov, ki preverja identifikatorje v obliku "W-NEW" ali "W-OLD" s pomočjo RegEx, energetsko učinkovitost s črkami od "A" do "G", garancijsko dobo in cikel zamenjave s številčnimi vrednostmi v letih ↴

💡 Odgovor LLM :



The screenshot shows a code editor window titled "Pipeline.py". The code is as follows:

```
1 # Regex
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r"^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$",
7     'Energy_Efficiency': r"^[A-G]\w*$",
8     'Warranty_Period': r"\d+\w{3}\d{2}\w{3}\d{2}$",
9     'Replacement_Cycle': r"\d+\w{3}\d{2}\w{3}\d{2}$"
10 }
11
12 # Define a verification function using regex
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df['verified_' + column] = verify_data(df[column], pattern)
20
```

The interface includes a toolbar with icons for Create in LLM Chat, Run in IDE, and other development tools.

Slika 7.2-8 Koda automatizira postopek potrjevanja z uporabo predlog RegEx za stolpce parametrov podatkovnega okvira .

Zgornja samodejno ustvarjena koda v jeziku Python (slika 7.2-8) uporablja knjižnico "re" (regularni izrazi RegEx) za opredelitev funkcije, ki preveri vsak atribut podatkovnega elementa v DataFrame. Za vsak določen stolpec (atribut) funkcija uporabi vzorec RegEx, da preveri, ali je vsak zapis skladen s pričakovano obliko, in rezultate doda kot nove vrednosti (False/True) v nov stolpec z atributom v DataFrame.

Takšna samodejna potrditev zagotavlja formalno skladnost podatkov in se lahko uporablja kot del sistema nadzora kakovosti v fazi preoblikovanja.

Po uspešno opravljenem koraku pretvorbe in preverjanju kakovosti so podatki pripravljeni za prenos v ciljne sisteme. Preoblikovane in potrjene podatke lahko za nadaljnjo uporabo naložite v CSV, JSON, Excel, podatkovne zbirke in druge formate. Odvisno od naloge se lahko rezultati predstavijo tudi v poročilih, grafih ali analitičnih nadzornih ploščah.

## ETL Load: vizualizacija rezultatov v preglednicah in grafih

Po končani fazi Transform, ko so podatki strukturirani in preverjeni, je zadnja faza Load, v kateri je mogoče podatke naložiti v ciljni sistem in jih vizualizirati za analizo. Vizualna predstavitev podatkov omogoča hitro ugotavljanje odstopanj, analizo porazdelitev in sporočanje ključnih ugotovitev vsem udeležencem projekta, tudi tistim brez tehničnega znanja.

Namesto predstavitev informacij v obliki tabel in številk lahko uporabimo infografike, grafe in nadzorne plošče (dashboards). Eno najpogostejših in najbolj prilagodljivih orodij za vizualizacijo strukturiranih podatkov v jeziku Python je knjižnica Matplotlib (slika 7.2-9, slika 7.2-10). Omogoča ustvarjanje statičnih, animiranih in interaktivnih grafov ter podpira širok nabor vrst grafov.

- Za vizualizacijo rezultatov preverjanja atributov iz sistema RPM (slika 7.2-7), lahko uporabite naslednjo poizvedbo za jezikovni model:

Napišite kodo za vizualizacijo podatkov DataFrame, zgoraj (slika 7.2-7), s histogramom za rezultate, ki prikazuje pogostost napak v atributu ↴

- ❸ Odgovor LLM v obliki kode in pripravljena vizualizacija rezultatov izvajanja kode neposredno v klepetalnici LLM :

The screenshot shows a Jupyter Notebook cell with the following Python code:

```

1 # Re-importing necessary libraries for visualization
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Visualization 1: Bar Chart
6 plt.figure(figsize=(10, 6))
7 df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
8 plt.title('Data Verification Summary - Bar Chart')
9 plt.xlabel('Data Categories')
10 plt.ylabel('Count')
11 plt.xticks(rotation=45)
12 plt.tight_layout()
13 plt.show()

```

Below the code cell is a "Run in IDE" button.

The resulting visualization is a stacked bar chart titled "Data Verification Summary - Bar Chart". The Y-axis is labeled "Count (of values checked)" and ranges from 0 to 5. The X-axis is labeled "Data Categories" and includes "Window\_ID", "Energy\_Efficiency", "Warranty\_Period", and "Replacement\_Cycle". The legend indicates "Passed" (green) and "Failed" (red). The chart shows that for each category, the "Passed" count is consistently higher than the "Failed" count, with the total count for "Window\_ID" being the highest (around 5) and "Warranty\_Period" being the lowest (around 4.5).

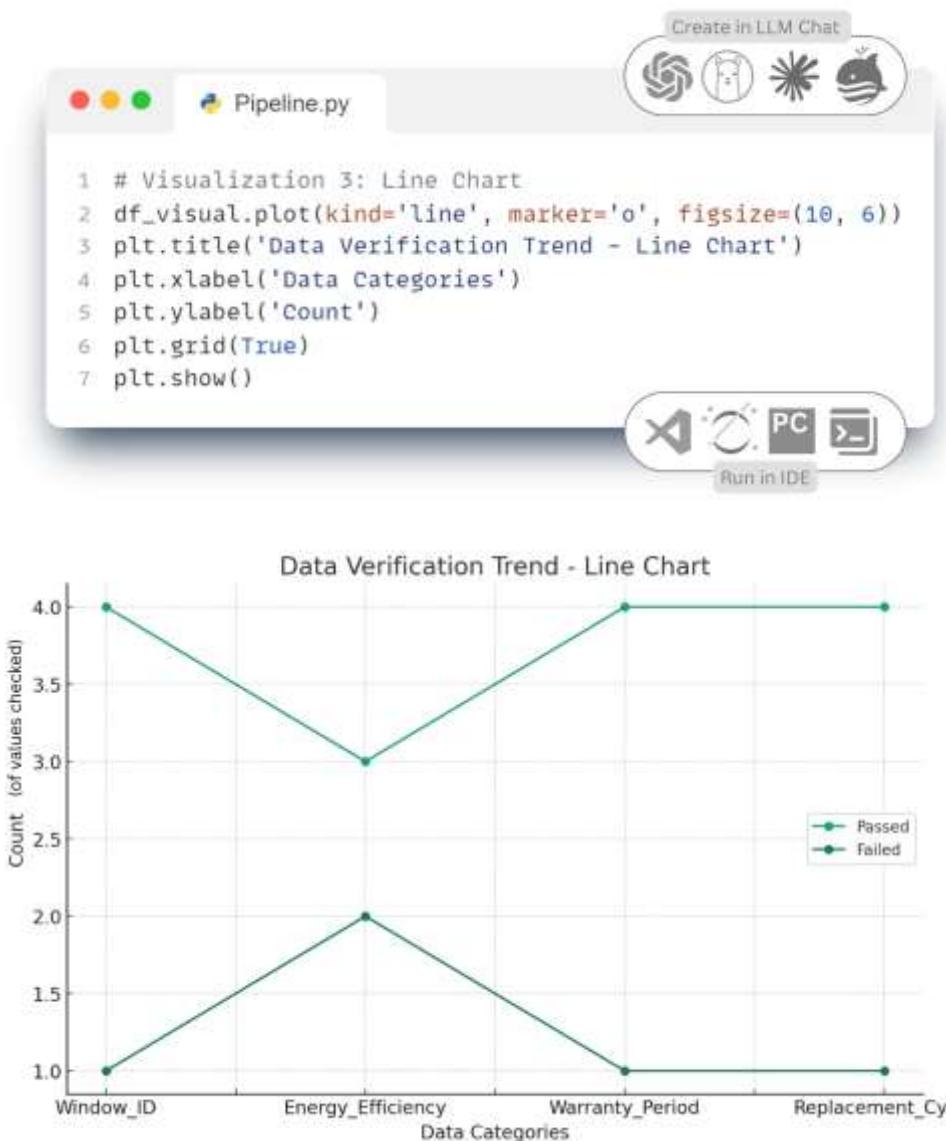
Data Category	Passed (Count)	Failed (Count)
Window_ID	~4.0	~1.0
Energy_Efficiency	~3.0	~2.0
Warranty_Period	~4.0	~0.5
Replacement_Cycle	~3.5	~1.0

Slika 7.2-9 Vizualizacija rezultatov koraka Transform preverjanja vrednosti atributov iz sistema RPM (slika 7.2-7) kot histogram v koraku Load .

- 2 Obstaja veliko odprtakodnih in brezplačnih knjižnic za vizualizacijo, ki omogočajo predstavitev strukturiranih podatkov v različnih oblikah. Nadalujmo z vizualizacijo rezultatov z drugo vrsto grafa z naslednjim pozivom v klepetu:

Iste podatke narišite kot graf črt. ↪

- 2 Odgovor LLM :



Slika 7.2-10 Vizualizacija validacijskih podatkov (slika 7.2-8) v obliki linijskega diagrama, pridobljenega z uporabo knjižnice Matplotlib .

Obstaja veliko odprtakodnih in brezplačnih knjižnic za vizualizacijo, kot so:

- Seaborn - za statistične grafe (slika 7.2-11)
- Plotly - za interaktivne spletne vizualizacije (Slika 7.2-12, Slika 7.1-6)
- Altair - za deklarativno vizualizacijo
- Dash ali Streamlit - za ustvarjanje popolnih nadzornih plošč

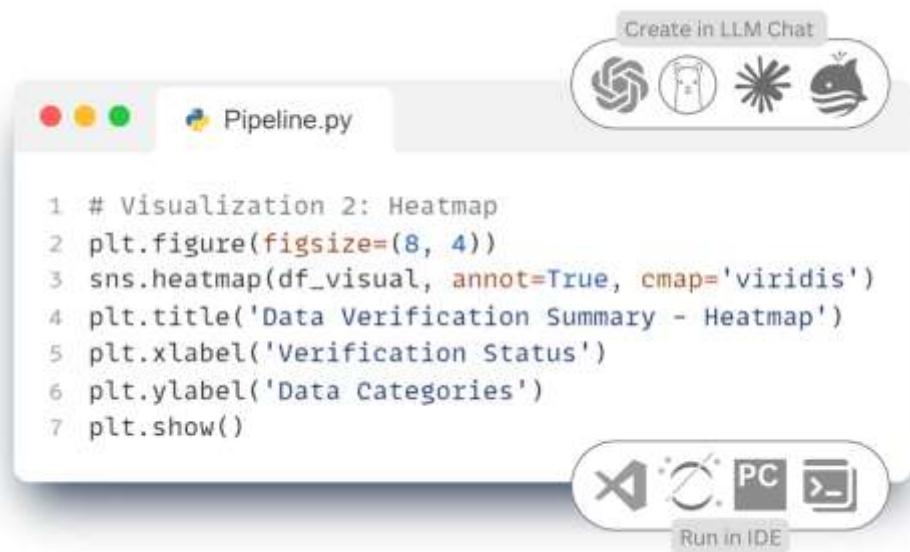
Poznavanje posebnih knjižnic za vizualizacijo ni potrebno - sodobna orodja, vključno z LLM, omogočajo samodejno ustvarjanje kode za grafe in celotne aplikacije na podlagi opisa naloge.

Izbira orodja je odvisna od ciljev projekta: ali gre za poročilo, predstavitev ali spletno nadzorno ploščo. Odprtakodna knjižnica Seaborn je na primer še posebej dobra za delo s kategoričnimi podatki, saj pomaga prepoznati vzorce in trende.

- Če si želite ogledati delajočo knjižnico Seaborn, lahko za uporabo želene knjižnice zaprosite neposredno MŠZŠ ali pa mu pošljete podobno besedilno zahtevo na naslov:

Prikaži topotno karto za ↗

- Odgovor LLM v obliki kode in dokončanega grafa, katerega kodo za izris lahko zdaj kopirate v IDE, sam graf pa lahko kopirate ali shranite za vstavljanje v dokument:



```

1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

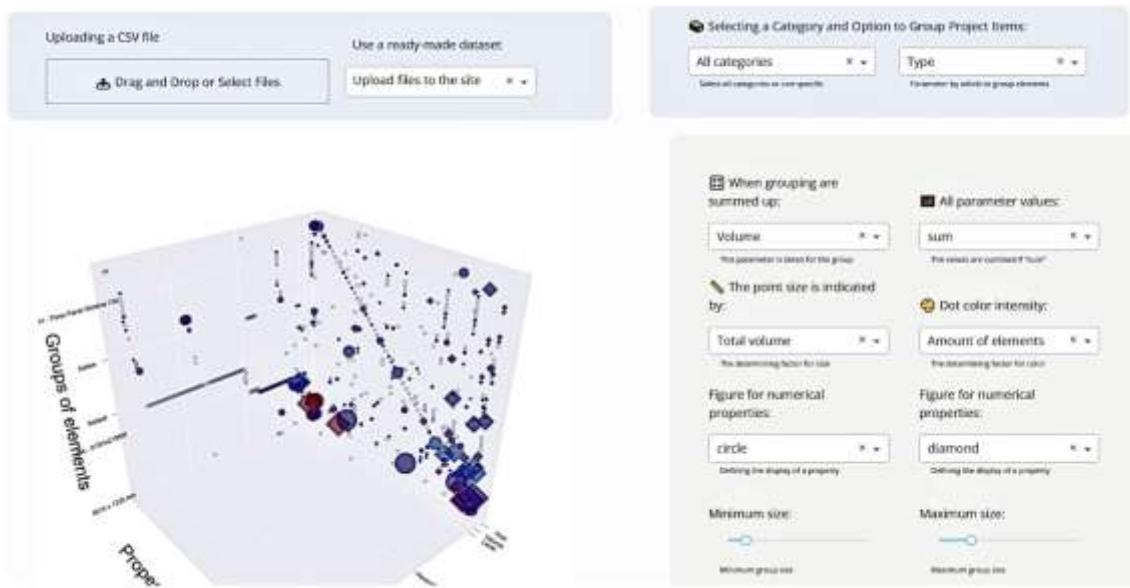
```

The screenshot shows a Streamlit application window titled "Pipeline.py". At the top right are icons for "Create in LLM Chat", a large blue "Run" button, and smaller icons for "PC" and "Email". Below the title bar, there's a toolbar with three colored dots (red, yellow, green) and a "Pipeline.py" icon. The main area contains the Python code for generating a heatmap. At the bottom right is a "Run in IDE" button.



Slika 7.2-11 Prikaz rezultatov potrjevanja (slika 7.2-8) podatkov z uporabo knjižnice Seaborn.

Za tiste, ki imajo raje interaktivni pristop, obstajajo orodja, ki omogočajo ustvarjanje dinamičnih grafikonov in plošč z možnostjo interakcije. Knjižnica Plotly (slika 7.1-6, slika 7.2-12) ponuja možnost ustvarjanja zelo interaktivnih grafov in plošč, ki jih je mogoče vgraditi v spletnne strani in uporabniku omogočajo interakcijo s podatki v realnem času.



Slika 7.2-12 Interaktivna 3D -vizualizacija atributov elementov iz projekta CAD- (BIM-) z uporabo knjižnice Plotly.

Specializirane odprtakodne knjižnice Bokeh, Dash in Streamlit zagotavljajo priročen način predstavitve podatkov, ne da bi za to potrebovali poglobljeno znanje spletnega razvoja. Bokeh je primeren za kompleksne interaktivne grafe, Dash se uporablja za izdelavo polnovrednih analitičnih nadzornih plošč, Streamlit pa omogoča hitro izdelavo spletnih aplikacij za analizo podatkov.

S takšnimi orodji za vizualizacijo lahko razvijalci in analitiki učinkovito posredujejo rezultate sodelavcem in zainteresiranim stranem, kar omogoča intuitivno interakcijo s podatki in poenostavlja sprejemanje odločitev.

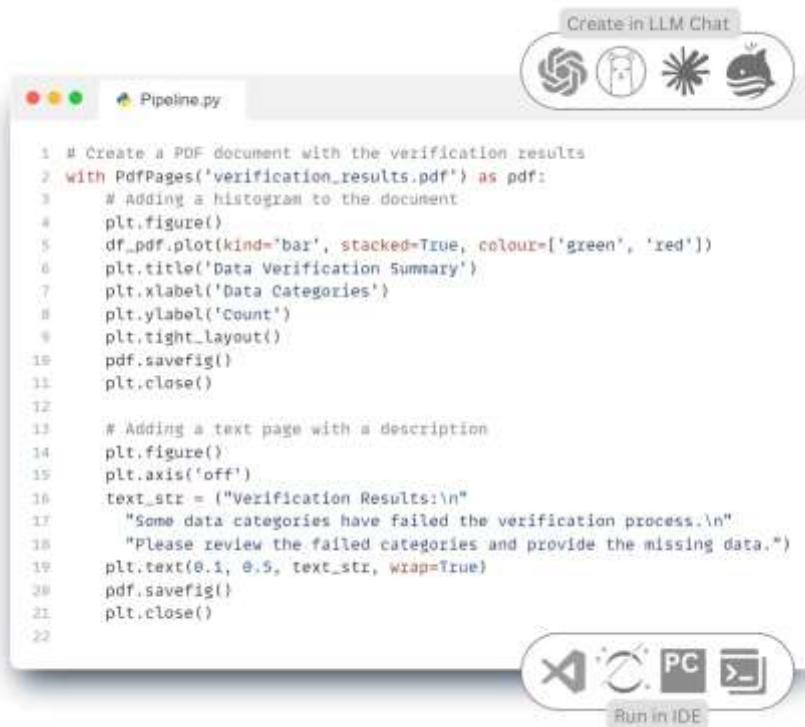
## ETL Load: samodejno ustvarjanje dokumentov PDF

V fazi nalaganja podatkov lahko podatke ne le vizualizirate, jih prenesete v tabele ali podatkovne zbirke, temveč tudi samodejno ustvarite poročila, vključno s potrebnimi grafikoni, diagrami in ključnimi analitičnimi kazalniki, ki jih prejme vodja ali strokovnjak, ki čaka na rezultate pregleda. Avtomatizirana poročila lahko vsebujejo tako komentarje kot besedilno razlago podatkov in elemente vizualizacije - tabele, grafe.

- Za izdelavo poročila PDF s histogramom (slika 7.2-9) in opisom analize na podlagi preverjanja, ki smo ga opravili v prejšnjih poglavjih, je dovolj, da v nadaljevanju dialoga z LLM oblikujete zahtevo, npr.:

Napišite kodo za ustvarjanje datoteke PDF s histogramom in opisom rezultatov zgornjega preverjanja podatkov (klepetu) ter napišite besedilno opozorilo, da nekatere kategorije niso bile preverjene in da je treba dopolniti manjkajoče. ↴

- LLM-jev odgovor kot koda in dokončan PDF z rezultati:



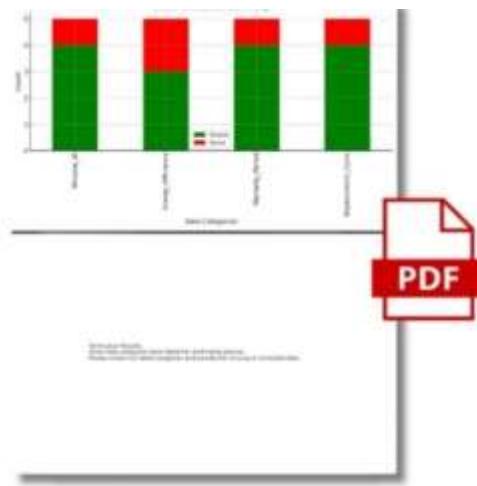
The screenshot shows a code editor window titled "Pipeline.py". The code is as follows:

```

1 # Create a PDF document with the verification results
2 with PdfPages('verification_results.pdf') as pdf:
3     # Adding a histogram to the document
4     plt.figure()
5     df_pdf.plot(kind='bar', stacked=True, colour=['green', 'red'])
6     plt.title('Data Verification Summary')
7     plt.xlabel('Data Categories')
8     plt.ylabel('Count')
9     plt.tight_layout()
10    pdf.savefig()
11    plt.close()
12
13    # Adding a text page with a description
14    plt.figure()
15    plt.axis('off')
16    text_str = ("Verification Results:\n"
17                "Some data categories have failed the verification process.\n"
18                "Please review the failed categories and provide the missing data.")
19    plt.text(0.1, 0.5, text_str, wrap=True)
20    pdf.savefig()
21    plt.close()
22

```

At the top of the editor, there are icons for "Create in LLM Chat" and "Run in IDE". Below the code editor, there are icons for "VS Code", "PC", and "Server".



**Slika 7.2-13 Avtomatizirana koda ustvari dokument PDF, ki vsebuje histogram s podatki o preskušu in besedilo z rezultati preskusa.**

Samodejno napisana rešitev z le 20 vrsticami kode z uporabo LLM takoj ustvari želeni dokument PDF (ali DOC) z vizualizacijo v obliki histograma atributov (slika 7.2-13), ki prikazuje število podatkov, ki so uspešno in neuspešno opravili preverjanje, ter z dodatnim besedilnim blokom, ki povzema rezultate in priporočila nadaljnje ukrepe.

Avtomatizirano ustvarjanje dokumentov je ključni element faze Load, zlasti v projektnem okolju, kjer sta hitrost poročanja in natančnost ključnega pomena.

## **ETL Load: samodejno ustvarjanje dokumentov iz FPDF**

Avtomatizacija poročanja v fazi ETL Load je pomemben korak pri obdelavi podatkov, zlasti kadar je treba rezultate analiz predstaviti v obliki, ki je enostavna za sporočanje in razumevanje. V gradbeništvu je to pogosto pomembno za poročila o napredku, statistične podatke o projektu, poročila o zagotavljanju kakovosti ali finančno dokumentacijo.

Eno najpriročnejših orodij za takšna opravila je odprtokodna knjižnica, FPDF, ki je na voljo za Python in PHP.

Odprtokodna knjižnica **FPDF** zagotavlja prilagodljiv način ustvarjanja dokumentov s kodo, ki omogoča dodajanje glave, besedila, tabel in slik. Uporaba kode namesto ročnega urejanja zmanjša število napak in pospeši postopek priprave poročil v formatu PDF.

Eden ključnih korakov pri ustvarjanju dokumenta PDF je dodajanje naslovov in glavnega besedila v obliki komentarjev ali opisov. Pri ustvarjanju poročila pa ni pomembno le dodajanje besedila, temveč tudi njegovo pravilno strukturiranje. Naslovi, odmiki, razmiki med vrsticami - vse to vpliva na berljivost dokumenta. Z uporabo programa FPDF, lahko nastavite parametre oblikovanja, nadzorujete razporeditev elementov in prilagodite slog dokumenta.

FPDF je načeloma zelo podoben HTML. Tisti, ki že poznajo HTML, lahko z uporabo FPDF zlahka

ustvarijo poljubno zapletene dokumente PDF, saj je struktura kode zelo podobna označevanju HTML: glave, besedilo, slike in tabele se dodajajo na podoben način. Tistim, ki ne poznajo HTML, ni treba skrbeti - uporabite lahko LLM, ki vam bo takoj pomagal sestaviti kodo za ustvarjanje želene postavitve dokumenta.

- Naslednji primer prikazuje, kako ustvariti poročilo z glavo in osnovnim besedilom. Izvedba te kode v katerem koli IDE s podporo za Python ustvari datoteko PDF, ki vsebuje želeno glavo in besedilo:

```
from fpdf import FPDF      # Uvoz knjižnice FPDF
pdf = FPDF()    # Ustvari PDF -dokument
pdf.add_page()   # Dodaj stran

pdf.set_font("Arial", style='B', size=16) # Nastavite pisavo: Arial, krepko, velikost 16
pdf.cell(200, 10, "Poročilo o projektu", ln=True, align='C') # Ustvari naslov in ga osredini
pdf.set_font("Arial", size=12) # Spremeni pisavo na navadno Arial, velikost 12
pdf.multi_cell(0, 10, "Ta dokument vsebuje podatke o rezultatih preverjanja projektne datoteke . . . ")
# Dodajte večvrstično besedilo
pdf.output(r "C:\reports\report.pdf") # Shranite PDF -file
```



**Slika 7.2-14 Z nekaj vrsticami kode Python lahko samodejno ustvarimo besedilni dokument PDF, ki ga potrebujemo.**

Pri pripravi poročil je treba upoštevati, da so podatki, iz katerih je dokument sestavljen, redko statični. Glave, besedilni bloki (slika 7.2-14) so pogosto oblikovani dinamično, saj dobijo vrednosti na stopnji Transformacija v procesu ETL.

Z uporabo kode lahko ustvarite dokumente, ki vsebujejo najnovejše informacije: ime projekta, datum izdelave poročila ter informacije o udeležencih ali trenutnem stanju. Uporaba spremenljivk v kodi omogoča samodejno vstavljanje teh podatkov na zahtevana mesta v poročilu, kar popolnoma odpravi potrebo po ročnem urejanju pred pošiljanjem.

Poleg preprostega besedila in naslovov imajo v projektni dokumentaciji posebno mesto tudi tabele. Skoraj vsak dokument vsebuje strukturirane podatke: od opisov objektov do rezultatov pregledov. Samodejno generiranje tabel na podlagi podatkov iz faze Transform omogoča ne le pospešitev postopka priprave dokumenta, temveč tudi zmanjšanje napak pri prenosu informacij. FPDF omogoča vstavljanje tabel v datoteke PDF (kot besedilo ali slike), nastavljanje robov celic, velikosti stolpcev in

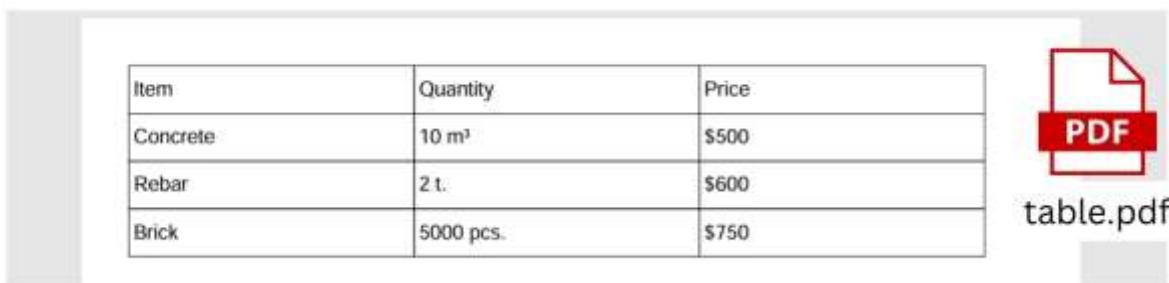
pisav (slika 7.2-15). To je še posebej priročno pri delu z dinamičnimi podatki, ko se število vrstic in stolpcev lahko spreminja glede na naloge dokumenta.

- Naslednji primer prikazuje, kako avtomatizirati ustvarjanje tabel, npr. s seznammi materialov, ocenami ali rezultati preskusov parametrov:

```
podatki = [
    [ "Artikel", "Količina", "Cena"], # Naslovi stolpcev
    [ "Beton", "10 m³", "$ 500."], # Podatki iz prve vrstice
    [ "Armatura", "2 toni", "$ 600"], # Podatki iz druge vrstice .
    [ "opeka", "5.000 kosov", "$ 750."], # Podatki tretje vrstice .
]

pdf = FPDF () # Ustvari PDF -dokument
pdf.add_page() # Dodaj stran
pdf.set_font("Arial", size=12) # Nastavi pisavo

za vrstico v podatkih: # Iskanje vrstic table
    za element v vrstici: # Pojdi skozi celice v vrstici
        pdf.cell(60, 10, item, border=1) # Ustvari celico z robom, širino 60 in višino 10
    pdf.ln() # Prehod na naslednjo vrstico
pdf.output(r "C:\reports\table.pdf") # Shranite PDF -file
```



The screenshot shows a Python script on the left and its output on the right. The script uses the FPDF library to create a table from a list of lists. The output on the right shows a PDF file named 'table.pdf' containing a table with three rows and three columns. The columns are labeled 'Item', 'Quantity', and 'Price'. The rows contain data for Concrete, Rebar, and Brick.

Item	Quantity	Price
Concrete	10 m <sup>3</sup>	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

**Slika 7.2-15 Iz koraka preoblikovanja v PDF lahko samodejno ustvarite ne le besedilo, temveč tudi vse informacije iz tabele.**

V dejanskih scenarijih poročanja so tabele običajno dinamično ustvarjene informacije, pridobljene v fazi preoblikovanja podatkov. V danem primeru (slika 7.2-15) je tabela vstavljena v dokument PDF - dokument v statični obliki: podatki za primer so bili postavljeni v podatkovni slovar (prva vrstica kode), v realnih razmerah pa se takšna podatkovna spremenljivka samodejno izpolni po npr. grupiranju podatkovnega okvira.

V praksi so take tabele pogosto sestavljene na podlagi strukturiranih podatkov iz različnih dinamičnih virov: podatkovnih zbirk, datotek Excel, vmesnikov API ali rezultatov analitičnih izračunov. Najpogosteje se v fazi preoblikovanja (ETL) podatki združujejo, grupirajo ali filtrirajo - in se šele nato preoblikujejo v seštevke v obliki grafov ali dvodimensionalnih tabel, prikazanih v poročilih. To pomeni, da se lahko vsebina tabele spreminja glede na izbrane parametre, obdobje analize, projektne filtre ali uporabniške nastavitev.

Uporaba dinamičnih podatkovnih okvirov in podatkovnih nizov v fazi Transform omogoča, da je postopek poročanja v fazi Load čim bolj prilagodljiv, razširljiv in enostavno ponovljiv brez potrebe po ročnem posredovanju.

Poleg tabel in besedila podpira tudi dodajanje grafov tabelaričnih podatkov, kar vam omogoča, da v poročilo vstavite slike, ustvarjene s knjižnico Matplotlib ali drugimi knjižnicami za vizualizacijo, ki smo jih obravnavali zgoraj. Vse grafe, diagrame in diagrame lahko v dokument dodate s pomočjo kode.

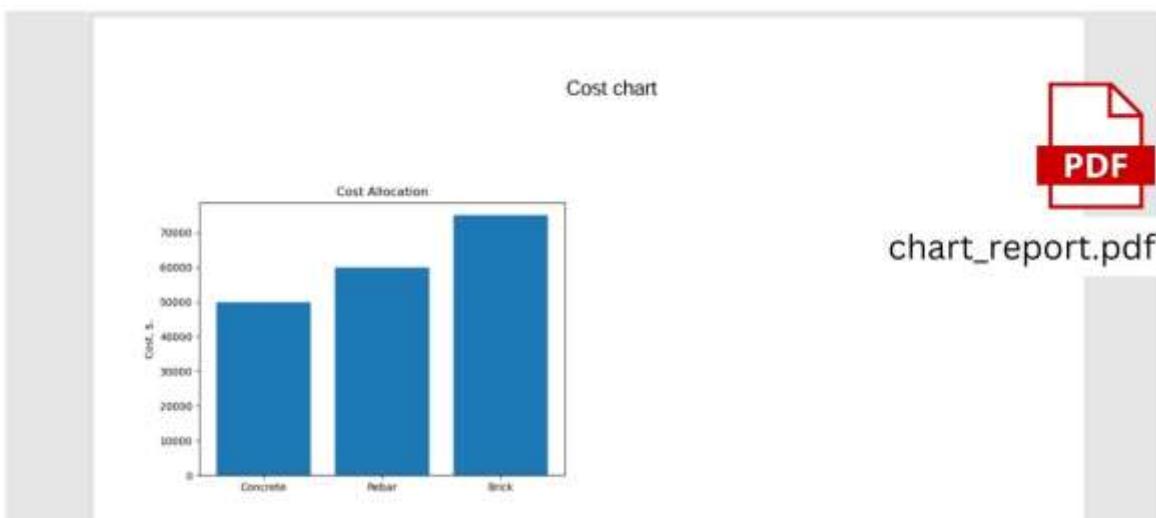
- Z uporabo knjižnice Python FPDF, dodajmo v dokument PDF graf, ki je bil predhodno ustvarjen s programom Matplotlib:

```
uvoz matplotlib.pyplot kot plt # Uvozite Matplotlib za ustvarjanje grafov

fig, ax = plt.subplots() # Ustvari sliko in osi grafa
categories = ["Concrete", "Rebar", "Brick"] # Imena kategorij
values = [50000, 60000, 75000] # Vrednosti kategorij
ax.bar(categories, values) # Ustvari stolpčni graf
plt.ylabel("Vrednost,$.") # Označite os Y
plt.title("Porazdelitev stroškov") # Dodajte naslov
plt.savefig(r "C:\reports\chart\chart\chart.png") # Shranite graf kot sliko

pdf = FPDF () # Ustvari PDF -dokument
pdf.add_page() # Dodaj stran
pdf.set_font("Arial", size=12) # Nastavi pisavo
pdf.cell(200, 10, "Cost Chart", ln=True, align='C') # Dodajte glavo

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100) # Vstavite sliko v
PDF (x, y - koordinate, w - širina)
pdf.output(r "C:\reports\chart_report.pdf") # Shranite datoteko PDF
```



**Slika 7.2-16 Z ducatom vrstic kode lahko ustvarite graf, ga shranite in nato prilepite v dokument PDF.**

S FPDF postane postopek priprave dokumentov in logike pregleden, hiter in priročen. Predloge, vgrajene v kodo, omogočajo generiranje dokumentov z aktualnimi podatki, s čimer se odpravi potreba po ročnem izpolnjevanju.

Uporaba avtomatizacije ETL - namesto zamudnega ročnega poročanja se lahko strokovnjaki osredotočijo na analiziranje podatkov in sprejemanje odločitev, namesto da izbirajo pravo orodje za delo z določenim podatkovnim silosom z jasnim uporabniškim vmesnikom.

Knjižnica FPDF tako zagotavlja prilagodljivo orodje za avtomatizirano ustvarjanje dokumentov vseh zahtevnosti - od kratkih tehničnih poročil do kompleksnih analitičnih povzetkov s tabelami in diagrami, kar omogoča ne le hitrejši pretok dokumentov, temveč tudi znatno zmanjša verjetnost napak, povezanih z ročnim vnosom podatkov in oblikovanjem.

## ETL Load: poročanje in nalaganje v druge sisteme

V fazi Load so bili rezultati pripravljeni v obliki preglednic, grafov in končnih poročil PDF, pripravljenih v skladu z določenimi zahtevami. Nadalje je mogoče te podatke izvoziti v strojno berljive formate (npr. CSV), kar je potrebno za integracijo z zunanjimi sistemi, kot so ERP, CAFM, CPM, platforme BI in druge korporativne ali industrijske rešitve. Poleg CSV je mogoče prenesti tudi v XLSX, JSON, XML ali neposredno v podatkovne zbirke, ki podpirajo samodejno izmenjavo informacij.

- 💡 Če želite ustvariti ustrezeno kodo za avtomatizacijo koraka Load, preprosto poizvedite po vmesniku LLM, na primer: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude ali QWEN:

Napišite kodo za ustvarjanje poročila o rezultatih preverjanja podatkov v podatkovnem okviru, kjer se stolpci s predpono "preverjeno\_" preštejejo, preimenujejo v "uspešno" in "neuspešno", manjkajoče vrednosti se nadomestijo z 0, nato pa se v datoteko CSV izvozijo samo tiste vrstice, ki so uspešno opravile vsa preverjanja.

 Odgovor LLM:



```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a CSV file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

The screenshot shows a Jupyter Notebook cell with the file name 'Pipeline.py' at the top. The cell contains Python code for data validation and export. At the top right of the notebook interface, there are icons for 'Create in LLM Chat', a profile picture, a flower icon, and a whale icon. Below the notebook, there are icons for a terminal, a clock, 'PC', and a file, with a 'Run in IDE' button at the bottom.

Slika 7.2-17 Potrjeni podatki, pridobljeni v koraku preoblikovanja iz končnega podatkovnega okvira, se izvozijo v datoteko CSV za integracijo z drugimi sistemi.

Zgornja koda (slika 7.2-17) izvaja zadnjo fazo postopka ETL - proces - Load, med katero se preverjeni podatki shranijo v obliki CSV, ki je združljiva z večino zunanjih sistemov in podatkovnih zbirk. S tem smo zaključili celoten cikel procesa ETL, vključno z ekstrakcijo, transformacijo, vizualizacijo, dokumentiranjem in izvozom podatkov v potrebne sisteme in formate, kar zagotavlja ponovljivost, preglednost in avtomatizacijo dela z informacijami.

Cevovod ETL se lahko uporablja tako za obdelavo posameznih projektov kot tudi za obsežne aplikacije - pri analizi na stotine ali tisoče prejetih podatkov v obliki dokumentov, slik, skeniranj, projektov CAD, oblakov točk, datotek PDF ali drugih virov, ki prihajajo iz porazdeljenih sistemov. Zaradi zmožnosti popolne avtomatizacije postopka ETL ni le orodje za tehnično obdelavo, temveč temelj digitalne informacijske infrastrukture v gradbeništvu.

## ETL z LLM: vizualizacija podatkov iz dokumentov PDF

Čas je, da preidete na oblikovanje celovitega postopka ETL, ki zajema vse ključne faze obdelave podatkov v enem samem scenariju - ekstrakcijo, preoblikovanje in nalaganje. Zgradimo avtomatiziran ETL-poteznič, ki vam omogoča obdelavo dokumentov PDF brez ročnega dela - pridobivanje podatkov iz dokumentov, vizualizacijo, analizo in prenos v druge sisteme.

Postopek ETL v našem primeru bo opisan s pozivi, ki bodo morali jezikovnemu modelu (LLM) pojasniti vse postopke ETL z opisom končnega rezultata, ki ga je treba dobiti. V tem primeru je naloga poiskati vse datoteke PDF v določeni mapi in njenih podmapah, iz njih pridobiti ustrezne informacije - npr. imena materialov, količine in stroške - ter rezultat predstaviti kot strukturirano tabelo (DataFrame) za nadaljnjo analizo.

- ❶ Prva besedilna poizvedba v LLM za samodejno pridobivanje podatkov iz več dokumentov PDF in ustvarjanje podatkovnega okvira podatkov za korak Izvleček:

Napišite kodo za pridobivanje informacij o gradivu iz datotek PDF v dani mapi in njenih podmapah. Podatki v datoteki PDF vključujejo ime materiala, količino in stroške. Rezultat je treba shraniti v podatkovni okvir ↵

- ❷ Odgovor LLM :



```

1 import os
2 import pdfplumber
3 import pandas as pd
4
5 def extract_data(base_folder_path):
6     """
7         Extract material data from PDF files in the specified base
8         folder and all its subfolders. Assumes each line in the PDF contains
9         'Material' followed by material name, quantity, and cost.
10    """
11    data = []
12    # Walk through each folder and subfolder in the base folder
13    for folder_path, _, files in os.walk(base_folder_path):
14        for file in filter(lambda f: f.endswith('.pdf'), files):
15            with pdfplumber.open(os.path.join(folder_path, file)) as pdf:
16                # Process each page in the PDF
17                for page in pdf.pages:
18                    # Splitting text into lines
19                    for line in page.extract_text().split('\n'):
20                        # Check if the line contains material data
21                        if 'Material' in line:
22                            parts = line.split()
23                            # Append data to list; convert quantity and cost to float
24                            data.append({'file': file, 'material': parts[1],
25                                         'quantity': float(parts[2]), 'cost': float(parts[3])})
26
27    # Convert list of dictionaries to pandas DataFrame
28    return pd.DataFrame(data)
29
30 # Specify the base folder path containing PDF files and subfolders with PDFs
31 df = extract_data(r'C:\')

```

Slika 7.2-18 LLM ustvari kodo Python za pridobivanje podatkov iz datotek PDF v določeni mapi in vseh njenih podmapah.

Odziv LLM (slika 7.2-18) je pripravljen skript Python, ki samodejno preleti vse mape, odpre najdene datoteke PDF, iz njih pridobi besedilne informacije in jih pretvori v tabelo. Dobljeno klepetalno kodo lahko zaženete v enem od priljubljenih IDE PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z vtičnikom PyDev, Thonny, Wing IDE, IntelliJ IDEA z vtičnikom Python, JupyterLab ali priljubljenih spletnih orodij Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Za poenostavitev postopka zbiranja lahko namesto kopiranja kode iz programa LLM in uporabe kode v IDE naložimo tudi ducat datotek PDF -file neposredno v klepetalnico LLM (slika 7.2-19) in kot izhod dobimo tabelo, ne da bi nam bilo treba videti kodo ali jo zagnati. Rezultat izvajanja te kode bo tabela z atributi, ki smo jih izbrali.

ChatGPT 4 ~

You  
Write code to extract material information from PDF files in a given folder. The data in the PDF includes material name, quantity and cost. The result should be saved in a pandas DataFrame and displayed on the screen.

ChatGPT  
Below is a sample Python script that demonstrates this process using 'PyPDF2'. Please note that you might need to adjust the text parsing part based on the specific format of your PDF files.

```
python
import os
import PyPDF2
import pandas as pd

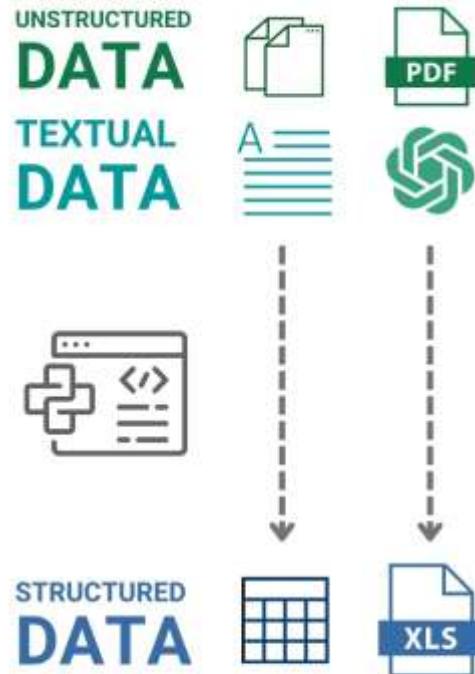
# Create a DataFrame from the extracted data
df = pd.DataFrame(data)

# Display the DataFrame
print(df)
```

Here's an example of how the output would look in a pandas DataFrame after extracting and processing the material information from the PDF files:

	Material Name	Quantity	Cost
0	Copper Wire	100	250.50
1	Aluminum Sheet	50	150.75
2	Plastic Pipes	200	100.00
3	Steel Rods	80	300.00

This table shows the material name, quantity, and cost as columns, with each row representing the data extracted from a PDF file. The actual content and structure of your DataFrame will depend on the specific format and data in your PDF files. [-]

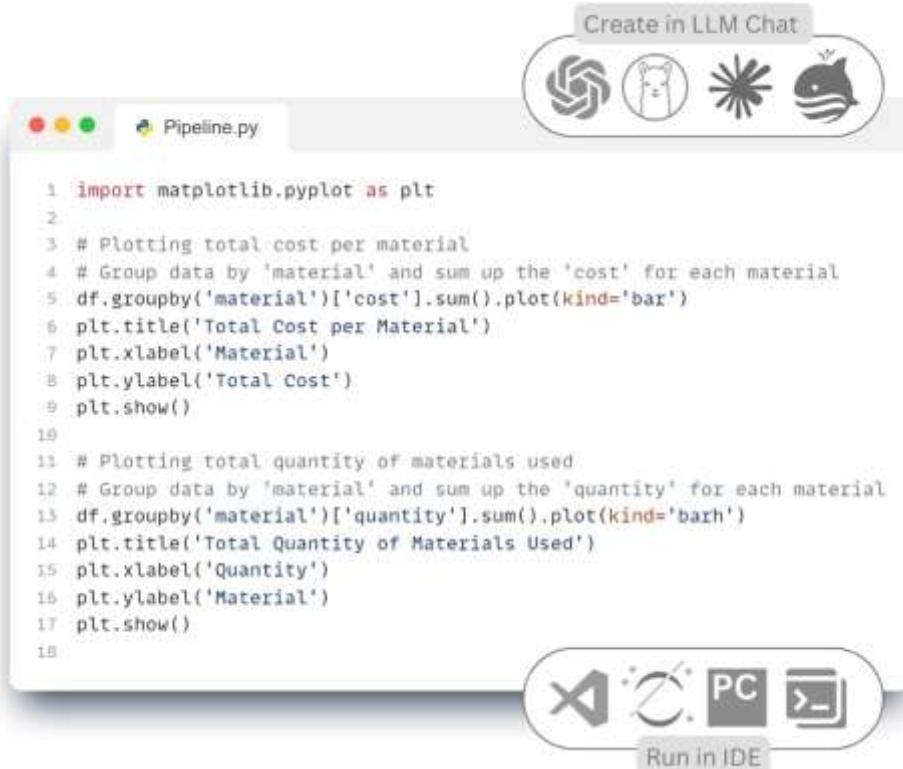


Slika 7.2-19 Rezultat izvajanja kode v LLM, ki podatke iz datotek PDF pridobi v strukturiranem pogledu podatkovnega okvira z izbranimi atributimi.

V naslednjem koraku zaprosimo za jezikovni model pridobljenih podatkov - na primer za primerjavo stroškov in obsega uporabe materiala ter oblikovanje nekaterih primerov vizualizacij, ki bodo služili kot podlaga za nadaljnjo analizo.

- ❷ V nadaljnjem pogоворu z LLM prosite, da iz tabel, ki so bile izdelane v koraku Transformacija (slika 7.2-18), izrišete nekaj grafov:

Prikažite skupne stroške in količino vsakega materiala iz podatkovnega okvira (slika 7.2-18) ↴



The screenshot shows a code editor window titled "Pipeline.py". The code uses the matplotlib library to plot total cost and quantity for five materials: Wood, Steel, Concrete, Glass, and Plastic. The first plot, titled "Total Cost per Material", is a bar chart showing costs ranging from approximately 1500 to 5000. The second plot, titled "Total Quantity of Materials Used", is a horizontal bar chart showing quantities ranging from approximately 150 to 450. A third plot, titled "Cost Distribution by Material", is a pie chart showing the percentage distribution of costs among the materials.

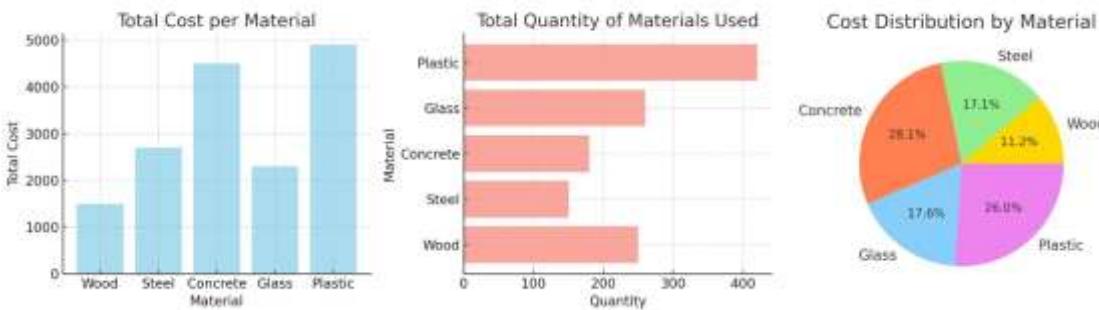
```

1 import matplotlib.pyplot as plt
2
3 # Plotting total cost per material
4 # Group data by 'material' and sum up the 'cost' for each material
5 df.groupby('material')['cost'].sum().plot(kind='bar')
6 plt.title('Total Cost per Material')
7 plt.xlabel('Material')
8 plt.ylabel('Total Cost')
9 plt.show()
10
11 # Plotting total quantity of materials used
12 # Group data by 'material' and sum up the 'quantity' for each material
13 df.groupby('material')['quantity'].sum().plot(kind='barh')
14 plt.title('Total Quantity of Materials Used')
15 plt.xlabel('Quantity')
16 plt.ylabel('Material')
17 plt.show()
18

```

Slika 7.2-20 Odziv LLM -modeli kot koda Pythona za vizualizacijo podatkov iz podatkovnega okvira z uporabo knjižnice matplotlib.

LLM samodejno ustvari in izvede kodo Python (slika 7.2-20) z uporabo knjižnice matplotlib. Po izvedbi te kode dobimo grafe stroškov in porabe materiala v gradbenih projektih neposredno v klepetu (slika 7.2-21), kar močno poenostavi analitično delo.

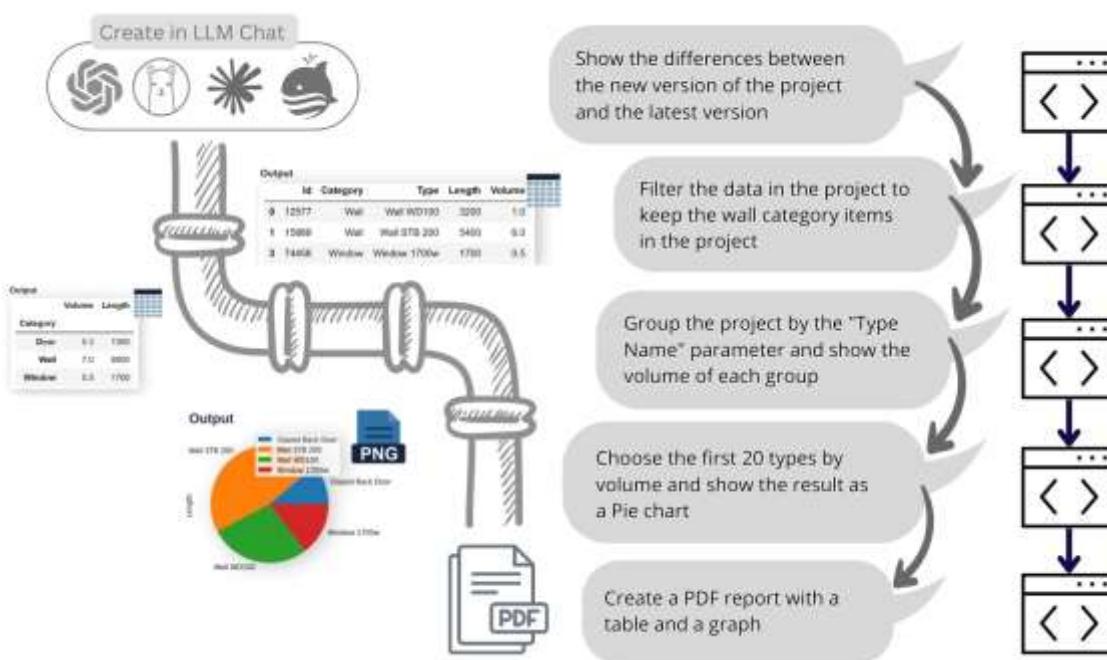


Slika 7.2-21 Vizualizacija odziva LLM v obliki grafov na podlagi podatkov, zbranih v okviru podatkov DataFrame.

Podpora pri razvijanju zamisli za pisanje kode ETL, analizi in izvajjanju kode ter vizualizaciji rezultatov

je na voljo s preprostimi besedilnimi poizvedbami v LLM, ne da bi se bilo treba naučiti osnov programiranja. Pojav orodij umetne inteligence, kot je LLM, vsekakor spreminja pristop k programiraju in avtomatizaciji obdelave podatkov (slika 7.2-22).

V poročilu družbe PwC je zapisano: "Kakšna je resnična vrednost umetne inteligence za vaše podjetje in kako jo lahko izkoristite?" (2017) [139] bosta avtomatizacija procesov in izboljšanje produktivnosti glavni gonili gospodarske rasti. Izboljšanje produktivnosti pa naj bi med letoma 2017 in 2030 predstavljalvo več kot 55% celotne rasti BDP, ki jo poganja umetna inteligencia."



Slika 7.2-22 AI LLM pomaga ustvariti osnutek kode, ki se uporablja v prihodnjih projektih, ne da bi bil potreben LLM.

Z orodji, kot so ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok, ter odprtimi podatki in odprtokodno programsko opremo, lahko avtomatiziramo postopke, ki so bili prej izvedeni le s specializiranimi, dragimi in težko vzdrževanimi modularnimi lastniškimi sistemmi.

V gradbeništvu to pomeni, da bodo podjetja, ki bodo prva uvela avtomatizirane cevovodne - podatkovne procese, imela velike koristi, od izboljšane učinkovitosti upravljanja projektov do zmanjšanja finančnih izgub in odprave razdrobljenih aplikacij in ločenih podatkovnih skladišč.

Opisana logika izvajanja poslovnih nalog v procesu ETL je ključni del avtomatizacije analitike in procesov obdelave podatkov, ki je posebna različica širšega koncepta - cevovodov (Pipelines).

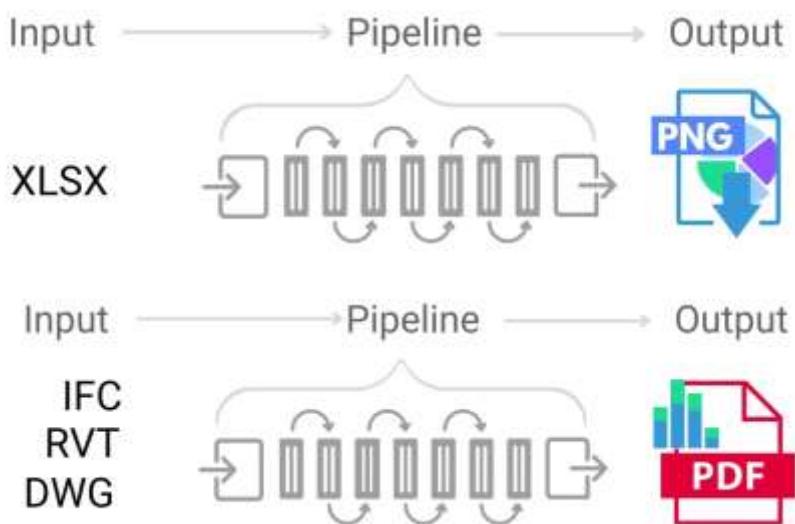
## POGLAVJE 7.3.

### AVTOMATSKI TRANSPORTER ETL (CEVOVOD)

#### Cevovod: Samodejno ETL transporter podatkov.

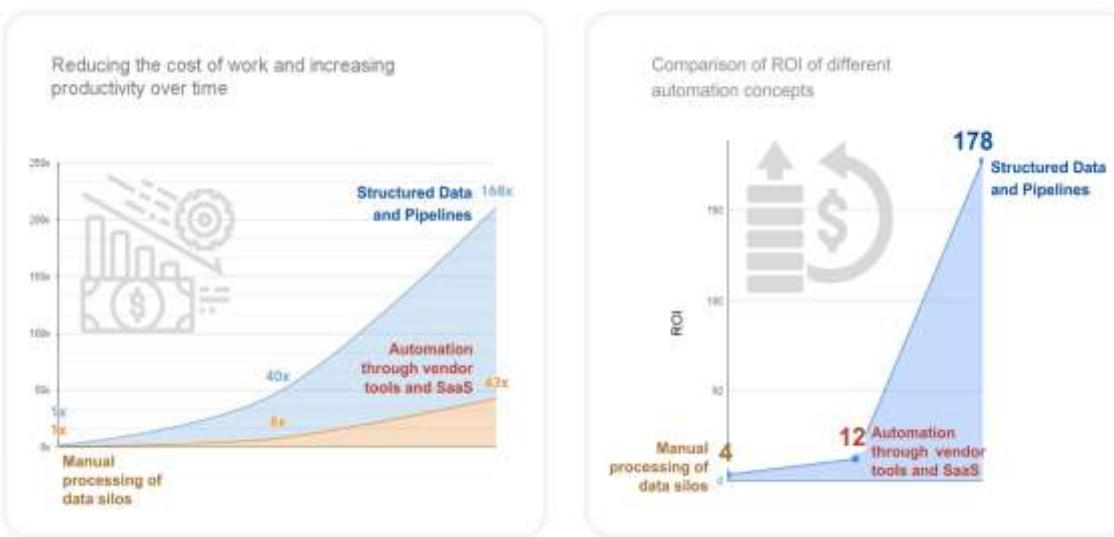
Postopek ETL se tradicionalno uporablja za obdelavo podatkov v analitičnih sistemih, ki zajemajo strukturirane in nestrukturirane vire. Vendar se v današnjem digitalnem okolju vse pogosteje uporablja širši izraz - Pipeline (transporter), ki opisuje vsako zaporedno verigo obdelave, pri kateri izhod ene stopnje postane vhod za naslednjo.

Ta pristop ne velja le za podatke, temveč tudi za druge vrste avtomatizacije: obdelavo nalog, gradnjo poročil, integracijo s programsko opremo in digitalni potek dela (slika 7.3-1).



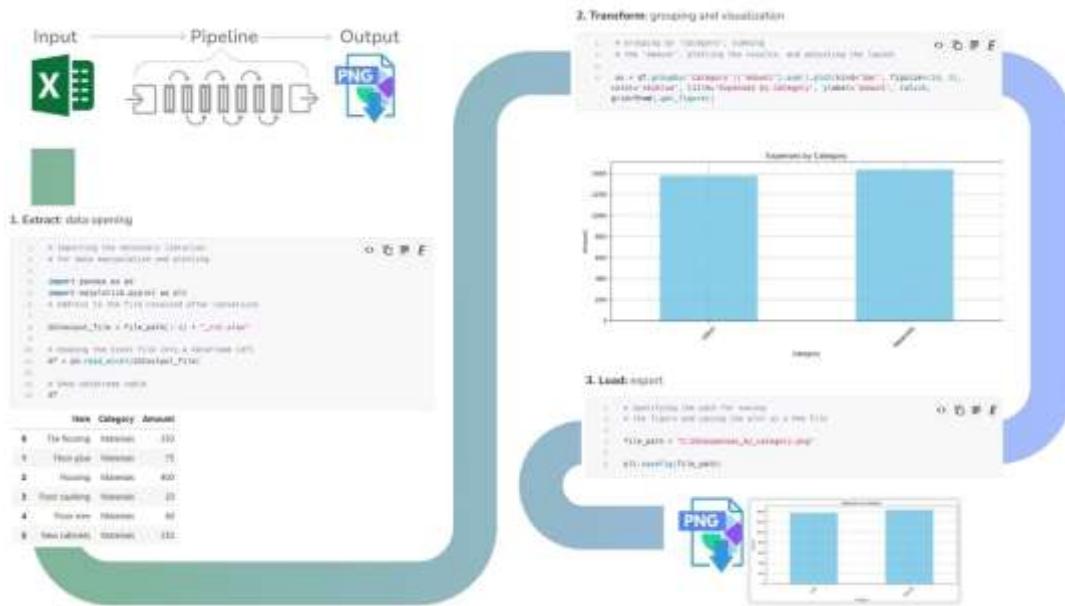
Slika 7.3-1 Cevovod je zaporedje obdelave, v katerem izhod ene stopnje postane vhod za naslednjo stopnjo.

Uporaba cevovodov je eden glavnih elementov avtomatizacije, zlasti pri delu z velikimi količinami heterogenih podatkov. Arhitektura cevovodov omogoča, da so zapleteni koraki obdelave organizirani v modularni, dosledni in obvladljivi oblici, kar povečuje berljivost, poenostavlja vzdrževanje kode ter omogoča postopno odpravljanje napak in skalabilno testiranje.



Slika 7.3-2 ROI Postopek potrjevanja podatkov s cevovodom skrajša čas izvajanja za deset- in stokrat v primerjavi z obdelavo s klasičnimi orodji [74].

Za razliko od ročnega dela v lastniških sistemih (ERP, PMIS, CAD itd.) vam pipelining omogoča znatno (slika 7.3-2) povečanje hitrosti opravil, preprečevanje ponavljanjačega se dela in avtomatizacijo začetka procesov ob pravem času (slika 7.3-3).

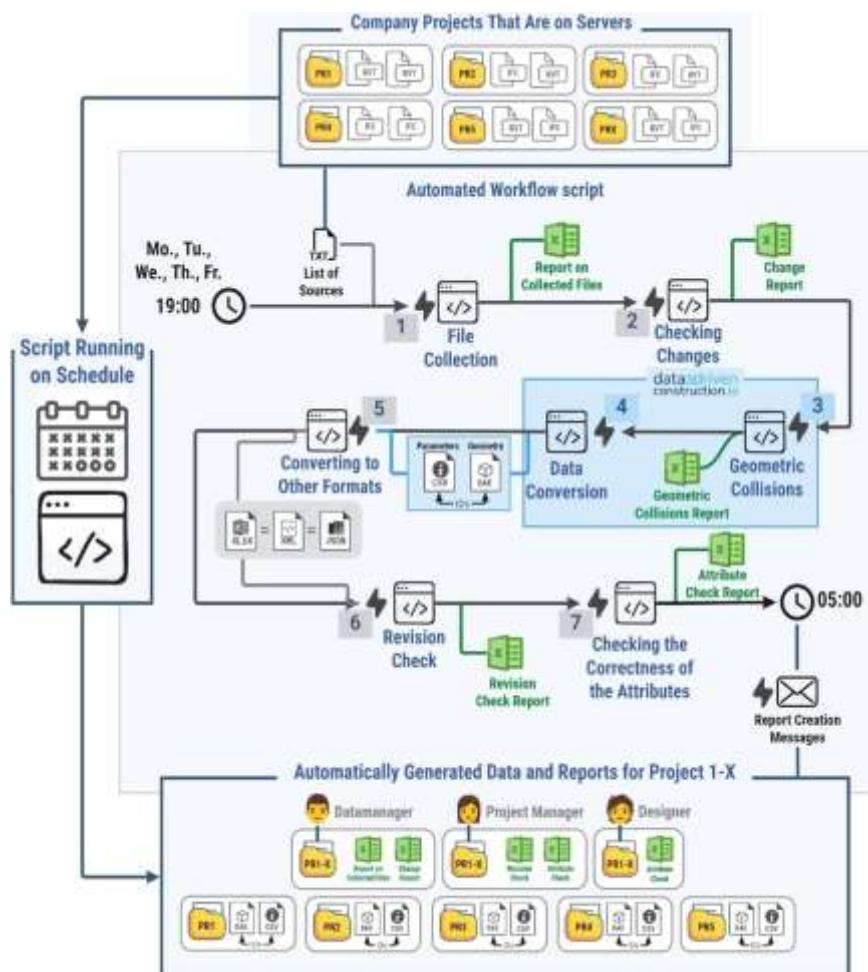


**Slika 7.3-3 Primer ETL Cevovod o samodejnem pridobivanju grafa iz tabelarnih podatkov v datoteki XLSX brez odpiranja programa Excel.**

Če želite obdelati pretočne podatke in zgraditi avtomatizirano cevovodno linijo, podobno postopku ETL, morate vnaprej določiti vire podatkov in časovni okvir za njihovo zbiranje - bodisi za določen poslovni proces bodisi za celotno podjetje.

Pri gradbenih projektih podatki prihajajo iz številnih heterogenih virov z različnimi intervali posodabljanja. Da bi ustvarili zanesljivo podatkovno vitrino, je ključnega pomena beležiti, kdaj se informacije pridobivajo in posodabljujo. To omogoča pravočasno sprejemanje odločitev in izboljša učinkovitost upravljanja projekta.

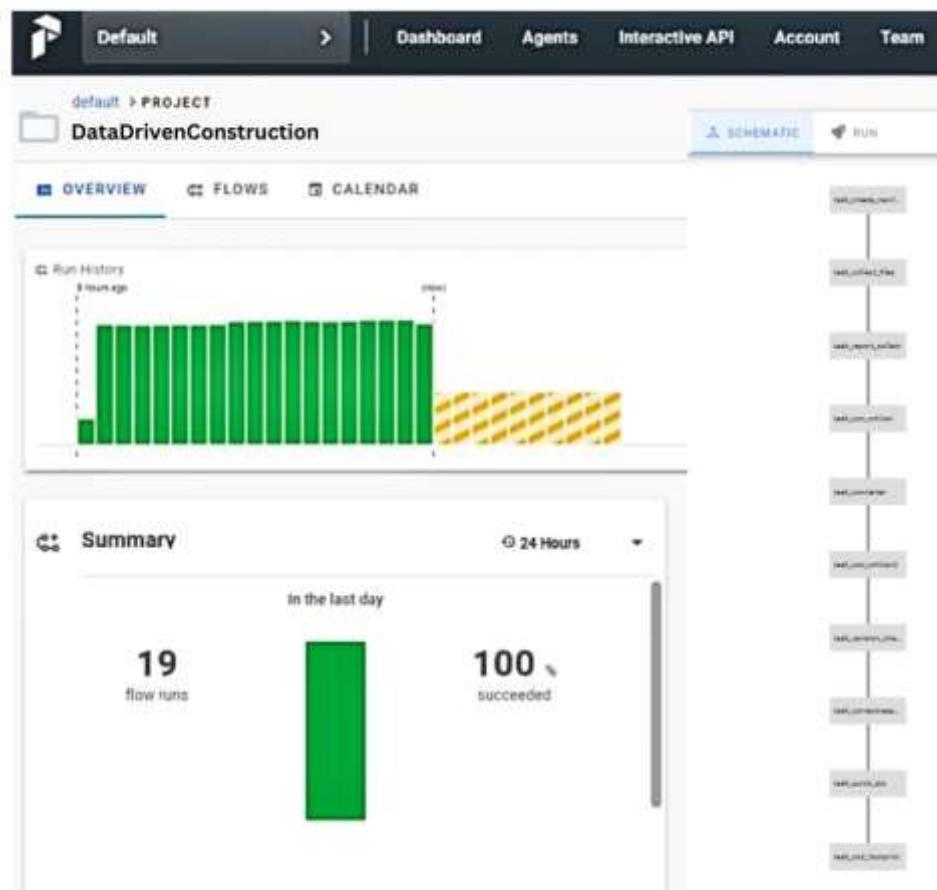
Ena od možnosti je, da se postopek montaže začne ob določenem času - na primer ob 19.00, ob koncu delovnega dne. Na tej točki se aktivira prva skripta, odgovorna za združevanje podatkov iz različnih sistemov in skladišč (slika 7.3-4 korak 1). Sledi samodejna obdelava in preoblikovanje podatkov v strukturirano obliko, primerno za analitiko (slika 7.3-4 korak 2-4). Na zadnji stopnji se s pomočjo pripravljenih podatkov samodejno ustvarijo poročila, nadzorne plošče in drugi izdelki, opisani v prejšnjih poglavjih (slika 7.3-4 korak 6-7). Tako imajo vodje do 5. ure zjutraj že na voljo aktualna poročila o stanju projekta v zahtevani obliki (slika 7.3-5).



Slika 7.3-4 Podatki v cevovodu, ki se samodejno zbirajo zvečer, se čez noč obdelajo, tako da imajo vodje do jutra na voljo posodobljena poročila in sveža poročila.

Pravočasno zbiranje podatkov, opredelitev ključnih kazalnikov uspešnosti, avtomatizacija procesov preoblikovanja in vizualizacija z nadzornimi ploščami so ključni elementi uspešnega odločanja na podlagi podatkov.

Takšni avtomatizirani procesi (slika 7.3-4) se lahko izvajajo popolnoma autonomno: izvajajo se po urniku, obdelujejo podatke brez nadzora in so lahko nameščeni v oblaku ali na lastnem strežniku podjetja (slika 7.3-5). To omogoča, da se taki cevovodi ETL vključijo v obstoječo infrastrukturo IT, pri čemer se ohrani nadzor nad podatki in zagotovi prilagodljivost pri razširjanju.



Slika 7.3-5 Avtomatski procesi ETL -prevoznik (slika 7.3-4) na platformi Prefect, v kateri se vsak delovni dan po 19. uri izmenično izvaja 10 skript python.

Avtomatizacija delovnih postopkov ne le poveča produktivnost ekipe, saj sprosti čas za bolj smiselne in manj rutinske naloge, temveč je tudi pomemben prvi korak k vključitvi tehnologij umetne inteligence (AI) v poslovne procese, kar bomo podrobnejše obravnavali v poglavju o napovedni analitiki in strojnem učenju.

## Cevovod - Postopek potrjevanja podatkov v skladu s standardom LLM

V prejšnjih poglavjih o oblikovanju podatkovnih zahtev in avtomatizaciji ETL smo korak za korakom razčlenili postopek priprave, preoblikovanja, potrjevanja in vizualizacije podatkov. Te dejavnosti so bile izvedene kot ločeni bloki kode (slika 7.2-18 - slika 7.2-20), od katerih vsak izvaja določeno nalogu.

Zdaj imamo naslednji cilj - združiti te elemente v enoten, skladen in avtomatiziran cevovod obdelave podatkov - cevovod, ETL -Pipeline - v katerem se vse faze (nalaganje, preverjanje, vizualizacija, izvoz) izvajajo zaporedno v eni sami samodejno izvedljivi skripti.

V naslednjem primeru bo izведен celoten cikel obdelave podatkov: od nalaganja izvirne datoteke CSV → do preverjanja strukture in vrednosti z uporabo regularnih izrazov → izračuna rezultatov → izdelave vizualnega poročila v obliki PDF.

💡 Za pridobitev ustrezne kode lahko uporabite naslednjo besedilno poizvedbo v LLM:

Napišite vzorec kode, ki naloži podatke iz CSV, potrdi DataFrame z uporabo regularnih izrazov, preveri identifikatorje v obliki "W-NEW" ali "W-OLD", energetsko učinkovitost s črkami od "A" do "G", garancijsko obdobje in cikel zamenjave s številčnimi vrednostmi v letih ter na koncu ustvari poročilo s številom pozitivnih in negativnih vrednosti, ustvari PDF s histogramom rezultatov in doda besedilni opis. ↴

■ Odgovor LLM :



```

1 # Importing necessary libraries
2 import pandas as pd
3 import re
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6
7
8 # Load data
9 df = pd.read_csv('path_to_csv_file.csv')
10
11 # Define patterns and verify data
12 patterns = {
13     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2}\$',
14     'Energy_Efficiency': r'^[A-G]\+\$',
15     'Warranty_Period': r'^\d+\syears\$',
16     'Replacement_Cycle': r'^\d+\syears\$'
17 }
18
19 for column, pattern in patterns.items():
20     df[f'verified_{column}'] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))
21
22 # Create report
23 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
24 report.columns = ['Passed', 'Failed']
25
26 # Generate PDF
27 pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
28 with PdfPages(pdf_path) as pdf:
29     # Plot and save bar chart
30     report.plot(kind='bar', stacked=True, color=['green', 'red'])
31     plt.title('Data Verification Summary')
32     plt.xlabel('Data Categories')
33     plt.ylabel('Count')
34     plt.tight_layout()
35     pdf.savefig()
36     plt.close()
37
38 # Add text page
39 plt.figure()
40 plt.axis('off')
41 text_str = "Verification Results Summary"
42 plt.text(0.1, 0.5, text_str, wrap=True)
43 pdf.savefig()
44 plt.close()
45
46 pdf_path
47

```

Slika 7.3-6 Pipeline (ETL) avtomatizira celoten cikel obdelave podatkov: od prenosa in preverjanja do izdelave strukturiranega poročila v formatu PDF .

Avtomatizirana koda (slika 7.3-6) v klepetalnici LLM ali v DIE bo po kopiranju kode potrdila podatke iz datoteke CSV z uporabo določenih regularnih izrazov, ustvarila poročilo o številu pozitivnih in negativnih zapisov ter nato shranila rezultate preverjanja kot datoteko PDF.

Ta struktura ETL, kjer je vsak korak - od nalaganja podatkov do izdelave poročila - izveden kot ločen modul, zagotavlja preglednost, razširljivost in ponovljivost. Predstavitev logike potrjevanja v obliki lahko berljive kode Python omogoča, da je postopek pregleden in razumljiv ne le za razvijalce, temveč tudi za strokovnjake za upravljanje podatkov, kakovost in analitiko.

Pipeline - pristop k avtomatizaciji obdelave podatkov vam omogoča standardizacijo procesov, povečanje njihove ponovljivosti in poenostavitev prilagajanja novim projektom. S tem ustvarite enotno metodologijo za analizo podatkov, ne glede na vir ali vrsto naloge - naj gre za testiranje skladnosti, poročanje ali prenos podatkov v zunanje sisteme.

Takšna avtomatizacija zmanjšuje število človeških napak, zmanjšuje odvisnost od lastniških rešitev ter povečuje natančnost in zanesljivost rezultatov, zato so primerni tako za operativno analitiko na ravni projekta kot za strateško analitiko na ravni podjetja.

## Cevovod -ETL: preverjanje podatkov in informacij o elementih projekta v CAD (BIM)

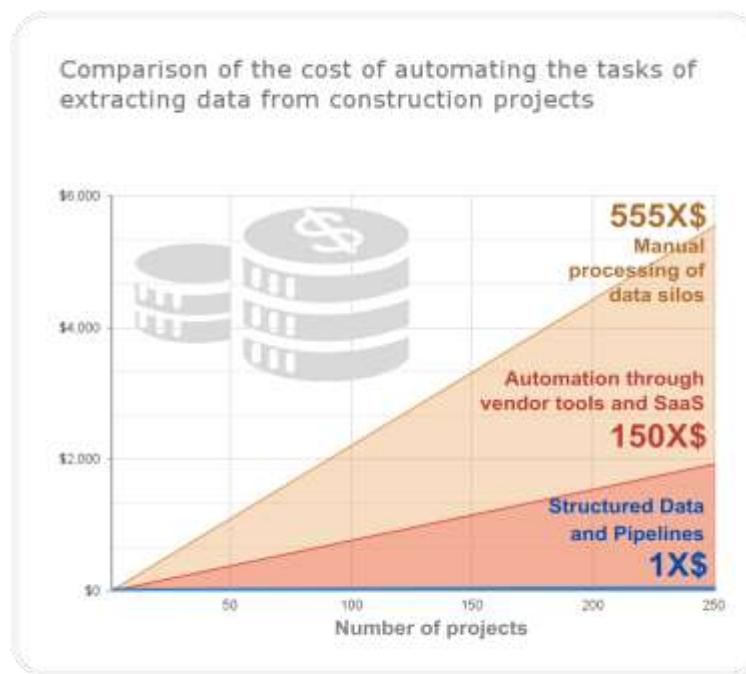
Podatki iz sistemov CAD in podatkovnih zbirk (BIM) so med najbolj izpopolnjenimi in dinamično posodobljenimi viri podatkov v poslovanju gradbenih podjetij. Te aplikacije projekta ne opisujejo le z geometrijo, temveč ga dopolnjujejo tudi z večplastnimi besedilnimi informacijami: prostorninami, lastnostmi materialov, dodelitvami prostorov, stopnjami energetske učinkovitosti, tolerancami, življenjskimi cikli in drugimi atributi.

Atributi, dodeljeni entitetam v modelih CAD, se oblikujejo v fazi načrtovanja in postanejo osnova za nadaljnje poslovne procese, vključno s kalkulacijo stroškov, načrtovanjem, oceno življenjskega cikla in integracijo s sistemi ERP in CAFM, kjer je učinkovitost procesov v veliki meri odvisna od kakovosti podatkov, ki prihajajo iz oblikovalskih oddelkov.

Tradicionalni pristop k potrjevanju atributov v modelih CAD- (BIM-) vključuje ročno potrjevanje (slika 7.2-1), ki postane dolgotrajen in drag postopek, če je obseg modelov velik. Glede na obseg in število sodobnih gradbenih projektov ter njihove redne posodobitve postane postopek potrjevanja in preoblikovanja podatkov nevzdržen in cenovno nedostopen.

Generalni izvajalci in vodje projektov se soočajo s potrebo po obdelavi velikih količin podatkov o projektih, vključno z več različicami in deli istih modelov. Podatki prihajajo od projektantskih organizacij v formatih RVT, DWG, DGN, IFC, NWD in drugih formatih (slika 3.1-14) ter zahtevajo redno preverjanje skladnosti z industrijskimi in korporativnimi standardi

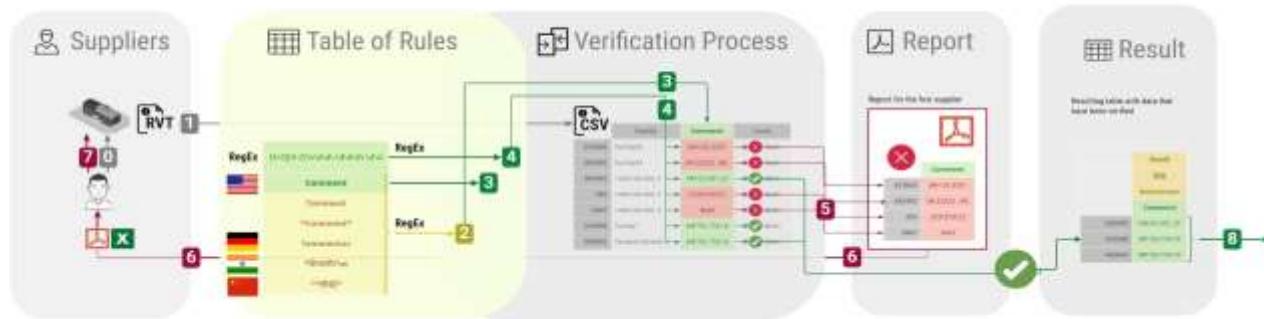
Zaradi odvisnosti od ročnih dejanj in specializirane programske opreme je postopek potrjevanja podatkov ozko grlo v delovnih tokovih, povezanih s podatki iz modelov za celotno podjetje. Avtomatizacija in uporaba strukturiranih zahtev lahko odpravita to odvisnost ter močno povečata hitrost in zanesljivost potrjevanja podatkov (slika 7.3-7).



Slika 7.3-7 Avtomatizacija poveča hitrost preverjanja in obdelave podatkov, kar za desetkrat zmanjša stroške dela [140].

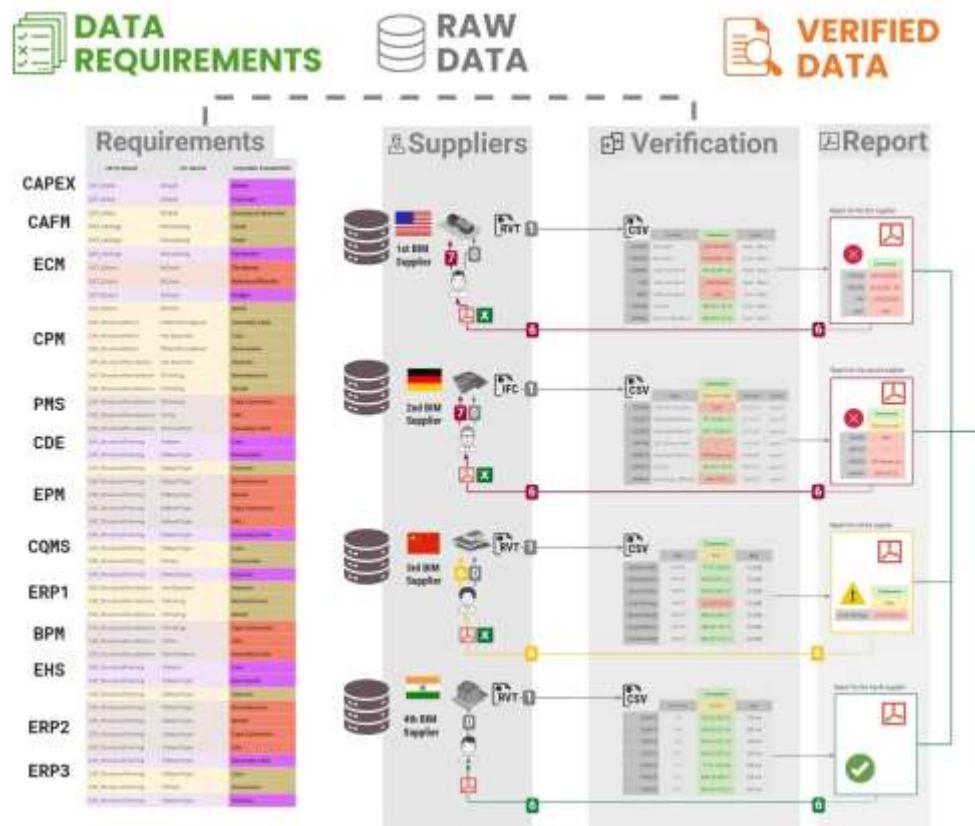
Postopek potrjevanja podatkov CAD vključuje pridobivanje podatkov (faza ETL Extract) iz različnih zaprtih (RVT, DWG, DGN, NWS itd.) ali odprtih polstrukturiranih in parametričnih formatov (IFC, CPXML, USD.) ali odprtih polstrukturiranih in parametričnih formatov (IFC, CPXML, USD), v katerih lahko za vsak atribut in njegove vrednosti (faza Transform) uporabimo tabele pravil z uporabo regularnih izrazov RegEx (slika 7.3-8), postopek, ki smo ga podrobno obravnavali v četrtem delu knjige.

Oblikovanje poročila o napakah v formatu PDF za in uspešno potrjene zapise je treba zaključiti z izpisom (korak Load) v strukturiranih formatih, ki upoštevajo samo potrjene entitete, ki jih je mogoče uporabiti za nadaljnje postopke.



Slika 7.3-8 Postopek potrjevanja podatkov od ponudnikov projektnih podatkov do končnega poročila, potrjenih z uporabo regularnih izrazov.

Avtomatizacija potrjevanja podatkov iz sistemov CAD (BIM) s strukturiranimi zahtevami in pretakanjem novih podatkov, ki se obdelujejo prek cevovodov ETL (slika 7.3-9), zmanjša potrebo po ročnem vključevanju v postopek potrjevanja (vsak postopek potrjevanja in zahteve glede podatkov je bil obravnavan v prejšnjih poglavjih).



Slika 7.3-9 Avtomatizacija potrjevanja podatkov s pomočjo ETL poenostavi vodenje gradbenih projektov, saj pospeši procese.

Tradicionalno lahko potrjevanje modelov, ki jih zagotovijo izvajalci in strokovnjaki CAD (BIM), traja od

nekaj dni do nekaj tednov. Z uvedbo avtomatiziranih postopkov ETL pa je to mogoče skrajšati na nekaj minut. V tipičnem primeru izvajalec navede: "*Model je potrjen in skladen.*" S to izjavo se začne veriga preverjanja izvajalčeve trditve o kakovosti podatkov:

- ☛ Vodja projekta - "*Izvajalec trdi: 'Model je bil preizkušen, vse je v redu.'*"
- ☛ Upravljavec podatkov - Preverjanje obremenitve:
  - Enostavna skripta v programu Pandas zazna kršitev v nekaj sekundah. Avtomatizacija odpravlja spore:
    - Kategorija: OST\_StructuralColumns, Parameter: FireRating IS NULL.
    - Ustvarite seznam ID-jev kršitev → izvozite v Excel/PDF.

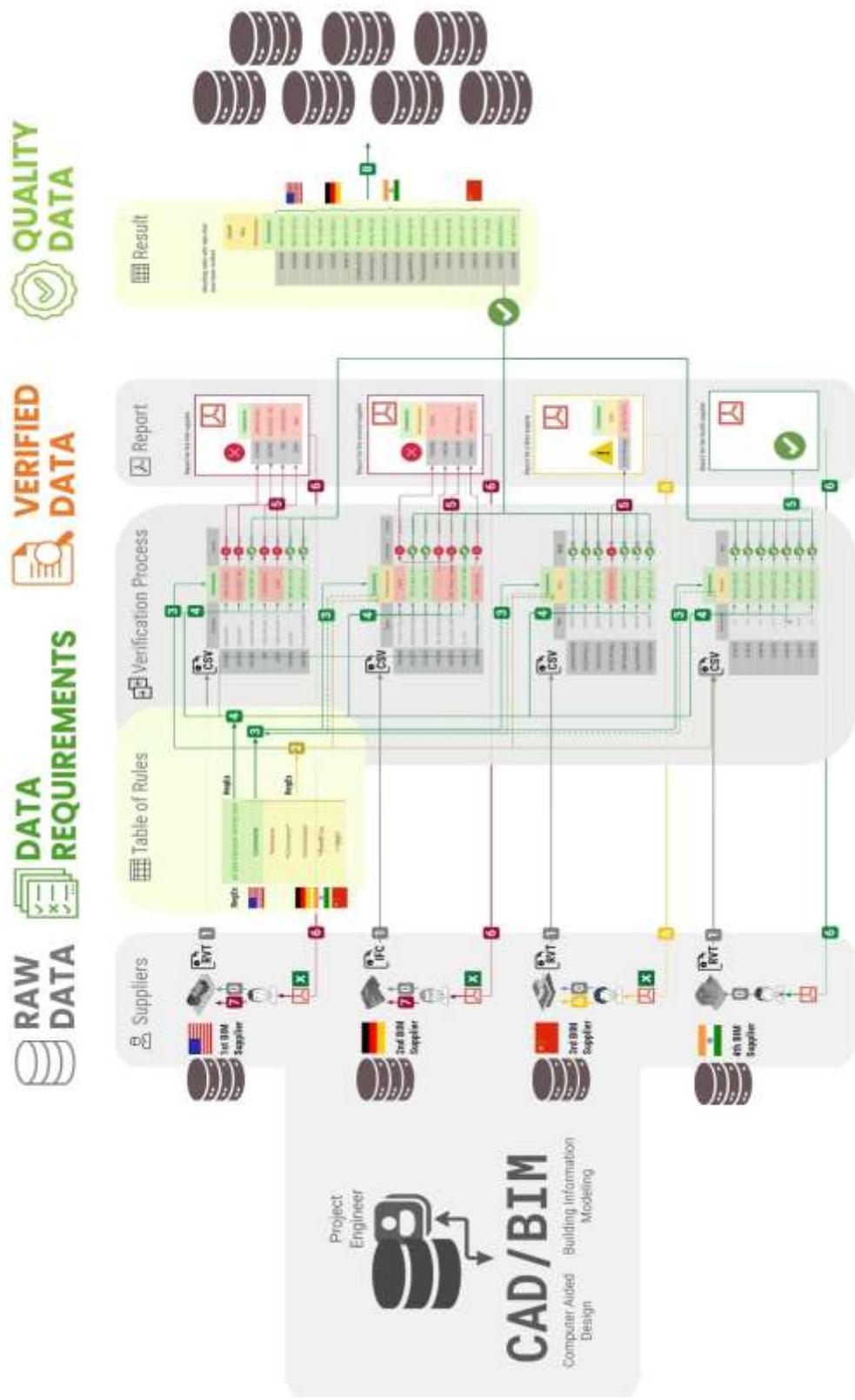
Enostavna skripta v programu Pandas odkrije kršitev v nekaj sekundah:

```
df = model_data[model_data["Kategorija"] == "OST_StructuralColumns"] # Filtriranje
issues = df[df["FireRating"].isnull()] # Prazne vrednosti
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Izvozi ID
```

- ☛ Vodja podatkov vodi projekta - "*Preverjanje spletne strani je pokazalo, da 18 stolpcev nima izpolnjenega parametra FireRating.*"
- ☛ Vodja projekta izvajalcu - "*Model se vrne v revizijo: parameter FireRating je obvezen, brez njega prevzem ni mogoč*"

Tako model CAD ni potrjen, avtomatizacija odpravlja spore, izvajalec pa skoraj takoj prejme strukturirano poročilo s seznamom identifikatorjev problematičnih elementov. Na ta način postane postopek potrjevanja pregleden, ponovljiv in zaščiten pred človeškimi napakami (slika 7.3-10).

S tem pristopom postane postopek potrjevanja podatkov inženirska funkcija in ne ročni nadzor kakovosti. S tem se ne le poveča produktivnost, temveč je mogoče isto logiko uporabiti pri vseh projektih podjetja, kar omogoča celovito digitalno preobrazbo procesov, od načrtovanja do delovanja.



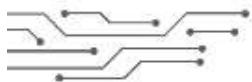
Slika 7.3-10 Avtomatizacija preverjanja atributov elementov odpravlja človeške napake in zmanjšuje verjetnost napak.

Z uporabo avtomatiziranih cevovodov (slika 7.3-10) lahko uporabniki sistema, ki pričakujejo kakovostne podatke iz sistemov CAD- (BIM-), takoj dobijo izhodne podatke, ki jih potrebujejo - tabele, dokumente, slike - in jih hitro vključijo v svoje delovne naloge.

Avtomatizacija nadzora, obdelave in analize povzroča spremembe v načinu upravljanja gradbenih projektov, zlasti interoperabilnost različnih sistemov brez uporabe zapletenih in dragih modularnih lastniških sistemov ali rešitev zaprtih dobaviteljev.

Medtem ko koncepti in marketinške kratice prihajajo in odhajajo, bodo sami postopki potrjevanja podatkovnih zahtev za vedno ostali sestavni del poslovnih procesov. Namesto oblikovanja vedno bolj specializiranih formatov in standardov bi se morala gradbena industrija posvetiti orodjem, ki so se že izkazala za učinkovita v drugih panogah. Danes obstajajo zmogljive platforme za avtomatizacijo obdelave podatkov in integracijo procesov, ki podjetjem omogočajo, da bistveno skrajšajo čas, porabljen za rutinske operacije, in zmanjšajo število napak pri ekstrakciji, transformaciji in nalaganju.

Eden od priljubljenih primerov rešitev za avtomatizacijo in orkestracijo procesov ETL je Apache Airflow, ki omogoča organizacijo zapletenih računskih procesov in upravljanje cevovodov ETL. Poleg Airflowa se aktivno uporabljajo tudi druge podobne rešitve, kot sta Apache NiFi za usmerjanje in pretakanje podatkov ter n8n za avtomatizacijo poslovnih procesov.



## POGLAVJE 7.4.

# ORKESTRACIJA ETL IN DELOVNIH TOKOV: PRAKTIČNE REŠITVE

## DAG in Apache Airflow: avtomatizacija in orkestracija poteka dela

Apache Airflow je brezplačna in odprtakodna platforma, zasnovana za avtomatizacijo, orkestriranje in spremljanje delovnih tokov (ETL -prevozniki).

Delo z velikimi količinami podatkov je potrebno vsak dan:

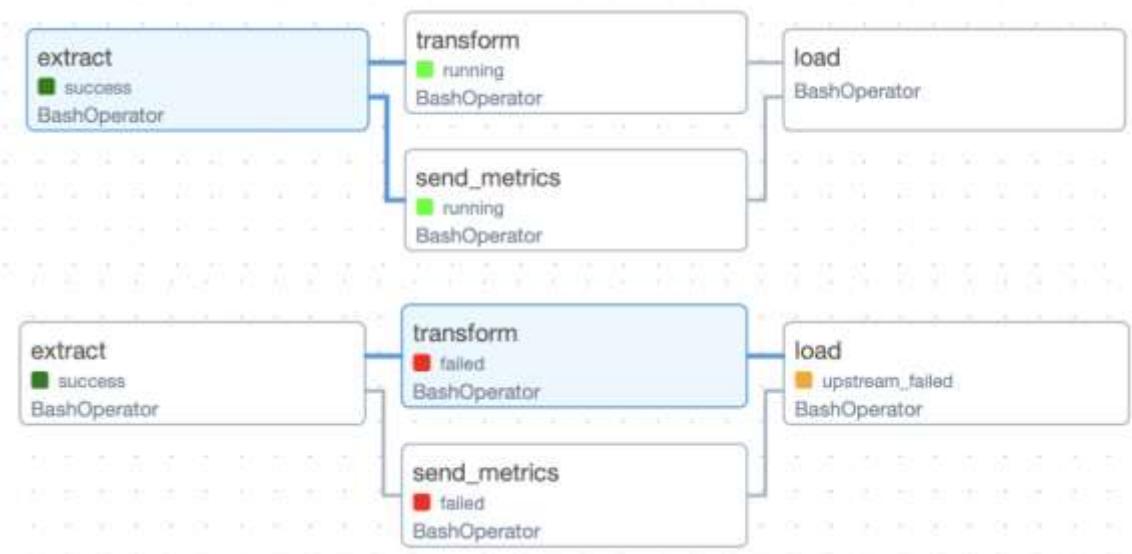
- Prenos datotek iz različnih virov - Izvleček (na primer od dobaviteljev ali strank).
- preoblikovanje teh podatkov v zahtevano obliko - preoblikovanje (strukturiranje, čiščenje in potrjevanje).
- pošiljanje rezultatov v preverjanje in ustvarjanje poročil - nalaganje (nalaganje v zahtevane sisteme, dokumente, zbirke podatkov ali nadzorne plošče).

Ročno izvajanje takšnih postopkov ETL zahteva veliko časa in povzroča tveganje človeških napak. Sprememba vira podatkov ali napaka v enem od korakov lahko povzroči zamude in napačne rezultate.

Orodja za avtomatizacijo, kot je Apache Airflow, omogočajo izgradnjo zanesljivega ETL -prevoznika, zmanjšanje napak, skrajšanje časa obdelave in zagotavljanje pravilnosti podatkov na vsaki stopnji. V središču Apache Airflow je koncept DAG (Directed Acyclic Graph) - usmerjeni aciklični graf, v katerem je vsako opravilo (operator) povezano z drugimi odvisnostmi in se izvaja strogo v določenem zaporedju. DAG odpravlja cikle, kar zagotavlja logično in predvidljivo strukturo izvajanja nalog.

Airflow skrbi za orkestracijo - upravljanje odvisnosti med opravili, nadzor urnikov izvajanja, spremljanje stanja in samodejno odzivanje na napake. Ta pristop zmanjšuje ročne posege in zagotavlja zanesljivost celotnega procesa.

**Orkestrator opravil** je orodje ali sistem, namenjen upravljanju in nadzoru izvajanja opravil v kompleksnih računalniških in informacijskih okoljih. Olajša postopek uvajanja, avtomatizacije in upravljanja izvajanja opravil za izboljšanje zmogljivosti in optimizacijo virov.



Slika 7.4-1 Apache Airflow zagotavlja uporabniku prijazen vmesnik, v katerem lahko vizualizirate DAG -ETL, si ogledate dnevnike izvajanja, stanje zagona opravil in drugo.

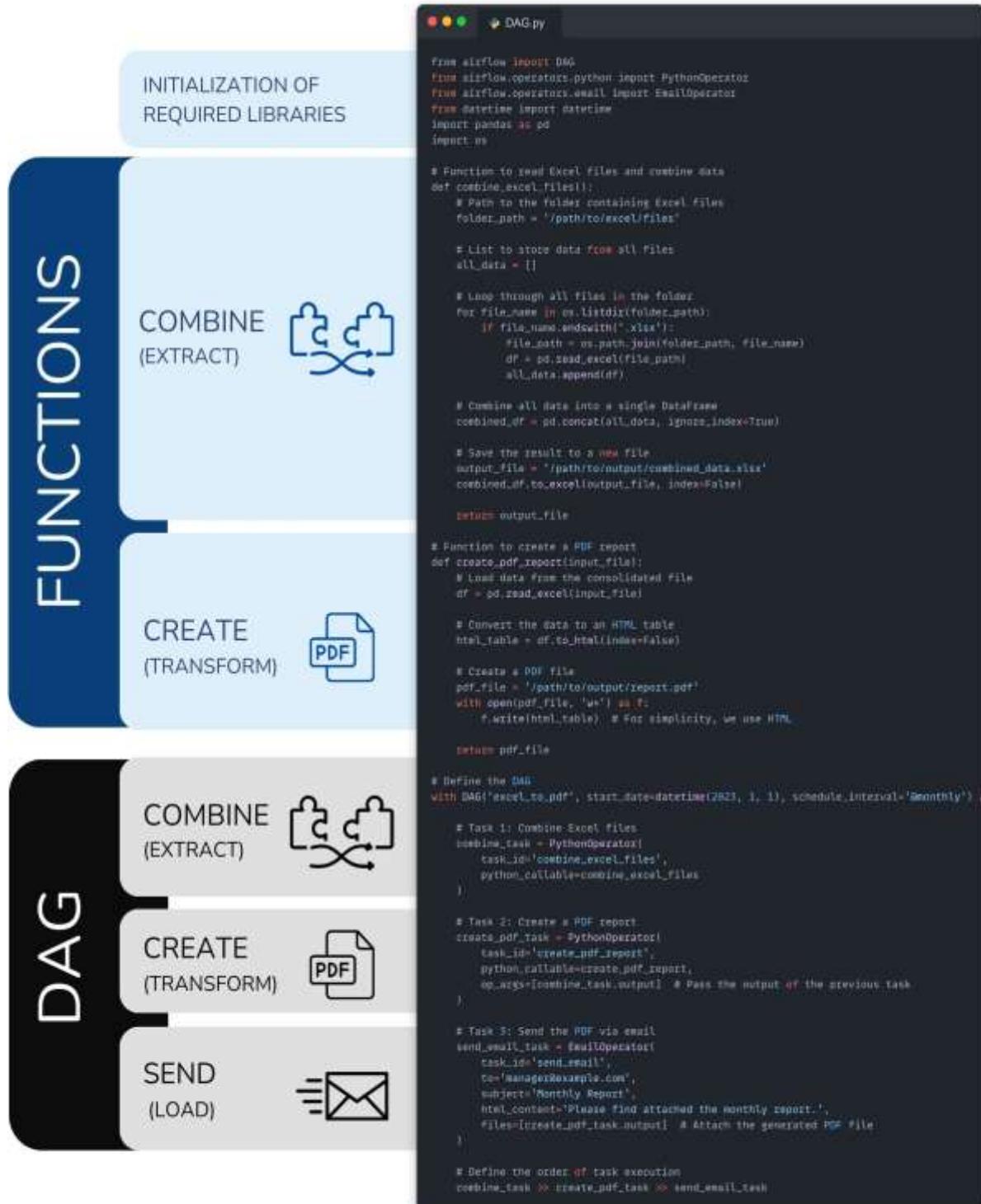
Airflow se pogosto uporablja za orkestracijo in avtomatizacijo porazdeljenega računalništva, obdelavo podatkov, upravljanje procesov ETL (Extract, Transform, Load), razporejanje nalog in druge podatkovne scenarije. Privzeto Apache Airflow kot podatkovno zbirko uporablja SQLite.

Primer enostavne skupine DAG, podobne skupini ETL, je sestavljen iz nalog - Izvleček, Preoblikovanje in Nalaganje. V grafu, ki se upravlja prek uporabniškega vmesnika (slika 7.4-1), je določen vrstni red izvajanja opravil (delov kode): na primer najprej se izvede izvleček, nato preoblikovanje (in pošiljanje\_metrike), delo pa zaključi opravilo nalaganje. Ko so opravljena vsa opravila, se šteje, da je postopek nalaganja podatkov uspešen.

## Apache Airflow: praktična uporaba pri avtomatizaciji ETL

Apache Airflow se pogosto uporablja za organizacijo zapletenih postopkov obdelave podatkov, kar omogoča gradnjo prilagodljivih ETL -prevoznikov. Apache Airflow je mogoče zagnati prek spletnega vmesnika ali programsko prek kode Python (slika 7.4-2). V spletnem vmesniku (slika 7.4-3) lahko skrbniki in razvijalci vizualno spremljajo skupine DAG, izvajajo naloge in analizirajo rezultate izvajanja.

Z uporabo DAG lahko določite jasno zaporedje opravil, upravljate odvisnosti med njimi in se samodejno odzivate na spremembe v izvornih podatkih. Oglejmo si primer uporabe sistema Airflow za avtomatizacijo obdelave poročil (slika 7.4-2).



Slika 7.4-2 Koncept ETL -prevoznik za obdelavo podatkov z Apache Airflow.

V tem primeru (slika 7.4-2) je obravnavana skupina DAG, ki opravlja ključne naloge v okviru ETL -

prevoznika:

**■ Preberite datoteke Excel (Izvleček):**

- Zaporedni pregled vseh datotek v danem imeniku.
- Preberi podatke iz vsake datoteke z uporabo knjižnice pandas.
- Združitev vseh podatkov v en sam podatkovni okvir.

**■ Ustvarite dokument PDF (Transform):**

- Preoblikovanje združenega okvira podatkov v tabelo HTML.
- Tabelo shranite kot PDF (v demo različici - prek HTML).

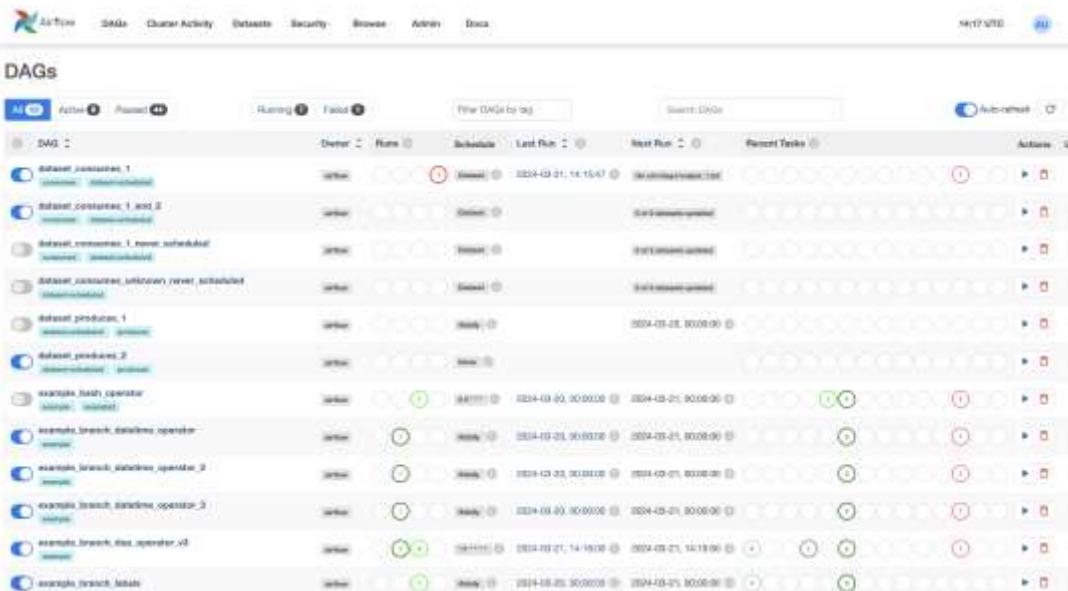
**■ Pošiljanje poročila po elektronski pošti (Load):**

- Uporabite EmailOperator za pošiljanje dokumenta PDF po e-pošti.

**■ Konfiguracija skupine DAG:**

- Opredelitev zaporedja opravil: pridobivanje podatkov → generiranje poročila → pošiljanje.
- Dodelitev urnika zagona (@monthly - prvi dan vsakega meseca).

Avtomatizirani primer ETL (slika 7.4-2) prikazuje, kako zbrati podatke iz datotek Excel, ustvariti dokument PDF in ga poslati po e-pošti. To je le eden od številnih možnih primerov uporabe programa Airflow. Ta primer je mogoče prilagoditi katerikoli posebni nalogi za poenostavitev in avtomatizacijo obdelave podatkov.



Slika 7.4-3 Pregled vseh DAG v okolju z informacijami o zadnjih izvedbah.

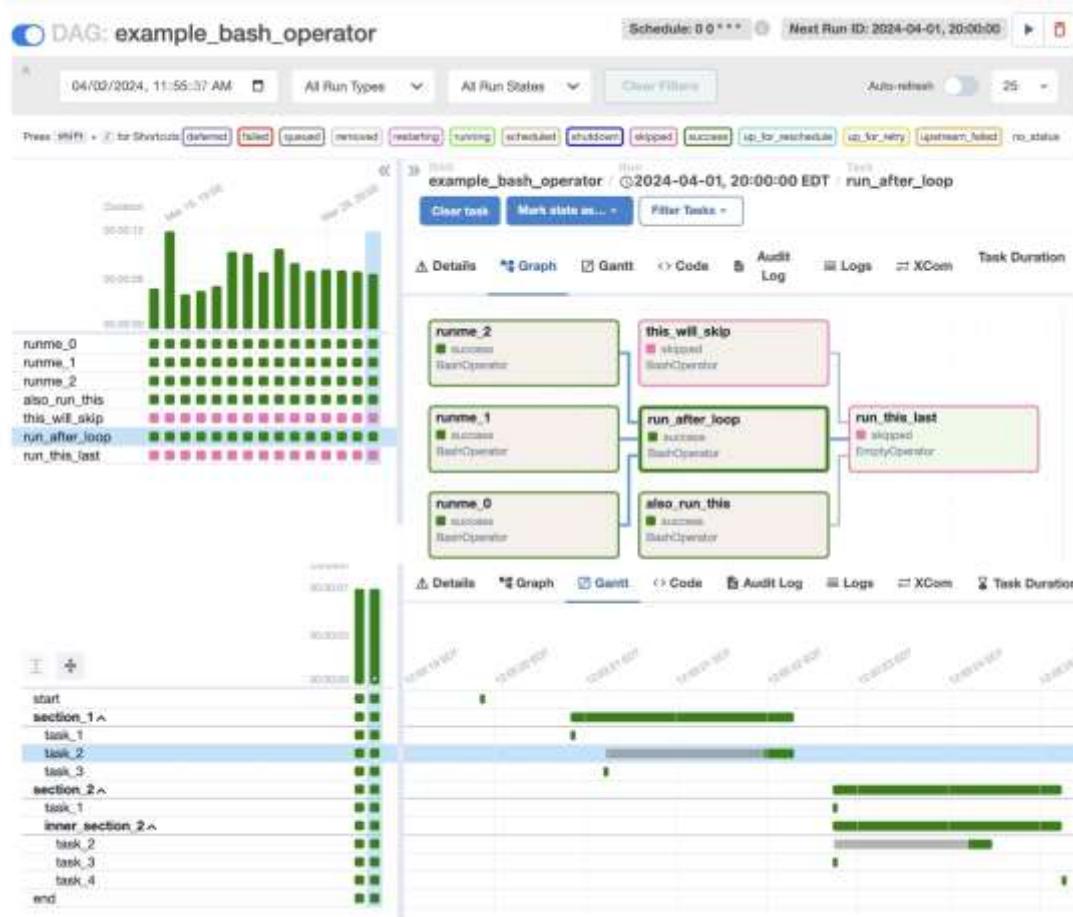
Spletni vmesnik Apache Airflow (slika 7.4-3) zagotavlja celovito vizualno okolje za upravljanje delovnih tokov podatkov. Prikazuje DAG kot interaktivne grafe, kjer vozlišča predstavljajo opravila, robovi pa odvisnosti med njimi, kar olajša spremeljanje zapletenih podatkovnih delovnih tokov. Vmesnik vključuje nadzorno ploščo z informacijami o stanju izvajanja nalog, zgodovini izvajanja, podrobnih dnevnikih in metrikah zmogljivosti. Skrbniki lahko ročno zaženejo opravila, ponovno zaženejo neuspešne operacije, začasno ustavijo skupine DAG in konfigurirajo okolske spremenljivke,

vse to prek intuitivnega uporabniškega vmesnika.

Takšno arhitekturo je mogoče dopolniti s potrjevanjem podatkov, obvestili o stanju izvajanja, integracijo z zunanjimi API-ji ali podatkovnimi zbirkami. Airflow omogoča prilagodljivo prilagajanje DAG: dodajanje novih nalog, spremjanje njihovega vrstnega reda, združevanje verig - zaradi česar je učinkovito orodje za avtomatizacijo zapletenih procesov obdelave podatkov. Pri izvajanju DAG v spletnem vmesniku Airflow (slika 7.4-3, slika 7.4-4) lahko spremljate stanje nalog. Sistem uporablja barvno indikacijo:

- Zelena - naloga je bila uspešno opravljena.
- Rumena - postopek je v teku.
- Rdeča - napaka med izvajanjem naloge.

V primeru napak (npr. manjkajoča datoteka ali poškodovana podatkovna struktura) sistem samodejno pošlje obvestilo.



Slika 7.4-4 Apache Airflow močno poenostavi diagnosticiranje težav, optimizacijo procesov in skupinsko sodelovanje pri zapletenih cevovodih za obdelavo podatkov.

Apache Airflow je priročen, ker avtomatizira rutinska opravila in jih ni treba opravljati ročno. Zanesljivost zagotavlja s spremeljanjem izvajanja procesov in takojšnjim obveščanjem o napakah. Prilagodljivost sistema omogoča enostavno dodajanje novih nalog ali sprememjanje obstoječih, s čimer se delovni tokovi prilagodijo spreminjačim se zahtevam.

Poleg Apache Airflow obstajajo podobna orodja za orkestriranje delovnih tokov. Na primer odprtokodno in brezplačno orodje Prefect (slika 7.3-5) ponuja preprosteho sintakso in se bolje povezuje s Pythonom, Luigi, ki ga je razvil Spotify, pa zagotavlja podobno funkcionalnost in dobro deluje z velikimi podatki. Omeniti velja tudi Kronos in Dagster, ki ponujata sodobne pristope k izgradnji cevovodov s poudarkom na modularnosti in skalabilnosti. Izberite orodje za orkestracijo opravil je odvisna od specifičnih potreb projekta, vendar vsa pomagajo avtomatizirati zapletene podatkovne procese ETL.

Posebej velja omeniti Apache NiFi, odprtokodno platformo, ki je zasnovana za pretakanje in usmerjanje podatkov. Za razliko od Airflow, ki se osredotoča na paketno obdelavo in upravljanje odvisnosti, se NiFi osredotoča na transformacijo podatkov v realnem času, ki poteka sproti, in prilagodljivo usmerjanje med sistemi.

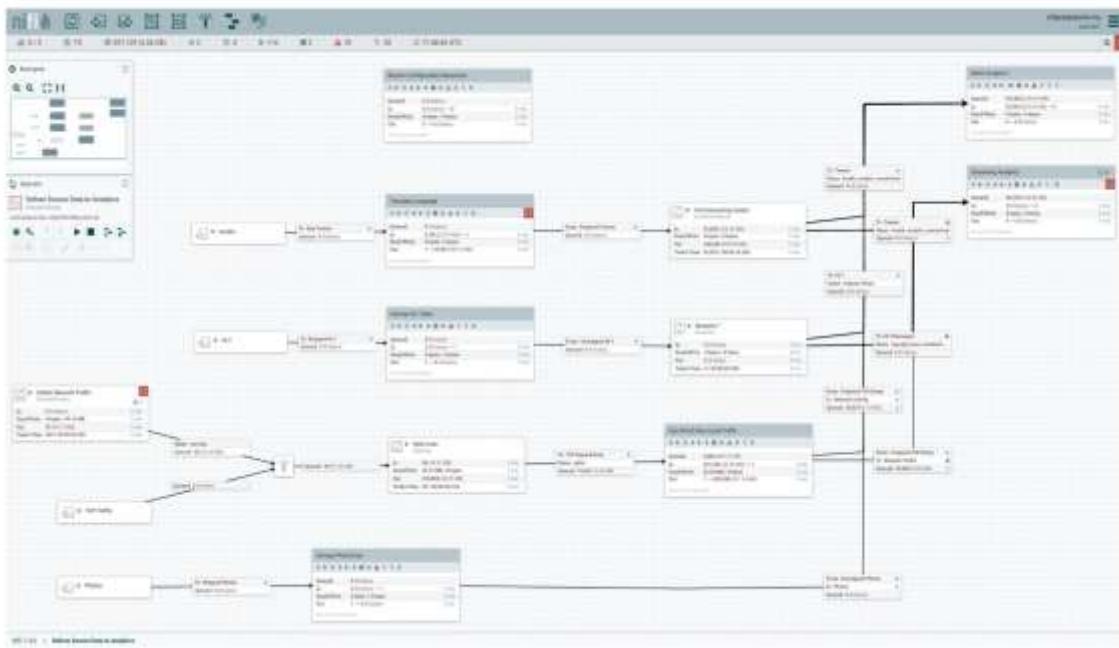
## Apache NiFi za usmerjanje in pretvorbo podatkov

Apache NiFi je zmogljiva odprtokodna platforma, namenjena avtomatizaciji podatkovnih tokov med različnimi sistemami. Prvotno jo je leta 2006 razvila ameriška Agencija za nacionalno varnost (NSA) pod imenom "Niagara Files" za notranjo uporabo. Leta 2014 je bil projekt odprt in prenesen na fundacijo Apache Software Foundation ter tako postal del njenih pobud za prenos tehnologije [141].

Apache NiFi je zasnovan za zbiranje, obdelavo in prenos podatkov v realnem času. Za razliko od sistema Airflow, ki deluje s paketnimi nalogami in zahteva natančno določene urnike, deluje NiFi v načinu pretočne obdelave, ki omogoča neprekinjen prenos podatkov med različnimi storitvami.

Apache NiFi je idealen za integracijo z napravami IoT, gradbenimi senzorji, nadzornimi sistemmi in npr. za pretočno preverjanje formatov CAD na strežniku, kjer je potreben takojšen odziv na spremembe podatkov.

Z vgrajenimi orodji za filtriranje, preoblikovanje in usmerjanje NiFi omogoča standardizacijo podatkov (Transform) pred prenosom (Load) v sisteme za shranjevanje ali analitične sisteme. Ena njegovih glavnih prednosti je vgrajena varnostna podpora in nadzor dostopa, zaradi česar je zanesljiva rešitev za ravnanje z občutljivimi informacijami.



Slika 7.4-5 Grafični prikaz pretoka podatkov v vmesniku Apache NiFi.

Apache NiFi učinkovito opravlja naloge pretakanja, filtriranja in usmerjanja podatkov v realnem času. Idealen je za tehnično intenzivne scenarije, kjer sta pomembna stabilen prenos informacij med sistemi in visoka prepustnost.

Kadar pa je glavni cilj povezovanje različnih storitev, avtomatizacija rutinskih postopkov in hitra vzpostavitev delovnih tokov brez poglobljenega programerskega znanja, so iskane rešitve z nizkim vstopnim pragom in največjo prilagodljivostjo. Eno takih orodij je n8n - platforma razreda Low-Code /No-Code, osredotočena na avtomatizacijo poslovanja in vizualno orkestracijo procesov.

## n8n Low-Code, No-Code orkestracija procesov

**n8n** je odprtakodna nizkokodna / nekodna platforma za izdelavo avtomatiziranih delovnih tokov, ki jo odlikujejo enostavna uporaba, prilagodljivost in možnost hitre integracije s številnimi zunanjimi storitvami.

**No-Code** je metoda ustvarjanja digitalnih izdelkov brez pisanja kode. Vsi elementi procesa - od logike do vmesnika - se izvajajo izključno s pomočjo vizualnih orodij. Platforme No-Code so namenjene uporabnikom brez tehničnega predznanja in omogočajo hitro ustvarjanje avtomatizacij, obrazcev, integracij in spletnih aplikacij. Primer: uporabnik vzpostavi samodejno pošiljanje obvestil ali integracijo z Google Sheets prek vmesnika "povleci in spusti" brez znanja programiranja.

Z odprtakodnimi in lokalnimi možnostmi uvajanja n8n v procesih avtomatizacije in ustvarjanja

cevovodov ETL omogoča podjetjem popoln nadzor nad njihovimi podatki, hkrati pa zagotavlja varnost in neodvisnost od ponudnikov storitev v oblaku.

V nasprotju z Apache Airflow, ki je usmerjen v računske naloge s togo orkestracijo in zahteva znanje Pythona, n8n zagotavlja vizualni urejevalnik, ki omogoča pisanje skript brez znanja programskih jezikov (slika 7.4-6). Čeprav njegov vmesnik omogoča ustvarjanje avtomatiziranih procesov brez pisanja kode (No-Code), lahko uporabniki v bolj zapletenih scenarijih dodajo lastne funkcije JavaScript in Python -funkcije za razširitev zmogljivosti (Low-Code).

**Low-Code** je pristop k razvoju programske opreme, pri katerem se osnovna logika aplikacije ali procesa ustvari z uporabo grafičnega vmesnika in vizualnih elementov, programska koda pa se uporablja le za prilagajanje ali razširitev funkcionalnosti.

Platforme Low-Code omogočajo bistveno hitrejši razvoj rešitev, saj poleg programerjev vključujejo tudi poslovne uporabnike z osnovnimi tehničnimi znanji. Primer: uporabnik lahko zgradi poslovni proces iz pripravljenih blokov in po potrebi doda lastno skripto v jeziku JavaScript ali Python.

Čeprav je n8n postavljen kot platforma z nizkim vstopnim pragom, so potrebni osnovno znanje programiranja, razumevanje spletnih tehnologij in spremnosti pri delu z API. Prilagodljivost sistema omogoča, da ga je mogoče prilagoditi širokemu naboru nalog - od samodejne obdelave podatkov do integracije s sporočilniki, napravami IoT in storitvami v oblaku.

Glavne značilnosti in prednosti uporabe telefona n8n:

- **Odprtokodne** in lokalne možnosti namestitve zagotavljajo popoln nadzor nad podatki, skladnost z varnostjo in neodvisnost od ponudnikov storitev v oblaku.
- **Integracija z več kot 330 storitvami**, vključno s storitvami CRM, ERP, e-trgovino, platformami v oblaku, sporočilniki in zbirkami podatkov.
- **Prilagodljivost scenarijev**: od preprostih obvestil do zapletenih verig z obdelavo zahtevkov API, logiko odločanja in povezavo storitev umetne inteligence.
- **Podpora za JavaScript in Python**: uporabniki lahko po potrebi vgradijo kodo po meri in tako razširijo možnosti avtomatizacije.
- **Intuitiven vizualni vmesnik**: omogoča hitro konfiguriranje in vizualizacijo vseh korakov procesa.

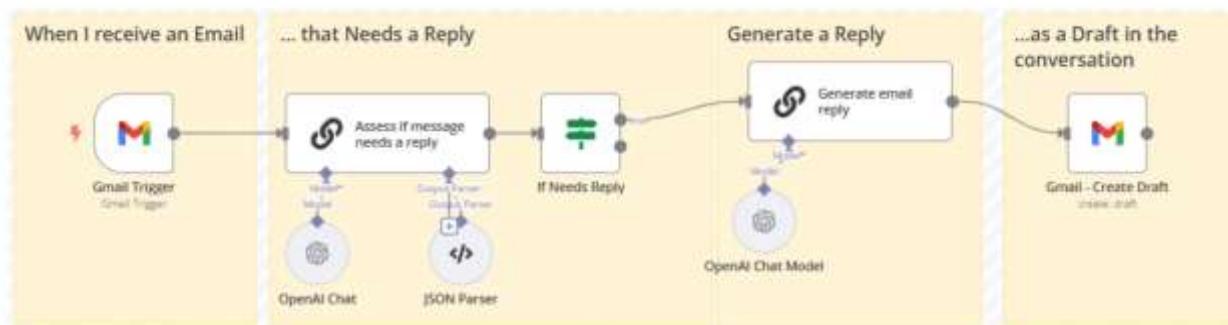
Platforme razreda Low-Code zagotavljajo orodja za ustvarjanje digitalnih rešitev z minimalno kodo, zato so idealne za ekipe, ki nimajo globokega tehničnega znanja, vendar morajo avtomatizirati procese.

V gradbeništvu se lahko n8n uporablja za avtomatizacijo različnih procesov, kot so povezovanje s sistemi za vodenje projektov, preverjanje tokov, pisanje izhodnih poročil in dopisov, samodejno posodabljanje podatkov o zalogah materiala, pošiljanje obvestil o stanju nalog ekipam in drugo. Prilagojeni cevovod v sistemu n8n lahko večkratno zmanjša število ročnih postopkov, zmanjša verjetnost napak in pospeši sprejemanje odločitev za izvedbo projekta.

Izbirate lahko med skoraj dva tisoč pripravljenimi, brezplačnimi in odprtakodnimi n8n Pipeline, ki so na voljo na: [n8n.io/workflows](https://n8n.io/workflows), za avtomatizacijo gradbenih delovnih procesov in osebnih opravil ter zmanjšanje rutinskih operacij.

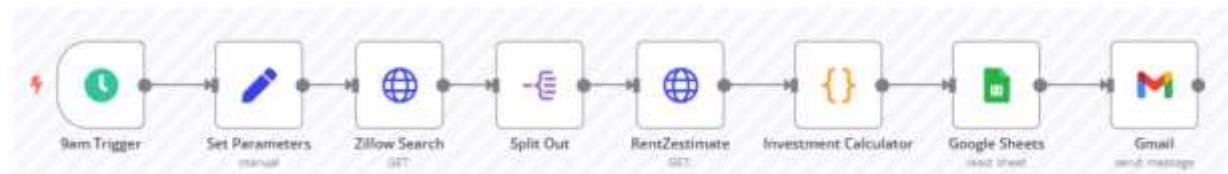
Vzemite eno od pripravljenih predlog Pipeline, ki je brezplačno na voljo na [n8n.io](https://n8n.io) [142] in samodejno ustvarja osnutke odgovorov v storitvi Gmail (slika 7.4-6) ter tako pomaga uporabnikom, ki prejemajo veliko število e-poštnih sporočil ali imajo težave pri sestavljanju odgovorov.

Ta predloga n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" (Slika 7.4-6) analizira prejeta e-poštna sporočila z uporabo LLM iz ChatGPT, določi, ali je potreben odgovor, ustvari osnutek iz ChatGPT in pretvori besedilo v HTML ter ga doda v verigo sporočil v Gmail. Pri tem se e-poštno sporočilo ne pošlje samodejno, saj lahko odgovor ročno uredite in odobrite. Nastavitev traja približno 10 minut in vključuje konfiguracijo OAuth za Gmail API in integracijo OpenAI API. Rezultat je priročna in brezplačna rešitev za avtomatizacijo rutinske e-poštne komunikacije brez izgube nadzora nad vsebino e-pošte.



Slika 7.4-6 Avtomatiziran postopek ustvarjanja e-poštnega odgovora z uporabo n8n.

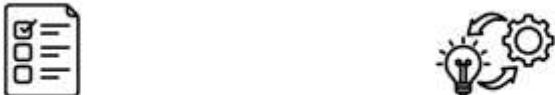
Drug primer avtomatizacije z n8n je iskanje ugodnih ponudb na nepremičinskem trgu [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", dnevno zbirja ustrezne ponudbe, ki ustrezajo danim merilom, z uporabo Zillow API. Samodejno izračuna ključne naložbene metrike (Cash on Cash ROI, Monthly Cash Flow, Down Payment), posodobi Google Sheets in pošlje zbirno poročilo po e-pošti (slika 7.4-7), kar vlagateljem omogoča prihranek časa in hiter odziv na najboljše ponudbe.



Slika 7.4-7 Avtomatiziran postopek za ocenjevanje investicijske privlačnosti nepremičnin.

Zaradi svoje prilagodljivosti in razširljivosti je n8n dragoceno orodje za podjetja, ki se želijo digitalno preoblikovati in postati bolj konkurenčna na trgu z razmeroma preprostimi in brezplačnimi odprtakodnimi orodji.

Orodja, kot so Apache NiFi, Airflow in n8n, si lahko predstavljamo kot tri plasti obdelave podatkov (slika 7.4-8). NiFi upravlja pretok podatkov ter zagotavlja njihovo dostavo in preoblikovanje, Airflow orkestrira izvajanje nalog z združevanjem podatkov v obdelovalne cevovode, n8n pa avtomatizira integracijo z zunanjimi storitvami in upravlja poslovno logiko.



	The main task	Approach
<b>Apache NiFi</b>	Streaming and data transformation	Real-time stream processing
<b>Apache Airflow</b>	Task orchestration, ETL pipelines	Batch planning, DAG processes
<b>n8n</b>	Integration, automation of business logic	Low-code visual orchestration

Slika 7.4-8 Apache Airflow, Apache NiFi in n8n lahko obravnavamo kot tri dopolnjujoče se plasti sodobne arhitekture za upravljanje podatkov.

Ta brezplačna in odprtakodna orodja skupaj predstavljajo primer učinkovitega ekosistema za upravljanje podatkov in procesov v gradbeništvu, ki podjetjem omogoča uporabo informacij za sprejemanje odločitev in automatizacijo procesov.

### Naslednji koraki: prehod z ročnih postopkov na rešitve, ki temeljijo na analitiki

Današnja gradbena podjetja delujejo v okolju velike negotovosti: spremenljive cene materialov, zamude pri dobavi, pomanjkanje delovne sile in kratki roki projektov. Uporaba analitičnih nadzornih plošč, prenosnikov ETL in sistemov BI podjetjem pomaga hitro prepoznati problematična področja, oceniti učinkovitost virov in predvideti spremembe, preden te privedejo do finančnih izgub.

Za povzetek tega dela velja izpostaviti glavne praktične korake, ki vam bodo pomagali pri uporabi obravnavanih tehnologij pri vsakodnevnih opravilih:

- izvajanje vizualizacij podatkov in analitičnih nadzornih plošč
  - Obvladajte proces ustvarjanja nadzornih plošč za spremljanje ključnih kazalnikov uspešnosti (KPI).
  - Uporaba orodij za vizualizacijo podatkov (Power BI, Tableau, Matplotlib, Plotly)
- Avtomatizacija obdelave podatkov s postopki ETL.
  - Nastavite samodejno zbiranje podatkov iz različnih virov (dokumentacija, tabele, CAD) s postopki ETL.
  - organiziranje preoblikovanja podatkov (npr. preverjanje regularnih izrazov ali izračun) z

uporabo skript Python.

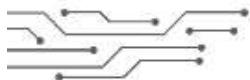
- S knjižnico FPDF poskusite nastaviti samodejno poročanje v PDF (ali DOC), pri čemer uporabite podatke iz datotek Excel ali pridobite informacije iz drugih dokumentov PDF.

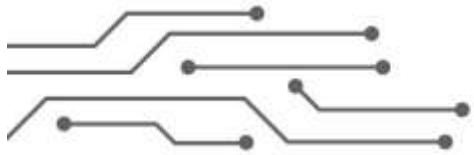
#### ■ Uporaba jezikovnih modelov (LLM) za avtomatizacijo

- uporaba obsežnih jezikovnih modelov (LLM) za ustvarjanje kode za pomoč pri pridobivanju in analiziraju podatkov iz nestrukturiranih dokumentov.
- Spoznajte orodje za avtomatizacijo n8n ter raziščite pripravljene predloge in študije primerov na njihovem spletnem mestu. Ugotovite, katere procese pri svojem delu lahko popolnoma avtomatizirate z uporabo pristopa brez kode/nizko-kodnega pristopa.

Analitični pristop k avtomatizaciji podatkov in procesov ne le skrajšuje čas, porabljen za rutinske postopke, temveč tudi izboljšuje kakovost odločanja. Podjetja, ki uvedejo orodja za vizualno analitiko in ETL -prevozniki, dobijo priložnost, da se hitro odzovejo na spremembe

Avtomatizacija poslovnih procesov z orodji, kot so n8n, Airflow in NiFi, je le prvi korak k digitalni zrelosti. Naslednji korak je kakovostno shranjevanje in upravljanje samih podatkov, ki so podlaga za avtomatizacijo. V osmem delu si bomo podrobno ogledali, kako lahko gradbena podjetja zgradijo trajnostno arhitekturo za shranjevanje podatkov in preidejo od kaosa dokumentov in datotek različnih formatov k centraliziranim platformam za shranjevanje in analitiko.





## VIII DEL

# **SHRANJEVANJE IN UPRAVLJANJE PODATKOV V GRADBENIŠTVU**

8. del obravnavajo sodobne tehnologije za shranjevanje in upravljanje podatkov v gradbeništvu. Analizira učinkovite formate za obdelavo velikih količin informacij - od preprostih CSV in XLSX do zmogljivejših Apache Parquet in ORC s podrobno primerjavo njihovih zmogljivosti in omejitev. Obravnavani so koncepti podatkovnih skladišč (DWH), podatkovnih jezer) in njihovih hibridnih rešitev (Data Lakehouse) ter načela upravljanja podatkov) in podatkovnega minimalizma). Podrobno so obravnavane težave podatkovnega močvirja) in strategije za preprečevanje kaosa v informacijskih sistemih. Predstavljeni so novi pristopi k delu s podatki, vključno z vektorskimi podatkovnimi zbirkami in njihovo uporabo pri gradnji s pomočjo koncepta Bounding Box. V tem delu sta obravnavani tudi metodologiji DataOps in VectorOps kot nova standarda za organizacijo delovnih postopkov s podatki.

## POGLAVJE 8.1.

### PODATKOVNA INFRASTRUKTURA: OD OBLIK SHRANJEVANJA DO DIGITALNIH REPOZITORIJEV

#### **Podatkovni atomi: temelj učinkovitega upravljanja informacij**

Vse v vesolju je sestavljeno iz najmanjših gradnikov - atomov in molekul, vse živo in neživo pa se sčasoma neizogibno vrne v to začetno stanje. V naravi ta proces poteka z osupljivo hitrostjo, ki jo poskušamo prenesti na procese, ki jih nadzoruje človek.

V gozdu se vsi živi organizmi sčasoma spremenijo v hranljivo snov, ki je osnova za nove rastline. Te rastline pa postanejo hrana za nova živa bitja, sestavljena iz istih atomov, ki so pred milijoni let ustvarili vesolje.

Tudi v poslovnem svetu je pomembno, da zapletene, večplastne strukture razčlenimo na najbolj temeljne, minimalno obdelane enote podobno kot atomi in molekule v naravi. To omogoča učinkovito shranjevanje in upravljanje atomov podatkov, ki se spremenijo v bogato, plodno podlago, ki postane ključni vir za rast analitike in kakovost odločanja.



Slika 8.1-1 Analiza in odločanje temeljita na ponovno uporabljenih podatkih, ki so bili nekoč obdelani in shranjeni.

Glasbene kompozicije so sestavljene iz not, ki se združujejo v kompleksne glasbene skladbe, medtem ko besede nastanejo iz osnovne enote, črke in zvoka. Naj gre za naravo, znanost, ekonomijo, umetnost

ali tehnologijo, svet kaže izjemno enotnost in harmonijo v svojem prizadevanju za uničenje, strukturo, kolesarjenje in ustvarjanje. Podobno so procesi v sistemih za izračun stroškov razdeljeni na majhne strukturirane enote - postavke virov - na ravni stroškovnih kalkulacij in razporedov. Te enote se nato, podobno kot zapiski, uporabljajo za oblikovanje kompleksnejših kalkulacij in razporedov. Enako načelo uporabljajo sistemi računalniško podprtega projektiranja, v katerih so kompleksni arhitekturni in inženirski projekti zgrajeni iz osnovnih elementov - posameznih elementov in knjižničnih komponent, iz katerih se ustvari celoten 3D -model projekta kompleksne stavbe ali strukture.

Koncept cikličnosti in strukture, ki je značilen za naravo in znanost, se odraža tudi v sodobnem svetu podatkov. Tako kot se v naravi vsa živa bitja vračajo k atomom in molekulam, se tudi v svetu sodobnih orodij za obdelavo podatkov informacije ponavadi vračajo k svoji najbolj primitivni obliki.

Najmanjši elementi s svojo končno nedeljivostjo so osnovni gradniki poslovnih procesov. Že na začetku je treba skrbno razmisljiti, kako te drobne gradnike iz različnih virov zbrati, strukturirati (razdeliti na atome) in shraniti. Pri organizaciji in shranjevanju podatkov ne gre le za njihovo razčlenitev na sestavne dele. Enako pomembno je zagotoviti, da so integrirani in shranjeni na strukturiran način, tako da je podatke mogoče zlahka priklicati, analizirati in uporabiti za odločanje, kadar koli so potrebni.

Za učinkovito obdelavo informacij je treba skrbno izbrati obliko in način shranjevanja podatkov - tako kot je treba pripraviti tla za rast dreves. Hraničnice podatkov je treba organizirati tako, da se zagotovita visoka kakovost in ustreznost informacij, pri čemer je treba izločiti odvečne ali nepomembne podatke. Bolje ko je ta "informacijska zemlja" strukturirana, hitreje in natančneje lahko uporabniki najdejo prave podatke in rešujejo analitične probleme.

## **Shranjevanje informacij: datoteke ali podatki**

Podatkovna skladišča podjetjem omogočajo zbiranje in združevanje informacij iz različnih sistemov ter tako ustvarjajo enotno središče za nadaljnjo analitiko. Zbrani zgodovinski podatki omogočajo ne le poglobljeno analizo procesov, temveč tudi prepoznavanje vzorcev, ki lahko vplivajo na uspešnost poslovanja.

Recimo, da se podjetje ukvarja z več projekti hkrati. Inženir želi ugotoviti, koliko betona je bilo vgrajenega in koliko ga je treba še kupiti. Pri tradicionalnem pristopu bi moral ročno poiskati strežnik in odpreti več tabel s predračuni, jih primerjati s potrdili o opravljenih delih in preveriti trenutna stanja zalog. To traja več ur ali celo dni. Tudi s postopki ETL in samodejnjimi skriptami ostaja naloga napol ročna: inženir mora še vedno ročno določiti pot do map ali določenih datotek v strežniku. To zmanjšuje splošni učinek avtomatizacije, saj še naprej jemlje dragoceni delovni čas.

Pri prehodu na upravljanje podatkov inženir namesto dela z datotečnim sistemom strežnika dobí dostop do enotne strukture shranjevanja, kjer se informacije posodablja v realnem času. Z eno samo poizvedbo - v obliki kode, poizvedbe SQL ali celo klica LLM -agenta - je mogoče takoj pridobiti natančne podatke o trenutnih stanjih, opravljenem delu in prihajajočih dobavah, če so bili podatki vnaprej pripravljeni in združeni v podatkovnem skladišču, kjer ni treba brskati po mapah, odpirati na desetine datotek in ročno primerjati vrednosti.

Gradbena podjetja so dolgo uporabljala dokumente PDF, risbe DWG, modele RVT ter na stotine in tisoče tabel Excel in druge različne formate, ki so shranjeni v posebnih mapah na strežnikih podjetja, kar otežuje iskanje informacij, njihovo preverjanje in analizo. Zato se datoteke, ki ostanejo po zaključku projektov, najpogosteje prenesejo nazaj na strežnik v arhivske mape, ki se v prihodnosti praktično ne uporablajo. Takšna tradicionalna hramba podatkov na podlagi datotek izgublja pomen s povečevanjem pretoka podatkov, saj je ranljiva za človeške napake.

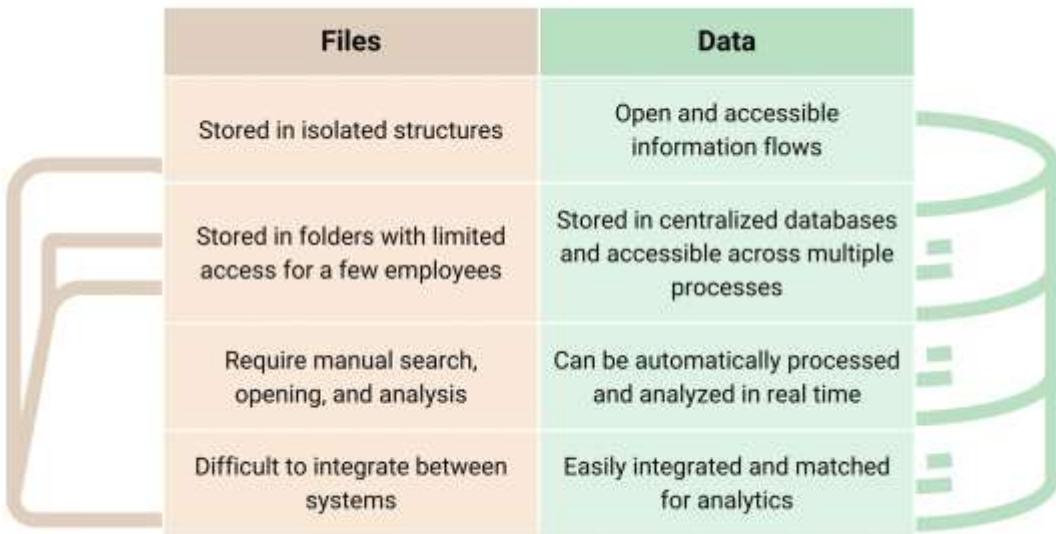
**Datoteka** je le izoliran vsebnik, v katerem so shranjeni podatki. Datoteke so ustvarjene za ljudi in ne za sisteme, zato jih je treba ročno odpirati, brati in razlagati. Primeri so Excelova -tabela, PDF -dokument ali CAD -črt, ki ga je treba posebej odpreti v določenem orodju, da bi prišli do želenih informacij. Brez strukturiranega prikaza in obdelave ostanejo informacije v njem neizkoriščene.

**Podatki** pa so strojno berljive informacije, ki se samodejno povezujejo, posodabljajo in analizirajo. V posameznem podatkovnem skladišču (npr. podatkovni bazi, DWH ali podatkovnem jezeru) so informacije predstavljene v obliki tabel, zapisov in povezav. To omogoča enotno shranjevanje, samodejne poizvedbe, vrednostne analize in poročanje v realnem času.

Uporaba podatkov namesto datotek (slika 8.1-1) omogoča odpravo postopka ročnega iskanja in poenotenje obdelave. Podjetja, ki so tak pristop že uvedla, pridobijo konkurenčno prednost zaradi hitrosti dostopa do informacij in možnosti hitrega vključevanja informacij v poslovne procese.

Prehod od uporabe datotek k uporabi podatkov je neizogibna sprememba, ki bo zaznamovala prihodnost gradbene industrije.

Vsako podjetje v gradbeništvu se bo moralo odločiti, ali bo še naprej hraniло informacije v različnih datotekah in silosih, ki jih morajo prebrati ljudje s posebnimi programi, ali pa jih bo v prvih fazah obdelave preoblikovalo v strukturirane podatke in ustvarilo enotno integrirano digitalno podlago za avtomatizirano upravljanje projektov.



**Slika 8.1-1 Razvoj pretoka informacij: od izoliranih datotek do integriranih podatkov.**

Zaradi eksplozije informacij so tradicionalne metode shranjevanja in obdelave datotek vse manj učinkovite. Tako v gradbeništvu kot v drugih panogah ni več dovolj, da se zanašamo na različne mape z datotekami v različnih formatih ali nepovezane podatkovne zbirke.

Podjetja, ki želijo ostati konkurenčna v digitalni dobi, se bodo neizogibno usmerila v integrirane digitalne platforme, uporabljala tehnologije velikih količin podatkov in avtomatizirane analitične sisteme.

Prehod od shranjevanja datotek k delovnim tokovom, ki temeljijo na podatkih, bo zahteval ponoven razmislek o pristopih k upravljanju informacij in zavestno izbiro formatov, primernih za nadaljnjo integracijo v centralizirane repozitorije. Od te izbire bo odvisno, kako učinkovito bo mogoče obdelovati podatke, kako hitro bo mogoče do njih dostopati in kako enostavno jih bo mogoče vključiti v digitalne procese podjetja.

## Shranjevanje velikih količin podatkov: analiza priljubljenih formatov in njihove učinkovitosti

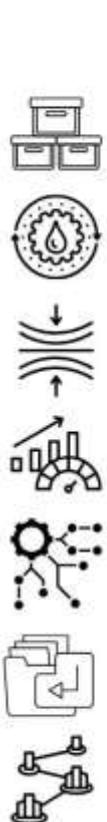
Formati za shranjevanje imajo ključno vlogo pri skalabilnosti, zanesljivosti in zmogljivosti analitične infrastrukture. Za analizo in obdelavo podatkov - kot so filtriranje, združevanje v skupine in agregiranje - smo v naših primerih uporabili Pandas DataFrame - priljubljeno strukturo za delo s podatki v pomnilniku RAM.

Vendar pa Pandas DataFrame nima lastnega formata za shranjevanje, zato se po končani obdelavi podatki izvozijo v enega od zunanjih formatov - najpogosteje CSV ali XLSX. Ti tabelarni formati so enostavni za izmenjavo in združljivi z večino zunanjih sistemov, vendar imajo številne omejitve: majhno učinkovitost shranjevanja, pomanjkanje stiskanja in slabo podporo različicam:

- **CSV** (Vrednosti, ločene z vejico): preprosta oblika besedila, ki jo podpirajo različne platforme in orodja. Je enostaven za uporabo, vendar ne podpira zapletenih podatkovnih vrst in stiskanja.
- **XLSX** (Excel Open XML Spreadsheet): format datoteke Microsoft Excel, ki podpira zahtevne funkcije, kot so formule, diagrami in oblikovanje. Čeprav je uporaben za ročno analizo in vizualizacijo podatkov, ni optimiziran za obsežno obdelavo podatkov.

Poleg priljubljenih tabelaričnih formatov XLSX in CSV obstaja več priljubljenih formatov za učinkovito shranjevanje strukturiranih podatkov (slika 8.1-2), od katerih ima vsak edinstvene prednosti glede na posebne zahteve za shranjevanje in analizo podatkov:

- **Apache Parquet**: datotečni format za shranjevanje podatkov v obliki stolpcev, optimiziran za uporabo v sistemih za analizo podatkov. Ponuja učinkovite sheme stiskanja in kodiranja podatkov, zato je idealen za kompleksne podatkovne strukture in obdelavo velikih količin podatkov.
- **Apache ORC** (Optimised Row Columnar): Podobno kot Parquet, ORC zagotavlja visoko stopnjo stiskanja in učinkovito shranjevanje podatkov. Optimiziran je za zahtevne operacije branja in je primeren za shranjevanje podatkovnih jezer.
- **JSON** (JavaScript Object Notation): čeprav JSON v primerjavi z binarnimi formati, kot sta Parquet ali ORC, ni tako učinkovit pri shranjevanju podatkov, je zelo dostopen in enostaven za delo, zato je idealen za skripte, kjer sta pomembni berljivost in spletna združljivost.
- **Feather**: hiter, lahek in enostaven za uporabo analitično usmerjen binarni stolpčni format za shranjevanje podatkov. Zasnovan je za učinkovit prenos podatkov med programoma Python (Pandas) in R, zato je odlična izbira za projekte, ki vključujejo ta programska okolja.
- **HDF5** (Hierarhični podatkovni format različice 5): zasnovan za shranjevanje in urejanje velikih količin podatkov. Podpira širok nabor podatkovnih tipov in je primeren za delo s kompleksnimi zbirkami podatkov. HDF5 je še posebej priljubljen v znanstvenem računalništvu zaradi svoje sposobnosti učinkovitega shranjevanja in dostopa do velikih zbirk podatkov.



	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)

Slika 8.1-2 Primerjava podatkovnih formatov, ki prikazuje glavne razlike v vidikih shranjevanja in obdelave.

Za primerjalno analizo formatov, ki se uporabljajo v fazi Load v postopku ETL, je bila izdelana tabela z velikostjo datotek in časom branja (slika 8.1-3). V študiji so bile uporabljene datoteke z enakimi podatki: tabela je vsebovala 10 000 vrstic in 10 stolpcev, napolnjenih z naključnimi vrednostmi.

V študijo so vključeni naslednji formati shranjevanja: CSV, Parquet, XLSX in HDF5 ter njihove stisnjene različice v arhivih ZIP. Neobdelani podatki so bili ustvarjeni z uporabo knjižnice NumPy in predstavljeni kot struktura Pandas DataFrame. Postopek testiranja je bil sestavljen iz naslednjih korakov:

- Shranjevanje datotek: podatkovni okvir je shranjen v štirih različnih formatih: CSV, Parquet, XLSX in HDF5. Vsak format ima edinstvene lastnosti pri načinu shranjevanja podatkov, kar vpliva na velikost datoteke in hitrost branja.
- Stiskanje datotek ZIP: za analizo učinkovitosti standardnega stiskanja je bila vsaka datoteka dodatno stisnjena v arhiv ZIP.
- Branje datotek (ETL - Load): čas branja je bil izmerjen za vsako datoteko po razpakirjanju iz ZIP. To omogoča oceno hitrosti dostopa do podatkov po izvleku iz arhiva.

Pomembno je omeniti, da se Pandas DataFrame ni uporabljal neposredno pri analizi velikosti ali časa branja, saj sam po sebi ne predstavlja oblike shranjevanja. Služil je le kot vmesna struktura za ustvarjanje in poznejše shranjevanje podatkov v različne formate.



Slika 8.1-3 Primerjava formatov za shranjevanje po velikosti in hitrosti branja.

Datoteke CSV in HDF5 kažejo (slika 8.1-3) visoko učinkovitost stiskanja, saj znatno zmanjšajo svojo velikost, ko so zapakirane v ZIP, kar je lahko še posebej uporabno v scenarijih, ki zahtevajo optimizacijo shranjevanja. Po drugi strani pa datotek XLSX praktično ni mogoče stisniti in njihova velikost v ZIP-u ostane primerljiva z izvirnikom, zaradi česar so manj primerne za uporabo pri velikih količinah podatkov ali v okoljih, kjer je pomembna hitrost dostopa do podatkov. Poleg tega je čas branja datotek XLSX v primerjavi z drugimi formati bistveno daljši, zaradi česar so manj primerne za hitre operacije branja podatkov. Apache Parquet je zaradi svoje stolpčne strukture pokazal visoko zmogljivost za analitične naloge in velike količine podatkov.

## Optimizacija shranjevanja z Apache Parquet

Eden od priljubljenih formatov za shranjevanje in obdelavo velikih količin podatkov je Apache Parquet. Ta format je zasnovan posebej za shranjevanje v obliki stolpcev (podobno kot Pandas), kar omogoča znatno zmanjšanje pomnilniškega odtisa in povečanje hitrosti analitičnih poizvedb. Za razliko od tradicionalnih formatov, kot sta CSV in XLSX, Parquet podpira nativno stiskanje in je optimiziran za sisteme velikih podatkov, vključno s Spark, Hadoop in shranjevanjem v oblaku.

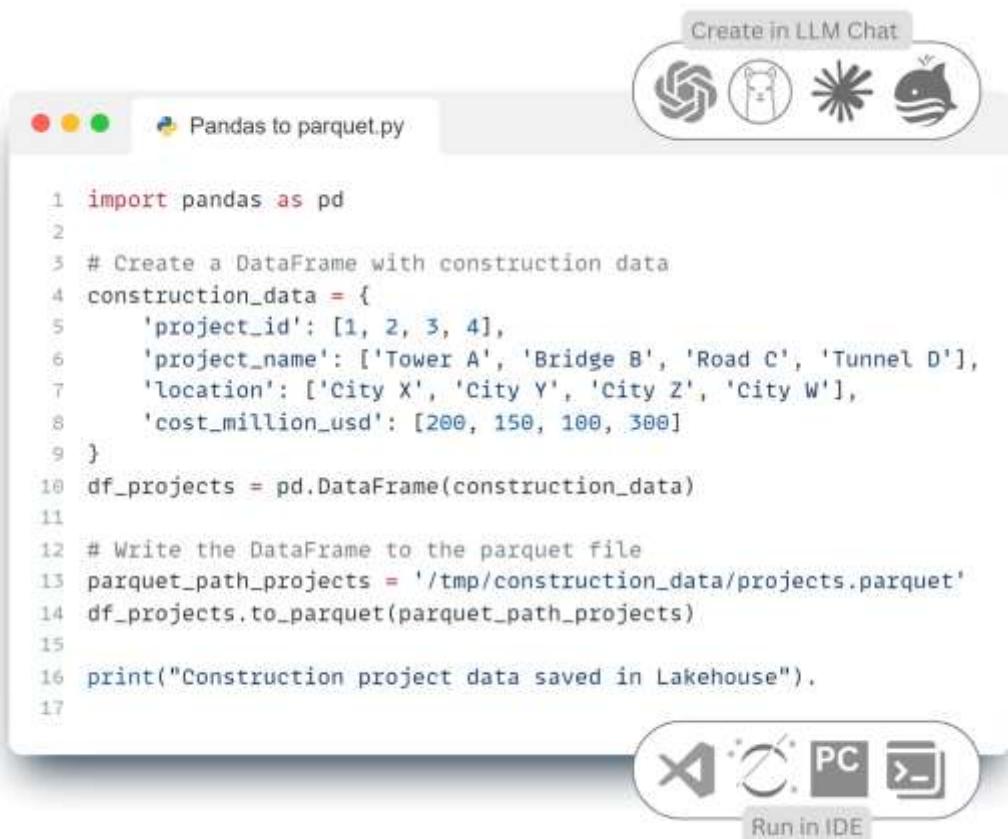
Ključne značilnosti sistema Parquet vključujejo podporo za stiskanje in kodiranje podatkov, kar znatno zmanjša velikost shrambe in pospeši operacije branja podatkov z neposrednim delom na želenih stolpcih in ne na vseh vrsticah podatkov.

Za nazoren primer, kako enostavno je pridobiti potrebno kodo za pretvorbo podatkov v Apache Parquet, uporabimo LLM.

- ❷ Pošljite besedilno zahtevo v klepet LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Napišite kodo za shranjevanje podatkov iz Pandas DataFrame v Apache Parquet. ↴

- ❸ Odgovor LLM :



The screenshot shows a user interface for generating code. At the top right is a button labeled "Create in LLM Chat". Below it is a row of icons representing different AI models: a blue circle, a cat face, a starburst, and a whale. The main area contains a code editor with the following Python script:

```

1 import pandas as pd
2
3 # Create a DataFrame with construction data
4 construction_data = {
5     'project_id': [1, 2, 3, 4],
6     'project_name': ['Tower A', 'Bridge B', 'Road C', 'Tunnel D'],
7     'location': ['City X', 'City Y', 'City Z', 'City W'],
8     'cost_million_usd': [200, 150, 100, 300]
9 }
10 df_projects = pd.DataFrame(construction_data)
11
12 # Write the DataFrame to the parquet file
13 parquet_path_projects = '/tmp/construction_data/projects.parquet'
14 df_projects.to_parquet(parquet_path_projects)
15
16 print("Construction project data saved in Lakehouse").
17

```

At the bottom right of the code editor is a "Run in IDE" button with icons for VS Code, PyCharm, and a terminal.

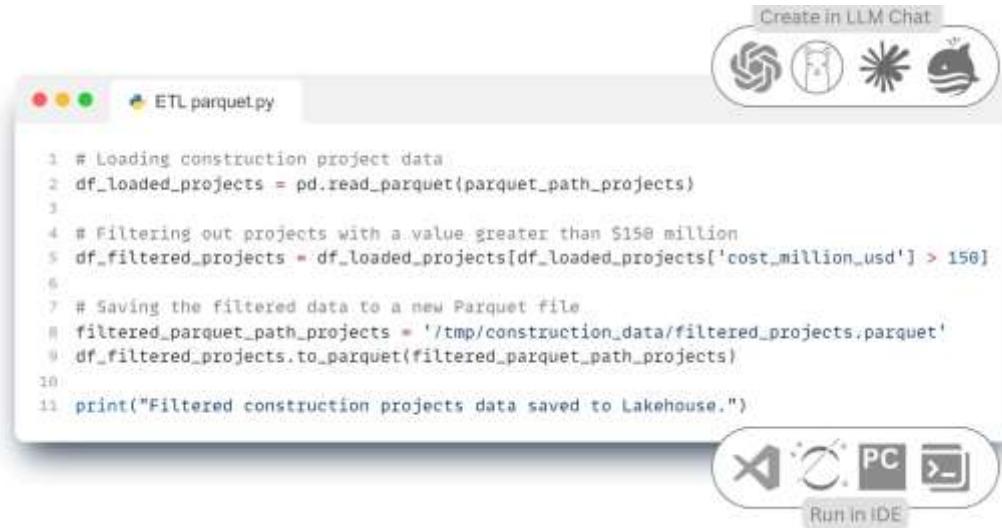
**Slika 8.1-4 Prenos podatkov podatkovnih okvirov iz RAM-a v pomnilniško učinkovit format Apache Parquet z nekaj vrsticami Pythona .**

Naslednji primer: simulirajmo postopek ETL s podatki, shranjenimi v formatu Parquet, za filtriranje projektov po določeni vrednosti enega od atributov "cost\_million\_usd" (slika 8.1-4).

- ❷ Ko se klepet nadaljuje, pošljite besedilno zahtevo na naslov LLM:

Napišite kodo, v kateri želimo filtrirati podatke v preglednici in shraniti samo tiste projekte (vrstice preglednice) iz podatkov Apache Parquet, katerih stroški (parameter cost\_million\_usd) presegajo 150 milijonov USD. ↴

💡 Odgovor LLM :



```

1 # Loading construction project data
2 df_loaded_projects = pd.read_parquet(parquet_path_projects)
3
4 # Filtering out projects with a value greater than $150 million
5 df_filtered_projects = df_loaded_projects[df_loaded_projects['cost_million_usd'] > 150]
6
7 # Saving the filtered data to a new Parquet file
8 filtered_parquet_path_projects = '/tmp/construction_data/filtered_projects.parquet'
9 df_filtered_projects.to_parquet(filtered_parquet_path_projects)
10
11 print("Filtered construction projects data saved to Lakehouse.")

```

The screenshot shows a Jupyter Notebook cell with the title "ETL parquet.py". The cell contains Python code for reading a Parquet file, filtering it to exclude projects costing more than \$150 million, and then saving the filtered data back to a new Parquet file. The notebook interface includes a toolbar with icons for "Create in LLM Chat", "Run in IDE", and other standard Jupyter functions.

Slika 8.1-5 Postopek ETL pri delu s podatki v formatu Apache Parquet je videti enako kot pri drugih strukturiranih formatih.

Uporaba formata Parquet (v primerjavi z XLSX, CSV itd.) znatno zmanjša količino shranjenih informacij in pospeši iskanje. Zaradi tega je odličen tako za shranjevanje kot za analizo podatkov. Parquet se povezuje z različnimi sistemi obdelave in zagotavlja učinkovit dostop v hibridnih arhitekturah.

Vendar je učinkovita oblika shranjevanja le en element celovite podatkovne izkušnje. Za vzpostavitev trajnostnega in razširljivega okolja je potrebna dobro zasnovana arhitektura za upravljanje podatkov. To funkcijo opravljajo sistemi razreda DWH (podatkovno skladišče). Zagotavljajo združevanje podatkov iz heterogenih virov, preglednost poslovnih procesov in možnost kompleksnih analiz z orodji BI in algoritmi strojnega učenja.

## DWH: podatkovno skladišče podatkovna skladišča

Tako kot je format Parquet optimiziran za učinkovito shranjevanje velikih količin informacij, je podatkovno skladišče optimizirano za povezovanje in strukturiranje podatkov za podporo analitiki, napovedovanju in odločanju vodstva.

V današnjih podjetjih podatki prihajajo iz številnih različnih virov: sistemov ERP, CAFM, CPM, CRM, računovodskih in skladiščnih sistemov, digitalnih modelov stavb CAD, senzorjev IoT in drugih rešitev. Da bi dobili celostno sliko, ni dovolj zgolj zbiranje podatkov - treba jih je organizirati, standardizirati in centralizirati v enem samem skladišču. Ravno to omogoča DWH - centraliziran sistem shranjevanja, ki

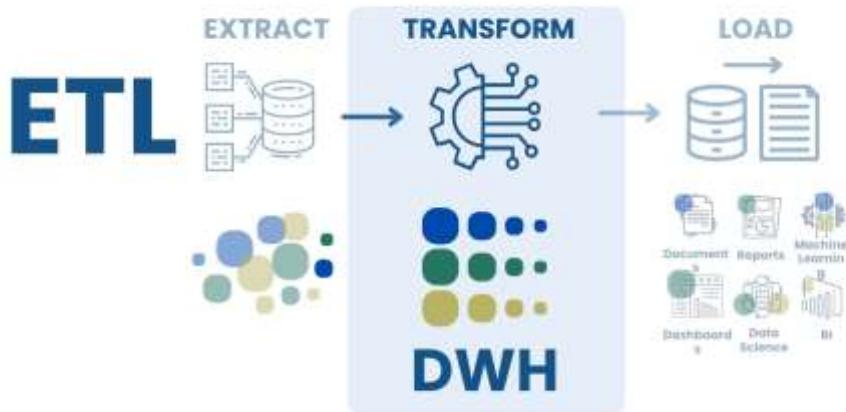
omogoča združevanje podatkov iz različnih virov, njihovo strukturiranje in dajanje na voljo za analitiko in strateško upravljanje.

**DWH (Data Warehouse)** je centraliziran sistem podatkovnega skladišča, ki združuje informacije iz različnih virov, jih strukturira in daje na voljo za analitiko in poročanje.

V številnih podjetjih so podatki razpršeni po različnih sistemih, kar smo obravnavali v prvih delih knjige (slika 1.2-4). DWH te vire združuje ter tako zagotavlja popolno preglednost in zanesljivost informacij. Podatkovno skladišče DWH je specializirana podatkovna zbirka (velika podatkovna zbirka), ki zbira, obdeluje in shranjuje podatke iz več virov. Glavne značilnosti podatkovne zbirke DWH so:

- **Uporaba procesov ETL** (Extract, Transform, Load) - pridobivanje podatkov iz virov, njihovo čiščenje, preoblikovanje, nalaganje v skladišče in avtomatizacija teh procesov, ki so bili obravnavani v sedmem delu knjige.
- **Granularnost podatkov** - podatki v DWH so lahko shranjeni v agregirani obliki (zbirna poročila) in v granularni obliki (neobdelani podatki). Od leta 2024 naprej so prav ponudniki CAD- začeli govoriti o granularnih podatkih [125], kar morda nakazuje, da se industrija pripravlja na prehod na specializirano shranjevanje v oblaku za obdelavo podatkov digitalnih modelov stavb.
- **Podpora analitiki in napovedovanju** - podatkovna skladišča so osnova za orodja BI, analizo velikih podatkov in strojno učenje.

DWH služi kot osnova za poslovno obveščanje, ki omogoča analizo ključnih kazalnikov uspešnosti, napovedovanje prodaje, nabave in stroškov ter avtomatizirano poročanje in vizualizacijo podatkov (slika 8.1-6).



Slika 8.1-6 V procesu ETL lahko DWH deluje kot osrednje skladišče, v katerem se podatki, pridobljeni iz različnih sistemov, preoblikujejo in raztovarjajo.

DWH ima ključno vlogo pri povezovanju, čiščenju in strukturiranju informacij ter zagotavlja trdno podlago za poslovno obveščanje in procese odločanja. Vendar je v današnjem okolju, kjer količina podatkov hitro narašča, viri podatkov pa postajajo vse bolj raznoliki, tradicionalni pristop DWH k

shranjevanju informacij pogosto zahteva razširitev v obliki ELT in podatkovnega jezera

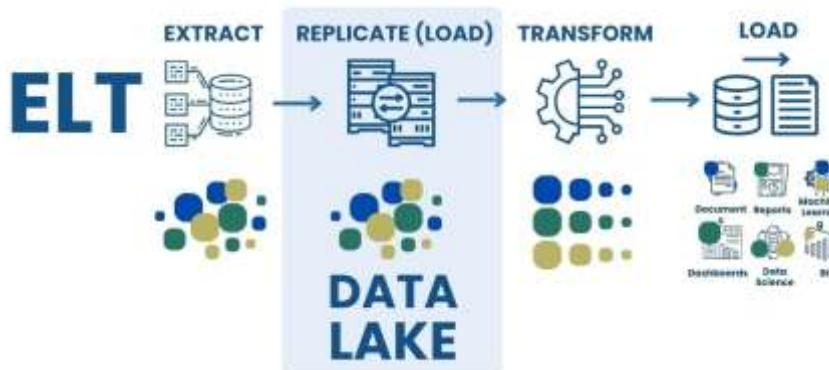
## **Podatkovno jezero - razvoj ETL do ELT: od tradicionalnega čiščenja do prilagodljive obdelave**

Klasična DWH - podatkovna skladišča, zasnovana za shranjevanje strukturiranih podatkov v oblikih, optimizirani za analitične poizvedbe, se soočajo z omejitvami pri obravnavi nestrukturiranih podatkov in skalabilnosti. Kot odgovor na te izzive so se pojavila podatkovna jezera), ki ponujajo prilagodljivo shranjevanje velikih količin heterogenih podatkov.

**Data Lake** ponuja alternativni pristop DWH, ki omogoča delo z nestrukturiranimi, polstrukturiranimi in neobdelanimi podatki brez predhodne toge sheme. Ta način shranjevanja je pogosto pomemben za obdelavo podatkov v realnem času, strojno učenje in napredno analitiko. Za razliko od DWH, ki podatke pred nalaganjem strukturira in združuje, podatkovno jezero omogoča shranjevanje podatkov v njihovi surovi oblikah, s čimer zagotavlja prilagodljivost in skalabilnost

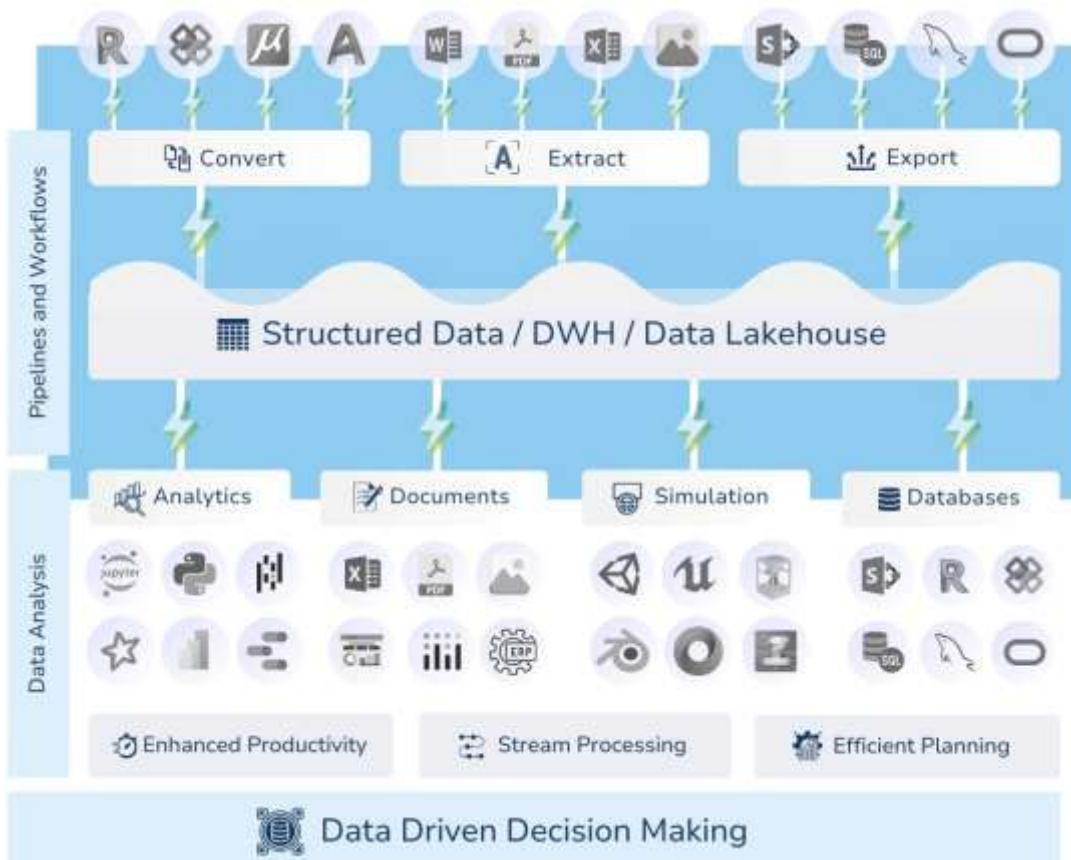
Prav razočaranje nad tradicionalnimi podatkovnimi skladišči (RDBMS, DWH) in zanimanje za "velike podatke" sta privedla do pojava podatkovnih jezer, kjer se namesto zapletenega ETL podatki zdaj preprosto naložijo v ohlapno strukturirano skladišče, obdelava pa poteka v fazi analize:

- V tradicionalnih podatkovnih skladiščih se podatki pred nalaganjem v skladišče običajno predhodno obdelajo, preoblikujejo in očistijo (ETL - Extract, Transform, Load) (slika 8.1-6). To pomeni, da so podatki strukturirani in optimizirani za posebne prihodnje analitične in poročevalske naloge. Poudarek je na ohranjanju visoke zmogljivosti poizvedb in celovitosti podatkov. Vendar je ta pristop lahko drag in manj prilagodljiv v smislu vključevanja novih vrst podatkov in hitro spreminjačih se podatkovnih schem.
- Podatkovna jezera pa so namenjena shranjevanju velikih količin neobdelanih podatkov v izvirni oblikah (slika 8.1-7). Postopek ETL (Extract, Transform, Load), nadomešča postopek ELT (Extract, Load, Transform), pri katerem se podatki najprej naložijo v skladišče "taki, kot so", in šele nato se lahko po potrebi preoblikujejo in analizirajo. To zagotavlja večjo prožnost in možnost shranjevanja heterogenih podatkov, vključno z nestrukturiranimi podatki, kot so besedilo, slike in dnevniki.



**Slika 8.1-7** Za razliko od ETL podatkovno jezero uporablja ELT, pri katerem se informacije najprej naložijo v "surovi" obliki, preoblikovanje pa se izvede v fazi nalaganja.

Tradicionalna podatkovna skladišča se osredotočajo na predhodno obdelavo podatkov, da bi zagotovili visoko zmogljivost poizvedb, medtem ko podatkovna jezera dajejo prednost prilagodljivosti: hranijo surove podatke in jih po potrebi preoblikujejo (slika 8.1-8).



**Slika 8.1-8** Sodobni koncepti shranjevanja so namenjeni shranjevanju in obdelavi vseh vrst podatkov za namene odločanja.

Kljud vsem prednostim pa podatkovna jezera niso brez pomanjkljivosti. Pomanjkanje stroge strukture in zapletenost upravljanja informacij lahko povzročita kaos, v katerem se podatki podvajajo, si

nasprotujejo ali postanejo nepomembni. Poleg tega iskanje in analiza podatkov v takšnem skladišču zahtevata precejšen napor, zlasti pri obravnavi heterogenih informacij. Za odpravo teh omejitev in združitev najboljših lastnosti tradicionalnih podatkovnih skladišč in podatkovnih jezer je bila razvita arhitektura Data Lakehouse.

## Arhitektura podatkovnega skladišča: sinergija skladišč in podatkovnih jezer

Da bi združili najboljše lastnosti DWH (strukturirana, upravljiva, visoko zmogljiva analitika) in podatkovnega jezera (skalabilnost, obravnavo heterogenih podatkov), je bil razvit pristop Data Lakehouse. Ta arhitektura združuje prilagodljivost podatkovnih jezer z zmogljivimi orodji za obdelavo in upravljanje, značilnimi za tradicionalna skladišča, ter vzpostavlja ravnovesje med shranjevanjem, analitiko in strojnim učenjem. Data Lakehouse je sinteza podatkovnih jezer in podatkovnih skladišč, ki združuje prilagodljivost in skalabilnost prvih z obvladljivostjo in optimizacijo poizvedb drugih.

**Podatkovno skladišče** je arhitekturni pristop, ki želi združiti prilagodljivost in skalabilnost podatkovnih jezer z obvladljivostjo in zmogljivostjo poizvedb podatkovnih skladišč (slika 8.1-9).

Ključne značilnosti Data Lakehouse vključujejo:

- **Odprta oblika shranjevanja podatkov:** uporaba odprtih oblik za shranjevanje podatkov, kot je Apache Parquet, zagotavlja učinkovitost in optimizirane poizvedbe.
- **Shema samo za branje:** v nasprotju s tradicionalnim pristopom sheme samo za pisanje v DWH, Lakehouse podpira shemo samo za branje, ki omogoča večjo prilagodljivost pri upravljanju podatkovne strukture.
- **Prilagodljiv in razširljiv:** podpira shranjevanje in analizo strukturiranih in nestrukturiranih podatkov ter zagotavlja visoko zmogljivost poizvedb z optimizacijo na ravni shranjevanja.

Data Lakehouse ponuja kompromisno rešitev, ki združuje prednosti obeh pristopov, zato je idealna za sodobne analitične delovne obremenitve, ki zahtevajo prilagodljivost pri obdelavi podatkov.



Slika 8.1-9 Data Lakehouse je naslednja generacija sistemov za shranjevanje podatkov, zasnovana za izpolnjevanje kompleksnih in nenehno spremenljajočih se zahtev.

Ideja sodobnih podatkovnih skladišč se zdi preprosta: če so vsi podatki na enem mestu, jih je lažje analizirati. Vendar pa v praksi ni vse tako gladko. Predstavljajte si, da se podjetje odloči popolnoma opustiti običajne računovodske in upravljaške sisteme (ERP, PMIS, CAFM ali druge) in jih nadomestiti z enim velikim podatkovnim jezerom, do katerega imajo dostop vsi. Kaj se bo zgodilo? Najverjetneje bo nastal kaos: podatki se bodo podvajali, si nasprotovali, ključne informacije pa se bodo izgubile ali poškodovale. Tudi če se podatkovno jezero uporablja le za analitiko, bo brez ustreznega upravljanja resno ogroženo:

- Podatke je težko razumeti: v običajnih sistemih imajo podatki jasno strukturo, v jezeru pa so le ogromna kopica datotek in tabel. Če želi nekaj najti, mora ugotoviti, za kaj je odgovorna vsaka vrstica in stolpec.
- Podatki so lahko netočni: če je na enem mestu shranjenih več različic istih informacij, je težko ugotoviti, katera različica je posodobljena. Zato se odločitve sprejemajo na podlagi zastarelih ali napačnih podatkov.
- Težko je pripraviti podatke za delo: podatkov je treba ne le shraniti, temveč tudi predstaviti v primerni obliki - v obliki poročil, grafov, tabel. V tradicionalnih sistemih je to opravljeno samodejno, v podatkovnih jezerih pa je potrebna dodatna obdelava.

Zato ima vsak koncept podatkovnega skladiščenja svoje značilnosti, pristope k obdelavi in poslovne aplikacije. Tradicionalne podatkovne zbirke se osredotočajo na transakcijske operacije, podatkovna skladišča (DWH) zagotavljajo strukturo za analitiko, podatkovna jezera (Data Lake) shranjujejo informacije v surovi obliki, hibridna skladišča (Data Lakehouse) pa združujejo prednosti DWH in Data Lake (slika 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

**Slika 8.1-10 DWH, Data Lake in Data Lakehouse: ključne razlike v vrstah podatkov, scenarijih uporabe, metodah obdelave in pristopih shranjevanja.**

Izbira arhitekture za shranjevanje je zapleten proces, ki je odvisen od poslovnih potreb, obsega informacij in analitičnih zahtev. Vsaka rešitev ima svoje prednosti in slabosti: DWH zagotavlja strukturo, Data Lake zagotavlja prilagodljivost, Lakehouse pa zagotavlja ravnovesje med obema. Organizacije so le redko omejene na eno samo podatkovno arhitekturo.

Ne glede na izbrano arhitekturo so avtomatizirani sistemi za upravljanje podatkov bistveno boljši od ročnih metod. Zmanjšujejo človeške napake, pospešujejo obdelavo informacij ter zagotavljajo preglednost in sledljivost podatkov v vseh fazah poslovnih procesov.

Medtem ko so centralizirana podatkovna skladišča na številnih področjih gospodarstva že postala industrijski standard, je stanje v gradbeništvu še vedno razdrobljeno. Podatki tukaj so porazdeljeni po različnih platformah (CDE, PMIS, ERP itd.), zato je težko ustvariti enotno sliko dogajanja in so potrebne arhitekture, ki so sposobne združiti te vire v celovito, analitično uporabno digitalno okolje.

## CDE, PMIS, ERP ali DWH in podatkovno jezero

Nekatera gradbena in inženirska podjetja že uporabljajo koncept skupnega podatkovnega okolja (CDE) v skladu s standardom ISO 19650. CDE v bistvu opravlja enake funkcije kot podatkovno skladišče (DWH) v drugih panogah: centralizira informacije, zagotavlja nadzor nad različicami in dostop do potrjenih informacij.

**Skupno podatkovno okolje (CDE)** je centraliziran digitalni prostor, ki se uporablja za upravljanje, shranjevanje, izmenjavo in sodelovanje pri projektnih informacijah v vseh fazah življenjskega cikla objekta. CDE se pogosto izvaja s tehnologijami v oblaku in je integrirano s sistemmi CAD (BIM).

Finančni, maloprodajni, logistični in industrijski sektorji že desetletja uporabljajo centralizirane sisteme za upravljanje podatkov, ki združujejo informacije iz različnih virov, nadzorujejo njihovo pomembnost in zagotavljajo analitiko. CDE ta načela nadgrajuje in jih prilagaja izzivom načrtovanja stavb in upravljanja življenjskega cikla.

Podobno kot DWH, tudi CDE strukturira podatke, zajema spremembe in zagotavlja enotno točko dostopa do preverjenih informacij. S prehodom v oblak in integracijo z analitičnimi orodji so razlike med njima vse manj očitne. Če k CDE dodamo še granularne podatke, o konceptu katerih so prodajalci CAD razpravljali že leta 2023[93, 125], lahko vidimo še več vzporednic s klasičnim DWH.

V poglavju "Gradbeni sistemi ERP in PMIS" smo že obravnavali informacijski sistem za vodenje projektov (PMIS) in ERP (Enterprise Resource Planning). Pri gradbenih projektih CDE in PMIS delujeta skupaj: CDE služi kot skladišče podatkov, vključno z risbami, modeli in projektno dokumentacijo, PMIS pa upravlja procese, kot so nadzor rokov, nalog, virov in proračunov.

Sistem ERP, ki je odgovoren za upravljanje poslovanja kot celote (finance, nabava, osebje, proizvodnja), se lahko poveže s sistemom PMIS, ki zagotavlja nadzor stroškov in proračuna na ravni podjetja. Za analitiko in poročanje se lahko DWH uporablja za zbiranje, strukturiranje in združevanje podatkov iz CDE, PMIS in ERP za ocenjevanje finančnih ključnih kazalnikov uspešnosti (ROI) in ugotavljanje vzorcev. Podatkovno jezero (DL) pa lahko dopolnjuje DWH s shranjevanjem surovih in nestrukturiranih podatkov (npr. dnevnički, podatki senzorjev, slike). Ti podatki se lahko obdelajo in naložijo v DWH za nadaljnjo analizo.

Tako se CDE in PMIS osredotočata na vodenje projektov, ERP na poslovne procese, DWH in Data Lake pa na analitiko in podatkovne operacije.

Pri primerjavi sistemov CDE, PMIS in ERP s sistemi DWH in Data Lake so vidne pomembne razlike v glede neodvisnosti od prodajalca, stroškov, prilagodljivosti integracije, neodvisnosti podatkov, hitrosti prilagajanja spremembam in analitičnih zmogljivosti (slika 8.1-11). Tradicionalni sistemi, kot so CDE, PMIS in ERP, so pogosto vezani na rešitve in standarde določenih prodajalcev, zaradi česar so manj prilagodljivi, njihovi stroški pa se povečujejo zaradi licenc in podpore. Poleg tega so podatki v takih sistemih pogosto zaprti v lastniške, zaprte formate, kar omejuje njihovo uporabo in analizo.

		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)

**Slika 8.1-11 DWH in podatkovno jezero ponujata večjo prilagodljivost in neodvisnost podatkov kot sistemi, kot so CDE, PMIS in ERP.**

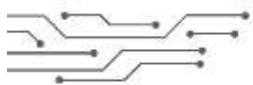
V nasprotju s tem pa DWH in Data Lake zagotavljata večjo prilagodljivost pri povezovanju z različnimi viri podatkov, njuna uporaba odprtih tehnologij in platform pa pomaga zmanjšati skupne stroške lastništva. Poleg tega DWH in Data Lake podpirata široko paletu analitičnih orodij, kar izboljšuje možnosti analize in upravljanja.

Z razvojem orodij za povratno inženirstvo za formate CAD in dostopom do podatkovnih zbirk aplikacij CAD postaja vse bolj pereče vprašanje: kako upravičena je nadaljnja uporaba zaprtih, izoliranih platform, če morajo biti projektni podatki na voljo širokemu krogu strokovnjakov, ki delajo pri več deset izvajalcih in projektantskih organizacijah?

Ta tehnološka odvisnost od dobavitelja lahko bistveno omeji prilagodljivost upravljanja podatkov, upočasni odzivanje na spremembe projekta in ovira učinkovito sodelovanje med udeleženci.

Tradisionalni pristopi k upravljanju podatkov - vključno z DWH, Data Lake, CDE in PMIS - so se osredotočali predvsem na shranjevanje, strukturiranje in obdelavo informacij. Z razvojem umetne inteligence in strojnega učenja pa narašča potreba po novih načinih organiziranja podatkov, ki ne le združujejo, temveč tudi razkrivajo zapletene povezave, iščejo skrite vzorce in zagotavljajo takojšen dostop do najpomembnejših informacij.

Posebno vlogo v tej smeri začenjajo igrati vektorske podatkovne zbirke - nova vrsta shrambe, optimizirana za visokodimenzionalne vgradnje.



## POGLAVJE 8.2.

# UPRAVLJANJE PODATKOVNEGA SKLADIŠČA IN PREPREČEVANJE KAOSA

### Vektorske zbirke podatkov in omejitveno polje

Vektorske podatkovne zbirke so nov razred shramb, ki ne hranijo le podatkov, temveč omogočajo iskanje po pomenu, primerjanje predmetov glede na semantično bližino in ustvarjanje inteligentnih sistemov: od priporočil do samodejne analize in ustvarjanja konteksta. Za razliko od tradicionalnih podatkovnih zbirk, ki se osredotočajo na natančna ujemanja, vektorske podatkovne zbirke iščejo podobne predmete na podlagi atributov - tudi če ni natančnega ujemanja

**Vektorska** podatkovna zbirka je specializirana vrsta podatkovne zbirke, ki hrani podatke kot večdimensionalne vektorje, od katerih vsak predstavlja določene značilnosti ali lastnosti. Ti vektorji imajo lahko različno število dimenzij, odvisno od kompleksnosti podatkov (v enem primeru je to lahko nekaj dimenzij, v drugem pa — tisoč).

Glavna prednost vektorskih podatkovnih zbirk je iskanje po semantični ustreznosti in ne po natančnem ujemanju vrednosti. Namesto poizvedb SQL in Pandas s filtri "equals" ali "contains" se uporablja iskanje najbližjih sosedov (k-NN) (o k-NN bomo več govorili v naslednjem delu knjige) v prostoru funkcij.

Z razvojem LLM (Large Language Models) in generativnih modelov se interakcija s podatkovnimi zbirkami začenja spreminjati. Zdaj je mogoče poizvedovati po podatkih v naravnem jeziku, dobiti semantična iskanja po dokumentih, samodejno izluščiti ključne izraze in zgraditi kontekstualne odnose med objekti - vse to brez potrebe po znanju jezika SQL ali poznavanju strukture tabel. To je bilo podrobneje obravnavano v poglavju "MŠO in njihova vloga pri obdelavi podatkov in poslovnih procesih".

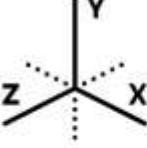
Vendar pa se je treba zavedati, da MDV ne strukturira in ureja informacij samodejno. Model samo preleti podatke in na podlagi konteksta poizvedbe poišče najustreznejši del podatkov. Če podatki niso bili predhodno očiščeni ali preoblikovani, bo poglobljeno iskanje podobno iskanju odgovora v digitalnih "smeteh" - morda bo delovalo, vendar bo kakovost rezultatov slabša. Najbolje je, če je mogoče podatke strukturirati (npr. dokumente prevesti v Markdown) in jih naložiti v vektorsko podatkovno zbirko. S tem se bistveno povečata natančnost in ustreznost rezultatov.

Sprva so se vektorske zbirke podatkov uporabljale pri strojnem učenju, danes pa se vse pogosteje uporabljajo tudi zunaj njega - v iskalnikih, personalizaciji vsebine in inteligentni analitiki.

Eden od najbolj očitnih primerov vektorskega pristopa v konstrukciji je omejujoča škatla (omejujoči paralelepiped). Gre za geometrijsko konstrukcijo, ki opisuje meje predmeta v trirazsežnem prostoru.

Bounding Box je opredeljen z minimalnimi in maksimalnimi koordinatami X, Y in Z, ki tvorijo "škatlo" okoli predmeta. S to metodo lahko ocenite velikost in umestitev elementa, ne da bi morali analizirati celotno geometrijo.

Vsako omejitveno polje lahko predstavimo kot vektor v večdimenzionalnem prostoru: na primer [x, y, z, širina, višina, globina] - že 6 dimenzij (slika 8.2-1)



### Bounding Box

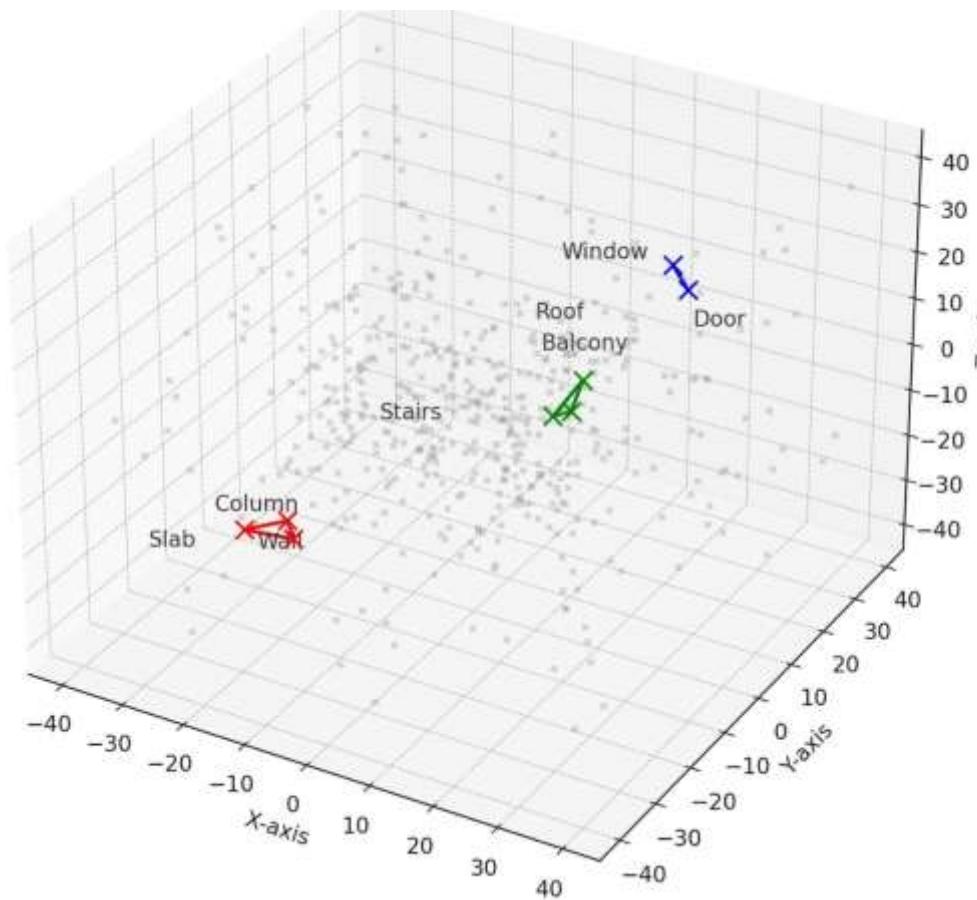
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Slika 8.2-1 Bounding Box -informacije o koordinatah elementov in njihovi lokaciji v modelu projekta so analogne vektorski podatkovni zbirki.

Ta predstavitev podatkov olajša številne naloge, med drugim preverjanje stičišč med objekti, načrtovanje prostorske razporeditve gradbenih elementov in izvajanje samodejnih izračunov. Bounding Box lahko služi kot most med zapletenimi 3D modeli in tradicionalnimi vektorskimi podatkovnimi zbirkami ter omogoča učinkovito uporabo prednosti obeh pristopov pri arhitektturnem in inženirskem modeliranju

Bounding Box je "vektorizacija geometrije", embedding (način preoblikovanja nečesa abstraktnega) pa je "vektorizacija pomena". Oba pristopa omogočata prehod od ročnega iskanja k inteligentnemu iskanju, pa naj gre za 3D objekte v modelu projekta ali koncepte v besedilu.

Iskanje objektov v projektu (na primer "poišči vsa okna s širino > 1,5 m") je podobno iskanju najbližjih sosedov (k-NN) v vektorski zbirki podatkov, kjer merila določajo "območje" v prostoru značilnosti. (več o iskanju najbližjih sosedov k-NN bomo povedali v naslednjem delu o strojnem učenju) (slika 8.2-2). Če atributom omejitvenega območja dodamo dodatne parametre (material, težo, čas izdelave), se tabela spremeni v visokodimensionalni vektor, kjer je vsak atribut nova dimenzija. To je bližje sodobnim vektorskim bazam, kjer se dimenzijs stejejo v stotinah ali tisočih (npr. vgradnja iz nevronskeih mrež).



Slika 8.2-2 Iskanje predmetov v projektu s pomočjo vektorskih podatkovnih zbirk.

Pristop, ki se uporablja v Bounding Box, se ne uporablja le za geometrijske objekte, temveč tudi za analizo besedila in jezika. Vektorske predstavitev podatkov se že aktivno uporabljam pri obdelavi naravnega jezika (NLP). Tako kot lahko predmete v gradbenem projektu razvrstimo v skupine glede na njihovo prostorsko bližino (slika 8.2-2), lahko besede v besedilu analiziramo glede na njihovo semantično in kontekstualno bližino.

Na primer, besede "arhitekt", "gradnja" in "oblikovanje" bodo v vektorskem prostoru druga ob drugi, ker imajo podoben pomen. V programu LLM ta mehanizem omogoča samodejno, brez potrebe po ročnem razvrščanju:

- prepoznavanje vsebine besedila
- izvajanje semantičnih iskanj po vsebini dokumentov
- Ustvarjanje samodejnih opomb in povzetkov besedila
- Poiščite sinonime in sorodne izraze

Vektorske zbirke podatkov vam omogočajo analizo besedila in iskanje povezanih izrazov v njem na enak način kot zbirka Bounding Box omogoča analizo prostorskih objektov v 3D -modelu. Primer Bounding Box elementov projekta pomaga razumeti, da vektorska predstavitev ni zgolj "umetni"

koncept iz ML, temveč naravni način strukturiranja podatkov za reševanje aplikativnih problemov, pa naj gre za iskanje stolpcev v projektu CAD ali za semantično blizu slike v podatkovni zbirkvi.

Strokovnjaki, ki delajo s podatkovnimi zbirkami, morajo biti pozorni na vektorske shrambe. Njihov razmah kaže na novo fazo v razvoju podatkovnih zbirk, kjer se začenjajo prepletati klasični relacijski sistemi in tehnologije, usmerjene v umetno inteligenco, ki tvorijo hibridne rešitve prihodnosti.

Uporabniki, ki razvijajo zapletene in obsežne aplikacije umetne inteligence, bodo za vektorsko iskanje uporabljali specializirane zbirke podatkov. Hkrati bodo tisti, ki potrebujejo le ločene funkcije umetne inteligence za vključitev v obstoječe aplikacije, najverjetneje izbrali vgrajene zmogljivosti vektorskega iskanja v podatkovnih zbirkah, ki jih že uporablja (PostgreSQL, Redis).

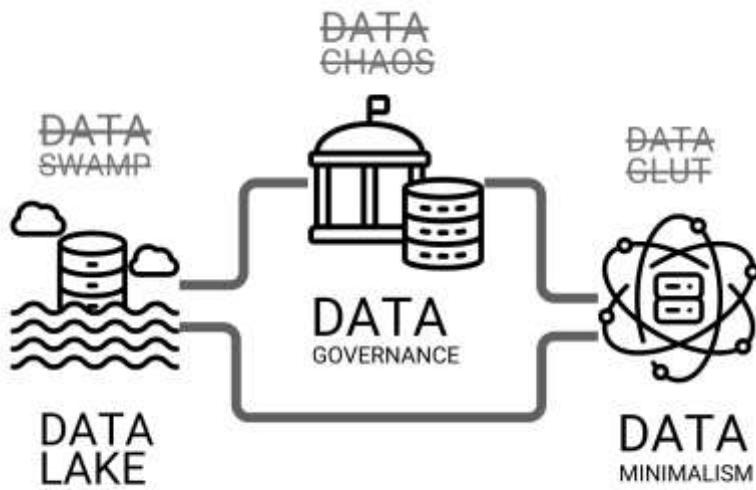
Čeprav sistemi, kot so DWH, Data Lake, CDE, PMIS, vektorske podatkovne zbirke in drugi, ponujajo različne pristope k shranjevanju in upravljanju podatkov, njihova učinkovitost ni odvisna le od njihove arhitekture, temveč tudi od tega, kako dobro so organizirani in upravljeni sami podatki. Tudi pri uporabi sodobnih rešitev - pa naj gre za vektorske podatkovne zbirke, klasične relacijske sisteme DBMS ali skladišča tipa Data Lake - lahko pomanjkanje jasnih pravil za upravljanje, strukturiranje in posodabljanje podatkov privede do enakih težav, s katerimi se srečujejo uporabniki, ki delajo z raznovrstnimi datotekami in podatki različnih formatov.

Brez upravljanja podatkov) lahko tudi najzmožljivejše rešitve postanejo kaotične in nestrukturirane, kar podatkovna jezera spremeni v podatkovna močvirja). Da bi se temu izognila, morajo podjetja ne le izbrati pravo arhitekturo shranjevanja, temveč tudi izvajati strategije za minimalizacijo podatkov), upravljanje dostopa in nadzor kakovosti, da bi podatke spremenila v učinkovito orodje za odločanje.

### **Upravljanje podatkov), podatkovni minimalizem) in podatkovno močvirje)**

Razumevanje in izvajanje konceptov upravljanja podatkov), podatkovnega minimalizma) in preprečevanja podatkovnega močvirja) sta ključna za uspešno upravljanje podatkovnih skladišč in zagotavljanje poslovne vrednosti (slika 8.2-3).

Po podatkih študije družbe Gartner (2017) je 85% projektov velikih podatkov neuspešnih, eden od ključnih razlogov pa je nezadostna kakovost podatkov in nezadostno upravljanje podatkov [144].



Slika 8.2-3 Nekateri ključni vidiki upravljanja podatkov so upravljanje podatkov in podatkovni minimalizem.

Upravljanje podatkov (Data Governance) je temeljna sestavina upravljanja podatkov, ki zagotavlja, da se podatki ustrezno in učinkovito uporabljajo v vseh poslovnih procesih. Pri tem ne gre le za vzpostavitev pravil in postopkov, temveč tudi za zagotavljanje razpoložljivosti, zanesljivosti in varnosti podatkov:

- Opredelitev in razvrstitev podatkov: jasna opredelitev in razvrstitev entitet omogoča organizacijam, da razumejo, katere entitete so v podjetju potrebne, in določijo, kako jih je treba uporabljati.
- Pravice dostopa in upravljanje: razvoj politik in postopkov za dostop do podatkov in njihovo upravljanje zagotavlja, da lahko do določenih podatkov dostopajo le pooblaščeni uporabniki.
- Zaščita podatkov pred zunanjimi grožnjami: zaščita podatkov pred zunanjimi grožnjami je ključni vidik upravljanja podatkov. To ne vključuje le tehničnih ukrepov, temveč tudi usposabljanje zaposlenih o osnovah informacijske varnosti.

Minimalizem podatkov (Data Minimalism) je pristop za zmanjšanje podatkov na najbolj dragocene in smiselne atributte in entitete v formaciji (slika 8.2-4), s čimer se zmanjšajo stroški in izboljša uporaba podatkov:

- Poenostavitev odločanja: zmanjšanje števila predmetov in njihovih atributov na najpomembnejše poenostavi odločanje, saj zmanjša čas in sredstva, potrebna za analizo in obdelavo podatkov.
- Osredotočanje na pomembno: z izbiro najpomembnejših entitet in atributov se lahko osredotočite na informacije, ki so resnično pomembne za poslovanje, ter izločite šum in nepotrebne podatke.
- Učinkovito razporejanje virov: minimizacija podatkov omogoča učinkovitejše razporejanje virov, zmanjšanje stroškov shranjevanja in obdelave podatkov ter izboljšanje kakovosti in varnosti podatkov.

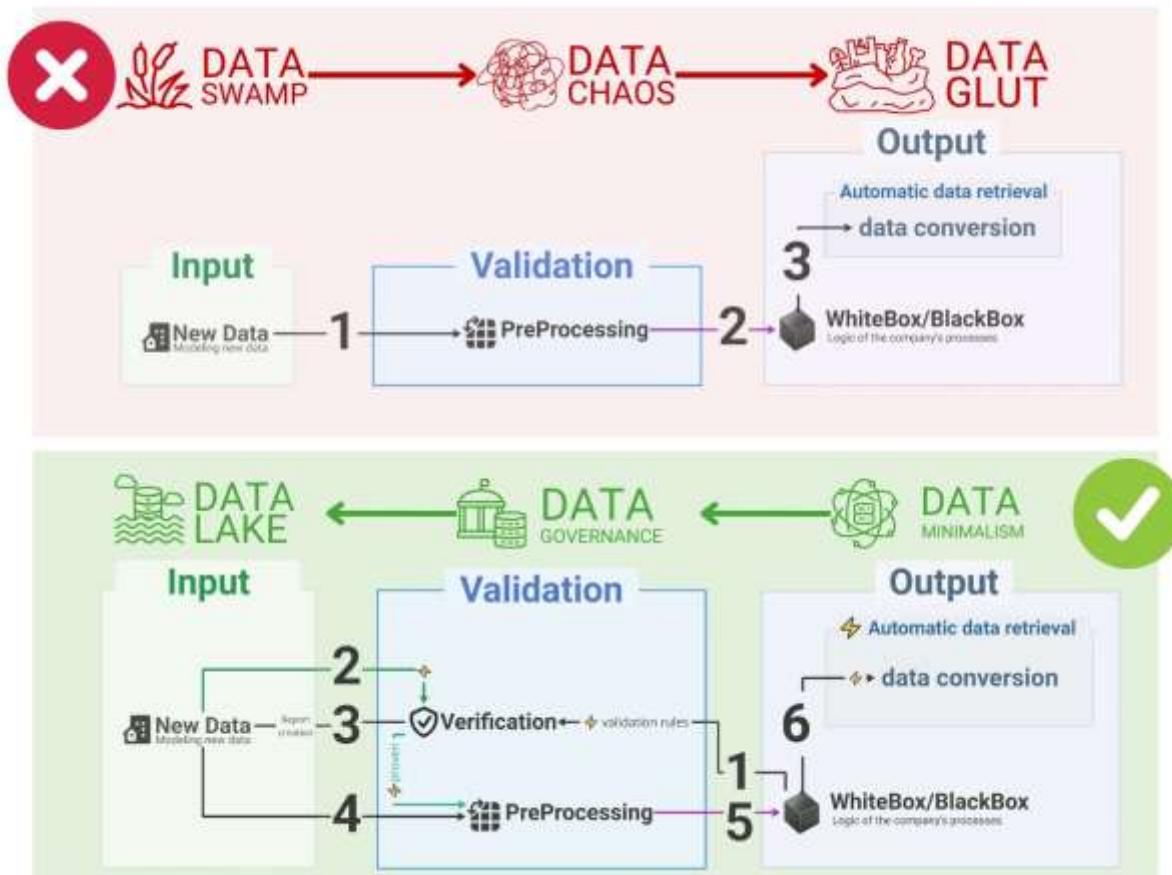
Logika dela s podatki se ne sme začeti z njihovim ustvarjanjem kot takim (slika 8.2-4), temveč z razumevanjem prihodnjih scenarijev uporabe teh podatkov še pred začetkom postopka ustvarjanja. Ta pristop omogoča vnaprejšnjo opredelitev minimalnih potrebnih zahtev za attribute, njihove vrste in mejne vrednosti. Te zahteve so podlaga za ustvarjanje pravilnih in stabilnih entitet v informacijskem modelu. Predhodno razumevanje namenov in uporabe podatkov prispeva k oblikovanju strukture, primerne za analizo. Več podrobnosti o pristopih k modeliranju podatkov na konceptualni, logični in fizični ravni je bilo obravnavanih v poglavju "Modeliranje podatkov: konceptualni, logični in fizični model".

V tradicionalnih poslovnih procesih gradbenih podjetij obdelava podatkov pogosteje spominja na odlaganje podatkov v močvirje, kjer se podatki najprej ustvarijo, nato pa jih strokovnjaki poskušajo vključiti v druge sisteme in orodja.

Podatkovno močvirje (Data Swamp) je posledica nenadzorovanega zbiranja in shranjevanja podatkov brez ustrezne organizacije, strukturiranja in upravljanja, zaradi česar so podatki nestrukturirani, težko uporabni in imajo majhno vrednost.

Kako preprečiti, da bi pretok informacij postal močvirje:

- **Upravljanje strukture podatkov:** zagotavljanje strukturiranosti in kategorizacije podatkov na spletnem mestu pomaga preprečiti kopiranje podatkov, saj so ti urejeni in lahko dostopni.
- **Razumevanje in razlaganje podatkov:** jasen opis izvora, sprememb in pomenov podatkov zagotavlja, da se podatki pravilno razumejo in razlagajo.
- **Vzdrževanje kakovosti podatkov:** redno vzdrževanje in čiščenje podatkov pomaga ohranjati kakovost, ustreznost in vrednost podatkov za analitiko in poslovne procese.



Slika 8.2-4 Da bi se izognili neredu v podatkovnem skladišču, morate postopek ustvarjanja podatkov začeti z zbiranjem zahtev za attribute.

Z vključevanjem načel upravljanja podatkov in podatkovnega minimalizma v postopke upravljanja podatkov ter z aktivnim preprečevanjem, da bi podatkovna skladišča postala podatkovna močvirja, lahko organizacije kar najbolj izkoristijo potencial svojih podatkov.

Naslednja stopnja v razvoju dela s podatki, po rešitvi vprašanj upravljanja in minimalizma, je standardizacija samodejne obdelave, zagotavljanje kakovosti in izvajanje metod, ki omogočajo uporabo podatkov za analizo, preoblikovanje in odločanje. To počneta metodologiji DataOps in VectorOps, ki postajata pomembni orodji za podjetja, ki se ukvarjajo z velikimi podatki in strojnim učenjem.

## DataOps in VectorOps: novi podatkovni standardi

Data Governance je odgovoren za nadzor in organizacijo podatkov, DataOps pa pomaga zagotavljati njihovo točnost, doslednost in nemoten pretok znotraj podjetja. To je še posebej pomembno za številne poslovne primere v gradbeništvu, kjer podatki nastajajo neprekinjeno in jih je treba pravočasno obdelati. Na primer v primerih, ko je treba v enem delovnem dnevu med različnimi sistemi sinhronizirati

informacijske modele stavb, zahteve projekta in analitična poročila, je vloga DataOps lahko ključna. Omogoča vzpostavitev stabilnih in ponovljivih procesov obdelave podatkov, kar zmanjšuje tveganje zamud in izgube relevantnosti informacij.

Samo upravljanje podatkov ni dovolj - bistveno je, da se podatki ne le shranjujejo, temveč tudi aktivno uporabljajo pri vsakodnevni poslovanju. Tu pride na vrsto DataOps - metodologija, ki se osredotoča na avtomatizacijo, integracijo in neprekinjen pretok podatkov.

**DataOps** se osredotoča na izboljšanje sodelovanja, integracije in avtomatizacije podatkovnih tokov v organizacijah. Sprejemanje praks DataOps spodbuja točnost, doslednost in razpoložljivost podatkov, kar je ključnega pomena za podatkovno usmerjene aplikacije.

Ključni orodji v ekosistemu DataOps sta Apache Airflow (slika 7.4-4) za orkestracijo delovnih tokov in Apache NiFi (slika 7.4-5) za usmerjanje in preoblikovanje podatkovnih tokov. Te tehnologije skupaj omogočajo prilagodljive, zanesljive in skalabilne podatkovne cevovode za samodejno obdelavo, nadzor in povezovanje informacij med sistemi (več podrobnosti v poglavju "Samodejni ETL -prevajalnik"). Pri izvajanju pristopa DataOps v gradbenih procesih je treba upoštevati štiri temeljne vidike:

1. **Ljudje in orodja so pomembnejši od podatkov:** ločene podatkovne shrambe se morda zdijo velika težava, vendar je realnost bolj zapletena. Poleg razdrobljenosti podatkov imajo pomembno vlogo tudi izoliranost ekip in različna orodja, ki jih uporabljajo. V gradbeništvu s podatki delajo strokovnjaki z različnih področij: podatkovni inženirji in analitiki, ekipe za BI in vizualizacijo ter strokovnjaki za vodenje projektov in kakovost. Vsak od njih ima različne načine dela, zato je pomembno ustvariti ekosistem, v katerem se podatki prosto pretakajo med udeleženci in zagotavlja enotno, dosledno različico informacij.
2. **Avtomatizirano testiranje in odkrivanje napak:** Gradbeni podatki vedno vsebujejo napake, bodisi da gre za netočnosti v modelih, računske napake ali zastarele specifikacije. Z rednim testiranjem podatkov in odpravljanjem ponavljajočih se napak lahko znatno izboljšate kakovost podatkov. Kot del sistema DataOps morate uvesti avtomatizirane kontrole in mehanizme potrjevanja, ki spremljajo pravilnost podatkov, analizirajo napake in prepoznavajo vzorce ter zajemajo in odpravljajo sistemski napake v vsakem delovnem postopku. Višja kot je stopnja samodejnega potrjevanja, večja je splošna kakovost podatkov in manjša verjetnost napak v končnih fazah.
3. **Podatke je treba preverjati enako kot programsko kodo:** večina gradbenih aplikacij temelji na obdelavi podatkov, vendar je njihov nadzor pogosto prepuščen sekundarnim vlogam. Če se modeli strojnega učenja usposabljam na netočnih podatkih, to vodi do napačnih napovedi in finančnih izgub. V okviru DataOps je treba podatke podvreči enako strogemu nadzoru kot programsko kodo: preverjanju logike, stresnim testom in ocenjevanju obnašanja modelov ob spremembah vhodnih vrednosti. Samo potrjeni in zanesljivi podatki se lahko uporabljajo kot podlaga za odločitve vodstva.
4. **Opazovanje podatkov brez ogrožanja učinkovitosti:** spremljanje podatkov ni le zbirka metrik, temveč strateško orodje za upravljanje kakovosti. Za učinkovito delovanje sistema DataOps je treba opazljivost vgraditi v vse faze ravnjanja s podatki, od zasnove do delovanja. Hkrati je pomembno, da spremljanje ne upočasnjuje sistema. Pri gradbenih projektih je ključnega

pomena, da se podatki ne le zbirajo, temveč da se zbirajo tako, da nikakor ne motijo dela strokovnjakov (npr. projektantov), ki ustvarjajo podatke. To ravnovesje omogoča nadzor kakovosti podatkov, ne da bi pri tem ogrozili produktivnost.

DataOps ni dodatno breme za podatkovne znanstvenike, temveč temelj njihovega dela. Z uvedbo DataOps lahko gradbena podjetja preidejo od kaotičnega upravljanja podatkov k učinkovitemu ekosistemu, v katerem podatki delajo za poslovanje.

VectorOps pa predstavlja naslednjo stopnjo v razvoju DataOps, ki se osredotoča na obdelavo, shranjevanje in analizo večdimenzionalnih vektorskih podatkov (ki so bili obravnavani v prejšnjem poglavju). To je zlasti pomembno na področjih, kot so digitalni dvojčki, modeli nevronskih mrež in semantično iskanje, ki začenjajo prihajati v gradbeno industrijo. VectorOps temelji na vektorskih podatkovnih zbirkah za učinkovito shranjevanje, indeksiranje in iskanje večdimenzionalnih predstavitev objektov.

**VectorOps** je naslednji korak po DataOps, osredotočen na obdelavo, analizo in uporabo vektorskih podatkov v gradbeništvu. Za razliko od DataOps, ki se osredotoča na pretok, doslednost in kakovost podatkov, se VectorOps osredotoča na upravljanje večdimenzionalnih predstavitev objektov, potrebnih za strojno učenje.

Za razliko od tradicionalnih pristopov vam VectorOps omogoča natančnejše opise objektov, kar je ključnega pomena za digitalne dvojčke, sisteme generativnega načrtovanja in samodejno odkrivanje napak v podatkih CAD, pretvorjenih v vektorsko obliko. Skupna implementacija DataOps in VectorOps tvori trdno podlago za skalabilno, avtomatizirano delo z velikimi količinami informacij - od klasičnih tabel do semantično bogatih prostorskih modelov.

## Naslednji koraki: od kaotičnega do strukturiranega shranjevanja

Tradicionalni pristopi k gradnji podatkovnih shramb pogosto vodijo do ustvarjanja različnih "silosov informacij", kjer so pomembna spoznanja nedostopna za analizo in sprejemanje odločitev. Sodobni koncepti shranjevanja, kot so podatkovno skladišče, podatkovno jezero in njihovi hibridi, omogočajo poenotenje raznovrstnih informacij in njihovo centralizirano dostopnost za pretakanje podatkov in poslovno obveščanje. Pomembno je ne le izbrati pravo arhitekturo shranjevanja, temveč tudi izvajati upravljanje podatkov) in podatkovni minimalizem), da se prepreči, da bi shrambe postale neobvladljiva podatkovna močvirja).

Za povzetek tega dela velja izpostaviti glavne praktične korake, ki vam bodo pomagali pri uporabi obravnavanih konceptov pri vsakodnevnih nalogah:

- Izbera učinkovitih oblik shranjevanja podatkov
  - Prehod od formatov CSV in XLSX k učinkovitejšim formatom (Apache Parquet, ORC) za shranjevanje velikih količin podatkov.
  - uvedba sistema za različico podatkov za sledenje spremembam
  - uporaba metapodatkov za opis strukture in izvora informacij

**■ Ustvarite enotno podatkovno arhitekturo podjetja**

- Primerjajte različne arhitekture shranjevanja: RDBMS, DWH in Data Lake. Izberite tisto, ki najbolj ustreza vašim potrebam po skalabilnosti, integraciji virov in analitični obdelavi.
- Oblikujte načrt postopka za pridobivanje, nalaganje in preoblikovanje podatkov (ETL) iz različnih virov za svoje naloge. Za vizualizacijo ključnih korakov in integracijskih točk uporabite orodja za vizualizacijo, kot so Miro, Lucidchart ali Draw.io

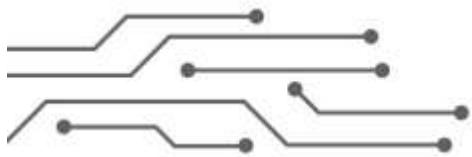
**■ Izvajanje praks upravljanja podatkov in podatkovnega minimalizma**

- upoštevajte pristop minimalizma podatkov - shranjujte in obdelujte le tisto, kar je resnično dragoceno.
- Izvajanje načel upravljanja podatkov - opredelitev odgovornosti za podatke, zagotavljanje kakovosti in preglednosti.
- Več o politikah upravljanja podatkov in konceptih DataOps, VectorOps
- Opredelitev merit kakovosti podatkov in postopkov za preverjanje podatkov v okviru DataOps.

Dobro organizirano shranjevanje podatkov je podlaga za centralizacijo analitičnih procesov podjetja. Prehod od kaotičnega kopiranja datotek k strukturiranim shrambam omogoča, da se informacije spremenijo v strateško sredstvo, ki pomaga sprejemati informirane odločitve in izboljšati učinkovitost poslovnih procesov.

Ko so postopki zbiranja, preoblikovanja, analize in strukturiranega shranjevanja podatkov avtomatizirani in standardizirani, je naslednja stopnja digitalne preobrazbe polnopravno ravnanje z velikimi podatki.





## IX DEL

# **VELIKI PODATKI, STROJNO UČENJE IN NAPOVEDI**

Deveti del se osredotoča na velike podatke, strojno učenje in napovedno analitiko v gradbeništvu. Raziskuje prehod od intuitivnega sprejemanja odločitev k objektivnim analizam na podlagi preteklih podatkov. S praktičnimi primeri so prikazane analize velikih podatkov v gradbeništvu - od razčlenjevanja podatkovnega niza gradbenih dovoljenj v San Franciscu do obdelave CAD - projektov z milijoni elementov. Posebna pozornost je namenjena metodam strojnega učenja za napovedovanje stroškov in časovnega razporeda gradbenih projektov, s podrobno obravnavo linearne regresije in algoritmov k-najbližjih sosedov. Prikazano je, kako strukturirani podatki postanejo osnova za napovedne modele za ocenjevanje tveganj, optimizacijo virov in izboljšanje učinkovitosti upravljanja projektov. Del vsebuje tudi priporočila za izbiro reprezentativnih vzorcev podatkov in pojasnjuje, zakaj za učinkovite analize niso vedno potrebni veliki podatkovni nizi.

## POGLAVJE 9.1.

# VELIKI PODATKI IN NJHOVA ANALIZA

### **Veliki podatki v gradbeništvu: od intuicije do predvidljivosti**

Izraz "veliki podatki" nima stroge opredelitve. Pojem se je prvotno pojavil, ko je količina informacij začela presegati zmogljivosti tradicionalnih metod njihove obdelave. Danes sta se količina in kompleksnost podatkov v številnih panogah, tudi v gradbeništvu, tako povečali, da jih ni mogoče spraviti v lokalni pomnilnik računalnikov in je za njihovo obdelavo treba uporabiti nove tehnologije.

Bistvo dela z velikimi podatki ni le shranjevanje in obdelava, temveč tudi napovedovanje. V gradbeništvu veliki podatki odpirajo pot od intuitivnih odločitev, ki temeljijo na subjektivni razlagi tabel in vizualizacij (kot je bilo obravnavano prej), do utemeljenih napovedi, podprtih z resničnimi opažanji in statističnimi podatki.

V nasprotju s splošnim prepričanjem cilj dela z velikimi količinami podatkov ni "narediti stroj, ki bo razmišljal kot človek", temveč uporaba matematičnih modelov in algoritmov za analizo velikih količin podatkov, da bi prepoznali vzorce, napovedali dogodke in optimizirali procese.

**Veliki podatki** niso hladen svet algoritmov brez človeškega vpliva. Nasprotno, veliki podatki delujejo v povezavi z našimi instinkti, napakami in ustvarjalnostjo. Prav nepopolnost človeškega razmišljanja nam omogoča, da najdemo nestandardne rešitve in dosežemo preboj.

Z razvojem digitalne tehnologije je gradbena industrija začela dejavno uporabljati tehnike obdelave podatkov, ki so prišle iz sektorja IT. Zahvaljujoč orodjem, kot sta Pandas in Apache Parquet, je mogoče združiti strukturirane in nestrukturirane podatke, kar poenostavi dostop do informacij in zmanjša izgubo pri analizi, medtem ko veliki podatkovni nizi iz dokumentov ali projektov CAD (slika 9.2-10 - slika 9.2-12) omogočajo zbiranje, analizo in napovedovanje podatkov v vseh fazah življenjskega cikla projekta.

Veliki podatki spreminjajo gradbeno industrijo in nanjo vplivajo na različne načine. Uporaba tehnologij velikih količin podatkov prinaša rezultate na številnih ključnih področjih, na primer na naslednjih:

- **Analiza naložbenega potenciala** - napovedovanje donosnosti in vračilne dobe projektov na podlagi podatkov iz prejšnjih objektov.
- **Prediktivno vzdrževanje** - prepoznavanje verjetnih okvar opreme, preden do njih dejansko pride, kar zmanjšuje čas izpada.
- **Optimizacija dobavne verige** - napovedovanje motenj in izboljšanje logistične učinkovitosti.
- **Analize energetske učinkovitosti** - pomoč pri načrtovanju nizkoenergijskih stavb.
- **Varnostno spremmljanje** - uporaba senzorjev in nosljivih naprav za spremmljanje razmer na gradbišču.

- **Nadzor kakovosti** - spremjanje skladnosti s procesnimi standardi v realnem času.
- **Upravljanje delovne sile** - analiza uspešnosti in napovedovanje kadrovskih potreb.

V gradbeništvu je težko najti področje, kjer podatkovna analitika in napovedi ne bi bile iskane. Glavna prednost algoritmov za napovedovanje je njihova zmožnost samoučenja in nenehnega izboljševanja ob kopičenju podatkov.

V bližnji prihodnosti umetna inteligenca ne bo le pomagala gradbenikom, temveč bo sprejemala ključne odločitve - od postopkov načrtovanja do vprašanj, povezanih z obratovanjem stavb.

Več o tem, kako se ustvarjajo napovedi in uporabljajo učni modeli, bo obravnavano v naslednjem delu knjige, "Strojno učenje in napovedi".

Za prehod na polnopravno delo z velikimi podatki je treba spremeniti sam pristop k analitiki. Medtem ko so se klasični sistemi, ki smo jih obravnavali doslej, osredotočali na vzročno-posledične odnose, se analitika velikih podatkov preusmerja na iskanje statističnih vzorcev in korelacij, ki lahko razkrijejo skrite odnose in predvidijo obnašanje objektov tudi brez popolnega razumevanja vseh dejavnikov.

## **Vprašanje izvedljivosti velikih podatkov: korelacija, statistika in vzorčenje podatkov**

Tradicionalno je gradnja temeljila na subjektivnih hipotezah in osebnih izkušnjah. Inženirji so z določeno stopnjo verjetnosti predvidevali, kako se bo material obnašal, kakšne obremenitve bo konstrukcija prenesla in kako dolgo bo projekt trajal. Te predpostavke so se preverjale v praksi, pogosto na račun časa, sredstev in prihodnjih tveganj.

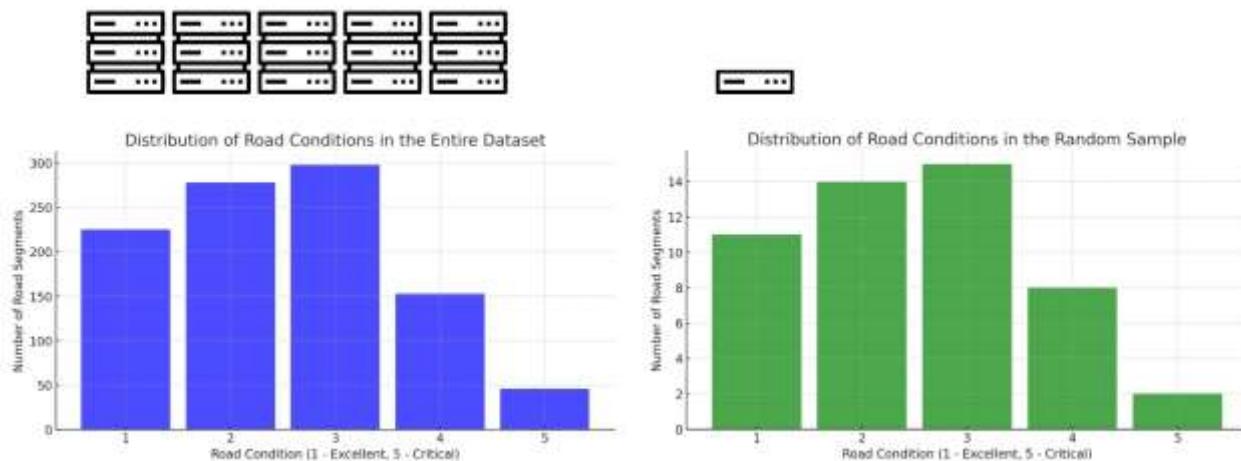
S pojavom velikih količin podatkov se pristop močno spreminja: odločitve se ne sprejemajo več na podlagi intuitivnih slutenj, temveč na podlagi analize obsežnih podatkovnih nizov. Gradbeništvo postopoma preneha biti umetnost intuicije in postaja natančna znanost napovedovanja.

Pri prehodu na idejo o uporabi velikih količin podatkov se neizogibno pojavi pomembno vprašanje: kako pomembna je količina podatkov in koliko informacij resnično potrebujemo za zanesljivo napovedno analitiko? Splošno razširjeno prepričanje, da "več kot je podatkov, večja je natančnost", se v praksi ne izkaže vedno za statistično veljavno.

Že leta 1934 je statistik Jerzy Neumann dokazal [145], da ključ do natančnosti statističnega sklepanja ni toliko v količini podatkov kot v njihovi reprezentativnosti in naključnosti vzorčenja.

To še posebej velja za gradbeno industrijo, kjer se velike količine podatkov zbirajo z uporabo senzorjev, skenerjev, nadzornih kamer, dronov in celo večformatnih modelov CAD, kar povečuje tveganje za nastanek slepih peg, odstopanj in izkriviljanja podatkov.

Poglejmo primer spremljanja stanja cestne površine. Celoten nabor podatkov o vseh cestnih odsekih lahko obsega X GB, njegova obdelava pa traja približno en dan. Hkrati bi naključni vzorec, ki bi vključeval le vsak 50. odsek ceste, potreboval le  $X/50$  GB in bi bil obdelan v pol ure, hkrati pa bi zagotavljal podobno natančnost ocen za nekatere izračune (slika 9.1-1).



Slika 9.1-1 Histogrami stanja vozišča: celoten niz podatkov in naključno vzorčenje kažeta enake rezultate.

Zato ključ do uspešne analize podatkov pogosto ni količina podatkov, temveč reprezentativnost vzorca in kakovost uporabljenih metod obdelave. Prehod na naključno vzorčenje in bolj selektiven pristop zahtevata spremembo razmišljanja v gradbeništvu. V preteklosti so podjetja sledila logiki "čim več podatkov, tem bolje", saj so verjela, da bo zajemanje vseh možnih kazalnikov povečalo natančnost.

Ta pristop spominja na priljubljeno napačno predstavo iz projektnega vodenja: "Več strokovnjakov kot bom pritegnil, bolj učinkovito bo delo". Vendar sta tako kot pri človeških virih kakovost in orodja pomembnejša od količine. Brez upoštevanja medsebojnih povezav (korelacijs) med podatki ali udeleženci projekta lahko povečevanje obsega privede le do šuma, izkrivljanja, podvajanja in nepotrebne zapravljanja.

Na koncu se pogosto izkaže, da je veliko bolj produktivno imeti manjši, vendar kakovostno pripravljen nabor podatkov, ki omogoča pripravo stabilnih in razumnih napovedi, kot pa se zanašati na obsežne, vendar kaotične informacije, ki vsebujejo veliko nasprotujočih si signalov.

Prevelika količina podatkov ne zagotavlja le večje natančnosti, temveč lahko privede tudi do izkrivljenih zaključkov zaradi prisotnosti šuma, odvečnih lastnosti, skritih korelacji in nepomembnih informacij. V takšnih okoliščinah se poveča tveganje pretiranega prilagajanja modelov in zmanjša zanesljivost analitičnih rezultatov.

V gradbeništvu je glavni izviv pri obravnavi velikih količin podatkov določitev optimalne količine in kakovosti podatkov. Na primer, pri spremeljanju stanja betonskih konstrukcij lahko uporaba na tisoče senzorjev in zbiranje informacij vsako minuto preobremenita sistem za shranjevanje in analizo. Če pa izvedete korelacijsko analizo in izberete 10% najbolj informativnih senzorjev, lahko dosežete skoraj

identično natančnost napovedi, pri čemer porabite mnogokrat, včasih desetkrat in stokrat, manj sredstev.

Z uporabo manjšega podskupine podatkov se zmanjšata količina potrebnega pomnilnika in čas obdelave, kar znatno zmanjša stroške shranjevanja in analiziranja podatkov, zato je naključno vzorčenje pogosto idealna rešitev za napovedno analitiko, zlasti pri velikih infrastrukturnih projektih ali pri delu v realnem času. Navsezadnje učinkovitost gradbenih procesov ni odvisna od količine zbranih podatkov, temveč od kakovosti njihove analize. Brez kritičnega pristopa in skrbne analize lahko podatki privedejo do napačnih zaključkov.

*Po določeni količini podatkov vsaka nova informacijska enota daje vedno manj uporabnih rezultatov. Namesto neskončnega zbiranja informacij se je treba osredotočiti na njihovo reprezentativnost in metode analize (slika 9.2-2).*

Ta pojav je dobro opisal Allen Wallis [146], ki je uporabo statističnih metod ponazoril s primerom testiranja dveh alternativnih zasnov izstrelkov ameriške mornarice.

*Vojna mornarica je preizkusila dve alternativni zasnovi izstrelkov (A in B) s serijo parnih strelov. V vsakem krogu A dobi oceno 1 ali 0, odvisno od tega, ali je njegova zmogljivost boljša ali slabša od zmogljivosti B, in obratno. Standardni statistični pristop vključuje izvedbo določenega števila poskusov (npr. 1000) in določitev zmagovalca na podlagi odstotne porazdelitve (npr. če A dobi 1 v več kot 53% primerov, velja za najboljšega). Ko je Allen Wallis o takšnem problemu razpravljal s (mornariškim) kapitanom Garrettom L. Schuylerjem, je kapitan ugovarjal, da bi bil takšen test, če citiram Allenovo zgodbo, lahko neuporaben. Če bi bil na mestu moder in izkušen častnik za strelivo, kot je bil Schuyler, bi po prvih nekaj sto [strelah] videl, da poskusa ni treba prekiniti, bodisi zato, ker je nova metoda očitno slabša, bodisi zato, ker je očitno boljša od tiste, na katero se je upalo [146].*

- Statistična raziskovalna skupina vlade ZDA na Univerzi Columbia, obdobje druge svetovne vojne

To načelo se pogosto uporablja v različnih panogah. V medicini se na primer klinična preskušanja novih zdravil izvajajo na naključnih vzorcih bolnikov, kar omogoča pridobitev statistično pomembnih rezultatov, ne da bi bilo treba zdravilo testirati na celotni populaciji ljudi, ki živijo na planetu. V ekonomiji in sociologiji se izvajajo reprezentativne raziskave, ki odražajo mnenje družbe, ne da bi bilo treba anketirati vse prebivalce države.

Tako kot vlade in raziskovalne organizacije izvajajo raziskave na majhnih populacijah, da bi razumele splošne družbene tendre, lahko podjetja v gradbeništvu uporabijo naključne vzorce podatkov za učinkovito spremmljanje in oblikovanje napovedi za vodenje projektov (slika 9.1-1).

*Veliki podatki lahko spremenijo pristop k družboslovju, vendar ne bodo nadomestili statistične zdrave pameti [147].*

- Thomas Landsall-Welfair, "Napovedovanje trenutnega razpoloženja naroda", *Significance* v. 9(4), 2012 .

Z vidika varčevanja z viri je pri zbiranju podatkov za prihodnje napovedi in odločanje pomembno odgovoriti na vprašanje: ali je smiselno porabiti veliko sredstev za zbiranje in obdelavo velikih podatkovnih nizov, če je mogoče uporabiti veliko manjši in cenejši testni podatkovni niz, ki ga je mogoče postopoma povečevati? Učinkovitost naključnega vzorčenja kaže, da lahko podjetja z izbiro metod zbiranja podatkov, ki ne zahtevajo celovite pokritosti, vendar še vedno zagotavljajo zadostno natančnost in reprezentativnost, zmanjšajo stroške zbiranja in usposabljanja modelov za desetkrat ali celo tisočkrat. Ta pristop omogoča tudi majhnim podjetjem, da z bistveno manjšimi sredstvi in količino podatkov dosežejo rezultate, enakovredne velikim korporacijam, kar je pomembno za podjetja, ki želijo z majhnimi sredstvi optimizirati stroške in pospešiti informirano odločanje. V naslednjih poglavjih raziščite primere analitike in napovedne analitike na podlagi javnih podatkovnih zbirk z uporabo orodij za obdelavo velikih količin podatkov.

## **Veliki podatki: analiza podatkov iz zbirke podatkov o milijonih gradbenih dovoljenj v San Franciscu**

Delo z odprtimi zbirkami podatkov je edinstvena priložnost za praktično uporabo načel, obravnavanih v prejšnjih poglavjih: preudarna izbira značilnosti, reprezentativno vzorčenje, vizualizacija in kritična analiza. V tem poglavju bomo raziskali, kako je mogoče z odprtimi podatki raziskati zapletene pojave, kot je gradbena dejavnost v velikem mestu - zlasti več kot milijon zapisov o gradbenih dovoljenjih v San Franciscu

Javno dostopni podatki o več kot milijonu gradbenih dovoljenj (slika 9.1-2) (zapisi v dveh zbirkah podatkov v formatu CSV) iz "San Francisco Department of Buildings". [148] nam omogočajo, da neobdelano tabelo CSV uporabimo ne le za analizo gradbene dejavnosti v mestu, temveč tudi za kritično analizo nedavnih trendov in zgodovine gradbeništva v San Franciscu v zadnjih 40 letih, od leta 1980 do leta 2019.

Primeri kode, ki so bili uporabljeni za izdelavo vizualizacij podatkovnih nizov (slike 9.1-3- Slike 9.1-8), ter vizualni grafi s kodo, pojasnili in komentarji so na voljo na platformi Kaggle z iskanjem "San Francisco. Gradbeni sektor 1980-2019." [149].

**count** 1.137695e+06

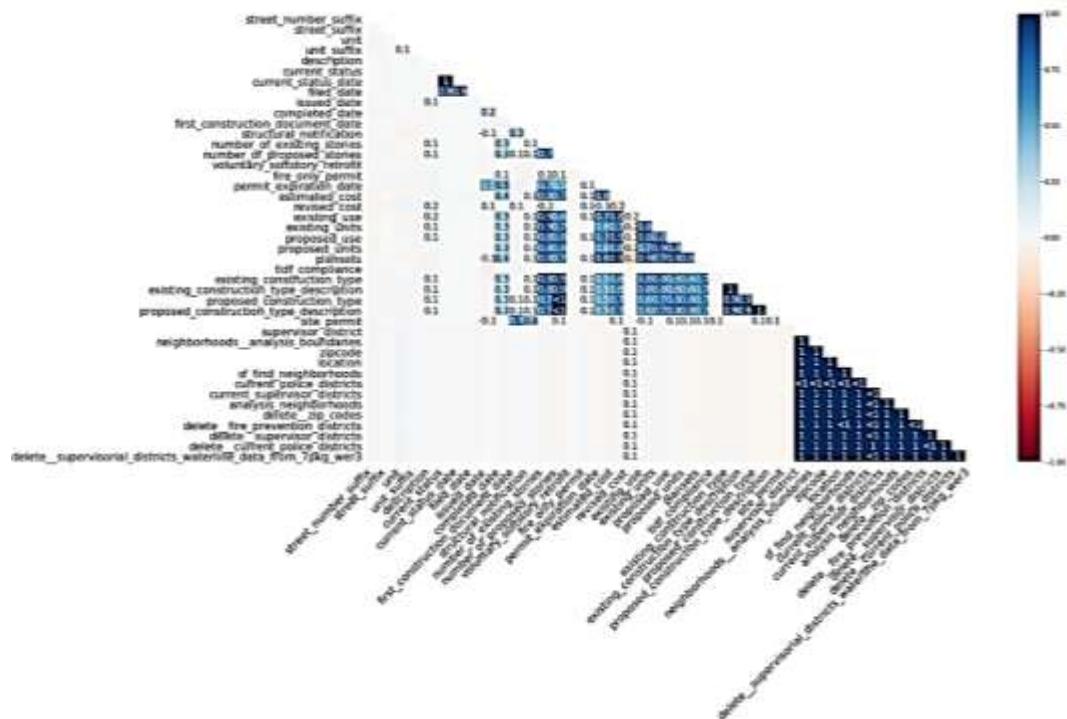
Building Permits on or after January 1, 2013

Building Permits before January 1, 2013

June 13, 2020 2,237 Views June 13, 2020 800 Views

permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date
07/01/1998	repair stucco	complete		07/07/1998	07/01/1998	07/01/1998
12/13/2004	reroofing	expired		01/24/2006	12/13/2004	12/13/2004
02/18/1992	install auto fire spks.	complete		06/29/1992	02/18/1992	03/18/1992
permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location
362780	9812394	11/01/1998	780.0	NaN	94123.0	(37.7962468780498, -122.4322541443574)
570817	200412131233	06/13/2005	9000.0	9000.0	94127.0	(37.729258516008388, -122.4644245957462)
198411	9202396	09/18/1992	9000.0	NaN	94111.0	(37.795086002552974, -122.39593224461805)

Slika 9.1-2 Zbirke podatkov vsebujejo informacije o izdanih gradbenih dovoljenjih z različnimi atributi objektov.

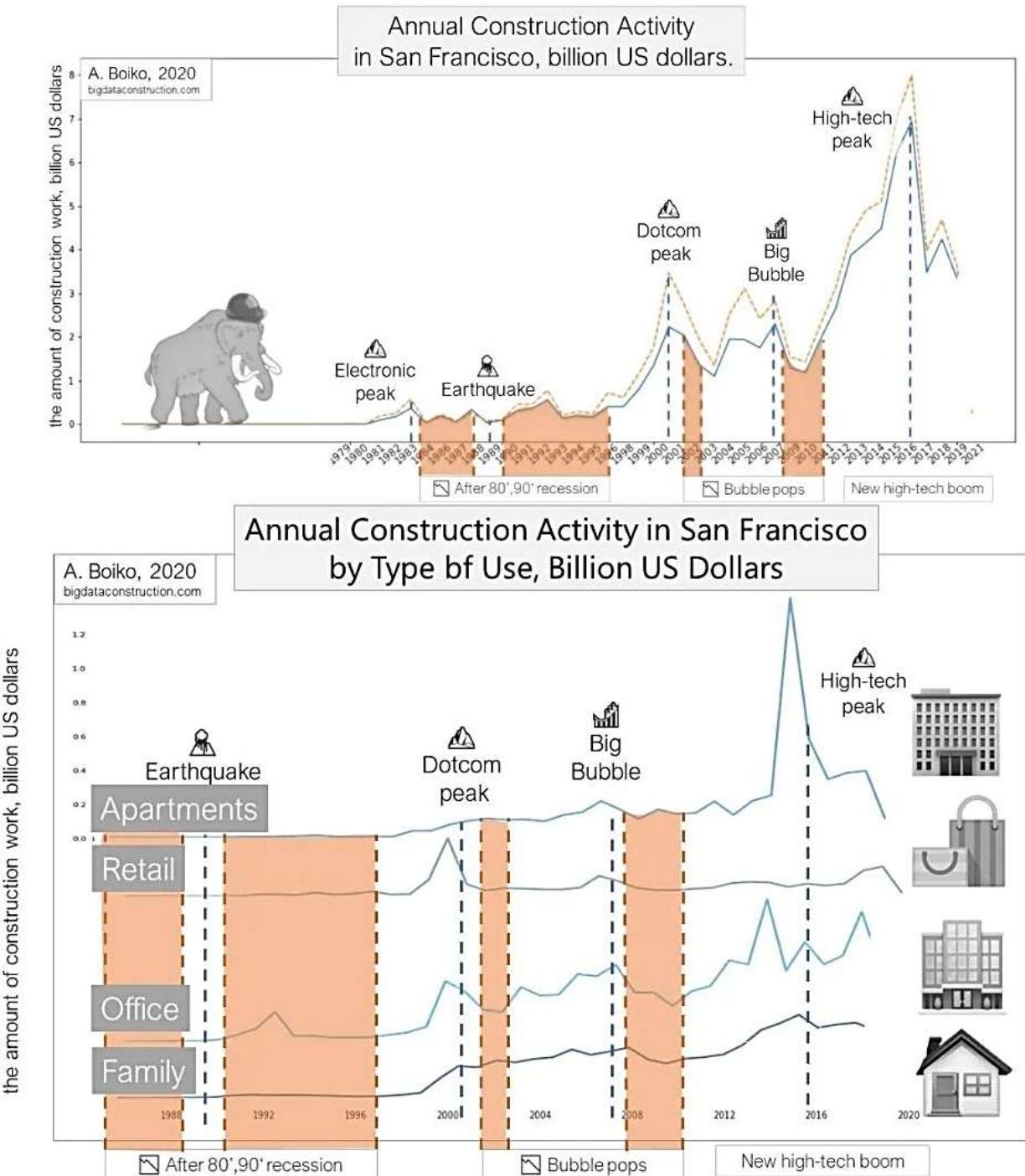


Slika 9.1-3 Toplotni zemljevid (Pandas in Seaborn), ki vizualizira vse atribute nabora podatkov in pomaga prepozнатi povezave med pari atributov.

Iz preglednice, ki jo je predložil Oddelek za stavbe v San Franciscu (slika 9.1-2), niso razvidni nobeni trendi ali zaključki. Suhe številke v obliki tabele niso podlaga za sprejemanje odločitev. Da bi bili podatki vizualno razumljivi, kot je podrobno obravnavano v poglavjih o vizualizaciji podatkov, jih je treba vizualizirati z uporabo različnih knjižnic, obravnavanih v sedmem delu knjige na temo "ETL in vizualizacija rezultatov v obliki grafov".

Z analizo podatkov z uporabo knjižnic Pandas DataFrame in Python za vizualizacijo vrednosti 1 137 695 dovoljenj [148] lahko ugotovimo, da je gradbena dejavnost v San Franciscu tesno povezana z gospodarskimi cikli, zlasti v cvetoči tehnološki industriji Silicijeve doline (slika 9.1-4).

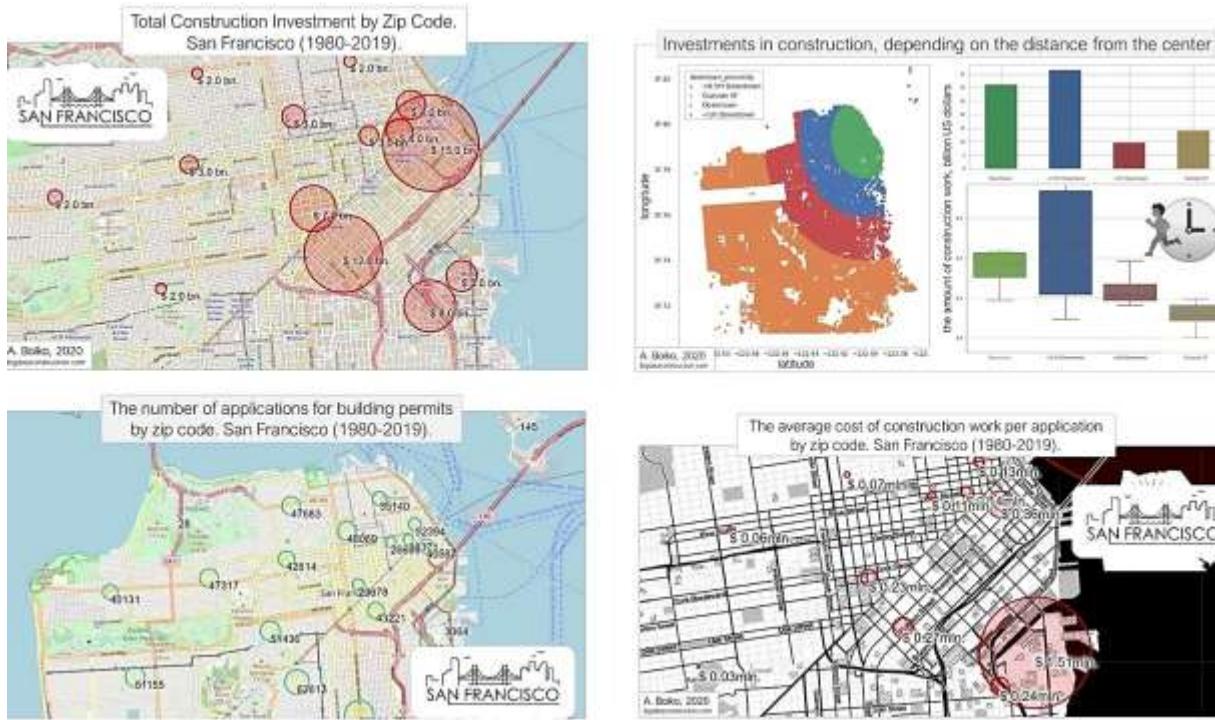
Gospodarski vzponi in padci pomembno vplivajo na število in vrednost gradbenih projektov. Na primer, prvi vrhunec gradbene dejavnosti je sovpadal z razcvetom elektronike sredi osemdesetih let prejšnjega stoletja (uporabljena Pandas in Matplotlib), naslednji vrhovi in padci pa so bili povezani z balonom dot-com in tehnološkim razcvetom v zadnjih letih.



Slika 9.1-4 V nepremičninskem sektorju San Francisca so naložbe povezane s tehnološkim razvojem Silicijeve doline.

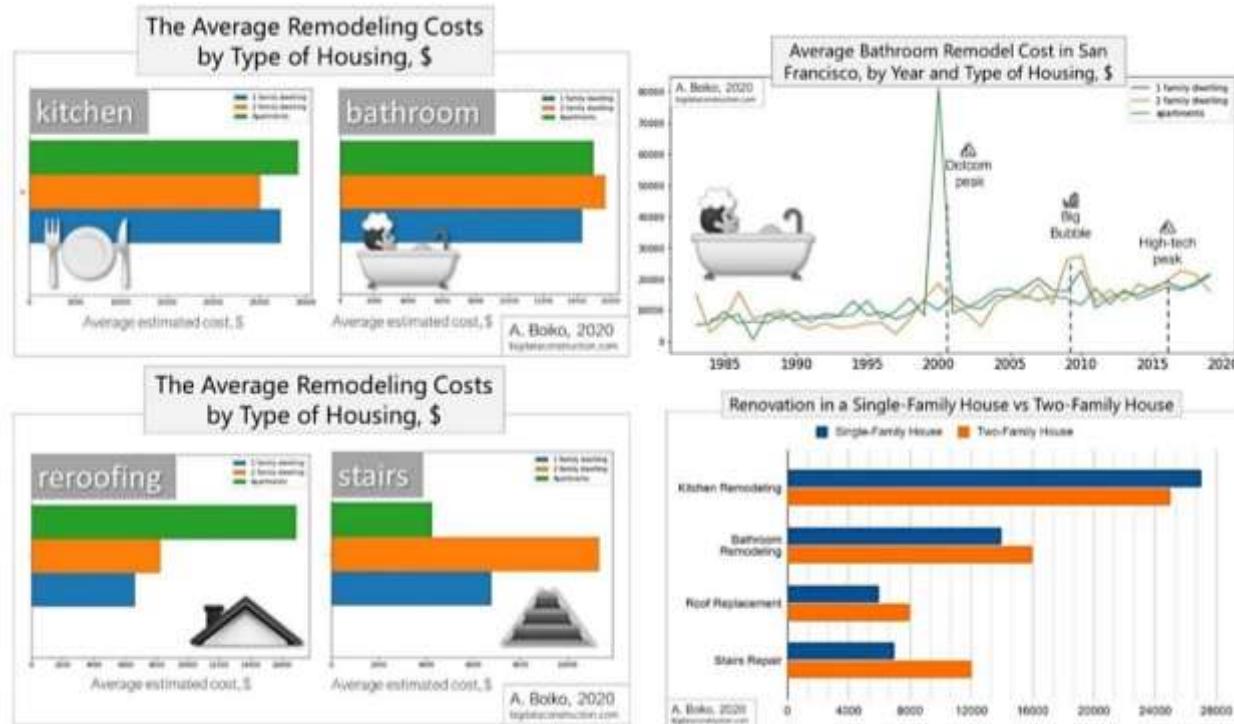
Analiza podatkov kaže, da je v San Franciscu večina 91,5 milijarde dolarjev, vloženih v gradnjo in prenovo v zadnjem desetletju, tj. skoraj 75%, skoncentrirana v središču mesta (slika 9.1-5 - uporabljeni sta Pandas in vizualizacijska knjižnica Folium) in v radiju 2 km od središča mesta, kar odraža večjo gostoto naložb na teh osrednjih območjih.

Povprečni stroški gradbenih dovoljenj se precej razlikujejo po soseskah, pri čemer so vloge v mestnem središču trikrat dražje kot vloge zunaj mestnega središča zaradi višjih stroškov zemljišč, dela in materialov ter strogih gradbenih predpisov, ki zahtevajo dražje materiale za izboljšanje energetske učinkovitosti.



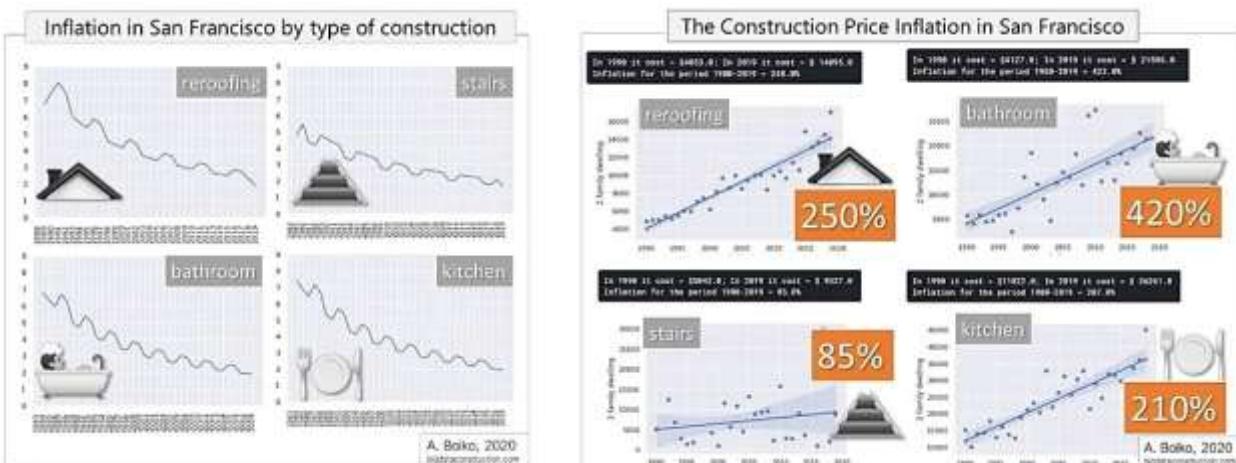
**Slika 9.1-5 V San Franciscu je 75 odstotkov gradbenih investicij (91,5 milijarde dolarjev) skoncentriranih v središču mesta.**

Nabor podatkov omogoča tudi izračun povprečnih cen popravil ne le po vrstah hiš, temveč tudi po mestnih četrtih in posameznih naslovih (poštih številkah). V San Franciscu dinamika stroškov obnove stanovanj kaže različne tende za različne vrste obnov in stanovanj (slika 9.1-6 - uporabljena Pandas in Matplotlib). Prenova kuhinje je opazno dražja od prenove kopalnice: povprečna prenova kuhinje v enodružinski hiši stane približno 28.000 USD, medtem ko v dvostanovanjski hiši 25.000 USD.



Slika 9.1-6 V SF stane prenova kuhinje skoraj dvakrat več kot prenova kopalnice, lastniki stanovanj pa morajo za kritje stroškov večjih stanovanjskih popravil 15 let vsak mesec odlagati na stran \$ 350.

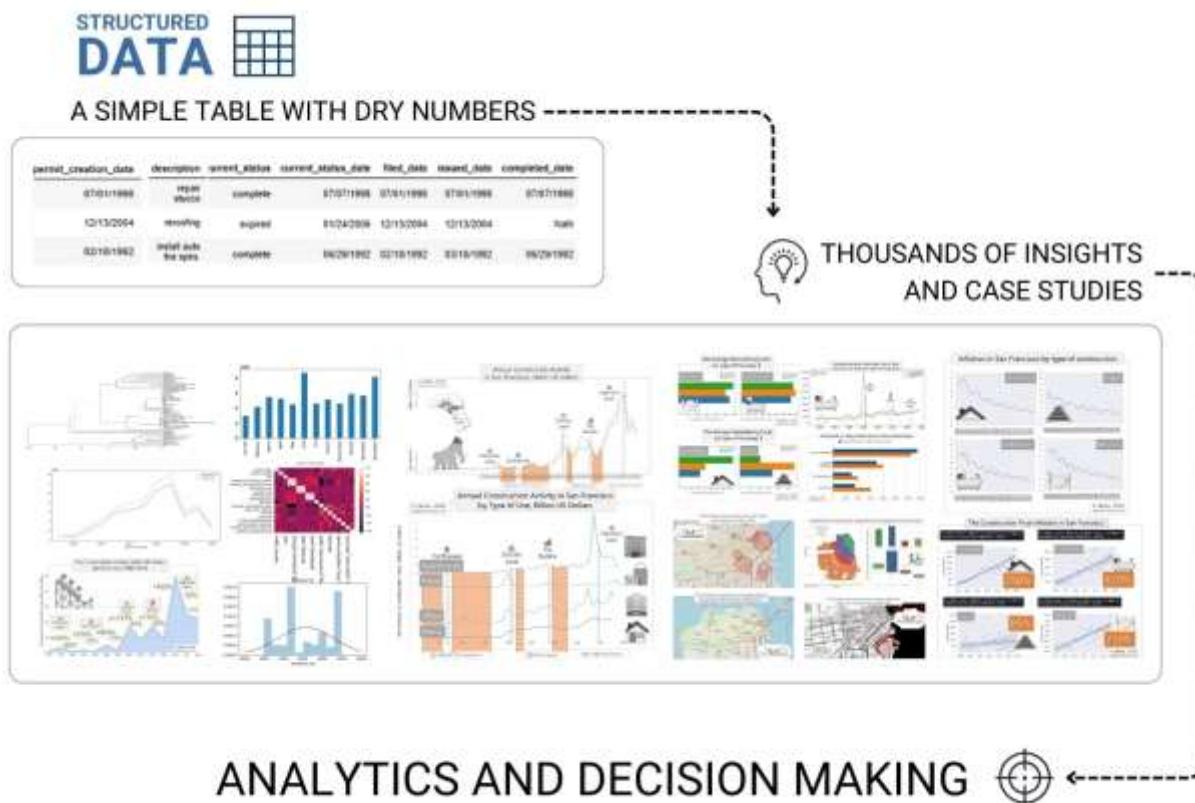
Inflacijo gradbenih stroškov v San Franciscu skozi leta je mogoče zasledovati z analizo podatkov, razvrščenih po vrsti stanovanj in letih (slika 9.1-7 - uporabljena v publikacijah Pandas in Seaborn), ki kaže stalno rast povprečnih stroškov popravil od leta 1990 in razkriva kratkoročne triletne cikle pri stroških popravil večstanovanjskih stavb.



Slika 9.1-7 Od leta 1980 do leta 2019 so se stroški prenove kopalnice v SF povečali za petkrat, medtem ko so se stroški prenove strehe in kuhinje potrojili, stroški prenove stopnišča pa povečali le za 85%.

Študija javnih podatkov gradbenega oddelka San Francisca (slika 9.1-3) razkriva, da so stroški gradnje v mestu zelo spremenljivi in pogosto nepredvidljivi, na kar vplivajo različni dejavniki. Ti dejavniki vključujejo gospodarsko rast, tehnološke inovacije in edinstvene zahteve različnih vrst stanovanj.

V preteklosti so takšne analize zahtevale poglobljeno znanje programiranja in analitike. S pojavom orodij LLM- pa je postopek postal dostopen in razumljiv širokemu krogu strokovnjakov v gradbeništvu, od inženirjev v projektnih oddelkih do višjega vodstva.



Slika 9.1-8 Prehod na vizualno razumljive podatke omogoča avtomatizirano odločanje s prepoznavanjem skritih vzorcev.

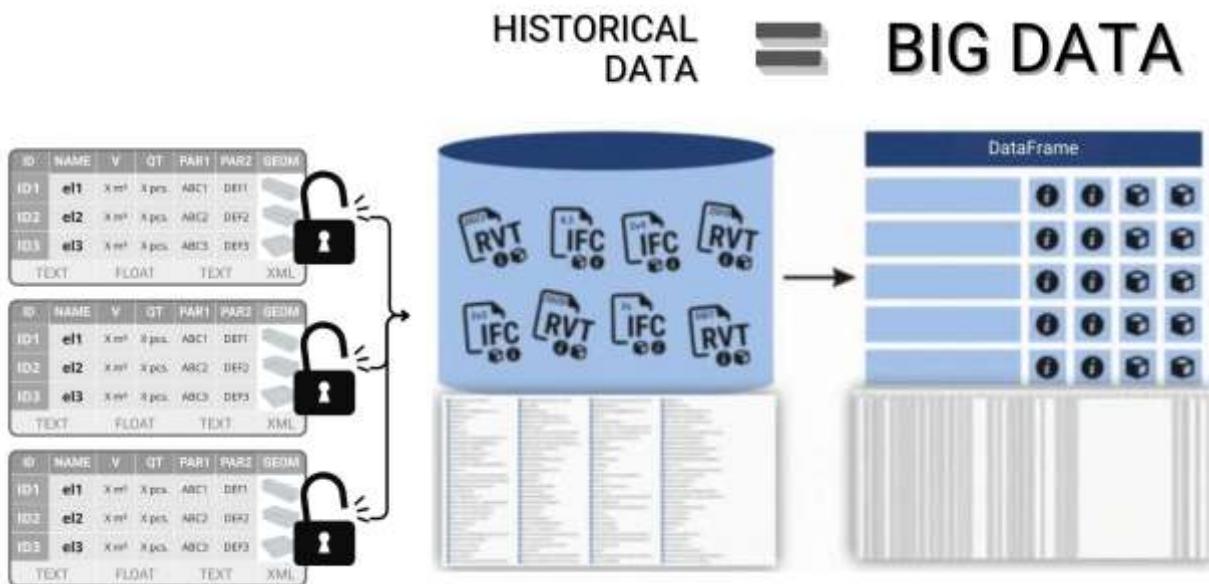
Tako kot smo analizirali podatke iz tabeličnega nabora podatkov "San Francisco Building Authority", lahko vizualiziramo in analiziramo kateri koli nabor podatkov - od slik in dokumentov do podatkov IoT ali podatkov iz izpeljanih podatkovnih zbirk CAD.

## Primer velikih podatkov na podlagi podatkov CAD (BIM)

V naslednjem primeru bomo analizirali velik nabor podatkov z uporabo podatkov iz različnih orodij CAD (BIM). Za zbiranje in oblikovanje velikega nabora podatkov je bil uporabljen specializiran samodejni spletni pregledovalnik (skripta), ki je bil konfiguriran za samodejno iskanje in zbiranje projektnih datotek s spletnih strani, ki ponujajo brezplačne arhitekturne modele v formatih RVT in IFC. V nekaj dneh je iskalnik uspešno našel in prenesel 4 596 datotek IFC in 6 471 datotek RVT ter 156 024 datotek

DWG[149].

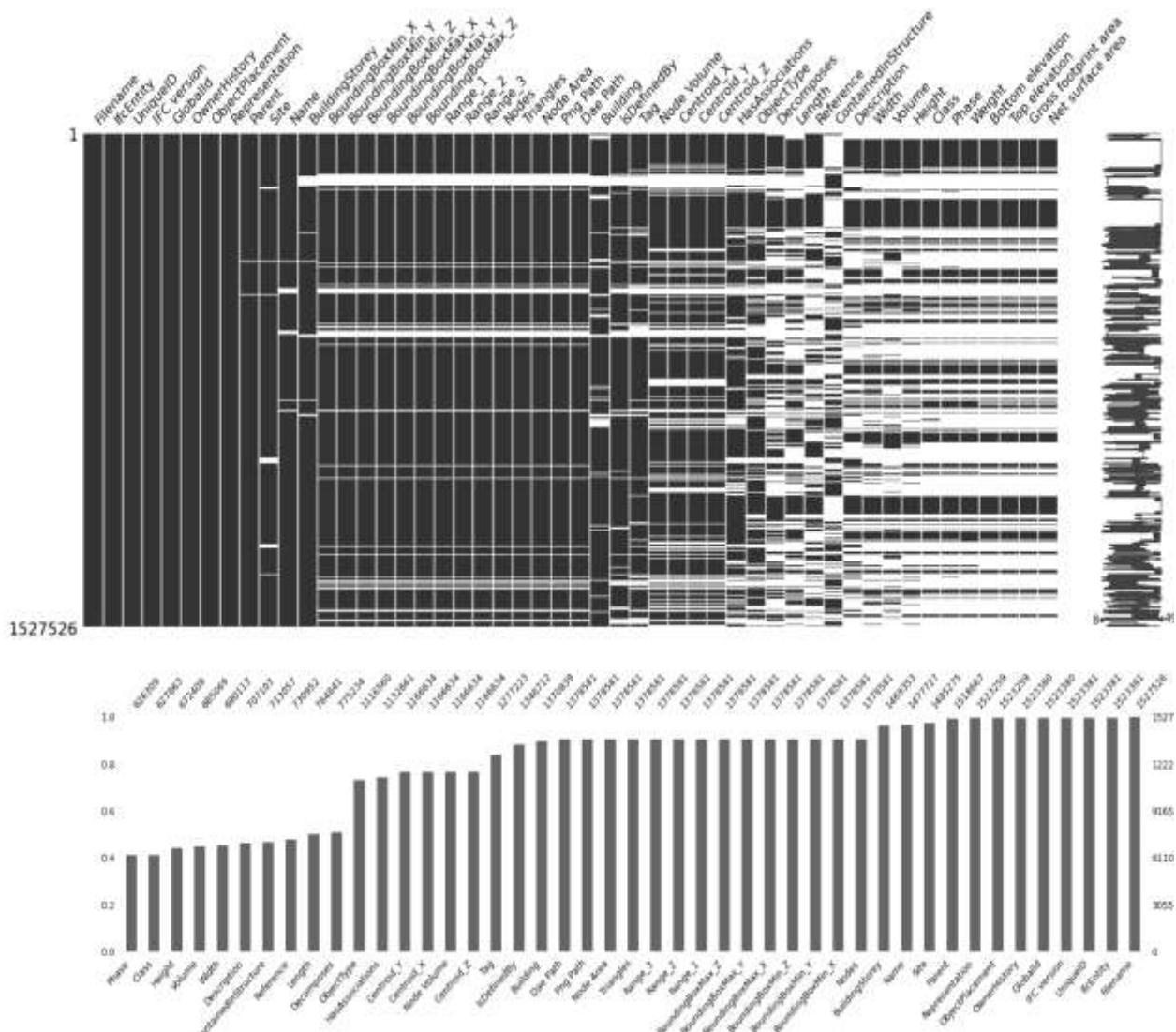
Po zbiranju projektov v formatih RVT in IFC različnih različic in pretvorbi v strukturiran format CSV z uporabo brezplačnih kompletov SDK za povratni inženiring je bilo skoraj 10 tisoč projektov RVT in IFC zbranih v eno veliko datoteko tabele Apache Parquet in za analizo prenesenih v datoteko Pandas DataFrame (slika 9.1-9).



**Slika 9.1-9 Strukturirani podatki podatki o projektih omogočajo združevanje poljubnega števila projektov v eno samo dvodimenzionalno tabelo.**

Podatki iz te obsežne zbirke vsebujejo naslednje informacije: zbirka datotek IFC vsebuje približno 4 milijone entitet (vrstic) in 24 962 atributov (stolpcev), zbirka datotek RVT, sestavljena iz približno 6 milijonov entitet (vrstic), pa vsebuje 27 025 različnih atributov (stolpcev).

Ti sklopi informacij (slika 9.1-10) zajemajo milijone elementov, za vsakega od njih pa so bile dodatno pridobljene koordinate geometrije Bounding Box (pravokotnik, ki določa meje objekta v projektu) in dodane v skupno preglednico - koordinate geometrije Bounding Box (pravokotnik, ki določa meje objekta v projektu) ter slike vsakega elementa v formatu PNG in geometrije v odprttem formatu XML - DAE (Collada).



Slika 9.1-10 Subset 1,5 milijona elementov in vizualizacija (knjižnica missingno) zasedenosti prvih 100 atributov v obliki histograma.

Tako smo pridobili vse informacije o desetinah milijonov elementov iz 4 596 projektov IFC in 6 471 projektov RVT, kjer so bili vsi atributi - lastnosti vseh elementov entitet in njihova geometrija (Bounding Box) prevedeni v strukturirano obliko ene same tabele (DataFrame) (slika 9.1-10 - podatki na populacijah podatkovnega okvira so prikazani kot histogrami).

Histogrami (slika 9.1-10, slika 9.2-6, slika 9.2-7), izrisani med postopkom analize, omogočajo hitro oceno gostote podatkov in pogostosti pojavljanja vrednosti v stolpcih. Tako dobimo prvi vpogled v porazdelitev lastnosti, prisotnost izstopajočih vrednosti in potencialno uporabnost posameznih atributov pri analizi in gradnji modelov strojnega učenja.

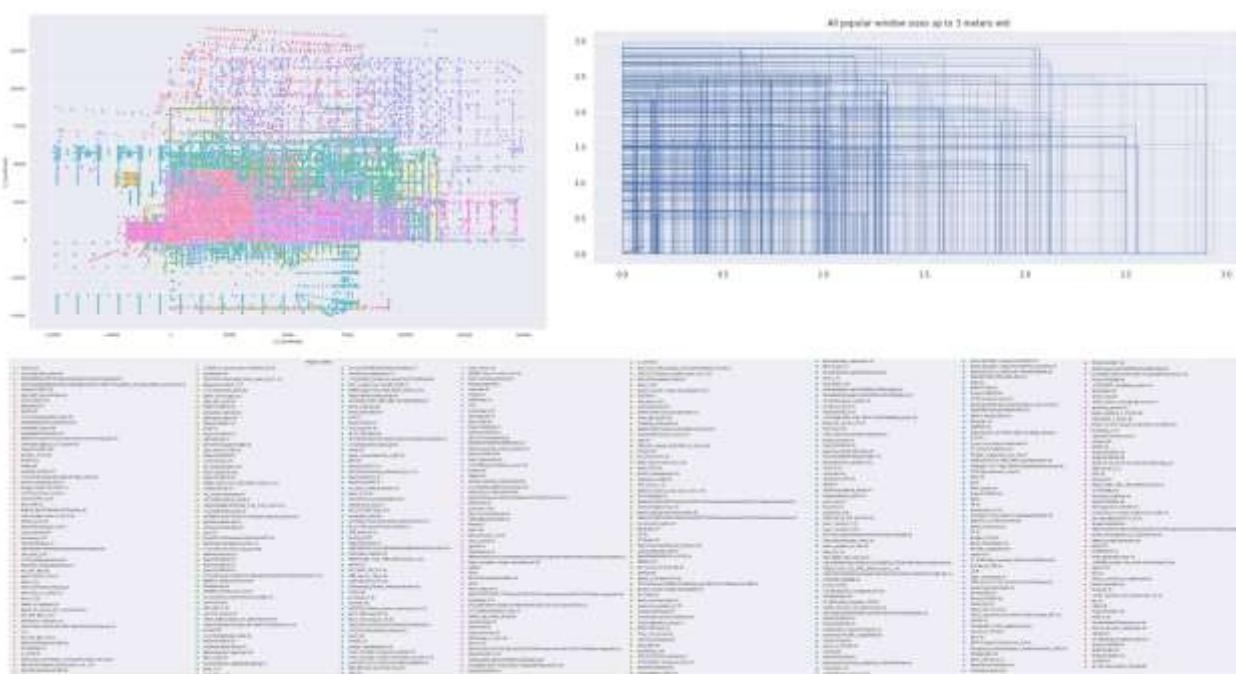
Primer praktične uporabe tega nabora podatkov (slika 9.1-10) je projekt "5000 IFC in RVT ". [149], ki je

na voljo na platformi Kaggle. Predstavlja beležnico Jupyter s celovito rešitvijo Pipeline: od predobdelave in analize podatkov do vizualizacije rezultatov z uporabo knjižnic Python - pandas, matplotlib, seaborn, folium in drugih (slika 9.1-11).



Slika 9.1-11 Primeri analize podatkov iz formatov CAD (BIM) z uporabo Pythonovih vizualizacijskih knjižnic in knjižnice pandas.

Na podlagi metainformacij je mogoče ugotoviti, v katerih mestih so bili razviti določeni projekti, in to prikazati na zemljevidu (npr. z uporabo knjižnice folium). Poleg tega časovni žigi v podatkih omogočajo raziskovanje vzorcev, kdaj so bile datoteke shranjene ali urejene: po dnevih v tednu, urah dneva in mesecih.



Slika 9.1-12 Vizualizacija geometrijskega položaja vseh stebrov in dimenzijskih vseh oken do 3 metrov v projektih s seznama na dnu dijagrama.

Geometrijski parametri v obliki omejitvenega polja, pridobljeni iz modelov, so prav tako primerni za skupne analize. Na sliki 9.1-12 sta na primer prikazana dva grafa: levi prikazuje porazdelitev razdalj med stolpci za vse projekte glede na ničelno točko, desni pa dimenzije vseh oken, visokih do 3 m, v vzorcu več deset tisoč okenskih elementov (po združevanju celotnega nabora podatkov po parametru "Kategorija" z vrednostjo "OST\_Windows", "IfcWindows").

Koda za analizo Pipeline za ta primer in sam nabor podatkov sta na voljo na spletnem mestu Kaggle pod naslovom "5000 IFC in RVT | DataDrivenCo-nstruction.io projects" [149]. Ta dokončani Pipeline skupaj z naborom podatkov lahko kopirate in brezplačno zaženete na spletu na portalu Kaggle ali brez povezave v enem od priljubljenih IDE: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z vtičnikom PyDev, Thonny, Wing IDE, IntelliJ IDEA z vtičnikom Python, JupyterLab ali priljubljenih spletnih orodij Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Analitični vpogledi, pridobljeni z obdelavo in preučevanjem velikih količin strukturiranih podatkov, bodo imeli ključno vlogo v procesih odločanja v gradbeništvu.

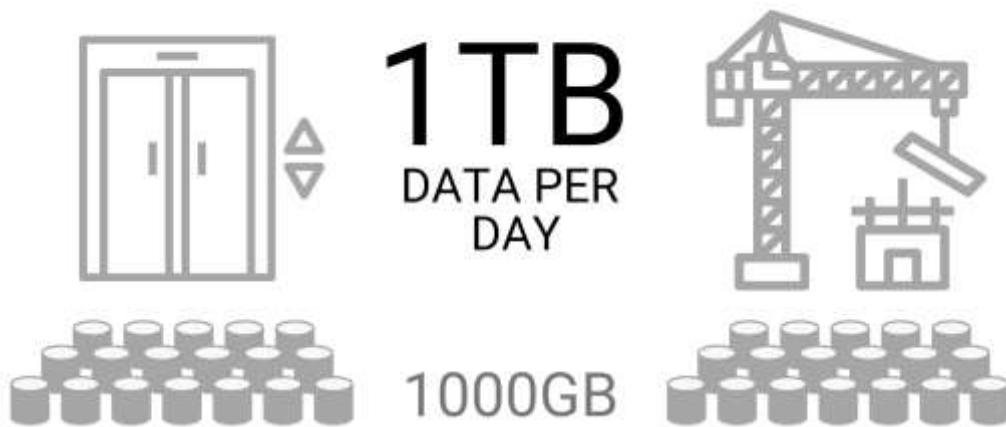
S tovrstno analizo informacij, ki temelji na preteklih projektih, lahko strokovnjaki učinkovito predvidijo na primer potrebe po materialu in delu ter optimizirajo projektne rešitve pred začetkom gradnje.

Medtem ko so projektni podatki ali gradbena dovoljenja relativno statične informacije, ki se spreminjajo razmeroma počasi, pa je sam proces gradnje hitro zasičen z različnimi senzorji in napravami IoT: kamerami, avtomatiziranimi nadzornimi sistemi, ki prenašajo podatke v realnem času - vse to gradbišče spreminja v dinamično digitalno okolje, kjer je treba podatke analizirati v realnem času.

## **Internet stvari Internet stvari in pametne pogodbe**

Internet stvari Internet stvari predstavlja nov val digitalne preobrazbe, v katerem vsaka naprava dobí svoj IP-naslov in postane del globalnega omrežja. Internet stvari je koncept, ki vključuje povezavo fizičnih predmetov z internetom za zbiranje, obdelavo in prenos podatkov. V gradbeništvu to pomeni možnost nadzora gradbenih procesov v realnem času, zmanjševanje izgube materiala, napovedovanje obrabe opreme in avtomatizacijo odločanja.

V članku CFMA "Priprava na prihodnost s povezano gradnjo" je navedeno [150] bo gradbena industrija v naslednjem desetletju doživelja veliko digitalno preobrazbo, ki se bo končala s konceptom povezanega gradbeništva - popolnoma integriranega in avtomatiziranega gradbišča.



Slika 9.1-13 IoT ali podatkovne naprave na gradbišču lahko dnevno ustvarijo in posredujejo terabajte podatkov.

Digitalno gradbišče pomeni, da bodo vsi elementi gradnje - od načrtovanja in logistike do izvajanja del in nadzora kakovosti na gradbišču z uporabo stacionarnih kamer in kvadrokopterjev - vključeni v enoten dinamičen digitalni ekosistem. Že prej, v sedmem delu te knjige, smo si ogledali zmogljivosti Apache NiFi (slika 7.4-5), brezplačnega in odprtakodnega orodja, ki omogoča pretok podatkov v realnem času - od zbiranja iz različnih virov do prenosa na platforme za shranjevanje ali analitiko.

Podatki o poteku gradnje, porabi materiala, stanju opreme in varnosti se bodo v realnem času prenašali v analitične sisteme (slika 9.1-13). To omogoča predvidevanje morebitnih tveganj, takojšnje odzivanje na odstopanja in optimizacijo procesov na gradbišču. Ključni elementi digitalnega gradbišča vključujejo:

- IoT - senzorji - spremljanje okoljskih parametrov, spremljanje gradbene opreme in nadzor delovnih pogojev.
- Digitalni dvojčki - virtualni modeli stavb in infrastrukture za napovedovanje morebitnih odstopanj in preprečevanje napak.
- Avtomatizirani logistični sistemi - upravljanje dobavne verige v realnem času za zmanjšanje izpadov in stroškov.
- Robotski gradbeni sistemi - uporaba avtonomnih strojev za opravljanje rutinskih in nevarnih nalog.

Robotizacija uporaba interneta stvari in koncept digitalnega gradbišča Connected Site (Construction) ne bodo le povečali učinkovitosti in zmanjšali stroškov, temveč tudi uvedli novo dobo varnosti, trajnostne gradnje in napovednega upravljanja projektov.

Ena najpomembnejših komponent interneta stvari so tudi oznake RFID (radiofrekvenčna identifikacija). Uporabljajo se za identifikacijo in sledenje materiala, strojev in celo osebja na gradbišču, kar povečuje preglednost in nadzor nad projektnimi viri.

RFID -tehnologija se uporablja za samodejno prepoznavanje predmetov s pomočjo radijskih signalov. Sestavljajo jo trije ključni elementi:

- RFID - oznake (pasivne ali aktivne) - vsebujejo edinstven identifikator in so pritrjene na materiale, orodja ali stroje.
- Skenerji so naprave, ki z oznak preberejo informacije in jih pošljejo v sistem.
- Centralizirana podatkovna zbirka - shranjuje informacije o lokaciji, statusu in gibanju predmetov.

Uporaba RFID v gradbeništvu:

- Samodejno evidentiranje materiala - oznake na pripravljenih betonskih izdelkih, armaturi ali paketih sendvičastih plošč omogočajo nadzor zalog in preprečujejo krajo.
- Nadzor dela osebja - RFID - priponke za zaposlene beležijo čas začetka in konca izmene, kar zagotavlja evidenco delovnega časa.
- Spremljanje opreme - RFID - sistem spremlja gibanje opreme, preprečuje izpade in izboljšuje učinkovitost logistike.

Ta sklop tehnologij dopolnjujejo pametne pogodbe, ki temeljijo na verigi blokov in avtomatizirajo plačila, nadzor dostave in skladnost pogodb brez posrednikov, kar zmanjšuje tveganje goljufij in zamud.

Ker danes ni skupnega podatkovnega modela, so pametne pogodbe preprosto koda, o kateri se udeleženci dogovorijo. S podatkovno usmerjenim pristopom pa je mogoče ustvariti skupni model pogodbenih parametrov, ga zakodirati v verigi blokov in avtomatizirati izpolnjevanje pogojev.

Na primer, v sistemu za upravljanje dobavne verige lahko pametna pogodba spremi dostojo pošiljke na podlagi senzorjev IoT in oznak RFID ter samodejno prenese plačilo, ko pošiljka prispe. Podobno lahko pametna pogodba na gradbišču na podlagi podatkov iz dronov ali gradbenih senzorjev zabeleži zaključek delovne faze - na primer namestitev armaturne palice ali betoniranje temeljev - in samodejno sproži naslednje plačilo izvajalcu brez potrebe po ročnih pregledih in papirnatih potrdilih.

Kljub novim tehnologijam in prizadevanjem mednarodnih organizacij za standardizacijo pa množica konkurenčnih standardov otežuje dogajanje na področju interneta stvari.

Glede na študijo družbe Cisco, objavljeno leta 2017 [151], se skoraj 60% pobud za internet stvari (IoT) ustavi v fazi preverjanja koncepta, le 26% podjetij pa meni, da so njihovi projekti interneta stvari v celoti uspešni. Poleg tega tretjina zaključenih projektov ne dosega zastavljenih ciljev in tudi po izvedbi niso priznani kot uspešni.

Eden ključnih razlogov je pomanjkanje interoperabilnosti med platformami, ki obdelujejo podatke iz različnih senzorjev. Zato podatki ostajajo izolirani v ločenih rešitvah. Alternativa temu pristopu, tako kot v drugih podobnih primerih (ki smo jih obravnavali v tej knjigi), je arhitektura, ki temelji na samih podatkih kot glavnem sredstvu.

Senzorji interneta stvari imajo ključno vlogo ne le pri spremljanju tehničnega stanja opreme, temveč tudi pri napovedni analitiki za zmanjšanje tveganj na gradbišču in izboljšanje splošne učinkovitosti procesov s predvidevanjem napak in odstopanj.

Podatke, ki jih zbirajo senzorji interneta stvari in oznake RFID, lahko v realnem času obdelajo algoritmi strojnega učenja, ki lahko zaznajo nepravilnosti in inženirje vnaprej opozorijo na morebitne okvare. Te lahko segajo od mikrorazpok v betonskih strukturah do neobičajnih prekinitev v delovanju stolpnih žerjavov, ki kažejo na tehnične okvare ali kršitve predpisov. Poleg tega lahko napredni algoritmi za analizo vedenja zajamejo vedenjske vzorce, ki lahko kažejo na primer na fizično utrujenost osebja, kar izboljša proaktivno upravljanje varnosti in dobrega počutja zaposlenih na gradbišču.

V gradbeništvu se nesreče in okvare - bodisi strojev bodisi ljudi - redko zgodijo nenadoma. Običajno jih spremljajo manjša odstopanja, ki ostanejo neopažena. Prediktivna analitika in strojno učenje omogočata zgodnje zaznavanje teh signalov, še preden pride do kritičnih posledic.

Dokumenti, projektne datoteke in podatki iz naprav IoT in oznak RFID tvorijo digitalni odtis gradbenih projektov, strojno učenje pa lahko iz njih pridobi koristne informacije. Z naraščanjem količine podatkov in demokratizacijo dostopa do njih gradbena industrija pridobiva nove priložnosti na področju analitike, napovedne analitike in aplikacij umetne inteligence.

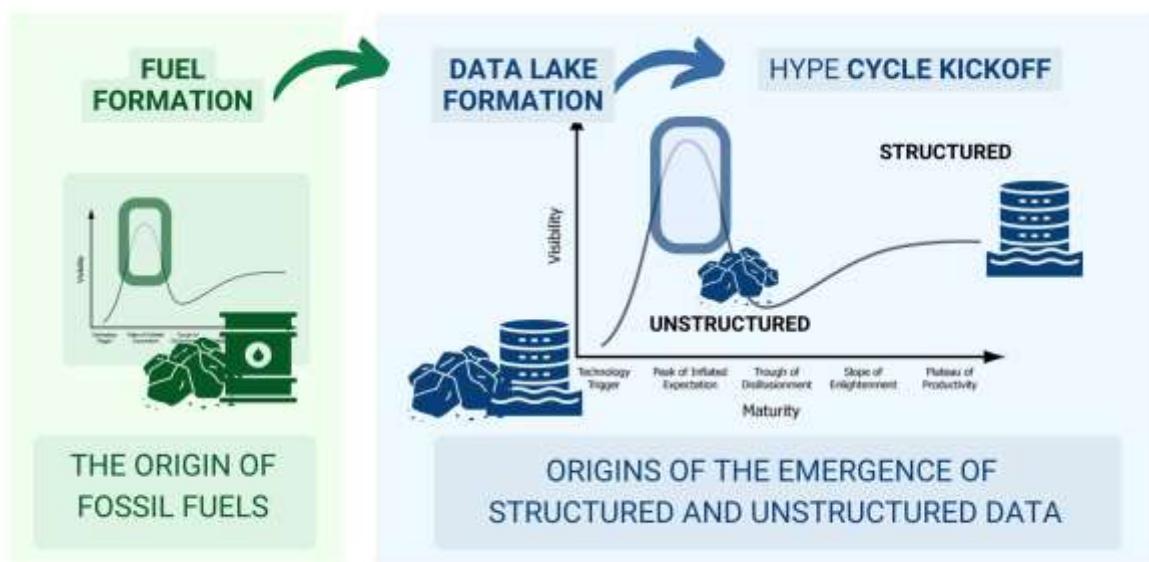


## POGLAVJE 9.2.

# STROJNO UČENJE IN NAPOVEDI

### Strojno učenje in umetna inteligenco bosta spremenila način gradnje

Podatkovne zbirke različnih sistemov v gradbeništvu - z neizogibno razpadajočo in vse bolj zapleteno infrastrukturo - postajajo plodna podlaga za prihodnje rešitve. Strežniki podjetij so kot gozd bogati z biomaso pomembnih informacij, ki so pogosto skrite pod zemljo, v drobovju map in strežnikov. Množice podatkov iz različnih sistemov, ki nastajajo danes po uporabi, po padcu na dno strežnika in po letih fosilizacije - bodo v prihodnosti pogon za strojno učenje in jezikovne modele. Interne klepetalnice v podjetju (npr. ločen primerek lokalno konfiguriranega ChatGPT, LlaMa, Mistral, DeepSeek) bodo zgrajene na podlagi teh internih modelov, ki bodo uporabljali centralizirano hrambo za hitro in priročno pridobivanje informacij ter izdelavo potrebnih grafov, nadzornih plošč in dokumentov.



Slika 9.2-1 Tako kot se drevesa spreminja v premog, se tudi informacije pod pritiskom časa in analitike sčasoma spremenijo v dragoceno poslovno energijo.

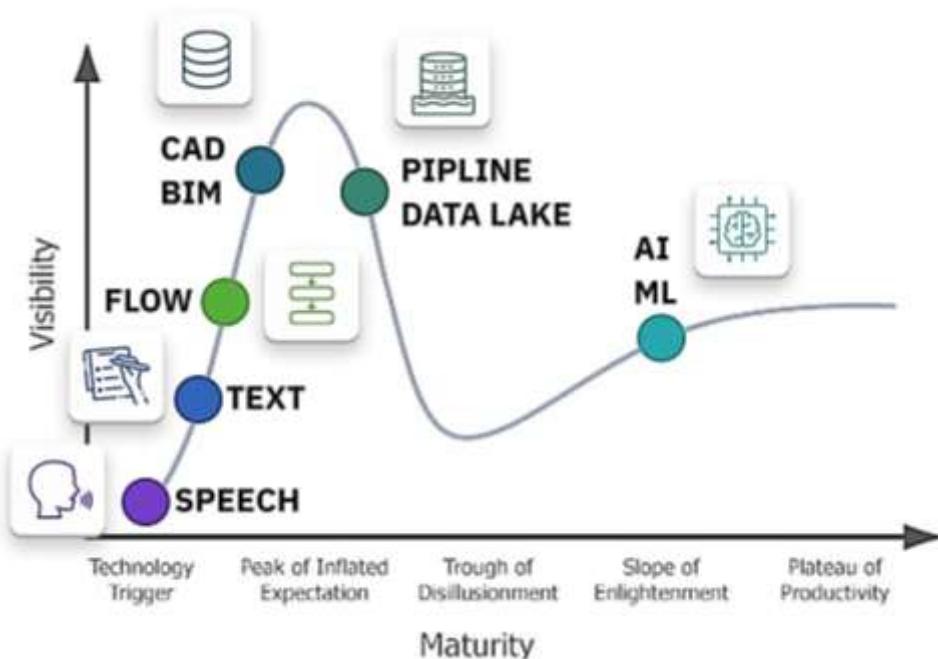
Fosilizacija rastlinske mase v kombinaciji s tlakom in temperaturo ustvari homogeno in edinstveno strukturirano homogeno maso dreves različnih vrst, ki so živela v različnih časih - oglje [152]. Na enak način informacije, zapisane na trdih diskih v različnih formatih in ob različnih časih pod pritiskom analitičnih oddelkov in temperaturo upravljanja kakovosti, sčasoma tvorijo homogeno strukturirano maso dragocenih informacij (slika 9.2-1).

Te plasti (ali pogosteje izolirane samorodne drobce) informacij ustvarijo izkušeni analitiki s skrbno organizacijo podatkov, ki začnejo postopoma iz na videz dolgo nepomembnih podatkov izluščiti dragocene informacije.

V trenutku, ko ti zreli podatkovni sloji niso več le "zažgani" v poročilih, temveč začnejo krožiti v

poslovnih procesih, bogatiti odločitve in izboljševati procese, je podjetje pripravljeno na naslednji korak - prehod na strojno učenje in umetno inteligenco (slika 9.2-2).

Strojno učenje (ML - Machine learning) je razred metod za reševanje problemov umetne inteligence. Algoritmi strojnega učenja prepoznavajo vzorce v velikih zbirkah podatkov in jih uporabljajo za lastno učenje. Z vsakim novim naborom podatkov se matematični algoritmi izboljšajo in prilagodijo glede na pridobljene informacije, kar omogoča stalno izboljševanje natančnosti priporočil in napovedi.



**Slika 9.2-2 Bledenje tehnologij za ustvarjanje podatkov in uporaba analitičnih orodij odpira vrata temi strojnega učenja.**

Kot je v intervjuju leta 2023 dejal vplivni direktor največjega svetovnega investicijskega sklada (ki ima v lasti ključne deleže v skoraj vseh največjih podjetjih, ki izdelujejo programsko opremo za gradbeništvo, in v podjetjih, ki imajo v lasti največ nepremičnin na svetu [55]) - strojno učenje bo spremenilo svet gradbeništva.

*Umetna inteligenca ima ogromen potencial. Spremenila bo naš način dela in življenja. UI in robotika bosta spremenili način dela in gradnje, UI in robotiko pa bomo lahko uporabili kot sredstvo za ustvarjanje veliko večje produktivnosti [153].*

- Izvršni direktor največjega svetovnega investicijskega sklada, intervju, september 2023.

Strojno učenje (ML) deluje tako, da obdeluje velike količine podatkov in s statističnimi tehnikami posnema vidike človeškega razmišljanja. Vendar večina podjetij nima takšnih zbirk podatkov, če pa jih

ima, pogosto niso dovolj označene. Pri tem lahko pomagajo semantične tehnologije in transferno učenje, tehnika, ki omogoča večjo učinkovitost ML pri obravnavi majhnih količin podatkov in katere izvedljivost je bila obravnavana v prejšnjih poglavijih tega dela.

Bistvo transfernega učenja je v tem, da namesto da bi vsako nalogu obdelali od začetka, lahko uporabite znanje, pridobljeno na sorodnih področjih. Zavedati se je treba, da je mogoče vzorce in odkritja iz drugih panog prilagoditi in uporabiti v gradbeništvu. Na primer, metode optimizacije logističnih procesov, razvite v maloprodaji, pomagajo izboljšati učinkovitost upravljanja dobavne verige v gradbeništvu. Analizo velikih količin podatkov, ki se aktivno uporablja v financah, je mogoče uporabiti za napovedovanje stroškov in obvladovanje tveganj pri gradbenih projektih. Tehnologije računalniškega vida in robotike, ki se razvijajo v industriji, pa se že uporabljajo pri samodejnem nadzoru kakovosti, spremeljanju varnosti in upravljanju objektov na gradbiščih.

Učenje s prenosom omogoča ne le hitrejše uvajanje inovacij, temveč tudi zmanjšanje stroškov njihovega razvoja z uporabo že zbranih izkušenj drugih panog.

$$\text{labor productivity in construction} = f(\text{AI})$$

**Slika 9.2-3 Umetna inteligenco tehnologije in robotika bodo v prihodnosti glavna gonilna sila za povečanje produktivnosti v gradbeništvu.**

Človeško razmišljanje je organizirano po podobnem načelu: pri reševanju novih problemov gradimo na predhodno pridobljenem znanju (slika 4.4-19, slika 4.4-20, slika 4.4-21). Tudi pri strojnem učenju ta pristop deluje - s poenostavitevijo podatkovnega modela in njegovo eleganco lahko zmanjšamo zapletenost problema za algoritme ML. To pa zmanjša potrebo po velikih količinah podatkov in zniža računske stroške.

### Od subjektivne ocene do statistične napovedi

Obdobje, ko so bile strateške odločitve odvisne od intuicije posameznih vodij (slika 9.2-4), je preteklost. V vse bolj konkurenčnem in zahtevnem gospodarskem okolju postaja subjektivni pristop preveč tvegan in neučinkovit. Podjetja, ki se še naprej zanašajo na osebna mnenja, namesto da bi objektivno analizirala podatke, izgubljajo sposobnost hitrega odzivanja na spremembe.

Konkurenčno okolje zahteva natančnost in ponovljivost na podlagi podatkov, statističnih vzorcev in izračunljive verjetnosti. Odločitve ne morejo več temeljiti na občutkih, temveč morajo temeljiti na korelacijah, trendih in napovednih modelih, ki izhajajo iz analitike in strojnega učenja. Ne gre le za spremembo orodij, temveč za spremembo logike razmišljanja: od domnev k dokazom, od subjektivnih verjetnosti k statistično izračunanim odstopanjem, od občutkov k dejstvu.



Slika 9.2-4 Obdobje odločitev, ki jih je sprejemal HiPPO (mnenje najbolje plačanega zaposlenega), bo s prihodom velikih podatkov in strojnega učenja postalo preteklost.

Vodje, ki so se prej zanašali le na lastne občutke, se bodo neizogibno soočili z novo resničnostjo: avtoriteta ne določa več izbire. V središču upravljanja so zdaj sistemi, ki analizirajo na milijone parametrov in vektorjev, prepoznavajo skrite vzorce in predlagajo optimalne strategije.

Glavni razlog, zakaj se podjetja danes še vedno izogibajo uvajanju ML, je pomanjkanje preglednosti. Večina modelov deluje kot "črne skrinjice" za menedžerje, ne da bi pojasnili, kako točno pridejo do svojih zaključkov. To povzroča težave: algoritmi lahko krepijo stereotipe in celo ustvarjajo humorne situacije, kot v primeru Microsoftovega klepetalnega robota, ki se je hitro spremenil v toksično komunikacijsko orodje [154].

Garry Kasparov, nekdanji svetovni prvak v šahu, v knjigi Deep Thinking razmišlja o svojem porazu z IBM-ovim računalnikom Big Blue [155]. Trdi, da prava vrednost umetne inteligence ni v kopiranju človeške inteligence, temveč v dopolnjevanju naših sposobnosti. UI bi morala opravljati naloge, pri katerih so ljudje šibki, medtem ko bi ljudje vnesli ustvarjalnost.

Računalniki so spremenili tradicionalni pristop k analizi šaha. Namesto da bi ustvarjali fascinantne zgodbe o partijah, računalniški šahovski programi nepristransko ocenjujejo vsako potezo le na podlagi njene dejanske moči ali slabosti. Kasparov opozarja, da človeška težnja, da bi na dogodke gledali kot na povezane zgodbe in ne kot na posamezna dejanja, pogosto pripelje do napačnih sklepov - ne le v šahu, temveč tudi v življenju na splošno.

Če nameravate uporabljati strojno učenje za napovedovanje in analizo, je zato pomembno, da razumete njegova osnovna načela - kako delujejo algoritmi in kako se obdelujejo podatki - preden začnete pri svojem delu uporabljati orodja za strojno učenje in AI. Najboljši način za začetek je pridobivanje praktičnih izkušenj.

Eno najprimernejših orodij za začetni uvod v temo strojnega učenja in napovedovanja je beležnica Jupyter in priljubljen klasični nabor podatkov Titanic, ki bo omogočil vizualni uvod v ključne metode analize podatkov in gradnje modelov strojnega učenja.

## Zbirka podatkov Titanik: Pozdravljen svet v svetu analitike podatkov in velikih podatkov

Eden najbolj znanih primerov uporabe ML v podatkovni analitiki je analiza nabora podatkov o Titaniku, ki se pogosto uporablja za preučevanje verjetnosti preživetja potnikov. Učenje te tabele je analogno programu "Hello World" pri učenju programskih jezikov.

Ob potopu ladje RMS Titanic leta 1912 je umrlo 1502 od 2224 ljudi. Nabor podatkov o Titaniku ne vsebuje le informacij o tem, ali je potnik preživel, temveč tudi atributi, kot so: starost, spol, razred vozovnice in drugi parametri. Ta nabor podatkov je na voljo brezplačno ter ga je mogoče odpreti in analizirati na različnih offline in spletnih platformah.

Povezava do nabora podatkov o Titaniku:

<https://raw.githubusercontent.com/datasets/titanic/master/titanic.csv>

V poglavju "IDE s podporo LLM in prihodnje programske spremembe" smo že obravnavali beležnico Jupyter - eno najbolj priljubljenih razvojnih okolij za podatkovno analizo in strojno učenje. Brezplačni oblačni analogiji beležnice Jupyter Notebook sta platformi Kaggle in Google Collab, ki omogočata izvajanje kode Python brez namestitve programske opreme in zagotavljata brezplačen dostop do računalniških virov.

Kaggle je največja tekmovalna platforma za podatkovno analitiko in strojno učenje z integriranim okoljem za izvajanje kode. Od oktobra 2023 ima Kaggle več kot 15 milijonov uporabnikov [156] iz 194 držav.

Prenesite in uporabite nabor podatkov o Titaniku na platformi Kaggle (slika 9.2-5), da shranite nabor podatkov (njegovo kopijo) in zaženete kodo Python s predhodno nameščenimi knjižnicami neposredno v brskalniku, ne da bi morali namestiti namenski IDE.



**Slika 9.2-5 Statistika tabele Titanik - najbolj priljubljena podatkovna zbirka za učenje podatkovne analitike in strojnega učenja.**

Zbirka podatkov o Titaniku vključuje podatke o 2 224 potnikih na krovu ladje RMS *Titanik* ob njenem brodolому leta 1912. Nabor podatkov je predstavljen kot dve ločeni tabeli, vzorec za usposabljanje (train.csv) in vzorec za testiranje (test.csv), kar omogoča uporabo tako za usposabljanje modelov kot za ocenjevanje njihove natančnosti na novih podatkih.

Podatkovna zbirka za usposabljanje vsebuje atribute - atribute potnikov (starost, spol, razred vozovnice in drugo) ter informacije o tem, kdo je preživel (stolpec z binarno vrednostjo "Preživel"). Nabor podatkov za usposabljanje (slika 9.2-6 - datoteka train.csv) se uporablja za usposabljanje modela. Testni nabor podatkov (slika 9.2-7 - datoteka test.csv) vključuje samo atribute potnikov brez informacij o preživelih (brez enega stolpca "Preživel"). Testni nabor podatkov je namenjen testiranju modela na novih podatkih in ocenjevanju njegove natančnosti.

Tako imamo skoraj enake atribute potnikov v učnih in testnih zbirkah podatkov. Edina ključna razlika je, da imamo v testni zbirki podatkov seznam potnikov, ki nimajo stolpca "Preživeli" - ciljne spremenljivke, ki se jo želimo naučiti napovedovati z uporabo različnih matematičnih algoritmov. Po izgradnji modela pa bomo lahko primerjali rezultate našega modela z dejanskim parametrom "Survivor" iz testnega nabora podatkov, ki ga bomo upoštevali pri vrednotenju rezultatov.

Glavni stolpci tabele, parametri potnikov v naboru podatkov za usposabljanje in testiranje:

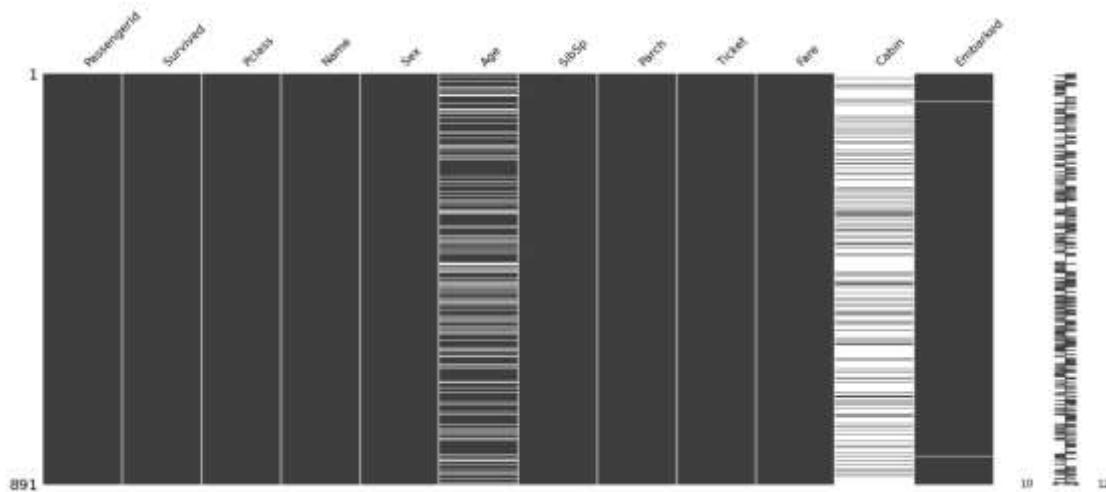
- **PassengerId** - edinstven identifikator potnika
- **Preživel** - 1, če je potnik preživel, 0, če je mrtev (ni na voljo v testnem nizu)
- **Pclass** - razred vozovnice (1, 2 ali 3)
- **Ime** - ime potnika
- **Spol** - spol potnika (moški/ženska)
- Starost
- **SibSp** - število bratov/sester ali zakoncev na krovu
- **Parch** - število staršev ali otrok na krovu
- **Vstopnica** - številka vstopnice
- **Prevoznina** - cena vozovnice
- Kabina - številka kabine (veliko podatkov manjka)
- **Vkrcano** je pristanišče vkrcanja (C = Cherbourg, Q = Queenstown, S = Southampton).

Za vizualizacijo manjkajočih podatkov v obeh preglednicah lahko uporabite knjižnico missingno (slika 9.2-6, slika 9.2-7), ki prikazuje manjkajoče vrednosti v obliki histograma, kjer bela polja označujejo manjkajoče podatke. Ta vizualizacija omogoča hitro oceno kakovosti podatkov pred obdelavo.

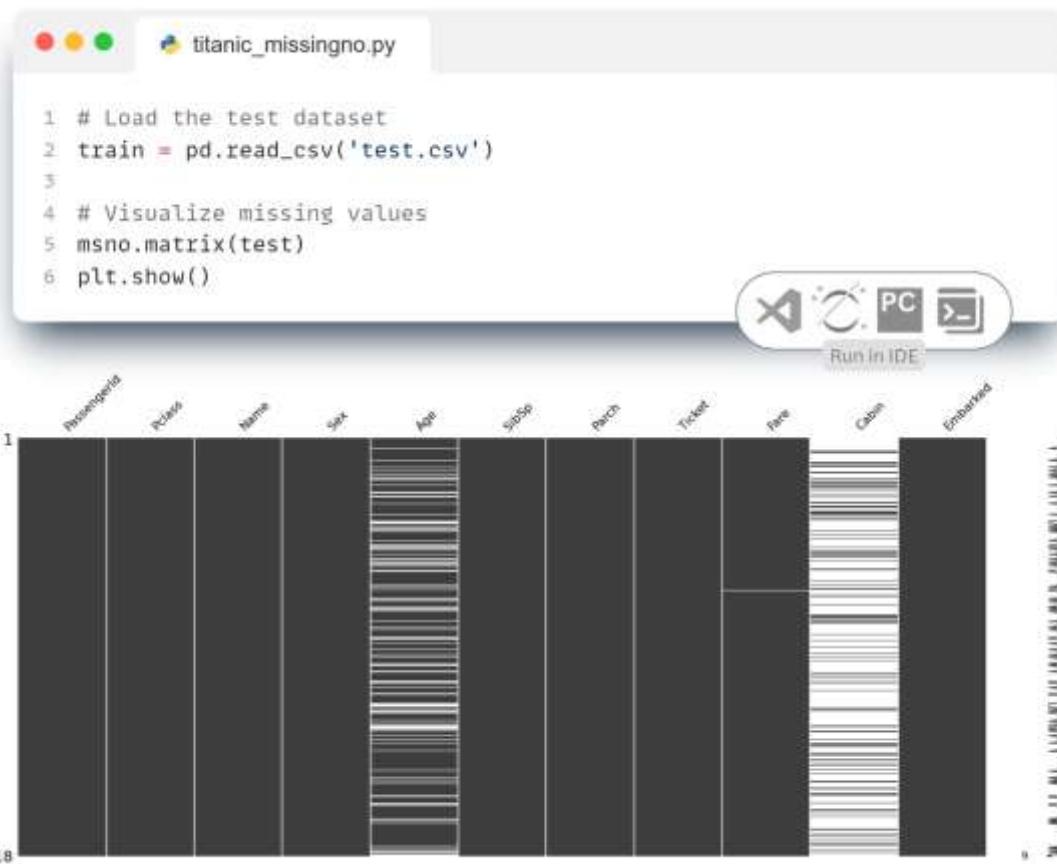
```
titanic_missingno.py
```

```
1 import pandas as pd
2 import missingno as msno
3 import matplotlib.pyplot as plt
4
5 # Load the dataset
6 train = pd.read_csv('train.csv')
7
8 # Visualize missing values
9 msno.matrix(train)
10 plt.show()
```

Run in IDE



Slika 9.2-6 Nekaj vrstic kode je uporabljenih za vizualizacijo manjkajočih podatkov v naboru podatkov za usposabljanje za Titanik, kjer je ključni parameter za usposabljanje parameter "Preživel".



Slika 9.2-7 Vizualizacija manjkajočih podatkov v testnem naboru podatkov o Titaniku, ki vsebuje samo značilnosti potnikov brez informacij.

Pred oblikovanjem hipotez in napovedi na podlagi nabora podatkov vizualne analize pomagajo prepoznati ključne vzorce v podatkih, oceniti njihovo kakovost in ugotoviti morebitne odvisnosti. Obstajajo številne tehnike vizualizacije, ki vam lahko pomagajo bolje razumeti nabor podatkov o Titaniku. Za analizo starostnih skupin potnikov lahko uporabite diagrame porazdelitve, grafe preživetja po spolu in razredu ter matrike manjkajočih podatkov, da ocenite kakovost informacij in razumete podatke.

- ❷ Prosimo LLM, da nam pomaga vizualizirati podatke iz nabora podatkov o Titaniku, tako da kateremu koli modelu LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ali kateremu koli drugemu) pošljemo naslednjo besedilno zahtevo:

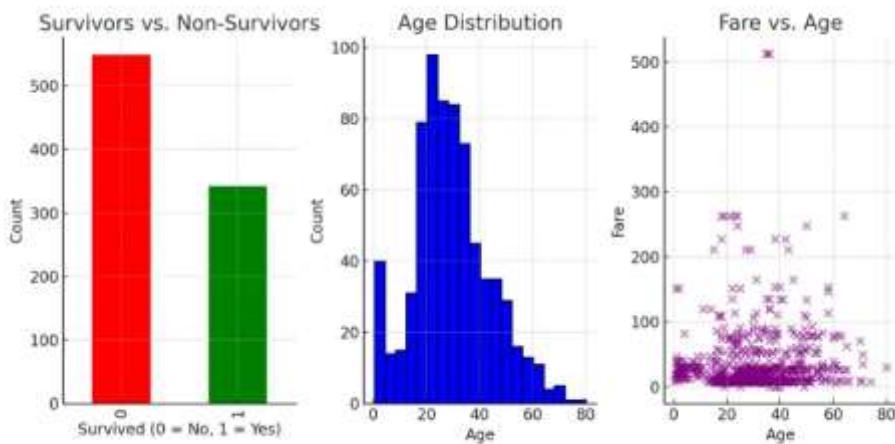
Prosimo, pokažite nekaj preprostih grafov za nabor podatkov o Titaniku. Sami prenesite nabor podatkov in prikažite ↵

- ❸ Odziv LLM v obliki pripravljene kode in grafov, ki prikazujejo parametre podatkovnega niza

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # Load the dataset
5 file_path = "/mnt/data/titanic.csv"
6 df = pd.read_csv(file_path)
7
8 # Convert column names to lowercase for consistency
9 df.columns = df.columns.str.lower()
10
11 # Basic Plots
12 plt.figure(figsize=(10, 5))
13
14 # Bar Chart: Survival Count
15 plt.subplot(1, 3, 1)
16 df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
17 plt.title("Survivors vs. Non-Survivors")
18 plt.xlabel("Survived (0 = No, 1 = Yes)")
19 plt.ylabel("Count")
20
21 # Histogram: Age Distribution
22 plt.subplot(1, 3, 2)
23 df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
24 plt.title("Age Distribution")
25 plt.xlabel("Age")
26 plt.ylabel("Count")
27
28 # Scatter Plot: Fare vs. Age
29 plt.subplot(1, 3, 3)
30 plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
31 plt.title("Fare vs. Age")
32 plt.xlabel("Age")
33 plt.ylabel("Fare")
34
35 # Show the plots
36 plt.tight_layout()
37 plt.show()

```



Slika 9.2-8 LLM vam pomaga pri takojšnji vizualizaciji podatkov vašega nabora podatkov.

Vizualizacija podatkov je pomemben korak pri pripravi podatkovne zbirke za poznejšo izdelavo modela strojnega učenja, do katerega je mogoče priti le z razumevanjem podatkov.

## Strojno učenje v akciji: od potnikov na Titaniku do upravljanja projektov

Glavna hipoteza, uporabljena za raziskovanje okvira strojnega učenja na podlagi nabora podatkov o Titaniku, je, da so imele nekatere skupine potnikov večje možnosti za preživetje.

Majhna tabela potnikov Titanika je postala priljubljena po vsem svetu, na milijone ljudi pa jo uporablja za usposabljanje, eksperimentiranje in testiranje modelov, da bi ugotovili, kateri algoritmi in hipoteze bodo na podlagi nabora podatkov za usposabljanje potnikov Titanika ustvarili najnatančnejši model napovedi preživetja.

Zbirka podatkov o Titaniku je privlačna zaradi svoje kompaktnosti: z več sto vrsticami in dvanajstimi stolpcji (slika 9.2-6) ponuja veliko možnosti za analizo. Zbirka podatkov je, relativno preprosto, klasičen primer rešitve binarne klasifikacije, kjer je cilj problema - preživetje - izražen v priročni obliki 0 ali 1.

John Wheeler v filmu "It from Bit" [7] trdi, da vesolje temelji na binarnih izbirah. Podobno tudi podjetje, ki ga vodijo ljudje, sestavljeni iz molekul, dejansko temelji na vrsti binarnih izbir.

Poleg tega podatki temeljijo na resničnem zgodovinskem dogodku, zato so za razliko od umetno ustvarjenih primerov dragoceni za raziskave. Samo na platformi Kaggle, eni največjih platform za podatkovne cevovode in ETL, je v izzivih, ki temeljijo na naboru podatkov o Titaniku, sodelovalo 1 355 998 ljudi, ki so razvili 53 963 edinstvenih rešitev podatkovnih cevovodov [157] (slika 9.2-9).

Zdi se neverjetno, a samo 1000 vrstic podatkov o potnikih s Titanika z 12 parametri je postalo polje za milijone hipotez, logičnih verig in edinstvenih podatkovnih cevovodov. Iz majhnega nabora podatkov se rodi neskončno vpogledov, hipotez in razlag - od preprostih modelov preživetja do kompleksnih sklopov, ki upoštevajo skrite vzorce in zapletene labirinte skelepanja.

## Machine Learning from Disaster

[Submit Prediction](#)

[Data](#) [Code](#) [Models](#) [Discussion](#) [Leaderboard](#) [Rules](#)

### Titanic Tutorial

Updated 3y ago

29858 comments · Titanic - Machine Learning from Disaster

16916

Gold \*\*\*

### Titanic competition w/ TensorFlow Decision Forests

Updated 2y ago

Score: 0.80143 · 318 comments · Titanic - Machine Learning from Disaster

1098

Gold \*\*\*

### Titanic Data Science Solutions

Updated 6y ago

2590 comments · Titanic - Machine Learning from Disaster

10723

Gold \*\*\*

### Exploring Survival on the Titanic

Updated 7y ago

Score: 0.80382 · 1072 comments · Titanic - Machine Learning from Disaster

3968

Gold \*\*\*

Slika 9.2-9 Prvih pet rešitev od skupno 53.963 pripravljenih in odprtakodnih rešitev Pipeline. Samo na portalu Kaggle [157] je ta problem poskušalo rešiti že skoraj 1,5 milijona ljudi [157].

Če lahko že tako majhna tabela ustvari na milijone edinstvenih rešitev (slika 9.2-9), kaj lahko rečemo o resničnih naborih podatkov o industrijski gradnji, kjer se parametri merijo v desettisočih?

Standardni CAD -projekt relativno majhne stavbe vsebuje več deset tisoč entitet s tisočimi parametri - od geometrijskih značilnosti do stroškovnih in časovnih atributov. Predstavljajte si, koliko potencialnih vpogledov, odnosov, napovedi in upravljavskih hipotez se skriva v podatkih vseh projektov vašega podjetja, zbranih v preteklih letih. Zgodovinski podatki o projektih niso le arhiv - so živi spomin organizacije, njen digitalni odtis, ki ga je mogoče analizirati za oblikovanje številnih edinstvenih hipotez.

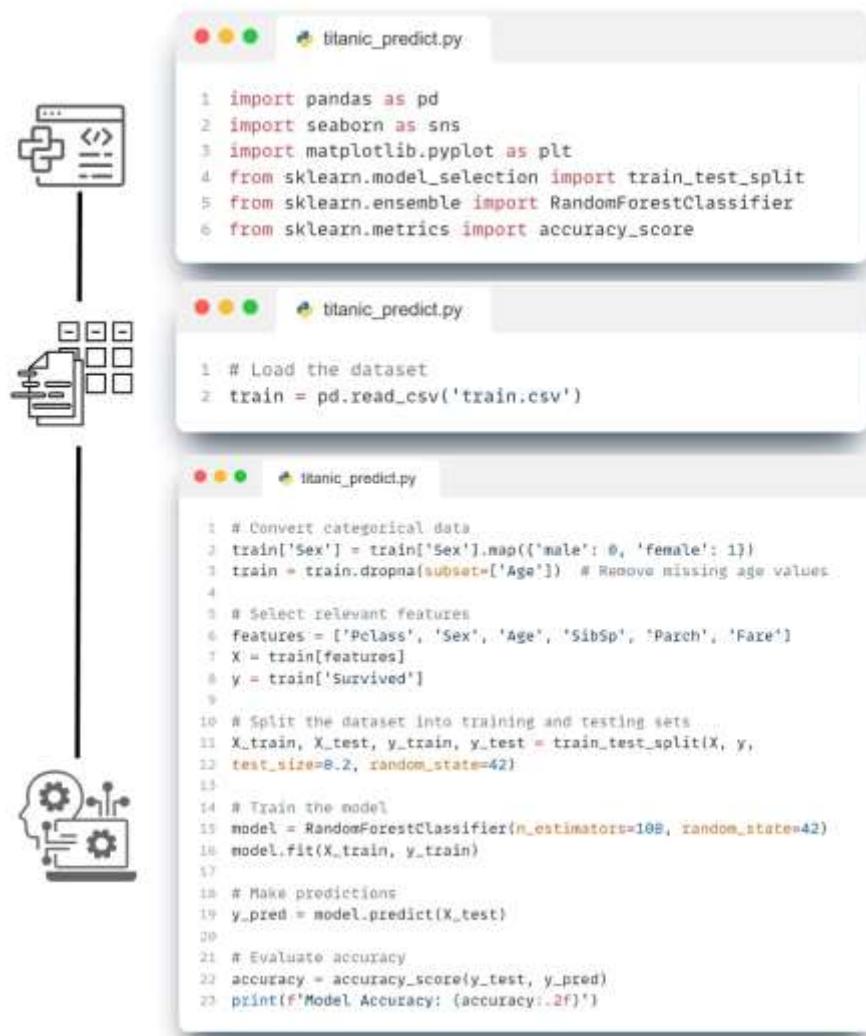
Najpomembnejše pa je, da vam ni treba čakati, da se skupnost Kaggle začne zanimati za vaše podjetje ali vaše podatke. S tem, kar imate, lahko začnete delati že danes: izvajajte analitiko na lastnih podatkih, usposabljaljajte modele na lastnih podatkih, prepoznavajte ponovitve, anomalije in vzorce. Če so bila včasih potrebna leta eksperimentiranja in dragega svetovanja, so zdaj dovolj samoiniciativnost,

magisterij, odprt pristop k podatkom in pripravljenost na učenje.

- ❷ Za izgradnjo algoritma strojnega učenja, ki bo na podlagi podatkovne zbirke train.csv za usposabljanje potnikov napovedal stopnjo preživetja potnikov, prosimo LLM, da za nas reši ta problem:

Na podlagi nabora učnih podatkov o potnikih iz Titanika izdelajte model strojnega učenja za napovedovanje stopnje preživetja. ↵

- ❸ Odgovor LLM :



```

titanic_predict.py
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.ensemble import RandomForestClassifier
6 from sklearn.metrics import accuracy_score

# Load the dataset
train = pd.read_csv('train.csv')

# 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}')

```

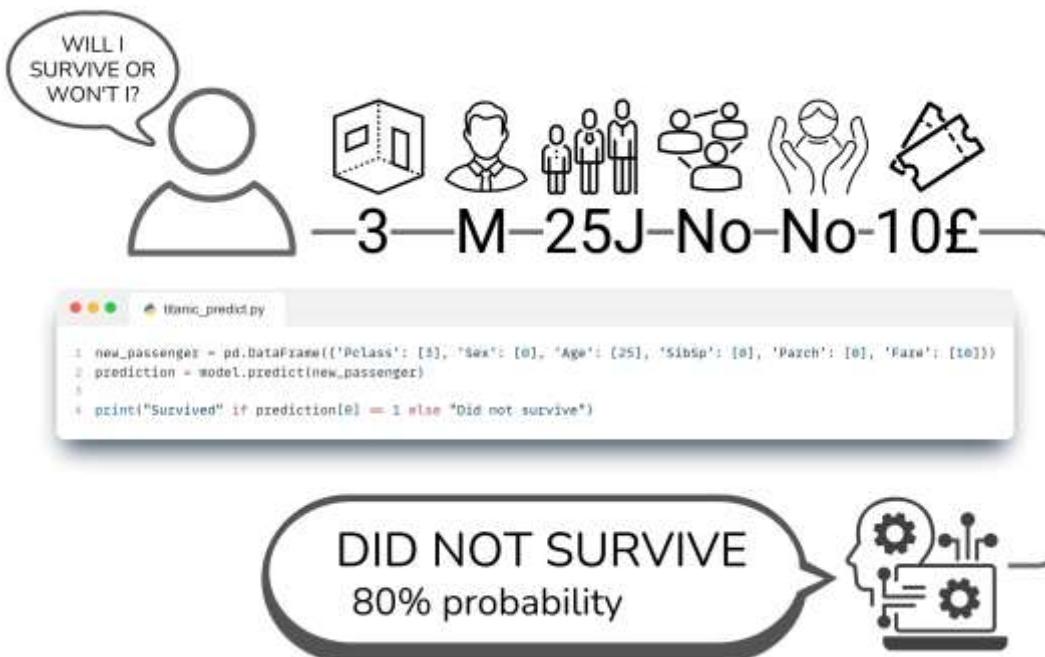
Slika 9.2-10 LLM je z algoritmom strojnega učenja Random Forest izdelal napoved preživelih na Titaniku.

Dobljena koda iz LLM (slika 9.2-10) naloži podatke o Titanikovih potnikih, jih očisti, pretvori kategorične spremenljivke (npr. spol v številčno obliko) in z algoritmom RandomForestClassifier usposobi model za napovedovanje, ali je potnik preživel ali ne (več o priljubljenih algoritmih bomo govorili v naslednjih poglavjih).

Koda podatke za usposabljanje razdeli na učne in testne nize (na spletnem mestu Kaggle sta že pripravljena pripravljena test.csv (slika 9.2-7) in train.csv (slika 9.2-6) za usposabljanje, nato se model usposobi na podatkih za usposabljanje in testira na testnih podatkih, da se ugotovi, kako dober je določen model napovedi. Po usposabljanju se testni podatki iz test.csv (z resničnimi podatki o tistih, ki so preživeli ali niso preživeli) vnesejo v model in ta napoveduje, kdo je preživel in kdo ne. V našem primeru je natančnost modela strojnega učenja, ki smo ga dobili, približno 80%, kar kaže, da precej dobro zajame vzorce.

Strojno učenje lahko primerjamo z otrokom, ki poskuša pravokotno kocko vstaviti v okroglo luknjo. V začetnih fazah algoritom poskuša številne pristope, pri čemer naleti na napake in nedoslednosti. Ta proces se morda zdi neučinkovit, vendar omogoča pomembno učenje: z analizo vsake napake model izboljšuje svoje napovedi in sprejema vse natančnejše odločitve.

Zdaj lahko ta model (slika 9.2-10) uporabite za napovedovanje stopnje preživetja novih potnikov in na primer, če ga oskrbite z informacijami o potnikih z uporabo funkcije model.predict s parametri: "moški", "3. razred", "star 25 let", "brez sorodnikov na krovu", bo model pripravil napoved - da potnik z 80-odstotno verjetnostjo ne bo preživel katastrofe, če je bil na Titaniku leta 1912 (slika 9.2-11).



Slika 9.2-11 Z modelom, ki smo ga ustvarili zgoraj, lahko zdaj z 80-odstotno verjetnostjo napovemo, ali bo novi potnik Titanika preživel ali ne.

Model napovedi preživetja potnikov na Titaniku ponazarja veliko širši koncept: na tisoče strokovnjakov v gradbeništvu vsak dan sprejema podobne "dvojne" odločitve - o življenju ali smrti odločitve, projekta, ocene, orodja, dobička ali izgube, varnosti ali tveganja. Tako kot v primeru Titanika, kjer je bil izid odvisen od dejavnikov (spol, starost, razred), tudi v gradbeništvu na vsak vidik odločitve vplivajo številni lastni dejavniki in spremenljivke (stolpci tabel): stroški materiala, usposobljenost delavcev, roki, vreme, logistika, tehnična tveganja, pripombe in na stotisoče drugih parametrov.

Strojno učenje v gradbeništvu temelji na enakih načelih kot na drugih področjih: modeli se usposabljam na preteklih podatkih - iz projektov, pogodb, ocen -, da se preverijo različne hipoteze in poiščejo najučinkovitejše rešitve. Ta proces je podoben učenju otroka s poskusi in napakami: z vsakim ciklom se modeli prilagodijo in postanejo natančnejši.

Uporaba zbranih podatkov odpira nova obzorja v gradbeništvu. Namesto zamudnih ročnih izračunov je mogoče usposobiti modele, ki lahko z veliko natančnostjo napovedujejo ključne značilnosti prihodnjih projektov. Na ta način napovedna analitika spreminja gradbeno industrijo v prostor, v katerem lahko ne le načrtujete, temveč tudi zanesljivo napovedujete razvoj dogodkov.

## Napovedi in napovedi na podlagi preteklih podatkov

Podatki, zbrani o projektih podjetja, omogočajo izdelavo modelov, s katerimi je mogoče predvideti stroškovne in časovne značilnosti prihodnjih, še ne realiziranih objektov - brez zamudnih ročnih izračunov in primerjav. To omogoča znatno pospešitev in poenostavitev postopkov vrednotenja, pri-

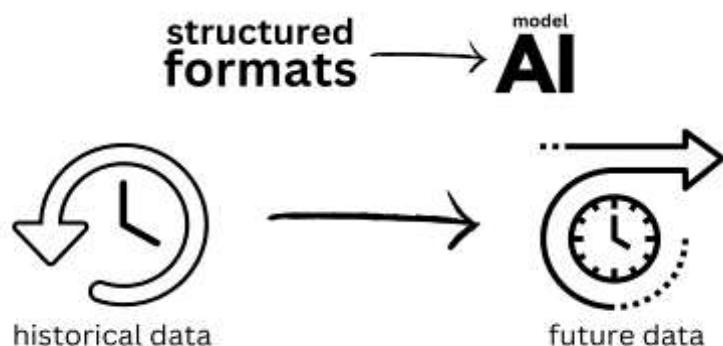
čemer se ne zanašajo na subjektivne predpostavke, temveč na zanesljive matematične napovedi.

Prej, v četrtem delu knjige, smo podrobno obravnavali tradicionalne metode ocenjevanja stroškov projekta, vključno z metodo, ki temelji na virih, omenili pa smo tudi parametrične in ekspertrne pristope. Te metode so še vedno pomembne, vendar se v sodobni praksi začenjajo bogatiti z orodji statistične analize in strojnega učenja, ki lahko bistveno izboljšajo natančnost in ponovljivost ocen.

Postopke ročnega in polavtomatskega izračunavanja cen in časovnih atributov bodo v prihodnosti dopolnjevali mnenja in napovedi modelov ML, ki lahko analizirajo zgodovinske podatke, iščejo skrite vzorce in predlagajo utemeljene odločitve.

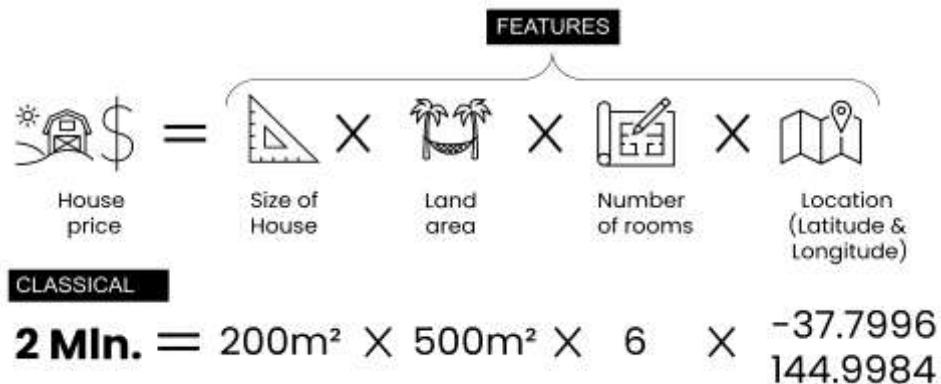
Novi podatki in scenariji se bodo samodejno ustvarjali iz že razpoložljivih informacij - podobno kot jezikovni modeli (LLM) ustvarjajo besedila, slike in kode na podlagi podatkov, ki so se z leti zbirali iz odprtih virov [158].

Tako kot se danes ljudje pri ocenjevanju prihodnjih dogodkov zanašamo na izkušnje, intuicijo in notranjo statistiko, bo v prihodnjih letih prihodnost gradbenih projektov vse bolj določala kombinacija zbranega znanja in matematičnih modelov strojnega učenja.



**Slika 9.2-12 Kvalitativni in strukturirani zgodovinski podatki o podjetju so gradivo, na katerem se gradijo modeli strojnega učenja in napovedi.**

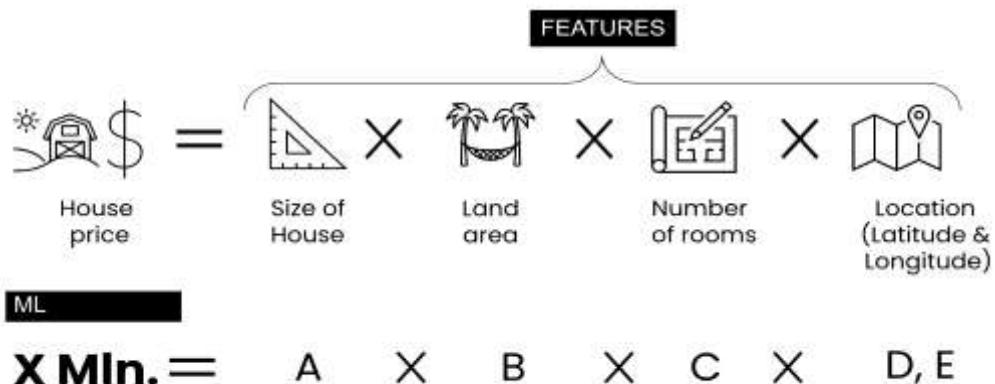
Poglejmo preprost primer: napovedovanje cene hiše na podlagi njene površine, velikosti parcele, števila sob in geografskega položaja. Eden od pristopov je izdelava klasičnega modela, ki analizira te parametre in izračuna pričakovano ceno (slika 9.2-13). Ta pristop zahteva natančno in vnaprej znano formulo, kar je v realni praksi praktično nemogoče.



Slika 9.2-13 Za oceno vrednosti hiše s fiksno formulo, ki jo je treba najti, lahko uporabimo klasični algoritem.

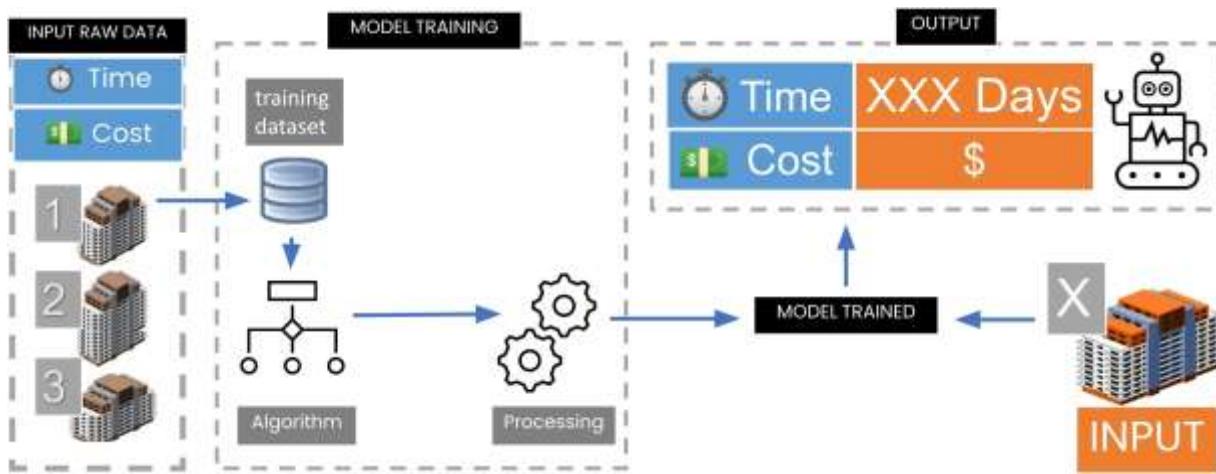
Strojno učenje odpravlja ročno iskanje formul in jih nadomešča z usposobljenimi algoritmi, ki neodvisno prepoznajo odvisnosti, ki so večkrat natančnejše od vseh vnaprej določenih enačb. Druga možnost je, da ustvarimo algoritem strojnega učenja, ki bo ustvaril model na podlagi predhodnega razumevanja problema in zgodovinskih podatkov, ki so lahko nepopolni (slika 9.2-14).

Na primeru cen lahko s strojnim učenjem ustvarimo različne vrste matematičnih modelov, ki ne zahtevajo poznavanja natančnega mehanizma oblikovanja stroškov. Model se "uči" iz podatkov o prejšnjih projektih in se prilagaja dejanskim vzorcem med gradbenimi parametri, njihovimi stroški in roki.



Slika 9.2-14 V nasprotju s klasičnim ocenjevanjem na podlagi formul se algoritem strojnega učenja uči na podlagi preteklih podatkov.

V okviru nadzorovanega strojnega učenja vsak projekt v naboru podatkov za usposabljanje vsebuje vhodne atributte (npr. stroške in čas podobnih stavb) in pričakovane izhodne vrednosti (npr. stroške ali čas). Podoben nabor podatkov se uporablja za ustvarjanje in prilagajanje modela strojnega učenja (slika 9.2-15). Večji kot je nabor podatkov in boljša kot je kakovost podatkov v njem, natančnejši bo model in natančnejši bodo rezultati napovedi.



Slika 9.2-15 Model ML, usposobljen na podatkih o stroških in časovnem načrtu preteklih projektov, bo z določeno verjetnostjo določil stroške in časovni načrt novega projekta.

Ko je model ustvarjen in usposobljen za ocenjevanje gradnje novega projekta, mu preprosto predložite nove atribute za nov projekt in model bo na podlagi predhodno naučenih vzorcev z določeno verjetnostjo zagotovil ocenjene rezultate.

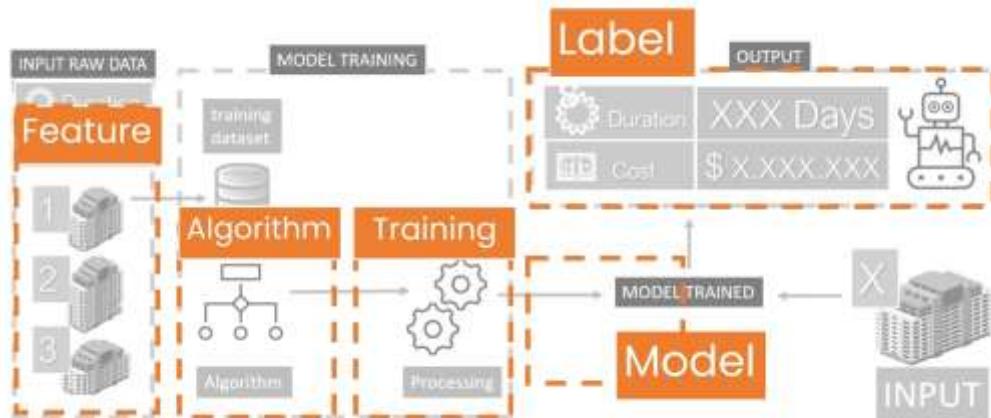
## Ključni koncepti strojnega učenja

Strojno učenje ni čarownija, temveč le matematika, podatki in iskanje vzorcev. Nima prave intelligence, temveč je program, ki je usposobljen na podatkih, da prepozna vzorce in sprejema odločitve brez stalnega človeškega vpletanja.

Strojno učenje uporablja več ključnih pojmov za opis svoje strukture (slika 9.2-16):

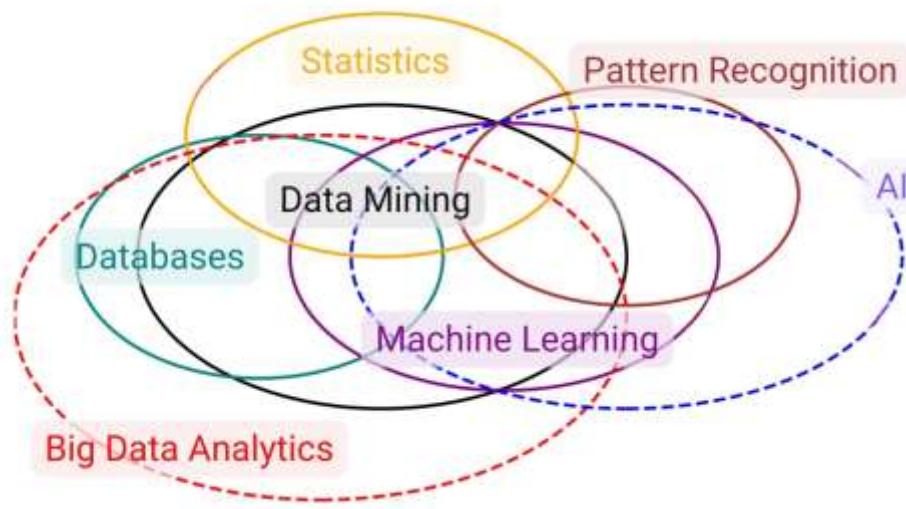
- **Oznake** so ciljne spremenljivke ali atributi (parameter "Survivor" v naboru podatkov o Titaniku), ki jih mora model napovedati. Primer: stroški gradnje (npr. v dolarjih), trajanje gradbenih del (npr. v mesecih).
- **Značilnosti** so neodvisne spremenljivke ali atributi, ki služijo kot vhodni podatki za model. V modelu za napovedovanje se uporabljajo za napovedovanje oznak. Primeri: velikost parcele (v kvadratnih metrih), število nadstropij stavbe, skupna površina stavbe (v kvadratnih metrih), geografska lega (zemljepisna širina in dolžina), vrsta materialov, uporabljenih pri gradnji. Število značilnosti določa tudi razsežnost podatkov.
- **Model** je niz različnih hipotez, od katerih ena približuje ciljno funkcijo, ki jo je treba napovedati ali približati. Primer: model strojnega učenja, ki uporablja tehnike regresijske analize za napovedovanje stroškov in časa gradnje.
- **Algoritem učenja** je postopek iskanja najboljše hipoteze v modelu, ki natančno ustreza ciljni funkciji, z uporabo niza podatkov za učenje. Primer: Algoritem linearne regresije, KNN ali naključnega gozda, ki analizira podatke o stroških in času gradnje, da bi ugotovil povezave in vzorce.

- Usposabljanje** - med postopkom usposabljanja algoritem analizira podatke za usposabljanje, da bi našel vzorce, ki ustrežajo razmerju med vhodnimi atributi in ciljnimi oznakami. Rezultat tega postopka je usposobljen model strojnega učenja, ki je pripravljen za napovedovanje. Primer: postopek, v katerem algoritem analizira pretekle podatke o gradnji (stroški, čas, značilnosti objekta), da ustvari napovedni model.



Slika 9.2-16 ML uporablja oznake in atribute za ustvarjanje modelov, ki se usposabljajo na podatkih s pomočjo algoritmov za napovedovanje rezultatov.

Strojno učenje ne obstaja izolirano, temveč je del širšega ekosistema analitičnih disciplin, vključno s statistiko, podatkovnimi bazami, podatkovnim rudarjenjem, prepoznavanjem vzorcev, analitiko velikih količin podatkov in umetno inteligenco. Slika 9.2-17 prikazuje, kako se ta področja prekrivajo in dopolnjujejo ter zagotavljajo celovit okvir za sodobne sisteme odločanja in avtomatizacije.

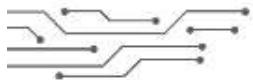


Slika 9.2-17 Razmerje med različnimi področji analize podatkov: statistika, strojno učenje, umetna inteligenco, veliki podatki, prepoznavanje vzorcev in podatkovno rudarjenje.

Glavni cilj strojnega učenja je opremiti računalnike s sposobnostjo samodejnega učenja znanja brez človeškega posredovanja ali pomoči in ustrezno prilagoditi svoje delovanje [159].

V prihodnosti bo torej vloga človeka le ta, da bo stroju zagotavljal kognitivne zmogljivosti - določil bo pogoje, uteži in parametre, vse ostalo pa bo opravil model strojnega učenja.

V naslednjem poglavju bomo obravnavali konkretne primere uporabe algoritmov. Uporabljeni bodo dejanske tabele in poenostavljeni modeli, ki bodo pokazali, kako se napoved gradi korak za korakom.



## POGLAVJE 9.3.

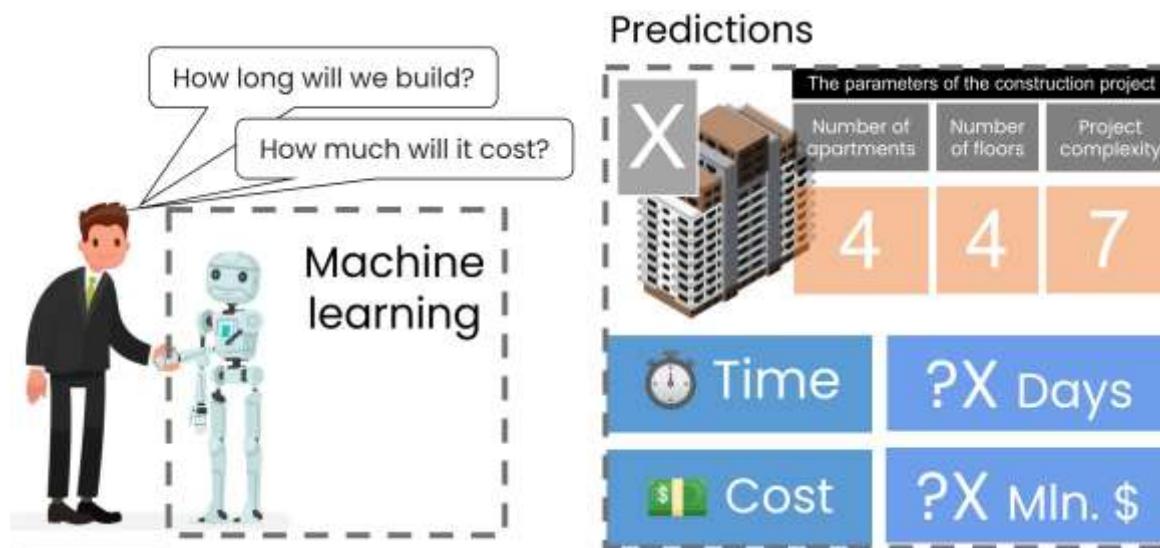
### NAPOVEDOVANJE STROŠKOV IN ROKOV Z UPORABO STROJNEGA UČENJA

#### Primer uporabe strojnega učenja za iskanje stroškov in časovnice projekta

Ocenjevanje časa in stroškov gradnje je eden ključnih procesov v dejavnostih gradbenega podjetja. Tradicionalno take ocene pripravljajo strokovnjaki na podlagi izkušenj, referenčnih knjig in regulativnih podatkovnih zbirk. Z digitalno preobrazbo in vse večjo razpoložljivostjo podatkov pa je zdaj mogoče uporabiti modele strojnega učenja (ML) za izboljšanje natančnosti in avtomatizacijo takšnih ocen.

Uvedba strojnega učenja v proces izračunavanja stroškov in časovnega okvira gradnje ne omogoča le učinkovitejšega načrtovanja, temveč postane tudi izhodišče za vključevanje inteligentnih modelov v druge poslovne procese - od upravljanja tveganj do optimizacije logistike in nabave.

Pomembno je, da lahko hitro ugotovite, koliko časa bo trajala gradnja projekta in kakšni bodo njegovi skupni stroški. Ta vprašanja o času in stroških projekta so bila tradicionalno v ospredju tako pri naročnikih kot pri gradbenih podjetjih vse od nastanka gradbene industrije.



Slika 9.3-1 Pri gradbenih projektih sta hitrost in kakovost ocenjevanja časa gradnje in stroškov ključna dejavnika uspeha.

V naslednjem primeru bodo ključni podatki iz preteklih projektov pridobljeni in uporabljeni za razvoj modela strojnega učenja, ki nam bo omogočil uporabo modela za oceno stroškov in časovnega okvira novih gradbenih projektov z novimi parametri (slika 9.3-1).

Upoštevajte tri projekte s tremi ključnimi atributmi: številom stanovanj (pri čemer je 100 stanovanj zaradi lažje vizualizacije enakovredno številu 10), številom nadstropij in pogojnim merilom zahtevnosti

gradnje na lestvici od 1 do 10, pri čemer je 10 najvišja ocena zahtevnosti. V strojnem učenju se postopek pretvorbe in poenostavitev vrednosti, kot je 100 v 10 ali 50 v 5, imenuje "normalizacija".

**Normalizacija** pri strojnem učenju je postopek približevanja različnih numeričnih podatkov skupni lestvici, da se olajša obdelava in analiza. Ta postopek je še posebej pomemben, kadar imajo podatki različne lestvice in merske enote.

Recimo, da je imel prvi projekt (slika 9.3-2) 50 stanovanj (po normalizaciji 5), 7 nadstropij in oceno zahtevnosti 2, kar pomeni razmeroma preprosto gradnjo. Drugi projekt je imel že 80 stanovanj, 9 nadstropij in razmeroma zapleteno konstrukcijo. Pod temi pogoji je gradnja prve in druge stanovanjske stavbe trajala 270 oziroma 330 dni, skupni stroški projekta pa so znašali 4,5 oziroma 5,8 milijona dolarjev.

Construction project	The parameters of the construction project			The key parameters of the project	
	Number of apartment	Number of floors	Project complexity	Time	Cost
1	5	7	2	270	\$ 4.502.000
2	8	9	6	330	\$ 5.750.000
3	3	5	3	230	\$ 3.262.000
X	4	4	7	?X	\$ ?X. XXX.XXX

Slika 9.3-2 Primer niza preteklih projektov, ki se bo uporabil za oceno časa in stroškov prihodnjega projekta X.

Pri gradnji modela strojnega učenja za takšne podatke je glavna naloga določiti kritične atributte (ali oznake) za napovedovanje, v tem primeru čas gradnje in stroške. Z majhnim naborom podatkov bomo podatke o prejšnjih gradbenih projektih uporabili za načrtovanje novih: z algoritmi strojnega učenja moramo napovedati stroške in trajanje gradnje novega projekta X na podlagi danih atributov novega projekta, kot so 40 stanovanj, 4 nadstropja in relativno visoka zahtevnost projekta 7 (slika 9.3-2). V realnem okolju je lahko število vhodnih parametrov veliko večje, od nekaj deset do več sto faktorjev. Ti lahko vključujejo: vrsto gradbenih materialov, podnebno območje, raven usposobljenosti izvajalcev, razpoložljivost komunalnih storitev, vrsto temeljev, sezono začetka del, pripombe mojstrov itd.

Če želimo ustvariti napovedni model strojnega učenja, moramo izbrati algoritom za njegovo ustvarjanje. Algoritom pri strojnem učenju je kot matematični recept, ki računalnik nauči, kako naj na podlagi podatkov napoveduje (meša v pravem vrstnem redu parametre) ali sprejema odločitve.

Za analizo podatkov o preteklih gradbenih projektih ter napovedovanje časa in stroškov prihodnjih projektov (slika 9.3-2) je mogoče uporabiti en priljubljen algoritem strojnega učenja:

- **Linearna regresija (Linearna regresija):** ta algoritem poskuša najti neposredno povezavo med atributi, na primer med številom nadstropij in stroški gradnje. Cilj algoritma je poiskati linearno enačbo, ki najbolje opisuje to razmerje in omogoča izdelavo napovedi.
- **Algoritem k-najbližji sosedje (k-NN):** ta algoritem primerja nov projekt s preteklimi projekti, ki so bili podobni po velikosti ali zahtevnosti. K-NN razvršča podatke na podlagi tega, kateri od k (števila) učnih primerov so mu najbližji. V okviru regresije je rezultat povprečje ali mediana k najbližjih sosedov.
- **Odločitvena drevesa:** je model napovednega modeliranja, ki podatke na podlagi različnih pogojev razdeli v podskupine z uporabo drevesne strukture. Vsako vozlišče drevesa predstavlja pogoj ali vprašanje, ki vodi do nadaljnje delitve podatkov, vsak list pa predstavlja končno napoved ali izid. Algoritem podatke razdeli v manjše skupine na podlagi različnih značilnosti, na primer najprej po številu zgodb, nato po zapletenosti in tako naprej, da se pripravi napoved.

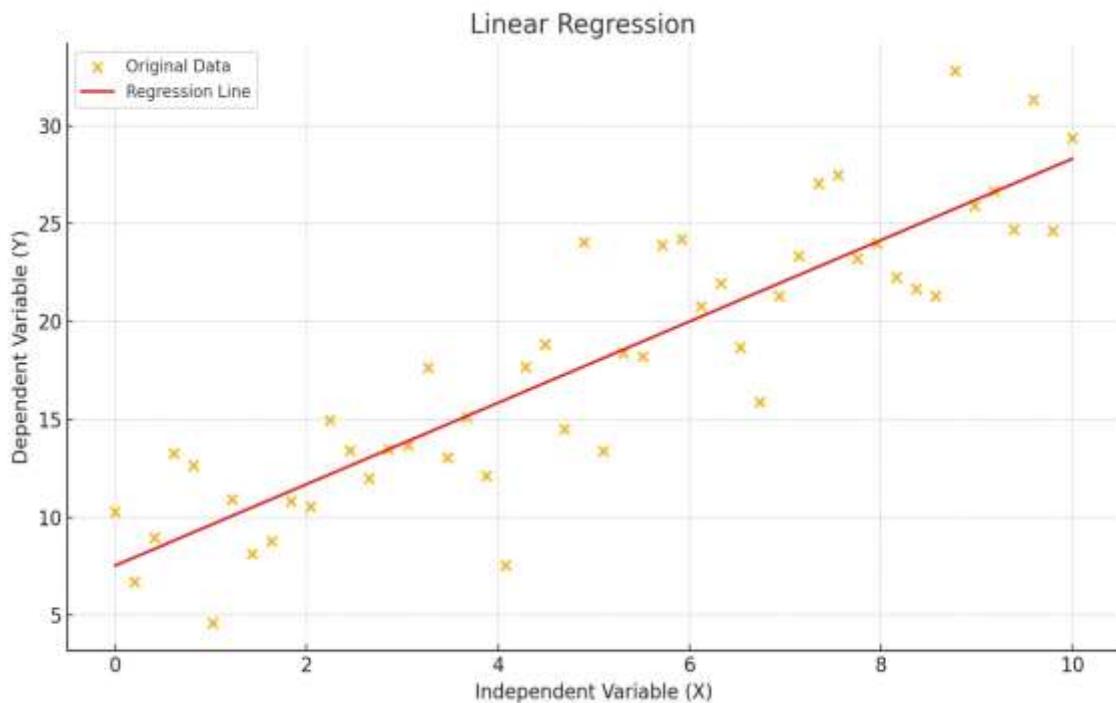
Oglejmo si algoritme strojnega učenja za ocenjevanje stroškov novega projekta na primeru dveh priljubljenih algoritmov: linearne regresije in algoritma K-najbližjih sosedov.

## Napovedovanje stroškov in časa projekta z uporabo linearne regresije

Linearna regresija je temeljni algoritem za analizo podatkov, ki napoveduje vrednost spremenljivke na podlagi linearne povezave z eno ali več drugimi spremenljivkami. Ta model predpostavlja, da med odvisno spremenljivko in eno ali več neodvisnimi spremenljivkami obstaja neposredna linearna povezava, cilj algoritma pa je najti to povezavo.

Zaradi preprostosti in jasnosti linearne regresije je postala priljubljeno orodje na različnih področjih. Pri obravnavi ene spremenljivke gre pri linearni regresiji za iskanje najbolje prilegajoče se premice skozi podatkovne točke.

Linearna regresija poišče najboljšo premico (rdeča črta), ki aproksimira odvisnost med vhodno spremenljivko X in izhodno spremenljivko Y. Ta premica omogoča napovedovanje vrednosti Y za nove vrednosti X na podlagi ugotovljene linearne povezave (slika 9.3-3).



Slika 9.3-3 Načelo linearne regresije je poiskati najboljšo premico skozi učne vrednosti.

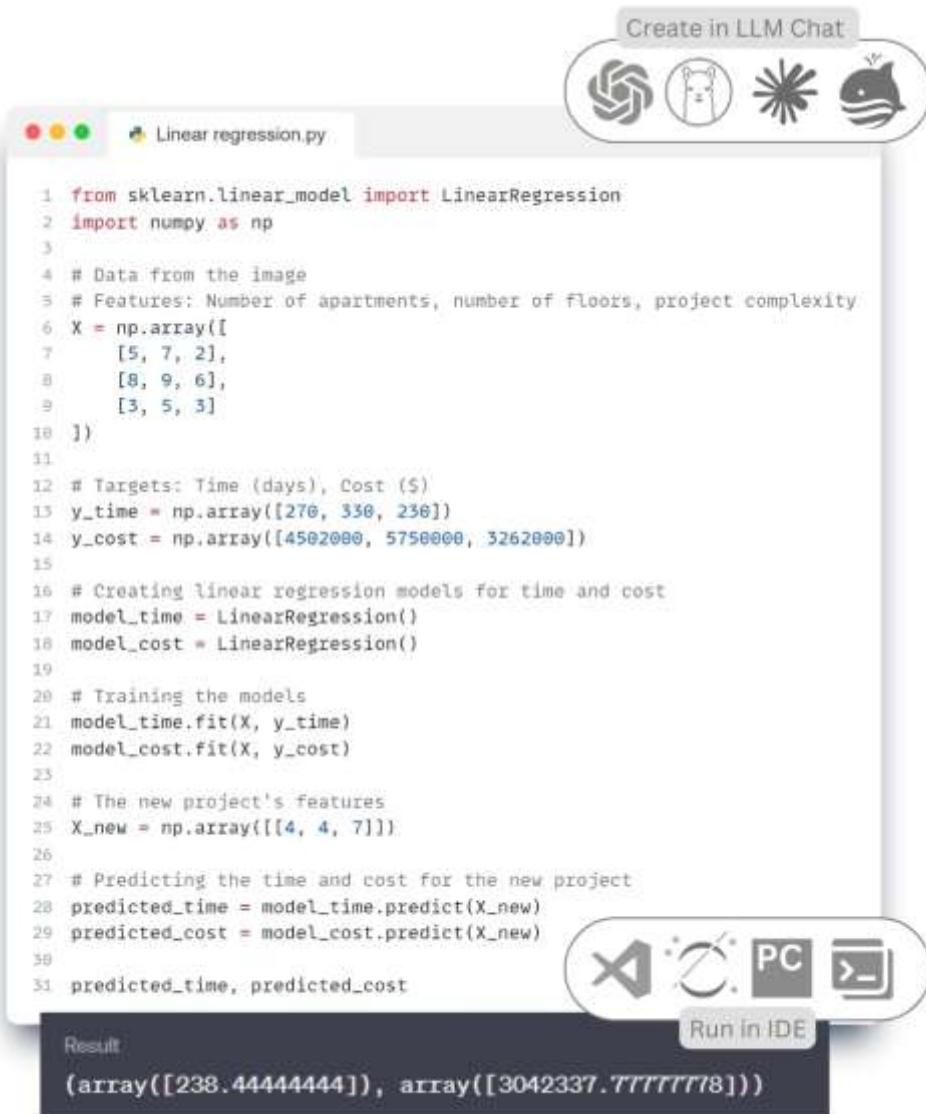
To črto predstavlja enačba, v kateri se z vnosom vrednosti neodvisne spremenljivke ( $X$ ) dobí napovedana vrednost odvisne spremenljivke ( $Y$ ). Ta postopek omogoča učinkovito napovedovanje  $Y$  na podlagi znanih vrednosti  $X$  z uporabo linearne povezave med njima. Primer iskanja takšne statistično povprečne premice je razviden iz ocene podatkov o gradbenih dovoljenjih v San Franciscu (slika 9.1-7), kjer je bila inflacija z uporabo linearne regresije izračunana za različne vrste objektov.

Tabelo s podatki o projektu (slika 9.3-2 iz prejšnjega poglavja) naložimo neposredno v program LLM in ga prosimo, da za nas sestavi preprost model strojnega učenja.

- ✍ Pošljite zahtevo za besedilo v klepet LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Prikazati moramo izdelavo preprostega modela strojnega učenja za napovedovanje stroškov in časa za izvedbo novega projekta X (slika 9.3-2 kot priložena slika). ↴

■ Odgovor LLM :



```

1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

Result  
`(array([238.44444444]), array([3042337.77777778]))`

Slika 9.3-4 LLM je izbral linearno regresijo, da bi ustvaril model strojnega učenja za napovedovanje stroškov in časa projekta.

Program LLM je samodejno prepoznal tabelo iz priložene slike in pretvoril podatke iz vizualne oblike v tabelsko polje (slika 9.3-4 - vrstica 6). To polje je bilo uporabljeno kot osnova za oblikovanje lastnosti in oznak, iz katerih je bil ustvarjen model strojnega učenja (slika 9.3-4 - 17.-22. vrstica), ki je uporabljal linearno regresijo.

Z uporabo osnovnega linearnega regresijskega modela, ki je bil usposobljen na "izredno majhnem" naboru podatkov, so bile izdelane napovedi za nov hipotetični gradbeni projekt z oznako Projekt X. V

našem problemu je za ta projekt značilno 40 stanovanj, 4 nadstropja in stopnja zahtevnosti 7 (slika 9.3-2).

Kot napoveduje linearni regresijski model na podlagi omejenega in majhnega nabora podatkov za novi projekt X (slika 9.3-4 - vrstica 24-29):

- **Gradnja bo trajala** približno 238 dni (238.444444)
- **Skupni odhodki** bodo znašali približno \$ 3.042.338 (3042337,777).

Za nadaljnje raziskovanje hipoteze o stroških projekta je koristno eksperimentirati z različnimi algoritmi in metodami strojnega učenja. Zato predvidimo enake vrednosti stroškov in časa za nov projekt X na podlagi majhnega nabora preteklih podatkov z uporabo algoritma K-Nearest Neighbours (k-NN).

## Napovedi stroškov in časa projekta z uporabo algoritma K-najbližji sosed (k-NN)

Algoritem k-najbližjih sosedov (k-NN) uporabljammo kot dodatni napovedovalec za oceno stroškov in trajanja novega projekta. Algoritem k-Nearest Neighbours (k-NN) je metoda nadzorovanega strojnega učenja (supervised machine learning) za klasifikacijo in regresijo. Algoritem k-NN smo prej obravnavali tudi v kontekstu iskanja po vektorskih podatkovnih zbirkah (slika 8.2-2), kjer se uporablja za iskanje najbližjih vektorjev (npr. besedil, slik ali tehničnih opisov). Pri tem pristopu je vsak projekt predstavljen kot točka v večdimenzionalnem prostoru, kjer vsaka dimenzija ustreza določeni lastnosti projekta.

V našem primeru jih bomo glede na tri atribute vsakega projekta predstavili kot točke v tridimenzionalnem prostoru (slika 9.3-5). Tako bo naš prihajajoči projekt X lokaliziran v tem prostoru s koordinatami ( $x=4$ ,  $y=4$ ,  $z=7$ ). Opozoriti je treba, da sta v realnih razmerah število točk in razsežnost prostora lahko za red velikosti večja.

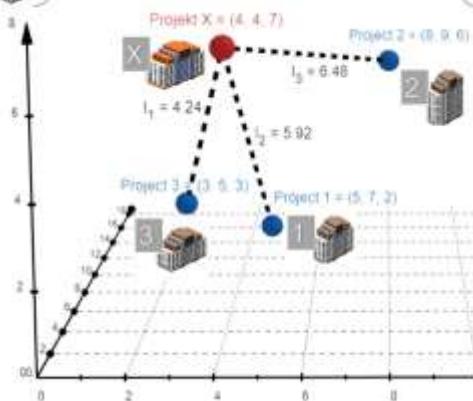
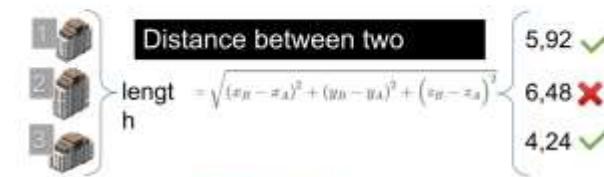
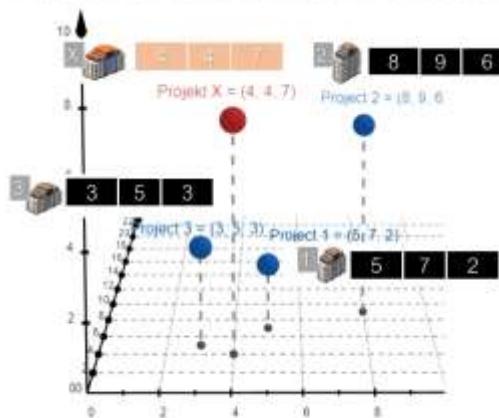
Algoritem K-NN (k-najbližjih sosedov) deluje tako, da meri razdaljo med želenim projektom X in projekti v učni bazi podatkov. S primerjavo teh razdalj algoritem določi projekte, ki so najbližje točki novega projekta X.

Če je na primer drugi projekt ( $x=8$ ,  $y=9$ ,  $z=6$ ) iz našega prvotnega nabora podatkov veliko bolj oddaljen od X (slika 9.3-5) kot drugi projekti, ga lahko izključimo iz nadaljnje analize. Posledično lahko za izračune uporabimo le dva ( $k=2$ ) najbližja projekta, na podlagi katerih bomo določili povprečno vrednost.

Takšna metoda z iskanjem po soseskah omogoča ocenjevanje podobnosti med projekti, kar posledično pomaga pri oblikovanju sklepov o verjetnih stroških in časovnem načrtu novega projekta na podlagi podobnih projektov, ki so bili izvedeni v preteklosti.

### k-nearest neighbors algorithm

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.



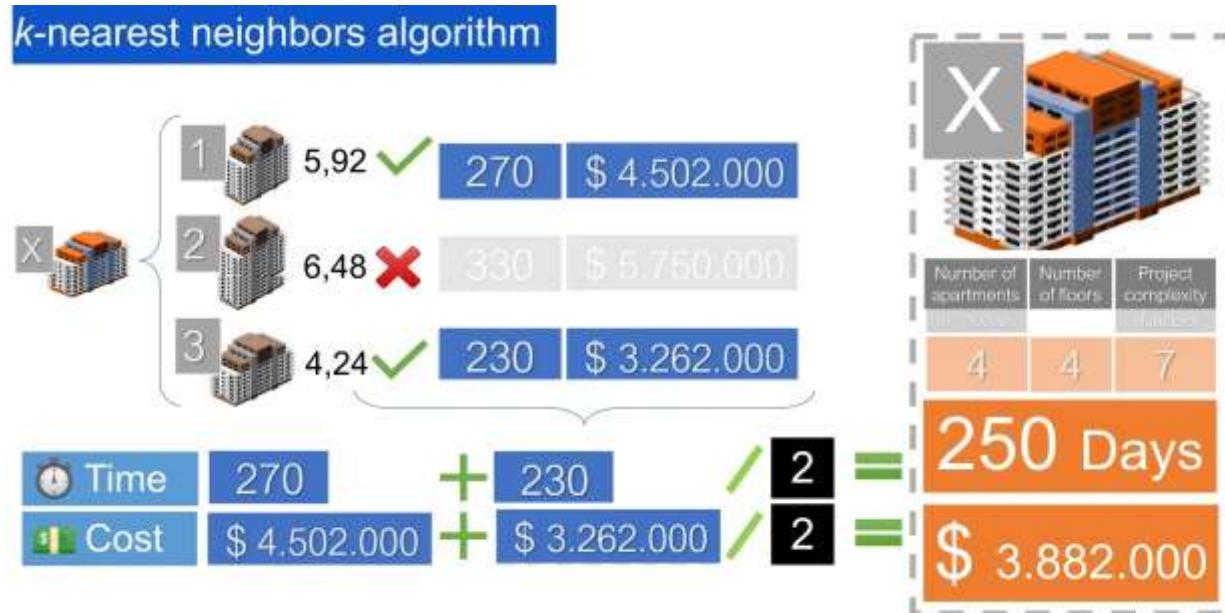
Slika 9.3-5 V algoritmu K-NN so projekti predstavljeni kot točke v večdimensionalnem prostoru, najbližji projekti pa so izbrani na podlagi razdalj za oceno podobnosti in napovedovanje.

Delo k-NN vključuje več ključnih korakov:

- **Priprava podatkov:** najprej se naložita nabora podatkov za usposabljanje in testiranje. Podatki za usposabljanje se uporabljajo za "usposabljanje" algoritma, testni podatki pa za preverjanje njegove učinkovitosti.
- **Izbira parametra K:** izbere se število K, ki določa, koliko najbližjih sosedov (podatkovnih točk) je treba upoštevati v algoritmu. Vrednost "K" je zelo pomembna, saj vpliva na rezultat.
- Postopek razvrščanja in regresija za testne podatke:
  - **Izračun razdalj:** za vsak element iz testnih podatkov se izračuna razdalja do vsakega elementa iz podatkov za usposabljanje (slika 9.3-5). Za to se lahko uporabijo različne metode merjenja razdalj, kot so evklidska razdalja (najpogostejsa metoda), manhattanska razdalja ali Hammingova razdalja.
  - **Razvrščanje in izbira K najbližjih sosedov:** po izračunu razdalj jih razvrstimo in izberemo K najbližjih točk testni točki.
  - **Določitev razreda ali vrednosti testne točke:** če gre za nalogo razvrščanja, se razred testne točke določi na podlagi najpogostejšega razreda med K izbranimi sosedmi. Če gre za nalogo regresije, se izračuna povprečje (ali druga mera osrednje tendenčne) vrednosti K sosedov.
- **Zaključek postopka:** ko so razvrščeni vsi testni podatki ali so zanje izdelane napovedi, je postopek zaključen.

Algoritem k-najbližjih sosedov (k-NN) je učinkovit v številnih praktičnih aplikacijah in je eno glavnih orodij v arzenalu strokovnjakov za strojno učenje. Ta algoritem je priljubljen zaradi svoje preprostosti in učinkovitosti, zlasti pri nalogah, kjer je odnose med podatki enostavno razlagati.

V našem primeru sta bila po uporabi algoritma K-najbližji sosedje določena dva projekta (iz našega majhnega vzorca) z najkrajšo razdaljo do projekta X (slika 9.3-5). Na podlagi teh projektov algoritem določi povprečje njune cene in trajanja gradnje. Po analizi (slika 9.3-6) algoritem s povprečenjem najbližjih sosedov ugotovi, da bo projekt X stal približno \$ 3.800.000 USD in bo trajal približno 250 dni.



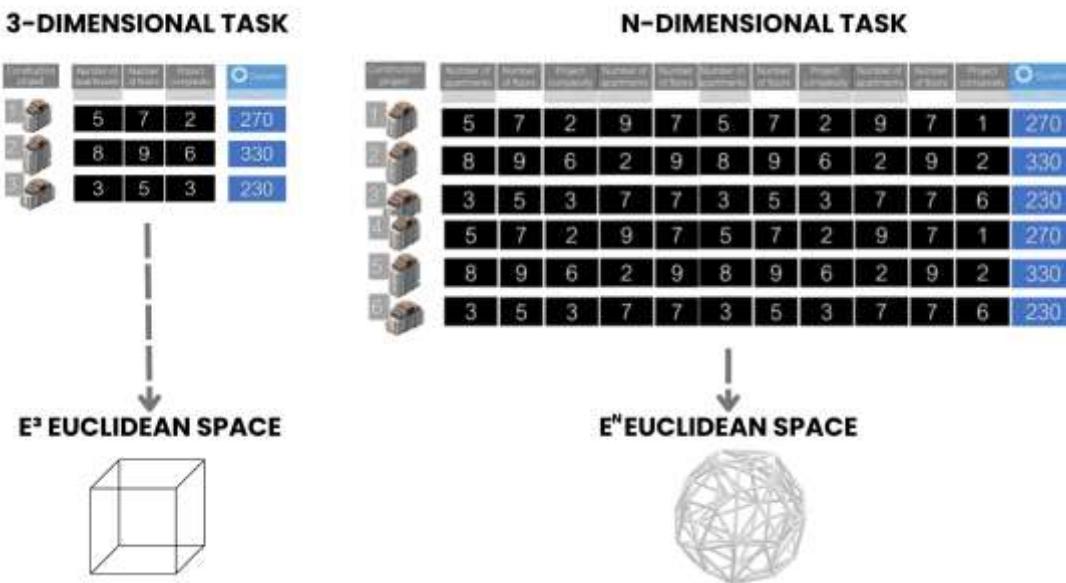
Slika 9.3-6 Algoritem K-najbližjih sosedov določi stroške in časovni razpored projekta X z analizo dveh najbližjih projektov v vzoru.

Algoritem k-najbližjih sosedov (k-NN) je še posebej priljubljen pri nalogah klasifikacije in regresije, kot so sistemi priporočanja, kjer se uporablja za predlaganje izdelkov ali vsebin na podlagi preferenc, podobnih interesom določenega uporabnika. Poleg tega se k-NN pogosto uporablja v medicinski diagnostiki za razvrščanje vrst bolezni na podlagi simptomov pacienta, pri prepoznavanju vzorcev in v finančnem sektorju za ocenjevanje kreditne sposobnosti strank.

Tudi z omejenimi podatki lahko modeli strojnega učenja zagotovijo koristne napovedi in bistveno izboljšajo analitično komponento upravljanja gradbenih projektov. Z razširitvijo in čiščenjem zgodovinskih podatkov je mogoče preiti na zahtevnejše modele - na primer z upoštevanjem vrste gradnje, lokacije, sezone začetka gradnje in drugih dejavnikov.

V naši poenostavljeni nalogi smo za vizualizacijo v 3D prostoru uporabili tri atribute, vendar resnični projekti v povprečju vključujejo na stotine ali tisoče atributov (glej nabor podatkov iz poglavja "Primer

velikih podatkov na podlagi CAD (BIM)", kar znatno poveča dimenzionalnost prostora in zapletenost predstavitev projektov kot vektorjev (slika 9.3-7).



Slika 9.3-7 V poenostavljenem primeru so bili za 3D -vizualizacijo uporabljeni trije atributi, medtem ko jih imajo resnični projekti več.

Z uporabo različnih algoritmov za isti nabor podatkov za projekt X, ki ima 40 stanovanj, 4 nadstropja in 7. stopnjo zahtevnosti, so bile dobljene različne napovedane vrednosti. Algoritem linearne regresije je napovedal čas dokončanja 238 dni in stroške \$ 3.042.338 (slika 9.3-4), medtem ko je algoritem k-NN napovedal 250 dni in \$ 3.882.000 (slika 9.3-6).

Natančnost napovedi, pridobljenih z modeli strojnega učenja, je neposredno odvisna od količine in kakovosti vhodnih podatkov. Več kot je projektov, vključenih v usposabljanje, in bolj kot so njihove značilnosti (atributi) in rezultati (oznake) predstavljeni popolno in natančno, večja je verjetnost za pridobitev zanesljivih napovedi z minimalnimi vrednostmi napak.

Pri tem imajo pomembno vlogo tehnike predobdelave podatkov, ki vključujejo:

- Normalizacija za uskladitev značilnosti na skupni lestvici;
- Odkrivanje in odpravljanje odstopanj, odpravljanje popačenosti modela;
- Kodiranje kategoričnih atributov, ki omogoča manipulacijo z besedilnimi podatki;
- Dopolnjevanje manjkajočih vrednosti, s čimer se poveča robustnost modela.

Poleg tega se uporabljajo metode navzkrižnega preverjanja za oceno splošne uporabnosti modela in njegove odpornosti na nove nabore podatkov, da se odkrije pretirano prilagajanje in izboljša zanesljivost napovedi.

*Kaos je red, ki ga je treba razvozlati [160].*

- José Saramago, "Dvojnik"

Tudi če se vam zdi, da kaosa vaših nalog ni mogoče formalno opisati, morate vedeti, da za vsak dogodek v svetu in zlasti za gradbene procese veljajo matematični zakoni, ki morda potrebujejo podporo pri izračunu vrednosti ne s strogimi formulami, temveč s pomočjo statističnih in zgodovinskih podatkov.

Tako tradicionalne ocene stroškovnega oddelka kot modeli strojnega učenja se neizogibno soočajo z negotovostjo in možnimi viri napak. Vendar pa lahko modeli strojnega učenja, če je na voljo dovolj kakovostnih podatkov, pokažejo primerljivo in včasih celo večjo natančnost napovedi kot strokovne ocene.

Strojno učenje bo verjetno postal zanesljivo dopolnilno orodje za analize, ki lahko: izpopolni izračune, predlaga alternativne scenarije in ugotavlja skrite odvisnosti med parametri projekta. Takšni modeli ne bodo trdili, da so univerzalni, vendar bodo kmalu zavzeli pomembno mesto v postopkih izračunov in odločanja. Tehnologije strojnega učenja ne bodo izključile sodelovanja inženirjev, ocenjevalcev in analitikov, temveč bodo, nasprotno, razširile njihove zmožnosti, saj bodo ponudile dodatno stališče, ki bo temeljilo na zgodovinskih podatkih.

Če bo strojno učenje ustrezno vključeno v poslovne procese gradbenih podjetij, lahko postane pomemben element sistema za podporo odločanju vodstva - ne kot nadomestilo za ljudi, temveč kot razširitev njihove strokovne intuicije in inženirske logike.

## Naslednji koraki: od shranjevanja do analize in napovedovanja

Sodobni pristopi k delu s podatki začenjajo spominjati način sprejemanja odločitev v gradbeništvu. Prehod od intuitivnih ocen k objektivni analizi podatkov ne izboljšuje le natančnosti, temveč odpira tudi nove možnosti za optimizacijo procesov. Za povzetek tega dela velja izpostaviti glavne praktične korake, ki vam bodo pomagali pri uporabi obravnnavanih metod pri vsakodnevnih nalogah:

- Vzpostavitev trajnostne infrastrukture za skladiščenje
  - Poskusite združiti različne dokumente in projektne podatke v enoten tabelarični model in združiti ključne informacije v enoten podatkovni okvir za nadaljnjo analizo.
  - uporaba učinkovitih formatov za shranjevanje podatkov - na primer stolpčnih formatov, kot je Apache Parquet, namesto formatov CSV ali XLSX - zlasti za tiste nize, ki bi se lahko v prihodnosti uporabili za učenje modelov strojnega učenja.
  - Vzpostavitev sistema različic podatkov za sledenje spremembam med projektom.
- Izvajanje orodij za analizo in avtomatizacijo

- Začnite analizirati pretekle podatke o projektih - z dokumentacijo, modeli, ocenami - da bi ugotovili vzorce, trende in anomalije.
- Glavni procesi ETL (Ekstrakcija, Transformacija, Nalaganje) za samodejno nalaganje in pripravo podatkov.
- Naučite se, kako vizualizirati ključne metrike z uporabo različnih brezplačnih knjižnic za vizualizacijo v Pythonu.
- Začetek uporabe statističnih metod in naključnega vzorčenja za pripravo reprezentativnih in ponovljivih analitičnih rezultatov.

#### ■ Večja zrelost pri delu s podatki

- Naučite se nekaj osnovnih algoritmov strojnega učenja s preprostimi in enostavnimi primeri, kot je nabor podatkov o Titaniku.
- Analizirajte trenutne procese in ugotovite, kje lahko preidete od toge logike vzrokov in posledic k statističnim metodam napovedovanja in ocenjevanja.
- Začnite obravnavati podatke kot strateško sredstvo in ne kot stranski proizvod: gradite procese odločanja na podatkovnih modelih in ne na specifičnih programskih rešitvah.

Gradbena podjetja, ki so spoznala vrednost podatkov, vstopajo v novo fazo razvoja, v kateri konkurenčne prednosti ne določa količina virov, temveč hitrost odločanja na podlagi analitike.



## NAJVEČJE UDOBJE S TISKANO RAZLIČICO

V rokah imate brezplačno digitalno različico knjige **Data-Driven Construction**. Za udobnejše delo in hiter dostop do gradiva vam priporočamo, da se posvetite **tiskani izdaji**:



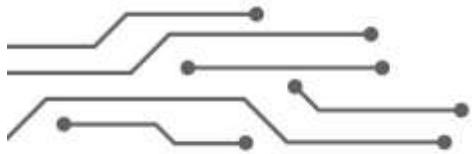
**Vedno pri roki:** knjiga v tiskani obliki bo postala zanesljivo delovno orodje, s katerim boste lahko hitro našli in uporabili prave vizualizacije in diagrame v vsaki delovni situaciji.

**Visoka kakovost ilustracij:** vse slike in grafike v tiskani izdaji so predstavljene v najvišji kakovosti.

**Hiter dostop do informacij:** enostavna navigacija, možnost beleženja opomb, zaznamkov in dela s knjigo kjer koli.

Z nakupom polne tiskane različice knjige dobite priročno orodje za udobno in učinkovito delo z informacijami: možnost hitre uporabe vizualnega gradiva pri vsakodnevnih opravilih, hitrega iskanja potrebnih shem in zapisovanja. Poleg tega s svojim nakupom podpirate širjenje odprtega znanja.

Tiskano različico knjige naročite na: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



X DEL

## **GRADBENIŠTVO V DIGITALNI DOBI. PRILOŽNOSTI IN IZZIVI**

Zadnji, deseti del je celovit pogled na prihodnost gradbeništva v dobi digitalne preobrazbe. V njem je analiziran prehod od vzročno-posledične analize k delu s korelacijami velikih količin podatkov. Povzete so vzporednice med razvojem likovne umetnosti in razvojem dela s podatki v gradbeništvu, ki kažejo, kako se panoga premika od podrobnega nadzora k celostnemu razumevanju procesov. Raziskan je koncept "überizacije" gradbene industrije, kjer lahko preglednost podatkov in avtomatizirani izračuni radikalno spremenijo tradicionalne poslovne modele, odpravijo potrebo po posrednikih in zmanjšajo možnosti za špekulacije. Podrobno so obravnavana nerešena vprašanja, kot je univerzalna klasifikacija elementov, ki gradbenim podjetjem dajejo čas, da se prilagodijo novemu okolju. Del se zaključi s konkretnimi priporočili za oblikovanje strategije digitalne preobrazbe, ki vključuje analizo ranljivosti in razširitev storitev, da bi ostali konkurenčni v spremenjajoči se panogi.

## POGLAVJE 10.1.

# STRATEGIJE PREŽIVETJA: USTVARJANJE KONKURENČNIH PREDNOSTI

### Korelacijske namesto izračunov: prihodnost gradbene analitike

Zaradi hitre digitalizacije informacij (slika 1.1-5) sodobno gradbeništvo doživlja temeljno preobrazbo, pri kateri podatki ne postajajo le orodje, temveč strateško sredstvo, ki lahko temeljito spremeni tradicionalne pristope k upravljanju projektov in poslovanja.

Že tisočletja se gradbeništvo opira na deterministične metode - natančne izračune, izdelavo detajlov in strog nadzor parametrov. V prvih stoletjih našega štetja so rimski inženirji uporabljali matematična načela pri gradnji akvaduktov in mostov. V srednjem veku so si arhitekti prizadevali za idealna razmerja gotskih katedral, v industrializaciji 20. stoletja pa so se oblikovali sistemi standardiziranih normativov in predpisov, ki so postali osnova za množično gradnjo.

Danes se vektor razvoja od iskanja strogih vzročno-posledičnih povezav preusmerja k verjetnostni analizi, iskanju korelacij in skritih vzorcev. Industrija vstopa v novo fazo - podatki postajajo ključni vir, na njih temelječa analitika pa nadomešča intuitivne in lokalno optimizirane pristope.



Slika 10.1-1 Skriti potencial gradbenih podatkov: obstoječi izračuni v podjetju so le vrh ledene gore, ki je na voljo vodstvu za analizo.

Informacijski sistem podjetja je kot ledena gora (slika 10.1-1): vodstvu podjetja je viden le majhen del potenciala podatkov, medtem ko je glavna vrednost skrita v globini. Pomembno je, da podatkov ne ocenjujemo le glede na njihovo trenutno uporabo, temveč tudi glede na priložnosti, ki jih bodo odprli v prihodnosti. Trajno konkurenčno prednost bodo ustvarila tista podjetja, ki se bodo naučila iz podatkov izluščiti skrite vzorce in ustvariti nova spoznanja

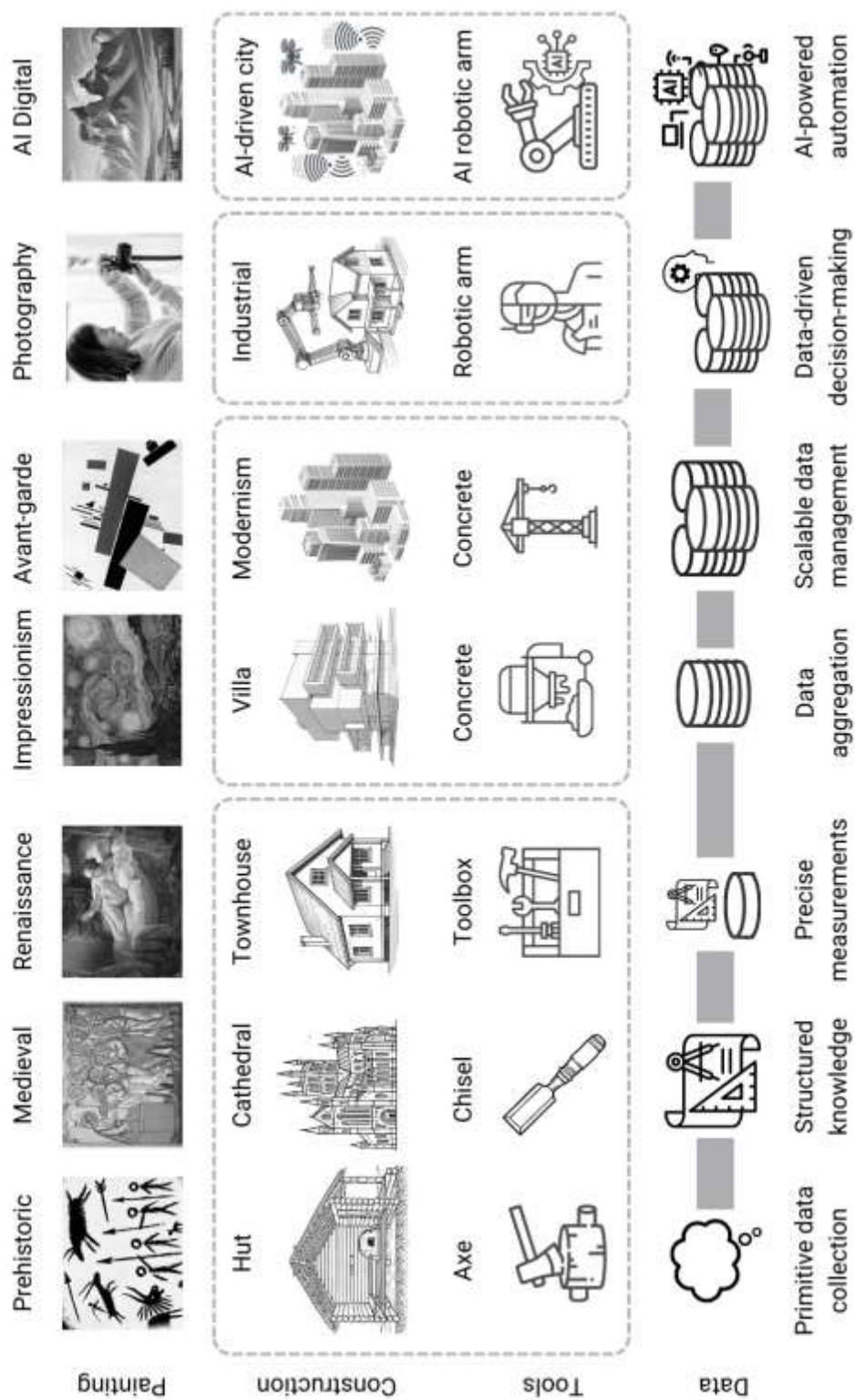
Iskanje skritih vzorcev in osmišljanje podatkov ni le delo s številkami, temveč ustvarjalni proces, ki zahteva abstraktno razmišljanje in sposobnost videti celotno sliko za različnimi elementi. V tem smislu lahko razvoj dela s podatki primerjamo z razvojem likovne umetnosti (slika 10.1-2).

Razvoj gradbeništva je zelo podoben razvoju likovne umetnosti. V obeh primerih je človeštvo napredovalo od primitivnih metod do izpopolnjenih tehnologij vizualizacije in analize. V prazgodovini so ljudje za reševanje vsakodnevnih nalog uporabljali jamske risbe in primitivna orodja. V srednjem veku in renesansi se je raven izpopolnjenosti arhitekture in umetnosti močno povečala. Do zdajnjega srednjega veka so se gradbena orodja razvila od preproste sekire do obsežnih kompletov orodij, ki simbolizirajo rast tehničnega znanja.

Doba realizma je bila prva revolucija v likovni umetnosti: umetniki so se naučili, kako reproducirati najmanjše podrobnosti in doseči največjo verodostojnost. V gradbeništvu so bile ustreznica tega obdobja natančne inženirske tehnike, podrobne risbe in strogo predpisani izračuni, ki so za stoletja postali osnova projektantske prakse.

Kasneje je impresionizem spremenil samo dojemanje umetniške resničnosti: namesto dobesednega upodabljanja oblike so umetniki začeli zajemati razpoloženje, svetlobo in dinamiko ter si prizadevali odražati splošni vtis in ne absolutno natančnost. Podobno se tudi strojno učenje v gradbeni analitiki odmika od togih logičnih modelov k prepoznavanju vzorcev in verjetnostnih vzorcev, ki omogočajo "videti" skrite odvisnosti v podatkih, ki so pri klasični analizi nedostopne. Ta pristop je odmev na ideje Bauhausa o minimalizmu in funkcionalnosti, kjer je pomen (funkcija) pomembnejši od oblike. Bauhaus si je prizadeval odstraniti odvečno, opustiti okrasje zaradi jasnosti, uporabnosti in množične privlačnosti. Stvari so morale biti razumljive in uporabne, brez presežkov - estetika se je rodila iz logike oblikovanja in namena.

S pojavom fotografije konec 19. stoletja je umetnost dobila novo orodje, s katerim je lahko z nezaslišano natančnostjo ujela resničnost, in obrnila odnos do likovne umetnosti na glavo. Podobno se v gradbeništvu zaradi industrijske revolucije v 21. stoletju uporabljajo robotske tehnologije, laserji, internet stvari, RFID in koncepti, kot je Connected Construction, kjer se je zbiranje posameznih parametrov razvilo v skalabilni inteligentni zajem celotne resničnosti gradbišča.



Slika 10.1-2 Era likovnega razvoja je skladna z pristopov k delu s podatki v gradbeništvu.

Tako kot se danes v likovni umetnosti s prihodom orodij umetne inteligence in LLM na novo razmišlja,

se v gradbeništvu dogaja še en kvantni preskok: inteligentni sistemi, ki jih poganja umetna inteligenca (AI), LLM chat omogočajo napovedovanje, optimizacijo in generiranje rešitev z minimalnim človeškim posredovanjem.

Vloga podatkov pri načrtovanju in upravljanju se je korenito spremenila. Medtem ko se je včasih znanje prenašalo ustno in je bilo empirične narave - tako kot je bila resničnost do 19. stoletja ujeta z ročno naslikanimi slikami -, je danes v ospredju popoln digitalni zajem gradbene "slike". S pomočjo algoritmov strojnega učenja se ta digitalna slika pretvori v impresionistično predstavitev gradbene realnosti - ne v natančno repliko, temveč v posplošeno, verjetnostno razumevanje procesov.

Hitro se približujemo obdobju, v katerem bodo procese načrtovanja, gradnje in obratovanja stavb ne le nadgradili, temveč jih bodo v veliki meri vodili sistemi umetne inteligence. Tako kot sodobna digitalna umetnost nastaja brez čopiča - s pomočjo besedilnih napotkov in generativnih modelov - bodo arhitekturne in inženirske rešitve prihodnosti oblikovane s ključnimi poiizvedbami in parametri, ki jih bo določil uporabnik.

V 21. stoletju postajajo dostop do podatkov, njihova interpretacija in kakovost analitike nepogrešljivi za uspeh projekta. Vrednosti podatkov pa ne določa njihova količina, temveč sposobnost strokovnjakov, da jih analizirajo, preverijo in spremenijo v dejanja.

## Na podatkih temelječ pristop v gradbeništvu: nova raven infrastrukture

V zgodovini človeštva je vsak tak tehnološki preskok prinesel temeljne spremembe v gospodarstvu in družbi. Danes smo priča novemu valu preobrazbe, ki je po obsegu primerljiv z industrijsko revolucijo v 19. stoletju. Vendar so bile pred sto leti glavno gonilo sprememb mehanske sile in energetske tehnologije, zdaj pa so to podatki in umetna inteligenca.

Agenti strojnega učenja, LLM in umetne inteligence spreminjajo samo bistvo aplikacij, zaradi česar tradicionalni skladi programske opreme (obravnavani v drugem delu knjige) niso več potrebni (slika 2.2-3). Vsa podatkovna logika je osredotočena v agentih umetne inteligence in ne v trdno zakodiranih poslovnih pravilih (slika 2.2-4).

V podatkovni dobi se tradicionalni pogledi na aplikacije temeljito spreminjajo. Prehajamo na model, v katerem bodo obsežni, modularni sistemi za podjetja neizogibno zamenjali odprte, lahke in prilagojene rešitve.

*V prihodnosti bo ostala le osnovna podatkovna struktura, vsa interakcija z njo pa bo potekala prek agentov, ki bodo delali neposredno s podatkovno bazo. Resnično verjamem, da bo izginila celotna aplikacijska struktura, saj je umetna inteligenca v neposredni interakciji z osnovno podatkovno bazo preprosto ne bo potrebovala. Celotno svojo kariero sem preživel v podjetjih SaaS - gradil sem podjetja, delal v njih in če sem iskren, verjetno zdaj ne bi zagnal novega podjetja SaaS. In verjetno zdaj ne bi vlagal v podjetja SaaS. Razmere so preveč negotove. To ne pomeni, da v prihodnosti ne bodo obstajala podjetja za programsko opremo, le da bodo videti povsem drugače. Prihodnji sistemi bodo podatkovne zbirke s poslovno logiko, ki bo vključena v agente [AI]. Ti agenti bodo delali z več podatkovnimi skladišči hkrati in ne bodo omejeni na eno samo podatkovno zbirko. Vsa logika se bo preselila v plast umetne inteligence [46].*

- Matthew Berman, izvršni direktor Forward Future

Ključna razlika nove paradigme je zmanjšanje tehnološkega balasta. Namesto monumentalnih zapletenih in zaprtih programskih sistemov bomo dobili prilagodljive, odprte in hitro prilagodljive module, ki dobesedno "živijo" znotraj podatkovnega toka (slika 7.4-1 - Apache Airflow, NiFi). Arhitektura prihodnjega upravljanja procesov predvideva uporabo mikroaplikacij - kompaktnih, namensko zasnovanih orodij, ki se bistveno razlikujejo od obsežnih in zaprtih sistemov ERP, PMIS, CDE, CAFM. Novi agenti bodo čim bolj prilagodljivi, integrirani in poslovno specifični (npr. nizkokodni/No-Code Slika 7.4-6).

*Vsa poslovna logika bo prenesena na te agente [AI], ki bodo izvajali operacije CRUD [Create, Read, Update in Delete] na več skladiščih, kar pomeni, da ne bodo razlikovali, katero zaledje se uporablja. Posodabljalci bodo več podatkovnih zbirk, vsa logika pa bo končala na tako imenovani ravni AI. In ko bo raven umetne inteligence tista, kjer bo vsa logika, bodo ljudje začeli zamenjevati zaledne sisteme. Že zdaj opažamo precej visok odstotek tržnih zmag pri uporabi zalednih strežnikov in agentov Dynamics, zato se bomo agresivno gibali v tej smeri in poskušali vse to združiti. Ne glede na to, ali gre za storitve za stranke ali druga področja, na primer ne le CRM, temveč tudi naše finančne in operativne rešitve. Ljudje namreč želijo več poslovnih aplikacij, ki jih poganja umetna inteligenca, kjer lahko logično raven poganjajo umetna inteligenca in agenti umetne inteligence. [...]. Ena najbolj vznemirljivih stvari zame je Excel s Pythonom, kar je primerljivo z GitHubom s Copilotom. To pomeni, kaj smo naredili: zdaj, ko imate Excel, ga preprosto odprite, zaženite program Copilot in se začnite igrati z njim. Ne gre več samo za razumevanje razpoložljivih številk - sam bo izdelal načrt. Tako kot delovni prostor GitHub Copilot ustvari načrt in ga nato izvede, je kot podatkovni analistik, ki uporablja Excel kot orodje za vizualizacijo vrstic in stolpcev za analizo. Copilot torej uporablja Excel kot orodje z vsemi njegovimi zmožnostmi, saj lahko ustvarja podatke in ima tolmač Python.*

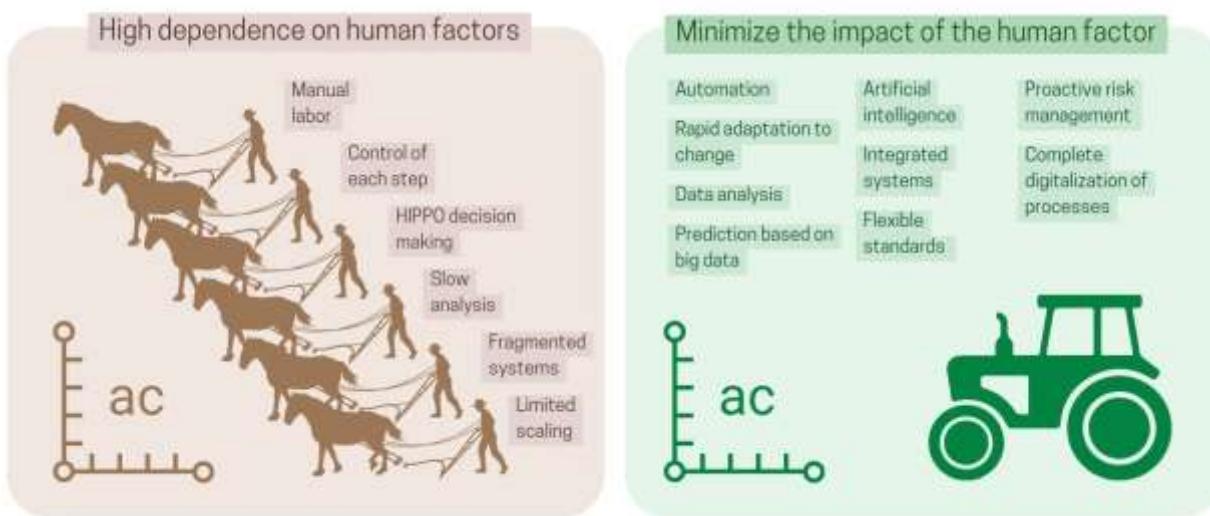
- Satya Nadella, izvršni direktor Microsoft, intervju s kanalom BG2, december 2024. [28]

Preobrazba, ki smo ji priča v logiki pisarniških aplikacij - prehod z modularnih, zaprtih sistemov na agente umetne inteligence, ki neposredno delajo z odprtimi podatki - je le del veliko večjega procesa. Ne gre le za spreminjanje vmesnikov ali arhitekture programske opreme: spremembe bodo vplivale na

temeljna načela organizacije dela, odločanja in upravljanja podjetij. V gradbeništvu bo to pivedlo do podatkovno usmerjene logike, v kateri bodo podatki postali osrednji element procesov od načrtovanja do upravljanja virov in spremeljanja gradnje.

## Digitalna pisarna naslednje generacije: kako umetna inteligenca spreminja delovni prostor

Pred skoraj stoletjem je človeštvo že doživelovalo podobno tehnološko revolucijo. Prehod s parnih strojev na električne motorje je trajal več kot štiri desetletja, vendar je na koncu spodbudil izjemno rast produktivnosti - predvsem zaradi decentralizacije energetskih zmogljivosti in prilagodljivosti novih rešitev. Ta premik ni spremenil le poteka zgodovine, saj se je večina prebivalstva preselila s podeželja v mesta, temveč je postavil tudi temelje sodobnega gospodarstva. Zgodovina tehnologije je pot od fizičnega dela do avtomatizacije in inteligentnih sistemov. Tako kot je traktor nadomestil številne obdelovalce zemlje, sodobna digitalna tehnologija nadomešča tradicionalne pisarniške metode vodenja gradnje (slika 10.1-3). Še na začetku 20. stoletja je večina svetovnega prebivalstva zemljo obdelovala ročno, dokler se v tridesetih letih 20. stoletja ni začela mehanizacija dela s stroji in traktorji.



Slika 10.1-3 Tako kot je traktor v začetku 20. stoletja nadomestil več deset ljudi, bo strojno učenje v 21. stoletju nadomestilo tradicionalne metode upravljanja podjetij in projektov.

Tako kot je človeštvo pred sto leti prešlo od obdelovanja posameznih zemljišč s primitivnimi orodji k obsežnemu kmetovanju z uporabo strojev, danes prehaja od obdelave različnih "silosov" informacij k delu z množicami podatkov s pomočjo zmogljivih "traktorjev" - cevovodov ETL in algoritmов umetne inteligence.

Smo na pragu podobnega preskoka, vendar na digitalni ravni: od tradicionalnega, ročnega upravljanja poslovanja k modelom, ki temelji na podatkih.

Pot do popolne podatkovno usmerjene arhitekture bo zahtevala čas, naložbe in organizacijska prizadevanja. Vendar ta pot ne odpira poti le do postopnega izboljšanja, temveč do kakovostnega

preskoka v smeri večje učinkovitosti, preglednosti in obvladljivosti gradbenih procesov. Vse to je pogojeno s sistematično uvedbo digitalnih orodij in opustitvijo zastarelih poslovnih praks.

Parametrizacija nalog, ETL, LLM, komponente interneta stvari, RFID, tokenizacija, veliki **podatki** in strojno učenje bodo tradicionalno gradbeništvo spremenili v **podatkovno vodeno gradbeništvo**, pri katerem bodo vsako podrobnost projekta in gradbenega poslovanja nadzorovali in optimizirali podatki.

Včasih je bilo za analizo informacij potrebnih na tisoče delovnih ur. Zdaj te naloge opravljajo algoritmi in programi LLM, ki s pomočjo pozivov spremenijo različne podatkovne nize v strateške vire. V tehnološkem svetu se dogaja isto, kar se je zgodilo v kmetijstvu: od motik prehajamo k avtomatiziranemu kmetijstvu. Tako tudi pisarniško delo v gradbeništvu - od Excelovih datotek in ročnega povzemanja - prehaja k inteligentnemu sistemu, kjer se podatki zbirajo, čistijo, strukturirajo in spreminjajo v vpoglede.

Podjetja morajo že danes začeti "gojiti" informacijska polja s kakovostnim zbiranjem podatkov in strukturiranjem informacij, jih "gnojiti" z orodji za čiščenje in normalizacijo, nato pa jih "pobirati" v obliki napovedne analitike in avtomatiziranih rešitev. Če lahko sodobni kmet s strojem nadomesti sto obdelovalcev zemlje, bodo inteligentni algoritmi lahko zaposlenim odvzeli rutino in jih prenesli v vlogo strateških upravljavcev informacijskih tokov.

Vendar pa je treba razumeti, da vzpostavitev resnično podatkovno usmerjene organizacije ni hiter proces. Gre za dolgoročno strateško usmeritev, podobno kot pri ustvarjanju novega mesta za gojenje novega gozda (slika 1.2-5) sistemov, kjer je vsako "drevo" v tem ekosistemu ločen proces, kompetenca ali orodje, ki potrebuje čas za rast in razvoj. Tako kot v primeru pravega gozda uspeh ni odvisen le od kakovosti sadilnega materiala (tehnologija), temveč tudi od tal (kultura podjetja), podnebja (poslovno okolje) in nege (sistemski pristop).

Podjetja se ne bodo mogla več zanašati le na zaprte, vnaprej pripravljene rešitve. V nasprotju s prejšnjimi fazami tehnološkega razvoja sedanjega prehoda - na odprt dostop do podatkov, umetno inteligenco in odprto kodo - veliki ponudniki verjetno ne bodo podprli, saj neposredno ogroža njihove uveljavljene poslovne modele in osnovne tokove prihodkov.

Kot je pokazala študija Harvard Business School [40], ki je bila obravnavana že v poglavju o četrti in peti tehnološki revoluciji, bi stroški ustvarjanja najbolj uporabljenih odprtakodnih rešitev iz nič za vsa podjetja znašali približno 4,15 milijarde dolarjev. Če pa si predstavljamo, da bi vsako podjetje razvilo svoje lastne alternative brez dostopa do obstoječih odprtakodnih orodij, kar se je dogajalo zadnja desetletja, bi lahko skupni poslovni stroški dosegli kolosalnih 8,8 bilijona dolarjev - to je cena iracionalnega povpraševanja, na katero je mogoče ovrednotiti trg programske opreme.

Tehnološki napredek bo neizogibno privедel do ponovnega premisleka o uveljavljenih poslovnih modelih. Medtem ko so podjetja včasih lahko zaslužila na podlagi zapletenih, nepreglednih procesov in zaprtih podatkov, z razvojem umetne intelligence in analitike ta pristop postaja vse manj smiseln.

Zaradi demokratizacije dostopa do podatkov in orodij se lahko tradicionalni trg prodaje programske opreme precej skrči. Hkrati pa se bo povečal nov trg - trg digitalnega strokovnega znanja, prilagajanja, integracije in oblikovanja rešitev. Tu vrednost ne bo izhajala iz prodaje licenc, temveč iz zmožnosti vzpostavitev prožnih, odprtih in prilagodljivih digitalnih procesov. Tako kot sta elektrifikacija in prihod traktorjev rodila nove industrije, tako bo tudi uporaba velikih podatkov, umetne inteligence in LLM odprla povsem nova obzorja za poslovanje v gradbeništvu, kar pa ne bo zahtevalo le tehnoloških naložb, temveč tudi temeljito preoblikovanje miselnosti, procesov in organizacijskih struktur. In tista podjetja in strokovnjaki, ki se bodo tega zavedali in začeli ukrepati že danes, bodo vodilni v prihodnosti.

V svetu, v katerem odpti podatki postajajo glavna prednost, bo razpoložljivost informacij spremenila pravila igre. Vlagatelji, stranke in regulatorji bodo vse bolj zahtevali preglednost, algoritmi strojnega učenja pa bodo lahko samodejno ugotavljali neskladja v ocenah, časovnih okvirih in stroških. To ustvarja pogoje za novo stopnjo digitalne preobrazbe, ki nas postopoma vodi v "uberizacijo" gradbeništva.

## **Odpti podatki in uberizacija ogrožajo obstoječa gradbena podjetja**

Gradbeništvo postaja proces upravljanja informacij. Natančnejši, kakovostnejši in popolnejši kot so podatki, učinkovitejši so projektiranje, izračuni, ocene stroškov, gradnja in obratovanje stavb. V prihodnosti ključni vir ne bodo žerjav, beton in armatura, temveč sposobnost zbiranja, analiziranja in uporabe informacij.

V prihodnosti bodo stranke gradbenih podjetij - investitorji in stranke, ki financirajo gradnjo - neizogibno izkoristile vrednost odprtih podatkov in analitike preteklih podatkov. S tem se bodo odprle možnosti za avtomatiziran izračun časovnega načrta projekta in stroškov, ne da bi gradbena podjetja vključevala v vprašanja izračunavanja stroškov, kar bo pripomoglo k nadzoru stroškov in hitrejšemu odkrivanju nepotrebnih stroškov.

Predstavljajte si gradbišče, na katerem laserski skenerji, kvadrokopterji in fotogrametrični sistemi v realnem času zbirajo natančne podatke o količini porabljenega betona. Te informacije se samodejno pretvorijo v preproste ploščate modele MESH z metapodatki, pri čemer se zaobidejo okorni sistemi CAD (BIM), brez odvisnosti od zapletenih geometrijskih jeder, ERP ali PMIS. Ti podatki, zbrani na gradbišču, se centralno prenesejo v enoto strukturirano shrambo, ki je naročniku na voljo za neodvisno analizo, kamor se naložijo dejanske cene iz različnih gradbenih trgovin s parametri, ki segajo od stopnje kreditnega financiranja do dinamično spreminjačih se dejavnikov, kot so vremenske razmere, borzne kotacije gradbenih materialov, logistične tarife in statistična sezonska nihanja cen dela. V takšnih razmerah so vsa odstopanja med načrtovanimi in dejanskimi količinami materiala takoj očitna, kar onemogoča manipuliranje z ocenami tako v fazi načrtovanja kot ob predaji projekta. Preglednosti gradbenega procesa tako ne zagotavlja vojska nadzornikov in menedžerjev, temveč objektivni digitalni podatki, ki zmanjšajo človeški dejavnik in možnost špekulacij.

V prihodnosti bodo tovrstno delo nadzora podatkov raje opravljali upravljavci podatkov na strani stranke (slika 1.2-4 Upravljač CQMS). To še posebej velja za kalkulacije in projektne ocene: tam, kjer je bil včasih celoten oddelek ocenjevalcev, bodo že jutri na voljo orodja za strojno učenje in napovedovanje, ki bodo gradbenim podjetjem določala cenovne omejitve, v katere se bodo lahko umestila.

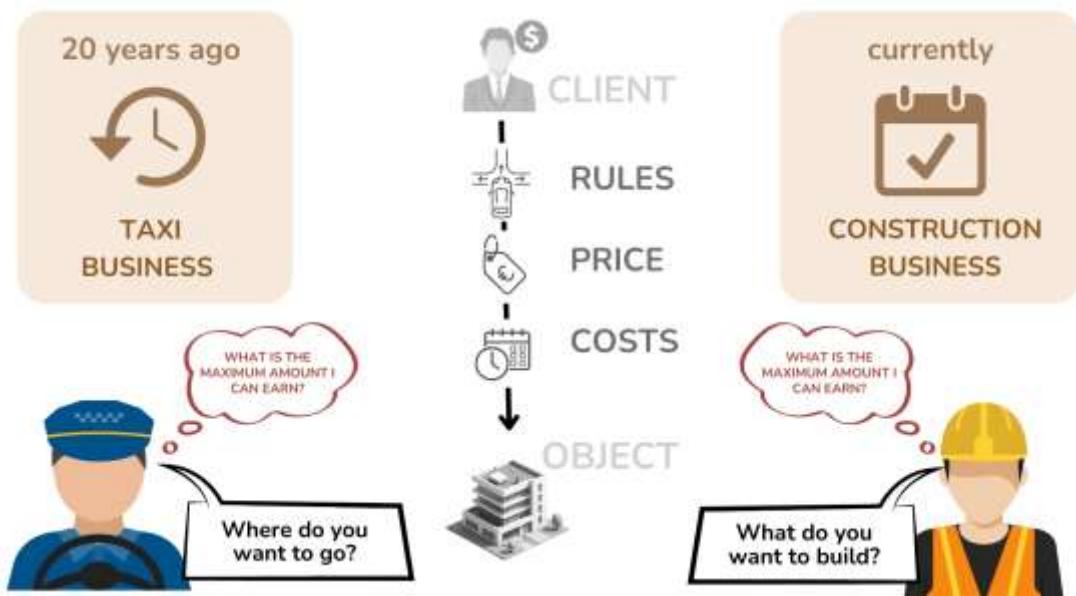
*Glede na razdrobljenost [gradbene] industrije, kjer večino sistemov in podsistemov dobavljajo mala in srednje velika podjetja, mora digitalna strategija izhajati od strank. Naročniki morajo ustvariti pogoje in mehanizme za sprostitev digitalnih zmogljivosti dobavne verige [20].*

- Andrew Davis in Giuliano Denicol, Accenture "Ustvarjanje večje vrednosti s kapitalskimi projekti"

Takšna odprtost in preglednost podatkov ogrožata gradbena podjetja, ki so navajena zaslужiti z nepreglednimi postopki in zmedenimi poročili, kjer se špekulacije in skriti stroški lahko skrivajo za zapletenimi in zaprtimi formati ter modularnimi lastniškimi podatkovnimi platformami. Zato je malo verjetno, da bodo gradbena podjetja, tako kot prodajalci, ki spodbujajo odprtakodne rešitve, zainteresirana za popolno uvedbo odprtih podatkov v svoje poslovne procese. Če so podatki na voljo in jih je za stranko enostavno obdelati, jih je mogoče samodejno preveriti, kar bo odpravilo možnost precenjevanja količin in manipuliranja s predračuni.

V skladu s poročilom Svetovnega gospodarskega foruma "Oblikovanje prihodnosti gradbeništva" 2016) [5] ostaja eden od ključnih izzivov panoge pasivna vloga naročnika. Kljub temu so naročniki tisti, ki bi morali prevzeti večjo odgovornost za rezultate projektov - od zgodnjega načrtovanja, izbire trajnostnih modelov interakcije do spremljanja uspešnosti. Brez aktivnega sodelovanja lastnikov projektov sistemsko preobrazba gradbene industrije ni mogoča.

Izguba nadzora nad izračunom obsega in stroškov je v zadnjih 20 letih že spremenila druge panoge, saj je strankam omogočila, da neposredno, brez posrednikov, ovirajo svoje cilje. Digitalizacija in preglednost podatkov sta preoblikovali številne tradicionalne poslovne modele, kot so taksisti s pojavom Uberja (slika 10.1-4), hotelirji s prihodom Airbnb ter trgovci na drobno in trgovine z vzponom Amazona ter banke z vzponom nebank in decentraliziranih fintech ekosistemov, kjer sta neposreden dostop do informacij ter avtomatizacija izračunov časa in stroškov bistveno zmanjšala vlogo posrednikov.



Slika 10.1-4 Gradbeništvo se bo soočilo z uberizacijo, s katero so se pred desetimi leti soočali taksisti, hotelirji in trgovci na drobno.

Proces demokratizacije dostopa do podatkov in orodij za njihovo obdelavo je neizogiben, sčasoma pa bodo odprti podatki o vseh komponentah projekta postali zahteva strank in nov standard. Zato bodo vprašanja uvedbe odprtih formatov in preglednih izračunov spodbujali investitorji, stranke, banke in skladi zasebnega kapitala (private equity) - tisti, ki so končni uporabniki zgrajenih objektov in nato objekt upravljajo desetletja.

Večji vlagatelji, stranke in banke že zahtevajo preglednost v gradbeništvu. Glede na Accenturino študijo "Ustvarjanje veče vrednosti s kapitalskimi projektmi" (2020) [20] postajajo pregledni in zanesljivi podatki odločilni dejavnik pri odločanju o naložbah v gradbeništvu. Kot poudarjajo strokovnjaki, zaupanja vredno in učinkovito vodenje projektov ni mogoče brez preglednosti, zlasti v kriznih časih. Poleg tega se lastniki sredstev in izvajalci vse bolj usmerjajo k pogodbam, ki spodbujajo izmenjavo podatkov in sodelovalno analitiko, kar odraža vse večje zahteve vlagateljev, bank in regulatorjev po odgovornosti in preglednosti.

Gibanje investitorja, stranke od ideje do dokončane stavbe, bo v prihodnosti podobno potovanju z avtopilotom - brez voznika v obliki gradbenega podjetja, obljudbla, da bo postalo neodvisno od špekulacij in negotovosti.

Doba odprtih podatkov in avtomatizacije bo neizogibno spremenila gradbeništvo, tako kot se je to že zgodilo v bančništvu, trgovini, kmetijstvu in logistiki. V teh panogah se vloga posrednikov in tradicionalni načini poslovanja umikajo avtomatizaciji in robotizaciji, pri čemer ni več prostora za neupravičene pribitke in špekulacije.

Podatki in procesi v vseh človekovih gospodarskih dejavnostih se v ničemer ne razlikujejo od tistih, s

katerimi se srečujejo strokovnjaki v gradbeništvu. Dolgoročno lahko gradbena podjetja, ki danes prevladujejo na trgu z določanjem standardov cen in kakovosti storitev, izgubijo vlogo ključnega posrednika med stranko in njenim gradbenim projektom.

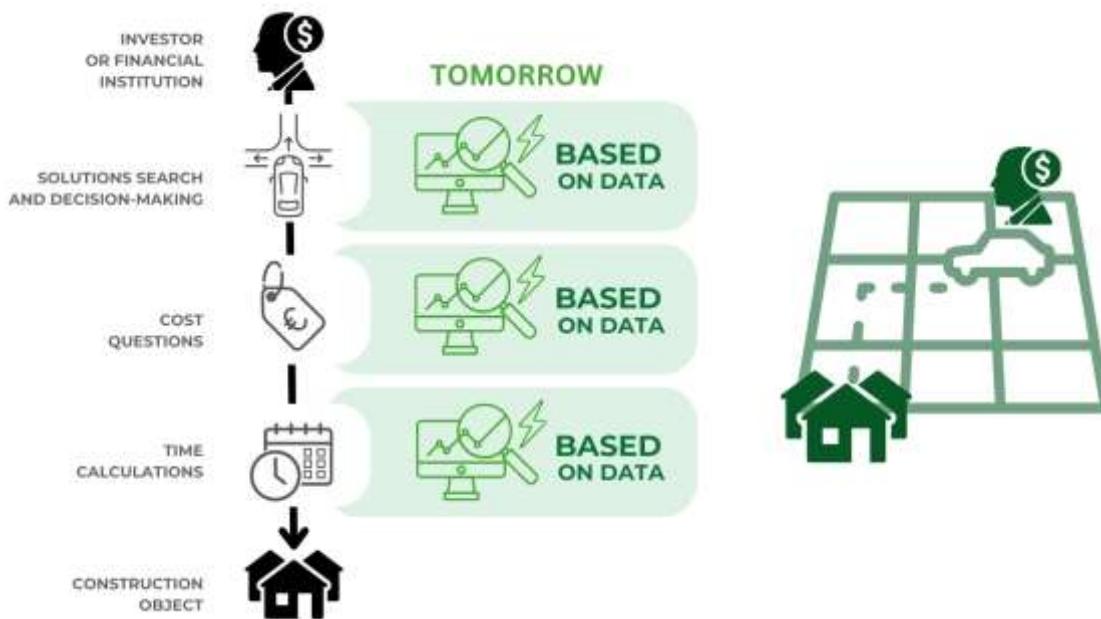
## **Nerešeni problemi uberizacije kot zadnja priložnost, da izkoristimo čas za preobrazbo**

Toda vrnimo se k realnim razmeram v gradbeništvu. Medtem ko se v nekaterih gospodarskih sektorjih pojavljajo samovozeči avtomobili, decentralizirani finančni sistemi in rešitve, ki temeljijo na umetni inteligenci, velik delež gradbenih podjetij še vedno ostaja papirnata organizacija, kjer se ključne odločitve sprejemajo bolj na podlagi intuicije in izkušenj posameznih strokovnjakov.

V tej paradigmni lahko sodobno gradbeno podjetje primerjamo z 20 let starim taksijem, ki nadzoruje vire, poti in dobavne roke ter je odgovoren za čas in stroške "potovanja" - od zamisli o projektu (logistični in namestitveni proces) do izvedbe projekta. Tako kot so GPS (v gradbeništvu IoT, RFID) in algoritmi strojnega učenja pri izračunih časa/stroškov nekoč preoblikovali področje transporta, imajo podatki, algoritmi in umetna inteligenca -agenti možnost preoblikovati upravljanje gradbeništva - od intuitivnih ocen do napovednih, vodenih modelov. V zadnjih 20 letih je v številnih panogah - v financah, kmetijstvu, maloprodaji in logistiki - postopoma izginila možnost špekuliranja zaradi nepreglednosti podatkov. Cene, stroški dostave ali finančne transakcije se izračunajo samodejno in statistično utemeljeno - v le nekaj sekundah na digitalnih platformah.

V prihodnosti se morajo gradbena podjetja zavedati, da bo demokratizacija dostopa do podatkov in orodij za njihovo analizo prekinila tradicionalni pristop k ocenjevanju stroškov in časovnih okvirov projekta ter odpravila možnost špekuliranja na podlagi nepreglednih podatkov o količini in ceni.

Podobno kot vožnja po urejeni cesti brez posredovanja voznika bodo tudi gradbeni postopki v prihodnosti vse bolj podobni sistemu "Uber" - z avtomatizirano oceno časa in stroškov, preglednim usmerjanjem nalog in minimalno odvisnostjo od človeških dejavnikov. To bo spremenilo samo naravo "potovanja" od ideje do realizacije - postal bo bolj predvidljivo, obvladljivo in podatkovno vodeno.



**Slika 10.1-5 Stroški in "potovalni" čas med gradnjo bodo določeni z uporabo strojnega učenja in statističnih orodij.**

S postopnim uvajanjem novih predpisov in zahtev v skoraj vseh državah sveta, ki zahtevajo prenos modelov CAD- (BIM)- naročnikom ali bankam, ki financirajo gradbene projekte, imata naročnik in stranka možnost, da neodvisno zagotovita preglednost izračunov stroškov in obsega. To je še posebej pomembno za velike naročnike in investorje, ki imajo dovolj kompetenc in orodij za sprotno analizo obsega in spremljanje tržnih cen. Za podjetja, ki izvajajo velike standardne projekte - trgovine, poslovne stavbe, stanovanjske komplekse -, postajajo takšne prakse standard.

Ker je informacijska vsebina modelov vse bolj popolna in standardizirana, možnost manipulacije in špekulacij praktično izginja. Digitalna transformacija postopoma spreminja pravila igre v gradbeništvu in podjetja, ki se tem spremembam ne bodo prilagodila, se lahko soočijo z resnimi izvivi.

Večja konkurenca, tehnološke motnje in zmanjševanje marž lahko vplivajo na trajnost poslovanja. Zaradi omejene likvidnosti se vse več akterjev v panogi obrača k avtomatizaciji, analitiki in podatkovnim tehnologijam kot načinu za izboljšanje učinkovitosti in preglednosti procesov. Ta orodja postajajo pomemben vir za ohranjanje konkurenčnosti v spremenljajočem se gospodarskem okolju.

Morda ne bi smeli čakati, da nas zunanje okoliščine prisilijo k nujnemu ukrepanju - veliko bolj učinkovito je, da se začnemo pripravljati že danes s krepitevijo digitalnih kompetenc, uvajanjem sodobnih rešitev in oblikovanjem kulture, osredotočene na podatke.

Ena od zadnjih ključnih tehnoloških ovir za obsežno digitalno preobrazbo gradbene industrije, ki bo v prihodnjih letih prizadela vsako podjetje, je problem samodejne klasifikacije elementov gradbenih projektov.

Brez zanesljivega, natančnega in skalabilnega razvrščanja ni mogoče ustvariti podlage za polnopravno analitiko, avtomatizacijo procesov in upravljanje življenjskega cikla z uporabo umetne inteligence in napovednih modelov. Dokler je klasifikacija objektov še vedno odvisna od ročne razlage izkušenih strokovnjakov - mojstrov, projektantov, ocenjevalcev - ima gradbena industrija še vedno priložnost. Ta čas lahko izkoristimo za pripravo na neizogibne spremembe: vse večje zahteve po preglednosti, demokratizacijo orodij in podatkov ter pojav sistemov samodejne klasifikacije, ki bodo korenito spremenili pravila igre.

Naloga samodejnega razvrščanja elementov gradbenega sveta je po svoji kompleksnosti primerljiva s prepoznavanjem predmetov v sistemih vožnje brez posadke, kar je eden glavnih izvivov. Predstavljajmo si brezpilotno vozilo, ki potuje od točke A do točke B (slika 10.1-5). Trenutni sistemi samodejne vožnje se spopadajo s problemom razvrščanja predmetov, ki jih prepoznajo lidarji in kamere. Ni dovolj, da avtomobil zgolj "vidi" oviro ali orientacijski znak, temveč mora biti sposoben prepozнатi, ali gre za pešca, prometni znak ali koš za smeti.

S podobnim temeljnim izvivom se sooča celotna gradbena industrija. Projektne elemente, kot so okna, vrata ali stebri, je mogoče zabeležiti v dokumentaciji, prikazati v modelih CAD, fotografirati na gradbišču ali prepoznati v oblakih točk laserskega skeniranja. Vendar njihova vizualna ali groba geometrijska prepoznavna ne zadostuje za izgradnjo resnično avtomatiziranega sistema za upravljanje projektov. Zagotoviti je treba, da je vsak element natančno in dosledno kategoriziran v tip, ki bo enolično prepoznaven v vseh nadaljnjih procesih - od predračunov in specifikacij do logistike, kontrole zalog in, kar je najpomembnejše, operacij (slika 4.2-6).

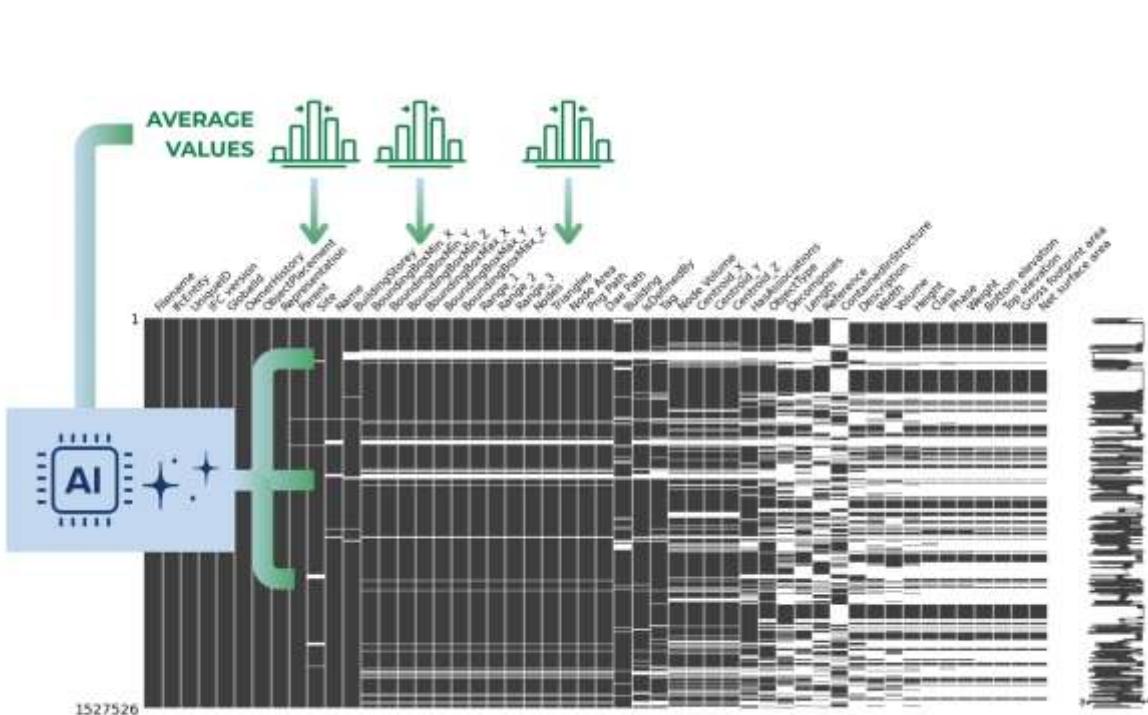
Na tej stopnji - prehod od prepoznavanja k smiselnemu klasifikaciji - se pojavi ena od ključnih ovir. Čeprav so digitalni sistemi tehnično sposobni razlikovati in prepoznavati objekte v modelih in na gradbišču, je glavna težava v pravilni in kontekstualno stabilni opredelitevi tipa elementa za različna programska okolja.. Projektant lahko na primer vrata v modelu CAD označi kot element kategorije "vrata", pri prenosu v sistem ERP ali PMIS pa so lahko napačno označena, bodisi zaradi napake projektanta bodisi zaradi nedoslednosti med sistemi. Poleg tega element med izvozom in uvozom podatkov pogosto izgubi nekatere pomembne atribute ali izgine iz sistemskega računovodstva. To povzroči vrzel v pretoku podatkov in spodbopava načelo celovite digitalizacije gradbenih procesov. Tako nastane kritična vrzel med "vidnim" in "razumljivim" semantičnim pomenom, ki spodbopava celovitost podatkov in znatno otežuje avtomatizacijo procesov v celotnem življenjskem ciklu gradbenega projekta.

Rešitev izizza univerzalne klasifikacije gradbenih elementov z uporabo tehnologij velikih podatkov in strojnega učenja (slika 10.1-6) bo katalizator za preoblikovanje celotne industrije - in morda nepričakovano odkritje za številna gradbena podjetja. Enoten, učljiv klasifikacijski sistem bo temelj za skalabilno analitiko, digitalno upravljanje in sprejetje umetne inteligence v vsakodnevno prakso gradbenih organizacij.

NVIDIA in drugi vodilni tehnološki ponudniki v drugih panogah že ponujajo rešitve, ki lahko samodejno kategorizirajo in strukturirajo velike količine besedilnih in vizualnih informacij.

Model NeMo Curator družbe NVIDIA [161] je na primer specializiran za samodejno razvrščanje in kategoriziranje podatkov v vnaprej določene kategorije, kar ima ključno vlogo pri optimizaciji cevovodov za obdelavo informacij za natančno nastavljanje in predhodno usposabljanje generativnih modelov umetne inteligenčne. Platforma Cosmos se usposablja na realnih video posnetkih in 3D prizorih [162], kar zagotavlja podlago za avtonomne sisteme in digitalne dvojčke, ki že nastajajo v ekosistemu NVIDIA. NVIDIA Omniverse, ki je do leta 2025 postal vodilno orodje za delo z obliko USD, univerzalnim opisom scene, ki bi sčasoma lahko nadomestil obliko IFC v procesih prenosa informacij o projektiranju. Skupaj z Isaac Sim, simulatorjem robotskih procesov [163], predstavljajo rešitve, kot so NeMo Curator, Cosmos in Omniverse, novo raven avtomatizacije: od čiščenja in filtriranja podatkov do ustvarjanja učnih nizov, modeliranja lastnosti objektov in usposabljanja robotov na gradbišču. Vsa ta orodja so brezplačna in odprtakodna, kar znatno zmanjšuje ovire za njihovo sprejetje v inženirski in gradbeni praksi.

Samodejno razvrščanje podatkov na ravni strukturiranih tabel ni tako zahtevna naloga, kot se morda zdi na prvi pogled. Kot smo pokazali v prejšnjem poglavju (slika 9.1-10), je mogoče manjkajoče ali napačne vrednosti razredov nadomestiti na podlagi podobnih parametrov drugih elementov, če so na voljo zbrani zgodovinski podatki. Če so bili elementi s podobnimi značilnostmi pravilno razvrščeni že v več zaključenih projektih, lahko sistem z veliko verjetnostjo predlaga ustrezno vrednost za nov ali nedokončan element (slika 10.1-6). Takšna logika, ki temelji na povprečnih vrednostih in analizi konteksta, je lahko še posebej učinkovita pri množični obdelavi tabelarnih podatkov, ki izhajajo iz ocen, specifikacij ali modelov CAD.



Slika 10.1-6 Strojno učenje vam lahko pomaga samodejno poiskati povprečne vrednosti za neizpolnjene (bela polja) parametre tabele na podlagi preteklih projektov.

Ob tako hitrem napredku na področju strojnega učenja je jasno: leta 2025 bi bilo naivno verjeti, da bo problem samodejnega razvrščanja gradbenih elementov še dolgo nerešen. Da, trenutni algoritmi še niso dosegli popolne zrelosti, zlasti pri nepopolnih ali heterogenih podatkih, vendar se okno priložnosti za prilaganje hitro zapira.

Podjetja, ki že vlagajo v zbiranje, čiščenje in urejanje podatkov ter uporabljajo orodja za avtomatizacijo ETL, bodo v prednosti. Ostali tvegajo, da bodo zaostali - tako kot podjetja nekoč niso bila kos izzivom digitalne preobrazbe v prometni in finančni industriji

Tisti, ki se še naprej zanašajo na ročno upravljanje podatkov in tradicionalne metode ocenjevanja stroškov in časa, tvegajo, da se bodo znašli v položaju taxi voznih parkov iz leta 2000, ki se do začetka leta 2020 ne bodo mogli prilagoditi dobi mobilnih aplikacij in samodejnih izračunov poti.



## POGLAVJE 10.2.

### PRAKTIČNI VODNIK ZA IZVAJANJE PRISTOPA, KI TEMELJI NA PODATKIH

#### Od teorije k praksi: načrt za digitalno preobrazbo v gradbeništvu

Gradbeništvo postopoma vstopa v novo fazo razvoja, v kateri znane procese vse bolj dopolnjujejo - in včasih celo nadomeščajo - digitalne platforme in pregledni modeli interakcije. To za podjetja ne pomeni le izzivov, temveč tudi velike priložnosti. Organizacije, ki že danes gradijo dolgoročno digitalno strategijo, bodo lahko ne le ohranile svoj položaj na trgu, temveč ga tudi razširile, saj bodo strankam ponudile sodobne pristope in zanesljive, tehnološko podprte rešitve.

Zavedati se je treba, da je poznavanje konceptov in tehnologij le izhodišče. Vodje in strokovnjaki se soočajo s praktičnim vprašanjem: kje začeti z izvajanjem in kako teoretične zamisli spremeniti v resnično vrednost. Poleg tega se vse pogosteje pojavlja vprašanje: na čem bo temeljilo poslovanje, če lahko stranka kadar koli spremeni tradicionalne metode obračunavanja stroškov in časovnega načrtovanja.

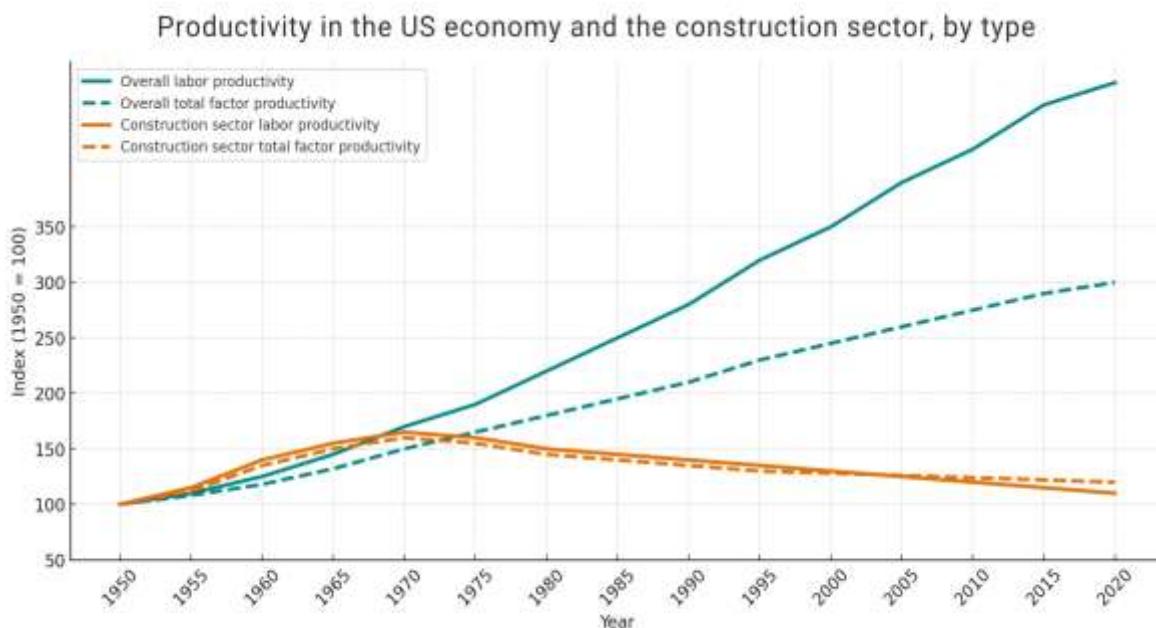
Odgovor verjetno ni toliko v tehnologiji, temveč v oblikovanju nove poklicne kulture, kjer se delo s podatki dojema kot sestavni del vsakdanje prakse. Prav pomanjkanje pozornosti digitalni tehnologiji in inovacijam je gradbeništvo cepilo v hudo zaostalost, ki smo jo opazili v zadnjih desetletjih [43].

Po podatkih družbe McKinsey znašajo izdatki za raziskave in razvoj v gradbeništvu manj kot 1% prihodkov, v avtomobilski in letalski industriji pa 3,5-4,5%. Podobno tudi stroški informacijske tehnologije v gradbeništvu ostajajo manjši od 1% skupnih prihodkov [107].

Posledično se v gradbeništvu ne zmanjšuje le stopnja avtomatizacije, temveč tudi produktivnost dela, in do leta 2020 bo gradbeni delavec proizvedel manj kot pred pol stoletja (slika 10.2-1)

Takšne težave s produktivnostjo v gradbenem sektorju so značilne za večino razvitih držav in držav v razvoju (produktivnost v gradbeništvu se je zmanjšala v 16 od 29 držav OECD (slika 2.2-1)) in ne kažejo le na pomanjkanje tehnologije, temveč tudi na potrebo po sistemskih spremembah v samem pristopu k upravljanju, usposabljanju in inovacijam.

Uspeh digitalne preobrazbe ni toliko odvisen od števila in razpoložljivosti orodij, temveč od sposobnosti organizacij, da pregledajo svoje procese in razvijejo kulturo, ki je odprta za spremembe. Ni ključna sama tehnologija, temveč ljudje in procesi, ki zagotavljajo njenu učinkovito uporabo, podpirajo stalno učenje in spodbujajo sprejemanje novih idej.



Slika 10.2-1 Paradoks produktivnosti dela in skupne produktivnosti virov v gospodarstvu ZDA in gradbenem sektorju (1950-2020) (na podlagi [43]).

V prvih delih knjige je bil model poslovnega okolja primerjan z gozdnim ekosistemom (slika 2.1-2, slika 1.2-4, slika 1.3-2). V zdravem gozdu imajo občasni požari kljub vsej svoji uničujoči moči ključno vlogo pri dolgoročni obnovi. Iz tal odstranijo staro vegetacijo, vrnejo shranjena hranila in ustvarijo prostor za novo življenje. Nekatere rastlinske vrste so se celo razvile tako, da se njihova semena odprejo šele, ko so izpostavljena visokim temperaturam ognja - naravni mehanizem, ki zagotavlja idealen čas za kalitev.

Podobno velja za podjetja, kjer lahko krize delujejo kot "nadzorovana izgorelost" in spodbujajo nastanek novih pristopov in podjetij, ki niso vezana na zastarele sisteme. Takšna obdobja prisilijo k opuščanju neučinkovitih praks, kar sprosti vire za inovacije. Tako kot se gozd po požaru začne s pionirskimi rastlinami, tako tudi podjetje po krizi oblikuje nove, prožne procese, ki postanejo osnova za zrelo informacijsko okolje.

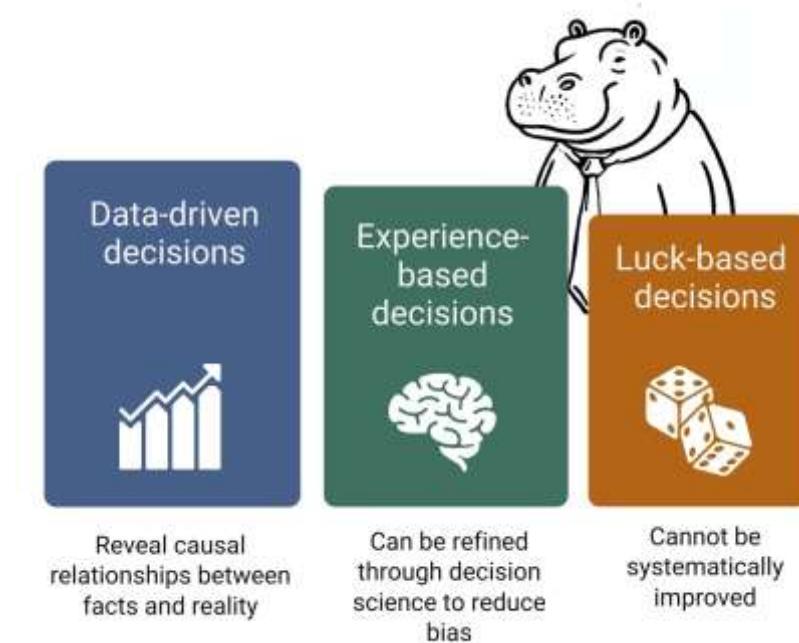
Podjetja, ki jim bo uspelo pravilno interpretirati te "ognjene signale" in preoblikovati njihovo motečo energijo v konstruktivne spremembe, bodo dosegla novo raven uspešnosti - z bolj preglednimi in prilagodljivimi podatkovnimi procesi, ki krepijo naravno sposobnost organizacije, da se obnavlja in raste.

Vse večji vpliv umetne inteligence in strojnega učenja na poslovno okolje ni več vprašljiv. Ne gre le za začasen trend, temveč za strateško nujnost. Podjetja, ki ignorirajo umetno inteligenco, tvegajo izgubo konkurenčnosti na trgu, ki vse bolj spodbuja inovativnost in prilagodljivost.

Prihodnost pripada tistim, ki na umetno inteligenco ne gledajo le kot na orodje, temveč kot na priložnost za premislek o vseh vidikih poslovanja - od optimizacije procesov do vodstvenih odločitev.

## Postavitev digitalnih temeljev: 1-5 korakov do digitalne zrelosti

V tem poglavju si ogledamo načrt digitalne preobrazbe in opredelimo ključne korake, ki so potrebni za uvedbo podatkovno usmerjenega pristopa, ki lahko pomaga preoblikovati tako kulturo podjetja kot tudi njegov informacijski ekosistem.



Slika 10.2-2 Nadzorovano posodabljanje in izbira strategije: primer, izkušnje ali podatki.

Po podatkih McKinseyjeve študije "Zakaj digitalne strategije ne uspejo" (2018) obstaja vsaj pet razlogov [164], zakaj podjetjem ne uspe doseči ciljev digitalne preobrazbe

- **Nejasne opredelitve:** Vodstveni delavci in menedžerji različno razumejo, kaj pomeni "digitalno", kar povzroča nesporazume in nedoslednosti.
- **Nerazumevanje digitalnega gospodarstva:** številna podjetja podcenjujejo obseg sprememb, ki jih digitalizacija prinaša v poslovne modele in dinamiko industrije (slika 10.1-6).
- **Zanemarjanje ekosistemov:** podjetja se osredotočajo na posamezne tehnološke rešitve (podatkovne silose) in spregledajo potrebo po vključevanju v širše digitalne ekosisteme (slika 2.2-2, slika 4.1-12).
- **Podcenjevanje digitalizacije konkurentov:** vodstveni delavci ne upoštevajo, da tudi konkurenčni dejavniki uvajajo digitalne tehnologije, kar lahko privede do izgube konkurenčne prednosti.
- **pogrešanje dvojnosti digitalizacije:** izvršni direktorji prenašajo odgovornost za digitalno preobrazbo na druge vodstvene delavce, kar birokratizira nadzor in upočasnuje proces sprememb.

Reševanje teh izzivov zahteva jasno razumevanje in usklajevanje digitalnih strategij na vseh ravneh

organizacije. Pred oblikovanjem digitalne strategije je pomembno razumeti izhodišče. Mnoge organizacije običajno sprejmejo nova orodja in platforme, ne da bi imele popolno sliko trenutnega stanja.

### **Korak 1: Izvedite revizijo trenutnih sistemov in podatkov.**

Pred spremnjanjem procesov je treba razumeti, kaj je že vzpostavljen. Z revizijo lahko ugotovite pomanjkljivosti pri upravljanju podatkov in razumete, katere vire lahko uporabite. Revizija je nekakšen "rentgenski posnetek" vaših poslovnih procesov. Z njo lahko prepozname področja tveganja in ugotovite, kateri podatki so za vaš projekt ali poslovanje ključnega pomena in kateri so drugotnega pomena.

Glavna dejanja:

- Načrt okolja IT (v programih Draw.io, Lucidchart, Miro, Visio ali Canva). Navedite uporabljenе sisteme (ERP, CAD, CAFM, CPM, SCM in druge), ki so vključeni v vaše procese in ki smo jih obravnavali v poglavju "Tehnologije in sistemi upravljanja v sodobni gradnji" (slika 1.2-4)
- Ocenite vprašanja kakovosti podatkov za vsak sistem glede pogostosti podvajanj, morebitnih manjkajočih vrednosti in nedoslednosti formata v vsakem sistemu.
- Opredelite "boleče točke" - mesta, kjer se procesi lahko prekinejo ali pogosto zahtevajo ročno posredovanje - uvoz, izvoz in dodatni procesi potrjevanja.

Če želite, da ekipa zaupa poročilom, morate že na začetku zagotoviti, da so podatki pravilni.

Revizija kakovosti podatkov vam bo pokazala, kateri podatki:

- Potreben je nadaljnji razvoj (vzpostaviti je treba samodejne postopke čiščenja ali dodatne transformacije).
- So "smeti", ki le zamašijo sisteme in se jih lahko znebite tako, da jih ne uporabljate več v procesih.

Takšno revizijo lahko opravite tudi sami. Včasih pa je koristno angažirati zunanjega svetovalca - zlasti iz drugih panog: svež pogled in neodvisnost od gradbenih "posebnosti" bosta pomagala trezno oceniti obstoječe stanje in se izogniti značilnim pastem pristranskosti do določenih rešitev in tehnologij.

### **Korak 2: Opredelitev ključnih standardov za usklajevanje podatkov.**

Po reviziji je treba oblikovati skupna pravila za delo s podatki. Kot smo obravnavali v poglavju "Standardi: Od naključnih datotek do inteligentnega podatkovnega modela", bo to pomagalo odpraviti ločene tokove podatkov.

Brez enotnega standarda bo vsaka ekipa še naprej delala "po svoje", vi pa boste ohranili "živalski vrt" integracij, kjer se podatki izgubljajo z vsako pretvorbo.

#### Glavna dejanja:

- Izberite podatkovne standarde za izmenjavo informacij med sistemi:
  - Za tabelarne podatke so to lahko strukturirani formati, kot so CSV, XLSX ali učinkovitejši formati, kot je Parquet
  - Za izmenjavo ohlapno strukturiranih podatkov in dokumentov: JSON ali XML
- obvladate delo s podatkovnimi modeli:
  - Začnite s parametriranjem nalog na ravni konceptualnega podatkovnega modela - kot je opisano v poglavju "Modeliranje podatkov: konceptualni, logični in fizični model" (slika 4.3-2).
  - Ko se poglobite v logiko poslovnega procesa, preidite na formalizacijo zahtev z uporabo parametrov v logičnih in fizičnih modelih (slika 4.3-6).
  - Določite ključne entitete, njihove attribute in odnose znotraj procesov ter vizualizirajte te odnose - tako med entitetami kot med parametri (slika 4.3-7).
- Uporabite regularne izraze (RegEx) za preverjanje in standardizacijo podatkov (slika 4.4-7), kot smo obravnavali v poglavju "Strukturirane zahteve in regularni izrazi RegEx". RegEx ni zapletena, je pa izredno pomembna tema pri delu oblikovanja zahtev na ravni fizičnih podatkovnih modelov.

Brez standardov na ravni podatkov in vizualizacije procesov ni mogoče zagotoviti doslednega in razširljivega digitalnega okolja. Zapomnите si, da so "slabi podatki dragi". Stroški napak se povečujejo, ko projekt ali organizacija postaja bolj zapletena. Poenotenje formatov, opredelitev poimenovanja, strukture in pravil potrjevanja je naložba v stabilnost in skalabilnost prihodnjih rešitev.

#### Korak 3. Izvedite DataOps in avtomatizirajte procese.

Brez dobro opredeljene arhitekture se bodo podjetja neizogibno soočala z raznovrstnimi podatki v ločenih informacijskih sistemih. Podatki bodo neintegrirani, podvojeni na več lokacijah, njihovo vzdrževanje pa bo drago.

Predstavljajte si, da so podatki voda, podatkovna arhitektura pa zapleten sistem cevovodov, ki to vodo prenaša od vira shranjevanja do mesta uporabe. Podatkovna arhitektura določa, kako se informacije zbirajo, shranjujejo, preoblikujejo, analizirajo in dostavljajo končnim uporabnikom ali aplikacijam.

DataOps (Data Operations) je metodologija, ki združuje zbiranje, čiščenje, potrjevanje in uporabo podatkov v enoten avtomatiziran procesni tok, kot smo podrobno obravnavali v 8. delu knjige.

### Glavna dejanja:

- Ustvarjanje in prilagajanje ETL -prevozniki za avtomatizacijo procesov:
  - Izvleček: organizirajte samodejno zbiranje podatkov iz dokumentov PDF (Slika 4.1-2, Slika 4.1-5, Slika 4.1-7), preglednic Excel, modelov CAD (Slika 7.2-4), sistemov ERP in drugih virov, s katerimi delate.
  - Preoblikovanje: nastavite samodejne procese za preoblikovanje podatkov v enotno strukturirano obliko in avtomatiziranje izračunov, ki se bodo izvajali zunaj zaprtih aplikacij (slika 7.2-8).
  - Nalaganje: poskusite ustvariti samodejno nalaganje podatkov v zbirne tabele, dokumente ali centralizirane shrambe (Slika 7.2-9, Slika 7.2-13, Slika 7.2-16).
- Avtomatizirajte postopke izračuna in QTO (Quantity Take-Off), kot smo obravnavali v poglavju "QTO Quantity Take-Off: grupiranje podatkov o projektu po atributih":
  - Nastavite samodejno pridobivanje volumnov iz modelov CAD z uporabo vmesnikov API, vtičnikov ali orodij za obratno inženirstvo (slika 5.2-5).
  - Ustvarite pravila za združevanje elementov za različne razrede po atributih v obliki tabel (slika 5.2-12).
  - Poskusite avtomatizirati pogosto ponavljajoče se izračune obsega in stroškov zunaj modularnih zaprtih sistemov (slika 5.2-15).
- Začnite uporabljati Python in Pandas za obdelavo podatkov, kot smo obravnavali v poglavju "Python Pandas: nepogrešljivo orodje za delo s podatki":
  - Uporabite DataFrame za delo z datotekami XLSX in avtomatizirajte obdelavo tabelaričnih podatkov (slika 3.4-6).
  - Avtomatizacija združevanja in preoblikovanja informacij z različnimi knjižnicami Python
  - Za poenostavitev pisanja pripravljenih blokov kode in celotnih cevovodov uporabite LLM (slika 7.2-18).
  - Poskusite zgraditi Pipeline v Pythonu, ki najde napake ali opazi nepravilnosti in pošlje obvestilo odgovorni osebi (npr. vodji projekta) (slika 7.4-2).

Avtomatizacija, ki temelji na načelih DataOps, vam omogoča prehod od ročnega in razdrobljenega ravnjanja s podatki k trajnostnim in ponovljivim postopkom. S tem se ne le zmanjša obremenitev zaposlenih, ki se vsak dan ukvarjajo z enakimi pretvorbami, temveč se tudi bistveno povečajo zanesljivost, razširljivost in preglednost celotnega informacijskega sistema.

### Korak 4: Ustvarite ekosistem za upravljanje odprtih podatkov.

Kljud razvoju zaprtih modularnih sistemov in njihovi integraciji z novimi orodji se podjetja soočajo z resno težavo - vse večja kompleksnost teh sistemov presega njihovo uporabnost. Prvotna zamisel o oblikovanju enotne lastniške platforme, ki bi pokrivala vse poslovne procese, je privredla do pretirane

centralizacije, pri kateri so za vsako spremembo potrebna velika sredstva in čas za prilagoditev.

Kot smo obravnavali v poglavju "Korporativni micelij: kako podatki povezujejo poslovne procese", je za učinkovito upravljanje podatkov potreben odprt in enoten ekosistem, ki povezuje vse vire informacij.

Ključni elementi ekosistema:

■ Izberite ustrezeno podatkovno shrambo:

- Za tabele in izračune uporabite podatkovne zbirke, na primer PostgreSQL ali MySQL (slika 3.1-7).
- Za dokumente in poročila je primerna shramba v oblaku (Google Drive, OneDrive) ali sistemi, ki podpirajo obliko JSON.
- Preverite zmogljivosti podatkovnih skladišč, podatkovnih jezer in drugih orodij za centralizirano shranjevanje in analizo velikih količin informacij (slika 8.1-8).

■ izvajanje rešitev za dostop do lastniških podatkov:

- Če uporabljate lastniške sisteme, konfigurirajte dostop do njih prek vmesnika API ali kompleta SDK za pridobivanje podatkov za zunanjou obdelavo (slika 4.1-2).
- Seznanite se z možnostmi orodij za povratni inženiring za formate CAD (slika 4.1-13).
- Vzpostavite cevovode ETL, ki periodično zbirajo podatke iz aplikacij ali strežnikov, jih pretvarjajo v odprte strukturirane formate in shranjujejo v skladišča (slika 7.2-3).
- v skupini razpravljajte o tem, kako zagotoviti dostop do podatkov, ne da bi za to potrebovali zaščiteno programsko opremo.
- Ne pozabite: podatki so pomembnejši od vmesnikov. Dolgoročno vrednost zagotavljata struktura in razpoložljivost informacij ter ne posebna orodja uporabniških vmesnikov.

■ Razmislite o ustanovitvi centra odličnosti (CoE) za podatke, kot smo obravnavali v poglavju "Center odličnosti (CoE) za modeliranje podatkov", ali o tem, kako lahko zagotovite strokovno znanje o podatkih na druge načine (slika 4.3-9).

Ekosistem za upravljanje podatkov ustvarja enoten informacijski prostor, v katerem vsi udeleženci projekta delajo z doslednimi, posodobljenimi in preverjenimi informacijami. Je osnova za razširljive, prilagodljive in zanesljive digitalne procese

## Izkoriščanje potenciala podatkov: 5-10 korakov do digitalne zrelosti

Poleg tehnične integracije je pomemben dejavnik uspešnega izvajanja digitalnih rešitev tudi to, da jih sprejmejo končni uporabniki. Vključevanje strank ali uporabnikov v merjenje uspešnosti je hkrati izliv za izboljšanje uporabniške izkušnje in upravljanje sprememb v podjetju. Če se rešitev ne prilega znanemu delovnemu procesu ali ne rešuje resničnih težav uporabnikov ali strank, se ne bo uporabljala in tega ne bodo popravili nobeni dodatni ukrepi in spodbude.

Preoblikovanje je ponavljajoč se proces, ki temelji na analizi podatkov o interakciji uporabnikov z

novimi procesi, s pogostimi cikli testiranja, stalnimi povratnimi informacijami in izboljšavami.

### **Korak 5: Vzpostavite kulturo podatkov, usposabljamte osebje in zbirajte povratne informacije**

Še tako napreden sistem ne bo deloval brez zavzetosti zaposlenih. Ustvariti morate okolje, v katerem se podatki uporabljajo vsak dan in v katerem ekipa razume njihovo vrednost.

V objavljenem poročilu vlade Združenega kraljestva za leto 2024 "Data Analytics and AI in Government Project Delivery" [83] je zapisano, da je za uspešno izvajanje podatkovne analitike in AI ključnega pomena usposabljanje strokovnjakov s potrebnimi kompetencami na področju obdelave in interpretacije podatkov.

Pomanjkanje strokovnega znanja in izkušenj na področju podatkovne analitike je ena od ključnih težav, ki omejujejo digitalno preobrazbo. Vodstveni delavci so navajeni na ustaljene rutine: četrтletne cikle, prednostne pobude in tradicionalne načine napredovanja projektov. Spremembe zahtevajo posebnega vodja - dovolj visokega, da ima vpliv, vendar ne tako visokega, da bi imel dovolj časa in motivacije za vodenje dolgoročnega projekta preobrazbe.

Glavna dejanja:

- spoznanje, da je treba preiti od subjektivnih odločitev, ki temeljijo na mnenju visoko plačanega zaposlenega (HiPPO), h kulti odločanja, ki temelji na dejstvih in podatkih, kot je obravnavano v poglavju "HiPPO ali nevarnost mnenja pri odločanju" (slika 2.1-9).
- Organizirajte sistematično usposabljanje:
  - organizirajte usposabljanja o uporabi strukturiranih podatkov in povabite strokovnjake iz drugih panog, ki niso pristranski do izdelkov in konceptov, ki so danes priljubljeni v gradbeništvu.
  - s kolegi razpravljati o pristopih in orodjih za analizo podatkov ter se samostojno učiti praktičnega dela z orodji, kot so Python, pandas in LLM (Slika 4.1-3, Slika 4.1-6).
  - Ustvarite knjižnico učnih gradiv (po možnosti s kratkimi videoposnetki) na temo strukturiranja podatkov (slika 3.2-15) in ustvarjanja podatkovnih modelov (slika 4.3-6, slika 4.3-7).
- uporaba sodobnih učnih tehnologij:
  - uporaba jezikovnih modelov (LLM) za podporo manipulaciji s kodo in podatki, vključno z generiranjem, preoblikovanjem in analizo kode ter obdelavo in interpretacijo tabelaričnih informacij (slika 3.4-1).
  - Raziščite, kako je mogoče kodo, ustvarjeno v programu LLM, prilagoditi in vključiti v celovito rešitev Pipeline pri delu v okolju za razvoj brez povezave (IDE) (slika 4.4-14, slika 5.2-13).

Če vodja še naprej sprejema odločitve po starem, nobeno usposabljanje ne bo prepričalo ljudi, da analitiko jemljejo resno.

Vzpostavitev podatkovne kulture ni mogoča brez stalnih povratnih informacij. S povratnimi informacijami lahko ugotovite vrzeli v procesih, orodjih in strategijah, ki jih ni mogoče odkriti z internimi poročili ali formalnimi KPI metrikami. Pohvalne pripombe uporabnikov vaših rešitev ne bodo imele praktične vrednosti. Dragocene so kritične povratne informacije, zlasti če temeljijo na konkretnih opažanjih in dejstvih. Vendar pridobivanje takšnih informacij zahteva napor: vzpostaviti morate procese, v katerih lahko udeleženci - tako notranji kot zunanji - delijo pripombe (morda je smiselno, da to počnejo anonimno) brez izkrivljanja in brez strahu, da bi njihova mnenja vplivala na njihovo lastno delo. Pomembno je, da to storijo brez izkrivljanja in brez strahu pred negativnimi posledicami zase.

*Vsako učenje je na koncu samoučenje [165].*

- Milton Friedman, ameriški ekonomist in statistik

Izvajanje analitičnih orodij mora spremljati redno preverjanje njihove učinkovitosti v praksi (ROI, KPI), kar je mogoče doseči le s strukturiranimi povratnimi informacijami zaposlenih, strank in partnerjev. Tako se podjetja ne le izognejo ponavljanju napak, temveč se tudi hitreje prilagodijo spremembam v okolju. Obstoj mehanizma za zbiranje in analiziranje povratnih informacij je eden od znakov zrelosti organizacije, ki prehaja od občasnih digitalnih pobud k trajnostnemu modelu stalnih izboljšav (slika 2.2-5).

## Korak 6: Od pilotnih projektov do razširitve

*Izberite dovolj velike bitke, da bodo pomembne, in dovolj majhne, da jih boste dobili.*

- Jonathan Kozol

Začetek digitalne preobrazbe "naenkrat in povsod" je zelo tvegan. Učinkovitejši pristop je začeti s pilotnimi projekti in postopoma širiti uspešne prakse.

Glavna dejanja:

■ Izberite pravi projekt za pilotni projekt:

- Opredelitev posebnega poslovnega cilja ali procesa z merljivimi rezultati (KPI, ROI) (slika 7.1-5).
- Izberite postopek avtomatizacije ETL, na primer samodejno preverjanje podatkov ali izračun delovne obremenitve (QTO) z uporabo programov Python in Pandas (slika 5.2-10)
- Določite jasne metrike za uspeh (npr. skrajšajte čas za pripravo specifikacij za

pregled ali poročil o potrditvi podatkov s tedna na dan).

■ Uporabite iterativne pristope:

- Začnite s preprostimi procesi pretvorbe podatkov in ustvarite pretočne pretvorbe podatkov več formatov v formate, ki jih potrebujete za svoje procese (Slika 4.1-2, Slika 4.1-5).
- Postopoma povečajte zahtevnost nalog in razširite avtomatizacijo procesov z oblikovanjem celotne cevovodne linije v IDE na podlagi dokumentiranih blokov kode (slika 4.1-7, slika 7.2-18).
- dokumentirajte in snemajte (po možnosti s kratkimi videoposnetki) uspešne rešitve ter jih delite s sodelavci ali v strokovnih skupnostih.

■ pripravite predloge in spremljajočo dokumentacijo za ponovitev takšnih rešitev, da jih bodo lahko učinkovito uporabljali vaši sodelavci (ali člani strokovne skupnosti, vključno z uporabniki družbenih medijev).

S postopnim "roll-upom" lahko ohranite visoko kakovost sprememb in ne zapadete v kaos vzporednega izvajanja. Strategija "od majhnega do velikega" zmanjšuje tveganja in omogoča učenje na majhnih napakah, ne da bi te prerasle v kritične težave.

Prehod od projektnega pristopa, pri katerem zaposleni sodelujejo le delno, k oblikovanju stalnih skupin (npr. strokovnih centrov) pomaga zagotoviti trajnostni razvoj izdelka tudi po izidu prve različice. Takšne ekipe ne le podpirajo obstoječe rešitve, temveč jih tudi še naprej izboljšujejo.

To zmanjšuje odvisnost od dolgotrajnih odobritev: člani ekipe so pooblaščeni za sprejemanje odločitev v okviru svojih pristojnosti. Tako se vodje osvobodijo potrebe po mikroupravljanju, ekipe pa se lahko osredotočijo na ustvarjanje prave vrednosti.

Razvoj novih rešitev ni sprint, temveč maraton. Uspešni so tisti, ki si na začetku prizadevajo za dolgoročno in dosledno delo.

Zavedati se je treba, da tehnologija zahteva stalen razvoj. Vlaganje v dolgoročni razvoj tehnoloških rešitev je osnova za uspešno poslovanje.

### Korak 7: Uporaba odprtih podatkovnih formatov in rešitev

Kot smo obravnavali v poglavjih o modularnih platformah (ERP, PMIS, CAFM, CDE itd.), se je treba osredotočiti na odprte in univerzalne podatkovne formate, ki zagotavljajo neodvisnost od rešitev prodajalcev in povečujejo razpoložljivost informacij za vse zainteresirane strani.

Glavna dejanja:

- Prehod z zaprtih formatov na odprte formate:
- Namesto lastniških formatov uporablajte odprte formate ali poiščite način za samodejno nalaganje ali pretvorbo zaprtih formatov v odprte formate (slika 3.2-15).

- Implementacija orodij za delo s paketi Parquet, CSV, JSON, XLSX, ki so standardi za izmenjavo med večino sodobnih sistemov (slika 8.1-2).
- Če ima delo s 3D geometrijo pomembno vlogo v vaših procesih, razmislite o uporabi odprtih formatov, kot so USD, glTF, DAE ali OBJ (slika 3.1-14).

■ Uporaba vektorskih podatkovnih zbirk podatkovnih zbirk za učinkovito analizo in iskanje informacij:

- Za poenostavitev 3D -geometrije uporabljajte metodo Bounding Box in druge metode (slika 8.2-1).
- Razmislite, kje lahko izvedete vektorizacijo podatkov - pretvorbo besedil, predmetov ali dokumentov v številčne predstavitve (slika 8.2-2).

■ Uporaba orodij za analizo velikih količin podatkov:

- Organizirajte shranjevanje zbranih zgodovinskih podatkov (npr. PDF, XLSX, CAD) v ustreznih formatih za analizo (Apache Parquet, CSV, ORC) (slika 8.1-2).
- Začnite uporabljati osnovne statistične metode in delati z reprezentativnimi vzorci - ali se vsaj seznanite s temeljnimi načeli statistike (slika 9.2-5).
- uvajanje in učenje orodij za vizualizacijo podatkov in odnosov med podatki za vizualizacijo rezultatov analize. Brez dobre vizualizacije je nemogoče v celoti razumeti same podatke ali procese, ki temeljijo na njih (slika 7.1-4).

Prehod na odprte oblike podatkov ter uvedba orodij za analizo, shranjevanje in vizualizacijo informacij je temelj za trajnostno in neodvisno digitalno upravljanje. To ne le zmanjšuje odvisnosti od prodajalcev, temveč tudi zagotavlja enakopraven dostop do podatkov za vse zainteresirane strani.

### Korak 8. Začnite izvajati strojno učenje za napovedovanje

Številna podjetja so zbrala velike količine podatkov - nekakšne "informacijske gejzirje", ki so še vedno neizkoriščeni. Ti podatki so bili zbrani v več sto ali tisoč projektih, vendar so bili v nadaljnjih procesih pogosto uporabljeni le enkrat ali pa sploh ne. Dokumenti in modeli, shranjeni v zaprtih formatih in sistemih, se pogosto dojemajo kot zastareli in neuporabni. V resnici pa so dragocen vir - podlaga za analizo storjenih napak, avtomatizacijo rutinskih postopkov in razvoj inovativnih rešitev za samodejno razvrščanje in prepoznavanje značilnosti v prihodnjih projektih.

Ključni izliv je naučiti se, kako pridobiti te podatke in jih pretvoriti v uporabne vpoglede. Kot je obravnavano v poglavju o strojnem učenju in napovedih, lahko tehnike strojnega učenja bistveno izboljšajo natančnost ocen in napovedi v različnih procesih, povezanih z gradbeništvom. Popolna uporaba zbranih podatkov odpira pot k izboljšanju učinkovitosti, zmanjšanju tveganja in izgradnji trajnostnih digitalnih procesov.

Glavna dejanja:

- Začnite s preprostimi algoritmami:
- Poskusite uporabiti linearno regresijo - z uporabo namigov iz programa LLM - za

napovedovanje ponavljajoče se uspešnosti v naborih podatkov, kjer odvisnosti od velikega števila dejavnikov ni ali so minimalne (slika 9.3-4).

- Razmislite, v katerih fazah vaših procesov bi lahko teoretično uporabili algoritem k-najbližjih sosedov (k-NN) - na primer za naloge razvrščanja, ocenjevanje podobnosti predmetov ali napovedovanje na podlagi zgodovinskih analogij (slika 9.3-5).

■ **Zbiranje in strukturiranje podatkov za usposabljanje modelov:**

- Zbiranje preteklih podatkov o projektu na enem mestu in v enotni obliki (slika 9.1-10).
- delo na kakovosti in reprezentativnosti učnih vzorcev s pomočjo samodejnega ETL (slika 9.2-8).
- Naučite se podatke razdeliti na učne in testne množice, kot smo to storili v primeru podatkovne množice Titanik (Slika 9.2-6, Slika 9.2-7).
- razmislite o razširitvi uporabe tehnik strojnega učenja na številne naloge, od napovedovanja časovnih okvirov projektov do optimizacije logistike, upravljanja virov in zgodnjega prepoznavanja morebitnih težav.

Strojno učenje je orodje, s katerim lahko arhivirane podatke spremenite v dragoceno sredstvo za napovedovanje, optimizacijo in informirano odločanje. Začnite z majhnimi nabori podatkov (slika 9.2-5) in preprostimi modeli ter postopoma povečujte njihovo kompleksnost.

### **Korak 9. Integracija interneta stvari in sodobnih tehnologij za zbiranje podatkov**

Svet gradbeništva hitro postaja digitalen: vsaka gradbena fotografija, vsaka objava na Teamu je že del večjega procesa parametriziranja in označevanja resničnosti. Tako kot je GPS nekoč spremenil logistiko, tudi internet stvari, RFID in tehnologije za samodejno zbiranje podatkov spreminjajo gradbeno industrijo. Kot je obravnavano v poglavju "Internet stvari in pametne pogodbe", je digitalno gradbišče s senzorji in samodejnim nadzorom prihodnost panoge.

Glavna dejanja:

- Izvedba interneta stvari - naprav, RFID - oznak in podrobno opisovanje procesov, povezanih z njimi:
  - Ocenite, na katerih področjih ali fazah projekta je mogoče doseči največjo donosnost naložbe (ROI) - na primer za spremljanje temperature, vibracij, vlage ali gibanja.
  - Razmislite o uporabi RFID za sledenje materialov, orodij in opreme v celotni dobavni verigi.
  - Razmislite, kako lahko zbrane podatke vključite v enoten informacijski sistem, kot je Apache NiFi, za samodejno obdelavo in analizo v realnem času (slika 7.4-5).
- Vzpostavite sistem za spremljanje v realnem času:
  - Razvijanje nadzornih plošč za spremljanje ključnih procesnih ali projektnih metrik z uporabo orodij za vizualizacijo, kot so Streamlit, Flask ali Power BI)
  - Nastavite samodejna obvestila za opozarjanje na kritična odstopanja od načrta ali

normativov (slika 7.4-2).

- Ocenite možnosti za napovedno vzdrževanje opreme na podlagi zbranih podatkov in ugotovljenih vzorcev (slika 9.3-6).

■ Združite podatke iz različnih virov:

- Začnite z vizualizacijo podatkovnega modela na fizični ravni - prikažite strukturo informacijskih tokov in ključnih parametrov iz sistemov CAD, naprav IoT in platform ERP (slika 4.3-1).
- Začnite z oblikovanjem grobe zasnove enotne platforme za analizo podatkov in podporo upravljalnim odločitvam. Zabeležite ključne funkcije, vire podatkov, uporabnike in predvidene scenarije uporabe (slika 4.3-7).

Prej ko začnete dejanske procese povezovati z digitalnim svetom, prej jih boste lahko upravljali s podatki - učinkovito, pregledno in v realnem času.

#### Korak 10. Pripravite se na prihodnje spremembe v panogi

Gradbena podjetja so nenehno pod pritiskom zunanjega okolja: gospodarske krize, tehnološki preskok, zakonodajne spremembe. Podobno kot gozd, ki se mora upirati dežju, snegu, suši in žgočemu soncu, tudi podjetja živijo v pogojih nenehnega prilagajanja. In tako kot drevesa z globokim koreninskim sistemom postanejo odporna na mraz in sušo, ostanejo uspešne in konkurenčne le tiste organizacije, ki imajo trdne temelje avtomatiziranih procesov, sposobnost predvidevanja sprememb in prožno prilagajanje strategij.

Kot je navedeno v poglavju "Strategije preživetja: ustvarjanje konkurenčnih prednosti", gradbena industrija vstopa v fazo radikalnega preoblikovanja. Interakcija med naročnikom in izvajalcem se premika k modelu prepričevanja, kjer preglednost, predvidljivost in digitalna orodja nadomeščajo tradicionalne pristope. V tej novi realnosti ne zmagujejo največji, temveč najbolj prilagodljivi in tehnološko zreli.

Glavna dejanja:

■ analizirajte poslovne ranljivosti v kontekstu odprtih podatkov:

- Ocenite, kako bi lahko demokratizacija dostopa do podatkov kot del uberizacije uničajoče vplivala na vašo konkurenčno prednost in posovanje (slika 10.1-5).
- Razmislite o strategiji za prehod od nepreglednih in ločenih procesov k poslovnim modelom, ki temelji na odprtih rešitvah, interoperabilnosti sistemov in preglednosti podatkov (slika 2.2-5).

■ Razvijte dolgoročno digitalno strategijo:

- Določite, ali želite biti vodilni na področju inovacij ali raje izberete scenarij "dohitevanja", pri katerem boste varčevali z viri.
- Opišite faze: kratkoročna (avtomatizacija procesov, strukturiranje podatkov), srednjeročna (izvajanje LLM in ETL), dolgoročna (digitalni ekosistemi, centralizirani

repozitoriji).

■ Razmislite o razširitvi svojega portfelja storitev:

- Razmislite o ponudbi novih storitev (osredotočenih na energetsko učinkovitost, ESG, podatkovne storitve). O novih poslovnih modelih bomo več govorili v naslednjem poglavju
- Prizadevajte si, da se predstavite kot zanesljiv tehnološki partner, ki podpira celoten življenjski cikel objekta, od načrtovanja do delovanja. Zaupanje v vas mora temeljiti na sistematičnem pristopu, preglednih postopkih in sposobnosti zagotavljanja trajnostnih tehnoloških rešitev.

V transformacijskem okolju ne zmagujejo tisti, ki se zgolj odzivajo na spremembe, temveč tisti, ki so proaktivni. Prilagodljivost, odprtost in digitalna zrelost so temelji trajnosti pri gradnji jutrišnjega dne.

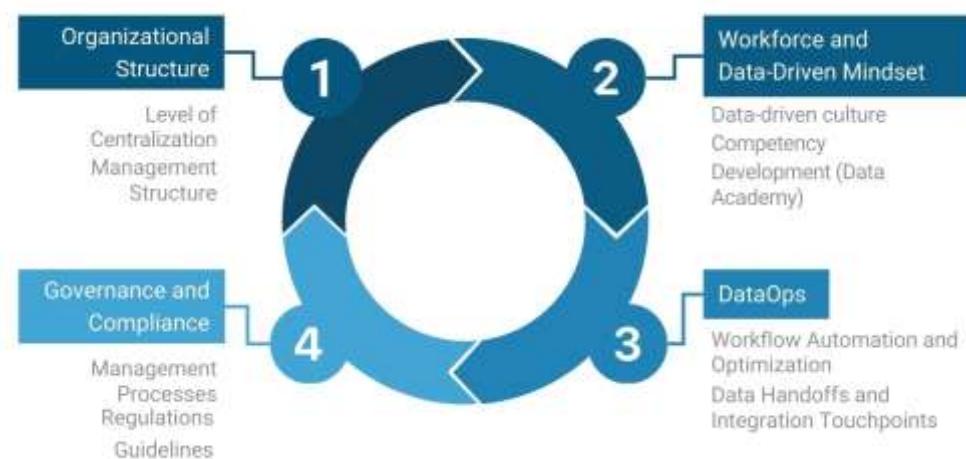
## Načrt preobrazbe: od kaosa do podatkovno usmerjenega podjetja

Naslednji načrt vam lahko služi kot začetno merilo - izhodišče za oblikovanje lastne strategije digitalne preobrazbe, ki temelji na podatkih:

- **Revizija in standardi:** analiza trenutnega stanja, poenotenje podatkov
- **strukturiranje in razvrščanje podatkov:** avtomatizirano preoblikovanje nestrukturiranih in ohlapno strukturiranih podatkov
- **Avtomatizacija združevanj,** izračunov in kalkulacij: uporaba odprtakodnih orodij in knjižnic za avtomatizacijo
- **Ekosistem in COE:** vzpostavitev notranje ekipe za oblikovanje enotnega podatkovnega ekosistema v podjetju.
- **Kultura in učenje:** prehod od rešitev HiPPO k rešitvam, ki temeljijo na podatkih
- **Pilotni projekti, povratne informacije in razširjanje:** delujte iterativno: preizkušajte nove metode v omejenem obsegu, zbirajte veljavne povratne informacije in postopoma razširjajte rešitve.
- **Odprti formati:** uporaba univerzalnih in odprtih formatov za neodvisnost od prodajalcev programske opreme
- **Strojno učenje:** vgradnja algoritmov ML v procese za napovedovanje in optimizacijo
- **IoT in digitalno gradbišče:** vključevanje sodobnih tehnologij zbiranja podatkov v procese
- **Strateško prilaganje:** priprava na prihodnje spremembe v industriji

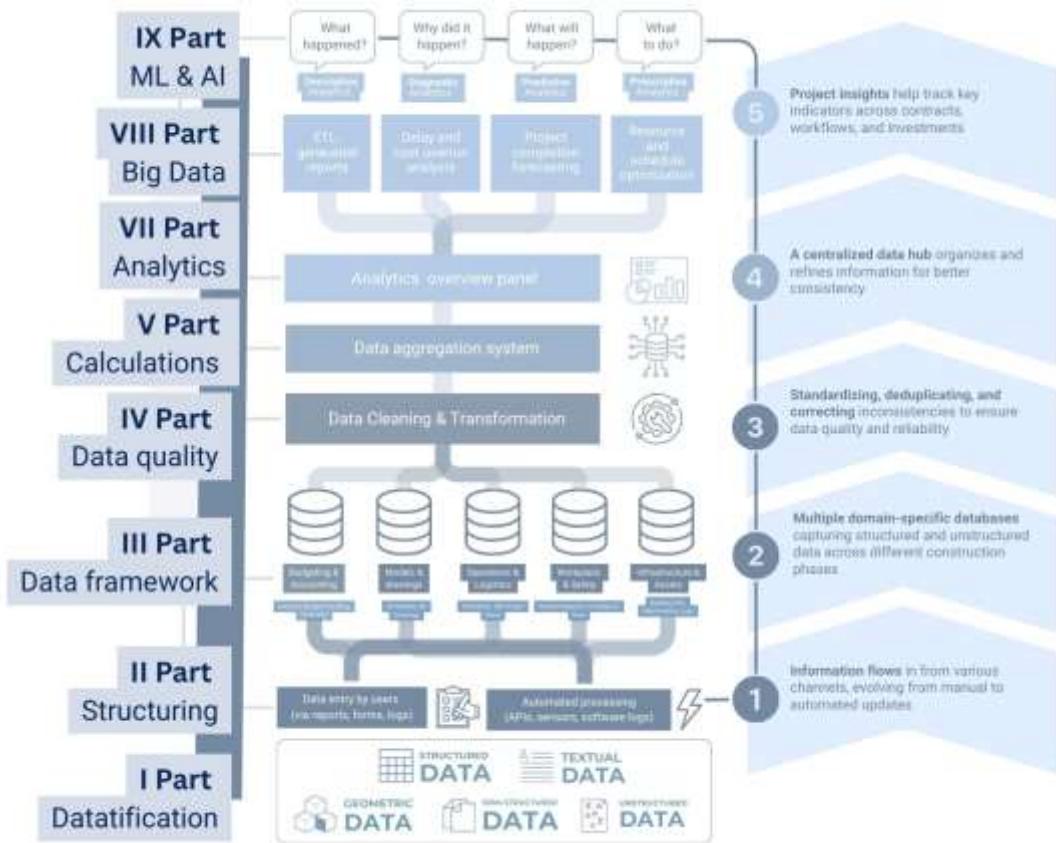
Predvsem pa ne pozabite, da "podatki sami ne spremenijo podjetja: podjetje spremenijo ljudje, ki vedo, kako delati s temi podatki". Poudarjajte kulturo, pregledne procese in zavezanost k nenehnim izboljšavam

Sistemski pristop omogoča prehod od ločenih digitalnih pobud k popolnemu modelu upravljanja, ki temelji na podatkih, kjer odločitve ne temeljijo na intuiciji ali predpostavkah, temveč na podatkih, dejstvih in matematično izračunanih verjetnostih. Pri digitalni preobrazbi gradbene industrije ne gre le za sprejemanje tehnologije, temveč za oblikovanje poslovnega ekosistema, v katerem se informacije o projektu nemoteno in iterativno prenašajo med različnimi sistemmi. Pri tem algoritmi strojnega učenja zagotavljajo samodejno, neprekinjeno analizo, napovedovanje in optimizacijo procesov. V takšnem okolju postanejo špekulacije in skriti podatki nepomembni - ostanejo le preverjeni modeli, pregledni izračuni in predvidljivi rezultati.



**Slika 10.2-3 Ključni elementi uspešnega upravljanja podatkov na ravni podjetja.**

Vsak del knjige ustreza določeni fazi obdelave in analize podatkov v gradbenih projektih (slika 2.2-5). Če se želite vrniti k eni od prej obravnavanih tem in si jo ogledati s celostnega razumevanja poteka uporabe podatkov, se lahko sklicujete na naslove delov na sliki 10.2-4.



Slika 10.2-4 Deli knjige v okviru cevovoda za obdelavo podatkov (slika 2.2-5): od digitalizacije informacij do analitike in umetne inteligence.

Ne glede na velikost organizacije, stopnjo tehnološke zrelosti ali proračun lahko že danes začnete uporabljati pristop, ki temelji na podatkih. Tudi majhni koraki v pravo smer bodo sčasoma prinesli rezultate.

Preobrazba na podlagi podatkov ni enkraten projekt, temveč neprekinjen, ponavljajoč se proces izboljšav, ki vključuje uvajanje novih orodij, preoblikovanje procesov in razvijanje kulture odločanja na podlagi podatkov.

## Gradnja v industriji 5.0: Kako zaslužiti, ko se ne morete več skrivati

Gradbena podjetja že dolgo zaslužijo z nepreglednostjo postopkov. Glavni poslovni model so bile špekulacije - precenjevanje stroškov materiala, obsega del in odstotnih pribitkov v zaprtih sistemih ERP - in PMIS -, ki so nedostopni za zunanje revizije. Omejen dostop strank in njihovih zaupnikov do izvirnih podatkov o projektu je ustvaril podlago za sheme, v katerih je bilo skoraj nemogoče preveriti zanesljivost izračunov.

Vendar ta model hitro izgublja pomen. Z demokratizacijo dostopa do podatkov, nastankom univerzitetnih diplomantov, prihodom odprtih podatkov in orodij za avtomatizacijo ETL se industrija premika k novim standardom dela.

Zato nepreglednost ni več konkurenčna prednost - kmalu bo postala breme, ki se ga bo težko znebiti. Preglednost se iz možnosti spremeni v pogoj za obstanek na trgu.

S kom bodo stranke - banke, vlagatelji, fizične stranke, stranke zasebnega kapitala, vladne stranke - sodelovale v novi digitalni resničnosti? Odgovor je očiten: s tistimi, ki so sposobni zagotoviti ne le rezultat, temveč tudi utemeljitev vsakega koraka na poti do njega. Ker se količina odprtih podatkov povečuje, bodo partnerji in stranke izbirali podjetja, ki zagotavljajo preglednost, natančnost in predvidljivost rezultatov.

Glede na to se pojavljajo novi poslovni modeli, ki temeljijo na upravljanju podatkov in zaupanju, ne pa na špekulacijah:

- **Prodaja procesov namesto kvadratnih metrov:** ključna prednost postaneta zaupanje in učinkovitost, ne pa konkretni sporazumi s popustom. Glavna vrednost bo predvidljivost rezultata na podlagi zanesljivih in preverjenih podatkov. Sodobna podjetja ne bodo prodajala gradbenega objekta kot takega, temveč:
  - natančni roki in pregledni urniki dela;
  - razumne ocene, podprte z izračuni;
  - popolno digitalno sledljivost in nadzor v vseh fazah projekta.
- **Inženiring in analitika kot storitev:** model "Data-as-a-Service" (način zagotavljanja pripravljenih podatkov uporabnikom prek interneta kot storitev), kjer vsak projekt postane del digitalne podatkovne verige, poslovna vrednost pa je v sposobnosti upravljanja te verige. Podjetja se spreminja v intelligentne platforme, ki ponujajo rešitve, temelječe na avtomatizaciji in analitiki:
  - Avtomatizirana in pregledna priprava ocen in načrtov;
  - ocena tveganja in časovnega okvira na podlagi algoritmov strojnega učenja;
  - izračun okoljskih kazalnikov (ESG, CO<sub>2</sub>, energetska učinkovitost);
  - izdelavo poročil iz revidiranih odprtih virov.
- **Proizvodnja inženirskega strokovnega znanja:** razvoj podjetja se lahko večkrat uporabi v

podjetju in distribuira kot ločen izdelek, kar predstavlja dodaten vir dohodka z digitalnimi storitvami. V novem okolju podjetja ne ustvarjajo le projektov, temveč tudi digitalna sredstva:

- knjižnice komponent in predloge za ocenjevanje;
- avtomatizirani moduli za preverjanje;
- Odprtokodni vtičniki in skripte (prodaja svetovanja) za delo s podatki.

■ **Nova vrsta podjetja: integrator, ki temelji na podatkih:** udeleženec na trgu, ki ni odvisen od določenih ponudnikov programske opreme ali modularnih sistemov in ni "zaklenjen" v en sam programski vmesnik. S podatki posluje svobodno - in na njih gradi svojo konkurenčnost. Gradbeno podjetje prihodnosti ni le izvajalec, temveč informacijski integrator, ki lahko za naročnika opravlja naslednje funkcije:

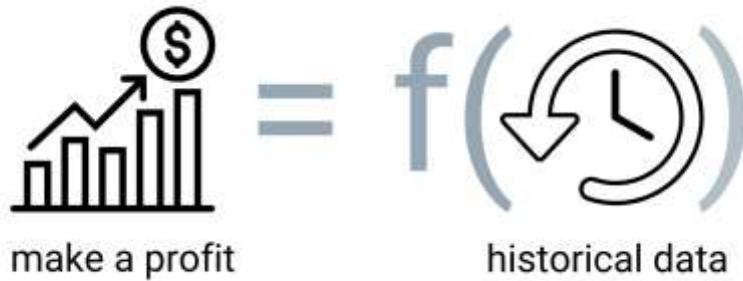
- Združite podatke iz različnih virov in izvajajte analitiko;
- Zagotavljanje preglednosti in verodostojnosti postopkov;
- svetovanje o optimizaciji poslovnih procesov;
- razvoj orodij, ki delujejo v ekosistemu odprtih podatkov, LLM, ETL in Pipelines.

Industrija 5.0 (slika 2.1-12) pomeni konec "obdobja ročnega povprečenja razmerij" in večernih sestankov med direktorji ter oddelkom za ocenjevanje in računovodstvom. Vse, kar je bilo prej skrito - izračuni, ocene, količine - postane odprto, preverljivo in razumljivo tudi nestrokovnjakom. Zmagovalci bodo tisti, ki se bodo prvi preusmerili. Vsi ostali bodo izpadli iz novega digitalnega gospodarstva gradbenega sektorja.

## ZAKLJUČEK

Gradbeništvo vstopa v obdobje korenitih sprememb. Od prvih zapisov na glinenih tablicah do velikih količin digitalnih podatkov, ki se pretakajo s projektnih strežnikov in gradbišč, je zgodovina upravljanja informacij v gradbeništvu vedno odražala zrelost tehnologije svojega časa. Danes, s prihodom avtomatizacije, odprtih formatov in inteligentnih analitičnih sistemov, se panoga ne sooča s postopnim razvojem, temveč s hitro digitalno preobrazbo.

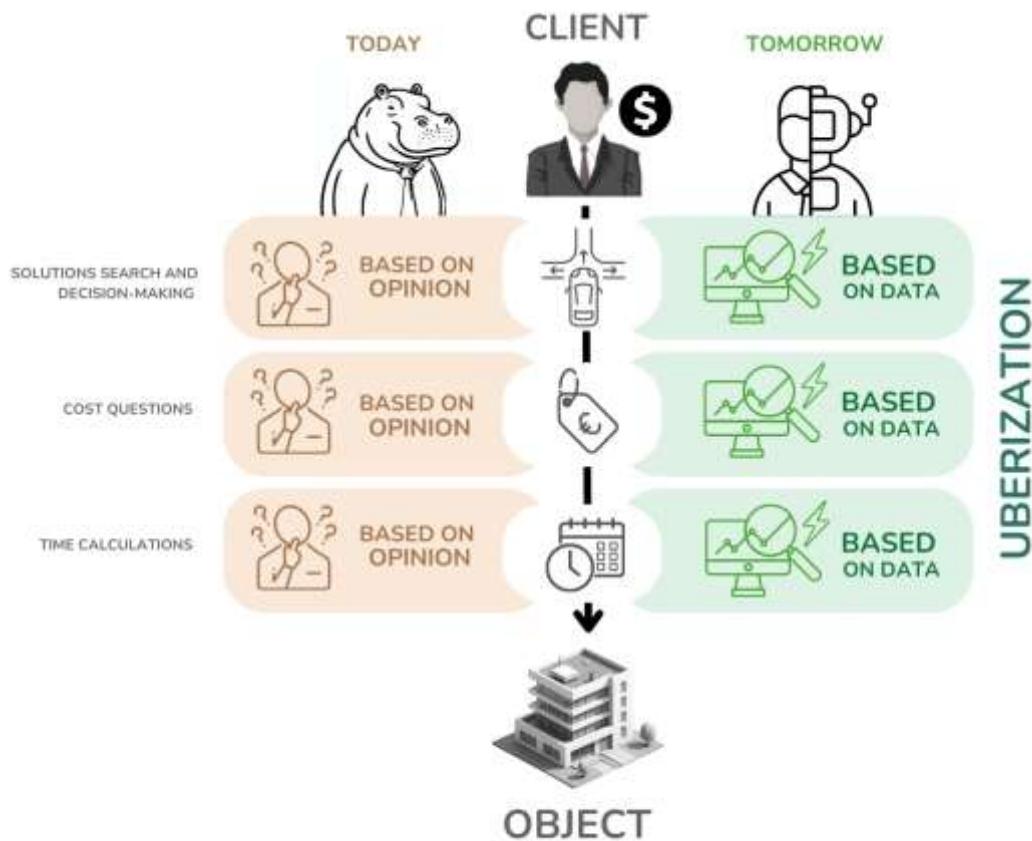
Tako kot v drugih gospodarskih panogah bo treba tudi v gradbeništvu na novo razmisliči ne le o orodjih, temveč tudi o načelih dela. Podjetja, ki so včasih narekovala trg in bila glavni posrednik med naročnikom in projektom, izgubljajo svoj edinstveni položaj. V ospredje stopata zaupanje in sposobnost dela s podatki - od njihovega zbiranja in strukturiranja do analiziranja, napovedovanja in avtomatizacije odločitev.



Slika 10.2-1 Strukturirani zgodovinski podatki so gorivo za učinkovito in obvladljivo poslovanje.

Ta knjiga podrobno opisuje ključna načela upravljanja podatkov v gradbeništvu - od revizije in standardizacije do avtomatizacije procesov, uporabe orodij za vizualizacijo in izvajanja intelligentnih algoritmov. Ogledali smo si, kako lahko tudi z omejenimi sredstvi zgradite delajočo podatkovno arhitekturo in začnete sprejemati odločitve na podlagi preverljivih dejstev in ne intuicije. Delo s podatki ni več le naloga oddelka IT - postane temelj kulture upravljanja, ki določa prožnost, prilagodljivost in dolgoročno vzdržnost podjetja.

Uporaba tehnologij strojnega učenja, sistemov za samodejno obdelavo, digitalnih dvojčkov in odprtih formatov že danes omogoča odpravo človeškega dejavnika tam, kjer je bil prej ključnega pomena. Gradbeništvo se pomika v smeri avtonomije in obvladljivosti, pri čemer lahko gibanje od ideje do realizacije projekta primerjamo z navigacijo v načinu avtopilotova: brez odvisnosti od subjektivnih odločitev, brez potrebe po ročnem posredovanju v vsaki fazi, vendar s popolno digitalno sledljivostjo in nadzorom (slika 10.2-2).



Slika 10.2-2 Prehod od odločanja na podlagi mnenj pomembnih strokovnjakov (HiPPO) k analizi podatkov bo spodbujala predvsem stranka.

S spoznavanjem metod, načel in orodij, predstavljenih v tej knjigi, boste lahko v svojem podjetju začeli sprejemati odločitve, ki temeljijo na podatkih, in ne na intuiciji. Prav tako boste lahko izvajali verige modulov v programih LLM, kopirali izhodiščne cevovode ETL v svoje razvojno okolje (IDE) in samodejno obdelovali podatke, da bi dobili informacije, ki jih potrebujete, v želeni obliki. Na podlagi poglavij v knjigi o velikih količinah podatkov in strojnem učenju boste nato lahko izvajali bolj zapletene scenarije - pridobivali boste nova spoznanja iz zgodovinskih podatkov in uporabljali algoritme strojnega učenja za napovedovanje in optimizacijo svojih procesov.

Odprti podatki in postopki bodo podlaga za natančnejše ocene stroškov in časovnih okvirov projektov, gradbenim podjetjem pa bodo preprečili špekuliranje na podlagi nepreglednih podatkov. To je hkrati izziv in priložnost za industrijo, da ponovno razmisli o svoji vlogi in se prilagodi novemu okolju, v katerem bosta preglednost in učinkovitost postala ključna dejavnika uspeha.

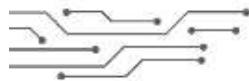
Pripravljenost prevzeti znanje in ga uporabiti v praksi je ključna za uspeh v dobi digitalne preobrazbe.

Podjetja, ki bodo to spoznala prva, bodo imela prednost v novi digitalni konkurenči. Vendar se je treba zavedati, da samo podatki ne bodo ničesar spremenili. Veliko ljudi bo moralo spremeniti način razmišljanja, za kar je potrebna spodbuda. Vaše podjetje mora ponovno razmisli o načinu deljenja

podatkov

Podjetje spremenijo tisti, ki znajo delati s temi podatki, jih interpretirati, uporabiti za optimizacijo in na njihovi podlagi ustvariti novo arhitekturo procesov.

Če berete te vrstice, ste pripravljeni na spremembe in ste že korak pred njimi. Hvala, ker ste se odločili za to pot. Dobrodošli v dobi digitalne preobrazbe!



## O AVTORJU

Moje ime je Artyom Boiko. Moja pot na gradbišču se je začela leta 2007 - z delom rudarja v rudniku naftnega skrilavca v mojem rojstnem mestu, medtem ko sem študiral na Univerzi za rudarstvo v Sankt Peterburgu, kjer sem se specializiral za rudarstvo in podzemno gradnjo. Na zadnji platnici te knjige lahko vidite eksplozionista v čeljusti, kjer smo kopali in razstrelili na stotine kubikov naftnega skrilavca. Moja poklicna pot me je vodila v različne smeri, od dela rudarskega in podzemnega gradbenika do industrijskega plezalca, krovca in monterja dvižne opreme. Imel sem čast sodelovati pri projektih različnega obsega, od gradnje zasebnih hiš do velikih industrijskih objektov v različnih regijah sveta.



Sčasoma se je moje delo preusmerilo od fizične gradnje k upravljanju informacij in digitalnim procesom. Od leta 2013 sem delal na različnih delovnih mestih v malih, srednjih in velikih gradbenih podjetjih v več regijah Nemčije, od projektanta do vodje upravljanja podatkov. Kar zadeva upravljanje podatkov, moje izkušnje obsegajo delo s podatki v različnih sistemih ERP, CAD (BIM), MEP, FEM, CMS. Sodeloval sem pri optimizaciji, avtomatizaciji procesov ter analizi, strojnem učenju in obdelavi podatkov v fazah načrtovanja, izračuna in izvedbe gradbenih del v industrijskih, stanovanjskih, infrastrukturnih in komunalnih gradbenih podjetjih.

Z odprtakodno programsko opremo in odptimi podatki se ukvarjam od leta 2003. V tem času sem izvedel številne spletne projekte - od spletnih strani in spletnih trgovin do celovitih spletnih aplikacij z uporabo odprtakodnih rešitev in odptih sistemov CMS. Te platforme, ki so v marsičem podobne sodobnim gradbenim ERP, imajo modularno arhitekturo, visoko prilagodljivost in dostopnost. Te izkušnje so postavile temelje mojemu strokovnemu pristopu - osredotočenost na odprtakodne tehnologije in kulturo sodelovalnega razvoja. Spoštovanje odprte kode in proste izmenjave znanja je nekaj, kar si prizadevam spodbujati v gradbeništvu. Moje prizadevanje za izboljšanje dostopnosti podatkov v gradbeništvu se je odrazilo v oblikovanju več skupnosti na družbenih medijih za razpravo o odprtosti podatkov in uporabi odprte kode v gradbeništvu ter zagonu več zagonskih podjetij, ki razvijajo rešitve za zagotavljanje dostopa do podatkov iz različnih zaprtih sistemov in platform.

Moj prispevek k strokovni skupnosti je sodelovanje kot govornika na konferencah, ki obravnavajo interoperabilnost CAD (BIM), ERP, 4D-5D, strojno učenje LLM in umetno inteligenco, ter članki, objavljeni v evropskih publikacijah s področja gradbeništva. Eden od mojih pomembnih dosežkov je oblikovanje "Zgodovine BIM". [111], celovitega zemljevida pomembnih programskih rešitev za upravljanje podatkov v gradbeništvu. Moja serija člankov v 7 delih "Razvoj BIM in lobistične igre", prevedena v več jezikov, je bila splošno priznana kot poskus osvetlitve skrite dinamike digitalnih standardov.

Tako sem se od rudarjenja kamnin preusmeril k rudarjenju in sistematizaciji gradbenih podatkov. Vedno sem odprt za strokovni dialog, nove ideje in skupne projekte. Hvaležno bom sprejel vsako povratno informacijo in se veselim vaših sporočil ali pa vas vidim med svojimi sledilci na družbenih omrežjih. Najlepša hvala, ker ste to knjigo prebrali do konca! Vesel bom, če vam bo ta knjiga pomagala bolje razumeti temo podatkov v gradbeništvu.

## NADALJNJI ODNOS

Mnenja bralcev imajo pomembno vlogo pri nadalnjem razvoju publikacij in izbiri prednostnih tem. Posebej dragocene so pripombe o tem, katere zamisli so se izkazale za koristne in katere vzbujajo dvome ter zahtevajo dodatna pojasnila ali navedbo virov. Knjiga vključuje širok spekter gradiva in analiz, od katerih se lahko nekatere zdijo sporne ali subjektivne. Če med branjem najdete netočnosti, napačno navedene vire, logične nedoslednosti ali tipkarske napake, vam bom hvaležen za vaše pripombe, razmišljanja ali kritike, ki jih lahko pošljete na naslov:[boikoartem@gmail.com](mailto:boikoartem@gmail.com). Ali prek sporočil na omrežju LinkedIn:[linkedin.com/in/boikoartem](https://linkedin.com/in/boikoartem)

Hvaležen bom za vsako omembo knjige Data-Driven Construction v družbenih medijih - deljenje bralnih izkušenj pomaga širiti informacije o odprtih podatkih in orodjih ter podpira moje delo.

## KOMENTAR PREVODA

Ta knjiga je bila prevedena s tehnologijo umetne inteligence. Tako je bil postopek prevajanja veliko hitrejši. Vendar pa lahko, tako kot pri vsakem tehnološkem postopku, pride do napak ali netočnosti. Če opazite kar koli, kar se vam zdi napačno ali nepravilno prevedeno, mi prosim pišite po elektronski pošti. Vaše pripombe bodo pripomogle k izboljšanju kakovosti prevoda.

## PODATKOVNO PODPRTE GRADBENE SKUPNOSTI

To je prostor, kjer lahko svobodno postavljate vprašanja ter delite svoje težave in rešitve:

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>

## **DRUGA ZNANJA IN VEŠČINE TER KONCEPTI.**

Poleg ključnih načel dela s podatki v gradbeništvu obravnava DataDrivenConstruction tudi številne dodatne koncepte, programe in spretnosti, ki so bistvenega pomena za strokovnjaka, ki uporablja podatke. Nekatera od njih so predstavljena le v pregledu, vendar so za prakso ključnega pomena.

Zainteresirani bralec lahko obišče spletno mesto DataDrivenConstruction.io, kjer so na voljo povezave do dodatnega gradiva o ključnih znanjih. Ta gradiva vključujejo delo s programoma Python in Pandas, izgradnjo procesov ETL, primere obdelave podatkov v gradbenih projektih CAD, sisteme velikih podatkov ter sodobne pristope k vizualizaciji in analitiki gradbenih podatkov.

Pri pripravi knjige "DataDrivenConstruction" in vseh študij primerov so bila uporabljena številna odprtakoda orodja in programska oprema. Avtor se zahvaljuje razvijalcem in soavtorjem naslednjih rešitev:

- Python in Pandas - hrbtenica manipulacije s podatki in avtomatizacije
- Scipy, NumPy, Matplotlib in Scikit-Learn - knjižnice za analizo podatkov in strojno učenje
- SQL in Apache Parquet - orodji za shranjevanje in obdelavo velikih količin gradbenih podatkov
- Odprta koda CAD (BIM) orodja za odprte podatke v odprtih formatih
- N8n, Apache Airflow, Apache NiFi - sistemi za orkestracijo in avtomatizacijo delovnih tokov
- DeepSeek, LLaMa, Mistral - Odprta koda LLM

Posebna zahvala gre vsem udeležencem razprav na temo odprtih podatkov in orodij v poklicnih skupnostih in družbenih omrežjih, ki so s svojimi kritikami, pripombami in idejami pripomogli k izboljšanju vsebine in strukture te knjige.

Spremljajte razvoj projekta na spletni strani DataDrivenConstruction.io, kjer so poleg posodobitev in popravkov knjige objavljena tudi nova poglavja, navodila in praktični primeri uporabe opisanih tehnik.

## NAJVEČJE UDOBJE S TISKANO RAZLIČICO

V rokah imate brezplačno digitalno različico knjige **Data-Driven Construction**. Za udobnejše delo in hiter dostop do gradiva vam priporočamo, da se posvetite **tiskani izdaji**:



■ **Vedno pri roki:** knjiga v tiskani obliki bo postala zanesljivo delovno orodje, s katerim boste lahko hitro našli in uporabili prave vizualizacije in diagrame v vsaki delovni situaciji.

■ **Visoka kakovost ilustracij:** vse slike in grafike v tiskani izdaji so predstavljene v najvišji kakovosti.

■ **Hiter dostop do informacij:** enostavna navigacija, možnost beleženja opomb, zaznamkov in dela s knjigo kjer koli.

Z nakupom polne tiskane različice knjige dobite priročno orodje za udobno in učinkovito delo z informacijami: možnost hitre uporabe vizualnega gradiva pri vsakodnevnih opravilih, hitrega iskanja potrebnih shem in zapisovanja. Poleg tega s svojim nakupom podpirate širjenje odprtega znanja.

Tiskano različico knjige naročite na: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## EDINSTVENA PRILOŽNOST ZA STRATEŠKO POZICIONIRANJE

V brezplačni različici programa DataDrivenConstruction vam ponujamo namestitev oglasnega gradiva. Plačljiva različica publikacije je že prvo leto po izidu pritegnila pozornost strokovnjakov iz več kot 50 držav sveta - od Latinske Amerike do azijsko-paciške regije. Če se želite pogovoriti o individualnih pogojih sodelovanja in pridobiti podrobne informacije o možnostih plasiranja, izpolnite obrazec za povratne informacije na uradnem portalu [datadrivenconstruction.io](http://datadrivenconstruction.io) ali pišite na kontaktne osebe, navedene na koncu knjige.



## POGLAVJA KNJIGE SO NA VOLJO NA DATADRIVENCONSTRUCTION.IO

Poglavia iz knjige Data-Driven Construction lahko preberete na spletnem mestu Data-Driven Construction, kjer so deli knjige objavljeni postopoma, tako da lahko hitro najdete informacije, ki jih potrebujete, in jih uporabite pri svojem delu. Našli boste tudi številne druge publikacije o podobnih temah ter primere aplikacij in rešitev, ki vam bodo v pomoč pri razvijanju znanja in uporabi podatkov v gradbeništvu.



## NAJNOVEJŠE RAZLIČICE KNJIGE PRENESITE Z URADNE SPLETNE STRANI

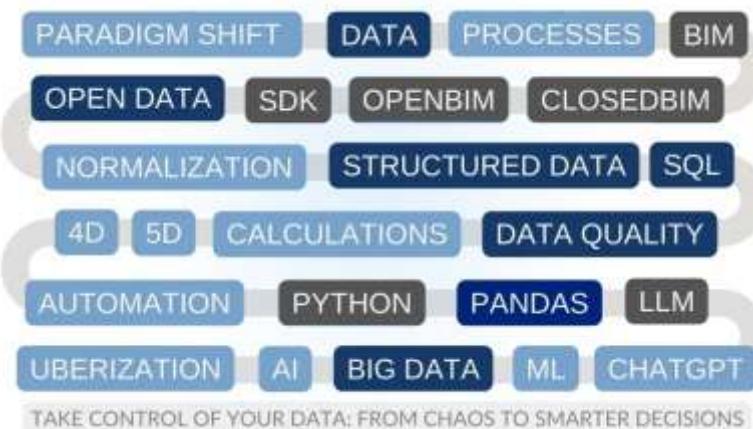
Aktualne in najnovejše različice knjige DataDrivenConstruction so na voljo za prenos na [datadrivenconstruction.io](http://datadrivenconstruction.io). Če želite prejemati posodobitve z novimi poglavji knjige, praktičnimi nasveti ali pregledi novih aplikacij, se naročite na novice:

- Kot prvi se boste seznanili z novimi deli knjige
- Pridobite praktične študije primerov in nasvete o analitiki in avtomatizaciji v gradbeništvu
- Spremljajte najnovejše tende, publikacije in primere uporabe

Pojdite na [datadrivenconstruction.io](http://datadrivenconstruction.io) in se naročite!

# PODATKOVNO PODPRTA GRADNJA: SVETOVANJE, DELAVNICE IN USPOSABLJANJE

Programi usposabljanja in svetovanje podjetja DataDrivenConstruction so pomagali več deset vodilnim gradbenim podjetjem po vsem svetu povečati učinkovitost, zmanjšati stroške in izboljšati kakovost rešitev. Med strankami podjetja DataDrivenConstruction so nekateri največji akterji na milijardnem trgu, vključno z gradbenimi, svetovalnimi in informacijskimi podjetji.



Zakaj izbrati nas?

- **Ustreznost:** govor o glavnih trendih in spoznanjih v panogi
- **Praktično:** pomoč strokovnjakom pri učinkovitem reševanju vsakodnevnih nalog s pomočjo PoC.
- **Osebni pristop:** prilagojen vašemu podjetju, kar najboljša korist usposabljanja in svetovanja

Ekipa DataDrivenConstruction se osredotoča na naslednja glavna področja:

- **Upravljanje kakovosti podatkov:** pomoč pri parametriziranju nalog, zbiranju zahtev, potrjevanju in pripravi podatkov za avtomatizirano obdelavo.
- **Pridobivanje podatkov - pridobivanje in strukturiranje podatkov:** vzpostavite postopke ETL in pridobivajte podatke iz e-pošte, PDF, Excel, slik in drugih virov.
- **BIM in CAD analitika:** zbiranje, strukturiranje in analiziranje informacij iz datotek RVT, IFC, DWG in drugih formatov CAD (BIM).
- **Analitika in preoblikovanje podatkov:** spremenjanje različnih informacij v strukturirane podatke, analitiko, vpoglede in odločitve.
- **Integracija podatkov in avtomatizacija procesov:** od samodejnega ustvarjanja dokumentov do integracije z notranjimi sistemi in zunanjimi zbirkami podatkov.

Kontaktirajte DataDrivenConstruction.io in se pozanimajte, kako lahko z uporabo avtomatizacije vaše podjetje doseže oprijemljive poslovne rezultate.

## GLOSARIJ

**UI (umetna inteligenco)** - umetna inteligenco; zmožnost računalniških sistemov, da opravljajo naloge, ki običajno zahtevajo človeško inteligenco, kot so prepoznavanje vzorcev, učenje in odločanje.

**Apache Airflow** je odprtakodna platforma za orkestracijo delovnih tokov, ki omogoča programsko ustvarjanje, načrtovanje in spremljanje delovnih tokov in ETL z uporabo usmerjenih acikličnih grafov (DAG).

**Apache NiFi** je orodje za avtomatizacijo podatkovnih tokov med sistemi, specializirano za usmerjanje in preoblikovanje podatkov.

**Apache Parquet** je učinkovit datotečni format za shranjevanje podatkov v obliki stolpcev, optimiziran za uporabo v sistemih za analizo velikih količin podatkov. Omogoča znatno stiskanje in hitro obdelavo.

**API (Application Programming Interface)** - formaliziran vmesnik, ki enemu programu omogoča interakcijo z drugim brez dostopa do izvorne kode, izmenjavo podatkov in funkcionalnosti prek standardiziranih zahtev in odzivov.

**Atribut** - značilnost ali lastnost predmeta, ki opisuje njegove lastnosti (npr. površino, prostornino, ceno, material).

**Podatkovne zbirke** so organizirane strukture za shranjevanje, upravljanje in dostop do informacij, ki se uporabljajo za učinkovito iskanje in obdelavo podatkov.

**BEP (BIM Execution Plan)** - Načrt izvajanja informacijskega modeliranja stavb, ki opredeljuje cilje, metode in postopke za izvajanje BIM v projektu.

**Veliki podatki** - množice informacij velike količine, raznolikosti in hitrosti posodabljanja, ki zahtevajo posebne tehnologije za obdelavo in analizo.

**BI (Business Intelligence)** - poslovna inteligenco; procesi, tehnologije in orodja za preoblikovanje podatkov v smiselne informacije za odločanje.

**BIM (informacijsko modeliranje stavb)** - informacijsko modeliranje stavb; postopek ustvarjanja in upravljanja digitalnih predstavitev fizičnih in funkcionalnih značilnosti gradbenih projektov, ki poleg 3D modelov vključuje tudi informacije o zmogljivosti, materialih, časovnem načrtu in stroških.

**BlackBox/WhiteBox** - pristopi k razumevanju sistema: v prvem primeru je notranja logika skrita, vidni so le vhodi in izhodi; v drugem primeru je procesiranje pregledno in na voljo za analizo.

**Bounding Box** je geometrijska konstrukcija, ki opisuje meje predmeta v tridimenzionalnem prostoru z najmanjšimi in največjimi koordinatami X, Y in Z, s čimer se okoli predmeta ustvari "polje".

**BREP (Boundary Representation)** je geometrijska predstavitev predmetov, ki jih opredeljuje z mejami površin.

**CAD (Computer-Aided Design)** je sistem za računalniško podprt načrtovanje, ki se uporablja za ustvarjanje, urejanje in analizo natančnih risb in 3D modelov v arhitekturi, gradbeništву, inženirstvu in drugih panogah.

**CAFM (Computer-Aided Facility Management)** je programska oprema za upravljanje nepremičnin in infrastrukture, ki vključuje načrtovanje prostora, upravljanje sredstev, vzdrževanje in spremljanje stroškov.

**CDE (Common Data Environment)** - centraliziran digitalni prostor za upravljanje, shranjevanje, izmenjavo in sodelovanje s projektnimi informacijami v vseh fazah življenjskega cikla objekta.

**Center odličnosti (CoE)** je specializirana struktura v organizaciji, ki je odgovorna za razvoj določenega področja znanja, razvoj standardov in najboljših praks, usposabljanje osebja in podpora pri uvajanju inovacij.

**CoClass** je sodoben sistem za razvrščanje gradbenih elementov tretje generacije.

**Konceptualni podatkovni model** je predstavitev osnovnih entitet in njihovih povezav na visoki ravni brez podrobnosti o atributih, ki se uporablja v začetnih fazah načrtovanja podatkovne zbirke.

**CRM (Customer Relationship Management)** je sistem za upravljanje odnosov s strankami, ki se uporablja za avtomatizacijo prodajnih in storitvenih procesov.

**DAG (Directed Acyclic Graph)** je usmerjeni aciklični graf, ki se uporablja v sistemih za orkestracijo podatkov (Airflow, NiFi) za določanje zaporedij opravil in odvisnosti.

**Dash** je ogrodje Python za ustvarjanje interaktivnih spletnih vizualizacij podatkov.

**Nadzorna plošča - Nadzorna plošča**, ki vizualno prikazuje ključne kazalnike uspešnosti in metrike v realnem času.

**Podatkovno usmerjen pristop** je metodologija, ki daje prednost podatkom pred aplikacijami ali programsko kodo, tako da podatki postanejo osrednje premoženje organizacije.

**Upravljanje podatkov** - sklop praks, procesov in politik, ki zagotavljajo ustrezeno in učinkovito uporabo podatkov v organizaciji, vključno z nadzorom dostopa, kakovosti in varnosti.

**Podatkovno jezero** je skladišče, namenjeno shranjevanju velikih količin neobdelanih podatkov v izvirni obliki, dokler se ne uporabijo.

**Data Lakehouse** je arhitekturni pristop, ki združuje prilagodljivost in skalabilnost podatkovnih jezer (**Data Lake**) z upravljanjem in zmogljivostjo podatkovnih skladišč (DWH).

**Podatkovno vodena gradnja** je strateški pristop, pri katerem so vse faze življenjskega cikla objekta - od načrtovanja do obratovanja - podprte z avtomatiziranimi, medsebojno povezanimi sistemi. Ta pristop omogoča nenehno učenje na podlagi dejstev, zmanjšuje negotovost in podjetjem omogoča doseganje trajnostnega vodilnega položaja v panogi.

**Integrator, ki temelji na podatkih** - podjetje, specializirano za združevanje podatkov iz različnih virov in njihovo analiziranje za sprejemanje upravljavskih odločitev.

**Pristop, ki temelji na podatkih** - metodologija, pri kateri se podatki obravnavajo kot strateško sredstvo, odločitve pa se sprejemajo na podlagi objektivne analize informacij in ne na podlagi subjektivnih mnenj.

**Minimalizem podatkov** - pristop k zmanjševanju podatkov na najbolj dragocene in pomembne, kar omogoča poenostavljeno obdelavo in analizo informacij.

**Podatkovno močvirje** - razpršena množica nestrukturiranih podatkov, ki nastane, ko se informacije zbirajo in shranjujejo nenadzorovano, brez ustrezne organizacije.

**DataOps** je metodologija, ki združuje načela DevOps, podatke in analitiko ter se osredotoča na izboljšanje sodelovanja, integracije in avtomatizacije podatkovnih tokov.

**Informacijska** digitalizacija je proces pretvorbe vseh vidikov gradbenih dejavnosti v digitalno obliko, primerno za analizo, interpretacijo in avtomatizacijo.

**DataFrame** - dvodimenzionalna tabelarna podatkovna struktura v knjižnici Pandas, kjer vrstice predstavljajo posamezne zapise ali predmete, stolpci pa njihove značilnosti ali attribute.

**Opisna** analitika - analiza preteklih podatkov za razumevanje preteklih dogodkov.

**Diagnostična** analitika - analiziranje podatkov, da bi ugotovili, zakaj se je nekaj zgodilo.

**Gantov diagram** je orodje za načrtovanje projektov, ki naloge prikazuje kot vodoravne črte na časovnem traku in omogoča vizualizacijo zaporedja in trajanja dela.

**DWH (Data Warehouse)** je centraliziran sistem podatkovnega skladišča, ki združuje informacije iz različnih virov, jih strukturira in daje na voljo za analitiko in poročanje.

**ESG (Environmental, Social, Governance)** - sklop meril za ocenjevanje okoljskih, socialnih in upravlja-vskih vplivov podjetja ali projekta.

**ELT (Extract, Load, Transform)** je postopek, pri katerem se podatki najprej pridobijo iz virov in naložijo v skladišče, nato pa se preoblikujejo za analitične namene.

**ETL (Extract, Transform, Load)** je postopek pridobivanja podatkov iz različnih virov, njihovega preoblikovanja v želeno obliko in nalaganja v ciljno shrambo za analizo.

**ER-diagram (Entity-Relationship)** - vizualni diagram, ki prikazuje entitete, njihove atributе in odnose med njimi ter se uporablja pri modeliranju podatkov.

**ERP (Enterprise Resource Planning)** je celovit modularni sistem za načrtovanje virov podjetja, ki se uporablja za upravljanje in optimizacijo različnih vidikov gradbenega procesa.

**Lastnosti** - V strojnem učenju so neodvisne spremenljivke ali atributi, ki se uporabljajo kot vhodni podatki za model.

**Fizični podatkovni model** - podroben prikaz strukture podatkovne zbirke, vključno s tabelami, stolpcji, podatkovnimi tipi, ključi in indeksi, optimiziran za določen sistem DBMS.

**FPDF** je Pythonova knjižnica za ustvarjanje dokumentov PDF.

**Geometric Core** je programska komponenta, ki zagotavlja osnovne algoritme za ustvarjanje, urejanje in analizo geometrijskih objektov v aplikacijah CAD, BIM in drugih inženirskih aplikacijah.

**HiPPO (Highest Paid Person's Opinion)** - pristop k odločanju, ki temelji na mnenju najbolje plačane osebe v organizaciji in ne na objektivnih podatkih.

**IDE (Integrated Development Environment)** - integrirano razvojno okolje, celovito orodje za pisanje, testiranje in odpravljanje napak v kodi (npr. PyCharm, VS Code, Jupyter Notebook).

**IDS (Information Delivery Specification)** je specifikacija zagotavljanja informacij, ki opredeljuje zahteve po podatkih v različnih fazah projekta.

**IFC (Industry Foundation Classes)** je format za izmenjavo podatkov BIM, ki zagotavlja interoperabilnost med različnimi programskimi rešitvami.

**Industrija 5.0** je koncept industrijskega razvoja, ki združuje zmogljivosti digitalizacije, avtomatizacije in umetne inteligence s človeškim potencialom in okoljsko trajnostjo.

**Integracija podatkov** je postopek združevanja podatkov iz različnih virov v enoten, skladen sistem, ki zagotavlja enoten pogled na informacije.

**Informacijski silosi** so izolirani sistemi za shranjevanje podatkov, ki si ne izmenjujejo informacij z drugimi sistemi, kar ovira učinkovito uporabo podatkov.

**IoT (internet stvari)** je koncept povezovanja fizičnih predmetov z internetom za zbiranje, obdelavo in prenos podatkov.

**k-NN (k-Nearest Neighbors)** je algoritem strojnega učenja, ki razvršča predmete na podlagi podobnosti z najbližjimi sosedji v učnem vzorcu.

**Kaggle** je platforma za tekmovanja v analizi podatkov in strojnem učenju.

**Kalkulacija** - izračun stroškov gradbenih del ali postopkov za določeno mersko enoto (npr. 1 m<sup>2</sup> stene iz mavčnih plošč, 1 m<sup>3</sup> betona).

**KPI (Key Performance Indicators)** so ključni kazalniki uspešnosti, merljivi kazalniki, ki se uporabljajo za ocenjevanje uspešnosti podjetja ali določenega projekta.

**Oznake** - V strojnem učenju so ciljne spremenljivke ali atributi, ki jih mora model napovedati.

**Algoritem učenja** - postopek iskanja najboljše hipoteze v modelu, ki ustreza ciljni funkciji, z uporabo niza podatkov za učenje.

**Linearna regresija** - statistična metoda modeliranja razmerja med odvisno spremenljivko in eno ali več neodvisnimi spremenljivkami.

**LLM (Large Language Model)** - Large Language Model, umetna inteligenca, usposobljena za razumevanje in ustvarjanje besedila iz velikih količin podatkov, ki lahko analizira kontekst in piše programsko kodo.

**LOD (Level of Detail/Development)** - raven podrobnosti modela, ki določa stopnjo geometrijske natančnosti in informacijske vsebine.

**Logični podatkovni model** je podrobni opis entitet, atributov, ključev in odnosov, ki odražajo poslovne informacije in pravila ter so vmesna stopnja med konceptualnim in fizičnim modelom.

**Strojno učenje** - razred tehnik umetne inteligence, ki računalniškim sistemom omogočajo učenje in napovedovanje na podlagi podatkov brez izrecnega programiranja.

**Masterformat** je klasifikacijski sistem prve generacije, ki se uporablja za strukturiranje gradbenih specifikacij v oddelke in discipline.

**MEP (Mechanical, Electrical, Plumbing)** - sistemi gradbene tehnike, ki vključujejo mehanske, električne in vodovodne komponente.

**Mreža** je mrežna predstavitev 3D predmetov, sestavljena iz vrhov, robov in površin.

**Model** - Pri strojnem učenju je niz različnih hipotez, od katerih ena aproksimira ciljno funkcijo, ki jo je treba napovedati ali aproksimirati.

**Modeliranje podatkov** je postopek oblikovanja strukturirane predstavitve podatkov in njihovih povezav za uporabo v informacijskih sistemih, vključno s konceptualno, logično in fizično ravnjo.

**n8n** je odprtokodno orodje za avtomatizacijo delovnih tokov in integracijo aplikacij z nizkokodnim pristopom.

**Normalizacija** - v strojnem učenju je postopek približevanja različnih številčnih podatkov skupni lestvici za lažjo obdelavo in analizo.

**Povratni inženiring** - postopek preučevanja naprave, delovanja in proizvodne tehnologije predmeta z analizo njegove strukture, funkcij in delovanja. V kontekstu podatkov - pridobivanje informacij iz lastniških formatov za uporabo v odprtih sistemih.

**OCR (optično prepoznavanje znakov)** je tehnologija **optičnega** prepoznavanja **znakov**, ki pretvori besedilne slike (skenirane dokumente, fotografije) v strojno berljivo obliko besedila.

**OmniClass** je mednarodni klasifikacijski standard druge generacije za upravljanje gradbenih informacij.

**Ontologija** - sistem medsebojnih povezav pojmov, ki formalizira določeno področje znanja.

**Odprta koda** - model za razvoj in distribucijo odprtokodne programske opreme, ki je na voljo za prostouporabo, preučevanje in spremjanje.

**Open BIM** je koncept odprtega BIM, ki vključuje uporabo odprtih standardov in formatov za izmenjavo podatkov med različnimi programskimi rešitvami.

**Odprti standardi** - javno dostopne specifikacije za izvedbo določene naloge, ki različnim sistemom omogočajo interakcijo in izmenjavo podatkov.

**Pandas** je odprtakodna knjižnica Pythona za obdelavo in analizo podatkov, ki zagotavlja podatkovne strukture DataFrame in Series za učinkovito obdelavo tabelaričnih podatkov.

**Paradigma odprtih podatkov** je pristop k obdelavi podatkov, pri katerem so informacije prosto dostopne za uporabo, ponovno uporabo in razširjanje komur koli.

**Parametrična** metoda je metoda ocenjevanja gradbenih projektov, ki uporablja statistične modele za ocenjevanje vrednosti na podlagi parametrov projekta.

**PIMS (Project Information Model)** je digitalni sistem, namenjen organizaciji, shranjevanju in izmenjavi vseh informacij o projektu.

**Potek** - zaporedje postopkov obdelave podatkov, od pridobivanja in preoblikovanja do analize in vizualizacije.

**PMIS (Project Information Management System)** je sistem za vodenje projektov, namenjen podrobnemu nadzoru nalog na ravni posameznega gradbenega projekta.

**Prediktivna** analitika je del analitike, ki uporablja statistične metode in strojno učenje za napovedovanje prihodnjih rezultatov na podlagi preteklih podatkov.

**Prepisovalna** analitika - del analitike, ki ne le napoveduje prihodnje rezultate, temveč tudi predlaga optimalne ukrepe za doseganje želenih rezultatov.

**Lastniški** formati so zaprti podatkovni formati, ki jih nadzoruje določeno podjetje, kar omejuje možnost izmenjave informacij in povečuje odvisnost od določene programske opreme.

**QTO (Quantity Take-Off)** je postopek pridobivanja količin elementov iz projektne dokumentacije za izračun količin materialov, potrebnih za izvedbo projekta.

**Sistem vodenja kakovosti** - sistem vodenja kakovosti, ki zagotavlja, da procesi in rezultati izpolnjujejo določene zahteve.

**RAG (Retrieval-Augmented Generation)** je metoda, ki združuje generativne zmogljivosti jezikovnih modelov z ekstrakcijo ustreznih informacij iz podatkovnih zbirk podjetij ter tako izboljšuje natančnost in ustreznost odgovorov.

**RDBMS (Relational Database Management System)** je sistem za upravljanje relacijskih podatkovnih zbirk, ki organizira informacije v obliki medsebojno povezanih tabel.

**RegEx (Regular Expressions)** je formaliziran jezik za iskanje in obdelavo nizov, ki omogoča določanje predlog za preverjanje skladnosti besedilnih podatkov z določenimi merili.

**Regresija** je statistična metoda za analizo razmerja med spremenljivkami.

**Izračuni CO<sub>2</sub>** so metoda ocenjevanja emisij ogljikovega dioksida, povezanih s proizvodnjo in uporabo gradbenih materialov in postopkov.

**Metoda virov** je metoda ocenjevanja, ki temelji na podrobni analizi vseh virov (materialov, dela, opreme), potrebnih za izvedbo gradbenih del.

**RFID (radiofrekvenčna identifikacija)** je tehnologija za samodejno identifikacijo predmetov z radijskimi signali, ki se uporablja za sledenje materiala, strojev in osebja.

**ROI (Return on Investment)** je kazalnik, ki odraža razmerje med dobičkom in vloženimi sredstvi ter se uporablja za oceno učinkovitosti naložb.

**SaaS (programska oprema kot storitev)** je model programske opreme kot storitve, pri katerem aplikacije gosti ponudnik, uporabnikom pa so na voljo prek interneta.

**SCM (Supply Chain Management)** - upravljanje dobavne verige, ki vključuje usklajevanje in optimizacijo vseh procesov od nabave materialov do dobave končnih izdelkov.

**Podatkovni silosi** so izolirana skladišča informacij v organizaciji, ki niso povezana z drugimi sistemi, kar otežuje izmenjavo podatkov in je neučinkovito.

**SQL (Structured Query Language)** je strukturiran poizvedovalni jezik, ki se uporablja za delo z relacijskimi podatkovnimi zbirkami.

**SQLite** je lahek, vgradljiv, medplatformni sistem DBMS, ki ne potrebuje ločenega strežnika in podpira osnovne funkcije SQL ter se pogosto uporablja v mobilnih aplikacijah in vgrajenih sistemih.

**Strukturirani podatki** - informacije, organizirane v posebni obliki z jasno strukturo, na primer v relacijskih podatkovnih zbirkah ali tabelah.

**ohlapno strukturirani podatki** - informacije z delno organiziranostjo in prilagodljivo strukturo, kot sta JSON ali XML, kjer lahko različni elementi vsebujejo različne sklope atributov.

Entiteta je konkreten ali abstrakten predmet realnega sveta, ki ga je mogoče enolično prepoznati, opisati in predstaviti v obliki podatkov.

**Nadzorovano učenje** - vrsta strojnega učenja, pri kateri se algoritem uči na razdeljenih podatkih, pri čemer je želeni rezultat znan za vsak primer.

**Taksonomija** je hierarhični klasifikacijski sistem, ki se uporablja za sistematično razvrščanje elementov na podlagi skupnih značilnosti.

**Zbirka podatkov Titanic** je priljubljena zbirka podatkov za usposabljanje in testiranje modelov strojnega učenja.

**Usposabljanje** - postopek, v katerem algoritem strojnega učenja analizira podatke za prepoznavanje vzorcev in oblikovanje modela.

**Prenosno učenje** je tehnika strojnega učenja, pri kateri se model, izurjen za eno nalogu, uporabi kot izhodišče za drugo nalogu.

**Preoblikovanje** - postopek spremjanja oblike, strukture ali vsebine podatkov za poznejšo uporabo.

**Podatkovne zahteve** so formalizirana merila, ki opredeljujejo strukturo, obliko, popolnost in kakovost informacij, potrebnih za podporo poslovnim procesom.

**Uberizacija gradbeništva** je proces preoblikovanja tradicionalnih poslovnih modelov v gradbeništvu pod vplivom digitalnih platform, ki omogočajo neposredno interakcijo med strankami in izvajalci brez posrednikov.

**Uniclass** je sistem razvrščanja gradbenih elementov druge in tretje generacije, ki se pogosto uporablja v Združenem kraljestvu.

**USD (Universal Scene Description)** je podatkovni format, ki je bil razvit za računalniško grafiko, vendar se zaradi svoje preproste strukture in neodvisnosti od geometrijskih jeder uporablja tudi v inženirskeih sistemih.

**Potrjevanje podatkov** je postopek preverjanja informacij glede na določena merila in zahteve, da se zagotovi točnost, popolnost in doslednost podatkov.

**Vektorska** podatkovna zbirka - specializirana vrsta podatkovne zbirke, ki hrani podatke kot večdimensionalne vektorje za učinkovito semantično iskanje in primerjavo predmetov.

**Vektorska predstavitev (vgradnja)** je metoda pretvorbe podatkov v večdimensionalne numerične vektorje, ki strojnim algoritmom omogoča učinkovito obdelavo in analizo informacij.

**VectorOps** je metodologija, ki se osredotoča na obdelavo, shranjevanje in analizo večdimenzionalnih vektorskih podatkov, zlasti na področjih, kot so digitalni dvojčki in semantično iskanje.

**Vizualizacija** - grafični prikaz podatkov za boljše zaznavanje in analizo informacij.

Izrazi so bili razvrščeni po abecednem redu glede na njihovo angleško ime.

## SEZNAM LITERATURE IN SPLETNEGA GRADIVA

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics" (Podatki o ključnih metrikah IT 2017: indeks objavljenih dokumentov in metrik), 12. december 2016. [Online]. Dostopno prek: <https://www.gartner.com/en/documents/3530919>. [Datum dostopa: 1. marec 2025].
- [2] KPMG, "Znani izzivi - novi pristopi. Globalna raziskava o gradbeništvu 2023, 1. januar 2023. [Online]. Dostopno na: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Dostop: 5. marec 2025].
- [3] F. R. Barnard, "Slika je vredna tisoč besed", 10 Mapi 1927. [Online]. Dostopno prek: [https://en.wikipedia.org/wiki/A\\_picture\\_is\\_worth\\_a\\_thousand\\_words](https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words). [Datum objave: 15. marec 2025].
- [4] M. Bastian, "Microsoftov izvršni direktor Satya Nadella pravi, da je samooklicanje AGI "nesmiselno hekanje po merilih", 21. februar 2025. [Online]. Dostopno prek: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [Datum dostopa: 15. marec 2025].
- [5] W. E. Forum, "Forum oblikuje prihodnost gradbeništva - pokrajina v preobrazbi:", 1. januar 2016. [Online]. Dostopno prek: [https://www3.weforum.org/docs/WEF\\_Shaping\\_the\\_Future\\_of\\_Construction.pdf](https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf). [Datum dostopa: 2. marec 2025].
- [6] C. D. Gillespie, "Clay: Earth's entanglement in the age of clay", 2024. [Online]. Dostopno prek: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Papirus, 3. stoletje pr. n. št. Jezik je grški," 2024. [Online]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Spremljanje: uporaba razpoložljivih orodij", 1980. [Online]. Dostopno na: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Datum objave: 15. marec 2025].
- [9] PWC, "Na podatkih temelječe informacije, ki jih študenti potrebujejo za uspeh v hitro spreminjačem se poslovнем svetu", 15. februar 2015. [Online]. Dostopno prek: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Datum dostopa: 15. marec 2025].
- [10] Skanska ZDA, "Trendi na jesenskem gradbenem trgu", 2. november 2023. [Online]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.

- [11] "Oxford Essential Quotations (4. izdaja)", Oxford University Press, 2016. [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-or-ed4-00006236>. [Datum dostopa: 1. marec 2025].
- [12] "Citat: Sondergaard o podatkovni analitiki," [na spletu]. Dostopno prek: <https://www.causeweb.org/cause/resources/library/r2493>. [Datum dostopa: 15. marec 2025].
- [13] "Kako globalno zanimanje za umetno inteligenco spodbuja trg upravljanja podatkov", 28. maj 2024. [Online]. Dostopno prek: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Datum dostopa: 15. marec 2025].
- [14] И. McCue, "Zgodovina ERP", 2024. [Online]. Dostopno na: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "Cena ERP: Koliko stane sistem ERP?", [na spletu]. Dostopno na: <https://erpscout.de/en/erp-costs/>. [Datum dostopa: 15. marec 2025].
- [16] softwarepath, "Kaj nam 1384 projektov ERP pove o izbiri ERP (poročilo ERP 2022)", 18. januar 2022. [Online]. Dostopno na: <https://softwarepath.com/guides/erp-report>. [Datum dostopa: 15. marec 2025].
- [17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16. december 2016. [Online]. Dostopno prek: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Datum naslavljanja: 1. marec 2025].
- [18] McKinsey, "Podatkovno vodeno podjetje leta 2025", 28. januar 2022. [Online]. Dostopno na: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Datum objave: 22. maj 2024].
- [19] Wikipedija, "Moorov zakon," [na spletu]. Dostopno prek: [https://en.wikipedia.org/wiki/Moore%27s\\_law](https://en.wikipedia.org/wiki/Moore%27s_law). [Datum dostopa: 15. marec 2025].
- [20] Accenture, "Ustvarjanje večje vrednosti s kapitalskimi projekti", 1. januar 2020. [Online]. Dostopno na: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf>. [Datum dostopa: 3. marec 2024].
- [21] Б. Marr, "Koliko podatkov ustvarimo vsak dan? Razburljivi statistični podatki, ki bi jih moral prebrati vsakdo," 2018. [Online]. Dostopno prek: <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] "Koliko podatkov se ustvari vsak dan?", 2024. [na spletu]. Dostopno na: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.
- [23] T. Sullivan, "AI and the global 'datasphere': how much information will humanity have by 2025?", 2024. [Online]. Dostopno na: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "Total number of printed books produced in various regions of Western Europe in each half century between 1454 and 1800," [Internet]. [na spletu]. Dostopno prek: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [Datum objave: 1. marec 2025].
- [25] "Primeri oblikovanja cen", 2024. [Online]. Dostopno na: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Podjetja zaradi vse večje kompleksnosti oddajajo shranjevanje podatkov v zunanje izvajanje," 10. maj 2024. [Online]. Dostopno prek: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Datum dostopa: 15. marec 2025].
- [27] JETSOFTPRO, "SaaS je mrtev? Razlaga šokantne napovedi Microsoftovega izvršnega direktorja," 13. januar 2025. [Online]. Dostopno na: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12. december 2024. [Online]. Dostopno prek: [https://www.youtube.com/watch?v=9NtsnzRFJ\\_o](https://www.youtube.com/watch?v=9NtsnzRFJ_o). [Datum dostopa: 15. marec 2025].
- [29] GoodReads, "Tim Berners-Lee," [Online]. Dostopno na: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Datum dostopa: 15. marec 2025].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1. januar 2023. [Online]. Dostopno na: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Dostop: 5. marec 2025].
- [31] I. Deininger, B. Koch, R. Bauknecht in 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. [na spletu]. Dostopno na: [https://www.researchgate.net/publication/374023998\\_Using\\_Digital\\_Models\\_to\\_Decarbonize\\_a\\_Production\\_Site\\_A\\_Case\\_Stud...](https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Stud...)
- [32] McKinsey, "REINVENTIFICIRANJE GRADBENIŠTVA: POT K VELIKI PROIZVODNOSTI", 1. februar 2017. [Online]. Dostopno prek: <https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insi>

- ghts/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf.
- [33] Delovna skupina za gradbeništvo za namestnika predsednika vlade, "Rethinking Construction", 1. oktober 2014. [Online]. Dostopno prek: [https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking\\_construction\\_report.pdf](https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf).
- [34] Forbes, "Brez mnenja ste le še ena oseba s podatki", 15. marec 2016. [Online]. Dostopno prek: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Datum dostopa: 15. marec 2025].
- [35] Wikiquote, "Charles Babbage," [Online]. Dostopno prek: [https://en.wikiquote.org/wiki/Charles\\_Babbage](https://en.wikiquote.org/wiki/Charles_Babbage). [Datum dostopa: 15. marec 2025].
- [36] SAP, "Nova raziskava ugotavlja, da skoraj polovica vodstvenih delavcev zaupa umetni inteligenci bolj kot sebi", 12. marec 2025. [Online]. Dostopno prek: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Datum dostopa: 15. marec 2025].
- [37] Kanadsko gradbeno združenje in KPMG v Kanadi, 2021, "Gradbeništvo v digitalnem svetu", 1. maj 2021. [Online]. Dostopno prek: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Datum dostopa: 5. marec 2025].
- [38] ZCS, "Dekodiranje pete industrijske revolucije". [na spletu]. Dostopno na: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Datum dostopa: 15. marec 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, "Vrednost odprtakodne programske opreme", 24. januar 2024. [Online]. Dostopno prek: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Datum dostopa: 15. marec 2025].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1. september 2008. [Online]. Dostopno na: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] McKinsey, "Izboljšanje produktivnosti v gradbeništvu," [na spletu]. Dostopno na: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Datum objave: 15. marec 2025].

- [43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector", 19. januar 2023. [Online]. Dostopno prek: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [Datum dostopa: 1. marec 2025].
- [44] McKinsey, "Uresničevanje produktivnosti v gradbeništvu ni več neobvezno", 9. avgust 2024. [Online]. Dostopno prek: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Datum dostopa: 5. marec 2025].
- [45] ING Group, "Zaostajanje produktivnosti v gradbeništvu povečuje stroške gradnje", 12. december 2022. [Online]. Dostopno na: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Datum obravnave: 15. marec 2025].
- [46] M. Berman, "Šokantna napoved Microsoftovega izvršnega direktorja: "Agenti bodo nadomestili vso programsko opremo", 19. december 2024. [Online]. Dostopno prek: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Datum dostopa: 15. marec 2025].
- [47] Business Insider, "Anthropic's CEO says that in 3 to 6 months, AI will be writing 90% of the code of software developers were charge of," 15. marec 2025. [Online]. Dostopno prek: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Datum dostopa: 30. marec 2025].
- [48] Statista, "Primerjava priljubljenosti sistemov za upravljanje podatkovnih zbirk (DBMS) po svetu junija 2024, po kategorijah", junij 2024. [Online]. Dostopno prek: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Datum objave: 15. marec 2025].
- [49] DB-Engines, "DB-Engines Ranking," [na spletu]. Dostopno na: <https://db-engines.com/en/ranking>. [Datum objave: 15. marec 2025].
- [50] "Raziskava razvijalcev Stack Overflow 2023," 2024. [Online]. Dostopno prek: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [Online]. Dostopno na: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Strukturirani in nestrukturirani podatki: kakšna je razlika?", 2024. [Online]. Dostopno na: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "PDF PRIMERJAVA PODATKOVNIH FORMATOV ZA GRADBENE PROJEKTE", 23. 4. 2024. [Online]. Dostopno na: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.

- [54] "Spletna stran z informacijskim modeliranjem stavb," 2003. [Online]. Dostopno na: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.
- [55] A. Boyko, "Lobistične vojne in razvoj BIM. 5. del: BlackRock je mojster vseh tehnologij. Kako korporacije nadzorujejo odprtakodno kodo," 2024. [Online]. Dostopno prek: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] D. Ušakov, "Neposredno modeliranje - kdo in zakaj ga potrebuje? Pregled konkurenčnih tehnologij", 14. 11. 2011. [Online]. Dostopno prek: [https://isicad.net/articles.php?article\\_num=14805](https://isicad.net/articles.php?article_num=14805). [Datum objave: 02 2025].
- [57] C. Eastman in A. Cthers, "Eastman, Charles; And Cthers", september 1974. [Online]. Dostopno prek: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Datum dostopa: 15. marec 2025].
- [58] D. Ušakov, "Neposredno modeliranje - kdo in zakaj ga potrebuje? Pregled konkurenčnih tehnologij", 11. november 2011. [Online]. Dostopno prek: [https://isicad.net/articles.php?article\\_num=14805](https://isicad.net/articles.php?article_num=14805). [Datum objave: 15. marec 2025].
- [59] D. Weisberg, "Zgodovina CAD," 12. december 2022. [Online]. Dostopno na: [https://www.shapr3d.com/blog/history-of-cad?utm\\_campaign=cadhISTORYnet](https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhISTORYnet). [Datum sklica: 15. marec 2025].
- [60] ADSK, "Bela knjiga o informacijskem modeliranju stavb", 2002. [Online]. Dostopno na: [https://web.archive.org/web/20060512180953/http://images.adsk.com/apac\\_sapac\\_main/files/4525081\\_BIM\\_WP\\_Rev5.pdf#expand](https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand). [Datum objave: 15. marec 2025].
- [61] ADSK, "Bela knjiga Informacijsko modeliranje stavb v praksi," [na spletu]. Dostopno na: [https://web.archive.org/web/20060512181000/http://images.adsk.com/apac\\_sapac\\_main/files/4525077\\_BIM\\_in\\_Practice.pdf](https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf). [Datum dostopa: 15. marec 2025].
- [62] A. Boyko, "Lobistične vojne in razvoj BIM. 2. del: odprti BIM in zaprti BIM. Evropa in preostali svet," 2024. [Online]. Dostopno prek: <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]. Dostopno prek: [https://youtu.be/S-TNdUgfHxk?si=evM\\_v28KQbGOG0k&t=1360](https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360).
- [64] ADSK, "Bela knjiga BIM", 2002. [Online]. Dostopno na: [https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac\\_sapac\\_main/files/4525081\\_BIM\\_WP\\_Rev5.pdf#expand](https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand). [Datum objave: 15. marec 2025].

- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [Online]. Dostopno na:  
[https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489\\_Benefits\\_of\\_MAI.pdf](https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf). [Datum dostopa: 15. marec 2025].
- [66] M. Shacklett, "Strukturirani in nestrukturirani podatki: 2024. [Online]. Dostopno na:  
<https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Dostopno na:  
<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]. Dostopno prek: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf> . [Datum dostopa: 02 2025].
- [69] CrowdFlower, "Poročilo o podatkovni znanosti 2016", 2016. [Online]. Dostopno na:  
[https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower\\_DataScienceReport\\_2016.pdf](https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf). [Datum dostopa: 15. marec 2025].
- [70] Analyticsindiamag, "6 najbolj zamudnih nalog za podatkovne znanstvenike", 15. maj 2019. [Online]. Dostopno prek: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] BizReport, "Poročilo: Podatkovni znanstveniki porabijo večino časa za čiščenje", 6. julij 2015. [Online]. Dostopno prek:  
<https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Datum dostopa: 5. marec 2025].
- [72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27. julij 2015. [Online]. Dostopno prek:  
[https://www.reddit.com/r/science/comments/3nyn5i/science\\_ama\\_series\\_stephen\\_hawking\\_a\\_ma\\_answers/](https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_a_ma_answers/). [Datum dostopa: 15. marec 2025].
- [73] B. Cyphers in K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [Online]. Dostopno na: <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]. Dostopno prek:  
<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Datum dostopa: 15. marec 2025].
- [75] A. Boyko, "Boj za odprte podatke v gradbeništvu. Zgodovina programov AUTOLISP, intelliCAD, openDWG, ODA in openCASCADE," 15 05 2024. [Online]. Dostopno prek:  
<https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry->

- 2b97200e6393. [Datum objave: 16. 02. 2025].
- [76] Wikipedija, "Microsoft in odprta koda," [na spletu]. Dostopno prek: [https://en.wikipedia.org/wiki/Microsoft\\_and\\_open\\_source](https://en.wikipedia.org/wiki/Microsoft_and_open_source). [Datum dostopa: 15. marec 2025].
- [77] TIME, "Razlika med odprtimi in zaprtimi modeli umetne inteligence se morda zmanjšuje. Zakaj je to pomembno?", 5. november 2024. [Online]. Dostopno prek: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Datum dostopa: 15. marec 2025].
- [78] The Verge, "Več kot četrtino nove kode v Googlu ustvari umetna inteligenca", 29. oktober 2024. [Online]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Datum dostopa: 15. marec 2025].
- [79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15. december 2020. [Online]. Dostopno na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-obdelava>. [Dostop: 15. marec 2025].
- [80] PwC, "PwC Open Source Monitor 2019", 2019. [Online]. Dostopno na: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Dostopno 15. marca 2025].
- [81] Travers Smith, "Odprta skrivnost: odprtakodna programska oprema", 2024. [Online]. Dostopno prek: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Datum dostopa: 15. marec 2025].
- [82] Deloitte, "Postopek prenosa podatkov pri preoblikovanju podjetij", 2021. [Online]. Dostopno na: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Datum objave: 15. marec 2025].
- [83] gov.uk, "Data Analytics and AI in Government Project Delivery", 20. marec 2024. [Online]. Dostopno na: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [Datum dostopa: 5. marec 2025].
- [84] "Izvorni citat: Vse bi moralo biti čim bolj preprosto, vendar ne preprostejše," 13. maj 2011. [Online]. Dostopno prek: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Datum dostopa: 15. marec 2025].
- [85] "Transformator (arhitektura globokega učenja)," [na spletu]. Dostopno na: [https://en.wikipedia.org/wiki/Transformer\\_\(arhitektura\\_globokega\\_učenja\)](https://en.wikipedia.org/wiki/Transformer_(arhitektura_globokega_učenja)). [Datum dostopa: 15. marec 2025].
- [86] "Statistika prenosa paketov Python," 2024. [Online]. Dostopno na:

- [https://www.pepy.tech/projects/pandas.](https://www.pepy.tech/projects/pandas)
- [87] Intervju Bit, "10 najboljših knjižnic Python", 2023. [Online]. Dostopno na: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more>. [Datum dostopa: 30. marec 2025].
- [88] "NVIDIA in HP pospešuje podatkovno znanost in generativno umetno inteligenco na delovnih postajah," 7. marec 2025. [Online]. Dostopno na: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Datum dostopa: 15. marec 2025].
- [89] P. Orac, "Kako v nekaj sekundah obdelati podatkovni okvir z milijoni vrstic", 2024. [Online]. Dostopno prek: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Kaj je Kaggle?", 2024. [Online]. Dostopno prek: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "Glavni izvršni direktor družbe NVIDIA Jensen Huang: Keynote na sejmu COMPUTEX 2024", 2. junij 2024. [Online]. Dostopno prek: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Dostopno 15. marca 2025].
- [92] "Člani: ustanovitelji in korporativni člani", 2024. [Online]. Dostopno prek: <https://www.opendesign.com/member-showcase>.
- [93] A. Bojko, "Doba sprememb: IFC je stvar preteklosti ali zakaj so se ADSK in drugi ponudniki CAD pripravljeni odpovedati IFC za USD v 14 ključnih dejstvih", 24. november 2024. [Online]. Dostopno prek: <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-3f9a82cccd10a>. [Dostop: 23. februar 2025].
- [94] A. Boyko, "The post-BIM world. Prehod na podatke in procese ter ali gradbena industrija potrebuje semantiko, formate in interoperabilnost," 20. dec. 2024. [Online]. Dostopno prek: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Datum obravnave: 23. februar 2025].
- [95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE", 2016. [Online]. Dostopno prek: [https://datascience.nih.gov/sites/default/files/NIH\\_Strategic\\_Plan\\_for\\_Data\\_Science\\_Final\\_508.pdf](https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf). [Datum dostopa: 23. februar 2025].
- [96] Harvard Business Review, "Bad Data Costs the U.S. \$3 Trillion Per Year", 22. september 2016. [Online]. Dostopno prek: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Učinki kakovosti podatkov", 1. januar 2025. [Online]. Dostopno na:

- [https://delpha.io/impacts-of-data-quality/.](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" (Oblikovanje za vzdrževanje: pomen upoštevanja obratovanja in vzdrževanja v fazi načrtovanja gradbenih projektov)," [na spletu]. Dostopno na: <https://www.wbdg.org/resources/design-for-maintainability>. [Datum dostopa: 15. marec 2025].
- [99] O. o. D. C. P. a. Oversight, "Priročnik za načrtovanje preprečevanja in obvladovanja korozije za vojaške sisteme in opremo", april 2014. [Online]. Dostopno na: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guide book%204%20Feb%202014.pdf>. [Datum dostopa: 15. marec 2025].
- [100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1. januar 2025. [Online]. Dostopno na: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "V iskanju žeblja," [na spletu]. Dostopno prek: [https://en.wikipedia.org/wiki/For\\_Want\\_of\\_a\\_Nail](https://en.wikipedia.org/wiki/For_Want_of_a_Nail). [Datum dostopa: 15. marec 2025].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", oktober 2013. [Online]. Dostopno na: [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](https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf). [Datum dostopa: 15. marec 2025].
- [103] EY, "Pot do ogljične neutralnosti", 10. marec 2023. [Online]. Dostopno na: [https://www.ey.com/ru\\_kz/services/consulting/the-path-to-carbon-neutrality](https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality). [Datum dostopa: 15. marec 2025].
- [104] PWC, "ESG Awareness", 1. julij 2024. [Online]. Dostopno na: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Datum dostopa: 15. marec 2025].
- [105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [Online]. Dostopno na: <https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO<sub>2</sub>\_izračunavanje vsebovanega ogljika", 2024. [Online]. Dostopno na: [https://github.com/datadrivenconstruction/CO2\\_calculating-the-embodied-carbon](https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon).
- [107] McKinsey, "Imagining Construction's Digital Future", 24. junij 2016. [Online]. Dostopno prek: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [Datum dostopa: 25. februar 2025].

- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10. oktober 2024. [Online]. Dostopno prek: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [Datum objave: 15. marec 2025].
- [109] SAS, "Podatkovno jezero in podatkovno skladišče - spoznajte razliko," [na spletu]. Dostopno na: [https://www.sas.com/en\\_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html](https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html). [Datum objave: 15. marec 2025].
- [110] ADSK, "Informacijsko modeliranje stavb", 2002. [Online]. Dostopno na: [https://www.laiserin.com/features/bim/autodesk\\_bim.pdf](https://www.laiserin.com/features/bim/autodesk_bim.pdf). [Datum objave: 15. marec 2025].
- [111] A. Boyko, "Zgodovinski zemljevid BIM", 2024. [Online]. Dostopno na: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "Opredelitve BIM po organizacijah in standardih", 27. december 2023. [Online]. Dostopno na: <https://encyclopedia.pub/entry/53149>. [Datum obravnave: 5. marec 2025].
- [113] Prodajalec CAD, "Program OPEN BIM", 2012. [Online]. Dostopno na: [https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open\\_bim/](https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/). [Datum objave: 30. marec 2025].
- [114] Wikipedija, "Industry Foundation Classes," [na spletu]. Dostopno na: [https://en.wikipedia.org/wiki/Industry\\_Foundation\\_Classes](https://en.wikipedia.org/wiki/Industry_Foundation_Classes). [Datum dostopa: 15. marec 2025].
- [115] Wikipedija, "IGES," [na spletu]. Dostopno na: <https://en.wikipedia.org/wiki/IGES>. [Datum objave: 30. marec 2025].
- [116] A. Bojko, "Zgodovina CAD (BIM)", 15. december 2021. [Online]. Dostopno prek: [https://miro.com/app/board/o9J\\_laML2cs=/](https://miro.com/app/board/o9J_laML2cs=/). [Datum nagovora: 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. 1. 2021. [Online]. Dostopno prek: <https://arxiv.org/pdf/2007.10951.pdf>. [Datum dostopa: 5. marec 2025].
- [118] И. Rogačev, "Pogovorimo se o BIM: Maxim Nechiporenko | Renga | IFC | Domači BIM," 13. april 2021. [Online]. Dostopno prek: <https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be>. [Datum nagovora: 5. marec 2025].
- [119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth", 17. december 2024. [Online]. Dostopno prek: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [Datum dostopa: 5. marec 2025].

- [120] "Flex token cost", 2024. [Online]. Dostopno na: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] A. Boyko, "Pozabite na BIM in demokratizirajte dostop do podatkov (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024. [Online]. Dostopno prek: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Baprozessmanagment-und-Bauindustrie.pdf>.
- [122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh in F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [Online]. Dostopno prek: <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]. Dostopno prek: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Datum dostopa: 2. marec 2025].
- [124] Apple.com, "Pixar, Adobe, Apple in NVIDIA so ustanovili zavezništvo OpenUSD za spodbujanje odprtih standardov za 3D vsebine," 1. avgust 2023. [Online]. Dostopno prek: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Dostop: 2. marec 2025].
- [125] AECmag, "ADSK-jeva granularna podatkovna strategija", 25. jul. 2024. [Online]. Dostopno prek: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [Datum dostopa: 15. marec 2025].
- [126] A. Bojko, "Doba sprememb: IFC je stvar preteklosti ali zakaj so se ADSK in drugi ponudniki CAD pripravljeni odpovedati IFC za USD v 14 ključnih dejstvih", 24. 11. 2024. [Online]. Dostopno prek: <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-3f9a82cccd10a>. [Datum obravnave: 23 fevrial 2025].
- [127] A. Boyko, "ENG BIM Cluster 2024 | Bitka za podatke in uporabo LLM in ChatGPT v gradbeništvu", 7. avgust 2024. [Online]. Dostopno prek: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [Dostopno: 15. marec 2025].
- [128] "Jeffrey Zeldman predstavlja", 6. maj 2008. [Online]. Dostopno prek: <https://zeldman.com/2008/05/06/content-precedes-design/>. [Datum dostopa: 15. marec 2025].
- [129] A. Boyko, "DWG Analyse with ChatGPT | DataDrivenConstruction", 5. marec 2024. [Online]. Dostopno na: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [Datum objave: 15. marec 2025].

- [130] McKinsey , "McKinseyjev vodnik za doseganje konkurenčnosti v dobi digitalne in umetne intelligence", 2023. [Online]. Dostopno prek: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Datum objave: 30. marec 2025].
- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31. marec 2016. [Online]. Dostopno prek: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Datum dostopa: 15. marec 2025].
- [132] J. Bertin, "Grafika in grafična obdelava informacij", 8. september 2011. [Online]. Dostopno na: [https://books.google.de/books/about/Graphics\\_and\\_Graphic\\_Information\\_Process.html?id=csqX\\_xnm4tcC&redir\\_esc=y](https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y). [Datum dostopa: 15. marec 2025].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking," [Online]. Dostopno na: <https://www.causeweb.org/cause/resources/library/r1266>. [Datum dostopa: 15. marec 2025].
- [134] Ministrymagazine, "Kako je znanost odkrila stvarstvo", januar 1986. [Online]. Dostopno na: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [Datum objave: 15. marec 2025].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23. maj 2017. [Online]. Dostopno prek: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Dostopno 15. maja 2024].
- [136] "Kako zgraditi podatkovno arhitekturo za spodbujanje inovacij - danes in jutri", 3. junij 2020. [Online]. Dostopno prek: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Datum nagovora: 15. marec 2025].
- [137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-ored6-00011630>. [Datum nagovora: 15. marec 2025].
- [138] "Prevorniki", 2024. [Online]. Dostopno na: <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]. Dostopno prek: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Datum dostopa: 18. februar 2025].
- [140] "Cevovod v gradnji", 2024. [Online]. Dostopno na: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Wikipedija, "Apache NiFi", 1. januar 2025. [Online]. Dostopno prek:

- [https://de.wikipedia.org/wiki/Apache\\_NiFi](https://de.wikipedia.org/wiki/Apache_NiFi). [Datum dostopa: 5. marec 2025].
- [142] n8n, "Gmail AI Auto-Responder: ustvarjanje osnutkov odgovorov na prejeta e-poštna sporočila", 1. maj 2024. [Online]. Dostopno prek: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [Dostopno 15. marca 2025].
- [143] n8n, "Avtomatizacija nepremičninskih dnevnih ponudb z Zillow API, Google Sheets in Gmail", 1. marec 2025. [Online]. Dostopno na: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [Datum dostopa: 15. marec 2025].
- [144] B. T. O'Neill, "Stopnja neuspešnosti projektov analitike, umetne inteligence in velikih podatkov = 85 % - jo!", 1. januar 2025. [Online]. Dostopno prek: <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]. Dostopno prek: [https://julia.quantecon.org/dynamic\\_programming/wald\\_friedman.html](https://julia.quantecon.org/dynamic_programming/wald_friedman.html). [Datum dostopa: 1. maj 2024].
- [147] T. Landsall-Welfair, Napovedovanje trenutnega razpoloženja naroda, *Significance*, 2012.
- [148] A. Boyko, "San Francisco. Gradbeni sektor 1980-2019," 2024. [Online]. Dostopno prek: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC Files 5000 Projects", 2024. [Online]. Dostopno prek: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Priprava na prihodnost s povezano gradnjo". [na spletu]. Dostopno na: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [Datum dostopa: 15. marec 2025].
- [151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing", 22. maj 2017. [Online]. Dostopno na: <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] "Pogoji, potrebni za ohranitev rastlinskih fosilov", 2024. [Online]. Dostopno na: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "Fink iz družbe BlackRock o obveznicah, združtvah in prevzemih, recesiji v ZDA in voltvah: celoten intervju," 2023. [Online]. Dostopno prek:

- <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 znanih katastrof umetne inteligence", 02. oktober 2024. [Online]. Dostopno prek: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [Datum dostopa: 15. marec 2025].
- [155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.
- [156] Wikipedija, "Kaggle", 1. januar 2025. [Online]. Dostopno prek: <https://en.wikipedia.org/wiki/Kaggle>. [Datum dostopa: 15. marec 2025].
- [157] Kaggle, "Titanik - strojno učenje iz katastrofe", 1. januar 2025. [Online]. Dostopno na: <https://www.kaggle.com/competitions/titanic/overview>. [Dostopno 10. marca 2025].
- [158] Ш. Johri, "Ustvarjanje ChatGPT: od podatkov do dialoga", 2024. [Online]. Dostopno na: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Domingos, "Nekaj koristnih stvari o strojnem učenju", 2024. [Online]. Dostopno prek: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Citatni citat", [na spletu]. Dostopno prek: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Datum dostopa: 17. marec 2025].
- [161] NVIDIA, "Izboljšajte svoje podatke za usposabljanje z novimi modeli klasifikatorja NVIDIA NeMo Curator", 19. december 2024. [Online]. Dostopno na: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Dostopno 25. marca 2025].
- [162] "NVIDIA napoveduje večjo izdajo modelov fundacije Cosmos World in orodij za fizikalne podatke umetne inteligence," 18. marec 2025. [Online]. Dostopno prek: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Datum dostopa: 25. marec 2025].
- [163] NVIDIA, "NVIDIA Isaac Sim," [na spletu]. Dostopno na: <https://developer.nvidia.com/isaac/sim>. [Dostopno 25. marca 2025].
- [164] M. Quarterly, "Zakaj digitalne strategije ne uspejo", 25. januar 2018. [Online]. Dostopno prek: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Datum dostopa: 15. marec 2025].
- [165] M. J. Perry, "Moji najljubši citati Miltona Friedmana", 17. november 2006. [Online]. Dostopno prek: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Datum dostopa: 1.

marec 2025].

- [166] J. A. Wheeler, "Informacije, fizika, kvantni elementi: iskanje povezav", 1990.
- [169] A. Boyko, "Lobistične vojne in razvoj BIM. 5. del: BlackRock je mojster vseh tehnologij. Kako korporacije nadzorujejo odprtakodno kodo," 2024. [Online]. Dostopno prek: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen in J. Beetz, "A SPARQL query engine for binary-formatted IFC building models", *Napredna inženirska informatika*, 2024.
- [171] "Število podjetij v gradbenem sektorju v Združenem kraljestvu leta 2021 po velikosti podjetja", 2024. [na spletu]. Dostopno prek: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5.000 projektov IFC&RVT," 2024. [Online]. Dostopno na: <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]. Dostopno prek: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Datum objave: februar 2025].
- [174] *Lobistične vojne za podatke v gradbeništvu / Tehnofeodalizem in zgodovina skrite preteklosti BIM*. [Film]. Nemčija: Artyom Boiko, 2023.
- [175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Automatic retrieval of documents and data from projects", 16. november 2023. [Online]. Dostopno na: [https://www.youtube.com/watch?v=ASXolti\\_YPs&t](https://www.youtube.com/watch?v=ASXolti_YPs&t). [Datum dostopa: 2. marec 2025].
- [176] M. & Company, "Three new mandates for capturing a digital transformation's full value", 22. januar 2022. [Online]. Dostopno prek: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Datum dostopa: 15. februar 2025].
- [177] KPMG, "Gradbeništvo v digitalnem svetu", 1. maj 2021. [Online]. Dostopno na: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Datum dostopa: 5. april 2024].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17. marec 2023. [Online]. Dostopno na: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Datum dostopa: 15. februar 2025].
- [179] O. Business, "Satya Nadella razkriva, kako bodo agenti umetne inteligence razbili modele SaaS",

10. januar 2025. [Online]. Dostopno prek: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Datum dostopa: 15. marec 2025].
- [180] Forbes, "Čiščenje velikih podatkov: najbolj zamudna in najmanj prijetna naloga podatkovne znanosti, pravi raziskava", 23. marec 2016. [Online]. Dostopno prek: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [Datum dostopa: 15. marec 2025].
- [181] Urad Združenega kraljestva za zunanje zadeve, Commonwealth in razvoj, "Strategija digitalnega razvoja 2024-2030", 18. marec 2024. [na spletu]. Dostopno prek: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [Datum dostopa: 15. marec 2025].
- [182] "Vizija in strategija v industriji oblikovanja stavb," 7. november 2003. [Online]. Dostopno na: <https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Datum nagovora: 5. marec 2025].
- [183] M. Bocharov, "Informacijsko modeliranje", marec 2025. [Online]. Dostopno na: <https://www.litres.ru/book/mihail-evgenievich-bocharov/informacionnoe-modelirovaniye-v-rossii-71780080/chitat-onlayn/?page=5>. [Datum objave: 15. marec 2025].
- [184] "Integrated Design-Through-Manufacturing: Benefits and Rationale," 2000. [Online]. Dostopno na: [https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489\\_Benefits\\_of\\_MAI.pdf](https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf). [Datum objave: 25. marec 2025].
- [185] CAD Vendor, "Program Open BIM je marketinška kampanja," 12. marec 2012. [Online]. Dostopno prek: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [Datum objave: 30. marec 2025].

# INDEKS SUBJEKTOV

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 Parquet, 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

Veliki podatki, 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, 367, 372, 376, 383, 389, 402, 403, 405, 406, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

475, 476, 492, 494, 497

BlackBox, 240, 242, 243.

BMS, 8

Bokeh, 320, 337

BOM, 76, 77, 79, 257, 263

Omejevalno polje, 234, 373, 392, 393, 394, 414, 415, 480

BREP, 142, 234, 263, 264, 276, 283, 284

## C

KAD, 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, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 406, 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<sub>2</sub>, 72

CO<sub>2</sub>, 229, 234, 235, 236, 237, 238, 239, 243

COBie, 156, 289, 292.

SE, 56, 168, 169, 170, 477

Kopilot, 114, 116, 459

CPIXML, 143, 272, 273, 276, 277, 279, 285, 296

CPM, 14, 17, 62, 166, 175, 233, 326, 473

CQMS, 14, 84, 177, 178, 462

CRM, 109, 369, 459.

CRUD, 51, 459

CSG, 263

CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168,

268, 272, 280, 329, 333, 344, 354, 354, 356, 373, 377,

378, 379, 380, 401, 407, 414, 474, 480

## D

DAE, 276, 277, 278, 280, 281, 284, 285, 414  
 DAG, 362, 363, 365, 366  
 Dash, 320, 336, 337.  
 Upravljanje podatkov, 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  
 Minimalizem podatkov, 373, 395, 396, 400, 401  
 Data Swamp, 373, 395, 397, 401  
 podatkovno skladišče, 382, 383, 400  
 Podatki kot storitev, 487  
 na podlagi podatkov, 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  
 Izvleček, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

## F

Perje, 62, 123, 330, 378  
 FPDF, 339, 340, 341, 342, 343

## G

GDPR, 109  
 GIS, 58  
 GLTF, 143, 278  
 Google preglednice, 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, 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  
 beležnica Jupyter, 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  
**obremenitev**, 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  
**Nizka koda**, 368, 369

## M

**Matplotlib**, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494  
**MCAD**, 77, 78, 257, 284  
**MEP**, 14, 175, 492  
**MREŽE**, 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  
**No-Code**, 368, 369  
**NURBS**, 142, 282, 283, 284, 285  
**NWC**, 8, 276

## O

**OBJ**, 143, 273, 276, 277, 278, 280, 281, 284, 285  
**OCCT**, 273  
**OCR**, 69, 128, 132, 134  
**OmniClass**, 154, 155, 156.  
**odprt BIM**, 142, 216, 219, 256, 261, 278, 291  
**Odprta koda**, 45, 55, 97, 98, 107, 108, 273, 275, 494  
**SOVA**, 267, 268, 269.

## P

**Pande**, 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  
**Cevovod**, 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  
**zasebni kapital**, 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  
**RTV**, 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  
OBSEG, 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  
KORAK, 261, 262, 263, 266, 272, 276, 277, 292  
Streamlit, 336, 337  
SVF, 142, 276, 285

## T

Transformacija, 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, 291, 299, 329, 378, 414, 474