



2. VÄLJAANNE
UUENDATUD JA
LAIENDATUD

DATA DRIVEN CONSTRUCTION

ANDMEAJASTUL NAVIGEERIMINE
EHITUSTÖÖSTUSES

TEHISINTELLEKT JA LLM KOOS
KASUTUSJUHTUMITEGA

Artem Boiko



100+

KEY DATA-RELATED TOPICS



50+

PRACTICAL PROBLEMS SOLVED



300+

UNIQUE ILLUSTRATIONS AND GRAPHS

DATA-DRIVEN CONSTRUCTION

Navigeerimine andmeajastul
ehitussektoris

Teine, läbivaadatud ja täiendatud väljaanne

ARTEM BOIKO



"Boiko on IT James Carville - viimase palju tsiteeritud "It's the economy, stupid", selle kuulsa raamatu puhul tuleb vaid üks sõna välja vahetada. "It's the data, stupid." (Ja et leida oma tee andmeuniversumis, kehtib tänaseni kreeka keelest pärit vanade roomlaste ütlus: "Navigare necesse est". Autor navigeerib oma lugejaid läbi kõigi andmemere sügavuste ja madalike kindla käe ja vankumatu kompassiga, rääkimata põhjalikust ajaloolisest lähenemisest ja, mis pole sugugi vähem tähtis, väga originaalsest graafikast ning heast huumorimeelest, mis ei ilmne alles teisel pilgul. Rahvusvaheline vastukaja Boiko raamatule ulatub eufoorilisest heakskiidust üsna sapise skeptitsismini, mis on raamatu teisele saksakeelsele väljaandele hästi mõjunud. Boiko on originaalne ja jaogmaatiline andmemõtteleja. Ta esitab lugejale põnevaid teadmisi ja alati julgeid, isegi provokatiivseid teese, mis inspireerivad edasi mõtlema. Suurepärane ravim saksa latentse konsensususe haiguse vastu. Muide, üldtõõdud ladinakeelsel vanasõnal on täiendus: "vivere non est necesse". See ei kehti Boiko lähenemise kohta andmete maailmale - andmed elavad ja nende elu on vajalik, et mitte öelda ülioluline."

- Dr. Burkhard Talebitari, vabakutseline toimetaja - sealhulgas ajakirja: BIM, mida Ernst & Sohn annab alates 2013. aastast igal aastal välja.

“

"Artem Boiko raamat on verstapostiks digiteerimise demokrati-seerimisel ehitussektoris - ja tõeliseks mängumuutjaks väikeste ja keskmise suurusega ettevõtete (VKEd) jaoks. Eriti murrangu-line: kasutades kaasaegseid avatud lähtekoodiga madala koodiga ja koodita vahendeid, saavad ettevõtted juba praegu tõhusalt integreerida andmeid oma äriprotsessidesse ja neid ka-sumlikult analüüsida - ilma süvendatud programmeerimisoskus-teta. See muudab tülivate kommertslike tarkvarapakettide kalli kasutamise üleliigseks. See raamat on üleskutse tegutseda! See on väärtuslik juhend kõigile, kes ei taha mitte ainult mõista ehi-tussektori digitaalset muutust, vaid ka seda aktiivselt kujundada - pragmaatiliselt, tõhusalt ja tulevikku suunatud viisil. Nüüd on aeg teha koostööd, et jagada neid teadmisi ja suurendada jät-kusuutlikult ehitussektori tootlikkust."

-**Dr. Michael Max Buehler**, HTWG Konstanz'i ehitusjuhtimise professor, GemeinWerk Ventures'i kaasomanik ja Devv-Stream'i sõltumatu direktor.

“

"DataDrivenConstruction" raamat on üks esimesi samme, mis ületab ehitajate tavapärase maailma piire, nende keeruliste pro-jekteerimis- ja juhtimissüsteemidega, kui tundub, et andmete keerukus ja küllastumine ei anna isegi võimalust radikaalseks lihtsustamiseks ja töö läbipaistvuse suurendamiseks ehitusand-metega. Artem näitab oma raamatus lihtsas keeles, millised või-malused avab meie ees kaasaegne andmetega töötamise tehnoloogia, ja annab sõna otseses mõttes konkreetseid samme, mida saab oma töös kohe rakendada. Ma kutsun kõiki, kes tahavad mõista, kuhu automatiseerimissüsteemid ehituses arenevad, tungivalt üles seda raamatut hoolikalt uurima, et mõista, et andmerevolutsioon ehituses koputab juba meie uk-sele. Praegu pakub see huvi vaid geekidele, kuid mõne aasta pärast, nagu BIM, on sellised lähenemisviisid ja tarkvara kõikjal olemas!"

- **Ihor Rogachew**, IMT kompetentsikeskuse juht, BIM ja digi-taalne ümberkujundamine RGD-s ning InfraBIM.Pro asutaja.

“

"Ma soovitan väga DataDrivenConstruction raamatut, mis käsit-leb, nagu pealkiri ütleb, andmepõhist teabehalduse lähenemis-viisi AECO jaoks. Kasutan seda praegu, et aidata alkatada mit-meid arutelusid erinevate rühmadega. Olen leidnud, et see on väga kättesaadav viide. Lisaks põhjalikule ülevaatele AECO va-hendite ajaloost, andmetest ja mitmete võtmetehnoloogiate tut-vustamisele sisaldab raamat mitmeid väga kasulikke skeeme, mis kirjeldavad andmeallikate ja lõppkasutaja artefaktide ula-tust koos näidistööprotsessidega. Mulle tundub, et just selliseid skeeme on meil teabestrategiate arendamisel ja jälgimisel roh-kem vaja ja need aitavad kaasa BEP'ile - määratledes ettevõtte üldise andmemudeli, millele saab peale panna PIMi ja AIMi piiri."

- **Paul Ransley**, Acmena peakonsultant ja Londoni transpordi-ameti süsteemide integreerimise insener.

“

"Kui "andmed on uus nafta", siis peame õppima neid määrat-lema, leidma, kaevandama ja rafineerima, et muuta need väär-tuslikuks. Minu arvates on raamat "DataDrivenConstruction" väga informatiivne ja sisukas. Raamat pakub kasulikku ajaloo-list tausta ja selgitab tööd andmetega lihtsas keeles. Neile, kes on huvitatud digitaalsest transformatsioonist, annab see hea a-rusaama andmetest - kuidas need toimivad, kuidas need on struktureeritud ja kuidas neid saab kasutada."

- **Ralph Montague**, ArcDoxi direktor, BIM koordinaatorite tipp-kohtumise direktor ja BIMi riikliku peegelkomitee esimees Iiri-maa riiklikus standardimisasutuses.

“

"Nagu raamatus rõhutati, on teave ehitussektori jaoks väga olu-line vara ning selle kättesaadavus lihtsustab oluliselt täpsete otsuste tegemist ja kiirendab projektide ajakava. Raamat pakub neutraalset ja tõhusat lähenemist sellele allikale juurdepääsuks ja selle ärakasutamiseks otsuste tegemisel. Raamatus esitatud metoodika kasutab kaasaegset lähenemist, mis ühendab tehi-sintellektipõhise programmeerimise ja juurdepääsetavav avatud lähtekoodiga tööriistad. Kasutades tehisintellekti võim-sust ja kasutades avatud lähtekoodiga tarkvara, on metoodika eesmärk suurendada automatiseerimist, optimeerida protsesse ning edendada ligipääsetavust ja koostööd selles valdkonnas. Raamatu keelekasutus on selge ja kergesti jälgitav."

- **Dr. Salih Ofluoğlu**, Antalya Bilimi ülikooli kunsti- ja arhitek-tuuriteaduskonna dekaan ja Euraasia BIM-foorumi korraldaja.

“

"Kõik, mida ma võin öelda, on WOW! See, kuidas te ühendasite ajalugu, LLM, graafika ja üldine lihtsus mõista oma punkte on tõeliselt tähelepanuväärne. Raamatu voolavus on hämmastav. Selles raamatus on nii palju geniaalseid aspekte; see on tõeliselt mängu muutev. See on suurepärane teabeallikas ja ma kiidan teid selle eest, kui palju vaeva ja kirge te olete sellesse pannud. Palju õnne sellise tähelepanuväärse teose loomise eest. Ma võiksin jätkata, kuid piisab, kui ütlen, et olen usumatult muljetavaldav!"

- **Natasha Prinsloo**, Energylab_i digitaalse praktika juht.

“

"Kõigile ehitussektori töötajatele, alates algajaist kuni kogenud professionaalideni, on see raamat mängude muutja! See ei ole teie tüüpiline tolmune lugemine - see on täis teadmisi, strateegi-aid ja pisut huumorit, mis hoiab teid tegevuses. See hõlmab and-mete kasutamise arengut ehituses alates iidsetest andmete sal-vestamise meetoditest kuni tipptasemel digitaal tehnoloogia-teni. See on nagu ajamasinaga reisimine läbi ehitusandmete evolutsiooni. Olenemata sellest, kas olete arhitekt, insener, pro-jektijuht või andmeanalüütik, see põhjalik juhend muudab revo-lutsiooniliselt teie lähenemise projektidele. Valmistuge protsesside optimeerimiseks, otsuste tegemise tõhustamiseks ja projektide juhtimiseks nagu kunagi varem!"

- **Pierpaolo Vergati**, Rooma Sapienza Ülikooli lektor ja Fintecna ehitusprojektijuht.

“

"Lugesin raamatu läbi ühe hingetõmbega, vähem kui 6 tunniga. Raamatu tootmiskvaliteet on suurepärane, tihe läikiv paber, värvitoonid, meeldiv kirjatüüp. Suur hulk praktilisi näiteid selle kohta, kuidas töötada LLMiga konkreetselt ehitussektoris, säästab teid kuude, kui mitte aastatepikkusest iseõppimisest. Töö näited on väga mitmekesised, alates lihtsatest kuni keeruliste näidetenäideni, ilma et peaksite ostma keerulist ja kallist tarkvara. Raamat võimaldab iga ehitussektori ettevõtte omanikel vaadata värskest oma äristrateegiat, digitaliseerimist ja arenguperspektiive. Ja väiksematel ettevõtetel suurendada tõhusust taskukohaste ja tasuta tööriistade abil."

- **Mihhail Kosarev**, TIM-ASG lektor ja konsultant ehitussektori digitaalse ümberkujundamise alal.

“

"DATA DRIVEN CONSTRUCTION" raamat, mis muudab mängu kõigi jaoks, kes on uudishimulikud, kuhu ehitussektor andmete ajastul liigub. Artem ei kriipsuta lihtsalt pinda; ta süveneb sügavale praegustesse arengutesse, väljakutsetesse ja paljulubavatesse võimalustesse ehituses. Selle raamatu teeb eriliseks selle ligipääsetavus - Artem selgitab keerulisi ideid, kasutades arusaadavaid analoogiaid, mis muudavad sisu kergesti mõistetavaks. Minu arvates on raamat usumatult informatiivne ja samas kõitev. Kokkuvõttes on Artem koostanud väärtusliku abimaterjali, mis mitte ainult ei teavita, vaid ka inspireerib. Olenevata sellest, kas olete kogenud spetsialist või ehituslane uustulnuk, see raamat avarab teie vaatenurka ja süvendab teie arusaamist sellest, kuhu tööstus liigub. Väga soovitatav!"

- **Moayad Saleh**, arhitekt ja BIM rakendamise juht TMM GROUP Gesamtplanungs GmbH-s.

“

"Pean ütlema, et "Andmepõhine ehitus" on väärt, et seda õpetataks õpikuna ülikoolides, ja see on raamat, mis annab väärtusliku panuse arenevasse BIM-valdkonda. Data-Driven Construction raamat sisaldab tehnilist sõnastikku, mis selgitab mõisteid väga hästi. Äärmiselt raskesti seletatavad teemad on väga ilusa visuaalse keelega lihtsaks ja arusaadavaks tehtud. Arvan, et see, mida tahetakse selgitada visuaalselt, tuleb lugejale ka lühidalt väljendada. Mõne visuaali arusaadavus ehk teisisõnu, visuaali lugemine nõuab eraldi teavet. Tahaksin ka öelda, et tutvustan Artem Boiko väärtuslikke töid hea meelega oma loengutes ja seminarides ülikoolides."

- **Dr. Ediz Yazicioglu**, ArchCube'i omanik ja Istanbuli Tehnikaülikooli ja Medipoli Ülikooli arhitektuuriosakonna ehitusprojekti juhtimise lektor.

“

"Andmepõhine ehitus" annab elavalt edasi ehitusandmetega tehtava teabepõhise töö põhitõed. Raamat, mis käsitleb infovooge ja majanduslikke põhimõisteid ning eristub seega teistest BIM raamatutest, sest see ei esinda mitte ainult tarkvaratootja vaatenurka, vaid püüab edasi anda ka põhimõisteid. Raamat, mida tasub lugeda ja näha."

- **Jakob Hirn**, Build Informed GmbH tegevjuht ja kaasasutaja ning innovatsioonifoorumi "On Top With BIM" algataja.

“

"Andmed on uus nafta", nagu öeldakse, nii et selle otsijatel või kaevuritel peaksid olema õiged vahendid ja mõtteviis, et sellest 21. sajandi ressursist väärtust ammutada. Ehitustööstus on liiga kaua olnud "3D-teabele" põhinevate protsesside libedal teel, mille puhul projekti elluviimine põhineb kellegi teise küpsetatud teabel (nt nad on juba joonistanud piruka või tulpdiagrammi), samas kui selle aluseks olevad "andmed" (nt töötlemata tabel) suudavad pakkuda palju rohkem, eriti kuna andmete liitmine ja tehisintellekt pakuvad piiramatuid võimalusi. Kui te toimetate (või õpetate/teadete) ehitust, on see raamat teie parim - ja seni ainus - abimaterjal, et orienteeruda andmetel põhinevas maailmas, mille me oleme leidnud."

- **Dr Zulfikar Adamu**, ehituslase strateegilise IT dotsent LSBUs, Ühendkuningriik.

“

Artem Boiko "Data-Driven Construction" on muljetavaldav teos, mis pakub ehitussektorile kindlaid aluseid pidevalt kasvavate tehnoloogiate ja infovõimaluste ajastul. Boiko suudab keerulisi teemasid arusaadavalt esitada, tutvustades samas ka visioonilisi ideid. Raamat on hästi läbimõeldud kogumik, mis mitte ainult ei tõsta esile praeguseid arenguid, vaid annab ka väljavaateid tulevaste uuendusteks. Raamat on väga soovitatav kõigile, kes soovivad tutvuda andmepõhise ehituse planeerimise ja teostamisega."

- **Markus Eiberger**, Stuttgarti rakenduskõrgkooli õppejõud, Konstruktionsgruppe Bauen'i vanemprojektijuht ja filiaalijuhi asetäitja, BIM-klastri Baden-Württembergi assotsiatsiooni juhatuse liige.



Teine väljaanne, märts 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-02-7



Artem Boiko Copyright

boikoartem@gmail.com

info@datadrivenconstruction.io

Käesoleva raamatu ühtegi osa ei tohi reprodutseerida ega edastada mis tahes kujul või viisil, elektrooniliselt või mehaaniliselt, kaasa arvatud fotokopeerimine, salvestamine või mis tahes teabe salvestus- ja otsingusüsteem, ilma autori kirjaliku loata - välja arvatud mitteärilistel eesmärkidel muutmata kujul levitamiseks. Raamatut levitatakse tasuta ja seda võib vabalt jagada teiste kasutajatega isiklikel, hariduslikel või teaduslikel eesmärkidel, tingimusel et autorsus ja viited originaalile jäävad alles. Autorile jäävad kõik mitteomandiõigused tekstile ja ta ei anna mingeid otseseid ega kaudseid garantiisid. Raamatus mainitud ettevõtted, tooted ja nimed võivad olla väljamõeldud või kasutada neid näidetena. Autor ei vastuta esitatud teabe kasutamisest tulenevate tagajärgede eest. Raamatus sisalduv teave on esitatud "nagu on", ilma garantiita selle täielikkuse või asjakohasuse kohta. Autor ei vastuta käesolevas raamatus sisalduva teabe, koodi või programmide kasutamisest tuleneva juhusliku või kaudse kahju eest. Käesolevas raamatus esitatud koodinäidised on mõeldud ainult õppimise eesmärgil. Lugejad kasutavad neid omal vastutusel. Autor soovib kõiki tarkvaralahendusi enne tootmiskeskonnas kasutamist testida. Kõik tekstis mainitud kaubamärgid ja tootenimed on vastavate firmade kaubamärgid, registreeritud kaubamärgid või teenusemärgid ning kuuluvad nende omanikele. Nende nimede kasutamine raamatus ei tähenda mingit seost nende omanikega ega nende heakskiitu. Kolmandate osapoolte toodete või teenuste mainimine ei kujuta endast soovist ega tähenda heakskiitu. Näidetes kasutatud ettevõtete ja toodete nimed võivad olla nende omanike kaubamärgid. Lingid kolmandate isikute veebilehtedele on esitatud mugavuse huvides ja ei tähenda, et autor toetab nendel veebilehtedel esitatud teavet. Kõik viidatud statistilised andmed, tsitaadid ja uuringud olid kirjutamise ajal ajakohased. Andmed võivad aja jooksul muutuda.

Seda raamatut levitatakse Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International litsentsi (CC BY-NC-ND 4.0) alusel. Te võite seda kopeerida ja levitada mittekaubanduslikel eesmärkidel, tingimusel, et te säilitate autorsuse ja ei tee muudatusi.



CC BY-NC-ND 4.0

© 2024 Artjom Boiko. Esimene väljaanne.

© 2025 Artjom Boiko. Teine, parandatud ja täiendatud väljaanne.

Kõik õigused kaitstud.

TEISE VÄLJAANDE EESSÕNA

See raamat on professionaalse kogukonnaga peetud elava dialoogi tulemus. See põhineb arvukatel erialastel aruteludel andmehalduse kohta ehitussektoris, mis toimusid erinevatel erialaplatvormidel ja sotsiaalmeediaplatvormidel. Need arutelud said aluseks artiklitele, väljaannetele ja visuaalsetele materjalidele, mis on tekitanud laia vastukaja erialases kogukonnas. Autori sisu kogub igal aastal miljonite vaatamiste arvu erinevatel platvormidel ja keeltes, tuues kokku ehitusvaldkonna digitaliseerimise valdkonna professionaalid.

Aasta jooksul pärast esimese väljaande ilmumist tellisid raamatut eksperdid rohkem kui 50 riigist, Brasiiliast ja Peruust Mauritiuse ja Jaapanini. Raamatu teine väljaanne, mida te nüüd käes hoiate, on ekspertide tagasiside, esimese väljaande kohta esitatud kriitika ja erialaringkondades toimunud arutelude põhjal läbi vaadatud ja täiendatud. Tänu tagasisidele on teist väljaannet oluliselt täiendatud: lisatud on uued peatükid CAD (BIM) tehnoloogiate ja tõhusate ETL -protsesside loomise kohta. Samuti on oluliselt suurendatud praktiliste näidete ja juhtumiuuringute arvu. Eriti väärtuslik on tagasiside ehitussektori juhtidelt, konsultatsioonifirmadelt ja suurtelt IT-ettevõtetelt, kes pöördusid autori poole digitaliseerimise ja koostalitlusvõime küsimustega nii enne kui ka pärast raamatu esimese versiooni ilmumist. Paljud neist on juba rakendanud raamatus kirjeldatud lähenemisviise või kavatsesid seda lähitulevikus teha

Te hoiate käes raamatut, mis on loodud arutelu ja aktiivse arvamuste vahetamise teel. Edusammud sünnivad dialoogis, arvamuste kokkupõrkes ja avatuses uutele lähenemisviisidele. Täna teid, et olete osa sellest dialoogist. Teie konstruktiivne kriitika on tulevaste paranduste alus. Kui tekstis on vigu või kui soovite jagada ideid ja ettepanekuid, on igasugune tagasiside teretulnud. Kontaktandmed on esitatud raamatu lõpus

MIKS ON RAAMAT TASUTA?

See raamat on mõeldud avatud õppevahendina, mille eesmärk on levitada ehitussektori andmehalduse kaasaegseid lähenemisviise. Raamatu esimene versioon oli aluseks erialaseltside kommentaaride ja ettepanekute kogumisele, mis võimaldas parandada materjali struktuuri ja sisu. Kõik kommentaarid, ettepanekud ja ideed on hoolikalt analüüsitud ja lisatud käesolevasse läbivaadatud versiooni. Raamatu eesmärk on aidata ehitusspetsialistidel mõista, kui oluline on töötada andmetega: süstemaatiliselt, teadlikult ja pidades silmas teabe pikaajalist väärtust. Autor on kogunud näiteid, illustatsioone ja praktilisi tähelepanekuid enam kui 10-aastasest tööst ehituse digitaliseerimise valdkonnas. Suurem osa sellest materjalist on sünninud reaalsete projektide, inseneride ja arendajatega peetud arutelude, rahvusvahelistes algatustes osalemise ja koolitusseminaride käigus. Raamat on katse struktureerida kogutud kogemusi ja jagada neid arusaadaval viisil. Kui soovite toetada raamatu ideede edasist levitamist ja saada mugavat formaati lugemiseks, näidete ja visuaalsete materjalidega töötamiseks - saate osta [trüki versiooni](#).

KASUTAMISE ÕIGUSED

Kõiki käesoleva raamatu materjale, illustatsioone ja katkendeid võib reprodutseerida, tsiteerida või kasutada mis tahes vormis ja mis tahes meedias tingimusel, et allikas on märgitud Artem Boiko ja raamatu pealkiri on "Data-Driven Construction" (andmepõhine ehitus). Täname teid töö austamise ja teadmiste levitamise eest.

Pühendan selle siira tänulikkusega oma perekonnale, kes sisendas minusse juba varases eas sügavat armastust ehituse vastu, minu kodusele kaevanduslinnale, kus sain õppida vastupidavust, ja minu maamõõtja abikaasale, kelle vankumatu toetus on olnud minu pidevaks innustajaks.

KELLELE SEE RAAMAT ON MÕELDUD

Kättesaadavas keeles kirjutatud raamat on suunatud paljudele ehitussektori lugejatele - alates üliõpilastest ja algajatest, kes soovivad mõista kaasaegsete ehitusprotsesside põhitõdesid, kuni spetsialistideni, kes vajavad ajakohast metoodikat andmete haldamiseks ehituses. Olenemata sellest, kas olete arhitekt, insener, töödejuhataja, ehitusjuht või andmeanalüütik, pakub see paljude unikaalsete illustratsioonide ja graafikutega põhjalik juhend väärtuslikke teadmisi sellest, kuidas kasutada andmeid äritegevuses protsesside optimeerimiseks ja automatiseerimiseks, otsuste tegemise parandamiseks ja ehitusprojektide juhtimiseks eri tasanditel, kasutades kaasaegseid vahendeid

Raamat on põhjalik juhend, mis ühendab teoreetilised alused ja praktilised soovituselised andmehaldustehnikate integreerimiseks ehitusprotsessidesse. Raamat keskendub teabe strateegilisele kasutamisele, et optimeerida toiminguid, automatiseerida protsesse, parandada otsuste langetamist ja juhtida projekte tõhusalt, kasutades kaasaegseid digivahendeid.

See raamat hõlmab ehitustööstuses teabega töötamise teoreetilisi ja praktilisi aspekte. Üksikasjalike näidete kaudu uuritakse ülesannete parameetrisseerimise, nõuete kogumise, struktureerimata ja mitmeformaadiliste andmete töötlemise ning nende muutmise metoodikat ehitusettevõtete jaoks tõhusateks lahendusteks.

Lugeja läbib järk-järgult tee nõuete kujundamisest ja põhiliste andmemudelite väljatöötamisest kuni heterogeensete teabeallikate integreerimise keerukamate protsessideni, ETL -protsesside loomiseni, teabepiipeliinide ja masinõppe mudelite ehitamiseni. Järjestikune lähenemisviis võimaldab selgelt näidata ehitusettevõtete äriprotsesside ja otsustamist toetavate süsteemide korraldamise ja automatiseerimise mehhanisme. Raamatu iga osa lõpeb praktilise peatükiga, mis sisaldab samm-sammult juhiseid, mis võimaldavad omandatud teadmiste kohest rakendamist reaalsetes projektides.

RAAMATU OSADE KOKKUVÕTE

See raamat on üles ehitatud andmete ümberkujundamise kontseptsioonile väärtusahelas: andmete kogumisest ja kvaliteedi tagamisest analüütilise töötlemise ja väärtuslike praktiliste lahenduste leidmiseni, kasutades kaasaegseid vahendeid ja metoodikaid.

1. osa: Digitaalne areng ehituses - jälgib andmehalduse ajaloolist muutumist savitahvlitelt kaasaegsete digitaalsüsteemideni, analüüsides moodulsüsteemide tekkimist ja teabe digitaliseerimise kasvavat tähtsust tööstusrevolutsiooni kontekstis.

2. osa: Ehitustööstuse infoprobleemid - uurib andmete killustatuse, teabesilode, HiPPO lähenemisviisi mõju otsuste tegemisele ja konfidentsiaalsete vormingute piiranguid, soovitades kaaluda üleminekut tehisintellekti ja elukestva õppe ökosüsteemidele.

3. osa: andmete süstematiseerimine ehituses - moodustab ehitusandmete tüpoloogiat, kirjeldab nende korraldamise meetodeid, integreerimist ettevõtte süsteemidega ja käsitleb pädevuskeskuste loomist infoprotsesside standardiseerimiseks.

4. osa: Andmete kvaliteedi tagamine - näitab meetodeid, kuidas muuta ebaühtlane teave kvaliteetseteks, struktureeritud andmeteks, sealhulgas andmete ekstraheerimine erinevatest allikatest, valideerimine ja modelleerimine LLMi abil.

5. osa: Kulu- ja ajakuluarvutused - käsitleb kulude ja planeerimisarvutuste digitaliseerimist, mahtude saamise automatiseerimist CAD-mudelitest (BIM), 4D-8D modelleerimistehnoloogiat ja ehitusprojektide ESG-arvutusi.

6. osa: CAD ja BIM - analüüsib kriitiliselt projekteerimistehnoloogiate arengut, süsteemide koostalitlusvõime probleeme, suundumusi avatud andmeformaate suunas ning tehisintellekti rakendamise väljavaateid projekteerimisel.

7. osa: andmeanalüüs ja automatiseerimine - käsitletakse teabe visualiseerimise põhimõtteid, peamisi tulemusnäitajaid, ETL-protsesse, töövoogude orkestreerimise vahendeid ja keelemudelite rakendamist rutiinsete ülesannete automatiseerimiseks.

Osa 8: Andmete säilitamine ja haldamine - uurib andmete säilitamise vorminguid, andmelao ja andmejärve kontseptsioone, andmehalduse põhimõtteid ja uusi lähenemisviise, sealhulgas vektorandmebaase ning DataOps- ja VectorOps-metoodikaid.

9. osa: suured andmed ja masinõpe - keskendub üleminekule objektiivsele analüüsile, mis põhineb ajaloolistel andmetel, asjade internetile ehitusplatsidel ning masinõppe algoritmide rakendamisele projekti kulude ja tähtaegade prognoosimiseks.

10. osa: Ehitustööstus digitaalsete andmete ajastul - tutvustab ehitustööstuse tulevikku, analüüsides üleminekut põhjuslikust analüüsist korrelatsioonidega töötamisele, ehituse "Uberiseerimise" kontseptsiooni ja digitaalse ümberkujundamise strateegiaid.

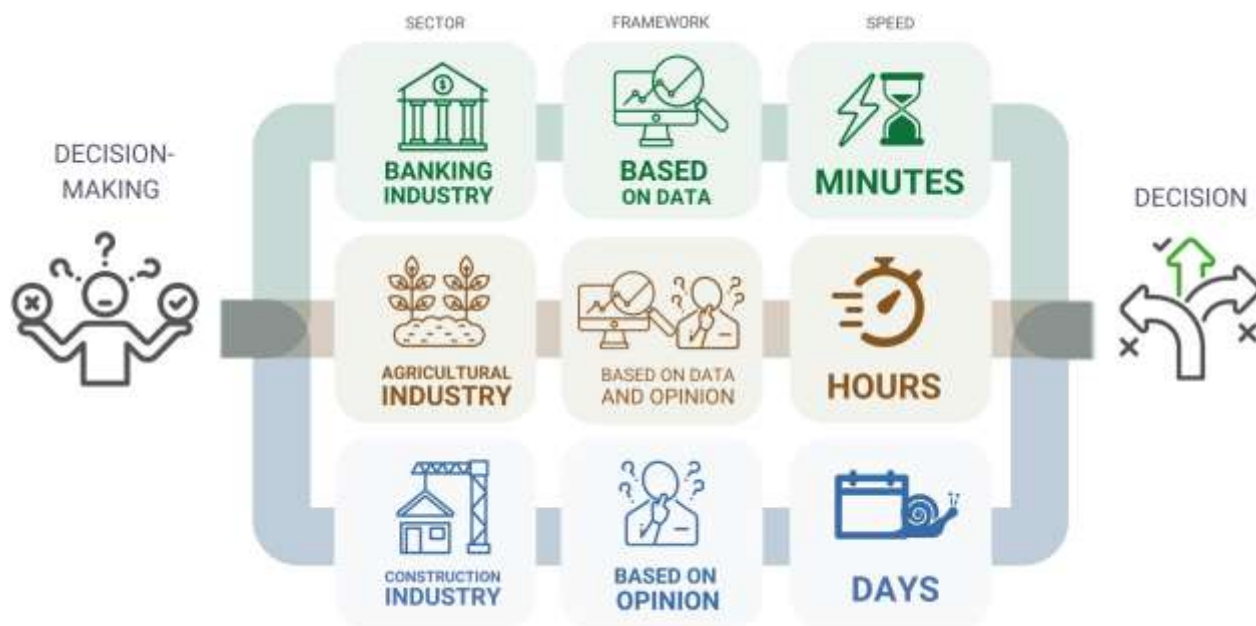
What is meant by **data-driven construction** ?



SISSEJUHATUS

Kui kaua suudab teie ettevõtte jääda konkurentsivõimeliseks maailmas, kus tehnoloogia areneb kiiresti ja iga äriaspekt, alates ajastamisest ja kuluarvestusest kuni riskianalüüsini, automatiseeritakse masinõppe mudelite abil?

Ehitustööstus, mis on eksisteerinud sama kaua kui inimkond ise, on revolutsiooniliste muutuste lävel, mis lubavad täielikult muuta meie mõtteviisi traditsioonilisest ehitusest. Juba teistes majandussektorites ei muuda digitaliseerimine lihtsalt reegleid, vaid tõrjub turult halastamatult välja ettevõtted, kes ei ole suutnud kohaneda uue andmetöötluskeskkonnaga ega suuda parandada otsuste tegemise kiirust (joonis 1).



Joonis 1 Ehitustööstuses sõltub otsuste tegemise kiirus inimtegurist sagedamini kui teistes tööstusharudes.

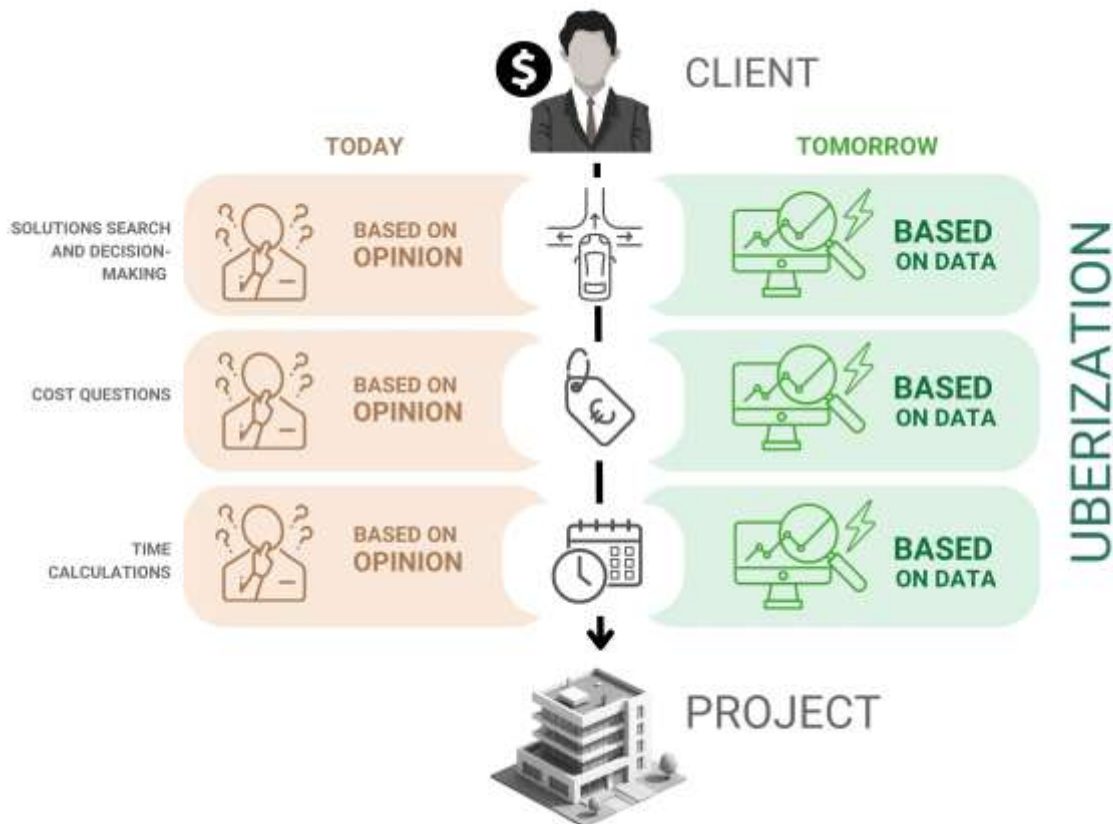
Pangandus, jaekaubandus, logistika ja põllumajandus on kiiresti liikumas täieliku digitaliseerimise suunas, kus ebatäpsustel ja subjektiivsetel arvamustel ei ole enam kohta. Kaasaegsed algoritmid suudavad analüüsida tohutuid andmehulki ja pakkuda klientidele täpseid prognoose - olgu selleks siis laenu tagasimaksmise tõenäosus, optimaalsed tarneteed või riskiprognosid.

Ehitus on üks viimaseid tööstusharusid, kus on toimunud vältimatu üleminek kõrgepalgaliste spetsialistide arvamustel põhinevatelt lahendustelt andmepõhiste lahendustele. See üleminek ei ole tingitud mitte ainult uuest tehnoloogilistest võimalustest, vaid ka turu ja klientide suurenenud nõudmistest läbipaistvuse, täpsuse ja kiiruse järele.

Robotiseerimine, protsesside automatiseerimine, avatud andmed ja nendel põhinevad prognoosid - kõik see ei ole enam ainult võimalus, vaid paratamatus. Enamik ehitussektori ettevõtteid, mis veel hiljuti vastutasid tellija ees projektide mahu, maksumuse, aja ja kvaliteedikontrolli eest, on nüüd oht, et

nad muutuvad pelgalt tellimuste täitjateks, kes ei tee olulisi otsuseid (joonis 2).

Tänu arvutusvõimsuse arengule, masinõppe algoritmidele ja demokraatlikule juurdepääsule andmetele on nüüd võimalik automaatselt kombineerida eri allikatest pärinevaid andmeid, mis võimaldab sügavamat protsessianalüüsi, riskide prognoosimist ja kulude optimeerimist ehitusprojekti arutelufaasis. Need tehnoloogiad võivad radikaalselt parandada tõhusust ja vähendada kulusid kogu sektoris.



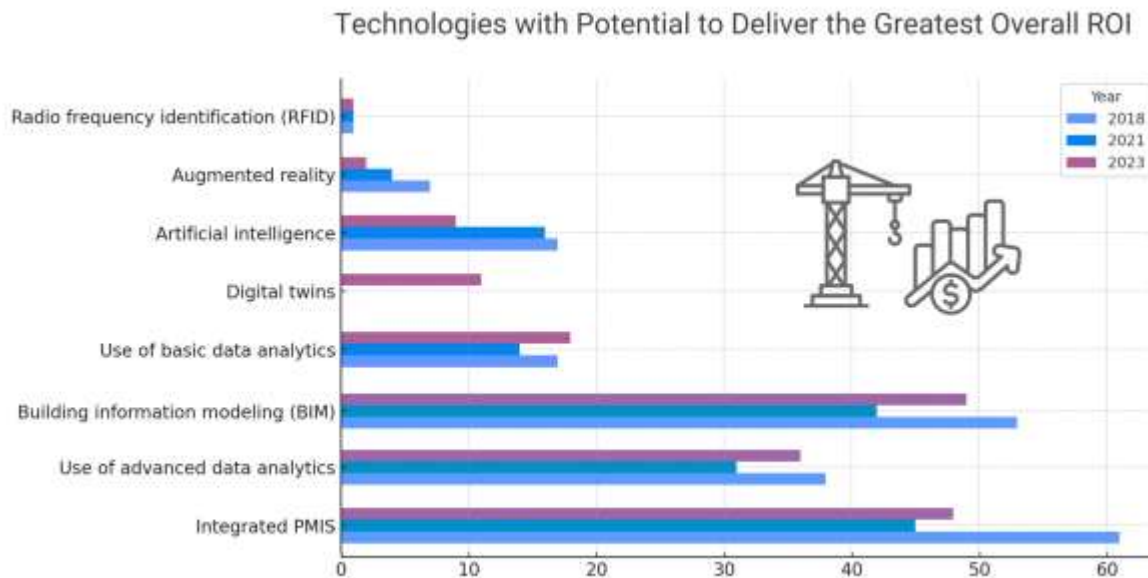
Joonis 2 Klient ei ole huvitatud liigsetest inimteguritest oma projekti elluviimisel.

Hoolimata kõigist uute vahendite ja kontseptsioonide eelistest on ehitussektor uute tehnoloogiate kasutuselevõtu osas teistest majandussektoritest kaugelt maha jäänud.

IT Metrics Key Data 2017 aruande kohaselt on ehitussektor IT-kulutuste poolest 19 tööstusharu seas viimasel kohal [1].

Andmemahu kiire kasv ja protsesside keerukus on muutumas ettevõtte juhtkonnale peavaluks ning peamine probleem uute tehnoloogiate kasutamisel on see, et andmed jäävad vaatamata nende rohkusele killustatuks, struktureerimata ja sageli erinevate süsteemide ja tarkvaratoodete vahel kokkusobimatuks. Seetõttu on paljud ehitussektori ettevõtted praegu eelkõige mures andmete kvaliteediprobleemide pärast, mida saab lahendada ainult tõhusate, automatiseeritud juhtimis- ja analüüsisüsteemide rakendamisega.

KPMG® poolt 2023. aastal ehitusjuhtide seas läbi viidud uuringu [2] kohaselt on projektijuhtimise infosüsteemidel (PMIS), täiustatud ja põhilistel andmeanalüüsidel ja ehitusinfo modelleerimisel (BIM) suurim potentsiaal projektide tasuvuse parandamiseks (joonis 3).



Joonis 3 Ehitusettevõtete juhtide seas läbi viidud küsitlus: millised tehnoloogiad tagavad investeerimisprojektide puhul suurima investeeringutasuvuse (ROI)? (materjalide [2] põhjal).

Andmete äriprotsessidesse integreerimisega seotud probleemide lahenduseks on tagada kvaliteetne teave, kasutada sobivaid andmeformaate ja rakendada tõhusaid meetodeid andmete loomiseks, säilitamiseks, analüüsimiseks ja töötlemiseks.

Andmete väärtuse mõistmine sunnib erinevaid tööstusharusid eemalduma siilustatud rakendustest ja keerulistest bürokraatlikest juhtimisstruktuuridest. Selle asemel keskendutakse uute lähenemisviiside loomisele infoarhitektuurile, muutes ettevõtte kaasaegseteks andmepõhisteks ettevõteteks. Varem või hiljem astub ka ehitussektor ise selle sammu, liikudes järkjärgulise digitaalse arengu asemel tõelise digitaalseks revolutsiooniks, mis mõjutab kõiki ettevõtteid.

Üleminek andmepõhistele äriprotsessidele ei ole lihtne. Paljud ettevõtteid seisavad silmitsi probleemidega, sest juhid ei mõista alati, kuidas kasutada kaootilisi andmekogumeid tõhususe ja ärikasvu parandamiseks.

Selles raamatus süvenetakse andmete maailma, kus teave on muutumas oluliseks strateegiliseks ressursiks, mis määrab äriprotsesside tõhususe ja jätkusuutlikkuse. Seoses teabe kiire kasvuga seisavad ettevõtteid silmitsi uute väljakutsetega. Digitaalne ümberkujundamine ei ole enam lihtsalt moesõna - see on muutumas hädavajalikuks.



Joonis 4 Andmed ja protsessid on ehituse alus.

Ümberkujundamise mõistmine tähendab võimet selgitada keerulist asja lihtsate sõnadega. Seepärast on raamat kirjutatud arusaadavas keeles ja sellega on kaasas autori illustatsioonid, mis on loodud spetsiaalselt selleks, et seletada võtmemõisteid arusaadavalt. Need diagrammid, graafikud ja visualiseeringud on loodud selleks, et kõrvaldada tõkked arusaamise eest ja muuta materjal arusaadavaks ka neile, kes varem pidasid selliseid teemasid liiga keeruliseks. Kõik selles raamatus olevad illustatsioonid, diagrammid ja graafikud on loodud autori poolt ja spetsiaalselt selleks, et visualiseerida tekstis kirjeldatud põhimõisteid.

Üks pilt ütleb rohkem kui tuhat sõna [3].

- Fred R. Barnard, inglise illustraator, 1927.

Teooria ja praktika ühendamiseks kasutame tehisintellekti vahendeid (eelkõige keelemudeleid), mis võimaldavad teil arendada lahendusi, ilma et teil oleks vaja sügavaid programmeerimisalaseid teadmisi. Kui olete orienteeritud praktilisele materjalile ja teid huvitab rohkem praktiline töö andmetega, võite esimese sissejuhatava osa vahele jätta ja minna otse raamatu teise osa juurde, kus algab konkreetsete näidete ja juhtumite kirjeldamine.

Ärge pange siiski liigseid ootusi AI (tehisintellekt), masinõppe ja LLM (Large Language Models) vahenditele üldiselt. Ilma kvaliteetsete sisendandmete ja teemat põhjalikult tundmata ei suuda isegi kõige arenenumad algoritmid anda usaldusväärseid ja sisukaid tulemusi.

Microsofti tegevjuht Satya Nadella hoiatab 2025. aasta alguses tehisintellekti muli ohu eest [4], võrreldes praegust hüpet dot-com-mulliga. Ta rõhutab, et väited AGI (Artificial General Intelligence) verstapostide saavutamise kohta ilma nõuetekohase põhjendamiseta on "mõttetu manipuleerimine mõõdikutega". Nadella usub, et tehisintellekti tegelikku edu tuleks mõõta pigem selle panuse kaudu ülemaailmsesse SKP kasvu kui liigse rõhuasetusega buzzwordidele.

Kõigi uute tehnoloogiate ja kontseptsioonide taga on keeruline ja vaearikas töö andmete kvaliteedi tagamiseks, äriprotsesside parameetriseerimine ja vahendite kohandamine tegelikele ülesannetele.

Andmepõhine lähenemisviis ei ole toode, mida saab lihtsalt alla laadida või osta. See on strateegia, mis tuleb üles ehitada. See algab olemasolevate protsesside ja probleemide värske vaatlusega ning nõuab seejärel distsiplineeritud liikumist valitud suunas.

Juhtivad tarkvaraarendajad ja rakenduste müüjad ei ole ehitussektori muutuste mootoriks, sest paljudele neist on andmepõhine lähenemine oht nende väljakujunenud ärimudelile.

Teised tööstusharud [erinevalt ehitusest], nagu näiteks autotööstus, on juba läbinud radikaalseid ja murrangulisi muutusi ning nende digitaalne ümberkujundamine on hästi edenenud. Ehitusettevõtted peavad tegutsema kiiresti ja otsustavalt: osavamad ettevõtted saavad tohutut kasu, samas kui need, kes kõhklevad, satuvad tõsisesse ohtu. Mõelge murrangule, mida digitaalne fotograafia on selles tööstusharus põhjustanud [5].

- Maailma Majandusfoorumi aruanne "Shaping the Future of Construction", 2016.

Need ettevõtted, kes mõistavad õigeaegselt uue lähenemisviisi võimalusi ja eeliseid, saavad püsiva konkurentsieelise ning suudavad areneda ja kasvada sõltumata suurte tarnijate lahendustest.

See on teie võimalus mitte ainult mitte tulla vastu tulevasele teabe digitaliseerimise tormile, vaid võtta see ka enda kontrolli alla. Sellest raamatust leiate mitte ainult tööstuse hetkeseisu analüüsi, vaid ka konkreetseid soovitusi oma protsesside ja ettevõtte ümbermõtestamiseks ja ümberkorraldamiseks, et saada uue ehitusajastu liidriks ja parandada oma töökogemust.

Ehituse digitaalne tulevik ei tähenda mitte ainult uute tehnoloogiate ja programmide kasutamist, vaid ka andmete käitlemise ja ärimudelite põhjalikku ümbermõtestamist.

Kas teie ettevõtte on selleks strateegiliseks muutuseks valmis?

SISUKORD

SISSEJUHATUS.....	1
SISUKORD.....	1
II OSA SAVITAHVLITEST KUNI DIGITAALSE REVOLUTSIOONINI: KUIDAS TEAVE ON ARENENUD EHITUSES	2
PEATÜKK 2.1. ANDMETE KASUTAMISE ARENG EHITUSSEKTORIS	3
Andmeajastu sünni ehituses.....	3
Alates savist ja papüürusest kuni digitaaltehnoogiiani	4
Protsess kui andmepõhise kogemuse vahend.....	5
Ehitusprotsesside digitaliseerimine	7
PEATÜKK 2.2. TEHNOLOOGIAD JA JUHTIMISSÜSTEEMID KAASAEGSES EHITUSES	11
Digitaalne revolutsioon ja modulaarsete MRP/ERP -süsteemide teke.	11
Andmehaldussüsteemid: andmekaevandamisest äritegevuse väljakutseteni	13
Ettevõtte mütseel: kuidas andmed on seotud äriprotsessidega	17
PEATÜKK 2.3. DIGITAALNE REVOLUTSIOON JA ANDMETE PLAHVATUSLIK KASV	20
Andmabuumi algus kui evolutsioonilaine	20
Kaasaegses ettevõttes tekkivate andmete hulk.....	22
Andmesalvestuse maksumus: majanduslik aspekt	23
Andmete kogumise piirid: massist tähenduseni	25
Järgmised sammud: andmete teooriast praktiliste muutusteni	27
III OSA KUIDAS EHITUSETTEVÕTLUS UPPUB ANDMEKAOSSESSE	28
PEATÜKK 3.1. ANDMETE KILLUSTATUS JA SILOD	29
Mida rohkem vahendeid, seda tõhusam on äri?	29
Andmesilod ja nende mõju ettevõtte tulemuslikkusele	31
dubleerimine ja andmete kvaliteedi puudumine kui üksmeelsuse puudumine.....	34
HiPPO ehk arvamuste oht otsuste tegemisel.....	36
Äriprotsesside keerukuse ja dünaamilisuse pidev suurenemine	39
Neljas tööstusrevolutsioon (tööstus 4.0) ja viies tööstusrevolutsioon (tööstus 5.0) ehituses. ...	42
PEATÜKK 3.2. KAOSE MUUTMINE KORRASTAMISEKS JA KEERUKUSE VÄHENDAMINE.....	45
Üleliigne kood ja suletud süsteemid, mis takistavad tootlikkuse parandamist.....	45

Siiludest ühtse andmelattu.....	47
Integreeritud salvestussüsteemid võimaldavad üleminekut tehisintellekti agentidele	48
Andmete kogumisest otsuste tegemiseni: tee automatiseerimiseni.....	51
Järgmised sammud: kaose muutmine hallatavaks süsteemiks	53
IV OSA ANDMERAAMISTIK EHTUSALASTES ÄRIPROTSESSIDES	55
PEATÜKK 4.1. ANDMETÜÜBID EHTUSES	56
Kõige olulisemad andmetüübid ehitussektoris	56
Struktureeritud andmed.....	60
Relatsioonilised andmebaasid RDBMS ja SQL päringukeel.....	62
SQL - päringud andmebaasides ja uued suundumused.....	64
Struktureerimata andmed	66
Tekstiandmed: struktureerimata kaose ja struktureeritud kaose vahel y	67
Poolstruktureeritud ja vabalt struktureeritud andmed	68
Geomeetrilised andmed ja selle rakendamine.....	69
CAD andmed: projekteerimisest andmete säilitamiseni.....	72
BIM (BOM) kontseptsiooni tekkimine ja CAD kasutamine protsessides.....	75
PEATÜKK 4.2. ANDMETE ÜHTLUSTAMINE JA STRUKTUREERIMINE	81
Süsteemide täitmine andmetega ehitussektoris	81
Andmete ümberkujundamine: kaasaegse ärianalüüsi kriitiline alus	84
Andmemudelid: seosed andmetes ja elementide vahelised seosed	87
Omane formaat ja selle mõju digitaalsetele protsessidele.....	91
Avatud formaadid muudavad lähenemisviisi digitaliseerimisele	95
Paradigmamuutus: avatud lähtekood kui tarkvaratootjate domineerimise ajastu lõppemine.	96
Struktureeritud avatud andmed: digitaalse ümberkujundamise alus.....	99
PEATÜKK 4.3. LLM JA NENDE ROLL ANDMETÖÖTLUSES JA ÄRIPROTSESSIDES.....	102
LLM jututubades: ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok andmeprotsesside automatiseerimiseks.	102
Suured LLM keelemudelid: kuidas see toimib	103
Kohalike LLMide kasutamine ettevõtte tundlike andmete jaoks.....	106
Täielik kontroll tehisintellekti üle ettevõttes ja kuidas oma LLM-i kasutusele võtta.....	108
RAG: Intelligent LLM - assistendid, kellel on juurdepääs ettevõtte andmetele.....	110
PEATÜKK 4.4. IDE KOOS LLM-TOEGA JA TULEVASED PROGRAMMIMUUDATUSED	113

IDE valimine: LLM-ist eksperimendid ärilahendusteni	113
LLM-i toetav IDE ja tulevased programmeerimise muudatused	115
Python Pandas: asendamatu tööriist andmetega töötamiseks	116
DataFrame: universaalne tabeliandmete formaat	120
Järgmised sammud: jätkusuutliku andmeraamistiku loomine	123
V OSA ANDMETE KVALITEET: KORRALDUS, STRUKTUREERIMINE, MODELLEERIMINE.....	125
PEATÜKK 5.1. ANDMETE TEISENDAMINE STRUKTUREERITUD KUJUL.....	126
Õppimine, kuidas muuta dokumendid, PDF, pildid ja tekstid struktureeritud vorminguteks.	126
Näide PDF-dokumendi -dokumendi konverteerimisest tabeliks.....	127
JPEG, PNG pildi teisendamine struktureeritud kujul	131
Tekstiandmete konverteerimine struktureeritud kujul	134
CAD-andmete (BIM) konverteerimine struktureeritud kujul.....	137
CAD lahenduste müüjad liiguvad struktureeritud andmete suunas.....	142
PEATÜKK 5.2. KLASSIFITSEERIMINE JA INTEGREERIMINE: ÜHINE KEEL EHITUSANDMETE JAKS	145
Otsuste tegemise kiirus sõltub andmete kvaliteedist	145
Andmete standardimine ja integreerimine.....	146
Digitaalne koostalitlusvõime algab nõuetest.....	149
Ühine ehituskeel: klassifikaatorite roll digitaalses ümberkujundamises	151
Masterformat, OmniClass, Uniclass ja CoClass: klassifitseerimissüsteemide areng.	154
PEATÜKK 5.3. ANDMETE MODELLEERIMINE JA TIPPKESKUS.....	159
Andmete modelleerimine: kontseptuaalne, loogiline ja füüsiline mudel	159
Praktiline andmemodelleerimine ehituse kontekstis	163
Andmebaasi loomine LLM-i abil	165
Andmete modelleerimise tippkeskus (CoE).....	167
PEATÜKK 5.4. NÕUETE SÜSTEMATISEERIMINE JA TEABE VALIDEERIMINE	170
Nõuete kogumine ja analüüs: teabevahetuse muutmine struktureeritud andmeteks.	170
Protsesside vooskeemid ja kontseptuaalsete raamistike tõhusus.....	174
Struktureeritud nõuded ja RegEx regulaaravaldised	176
Andmete kogumine kontrollimiseks.....	181
andmete kontrollimine ja kontrollimise tulemused	183
kontrollitulemuste visualiseerimine.....	188

Andmete kvaliteedikontrolli võrdlus ja inimelu vajaduste võrdlus	190
Järgmised sammud: andmete muutmine täpseteks arvutusteks ja plaanideks	192
VI OSA KULUDE JA AJA ARVUTAMINE: ANDMETE LISAMINE EHITUSPROTSESSIDESSE.....	194
PEATÜKK 6.1. EHITUSPROJEKTIDE KULUARVUTUSED JA KALKULATSIOONID	195
Ehituse põhitõed: koguse, maksumuse ja aja hindamine	195
Projektide hinnangulise maksumuse arvutamise meetodid	196
Ressursipõhine meetod kulukalkulatsioon ja kalkulatsioon ehituses.....	197
Ehitusressursside andmebaas: ehitusmaterjalide ja -tööde kataloog	197
Arvutuste koostamine ja tööde kulukalkulatsioon ressursibaasi alusel	199
Projekti lõplik kulukalkulatsioon: kalkulatsioonist eelarvesse	204
PEATÜKK 6.2. KOGUSELINE ARVESTUS NING KALKULATSIOONIDE JA GRAAFIKUTE AUTOMAATNE KOOSTAMINE	208
Üleminek 3D-lt 4D-le ja 5D-le: ruumiliste ja kvantitatiivsete parameetrite kasutamine.....	208
5D atribuudid ja atribuutide mahtude saamine CADist	208
QTO Quantity Take-Off: projektandmete rühmitamine atribuutide järgi	212
QTO automatiseerimine kasutades LLM ja struktureeritud andmeid	217
QTO kogu projekti arvutamine, kasutades Exceli tabelis esitatud gruppireegleid.	221
PEATÜKK 6.3. 4D, 6D -8D JA SÜSINIKDIOKSIIDI HEITKOGUSTE ARVUTAMINE CO ₂	227
4D-mudel: aja integreerimine ehitushinnangutesse	227
Ehitusgraafik ja selle automatiseerimine kulukalkulatsiooni andmete alusel	228
Laiendatud atribuutikihid 6D -8D: energiatõhususest ohutuse tagamiseni.	230
CO ₂ hindamine ja ehitusprojektide süsinikdioksiidi heitkoguste arvutamine	233
PEATÜKK 6.4. EHITUS ERP JA PMIS SÜSTEEMID	238
Ehitus ERP -süsteemid arvutuste ja kalkulatsioonide näitel	238
PMIS: vahendaja ERP ja ehitusplatsi vahel	243
Spekuleerimine, kasum, isoleeritus ja vähene läbipaistvus ERP ja PMISi puhul	244
Suletud ERP-ajastu lõpp /PMIS: ehitustööstus vajab uusi lähenemisviise.....	247
Järgmised sammud: projektandmete tõhus kasutamine	248
VII OSA CAD JA BIM: TURUNDUS, TEGELIKKUS JA PROJEKTEERIMISANDMETE TULEVIK EHITUSES	251
PEATÜKK 7.1. BIM- KONTSEPTSIOONIDE TEKKIMINE EHITUSSEKTORIS	252
BIM-i ja avatud BIM-i kui CAD- müüjate turunduskontseptsioonide tekkimise ajalugu.	252
BIM-i tegelikkus: integreeritud andmebaaside asemel - suletud moodulsüsteemid	255

Avatud formaadis IFC ilmumine ehitussektoris	257
IFC formaadi probleem sõltuvalt geomeetrisest tuumast	259
Ilmumine semantika teema ülesehitamisel ja ontoloogia	262
Miks semantilised tehnoloogiad ei vasta ehitussektoris ootustele	264
PEATÜKK 7.2. SULETUD PROJEKTIFORMAADID JA KOOSTALITLUSVÕIME PROBLEEMID	268
Suletud andmed ja tootlikkuse langus: CAD-tööstuse ummikseis (BIM)	268
Müüt koostalitlusvõime kohta CAD-süsteemide vahel	270
Mine USD ja granuleeritud andmed m.....	274
PEATÜKK 7.3. EHITUSGEOMEETRIA: JOONTEST KUUPMEETRITENI	278
Kui jooned muutuvad rahaks või miks ehitajad vajavad geometriat.....	278
Joontest mahtudeks: kuidas pindala ja maht muutuvad andmeteks	278
Üleminek MESH, USD ja hulknurksed: geomeetria tessellatsiooni kasutamine.....	281
LOD, LOI, LOMD - detailide unikaalne liigitus CADis (BIM).....	282
Uued CAD standardid (BIM) - AIA, BEP, IDS, LOD, COBie	285
PEATÜKK 7.4. PROJEKTEERIMISPARAMEETRITE MÄÄRAMINE JA KASUTAMINE CAD-TÖÖDE JAKS.	290
CAD-andmete unikaalsuse illusioon (BIM): tee analüütika ja avatud vormingute poole	290
Projekteerimine läbi parameetrite: CADi tulevik ja BIM.....	293
LLM-i tekkimine projekteerimise CAD andmetöötlusprotsessides	296
DWG -failide automatiseeritud analüüs LLM ja Pandasiga.....	299
Järgmised sammud: üleminek suletud vormingutelt avatud andmetele	305
VIII OSA ANDMEPÕHINE OTSUSTE TEGEMINE, ANALÜÜTIKA, AUTOMATISEERIMINE JA MASINÕPE	307
PEATÜKK 8.1. ANDMEANALÜÜS JA ANDMEPÕHINE OTSUSTE TEGEMINE	308
Andmed kui ressurss otsuste tegemisel.....	308
Andmete visualiseerimine: mõistmise ja otsuste tegemise võti.....	311
KPId ja ROI.....	313
Armatuurilauad ja armatuurilauad: mõõdikutega visualiseerimine tõhusa juhtimise eesmärgil..	315
Andmeanalüüs ja küsimuste esitamise kunst	317
PEATÜKK 8.2. ANDMEVOOG ILMA MANUAALSE VAEVATA: MIKS ON VAJA ETL.....	320
ETL automatiseerimine: kulude vähendamine ja andmekäitluse kiirendamine	320
ETL Extract: andmete kogumine.....	324
ETL Transform: valideerimis- ja transformatsioonireeglite rakendamine	327

ETL Load: Visualiseeri tulemused graafikutes ja graafikutes.....	329
ETL Load: PDF dokumentide automaatne loomine	334
ETL Load: dokumentide automaatne genereerimine FPDF-st.....	335
ETL Load: aruandlus ja laadimine teistesse süsteemidesse	339
ETL koos LLMiga: andmete visualiseerimine PDF -dokumentid	340
PEATÜKK 8.3. AUTOMAATNE ETL KONVEIER (TORUSTIK).....	345
Pipeline: Automaatne ETL konveier andmed	345
Torustik -ETL andmete valideerimisprotsess LLMiga.....	349
Torustik -ETL: projektielementide andmete ja teabe kontrollimine CADis (BIM).	351
PEATÜKK 8.4. ETL ORKESTREERIMINE JA TÖÖVOOD: PRAKTILISED LAHENDUSED	357
DAG ja Apache Airflow: töövoogude automatiseerimine ja orkestreerimine	357
Apache Airflow: praktiline rakendus ETLi automatiseerimisel	358
Apache NiFi marsruutimiseks ja andmete teisendamiseks.....	362
n8n Low-Code, No-Code protsessi orkestreerimine	363
Järgmised sammud: üleminek manuaalsetelt toimingutelt analüütilistele lahendustele	366
IX OSA ANDMETE SÄILITAMINE JA HALDAMINE EHITUSES	368
PEATÜKK 9.1. ANDMETARISTU: SALVESTUSFORMAATIDEST DIGITAALSETE HOIDLATENI	369
Andmeatomid: tõhusa teabehalduse alus	369
Teabe säilitamine: failid või andmed	370
Suurandmete säilitamine: populaarsete vormingute ja nende tõhususe analüüs	372
Salvestusruumi optimeerimine Apache Parquet'iga	375
andmeladu.....	377
Andmejärv - ETLi areng kuni ELT: traditsioonilisest puhastamisest paindliku töötlemiseni.....	379
Data Lakehouse'i arhitektuur: andmeaida ja andmejärvede sünergia	381
CDE, PMIS, ERP või DWH ja Data Lake	383
PEATÜKK 9.2. ANDMELAO HALDAMINE JA KAOSE VÄLTIMINE	387
Vektorandmebaasid ja Bounding Box	387
Andmehaldus), andmeminimalism) ja andmesoo)	390
DataOps ja VectorOps: uued andmestandardid	393
Järgmised sammud: kaootilisest ladustamisest struktureeritud ladustamiseni	395
X OSA SUURED ANDMED, MASINÕPE JA PROGNOOSID	397
PEATÜKK 10.1. SUURED ANDMED JA NENDE ANALÜÜS	398

Suurandmed ehituses: intuitsioonist prognoositavuseni.....	398
Suurandmete teostatavuse kahtluse alla seadmine: korrelatsioon, statistika ja andmete valimi moodustamine.....	399
Suured andmed: San Francisco miljonite ehituslubade andmekogumi andmete analüüsimine	402
Näide CAD-andmetel põhinevate suurandmete kohta (BIM).....	408
asjade internet Asjade internet ja arukad lepingud	412
PEATÜKK 10.2. MASINÕPE JA PROGNOOSID	416
Masinõpe ja tehisintellekt muudavad seda, kuidas me ehitame.....	416
Subjektiivsest hinnangust statistilise prognoosini	418
Titanic data set: Hello World analüütika maailmas andmed ja suurandmed	420
Masinõpe tegevuses: Titanicu reisijatest projektijuhtimiseni	425
Ajaloolistel andmetel põhinevad prognoosid ja prognoosid	429
Masinõppe põhimõisted.....	431
PEATÜKK 10.3. KULUDE JA TÄHTAEGADE PROGNOOSIMINE MASINÕPPE ABIL	434
Näide masinõppe kasutamise kohta projekti kulude ja tähtaegade leidmiseks.	434
Projekti kulude ja aja prognoosimine lineaarse regressiooni abil	436
Projekti kulude ja aja prognoosimine K-lähedaseima naabri algoritmi abil (k-NN).....	439
Järgmised sammud: ladustamisest analüüsi ja prognoosimiseni	443
XI OSA E HITUSTÖÖSTUS DIGITAALAJASTUL. VÕIMALUSED JA VÄLJAKUTSED	446
PEATÜKK 11.1. ELLUJÄÄMISSTRATEEGIAID: KONKURENTSIEELISTE LOOMINE	447
Korrelatsioonid arvutuste asemel: ehitusanalüütika tulevik	447
Andmepõhine lähenemisviis ehituses: uus infrastruktuuritase.....	450
Järgmise põlvkonna digitaalne kontor: kuidas tehisintellekt muudab tööruumi.	452
Avatud andmed ja Uberiseerimine on oht olemasolevale ehitusettevõttele.....	454
Ümberkorralduste lahendamata probleemid kui viimane võimalus kasutada aega ümberkujundamiseks	457
PEATÜKK 11.2. PRAKTILINE JUHEND ANDMEPÕHISE LÄHENEMISVIISI RAKENDAMISEKS.....	462
Teoriast praktikasse: teekaart digitaalseks ümberkujundamiseks ehituses	462
Digitaalse vundamendi rajamine: 1-5 sammu digitaalse küpsuse saavutamiseks	464
Andmete potentsiaali avamine: 5-10 sammu digitaalse küpsuse saavutamiseni	468
Ümberkujundamise tegevuskava: kaosest andmepõhiseks ettevõtteks.....	475
Ehitus tööstuses 5.0: kuidas teenida raha, kui sa ei saa enam varjata.....	478
KOKKUVÕTE.....	480

AUTORI KOHTA.....	483
EDASINE SUHE.....	484
TÕLKEKOMMENTAAR.....	484
MUUD OSKUSED JA MÕISTED	485
GLOSSARIUM	489
VIIDETE JA VEEBIMATERJALIDE LOETELU	496
AINETE INDEKS.....	513

MAKSIMAALNE MUGAVUS TRÜKITUD VERSIOONIGA

Sul on käes tasuta digitaalne versioon raamatust **Andmepõhine ehitus**. Mugavamaks tööks ja kiireks juurdepääsuks materjalidele soovitame pöörata tähelepanu **trükitud väljaandele**:



■ **Alati käepärast:** trükitud raamatust saab usaldusväärne töövahend, mis võimaldab teil igas töösituatsioonis kiiresti leida ja kasutada õigeid visualiseeringuid ja skeeme.

■ **Kvaliteetsed illustratsioonid:** kõik trükiväljaande pildid ja graafika on esitatud maksimaalse kvaliteediga.

■ **Kiire juurdepääs teabele:** lihtne navigeerimine, võimalus teha märkmeid, järjehoidjaid ja töötada raamatuga igal pool.

Ostes raamatu trükitud täisversiooni, saate mugava abivahendi mugavaks ja tõhusaks tööks teabega: võimaluse kasutada visuaalseid materjale kiiresti igapäevastes ülesannetes, leida kiiresti vajalikud skeemid ja teha märkmeid. Lisaks toetab teie ostuga avatud teadmiste levitamist.

Tellige raamatu trükitud versioon aadressil: datadrivenconstruction.io/books



I OSA

SAVITAHVLITEST KUNI DIGITAALSE REVOLUTSIOONINI: KUIDAS TEAVE ON ARENENUD EHITUSES

Raamatu esimeses osas vaadeldakse andmehalduse ajaloolist arengut ehitussektoris, alates füüsilisel andmekandjal olevate algsete andmete säilitamisest kuni tänapäevaste digitaalsete ökosüsteemideni. Selles analüüsitakse andmehaldustehnoloogiate muutumist, ERP - süsteemide tekkimist ja andmete killustatuse mõju äriprotsesside tõhususele. Erilist tähelepanu pööratakse teabe digitaliseerimise protsessile ja objektiivse analüüsi kasvavale tähtsusele subjektiivse eksperthinnangu asemel. Üksikasjalikult uuritakse kaasaegse ehitussektori ees seisvat teabe mahu eksponentsiaalset kasvu ja sellega seotud väljakutseid ettevõtlussüsteemidele. Uuritakse ehitussektori positsiooni neljanda ja viienda tööstusrevolutsiooni kontekstis ning uuritakse tehisintellekti ja andmekeskse lähenemise potentsiaali jätkusuutliku konkurentsieelise loomiseks.

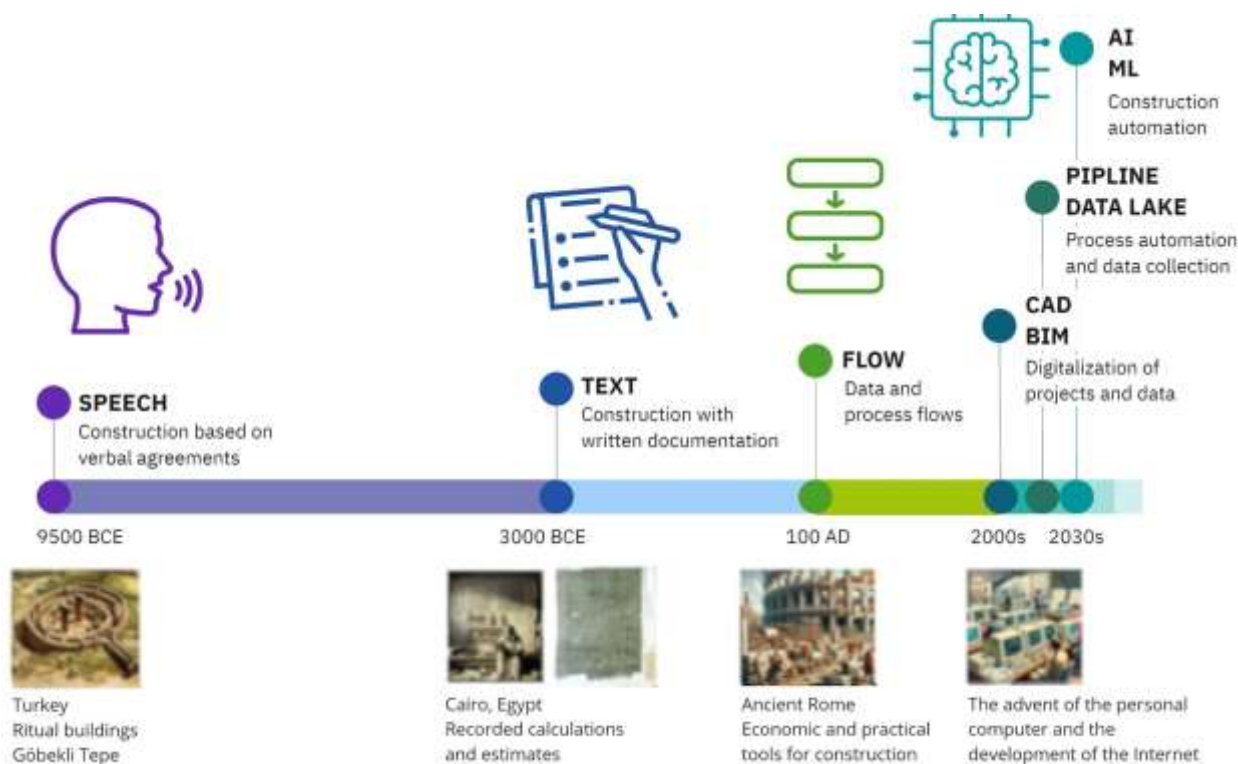
PEATÜKK 1.1.

ANDMETE KASUTAMISE ARENG EHTUSSEKTORIS

Andmeajastu sündi ehituses

Umbes 10 000 aastat tagasi, neoliitikumis, tegi inimkond oma arengus revolutsioonilise ülemineku, loobudes nomaadlikust eluviisist ja asudes elama sessiivselt, mis tõi kaasa esimeste savist, puidust ja kivist ehitatud primitiivsete ehitiste tekkimise [6]. Sellest hetkest algab ehitustööstuse ajalugu.

Tsivilisatsioonide arenedes muutus arhitektuur üha keerulisemaks, mis viis esimeste rituaalsete templite ja avalike ehitiste rajamiseni. Arhitektuursete projektide suurenev keerukus nõudis antiikajal inseneridelt ja juhtidelt esimeste protokollide ja arvutuste koostamist. Esimesed savitahvlitele ja papüürustele kantud üleskirjutused sisaldasid sageli kirjeldust, kuidas arvutati vajalike ehitusmaterjalide kogust, nende maksumust ja tehtud tööde eest tasumise arvutamist [7]. Nii algas ehituses andmete kasutamise ajastu - kaua enne kaasaegse digitaaltehnoloogia tulekut (joonis 1.1-1).

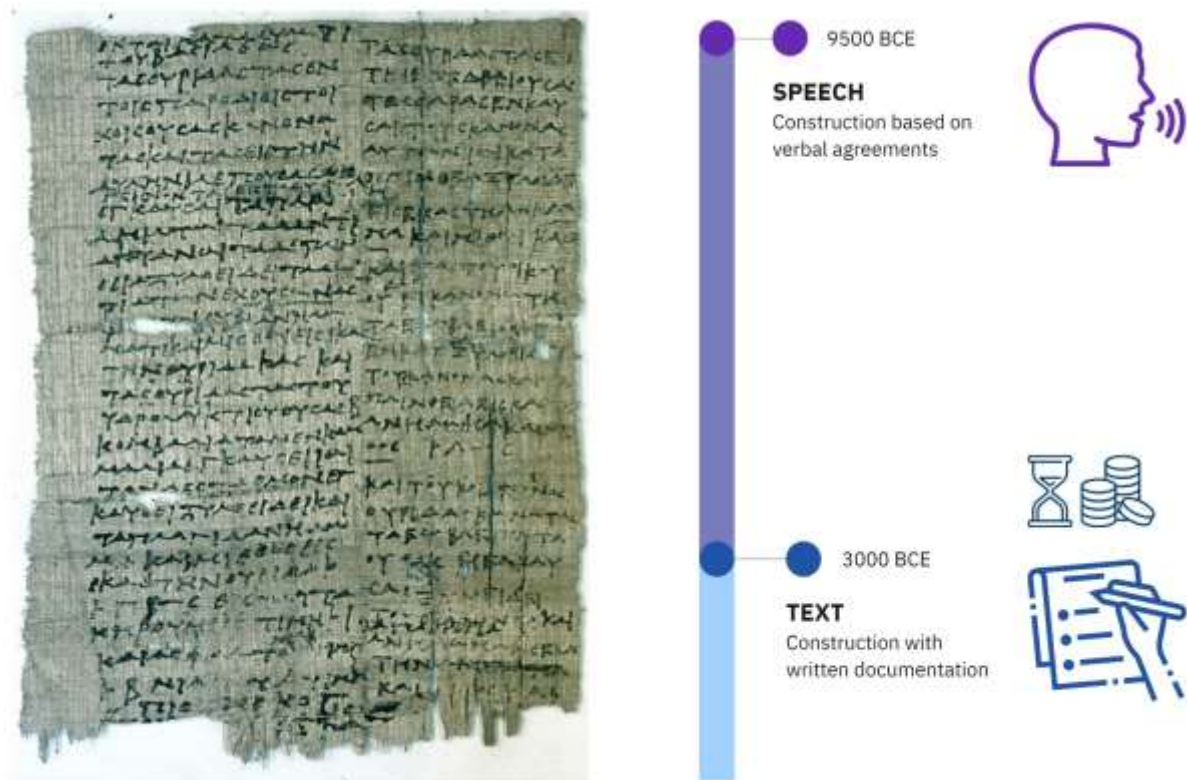


Joonis 1.1-1 Infotehnoloogia arengu kronoloogia ehituses: verbaalsest teabest tehisintellektiini.

Alates savist ja papüürusest kuni digitaaltehnoloogiani

Esimesed dokumentaalsed tõendid ehitamise kohta pärinevad püramiidide ehitamise ajast, umbes 3000-4000 eKr[7]. Sellest ajast alates on kirjalike dokumentide pidamine hõlbustanud ja saatnud ehitussektori arengut, võimaldades teadmiste kogumist ja süstematiseerimist, mis järgmise 10 000 aasta jooksul viisid märkimisväärsete uuendusteni ehitusmeetodites ja arhitektuuris.

Esimeste füüsiliste andmekandjate, nagu aastatuhandete taguste savitahvlite, papüüruse (joonis 1.1-2) või 1980. aastate "A0" paberi kasutamine ehituses andmete salvestamiseks ei olnud algselt mõeldud selle teabe rakendamiseks uutes projektides. Selliste kirjade peamine eesmärk oli täpsustada projekti hetkeseisu, sealhulgas vajalike materjalide ja tööde maksumuse arvutused. Samamoodi ei taga tänapäeval digitaalsete projekteerimisandmete ja mudelite kättesaadavus alati nende kasutamist tulevastes projektides ning sageli on see peamiselt teave vajalike materjalide ja ehitusmaksumuse jooksvate arvutuste jaoks.



Joonis 1.1-2 3. sajandil eKr. koostatud papüürus, kus kirjeldatakse kuningliku palee eri tüüpi akende värvimise maksumust, kasutades enkaustikatehnikat.

Inimkonnal kulus umbes 5000 aastat, et minna ehitusprojektide juhtimisel suulistelt vestlustelt üle kirjalikele dokumentidele, ja sama palju aega kulus ka paberilt digitaalsetele andmetele kui esmasele planeerimis- ja kontrollivahendile.

Nii nagu kaubandus- ja rahasuhete areng stimuleeris kirjanduse ja esimeste juristide tekkimist vaidluste lahendamiseks, nii viisid esimesed materjalikulude ja töömahtude registreerimine ehituses esimeste juhtide tekkimiseni ehitussektoris, kelle ülesandeks oli dokumenteerida, jälgida ja vastutada projekti ajakava ja kulusid käsitleva põhiteabe eest.

Tänapäeval mängivad andmed palju olulisemat rolli: need ei salvesta mitte ainult tehtud otsuseid, vaid muutuvad ka tuleviku prognoosimise ja modelleerimise vahendiks. See on vundament, millele on rajatud kaasaegne protsessipõhine lähenemine projektijuhtimises - kogutud kogemuste muutmine struktureeritud ja kontrollitavatel andmetel põhinevaks otsuste tegemise süsteemiks.

Protsess kui andmepõhise kogemuse vahend

Iga protsessi keskmeks on mineviku kogemuste muutmine tuleviku planeerimise vahendiks. Kogemused tänapäevases mõttes on struktureeritud andmekogum, mida saab analüüsida, et teha teadlikke prognoose.

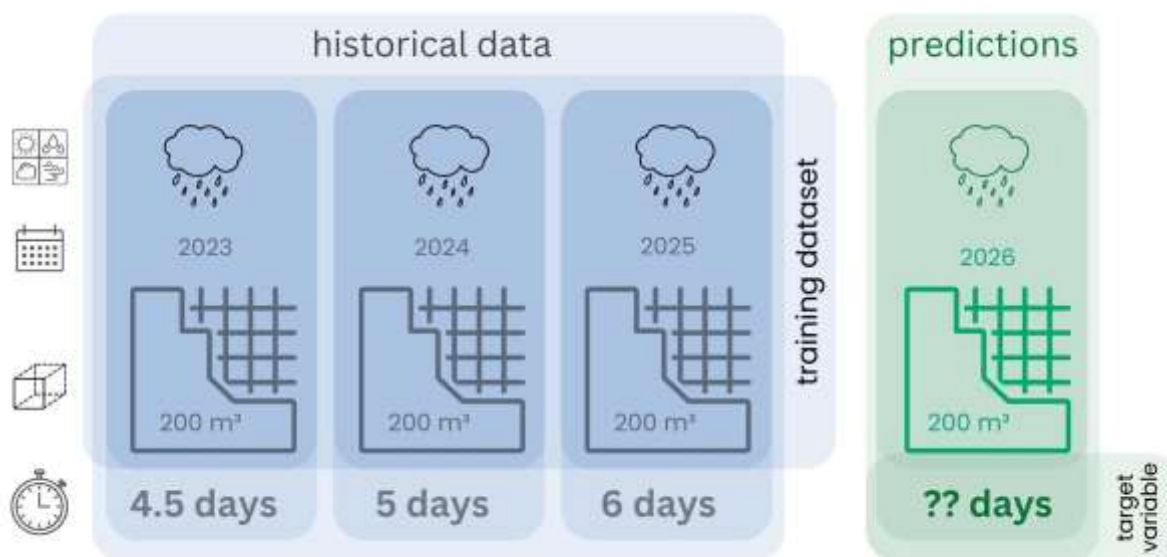
Prognoosimise aluseks on ajaloolised andmed, sest need näitavad selgelt tehtud töö tulemusi ja annavad ülevaate neid tulemusi mõjutavatest teguritest.

Võtame konkreetse näite monoliitse ehituse kohta: tavaliselt võetakse tööde ajakava planeerimisel arvesse betooni mahtu, konstruktsiooni keerukust ja ilmastikutingimusi. Oletame, et konkreetse ehitusplatsi töödejuhataja või ettevõtte viimase kolme aasta (2023-2025) ajaloolised andmed näitavad, et 200 m² suuruse monoliitkonstruktsiooni valamine vihmase ilmaga võttis aega 4,5-6 päeva (joonis 1.1-3). Just need kogutud statistilised andmed saavad aluseks, kui tulevaste projektide sarnaste tööde planeerimisel prognoositakse tööde teostamise aega ja ressursside maksumust. Nende ajalooliste andmete põhjal saab töödejuhataja või kalkulaator teha kogemustele tuginedes teadliku prognoosi selle kohta, kui palju aega kulub 2026. aastal sarnaste tööde teostamiseks sarnastes tingimustes.

Antud ajahinnangute puhul toimib analüütiline protsess mehhanismina, mis muudab ebaühtlased andmed struktureeritud kogemusteks ja seejärel täpselt planeerimisvahendiks. Andmed ja protsessid moodustavad ühtse ökosüsteemi, kus üks ei saa eksisteerida ilma teiseta.

Arvuta, mis on mõõdetav, mõõda, mis on mõõdetav, ja tee mõõdetavaks see, mis ei ole mõõdetav [8].

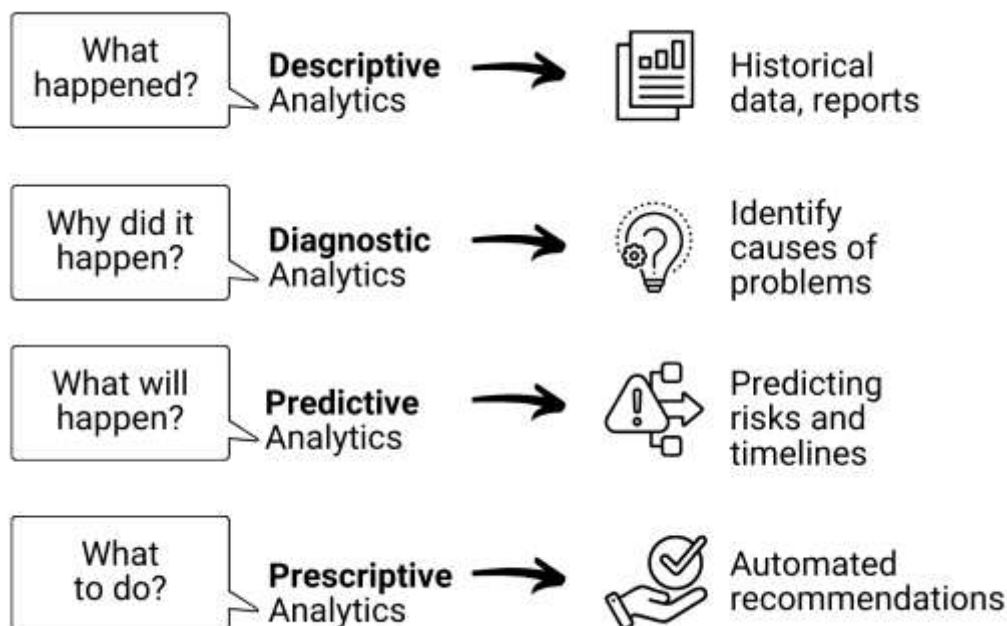
- Galileo Galilei



Joonis 1.1-3 Ajaloolised andmed toimivad koolitusandmestikuna ühe tulevase väärtuse prognoosimiseks.

Tänapäeva ärimaastikul on andmeanalüüs muutumas tõhusa projektijuhtimise, protsesside optimeerimise ja strateegiliste otsuste tegemise kriitiliseks komponendiks. Ehitustööstus omandab järk-järgult neli peamist analüütika taset, millest igaüks vastab konkreetsele küsimusele ja pakub ainulaadset kasu (joonis 1.1-4):

- **Kirjeldav analüüs** - vastab küsimusele "*mis juhtus?*" ning annab ajaloolisi andmeid ja aruandeid varasemate sündmuste ja tulemuste kohta: viimase kolme aasta jooksul (2023-2025) kulus vihmase ilmaga 200m² monoliitse konstruktsiooni valamiseks 4,5-6 päeva.
- **Diagnostiline analüüs** - vastab küsimusele "*miks see juhtus?*", tuvastades probleemide põhjused: analüüs näitab, et monoliitkonstruktsiooni valamisaeg pikenes vihmase ilma tõttu, mis aeglustas betooni kõvenemisprotsessi.
- **Ennustav analüüs** - tulevikku suunatud, võimalike riskide ja ettevalmistusaegade prognoosimine, vastates küsimusele "*mis juhtub?*": ajalooliste andmete põhjal prognoositakse, et 2026. aastal võtab sarnase 200 m² suuruse monoliitse konstruktsiooni valamine vihmase ilmaga umbes 5,5 päeva, võttes arvesse kõiki teadaolevaid tegureid ja suundumusi.
- **Ettekirjutav analüüs** - annab automatiseeritud soovitusi ja vastab küsimusele "*mida teha?*", võimaldades ettevõtetel valida optimaalseid tegevusi: Tööde optimeerimiseks soovitatakse näiteks: kasutada spetsiaalseid lisandeid, mis kiirendavad betooni kõvenemist kõrge õhuniiskuse tingimustes; planeerida valamine perioodidesse, kus sademete tõenäosus on kõige väiksem; korraldada konstruktsioonile ajutised varjendid, mis lühendavad tööaega 4-4,5 päevani ka ebasoodsate ilmastikuolude korral.



Joonis 1.1-4 Analüütika põhitüübid: mineviku kirjeldamisest automatiseeritud otsuste tegemiseni.

Täielik digitaalne ümberkujundamine, mis eeldab üleminekut süsteemianalüütikale ja andmepõhisele juhtimisele, nõuab mitte ainult allhankeid, vaid ka pädeva sisemise meeskonna moodustamist. Sellise meeskonna võtmeisikud peaksid olema tootejuhid, andmetöötledajad, analüütikud ja arendajad, kes teevad tihedat koostööd äriüksustega (joonis 4.3-9). Selline koostöö on vajalik intelligentsete analüütiliste küsimuste esitamiseks ja äriotsuste tegemise ülesannete tõhusaks parameetriseerimiseks. Infoühiskonnas muutuvad andmed mitte ainult abivahendiks, vaid prognoosimise ja optimeerimise aluseks.

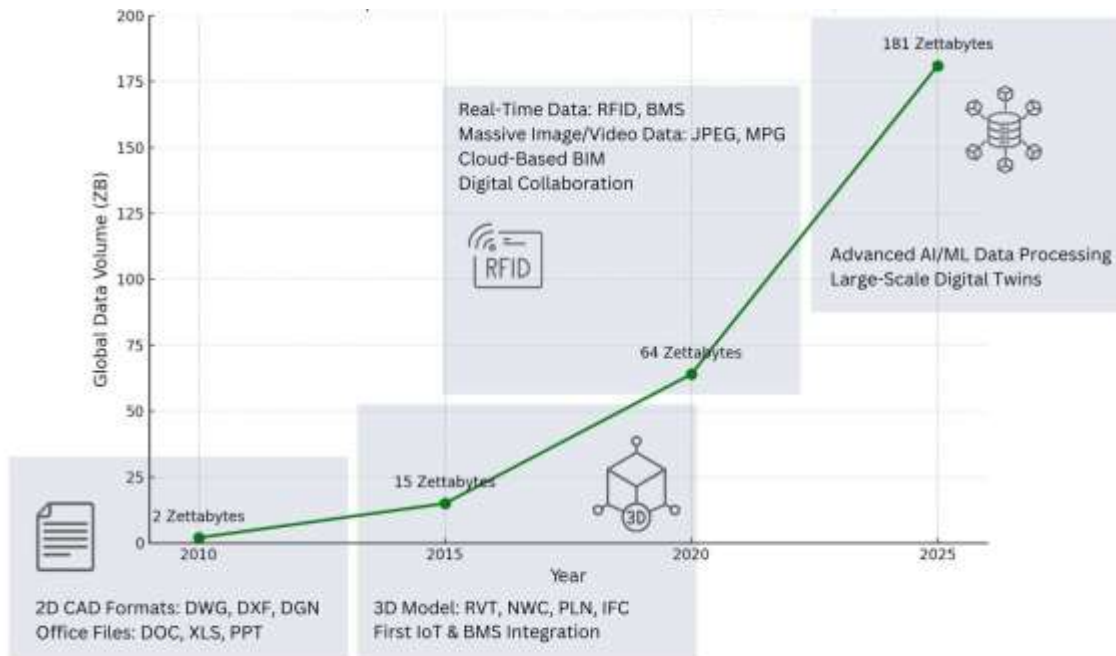
Ehituses muudab digitaalne ümberkujundamine põhjalikult seda, kuidas rajatise projekteeritakse, hallatakse ja käitatakse. Seda protsessi nimetatakse teabe digitaliseerimiseks, mille puhul kõik ehitusprotsessi aspektid digitaliseeritakse analüüsiks sobivasse digitaalsesse vormi.

Ehitusprotsesside digitaliseerimine

Ehituses salvestatud teabe hulk on aastatuhandeid peaaegu muutumatuna püsinud, kuid viimastel aastakümnetel on see kiiresti kasvanud (joonis 1.1-5).

Vastavalt PwC uuringule® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] on 90% kõigist maailma andmetest loodud viimase kahe aasta jooksul (2015. aasta seisuga). Siiski ei kasuta enamik ettevõtteid neid andmeid täielikult ära, kuna need jäävad kas siloorsetesse süsteemidesse või lihtsalt arhiveeritakse ilma tegeliku analüüsita.

Andmemahu suurenemine on viimastel aastatel ainult kiirenenud, kahekordistudes 15 zettabaidilt 2015. aastal 181 zettabaidini 2025. aastal [10]. Iga päev täituvad ehitus- ja projekteerimisfirmade serverid projektdokumentatsiooni, tööplaanide, arvutuste ja kalkulatsioonide, finantsaruannetega. 2D/3D -jooniste jaoks kasutatakse DWG, DXF ja DGN formaate ning 3D-mudelite jaoks RVT, NWC, PLN ja IFC™. Tekstidokumendid, tabelid ja esitlused salvestatakse DOC, XLSX ja PPT formaadis. Lisaks videotele ja piltidele ehitusplatsilt - MPG ja JPEG formaadis, reaajas andmed asjade interneti komponentidest, RFID® sildid (identifitseerimine ja jälgimine) ja BMS hoonete juhtimissüsteemid (järelevalve ja kontroll)



Joonis 1.1-5 Paraboolsete andmete kasv 2010-2025 (põhineb [10]).

Seoses teabe kiire kasvuga seisab ehitussektor silmitsi vajadusega mitte ainult koguda ja säilitada andmeid, vaid ka tagada nende kontrollimine, valideerimine, mõõdetavus ja analüütiline töötlemine. Tänapäeval on tööstusharu läbimas aktiivset teabe digitaliseerimise etappi - ehitustegevuse kõigi aspektide süstemaatilist muutmist digitaalseks vormiks, mis sobib analüüsiks, tõlgendamiseks ja automatiseerimiseks.

Teabe digitaliseerimine tähendab, et võetakse teave kõigi ehitusprojekti ja ehitusprotsessi enda üksuste ja elementide kohta - sealhulgas nende kohta, mida me varem ei pidanud üldse informatsiooniks - ja teisendatakse see andmeformaati, et muuta teave mõõdetavaks ja kergesti analüüsitavaks.

Ehituse kontekstis tähendab see, et analüütiliste mudelite loomiseks tuleb koguda ja digiteerida teavet projektide kõigi elementide ja protsesside kohta - alates masinate ja inimeste liikumisest ehitusplatsil kuni ehitusplatsi ilmastiku- ja ilmastikutingimuste, jooksvate materjalihindade ja keskpanga intressimääradeni.

Kui te suudate mõõta seda, millest te räägite, ja väljendada seda numbrites, siis te teate midagi sellest teemast. Aga kui te ei suuda seda kvantitatiivselt väljendada, on teie teadmised äärmiselt piiratud ja mitterahuldavad. See võib olla lähtepunkt, kuid see ei ole tõelise teadusliku teadmise tase. [11].

- W. Thomson (lord Kelvin), 1824-1907, Briti teadlane

Teabe digitaliseerimine läheb kaugemale traditsioonilisest lähenemisviisist teabe kogumisele, kus registreeriti ainult põhilised näitajad, nagu töötunnid või tegelikud materjalikulud. Tänapäeval saab praktiliselt iga sündmuse muuta andmevooks, mis sobib põhjalikuks analüüsiks, kasutades täiustatud analüütilisi vahendeid ja masinõppe meetodeid. Ehitustööstuses on toimunud põhimõtteline üleminek paberjoonistelt, Exceli tabelitelt ja suulistelt juhistelt digitaalsetele süsteemidele (joonis 1.2-4), kus iga projekti element muutub andmeallikaks. Isegi töötajaid - alates inseneridest kuni ehitustöötajateni ehitusplatsil - vaadeldakse nüüd kui digitaalsete muutujate ja andmekogumite kogumit.

KPMG uuringu "Tuttavad väljakutsed - uued lähenemisviisid: ülemaailmne ehitusuuring 2023" kohaselt on digitaalsed kaksikud, tehisintellekt (AI) ja suurandmed, kujunemas projektide kasumlikkuse parandamise peamisteks tõukejõududeks [2].

Kaasaegsed tehnoloogiad mitte ainult ei lihtsusta teabe kogumist, muutes selle suures osas automaatseks, vaid vähendavad ka radikaalselt andmete säilitamise kulusid. Selle tulemusena on ettevõtted eemaldunud valikulisest lähenemisviisist ja eelistavad kogu teabe kogumist hilisemaks analüüsiks (joonis 2.1-5), mis avab potentsiaalsed võimalused protsesside optimeerimiseks tulevikus.

Teabe digitaliseerimine ja digitaliseerimine võimaldavad avastada teabe varjatud, seni kasutamata väärtust. Kui andmed on õigesti korraldatud, saab neid uuesti kasutada, ümber tõlgendada ja integreerida uutesse teenustesse ja lahendustesse.

Tulevikus toob teabe digitaliseerimine tõenäoliselt kaasa dokumendihalduse täieliku automatiseerimise, isejuhtivate ehitusprotsesside kasutuselevõtu ja uute ametite - ehitusandmete analüütikute, tehisintellekti projektijuhtimise ekspertide ja digitaalinseneride - tekkimise. Ehitusprojektid muutuvad dünaamilisteks teabeallikateks ning otsuste tegemine ei põhine enam intuitsioonil või subjektiivsel kogemusel, vaid usaldusväärsetel ja korratavatel digitaalsetel faktidel

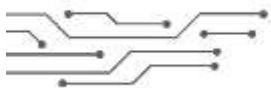
Teave on 21. sajandi nafta ja analüütika on sisepõlemismootor [12].

- Peter Sondergaard, Gartneri vanem asepresident.®

Uuringu IoT Analytics 2024 [13] kohaselt kasvavad ülemaailmsed kulutused andmehaldusele ja -analüüsile eeldatavasti järsult 185,5 miljardilt dollarilt 2023. aastal 513,3 miljardi dollarini 2030. aastaks, mis tähendab 16%-list aastakasvu. Siiski ei kasva kõik komponendid ühesuguse kiirusega: analüütika kasvab kiiresti, samas kui salvestusruumi kasv aeglustub. Andmehalduse ökosüsteemis kasvab analüütika kõige kiiremini: prognooside kohaselt kasvab see 60,6 miljardilt dollarilt 2023. aastal 227,9 miljardi dollarini 2030. aastaks, mis on 27%-line aastane kasvutempo.

Seoses teabe kiireneva digitaliseerimisega ja teabemahtude kiire kasvuga seisavad ehitusprojektide ja ettevõtete juhtimine silmitsi vajadusega süstemaatiliselt salvestada, analüüsida ja töödelda mitmekesiseid, sageli heterogeenseid andmeid. Vastuseks sellele väljakutsele algas alates 1990ndate keskpaigast tööstusharus ulatuslik üleminek dokumentatsiooni elektroonilisele loomisele, säilitamisele ja haldamisele - alates tabelitest ja projekteerimisarvutustest kuni jooniste ja lepinguteni.

Traditsioonilised paberdokumendid, mis nõuavad allkirju, füüsilist säilitamist, regulaarset läbivaatamist ja arhiveerimist kappides, asendatakse järk-järgult digitaalsüsteemidega, mis salvestavad andmeid struktureeritult - spetsiaalsete rakenduste andmebaasides.



PEATÜKK 1.2.

TEHNOLOOGIAD JA JUHTIMISSÜSTEEMID KAASAEGSES EHTUSES

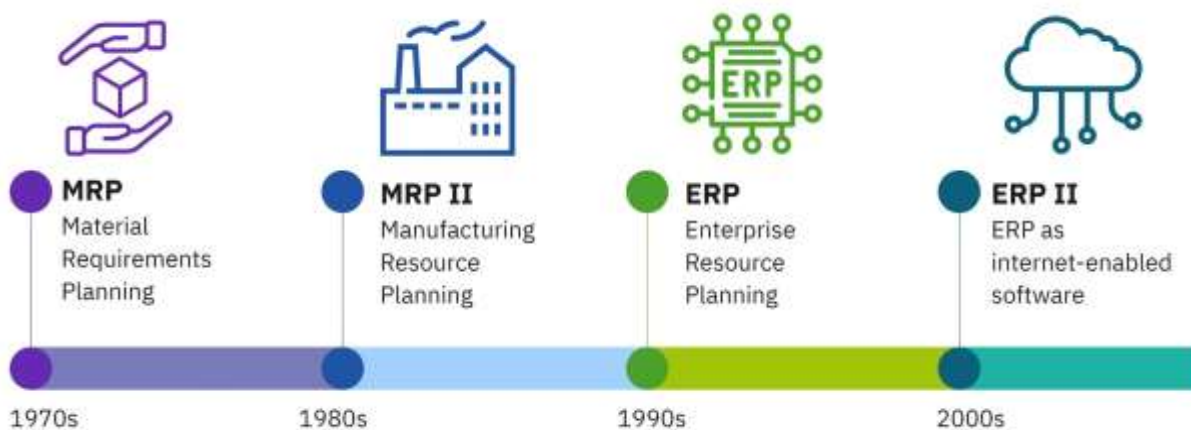
Digitaalne revolutsioon ja modulaarsete MRP/ERP -süsteemide teke.

Kaasaegse digitaalse andmesalvestuse ja -töötamise ajastu algas 1950. aastatel magnetlinde kasutuselevõtuga, mis avas võimaluse salvestada ja kasutada suuri andmehulki. Järgmine läbimurre oli kettaseadmete tulek, mis muutis radikaalselt andmehalduse käsitlust ehitustööstuses.

Andmelao arendamise tõttu on suur hulk ettevõtteid sisenenud lahenduste turule ja hakanud arendama moodulitarkvara andmete loomiseks, salvestamiseks, töötlemiseks ja rutiinsete ülesannete automatiseerimiseks

Teabe ja tööriistade hüppeline kasv on toonud kaasa vajaduse integreeritud, modulaarsete lahenduste järele, mis ei tööta üksikute failidega, vaid aitavad hallata ja kontrollida andmevooge protsesside ja projektide lõikes.

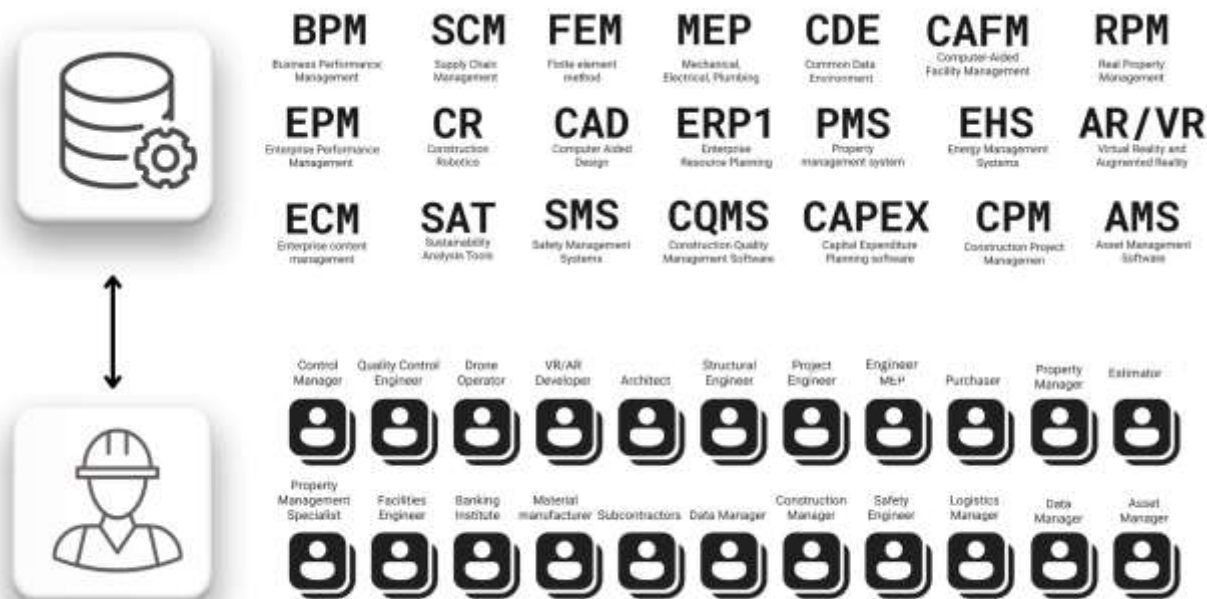
Esimesed terviklikud platvormi tööriistad pidid mitte ainult dokumente säilitama, vaid ka dokumenteerima kõik muudatuste taotlused ja toimingud protsessides: kes need algatas, milline oli taotluse ulatus ja mis lõpuks väärtuse või atribuutina salvestati. Selleks oli vaja süsteemi, mis suudaks jälgida täpseid arvutusi ja tehtud otsuseid (joonis 1.2-1). Sellisteks platvormideks olid esimesed MRP (Material Requirements Planning) ja ERP (Enterprise Resource Planning) süsteemid, mis muutusid populaarseks alates 1990. aastate algusest [14]



Joonis 1.2-1 Andmesalvestustehnoloogia areng viis ERP -süsteemide tekkimiseni 1980. aastatel.

Esimesed MRP - ja ERP - süsteemid panid aluse äriprotsesside ja ehitusprojektide juhtimise digitaliseerimise ajastule. Algselt põhiliste äriprotsesside automatiseerimiseks loodud moodulsüsteemid integreeriti lõpuks täiendavate, paindlikumate ja kohandatavate tarkvaralahendustega.

Need lisalahendused olid mõeldud andmete töötlemiseks ja projekti sisu haldamiseks (joonis 1.2-2), nad kas asendasid suurte süsteemide teatud moodulid või täiendasid neid tõhusalt, laiendades kogu süsteemi funktsionaalsust.



Joonis 1.2-2 Uued tarkvaralahendused on meelitanud ettevõttesse andmevoogude haldamiseks terve armee juhte.

Viimastel aastakümnetel on ettevõtted investeerinud palju moodulsüsteemidesse [15], pidades neid pikaajalisteks integreeritud lahendusteks.

Software Path 2022. aasta aruande [16] kohaselt on ERP - süsteemi keskmine eelarve kasutaja kohta 9000 dollarit. Keskmiselt kasutab selliseid süsteeme umbes 26% ettevõtte töötajatest. Seega ulatub ERP-süsteemi juurutamise kogumaksumus 100 kasutajaga organisatsiooni puhul ligikaudu 900 000 dollarini.

Investeeringud suletud ja moodulitest koosnevatesse suletud lahendustesse on üha vähem õigustatud kaasaegsete, paindlike ja avatud tehnoloogiate kiire arengu taustal. Kui sellised investeeringud on juba tehtud, on oluline objektiivselt ümber hinnata olemasolevate süsteemide rolli: kas need on pikas perspektiivis endiselt olulised või saab nende funktsioone üle vaadata ja tõhusamalt ja läbipaistvamalt rakendada.

Üks tänaste moodulipõhiste andmeplatvormide põhiprobleemidest on see, et nad tsentraliseerivad andmealduse suletud rakendustes. Selle tulemusel muutuvad andmed - ettevõtte põhivara - pigem sõltuvaks konkreetsetest tarkvaralahendustest kui vastupidi. See piirab teabe taaskasutamist, raskendab migratsiooni ja vähendab ettevõtte paindlikkust kiiresti muutuval digitaalsel maastikul.

Kui on tõenäoline, et suletud moodularhitektuuri tähtsus või asjakohasus tulevikus väheneb, on mõistlik tunnista täna tehtud kulutusi kui uppunud kulusid ja keskenduda strateegilisele üleminekule avatumale, skaleeritavamale ja kohanemisvõimelisemale digitaalsele ökosüsteemile.

Kaitstud tarkvara iseloomustab arendaja ainukontroll lähtekoodi ja kasutajaandmete üle, mis on loodud selliste lahenduste kasutamise käigus. Erinevalt avatud lähtekoodiga tarkvarast ei ole kasutajatel juurdepääsu rakenduse sisemisele struktuurile ning nad ei saa seda iseseisvalt läbi vaadata, muuta või kohandada oma vajadustele. Selle asemel peavad nad ostma litsentsid, mis annavad õiguse kasutada tarkvara müüja kehtestatud piirides.

Kaasaegne andmekeskne lähenemisviis pakub teistsugust paradigmat: andmeid tuleks vaadelda kui strateegilist põhivara - sõltumatut, püsivat ja konkreetsetest tarkvaralahendustest eraldiseisvat. Rakendused omakorda muutuvad pelgalt andmetötlusvahenditeks, mida saab vabalt asendada, ilma et tekiks oht, et kriitiline teave läheb kaduma.

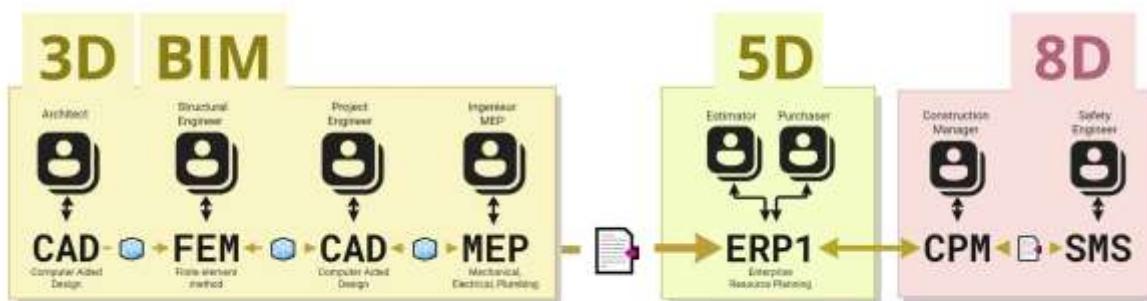
ERP- ja MRP-süsteemide väljatöötamine 1990ndatel (joonis 1.2-1) andis ettevõtetele võimsad vahendid protsesside juhtimiseks, kuid selle tahtmatu tagajärjeks oli ka infovoogude haldamisega seotud inimeste arvu märkimisväärne suurenemine. Selle asemel, et automatiseerida ja lihtsustada tööülesandeid, tekitasid need süsteemid sageli uue taseme keerukust, bürokraatiat ja sõltuvust sisemistest IT-ressurssidest.

Andmehaldussüsteemid: andmekaevandamisest äritegevuse väljakutseteni

Tänapäeva ettevõtted seisavad silmitsi vajadusega integreerida mitu andmehaldussüsteemi. Andmehaldussüsteemide valimine, nende süsteemide mõistlik haldamine ja erinevate andmeallikate integreerimine on muutunud äritegevuse tulemuslikkuse seisukohalt kriitilise tähtsusega.

2020. aastate keskel võib leida sadu (suurtes ehitusettevõtetes tuhandeid) erinevaid süsteeme (joonis 1.2-3), mis peavad töötama harmooniliselt, et ehitusprotsessi kõik aspektid sujuvalt ja ühtselt toimiksid.

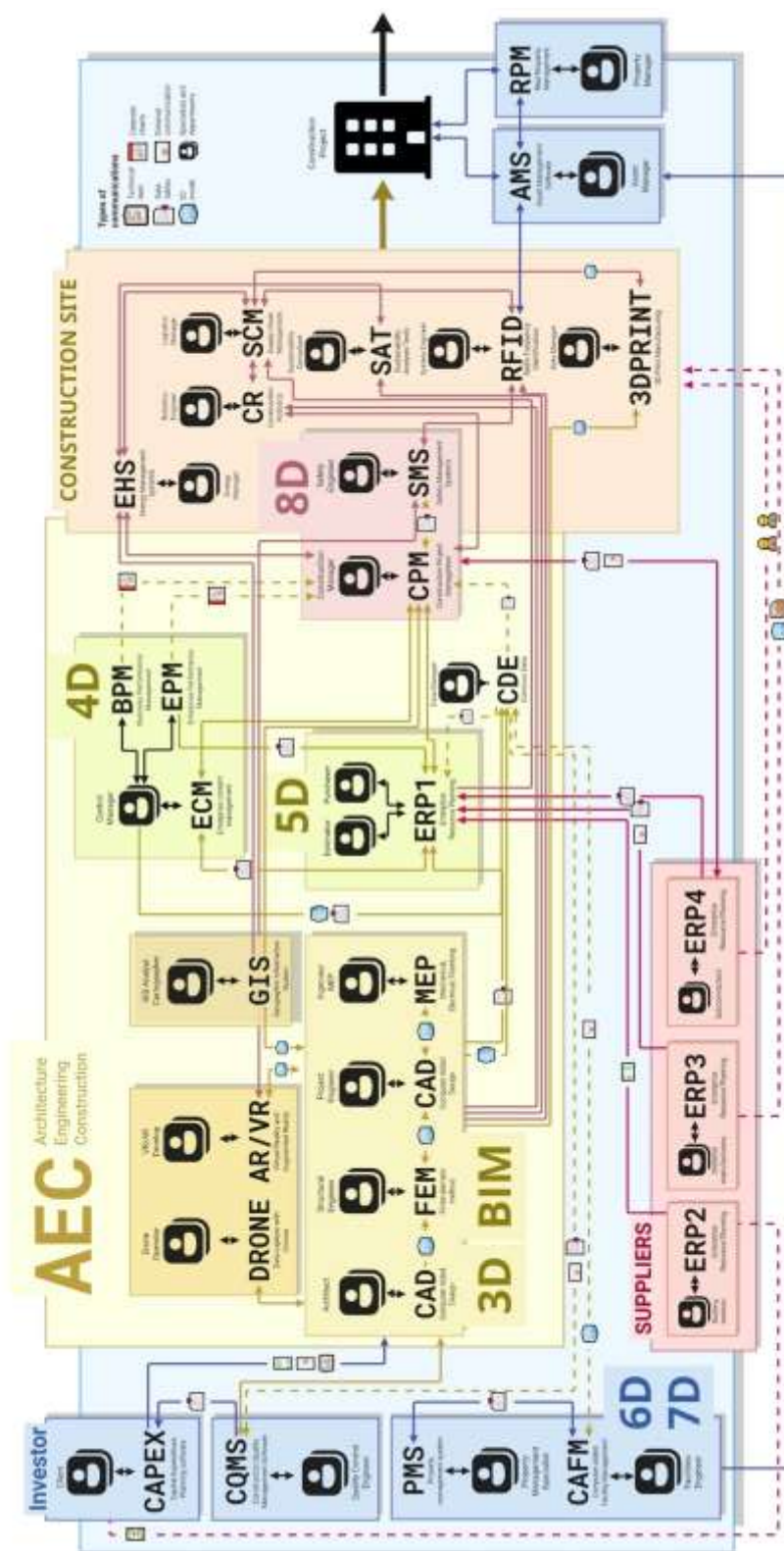
Deloitte'i 2016. aasta uuringu® "Data-Driven Management in Digital Capital Projects" kohaselt kasutab keskmine ehitusspetsialist päevas 3,3 tarkvararakendust, kuid ainult 1,7 neist on omavahel integreeritud [17].



Joonis 1.2-3 Iga ärisüsteem vajab kvaliteetseks andmehalduseks professionaalset meeskonda ja vastutavat juhti.

Järgnevalt on esitatud loetelu populaarsetest süsteemidest, mida kasutatakse ehitussektori keskmise suurusega ja suurte ettevõtete jaoks ehitusprojektide tõhusaks juhtimiseks:

- **ERP (Enterprise Resource Planning)** - pakub äriprotsesside, sealhulgas raamatupidamise, hangete ja projektijuhtimise integreerimist.
- **CAPEX (Capital Expenditure Planning Software)** - kasutatakse ehitusprojektide finantsinvesteeringute eelarvestamiseks ja haldamiseks, aitab määrata põhivara ja investeeringute maksumust pikaajalistesse varadesse.
- **CAD (Computer-Aided Design)** ja **BIM (Building Information Modeling)** - kasutatakse projektide üksikasjalike ja täpsete tehniliste jooniste ja 3D -mudelite loomiseks. Nende süsteemide põhirõhk on töö geomeetrilise teabega.
- **MEP (Mehaanika, elektri- ja veevarustus)** - tehnilised süsteemid, mis hõlmavad mehaanilisi, elektri- ja veevarustuskomponente ning kirjeldavad üksikasjalikult projekti sisemist "ringlussüsteemi".
- **GIS (Geograafilised infosüsteemid)** - kasutatakse maastiku analüüsiks ja planeerimiseks, sealhulgas kartograafia ja ruumiline analüüs.
- **CQMS (ehituskvaliteedi juhtimise tarkvara)** - tagab, et ehitusprotsessid vastavad kehtestatud standarditele ja eeskirjadele, aidates kõrvaldada puudusi.
- **CPM (ehitusprojektide juhtimine)** - hõlmab ehitusprotsesside planeerimist, koordineerimist ja kontrolli.
- **CAFM (Computer-Aided Facility Management)** - hoonete haldus- ja hooldussüsteemid.
- **SCM (Supply Chain Management)** on vajalik, et optimeerida materjalide ja teabe liikumist tarnijate ja ehitusplatsi vahel.
- **EPM (Enterprise Performance Management)** - eesmärk on parandada äriprotsesse ja tulemuslikkust.
- **AMS (Asset Management Software)** - kasutatakse seadmete ja infrastruktuuri kasutamise, haldamise ja hoolduse optimeerimiseks kogu vara elutsükli jooksul.
- **RPM (kinnisvarahaldus)** - hõlmab ülesandeid ja protsesse, mis on seotud hoonete ja maa ning nendega seotud ressursside ja varade haldamise ja käitamisega.



Joonis 1.2-4 Süsteemide omavaheline seotus, mis ühendab ettevõtte protsessid ja teabevoe eri osakondade vahel.

- **CAE (Computer-Aided Engineering)** - Arvutipõhine projekteerimine, hõlmab arvutuslikke ja simulatsioonisüsteeme, nagu lõplike elementide analüüs (FEA) ja arvutuslik vedeliku dünaamika (CFD).
- **CFD (Computational Fluid Dynamics)** - arvutuslik vedeliku dünaamika, vedeliku- ja gaasivoolude modelleerimine. CAE alamkategooria.
- **CAPP (Computer-Aided Process Planning)** - arvutipõhine protsesside planeerimine. Seda kasutatakse marsruutide ja protsessikaartide koostamiseks.
- **CAM (Computer-Aided Manufacturing)** - arvutipõhine tootmine, CNC-pinkide juhtimisprogram-
mide koostamine.
- **PDM (Product Data Management)** - tooteandmete haldamine, tehnilise dokumentatsiooni säili-
tamise ja haldamise süsteem.
- **MES (Manufacturing Execution System)** on reaalajas toimiv tootmisprotsessi
juhtimissüsteem.
- **PLM (Product Lifecycle Management)** - projektielemendi elutsükli haldamine, integreerib PDM,
CAPP, CAM ja muud süsteemid toote täielikuks kontrollimiseks arendusest kuni kõrval-
damiseni.

Need ja paljud teised süsteemid, mis sisaldavad erinevaid tarkvaralahendusi, on muutunud kaasaegse ehitustööstuse lahutamatuks osaks (joonis 1.2-4). Nende süsteemide tuum on spetsialiseeritud andmebaasid, millel on intuiitsed kasutajaliidesed, mis võimaldavad tõhusat teabe sisestamist, töötlemist ja analüüsi kõigis projekteerimis- ja ehitusetappides. Digitaalsete vahendite omavaheline integreerimine ei aita mitte ainult optimeerida tööprotsesse, vaid parandab oluliselt ka otsuste täpsust, mis mõjutab positiivselt projekti elluviimise aega ja kvaliteeti.

Kuid pooltel juhtudel puudub integratsioon. Statistika kohaselt on ainult iga teine rakendus või süsteem integreeritud teiste lahendustega [17]. See näitab digitaalse keskkonna jätkuvat killustatust ja rõhutab vajadust töötada välja avatud standardid ja ühtsed liidesed, et tagada läbiv teabevahetus ehitusprojekti.

Üheks peamiseks integreerimise väljakutseks kaasaegsete ettevõtete jaoks on endiselt digi-
taalsüsteemide suur keerukus ja nõuded kasutajate pädevusele, mis on vajalikud teabe tõhusaks
leidmiseks ja tõlgendamiseks. Iga ettevõttes kasutusele võetud süsteemi toetamiseks moodustatakse
spetsialistide meeskond, mida juhib võtmejuht (joonis 1.2-2).

Võtmesüsteemi halduril on otsustav roll andmevoo õiges suunas ja ta vastutab lõpliku
teabe kvaliteedi eest, nii nagu esimesed haldurid vastutasid tuhandeid aastaid tagasi pa-
püürusele või savitahvlitele kirjutatud numbrite eest.

Et muuta erinevad infovood juhtimisvahendiks, on oluline oskus andmeid süstemaatiliselt integreerida
ja hallata. Selles arhitektuuris peavad juhid tegutsema ühtse võrgustiku elementidena - nagu mütseel,
mis ühendab ettevõtte üksikud osad ühtseks elavaks organismiks, mis on võimeline kohanema ja
arenema.

Ettevõtte mütseel: kuidas andmed on seotud äriprotsessidega

Andmete integreerimine rakendustesse ja andmebaasidesse põhineb erinevatest allikatest, sealhulgas erinevatest osakondadest ja spetsialistidest pärineva teabe koondamisel (joonis 1.2-4). Spetsialistid otsivad asjakohaseid andmeid, töötlevad neid ja edastavad need oma süsteemidesse ja rakendustesse edasiseks kasutamiseks.

Iga ettevõtte süsteem, mis koosneb tööriistade, tehnoloogiate ja andmebaaside kogumist, on teadmiste puu, mis on juurdunud ajalooliste andmete pinnasesse ja kasvab, et kanda uusi vilju valmislahenduste kujul: dokumendid, arvutused, tabelid, graafikud ja armatuurlauad (joonis 1.2-5). Ettevõtte süsteemid, nagu puud konkreetsetes metsas, suhtlevad ja suhtlevad omavahel, moodustades keerulise ja hästi struktureeritud süsteemi, mida toetavad ja juhivad asjatundlikud juhid.

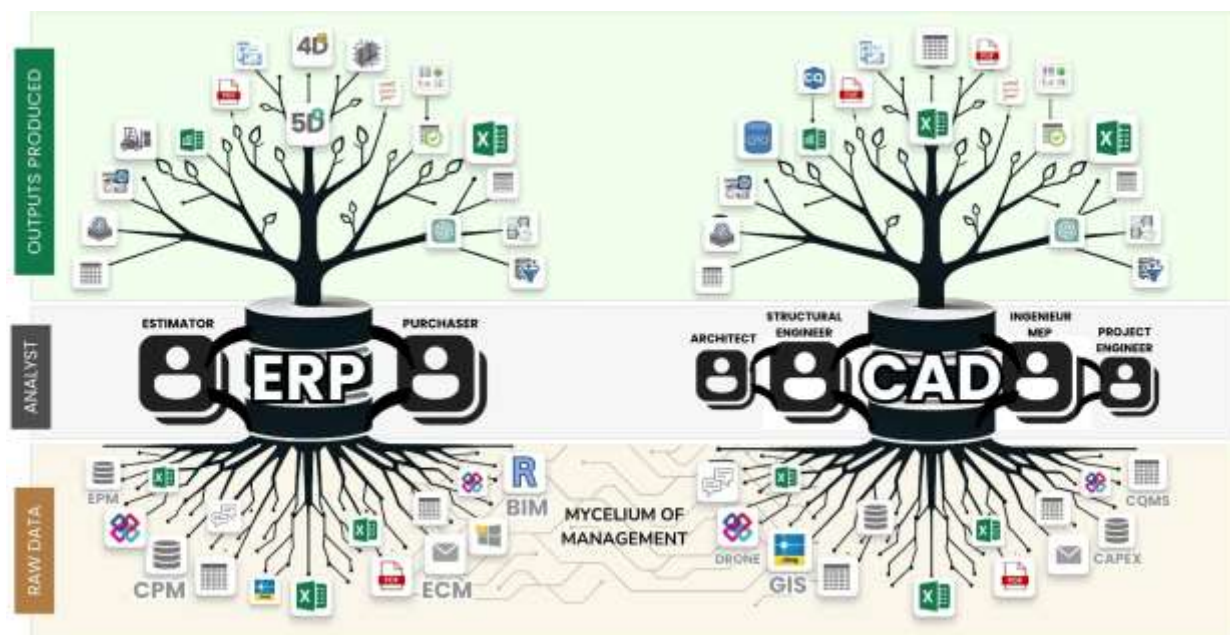
Ettevõtte infootsingu- ja teabeedastussüsteem toimib nagu keerukas metsavõrk, mis koosneb puudest (süsteemid) ja seenestest (juhid), mis toimivad juhtidena ja taaskasutajatena, tagades, et teave edastatakse ja liigub õigetes süsteemidesse. See aitab säilitada tervislikku ja tõhusat andmevoolu ja -jaotust ettevõttesiseselt.

Eksperdid, nagu juured, imavad projekti algstaadiumis toorandmeid, muutes need ettevõtte ökosüsteemi toitaineteks. Andme- ja sisuhaldussüsteemid (joonis 1.2-4 - ERP, CPM, BIM jne) toimivad võimsate infoteedena, mille kaudu need teadmised liiguvad läbi ettevõtte kõigi tasandite.

Nii nagu looduses, kus iga ökosüsteemi element mängib oma rolli, aitab ka ettevõtte ärimaastikul iga protsessis osaleja - alates insenerist kuni analüütikuni - kaasa teabekeskkonna kasvule ja viljakusele. Need süsteemsed "andmepuud" (joonis 1.2-5) ei ole lihtsalt teabe kogumise mehhanismid, vaid konkurentsieelis, mis tagab ettevõtte jätkusuutlikkuse.

Metsade ökosüsteemid peegeldavad üllatavalt täpselt seda, kuidas digitaalsed ettevõtete struktuurid on korraldatud. Sarnaselt metsa astmelisele struktuurile - aluskasvust kuni puude ladvuni - jaotab ettevõtte juhtimine ülesanded vastutustasandite ja funktsionaalsete osakondade vahel.

Sügavad ja hargnevad puujuured tagavad vastupidavuse ja juurdepääsu toitainetele. Samamoodi toetavad kindel organisatsiooniline struktuur ja stabiilsed protsessid kvaliteetsete andmetega töötamiseks kogu ettevõtte infoökosüsteemi, aidates kaasa selle jätkusuutlikule kasvule ja arengule isegi (suure tuule) turu ebastabiilsuse ja kriiside ajal.



Joonis 1.2-5 Andmete integreerimine erinevate süsteemide kaudu on nagu mütseel, mis ühendab juhid ja spetsialistid ühtseks infovõrgustikuks.

Kaasaegne arusaam mastaabist äris on arenenud. Tänapäeval ei määra ettevõtte väärtust mitte ainult selle nähtav osa - "kroonid" lõppdokumentide ja aruannete kujul - vaid ka kvalitatiivselt kogutud ja süstemaatiliselt töödeldud andmete "juursüsteemi" sügavus. Mida rohkem teavet on võimalik koguda ja töödelda, seda suurem on ettevõtte väärtus. Ettevõtted, kes koguvad metoodiliselt juba töödeldud andmete "komposti" ja suudavad sellest kasulikke teadmisi ammutada, saavad strateegilise eelise

Ajaloolisest teabest on saamas uut liiki kapital, mis võimaldab kasvu, protsesside optimeerimist ja konkurentsieeliseid. Andmetel põhinevas maailmas ei ole oluline, kellel on rohkem, vaid kes teab rohkem, see võidab.

Ehitustööstuse jaoks tähendab see üleminekut reaajas toimuvale projektijuhtimisele, kus kõik protsessid - alates projekteerimisest ja hangetest kuni töövõtjate koordineerimiseni - põhinevad asjakohastel, igapäevaselt ajakohastatud andmetel. Erinevatest allikatest (ERP -süsteemid, CAD - mudelid, andurid IoT ehitusplatsidel, RFID) pärineva teabe integreerimine võimaldab teha täpsemaid prognoose, reageerida kiiresti muutustele ja vältida ajakohaste andmete puudumisest tingitud viivitusi.

Vastavalt McKinsey & Company andmekesksele ettevõttele 2025 (McKinsey & Company®, 2022 [18]) tuginevad tuleviku edukad ettevõtted andmetele kõigis oma tegevuse võtmeaspektides, alates strateegilistest otsustest kuni operatiivse suhtlemiseni.

Andmed ei ole enam pelgalt analüüsivahend, vaid muutuvad kõigi äriprotsesside lahutamatuks osaks, tagades läbipaistvuse, kontrolli ja juhtimise automatiseerimise. Andmetel põhinev liikumine võimaldab organisatsioonidel vähendada inimtegu mõju, vähendada tegevusriske ning suurendada otsuste

tegemise läbipaistvust ja tõhusust.

21. sajand on keernud majandusliku paradigma pea peale: kui varem nimetati naftat "mustaks kullaks" selle võime tõttu masinate ja transpordi käivitamiseks, siis tänapäeval on ajalise surve all kokkusurutud ajaloolised andmed muutumas uueks strateegiliseks ressursiks, mis toidab mitte masinaid, vaid otsuste tegemise algoritme, mis hakkavad ettevõtlust juhtima.



PEATÜKK 1.3.

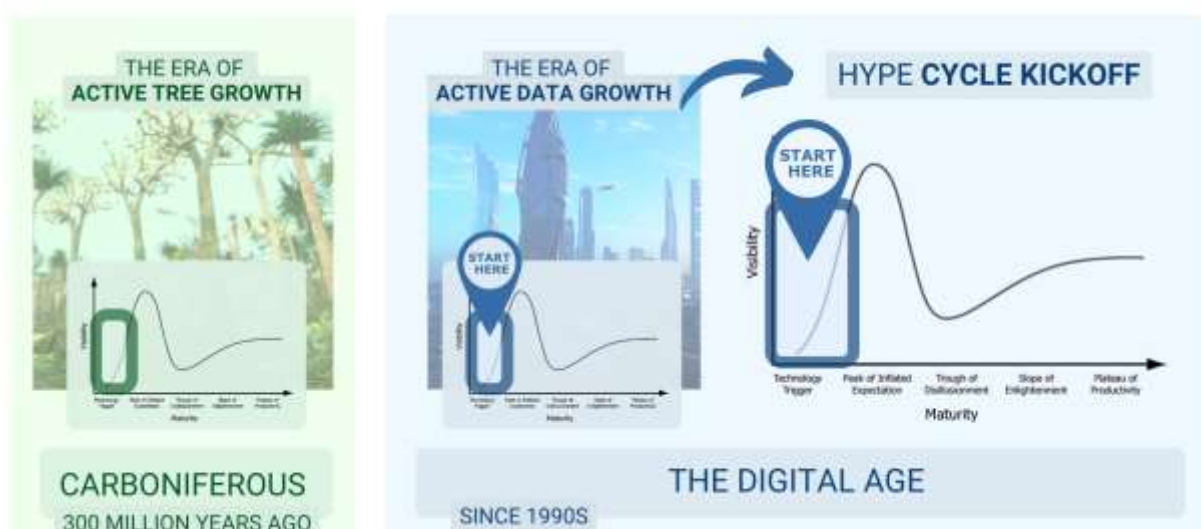
DIGITAALNE REVOLUTSIOON JA ANDMETE PLAHVATUSLIK KASV

Andmabuumi algus kui evolutsioonilaine

Ehitustööstus kogeb enneolematut infoplahvatust. Kui me mõtleme ettevõtlusest kui teadmistepuust (joonis 1.2-5), mida toidavad andmed, võib digitaliseerimise praegust etappi võrrelda taimestiku kiire kasvuga süsinikdioksiidi perioodil, ajastul, mil Maa biosfääri muutis kiire biomassi kogunemine (joonis 1.3-1).

Ülemaailmse digitaliseerimisega kahekordistub ehitussektoris igal aastal teabe hulk. Kaasaegne tehnoloogia võimaldab andmeid koguda taustal, analüüsida neid reaalsajas ja kasutada neid sellises ulatuses, mis veel mõni aeg tagasi tundus võimatu.

Vastavalt Gordon Moore'i seadusele, mille sõnastas Gordon Moore (Inteli® kaasasutaja), kahekordistub integraallülituste tihedus ja keerukus ning töödeldavate ja salvestatavate andmete hulk umbes iga kahe aasta tagant [19].



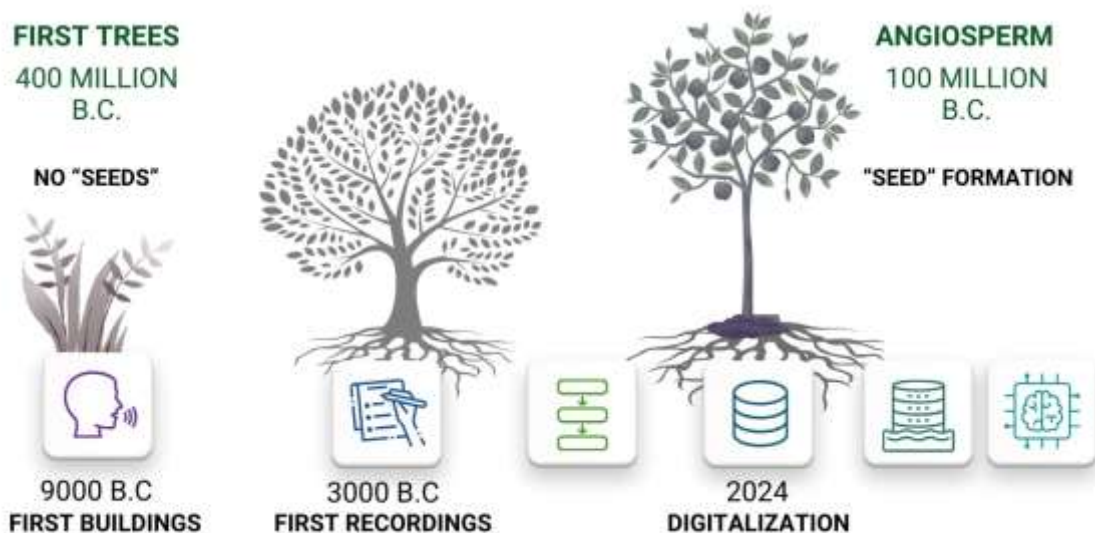
Joonis 1.3-1 Digitaliseerimise algus on toonud kaasa andmete eksponentsiaalse kasvu, nagu taimestiku hüppeline kasvamine söeajastul.

Kui muistsed megaliithitised, nagu Göbekli Tepe (Türgi), ei jätnud endast maha dokumenteeritud teadmisi, mis sobiksid taaskasutamiseks, siis tänapäeval võimaldavad digitaaltehnoloogiad koguda ja taaskasutada teavet. Seda võib võrrelda evolutsioonilise üleminekuga spoortaimedelt seemnetaimedele (angiospermidele): seemnete tekkimine andis aluse elu laialdasele levikule planeedil. (Joonis 1.3-2).

Sarnaselt muutuvad varasemate projektide andmed omamoodi "digitaalseks seemneks" - DNA teadmistekandjateks, mida saab skaleerida ja kasutada uutes projektides ja projektides duces. Kaasaegsete tehisintellekti vahendite - masinõppe ja suurte keelemudelite (LLM), nagu ChatGPT,

LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok - teke võimaldab andmeid automaatselt ekstraheerida, tõlgendada ja rakendada uutes kontekstides

Nii nagu seemned muutsid elu leviku algselt eluta planeedil revolutsiooniliseks, muutuvad "andmeseemned" uute teabestruktuuride ja teadmiste automaatse tekkimise aluseks, võimaldades digitaalsetel ökosüsteemidel iseseisvalt areneda ja kohaneda muutuvate kasutajanõudmistega.



Joonis 1.3-2 Digitaalsed "andmeseemned" mängivad evolutsiooniliselt sama rolli kui angiospermid, õistaimed, mis muutsid Maa ökosüsteemi.

Seisame ehituses uue ajastu künnisel, kus andmete plahvatuslik kasv ja "andmeseemnete" - varasematest ja käimasolevatest projektidest saadud struktureeritud teabe - aktiivne levitamine on tööstuse digitaalse tuleviku aluseks. Nende "tolmeldamine" suurte andmemudelite kaudu võimaldab meil mitte ainult jälgida digitaalseid muutusi, vaid osaleda aktiivselt iseõppivate, kohanemisvõimeliste ökosüsteemide loomisel. See ei ole evolutsioon - see on digitaalne revolutsioon, kus andmed muutuvad uue reaalsuse peamiseks ehitusplokiks

Andmete hulk ehitussektoris suureneb järsult tänu eri valdkondade teabele kogu ehitusprojekti elutsükli jooksul. See tohutu andmete kogunemine on lükanud ehitussektori suurandmete ajastu poole [20].

- Prof Hang Yang, Wuhan Tehnoloogiaülikooli tsiviilehituse ja arhitektuuri osakond, Wuhan,.

Andmete kasv infoajastul meenutab evolutsioonilisi protsesse looduses: nii nagu metsade areng muutis planeedi iidset maastikku, muudab praegune infoplahvatus kogu ehitussektori maastikku.

Kaasaegses ettevõttes tekkivate andmete hulk

Viimase kahe aasta jooksul on 90% kõigist maailmas olemasolevatest andmetest loodud [21]. Alates 2023. aastast tekitab iga inimene, sealhulgas ehitussektori spetsialistid, umbes 1,7 megabaiti andmeid sekundis [22] ja kogu andmemaht maailmas jõuab 2023. aastal 64 zettabaiti ning prognooside kohaselt ületab see 2025. aastaks 180 zettabaiti ehk $180 \cdot 10^{15}$ megabaiti [23].

Sellisel infoplahvatusel on ajalooline pretsedent - trükipressi leiutamine Johannes Gutenbergi poolt XV sajandil. Vaid viiskümmend aastat pärast selle kasutuselevõttu kahekordistus raamatute arv Euroopas: mõne aastakümnega trükiti sama palju raamatuid kui oli käsitsi loodud eelneva 1200 aasta jooksul [24]. Tänapäeval oleme tunnistajaks veelgi kiiremale kasvule: andmete hulk maailmas kahekordistub iga kolme aasta tagant.

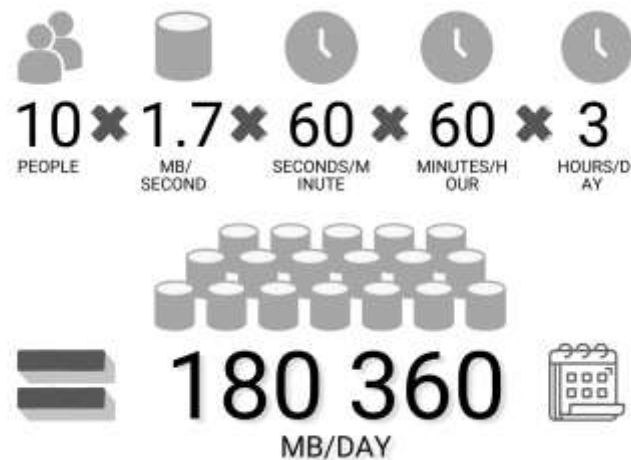
Arvestades praegust andmete kasvutempot, on ehitussektoril potentsiaali luua järgmise paarikümne aasta jooksul sama palju teavet, kui on kogunenud kogu senise ajaloo jooksul



Joonis 1.3-3 Iga töötaja igapäevane andmete salvestamine ettevõtte serveritesse aitab kaasa andmete pidevale kasvule.

Tänapäeva ehitusettevõtete maailmas tekitavad isegi väikesed ettevõtted igapäevaselt tohutult palju multiformaatseid andmeid ning isegi väikese ehitusettevõtte digitaalne jalajälg võib ulatuda kümnetesse gigabaitidesse päevas - alates mudelitest ja joonistest kuni fotodokumentide ja anduriteni kohapeal. Kui eeldada, et iga tehnik tekitab keskmiselt umbes 1,7 MB andmeid sekundis, siis vastab see umbes 146 GB-le päevas või 53 TB-le aastas (joonis 1.3-3).

Kui 10-liikmeline meeskond töötab aktiivselt vaid 3 tundi päevas, ulatub päevas kogutud teabe hulk 180 gigabaiti (joonis 1.3-4).



Joonis 1.3-4 10-liikmeline ettevõtte toodab päevas umbes 50-200 gigabaiti andmeid.

Eeldades, et 30% tööandmetest on uued (ülejäänud kirjutatakse üle või kustutatakse), võib 10-liikmeline ettevõtte luua umbes sadu gigabaiti uusi andmeid kuus (tegelik arv sõltub ettevõtte äritegevusest)

Seega on selge, et me ei genereeri lihtsalt üha rohkem andmeid - meil on üha suurem vajadus nende töhuse haldamise, säilitamise ja pikaajalise kättesaadavuse järele. Ja kui varem võisid andmed tasuta "lebada" kohalikes serverites, siis digitaalse ümberkujundamise kontekstis hakkab üha rohkem ettevõtteid kasutama oma infoinfrastruktuuri alusena pilvelahendusi.

Andmesalvestuse maksumus: majanduslik aspekt

Viimastel aastatel on üha rohkem ettevõtteid ostanud andmete säilitamise pilveteenustesse. Näiteks kui ettevõtte majutab pool oma andmetest pilves, mille keskmine hind on 0,015 dollarit gigabaidi kohta kuus, võivad tema salvestuskulud kasvada 10-50 dollarit [25] kuus.

Väikese ettevõtte jaoks, kellel on tüüpilised andmetöötlusmustrid, võivad pilve salvestuskulud ulatuda sadadest kuni potentsiaalselt üle tuhande dollari kuus (joonis 1.3-5) mõne aasta jooksul, tekitades potentsiaalselt märkimisväärse finantskoormuse.

Forresteri uuringu "Enterprises Outsource Data Storage as Complexity Grows" [26] kohaselt, mille käigus küsitleti 214 tehnoloogilise infrastruktuuri eest vastutavat otsustajat. [26], milles küsitleti 214 tehnoloogiainfrastruktuuri otsustajat, on üle kolmandiku organisatsioonidest andmesalvestuse kasvava mahu ja keerukusega toimetulekuks ostanud andmesalvestuse allhanke, kusjuures peaaegu kaks kolmandikku ettevõtetest eelistab tellimuspõhist mudelit.



Joonis 1.3-5 Andmete kolimine pilve võib suurendada igakuiseid salvestuskulusid kuni 2000 dollarit, isegi kui ettevõttel on vaid 10 töötajat.

Olukorda raskendab veelgi pilvepõhiste tehnoloogiate, nagu CAD (BIM), CAFM, PMIS ja ERP - süsteemide kiirenenud kasutuselevõtt, mis suurendab veelgi andmete säilitamise ja töötlemise kulusid. Selle tulemusena on ettevõtted sunnitud otsima võimalusi kulude optimeerimiseks ja pilveteenuste pakujatest sõltuvuse vähendamiseks.

Alates 2023. aastast on suurte keelemudelite (LLM) aktiivse arendamisega hakanud muutuma lähenemisviisid andmete säilitamisele. Üha enam ettevõtteid mõtleb oma andmete üle kontrolli tagasi võtta, kuna teabe töötlemine oma serverites muutub turvalisemaks ja kasumlikumaks.

Sellega seoses tuleb esile suundumus, et pilvepõhisest salvestamisest ja ainult vajalike andmete töötlemisest loobutakse ja eelistatakse ettevõtte LLM ja AI -lahenduste kohalikku kasutuselevõttu. Nagu Microsofti tegevjuht ühes intervjuus [27] märkis, hakkavad AI-agendid selle asemel, et tugineda erinevate ülesannete täitmiseks mitmetele eraldi rakendustele või pilvepõhistele SaaS lahendustele, haldama protsesse andmebaasides, automatiseerides erinevate süsteemide funktsioone.

[...] vana lähenemine sellele [andmetöötluste] probleemile oli: kui te mõtlete tagasi sellele, kuidas erinevad ärirakendused integratsiooni käsitlesid, siis kasutasid nad ühendajaid. Ettevõtted müüsid nende ühendajate litsentse ja ärimudel kujunes selle ümber. SAP [ERP] on üks klassikaline näide: SAPi andmetele sai ligi ainult siis, kui oli olemas õige ühendaja. Seega tundub mulle, et midagi sarnast tekib ka [AI] agentide suhtluse puhul [...]. Vähemalt meie lähenemine on järgmine: ma arvan, et [AI] agentide ajastul kukub tõenäoliselt kokku ärirakenduste olemasolu kontseptsioon. Sest kui te mõtlete selle peale, siis on nad sisuliselt andmebaasid koos hunniku ärioloogikaga

- Satya Nadella, Microsofti tegevjuht, intervjuu BG2 kanalile, 2024. [28]

Selles paradigmas läheb andmepõhine LLM lähenemine kaugemale klassikalistest süsteemidest. Tehisintellektist saab vahendaja kasutaja ja andmete vahel (joonis 2.2-3, joonis 2.2-4), kõrvaldades vajaduse mitmete vahendusliideste järele ja suurendades äriprotsesside tõhusust. Sellest

lähenedamisest andmetega töötamisele räägime lähemalt peatükis "Kaoe muutmine korrastamiseks ja keerukuse vähendamine".

Kuigi tuleviku arhitektuur on alles kujunemas, seisavad ettevõtted juba praegu silmitsi mineviku otsuste tagajärgedega. Viimaste aastakümnete massiline digitaliseerimine, millega kaasnesid erinevad süsteemid ja andmete kontrollimatu kogunemine, on toonud kaasa uue probleemi - teabe ülekoormuse.

Andmete kogumise piirid: massist tähenduseni

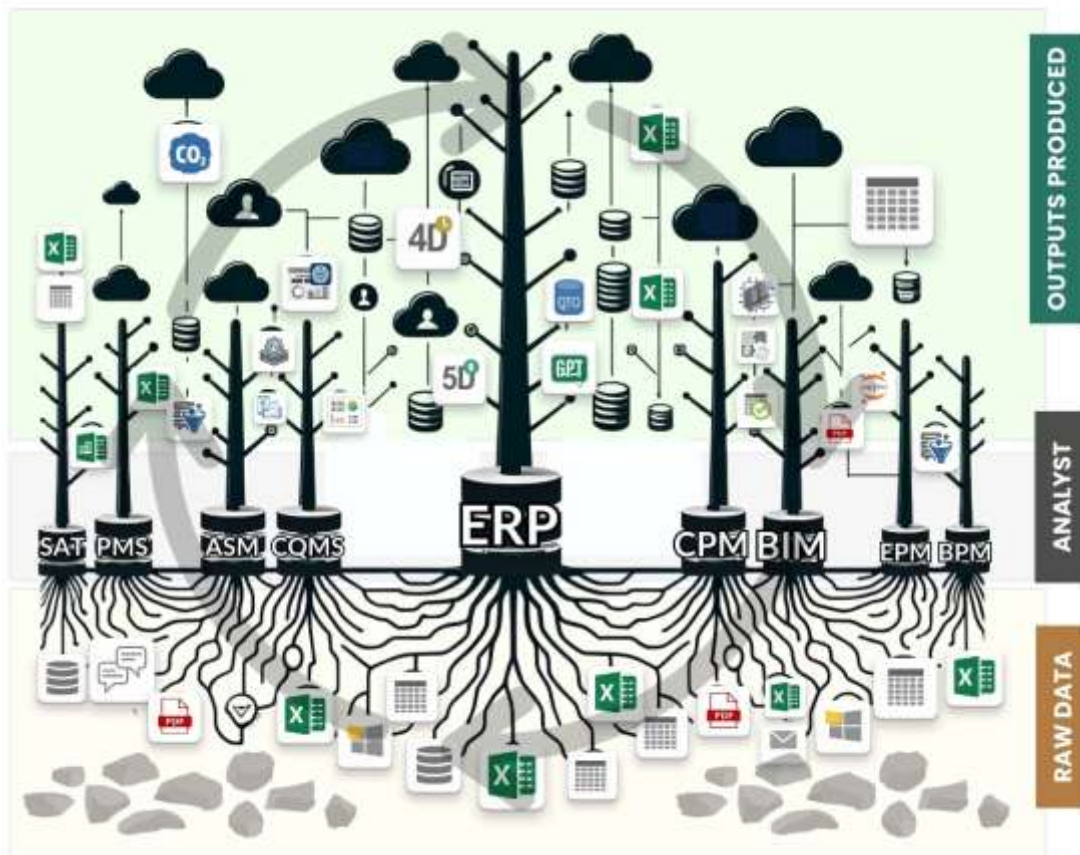
Kaasaegsed ettevõtte süsteemid arenevad ja toimivad edukalt juhitud kasvu tingimustes, kui andmemahut ja rakenduste arv on tasakaalus IT-osakondade ja juhtide võimekusega. Viimastel aastakümnetel on aga digitaliseerimine viinud andmete mahu ja keerukuse kontrollimatule kasvule, mis on põhjustanud ettevõtete infoökosüsteemis üleküllastumise efekti.

Tänapäeval on serverid ja salvestusruumid koormatud enneolematu hulga töötlemata ja mitmevormingulise teabe sissevooluga, millel ei ole aega muutuda kompostiks ja mis muutub kiiresti ebaoluliseks. Ettevõtte piiratud ressursid ei suuda selle tulvaga toime tulla ja andmed kogunevad isoleeritud silodesse (nn "silod"), mis vajavad kasuliku teabe väljavõtmiseks käsitsi töötlemist.

Selle tulemusel kannatavad kaasaegsed ettevõtte juhtimissüsteemid sageli teabe ülekoormuse all, nagu murakastunud ja hallitusega kaetud mets. Ettevõtte ökosüsteemi südames oleva toitva infohüümuse asemel tekivad isoleeritud eriformaadiliste andmete alad, mis paratamatult viib äriprotsesside üldise tõhususe vähenemiseni.

Viimase 40 aasta pikkale eksponentsiaalse andmete kasvu perioodile järgneb paratamatult küllastumis- ja sellele järgnev jahtumishaas. Kui mälu saavutab oma piirid, toimub kvalitatiivne muutus: andmed ei ole enam lihtsalt mäluobjekt, vaid strateegiline ressurss.

Tehisintellekti ja masinõppe arenguga on ettevõtetel võimalus vähendada infotöötluskulusid ja minna kvantitatiivselt kasvult üle andmete kvalitatiivsele kasutamisele. Järgmise kümnendi jooksul peab ehitustööstus keskenduma üha suurema hulga andmete loomiselt nende struktuuri, terviklikkuse ja analüütilise väärtuse tagamisele.



Joonis 1.3-6 Isoleeritud andmeallikad takistavad infosüsteemide vahelist teabevahetust.

Peamine väärtus ei seisne enam mitte teabemahus, vaid suutlikkuses seda automaatselt tõlgendada ja muuta see rakenduslike otsuste tegemisel kasulikuks teadmiseks. Selleks, et andmed oleksid tõeliselt kasulikud, tuleb neid nõuetekohaselt hallata: koguda, kontrollida, struktureerida, salvestada ja analüüsida konkreetsete äritegevusega seotud ülesannete kontekstis.

Andmeanalüüsi protsess ettevõttes on sarnane metsas olevate puude elutsükli ja lagunemise ning uute noorte ja tugevate puude tekkimise tsükliga: küpsed puud surevad ära, lagunevad ja muutuvad uue kasvu kasvulavaks. Lõpetatud ja lõpetatud protsessid muutuvad pärast nende lõpetamist ettevõtte teabeökosüsteemi osaks, muutudes lõpuks teabehuumuseks, mis toidab uute süsteemide ja andmete tulevast kasvu.

Praktikas on see tsükkel aga sageli katkenud. Orgaanilise uuenemise asemel tekib kihiline kaos, nagu geoloogilised kihid, kus uued süsteemid ladestuvad vanade peale ilma sügava integratsiooni ja struktureerimiseta. Selle tulemusel tekivad erinevad teabesilod, mis takistavad teadmiste ringlust ja raskendavad andmehaldust.

Järgmised sammud: andmete teooriast praktiliste muutusteni

Andmete areng ehituses on teekond savitahvlitelt kaasaegsete moodulplatvormideni. Tänapäeval ei ole väljakutse mitte teabe kogumine, vaid raamistiku loomine, mis muudab erinevad ja erinevad andmed strateegiliseks ressursiks. Olenemata sellest, kas teie roll on ettevõtte juhi või inseneri roll, andmete väärtuse mõistmine ja nendega töötamine on tulevikus üks võtmeoskusi.

Selle osa kokkuvõtteks tasub välja tuua peamised praktilised sammud, mis aitavad teil rakendada käsitletud lähenemisviise oma igapäevastes ülesannetes:

■ Teabevoogude isikliku auditi läbiviimine

- ☐ Koostage nimekiri kõigist süsteemidest ja rakendustest, millega te igapäevaselt töötate.
- ☐ Märkige, kus te kulutate kõige rohkem aega andmete otsimisele või topeltkontrollile.
- ☐ Määrake kindlaks oma peamised teabeallikad
- ☐ Analüüsige oma praegust rakenduste maastikku, et leida üleliigseid ja dubleerivaid funktsioone.

■ Püüdke liikuda protsesside kaudu analüütilise küpsuse tasemete kaupa.

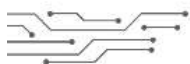
- ☐ Alustage oma ülesandeid kirjeldava analüütikaga (mis juhtus?)
- ☐ järk-järgult kasutusele diagnostika (miks see juhtus?)
- ☐ Mõelge, kuidas protsessides saab liikuda prognoosiva (mis juhtub?) ja ettekirjutava (mida teha?) analüütika suunas.

■ Alustage oma tööandmete struktureerimist

- ☐ Võtke kasutusele ühtne süsteem oma töös sageli kasutatavate failide ja kaustade nimetamiseks.
- ☐ Looge sageli kasutatavate dokumentide ja aruannete mallid
- ☐ Arhiveerige regulaarselt lõpetatud projekte selge struktuuriga

Isegi kui te ei saa muuta kogu oma meeskonna või ettevõtte infoinfrastruktuuri, alustage omaenda protsessidest ja väikestest parandustest oma igapäevatöös. Pidage meeles, et andmete tegelik väärtus ei seisne nende mahus, vaid suutlikkuses saada neist kasulikke teadmisi. Isegi väikestel, kuid korralikult struktureeritud ja analüüsitud andmekogumitel võib olla märkimisväärne mõju, kui need integreeritakse otsustusprotsessidesse.

Raamatu järgmistes osades liigume edasi konkreetsete andmetega töötamise meetodite ja tööriistade juurde, vaatleme võimalusi struktureerimata teabe muutmiseks struktureeritud kogumiteks, uurime analüütika automatiseerimise tehnoloogiaid ja kirjeldame üksikasjalikult, kuidas ehitada ehitusettevõttes üles tõhus analüütika ökosüsteem.





II OSA

KUIDAS EHITUSETTEVÕTLUS UPPUB ANDMEKAOSSESSE

Teine osa on pühendatud ehitusettevõtete ees seisvate probleemide kriitilisele analüüsile, mis on seotud suurenevate andmemahutude käsitlemisega.

Üksikasjalikult uuritakse teabe killustatuse tagajärgi ja "andmete silo" nähtust, mis takistab tõhusat otsustamist. Uuritakse HiPPO -lähenemisviisi (Highest Paid Person's Opinion) probleeme ja selle mõju juhtimisotsuste kvaliteedile ehitusprojektides. Hinnatakse dünaamiliste äriprotsesside ja nende kasvava keerukuse mõju teabevoogudele ja tegevuse tõhususele. Tuuakse konkreetseid näiteid selle kohta, kuidas süsteemi liigne keerukus suurendab kulusid ja vähendab organisatsioonide paindlikkust. Erilist tähelepanu pööratakse piirangutele, mis on tingitud konfidentsiaalsetest vormingutest, ja avatud standardite kasutamise väljavaadetele ehitussektoris. Esitatakse kontseptsioon, mis käsitleb üleminekut tehisintellekti- ja LLM-põhiste tarkvaraökosüsteemidele, mis vähendavad liigset keerukust ja tehnilisi takistusi.

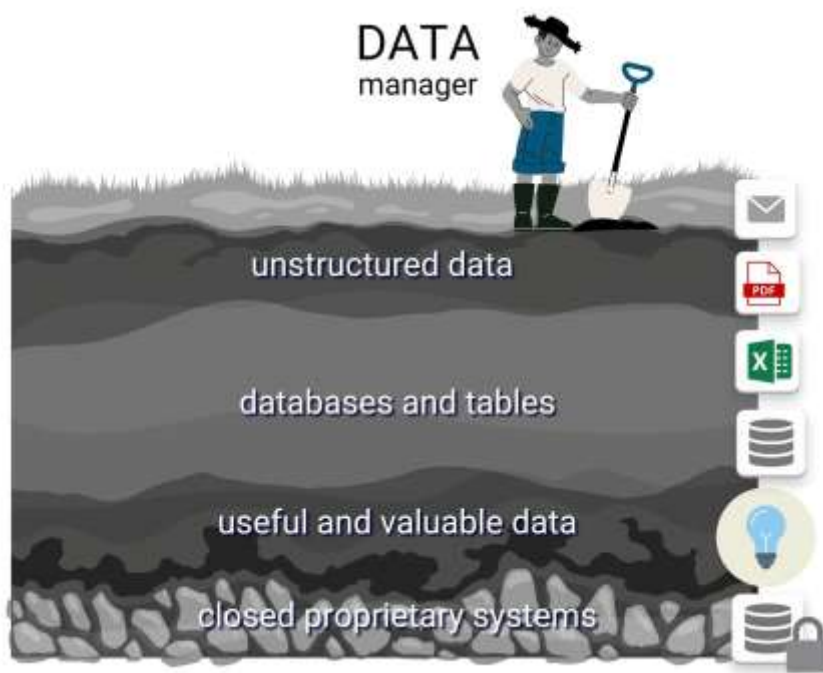
PEATÜKK 2.1.

ANDMETE KILLUSTATUS JA SILOD

Mida rohkem vahendeid, seda tõhusam on äri?

Esmapilgul võib tunduda, et rohkem digitaalseid vahendeid toob kaasa suurema tõhususe. Tegelikuses see siiski nii ei ole. Iga uue lahendusega, olgu see siis pilveteenus, pärandipõhine süsteem või järjekordne Exceli aruanne, lisab ettevõtte oma digitaalsele maastikule veel ühe kihi - kihi, mis sageli ei ole integreeritud teistega (joonis 2.1-1).

Andmeid võib võrrelda kivisöe või naftaga: nende kogunemine võtab aastaid, need on tihenenud kaose, vigade, struktureerimata protsesside ja unustatud vormingute kihtide alla. Et sellest tõeliselt kasulikku teavet välja võtta, peavad ettevõtted sõna otseses mõttes läbi tungima läbi aegunud lahenduste ja digitaalse müra kihtide.



Joonis 2.1-1 Mitmesugused andmed moodustavad killustunud kihid - isegi "kuldseid" arusaamad kaovad süsteemse keerukuse geoloogilistes kivimites.

Iga uus rakendus jätab endast maha jälje: faili, tabeli või terve isoleeritud "silo" serveris. Üks kiht on savi (vananenud ja unustatud andmed), teine on liiv (erinevad tabelid ja aruanded) ja kolmas on graniit (suletud patenteeritud vormingud, mida ei saa integreerida). Aja jooksul meenutab ettevõtte digitaalne keskkond üha enam kontrollimatult kuhjunud teabe reservuaari, kus väärtus on kadunud sügavale ettevõtte serveritesse.

Iga uue projektiga ja iga uue süsteemiga muutub keerulisemaks mitte ainult infrastruktuur, vaid ka tee

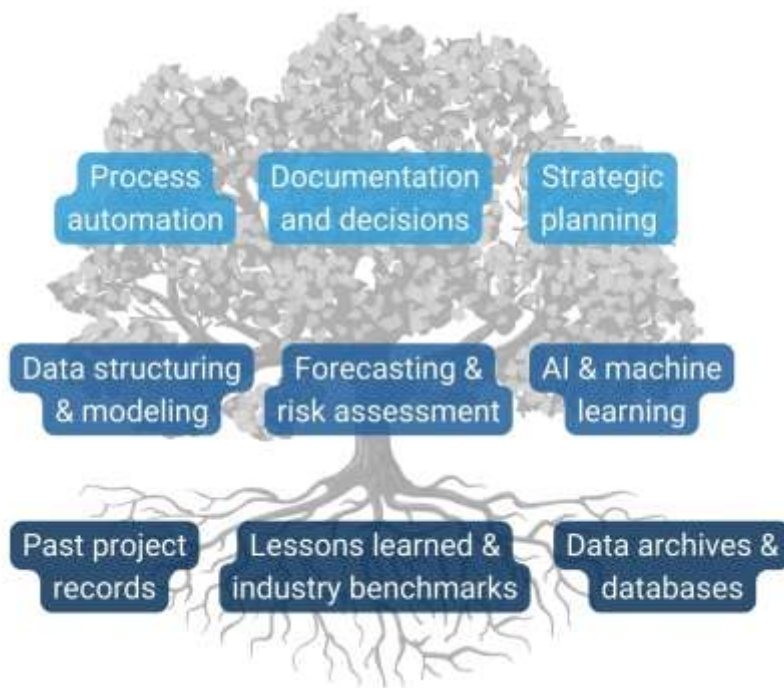
kasuliku kvaliteediga andmetele. Väärtusliku "kivini" jõudmine nõuab põhjalikku puhastamist, teabe struktureerimist, "tükeldamist", rühmitamist mõttekateks tükkideks ja strateegiliselt oluliste arusaamade väljatoomist analüütika ja andmemodelleerimise abil.

Andmed on väärtuslik asi ja need kestavad kauem kui süsteemid [mis andmeid töötlevad] ise [29].

- Tim Berners-Lee, World Wide Webi isa ja esimese veebilehe looja

Enne kui andmed saavad "väärtuslikuks asjaks" ja usaldusväärseks aluseks otsuste tegemisel, tuleb neid hoolikalt ette valmistada. Just korralik eeltöötlus muudab ebaühtlased andmed struktureeritud kogemusteks, kasulikuks teabehuumuseks, millest saab prognoosimis- ja optimeerimisvahend.

On levinud väärarusaam, et analüüsi alustamiseks on vaja täiesti puhtaid andmeid, kuid praktikas on mustade andmetega töötamine protsessi oluline osa.



Joonis 2.1-2 Andmed on äritegevuse juursüsteem ja alus, mis omakorda põhineb otsustusprotsessidel.

Kuna tehnoloogia areneb pidevalt edasi, peab ka teie ettevõtte edasi liikuma ja õppima, kuidas andmete abil väärtust luua. Nii nagu nafta- ja söefirmad ehitavad infrastruktuuri maavarade kaevandamiseks, peavad ka ettevõtted õppima, kuidas hallata uue teabe voogu oma serverites ja ammutada

väärtuslikke teadmisi kasutamata, vormimata ja vananenud andmetest, muutes need strateegiliseks ressursiks.

Esimene samm on väljade loomine (andmelaod). Isegi kõige võimsamad tööriistad ei lahenda andmete eraldatuse ja mitmeformaadiliste andmete probleemi, kui ettevõtted jätkavad siloorsete süsteemidega töötamist. Kui andmed eksisteerivad üksteisest eraldi, ilma et need ristuksid ja jagaksid teavet, seisavad ettevõtted silmitsi "andmesilo" efektiga. Ühtse ja järjepideva infrastruktuuri asemel on ettevõtted sunnitud kulutama ressursse andmete ühendamisele ja sünkroniseerimisele.

Andmesilod ja nende mõju ettevõtte tulemuslikkusele

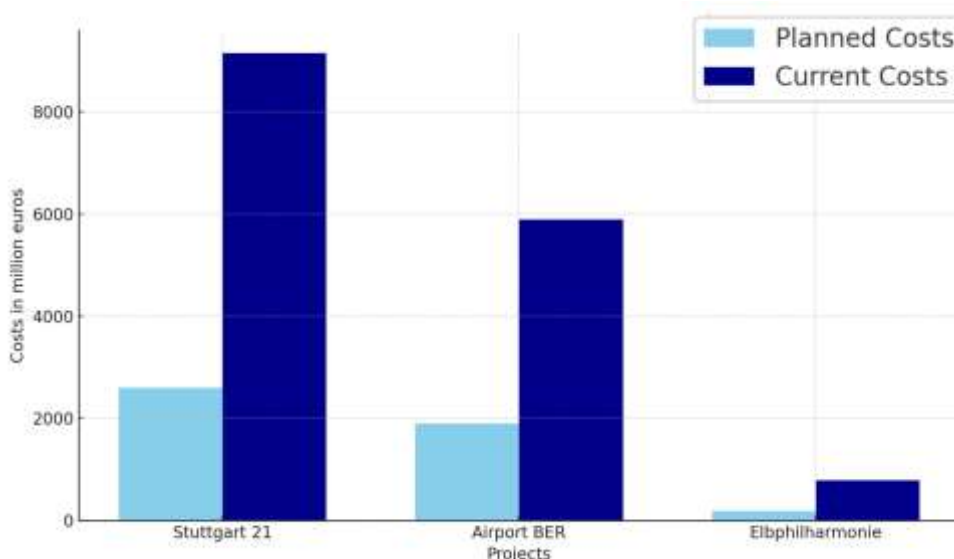
Kujutage ette, et te ehitate elamurajooni, kuid igal meeskonnal on oma projekt. Mõned ehitavad seinu, teised rajavad kommunikatsioone ja kolmandad sillutavad teid, ilma et nad teineteisega konsulteeriksid. Selle tulemusel ei vasta torud seinte avadele, liftišahtid ei vasta korrustele ja teed tuleb lammutada ja uuesti rajada.

Selline olukord ei ole pelgalt hüpoteetiline stsenaarium, vaid paljude kaasaegsete ehitusprojektide tegelikkus. Erinevate süsteemidega töötavate peatöövõtjate ja alltöövõtjate suure arvu tõttu ning ilma ühtse koordineerimiskeskuseta muutub protsess lõputute kooskõlastuste, ümbertööde ja konfliktide seeriaks. Kõik see toob kaasa märkimisväärsed viivitusi ja mitmekordseid projektkulusid.

Klassikaline olukord ehitusplatsil on lihtne: šabloon on valmis, kuid armatuuri tarne ei ole õigeaegselt kohale jõudnud. Erinevates süsteemides teabe kontrollimisel on suhtlus umbes järgmine:

- 🔊 Ehitusplatsi **töödejuhataja** kirjutab 20. kuupäeval projektijuhile: *"Me oleme lõpetanud šalongi paigaldamise, kus on raudbetoon?"*.
- 🔊 **Projektijuht** (PMIS) hankeosakonnale: - *"Šabloon on valmis. Minu süsteemis [PMIS] on kirjas, et betoonraudad pidid saabuma 18. kuupäeval. Kus on betoonraudad?"*
- 🔊 **Tarneahela spetsialist** (ERP): - *"Meie ERP ütleb, et tarne toimub 25. kuupäeval"*.
- 🔊 **Andmeinsener** või IT-osakond (vastutab integratsioonide eest): - *PMISis on kuupäev 18., ERPis 25. kuupäev. ERP ja PMISi vahel puudub OrderID-link, seega ei ole andmed sünkroonitud. See on tüüpiline näide infolüngast.*
- 🔊 **Projektijuht** peadirektorile - *"Paigaldiste tarnimine on hilineanud, ehitusplats seisab ja pole selge, kes vastutab"*.

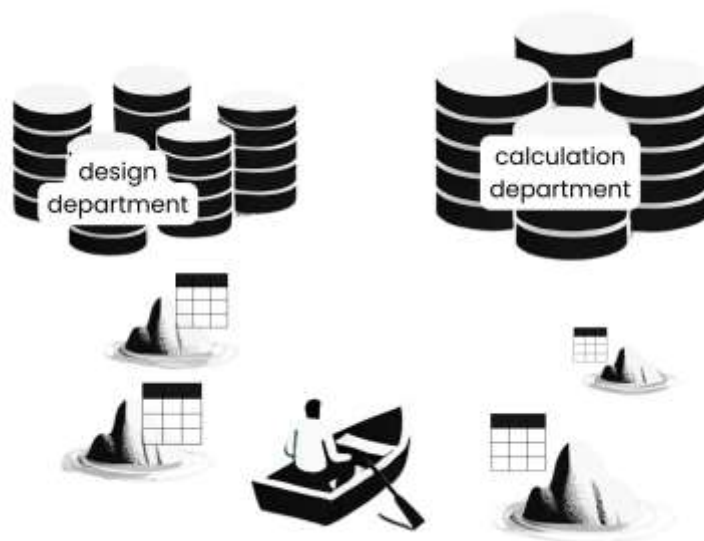
Intsidendi põhjuseks oli andmete isoleerimine eri süsteemides. Andmeallikate integreerimise ja ühendamise, ühtse teabehoidla loomisega ja automatiseerimisega ETL -tööriistade (Apache NiFi, Airflow või n8n) abil saab kõrvaldada süsteemide vahelised silod. Neid ja teisi meetodeid ja vahendeid käsitletakse üksikasjalikult raamatu hilisemates osades.



Joonis 2.1-3 Saksamaa suuremate infrastruktuuriprojektide kavandatud ja tegelike kulude võrdlus.

Sama juhtub ka ettevõtlussüsteemidega: kõigepealt luuakse eraldiseisvad lahendused ja seejärel tuleb nende integreerimiseks ja ühtlustamiseks kulutada tohutuid eelarvevahendeid. Kui andme- ja kommunikatsioonimudelid oleksid algusest peale läbi mõeldud, ei oleks integratsiooni üldse vaja. Siilotud andmed tekitavad digitaalses maailmas kaose, nagu koordineerimata ehitusprotsess.

KPMG 2023. aasta uuringu "Cue construction 4.0: Time to make or break" kohaselt jagab ainult 36% ettevõtetest andmeid tõhusalt osakondade vahel, samas kui 61% seisab silmitsi tõsiste probleemidega, mis tulenevad isoleeritud andmesilodest [30].



Joonis 2.1-4 Aastaid kogutud raskesti kogutavad andmed kogunevad isoleeritud salvestusruumidesse, mida ei kasutata kunagi.

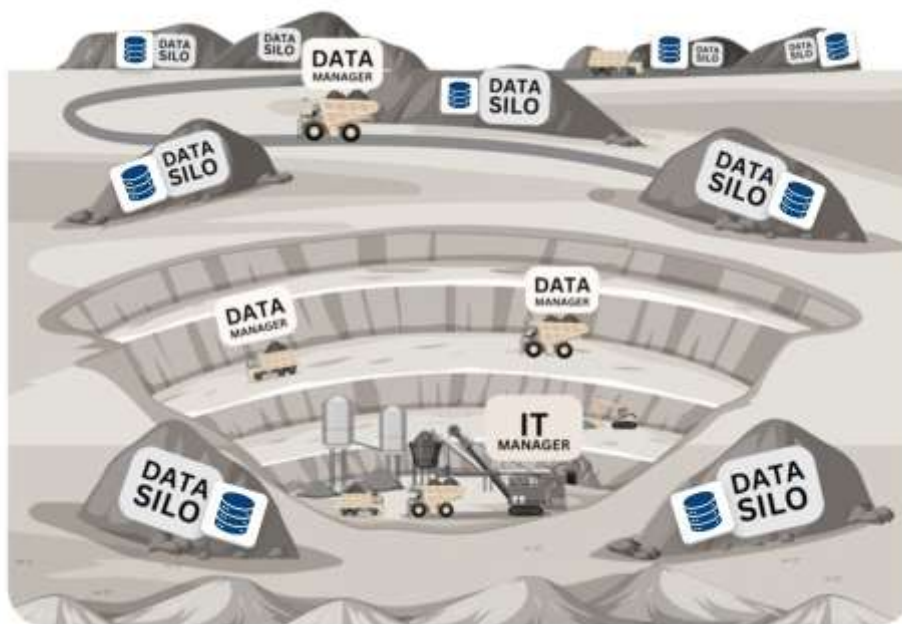
Ettevõtte andmeid hoitakse isoleeritud süsteemides, nagu üksikuid puid, mis on laiali maastikul. Igaüks neist sisaldab väärtuslikku teavet, kuid nendevaheliste ühenduste puudumine takistab ühtse, omavahel ühendatud ökosüsteemi loomist. Selline eraldatus takistab andmete liikumist ja piirab organisatsiooni võimet näha tervikpilti. Nende silode ühendamise on äärmiselt pikk ja keeruline protsess, mille käigus kasvatatakse juhtimistasandil seenemütseeli, et õppida, kuidas üksikuid teabetükke süsteemide vahel üle kanda.

WEFi 2016. aasta uuringu kohaselt on üks peamisi takistusi digitaalsel ümberkujundamisel ühiste andmestandardite puudumine ja killustatus.

Ehitustööstus on üks kõige killustatumatest maailmas ja sõltub kõigi väärtusahelas osalejate sujuvast koostööst [5].

- Maailma Majandusfoorum 2016: tuleviku kujundamine

Disainerid, juhid, koordinaatorid ja arendajad eelistavad sageli töötada iseseisvalt, vältides koordineerimise keerukust. See loomulik kalduvus viib teabe "silode" tekkimiseni, kus andmed on eraldatud eraldi süsteemidesse. Mida rohkem on selliseid isoleeritud süsteeme, seda raskem on neid koos töötama panna. Aja jooksul saab iga süsteem oma andmebaasi ja spetsiaalse tugiosakonna (joonis 1.2-4), mis muudab integreerimise veelgi keerulisemaks.



Joonis 2.1-5 Iga süsteem kipub looma oma unikaalse silo andmeid, mida tuleb töödelda sobivate vahenditega [31].

Ettevõtete süsteemide nõiarõng näeb välja järgmiselt: ettevõtted investeerivad keerukatesse isoleeritud lahendustesse, seepärast seisavad nad silmitsi nende integreerimise kõrgete kuludega ning arendajad, kes mõistavad süsteemide ühendamise keerukust, eelistavad töötada oma suletud

ökosüsteemides. Kõik see suurendab IT-maastiku killustatust ja raskendab üleminekut uutele lahendustele (joonis 2.1-5). Juhid kritiseerivad lõpuks andmesilosid, kuid analüüsivad harva nende põhjuseid ja seda, kuidas neid vältida. Juhid kurdavad vananenud IT-süsteemide üle, kuid nende asendamine nõuab märkimisväärsed investeeringuid ja annab harva oodatud tulemusi. Selle tulemusel muudavad isegi katsed probleemi vastu võidelda sageli olukorda halvemaks.

Peamine põhjus, miks ühendus puudub, on rakenduste eelistamine andmetele. Ettevõtted arendavad esmalt eraldi süsteeme või ostavad tarnijatelt valmislahendusi ning püüavad neid siis ühendada, luues dubleerivaid ja ühildumatuid salvestusrume ja andmebaase.

Killustatuse probleemi ületamine nõuab radikaalset uut lähenemist - andmete eelistamist rakendustele. Ettevõtted peavad kõigepealt välja töötama andmehaldusstrateegiad ja andmemudelid ning seejärel looma süsteemid või ostma lahendused, mis töötavad ühtse teabekogumiga, mitte looma uusi tõkkeid.

Oleme sisenemas uude maailma, kus andmed võivad olla olulisemad kui tarkvara.

- Tim O'Reilly, O'Reilly Media, Inc. tegevjuht.

McKinsey Global Institute'i uuring "Rethinking Construction: the path to improved productivity" (2016) näitab, et ehitussektor on digitaalses ümberkujundamises teistest sektoritest maha jäänud [32]. Aruande kohaselt võib automatiseeritud andmehalduse ja digitaalsete platvormide kasutuselevõtt oluliselt parandada tootlikkust ja vähendada protsesside ebajärjekindlusest tulenevaid kahjusid. Seda vajadust digitaalse ümberkujundamise järele rõhutab ka Egani (Ühendkuningriik, 1998) aruanne [33], milles rõhutatakse integreeritud protsesside ja koostööpõhise lähenemisviisi võtmerolli ehituses.

Selle tulemusena, kui viimase 10 000 aasta jooksul oli andmehaldurite peamine probleem andmete vähesus, siis andmete ja andmehaldussüsteemide laviiniga seisavad kasutajad ja haldajad silmitsi probleemiga - andmete üleküllusega, mis raskendab õiguslikult korrektse ja kvaliteetse teabe leidmist.

Andmete ebaühtlased silosid viivad paratamatult andmete kvaliteedi vähenemise tõsise probleemi. Mitme sõltumatu süsteemi puhul võivad samad andmed eksisteerida eri versioonides, sageli vastuoluliste väärtustega, mis tekitab lisaraskusi kasutajatele, kes peavad kindlaks tegema, milline teave on asjakohane ja usaldusväärne.

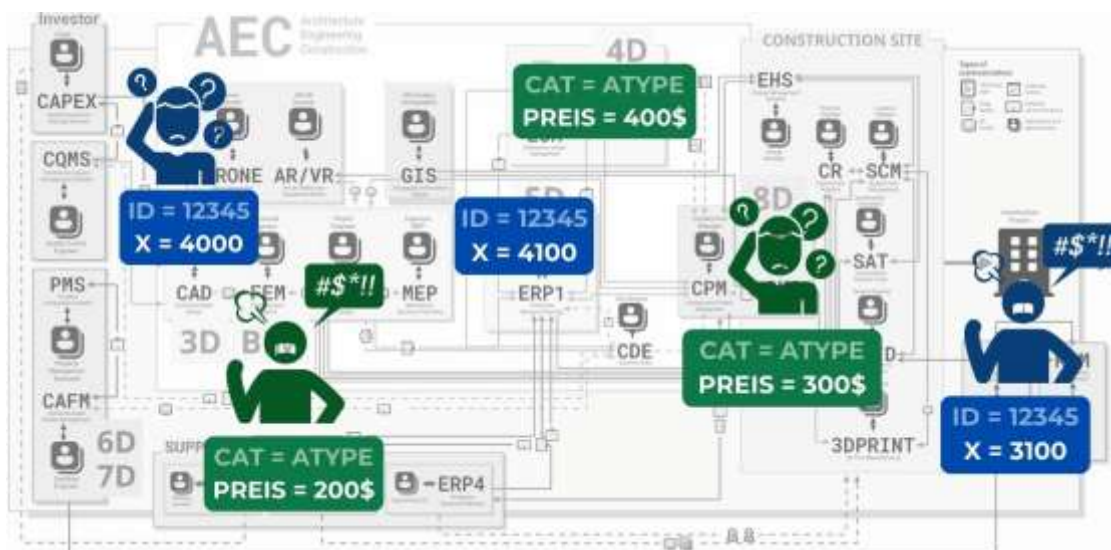
dubleerimine ja andmete kvaliteedi puudumine kui üksmeelsuse puudumine.

Andmesilode probleemi tõttu peavad juhid kulutama palju aega andmete otsimisele ja kooskõlastamisele. Kvaliteediprobleemide vastu kaitsmiseks loovad ettevõtted keerulisi teabehaldusstruktuure, kus andmete otsimise, kontrollimise ja kooskõlastamise eest vastutab juhtide vertikaalne rühm. Selline lähenemisviis suurendab aga ainult bürokraatiat ja aeglustab otsuste

tegemist. Mida rohkem on andmeid, seda raskem on neid analüüsida ja tõlgendada, eriti kui puudub ühtne standard nende säilitamiseks ja töötlemiseks.

Seoses tarkvararakenduste ja -süsteemide rohkusega, mis on viimasel kümnendil kasvanud nagu seened pärast vihma, on silode ja ebasobiva andmekvaliteedi probleem muutunud lõppkasutajate jaoks üha olulisemaks. Samu andmeid, kuid erinevate väärtustega, võib nüüd leida erinevates süsteemides ja rakendustes (joonis 2.1-6). See tekitab lõppkasutajatele raskusi, kui nad püüavad kindlaks teha, milline andmete versioon on paljude olemasolevate andmete hulgast asjakohane ja õige. See põhjustab vigu analüüsides ja lõppkokkuvõttes otsuste tegemisel.

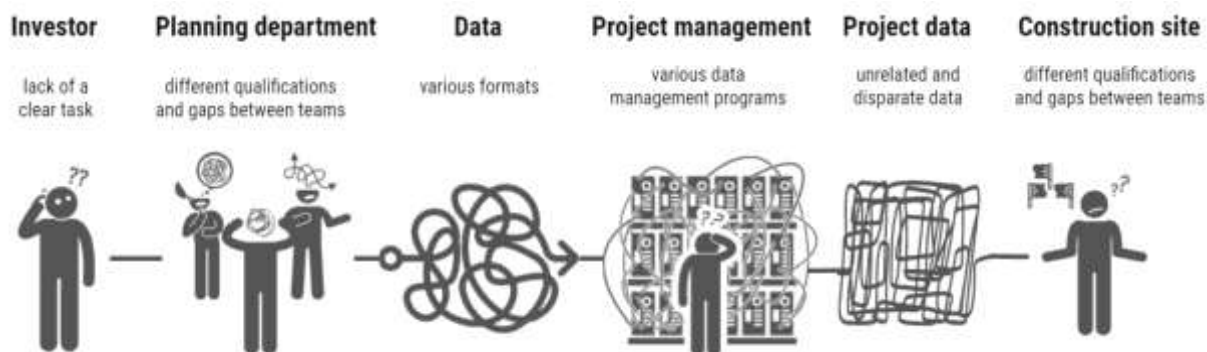
Et kindlustada õigete andmete leidmisega seotud probleemide vastu, loovad ettevõtte juhid mitmetasandilise bürokraatia, mis koosneb kontrollijuhtidest. Nende ülesanne on olla võimeline kiiresti leidma, kontrollima ja saatma nõutud andmeid tabelite ja aruannete kujul, orienteerudes erinevate süsteemide labürintis.



Joonis 2.1-6 Püüdes leida õigeid andmeid, peavad juhid tagama erinevate süsteemide vahel andmete kvaliteedi ja õiguskindluse.

Praktikas tekitab see mudel aga uusi keerukusi. Kui andmeid hallatakse käsitsi ja teave on hajutatud paljude omavahel mitteseotud otsuste vahel, muutub iga katse saada täpset ja ajakohast teavet otsustajate püramiidi kaudu (joonis 2.1-7) kitsaskohaks - see on aeganõudev ja vigadega seotud.

Olukorda raskendab digitaalsete lahenduste laaviin. Tarkvaraturg on jätkuvalt üle ujutatud uutest vahenditest, mis tunduvad paljutöotavad. Kuid ilma selge andmehaldusstrateegiata ei integreeru need lahendused ühtseks süsteemiks, vaid loovad hoopis täiendavaid keerukuse ja dubleerimise kihte. Selle tulemusena leiavad ettevõtted end protsesside lihtsustamise asemel veelgi killustatumast ja kaootilisemast teabekeskonnast.



Joonis 2.1-7 Süsteemide keerukus ja andmevormingute mitmekesisus viivad järjepidevuse kadumiseni ehitusprotsessis.

Kõik need probleemid, mis on seotud paljude erinevate lahenduste haldamisega, viivad ettevõtte juhtkonna varem või hiljem olulise äratundmiseni: küsimus ei ole andmemahus ega järgmise "üheainsa" andmetöötlusvahendi otsimises. Tegelik põhjus peitub andmete kvaliteedis ja selles, kuidas organisatsioon neid loob, vastu võtab, säilitab ja kasutab.

Jätkusuutliku edu võti ei seisne mitte uute "võlurakenduste" tagaajamises, vaid andmekultuuri loomises ettevõttesiseselt. See tähendab, et andmeid tuleb käsitleda strateegilise väärtusena ning andmete kvaliteet, terviklikkus ja asjakohasus tuleb seada prioriteediks organisatsiooni kõigil tasanditel.

Lahendus kvaliteedi ja kvantiteedi dilemmale peitub ühtse andmestruktuuri loomises, mis kõrvaldab dubleerimise, kõrvaldab vastuolud ja ühtlustab infovooge. Selline ülesehitus pakub ühtset, usaldusväärset andmeallikat, mille põhjal saab teha teadlikke, täpseid ja õigeaegseid otsuseid.

Vastasel juhul, nagu ikka veel sageli juhtub, tuginevad ettevõtted jätkuvalt pigem HiPPO ekspertide subjektiivsetele arvamustele ja intuiitivsetele hinnangutele kui usaldusväärsetele faktidele. Eriti märgatav on see ehitussektoris, kus ekspertiisil on traditsiooniliselt oluline roll.

HiPPO ehk arvamuste oht otsuste tegemisel

Traditsiooniliselt tehakse ehitussektoris olulisi otsuseid kogemuste ja subjektiivse hinnangu alusel. Ilma õigeaegsete ja usaldusväärsete andmeteta peavad ettevõtte juhid tegutsema pimesi, tuginedes pigem kõrgeima palgaga töötajate intuitsioonile (HiPPO - Highest Paid Person's Opinion) kui objektiivsetele faktidele (joonis 2.1-8).

NO ANALYTICS?
WELCOME TO THE HIPPO*



*HIGHEST PAID PERSON'S OPINION

Joonis 2.1-8 Analüütika puudumisel sõltub äri kogunud spetsialistide subjektiivsest arvamusest.

Selline lähenemisviis võib olla õigustatud stabiilses ja aeglaselt muutuv keskkonnas, kuid digitaalse ümberkujundamise ajastul muutub see tõsiseks riskiks. Intuitsioonil ja oletustel põhinevad otsused on alid moonutustele, põhinevad sageli põhjendamatutel hüpoteesidel ega võta arvesse andmetes kajastuvat keerulist pilti

See, mida ettevõtte otsustustasandil peetakse intelligentseks aruteluks, ei põhine sageli millelgi konkreetsele. Ettevõtte edu ei peaks sõltuma mitte ekspertide autoriteedist ja palgast, vaid oskusest töötada tõhusalt andmetega, tuvastada mustreid ja teha teadlikke otsuseid.

Oluline on loobuda arusaamast, et autoriteet või kogemus tähendab automaatselt, et otsus on õige. Andmepõhine lähenemine muudab mängu: otsuste tegemise aluseks on nüüd andmed ja analüüsid, mitte ametikoht ja palk. Suurandmed, masinõpe ja visuaalne analüüs võimaldavad meil tuvastada mustreid ja tugineda faktidele, mitte oletustele (joonis 1.1-4).

Ilma andmeteta olete lihtsalt üks inimene, kellel on oma arvamus [34].

- W. Edwards Deming, teadlane ja juhtimiskonsultant

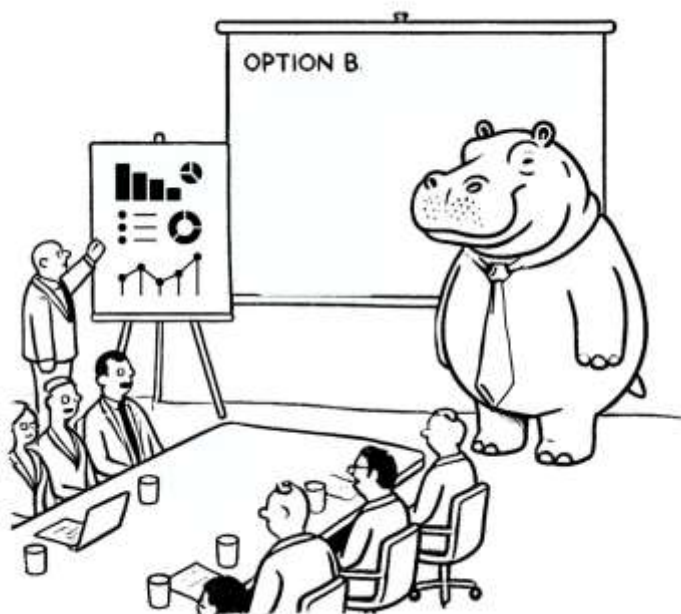
Kaasaegsed andmehaldusmeetodid tagavad ka teadmiste järjepidevuse ettevõttes. Selgelt kirjeldatud protsessid, automatiseerimine ja süstemaatiline lähenemine võimaldavad isegi võtmerollide üleandmist ilma tõhusust kaotamata.

Kuid pime usaldus andmete suhtes võib põhjustada ka tõsisid vigu. Andmed ise on lihtsalt numbrite kogum. Ilma nõuetekohase analüüsi, konteksti ja võimeta tuvastada mustreid, ei ole neil mingit väärtust ja nad ei saa juhtida protsesse. Edu võti ei seisne mitte selles, et valida HiPPO intuitsiooni ja analüüsi vahel, vaid selles, et luua intelligentsed vahendid, mis muudavad ebaühtlase teabe hallatavateks, teadlikeks otsusteks.

Digitaalses ehituskeskkonnas ei ole otsustavaks eduteguriks mitte ametiaeg ja koht hierarhias, vaid reageerimisvõime, otsuste täpsus ja ressursitõhusus

Andmed on vahendid, mitte absoluutsed tõed. Need peaksid täiendama, mitte asendama inimese mõtlemist. Vaatamata analüütika eelistele, ei saa andmed täielikult asendada inimese intuitsiooni ja kogemust. Nende roll on aidata teha täpsemaid ja teadlikumaid otsuseid.

Konkurentsieeliseid ei saavutata mitte ainult standardite täitmisega, vaid ka sellega, et suudetakse konkurente edestada ressursside tõhusas kasutamises, mis on kõigile ühesugused. Tulevikus muutuvad andmeoskused sama oluliseks kui kunagi kirjaoskused või matemaatikaoskused. Spetsialistid, kes oskavad andmeid analüüsida ja tõlgendada, suudavad teha täpsemaid otsuseid, tõrjudes välja need, kes tuginevad ainult isiklikule kogemusele (joonis 2.1-9).



Joonis 2.1-9 Otsused peaksid põhinema objektiivsel analüüsil, mitte kõrgeima palgaga töötaja arvamusel.

Juhid, spetsialistid ja insenerid tegutsevad andmeanalüütikutena, uurides projektide struktuuri, dünaamikat ja põhinäitajaid. Inimressurssidest saavad süsteemi elemendid, mis vajavad paindlikku andmepõhist kohandamist, et maksimeerida tõhusust.

Ebapiisavate andmete kasutamisel on vead palju väiksemad kui andmete puudumisel [35].

- Charles Babbage, esimese analüütilise arvutusmasina leiutaja.

Suurandmete tekkimine ja LLM (Large Language Models) kasutuselevõtt on radikaalselt muutnud mitte ainult meie analüüsimeetodeid, vaid ka otsuste tegemise olemust. Kui varem keskenduti põhjuslikkusele (miks midagi juhtus - diagnostiline analüüs) (joonis 1.1-4), siis tänapäeval on esiplaanile tõusmas võime ennustada tulevikku (prognoosiv analüüs) ja tulevikus preskriptiivne analüüs, kus masinõpe ja tehisintellekt pakuvad otsustusprotsessis parimat valikut.

Vastavalt uuele SAP™ uuringule "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36] oleks 44% tippjuhtidest valmis muutma oma varasemat otsust AI nõuannete põhjal ja 38% usaldaks AI-d nende nimel äriotsuste tegemisel. Samal ajal ütles 74% juhtidest, et nad usaldavad tehisintellekti nõuandeid rohkem kui oma sõpru ja perekonda, ning 55% töötab ettevõtetes, kus tehisintellekti abil saadud teadmised asendavad traditsioonilisi otsustusmeetodeid või lähevad neist sageli mööda - eriti organisatsioonides, mille aastane tulu on üle 5 miljardi dollari. Lisaks kasutab 48% vastanutest igapäevaselt tehisintellekti genereerivaid vahendeid, sealhulgas 15% neist kasutab neid mitu korda päevas.

Elukorralduslike ja automatiseeritud andmehaldussüsteemide arenguga tekib uus väljakutse: kuidas kasutada teavet tõhusalt, kaotamata selle väärtust ühildumatute vormingute ja heterogeensete allikate kaoses, mida täiendab äriprotsesside kasvav keerukus ja dünaamilisus.

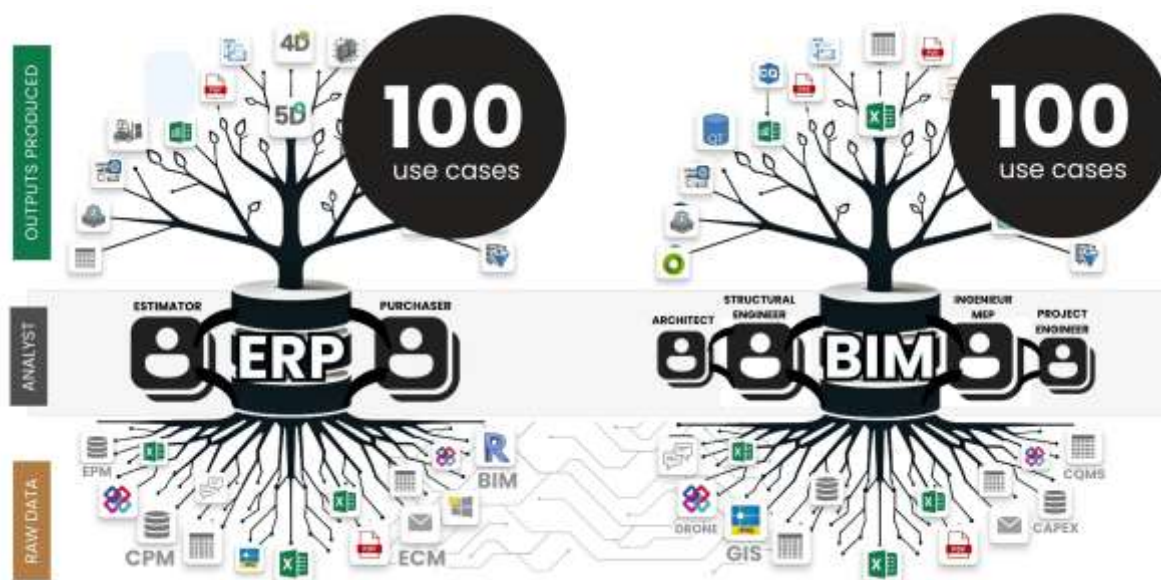
Äriprotsesside keerukuse ja dünaamilisuse pidev suurenemine

Ehitustööstus seisab tänapäeval silmitsi tõsiste väljakutsetega andmete ja protsesside haldamisel. Peamised probleemid on siilutud infosüsteemid, liigne bürokraatia ja digitaalsete vahendite vahelise integratsiooni puudumine. Need väljakutsed süvenevad, kuna äriprotsessid ise muutuvad keerulisemaks - seda tingivad tehnoloogia, muutuvad kliendinõuded ja arenevad eeskirjad.

Ehitusprojektide eripära ei tulene mitte ainult nende tehnilistest iseärasustest, vaid ka riiklike standardite ja regulatiivsete nõuete erinevustest eri riikides (joonis 4.2-10, joonis 5.1-7). See nõuab paindlikku, individuaalset lähenemist igale projektile, mida on raske rakendada traditsiooniliste moodulitest koosnevate kontrollisüsteemide raames. Protsesside keerukuse ja suure andmehulga tõttu pöörduvad paljud ettevõtted spetsialiseeritud lahendusi pakuvate müüjate poole. Kuid turg on ülekoormatud - paljud idufirmad pakuvad sarnaseid tooteid, mis keskenduvad kitsastele ülesannetele. Selle tulemusel jääb andmehalduse terviklik lähenemine sageli kaduma.

Uute tehnoloogiate ja turunõuetega kohanemine on muutumas konkurentsivõime kriitiliseks teguriks. Olemasolevad patenteeritud rakendused ja moodulsüsteemid on aga vähese kohanemisvõimega - mis tahes muudatused nõuavad sageli pikki ja kulukaid muudatusi arendajate poolt, kes ei mõista alati ehitusprotsesside eripärasid.

Ettevõtted leiavad end tehnoloogilise mahajäämuse pantvangist, oodates uusi uuendusi, selle asemel et kiiresti rakendada uuenduslikke integreeritud lähenemisviise. Selle tulemusel on ehitusorganisatsioonide sisemine struktuur sageli omavahel seotud hierarhiliste ja sageli suletud süsteemide keerukas ökosüsteem, mida koordineerib mitmetasandiline juhtide võrgustik (joonis 2.1-10).

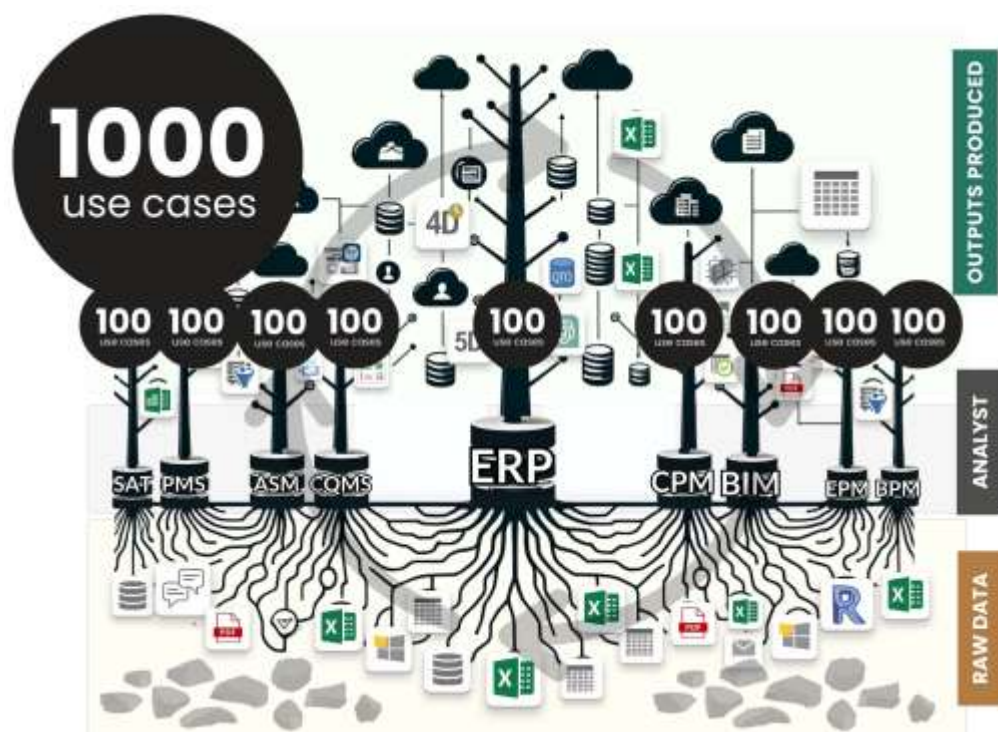


Joonis 2.1-10 Ettevõtte koosnevad omavahel ühendatud süsteemidest, mille omavaheline ühendamine moodustab protsessid, mis vajavad automatiseerimist.

Kanada ehitusliidu ja KPMG Canada 2021 uuringu [37] kohaselt usub vaid 25% ettevõtetest, et nad on tehnoloogia kasutuselevõtu või digitaalsete lahenduste osas konkurentidega võrreldes olulises või erinevas positsioonis. Ainult 23% vastanutest teatas, et nende lahendused on märkimisväärselt või suurel määral andmepõhised. Samal ajal iseloomustas enamik küsitlusele vastanutest, et nad kasutavad mitmesuguseid muid tehnoloogiaid üksnes eksperimentaalselt või tunnistasid, et ei kasuta neid üldse.

See vastumeelsus tehnoloogilistes eksperimentides osalemise suhtes on eriti ilmne suurte infrastruktuuriprojektide puhul, kus vead võivad maksta miljoneid dollareid. Isegi kõige arenenumad tehnoloogiad - digitaalsed kaksikud, prognoositav analüüs - ei kohta sageli vastuseisu mitte nende tõhususe, vaid tegelikes projektides tõestatud usaldusväärsuse puudumise tõttu.

Maailma Majandusfoorumi (WEF) aruande "Shaping the Future of Construction" [5] kohaselt seisab uute tehnoloogiate kasutuselevõtt ehituses silmitsi mitte ainult tehniliste raskustega, vaid ka psühholoogiliste takistustega klientide poolt. [5] kohaselt ei ole uute tehnoloogiate kasutuselevõtul ehituses mitte ainult tehnilisi raskusi, vaid ka psühholoogilisi takistusi klientide poolt. Paljud kliendid kardavad, et täiustatud lahenduste kasutamine muudab nende projektid katsealaks ja teeb neist "katsekoerad" ning ettearvamatud tagajärjed võivad põhjustada lisakulusid ja -riske.



Joonis 2.1-11 Iga andmekasutuse jaoks pakub lahenduste turg rakendusi protsesside optimeerimiseks ja automatiseerimiseks.

Ehitustööstus on väga mitmekesine: erinevatel projektidel on erinevad nõuded, piirkondlikud eripärad, seadusjärgsed klassifitseerimiseeskirjad (joonis 4.2-10), arvutamisstandardid (joonis 5.1-7) jne. Seetõttu on praktiliselt võimatu luua ühte universaalset rakendust või süsteemi, mis sobiks ideaalselt kõigi nende nõuetega ja projekti eripäradega.

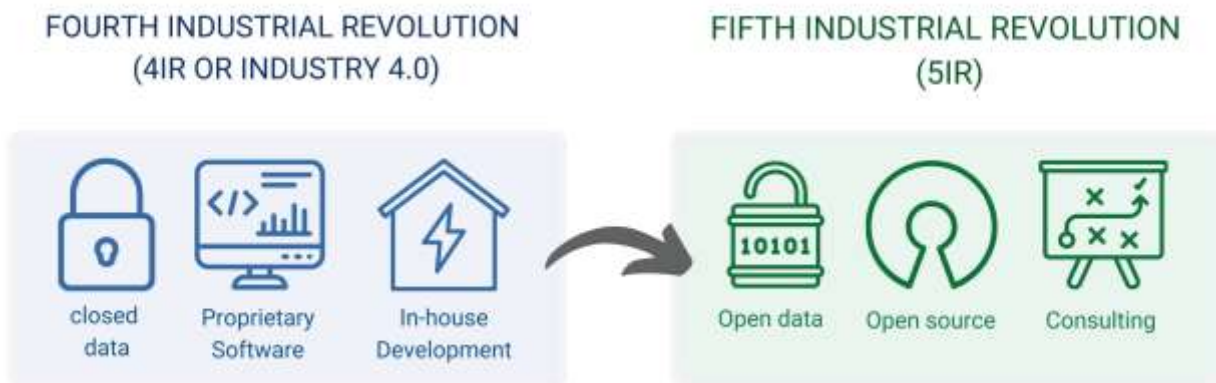
Püüdes tulla toime süsteemide kasvava keerukusega ja sõltuvusega tarkvaratootjatest, mõistetakse üha enam, et tõhusa andmehalduse võti ei ole mitte ainult avatus ja standardimine, vaid ka protsessiarhitektuuri enda lihtsustamine. Äriprotsesside kasvav keerukus ja dünaamilisus nõuab uusi lähenemisviise, kus prioriteet nihkub andmete kogumise asemel nende struktureerimisele ja organiseerimisele. Just see nihe on järgmine samm ehitussektori arengus, mis tähistab tarkvaratootjate domineerimise ajastu lõppu ja sisulise teabe korraldamise ajastu algust.

Ühekordsete lahenduste piiratuse ja haavatavuse mõistmine kasvava keerukuse suhtes toob kaasa prioriteetide nihkumise suletud platvormidelt ja andmete varumisest läbipaistvuse, kohandatavuse ja struktureeritud teabekäitluse suunas. See mõtteviisi muutus peegeldab laiemaid muutusi maailmamajanduses ja tehnoloogias, mida kirjeldatakse nn "tööstusrevolutsioonide" kaudu. Et mõista, kuhu ehitus liigub ja kuhu see tulevikus suundub, on vaja kaaluda tööstuse kohta neljanda ja viienda tööstusrevolutsiooni kontekstis - alates automatiseerimisest ja digitaliseerimisest kuni personaliseerimise, avatud standardite ja teenuspõhise andmemudelini.

Neljas tööstusrevolutsioon (tööstus 4.0) ja viies tööstusrevolutsioon (tööstus 5.0) ehituses.

Tehnoloogilised ja majanduslikud etapid on teoreetilised mõisted, mida kasutatakse ühiskonna ja majanduse arengu kirjeldamiseks ja analüüsimiseks erinevates arenguetappides. Erinevad teadlased ja eksperdid võivad neid erinevalt tõlgendada.

- **Neljas tööstusrevolutsioon** (4IR või Industry 4.0) on seotud infotehnoloogia, automatiseerimise, digitaliseerimise ja globaliseerumisega. Üks selle põhielemente on patenteeritud tarkvaralahenduste, st konkreetsete ülesannete ja ettevõtete jaoks loodud spetsiaalsete digitaalsete toodete loomine. Need lahendused muutuvad sageli IT-infrastruktuuri oluliseks osaks, kuid on ilma täiendavate muudatusteta halvasti skaleeritavad.
- **Viies tööstusrevolutsioon (5IR)** on praegu kontseptsiooni ja arengu varasemas etapis kui 4IR. Selle põhiprintsiibid hõlmavad toodete ja teenuste suuremat isikupärastamist. 5IR on liikumine kohanemisvõimelisema, paindlikuma ja personaalsema majandustegevuse suunas, keskendudes isikupärastamisele, nõustamisele ja teenustele orienteeritud mudelitele. Viienda majandusviisi põhiaspekt on andmete kasutamine otsuste tegemisel, mis on praktiliselt võimatu ilma avatud andmete ja avatud vahendite kasutamiseta (joonis 2.1-12).



Joonis 2.1-12 Neljas muster keskendub lahendustele, viies muster aga isikupärastamisele ja andmetele.

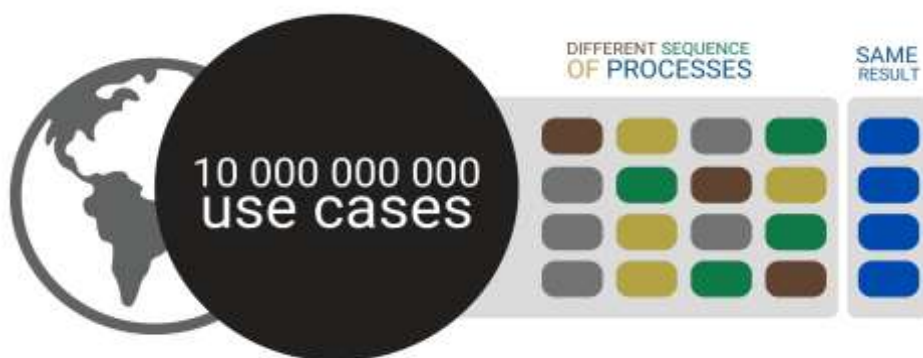
Rakenduse loomine ehitussektori ettevõtetele kasutamiseks kümnes või sajas organisatsioonis ei garanteeri selle edukat laiendamist teistele ettevõtetele, piirkondadele või riikidele ilma oluliste muudatuste ja täiendusteta. Selliste lahenduste eduka skaleerimise tõenäosus jääb väikeseks, sest igal organisatsioonil on ainulaadsed protsessid, nõuded ja tingimused, mis võivad nõuda individuaalseid kohandusi.

Oluline on mõista, et juba praegu eeldab tehnoloogiliste lahenduste edukas integreerimine sügavalt personaliseeritud lähenemist igale protsessile, projektile ja ettevõttele. See tähendab, et isegi kui universaalne raamistik, vahend või programm on välja töötatud, tuleb seda üksikasjalikult kohandada ja kohandada iga konkreetse ettevõtte ja projekti unikaalsetele nõuetele ja tingimustele.

Vastavalt PwC aruandele "Viienda tööstusrevolutsiooni dekodeerimine" [38] [38] tugineb umbes 50% erinevate tööstusharude tippjuhtidest sel aastal arenenud tehnoloogia ja inimteadmiste integreerimisele. Selline lähenemine võimaldab neil kiiresti kohaneda muutustega tootedisainis või klientide nõudmistes, luues personaalse tootmise.

Iga protsess nõuab unikaalse funktsiooni või rakenduse arendamist, mis, arvestades ülemaailmse ehitussektori suurust ja projektide mitmekesisust, toob kaasa tohutu hulga ärijuhtumite olemasolu, millest igaüks esindab unikaalset Pipeline loogikat (joonis 2.1-13). Igal sellisel juhtumil on oma eripärad ja see nõuab kohandatud lähenemisviisi. Üksikasjalikumalt vaatleme sama analüütilise probleemi võimalike lahenduste mitmekesisust erinevate lähenemisviiside kontekstis peatükis, mis on pühendatud masinõppele ja Titanicu andmekogumi analüüsimisele (joonis 9.2-9).

Torustik on digitaalsete protsesside kontekstis tegevuste, protsesside ja vahendite jada, mis võimaldab automatiseeritud või struktureeritud andme- ja töövoolu läbi projekti elutsükli eri etappide.



Joonis 2.1-13 Ärijuhtumite individuaalsus ja varieeruvus muudab võimaluks katsed luua skaleeritavaid suletud platvorme ja vahendeid.

Meie elu on juba mitmel moel muutunud digitaalse transformatsiooni mõjul ja täna võime rääkida uue etapi algusest ehitussektori majanduslikus arengus. Selles "uues majanduses" korraldatakse konkurents teistsuguste reeglite järgi: kes suudab tõhusalt muuta avalikud teadmised ja avatud andmed nõutavateks toodeteks ja teenusteks, saab viienda tööstusrevolutsiooni tingimustes olulise eelise.

Nagu märgib majandusteadlane Kate Maskus raamatus "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39]. 2012 [39], "elame globaalses teadmispõhises majanduses ja tulevik kuulub neile, kes oskavad teaduslikke avastusi kaubaks muuta".

Üleminek viiendale majandusrežiimile tähendab keskendumist suletud IT-lahendustelt avatud standarditele ja platvormidele. Ettevõtted hakkavad loobuma traditsioonilistest tarkvaratoodetest ja eelistavad teenustele orienteeritud mudeleid, kus peamiseks varaks on andmed, mitte patenteeritud tehnoloogiad.

Harvard Business School 2024 uuring [40] näitab avatud lähtekoodiga tarkvara (Open Source Software, OSS) tohutut majanduslikku väärtust. Uuringu kohaselt on OSS 96% kõigist tarkvarakoodidest ja mõned kommertstarkvarad koosnevad 99,9% ulatuses OSS-komponentidest. Ilma OSSita kulutaksid ettevõtted 3,5 korda rohkem tarkvara peale.

Ettevõtete ökosüsteemide ehitamine, järgides ülemaailmseid suundumusi, läheb järk-järgult üle viiendale majanduslikule paradigmale, kus andmekeskne analüüs ja nõustamisteenused muutuvad tähtsamaks kui isoleeritud suletud lahendused, mille kasutusjuhud on jäigalt määratletud.

Digitaliseerimise ajastu muudab tööstuse jõudude vahekorda: selle asemel, et toetuda müüja lahendustele, hakkavad ettevõtted oma konkurentsivõimet rajama oma võimele kasutada andmeid tõhusalt. Selle tulemusel liigub ehitussektor vanadelt jäikadelt süsteemidelt paindlikele, kohanduvatele ökosüsteemidele, kus avatud standardid ja koostalitlusvõimelised vahendid on projekti juhtimise aluseks. Rakendustootjate domineerimise ajastu lõppemine loob uue keskkonna, kus väärtust ei määratleta mitte suletud lähtekoodi ja spetsialiseeritud ühenduspunktide omamisega, vaid võimega muuta andmed strateegiliseks eeliseks.



PEATÜKK 2.2.

KAOSE MUUTMINE KORRASTAMISEKS JA KEERUKUSE VÄHENDAMINE

Üleliigne kood ja suletud süsteemid, mis takistavad tootlikkuse parandamist

Viimastel aastakümnetel on tehnoloogilisi muutusi IT-sfääris juhtinud peamiselt tarkvaratootjad. Nad määravad arengu suuna, määrates kindlaks, millised tehnoloogiad ettevõtte peaksid kasutusele võtma ja millised tuleks maha jätta. Ajastul, mil siilustatud lahendustelt mindi üle tsentraliseeritud andmebaasidele ja integreeritud süsteemidele, edendasid müüjad litsentseeritud tooteid, pakkudes kontrolli juurdepääsu ja skaleeritavuse üle. Hiljem, koos pilvetechnoloogiate ja SaaS-mudelite (Software as a Service) tulekuga, muutus see kontroll tellimusudeliks, mis kinnistas kasutajad digitaalsete teenuste lojaalsete klientidena.

Selline lähenemine on tekitanud paradoksi: vaatamata enneolematult suurele hulgale loodud programmikoodile kasutatakse tegelikult vaid väikest osa sellest. Võib-olla on koodi sadu või tuhandeid kordi rohkem kui vaja, sest samu äriprotsesse kirjeldatakse ja dubleeritakse kümnetes või sadades programides erinevalt, isegi ühe ja sama ettevõtte sees. Samal ajal on arenduskulud juba tasutud ja need kulud on tagastamatud. Sellest hoolimata jätkab tööstus selle tsükli kordamist, luues uusi tooteid, millel on lõppkasutajale minimaalne lisaväärtus, sagedamini turu ootuste kui tegelike vajaduste survel.

Kaitsevæe hankeülikooli (DAU) tarkvaraarenduse kulude hindamise juhendi [41] kohaselt võivad tarkvaraarenduse kulud erineda märkimisväärselt sõltuvalt mitmest tegurist, sealhulgas süsteemi keerukusest ja valitud tehnoloogiast. Ajalooliselt on 2008. aasta arenduskulud olnud umbes 100 dollarit lähtekoodirea kohta (SLOC), samas kui hoolduskulud võivad ulatuda kuni 4000 dollarini SLOCi kohta.

Ainult üks CAD-rakenduste komponentidest - geomeetiline tuum - võib sisaldada kümneid miljoneid koodiridu (joonis 6.1-5). Sarnane olukord on ka ERP-süsteemides (joonis 5.4-4), mille keerukuse käsitlemise juurde pöördume tagasi raamatu viiendas osas. Lähemal vaatlusel selgub siiski, et suur osa sellest koodist ei anna lisaväärtust, vaid toimib üksnes "postiljonina" - mehaaniliselt liigutab andmeid andmebaasi, API, kasutajaliidese ja süsteemi teiste tabelite vahel. Hoolimata populaarsest müüdist nn äriloogika kriitilisest tähtsusest, on karm reaalsus palju proosalisem: kaasaegsed koodibaasid on täis aegunud malliblokke (pärandkoodi), mille ainus eesmärk on tagada andmete edastamine tabelite ja komponentide vahel, mõjutamata seejuures otsuste tegemist või äritegevuse tõhusust.

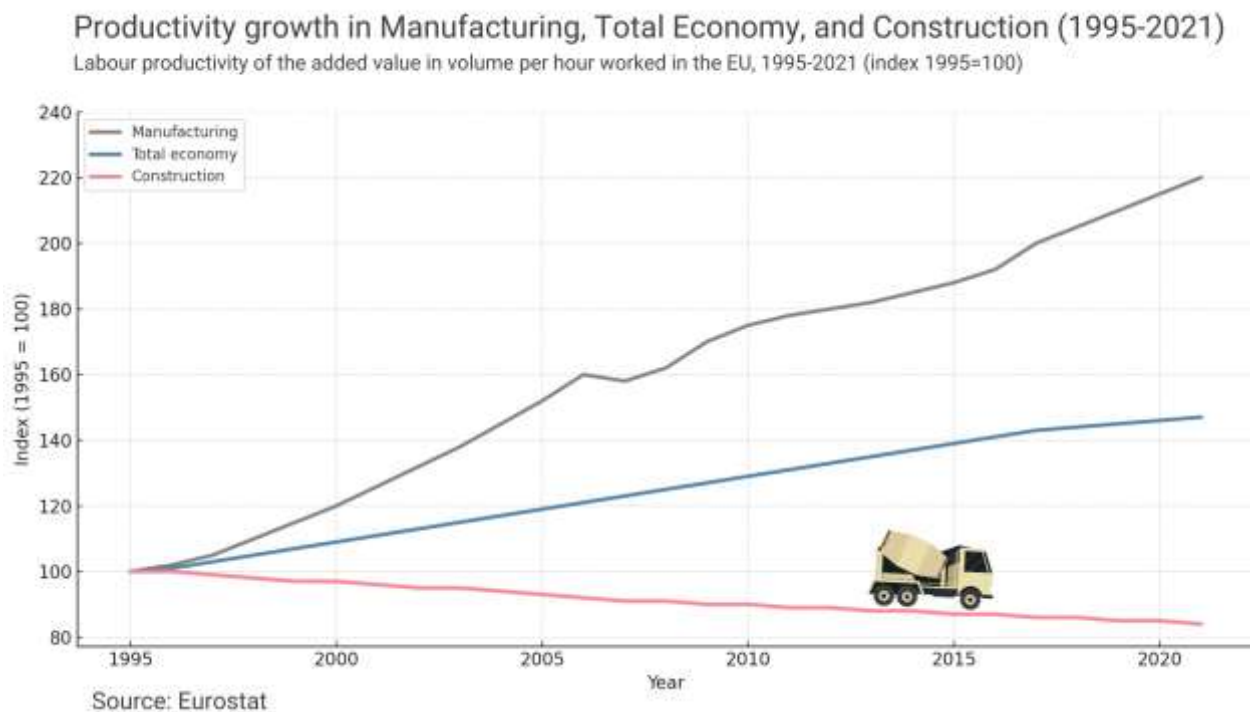
Selle tulemusena muutuvad suletud lahendused, mis töötlevad eri allikatest pärinevaid andmeid, paratamatult segadusttekitavaks "spageti-ökosüsteemiks". Neid keerulisi, omavahel põimunud süsteeme saab hallata ainult poolrutiinselt töötavate juhtide armee. Selline andmehalduse korraldus ei ole mitte ainult ressursside seisukohast ebatõhus, vaid tekitab ka kriitilisi nõrkusi äriprotsessides, muutes ettevõtte sõltuvaks kitsast spetsialistide ringist, kes mõistavad, kuidas see tehnoloogiline labürint toimib.

Koodimahu ja rakenduste arvu pidev kasv ning müüjate pakutavate kontseptsioonide suurenev kee-

rukus on toonud kaasa loomuliku tulemuse - ehituses kasutatava IT-ökosüsteemi keerukuse suurenemise. See on muutnud digitaliseerimise praktilise rakendamise tööstusharu rakenduste arvu suurendamise kaudu ebaefektiivseks. Kasutajate vajadusi arvestamata loodud tarkvaratooted nõuavad sageli märkimisväärsed ressursse rakendamiseks ja toetuseks, kuid ei too oodatud tulu.

Vastavalt McKinsey uuringule "Ehituse tootlikkuse suurendamine" [42] on viimase kahe aastakümne jooksul tööjõu tootlikkuse kasv ehituses olnud maailmas keskmiselt vaid 1% aastas, võrreldes 2,8% kasvuga maailma majanduses tervikuna ja 3,6% kasvuga töötlevas tööstuses. Ameerika Ühendriikides on tööviljakus ehituses töötaja kohta alates 1960ndatest aastatest vähenenud poole võrra [43].

Süsteemide kasvav keerukus, isoleeritus ja suletud andmed on takistanud spetsialistide vahelist suhtlust, mistõttu ehitussektor on üks kõige vähem tõhusamaid (joonis 2.2-1). 2040. aastaks 22 triljoni dollarini, mis nõuab märkimisväärselt tõhususe suurendamist.



Joonis 2.2-1 Suletud ja keerulised andmed ning sellest tulenevalt kehv kommunikatsioon spetsialistide vahel viisid ehitussektori üheks kõige vähem tõhusaks majandusharuks (põhineb [44], [45]).

Nagu rõhutatakse McKinsey (2024) uuringus "Ehituse tootlikkuse tagamine ei ole enam vabatahtlik", ei saa ehitusettevõtete kasvava ressurside nappuse ja tööstuse püüdluse tõttu kahekordistada oma kasvutempot enam lubada endale praegusele tootlikkuse tasemele jäämist [44]. Prognooside kohaselt tõusevad ülemaailmsed ehituskulud 13 triljonilt USA dollarilt 2023. aastal palju kõrgemale tasemele kümnendi lõpuks, mistõttu on tõhususe küsimus mitte lihtsalt asjakohane, vaid kriitilise tähtsusega.

Üks peamisi võimalusi tõhususe parandamiseks on rakendusstruktuuride ja andmete ökosüsteemi

arhitektuuride vältimatu ühtlustamine ja lihtsustamine. Selline ratsionaliseerimise lähenemisviis kaotab üleliigsed abstraktsioonikihtide ja tarbetu keerukuse, mis on aastate jooksul ettevõtte süsteemides kogunenud.

Siiludest ühtse andmelattu

Mida rohkem andmeid organisatsioon kogub, seda raskem on neist tegelikku väärtust välja võtta. Kuna teave on killustatud ja salvestatud isoleeritud silodes, on tänapäeva ettevõtete äriprotsessid nagu ehitajad, kes üritavad ehitada pilvelõhkujat tuhandetes eri ladudes ladustatud materjalidest. Informatsiooni üleküllus mitte ainult ei raskenda juurdepääsu õiguslikult asjakohasele teabele, vaid aeglustab ka otsuste tegemist: iga sammu tuleb korduvalt kontrollida ja kinnitada.

Iga ülesanne või protsess on kindlalt seotud eraldi tabeli või andmebaasiga ning andmevahetus süsteemide vahel nõuab keerulisi integratsioone. Vead ja ebakõlad ühes süsteemis võivad põhjustada ahelarikkeid teistes süsteemides. Väärad väärtused, hilinenud uuendused ja dubleeriv teave sunnivad töötajaid kulutama märkimisväärselt aega andmete käsitsi kooskõlastamisele ja kooskõlastamisele. Selle tulemusena kulub organisatsioonil rohkem aega killustatuse tagajärgedega tegelemisele kui protsesside arendamisele ja optimeerimisele

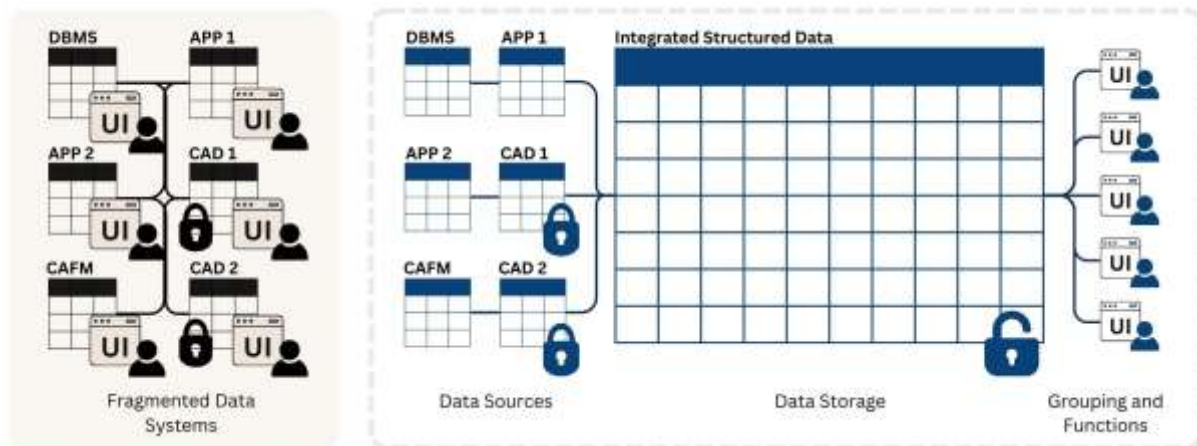
See probleem on universaalne: mõned ettevõtted võitlevad jätkuvalt kaosega, samas kui teised leiavad lahenduse integratsioonis - infovoogude viimisel tsentraliseeritud salvestussüsteemi. Mõelge sellest kui ühest suurest tabelist, kuhu saab salvestada kõik ülesannete, projektide ja objektidega seotud üksused. Kümnete erinevate tabelite ja vormingute asemel tekib üks ühtne ühtne hoidla (joonis 2.2-2), mis võimaldab:

- minimeerida andmekaaotust;
- kaotada vajadus teabe pideva ühtlustamise järele;
- parandada andmete kättesaadavust ja kvaliteeti;
- lihtsustada analüütilist töötlemist ja masinõpet

Andmete viimine ühtsesse standardisse tähendab, et sõltumata allikast teisendatakse teave ühtsesse ja masinloetavasse vormingusse. Selline andmete korraldus võimaldab kontrollida nende terviklikkust, analüüsida neid reaajas ja kasutada neid kiiresti juhtimisotsuste tegemiseks.

Integreeritud salvestussüsteemide kontseptsiooni ja nende rakendamist analüütikas ja masinõppes käsitletakse üksikasjalikumalt peatükis "Suurandmete salvestamine ja masinõpe". Andmete modelleerimise ja struktureerimise teemasid käsitletakse üksikasjalikult peatükkides "Andmete muutmine struktureeritud kujul" ja "Kuidas standardid muudavad mängu: juhuslikest failidest

läbimõeldud andmemudelini".



Joonis 2.2-2 Andmete integreerimine kõrvaldab silod, parandab teabe kättesaadavust ja optimeerib äriprotsesse.

Kui andmed on struktureeritud ja ühendatud, on järgmine loogiline samm nende valideerimine. Ühtse integreeritud repositooriumi abil on see protsess oluliselt lihtsustatud: ei ole enam mitmeid vastuolulisi skeeme, dubleerivaid struktuure ja keerulisi seoseid tabelite vahel. Kogu teave viiakse vastavusse ühtse andmemudeliga, mis kõrvaldab sisemised vastuolud ja kiirendab valideerimisprotsessi. valideerimine ja andmete kvaliteedi tagamine on kõigi äriprotsesside nurgakiviks ja me vaatleme neid üksikasjalikumalt raamatu asjakohastes peatükkides.

Viimases etapis rühmitatakse, filtreeritakse ja analüüsitakse andmeid. Nende suhtes rakendatakse mitmesuguseid funktsioone: liitmine (liitmine, korrutamine), tabelite, veergude või ridade vahelised arvutused (joonis 2.2-4). Andmetega töötamine muutub etappide jadaks: kogumine, struktureerimine, valideerimine, teisendamine, analüütiline töötlemine ja lõpprakendustesse viimine, kus teavet kasutatakse praktiliste probleemide lahendamiseks. Selliste stsenaariumide loomisest, etappide automatiseerimisest ja töötlemisvoogude loomisest räägime lähemalt peatükkides, mis käsitlevad ETL -protsesse ja andmeportaali läheneviisi.

Seega ei tähenda digitaalne ümberkujundamine mitte ainult teabe käitlemise lihtsustamist. Tegemist on andmehalduse liigse keerukuse kõrvaldamisega, üleminekuga kaosest prognoositavuse suunas, mitmest süsteemist hallatava protsessi suunas. Mida väiksem on arhitektuuri keerukus, seda vähem koodi on vaja selle toetamiseks. Ja tulevikus võib kood kui selline kaduda täielikult, andes ruumi intelligentsetele agentidele, kes iseseisvalt analüüsivad, süstematiseerivad ja teisendavad andmeid.

Integreeritud salvestussüsteemid võimaldavad üleminekut tehisintellekti agentidele

Mida vähem keerulised on andmed ja süsteemid, seda vähem koodi on vaja kirjutada ja hooldada. Ja kõige lihtsam viis arenduse kokkuhoiduks on vabaneda koodist täielikult, asendades selle andmetega. Kui rakenduse koodi arendamisel liigutakse koodilt andmemudelitele, toimub paratamatult üleminek

andmekesksele (andmepõhisele) lähenemisviisile, sest nende kontseptsioonide taga on täiesti erinev mõtteviis.

Kui inimene otsustab töötada keskmises olevate andmetega, hakkab ta nägema nende rolli teistmoodi. Andmed ei ole enam lihtsalt rakenduste "toormaterjal" - need on nüüd alus, millele arhitektuur, loogika ja interaktsioon on ehitatud.

Traditsiooniline lähenemine andmehaldusele algab tavaliselt taotluse tasandil ja sarnaneb ehituslikult kohmakale bürokraatlikule süsteemile: mitmetasandilised kinnitused, käsitsi tehtavad kontrollid, dokumentide lõputud versioonid läbi asjakohaste tarkvaratoodete. Digitaalsete tehnoloogiate arenguga on üha enam ettevõtteid sunnitud minimeerimise põhimõttele üle minema - salvestada ja kasutada ainult seda, mis on tõesti vajalik ja mida kasutatakse.

Minimeerimise loogika on üle võetud müüjate poolt. Andmete säilitamise ja töötlemise lihtsustamiseks viiakse kasutajate töö offline-rakendustelt ja -vahenditelt üle pilveteenustele ja nn SaaS-lahendustele.

SaaS-kontseptsioon (Software as a Service ehk "tarkvara kui teenus") on üks kaasaegsete IT-infrastruktuuride peamisi suundumusi, mis võimaldab kasutajatel kasutada rakendusi interneti kaudu, ilma et nad peaksid tarkvara ise oma arvutisse paigaldama ja hooldama.

Ühest küljest on SaaS lihtsustanud skaleerimist, versioonikontrolli ning vähendanud tugi- ja hoolduskulusid, kuid teisest küljest on see lisaks sõltuvusele konkreetse rakenduse loogikast muutnud kasutaja täielikult sõltuvaks ka teenusepakkuja pilveinfrastruktuurist. Kui teenus läheb katki, võib juurdepääs andmetele ja äriprotsessidele olla ajutiselt või isegi jäädavalt blokeeritud. Lisaks salvestatakse kõik kasutaja andmed SaaS-rakendustega töötades teenusepakkuja serverites, mis tekitab turvariske ja regulatiivsetele nõuetele vastavuse riske. Tariifide või kasutustingimuste muutmine võib samuti kaasa tuua kulude suurenemise või vajaduse kiireks üleminekuks.

Tehisintellekti, LLM -agentide ja andmekeskse lähenemisviisi areng on seadnud kahtluse alla rakenduste tuleviku nende traditsioonilisel kujul ja SaaS teostamise. Kui varem pidid rakendused ja teenused haldama äriloogikat ja töötleva andmeid, siis AI-agentide tulekuga võivad need funktsioonid minna üle intelligentsetele süsteemidele, mis töötavad otse andmetega.

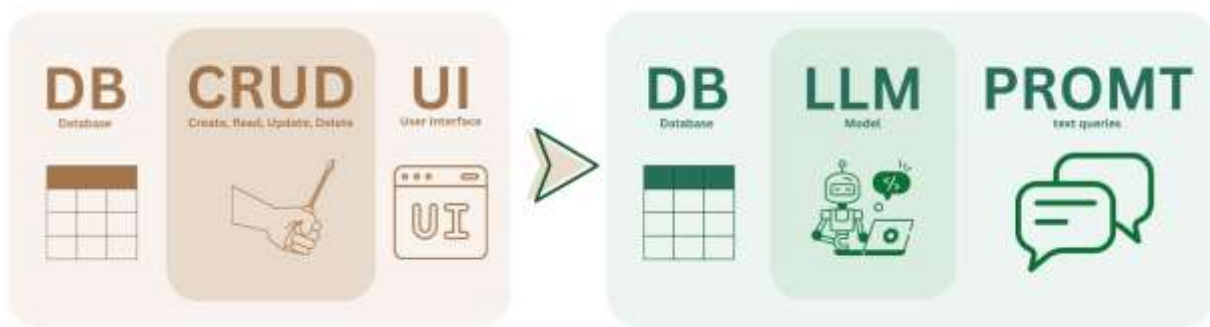
Seetõttu arutatakse IT-osakondades ja juhtimistasandil üha enam hübriidarhitektuuride üle, kus tehisintellekti -agendid ja kohapealsed lahendused täiendavad pilveteenuseid, vähendades sõltuvust SaaS-platvormidest.

Meie lähenemisviis tunnistab, et traditsioonilised ärirakendused või SaaS -rakendused võivad agentide ajastul oluliselt muutuda. Need rakendused on sisuliselt CRUD [create, read, update and delete] andmebaasid koos äriloogikaga. Kuid tulevikus võtavad selle loogika üle tehisintellekti agendid [46].

- Satya Nadella, Microsofti tegevjuht, 2024.

Andmekeskne lähenemisviis ja tehisintellekti/LLM-agentide kasutamine võib vähendada üleliigseid protsesse ja seega vähendada töötajate töökoormust. Kui andmed on korralikult korraldatud, on neid lihtsam analüüsida, visualiseerida ja otsuste tegemisel rakendada. Lõputute aruannete ja kontrollide asemel saavad spetsialistid paari klikiga või LLM-agentide abil automaatselt ligipääsu ajakohasele teabele valmis dokumentide ja armatuurlaudade kujul.

Meid abistavad andmete töötlemisel tehisintellekti vahendid (AI) ja LLM vestlused. Viimastel aastatel on täheldatud suundumust traditsioonilistelt CRUD-operatsioonidelt (create, read, update, delete) suurte keelemudelite (LLM) kasutamise suunas andmehalduses. LLMid on võimelised tõlgendama loomulikku keelt ja genereerima automaatselt sobivaid andmebaasi päringuid, mis lihtsustab andmehaldussüsteemidega suhtlemist (joonis 2.2-3).



Joonis 2.2-3 AI asendab ja integreerib salvestus- ja andmebaasilahendused, tõrjudes järkjärgult välja traditsioonilised rakendused ja CRUD -operatsioonid.

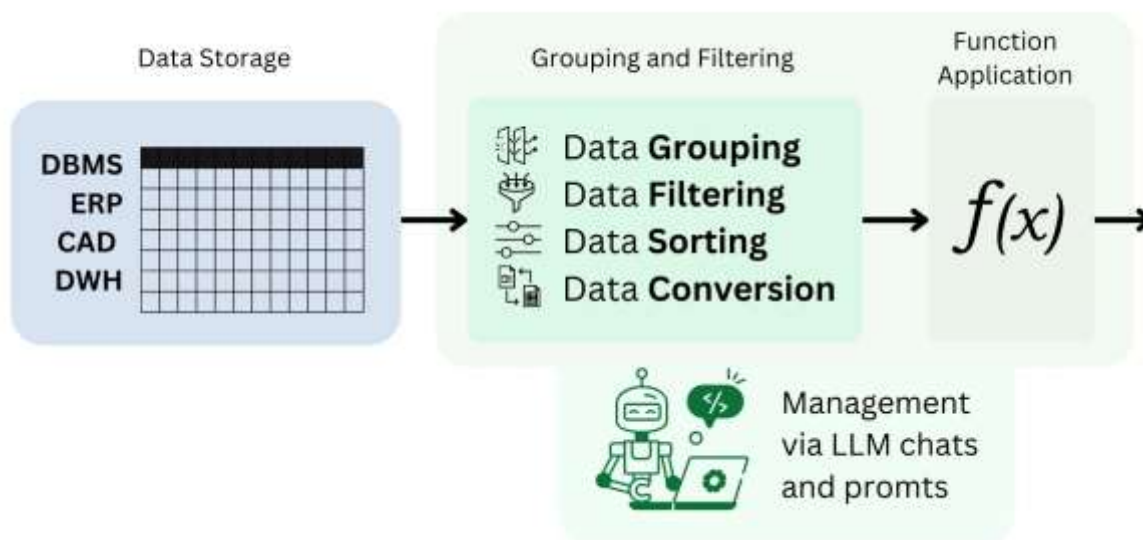
Järgmise 3-6 kuu jooksul kirjutab AI 90% koodist ja 12 kuu pärast võib peaaegu kogu kood olla loodud tehisintellekti abil [47].

- Dario Amodei, LLM tegevjuht Anthropic, märts 2025.

Vaatamata tehisintellekti arendusvahendite (nt GitHub Copilot) kiirele arengule, on 2025. aastal arendajatel selles protsessis endiselt võtmeroll. Tehisintellekti agendid muutuvad üha kasulikumateks assistentideks: nad tõlgendavad automaatselt kasutaja päringuid, genereerivad SQL- ja Pandas-päringuid (sellest lähemalt järgmistes peatükkides) või kirjutavad andmete analüüsimiseks koodi. Nii asendab tehisintellekt järkjärgult traditsioonilisi rakenduste kasutajaliideseid.

Tehisintellekti mudelite, näiteks keelemudelite, levik soodustab hübriidarhitektuuride arendamist. Selle asemel, et täielikult loobuda pilvelahendustest ja SaaS toodetest, võime näha pilveteenuste integreerimist kohalike andmehaldussüsteemidega. Näiteks võimaldab föderatiivne õppimine võimsaid tehisintellekti mudeleid, ilma et oleks vaja tundlikke andmeid pilve viia. Sel viisil saavad ettevõtted säilitada kontrolli oma andmete üle, saades samal ajal juurdepääsu arenenud

tehnoloogiatele.



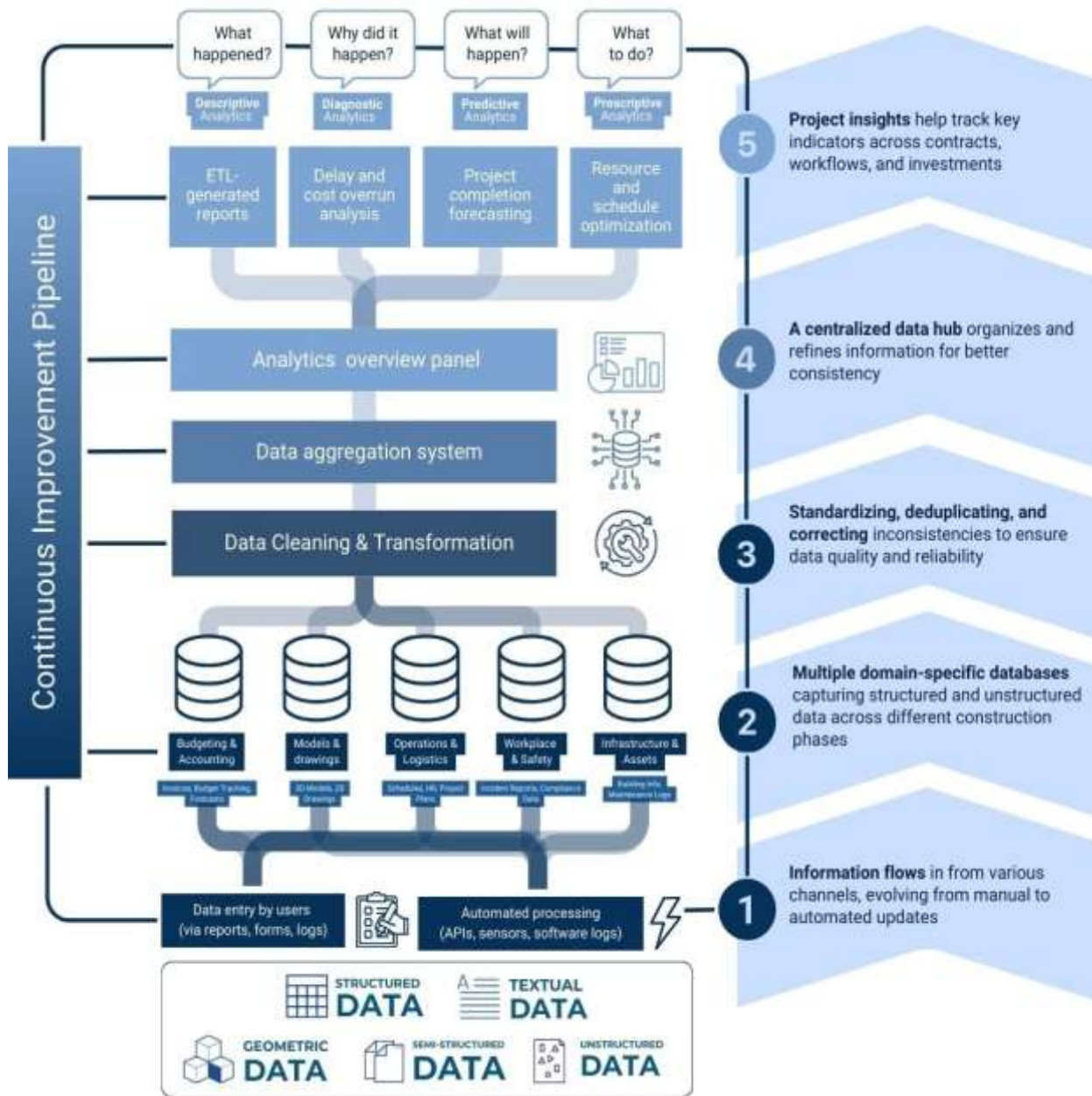
Joonis 2.2-4 Grupeerimise, filtreerimise ja sorteerimise põhitoiminguid, millele järgneb funktsioonide rakendamine, teostab LLM vestlused.

Ehitustööstuse tulevik põhineb kohapealsete lahenduste, pilveteenuste ja intelligentsete mudelite kombinatsioonil, mis töötavad koos, et luua tõhusad ja turvalised andmehaldussüsteemid. LLM võimaldab kasutajatel, kellel puuduvad põhjalikud tehnilised teadmised, suhelda andmebaaside ja andmeladudega, sõnastades oma päringud loomulikus keeles. LLM- ja AI-agentidest ja nende toimimisest räägime lähemalt peatükis "LLM-agendid ja struktureeritud andmeformaadid".

Korralikult organiseeritud andmed ja lihtsad, hõlpsasti kasutatavad LLM-võimekusega analüüsivahendid ei lihtsusta mitte ainult tööd teabega, vaid aitavad ka vähendada vigu, suurendada tõhusust ja automatiseerida protsesse.

Andmete kogumisest otsuste tegemiseni: tee automatiseerimiseni

Raamatu hilisemates osades vaatleme üksikasjalikult, kuidas spetsialistid omavahel suhtlevad ja kuidas andmed muutuvad otsuste tegemise, automatiseerimise ja tegevuse tõhususe aluseks. Joonisel 2.2-5 on näidisdiagramm, mis näitab andmetöötlustappide järjestust andmekeskse lähenemisviisi puhul. See diagramm illustreerib pideva täiustamise torujuhtme), mille osasid käsitletakse üksikasjalikult hiljem raamatus.



Joonis 2.2-5 Näide andmete pideva täiustamise torujuhtmest: andmete töötlemise ja analüüsi voog ehitusprojektides.

Keskmise suurusega ettevõtte äriprotsesse kirjeldav süsteem on üles ehitatud mitmetasandilise põhimõtte alusel. See hõlmab: andmete kogumine, puhastamine, koondamine, analüütiline töötlemine ja tulemuste põhjal otsuste tegemine. Kõiki neid etappe uurime hiljem raamatus - nii teoreetilises kontekstis kui ka praktiliste näidete kaudu:

- Esimesel tasandil toimub **andmete sisestamine** (joonis 3.1-1). Teavet saadakse nii käsitsi (aruannete, vormide, logide kaudu) kui ka automatiseeritud kujul (API-lt, anduritelt, tarkva-

rasüsteemidelt). Andmed võivad olla erineva struktuuriga: geomeetrilised, tekstilised, struktureerimata. Selles etapis on vaja standardiseerida, struktureerida ja ühtlustada teabevooge.

- Järgmine tase on **andmete töötlemine ja ümberkujundamine**. See hõlmab puhastamist, duplikaatide eemaldamist, vigade parandamist ja teabe ettevalmistamist edasiseks analüüsiks (joonis 4.2-5). See etapp on kriitilise tähtsusega, sest analüüside kvaliteet sõltub otseselt andmete puhtusest ja täpsusest.
- Seejärel **liiguvad andmed spetsiaalsetesse tabelitesse, andmeraamidesse või andmebaasidesse**, mis on jaotatud funktsionaalsete valdkondade kaupa: eelarvestamine ja raamatupidamine, mudelid ja joonised, logistika, turvalisus ja infrastruktuur. Selline jaotamine võimaldab lihtsat juurdepääsu ja teabe ristanalüüsi.
- Seejärel **koondatakse andmed ja kuvatakse analüütilisel armatuurlaual** (showcase). Siinkohal rakendatakse kirjeldavaid, diagnostilisi, prognoosivaid ja ettekirjutavaid analüüsimeetodeid. See aitab vastata põhiküsimustele (joonis 1.1-4): mis juhtus, miks see juhtus, mis juhtub tulevikus ja milliseid meetmeid tuleb võtta. Näiteks saab süsteem tuvastada viivitusi, prognoosida projektide lõpuleviimist või optimeerida ressursse.
- Viimasel tasandil tehakse **analüütilised järeldused ja** koostatakse **põhinäitajad**, mis aitavad jälgida lepingu täitmist, juhtida investeeringuid ja parandada äriprotsesse (joonis 7.4-2). Sellest teabest saab otsuste tegemise ja ettevõtte arengustrategia alus.

Samamoodi läheb andmete kogumine strateegilises juhtimises kasutamisele. Raamatu järgmistes osades vaatleme iga etappi üksikasjalikult, keskendudes andmetüüpidele, andmetöötlustehnikele, analüüsivahenditele ja tegelikele juhtumitele, kuidas neid lähenemisviise ehitustööstuses kasutatakse.

Järgmised sammud: kaose muutmine hallatavaks süsteemiks

Selles osas uurisime teabesilode probleeme ja vaatlesime süsteemi liigse keerukuse mõju ettevõtte tulemuslikkusele, analüüsides üleminekut neljandast tööstusrevolutsioonist viiendasse, kus kesksel kohal on pigem andmed kui rakendused. Nägime, kuidas siilutud infosüsteemid loovad takistusi teadmiste jagamisele ning kuidas IT-maastiku jätkuv keerukus vähendab tootlikkust ja pärsib innovatsiooni ehitustööstuses.

Selle osa kokkuvõtteks tasub välja tuua peamised praktilised sammud, mis aitavad teil rakendada käsitletud lähenemisviise oma igapäevastes ülesannetes:

- Visualiseerige oma teabemaastikku
 - ☐ Looge visuaalne kaart andmeallikatest (Miro, Figma, Canva), millega te regulaarselt töötate.
 - ☐ Lisage sellele kaardile süsteemid ja rakendused, mida te oma töös kasutate.
 - ☐ tuvastada võimalikud dubleerivad funktsioonid ja üleliigsed lahendused.
 - ☐ Määrata kindlaks kriitilised punktid, kus süsteemidevahelise andmeedastuse käigus võib tekkida andmekadu või -vigastus.
- Rakendada personaliseeritud andmehalduse tavasid
 - ☐ Üleminek rakendustelt andmete kui protsesside põhivarade asemel andmete

kasutamisele.

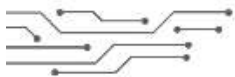
- ☐ läbipaistvuse tagamiseks dokumenteerida andmeallikad ja töötlemismeetodid
- ☐ Töötada välja mehhanismid andmete kvaliteedi hindamiseks ja parandamiseks
- ☐ Püüdke tagada, et andmed sisestatakse üks kord ja neid kasutatakse korduvalt - see on tõhusa protsessikorralduse alus.

■ Andmekeskse (andmepõhise) lähenemisviisi edendamine oma meeskonnas.

- ☐ Teha ettepanek kasutada standardiseeritud ja ühtseid vorminguid vastastikuse andmevahetuse jaoks.
- ☐ tõstatada meeskonnakoosolekutel regulaarselt andmete kvaliteedi ja kättesaadavusega seotud küsimusi.
- ☐ Tutvuge avatud lähtekoodiga alternatiividega, mida kasutate oma probleemide lahendamiseks.

Alustage väikselt - valige üks konkreetne protsess või andmekogum, mis on teie töö jaoks kriitilise tähtsusega, ja rakendage sellele andmekeskne lähenemine, nihutades fookuse tööriistadelt andmetele. Saavutades edu ühe katseprojekti käigus, saate mitte ainult praktilisi kogemusi, vaid ka selge ülevaate uue meetodika kasulikkusest oma meeskonnale. Enamiku nende sammude lõpuleviimisel, kui teil on küsimusi, võite küsida selgitusi ja abi mis tahes ajakohastatud LLM-ilt.

Raamatu järgmistes osades käsitleme üksikasjalikumalt andmete struktureerimise ja ühtlustamise tehnikaid ning uurime praktilisi lähenemisviise heterogeense teabe integreerimiseks. Erilist tähelepanu pööratakse üleminekule eraldiseisvatelt silodelt ühtsetele andmeökosüsteemidele, mis mängivad ehitussektori digitaalses ümberkujundamises võtmerolli.





III OSA

ANDMERAAMISTIK EHTUSALASTES ÄRIPROTSESSIDES

Kolmandas osas kujundatakse terviklik arusaam andmete tüpoloogiast ehituses ja nende tõhusa korraldamise meetoditest. Analüüsitakse struktureeritud, struktureerimata, poolstruktureeritud, tekstiliste ja geomeetriliste andmetega töötamise omadusi ja eripärasid ehitusprojektide kontekstis. Vaadatakse läbi kaasaegsed salvestusformaadid ja protokollid teabe vahetamiseks erinevate tööstuses kasutatavate süsteemide vahel. Kirjeldatakse praktilisi vahendeid ja tehnikaid mitmeformaadiliste andmete teisendamiseks ühtsesse struktureeritud keskkonda, sealhulgas seda, kuidas integreerida CAD (BIM) andmeid. Tehakse ettepanekuid lähenemisviiside kohta, kuidas tagada andmete kvaliteet standardimise ja valideerimise kaudu, mis on ehitusarvutuste täpsuse seisukohalt kriitilise tähtsusega. Üksikasjalikult analüüsitakse kaasaegsete tehnoloogiate (Python Pandas, LLM -mudelid) kasutamise praktilisi aspekte koos koodinäidetega, et lahendada tüüpilisi probleeme ehitussektoris. Põhjendatakse kompetentsikeskuse (CoE) kui teabehalduse lähenemisviiside koordineerimise ja standardiseerimise organisatsioonilise struktuuri loomise väärtust.

PEATÜKK 3.1.

ANDMETÜÜBID EBITUSES

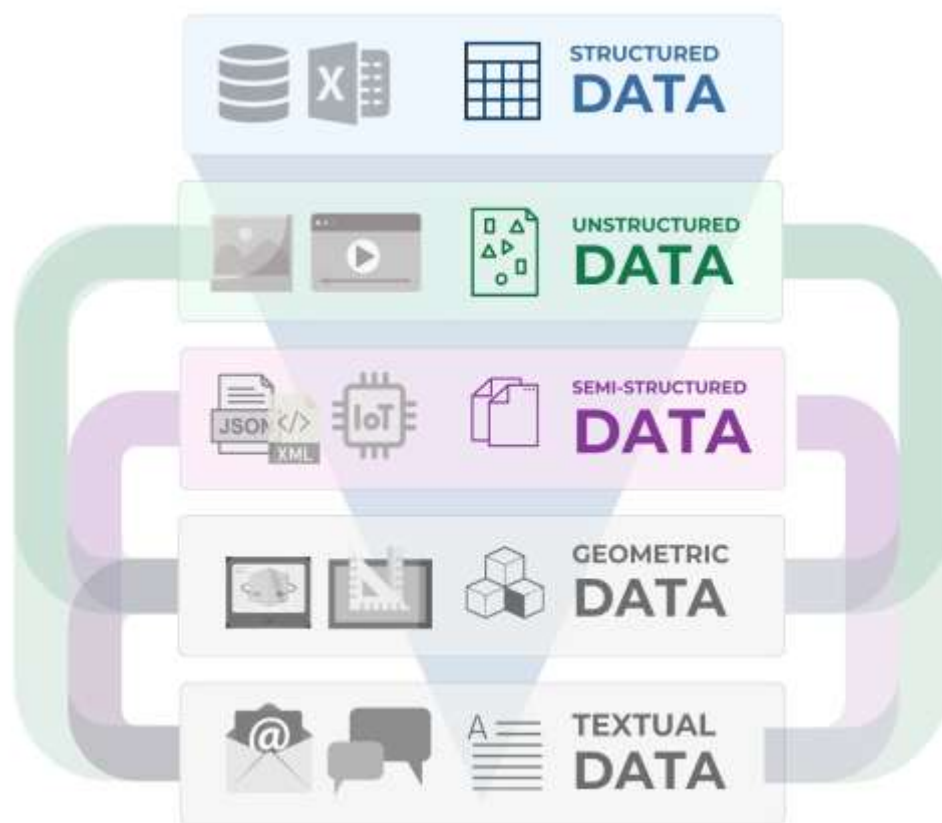
Kõige olulisemad andmetüübid ehitussektoris

Kaasaegses ehitustööstuses täidetakse ettevõtete süsteemid, rakendused ja andmelaod aktiivselt eri liiki ja vorminguga teavet ja andmeid (joonis 3.1-1). Vaatleme lähemalt peamisi andmetüüpe, mis moodustavad kaasaegse ehitustööstuses tegutseva ettevõtte infomaastiku:

- **Struktureeritud** andmed: nendel andmetel on selge organisatsiooniline struktuur, nt Exceli tabelid ja relatsioonilised andmebaasid.
- **Struktureerimata** andmed: see on teave, mis ei ole korraldatud rangete reeglite järgi. Sellised andmed on näiteks tekst, video, fotod ja helisalvestised.
- **Lahtiselt struktureeritud** andmed: need andmed asuvad struktureeritud ja struktureerimata andmete vahepealsel kohal. Need sisaldavad struktuurielemente, kuid see struktuur ei ole alati selge või on sageli kirjeldatud erinevate skeemide kaudu. Poolstruktureeritud andmete näited ehituses on: tehnilised kirjeldused, projektdokumentatsioon või arenguaruanded.
- **Tekstiandmed**: hõlmab kõike, mis on saadud suulisest ja kirjalikust suhtlusest, näiteks e-kirju, koosolekute ja kohtumiste protokollid.
- **Geomeetrilised** andmed: need andmed pärinevad CAD-programmidest, milles spetsialistid loovad projekti elementide geomeetrilisi andmeid visualiseerimiseks, mahuväärtuste kinnitamiseks või kokkupõrgete kontrollimiseks.

Oluline on märkida, et geomeetrilised ja tekstilised (tähtnumbrilised) andmed ei ole eraldi kategooria, vaid võivad esineda kõigis kolmes andmetüübis. Näiteks geomeetrilised andmed võivad olla osa nii struktureeritud andmetest (parameetrilised CAD-formaadid) kui ka struktureerimata andmetest (skaneeritud joonised). Tekstiandmed võivad samamoodi olla nii organiseeritud andmebaasides (struktureeritud andmed) kui ka selge struktuurita dokumentidena.

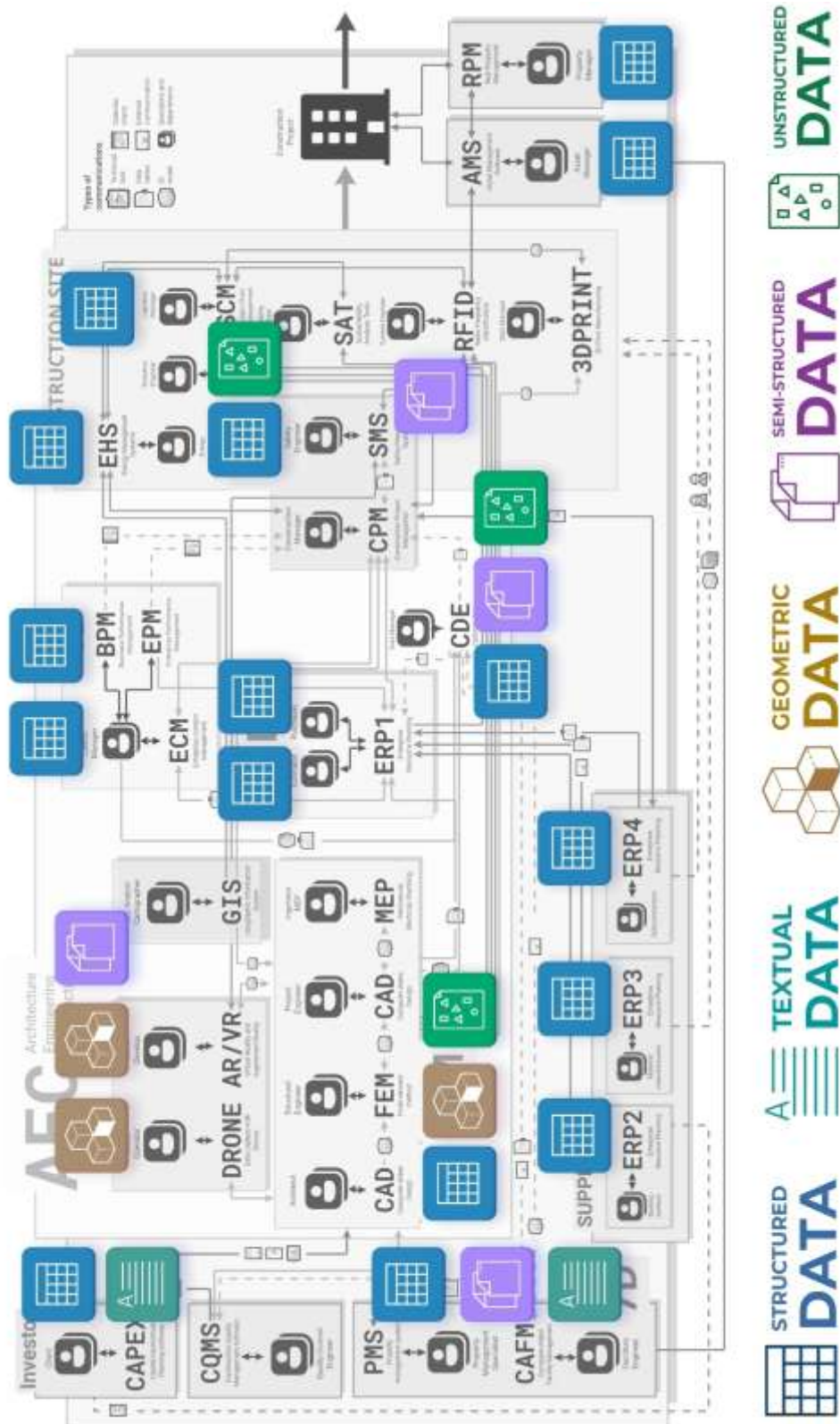
Ehitusettevõtte iga andmetüüp on ainulaadne element ettevõtte infovarade mosaiigis. Struktureerimata andmetest, nagu näiteks ehitusplatside pildid ja koosolekute helisalvestised, kuni struktureeritud andmestikuni, sealhulgas tabelid ja andmebaasid, mängib iga element ettevõtte teabemaastiku kujundamisel olulist rolli.



Joonis 3.1-1 Insenerid ja andmehaldurid peavad õppima töötama kõigi ehitustööstuses kasutatavate andmetüüpidega.

Järgnevalt on esitatud näidisloetelu vaid mõnedest ehituses kasutatavatest süsteemidest ja nendega seotud andmetüüpidest (joonis 3.1-2):

- **ERP** (Enterprise Resource Planning) - käitleb üldiselt struktureeritud andmeid, et aidata hallata ettevõtte ressursse ja integreerida erinevaid äriprotsesse.
- **CAD** (Computer-Aided Design) koos **BIMiga** (Building Information Modeling) - kasutab geomeetrilisi ja poolstruktureeritud andmeid ehitusprojektide projekteerimiseks ja modelleerimiseks, tagades projekteerimisfaasis teabe täpsuse ja järjepidevuse.
- **GIS** (Geinfosüsteemid) - töötab geomeetriliste ja struktureeritud andmetega, et luua ja analüüsida kartograafilisi andmeid ja ruumilisi seoseid.
- **RFID** (raadiosagedustuvastus) - kasutab poolstruktureeritud andmeid materjalide ja seadmete tõhusaks jälgimiseks ehitusplatsil, kasutades raadiosagedustuvastust.
- **ECM** (Engineering Content Management) on süsteem insenerandmete ja dokumentatsiooni, sealhulgas poolstruktureeritud ja struktureerimata andmete, näiteks tehniliste jooniste ja projekteerimisdokumentide haldamiseks.

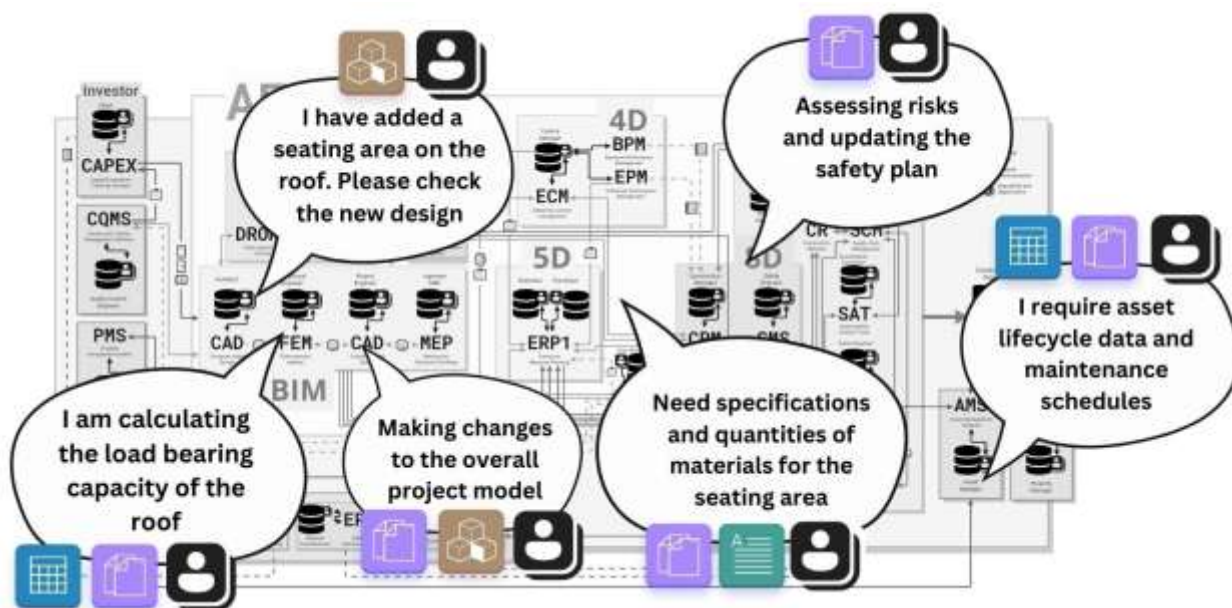


Joonis 3.1-2 Erinevad vormingud ja andmed täidavad erinevaid süsteeme, mis nõuavad tõlkimist keeruliseks integreerimiseks sobivasse vormi.

Need ja paljud teised ettevõtte süsteemid haldavad mitmesuguseid andmeid, alates struktureeritud tabeliandmetest kuni keeruliste geomeetriliste mudeliteni, pakkudes integreeritud koostööd projekteerimis-, planeerimis- ja ehitusjuhtimisprotsessides.

Lihtsustatud dialoogi näites (joonis 3.1-3) vahetatakse ehitusprojekti spetsialistide vahel eri liiki andmeid:

- 🔊 **Arhitekt:** "Võttes arvesse kliendi soovi, olen lisanud katusele istumisala. Palun vaadake uut kujundust" (geomeetrilised andmed - mudel).
- 🔊 **Ehitusinsener:** "Projekt on vastu võetud. Arvutan katuse kandevõime uue puhkeala jaoks" (struktureeritud ja poolstruktureeritud andmed - arvutustabelid).
- 🔊 **Ostujuht:** "Vajadus puhkeala materjalide spetsifikatsioonide ja koguste kohta, et korraldada ostu" (tekstilised ja poolstruktureeritud andmed - nimekirjad ja spetsifikatsioonid).
- 🔊 **Töötervishoiu ja tööohutuse insener:** "Sai andmeid uue ala kohta. Hindan riske ja ajakohastan ohutusplaani" (poolstruktureeritud andmed - dokumendid ja plaanid).
- 🔊 **BIM-spetsialist** -modelleerimine: "Muudatuste tegemine üldises projektimudelil, et kohendada töödokumentatsiooni" (geomeetrilised andmed ja poolstruktureeritud andmed).
- 🔊 **Projekti juht:** "Lisan uue puhkeala tööplaani. Uuendan ajakavasid ja ressursse projektijuhtimissüsteemis" (struktureeritud ja poolstruktureeritud andmed - ajakavad ja plaanid).
- 🔊 **Rajatiste hooldusspetsialist (FM):** "Koostan andmed puhkeala tulevase hoolduse kohta ja sissestan need kinnisvarahaldussüsteemi" (struktureeritud ja poolstruktureeritud andmed - juhised ja hooldusplaanid).



Joonis 3.1-3 Spetsialistide vaheline suhtlus toimub nii teksti kui ka andmete tasandil.

Iga spetsialist töötab erinevat tüüpi andmetega, et tagada meeskonna tõhus koostöö ja projekti edu. Struktureeritud, poolstruktureeritud ja struktureerimata andmete erinevuste mõistmine võimaldab teil ära tunda iga andmetüübi ainulaadset rolli digitaalsetes äriprotsessides. Oluline ei ole mitte ainult teada, et on olemas erinevad andmevormid, vaid ka mõista, kuidas, kus ja miks neid kasutatakse.

Veel mõni aeg tagasi tundus idee ühendada nii erinevaid andmeid ambitsioonikas, kuid raskesti realiseeritav. Tänapäeval on see juba osa igapäevasest praktikast. Erinevate skeemide ja struktuuride andmete integreerimine on muutunud kaasaegse infosüsteemide arhitektuuri lahutamatuks osaks.

Järgnevates peatükkides vaatleme üksikasjalikult peamisi standardeid ja lähenemisviise, mis võimaldavad struktureeritud, poolstruktureeritud ja struktureerimata andmeid ühendada ühtseks terviklikuks vaateks. Erilist tähelepanu pööratakse struktureeritud andmetele ja relatsioonilistele andmebaasidele kui peamistele mehhanismidele teabe säilitamiseks, töötlemiseks ja analüüsimiseks ehitussektoris.

Struktureeritud andmed

Ehitustööstuses pärineb teave paljudest allikatest - joonistest, spetsifikatsioonidest, graafikutest ja aruannetest. Et seda teabevoolu tõhusalt hallata, peab see olema struktureeritud. Struktureeritud andmed võimaldavad teavet korraldada mugavas, loetavas ja juurdepääsetavas vormis.

JB Knowledge'i 5. iga-aastase ehitustehnoloogia aruande [17] kohaselt jälgib ja hindab 67% ehitusprojektide juhtimise spetsialistidest töö tulemuslikkust käsitsi või kasutades arvutustabeleid.

Mõned kõige levinumad struktureeritud andmeformaadid on XLSX ja CSV. Neid kasutatakse laialdaselt teabe salvestamiseks, töötlemiseks ja analüüsimiseks tabelites. Sellistes tabelites on andmed esitatud ridade ja veergude kujul, mis teeb nende lugemise, töötlemise ja analüüsimise lihtsaks.

XLSX, Microsofti loodud **formaad**, mis põhineb XML -struktuuride kasutamisel ja mille archiveerimisel kasutatakse ZIP-algoritmi. Selle formaadi peamised omadused:

- Keeruliste valemite, graafikute ja makrode tugi.
- Võimalus salvestada andmeid erinevatel lehtedel ning vormindada teavet.
- Optimeeritud Microsoft Excelile, kuid ühildub ka teiste kontoriprogrammidega.

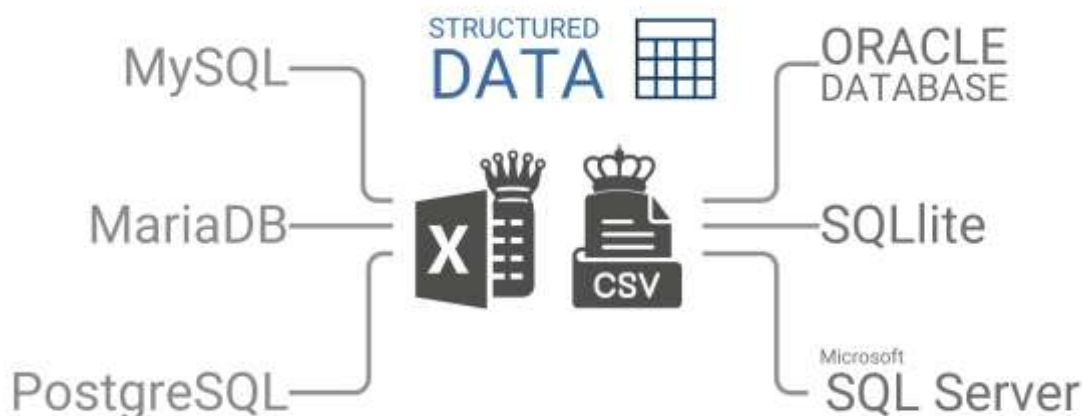
CSV-vorming on tavaline tekstifail, milles väärtused on eraldatud komadega, semikoolonitega või muude eraldusmärkidega. Peamised eelised:

- Universaalne ühilduvus erinevate programmide ja operatsioonisüsteemidega.
- Lihtne import/eksport andmebaasidesse ja analüütilistesse süsteemidesse.
- Lihtne töötlemine isegi tekstiredaktorites.

CSV ei toeta siiski vameid ja vorminguid, seega on selle peamine rakendusala andmevahetus süsteemide vahel ja teabe massiline ajakohastamine. Tänu oma mitmekülsusele ja platvormi sõltumatusle on CSV muutunud populaarseks vahendiks andmeedastuseks heterogeensetes IT-keskkondades.

Kaks formaati XLSX ja CSV toimivad ühendusülina erinevate struktureeritud andmeid käitlevate

süsteemide vahel (joonis 3.1-4). Need on eriti kasulikud ülesannetes, kus on oluline loetavus, käsitsi toimetamine ja põhiline ühilduvus



Joonis 3.1-4 XLSX ja CSV formaadid on ühenduslüliks erinevate süsteemide vahel, mis töötavad struktureeritud andmetega.

Platvormisõltumatus muudab CSV kõige populaarsemaks andmeedastusformaadiks heterogeensetes IT-keskkondades ja -süsteemides.

XLSX ja CSV ei ole siiski mõeldud suure jõudlusega arvutamiseks või suurte andmemahtude pikaajaliseks säilitamiseks. Sellistel eesmärkidel kasutatakse moodsamaid struktureeritud vorminguid, nagu Apache Parquet, Apache ORC, Feather, HDF5. Neid vorminguid käsitletakse üksikasjalikumalt käesoleva raamatu 9. osa peatükis "Suurandmete salvestamine: populaarsete vormingute ja nende tõhususe analüüs"

Praktikas kasutatakse Excel koos XLSX-vorminguga sagedamini väikeste ülesannete ja rutiinsete protsesside automatiseerimiseks. Keerulisemad stsenaariumid eeldavad selliste andmehaldussüsteemide kasutamist nagu ERP, PMIS CAFM, CPM, SCM ja muud (joonis 3.2-1). Need süsteemid salvestavad struktureeritud andmeid, millel põhineb ettevõtte infovoogude korraldamine ja juhtimine.

Ehitustööstuses kasutatavad kaasaegsed andmehalduse infosüsteemid põhinevad struktureeritud andmetel, mis on organiseeritud tabelite kujul. Suurte andmemahtude usaldusväärseks, skaleeritavaks ja terviklikuks haldamiseks pöörduvad rakenduste ja süsteemide arendajad relatsiooniliste andmebaaside haldussüsteemide (RDBMS) poole.

Relatsioonilised andmebaasid RDBMS ja SQL päringukeel

Andmete tõhusaks säilitamiseks, töötlemiseks ja analüüsimiseks on **relatsioonilised andmebaasid (RDBMS)** andmesalvestussüsteemid, mis korraldavad teavet tabelitesse, mille vahel on määratletud seosed.

Andmebaasides (RDBMS) korraldatud andmed ei ole lihtsalt digitaalne teave; need on aluseks tehingutele ja eri süsteemide vahelisele suhtlusele.

Järgnevalt on esitatud mõned kõige levinumad relatsiooniliste andmebaaside juhtimissüsteemid (RDBMS) (joonis 3.1-5):

- **MySQL** (avatud lähtekoodiga) on üks populaarsemaid RDBMS, mis on osa LAMP stackist (Linux, Apache, MySQL, PHP /Perl/Python). Seda kasutatakse laialdaselt veebiarenduses tänu selle lihtsusele ja suurele jõudlusele.
- **PostgreSQL** (avatud lähtekoodiga) on võimas objektrelatsiooniline süsteem, mis on tuntud oma usaldusväärsuse ja täiustatud funktsioonide poolest. See sobib keeruliste ettevõtete lahenduste jaoks.
- **Microsoft SQL Server** on Microsofti kommertssüsteem, mida kasutatakse laialdaselt ettevõtluskeskkondades tänu selle integreerimisele teiste ettevõtte toodetega ja kõrgele turvalisuse tasemele.
- **Oracle Database** on üks võimsamaid ja usaldusväärsemaid DBMSi, mida kasutatakse suuretegevõtetes ja kriitiliste rakenduste puhul.
- **IBM DB2** - suunatud suurte ettevõtetele, pakkudes suurt jõudlust ja veatolerantsust.
- **SQLite** (avatud lähtekoodiga) on kerge sisseehitatud andmebaas, mis sobib ideaalselt mobiilirakenduste ja iseseisvate süsteemide, näiteks CAD projekteerimistarkvara (BIM) jaoks.

Populaarsed andmebaaside haldussüsteemid ehitusvaldkonnas - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 ja SQLite - töötavad struktureeritud andmetega. Kõik need andmebaasisüsteemid on võimsad ja paindlikud lahendused mitmesuguste äriprotsesside ja rakenduste haldamiseks, alates väikestest veebilehtedest kuni suuremahuliste ettevõtlussüsteemideni (joonis 3.2-1).

Statista [48] andmetel moodustavad relatsioonilised andmebaaside juhtimissüsteemid (RDBMS) umbes 72% kõigist 2022. aastal kasutatavatest andmebaaside juhtimissüsteemidest.

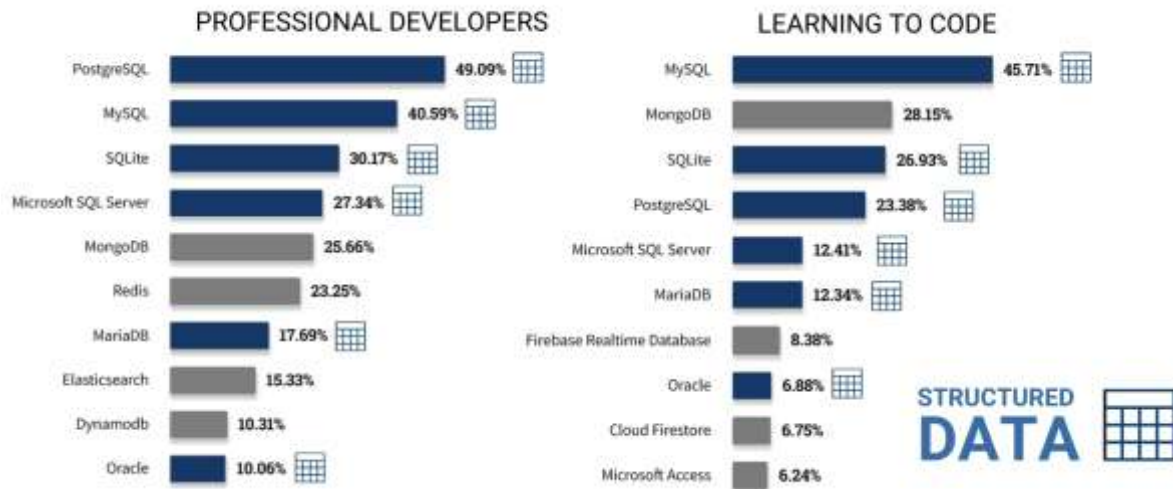
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open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

Joonis 3.1-5 Struktureeritud andmebaaside (sinisega märgitudkasutamise populaarsus DBMSide pingereas (põhineb [49])).

Avatud lähtekoodiga andmebaase on üsna lihtne paigaldada - isegi ilma ulatuslike tehniliste teadmisteta. Avatud lähtekoodiga süsteemid, nagu PostgreSQL, MySQL või SQLite, on saadaval tasuta ja töötavad enamikus operatsioonisüsteemides: Windows, macOS ja Linux. Tuleb vaid minna projekti ametlikule veebisaidile, laadida alla installija ja järgida juhiseid. Enamasti ei võta paigaldamine aega rohkem kui 10-15 minutit. Ühe sellise andmebaasi modelleerimise ja loome raamatu neljandas osas (joonis 4.3-8).

Kui teie ettevõtte kasutab pilveteenuseid (näiteks Amazon Web Services, Google Cloud või Microsoft Azure), saate andmebaasi paari klikiga kasutusele võtta - platvorm pakub teile paigaldamiseks valmis malle. Tänu koodi avatusele on selliseid andmebaase lihtne teie ülesannete jaoks kohandada ning suur kasutajaskond aitab teil alati leida lahenduse mis tahes probleemile.

RDBMS on endiselt aluseks paljudele ärirakendustele ja analüüsiplatvormidele (joonis 3.1-6), mis võimaldavad ettevõtetel andmeid tõhusalt salvestada, töödelda ja analüüsida ning seega teha teadlikke ja õigeaegseid otsuseid.



Joonis 3.1-6 StackOverFlow (suurim IT-foorum) arendajate küsitlus selle kohta, milliseid andmebaase nad kasutasid eelmisel aastal ja milliseid nad soovivad kasutada järgmisel aastal (RDBMSid on sinisega esile tõstetud) (põhineb [50]).

RDBMS pakub usaldusväärsust, andmete järjepidevust, tehingutoetust ja kasutab võimsat päringukeelt - SQL (Structured Query Language), mida kasutatakse sageli analüütikas ja mis võimaldab hõlpsasti hankida, muuta ja analüüsida andmebaasides salvestatud teavet. SQL on peamine vahend tööks andmetega relatsioonilistes süsteemides.

SQL - päringud andmebaasides ja uued suundumused

Relatsioonilistes andmebaasides sageli kasutatava SQL-keepe peamine eelis teiste teabehalduse viiside ees (näiteks klassikalise Exceli tabelite abil) on väga suurte andmebaaside mahtude toetamine suure päringute töötlemise kiirusega.

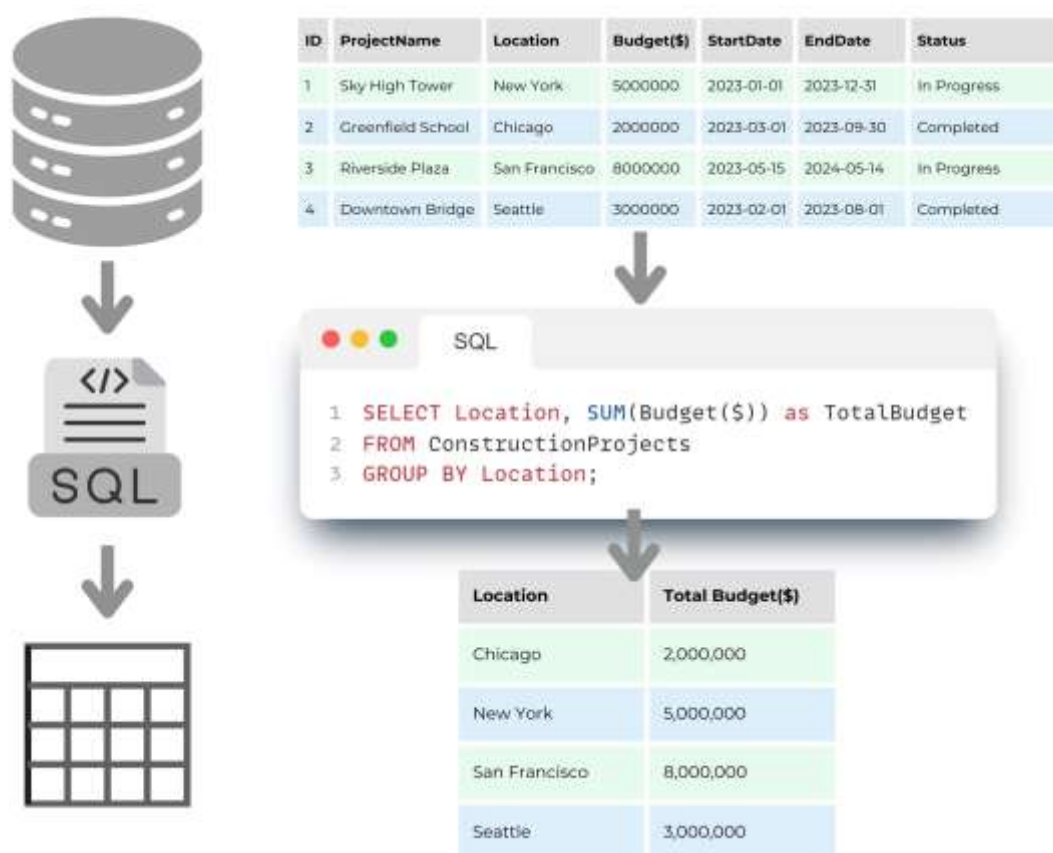
Struktureeritud päringukeel (SQL) on spetsiaalne programmeerimiskeel, mis on mõeldud teabe salvestamiseks, töötlemiseks ja analüüsimiseks relatsioonilistes andmebaasides. SQL-i kasutatakse andmete loomiseks, haldamiseks ja andmetele juurdepääsuks, võimaldades tõhusalt leida, filtreerida, kombineerida ja koondada teavet. See on peamine vahend andmetele juurdepääsuks, pakkudes mugavat ja formaliseeritud viisi teabehoidlatega suhtlemiseks.

SEQUEL-SQL süsteemide areng kulgeb läbi selliste oluliste toodete ja ettevõtete nagu Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL ja MySQL ning kulmineerub SQLite ja MariaDB [51] tekkimisega. SQL pakub tabelarvutusfunktsioone, mida Excelis ei leidu, muutes andmetöötluste skaleeritavamaks, turvalisemaks ja hõlpsasti automatiseeritavaks:

- **Andmestruktuuride loomine ja haldamine (DDL):** SQLis saab andmebaasis luua, muuta ja kustutada tabeleid, luua nendevahelisi seoseid ja määratleda andmesalvestusstruktuure. Excel seevastu töötab fikseeritud lehtede ja lahtritega, ilma selgelt määratletud seosteta

lehtede ja andmekogumite vahel.

- **Andmete manipuleerimine (DML):** SQL võimaldab andmete massilist lisamist, muutmist, kustutamist ja väljavõtete tegemist suure kiirusega, tehes keerulisi päringuid filtreerimise, sorteerimise ja tabelite ühendamise abil (joonis 3.1-7). Excelis nõuab suurte andmemahdade töötlemine käsitsi tehtavaid toiminguid või spetsiaalseid makrosid, mis aeglustab protsessi ja suurendab vigade tõenäosust.
- **Juurdepääsu kontroll (DCL):** SQL võimaldab teil eristada eri kasutajate juurdepääsuõigusi andmetele, piirates teabe muutmise või vaatamise võimalust. Excelis seevastu on juurdepääs kas jagatud (faili edastamisel) või nõuab keerulisi seadistusi koos õiguste jagamisega pilveteenuste kaudu.



Joonis 3.1-7 DMLi näide SQLis: kiire töötlemine, grupeerimine ja agregeerimine mõne koodireaga automaatselt andmetöötluks.

Excel teeb tööd andmetega lihtsaks tänu oma visuaalsele ja intuitiivsele struktuurile. Kuid kui andmemahd suureneb, väheneb Exceli jõudlus. Excelil on ka piirangud andmete mahule, mida ta saab salvestada - maksimaalselt üks miljon rida - ja jõudlus väheneb juba ammu enne selle piiri saavutamist. Seega, kuigi Excel tundub olevat parem väikeste andmehulkade visualiseerimiseks ja manipuleerimiseks, sobib SQL paremini suurte andmekogumite käsitlemiseks.

Järgmine etapp struktureeritud andmete arengus oli kolumniliste andmebaaside (Columnar Databases) teke, mis on alternatiiviks traditsioonilistele relatsioonilistele andmebaasidele, eriti kui

tegemist on oluliselt suuremate andmemahutude ja analüütiliste arvutustega. Erinevalt reaandmebaasidest, kus andmed salvestatakse rea kaupa, salvestavad veerandandmebaasid teavet veergude kaupa. Võrreldes klassikaliste andmebaasidega võimaldab see:

- Vähendage salvestusruumi, tihendades tõhusalt ühtseid andmeid veergudes.
- Kiirendab analüütilisi päringuid, kuna loetakse ainult vajalikke veerge, mitte kogu tabelit.
- Suurandmete optimeerimine ja andmete ladustamine, nt Data Lakehouse Architecture.

Rohkem räägime veergandmebaasidest, Pandasist DataFrame, Apache Parquet, HDF5, samuti nendel põhinevate Big Data -varamute loomisest andmeanalüüsi ja -töötamise eesmärgil käesoleva raamatu järgmistes peatükkides - "DataFrame: universaalne tabelandmeformaad" ja "Andmesalvestusformaadid ja töö Apache Parquet'iga: DWH -andmelaod ja Data Lakehouse arhitektuur".

Struktureerimata andmed

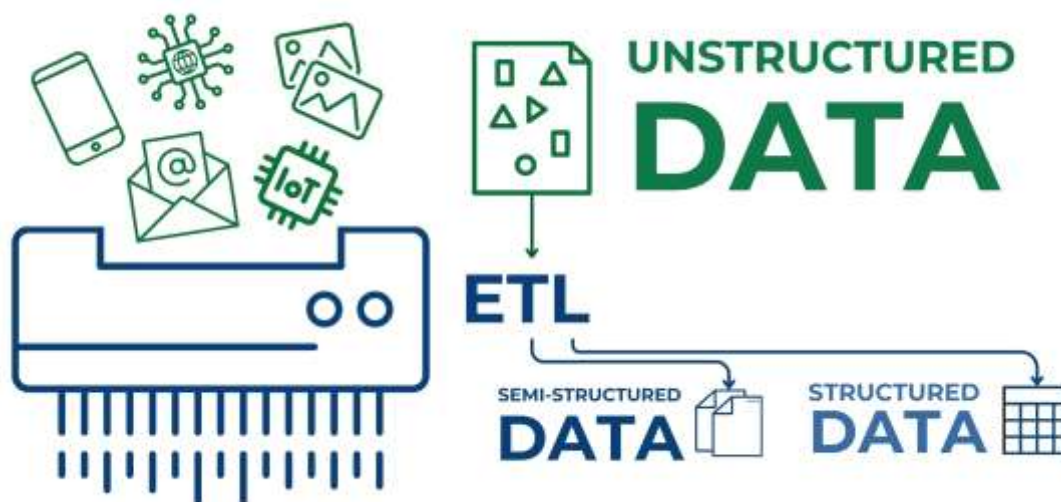
Kuigi enamik rakendustes ja infosüsteemides kasutatavatest andmetest on struktureeritud kujul, on suurem osa ehituses tekkivast teabest struktureerimata andmed - pildid, videod, tekstidokumendid, helisalvestised ja muud sisuvormid. See kehtib eriti ehitus-, käitamis- ja tehnilise järelevalve etapis, kus domineerivad visuaalne ja tekstiline teave.

Struktureerimata andmed on teave, millel puudub eelnevalt määratletud mudel või struktuur, mis ei ole organiseeritud traditsioonilistesse ridadeks ja veergudeks nagu andmebaasides või tabelites.

Üldiselt võib struktureerimata andmeid liigitada kahte kategooriasse:

- Inimese loodud struktureerimata andmed, mis hõlmavad erinevat tüüpi inimese loodud sisu: tekstidokumendid, e-kirjad, pildid, videod jne.
- Masinate poolt genereeritud struktureerimata andmeid loovad seadmed ja andurid: need hõlmavad näiteks logifaile, GPS-andmeid, asjade interneti tulemusi (IoT) ja muud telemeetrilisi andmeid ehitusplatsilt.

Erinevalt struktureeritud andmetest, mida on mugav korraldada tabelitesse ja andmebaasidesse, vajavad struktureerimata andmed enne nende integreerimist infosüsteemidesse täiendavaid töötlemisetappe (joonis 3.1-8). Selliste andmete automatiseeritud kogumise, analüüsi ja ümberkujundamise tehnoloogiate kasutamine avab uusi võimalusi ehitustõhususe parandamiseks, vigade vähendamiseks ja inimtegu mõju minimeerimiseks.



Joonis 3.1-8 Struktureerimata andmete töötlemine algab nende ümberkujundamisega poolstruktureeritud ja struktureeritud andmeteks.

Struktureerimata andmed moodustavad kuni 80% kogu teabest [52], millega spetsialistid ettevõtetes kokku puutuvad, mistõttu käsitleme nende liike ja töötlemist üksikasjalikult koos näidetega raamatu järgmistes peatükkides.

Arutelu hõlbustamiseks on tekstilised andmed liigitatud eraldi. Kuigi need on pigem struktureerimata andmete liik, nõuavad nende tähtsus ja levik ehitussektoris erilist tähelepanu.

Tekstiandmed: struktureerimata kaose ja struktureeritud kaose vahel y

Tekstiandmed hõlmavad ehitussektoris väga erinevaid vorminguid ja teabetüüpe, alates paberdokumentidest kuni mitteametlike kommunikatsioonimeetoditeni, nagu kirjad, vestlused, töökirjavahetus ja suulised kohtumised ehitusplatsil. Kõik need tekstilised andmed sisaldavad olulist teavet ehitusprojektide juhtimiseks, alates projekteerimisotsuste üksikasjadest ja plaanide muudatustest kuni ohutusega seotud arutelude ja läbirääkimisteni töövõtjate ja klientidega (joonis 3.1-9).



Joonis 3.1-9 Tekstiandmed, mis on üks populaarsemaid projektis osalejate vahelises suhtluses kasutatavaid teabetüüpe.

Tekstiline teave võib olla nii formaliseeritud kui ka struktureerimata. Formaalsete andmete hulka kuuluvad Wordi dokumendid (.doc,.docx), PDF, samuti koosolekute protokollide tekstifailid (.txt). Mitteametlikud andmed hõlmavad messengeri ja e-posti kirjavahetust, koosolekute protokolle (Teams, Zoom, Google Meet) ja arutelude helisalvestusi (.mp3,.wav), mis tuleb konverteerida tekstiks.

Kuid kui kirjalikud dokumendid, nagu ametlikud taotlused, lepingutingimused ja e-kirjad, on tavaliselt juba struktureeritud, siis suuline suhtlus ja tööalane kirjavahetus on sageli struktureerimata, mistõttu on neid raske analüüsida ja integreerida projektijuhtimissüsteemidesse.

Tekstiandmete tõhusa haldamise võti on nende teisendamine struktureeritud vormingusse. See võimaldab töödeldud teavet automaatselt integreerida olemasolevatesse süsteemidesse, mis juba töötavad struktureeritud andmetega.



Joonis 3.1-10 Tekstisisu teisendamine struktureeritud andmeteks.

Tekstilise teabe tõhusaks kasutamiseks tuleb see automaatselt struktureeritud kujul teisendada (joonis 3.1-10). See protsess hõlmab tavaliselt mitmeid etappe:

- **Tekstituvastamine (OCR)** - dokumentide ja jooniste kujutiste konverteerimine masinloetavasse formaati.
- **Tekstianalüüs (NLP)** - põhiparameetrite (kuupäevad, summad ja projekti jaoks olulised arvud) automaatne tuvastamine.
- **Andmete klassifitseerimine** - teabe kategoriseerimine (rahandus, logistika, riskijuhtimine).

Pärast äratundmist ja klassifitseerimist saab juba struktureeritud andmeid integreerida andmebaasidesse ja kasutada automatiseeritud aruandlus- ja juhtimissüsteemides.

Poolstruktureeritud ja vabalt struktureeritud andmed

Poolstruktureeritud andmed sisaldavad teatavat organiseerituse taset, kuid neil ei ole ranget skeemi ega struktuuri. Kuigi selline teave sisaldab struktureeritud elemente (nt kuupäevad, töötajate nimed ja täidetud ülesannete nimekirjad), võib esitusviis projektiti või isegi töötajati oluliselt erineda. Selliste andmete näideteks on ajaprotokollid, eduaruanded ja ajakavad, mida võib esitada erinevates vormingutes.

Poolstruktureeritud andmeid on lihtsam analüüsida kui struktureerimata andmeid, kuid need vajavad

täiendavat töötlemist, et neid saaks integreerida standardiseeritud projektijuhtimissüsteemidesse.

Töötamine poolstruktureeritud andmetega, mida iseloomustab pidevalt muutuv struktuur, kujutab endast märkimisväärset väljakutset. See tuleneb sellest, et andmete struktuuri muutlikkus nõuab eraldi individuaalset lähenemist iga poolstruktureeritud andmeallika töötlemiseks ja analüüsimiseks.

Kuid kui struktureerimata andmete käsitlemine nõuab palju vaeva, siis poolstruktureeritud andmete töötlemiseks on võimalik kasutada suhteliselt lihtsaid meetodeid ja vahendeid.

Nõrgalt struktureeritud andmed on üldisem termin, mis kirjeldab minimaalse või ebatäieliku struktuuriga andmeid. Enamasti on need tekstidokumendid, vestlused, e-kirjad, kus leidub mõningaid metaandmeid (nt kuupäev, saatja), kuid suurem osa teabest on esitatud kaootiliselt.

Ehituses leidub lahtiselt struktureeritud andmeid mitmesugustes protsessides. Näiteks võivad need sisaldada:

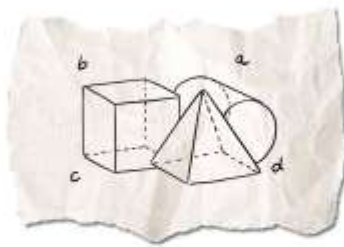
- Hinnangud ja hinnapakumised - tabelid materjalide, koguste ja kulude kohta, kuid ilma ühtse vorminguta.
- Joonised ja tehnilised diagrammid - failid PDF-formaadis või DWG-formaadis, mis sisaldavad tekstimärkusi ja metaandmeid, kuid millel puudub rangelt fikseeritud struktuur.
- Tööplaanid - andmed MS Project, Primavera P6 või muudest süsteemidest, millel võib olla erinev ekspordistruktuur.
- CAD (BIM -mudelid) - sisaldavad konstruktsiooni elemente, kuid andmete esitus sõltub tarkvarast ja projekti standardist.

Geomeetrilisi andmeid, mida toodavad CAD-süsteemid, võib liigitada samamoodi nagu poolstruktureeritud andmeid. Siiski liigitame geomeetrilised CAD- (BIM-) andmed eraldi andmetüübiks, sest neid, nagu ka tekstandmeid, saab ettevõtte protsessides sageli käsitleda eraldi andmetüübina.

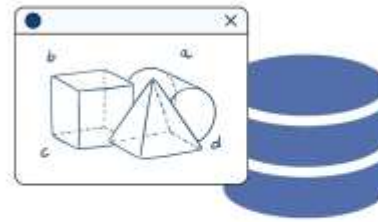
Geomeetrilised andmed ja selle rakendamine

Kui projekti elementide metaandmed on peaaegu alati salvestatud tabelite, struktureeritud või vabalt struktureeritud vormingute kujul, siis projekti elementide geomeetrilised andmed luuakse enamikul juhtudel spetsiaalsete CAD vahendite abil (joonis 3.1-11), mis võimaldavad projekti elemente üksikasjalikult visualiseerida joonte (2D) või geomeetriliste kehade (3D) kogumina.

3000 BCE - 1960s

physical medium
(artefact)

1960s to present day

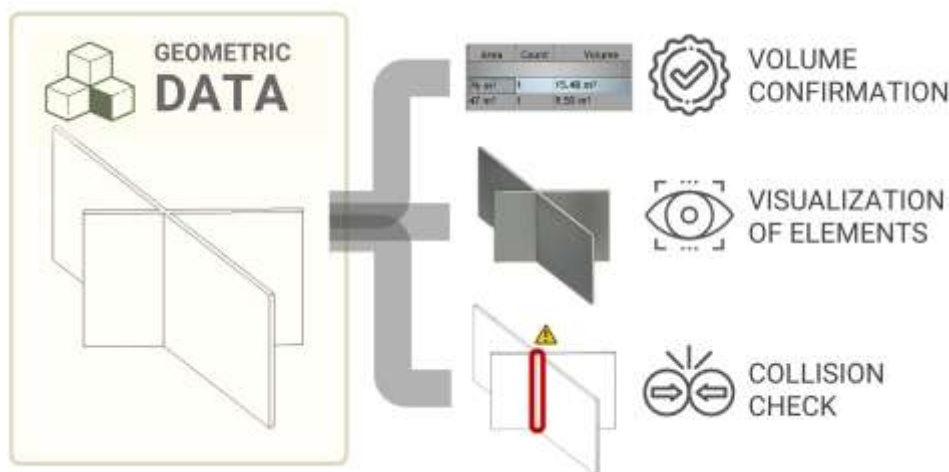
digital project data
(CAD data)

Joonis 3.1-11 CAD tööriistad on aidanud geomeetrilist teavet füüsilisest andmekandjast andmebaasi kujul üle viia.

Ehituses ja arhitektuuris geomeetriliste andmetega töötamisel võib välja tuua kolm peamist geomeetriliste andmete rakendust (joonis 3.1-12):

- **Mahtude kinnitamine:** geomeetrilised andmed, mis on genereeritud CAD-programmides (BIM), kasutades spetsiaalseid geomeetrilisi tuumasid, on vajalikud projekti elementide mahtude ja mõõtmete automaatseks ja täpseks määramiseks. Need andmed hõlmavad automaatselt arvutatud pindalasisid, mahtusid, pikkusi ja muid olulisi atribuute, mis on vajalikud planeerimiseks, eelarvestamiseks ning ressursside ja materjalide tellimiseks
- **Projekti visualiseerimine:** projekti muudatuste korral võimaldab elementide visualiseerimine automaatselt genereerida ajakohastatud joonised erinevates tasandites. Projekti visualiseerimine algstaadiumis aitab kiirendada kõigi osalejate vahelist arusaamist, et säästa ehitusprotsessi ajal aega ja ressursse.
- **Kokkupõrgete kontrollimine:** keerukates ehitus- ja insenerprojektides, kus on kriitilise tähtsusega mitme kategooria elementide (nt torud ja seinad) koostoimimine ilma "geomeetriliste konfliktideta", on kokkupõrgete kontrollimine võtmeroll. Kasutades kollisioonituvastustarkvara, saate ennetavalt tuvastada võimalikke geomeetrilisi konflikte projekti elementide vahel, vältides sellega kulukaid vigu ehitusprotsessi ajal.

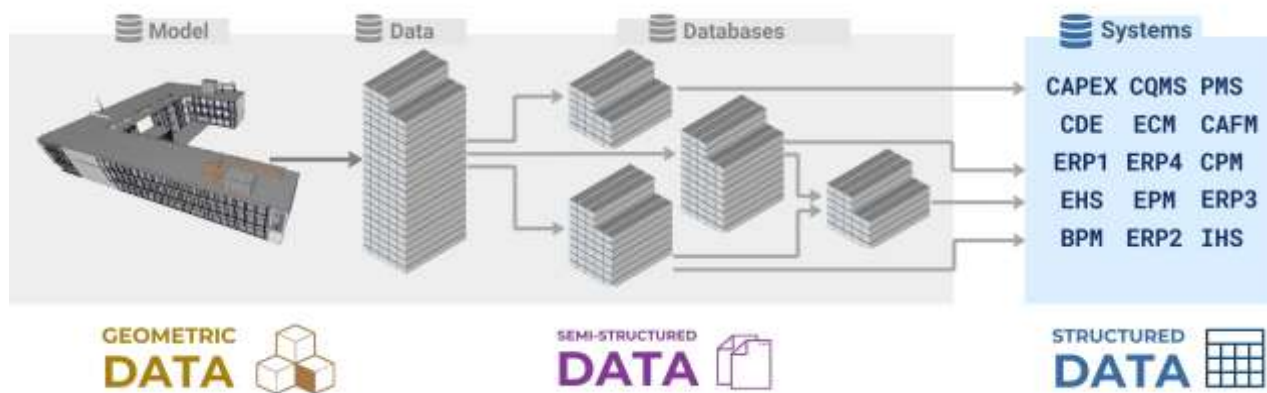
Alates inseneriplaneerimisbüroode algusest, alates esimeste keeruliste ehitiste ehitamisest, esitasid ehitusinsenerid geomeetrilist teavet jooniste, joonte ja lamedate geomeetriliste elementide kujul (papüürus, "A0" paberil või DWG, PDF, PLT formaadis), mille alusel töödejuhatajad ja kalkulaatorid (joonis 3.1-11), kogusid viimaste aastatuhandete jooksul joonte ja nurgamõõtjate abil atribuutmahte või elementide ja elementide rühmade arvu. 3.1-11), kogusid viimaste aastatuhandete jooksul joonlaudade ja transporterite abil atribuutmahte ehk elementide ja elementide rühmade koguseid.



Joonis 3.1-12 Geomeetria on aluseks elementide mahuparameetrite saamisel, mida kasutatakse projekti kulude ja ajakava arvutamiseks.

Tänapäeval on see manuaalne ja aeganõudev ülesanne lahendatud täieliku automatiseerimisega tänu moodsates CAD-vahendites (BIM) kasutusele võetud ruumilise modelleerimise võimalusele, mis võimaldab spetsiaalse geomeetrilise tuuma abil automaatselt saada iga elemendi ruumilised omadused, ilma et oleks vaja käsitsi arvutada ruumilisi parameetreid.

Kaasaegsed CAD-vahendid võimaldavad ka projekti elementide klassifitseerimist ja kategoriseerimist, et saaksite projektandmebaasist üles laadida spetsifikatsioonitabelid kasutamiseks erinevates süsteemides, näiteks kulude kalkulatsioonis, ajaplaneerimises või CO₂-arvutustes (joonis 3.1-13). Spetsifikatsioonide, QTO tabelite ja koguste hankimist ja praktilisi näiteid käsitleme peatükis "Koguste hankimine ja kvantifitseerimine".



Joonis 3.1-13 CAD-vahendid (BIM) salvestavad andmeid andmebaasidesse, mis on mõeldud teiste süsteemidega integreerimiseks ja koostöömiseks.

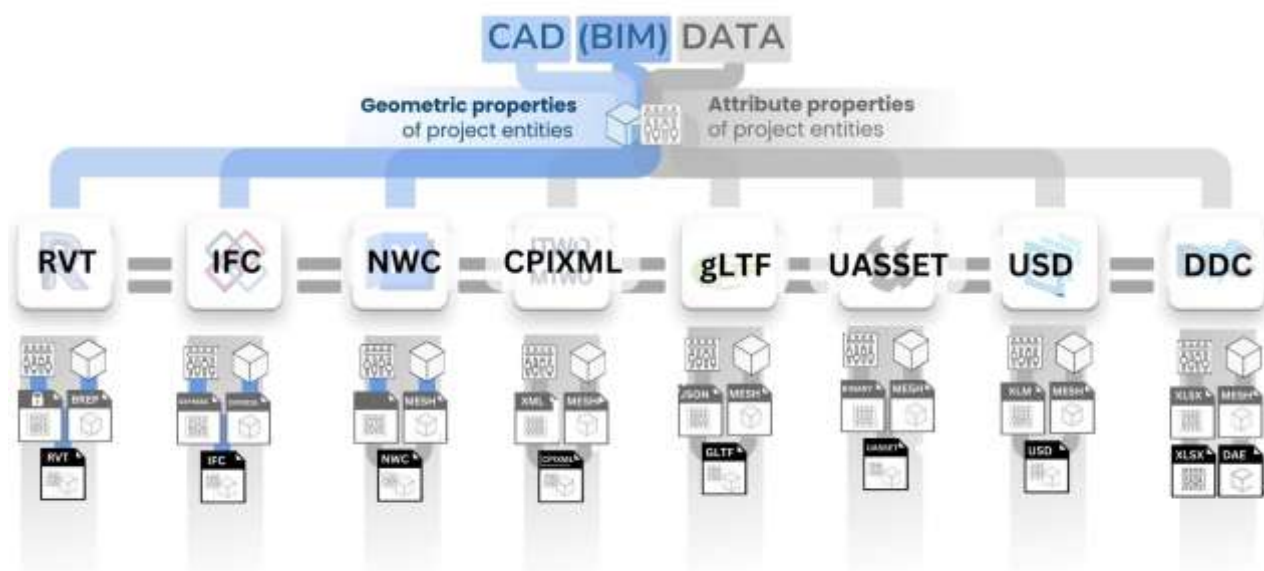
CAD-keskkonnas kasutatavate andmebaaside ja formaatide suletud olemuse tõttu on CAD-lahendustes loodud geomeetrilised andmed muutunud tegelikult eraldi teabetüübiks. See ühendab endas nii elementide geomeetriat kui ka metainfot (struktureeritud või poolstruktureeritud), mis on suletud spetsiaalsetesse failidesse ja vormingutesse.

CAD andmed: projekteerimisest andmete säilitamiseni

Kaasaegsed CAD- ja BIM-süsteemid salvestavad andmeid omaenda, sageli patenteeritud vormingutes: DWG, DXF, RVT, DGN, PLN ja muud. Need formaadid toetavad nii objektide 2D- kui ka 3D-kujutisi, säilitades mitte ainult geomeetria, vaid ka objektidega seotud atribuute. Siin on kõige levinumad:

- **DWG** on binaarne failivorming, mida kasutatakse kahemõõtmeliste (ja harvemini kolmemõõtmeliste) projekteerimisandmete ja metaandmete salvestamiseks.
- **DXF** on tekstivorming, mis on mõeldud 2D ja 3D -jooniste vahetamiseks CAD -süsteemide vahel. See sisaldab geomeetria-, kihtide ja atribuutide andmeid, toetab nii ASCII- kui ka binaarset esitust.
- **RVT** on binaarne formaat CAD mudelite salvestamiseks, sealhulgas 3D -geomeetria, elementide atribuudid, seosed ja disainiparameetrid.
- **IFC** on avatud tekstivorming ehitusandmete vahetamiseks CAD (BIM) süsteemide vahel. See sisaldab geomeetriat, objektide omadusi ja teavet nende seoste kohta.

Lisaks neile kasutatakse ka teisi formate: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Kuigi nad erinevad eesmärgilt ja avatuse tasemelt (joonis 3.1-14), võivad nad kõik esitada sama projekti infomudelit erineval kujul. Keerulistes projektides kasutatakse neid formate sageli paralleelselt, alates joonistamisest kuni projekti mudelite koordineerimiseni.



Joonis 3.1-14 Populaarsed CAD-salvestusformaadid kirjeldavad geomeetriat BREP- või MESH-parameetrite abil, mida täiendavad atribuutide andmed.

Kõik eespool nimetatud vormingud võimaldavad salvestada andmeid ehitusprojekti iga elemendi kohta ja kõik eespool nimetatud vormingud sisaldavad kahte põhilist andmetüüpi:

- **Geomeetrilised parameetrid** - kirjeldavad objekti kuju, asukohta ja mõõtmeid. Geomeetriat ja selle kasutamist käsitletakse üksikasjalikult raamatu kuuendas osas, mis on pühendatud CAD (BIM) lahendustele;

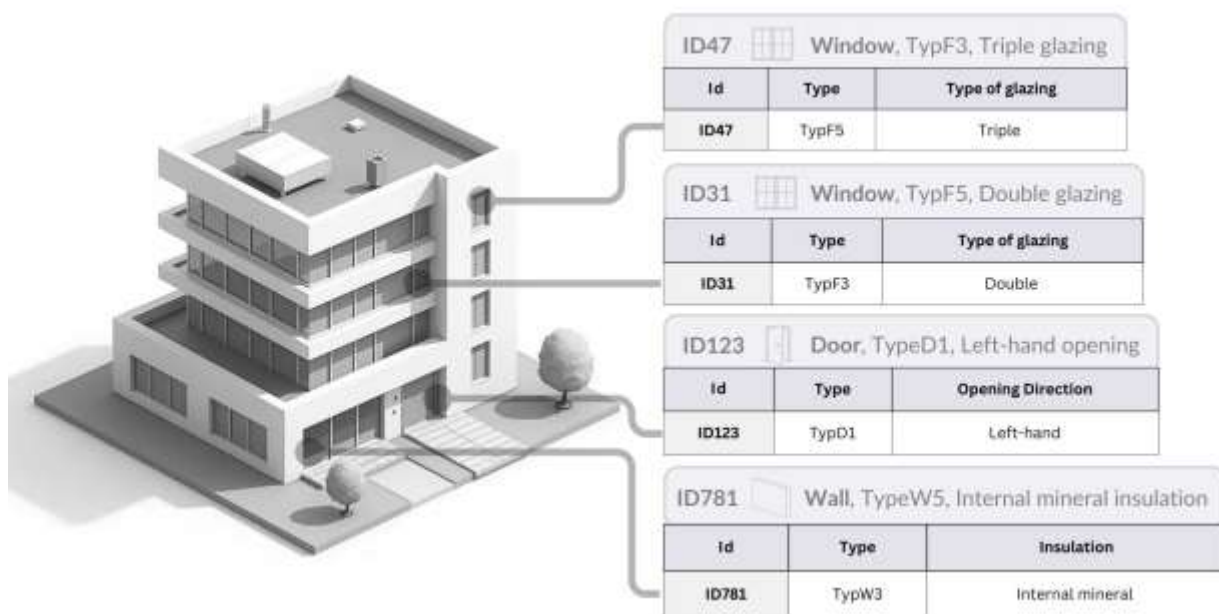
- **Atribuutide omadused** - sisaldavad mitmesugust teavet: materjalid, elementide tüübid, tehnilised omadused, unikaalsed identifikaatorid ja muud omadused, mida projekti elemendid võivad omada.

Atribuutide andmed on kaasaegsetes projektides eriti olulised, kuna need määratlevad objektide kasutusomadused, võimaldavad projekteerimist ja kuluarvestust ning tagavad projekteerimises, ehitamises ja käitamises osalejate vahelise läbiva koostoime. Näiteks:

- Akende ja uste puhul: konstruktsiooni tüüp, klaaside tüüp, avanemissuund (joonis 3.2-1).
- Seinte puhul registreeritakse teave materjalide, soojusisolatsiooni ja akustiliste omaduste kohta.
- Tehnosüsteemide puhul salvestatakse torustike, kanalite, kaablitrasside ja nende ühenduste parameetrid.

Need parameetrid võivad olla salvestatud nii CAD-(BIM -)failides endas kui ka välistes andmebaasides - ekspordi, konverteerimise või otsese juurdepääsu tulemusena sisemistele CAD-struktuuridele pöördprojekteerimise vahendite kaudu. Selline lähenemisviis hõlbustab projekteerimisalase teabe integreerimist teiste ettevõtte süsteemide ja platvormidega

CAD (BIM) kontekstis on **pöördprojekteerimine** digitaalse mudeli sisemise struktuuri väljavõtete tegemine ja analüüsimine, et taastada selle loogika, andmestruktuur ja sõltuvused ilma juurdepääsu algsetele algoritmidele või dokumentatsioonile.

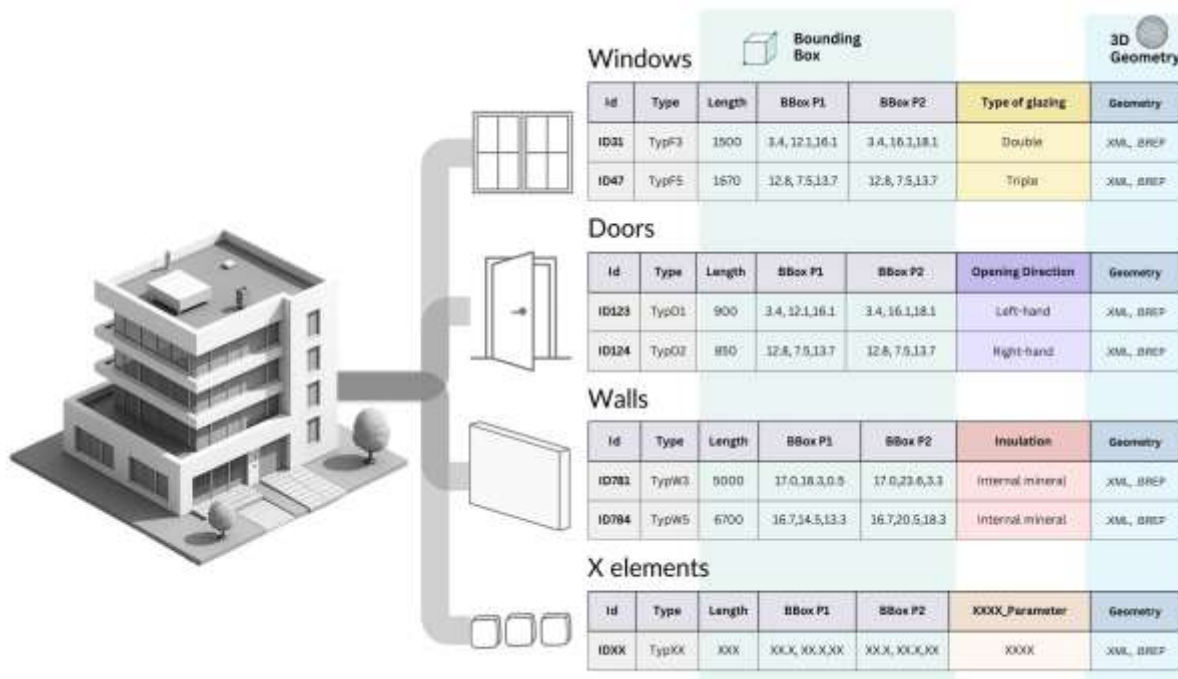


Joonis 3.1-15 Projekti element sisaldab lisaks parameetrilise või hulknurkse geomeetria kirjeldamisele ka teavet elementide parameetrite ja omaduste kohta.

Selle tulemusena moodustub iga elemendi ümber unikaalne parameetrite ja omaduste kogum, mis hõlmab nii iga objekti unikaalseid omadusi (nt identifikaator ja mõõtmed) kui ka elementide rühmade

ühiseid atribuute. See võimaldab mitte ainult analüüsida projekti üksikuid elemente-olendeid, vaid ka kombineerida neid loogilisteks rühmadeks, mida teised spetsialistid saavad seejärel kasutada oma ülesannetes ja arvutustes süsteemides ja andmebaasides.

Entiteet on tegeliku maailma konkreetne või abstraktne objekt, mida saab üheselt identifitseerida, kirjeldada ja esitada andmete kujul.

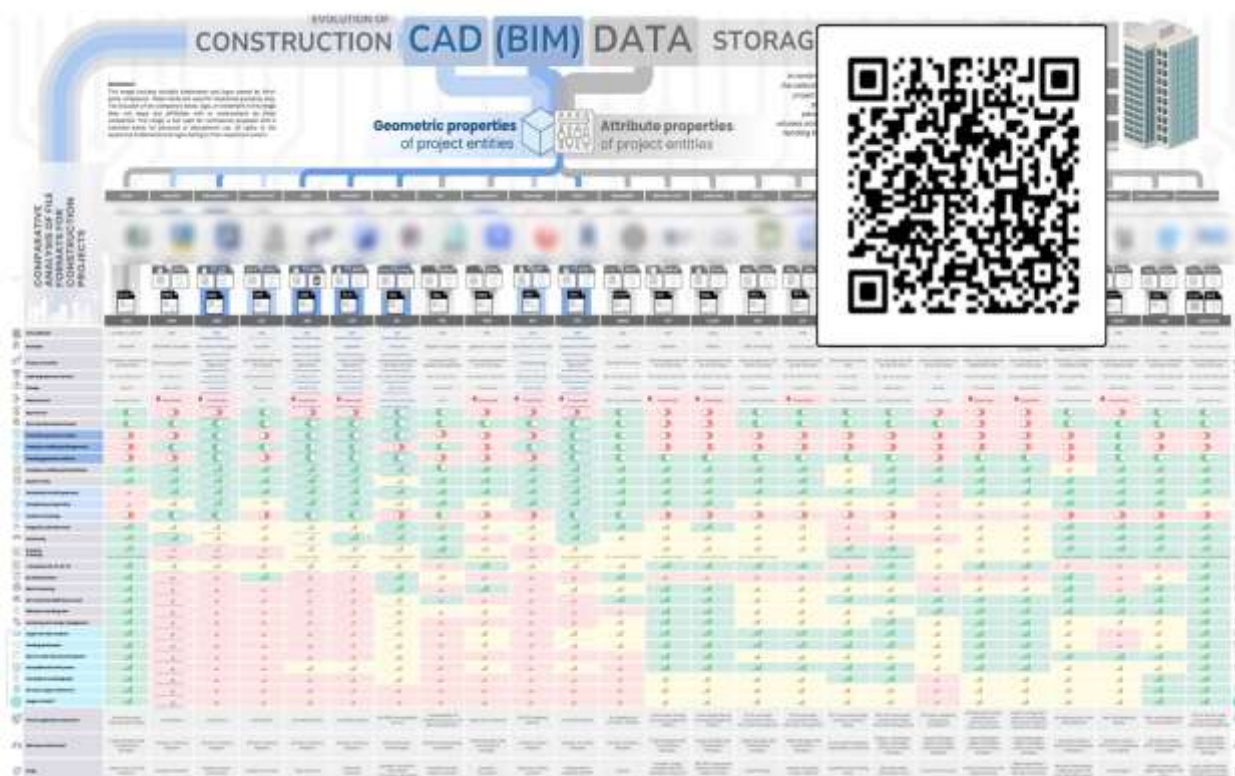


Joonis 3.1-16 Iga projektelement sisaldab atribuute, mis on kas projekteerija poolt sisestatud või arvutatud CAD-programmis.

Viimaste aastakümnete jooksul on ehitussektoris välja töötatud palju uusi CAD-formaate (BIM), mis lihtsustavad andmete loomist, säilitamist ja edastamist. Need formaadid võivad olla suletud või avatud, tabeli kujulised, parameetrilised või graafilised. Nende mitmekesisus ja killustatus raskendavad siiski märkimisväärselt andmehaldust projekti elutsükli kõikides etappides. Joonisel 3.1-17 on esitatud ehituses kasutatavate peamiste andmevahetuse vormingute võrdlustabel (täisversioon on saadaval QR-koodi abil).

Koostalitlusvõime ja CAD-andmetele juurdepääsu probleemide lahendamiseks on kaasatud BIM-juhid (BIM) ja koordinaatorid, kelle ülesanne on kontrollida eksporti, kontrollida andmete kvaliteeti ja integreerida CAD (BIM) andmete osad teistesse süsteemidesse.

Kuid formaatide suletud olemuse ja keerukuse tõttu on seda protsessi raske automatiseerida, mis sunnib spetsialiste teostama paljusid toiminguid käsitsi, ilma et nad suudaksid luua täieulatuslikke andmetöötlusprotsesse (pipeline).



Joonis 3.1-17 Tabel, milles võrreldakse peamisi andmeformaate, milles projekti elementide teavet säilitatakse [53].

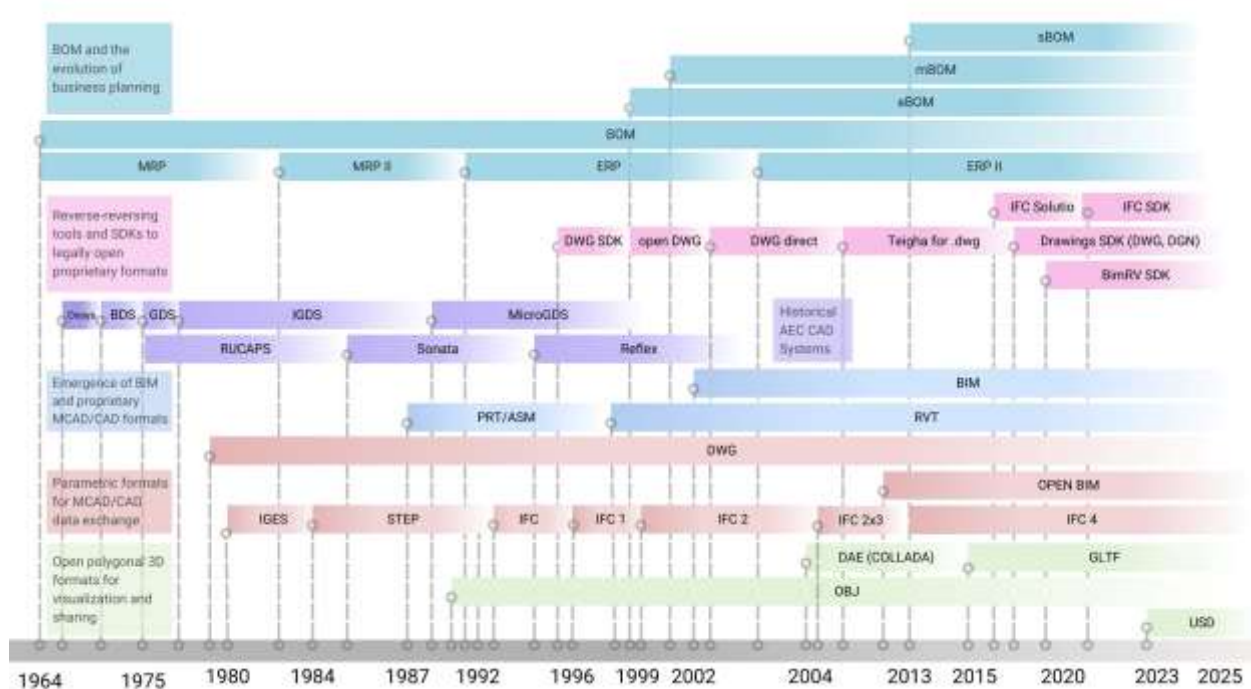
Et mõista, miks on nii palju erinevaid andmeformaate ja miks enamik neist on suletud, on oluline süveneda CAD (BIM) programmides toimuvatesse protsessidesse, mida käsitletakse üksikasjalikult raamatu kuuendas osas.

CAD-süsteemide arendajad töid geomeetria täiendava teabekihi sisse BIM-kontseptsiooniga (Building Information Modeling), mis on ehitussektoris alates 2002. aastast aktiivselt propageeritud turundustermin [54].

BIM (BOM) kontseptsiooni tekkimine ja CAD kasutamine protsessides.

Ehitusinformatsiooni modelleerimise (BIM) mõiste, mis esitati esmakordselt 2002. aasta BIM Whitepaperis [54], sai alguse CAD-tarkvara tootjate turundusalgatustest. See tekkis CAD-tarkvara arendajate turundusalgatustest ja oli katse kohandada masinaehituses juba väljakujunenud põhimõtteid ehitussektori vajadustele.

BIM sai inspiratsiooni BOM (Bill of Materials) - Bill of Materials - kontseptsioonist, mida on tööstuses laialdaselt kasutatud alates 1980. aastate lõpust. Masinaehituses võimaldas BOM siduda CAD-süsteemide andmed PDM (Product Data Management), PLM (Product Lifecycle Management) ja ERP-süsteemidega, võimaldades inseneriteabe terviklikku haldamist kogu toote elutsükli jooksul (joonis 3.1-8).



Joonis 3.1-18 Spetsifikatsioonide (BOM), infomodelleerimise (BIM) ja digitaalformaate areng inseneriehitustööstuses.

BOM-kontseptsiooni kaasaegne areng on viinud laiendatud raamistiku - XBOM (Extended BOM) - tekkimiseni, mis hõlmab lisaks toote koostisele ka käitumisstsenaariume, käitamisnõudeid, jätkusuutlikkuse parameetreid ja andmeid prognoositavaks analüüsiks. XBOM täidab sisuliselt sama rolli kui BIM ehituses: mõlema lähenemisviisi eesmärk on muuta digitaalne mudel ühtseks tööallikaks kõigi projektiosaliste jaoks kogu projekti elutsükli jooksul.

Oluliseks verstapostiks BOMi tekkimisel ehituses oli esimese spetsiaalselt ehitussektorile kohandatud parameetrilise CADi (MCAD) kasutuselevõtt 2002. aastal. Selle töötas välja meeskond, kes oli varem loonud Pro-E®, mis oli 1980ndate lõpus tekkinud revolutsiooniline MCAD-süsteem masinaehituse jaoks ja millest oli saanud tööstusstandard [55].

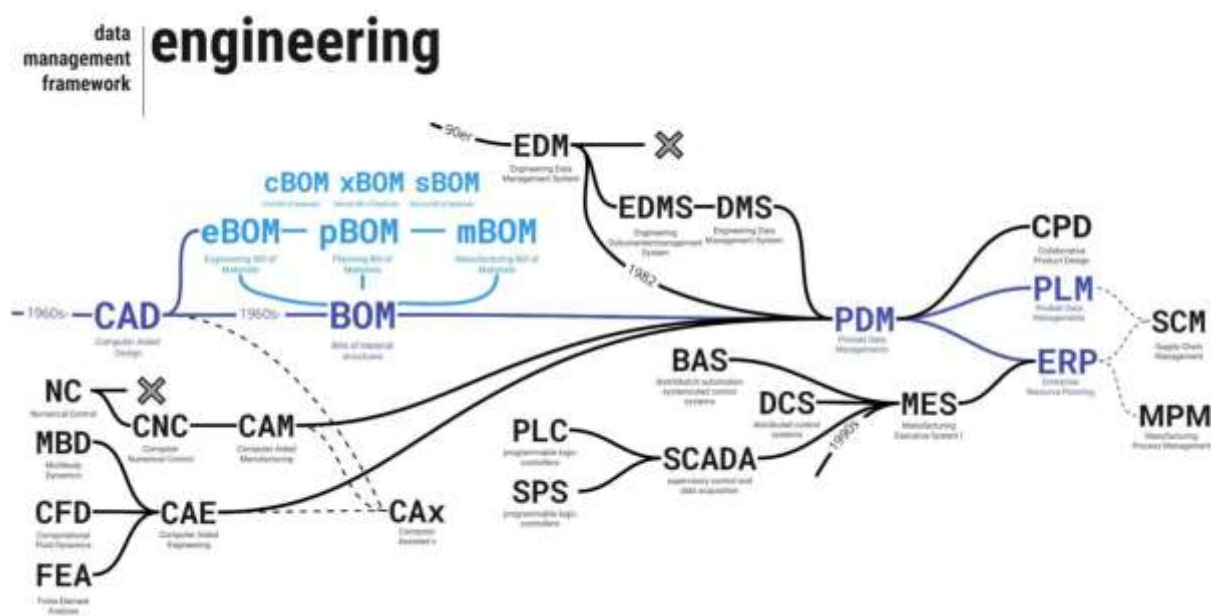
Juba 1980ndate lõpus püüti kõrvaldada tollaste CAD-programmide [56] piirangud -programmide piirangud. Peamine eesmärk oli vähendada töömahtu, mis on vajalik konstruktsioonelementide parameetrite muutmiseks, ning võimaldada mudeli uuendamist CAD-programmidest väljapoole jäävate andmete põhjal andmebaasi kaudu [57]. Kõige olulisemat rolli pidi selles mängima parameetriseerimine: omaduste automaatne hankimine andmebaasist ja nende kasutamine mudeli uuendamiseks CAD-süsteemides.

Pro-E ja selle aluseks olev elementaarse parameetrilise modelleerimise c BOM kontseptsioon on avaldanud märkimisväärset mõju CAD - ja MCAD - turu arengule [58]. See mudel on olnud tööstuses kasutusel juba 25 aastat ja paljud kaasaegsed süsteemid on saanud selle kontseptuaalseteks järeltulijateks.

Eesmärk on luua süsteem, mis on piisavalt paindlik, et julgustada inseneri kergesti kaaluma erinevaid konstruktsioone. Ja konstruktsiooni muutmise kulud peaksid olema võimalikult nullilähedased. Traditsiooniline CAD / CAM tarkvara piirab ebareaalselt odavate muudatuste tegemist ainult projekteerimisprotsessi alguses [59].

- Samuel Heisenberg, Parametric Technology Corporationi® asutaja, MCAD -toote Pro-E arendaja ja RVT-vormingut kasutava CAD-toote looja õpetaja

Masinaehituses on PDM-, PLM-, MRP- ja ERP-süsteemid muutunud peamisteks platvormideks. Nad mängivad kesket rolli andmete ja protsesside haldamisel, kogudes teavet CAX-süsteemidest (CAD, CAM, CAE) ja korraldades projekteerimistegevust toote struktuuri alusel (BOM: eBOM, pBOM, mBOM) (joonis 3.1-18). Selline integreerimine vähendab vigu, väldib andmete dubleerimist ja tagab läbiva jälgitavuse projekteerimisest kuni tootmiseni.



Joonis 3.1-19 Ajalooliselt tekkis BOM 1960. aastatel kui viis, kuidas struktureerida CAX-süsteemide andmeid ja edastada neid juhtimissüsteemidele.

Endise Pro-E meeskonna poolt välja töötatud ja BOM-lähenemisel põhineva CAD-lahenduse ostmist ühe juhtiva tootja poolt tähistas peaaegu kohene BIM Whitepaper'i seeria (2002-2003)[60][61] avaldamine. Juba alates 2000. aastate keskpaigast hakati ehitussektoris aktiivselt propageerima BIM-kontseptsiooni, mis suurendas märgatavalt huvi parameetrilise tarkvara vastu. Populaarsus kasvas nii kiiresti, et masinaehituse Pro-E - selle müüja poolt propageeritud parameetriline CAD - ehitusharu on tegelikult tõrjunud välja konkurendid arhitektuuri- ja ehituskonstruktsioonide projekteerimise segmendis (joonis 3.1-20). 2020. aastate alguseks on see de facto kindlustanud ülemaailmse domineerimise BIM (CAD) turul [62].



Joonis 3.1-20 Google'i otsingu populaarsus (RVT versus IFC): endise Pro-E meeskonna poolt loodud parameetriline CAD koos BOM toega -BIM on saavutanud populaarsuse peaaegu enamikus maailma riikides.

Viimase 20 aasta jooksul on lühend BIM omandanud mitmeid tõlgendusi, mille mitmetähenduslikkus tuleneb 2000. aastate alguses tekkinud algsetest turunduskontseptsioonidest. ISO 19650 standard, mis mängis olulist rolli termini populariseerimisel, kindlustas tegelikult BIMi kui "teaduspõhise" lähenemisviisi teabehaldusele. Standardi enda tekstis, mis on pühendatud andmete haldamisele objektide kogu elutsükli jooksul BIMi abil, mainitakse siiski lühendit BIM, kuid seda ei ole kunagi selgelt määratletud

Müüja algne veebisait, mis avaldas 2002. aastal[60] ja 2003. aastal[61] rea Whitepapers BIMi kohta, reprodutseeris tegelikult turundusmaterjale BOM (Bills of Materials) ja PLM (Product Lifecycle Management) kontseptsioonide kohta, mida varem, 1990. aastatel, kasutati Pro-E masinaehitustarkvaras [63].

Ehitusinformatsiooni modelleerimine, mis on uuenduslik uus lähenemisviis hoonete projekteerimisele, ehitamisele ja haldamisele, mille võttis 2002. aastal kasutusele..... [CAD-tootja nimi], on muutnud kogu maailma tööstuse spetsialistide mõtteviisi selle kohta, kuidas tehnoloogiat saab rakendada hoonete projekteerimisel, ehitamisel ja haldamisel.

- BIM Whitepaper, 2003 [61].

Need varajased väljaanded seostasid BIMi otseselt tsentraliseeritud integreeritud andmebaasi kontseptsiooniga. Nagu on märgitud 2003. aasta Whitepaperis, on BIM ehitusteabe haldamine, mille puhul kõik uuendused toimuvad ühtses hoidlas, mis tagab, et kõik joonised, lõiked ja spetsifikatsioonid (BOM - Bills of Materials) on sünkroonitud.

BIM on kirjeldatud kui ehitusinfo haldamine, kus kõik uuendused ja muudatused toimuvad andmebaasis. Nii et olenemata sellest, kas tegemist on skeemide, lõikude või lehtjoonistega, on kõik alati kooskõlastatud, järjepidev ja ajakohane.

- CAD-ettevõtte veebileht vendor with BIM Whitepaper, 2003 [54].

Idee juhtida projekteerimist ühtse integreeritud andmebaasi kaudu on laialdaselt arutatud juba 1980. aastate uuringutes. Näiteks Charles Eastmani BDS-kontseptsioonis [57] oli 43 viidet mõistele "andmebaas" (joonis 6.1-2). Aastaks 2004 oli see arv peaaegu poole võrra vähenenud 23-le 2002. aasta Whitepaperis BIMi kohta [64]. Ja 2000. aastate keskpaigaks oli andmebaaside teema praktiliselt kadunud müüjate turundusmaterjalidest ja digitaliseerimise tegevuskavast üldiselt.

Kuigi algselt oli BIM-süsteemi keskmeks mõeldud andmebaas ja juurdepääs sellele, on aja jooksul rõhuasetus nihkunud geomeetria, visualiseerimisele ja 3D-le. IFC-standardi registreerija 1994. aastal, kes avaldas 2002. aastal BIM Whitepaper'i - sama müüja - 2000. aastate alguse Whitepaper'is juhtis selgesõnaliselt tähelepanu neutraalsete formaatide, nagu IGES, STEP ja IFC, piirangutele ja vajadusele otsese juurdepääsu järel CAD-andmebaasidele:

Erinevad rakendused võivad olla vastuolus ja uuesti sisestatud andmed võivad olla ebatäpsed [...]. Traditsioonilise arvutipõhise projekteerimise [CAD] tulemus: suuremad kulud, pikem turuleviimise aeg ja madalam tootekvaliteet. Tänapäeval kasutavad kõik suuremad rakendused madala taseme andmevahetuseks tööstuse standardliideseid. Kasutades vanu IGES standardeid või uut STEPi [IFC on de facto ja de jure STEP/IGESi formaadi koopia], et vahetada andmeid eri tootjate rakenduste vahel, saavad kasutajad saavutada teatava andmete ühilduvuse parimate toodete vahel. Kuid IGES ja STEP töötavad ainult madalatel tasemetel ja nad ei saa vahetada nii rikkalikke andmeid kui tänapäeva juhtivate rakenduste poolt genereeritud teave [...]. Ja kuigi need ja teised standardid paranevad peaaegu iga päev, jäävad nad alati maha tänaste tootjate toodetest andmete rikkalikkuse poolest. [...] rakendusesisesed programmid peavad olema võimelised vahetama ja säilitama andmete rikkalikkust ilma neutraalsete tõlkijate, nagu IGES, STEP [IFC] või PATRAN, kasutamata. Selle asemel peaksid raamrakendused olema võimelised pääsema otse ligi aluseks olevale CAD-andmebaasile, nii et teabe detailsus ja täpsus ei läheks kaduma.

- CAD-tootjate valge raamat (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale," 2000 [65].

Seega peeti juba 1980ndatel ja 2000ndate alguses CAD-keskkonnas digitaalse projekteerimise põhiel-
emendiks pigem andmebaasi kui formaatfaili või neutraalset IFC-vormingut. Soovitati loobuda tõlkijat-
est ja rakendustele otsest juurdepääsu andmetele. Tegelikult hakkas BIM kontseptsioon aga 2020.
aastate keskpaigaks meenutama "jaga ja valitse" strateegiat, kus suletud geomeetrilisi tuumasid ka-
sutavate tarkvaratootjate huvid on eelistatud avatud infovahetuse arendamise ees.

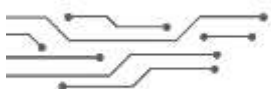
Tänapäeval peetakse BIMi ehitussektori lahutamatuks osaks. Kuid viimase kahe aastakümne jooksul on lubadused lihtsustatud koostööst ja andmete integreerimisest jäänud suures osas realiseerimata. Enamik lahendusi on endiselt seotud suletud või neutraalsete vormingute ja spetsialiseeritud tööriistadega. Vaatleme üksikasjalikult BIMi ajalugu, avatud BIMi ja IFCd ning koostalitlusvõime ja geomeetriliste tuumade küsimusi raamatu "CAD ja BIM: turundus, tegelikkus ja projekteerimisandmete tulevik ehituses" 6. osas.

Tänapäeval seisab tööstus silmitsi peamise väljakutsega, milleks on üleminek CAD (BIM) traditsioonilisest arusaamast kui modelleerimisvahendist selle kasutamisele täieõigusliku andmebaasina. See nõuab uusi lähenemisviise teabega töötamisele, loobumist suletud ökosüsteemide sõltuvusest ja avatud lahenduste rakendamist.

Tänu CAD-andmebaasidele juurdepääsu võimaldavate reverse engineering vahendite arengule ning avatud lähtekoodiga ja LLM-tehnoloogiate levikule eemalduvad ehitussektori kasutajad ja arendajad üha enam tarkvaratootjate ebaselgetest tingimustest. Selle asemel nihkub tähelepanu sellele, mis on tegelikult oluline: andmed (andmebaasid) ja protsessid.

Trendikate akronüümide ja visualiseerimiste taga on standardsed andmehalduse tavad: salvestamine, ülekandmine ja teisendamine - st klassikaline ETL-protsess (Extract, Transform, Load). Nagu teistes tööstusharudes, nõuab ka ehitussektori digitaliseerimine mitte ainult andmevahetusstandardeid, vaid ka heterogeense teabe selgelt struktureeritud käsitlemist.

CAD-andmete (BIM) potentsiaali täielikuks ärakasutamiseks peavad ettevõtted oma lähenemisviisi teabehaldusele ümber mõtlema. See toob paratamatult kaasa digitaalse ümberkujundamise võtmelemendi - ehitusspetsialistide igapäevaselt kasutatavate andmete ühtlustamise, standardiseerimise ja mõtestatud struktureerimise.

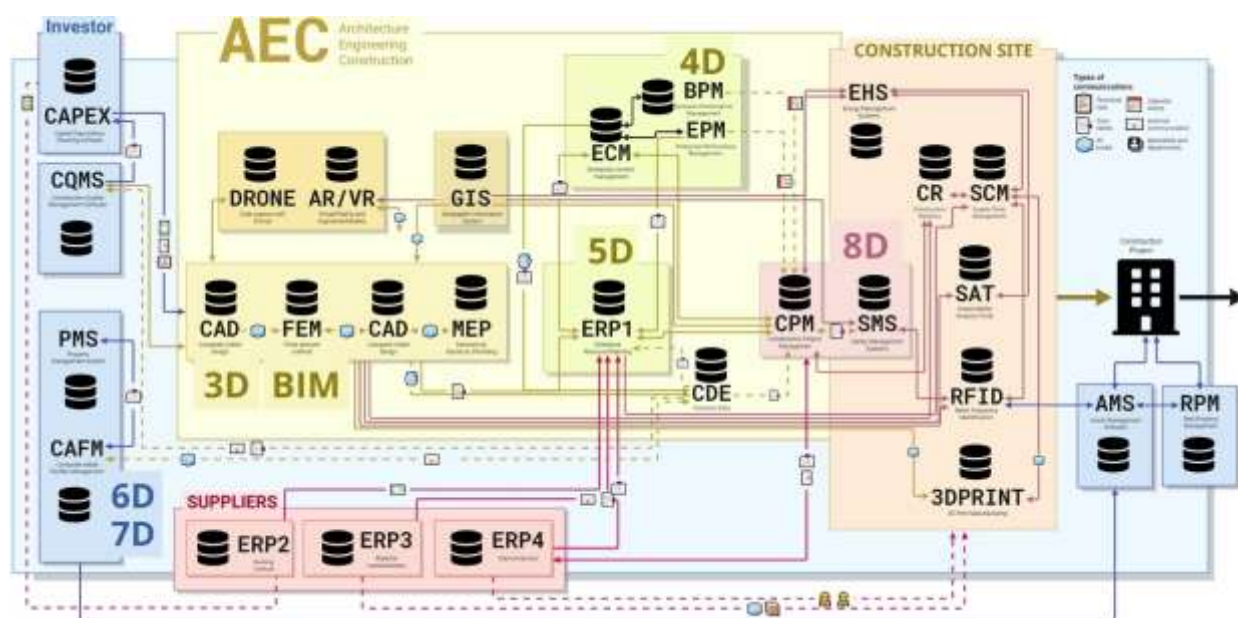


PEATÜKK 3.2.

ANDMETE ÜHTLUSTAMINE JA STRUKTUREERIMINE

Süsteemide täitmine andmetega ehitussektoris

Olgu tegemist suurte korporatsioonide või keskmise suurusega ettevõtetega, spetsialistid tegelevad igapäevaselt erinevate liidestega tarkvarasüsteemide ja andmebaaside täitmisega mitmeformaadilise teabega (joonis 3.2-1), mis peavad juhtide abiga omavahel koostööd tegema. Just see koostoimivate süsteemide ja protsesside kompleks loob lõppkokkuvõttes ettevõttele tulu ja kasumi.



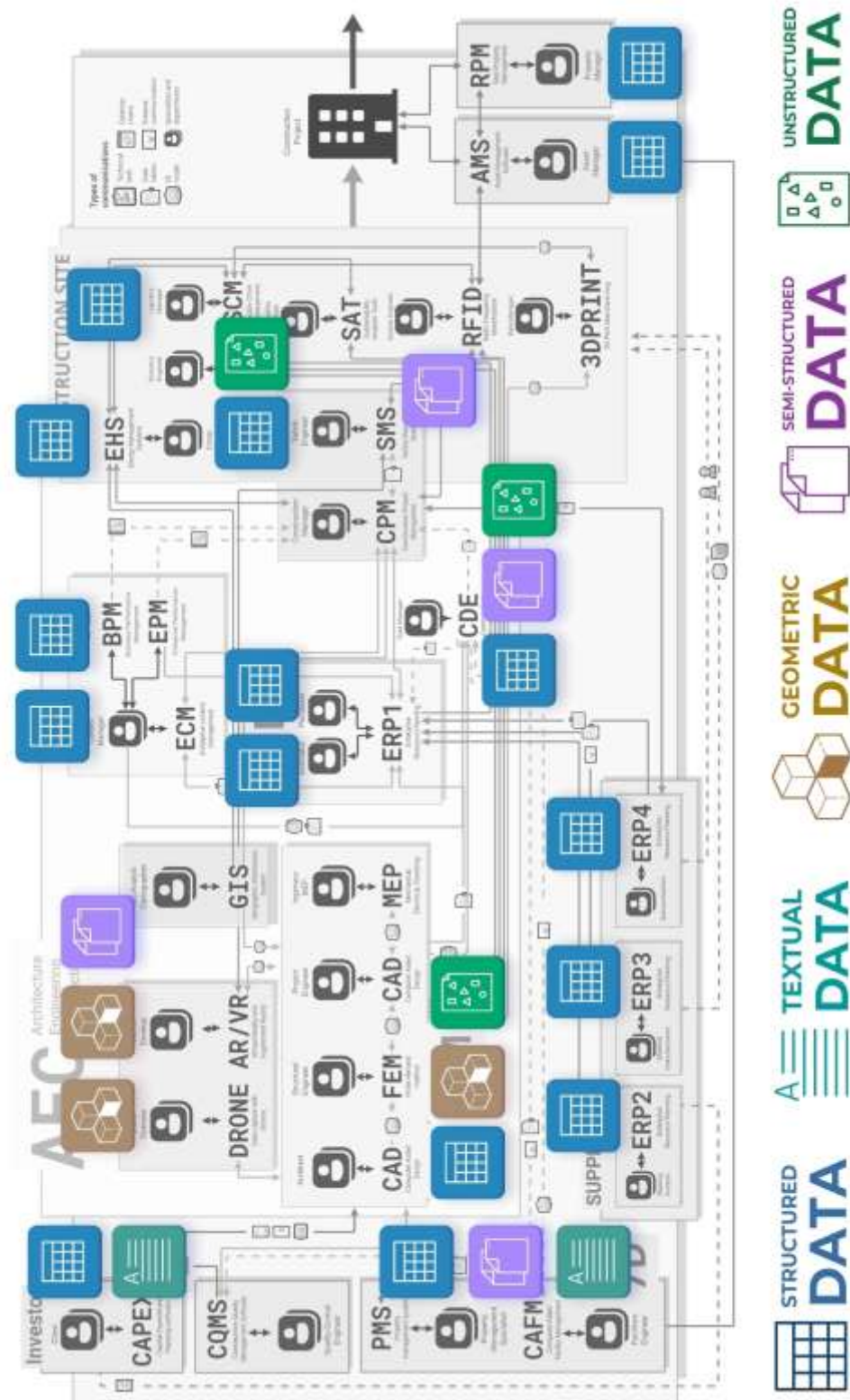
Joonis 3.2-1 Praktiliselt iga ehitusvaldkonnas kasutatav süsteem või rakendus sisaldab ühte populaarset RDBMS andmebaasi.

Iga eespool nimetatud ja ehitussektoris rakendatav süsteemikategooria töötab oma andmetüüpidega, mis vastavad nende süsteemide funktsionaalsele rollile. Et liikuda abstraktselt tasandilt konkreetsele tasandile, liigume andmetüüpidelt nende esitamisele vormingute ja dokumentidena.

Varem esitatud süsteemide loetelule (joonis 1.2-4) lisame nüüd konkreetsed formaadid ja dokumendiliigid, millega nad sageli töötavad:

■ Investor (CAPEX)

- Finantsandmed: eelarved, kuluproгноosid (struktureeritud andmed).
- Turutrendide andmed: turuanalüüsid (struktureeritud ja struktureerimata andmed).
- Õiguslikud ja lepingulised andmed: lepingud (tekstiandmed).



Joonis 3.2-2 Ehitustööstuses kasutatakse palju eri liidestega süsteeme, mis tegelevad eri liiki andmetega.

■ **Juhtimissüsteemid (PMS, CAFM, CQMS)**

- Projekti andmed: graafikud, ülesanded (struktureeritud andmed).
- Rajatiste hooldusandmed: hooldusplaanid (teksti- ja poolstruktureeritud andmed).
- Kvaliteedikontrolli andmed: standardid, inspekteerimisaruanded (tekstilised ja struktureerimata andmed).

■ **CAD, FEM ja BIM**

- Tehnilised joonised: arhitektuursed, struktuursed plaanid (geomeetrilised andmed, struktureerimata andmed).
- Ehitusmodelid: 3D -modelid, materjaliandmed (geomeetrilised ja poolstruktureeritud andmed).
- Tehnilised arvutused: koormusanalüüs (struktureeritud andmed).

■ **Ehitusplatsi juhtimissüsteemid (EHS, SCM)**

- Ohutus- ja terviseandmed: ohutusprotokollid (tekstilised ja struktureeritud andmed).
- Tarneahela andmed: varud, tellimused (struktureeritud andmed).
- Igapäevased aruanded: tööaeg, tootlikkus (struktureeritud andmed).

■ **Droonid, AR/VR, GIS, 3D -trükkimine**

- Geoandmed: topograafilised kaardid (geomeetrilised ja struktureeritud andmed).
- Reaalajas andmed: video ja fotod (struktureerimata andmed).
- Modelid 3D -trükkimiseks: digitaalsed joonised (geomeetrilised andmed).

■ **Täiendavad juhtimissüsteemid (4D BPM, 5D ERP1)**

- Aja- ja kuluandmed: ajakavad, kalkulatsioonid (struktureeritud andmed).
- Muudatuste haldamine: projekti muudatuste kirjed (tekst ja struktureeritud andmed).
- Tulemuslikkuse aruandlus: edukuse näitajad (struktureeritud andmed).

■ **Andmete integreerimine ja side (CDE, RFID, AMS, RPM)**

- Andmevahetus: dokumendivahetus, andmemudelid (struktureeritud ja tekstilised andmed).
- RFID ja jälgimisandmed: logistika, varahaldus (struktureeritud andmed).
- Järelevalve ja kontroll: andurid (struktureeritud ja struktureerimata andmed).

Seega töötab iga ehitussektori süsteem - alates ehitusplatsi juhtimissüsteemidest kuni operatiivandmebaasideni - oma tüüpi teabega: struktureeritud, tekstiline, geomeetiline ja muu. "Andmemaastik", millega spetsialistid peavad igapäevaselt töötama, on äärmiselt mitmekesine. Lihtne formaatide ülesloendamine ei näita siiski tegeliku teabega töötamise keerukust.

Praktikas seisavad ettevõtted silmitsi asjaoluga, et isegi süsteemidest saadud andmed ei ole valmis kasutamiseks "sellisel kujul". See kehtib eelkõige tekstide, piltide, PDF-failide, CAD-failide ja muude formaatide puhul, mida on raske analüüsida standardvahenditega. Seepärast on järgmine oluline samm andmete muundamine - protsess, ilma milleta ei saa töötlemist, analüüsi, visualiseerimist ja otsuste tegemist tõhusalt automatiseerida.

Andmete ümberkujundamine: kaasaegse ärianalüüsi kriitiline alus

Tänapäeval seisavad enamik ettevõtteid silmitsi paradoksiga: umbes 80% nende igapäevastest protsessidest tuginevad endiselt klassikalistele struktureeritud andmetele - tuttavatele Exceli tabelitele ja relatsioonilistele andmebaasidele (RDBMS) [66]. Samal ajal on aga 80% ettevõtete digitaalsesse ökosüsteemi sisenevast uuest teabest struktureerimata või lahtiselt struktureeritud (joonis 3.2-3) [52]. See hõlmab teksti, graafikat, geomeetriat, pilte, CAD -mudeleid, PDF-dokumentatsiooni, heli- ja videosalvestusi, elektroonilist kirjavahetust ja palju muud.

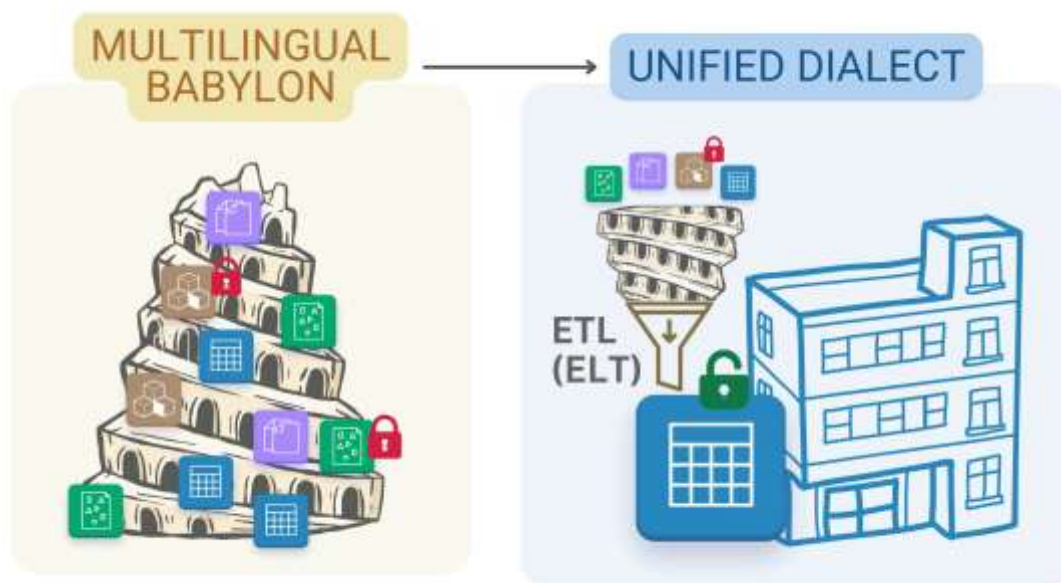
Lisaks sellele kasvab struktureerimata andmete maht jätkuvalt kiiresti - hinnanguliselt kasvab see igal aastal 55-65% [67]. Selline dünaamika tekitab tõsiseid raskusi uue teabe integreerimisel olemasolevatesse äriprotsessidesse. Selle multiformaatse andmevoo eiramine viib infolünkade tekkimiseni ja vähendab kogu ettevõtte digitaalse keskkonna juhitavust.



Joonis 3.2-3 Struktureerimata andmete iga-aastane kasv tekitab probleeme voogesituse integreerimisel äriprotsessidesse.

Keerukate struktureerimata ja segadust tekitava lahtise struktuuriga andmete eiramine automatiseerimisprotsessides võib põhjustada märkimisväärsed lünki ettevõtte teabemaastikul. Tänapäeva kontrollimatu ja laviinitaolise infoliikumise maailmas peavad ettevõtted võtma kasutusele hübriidse lähenemisviisi andmehaldusele, mis hõlmab tõhusaid meetodeid igat liiki andmete käsitlemiseks.

Tõhusa andmehalduse võti seisneb organiseerimises, struktureerimises ja liigitamises erinevat tüüpi andmete "Paabel" (sealhulgas struktureerimata, tekstilised ja geomeetrilised vormingud, struktureeritud või lahtiselt struktureeritud andmed). See protsess muudab kaootilised andmekogumid süsteemidesse integreeritavateks organiseeritud struktuurideks, võimaldades seeläbi nende põhjal otsuste langetamist (joonis 3.2-4).



Joonis 3.2-4 Andmehaldusosakondade peamine ülesanne on muuta mitmekesiste ja eri formaadis andmete "Babülon" struktureeritud ja kategoriseeritud süsteemiks.

Üks peamisi takistusi sellisel ühtlustamisel on endiselt erinevate digitaalsete platvormide vähene koostalitlusvõime - eelmistes peatükkides käsitletud "silod".

Aruande kohaselt rõhutab riiklik standardite ja tehnoloogiainstituut (NIST, USA) [68], et andmete vähene ühilduvus erinevate ehitusplatvormide vahel toob kaasa teabe kadumise ja märkimisväärsed lisakulud. Ainuüksi 2002. aastal põhjustasid tarkvara koostalitlusvõime probleemid USA kapitaliehituses aastas kokku 15,8 miljardit dollarit kahju, millest kaks kolmandikku kannavad hoonete omanikud ja operaatorid, eriti käitamise ja hoolduse ajal [68]. Uuringus märgitakse ka, et andmeformaadide standardiseerimine võib vähendada neid kahjusid ja parandada tõhusust kogu rajatise elutsükli jooksul.

2016. aasta CrowdFloweri uuringu [69] kohaselt, mis hõlmas 16 000 andmeteadlast üle maailma, on peamine probleem endiselt "määrdund" ja mitmeformaadilised andmed. Selle uuringu kohaselt ei ole kõige väärtuslikum ressurss mitte lõplikud andmebaasid või masinõppe mudelid, vaid teabe ettevalmistamisele kuluv aeg.

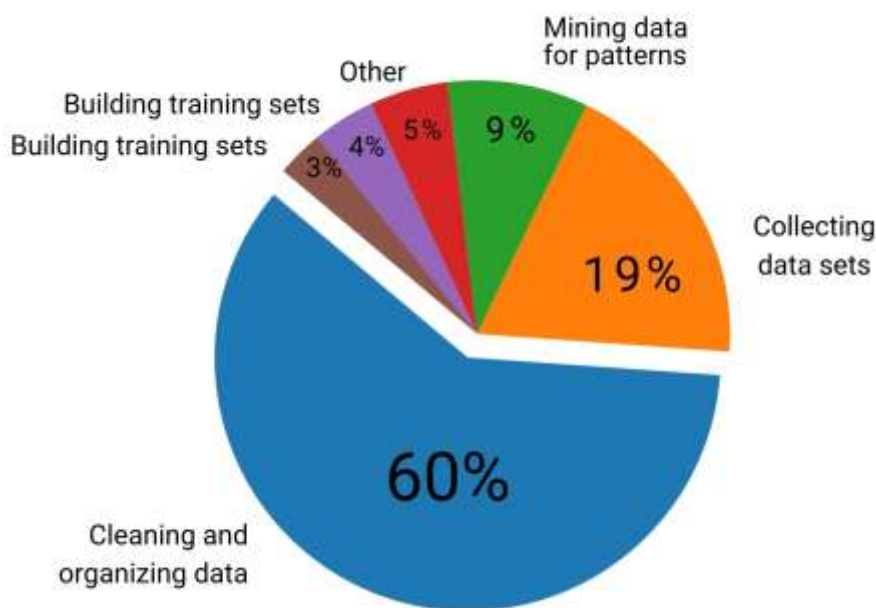
Puhastamine, vormindamine ja korrastamine võtab kuni 60 protsenti analüütiku ja andmehalduri ajast. Peaaegu viiendik kulub õigete andmekogumite otsimisele ja kogumisele, mis on sageli peidetud silodesse ja analüüsiks kättesaamatuks. Ja ainult umbes 9 protsenti ajast kulub otseselt modelleerimisele, analüüsile, prognooside tegemisele ja hüpoteeside testimisele. Ülejäänud aeg kulub teabevahetusele, visualiseerimisele, aruandlusele ja toetavate teabeallikate uurimisele

Keskmiselt jaguneb juhi andmetöö järgmiselt (joonis 3.2-5):

- **Andmete puhastamine ja korrastamine (60%):** puhaste ja struktureeritud andmete olemasolu võib oluliselt vähendada analüütiku tööaega ja kiirendada ülesannete täitmist.
- **Andmete kogumine (19%):** Andmeteaduse spetsialistide jaoks on oluline väljakutse

asjakohaste andmekogumite leidmine. Sageli on ettevõtte andmed kuhjatud kaootiliselt organiseeritud "silodesse", mis raskendab juurdepääsu vajalikule teabele.

- **Modelleerimine / masinõpe (9%):** sageli takistab seda klientide ebaselgus ärieesmärkide osas. Selge ülesandepüstituse puudumine võib muuta isegi parima mudeli potentsiaali olematuks.
- **Muud ülesanded (5%):** lisaks andmete töötlemisele peavad analüütikud tegelema ka uurimistööga, uurima andmeid eri vaatenurkadest, edastama tulemusi visualiseerimiste ja aruannete kaudu ning soovutama protsesside ja strateegiate optimeerimist.



Joonis 3.2-5 Millega tegelevad andmehaldurid kõige rohkem aega kulutavad (põhineb [70]).

Neid hinnanguid toetavad teised uuringud. 2015. aastal ajakirjas BizReport avaldatud Xplenty uuringu [71] kohaselt kulub 50-90% äriteabe (BI) spetsialistide (BI) ajast andmete ettevalmistamisele analüüsiks.

Andmete puhastamine, valideerimine ja korrastamine on kõigi järgnevate andme- ja analüüsiprotsesside kriitiline alus, mis võtab kuni 90% andmeteadlaste ajast.

See lõppkasutajale nähtamatu vaevarikas töö on ülioluline. Vead toorandmetes moonutavad paratamatult analüüsi, on eksitavad ja võivad põhjustada kulukaid juhtimisvigu. Seepärast on andmete puhastamise ja standardimise protsessid - alates dubleeringute kõrvaldamisest ja puuduste täitmisest kuni mõõtühikute ühtlustamise ja ühisele mudelile kohandamiseni - muutumas kaasaegse digitaalstrateegia nurgakiviks.

Seega ei võta andmete põhjalik ümberkujundamine, puhastamine ja standardimine mitte ainult suurema osa spetsialistide ajast (kuni 80% andmetega töötamisest), vaid määrab ka nende tõhusa kasutamise võimaluse kaasaegsete äriprotsesside raames. Kuid andmete korrastamine ja puhastamine üksi ei ammenda ettevõtte infovoogude optimaalse haldamise ülesannet. Korraldamise ja

struktureerimise etapis muutub sobiv andmemudeli valik, mis mõjutab otseselt teabega töötamise mugavust ja tõhusust järgmistes töötlemisetappides.

Kuna andmed ja ärieesmärgid on erinevad, on oluline mõista andmemudelite omadusi ja osata valida või luua õige struktuur. Sõltuvalt struktureerimise astmest ja elementidevaheliste suhete kirjeldamise viisist on kolm peamist mudelit: struktureeritud, vabalt struktureeritud ja graafilised. Iga mudel sobib erinevate ülesannete jaoks ja neil on omad tugevused ja nõrkused.

Andmemudelid: seosed andmetes ja elementide vahelised seosed

Andmed on infosüsteemides korraldatud erinevalt - sõltuvalt teabe säilitamise, töötlemise ja edastamise ülesannetest ja nõuetest. Peamine erinevus andmemudelite tüüpide vahel, st teabe salvestamise vorm, on struktureerimise tase ja viis, kuidas kirjeldatakse elementidevahelisi seoseid.

Struktureeritud andmetel on selge ja korratav skeem: need on korraldatud kindlaid veerge sisaldavate tabelitena. Selline vorming tagab prognoositavuse, lihtsa töötlemise ja tõhususe SQL päringute tegemisel, filtreerimisel ja summeerimisel. Näited - andmebaasid (RDBMS), Excel, CSV.

Vabalt struktureeritud andmed võimaldavad paindlikku struktuuri: erinevad elemendid võivad sisaldada erinevaid atribuute ja neid võib salvestada hierarhiatena. Näited on JSON, XML või muud dokumendivormingud. Sellised andmed on mugavad, kui on vaja modelleerida üksteise sisseehitatud objekte ja nende vahelisi seoseid, kuid teisest küljest raskendab see andmete analüüsi ja standardimist (joonis 3.2-6).

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

Joonis 3.2-6 Andmemudel on loogiline struktuur, mis kirjeldab, kuidas andmed on süsteemis organiseeritud, salvestatud ja töödeldud.

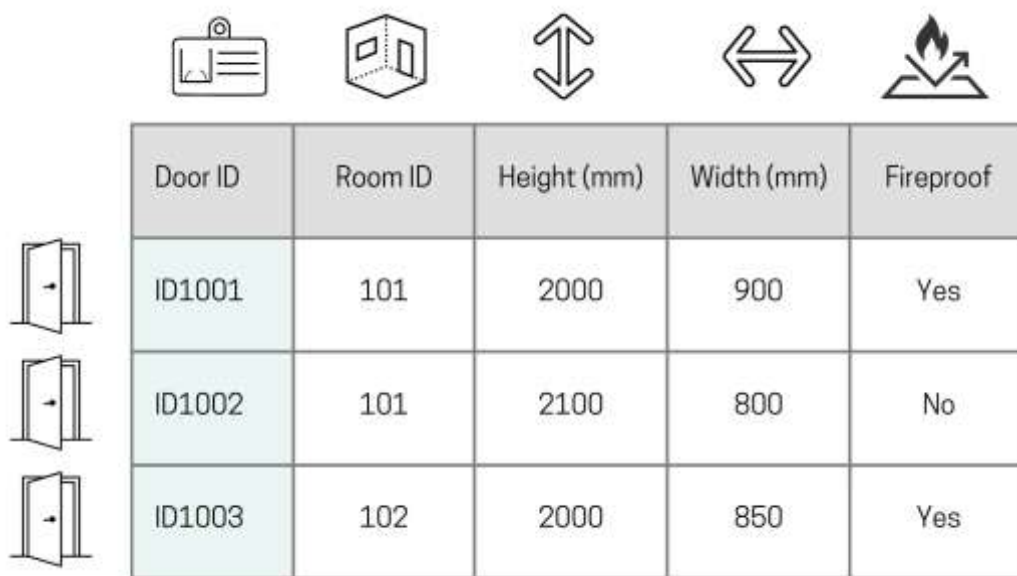
Sobiva formaadi valik sõltub eesmärkidest:

- Kui filtreerimise ja analüüsi kiirus on oluline - sobivad relatsioonilised tabelid (SQL, CSV, RDBMS, veergudega andmebaasid).
- Kui on vaja paindlikku struktuuri - on parem kasutada JSON või XML.

- Kui andmetel on keerulised seosed - graafiandmebaasid pakuvad nähtavust ja skaleeritavust.

Klassikalistes relatsioonilistes andmebaasides (RDBMS) esitatakse iga üksus (nt uks) reana ja selle omadused tabeli veergudena. Näiteks kategooria "Uksed" elementide tabel võib sisaldada väljad ID, kõrgus, laius, tulekindlus ja ruumi ID, mis näitab ruumi (joonis 3.2-7).

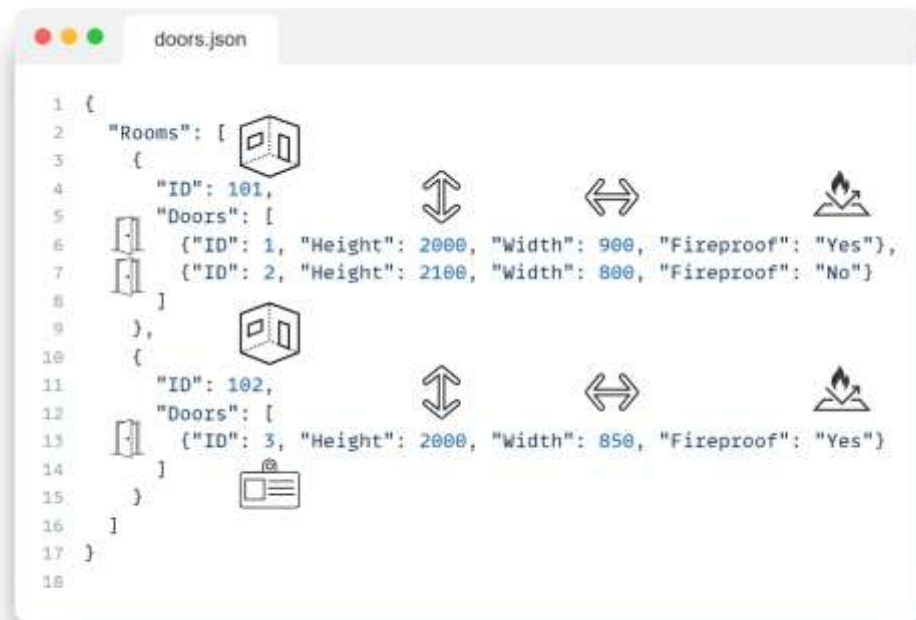
Klassikalistes relatsioonilistes andmebaasides (RDBMS) moodustatakse suhted tabelite kujul, kus iga kirje kujutab objekti ja veerud selle parameetreid. Tabeli kujul näevad andmed uste kohta projektis välja järgmiselt, kus iga rida esindab eraldi elementi - ust koos selle unikaalse identifikaatori ja atribuutidega ning seos ruumiga toimub parameetri "Room ID" kaudu.



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

Joonis 3.2-7 Andmed projekti kategooria "Uksed" kolme elemendi kohta tabeli kujul struktureeritud kujul.

Lahtiselt struktureeritud vormingutes, nagu JSON või XML, salvestatakse andmeid hierarhiliselt või nested kujul, kus elemendid võivad sisaldada teisi objekte ja nende struktuur võib erineda. See võimaldab modelleerida elementide vahelisi keerulisi seoseid. Sarnane teave projekti uste kohta, mis on salvestatud struktureeritud kujul (joonis 3.2-7), on esitatud vabalt struktureeritud vormingus (JSON) selliselt (joonis 3.2-8), et need muutuvad ruumide (Rooms - ID) sisemiste objektidena, mis loogiliselt peegeldab hierarhiat.



Joonis 3.2-8 Teave projekti kategooria "Uksed" elementide kohta JSON-vormingus.

Graafimudelis esitatakse andmed sõlmedena (tipud) ja nendevaheliste sidemetena (servad). See võimaldab visualiseerida objektide ja nende atribuutide vahelisi keerulisi seoseid. Projekti uste ja ruumide andmete puhul on graafi kujutis järgmine:

- **Sõlmed (nodes)** esindavad peamisi üksusi: ruumid (Room 101, Room 102) ja uksed (ID1001, ID1002, ID1003).
- **Ribid (lingid)** näitavad nende üksuste vahelisi seoseid, nt ukse kuulumine teatud ruumi.
- **Atribuudid** kaardistatakse sõlmedele ja sisaldavad üksuse omadusi (kõrgus, laius, uste tulekindlus).



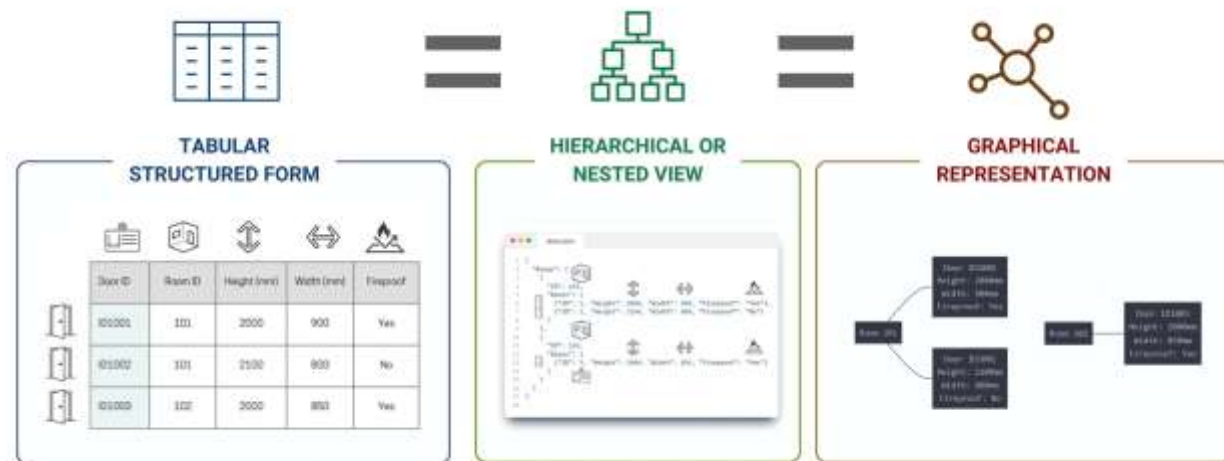
Joonis 3.2-9 Projekti ukseüksuse teave graafilises vaates.

Uksekirjelduse graafi andmemudelil on iga ruum ja iga uks eraldi sõlmed. Uksed on ruumidega seotud

servade kaudu, mis näitavad, et uks kuulub konkreetsesse ruumi. Uste omadused (kõrgus, laius, tulekindlus) on salvestatud vastavate sõlmede omadustena. Lähemalt käsitletakse graaformaate ja graafi semantika tekkimist ehituses peatükis "semantika ja ontoloogia tekkimine ehituses".

Graafiandmebaasid on tõhusad siis, kui olulised ei ole mitte niivõrd andmed ise, vaid nende vahelised suhted, näiteks soovitusüsteemides, marsruutimise süsteemides või kui modelleeritakse keerulisi suhteid rajatiste haldamise projektides. Graafivorming lihtsustab uute seoste loomist, võimaldades uute andmetüüpide lisamist graafi ilma salvestusstruktuuri muutmata. Võrreldes relatsioonitabelite ja struktureeritud vormingutega ei ole graafis siiski täiendavat andmete ühendatavust - kahemõõtmeliste andmebaasiandmete ülekandmine graafi ei suurenda seoste arvu ega võimalda saada uut teavet.

Andmete vorm ja skeem tuleks kohandada vastavalt konkreetsele kasutusjuhule ja täidetavatele ülesannetele. Äriprotsessides tõhusaks tööks on oluline kasutada neid vahendeid ja andmemudeleid, mis aitavad võimalikult kiiresti ja lihtsalt tulemusi saada.



Joonis 3.2-10 Sama teavet projekti elementide kohta saab salvestada erinevates vormingutes, kasutades erinevaid andmemudeleid.

Tänapäeval seisab enamik suurettevõtteid silmitsi andmete liigse keerukuse probleemiga. Iga sadadest või tuhandetest rakendustest kasutab oma andmemudelit, mis tekitab liigset keerukust - üksik mudel on sageli kümneid kordi keerulisem kui vaja ja kõigi mudelite kogum on tuhandeid kordi keerulisem. Selline liigne keerukus raskendab oluliselt nii arendajate kui ka lõppkasutajate tööd.

Selline keerukus seab ettevõtte süsteemide arendamisele ja hooldamisele tõsiseid piiranguid. Iga uus element mudelis nõuab täiendavat koodi, uue loogika rakendamist, põhjalikku testimist ja olemasolevate lahenduste kohandamist. Kõik see suurendab kulusid ja aeglustab ettevõtte automatiseerimismeeskonna tööd, muutes isegi lihtsad ülesanded kulukaks ja aeganõudvaks protsessiks.

Keerukus mõjutab kõiki andmearhitektuuri tasandeid. Relatsioonilistes andmebaasides väljendub see tabelite ja veergude kasvavas arvus, mis on sageli üleliigsed. Objektipõhistes süsteemides suurendab keerukust klasside ja omavahel seotud omaduste paljusus. Sellistes vormingutes nagu XML või JSON väljendub keerukus segadust tekitavates nested structures, unikaalsete võtmete ja ebajärjekindlate skeemide kaudu.

Andmemudelite liigne keerukus muudab süsteemid mitte ainult vähem tõhusaks, vaid ka lõppkasutajate jaoks raskesti mõistetavaks ning tulevikus suured keelemudelid ja LLM-agendid. Just andmemudelite ja andmetöötamise mõistmise ja keerukuse probleem tõstatab küsimuse: kuidas muuta andmed piisavalt lihtsaks, et need hakkaksid tegelikult kiiresti kasulikud olema.

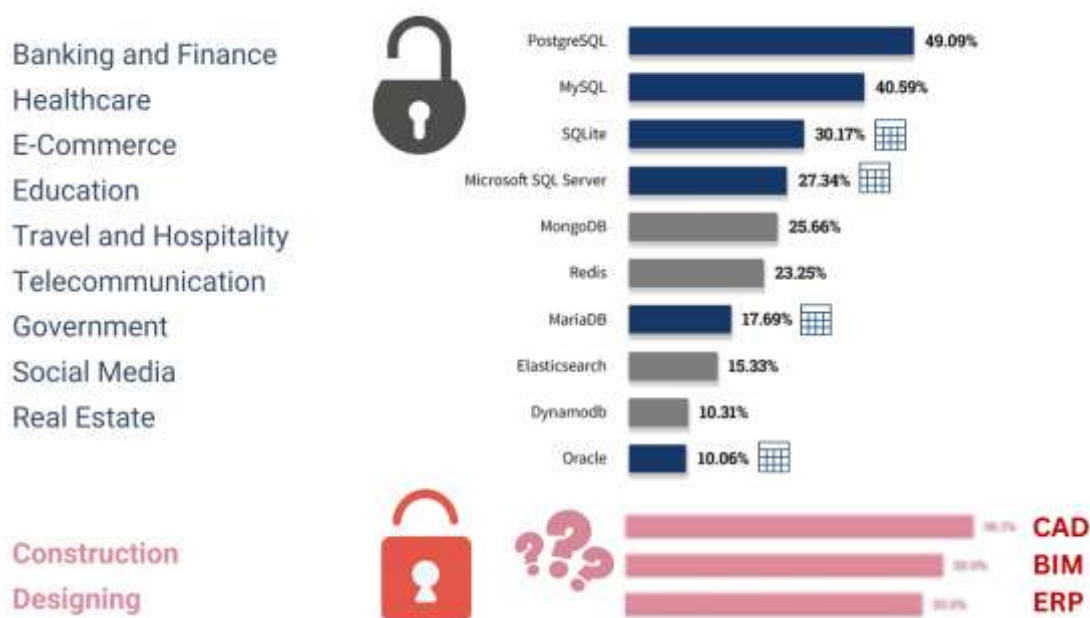
Isegi kui andmemudelid on targalt valitud, väheneb nende kasulikkus oluliselt, kui juurdepääs andmetele on piiratud. Omanikuvormingud ja suletud platvormid takistavad integratsiooni, raskendavad automatiseerimist ja võtavad ära kontrolli omandiõigusega seotud teabe üle, luues mitte ainult uute andmete silo, vaid lukustatud silo, millele on juurdepääs ainult müüja loal. Probleemi ulatuse mõistmiseks on oluline kaaluda täpselt, kuidas suletud süsteemid mõjutavad digitaalseid protsesse ehituses.

Omane formaat ja selle mõju digitaalsetele protsessidele

Üks peamisi väljakutseid, millega ehitusettevõtted digitaliseerimise käigus silmitsi seisavad, on piiratud juurdepääs andmetele. See raskendab süsteemide integreerimist, vähendab teabe kvaliteeti ja raskendab tõhusate protsesside korraldamist. Sageli on nende raskuste põhjuseks patenteeritud vormingute ja suletud tarkvaralahenduste kasutamine.

Kahjuks võimaldavad paljud ehitustööstuses kasutatavad programmid siiani kasutajal salvestada andmeid ainult konfidentsiaalsetes vormingutes või pilvemäludes, millele on võimalik ligi pääseda ainult rangelt piiratud liidestite kaudu. Samuti ei ole haruldane, et need lahendused on ehitatud sõltuvalt suuremate tarnijate veelgi suletumatest süsteemidest. Selle tulemusel on isegi need arendajad, kes sooviksid pakkuda avatumaid arhitektuure, sunnitud järgima suurte müüjate dikteeritud reegleid.

Kuigi kaasaegsed ehitusandmete haldamise süsteemid toetavad üha enam avatud formaate ja standardeid (joonis 3.1-5), on CAD- (BIM)-põhised andmebaasid ja nendega seotud ERP- ja CAFM-süsteemid endiselt eraldiseisvad, patenteeritud "saared" tööstuse digitaalsel maastikul (joonis 3.2-11).



Joonis 3.2-11 Andmete suletud ja kaitstud olemus loob tõkkeid andmete integreerimiseks ja juurdepääsuks.

Suletud ja monopoliseeritud formaadid ja protokollid ei ole probleemiks mitte ainult ehitussektoris. Paljudes majandussektorites algas võitlus suletud standardite ja piiratud juurdepääsuga andmetele innovatsiooni aeglustumisega (joonis 3.2-12), kunstlike tõkete olemasoluga uute osalejate sisenemisel ja süveneva sõltuvusega suurtest tarnijatest. Seoses andmete tähtsuse kiire kasvuga ei ole konkurentsiasutustel lihtsalt aega reageerida uute digitaalsete turgude väljakutsetele, mistõttu suletud formaadid ja piiratud juurdepääs andmetele muutuvad sisuliselt digitaalseteks "piirideks", mis piiravad teabevahetust ja majanduskasvu [63].

Kui masinad toodavad kõike, mida me vajame, siis sõltub meie olukord sellest, kuidas neid kaupu jaotatakse. Igaüks saab nautida jõukat elu ainult siis, kui masinate toodetud rikkust jagatakse. Või siis satub enamik inimesi elama äärmisesse vaesusse, kui autoomanikud saavad edukalt lobitööd teha rikkuse ümberjaotamise vastu. Siiani näib, et asjad lähevad teistmoodi, sest tehnoloogia viib üha suurema ebavõrdsuseni [72].

- Stephen Hawking, astrofüüsik, 2015

Monopolies or tight control over critical data formats

Telecommunications:
Proprietary Protocols

1970s-1980s

Computing Industry:
Open Source Movement

1980s

Document Formats:
PDFs and DOCs

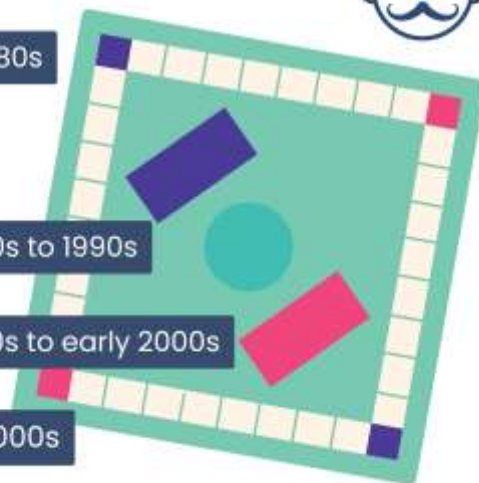
Late 1980s to 1990s

Web Browsing:
Browser Wars

Mid-1990s to early 2000s

Media:
Audio and Video Codecs

1990s-2000s



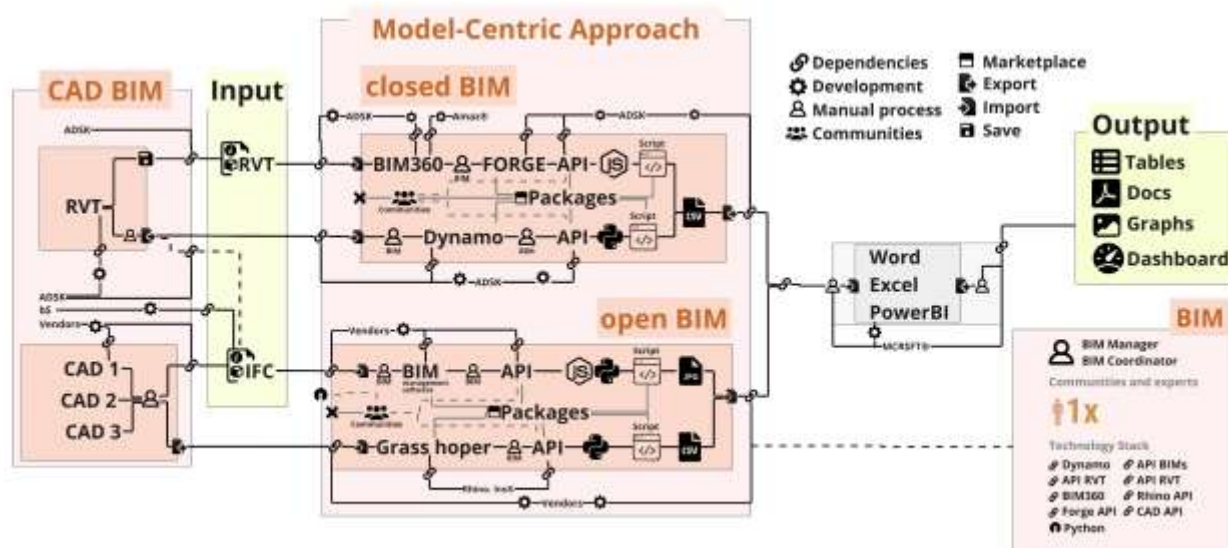
Joonis 3.2-12 Põhiliste andmevormingute ja -protokollide monopolne omandiõigus ei ole üksnes ehitussektori probleem.

Selle tulemusel on suletud juurdepääsu tõttu andmebaasidele programmid, andmehaldurid, andmeanalüütikud, IT-spetsialistid ja arendajad, kes loovad ehitussektori andmetele juurdepääsu, töötlemise ja automatiseerimise rakendusi, tänapäeval paljuski sõltuvuses tarkvaratootjatest (joonis 3.2-13). Need sõltuvused täiendavate juurdepääsukihtide näol nõuavad spetsiaalsete API -ühenduste ning spetsiaalsete tööriistade ja tarkvaraga lahenduste loomist.

API (Application Programming Interface) on ametlik liides, mille kaudu üks programm saab suhelda teise programmiga, vahetades andmeid ja funktsioone, ilma et oleks vaja juurdepääsu lähtekoodile. API kirjeldab, milliseid päringuid võib väline süsteem esitada, millises formaadis need peaksid olema ja milliseid vastuseid ta saab. See on standardiseeritud "leping" tarkvaramoodulite vahel.

Suur hulk sõltuvusi suletud lahendustest põhjustab selle, et kogu ettevõtte koodiarhitektuur ja äriprotsesside loogika muutub "spageti-arhitektuuriks", mille vahendid sõltuvad tarkvaratootja poliitikast, mis tagab kvaliteetse juurdepääsu andmetele.

Sõltuvus suletud lahendustest ja platvormidest toob kaasa mitte ainult paindlikkuse kaotuse, vaid ka tõelised äririskid. Litsentsitingimuste muutmine, andmetele juurdepääsu sulgemine, formaatide või API-struktuuri muutmine - kõik need võivad blokeerida kriitilisi protsesse. Äkki selgub, et ühe tabeli uuendamine nõuab terve integratsioonide ja ühenduste ploki ümbertöötamist (joonis 3.2-13) ning iga suuremahuline tarkvara või selle API tarnija uuendamine muutub potentsiaalseks ohuks kogu ettevõtte süsteemi stabiilsusele.



Joonis 3.2-13 Näide suurest arvust sõltuvustest CAD-töötlusel -andmed tekitavad ehitusettevõtete ökosüsteemis takistusi andmete integreerimisel.

Arendajad ja süsteemiarhitektid on sellistes tingimustes sunnitud töötama mitte ootuse, vaid ellujäämise nimel. Uute lahenduste rakendamise asemel kohanevad nad. Arendamise asemel püüavad nad säilitada ühilduvust. Protsesside automatiseerimise ja kiirendamise asemel kulutavad nad oma aega järgmiste suletud liidest, API dokumentatsiooni ja lõputu koodi ümberehitamise uurimisele.

Töötamine suletud formaatide ja süsteemidega ei ole ainult tehniline väljakutse - see on strateegiline piirang. Hoolimata kaasaegse automatiseerimise, tehisintellekti, LLM ja prognoosiva analüüsi pakutavatest ilmsetest võimalustest, ei suuda paljud ettevõtted oma potentsiaali täielikult ära kasutada. Ja tõkked, mis on püstitatud patenteeritud vormingute tõttu (joonis 3.2-13), keelavad ettevõtetel juurdepääsu oma andmetele. See ongi ehk ehituses toimuva digitaalse ümberkujundamise ironia.

Andmete läbipaistvus ja avatud süsteemid ei ole luksus, vaid kiiruse ja tõhususe eeltingimus. Ilma avatuseta on äriprotsessid täis tarbetut bürokraatiat, mitmekihilisi heakskiitmisahelaid ja kasvavat sõltuvust HiPPO põhimõttest - otsuste tegemine kõrgeima palgaga inimese arvamuse alusel.

Sellelgi poolt on silmapiiril kujunemas paradigmuuutus. Vaatamata patenteeritud lahenduste domineerimisele mõistab üha rohkem ettevõtteid neljanda tööstusrevolutsiooni inspireeritud arhitektuuride piiranguid. Praegu on vektor liikumas viienda revolutsiooni põhimõtete suunas, kus kesksel kohal on andmed kui strateegiline vara, avatud liidesed (APId) ja süsteemide tõeline koostalitlusvõime.

See üleminek tähistab üleminekut suletud ökosüsteemidest paindlike, modulaarsete digitaalsete

arhitektuuride suunas, kus avatud formaadid, standardid ja läbipaistev andmevahetus mängivad võtmerolli.

Avatud formaadid muudavad lähenemisviisi digitaliseerimisele

Ehitustööstus oli üks viimaseid, kes tegeles suletud ja konfidentsiaalsete andmete probleemiga. Erinevalt teistest majandussektoritest on digitaliseerimine siin aeglaselt arenenud. Selle põhjuseks on tööstuse traditsiooniline konservatiivsus, erinevate kohalike lahenduste levik ja paberipõhise dokumendihalduse sügavalt juurdunud olemus. Aastakümneid tuginesid peamised ehitusprotsessid füüsilistele joonistele, telefonikõnedele ja mittesünkroniseeritud andmebaasidele. Selles kontekstis on suletud formaate pikka aega peetud pigem normiks kui takistuseks.

Teiste tööstusharude kogemused näitavad, et suletud andmetele seatud tõkete kõrvaldamine toob kaasa innovatsiooni hüppelise kasvu, kiirema arengu ja suurema konkurentsi [73]. Teaduses võimaldab avatud andmete vahetamine kiirendada avastusi ja arendada rahvusvahelist koostööd. Meditsiinis võib see parandada diagnoosimise ja ravi tõhusust. Tarkvaraarenduses - luua koosloome ökosüsteeme ja kiiret tootearendust.

McKinsey aruande "Avatud andmed: Unlock Innovation and Productivity with Information Flow" 2013. [74] on avatud andmetel potentsiaali avada igal aastal 3 kuni 5 triljonit dollarit seitsmes võtmetähtsusega tööstusharus, sealhulgas ehituses, transpordis, tervishoius ja energeetikas. Sama uuringu kohaselt võimaldavad detsentraliseeritud andmete ökosüsteemid suurtel ehitusettevõtetel ja töövõtjatel vähendada tarkvara arendus- ja hoolduskulusid, kiirendades digitaalset kasutuselevõttu.

Üleminek avatud arhitektuuridele, mis on juba ammu alanud teistes majandussektorites, levib järkjärgult ka ehitussektorisse. Suurettevõtted ja avalikud tellijad ning eelkõige finantsorganisatsioonid, kes kontrollivad investeringuid ehitusprojektidesse, nõuavad üha enam avatud andmete kasutamist ja juurdepääsu arvutuste, arvutuste ja rakenduste lähtekoodile. Arendajatelt ei oodata enam ainult digitaalsete lahenduste loomist ja projekti lõplike arvude näitamist - neilt oodatakse läbipaistvust, reprodutseeritavust ja sõltumatust kolmandate osapoolte rakenduste müüjatest.

Avatud lähtekoodiga lahenduste kasutamine annab kliendile kindluse, et isegi kui välised arendajad lõpetavad koostöö või lahkuvad projektist, ei mõjuta see võimalust vahendeid ja süsteeme edasi arendada. Avatud andmete üks peamisi eeliseid on selle võime kaotada rakenduste arendajate sõltuvus konkreetsetest platvormidest, et pääseda ligi andmetele.

Kui ettevõtte ei saa täielikult loobuda patenteeritud lahendustest, on võimalik kompromiss pöördtehnika kasutamine. Need juriidiliselt ja tehniliselt usaldusväärsed meetodid võimaldavad muuta suletud formaadid kättesaadavaks, struktureeritumaks ja integreerimiseks sobivaks. See on eriti oluline, kui on vaja ühendada pärandisüsteeme või migreerida teavet ühest tarkvaramaastikust teise.

Üks eredamaid näiteid avatud vormingutele ülemineku ja pöördprojekteerimise (seaduslike süsteemide häkkimise) kohaldamise ajaloost ehituses on DWG-formaadi avamise eest peetud võitlus, mida kasutatakse laialdaselt arvutipõhistes projekteerimissüsteemides (CAD). Vastuseks ühe tarkvaratootja monopolile moodustasid 1998. aastal ülejäänud 15 CAD-tootjat uue liidu nimega "Open DWG", et pakkuda arendajatele tasuta ja sõltumatuid vahendeid DWG-vorminguga (de facto jooniste edastamise standard) töötamiseks ilma patenteeritud tarkvara või suletud API deta. See sündmus oli pöördepunkt, mis võimaldas kümnetel tuhandetel ettevõtetel saada vaba juurdepääsu 1980ndate lõpust kuni tänapäevani levinud CAD-lahenduse suletud formaadile ja luua ühilduvaid lahendusi, mis soodustasid konkurentsi CAD-turul [75]. Tänapäeval kasutatakse "Open DWG" SDK-d, mis loodi esimest korda juba 1996. aastal, peaaegu kõikides lahendustes, kus on võimalik importida, redigeerida ja eksportida DWG-vormingut, väljaspool DWG-vormingu arendaja ametlikku rakendust.

Teised tehnoloogiagigandid on sunnitud tegema samasuguseid muudatusi. Microsoft, mis kunagi oli omandiõiguse sümbol, avas .NET raamistiku lähtekoodi, hakkas kasutama Linuxi Azure'i pilveteenuse infrastruktuuris ja omandas GitHubi, et tugevdada oma positsiooni avatud lähtekoodiga kogukonnas. [76]. Meta (endine Facebook) avaldas avatud lähtekoodiga tehisintellekti mudelid, näiteks Llamade seeria, et edendada innovatsiooni ja koostööd tehisintellektiagentide arendamisel. Tegevjuht Mark Zuckerberg eeldab, et avatud lähtekoodiga platvormid on järgmise kümnendi jooksul tehnoloogilise arengu eestvedajateks [77].

Avatud lähtekood on tarkvara arendamise ja levitamise mudel, mille puhul lähtekood on avatud vabaks kasutamiseks, uurimiseks, muutmiseks ja levitamiseks.

Avatud andmed ja avatud lähtekoodiga lahendused ei ole muutumas mitte ainult trendiks, vaid digitaalse jätkusuutlikkuse aluseks. Need annavad ettevõtetele paindlikkuse, vastupidavuse, kontrolli oma otsuste üle ja võimaluse skaleerida digitaalseid protsesse sõltumata müüja poliitikast. Sama oluline on see, et need annavad ettevõtetele tagasi kontrolli 21. sajandi kõige väärtuslikuma ressursi - nende andmete - üle.

Paradigmamuutus: avatud lähtekood kui tarkvaratootjate domineerimise ajastu lõppemine.

Ehitustööstuses on toimumas muutus, mida ei saa tavapärasel viisil rahas väljendada. Andmepõhise, andmekeskse lähenemisviisi kontseptsioon ja avatud lähtekoodiga vahendite kasutamine viib mängureeglite ümbermõtestamiseni, millel turu tarkvaragigandid seisavad.

Erinevalt varasematest tehnoloogilistest ümberkorraldustest ei toeta müüjad seda üleminekut aktiivselt. Paradigmamuutus ohustab nende traditsioonilisi litsentseerimisel, tellimustel ja nõustamisel põhinevaid ärimudeleid. Uus reaalsus ei hõlma valmis toodet ega tasulist tellimust - see nõuab protsesside ja mõtlemise ümberkujundamist.

Avatud tehnoloogiatel põhinevate andmekeskuste lahenduste haldamiseks ja arendamiseks peavad ettevõtted sisemised protsessid ümber mõtlema. Eri osakondade spetsialistid peavad mitte ainult tegema koostööd, vaid ka mõtlema ümber, kuidas nad koos töötavad.

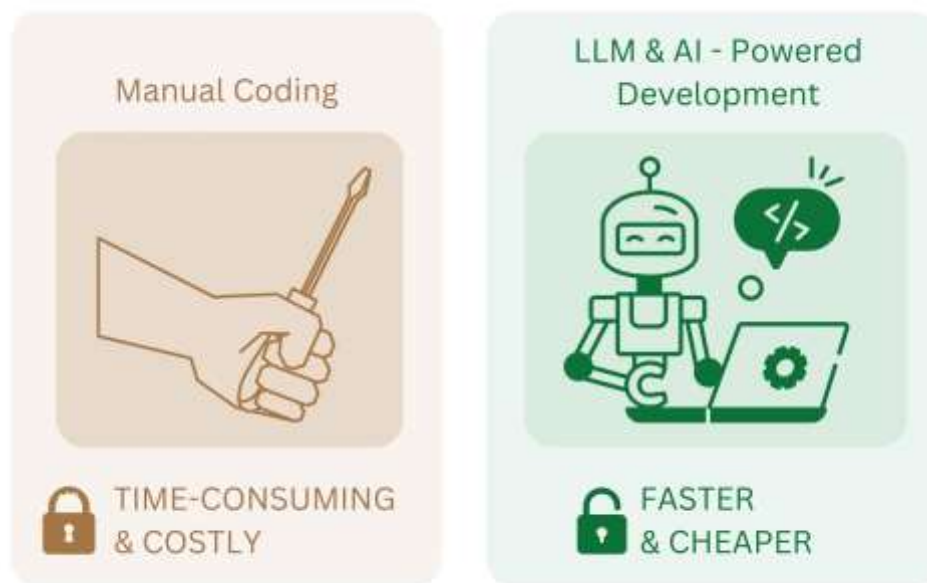
Uus paradigma eeldab avatud andmete ja avatud lähtekoodiga lahenduste kasutamist, kus tarkvarakoodi loomisel mängivad erilist rolli pigem tehisintellektil ja suurtel keelemudelitel (LLM) põhinevad vahendid kui programmeerijad. Juba 2024. aasta keskpaigaks luuakse üle 25 protsendi uuest koodist Google'is tehisintellekti abil [78]. Tulevikus tehakse LLMidega kodeerimisel 80% tööst vaid 20% ajaga (joonis 3.2-14).

McKinsey 2020. aasta uuringu [79] kohaselt asendavad GPU-d analüütikas üha enam protsessoreid tänu nende suurele jõudlusele ja kaasaegsete avatud lähtekoodiga vahendite toetusele. See võimaldab ettevõtetel kiirendada andmetöötlust ilma märkimisväärselte investeeringuteta kallisesse tarkvarasse või nappide spetsialistide palkamiseta.

Juhtivad konsultatsioonifirmad, nagu McKinsey, PwC ja Deloitte, rõhutavad avatud standardite, avatud lähtekoodiga rakenduste kasvavat tähtsust kõigis tööstusharudes.

PwC Avatud lähtekoodiga Monitor 2019 aruande [80] kohaselt kasutab 69% 100 või enama töötajaga ettevõtetest teadlikult avatud lähtekoodiga lahendusi. Eriti aktiivselt kasutatakse OSSi suurtes ettevõtetes: 71% 200-499 töötajaga ettevõtetes, 78% 500-1999 töötajaga ettevõtete kategoorias ja kuni 86% üle 2000 töötajaga ettevõtete seas. Synopsys OSSRA 2023 aruande kohaselt sisaldas 96% analüüsitud koodibaasidest avatud lähtekoodiga komponente [81].

Tulevikus ei ole arendaja roll mitte käsitsi koodi kirjutamine, vaid andmemudelite ja voogude arhitektuuride kavandamine ning tehisintellekti agentide haldamine, mis loovad nõudmise korral õigeid arvutusi. Kasutajaliidesed muutuvad minimalistlikuks ja interaktsioon muutub dialoogipõhiseks. Klassikaline programmeerimine annab teed kõrgetasemelisele disainile ja digitaalsete lahenduste orkestreerimisele (joonis 3.2-14). Praegused suundumused - nagu madala koodiga platvormid (joonis 7.4-6) ja LLM-võimelised ökosüsteemid (joonis 7.4-4) - vähendavad märkimisväärselt IT-süsteemide arendamise ja hooldamise kulusid.



Joonis 3.2-14 Kui täna luuakse rakendusi käsitsi programmeerijate poolt, siis tulevikus genereeritakse märkimisväärne osa koodist tehisintellekti ja LLM-põhiste lahenduste abil.

See üleminek ei erine eelmistest ning suur tarkvara tootja ei ole tõenäoliselt selle katalüsaatoriks.

Harvard Business School'i uuringus "The Value of Open Source Software" 2024 [40] on avatud lähtekoodiga tarkvara koguväärtus hinnatud kahest vaatenurgast. Ühest küljest, kui arvutada, kui palju kuluks kõigi olemasolevate avatud lähtekoodiga lahenduste loomiseks nullist, oleks see summa umbes 4,15 miljardit dollarit. Teisest küljest, kui kujutame ette, et iga ettevõtte arendab ise omal käel avatud lähtekoodiga lahenduste analooge (mis juhtub igal pool), ilma et tal oleks ligipääs olemasolevatele vahenditele, siis ulatuks ettevõtte kogukulu kolossaalsesse 8,8 triljoni dollarini - see on nõudluse hind.

Pole raske arvata, et ükski suur tarkvaratootja ei ole huvitatud 8,8 triljoni dollari suuruse potentsiaalse väärtusega tarkvaraturu kahanemisest vaid 4,15 miljardi dollarini. See tähendaks nõudluse mahu vähendamist rohkem kui 2000 korda. Selline ümberkujundamine on lihtsalt kahjumlik müüjatele, kelle ärimudelid on üles ehitatud aastatepikkusele klientide sõltuvuse säilitamisele suletud lahendustest. Seega võivad ettevõtted, kes ootavad, et keegi pakuks neile mugavat ja avatud valmislahendust, pettuda - need müüjad lihtsalt ei tule kohale.

Üleminek avatud digitaalsele arhitektuurile ei tähenda töökohtade või tulude kaotamist. Vastupidi, see loob tingimused paindlikele ja kohanemisvõimelistele ärimudelitele, mis võivad lõppkokkuvõttes tõrjuda traditsioonilise litsentsi- ja karbitarkvaraturu.

Litsentside müümise asemel - teenused, suletud formaatide asemel - avatud platvormid, müüjast sõltuvuse asemel - sõltumatus ja võimalus luua lahendusi tegelikele vajadustele. Need, kes varem lihtsalt kasutasid tööriistu, saavad nende kaasautoriteks. Ja need, kes oskavad töötada andmete,

modelite, stsenaariumide ja loogikaga, leiavad end tööstuse uue digitaalmajanduse keskmes. Neist muutustest ja sellest, millised uued rollid, ärimudelid ja koostöövormid on tekkimas avatud andmete ümber, räägime lähemalt raamatu viimases, kümnendas osas.

Avatud andmetel ja avatud koodil põhinevad lahendused võimaldavad ettevõtetel keskenduda äriprotsesside tõhususele, mitte vananenud API-de ja suletud süsteemide integreerimisele. Teadlik üleminek avatud arhitektuurile võib märkimisväärselt parandada tootlikkust ja vähendada sõltuvust müüjatest.

Üleminek uuele reaalsusele ei tähenda mitte ainult tarkvaraarenduse lähenemisviisi muutmist, vaid ka andmetega töötamise põhimõtte ümbermõtestamist. Selle muutuse keskmes ei ole mitte kood, vaid teave: selle struktuur, kättesaadavus ja tõlgendatavus. Siinkohal tõusevad esile avatud ja struktureeritud andmed, mis muutuvad uue digitaalse arhitektuuri lahutamatuks osaks.

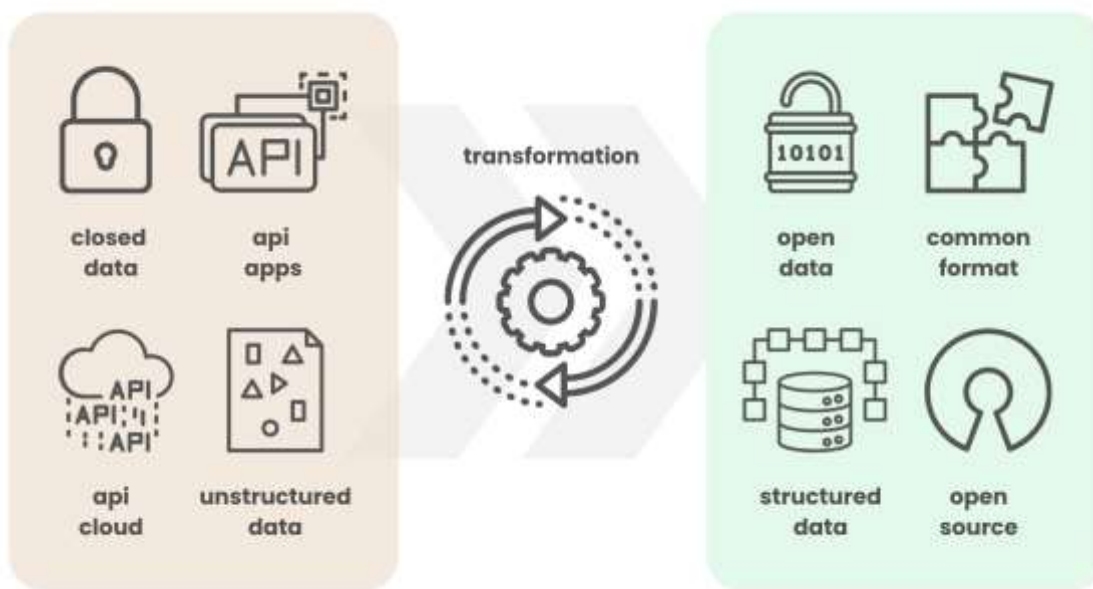
Struktureeritud avatud andmed: digitaalse ümberkujundamise alus

Kui varasematel aastakümnetel määras ettevõtte jätkusuutlikkuse suuresti tarkvaralahenduste valik ja sõltuvus konkreetsetest tarnijatest, siis tänapäeva digitaalmajanduses on võtmeteguriks andmete kvaliteet ja võime nendega tõhusalt töötada. Avatud lähtekood on uue tehnoloogilise paradigma oluline osa, kuid selle potentsiaali saab tõeliselt kasutada ainult siis, kui andmed on arusaadavad, korrastatud ja masinloetavad. Igasuguste andmemudelite hulgas on struktureeritud avatud andmed muutumas jätkusuutliku digitaalse ümberkujundamise nurgakiviks.

Struktureeritud avatud andmete peamine eelis on üheselt mõistetav tõlgendamine ja automatiseeritud töötlemise võimalus. See võimaldab märkimisväärset tõhususe kasvu nii üksikute toimingute kui ka kogu organisatsiooni tasandil.

Vastavalt Deloitte'i aruandele "The Data Transfer Process in Enterprise Transformation" ("Andmete edastamise protsess ettevõtte ümberkujundamisel") [82] on koostöö IT-ga struktureeritud andmete edastamise haldamiseks kriitilise tähtsusega. Ühendkuningriigi valitsuse aruande "Data Analytics and AI in Government Project Delivery" (2024) [83] kohaselt on projektijuhtimise tõhususe parandamisel võtmetähtsusega andmete jagamise takistuste kõrvaldamine erinevate projektide ja organisatsioonide vahel. Dokumendis rõhutatakse, et andmete vormingute standardiseerimine ja avatud andmete põhimõtete kasutuselevõtt võib vältida teabe dubleerimist, minimeerida ajakadu ja parandada prognooside täpsust.

Ehitustööstuses, kus traditsiooniliselt valitseb suur killustatus ja formaatide mitmekesisus, mängivad struktureeritud ühtlustamisprotsess ja struktureeritud avatud andmed olulist rolli ühtsete ja hallatavate protsesside kujundamisel (joonis 4.1-14). Need võimaldavad projektis osalejatel keskenduda pigem tootlikkuse parandamisele kui suletud platvormide, andmemudelite ja -vormingute vastuoludega seotud tehniliste probleemide lahendamisele.



Joonis 3.2-15 Avatud struktureeritud andmed vähendavad sõltuvust tarkvaralahendustest ja platvormidest ning kiirendavad innovatsiooni.

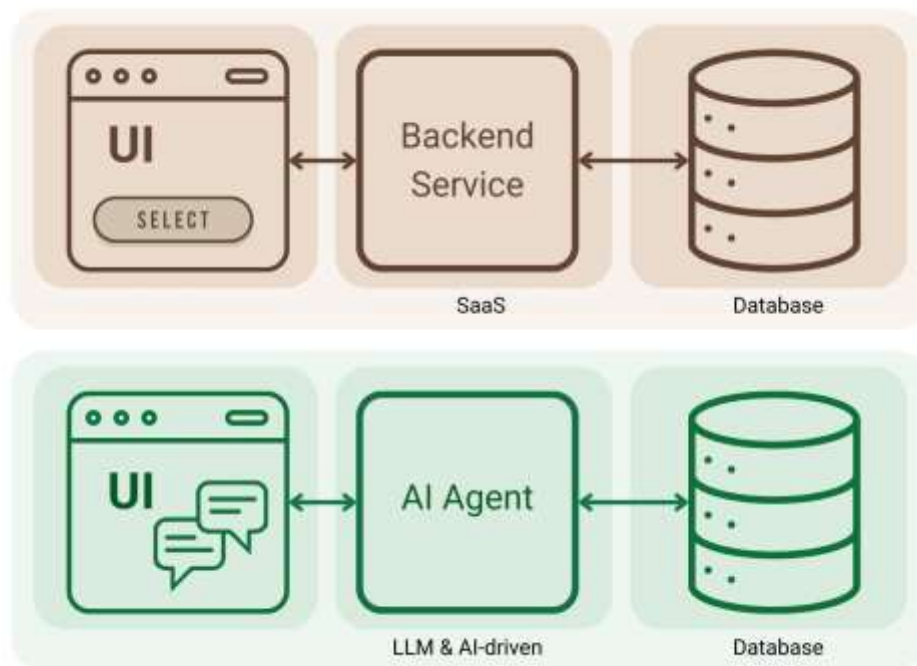
Kaasaegsed tehnoloogiavahendid, mida käsitleme üksikasjalikult hiljem raamatus, võimaldavad mitte ainult koguda teavet, vaid ka puhastada seda automaatselt: kõrvaldada dubleerimisi, parandada vigu ja normaliseerida väärtusi. See tähendab, et analüütikud ja insenerid ei tööta mitte lahtiste dokumentidega, vaid organiseeritud teadmistebaasiga, mis sobib analüüsiks, automatiseerimiseks ja otsuste tegemiseks.

Tehke see nii lihtsaks kui võimalik, kuid mitte lihtsamaks.

- Albert Einstein, teoreetiline füüsik (tsitaadi autorsus on vaieldav [84]).

Tänapäeval saab enamiku andmetega töötamise kasutajaliidest luua automaatselt - ilma et oleks vaja iga äritegevuse jaoks käsitsi koodi kirjutada. Selleks on vaja infrastruktuuri kihti, mis mõistab andmete struktuuri, mudelit ja loogikat ilma täiendavate juhisteta (joonis 4.1-15). Selline lähenemine on võimalik tänu struktureeritud andmetele: vormid, tabelid, filtrid ja vaated saab automaatselt luua minimaalse programmeerimisega.

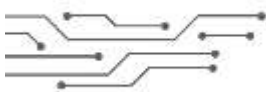
Kõige olulisemad kasutajakriitilised liidesed võivad endiselt vajada käsitsi täiustamist. Kuid enamikul juhtudel - ja see on 50-90 protsenti tööstsenariumidest - piisab rakenduste ja arvutuste automaatselt genereerimisest ilma selleks spetsiaalsete rakenduste kasutamiseta (joonis 3.2-16), mis vähendab oluliselt arendus- ja hoolduskulusid, vähendab vigu ja kiirendab digitaalsete lahenduste rakendamist.



Joonis 3.2-16 Andmetega töötamise arhitektuurimudelid: traditsiooniline rakendusarhitektuur ja tehisintellekti -orienteeritud mudel koos LLMiga.

Üleminek üksikutel rakendustel põhinevatelt arhitektuuridelt intelligentselt hallatavatele süsteemidele, mis põhinevad keelemudelitel (LLM), on järgmine samm digitaalses evolutsioonis. Sellise arhitektuuri puhul ei muutu struktureeritud andmed mitte ainult salvestusobjektiks, vaid ka suhtlemise aluseks tehisintellekti vahenditele, mis on võimelised analüüsima, tõlgendama ja soovitama tegevusi kontekstipõhiselt.

Järgnevates peatükkides vaatleme reaalseid näiteid avatud struktureeritud andmetel põhineva arhitektuuri rakendamise kohta ja näitame, kuidas keelemudeleid rakendatakse andmete automaatseks tõlgendamiseks, valideerimiseks ja töötlemiseks. Need praktilised juhtumid aitavad teil paremini mõista, kuidas uus digitaalne loogika toimib praktikas - ja millist kasu see toob muutusteks valmisolevatele ettevõtetele.



PEATÜKK 3.3.

LLM JA NENDE ROLL ANDMETÖÖTLUSES JA ÄRIPROTSESSIDES

LLM jututubades: ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok andmeotsesside automatiseerimiseks.

Suurte keelemudelite (LLM) tekkimine oli loomulikuks jätkuks liikumisele struktureeritud avatud andmete ja avatud lähtekoodiga filosoofia suunas. Kui andmed muutuvad organiseeritud, juurdepääsetavaks ja masinloetavaks, on järgmine samm tööriist, mis suudab selle teabega suhelda ilma keerulise koodi kirjutamise või tehniliste eriteadmiste omamiseta.

Elukestva õppe programmid on otsene tulemus avatusest: suured avatud andmekogumid, publikatsioonid ja avatud lähtekoodiga liikumine. Ilma avatud teadusartiklite, avalikult kättesaadavate tekstiliste andmete ja koostööpõhise arenduskultuurita ei oleks olemas ChatGPT-d ega muid elukestva õppe programme. LLM on teatud mõttes "destillaat" inimkonna kogutud digitaalsetest teadmistest, mis on kogutud ja haritud avatuse põhimõtete kaudu.

Kaasaegsed suured keelemudelid (LLM - Large Language Models) nagu ChatGPT® (OpenAI), LLaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™ annavad kasutajatele võimaluse sõnastada päringuid andmetele loomulikus keeles. See muudab teabega töötamise kättesaadavaks mitte ainult arendajatele, vaid ka analüütikutele, inseneridele, planeerijatele, juhtidele ja muudele spetsialistidele, kes varem olid programmeerimisest kaugel

LLM (Large Language Model) on tehisintellekt, mis on koolitatud mõistma ja genereerima teksti, mis põhineb suurel hulgal internetist kogutud andmetel. See suudab analüüsida konteksti, vastata küsimustele, pidada dialoogi, kirjutada tekste ja genereerida tarkvarakoodi.

Kui varem nõudis andmete visualiseerimine, töötlemine või analüüs spetsiaalse programmeerimiskeele - Python, SQL, R või Scala - tundmist, samuti oskust töötada selliste raamatukogudega nagu Pandas, Polars või DuckDB ja paljude teistega, siis alates 2023. aastast on olukord radikaalselt muutunud. Nüüd saab kasutaja lihtsalt kirjeldada, mida ta tahab saada - ja mudel ise genereerib koodi, täidab selle, kuvab tabeli või graafiku ja selgitab tulemust. Esimest korda aastakümnete jooksul ei ole tehnoloogia areng läinud mitte keerukuse, vaid radikaalse lihtsustamise ja juurdepääsetavuse teed.

See põhimõte - "töötle andmeid sõnadega (prompts)" - tähistas uut etappi teabega töötamise arengus, viies lahenduste loomise veelgi kõrgemale abstraktsioonitasandile. Nii nagu kunagi ei olnud enam vaja, et kasutajad mõistaksid interneti tehnilisi aluseid, et pidada veebipoe või luua veebisait WordPressi, Joomla ja teiste avatud lähtekoodiga moodulsüsteemide abil (av tor raamatus on selliste süsteemidega töötanud alates 2005. aastast, sealhulgas haridus- ja inseneriplatvormide

veebiplatvormidega). - See omakorda on toonud kaasa digitaalse sisu ja veebipõhise äri buumi - tänapäeval saavad insenerid, analüütikud ja juhid automatiseerida töövooge ilma programmeerimiskeelte tundmiseta. Seda soodustavad võimsad LLM-id - nii tasuta kui ka avatud lähtekoodiga, nagu LLaMA, Mistral, Qwen, DeepSeek ja teised -, mis teevad arenenud tehnoloogiad kättesaadavaks võimalikult laiale publikule.

Suured LLM keelemudelid: kuidas see toimib

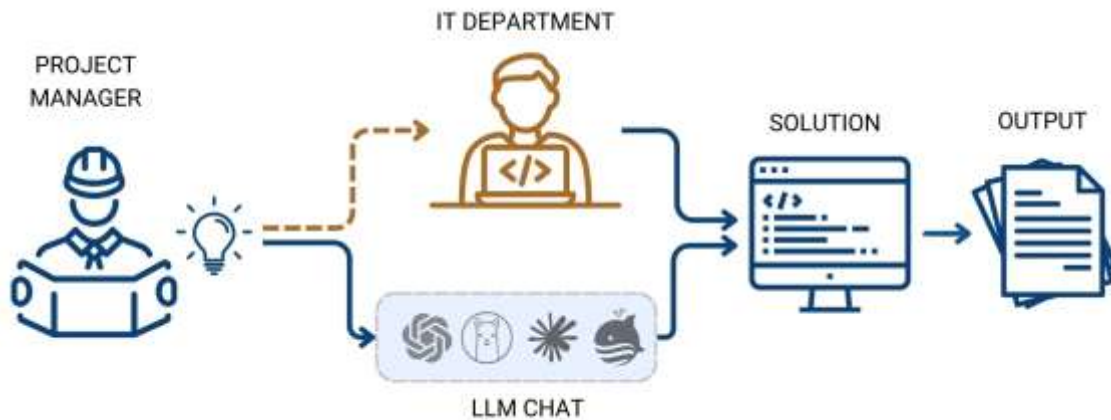
Suured keelemudelid (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) on neuronivõrgud, mis on koolitatud tohutu hulga tekstiandmete põhjal, mis pärinevad internetist, raamatutest, artiklitest ja muudest allikatest. Nende peamine ülesanne on mõista inimkõne konteksti ja genereerida mõtestatud vastuseid.

Kaasaegne LLM põhineb Google'i teadlaste poolt 2017. aastal välja pakutud Transformer-arhitektuuril [85]. Selle arhitektuuri põhikomponent on tähelepanumehhanism, mis võimaldab mudelil arvestada sõnade vahelisi seoseid sõltumata nende asukohast tekstis.

LLM-i õppimisprotsess on kaugeltki sarnane sellega, kuidas inimesed õpivad keelt, ainult et see on miljoneid kordi suurem. Mudel analüüsib miljardeid sõnade ja väljendite näiteid, tuvastades mustreid keele struktuuris ja semantiliste üleminekute loogikas. Kogu tekst jaotatakse märgideks - minimaalseteks semantilisteks ühikuteks (sõnad või nende osad), mis seejärel teisendatakse vektoriteks mitmemõõtmelises ruumis (joonis 8.2-2). Need vektoriesindused võimaldavad masinal "mõista" mõistete vahelisi varjatud seoseid, selle asemel, et lihtsalt tegutseda tekstiga kui sümbolite jadaga.

Suured keelemudelid ei ole ainult vahendid teksti genereerimiseks. Nad suudavad ära tunda tähendusi, leida seoseid mõistete vahel ja töötada andmetega, isegi kui need on esitatud erinevates vormingutes. Peamine on see, et teave tuleb jaotada arusaadavateks mudeliteks ja esitada märgenditena, millega LLM saab töötada.

Sama lähenemisviisi saab kohaldada ka ehitusprojektide suhtes. Kui me mõtleme projektist kui mingist tekstist, kus iga hoone, element või konstruktsioon on sümbol, siis võime hakata sellist teavet sarnaselt töötlema. Ehitusprojekte võib võrrelda raamatutega, mis on organiseeritud kategooriatesse, peatükkidesse ja paragrahvigruppidesse, mis koosnevad minimaalsetest märgenditest - ehitusprojekti elementidest (joonis 3.3-1). Andmemudeleid struktureeritud vormingusse tõlkides saame struktureeritud andmed tõlkida ka vektorbaasideks (joonis 8.2-2), mis on ideaalne allikas masinõppe ja selliste tehnoloogiate nagu LLM jaoks.



Joonis 3.3-2 LLMid võimaldavad kasutajatel kirjutada koodi ja saada tulemusi ilma programmeerimisoskusteta.

Wakefield Researchi poolt läbi viidud ja SAPi poolt sponsoreeritud 2024. aasta uuringu [36] kohaselt, mille käigus küsitleti 300 tippjuhti ettevõtetes, mille aastatulu on vähemalt 1 miljard dollarit USAs: 52% tippjuhtidest usaldab tehisintellekti andmete analüüsimiseks ja otsuste tegemiseks vajalike soovitude andmiseks. Veel 48% kasutab tehisintellekti varem arvestamata jäänud riskide tuvastamiseks ja 47% kasutab tehisintellekti alternatiivsete plaanide soovitamiseks. Lisaks kasutab 40% tehisintellekti uute toodete arendamiseks, eelarve planeerimiseks ja turu-uuringuteks. Uuring näitas ka tehisintellekti positiivset mõju isiklikule elule: 39% vastanutest teatas, et töö- ja eraelu tasakaal on paranenud, 38%, et vaimne tervis on paranenud, ja 31%, et stressitase on langenud.

Kogu oma võimsuse juures on elukestva õppe mudelid siiski endiselt vahend, mida on oluline kasutada teadlikult. Nagu igal tehnoloogial, on ka neil piirangud. Üks tuntumaid probleeme on nn "hallutsinatsioonid" - juhtumid, kus mudel annab kindlalt usutava, kuid faktiliselt vale vastuse. Seetõttu on väga oluline mõista, kuidas mudel töötab: milliseid andmeid ja andmemudeleid ta suudab vigadeta tõlgendada, kuidas ta päringuid tõlgendab ja kust ta oma teavet saab. Samuti tasub meeles pidada, et LLMi teadmised piirduvad tema väljaõppe kuupäevaga ning ilma ühenduseta väliste andmetega ei pruugi mudel võtta arvesse kehtivaid norme, standardeid, hindu või tehnoloogiaid.

Nende probleemide lahenduseks on vektorandmebaaside korrapärane ajakohastamine, asjakohaste allikatega ühendamine ja autonoomse tehisintellekti - agentide arendamine, mis ei vasta ainult küsimustele, vaid kasutavad andmeid proaktiivselt koolituseks, haldavad ülesandeid, tuvastavad riske, pakuvad optimeerimisvõimalusi ja jälgivad projekti tulemuslikkust.

Üleminek LLM -liidestele ehituses ei ole ainult tehnoloogiline uuendus. See on paradigmuuutus, mis kõrvaldab tõkked inimeste ja andmete vahel. See on võimalus töötada teabega sama lihtsalt kui me omavahel räägime - ja saada ikkagi täpseid, kontrollitud ja rakendatavaid tulemusi.

Need ettevõtted, kes hakkavad selliseid vahendeid teistest varem kasutama, saavad märkimisväärse konkurentsieelise. See hõlmab töö kiirendamist, kulude vähendamist ja projekteerimislahenduste

kvaliteedi parandamist tänu kiirele juurdepääsule andmeanalüüsile ja võimele leida kiiresti vastused keerulistele küsimustele. Kuid arvesse tuleb võtta ka julgeolekuküsimusi. Pilvepõhiste LLM -teenuste kasutamisega võib kaasneda andmete lekkimise oht. Seetõttu otsivad organisatsioonid üha enam alternatiivseid lahendusi, mis võimaldavad neil kasutada LLM-vahendeid oma infrastruktuuris - kohapeal, täieliku kaitse ja kontrolliga teabe üle.

Kohalike LLMide kasutamine ettevõtte tundlike andmete jaoks

Esimeste chat-LLMide ilmumine 2022 tähistas uut etappi tehisintellekti arengus. Kuid kohe pärast nende mudelite laialdast kasutuselevõttu tekkis õigustatud küsimus: kui turvaline on ettevõttega seotud andmete ja päringute edastamine pilve? Enamik pilvepõhiseid keelemudeleid salvestas oma serveritesse suhtlusajalugu ja üleslaaditud dokumente ning tundliku teabega tegelevate ettevõtete jaoks oli see tõsine takistus tehisintellekti kasutuselevõtul.

Üks kõige jätkusuutlikumaid ja loogilisemaid lahendusi sellele probleemile on olnud avatud lähtekoodiga LLM kohalik kasutuselevõtt ettevõtte IT-infrastruktuuris. Erinevalt pilveteenustest töötavad lokaalsed mudelid ilma internetiühenduseta, ei edasta andmeid välistesse serveritesse ja annavad ettevõtetele täieliku kontrolli teabe üle

Parim avatud mudel [Open Source LLM] on praegu võrreldav suletud mudelite [nagu ChatGPT, Claude] jõudlusega, kuid umbes aastase mahajäämusega [77].

- Ben Cottier, mittetulundusliku teadusasutuse Epoch AI juhtivteadur, 2024

Suuremad tehnoloogiaettevõtted on hakanud oma elukestva õppe programme kohalikuks kasutamiseks kättesaadavaks tegema. Meta avatud lähtekoodiga LLaMA-seeria ja Hiina kiiresti kasvav projekt DeepSeek olid näited avatud arhitektuuri suunas liikumisest. Nende kõrval on ka Mistral ja Falcon andnud välja võimsaid mudeleid, mis on vabastatud patenteeritud platvormide piirangutest. Need algatused ei ole mitte ainult kiirendanud ülemaailmse tehisintellekti arengut, vaid on andnud ka eraelu puutumatust arvestavatele ettevõtetele tõelisi alternatiive sõltumatuse, paindlikkuse ja turvalisuse tagamise osas.

Ettevõtte keskkonnas, eriti ehitussektoris, ei ole andmekaitse mitte ainult mugavuse, vaid ka õigusnormide täitmise küsimus. Töö pakkumisdokumentide, kalkulatsioonide, jooniste ja konfidentsiaalse kirjavahetusega nõuab ranget kontrolli. Siinkohal pakub kohalik LLM vajalikku kindlustunnet, et andmed jäävad ettevõtte piiridesse.

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

Joonis 3.3-3 Kohalikud mudelid pakuvad täielikku kontrolli ja turvalisust, samas kui pilvepõhised lahendused pakuvad lihtsat integreerimist ja automaatseid uuendusi.

Kohaliku avatud lähtekoodiga LLM:

- Täielik kontroll andmete üle. Kogu teave jääb ettevõttesse, välistades volitamata juurdepääsu ja andmete lekkimise.
- Offline-töö. Ei sõltu internetiühendusest, mis on eriti oluline isoleeritud IT-infrastruktuuride puhul. See tagab katkematu töö ka sanktsioonide või blokeeritud pilveteenuste korral.
- Rakenduse paindlikkus. Mudelit saab kasutada teksti genereerimiseks, andmeanalüüsiks, programmikoodi kirjutamiseks, projekteerimise toetamiseks ja äriprotsesside haldamiseks.
- Kohandamine ettevõtte eesmärkidega. LLM saab koolitada ettevõttesiseseid dokumente, mis võimaldab arvesse võtta ettevõtte töö eripära ja selle tööstusharu eripärasid. Kohalikku LLMi saab ühendada CRMi, ERPi või BI platvormidega, mis võimaldab automatiseerida

kliendipäringute analüüsi, aruannete koostamist või isegi trendide prognoosimist.

DeepSeeki tasuta ja avatud lähtekoodiga mudeli -R1-7B kasutuselevõtt serveris, millele on juurdepääs kogu kasutajaskonnale, maksab 1000 dollarit kuus ja võib potentsiaalselt maksta vähem kui pilve APIde, näiteks ChatGPT või Claude aastamaksud, ning võimaldab ettevõtetel oma andmeid täielikult kontrollida, välistab nende edastamise internetti ja aitab täita regulatiivseid nõudeid, nagu GDPR

Teistes tööstusharudes on kohalikud LLMid juba muutnud oma lähenemist automatiseerimisele. Tugiteenuste puhul vastavad nad sagedastele kliendipäringutele, vähendades operaatorite töökoormust. Personaliosakondades analüüsivad nad CVsid ja valivad välja sobivad kandidaadid. E-kaubanduses genereerivad nad personaliseeritud pakkumisi ilma kasutaja andmeid avaldamata.

Sarnast mõju on oodata ka ehitussektoris. Tänu LLM integreerimisele projektiandmete ja standarditega on võimalik kiirendada dokumentatsiooni koostamist, automatiseerida kalkulasioonide koostamist ja prognoosivaid kuluanalüüse. LLMi kasutamine koos struktureeritud tabelite ja andmeraamistikega on muutumas eriti paljulubavaks valdkonnaks.

Täielik kontroll tehisintellekti üle ettevõttes ja kuidas oma LLM-i kasutusele võtta

Kaasaegsed tööriistad võimaldavad ettevõtetel kasutada suurt keelemudelit (LLM) lokaalselt vaid mõne tunniga. See annab täieliku kontrolli andmete ja infrastruktuuri üle, välistades sõltuvuse välistest pilveteenustest ja minimeerides teabe lekkimise riski. See lahendus on eriti oluline organisatsioonide jaoks, kes töötavad tundlike projektdokumentidega või äriliselt tundlike andmetega.

Sõltuvalt ülesannetest ja ressurssidest on saadaval erinevad kasutuselevõtustsenaariumid, alates valmislahendustest kuni paindlikumate ja skaleeritavamate arhitektuurideni. Üks lihtsamaid vahendeid on Ollama, mis võimaldab keelemudeleid käivitada sõna otseses mõttes ühe klõpsuga, ilma et oleks vaja sügavaid tehnilisi teadmisi. Kiire algus Ollama abil:

1. Laadige oma operatsioonisüsteemi (Windows / Linux / macOS) jaoks mõeldud distributsioon alla ametlikult veebilehelt: ollama.com
2. Paigaldage mudel käsurea kaudu. Näiteks *Mistrali* mudeli puhul:

```
ollama run mistral
```

3. Pärast mudeli käivitamist on see valmis töötama - saate saata tekstiküsimusi terminali kaudu või integreerida selle teistesse tööriistadesse. Käivitage mudel ja käivitage päring:

ollama run mistral "Kuidas koostada arvutust koos kõigi ressurssidega 100 mm la-
iuse kipsplaadist vaheseina paigaldamise tööde kohta?".

Neile, kes eelistavad töötada tuttavas visuaalses keskkonnas, on olemas LM Studio, tasuta rakendus, mille kasutajaliides meenutab ChatGPT-i

- Paigaldage LM Studio, laadides alla distributsioonikomplekti ametlikult kodulehelt - lmstudio.ai
- Valige sisseehitatud kataloogi kaudu mudel (nt Falcon või GPT-Neo-X) ja laadige see alla.
- Töö mudeliga intuitiivse kasutajaliidese kaudu, mis meenutab ChatGPT, kuid on täielikult lokaliseeritud.

	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

Joonis 3.3-4 Populaarsete kohalike avatud lähtekoodiga LLM -mudelite võrdlus.

Mudeli valik sõltub kiiruse, täpsuse ja olemasolevate riistvaraliste võimaluste nõuetest (joonis 3.3-4). Väikesed mudelid, nagu Mistral 7B ja Baichuan 7B, sobivad kergete ülesannete ja mobiilsete seadmete jaoks, samas kui võimsad mudelid, nagu DeepSeek-V3, nõuavad märkimisväärsed arvutusressursse, kuid pakuvad suurt jõudlust ja toetust mitmetele keeltele. Lähiaastatel kasvab LLM-turg kiiresti - näeme üha rohkem kergemaid ja spetsialiseeritud mudeleid. Üldotstarbeliste LLMide asemel, mis katavad kogu inimsisu, tekivad kitsaste valdkondlike eksperditeadmiste põhjal koolitatud mudelid.

Näiteks võime oodata selliste mudelite tekkimist, mis on mõeldud üksnes insenerarvutuste, ehitushinnangute või CAD-formaadis andmete töötlemiseks. Sellised spetsialiseeritud mudelid on kiiremad, täpsemad ja turvalisemad kasutada - eriti professionaalsetes keskkondades, kus on oluline kõrge usaldusväärsus ja teematevaheline sügavus.

Kui kohalik LLM on käivitatud, saab seda kohandada vastavalt ettevõtte konkreetsetele ülesannetele. Selleks kasutatakse peenhäälestustehnikat, mille puhul mudelit täiendavalt treenitakse sisedokumentide, tehniliste juhendite, lepingumallide või projektdokumentatsiooni alusel.

RAG: Intelligent LLM - assistendid, kellel on juurdepääs ettevõtte andmetele.

Järgmine etapp LLM-rakenduse arengus äris on mudelite integreerimine tegelike reaalajas ettevõtte andmetega. Seda lähenemisviisi nimetatakse RAG (Retrieval-Augmented Generation) - väljavõtte- ja täiendatud põlvkond (Retrieval-Augmented Generation). Selles arhitektuuris muutub keelemudel mitte ainult dialoogiliideseks, vaid täieõiguslikuks intelligentseks assistendiks, mis suudab liikuda dokumentides, joonistes, andmebaasides ja anda täpseid, kontekstipõhiseid vastuseid.

RAGi peamine eelis on võimalus kasutada ettevõtte sisemisi andmeid, ilma et oleks vaja mudelit eelnevalt treenida, säilitades samas suure täpsuse ja paindlikkuse teabe käitlemisel.

RAG tehnoloogia ühendab endas kaks põhikomponenti:

- Väljaotsimine: mudel ühendub andmehoidlatega - dokumendid, tabelid, PDF -failid, joonised - ja hangib asjakohast teavet vastavalt kasutaja soovile.
- Täiendatud genereerimine: mudel genereerib väljavõetud andmete põhjal täpse ja teadliku vastuse, võttes arvesse päringu konteksti ja spetsiifilisust.

LLMi käivitamiseks koos RAGi toetusega tuleb järgida mõningaid samme:

- **Andmete ettevalmistamine:** vajalike dokumentide, jooniste, spetsifikatsioonide, tabelite kogumine. Need võivad olla erinevates formaatides ja struktuurides, alates PDF-st kuni Excelini.
- **Indekseerimine ja vektoriseerimine:** kasutades selliseid vahendeid nagu LlamaIndex või LangChain, teisendatakse andmed vektoriesindusteks, mis võimaldavad leida tekstifragmentide vahelisi semantilisi seoseid (rohkem vektorandmebaasidest ja suurte massiiviandmete tõlkimisest vektoriesinduseks, sealhulgas CAD projektidest, 8. osas).
- **Küsi assistendilt:** kui andmed on üles laaditud, saad esitada mudelile küsimusi ja see otsib vastuseid ettevõtte raamistikus, mitte internetist kogutud üldistes teadmistes.

Oletame, et ettevõttel on kaust constructionsite_docs, kuhu on salvestatud lepingud, juhendid, kalkulatsioonid ja tabelid. Pythoni skripti abil (joonis 3.3-5) saame seda kausta skaneerida ja luua vektorindeksit: iga dokument teisendatakse vektorite kogumiks, mis kajastab teksti semantilist sisu.

See muudab dokumendid omamoodi "täenduskaardiks", millel mudel saab tõhusalt liikuda ja leida seoseid terminite ja fraaside vahel.

Näiteks "mäletab" mudel, et sõnad "tagastamine" ja "kaebus" on sageli lepingu osas, mis käsitleb materjalide saatmist ehitusplatsile. Siis, kui esitatakse küsimus - näiteks: "Milline on meie tagastusperiood?". (joonis 3.3-5 - koodirida 11) - analüüsib LLM sisemisi dokumente ja leiab täpse teabe, tegutsedes nagu intelligentne assistent, kes suudab lugeda ja mõista kõigi ettevõtte failide sisu.



```

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)

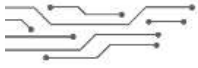
```

Joonis 3.3-5 LM loeb failidega kausta - sarnaselt sellega, kuidas inimene avab selle ja otsib soovitud dokumenti

Koodi saab käivitada mis tahes arvutis, kuhu on installeeritud Python. Pythoni ja IDE-de kasutamisest koodi käivitamiseks räägime lähemalt järgmises peatükis.

LLMi kohalik kasutuselevõtt ei ole lihtsalt trend, vaid strateegiline lahendus ettevõtetele, kes hindavad turvalisust ja paindlikkust. LLMi kasutuselevõtt kas ettevõtte kohalikes arvutites või veebipõhiste lahenduste abil on siiski ainult esimene samm. Selleks, et LLMi võimeid reaalses ülesannetes rakendada, peavad ettevõtted kasutama vahendeid, mis võimaldavad neil mitte ainult saada vestlusvastuseid, vaid ka salvestada loodud loogikat koodi kujul, mida saab käivitada väljaspool LLMi kasutamise konteksti. See on oluline lahenduste skaleerimiseks - korralikult korraldatud protsessid võimaldavad tehisintellekti arendusi rakendada korraga mitme projekti puhul või isegi kogu ettevõttes.

Selles kontekstis on oluline roll sobiva arenduskeskkonna (IDE) valikul. Kaasaegsed programmeerimisvahendid võimaldavad mitte ainult arendada LLM-põhiseid lahendusi, vaid ka integreerida neid olemasolevatesse äriprotsessidesse, muutes need automatiseeritud ETL-iks - Pipeline



PEATÜKK 3.4.

IDE KOOS LLM-TOEGA JA TULEVASED PROGRAMMIMUUDATUSED

IDE valimine: LLM-ist eksperimendid ärilahendusteni

Kui sukeldutakse automatiseerimise, andmeanalüüsi ja tehisintellekti maailma - eriti kui töötatakse suurte keelemudelitega (LLM) - on oluline valida õige integreeritud arenduskeskkond (IDE). See IDE on teie peamine töövahend: koht, kus LLMi genereeritud koodi käivitatakse nii kohalikus arvutis kui ka ettevõtte võrgus. IDE valik ei määra mitte ainult teie töö mugavust, vaid ka seda, kui kiiresti te suudate liikuda LLMi eksperimentaalsetest päringutest täisväärtuslike lahendusteni, mis on integreeritud tegelikesse äriprotsessidesse.

IDE (integreeritud arenduskeskkond) on mitmekülgne ehitusplokk teie arvutis protsesside automatiseerimiseks ja andmetöötluseks. Selle asemel, et hoida eraldi saagi, haamrit, puuri ja muid tööriistu, on teil üks seade, millega saab kõike seda teha - lõigata, kinnitada, puurida ja isegi kontrollida materjalide kvaliteeti. IDE programmeerijatele on üks ruum, kus saab kirjutada koodi (analoogiliselt ehitusega - luua jooniseid), testida selle tööd (ehitusmodeli kokkupanek), leida vigu (nagu konstruktsioonide tugevuse kontrollimine ehituses) ja käivitada valmis projekti (maja kasutuselevõtmine).

Ülevaade populaarsetest IDE-dest:

- **PyCharm®** (JetBrains) on võimas professionaalne IDE Pythonile. See sobib hästi tõsiste projektide jaoks tänu suurele hulgale sisseehitatud funktsioonidele. Interaktiivsete Jupyter-failide (IPYNB) põhitugi on siiski saadaval ainult tasulises versioonis ja algajatele võib kasutajaliides tunduda üle jõu käiv.

IPYNB laiendiga fail (Interactive Python Notebook) on interaktiivsete Jupyter® märkmikute formaat (joonis 3.4-1), kus kood, visualiseeringud ja selgitused on ühendatud ühte dokumenti. See vorming sobib ideaalselt aruannete, analüütika ja koolitusskenaariumide koostamiseks.

- **VS Code®** (Microsoft) on kiire, paindlik ja kohandatav tööriist, millel on tasuta IPYNB-tugi ja palju lisaseadmeid. Sobib nii algajatele kui ka professionaalidele. Võimaldab integreerida GitHub Copilot ja keelemudelite pistikprogramme, mistõttu on see suurepärane valik tehisintellekti ja andmeteaduse projektide jaoks.
- **Jupyter Notebook** - klassikaline ja populaarne valik eksperimenteerimiseks ja õppimiseks. See võimaldab kirjutada koodi, lisada selgitusi ja visualiseerida tulemusi ühes kasutajaliideses (joonis 3.4-1). Ideaalne hüpoteeside kiireks testimiseks, tööks LLMiga ja reprodutseeritavate metsiku andmeanalüüsi etappide loomiseks. Sõltuvuste ja raamatukogude haldamiseks soovitame kasutada Anaconda Navigatorit, mis on visuaalne kasutajaliides Pythoni keskkonna haldamiseks.

lahendusteks muuta.

Kõik eespool kirjeldatud IDEd võimaldavad luua andmetöötlusahelaid, st koodiplokkide moodulite ahelaid (mida võib genereerida LLM), millest igaüks vastutab näiteks eri etapi eest:

- analüütilised stsenaariumid,
- dokumentidest teabe väljavõtte ahelad,
- automaatne reageerimine RAGi alusel,
- aruannete ja visualiseerimiste genereerimine.

Tänu modulaarsele ülesehitusele saab iga sammu esitada eraldi plokina: andmete laadimine→ filtreerimine→ analüüs→ visualiseerimine→ tulemuste eksportimine. Neid plokkide saab taaskasutada, - kohandada ja kokku panna uuteks ahelateks nagu konstruktor, ainult andmete jaoks.

Inseneridele, juhtidele ja analüütikutele avab see võimaluse dokumenteerida otsustusloogika koodina, mida saab genereerida LLM-i abil. Selline lähenemisviis aitab kiirendada rutiinseid ülesandeid, automatiseerida tüüpilisi toiminguid ja luua korratavaid protsesse, kus iga samm on selgelt dokumenteeritud ja kõigile meeskonnaliikmetele läbipaistev.

Automatiseeritud ETL Pipelines (joonis 7.2-3), Apache Airflow (joonis 7.4-4), Apache NiFi (joonis 7.4-5) ja n8n (joonis 7.4-6) vahendeid protsesside automatiseerimise loogikablokkide ehitamiseks käsitletakse üksikasjalikumalt raamatu 7. ja 8. osas.

LLM-i toetav IDE ja tulevased programmeerimise muudatused

Tehisintellekti integreerimine arendusprotsessidesse muudab programmeerimismaastikku. Kaasaegsed keskkonnad ei ole enam lihtsalt tekstiredaktorid koos süntaksi esiletõstmisega - need on muutumas intelligentseteks assistentideks, mis suudavad mõista projekti loogikat, täiendada koodi ja isegi selgitada, kuidas konkreetne koodifragment töötab. Turule ilmuvad tooted, mis kasutavad tehisintellekti, et laiendada tavapärase arenduse piire:

- **GitHub Copilot** (integreeritav VS Code'iga, PyCharmiga): AI -assistent, mis genereerib koodi kommentaaride või osaliste kirjelduste põhjal, muutes tekstilised vihjed valmislahendusteks.
- **Cursor** (VS Code'i haru koos AI -kerneliga): võimaldab mitte ainult koodi kirjutamist lõpetada, vaid ka esitada küsimusi projektile, otsida sõltuvusi ja õppida koodibaasist.
- **JetBrains AI Assistant**: JetBrains IDE (sh PyCharm) plugin, mille funktsiooniks on keerulise koodi selgitamine, optimeerimine ja testide loomine.
- **Amazon CodeWhisperer**: analoog Copilotile, mis keskendub turvalisusele ja Amazoni AWS-i teenuste toetamisele.

Programmitöö muutub lähiaastatel järsult. Põhirõhk nihkub rutiinsest koodi kirjutamisest mudeli ja andmearhitektuuri kavandamisele - arendajad hakkavad rohkem osalema süsteemi kavandamises, samas kui tehisintellektuaal võtab üle malliülesanded: koodi genereerimine, testid, dokumentatsioon ja põhifunktsioonid. Programmeerimise tulevik on inimeste ja tehisintellekti koostöö, kus masinad võtavad üle tehnilise rutiini ja inimesed keskenduvad loovusele.

Loomulikus keeles programmeerimine muutub igapäevaseks. IDE personaliseerimine jõuab uuele tasemele - arenduskeskkonnad õpivad kohanema kasutaja tööstiili ja ettevõtetega, ennetades mustreid, pakkudes kontekstipõhiseid lahendusi ja õppides varasematest projektidest

See ei kaota arendaja rolli, kuid muudab seda radikaalselt: koodi kirjutamisest teadmiste, kvaliteedi ja protsesside juhtimisele. See areng mõjutab ka äriteabe analüüsi, kus aruannete, visualiseerimiste ja otsustamist toetavate rakenduste loomine hakkab üha enam toimuma koodi ja loogika genereerimise kaudu tehisintellekti ja LLMi, vestluse ja agentide kasutajaliidestest abil.

Kui ettevõtte on LLM-vestlused sisse seadnud ja valinud sobiva arenduskeskkonna, on järgmine oluline samm andmete korrastamine. See protsess hõlmab teabe eraldamist erinevatest allikatest, selle puhastamist, struktureeritud kujul teisendamist ja integreerimist ettevõtte süsteemidesse.

Kaasaegse andmekeskse andmehalduse lähenemisviisi peamine eesmärk on viia andmed ühtsesse universaalsesse vormi, mis ühildub paljude vahendite ja rakendustega. Struktureerimisprotsesside ja struktureeritud andmete käsitlemiseks on vaja spetsiaalseid raamatukogusid. Üks võimsamaid, paindlikumaid ja populaarsemaid on Pandase raamatukogu Pythonile. See võimaldab mugavalt töödelda tabeliandmeid: filtreerida, rühmitada, puhastada, lisada, koondada ja esitada aruandeid.

Python Pandas: asendamatu tööriist andmetega töötamiseks

Pandas on andmeanalüüsi ja automatiseerimise maailmas eriline koht. See on üks populaarsemaid ja laialdaselt kasutatavaid Pythoni programmeerimiskeele [86] raamatukogusid, mis on mõeldud tööks struktureeritud andmetega.

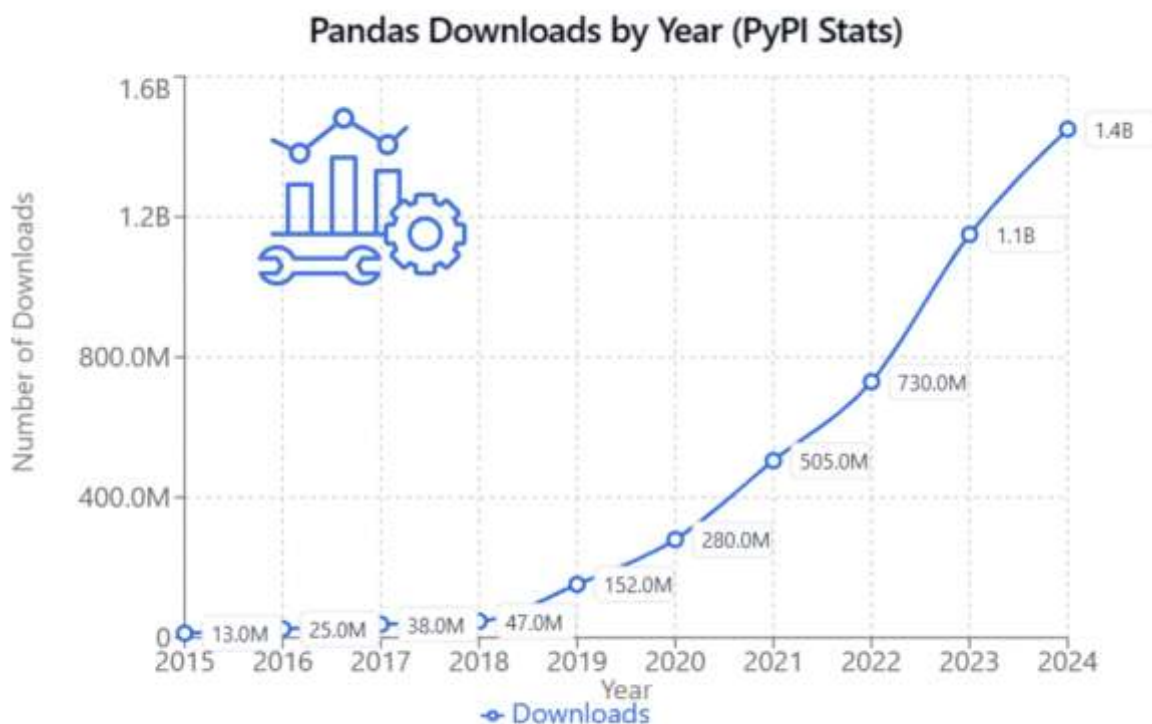
Raamatukogu on nagu valmis tööriistade kogum: funktsioonid, moodulid, klassid. Nii nagu ehitusplatsil ei pea iga kord haamrit või tasandit leiutama, võimaldavad raamatukogud programmeerimises probleeme kiiresti lahendada, ilma et peaksite põhifunktsioone ja -lahendusi uuesti leiutama.

Pandas on avatud lähtekoodiga Pythoni raamatukogu, mis pakub suure jõudlusega ja intuitiivseid andmestruktuure, eelkõige DataFrame, mis on universaalne formaat tabelitega töötamiseks. Pandas on Šveitsi nuga analüütikutele, inseneridele ja arendajatele, kes töötavad andmetega.

Python on lihtsa süntaksiga kõrgetasemeline programmeerimiskeel, mida kasutatakse aktiivselt analüüsi, automatiseerimise, masinõppe ja veebiarenduse valdkonnas. Selle populaarsus tuleneb selle koodi loetavusest, platvormideülesusest ja rikkalikust raamatukogude ökosüsteemist. Praeguseks on Pythonile loodud üle 137 000 avatud lähtekoodiga paketi [87] ja see arv kasvab peaaegu iga päev. Iga selline raamatukogu on omamoodi valmis funktsioonide hoidla: alates lihtsatest matemaatilistest operatsioonidest kuni keeruliste tööriistadeni pilditöötamiseks, suurandmete analüüsiks, neuronivõrkudeks ja integreerimiseks väliste teenustega.

Teisisõnu, kujutage ette, et teil on vaba ja avatud juurdepääs sadadele tuhandetele valmis tarkvaralahendustele - raamatukogudele ja tööriistadele, mida saate otse oma äriprotsessidesse integreerida. See on justkui tohtu kataloog rakendustest automatiseerimiseks, analüüsiks, visualiseerimiseks, integreerimiseks ja muuks - ja kõik need on kohe pärast Pythoni installimist kättesaadavad.

Pandas on üks populaarsemaid pakette Pythoni ökosüsteemis. Aastal 2022 ulatus Pandase raamatukogu keskmine allalaadimiste arv 4 miljonini päevas (joonis 3.4-3), 2025. aasta alguseks on see arv tõusnud 12 miljoni allalaadimiseni päevas, mis peegeldab selle kasvavat populaarsust ja laialdast kasutamist andmeanalüütikas ja LLM-vestluses [86]

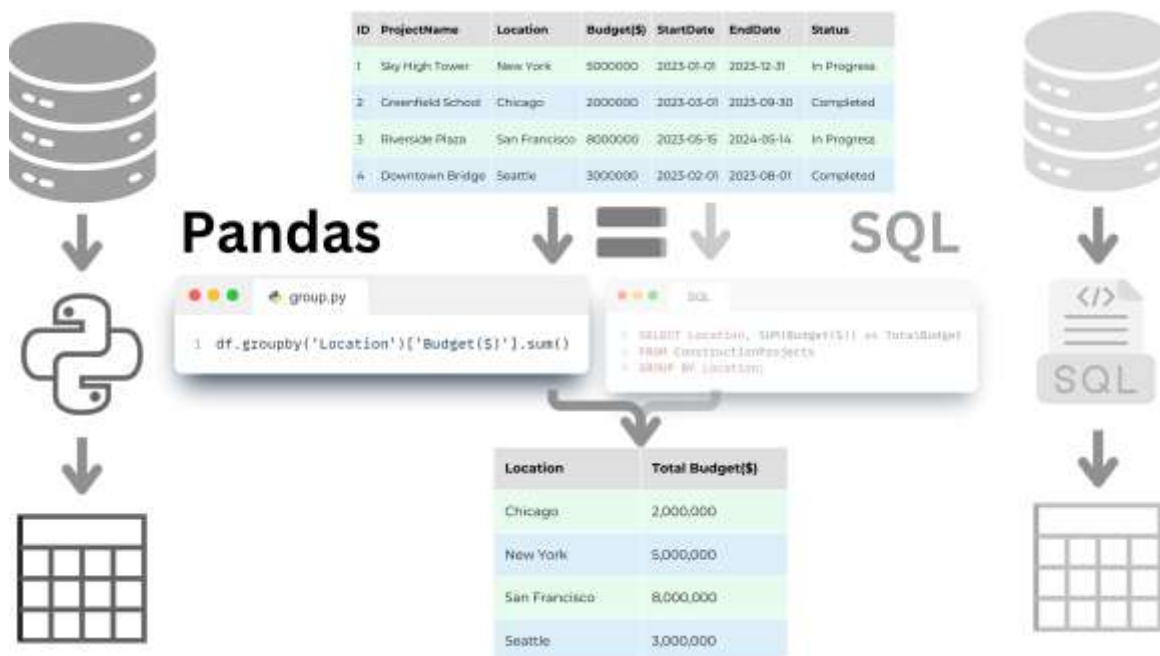


Joonis 3.4-3 Pandas on üks enim allalaaditud raamatukogusid. Aastal 2024 ületas selle aastane allalaadimiste arv 1,4 miljardit.

Pandase raamatukogu päringukeel on funktsionaalsuselt sarnane SQL päringukeelele, mida me käsitlesime peatükis "Relatsioonilised andmebaasid ja SQL päringukeel".

Analüütika ja struktureeritud andmete haldamise maailmas paistab Pandas silma oma lihtsuse, kiiruse ja võimsuse poolest, pakkudes kasutajatele laia valikut vahendeid teabe tõhusaks analüüsiks ja töötlemiseks.

Mõlemad tööriistad - SQL ja Pandas - pakuvad võimsaid andmemajandamisvõimalusi, eriti võrreldes traditsioonilise Exceliga. Nad toetavad selliseid toiminguid nagu valimine, filtreerimine (joonis 3.4-4), ainus erinevus on see, et SQL on optimeeritud tööks relatsiooniliste andmebaasidega, samas kui Pandas töötleb andmeid RAMis, mis võimaldab seda kasutada mis tahes arvutis, ilma et oleks vaja luua andmebaase ja kasutada eraldi infrastruktuuri.

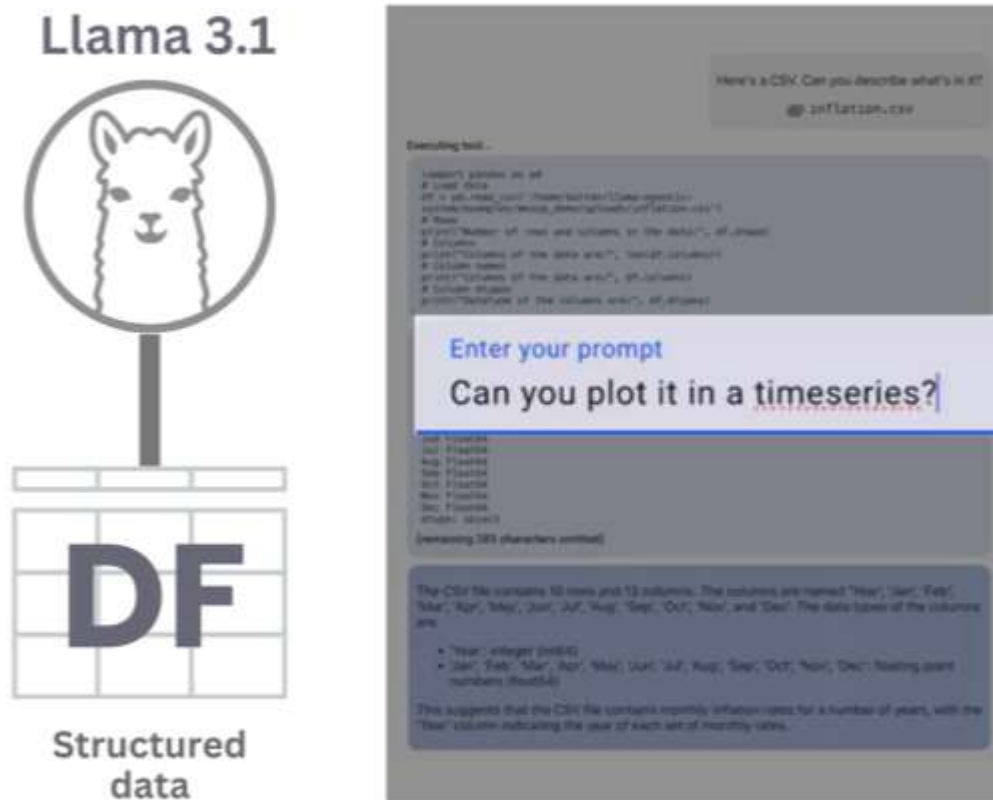


Joonis 3.4-4 Pandas on erinevalt SQL paindlik, et töötada erinevate andmeformaadidega, mis ei piirdu ainult andmebaasidega.

Pandas on sageli eelistatud teadusuuringuteks, protsesside automatiseerimiseks, torujuhtmete loomiseks (sealhulgas ETL) ja andmete manipuleerimiseks Pythonis, samas kui SQL on andmebaaside haldamise standard ja seda kasutatakse sageli ettevõtluskeskkondades suurte andmemahutude töötlemiseks.

Pythoni programmeerimiskeele Pandas raamatukogu võimaldab teil teha mitte ainult põhilisi toiminguid, nagu tabelite lugemine ja kirjutamine, vaid ka keerulisemaid ülesandeid, sealhulgas andmete ühendamine, andmete rühmitamine ja keeruliste analüütiliste arvutuste tegemine.

Tänapäeval kasutatakse Pandas raamatukogu mitte ainult akadeemilistes uuringutes ja ärianalüüsis, vaid ka koos LLM -mudelitega. Näiteks Meta® divisjon (Facebook™) pööras 2024. aastal uue avatud lähtekoodiga mudeli LLaMa 3.1 avaldamisel erilist tähelepanu tööle struktureeritud andmetega, tehes selle avaldamise üheks võtme- ja esimeseks juhtumiks just CSV-vormingus struktureeritud andmekaadrite (joonis 3.4-5) töötlemise ja integreerimise Pandase raamatukoguga otse chatis.



Joonis 3.4-5 Üks esimesi ja peamisi 2024. aastal LLaMa 3.1 tutvustatud Meta meeskonna juhtumeid oli rakenduste loomine Pandase abil.

Pandas on miljonite andmeteadlaste jaoks hädavajalik vahend, mis töötleb ja valmistab andmeid ette genereeriva tehisintellekti jaoks. Pandase kiirendamine ilma koodi muutmata on suur samm edasi. Andmeteadlased saavad töödelda andmeid tundide asemel minutite jooksul ja saada suurusjärgu võrra rohkem andmeid, et treenida genereerivaid tehisintellekti mudeleid [88].

- Jensen Huang, NVIDIA asutaja ja tegevjuht.

Kasutades Pandas, on võimalik hallata ja analüüsida andmekogumeid, mis ületavad Exceli võimalusi. Kui Excel suudab tavaliselt töödelda kuni 1 miljon rida andmeid, siis Pandas saab hõlpsasti hakkama andmekogumitega (joonis 9.1-2, joonis 9.1-10), mis sisaldavad kümneid miljoneid ridu [89]. See võime võimaldab kasutajatel teostada keerukat andmeanalüüsi ja visualiseerimist suurte andmekogumite kohta, andes sügavamaid teadmisi ja hõlbustades andmepõhist otsustamist. Lisaks on Pandas'il tugev kogukonna toetus [90]: sadu miljoneid arendajaid ja analüütikuid kogu maailmas (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) kasutavad seda iga päev online või offline, pakkudes suurt hulka valmislahendusi mis tahes äriprobleemide lahendamiseks.

Enamiku Pythoni analüütiliste protsesside keskmes on andmete struktureeritud vorm nimega

DataFrame, mida pakub Pandas raamatukogu. See on võimas ja paindlik vahend tabeliandmete korrastamiseks, analüüsimiseks ja visualiseerimiseks.

DataFrame: universaalne tabeliandmete formaat

DataFrame on Pandase raamatukogu keskne struktuur, mis on kahemõõtmeline tabel (joonis 3.4-6), kus read vastavad üksikutele objektidele või kirjetele ja veerud nende omadustele, parameetritele või kategooriatele. See struktuur meenutab visuaalselt Exceli tabelit, kuid on paindlikkuse, skaleeritavuse ja funktsionaalsuse poolest palju parem.

DataFrame on viis, kuidas kujutada ja töödelda tabeli kujulisi andmeid, mis on salvestatud arvuti RAM-i.

DataFrame on viis, kuidas kujutada ja töödelda tabeli kujulisi andmeid, mis on salvestatud arvuti RAM-i. Tabelis võivad read kajastada näiteks ehitusprojekti elemente ja veerud nende omadusi: kategooriad, mõõtmed, koordinaadid, maksumus, tingimused jne. Lisaks võib selline tabel sisaldada nii teavet ühe projekti kohta (joonis 4.1-13) kui ka andmeid miljonite objektide kohta tuhandetest erinevatest projektidest (joonis 9.1-10). Tänu Pandase vektoriseeritud operatsioonidele on selliseid andmemahte lihtne filtreerida, rühmitada ja koondada suure kiirusega.

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

Joonis 3.4-6 Ehitusprojekt kui andmeruut on kahemõõtmeline tabel, mille elemendid on ridades ja atribuudid veergudes.

Nvidia hinnangul kasutatakse juba praegu kuni 30 protsenti kõigist arvutiressurssidest struktureeritud andmete - andmekaadrite - töötlemiseks ja see osakaal kasvab jätkuvalt.

Andmetöötlus on see, millega igas ettevõttes tehakse tõenäoliselt kolmandik maailma arvutitööstusest. Enamiku ettevõtete andmetöötlus ja andmed on DataFrame'is, tabeliformaadis

| - Jensen Huang, Nvidia tegevjuht [91]

Loetleme mõned peamised omadused DataFrame Pandas:

- **Veerud:** DataFrame'is on andmed organiseeritud veergudeks, millel on unikaalne nimi. Atribuutide veerud võivad sisaldada eri tüüpi andmeid, sarnaselt andmebaaside veergudele või tabelite veergudele.
- **Pandas Seeria** on Pandas'e ühemõõtmeline andmestruktuur, mis sarnaneb tabeli loetelule või veerule, kus iga väärtus vastab erinevale indeksile.

Pandas Series on üle 400 atribuudi ja meetodi, mis muudab andmetega töötamise uskumatult paindlikuks. Saate veerule otse rakendada ühte neljasajast olemasolevast funktsioonist, teha matemaatilisi operatsioone, filtreerida andmeid, asendada väärtusi, töötada kuupäevade, stringidega ja palju muud. Lisaks toetab Series vektoriseeritud operatsioone, mis kiirendab suurte andmekogumite töötlemist võrreldes tsükliliste arvutustega oluliselt. Näiteks saate hõlpsasti korrutada kõik väärtused arvuga, asendada puuduvad andmed või rakendada keerulisi teisendusi ilma keerulisi tsükleid kirjutamata.

- **Rows:** DataFrame'is saab indekseerida unikaalsete väärtustega. See indeks võimaldab kiiresti muuta ja kohandada andmeid konkreetsetes ridades.
- **Indeks:** Vaikimisi määrab Pandas DataFrame'i loomisel igale reale indeksi vahemikus 0 kuni N-1 (kus N on kõikide ridade arv DataFrame'is). Indeksit saab siiski muuta, et lisada erimärgistused, näiteks kuupäevad või unikaalsed tunnused.
- Ridade **indekseerimine** DataFrame'is tähendab, et igale reale määratakse unikaalne nimi või märgis, mida nimetatakse DataFrame'i indeksiks.
- **Andmetüübid:** DataFrame toetab erinevaid andmetüüpe, sealhulgas: Int, Float, Bool, Datetime64 ja tekstandmete jaoks object. Igal DataFrame'i veerul on oma andmetüüp, mis määrab, milliseid operatsioone saab selle sisuga teha.
- **Andmeoperatsioonid:** DataFrame toetab mitmesuguseid andmetöötlusoperatsioone, sealhulgas liitmist (`groupby`), ühendamist (`merge` ja `join`), ühendamist (`concat`), jagamist-rakendamist-kombineerimist ja paljusid muid andmetöötlustehnikaid.
- **Suuruse manipuleerimine:** DataFrame võimaldab lisada ja eemaldada veerge ja ridu, muutes selle dünaamiliseks struktuuriks, mida saab muuta vastavalt teie andmeanalüüsi vajadustele.
- **andmete visualiseerimine:** kasutades sisseehitatud visualiseerimistehnikaid või suheldes populaarsete andmete visualiseerimise raamatukogudega, nagu Matplotlib või Seaborn, saab DataFrame hõlpsasti teisendada graafikuteks ja diagrammideks, et andmeid graafiliselt esitada.
- **Andmete sisestamine ja väljastamine:** Pandas pakub funktsioone andmete importimiseks ja eksportimiseks erinevatesse failivormingutesse, nagu CSV, Excel, JSON, HTML ja SQL, mis võib muuta DataFrame andmete kogumise ja levitamise keskseks sõlmpunktiks.

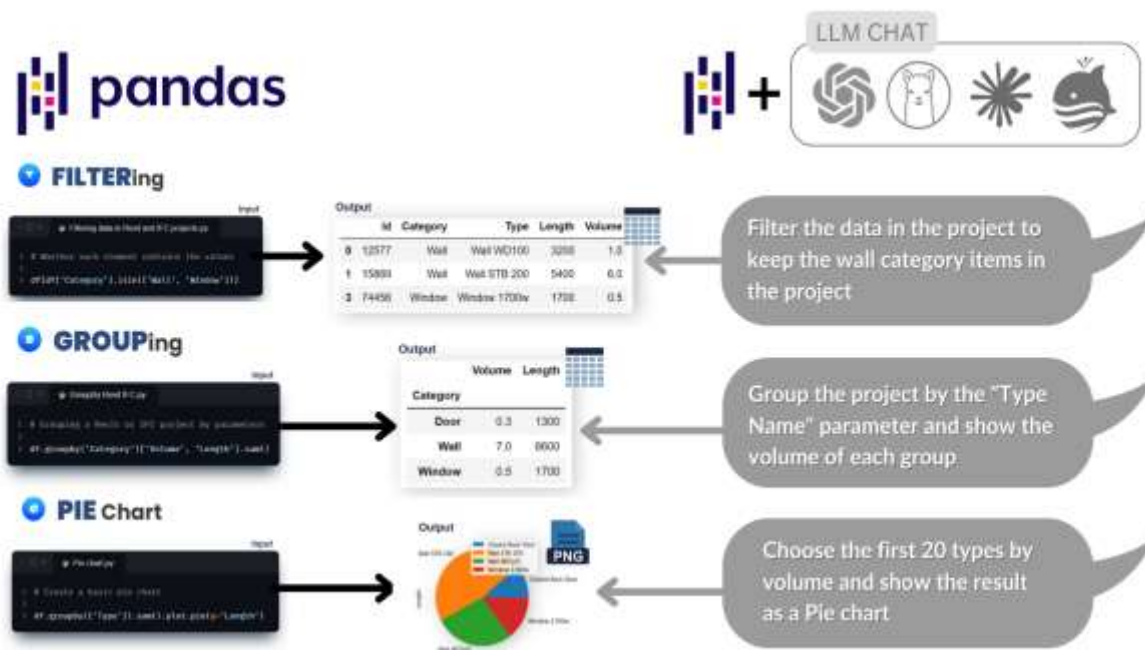
Erinevalt CSV ja XLSX pakub Pandas DataFrame andmetega töötamisel suuremat paindlikkust ja jõudlust: see suudab töödelda suuri andmehulki RAM-is, toetab laiendatud andmetüüpe (sh

kuupäevad, loogilised väärtused ja aegread) ning pakub ulatuslikke võimalusi andmete filtreerimiseks, koondamiseks, ühendamiseks ja visualiseerimiseks. Kuigi CSV ei salvesta teavet andmetüüpide ja -struktuuri kohta ning XLSX on sageli ülekoormatud vormindamisega ja madala skaleeritavusega, on DataFrame endiselt optimaalne valik kiireks analüüsiks, protsesside automatiseerimiseks ja integreerimiseks tehisintellekti -mudelitega (joonis 3.4-7). Järgnevates peatükkides käsitletakse üksikasjalikult kõiki neid andmete aspekte, samuti käsitletakse raamatu 8. osas üksikasjalikult sarnaseid formaate nagu Parquet, Apache Orc, JSON, Feather, HDF5 ja andmelaod (joonis 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)

Joonis 3.4-7 DataFrame on optimaalne valik suure jõudluse ja täiustatud andmetüüpide toega andmemajanduseks.

Tänu nende paindlikkusele, võimsusele ja kasutusmugavusele on Pandase raamatukogu ja DataFrame'i formaat muutunud Pythoni andmeanalüüsi de facto standardiks. Need on ideaalsed nii lihtsate aruannete loomiseks kui ka keerukate analüütiliste torujuhtmete loomiseks, eriti koos LLM-mudelitega.



Joonis 3.4-8 LLM lihtsustab Pandasega suhtlemist: koodi asemel piisab tekstpäringust.

Tänapäeval kasutatakse Pandast aktiivselt LLM-põhistes jututubades, nagu ChatGPT, LLaMa, DeepSeek, QWEN ja teised. Paljudel juhtudel, kui mudel saab päringu, mis on seotud tabelitöötluste, andmete valideerimise või analüüsiga, genereerib ta koodi täpselt Pandase raamatukogu abil. See teeb DataFrame'ist loomuliku "keele" andmete esitamiseks tehisintellektiga peetavates dialoogides (joonis 3.4-8).

Kaasaegsed andmetehnoloogiad, nagu Pandas, lihtsustavad andmete analüüsimist, automatiseerimist ja integreerimist äriprotsessidesse. Need annavad kiiresti tulemusi, vähendavad spetsialistide töökoormust ja tagavad korduvad toimingud.

Järgmised sammud: jätkusuutliku andmeraamistiku loomine

Selles osas vaatlesime ehitustööstuses kasutatavaid peamisi andmetüüpe, õppisime tundma nende salvestamise erinevaid formaate ja analüüsisime kaasaegsete vahendite, sealhulgas LLM ja IDEde rolli teabe töötlemisel. Õppisime, et tõhus andmehaldus on teadliku otsustamise ja äriprotsesside automatiseerimise alus. Organisatsioonid, kes suudavad oma andmeid struktureerida ja korrastada, saavad märkimisväärse konkurentsieelise andmete töötlemise ja ümberkujundamise etappides.

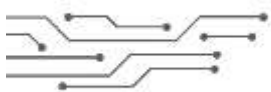
Selle osa kokkuvõtteks tasub välja tuua peamised praktilised sammud, mis aitavad teil rakendada käsitletud lähenemisviise oma igapäevastes ülesannetes:

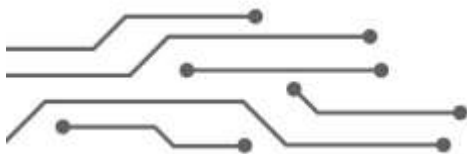
- Viige läbi oma protsesside andmeaudit
- Tehke inventuur kõigist andmetüüpidest, mida te oma projektides kasutate.

- ☐ Määrake kindlaks, millised andmetüübid ja -mudelid on teie äriprotsesside jaoks kõige kriitilisemad.
- ☐ Määrata kindlaks probleemsed valdkonnad, kus teave jääb sageli struktureerimata, halvasti struktureeritud või kättesaamatuks.
- Alustage andmehaldusstrateegia koostamist
 - ☐ tõstatada poliitikaküsimusi ja standardeid erinevate andmetüüpide käsitlemiseks.
 - ☐ Analüüsige, milliseid teie töövooge saab parandada, kui muuta struktureerimata andmed struktureeritud andmeteks.
 - ☐ Luua andmete säilitamise ja juurdepääsu poliitika, mis võtab arvesse turvalisust ja konfidentsiaalsust.
- Andmetega töötamise põhitööriistade paigaldamine ja omandamine
 - ☐ Valige sobiv IDE, mis vastab teie ülesannetele (nt installige VS Code või Jupyter Notebook).
 - ☐ Proovige luua kohalik LLM, et käsitleda teie isikuandmeid konfidentsiaalselt.
 - ☐ Alustage eksperimenteerimist Pandase raamatukoguga, et töödelda XLSX-tabeliandmeid.
 - ☐ Kirjeldage LLMile tüüpilisi ülesandeid, millega te tegelete tabelarvutusprogrammides või andmebaasides, ja paluge LLMil automatiseerida töö Pandase abil.

Nende sammude rakendamine võimaldab teil järk-järgult muuta oma lähenemist andmetega töötamisele, liikudes eraldiseisvatest, struktureerimata andmekogumitest ühtsesse ökosüsteemi, kus andmed muutuvad kättesaadavaks ja arusaadavaks varaks. Alustage väikselt - looge oma esimene DataFrame Pandas, käivitage kohalik LLM, automatiseerige oma esimene rutiinne ülesanne Pythoni abil (nt Exceli tabelid).

Raamatu neljandas osas keskendutakse andmete kvaliteedile, andmete korraldamisele, struktureerimisele ja modelleerimisele. Keskendume metoodikatele, mis muudavad erinevad andmeallikad - alates PDF-dokumentidest ja tekstidest kuni piltide ja CAD-mudeliteni - struktureeritud andmekogumiteks, mis sobivad analüüsiks ja automatiseerimiseks. Samuti uurime, kuidas vormistatakse andmenõudeid, kuidas ehitatakse kontseptuaalseid ja loogilisi mudeleid ehitusprojektides ning kuidas kaasaegsed keelemudelid (LLM) võivad selles protsessis abiks olla.





IV OSA

ANDMETE KVALITEET: KORRALDUS, STRUKTUREERIMINE, MODELLEERIMINE

Neljandas osas keskendutakse metoodikatele ja tehnoloogiatele, mille abil muudetakse ebaühtlane teave kvaliteetseteks struktureeritud andmekogumiteks. Üksikasjalikult käsitletakse andmevajaduste kujundamise ja dokumenteerimise protsesse, mis on aluseks tõhusale teabearhitektuurile ehitusprojektides. Esitatakse praktilised meetodid struktureeritud teabe eraldamiseks erinevatest allikatest (PDF -dokumendid, pildid, tekstifailid, CAD - mudelid) koos rakendamisnäidetega. Analüüsitakse regulaaravaldiste (RegEx) ja muude vahendite kasutamist andmete automaatseks valideerimiseks ja kontrollimiseks. Andmete modelleerimise protsessi kontseptuaalsel, loogilisel ja füüsilisel tasandil kirjeldatakse samm-sammult, võttes arvesse ehitussektori eripärasid. Näidatakse konkreetseid näiteid keelemudelite (LLM) kasutamise kohta teabe struktureerimise ja valideerimise protsesside automatiseerimiseks. Tehakse ettepanekuid tõhusate lähenemisviiside kohta analüüsitulemuste visualiseerimiseks, mis suurendab analüütilise teabe kättesaadavust ehitusprojektide juhtimise kõigil tasanditel

PEATÜKK 4.1.

ANDMETE TEISENDAMINE STRUKTUREERITUD KUJUL

Andmepõhise majanduse ajastul on andmed muutumas pigem otsuste tegemise aluseks kui takistuseks. Selle asemel, et pidevalt kohandada teavet igale uuele süsteemile ja selle vormingutele, püüavad ettevõtted üha enam moodustada ühtset struktureeritud andmemudelit, mis oleks universaalne tõeliskas kõigi protsesside jaoks. Kaasaegsed infosüsteemid ei ole loodud mitte formaatide ja liideste, vaid andmete tähenduse järgi - sest struktuur võib küll muutuda, kuid teabe tähendus jääb palju kauem samaks.

Andmetega tõhusa töötamise võti ei ole nende lõputu teisendamine ja ümberkujundamine, vaid nende õige korraldamine algusest peale: universaalse struktuuri loomine, mis võimaldab läbipaistvust, automatiseerimist ja integreerimist projekti elutsükli kõikides etappides.

Traditsiooniline lähenemisviis sunnib iga uue platvormi rakendamisega tegema käsitsi kohandusi: andmete migreerimine, atribuutide nimede muutmine, vormingute kohandamine. Need sammud ei paranda andmete kvaliteeti, vaid ainult varjavad probleeme, tekitades lõputute ümberkujundamiste nõiaringi. Selle tulemusena muutuvad ettevõtted sõltuvaks konkreetsetest tarkvaralahendustest ja digitaalne ümberkujundamine aeglustub.

Järgmistes peatükkides vaatleme, kuidas andmeid õigesti struktureerida ja seejärel kuidas luua universaalseid mudeleid, minimeerida platvormisõltuvust ja keskenduda sellele, mis on kõige olulisem - andmed kui strateegiline ressurss, mille ümber jätkusuutlikud protsessid ehitatakse.

Õppimine, kuidas muuta dokumendid, PDF, pildid ja tekstid struktureeritud vorminguteks.

Ehitusprojektide puhul on valdav osa teabest olemas struktureerimata kujul: tehnilised dokumendid, tööde kirjeldused, joonised, spetsifikatsioonid, ajakavad, protokollid. Nende mitmekesisus - nii formaadi kui ka sisu poolest - muudab integreerimise ja automatiseerimise keeruliseks.

Struktureeritud või poolstruktureeritud vormingutesse konverteerimise protsess võib erineda sõltuvalt sisendandmete tüübist ja soovitud töötlemistulemustest.

Struktureerimata andmete muutmine struktureeritud vormilt struktureeritud vormile on nii kunst kui ka teadus. See protsess sõltub lähteandmete tüübist ja analüüsi eesmärgist ning võtab sageli olulise osa andmetöötleja (joonis 3.2-5) ja analüütiku tööst, mille eesmärk on saada puhas ja korrastatud andmestik.



Joonis 4.1-1 Struktureerimata skaneeritud dokumendi teisendamine struktureeritud tabelivormingusse.

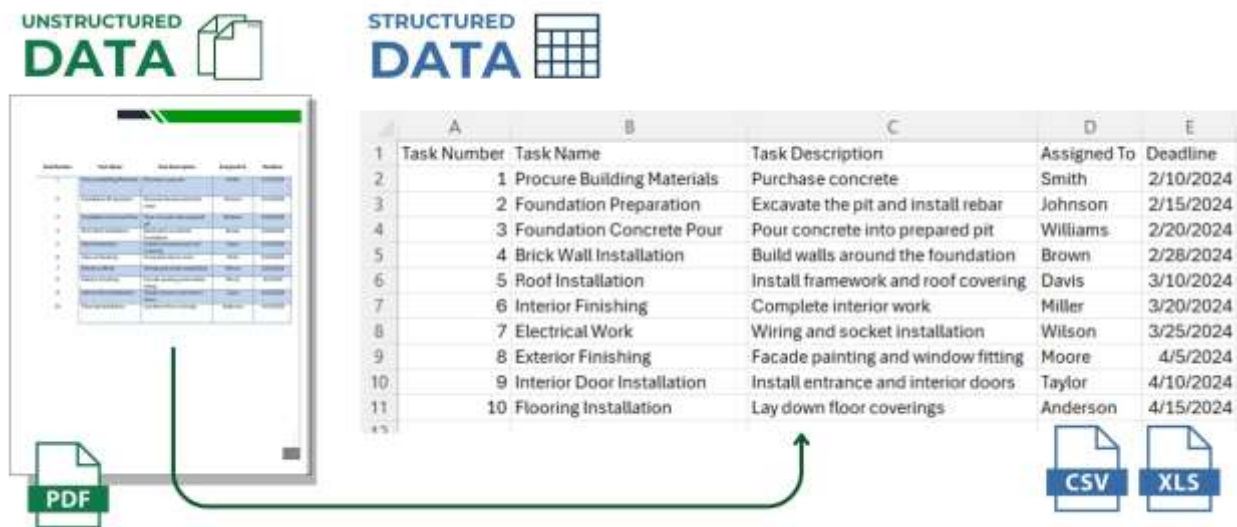
Dokumentide, PDF, piltide ja tekstide muutmine struktureeritud vormingusse (joonis 4.1-1) on järkjärguline protsess, mis hõlmab järgmisi samme:

- **Extract):** Selles etapis laaditakse algdokument või pilt, mis sisaldab struktureerimata andmeid. See võib olla näiteks PDF-dokument -dokument, foto, joonis või skeem.
- **Andmete ümberkujundamine (Transform):** Sellele järgneb samm, mille käigus struktureerimata andmed muudetakse struktureeritud vormingusse. See võib hõlmata näiteks teksti äratundmist ja tõlgendamist piltidelt, kasutades optilist tähemärgituvastust (OCR) või muid töötlemismeetodeid.
- **Andmete laadimine ja salvestamine (Load):** viimane samm hõlmab töödeldud andmete salvestamist edasiseks tööks erinevates vormingutes, näiteks CSV, XLSX, XML, JSON, kusjuures vormingu valik sõltub konkreetsetest nõuetest ja eelistustest.

See protsess, mida nimetatakse ETL (Extract, Transform, Load), mängib automaatses andmetöötlemises võtmerolli ja seda käsitletakse üksikasjalikumalt peatükis "ETL ja Pipeline: Extract, Transform, Load". Järgnevalt vaatleme näiteid selle kohta, kuidas eri formaadis dokumente struktureeritud andmeteks muundatakse.

Näide PDF-dokumendi -dokumendi konverteerimisest tabeliks

Üks kõige tavalisemaid ülesandeid ehitusprojektides on PDF-vormingus spetsifikaatide töötlemine. Struktureerimata andmetest struktureeritud vormingusse ülemineku näitlikustamiseks vaatleme praktilist näidet: PDF-dokumendist tabeli väljavõtmine ja selle teisendamine CSV- või Exceli-vormingusse (joonis 4.1-2).



Joonis 4.1-2 Erinevalt PDF on CSV ja XLSX formaadid laialt levinud ja kergesti integreeritavad erinevatesse andmehaldussüsteemidesse.

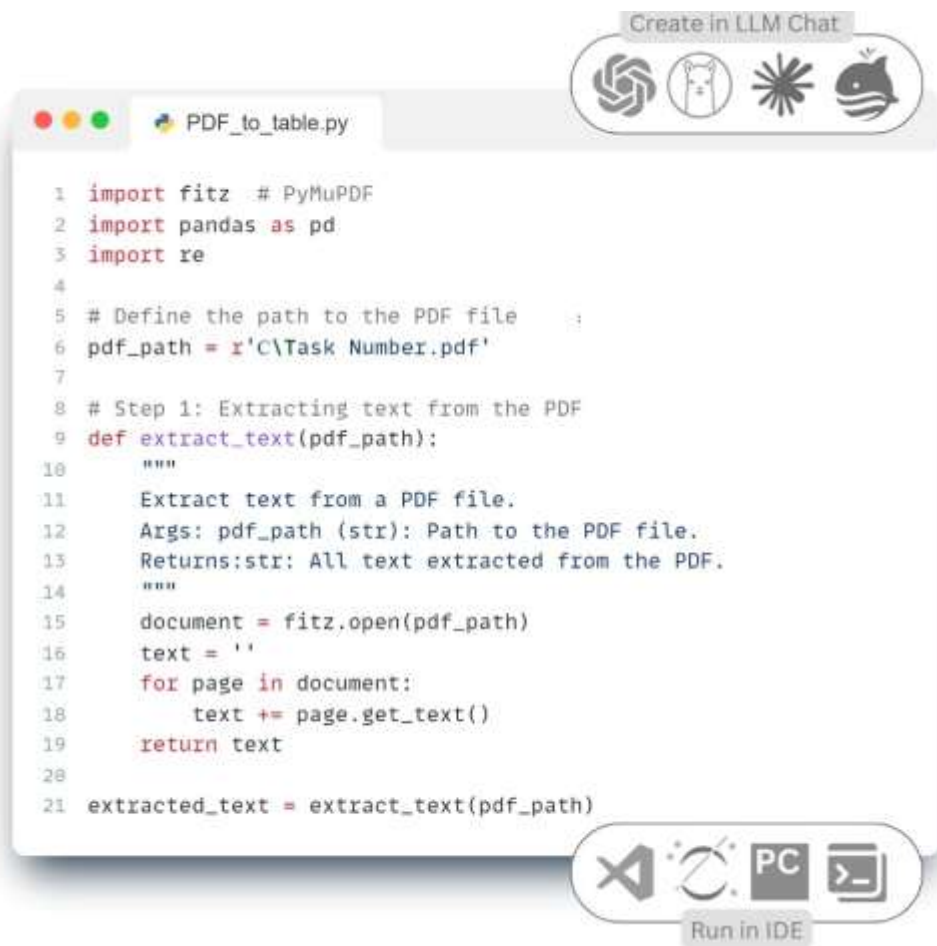
LLM-keelemudelid, nagu ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN lihtsustavad oluliselt andmeteadlaste tööd andmetega, vähendades vajadust programmeerimiskeelte süvaõppe järele ja võimaldades paljusid ülesandeid lahendada tekstiküsimuste abil.

Seega, selle asemel, et kulutada aega lahenduste otsimisele internetis (tavaliselt StackOverflow veebilehel või temaatilistes foorumites ja vestlustes) või pöörduda andmetöötamise spetsialistide poole, saame kasutada kaasaegsete online või kohalike LLMide võimalusi. Piisab, kui esitada päring ja mudel pakub valmis koodi PDF -dokumendi teisendamiseks tabeliformaati.

- 🗣️ Saatke järgmine tekstitaotlus mis tahes LLM -mudelile (CHATGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN või mõni muu):

Palun kirjutage kood teksti väljavõtmiseks PDF-failist -faili, mis sisaldab tabelit. Kood peaks võtma argumendina faili teekonda ja tagastama ekstraheeritud tabeli DataFrame'ina📄

- LLM -mudeli vastus on enamasti Python-koodina, kuna seda keelt kasutatakse laialdaselt andmetööluseks, automatiseerimiseks ja erinevate failivormingutega töötamiseks:



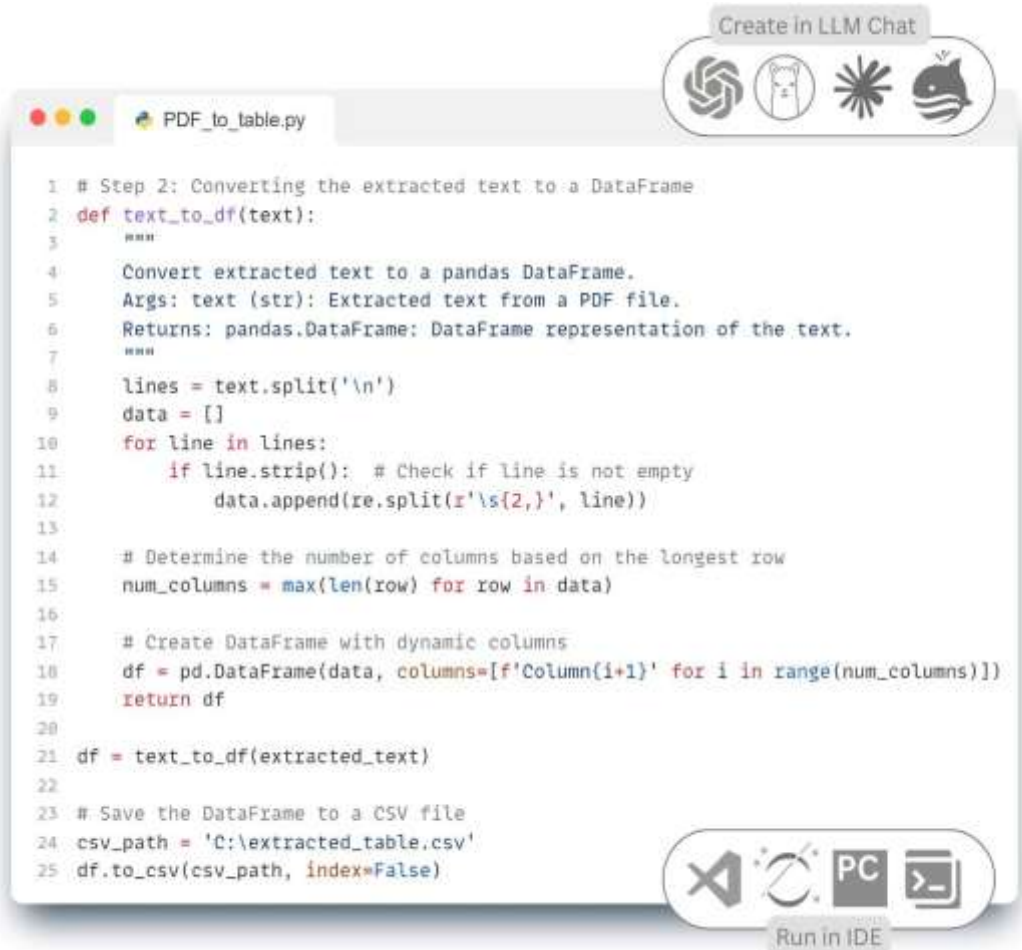
Joonis 4.1-3 LLM vastus Pythoni koodi kujul ja selle raamatukogud ja paketid (Pandas, Fitz) ekstraheerivad teksti PDF -failist.

Seda koodi (joonis 4.1-3) saab käivitada offline ühes eespool mainitud populaarsetest IDE-dest: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse koos PyDev pluginaga, Thonny, Wing IDE, IntelliJ IDEA koos Python pluginaga, JupyterLab või populaarsed online vahendid: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Sammus "Konverteerimine" kasutame populaarset Pandase raamatukogu (mida käsitlesime üksikasjalikult peatükis "Python Pandas: asendamatutööriist andmetega töötamiseks"), et lugeda väljavõetud tekst DataFrame'i ja salvestada DataFrame CSV tabelifaili või XLXS:

Mul on vaja koodi, mis teisendab PDF -failist saadud tabeli DataFrame'iks. Samuti lisada kood, et salvestada DataFrame CSV faili. ↵

LLMi vastus :



```

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)
  
```

Joonis 4.1-4 Väljavõetud tabeli konverteerimine PDF-st DataFrame'iks ja tabeli salvestamine CSV -faili.

Kui koodi täitmise ajal tekib viga (joonis 4.1-3, joonis 4.1-4) - nt puuduvate raamatukogude või vale failiposti tee tõttu - saab veateksti lihtsalt koos lähtekoodiga kopeerida ja uuesti esitada LLM -mudelile. Mudel analüüsib veateadet, selgitab, milles probleem seisneb, ja pakub välja parandused või lisameetmed.

Seega muutub suhtlemine AI LLM-iga täielikuks tsükl→test→tagasiside→parandamine - ilma, et oleks vaja sügavaid tehnilisi teadmisi.

Kasutades tavalist tekstiküsimustikku LLM chat ja kümme rida Python'i, mida saame käivitada lokaalselt mis tahes IDEs, teisendasime PDF -dokumendi tabeli kujuliseks CSV-vorminguks, mis erinevalt PDF-dokumendist on kergesti masinloetav ja kiiresti integreeritav mis tahes andmehaldussüsteemi.

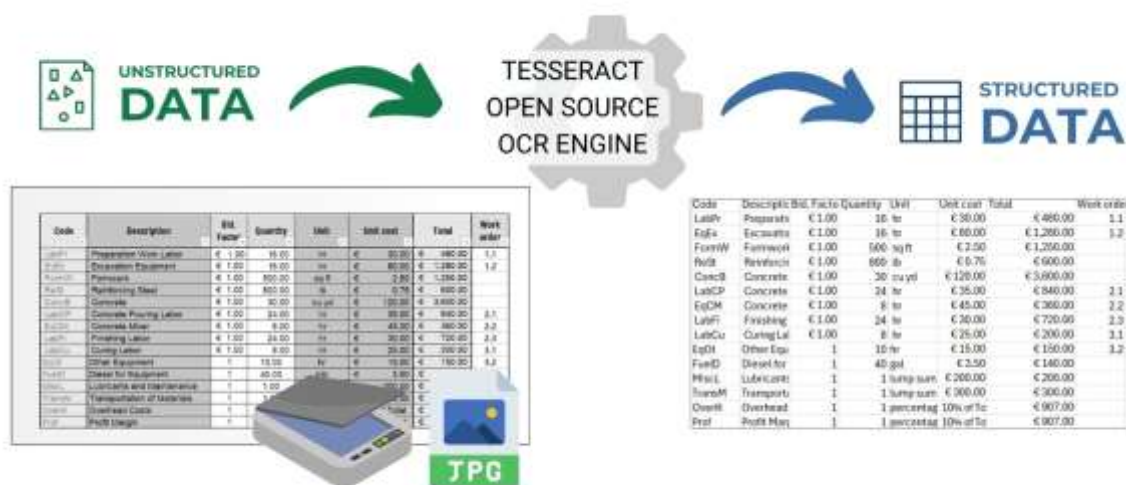
Me võime seda koodi (joonis 4.1-3, joonis 4.1-4) rakendada, kopeerides seda mis tahes LLMi jututuppa, kümnetele või tuhandetele uutele PDF-dokumentidele serveris, automatiseerides sellega struktureerimata dokumentide voo teisendamise protsessi struktureeritud CSV-tabeli formaati.

Kuid PDF-dokumendid ei sisalda alati teksti, enamasti on tegemist skaneeritud dokumentidega, mida tuleb töödelda piltidena. Kuigi pildid on oma olemuselt struktureerimata, võimaldab äratundmisraamatukogude arendamine ja kasutamine nende sisu ekstraheerida, töödelda ja analüüsida, võimaldades meil neid andmeid oma äriprotsessides täielikult ära kasutada.

JPEG, PNG pildi teisendamine struktureeritud kujul

Pildid on üks levinumaid struktureerimata andmete vorme. Ehituses ja paljudes muudes tööstusharudes säilitatakse suur hulk teavet skaneeritud dokumentide, skeemide, fotode ja jooniste kujul. Sellised andmed sisaldavad väärtuslikku teavet, kuid neid ei saa otseselt töödelda, näiteks Exceli tabelis või andmebaasis. Pildid sisaldavad palju keerulist teavet, sest nende sisu, värvid, tekstuurid on erinevad ja kasuliku teabe väljavõtmiseks on vaja spetsiaalset töötlemist.

Piltide kui andmeallikate kasutamisel on raskuseks struktuuride puudumine. Pildid ei anna tähendust edasi otsesel, kergesti mõõdetaval viisil, mida arvuti saab kohe mõista või töödelda, nagu seda teeb Exceli tabel või andmebaasi tabel. Struktureerimata pildiandmete teisendamiseks struktureeritud kujul tuleb kasutada spetsiaalseid raamatukogusid, mis suudavad tõlgendada neis sisalduvat visuaalset teavet (joonis 4.1-5).



Joonis 4.1-5 Skaneeritud dokumentide ja piltide teisendamine struktureeritud formaatidesse on võimalik spetsiaalsete OCR-vahenditega.

OCR (Optical Character Recognition) tehnoloogiat kasutatakse tekstide eraldamiseks piltidelt. See võimaldab tuvastada tähti ja numbreid dokumentide skaneeringutes, fotodes ja PDF -failides, muutes need töödeldavaks ja masinloetavaks tekstiks. OCR-tehnoloogiaid on juba ammu kasutatud dokumentide automatiseerimisel ja tänapäeval on neid lihtne integreerida mis tahes äriprotsessidesse ja Python -rakendustesse. Üks populaarsemaid OCR-vahendeid on Tesseract, avatud lähtekoodiga, mille algselt töötas välja HP™ ja mida nüüd toetab Google™. See toetab üle 100 keele ja on suure äratundmise täpsusega.

Palume LLM chat kirjutada näidiskood, et saada andmeid skaneeritud või pildistatud tabelist struktureeritud viisil.

- 🗨️ Saada tekstitaotlus LLM chat (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN või mõni muu):

Kirjutage kood, et teisendada tabelit sisaldav JPEG pilt DataFrame tabeliks ↩

- LLMi vastus soovib enamasti kasutada Pytesseracti raamatukogu, et tuvastada teksti piltidel :



Joonis 4.1-6 Piltide või fotode tabelist väljavõetud teksti teisendamine struktureeritud tabeli kujuliseks esituseks.

Selles näites - kood (joonis 4.1-6), mis on tuletatud LLM, kasutab pytesseract raamatukogu (Tesseract for Python), et teisendada pilt tekstiks, kasutades OCR (optiline märgituvastus) ja Pandas raamatukogu, et teisendada see tekst struktureeritud kujul, st DataFrame.

Konversiooniprotsess hõlmab tavaliselt eeltöötlust, et parandada pildi kvaliteeti, mille järel rakendatakse erinevaid algoritme mustrite tuvastamiseks, tunnuste eraldamiseks või objektide tuvastamiseks. Selle tulemusena muudetakse struktureerimata visuaalne teave struktureeritud andmeteks.

Kuigi PDF ja pildid on peamised struktureerimata teabe allikad, on tegelik meister mahu poolest e- kirjades, vestlustes, koosolekutel ja sõnumitoojates tekkiv tekst. Neid andmeid ei ole mitte ainult palju, vaid need on ka hajutatud, mitteametlikud ja äärmiselt halvasti struktureeritud.

Tekstiandmete konverteerimine struktureeritud kujul

Lisaks tabelitega PDF-dokumentidele (joonis 4.1-2) ja tabelivormide skaneeritud versioonidele (joonis 4.1-5) on oluline osa projektdokumentatsiooni teabest esitatud tekstina. See võib olla nii tekstidokumentides sisalduvad sidusad laused kui ka joonistele ja skeemidele hajutatud katkendlikud kirjed. Kaasaegsetes andmetöötlustingimustes on üks levinumaid ülesandeid sellise teksti teisendamine struktureeritud vormingusse, mis sobib analüüsiks, visualiseerimiseks ja otsuste tegemiseks.

Selle protsessi keskmes on taksonoomia, mis on klassifitseerimissüsteem, mis võimaldab teabe liigitamist kategooriatesse ja alamkategooriatesse ühiste tunnuste alusel.

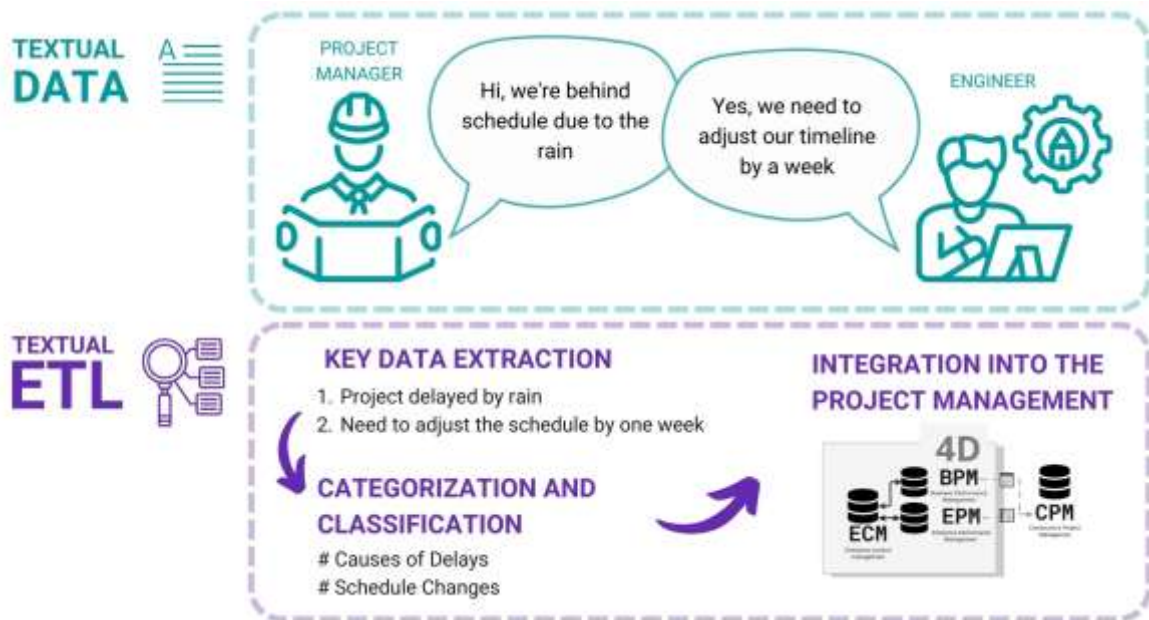
Taksonoomia on hierarhiline klassifitseerimisstruktuur, mida kasutatakse objektide rühmitamiseks ja organiseerimiseks. Tekstitötluse kontekstis on see aluseks objektide süstemaatilisele liigitamisele semantilistesse kategooriatesse, mis lihtsustab analüüsi ja parandab andmetötluse kvaliteeti.

Taksonoomia loomisega kaasnevad üksuste väljavõtte, kategoriseerimise ja kontekstualiseerimise etapid. Tekstiandmetest teabe ekstraheerimise protsessi modelleerimiseks on järgmised sammud sarnased nendega, mida oleme juba rakendanud PDF-dokumentidest andmete struktureerimisel:

- **Extract**): teil on vaja analüüsida tekstiandmeid, et saada teavet viivituste ja muudatuste kohta projekti ajakavas.
- **Kategoriseerimine ja klassifitseerimine** (Transform): saadud teabe kategoriseerimine, nt hilinemiste ja ajakava muudatuste põhjused.
- **Integreerimine** (Load): lõpus valmistame struktureeritud andmed ette integreerimiseks välistesse andmehaldussüsteemidesse.

Mõelgem olukorrale: meil on dialoog projektijuhi ja inseneri vahel, kes arutavad probleeme ajakavaga. Meie eesmärk on eraldada võtmeelemendid (viivituse põhjused, ajakava korrigeerimine) ja esitada need struktureeritud kujul (joonis 4.1-7).

Teostame väljavõtte eeldatavate märksõnade alusel, loome andmete väljavõtte simuleerimiseks DataFrame ja pärast teisendamist uue DataFrame tabeli, mis sisaldab veerge kuupäeva, sündmuse (nt hilinemise põhjus) ja tegevuse (nt ajakava muutus) jaoks.



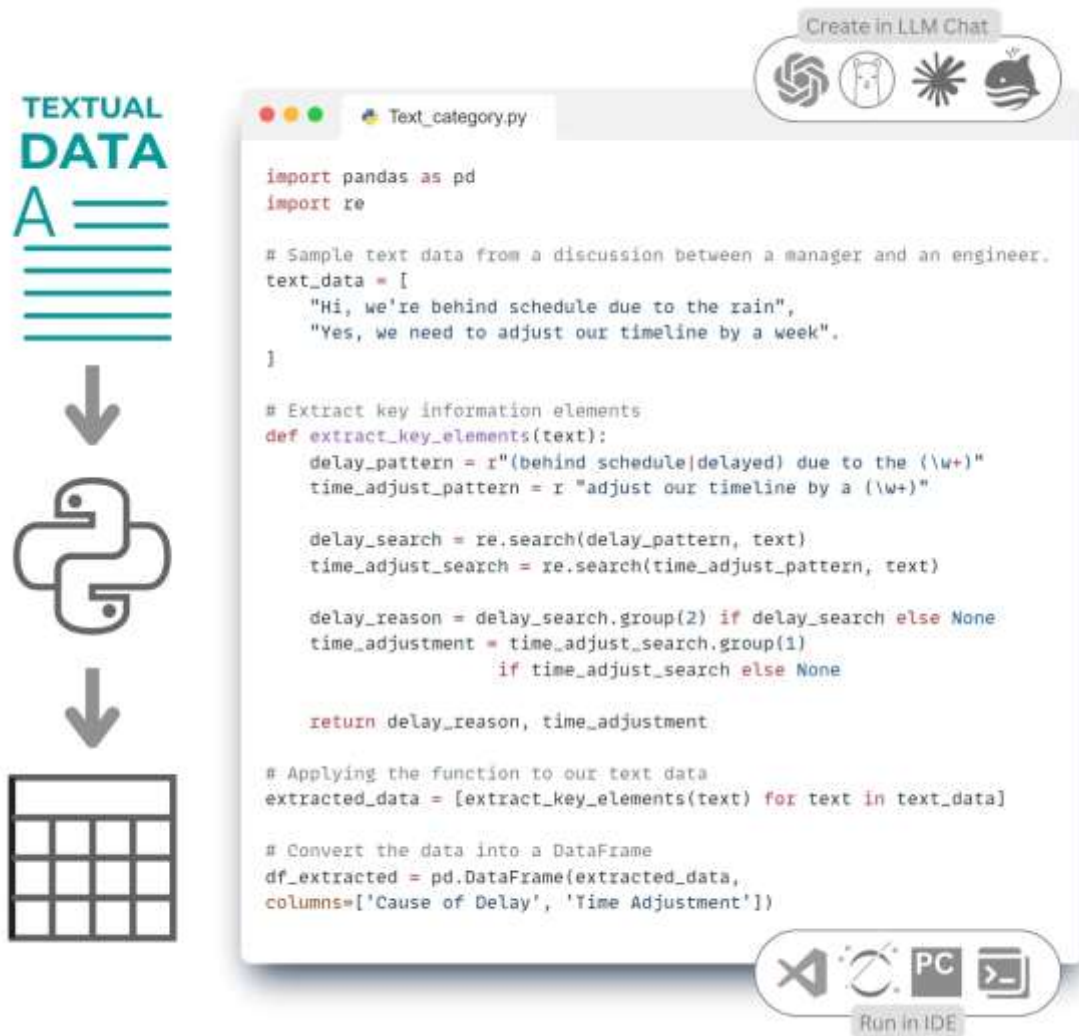
Joonis 4.1-7 Teksti põhiteabe esiletõstmine ajakava kohandamise ja muudatuste projektijuhtimissüsteemi integreerimise vajaduse kohta.

Siin on kood probleemi lahendamiseks, kasutades tekstiküsimustikku ühes keelemudelis nagu eelmistes näidetes.

🗨️ Saatke tekstitaotlus mis tahes LLM jututuppa:

Mul on vestlus juhi vahel: "Tere, me oleme vihma tõttu graafikust maha jäänud" ja inseneri vahel: "Jah, me peame tähtaega nädala võrra korrigeerima". Mul on vaja skripti, mis analüüsiks tulevasi sarnaseid tekstidialooge, võtaks neist välja viivituste põhjused ja vajalikud tähtaja kohandused ning genereeriks nendest andmetest DataFrame'i. DataFrame tuleks seejärel salvestada CSV -faili. ↩

- LLM-i vastus sisaldab tavaliselt Pythoni -koodi, mis kasutab regulaarseid väljendeid (re - Regex) ja Pandas (pd) raamatukogu :

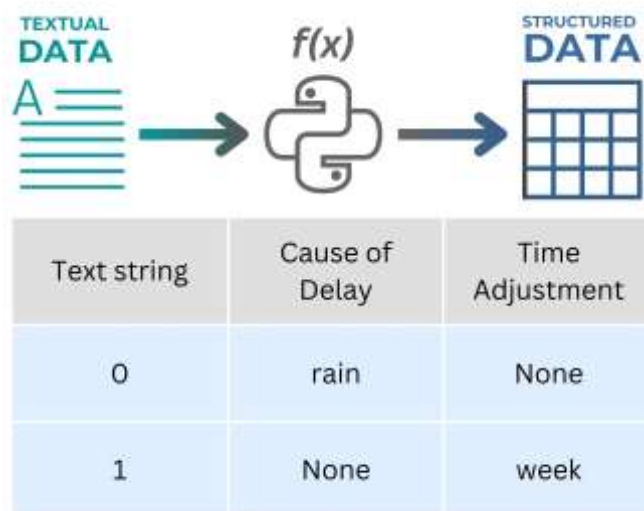


Joonis 4.1-8 Tähtaegade kohandamise vajadust käsitleva teksti põhiteabe esiletõstmine tabelis.

Selles näites (joonis 4.1-7) analüüsitakse projektijuhi ja inseneri vahelist kirjavahetust sisaldavaid tekstiandmeid, et tuvastada ja eraldada konkreetset teavet, mis võib mõjutada tulevaste sarnaste dialoogidega projektide juhtimist. Kasutades regulaaravaldisi (regulaaravaldistest räägime lähemalt peatükis "Struktureeritud nõuded ja RegExi regulaaravaldised"), tuvastatakse mustrite abil projekti hilinemise põhjused ja ajakava vajalikud kohandused. Selles näites kirjutatud funktsioon eraldab mustrite alusel stringidest kas viivituse põhjuse või ajakava kohandamise: valides sõna pärast "because of" kui viivituse põhjuse või sõna pärast "by" kui ajakava kohandamise.

Kui reas mainitakse ilmastiku tõttu tekkinud viivitust, tuvastatakse põhjuseks "vihm"; kui reas mainitakse ajakava korrigeerimist konkreetse perioodi kohta, siis see periood eraldatakse

ajakorrigeerimisena (joonis 4.1-9). Kui reas puudub mõni neist sõnadest, siis on vastava tunnuse veeru väärtus "None".



Joonis 4.1-9 DataFrame'ina saadud kokkuvõttev tabel sisaldab pärast koodi täitmist teavet viivituste olemasolu ja vajalike ajaliste kohanduste kohta.

Tingimuste struktureerimine ja parameetriseerimine tekstist (dialoog, kiri, dokument) võimaldab kiiresti kõrvaldada viivitusi ehitusel: näiteks tööliste puudus võib mõjutada töö tempot halva ilma korral, seega ettevõtte, teades viivituse parameetrid dialoogidest (joonis 4.1-9) ehitusplatsi töödejuhi ja projektijuhi vahel - eelnevalt, saavad ebasoodsa prognoosi korral meeskonda tugevdada.

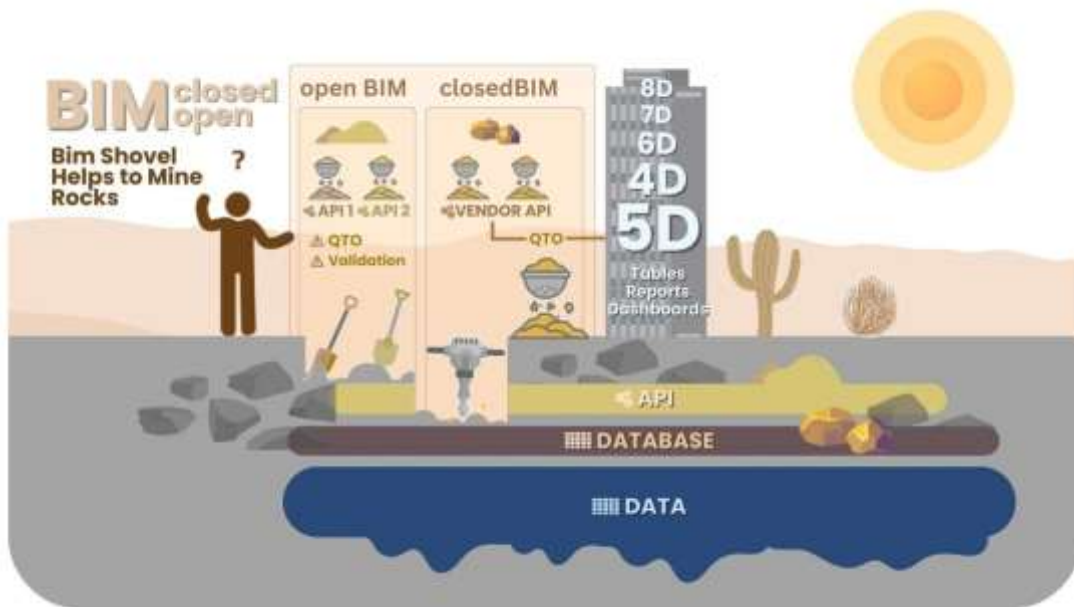
Dokumentide ja piltide teisendamine struktureeritud vormingusse on võimalik suhteliselt lihtsate, avatud ja tasuta kategoriseerimisega seotud vahendite abil.

Elementide kategoriseerimine on samuti oluline osa projektandmetega töötamisel, eriti CAD-tarkvara kasutamisel (BIM).

CAD-andmete (BIM) konverteerimine struktureeritud kujul.

CAD-andmete struktureerimine ja kategoriseerimine (BIM) on keerulisem ülesanne, sest CAD (BIM) andmebaasidest salvestatud andmed on peaaegu alati suletud või keerukates parameetrilistes vormingutes, mis sageli kombineerivad geomeetrilisi andmelemente (poolstruktureeritud) ja metainformatsioonelemente (poolstruktureeritud või struktureeritud andmed) samal ajal.

CAD (BIM) süsteemide algsed andmeformaadid on tavaliselt kaitstud ja otsekasutamiseks kättesaamatud, välja arvatud juhul, kui arendaja ise kasutab spetsiaalset tarkvara või API - liideseid (joonis 4.1-10). Selline andmete isoleerimine moodustab suletud salvestussilosid, mis piiravad vaba teabevahetust ja takistavad ettevõtte läbivate digitaalsete protsesside loomist.



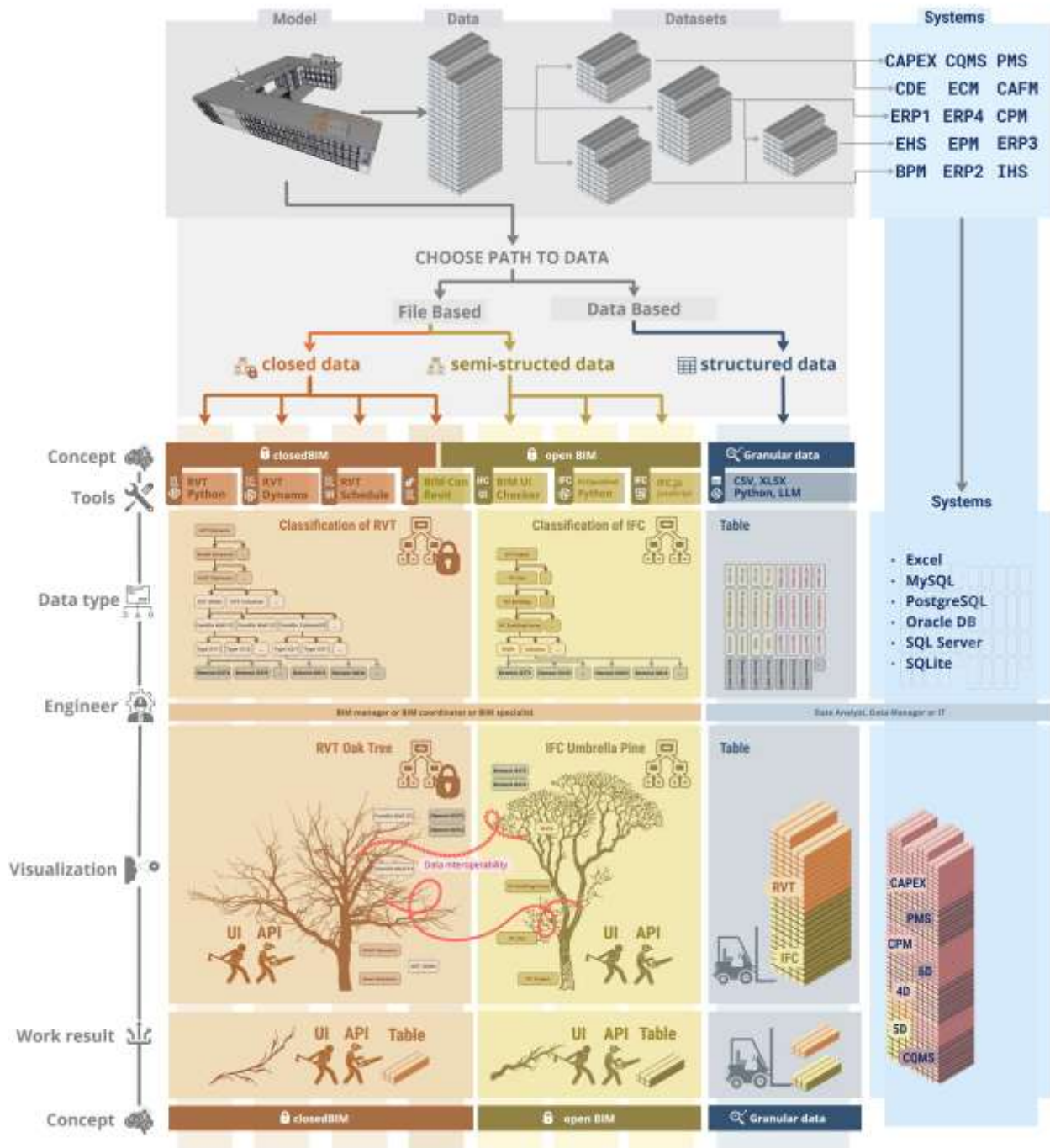
Joonis 4.1-10 CAD-spetsialistid (BIM) saavad ligipääsu algupärastele andmetele API - ühenduste või müüja tööriistade kaudu.

Spetsiaalsetes CAD (BIM) formaatides kogutakse teave projekti elementide omaduste ja atribuutide kohta hierarhilises klassifitseerimissüsteemis, kus vastavate omadustega üksused asuvad nagu viljapuu viljad andmete klassifitseerimise harude kõige uuemates sõlmedes (joonis 4.1-11).

Andmete väljavõtmine sellistest hierarhiatest on võimalik kahel viisil: kas käsitsi, klõpsates igale sõlmpunktile, justkui töödeldes puitu, lõigates valitud kategooriate ja tüüpide oksad kirvega maha. Teise võimalusena pakub rakenduste programmeerimisliidest (API) kasutamine tõhusamat, automatiseeritud lähenemist andmete hankimiseks ja rühmitamiseks, muutes need lõpuks struktureeritud tabeliks, mida saab kasutada teistes süsteemides.

Struktureeritud andmetabelite väljavõtmiseks CAD-projektidest (BIM) saab kasutada erinevaid vahendeid, nagu Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC või avatud lähtekoodiga lahendusi, nagu IfcOpSh või IFCjs IFC-vormingu jaoks.

Kaasaegsed andmeekspordi- ja konverteerimisvahendid võimaldavad andmete töötlemise ja ettevalmistamise lihtsustamiseks jagada CAD-mudelite sisu kaheks põhikomponendiks: geomeetriline teave ja atribuutide andmed (joonis 4.1-13) - metainfo, mis kirjeldab projekteerimiselementide omadusi (joonis 3.1-16). Need kaks andmekihti on omavahel seotud unikaalsete identifikaatorite kaudu, tänu millele on võimalik täpselt seostada iga geomeetrilise kirjeldusega element (parameetrite või hulknurkade kaudu) selle atribuutidega: nimi, materjal, valmidusaste, maksumus jne. Selline lähenemisviis tagab mudeli terviklikkuse ja võimaldab andmete paindlikku kasutamist nii visualiseerimiseks (geomeetrilise mudeli andmed) kui ka analüüsi- või haldusülesannete täitmiseks (struktureeritud või vabalt struktureeritud), töötades nende kahe andmetüübiga eraldi või paralleelselt.



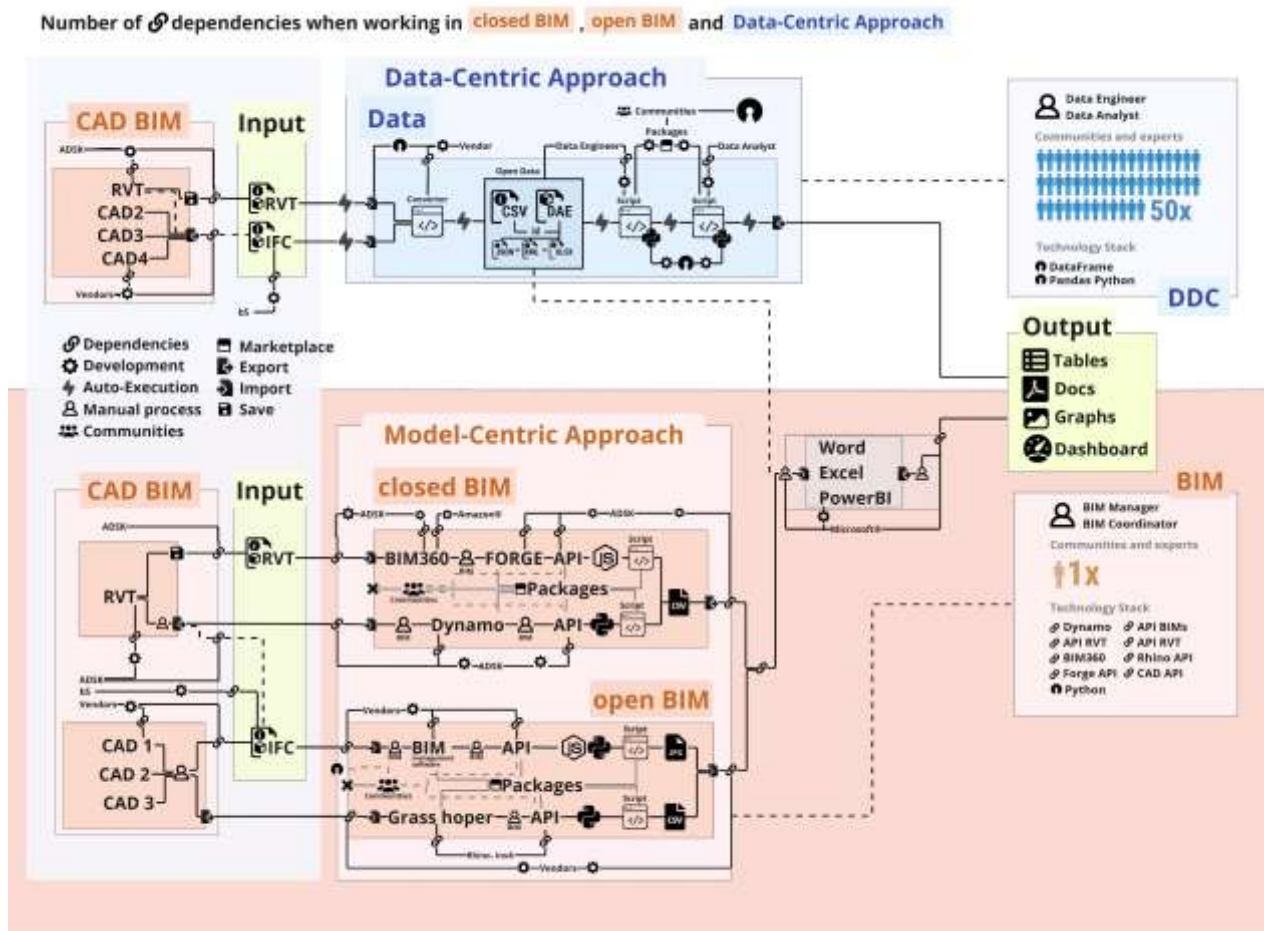
Joonis 4.1-11 CAD-andmebaasidest (BIM) pärineva teabe vaade esitatakse kasutajale klassifitseerimispuude kujul.

Tänu pöördtehnoloogiate arengule ja CAD-andmete konverteerimiseks mõeldud SDKde (Software Development Kit) tulekule on andmete kättesaadavus ja konverteerimine suletud CAD-tarkvara formaatidest (BIM) muutunud palju lihtsamaks. Nüüd on võimalik seaduslikult ja turvaliselt konverteerida andmeid suletud formaatidest universaalsesse formaatidesse, mis sobivad analüüsiks ja kasutamiseks teistes süsteemides. Esimeste pöördprojekteerimisvahendite ("Open

DWG") ajalugu ja võitlust CAD-tootjate formaatide üle domineerimise eest käsitleti peatükis "Struktureeritud andmed: digitaalse ümberkujundamise alus".

Reverse engineering vahendid võimaldavad andmete seaduslikku kättesaamist suletud patenteeritud formaatidest, jaotades teabe segatud CAD (BIM) formaadist kasutajale vajalikeks andmetüüpideks ja formaatideks, mis lihtsustab selle töötlemist ja analüüsimist.

Kasutades pöördprojekteerimist ja otsest juurdepääsu CAD-andmebaaside teabele, teeb teabe kättesaadavaks, võimaldades avatud andmeid ja avatud vahendeid, samuti andmeanalüüsi standardvahendite abil, aruannete koostamist, visualiseerimist ja integreerimist teiste digitaalsüsteemidega (joonis 4.1-12).



Joonis 4.1-12 Otsene juurdepääs CAD-andmetele minimeerib sõltuvust tarkvaraplatvormidest ja liigub andmekeskse lähenemisi suunas.

Alates 1996. aastast DWG-formaadi, alates 2008. aastast DGN-formaadi ja alates 2018. aastast RVT puhul on võimalik pöördprojekteerimisvahendite abil mugavalt ja tõhusalt konverteerida algselt suletud CAD-andmeformaate mis tahes muudesse formaatidesse, sealhulgas struktureeritud formaatidesse (joonis 4.1-13). Tänapäeval kasutavad peaaegu kõik suuremad CAD (BIM) ja suured

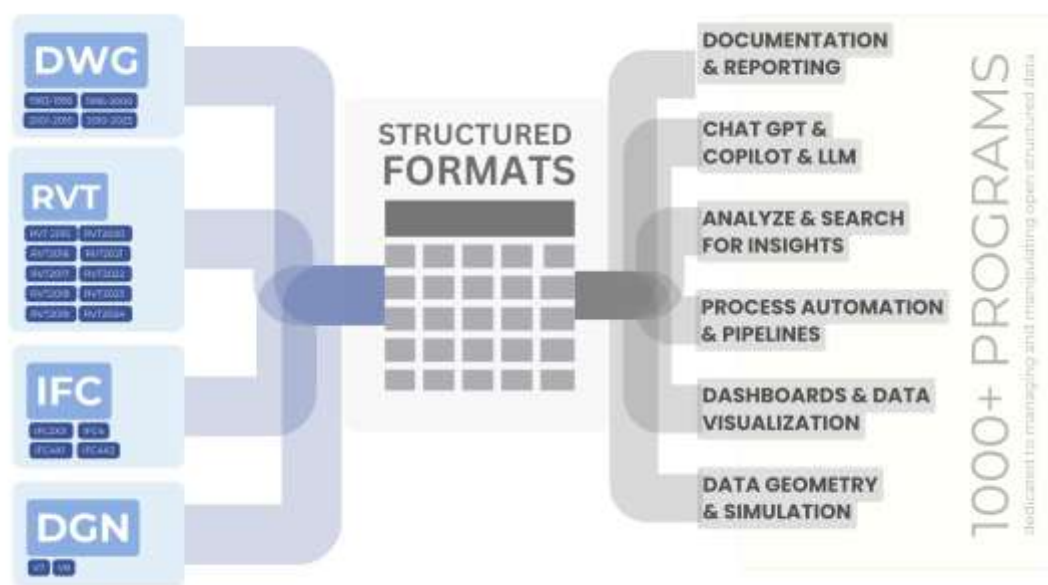
projekteerimisfirmad maailmas SDKsid - pöördprojekteerimise vahendeid, et eraldada andmeid suletud CAD (BIM) tootjate formaatidest [92].

IFC	RVT	DWG	DGN
STRUCTURED DATA	STRUCTURED DATA	STRUCTURED DATA	STRUCTURED DATA
...	...	...	...

Joonis 4.1-13 Tagasipööratud projekteerimisvahendite kasutamine võimaldab CAD (BIM) programmiandmebaasi konverteerida mis tahes sobivaks andmemudeliks.

Andmete konverteerimine suletud, patenteeritud vormingutest avatud vormingutesse ja segatud CAD (BIM) vormingute eraldamine geomeetrilisteks ja metainfo atribuutide andmeteks lihtsustab tööd nendega, muutes need kättesaadavaks analüüsiks, manipuleerimiseks ja integreerimiseks teiste süsteemidega (joonis 4.1-14).

Tänapäeva töös CAD-andmetega (BIM) oleme jõudnud punkti, kus me ei pea CAD (BIM) müüjatelt luba küsima, et pääseda ligi CAD-formaatide teabele.



Joonis 4.1-14 Kaasaegsed SDK-vahendid võimaldavad andmete seaduslikku konverteerimist patenteeritud CAD-andmebaasiformaatidest (BIM).

Praeguseid suundumusi CAD-andmete töötlemisel kujundavad jätkuvalt peamised turuosalised - CAD - müüjad, kes püüavad tugevdada oma positsiooni andmemaailmas ning luua uusi formaate ja

kontseptsioone.

CAD lahenduste müüjad liiguvad struktureeritud andmete suunas.

Alates 2024. aastast toimub projekteerimis- ja ehitustööstuses märkimisväärne tehnoloogiline muutus andmete kasutamisel ja töötlemisel. CAD -süsteemi müüjad keskenduvad projekteerimisandmete vaba juurdepääsu asemel järgmiste uute kontseptsioonide edendamisele. Sellised lähenemisviisid nagu BIM (loodud 2002. aastal) ja avatud BIM (loodud 2012. aastal) annavad järk-järgult teed kaasaegsetele tehnoloogilistele lahendustele, mida CAD-müüjad hakkavad edendama [93]:

- liikumine "granuleeritud" andmete kasutamise suunas, mis võimaldab tõhusat teabehaldust ja liikumist andmeanalüüsi suunas.
- USD formaadi tekkimine ja Entity-component-system lähenemisviisi (ECS) rakendamine andmete paindlikuks korraldamiseks.
- Tehisintellekti aktiivne kasutamine andmetöötluses, protsesside automatiseerimisel ja andmeanalüüsis.
- Koostalitlusvõime arendamine - erinevate programmide, süsteemide ja andmebaaside parem koostoime.

Kõiki neid aspekte käsitletakse üksikasjalikumalt raamatu "CAD ja BIM: turundus, tegelikkus ja projekteerimisandmete tulevik ehituses" kuuendas osas. Selles peatükis kirjeldame vaid lühidalt muutuste üldist vektorit: suuremad CAD-müüjad püüavad nüüd uuesti läbi mõelda, kuidas projekteerimisandmed on struktureeritud. Üks peamisi muutusi on loobumine klassikalisest failipõhisest salvestusmudelitest ja üleminek granuleeritud, analüüsile orienteeritud andmearhitektuurile, mis võimaldab pidevat juurdepääsu üksikutele mudelikomponentidele [93].

Toimuva asja tuum on see, et tööstusharu loobub järk-järgult kohmakatest, spetsialiseeritud ja parameetrilistest vormingutest, mis nõuavad geomeetrilisi tuumasid, ning eelistab universaalsemaid, masinloetavaid ja paindlikumaid lahendusi.

Üheks selliseks muutuste mootoriks on USD (Universal Scene Description) formaat, mis on algselt välja töötatud arvutigraafikatööstuses, kuid mida on juba tunnustatud insenerirakendustes tänu NVIDIA Omniverse (ja Isaac Sim) platvormi arendamisele simulatsioonide ja visualiseerimiste jaoks [93]. Erinevalt parameetrilisest IFC-st pakub USD lihtsamat struktuuri ja võimaldab geomeetria ja objekti omaduste kirjeldamist JSON-vormingus (joonis 4.1-15), mis lihtsustab teabe töötlemist ja kiirendab selle integreerimist digitaalsesse protsessi. Uus formaat võimaldab geomeetria salvestamist (lisaks BREP -NURBS - täpsemalt raamatu 6. osas) MESH-polygoonide kujul ja objekti omadusi JSONis, mis muudab selle mugavamaks automatiseeritud protsesside ja pilvede ökosüsteemides töötamise jaoks [94].

Mõned CAD- ja ERP-tootjad kasutavad juba sarnaseid formaate (nt NWD, SVF, CP2, CPIXML), kuid enamik neist on endiselt suletud ja väliskasutamiseks kättesaamatu, mis piirab andmete

integreerimise ja taaskasutamise võimalusi. Selles kontekstis võib USD mängida sama rolli, mida kunagi mängis DXF - avatud alternatiivi sellistele patenteeritud vormingutele nagu 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

Joonis 4.1-15 USD formaat, mis on CAD-tootjate katse vastata nõudlusele koostalitlusvõime ja projekteerimisandmete sõltumatusse järele geomeetrilistest tuumadest.

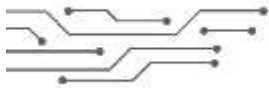
Peamiste arendajate üleminek avatud ja lihtsustatud USD, GLTF, OBJ, XML (suletud NWD, CP2, SVF, SVF2, CPIXML) ja sarnastele vormingutele (joonis 3.1-17) peegeldab ülemaailmset suundumust ja tööstuse nõudlust andmete lihtsustamise ja parema kättesaadavuse järele. Lähiaastatel on oodata järkjärgulist eemaldumist keerulistest parameetrilistest standarditest ja vormingutest, mis sõltuvad geomeetrilistest tuumadest, ning üleminekut lihtsamate ja struktureeritumate lahenduste kasuks. See üleminek kiirendab ehitussektori digitaliseerimist, hõlbustab protsesside automatiseerimist ja lihtsustab andmevahetust.

Hoolimata CAD -ettevõtjate strateegilistest plaanidest edendada uusi avatud formaate, saavad

ehitussektori spetsialistid täieliku juurdepääsu suletud CAD-süsteemide andmetele ka ilma CAD (BIM) vahenditeta, kasutades pöördprojekteerimisvahendeid.

Kõik need suundumused viivad paratamatult üleminekuni mahukatelt, monoliitsetelt 3D-mudelitelt universaalsetele, struktureeritud andmetele ja teistes tööstusharudes juba ammu tõestatud vormingute kasutamisele. Kui projektimeeskonnad hakkavad nägema CAD-mudeleid mitte ainult visuaalsete objektide või failikogumina, vaid teadmisi ja teavet sisaldavate andmebaasidena, muutub lähenemine projekteerimisele ja juhtimisele kardinaalselt.

Kui meeskonnad on õppinud struktureeritud andmete väljavõtteid dokumentidest, tekstidest, joonistest ja CAD-mudelitest ning neil on juurdepääs andmebaasidele, on järgmine oluline samm andmete modelleerimine ja kvaliteedi tagamine. Just see samm määrab suures osas ära sellise teabe töötlemise ja muundamise kiiruse, mida lõpuks kasutatakse otsuste tegemiseks konkreetsete rakendusülesannete puhul.



PEATÜKK 4.2.

KLASSIFITSEERIMINE JA INTEGREERIMINE: ÜHINE KEEL EHITUSANDMETE JAOKS

Otsuste tegemise kiirus sõltub andmete kvaliteedist

Tänapäeval on projekteerimisandmete arhitektuuris toimumas põhjalikud muutused. Tööstus liigub ära mahukatelt, isoleeritud mudelitelt ja suletud formaatidelt paindlikumate, masinloetavate struktuuride suunas, mis keskenduvad analüüsile, integratsioonile ja protsesside automatiseerimisele. Kuid üleminek uutele vormingutele üksi ei taga tõhusust - tähelepanu keskmes on paratamatult andmete kvaliteet.

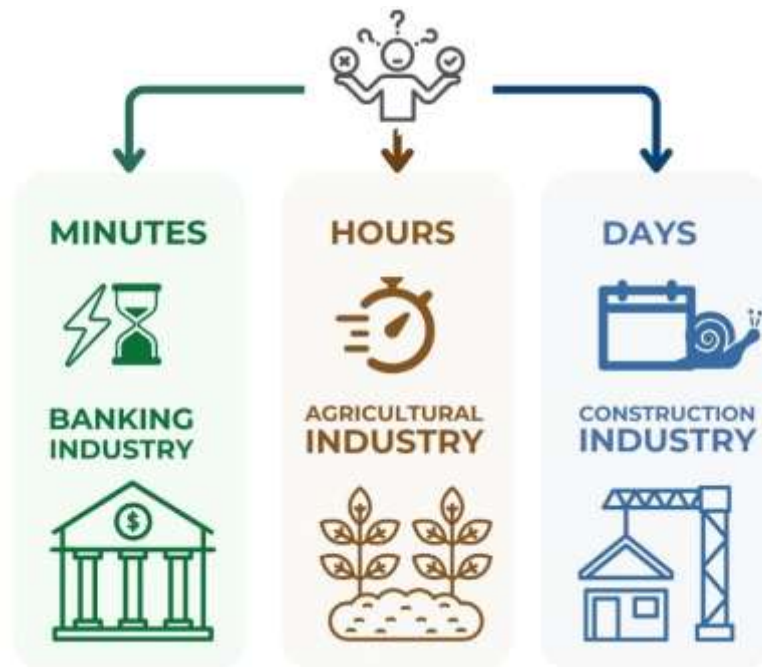
Selle raamatu lehekülgedel räägime palju formaatidest, süsteemidest ja protsessidest. Kuid kõik need jõupingutused on mõttetud ilma ühe põhielemendita: andmed, mida saab usaldada. Andmete kvaliteet on digitaliseerimise nurgakivi, mille juurde me järgnevates osades tagasi pöördume.

Kaasaegsed ehitusettevõtted - eriti suured - kasutavad kümneid ja mõnikord tuhandeid erinevaid süsteeme ja andmebaase (joonis 4.2-1). Neid süsteeme tuleb mitte ainult regulaarselt uue teabega täita, vaid ka omavahel tõhusalt suhelda. Kõik uued andmed, mis tekivad sissetuleva teabe töötlemise tulemusena, integreeritakse nendesse keskkondadesse ja neid kasutatakse konkreetsete äriülesannete lahendamiseks.

Ja kui varem langetasid otsuseid konkreetsete äriülesannete kohta tippjuhid - nn HiPPOd (joonis 2.1-9) - kogemuste ja intuitsiooni põhjal, siis tänapäeval, kui teabe hulk on järsult kasvanud, on selline lähenemine muutumas vastuoluliseks. Seda on asendamas automatiseeritud analüütika, mis töötab reaalajaliste andmetega.

"Traditsioonilised-manuaalsed" arutelud juhtkonna tasandil toimivate äriprotsesside üle lähevad üle operatiivanalüütikale, mis nõuab kiireid vastuseid äriküsimustele.

Ajastu, mil raamatupidajad, töödejuhatajad ja kalkulaatorid käsitsi koostatud aruannete ja kokkuvõtlike tabelite ning projektiandmete näidisvormide koostamine päevade ja nädalate jooksul on möödas. Tänapäeval on otsuste tegemise kiirus ja õigeaegsus muutumas konkurentsieelise võtmeteguriks.



Joonis 4.2-1 Ehitustööstuses kulub arvutuste tegemiseks ja otsuste tegemiseks päevi, erinevalt teistest tööstusharudest, kus see toimub tundide või minutitega.

Peamine erinevus ehitussektori ja digitaalselt arenenumate tööstusharude vahel (joonis 4.2-1) on andmete madal kvaliteet ja standardiseerimine. Vananenud lähenemisviisid teabe genereerimisele, edastamisele ja töötlemisele aeglustavad protsesse ja tekitavad kaose. Ühtsete andmekvaliteedi standardite puudumine takistab läbiva automatiseerimise rakendamist

Üheks peamiseks probleemiks on endiselt sisendandmete halb kvaliteet ning nende ettevalmistamise ja valideerimise formaliseeritud protsesside puudumine. Ilma usaldusväärsete ja järjepidevate andmeteta ei ole võimalik süsteeme tõhusalt integreerida. See toob kaasa viivitusi, vigu ja suuremaid kulusid projekti elutsükli igas etapis.

Raamatu järgmistes osades vaatleme üksikasjalikult, kuidas saate parandada andmete kvaliteeti, standardiseerida protsesse ja lühendada teekonda teabest kvaliteetsete, valideeritud ja järjepidevate andeteni.

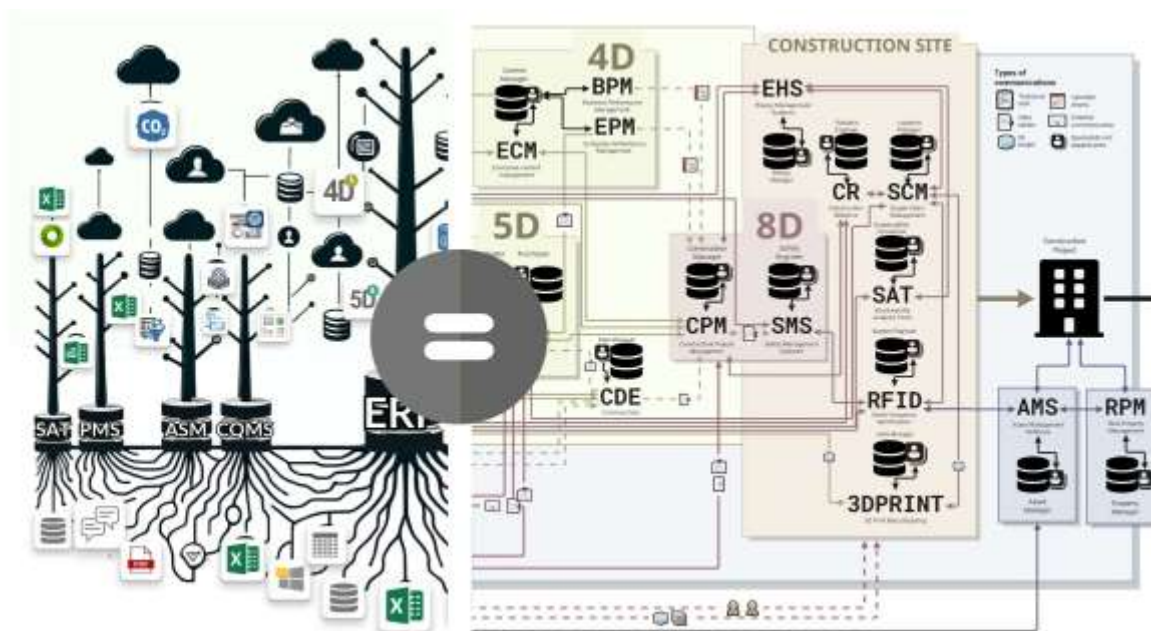
Andmete standardimine ja integreerimine

Tõhus andmehaldus nõuab selget standardimisstrateegiat. Ainult selged nõuded andmete struktuurile ja kvaliteedile võimaldavad automatiseerida andmete valideerimist, vähendada käsitsi tehtavaid toiminguid ja kiirendada teadlike otsuste tegemist projekti kõikides etappides.

Igapäevases praktikas peab ehitusettevõtte töötlema iga päev sadu faile: e-kirjad, PDF -dokumendid, CAD projekteerimisfailid, andmed IOT anduritelt, mis tuleb integreerida ettevõtte äriprotsessidesse.

Ettevõtte andmebaaside ja tööriistade ökosüsteemi mets (joonis 4.2-2) peab õppima, kuidas saabuvatest mitmeformaadilistest andmetest toitaineid tuletada, et saavutada ettevõtte soovitud tulemusi.

Andmevoogude tõhusaks haldamiseks ei ole tingimata vaja palgata terve armee juhte, vaid esmalt tuleb välja töötada ranged nõuded ja standardid andmetele ning kasutada asjakohaseid vahendeid nende automaatseks valideerimiseks, ühtlustamiseks ja töötlemiseks.



Joonis 4.2-2 Ettevõtte terve ökosüsteemi tagamine eeldab selle süsteemide kvaliteetset ja õigeaegset varustamist.

Andmete valideerimise ja ühtlustamise protsessi automatiseerimiseks (hilisemaks automaatseks integreerimiseks) peaksite alustama iga konkreetse süsteemi jaoks vajalike andmete miinimumnõuete kirjeldamisest. Need nõuded määratlevad:

- Mida täpselt on vaja hankida?
- Millisel kujul (struktuur, formaat)?
- Millised atribuudid on kohustuslikud?
- Millised täpsuse ja täielikkuse tolerantsid on vastuvõetavad?

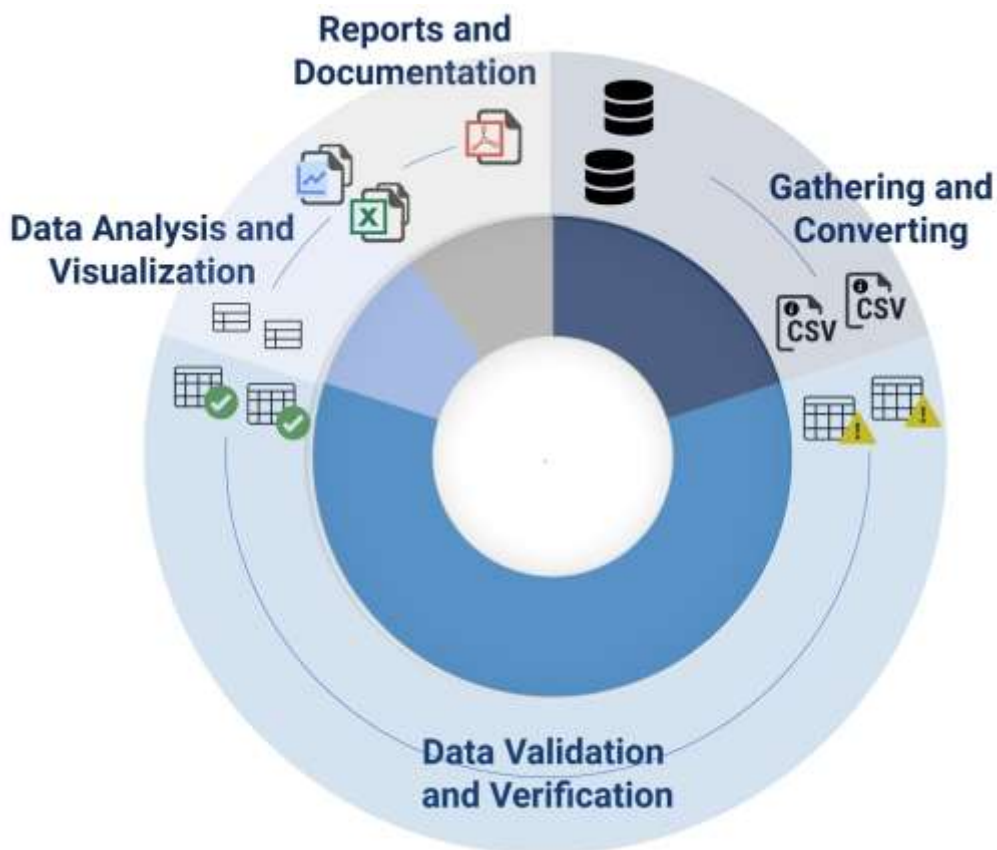
Andmenõuded kirjeldavad saadud ja töödeldud teabe kvaliteedi, struktuuri ja täielikkuse kriteeriume. Näiteks PDF-dokumentide tekstide puhul on oluline, et need oleksid täpselt vormindatud vastavalt tööstusstandarditele (joonis 7.2-14 - joonis 7.2-16). CAD -mudelite objektidel peavad olema õiged atribuudid (mõõtmed, koodid, lingid klassifikaatoritele) (joonis 7.3-9, joonis 7.3-10). Ja lepingute skaneerimisel on olulised selged kuupäevad ning võime automaatselt eraldada summa ja

põhitingimused (joonis 4.1-7 - joonis 4.1-10).

Andmenõuete sõnastamine ja nende vastavuse automaatne kontrollimine on üks kõige aeganõudvamaid, kuid kriitilisemaid samme. See on kõige aeganõudvam samm äriprotsessides.

Nagu käesoleva raamatu 3. osas mainitud, kulub 50% kuni 90% äriteabe (BI) spetsialistide ajast pigem andmete ettevalmistamisele kui analüüsile (joonis 3.2-5). See protsess hõlmab andmete kogumist, kontrollimist, valideerimist, ühtlustamist ja struktureerimist.

2016. aasta uuringu [95] kohaselt väitsid paljude erinevate laia spektriga valdkondade andmeteadlased, et nad kulutavad suurema osa oma tööajast (umbes 80%) sellele, mida neile kõige vähem meeldib teha (joonis 4.2-3): olemasolevate andmekogumite kogumine ja nende korrastamine (ühendamine, struktureerimine). Seega jääb vähem kui 20% nende ajast loomingulisteks ülesanneteks, näiteks uute arusaamade ja avastusteni viivate mustrite ja seaduspärasuste leidmiseks.



Joonis 4.2-3 kontrollimine ja andmete kvaliteedi tagamine on kõige kulukam, aeganõudvam ja keerulisem etapp andmete ettevalmistamisel teistesse süsteemidesse integreerimiseks.

Edukas andmehaldus ehitusettevõttes nõuab terviklikku lähenemisviisi, mis hõlmab ülesannete

parameetriseerimist, andmete kvaliteedinõuete sõnastamist ja sobivate vahendite kasutamist nende automatiseeritud valideerimiseks.

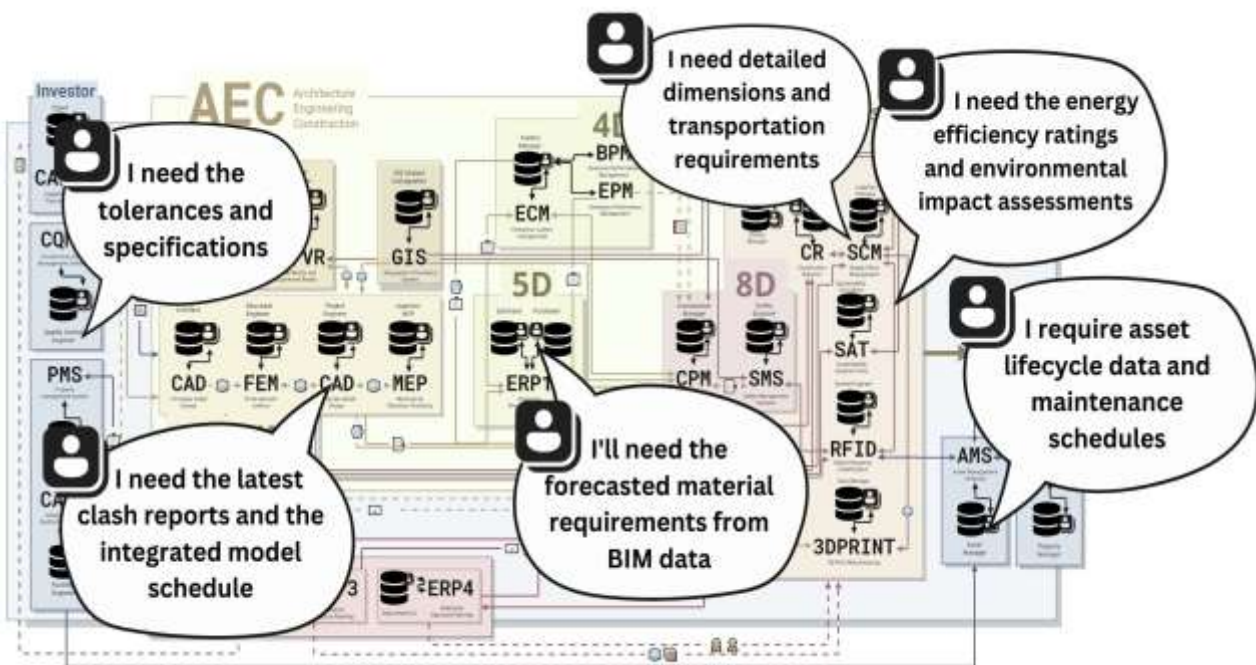
Digitaalne koostalitlusvõime alab nõuetest

Kuna digitaalsete süsteemide arv ettevõtetes kasvab, kasvab ka vajadus nende vaheliste andmete järjepidevuse järele. Erinevate IT-süsteemide eest vastutavad juhid ei suuda sageli pidada sammu üha suureneva andmemahu ja erinevate vormingutega. Sellises olukorras on nad sunnitud paluma spetsialistidel luua andmed sellisel kujul, et neid saaks kasutada teistes rakendustes ja platvormidel.

See omakorda nõuab, et andmete genereerimisega tegelevad insenerid ja töötajad kohaneksid paljude nõuetega, sageli ilma läbipaistvuse ja selge arusaamata sellest, kus ja kuidas andmeid tulevikus kasutatakse. Standardiseeritud lähenemisviiside puudumine teabe käsitlemisel toob kaasa ebatõhususe ja suuremad kulud kontrollimise etapis, mis andmete keerukuse ja mittestandardiseerituse tõttu on sageli käsitsi teostatav.

Andmete standardimise küsimus ei ole ainult mugavuse või automatiseerimise küsimus. See on otsene rahaline kahju. IBMi 2016. aasta aruande kohaselt on andmete kehvast kvaliteedist tulenev iga-aastane kahju USAs 3,1 triljonit dollarit [96]. Lisaks näitavad MIT ja teiste analüütiliste konsultatsioonifirmade uuringud, et andmete kehva kvaliteedi hind võib ulatuda 15-25%ni ettevõtte tuludest [97].

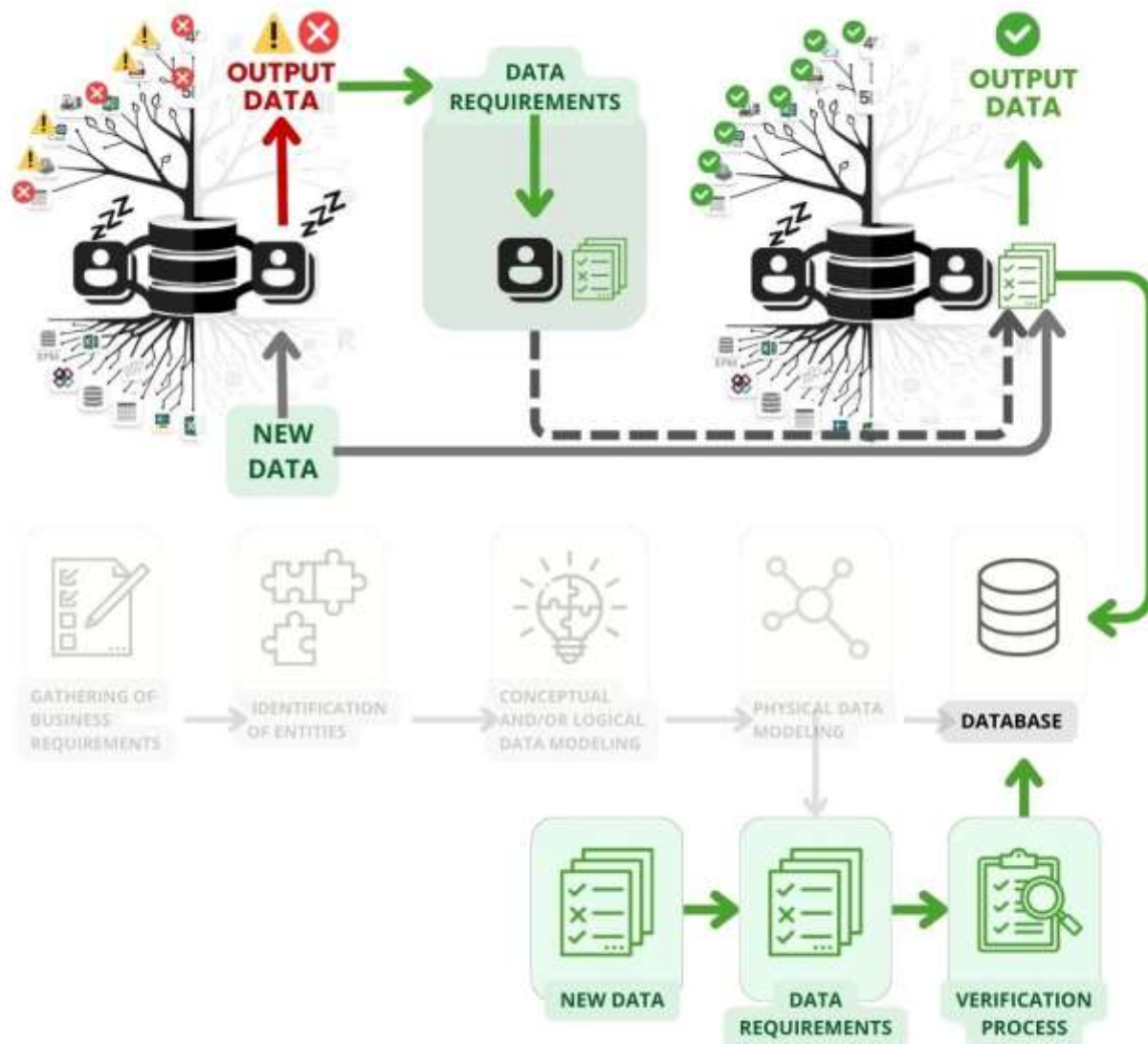
Sellistes tingimustes on väga oluline, et oleks selgelt määratletud andmenõuded ja kirjeldused selle kohta, millised parameetrid, millises formaadis ja millise detailsusega peaksid loodud objektid sisaldama. Ilma neid nõudeid vormistamata on võimatu tagada andmete kvaliteeti ja ühilduvust süsteemide ja projekti etappide vahel (joonis 4.2-4).



Joonis 4.2-4 Äri põhineb erinevate rollide koostoimel, millest igähele on vaja teatavaid parameetreid ja väärtusi, mis on ärieesmärkide täitmiseks kriitilise tähtsusega.

Selleks, et sõnastada õiged andmenõuded, peate mõistma äriprotsesse andmete tasandil. Ehitusprojektid on eri tüüpi, ulatusega ja osalejate arvuga ning iga süsteem - olgu see siis modelleerimine (CAD (BIM)), planeerimine (ERP 4D), kuluarvestus (ERP 5D) või logistika (SCM) - nõuab oma unikaalseid parameetreid sisendite (sisendüksuste-elementide) jaoks.

Sõltuvalt nendest vajadustest peavad ärijuhid kas kujundama uusi andmestruktuure, et vastata nõuetele, või kohandama olemasolevaid tabeleid ja andmebaase. Loodud andmete kvaliteet sõltub otseselt sellest, kui täpselt ja õigesti on nõuded sõnastatud (joonis 4.2-5).



Joonis 4.2-5 Andmete kvaliteet sõltub konkreetsete andmete kasutusjuhtumite jaoks loodud nõuete kvaliteedist.

Kuna igal süsteemil on oma spetsiifilised andmenõuded, on esimene samm üldiste nõuete sõnastamisel kõigi äriprotsessidega seotud elementide liigitamine. See tähendab vajadust jagada objektid klassidesse ja klassirühmadesse, mis vastavad konkreetsetele süsteemidele või rakendusülesannetele. Iga sellise rühma jaoks töötatakse välja eraldi nõuded andmete struktuurile, atribuutidele ja kvaliteedile.

Praktikas seisab selle lähenemisviisi rakendamine aga silmitsi suure probleemiga: andmete rühmitamise ühtse keele puudumine. Erinev klassifikatsioon, dubleerivad identifikaatorid ja kokkusobimatud vormingud viivad selleni, et iga ettevõtte, iga tarkvara ja isegi iga projekt moodustab omaette, isoleeritud andmemudelid ja -klassid. Tulemuseks on digitaalne "Paabeli torn", kus teabe edastamine süsteemide vahel nõuab mitmekordset konverteerimist õigetes andmemudelitesse ja -klassidesse, mida sageli tehakse käsitsi. Seda takistust saab ületada ainult universaalsete klassifikaatorite ja standardiseeritud nõuete kogumite suunas liikudes.

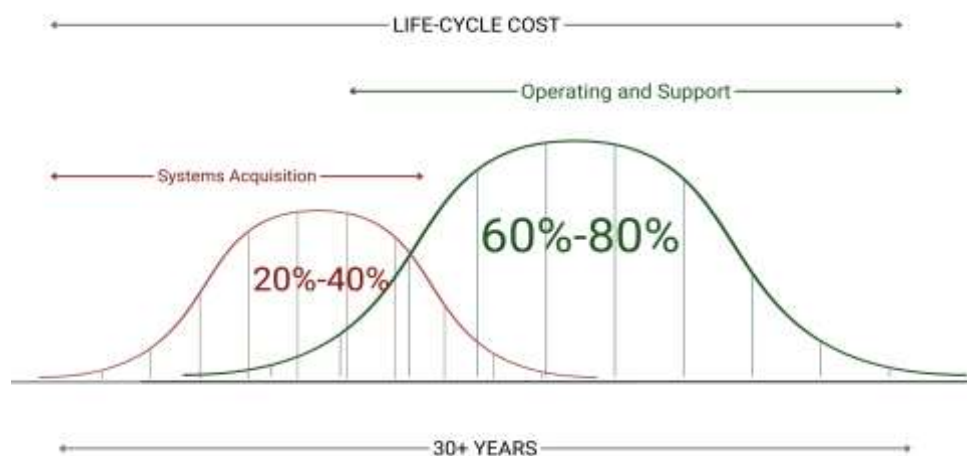
Ühine ehituskeel: klassifikaatorite roll digitaalses ümberkujundamises

Kontrolli- ja töötlemisprotsesside digitaliseerimise ja automatiseerimise kontekstis on klassifitseerimissüsteemidel eriline roll - omamoodi "digitaalsed sõnastikud", mis tagavad objektide kirjeldamise ja parameetriseerimise ühetaolisuse. Klassifikaatorid moodustavad "ühise keele", mis võimaldab andmeid rühmitada tähenduse järgi ja integreerida neid erinevate süsteemide, juhtimistasandite ja projekti elutsükli etappide vahel.

Klassifikaatorite kõige käegakatsutavam mõju avaldub hoone elutsükli majanduses, kus kõige olulisem aspekt on pikaajaliste tegevuskulude optimeerimine. Uuringud näitavad, et tegevuskulud moodustavad kuni 80% hoone kogukuludest, mis on kolm korda suuremad kui algsed ehituskulud (joonis 4.2-6) [98]. See tähendab, et otsus tulevaste kulude kohta kujuneb suures osas projekteerimise etapis

Seetõttu peaksid nõuded operatsiooniinseneride poolt (CAFM, AMS, PMS, RPM) saama lähtepunktiks andmenõuete koostamisel projekteerimisfaasis (joonis 1.2-4). Neid süsteeme ei tohiks vaadelda kui projekti lõppetappi, vaid kui projekti kogu digitaalse ökosüsteemi lahutamatu osa, alates kontseptsioonist kuni lammutamiseni

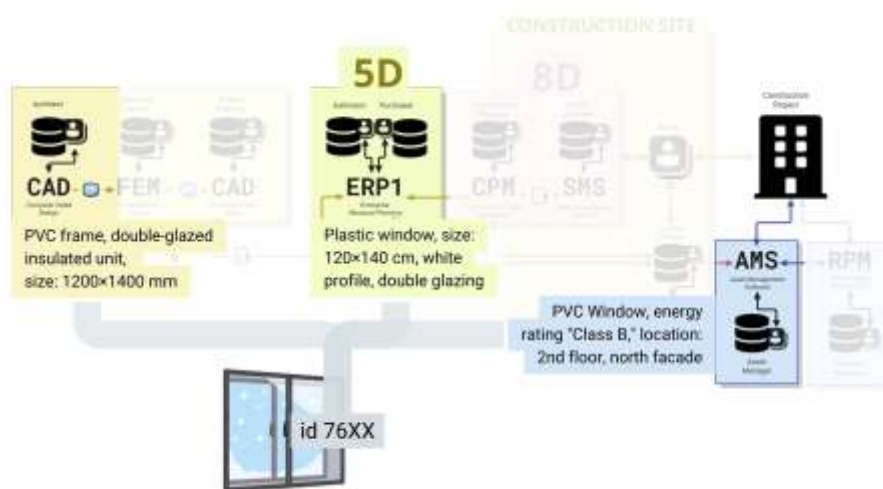
Kaasaegne klassifikaator ei ole lihtsalt koodide süsteem rühmitamiseks. See on mehhanism arhitektide, inseneride, kalkulaatorite, logistikute, hoolduse ja IT-süsteemide vastastikuse mõistmise jaoks. Nii nagu auto autopiloot peab teeobjekte üheselt ja täpselt ära tundma, peavad digitaalsed ehitussüsteemid ja nende kasutajad tõlgendama sama projektielementi üheselt erinevate süsteemide jaoks elementide klassi kaudu.



Joonis 4.2-6 Kasutus- ja hoolduskulud ületavad ehitusmaksumust kolmekordselt, moodustades 60-80% hoone elutsükli kogukuludest (vastavalt [99]).

Klassifikaatorite arendamise tase on otseselt seotud ettevõtte digitaliseerimise sügavusega ja selle digitaalse küpsusega. Madala digitaalse küpsuse tasemega organisatsioonid seisavad silmitsi killustatud andmete, kokkusobimatute infosüsteemide ja sellest tulenevalt kokkusobimatute ja ebatõhusate klassifikaatoritega. Sellistes ettevõtetes võib ühel ja samal elemendil olla sageli eri süsteemides erinevad rühmitustunnused, mis takistab kriitiliselt lõplikku integreerimist ja muudab protsesside automatiseerimise võimatuks.

Näiteks võib sama aken projektis olla erinevalt märgistatud CAD mudelis, kalkulatsiooni- ja hooldussüsteemis (joonis 4.2-7), sest protsessi eri osalejad tajuvad elemente mitmemõõtmeliselt. Hindaja jaoks on aknakategooria elemendi puhul olulised maht ja maksumus, hooldusteenuse jaoks kättesaadavus ja hooldatavus, arhitekti jaoks esteetilised ja funktsionaalsed omadused. Selle tulemusena võib sama element nõuda erinevaid parameetreid.



Joonis 4.2-7 Süsteemidevahelise ebajärjekindla klassifitseerimise korral kaotab element igal sammul, mis toimub tema üleminekul teise süsteemi, osa atribuutide kohta käivat teavet.

Kuna ehitusdetailide klassifitseerimist on raske üheselt määratleda, määravad eri valdkondade spetsialistid sageli samale elemendile kokkusobimatud klassid. See toob kaasa objekti ühtse käsitlemise kadumise, mis nõuab hilisemat käsitsi sekkumist, et ühtlustada erinevaid klassifitseerimissüsteeme ja luua järjepidevus eri spetsialistide poolt määratletud tüüpide ja klasside vahel.

Selle ebajärjekindluse tõttu ei saa hankeosakonda (ERP) saabunud tegevusdokumente, kui ehitusdetail ostetakse tootjalt, sageli õigesti seostada selle objekti klassifitseerimisega ehitusplatsil (PMIS, SCM). Selle tulemusena ei integreerita kriitilist teavet tõenäoliselt infrastruktuuri ja varahaldussüsteemidesse (CAFM, AMS), mis tekitab tõsiseid probleeme nii kasutuselevõtu kui ka hilisema hoolduse (AMS, RPM) või elemendi väljavahetamise ajal.

Kõrge digitaalse küpsusega ettevõtetes mängivad klassifikaatorid närvisüsteemi rolli, mis ühendab kõiki infovooge. Sama element saab unikaalse tunnuse, mis võimaldab seda edastada CAD, ERP, AMS ja CAFM -süsteemide ja nende klassifikaatorite vahel ilma moonutuste või kadudeta.

Tõhusate klassifikaatorite loomiseks tuleb mõista, kuidas andmeid kasutatakse. Üks ja sama insener võib eri projektides nimetada ja klassifitseerida elementi erinevalt. Ainult aastate jooksul kasutamisstatistikat kogudes saab välja töötada stabiilse klassifitseerimissüsteemi. Selles aitab masinõpe: algoritmid analüüsivad tuhandeid projekte (joonis 9.1-10), tuvastades masinõppe abil tõenäolised klassid ja parameetrid (joonis 10.1-6). Automaatne klassifitseerimine on eriti väärtuslik keskkondades, kus käsitsi klassifitseerimine ei ole andmemahu tõttu võimalik. Automaatsed klassifitseerimissüsteemid suudavad eristada põhikategooriaid minimaalselt täidetud esemeparameetrite alusel (täpsemalt raamatu üheksandas ja kümnendas osas).

Arendatud klassifikaatorisüsteemid muutuvad edasise digitaliseerimise katalüsaatoriteks, luues aluse:

- Projekti kulude ja tähtaegade automatiseeritud hindamine.
- Võimalike riskide ja konfliktide prognoositav analüüs
- Hankemenetluste ja logistikaahelate optimeerimine
- Hoonete ja struktuuride digitaalsete kaksikute loomine
- Integratsioonid aruka linna ja asjade interneti süsteemidega

Aeg ümberkujundamiseks on piiratud - masinõppe ja arvutinägemistehnoloogiate arenguga lahendatakse lähiaastatel aastakümneid lahendamatu automaatse klassifitseerimise probleem ning ehitus- ja projekteerimisettevõtted, kes ei suuda õigeaegselt kohaneda, riskivad digitaalsete platvormide poolt tõrjutud taksoparkide saatuse kordumisega.

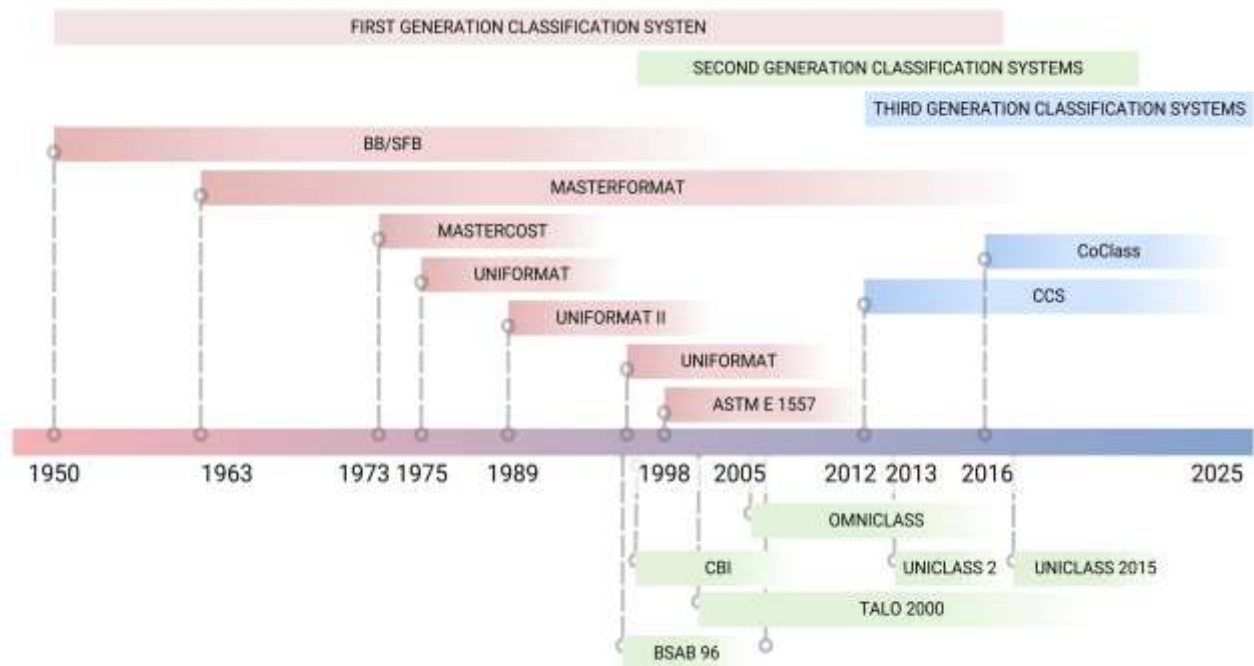
Kulu- ja ajakavaarvutuste automatiseerimist ning suurandmete ja masinõppe kasutamist käsitletakse üksikasjalikumalt raamatu viiendas ja üheksandas osas. Taksoparkide saatuse kordumise ohtu ja ehitussektori ümberiseerimist käsitletakse üksikasjalikult raamatu kümnendas osas.

Mõistes klassifikaatorite võtmerolli ehitussektori digitaalses ümberkujundamises, tuleb pöörduda nende arengu ajaloo juurde. Just ajalooline kontekst võimaldab meil mõista, kuidas on klassifikatsiooni lähenemisviisid arenenud ja millised suundumused määravad nende praeguse seisundi.

Masterformat, OmniClass, Uniclass ja CoClass: klassifitseerimissüsteemide areng.

Ajalooliselt on ehituselementide ja tööde klassifikaatorid arenenud kolmes põlvkonnas, millest igaüks kajastab olemasolevat tehnoloogiat ja tööstuse jooksvaid vajadusi konkreetsel ajaperioodil (joonis 4.2-8):

- **Esimene põlvkond** (1950ndate algusest kuni 1980ndate lõpuni) - paberkataloogid, kohalikud hierarhilised klassifikaatorid (nt Masterformat, Sfb).
- **Teise põlvkonna** (1990ndate lõpust kuni 2010ndate keskpaigani) moodustavad Exceli ja Accessi abil rakendatud tabelarvutid ja struktureeritud andmebaasid (ASTM E 1557, OmniClass, Uniclass 1997).
- **Kolmas põlvkond** (2010ndatest tänapäevani) - digitaalsed teenused ja APId -liidesed, integreerimine CAD (BIM), automatiseerimine (Uniclass 2015, CoClass).



Joonis 4.2-8 Ehitustööstuse klassifikaatorite kolm põlvkonda.

Viimastel aastakümnetel on klassifikaatorite hierarhiline keerukus vähenenud (joonis 4.2-9): kui varasemad süsteemid, nagu OmniClass, kasutasid kuni 7 astet, et kirjeldada 6887 klassi, siis

tänapäevased lahendused, nagu CoClass, piirduvad 3 astmega ja 750 klassiga. See lihtsustab tööd andmetega, säilitades samas vajaliku granulaarsuse. Uniclass 2015, mida kasutatakse sageli standardina Ühendkuningriigis, ühendab 7210 klassi vaid 4 tasandisse, mis teeb selle mugavaks CAD projektide ja riigihangete jaoks.

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

Joonis 4.2-9 Iga uue klassifikaatoripõlvkonnaga väheneb liigitamise keerukus järsult.

Erinevate riikide ehitushinnangute süsteemides võib isegi sellist tüüpilist elementi nagu betoonist vundamendisein kirjeldada väga erinevalt (joonis 4.2-10). Need erinevused kajastavad riiklikke ehitustavasid, kasutatavaid mõõtmissüsteeme, materjalide klassifitseerimise lähenemisviise ning igas riigis kehtivaid regulatiivseid ja tehnilisi nõudeid.



Joonis 4.2-10 Sama elementi kasutatakse eri riikide projektides erinevate kirjelduste ja klassifikatsioonide kaudu.

Samade elementide liigituste mitmekesisus raskendab rahvusvahelist koostööd ning muudab kulude ja tööde ulatuse võrdlemise rahvusvaheliste projektide raames aeganõudvaks ja mõnikord peaaegu võimatuks. Hetkel ei ole ülemaailmsel tasandil üht universaalset klassifikaatorit - iga riik või piirkond töötab välja oma süsteemid, mis põhinevad kohalikel normidel, keelel ja ärikultuuril:

- **CCS (Taani):** Kulude liigitussüsteem - süsteem kulude liigitamiseks kogu rajatise elutsükli jooksul (projekteerimine, ehitamine, käitamine). Põhirõhk on käitamise ja hoolduse loogikal, kuid hõlmab ka eelarve- ja ressurside haldamist.
- **NS 3451 (Norra):** liigitab rajatised funktsioonide, projekteerimiselementide ja elutsükli etappide

kaupa. Kasutatakse projektijuhtimiseks, kulude hindamiseks ja pikaajaliseks planeerimiseks.

- **MasterFormat** (USA): süsteem ehitusspetsifikaatide struktureerimiseks osadeks (nt betoon, elektritööd, viimistlus). Keskendub pigem erialadele ja tööliikidele kui funktsionaalsetele elementidele (erinevalt UniFormatist).
- **Uniclass 2** (UK): üks kõige üksikasjalikumaid klassifikaatoreid, mida kasutatakse riigihangete ja BIM-projektide puhul. Ühendab objektide, tööde, materjalide ja ruumide andmed ühtsesse süsteemi.
- **OmniClass**: rahvusvaheline standard (välja töötatud CSI poolt USAs) objektiteabe haldamiseks komponentide raamatukogudest elektrooniliste spetsifikatsioonideni. Sobib andmete pikaajaliseks säilitamiseks, ühildub CAD (BIM) ja muude digitaalsete vahenditega.
- **COBie**: Ehitus- ja ehitusinfo vahetamine on rahvusvaheline standard projekteerimis-, ehitus- ja kasutusetappide vahelise andmevahetuse jaoks. Sisaldab BS 1192-4:2014 osana kontseptsioonist "BIM - kasutusvalmis mudel". Keskendub teabe edastamisele (nt seadmete spetsifikatsioonid, garantiid, töövõtjate kontaktid).

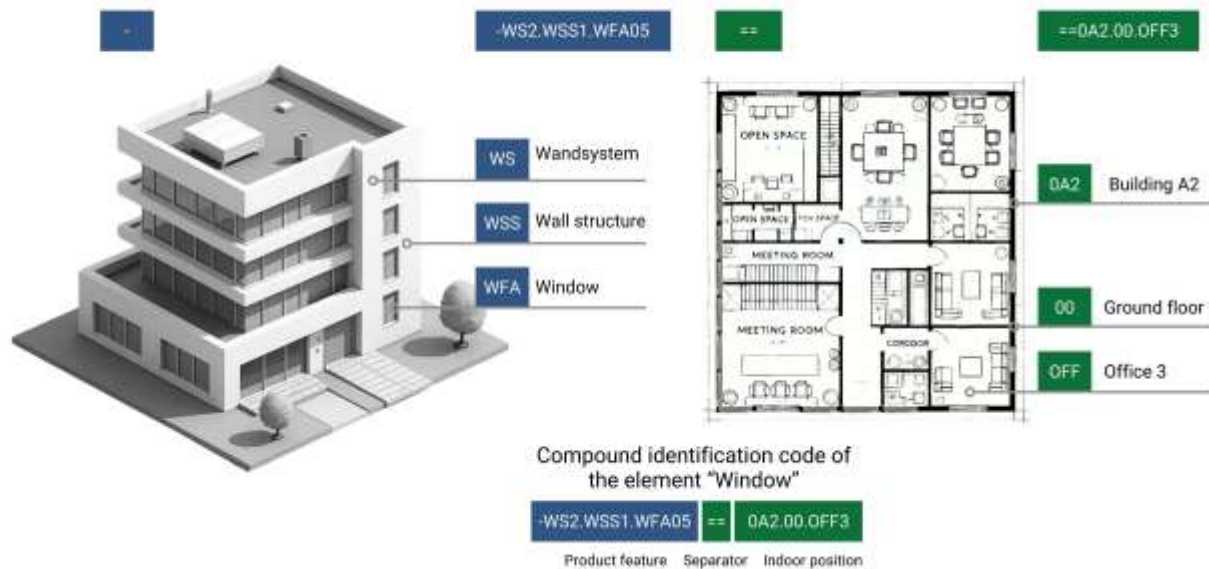
Ehitustööstuse globaliseerumine toob tõenäoliselt kaasa ehitusdetailide liigitussüsteemide järkjärgulise ühtlustamise, mis vähendab oluliselt sõltuvust kohalikest riiklikest standarditest. See protsess võib areneda sarnaselt Interneti-side arenguga, kus universaalsed andmeedastusprotokollid tõrjusid lõpuks välja erinevad kohalikud vormingud, tagades süsteemide ülemaailmse koostalitlusvõime.

Alternatiivne arengutee võib olla otsene üleminek masinõppe tehnoloogiatel põhinevatele automaatsetele klassifitseerimissüsteemidele. Sellistel tehnoloogiatel, mida praegu arendatakse peamiselt autonoomse transpordi valdkonnas, on märkimisväärne potentsiaal suurte CAD-konstruktsioonandmekogumite puhul (joonis 10.1-6).

Tänapäeval ei piirdu olukord ainult klassifikaatorite siseriikliku klastriga. Paljude eripärade tõttu, mida riiklikul tasandil ei võeta arvesse, peab iga ettevõtte ühtlustama ja standardiseerima elementide ja ressursside kategooriad, millega ta töötab.

Reeglina alustatakse seda protsessi väiksel - objektide kohalike tabelite või sisemiste märgistussüsteemidega. Strateegiline eesmärk on siiski liikuda kõigi elementide kirjeldamiseks ühtse keele suunas, mis oleks arusaadav mitte ainult ettevõtte sees, vaid ka väljaspool ettevõtet - ideaalis ühtlustatud rahvusvaheliste või tööstusharu klassifikaatoritega (joonis 4.2-8). Selline lähenemisviis hõlbustab integratsiooni väliste partnerite ja digitaalsete süsteemidega ning soodustab ühtsete läbivate protsesside kujunemist objektide elutsükli jooksul.

Enne automatiseerimisele ja skaleeritavatele IT-süsteemidele üleminekut on vaja kas kasutada riikliku tasandi klassifikaatoreid või luua oma, loogiline ja üheselt mõistetav elementide tuvastamise struktuur. Iga objekt - olgu see siis aken (joonis 4.2-11), uks või tehnosüsteem - peab olema kirjeldatud nii, et see oleks üheselt äratuntav igas ettevõtte digitaalsüsteemis. See on kriitilise tähtsusega üleminekul tasapinnalistelt joonistelt digitaalsetele mudelitele, mis hõlmavad nii projekteerimisfaasi kui ka ehitise käitamist.



Joonis 4.2-11 Näide aknaehituselementide kombineeritud identifikaatorist, mis põhineb klassifikatsioonil ja asukohal hoones.

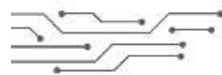
Üks näide siseklassifikaatorite kohta võiks olla liitkoodi väljatöötamine (joonis 4.2-11). Selline kood kombineerib mitu tasandit teavet: elemendi funktsionaalne otstarve (nt "aken seinas"), selle tüüp ja täpne ruumiline viide - hoone A2, korrus 0, ruum 3. Selline mitmetasandiline struktuur võimaldab luua ühtse süsteemi digitaalsetes mudelites ja dokumentides navigeerimiseks, eriti andmete kontrollimise ja teisendamise etappides, kus on vaja elementide ühemõttelist rühmitamist. Elementide ühemõtteline äratundmine tagab osakondade vahelise järjepidevuse ja vähendab dubleerimise, vigade ja teabe kadumise ohtu.

Hästi koostatud klassifikaator ei ole lihtsalt tehniline dokument, vaid ettevõtte digitaalse ökosüsteemi alus:

- tagab andmete ühilduvuse süsteemide vahel;
- vähendab teabe otsimise ja töötlemise kulusid;
- suurendab läbipaistvust ja hallatavust;
- loob aluse skaleerimiseks ja automatiseerimiseks.

Objektide standardiseeritud kirjeldus riiklike klassifikaatorite või patenteeritud kombineeritud tunnuscodeide abil saab aluseks ühtsetele andmetele, usaldusväärsele teabevahetusele ja järgnevale intelligentsete teenuste rakendamisele - alates automaatsetest hangetest kuni digitaalsete kaksikuteni.

Pärast mitmeformaadiliste andmete struktureerimise etapi lõpetamist ja elementide äratundmiseks ja rühmitamiseks kasutatava klassifikaatori valimist on järgmine samm andmete korrektne modelleerimine. See protsess hõlmab võtmeperameetrite kindlaksmääramist, loogilise



andmestruktuuri loomist ja elementidevaheliste seoste kirjeldamist.

PEATÜKK 4.3.

ANDMETE MODELLEERIMINE JA TIPPKESKUS

Andmete modelleerimine: kontseptuaalne, loogiline ja füüsiline mudel

(Meie poolt varem struktureeritud ja kategoriseeritud) andmete tõhus haldamine on võimatu ilma läbimõeldud salvestus- ja töötlemisstruktuurita. Et tagada teabe kättesaadavus ja järjepidevus salvestus- ja töötlemisetapis, kasutavad ettevõtted andmete modelleerimist, mis on metoodika, mis võimaldab neil kujundada tabelid, andmebaasid ja nendevahelised seosed vastavalt ärinõuetele.

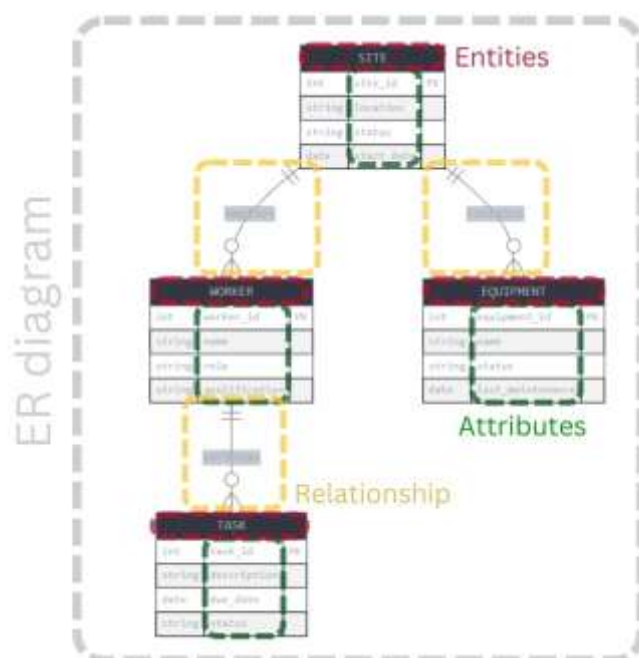
Andmete modelleerimine on alus, millele iga digitaalne ökosüsteem on üles ehitatud. Ilma süsteemide kirjelduse, nõuete ja andmemodelleerimiseta ei tea ega mõista andmeid loovad insenerid ja spetsialistid, kus nende loodud andmeid kasutatakse.

Nagu hoone ehitamisel, kus ei saa alustada telliskivide paigaldamist ilma plaanita, nõuab andmelaosüsteemi loomine selget arusaama sellest, milliseid andmeid kasutatakse, kuidas neid ühendatakse ja kes nendega töötavad. Ilma protsesside ja nõuete kirjelduseta kaotavad andmeid loovad insenerid ja spetsialistid silmist, kus ja kuidas andmeid tulevikus kasutatakse.

Andmemudel on sillaks äri ja IT vahel. See võimaldab vormistada nõuded, struktureerida teavet ja hõlbustada sidusrühmade vahelist suhtlust. Selles mõttes on andmemodelleerimine sarnane arhitekti tööga, kes vastavalt kliendi plaanile töötab välja ehitusplaani ja annab selle seejärel ehitajatele - andmebaasi administraatoritele ja arendajatele - rakendamiseks (andmebaasi loomiseks) üle.

Seega peab iga ehitusettevõtte lisaks elementide ja ressursside struktureerimisele ja kategoriseerimisele (joonis 4.2-11) valdama andmebaaside (tabelite) "ehitamise" kunsti ja õppima, kuidas luua nende vahel seoseid, justkui ühendades tellised ettevõtte andmetest usaldusväärseks ja tugevaks teadmiste seinaks. Andmete modelleerimise põhimõisted (joonis 4.3-1) on järgmised:

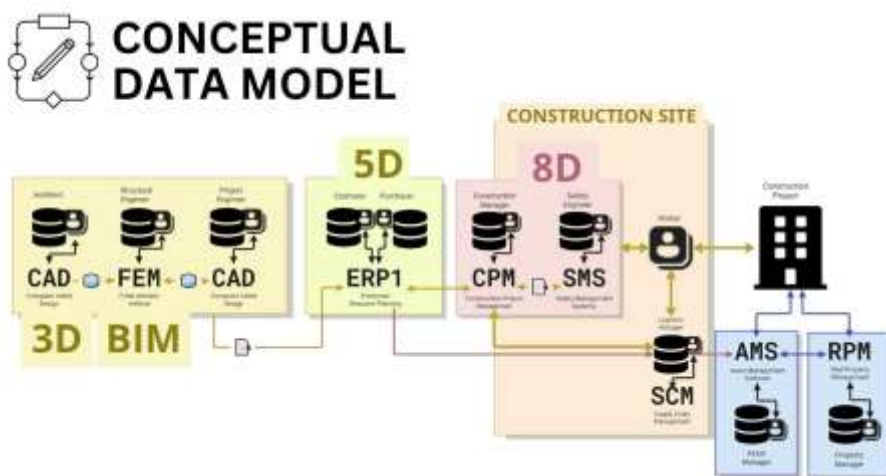
- **Entiteetidid** on objektid, mille kohta tuleb andmeid koguda. Varases projekteerimisfaasis võib üksus olla üks element (nt "uks"), hindamismudelil võib see olla kategoriseeritud elementide rühm (nt "siseuksed").
- **Atribuudid** on üksuste omadused, mis kirjeldavad olulisi üksikasju: mõõtmed, omadused, koostekulud, logistika ja muud parameetrid.
- **Suhted (lingid)** - näitavad, kuidas üksused omavahel suhtlevad. Need võivad olla ühte tüüpi: "üks ühele", "paljud ühele", "paljud paljudele".
- **ER-diagrammid** (Entity-Relationship diagramme) on visuaalsed diagrammid, mis näitavad olemite, atribuutide ja nendevaheliste suhete kohta. ER-diagrammid võivad olla kontseptuaalsed, loogilised ja füüsilised - igaüks neist kajastab erinevat detailsuse taset.



Joonis 4.3-1 ER diagramm kontseptuaalse andmebaasi struktuuri kohta koos üksuste, atribuutide ja suhetega.

Andmete kujundamise ja nendevaheliste seoste määratlemise protsess jaguneb traditsiooniliselt kolmeks peamiseks mudeliks. Igüks neist täidab teatavaid funktsioone, mis erinevad andmete struktuuri kujutamise üksikasjalikkuse ja abstraktsiooniastme poolest:

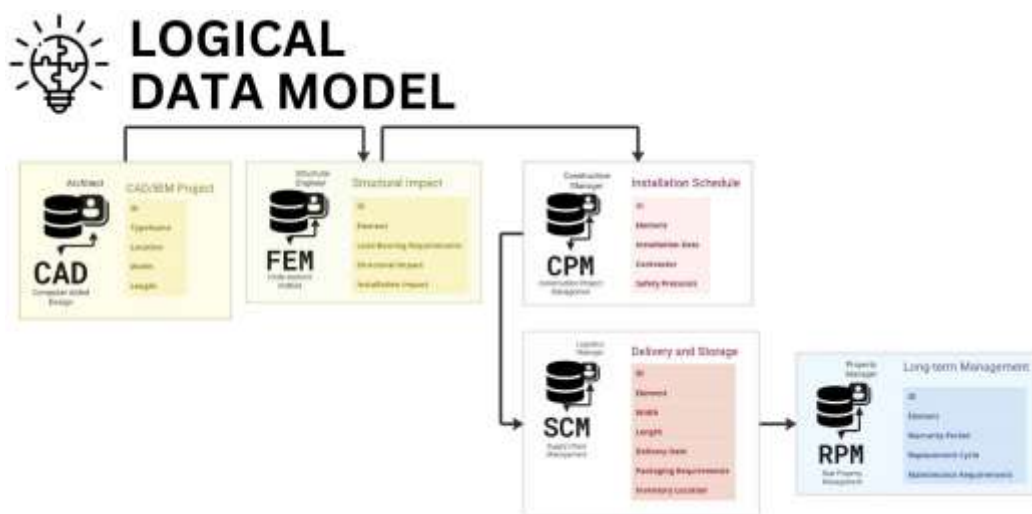
- **Kontseptuaalne andmemudel:** selles mudelis kirjeldatakse peamisi üksusi ja nende suhteid, ilma atribuutide üksikasjadesse laskumata. Seda kasutatakse tavaliselt planeerimise algfaasis. Selles etapis saab andmebaasidest ja süsteemidest visandada erinevate osakondade ja spetsialistide vahelisi seoseid.



Joonis 4.3-2 Kontseptuaalne diagramm kirjeldab süsteemi sisu: kõrgetasemeline suhete esitus, ilma tehniliste üksikasjadeta.

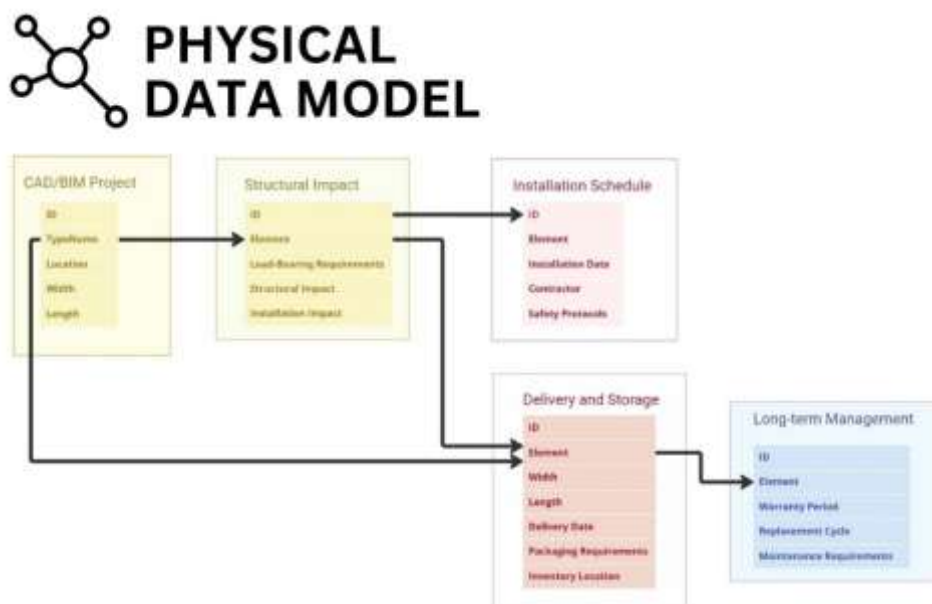
- **Loogiline andmemudel:** Kontseptuaalsel mudelil põhinev loogiline andmemudel sisaldab üksuste,

atribuutide, võtmete ja seoste üksikasjalikke kirjeldusi, äriteabe ja -reeglite kaardistamist.



Joonis 4.3-3 Loogiline andmemudel kirjeldab üksikasjalikult andmetüüpe, seoseid ja võtmeid, kuid ilma süsteemi rakendamiseta.

- **Füüsiline andmemudel:** See mudel kirjeldab andmebaasi rakendamiseks vajalikke struktuure, sealhulgas tabeleid, veerge ja seoseid. See keskendub andmebaasi jõudlusele, indekseerimisstrateegiatele ja füüsilisele salvestusele, et optimeerida andmebaaside füüsilist kasutuselevõttu.

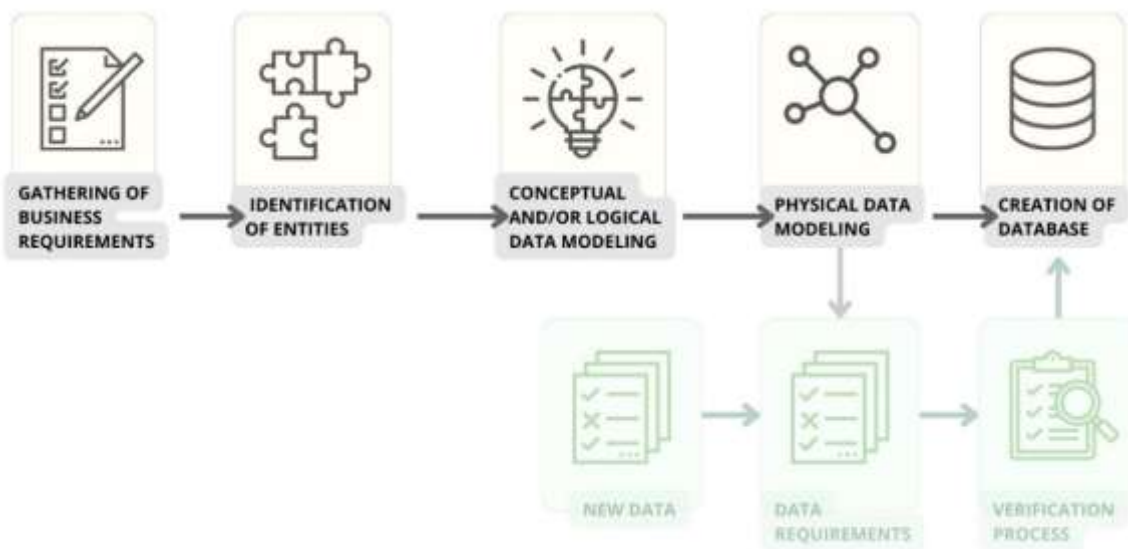


Joonis 4.3-4 Füüsiline andmemudel määratleb, kuidas süsteemi rakendatakse, sealhulgas andmebaasi tabelid ja konkreetsed üksikasjad.

Andmebaaside projekteerimisel ja tabelisuhete kavandamisel on tõhusa süsteemiarhitektuuri loomisel võtmeroll abstraktsioonitasandite mõistmisel.

Tõhus andmemodelleerimise metoodika võimaldab teil ühendada ärieesmärgid tehnilise rakendamisega, muutes kogu protsessiahela läbipaistvaks ja juhitavaks. Andmete modelleerimine ei ole ühekordne ülesanne, vaid protsess, mis hõlmab järjestikuseid etappe (joonis 4.3-5):

- **Ärinõuete kogumine:** määratletakse peamised ülesanded, eesmärgid ja teabevoold. See on ekspertide ja kasutajatega aktiivse suhtlemise etapp.
- **Entiteetide tuvastamine:** tuuakse välja peamised objektid, kategooriad ja andmetüübid, mida on oluline tulevases süsteemis arvesse võtta.
- **Kontseptuaalse ja loogilise mudeli väljatöötamine:** esmalt võetakse kirja põhilised üksused ja nende seosed, seejärel atribuudid, reeglid ja üksikasjalik struktuur.
- **Füüsiline modelleerimine:** kavandatakse mudeli tehniline rakendamine: tabelid, väljad, seosed, piirangud, indeksid.
- **Andmebaasi loomine:** viimane samm on füüsilise mudeli rakendamine valitud andmebaasisüsteemis, testimine ja töövalmidus.



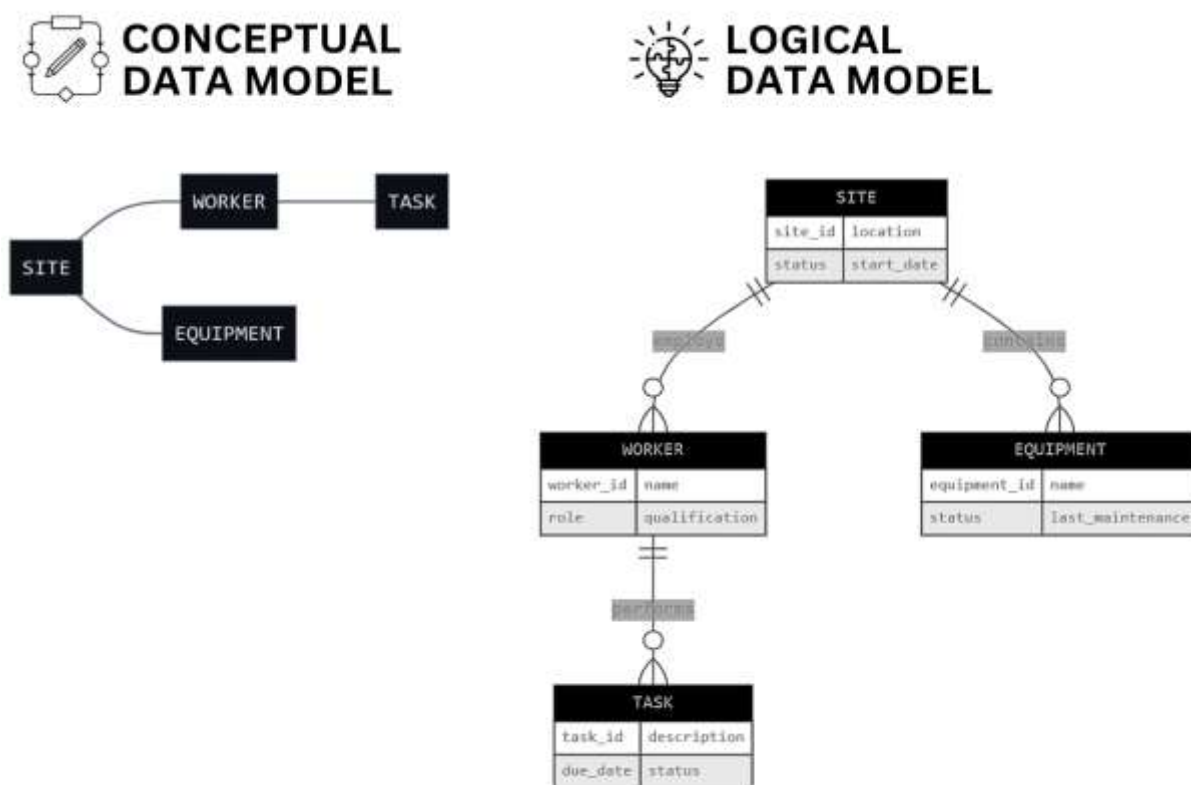
Joonis 4.3-5 Andmebaaside ja andmehaldussüsteemide loomine äriprotsesside jaoks algab nõuete koostamisest ja andmete modelleerimisest.

Nõuetekohaselt kavandatud andmemodelleerimisprotsessid võimaldavad läbipaistvat teabevahetust, mis on eriti oluline keerukates projektides, näiteks ehitusprojektide või ehitusplatside haldamisel. Vaatleme, kuidas üleminek kontseptuaalselt mudelilt loogilisele mudelile ja seejärel füüsilisele mudelile võib aidata protsesse sujuvamaks muuta.

Praktiline andmemodelleerimine ehituse kontekstis

Võtame andmemodelleerimise näitena ehitusplatsi juhtimise ülesande ja teisendame töödejuhataja nõuded struktureeritud loogikamudeliks. Lähtudes ehitusplatsi juhtimise põhivajadustest, määratleme : ehitusplats (SITE), töötajad (WORKER), seadmed (EQUIPMENT), ülesanded (TASK) ja seadmete kasutamine (EQUIPMENT_USAGE). Iga üksus sisaldab hulga atribuute, mis kajastavad olulisi omadusi. Näiteks TASKi puhul võib see olla ülesande kirjeldus, tähtaeg, staatus, prioriteet; WORKERi puhul võib see olla nimi, tema roll objektil, praegune töökoht jne.

Loogikamudel kehtestab nende üksuste vahelised seosed, näidates, kuidas need tegelikes tööprotsessides omavahel suhtlevad (joonis 4.3-6). Näiteks seos objekti ja töötajate vahel näitab, et ühel objektil võib töötada mitu töötajat, samas kui seos töötajate ja ülesannete vahel näitab, et üks töötaja võib täita mitu ülesannet.

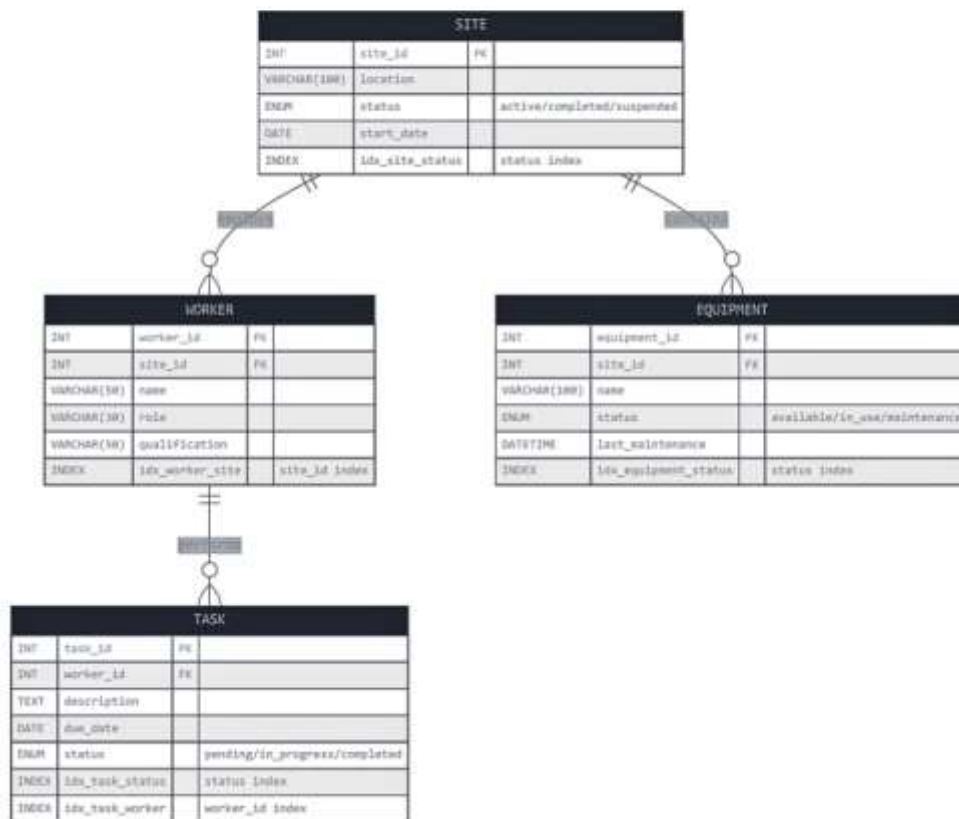


Joonis 4.3-6 Töödejuhi nõuete alusel loodud kontseptuaalne ja loogiline andmemudel ehitusplatsi protsesside kirjeldamiseks.

Füüsilisele mudelile üleminekul lisatakse tehnilised rakendamisdetailid: spetsiifilised andmetüübid (VARCHAR, INT, DATE), primaar- ja võõrvõtmed tabelite vaheliste suhete jaoks ning indeksid andmebaasi jõudluse optimeerimiseks (joonis 4.3-7).

Näiteks tuleks staatuste jaoks määratleda konkreetsed tüübid koos võimalike väärtustega ning lisada indeksid sellistele võtmeväljadele nagu staatus ja worker_id, et parandada otsingu tulemuslikkust. See muudab süsteemi loogilise kirjelduse konkreetseks andmebaasi rakendusplaaniks, mis on valmis loomiseks ja rakendamiseks.

PHYSICAL DATA MODEL



Joonis 4.3-7 Füüsiline andmemudel kirjeldab ehitusplatsi üksusi minimaalsete nõutavate parameetrite kaudu.

Füüsiline mudel erineb sageli loogilisest mudelist. Keskmiselt jaguneb modelleerimisele kuluv aeg järgmiselt: umbes 50% kulub kontseptuaalsele mudelile (nõuete kogumine, protsesside arutamine, üksuste kindlaksmääramine), 10% loogilisele mudelile (atribuutide ja seoste täpsustamine) ja 40% füüsilisele mudelile (rakendamine, testimine, kohandamine DBMSiga).

Selline tasakaal tuleneb asjaolust, et kontseptuaalne etapp paneb aluse andmestruktuurile, samas kui loogiline mudel määrab ainult seosed ja atribuudid. Füüsiline mudel nõuab kõige rohkem ressursse, sest selles etapis rakendatakse andmed konkreetsetele platvormidele ja vahenditele.

Andmebaasi loomine LLM-i abil

Kui meil on olemas andmemudel ja üksuste kirjeldus parameetrite kaudu, oleme valmis looma andmebaase - hoidlaid, kuhu me salvestame pärast struktureerimise etappi saabuvat teavet konkreetsete protsesside kohta.

Proovime luua näite lihtsast, kuid funktsionaalsest andmebaasist minimaalse koodimahuga SQLite abil Pythoni programmeerimiskeele näitel. Relatsioonilisi andmebaase käsitleti üksikasjalikult peatükis "Struktureeritud relatsioonilised andmebaasid ja SQL päringukeel".

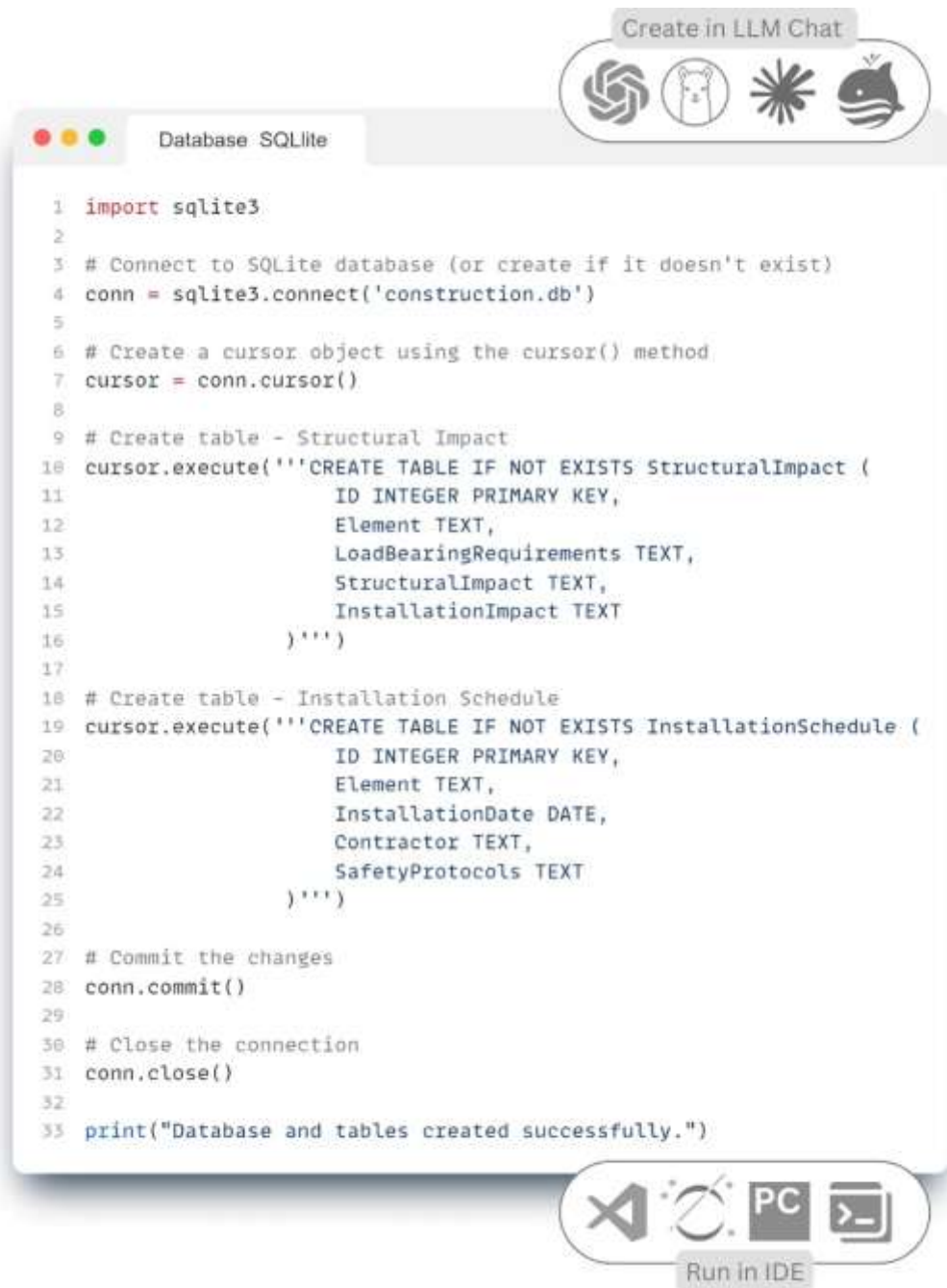
SQLite on kerge, avatud lähtekoodiga, sisseehitatav, platvormideülene andmebaas, mis toetab põhilisi SQL-funktsioone. SQLite'i kasutatakse laialdaselt CAD (BIM), ERP, EPM ja on olemas peaaegu igas nutitelefonis.

Palume LLM mudelil luua meile väike SQLite andmebaas, mis põhineb eespool loodud füüsilisel mudelil (joonis 4.3-4) FEM (struktuuriline mõju) ja CPM (koostegraafik) süsteemide jaoks.

- 🗨️ Saada tekstitaotlus LLM chat (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN või mõni muu):

Luua SQLite andmebaasid süsteemide "Konstruktiivne mõju" ja "Paigaldusgraafik" jaoks manuses oleva joonise järgi (kleebi skeem pildina vestlusesse (joonis 4.3-4). ↵

LLMi vastus :



Joonis 4.3-8 Genereeritud kood loob üleslaetud pildist andmebaasid ja tabelid, millesse saame nüüd andmeid laadida.

Seda andmebaasi koodi saab käivitada Pythoni keskkonnas koos SQLite'i toetusega, mis on tavaliselt Pythoni standardkeskkonnas olemas. Seda saab käivitada mis tahes kohalikul Python-masinal, mis on võimeline skripte käivitama ja faile looma. Pärast käivitamist ja täitmist loob see kood teie kõvakettale faili - SQLite'i andmebaasi nimega construction.db (sarnaselt Exceli failile koos töölehtedega), mis sisaldab tabeleid "Structural Impact" ja "Installation Schedule".

Kui uus **ehitus.db** andmebaas on nendes tabelites loodud, saab andmeid nendesse tabelitesse lisada SQL _päringute või impordi kaudu, mis võimaldab hiljem alustada automaatse andmetöötluse loomist. Andmeid saab SQLite'i andmebaasi importida CSV-failidest, Exceli tabelitest või eksportida teistest andmebaasidest ja hoidlatest API kaudu.

Jätkusuutliku andmemodelleerimise ja tõhusa andmebaaside haldamise protsessi loomiseks vajab ettevõtte selgelt määratletud strateegiat ning tehniliste ja ärieeskondade vahelist kooskõlastamist. Erinevate projektide ja mitmete andmeallikate puhul on sageli raske tagada järjepidevust, standardimist ja kvaliteedikontrolli kõigil tasanditel. Üheks peamiseks lahenduseks võib olla spetsiaalse andmemodelleerimise tippkeskuse loomine ettevõttes.

Andmete modelleerimise tippkeskus (CoE)

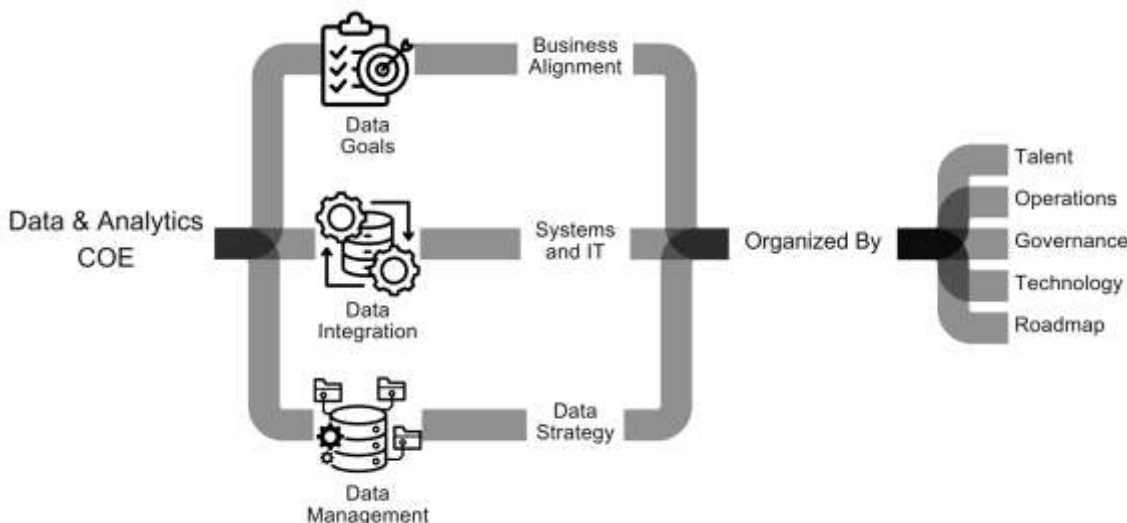
Kuna andmetest on saanud üks peamisi strateegilisi varasid, peavad ettevõtted tegema enamat kui lihtsalt teavet korrektseks koguma ja säilitama - oluline on õppida, kuidas andmeid süstemaatiliselt hallata. Klassifitseerimise ja andmemodelleerimise tippkeskus (CoE) on struktuuriüksus, mis tagab kogu andmekäitluse järjepidevuse, kvaliteedi ja tõhususe organisatsioonis.

Tippkeskus on eksperditoe tuum ja metoodiline alus ettevõtte digitaalseks ümberkujundamiseks. See loob andmepõhise kultuuri ja võimaldab organisatsioonidel luua protsessid, mis põhinevad otsuste tegemisel struktureeritud, valideeritud ja representatiivsetel andmetel, mitte intuitsioonil või kohalikul teabel.

Tippkeskuse moodustavad tavaliselt valdkonnaülesed meeskonnad, kes töötavad "kahe pitsat" põhimõttel. See põhimõte, mille pakkus välja Jeff Bezos, tähendab, et meeskonna suurus peaks olema selline, et seda saaks toita kahe pitsaga, st mitte rohkem kui 6-10 inimest. Selline lähenemine aitab vältida liigset bürokraatiat ja suurendab töö paindlikkust. CoE meeskonda peaksid kuuluma erinevate tehniliste oskustega töötajad, alates andmeanalüütikast ja masinõppe kuni spetsiifiliste ärivaldkondade ekspertideadmisteni. Oma sügavate tehniliste teadmistega peaksid andmetöötajad mitte ainult optimeerima protsesse ja modelleerima andmeid, vaid ka toetama kolleege, vähendades rutiinsete ülesannete täitmisele kuluvat aega (joonis 4.3-9).

Nii nagu looduses tagab ökosüsteemi vastupidavuse bioloogiline mitmekesisus, saavutatakse ka digitaalmaailmas paindlikkus ja kohanemisvõime andmete käitlemise eri lähenemisviiside mitmekesisuse kaudu. Seda mitmekesisust peavad siiski toetama ühised eeskirjad ja mõisted.

Tippkeskust (CoE) võib võrrelda metsaökosüsteemi "kliimatingimustega", mis määravad, millised andmed õitsevad ja millised automaatselt kõrvale jäetakse. Luues kvaliteetsete andmete jaoks soodsa "kliima", hõlbustab tippkeskus parimate tavade ja meetodite loomulikku valimist, mis seejärel muutuvad organisatsiooni standarditeks.



Joonis 4.3-9 Andmete ja analüüsi tippkeskus (CoE) koondab eksperditeadmisi andmete haldamise, integreerimise ja strateegia võtmeaspektide kohta.

Integratsioonitsükli kiirendamiseks ja paremate tulemuste saavutamiseks peaks Euroopa Nõukogu andma oma liikmetele piisava sõltumatuse otsuste tegemisel. See on eriti oluline dünaamilises keskkonnas, kus katsetamine, pidev tagasiside ja sagedased versioonid võivad tuua märkimisväärset kasu. Selline autonoomia on siiski tõhus ainult siis, kui on olemas selge teabevahetus ja toetus kõrgema juhtkonna poolt. Ilma strateegilise nägemuse ja tiptasandi koordineerimiseta võib isegi kõige pädevam meeskond oma algatuste rakendamisel takistuste ees seista.

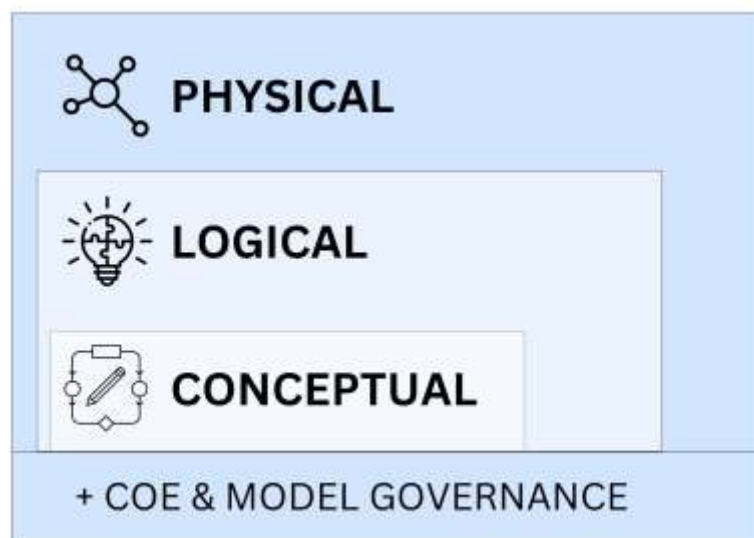
Ettevõtte juhtivtöötaja või kõrgem juhtkond vastutab selle eest, et andmete modelleerimise lähenemisviis ei piirduks ühe või kahe projektiga, vaid oleks integreeritud üldisesse teabehalduse ja äriprotsesside haldamise süsteemi.

Lisaks andmete modelleerimise ja andmehaldusega seotud ülesannetele vastutab ekspertkeskus ühiste standardite ja lähenemisviiside väljatöötamise eest andmete infrastruktuuri kasutuselevõtuks ja toimimiseks. Lisaks sellele edendab ta organisatsioonis pideva täiustamise, protsesside optimeerimise ja andmete tõhusa kasutamise kultuuri (joonis 4.3-10).

Andmete ja mudelite haldamise süstemaatilise lähenemisviisi Euroopa Nõukogu raames võib jagada mitmeks põhiplokiks:

- **Protsesside standardimine ja mudelite elutsükli haldamine:** Regioonide Komitee töötab välja ja rakendab metoodikaid andmemudelite loomise ja haldamise ühtlustamiseks. See hõlmab: struktuurimudelite, kvaliteedikontrolli meetodite ja versioonikontrolli süsteemide loomist, et tagada andmete järjepidevus kõigis tööetappides.

- **Rollide haldamine ja vastutuse määramine:** COE määratleb võtmerollid andmemodelleerimisprotsessis. Iga projektis osalejale määratakse selgelt määratletud rollid ja vastutusvaldkonnad, mis hõlbustab meeskonnatööd ja vähendab andmete vastuolude riski.
- **Kvaliteedikontroll ja auditeerimine:** ehitusandmete tõhus haldamine nõuab nende kvaliteedi pidevat jälgimist. Rakendatakse automatiseeritud mehhanisme andmete kontrollimiseks, vigade ja puuduvate tunnuste tuvastamiseks.
- **Metaandmete ja teabearhitektuuri haldamine:** Regioonide Komitee vastutab ühtse klassifikatsiooni ja identifikaatorite, nimetamise ja üksuste kirjeldamise standardite süsteemi loomise eest, mis on süsteemide vahelise integratsiooni jaoks kriitilise tähtsusega.



Joonis 4.3-10 Andmete modelleerimine ja andmekvaliteedi juhtimine on üks peamisi probleeme Euroopa Nõukogu

Andmete tippkeskus (CoE) ei ole lihtsalt ekspertide rühm, vaid süsteemne mehhanism, mis loob uue andmepõhise kultuuri ja tagab ühtse lähenemisviisi andmetega töötamisele kogu ettevõttes. Tänu modelleerimisprotsesside pidevale integreerimisele üldisesse teabehaldussüsteemi, standardimisele, klassifitseerimisele ja andmete kvaliteedikontrollile, aitab andmekeskuste komitee ettevõtetel pidevalt täiustada oma tooteid ja äriprotsesse, reageerida kiiremini turumuutustele ja teha usaldusväärsetel analüütilistel andmetel põhinevaid teadlikke otsuseid.

Sellised keskused on eriti tõhusad, kui neid kombineeritakse kaasaegsete DataOps-põhimõtetega - all, mis tagab andmete pideva tarnimise, automatiseerimise ja kvaliteedikontrolli. DataOpsist räägime lähemalt 8. osas, peatükis "Kaasaegsed andmetehnoloogiad ehitustööstuses".

Järgnevatel peatükkidel liigume strateegiast praktikasse - "muundume" tinglikult andmekeskuseks: vaatleme mitmeid näiteid, kuidas toimub ülesannete parameetriseerimine, nõuete kogumine ja automaatne valideerimisprotsess.



PEATÜKK 4.4.

NÕUETE SÜSTEMATISEERIMINE JA TEABE VALIDEERIMINE

Nõuete kogumine ja analüüs: teabevahetuse muutmine struktureeritud andmeteks.

Nõuete kogumine ja haldamine on esimene samm andmete kvaliteedi tagamisel. Vaatamata digitaalsete vahendite arengule, sõnastatakse enamik nõudeid endiselt struktureerimata viisil: kirjade, koosolekute protokollide, telefonikõnede ja suuliste arutelude kaudu. Selline suhtlusviis raskendab teabe automatiseerimist, valideerimist ja taaskasutamist. Selles peatükis vaatleme, kuidas tekstipõhised nõuded tõlkida formaalseteks struktuurideks, tagades, et ärinõuded on läbipaistvad ja süstemaatilised.

Gartneri uuringus "Data Quality: Best Practices for Accurate Insights" rõhutatakse andmete kvaliteedi kriitilist tähtsust edukate andmete ja analüütiliste algatuste jaoks [100]. Nad märgivad, et kehv andmete kvaliteet läheb organisatsioonidele aastas keskmiselt vähemalt 12,9 miljonit dollarit maksma ning et usaldusväärsed ja kvaliteetsed andmed on andmepõhise ettevõtte loomiseks hädavajalikud.

Struktureeritud nõuete puudumine toob kaasa selle, et sama element (üksus) ja selle parameetrid võivad olla salvestatud eri süsteemides erinevates variatsioonides. See mitte ainult ei vähenda protsesside tõhusust, vaid põhjustab ka aja raiskamist, teabe dubleerimist ja vajadust andmeid enne nende kasutamist uuesti valideerida. Selle tulemusel võib isegi üksainus puudus - kadunud parameeter või üksainus valesti kirjeldatud element - aeglustada otsuste tegemist ja põhjustada ressursside ebaefektiivset kasutamist.

*Küünte puudumisel jäi hobuseraua kaduma.
Hobuseraudade puudumisel oli hobune kadunud.
Hobuse puudumisel kaotati ratsanik.
Ratsaniku puudumisel läks kaduma sõnum.
Sõnumi puudumisel kaotati lahing.
Lahingu puudumisel kaotati kuningriik.
Kõik see toimus hobuseraua naela puudumise tõttu.*

- Vanasõna [101]

Andmete täitmise ja säilitamise protsessi nõuete analüüsimine ja kogumine algab kõigi sidusrühmade kindlakstegemisega. Nii nagu vanasõnaline ühe naela kaotamine toob kaasa kriitiliste tagajärgede ahela, võib äritegevuses ühe sidusrühma kaotamine, tähelepanuta jäetud nõue või isegi ühe parameetri kadumine mõjutada oluliselt mitte ainult üksikut äriprotsessi, vaid kogu projekti ökosüsteemi ja organisatsiooni kui tervikut. Seetõttu on väga oluline tuvastada isegi need elemendid, parameetrid ja rollid, mis esmapilgul tunduvad ebaolulised, kuid võivad hiljem osutuda äritegevuse jätkusuutlikkuse

seisukohalt kriitilise tähtsusega.

Kujutame ette, et ettevõttel on projekt, mille puhul klient esitab uue taotluse - "*lisage hoone põhjaküljele täiendav aken*". Väikeses protsessis "*kliendi taotlus lisada praegusele projektile uus aken*" osalevad arhitekt, klient, CAD-spetsialist (BIM), ehitusjuht, logistikajuht, ERP -analüütik, kvaliteedikontrolli insener, ohutusinsener, kontrollijuht ja kinnisvarajuhataja.

Isegi väike protsess võib hõlmata kümneid erinevaid spetsialiste. Iga protsessis osaleja peab mõistma spetsialistide nõudeid, kellega ta on andmete tasandil seotud.

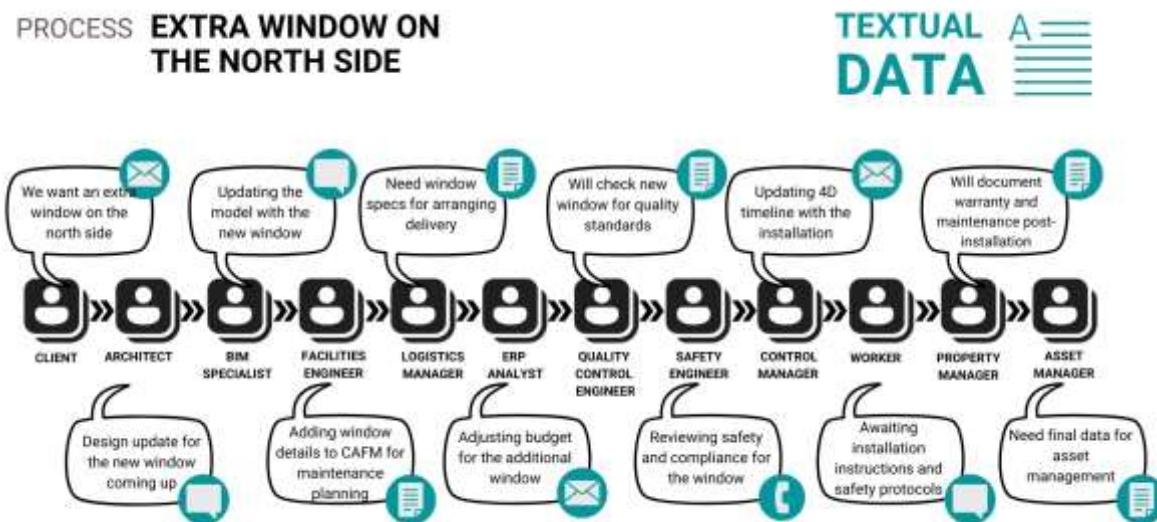
Teksti tasandil (joonis 4.4-1) toimub kliendi ja protsessiahela spetsialistide vaheline suhtlus järgmiselt:

- 🗣️ **Klient:** "*Otsustasime lisada põhjaküljele täiendava akna parema valgustuse tagamiseks. Kas seda on võimalik realiseerida?*"
- 🗣️ **Arhitekt:** "*Muidugi, ma vaatan projekti üle, et lisada uus aken, ja saadan ajakohastatud CAD-plaanid (BIM)*".
- 🗣️ **CAD-spetsialist (BIM):** "*Saanud uue projekti. Uuendan CAD (BIM) mudelit täiendava aknaga ja pärast kooskõlastamist FEM-inseneriga annan uue akna täpse asukoha ja mõõtmed*".
- 🗣️ **Ehitusjuht:** "*On saadud uus projekt. Korrigeerime 4D paigalduskuupäevi ja teavitame kõiki asjaomaseid alltöövõtjaid*".
- 🗣️ **Rajatiste insener (CAFM):** "*Sisestan 6D andmed uue akna kohta CAFM-süsteemi tulevase rajatiste haldamise ja hoolduse planeerimise jaoks*".
- 🗣️ **Logistikajuht:** "*Mul on vaja uue akna mõõtmeid ja kaalu, et korraldada akna kohale toimetamist*".
- 🗣️ **ERP -analüütik:** "*Mul on vaja 5D eelarve uuendamise jaoks meie ERP-süsteemis reguleerimisala tabeleid ja täpset aknatüüpi, et kajastada uue akna maksumust üldises projekti hinnangus*".
- 🗣️ **Kvaliteedikontrolliinsener:** "*Kui akende spetsifikatsioonid on valmis, veendun, et need vastavad meie kvaliteedi- ja materjalistandarditele*".
- 🗣️ **Ohutusinsener:** "*Ma hindan uue akna ohutusaspekte, keskendudes eelkõige nõuetele vastavusele ja evakueerimisele 8D-kava alusel*".
- 🗣️ **Kontrollijuht:** "*Lähtudes ERP täpsest töömahust, uuendame oma 4D-ajagraafikut, et kajastada uue akna paigaldamist, ja salvestame uued andmed projekti sisuhaldussüsteemis*".
- 🗣️ **Töötaja (paigaldaja):** "*Te vajate juhiseid paigaldamise, kokkupaneku ja tööde ajastamise kohta. Kas lisaks sellele on kehtestatud mingeid erilisi ohutuseeskirju, mida ma pean järgima?*"
- 🗣️ **Kinnisvarahaldur:** "*Pärast paigaldamist dokumenteerin garantii- ja hooldusteabe pikaajaliseks haldamiseks*".
- 🗣️ **Varahaldur:** "*Seadmete insener, palun saatke lõplikud andmed vara jälgimiseks ja elutsükli haldamiseks*".
- 🗣️ **Klient:** "*Oota, võib-olla mul on kiire ja akent ei olegi vaja. Võib-olla peaksin tegema rõdu*".

Selliste stsenaariumide puhul, mis toimuvad sageli, põhjustab isegi väike muutus ahelreaktsiooni mitme süsteemi ja rolli vahel. Sellisel juhul on peaaegu kogu suhtlus algfaasis tekstina: e-kirjad, vestlused, koosolekute protokollid (joonis 4.4-1).

Sellise ehitusprojekti tekstipõhise kommunikatsioonisüsteemi puhul on väga oluline kõigi andmevahetustoimingute ja tehtud otsuste õigusliku kinnitamise ja registreerimise süsteem. Sellega tagatakse, et iga tehtud otsus, juhised või muudatus on õiguslikult kehtiv ja jälgitav, vähendades tulevaste

"arusaamatuste" riski

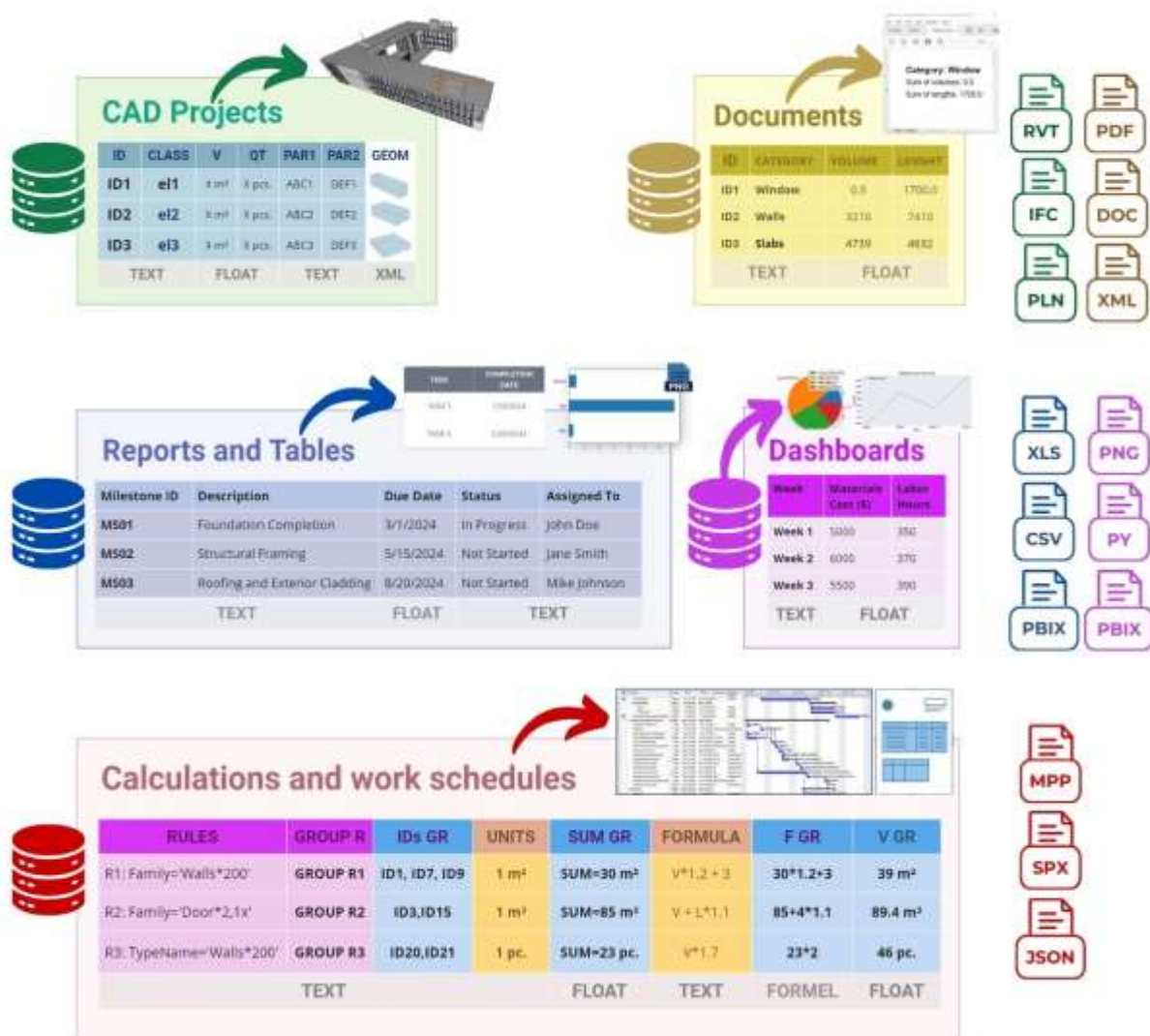


Joonis 4.4-1 Tellija ja töövõtja vaheline suhtlus projekti algfaasis sisaldab sageli mitmevormilisi tekstilisi andmeid.

Ehitusprojekti asjaomaste süsteemide õigusliku kontrolli ja otsuste kinnitamise puudumine võib põhjustada tõsiseid probleeme kõigile asjaosalistele. Iga otsus, korraldus või muudatus, mis on tehtud ilma nõuetekohase dokumentatsiooni ja kinnitamiseta, võib põhjustada vaidlusi (ja kohtuvaidlusi).

Kõigi otsuste juriidilist konsolideerimist tekstilise suhtluse teel saab tagada ainult suure hulga allkirjastatud dokumentide abil, mis langeb juhtkonna õlgadele, kes on kohustatud kõik tehingud registreerima. Kui iga osaleja peab iga toiminguga seotud dokumente allkirjastama, kaotab süsteem paindlikkuse ja muutub bürookraatlikuks labürindiks. Tehingukinnituste puudumine ei lükka mitte ainult projekti elluviimist edasi, vaid võib põhjustada ka rahalist kahju ja halvendada osalejate vahelisi suhteid, kuni juriidiliste probleemideni välja.

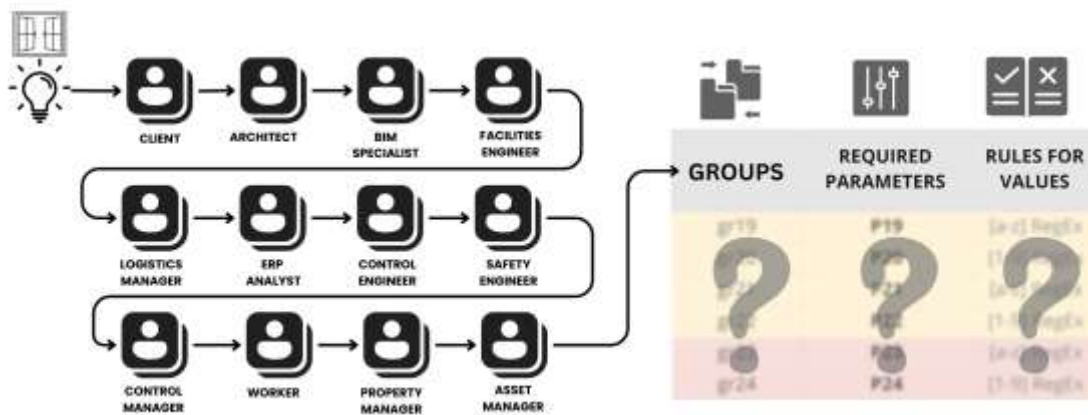
Selline tehingu heakskiitmise protsess, mis tavaliselt algab tekstipõhiste aruteludega, areneb järgmistes etappides järk-järgult mitmeformaadiliseks dokumendivahetuseks (joonis 4.4-2), mis muudab varem ainult teksti teel toimuva suhtluse oluliselt keerulisemaks. Ilma selgelt määratletud nõueteta muutub selliste protsesside automatiseerimine, mis on täis mitmeformaadilisi andmeid ja suurt hulka tekstilisi nõudeid, peaaegu võimatuks.



Joonis 4.4-2 Iga süsteem ehitusettevõtte maastikul toimib õiguslikult asjakohaste dokumentide allikana erinevates vormingutes.

Tekstisõnumid nõuavad, et iga spetsialist kas tutvuks kogu kirjavahetusega või osaleks regulaarselt kõigil koosolekutel, et saada aru projekti hetkeseisust.

Selle piirangu ületamiseks on vajalik üleminek tekstipõhiselt suhtluselt struktureeritud nõudlusmodelile. See on võimalik ainult süstemaatilise analüüsi, protsesside visualiseerimise ja koostoidete kirjeldamise kaudu vooskeemide ja andmemudelite kujul (joonis 4.4-3). Nii nagu andmete modelleerimisel (joonis 4.3-7), liikusime kontekstuaalselt ideede tasandilt kontseptuaalsele tasandile, lisades osalejate poolt kasutatavad süsteemid ja vahendid ning nendevahelised seosed.



Joonis 4.4-3 Selleks, et õppida valideerimisprotsessi juhtimist ja automatiseerimist, tuleb protsessid visualiseerida ja nõuded struktureerida.

Esimene samm nõuete ja suhete süstematiseerimisel on kõikide seoste ja suhete visualiseerimine kontseptuaalsete vooskeemide abil. Kontseptuaalsel tasandil ei tee mitte ainult kõigile protsessis osalejatele kogu protsessiahela mõistmist lihtsamaks, vaid näitab ka selgelt, miks ja kelle jaoks on andmeid (ja nõudeid) igas protsessietapis vaja.

Protsesside vooskeemid ja kontseptuaalsete raamistike tõhusus

Selleks, et ületada lõhe traditsioonilise ja kaasaegse andmehaldusviisi vahel, peavad ettevõtted teadlikult üle minema killustatud tekstikirjeldustelt struktureeritud protsesside esitusviisidele. Andmete areng - savitahvlitelt digitaalsete ökosüsteemideni - nõuab uusi mõtlemisvahendeid. Üks selline vahend on kontseptuaalne modelleerimine vooskeemide abil. Visuaalsete skeemide - vooskeemide, protsessiskeemide, koostoimeskeemide - loomine võimaldab projektis osalejatel mõista, kuidas nende tegevused ja otsused mõjutavad kogu otsustussüsteemi.

Kui protsessid nõuavad mitte ainult andmete säilitamist, vaid ka nende analüüsimist või automatiseerimist, siis tuleb hakata tegelema kontseptuaalse ja visuaalse nõuete kihi loomisega.

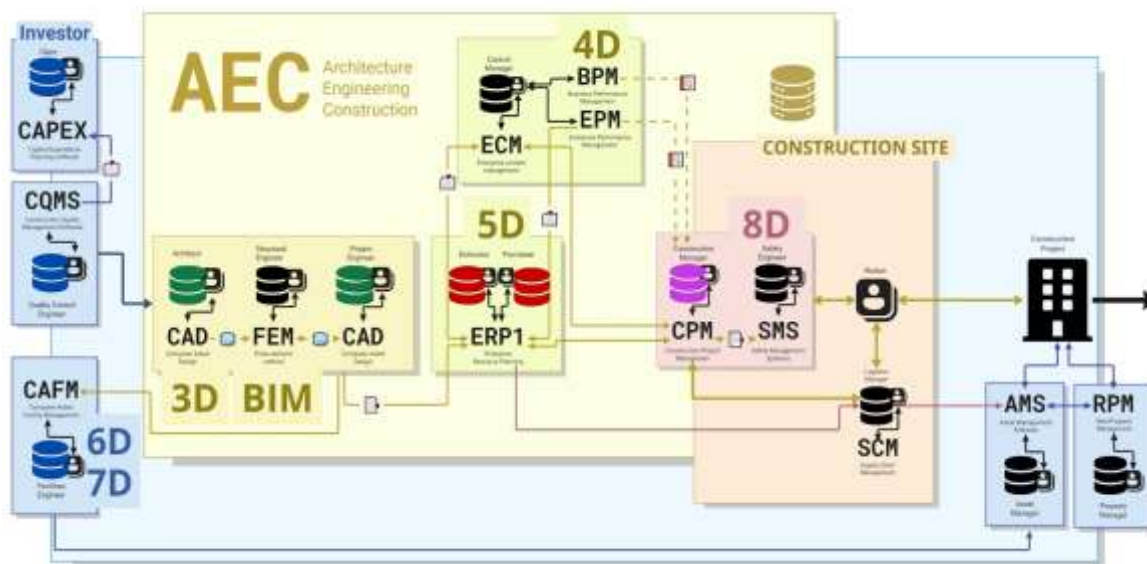
Meie näites (joonis 4.4-1) võib iga spetsialist kuuluda väikesesse meeskonda, kuid ka suuremasse osakonda, mis hõlmab kuni tosinat eksperti, kes on üldjuhi alluvuses. Iga osakond kasutab spetsiaalset rakendusandmebaasi (joonis 1.2-4, nt ERP, CAD, MEP, CDE, ECM, CPM jne), mida ajakohastatakse regulaarselt sissetuleva teabega, mis on vajalik dokumentide koostamiseks, otsuste õigusliku seisuga registreerimiseks ja protsesside haldamiseks.

Tehinguprotsess sarnaneb 4000 aastat tagasi toimunud iidsete juhtide tööga, kui otsuste õiguslikuks kinnitamiseks kasutati savitahvleid ja papüürusi. Erinevus tänapäeva süsteemide ja nende savist ja paberist eelkäijate vahel seisneb selles, et tänapäevased meetodid hõlmavad lisaks ka tekstilise teabe teisendamist digitaalsesse vormi, et seda muudes süsteemides ja töövahendites automaatselt edasi töödelda.

Protsessi visualiseerimine kontseptuaalsete vooskeemide kujul aitab kirjeldada iga sammu ja eri rollide vahelist koostoimet, muutes keerulise töövoogu selgeks ja lihtsaks.

Protsesside visualiseerimine tagab, et protsessiloomika on läbipaistev ja kõigile meeskonnaliikmetele kättesaadav.

Sama kommunikatiivne protsess akna lisamiseks projektile, mida kirjeldati teksti, sõnumite (joonis 4.4-1) ja plokkiskeemi kujul, on sarnane kontseptuaalse mudeliga, mida me arutasime andmete modelleerimise peatükis (joonis 4.4-4).



Joonis 4.4-4 Kontseptuaalne diagramm näitab projektis osalejaid kui andmebaasi kasutajaid, kelle päringud ühendavad erinevaid süsteeme.

Kuigi kontseptuaalsed diagrammid on oluline samm, piirduvad paljud ettevõtted selle tasemega, arvates, et protsesside mõistmiseks piisab visuaalsest diagrammist. See tekitab illusiooni juhitavusest: juhid saavad sellisel vooskeemil kergemini tajuda suurt pilti ja näha osalejate ja etappide vahelisi seoseid. Sellised skeemid ei anna aga selget ettekujutust sellest, milliseid andmeid iga osaleja kohta on vaja, millises formaadis need tuleb edastada ning millised parameetrid ja atribuudid on automaatika realiseerimiseks kohustuslikud. Kontseptuaalne vooskeem on pigem nagu marsruudikaart: see näitab, kes kellega suhtleb, kuid ei näita, mida selle suhtluse käigus edastatakse.

Isegi kui protsess on kontseptuaalsel tasandil üksikasjalikult kirjeldatud, kasutades vooskeeme, ei garanteeri see selle tõhusust. Visualiseerimine lihtsustab sageli juhtide tööd, võimaldades neil protsessi hõlpsamini jälgida samm-sammult aruandluse abil. Andmebaasinseneride jaoks ei pruugi kontseptuaalne esitus olla aga piisavalt selge ega anna selget arusaama sellest, kuidas protsessi parameetrite ja nõuete tasandil rakendada.

Kuna me liigume keerukamate andmete ökosüsteemide suunas, muutub kontseptuaalsete ja visuaalsete vahendite esialgne rakendamine kriitiliseks, et tagada, et andmeprotsessid ei oleks mitte ainult tõhusad, vaid ka kooskõlas organisatsiooni strateegiliste eesmärkidega. Selleks, et see akna lisamise protsess (joonis 4.4-1) täielikult andmenõuete tasandile üle kanda, tuleb minna tasandit sügavamale ja tõlkida protsessi kontseptuaalne visualiseerimine andmete, nõutavate atribuutide ja nende piirväärtuste loogilisele ja füüsilisele tasandile.

Struktureeritud nõuded ja RegEx regulaaravaldised

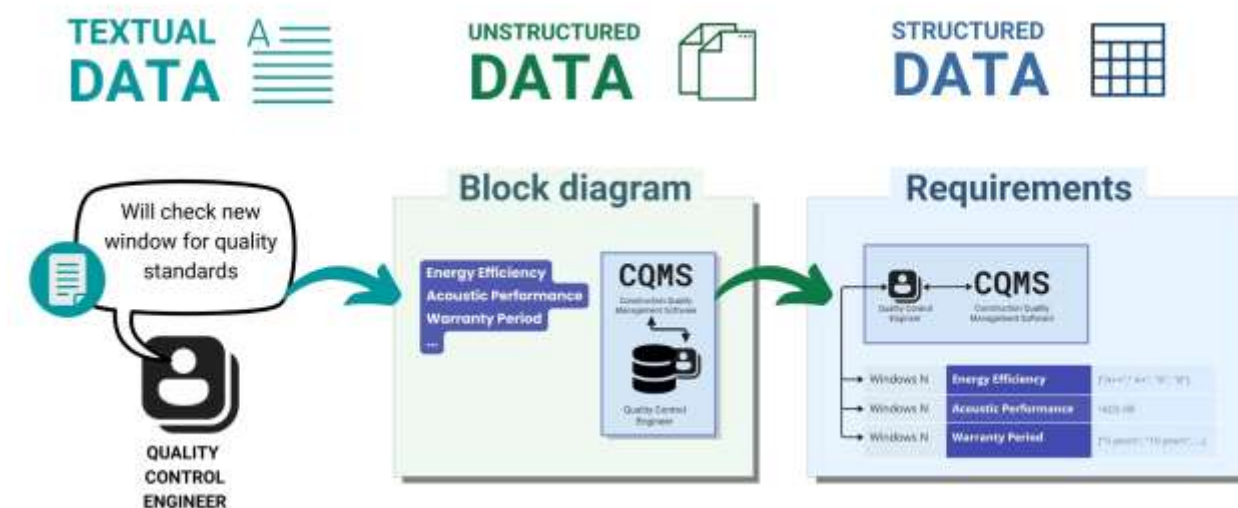
Kuni 80% ettevõtetes loodud andmetest on struktureerimata või poolstruktureeritud kujul [52] - tekst, dokumendid, kirjad, PDF -failid, vestlused. Selliseid andmeid (joonis 4.4-1) on raske analüüsida, kontrollida, süsteemidevaheliselt edastada ja automatiseerimisel kasutada.

Haldatavuse, läbipaistvuse ja automaatse valideerimise tagamiseks on vaja tõlkida tekstilised ja poolstruktureeritud nõuded hästi määratletud, struktureeritud vormingutesse. Struktureerimisprotsess ei puuduta mitte ainult andmeid (mida oleme üksikasjalikult käsitlenud raamatu käesoleva osa esimestes peatükkides), vaid ka nõudeid endid, mida projektis osalejad sõnastavad tavaliselt kogu projekti elutsükli jooksul vabas tekstis, sageli mõtlemata sellele, et neid protsesse saab automatiseerida.

Nii nagu me juba teisendasime andmed struktureerimata tekstilisest vormist struktureeritud vormi, teisendame ka nõuete tööprotsessis tekstilised nõuded struktureeritud "loogilise ja füüsilise kihi" formaati.

Akna lisamise näites (joonis 4.4-1) on järgmine samm andmenõuete kirjeldamine tabeli kujul. Me struktureerime teavet iga süsteemi kohta, mida projektis osalejad kasutavad, määrates kindlaks peamised atribuudid ja nende piirväärtused

Võtame näiteks ühe sellise süsteemi (joonis 4.4-5) - ehituskvaliteedi juhtimissüsteemi (CQMS), mida kasutab tellija poolne kvaliteedikontrolli insener. Selle abil kontrollib ta, kas projekti uus element - antud juhul "uus aken" - vastab kehtestatud standarditele ja nõuetele.



Joonis 4.4-5 Tekstipõhiste nõuete teisendamine tabelivormingusse koos olemite atribuutide kirjeldustega lihtsustab teiste spetsialistide jaoks arusaamist.

Näitena võib tuua mõned olulised nõuded CQMS -süsteemi "aknasüsteemid" tüüpi üksuste atribuutidele (joonis 4.4-6): energiatõhusus, akustilised omadused ja garantiiaeg. Iga kategooria sisaldab teatavaid standardeid ja spetsifikatsioone, mida tuleb arvestada aknasüsteemide projekteerimisel ja paigaldamisel.

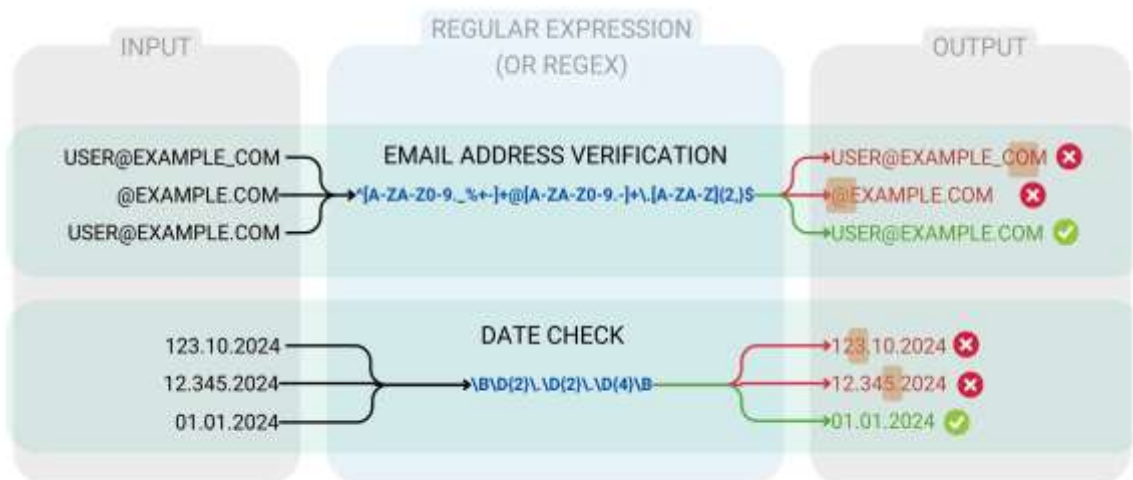


Joonis 4.4-6 Kvaliteedikontrollija peaks kontrollima uusi aknatüüpi elemente energiatõhususe, heliisolatsiooni ja garantiistandardite osas.

Andmenõuded, mida kvaliteedikontrolli insener määrab tabeli kujul, on näiteks järgmised piirväärtused:

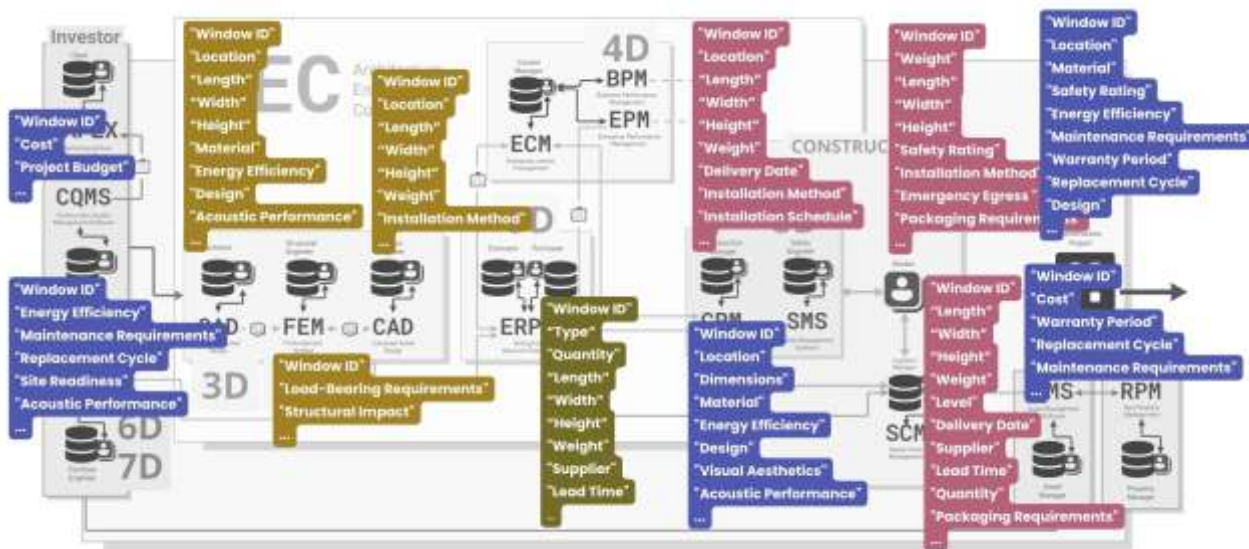
- **Akende energiatõhususe klass** ulatub "A++", mis tähistab kõrgeimat energiatõhusust, kuni "B", mida peetakse minimaalseks vastuvõetavaks tasemeks, ning need klassid on esitatud vastuvõetavate väärtuste loeteluga ["A++", "A+", "A", "A", "A", "A", "B"].

kontrollida kontrollimata jäänud andmed automaatselt kõrvale jätta ja kontrollitud andmed automaatselt süsteemidesse edasiseks töötlemiseks edastada.



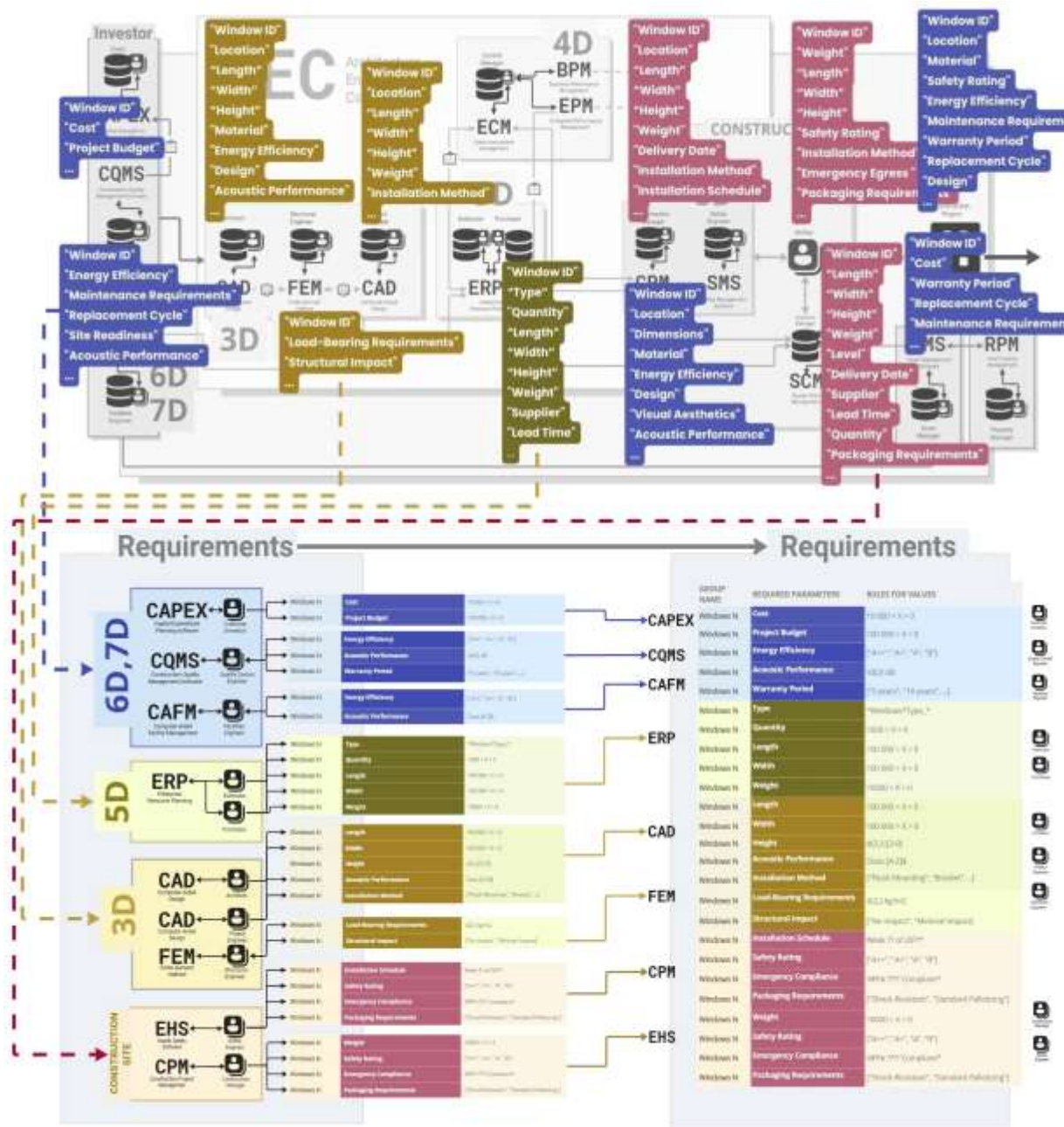
Joonis 4.4-7 Regulaaravaldiste kasutamine on äärmiselt oluline vahend tekstiandmete valideerimise protsessis.

Nüüd, liikudes nõudmistega töötamise kontseptuaalselt tasandilt loogilisele tasandile, teisendame meie uue akna paigaldamise protsessis (joonis 4.4-4) kõikide spetsialistide nõuded järjestatud loeteluks atribuutide kujul ja lisame need loetelud koos vajalike atribuutidega meie vooluskeemile iga spetsialisti kohta (joonis 4.4-8).



Joonis 4.4-8 Loogilise protsessi tasandil lisatakse iga spetsialisti poolt käsitletavad atribuudid vastavatesse süsteemidesse.

Lisades kõik atribuudid ühele ühisele protsessitabelile, muudame eelnevalt kontseptuaalsel tasandil (joonis 4.4-1) teksti ja dialoogina esitatud teabe struktureeritud ja süstematiseeritud kujul füüsilise tasandi tabeliteks (joonis 4.4-9).



Joonis 4.4-9 Struktureerimata spetsialistide dialoogi teisendamine struktureeritud tabeliteks aitab mõista nõudeid füüsilisel tasandil.

Nüüd tuleb andmenõuded edastada spetsialistidele, kes loovad teavet konkreetsete süsteemide jaoks. Näiteks kui te töötate CAD-andmebaasis, siis enne elementide modelleerimise alustamist peaksite koguma kõik vajalikud parameetrid, lähtudes andmete lõppkasutuse stsenaariumidest. See algab tavaliselt operatiivfaasist, millele järgneb ehitusplats, logistikaosakond, kalkulatsiooniosakond, konstruktsiooniaruutuste osakond jne. Alles pärast seda, kui olete võtnud arvesse kõigi nende valdkondade nõudeid, saate alustada andmete loomist - kogutud parameetrite põhjal. See võimaldab teil tulevikus automatiseerida andmete kontrollimist ja edastamist mööda ahelat.

Kui uued andmed vastavad nõuetele, integreeritakse need automaatselt ettevõtte andmete ökosüsteemi, suunates need otse kasutajatele ja süsteemidele, mille jaoks need on mõeldud. Andmete kontrollimine atribuutide ja nende väärtuste suhtes tagab, et teave vastab nõutud kvaliteedistandarditele ja on valmis kasutamiseks ettevõtte stsenaariumides.

Andmenõuded on määratletud ja nüüd, enne kui tõendamine saab alata, tuleb luua, hankida või koguda tõendatavad andmed või registreerida andmebaasides oleva teabe hetkeseis, et seda saaks kasutada tõendamisprotsessis.

Andmete kogumine kontrollimiseks

Enne valideerimise alustamist on oluline tagada, et andmed oleksid valideerimiseks sobival kujul kättesaadavad. See ei tähenda mitte lihtsalt teabe olemasolu, vaid selle ettevalmistamist: andmed tuleb koguda ja muuta struktureerimata, vabalt struktureeritud, tekstilisest ja geomeetrisest vormingust struktureeritud kujul. Seda protsessi on üksikasjalikult kirjeldatud eelmistes peatükkides, kus käsitleti eri liiki andmete teisendamise meetodeid. Kõigi teisenduste tulemusena võtavad saabuval andmed avatud struktureeritud tabelite kuju (joonis 4.1-2, joonis 4.1-9, joonis 4.1-13).

Nõuded ja struktureeritud tabelid vajalike parameetrite ja piirväärtustega (joonis 4.4-9), saame alustada andmete valideerimist - kas ühtse automatiseeritud protsessina (Pipeline) või iga sissetuleva dokumendi etapiisilise valideerimisena.

Kontrolli käivitamiseks on vaja kas saada sisendiks uus fail või fikseerida andmete hetkeseis - luua hetkeülesvõtte või eksportida praegused ja sissetulevad andmed või luua ühendus välise või sisemise andmebaasiga. Vaatlusaluses näites luuakse selline hetkeseis, teisendades automaatselt CAD-andmed aadressilt struktureeritud vormingusse, mis on salvestatud näiteks reedel, 29. märtsil 2024 kell 23:00:00, kui kõik projekteerijad on koju läinud.

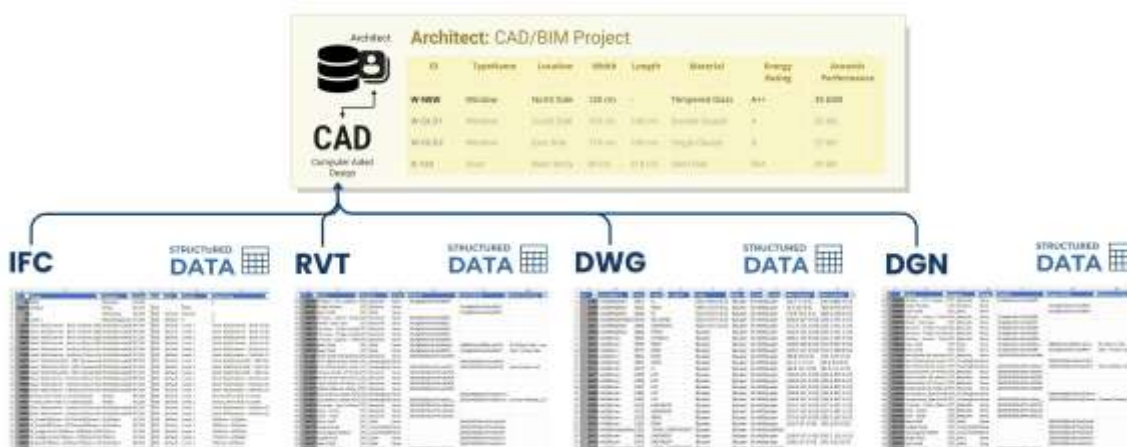


Joonis 4.4-10 CAD-andmebaasi hetkeülesvõtte (BIM), mis näitab uue, klassi "Window" kuuluva üksuse praegust atribuutteavet projektumodeli praeguses versioonis.

Tänu peatükis "CAD-andmete (BIM) teisendamine struktureeritud kujul" käsitletud pöördprojekteerimisvahenditele saab seda erinevatest CAD (BIM) vahenditest ja redaktoritest saadud teavet korraldada eraldi tabelitesse (joonis 4.4-11) või ühendada ühte ühisesse tabelisse, mis ühendab

projekti eri osad (joonis 9.1-10).

Selline tabel - andmebaas näitab akende ja uste unikaalseid identifikaatoreid (ID atribuut), tüübinimesid (TypeName), mõõtmeid (Width, Length), materjale (Material), samuti energia- ja akustilisi näitajaid ja muid omadusi. Sellise CAD-programmis (BIM) täidetud tabeli kogub projekteerija erinevatest osakondadest ja dokumentidest, moodustades projekti infomudeli.



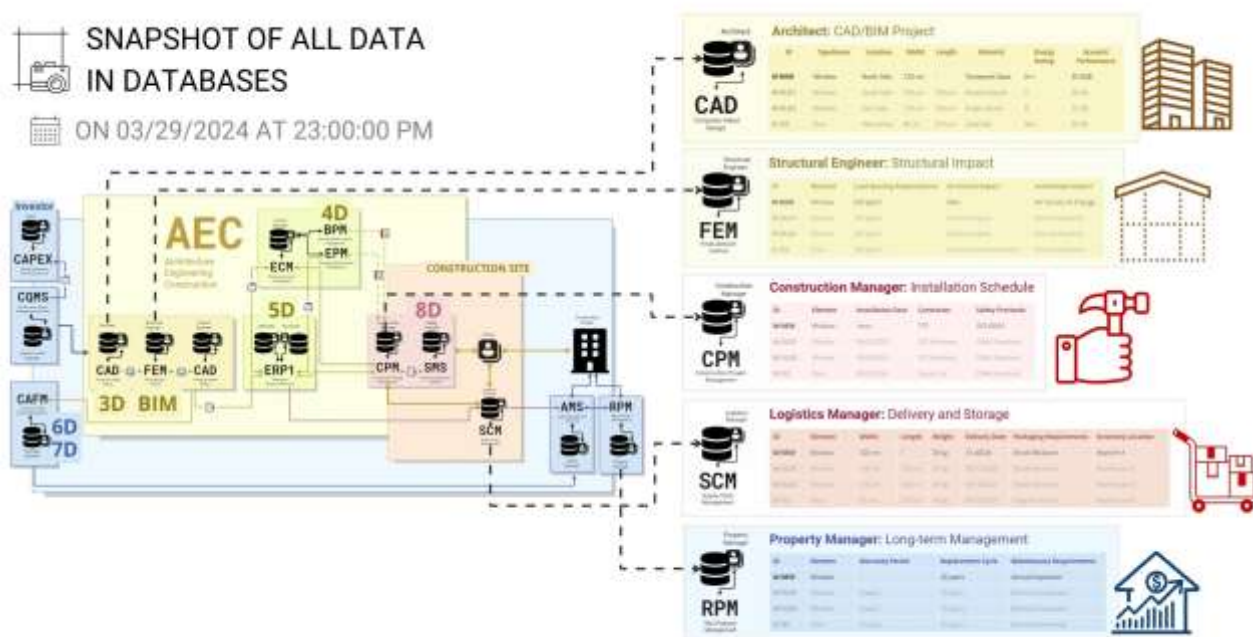
Joonis 4.4-11 Struktureeritud andmed CAD-süsteemidest võivad olla kahemõõtmeline tabel, mille veerud tähistavad elementide atribuute.

Reaalsed CAD (BIM) projektid sisaldavad kümneid või sadu tuhandeid elemente (joonis 9.1-10). CAD-formaadis olevad elemendid liigitatakse automaatselt tüübi ja kategooria järgi - alates akendest ja udest kuni plaatide, tahvlite ja seinteni. Unikaalsed identifikaatorid (nt algupärane ID, mille CAD-lahendus määrab automaatselt) või tüübi atribuudid (tüübi nimi, tüüp, perekond) võimaldavad sama objekti jälgimist erinevates süsteemides. Näiteks saab uue akna hoone põhjaseinas üheselt identifitseerida üheainsa identifikaatoriga "W-NEW" kõigis organisatsiooni asjaomastes süsteemides.

Kuigi üksuste nimed ja identifikaatorid peaksid olema kõikides süsteemides ühtsed, võivad nende üksustega seotud atribuutide ja väärtuste kogum sõltuvalt kasutuskontekstist oluliselt erineda. Arhitektid, ehitusinsenerid, ehitus- ja logistika- ja kinnisvarahoolduse spetsialistid tajuvad samu elemente erinevalt. Igaüks neist tugineb oma klassifikaatoritele, standarditele ja eesmärkidele: mõned vaatavad akent puhtalt esteetilisest vaatepunktist, hinnates selle kuju ja proportsioone, samas kui teised vaatavad seda tehnilisest või kasutusotstarbelisest vaatepunktist, analüüsides soojusjuhtivust, paigaldusmeetodit, kaalu või hooldusnõudeid. Seetõttu on andmete modelleerimisel ja elementide kirjeldamisel oluline võtta arvesse nende mitmekülgset kasutusvõimalust ja tagada andmete järjepidevus, võttes samal ajal arvesse tööstusharu eripärasid.

Iga rolli jaoks ettevõtte protsessides on olemas spetsiaalsed andmebaasid oma kasutajaliidesega - alates projekteerimisest ja arvutustest kuni logistika, paigaldamise ja hoone käitamiseni (joonis 4.4-12). Iga sellist süsteemi haldab professionaalne spetsialistide meeskond spetsiaalse kasutajaliidese või andmebaasi päringute kaudu, kus kõigi sisestatud väärtuste kohta tehtud otsuste summaarse

täitmise taga on süsteemi juht või osakonnajuhataja, kes vastutab sisestatud andmete õiguspärasuse ja kvaliteedi eest oma teisi süsteeme teenindavate vastaspoolte ees.



Joonis 4.4-12 Samal üksusel on erinevates süsteemides sama identifikaator, kuid erinevad atribuudid, mis on olulised ainult selles süsteemis.

Kui me oleme struktureeritud nõuete ja andmete kogumise loogilisel ja füüsilisel tasandil korraldanud, jääb meile üle luua protsess, mille abil saame automaatselt kontrollida erinevatest sissetulevatest dokumentidest ja erinevatest süsteemidest pärinevaid andmeid eelnevalt kogutud nõuete suhtes.

andmete kontrollimine ja kontrollimise tulemused

Kõik süsteemi sisenevad uued andmed - olgu need siis dokumendid, tabelid või andmebaasi kirjed kliendilt, arhitektilt, insenerilt, töödejuhatajalt, logistikult või kinnisvarahaldurilt - peavad olema valideeritud eelnevalt sõnastatud nõuete alusel (joonis 4.4-9). Valideerimisprotsess on kriitilise tähtsusega: mis tahes vead andmetes võivad põhjustada ebaõigeid arvutusi, ajakava hilinemist ja isegi rahalist kahju. Selliste riskide minimeerimiseks tuleks korraldada süstemaatiline ja korduv, iteratiivne andmete valideerimismenetlus.

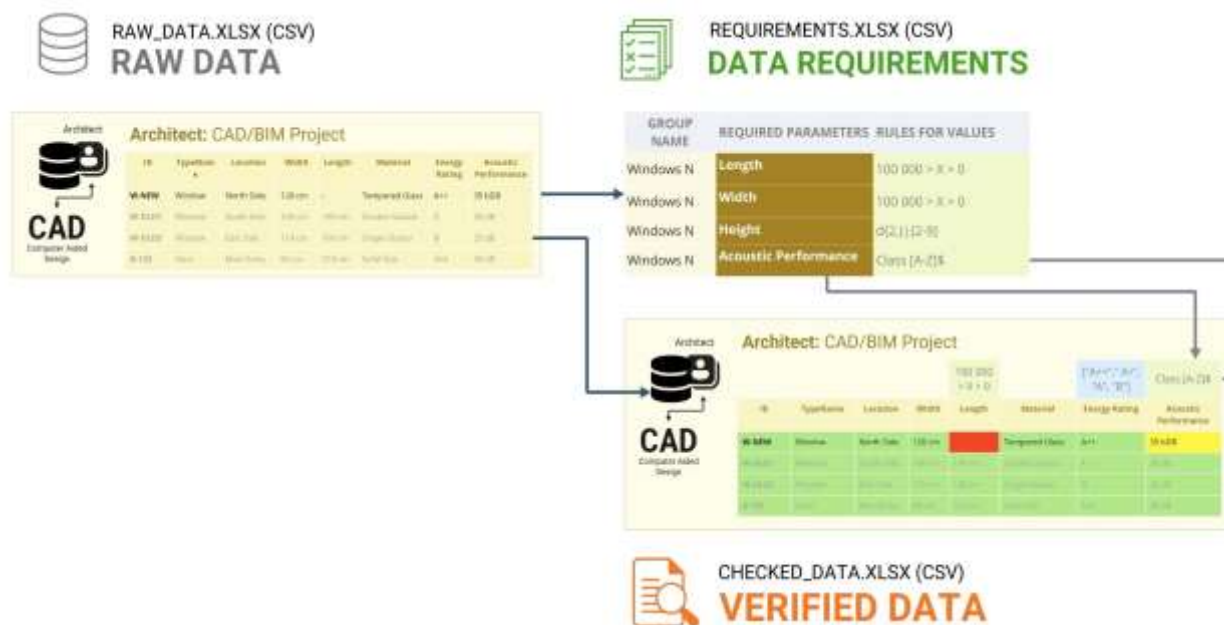
Süsteemi sisenevate uute - struktureerimata, tekstiliste või geomeetriliste - andmete valideerimiseks tuleb need teisendada vabalt struktureeritud või struktureeritud vormingusse. Seejärel peab valideerimisprotsess kontrollima andmeid nõutavate atribuutide ja nende lubatud väärtuste täieliku loetelu alusel.

Erinevat tüüpi andmete: teksti, piltide, PDF -dokumentide ja segatud CAD (BIM) andmete teisendamist

struktureeritud kujul käsitleti üksikasjalikult peatükis "Andmete teisendamine struktureeritud kujul".

Näitena võib tuua CAD-projektist (BIM) saadud tabeli (joonis 4.4-11). See sisaldab poolstruktureeritud geomeetrilisi andmeid ja struktureeritud atribuutteavet projektiüksuste kohta (joonis 3.1-14) - nt element klassist "Windows".

Valideerimise teostamiseks võrdleme atribuutide väärtusi (joonis 4.4-11) ekspertide poolt nõuete kujul määratletud võrdluspiiride väärtustega (joonis 4.4-9). Lõplikust võrdlustabelist (joonis 4.4-13) selgub, millised väärtused on vastuvõetavad ja milliseid tuleb korrigeerida, enne kui andmeid saab kasutada väljaspool CAD-rakendusi (BIM).



Joonis 4.4-13 Lõplikus valideerimistabelis on esile toodud need klassi "Windows" uue olemuse atribuutide väärtused, millele peaksite tähelepanu pöörama.

Rakendades sarnase lahenduse Pandas raamatukogu abil, mida kirjeldasime varem peatükis "Pandas: asendamatu tööriist andmete analüüsiks", valideerime andmeid CAD-failist (BIM) (RVT, IFC, DWG, NWS, DGN) (joonis 4.4-11) saadud tabeli kujulisest failist, kasutades nõudeid teisest tabeli kujulisest nõuete failist (joonis 4.4-9).

Koodi saamiseks tuleb LLM-i käsureas kirjeldada, et meil on vaja laadida andmed failist **raw_data.xlsx** (täielik andmekogum CAD-andmebaasist (BIM)), **kontrollida** neid ja salvestada tulemus uude faili **checked_data.xlsx** (joonis 4.4-13).

🗨️ Saame koodi kasutades LLM ilma Pandas raamatukogu mainimata:

Kirjutage kood, et kontrollida tabelit failist raw_data.xlsx ja valideerida need järgmiste valideerimisreeglite abil: veergude "Laius" ja "Pikkus" väärtused on suuremad kui null, "Energiaklass" sisaldub nimekirjas ["A++", "A+", "A", "B"] ja "Akustiline jõudlus" on muutuja, mille määrame hiljem - koos viimase valideerimissambaga, ja salvestage lõplik tabel uude Exceli faili checked_data.xlsx.

- LLM-i vastuses kirjeldatakse lühike näide Python-koodist, mida saab täiustada ja täiendada järgnevate üleskutsetega :

```

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)

```

Joonis 4.4-14 LLM- mudeli poolt genereeritud kood kontrollib teisendatud CAD (BIM) projekti vastavust piirväärtuste kujul esitatud atribuutide nõuetele.

LLM keelemudeliga genereeritud koodi saab kasutada mis tahes populaarses IDE-s või veebipõhises tööriistas: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse koos PyDev pluginaga, Thonny, Wing IDE, IntelliJ IDEA koos Python pluginaga, JupyterLab või populaarsetes veebipõhistes tööriistades Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Koodi täitmine (joonis 4.4-14) näitab, et CAD-andmebaasist (BIM) pärinevad "üksuse elemendid" W-OLD1, W-OLD2, D-122 (ja muud elemendid) vastavad atribuutide nõuetele: laius ja pikkus on suuremad kui null ja energiatõhususe klass on üks loeteluväärtustest "A++", "A", "B", "C" (joonis 4.4-15).

Vajalik ja hiljuti lisatud element W-NEW, mis vastutab uue elemendiklassi "Window" eest põhja poolel,

ei vasta nõuetele (atribuut "Requirements Met"), sest selle pikkus on null (väärtus "0.0" on meie "Width">0" reegli kohaselt vastuvõetamatu) ja see ei määra energiatõhususe klassi.



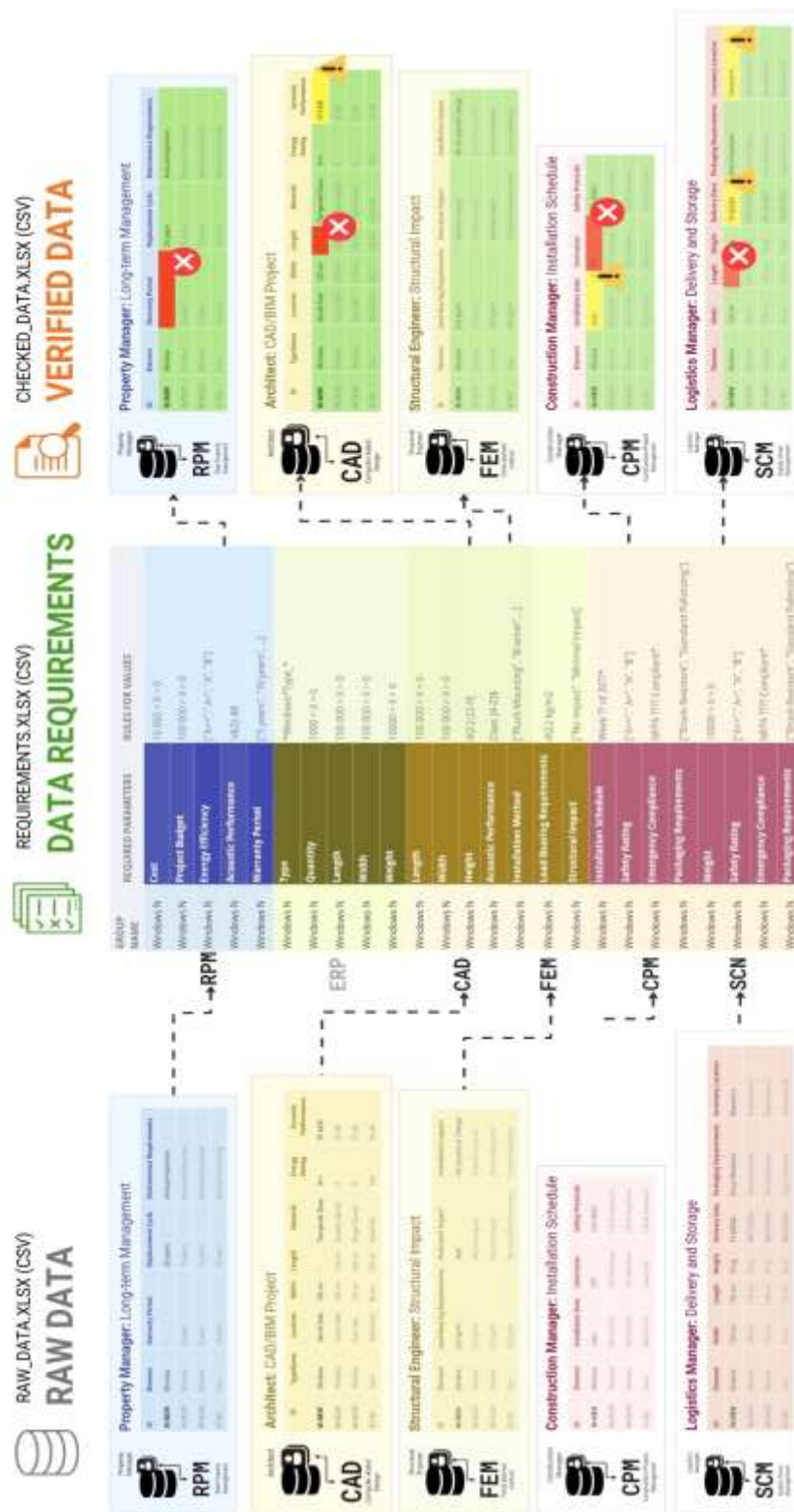
	ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

Joonis 4.4-15 Kontrollimine tuvastab üksused, mis ei ole läbinud kontrolliprotsessi, ja lisab tulemustele uue atribuudi väärtustega "False" või "True".

Samamoodi kontrollime akna lisamise protsessi käigus kõikidest erinevatelt spetsialistidelt saadud andmetest (joonis 4.4-1) kõikide projekti elementide (entiteetide) ja nõutavate atribuutide kooskõla iga süsteemi, tabeli või andmebaasi puhul.

Lõplikus tabelis on mugav visualiseerimiseks märkida kontrolli tulemused värviliselt: rohelisega tähistatakse kontrolli edukalt läbinud atribuute, kollasega - mittekriitiliste kõrvalekalletega väärtusi ja punasega - kriitilisi kõrvalekaldeid (joonis 4.4-16).

Valideerimise tulemusena (joonis 4.4-16) saame nimekirja usaldusväärsetest ja valideeritud elementidest koos nende identifikaatoritega, mille vastavust atribuutide nõuetele on kontrollitud. Valideeritud elemendid annavad kindlustunde, et need elemendid vastavad esitatud standarditele ja spetsifikatsioonidele kõigile süsteemidele, mis on kaasatud Window klassi või mõne muu klassi elementide lisamise protsessi (andmete valideerimise automatiseerimisest ja automatiseeritud ETL protsessi loomisest räägime lähemalt peatükis "ETL ja andmete valideerimise automatiseerimine").



Joonis 4.4-16 Kõigi süsteemide kontrollimise tulemus võimaldab kindlaks teha, millised andmed ei vasta ettevõtte nõuetele.

Üksused, mis on edukalt valideeritud, ei vaja tavaliselt palju tähelepanu. Nad liiguvad takistusteta edasi järgmistesse töötlemisetappidesse ja teistesse süsteemidesse integreerimiseks. Erinevalt "kvaliteetsetest" elementidest pakuvad kõige suuremat huvi just need elemendid, mille valideerimine ebaõnnestub. Teave selliste kõrvalekallete kohta on kriitilise tähtsusega: seda tuleks edastada mitte ainult tabeliaruannetes, vaid ka mitmesuguste visualiseerimisvahendite abil. Kontrollimise tulemuste graafiline esitus aitab kiiresti hinnata andmete kvaliteedi üldist seisundit, tuvastada probleemsed valdkonnad ja võtta kiiresti parandus- või parandusmeetmeid.

kontrollitulemuste visualiseerimine

Visualiseerimine on oluline vahend kontrolli tulemuste tõlgendamisel. Lisaks tavapärastele kokkuvõtivatele tabelitele võib see sisaldada teabepaneele, skeeme ja automaatselt genereeritud PDF-dokumente, mis rühmitavad projekti elemente nende inspekteerimise staatuse järgi. Värvikoodid võivad olla siinkohal abiks: roheline võib tähistada elemente, mis on edukalt valideeritud, kollane võib tähistada elemente, mis vajavad täiendavat tähelepanu, ja punane võib tähistada elemente, milles on kriitilisi vigu või mille kohta puuduvad olulised andmed.

Meie näites (joonis 4.4-1) analüüsime iga süsteemi andmeid samm-sammult: alates CADist (BIM) ja kinnisvarahaldusest kuni logistika ja paigaldusgraafikuteni (joonis 4.4-16). Pärast auditit genereeritakse iga spetsialisti jaoks automaatselt individuaalsed hoiatused või aruandlusedokumentid, nt PDF-vormingus (joonis 4.4-17). Kui andmed on õiged, saab spetsialist lühikese teate: "Täname teid koostöö eest". Kui leitakse lahknevusi, saadetakse üksikasjalik aruanne järgmises sõnastuses: "Selles dokumendis on loetletud elemendid, nende identifikaatorid, atribuudid ja väärtused, mille vastavust ei ole kontrollitud".



Joonis 4.4-17 Valideerimine ja automaatne aruande koostamine kiirendab andmete puuduste leidmist ja mõistmist andmeid koostava spetsialisti jaoks.

Tänu automaatsele valideerimisprotsessile - niipea, kui tuvastatakse viga või andmelünk, saadetakse kohene teade vestluskirja, e-posti või PDF-formaadis -dokumendi kujul isikule, kes vastutab asjaomaste elementide ja nende atribuutide loomise või töötlemise eest (joonis 4.4-18), koos nimekirja elementidest ja atribuutide kirjeldustest, mida ei ole valideeritud.



Joonis 4.4-18 Automaatsed kontrolliaruanded lihtsustavad vigade mõistmist ja kiirendavad projektiandmete täitmise tööd.

Näiteks kui dokument saabub kinnisvarahaldussüsteemi (pärast struktureerimist), kus atribuut "Garantiiperiood" on valesti täidetud, saab kinnisvarahaldur hoiatuse, mis sisaldab nimekirja atribuutidest, mida tuleb kontrollida ja parandada.

Samamoodi põhjustavad kõik puudused paigaldusgraafikus või logistikaandmetes automaatse aruande koostamise ja näiteks chat-teavituse või e-kirja kontrolli tulemuste saatmise asjaomasele spetsialistile.

Lisaks PDF-formaadis -dokumentidele ja tulemusi sisaldavatele graafikutele on võimalik luua armatuurlauad ja interaktiivsed 3D -mudelid (joonis 7.1-6, joonis 7.2-12), mis tõstavad esile puuduvate atribuutidega elemente, võimaldades kasutajatel visuaalselt kasutada elementide 3D-geomeetriat, et filtreerida ja hinnata nende elementide kvaliteeti ja täielikkust projektis.

inspekteerimistulemuste visualiseerimine automaatselt genereeritud dokumentide, graafikute või armatuurtahtvite kujul lihtsustab oluliselt andmete tõlgendamist ja hõlbustab tõhusat teabevahetust projektis osalejate vahel.

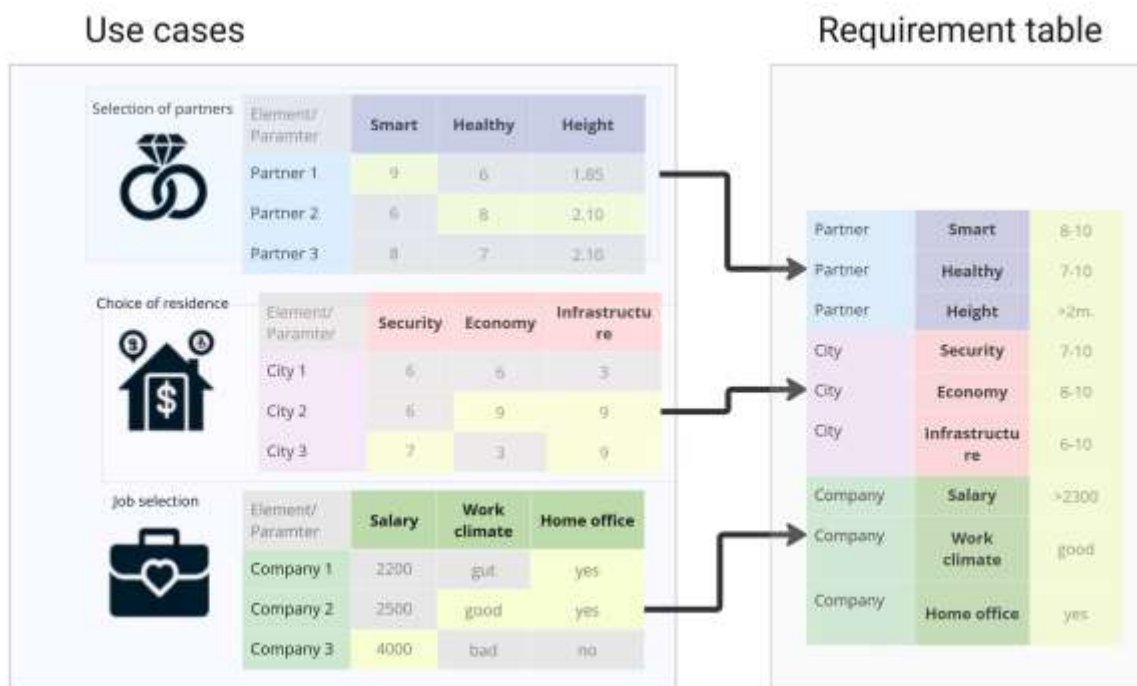
Erinevatest süsteemidest ja teabeallikatest pärit andmete automaatse kontrollimise protsessi võib võrrelda teadliku otsustusprotsessiga igapäevaelus. Nii nagu ehitussektori ettevõtted arvestavad paljude muutujatega - alates sisendandmete usaldusväärsusest kuni nende mõjuni projekti elluviimise ajale, maksumusele ja kvaliteedile -, kaalub inimene oluliste otsuste tegemisel, näiteks elukoha valikul, tervet rida tegureid: transpordi kättesaadavus, infrastruktuur, maksumus, turvalisus, elukvaliteet. Kõik need kaalutlused moodustavad kriteeriumide süsteemi, mis on aluseks lõplikele otsustele, mis moodustavad meie elu.

Andmete kvaliteedikontrolli võrdlus ja inimelu vajaduste võrdlus

Vaatamata andmete kvaliteedi kontrollimise meetodite ja vahendite pidevale arengule, jääb teabe vastavuse põhiprintsiip muutumatuks. See põhimõte on sisse ehitatud küpsete juhtimissüsteemide vundamenti, olgu see siis äris või igapäevaelus.

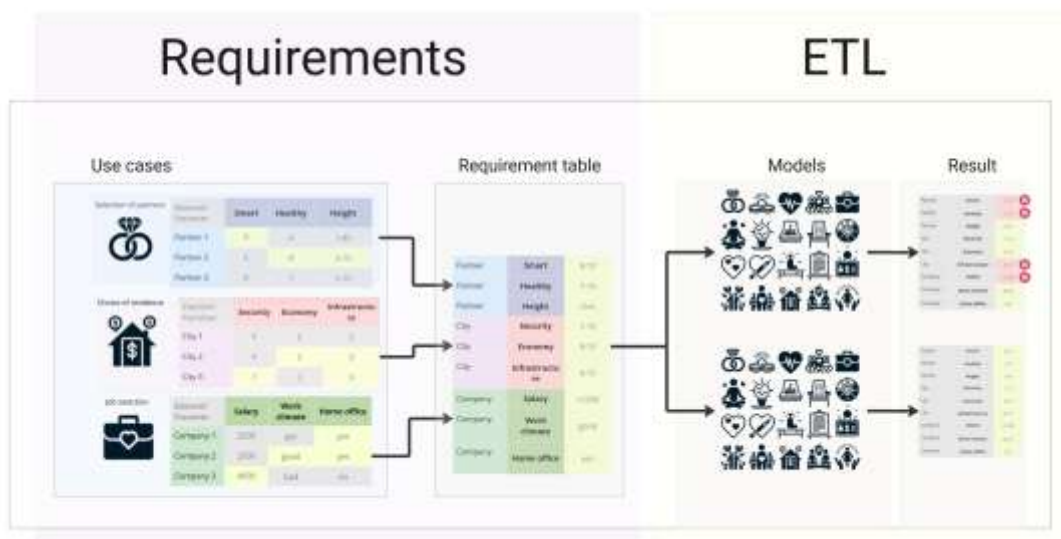
Iteratiivse andmete valideerimise protsess on paljuski sarnane otsustusprotsessiga, millega kõik inimesed igapäevaselt silmitsi seisavad. Mõlemal juhul toetume kogemustele, andmetele ja uuele teabele, kui see muutub kättesaadavaks. Ja üha enam elu- ja töölaseid otsuseid - alates strateegilistest kuni igapäevaste otsusteni - tehakse andmete põhjal.

Näiteks elukoha või elukaaslase valimisel moodustame me intuitiivselt oma mõtetes kriteeriumide ja omaduste tabeli, mille alusel võrdleme alternatiive (joonis 4.4-19). Need omadused - olgu selleks siis isiku isiklikud omadused või kinnisvara parameetrid - kujutavad endast atribuute, mis mõjutavad lõplikku otsust.



Joonis 4.4-19 Elukoha, töökoha või partnerluse valik põhineb individuaalsete omaduste nõuetel.

Struktureeritud andmete kasutamine ja formaliseeritud lähenemine vajaduste kirjeldamisele (joonis 4.4-20) aitab kaasa teadlikumatele ja teadlikumatele valikutele nii töö- kui ka eraelu valdkonnas.

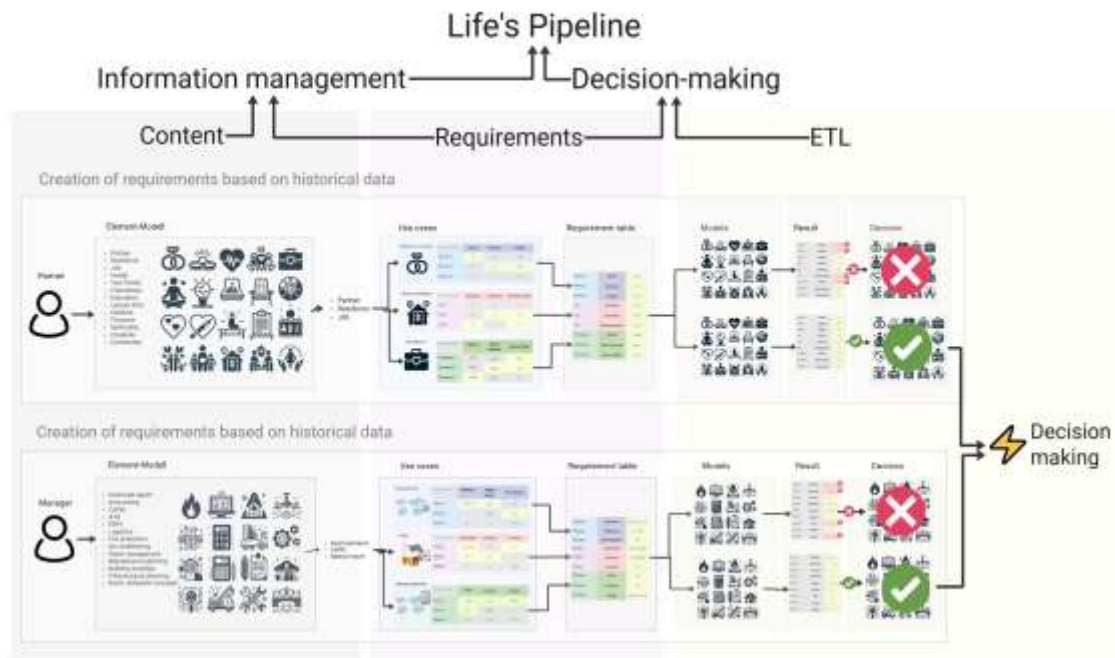


Joonis 4.4-20 Nõuete formaliseerimine võimaldab süstematiseerida elu- ja äriotsuste tajumist.

Andmepõhine otsustusmeetod ei ole ainult äriline vahend. See on sujuvalt integreeritud ka igapäevaellu, järgides tavalisi andmetöötlusetappe (joonis 4.4-21), mis sarnanevad ETL-protsessile (Extract, Transform, Load), mida me juba käesoleva osa alguses andmete struktureerimisel käsitlesime ja mida me üksikasjalikult uurime ülesannete automatiseerimise kontekstis raamatu seitsmendas osas:

- **Andmed kui alus (väljavõte):** Igas valdkonnas - olgu see siis töö või isiklik elu - kogume teavet. Äris võib see olla aruanded, arvud, turuandmed; isiklikus elus võib see olla isiklik kogemus, nõuanded lähedastelt, tagasiside, tähelepanekud.
- **Hindamiskriteeriumid (Transform):** kogutud teavet tõlgendatakse eelnevalt määratletud kriteeriumide alusel. Tööl on need tulemusnäitajad (KPI), eelarvepiirangud ja normid; isiklikus elus sellised parameetrid nagu hind, mugavus, usaldusväarsus, karismaatilisus jne.
- **Proгноosimine ja riskianalüüs (Load):** viimane etapp hõlmab otsuste tegemist, mis põhinevad muundatud andmete analüüsil ja võimalike tagajärgede võrdlemisel. See on sarnane äriprotsessidega, kus andmed läbivad ärioloogika ja riskifiltri.

Otsused, mida me teeme - alates triviaalsetest eelistustest, näiteks mida süüa hommikusöögiks, kuni suurte elusündmusteni, nagu karjääri või elukaaslase valik - on olemuslikult andmete töötlemise ja hindamise tulemus.



Joonis 4.4-21 Ettevõtlus ja elu üldiselt on rida andmepõhiseid otsuseid, kus otsuste tegemisel kasutatavate andmete kvaliteet on võtmetegur.

Kõik meie elus on omavahel seotud ja nii nagu elusorganismid, sealhulgas inimesed, järgivad loodusseadusi, arenevad ja kohanevad muutuvate tingimustega, nii peegeldavad ka inimlikud protsessid, sealhulgas see, kuidas me andmeid kogume ja analüüsime, neid looduslikke põhimõtteid. Looduse ja inimtegevuse vaheline tihe seos kinnitab mitte ainult meie sõltuvust loodusest, vaid ka meie soovi rakendada miljonite aastate jooksul evolutsiooni käigus lihvitud seadusi, et luua andmete arhitektuurid, protsessid ja süsteemid otsuste tegemiseks.

Uued tehnoloogiad, eriti ehituses, on suurepärane näide sellest, kuidas inimkond on ikka ja jälle saanud inspiratsiooni loodusest, et luua paremaid, säästvamaid ja tõhusamaid lahendusi.

Järgmised sammud: andmete muutmine täpseteks arvutusteks ja plaanideks

Selles osas vaatlesime, kuidas teisendada struktureerimata andmeid struktureeritud vormingusse, arendada andmemudeleid ja korraldada ehitusprojektide teabe kvaliteedi kontrollimise protsesse. Andmehaldus, standardimine ja klassifitseerimine on põhiprotsess, mis nõuab süstemaatilist lähenemist ja selget arusaamist ärinõuetest. Selles osas käsitletavat tehnikat ja vahendeid võimaldavad usaldusväärset integratsiooni erinevate süsteemide vahel kogu objekti elutsükli jooksul.

Selle osa kokkuvõtteks toome välja peamised praktilised sammud, mis aitavad teil rakendada käsitletud lähenemisviise oma igapäevastes ülesannetes:

- Alustage nõuete süstematiseerimisest
 - ☐ Looge oma projektide ja protsesside põhielementide atribuutide ja parameetrite register.
 - ☐ Dokumenteerida iga atribuudi piirväärtused

- ☐ Visualiseerida protsessid ja klasside, süsteemide ja atribuutide vahelised seosed, kasutades vooskeeme (nt Miro, Canva, Visio).

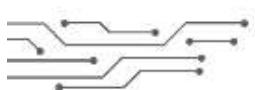
■ Automatiseerida andmete konverteerimine

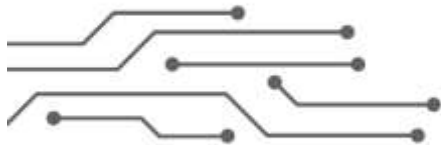
- ☐ Kontrollige, milliseid teie protsessides sageli kasutatavaid dokumente on võimalik OCR-raamatukogude abil digiteerida ja teisendada tabeli kujul.
- ☐ Vaadake pöördprojekteerimise vahendeid, et saada andmeid CADist (BIM).
- ☐ Proovige seadistada automaatne andmete otsimine dokumentidest või formaatidest, mida te oma töös sageli kasutate, tabelivormi.
- ☐ Automaatse konverteerimise seadistamine erinevate andmeformaatide vahel

■ Luua teadmistebaas kategoriseerimiseks

- ☐ Tööstusstandarditega kooskõlas oleva sisemise või olemasoleva elementide klassifikaatori väljatöötamine või kasutamine.
- ☐ dokumenteerida erinevate klassifitseerimissüsteemide vahelised seosed.
- ☐ Arutlege oma meeskonnaga ühtse identifitseerimissüsteemi kasutamise ja elementide üheselt mõistetava liigitamise teemat.
- ☐ Alustage protsessi loomist andmete automaatseks valideerimiseks - nii nende andmete, millega te töötate meeskonnas, kui ka nende andmete, mis edastatakse välissüsteemidesse.

Neid lähenemisviise kasutades saate oluliselt parandada oma andmete kvaliteeti ning lihtsustada nende hilisemat töötlemist ja ümberkujundamist. Käesoleva raamatu järgmistes osades vaatleme, kuidas rakendada juba struktureeritud ja ettevalmistatud andmeid automatiseeritud arvutuste, kulude hindamise, ajakava koostamise ja ehitusprojektide juhtimise jaoks.





V OSA

KULUDE JA AJA ARVUTAMINE: ANDMETE LISAMINE EHITUSPROTSESSIDESSE

Viies osa on pühendatud andmete kasutamise praktilistele aspektidele, et optimeerida ehitusprojektide kulukalkulatsiooni ja planeerimist. Üksikasjalikult analüüsitakse kulukalkulatsiooni ressursipõhist meetodit ja kalkulatsiooniprotsesside automatiseerimist. Käsitletakse koguseliste andmete automatiseeritud hankimise meetodeid) CAD (BIM)-mudelitest ja nende integreerimist arvutussüsteemidega. Uuritakse 4D ja 5D modelleerimismeetodeid ajaplaneerimise ja ehituskulude juhtimise jaoks ning tuuakse konkreetseid näiteid nende rakendamise kohta. Analüüsitakse laiendatud teabekihte 6D -8D, mis pakuvad integreeritud lähenemisviisi kinnisvaraobjektide jätkusuutlikkuse, toimimise ja ohutuse hindamiseks. Üksikasjalikult käsitletakse ehitusprojektide süsinikdioksiidi jalajälje ja ESG -näitajate arvutamise meetodeid kaasaegsete keskkonnanõuete ja -standardite kontekstis. Kriitiliselt hinnatakse traditsiooniliste ERP ja PMIS süsteemide võimalusi ja piiranguid ehitusprotsesside juhtimises, analüüsides nende mõju hinnakujunduse läbipaistvusele. Prognoositakse väljavaateid üleminekuks suletud lahendustelt avatud standarditele ja paindlikele andmeanalüüsi vahenditele, mis suudavad tagada ehitusprotsesside suurema tõhususe.

PEATÜKK 5.1.

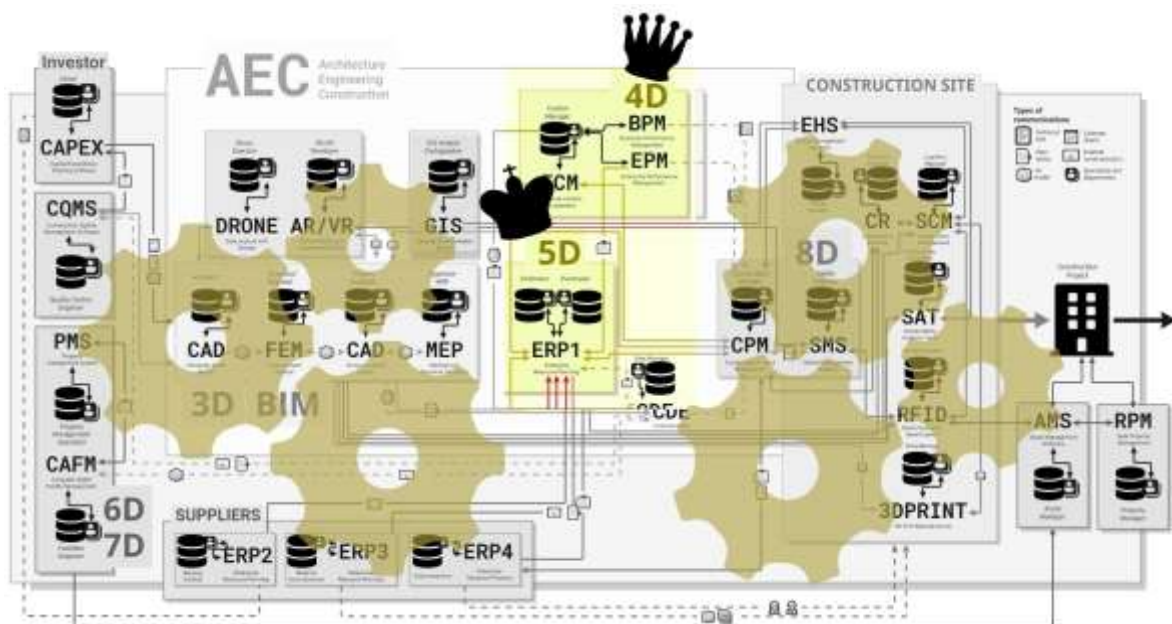
EHITUSPROJEKTIDE KULUARVUTUSED JA KALKULATSIOONID

Ehituse põhitõed: koguse, maksumuse ja aja hindamine

Paljude äriprotsesside hulgas, mis määravad ehitussektori ettevõtte jätkusuutlikkuse, on eriti olulised - nagu ka tuhandeid aastaid tagasi - protsessid, mille abil hinnatakse täpselt elementide arvu, projekti maksumust ja tarneaega (joonis 5.1-1).

Kirjanduse areng oli mitmete tegurite tulemus, sealhulgas vajadus registreerida varajaste ühiskondade majandustehinguid, kaubandust ja ressursside haldamist. Esimesed õiguslikult olulised dokumendid, savitahvlid, millel olid materjalikulude ja tööjõukulude arvutused, olid kasutusel kaubanduse ja ehitusega seotud valdkondades. Need tahvlid registreerisid osapoolte kohustused ehitiste ehitamisel ning neid säilitati lepingute ning raha- ja kaubandussuhete tõendusmaterjalina.

Tuhandeid aastaid on hinnangute koostamise meetod püsinud suuresti muutumatuna: arvutused tehti käsitsi, tuginedes hindamisinseneri kogemusele ja intuitsioonile. Modulaarsete ERP-süsteemide ja CAD-vahendite tulekuga hakkas aga traditsiooniline lähenemine koguste, kulude ja aja hindamisele kiiresti muutuma. Tänapäeva digitaaltehnoloogiad võimaldavad aja ja maksumuse põhilisi arvutusi täielikult automatiseerida, võimaldades ehitusprojektide ressursside planeerimisel suuremat täpsust, kiirust ja läbipaistvust.



Joonis 5.1-1 Paljudest erinevatest süsteemidest on äritegevuses kõige olulisemad vahendid, mis vastutavad mahu, kulude ja aja mõõtmise eest.

Ehitusettevõtete tähelepanu keskmes on täpsed aja- ja kuluandmed. Need omakorda sõltuvad kasutatud materjalide ja tööjõu hulgast ning nende läbipaistvus mõjutab kasumlikkust. Arvutusprotsesside keerukus ja nende läbipaistmatus viivad aga sageli suuremate projektkulude, tähtaegadest mahajäämise ja isegi pankrotini.

KPMG aruande "Tuttavad probleemid - uued lähenemisviisid" (2023) kohaselt valmib ainult 50% ehitusprojektidest õigeaegselt ja 87% ettevõtetest teatab, et kapitaliprojektide ökonoomsus on suurenenud. Peamised probleemid on seotud kvalifitseeritud personali puudumise ja riskide prognoosimise keerukusega [2].

Varasemate projektide ehitamise ajal kogutakse kogu ehitusettevõtte eluea jooksul ajaloolist kuluarvestust ja protsessiaegade andmeid ning sisestatakse need erinevate süsteemide andmebaasidesse (ERP, PMIS BPM, EPM jne).

Kvaliteetsete varasemate kuluandmete olemasolu on ehitusorganisatsiooni jaoks oluline konkurentsieelis, mis mõjutab otseselt tema elujulisust.

Ehitus- ja inseneriettevõtete kalkulatsiooni- ja kuluarvestusosakonnad on loodud selleks, et koguda, salvestada ja ajakohastada ajaloolisi andmeid projektikalkulatsioonide kohta. Nende peamine ülesanne on koguda ja süstematiseerida ettevõtte kogemusi, mis võimaldab aja jooksul parandada uute projektide ulatuse, ajakava ja maksumuse hindamise täpsust. Selline lähenemine aitab vähendada vigu tulevastes arvutustes, mis põhinevad juba teostatud projektide praktilisel tulemusel.

Projektide hinnangulise maksumuse arvutamise meetodid

Kulukalkulatsioonispetsialistid kasutavad erinevaid hindamismeetodeid, millest igaüks keskendub konkreetsele andmetüübile, teabe kättesaadavusele ja projekti detailsuse tasemele. Kõige levinumad on järgmised:

- **Ressursipõhine meetod:** projekti hinnanguline maksumus, mis põhineb kõigi vajalike ressursside, näiteks materjalide, seadmete ja tööjõu üksikasjalikul analüüsil. See meetod nõuab kõigi ülesannete ja iga ülesande täitmiseks vajalike ressursside üksikasjalikku loetelu, millele järgneb nende maksumuse arvutamine. See meetod on väga täpne ja seda kasutatakse kulude hindamisel laialdaselt.
- **Parameetriline meetod:** kasutab statistilisi mudeleid, et hinnata kulusid projekti parameetrite alusel. See võib hõlmata kulude analüüsimist mõõõtühiku, näiteks ehituspinna või töömahu kohta, ja nende väärtuste kohandamist konkreetsete projektitingimustega. See meetod on eriti tõhus varajases etapis, kui üksikasjalik teave ei ole veel kättesaadav.
- **Ühikumeetod (ühikukulude meetod):** arvutab projekti hinnangulise maksumuse, mis põhineb maksumusel mõõõtühiku kohta (nt ruutmeetri või kuupmeetri kohta). See annab kiire ja lihtsa võimaluse võrrelda ja analüüsida erinevate projektide või projektiosade maksumust.
- **Ekspertide hinnang (Delfi meetod):** põhineb ekspertide arvamustel, kes kasutavad oma kogemusi ja teadmisi projekti väärtuse hindamiseks. See meetod on kasulik, kui täpsed lähteandmed ei ole kättesaadavad või kui projekt on ainulaadne.

Tasub märkida, et parameetrilist meetodit ja ekspertide hinnanguid saab kohandada masinõppe mudelitele. See võimaldab projekti kulude ja ajakava prognooside automaatset genereerimist treeningvalimite põhjal. Näiteid selliste mudelite kohaldamise kohta on üksikasjalikumalt käsitletud peatükis "Näide masinõppe kasutamise kohta projekti kulude ja ajakava leidmiseks" (joonis 9.3-5).

Sellest hoolimata on ressursipõhine meetod endiselt kõige populaarsem ja laialdaselt kasutatav maailma praktikas. See ei anna mitte ainult täpset hinnangut hinnangulise maksumuse kohta, vaid võimaldab ka arvutada ehitusplatsil toimuvate üksikute protsesside ja kogu projekti kui terviku kestust (täpsemalt peatükis "Ehitusgraafikud ja 4D -projekti andmed").

Ressursipõhine meetod kulukalkulatsioon ja kalkulatsioon ehituses

Ressursipõhine kuluarvestus on juhtimisarvestuse meetod, mille puhul projekti maksumus põhineb kõigi kaasatud ressursside otsesel arvestusel. Ehituses hõlmab see lähenemisviis kõigi tööde teostamiseks vajalike materjali-, tööjõu- ja tehniliste ressursside üksikasjalikku analüüsi ja hindamist.

Ressursipõhine meetod, tagab eelarve planeerimisel suure läbipaistvuse ja täpsuse, kuna see keskendub ressursside tegelikele hindadele hindamise ajal. See on eriti oluline ebastabiilses majanduskeskkonnas, kus hinnakõikumised võivad oluliselt mõjutada projekti üldkulusid.

Järgmistes peatükkides vaatleme üksikasjalikult ressursipõhist kuluarvestuse protsessi. Et paremini mõista selle põhimõtteid ehituses, tõmbame analoogia restoranis toimuva õhtusöögi kuluarvestusega. Restoranijuht koostab õhtut planeerides nimekirja vajalikest toodetest, võtab arvesse iga roa valmistamise aega ja korrutab seejärel kulud külaliste arvuga. Ehituses on protsess sarnane: iga projekti elementide (objektide) kategooria kohta koostatakse üksikute kulude kalkulatsioonid Recipes, ja projekti kogumaksumus määratakse kindlaks, kui kõik kulud liidetakse kokku koondarveks - lõplikuks kalkulatsiooniks kategooriate kaupa.

Ressursipõhise lähenemise peamine ja esialgne etapp on ettevõtte esialgse andmebaasi loomine. Kulukalkulatsiooni esimeses etapis koostatakse struktureeritud loetelu kõigist esemetest, materjalidest, tööliikidest ja ressurssidest, mis on ettevõtte käsutuses ehitusprojektide raames - alates naelast laos kuni inimeste kirjelduse kaudu nende kvalifikatsiooni ja tunnitasu. See teave süstematiseeritakse ühtsesse "Ehitusressursside ja materjalide andmebaasi" - tabeli kujulisse registrisse, mis sisaldab andmeid nimetuste, omaduste, mõõtühikute ja kehtivate hindade kohta. See andmebaas on peamine ja peamine teabeallikas kõigi järgnevate ressursside arvutuste jaoks - nii tööde maksumuse kui ka ajakava jaoks.

Ehitusressursside andmebaas: ehitusmaterjalide ja -tööde kataloog

Ehitusressursside ja -materjalide andmebaas või tabel - sisaldab üksikasjalikku teavet iga ehitusprojektis kasutatava elemendi - toote, eseme, materjali või teenuse - kohta, sealhulgas selle nimetus, kirjeldus, mõõtühik ja ühikuhind, mis on salvestatud struktureeritud kujul. Selles tabelis on

kõik alates erinevatest projektides kasutatavatest kütustest ja materjalidest kuni spetsialistide üksikasjalike loeteludeni erinevate kategooriate kujul koos tunnitase kirjeldustega (joonis 5.1-2).

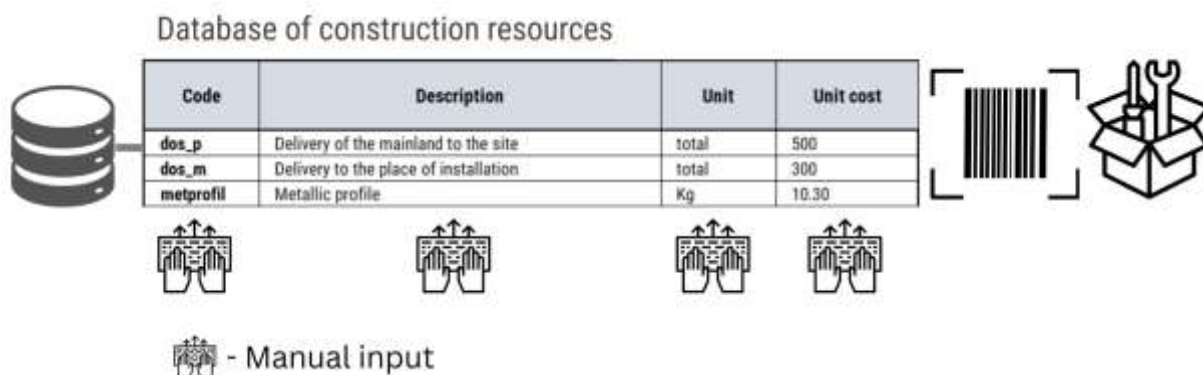
Database of resources	
 1st grade potatoes 1 kg \$2,99	 Sand lime bricks 1 pcs \$1
 Black Angus marble beef 1 kg \$26,99	 JCB 3CX backhoe loader 1 h \$150
 Broccoli 1 pcs \$1,99	 Laborer of the 1st category 1 h \$30

Joonis 5.1-2 Ressursitabel on koostisosade loetelu, mis kirjeldab materjali ja teenust koos ühikuhinnaga.

"Ressursside andmebaas" on sarnane veebipoe tootekataloogiga, kus iga toote omaduste üksikasjalik kirjeldus on olemas. See lihtsustab kulukalkulaatorite jaoks õigete ressursside valimist (nagu toodete valimine ostukorvi lisamisel), mis on vajalikud konkreetsete ehitusprotsesside arvutamiseks arvutuste kujul (lõplik tellimus veebipoes).

Ressursiandmebaasi võib pidada ka kõikide koostisosade loeteluks restorani kokaraamatus. Iga ehitusmaterjal, varustus ja teenus on sarnane retseptides kasutatavate koostisosadega. "Ressursiandmebaas" on kõigi koostisosade - ehitusmaterjalide ja teenuste - üksikasjalik loetelu, sealhulgas nende maksumus ühiku kohta: tükk, meeter, tund, liiter jne.

Uusi üksuse elemente saab lisada tabelisse "Ehitusressursside andmebaasid" kahel viisil - käsitsi (joonis 5.1-3) või automaatselt, integreerides neid ettevõtte varude haldamise süsteemidega või tarnijate andmebaasidega.



Joonis 5.1-3 Ressursiandmebaasi täidetakse käsitsi või võetakse automaatselt andmeid teistest andmebaasidest.

Tüüpiline keskmise suurusega ehitusettevõtte kasutab andmebaasi, mis sisaldab tuhandeid ja mõnikord kümneid tuhandeid ehitusprojektides kasutatavaid üksikasjalike kirjeldustega objekte. Neid andmeid kasutatakse automaatselt lepingutes ja projektdokumentatsioonis, et kirjeldada täpselt tööde ja protsesside koostist

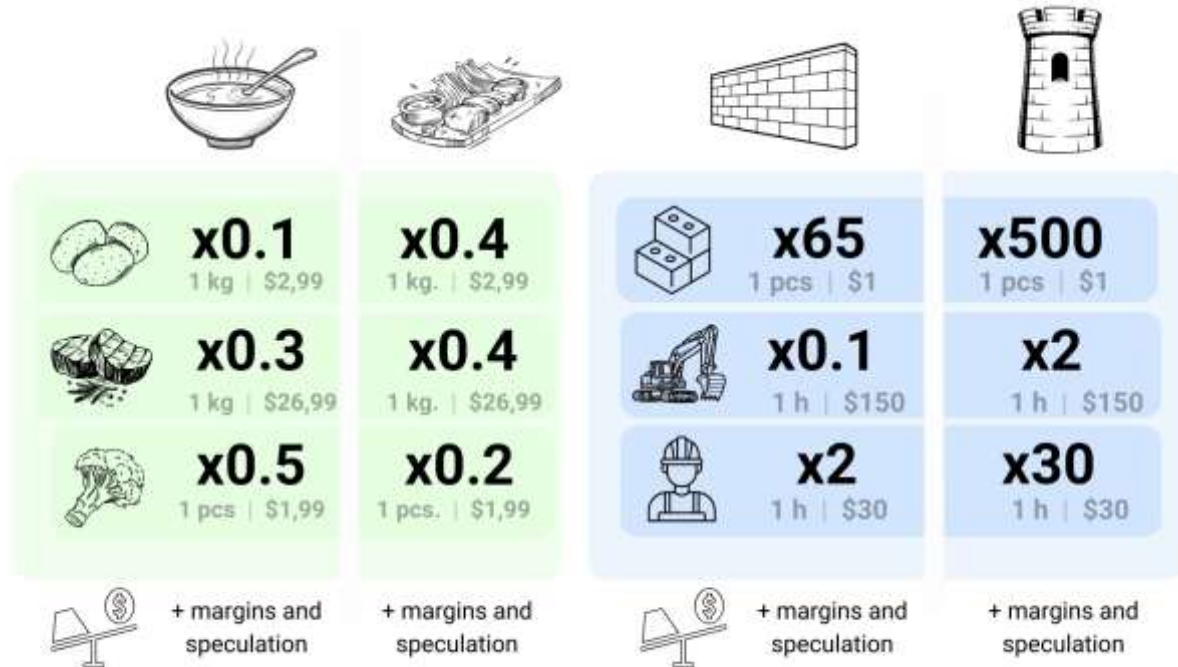
Muutuvate turutingimustega, näiteks inflatsiooniga kaasas käimiseks ajakohastatakse ressursside andmebaasis (joonis 5.1-3) iga toote (kauba või teenuse) "ühikuhindade" atribuuti regulaarselt käsitsi või laadides automaatselt alla jooksvad hinnad teistest süsteemidest või veebiplatvormidest.

Ressursi ühikuhinna ajakohastamine võib toimuda kord kuus, kord kvartalis või kord aastas - sõltuvalt ressursi laadist, inflatsioonist ja välisest majanduskliimast. Selline ajakohastamine on vajalik arvutuste ja hinnangute täpsuse säilitamiseks, sest need põhielemendid on kuluhindade koostajate töö lähtepunktiks. Ajakohastatud andmeid kasutatakse selleks, et koostada kalkulatsioonid, eelarved ja ajakavad, mis kajastavad tegelikke turutingimusi ja vähendavad vigade riski hilisemates projektikalkulatsioonides.

Arvutuste koostamine ja tööde kulukalkulatsioon ressursibaasi alusel

Olles täitnud "Ehitusressursside andmebaasi" (joonis 5.1-3) minimaalsete ühikute üksustega, saate hakata looma arvutusi, mis arvutatakse iga protsessi või töö kohta ehitusplatsil teatud mõõtühikute kohta: näiteks ühe kuupmeetri betooni, ühe ruutmeetri kipsplaadiseina, ühe meetri kõnnitee või ühe akna paigaldamise kohta.

Näiteks 1 m² suuruse tellisseina ehitamiseks (joonis 5.1-4) on varasemate projektide kogemuste põhjal vaja umbes 65 telliskivi (üksus "Silikaattellis"), mille maksumus on 1 \$ tükk (atribuut "Maksumus tüki kohta"), kokku 65 \$. Samuti on minu kogemuse kohaselt vaja kasutada 10 minutit ehitusseadmeid (üksus "JCB 3CX Loader"), mis paigutavad tellised tööpiirkonna lähedale. Kuna seadme rentimine maksab 150 dollarit tunni kohta, siis 6 minutit kasutamist maksaks umbes 15 dollarit. Lisaks on vaja telliskivide paigaldamise töövõtjat 2 tunniks, mille tunnitasu on 30 dollarit ja kokku 60 dollarit.

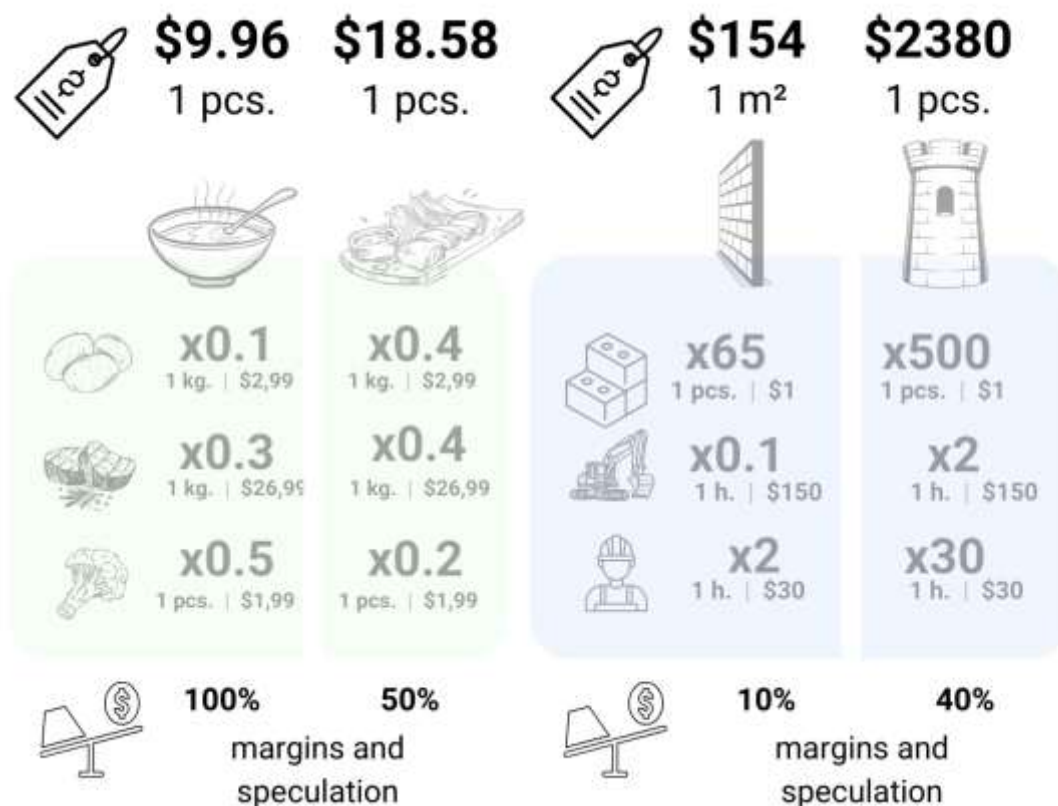


Joonis 5.1-4 Kulukalkulatsioonides on esitatud tööde ja protsesside teostamiseks vajalike ehitusmaterjalide ja -teenuste üksikasjalik loetelu.

Arvutuste koostis (nn "retseptid") kujuneb ettevõtte poolt suure hulga sarnaste tööde tegemise käigus kogutud ajaloolise kogemuse põhjal. See praktiline kogemus koguneb tavaliselt ehitusplatsilt saadud tagasiside kaudu. Eelkõige kogub töödejuhataja teavet otse ehitusplatsil, registreerides tegelikud tööjõukulud, materjalikulu ja tehnoloogiliste toimingute nüansid. Koostöös hindamisosakonnaga täpsustatakse seda teavet seejärel iteratiivselt: protsessi kirjeldusi täpsustatakse, ressursside kombinatsiooni kohandatakse ja kuluarvestust ajakohastatakse, et kajastada hiljutiste projektide tegelikke andmeid.

Nii nagu retsept kirjeldab roa valmistamiseks vajalikke koostisosi ja koguseid, sisaldab kalkulatsioonileht üksikasjalikku nimekirja kõigist ehitusmaterjalidest, ressurssidest ja teenustest, mida on vaja konkreetse töö või protsessi lõpuleviimiseks.

Regulaarselt tehtud töö võimaldab töölistel, töödejuhatajatel ja kalkulaatoritel orienteeruda vajalike ressursside koguses: materjalid, kütus, tööaeg ja muud parameetrid, mis on vajalikud tööüksuse täitmiseks (joonis 5.1-5). Need andmed sisestatakse kalkulatsioonisüsteemidesse tabelite kujul, kus iga ülesanne ja toiming on kirjeldatud ressursibaasi miinimumelementide kaudu (koos pidevalt ajakohastatud hindadega), mis tagab arvutuste täpsuse.



Joonis 5.1-5 Iga töö kohta kogutakse ühikuhinnad, kus üksuse mahu atribuut korrutatakse selle kogusega, millele lisatakse kasumiprotsent.

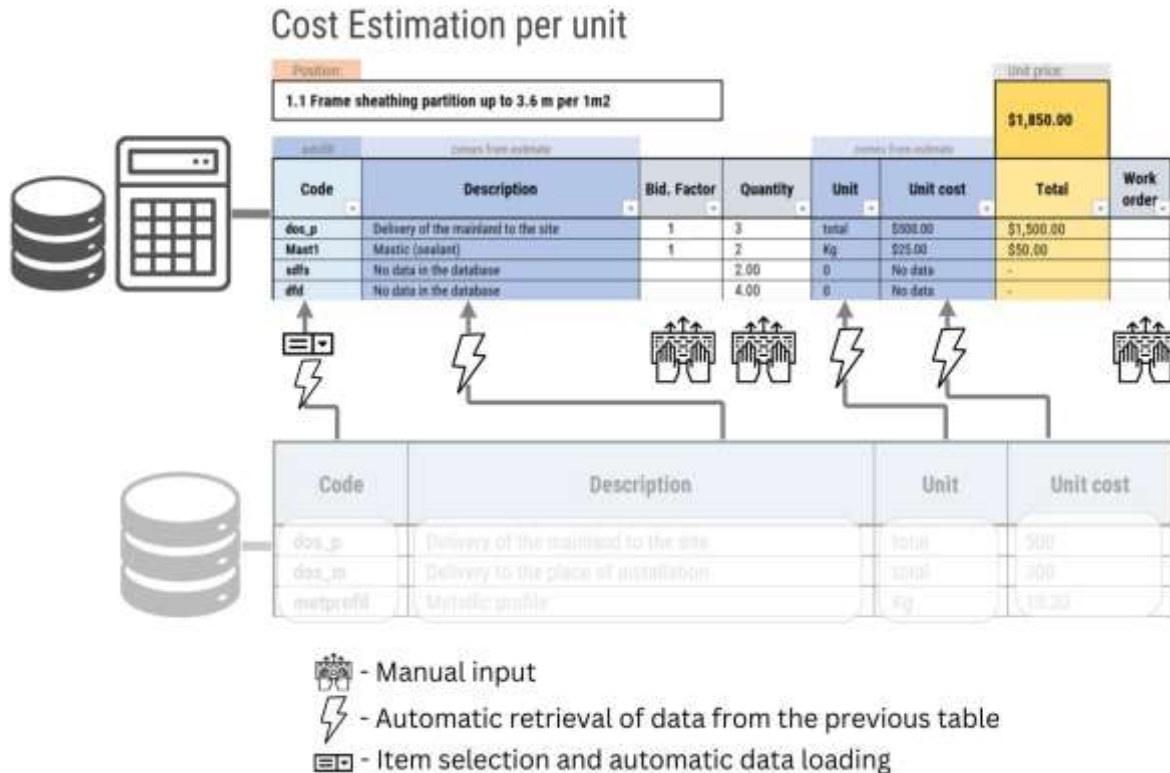
Iga protsessi või tegevuse (kuluobjekti) kogumaksumuse saamiseks korrutatakse kuluatribuut selle arvuga ja teguritega. Koefitsiendid võivad võtta arvesse erinevaid tegureid, näiteks tööde keerukust, piirkondlikke omadusi, inflatsioonimäära, võimalikke riske (eeldatav üldkulude protsent) või spekulatsioonide (lisakasumi tegur).

Kalkulaator kui analüütik teisendab töödejuhi kogemused ja soovitusel standardiseeritud kalkulatsioonideks, kirjeldades ehitusprotsesse ressursiüksuste kaudu tabeli kujul. Sisuliselt on hindaja ülesanne koguda ja struktureerida ehitusplatsilt tulevat teavet parameetrite ja koefitsientide kaudu.

Seega sisaldab lõplik maksumus tööühiku (nt ruut- või kuupmeetri või ühe seadme paigaldamise kohta) mitte ainult materjalide ja tööjõu otsekulusid, vaid ka ettevõtte hinnalisandit, üldkulusid, kindlustust ja muid tegureid (joonis 5.1-6)

Samal ajal ei pea me enam muretsema tegelike hindade pärast (retseptide) arvutustes, sest tegelikud hinnad kajastuvad alati "ressursibaasis" (koostisosade tabelis). Arvutuste tasandil laetakse tabelisse automaatselt andmed ressursibaasist (nt esemekoodi või selle unikaalse tunnuse järgi), mis laadivad kirjelduse ja tegeliku ühikuhinna, mida omakorda saab automaatselt laadida veebiplatvormidelt või

ehitusmaterjalide veebipoest. Kalkulaator arvutuste tasandil tööde puhul tuleb tööd või protsessi kirjeldada ainult atribuudi "ressursside kogus" ja täiendavate tegurite kaudu.



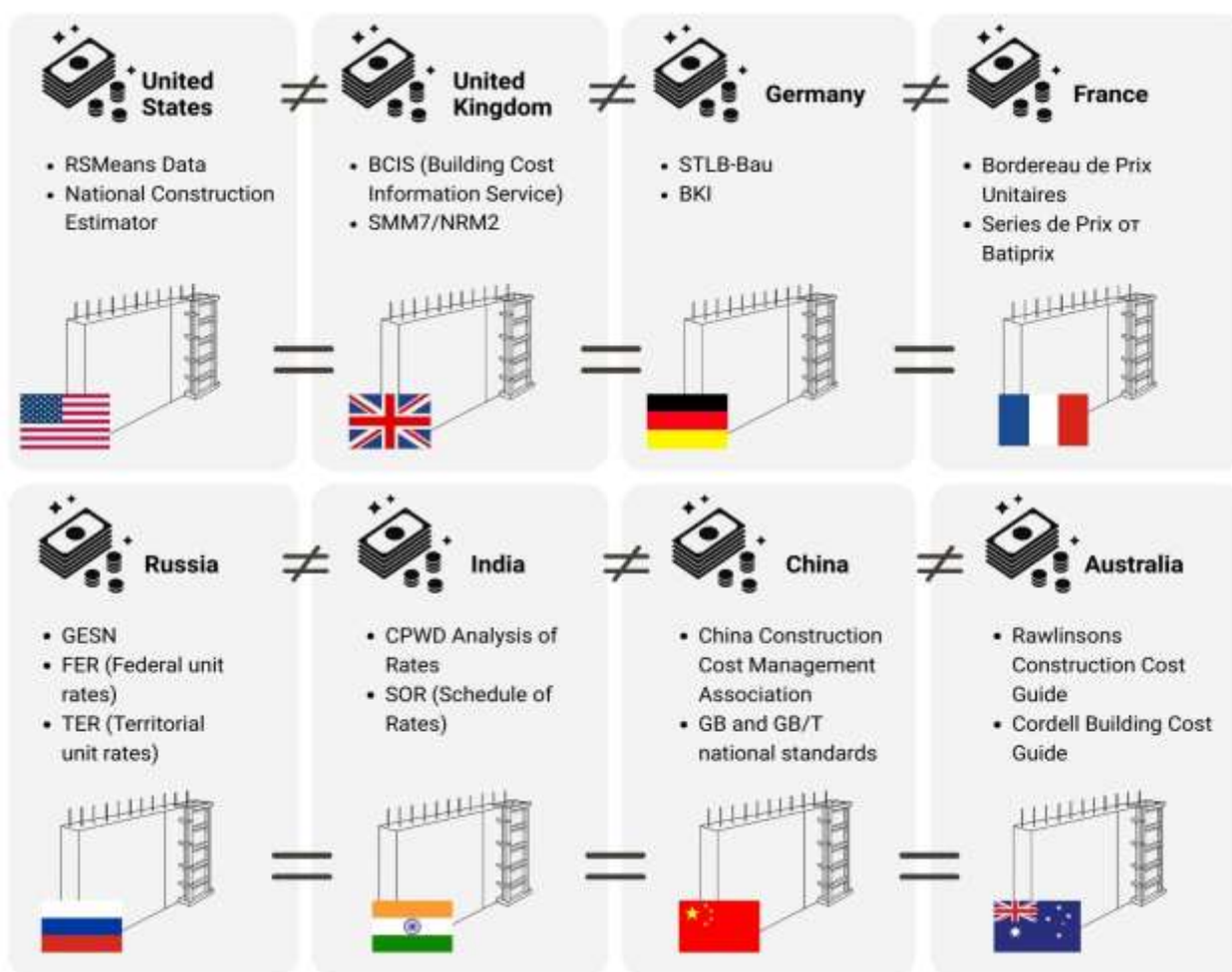
Joonis 5.1-6 Töö ühikuhinna arvutamise etapis täidetakse ainult vajalike ressursside koguse atribuudid, kõik muu laaditakse automaatselt ressursside andmebaasist.

Loodud tööde kalkulatsioonid on salvestatud tüüpiliste projektide näidistabelite kujul, mis on otseselt seotud ehitusressursside ja materjalide andmebaasiga. Need mallid kujutavad endast tulevaste projektide jaoks korduvate tööde tüüpide standardiseeritud retsepte, mis tagavad arvutuste ühtsuse kogu ettevõttes.

Kui mõne ressursi maksumus muutub andmebaasis (joonis 5.1-3) - kas käsitsi või automaatselt jooksvate turuhindade allalaadimise kaudu (nt inflatsioonitingimustes) - kajastuvad uuendused kohe kõigis seotud kulukalkulatsioonides (joonis 5.1-6). See tähendab, et muuta tuleb ainult ressursibaasi, samas kui kuluarvestuse mallid ja kalkulatsioonid jäävad aja jooksul muutumatuks. Selline lähenemisviis tagab arvutuste stabiilsuse ja reprodutseeritavuse mis tahes hinnakõikumiste puhul, mida võetakse arvesse ainult suhteliselt lihtsas ressursitabelis (joonis 5.1-3).

Iga uue projekti jaoks luuakse standardkulude kalkulatsioonimudeli koopiat, mis võimaldab teha muudatusi ja kohandada tegevusi vastavalt konkreetsetele nõuetele, muutmata ettevõtte poolt vastu võetud algset malli. Selline lähenemine tagab paindlikkuse arvutuste kohandamisel: saate võtta arvesse ehitusplatsi eripära, kliendi soove, võtta kasutusele riski- või tasuvus(spekulatsioon)koefitsiente - ja seda kõike ilma ettevõtte standardeid rikkumata. See aitab ettevõttel leida tasakaalu kasumi maksimeerimise, kliendi rahulolu ja konkurentsivõime säilitamise vahel.

Mõnes riigis on sellised aastakümnete jooksul kogunenud kuluarvestuse mallid riiklikul tasandil standardiseeritud ja saanud osaks riiklikest ehituskuluarvestussüsteemi standarditest (joonis 5.1-7).



Joonis 5.1-7 Erinevatel riikidel üle maailma on oma kuluarvestuse eeskirjad koos oma (ettekirjutavate) ehitustööde kompendiumide ja standarditega sama elemendi kuluarvestuseks.

Selliseid standardiseeritud ressursside hinnangu aluseid (joonis 5.1-7) on kohustuslik kasutada kõigil, eriti riiklikult rahastatavate projektide puhul. Selline standardiseerimine tagab läbipaistvuse, võrreldavuse ja õigluse hindade kujundamisel ja lepinguliste kohustuste võtmisel kliendi jaoks.

Projekti lõplik kulukalkulatsioon: kalkulatsioonist eelarvesse

Riiklikel ja tööstusharupõhistel hindamisstandarditel on eri riikides ehitustegevuses erinev roll. Kui mõnes riigis nõutakse rangelt ühe standardi järgimist, siis enamikus arenenud riikides kasutatakse paindlikumat lähenemisviisi. Turumajandusega riikides on riiklikud ehitusstandardid tavaliselt ainult lähtealuseks. Ehitusettevõtted kohandavad neid standardeid oma tegevusmudelitele või vaatavad need täielikult läbi, täiendades neid oma kohandatud teguritega. Need kohandused kajastavad ettevõtte kogemusi, ressursside juhtimise tõhusust ja sageli ka tegureid, mille puhul võib arvesse võtta näiteks ettevõtte spekulatiivset kasumit.

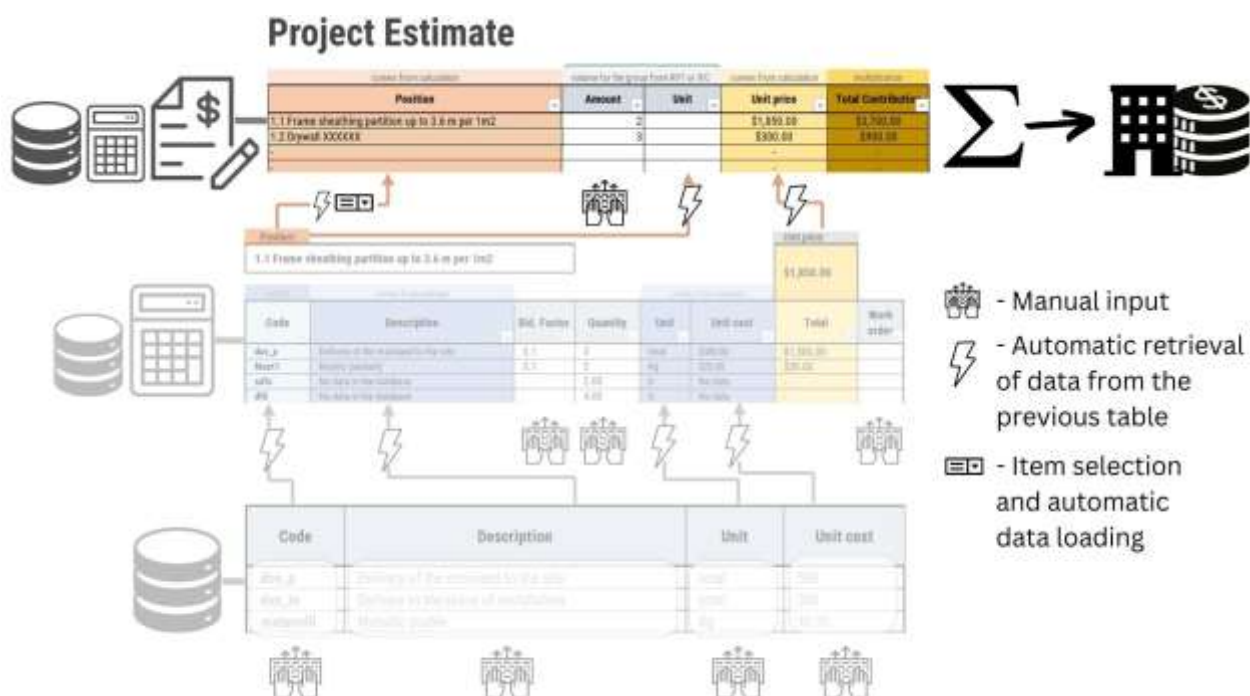
Selle tulemusena võivad konkurentsi tase, turunõudlus, sihtmarginaalid ja isegi suhted konkreetsete klientidega põhjustada märkimisväärsed kõrvalekaldeid standardiseeritud normidest. Selline praktika tagab turu paindlikkuse, kuid raskendab ka erinevate töövõtjate pakkumiste läbipaistvat võrdlemist, tuues ehitussektoris kalkulatsiooniprotsessi selles etapis sisse spekulatiivse hinnakujunduse elemendi.

Kui üksikute tegevuste ja protsesside arvutusmallid on koostatud - või sagedamini lihtsalt kopeeritud valitsuse standardhinnangutest (joonis 5.1-7), millele on lisatud koefitsiendid konkreetse ettevõtte "eripärade" kajastamiseks -, tuleb viimaseks sammuks korrutada iga kirje maksumus uue projekti töömahu või protsesside vastava atribuudiga.

Uue ehitusprojekti kogumaksumuse arvutamisel on peamine samm kõigi kuluartiklite kulude kokkuvõtmine, mis korrutatakse nende tööde mahuga projektis.

Projekti kogumaksumuse moodustamiseks alustame meie lihtsustatud näites ühe ruutmeetri seina ehitamise maksumuse arvutamisest ja korrutame selle arvutamise maksumuse (nt töö "1m² seinaelementide standardpaigaldamine") projekti seinte ruutmeetrite koguarvuga (nt projekti CAD-ist või töödejuhi arvutustest saadud "Seinatüüpi" üksuse "Pindala" või "Kogus" atribuut (joonis 5.1-8)).

Arvutame kõigi projekti elementide maksumuse samamoodi (joonis 5.1-8): võtame tööühiku maksumuse ja korrutame selle konkreetse projekti elemendi või elementide rühma mahuga. Kalkulaator peab ainult sisestama nende elementide, tegevuste või protsesside arvu projektis mahu või koguse kujul. See võimaldab automaatselt genereerida täieliku ehitushinnangu.



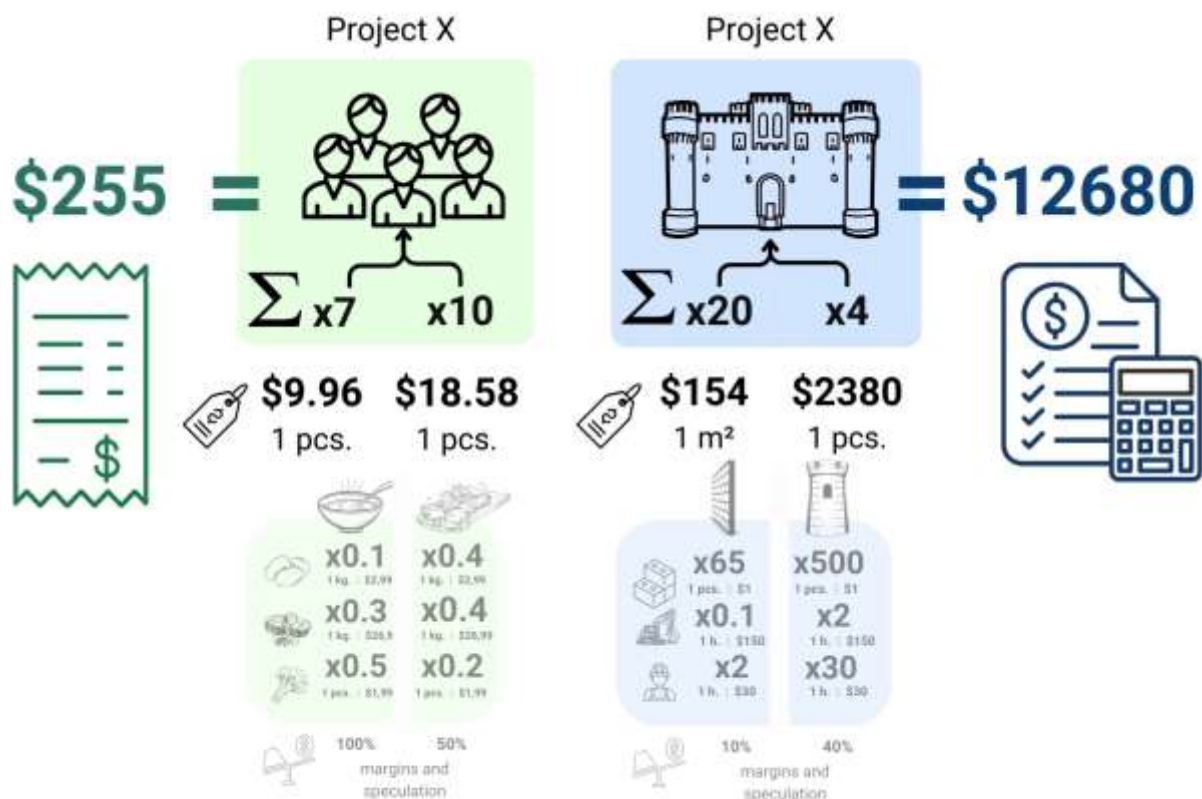
Joonis 5.1-8 Hinnangu koostamise etapis sisestame ainult tööde mahu.

Nagu arvutuste puhul, laadime sellel tasandil automaatselt valmis arvutatud elemendid (arvutuste mallist või uued, mallist kopeeritud ja redigeeritud elemendid), mis toovad automaatselt kaasa jooksva tööühiku maksumuse (mida uuendatakse automaatselt ressursside andmebaasist (joonis 5.1-8 alumine tabel)). Seega, kui ressursiandmebaasi või kulukalkulatsiooni tabelite andmed muutuvad - ajakohastatakse automaatselt kalkulatsiooni andmeid jooksva päeva kohta, ilma et oleks vaja muuta kulukalkulatsiooni või kalkulatsiooni ennast.

Restorani kontekstis arvutatakse ürituse lõppmaksumus sarnaselt ja see võrdub kogu õhtusöögi lõppmaksumusega, kus iga roa maksumus korrutatud külaliste arvuga annab kokku tšeki kogumaksumuse (joonis 5.1-9). Ja nii nagu ehituses, ei pruugi ka restoranis toiduvalmistamise retseptid aastakümneid muutuda. Erinevalt hindadest, kus koostisainete hind võib muutuda iga tunni tagant.

Nii nagu restoraniomanik korrutab iga söögi maksumuse portsjonite ja inimeste arvuga, et määrata ürituse kogumaksumus, liidab kulude kalkulatsiooni juht kõigi projekti komponentide maksumuse kokku, et saada täielik ehitushinnang.

Seega määratakse projekti iga tegevuse jaoks selle lõplik maksumus (joonis 5.1-9), mis korrutatakse sellele tegevusele vastava üksuse atribuutmahuga - see annab tegevuste rühmade maksumuse, millest saadakse kogu projekti lõplik maksumus.



Joonis 5.1-9 Lõpliku hinnangu arutamiseks liidetakse iga elemendi töökulu atribuut ja selle ulatuse atribuut kokku.

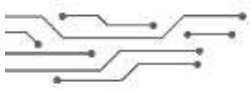
Projekti kogumaksumus (joonis 5.1-8) annab projekti finantspildi, mis võimaldab tellijatel, investoritel või finantseerimisasutustel mõista projekti elluviimiseks vajalikku kogueelarvet ja rahalisi vahendeid igal päeval, võttes arvesse jooksvaid hindu.

Ja kui ressursibaaside, arvutuste ja kalkulatsioonide koostamise protsessid (protsessiretseptid) on juba kümnete tuhandete aastate jooksul välja töötatud, poolautomatiseeritud ja lihvitud ning riigi tasandil salvestatud, siis kvalitatiivse teabe saamine elementide mahu ja koguse kohta lõpliku kalkulatsiooni viimase etapi jaoks - on tänapäeval endiselt kitsaskoht kõigi projekti kulu- ja ajaatribuutide arvutamise protsessides ja üldiselt projekti üldelarves.

Tuhandeid aastaid on mahtude arvutamise traditsiooniline meetod olnud käsitsi mõõtmine, mille puhul on kasutatud lamejooniseid. Digitaalajastu saabudes on ettevõtted avastanud, et mahu- ja kogusteavet saab nüüd automaatselt välja võtta CAD-mudelites sisalduvatest geomeetrilistest andmetest, mis muudab aastatuhandeid vanad kvantitatiivsete andmete saamise viisid.

Kaasaegsed lähenemisviisid protsesside hindamisele ja kalkuleerimisele hõlmavad automaatset mahuliste ja kvantitatiivsete tunnuste väljavõtmist CAD-andmebaasidest, mida saab üles laadida ja ühendada kalkulatsiooniprotsessiga, et saada ajakohaseid projektirühmade mahte igas

projekteerimise etapis kuni kasutamiseni.



PEATÜKK 5.2.

KOGUSELINE ARVESTUS NING KALKULATSIOONIDE JA GRAAFIKUTE AUTOMAATNE KOOSTAMINE

Üleminek 3D-lt 4D-le ja 5D-le: ruumiliste ja kvantitatiivsete parameetrite kasutamine

Kui käes on kuluarvestustabelid koos kirjeldatud protsessidega ressursside kaudu (joonis 5.1-8), on järgmine samm arvutuste ja lõpliku hinnangu jaoks vajalike elementide rühma mahu- või koguseparameetrite automaatne saamine.

Projekti elementide - nt seinte või plaatide - mahulised omadused saab automaatselt CAD-andmebaasidest välja võtta. CAD-programmides loodud parameetrilised objektid teisendatakse geomeetrilise tuuma abil pikkuse, laiuse, pindala, mahu ja muude parameetrite arvulisteks väärtusteks. 3D-geomeetria põhinevat mahtude saamise protsessi käsitletakse üksikasjalikumalt järgmises, kuuendas osas (joonis 6.3-3), mis on pühendatud tööle CADiga (BIM). Lisaks mahtudele saab CAD-mudeli andmebaasist saada ka sarnaste elementide arvu, filtreerides ja grupeerides objekte kategooriate ja omaduste järgi. Need parameetrid, mis võimaldavad rühmitamist, saavad aluseks projekti elementide sidumiseks ressursiarvutuste kaudu kogu projekti arvutuste, lõplike kalkulatsioonide ja eelarvega.

Seega täiendatakse 3D (CAD) mudelist saadud andmemudelit uute parameetrikihistidega, mida nimetatakse 4D ja 5D. Uutes üksuse atribuutide kihtides 4D (aeg) ja 5D (kulud) kasutatakse 3D geomeetrilisi andmeid üksuse mahu atribuutide väärtuste allikana.

- **4D** on parameetrite infokihi, mis lisab elementide 3D-parameetritele teavet ehitustoimingute kestuse kohta. Need andmed on olulised ajakavade planeerimiseks ja projekti tähtaegade haldamiseks
- **5D** on andmemudeli järgmine laiendustase, kus elemendid on täiendatud kulukirjeldustega. See lisab geomeetrilisele teabele finantsaspekti: materjalide, tööde ja seadmete maksumus, mis võimaldab eelarve arvutusi, tasuvusanalüüsi ja kulude juhtimist ehitusprotsessi ajal.

Projektiüksuste rühmade kulude ja 3D, 4D ja 5D atribuutide andmeid kirjeldatakse sarnaselt modulaarsete ERP-, PIMS -süsteemide (või Exceli sarnaste vahendite) arvutustega ning neid kasutatakse nii üksikute rühmade kui ka kogu projekti eelarve automaatseks kulukalkulatsiooniks ja eelarve planeerimiseks.

5D atribuudid ja atribuutide mahtude saamine CADist

Ehitusprojekti lõpliku kalkulatsiooni koostamisel, mille koostamist oleme käsitlenud eelmistes peatükkides (joonis 5.1-8), kogutakse iga projektielementide kategooria mahuatribuudid kas käsitsi või võetakse need CAD-tarkvara poolt esitatud mahuatribuutide spetsifikatsioonidest.

Traditsiooniline manuaalne meetod koguste arvutamiseks hõlmab töödejuhataja ja kalkulaatori tööd, kes analüüsivad jooniseid, mida on tuhandeid aastaid esitatud joontena paberil ja viimased 30 aastat digitaalsetes vormingutes, nagu PDF (PLT) või DWG. Tuginedes erialasele kogemusele, mõõdavad nad vajalike tööde ja materjalide kogused, sageli joonlauda ja nurksulguriga. See meetod nõuab märkimisväärseid jõupingutusi ja aega ning erilist tähelepanu üksikasjadele.

Projekti ulatuse atribuutide kindlaksmääramine sel viisil võib sõltuvalt projekti ulatusest võtta aega mõnest päevast kuni mitme kuuni. Lisaks sellele, kuna kõik mõõtmised ja arvutused tehakse käsitsi, on oht, et inimlikud vead võivad põhjustada ebatäpseid andmeid, mis omakorda mõjutavad vigu projekti aja ja kulude hindamisel, mille eest vastutab kogu ettevõtte.

Kaasaegsed meetodid, mis põhinevad CAD-andmebaaside kasutamisel, lihtsustavad oluliselt mahtude arvutamist. CAD-mudelites sisaldab elementide geomeetria juba ruumalaatribuute, mida saab automaatselt arvutada (geomeetrilise tuuma kaudu (joonis 6.3-3)) ja esitada või eksportida tabelina.

Sellise stsenaariumi korral küsib kalkulatsiooniosakond CAD-konstruktorilt andmeid projekti elementide koguse ja mahu omaduste kohta. Need andmed eksporditakse tabelina või integreeritakse otse kalkulatsiooniandmebaasidesse - olgu selleks siis Excel, ERP või PMIS - süsteemidesse. See protsess ei alga sageli mitte ametliku taotlusega, vaid lühikese dialoogiga kliendi (algataja) ning ehitus- või projekteerimisfirma arhitekti kalkulaatori vahel. Allpool on esitatud lihtsustatud näide, mis näitab, kuidas igapäevasest suhtlusest moodustub struktureeritud tabel automaatseteks arvutusteks (QTO):

- 🗨 Klient - *"Soovin lisada hoonele veel ühe korruse, samasuguses konfiguratsioonis nagu esimene korrus"*
- 🗨 Arhitekt (CAD) - *"Kolmanda korruse lisamine, konfiguratsioon on sama nagu teisel korrusel"*. Ja pärast seda sõnumit saadab kalkulaatorile projekti uue CAD-versiooni.
- 🗨 Kalkulaator teostab automaatselt grupeerimise ja arvutused (ERP, PMIS, Excel) - *"Ma ajan projekti läbi Exceli tabelis QTO reeglitega (ERP, PMIS), saan uue korruse mahud kategooriate kaupa ja genereerin kalkulatsiooni"*

Selle tulemusel muudetakse tekstidialoog tabelistruktuuriks koos rühmitamisreeglitega:

Element	Kategooria	Põrand
Kattuvus	OST_Floors	3
Veerg	OST_StructuralColumns (struktuurisambad)	3
trepikoda	OST_Stairs	3

Pärast projekteerijalt saadud CAD-mudeli automaatset grupeerimist vastavalt kalkulaatori QTO reeglitele ja mahtude automaatset korrutamist ressursiarvutustega (joonis 5.1-8) saame järgmised tulemused, mis saadetakse tellijale:

Element	Kööde	Põrand	Hind ühiku kohta.	Kogumaksumus
Kattuvus	420 m ²	3	150 €/m ²	63 000 €
Veerg	4 tk.	3	2450 €/tükk.	9 800 €
trepikoda	2 tk.	3	4300 €/tükk.	8 600 €
KOKKU:	-	-	-	81 400 €

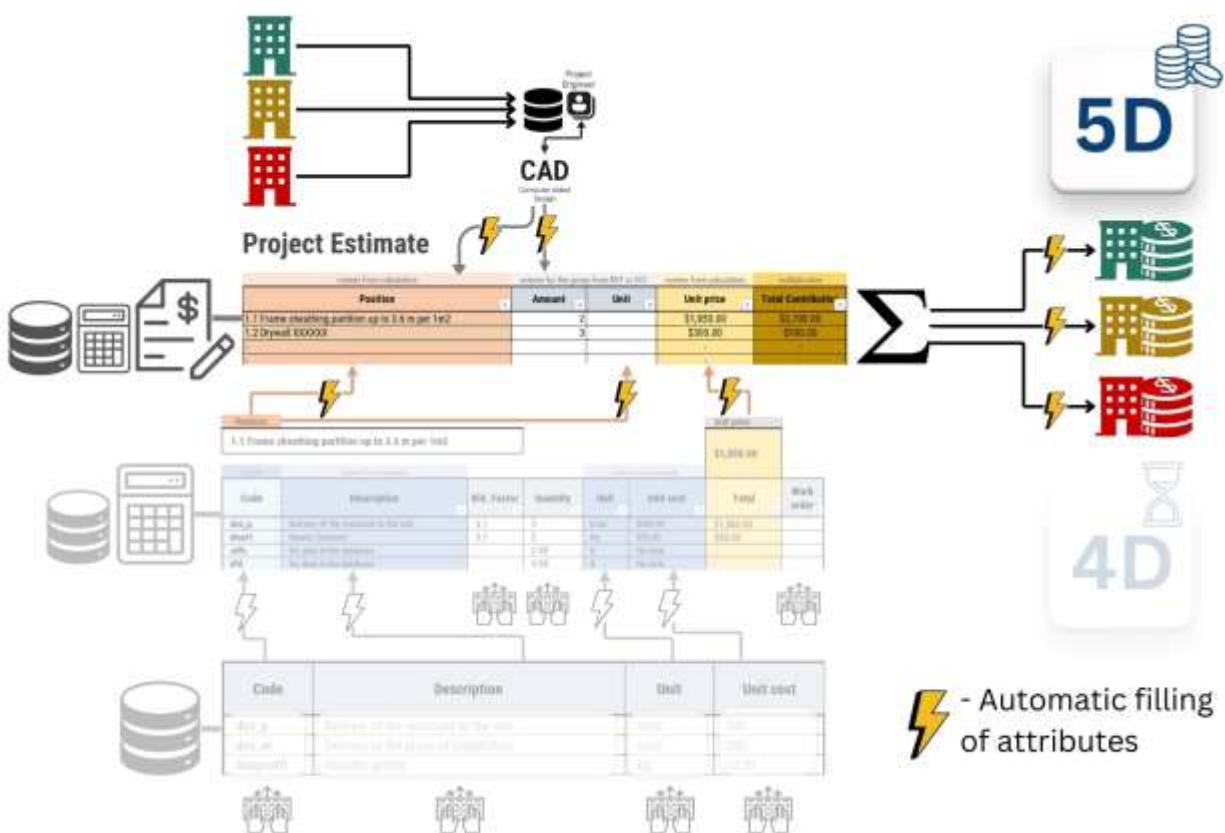
🗣️ Klient - "Täna teid, see on päris palju, me peame mõned ruumid maha lõikama". Ja tsükel kordub mitu korda.

See stsenaarium võib korduda mitu korda, eriti heakskiitmise etapis, kus klient ootab kohest tagasisidet. Tegelikult võivad sellised protsessid aga venida päevade või isegi nädalate kaupa. Tänapäeval peaks tänu automaatse rühmitamise ja arvutamisreeglite kasutuselevõtule toimingud, mis varem võtsid palju aega, olema tehtud minutitega. Koguste automatiseeritud kogumine rühmitamisreeglite abil mitte ainult ei kiirenda arvutusi ja kalkulatsioone, vaid vähendab inimteguri minimeerimisega ka vigade tõenäosust, tagades läbipaistva ja täpse hinnangu projekti kuludele.

Kui 3D-mudeli loomisel CAD-süsteemis võeti algsest arvesse hindamisosakonna nõudeid (mis on praktikas veel haruldane) ning elementide rühmade nimed, identifikaatorid ja nende klassifitseerimisatribuudid on määratud parameetrite kujul, mis langevad kokku hindamisrühmade ja -klasside struktuuridega, siis saab ruumilised atribuudid automaatselt ilma täiendavate teisendusteta hindamissüsteemidesse üle kanda.

Automaatne mahuandmete väljavõtmine CADist spetsifikatsioonitabelite kujul võimaldab kiiresti saada ajakohaseid andmeid üksikute tööde ja projekti kui terviku maksumuse kohta (joonis 5.2-1). Uuendades arvutamisprotsessis või arvutussüsteemis ainult CAD-faili koos projekti mahtudega, saab ettevõtte kalkulatsiooni kiiresti ümber arvutada, võttes arvesse viimaseid muudatusi, tagades kõigi

järgnevate arvutuste suure täpsuse ja järjepidevuse.



Joonis 5.2-1 CAD-tabelitest või andmebaasidest saadud mahu atribuudid sisestatakse automaatselt kalkulasiooni, mis võimaldab kohe arvutada projekti kogumaksumuse.

Kapitaliprojektide keerukuse kasvades muutub kogu eelarve arvutamine ja projektide kogumaksumuse analüüsimine sellise stsenaariumi korral (joonis 5.2-1) - teadliku otsustamise võtmetähtsusega vahendiks.

Accenture'i uuringu "Creating More Value through Capital Projects (2024)" [20] kohaselt integreerivad juhtivad ettevõtted aktiivselt andmeanalüüsi digitaalsete algatuste puhul, kasutades varasemat teavet tulemuste prognoosimiseks ja optimeerimiseks. Uuringud näitavad, et üha enam omanike-ettevõtjaid rakendavad suurandmete analüüsi, et prognoosida turusuundumusi ja hinnata ärilist elujõulisust enne projekteerimise alustamist. See saavutatakse olemasolevate projektide portfelli andmeaitade analüüsimisega. Lisaks sellele rakendab 79 protsenti omanik-ettevõtjatest "jõulist" prognoosivat analüüsi, et hinnata projektide tulemuslikkust ja toetada reaajas operatiivsete otsuste tegemist.

Ehitusprojektide kaasaegne tõhus juhtimine on lahutamatult seotud suure hulga teabe töötlemise ja analüüsiga kõigis projekteerimise ja projekteerimisele eelnevates etappides. andmeaida, ressursiarvutuste, prognoosimudelite ja masinõppe kasutamine võimaldab mitte ainult minimeerida riske arvutustes, vaid ka teha strateegilisi otsuseid projekti rahastamise kohta projekteerimise varajases etapis. Andmevarudest ja prognoosimudelitest, mis täiendavad arvutusi, räägime lähemalt

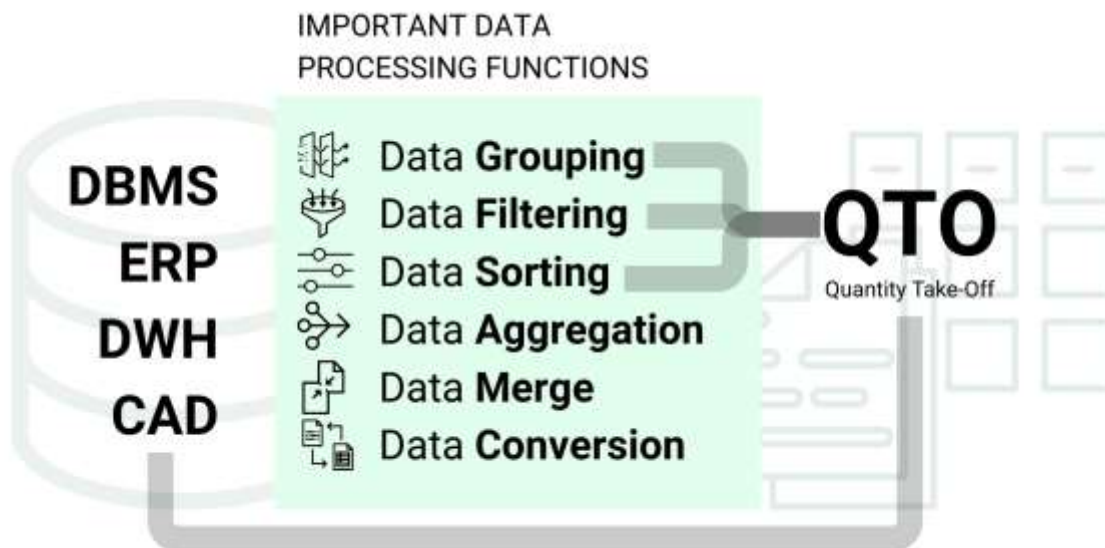
raamatu üheksandas osas.

CAD-projektide elementide mahuparameetrite automaatne saamine, mis on vajalikud kalkulatsioonide koostamiseks, toimub grupeerimisvahendite QTO (Quantity Take-Off) abil. QTO tööriistad töötavad, grupeerides kõik projekti objektid spetsiaalsete elementide identifikaatorite või elementide atribuutide parameetrite järgi, kasutades CAD-andmebaasis loodud spetsifikatsioone ja tabeleid.

QTO Quantity Take-Off: projektandmete rühmitamine atribuutide järgi

QTO (Quantity Take-Off) on ehituses projekti elluviimiseks vajalike elementide koguste väljavõtmine. Tegelikult on QTO sageli poolmanuaalne protsess, mis hõlmab andmete kogumist erinevatest allikatest: PDF-dokumentidest, DWG-joonistest ja digitaalsetest CAD-mudelitest.

CAD-andmebaasidest saadud andmetega töötades on QTO protsess teostatud filtreerimis-, sorteerimis-, rühmitamis- ja koondamisoperatsioonide jadana. Mudelielemendid valitakse välja klasside, kategooriate ja tüüpide parameetrite järgi, misjärel nende kvantitatiivsed omadused - nagu maht, pindala, pikkus või kogus - võetakse kokku vastavalt arvutusloogikale (joonis 5.2-2).



Joonis 5.2-2 Andmete rühmitamine ja filtreerimine on kõige populaarsemad andmebaasides ja andmeladudes kasutatavad funktsioonid.

QTO (filtreerimine ja rühmitamine) võimaldab süstematiseerida andmeid, koostada spetsifikaate ja valmistada ette sisendinformatsiooni kalkulatsioonide, ostude ja tööplaanide arvutamiseks. QTO aluseks on elementide klassifitseerimine mõõdetavate tunnuste tüübi järgi. Iga elemendi või elementide rühma jaoks valitakse vastav kvantitatiivne mõõteparameeter. Näiteks:

- **Pikkuse atribuut** (äärekiivi - meetrites)
- **Pindala atribuut** (kipsplaaditööd - ruutmeetrites)
- **Mahtude atribuut** (betoonitööd - kuupmeetrites)

■ Koguse atribuut (aknad - tüki kohta)

Lisaks geomeetria alusel matemaatiliselt genereeritud mahukarakteristikutele rakendatakse pärast QTO grupeerimist arvutustes sageli ka ületamiskoeffitsiente (joonis 5.2-12, nt 1,1, et võtta arvesse 10% logistika ja paigalduse eest) - parandusväärtusi, mis võtavad arvesse kadusid, paigaldust, ladustamist või transporti. See võimaldab täpsemalt prognoosida materjalide tegelikku tarbimist ja vältida nii puudujääke kui ka ülevarusid ehitusplatsil.

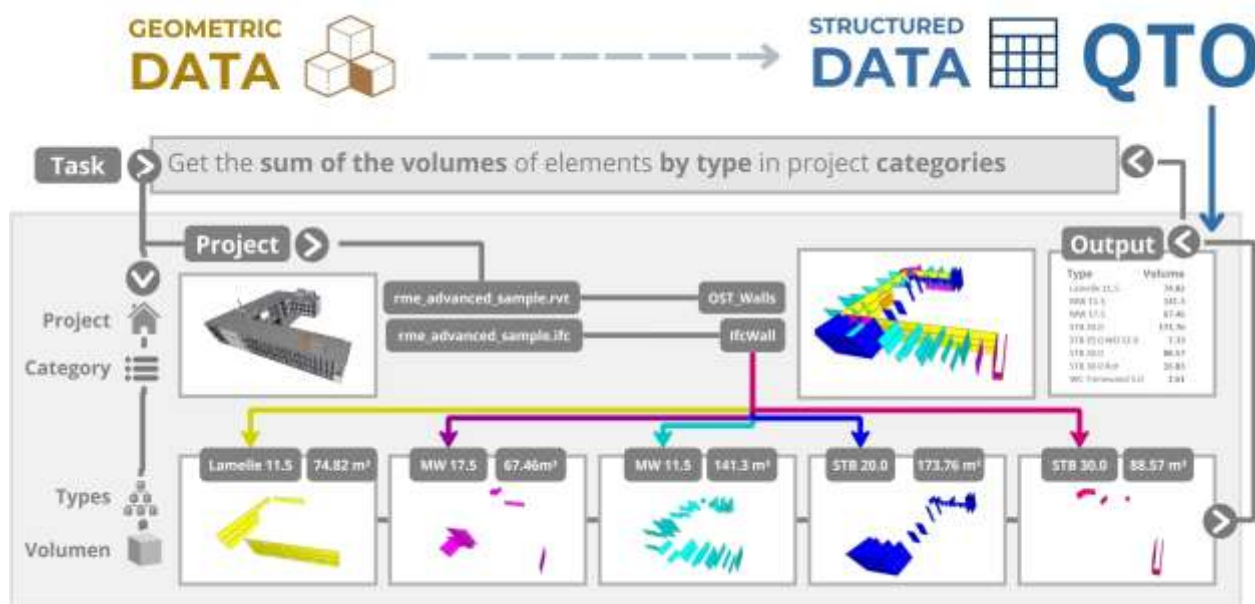
Automaatne koguseline arvestusprotsess (QTO) on oluline täpsete arvutuste ja hinnangute koostamiseks, inimlike vigade vähendamiseks mahu täpsustamise protsessides ja materjalide üle- või alatellimise vältimiseks.

QTO-protsessi näitena vaatleme tavalist juhtumit, kui CAD-andmebaasist on vaja näidata teatud kategooria, elementide klasside jaoks tabelis esitatud mahud elementide tüüpide kaupa. Rühmitame kõik projekti elemendid tüübi järgi CAD-projekti seinakategooriast ja võtame kokku iga tüübi mahu atribuudid, et esitada tulemus QTO mahutabelina (joonis 5.2-3).

Tüüpilise CAD-projekti näites (joonis 5.2-3) on kõik CAD-andmebaasis olevad seinakategooria elemendid rühmitatud seinatüübi järgi, nt "Lamelle 11.5", "MW 11.5" ja "STB 20.0", ning neil on selgelt määratletud ruumalaatribuudid, mis on esitatud meetriliste kuubikutena.

Projekteerijate ja arvutusspetsialistide vahelisel liidesel asuva juhi eesmärk on saada valitud kategooria elementide tüübiti automaatne mahutabel. Mitte ainult konkreetse projekti jaoks, vaid ka universaalsel kujul, mis on kohaldatav teiste sarnase mudelistruktuuriga projektide puhul. See võimaldab lähenemist skaleerida ja andmeid taaskasutada ilma dubleerimiseta.

Möödas on ajad, mil kogenud projekteerijad ja kalkulaatorid relvastusid joonlauaga, mõõtes hoolikalt iga joont paberil või PDF -plaanidel - traditsioon, mis ei ole aastatuhandete jooksul muutunud. 3D -modelleerimise arenguga, kus iga elemendi geomeetria on nüüd otseselt seotud automaatselt arvutatud mahuomadustega, on mahtude ja QTO koguste määramise protsess muutunud automatiseerituks.



Joonis 5.2-3 QTO mahu ja koguse atribuutide hankimine projektist hõlmab projekti elementide rühmitamist ja filtreerimist.

Meie näites on ülesanne "valida projekti seinte kategooria, rühmitada kõik elemendid tüübi järgi ja esitada teave mahu atribuutide kohta struktureeritud tabelina", nii et seda tabelit saaksid kasutada kümned teised spetsialistid kalkulatsioonide, logistika, tööplaanide ja muude äritegevuste arvutamiseks (joonis 6.1-3).

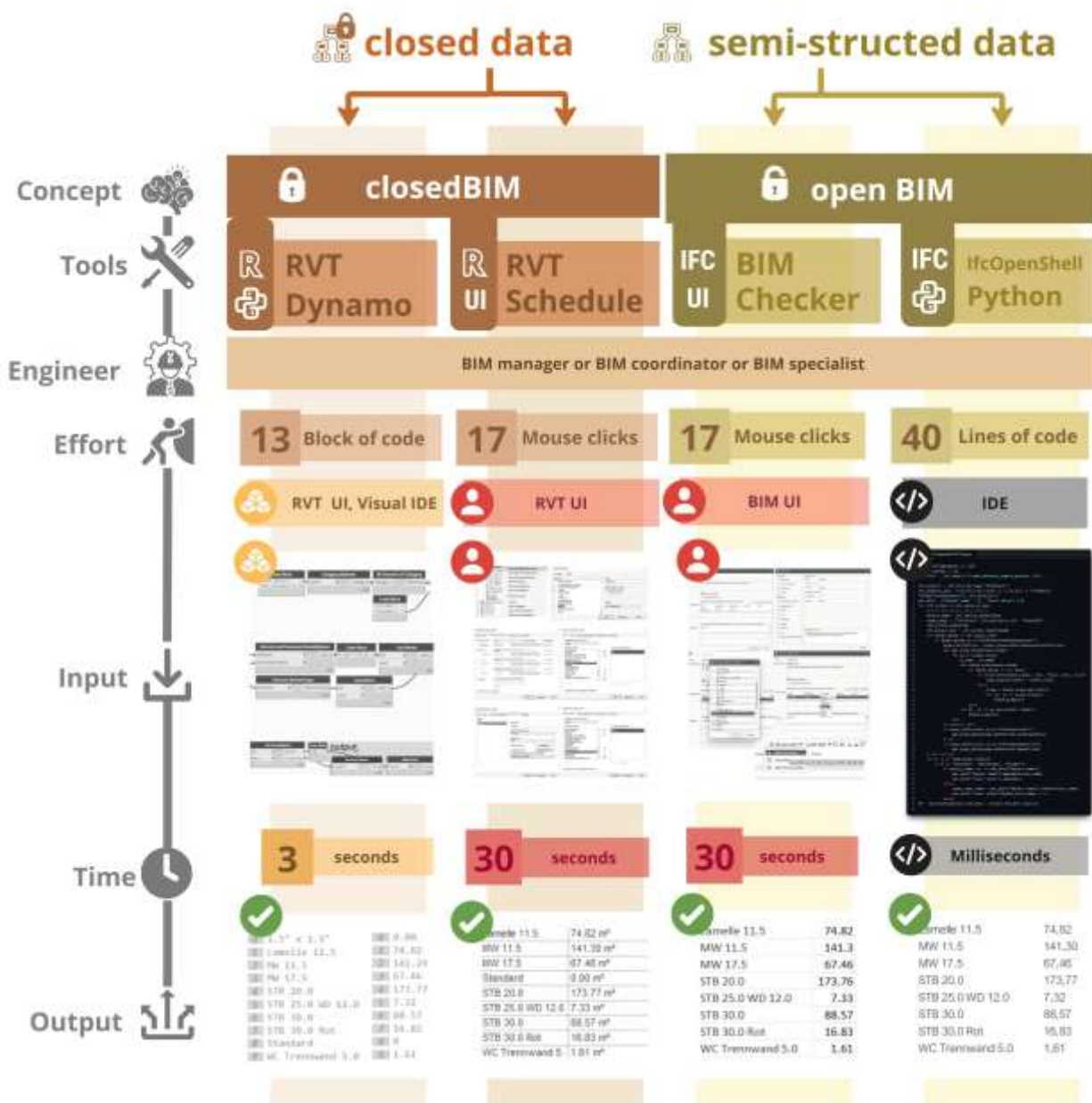
CAD-andmete suletud olemuse tõttu ei saa tänapäeval iga spetsialist kasutada otsest juurdepääsu CAD-andmebaasile (juurdepääsu probleemi põhjused ja lahendused on üksikasjalikult kirjeldatud raamatu kuuendas osas). Seetõttu peavad paljud inimesed pöörduma spetsiaalsete BIM-vahendite poole, mis põhinevad avatud BIMi ja suletud BIMi kontseptsioonidel [63]. Kui töötatakse spetsialiseeritud BIM tööriistadega või otse CAD-tarkvara keskkonnas, võib QTO (Quantity Take-Off) tulemustega tabelit genereerida erinevalt - sõltuvalt sellest, kas kasutatakse käsitsi kasutatavat liidest või tarkvara automatiseerimist.

Näiteks piisab CAD (BIM) tarkvara kasutajaliidese kasutamisel umbes 17 toimingust (nupuvajutusest), et saada valmis mahtude tabel (joonis 5.2-4). Siiski peab kasutaja olema hästi kursis CAD (BIM) tarkvara mudeli struktuuri ja funktsioonidega.

Kui automatiseerimist rakendatakse programmikoodi või CAD-programmide pluginate ja API-vahendite kaudu, väheneb mahtude tabelite saamiseks vajalike käsitsi tehtavate sammude arv, kuid sõltuvalt kasutatavast raamatukogust või vahendist tuleb kirjutada 40-150 rida koodi:

- **IfcOpSh (avatud BIM)** või **Dynamo IronPython (suletud BIM)** - võimaldab teil saada QTO tabeli CAD-formaadist või CAD-programmist vaid ~40 koodireaga.
- **IFC_js (avatud BIM)** - nõuab ligikaudu 150 rida koodi, et eraldada IFC-mudelst mahukaid atribuute.

- **Interface CAD tööriistad (BIM)** - võimaldab teil saada sama tulemuse käsitsi, 17 hiireklikiga.



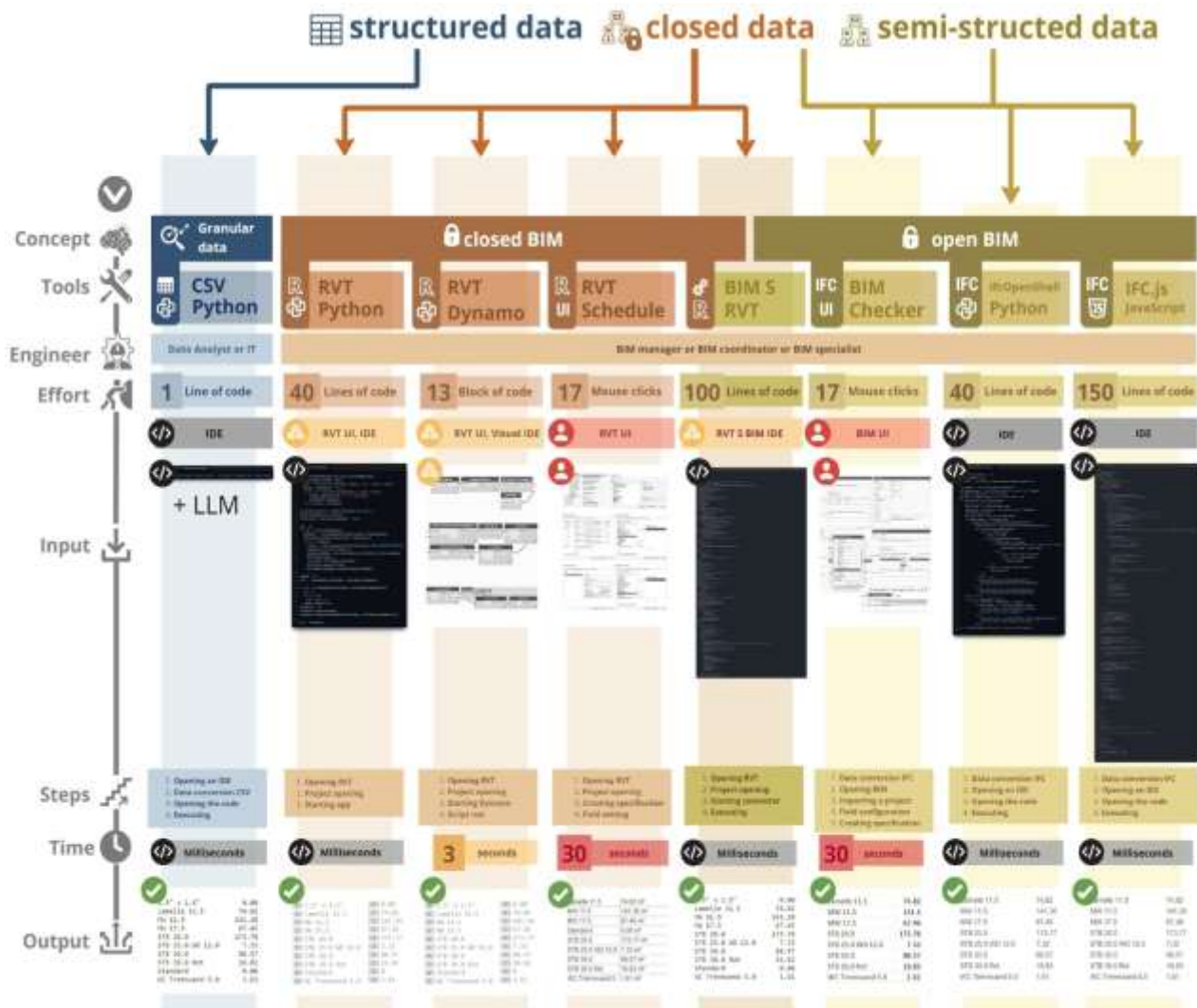
Joonis 5.2-4 CAD (BIM) projekteerijad ja juhid kasutavad QTO tabelite loomiseks 40-150 rida koodi või kümnekond klahvivajutust.

Tulemus on sama - struktureeritud tabel mahu atribuutidega elementide rühma jaoks. Ainus erinevus on tööjõukulu ja kasutajalt nõutavate tehniliste teadmiste tase (joonis 5.2-4). Kaasaegsed vahendid kiirendavad võrreldes käsitsi kogutud mahtude kogumisega oluliselt QTO protsessi ja vähendavad vigade tõenäosust. Need võimaldavad andmeid võtta otse projektimudelitest, välistades vajaduse arvutada mahud käsitsi ümber joonistest, nagu varem tehti.

Sõltumata kasutatavast meetodist - kas avatud BIM või suletud BIM - on võimalik saada identne QTO

- tabel projekti elementide mahtudega (joonis 5.2-4). Kui aga töötatakse projektandmetega CAD - (BIM-) kontseptsioonides, sõltuvad kasutajad spetsiaalsetest tööriistadest ja APIdest, mida pakuvad tarnijad (joonis 3.2-13). See tekitab täiendavaid sõltuvuskihti ja nõuab unikaalsete andmeskeemide õppimist, piirates samas otsest juurdepääsu andmetele.

CAD-andmete suletud olemuse tõttu raskendab QTO-tabelite ja muude parameetrite hankimine arvutuste automatiseerimist ja integreerimist välissüsteemidega. Kasutades vahendeid, mis võimaldavad otsest juurdepääsu andmebaasidele ja kandes projekti CAD -andmed pöördprojekteerimise vahendite abil avatud struktureeritud andmevormingusse (joonis 4.1-13), saab ühe koodireaga saada identse QTO tabeli (joonis 5.2-5 - granuleeritud andmetega variant).



Joonis 5.2-5 Erinevad tööriistad annavad samad tulemused projektiüksuste atribuutide tabelite kujul, kuid erineva tööjõukuluga.

Kui kasutatakse avatud struktureeritud andmeid CAD-projektidest, nagu on mainitud peatükis "CAD (BIM) andmete teisendamine struktureeritud kujul", on grupeerimisprotsess, QTO, oluliselt lihtsustatud.

Avatud struktureeritud andmete kasutamisel või otsesel juurdepääsul CAD-mudelite andmebaasidele põhinevad lähenemisviisid on vabad turunduspiirangutest, mis on seotud akronüümiga BIM. Need tuginevad teistes tööstusharudes juba ammu kasutatavatele ja tõestatud vahenditele (joonis 7.3-10 ETL protsess).

McKinsey uuringu "Avatud andmed: Unleash Innovation and Productivity with Streaming Information". [102], mis viidi läbi 2013. aastal, võib avatud andmete kasutamine luua võimalusi 30-50 miljardi dollari suuruse kokkuhoiu saavutamiseks aastas energiatootmisrajatiste projekteerimisel, projekteerimisel, hankimisel ja ehitamisel. See tähendab 15% kokkuhoidu ehituskapitali kuludes.

Töötamine avatud struktureeritud (granuleeritud) andmetega lihtsustab teabe otsimist ja töötlemist, vähendab sõltuvust spetsialiseeritud BIM-platvormidest ja avab tee automatiseerimisele, ilma et oleks vaja kasutada patenteeritud süsteeme või CAD-formaatide parameetrilisi ja keerulisi andmemudeleid.

QTO automatiseerimine kasutades LLM ja struktureeritud andmeid

Struktureerimata andmete teisendamine struktureeritud vormi parandab oluliselt erinevate protsesside tõhusust: see lihtsustab andmete töötlemist (joonis 4.1-1, joonis 4.1-2) ja kiirendab valideerimisprotsessi, kuna muudab nõuded selgemaks ja läbipaistvamaks, nagu me juba eelmistes peatükkides arutasime. Samamoodi lihtsustab CAD-andmete (BIM) tõlkimine struktureeritud avatud vormi (joonis 4.1-12, joonis 4.1-13) atribuutide rühmitamise protsessi ja QTO protsessi.

QTO atribuutide tabelil on struktureeritud vorm, nii et struktureeritud CAD andmete kasutamisel töötame ühe andmemudeliga (joonis 5.2-5), mis välistab vajaduse konverteerida ja tõlkida projekti andmemudeleid ja rühmitamisreegleid ühisele nimetaja. See võimaldab meil rühmitada andmeid ühe või mitme atribuudi järgi vaid ühe koodireaga. Seevastu avatud BIMis ja suletud BIMis, kus andmed on salvestatud poolstruktureeritud, parameetrilises või suletud vormingus, nõuab töötlemine kümneid või isegi sadu koodiridu ja API kasutamist, et suhelda geomeetria- ja atribuutide teabega.

- 🗨️ Näide QTO struktureeritud projekti grupeerimisest ühe atribuudi järgi. Teksti päring mis tahes LLM jututuba (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN või mõni muu):

Mul on CAD -projekt kui DataFrame - palun filtreerige projekti andmed, et saada elemendid, mille parameeter "Type" sisaldab ainult "Type 1" väärtust. ↩

- LLMi vastus on väga tõenäoliselt Pythoni koodi kujul kasutades Pandas :



Joonis 5.2-6 Üks rida koodi, mis on kirjutatud LLMiga , võimaldab rühmitada kogu CAD projekti atribuudi "Type" järgi ja saada soovitud elementide rühma.

Tänu kahemõõtmelise DataFrame lihtsale struktuurile ei ole meil vaja selgitada LLM skeemi ja andmemudelit, mis lühendab tõlgendamise etappe ja kiirendab lõplike lahenduste loomist. Varem nõudis isegi lihtsa koodi kirjutamine programmeerimiskeelte õppimist, kuid nüüd võimaldavad moodsad keelemudelid (LLM) struktureeritud andmetega töötades tekstpäringute abil automaatselt protsessiloogika koodiks tõlkida.

LLM automatiseerimine ja keelemudelid võivad täielikult kõrvaldada CAD (BIM) andmete rühmitamise ja töötlemisega tegelevate spetsialistide vajaduse õppida programmeerimiskeeli või BIM-vahendeid, pakkudes võimalust lahendada probleeme tekstiküsimuste abil.

Sama päring - kõikide projektielementide rühmitamine kategooriast "seinad" ja iga tüübi mahu arvutamine (joonis 5.2-5) - mis CAD-keskkonnas (BIM) nõuab 17 klõpsu kasutajaliideses või 40 rida koodi kirjutamist, näeb avatud andmetöötlusvahendites (nt SQL või Pandas) välja lihtsa ja intuitiivse päringuna:

- Ühe reaga Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

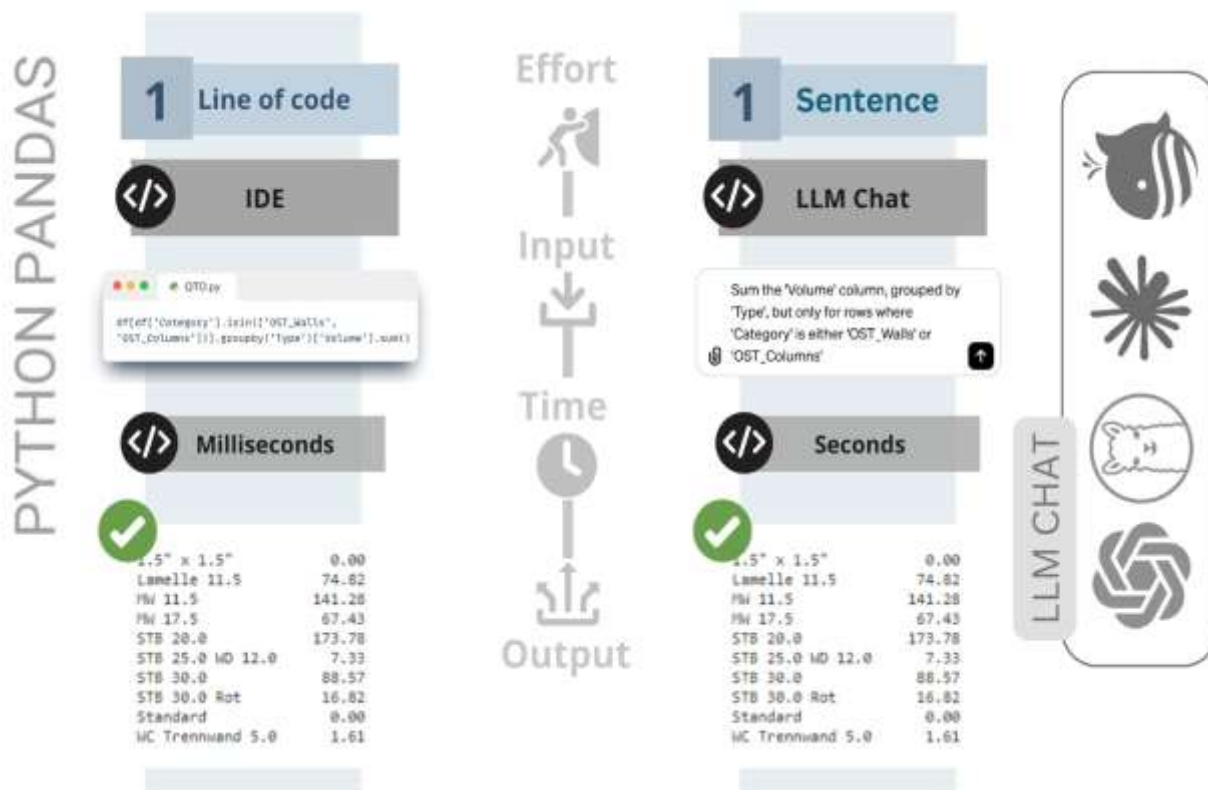
Koodi dekodeerimine: võtta df-st (DataFrame) elemendid, mille atribuut-sammas "Category" on väärtusega "OST_Walls", rühmitada kõik saadud elemendid atribuut-samba "Type" järgi ja summeerida saadud elementide rühma jaoks atribuut "Volume".

- CAD-ist saadud struktureeritud projekti rühmitamine SQL-i abil:


```
SELECT Type, SUM(Volume) AS TotalVolume
FROM elemendid
WHERE Kategooria = "OST_Seinad".
GROUP BY Type;
```

- LLMi abil saame kirjutada rühmitamise taotluse projekti andmebaasi lihtsa tekstiviidetena - palve (joonis 5.2-7):

Projekti andmeraamistiku puhul rühmitage elemendid parameetri "Type" järgi, kuid ainult nende elementide puhul, mille parameeter "Category" on võrdne "OST_Walls" või "OST_Columns", ja summeerige saadud veeru parameeter "Volume". ↵

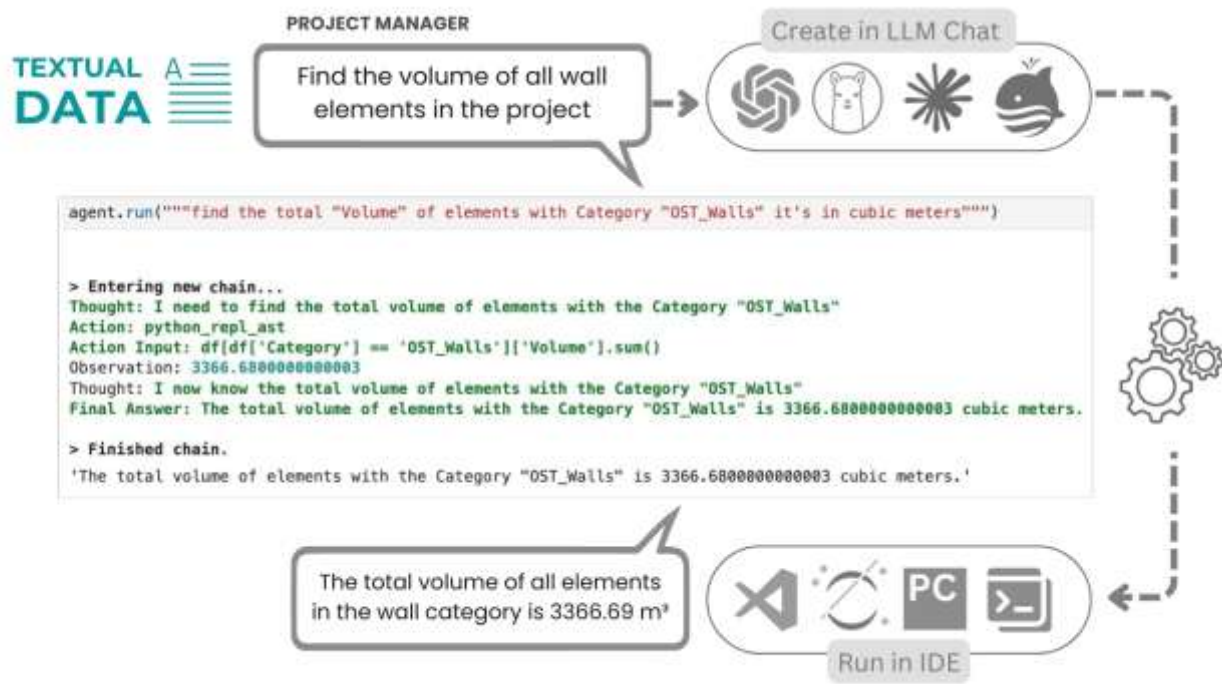


Joonis 5.2-7 Kasutades SQL, Pandas ja LLM andmetötluse automatiseerimine on nüüd võimalik mõne koodirea ja tekstipäringuga.

QTO saamine CAD-andmetest kasutades LLM-vahendeid (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), muudab oluliselt traditsioonilisi meetodeid atribuutteabe, kvantitatiivsete ja mahuliste andmete eraldamiseks üksikute objektide ja objektide rühmade kohta.

Nüüd saavad isegi projektijuhid, kalkulatsiooni- või logistikaspetsialistid, kellel ei ole süvateadmisi projekteerimisest ja kellel ei ole spetsiaalset CAD-tarkvara - (BIM-) müüjad, kellel on juurdepääs CAD-andmebaasile, sekundi ajaga, lihtsalt päringut kirjutades või dikteerides, teada seina- või muude objektide kategooria elementide kogumahu.

Tekstpäringute puhul (joonis 5.2-8) töötleb mudeli LLM-agent kasutaja rakendada teatud funktsiooni ühele või mitmele parameetrile - tabeli veergudele. Selle tulemusena saab kasutaja LLM-iga suheldes kas uue veeruparameetri uute väärtustega või ühe konkreetse väärtuse pärast rühmitamist



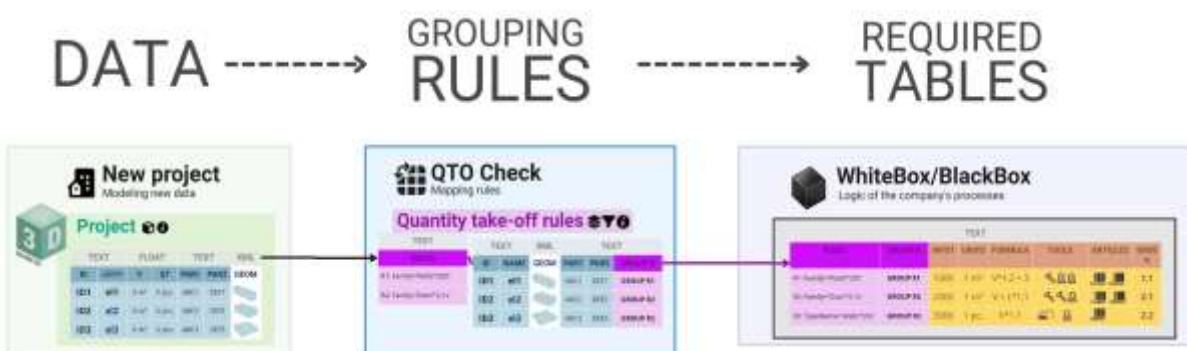
Joonis 5.2-8 LLM mudel, mis töötab struktureeritud andmetega, mõistab teksti päringu kontekstist, millist rühmitust ja atribuute kasutaja küsib.

Kui on vaja saada koguseid ainult ühe elementide rühma kohta, piisab lihtsast QTO päringust (joonis 5.2-7) CAD-mudeli andmete kohta. Kui aga arvutatakse eelarvet või kalkulatsiooni kogu projekti jaoks, mis koosneb paljudest elementide rühmadest, on sageli vaja saada kogused kõigi elementide tüüpide (klasside) kohta, kus iga elementide kategooriat töödeldakse eraldi - rühmitades neid asjakohaste atribuutide järgi.

Hindajate ja hindajate praktikas kasutatakse eri tüüpi objektide puhul individuaalseid rühmitamis- ja arvutamise reegleid. Näiteks aknad grupeeritakse tavaliselt korruste või tsoonide järgi (grupeerimisparameeter - atribuut Level, Rooms) ja seinad - materjali või konstruktsiooni tüübi järgi (parameeter Material, Type). Grupeerimisprotsessi automatiseerimiseks on sellised reeglid eelnevalt kirjeldatud rühmitamisreeglite tabelite kujul. Need tabelid toimivad konfiguratsioonimustritena, mis määravad, milliseid atribuute tuleb projekti iga elementide rühma arvutustes kasutada.

QTO kogu projekti arvutamine, kasutades Exceli tabelis esitatud grupireegleid.

Reaalsetes ehitusprojektides on sageli vaja teostada ühe elementide rühma piires samaaegselt mitme atribuudi järgi koondamist. Näiteks kui töötate kategooriaga "Aknad" (kus atribuut Kategooria sisaldab väärtusi nagu OST_Avad või IfcAvad), saab elemente rühmitada mitte ainult tüübi järgi - näiteks väljal Tüübinimi või Tüüp oleva väärtuse järgi - vaid ka täiendavate omaduste, näiteks vastava atribuudiga määratud soojusjuhtivuse taseme järgi. Selline mitmemõõtmeline rühmitamine võimaldab konkreetse rühma jaoks täpsemaid tulemusi. Samamoodi saab seina- või põrandakategooriate arvutamisel kasutada filtreid või rühmitamiskriteeriumidena atribuutide - näiteks materjali, taseme, põranda, tulekindluse ja muude parameetrite - suvalisi kombinatsioone (joonis 5.2-9).



Joonis 5.2-9 Iga projekti üksuste rühma või kategooria jaoks on olemas erinev grupeerimisvalem, mis koosneb ühest või mitmest kriteeriumist.

Selliste rühmitamisreeglite määratlemise protsess on sarnane andmenõuete loomise protsessiga, mida on kirjeldatud peatükis "Nõuete loomine ja kvaliteedi kontrollimine andmed" (joonis 4.4-5), kus me käsitlesime üksikasjalikult tööd andmemudelitega. Sellised grupeerimis- ja arvutusreeglid tagavad tulemuste täpsuse ja asjakohasuse, et arvutada automaatselt mingi üksuste kategooria koguse või mahu kogumisatribuute, võttes arvesse kõiki vajalikke tingimusi, mida tuleb arvutustes ja arvutustes arvesse võtta.

- Järgnev koodinäide filtreerib projektide tabelit nii, et saadud andmekogum sisaldab ainult selliseid üksusi, mille atribuudi "Kategooria" veerg sisaldab väärtusi "OST_Windows" või "IfcWindows" ja samal ajal sisaldab atribuudi "Tüüp" veerg väärtust "Tüüp 1":

Mul on DataFrame projekt - filtreerin andmeid nii, et andmekogumisse jääksid ainult need elemendid, mille atribuut "Kategooria" sisaldab väärtusi "OST_Windows" või "IfcWindows" ja samal ajal sisaldab atribuut Type väärtust "Type 1"

LLMi vastus :



Joonis 5.2-10 Üks koodirida, mis sarnaneb Exceli valemiga , võimaldab teil rühmitada kõiki projektiüksusi mitme atribuudi järgi.

Saadud kood (joonis 5.2-10) pärast CAD-andmete tõlkimist struktureeritud avatud formaatidesse (joonis 4.1-13) saab käivitada mõnes populaarses IDE-s (integreeritud arenduskeskkonnas), mida me eespool mainisime, offline-režiimis: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse koos PyDev pluginaga, Thonny, Wing IDE, IntelliJ IDEA koos Python pluginaga, JupyterLab või populaarsed veebipõhised vahendid: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Et saada QTO DataFrame vormil kategoorias "Windows" ainult konkreetse soojusjuhtivuse väärtusega projektiüksused, saame kasutada järgmist päringut LLMile:

Mul on DataFrame projekt - filtreerin andmeid nii, et andmekogumisse jääksid ainult kirjed, mille "Kategooria" sisaldab "OST_Windows" või "IfcWindows" väärtusi, ja samal ajal peaks veerus ThermalConductivity olema väärtus 0,4

LLMi vastus :



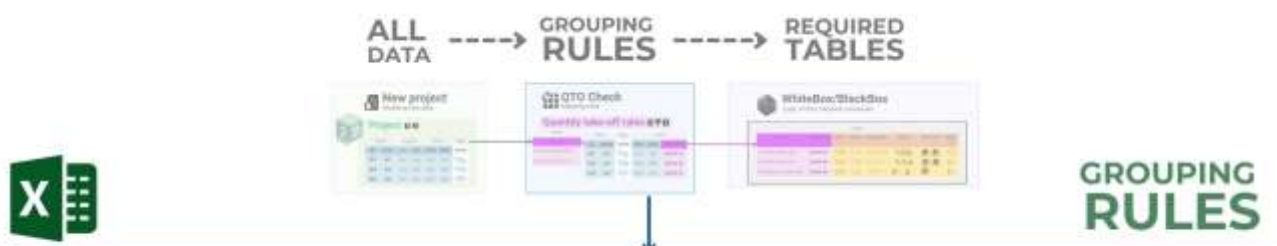
Joonis 5.2-11 Äärmiselt lihtne Pandas päringukeel Python võimaldab teil käivitada QTOsid suvalise arvu projektide jaoks samaaegselt.

LLMilt saadud vastuses (joonis 5.2-11) kasutatakse loogilist tingimust "&" kahe kriteeriumi ühendamiseks: soojusjuhtivuse väärtus ja kuulumine ühte kahest kategooriast. Meetodiga "isin" kontrollitakse, kas atribuudi veeru "Kategooria" väärtus sisaldub esitatud loetelus.

Suure hulga elementide rühmade ja erineva rühmitamisloogikaga projektide puhul - iga projektiüksuse kategooria (nt aknad, ukSED, plaadid) jaoks tuleks kehtestada individuaalsed rühmitamisreeglid, mis võivad sisaldada täiendavaid koefitsiente või atribuutide summaarse arvutamise valemeid. Need valemid (joonis 5.2-12 atribuudi "formel", nt x-väärtus koguse ja y-maht grupi kohta) ja koefitsiendid võtavad arvesse näiteks iga grupi unikaalseid omadusi:

- %-lised lisad materjali mahule, et võtta arvesse ületamisi.
- fikseeritud täiendav materjalikogus
- võimalike riskide ja arvutusvigadega seotud kohandused valemities

Kui filtreerimis- ja grupeerimisreeglid on iga kirje kategooria jaoks parameetrivalemite kujul sõnastatud, võib need salvestada rida-realt tabelina - näiteks Exceli formaadis (joonis 5.2-12). Nende reeglite struktureeritud kujul salvestamise abil saab projektiandmete väljavõtte, filtreerimise ja rühmitamise protsessi täielikult automatiseerida. Selle asemel, et käsitsi kirjutada palju eraldi päringuid, loeb süsteem lihtsalt parameetrite tabelit ja kohaldab asjakohaseid reegleid mudelile (projekti üldisele andmeraamistikule (joonis 4.1-13)), genereerides lõplikud QTO tabelid iga projektielementide kategooria jaoks.



pos	code	description	Parametr 1	Expression1	Parametr 2	Expression2	Aggregate Parameter	formel
F1	3	Walls						
W1.1	3.1.1	Interior - 165	Category	OST_Walls	Type Name	Interior - 165*	Area	x*1.5+y*0.1
W1.2	3.1.2	SIP Wall 202mm	Category	OST_Walls	Type Name	SIP 202mm Wall*	Volume	x/1.2-y+0.4
W1.3	3.1.2	CL W1	Category	OST_Walls	Type Name	CL_W1*	Area	x*0.5*y+2
W2.1	3.2.2	Retaining 300mm	Category	OST_Walls	Type Name	Retaining - 300mm*	Volume	x*1.1+8*y*0.2

Joonis 5.2-12 QTO atribuutide rühmitamise tabel kehtestab reeglid projekti elementide rühmitamiseks, andes iga kategooria jaoks täpse koguarvu ja ulatuse.

Kogutud reeglid võimaldavad teil kogu projekti rühmitada ja teostada kõik vajalikud arvutused, sealhulgas mahu atribuutide kohandamine. Selle tulemusena viiakse mahud "tegelikku mahtu", mida kasutatakse arvutuste ja arvutuste tegemiseks, mitte neid, mis olid algsest CAD-mudelisis projekteerimise etapis.

QTO mahtude tabelite automaatse loomise protsessis kogu projekti jaoks peab rakendus läbima kõik grupeerimisreeglite tabeli kategooriad, võtma grupeerimisatribuudid, grupeerima kõik projekti elemendid nende järgi ja koondama selle grupi mahtude atribuudi, korrutades selle lisaks täpsustava teguriga või koefitsiendiga.

Palume LLM kirjutada meile sellise lahenduse jaoks kood, kus kood peaks laadima kaks tabelit - grupeerimisreeglite tabelit (joonis 5.2-12) ja projekti enda andmete tabelit (joonis 4.1-13), seejärel rakendama grupeerimisreegleid, grupeerima elemendid vastavalt antud reeglitele, arvutama agregeeritud väärtused ja salvestama tulemused uude Exceli faili.

🗨️ Saatke tekstitaotlus aadressil vestluskeskkonnas LLM:

Mul on vaja koodi, et lugeda projekti andmeid failist 'basic_sample_project.xlsx' ja seejärel reegleid failist 'Grouping_rules_QTO.xlsx', mille abil rühmitada kõik andmed 'Parameter 1' ja 'Parameter 2' järgi, agregeerida 'Aggregate Parameter', filtreerida 'Expression2' järgi, teha arvutusi 'Formel1' järgi ja salvestada QTO tabelisse 'QTQ_table2.xlsx'📄

LLMi vastus :



```

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 ].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('{{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["formel"], {'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)

```

Joonis 5.2-13 Pythoni kood Pandas töötleb projekteerimisandmeid CAD-projektist , grupeerides neid vastavalt Excelis määratletud reeglitele.

Koodi täitmise lõpptulemus (joonis 5.2-13) on üksuse-grupi tabel, mis sisaldab mitte ainult algse CAD-(BIM-) mudeli üldistatud ruumalaatribuute, vaid ka uut reaalselt ruumalaatribuuti, mis võtab arvesse kõiki nõudeid arvutuste ja hinnangute korrektseks koostamiseks (näide joonis 5.2-14).

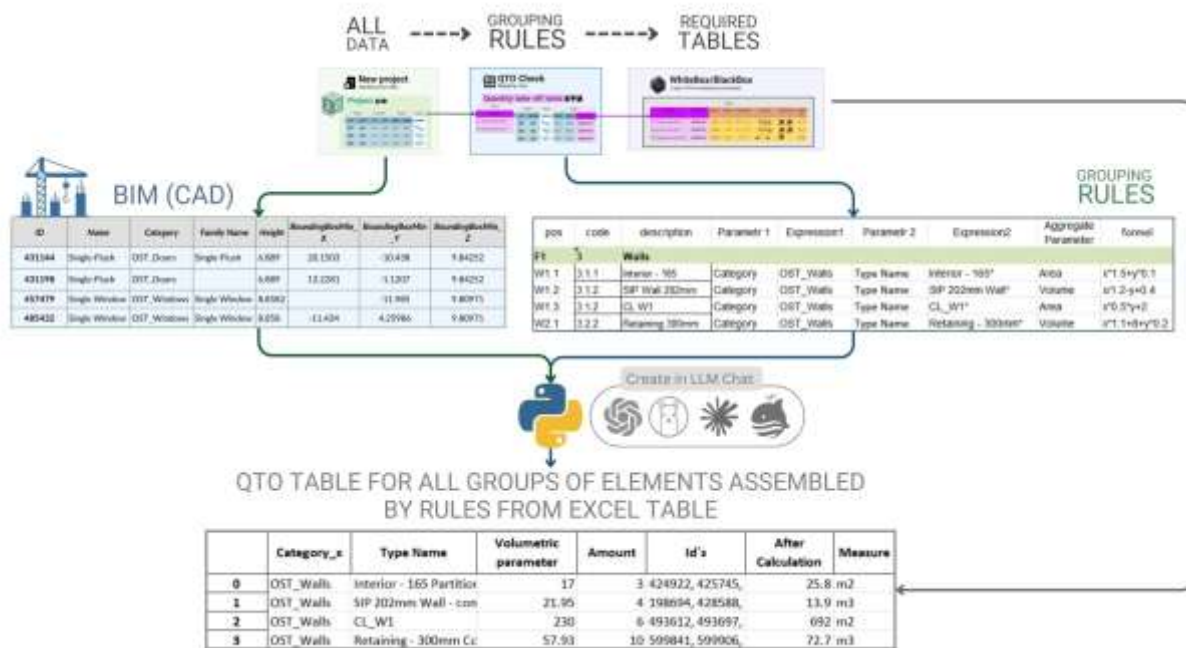
 QTO TABLE FOR ALL GROUPS OF ELEMENTS
ASSEMBLED BY RULES FROM EXCEL TABLE

	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partitor	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	

Joonis 5.2-14 Atribuut "Pärast arvutust" lisatakse kokkuvõtte tabelisse pärast koodi täitmist, mis automaatselt arvutab tegeliku mahu.

Saadud koodi (joonis 5.2-13) saab käivitada mõnes populaarses IDE-s (mida me eespool mainisime) ja rakendada ükskõik kui paljudele olemasolevatele või uutele sissetulevatele projektidele (RVT, IFC,

DWG, NWS, DGN jne), olgu see siis mõni projekt või ehk sadu projekte erinevates formaatides struktureeritud kujul (joonis 5.2-15).



Joonis 5.2-15 Automaatne ehitusandmete rühmitamise protsess seob BIM-andmed (CAD) QTO tabelitega reeglite kaudu Exceli tabelist.

Kohandatud ja parametrizeeritud mahuandmete kogumise protsess (joonis 5.2-15) võimaldab täielikult automatiseeritud andmete kogumist projekti elementide kvantitatiivsete tunnuste ja mahtude kohta, et nendega edasi töötada, sealhulgas kulude hindamist, logistikat, tööplaanide ja süsiniku jalajälje arvutamist ning muid analüütilisi ülesandeid.

Olles õppinud vahendeid, mis võimaldavad meil hõlpsasti korraldada ja rühmitada projektielementide gruppe teatud tunnuste järgi, oleme nüüd valmis integreerima rühmitatud ja filtreeritud projekte erinevate ettevõtte arvutuste ja äristsenariumidega.



PEATÜKK 5.3.

4D, 6D -8D JA SÜSINIKDIOKSIIDI HEITKOGUSTE ARVUTAMINE CO₂.

4D-mudel: aja integreerimine ehitushinnangutesse

Lisaks kalkulatsioonile on projekteerimisandmete üks peamisi rakendusi ehituses ajaparametrite määramine - nii üksikute ehitustööde kui ka kogu projekti jaoks. Ressursipõhist hindamismeetodit ja sellega seotud arvutusandmebaasi, mida käsitleti üksikasjalikult eelmises peatükis "Ehitusprojektide arvutused ja hinnangud", kasutatakse sageli aja automatiseeritud arvutamise ja tööde teostamise ajakava koostamise alusena.

Ressursipõhine lähenemisviis võtab arvesse mitte ainult materjalikulusid, vaid ka ajalisi ressursse. Kulukalkulatsioonis saab igale protsessile määrata töökorralduse atribuudi (joonis 5.3-1 - Töökorralduse parameeter) ning selle protsessi teostamisega seotud aja- ja maksumuse. Need parameetrid on eriti olulised selliste toimingute kirjeldamisel, millel ei ole fikseeritud turuhinda ja mida ei saa otseselt osta - näiteks ehitusseadmete kasutamine, tööliste kasutamine või logistilised protsessid (mida tavaliselt väljendatakse tundides). Sellistel juhtudel ei määra kulusid kindlaks hankeosakond, vaid otse rakendusetevõtte, lähtudes sisemistest normidest või tootmismahitudest (joonis 5.3-1).

Position: **Concrete Foundation Block per 1 piece.** Unit price: **\$11,934.00**

Code	Description	Bid. Factor	Quantity	Unit	Unit cost	Total	Work order
LabPr	Preparation Work Labor	1	16.00	hr	€ 30.00	€ 480.00	1.1
EqEx	Excavation Equipment	1	16.00	hr	€ 80.00	€ 1,280.00	1.2
FormW	Formwork	1	500.00	sq ft	€ 2.50	€ 1,250.00	
ReSt	Reinforcing Steel	1	800.00	lb	€ 0.75	€ 600.00	
ConcB	Concrete	1	30.00	cu yd	€ 120.00	€ 3,600.00	
LabCP	Concrete Pouring Labor	1	24.00	hr	€ 35.00	€ 840.00	2.1
EqCM	Concrete Mixer	1	8.00	hr	€ 45.00	€ 360.00	2.2
LabFI	Finishing Labor	1	24.00	hr	€ 30.00	€ 720.00	2.3
LabCu	Curing Labor	1	8.00	hr	€ 25.00	€ 200.00	3.1
EqOt	Other Equipment	1	10.00	hr	€ 15.00	€ 150.00	3.2
FuelD	Diesel for Equipment	1	40.00	gal	€ 3.50	€ 140.00	
MiscL	Lubricants and Maintenance	1	1.00	lump sum	€ 200.00	€ 200.00	
TransM	Transportation of Materials	1	1.00	lump sum	€ 300.00	€ 300.00	
OverH	Overhead Costs	1	1.00	percentage	10% of Total	€ 907.00	
Prof	Profit Margin	1	1.00	percentage	10% of Total	€ 907.00	

Code	Description	Unit	Unit cost
Exc_d	Excavation of the foundation to the top	hour	80.00
Exc_m	Excavation to the place of foundation	hour	100.00
Transport	Material profit	%	10.00

4D

Joonis 5.3-1 Ressursipõhise hindamismeetodi tööarvutused hõlmavad töötundide ajastamist.

Seega ei hõlma kalkulatsioonid kulude tasandil mitte ainult kütuse- ja materjalikulusid (ostukulud), vaid ka juhtide, tehnikute ja abitöötajate poolt kohapeal kulutatud aega. Antud näites (joonis 5.3-1) on

kulude tabelis arvatud vundamendiploki paigaldamise maksumus, mis sisaldab töö etappe, nagu ettevalmistus, karkassi paigaldamine ja betooni valamine, samuti vajalikke materjale ja tööjõudu. Üksikud toimingud, näiteks ettevalmistustööd, ei pruugi siiski sisaldada materjalikulusid, kuid võivad sisaldada märkimisväärsed ajutisi tööjõukulusid, mis on väljendatud töötundides.

Tööde järjekorra (tööde ajakava) planeerimiseks ehitusplatsil lisatakse arvutustabelisse käsitsi atribuut "Töökorraldus" (joonis 5.3-1). See atribuut määratakse lisasambas ainult nende objektide puhul, mille mõõtühik on väljendatud ajas (tund, päev). See atribuut on lisaks töö koodile, kirjeldusele, kogusele, mõõtühikule (parameeter "Unit") ja kuludele. Tegevuste numbriline järjekord (parameeter "Work order") võimaldab määrata tööde teostamise järjekorra ehitusplatsil ja kasutada seda ajakava koostamisel.

Ehitusgraafik ja selle automatiseerimine kulukalkulatsiooni andmete alusel

Ehitusgraafik on projekti elluviimise käigus teostatavate tööde ja protsesside visuaalne esitus. See koostatakse üksikasjalike ressursiarvutuste alusel (joonis 5.3-1), kus iga ülesanne-töö on lisaks ressursside maksumusele planeeritud aja ja järjekorra järgi.

Erinevalt keskmistamise meetodist, kus ajakulu hinnangud põhinevad materjalide või seadmete paigaldamise tüüpilistel tundidel, põhineb planeerimine ressursipõhise meetodi puhul tegelikel andmetel kulukalkulatsioonis. Iga tööjõukuludega seotud kalkulatsiooni kirje põhineb kohaldatud kalendril, mis võtab arvesse ressursside kasutamise tegelikke tingimusi tööperioodi jooksul. Tootlike töötundide kohandamine koefitsientide abil kulukalkulatsiooni tasandil (joonis 5.3-1 parameeter "Bid. Factor"), võimaldab võtta arvesse erinevusi tootlikkuses ja hooajalisi eripärasid, mis mõjutavad tööde ajastust.

Selleks, et määrata Gantt diagrammi ehitusgraafiku protsessi algus- ja lõpukuupäevad, võtame vundamendiplokkide kalkulatsioonist iga elemendi ajaatribuudi väärtused ja korrutame need plokkide arvuga (antud juhul betoonvundamendiplokkide arvuga). See arvutus annab iga ülesande kestuse. Seejärel joonistame need kestused graafiku loomiseks ajaskaalal, alustades projekti alguskuupäevast, ning tulemuseks on visuaalne esitus, mis näitab, millal iga ülesanne peaks algama ja lõppema. Protsesside parameeter "Töökorraldus" võimaldab meil lisaks mõista, kas tööprotsess toimub paralleelselt ("Töökorraldus" nt 1.1-1.1) või järjestikku (1.1-1.2).

Gantt diagramm on graafiline tööriist planeerimiseks ja projektijuhtimiseks, mis kujutab ülesandeid horisontaalsete ribadena ajaskaalal. Iga riba tähistab ülesande kestust, selle algust ja lõppu.

Ajagraafik ehk Gantt'i diagramm aitab projektijuhtidel ja töötajatel selgelt mõista, millal ja millises järjekorras tuleb erinevad ehitusetapid läbi viia, tagades nii ressursside tõhusa kasutamise ja tähtaegadest kinnipidamise.

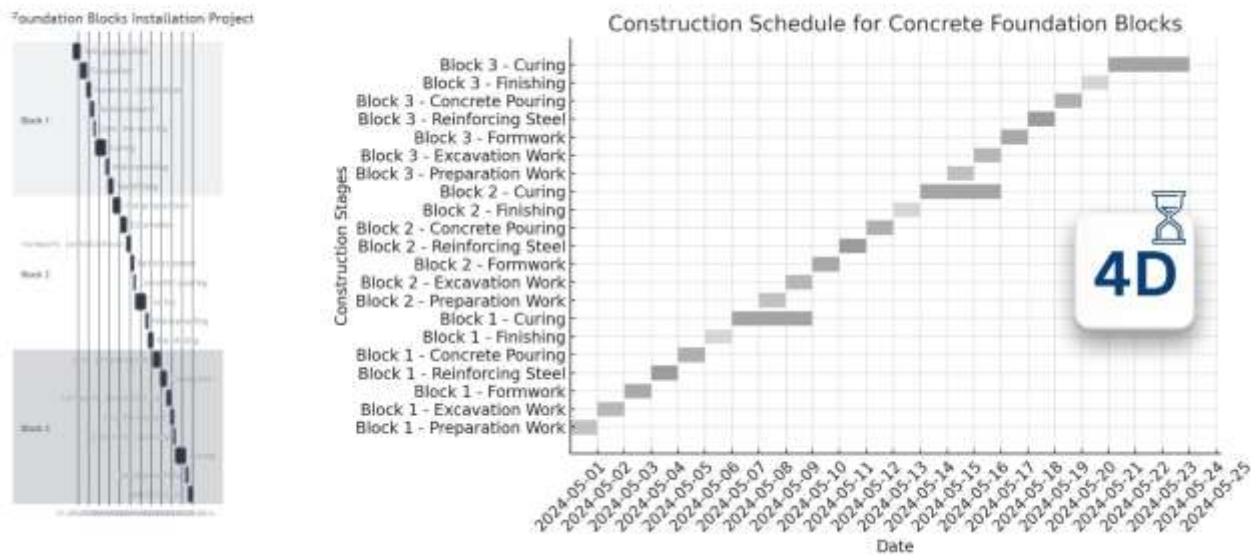
Esitame kolme betoonvundamendiploki paigaldamise ajakava, kasutades ülaltoodud tabelis esitatud arvutusi. Kasutades ülaltoodud näite kulutabeli (joonis 5.3-1), palume LLMil koostada ajakava kolme vundamendiploki elemendi paigaldamiseks, näiteks 2024. aasta esimeseks maiks.

Kalkulatsiooni saatmiseks LLM-i, saame laadida üles kalkulatsioonitabelit XLSX-formaadis või lihtsalt sisestada kalkulatsiooni JPEG-kujutise ekraanipildi otse LLM-i vestlusesse (joonis 5.3-1). LLM leiab iseseisvalt raamatukogu tabeli pildi visualiseerimiseks ja lisab kõik andmed graafikusse, korrutades tegevuste ajaatribuudid tabelist nende mahuga.

🗨️ Saada tekstitaotlus aadressile LLM:

Koostage Gantti diagrammi kujul tööplaani, kasutades 3 vundamendiploki paigaldamise jaoks ajaga seotud tabeli väärtusi (joonis 5.3-1 JPEG-formaadis lisas). Iga ploki tööd tuleb teha järjestikku. Tööde alguseks määratakse 01/05/2024. ⏸️

🗨️ LLMi vastus

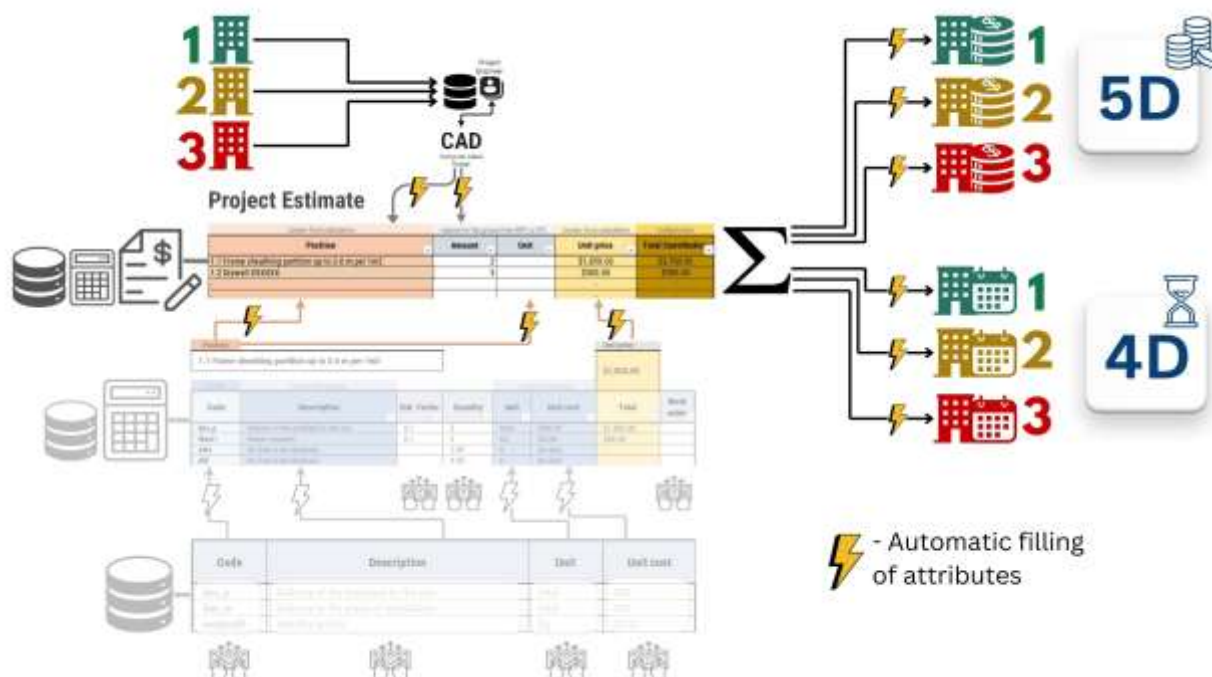


Joonis 5.3-2 Automaatselt genereeritud mitu LLM-i Gantti diagramm näitab kolme betoonploki ehitamise etappe vastavalt ülesande tingimustele.

Saadud graafik (joonis 5.3-2) on ajadiagramm, kus iga horisontaalne riba vastab teatavale tööetapile vundamendiplokil ja näitab tööde järjestust (parameeter "Tööde järjekord"), nagu ettevalmistus, kaevamine, šali paigaldamine, armeerimine, betoonivalu ja viimistlemine, st need protsessid, mille ajalised parameetrid ja järjestus on arvutustes täidetud.

Selline ajakava (joonis 5.3-2) ei võta arvesse tööpäevade, vahetuste või tööajanormidega seotud piiranguid, vaid on mõeldud üksnes protsessi kontseptuaalseks visualiseerimiseks. Täpset ajakava, mis kajastab töö paralleelsust, saab täiendada asjakohaste promtide või täiendavate juhistega jututoa sees

Kasutades ühtset kuluarvestust (joonis 5.3-1), on tänu 3D -geomeetriast saadud mahuaatribuutidele võimalik automaatselt hinnata nii projekti maksumust automatiseeritud hinnangute kaudu kui ka samal ajal arvutada tabelite või graafikute kujul rühmade ajalisi omadusi erinevate projektivariantide jaoks (joonis 5.3-3).



Joonis 5.3-3 Automaatne arvutus, võimaldab koheselt ja automaatselt prognoosida kulusid ja aega erinevate projektivariantide jaoks.

Kaasaegsed modulaarsed ERP -süsteemid (joonis 5.4-4), mis laadivad andmeid CAD mudelistest, kasutavad sarnaseid automaatseid ajaarvutusmeetodeid, mis vähendavad oluliselt otsustusprotsessi. See võimaldab koheselt ja täpselt planeerida tööplaan ning arvutada projekti realiseerimise kõigi ülesannete täitmiseks kuluvat koguaega, võttes arvesse reaalseid hindu.

Laiendatud atribuutkihid 6D -8D: energiatõhususest ohutuse tagamiseni.

6D, 7D ja 8D on infomodelleerimise laiendatud tasemed, millest igaüks lisab projekti terviklikku infomodelisse täiendavaid atribuutide kihte, mille aluseks on 3D -mudeli atribuudid koos nende arvu ja ulatusega. Iga lisakiht toob sisse konkreetsed parameetrid, mida on vaja edasiseks rühmitamiseks või edasiseks identifitseerimiseks teistes süsteemides, nagu näiteks kinnisvarahaldussüsteemid (PMS), arvutipõhine rajatiste haldamine (CAFM), ehitusprojektide haldamine (CPM) ja ohutusjuhtimissüsteemid (SMS).



Joonis 5.3-4 Atribuudid 6D, 7D ja 8D andmemudelisse laiendavad projekti erinevate aspektide arvestamist, alates energiatõhususest kuni ohutusele.

- **6D-s** lisatakse lisaks projekti andmebaasile (või andmeraamistikule (joonis 4.1-13)), mis sisaldab elementide geomeetrilisi ja ruumilisi atribuute, teave (atribuutide veerud) keskkonnasäästlikkuse kohta. See hõlmab teavet energiatõhususe, süsiniku jalajälje, materjalide ringlussevõetavuse ja keskkonnasõbralike tehnoloogiate kasutamise kohta. Need andmed võimaldavad hinnata projekti keskkonnamõju, optimeerida projektiotsuseid ja saavutada säästva arengu eesmärgi.
- **7D** atribuudid täiendavad hoone hoolduse haldamiseks vajalikke atribuute. Need on andmed hooldusgraafikute, komponentide elutsükli, tehnilise dokumentatsiooni ja remondiajaloo kohta. See teabekogum tagab mudeli integreeritavuse hooldussüsteemidega (CAFM, AMS), võimaldab hoolduse tõhusat planeerimist, seadmete väljavahetamist ja pakub tuge kogu rajatise elutsükli jooksul.
- **8D** täiendav atribuutide kiht, - sisaldab ohutusega seotud teavet - nii ehitusetapis kui ka hilisema käitamise ajal. Mudel sisaldab töötajate ohutusmeetmeid, hädaolukorra juhiseid, evakuatsioonisüsteeme ja tuleohutuse nõudeid. Nende andmete integreerimine digitaalsesse mudelisse aitab riske eelnevalt arvesse võtta ning töötada välja arhitektuurilised, tehnilised ja organisatsioonilised lahendused, mis võtavad arvesse tervisekaitse- ja ohutuse nõudeid.

Struktureeritud tabeli kujul kujutavad kihid 4D kuni 8D täiendavaid atribuute täidetud väärtustega veergude kujul (joonis 5.3-5), mis on lisatud juba täidetud 3D -mudeli atribuutidele, nagu nimi, kategooria, tüüp ja mahulised omadused. Atribuutide 6D, 7D ja 8D kihtide väärtused sisaldavad täiendavaid tekstilisi ja numbrilisi andmeid, nagu ringlussevõtu protsent, süsiniku jalajälg, garantiaeg, asendustsükkel, paigalduskuupäev, ohutusprotokollid jne.



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 ₂ e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO ₂ e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO ₂ e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO ₂ e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

Joonis 5.3-5 6D -8D lisab atribuutide kihid andmemudelile, mis sisaldab juba 3D -mudeli geomeetrilisi ja ruumilisi atribuute.

Meie uue akna (joonis 4.4-1) puhul võib elemendil, mille identifikaator on W-NEW (joonis 5.3-5), olla järgmised atribuudid 3D -8D:

3D - atribuudid - CAD süsteemidest saadud geomeetriline teave:

- "Type name" - element "Window"
- "Laius" - 120 cm
- Lisaks saate lisada elemendi "Bounding Box " punktid või selle "geomeetria BREP / MESH " eraldi atribuudina.

6D atribuudid - keskkonnasäästlikkus:

- Taaskasutamise määr on 90 protsenti
- "Süsinikdioksiidi jalajälg" - 1622 kg CO₂.

Atribuudid 7D - objektide haldamise andmed:

- "Garantiaaeg" - 8 aastat
- "Asendustsükkel" on 20 aastat vana
- "Hooldus" - nõutav igal aastal

8D atribuudid - ehitiste ohutu kasutamise ja toimimise tagamine:

- installeeritud" aken - firma "XYZ Windows" poolt
- "Ohutusstandard" - vastab ISO 45001 standardile.

Kõik andmebaasis või andmekogumisse (joonis 5.3-5) salvestatud parameetrid on eri osakondade spetsialistidele vajalikud rühmitamiseks, otsinguteks või arvutusteks. Selline projektiobjektide mitmemõõtmeline atribuutidel põhinev kirjeldus annab täieliku pildi nende elutsüklist, tegevusnõuetest ja paljudest muudest projekti projekteerimiseks, ehitamiseks ja käitamiseks vajalikest aspektidest.

CO₂ hindamine ja ehitusprojektide süsinikdioksiidi heitkoguste arvutamine

Ehitusprojektide jätkusuutlikkuse teema kõrval etapis 6D (joonis 5.3-5) keskendub kaasaegne ehitus projektide keskkonnasäästlikkusele, kus üheks oluliseks aspektiks saab projekti elutsükli etappide (nt tootmine ja paigaldamine) käigus tekkivate süsinikdioksiidi CO₂ heitkoguste hindamine ja minimeerimine.

Ehitusmaterjalide süsinikdioksiidi heitkoguste hindamine ja arvutamine on protsess, mille käigus määratakse süsinikdioksiidi koguhulk kindlaks, korrutades projektis kasutatud elemendi või elementide rühma mahuomadused vastava kategooria jaoks sobiva süsinikdioksiidi heitmekoeffitsiendiga.

Süsinikdioksiidi heitkoguste arvestamine ehitusprojektide hindamisel osana laiematest ESG kriteeriumidest (keskkonna-, sotsiaal- ja juhtimiskriteeriumid) lisab analüüsile uue keerukuse taseme. See on eriti oluline kliendi ja investori jaoks asjakohaste sertifikaatide, nagu LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) või DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen), saamisel. Ühe nimetatud sertifikaadi omandamine võib oluliselt suurendada kinnisvara turustatavust, lihtsustada kasutuselevõtmist ja tagada säästvusele orienteeritud üürnike nõuete täitmist (ESG). Sõltuvalt projekti nõuetest võib kasutada ka HQE (Haute Qualité Environnementale, Prantsuse rohelise hoone standard), WELL (WELL Building Standard, mis keskendub kasutajate tervisele ja mugavusele) ja GRESB (Global Real Estate Sustainability Benchmark)

Keskkonna-, sotsiaalne ja valitsemistava **ESG** (keskkonna-, sotsiaalne ja valitsemistava) on lai põhimõtete kogum, mida saab kasutada ettevõtte juhtimise, sotsiaalse ja keskkonnamõju hindamiseks nii ettevõttesiseselt kui ka väliselt.

ESG, mis algelt töötati välja 2000. aastate alguses finantsfondide poolt, et anda investoritele teavet üldiste keskkonna-, sotsiaal- ja juhtimiskriteeriumide kohta, on kujunenud oluliseks näitajaks nii ettevõtete kui ka projektide, sealhulgas ehitusprojektide hindamisel. Suuremate konsultatsioonifirmade uuringute kohaselt on keskkonna-, sotsiaal- ja juhtimisküsimused muutumas ehitussektori lahutamatuks osaks.

EY (2023) "The Path to Carbon Neutrality" kohaselt ei vähenda ESG -põhimõtteid aktiivselt rakendavad ettevõtted mitte ainult pikaajalisi riske, vaid suurendavad ka oma ärimudelite tõhusust, mis on eriti oluline turgude globaalsel ümberkujundamisel [103]. PwC aruandes "ESG teadlikkus" märgitakse, et ettevõtete teadlikkus ESG -tegurite olulisusest ulatub 67% ja 97% vahel, kusjuures enamik organisatsioone peab neid suundumusi tulevase jätkusuutlikkuse seisukohalt võtmetähtsusega [104] ja et enamasti näevad ettevõtted sidusrühmade poolt märkimisväärset survet ESG põhimõtete integreerimiseks.

Seega ei aita ESG -põhimõtete integreerimine ehitusprojektidesse mitte ainult kaasa rahvusvaheliste jätkusuutlikkuse sertifikaatide, nagu LEED, BREEAM, DGNB, saamisele, vaid tagab ka tööstuse ettevõtete pikaajalise jätkusuutlikkuse ja konkurentsivõime.

Üks olulisemaid tegureid, mis mõjutab ehitusprojekti üldist süsinikujalajälge, on ehitusmaterjalide ja -komponentide tootmis- ja logistikaetapid. Kohapeal kasutatavad materjalid mõjutavad sageli otsustavalt kogu CO₂-heidet, eriti projekti elutsükli varajases etapis - alates tooraine kaevandamisest kuni ehitusplatsile tarnimiseni.

Heitkoguste arvutamine ehitusdetailide kategooriate või liikide kaupa nõuab süsinikdioksiidi heitmekoeffitsientide kasutamist, mis kajastavad erinevate materjalide tootmisel tekkiva CO₂ kogust. Selliste materjalide hulka kuuluvad betoon, tellised, ringlussevõetud teras, alumiinium ja muud. Need väärtused võetakse tavaliselt usaldusväärsetest allikatest ja rahvusvahelistest andmebaasidest, näiteks Ühendkuningriigi ICE 2015 (Inventory of Carbon and Energy) ja USA EPA 2006 (U.S. Environmental Protection Agency) [105]. Järgnevas tabelis (joonis 5.3-6) on esitatud kokkuvõtte mitmete levinud ehitusmaterjalide lähteheitmekoeffitsientidest. Iga materjali kohta on esitatud kaks põhiparameetrit: CO₂ eriheide (kilogrammides materjali kilogrammi kohta) ja mahu ja kaalu ümberarvestustegurid (kilogrammides kuupmeetri kohta), mis on vajalikud arvutuste integreerimiseks projekteerimismudelisse ja ühendamiseks QTO andmete rühmitusega.



Carbon Emitted in Production

Material	Abbreviated	UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m ³ to kg
		Process Emissions (kg CO ₂ e/ kg of product) (K1)	Process Emissions (kg CO ₂ e/ kg of product) (K2)	kg / m ³ (K3)
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

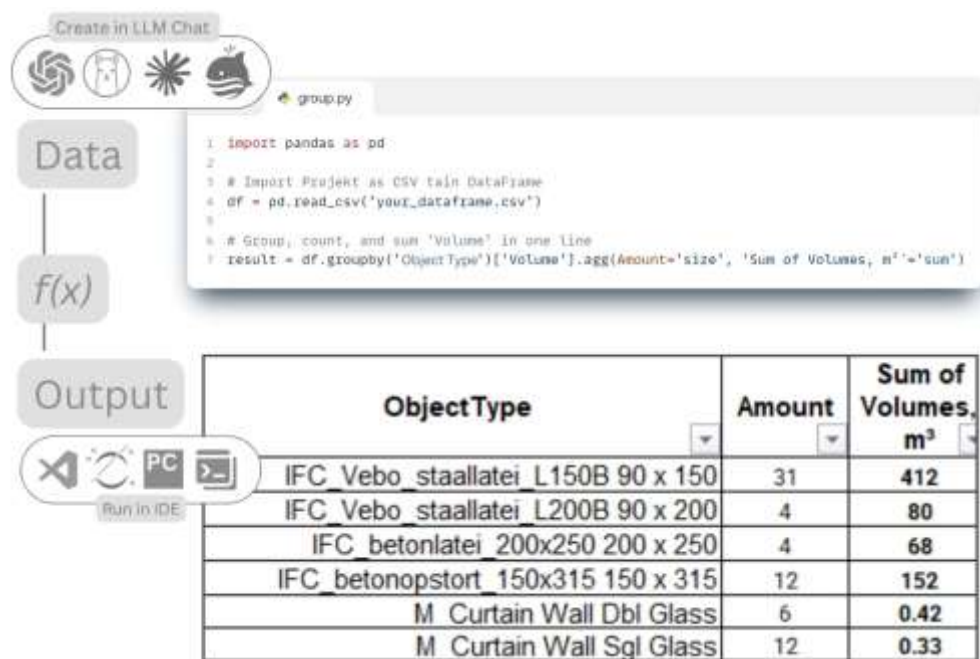
Joonis 5.3-6 Erinevate ehitusmaterjalide tootmisel eraldunud süsiniku hulk vastavalt Ühendkuningriigi ICE ja USA EPA andmebaasile.

Projekti CO₂ koguheitmete arvutamiseks, nagu ka 4D ja 5D arvutuste puhul, on vaja määrata iga rajatiste rühma atribuutide mahud. Seda saab teha kvantitatiivse analüüsi vahendite (QTO) abil, saades omaduste mahud kuupmeetrites, nagu on üksikasjalikult käsitletud jaotises Koguse võtmine. Seejärel korrutatakse need mahud iga materjalirühma atribuudi "CO₂-protsessi heitkogused" asjakohaste koeffitsientidega.

- 🔊 Ekstraheerime automaatselt CAD-projektist (BIM) mahtude tabeli elementide tüübi järgi, grupeerides kõik projekti andmed, nagu juba eelmistes peatükkides tehtud. Selle ülesande täitmiseks kasutame LLMI.

Palun rühmitage DataFrame tabel projektist CAD (BIM) veeru parameetri "Objekti nimi" (või "tüüp") järgi ja näidake igas rühmas olevate objektide arvu ning tehke kokkuvõtte parameetri "maht" kohta kõigi tüübi objektide kohta. ↵

LLMi vastus :

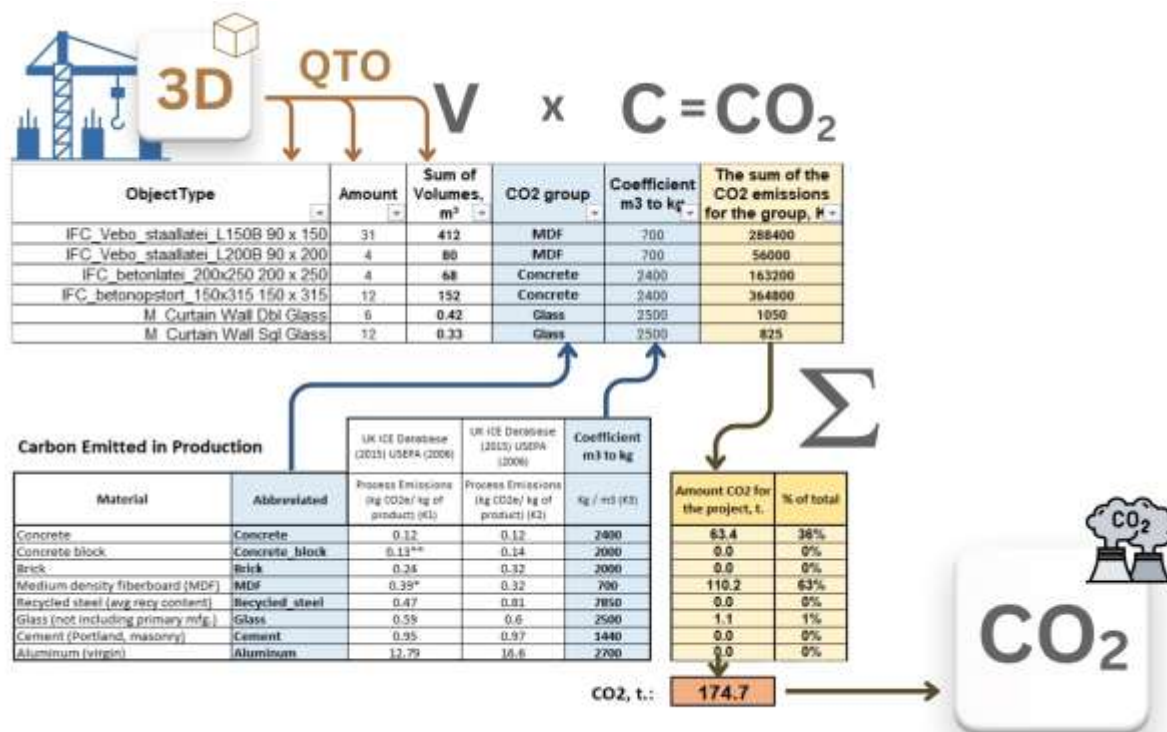


Joonis 5.3-7 LLM-i genereeritud kood on rühmitanud projektiüksused meile tüübi (ObjectType) järgi, millel on kokkuvõtlik atribuut "Volume".

Kogu projekti CO₂ koguheitmete arvutamise automatiseerimiseks tuleb lihtsalt seadistada tabelis automaatne andmete seostamine või seostada käsitsi elementide tüübid (joonis 5.3-7) vastavate materjalitüüpidega (joonis 5.3-6) heitekoefitsientide tabelist. Valmis tabel koos heitekoefitsientide ja valemitega, samuti kood, millega saab CAD-formaatidest (BIM) mahte kätte ja CO₂ määramist automatiseerida, on leitav GitHubis, otsides "CO₂-calculating-the-embodied-carbon. DataDrivenConstruction." [106].

Seega võimaldab andmete integreerimine pärast QTO elementide rühmitamist CAD andmebaasist süsinikdioksiidi heitkoguste automaatset arvutamist (joonis 5.3-8) erinevate konstruktsioonivariantide puhul. See võimaldab analüüsida erinevate materjalide mõju erinevates variantides ja valida ainult need lahendused, mis vastavad CO₂-emissiooni nõuetele, et saada hoone kasutuselevõtmisel konkreetne sertifikaat

CO₂-heite hindamine, korrutades tegureid rühmitatud projektielementide mahtudega, on tüüpiline näide ehitusettevõtte ülesande kohta, mis on seotud ESG-reitingu saamisega (nt LEED-sertifitseerimine) rajatisele.

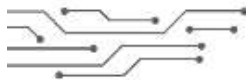


Joonis 5.3-8 QTO rühmade integreerimine CAD-andmebaasidest tagab täpsuse ja automatiseerituse CO₂ lõplike heitkoguste hinnangute tuletamisel.

Samamoodi saame elementide rühmade mahtude määratlemisel teha arvutusi materjalide kontrolli ja logistika, kvaliteedi jälgimise ja juhtimise, energia modelleerimise ja analüüsi ning paljude muude ülesannete jaoks, et saada uus atribuutide staatus (parameeter tabelis) nii üksikute elementide rühmade kui ka kogu projekti jaoks.

Kui selliste arvutusprotsesside arv ettevõttes hakkab kasvama, tekib küsimus, kas selliseid arvutusi on vaja automatiseerida ja arvutuste tulemusi ettevõtte protsessidesse ja andmehaldussüsteemidesse rakendada.

Täieliku lahenduse keerukuse tõttu tellivad ehitussektori keskmised ja suured ettevõtted sellise automatiseerimise ERP (või PMIS) süsteemi arendusettevõtelt. Arendusettevõtte loovad suurtele klientidele ühe tervikliku moodulsüsteemi, et hallata paljusid erinevaid teabekihte, sealhulgas materjali- ja ressursikalkulatsioone.



PEATÜKK 5.4.

EHITUS ERP JA PMIS SÜSTEEMID

Ehitus ERP -süsteemid arvutuste ja kalkulatsioonide näitel

Modulaarsed ERP-süsteemid integreerivad erinevad atribuutide (teabe) kihid ja andmevood ühtsesse terviklikku süsteemi, mis võimaldab projektijuhtidel hallata ressursse, finantse, logistikat ja muid projekti aspekte sünkroonselt ühe platvormi raames. Ehituse ERP-süsteem toimib ehitusprojektide "ajuna", lihtsustades automatiseerimise abil korduvaid protsesse, tagades läbipaistvuse ja kontrolli kogu ehitusprotsessi jooksul.

Ehituse ERP - süsteemid (Enterprise Resource Planning) on terviklikud tarkvaralahendused, mis on mõeldud ehitusprotsessi erinevate aspektide haldamiseks ja optimeerimiseks. Ehituse ERP-süsteemide keskmeks on kulukalkulatsiooni ja ajakava haldamise moodulid, mis teeb neist olulise vahendi ressursside tõhusaks planeerimiseks.

ERP-moodulid - süsteemid võimaldavad kasutajatel sisestada, töödelda ja analüüsida struktureeritud viisil andmeid, mis hõlmavad projekti erinevaid aspekte, mis võivad hõlmata materjali- ja tööjõukulude arvestust, seadmete kasutamist, logistika juhtimist, inimressursse, kontakte ja muid ehitustegevusi.

Üks süsteemi funktsionaalsetest plokkidest on äriloogika automatiseerimise moodul - BlackBox/WhiteBox, mis täidab protsessi juhtimiskeskuse rolli.

BlackBox /WhiteBox võimaldab spetsialistidel, kes kasutavad ERP -süsteemi, hallata paindlikult juurdepääsuõiguste kaudu ettevõtte erinevaid aspekte, mis on teiste kasutajate või administraatorite poolt juba eelkonfigureeritud. ERP-süsteemide kontekstis viitavad terminid *BlackBox* ja *WhiteBox* süsteemi sisemise loogika läbipaistvuse ja kontrollitavuse tasemele:

- **BlackBox** ("must kast") - kasutaja suhtleb süsteemiga liidese kaudu, ilma juurdepääsuta protsessi täitmise sisemisele loogikale. Süsteem teeb arvutusi ise, lähtudes eelnevalt määratletud reeglitest, mis on lõppkasutaja eest varjatud. Ta sisestab andmed ja saab tulemuse, teadmata, milliseid atribuute või koefitsiente sees kasutati.
- **WhiteBox** ("valge kast") - protsessiloogika on vaatamiseks, kohandamiseks ja muutmiseks kättesaadav. Edasijõudnud kasutajad, administraatorid või integraatorid saavad käsitsi määratleda andmetöötlusalgoritme, arvutusreegleid ja projektiüksuste vahelisi suhtlusstsenaariume.

Näiteks võib kogenud kasutaja või administraator kehtestada reegli: millised hinnangu atribuudid tuleb omavahel korrutada või grupeerida teatud tunnuse järgi ja kuhu lõpptulemus tuleb salvestada. Hiljem laadivad vähem koolitatud spetsialistid, näiteks hindamisinsenerid, lihtsalt uusi andmeid ERPi kasutajaliidese kaudu - ja saavad valmis kalkulatsioonid, ajakavad või spetsifikatsioonid, ilma et nad peaksid koodi kirjutama või loogika tehnilistest üksikasjadest aru saama.

Eelmistes peatükkides käsitleti arvutus- ja loogikamooduleid LLM-i koostoimimise kontekstis. ERP-keskkonnas toimuvad sarnased arvutused ja teisendused moodulites, mis on peidetud nuppude ja vormide kasutajaliidese taha.

Järgnevas näites (joonis 5.4-1) on ERP -süsteemi administraator BlackBox /WhiteBox moodulis määratlenud reeglid, mille abil sobitatakse hinnangute üksuste atribuute QTO grupeerimise atribuutidega. Tänu sellele (juhi või administraatori poolt konfigureeritud BlackBox/WhiteBox moodulile saab kasutaja (kalkulaator või insener), lisades käsitsi koguse või mahu atribuudi ERP kasutajaliidese kaudu, automaatselt valmis hinnangud ja tööplaanid. Sel viisil muutuvad eelmistes peatükkides käsitletud arvutus- ja hinnangute koostamise protsessid ERP-siseselt koodi abil poolautomaatseks konveieriks.

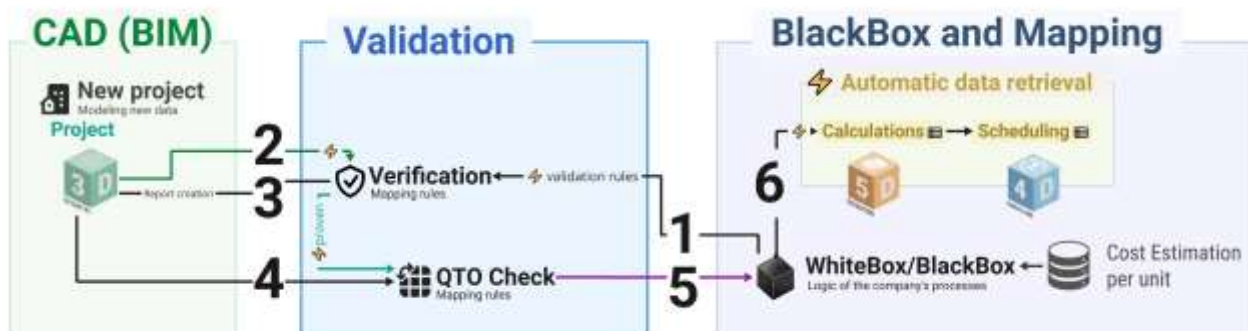
Selle poolautomaatse protsessi ühendamise CAD (BIM) mudelite (joonis 4.1-13) ruumiliste atribuutidega, näiteks CAD-projekti laadimisega ERP-moodulisse, mis on selleks eelnevalt konfigureeritud, muudab andmevoo sünkroniseeritud mehhanismiks, mis suudab iseseisvalt ja koheselt ajakohastada üksikute elementide rühmade või kogu projekti väärtust vastavalt projekteerimisfaasis toimunud muudatustele, kui CAD-mudelit ERP-sse laaditakse.

Selleks, et luua automatiseeritud andmevoog (joonis 5.4-2) CAD (BIM) ja ERP süsteemide vahel, tuleb CAD (BIM) mudelite andmebaaside andmete põhiprotsessid ja nõuded struktureeritud viisil määratleda, nagu me juba eespool peatükis "Nõuded ja andmete kvaliteedi tagamine " käsitlesime. See protsess ERP-s jaguneb sarnasteks etappideks:

- **Valideerimisreeglite loomine (1)**, millel on oluline roll ERP -süsteemi sisestatud andmete õigsuse tagamisel. Valideerimisreeglid toimivad filtritena, mis valideerivad üksusi ja nende atribuute, lubades süsteemi siseneda ainult nendel elementidel, mis vastavad nõuetele. Kontrollimisest ja valideerimisest saate rohkem teada peatükis "Nõuete loomine ja andmete kvaliteedi valideerimine ".
- Seejärel toimub ERP-süsteemis **kontrolliprotsess (2)**, mis kinnitab, et kõik projektiüksuse elemendid koos nende atribuutide ja väärtustega on õigesti loodud ja on valmis järgmisteks töötlemisetappideks.
- Kui esineb probleeme mittetäielike atribuutide andmetega, **koostatakse aruanne (3)** ja projekt koos parandusjuhistega saadetakse läbivaatamiseks, kuni see on valmis järgmiseks iteratsiooniks.
- Kui projekti andmed on valideeritud ja kontrollitud, kasutatakse neid teises ERP-moodulis **(4), et luua Quantity Take-Off tabelid (QTO)**, mis loovad koguste atribuudid üksuste gruppidele, materjalidele ja ressurssidele vastavalt eelnevalt loodud reeglitele (WhiteBox/BlackBox).
- Sobitusreeglite või QTO järgi rühmitatud andmed **integreeritakse** automaatselt **arvutustega**

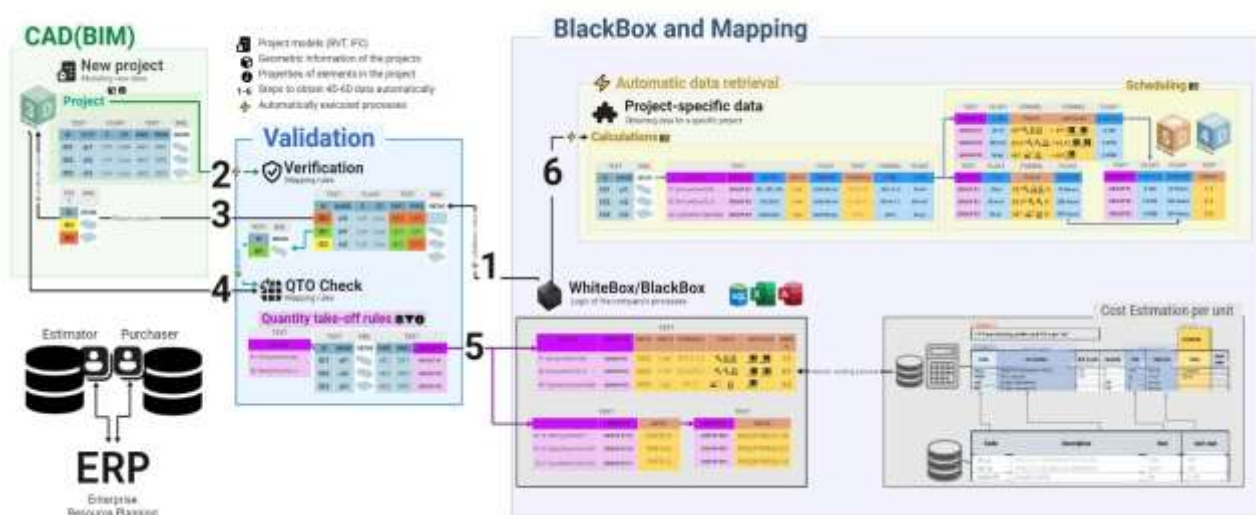
(nt kulud ja aeg) (5).

- ERP -süsteemi viimases etapis korrutab kasutaja, korrutades QTO tabelist pärinevad reguleerimisala atribuudid protsessitabelite atribuutidega (nt hinnangulised objektid), **automaatselt arvutustulemused** (6) (nt kulud hinnangud, tööplaanid või CO₂ heitkogused) iga üksuse rühma ja projekti kui terviku kohta.

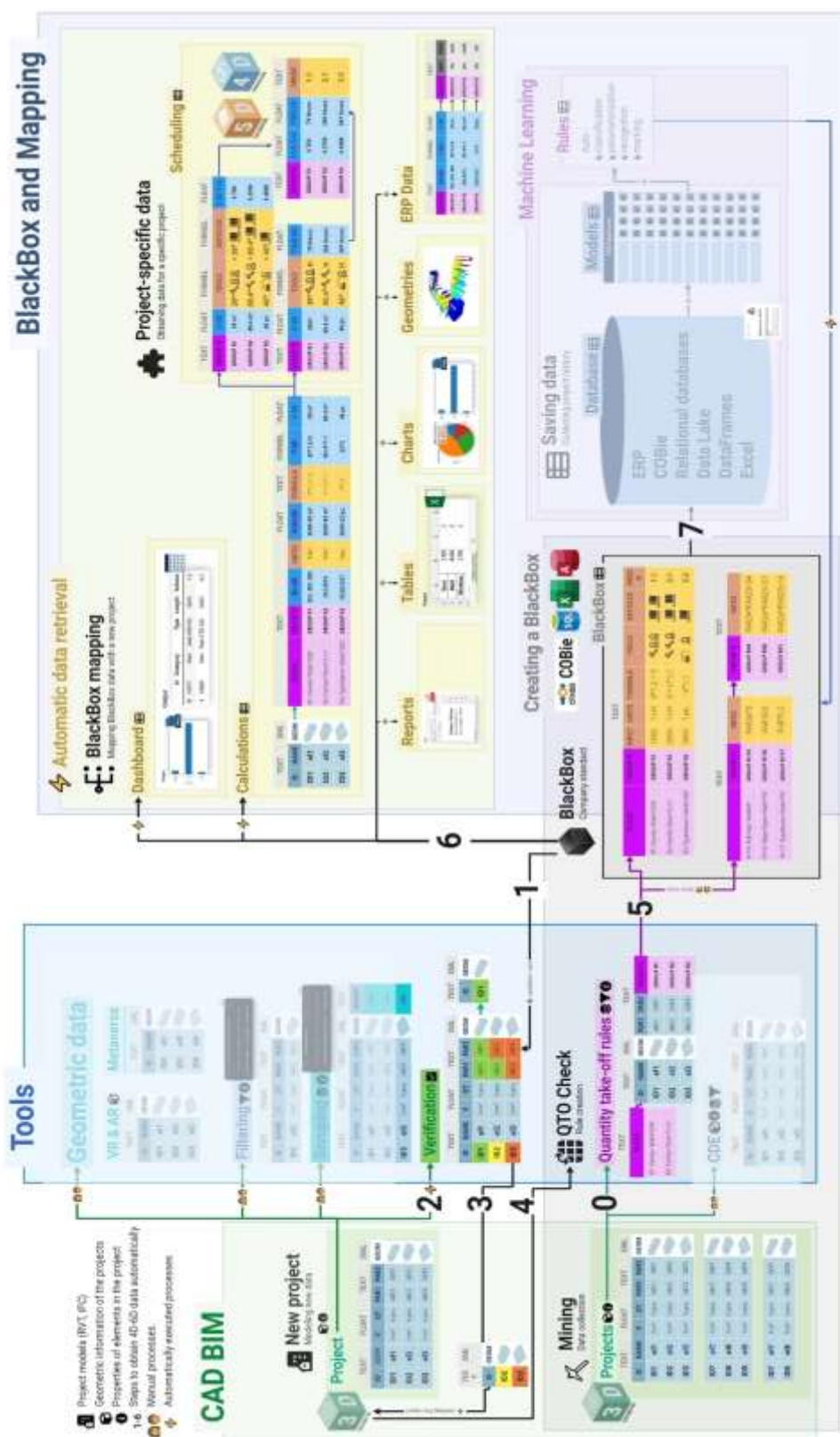


Joonis 5.4-2 Ehituse ERP -süsteemi arhitektuur koos CAD (BIM), alates valideerimisreeglite loomisest (1) kuni kulude ja tööplaanide automaatse arvutamiseni (5-6).

Modulaarses ERP -süsteemis on protsessid integreeritud, kasutades tarkvara, mis sisaldab kasutajaliidest. Kasutajaliidese taga on back-end, kus struktureeritud tabelid töötlevad andmeid, sooritades erinevaid operatsioone, mis on juhi või administraatori poolt eelnevalt konfigureeritud. Selle tulemusena saab kasutaja tänu eelnevalt määratletud ja kohandatud automatiseerimisloogikale (moodulites BlackBox /WhiteBox) poolautomaatselt ettevalmistatud dokumendid, mis vastavad tema ülesannetele



Joonis 5.4-3 ERP - süsteem aitab juhtidel ja kasutajatel uute andmete genereerimiseks erialaste tabelite vahel liikuda.



Joonis 5.4-4 ERP -süsteem on integreeritud analüütiliste vahenditega ja automatiseerib ettevõtte otsustusprotsessi.

Samamoodi on ERP -süsteemide protsessid, alates algusest kuni lõpliku arvutamiseni (sammud 1-6, joonis 5.4-3), omavahel seotud sammude ahel, mis lõppkokkuvõttes tagab planeerimise läbipaistvuse, tõhususe ja täpsuse.

Kaasaegsed ehitus-ERP -süsteemid sisaldavad lisaks kulude ja ajakava arvutamise moodulitele ka kümneid teisi eelkonfigureeritud mooduleid, mis tavaliselt hõlmavad dokumendihaldust, projekti edenemise jälgimist, lepingute haldamist, tarneahelat ja logistikat, samuti integratsiooni teiste ärisüsteemide ja platvormidega. ERP integreeritud analüüsivahendid võimaldavad kasutajatel automatiseerida armatuurtahvlite loomist, et jälgida projekti tulemuslikkuse põhinäitajaid (KPI - key performance indicators). See võimaldab ehitusprojekti kõigi aspektide tsentraliseeritud ja järjepidevat haldamist, kusjuures püütakse ühendada suur hulk rakendusi ja süsteeme ühele platvormile.

Tulevikus kasutatakse ERP -analüütikat koos masinõppega, et suurendada täpsust ja optimeerida tulevaste projekti omaduste arvutamise protsessi. Suurandmete (joonis 5.4-4) abil analüüsitud ja ERP-süsteemidest kogutud andmed ja tunnused on tulevikus aluseks prognoosimudelite loomisele, mis võimaldavad täpselt ette näha võimalikke viivitusi, riske või näiteks võimalikke muutusi materjalikuludes.

Alternatiivina ERP-le kasutatakse ehitussektoris sageli PMIS (Project Management Information System), mis on projektijuhtimissüsteem, mis on mõeldud ülesannete üksikasjalikuks kontrollimiseks üksikute ehitusprojektide tasandil.

PMIS: vahendaja ERP ja ehitusplatsi vahel

Erinevalt ERP, mis hõlmab kogu ettevõtte äriprotsesside ahelat, keskendub PMIS konkreetse projekti haldamisele, tähtaegade, eelarvete, ressursside ja dokumentatsiooni jälgimisele.

PMIS (projektijuhtimise infosüsteem) on ehitusprojektide haldamise tarkvara, mis on loodud projekti kõigi aspektide planeerimiseks, jälgimiseks, analüüsimiseks ja aruandluseks.

PMIS võimaldab teil hallata dokumente, ajakavasid, eelarveid ja esmapilgul võib PMIS tunduda ERP dubleeriva lahendusena, kuid peamine erinevus seisneb juhtimise tasemes:

- **ERP** keskendub ettevõtte kui terviku äriprotsessidele: kulude, lepingute, hangete, inimressursside ja ressursside juhtimine ettevõtte tasandil.
- **PMIS** keskendub individuaalsele projektijuhtimisele, pakkudes üksikasjalikku planeerimist, muudatuste kontrolli, aruandlust ja osalejate koordineerimist.

Paljudel juhtudel on just ERP - süsteemid juba piisava funktsionaalsusega ning PMIS rakendamine muutub pigem mugavuse ja ettevõtte eelistuste küsimuseks. Paljud töövõtjad ja kliendid kasutavad PMISi mitte sellepärast, et see on vajalik, vaid sellepärast, et müüja või suur klient, kes soovib andmeid koondada konkreetsele platvormile, on seda peale surunud.

Tuleks mainida, et ehitusprojektide juhtimise rahvusvahelises terminoloogias on olemas ka teised eraldi populaarsed mõisted, nagu PLM (Product Lifecycle Management) ja EPC ja EPC-M (Engineering,

Procurement and Construction Management) - ehitussektori lepingute sõlmimise meetodid.

Kui ettevõtte kasutab juba ERP koos projektijuhtimise moodulitega, võib PMIS kasutuselevõtt olla tarbetu ühenduslüli, mis dubleerib funktsionaalsust. Kui aga protsessid ei ole automatiseeritud ja andmed on killustatud, võib PMIS olla mugavam ja lihtsamini hooldatav vahend.

Spekuleerimine, kasum, isoleeritus ja vähene läbipaistvus ERP ja PMISi puhul

Vaatamata liideste ja menetluste välisele lihtsusele, on ehitus-ERP ja PMIS - süsteemid enamasti suletud ja paindumatud lahendused. Sellised süsteemid tarnitakse tavaliselt ühe müüja eelkonfigureeritud tarkvarapaketina, millel on piiratud juurdepääs sisemistele andmebaasidele ja protsessiloomikale.

CAD-(BIM-) tarnijad võtavad üha enam üle selliste süsteemide arendamise ja kontrolli, kuna nende andmebaasid sisaldavad ERP-süsteemidele vajalikku teavet: projektielementide kvantitatiivseid ja mahulisi omadusi. Kuid selle asemel, et pakkuda juurdepääsu nendele andmetele avatud või masinloetaval kujul, pakuvad müüjad ainult piiratud kasutusstsenaariume ja suletud töötlemisloogikat - mis on eelnevalt määratletud BlackBoxi moodulites. See vähendab süsteemi paindlikkust ja takistab selle kohandamist konkreetsetele projektitingimustele.

Andmete piiratud läbipaistvus on endiselt üks peamisi probleeme digitaalsete protsesside puhul ehituses. Suletud andmebaasiarhitektuur, juurdepääsu puudumine ehituselementide täielike atribuutide kogumitele, keskendumine *BlackBoxi* automatiseerimismoodulitele ja avatud liideste puudumine suurendavad oluliselt dokumendibürokraatia riske. Sellised piirangud tekitavad otsustusprotsessis kitsaskohti, raskendavad teabe kontrollimist ja avavad ukse andmete varjamisele või spekuleerimisele ERP/PMIS-süsteemides. Kasutajad saavad tavaliselt ainult piiratud juurdepääsu - olgu see siis vähendatud liides või osaline API -, ilma et neil oleks võimalik otse suhelda esmaste andmeallikatega. See on eriti kriitiline, kui tegemist on CAD-projektidest automaatselt genereeritud parameetritega, nagu mahud, pindalad ja kogused, mida kasutatakse QTO arvutustes.

Selle tulemusena keskenduvad paljud ehitusettevõtted selle asemel, et otsida tõhusust protsesside automatiseerimise, avatud andmete, tehingukulude vähendamise ja uute ärimudelite loomise kaudu, väliste parameetrite haldamisele - projekti kulusid mõjutavate tegurite, korrigeerimistegurite ja arvutusmeetodite manipuleerimisele suletud ERP/PMIS platvormides. See loob ruumi spekuleerimiseks, moonutab tegelikke tootmiskulusid ja vähendab usaldust kõigi ehitusprotsessis osalejate vahel

Ehituses moodustub kasum valminud projektist saadava tulu ja muutuvkulude, mis hõlmavad projekteerimist, materjale, tööjõudu ja muid projekti elluviimisega otseselt seotud otseseid kulusid, vahena. Nende kulude väärtust mõjutav võtmetegur ei ole aga mitte ainult tehnoloogia või logistika, vaid ka arvutuste kiirus ja täpsus ning juhtimisotsuste kvaliteet ettevõttes.

Probleemi süvendab asjaolu, et enamikus ehitusettevõtetes jäävad kuluarvestusprotsessid läbipaistmatuks mitte ainult klientide, vaid ka töötajate endi jaoks, kes ei kuulu kalkultatsiooni- või finantsosakonda. Selline suletus soodustab ettevõttes privilegeeritud spetsialistide rühma - "finantsteadmiste" kandjate - moodustumist, kellel on ainuõigus muuta atribuute ja parandustegureid ERP/PMIS -süsteemides. Need töötajad võivad koos ettevõtete juhtidega tegelikult kontrollida projekti finantsloogikat.

Sellistes tingimustes muutuvad kalkulaatorid "finantsjongleerijateks", kes peavad tasakaalustama ettevõtte kasumi maksimeerimise ja kliendi jaoks konkurentsivõimelise hinna säilitamise vahel. Samal ajal peavad nad vältima jultunud ja jõhkraid manipulatsioone, et mitte kahjustada ettevõtte mainet. Selles etapis kehtestatakse koefitsiendid, et varjata ülehinnatud mahte või materjalide ja tööde kulusid.

Selle tulemusena ei ole ehitussektoris tegutsevate ettevõtete peamine skeem tõhususe ja kasumlikkuse suurendamiseks mitte otsustusprotsesside automatiseerimine ja kiirendamine, vaid materjalide ja tööde hindadega spekulatsioon (joonis 5.4-5). Tööde ja materjalide maksumuse ülehindamine toimub suletud ERP /PMIS - süsteemides "halli" arvestuse abil, paisutades protsendimäärad üle materjalide või tööde mahu keskmiste turuhindade koefitsientide abil (joonis 5.1-6), mida käsitleti peatükis "Arvutuste koostamine ja tööde maksumuse arvutamine ressursibaasi alusel".

Selle tulemusena saab klient arvutuse, mis ei kajasta tegelikku maksumust ega tööde mahtu, vaid on paljude varjatud sisemiste koefitsientide tulemus. Samal ajal on alltöövõtjad, püüdes vastata peatöövõtja poolt kehtestatud alahinnatud määradele, sageli sunnitud ostma odavamaid ja madalama kvaliteediga materjale, mis halvendab ehitamise lõppkvaliteeti.

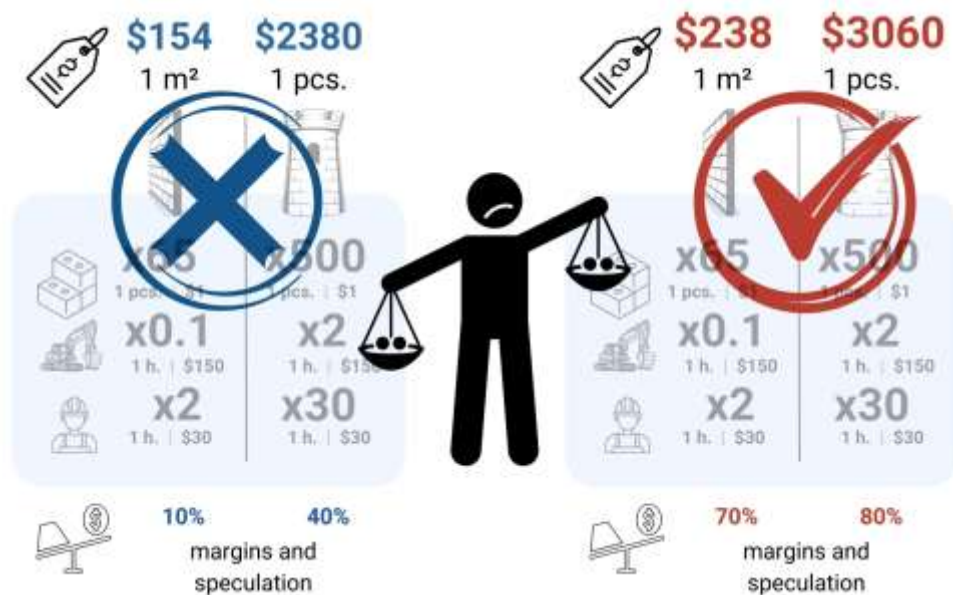
Spekulatiivne protsess, mille käigus otsitakse kasumit õhust välja, kahjustab lõpuks nii kliente, kes saavad ebausaldusväärseid andmeid, kui ka juhte, kes on sunnitud leidma üha rohkem spekulatsioonimudeleid

Sellest tulenevalt on bürokraatia tase andmete ja protsesside haldamisel seda suurem, mida suurem on projekt. Iga samm ja iga moodul kätkeb sageli läbipaistmatuid koefitsiente ja lisatasusid, mis on varjatud arvutusalgorithmides ja sisemistes menetlustes. See mitte ainult ei tee auditeerimist keeruliseks, vaid moonutab oluliselt ka projekti finantspilti. Suurtes ehitusprojektides viib selline praktika sageli lõppkulude mitmekordse (mõnikord kuni kümnekordse) suurenemiseni, samas kui

tegelikud mahud ja kulud jäävad tellija tegelikust kontrollist välja (joonis 2.1-3 Saksamaa suurte infrastruktuuriprojektide kavandatud ja tegelike kulude võrdlus).

McKinsey & Company aruande "Imagining the Digital Future of Construction" (2016) kohaselt viiakse suured ehitusprojektid lõpule keskmiselt 20% hiljem kui planeeritud ja kuni 80% üle eelarve [107].

Hinnangu koostamise ja eelarvestamise osakonnad on muutumas ettevõtte kõige enam valvatud lüliks. Juurdepääs neile on rangelt piiratud isegi sisespetsialistide jaoks ning suletud loogika ja andmebaasi struktuuride tõttu on võimatu objektiivselt ja moonutusteta hinnata projektiotsuste tõhusust. Läbipaistvuse puudumine viib selleni, et ettevõtted on sunnitud mitte optimeerima protsesse, vaid võitlema ellujäämise eest arvude ja tegurite "loomingulise" haldamisega (joonis 5.3-1, joonis 5.1-6 - näiteks parameeter "Bid. Factor").



Joonis 5.4-5 Arvelduste tasandi spekulatsioonimäärad on ettevõtete peamine kasum ja žongleerimise kunst töö kvaliteedi ja maine vahel.

Kõik see seab kahtluse alla suletud ERP/PMIS - süsteemide jätkuva kasutamise ehituses. Digitaalse ümberkujundamise ja klientide kasvavate läbipaistvusnõuete kontekstis (joonis 10.2-3) on ebatõenäoline, et projektide elluviimine jääb pikemas perspektiivis sõltuvusse suletud lahendustest, mis piiravad paindlikkust, takistavad integratsiooni ja pärsivad äri arengut.

Ja ükskõik kui soodne on ehitusettevõtetele töötada suletud andmebaasides asuvate andmesilode ja läbipaistmatute andmetega - paratamatult hõlmab ehitussektori tulevik üleminekut avatud platvormidele, masinloetavatele ja läbipaistvatele andmestruktuuridele ning usaldusel põhinevale automatiseerimisele. Seda muutust juhitakse ülevalt poolt - klientide, reguleerivate asutuste ja ühiskonna surve, kes nõuavad üha enam aruandekohustust, jätkusuutlikkust, läbipaistvust ja majanduslikku otstarbekust.

Suletud ERP-ajastu lõpp /PMIS: ehitustööstus vajab uusi lähenemisviise

Mahukad modulaarsed ERP/PMIS - süsteemid, mis koosnevad kümnetest miljonitest koodiridadest, muudavad igasugused muudatused neis äärmiselt keeruliseks. Sellisel juhul kujuneb üleminek uuele platvormile ettevõtte jaoks juba eelkonfigureeritud moodulite, kümnete tuhandete artiklite ressursiandmebaaside (joonis 5.1-3) ja tuhandete valmis arvutuste (joonis 5.1-6) olemasolul kulukaks ja pikaajaliseks protsessiks. Mida rohkem koodi ja pärandarhitektuuri - seda suurem on sisemise ebaefektiivsuse tase ja iga uus projekt teeb asja ainult hullemaks. Paljudes ettevõtetes muutuvad andmemigratsioon ja uute lahenduste integreerimine mitmeaastasteks eeposteks, millega kaasnevad pidevad ümbertöötamised ja lõputud kompromisside otsingud. Tulemuseks on sageli tagasipöördumine vanade, tuttavate platvormide juurde, hoolimata nende piirangutest.

Nagu rõhutati Saksamaa aruandes "Must raamat" [108], mis käsitleb süsteemseid puudusi ehitusandmete haldamisel, on teabe killustatus ja tsentraliseeritud lähenemisviisi puudumine selle haldamisel üks peamisi ebatõhususe põhjusi. Ilma standardiseerimise ja integreerimiseta kaotavad andmed oma väärtuse, muutudes pigem arhiiviks kui juhtimisvahendiks.

Andmekvaliteedi kadumise peamine põhjus on ehitusprojektide ebapiisav planeerimine ja kontroll, mis toob sageli kaasa märkimisväärse kulude kasvu. Mustas raamatus "Focus: The Cost Explosion" analüüsitakse neid soovimatuid tagajärgi põhjustavaid põhitegureid. Nende hulka kuuluvad ebapiisavad vajaduste analüüsid, teostatavusuuringute puudumine ja kooskõlastamata planeerimine, mis põhjustavad lisakulusid, mida oleks võinud vältida.

Ettevõtte küpses IT-ökosüsteemis on vananenud süsteemi väljavahetamine võrreldav juba ehitatud hoone kandva samba väljavahetamisega. Ei piisa lihtsalt vana eemaldamisest ja uue paigaldamisest - oluline on teha seda nii, et hoone jääks stabiilseks, laed ei variseks kokku ja kõik kommunikatsioonid töötaksid jätkuvalt. Siin peituvadki raskused: iga viga võib põhjustada tõsiseid tagajärgi kogu ettevõtte süsteemile.

Sellel põhjal kasutavad ehitussektori suurte ERP-toodete arendajad jätkuvalt kirjutatud koodi hulka oma platvormi kasuks rääkiva argumendina. Eriala konverentsidel võib ikka veel kuulda selliseid fraase nagu: "Sellise süsteemi uuesti loomiseks kuluks 150 inimaastat", hoolimata sellest, et suurem osa selliste süsteemide funktsionaalsusest peidab endas andmebaase ja üsna lihtsaid funktsioone tabelitega töötamiseks, mis on pakendatud spetsiaalsesse fikseeritud kasutajaliidesesse. Praktikas muutub "150 inimaasta" koodimaht pigem koormaks kui konkurentsieeliseks. Mida rohkem koodi - seda suuremad on toetuskulud, seda raskem on kohaneda uute tingimustega ja seda kõrgem on uute arendajate ja klientide sisenemislävi.

Paljud moodulhoonete süsteemid meenutavad tänapäeval kohmakaid ja aegunud "Frankensteini konstruktsioone", kus iga hooletu muudatus võib põhjustada tõrkeid. Iga uus moodul muudab niigi ülekoormatud süsteemi veelgi keerulisemaks, muutes selle labürindiks, millest saavad aru vaid vähesed spetsialistid, mis muudab selle hooldamise ja ajakohastamise veelgi keerulisemaks.

Keerukust mõistavad ka arendajad ise, kes teevad perioodiliselt pausi refaktoormiseks - arhitektuuri läbivaatamiseks, et võtta arvesse uute tehnoloogiate tekkimist. Kuid isegi kui refaktooringut tehakse regulaarselt, kasvab keerukus paratamatult. Selliste süsteemide arhitektid harjuvad kasvava keerukusega, kuid uute kasutajate ja spetsialistide jaoks muutub see ületamatuks takistuseks. Selle tulemusel koondub kogu oskusteave mõne arendaja kätte ja süsteem ei ole enam skaleeritav. Lühiajaliselt on sellised eksperdid kasulikud, kuid pikemas perspektiivis muutuvad nad osaks probleemist.

Organisatsioonid hakkavad jätkuvalt integreerima "väikseid" andmeid oma suurte andmete vastastega ning on rumal arvata, et üks rakendus - olgu see kui kallid või töökindel tahes - suudab kõike töödelda [109].

- Phil Simon, "Conversations About Collaboration" podcasti saatejuht.

Tekib õigustatud küsimus: kas meil on tõesti vaja nii kohmakaid ja suletud süsteeme tööde maksumuse ja ajastuse arvutamiseks tabelite kujul, kui teised tööstusharud on juba ammu suutnud sarnaste ülesannetega toime tulla, kasutades avatud andmete ja läbipaistva loogikaga analüütilisi vahendeid?

Praegu on ehituses endiselt nõudlus suletud moodulplatvormid, eelkõige kuluarvestuse eripära tõttu (joonis 5.1-7). Selliseid süsteeme kasutatakse sageli hallide või läbipaistmatute skeemide juhtimiseks, mis võimaldab tegelikke kulusid tellija eest varjata. Kuna aga tööstus areneb digitaalseks, eelkõige kliendid, ja liigub nn Uberiseeritud ajastusse, kaotavad vahendajad, nimelt ehitusettevõtted oma ERP-idega, oma tähtsust aja- ja kuluarvestuses. See muudab ehitussektori nägu igaveseks. Loe lähemalt raamatu viimasest osast ja peatükist "Ehitus 5.0: kuidas teenida raha, kui ei saa enam varjata".

Viimase 30 aasta jooksul kogunenud tuhanded pärandlahendused, mille arendamisse on investeeritud tuhandeid inimaastaid, hakkavad kiiresti kaduma. Üleminek avatud, läbipaistva ja paindliku andmehalduse suunas on vältimatu. Küsimus on vaid selles, millised ettevõtted suudavad nende muutustega kohaneda ja millised jäävad vana mudeli pantvangi.

Sarnane olukord on täheldatav CAD (BIM -) vahendite valdkonnas, mille andmed täidavad tänapäeval ERP/PMIS -süsteemide projekteerimisüksuste mahuparameetrid. Algselt põhines BIMi idee (mis töötati välja juba 2002. aastal [110]) ühe integreeritud andmebaasi kontseptsioonil, kuid praktikas nõuab tänapäeval BIMiga töötamine tervet rida spetsiaalset tarkvara ja formaate. See, mis pidi lihtsustama projekteerimise ja ehituse juhtimist, on muutunud järjekordseks korraks omaloodud lahendusi, mis raskendavad integreerimist ja vähendavad ettevõtte paindlikkust.

Järgmised sammud: projektiandmete tõhus kasutamine

Selles osas näitasime, kuidas struktureeritud andmed saavad aluseks ehitusprojektide täpsetele

kulude ja ajakava arvutustele. QTO, ajakava ja kalkulatsiooniprotsessi automatiseerimine vähendab tööjõukulusid ja parandab oluliselt tulemuste täpsust.

Selle osa kokkuvõtteks tasub välja tuua peamised praktilised sammud, mis aitavad teil rakendada käsitletud lähenemisviise oma igapäevaste ülesannete täitmisel. Need lähenemisviisid on universaalsed - need on kasulikud nii ettevõtte digitaalsel ümberkujundamisel kui ka arvutustega tegelevate spetsialistide igapäevatoos:

■ Automatiseerida rutiinseid arvutusi

- ☐ Püüdke leida standardseid tööde maksumusi, mida saate oma töös seostada.
- ☐ Analüüsige, milliseid meetodeid kasutatakse teie riigi ehitusplatsil tööde või protsesside maksumuse või kalkulatsiooni arvutamiseks (joonis 5.1-7).
- ☐ Kui töötate CAD-süsteemiga - õppige tundma spetsifikatsioonide ja QTO andmete automaatse väljavõtte funktsioone oma CAD- (BIM-) tarkvaras.
- ☐ Kasutage LLM-i, et kirjutada arvutuste automatiseerimiseks koodi eelnõu

■ Arendage oma QTO tööriistad

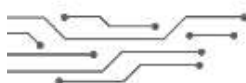
- ☐ Looge skriptid või tabelid mahu lugemise automatiseerimiseks
- ☐ Standardiseerida kategooriad ja elementide rühmad, et saavutada ühtne lähenemisviis hindamisele.
- ☐ dokumenteerida arvutusmetoodika, et tagada tulemuste korratavus uutes projektides.

■ Integreerige projekti erinevad aspektid oma töösse

- ☐ Kui töötate moodulsüsteemidega, püüdke visualiseerida oma protsesse mitte ainult diagrammide või graafikutena, vaid ka andmete tasandil - eelkõige tabelite kujul.
- ☐ CAD-andmebaasidest saadud andmete automaatne ühendamine arvutustega - Pythoni koodiga, kasutades rühmitamist, filtreerimist ja summeerimist.
- ☐ Luua selgeid visualiseeringuid QTO rühmade kohta, et esitada kolleegidele ja klientidele keerulist teavet.

Need sammud aitavad luua jätkusuutliku arvutussüsteemi, mis põhineb automatiseerimisel ja andmete standardimisel. Selline lähenemisviis parandab täpsust ja vähendab igapäevaste arvutusprobleemide rutiini.

Järgmistes peatükkides keskendutakse CAD - (BIM-) toodete tehnilistele aspektidele ja põhjustele, miks CAD-andmebaase on ikka veel raske integreerida ettevõtete äriprotsessidesse. Kui teid ei huvita praegu BIM-i rakendamise ajalugu ehituses, CAD-vahendite areng ja nende tehnoloogiatega töötamise tehnilised üksikasjad, võite minna otse raamatu seitsmenda osa "Andmepõhine otsuste tegemine" juurde.



MAKSIMAALNE MUGAVUS TRÜKITUD VERSIOONIGA

Sul on käes tasuta digitaalne versioon raamatust **Andmepõhine ehitus**. Mugavamaks tööks ja kiireks juurdepääsuks materjalidele soovitame pöörata tähelepanu **trükitud väljaandele**:



■ **Alati käepärast:** trükitud raamatust saab usaldusväärne töövahend, mis võimaldab teil igas töösituatsioonis kiiresti leida ja kasutada õigeid visualiseeringuid ja skeeme.

■ **Kvaliteetsed illustratsioonid:** kõik trükiväljaande pildid ja graafika on esitatud maksimaalse kvaliteediga.

■ **Kiire juurdepääs teabele:** lihtne navigeerimine, võimalus teha märkmeid, järjehoidjaid ja töötada raamatuga igal pool.



Ostes raamatu trükitud täisversiooni, saate mugava

abivahendi mugavaks ja tõhusaks tööks teabega: võimaluse kasutada visuaalseid materjale kiiresti igapäevastes ülesannetes, leida kiiresti vajalikud skeemid ja teha märkmeid. Lisaks toetab teie ostuga avatud teadmiste levitamist.

Tellige raamatu trükitud versioon aadressil: datadrivenconstruction.io/books



VI OSA

CAD JA BIM: TURUNDUS, TEGELIKKUS JA PROJEKTEERIMISANDMETE TULEVIK EHTUSES

Raamatu kuuendas osas analüüsitakse kriitiliselt CAD ja BIM -tehnoloogiate arengut ja nende mõju andmehaldusprotsessidele ehituses. Selles jälgitakse BIM-kontseptsiooni ajaloolist muutumist algsest integreeritud andmebaasi ideest kuni praeguste turunduskonstruksioonideni, mida tarkvaratootjad propageerivad. Hinnatakse patenteeritud vormingute ja suletud süsteemide mõju projektide andmekäitluse tõhususele ja ehitussektori üldisele tulemuslikkusele. Üksikasjalikult analüüsitakse erinevate CAD-süsteemide ühilduvuse probleeme ja nende integreerimise raskusi ehitusettevõtete äriprotsessidega. Arutatakse praegusi suundumusi lihtsustatud avatud andmeformaate, näiteks USD, suunas ja nende võimalikku mõju tööstusele. Tutvustatakse alternatiivseid lähenemisviise teabe väljavõtmiseks suletud süsteemidest, sealhulgas tagasipööramise tehnikat. Analüüsitakse tehisintellekti ja masinõppe rakendamise väljavaateid projekteerimise ja andmeanalüüsi protsesside automatiseerimiseks ehituses. Sõnastatakse prognoosid projekteerimistehnoloogiate arendamiseks, mis on orienteeritud pigem kasutajate tegelikele vajadustele kui tarkvara tarnijate huvidele.

PEATÜKK 6.1.

BIM- KONTSEPTSIOONIDE TEKKIMINE EHITUSSEKTORIS

Algselt ei olnud see kuues osa, mis oli pühendatud CADile (BIM), raamatu esimeses versioonis veel olemas. Kaitstud vormingute, geomeetriliste tuumade ja suletud süsteemide teemad on liiga tehnilised, ülekoormatud üksikasjadega ja näiliselt kasutud neile, kes tahavad lihtsalt aru saada, kuidas andmetega töötada. Tagasiside ja taotlused lisada raamatu esimesse versiooni selgitusi näitasid siiski, et ilma CAD-süsteemide sisemise toimimise keerukuse, geomeetriliste tuumade, erinevate vormingute ja samade andmete ühildumatute salvestusskeemide mõistmiseta on võimatu tõeliselt mõista, miks müüjate poolt propageeritud kontseptsioonid sageli raskendavad teabega töötamist ja takistavad üleminekut avatud parameetrilisele projekteerimisele. Seepärast on see osa võtnud oma koha raamatu ülesehituses. Kui CAD (BIM) ei ole teie jaoks esmatähtis, võite otse edasi minna järgmisesse ossa - "VII OSA: Andmepõhine otsuste tegemine, analüütika, automatiseerimine ja masinõpe".

BIM-i ja avatud BIM-i kui CAD- müüjate turunduskontseptsioonide tekkimise ajalugu.

Digitaalsete andmete tulekuga 1990ndatel võeti arvutitehnoloogia kasutusele mitte ainult äriprotsessides, vaid ka projekteerimisprotsessides, mis viis selliste kontseptsioonide nagu CAD (arvutipõhised projekteerimissüsteemid) ja hiljem BIM (ehitusinfo modelleerimine) tekkimiseni

Kuid nagu iga innovatsioon, ei ole ka need siiski arengu lõpp-punkt. Kontseptsioonidest nagu BIM on saanud oluline verstapost ehitussektori ajaloos, kuid varem või hiljem võivad need anda teed parematele vahenditele ja lähenemisviisidele, mis vastavad paremini tuleviku väljakutsetele.

CAD müüjate mõju all ja omaenda rakendamise keerukuse tõttu segaduses, ei pruugi 2002. aastal ilmunud BIM kontseptsioon oma kolmekümnnendat aastapäeva ära elada, nagu rokkstaar, mis vilksatas eredalt, kuid hääbus kiiresti. Põhjus on lihtne: andmeteadlaste nõudmised muutuvad kiiremini, kui CAD-müüjad suudavad nendega kohaneda.

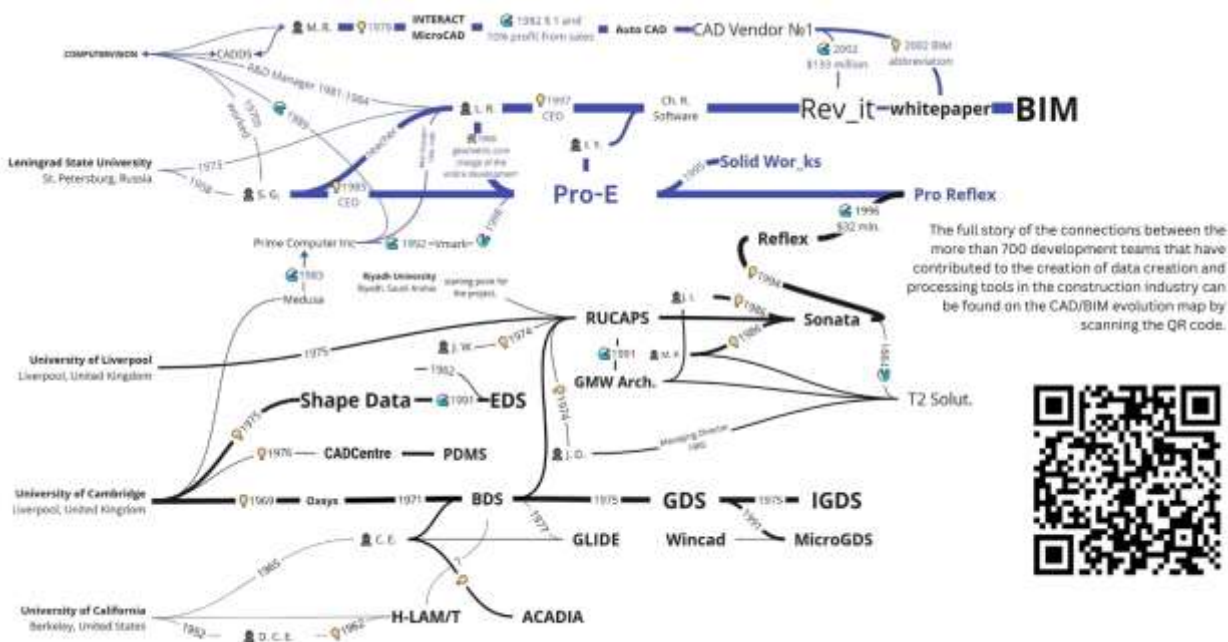
Seistes silmitsi kvaliteetsete andmete puudumisega, nõuavad tänapäeva ehitussektori spetsialistid platvormideülest koostalitlusvõimet ja juurdepääsu CAD- projektide avatud andmetele, et lihtsustada nende analüüsi ja töötlemist. CAD-andmete keerukus ja CAD-andmete segane töötlemine mõjutab negatiivselt kõiki ehitusprotsessis osalejaid: projekteerijaid, projektijuhte, ehitustöötajaid kohapeal ja lõpuks ka klienti.

Täieliku andmekogumi asemel saavad klient ja investor täna konteinerid CAD- formaadis, mis nõuavad keerulisi geomeetrilisi tuumasid, andmeskeemide mõistmist, igal aastal ajakohastatavat API - dokumentatsiooni ja spetsiaalset CAD-tarkvara (BIM), et andmetega töötada. Samal ajal jääb suur osa projekteerimisandmetest kasutamata.

Tänapäeva projekteerimis- ja ehitusmaailmas viib CAD-andmete kättesaadavuse keerukus projektijuhtimise liigsele insenerlikkusele. Keskmised ja suured ettevõtted, kes töötavad CAD-andmetega või arendavad BIM-lahendusi, on sunnitud kas säilitama tihedaid suhteid CAD-tootjatega lahendused, et pääseda andmetele ligi APIde kaudu, või mööda CAD-tootjate piirangutest, kasutades kalleid SDK-konvertereid, et pöördprojekteerida, et saada avatud andmeid [75].

Varaliste andmete kasutamine on aegunud ja ei vasta enam tänapäeva digitaalse keskkonna nõuetele. Tulevikus jagunevad ettevõtted kahte tüüpi: need, kes kasutavad avatud andmeid tõhusalt, ja need, kes lahkuvad turult.

Kontseptsioon BIM (Building Information Modeling), ilmus ehitustööstuses ühe suurema CAD-tootja poolt 2002. aastal avaldatud Whitepaper BIM [54], mida täiendas masinaehituse kontseptsioon BOM (Bills of Materials), mis sai alguse parameetrisest lähenemisest projektiandmete loomisele ja töötlemisele (joonis 6.1-1). Parameetiline lähenemine projekteerimisandmete loomisele ja töötlemisele oli üks esimesi, mida rakendati masinaehituse projekteerimise süsteemis Pro-E (MCAD). See süsteem sai [111] paljudele kaasaegsetele CAD lahendustele, sealhulgas neile, mida tänapäeval kasutatakse ehitustööstuses.



Joonis 6.1-1 BIM-kontseptsiooni ja sarnaste kontseptsioonide ajaloo kaart.

Ajakirjanikud ja AEC-konsultandid, kes kuni 2000ndate alguseni propageerisid CAD-vahendeid - pakkujad, pöörasid alates 2002. aastast oma tähelepanu Whitepaper BIMile. Just Whitepaper BIM 2002-2004 ja aastatel 2002, 2003, 2005 ja 2007 avaldatud artiklid mängisid võtmerolli BIM-kontseptsiooni populariseerimisel ehitussektoris [112].

*Ehitusinfo modelleerimine on strateegia..... [CAD-müüja ettevõtte nimi]
infotehnoloogia rakendamiseks ehitussektoris.*

- BIM Whitepaper, 2002 [60].

2000ndate keskpaigaks hakkasid "teadlased" siduma CAD- müüja poolt 2002. aastal avaldatud BIM-kontseptsiooni varasemate teadustööde, näiteks Charles Eastmani BDS, mis sai aluseks sellistele süsteemidele nagu GLIDE, GBM, BPM, RUCAPS. Charles Eastman pani oma teedrajas teoses Building Description System (1974) kaasaegse infomodelleerimise teoreetilised alused. Termin "andmebaas" esineb tema teoses 43 korda (joonis 6.1-2) - sagedamini kui ükski teine, välja arvatud sõna "hoone".

Eastmani põhiidee oli, et kogu teave hoone kohta - alates geometriast kuni elementide omaduste ja nende omavaheliste seosteni - tuleks salvestada ühtsesse struktureeritud andmebaasi. Sellest andmebaasist saab automaatselt genereerida ja analüüsida jooniseid, spetsifikatsioone, arvutusi ja eeskirjadele vastavust. Eastman kritiseeris selgesõnaliselt jooniseid kui vananenud ja üleliigset kommunikatsioonimeetodit, viidates teabe dubleerimisele, ajakohastamisega seotud probleemidele ja vajadusele muudatuste tegemisel käsitsi ajakohastada. Selle asemel pakkus ta välja üheainsa digitaalse mudeli kasutamise andmebaasis, kus iga muudatus tehakse üks kord ja see kajastub automaatselt kõigis vaadetes.

On märkimisväärne, et Eastman ei seadnud visualiseerimist oma kontseptsiooni keskmesse. Tema süsteemis oli kesksel kohal teave: parameetrid, seosed, atribuudid, analüüs ja automatiseerimisvõimalused. Joonised olid tema käsitluses vaid üks andmebaasi andmete kuvamise vormidest, mitte peamine projekteerimisalase teabe allikas.

Esimeses BIMi käsitlevas Whitepaperis juhtivalt CAD pakkujalt kasutati väljendit "andmebaas" sama sageli kui Charles Eastmani BDSis - 23 korda [60] seitsme lehekülje jooksul ja see oli üks populaarsemaid sõnu dokumendis pärast sõnu "Building", "Information", "Modeling" ja "Design". Kuid 2003. aastaks esineb termin "andmebaas" sarnastes dokumentides ainult kaks korda [61] ning 2000. aastate lõpuks oli andmebaaside teema disainandmete arutelust praktiliselt kadunud. Selle tulemusel ei ole "ühtse integreeritud andmebaasi visuaalse ja kvantitatiivse analüüsi jaoks" kontseptsioon kunagi täielikult realiseerunud.

Nii on ehitussektor jõudnud Charles Eastmani progressiivsest BDS-kontseptsioonist, milles rõhutati andmebaase ja Samuel Geisbergi ideid projekteerimisandmete automaatsest ajakohastamisest andmebaasidest masinaehitustootes Pro-E (tänapäeval ehituses kasutatavate populaarsete CAD - lahenduste eelkäija) kuni praeguse turustatud BIM, kus andmehaldust andmebaaside kaudu vaevu mainitakse, hoolimata sellest, et see oli algse teoreetilise kontseptsiooni aluseks.

11. - Integrated Basis of a Design by the Descriptive System

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

12. - Design

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

It is not sufficient to state that a computerized description of a building must facilitate its function in all various situations of design, or to state that the building design, construction and operation as well as the maintenance of these various functions. But design and construction must be integrated and must work within the architectural, structural, and mechanical description of a building. The design must be integrated with the construction and the operation as well as the maintenance of the building. In addition, space as well as time must be explicitly mentioned. The design must provide a single description of each building element.

Figure 11 depicts the **integrated** and its basic structure on design. In the various situations, each of the building's operations and values are defined. The operations as the origin are shown. Each level is shown as a separate data element.

A user must be able to conveniently access a new pattern, new expression for an existing pattern, or new values for an existing expression/pattern available. Therefore, consideration must be given to editing and redefining of existing patterns, expressions, and values already stored. Also in certain elements, there will be no need to define certain conditions through expressions and values because all instances will have the same shape and dimensions. In this case, consideration should be entered directly, without underlying expressions. There are only **simple** elements. All the elements described are provided by the **integrated**, as shown in Figure 11.

Within the **integrated**, elements in the different data elements are through a common structure. Each pattern, expression, and instance have a unique value and value. Within the **integrated**, all expressions based on a common pattern are shown. Within all instances based on common expressions are shown. Also, these instances without expressions that are directly defined are linked to the pattern that they are associated with. These linkages duplicate the relations shown in Figure 11 and allow operations on these sets of elements related by the hierarchy.

The details of the **integrated** are provided in a separate report. The **integrated** has been implemented and is now undergoing preliminary testing.

12. - Spatial Interplay

Efforts were made at the onset of the building description system project to find ways to access the **integrated** according to the spatial organization of elements. Particularly needed in the capability to access all elements developing a spatial sense of interest. Two significant advances were made on this problem resulting in very efficient spatial access being achieved by 3D or more and also efficient methods regarding movement to only those selected building elements of interest. In addition, also allows searching of elements to be based on the size of objects, which is extremely useful for the generation of displays or drawings. The algorithms for these related searches have been tested and both classes of algorithms are presented elsewhere.

If current architectural practices continue, it is imperative that a designer be able to enter elements of design. The use of multiple a computer used for defining such elements made of concrete, plastic, woodwork, and other locally fabricated materials. In fact, there are only a few hundred elements available for entering such data into a computer **integrated**.

Joonis 6.1-2 Charles Eastman 1974. aastal kirjeldatud BDS-kontseptsioonis kasutati väljendit "andmebaas" (kollasega esile tõstetud) 43 korda.

BDS ja sarnased kontseptsioonid kuni 2000. aastateni töötati välja pigem hoonete digitaalse andmebaasi kui visualiseerimisvahendina. BIM muutus 2002. aastal projekteerimisvahendiks, kus andmebaas jäi tagaplaanile. Mida oleme kaotanud üleminekul BDS-ilt ja sarnastelt kontseptsioonidelt 1990ndatel BIMile 2010ndate keskpaigaks:

- Avatud andmebaasid: BDS ja muud sarnased kontseptsioonid rõhutasid analüütilisust, BIM rõhutas projekteerimist.
- Andmetega töötamise paindlikkust: BDS rõhutas andmeanalüüsi, BIM rõhutas protsesse, mis põhinevad ebaselgetel andmetel.
- Läbipaistvus: BDS pidi olema avatud integreeritud andmebaas, samal ajal kui BIM CAD-müüjad on teinud oma andmebaasid täiesti kinniseks ja on 20 aastat edutult võidelnud pöördprojekteerimise vahendite vastu, mis avavad patenteeritud vorminguid.

Viimase 30 aasta jooksul ei ole projekteerijatel kunagi olnud juurdepääsu "integreeritud andmebaasile" ja pärast kakskümmend aastat kestnud BIM -tööriistade turundamise eufooriat hakkab ehitussektor mõistma selle moeröögatuse tagajärgi.

BIM-i tegelikkus: integreeritud andmebaaside asemel - suletud moodulsüsteemid

Selle asemel, et keskenduda andmetele, neid struktureerida ja integreerida ühtsesse protsessidesse, on CAD - (BIM-) süsteemide kasutajad sunnitud töötama killustunud varaliste lahenduste kogumiga, millest igaüks dikteerib oma mängureeglid:

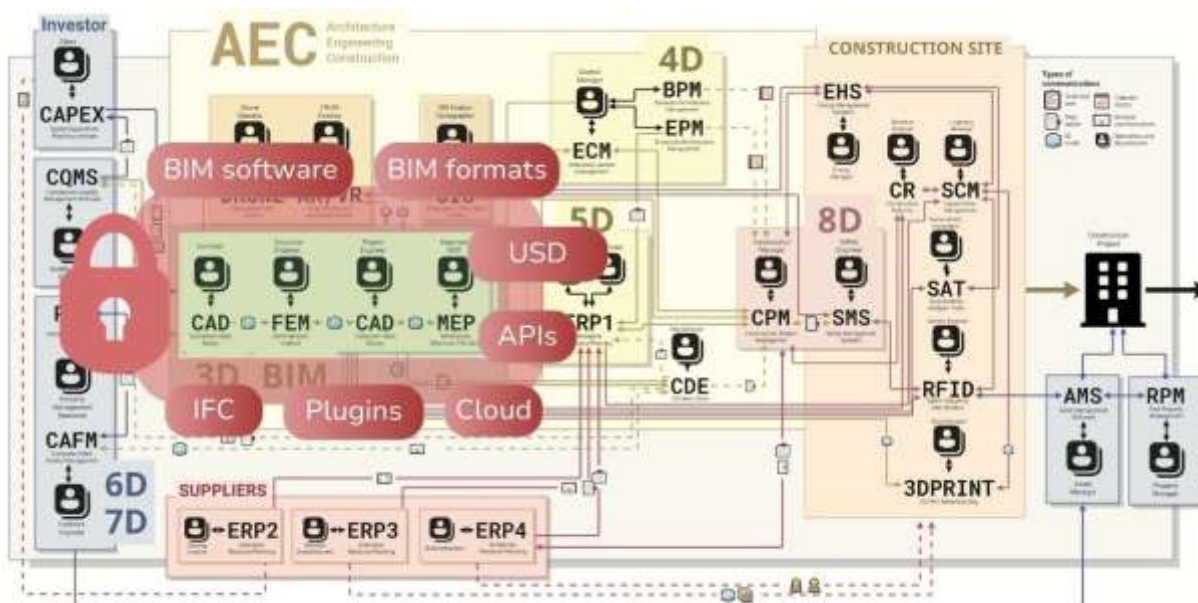
- Ühtne andmebaas, mida käsitleti esimeses BIM Whitepaperis, on jäänud müüdiks. Hoolimata kõlavatest väidetest on juurdepääs andmetele endiselt piiratud ja jaotatud suletud süsteemide vahel.

- **BIM - mudelid on muutunud pigem suletud ökosüsteemiks** kui tööriistaks. Läbipaistva teabevahetuse asemel on kasutajad sunnitud maksma tellimuste eest ja kasutama patenteeritud APIsid.
- **Andmed kuuluvad müüjatele, mitte kasutajatele.** Projektiteave on lukustatud patenteeritud vormingutesse või pilveteenustesse, selle asemel et olla kättesaadav avatud ja sõltumatutes vormingutes.

Projekteerijatel ja projektijuhtidel ei ole sageli juurdepääsu CAD-andmebaasile ega formaadile, milles nende enda projektiandmed on salvestatud. See muudab võimatuks teabe kiire kontrollimise või andmete struktuurile ja kvaliteedile esitatavate nõuete sõnastamise (joonis 6.1-3). Juurdepääs sellistele andmetele nõuab tervet rida spetsiaalseid programme, mis on seotud APIde ja pistikprogrammide kaudu, mis viib ehitussektori protsesside liigse bürokraatiseerumiseni. Samal ajal kasutavad neid andmeid samaaegselt kümned infosüsteemid ja sajad spetsialistid.

*Me peame suutma hallata kõiki neid andmeid [CAD (BIM)] digitaalselt salvestada ning müüa elutsükli- ja protsessihaldustarkvara, **sest iga inseneri** [projekteerija] **kohta**, kes loob midagi [CAD-tarkvaras], **on kümme inimest**, kes töötavad nende andmetega" [41].*

- CAD - BIM-kontseptsiooni, 2005. aastal loonud tootja tegevjuht.



Joonis 6.1-3 CAD- (BIM-) andmebaasid on endiselt üks viimaseid suletud süsteeme IT-osakondade ja andmehaldurite jaoks ehitusettevõtete ökosüsteemis.

Kui selgub, et BIM on pigem vahend andmebaaside kommertsialiseerimiseks kui täieõiguslik andmebaaside haldamise vahend, tekib loogiline küsimus: kuidas me saame taastada kontrolli andmete üle? Vastus on kasutada avatud andmestruktuure, kus kasutaja, mitte tarkvaratootja, saab teabe omanikuks.

Ehitustööstuse kasutajad ja lahenduste arendajad, nagu ka nende kolleegid teistes tööstusharudes, eemalduvad paratamatult viimased 30 aastat domineerinud ebamäärasest tarkvaratootjate terminoloogiast, keskendudes digitaliseerimise põhiaspektidele - "andmed" ja "protsessid".

Veel 1980ndate lõpus esitati ehituses digitaalse arengu võtmevaldkonda kui andmetele juurdepääsu ja projektiteabe haldamise küsimust. Aja jooksul on aga rõhuasetus nihkunud. Selle asemel, et arendada läbipaistvaid ja kättesaadavaid lähenemisviise andmetega töötamiseks, propageeriti aktiivselt IFC-vormingut ja avatud BIM-kontseptsiooni kui katseid juhtida spetsialistide tähelepanu kõrvale projektiandmebaaside haldamise teemadelt.

Avatud formaadis IFC ilmumine ehitussektoris

Niinimetatud avatud formaat IFC (Industry Foundation Classes) on standardiks, mis tagab koostalitlusvõime erinevate CAD (BIM -) süsteemide vahel. Selle väljatöötamine toimus organisatsioonide raames, mis on loodud ja mida kontrollivad suured CAD-müüjad. IFC-formaadi alusel töötasid kaks CAD- ettevõtet 2012. aastal välja turunduskontseptsiooni OPEN BIM[63].

IFC (Industry Foundation Classes) on avatud standard andmevahetuseks ehitustööstuses, mis on mõeldud erinevate CAD - (BIM-) süsteemide koostalitlusvõime tagamiseks.

Avatud BIM - kontseptsioon hõlmab tööd CAD-andmebaaside teabega ja teabevahetust süsteemide vahel avatud CAD-andmete vahetamise formaadi - IFC - kaudu.

Open BIM programm on turunduskampaania, mille algatasid... [1 CAD-tootja]... [2 CAD-tootja] ja teised ettevõtted, et julgustada ja hõlbustada OPEN BIM kontseptsiooni ülemaailmset koordineeritud edendamist kogu AEC-tööstuses, kusjuures programmis osalejatele on kättesaadav ühtne kommunikatsioon ja ühine kaubamärk.

- CAD-müüja veebisaidilt, OPEN BIM programm, 2012 [113].

IFC kohandati Müncheni Tehnikaülikooli poolt 1980ndate lõpus masinaehitusformaadist STEP, ning hiljem registreeriti see ühe suure projekteerimisfirma ja ühe suure CAD- tootja poolt, et moodustada 1994. aastal IAI (Industry Alliance for Interoperability) [114] (joonis 6.1-4). IFC formaat töötati välja selleks, et tagada koostalitlusvõime erinevate CAD süsteemide vahel, ning see põhines põhimõtetel, mis on sätestatud masinaehitusformaadis STEP, mis omakorda sai alguse IGES formaadist, mis loodi 1979. aastal CAD kasutajate ja tarnijate rühma poolt NIST (The National Institute of Standards and

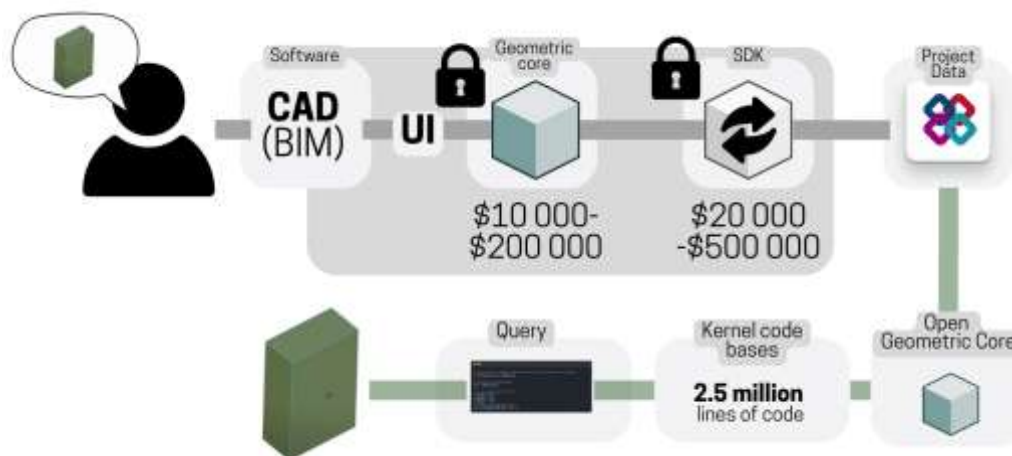
Ehitusandmete loomise ja töötlemise vahendite loomisel osalenud enam kui 700 arendusmeeskonna koostöömise üksikasjalik ajalugu on esitatud kaardil "The Evolution of CAD (BIM)" [116]. [116].

Avatud kujul IFC koosneb konstruktsioonelementide geomeetrisest kirjeldusest ja metainformatsiooni kirjeldusest. IFC-vormingus geomeetria kujutamiseks kasutatakse erinevaid meetodeid, näiteks CSG ja Swept Solids: parameetiline kujutamine BREP on siiski muutunud juhtivaks standardiks elementide geomeetria edastamisel IFC-vormingus, kuna see vorming on toetatud CAD-(BIM-) programmide eksportimisel ja võimaldab elementide võimalikku redigeerimist IFC-formaadis CAD-programmidesse tagasi importimisel.

IFC formaadi probleem sõltuvalt geomeetrisest tuumast

Enamasti, kui geomeetria on IFC-s määratletud parameetriselt (BREP), muutub võimatuks geomeetriseliste omaduste, näiteks projektiüksuste ruumala või pindala visualiseerimine või väljavõtete tegemine ainult IFC-faili abil, sest sellisel juhul on geomeetriaga töötamiseks ja selle visualiseerimiseks vaja geomeetriselise tuuma (joonis 6.1-5), mis esialgu puudub.

Geomeetria tuum on tarkvarakomponent, mis pakub põhilisi algoritme geomeetriseliste objektide loomiseks, redigeerimiseks ja analüüsimiseks CAD (CAD), BIM ja muudes insenerirakendustes. See vastutab 2D ja 3D -geomeetria koostamise ja sellega tehtavate operatsioonide eest, näiteks: Bool'i operatsioonid, silumine, lõiked, transformatsioonid ja visualiseerimine.



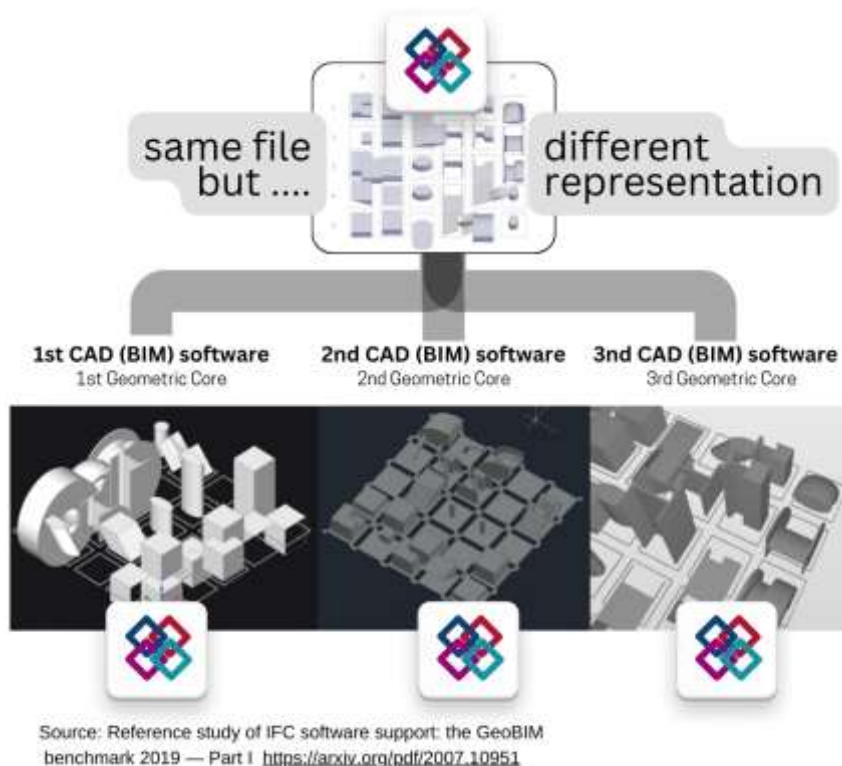
Joonis 6.1-5 Geomeetria loomine CAD- tarkvara abil toimub tänapäeval patenteeritud geomeetria tuumade ja SDKde kaudu, mis sageli ei kuulu CAD- tarnijatele.

Igal CAD-programmil ja kõigil programmidel, mis töötavad parameetriseliste või IFC-vormingutega, on oma või ostetud geomeetriselise tuum. Ja kui IFC -BREP formaadis olevate primitiivsete elementidega ei ole probleeme ja erinevate geomeetriseliste tuumadega programmides saab neid elemente kuvada sarnaselt, kuid lisaks probleemidele erinevate geomeetriseliste tuumade mootoritega, on piisavalt elemente, millel on oma eripärad korrektseks kuvamiseks. Seda probleemi käsitletakse üksikasjalikult

2019. aastal avaldatud rahvusvahelises uuringus "A reference study of IFC software support" [117].

Samad standardiseeritud andmekogumid annavad vastuolulisi tulemusileitakse vähe ühiseid mustreid ning standardi toetamisel on leitud tõsiseid probleeme [IFC], mis on ilmselt tingitud standardi andmemudeli väga suurest keerukusest. Osaliselt on selles süüdi ka standardid ise, kuna need jätavad sageli mõned üksikasjad määratlemata, millel on suured vabadusastmed ja erinevad tõlgendusvõimalused. Need võimaldavad objektide organiseerimise ja säilitamise suurt keerukust, mis ei soodusta tõhusat universaalset mõistmist, unikaalseid rakendusi ja järjepidevat andmemodelleerimist [117].

- IFC-tarkvara toetuse võrdlusuuring, 2021



Joonis 6.1-6 Erinevad geomeetrilised tuumad annavad sama parameetriliselt kirjeldatud geomeetria erinevaid esitusi (põhineb [117]).

"Teatavate sätete" õiget mõistmist saavad IFC-d välja töötavate eriorganisatsioonide tasulised liikmed. Sellest tulenevalt püüab igaüks, kes soovib saada juurdepääsu olulistele teadmistele IFC teatavate omaduste kohta, teha koostööd suurte CAD- müüjatega või jõuda oma uurimistööga omaduste kvalitatiivsele käsitlemisele

Sa komistad küsimuse peale andmete importimise ja eksportimise kohta IFC-vormingus ja küsid kolleegide käest: "Miks on IFC-failis nii teave ruumide parameetrilise ülekande kohta? Avatud spetsifikatsioon ei ütle selle kohta midagi". Vastus "asjatundlikumatelt" Euroopa müüjatelt: "Jah, seda ei ole öeldud, kuid see on lubatud".

- CAD 2021 arendaja intervjuust [118].

IFC kirjeldab geomeetriat parameetriliste primitiivide kaudu, kuid ei sisalda varjatud tuuma - selle rolli täidab CAD-programm, mis koostab geomeetria geomeetrilise tuuma kaudu. Geomeetria tuum teostab matemaatilised arvutused ja määratleb lõikepunktid, samas kui IFC annab ainult andmed selle tõlgendamiseks. Kui IFC sisaldab ebaõigeid pindu, võivad erinevad programmid, millel on erinevad geomeetria tuumad, neid kas ignoreerida või tekitada vigu, sõltuvalt tuumast.

Sellest tulenevalt on IFC-vorminguga töötamiseks vaja vastata põhiküsimusele, millele on raske leida üheselt mõistetavat vastust - millist vahendit, millise geomeetrilise tuumaga tuleks kasutada, et saada algselt CAD-programmis, millest IFC on saadud, projektis kasutatud andmete kvaliteet?

Andmete kvaliteediprobleemid ja IFC formaadi keerukus ei võimalda projekteerimisandmete otsest kasutamist protsesside automatiseerimiseks, analüüsiks ja andmetöötluseks, mis viib arendajad sageli paratamatult vajadusele kasutada suletud CAD -lahendusi, millel on "kvaliteetne" juurdepääs andmetele[63], millest on kirjutanud IFC 1994. aastal registreeritud IFC müüja ise [65].

Kõik IFC parameetrite kaardistamise ja genereerimise eripärad geomeetria tuumas on võimalik realiseerida ainult suurte arendusmeeskondade poolt, kellel on geomeetria tuumaga töötamise kogemus. Seetõttu on IFC-vormingu eripärade ja keerukuse praegune praktika kasulik eelkõige CAD-müüjatele ja sellel on palju ühist suurte tarkvaratootjate strateegiaga "võta vastu, laienda, hävita", kui standardi kasvav keerukus tegelikult loob tõkkeid väikestele turuosalistele [94].

Suurte strateegia võib olla avatud standardite kohandamine, omandiõigusega laienduste ja funktsioonide lisamine, et luua kasutajate sõltuvus nende toodetest, et seejärel konkurendid välja suruda.

IFC-vorming, mis on mõeldud universaalseks sillaks erinevate CAD- (BIM-) süsteemide vahel, täidab tegelikult erinevate CAD-platvormide geomeetriliste tuumade ühilduvusprobleemide näitaja rolli, sarnaselt formaadiga STEP, millest see algselt tekkis.

Selle tulemusel on tänapäeval IFC ontoloogia täielik ja kvaliteetne rakendamine teostatav suurte CAD-tootjate jaoks, kes saavad investeerida märkimisväärsed ressursse, et toetada kõiki üksusi ja nende seostamist oma sisemise geomeetria tuumaga, mida IFC kui standardi puhul ei ole olemas. Suurtel tarnijatel on ka võimalus kooskõlastada omavahel funktsioonide tehnilisi üksikasju, mis ei pruugi olla kättesaadavad isegi kõige aktiivsemale osalejale IFC-vormingu arendusorganisatsioonides.

Väikeste sõltumatute meeskondade ja avatud lähtekoodiga projektide jaoks, mis püüavad toetada koostalitlusvõimeliste vormingute arendamist, muutub sisemise geomeetria tuumiku puudumine tõsiseks probleemiks. Ilma selleta on praktiliselt võimatu võtta arvesse kõiki paljusid platvormideülese andmevahetusega seotud nüansse ja nüansse.

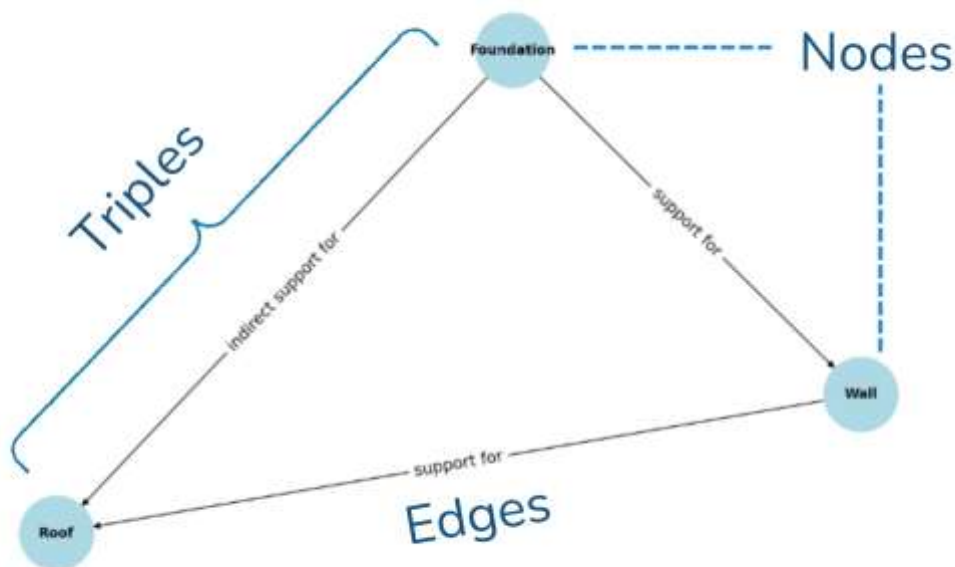
Seoses IFC parameetrilise formaadi ja avatud BIM-kontseptsiooni väljatöötamisega on ehitussektoris intensiivistunud arutelud ontoloogia ja semantika rolli üle andmete ja protsesside haldamisel.

Ilumine semantika teema ülesehitamisel ja ontoloogia

Tänu semantilise interneti ideedele 1990ndate ja IFC-vormingu väljatöötamises osalenud organisatsioonide jõupingutustele on semantika ja ontoloogiad muutunud 2020ndate keskpaigaks ehitussektoris arutatava standardimise üheks põhielemendiks.

Semantilised tehnoloogiad on suurte heterogeensete andmemassiivide ühtlustamine, standardiseerimine ja muutmine ning keeruliste otsingute teostamine.

Semantiliste andmete salvestamiseks kasutatakse OWL (Web Ontology Language), mis on esitatud RDF-graafidena -tripletid (Resource Description Framework) (joonis 6.1-7). OWL viitab graafi andmemudelitele, mille tüüpe käsitlesime lähemalt peatükis "Andmemudelid: andmesuhted ja elementidevahelised suhted".



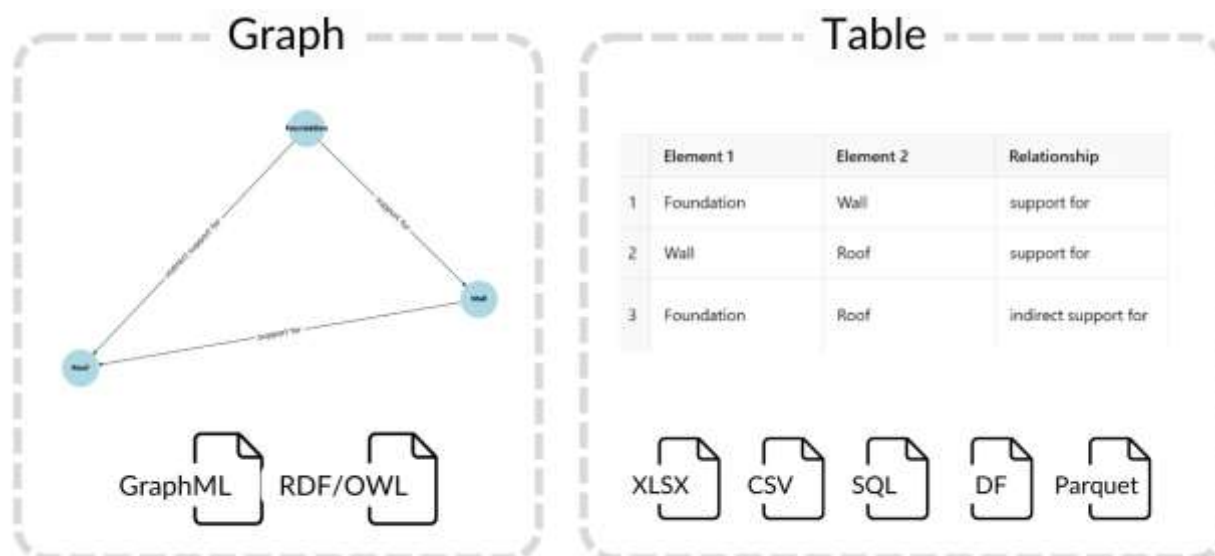
Joonis 6.1-7 RDF andmemudel: sõlmed, servad ja kolmikud, mis illustreerivad ehitusplokkide vahelisi seoseid.

Teoreetiliselt võimaldab risonerite (automaatse loogilise järeldamise tarkvara) loogiline järeldamine tuletada ontoloogiatest uusi väiteid. Näiteks kui ehitusontoloogias on kirjas, et "vundament on seina tugi" ja "sein on katuse tugi" (joonis 6.1-7), suudab risoner automaatselt järeldada, et "vundament on katuse tugi".

Selline mehhanism on kasulik andmete analüüsi optimeerimiseks, sest sellega välditakse kõigi sõltuvuste selgesõnalist määramist. Siiski ei loo see uusi teadmisi, vaid üksnes tuvastab ja struktureerib juba teadaolevad faktid.

Semantika ei loo iseenesest uut tähendust või teadmisi ning ei ole selles aspektis teistest andmesalvestus- ja andmetöötlustehnoloogiatest parem. Relatsiooniliste andmebaaside andmete esitamine kolmikutena ei muuda neid tähenduslikumaks. Tabelite asendamine graafstruktuuridega võib olla kasulik andmemudelite ühtlustamiseks, lihtsaks otsimiseks ja turvaliseks redigeerimiseks, kuid see ei tee andmeid "targemaks" - arvuti ei hakka nende sisu paremini mõistma.

Andmete loogilisi seoseid saab korraldada ilma keeruliste semantiliste tehnoloogiateta (joonis 6.1-8). Traditsioonilised relatsioonilised andmebaasid (SQL) ning CSV või XLSX formaadid võimaldavad sarnaste sõltuvuste loomist. Näiteks saab veergandmebaasis lisada välja "katuse tugi" ja seina loomisel automaatselt seostada katuse vundamendiga. See lähenemisviis on rakendatud ilma RDF, OWL, graafide või riisonite kasutamiset, jäädes lihtsaks ja tõhusaks lahenduseks andmete salvestamiseks ja analüüsimiseks.



Joonis 6.1-8 Graafi ja tabeli andmemudelite võrdlus samade loogiliste seoste kujutamiseks.

Mitmete suurte ehitusettevõtete ja IFC-vormingu arendusorganisatsiooni [94] otsus järgida semantilise veebi kontseptsiooni, mis tundus 1990ndate lõpus paljulubav, on avaldanud olulist mõju ehitussektori standardite arengule.

Paradoksaalne on aga see, et algselt Interneti jaoks mõeldud semantilise veebi kontseptsiooni ei ole laialdaselt vastu võetud isegi selle kodumaises keskkonnas. Hoolimata RDF ja OWL arendamisest, ei ole täieõiguslik semantiline veeb oma algses kontseptsioonis ilmunud ja selle loomine on juba praegu ebatõenäoline.

Miks semantilised tehnoloogiad ei vasta ehitussektoris ootustele

Teistes tööstusharudes on semantika kasutamise tehnoloogiate piirangutega kokku puutunud. Mängutööstuses on katsed kirjeldada mänguobjekte ja nende interaktsioone ontoloogiate abil osutunud ebaefektiivseks muutuste suure dünaamilisuse tõttu. Selle tulemusena on eelistatud lihtsamaid andmeformaate, nagu XML ja JSON, koos algoritmiliste lahendustega. Kinnisvarasektoris oli olukord sarnane: piirkondlike terminoloogiliste erinevuste ja sagedaste turumuutuste tõttu osutus ontoloogiate kasutamine liiga keeruliseks, samas kui lihtsad andmebaasid ja standardid, nagu RETS [119], said andmevahetusega seotud probleemidega paremini hakkama.

Tehnilised raskused, nagu märgistuse keerukus, suur tööjõumahukus ja arendajate vähene motivatsioon, pidurdasid semantilise veebi kasutuselevõttu ja teistes majandussektorites. RDF (Resource Description Framework) ei saanud massistandardiks ning ontoloogiad osutusid liiga keeruliseks ja majanduslikult põhjendamatuks.

Selle tulemusel jäi ambitsioonikas idee luua ülemaailmne semantiline veeb realiseerimata. Kuigi mõned selle tehnoloogia elemendid, näiteks ontoloogiad ja SPARQL, on leidnud tee

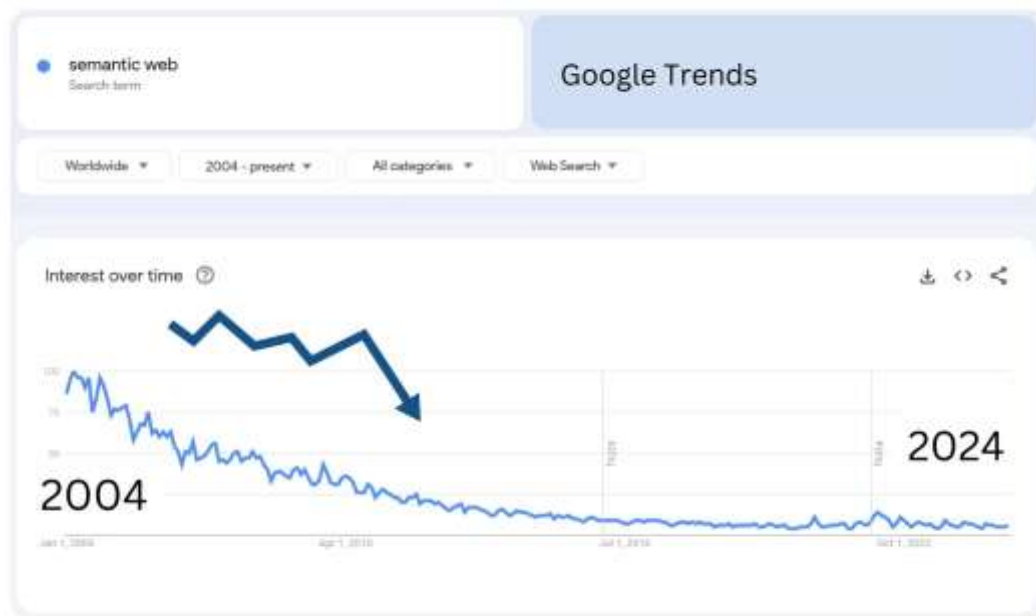
ettevõtluslahendustesse, ei ole algset eesmärki, milleks oli ühtse tervikliku andmestruktuuri loomine, saavutatud.

Interneti kontseptsioon, kus arvutid suudavad sisu mõtestada, on osutunud tehniliselt keeruliseks ja äriliselt kahjumlikuks. Seetõttu vähendasid seda ideed toetanud ettevõtted lõpuks selle kasutamist üksikutele kasulikele vahenditele, jättes RDF ja OWL pigem väga spetsiifiliste ettevõtete kui interneti kui terviku vajaduste jaoks. Google Trends'i (joonis 6.1-9) analüüs viimase 20 aasta jooksul näitab, et semantilise veebi jaoks ei pruugi olla enam mingeid väljavaateid.

Ei ole vaja üksusi asjatult paljundada. Kui mingi nähtuse kohta on mitu loogiliselt järjepidevat seletust, mis seletavad seda võrdsest hästi, tuleks, kui kõik muud asjad on võrdsed, eelistada neist kõige lihtsamat.

's Razor

Siinkohal tekib loogiline küsimus: milleks üldse kasutada konstruktsioonis kolmikuid, tõusjaid ja SPARQL, kui andmeid saab töödelda ka populaarsete struktureeritud päringute abil (SQL, Pandas, Apache®)? Ettevõtete rakendustes on SQL andmebaasidega töötamise standard. SPARQL seevastu nõuab keerulisi graafistruktuure ja spetsiaalset tarkvara ning Google'i suundumuste kohaselt ei ärata arendajate huvi.



Joonis 6.1-9 Huvi "semantilise interneti" päringute vastu vastavalt Google'i statistikale.

Graafiandmebaasid ja klassifitseerimispuud võivad mõnel juhul olla kasulikud, kuid nende kasutamine ei ole alati põhjendatud enamiku igapäevaste ülesannete puhul. Seetõttu on teadmiste graafide loomine ja semantilise veebi tehnoloogiate kasutamine mõttekas ainult siis, kui on vaja ühendada eri allikatest pärinevaid andmeid või teha keerulisi loogilisi järeldusi.

Üleminek tabelitelt graafiliste andmemudelite juurde parandab otsingut ja ühtlustab teabevoolu, kuid ei muuda andmeid masinate jaoks sisukamaks. Küsimus ei ole mitte selles, kas semantilisi tehnoloogiaid tuleks kasutada, vaid selles, kus need tõesti midagi muudavad. Enne ontoloogia, semantika ja graafandmebaaside rakendamist oma ettevõttes uurige, millised ettevõtted juba kasutavad neid tehnoloogiaid edukalt ja kus need on ebaõnnestunud.

Hoolimata ambitsioonikatest ootustest ei ole semantilised tehnoloogiad kunagi saanud universaalseks lahenduseks andmete struktureerimiseks ehitussektoris. Praktikas ei ole need tehnoloogiad viinud universaalse lahenduseni, vaid on vaid lisanud uusi keerukusi, ning need jõupingutused peegeldavad semantilise interneti kontseptsiooni realiseerimata jäänud ambitsioone, kus ootused ületasid kaugeltki tegelikkust.



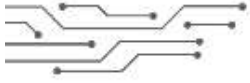
Joonis 6.1-10 Geomeetria ja teave ehitusprotsessides: keerulistest CAD- ja BIM- süsteemidest lihtsustatud andmeteks analüütika jaoks.

Kui IT-sektoris on semantilise veebi ebaõnnestumisi kompenseeritud uute tehnoloogiate (suured andmed, asjade internet, masinõpe, AR/VR) esilekerkimisega, siis ehitussektoris selliseid võimalusi ei ole.

Lisaks probleemidele, mis on seotud mõistete kasutamisega projekti elementide vaheliste andmesuhete edastamiseks, on veel üks põhiprobleem - nende andmete kättesaadavus. Ehitustööstuses domineerivad endiselt suletud süsteemid, mis raskendab tööd andmetega, teabe jagamist ja protsesside tõhususe parandamist.

Andmete suletud olemus on muutumas üheks peamiseks takistuseks, mis takistab digitaalsete lahenduste arendamist ehituses. Erinevalt IT-tööstusest, kus avatud ja ühtlustatud andmeformaadid

on muutunud standardiks, kasutab CAD-sektoris (BIM) iga tarkvara oma formaati, mis loob suletud ökosüsteeme ja piirab kunstlikult kasutajaid.



PEATÜKK 6.2.

SULETUD PROJEKTIFORMAADID JA KOOSTALITLUSVÕIME PROBLEEMID

Suletud andmed ja tootlikkuse langus: CAD-tööstuse ummikseis (BIM)

CAD - süsteemide omandiõiguse tõttu on igal programmil oma unikaalne andmeformaad, mis on kas suletud ja väljastpoolt ligipääsmatu - RVT, PLN, DWG, NDW, NWD, SKP, või on kättesaadav poolstruktureeritud kujul üsna keerulise konverteerimisprotsessi kaudu - JSON, XML (CPIXML), IFC, STEP ja ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV jne..

Erinevad andmeformaadid, milles samu andmeid samade projektide kohta saab salvestada, ei erine mitte ainult struktuuri poolest, vaid sisaldavad ka sisemiste märgendite erinevaid versioone, mida arendajad peavad arvestama, et tagada rakenduste ühilduvus. Näiteks 2025. aasta CAD-formaad avaneb 2026. aasta CAD-programmis, kuid sama projekt ei avane kunagi kõigis CAD-programmi versioonides, mis võisid olla kättesaadavad enne 2025. aastat.

Kuna ehitussektori tarkvarapakkuja ei paku otsest juurdepääsu andmebaasidele, loob ta sageli oma unikaalse formaadi ja vahendid, mida spetsialist (projekteerija või andmehaldur) peab kasutama andmetele juurdepääsuks, nende importimiseks ja eksportimiseks.

Selle tagajärjel tõstavad põhiliste CAD (BIM) ja sellega seotud lahenduste (nt ERP/PMIS)) müüjad pidevalt toodete kasutamise hindu ning tavakasutajad on sunnitud maksma "komisjonitasu" igas andmeedastuse etapis formaatide kaupa [63]: ühendamise, impordi, ekspordi ja kasutajate endi loodud andmetega töötamise eest.

Populaarsete CAD - (BIM-) toodete pilvemälu andmete kasutamise hind ulatub 2025. aastal 1 dollarini tehingu kohta [120] ja keskmise suurusega ettevõtete ehitus-ERP -toodete tellimused ulatuvad viie- ja kuuekohaliste summadeni aastas [121].

Kaasaegse ehitustarkvara olemus seisneb selles, et mitte automatiseerimine või tõhususe suurendamine, vaid inseneride võime mõista konkreetset väga spetsiifilist tarkvara mõjutab ehitusprojektide andmetöötamise kvaliteeti ja kulusid, samuti ehitusprojekte tegevate ettevõtete kasumit ja pikaajalist ellujäämist.

Juurdepääsu puudumine andmebaasidele CAD - süsteemid, mida kasutatakse kümnetes teistes süsteemides ja sadades protsessides [63], ning sellest tulenev kvaliteetse suhtluse puudumine üksikute spetsialistide vahel on viinud ehitussektori tootlikkuse poolest ühe kõige ebaefektiivsema majandussektori staatusesse [44].

Viimase 20 aasta jooksul on CAD- (BIM-) projekteerimisrakenduste, uute süsteemide (ERP), uute ehitustehnoloogiate ja -materjalide tekkimisega kogu ehitussektori tootlikkus langenud 20% (joonis 2.2-1), samas kui kõigi nende majandussektorite üldine tootlikkus, kus ei ole suuri probleeme

andmebaaside ja turunduslike BIM-kontseptsioonide kasutamisega, on kasvanud 70% (96% töötlevas tööstuses) [122].



Joonis 6.2-1 Projektide andmete eraldatuse ja keerukuse tõttu, millest sõltuvad ehitustööstuses kümned osakonnad ja sajad protsessid, on otsuste tegemise kiirus mitu korda aeglasem kui teistes tööstusharudes.

Siiski on ka üksikuid näiteid alternatiivsetest lähenemisviisidest CAD-lahenduste vahelise koostalitlusvõime loomiseks. Euroopa suurim ehitusettevõtte näitab juba 2018. aastal alustatud SCOPE projektiga [123], kuidas on võimalik minna kaugemale CAD- (BIM-) süsteemide klassikalisest loogikast. Selle asemel, et püüda allutada IFC või tugineda patenteeritud geomeetria tuumadele, kasutavad SCOPE arendajad APIsid ja SDKsid pöördprojekteerimist, et eraldada andmeid erinevatest CAD-programmidest, teisendada need neutraalsetesse vormingutesse, nagu OBJ või CPIXML, mis põhinevad ainsal avatud lähtekoodiga geomeetria tuumal OCCT, ning rakendada neid edasi sadades ehitus- ja projekteerimisettevõtete äriprotsessides. Hoolimata idee edumeelsusest seisavad sellised projektid siiski silmitsi vabade geomeetriliste tuumade piirangute ja keerukusega ning nad jäävad endiselt ühe ettevõtte suletud ökosüsteemi osaks, mis taastoodab monovarustaja lahenduste loogikat.

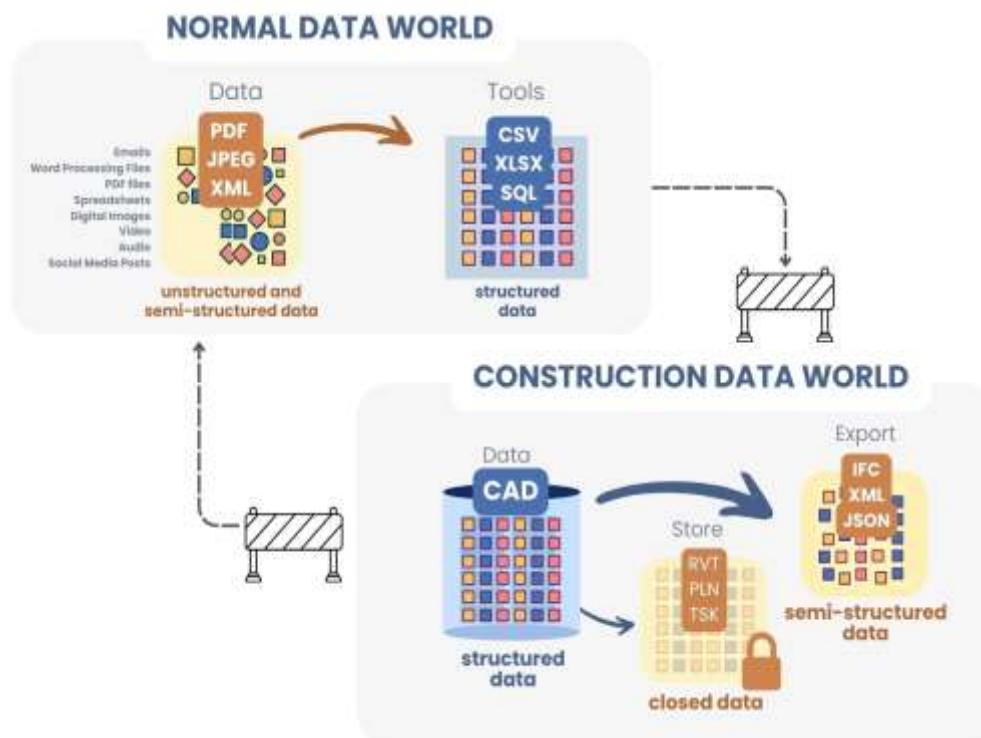
Suletud süsteemide piirangute ja andmeformaate erinevuste ning nende ühendamiseks vajalike tõhusate vahendite puudumise tõttu seisavad CAD-formaatidega töötavad ettevõtted silmitsi märkimisväärsede andmemahutude kogumisega, mille struktuur ja sulgemine on erinev. Neid andmeid ei kasutata nõuetekohaselt ja need kaovad arhiividesse, kus nad jäävad igaveseks unustatud ja kasutamata.

Andmed, mis on saadud projekteerimisfaasis märkimisväärsete jõupingutustega, muutuvad edasiseks kasutamiseks kättesaamatuks nende keerukuse ja suletud olemuse tõttu.

Selle tulemusena on ehitussektori arendajad viimase 30 aasta jooksul olnud sunnitud ikka ja jälle silmitsi seisma sama probleemiga: iga uus suletud formaat või patenteeritud lahendus tekitab vajaduse integreerida olemasolevaid avatud ja suletud CAD-süsteeme. Need pidevad katsed tagada koostalitlusvõime erinevate CAD- ja BIM-lahenduste vahel muudavad andmete ökosüsteemi ainult keerulisemaks, selle asemel et aidata kaasa selle lihtsustamisele ja standardiseerimisele.

Müüt koostalitlusvõime kohta CAD-süsteemide vahel

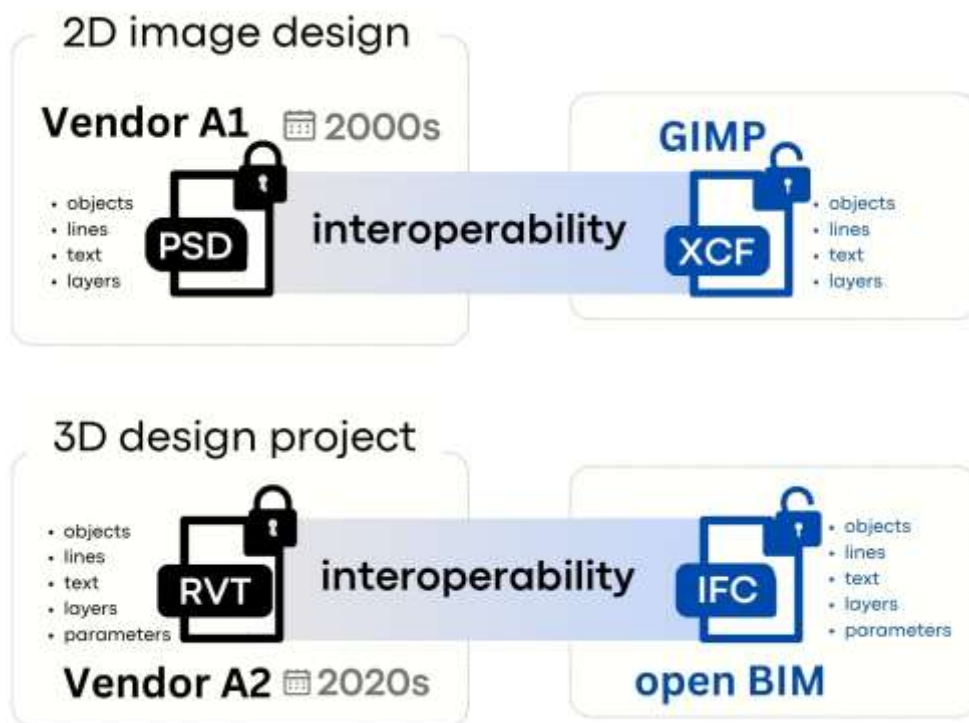
Kui 1990ndate keskel oli CAD-keskkonna koostalitlusvõime arendamise peamine suund patenteeritud DWG-formaadi murdmine - mis kulmineerus Open DWG alliansi [75] võiduga ja kogu ehitussektori kõige populaarsema joonistusformaadi tegeliku avamisega -, siis 2020ndate keskpaigaks on rõhuasetus muutunud. Ehitustööstuses on jõudmas uus trend: arvukad arendusmeeskonnad keskenduvad nn "sildade" loomisele suletud CAD-süsteemide (closed BIM), IFC-vormingu ja avatud lahenduste (open BIM) vahel. Enamik neist algatustest põhineb IFC-formaadi ja OCCT-geomeetria tuumiku kasutamisel, pakkudes tehnilist silda erinevate platvormide vahel. Seda lähenemisviisi peetakse paljulubavaks suunaks, mis võib oluliselt parandada andmevahetust ja tarkvaravahendite koostalitlusvõimet.



Joonis 6.2-2 Kui teised tööstusharud töötavad avatud andmetega, siis ehitustööstus peab töötama suletud või vabalt struktureeritud CAD-formaatidega (BIM).

Sellisel lähenemisviisil on ajaloolised paralleelid. 2000ndatel aastatel püüdsid arendajad, püüdes ületada suurima graafiliste redaktorite tootja (2D world) domineerimist, luua sujuvat integratsiooni oma patenteeritud lahenduse ja vaba avatud lähtekoodiga vahel - alternatiiv GIMPile (joonis 6.2-3). Toona, nagu ka tänapäeval konstrueerimisel, oli tegemist katsega ühendada suletud ja avatud süsteemid, säilitades samal ajal keerulised parameetrid, kihid ja sisemise tarkvaraloogika.

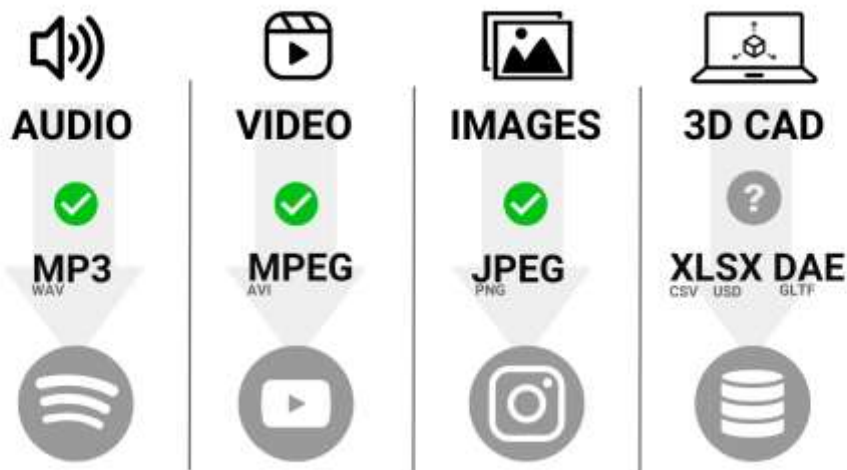
Kuid tegelikult otsisid kasutajad lihtsaid lahendusi - lamedad, avatud andmed ilma liigselt keeruliste kihtide ja programmiparameetriteta (analoogid CADi geomeetrilise tuuma). Kasutajad otsisid lihtsaid ja avatud andmeformaate, mis ei sisaldaks liigset loogikat. JPEG, PNG ja GIF said sellisteks graafikaformaatideks. Tänapäeval kasutatakse neid sotsiaalvõrgustikes, veebisaitidel, rakendustes - neid on lihtne töödelda ja tõlgendada, sõltumata platvormist või tarkvaratootjast.



Joonis 6.2-3 Andmeformaate koostalitlusvõime ehituses sarnaneb 2000. aastatel toimunud katsetega ühendada ühe populaarse tootja patenteeritud toode ja avatud lähtekoodiga GIMP.

Selle tulemusena ei kasuta peaaegu keegi pilditööstuses tänapäeval suletud formaate, nagu PSD või avatud XCF, rakenduste, sotsiaalvõrgustike, nagu Facebook ja Instagram, või veebisaitide sisuna. Selle asemel kasutatakse enamasti lamedat ja avatud JPEG-, PNG- ja GIF-formaati, mis on lihtsa kasutuse ja laialtlevinud ühilduvuse huvides. Avatud formaadid, nagu JPEG ja PNG, on muutunud pildi jagamise standardiks tänu nende mitmekülsusele ja laialdasele toetusele, mis teeb nende kasutamise erinevatel platvormidel lihtsaks. Sarnast üleminekut võib täheldada ka muude

andmevahetusformaate, näiteks video ja heli puhul, kus universaalseid formate, nagu MPEG ja MP3, rõhutatakse nende pakkimise tõhususe ja laiaulatusliku ühilduvuse tõttu. Selline liikumine standardiseerimise suunas on lihtsustanud sisu ja teabe jagamist ja taasesitamist, muutes need kõigile kasutajatele kättesaadavaks mitmetel platvormidel (joonis 6.2-4).



Joonis 6.2-4 Lihtsustatud vormingud, millel puuduvad keerulised redigeerimisfunktsioonid, on muutunud populaarseks andmete jagamiseks ja kasutamiseks.

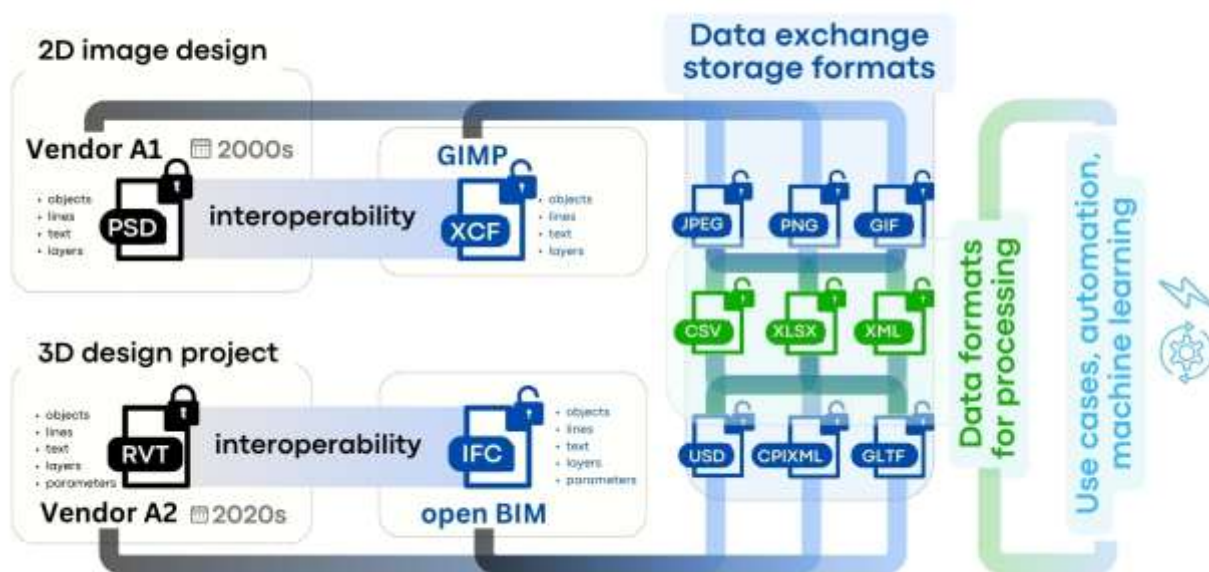
Sarnased protsessid toimuvad 3D modelleerimisel. Lihtsad ja avatud vormingud nagu USD, OBJ, glTF, DAE, DXF, SQL ja XLSX kasutatakse projektides üha enam andmevahetuseks väljaspool CAD-keskkonda (BIM). Need vormingud salvestavad kogu vajaliku teabe, sealhulgas geomeetria ja metaandmed, ilma et oleks vaja kasutada keerukat BREP-struktuuri, geomeetria tuumasid või tarnijaspetsiifilisi sisemisi klassifikaatoreid. Ka juhtivate tarkvaratootjate pakutavad omased vormingud nagu NWC, SVF, SVF2, CPIXML ja CP2 täidavad sarnaseid funktsioone, kuid jäävad, erinevalt avatud standarditest, suletud vorminguteks.

On märkimisväärne (ja tasub uuesti meelde tuletada, nagu eelmises peatükis juba mainitud), et seda ideed - neutraalsete ja parameetriliste vaheformaate, nagu IGES, STEP ja IFC, tagasilükkamist - toetas juba 2000. aastal peamine CAD tarnija, kes koostas BIM Whitepaper ja registreeris IFC-vormingu 1994. aastal. Whitepaper 2000 "Integrated Design and Manufacturing" (integreeritud projekteerimine ja tootmine). [65] rõhutab CAD-müüja, et teabe täielikkuse ja täpsuse säilitamiseks on oluline omalaadne juurdepääs CAD-andmebaasile tarkvarakeskkonnas, ilma et oleks vaja kasutada vahepealseid tõlkeid ja parameetrilisi vorminguid.

Ehitustööstus ei ole veel kokku leppinud ei CAD-andmebaasidele juurdepääsu vahendites või nende sunnitud tagasipööramises, ega ka ühise lihtsustatud andmeformaadi vastuvõtmises, mida kasutatakse väljaspool CAD-platvorme (BIM). Näiteks kasutavad paljud Kesk-Euroopa ja saksakeelsete piirkondade suured ehitussektoris tegutsevad ettevõtted oma ERP -süsteemides CPIXML formaati [121]. See patenteeritud formaat, mis on omamoodi XML, ühendab CAD (BIM)

projektandmed, sealhulgas geomeetrilised ja metaandmed, ühtsesse organiseeritud lihtsustatud struktuuri. Suured ehitusettevõtted loovad ka omaenda uusi formaate ja süsteeme, nagu SCOPE projekt, mida me arutasime eelmises peatükis

Parameetriliste CAD-vormingute või keeruliste parameetriliste failide IFC (STEP) suletud loogika on enamikus äriprotsessides üleliigne. Kasutajad otsivad lihtsustatud ja lame formaate, nagu USD, CPIXML, XML & OBJ, DXF, glTF, SQLite, DAE & XLSX, mis sisaldavad kogu vajalikku elementide teavet, kuid ei ole koormatud üleliigse BREP-geomeetria loogikaga, sõltuvusega geomeetristest tuumadest ja konkreetsete CAD ja BIM -toodete siseklassifikatsioonidega (joonis. 6.2-5).



Joonis 6.2-5 Enamiku kasutusjuhtumite puhul valivad kasutajad võimalikult lihtsad vormingud, mis ei sõltu müüja programmist.

Selliste lamedate pildivormingute nagu JPEG, PNG ja GIF, mis on vabastatud tootjate sisemootorite üleliigsest loogikast, kasutuselevõtt on soodustanud tuhandete koostalitlusvõimeliste lahenduste väljatöötamist graafika töötlemiseks ja kasutamiseks. See on viinud rakenduste tekkimiseni alates retušeerimis- ja filtreerimisvahenditest kuni sotsiaalvõrgustikeni, nagu Instagram, Snapchat ja Canva, kus neid lihtsustatud andmeid saab kasutada ilma konkreetse tarkvaraarendajaga sidumata.

Projekteerimise CAD -formaate standardiseerimine ja lihtsustamine soodustab paljude uute kasutajasõbralike ja sõltumatute töövahendite tekkimist ehitusprojektidega töötamiseks.

Üleminek suletud geomeetriliste tuumadega seotud müüja rakenduste keerukast loogikast universaalsete avatud vormingute suunas, mis põhinevad lihtsustatud elementide raamatukogudel, loob eeldused paindlikumaks, läbipaistvamaks ja tõhusamaks andmekäitluseks. See avab ka juurdepääsu teabele kõigile ehitusprotsessis osalevatele osapooltele - alates projekteerijatest kuni klientide ja hooldusteenusteni.

Sellegipoolest on väga tõenäoline, et lähiaastatel püüavad CAD tarnijad uuesti nihutada arutelu koostalitlusvõime ja juurdepääsu üle CAD-andmebaasidele. See puudutab juba praegu "uusi" kontseptsioone - nagu granulaarsed andmed, intelligentsed graafid, "föderaalsed mudelid", digitaalsed kaksikud pilvehoidlates - ning samuti tööstusliitude ja standardite loomist, mis jätkavad BIMi ja avatud BIMi teed. Vaatamata atraktiivsele terminoloogiale võivad sellised algatused taas kord muutuda vahenditeks, mis hoiavad kasutajaid omanike ökosüsteemides. Üks näide on USD (Universal Scene Description) formaadi aktiivne propageerimine "uue standardina" platvormideülese CAD (BIM) koostöö jaoks alates 2023. aastast.

Mine USD ja granuleeritud andmed m

AOUSD-liidu [124] tekkimine 2023. aastal tähistab olulist pööret ehitussektoris. Me oleme tunnistajaks uue reaalsuse algusele, mida kujundavad CAD-müüjad, ehitusandmetega tegelemisel mitme olulise muudatuse kaudu. Esimene oluline muutus puudutab CAD -andmete tajumist. Kontseptuaalse projekteerimise varajases etapis osalevad spetsialistid mõistavad üha enam, et CAD-keskkonnas projekti loomine on vaid alguspunkt. Projekteerimisprotsessi käigus loodud andmetest saab lõpuks objektide analüüsi, käitamise ja haldamise alus. See tähendab, et need peavad olema kättesaadavad ja kasutatavad ka muudes süsteemides kui traditsioonilised CAD-vahendid.

Samal ajal toimub revolutsioon juhtivate arendajate lähenemisviisis. Tööstuse juhtiv CAD- müüja, BIM-kontseptsiooni ja IFC-formaadi looja, võtab oma strateegias ootamatu pöörde. Alates 2023. aastast eemaldub ettevõtte traditsioonilisest andmete säilitamisest üksikutes failides, keskendudes tööle granuleeritud (normaliseeritud ja struktureeritud) andmetega ning minnes üle andmekeskse lähenemisviisi juurde [125].

Tootjad järgivad teiste tööstusharude ajaloolisi suundumusi: enamik kasutajaid ei vaja suletud CAD-vorminguid (sarnaselt PSD-le) ega keerulisi parameetrilisi IFC-faile (sarnaselt GIMPile koos kihiloogikaga). Nad vajavad lihtsaid objektipilte, mida saab kasutada CAFMis (ehitusinstrumendis), ERPis (Facebook) ja tuhandetes muudes protsessides, mis on täidetud Exceli tabelite ja PDF dokumentidega.

Praegused suundumused ehitustööstuses loovad potentsiaalselt eelduse järkjärguliseks eemaldumiseks parameetrilistest ja keerulistest vormingutest universaalsemate ja sõltumatute vormingute kasuks: USD, GLTF, DAE, OBJ (metainformatsiooniga nii hübriidvormingus kui ka eraldi struktureeritud või vabalt struktureeritud vormingutes). Ajaloolised liidrid, sealhulgas suured projekteerimisfirmad, kes kunagi 1990ndate keskel aktiivselt propageerisid IFC-d, propageerivad nüüd avalikult uut USD-vormingut [93], rõhutades selle lihtsust ja mitmekülgsust (joonis 6.2-6). USD massiline kasutuselevõtt toodetes, GLTF-ühilduvus ja aktiivne integreerimine sellistesse tööriistadesse nagu Blender, Unreal Engine ja Omniverse näitavad uue paradigma algust andmetega töötamisel. Koos kohalike lahendustega, nagu Euroopa ühtse USD-vormingu CPIXML, mida kasutatakse populaarsetes Euroopa ERP-s, populaarsus võib potentsiaalselt tugevdada USD positsiooni Kesk-Euroopas. IFC-vormingu arendamisega seotud organisatsioonid kohandavad juba praegu oma strateegiat USD-le [126], mis kinnitab vaid ülemineku vältimatust.

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

Joonis 6.2-6 IFC ja USD formaatide tehniliste näitajate võrdlus.

Selle taustal on USD potentsiaal saada de facto standardiks, mis lubab ületada mitmeid praeguseid piiranguid, mis on peamiselt seotud olemasolevate CAD - (BIM-) formaatide keerukusega ja nende tõlgendamise sõltuvusega geomeetrilistest tuumadest.

Parameetriliste ja keeruliste CAD -vormingute ja IFC asemel saavad lihtsustatud andmeformaadid USD, glTF, DAE, OBJ koos elementide metainformatsiooniga CSV, XLSX, JSON, XML oma lihtsuse ja paindlikkuse tõttu ehitussektoris oma koha.

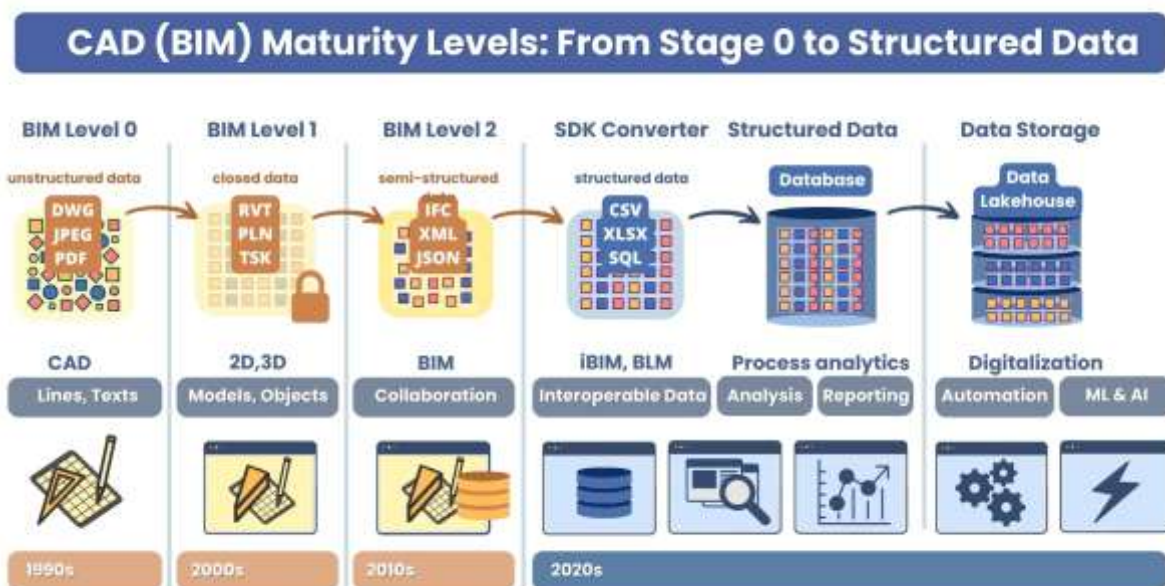
Praegused muutused ehitussektoris näivad esmapilgul nagu tehnoloogiline läbimurre, mis on seotud üleminekuga vananenud IFC-lt kaasaegsemale USA dollarile. Siiski tasub arvestada, et juba 2000.

aastal kirjutas sama CAD müüja, kes töötas välja IFC, selle probleemidest ja vajadusest pääseda ligi andmebaasile [65], ning nüüd propageerib aktiivselt üleminekut uuele standardile - USD.

Veel ühe "avatud andmete" fassaadi taga võib peituda USD ja "uued" kontseptsioonid granuleeritud andmete haldamiseks pilvepõhiste rakenduste kaudu, mida CAD müüjad hakkavad reklaamima, müüjate kavatsus monopoliseerida projektiandmete haldamine, kus kasutajad leiavad end olukorras, kus formaadi valik on rohkem seotud ettevõtte huvidega kui tegelike vajadustega.

Oluliste faktide analüüs [93] näitab, et nende muudatuste peamine eesmärk on pigem vähem kasutaja mugavus kui eelkõige kontrolli säilitamine ökosüsteemide ja andmevoogude üle tootjate huvides, kes 40 aasta jooksul ei ole kunagi suutnud pakkuda juurdepääsu CAD-andmebaasidele.

Võib-olla on nüüd ettevõtetel aeg lõpetada uute kontseptsioonide ootamine tarkvaratootjatelt ja keskenduda enda arendamisele andmekeskse suunas. Olles vabanenud andmete juurdepääsu probleemidest pöördtehnoloogiliste vahendite abil, saab tööstus iseseisvalt liikuda kaasaegsete, tasuta ja mugavate töö- ja analüüsvahendite suunas ilma uusi kontseptsioone peale surumata.



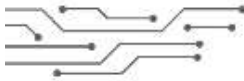
Joonis 6.2-7 CAD-küpsuse tase (BIM): struktureerimata andmetest struktureeritud andmete ja hoidlateni.

Juurdepääs andmebaasidele, avatud andmetele ja vormingutele muutub ehitussektoris paratamatult standardiks, hoolimata müüjate katsetest seda protsessi takistada - see on vaid aja küsimus (joonis 6.2-7). Selle ülemineku tempo võib oluliselt kiireneada, kui üha enam spetsialiste tutvub avatud vormingute, andmebaasivahenditega ja kättesaadavate pöördprojekteerimise SDKdega, mis võimaldavad otsest juurdepääsu CAD-andmetele -süsteemidele [92].

Tulevik peitub avatud, ühtsetes ja analüütiliselt juurdepääsetavates andmetes. Et vältida sõltuvust müüja lahendustest ja kinniste ökosüsteemide pantvangi sattumist, peavad ehitus- ja insenerettevõtted varem või hiljem toetuma avatusele ja sõltumatusele, valides vorminguid ja lahendusi, mis võimaldavad täielikku kontrolli andmete üle.

Ehitustööstuses täna loodavad andmed on tulevikus äriotsuste tegemisel võtmeressurss. Need toimivad strateegilise "kütusena", mis toidab ehitusettevõtete arengut ja tõhusust. Ehitustööstuse tulevik seisneb oskuses töötada andmetega, mitte andmeformaaside või -mudelite valikus.

Et mõista erinevust avatud vormingute USD, glTF, DAE, OBJ ja patenteeritud parameetriliste CAD vormingute vahel, on oluline käsitleda ühte kõige keerulisemat ja olulisemat andmeelementi visualiseerimisel ja projekteerimisarvutustes - geomeetriat ja selle genereerimisprotsessi. Ja selleks, et mõista, kuidas geomeetrilised andmed muutuvad ehituses analüütika ja arvutuste aluseks, on vaja uurida sügavamalt geomeetria loomise, teisendamise ja salvestamise mehhanisme.



PEATÜKK 6.3.

EHITUSGEOMEETRIA: JOONTEST KUUPMEETRITENI

Kui jooned muutuvad rahaks või miks ehitajad vajavad geomeetriat

Geomeetria ei ole ehituses mitte ainult visualiseerimine, vaid ka täpsete kvantitatiivsete arvutuste alus. Projektmodelis täiendab geomeetria elementide parameetrite loetelusid (joonis 3.1-16) oluliste ruumiliste näitajatega, nagu pikkus, pindala ja ruumala. Need mahuparameetrite väärtused arvutatakse automaatselt geomeetria tuumade abil ja need on hinnangute, ajakavade ja ressursimudelite lähtepunktiks. Nagu me oleme juba käsitlenud käesoleva raamatu 5. osas ja peatükis "Kulude arvestamine ja eelarvestamine ehitusprojektid", on just CAD -mudelitest pärit objektirühmade ruumilised parameetrid aluseks kaasaegsetele ERP, PMIS -süsteemidele/ Geomeetria mängib olulist rolli mitte ainult projekteerimisfaasis, vaid ka projekti rakendamise juhtimises, ajakava kontrollimisel, eelarvestamisel ja käitamisel. Nii nagu tuhandeid aastaid tagasi, kui Egiptuse püramiidide ehitamisel sõltus projekti täpsus pikkuse mõõtmetest, nagu küünarnukid ja kuubitsad, mõjutab tänapäeval CAD -programmide geomeetria tõlgendamise täpsus otseselt tulemust: alates eelarvest ja tähtaegadest kuni töövõtja valimise ja tarnelogistikani

Väga konkurentsitihedas ja eelarvepiirangutega keskkonnas muutub mahuarvutuste täpsus, mis sõltub otseselt geomeetrist, ellujäämisteguriks. Kaasaegsed ERP -süsteemid sõltuvad otseselt CAD - ja BIM - mudelitest saadud korrektsetest mahuomadustest. Seetõttu ei ole elementide täpne geomeetiline kirjeldus mitte ainult visualiseerimine, vaid oluline vahend kulude ja ehitusaja juhtimiseks.

Ajalooliselt on geomeetria olnud inseneriteaduse peamine suhtluskeel. Alates joontest papipüürusel kuni digitaalsete mudeliteni on joonised ja geomeetrilised kujutised olnud teabevahetuse vahendiks projekteerijate, järelevalveametnike ja kalkulaatorite vahel. Enne arvutite kasutuselevõttu tehti arvutusi käsitsi, kasutades joonlaudu ja nurgamõõtjaid. Tänapäeval on see ülesanne automatiseeritud tänu ruumilisele modelleerimisele: CAD-tarkvara geomeetrilised tuumad muudavad jooned ja punktid kolmemõõtmelisteks kehadeks, millest kõik vajalikud omadused automaatselt välja võetakse.

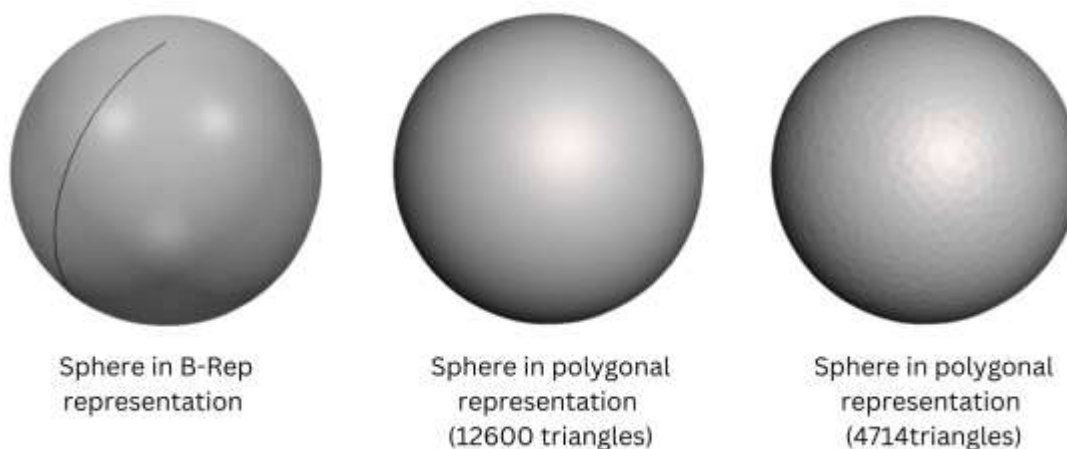
CAD- (BIM-) programmide kasutajaliidese kaudu toimub arvutuste jaoks vajalike geomeetriliste elementide loomine CAD- (BIM-) -programmides. Punktide ja joonte teisendamiseks ruumilisteks kehadeks kasutatakse geomeetrilist tuuma, mis täidab põhiülesannet - geomeetria teisendamist ruumilisteks mudeliteks, millest pärast ühtlustamist arvutatakse automaatselt elemendi ruumilised omadused.

Joontest mahtudeks: kuidas pindala ja maht muutuvad andmeteks

Inseneripraktikas arvutatakse mahud ja pindalad geomeetrilistelt pindadelt, mida kirjeldatakse analüütiliselt või parameetriliste mudelite abil, nagu NURBS (mitteühtlane ratsionaalne B-splain) BREP (boundary element representation) raamistikus.

NURBS (Non-Uniform Rational B-Splines) on matemaatiline viis kõverate ja pindade kirjeldamiseks, samas kui BREP on raamistik objekti täieliku kolmemõõtmelise geomeetria, sealhulgas selle piiride kirjeldamiseks, mida saab NURBSi abil määratleda.

Hoolimata BREPi ja NURBSi täpsusest, nõuavad nad võimsaid arvutusressursse ja keerulisi algoritme. Otsene arvutamine sellistest matemaatilistelt täpsetest kirjeldustest on aga sageli arvutuslikult raske, mistõttu praktikas kasutatakse peaaegu alati tessellatsiooni - pindade muundamist kolmnurkade ruudustikuks -, mis lihtsustab hilisemaid arvutusi. Tessellatsioon on kompleksse pinna jaotamine kolmnurkadeks või hulknurkadeks. CAD /CAE keskkondades kasutatakse seda meetodit visualiseerimiseks, ruumala arvutamiseks, kokkupõrkeotsinguks, ekspordiks sellistesse vormingutesse nagu MESH ja kokkupõrkeanalüüsiks. Näide loodusest on mesilase meekärg, kus keerukas kuju on jaotatud korrapäraseks ruudustikuks (joonis 6.3-1).



Joonis 6.3-1 Sama kera parameetrilises kirjelduses BREP ja hulknurga kujutis erineva arvu kolmnurkadega.

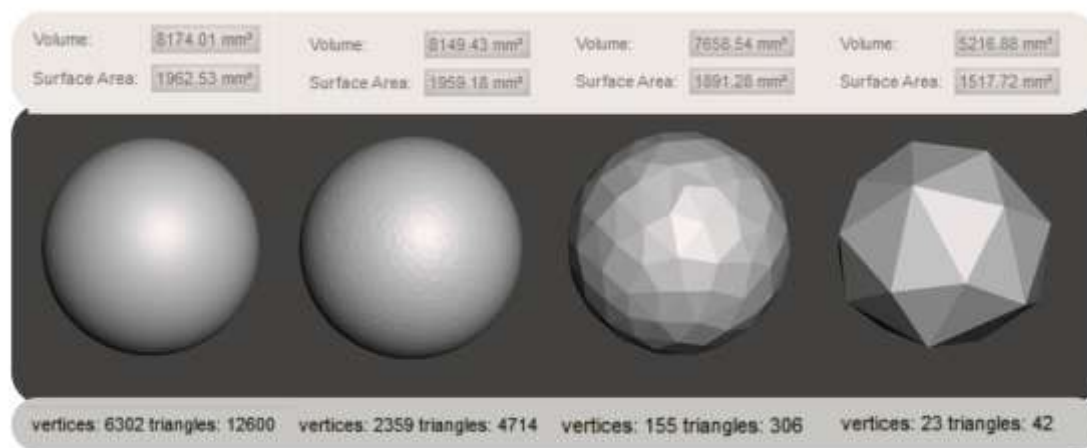
BREP (NURBS), mida kasutatakse CADis, ei ole geomeetria põhimudel. See loodi mugava vahendina ringide ja ratsionaalsete splainide kujutamiseks ning geomeetriliste andmete salvestamise minimeerimiseks. Sellel on siiski piirangud - näiteks suutmatust kirjeldada täpselt spiraaljoonte ja pindade aluseks olevat sinusoidi ning vajadus kasutada keerulisi geomeetrilisi tuumasid.

Seevastu kolmnurkvõrke ja parameetriliste kujundite tessellatsiooni iseloomustab lihtsus, tõhus mälukasutus ja võime töödelda suuri andmehulki (joonis 6.3-2). Need eelised võimaldavad loobuda geomeetriliste kujundite arvutamisel keerulistest ja kallistest geomeetrilistest tuumadest ja neisse põimitud kümnetest miljonitest koodiridadest.

Enamikul juhtudel ei ole oluline, kuidas täpselt on ruumilised omadused määratletud - kas parameetriliste mudelite (BREP, IFC) või hulknurkade (USD, glTF, DAE, OBJ) kaudu. Geomeetria jääb lähendavaks vormiks: kas NURBS või MESH kaudu, tegemist on alati kuju ligikaudse kirjeldusega.

Polügoonidena või BREP (NURBS) määratletud geomeetria jääb teataval määral ainult viisiks, kuidas ligikaudselt kirjeldada pidevat kuju. Nii nagu Fresneli integraalidel ei ole täpset analüütilist väljendit, on ka geomeetria diskretiseerimine hulknurkade või NURBSi abil alati ligikaudne, nagu ka kolmnurkne MESH.

Parameetiline geomeetria BREP formaadis on vajalik peamiselt siis, kui oluline on andmete minimaalne suurus ning selle töötlemiseks ja kuvamiseks on võimalik kasutada ressursimahukaid ja kalleid geomeetrilisi tuumasid. Enamasti on see iseloomulik CAD -programmide arendajatele, kes kasutavad selleks oma toodetes MCAD -pakkujate geomeetrilisi kerneleid. Sellisel juhul muudetakse BREP-mudelid isegi nendes programmides tessellatsiooni käigus visualiseerimiseks ja arvutusteks sageli kolmnurkadeks (sarnaselt viisiga, kuidas PSD-faile lihtsustatakse JPEG-formaadis).



Joonis 6.3-2 Erineva hulknurkade arvuga jooniste mahukarakteristikute erinevus.

Polügonaausel MESHil, samuti parameetrilisel BREPil, on omad eelised ja piirangud, kuid eesmärk on sama - kirjeldada geomeetriat, võttes arvesse kasutaja ülesandeid. Lõppkokkuvõttes ei sõltu geomeetrilise mudeli täpsus mitte ainult selle kujutamise meetodist, vaid ka konkreetse ülesande nõuetest.

Enamiku konstruktsiooniprobleemide puhul võib parameetrilise geomeetria ja keeruliste geomeetriliste tuumade vajadus olla üleliigne.

Iga konkreetse arvutuste automatiseerimise ülesande puhul tasub kaaluda, kas CAD-arendajad, kes on huvitatud oma tarkvaratoodete reklaamimisest ja müügist, liialdavad parameetrilise geomeetria tähtsust.

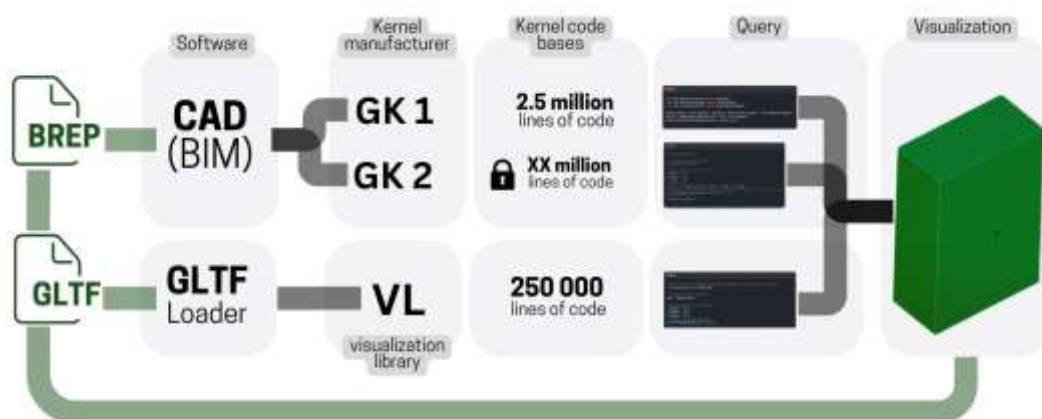
Üleminek MESH, USD ja hulknurksed: geomeetria tessellatsiooni kasutamine

Ehitustööstuses on projekteerimisteabe ja geomeetriaga töötavate süsteemide, andmebaaside või protsesside automatiseerimisel oluline püüelda sõltumatuse poole konkreetsetest CAD-redaktoritest ja geomeetrilistest tuumadest.

Nii arvutusosakondades kui ka ehitusplatsil kasutatav andmevahetusformaad ei tohiks põhineda konkreetsetel CAD- (BIM-) programmil. Geomeetiline teave tuleks esitada formaadis otse tessellatsiooni abil, viitamata geomeetrilisele tuumale või CAD-arhitektuurile.

Parameetrilist geomeetria CADist võib pidada vahepealseks allikaks, kuid mitte universaalse formaadi aluseks. Enamik parameetrilisi kirjeldusi (sealhulgas BREP ja NURBS) teisendatakse edasiseks töötlemiseks igal juhul polügonaalseteks MESH. Kui tulemus on sama (tessellatsioon ja hulknurksus) ja protsess on lihtsam, on valik ilmselge. See on sarnane valikuga graafontoloogiatega ja struktureeritud tabelite vahel (mida me arutasime neljandas osas): liigne keerukus on harva põhjendatud (joonis 3.2-10, joonis 6.1-8).

Avatud formaadid, nagu OBJ, STL, glTF, SVF, CPIXML, USD ja DAE, kasutavad universaalset kolmnurkse võrgu struktuuri, mis annab neile märkimisväärsed eeliseid. Need vormingud on suurepärase koostalitlusvõimega - neid on lihtne lugeda ja visualiseerida, kasutades olemasolevaid avatud lähtekoodiga raamatukogusid, ilma et oleks vaja keerulisi spetsiaalseid geomeetria tuumasid, mis sisaldavad miljardeid koodiridu (joonis 6.3-3). Neid mitmekülgseid geomeetriaformaate kasutatakse mitmesugustes rakendustes alates suhteliselt lihtsatest kõõgi projekteerimisvahenditest IKEA™-s kuni keerukate objektide visualiseerimissüsteemideni kino- ja VR-rakendustes -rakendustes. Oluline eelis on see, et nende formaatidega töötamiseks on olemas suur hulk tasuta ja avatud lähtekoodiga raamatukogusid, mis on kättesaadavad enamiku platvormide ja programmeerimiskeelte jaoks.



Joonis 6.3-3 Sama geometria kujutamine saavutatakse parameetriliste vormingute ja geometriliste tuumade või trianguleeritud vormingute ja avatud lähtekoodiga visualiseerimisraamatukogude abil.

Nii nagu kasutajad ise, seisavad ka CAD -pakkujad silmitsi probleemidega välismaa parameetriliste CAD-vormingute või avatud IFC tõlgendamisel, sest nende geometrilised tuumad on erinevad. Praktikas kasutavad eranditult kõik CAD -pakkujad süsteemide vaheliseks andmeedastuseks pöördprojekteerimise SDK ning ükski neist ei toetu koostalitlusvõime eesmärgil sellistele formaatidele nagu IFC või USD [93].

Selle asemel, et kasutada CAD- müüjate liitude poolt propageeritud kontseptsioone, mida nad ise ei kasuta, on CAD-lahenduste arendajate ja kasutajate jaoks produktiivsem keskenduda iga lähenemisviisi eeliste mõistmisele konkreetses kontekstis ja valida üks või teine geometriatüüp sõltuvalt kasutusjuhust. Erinevate geometriliste esitusviiside vahel valimine on kompromiss täpsuse, arvutustõhususe ja konkreetse ülesande praktiliste vajaduste vahel.

Geomeetriliste tuumade kasutamisega seotud keerukus, mille suured tarnijad on projekteerimisandmete töötlemisel ehitustööstusele traditsiooniliselt peale sundinud, osutub sageli üleliigseks. MESH-geomeetrial põhinev USD-vorming võib saada tööstusele omamoodi "Pandora laekaks", mis avab arendajatele uusi võimalusi andmevahetuse korraldamiseks - väljaspool IFC ja CAD-müüjatele tüüpiliste parameetriliste BREP-struktuuride raamistikku.

Pärast USD, DAE, gLTF, OBJ jne. struktuuri lähemal vaatlemisel selgub, et on olemas lihtsamad, avatud formaadid, mis võimaldavad geomeetrilise teabe edastamist ja kasutamist tõhusalt korraldada, ilma et oleks vaja tugineda keerulisele parameetrikale ja suletud geomeetrilistele tuumadele. Selline lähenemisviis mitte ainult ei alanda arendajate jaoks tehnilist sisenemislävendit, vaid soodustab ka paindlike, skaleeritavate ja tõeliselt avatud lahenduste väljatöötamist digitaalehituse jaoks.

LOD, LOI, LOMD - detailide unikaalne liigitus CADis (BIM).

Lisaks geomeetrilistele esitusformaatidele pakuvad CAD - (BIM-) meetodikad maailmas, kus eri

tööstusharud kasutavad eri detailsuse ja andmete sügavuse tasemeid, oma unikaalseid klassifitseerimissüsteeme, mis struktureerivad lähenemist ehitusmodelite teavitamisele.

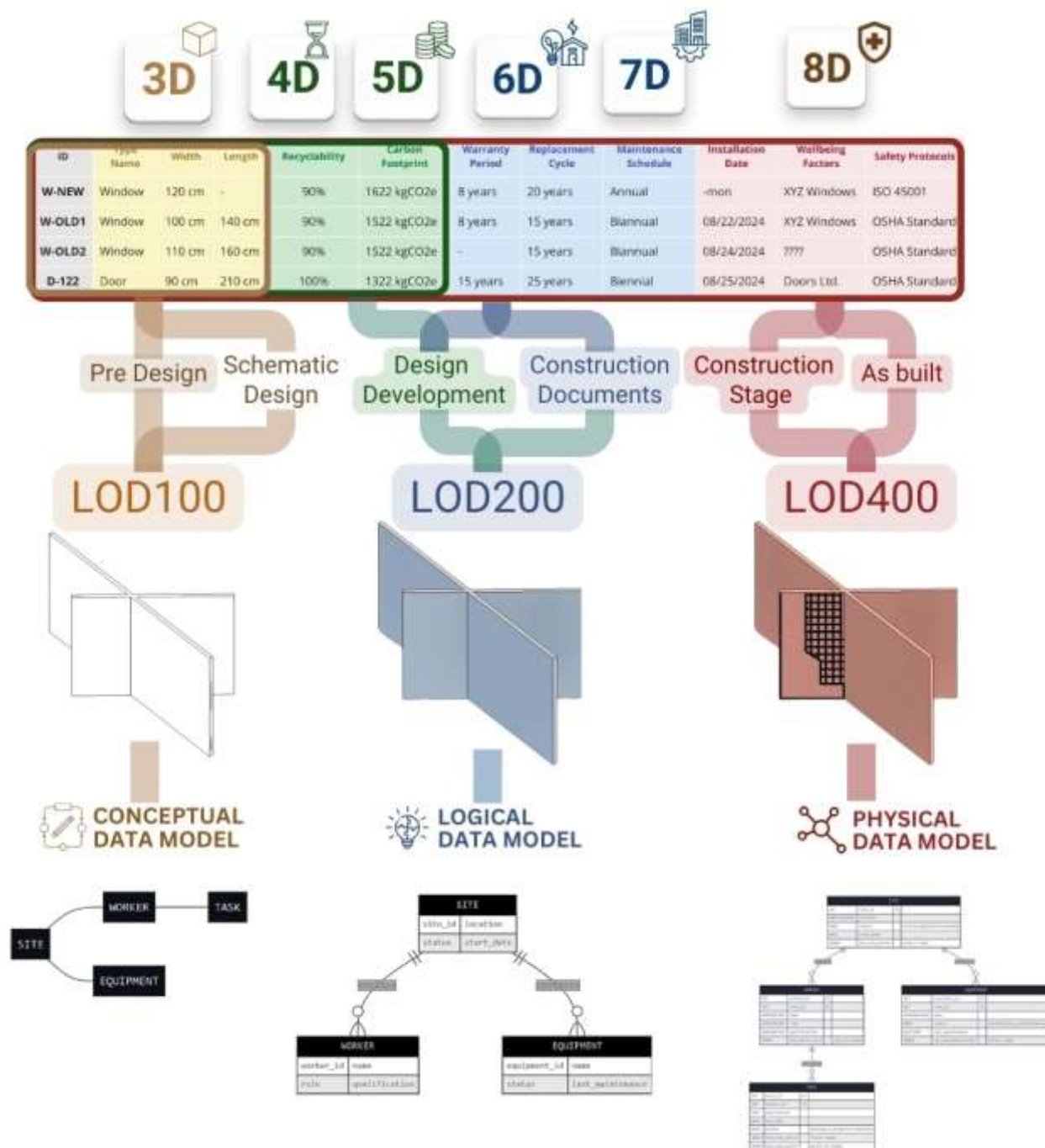
Üks näide uutest standardiseerimise lähenemisviisidest on mudeliarenduse tasemete kehtestamine, mis kajastab nii graafiliste kui ka teabekomponentide valmisoleku ja usaldusväärsuse astet. Teabesisu eristamiseks töös CAD - (BIM-) andmetega ilmus LOD (Level Of Detail) - mudeli graafilise osa detailsuse tase ja LOI (Level Of Information) - andmete väljatöötamise tase. Lisaks võeti integreeritud lähenemisviisi jaoks kasutusele mõiste LOA (Level of Accuracy) - esitatud elementide täpsus ja LOG (Level of Geometry) graafilise kujutamise täpsuse määramiseks.

Detailsusastmed (LOD) on märgitud numbritega 100 kuni 500, mis kajastavad mudeli arenguastet. LOD 100 on kontseptuaalne mudel, millel on üldised kujundid ja mõõtmed. LOD 200 sisaldab täpsemaid mõõtmeid ja kujundeid, kuid tingimusliku detailsusega. LOD 300 on üksikasjalik mudel, millel on täpsed mõõtmed, kujud ja elementide asukohad. LOD 400 sisaldab üksikasjalikku teavet, mis on vajalik elementide valmistamiseks ja paigaldamiseks. LOD 500 kajastab rajatise tegelikku seisundit pärast ehitamist ning seda kasutatakse käitamiseks ja hoolduseks. Need tasemed kirjeldavad CAD (BIM) mudeli infoküllastuse struktuuri elutsükli eri etappides, sealhulgas 3D, 4D, 5D ja kaugemalgi.

Reaalsetes projektides on kõrge detailsusaste (LOD400) sageli liigne ja piisab LOD100 geomeetriast või isegi tasapinnalistest joonistest, samas kui ülejäänud andmed võib saada kas arvutuslikult või seotud elementidest, millel ei pruugi olla eraldi geomeetriat. Näiteks ruumidel ja ruumielementidel (ruumielementide kategooriad) ei pruugi olla visuaalset geomeetriat, kuid need sisaldavad märkimisväärset hulka teavet ja andmebaase, mille ümber paljud äriprotsessid on üles ehitatud.

Seetõttu on oluline enne projekteerimise alustamist selgelt määratleda nõutav detailsus. 4D -7D kasutusjuhtumite puhul piisab sageli isegi DWG joonistest ja minimaalsest LOD100 geomeetriast. Nõuete koostamise protsessi peamine ülesanne on leida tasakaal mudeli rikkalikkuse ja praktilisuse vahel.

Sisuliselt, kui me käsitleme CAD (BIM) andmeid kui andmebaasi (mida nad ka on), ei ole mudeli küllastamise kirjeldus uute akronüümide kaudu midagi muud kui infosüsteemide andmete samm-sammuline modelleerimine, alustades kontseptuaalsest tasandist ja lõpetades füüsilise tasandiga (joonis 6.3-4), mida käsitleti üksikasjalikult raamatu kolmandas ja neljandas osas. Iga LOD ja LOI suurendamine tähendab uute ülesannete jaoks vajaliku teabe lisamist: arvutused, ehitusjuhtimine, käitamine ja seda iseloomustab mudeli järjestikune rikastamine täiendavate teabekihtidega (3D -8D) erinevate parameetrite kujul, mida käsitleti raamatu viiendas osas.



Joonis 6.3-4 Projekti detailide täiendamise protsess on identne andmete modelleerimisega kontseptuaalsest andmemudelilt füüsilise andmemudelini.

Geomeetria on vaid osa projekteerimisandmetest, mille vajadus ei ole ehitusprojektide puhul alati põhjendatud, ning CAD- () andmetega töötamise põhiküsimus ei ole mitte niivõrd selles, kuidas mudeleid visualiseeritakse, vaid pigem selles, kuidas nende mudelite andmeid saab kasutada väljaspool CAD- (BIM -) programme.

2000. aastate keskpaigaks seisis ehitussektor enneolematu väljakutse ees, sest juhtimis- ja

andmetöötlussüsteemides olevate andmete hulk kasvas kiiresti, eriti CAD-osakondadest (BIM). See andmemahu järsk suurenemine tabas ettevõtete juhte ootamatult ja nad ei olnud valmis kasvavateks nõudmisteks andmete kvaliteedile ja haldamisele.

Uued CAD standardid (BIM) - AIA, BEP, IDS, LOD, COBie

Kasutades ära avatud juurdepääsu puudumist CAD-andmebaasidele ja piiratud konkurentsi andmetöötlusturul ning kasutades uue akronüümiga BIM seotud turunduskampaaniaid, on CAD-andmetega tegelemisega tegelevad organisatsioonid hakanud looma uusi standardeid ja kontseptsioone, mis de jure peaksid olema suunatud andmehalduse tavade parandamisele.

Kuigi peaaegu kõik CAD-müüjate ja arendajate (BIM) poolt otseselt või kaudselt toetatud algatused on olnud suunatud tööprotsesside optimeerimisele, on nende tulemuseks olnud hulgaliselt standardeid, mida erinevad sidusrühmad on lobitsenud, mis on ehitussektoris tekitanud mõningast ebaselgust ja segadust andmeprotsesside osas.

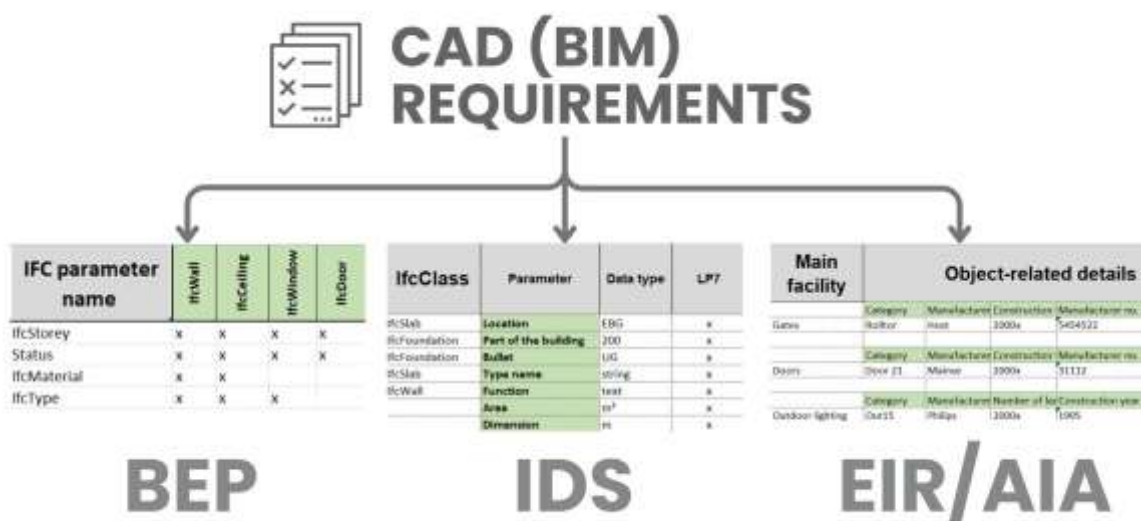
Loetleme mõned uued andmestandardid, lisaks LOD, LOI, LOA, LOG, mis on viimastel aastatel ehitussektoris esile kerkinud:

- **BEP** (BIM Execution Plan) - kirjeldab, kuidas integreerida ja kasutada CAD (BIM) projektis, määratledes andmekäitlusmeetodid ja -protsessid.
- **EIR dokument /AIA** (tellija infonõuded) - koostab tellija enne pakkumismenetluse väljakuulutamist ja sisaldab nõudeid töövõtjale teabe koostamiseks ja esitamiseks. See on vastava projekti BEPi aluseks.
- **AIM** (varade infomudel) on osa BIM-protsessist. Kui projekt on tarnitud ja lõpetatud, nimetatakse andmemudelit varade infomudeliks või AIMiks. AIMi eesmärk on hallata, hooldada ja käitada realiseeritud vara.
- **IDS** (Information Delivery Specification) - määratleb nõuded ning selle, milliseid andmeid ja millises formaadis on ehitusprojekti eri etappides vaja.
- **iLOD** on detailsuse tase LOD, millega teave on esitatud BIM -mudelis. See määratleb, kui üksikasjalik ja täielik on mudelis sisalduv teave, alates geomeetrilistest põhiesitlustest kuni üksikasjalike spetsifikatsioonide ja andmeni välja.
- **eLOD** - LOD CAD-mudeli üksikute elementide detailsusaste (BIM). See määratleb, millises ulatuses on iga element modelleeritud, ning sellega seotud teavet, nagu mõõtmed, materjalid, toimivusomadused ja muud asjakohased omadused.
- **APS** (Platform Services) ja muud tooted suurematelt CAD-tootjatelt (BIM) - kirjeldab vahendeid ja infrastruktuuri, mis on vajalikud seotud ja avatud andmemudelite loomiseks.

Kuigi CAD (BIM) standardite - nagu LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - rakendamise deklareeritud eesmärk on parandada andmehalduse kvaliteeti ja laiendada automatiseerimisvõimalusi, viib nende kasutamine praktikas sageli protsesside liigse keerukuse ja

killustatuseni. Kui käsitleda CAD (BIM) mudelit kui omamoodi andmebaasi, siis selgub, et paljud neist standarditest dubleerivad teistes tööstusharudes infosüsteemidega töötamisel juba ammu kasutusel olevaid ja tõhusaid lähenemisi. Lihtsustamise ja ühtlustamise asemel tekitavad sellised algatused sageli täiendavat terminoloogilist koormust ja takistavad tõeliselt avatud ja paindlike lahenduste rakendamist.

Paljud neist uutest kontseptsioonidest asendavad tegelikult modelleerimis- ja andmete valideerimisprotsesse, mida käsitleti üksikasjalikult raamatu esimestes osades ja mida on teistes majandussektorites juba ammu kasutatud. Ehituses seevastu liigub standardimisprotsess sageli vastupidises suunas - luuakse uusi andmekirjeldusformaate, uusi standardeid ja uusi andmete valideerimise kontseptsioone, mis ei vii alati tegeliku ühtsuse ja praktilise rakendatavuse saavutamiseni. Selle tulemusel seisab tööstusharu töötlemise lihtsustamise ja automatiseerimise asemel silmitsi täiendava regulatsiooni ja bürokraatiaga (joonis 6.3-1), mis ei ole alati tõhususe suurendamiseks soodne.



Joonis 6.3-1 Andme- ja teabesisu nõuded taandatakse tabelite abil kirjeldatud atribuutide ja nende piirväärtuste kirjeldusele.

Selle asemel, et lihtsustada andmetöötlust, tekitavad uued CAD (BIM) andmetega seotud mõisted sageli täiendavaid keerukusi ja vaidlusi juba tõlgendamise ja põhimääratluste etapis.

Üks viimaseid näiteid uutest kontseptsioonidest on IDS-vorming (kasutusele võetud 2020. aastal), mis võimaldab kirjeldada nõudeid infomudeli atribuutide koostisele avatud BIM-kontseptsioonis. IDS-i nõuded kirjeldavad teavet atribuutide ja nende piirväärtuste kohta struktureeritud tabeli kujul (Excel või MySQL), mis seejärel tõlgitakse poolstruktureeritud XML-formaadi märgistuseks, mis on ümber nimetatud XML-i spetsiaalseks lühendiks IDS.

Andmete kogumise, ettevalmistamise ja analüüsimise meetodid ja vahendid ehituses ei tohiks põhimõtteliselt erineda teiste majandussektorite spetsialistide poolt kasutatavatest meetoditest ja vahenditest.

Tööstus on välja töötanud omaette terminoloogilise ökosüsteemi, mis nõuab kriitilist järelemõtlemist ja ümberhindamist:

- STEP-formaat on paigutatud uue nime all IFC, mida on täiendatud konstruktsiooni kategoriseerimisega, arvestamata STEP-formaadi enda piiranguid.
- Parameetrilist formaati IFC kasutatakse andmesideprotsessides, vaatamata visualiseerimiseks ja arvutamiseks vajaliku ühtse geomeetrilise tuuma puudumisele.
- Juurdepääsu andmebaasidele CAD - süsteemidele reklaamitakse termini "BIM " all, ilma et arutataks nende andmebaaside spetsiifikat ja juurdepääsu neile.
- Tarnijad propageerivad koostalitlusvõimet IFC ja USD formaatide kaudu, sageli ilma neid praktikas rakendamata, kasutades kulukat pöördprojekteerimist, millega nad ise on vaeva näinud.
- Terminid LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD on üldkasutatavad samade olendiparameetrite kirjeldamiseks, viitamata teistes tööstusharudes juba ammu kasutatavatele modelleerimis- ja kontrollivahenditele.

Ehitustööstus näitab, et kuigi see kõlab kummaliselt, on kõik eelnimetatu ehitustööstuses võimalik - eriti kui peamine eesmärk on teenida raha igas andmetöötlustapis spetsialiseeritud teenuste ja tarkvara müügi kaudu. Ärilisest vaatepunktist ei ole selles midagi halba. Siiski jääb lahtiseks küsimuseks, kas sellised akronüümid ja CAD-iga (BIM) seotud lähenemisviisid tõesti lisavad väärtust ja lihtsustavad erialaseid protsesse.

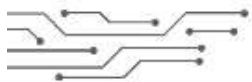
Ehitustööstuses toimib selline süsteem, sest tööstusharu ise teenib suurema osa oma spekulatiivsest kasumist selles süsteemide ja akronüümide labüündis. Läbipaistvatest protsessidest ja avatud andmetest huvitatud ettevõtted on haruldased. Selline keeruline olukord jätkub tõenäoliselt lõputult - kuni kliendid, kliendid, investorid, pangad ja erakapitali investeerimisfondid hakkavad nõudma selgemaid ja teadlikumaid lähenemisviise teabehaldusele.

Tööstuses on kogunenud liigne arv akronüüme, kuid need kõik kirjeldavad erineval määral samu protsesse ja andmenõudeid. Nende tegelik kasulikkus töövoogude lihtsustamisel jääb küsitavaks.

Kuigi mõisted ja turunduslikud akronüümid tulevad ja lähevad, jäävad andmenõuete valideerimisprotsessid ise igaveseks äriprotsesside lahutamatuks osaks. Selle asemel, et luua üha rohkem spetsialiseeritud formaate ja eeskirju, peaks ehitussektor kasutama vahendeid, mis on juba tõestanud oma tõhusust teistes valdkondades, näiteks rahandus, tööstus ja IT.

Terminite, akronüümide ja formaatide rohkus loob illusiooni põhjalikult läbimõeldud digitaalsetest ehitusprotsessidest. Turunduskontseptsioonide ja keerulise terminoloogia taga peitub aga sageli lihtne, kuid ebamugav tõde: andmed on endiselt raskesti kättesaadavad, halvasti dokumenteeritud ja jäigalt seotud konkreetsete tarkvaralahendustega.

Selleks, et väljuda sellest akronüümide ja formaatide nõiaringist formaatide pärast, on vaja vaadata CAD (BIM) süsteeme mitte kui maagilisi teabehaldusvahendeid, vaid kui seda, mis nad tegelikult on - spetsialiseeritud andmebaasid. Just selle prisma kaudu saab mõista, kus lõpeb turundus ja algab tegelik töö teabega.



PEATÜKK 6.4.

PROJEKTEERIMISPARAMEETRITE MÄÄRAMINE JA KASUTAMINE CAD-TÖÖDE JAOKS.

CAD-andmete unikaalsuse illusioon (BIM): tee analüütika ja avatud vormingute poole

Kaasaegsed CAD (BIM) platvormid on oluliselt muutnud lähenemist projekteerimis- ja ehitusinfo haldamisele. Kui varem kasutati neid vahendeid peamiselt jooniste ja 3D-mudelite loomiseks, siis tänapäeval toimivad need täieõiguslikena projektandmete hoidlatena. Ühtse tõeallika kontseptsiooni kohaselt muutub parameetiline mudel üha enam peamiseks ja sageli ka ainsaks projektiteabe allikaks, tagades selle terviklikkuse ja asjakohasuse kogu projekti elutsükli jooksul.

Peamine erinevus CAD - (BIM -) platvormide ja muude ehitusandmete haldamise süsteemide vahel on vajadus spetsiaalsete tööriistade ja APIde järele, et pääseda ligi teabele (ainus tõeallikas). Need andmebaasid ei ole traditsioonilises mõttes universaalsed: avatud struktuuri ja paindliku integreeritavuse asemel on need suletud keskkond, mis on kindlalt seotud konkreetse platvormi ja formaadiga.

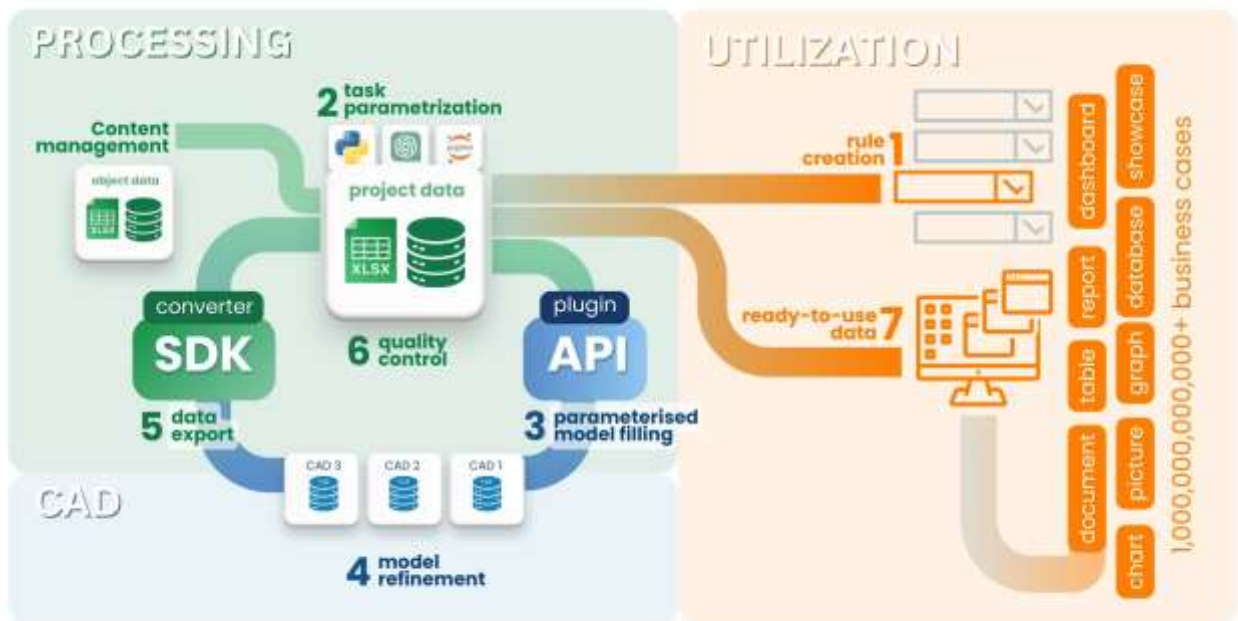
Hoolimata CAD -andmetega töötamise keerukusest, on oluline küsimus, mis läheb tehnilisest teostusest kaugemale: mis on CAD-andmebaasid (BIM) tegelikult? Sellele küsimusele vastamiseks on vaja minna kaugemale tavapärastest akronüümidest ja mõistetest, mida tarkvaraarendajad peale suruvad. Selle asemel tasub keskenduda projektiteabega töötamise põhiolemusele: andmetele ja nende töötlemisele.

Ehituse äriprotsess ei alga mitte tööga CAD - või BIM - vahenditega, vaid projekti nõuete ja andmete modelleerimisega. Kõigepealt määratletakse ülesande parameetrid: üksuste loetelu, nende algused omadused ja piirväärtused, mida tuleb konkreetse ülesande lahendamisel arvesse võtta. Alles seejärel luuakse CAD (BIM) süsteemides mudelid ja elemendid määratud parameetrite alusel

Protsess, mis eelneb teabe loomisele CAD - (BIM-) andmebaasides, on täielikult sama, mis andmete modelleerimise protsess, mida käsitleti üksikasjalikult raamatu neljandas osas ja peatükis "Andmete modelleerimine: kontseptuaalne, loogiline ja füüsiline mudel" (joonis 4.3-1).

Nii nagu andmete modelleerimisel loome nõuded andmetele, mida me hiljem andmebaasis töödelda tahame, koostavad CAD-andmebaaside juhid projekteerimisnõudeid mitme tabeli veergude või võtmeväärtuspaaride loetelude kujul (joonis 6.4-1, sammud 1-2). Ja ainult nende algsete parameetrite alusel, kasutades API automaatselt või käsitsi, loob (või pigem täiendab) projekteerija CAD- (BIM) andmebaasides objekte (sammud 3-4), mille järel kontrollitakse uuesti nende vastavust algsetele nõuetele (sammud 5-6). Seda protsessi - määratlemine → loomine → valideerimine →

korrigeerimine (sammud 2-6) - korratakse iteratiivselt, kuni andmete kvaliteet jõuab sarnaselt andmete modelleerimisele soovitud tasemele, mis on ette nähtud sihtsüsteemi jaoks - dokumendid, tabelid või armatuurlauad (samm 7).

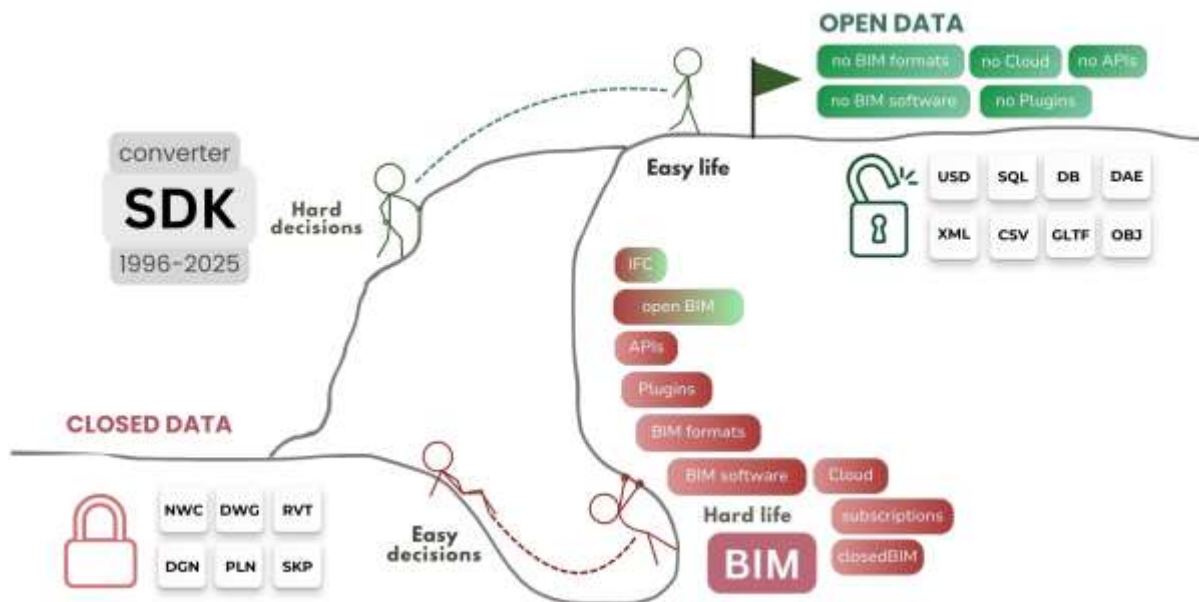


Joonis 6.4-1 Ehitusprojektide elluviimise äriprotsesside andmebaaside infoküllastuse tsükkel.

Kui me käsitleme CAD (BIM) kui parameetrite edastamise mehhanismi võtmeväärtuspaaride kogumi kujul, mis on loodud väljaspool projekteerimiskeskonda määratletud nõuete alusel (joonis 6.4-1, sammud 1-2), siis nihkub arutelu fookus konkreetsetest tarkvaralahendustest ja nende piirangutest põhimõttelisematele aspektidele - andmete struktuurile, andmemudelitele ja andmenõuetele. Sisuliselt räägime andmebaasi parameetrite küllastamisest ja klassikalisest andmemodelleerimisprotsessist (sammud 2-3 ja 5-6). Ainus erinevus seisneb selles, et CAD-andmebaaside suletud olemuse ja kasutatavate vormingute eripära tõttu kaasneb see protsess spetsiaalsete BIM-tööriistade kasutamisega. Tekib küsimus: milles seisneb BIMi ainulaadsus, kui teistes tööstusharudes ei ole sarnaseid lähenemisviise?

Viimase 20 aasta jooksul on BIM olnud rohkem kui lihtsalt üks andmeallikas. CAD -BIM paketti turustatakse sageli kui parameetrilist vahendit, millel on integreeritud andmebaas [64] ja mis on võimeline automatiseerima ehitusprojektide projekteerimise, modelleerimise ja elutsükli juhtimise protsesse. Tegelikult on BIM siiski muutunud pigem vahendiks, mis hoiab kasutajaid müüja platvormil, kui mugavaks andmete ja protsesside haldamise meetodiks.

Selle tulemusena on CAD- (BIM-) andmed nende platvormides isoleeritud, varjates projektiteavet oma APIde ja geomeetriliste tuumade taha. See on võtnud kasutajatelt võimaluse iseseisvalt ligi pääseda andmebaasidele ning andmeid ekstraheerida, analüüsida, automatiseerida ja edastada teistesse süsteemidesse, vältides müüjate ökosüsteeme.



Joonis 6.4-2 Ehituses nõuavad kaasaegsed vormingud keerukaid geomeetrilisi tuumasid, igal aastal uuendatud API ja spetsiaalseid litsentse CAD -(BIM -) tarkvara jaoks.

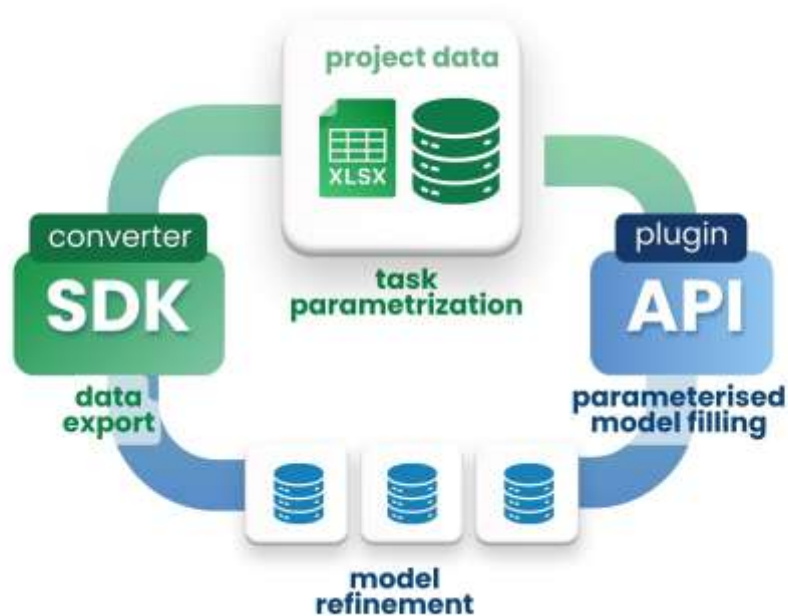
Kaasaegsete CAD-vahenditega töötavad ettevõtted peaksid kasutama sama lähenemisviisi andmetega töötamiseks, mida kõik CAD-vahendimüüjad ise eranditult praktikas kasutavad: andmete teisendamine SDK - pöördprojekteerimise vahendite abil, mille vastu CAD-vahendimüüjad on alates 1995. aastast võidelnud [75]. Omades täielikku juurdepääsu CAD-andmebaasile ja kasutades pöördprojekteerimisvahendeid, saame [127] lamedate objektide kogumi koos atribuutidega ja eksportida neid mis tahes sobivasse avatud formaati (joonis 6.4-2), sealhulgas nii geomeetria kui ka projekteerimiselementide parameetrid. Selline lähenemisviis muudab põhimõtteliselt teabega töötamise paradigmat - failikeskse arhitektuuri asemel kasutatakse andmekeskseid struktuure:

- Sellised andmeformaadid nagu RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET ja teised sisaldavad identset teavet sama projekti elementide kohta. See tähendab, et konkreetse formaadi ja selle skeemi tundmine ei tohiks olla takistuseks andmetega töötamisel.
- Mis tahes formaadis andmeid saab kombineerida üheks avatud struktureeritud ja granuleeritud struktuuriks (joonis 9.1-10), mis sisaldab MESH-i kolmnurkse geomeetria ja kõigi objektide omadusi, ilma geomeetriliste tuumade piiranguteta.
- Andmeanalüütika püüab saavutada universaalsust: avatud andmeid kasutades saab töötada projektiandmetega sõltumata kasutatavast formaadist.
- Minimeerimine ning sõltuvus APIdest ja müüja pistikprogrammidest: töö andmetega ei sõltu enam API-oskustest.

Kui ja CAD -andmete nõuded on teisendatud kergesti analüüsitavateks struktureeritud esitusformaatideks, ei sõltu arendajad enam konkreetsetest andmeskeemidest ja suletud ökosüsteemidest.

Projekteerimine läbi parameetrite: CADi tulevik ja BIM

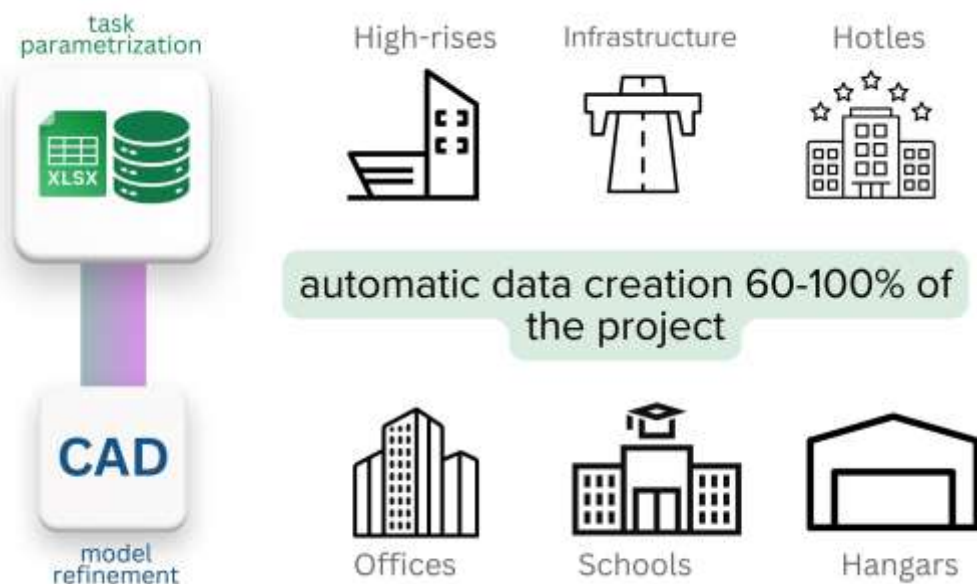
Ükski ehitusprojekt maailmas ei ole kunagi alanud CAD-programmiga. Enne kui joonis või mudel võtab CADis kuju, läbib see kontseptsiooni loomise etapi (joonis 6.4-1, etapid 1-2), kus keskendutakse parameetritele, mis määratlevad tulevase objekti põhiidee ja loogika. See etapp vastab andmete modelleerimise kontseptuaalsele tasandile (joonis 4.3-6). Parameetrid võivad eksisteerida ainult projekteerija mõtetes, kuid ideaalis on need paigutatud struktureeritud nimekirjade, tabelite või andmebaaside kujul (joonis 6.4-3), mis võimaldab läbipaistvust, reprodutseeritavust ja projekteerimisprotsessi edasist automatiseerimist.



Joonis 6.4-3 Konstrueerimisprotsess on iteratiivne protsess, mille käigus täidetakse CAD-andmebaasi väljastpoolt saadud teabega, kasutades nõudeid väärtusahelas.

Enne CAD-modelleerimise enda alustamist (andmete modelleerimise loogiline ja füüsiline faas (joonis 4.3-7)) on oluline määratleda projekti aluseks olevad piirparameetrid. Need atribuudid, nagu ka muud nõuded, kogutakse andmete kasutamise ahela (nt süsteemide) lõpust ning nende kaudu on juba määratletud projekti tulevaste objektide piirangud, eesmärgid ja põhilised omadused.

Modelleerimine ise võib olla täielikult automatiseeritud 60-100% ulatuses parameetriliste modelleerimisvahendite abil (joonis 6.4-3), kui nõuded on hästi määratletud. Niipea, kui projekt on kirjeldatud parameetrite kujul, muutub selle moodustamine tehniliselt teostatavaks, näiteks visuaalsete programmeerimiskeelte abil, nagu Grasshopper Dynamo, mis on integreeritud kaasaegsetesse CAD -keskkondadesse või tasuta lahendused Blender, UE, Omniverse.



Joonis 6.4-4 Enamik tüpiseeritud projekte luuakse juba täielikult automaatselt tänu parameetrilistele programmeerimisvahenditele.

Juba praegu ei looda suuri tööstuslikke ja tüpiseeritud projekte mitte disainiosakonna käe läbi, vaid parameetriliste tööriistade ja visuaalse programmeerimise abil. See võimaldab luua mudelit, mis põhineb pigem andmetel kui konkreetse disaineri või juhi subjektiivsetel otsustel.

Sisu eelneb kujundusele. Disain ilma sisuta ei ole disain, vaid kaunistus [128].

- Jeffrey Zeldman, veebidisainer ja ettevõtja

Protsess ei alga joonistamise või 3D -modelleerimisega, vaid nõuete koostamisega. Nõuded määravad, milliseid elemente projektis kasutatakse, milliseid andmeid tuleb edastada teistele osakondadele ja süsteemidele. Ainult struktureeritud nõuete olemasolu võimaldab regulaarselt (näiteks isegi iga 10 minuti tagant, ilma et see häiriks projekteerija tööd) mudelite automaatset .

Võib-olla saab tulevikus CAD- (BIM-) süsteemist lihtsalt andmebaasi täitmise liides ja see, millises CAD-vahendis modelleerimine (füüsilisel tasandil), ei ole oluline.

Sarnaselt kasutatakse masinaehituses sageli 3D-modelleerimist, kuid see ei ole projekti vajalik või kohustuslik element. Enamasti piisab klassikalisest 2D-dokumentatsioonist ja selle alusel luuakse vajalik infomudel. See mudel pannakse kokku tööstusstandardite kohaselt struktureeritud komponentidest ja see sisaldab kogu vajalikku teavet projekteerimise ja tootmiskorralduse mõistmiseks. Tehase infomudeli alusel luuakse seejärel tehase infomudel, millele lisatakse konkreetsed tooted ja vooluskeemid, mis on juba orienteeritud tehnoloogide vajadustele. Kogu protsessi saab korraldada ilma tarbetu keerukuseta, koormamata süsteemi üle 3D-graafikaga, kus see ei anna tegelikku kasu.

Oluline on mõista, et 3D-mudel ise ja CAD -süsteem ei peaks mängima peamist rolli - see on vaid vahend kvantitatiivseks ja geomeetriliseks analüüsiks. Kõik muud parameetrid, välja arvatud geomeetria, mis kirjeldavad üksust, tuleks võimaluse korral salvestada ja töödelda väljaspool CAD-keskkonda (BIM).

Disainimine parameetrite abil ei ole lihtsalt trend, vaid ehitussektori vältimatu tulevik. Selle asemel, et luua keerulisi 3D -mudeleid käsitsi, hakkavad projekteerijad töötama andmetega, valideerima neid ja automatiseerima protsesse, tuues ehituse lähemale programmeerimise maailmale. Aja jooksul ehitatakse projekteerimisprotsessid üles tarkvaraarenduse põhimõtetele:

- Nõuete loomine → Mudeli loomine → Upload serverisse → Validate changes → Pull request
- Pull-taotluste puhul viiakse automaatselt läbi mudelikontroll nõuete suhtes, mis on loodud enne või projekteerimise ajal Pull-taotluse osana.
- Pärast andmete kvaliteedi kontrollimist ja heakskiitmist viiakse muudatused projekti, ühisandmebaasi või kantakse automaatselt üle teistesse süsteemidesse.

Juba praegu alustatakse masinaehituses selliseid konstruktsioonimuudatusi muudatusteate koostamisega. Sarnane skeem ootab ehitustööstust: projekteerimine on iteratiivne protsess, mille iga sammu toetavad parameetrilised nõuded. Selline süsteem võimaldab projekteerijatel luua automaatseid kontrole ja automaatseid tõmbetaotlusi konkreetsete nõuete jaoks.

Tuleviku disainer on eelkõige andmete kasutaja, mitte käsitsi modelleerija. Tema ülesanne on täita projekt parameetriliste objektidega, kus geomeetria on ainult üks atribuutidest.

Just arusaam andmete modelleerimise, klassifitseerimise ja standardiseerimise tähtsusest, mida on üksikasjalikult käsitletud raamatu eelmistes peatükkides, mängib ümberkujundamisel olulist rolli. Tuleviku kujundusnormid vormistatakse XLSX või XML -skeemide kujul võtmeväärtusparameetrite paaridena.

Ehitustööstuse tulevik on andmete kogumine, nende analüüsimine, valideerimine ja protsesside automatiseerimine analüütiliste vahendite abil. BIM (või CAD) ei ole lõppeesmärk, vaid ainult arenguetapp. Kui spetsialistid mõistavad, et nad saavad töötada otse andmetega, möödudes traditsioonilistest CAD-vahenditest, annab mõiste "BIM" ise järk-järgult teed mõistetele, mis käsitlevad struktureeritud ja granuleeritud ehitusprojektide andmete kasutamist.

Üks peamisi muutusi kiirendavaid tegureid on olnud suurte keelemudelite (LLM) ja nendel põhinevate vahendite teke. Need tehnoloogiad muudavad projekteerimisandmete käsitlemise viisi, võimaldades juurdepääsu teabele ilma, et oleks vaja põhjalikke teadmisi APIdest või müüja lahendustest. LLMide abil muutub nõuete loomise protsess ja suhtlemine CAD andmetega intuitiivseks ja ligipääsetavaks.

LLM-i tekkimine projekteerimise CAD andmetötlusprotsessides

Lisaks CAD-andmebaasidele juurdepääsu vahendite ning avatud ja lihtsustatud CAD -vormingute arendamisele on LLM -tööriistade (Large Language Models) teke muutmas revolutsiooniliselt projekteerimisandmete töötlemist. Kui varem oli juurdepääs teabele peamiselt keerukate liidest kaudu ning nõudis programmeerimisoskust ja API tundmist, siis nüüd on võimalik andmetega suhelda loomuliku keele abil.

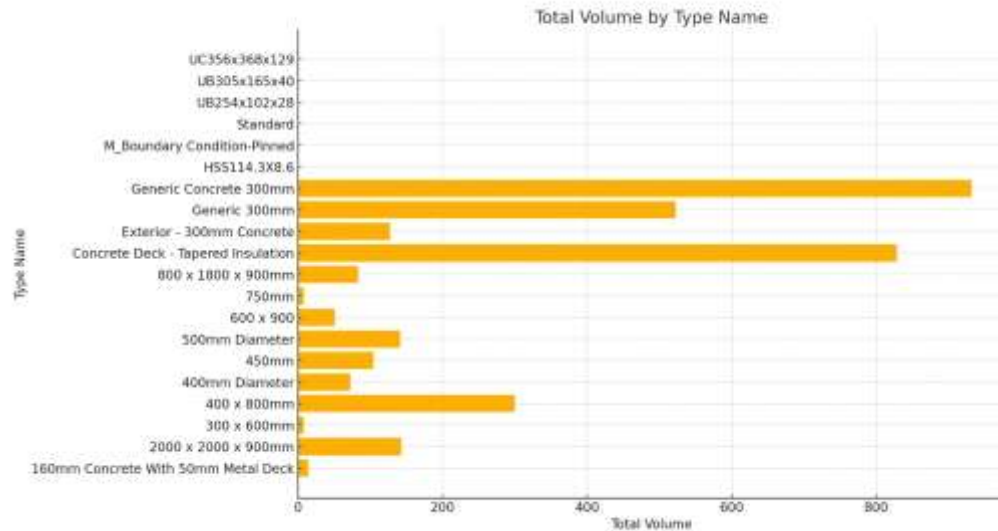
Insenerid, juhid ja planeerijad, kellel puudub tehniline taust, saavad projektiandmetest vajalikku teavet, sõnastades päringuid tavakeeles. Eeldusel, et andmed on struktureeritud ja kättesaadavad (joonis 4.1-13), piisab, kui esitada LLM jututuppa küsimus, näiteks: *"Näita tabelis rühmitusega tüübi järgi kõik seinad, mille maht on üle 10 kuupmeetri"* - ja mudel teisendab selle päringu automaatselt SQLi või Pandase koodiks, genereerides kokkuvõtliku tabeli, graafiku või isegi valmis dokumendi.

Allpool on toodud mõned näited tegelikust elust, kuidas LLM -mudelid suhtlevad eri CAD- (BIM-) formaatides esitatud projekteerimisandmetega.

- 🗣️ Näide päringu kohta LLM vestluses CAD-projektile RVT formaadis pärast konverteerimist (joonis 4.1-13) tabeliandmekavasse (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN või mõni muu):

Rühmitada andmed RVT-failist saadud andmeraamistikus "Tüübi nimi" parameetri "Volume" summeerimisel ja näidata rühmas olevate elementide arvu. Ja palun näidake seda kõike horisontaalse histogrammina ilma nullväärtusteta.

LLM vastus horisontaalse tulpdiagrammina (PNG-vormingus):



Joonis 6.4-5 Selle asemel, et kasutada 17 hiireklikki või 40 rida koodi pluginate abil, saame LLM-is koheselt teksti päringu abil välja QTO tabeli.

Selleks, et luua QTO tabel seinatüüpide kohta koos kogupinna ja kogusega kategooriast "Seinad", sõnastame tekstipäringu LLM -chat:

Võtke projekti andmeraamist ainult need elemendid, mille parameetris "Kategooria" on "OST_Seinad", grupeerige need "Tüübi nimi" järgi, summeerige veeru "Pindala" väärtus, lisage kogus ja kuvage need tabelis, eemaldades nullväärtused.

LLM vastus valmis QTO tabelina:

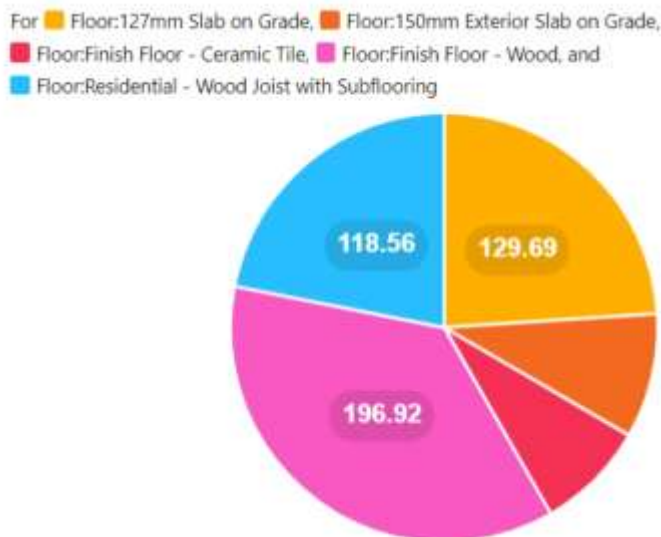
Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

Joonis 6.4-6 QTO tabeli loomine loomulikus keeles annab sama kvaliteetse tulemuse kui CAD - (BIM-) vahendite kasutamisel.

Teeme päringu projektile IFC formaadis pärast teisendamist tabeli kujuliseks andmeraamatuks ja sisestame sarnase tekstpäringu mis tahes LLM vestlusesse:

Võtke projektist ainult need elemendid, millel on parameetris "Parent" väärtused Level 1 ja Level 2, ja võtke elemendid, millel on parameetris "Category" väärtused IfcSlab, seejärel grupeerige need elemendid parameetri "ObjectType" järgi, summeerige väärtused parameetris "PSet_RVT _Dimensions Area" ja näidake neid ringdiagrammina.

LLM vastus kui valmis Tükigraafik elementide rühmade kohta IFC andmetest :



Joonis 6.4-7 IFC struktureeritud kujul olevate andmete päringu tulemus võib olla mis tahes tüüpi graafik, mis on andmete mõistmiseks mugav.

Iga saadud valmislahenduse (joonis 6.4-5 - joonis 6.4-7) taga on kümnekond rida Pythoni koodi, mis kasutab Pandas raamatukogu. Saadud koodi saab kopeerida LLMi jututuppa ja kasutada mis tahes kohalikus või veebipõhises IDE-s, et saada identseid tulemusi väljaspool LLMi jututuba.

Samas LLM-chatis saame töötada mitte ainult 3D-CAD (BIM) formaatidest saadud projektidega, vaid ka DWG-formaadis tasandusjoonistega, millele saame esitada päringu LLM chatis, et kuvada näiteks andmeid elementide rühmade kohta joonte või 3D-geomeetria kujul pärast teisendamist struktureeritud kujul.

DWG -failide automatiseeritud analüüs LLM ja Pandasiga

Andmete töötlemine DWG -failidest on alati olnud keerukas ülesanne, mis nõuab spetsiaalset tarkvara ja sageli käsitsi analüüsi. Tehisintellekti ja LLM-vahendite arenguga on siiski võimalik automatiseerida paljusid etappe, seda tänapäeval enamasti manuaalset protsessi. Vaatleme tõelist torustikku LLMile (antud näites ChatGPT) esitatud taotlusi, mis võimaldavad töötada DWG joonistega:

- Filtreeri DWG andmeid kihi, ID ja koordinaatide järgi.
- Visualiseerida elementide geomeetria
- Automaatne märkuste tegemine joonistele parameetrite alusel
- Laiendage seina polügoonid horisontaaltasandile

- Loo interaktiivne 3D -visualiseerimine tasapinnalistest andmetest.
- Struktureerida ja analüüsida ehitusandmeid ilma keeruliste CAD -tööriistadeta.

Meie puhul algab torujuhtme koostamise protsess järjestikuse koodi genereerimisega LLM kaudu. Kõigepealt genereeritakse ülesannet kirjeldav päring. ChatGPT genereerib Python -koodi, mida täidetakse ja analüüsitakse, näidates tulemust jututoa sees. Kui tulemus ei vasta ootustele, parandatakse päringut ja protsessi korratakse

Pipeline on andmete töötlemiseks ja analüüsimiseks tehtavate automatiseeritud etappide jada. Sellises protsessis võtab iga samm andmed sisendiks, teeb ümberkujundusi ja edastab tulemuse järgmisele sammule.

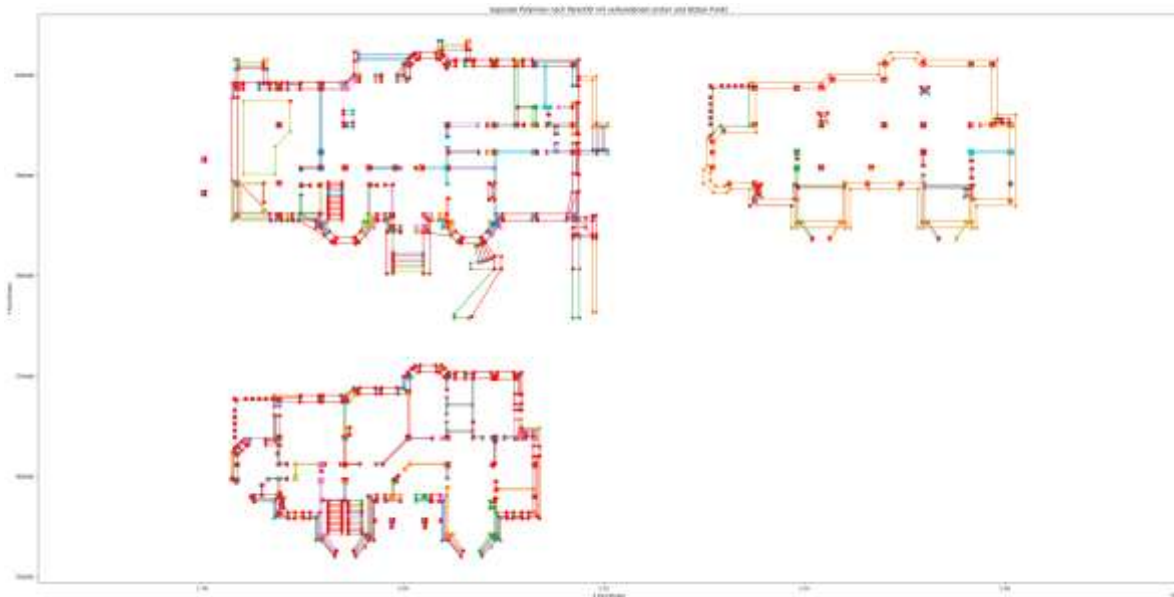
Pärast soovitud tulemuse saamist kopeeritakse kood LLMist ja kleebitakse see plokkide kujul koodiks mõnes mugavas IDE-s, meie puhul Kaggle'i platvormil.com. Saadud koodifragmendid ühendatakse üheks Pipeline, mis automatiseerib kogu protsessi - alates andmete laadimisest kuni nende lõpliku analüüsini. Selline lähenemisviis võimaldab analüütiliste protsesside kiiret arendamist ja skaleerimist ilma sügavate programmeerimisoskusteta. Kõigi allpool esitatud fragmentide täielik kood koos näidisküsimustega on leitav Kaggle.com platvormil, kui otsida "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. [129].

Alustame tööd DWG-andmetega pärast teisendamist struktureeritud kujul (joonis 4.1-13), klassikalise sammuga - kõigi meie ülesande jaoks vajalike jooniseandmete, täpsemalt polügoonide (parameeter "ParentID" võimaldab grupeerida jooned gruppidesse) rühmitamise ja filtreerimisega, mille parameetris (andmeraami veerus) "Layer" on string väärtus, mis sisaldab järgmist tähekombinatsiooni (Regex) - "wall".

- 🗨️ Sarnase ülesande koodi ja tulemuse saamiseks pildi kujul tuleks kirjutada järgmine päring LLM-is:

Kõigepealt kontrollige, kas DWG-st saadud andmeraamistik sisaldab määratletud veerge: "Layer", "ID", "ParentID" ja "Point". Seejärel filtreerige veerust "Layer" välja ID-d, mis sisaldavad stringi "wall". Leidke veerus "ParentID" elemendid, mis vastavad nendele identifikaatoritele. Määrake funktsioon, et puhastada ja jagada andmed veerus "Point". See hõlmab sulgude eemaldamist ja väärtuste jagamist x-, y- ja z-koordinaatideks. Joonistage andmed, kasutades matplotlib'i. Iga unikaalse "ParentID" jaoks joonistatakse eraldi polügoon, mis ühendab "Point"-koordinaadid. Veenduge, et esimene ja viimane punkt on võimalusel ühendatud. Määrake asjakohased sildid ja pealkirjad, tagades, et x- ja y-teljed on võrdselt skaleeritud.

- 🖼️ Vastus LLM annab teile valmis pildi, mille taga on peidetud Python-kood, mis selle genereeris:

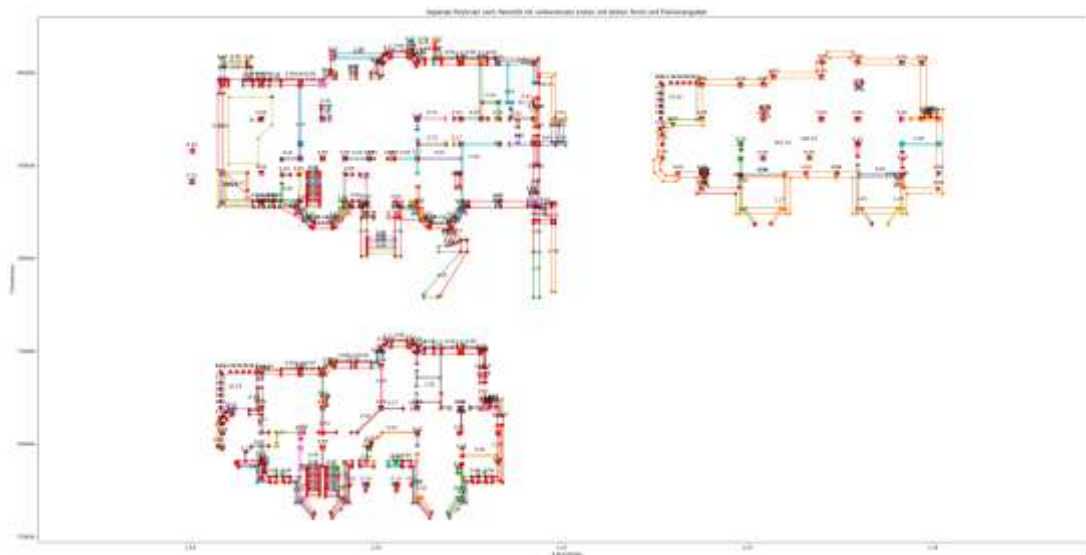


Joonis 6.4-8 LLM kood eraldas DWG failist kõik "seina" kihi jooned, puhastas nende koordinaadid ja konstrueeris ühe Pythoni raamatukogu abil polügoonid.

- 🗨️ Nüüd lisame joontele pindala parameetri, mis on iga polügooni omadustes (ühes andmeraami veerus):

Nüüd saada igast polügoonist ainult üks "ParentID" - leia see ID veerust "ID", võta "Area" väärtus, jaga see 1,000,000-ga ja lisa see väärtus graafikule.

LLM vastus näitab uut graafikut, kus igal polügraanil on pindala koos pealkirjaga:

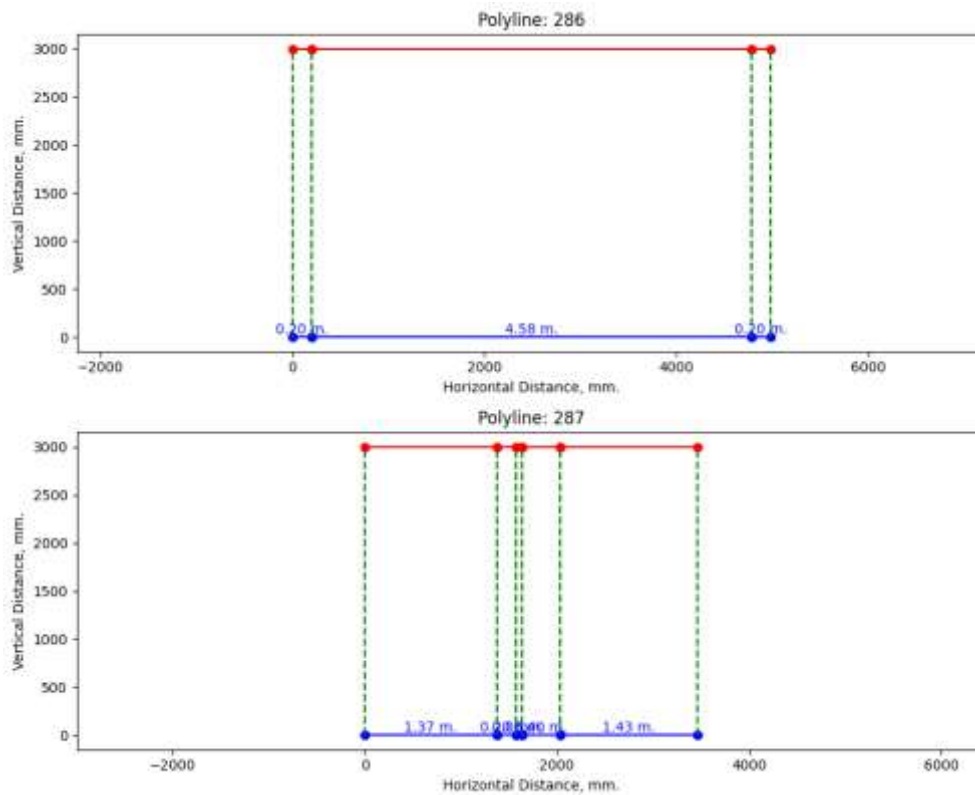


Joonis 6.4-9 LLM on lisanud koodi, mis võtab iga polügooni pindala väärtused ja lisab need pildile koos joone visualiseerimisega.

Seejärel teisendame iga polügooni horisontaalseks jooneks, lisame 3000 mm kõrgusel oleva paralleelse joone ja ühendame need üheks tasapinnaks, et näidata sel viisil seinaelementide pindade paigutust:

Sa pead võtma kõik elemendid veerust "Layer", mille väärtus on "wall". Võtke need ID-d loendina veerust "ID" ja leidke need ID-d kogu andmeraamist veerus "ParentID". Kõik elemendid on jooned, mis on kombineeritud üheks polügooniks. Igal joonel on erinev x, y geomeetria esimene punkt veerus "Point". Te peate võtma iga polügooni kordamööda ja punktist 0,0 horisontaalselt joonistama iga polügooni segmendi pikkuse. polügooni iga segmendi pikkuse üheks jooneks. Seejärel joonistage täpselt samad jooned ainult 3000 võrra kõrgemale, ühendage kõik punktid üheks tasapinnaks.

LLM-vastus väljastab koodi, mis võimaldab joonistada seinajooniseid tasapinnal:

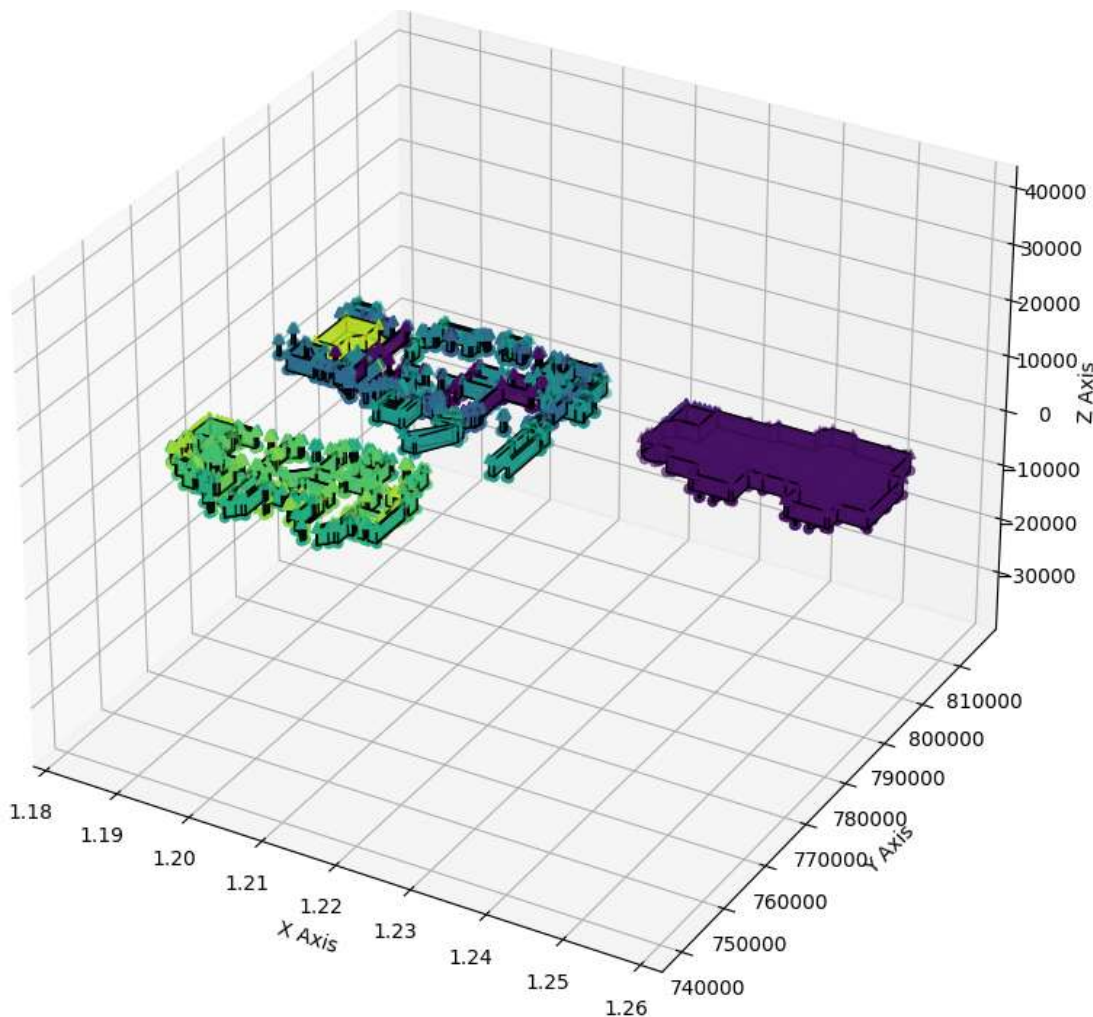


Joonis 6.4-10 Muudame iga polügooni käskluste abil paigutuseks, mis visualiseerib seina tasapinnad otse LLM-i vestluses.

Nüüd liigume 2D-projektsioonist 3D - modelleerime seinad lamedatest joontest, ühendades polüjoonte ülemise ja alumise kihi:

Visualiseeri seinaelemendid 3D-s, ühendades polüjooned kõrgusel $z = 0$ ja $z = 3000$ mm. Luua suletud geomeetria, mis kujutab hoone seinu. Kasutage Matplotlib 3D graafikavahendit.

- LLM genereerib interaktiivse 3D -graafiku, kus iga polügraafik on kujutatud tasandite kogumina. Kasutaja saab vabalt liikuda elementide vahel arvutihiirega, uurides mudelit 3D-režiimis, kopeerides koodi vestlusest IDE-sse:



Joonis 6.4-11 LLM aitas ehitada koodi [129], et visualiseerida lamedad joonistusjooned 3D vaateks, mida saab uurida 3D vaates IDE-siseses 3D vaatajas.

Loogilise ja korratava torujuhtme loomiseks - alates DWG -faili esialgsest konverteerimisest ja laadimisest kuni lõpptulemuseni - on soovitatav pärast iga sammu kopeerida genereeritud LLM - koodiplokk IDE-sse. Sel viisil ei saa mitte ainult kontrollida tulemust chatis, vaid seda ka kohe oma arenduskeskkonnas käivitada. See võimaldab teil protsessi järjestikku üles ehitada, siluda ja kohendada seda vastavalt vajadusele.

Kõigi fragmentide (joonised 6.4-8 kuni joonised 6.4-11) täieliku Pipeline koodi koos näidisküsimustega leiate Kaggle platvormilt.com otsides "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. Kaggle'is saab mitte ainult vaadata koodi ja kasutatud päringuid, vaid

ka kopeerida ja testida kogu Pipeline'i koos originaal DWG andmeraamidega pilves tasuta, ilma et oleks vaja paigaldada lisatarkvara või IDEd ennast.

Selles peatükis esitatud lähenemisviis võimaldab teil täielikult automatiseerida DWG -projektidel põhinevate dokumentide kontrollimist, töötlemist ja genereerimist. Välja töötatud Pipeline sobib nii üksikute jooniste töötlemiseks kui ka kümnete, sadade ja tuhandete DWG-failide partiitöötamiseks koos iga projekti jaoks vajalike aruannete ja visualiseeringute automaatse genereerimisega.

Protsessi saab korraldada järjestikku ja läbipaistvalt: kõigepealt konverteeritakse CAD -faili andmed automaatselt XLSX-formaati, seejärel laaditakse andmeraamidesse, millele järgneb rühmitamine, kontrollimine ja tulemuste genereerimine - kõik see on rakendatud ühes Jupyter notebookis või Python -scriptis, mis tahes populaarses IDE-s. Vajaduse korral saab protsessi hõlpsasti laiendada, integreerides seda projektdokumentatsiooni haldussüsteemidega: CAD-faile saab automaatselt välja otsida vastavalt kindlaksmääratud kriteeriumidele, tulemusi saab tagastada salvestussüsteemi ja kasutajaid saab teavitada, kui tulemused on valmis - e-posti või sõnumitooja kaudu.

LLM vestluste ja agentide kasutamine projekteerimisandmetega töötamiseks vähendab sõltuvust spetsiaalsetest CAD -programmidest ning võimaldab teostada arhitektuursete projektide analüüsi ja visualiseerimist, ilma et oleks vaja käsitsi liidesega suhelda - ilma hiireklõpsudeta ja keerulise menüüst navigeerimise meespidamiseta.

Iga päevaga kuuleb ehitussektoris üha rohkem LLMist, granuleeritud struktureeritud andmetest, DataFramesist ja kolonniandmebaasidest. Erinevatest andmebaasidest ja CAD-vormingutest moodustatud ühtsed kahemõõtmelised DataFrame'id on ideaalne kütus kaasaegsetele analüüsivahenditele, millega tegelevad aktiivselt teiste tööstusharude spetsialistid.

Automatiseerimisprotsess ise lihtsustub oluliselt - selle asemel, et uurida suletud nišitoodete API ja kirjutada keerulisi skripte parameetrite analüüsimiseks või teisendamiseks, piisab nüüd ülesande sõnastamisest üksikute tekstikäskude kogumina, mis pannakse vajaliku programmeerimiskeele jaoks vajaliku Pipeline või Workflow-protsessi sisse, mis töötab tasuta peaaegu igas seadmes. Enam ei pea ootama uusi tooteid, formaate, lisaseadmeid või uuendusi CAD- (BIM-) tööriistade tarnijatelt. Insenerid ja ehitajad saavad iseseisvalt töötada andmetega, kasutades lihtsaid, tasuta ja kergesti arusaadavaid vahendeid, mida abistavad LLM vestlused ja agendid.

Järgmised sammud: üleminek suletud vormingutelt avatud andmetele

Tuleviku projekteerimisandmetega töötades on ebatõenäoline, et keegi peab tõesti mõistma patenteeritud tööriistade geomeetrilisi tuumasid või õppima sadu ühildumatuid formaate, mis sisaldavad sama teavet. Samas on raske argumenteerida uute tasuta tööriistade, avatud andmete ja lähenemisviiside kasutamise poolt, mida tarkvaratootjad tõenäoliselt ei propageeri, ilma et mõistaks, miks on oluline liikuda avatud struktureeritud andmetele.

Selles peatükis oleme käsitlenud CAD (BIM) andmete peamisi omadusi, nende piiranguid ja võimalusi ning seda, et hoolimata müüjate turunduslubadustest seisavad insenerid ja projekteerijad iga päev silmitsi raskustega projekteerimisandmete väljavõtte, edastamise ja analüüsimisega. Nende

süsteemide ülesehituse mõistmine ja alternatiivsete lähenemisviiside tundmaõppimine - mis põhinevad avatud vormingutel ja automatiseerimisel LLM kaudu - võib muuta elu palju lihtsamaks isegi üksikute spetsialistide, rääkimata ettevõtetest. Selle osa kokkuvõtteks tasub esile tuua peamised praktilised sammud, mis aitavad teil rakendada käesolevas osas käsitletud lähenemisviise oma igapäevastes ülesannetes:

■ Laiendage oma tööriistakomplekti projektandmetega töötamiseks

- ☐ Tutvu olemasolevate lisandmoodulite ja utiliitidega, et saada andmeid kasutatavatest CAD - (BIM-) süsteemidest.
- ☐ Tutvuge olemasolevate SDK-de ja API-dega, mis võimaldavad automatiseerida andmete väljavõtteid suletud formaatidest, ilma et peaksite käsitsi avama spetsiaalset tarkvara.
- ☐ Põhioskused avatud mitteparameetriliste geomeetriliste vormingute (OBJ, glTF, USD, DAE) ja vastavate avatud lähtekoodiga raamatukogudega töötamisel.
- ☐ Püüdke mõelda süsteemile, mis salvestaks projekti metaandmeid eraldi geomeetriast väljaspool CADi (BIM) lahendusi, et lihtsustada analüüsi ja integreerimist teiste süsteemidega.
- ☐ Kasutage LLM, et automatiseerida andmete konverteerimise probleeme vormingute vahel.

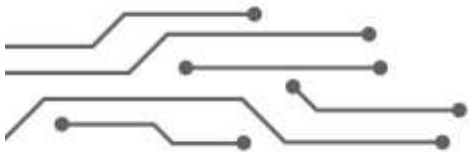
■ Looge oma protsessid projektiteabe käsitlemiseks

- ☐ Alustada ülesannete ja modelleerimisnõuete kirjeldamist parameetrite ja nende väärtuste kaudu lihtsas ja struktureeritud vormis.
- ☐ Luua isiklik skriptide või koodiplokkide raamatukogu sageli tehtavate toimingute jaoks.

■ Edendada avatud standardite kasutamist oma töös

- ☐ kutsuda kolleege ja partnereid üles jagama andmeid avatud vormingus, mida ei piira tarkvaratootjate ökosüsteem.
- ☐ demonstreerida struktureeritud andmete kasutamise eeliseid konkreetsete näidetega.
- ☐ Algatada arutelusid suletud formaatidega seotud probleemide ja võimalike lahenduste üle.

Isegi kui te ei saa muuta oma ettevõtte poliitikat seoses CAD - (BIM-) platvormidega, võimaldab avatud formaadis projektandmetega töötamise põhimõtete isiklik mõistmine teil oma töö tõhusust märkimisväärselt suurendada. Luues oma vahendid ja meetodid andmete ekstraheerimiseks ja teisendamiseks erinevatest vormingutest, ei optimeeri te mitte ainult oma töövooge, vaid saate ka paindlikkuse, et mööda minna standardtarkvaralahenduste piirangutest.



VII OSA

ANDMEPÕHINE OTSUSTE TEGEMINE, ANALÜÜTIKA, AUTOMATISEERIMINE JA MASINÕPE

Seitsmes osa keskendub andmeanalüütikale ja protsesside automatiseerimisele ehitustööstuses. Selles käsitletakse, kuidas andmed muutuvad otsuste tegemise aluseks, ja selgitatakse teabe visualiseerimise põhimõtteid tõhusa analüüsi jaoks. Üksikasjalikult kirjeldatakse tulemuslikkuse põhinäitajaid (KPI), investeeringutasuvuse hindamise meetodeid (ROI) ja näidislaudade loomist projektide jälgimiseks. Erilist tähelepanu pööratakse ETL-protsessidele (Extract, Transform, Load) ja nende automatiseerimisele torujuhtmete (Pipeline) abil, et muuta erinevad andmed analüüsiks struktureeritud teabeks. Käsitletakse töövoogude korraldamise vahendeid, nagu Apache Airflow, Apache NiFi ja n8n, mis võimaldavad luua automatiseeritud andmepiire ilma sügavate programmeerimisoskusteta. Olulist rolli mängivad suured keelemudelid (LLM) ja nende kasutamine andmeanalüüsi lihtsustamiseks ja rutiinsete ülesannete automatiseerimiseks

PEATÜKK 7.1.

ANDMEANALÜÜS JA ANDMEPÕHINE OTSUSTE TEGEMINE

Pärast teabe kogumise, struktureerimise, puhastamise ja kontrollimise etappe on saadud ühtne ja analüüsitav andmekogum. Raamatu eelmistes osades käsitleti heterogeensete allikate süstematiseerimist ja struktureerimist - alates PDF-dokumentidest ja koosolekute tekstikirjeldustest kuni CAD-mudelite ja geomeetriliste andmeteni. Üksikasjalikult kirjeldatakse teabe kontrollimist ja vastavusse viimist erinevate süsteemide ja klassifikaatorite nõuetega, dubleeringute ja vastuolude kõrvaldamist.

Kõik arvutused, mis tehakse nende andmete põhjal (raamatu kolmas ja neljas osa) - alates lihtsatest teisendustest kuni aja-, kulu- ja ESG-näitajate arvutusteni (viies osa) - on agregeeritud analüüsiülesanded. Need on aluseks projekti hetkeseisu mõistmisele, selle parameetrite hindamisele ja seejärel otsuste tegemisele. Selle tulemusel muutuvad andmed arvutuste tulemusel lahtistest kirjetest hallatavaks ressursiks, mis on võimeline vastama olulistele äriküsimustele.

Eelmistes peatükkides on üksikasjalikult kirjeldatud andmete kogumise ja kvaliteedikontrolli protsesse, mida kasutatakse tüüpilistes äritegevustes ja ehitussektori spetsiifilistes protsessides. Analüütika selles kontekstis on paljuskümnelt sarnane teiste tööstusharude rakendustega, kuid sellel on mitmeid eripärasid.

Järgnevatel peatükkides kirjeldatakse üksikasjalikult andmeanalüüsi protsessi, sealhulgas automatiseerimise etappe alates teabe esialgsest hankimisest ja selle ümberkujundamisest kuni selle hilisema edastamiseni sihtsüsteemidesse ja dokumentidesse. Kõigepealt esitatakse teoreetiline osa, milles keskendutakse andmepõhise otsustamise valitud aspektidele. Seejärel käsitletakse järgmistel peatükkides praktilist osa, mis on seotud ETLi automatiseerimise ja ehitamisega -Pipeline.

Andmed kui ressurss otsuste tegemisel

Andmepõhine otsustamine on sageli iteratiivne protsess, mis algab süstemaatilise teabe kogumisega erinevatest teabeallikatest. Nagu looduslik tsükkel, langevad üksikud andmeelemendid ja terved infosüsteemid järk-järgult pinnasesse - kogunevad ettevõtete teabehoidlatesse (joonis 1.3-2). Aja jooksul muutuvad need andmed, nagu langenud lehed ja oksad, väärtuslikuks materjaliks. Andmeinseneride ja analüütikute mütseel korrastab ja valmistab teavet edaspidiseks kasutamiseks ette ning muudab langenud andmed ja süsteemid väärtuslikuks kompostiks, et kasvatada uusi võrseid ja uusi süsteeme (joonis 1.2-5).

Analüütika laialdase kasutamise suundumused erinevates tööstusharudes tähistavad uue ajastu algust, kus töö andmetega muutub kutsetegevuse aluseks (joonis 7.1-1). Ehitustööstuse spetsialistidel on oluline kohaneda nende muutustega ja olla valmis liikuma uude ajastusse - andmete ja analüütika ajastusse

Andmete käsitsi liigutamine tabelite vahel ja arvutuste käsitsi teostamine on järk-järgult muutumas minevikuks, andes teed automatiseerimisele, andmevoogude analüüsile, analüütikale ja masinõppele. Need vahendid on muutumas kaasaegsete otsustustugisüsteemide põhielementideks.

McKinsey raamatus "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] tsiteeritakse uuringut, mis viidi läbi 1330 tippjuhi seas erinevatest piirkondadest, tööstusharudest ja funktsionaalsetest valdkondadest, mis viidi läbi 2022. aastal. [130], tsiteerib uuringut, mis viidi läbi 2022. aastal 1330 tippjuhi seas erinevatest piirkondadest, tööstusharudest ja funktsionaalsetest valdkondadest. Selle tulemuste kohaselt kasutab 70 protsenti juhtidest täiustatud analüütikat oma ideede genereerimiseks ja 50 protsenti rakendab tehisintellekti otsustusprotsesside parandamiseks ja automatiseerimiseks.



Joonis 7.1-1 Andmeanalüüs ja analüütika on peamine vahend, mis aitab suurendada otsuste tegemise kiirust ettevõttes.

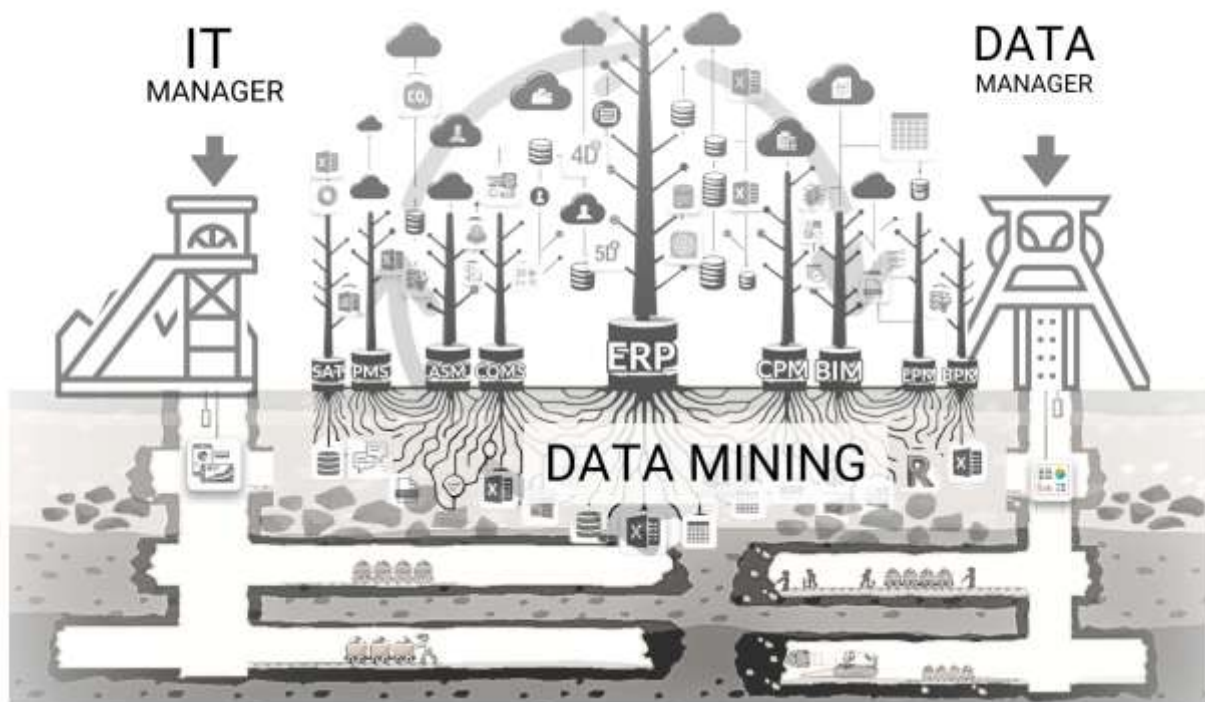
Andmeanalüüs tungib sarnaselt mütseeli levikuga mineviku otsuste huumusesse, aidates ühendada üksikuid süsteeme ja suunates juhte väärtuslikele arusaamadele. Need teadmised, nagu toitud lagunenud andmesüsteemi puudest, toidavad uusi otsuseid ettevõttes, viies tõhusate muudatuste ja kvaliteetse teabe kasvu, nagu uued võrsed ja idud, mis kasvavad rikkalikust ja tervest mullast (joonis 1.2-5).

Numbritel on oluline lugu. Nad loodavad, et te annate neile selge ja veenva hääle [131].

- Stephen Few, andmete visualiseerimise ekspert

Keskmise suurusega ja väikestes ettevõtetes on teabe kaevandamine ja ettevalmistamine edasiseks analüüsiks tänapäeval äärmiselt töömahukas protsess (joonis 7.1-2), mis on võrreldav 18. sajandi söekaevandusega. Kuni viimase ajani oli andmete kaevandamise ja ettevalmistamise töö pigem reserveeritud väga spetsialiseerunud nišis töötavatele seiklejatele, kellel oli väike ja piiratud hulk vahendeid, et töötada erinevat tüüpi andmetega struktureerimata, lahtiselt struktureeritud, segatud ja suletud allikatest.

Otsustajad ja juhid on sageli kogenematud heterogeensete andmete ja süsteemidega, kuid peavad siiski tegema nende põhjal otsuseid. Selle tulemusena on andmetel põhinev otsustamine tänapäeva ehitustööstuses viimastel aastakümnetel olnud vähem automatiseeritud protsess ja rohkem sarnaneb see kaevurite mitmepäevase käsitsi tehtava tööga varajastes söekaevandustes.



Joonis 7.1-2 Andmete kaevandamise protsessis läbivad eksperdid andmete ettevalmistamise keerulise tee - alates puhastamisest kuni struktureerimiseni hilisemaks analüüsiks.

Kuigi ehitussektori kaasaegsed andmekogumismeetodid on kindlasti arenenumad kui 12. sajandi kaevurite primitiivsed tehnikad, on see siiski keeruline ja kõrge riskiga ülesanne, mis nõuab märkimisväärsed ressursse ja teadmisi, mida saavad endale lubada vaid suured ettevõtted. Varasemate projektide kogunenud pärandist andmete väljavõtte ja analüüsiga on kuni viimase ajani tegelenud valdavalt suured, tehnoloogiliselt arenenud ettevõtted, kes on aastakümneid järjepidevalt andmeid kogunud ja säilitanud

Varem oli analüütika valdkonnas juhtiv roll tehnoloogiliselt küpsetel ettevõtetel, kes olid aastakümneid andmeid kogunud. Tänapäeval on olukord muutumas: juurdepääs andmetele ja andmetöötlusvahenditele on demokratiseerumas - varem keerulised lahendused on nüüd kõigile tasuta kättesaadavad.

Analüütika rakendamine võimaldab ettevõtetel teha täpsemaid ja teadlikumaid otsuseid reaalajas. Järgnev juhtumiuuring illustreerib, kuidas ajaloolised andmed aitavad teha rahaliselt põhjendatud

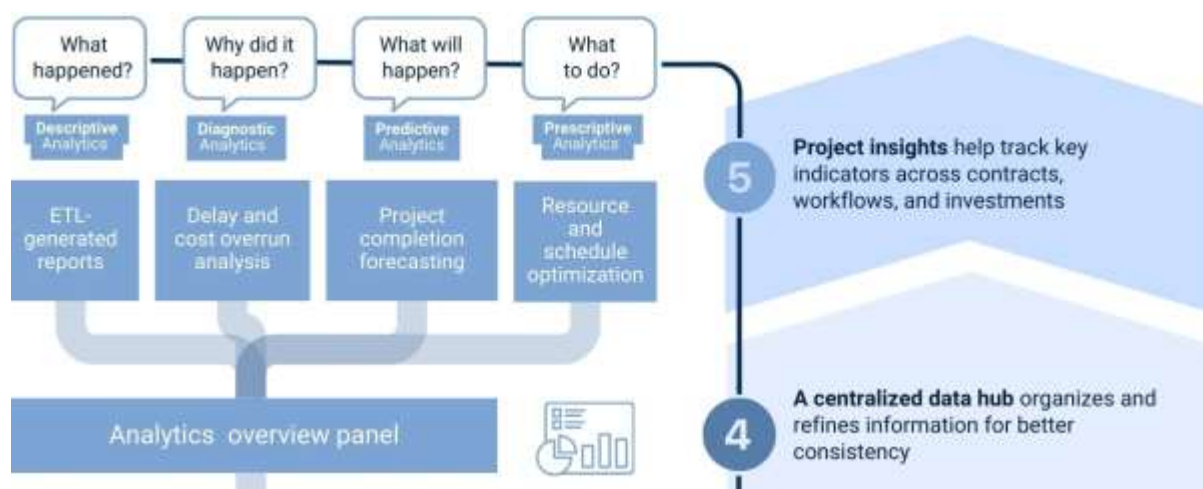
otsuseid:

- 🔊 **Projektijuht** - "Praegu on betooni keskmine hind linnas 82€/m³, meil on prognoosis 95 €/m³."
- 🔊 **Kalkulaator** - "Eelmiste projektide puhul oli ületamine umbes 15%, nii et ma tegin tagasilöögi."
- 🔊 **Andmehaldur või kliendipoolne kontrolliinsener** - "Vaatame viimase kolme pakkumise analüütikat."

Pärast varasemate projektide DataFrame'i analüüsimist saame:

- **Keskmine tegelik ostuhind:** 84,80 /m³€
- **Keskmine ülekulude määr:** +4,7%
- 🔊 **Soovitav määr hinnanguliselt:** ~ 85 /m³€

Selline otsus ei põhine enam subjektiivsetel tunnetel, vaid konkreetsetel ajaloolisel statistikal, mis aitab vähendada riske ja suurendada pakkumuse kehtivust. Varasemate projektide andmete analüüs muutub omamoodi "orgaaniliseks väetiseks", millest idanevad uued, täpsemad lahendused.



Joonis 7.1-3 Andmeanalüüs vastab kolmele põhiküsimusele: mis juhtus, miks see juhtus ja mida tuleks teha edasi.

Otsustajad ja juhid seisavad sageli silmitsi vajadusega töötada heterogeensete andmete ja süsteemidega ilma piisavate tehniliste eriteadmisteta. Sellistes olukordades on visualiseerimine, mis on analüüsiprotsessi üks esimesi ja kõige olulisemaid samme, andmete mõistmise peamine abivahend. See võimaldab esitada teavet visuaalsel ja arusaadaval kujul.

Andmete visualiseerimine: mõistmise ja otsuste tegemise võti

Tänapäeva ehitustööstuses, kus projektiandmeid iseloomustab keerukus ja mitmetasandiline struktuur, mängib visualiseerimine võtmerolli. andmete visualiseerimine võimaldab projektijuhtidel ja inseneridel visualiseerida suurtes, heterogeensetes andmemahtudes peituvaid keerulisi mustreid ja suundumusi.

andmete visualiseerimine lihtsustab projekti seisu mõistmist: ressursside jaotamine, kulusuundumused või materjalikasutus. Graafikud ja diagrammid muudavad keerulise ja kuiva teabe kättesaadavaks ja arusaadavaks, võimaldades kiiresti tuvastada tähelepanu vajavaid võtmevaldkondi ja märgata võimalikke probleeme.

andmete visualiseerimine mitte ainult ei lihtsusta teabe tõlgendamist, vaid on oluline samm analüüsiprotsessis ja teadlike juhtimisotsuste tegemisel, aidates vastata küsimustele "mis juhtus?" ja "kuidas see juhtus?" (joonis 2.2-5). (Joonis 2.2-5).

Graafika on visuaalne vahend loogiliste probleemide lahendamiseks [132].

- Jacques Bertin, "Graafika ja graafiline infotöötlus

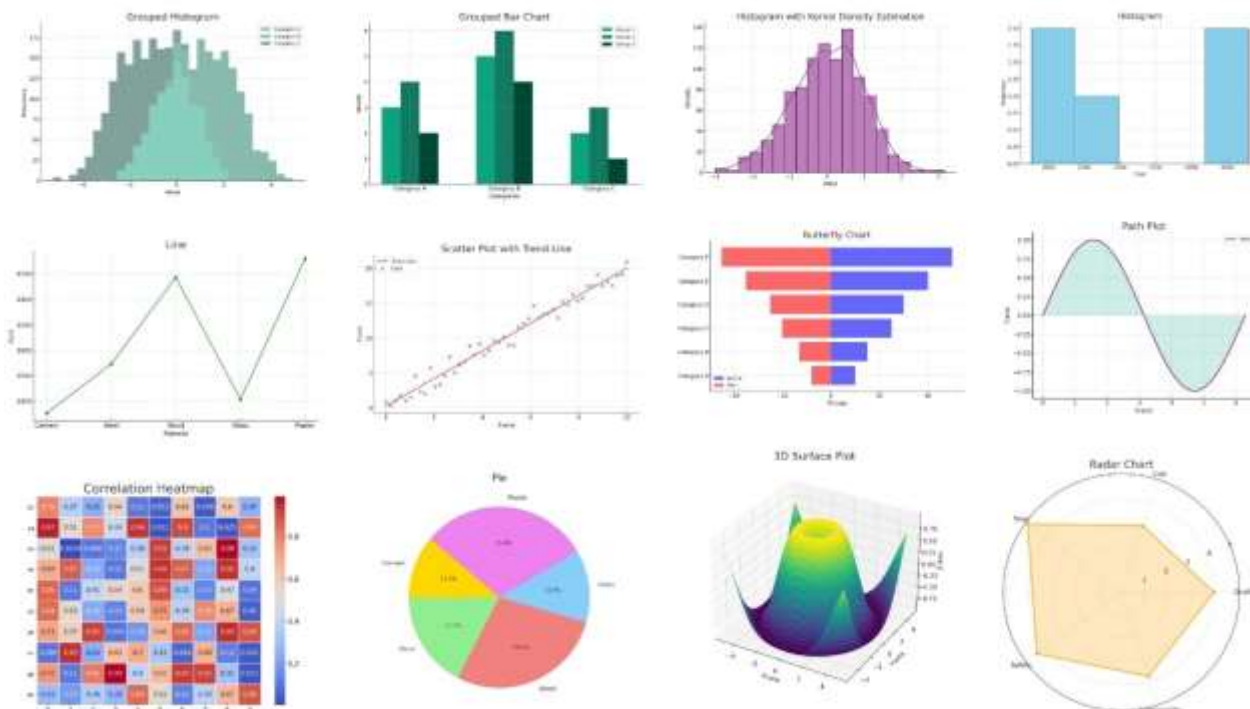
Enne oluliste otsuste tegemist kasutavad projektijuhid pigem andmete visuaalset esitust kui kuivi ja raskesti tõlgendatavaid arvandmeid tabelitest või tekstisõnumitest.

Andmed ilma visualiseerimiseta on nagu ehitusplatsil juhuslikult laiali laotatud ehitusmaterjalid: nende potentsiaal on ebaselge. Alles siis, kui need on selgelt visualiseeritud, nagu telliskividest ja betoonist ehitatud maja, saab nende väärtus selgeks. Kuni maja ehitamiseni on võimatu öelda, kas materjalihunnikust saab väike majake, luksusvilla või pilvelõhkuja.

Ettevõtetal on andmeid erinevatest süsteemidest (joonis 1.2-4 kuni joonis 2.1-10), finantstehinguid ja ulatuslikke tekstiandmeid. Nende andmete kasutamine ärilisel eesmärgil on aga sageli keeruline. Sellistes olukordades muutub visualiseerimine oluliseks vahendiks andmete tähenduse edastamisel, aidates esitada teavet mis tahes asjatundjale arusaadavas vormis, nagu näiteks armatuurlauad, graafikud ja diagrammid.

PwC uuringus "What Students Need to Succeed in a Rapidly Changing Business World" (2015) rõhutatakse [9], et edukad ettevõtted ei piirdu andmeanalüüsiga, vaid kasutavad aktiivselt interaktiivseid visualiseerimisvahendeid, nagu graafikud, infograafiad ja analüütilised armatuurlauad, et toetada otsuste tegemist. Aruande kohaselt - andmete visualiseerimine aitab klientidel mõista lugu, mida andmed räägivad graafikute, graafikute, diagrammide, armatuurtahvlite ja interaktiivsete andmemudelite abil.

Teabe teisendamine visuaalseteks graafilisteks vormideks, näiteks graafikuteks, diagrammideks ja diagrammideks, parandab inimese aju arusaamist ja andmete tõlgendamist (joonis 7.1-4). See võimaldab projektijuhtidel ja analüütikutel kiiremini hinnata keerulisi stsenaariume ja teha teadlikke otsuseid, mis põhinevad pigem visuaalselt äratuntavatel suundumustel ja mustritel kui intuitsioonil.



Joonis 7.1-4 Erinevad visualiseerimisviisid on mõeldud selleks, et aidata inimese ajal paremini mõista ja mõtestada numbrite kuiva teavet.

Visualiseeringute loomist andmetest ja erinevate tasuta visualiseerimisraamatukogude kasutamist käsitletakse üksikasjalikumalt järgmises peatükis, mis käsitleb ETL -protsesse.

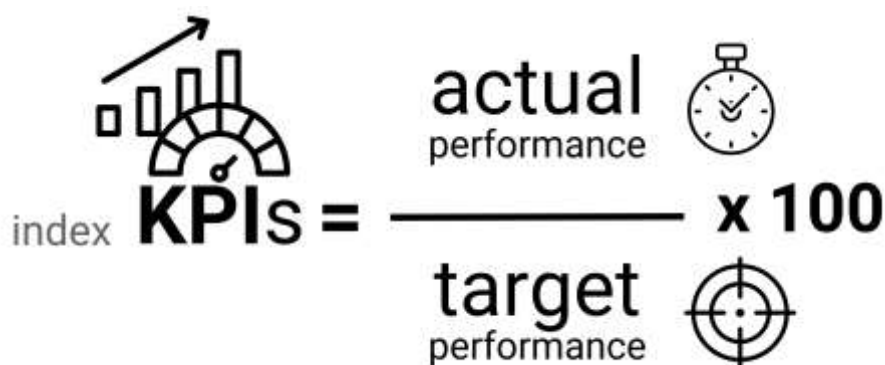
Visualiseerimine on muutumas ehitussektoris andmetega töötamise lahutamatuks osaks - see aitab mitte ainult "näha" andmeid, vaid ka mõista nende tähendust juhtimisülesannete kontekstis. Selleks, et visualiseerimine oleks tõeliselt kasulik, tuleb siiski eelnevalt kindlaks määrata, mida täpselt tuleks visualiseerida ja millised mõõdikud on projekti tulemuslikkuse hindamisel tõesti olulised. Siinkohal tulevadki mängu sellised tulemuslikkuse mõõdikud nagu tulemuslikkuse põhinäitajad ja tasuvusnäitajad (ROI). Ilma nendeta on isegi kõige ilusamad armatuurtahvlid oht, et need on lihtsalt "infomüra".

KPId ja ROI

Tänapäeva ehitustööstuses on tulemuslikkuse näitajate (KPI ja ROI) haldamine ja nende visualiseerimine aruannete ja näidisplaatide kaudu võtmeroll tootlikkuse ja projektijuhtimise tõhususe parandamisel.

Nagu igas äris, on ka ehituses vaja selgelt määratleda mõõdikud, mille alusel mõõdetakse edu, investeeringu tasuvust ja tulemuslikkust. Erinevate protsesside kohta andmete hankimisel peab andmepõhine organisatsioon kõigepealt õppima kindlaks määrama peamised **tulemuslikkuse põhinäitajad (Key Performance Indicators)** - kvantitatiivsed mõõdikud, mis kajastavad strateegiliste ja tegevuseesmärkide saavutamise ulatust.

Peamiste tulemusnäitajate arvutamiseks kasutatakse tavaliselt valemit (joonis 7.1-5), mis hõlmab tegelikke ja kavandatud näitajaid. Näiteks projekti, töötaja või protsessi individuaalse tulemuslikkuse näitaja arvutamiseks jagatakse tegelik tulemus kavandatud tulemuslikkusega ja korrutatakse tulemus 100%-ga.



$$\text{index KPIs} = \frac{\text{actual performance}}{\text{target performance}} \times 100$$

Joonis 7.1-5 Peamised tulemusnäitajad kasutatakse projekti või protsessi edukuse mõõtmiseks põhieesmärkide saavutamisel.

Saitide tasandil saab kasutada üksikasjalikumaid põhinäitajaid mõõdikuid:

- **Peamiste vaheetappide** (vundament, paigaldus, viimistlus) **ajastus** - võimaldab kontrollida tööplaanide järgimist.
- **Materjalide ülejäägi protsent** - aitab hallata hankeid ja vähendada raiskamist.
- **Planeerimata masinate seisakute arv** - mõjutab tootlikkust ja kulusid.

Vale mõõdikute valimine võib viia ekslike "*mida teha?*" otsuste tegemiseni (joonis 2.2-5). Näiteks kui ettevõtte keskendub ainult ruutmeetri maksumusele, kuid ei arvesta muudatuste maksumust, võib materjalide kokkuhoid tuua kaasa kehvema kvaliteedi ja suuremad kulud tulevastes projektides.

Eesmärkide seadmisel on oluline, et oleks selge, mida mõõdetakse. Ebaselge sõnastus viib valede järeldusteni ja raskendab kontrolli. Vaatame näiteid edukatest ja ebaõnnestunud põhinäitajatest ehituses.

Head tulemuslikkuse näitajad:

- 🗣️ "*Vähendage aasta lõpuks remonditööde osakaalu 10 protsendi võrra.*"
- 🗣️ "*Suurendame fassaadide paigaldamise kiirust 15 protsenti, ilma et kvaliteet kannataks järgmise kvartali jooksul*"

🗣️ "Vähendada masinate seisakuid 20% võrra, optimeerides tööplaane aasta lõpuks"

Need mõõdikud on selgelt mõõdetavad, neil on konkreetsed väärtused ja ajakava.

Halvad tulemuslikkuse näitajad:

- 🗣️ "Me ehitame kiiremini" (Kui palju kiiremini? Mida tähendab "kiiremini"?).
- 🗣️ "Me parandame betoonitööde kvaliteeti" (Kuidas täpselt kvaliteeti mõõdetakse?)
- 🗣️ "Me parandame töövõtjate suhtlemist kohapeal" (Millised kriteeriumid näitavad paranemist?)

Hea tulemuslikkuse näitaja on selline, mida saab mõõta ja objektiivselt hinnata. Ehituses on see eriti oluline, sest ilma selgete näitajateta on võimatu jälgida tulemuslikkust ja saavutada stabiilseid tulemusi.

Lisaks põhinäitajatele on investeringute tõhususe hindamiseks olemas täiendav mõõdik: **ROI (Return on Investment)** - investeringu tasuvuse näitaja, mis kajastab kasumi ja investeeritud vahendite suhet. ROI aitab hinnata, kas uute meetodite, tehnoloogiate või vahendite kasutuselevõtt on õigustatud: alates digitaalsetest lahendustest ja automatiseerimisest (nt joonis 7.3-2) kuni uute ehitusmaterjalide kasutamiseni. See näitaja aitab teha teadlikke otsuseid edasiste investeringute kohta, lähtudes nende tegelikust mõjust ettevõtte kasumlikkusele

Ehitusprojektide juhtimise kontekstis võib ROI (investeeringutasuvus) olla üks peamisi tulemusnäitajaid (KPI), kui ettevõtte eesmärk on mõõta projekti, tehnoloogia või protsessi täiustamise investeringu tasuvust. Näiteks kui rakendatakse uut ehitusjuhtimise tehnikat, võib ROI näidata, kui palju see on parandanud kasumlikkust.

Erinevatest allikatest kogutud andmete, näiteks materjalikulu, töötundide ja kulude, alusel toimuv regulaarne tulemusnäitajate ja tasuvusnäitajate mõõtmine võimaldab projektijuhtkonnal tõhusalt juhtida ressursse ja teha kiireid otsuseid. Nende andmete pikaajaline säilitamine võimaldab analüüsida tulevasi suundumusi ja optimeerida protsesse.

Peamiste tulemusnäitajate, tasuvusnäitajate ja muude näitajate visualiseerimiseks kasutatakse mitmesuguseid graafikuid ja graafikuid, mis tavaliselt ühendatakse armatuurlaudadeks.

Armatuurlaud ja armatuurlaud: mõõdikutega visualiseerimine tõhusa juhtimise eesmärgil

Näitajate ja mõõdikute visualiseerimiseks kasutatakse mitmesuguseid graafikuid ja graafikuid, mis tavaliselt kombineeritakse andmeväljapanekuteks ja armatuurlaudadeks. Need armatuurlaudad annavad tsentraliseeritud ülevaate projekti või projekti osade seisust, näidates põhinäitajaid (ideaalis reaalajas). Ajakohastatud ja pidevalt ajakohastatavad armatuurlaudad võimaldavad meeskonnal kiiresti reageerida muutustele.

Tulemustabelid on vahendid, mis visualiseerivad kvantitatiivseid hinnanguid, muutes need kõigile projektis osalejatele kergesti kättesaadavaks ja arusaadavaks.



Joonis 7.1-6 Peamiste tulemusnäitajate haldamine ja nende visualiseerimine armatuurlaudade kaudu on projekti tootlikkuse ja tõhususe parandamise võti.

Siin on mõned näited populaarsetest tööriistadest, millega saate luua armatuurlaudad:

- **Power BI** on Microsofti vahend interaktiivsete aruannete ja armatuurlaudade loomiseks.
- **Tableau ja Google Data Studio** on võimsad tööriistad andmete visualiseerimiseks ja armatuurlaudade loomiseks ilma koodi kirjutamata.
- **Plotly** (joonis 7.1-6, joonis 7.2-12) on raamatukogu interaktiivsete graafikute loomiseks ja Dash on raamistik andmete analüüsimiseks mõeldud veebirakenduste loomiseks. Neid saab kasutada koos interaktiivsete armatuurtahvlite loomiseks.
- **Arvukad Pythoni raamatukogud** (joonis 7.2-9 - joonis 7.2-11) - Pythonis on palju avatud lähtekoodiga ja tasuta raamatukogusid andmete visualiseerimiseks, näiteks Matplotlib, Seaborn, Plotly, Bokeh ja teised. Neid saab kasutada graafikute loomiseks ja nende integreerimiseks veebirakendusse, kasutades selliseid raamistikke nagu Flask või Django.
- **JavaScripti raamatukogud:** võimaldab luua interaktiivseid armatuurlaudu, kasutades avatud lähtekoodiga JavaScripti raamatukogusid, nagu D3.js või Chart.js, ja integreerida need veebirakendusse.

Peamiste tulemusnäitajate hindamiseks ja ülevaatlilike tahvlite loomiseks on vaja ajakohaseid andmeid ja selget ajakava teabe kogumiseks ja analüüsimiseks.

Üldiselt on ehitussektoris analüütilise lähenemise aluseks projektijuhtimise analüütiline lähenemine, mille aluseks on põhinäitajad, ROI ja armatuurtahvlid. Need ei aita mitte ainult jälgida ja hinnata hetkeolukorda, vaid annavad ka väärtuslikke teadmisi tulevaste planeerimis- ja optimeerimisprotsesside jaoks - protsessid, mis sõltuvad otseselt andmete tõlgendamisest ning õigete ja õigeaegsete küsimuste esitamisest.

Andmeanalüüs ja küsimuste esitamise kunst

Andmete tõlgendamine on analüüsi viimane etapp, kus teave saab mõtestatud ja hakkab "rääkima". See on koht, kus sõnastatakse vastused põhiküsimustele: *"mida teha?"* ja *"kuidas teha?"* (joonis 2.2-5). Selles etapis on võimalik teha kokkuvõtteid tulemustest, tuvastada mustreid, luua põhjus-tagajärg seoseid ja teha järeldusi visualiseerimise ja statistilise analüüsi põhjal.

Võib-olla ei ole kaugel aeg, mil hakatakse mõistma, et selleks, et saada täielikult tõhusaks kodanikuks ühes uuestest suurtest keerulistest maailmariikidest, mis praegu arenevad, on sama vajalik osata arvutada, mõelda keskmistes, maksimumides ja miinimumides, nagu praegu on vaja osata lugeda ja kirjutada [133].

- Samuel S. Wilkes, tsiteeritud 1951. aasta presidendi pöördumises Ameerika Statistikaassotsiatsioonile.

Vastavalt Ühendkuningriigi valitsuse avaldatud aruandele "Andmeanalüüs ja tehisintellekt valitsusprojektide elluviimisel" (2024) [83] võib analüütika rakendamine andmete ja tehisintellekti (AI) märkimisväärselt parandada projektijuhtimise protsesse, suurendades aja- ja kuluproгноoside täpsust ning vähendades riske ja ebakindlust. Dokumendis rõhutatakse, et avaliku sektori organisatsioonid, kes kasutavad täiustatud analüütilisi vahendeid, saavutavad infrastruktuuri algatustes suuremat tulemuslikkust.

Neljanda tööstusrevolutsiooni väga konkurentsitihedas ja madala marginaaliga keskkonnas tegutsevat kaasaegset ehitusettevõtlust võib võrrelda sõjaliste operatsioonidega. Siin sõltub ettevõtte ellujäämine ja edu ressursside ja kvaliteetse teabe kiirest hankimisest - ja seega õigeaegsetest ja teadlikest otsustest (joonis 7.1-7).

Kui andmete visualiseerimine on "luure", mis annab ülevaate, siis andmeanalüütika on "laskemoona", mida on vaja tegutsemiseks. See annab vastused küsimustele: *mida* ja *kuidas teha*, ning on aluseks konkurentsieelise saavutamisele turul.

Analüütika muudab erinevad andmed struktureeritud ja sisukaks teabeks, mille põhjal saab teha otsuseid.

Analüütikute ja juhtide ülesanne ei ole lihtsalt tõlgendada teavet, vaid pakkuda teadlikke otsuseid, tuvastada suundumusi, määrata kindlaks seosed eri liiki andmete vahel ja liigitada neid vastavalt projekti eesmärkidele ja spetsiifikale. Kasutades visualiseerimisvahendeid ja statistilise analüüsi

meetodeid, muudavad nad andmed ettevõtte strateegiliseks varaks.



Joonis 7.1-7 Andmete analüüs on see, mis lõpuks muudab kogutud teabe otsuste tegemise allikaks.

Selleks, et teha tõeliselt teadlikke otsuseid analüüsiprotsessis, on vaja õppida, kuidas õigesti sõnastada küsimustele, mida esitatakse andmetele. Nende küsimuste kvaliteet mõjutab otseselt saadud arusaamade sügavust ja sellest tulenevalt juhtimisotsuste kvaliteeti.

Minevik eksisteerib ainult niivõrd, kuivõrd see on olemas tänastes dokumentides. Ja seda, mida need andmed kujutavad, määravad küsimused, mida me esitame. Ei ole olemas muud ajalugu kui see [134].

- John Archibald Wheeler, füüsik 1982

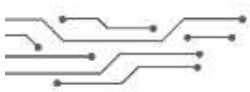
Sügavate küsimuste esitamise ja kriitilise mõtlemise kunst on andmetega töötamisel oluline oskus. Enamik inimesi kipub küsima lihtsaid, pealiskaudseid küsimusi, mille vastamine nõuab vähe pingutust. Tõeline analüüs algab aga mõtestatud ja läbimõeldud küsimustega, mis võivad paljastada varjatud seoseid ja põhjus-tagajärg seoseid teabes, mis võib olla peidetud mitme arutluskihi taha.

Uuringu "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135] kohaselt nõuab edukas digitaalne ümberkujundamine investeringuid analüütilisse võimekusse, muudatuste juhtimise programmidesse ja ärieesmärkide kooskõlastamist IT-algatustega. Ettevõtted, kes loovad andmepõhise kultuuri, peaksid investeerima andmeanalüüsi võimekusse ja käivitama muutuste juhtimise programme, et juurutada uut mõtlemist, käitumist ja tööviise.

Ilma investeringuteta analüütilise kultuuri arendamisse, andmevahendite täiustamisse ja spetsialistide koolitamisest riskivad ettevõtted jätkuvalt sellega, et nad teevad otsuseid, mis põhinevad vananenud või ebatäielikul teabel - või tuginevad HiPPO juhtide subjektiivsetele arvamustele (joonis 2.1-9).

Analüütika asjakohasuse ja pideva ajakohastamise vajaduse mõistmine ja armatuurlauad viivad paratamatult juhtkonna arusaamisele analüütiliste protsesside automatiseerimise olulisusest. Automatiseerimine suurendab otsuste tegemise kiirust, vähendab inimteguri mõju ja tagab andmete asjakohasuse. Teabemahtude eksponentsiaalse kasvuga muutub kiirus mitte ainult konkurentsieeliseks, vaid jätkusuutliku edu võtmeteguriks.

Andmete analüüsi- ja töötlemisprotsesside automatiseerimine üldiselt on lahutamatult seotud ETLi teemaga (Extract, Transform, Load). Nii nagu automatiseerimisprotsessis tuleb andmeid teisendada, nii ka ETL-protsessis ekstraheeritakse andmed erinevatest allikatest, teisendatakse vastavalt vajalikele nõuetele ja laetakse edasiseks kasutamiseks sihtsüsteemidesse.



PEATÜKK 7.2.

ANDMEVOOG ILMA MANUAALSE VAEVATA: MIKS ON VAJA ETL

ETL automatiseerimine: kulude vähendamine ja andmekäitluse kiirendamine

Kui peamised tulemusnäitajad (KPId) ei kasva enam, vaatamata andmemahu ja meeskonna suuruse suurenemisele, jõuab ettevõtte juhtkond paratamatult arusaamisele, et protsessid tuleb automatiseerida. Varem või hiljem saab see arusaam ajendiks käivitada kompleksne automatiseerimine, mille peamine eesmärk on vähendada protsesside keerukust, kiirendada töötlemist ja vähendada sõltuvust inimtegurist.

McKinsey uuringu "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136] kohaselt saavad voogandmete arhitektuuri kasutavad ettevõtted märkimisväärse eelise, sest nad saavad teavet reaajas analüüsida. Voogtehnoloogiad võimaldavad reaajas sõnumite otsest analüüsi ja prognoosiva hoolduse rakendamist tootmises, analüüsides reaajas andurite andmeid.

Protsesside lihtsustamine on automatiseerimine, kus traditsioonilised käsitsi tehtavad funktsioonid asendatakse algoritmide ja süsteemidega.

Automatiseerimine või õigemini "inimeste rolli minimeerimine andmetöötluses" on iga ettevõtte jaoks pöördumatu ja väga tundlik protsess. Mis tahes eriala spetsialistid kõhklevad sageli oma töömeetodite ja peensuste täielikku avalikustamist optimeerijatele, mõistes ohtu kaotada oma töökoht kiiresti arenevas tehnoloogilises keskkonnas.

Kui tahad endale vaenlasi teha, siis proovi asju muuta [137].

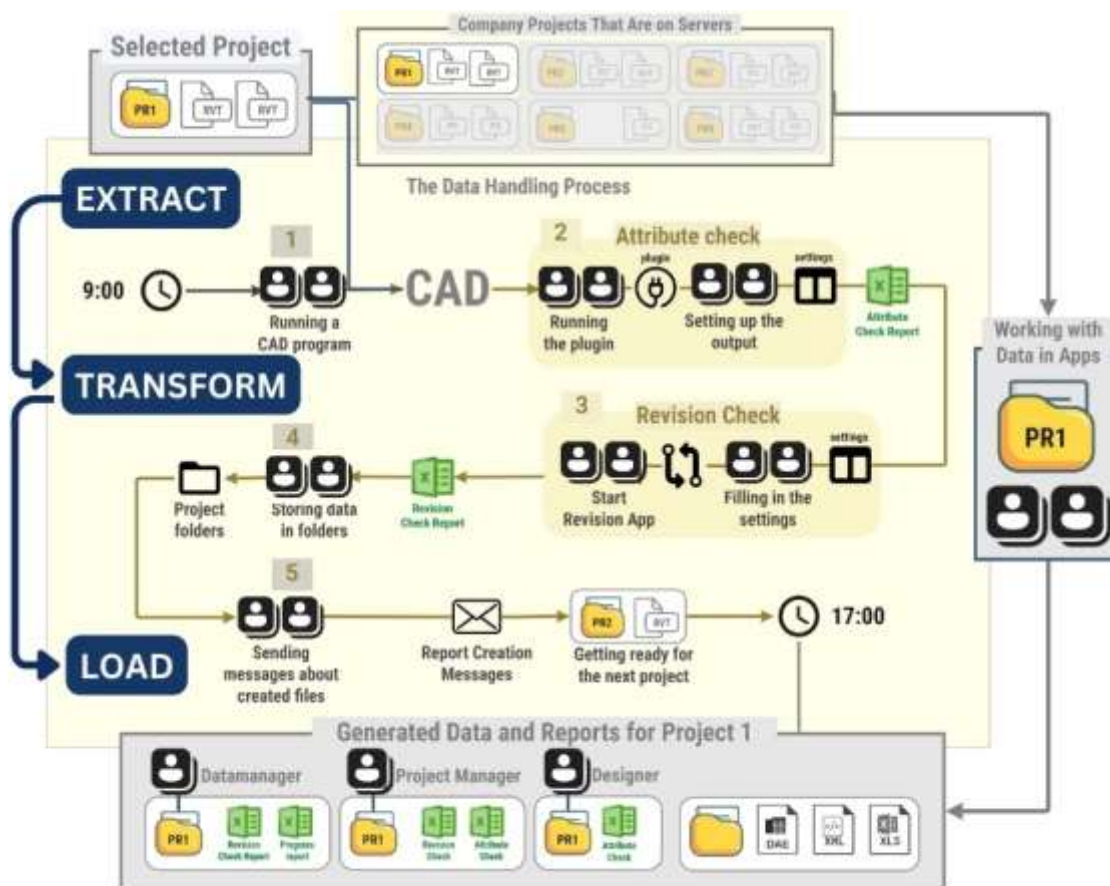
- Woodrow Wilson, kõne müügimeeste kongressil, Detroit, 1916

Vaatamata automatiseerimise ilmsetele eelistele on paljude ettevõtete igapäevatöös endiselt suur osakaal manuaalset tööd, eriti tehniliste andmete valdkonnas. Praeguse olukorra illustreerimiseks vaadake tüüpilist näidet järjestikuse andmetöötluse kohta sellistes protsessides.

Manuaalset andmetöötlust saab illustreerida CAD-andmebaasidest saadud teabe koostoitimise näitel. Traditsiooniline andmetöötlus ("käsitsi" ETL -protsess) CAD (BIM) osakondades atribuutide tabelite loomiseks või projekteerimisandmete põhjal dokumentatsiooni loomiseks toimub järgmises järjekorras (joonis 7.2-1):

1. Manuaalne **väljavõtte tegemine (Extract)**: kasutaja avab projekti käsitsi - käivitades CAD-rakenduse (BIM) (joonis 7.2-1 samm 1).
2. **Kontrollimine**: järgmine samm hõlmab tavaliselt mitmete pluginate või abirakenduste käsitsi käivitamist andmete ettevalmistamiseks ja nende kvaliteedi hindamiseks (joonis 7.2-1 samm

- 2-3).
3. Manuaalne **teisendamine (Transform)**: pärast ettevalmistamist algab andmete töötlemine, mis nõuab erinevate tarkvaravahendite käsitsi kasutamist, milles andmed valmistatakse ette üleslaadimiseks (joonis 7.2-1 samm 4).
 4. Käsitsi **üleslaadimine (Load)**: teisendatud andmete käsitsi üleslaadimine välissüsteemidesse, andmeformaatidesse ja dokumentidesse (joonis 7.2-1 samm 5).



Joonis 7.2-1 Traditsiooniline käsitsi töötlemine ETL on piiratud üksikute tehnikute soovide ja füüsiliste võimalustega.

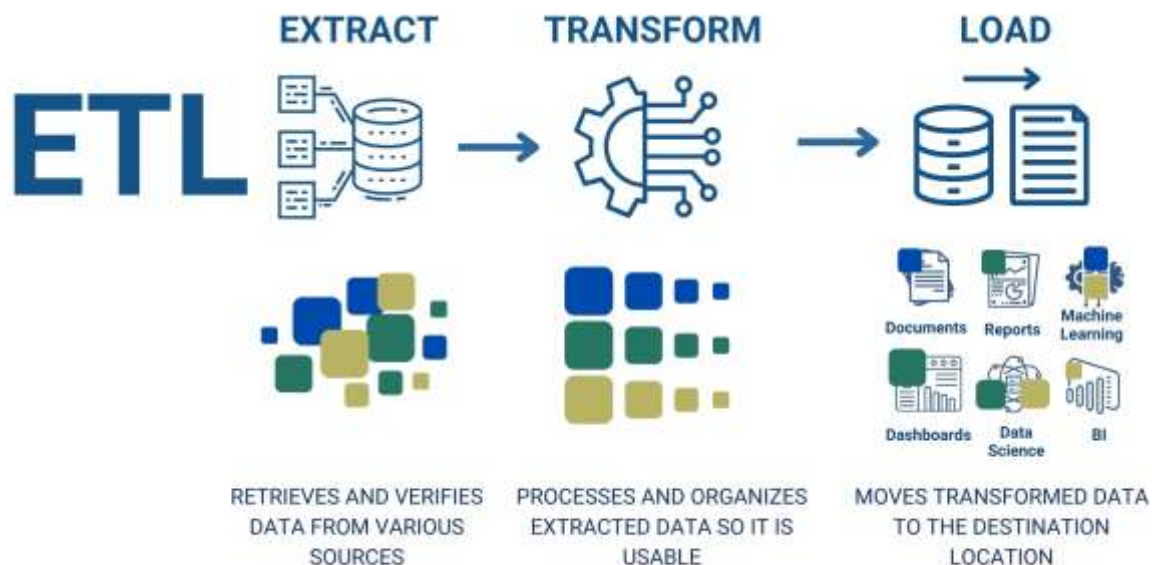
Selline töövoog on näide klassikalisest ETL -protsessist - ekstraheerimine, transformatsioon ja laadimine (ETL). Erinevalt teistest tööstusharudest, kus automaatsed ETL-putkad on juba ammu standardiks, domineerib ehitustööstuses ikka veel käsitsi töötamine, mis aeglustab protsesse ja suurendab kulusid

ETL (Extract, Transform, Load) on protsess, mille käigus andmed ekstraheeritakse erinevatest allikatest, muudetakse soovitud vormingusse ja laaditakse edasiseks analüüsiks ja kasutamiseks sihtsüsteemi.

ETL on protsess, mis tähistab kolme peamist andmetöötamise komponenti: Extract, Transform ja Load (joonis 7.2-2):

- Väljavõte - andmete väljavõtmine erinevatest allikatest (failid, andmebaasid, API).
- **Transform** - andmete puhastamine, koondamine, normaliseerimine ja loogiline töötlemine.
- **Laadige** - laadige struktureeritud teavet andmelattu, aruandesse või BI-süsteemi.

Varasemalt on ETL-i mõistet käsitletud vaid aeg-ajalt: struktureerimata skaneeritud dokumendi teisendamisel struktureeritud tabeliformaati (joonis 4.1-1), nõuete formaliseerimise kontekstis, et süstematiseerida nii elu- kui ka äriprotsesside tajumist (joonis 4.4-20), ning andmete valideerimise ja andmete töötlemise automatiseerimisel CAD-lahendustest. Vaatleme nüüd ETLi üksikasjalikumalt tüüpiliste töövoogude kontekstis.



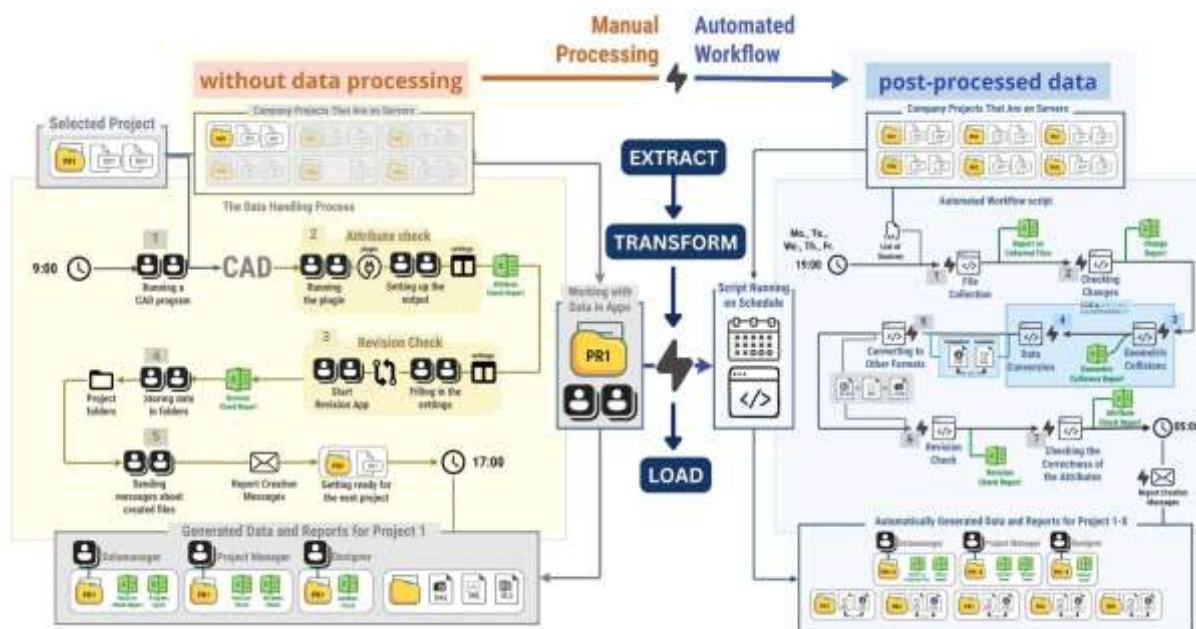
Joonis 7.2-2 ETL automatiseerib korduvaid andmetöötlusülesandeid.

Manuaalne või poolautomaatne ETL -protsess eeldab, et juht või tehnik juhib kõiki etappe käsitsi - alates andmete kogumisest kuni aruande koostamiseni. Selline protsess võtab palju aega, eriti kui tööaeg on piiratud (nt 9:00-17:00).

Sageli püüavad ettevõtted lahendada madala tõhususe ja aeglase kiiruse probleemi, ostes moodulitest koosnevaid integreeritud lahendusi (ERP, PMIS, CPM, CAFM jne), mida seejärel arendavad edasi välised müüjad ja konsultandid. Kuid need müüjad ja kolmandate isikute arendajad muutuvad sageli kriitiliseks sõltuvuspunktiks: nende tehnilised piirangud mõjutavad otseselt kogu süsteemi ja ettevõtte kui terviku tulemuslikkust, nagu on üksikasjalikult kirjeldatud eelmistes peatükkides, mis käsitlevad patenteeritud süsteeme ja vorminguid. Killustatusest ja sõltuvusest tulenevaid probleeme käsitleti üksikasjalikult peatükis "Kuidas ehitusettevõtted uppuvad andmekaosesse".

Kui ettevõtte ei ole valmis rakendama ühe müüja suurt moodulplatvormi, hakkab ta otsima alternatiivseid võimalusi automatiseerimiseks. Üks neist on arendada oma modulaarseid avatud ETL -konveierid, kus iga etapp (ekstraheerimine, transformatsioon, valideerimine, laadimine) on rakendatud skriptidena, mida täidetakse graafiku alusel.

Sama ETL-tööprotsessi automatiseeritud versioonis (joonis 7.2-1) näeb tööprotsess välja nagu moodulkood, mis algab andmete töötlemisega ja nende tõlkimisega avatud struktureeritud vormi. Kui struktureeritud andmed on saadud, käivitatakse automaatselt ja plaanipäraselt erinevaid skripte või moduleid, et kontrollida muudatusi, teisendada ja saata teateid (joonis 7.2-3).



Joonis 7.2-3 Vasakul on käsitsi töötlemine, paremal on automaatne protsess, mis erinevalt traditsioonilisest käsitsi töötlustest ei ole piiratud kasutaja võimalustega.

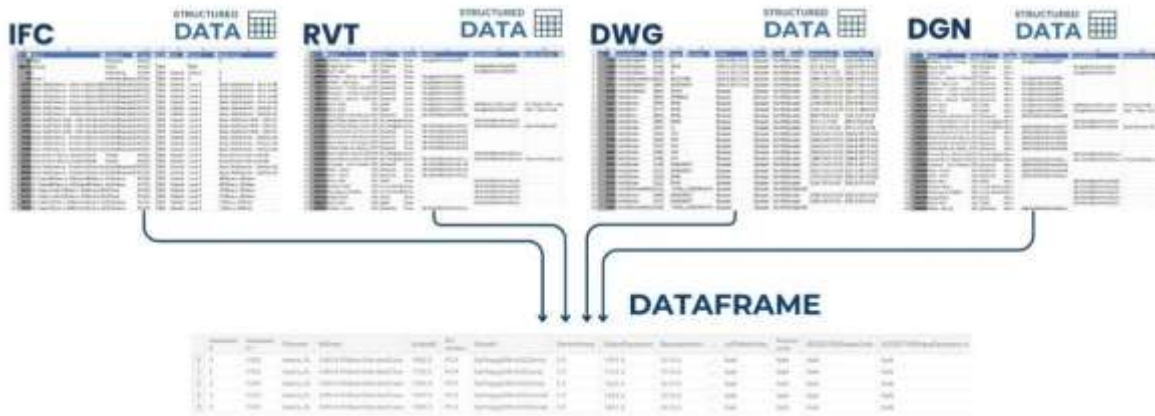
Automatiseeritud tööprotsessis lihtsustab andmete töötlemist ET(L) andmete eeltöötlus: struktureerimine ja ühendamise.

Traditsiooniliste töötlemismeetodite puhul töötavad spetsialistid andmetega "nii nagu need on" - nii nagu need on süsteemidest või tarkvarast välja otsitud. Seevastu automatiseeritud protsessides läbivad andmed sageli kõigepealt ETL -plaani, kus need viiakse ühtsesse struktuuri ja vormingusse, mis sobib edasiseks kasutamiseks ja analüüsiks.

Võtame praktilise ETLi näite, mis demonstreerib peatükis " andmete valideerimine ja valideerimistulemused" kirjeldatud andmetabeli valideerimisprotsessi (joonis 4.4-13). Selleks kasutame Pandas raamatukogu koos LLMiga andmete automatiseeritud analüüsi- ja töötlemisprotsesside jaoks.

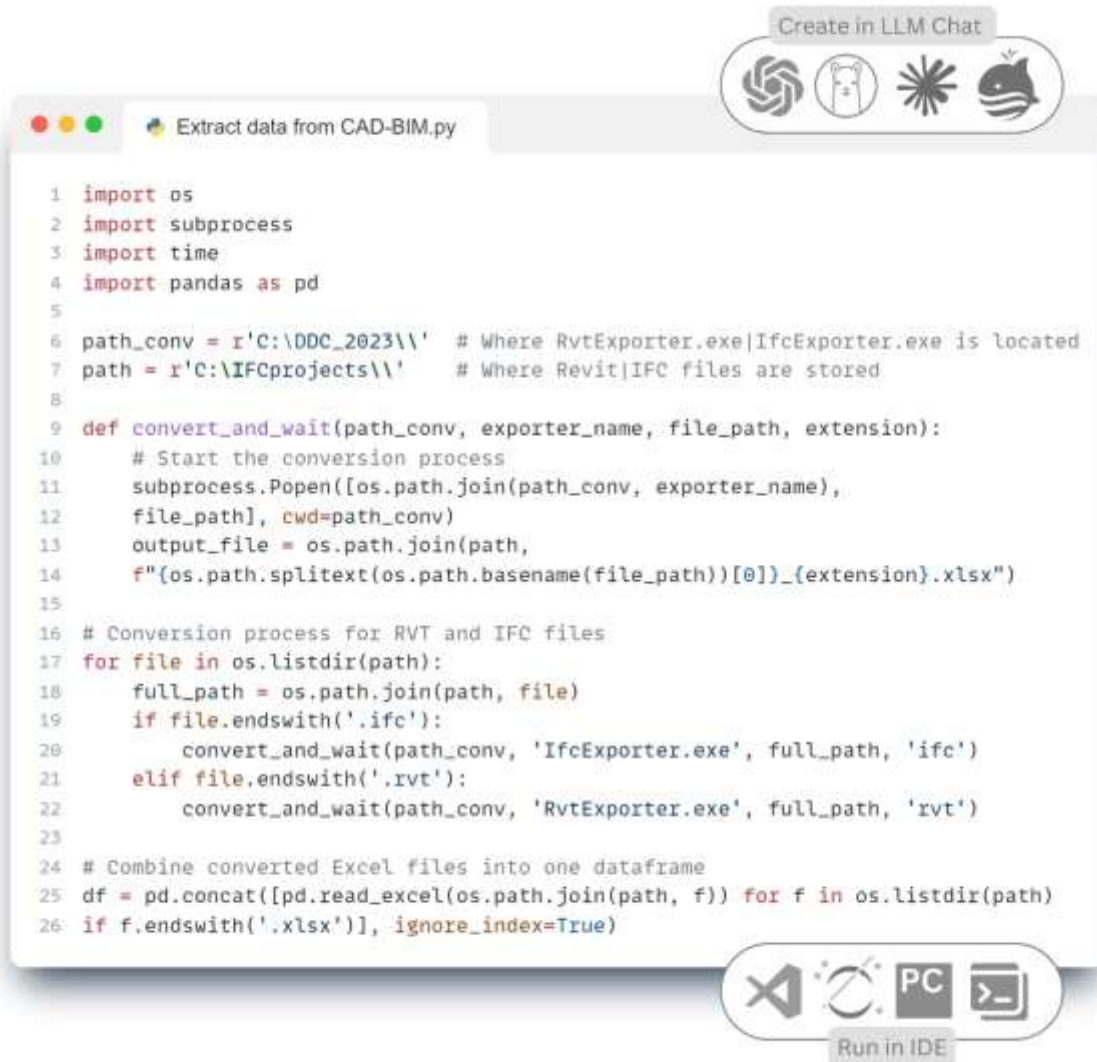
ETL Extract: andmete kogumine

ETL protsessi esimene etapp - Extract) - algab koodi kirjutamisega, et koguda andmekogumeid edasiseks kontrollimiseks ja töötlemiseks. Selleks skaneeritakse kõik tootmisserveri kaustad, kogutakse kindla formaadi ja sisuga dokumendid ning teisendatakse need seejärel struktureeritud kujul. Seda protsessi käsitletakse üksikasjalikult peatükkides "Struktureerimata ja tekstiliste andmete teisendamine struktureeritud kujul" ja "CAD-andmete (BIM) teisendamine struktureeritud kujul" (joonis 4.1-1 - joonis 4.1-12).



Joonis 7.2-4 Konverteeri CAD-andmed (BIM) üheks suureks andmeraamistikuks, mis sisaldab kõiki projektilõike.

Illustreeriva näitena kasutame andmete laadimise sammu Extract ja saame kõigi CAD- (BIM-) projektide tabeli (joonis 7.2-4), kasutades RVT- ja IFC-formaatide jaoks pöördprojekteerimist võimaldavaid konverteereid [138], et saada kõigist projektidest struktureeritud tabelid ja ühendada need üheks suureks DataFrame-tabeliks.



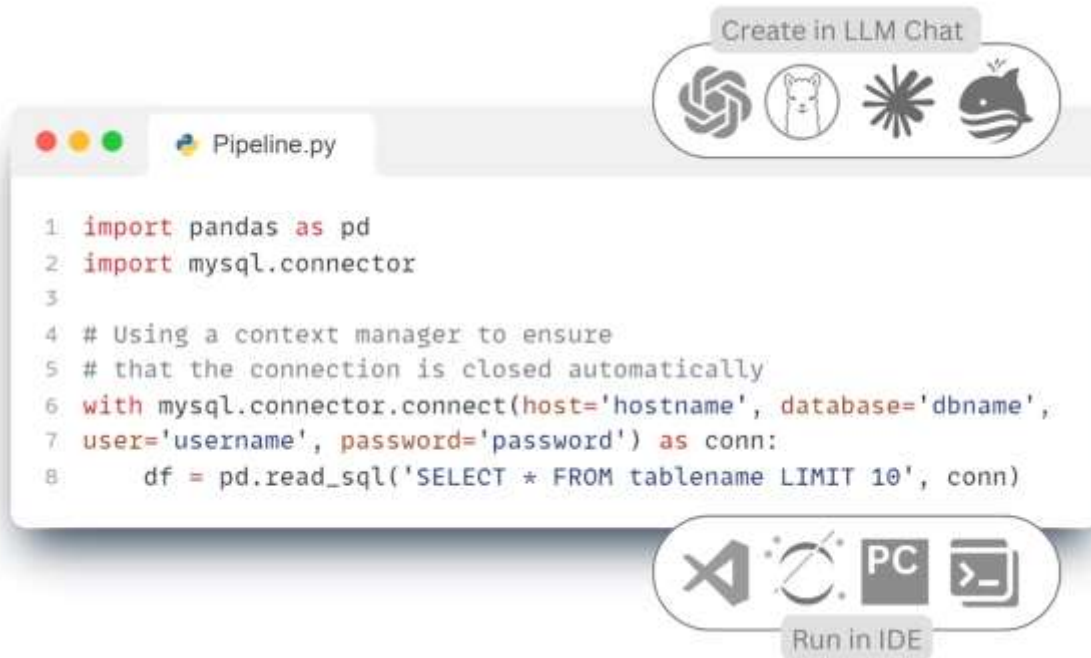
Joonis 7.2-5 Konvertimine Python koodi ja SDK pöördtehnoloogiavahendi abil RVT ja IFC failid üheks suureks struktureeritud (df) DataFrame.

Pandas DataFrame saab laadida andmeid erinevatest allikatest, sealhulgas CSV tekstifailidest, Exceli tabelitest, JSON - ja XML - failidest, suurte andmete salvestusformaadidest nagu Parquet ja HDF5 ning MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle ja muudest andmebaasidest. Lisaks toetab Pandas andmete laadimist APIdest, veebilehtedest, pilveteenustest ja salvestussüsteemidest, nagu Google BigQuery, Amazon Redshift ja Snowflake.

- ☛ Selleks, et kirjutada koodi andmebaasidega ühenduse loomiseks ja andmete kogumiseks, saatke sarnane tekstitaotlus LLM jututuppa (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN või mõni muu):

Palun kirjutage näide MySQLiga ühendamise ja andmete teisendamise kohta. ↩

LLMi vastus :



Joonis 7.2-6 Näide Pythoni kaudu MySQL andmebaasi ühendamise ja andmete importimisest MySQL andmebaasist DataFrame'i .

Saadud kood (joonis 7.2-5, joonis 7.2-6) saab käivitada mõnes populaarses IDE-s (integreeritud arenduskeskkonnas), mida me eespool mainisime, offline-režiimis: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse koos PyDev pluginaga, Thonny, Wing IDE, IntelliJ IDEA koos Python pluginaga, JupyterLab või populaarsed online-vahendid: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Laadides multiformaadi andmed muutujasse "df" (joonis 7.2-5 - rida 25; joonis 7.2-6 - rida 8), teisendasime andmed Pandas DataFrame formaati, mis on üks populaarsemaid andmetöötluste struktuure, mis on ridade ja veergudega kahemõõtmeline tabel. Teistest ETL -ipeliinides kasutatavatest salvestusformaatidest, nagu Parquet, Apache ORC, JSON, Feather, HDF5, samuti kaasaegsetest andmeladudest räägime lähemalt peatükis "Andmete salvestamine ja haldamine ehituses" (joonis 8.1-2).

Pärast andmete ekstraheerimise ja struktureerimise etappi (Extract) moodustub ühtne teabemassiiv (joonis 7.2-5, joonis 7.2-6), mis on valmis edasiseks töötlemiseks. Enne nende andmete laadimist sihtsüsteemidesse või nende kasutamist analüüsiks tuleb siiski tagada nende kvaliteet, terviklikkus ja vastavus kindlaksmääratud nõuetele. Selles etapis toimub andmete ümberkujundamine (Transform), mis on oluline samm edasiste järeltöötluste ja otsuste usaldusväärsuse tagamiseks.

ETL Transform: valideerimis- ja transformatsioonireeglite rakendamine

Transform on samm, kus andmeid töödeldakse ja teisendatakse. See protsess võib hõlmata õigsuse kontrollimist, normaliseerimist, puuduvate väärtuste täitmist ja valideerimist automatiseeritud vahendite abil

Vastavalt PwC uuringule "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] on kaasaegsed audiitorfirmad eemaldunud juhuslikust andmete kontrollimisest ja liiguvad massilise hulga teabe analüüsimise suunas, kasutades selleks automatiseeritud vahendeid. Selline lähenemine võimaldab mitte ainult tuvastada lahknevusi aruandluses, vaid ka pakkuda soovitusi äriprotsesside optimeerimiseks.

Ehituses saab sarnaseid meetodeid kasutada näiteks projekteerimisandmete automaatseks valideerimiseks, ehituskvaliteedi kontrollimiseks ja töövõtjate tegevuse hindamiseks. Üks vahend andmetöötluse automatiseerimiseks ja kiirendamiseks on regulaaravaldiste (Regex) kasutamine ETL-protsessi andmetöötlusetapis (Transform). Regex võimaldab tõhusalt valideerida andmesõnu, tuvastada vastuolusid ja tagada teabe terviklikkus minimaalse ressursikuluga. RegExi (joonis 4.4-7) kohta räägiti lähemalt peatükis "Nõuete teisendamine struktureeritud kujul".

Vaatleme praktilist näidet: objektide haldussüsteemis (RPM) seab haldaja objektide põhiomadustele nõuded (joonis 7.2-7). Ümberkujundamise etapis tuleb kinnitada järgmised parameetrid:

- objekti identifikaatori vormingute kontrollimine (atribuut "ID")
- garantiiperioodi asendusväärtuste kontrollimine (atribuut "Garantiiperiood")
- elemendi asendustsükli kontrollimine (atribuut "hooldusnõuded")



Property Manager:
Long-term Management

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 Varnishing

Joonis 7.2-7 Kontrollimine kvaliteet algab atribuutide ja nende piirväärtuste nõuete kehtestamisega.

Parameetri valideerimise piirväärtuste määramiseks oletame näiteks, et me teame oma kogemustest, et atribuudi "ID" kehtivad väärtused võivad sisaldada ainult string-väärtusi "W-NEW", "W-OLD1" või "D-122" või sarnaseid väärtusi, kus esimene täht on täht, millele järgneb kriips ja seejärel kolm tähemärki "NEW", "OLD" või mõni kolmekohaline number (joonis 7.2-7). Nende identifikaatorite valideerimiseks

saab kasutada järgmist regulaaravaldist (RegEx):

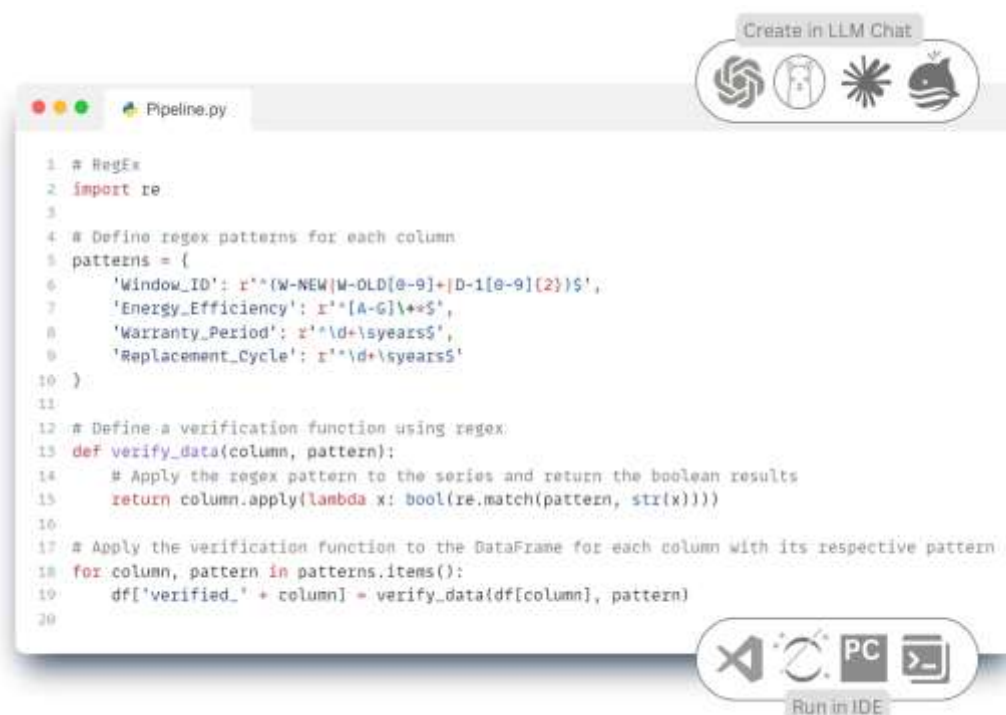
```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

See mall tagab, et kõik andmetes olevad identifikaatorid vastavad kindlaksmääratud kriteeriumidele. Kui mõni väärtus ei vasta kontrollile, registreerib süsteem veateate. Andmete teisendamiseks Python-koodi loomiseks ja saadud andmete kasutamiseks tulemustabeli loomiseks tuleb lihtsalt sõnastada päring LLM vestluses.

💬 Tekstitaotlus aadressile LLM:

Kirjutage kood DataFrame'i veergude valideerimiseks regulaaravaldiste abil, mis kontrollib RegExi abil identifikaatoreid formaadis "W-NEW" või "W-OLD", energiatõhusust tähtedega "A" kuni "G", garantiiaga ja asendustsükli numbriliste väärtustega aastates. 📄

📄 LLMi vastus :



```
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}$',
7     'Energy_Efficiency': r'^[A-G]\+\+$',
8     'Warranty_Period': r'^\d+\syears$',
9     'Replacement_Cycle': r'^\d+\syears$'
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
```

Joonis 7.2-8 Kood automatiseerib valideerimisprotsessi, rakendades RegEx-malle andmeraami parameetri veergudele .

Ülaltoodud, automaatselt genereeritud Python -kood (joonis 7.2-8) kasutab "re" raamatukogu (RegEx regular expressions), et defineerida funktsioon, mis kontrollib DataFrame andmemeendi iga

atribuuti. Iga määratud veeru (atribuudi) puhul rakendab funktsioon RegEx-mustrit, et kontrollida, kas iga kirje vastab oodatavale vormingule, ja lisab tulemused uute väärtustena (False/True) DataFrame'i uude atribuut-sambasse.

Selline automatiseeritud valideerimine tagab andmete ametliku vastavuse ja seda saab kasutada kvaliteedikontrolli süsteemi osana ümberkujundamise etapis.

Pärast seda, kui transform ja kvaliteedikontroll on edukalt lõpule viidud, on andmed valmis üleslaadimiseks sihtsüsteemidesse. Transformeeritud ja valideeritud andmeid saab edasiseks kasutamiseks üles laadida CSV, JSON, Excel, andmebaasidesse ja muudesse vormingutesse. Sõltuvalt ülesandest saab tulemusi esitada ka aruannetes, graafikutes või analüütilistes armatuurtahvlites.

ETL Load: Visualiseeri tulemused graafikutes ja graafikutes.

Pärast Transform etapi lõpetamist, kui andmed on struktureeritud ja kontrollitud, on viimane etapp Load, kus andmeid saab nii laadida sihtsüsteemi kui ka visualiseerida analüüsiks. Andmete visuaalne esitamine võimaldab kiiresti tuvastada kõrvalekaldeid, analüüsida jaotusi ja edastada peamised järeldused kõigile projektis osalejatele, sealhulgas neile, kellel puudub tehniline taust.

Selle asemel, et esitada teavet tabelite ja numbritena, saame kasutada infograafikat, graafikuid ja armatuurlaudu (dashboard). Üks levinumaid ja paindlikumaid vahendeid struktureeritud andmete visualiseerimiseks Pythonis on Matplotlib'i raamatukogu (joonis 7.2-9, joonis 7.2-10). See võimaldab luua staatilisi, animeeritud ja interaktiivseid graafikuid ning toetab mitmesuguseid graafikutüüpe.

- 📌 RPM-süsteemi (joonis 7.2-7) atribuutide kontrollimise tulemuste visualiseerimiseks saate kasutada järgmist päringut keelemudelile:

Kirjutage kood eespool esitatud DataFrame andmete visualiseerimiseks (joonis 7.2-7) koos tulemuste histogrammiga, mis näitab vigade sagedust atribuudis. ↗

- LLM vastus koodina ja valmis visualiseerimine otse LLMi jututuppa koodi täitmise tulemuste kohta :

[Create in LLM Chat](#)

●
●
●
Pipeline.py

```

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()

```

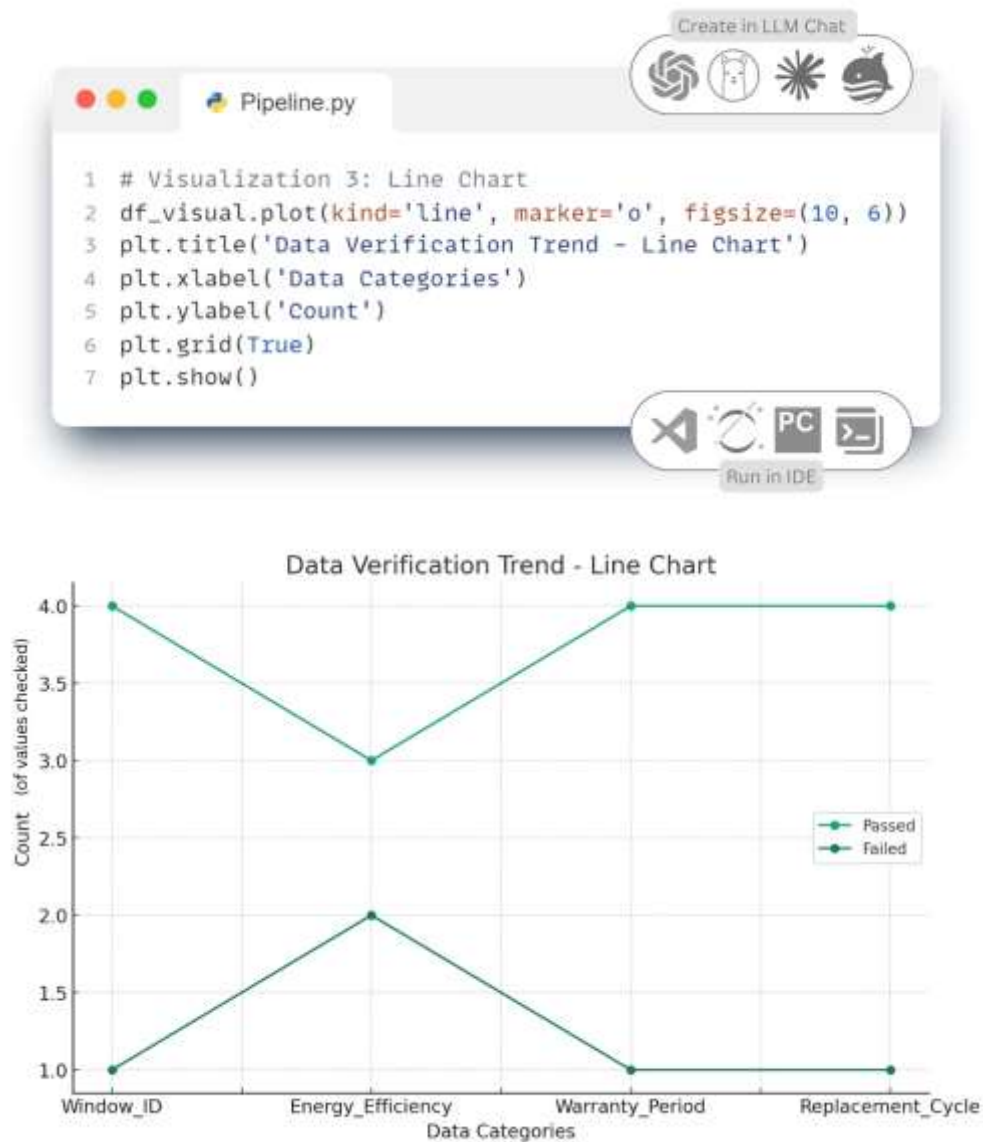


Joonis 7.2-9 Visualiseerimine transformeerimise sammu tulemuste visualiseerimine (joonis 7.2-7) atribuutide väärtuste kontrollimisel pöörlemissüsteemist kui histogramm koormuse sammus .

- 🗨️ On olemas palju avatud lähtekoodiga ja tasuta visualiseerimisraamatukogusid, mis võimaldavad esitada struktureeritud andmeid erinevates vormingutes. Jätkame tulemuste visualiseerimist teist tüüpi graafiku abil, mille kohta on järgmine üleskutse vestluses:

Joonistage samad andmed joonte graafikuna. ↩

🗨️ LLMi vastus :



Joonis 7.2-10 Valideerimisandmete visualiseerimine (joonis 7.2-8) joondigrammina, mis on saadud Matplotlib'i raamatukogu abil.

On olemas palju avatud lähtekoodiga ja tasuta visualiseerimisraamatukogusid, näiteks:

- Seaborn - statistiliste graafikute jaoks (joonis 7.2-11).
- Plotly - interaktiivsete veebi visualiseerimiste jaoks (joonis 7.2-12, joonis 7.1-6).
- Altair - deklaratiivseks visualiseerimiseks
- Dash või Streamlit - täielike armatuurlaudade loomiseks

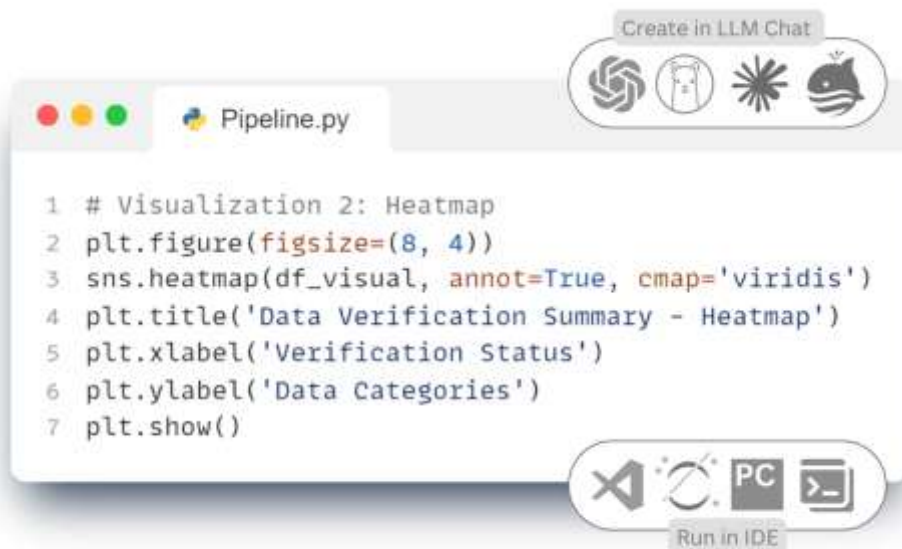
Konkreetsete visualiseerimise raamatukogude tundmine ei ole vajalik - kaasaegsed vahendid, sealhulgas LLM, võimaldavad automaatselt genereerida graafikute ja tervete rakenduste koodi ülesande kirjelduse põhjal.

Tööriista valik sõltub projekti eesmärkidest: kas tegemist on aruande, esitluse või veebipõhise armatuurlauaga. Näiteks avatud lähtekoodiga raamatukogu Seaborn on eriti hea kategooriliste andmetega töötamiseks, aidates tuvastada mustreid ja suundumusi.

- 🗉 Seaborn'i raamatukogu töötavaid raamatukogusid saab vaadata, siis võib soovitud raamatukogu kasutada kas otse LLM-i või saata sarnane tekstitaotlus LLM-i aadressil:

Näita soojuskaarti ↗

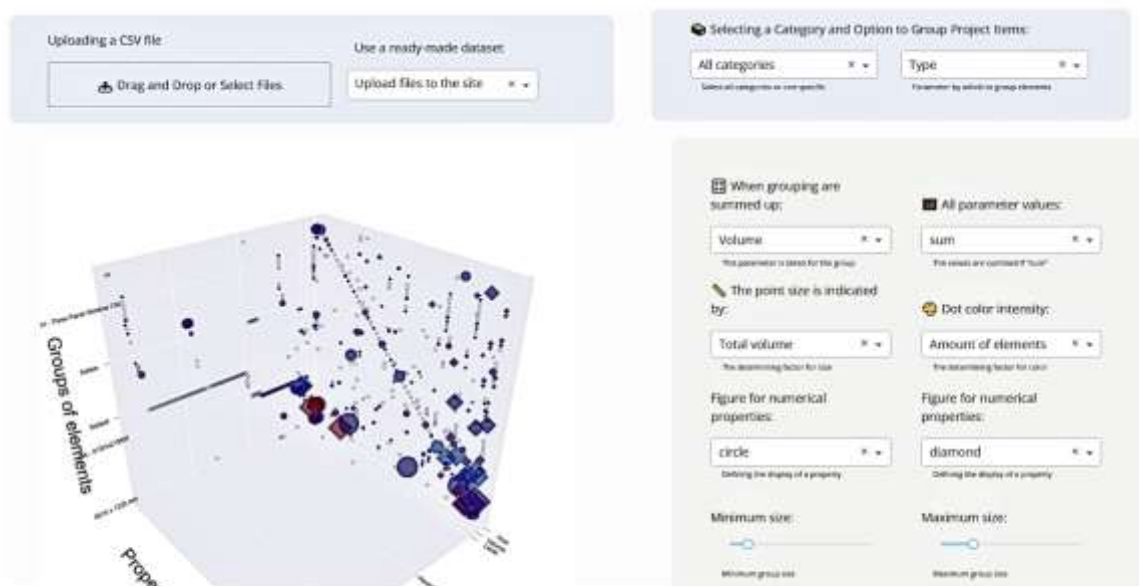
- 🗉 LLM-i vastus koodi ja valmis graafiku kujul, mille joonistamise koodi saab nüüd IDE-sse kopeerida ja graafiku ise saab kopeerida või salvestada, et seda dokumendisse kleepida:





Joonis 7.2-11 andmete valideerimise tulemuste visualiseerimine (joonis 7.2-8), kasutades Seaborn raamatukogu.

Nende jaoks, kes eelistavad interaktiivset lähenemist, on olemas vahendid, mis võimaldavad luua dünaamilisi skeeme ja paneele, mis võimaldavad interaktiivsust. Raamatukogu Plotly (joonis 7.1-6, joonis 7.2-12) pakub võimalust luua väga interaktiivseid diagramme ja paneele, mida saab põimida veebilehtedesse ja mis võimaldavad kasutajal reaalajas andmetega suhelda.



Joonis 7.2-12 Interaktiivne 3D -elemendi atribuutide visualiseerimine CAD- (BIM-) projektist, kasutades Plotly raamatukogu.

Spetsiaalsed avatud lähtekoodiga raamatukogud Bokeh, Dash ja Streamlit pakuvad mugavat viisi andmete esitamiseks, ilma et oleks vaja sügavaid teadmisi veebiarendusest. Bokeh sobib keeruliste interaktiivsete graafikute koostamiseks, Dash'i abil saab luua täisväärtuslikke analüütilisi armatuurtahtleid ja Streamlit võimaldab kiiresti luua andmeanalüüsi veebirakendusi.

Selliste visualiseerimisvahendite abil saavad arendajad ja analüütikud tõhusalt levitada tulemusi kolleegidele ja sidusrühmadele, võimaldades intuiitivset suhtlemist andmetega ja lihtsustades otsuste tegemist.

ETL Load: PDF dokumentide automaatne loomine

Andmete laadimise etapis saate mitte ainult visualiseerida andmeid, laadida neid tabelitesse või andmebaasidesse, vaid ka automaatselt genereerida aruandeid, sealhulgas vajalikke graafikuid, graafikuid ja peamisi analüütilisi näitajaid, mida juht või spetsialist saab auditi tulemusi oodates. Automatiseeritud aruanded võivad sisaldada nii kommentaare kui ka andmete ja visualiseerimiselementide - tabelite, graafikute - tekstilist tõlgendust.

- ☞ Selleks, et luua PDF -aruanne koos histogrammiga (joonis 7.2-9) ja analüüsi kirjeldus, mis põhineb eelmistes peatükkides tehtud valideerimisel, piisab, kui sõnastada taotlus dialoogis jätkates LLM-i, nt:

Kirjutage kood, et luua PDF -fail koos histogrammi ja kirjeldusega andmete valideerimise tulemuste kohta eespool (vestluses) ning kirjutada tekstiline hoiatus, et mõned kategooriad ei ole valideeritud ja et puuduvad tuleb täita. ↵

- ☞ LLM-i vastus koodina ja valmis PDF koos tulemustega:

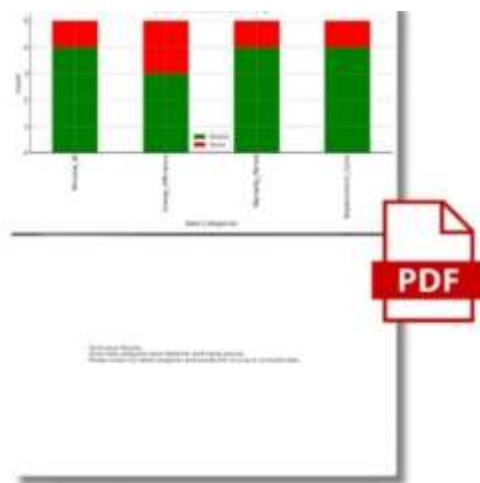
[Create in LLM Chat](#)

Pipeline.py

```

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

```



Joonis 7.2-13 Automaatne kood loob PDF-formaadis -dokumendi, mis sisaldab histogrammi koos katseandmetega ja teksti katsetulemustega.

Automaatselt kirjutatud lahendus, mis koosneb vaid 20 koodireast, kasutades LLM, loob kohe soovitud PDF (või DOC) dokumendi koos visualiseerimisega atribuutide histogrammi kujul (joonis 7.2-13), mis näitab valideerimise läbinud ja läbikukkunud andmete arvu, ning lisades tekstiploki, mis võtab tulemused kokku ja annab soovitusi edasisteks meetmeteks.

Automaatne dokumentide genereerimine on koormusetapi põhielement, eriti projektikeskkonnas, kus aruandluse kiirus ja täpsus on kriitilise tähtsusega.

ETL Load: dokumentide automaatne genereerimine FPDF-st

Aruandluse automatiseerimine ETL-etapis Load on oluline samm andmetöötluses, eriti kui analüüside tulemused tuleb esitada kergesti edastatavas ja arusaadavas vormis. Ehitustööstuses on see sageli asjakohane eduaruannete, projektiandmete statistika, kvaliteedi tagamise aruannete või finantsdokumentatsiooni puhul.

Üks kõige mugavamaid vahendeid selliste ülesannete jaoks on avatud lähtekoodiga raamatukogu, FPDF, mis on saadaval nii Pythonile kui ka PHP-le.

Avatud lähtekoodiga **FPDF**-i raamatukogu pakub paindlikku võimalust dokumentide genereerimiseks koodi abil, võimaldades lisada pealkirju, teksti, tabeleid ja pilte. Koodi kasutamine käsitsi redigeerimise asemel vähendab vigu ja kiirendab aruannete koostamist PDF-formaadis.

Üks peamisi samme PDF-dokumendi loomisel on pealkirjade ja põhiteksti lisamine kommentaaride või kirjelduste kujul. Aruande loomisel on aga oluline mitte ainult teksti lisamine, vaid ka selle õige struktureerimine. Pealkirjad, taanded, reavahe - kõik see mõjutab dokumendi loetavust. Kasutades FPDFi, saate määrata vormindamisparameetrid, kontrollida elementide paigutust ja kohendada dokumendi stiili.

FPDF on põhimõtteliselt väga sarnane HTML. Need, kes on juba tuttavad HTMLiga, saavad FPDFi abil hõlpsasti luua PDF mis tahes keerukusega dokumente, sest koodi struktuur on väga sarnane HTMLi märgendiga: pealkirjad, tekst, pildid ja tabelid lisatakse sarnasel viisil. Need, kes ei tunne HTML-i, ei pea muretsema - võite kasutada LLM-i, mis aitab teil kohe koostada koodi soovitud dokumendi kujunduse loomiseks.

- 📄 Järgnev näide näitab, kuidas luua aruande pealkirja ja tekstiga aruanne. Selle koodi täitmine mis tahes Python-toega IDE-s loob PDF -faili, mis sisaldab soovitud päist ja teksti:

```
from fpdf import FPDF    # Import FPDF raamatukogu
pdf = FPDF()            # Loo PDF -document
pdf.add_page()          # Lisa lehekülg.

pdf.set_font("Arial", style='B', size=16) # Set font: Arial, bold, suurus 16
pdf.cell(200, 10, "Projektiaruanne", ln=True, align='C') # Loo pealkiri ja tsentreeri see.
pdf.set_font("Arial", size=12) # Muuda kirjastiili tavaliseks Arial, suurus 12
pdf.multi_cell(0, 10, "See dokument sisaldab andmeid projektifaili kontrollimise tulemuste kohta . . .")
# Lisage mitmerealine tekst
pdf.output(r "C:\reports\report.pdf") # Salvesta PDF -file
```



Joonis 7.2-14 Mõne Pythoni koodireaga saame automaatselt genereerida vajaliku PDF tekstidokumendi.

Aruannete koostamisel on oluline arvestada, et andmed, millest dokument moodustatakse, on harva staatilised. Pealkirjad, tekstiplokid (joonis 7.2-14) moodustatakse sageli dünaamiliselt, saades ETL-protssis väärtused transformatsioonietapis.

Koodi kasutamine võimaldab teil luua dokumente, mis sisaldavad ajakohast teavet: projekti nimi, aruande koostamise kuupäev, samuti teave osalejate või hetkeseisu kohta. Muutujate kasutamine koodis võimaldab teil automaatselt sisestada need andmed nõutud kohtadesse aruandes, välistades täielikult vajaduse käsitsi toimetada neid enne saatmist.

Lisaks lihtsale tekstile ja pealkirjadele on projektdokumentatsioonis eriline koht tabelitel. Peaaegu iga dokument sisaldab struktureeritud andmeid: alates objektide kirjeldustest kuni inspekteerimistulemusteni. Transform etapis saadud andmete põhjal automaatselt genereeritud tabelid võimaldavad mitte ainult kiirendada dokumendi koostamise protsessi, vaid ka minimeerida

vigu teabe edastamisel. FPDF võimaldab lisada PDF-failidesse -faile tabeleid (teksti või piltidena), määrates lahtrite piire, veergude suurust ja kirjastiili (joonis 7.2-15). See on eriti mugav dünaamiliste andmetega töötamisel, kui ridade ja veergude arv võib sõltuvalt dokumendi ülesannetest muutuda.

- 📌 Järgnev näide näitab, kuidas automatiseerida tabelite loomist, nt materjalide nimekirjade, hinnangute või parameetrite katsetulemuste kohta:

```
data = [
    ["Item", "Quantity", "Price"], # Column headings
    ["Beton", "10 m³", "$ 500."], # Esimese rea andmed
    ["Rebar", "2 tonni", "$ 600"], # Teise rea andmed.
    ["Telliskivi", "5000 tk", "$ 750."], # Kolmanda rea andmed.
]

pdf = FPDF () # Loo PDF -document
pdf.add_page() # Lisa lehekülg
pdf.set_font("Arial", size=12) # Kirjastiili määramine

for row in data: # Otsi tabeli read
    for item in row: # Käi läbi rea lahtrid
        pdf.cell(60, 10, item, border=1) # Loo raamiga lahter, mille laius on 60 ja kõrgus 10.
    pdf.ln() # Liigu järgmisele reale
pdf.output(r "C:\reports\table.pdf") # Salvesta PDF -file
```

Item	Quantity	Price
Concrete	10 m³	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750



Joonis 7.2-15 Saate automaatselt genereerida mitte ainult teksti, vaid ka mistahes tabelite andmeid PDF transformeerimise sammust.

Reaalsetes aruandlusstsenaariumides on tabelid tavaliselt dünaamiliselt genereeritud teave, mis saadakse andmete teisendamise etapis. Antud näites (joonis 7.2-15) on tabel sisestatud PDF -dokumendile staatilisel kujul: näite andmed paigutati andmesõnaraamatusse (koodi esimene rida), reaalsetes tingimustes täidetakse selline andmemuutuja automaatselt pärast nt andmeraamistiku grupeerimist.

Praktikas ehitatakse sellised tabelid sageli struktureeritud andmete põhjal, mis pärinevad erinevatest dünaamilistest allikatest: andmebaasidest, Exceli -failidest, API -liidestest või analüütiliste arvutuste tulemustest. Kõige sagedamini muudetakse andmed (ETL) etapis agregeeritud, rühmitatud või filtreeritud ning alles seejärel teisendatakse need aruannetes esitatud graafikute või kahemõõtmeliste tabelite kujul summadeks. See tähendab, et tabeli sisu võib muutuda sõltuvalt valitud parameetritest, analüüsiperioodist, projekti filtritest või kasutaja seadetest.

Dünaamiliste andmeraamide ja andmekogumite kasutamine Transform etapis muudab aruandlusprotsessi Load etapis võimalikult paindlikuks, skaleeritavaks ja kergesti korratavaks, ilma et oleks vaja käsitsi sekkuda.

Lisaks tabelitele ja tekstile toetab FPDF ka tabeliandmete graafikute lisamist, mis võimaldab lisada aruandesse Matplotlibiga või muude eespool käsitletud visualiseerimisraamatukogudega genereeritud pilte. Koodi abil saab dokumenti lisada mis tahes graafikuid, graafikuid ja diagramme.

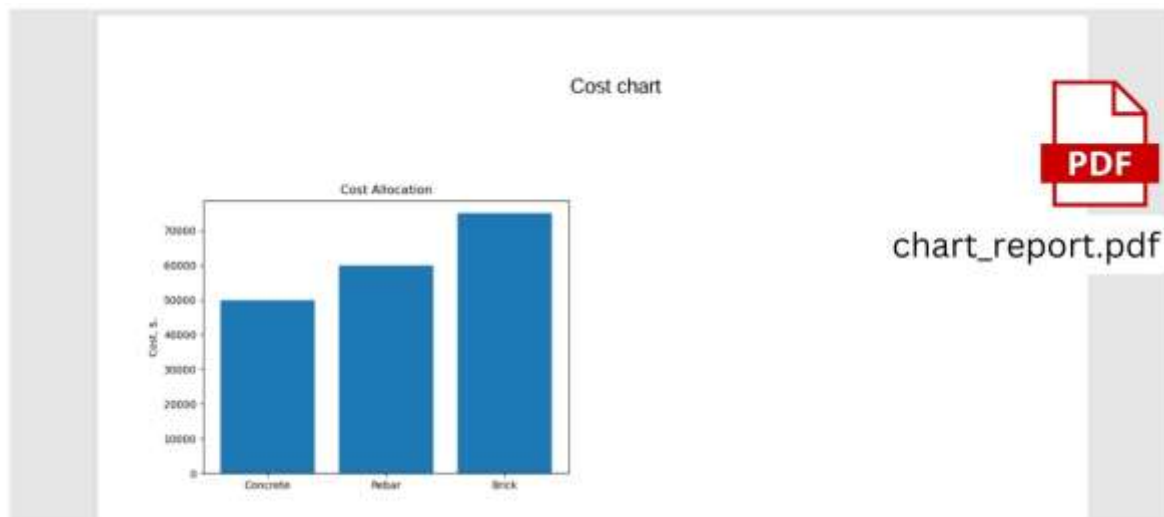
- 📄 Kasutades Pythoni raamatukogu FPDF, lisame PDF-dokumendile Matplotlib. abil eelnevalt genereeritud graafiku:

```
import matplotlib.pyplot as plt # Import Matplotlib graafikute loomiseks

fig, ax = plt.subplots() # Loo graafiku joonis ja teljed
categories = ["Betoon", "Rebar", "Brick"] # Kategooriate nimed
values = [50000, 60000, 75000] # Kategooria väärtused
ax.bar(categories, values) # Luua tulpdiagramm
plt.ylabel("Väärtus,$.") # Märkige Y-telgile
plt.title("Kulude jaotus") # Lisa pealkiri
plt.savefig(r "C:\reports\chart\chart\chart.png") # Salvestage graafik pildina

pdf = FPDF () # Loo PDF -dokumendi.
pdf.add_page() # Lisa lehekülg
pdf.set_font("Arial", size=12) # Kirjastiili määramine
pdf.cell(200, 10, "Cost Chart", ln=True, align='C') # Lisame päise.

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100) # Sisesta pilt
PDF-i (x, y - koordinaadid, w - laius)
pdf.output(r "C:\reports\chart_report.pdf") # PDF-faili salvestamine
```



Joonis 7.2-16 Kümnekonna koodireaga saate luua graafiku, salvestada selle ja seejärel kleepida selle PDF-dokumendi.

FPDFi abil muutub dokumentide ettevalmistamise ja loogika protsess läbipaistvaks, kiireks ja mugavaks. Koodi sisseehitatud mallid võimaldavad genereerida ajakohaste andmetega dokumente, kaotades vajaduse käsitsi täitmise järele.

ETL-automaatika kasutamine - aeganõudva käsitsi aruandluse asemel saavad spetsialistid keskenduda andmete analüüsimisele ja otsuste tegemisele, selle asemel, et valida õige vahend konkreetse andmesiloga töötamiseks selge kasutajaliidesega.

Seega pakub FPDFi raamatukogu paindlikku vahendit mis tahes keerukusega dokumentide automatiseeritud loomiseks - alates lühikestest tehnilistest aruannetest kuni keerukate analüütiliste kokkuvõtete, tabelite ja diagrammide koostamiseni, mis võimaldab mitte ainult kiirendada dokumendivoolu, vaid ka vähendada oluliselt käsitsi sisestatud andmete ja vormindamisega seotud vigade tõenäosust.

ETL Load: aruandlus ja laadimine teistesse süsteemidesse

Laadimise etapis koostati tulemused tabelite, graafikute ja PDF-lõpparuannete kujul, mis koostati vastavalt kehtestatud nõuetele. Lisaks on võimalik neid andmeid eksportida masinloetavatesse vormingutesse (nt CSV), mis on vajalik integratsiooniks väliste süsteemidega, nagu ERP, CAFM, CPM, BI platvormid ja muud ettevõtte või tööstuse lahendused. Lisaks CSV-formaadile saab andmeid üles laadida ka XLSX-, JSON- ja XML-formaadis või otse andmebaasidesse, mis toetavad automaatset teabevahetust.

- 🗣️ Laadimise sammu automatiseerimiseks vajaliku koodi genereerimiseks küsige lihtsalt näiteks LLM-i -liidest: ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude või QWEN:

Kirjutage kood andmete valideerimise tulemuste aruande koostamiseks DataFrame'is, kus veerud, mille eesliide on "verified_", loetakse, nimetatakse ümber "Passed" ja "Failed", puuduvad väärtused asendatakse 0-ga ja seejärel eksporditakse ainult need read, mis läbivad kõik valideerimised, CSV-faili.

 LLMi vastus:



Joonis 7.2-17 Transformeerimise etapis saadud valideeritud andmed lõplikust andmeraamist eksporditakse CSV -faili, et neid saaks integreerida teiste süsteemidega.

Ülaltoodud kood (joonis 7.2-17) rakendab ETL -protsessi viimast etappi - laadimist, mille käigus kontrollitud andmed salvestatakse CSV-vormingus, mis ühildub enamiku välissüsteemide ja andmebaasidega. Seega oleme lõpule viinud ETL -protsessi täieliku tsükli, sealhulgas andmete väljavõtte, teisendamise, visualiseerimise, dokumenteerimise ja eksportimise vajalikesse süsteemidesse ja vormingutesse, mis tagab reprodutseeritavuse, läbipaistvuse ja teabega töötamise automatiseerimise.

ETL torujuhtme saab kasutada nii üksikute projektide töötlemiseks kui ka suuremahuliste rakenduste jaoks - kui analüüsitakse sadu või tuhandeid sissetulevaid andmeid dokumentide, piltide, skaneeringute, CAD projektide, punktpilvede, PDF failide või muude hajutatud süsteemidest pärinevate allikate kujul. Võimalus protsessi täielikult automatiseerida muudab ETLi mitte ainult tehniliseks töötlemisvahendiks, vaid ehitusteabe digitaalse infrastruktuuri aluseks.

ETL koos LLMiga: andmete visualiseerimine PDF -dokumentidest -dokumendid

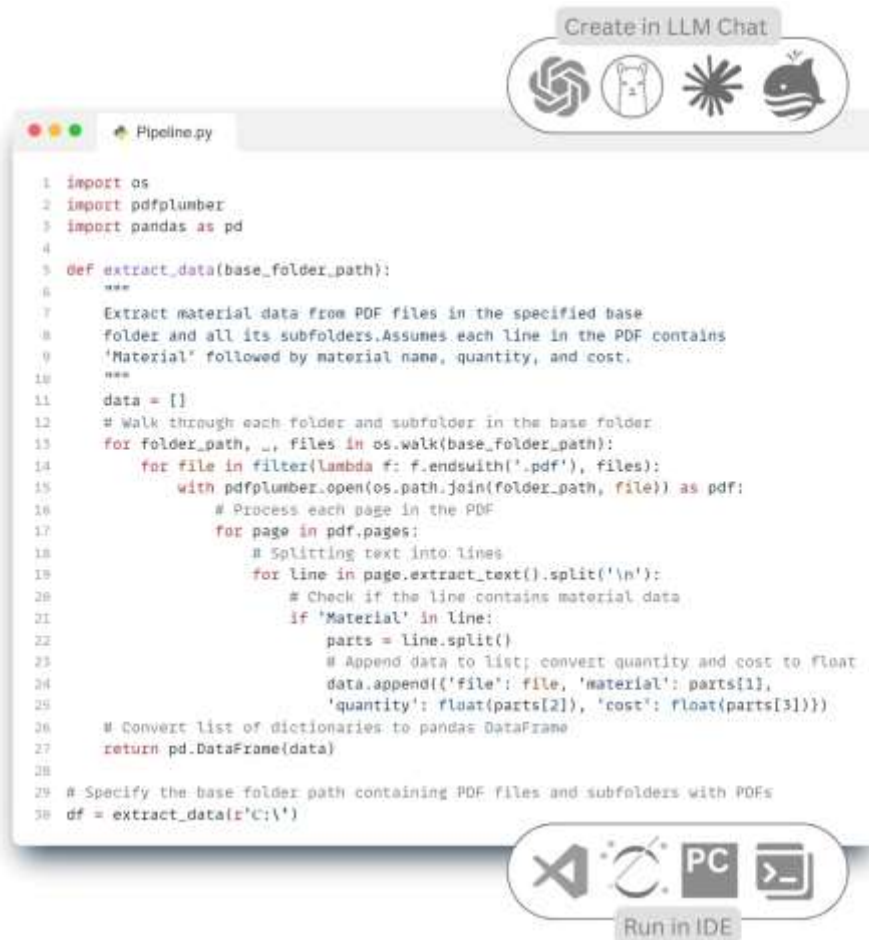
On aeg liikuda edasi täisväärtusliku ETL-protsessi loomiseks, mis hõlmab kõiki andmekäitluse põhietappe ühes stsenaariumis - ekstraheerimist, teisendamist ja laadimist. Ehitame automatiseeritud ETL-piibliini, mis võimaldab teil töödelda PDF-dokumente ilma käsitsi töötlemata - ekstraheerida andmeid dokumentidest, visualiseerida, analüüsida ja edastada teistesse süsteemidesse.

Meie näites kirjeldatakse ETL-protsessi käskluste abil, mis peavad selgitama keelemudelile (LLM) kõiki ETL-protsesse koos saadava lõpptulemuse kirjeldusega. Antud juhul on ülesanne leida kõik PDF-failid kindlaksmääratud kaustas ja selle alamkaustades, eraldada neist asjakohane teave - nt materjalide nimed, kogused ja kulud - ning esitada tulemus struktureeritud tabelina (DataFrame) edasiseks analüüsiks

- 🗨️ Esimene tekstipäring LLM-is, et automaatselt eraldada andmeid mitmest PDF-dokumendist ja luua andmete raamistik väljavõtte sammu jaoks:

Kirjutage kood, et eraldada materjaliinfo PDF-failidest antud kaustas ja selle alamkaustadest. PDF-is sisalduvad andmed hõlmavad materjali nime, kogust ja maksumust. Tulemus tuleks salvestada DataFrame'isse📄

- 🗨️ LLMi vastus :



Joonis 7.2-18 LLM loob Pythoni -koodi, et ekstraheerida andmeid PDF -failidest konkreetses kaustas ja kõigis selle alamkaustades.

LLM-vastus (joonis 7.2-18) on valmis Python -scripti, mis läbib automaatselt kõik kaustad, avab leitud PDF -failid, võtab neist tekstilise teabe ja teisendab selle tabeliks. Saadud jutukoodi saab käivitada ühes populaarsetest IDE-dest PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse koos PyDev pluginaga, Thonny, Wing IDE, IntelliJ IDEA koos Python pluginaga, JupyterLab või populaarsed veebipõhised tööriistad Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Alternatiivina, et lihtsustada kogumisprotsessi, saame koodi kopeerimise asemel LLM ja koodi kasutamise asemel IDE-s, laadida ka kümmekond PDF -faili otse LLM-i vestlusesse (joonis 7.2-19) ja saada väljundiks tabeli, ilma et peaksime koodi nägema või käivitama. Selle koodi täitmise tulemuseks on tabel meie valitud atribuutidega.

The diagram illustrates the workflow of converting unstructured data into structured data using LLM and Python. On the left, a ChatGPT 4 interface shows a user prompt asking for a Python script to extract material information from PDF files. The response includes a sample Python script using 'PyPDF2' and 'pandas' to extract data from PDF files and display it as a pandas DataFrame. Below the code, an example of the output DataFrame is shown, containing columns for Material Name, Quantity, and Cost, with rows for Copper Wire, Aluminum Sheet, Plastic Pipes, and Steel Rods. On the right, a flowchart shows 'UNSTRUCTURED DATA TEXTUAL DATA' (represented by PDF icons) being processed by an LLM (represented by a GPT icon) to produce 'STRUCTURED DATA' (represented by an XLS icon). A central icon of a code editor with a plus sign and code symbols indicates the integration of the LLM and the Python script.

```
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)
```

	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	320.20

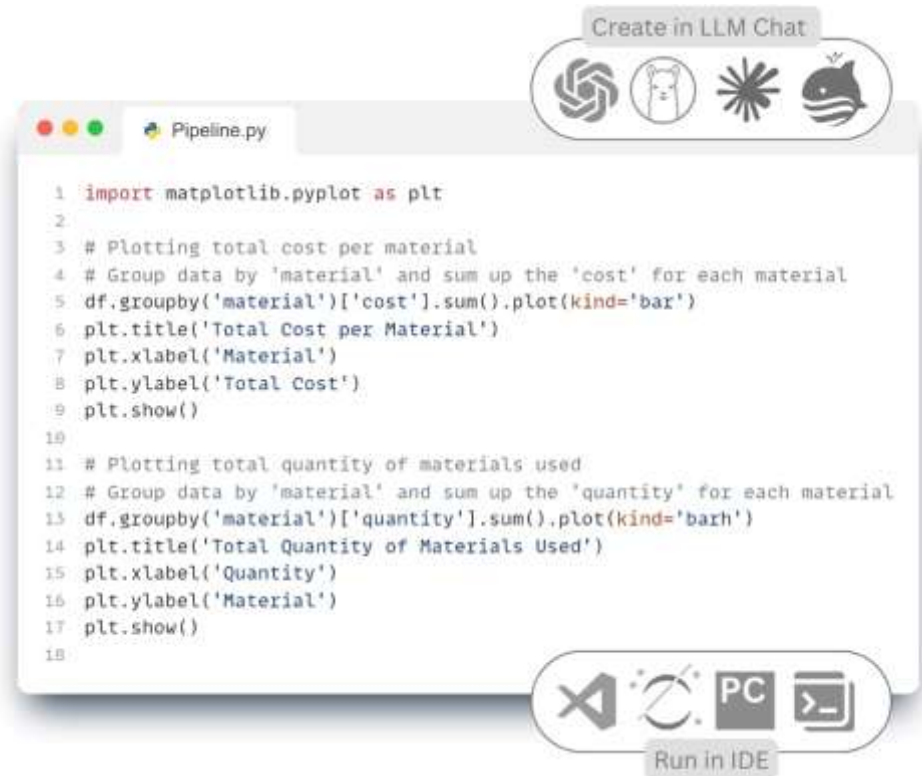
This table shows the material name, quantity, and cost as columns, with each row representing the data extracted from a PDF file. The actual content and structure of your DataFrame will depend on the specific format and data in your PDF files. [↩]

Joonis 7.2-19 LLM-i koodi täitmise tulemus, mis ekstraheerib andmed PDF -failidest struktureeritud andmeraamistiku vaates koos valitud atribuutidega.

Järgmises etapis palume saadud andmete põhjal koostada keelemudeli - näiteks võrrelda materjali kasutamise maksumust ja mahtu ning luua mõned näidisvisualiseeringud, mis on aluseks edasisele analüüsile.

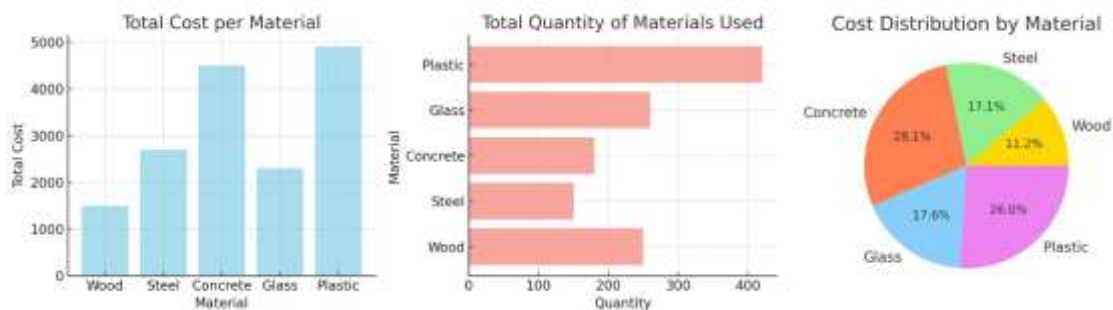
- Paluge jätkuvas vestluses LLM joonistada mõned graafikud tabelitest, mis koostati sammus Transform (Joonis 7.2-18):

Visualiseerida iga materjali kogumaksumus ja -kogus DataFrame'ist (joonis 7.2-18)



Joonis 7.2-20 LLM vastus -mudelid Pythoni koodina andmete visualiseerimiseks andmeraamistikust, kasutades matplotlib raamatukogu.

LLM genereerib ja täidab automaatselt Pythoni -koodi (joonis 7.2-20), kasutades matplotlib raamatukogu. Pärast selle koodi täitmist saame ehitusprojektide kulude ja materjalikasutuse graafikud otse chatis (joonis 7.2-21), mis lihtsustab oluliselt analüütilist tööd.

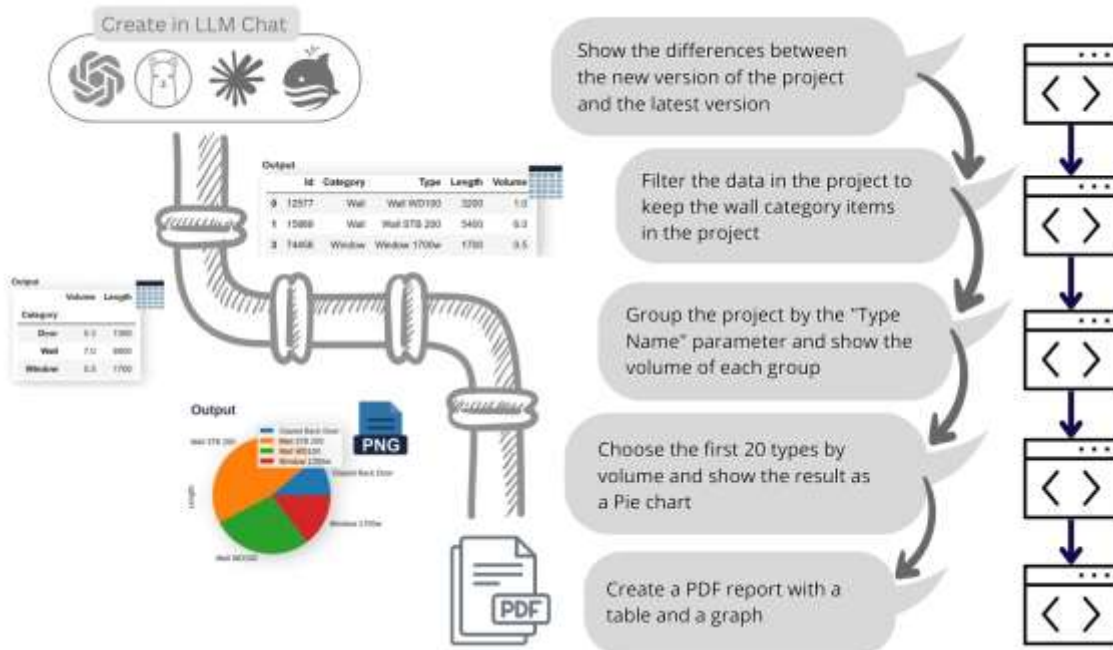


Joonis 7.2-21 LLM-vastuse visualiseerimine graafikutena DataFrame'i kogutud andmete põhjal.

Toetus ETL-koodi kirjutamise ideede väljatöötamiseks, koodi analüüsimiseks ja täitmiseks ning tulemuste visualiseerimiseks on saadaval lihtsate tekstiküsimuste kaudu LLM, ilma et oleks vaja

õppida programmeerimise põhitõdesid. Tehisintellekti vahendite, nagu LLM, tekkimine muudab kindlasti lähenemisviisi programmeerimisele ja andmetöötuse automatiseerimisele (joonis 7.2-22).

PwC aruande "Milline on tehisintellekti tegelik väärtus teie ettevõtte jaoks ja kuidas seda ära kasutada?" kohaselt. (2017) [139] on protsesside automatiseerimine ja tootlikkuse suurendamine peamised majanduskasvu edasiviijad. Ja tootlikkuse paranemine moodustab eeldatavasti enam kui 55% kogu tehisintellektipõhisest SKP kasvust aastatel 2017-2030."



Joonis 7.2-22 AI LLM aitab genereerida koodi eelnõud, mida rakendatakse tulevastes projektides ilma LLMi vajaduseta.

Kasutades selliseid vahendeid nagu ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok, samuti avatud andmeid ja avatud lähtekoodiga tarkvara, saame automatiseerida protsesse, mida varem tehti ainult spetsiaalsete, kulukate ja raskesti hooldatavate modulaarsete varaliste abil.

Ehituse kontekstis tähendab see, et ettevõtted, kes võtavad esimesena kasutusele automatiseeritud Pipeline -andmeprotsessid, saavad märkimisväärset kasu, alates projektijuhtimise tõhususe paranemisest kuni finantskahjude vähenemiseni ning killustatud rakenduste ja silohoidlate kõrvaldamiseni.

ETL-protsessi kirjeldatud äritegevuse ülesannete täitmise loogika on analüütika ja andmetöötusprotsesside automatiseerimise oluline osa, mis on laiema mõiste - torujuhtmete (Pipelines) - spetsiifiline variant.

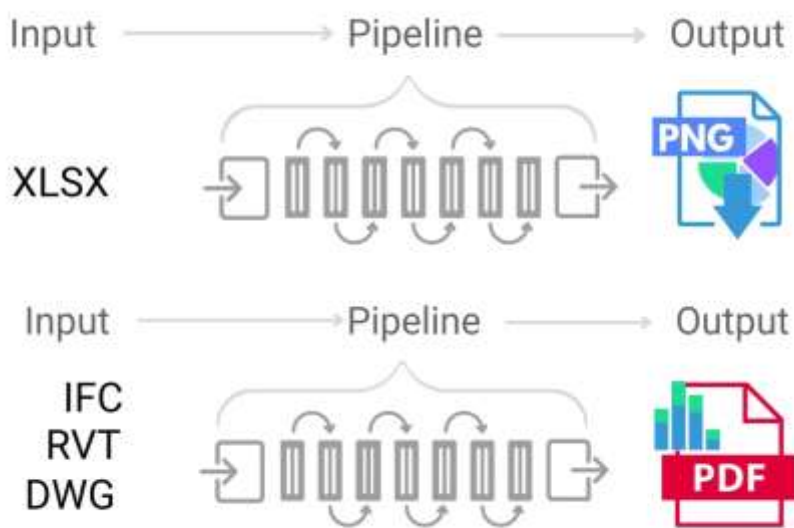
PEATÜKK 7.3.

AUTOMAATNE ETL KONVEIER (TORUSTIK)

Pipeline: Automaatne ETL konveier andmed

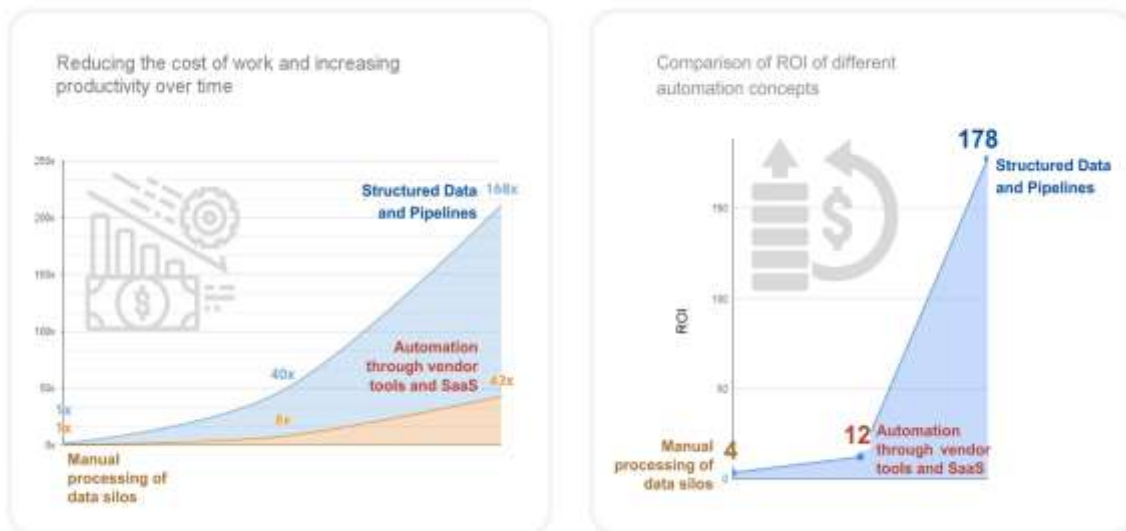
ETL-protsessi on traditsiooniliselt kasutatud andmete töötlemiseks analüütilistes süsteemides, hõlmates nii struktureeritud kui ka struktureerimata allikaid. Tänapäeva digitaalkeskkonnas kasutatakse siiski üha enam laiemat terminit - Pipeline (konveier), mis kirjeldab mis tahes järjestikust töötlemisahelat, kus ühe etapi väljundist saab järgmise etapi sisend.

See lähenemisviis ei kehti mitte ainult andmete, vaid ka muud liiki automatiseerimise kohta: ülesannete töötlemine, aruandluse koostamine, integreerimine tarkvaraga ja digitaalne tööprotsess (joonis 7.3-1).



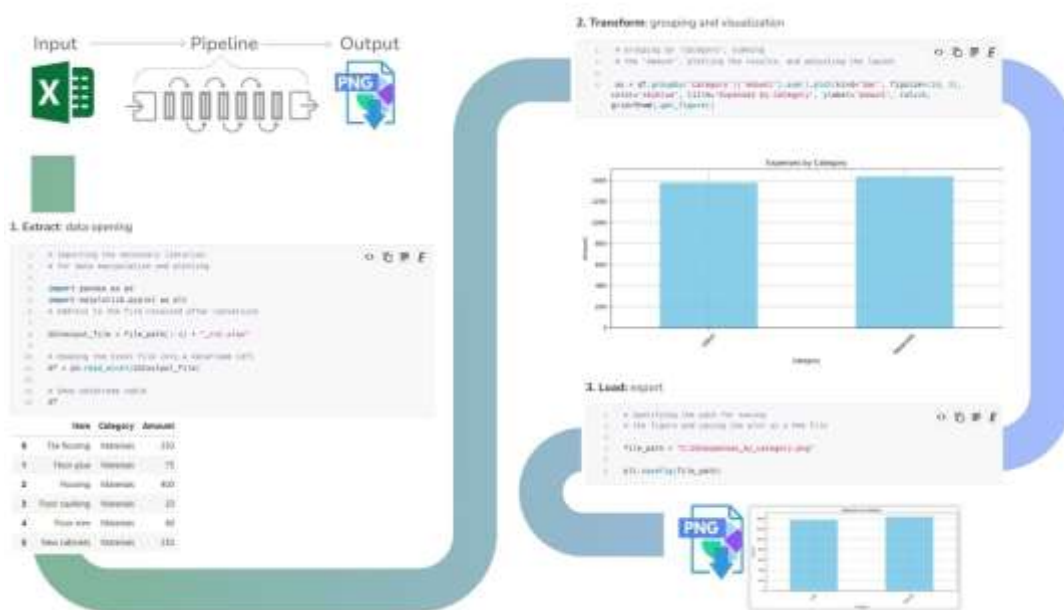
Joonis 7.3-1 Pipeline on töötlemisjärjekord, kus ühe etapi väljundist saab järgmise etapi sisend.

Torustiku kasutamine on üks peamisi automatiseerimise elemente, eriti kui töötatakse suurte heterogeensete andmemahutudega. Pipeline'i arhitektuur võimaldab korraldada keerukaid töötlemisetappe modulaarselt, järjepidevalt ja hallatavalt, mis suurendab loetavust, lihtsustab koodi hooldust ning võimaldab inkrementaalset silumist ja skaleeritavat testimist.



Joonis 7.3-2 ROI Pipeline andmete valideerimise protsess vähendab täitmisaega kümneid ja sadu kordi võrreldes klassikaliste vahenditega tehtava töötlemisega [74].

Erinevalt käsitsi tehtavast tööst oma süsteemides (ERP, PMIS, CAD jne) võimaldab pipelining oluliselt (joonis 7.3-2) suurendada ülesannete kiirust, vältida korduvat tööd ja automatiseerida protsesside käivitamist õigel ajal (joonis 7.3-3).

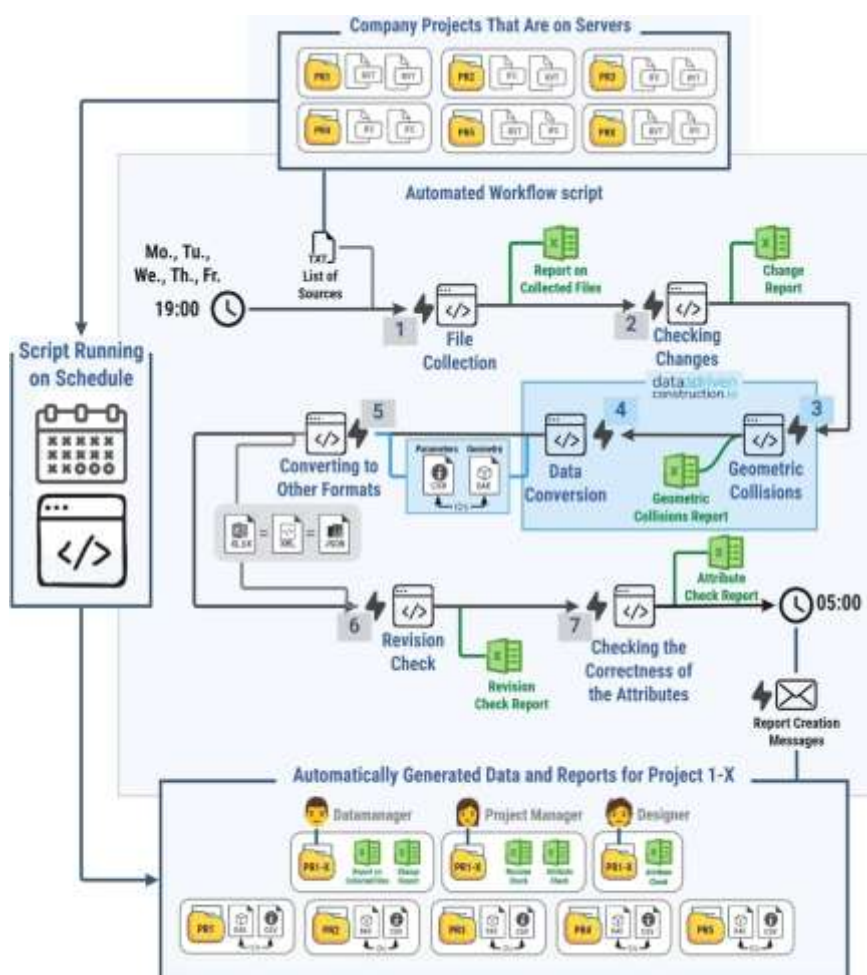


Joonis 7.3-3 ETL näide Pipeline, et automaatselt tuletada graafik XLSX-faili tabeliandmetest ilma Exceli avamata.

Andmete voogedastuse töötlemiseks ja automatiseeritud torujuhtme loomiseks, sarnaselt ETL-protsessiga, tuleb eelnevalt kindlaks määrata andmete allikad ja nende kogumise ajakava - kas konkreetse äriprotsessi või kogu ettevõtte jaoks.

Ehitusprojektide puhul pärinevad andmed paljudest heterogeensetest allikatest, mille uuendamisintervall on erinev. Usaldusväärse andmeväljapaneku loomiseks on väga oluline registreerida, millal teavet hangitakse ja ajakohastatakse. See võimaldab õigeaegset otsuste tegemist ja parandab projektijuhtimise tõhusust.

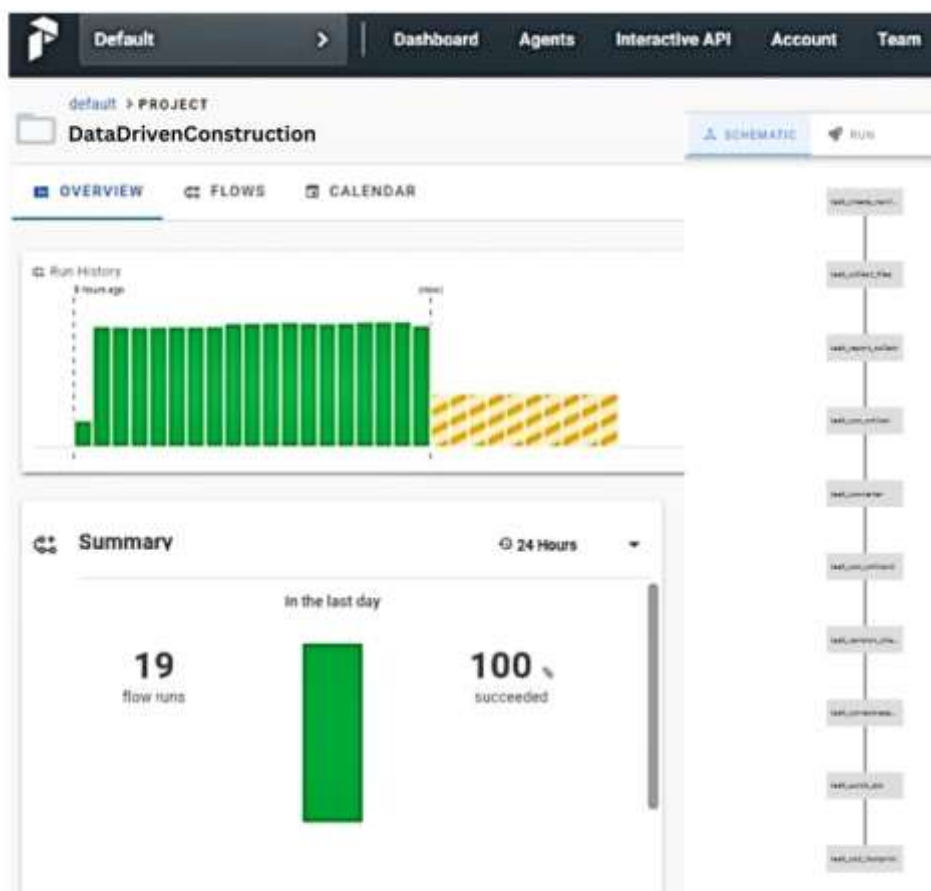
Üks võimalus on alustada montaažiprotsessi kindlal kellaajal - näiteks kell 19:00, tööpäeva lõpus. Sel hetkel aktiveeritakse esimene skript, mis vastutab erinevate süsteemide ja salvestiste andmete koondamise eest (joonis 7.3-4 samm 1). Sellele järgneb andmete automaatne töötlemine ja muundamine analüüsiks sobivasse struktureeritud vormingusse (joonis 7.3-4 samm 2-4). Viimases etapis genereeritakse ettevalmistatud andmeid kasutades automaatselt aruanded, armatuurlauad ja muud eelmistes peatükkides kirjeldatud tooted (joonis 7.3-4 samm 6-7). Selle tulemusena on juhtidel hommikul kell 05:00 juba ajakohased ja vajalikus vormis projekti seisuaruanded (joonis 7.3-5).



Joonis 7.3-4 Õhtul automaatselt kogutud andmeid torujuhtmes töödeldakse öösel, nii et juhtidel on hommikuks ajakohased aruanded ja värsked aruanded.

Õigeaegne andmete kogumine, põhinäitajate määratlemine, ümberkujundamisprotsesside automatiseerimine ja visualiseerimine armatuurlaudade abil on eduka andmepõhise otsustusprotsessi võtmelemendid.

Selliseid automatiseeritud protsesse (joonis 7.3-4) saab teostada täiesti autonoomselt: need töötavad graafiku alusel, töötlevad andmeid järelevalveta ja neid saab kasutada kas pilves või ettevõtte enda serveris (joonis 7.3-5). See võimaldab selliseid ETL-pileteid integreerida olemasolevasse IT-infrastruktuuri, säilitades kontrolli andmete üle ja pakkudes paindlikku skaleerimist.



Joonis 7.3-5 Automaatne ETL -konveier protsessid (joonis 7.3-4) Prefekti platvormil, kus 10 python-skripti käivitatakse vaheldumisi pärast 19:00 igal tööpäeval.

Tööprotsesside automatiseerimine mitte ainult ei suurenda meeskonna tootlikkust, vabastades aega sisukamateks ja vähem rutiinseteks ülesanneteks, vaid on ka oluline esimene samm tehisintellekti (AI) tehnoloogiate integreerimise suunas äriprotsessidesse, mida käsitleme üksikasjalikumalt peatükis "Ennustav analüüs ja masinõpe".

Torustik -ETL andmete valideerimisprotsess LLMiga

Eelmistes peatükkides, mis käsitlesid andmenõuete koostamist ja ETLi automatiseerimist, kirjeldasime samm-sammult andmete ettevalmistamise, teisendamise, valideerimise ja visualiseerimise protsessi. Need tegevused rakendati eraldi koodiplokkidena (joonis 7.2-18 - joonis 7.2-20), millest igaüks täidab konkreetset ülesannet.

Nüüd on meil järgmine eesmärk - ühendada need elemendid ühtseks, sidusaks ja automatiseeritud andmetöötluseks - torujuhtmeks, ETL -Pipeline -, kus kõik etapid (laadimine, valideerimine, visualiseerimine, eksport) viiakse järjestikku läbi ühes automaatselt täidetavas skriptis.

Järgnevas näites rakendatakse täielik andmetöötlustsükkel: alates CSV lähtefaili →laadimisest kuni struktuuri ja väärtuste kontrollimiseni regulaaravaldiste abil →tulemuste arvutamiseni→ visuaalse aruande genereerimiseni PDF-formaadis.

 Sobiva koodi saamiseks võite kasutada järgmist tekstiküsimustikku LLMile:

Palun kirjutage koodinäide, mis laeb andmeid CSV-st, valideerib DataFrame kasutades regulaarseid väljendeid, kontrollib identifikaatoreid formaadis 'W-NEW' või 'W-OLD', energiatõhusust tähtedega 'A' kuni 'G', garantiiperioodi ja asendustsükli numbriliste väärtustega aastates ning loob lõpuks aruande koos läbitud ja läbikukkunud väärtuste arvuga, genereerib PDF-i koos tulemuste histogrammiga ja lisab tekstilise kirjelduse. ↩

 LLMi vastus :

[Create in LLM Chat](#)


Pipeline.py

```

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['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

```



Joonis 7.3-6 Pipeline (ETL) automatiseerib kogu andmetötluse tsükli: alates allalaadimisest ja valideerimisest kuni PDF-formaadis struktureeritud aruande koostamiseni .

Automaatne kood (joonis 7.3-6) LLMi jututoa sees või DIE-s, pärast koodi kopeerimist, valideerib CSV-faili andmed, kasutades määratud regulaarseid väljendeid, loob aruande läbitud ja läbikukkunud kirjade arvu kohta ning salvestab valideerimise tulemused PDF-failina.

Selline ETL-konveieri struktuur, kus iga samm - alates andmete laadimisest kuni aruande koostamiseni - on rakendatud eraldi moodulina, tagab läbipaistvuse, skaleeritavuse ja reprodutseeritavuse. Valideerimisloogika esitamine kergesti loetava Python-koodina muudab protsessi läbipaistvaks ja arusaadavaks mitte ainult arendajatele, vaid ka andmehalduse, kvaliteedi ja analüüsi spetsialistidele.

Pipeline'i - lähenemine andmetöötluse automatiseerimisele võimaldab standardiseerida protsesse, suurendada nende korratavust ja lihtsustada kohandamist uute projektidega. See loob ühtse metoodika andmete analüüsimiseks, olenemata andmeallikast või ülesande tüübist - olgu see siis vastavustesti, aruannete koostamine või andmete edastamine välissüsteemidesse.

Selline automatiseerimine vähendab inimlikke vigu, vähendab sõltuvust patenteeritud lahendustest ning suurendab tulemuste täpsust ja usaldusväärsust, mistõttu need sobivad nii operatiivanalüütikaks projekti tasandil kui ka strateegiliseks analüütikaks ettevõtte tasandil.

Torustik -ETL: projektielementide andmete ja teabe kontrollimine CADis (BIM).

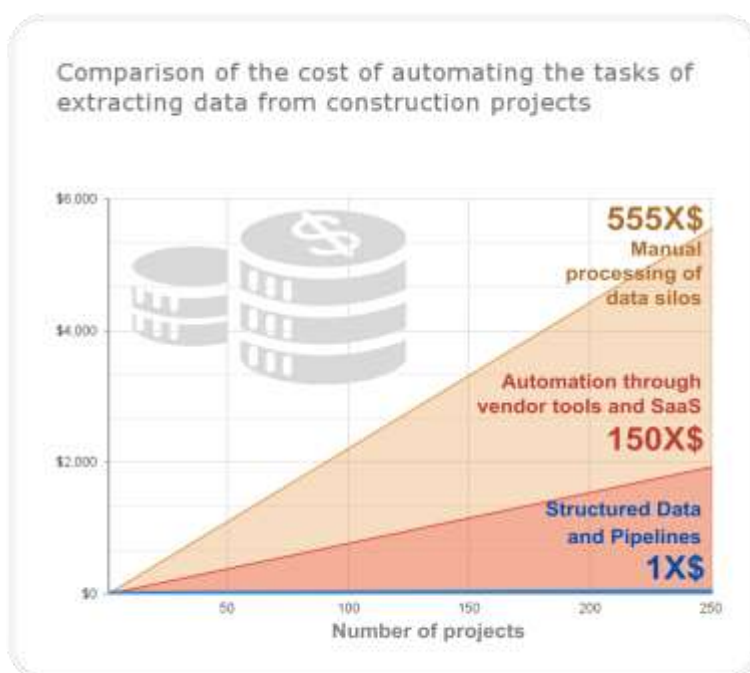
CAD-süsteemide ja andmebaaside (BIM) andmed on ehitusettevõtete jaoks ühed kõige keerukamad ja dünaamiliselt ajakohastatavad andmeallikad. Need rakendused ei kirjelda projekti mitte ainult geomeetria abil, vaid täiendavad seda ka mitmel tasandil tekstilise teabega: mahud, materjaliomadused, ruumide määramine, energiatõhususe tasemed, tolerantsid, eeldatav eluiga ja muud atribuudid.

CAD-mudelites üksustele määratud atribuudid moodustatakse projekteerimisetapis ja need saavad aluseks edasistele äriprotsessidele, sealhulgas kulukalkulatsioonile, ajakava koostamisele, elutsükli hindamisele ja integreerimisele ERP ja CAFM-süsteemidega, kus protsesside tõhusus sõltub suuresti projekteerimisosakondadest pärinevate andmete kvaliteedist.

Traditsiooniline lähenemisviis CAD- (BIM-) mudelite atribuutide valideerimiseks hõlmab käsitsi valideerimist (joonis 7.2-1), mis muutub pikaajaliseks ja kulukaks protsessiks, kui mudelite maht on suur. Arvestades kaasaegsete ehitusprojektide mahtu ja arvu ning nende regulaarset ajakohastamist, muutub andmete valideerimise ja ümberkujundamise protsess jätkusuutmatuks ja kättesaamatuks.

Peatöövõtjad ja projektijuhid seisavad silmitsi vajadusega töödelda suures koguses projektiandmeid, sealhulgas samade mudelite mitmeid versioone ja fragmente. Andmed tulevad projekteerimisasutustelt RVT, DWG, DGN, IFC, NWD ja muudes vormingutes (joonis 3.1-14) ning nõuavad regulaarset kontrollimist, et need vastaksid tööstus- ja ettevõtte standarditele

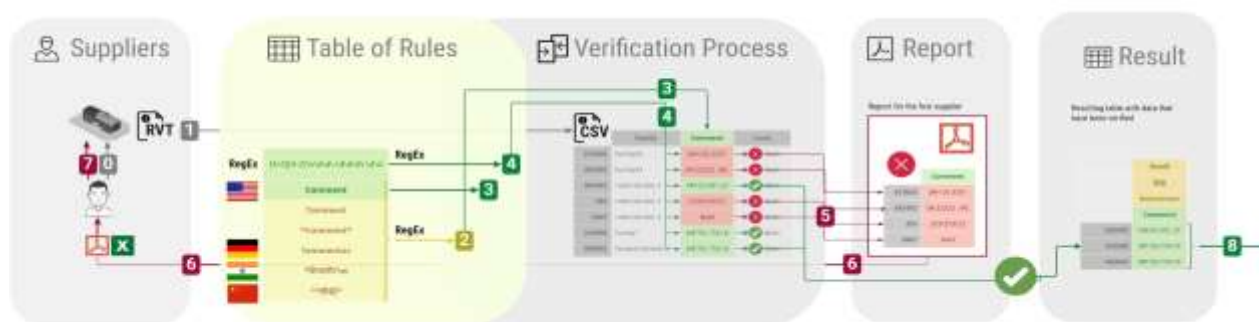
Sõltuvus käsitsi tehtavatest toimingutest ja spetsiaalsest tarkvarast muudab andmete valideerimise protsessi kitsaskohaks kogu ettevõtet hõlmavate mudelite andmetega seotud tööprotsessides. Automatiseerimine ja struktureeritud nõuete kasutamine võib selle sõltuvuse kõrvaldada, suurendades oluliselt andmete valideerimise kiirust ja usaldusväärsust (joonis 7.3-7).



Joonis 7.3-7 Automatiseerimine suurendab andmete kontrollimise ja töötlemise kiirust, mis vähendab tööde maksumust kümneid kordi [140].

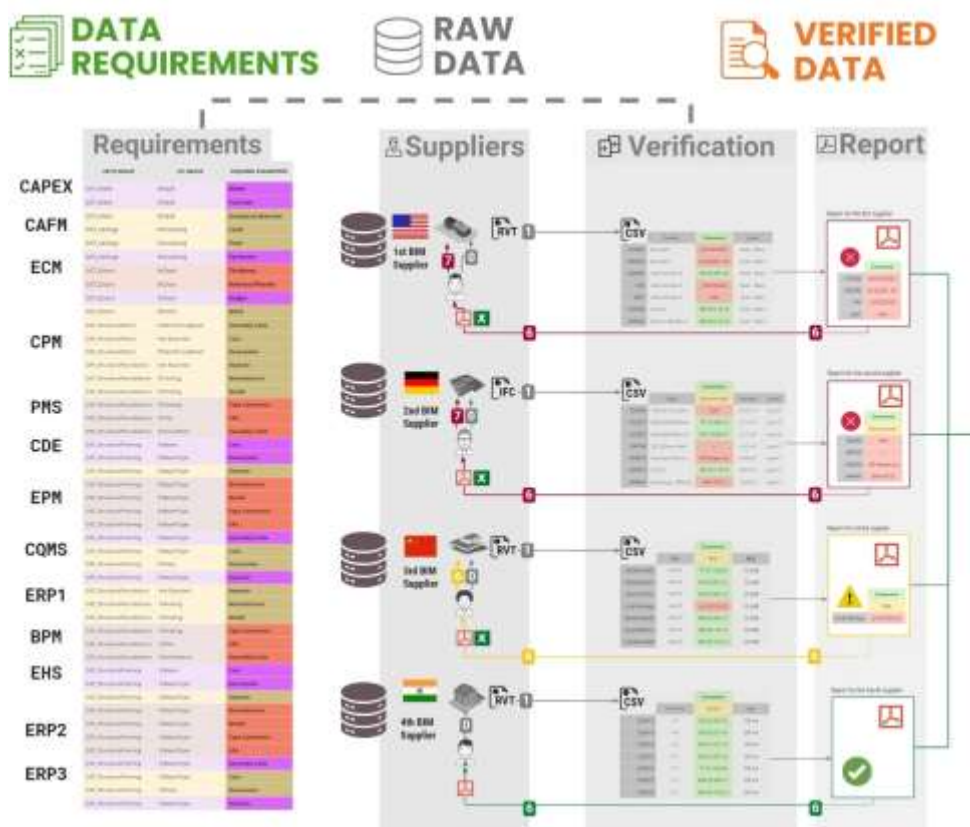
CAD-andmete valideerimisprotsess hõlmab andmete ekstraheerimist (ETL etapp Extract) erinevatest suletud (RVT, DWG, DGN, NWS jne) või avatud poolstruktureeritud ja parameetrilistest vormingutest (IFC, CPXML, USD.) või avatud poolstruktureeritud ja parameetrilistest vormingutest (IFC, CPXML, USD), kus igale atribuudile ja selle väärtustele saab kohaldada reeglitableid (Transform etapp), kasutades regulaarseid väljendeid RegEx (joonis 7.3-8), protsessi, mida me käsitlesime üksikasjalikult raamatu neljandas osas.

PDF-vigaaruande loomine ja edukalt valideeritud kirjade kohta peaks olema lõpule viidud väljundiga (laadimissamm) struktureeritud vormingus, mis võtab arvesse ainult valideeritud üksusi, mida saab kasutada edasistes protsessides.



Joonis 7.3-8 Andmete valideerimisprotsess projektandmete esitajalt regulaaravaldiste abil valideeritud lõpparuandeni.

CAD-süsteemidest (BIM) pärit andmete valideerimise automatiseerimine koos struktureeritud nõuetega ja uute andmete voogedastus, mida töödeldakse ETL-pipeliinide kaudu (joonis 7.3-9), vähendab vajadust valideerimisprotsessis käsitsi osalemise järele (iga valideerimis- ja andmenõuete protsessi on käsitletud eelmistes peatükkides).



Joonis 7.3-9 Andmete valideerimise automatiseerimine ETLi abil lihtsustab ehitusprojektide haldamist, kiirendades protsesse.

Traditsiooniliselt võib töövõtjate ja CAD (BIM) spetsialistide esitatud mudelite valideerimine võtta

päevi või nädalaid. Automaatsete ETL-protsesside kasutuselevõttuga saab seda aga vähendada mõne minutini. Tüüpilises olukorras väidab töövõtja: *"Mudel on valideeritud ja vastab nõuetele."* See avaldus käivitab töövõtja andmekvaliteedi väite kontrollimise ahela:

🗨️ Projektijuht - *"Töövõtja väidab: "Mudel on testitud, kõik on korras"."*

🗨️ Andmehaldur - Laadimise valideerimine:

■ Lihtne skript Pandases tuvastab rikkumise sekunditega. Automatiseerimine välistab vaidlused:

- Kategooria: OST_StructuralColumns, parameeter: FireRating IS NULL.
- Loo nimekiri rikkumiste ID-dest → eksportida Excelisse/PDFi.

Lihtne skript Pandas tuvastab rikkumise sekunditega:

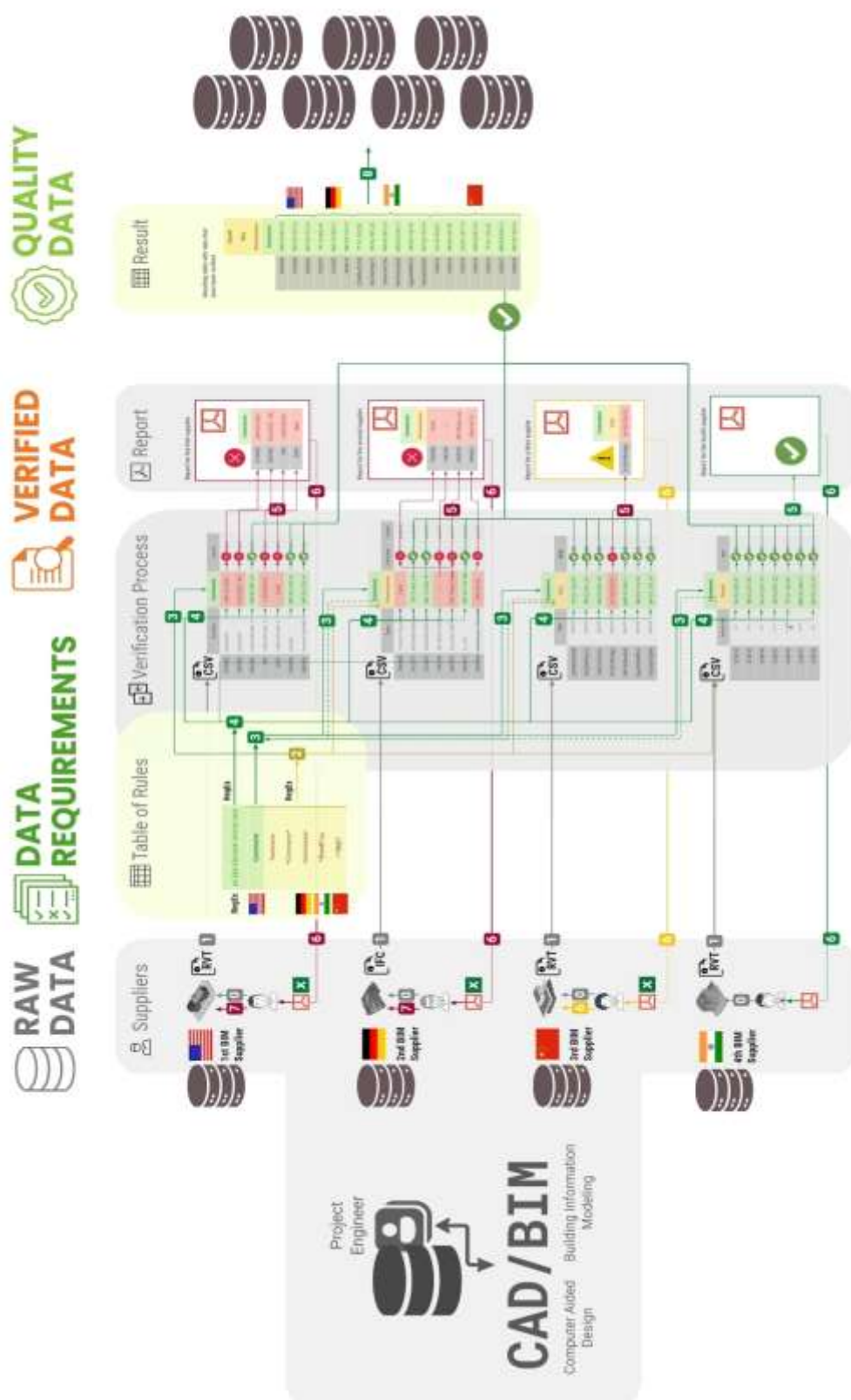
```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # Filtering
issues = df[df["FireRating"].isnull()] # Tühjad väärtused
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Eksporti tunnused
```

🗨️ Andmehaldur projektijuhile - *"kontrollimine näitab, et 18 veerus ei ole FireRating parameetrit täidetud."*

🗨️ Projektijuht töövõtjale - *"Mudel tagastatakse läbivaatamiseks: parameeter FireRating on kohustuslik, ilma selleta on vastuvõtmine võimatu"*

Selle tulemusena ei toimu CAD-mudeli valideerimist, automatiseerimine välistab vaidlused ja töövõtja saab peaaegu koheselt struktureeritud aruande probleemsete elementide ID-de loeteluga. Sel viisil muutub valideerimisprotsess läbipaistvaks, korratavaks ja inimlike vigade eest kaitstuks (joonis 7.3-10).

Selline lähenemisviis muudab andmete valideerimise protsessi pigem tehniliseks funktsiooniks kui käsitsi teostatavaks kvaliteedikontrolliks. See mitte ainult ei suurenda tootlikkust, vaid võimaldab ka kohaldada sama loogikat kõigi ettevõtte projektide suhtes, võimaldades protsesside läbivat digitaalset ümberkujundamist alates projekteerimisest kuni operatsioonini.



Joonis 7.3-10 Elementide atribuutide kontrollimise automatiseerimine välistab inimliku eksimuse ja vähendab vigade tõenäosust.

Automatiseeritud torujuhtmete (joonis 7.3-10) abil saavad süsteemi kasutajad, kes ootavad CAD- (BIM-) süsteemidest kvaliteetseid andmeid, kohe kätte vajalikud väljundandmed - tabelid, dokumendid, pildid - ja integreerivad need kiiresti oma tööülesannetesse.

Kontrolli, töötlemise ja analüüsi automatiseerimine toob kaasa muutusi ehitusprojektide juhtimise käsitluses, eelkõige erinevate süsteemide koostalitlusvõime osas, ilma et kasutataks keerulisi ja kalleid modulaarseid varasüsteeme või suletud tarnijate lahendusi.

Kuigi mõisted ja turunduslikud akronüümid tulevad ja lähevad, jäävad andmenõuete valideerimisprotsessid ise igaveseks äriprotsesside lahutamatuks osaks. Selle asemel, et luua üha rohkem spetsialiseeritud formaate ja standardeid, peaks ehitussektor kasutama vahendeid, mis on juba tõestanud oma tõhusust teistes tööstusharudes. Tänapäeval on olemas võimsad platvormid andmetöötlemise ja protsesside integreerimise automatiseerimiseks, mis võimaldavad ettevõtetel oluliselt vähendada rutiinsetele toimingutele kuluvat aega ja minimeerida vigu Extract, Transform ja Load.

Üks populaarne näide ETL-protsesside automatiseerimise ja orkestreerimise lahendustest on Apache Airflow, mis võimaldab korraldada keerulisi arvutusprotsesse ja hallata ETL-pileteid. Lisaks Airflow'le kasutatakse aktiivselt ka teisi sarnaseid lahendusi, nagu Apache NiFi andmete marsruutimiseks ja voogedastuseks ning n8n äriprotsesside automatiseerimiseks.



PEATÜKK 7.4.

ETL ORKESTREERIMINE JA TÖÖVOOD: PRAKTILISED LAHENDUSED

DAG ja Apache Airflow: töövoogude automatiseerimine ja orkestreerimine

Apache Airflow on tasuta ja avatud lähtekoodiga platvorm, mis on loodud töövoogude (ETL - konveierid) automatiseerimiseks, orkestreerimiseks ja jälgimiseks.

Suurte andmemahtudega töötamine on vajalik iga päev:

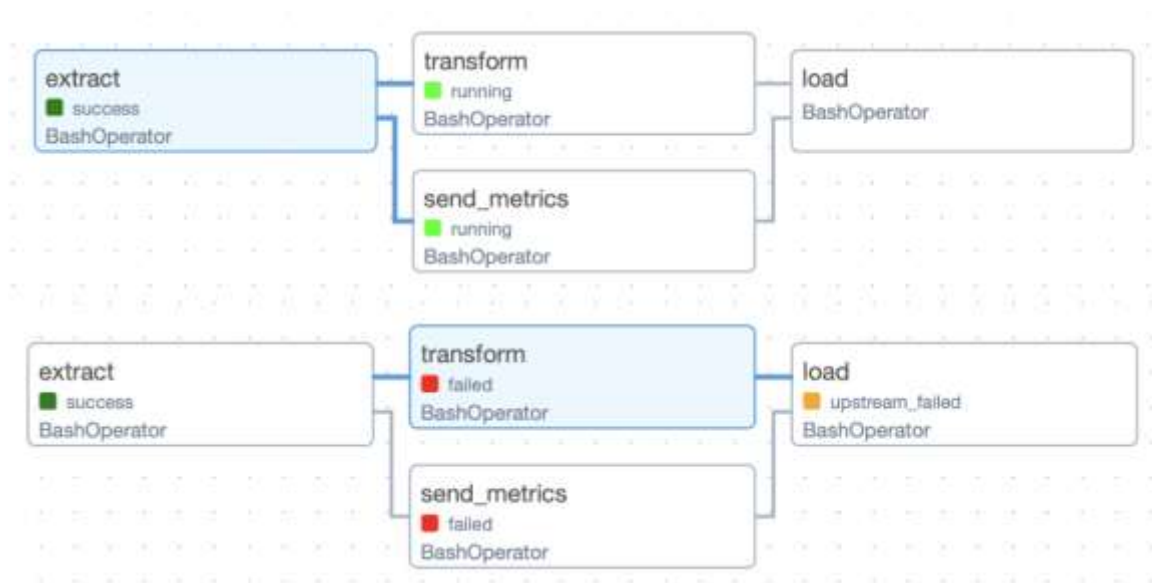
- Failide allalaadimine erinevatest allikatest - Väljavõte (näiteks tarnijatelt või klientidelt).
- Andmete teisendamine nõutavasse vormingusse - Transform (struktureerimine, puhastamine ja valideerimine).
- Tulemuste saatmine kontrollimiseks ja aruannete koostamine - laadimine (üleslaadimine nõutavatesse süsteemidesse, dokumentidesse, andmebaasidesse või armatuurlaudadele.

Selliste ETL-protsesside käsitsi teostamine võtab palju aega ja toob kaasa inimliku vea ohu. Muutus andmeallikas või tõrge mõnes etapis võib põhjustada viivitusi ja ebaõigeid tulemusi.

Automatiseerimisvahendid, näiteks Apache Airflow, võimaldavad luua usaldusväärse ETL -konveieri, minimeerida vigu, vähendada töötlemisaega ja tagada, et andmed on igas etapis korrektsed. Apache Airflow keskmes on DAG (Directed Acyclic Graph) kontseptsioon - suunatud atsükliline graaf, milles iga ülesanne (operaator) on seotud teiste sõltuvustega ja mida täidetakse rangelt kindlaksmääratud järjekorras. DAG välistab tsüklid, mis tagab ülesannete täitmise loogilise ja prognoositava struktuuri.

Airflow hoolitseb orkestreerimise eest - ülesannete vaheliste sõltuvuste haldamine, täitmise ajakavade kontrollimine, oleku jälgimine ja automaatne reageerimine tõrgetele. Selline lähenemisviis vähendab käsitsi sekkumist ja tagab kogu protsessi usaldusväärsuse.

Task Orchestrator on vahend või süsteem, mis on loodud ülesannete täitmise haldamiseks ja kontrollimiseks keerukates arvuti- ja teabekeskondades. See hõlbustab ülesannete täitmise juurutamise, automatiseerimise ja haldamise protsessi, et parandada jõudlust ja optimeerida ressursse.



Joonis 7.4-1 Apache Airflow pakub kasutajasõbralikku kasutajaliidest, kus saab visualiseerida DAG -ETL, vaadata täitmisprotokolle, ülesannete käivitamise olekut ja muud.

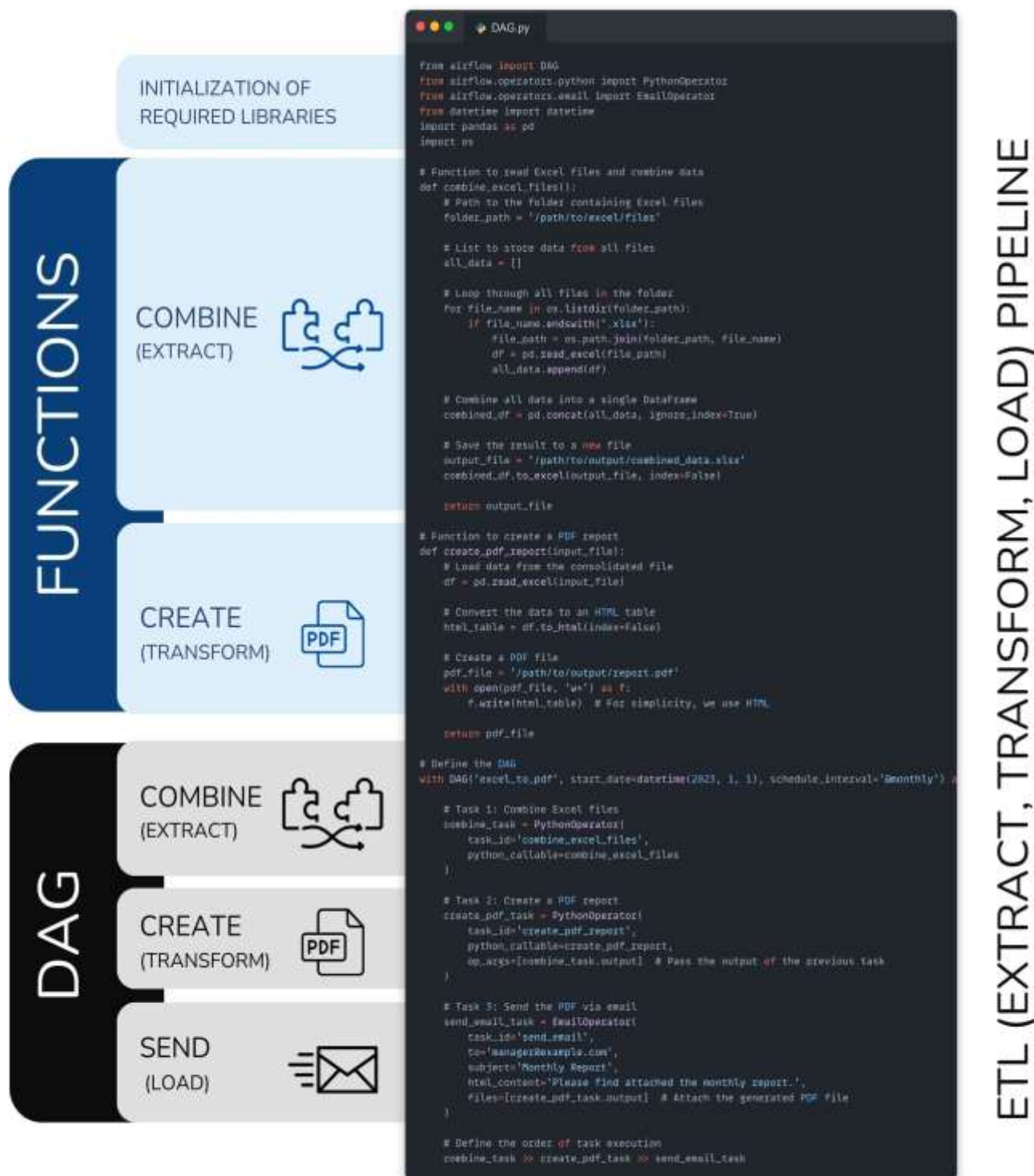
Airflow't kasutatakse laialdaselt hajutatud arvutuste, andmetöötuse, ETL (Extract, Transform, Load) protsesside haldamise, ülesannete ajakava ja muude andmestsenaariumide orkestreerimiseks ja automatiseerimiseks. Vaikimisi kasutab Apache Airflow andmebaasina SQLite.

Näide lihtsast DAGist, mis sarnaneb ETL, koosneb ülesannetest - Extract, Transform ja Load. Graafis, mida juhitakse kasutajaliidese kaudu (joonis 7.4-1), on määratud ülesannete (koodifragmentide) täitmise järjekord: näiteks esimesena täidetakse Extract, seejärel transform (ja sending_metrics) ning töö lõpetab ülesanne Load. Kui kõik ülesanded on lõpetatud, loetakse andmete laadimine edukaks.

Apache Airflow: praktiline rakendus ETLi automatiseerimisel

Apache Airflow on laialdaselt kasutusel keerukate andmetöötlusprotsesside korraldamiseks, võimaldades luua paindlikke ETL -konveiereid. Apache Airflow't saab käivitada kas veebiliidese kaudu või programmeeritult Python-koodi abil (joonis 7.4-2). Veebiliidese (joonis 7.4-3) saavad administraatorid ja arendajad visuaalselt jälgida DAGe, käivitada ülesandeid ja analüüsida täitmise tulemusi.

Kasutades DAGi, saate määrata ülesannete selge järjestuse, hallata nende vahelisi sõltuvusi ja reageerida automaatselt lähteandmete muutustele. Vaatleme näidet Airflow'i kasutamisest aruandluse töötlemise automatiseerimiseks (joonis 7.4-2).



Joonis 7.4-2 ETL-kontseptsioon -konveier andmete töötlemiseks Apache Airflow'i abil.

Selles näites (joonis 7.4-2) vaadeldakse DAGi, mis täidab ETLi -konveieris põhiülesandeid:

■ Exceli -failide lugemine (Extract):

- Kõigi antud kataloogis olevate failide järjestikune läbimine.
- Loe andmeid igast failist, kasutades pandas raamatukogu.
- Kõik andmed ühendatakse ühte DataFrame'i.

■ Loo PDF -dokument (Transform):

- Transformeerib ühendatud DataFrame HTML -tabeliks.
- Salvestage tabel PDF-formaadis (demoversioonis - HTML-i kaudu).

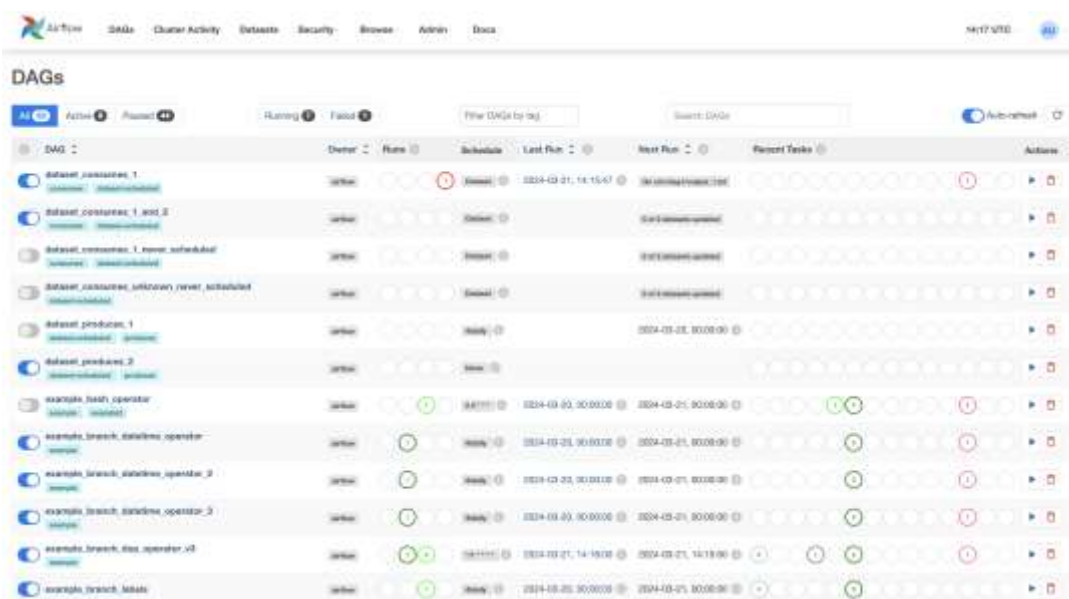
■ Aruande saatmine e-posti teel (Load):

- Rakendage EmailOperator PDF -dokumendi saatmiseks e-posti teel.

■ DAG-i konfigureerimine:

- Ülesannete järjestuse määratlemine: andmete väljavõtmine → aruande koostamine → saatmine.
- Käivitamise ajakava määramine (@kuu - iga kuu esimene päev).

Automaatne ETL -näide (joonis 7.4-2) näitab, kuidas koguda andmeid Exceli -failidest, luua PDF -dokument ja saata see e-postiga. See on vaid üks paljudest Airflow'i võimalikest kasutusviisidest. Seda näidet saab kohandada mis tahes konkreetse ülesandega, et lihtsustada ja automatiseerida andmetöötlust.



DAG	Owner	Name	Schedule	Last Run	Next Run	Recent Tasks
data_loader_operator_1	airflow	data_loader_operator_1	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	data_loader_operator_1
data_loader_operator_1_and_2	airflow	data_loader_operator_1_and_2	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	data_loader_operator_1_and_2
data_loader_operator_1_and_2_scheduler	airflow	data_loader_operator_1_and_2_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	data_loader_operator_1_and_2_scheduler
data_loader_operator_scheduler_scheduler	airflow	data_loader_operator_scheduler_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	data_loader_operator_scheduler_scheduler
data_loader_operator_1_scheduler	airflow	data_loader_operator_1_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	data_loader_operator_1_scheduler
data_loader_operator_2_scheduler	airflow	data_loader_operator_2_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	data_loader_operator_2_scheduler
example_branch_operator	airflow	example_branch_operator	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	example_branch_operator
example_branch_scheduler_operator_scheduler	airflow	example_branch_scheduler_operator_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	example_branch_scheduler_operator_scheduler
example_branch_scheduler_operator_scheduler_scheduler	airflow	example_branch_scheduler_operator_scheduler_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	example_branch_scheduler_operator_scheduler_scheduler
example_branch_scheduler_operator_scheduler_scheduler_scheduler	airflow	example_branch_scheduler_operator_scheduler_scheduler_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	example_branch_scheduler_operator_scheduler_scheduler_scheduler
example_branch_scheduler_scheduler_scheduler_scheduler_scheduler	airflow	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler
example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler	airflow	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler
example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler	airflow	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler
example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler	airflow	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler
example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler	airflow	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler	Interval: 1	2024-03-21, 14:15:47	2024-03-21, 14:15:47	example_branch_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler_scheduler

Joonis 7.4-3 Ülevaade kõigist DAGidest keskkonnas koos teabega viimaste käivituste kohta.

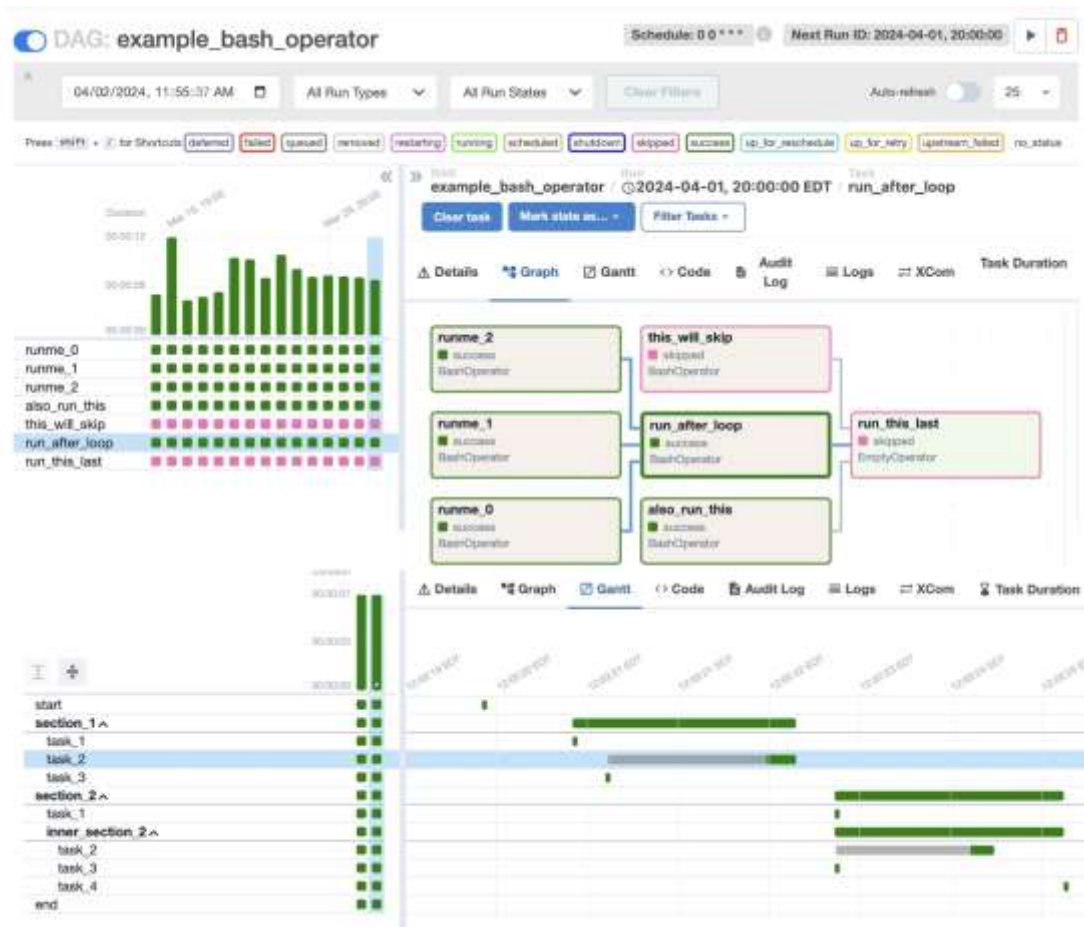
Apache Airflow'i veebiliides (joonis 7.4-3) pakub põhjalikku visuaalset keskkonda andmete töövoogude haldamiseks. See kuvab DAG-d interaktiivsete graafidena, kus sõlmed kujutavad ülesandeid ja servad nende vahelisi sõltuvusi, muutes keeruliste andmevoogude jälgimise lihtsaks. Kasutajaliides sisaldab armatuurlauda, mis sisaldab teavet ülesannete täitmise oleku, jooksva ajaloo, üksikasjalike logide ja jõudlusnäitajate kohta. Administraatorid saavad intuitiivse kasutajaliidese kaudu käsitsi käivitada ülesandeid, taaskäivitada ebaõnnestunud toiminguid, peatada DAGid ja

konfigureerida keskkonnamuutujaid.

Sellist arhitektuuri saab täiendada andmete valideerimise, täitmise olekust teavitamise, integratsiooni väliste APIdega või andmebaasidega. Airflow võimaldab paindlikult kohandada DAGi: lisada uusi ülesandeid, muuta nende järjekorda, kombineerida ahelad - see teeb sellest tõhusa vahendi keerukate andmetöötlusprotsesside automatiseerimiseks. DAGi käivitamisel Airflow'i veebiliideses (joonis 7.4-3, joonis 7.4-4) saate jälgida ülesannete staatust. Süsteem kasutab värvilist märgistust:

- Roheline - ülesanne on edukalt täidetud.
- Kollane - protsess on käimas.
- Punane - viga ülesande täitmisel.

Vigade korral (nt puuduv fail või katkine andmestruktuur) algatab süsteem automaatselt teate saatmise.



Joonis 7.4-4 Apache Airflow lihtsustab oluliselt probleemide diagnoosimist, protsesside optimeerimist ja meeskonna koostööd keerukate andmetöötlusliinide puhul.

Apache Airflow on mugav, sest see automatiseerib rutiinseid ülesandeid, välistades vajaduse neid käsitsi täita. See tagab töökindluse, jälgides protsesside täitmist ja andes kohese veateate. Süsteemi paindlikkus võimaldab hõlpsasti lisada uusi ülesandeid või muuta olemasolevaid, kohandades töövooge vastavalt muutuvatele nõuetele.

Lisaks Apache Airflow'le on olemas sarnased vahendid töövoogude orkestreerimiseks. Näiteks avatud lähtekoodiga ja tasuta Prefect (joonis 7.3-5) pakub lihtsamat süntaksit ja integreerub paremini Pythoniga, Spotify poolt välja töötatud Luigi pakub sarnast funktsionaalsust ja töötab hästi suurandmete puhul. Märkimist väärivad ka Kronos ja Dagster, mis pakuvad kaasaegseid lähenemisviise Pipeline'i ehitamiseks, keskendudes modulaarsusele ja skaleeritavusele. Ülesannete orkestreerimisvahendi valik sõltub projekti konkreetsetest vajadustest, kuid need kõik aitavad automatiseerida keerulisi ETL andmeprotsesse.

Eriti tähelepanuväärne on Apache NiFi, avatud lähtekoodiga platvorm, mis on mõeldud andmete voogedastuseks ja marsruutimiseks. Erinevalt Airflow'st, mis keskendub partiitöötlusele ja sõltuvuste haldamisele, keskendub NiFi andmete reaajas, jooksvalt toimuvale ümberkujundamisele ja süsteemide vahelisele paindlikule marsruutimisele.

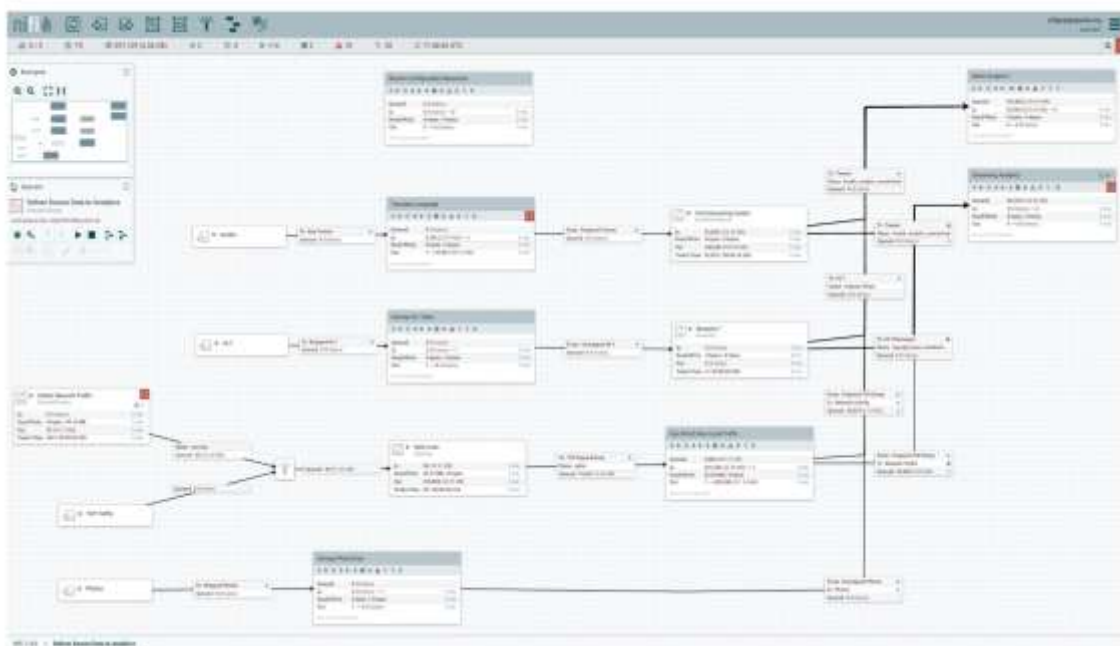
Apache NiFi marsruutimiseks ja andmete teisendamiseks

Apache NiFi on võimas avatud lähtekoodiga platvorm, mis on loodud erinevate süsteemide vaheliste andmevoogude automatiseerimiseks. Algselt arendas seda 2006. aastal USA Riiklik Julgeolekuagentuur (NSA) "Niagara Files" nime all sisekasutuseks. 2014. aastal anti projekt avatud lähtekohaga üle Apache Software Foundationile, muutudes osaks nende tehnoloogiasiirde algatustest [141].

Apache NiFi on mõeldud andmete kogumiseks, töötlemiseks ja edastamiseks reaajas. Erinevalt Airflow'st, mis töötab partiiülesannetega ja nõuab täpselt määratletud ajakava, töötab NiFi voogtöötamise režiimis, mis võimaldab andmete pidevat edastamist erinevate teenuste vahel.

Apache NiFi sobib ideaalselt IoT seadmete, ehitusandurite ja seiresüsteemide integreerimiseks ning näiteks CAD formaatide voogvõrgus valideerimiseks serveris, kus võib olla vajalik kohene reageerimine andmete muutustele.

Sisseehitatud filtreerimis-, teisendamis- ja marsruutimisvahendite abil võimaldab NiFi andmeid standardiseerida (Transform) enne nende edastamist (Load) salvestus- või analüüsisüsteemidesse. Üks selle peamisi eeliseid on sisseehitatud turvatugi ja juurdepääsukontroll, mis teeb sellest usaldusväärse lahenduse tundliku teabe käitlemiseks.



Joonis 7.4-5 Andmevoo graafiline kujutamine Apache NiFi liideses.

Apache NiFi tegeleb tõhusalt reaajas andmete voogedastuse, filtreerimise ja marsruutimise ülesannetega. See on ideaalne tehniliselt intensiivsete stsenaariumide jaoks, kus oluline on stabiilne teabeedastus süsteemide vahel ja suur läbilaskevõime.

Kui aga peamine eesmärk on integreerida erinevaid teenuseid, automatiseerida rutiinseid toiminguid ja luua kiiresti töövooge ilma sügavate programmeerimisoskusteta, on nõutud madala sisenemislävega ja maksimaalse paindlikkusega lahendused. Üks selline vahend on n8n - Low-Code /No-Code klassi platvorm, mis on keskendunud äritegevuse automatiseerimisele ja protsesside visuaalsele orkestreerimisele.

n8n Low-Code, No-Code protsessi orkestreerimine

n8n on avatud lähtekoodiga Low-Code / No-Code platvorm automatiseeritud töövoogude loomiseks, mida iseloomustab kasutusmugavus, paindlikkus ja võime kiiresti integreeruda paljude väliste teenustega.

No-Code on meetod digitaalsete toodete loomiseks ilma koodi kirjutamata. Kõik protsessi elemendid - alates loogikast kuni kasutajaliideseni - realiseeritakse ainult visuaalsete vahendite abil. No-Code platvormid on suunatud tehnilise taustata kasutajatele ja võimaldavad kiiresti luua automaatikaid, vorme, integratsioone ja veebirakendusi. Näide: kasutaja seab automaatse teavituste saatmise või Google Sheetsi integratsiooni drag-and-drop-liidese kaudu ilma programmeerimisoskusteta.

Avatud lähtekoodiga ja kohaliku kasutuselevõtu võimalustega n8n automatiseerimise ja ETL-pipeliinide loomise protsessides annab ettevõtetele täieliku kontrolli oma andmete üle, tagades samal ajal turvalisuse ja sõltumatuse pilvepakkujatest.

Erinevalt Apache Airflow'st, mis on orienteeritud jäiga orkestreerimisega arvutusülesannetele ja nõuab Pythoni keele oskust, pakub n8n visuaalset redaktorit, mis võimaldab skriptide koostamist ilma programmeerimiskeelte tundmist nõudmata (joonis 7.4-6). Kuigi selle kasutajaliides võimaldab luua automatiseeritud protsesse ilma koodi kirjutamata (No-Code), saavad kasutajad keerulisemates stsenaariumides lisada oma JavaScript ja Python -funktsioone, et laiendada võimalusi (Low-Code).

Low-Code on tarkvaraarenduse lähenemisviis, mille puhul rakenduse või protsessi põhiloogika luuakse graafilise kasutajaliidese ja visuaalsete elementide abil ning programmikoodi kasutatakse ainult funktsionaalsuse kohandamiseks või laiendamiseks. Low-Code platvormid võimaldavad oluliselt kiirendada lahenduste väljatöötamist, kaasates lisaks programmeerijatele ka põhiliste tehniliste oskustega ärikasutajaid. Näide: kasutaja võib ehitada äriprotsessi valmisplokkidest ja vajadusel lisada oma skripti JavaScriptis või Pythonis.

Kuigi n8n on positsioneeritud platvormina, mille puhul on madal sisenemislävi, on vaja põhilisi programmeerimisalaseid teadmisi, arusaamist veebitehnoloogiatest ja oskusi APIga töötamisel. Süsteemi paindlikkus võimaldab seda kohandada mitmesuguste ülesannete täitmiseks - alates automaatsest andmetöötlustest kuni sõnumitooja, asjade interneti -seadmete ja pilveteenustega integreerimiseni.

Peamised funktsioonid ja eelised, mis tulenevad n8n kasutamisest:

- **Avatud lähtekoodiga** ja kohalikud kasutuselevõtuvõimalused tagavad täieliku andmekontrolli, turvalisuse vastavuse ja sõltumatuse pilvepakkujatest.
- **Integratsioon üle 330 teenusega**, sealhulgas CRM, ERP, e-kaubandus, pilveplatvormid, sõnumitooja ja andmebaasid.
- **Stsenaariumide paindlikkus**: alates lihtsatest teadetest kuni keeruliste ahelateni koos -päringute API töötlemisega, otsustusloogikaga ja AI -teenuste ühendamisega.
- **JavaScript ja Python tugi**: kasutajad saavad vajaduse korral manustada kohandatud koodi, laiendades automatiseerimisvõimalusi.
- **Intuiitiivne visuaalne kasutajaliides**: võimaldab teil kiiresti konfigureerida ja visualiseerida kõiki protsessietappe.

Low-Code klassi platvormid pakuvad vahendeid, et luua digitaalseid lahendusi minimaalse koodiga, mistõttu on need ideaalsed meeskondadele, kellel puuduvad põhjalikud tehnilised teadmised, kuid kes vajavad protsesside automatiseerimist.

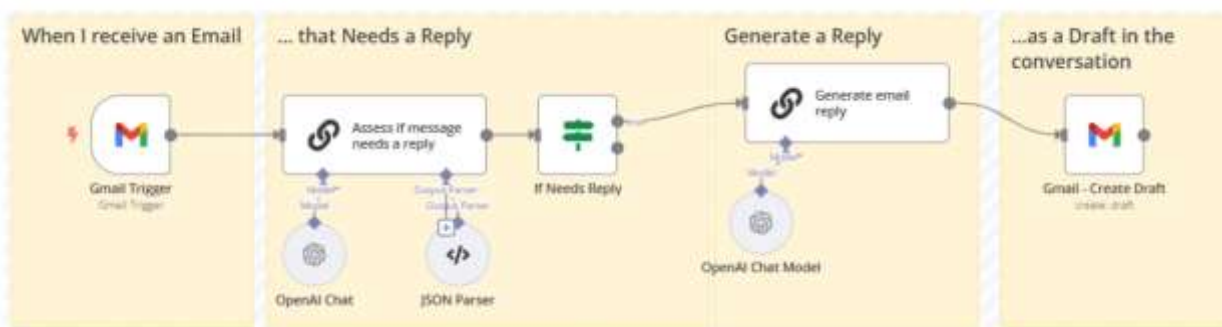
Ehituses saab n8n abil automatiseerida mitmesuguseid protsesse, näiteks integreeruda projektijuhtimissüsteemidega, kontrollida töövooge, kirjutada valmis aruandeid ja kirju, ajakohastada automaatselt materjalide inventuuriandmeid, saata meeskonnale teateid ülesannete seisu kohta ja

palju muud. Kohandatud Pipeline n8nis võib mitmekordselt vähendada käsitsi tehtavaid toiminguid, vähendada vigade tõenäosust ja kiirendada otsuste tegemist projekti teostamiseks.

Saate valida ligi kahe tuhande valmis, tasuta ja avatud lähtekoodiga n8n Pipeline, mis on saadaval aadressil: n8n.io/workflows, et automatiseerida nii ehitustööde kui ka isiklike ülesannete täitmist, vähendades rutiinseid toiminguid.

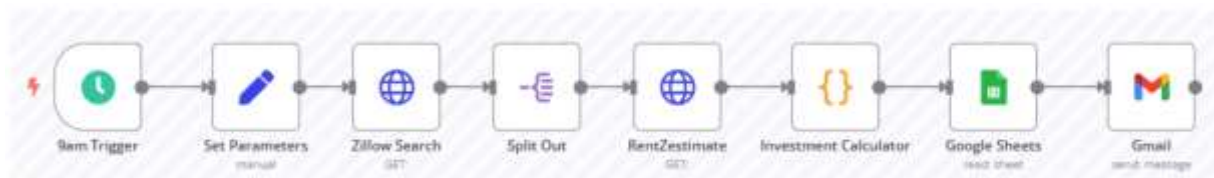
Võtame näiteks ühe valmis Pipeline'i malli, mis on tasuta saadaval aadressil n8n.io [142], mis loob Gmailis automaatselt vastuste eelnõud (joonis 7.4-6), aidates kasutajatel, kes saavad suure hulga e-kirju või kellel on raskusi vastuste koostamisega.

See n8n "Gmail AI Auto-Responder: Loo vastusevariantide eelnõud sissetulevatele e-kirjadele" mall (joonis 7.4-6) analüüsib sissetulevaid e-kirju, kasutades LLM ChatGPT-st, määrab kindlaks, kas vastus on vajalik, genereerib eelnõu ChatGPT-st ja konverteerib teksti HTML-ks ning lisab selle Gmaili sõnumite ahelasse. See ei saada automaatselt e-kirja, võimaldades teil vastust käsitsi muuta ja heaks kiita. Seadistamine võtab umbes 10 minutit ja hõlmab Gmaili API ja OpenAI API integreerimise OAuth- konfiguratsiooni. Tulemuseks on mugav ja tasuta lahendus rutiinse e-kirjavahetuse automatiseerimiseks, kaotamata seejuures kontrolli e-kirjade sisu üle.



Joonis 7.4-6 Automaatne e-posti vastuse genereerimise protsess, kasutades n8n.

Teine näide automatiseerimisest n8n abil on heade pakkumiste leidmine kinnisvaraturul [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", kogub igapäevaselt antud kriteeriumidele vastavaid asjakohaseid pakkumisi kasutades Zillow API. See arvutab automaatselt välja peamised investeerimisnäitajad (Cash on Cash ROI, Monthly Cash Flow, Down Payment), uuendab Google Sheets ja saadab kokkuvõtliku aruande e-posti teel (joonis 7.4-7), mis võimaldab investoritel säästa aega ja reageerida kiiresti parimatele pakkumistele.



Joonis 7.4-7 Kinnisvara investeringute atraktiivsuse hindamise automatiseeritud protsess.

Selle paindlikkus ja laiendatavus muudavad n8n väärtuslikuks vahendiks ettevõtetele, kes soovivad

digitaalselt ümber kujundada ja muutuda turul konkurentsivõimelisemaks suhteliselt lihtsate ja tasuta avatud lähtekoodiga vahendite abil.

Selliseid vahendeid nagu Apache NiFi, Airflow ja n8n võib käsitleda kolme andmetöötluskihina (joonis 7.4-8). NiFi haldab andmevoolu, tagades selle edastamise ja muundamise, Airflow orkestreerib ülesannete täitmist, koondades andmed töötlemisliinidesse, ja n8n automatiseerib integratsiooni väliste teenustega ja haldab äriloogikat.




	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Joonis 7.4-8 Apache Airflow, Apache NiFi ja n8n võib vaadelda kaasaegse andmehaldusarhitektuuri kolme üksteist täiendava kihina.

Koos moodustavad need tasuta ja avatud lähtekoodiga tööriistad potentsiaalselt näite tõhusast ökosüsteemist andmete ja protsesside haldamiseks ehitustööstuses, mis võimaldab ettevõtetel kasutada teavet otsuste tegemiseks ja protsesside automatiseerimiseks.

Järgmised sammud: üleminek manuaalsetelt toimingutelt analüütilistele lahendustele

Tänapäeva ehitusettevõtted tegutsevad suure ebakindluse keskkonnas: muutuvad materjalihinnad, hilinevad tarned, tööjõupuudus ja pingelised projektitähitajad. Analüütiliste armatuurtahvlite, ETL - konveierite ja BI-süsteemide kasutamine aitab ettevõtetel kiiresti tuvastada probleemseid valdkondi, hinnata ressursitõhusust ja prognoosida muutusi enne, kui need toovad kaasa rahalisi kahjusid.

Selle osa kokkuvõtteks tasub välja tuua peamised praktilised sammud, mis aitavad teil rakendada käsitletud tehnoloogiaid oma igapäevastes ülesannetes:

- Andmete visualiseerimise ja analüütiliste armatuurtahvlite rakendamine
 - ☐ Peamiste tulemusnäitajate (KPId) jälgimiseks vajalike näidikute loomise protsessi omandamine.
 - ☐ Kasutage oma andmete visualiseerimise vahendeid (Power BI, Tableau, Matplotlib, Plotly).

■ Automatiseerida andmetöötlust ETL -protsesside kaudu.

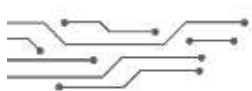
- ☐ Seadistage automaatne andmete kogumine erinevatest allikatest (dokumentatsioon, tabelid, CAD) ETL-protsesside kaudu.
- ☐ Andmete teisendamise korraldamine (nt regulaaravaldiste kontrollimine või arvutamine) Pythoni skriptide abil.
- ☐ Proovige FPDF-i raamatukogu abil luua automaatne PDF (või DOC) aruandlus, kasutades andmeid Exceli failidest või ekstraheerides teavet teistest PDF-dokumentidest.

■ Keelemudelite kasutamine (LLM) automatiseerimiseks

- ☐ kasutada suuri keelemudeleid (LLM), et luua koodi, mis aitab andmeid struktureerimata dokumentidest välja võtta ja analüüsida.
- ☐ Tutvuge n8n automatiseerimisvahendiga ning tutvuge nende veebisaidil olevate valmis mallide ja juhtumiuuringutega. Tehke kindlaks, milliseid teie tööprotsesse saab täielikult automatiseerida, kasutades No-Code/Low-Code lähenemist.

Analüütiline lähenemine andmete ja protsesside automatiseerimisele mitte ainult ei vähenda rutiinsetele toimingutele kuluvat aega, vaid parandab ka otsuste tegemise kvaliteeti. Ettevõtted, kes rakendavad visuaalse analüüsi vahendeid ja ETL -konveierid, saavad võimaluse reageerida muutustele kiiresti

Äriprotsesside automatiseerimine selliste vahendite abil nagu n8n, Airflow ja NiFi on alles esimene samm digitaalse küpsuse suunas. Järgmine samm on automatiseerimise aluseks olevate andmete kvaliteetne säilitamine ja haldamine. 8. osas vaatleme põhjalikult, kuidas ehitusettevõtted saavad luua jätkusuutliku andmesalvestusarhitektuuri, liikudes dokumentide ja mitmeformaadiliste failide kaosest tsentraliseeritud salvestus- ja analüüsiplatvormide juurde.





VIII OSA

ANDMETE SÄILITAMINE JA HALDAMINE EHITUSES

8. osas uuritakse kaasaegseid andmesalvestus- ja andmehaldustehnoloogiaid ehitussektoris. Selles analüüsitakse tõhusaid vorminguid suurte andmemahutude töötlemiseks - alates lihtsast CSV-st ja XLSX-st kuni suurema jõudlusega Apache Parquet'i ja ORC-ni, kusjuures võrreldakse üksikasjalikult nende võimalusi ja piiranguid. Käsitletakse andmeaida (DWH), andmejärvede) ja nende hübriidlahenduste (Data Lakehouse) kontseptsioone, samuti andmete haldamise) ja andmeminimalismi) põhimõtteid. Üksikasjalikult käsitletakse andmesoo probleeme) ja strateegiaid, kuidas vältida kaost infosüsteemides. Tutvustatakse uusi lähenemisviise andmetega töötamiseks, sealhulgas vektorandmebaasid ja nende rakendamine ehituses Bounding Boxi kontseptsiooni kaudu. Selles osas käsitletakse ka DataOps ja VectorOps metoodikaid kui uusi standardeid andmete töövoogude korraldamiseks.

PEATÜKK 8.1.

ANDMETARISTU: SALVESTUSFORMAATIDEST DIGITAALSETE HOIDLATENI

Andmeatomid: tõhusa teabehalduse alus

Kõik universumis koosneb kõige väiksematest ehitusplokkidest - aatomitest ja molekulidest - ning aja jooksul pöörduvad kõik elus ja eluta asjad paratamatult tagasi sellesse algseisundisse. Looduses toimub see protsess hämmastava kiirusega, mida me püüame üle kanda inimese poolt juhitavatele protsessidele.

Metsas muutuvad kõik elusorganismid lõpuks toitevaks aineks, mis on aluseks uutele taimedele. Need taimed omakorda saavad toiduks uutele elusolenditele, mis koosnevad samadest aatomitest, mis löid miljonite aastate eest Universumi.

Ärimaailmas on samuti oluline jaotada keerulised, mitmekihilised struktuurid kõige põhilisemateks, minimaalselt töödeldud üksusteks - nagu aatomid ja molekulid looduses. See võimaldab andmete aatomeid tõhusalt salvestada ja hallata, muutes need rikkalikuks, viljakaks aluseks, millest saab peamine ressurss analüütika ja otsuste kvaliteedi kasvuks.



Joonis 8.1-1 Analüüs ja otsuste tegemine põhineb korduvalt töödeldud ja salvestatud andmetel.

Muusikakompositsioonid koosnevad nootidest, mis kombineeruvad keerukate muusikapalade loomiseks, samas kui sõnad luuakse primitiivsest üksusest, tähest-häälest. Olgu tegemist looduse,

teaduse, majanduse, kunsti või tehnoloogiaga, maailm ilmutab märkimisväärset ühtsust ja harmooniat oma püüdlustes hävitada, struktureerida, tsirkuleerida ja luua. Samamoodi jaotatakse protsessid kulukalkulatsioonisüsteemides tillukesteks struktureeritud üksusteks - ressursiobjektideks - kulukalkulatsioonide ja ajakavade tasandil. Neid üksusi, nagu märkmeid, kasutatakse seejärel keerukamate kalkulatsioonide ja graafikute koostamiseks. Sama põhimõtet kasutavad ka arvutipõhised projekteerimissüsteemid, kus keerukad arhitektuuri- ja inseneriprojektid ehitatakse üles põhielementidest - üksikutest elementidest ja raamatukogu komponentidest, millest luuakse kompleksse hoone või ehitise projekti täielik 3D -model.

Loodusele ja teadusele omane tsüklilisuse ja struktuuri kontseptsioon kajastub ka kaasaegses andmemailmas. Nii nagu looduses pöörduvad kõik elusolendid tagasi aatomite ja molekulide juurde, nii kipub ka kaasaegsete andmetöötlusvahendite maailmas teave pöörduma tagasi oma kõige primitiivsemasse vormi.

Väikseimad elemendid, mis on lõplikult jagamatud, on äriprotsesside põhielemendid. Oluline on algusest peale hoolikalt kaaluda, kuidas neid eri allikatest pärit pisikesi ehitusplokke koguda, struktureerida (aatomiteks jaotada) ja salvestada. Andmete organiseerimine ja säilitamine ei ole lihtsalt nende osadeks jaotamine. Sama oluline on tagada nende integreerimine ja struktureeritud säilitamine, et andmeid oleks võimalik hõlpsasti leida, analüüsida ja kasutada otsuste tegemiseks, kui neid on vaja.

Andmete tõhusaks töötlemiseks tuleb hoolikalt valida andmete säilitamise vorm ja meetodid - nii nagu pinnas peab olema ette valmistatud puude kasvuks. Andmehoidlad tuleb korraldada nii, et tagada teabe kõrge kvaliteet ja asjakohasus, kõrvaldades üleliigsed või ebaolulised andmed. Mida paremini on see "teabemuld" struktureeritud, seda kiiremini ja täpsemalt suudavad kasutajad leida õiged andmed ja lahendada analüütilisi probleeme.

Teabe säilitamine: failid või andmed

Andmelaod võimaldavad ettevõtetel koguda ja kombineerida teavet erinevatest süsteemidest, luues ühtse keskuse edasiseks analüüsiks. Kogutud ajaloolised andmed võimaldavad mitte ainult protsesside sügavat analüüsi, vaid ka äritegevust mõjutavate mustrite tuvastamist.

Oletame, et ettevõtte töötab korraga mitme projektiga. Insener tahab aru saada, kui palju betooni on valatud ja kui palju tuleb veel osta. Traditsioonilise lähenemise korral peaks ta käsitsi serverist otsima ja avama mitu kalkulatsioonitabelit, võrdlema neid lõpetatud tööde sertifikaatidega ja kontrollima jooksvaid varude saldosid. See võtab tunde või isegi päevi. Isegi ETL-protsesside ja automaatsete skriptide puhul jääb ülesanne pooleldi manuaalseks: insener peab ikka veel käsitsi määrama serveris olevate kaustade või konkreetsete failide tee. See vähendab automatiseerimise üldist mõju, sest see võtab jätkuvalt väärtuslikku tööaega.

Andmehaldusele üleminekul saab insener serveri failisüsteemi asemel juurdepääsu ühtsele salvestusstruktuurile, kus teavet ajakohastatakse reaajajas. Üks päring - koodi, SQL -päringu või isegi LLM -agendi kõne kujul - võib koheselt anda täpseid andmeid jooksvate saldode, tehtud tööde ja eelseisvate tarnete kohta, kui andmed on eelnevalt ette valmistatud ja ühendatud andmelaos, kus ei

ole vaja rändada läbi kaustade, avada kümneid faile ja võrrelda väärtusi käsitsi.

Pikka aega kasutasid ehitusettevõtted PDF -dokumente, DWG -jooniseid, RVT -mudeleid ja sadu ja tuhandeid Exceli -tabeleid ja muid erinevaid formaate, mis on salvestatud ettevõtte serverites kindlatesse kaustadesse, mis raskendab teabe otsimist, kontrollimist ja analüüsimist. Selle tulemusel viiakse pärast projektide lõpetamist allesjäänud failid enamasti tagasi serverisse arhiveerimiskaustadesse, mida praktiliselt ei kasutata tulevikus. Selline traditsiooniline failipõhine andmesalvestus kaotab andmevoo kasvades oma asjakohasuse, kuna see on inimlikele vigadele vastuvõtlik.

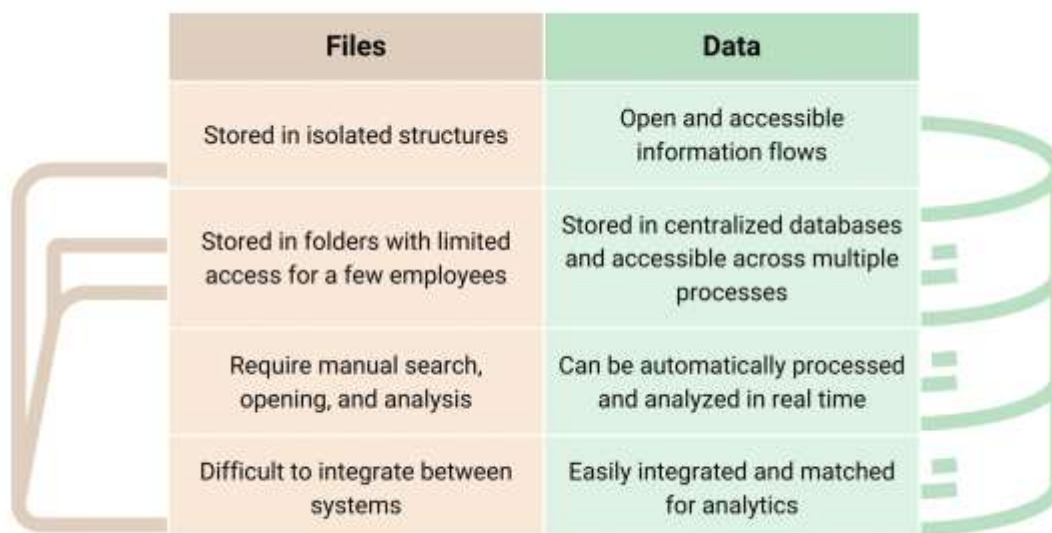
Fail on lihtsalt isoleeritud konteiner, kuhu andmed salvestatakse. Failid on loodud inimeste, mitte süsteemide jaoks, seega vajavad nad käsitsi avamist, lugemist ja tõlgendamist. Näiteks Excel -tabel, PDF -dokument või CAD -joonis, mis tuleb soovitud teabe saamiseks avada konkreetsetes tööriistas. Ilma struktureeritud otsingu ja töötlemiseta jääb selles sisalduv teave kasutamata.

Andmed on omakorda masinloetav teave, mida seostatakse, ajakohastatakse ja analüüsitakse automaatselt. Ühes andmelaos (nt andmebaas, DWH või Data Lake) on teave esitatud tabelite, kirjete ja seoste kujul. See võimaldab ühtset säilitamist, automatiseeritud päringuid, väärtuste analüüsi ja reaaliajase aruandlust.

Andmete kasutamine failide asemel (joonis 8.1-1) välistab manuaalse otsinguprotsessi ja ühtlustab töötlemist. Ettevõtted, kes on sellist lähenemisviisi juba rakendanud, saavad konkurentsieelise tänu kiirele juurdepääsule teabele ja võimele seda kiiresti äriprotsessidesse integreerida.

Üleminek failide kasutamiselt andmete kasutamisele on vältimatu muutus, mis kujundab ehitussektori tulevikku.

Iga ehitussektori ettevõtte seisab olulise valiku ees: kas jätkata teabe säilitamist eraldiseisvates failides ja silodes, mida inimesed peavad eriprogrammide abil lugema, või muuta see töötlemise esimestes etappides struktureeritud andmeteks, luues ühtse integreeritud digitaalse aluse automatiseeritud projektijuhtimiseks.



Joonis 8.1-1 Teabevoog areng: eraldiseisvatest failidest integreeritud andmeteni.

Infahulga tõttu muutuvad traditsioonilised failide säilitamise ja töötlemise meetodid üha vähem tõhusaks. Ehitustööstuses, nagu ka teistes sektorites, ei ole enam piisav tugineda erinevate failivormingutega failikaustadele või ühendamata andmebaasidele.

Ettevõtted, kes soovivad jääda digitaalajastul konkurentsivõimeliseks, liiguvad paratamatult integreeritud digitaalsete platvormide suunas, kasutavad suurandmete tehnoloogiaid ja automatiseeritud analüüsisüsteeme.

Üleminek failipõhiselt säilitamiselt andmepõhistele töövoogudele nõuab teabehalduse lähenemisviiside ümbermõtestamist ja teadlikku formaatide valimist, mis sobivad edasiseks integreerimiseks tsentraliseeritud hoidlatesse. See valik määrab, kui tõhusalt saab andmeid töödelda, kui kiiresti saab neile juurde pääseda ja kui lihtsalt saab neid integreerida ettevõtte digitaalprotsessidesse.

Suurandmete säilitamine: populaarsete vormingute ja nende tõhususe analüüs

Salvestusformaadid mängivad võtmerolli analüütilise infrastruktuuri skaleeritavuse, usaldusväärsuse ja jõudluse osas. Andmete analüüsimiseks ja töötlemiseks - näiteks filtreerimiseks, rühmitamiseks ja agregeerimiseks - kasutati meie näidetes Pandas DataFrame - populaarne struktuur andmetega töötamiseks RAM-is.

Kuid Pandas DataFrame ei oma salvestusformaati, seega pärast töötlemise lõppu eksporditakse andmed ühte välisformaati - enamasti CSV või XLSX. Neid tabeliformaate on lihtne vahetada ja need ühilduvad enamiku välissüsteemidega, kuid neil on mitmeid piiranguid: madal salvestamise tõhusus, pakkimise puudumine ja kehv versioonitoetus:

- **CSV** (Comma-Separated Values): lihtne tekstivorming, mida toetavad laialdaselt erinevad platvormid ja tööriistad. Seda on lihtne kasutada, kuid see ei toeta keerulisi andmetüüpe ega

pakkimist.

- **XLSX** (Excel Open XML Spreadsheet): Microsoft Exceli failivorming, mis toetab selliseid keerulisi funktsioone nagu valemid, diagrammid ja kujundus. Kuigi see on kasulik andmete käsitsi analüüsimiseks ja visualiseerimiseks, ei ole see optimeeritud suuremahuliseks andmetöötamiseks.

Lisaks populaarsetele tabeliformaatidele XLSX ja CSV on struktureeritud andmete tõhusaks salvestamiseks mitu populaarset formaati (joonis 8.1-2), millest igaühel on sõltuvalt konkreetsetest andmete salvestamise ja analüüsi nõuetest unikaalsed eelised:

- **Apache Parquet**: veeruline andmesalvestusformaad, mis on optimeeritud kasutamiseks andmeanalüüsisüsteemides. See pakub tõhusat andmete pakkimist ja kodeerimisskeeme, mis muudab selle ideaalseks keeruliste andmestruktuuride ja suurte andmete töötlemiseks.
- **Apache ORC** (Optimized Row Columnar): Sarnaselt Parquet'ile pakub ORC kõrget pakkimist ja tõhusat andmesalvestust. See on optimeeritud raskete lugemisoperatsioonide jaoks ja sobib hästi andmeparkide salvestamiseks.
- **JSON** (JavaScript Object Notation): kuigi JSON ei ole andmete salvestamise seisukohast nii tõhus kui binaarsed vormingud, näiteks Parquet või ORC, on see väga kättesaadav ja sellega on lihtne töötada, mistõttu on see ideaalne skriptide jaoks, kus loetavus ja veebiühilduvus on olulised.
- **Feather**: kiire, kerge ja kergesti kasutatav analüüsile orienteeritud binaarne veeruline andmesalvestusformaad. See on loodud andmete tõhusaks edastamiseks Pythoni (Pandas) ja R-i vahel, mistõttu on see suurepärane valik neid programmeerimiskeskondi hõlmavate projektide jaoks.
- **HDF5** (Hierarchical Data Format version 5): mõeldud suurte andmemahtude salvestamiseks ja korraldamiseks. See toetab paljusid andmetüüpe ja sobib hästi töötamiseks keeruliste andmekogumitega. HDF5 on eriti populaarne teadusarvutites, kuna see võimaldab tõhusalt salvestada ja kasutada suuri andmekogumeid.

		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)

Joonis 8.1-2 Andmeformaate võrdlus, mis näitab peamisi erinevusi salvestamise ja töötlemise aspektides.

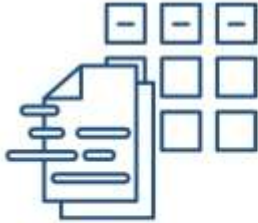
ETL-protsessi laadimisfaasis kasutatud vormingute võrdleva analüüsi tegemiseks koostati tabel, milles on esitatud failide suurused ja lugemisaeg (joonis 8.1-3). Uuringus kasutati identsete andmetega faile: tabel sisaldas 10 000 rida ja 10 veergu, mis olid täidetud juhuslike väärtustega.

Uuringusse on kaasatud järgmised salvestusformaadid: CSV, Parquet, XLSX ja HDF5, samuti nende tihendatud versioonid ZIP-arhiividenä. Toorandmed genereeriti NumPy raamatukogu abil ja esitati Pandas DataFrame'i struktuurina. Testimisprotsess koosnes järgmistest etappidest:

- Faili salvestamine: andmeraam salvestatakse neljas erinevas formaadis: CSV, Parquet, XLSX ja HDF5. Igal vormingul on unikaalsed omadused andmete salvestamisel, mis mõjutavad faili suurust ja lugemiskiirust.
- ZIP-faili pakkimine: standardse pakkimise tõhususe analüüsimiseks pakiti iga fail täiendavalt ZIP-arhiivi.
- Failide lugemine (ETL - Load): iga faili lugemisaeg mõõdeti pärast selle lahtipakkimist ZIP-ist. See võimaldab hinnata andmetele juurdepääsu kiirust pärast arhiivist väljavõtmist.

Oluline on märkida, et Pandas DataFrame'i ei kasutatud otseselt suuruse või lugemisaja analüüsimisel, kuna see ei ole omaette salvestusformaat. See oli ainult vahestruktuuriks andmete genereerimiseks ja hilisemaks salvestamiseks erinevatesse vormingutesse.



		Original Size (MB)	ZIP Size (MB)	Read Time (s)
Structured table 100000 rows x 10 columns 	CSV	18.82	8.73	0.48
	Parquet	9.67	9.22	0.17
	XLSX	12.59	12.59	41.73
	HDF5	8.50	7.62	0.13

Joonis 8.1-3 Salvestusformaatide võrdlus suuruse ja lugemiskiiruse järgi.

CSV ja HDF5 failid näitavad (joonis 8.1-3) suurt kompressiooni tõhusust, mis vähendab nende suurust märkimisväärselt, kui need pakitakse ZIP-varamusesse, mis võib olla eriti kasulik stsenaariumides, mis nõuavad salvestamise optimeerimist. XLSX failid seevastu on praktiliselt tihendamatud ja nende suurus ZIPis jääb võrreldavaks originaaliga, mistõttu on need vähem soodsad kasutamiseks suurtes andmemahitudes või keskkondades, kus on oluline andmete kättesaadavuse kiirus. Lisaks on XLSXi lugemisaeg võrreldes teiste formaatidega märkimisväärselt pikem, mistõttu on see vähem soodne andmete kiireks lugemiseks. Apache Parquet on näidanud suurt jõudlust analüütiliste ülesannete ja suurte andmemahutude puhul tänu oma veerulisele struktuurile.

Salvestusruumi optimeerimine Apache Parquet'iga

Üks populaarsemaid suurandmete salvestamise ja töötlemise formaate on Apache Parquet. See formaat on loodud spetsiaalselt veerulise salvestamise jaoks (sarnaselt Pandasile), mis võimaldab oluliselt vähendada mälu mahtu ja suurendada analüütiliste päringute kiirust. Erinevalt traditsioonilistest formaatidest, nagu CSV ja XLSX, toetab Parquet algupärast pakkimist ja on optimeeritud suurandmete süsteemide, sealhulgas Sparki, Hadoopi ja pilvesalvestuse jaoks.

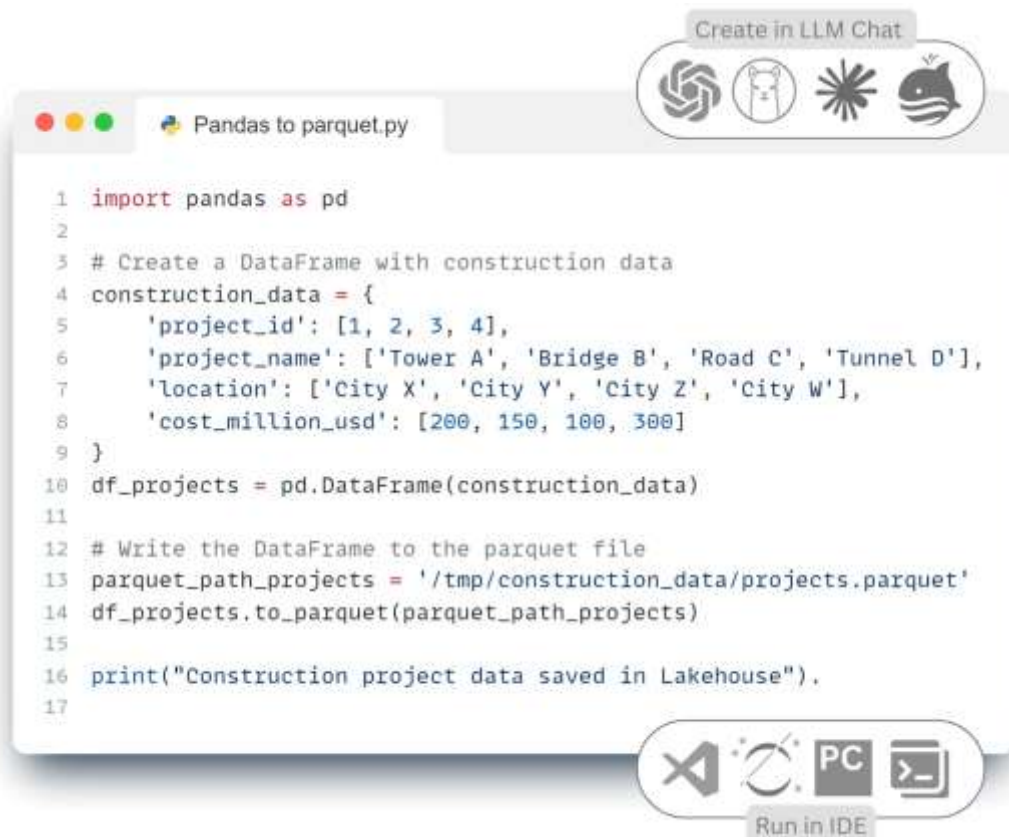
Parquet'i põhijoonte hulka kuulub andmete pakkimise ja kodeerimise tugi, mis vähendab oluliselt salvestusruumi suurust ja kiirendab andmete lugemisoperatsioone, töötades otse soovitud veergudel, mitte kõigil andmeridadel.

Illustreeriva näitena, kui lihtne on saada vajalik kood andmete konverteerimiseks Apache Parquet'ile, kasutame LLM-i.

🗨️ Saada tekstitaotlus LLM chat (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Kirjuta kood andmete salvestamiseks Pandasist DataFrame Apache Parquet'ile. ↩️

🗨️ LLMi vastus :



```

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

```

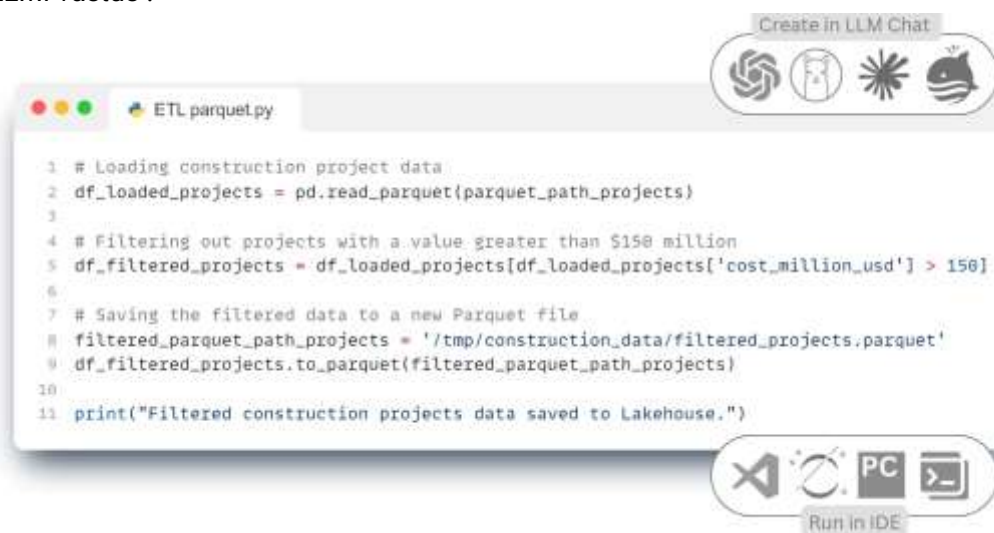
Joonis 8.1-4 Andmeraami andmete ülekandmine RAM-ist mälutõhusasse Apache Parquet formaati kasutades paari rida Pythoni .

Järgmine näide: simuleerime ETL-protsessi koos andmetega, mis on salvestatud kujul Parquet, et filtreerida projekte ühe atribuudi "cost_million_usd" teatud väärtuse järgi (joonis 8.1-4).

🗨️ Kui vestlus jätkub, saatke tekstisõnumitaotlus aadressile LLM:

Kirjutage kood, kus me tahame filtreerida andmeid tabelis ja salvestada ainult need projektid (tabeliread) Apache Parquet'i andmetest, mille maksumus (parameeter `cost_million_usd`) ületab 150 miljonit dollarit. ↵

🗨️ LLMi vastus :



```

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.")

```

Joonis 8.1-5 ETL-protsess Apache Parquet formaadis olevate andmetega töötamisel näeb välja sama nagu teiste struktureeritud formaatide puhul.

Parquet-vormingu kasutamine (võrreldes XLSX, CSV jne.) vähendab oluliselt salvestatava teabe hulka ja kiirendab otsingutoiminguid. See muudab selle suurepäraseks nii andmete säilitamiseks kui ka analüüsimiseks. Parquet integreerub erinevate töötlemissüsteemidega, pakkudes tõhusat juurdepääsu hübriidarhitektuurides.

Tõhus salvestusformaad on siiski ainult üks osa terviklikust andmekogemusest. Jätkusuutliku ja skaleeritava keskkonna loomiseks on vaja hästi kavandatud andmehaldusarhitektuuri. DWH (Data Warehouse) klassi süsteemid täidavad seda funktsiooni. Need pakuvad heterogeensetest allikatest pärit andmete koondamist, äriprotsesside läbipaistvust ja võimalust keeruliseks analüüsiks BI-vahendite ja masinõppe algoritmide abil.

andmeladu

Nii nagu Parquet formaad on optimeeritud suurte andmehulkade tõhusaks salvestamiseks, on andmeladu optimeeritud andmete integreerimiseks ja struktureerimiseks, et toetada analüütikat, prognoosimist ja juhtimisotsuste tegemist.

Tänapäeva ettevõtetes pärinevad andmed paljudest erinevatest allikatest: ERP, CAFM, CPM, CRM-süsteemid, raamatupidamis- ja laohaldus, hoonete digitaalsed CAD-mudelid, IoT-sensorid ja muud lahendused. Tervikliku pildi saamiseks ei piisa lihtsalt andmete kogumisest - need tuleb organiseerida,

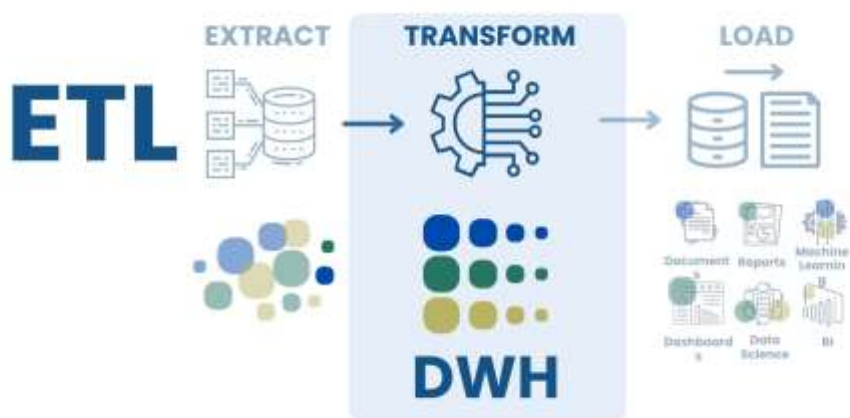
standardiseerida ja tsentraliseerida ühtsesse hoidlasse. Just seda teeb DWH - tsentraliseeritud salvestussüsteem, mis võimaldab koondada erinevatest allikatest pärit teavet, struktureerida seda ja teha see analüütika ja strateegilise juhtimise jaoks kättesaadavaks.

DWH (Data Warehouse) on tsentraliseeritud andmelaosüsteem, mis koondab teavet mitmest allikast, struktureerib seda ja teeb selle kättesaadavaks analüüsiks ja aruandluseks.

Paljudes ettevõtetes on andmed laiali erinevates süsteemides, mida me käsitlesime raamatu esimestes osades (joonis 1.2-4). DWH integreerib need allikad, tagades teabe täieliku läbipaistvuse ja usaldusväärsuse. DWH andmeladu on spetsialiseeritud andmebaas (suur andmebaas), mis kogub, töötleb ja säilitab andmeid mitmest allikast. DWH peamised omadused on järgmised:

- **Kasutades ETL -protsesse** (Extract, Transform, Load) - andmete väljavõtmine allikatest, puhastamine, teisendamine, laadimine repositooriumi ja nende protsesside automatiseerimine, mida käsitleti raamatu seitsmendas osas.
- **Andmete granulaarsus** - andmeid DWH-s saab salvestada nii agregeeritud kujul (koondaruanded) kui ka granuleeritud kujul (töötlemata andmed). Alates 2024. aastast on just CAD- müüjad hakanud rääkima granulaarsetest andmetest [125], mis võib-olla näitab, et tööstus valmistub üleminekuks spetsiaalsele pilvemajutusele digitaalsete ehitismudelite andmete töötlemiseks.
- **Analüütika toetamine ja ennustav** - andmelaod loovad aluse BI-vahenditele, Big Data - analüüsile ja masinõppe.

DWH on aluseks äriteabe, mis võimaldab põhiliste tulemusnäitajate analüüsi, müügi, ostude ja kulude prognoosimist ning automatiseeritud aruandlust ja andmete visualiseerimist (joonis 8.1-6).



Joonis 8.1-6 ETL -protsessis võib DWH toimida keskse repositooriumina, kus erinevatest süsteemidest ekstraheeritud andmed läbivad ümberkujundamise ja mahalaadimise etapid.

DWH mängib võtmerolli teabe integreerimisel, puhastamisel ja struktureerimisel, luues tugeva aluse äriteabe ja otsustusprotsesside jaoks. Kuid tänapäeva keskkonnas, kus andmemahud kasvavad

kiiresti ja andmeallikad muutuvad üha mitmekesisemaks, vajab traditsiooniline DWH lähenemine teabe säilitamisele sageli laiendamist ELT ja Data Lake näol

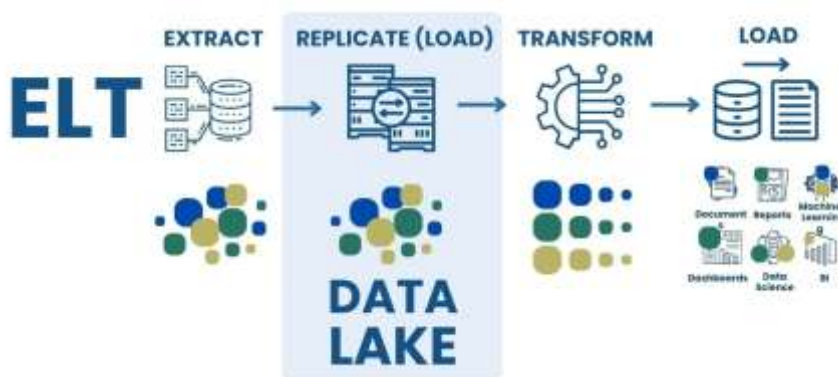
Andmejärv - ETLi areng kuni ELT: traditsioonilisest puhastamisest paindliku töötlemiseni.

Klassikalised andmelaod - andmelaod, mis on mõeldud struktureeritud andmete salvestamiseks analüütiliste päringute jaoks optimeeritud kujul, on kogenud piiranguid struktureerimata andmete käsitlemisel ja skaleeritavuses. Vastuseks nendele probleemidele on tekkinud andmejärv(d), mis pakuvad paindlikku ladustamist suurele hulgale heterogeensetele andmekogumitele.

Data Lake pakub alternatiivset DWH -lähenemist, mis võimaldab töötada struktureerimata, poolstruktureeritud ja töötlemata andmetega ilma eelneva jäiga skeemita. See salvestusmeetod on sageli asjakohane reaalajas andmete töötlemise, masinõppe ja täiustatud analüüsi jaoks. Erinevalt DWH-st, mis struktureerib ja agregeerib andmeid enne laadimist, võimaldab Data Lake salvestada teavet töötlemata kujul, pakkudes seega paindlikkust ja skaleeritavust

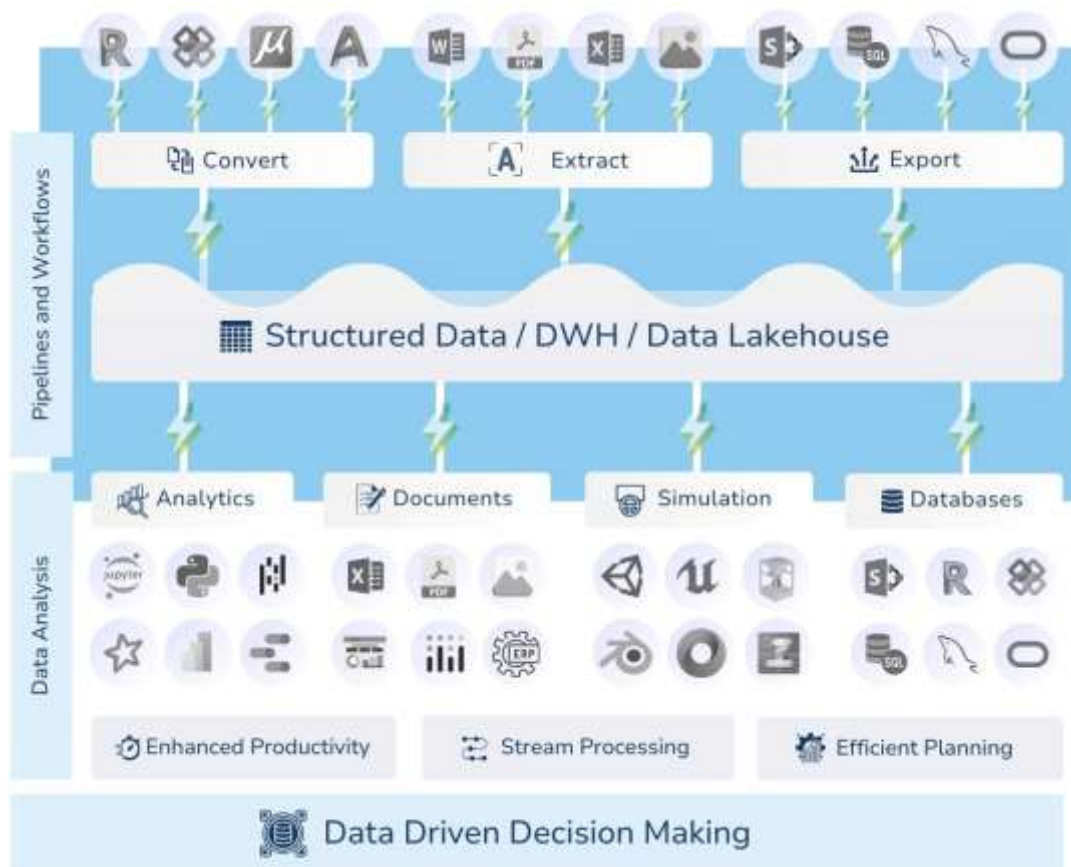
See oli pettumus traditsiooniliste andmelaodega (RDBMS, DWH) ja huvi "suurte andmete" vastu, mis viis andmete järvede tekkimiseni, kus keerulise ETLi asemel laaditakse andmed nüüd lihtsalt üles vabalt struktureeritud repositooriumi, kus töötlemine toimub analüüsi etapis:

- Traditsioonilistes andmeladudes töödeldakse, teisendatakse ja puhastatakse andmeid tavaliselt enne nende laoruumidesse laadimist (ETL - Extract, Transform, Load) (joonis 8.1-6). See tähendab, et andmed on struktureeritud ja optimeeritud konkreetsete tulevaste analüütiliste ja aruandlusülesannete jaoks. Põhirõhk on kõrge päringute jõudluse ja andmete terviklikkuse säilitamisel. Selline lähenemisviis võib siiski olla kulukas ja vähem paindlik uute andmetüüpide ja kiiresti muutuvate andmeskeemide integreerimise osas.
- Andmejärv(d) seevastu on mõeldud suurte koguste töötlemata andmete säilitamiseks nende algsel kujul (joonis 8.1-7). ETL (Extract, Transform, Load), protsess asendatakse ELT (Extract, Load, Transform), kus andmed laaditakse kõigepealt laoruumi "sellisena, nagu nad on" ja alles seejärel saab neid vastavalt vajadusele teisendada ja analüüsida. See tagab suurema paindlikkuse ja võimaluse salvestada heterogeenseid andmeid, sealhulgas struktureerimata andmeid, nagu tekst, pildid ja logid.



Joonis 8.1-7 Erinevalt ETList kasutab Data Lake ELT, kus teave laaditakse kõigepealt üles "toorel" kujul ja transformatsioon toimub üleslaadimise etapis.

Traditsioonilised andmelaod keskenduvad andmete eeltöötlemisele, et tagada päringute kõrge jõudlus, samas kui andmejärvad seavad esikohale paindlikkuse: nad salvestavad töötlemata andmeid ja muudavad neid vastavalt vajadusele (joonis 8.1-8).



Joonis 8.1-8 Kaasaegsete salvestuskontseptsioonide eesmärk on salvestada ja töödelda igat liiki andmeid otsuste tegemiseks.

Siiski, vaatamata kõigile eelistele ja andmejärvadele ei ole ilma puudusteta. Range struktuuri puudumine ja teabehalduse keerukus võivad viia kaoseni, kus andmed dubleerivad, on omavahel

vastuolus või muutuvad ebaoluliseks. Lisaks nõuab andmete otsimine ja analüüsimine sellises andmekogus märkimisväärselt pingutust, eriti kui tegemist on heterogeense teabega. Nende piirangute ületamiseks ning traditsiooniliste andmeida ja andmejärvede parimate omaduste ühendamiseks töötati välja Data Lakehouse'i arhitektuur.

Data Lakehouse'i arhitektuur: andmeida ja andmejärvede sünergia

DWH (struktureeritud, hallatav, suure jõudlusega analüütika) ja Data Lake (skaleeritavus, heterogeensete andmete käitlemine) parimate omaduste ühendamiseks töötati välja Data Lakehouse'i lähenemisviis. Selles arhitektuuris on ühendatud andmejärvede paindlikkus ja traditsioonilistele ladudele iseloomulikud võimsad töötlemis- ja haldusvahendid, saavutades tasakaalu salvestamise, analüüsi ja masinõppe vahel. Data Lakehouse on andmejärvede ja andmeladude süntees, ühendades esimeste paindlikkuse ja skaleeritavuse viimaste hallatavuse ja päringute optimeerimisega.

Data Lakehouse on arhitektuuriline lähenemisviis, mille eesmärk on ühendada andmejärvede paindlikkus ja skaleeritavus andmelao hallatavuse ja päringute sooritusvõimega (joonis 8.1-9).

Data Lakehouse'i peamised omadused on järgmised:

- **Avatud andmesalvestusformaat:** avatud formaatide kasutamine andmete salvestamiseks, näiteks Apache Parquet, tagab tõhususe ja optimeeritud päringud.
- Ainult lugemiseks mõeldud **skeem:** erinevalt DWH traditsioonilisest ainult kirjutamiseks mõeldud skeemi lähenemisviisist toetab Lakehouse ainult lugemiseks mõeldud skeemi, mis võimaldab suuremat paindlikkust andmestruktuuri haldamisel.
- **Paindlik ja skaleeritav:** toetab struktureeritud ja struktureerimata andmete salvestamist ja analüüsi, pakkudes suurt päringute jõudlust tänu salvestustasandi optimeerimisele.

Data Lakehouse pakub kompromisslahendust, mis ühendab mõlema lähenemisviisi eelised, mis muudab selle ideaalseks kaasaegsete analüütiliste töökoormuste jaoks, mis nõuavad paindlikkust andmete töötlemisel.



Joonis 8.1-9 Data Lakehouse on järgmise põlvkonna salvestussüsteemid, mis on loodud vastama keerulistele ja pidevalt muutuvatele nõuetele.

Kaasaegsete andmeladude idee tundub lihtne: kui kõik andmed on ühes kohas, on neid lihtsam analüüsida. Praktikas ei ole aga kõik nii sujuvalt. Kujutage ette, et ettevõtte otsustab täielikult loobuda tavapärastest raamatupidamis- ja juhtimissüsteemidest (ERP, PMIS, CAFM või muud), asendades need ühe tohutu andmepaiga, millele kõigil on juurdepääs. Mis juhtub? Tõenäoliselt tekib kaos: andmed dubleerivad, on vastuolulised ja kriitiline teave läheb kaduma või on rikutud. Isegi kui andmeparki kasutatakse ainult analüüsiks, on see ilma nõuetekohase haldamiseta tõsiselt ohustatud:

- Andmeid on raske mõista: tavapärastes süsteemides on andmetel selge struktuur, kuid järves on need lihtsalt tohutud failide ja tabelite kogumid. Et midagi leida, peab aru saama, mille eest iga rida ja veerg vastutab.
- Andmed võivad olla ebatäpsed: kui ühest ja samast teabest on salvestatud mitu versiooni, on raske teada, milline versioon on ajakohane. Selle tulemusel tehakse otsuseid vananenud või vigaste andmete põhjal.
- Andmete ettevalmistamine tööks on keeruline: andmeid tuleb mitte ainult salvestada, vaid ka esitada mugaval kujul - aruannete, graafikute, tabelite kujul. Traditsioonilistes süsteemides toimub see automaatselt, kuid andmejärvedes nõuab see täiendavat töötlemist.

Selle tulemusena on igal andmelao kontseptsioonil oma omadused, töötlemisviisid ja äirakendused. Traditsioonilised andmebaasid keskenduvad tehingutele, andmelaod (DWH) pakuvad struktuuri analüütika jaoks, andmejärved (Data Lake) salvestavad teavet töötlemata kujul ja hübriidlaod (Data Lakehouse) ühendavad DWH ja Data Lake eelised (joonis 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

Joonis 8.1-10 DWH, Data Lake ja Data Lakehouse: peamised erinevused andmetüüpide, kasutusstsenaariumide, töötlemismeetodite ja salvestusviiside vahel.

Salvestusarhitektuuri valimine on keeruline protsess, mis sõltub ärivajadustest, teabemahust ja analüütilistest nõuetest. Igal lahendusel on omad plussid ja miinused: DWH pakub struktuuri, Data Lake pakub paindlikkust ja Lakehouse pakub tasakaalu nende kahe vahel. Organisatsioonid on harva piiratud ühe andmearhitektuuriga.

Sõltumata valitud arhitektuurist on automatiseeritud andmehaldussüsteemid oluliselt paremad kui manuaalsed meetodid. Nad vähendavad inimlikke vigu, kiirendavad teabe töötlemist ning tagavad andmete läbipaistvuse ja jälgitavuse äriprotsesside kõikides etappides.

Ja kuigi tsentraliseeritud andmelaod on paljudes majandusvaldkondades juba muutunud tööstusstandardiks, on olukord ehituses endiselt killustatud. Andmed on siin hajutatud erinevatele platvormidele (CDE, PMIS, ERP jne), mis raskendab ühtse pildi loomist toimuvast ja nõuab arhitektuuri, mis suudab need allikad siduda ühtseks, analüütiliselt kasutatavaks digitaalseks keskkonnaks.

CDE, PMIS, ERP või DWH ja Data Lake

Mõned ehitus- ja insenerettevõtted kasutavad juba ühise andmekeskonna (CDE) kontseptsiooni vastavalt ISO 19650. Põhimõtteliselt täidab CDE samu funktsioone, mis andmeaida (DWH) teistes tööstusharudes: teabe tsentraliseerimine, versioonikontroll, juurdepääs valideeritud teabele.

Ühine andmekeskond (CDE) on tsentraliseeritud digitaalne ruum, mida kasutatakse projektiteabe haldamiseks, säilitamiseks, jagamiseks ja koostööks kogu rajatise elutsükli kõikides etappides. CDE on sageli rakendatud pilvepõhiste tehnoloogiate abil ja integreeritud CAD (BIM) süsteemidega.

Finants-, jaemüügi-, logistika- ja tööstussektorid on aastakümneid kasutanud tsentraliseeritud andmehaldussüsteeme, mis ühendavad eri allikatest pärit teavet, kontrollivad selle asjakohasust ja pakuvad analüütilist teavet. CDE viib need põhimõtted edasi, kohandades neid hoonete projekteerimise ja elutsükli juhtimise väljakutsetele.

Nagu DWH, struktureerib CDE andmeid, salvestab muudatused ja pakub ühtset juurdepääsu kontrollitud teabele. Pilveteenuste kasutuselevõttuga ja integreerimisega analüütiliste vahenditega muutuvad erinevused nende kahe süsteemi vahel üha vähem nähtavaks. Lisades CDE granuleeritud andmed, mille kontseptsiooni on CAD -pakkujad arutanud alates 2023[93, 125] , võib näha veelgi rohkem paralleele klassikalise DWH.

Varasemalt peatükis "Ehituse ERP ja PMIS süsteemid" oleme juba vaadelnud PMISi (projektijuhtimise infosüsteem) ja ERPi (ettevõtte ressursside planeerimine). Ehitusprojektides töötavad CDE ja PMIS koos: CDE on andmete, sealhulgas jooniste, mudelite ja projektdokumentatsiooni hoidla, PMIS aga haldab selliseid protsesse nagu tähtaegade, ülesannete, ressursside ja eelarvete kontrollimine.

ERP, mis vastutab ettevõtte kui terviku (rahandus, hanked, personal, tootmine) juhtimise eest, saab integreerida PMISiga, mis võimaldab kulude ja eelarve kontrolli ettevõtte tasandil. Analüüsi ja aruandluse jaoks saab DWH kasutada CDE, PMISi ja ERPi andmete kogumiseks, struktureerimiseks ja koondamiseks, et hinnata finantsnäitajaid (ROI) ja tuvastada mustreid. Data Lake (DL) võib omakorda täiendada DWH-d, salvestades töötlemata ja struktureerimata andmeid (nt logid, andurite andmed, pildid). Neid andmeid saab töödelda ja laadida DWH-sse edasiseks analüüsiks.

Seega keskenduvad CDE ja PMIS projektijuhtimisele, ERP äriprotsessidele ning DWH ja Data Lake analüütikale ja andmeoperatsioonidele.

Võrreldes CDE, PMIS ja ERP süsteemide DWH ja Data Lake süsteemidega, võib märgata olulisi erinevusi müüja sõltumatuse, kulude, integratsiooni paindlikkuse, andmete sõltumatuse, muudatustega kohanemise kiiruse ja analüütiliste võimete osas (joonis 8.1-11). Traditsioonilised süsteemid, nagu CDE, PMIS ja ERP, on sageli seotud konkreetsete müüjate lahendustega ja standarditega, mis muudab need vähem paindlikuks ja suurendab nende kulusid litsentside ja toetuse tõttu. Lisaks on andmed sellistes süsteemides sageli kapseldatud konfidentsiaalsetesse, suletud vormingutesse, mis piirab nende kasutamist ja analüüsi.

		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)

Joonis 8.1-11 DWH ja Data Lake pakuvad suuremat paindlikkust ja andmete sõltumatust kui sellised süsteemid nagu CDE, PMIS ja ERP.

Seevastu DWH ja Data Lake pakuvad suuremat paindlikkust erinevate andmeallikatega integreerimisel ning nende avatud tehnoloogiate ja platvormide kasutamine aitab vähendada kogukulu. Lisaks toetavad DWH ja Data Lake mitmesuguseid analüüsivahendeid, mis parandab analüüsi- ja haldusvõimalusi.

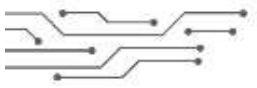
CAD-formaatide tagasipööramise vahendite ja CAD-rakenduste andmebaaside kättesaadavuse arenguga muutub üha teravamaks küsimus: kui põhjendatud on jätkata suletud, isoleeritud platvormide kasutamist, kui projekteerimisandmed peavad olema kättesaadavad laiale hulgale spetsialistidele, kes töötavad kümnetes töövõtjate ja projekteerimisorganisatsioonide juures?

Selline sõltuvus tarnijast sõltuvast tehnoloogiast võib oluliselt piirata andmehalduse paindlikkust, aeglustada reageerimist projekti muudatustele ja takistada osalejate tõhusat koostööd.

Traditsioonilised lähenemisviisid andmehaldusele - sealhulgas DWH, Data Lake, CDE ja PMIS - on keskendunud peamiselt teabe salvestamisele, struktureerimisele ja töötlemisele. Kuid koos tehisintellekti ja masinõppe arenguga kasvab vajadus uute andmete korraldamise viiside järele, mis mitte ainult ei koonda, vaid ka paljastavad keerulisi seoseid, leiavad varjatud mustreid ja võimaldavad

kohese juurdepääsu kõige asjakohasemale teabele.

Vektorandmebaasid - uut tüüpi salvestusruumid, mis on optimeeritud suuremõõtmeliste varjundite jaoks - hakkavad selles suunas erilist rolli mängima.



PEATÜKK 8.2.

ANDMELAO HALDAMINE JA KAOSE VÄLTIMINE

Vektorandmebaasid ja Bounding Box

Vektorandmebaasid on uus repositooriumide klass, mis ei salvesta lihtsalt andmeid, vaid võimaldavad otsingut tähenduse järgi, objektide võrdlemist semantilise läheduse järgi ja intelligentsete süsteemide loomist: alates soovitustest kuni automaatse analüüsi ja konteksti genereerimiseni. Erinevalt traditsioonilistest andmebaasidest, mis keskenduvad täpsetele kokkulangevustele, leiavad vektorandmebaasid sarnaseid objekte atribuutide alusel - isegi siis, kui täpne kokkulangevus puudub

Vektorandmebaas on spetsiaalne andmebaasi tüüp, mis salvestab andmeid mitmemõõtmeliste vektoritena, millest igaüks esindab teatavaid omadusi või omadusi. Nendel vektoritel võib olla erinev arv mõõtmeid, sõltuvalt andmete keerukusest (ühel juhul võib see olla mõni mõõde, teisel juhul — tuhandeid).

Vektorandmebaaside peamine eelis on pigem semantilise asjakohasuse kui väärtuste täpse vastavuse alusel tehtav otsing. SQL- ja Pandas -päringute asemel kasutatakse "võrdub" või "sisaldab" filtritega lähimate naabrite (k-NN) otsingut (k-NN-ist räägime lähemalt raamatu järgmises osas) tunnuste ruumis.

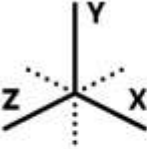
LLM (Large Language Models) ja generatiivsete mudelite arendamisega hakkab andmebaasidega suhtlemine muutuma. Nüüd on võimalik teha päringuid loomulikus keeles, teha dokumentidele semantilisi otsinguid, automaatselt võtmetermineid ja luua objektidevahelisi kontekstiüleseid seoseid - ja seda kõike ilma SQL oskuse või tabelistruktuuri tundmiseta. Seda käsitleti üksikasjalikumalt jaotises "LLMid ja nende roll andmetöötluses ja äriprotsessides".

Siiski on oluline mõista, et elukestva õppe dokumendid ei struktureeri ja sea teavet automaatselt korda. Mudel lihtsalt libiseb läbi andmete ja leiab päringu kontekstist lähtuvalt kõige asjakohasemad andmed. Kui andmeid ei ole eelnevalt puhastatud või teisendatud, on süvaotsing nagu katse leida vastus digitaalsest "prügist" - see võib küll toimida, kuid tulemuste kvaliteet on madalam. Ideaalis, kui andmeid saab struktureerida (nt tõlkida dokumendid Markdown'iks) ja laadida vektorandmebaasi. See suurendab oluliselt tulemuste täpsust ja asjakohasust.

Algselt kasutati vektorandmebaase masinõppes, kuid tänapäeval leiavad need üha rohkem rakendusi väljaspool seda - otsingumootorites, sisu personaliseerimises ja intelligentsetes analüüsis.

Üks ilmekamaid näiteid vektoripõhise lähenemise kohta ehituses on Bounding Box (piirnev paralleelpiloot). See on geomeetriline konstruktsioon, mis kirjeldab objekti piire kolmemõõtmelises ruumis. Bounding Box on määratletud minimaalsete ja maksimaalsete X-, Y- ja Z-koordinaatidega, mis moodustavad "kasti" ümber objekti. See meetod võimaldab hinnata elemendi suurust ja paigutust, ilma et oleks vaja analüüsida kogu geomeetriat.

Iga Bounding Box saab esitada vektorina mitmemõõtmelises ruumis: näiteks [x, y, z, laius, kõrgus, sügavus] - juba 6 mõõdet (joonis 8.2-1).



Bounding Box

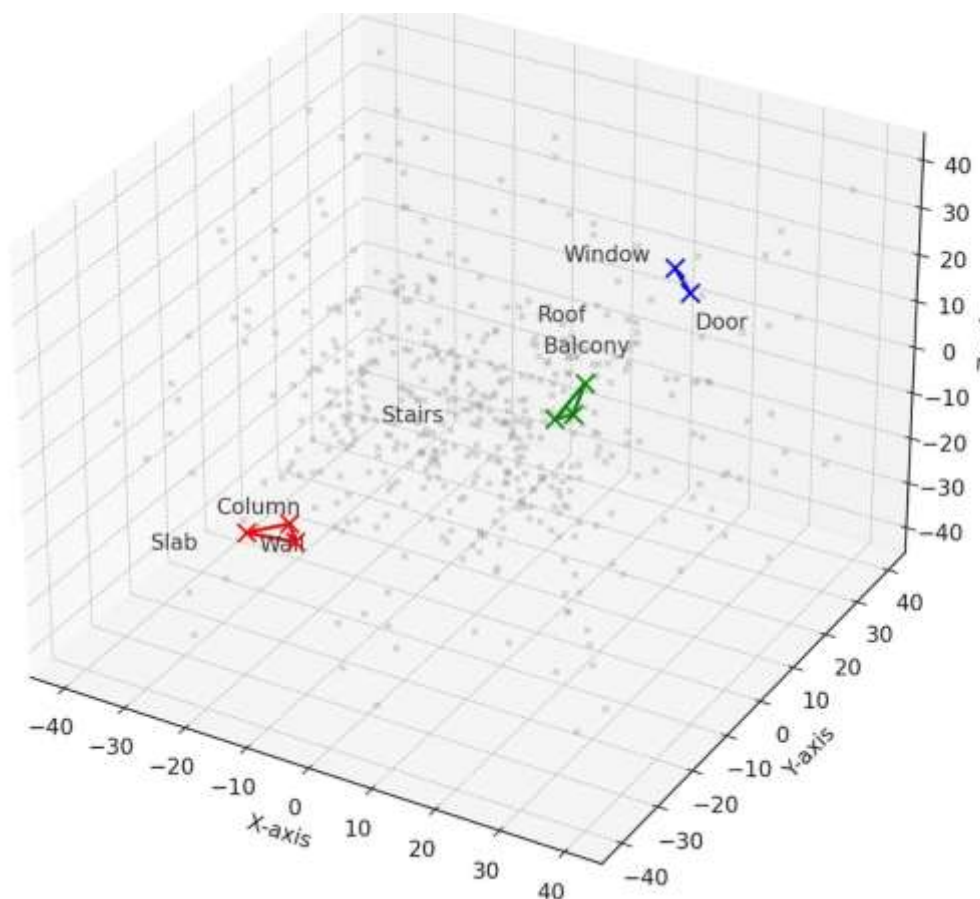
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Joonis 8.2-1 Bounding Box - elementide koordinaatide andmed ja nende asukoht projekti mudelis on analoogne vektorandmebaasiga.

See andmete esitusviis hõlbustab paljusid ülesandeid, sealhulgas objektide vaheliste ristumiste kontrollimist, ehitusdetailide ruumilise jaotuse planeerimist ja automaatsete arvutuste teostamist. Bounding Box võib olla sillaks keerukate 3D-mudelite ja traditsiooniliste vektorandmebaaside vahel, võimaldades teil tõhusalt kasutada mõlema lähenemisviisi eeliseid arhitektuuri- ja insenerimudelite modelleerimisel.

Bounding Box on "geomeetria vektoriseerimine" ja embedding (viis millegi abstraktse muutmiseks) on "täheenduse vektoriseerimine". Mõlemad lähenemisviisid võimaldavad liikuda käsitsi otsingult intelligentsele otsingule, olgu selleks siis 3D -objektid projekti mudelis või mõisted tekstis.

Objektide otsimine projektis (näiteks "leia kõik aknad, mille laius on > 1,5 m") on sarnane lähimate naabrite otsinguga (k-NN) vektorandmebaasis, kus kriteeriumid määratlevad "tsooni" tunnuste ruumis. (k-NN lähimate naabrite otsingust räägime lähemalt järgmises masinõpet käsitlevas osas) (joonis 8.2-2). Kui me lisame piiritletud kasti atribuutidele täiendavaid parameetreid (materjal, kaal, tootmisaeg), muutub tabel kõrgmõõtmeliseks vektoriks, kus iga atribuut on uus mõõde. See on lähemal kaasaegsetele vektorbaasidele, kus dimensioonid on loetud sadades või tuhandetes (nt neuronivõrkude sisseehitamine).



Joonis 8.2-2 Objektide otsimine projektis vektorandmebaaside abil.

Bounding Boxis kasutatav lähenemisviis ei ole kohaldatav mitte ainult geomeetriliste objektide, vaid ka teksti ja keele analüüsi puhul. Andmete vektoreid kasutatakse juba aktiivselt loomuliku keele töötlemisel (NLP). Nii nagu ehitusprojektis olevaid objekte saab rühmitada nende ruumilise läheduse järgi (joonis 8.2-2), saab tekstis olevaid sõnu analüüsida nende semantilise ja kontekstuaalse läheduse järgi.

Näiteks sõnad "arhitekt", "ehitus", "disain" on vektoriruumis üksteise kõrval, sest neil on sarnane tähendus. LLMis võimaldab see mehhanism automaatset, ei nõuta käsitsi kategoriseerimist:

- Teksti teema tuvastamine
- Dokumendi sisu semantilise otsingu teostamine
- Automaatsete märkuste ja tekstikokkuvõtete koostamine
- Leia sünonüümid ja nendega seotud mõisted

Vektorandmebaasid võimaldavad analüüsida teksti ja leida selles seotud mõisteid samamoodi, nagu Bounding Box aitab analüüsida ruumilisi objekte 3D -modelites. Bounding Boxi näide projektelementide kohta aitab mõista, et vektoriesitus ei ole puhtalt "kunstlik" kontseptsioon ML-st, vaid loomulik viis andmete struktureerimiseks rakenduslike probleemide lahendamiseks, olgu selleks

siis veerude otsimine CAD-projektis või semantiliselt lähedaste piltide otsimine andmebaasis.

Andmebaasidega töötavad spetsialistid peaksid pöörama tähelepanu vektormäludele. Nende levik näitab uut etappi andmebaaside arengus, kus klassikalised relatsioonilised süsteemid ja tehisintellektile orienteeritud tehnoloogiad hakkavad omavahel põimuma, moodustades tuleviku hübriidlahendusi.

Kasutajad, kes arendavad keerulisi ja suuremahulisi tehisintellekti -rakendusi, kasutavad vektorotsinguks spetsiaalseid andmebaase. Samal ajal valivad need, kes vajavad ainult eraldi AI-funktsioone olemasolevatesse rakendustesse integreerimiseks, pigem juba kasutatavatesse andmebaasidesse (PostgreSQL, Redis) sisseehitatud vektorotsinguvõimalusi.

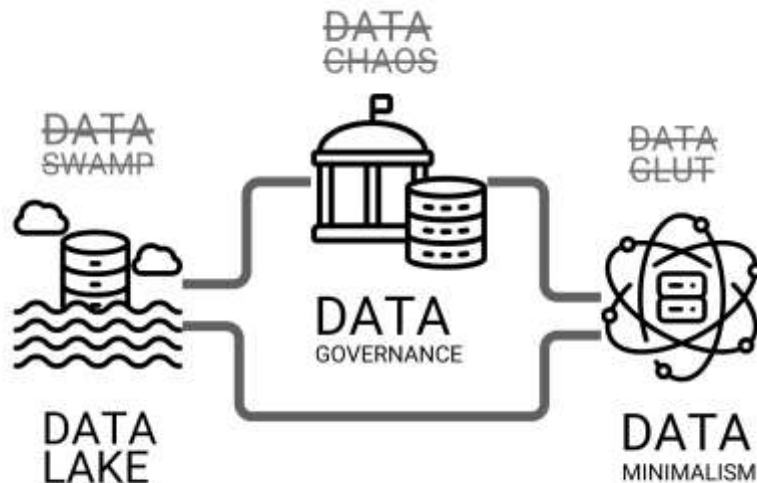
Kuigi sellised süsteemid nagu DWH, Data Lake, CDE, PMIS, vektorandmebaasid ja muud pakuvad erinevaid lähenemisviise andmete säilitamisele ja haldamisele, ei määra nende tõhusust mitte ainult nende arhitektuur, vaid ka see, kui hästi on andmed ise korraldatud ja hallatud. Isegi kui kasutatakse kaasaegseid lahendusi - olgu selleks siis vektorandmebaasid, klassikalised relatsioonilised andmebaasid või Data Lake'i tüüpi andmehoidlad - võib selgete reeglite puudumine andmete haldamiseks, struktureerimiseks ja ajakohastamiseks viia samade raskusteni, millega puutuvad kokku kasutajad, kes töötavad erinevate failide ja mitmeformaadiliste andmetega.

Ilma andmehalduseta) võivad isegi kõige võimsamad lahendused muutuda kaootiliseks ja struktureerimata, muutes andmejärved andmesoodeks). Selle vältimiseks peavad ettevõtted mitte ainult valima õige salvestusarhitektuuri, vaid ka rakendama andmeminimalismi), juurdepääsu haldust ja kvaliteedikontrolli strateegiaid, et muuta andmed tõhusaks otsustamisvahendiks.

Andmehaldus), andmeminimalism) ja andmesoo)

Andmehalduse), andmeminimalismi) ja andmesoo vältimise) mõistete mõistmine ja rakendamine on andmeaida edukaks haldamiseks ja ärilise väärtuse loomiseks võtmetähtsusega (joonis 8.2-3).

Gartneri (2017) uuringu kohaselt ebaõnnestub 85% suurtest andmeprojektidest ning üks peamisi põhjusi on ebapiisav andmete kvaliteet ja andmete juhtimine [144].



Joonis 8.2-3 Andmehalduse mõned peamised aspektid on andmehaldus ja andmeminimalism.

Andmehaldus (Andmehaldus) on andmehalduse põhikomponent, mis tagab, et andmeid kasutatakse kõikides äriprotsessides asjakohaselt ja tõhusalt. See ei tähenda mitte ainult eeskirjade ja menetluste kehtestamist, vaid ka andmete kättesaadavuse, usaldusväärsuse ja turvalisuse tagamist:

- andmete määratlemine ja klassifitseerimine: üksuste selge määratlemine ja klassifitseerimine võimaldab organisatsioonidel mõista, milliseid üksusi on ettevõttes vaja ja kuidas neid kasutada.
- Juurdepääsuõigused ja haldamine: andmetele juurdepääsu ja andmete haldamise poliitika ja korra väljatöötamine tagab, et ainult volitatud kasutajad saavad juurdepääsu teatud andmetele.
- Andmete kaitsmine väliste ohtude eest: andmete kaitsmine väliste ohtude eest on andmehalduse oluline aspekt. See ei hõlma mitte ainult tehnilisi meetmeid, vaid ka töötajate koolitust infoturbe aluste osas.

Andmete minimalism (Data Minimalism) on lähenemine, mille eesmärk on vähendada andmeid kõige väärtuslikumate ja mõttekamate atribuutide ja üksuste moodustamiseks (joonis 8.2-4), vähendades seeläbi kulusid ja parandades andmete kasutamist:

- Otsustamise lihtsustamine: objektide ja nende atribuutide arvu vähendamine kõige olulisematele lihtsustab otsuste tegemist, vähendades andmete analüüsimiseks ja töötlemiseks kuluvat aega ja ressursse.
- Keskendumine olulisele: kõige olulisemate üksuste ja atribuutide valimine võimaldab keskenduda äri jaoks tõesti olulisele teabele, kõrvaldades müra ja ebavajalikud andmed.
- Tõhus ressursside jaotamine: andmete minimeerimine võimaldab ressursside tõhusamat jaotamist, vähendades andmete säilitamise ja töötlemise kulusid, parandades andmete kvaliteeti ja turvalisust.

Andmetega töötamise loogika ei peaks algama mitte nende loomisest kui sellisest (joonis 8.2-4), vaid arusaamisest tulevatest stsenaariumidest, kuidas neid andmeid kasutada, juba enne andmete

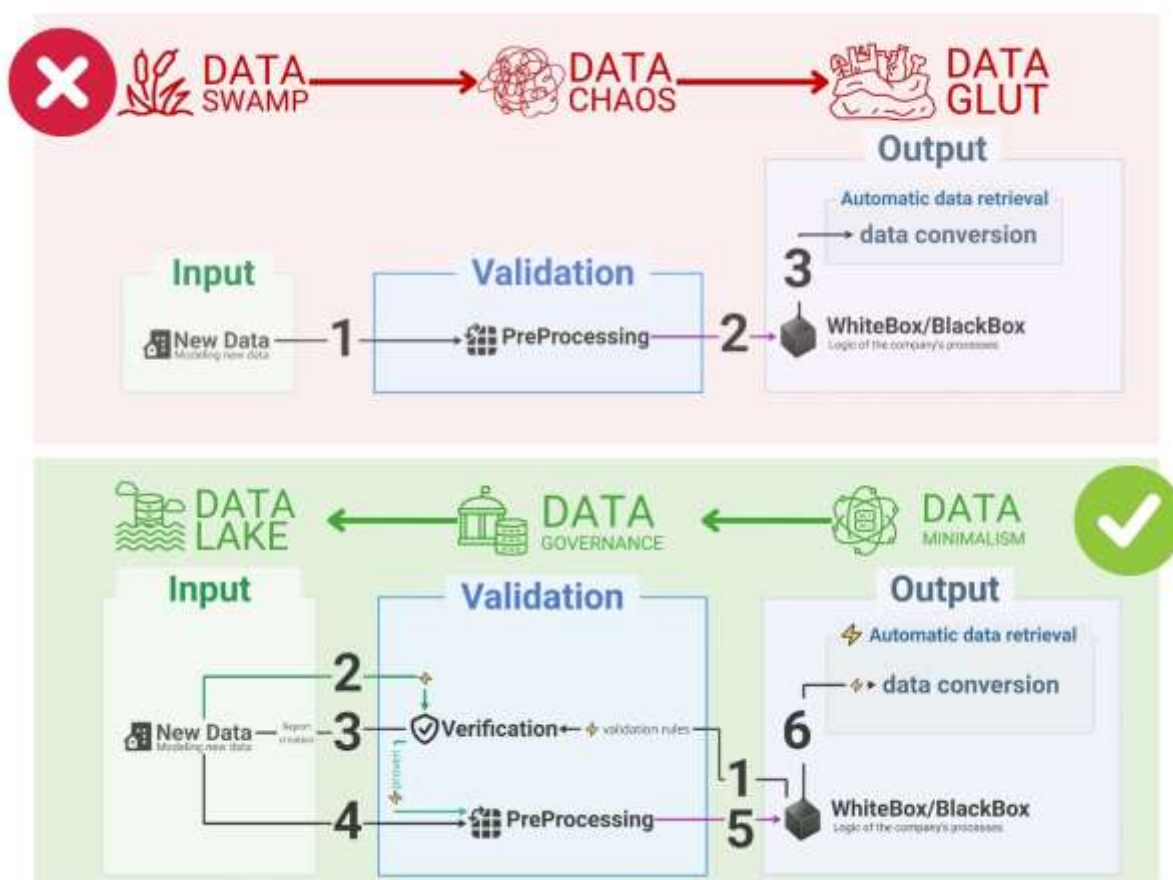
genereerimise protsessi algust. Selline lähenemisviis võimaldab eelnevalt määratleda minimaalsed vajalikud nõuded atribuutidele, nende tüüpidele ja piirväärtustele. Need nõuded on aluseks korrektsete ja stabiilsete üksuste loomisele infomudelil. Esialgne arusaam andmete eesmärkidest ja kasutusalaadest aitab kaasa analüüsiks sobiva struktuuri moodustamisele. Kontseptuaalse, loogilise ja füüsilise tasandi andmemodelleerimise lähenemisviise käsitleti üksikasjalikumalt peatükis "Andmemodelleerimine: kontseptuaalne, loogiline ja füüsiline mudel".

Ehitusettevõtete traditsioonilistes äriprotsessides sarnaneb andmetöötlus pigem andmete sootuks mahapanemisega, kus andmed luuakse kõigepealt ja seejärel püüavad spetsialistid neid integreerida teistesse süsteemidesse ja vahenditesse.

Andmesummi (Data Swamp) on andmete kontrollimatu kogumise ja säilitamise tulemus ilma nõuetekohase organiseerimise, struktureerimise ja haldamiseta, mille tulemuseks on struktureerimata, raskesti kasutatav ja väheväärtuslik teave.

Kuidas vältida, et teabevool ei muutuks sootuks:

- **Andmete struktuuri haldamine:** andmete struktureerimise ja kategoriseerimise tagamine aitab vältida andmete üleujutamist, muutes need korrastatuks ja kergesti kättesaadavaks.
- Andmete **mõistmine ja tõlgendamine:** andmete päritolu, muutuste ja tähenduse selge kirjeldus tagab andmete mõistmise ja õige tõlgendamise.
- **Andmete kvaliteedi säilitamine:** andmete korrapärane hooldus ja puhastamine aitab säilitada andmete kvaliteeti, asjakohasust ja väärtust analüütika ja äriprotsesside jaoks.



Joonis 8.2-4 Andmelao segaduse vältimiseks peaksite andmete loomise protsessi alustama atribuutide nõuete kogumisega.

Integreerides andmehalduse ja andmeminimalismi põhimõtted andmehaldusprotsessidesse ning ennetades aktiivselt, et andmelaod ei muutuks andmesoodeks, saavad organisatsioonid maksimeerida oma andmete potentsiaali.

Järgmine etapp andmetega töötamise arengus, pärast juhtimise ja minimalismi küsimuste lahendamist, on automaatse töötlemise standardimine, kvaliteedi tagamine ja selliste meetodite rakendamine, mis muudavad andmed kasutatavaks analüüsiks, ümberkujundamiseks ja otsuste tegemiseks. Sellega tegelevad DataOps ja VectorOps meetodid, mis on muutumas oluliseks tööriistaks suurandmete ja masinõppega töötavatele ettevõtetele.

DataOps ja VectorOps: uued andmestandardid

Kui andmehaldus vastutab andmete kontrollimise ja korraldamise eest, siis DataOps aitab tagada andmete täpsuse, järjepidevuse ja sujuva liikumise ettevõttes. See on eriti oluline mitmete ehitusettevõtete puhul, kus andmeid luuakse pidevalt ja need vajavad õigeaegset töötlemist. Näiteks olukordades, kus ehitusinfo mudelid, projekti nõuded ja analüütilised aruanded tuleb ühe tööpäeva

jooksul eri süsteemide vahel sünkroniseerida, võib DataOps'i roll olla võtmetähtsusega. See võimaldab luua stabiilseid ja korratavaid andmetöötlusprotsesse, vähendades viivituste ja teabe asjakohasuse kaotamise ohtu.

Andmehaldusest üksi ei piisa - oluline on, et andmeid mitte ainult ei säilitataks, vaid et neid kasutataks aktiivselt igapäevases tegevuses. Siinkohal tuleb mängu DataOps - metoodika, mis keskendub automatiseerimisele, integreerimisele ja pidevale andmevoolule.

DataOps keskendub organisatsioonide andmevoogude koostöö, integreerimise ja automatiseerimise parandamisele. DataOps-tavade kasutuselevõtt soodustab andmete täpsust, järjepidevust ja kättesaadavust, mis on andmekeskse rakenduse jaoks kriitilise tähtsusega.

DataOps ökosüsteemi peamised tööriistad on Apache Airflow (joonis 7.4-4) töövoogude orkestreerimiseks ja Apache NiFi (joonis 7.4-5) andmevoogude marsruutimiseks ja muundamiseks. Koos võimaldavad need tehnoloogiad paindlikke, usaldusväärseid ja skaleeritavaid andmepiire süsteemide vahelise teabe automaatseks töötlemiseks, kontrollimiseks ja integreerimiseks (täpsemalt peatükis "Automaatne ETL -konveier "). DataOps-lähenemise rakendamisel ehitusprotsessides on oluline arvestada nelja põhiaspekti:

1. **Inimesed ja vahendid on olulisemad kui andmed:** siilutud andmehoidlad võidakse pidada suureks probleemiks, kuid tegelikkus on keerulisem. Lisaks andmete killustatusele mängivad olulist rolli meeskondade eraldatus ja nende kasutatavad erinevad vahendid. Ehituses töötavad andmetega eri valdkondade spetsialistid: andmeinsenerid ja analüütikud, BI- ja visualiseerimismeeskonnad ning projektijuhtimise ja kvaliteedi eksperdid. Igaühel neist on erinevad tööviisid, seega on oluline luua ökosüsteem, kus andmed liiguvad vabalt osalejate vahel, pakkudes ühtset ja järjepidevat versiooni teabest.
2. **Testimise ja vigade tuvastamise automatiseerimine:** ehitusandmed sisaldavad alati vigu, olgu need siis mudelite ebatäpsused, arvutusviga või vananenud spetsifikatsioonid. Andmete korrapärane testimine ja korduvate vigade kõrvaldamine võib oluliselt parandada andmete kvaliteeti. DataOps raames tuleb rakendada automatiseeritud kontrolli- ja valideerimismehhanismid, mis jälgivad andmete õigsust, analüüsivad vigu ja tuvastavad mustreid ning registreerivad ja lahendavad süsteemi vead igas töövoos. Mida suurem on automatiseeritud valideerimise tase, seda parem on andmete üldine kvaliteet ja seda väiksem on vigade tõenäosus lõppetappides.
3. **Andmeid tuleks testida samamoodi nagu programmikoodi:** enamik ehitusrakendusi põhineb andmetöötlusel, kuid selle kontroll on sageli jäetud teisejärguliste rollidele. Kui masinõppe mudeleid treenitakse ebatäpsete andmete põhjal, toob see kaasa ebaõigeid prognoose ja rahalisi kahjusid. DataOpsis tuleks andmeid kontrollida sama rangelt kui tarkvarakoodi: loogikakontrollid, stressitestid, mudelite käitumise hindamine sisendväärtuste muutumisel. Juhtimisotsuste aluseks saab võtta ainult valideeritud ja usaldusväärseid andmeid.
4. **Andmete jälgitavus ilma tulemuslikkust ohustamata:** andmete jälgimine ei ole lihtsalt mõõdikute kogum, vaid strateegiline kvaliteedijuhtimise vahend. Selleks, et DataOps toimiks tõhusalt, peab jälgitavus olema sisse ehitatud andmekäitluse kõikidesse etappidesse, alates kavandamisest kuni kasutamiseni. Samal ajal on oluline, et jälgimine ei aeglustaks süsteemi.

Ehitusprojektide puhul on ülioluline mitte ainult koguda andmeid, vaid teha seda nii, et andmeid loovate spetsialistide (nt projekteerijate) töö ei katkeks kuidagi. Selline tasakaal võimaldab kontrollida andmete kvaliteeti ilma tootlikkust kahjustamata.

DataOps ei ole andmeteadlaste jaoks mitte lisakoormus, vaid nende töö selgroog. DataOps'i rakendades saavad ehitusettevõtted liikuda kaootilisest andmehaldusest tõhusa ökosüsteemi suunas, kus andmed töötavad ettevõtte heaks.

VectorOps on omakorda järgmine etapp DataOps arengus, mis keskendub mitmemõõtmeliste vektorandmete töötlemisele, salvestamisele ja analüüsile (mida käsitleti eelmises peatükis). See on eriti oluline sellistes valdkondades nagu digitaalsed kaksikud, närvivõrgumudelid ja semantiline otsing, mis hakkavad jõudma ehitussektorisse. VectorOps tugineb vektorandmebaasidele, et tõhusalt salvestada, indekseerida ja otsida objektide mitmemõõtmelisi esitusi.

VectorOps on järgmine samm pärast DataOps, mis keskendub vektorandmete töötlemisele, analüüsile ja kasutamisele ehituses. Erinevalt DataOpsist, mis keskendub andmevoogudele, järjepidevusele ja kvaliteedile, keskendub VectorOps masinõppe jaoks vajalike mitmemõõtmeliste objektide representatsioonide haldamisele.

Erinevalt traditsioonilistest lähenemisviisidest võimaldab VectorOps saavutada täpsemaid objektkirjeldusi, mis on kriitilise tähtsusega digitaalsete kaksikute, generatiivsete disainisüsteemide ja vektorformaati teisendatud CAD-andmete automaatse veatuvastuse jaoks. DataOps'i ja VectorOps'i kombineeritud rakendamine moodustab kindla aluse skaleeritavaks, automatiseeritud tööks suurte andmemahtudega - alates klassikalistest tabelitest kuni semantiliselt rikkalike ruumiliste mudeliteni

Järgmised sammud: kaootilisest ladustamisest struktureeritud ladustamiseni

Traditsioonilised lähenemisviisid andmesalvestuse loomisele toovad sageli kaasa erinevate "teabesilode" tekkimise, kus oluline teave on analüüsi ja otsuste tegemise jaoks kättesaamatu. Kaasaegsed salvestuskontseptsioonid, nagu andmeladu, andmejärv ja nende hübriidid, võimaldavad ühendada erinevad andmed ja teha need tsentraliseeritult kättesaadavaks andmete voogedastuse ja ärianalüüsi jaoks. Oluline on mitte ainult valida õige salvestusarhitektuur, vaid ka rakendada andmehaldust) ja andmeminimalismi), et vältida salvestusruumide muutumist kontrollimatuteks andmesoodeks).

Selle osa kokkuvõtteks tasub välja tuua peamised praktilised sammud, mis aitavad teil rakendada käsitletud kontseptsioone oma igapäevaste ülesannete täitmisel:

- Valige tõhusad andmesalvestusformaadid
 - ☐ Üleminek CSV ja XLSX asemel tõhusamate vormingute (Apache Parquet, ORC) kasutamine suurte andmemahtude salvestamiseks.
 - ☐ Andmete versioonisüsteemi rakendamine muudatuste jälgimiseks
 - ☐ Kasutage metaandmeid teabe struktuuri ja päritolu kirjeldamiseks.

■ Luua ettevõtte ühtne andmearhitektuur

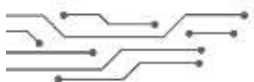
- ☐ Võrrelda erinevaid salvestusarhitektuure: RDBMS, DWH ja Data Lake. Valige see, mis vastab kõige paremini teie skaleeritavuse, allikate integreerimise ja analüütilise töötlemise vajadustele.
- ☐ Kujundage oma ülesannete jaoks protsessikaart andmete väljavõtmiseks, laadimiseks ja teisendamiseks (ETL) erinevatest allikatest. Kasutage visualiseerimisvahendeid, nagu Miro, Lucidchart või Draw.io, et visualiseerida põhietappe ja integratsioonipunkte.

■ Andmehalduspraktika rakendamine ja andmeminimalism

- ☐ Järgige andmeminimalismi lähenemisviisi - salvestage ja töötle ainult seda, mis on tõeliselt väärtuslik.
- ☐ Andmehalduse põhimõtete rakendamine - andmete eest vastutuse määratlemine, kvaliteedi ja läbipaistvuse tagamine.
- ☐ Lisateave andmehalduspõhimõtete ja DataOps-kontseptsioonide kohta, VectorOps
- ☐ Andmete kvaliteedikriteeriumide määratlemine ja andmete valideerimise menetlused DataOpsis.

Hästi korraldatud andmesalvestus loob aluse ettevõtte analüütiliste protsesside tsentraliseerimiseks. Üleminek failide kaootiliselt kogunemiselt struktureeritud salvestusruumidele võimaldab muuta teabe strateegiliseks varaks, mis aitab teha teadlikke otsuseid ja parandada äriprotsesside tõhusust.

Kui andmete kogumise, ümberkujundamise, analüüsi ja struktureeritud säilitamise protsessid on automatiseeritud ja standardiseeritud, on digitaalsete muutuste järgmine etapp suurandmete täielik käsitlemine.





IX OSA

SUURED ANDMED, MASINÕPE JA PROGNOOSID

Üheksandas osas keskendutakse suurtele andmetele, masinõppele ja prognoositavale analüüsile ehitustööstuses. Selles uuritakse üleminekut intuiitselt otsuste langetamiselt objektiivsetele analüüsidele, mis põhinevad ajaloolistel andmetel. Suurandmete analüüsi demonstreerimiseks ehituses kasutatakse praktilisi näiteid - alates San Francisco ehituslubade andmekogumi analüüsist kuni CAD-töötlemiseni - miljonite elementidega projektid. Erilist tähelepanu pööratakse masinõppe meetoditele ehitusprojektide maksumuse ja ajakava prognoosimiseks, kusjuures üksikasjalikult käsitletakse lineaarset regressiooni ja k-lähedaseima naabri algoritme. Näidatakse, kuidas struktureeritud andmed saavad aluseks riskide hindamiseks, ressursside optimeerimiseks ja projekti juhtimise tõhususe suurendamiseks mõeldud prognoosimudelitele. Osas antakse ka soovitusi, kuidas valida representatiivseid andmeväljaandeid, ja selgitatakse, miks tõhusate analüüside tegemiseks ei ole alati vaja suuri andmekogumeid.

PEATÜKK 9.1.

SUURED ANDMED JA NENDE ANALÜÜS

Suurandmed ehituses: intuitsioonist prognoositavuseni

Mõiste "suured andmed" ei ole rangelt määratletud. Mõiste tekkis algselt siis, kui teabe maht hakkas ületama selle töötlemise traditsiooniliste meetodite võimalusi. Tänapäeval on andmete maht ja keerukus paljudes tööstusharudes, sealhulgas ehituses, kasvanud nii palju, et need ei mahu arvutite kohalikku mälli ja nende töötlemiseks on vaja kasutada uusi tehnoloogiaid.

Suurte andmetega töötamise põhiolemus ei ole mitte ainult salvestamine ja töötlemine, vaid ka prognoosimisvõime. Ehitustööstuses avab Big Data tee intuiitivsetelt otsustelt, mis põhinevad tabelite ja visualiseeringute subjektiivsel tõlgendamisel (nagu eespool käsitletud), teadlike prognoosideni, mida toetavad reaalsed vaatlused ja statistika.

Vastupidiselt levinud arvamusele ei ole suurandmete töötamise eesmärk "panna masin mõtlema nagu inimene", vaid rakendada matemaatilisi mudeleid ja algoritme tohutute andmehulkade analüüsimiseks, et tuvastada mustreid, ennustada sündmusi ja optimeerida protsesse.

Suurandmed ei ole algoritmide külm maailm, kus puudub inimese mõju. Vastupidi, suurandmed töötavad koos meie instinktide, vigade ja loovusega. Just inimliku mõtlemise ebatäiuslikkus võimaldab meil leida ebestandardseid lahendusi ja teha läbimurdeid.

Digitaaltehnoloogia arenguga on ehitussektor hakanud aktiivselt kasutama IT-sektorist pärit andmetöötlustehnikaid. Tänu sellistele vahenditele nagu Pandas ja Apache Parquet saab struktureeritud ja struktureerimata andmeid kombineerida, lihtsustades juurdepääsu teabele ja vähendades analüüsikadu, samas kui dokumentide või CAD-projektide suured andmekogumid (joonis 9.2-10 - joonis 9.2-12) võimaldavad andmeid koguda, analüüsida ja prognoosida projekti elutsükli kõikides etappides.

Suurtel andmetel on ehitustööstust muutev mõju, mis mõjutab seda potentsiaalselt mitmel viisil. Suurandmete tehnoloogiate rakendamine annab tulemusi mitmes võtmevaldkonnas, sealhulgas näiteks järgmistes:

- **Investeeringipotentsiaali analüüs** - projektide kasumlikkuse ja tasuvusperioodide prognoosimine varasemate rajatiste andmete põhjal.
- **Ennetav hooldus** - tõenäoliste seadmete rikete tuvastamine enne nende tegelikku ilmnemist, mis vähendab seisakuid.
- **Tarneahela optimeerimine** - häirete prognoosimine ja logistika tõhususe parandamine.
- **Energiatõhususe analüüsid** - abi madala energiatarbega hoonete projekteerimisel.
- **Ohutusjärelvalve** - andurite ja kantavate seadmete kasutamine ehitusplatsi tingimuste jälgimiseks.

- **Kvaliteedikontroll** - protsessistandardite täitmise jälgimine reaalajas.
- **Tööjõu juhtimine** - tööjõuvajaduse analüüs ja personalivajaduse prognoosimine.

Ehituses on raske leida valdkonda, kus andmeanalüütika ja prognoosid ei ole nõutud. Ennustusalgoritmide peamine eelis on nende võime ise õppida ja andmete kogunemisel pidevalt täiustuda.

Lähitulevikus ei hakka tehisintellekt mitte ainult ehitajatele abiks olema, vaid hakkab tegema olulisi otsuseid - alates projekteerimisprotsessidest kuni hoonete käitamisega seotud küsimusteni.

Rohkem sellest, kuidas prognoose genereeritakse ja õppemudeleid kasutatakse, räägitakse raamatu järgmises osas "Masinõpe ja prognoosid".

Üleminek täiemahulisele tööle suurandmetega nõuab muutust analüütika käsitluses. Kui seni käsitletud klassikalised süsteemid keskendusid põhjus-tagajärg seostele, siis suurandmete analüüs keskendub nüüd statistiliste mustrite ja korrelatsioonide otsimisele, mis võimaldab tuvastada varjatud seoseid ja ennustada objektide käitumist isegi ilma kõigi tegurite täieliku mõistmiseta.

Suurandmete teostatavuse kahtluse alla seadmine: korrelatsioon, statistika ja andmete valimi moodustamine

Traditsiooniliselt põhines konstruktsioon subjektiivsetel hüpoteesidel ja isiklikel kogemustel. Insenerid eeldasid - teatud tõenäosusega -, kuidas materjal käitub, millistele koormustele konstruktsioon vastu peab ja kui kaua projekt kestab. Neid oletusi testiti praktikas, sageli aja, ressursside ja tulevaste riskide hinnaga.

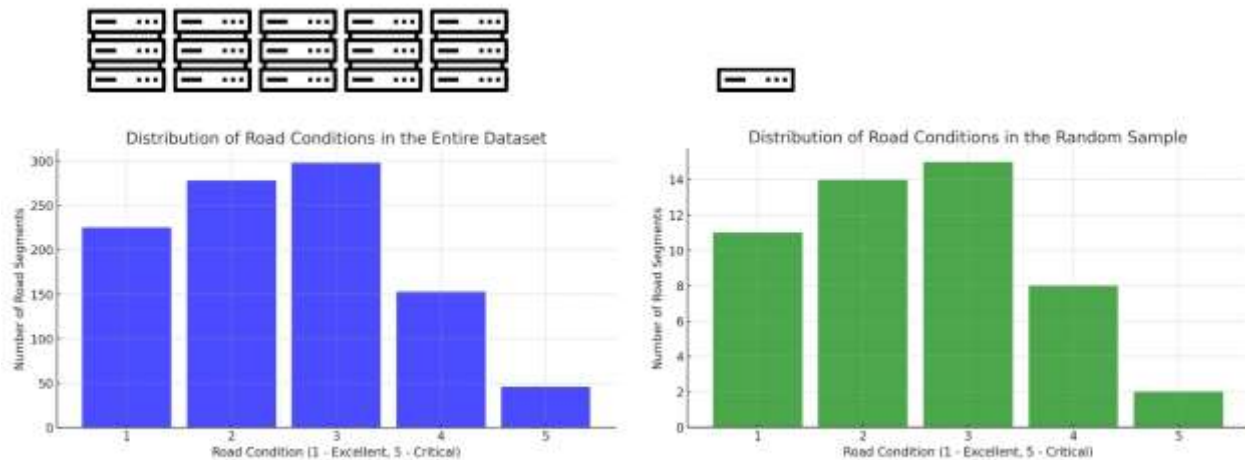
Suurte andmete tulekuga on lähenemine dramaatiliselt muutumas: otsuseid ei tehta enam intuiitivsete aimduste põhjal, vaid suurte andmekogumite analüüsi tulemusena. Ehitus on järk-järgult muutumas intuitsiooni kunstiks ja muutumas täpseks prognoosimise teaduseks.

Üleminek suurandmete kasutamise ideele tõstatab paratamatult olulise küsimuse: kui kriitiline on andmete hulk ja kui palju teavet on tegelikult vaja usaldusväärse prognoosiva analüüsi jaoks? Levinud uskumus, et "mida rohkem andmeid, seda suurem täpsus", ei osutu praktikas alati statistiliselt kehtivaks.

Juba 1934. aastal tõestas statistik Jerzy Neumann [145], et statistiliste järelduste täpsuse võti ei seisne mitte niivõrd andmete koguses, kuivõrd nende representatiivsuses ja valimi juhuslikkuses.

See kehtib eriti ehitustööstuses, kus kogutakse suuri andmemasse asjade interneti -andurite, skannerite, valvekaamerate, droonide ja isegi mitmeformaadiliste CAD -mudelite abil, mis suurendab pimedate kohtade, kõrvalekallete ja andmemoonutuste ohtu.

Võtame näiteks teekatte seisundi jälgimise. Kõikide teelõikude täielik andmekogum võib võtta X GB ja selle töötlemiseks kulub umbes üks päev. Samal ajal võtaks juhuslik valim, mis hõlmaks ainult iga 50. teelõiku, vaid X/50 GB ja seda saaks töödelda poole tunniga, andes samas teatud arvutuste puhul sarnase täpsusega hinnanguid (joonis 9.1-1).



Joonis 9.1-1 Teekatte seisundi histogrammid: kogu andmestik ja juhuslik valim näitavad identseid tulemusi.

Seega ei pruugi eduka andmeanalüüsi võtmeks sageli olla mitte andmete hulk, vaid valimi representatiivsus ja kasutatud töötlemismeetodite kvaliteet. Üleminek juhusliku valimi ja valikulisema lähenemise suunas nõuab ehitussektoris mõtteviisi muutust. Ajalooliselt on ettevõtted järginud loogikat "mida rohkem andmeid, seda parem", uskudes, et kõigi võimalike näitajate hõlmamine maksimeerib täpsust.

Selline lähenemine meenutab projektijuhtimise populaarset väärarusaama: "Mida rohkem spetsialiste ma kaasan, seda tõhusam on töö". Kuid nagu ka inimressursside puhul, on kvaliteet ja vahendid olulisemad kui kvantiteet. Ilma andmete või projektis osalejate vahelisi seoseid (korrelatsioone) arvestamata võib mahu suurendamine viia ainult müra, moonutuste, dubleerimise ja tarbetu raiskamiseni.

Lõppkokkuvõttes selgub sageli, et palju produktiivsem on väiksem, kuid kvalitatiivselt ettevalmistatud andmekogum, mis suudab toota stabiilseid ja mõistlikke prognoose, kui tugineda massilisele, kuid kaootilisele teabele, mis sisaldab palju vastuolulisi signaale.

Liigne andmemahut ei taga mitte ainult suuremat täpsust, vaid võib ka põhjustada moonutatud järeldusi müra, üleliigsete tunnuste, varjatud korrelatsioonide ja ebaolulise teabe tõttu. Sellises olukorras suureneb mudelite liigse sobitamise oht ja analüütiliste tulemuste usaldusväärsus väheneb.

Ehituses on suurandmete käsitlemisel suur väljakutse andmete optimaalse koguse ja kvaliteedi kindlaksmääramine. Näiteks betoonkonstruktsioonide seisundi jälgimisel võib tuhandete andurite kasutamine ja teabe kogumine iga minutiga koormata salvestus- ja analüüsisüsteemi. Kui aga teha

korrelatsioonianalüüsi ja valida välja 10% kõige informatiivsemaid andureid, saab peaaegu sama täpsusega prognoose, kulutades mitu korda, mõnikord kümneid ja sadu kordi vähem ressursse.

Väiksema osaandmete koguse kasutamine vähendab nii vajaliku salvestusruumi kui ka töötlemisaja mahtu, mis vähendab oluliselt andmete säilitamise ja analüüsimise kulusid ning muudab juhusliku valimi sageli ideaalseks lahenduseks prognoosiva analüüsi jaoks, eriti suurte infrastruktuuriprojektide või reaalarajas töötamise puhul. Lõppkokkuvõttes ei määra ehitusprotsesside tõhusust mitte kogutud andmete hulk, vaid nende analüüsi kvaliteet. Ilma kriitilise lähenemise ja hoolika analüüsita võivad andmed viia valede järeldusteni.

Pärast teatud andmehulga ületamist annab iga uus teabeühik üha vähem kasulikke tulemusi. Lõputu teabe kogumise asemel on oluline keskenduda selle representatiivsusele ja analüüsimeetoditele (joonis 9.2-2).

Seda nähtust on hästi kirjeldanud Allen Wallis [146], kes illustreerib statistiliste meetodite kasutamist kahe alternatiivse USA mereväe mürskude konstruktsiooni testimise näitel.

Merevägi katsetas kahte alternatiivset mürsukujundust (A ja B), viies läbi paarilise laskmise seeria. Igas voorus saab A 1 või 0, sõltuvalt sellest, kas selle tulemuslikkus on parem või halvem kui B-l, ja vastupidi. Standardne statistiline lähenemisviis hõlmab kindla arvu katsete läbiviimist (nt 1000) ja võitja määramist protsentuaalse jaotuse alusel (nt kui A saab 1 rohkem kui 53% juhtudest, peetakse seda parimaks). Kui Allen Wallis arutas sellist probleemi (mereväe) kapten Garrett L. Schuyleriga, oli kapten vastu, et selline katse võib Alleni jutu järgi olla kasutu. Kui selline tark ja kogenud laskurohvitser nagu Schuyler oleks olnud kohapeal, oleks ta pärast esimest paarsada [lasku] näinud, et katset ei ole vaja lõpetada, kas seetõttu, et uus meetod on selgelt halvem või seetõttu, et see on selgelt parem, kui loodeti [146].

- USA valitsuse statistiline uurimisrühm Columbia Ülikoolis, II maailmasõja periood

Seda põhimõtet kasutatakse laialdaselt erinevates tööstusharudes. Näiteks meditsiinis viiakse uute ravimite kliinilisi uuringuid läbi patsientide juhuvalimite peal, mis võimaldab saada statistiliselt olulisi tulemusi, ilma et oleks vaja katsetada ravimit kogu maailmas elava elanikkonna peal. Majanduses ja sotsioloogias viiakse läbi representatiivseid uuringuid, et kajastada ühiskonna arvamust, ilma et oleks vaja küsitleda kõiki elanikke.

Nii nagu valitsused ja teadusasutused küsitlevad väikeseid elanikkonnarühmi, et mõista üldisi ühiskondlikke suundumusi, saavad ehitussektori ettevõtted kasutada juhuslikke andmeproove, et tõhusalt jälgida ja koostada prognoose projekti juhtimiseks (joonis 9.1-1).

Suurandmed võivad muuta sotsiaalteaduste käsitlust, kuid ei asenda statistilist tervet mõistust [147].

- Thomas Landsall-Welfair, "Riigi praeguse meeleolu prognoosimine", Significance v. 9(4), 2012 .

Ressursisäästu seisukohast on tulevaste prognooside ja otsuste tegemiseks andmete kogumisel oluline vastata küsimusele: kas on mõttekas kulutada märkimisväärsed ressursse suurte andmekogumite kogumiseks ja töötlemiseks, kui kasutada palju väiksemaid ja odavamaid katseandmeid, mida saab järk-järgult suurendada? Juhusliku valimi tõhusus näitab, et ettevõtted saavad vähendada kümneid või isegi tuhandeid kordi kulusid mudelite kogumiseks ja treenimiseks, valides andmekogumismeetodid, mis ei nõua ulatuslikku katvust, kuid tagavad siiski piisava täpsuse ja representatiivsuse. Selline lähenemisviis võimaldab isegi väikestel ettevõtetel saavutada suurettevõtete võrdväärseid tulemusi, kasutades oluliselt vähem ressursse ja andmemahutu, mis on oluline ettevõtetele, kes soovivad kulusid optimeerida ja kiirendada teadlike otsuste tegemist väikeste ressursside abil. Järgmistes peatükkides uurige näiteid analüütika ja prognoosiva analüüsi kohta, mis põhinevad avalikel andmekogumitel, kasutades suurandmete tööriistu.

Suured andmed: San Francisco miljonite ehituslubade andmekogumi andmete analüüsimine

Töö avatud andmekogumitega annab ainulaadse võimaluse rakendada praktikas eelmistes peatükkides käsitletud põhimõtteid: mõistlik funktsioonide valik, representatiivne valim, visualiseerimine ja kriitiline analüüs. Selles peatükis uurime, kuidas saab avatud andmete abil uurida selliseid keerulisi nähtusi nagu ehitustegevus suures linnas - eelkõige üle miljoni ehitusloa kirje San Franciscos.

Avalikult kättesaadavad andmed enam kui miljoni ehitusloa kohta (joonis 9.1-2) (kirjed kahes CSV-formaadis andmekogumis) "San Francisco ehitusosakonnast". [148] võimaldavad meil kasutada toorest CSV -tabelit mitte ainult linna ehitustegevuse analüüsimiseks, vaid ka San Francisco ehitussektori hiljutiste suundumuste ja ajaloo kriitiliseks analüüsimiseks viimase 40 aasta jooksul, alates 1980. aastast kuni 2019. aastani.

Andmekogumi visualiseerimiseks kasutatud koodinäited (joonised 9.1-3- joonised 9.1-8), samuti visuaalsed graafikud koos koodi, selgituste ja kommentaaridega on leitavad Kaggle'i platvormil, otsides "San Francisco. Construction Sector 1980-2019". [149].

count 1.137695e+06

Public

Building Permits on or after
January 1, 2013

June 13, 2020

2,237 Views

Public

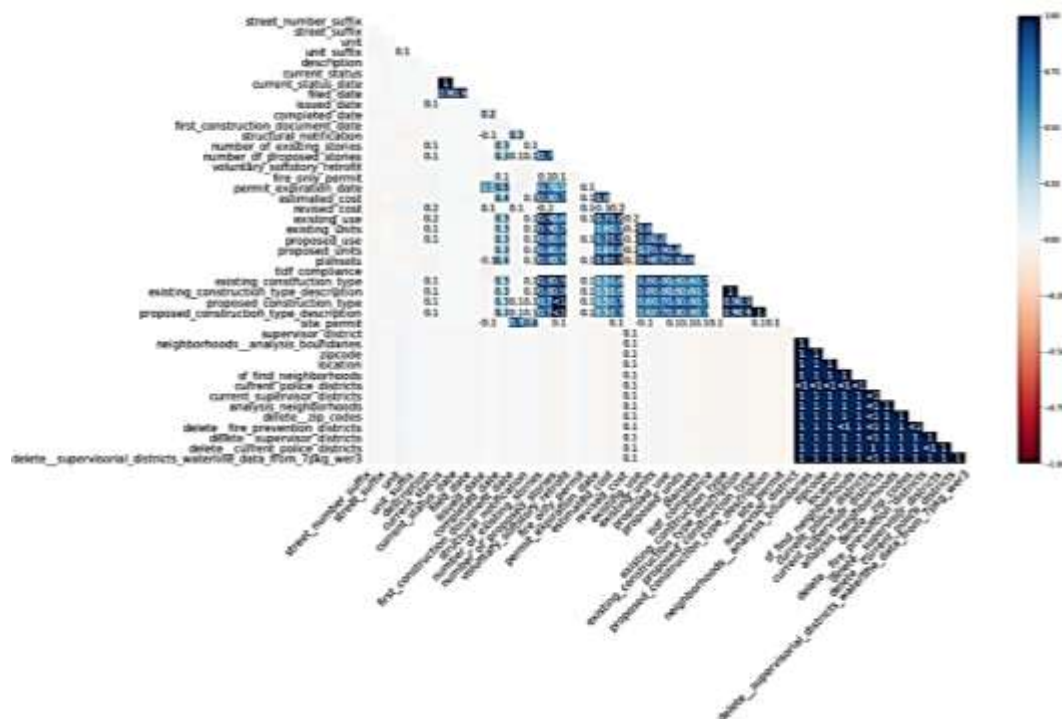
Building Permits before
January 1, 2013

June 13, 2020

860 Views

permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date
07/01/1998	repair stucco	complete	07/07/1998	07/01/1998	07/01/1998	07/07/1998
12/13/2004	reroofing	expired	01/24/2006	12/13/2004	12/13/2004	NaN
02/18/1992	install auto fire spks	complete	06/29/1992	02/18/1992	03/18/1992	06/29/1992
permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	94123.0 (37.7963468760498, -122.4322641443574)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0 (37.729258518008388, -122.4644245687462)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0 (37.79506002552974, -122.39593224461805)

Joonis 9.1-2 Andmekogumid sisaldavad teavet väljastatud ehituslubade kohta erinevate objekti atribuutidega.

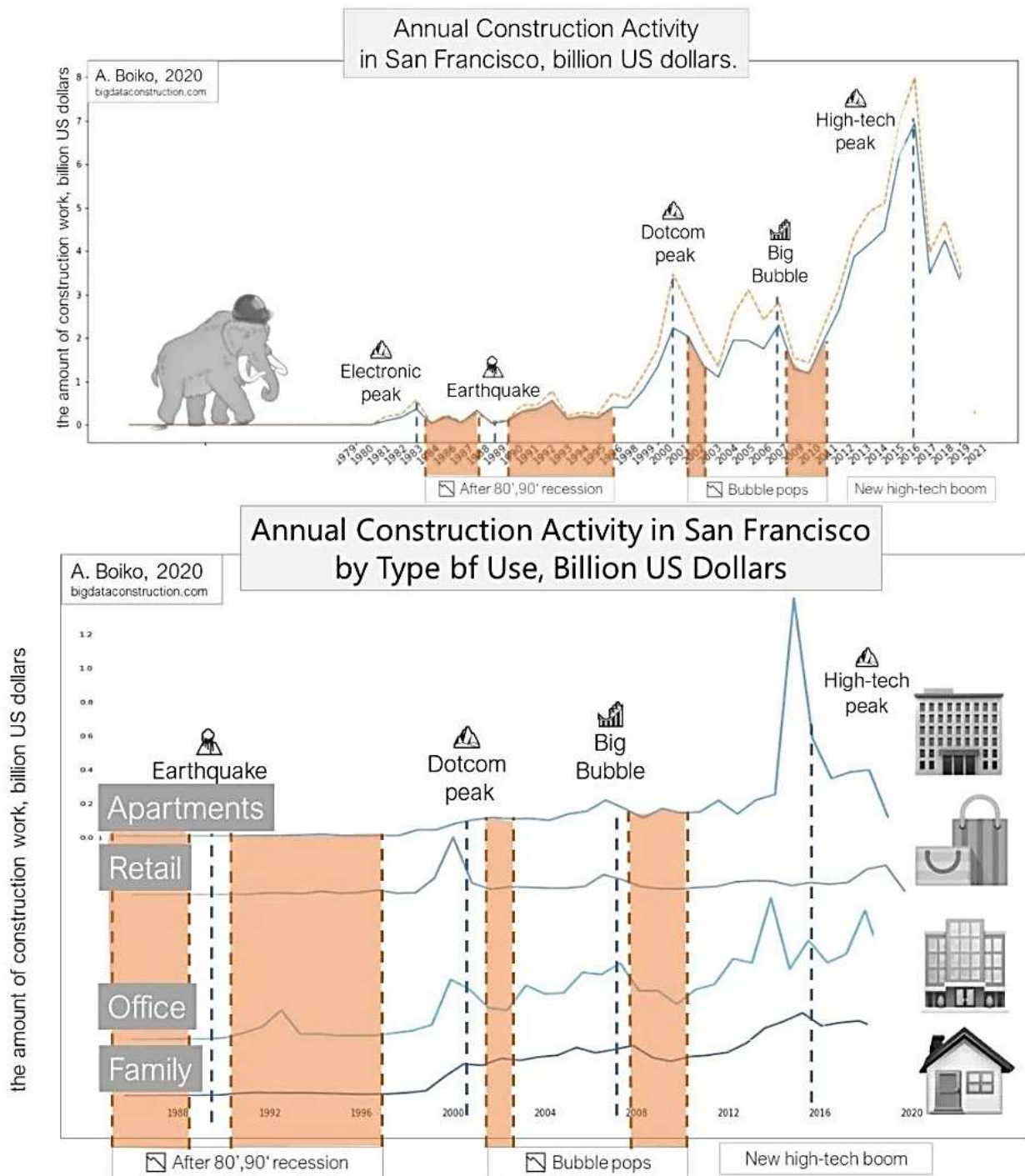


Joonis 9.1-3 Soojuskaart (Pandas ja Seaborn), mis visualiseerib andmekogumi kõiki atribuute ja aitab tuvastada atribuutide paaride vahelisi seoseid.

San Francisco ehitusosakonna esitatud tabelist (joonis 9.1-2) ei ilmne mingeid suundumusi ega järeltusi. Kuivad numbrid tabeli kujul ei ole otsuste tegemise aluseks. Et andmed oleksid visuaalselt arusaadavad, nagu on üksikasjalikult käsitletud andmete visualiseerimist käsitlevates peatükkides, tuleks need visualiseerida, kasutades erinevaid raamatukogusid, mida käsitletakse raamatu seitsmendas osas teemal "ETL ja tulemuste visualiseerimine graafikutena".

Analüüsid andmeid, kasutades Pandas DataFrame ja Pythoni visualiseerimisraamatukogusid, 1 137 695 loa väärtuse kohta [148], saame järeldada, et ehitustegevus San Franciscos on tihedalt seotud majandusüklitega, eriti õitsva Silicon Valley tehnoloogiatööstuse puhul (joonis 9.1-4).

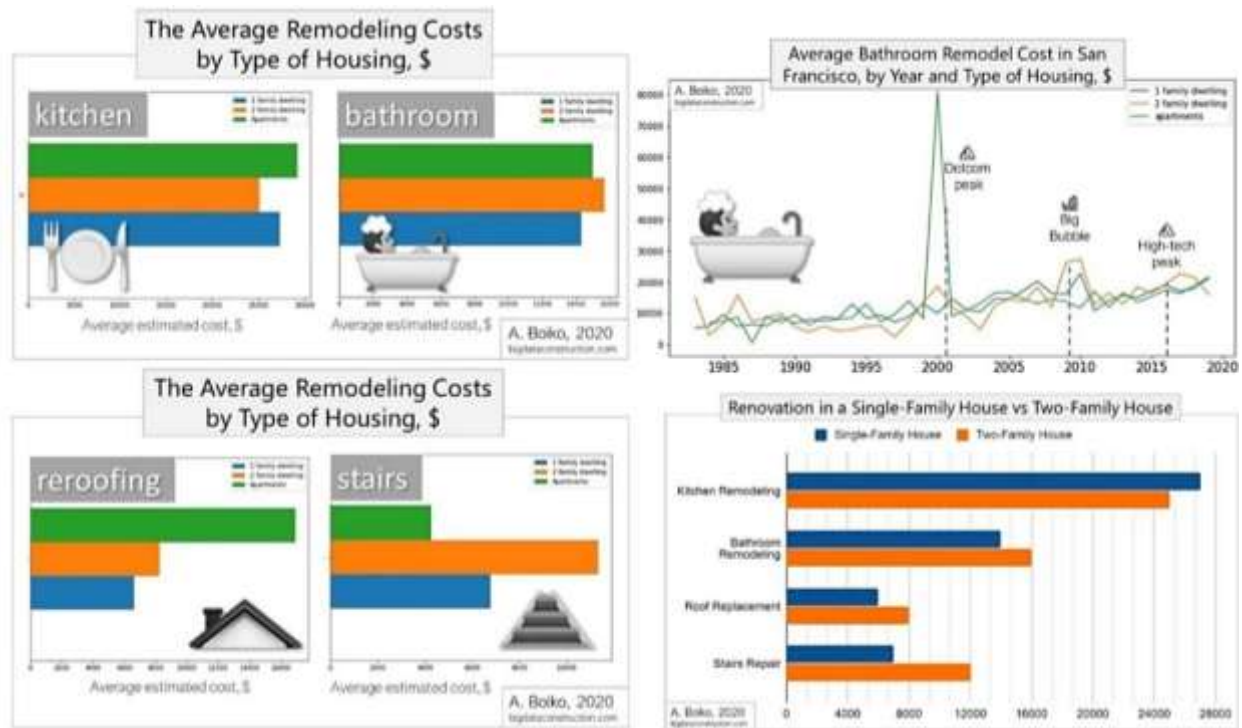
Majandusbuumid ja -langused mõjutavad oluliselt ehitusprojektide arvu ja väärtust. Näiteks langes ehitustegevuse esimene tippaeg kokku elektroonikabuumiga 1980ndate keskel (kasutatud Pandas ja Matplotlib) ning hilisemad tipud ja langused olid seotud dot-com-mulli ja viimaste aastate tehnoloogiabuumiga.



Joonis 9.1-4 San Francisco kinnisvarasektoris on investeeringud korrelatsioonis Silicon Valley tehnoloogilise arenguga.

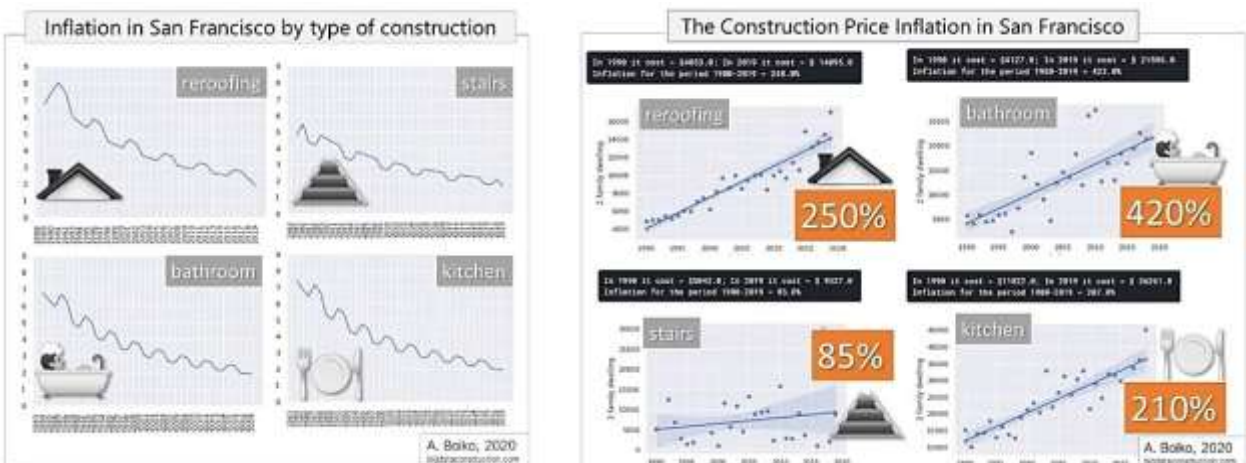
Andmeanalüüs näitab, et San Francisco on suurem osa viimase kümne aasta jooksul ehitusse ja saneerimisse investeeritud 91,5 miljardist dollarist - peaaegu 75% - koondunud kesklinna piirkonda (joonis 9.1-5 - kasutatud Pandas ja Foliumi visualiseerimisraamatukogu) ja kesklinna 2 km raadiuses, mis peegeldab investeeringute suuremat tihedust nendes kesklinna tsoonides.

Andmekogum võimaldab ka keskmiste remondihindade arvutamist mitte ainult majatüüpide, vaid ka linnaosade ja üksikute aadresside (postiindeksite) kaupa. San Franciscos näitab kodude remondikulude dünaamika eri remonditüüpide ja eluruumide puhul erinevaid suundumusi (joonis 9.1-6 - kasutatud Pandas ja Matplotlib). Köögi renoveerimine on märgatavalt kallim kui vannitoa renoveerimine: keskmine köögi renoveerimine ühepereelamu puhul maksab umbes 28 000 dollarit võrreldes 25 000 dollariga kahepereelamu puhul.



Joonis 9.1-6 SF-s maksab köögi renoveerimine peaaegu kaks korda rohkem kui vannitoa renoveerimine ja koduomanikud peavad 15 aasta jooksul iga kuu kõrvale panema \$ 350, et katta kodu kapitaalremondi kulud.

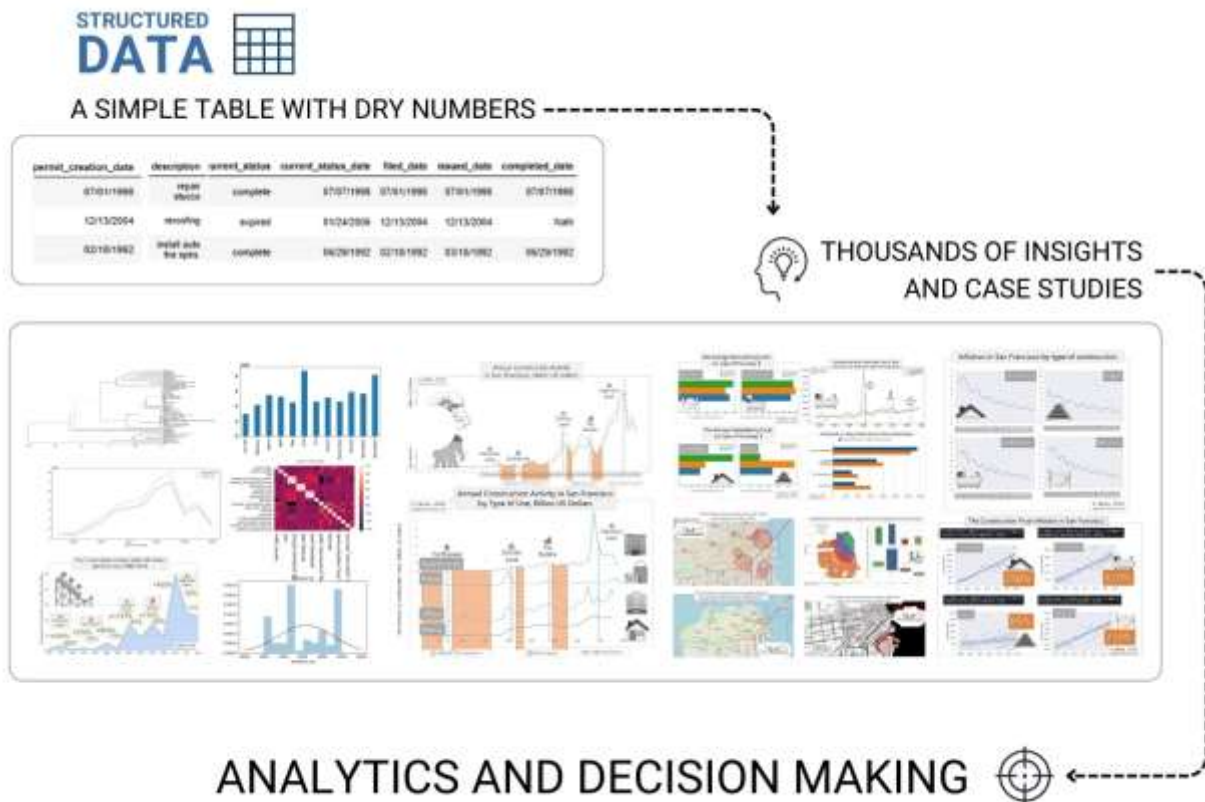
San Francisco ehituskulude inflatsiooni aastate jooksul saab jälgida, analüüsides andmeid elamu tüübi ja aasta järgi rühmitatult (joonis 9.1-7 - kasutatud Pandas ja Seaborn), mis näitab keskmiste remondikulude pidevat tõusu alates 1990. aastast ning näitab lühiajalisi kolmeaastaseid tsükleid mitmepereelamute remondikuludes.



Joonis 9.1-7 Alates 1980. aastast kuni 2019. aastani on vannitoa renoveerimise hind SF-s kasvanud viis korda, samas kui katuse ja köögi renoveerimine on kolmekordistunud ning trepikoja renoveerimine on kasvanud ainult 85%.

San Francisco ehitusosakonna avalike andmete uurimine (joonis 9.1-3) näitab, et linna ehituskulud on äärmiselt varieeruvad ja sageli ettearvamatud, mida mõjutavad mitmesugused tegurid. Nende tegurite hulka kuuluvad majanduskasv, tehnoloogilised uuendused ja eri elamutüüpide ainulaadsed nõuded.

Varem nõudsid sellised analüüsid põhjalikke teadmisi programmeerimisest ja analüütikast. LLM-vahendite kasutuselevõtuuga on see protsess aga muutunud kättesaadavaks ja arusaadavaks paljudele ehitussektori spetsialistidele, alates projekteerimisosakondade inseneridest kuni tippjuhtkonnani.



Joonis 9.1-8 Üleminek visuaalselt arusaadavatele andmetele võimaldab automatiseeritud otsuste langetamist varjatud mustrite äratundmise kaudu.

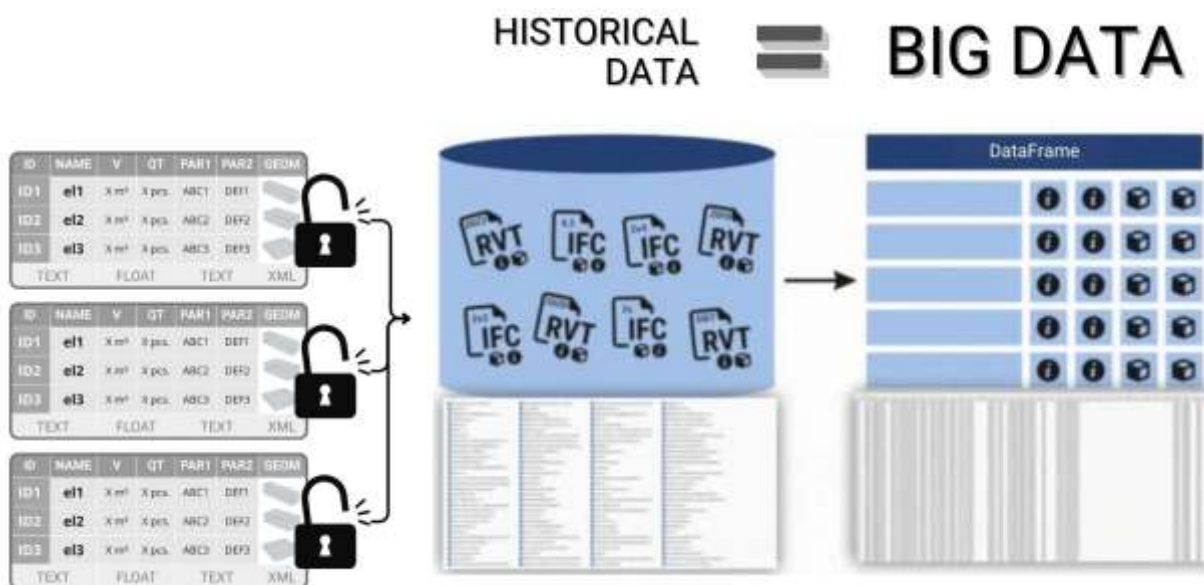
Nii nagu me analüüsisime andmeid "San Francisco ehitusameti" tabeliandmestikust, saame visualiseerida ja analüüsida mis tahes andmestikku - alates piltidest ja dokumentidest kuni asjade interneti andmestikuni või tuletatud CAD-andmebaasidest.

Näide CAD-andmetel põhinevate suurandmete kohta (BIM)

Järgnevas näites analüüsisime suurt andmestikku, kasutades erinevate CAD-vahendite (BIM) andmeid. Suure andmekogumi kogumiseks ja loomiseks kasutati spetsiaalset automatiseeritud veebiotsingumootorit (skripti), mis on konfigureeritud nii, et see otsib ja kogub automaatselt projekteerimisfaile veebisaitidelt, mis pakuvad tasuta arhitektuurimudeleid vormingutes RVT ja IFC. Mõne päeva jooksul leidis ja laadis see programm edukalt alla 4596 IFC-faili ja 6471 RVT-faili ning 156

024 DWG-faili[149].

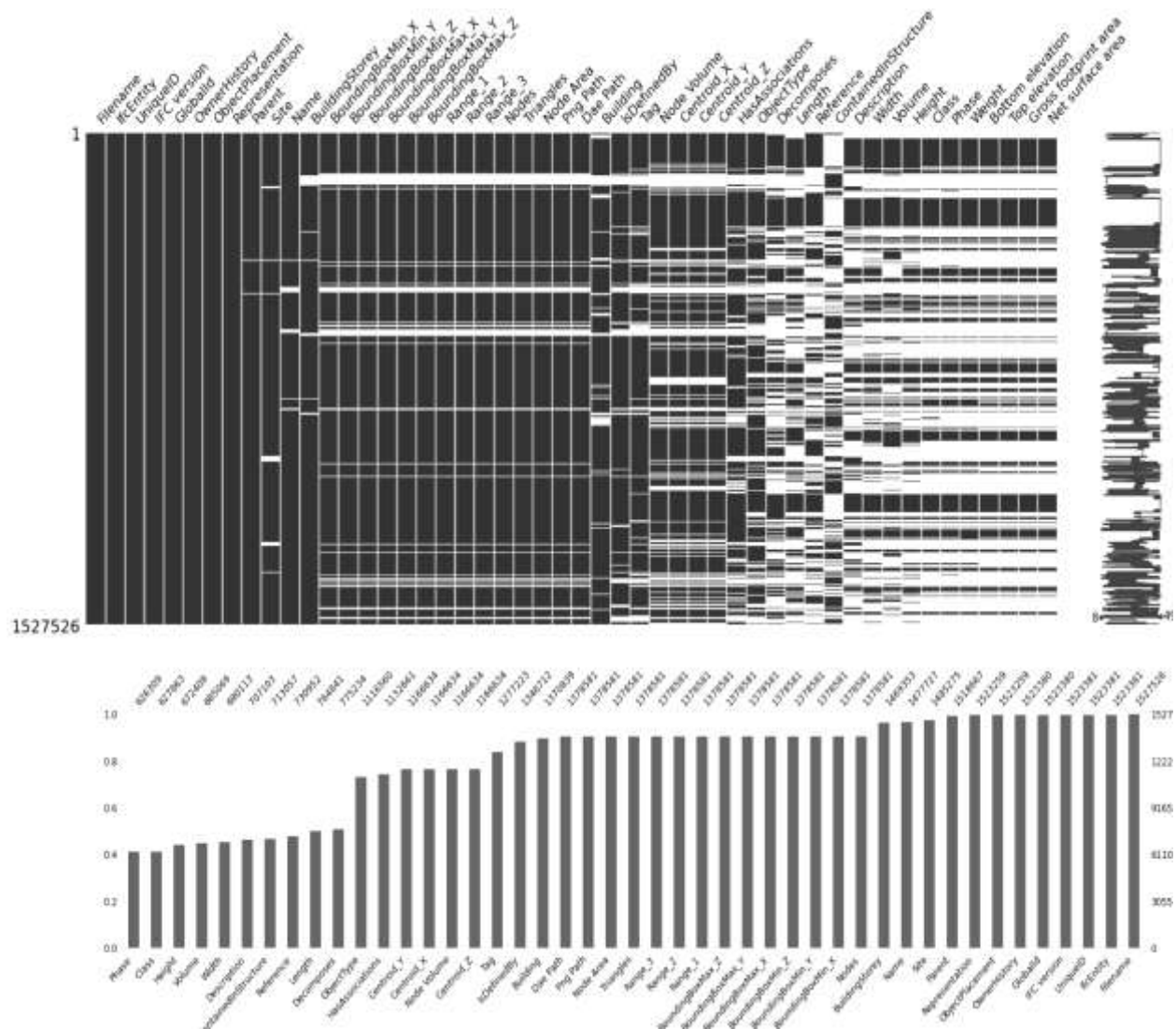
Pärast projektide kogumist RVT ja IFC eri versioonide formaatides ja nende teisendamist struktureeritud CSV-vormingusse kasutades tasuta reverse engineering SDK-d, koguti peaaegu 10 tuhat RVT ja IFC projekti ühte suurde Apache Parquet tabelifaili ja laaditi üles Pandas DataFrame analüüsi jaoks (joonis 9.1-9).



Joonis 9.1-9 Struktureeritud andmed Projektiandmed võimaldavad teil ühendada ükskõik kui palju projekte üheks kahemõõtmeliseks tabeliks.

Selle ulatusliku kogumiku andmed sisaldavad järgmist teavet: IFC-failikogum sisaldab umbes 4 miljonit üksust (rida) ja 24 962 atribuuti (veergu) ning RVT-failikogum, mis koosneb umbes 6 miljonist üksusest (rida), sisaldab 27 025 erinevat atribuuti (veergu).

Need infokogumid (joonis 9.1-10) hõlmavad miljoneid elemente, mille kohta saadi täiendavalt Bounding Box geomeetria koordinaadid (projekti objekti piire määratlev ristkülik) ja lisati ühisele tabelile - Bounding Box geomeetria koordinaadid (projekti objekti piire määratlev ristkülik) ja iga elemendi kujutised PNG-formaadis ning geomeetria avatud XML-formaadis - DAE (Collada).



Joonis 9.1-10 Sabset 1,5 miljonist elemendist ja esimese 100 atribuudi hõivatuse visualiseerimine (missingno library) histogrammina.

Seega saime kogu teabe kümnete miljonite elementide kohta 4596 IFC-projektist ja 6471 RVT-projektist, kus kõikide üksuste elementide kõik atribuudid-omadused ja nende geometria (Bounding Box) tõlgiti struktureeritud kujul ühte tabelisse (DataFrame) (joonis 9.1-10 - andmed dataframe'i populatsioonide kohta ilmuvad histogrammidena).

Analüüsi käigus koostatud histogrammid (joonis 9.1-10, joonis 9.2-6, joonis 9.2-7) võimaldavad kiiresti hinnata andmete tihedust ja väärtuste esinemissagedust veergudel. See annab esimese ülevaate tunnuste jaotusest, kõrvalekallete esinemisest ja üksikute tunnuste potentsiaalsest kasulikkusest masinõppe mudelite analüüsimisel ja koostamisel.

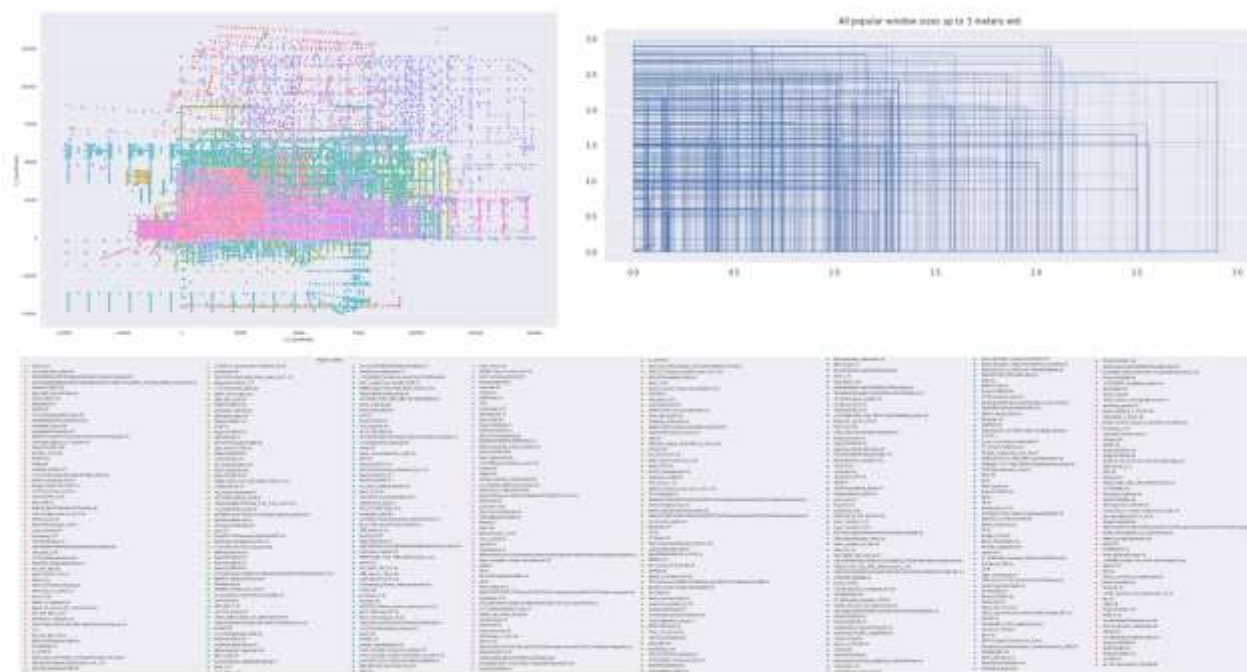
Üks näide selle andmekogumi praktilisest kasutamisest (joonis 9.1-10) on projekt "5000 IFC ja RVT".

[149], mis on kättesaadav Kaggle'i platvormil. Selles esitatakse Jupyter Notebook koos täieliku Pipeline'i lahendusega: alates andmete eeltöötlustest ja analüüsist kuni tulemuste visualiseerimiseni, kasutades Pythoni raamatukogusid - pandas, matplotlib, seaborn, folium jt (joonis 9.1-11).



Joonis 9.1-11 Näited CAD-vormingute (BIM) andmete analüüsist Pythoni visualiseerimisraamatukogude ja pandas-raamatukogu abil.

Metainfo põhjal on võimalik kindlaks teha, millistes linnades on teatud projektid välja töötatud, ja seda kaardil kuvada (nt foliumi raamatukogu abil). Lisaks võimaldavad andmetes olevad ajatemplid uurida mustreid, millal faile salvestati või redigeeriti: nädalapäevade, kellaaegade ja kuude kaupa.



Joonis 9.1-12 Visualiseerimine kõigi sammaste geomeetrilise asukoha ja kõigi kuni 3 meetri pikkuste akende mõõtmete projektid graafiku allosas olevast nimekirjast.

Mudelitest saadud geomeetrilised parameetrid Bounding Boxi kujul sobivad ka koondanalüüsiks. Näiteks joonisel 9.1-12 on esitatud kaks graafikut: vasakpoolne näitab kõikide projektide veergude vahemaade jaotust nullpunkti suhtes ja parempoolne näitab kõigi kuni 3 meetri kõrguste akende mõõtmeid kümnetest tuhandetest aknaelementidest koosnevas valimis (pärast kogu andmestiku rühmitamist parameetri "Category" järgi väärtusega "OST_Windows", "IfcWindows").

Selle näite torujuhtme analüüsi kood ja andmestik ise on saadaval Kaggle'i veebisaidil pealkirja "5000 IFC ja RVT | DataDrivenConstruction.io" all [149]. Seda valmis Pipeline'i koos andmestikuga saab kopeerida ja käivitada tasuta internetis Kaggle'is või offline ühes populaarsetest IDE-dest: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse koos PyDev pluginaga, Thonny, Wing IDE, IntelliJ IDEA koos Python pluginaga, JupyterLab või populaarsed online tööriistad Kaggle'is.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Suurte koguste struktureeritud andmete töötlemisel ja uurimisel saadud analüütilised teadmised mängivad otsustavat rolli ehitussektori otsustusprotsessides.

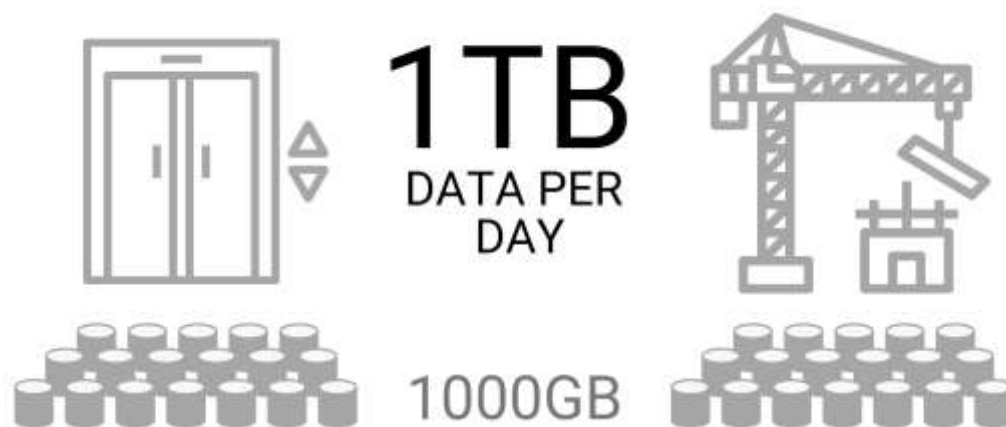
Sellise varasematel projektidel põhineva teabeanalüüsi abil saavad spetsialistid tõhusalt prognoosida näiteks materjali- ja tööjõuvajadust ning optimeerida projekteerimislahendusi enne ehituse algust

Kui projekteerimisandmed või ehitusload on suhteliselt staatiline teave, mis muutub suhteliselt aeglaselt, siis ehitusprotsess ise on kiiresti küllastumas mitmesuguste andurite ja asjade interneti - seadmetega: kaamerad, automaatsed seiresüsteemid, mis edastavad andmeid reaalajas - kõik see muudab ehitusplatsi dünaamiliseks digitaalseks keskkonnaks, kus andmeid tuleb analüüsida reaalajas.

asjade internet Asjade internet ja arukad lepingud

Asjade internet Asjade internet esindab uut digitaalsete muutuste lainet, kus iga seade saab oma IP-aadressi ja muutub osaks ülemaailmsest võrgustikust. Asjade internet on kontseptsioon, mis hõlmab füüsiliste objektide ühendamist interneti, et koguda, töödelda ja edastada andmeid. Ehituses tähendab see võimalust kontrollida ehitusprotsesse reaalajas, minimeerida materjalide raiskamist, prognoosida seadmete kulumist ja automatiseerida otsuste tegemist.

Vastavalt CFMA artiklile "Tuleviku ettevalmistamine ühendatud ehitusega" [150] läbib ehitustööstus järgmise kümnendi jooksul suure digitaalse muutuse, mis kulmineerub mõiste Connected Construction - täielikult integreeritud ja automatiseeritud ehitusplatsiga.



Joonis 9.1-13 IOT- või ehitusplatsi andmesideseadmed võivad toota ja edastada terabaite andmeid päevas.

Digitaalne ehitusplats tähendab, et kõik ehituselemendid - alates planeerimisest ja logistikast kuni tööde teostamise ja kvaliteedikontrolliga ehitusplatsil, kus kasutatakse statsionaarseid kaameraid ja kvadroskoope - integreeritakse ühtseks dünaamiliseks digitaalseks ökosüsteemiks. Varem, käesoleva raamatu 7. osas, vaatlesime juba Apache NiFi (joonis 7.4-5) võimalusi, mis on tasuta ja avatud lähtekoodiga vahend, mis võimaldab andmete reaajas voogedastamist - alates andmete kogumisest erinevatest allikatest kuni edastamiseni salvestus- või analüüsiplatvormidele.

Andmed ehituse edenemise, materjalikulu, seadmete seisundi ja ohutuse kohta edastatakse reaajas analüüsisüsteemidele (joonis 9.1-13). See võimaldab prognoosida võimalikke riske, reageerida kõrvalekalletele kiiresti ja optimeerida ehitusplatsi protsesse. Digitaalse ehitusplatsi põhikomponendid on järgmised:

- asjade Interneti -andurid - keskkonnaparameetrite jälgimine, ehitusseadmete jälgimine ja töötingimuste kontrollimine.
- Digitaalsed kaksikud - hoonete ja infrastruktuuri virtuaalsed mudelid võimalike kõrvalekallete prognoosimiseks ja vigade vältimiseks.
- Automatiseeritud logistikasüsteemid - tarneahela juhtimine reaajas, et vähendada seisakuid ja kulusid.
- Robootilised ehitussüsteemid - autonoomsete masinate kasutamine rutiinsete ja ohtlike ülesannete täitmiseks.

Robotiseerimine Interneti ja digitaalse ehitusplatsi (Connected Site) kontseptsioon ei suurenda mitte ainult tõhusust ja vähendab kulusid, vaid toob kaasa ka uue ohutuse, jätkusuutliku ehituse ja prognoosiva projekti juhtimise ajastu.

RFID (raadiosagedustuvastus) on samuti üks asjade interneti kõige olulisemaid komponente. Neid kasutatakse materjalide, masinate ja isegi töötajate identifitseerimiseks ja jälgimiseks ehitusplatsil, suurendades projekti ressursside läbipaistvust ja kontrolli.

RFID -tehnoloogiat kasutatakse objektide automaatseks äratundmiseks raadiosignaalide abil. See koosneb kolmest põhielemendist:

- RFID - märgised (passiivsed või aktiivsed) - sisaldavad unikaalset identifikaatorit ja kinnitatakse materjalide, tööriistade või masinate külge.
- Skannerid on seadmed, mis loevad teavet siltidelt ja edastavad selle süsteemi.
- Tsentraliseeritud andmebaas - salvestab teavet objektide asukoha, staatuse ja liikumise kohta.

RFID kasutamine ehituses:

- Automaatne materjaliarvestus - sildid valmisbetoontoodetel, armatuuril või sandwich-paneelide pakenditel võimaldavad varude kontrolli ja hoiavad ära vargused.
- Töötajate töö kontroll - RFID - töötajapiletid registreerivad vahetuse algus- ja lõpuaegu, mis võimaldab tööaja arvestust.
- Seadmete jälgimine - RFID - süsteem jälgib seadmete liikumist, vältides seisakuid ja parandades logistika tõhusust.

Seda tehnoloogiakomplekti täiendavad plokiahelapõhised arukad lepingud, mis automatiseerivad makseid, tarnekontrolli ja lepingute täitmist ilma vahendajate vajaduseta, vähendades pettuse ja viivituste riski.

Praegu on arukad lepingud ühise andmemudeli puudumisel lihtsalt kood, milles osalejad kokku lepivad. Andmekeskse lähenemisviisi abil on aga võimalik luua lepingu parameetrite ühine mudel, kodeerida see plokiahelasse ja automatiseerida tingimuste täitmine.

Näiteks tarneahela haldussüsteemis võib arukas leping jälgida saadetise kohaletoimetamist asjade interneti -andurite ja RFID -märgiste abil ning automaatselt makse üle kanda, kui saadetis saabub. Samamoodi võiks arukas leping ehitusplatsil registreerida tööetapi - näiteks armatuuri paigaldamine või vundamendi valamine - lõpetamist droonide või ehitusandurite andmete põhjal ning algsatada automaatselt järgmise makse töövõtjale, ilma et oleks vaja käsitsi kontrollida ja pabertõendeid esitada.

Kuid vaatamata uutele tehnoloogiatele ja rahvusvaheliste standardiorganisatsioonide jõupingutustele muudab asjade interneti maastikku keeruliseks konkureerivate standardite rohkus.

Cisco 2017. aastal avaldatud uuringu [151] kohaselt peatub peaaegu 60% asjade interneti algtustest (IoT) kontseptsiooni tõestamise etapis ja ainult 26% ettevõtetest peab oma asjade interneti projekte täielikult edukaks. Lisaks sellele ei saavuta kolmandik lõpetatud projektidest seatud eesmärgi ning neid ei tunnistata edukaks ka pärast rakendamist.

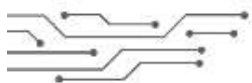
Üks peamisi põhjusi on erinevate andurite andmeid töötlevate platvormide koostalitlusvõime puudumine. Selle tulemusel jäävad andmed eraldi lahendustesse isoleerituks. Alternatiiviks sellisele lähenemisviisile, nagu ka muudel sarnastel juhtudel (mida oleme käesolevas raamatus käsitlenud), on arhitektuur, mis on üles ehitatud andmete endi kui peamise vara ümber.

Asjade interneti andurid mängivad võtmerolli mitte ainult seadmete tehnilise seisundi jälgimisel, vaid ka prognoosivas analüüsis, et vähendada riske ehitusplatsil ja parandada protsessi üldist tulemuslikkust, prognoosides rikkeid ja kõrvalekaldeid.

Asjade interneti andurite ja RFID-siltide kogutud andmeid saab reaajas töödelda masinõppe algoritmide abil, mis suudavad tuvastada kõrvalekaldeid ja hoiatada insenere võimalike rikete eest ette. Need võivad ulatuda mikropraodest betoonkonstruktsioonides kuni ebatüüpiliste pausideni tornkraana töös, mis viitavad tehnilistele riketele või regulatiivsetele rikkumistele. Lisaks sellele võivad täiustatud käitumisanalüüsi algoritmid tabada käitumismustreid, mis võivad viidata näiteks töötajate füüsilisele väsimusele, parandades ohutuse ja töötajate heaolu ennetavat juhtimist kohapeal.

Ehitustööstuses juhtuvad õnnetused ja tõrked - nii masinate kui ka inimestega - harva ootamatult. Neile eelnevad tavaliselt väikesed kõrvalekalded, mis jäävad märkamatuks. Ennustav analüüs ja masinõppe võimaldavad neid signaale varakult avastada, isegi enne kriitiliste tagajärgede tekkimist.

Kui dokumendid, projektitoimikud ning asjade interneti seadmetest ja RFID-siltidest pärinevad andmed moodustavad ehitusprojektide digitaalse jalajälje, siis masinõppe võib aidata sellest kasulikke teadmisi välja võtta. Seoses andmete arvu suurenemise ja andmetele juurdepääsu demokratiseerimisega avanevad ehitussektoris uued võimalused analüütika, prognoosiva analüüsi ja tehisintellekti rakenduste valdkonnas.

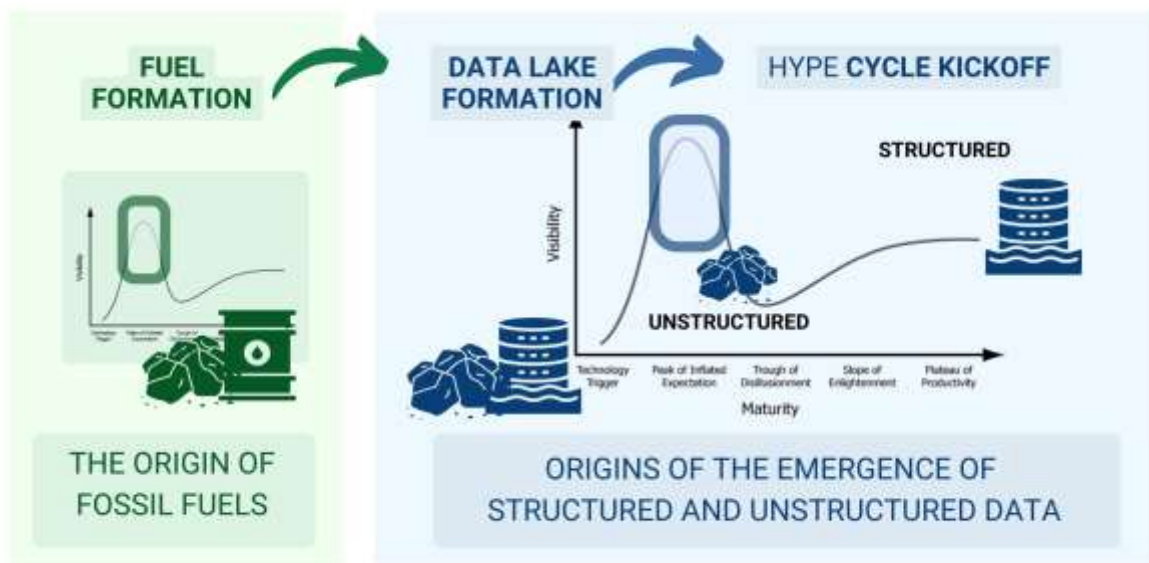


PEATÜKK 9.2.

MASINÕPE JA PROGNOOSID

Masinõpe ja tehisintellekt muudavad seda, kuidas me ehitame

Erinevate süsteemide andmebaasid ehitustegevuses - koos nende paratamatult laguneva ja üha keerulisemaks muutuva infrastruktuuriga - on muutumas tulevaste lahenduste kasvulavaks. Ettevõtete serverid, nagu mets, on rikas biomassitäie olulist teavet, mis on sageli peidetud maa alla, kaustade ja serverite sisemusse. Erinevatest süsteemidest täna tekkivad andmemassid - pärast kasutamist, pärast serveri põhja langemist ja pärast aastatepikkust kivistumist - toidavad tulevikus masinõppe ja keelemudeleid. Ettevõtte sisemised jututubasid (nt eraldi instantsina lokaalselt konfigureeritud ChatGPT, LLaMa, Mistral, DeepSeek) hakatakse ehitama nende ettevõttesiseste mudelite põhjal, kasutades tsentraliseeritud mälu, et kiiresti ja mugavalt teavet välja otsida ning luua vajalikke graafikuid, armatuurilaudu ja dokumente.



Joonis 9.2-1 Nii nagu puud muutuvad aja ja analüütika surve all söeks, muutub ka teave aja jooksul väärtuslikuks ettevõtlusenergiaks.

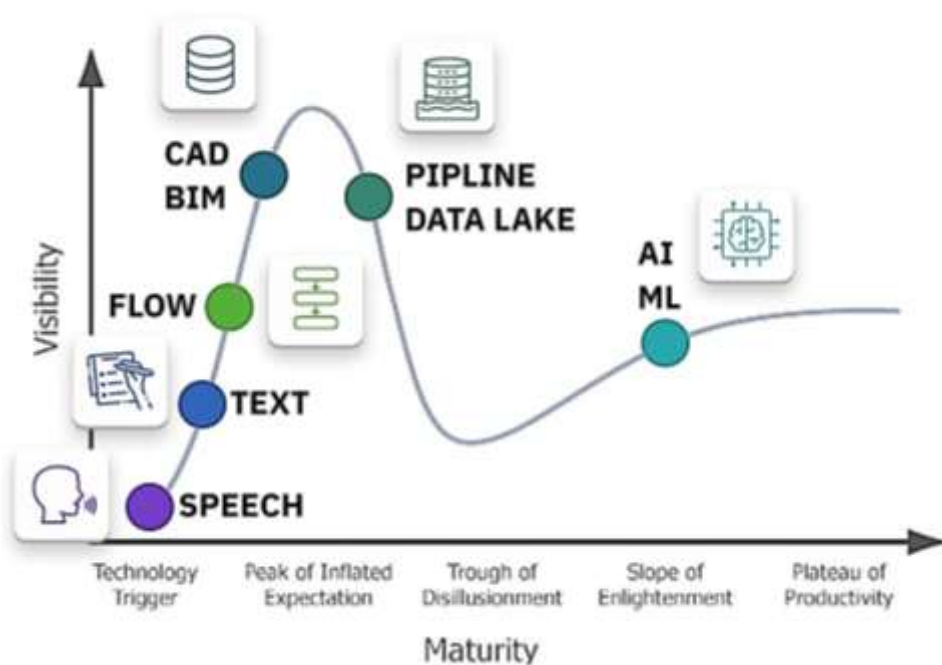
Taimede massi kivistumine koos rõhu ja temperatuuriga tekitab eri aegadel elanud eri liiki puude homogeense ja ainulaadse struktuuriga homogeense massi - söe [152]. Samamoodi moodustab eri formaadis ja eri aegadel kõvakettale salvestatud teave analüütikaosakondade surve ja kvaliteedijuhtimise temperatuuri all lõpuks homogeense struktureeritud väärtusliku teabe massi (joonis 9.2-1).

Need teabekihid (või sagedamini üksikud nuggiti) luuakse kogenud analüütikute vaearikka andmete organiseerimise teel, kes hakkavad järk-järgult eraldama väärtuslikku teavet näiliselt pikka aega ebaolulistest andmetest.

Hetkel, mil need küpsed andmekihid ei ole enam lihtsalt aruannetes "põlenud", vaid hakkavad ringluses

olema äriprotsessides, rikastades otsuseid ja parandades protsesse, on ettevõtte valmis järgmiseks sammuks - üleminekuks masinõppele ja tehisintellektile (joonis 9.2-2).

Masinõpe (ML - Machine learning) on meetodite klass tehisintellekti probleemide lahendamiseks. Masinõppe algoritmid tuvastavad suurtes andmekogumites mustreid ja kasutavad neid ise õppimiseks. Iga uus andmekogum võimaldab matemaatilistel algoritmidel paraneda ja kohaneda vastavalt saadud teabele, mis võimaldab pidevalt parandada soovitude ja prognooside täpsust.



Joonis 9.2-2 Andmete loomise tehnoloogiate hääbumine ja analüütiliste vahendite rakendamine avab ukse masinõppe teemale.

Nagu ütles maailma suurima investeerimisfondi mõjukas tegevjuht (kes omab põhiosalusi peaaegu kõigis suurimates ehitustarkvaraettevõtetes, samuti ettevõtetes, mis omavad suurimat hulka kinnisvara maailmas [55]) 2023. aasta intervjuus - masinõpe muudab ehitusmaailma.

Tehisintellekti on tohtu potentsiaal. See muudab meie töö- ja eluviise. Tehisintellekt ja robotika muudavad seda, kuidas me töötame ja kuidas me ehitame, ning me saame kasutada tehisintellekti ja robotikat vahendina palju suurema tootlikkuse loomiseks [153].

- Maailma suurima investeerimisfondi tegevjuht, intervjuu, september 2023.

Masinõpe (ML) töötab suurte andmehulkade töötlemisel, kasutades statistilisi meetodeid, et jäljendada inimese mõtlemise aspekte. Enamikul ettevõtetel ei ole aga selliseid andmekogumeid, ja

kui ongi, siis ei ole need sageli piisavalt märgistatud. Siinkohal võivad abiks olla semantilised tehnoloogiad ja ülekandeõpe, mis võimaldab ML-l olla tõhusam väikeste andmemahtude töötlemisel, mille teostatavust on käsitletud käesoleva osa eelmistes peatükkides.

Siirdeõppe põhiõlemus seisneb selles, et selle asemel, et õppida iga ülesannet nullist, saab kasutada seotud valdkondades omandatud teadmisi. On vaja mõista, et teistest tööstusharudest pärit mustreid ja avastusi saab kohandada ja rakendada ehitustööstuses. Näiteks aitavad jaekaubanduses välja töötatud logistiliste protsesside optimeerimise meetodid parandada ehitusalase tarneahela juhtimise tõhusust. Suurandmete analüüsi, mida kasutatakse aktiivselt finantsvaldkonnas, saab rakendada ehitusprojektide kulude prognoosimisel ja riskijuhtimisel. Tööstuses arendatavad arvutinägemise ja robotika tehnoloogiad leiavad juba praegu rakendusi automatiseeritud kvaliteedikontrolli, ohutuse jälgimise ja ehitusplatsi rajatiste haldamise valdkonnas.

Siirdeõpe võimaldab mitte ainult kiirendada uuenduste kasutuselevõttu, vaid ka vähendada nende väljatöötamise kulusid, kasutades teiste tööstusharude juba kogutud kogemusi.

$$\text{labor productivity in construction} = f(\text{AI})$$

Joonis 9.2-3 Tehisintellekt tehnoloogiad ja robotika on tulevikus peamine liikumapanev jõud, et suurendada tootlikkust ehitustööstuses.

Inimese mõtlemine on korraldatud sarnasel põhimõttel: me toetume varem omandatud teadmistele, et lahendada uusi probleeme (joonis 4.4-19, joonis 4.4-20, joonis 4.4-21). Ka masinõppe puhul toimib see lähenemine - lihtsustades andmemudelit ja muutes selle elegantsemaks, saame vähendada probleemi keerukust ML-algoritmide jaoks. See omakorda vähendab vajadust suurte andmehulkade järele ja vähendab arvutuskulusid.

Subjektiiivsest hinnangust statistilise prognoosini

Ajastu, mil strateegilised otsused sõltusid üksikute juhtide intuitsioonist (joonis 9.2-4), kuulub minevikku. Üha tihedama konkurentsi ja väljakutsetega majanduskeskkonnas muutub subjektiivne lähenemine liiga riskantseks ja ebatõhusaks. Ettevõtted, mis jätkuvalt tuginevad isiklikele arvamustele, selle asemel et objektiivselt analüüsida andmeid, kaotavad võime kiiresti muutustele reageerida.

Konkurentsikeskkond nõuab täpsust ja korratavust, mis põhineb andmetel, statistilistel mustritel ja arvutataval tõenäosusel. Otsused ei saa enam põhineda tunnetusel, vaid need peavad põhinema korrelatsioonidel, suundumustel ja analüütika ja masinõppe abil saadud prognoosimudelitel. See ei tähenda mitte ainult vahendite muutmist, vaid ka mõtlemise loogika muutmist: eeldustelt tõenditele, subjektiivsetelt tõenäosustelt statistiliselt arvutatud kõrvalekalletele, tunnetelt faktidele.



Joonis 9.2-4 HiPPO (kõrgeima palgaga töötaja arvamus) tehtud otsuste ajastu jääb suurte andmete ja masinõppe tulekuga minevikku.

Juhid, kes varem tuginesid ainult oma tunnetele, seisavad paratamatult silmitsi uue reaalsusega: autoriteet ei määra enam valikuid. Juhtimise keskmes on nüüd süsteemid, mis analüüsivad miljoneid parameetreid ja vektoreid, tuvastavad varjatud mustreid ja pakuvad optimaalseid strateegiaid.

Peamine põhjus, miks ettevõtted tänapäeval ikka veel väldivad ML-i rakendamist, on selle vähene läbipaistvus. Enamik mudeleid töötab juhtide jaoks "mustade kastidena", selgitamata, kuidas täpselt nad oma järeldusteni jõuavad. See toob kaasa probleeme: algoritmid võivad tugevdada stereotüüpe ja luua isegi humoorikaid olukordi, nagu Microsofti juturobot, mis muutus kiiresti mürgiseks suhtlusvahendiks [154].

Teoses Deep Thinking mõtiskleb endine malemaailma meister Garri Kasparov oma kaotuse üle IBMi Big Blue arvutile [155]. Ta väidab, et tehisintellekti tõeline väärtus ei seisne mitte inimese intelligentsuse kopeerimises, vaid meie võimete täiendamises. Tehisintellekt peaks täitma ülesandeid, milles inimene on nõrk, samal ajal kui inimene toob loovust. Arvutid on muutnud traditsioonilist lähenemist maleanalüüsile. Selle asemel, et luua põnevaid lugusid partiide kohta, hindavad arvutimänguprogrammid iga käiku erapooletult, lähtudes ainult selle tegelikust tugevusest või nõrkusest. Kasparov märgib, et inimese kalduvus näha sündmusi pigem sidusate lugudena kui üksikute käikudena viib sageli valede järeldusteni - mitte ainult males, vaid ka elus üldiselt.

Seega, kui plaanite kasutada masinõpet prognoosimiseks ja analüüsiks, on oluline mõista selle põhiprintsiipe - kuidas algoritmid töötavad ja kuidas andmeid töödeldakse - enne, kui hakkate oma töös kasutama masinõppe vahendeid ja AI. Parim viis alustamiseks on praktiline kogemus.

Üks kõige mugavamaid vahendeid masinõppe teemaga alustamiseks ja prognoosimiseks on Jupyter Notebook ja populaarne klassikaline Titanic andmestik, mis annab visuaalse sissejuhatuse andmete analüüsimise ja ML-mudelite koostamise põhimeetoditesse.

Titanic data set: Hello World analüütika maailmas andmed ja suurandmed

Üks kuulsamaid näiteid ML kasutamise kohta andmeanalüütikas on Titanicu andmestiku analüüs, mida kasutatakse sageli reisijate ellujäämise tõenäosuse uurimiseks. Selle tabeli õppimine on analoogne "Hello World" programmiga programmeerimiskeelte õppimisel.

RMS Titanic'i uppumise tagajärjel hukkus 1912. aastal 1502 inimest 2224-st. Titanicu andmekogum sisaldab mitte ainult teavet selle kohta, kas reisija jäi ellu, vaid ka selliseid tunnuseid nagu vanus, sugu, piletiklass ja muud parameetrid. See andmekogum on tasuta kättesaadav ning seda saab avada ja analüüsida erinevatel offline- ja veebipõhistel platvormidel.

Link Titanicu andmekogumile:

<https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv>

Varasemalt peatükis "LLM-võimelised IDE-d ja tulevased muudatused programmeerimises" käsitlesime juba Jupyter Notebook - ühte kõige populaarsemat arenduskeskkonda andmeanalüüsi ja masinõppe jaoks. Jupyter Notebooki tasuta pilvepõhised analoogid on Kaggle ja Google Collab platvormid, mis võimaldavad Pythoni koodi käivitada ilma tarkvara installeerimata ja pakuvad tasuta juurdepääsu arvutiressurssidele.

Kaggle on suurim andmeanalüütika, masinõppe võistlusplatvorm koos integreeritud koodi täitmise keskkonnaga. Kaggle'il on 2023. aasta oktoobri seisuga üle 15 miljoni kasutaja [156] 194 riigist.

Laadige alla ja kasutage Titanicu andmestikku Kaggle'i platvormil (joonis 9.2-5), et salvestada andmestik (selle koopia) ja käivitada Pythoni kood koos eelinstalleeritud raamatukogudega otse

brauseris, ilma et oleks vaja paigaldada spetsiaalset IDE-d.



Joonis 9.2-5 Titanic'u tabeli statistika - kõige populaarsem treeningandmestik andmeanalüüsi ja masinõppe õppimiseks.

Titanicu andmekogum sisaldab andmeid 2224 reisija kohta, kes olid *RMS Titanicu* pardal selle hukkumise ajal 1912. aastal. Andmekogum on esitatud kahe eraldi tabelina, treening- (train.csv) ja testvalimina (test.csv), mis võimaldab seda kasutada nii mudelite treenimiseks kui ka nende täpsuse hindamiseks uutel andmetel.

Koolitusandmestik sisaldab nii reisijate atribuute (vanus, sugu, piletiklass ja muud) kui ka teavet selle kohta, kes jäid ellu (veerg binaarsete väärtustega "Survived"). Mudeli treenimiseks kasutatakse treeningu andmestikku (joonis 9.2-6 - fail train.csv). Testandmekogum (joonis 9.2-7 - fail test.csv) sisaldab ainult reisijate atribuute ilma ellujäänute teabeta (ilma ühegi veeruga "Survivor"). Testandmekogum on mõeldud mudeli testimiseks uute andmetega ja selle täpsuse hindamiseks.

Seega on meil peaaegu identsed reisijate atribuudid treening- ja testandmekogumites. Ainus oluline erinevus on see, et testandmekogumis on meil nimekiri reisijatest, kellel puudub veerg "Survivor" - sihtmuutuja, mida me tahame õppida prognoosima erinevate matemaatiliste algoritmide abil. Ja pärast mudeli koostamist saame võrrelda meie mudeli väljundit testandmekogumi tegeliku parameetriga "Survivor", mida võtame tulemuste hindamisel arvesse.

Tabeli peamised veerud, reisijate parameetrid koolitus- ja testandmestikus:

- **PassengerId** - reisijate unikaalne identifikaator
- **Ellujäänud** - 1, kui reisija jäi ellu, 0, kui surnud (ei ole testikogumis saadaval).
- **Pclass** - piletiklass (1, 2 või 3)
- **Nimi** - reisija nimi
- **Sugu** - reisija sugu (mees/naine)

- Vanus
- **SibSp** - pardal olevate vendade/õdede või abikaasade arv
- **Pardal** - pardal olevate vanemate või laste arv
- **Pilet** - pileti number
- **Sõiduhind** - pileti hind
- Kabiin - kabiini number (paljud andmed puuduvad)
- **Pardale minek** on pardalemineku sadam (C = Cherbourg, Q = Queenstown, S = Southampton).

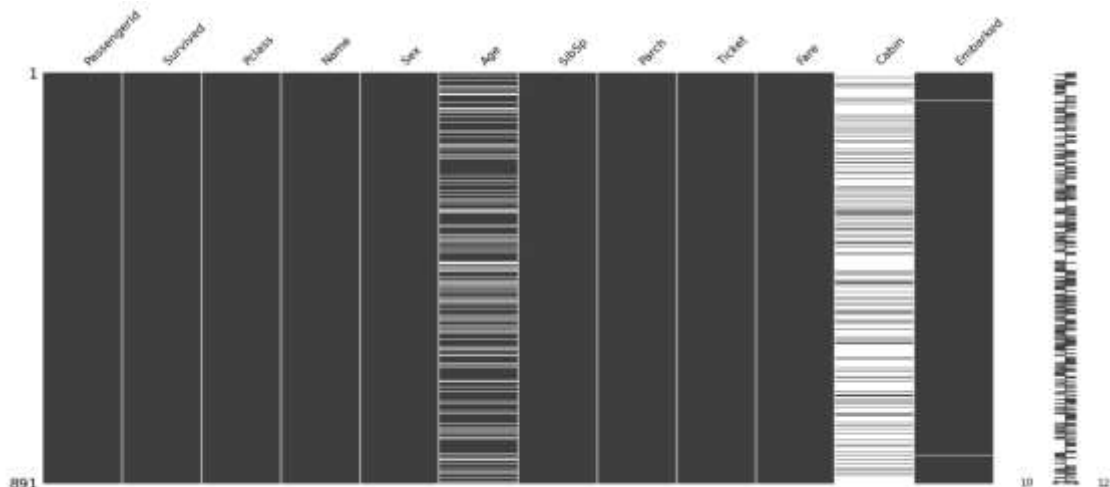
Puuduvate andmete visualiseerimiseks mõlemas tabelis saate kasutada missingno raamatukogu (joonis 9.2-6, joonis 9.2-7), mis kuvab puuduvad väärtused histogrammi kujul, kus valged väljad näitavad puuduvaid andmeid. See visualiseerimine võimaldab andmete kvaliteeti enne töötlemist kiiresti hinnata.



```

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()

```



Joonis 9.2-6 Mõned koodiread kasutatakse puuduvate andmete visualiseerimiseks Titanicu treeningu andmestikus, kus treeningu võtmevariabliks on parameeter "Survived".



Joonis 9.2-7 Visualiseerimine puuduvate andmete kohta Titanicu testandmestikus, mis sisaldab ainult reisijate omadusi, mille kohta puudub teave.

Enne hüpoteeside sõnastamist ja prognooside tegemist andmekogumi põhjal aitavad visuaalsed analüüsid tuvastada põhilisi mustreid andmetes, hinnata nende kvaliteeti ja tuvastada võimalikke sõltuvusi. On mitmeid visualiseerimistehnikaid, mis aitavad Titanicu andmestikku paremini mõista. Te võite kasutada jaotuse graafikuid reisijate vanuserühmade analüüsimiseks, ellujäämisgraafikuid soo ja klassi järgi ning puuduvate andmete maatriksid, et hinnata teabe kvaliteeti ja mõista andmeid.

- Palume LLMi aidata meil visualiseerida Titanicu andmestiku andmeid, saates järgmise tekstitaotluse mis tahes LLMi mudelile (CHATGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN või mõni muu):

Palun näidake mõned lihtsad graafikud Titanicu andmekogumi kohta. Laadige andmestik ise alla ja näidake. ↵

- LLM vastus valmis koodi ja graafikute kujul, mis visualiseerivad andmekogumi parameetreid

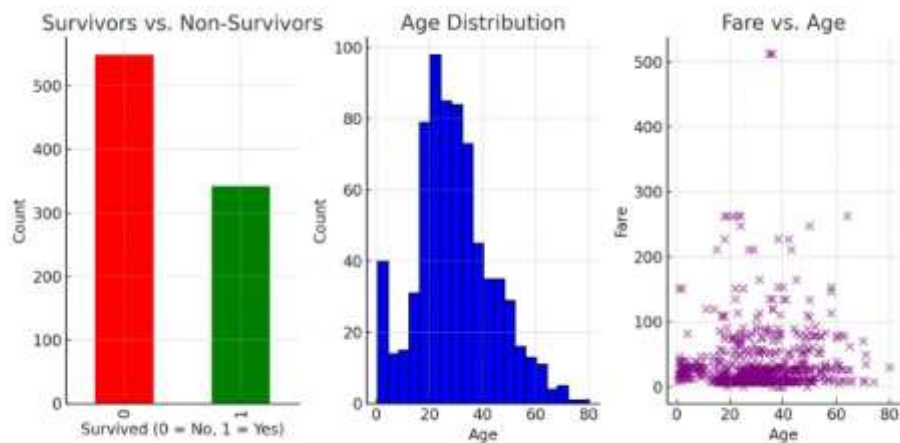
Create in LLM Chat

```

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()

```

Run in IDE



Joonis 9.2-8 LLM aitab teil kohe visualiseerida andmekogumi andmeid.

Andmete visualiseerimine on oluline samm, et valmistada andmestik ette hilisemaks masinõppe mudeli konstrueerimiseks, mida saab kasutada ainult andmete mõistmisel.

Masinõpe tegevuses: Titanicu reisijatest projektijuhtimiseni

Peamine hüpotees, mida kasutati masinõppe raamistiku uurimiseks Titanicu andmestiku põhjal, on, et teatavatel reisijate rühmadel oli suurem tõenäosus ellu jääda.

Titanicu reisijate väike tabel on muutunud populaarseks kogu maailmas ning miljonid inimesed kasutavad seda treenimiseks, katsetamiseks ja mudeli testimiseks, et välja selgitada, millised algoritmid ja hüpoteesid loovad Titanicu reisijate treeningandmestiku põhjal kõige täpsema ellujäämise prognoosimudeli.

Titanicu andmestiku atraktiivsus on tingitud selle kompaktsusest: mitmesaja rea ja kaheteistkümne veeruga (joonis 9.2-6) annab see piisavalt võimalusi analüüsiks. Andmekogum on suhteliselt lihtsalt klassikaline näide binaarsest klassifitseerimislahendusest, kus probleemi eesmärk - ellujäämine - on väljendatud mugavas formaadis 0 või 1.

John Wheeler filmis "It from Bit" [7] väidab, et universum põhineb binaarsetel valikutel. Samamoodi on molekulidest koosnevate inimeste juhitav äri tegelikult üles ehitatud binaarsetele binaarsetele valikutele.

Lisaks sellele põhinevad andmed tegelikul ajaloolisel sündmusel, mis muudab need erinevalt kunstlikult loodud näidetest uurimistööks väärtuslikuks. Ainuüksi Kaggle'i platvormil, mis on üks suurimaid andmeportaali ja ETL, osales Titanicu andmestikul põhinevates ülesannetes 1 355 998 inimest, kes arendasid 53 963 unikaalset andmeportaali lahendust [157] (joonis 9.2-9).

Tundub uskumatu, kuid vaid 1000 rida andmeid Titanicu reisijate kohta 12 parameetriga on muutunud miljonite hüpoteeside, loogiliste ahelate ja ainulaadsete andmepiipelide väljaks. Väikesest andmestikust sünnivad lõputud arusaamad, hüpoteesid ja tõlgendused - alates lihtsatest ellujäämismudelitest kuni keeruliste ansambliteni, mis võtavad arvesse varjatud mustreid ja keerukaid arutluslabürinte.

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

...

Joonis 9.2-9 Esimesed viis lahendust 53 963 valmis ja avatud lähtekoodiga Pipeline'i lahendusest. Ainuüksi Kaggle'is on seda probleemi juba peaaegu 1,5 miljonit inimest proovinud lahendada [157].

Kui isegi selline väike tabel võib genereerida miljoneid unikaalseid lahendusi (joonis 9.2-9), siis mida saame öelda reaalsete tööstuslike ehitusandmetike kohta, kus parameetrite arv on kümneid tuhandeid?

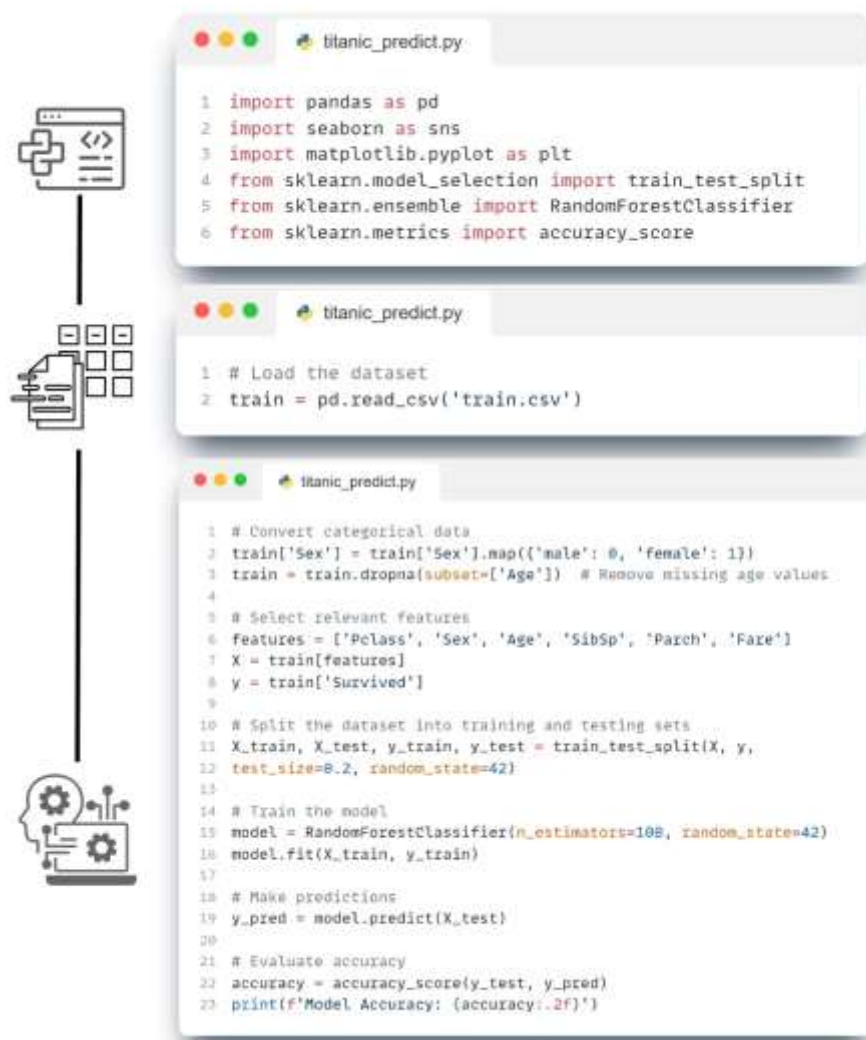
Standardne CAD - suhteliselt väikese hoone projekt sisaldab kümneid tuhandeid üksusi tuhandete parameetritega - alates geomeetrilistest omadustest kuni kulude ja aja atribuutideni. Kujutage ette, kui palju potentsiaalseid teadmisi, seoseid, prognoose ja juhtimishüpoteese on peidetud kõigi teie ettevõtte aastate jooksul kogutud projektide andmetesse. Ajaloolised projektiandmed ei ole lihtsalt arhiiv - need on organisatsiooni elav mälu, selle digitaalne jalajälg, mida saab analüüsida, et luua suur hulk unikaalseid hüpoteese.

Kõige tähtsam on see, et te ei pea ootama, et Kaggle'i kogukond teie ettevõtte või andmete vastu huvi tunneks. Võite alustada tööd sellega, mis teil juba täna olemas on: käivitage oma andmetega analüütikat, treenige oma andmetega mudeleid, tuvastage kordusi, anomaaliaid ja mustreid. Seal, kus varem oli vaja aastatepikkust katsetamist ja kallist nõustamist, piisab nüüd algatusvõimest, LLM-ist, avatud lähenemisest andmetele ja õppimisvalmidusest.

- 🗣️ Et luua masinõppe algoritm, mis ennustab reisijate ellujäämisprotsenti train.csv reisijate treeningandmestiku põhjal, palume LLM lahendada see probleem meie jaoks:

Luu Titanicu reisijate treeninguandmestiku põhjal masinõppe mudel, et ennustada ellujäämismäära. ↵

LLMi vastus :



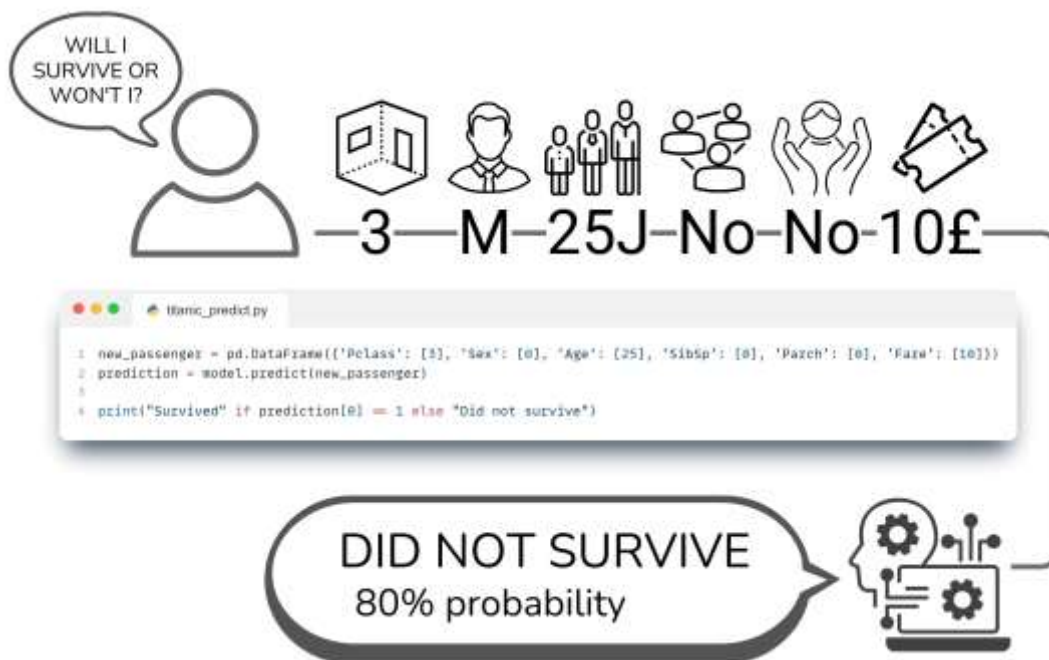
Joonis 9.2-10 LLM koostas Titanicu ellujäänute prognoosi, kasutades masinõppe algoritmi Random Forest.

LLM-i (joonis 9.2-10) saadud kood laeb Titanicu reisijate andmed, puhastab need, konverteerib kategoorilised muutujad (nt sugu numbriliseks) ja treenib mudeli RandomForestClassifier algoritmi abil, et ennustada, kas reisija jäi ellu või mitte (populaarsetest algoritmidest räägime lähemalt järgmistes peatükkides).

Kood jagab koolitusandmed koolitus- ja testkogumiteks (Kaggle'i veebilehel on juba loodud valmis test.csv (joonis 9.2-7) ja train.csv (joonis 9.2-6) koolituse jaoks, seejärel treenitakse mudel koolitusandmetel ja testitakse testandmetel, et näha, kui hea on konkreetne prognoosimudel. Pärast treenimist sisestatakse testandmed test.csv-st (koos tegelike andmetega nende kohta, kes jäid ellu või ei jäänud ellu) mudelisse ja see ennustab, kes jäid ellu ja kes mitte. Meie puhul on saadud masinõppe mudeli täpsus umbes 80%, mis näitab, et see kajastab mustreid üsna hästi.

Masinõpet võib võrrelda lapsega, kes üritab riskikülikukujulise klotsi ümmargusse auku sobitada. Algoritm proovib algstaadiumis mitmeid lähenemisviise, puutub kokku vigade ja vastuoludega. See protsess võib tunduda ebaefektiivne, kuid see pakub olulist õppimist: iga viga analüüsidest parandab mudel oma prognoose ja teeb üha täpsemaid otsuseid.

Nüüd saab seda mudelit (joonis 9.2-10) kasutada uute reisijate ellujäämismäära prognoosimiseks ja näiteks, kui sööta sellele reisijate teavet, kasutades funktsiooni model.predict parameetritega: "mees", "3. klass", "25-aastane", "ei ole sugulasi pardal", siis annab mudel prognoosi - et reisija, kes oli 1912. aastal Titanicul (joonis 9.2-11), ei ela õnnetust 80% tõenäosusega üle (joonis 9.2-11).



Joonis 9.2-11 Eespool loodud mudel suudab nüüd 80% tõenäosusega ennustada, kas mõni uus Titanicu reisija jääb ellu või mitte.

Titanicu reisijate ellujäämise prognoosimudel illustreerib palju laiemat kontseptsiooni: iga päev teevad tuhanded ehitussektori spetsialistid sarnaseid "topeltotsuseid" - elu või surma otsus, projekt, hinnang, tööriist, kasum või kahjum, ohutus või risk. Nagu Titanicu näites, kus tulemus sõltus teguritest (sugu, vanus, klass), mõjutavad ehituses iga otsuse aspekti paljud oma tegurid ja muutujad (tabelite veerud): materjalide maksumus, tööliste kvalifikatsioon, tähtajad, ilmastik, logistika, tehnilised riskid, kommentaarid ja sajad tuhanded muud parameetrid.

Ehitustööstuses järgib masinõpe samu põhimõtteid kui teistes valdkondades: mudelid treenitakse varasemate andmete - projektide, lepingute ja hinnangute - põhjal, et testida erinevaid hüpoteese ja leida kõige tõhusamad lahendused. See protsess on nagu lapse õpetamine katse ja eksituse kaudu: iga tsükliga kohanduvad mudelid ja muutuvad täpsemaks.

Kogutud andmete kasutamine avab ehituses uusi horisonte. Ajamahukate käsitsi tehtavate arvutuste asemel saab koolitada mudeleid, mis suudavad tulevaste projektide põhinäitajad suure täpsusega ennustada. Sel viisil muudab prognoosiv analüüs ehitussektori ruumiks, kus saab mitte ainult planeerida, vaid ka kindlalt arenguid prognoosida.

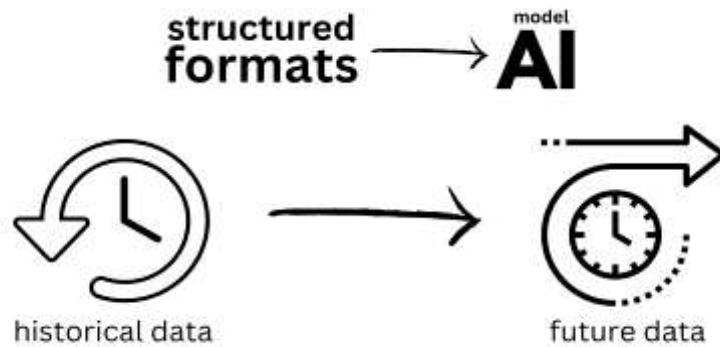
Ajaloolistel andmetel põhinevad prognoosid ja prognoosid

Ettevõtte projektide kohta kogutud andmed avavad võimaluse luua mudeleid, mis võimaldavad prognoosida tulevaste, veel realiseerimata objektide kulu- ja ajaomadusi - ilma aeganõudvate manuaalsete arvutuste ja võrdlusteta. See võimaldab oluliselt kiirendada ja lihtsustada hindamisprotsesse, tuginedes mitte subjektiivsetele oletustele, vaid usaldusväärsetele matemaatilistele prognoosidele.

Varem, raamatu neljandas osas, oleme üksikasjalikult käsitlenud traditsioonilisi projekti kulude hindamise meetodeid, sealhulgas ressursipõhist meetodit, ning maininud ka parameetrilisi ja ekspertide lähenemisviise. Need meetodid on endiselt asjakohased, kuid tänapäevases praktikas on neid hakatud rikastama statistilise analüüsi ja masinõppe vahenditega, mis võivad oluliselt parandada hinnangute täpsust ja reprodutseeritavust.

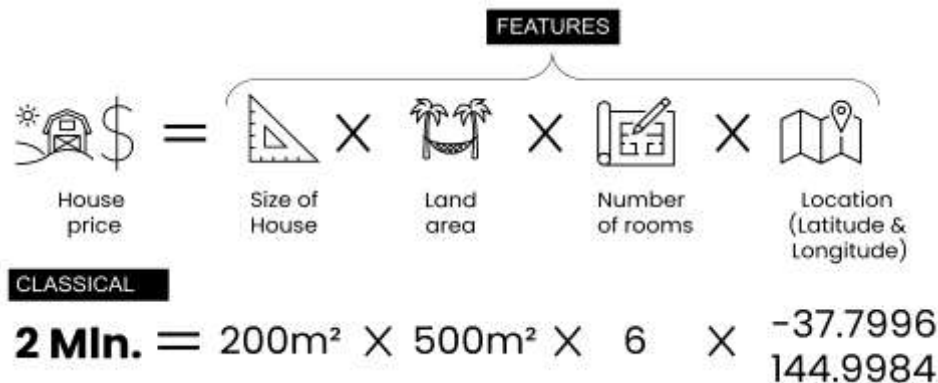
Hinna- ja ajaatribuutide käsitsi ja poolautomaatselt arvutamise protsesse täiendavad tulevikus ML-mudelite arvamus ja prognoosid, mis on võimelised analüüsima ajaloolisi andmeid, leidma varjatud mustreid ja tegema ettepanekuid teadlike otsuste tegemiseks. Uued andmed ja stsenaariumid luuakse automaatselt juba olemasolevast teabest - sarnaselt sellega, kuidas keelemudelid (LLM) loovad tekstid, pildid ja koodid aastate jooksul avatud allikatest kogutud andmete põhjal [158].

Nii nagu inimesed tuginevad täna tulevaste sündmuste hindamisel kogemustele, intuitsioonile ja sisemisele statistikale, määratakse lähiaastatel ehitusprojektide tulevik üha enam kindlaks kogutud teadmiste ja matemaatiliste masinõppe mudelite kombinatsiooniga.



Joonis 9.2-12 Kvalitatiivsed ja struktureeritud ajaloolised ettevõtte andmed on materjal, mille põhjal masinõppe mudelid ja prognoosid ehitatakse.

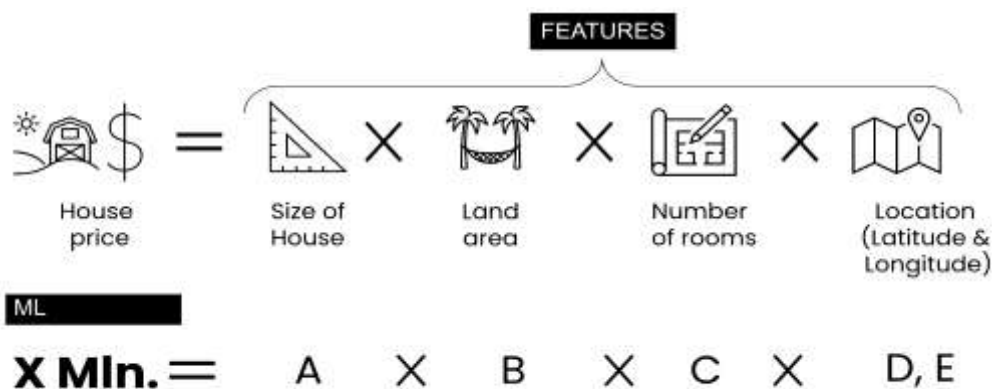
Võtame lihtsa näite: maja hinna ennustamine selle pindala, krundi suuruse, tubade arvu ja geograafilise asukoha alusel. Üks lähenemisviis on luua klassikaline mudel, mis analüüsib neid parameetreid ja arvutab eeldatava hinna (joonis 9.2-13). See lähenemisviis nõuab täpset ja eelnevalt teadaolevat valemit, mis on tegelikus praktikas praktiliselt võimatu.



Joonis 9.2-13 Klassikalise algoritmi abil saab hinnata maja väärtust, mille jaoks tuleb leida kindel valem.

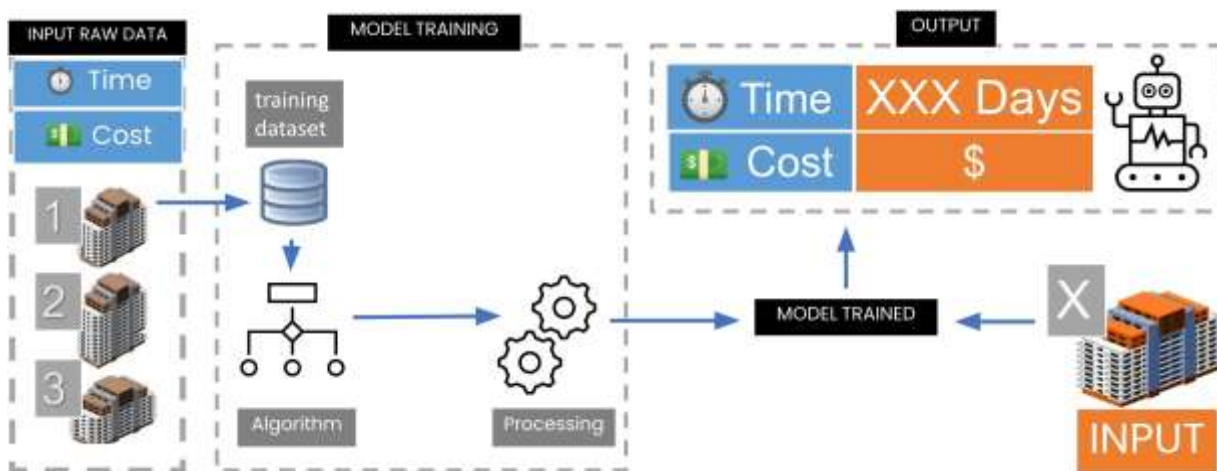
Masinõpe kõrvaldab käsitsi otsitavad valemid ja asendab need koolitatud algoritmidega, mis tuvastavad iseseisvalt sõltuvused, mis on mitu korda täpsemad kui mis tahes etteantud võrrandid. Alternatiivina loome masinõppe algoritmi, mis genereerib mudeli, mis põhineb probleemi eelneval mõistmisel ja ajaloolistel andmetel, mis võivad olla puudulikud (joonis 9.2-14).

Hinnakujunduse näitel võimaldab masinõpe luua eri tüüpi matemaatilisi mudeleid, mis ei nõua teadmisi kulude kujunemise täpse mehhanismi kohta. Mudel "õpib" eelmiste projektide andmete põhjal, kohandudes ehitusparameetrite, nende maksumuse ja tähtaegade vaheliste tegelike mustritega.



Joonis 9.2-14 Erinevalt klassikalisest valemipõhisest hindamisest on masinõppe algoritm treenitud ajalooliste andmete põhjal.

Järelevalvega masinõppe kontekstis sisaldab iga projekt treeningu andmekogumis nii sisendatribuute (nt sarnaste ehitiste maksumus ja aeg) kui ka oodatavaid väljundväärtusi (nt maksumus või aeg). Sarnast andmekogumit kasutatakse masinõppe mudeli loomiseks ja kohandamiseks (joonis 9.2-15). Mida suurem on andmekogum ja mida kvaliteetsemad on selles sisalduvad andmed, seda täpsem on mudel ja seda täpsemad on prognoositulemused.



Joonis 9.2-15 ML-mudel, mis on treenitud varasemate projektide kulude ja ajakava andmete põhjal, määrab uue projekti kulud ja ajakava teatud tõenäosusega.

Kui mudel on loodud ja treenitud uue projekti ehitamise hindamiseks, tuleb mudelile lihtsalt esitada uue projekti uued atribuudid ja mudel annab hinnangulised tulemused, mis põhinevad varem õpitud mustritel ja teatud tõenäosusega.

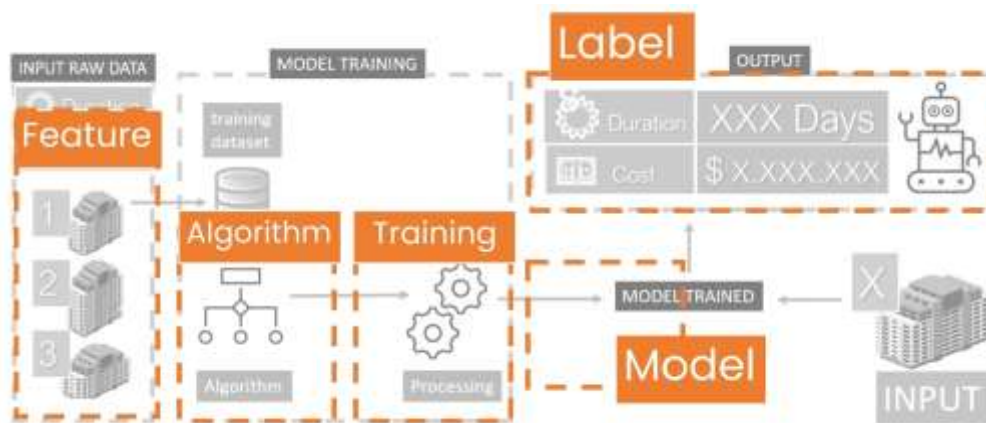
Masinõppe põhimõisted

Masinõpe ei ole maagia, see on lihtsalt matemaatika, andmed ja mustrite leidmine. Sellel ei ole tegelikku intelligentsust, vaid see on programm, mis on koolitatud andmete põhjal mustrite

äratundmiseks ja otsuste tegemiseks ilma inimese pideva osaluseta.

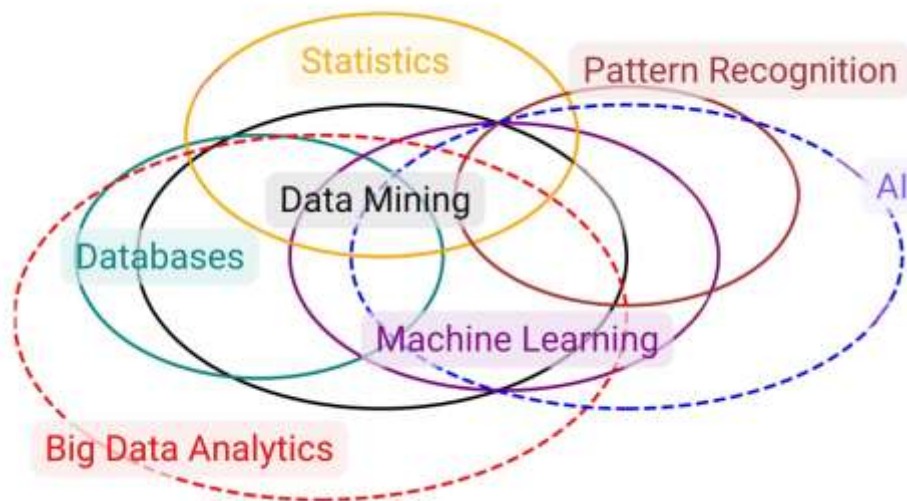
Masinõpe kasutab oma struktuuri kirjeldamiseks mitmeid põhimõisteid (joonis 9.2-16):

- **Märgised** on sihtm muutujad või atribuudid (Titanicu andmestikus parameeter "Survivor"), mida mudel peaks prognoosima. Näide: ehitusmaksumus (nt dollarites), ehitustööde kestus (nt kuudes).
- **Tunnused** on sõltumatud muutujad või atribuudid, mis on mudeli sisendiks. Prognoosimudelis kasutatakse neid tunnuste prognoosimiseks. Näited: krundi suurus (ruutmeetrites), hoone korruste arv, hoone kogupindala (ruutmeetrites), geograafiline asukoht (laius- ja pikkuskraad), ehituses kasutatud materjalide tüüp. Tunnuste arv määrab ka andmete mõõtmelisuse.
- **Mudel** on erinevate hüpoteeside kogum, millest üks lähendab prognoositavat või lähendatavat sihtfunktsiooni. Näide: masinõppe mudel, mis kasutab regressioonanalüüsi meetodeid, et ennustada ehitamise kulusid ja ajakava.
- **Õppimisalgoritm** **Õppimisalgoritm** on protsess, mille käigus leitakse mudelis parim hüpotees, mis vastab täpselt sihtfunktsioonile, kasutades treeninguandmete kogumit. Näide: lineaarne regressioon, KNN või juhuslik metsaalgoritm, mis analüüsib kulude ja ehitusaja andmeid, et tuvastada seoseid ja mustreid.
- **Koolitus** - Koolitusprotsessi käigus analüüsib algoritm koolitusandmeid, leides mustreid, mis vastavad sisendatribuutide ja sihtmärgiste vahelisele seosele. Selle protsessi tulemuseks on koolitatud masinõppe mudel, mis on valmis prognoosimiseks. Näide: protsess, mille käigus algoritm analüüsib ajaloolisi ehitusandmeid (kulud, aeg, rajatise omadused), et luua prognoositav mudel.



Joonis 9.2-16 ML kasutab märgiseid ja atribuute, et luua mudelid, mida treenitakse andmete põhjal algoritmide abil, et ennustada tulemusi.

Masinõpe ei eksisteeri isoleeritult, vaid on osa laiemast analüütiliste distsipliinide ökosüsteemist, mis hõlmab statistikat, andmebaase, andmekae vandamist, mustrituvastust, suurandmete analüüsi ja tehisintellekti. Joonis 9.2-17 näitab, kuidas need valdkonnad kattuvad ja täiendavad teineteist, pakkudes kaasaegsete otsuste tegemise ja automatiseerimise süsteemide terviklikku raamistikku.

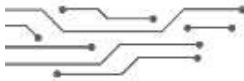


Joonis 9.2-17 Andmeanalüüsi eri valdkondade - statistika, masinõppe, tehisintellekti, suurandmete, mustrituvastuse ja andmekaeve - vaheline seos.

Masinõppe peamine eesmärk on anda arvutitele võime õppida automaatselt teadmisi ilma inimese sekkumiseta või abita ning kohandada oma tegevust vastavalt sellele [159].

Seega on inimese roll tulevikus ainult selles, et ta annab masinale kognitiivseid võimeid - ta määrab tingimused, kaalud ja parameetrid ning masinõppe mudel teeb kõik muu.

Järgmises peatükis käsitleme konkreetseid näiteid algoritmide rakendamise kohta. Kasutatakse reaalseid tabeleid ja lihtsustatud mudeleid, et näidata, kuidas prognoos samm-sammult üles ehitatakse.



PEATÜKK 9.3.

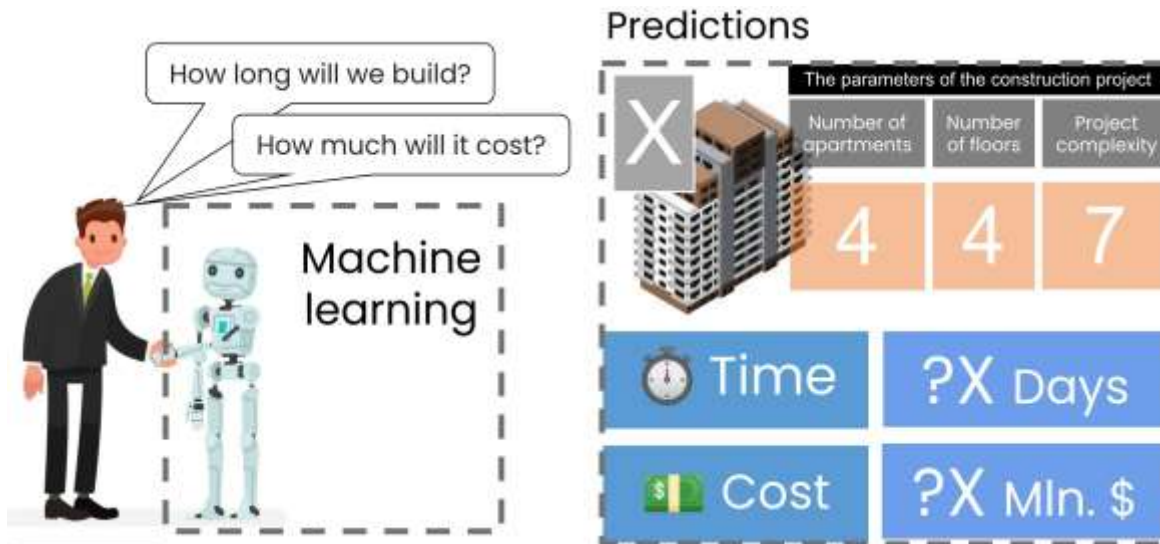
KULUDE JA TÄHTAEGADE PROGNOOSIMINE MASINÕPPE ABIL

Näide masinõppe kasutamise kohta projekti kulude ja tähtaegade leidmiseks.

Ehitusaja ja -kulude hindamine on üks peamisi protsesse ehitusettevõtte tegevuses. Traditsiooniliselt on selliseid hinnanguid koostanud eksperdid, tuginedes kogemustele, võrdlusraamatutele ja regulatiivsetele andmebaasidele. Kuid tänu digitaalsele ümberkujundamisele ja andmete kättesaadavuse suurenemisele on nüüd võimalik kasutada masinõppe (ML) mudeleid, et parandada selliste hinnangute täpsust ja automatiseerida neid.

Masinõppe kasutuselevõtt ehitusmaksumuse ja -aja arvutamise protsessis ei võimalda mitte ainult tõhusamat planeerimist, vaid saab ka lähtepunktiks intelligentsete mudelite integreerimiseks teistesse äriprotsessidesse - alates riskijuhtimisest kuni logistika ja hangete optimeerimiseni.

On oluline, et oleks võimalik kiiresti kindlaks teha, kui kaua võtab projekti ehitamine aega ja milline on selle kogumaksumus. Need küsimused projekti aja ja maksumuse kohta on traditsiooniliselt olnud nii klientide kui ka ehitusettevõtete tähelepanu keskmes juba ehitussektori sünnist saadik.



Joonis 9.3-1 Ehitusprojektide puhul on ehitusaja ja -kulude hindamise kiirus ja kvaliteet peamised edutegurid.

Järgnevas näites võetakse välja varasemate projektide põhiandmed ja neid kasutatakse masinõppe mudeli väljatöötamiseks, mis võimaldab kasutada mudelit uute parameetritega uute ehitusprojektide maksumuse ja ajakava hindamiseks (joonis 9.3-1).

Vaatleme kolme projekti, millel on kolm peamist omadust: korterite arv (kus 100 korterit on visualiseerimise lihtsustamiseks võrdne arvuga 10), korruste arv ja ehitamise keerukuse tingimuslik

mõõtühik skaalal 1-10, kus 10 on kõrgeim keerukuse hinne. Masinõppes nimetatakse väärtuste, näiteks 100 väärtuse teisendamist ja lihtsustamist väärtuseks 10 või 50 väärtuseks 5, "normaliseerimiseks".

Normaliseerimine on masinõppes protsess, mille käigus viiakse erinevad arvandmed ühtsele skaalale, et hõlbustada töötlemist ja analüüsi. See protsess on eriti oluline, kui andmed on erineva skaala ja mõõtühikutega.

Oletame, et esimene projekt (joonis 9.3-2) oli 50 korterit (pärast normaliseerimist 5), 7 korrust ja keerukuse hinne 2, mis tähendas suhteliselt lihtsat ehitust. Teisel projektil oli juba 80 korterit, 9 korrust ja suhteliselt keerukas konstruktsioon. Sellistel tingimustel võttis esimese ja teise korterelamu ehitamine aega 270 ja 330 päeva ning projekti kogumaksumus oli vastavalt 4,5 miljonit ja 5,8 miljonit dollarit.

Construction project	The parameters of the construction project			The key parameters of the project	
	Number of apartment s 100 = 1000 apt	Number of floors	Project complexity	Time Days	Cost The total cost of the project
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

Joonis 9.3-2 Näide varasemate projektide kogumist, mida kasutatakse tulevase projekti X aja ja kulude hindamiseks.

Selliste andmete jaoks masinõppe mudeli loomisel on peamine ülesanne tuvastada prognoosimiseks kriitilised tunnused (või sildid), käesoleval juhul ehitusaeg ja -kulud. Väikese andmestiku puhul kasutame uute ehitusprojektide kavandamisel teavet varasemate ehitusprojektide kohta: kasutades masinõppe algoritme, peame ennustama uue projekti X maksumust ja ehitusaega, tuginedes uue projekti antud atribuutidele, näiteks 40 korterit, 4 korrust ja projekti suhteline suur keerukus 7 (joonis 9.3-2). Reaalses olukorras võib sisendparameetrite arv olla palju suurem, ulatudes mitmetest kümnetest kuni sadade teguriteni. Nende hulka võivad kuuluda: ehitusmaterjalide tüüp, kliimavöönd, töövõtjate kvalifikatsioonitase, kommunaalteenuste kättesaadavus, vundamendi tüüp, tööde alguse aeg, töödejuhatajate märkused jne.

Ennustava masinõppe mudeli loomiseks peame valima algoritmi. Algoritm on masinõppes nagu matemaatiline retsept, mis õpetab arvutile, kuidas teha prognoose (parameetrite õiges järjekorras segamine) või teha andmete põhjal otsuseid.

Varasemate ehitusprojektide andmete analüüsimiseks ja tulevaste projektide aja ja maksumuse prognoosimiseks (joonis 9.3-2) saab kasutada ühte populaarset masinõppe algoritmi:

- **Lineaarne regressioon (lineaarne regressioon):** selle algoritmiga püütakse leida otsest seost tunnuste vahel, näiteks korruste arvu ja ehitusmaksumuse vahel. Algoritmi eesmärk on leida seda seost kõige paremini kirjeldav lineaarne võrrand, mis võimaldab teha prognoose.
- **Algoritm k-naabrid (k-NN):** see algoritm võrdleb uut projekti varasemate projektidega, mis olid suuruselt või keerukuselt sarnased. K-NN klassifitseerib andmed selle alusel, millised k (arv) treeningnäidetest on sellele kõige lähemal. Regressiooni kontekstis on tulemus k lähima naabri keskmine või mediaan.
- **Otsustuspuud:** on ennustav modelleerimismudel, mis jagab andmed erinevate tingimuste alusel alamhulkadeks, kasutades puude struktuuri. Puu iga sõlm esindab tingimust või küsimust, mis viib andmete edasise jaotamiseni, ja iga leht esindab lõplikku prognoosi või tulemust. Algoritm jagab andmed prognooside tegemiseks väiksemateks rühmadeks erinevate tunnuste alusel, näiteks kõigepealt lugude arvu, seejärel keerukuse jne alusel.

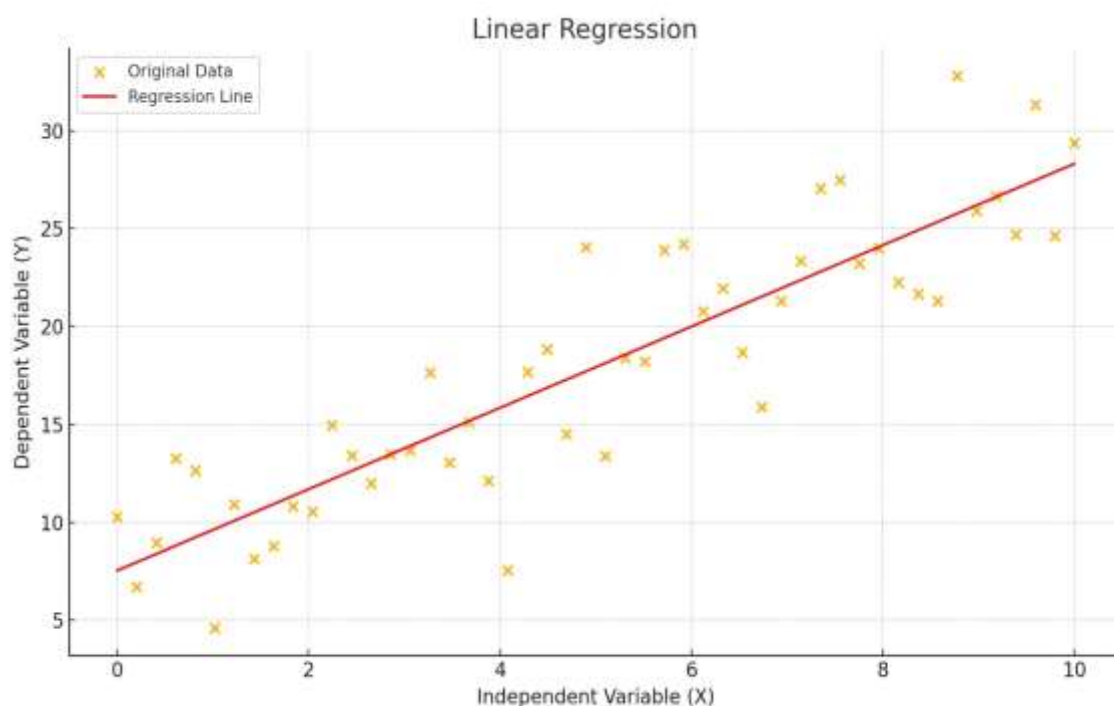
Vaatleme masinõppe algoritme uue projekti maksumuse hindamiseks kahe populaarse algoritmi näitel: lineaarne regressioon ja K-lähedaseimate naabrite algoritm.

Projekti kulude ja aja prognoosimine lineaarse regressiooni abil

Lineaarne regressioon on põhiline andmeanalüüsi algoritm, mis ennustab muutuja väärtust lineaarse seose alusel ühe või mitme teise muutujaga. See mudel eeldab, et sõltuva muutuja ja ühe või mitme sõltumatu muutuja vahel on otsene lineaarne seos ning algoritmi eesmärk on leida see seos.

Lineaarse regressiooni lihtsus ja selgus on muutnud selle populaarseks vahendiks paljudes valdkondades. Kui tegemist on ühe muutujaga, on lineaarse regressiooni eesmärk leida kõige paremini sobiv joon läbi andmepunktide.

Lineaarne regressioon leiab parima sirge (punane joon), mis lähendab sisendmuutuja X ja väljundmuutuja Y vahelist sõltuvust. See joon võimaldab ennustada Y väärtusi uute X väärtuste korral tuvastatud lineaarse seose alusel (joonis 9.3-3).



Joonis 9.3-3 Lineaarse regressiooni põhimõte on leida parim sirge läbi treeningväärtuste.

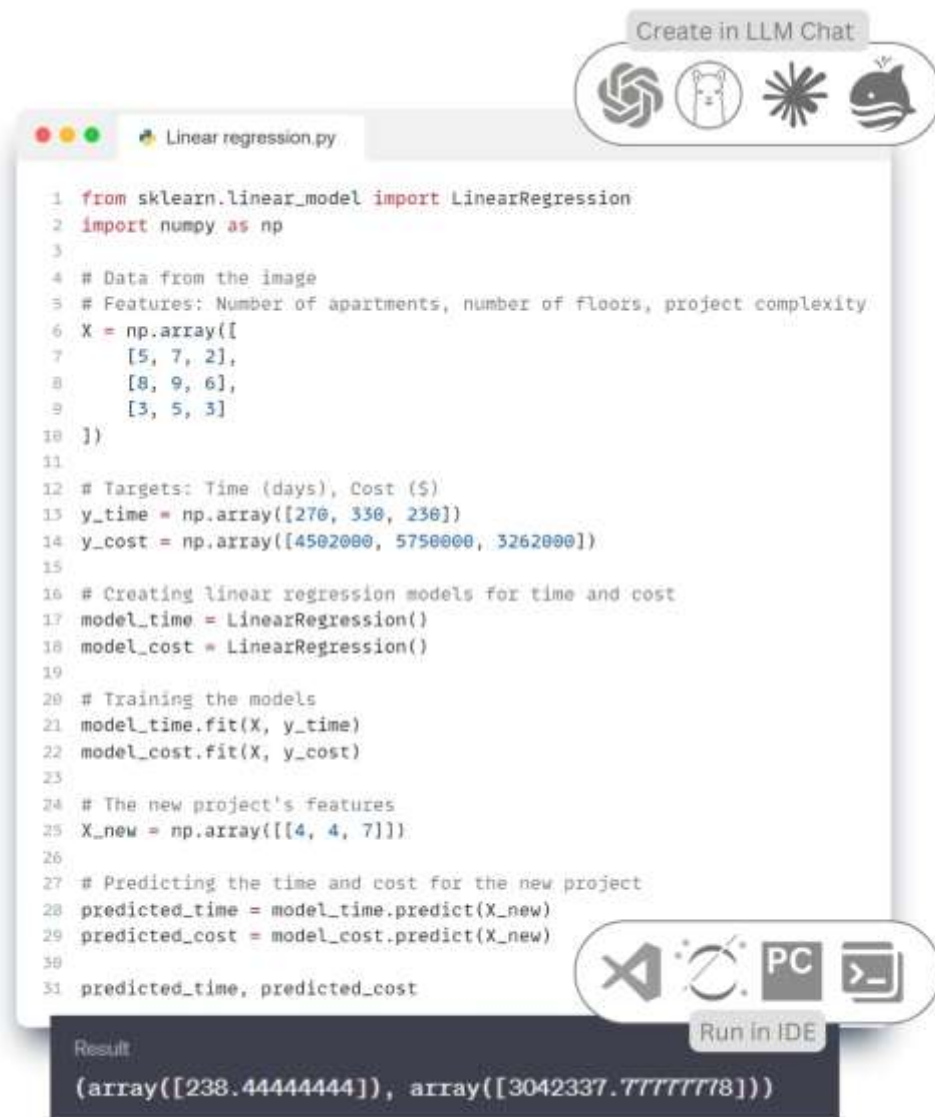
Seda joont kujutab võrrand, milles sõltumatu muutuja (X) väärtuse sisestamine annab sõltuva muutuja (Y) prognoositava väärtuse. See protsess võimaldab Y-i tõhusalt ennustada teadaolevate X-i väärtuste põhjal, kasutades nendevahelist lineaarset seost. Näide sellise statistiliselt keskmistatud joone leidmise kohta on esitatud San Francisco ehituslubade andmete hindamisel (joonis 9.1-7), kus lineaarse regressiooni abil arvutati inflatsioon eri tüüpi rajatiste jaoks.

Laadime projekti andmetabeli (joonis 9.3-2 eelmisest peatükist) otse LLM-i ja palume sellel luua meile lihtne masinõppe mudel.

🗨️ Saada tekstitaotlus LLM chat (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Peame näitama lihtsa masinõppe mudeli konstrueerimist, et ennustada uue projekti X elluviimise kulusid ja aega (joonis 9.3-2 lisatud pildil). ↵

LLMi vastus :



The image shows a code editor window titled "Linear regression.py". The code imports LinearRegression from sklearn and numpy. It defines features X as a 3x3 array and targets y_time and y_cost as 1x3 arrays. It then creates and fits two linear regression models, model_time and model_cost. Finally, it predicts the time and cost for a new project with features [4, 4, 7]. The output shows the predicted time as 238.4444444 and the predicted cost as 3042337.77777778.

```

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([238.4444444]), array([3042337.77777778]))

Joonis 9.3-4 LLM valis lineaarse regressiooni, et luua masinõppe mudel projekti kulude ja aja prognoosimiseks.

LLM tuvastas lisatud pildilt automaatselt tabeli ja konverteeris andmed visuaalsest formaadist tabelimassiivi (joonis 9.3-4 - rida 6). Selle massiivi alusel loodi tunnused ja sildid, mille põhjal loodi masinõppe mudel (joonis 9.3-4 - 17.-22. rida), milles kasutati lineaarset regressiooni.

Kasutades lineaarset regressioonimudelit, mis treeniti "äärmiselt väikese" andmekogumi põhjal, tehti prognoosid uue hüpoteetilise ehitusprojekti kohta, mis kandis nimetust "Projekt X". Meie probleemi puhul iseloomustab seda projekti 40 korterit, 4 korrust ja keerukusaste 7 (joonis 9.3-2).

Nagu ennustab piiratud ja väikesel andmestikul põhinev lineaarne regressioonimudel uue projekti X kohta (joonis 9.3-4 - rida 24-29):

- **Ehituse kestus** on ligikaudu 238 päeva (238.444444444).
- **Kogukulu** on ligikaudu \$ 3,042,338 (3042337.777).

Projekti kulude hüpoteesi edasiseks uurimiseks on kasulik katsetada erinevaid masinõppe algoritme ja meetodeid. Seepärast ennustame uue projekti X jaoks samu kulu- ja ajaväärtusi väikese hulga ajalooliste andmete põhjal, kasutades K-Nearest Neighbours algoritmi (k-NN).

Projekti kulude ja aja prognoosimine K-lähedaseima naabri algoritmi abil (k-NN).

Me kasutame uue projekti maksumuse ja kestuse hindamiseks täiendava prognoosijana k-Nearest Neighbours (k-NN) algoritmi. K-Nearest Neighbours (k-NN) algoritm on kontrollitud masinõppe (supervised machine learning) meetod nii klassifitseerimiseks kui ka regressiooniks. Oleme k-NN algoritmi varem käsitletud ka vektorite andmebaasiotsingu kontekstis (joonis 8.2-2), kus seda kasutatakse lähimate vektorite (nt tekstide, piltide või tehniliste kirjelduste) leidmiseks. Selle lähenemisviisi puhul esitatakse iga projekt punktina mitmemõõtmelises ruumis, kus iga mõõde vastab projekti konkreetsele atribuudile.

Meie puhul kujutame iga projekti kolm atribuuti kolmemõõtmelises ruumis punktidenä (joonis 9.3-5). Seega lokaliseeritakse meie tulevane projekt X selles ruumis koordinaatidega ($x=4$, $y=4$, $z=7$). Tuleb märkida, et reaalses tingimustes võib punktide arv ja ruumi dimensioonilisus olla suurusjärgu võrra suurem.

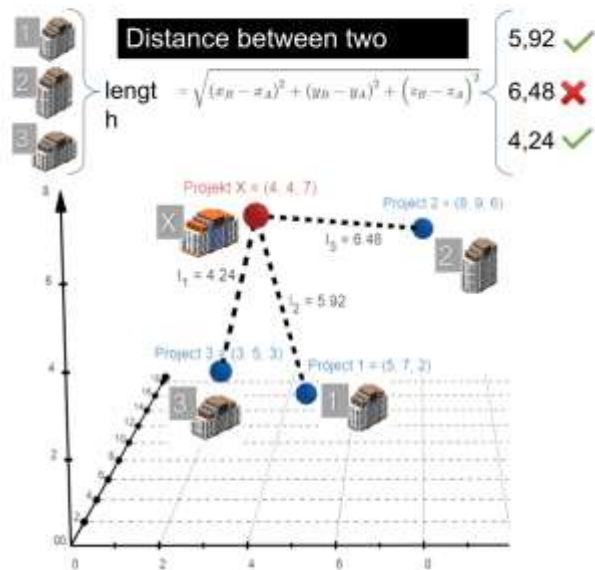
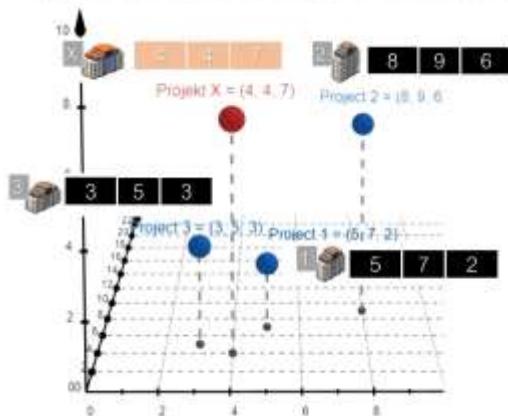
K-NN (k-nearest neighbours) algoritm töötab soovitud projekti X ja treeningandmebaasis olevate projektide vahelise kauguse mõõtmise teel. Neid vahemaid võrreldes määrab algoritm kindlaks projektid, mis on uue projekti X punktile kõige lähemal.

Näiteks kui meie esialgse andmestiku teine projekt ($x=8$, $y=9$, $z=6$) on X-ist palju kaugemal (joonis 9.3-5) kui teised projektid, võib selle edasisest analüüsist välja jätta. Selle tulemusena saab arvutustes kasutada ainult kahte ($k=2$) lähimat projekti, mille alusel määratakse keskmine väärtus.

Selline meetod võimaldab naabruskonna otsingu kaudu hinnata projektide sarnasusi, mis omakorda aitab teha järeldusi uue projekti tõenäoliste kulude ja ajakava kohta, tuginedes varem teostatud sarnastele projektidele.

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.



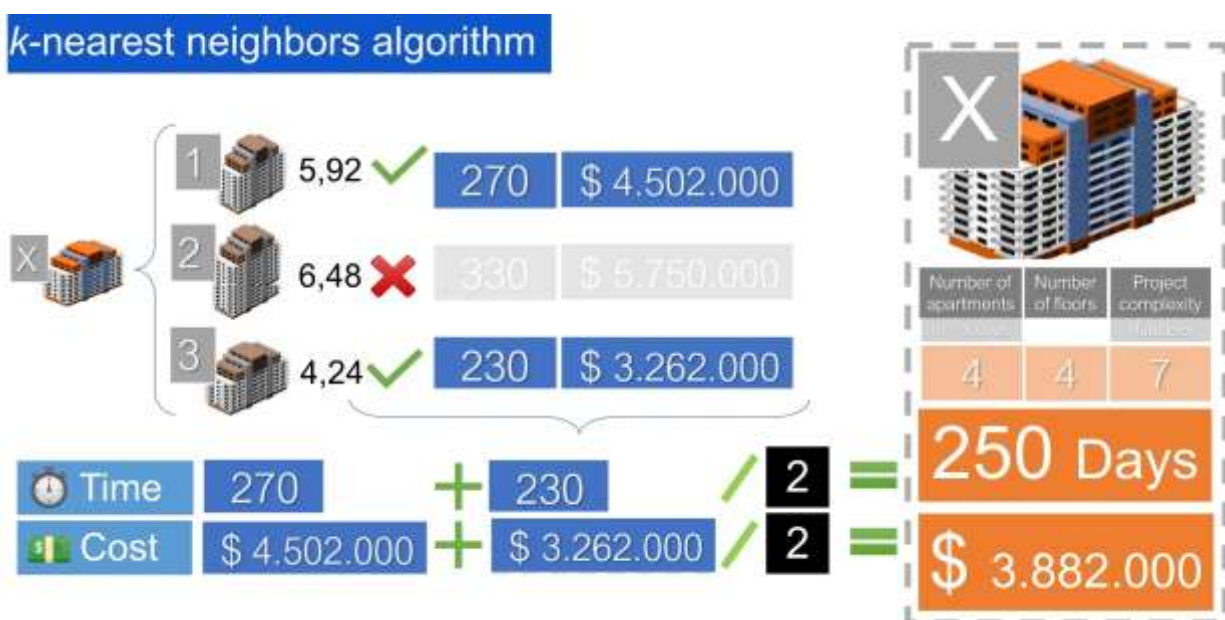
Joonis 9.3-5 K-NN algoritmi puhul esitatakse projektid punktidenä mitmemõõtmelises ruumis ning sarnasuse hindamiseks ja prognoosimiseks valitakse kauguste alusel välja lähimad projektid.

K-NN töö hõlmab mitmeid põhietappe:

- **Andmete ettevalmistamine:** kõigepealt laaditakse treening- ja testandmekogumid. Koolitusandmeid kasutatakse algoritmi "treenimiseks" ja testandmeid selle tõhususe kontrollimiseks.
- **Parameetri K valimine:** valitakse arv K, mis näitab, kui palju lähimaid naabreid (andmepunkte) tuleks algoritmis arvesse võtta. K väärtus on väga oluline, sest see mõjutab tulemust.
- Klassifitseerimisprotsess ja regressioon katseandmete puhul:
 - **Kauguste arvutamine:** igale elemendile testandmetest arvutatakse kaugus iga elemendiga treeningandmetest (joonis 9.3-5). Selleks võib kasutada erinevaid kauguse mõõtmise meetodeid, näiteks eukleidilist kaugust (kõige levinum meetod), Manhattani kaugust või Hammingi kaugust.
 - **Sorteerimine ja K lähima naabri valimine:** pärast vahemaade arvutamist sorteeritakse need ja valitakse K testitava punktile lähimat punkti.
 - **Katsepunkti klassi või väärtuse määramine:** kui tegemist on klassifitseerimisülesandega, määratakse katsepunkti klass K valitud naabrite kõige sagedasema klassi alusel. Kui tegemist on regressioonülesandega, arvutatakse K naabri väärtuste keskmine (või muu kesktendentsuse mõõt).
- **Protsessi lõpuleviimine:** kui kõik katseandmed on klassifitseeritud või nende kohta on tehtud prognoosid, on protsess lõpule viidud.

Algoritm k-naabrid (k-NN) on tõhus paljudes praktilistes rakendustes ja on üks peamisi vahendeid masinõppe spetsialistide arsenalis. See algoritm on populaarne oma lihtsuse ja tõhususe tõttu, eriti ülesannetes, kus andmete vahelisi seoseid on lihtne tõlgendada.

Meie näites tuvastati pärast K-lähedaseimate naabrite algoritmi rakendamist kaks projekti (meie väikesest valimist), mille kaugus projektist X oli kõige lühem (joonis 9.3-5). Nende projektide põhjal määrab algoritm nende hinna ja ehitusaja keskmise. Pärast analüüsi (joonis 9.3-6) järeldab algoritm lähimate naabrite keskmistamise teel, et projekt X maksab ligikaudu \$ 3 800 000 dollarit ja võtab aega umbes 250 päeva.



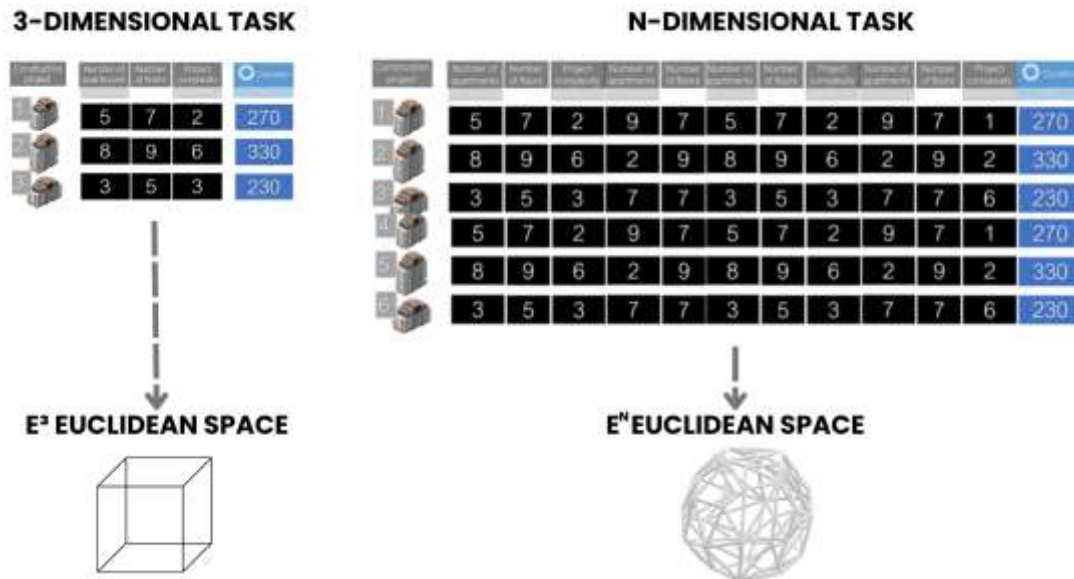
Joonis 9.3-6 Algoritm K-lähedaseimad naabrid määrab projekti X maksumuse ja ajakava, analüüsid kahte lähimat projekti valimis.

Algoritm k-Nearest Neighbors (k-NN) on eriti populaarne klassifitseerimis- ja regressioonülesannetes, näiteks soovitusüsteemides, kus seda kasutatakse toodete või sisu soovitamiseks, mis põhineb konkreetse kasutaja huvidega sarnastel eelistustel. Lisaks kasutatakse k-NNi laialdaselt meditsiinidiagnostikas, et klassifitseerida haigustüüpe patsiendi sümptomite alusel, mustrituvastuses ja finantssektoris klientide krediitvõimelisuse hindamiseks.

Isegi piiratud andmete puhul võivad masinõppe mudelid anda kasulikke prognoose ja parandada oluliselt ehitusprojektide juhtimise analüütilist komponenti. Kui ajaloolisi andmeid laiendatakse ja puhastatakse, on võimalik liikuda keerukamate mudelite suunas - näiteks võttes arvesse ehitustüüpi, asukohta, ehituse alguse hooaega ja muid tegureid.

Meie lihtsustatud ülesanne kasutas 3D-ruumis visualiseerimiseks kolme atribuuti, kuid reaalsed projektid sisaldavad keskmiselt sadu või tuhandeid atribuute (vt andmestik peatükist "Näide CAD-

põhistest suurtest andmetest (BIM)"), mis suurendab oluliselt ruumi dimensioonilisust ja projektide vektoritena kujutamise keerukust (joonis 9.3-7).



Joonis 9.3-7 Lihtsustatud näites kasutati 3D -visualiseerimiseks kolme atribuuti, samas kui reaalses projektides on neid rohkem.

Erinevate algoritmide rakendamine samale andmekogumile projekti X puhul, millel on 40 korterit, 4 korrust ja keerukusaste 7, andis erinevaid prognoosituid väärtusi. Lineaarse regressiooni algoritm ennustas valmimisaja pikkuseks 238 päeva ja maksumuseks \$ 3 042 338 (joonis 9.3-4), samas kui k-NN algoritm ennustas 250 päeva ja \$ 3 882 000 (joonis 9.3-6).

Masinõppe mudelite abil saadud prognooside täpsus sõltub otseselt sisendandmete mahust ja kvaliteedist. Mida rohkem projekte on kaasatud koolitusse ning mida täielikumalt ja täpsemalt on esindatud nende omadused (atribuudid) ja tulemused (sildid), seda suurem on tõenäosus saada usaldusväärseid prognoose minimaalsete veaväärtustega.

Selles protsessis on oluline roll andmete eeltöötlusmeetoditel, sealhulgas:

- Normaliseerimine, et viia tunnused ühisele skaalale;
- Kõrvalekalde tuvastamine ja kõrvaldamine, mudeli moonutamise kõrvaldamine;
- Kategooriliste atribuutide kodeerimine, et võimaldada tekstiandmetega manipuleerimist;
- Puuduvate väärtuste täitmine, mis suurendab mudeli usaldusväärsust.

Lisaks kasutatakse ristvalideerimismeetodeid, et hinnata mudeli üldistatavust ja selle vastupidavust uutele andmekogumitele, et tuvastada liigset kohandamist ja parandada prognoosi usaldusväärsust.

Kaos on korraldus, mida tuleb dešifreerida [160].

- José Saramago, "Topelt"

Isegi kui teile tundub, et teie ülesannete kaos ei ole formaalselt kirjeldatav, peaksite teadma, et mis tahes sündmused maailmas ja eriti ehitusprotsessid alluvad matemaatilistele seadustele, mis võivad vajada tuge väärtuste arvutamisel mitte rangete valemite, vaid statistika ja ajalooliste andmete abil.

Nii traditsioonilise kuluarvestusosakonna hinnangud kui ka masinõppe mudelid seisavad paratamatult silmitsi ebakindluse ja potentsiaalsete veaallikatega. Kui on olemas piisavalt kvaliteetsed andmed, võivad masinõppe mudelid siiski näidata võrreldavat ja mõnikord isegi suuremat ennustustäpsust kui ekspertide hinnangud.

Masinõppest saab tõenäoliselt usaldusväärne lisavahend analüüside tegemiseks, mis võimaldab: täpsustada arvutusi, pakkuda alternatiivseid stsenaariume ja tuvastada varjatud sõltuvusi projekti parameetrite vahel. Sellised mudelid ei pretendeeri universaalsusele, kuid varsti võtavad nad arvutustes ja otsustusprotsessides olulise koha. Masinõppe tehnoloogiad ei välista inseneride, kalkulaatorite ja analüütikute osalemist, vaid vastupidi, laiendavad nende võimalusi, pakkudes täiendavat vaatenurka, mis põhineb ajaloolistel andmetel.

Kui masinõppe integreeritakse korralikult ehitusettevõtete äriprotsessidesse, võib sellest saada oluline element juhtimisotsuste tugisüsteemis - mitte inimeste asendajana, vaid nende professionaalse intuitsiooni ja inseneriloogika laiendajana.

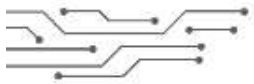
Järgmised sammud: ladustamisest analüüsi ja prognoosimiseni

Kaasaegsed lähenemisviisid andmetega töötamisele on hakanud muutma otsuste tegemise viisi ehitussektoris. Üleminek intuiitivetelt hinnangutelt objektiivsele andmeanalüüsile mitte ainult ei paranda täpsust, vaid avab ka uusi võimalusi protsesside optimeerimiseks. Selle osa kokkuvõtteks tasub välja tuua peamised praktilised sammud, mis aitavad teil käsitletud meetodeid oma igapäevastes ülesannetes rakendada:

- Jätkusuutliku ladustamisinfrastruktuuri loomine
 - ☐ Püüa ühendada erinevad dokumendid ja projektiandmed ühtseks tabeli-mudeliks, koondades põhiteabe ühte andmeraamistikku edasiseks analüüsiks.
 - ☐ Kasutage tõhusaid andmesalvestusformaate - näiteks CSV või XLSX asemel veerulisi formaate nagu Apache Parquet - eriti nende kogumite puhul, mida võidakse tulevikus kasutada masinõppe mudelite treenimiseks.
 - ☐ luua andmete versioonisüsteem, et jälgida muudatusi kogu projekti jooksul.
- Analüüsi- ja automatiseerimisvahendite rakendamine

- ☐ Alustage ajalooliste projektiandmete analüüsi - dokumentatsiooni, mudelite ja hinnangute alusel -, et tuvastada mustreid, suundumusi ja kõrvalekaldeid.
 - ☐ ETL-protsesside (Extract, Transform, Load) juhtimine andmete automaatseks laadimiseks ja ettevalmistamiseks.
 - ☐ Õppige, kuidas visualiseerida peamisi mõõdikuid, kasutades erinevaid tasuta Pythoni visualiseerimisraamatukogusid.
 - ☐ Alustada statistiliste meetodite ja juhuvalimi rakendamist, et saada representatiivseid ja korratavaid analüütilisi tulemusi.
- Andmetega töötamise küpsuse suurendamine
- ☐ Õppige ära mõned põhilised masinõppe algoritmid lihtsate ja lihtsate näidete abil, nagu Titanicu andmestik.
 - ☐ Analüüsige praeguseid protsesse ja tehke kindlaks, kus saab liikuda jäigast põhjustagajärg-loogikast statistiliste prognoosimis- ja hindamismeetodite juurde.
 - ☐ Andmete käsitlemine strateegilise vara, mitte kõrvalsaadusena: ehitage otsustusprotsessid pigem andmemudelite kui konkreetsete tarkvaralahenduste ümber.

Andmete väärtust mõistnud ehitusettevõtted on jõudmas uude arengufaasi, kus konkurentsieelise määrab mitte ressursside hulk, vaid analüütikal põhinevate otsuste tegemise kiirus.



MAKSIMAALNE MUGAVUS TRÜKITUD VERSIOONIGA

Sul on käes tasuta digitaalne versioon raamatust **Andmepõhine ehitus**. Mugavamaks tööks ja kiireks juurdepääsuks materjalidele soovitame pöörata tähelepanu **trükitud väljaandele**:



■ **Alati käepärast:** trükitud raamatust saab usaldusväärne töövahend, mis võimaldab teil igas töösituatsioonis kiiresti leida ja kasutada õigeid visualiseeringuid ja skeeme.

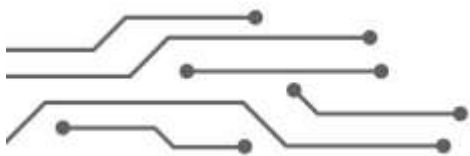
■ **Kvaliteetsed illustratsioonid:** kõik trükiväljaande pildid ja graafika on esitatud maksimaalse kvaliteediga.

■ **Kiire juurdepääs teabele:** lihtne navigeerimine, võimalus teha märkmeid, järjehoidjaid ja töötada raamatuga igal pool.

Ostes raamatu trükitud täisversiooni, saate mugava

abivahendi mugavaks ja tõhusaks tööks teabega: võimaluse kasutada visuaalseid materjale kiiresti igapäevastes ülesannetes, leida kiiresti vajalikud skeemid ja teha märkmeid. Lisaks toetab teie ostuga avatud teadmiste levitamist.

Tellige raamatu trükitud versioon aadressil: datadrivenconstruction.io/books



X OSA

EHITUSTÖÖSTUS DIGITAALAJASTUL. VÕIMALUSED JA VÄLJAKUTSED

Viimane kümnes osa on põhjalik ülevaade ehitussektori tulevikust digitaalse ümberkujundamise ajastul. Selles analüüsitakse üleminekut põhjuslikust analüüsist suurte andmekorrelatsioonidega töötamisele. Paralleelse tõmmatakse kujutava kunsti arengu ja andmetöötuse arengu vahel ehituses, näidates, kuidas tööstusharu liigub üksikasjalikult kontrollilt protsesside terviklikule mõistmisele. Uuritakse ehitussektori "uberiseerimise" kontseptsiooni, kus andmete läbipaistvus ja automatiseeritud arvutused võivad radikaalselt muuta traditsioonilisi ärimudeleid, kaotades vajaduse vahendajate järele ja vähendades spekulatsioonivõimalusi. Üksikasjalikult arutatakse lahendamata küsimusi, nagu näiteks elementide universaalne klassifitseerimine, mis annab ehitusettevõtetele aega uue keskkonnaga kohanemiseks. Osa lõpeb konkreetsete soovitusetega digitaalse ümberkujundamise strateegia kujundamiseks, mis hõlmab haavatavuste analüüsimist ja teenuste laiendamist, et jääda muutavas tööstusharus konkurentsivõimeliseks.

PEATÜKK 10.1.

ELLUJÄÄMISSTRATEEGIAD: KONKURENTSIEELISTE LOOMINE

Korrelatsioonid arvutuste asemel: ehitusanalüütika tulevik

Tänu teabe kiirele digitaliseerimisele (joonis 1.1-5) on kaasaegne ehitustegevus läbimas põhjalikku muutust, kus andmed muutuvad mitte lihtsalt tööriistaks, vaid strateegiliseks varaks, mis võib põhjalikult muuta traditsioonilisi lähenemisviise projekti- ja ärijuhtimisele.

Tuhandeid aastaid on ehitustegevus tuginenud deterministlikele meetoditele - täpsed arvutused, detailplaneeringud ja parameetrite range kontroll. Esimestel sajanditel pKr rakendasid Rooma insenerid matemaatilisi põhimõtteid akveduktide ja sildade ehitamisel. Keskajal püüdsid arhitektid gooti katedraalide ideaalsete proportsioonide poole ning 20. sajandi industrialiseerimise käigus kujunesid standardiseeritud normide ja eeskirjade süsteemid, mis said aluseks massiehitusele

Tänapäeval on arenguvektor nihkunud rangelt põhjus-tagajärg seoste otsimisest tõenäosusliku analüüsi, korrelatsioonide ja varjatud mustrite otsimise suunas. Tööstus on astumas uude faasi - andmed muutuvad peamiseks ressursiks ning nende põhjal tehtav analüüs asendab intuiitivseid ja kohalikke optimeeritud lähenemisviise.



Joonis 10.1-1 Ehitusandmete varjatud potentsiaal: ettevõttes olemasolevad arvutused on vaid jäämäe tipp, mida juhtkond saab analüüsida.

Ettevõtte infosüsteem on nagu jäämägi (joonis 10.1-1): ainult väike osa andmete potentsiaalst on ettevõtte juhtkonnale nähtav, samas kui peamine väärtus on peidetud sügavale. Oluline on hinnata andmeid mitte ainult nende praeguse kasutamise, vaid ka nende tulevaste võimaluste põhjal. Just need ettevõtted, kes õpivad, kuidas andmetest varjatud mustreid välja tuua ja uusi teadmisi luua, loovad püsiva konkurentsieelise

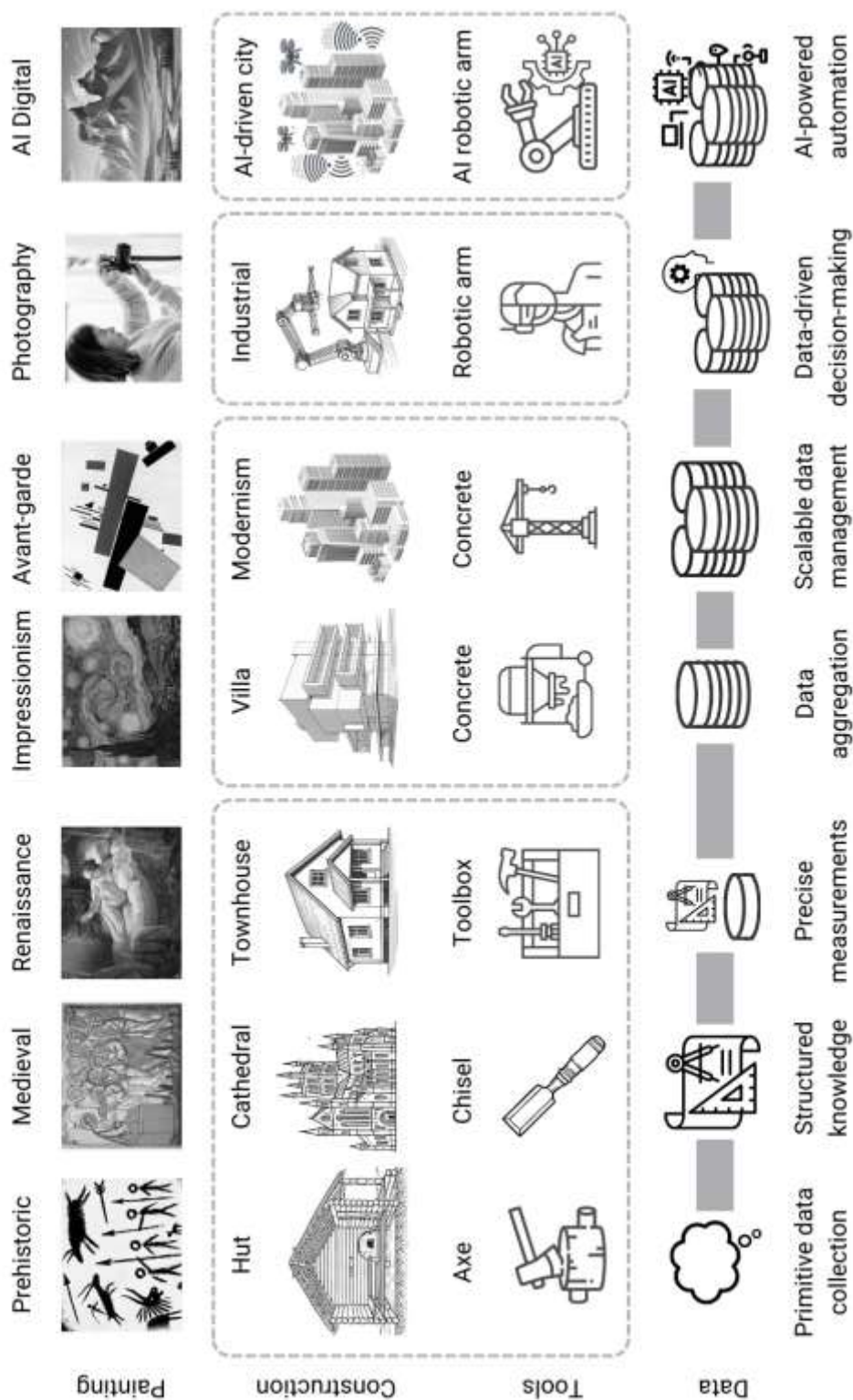
Varjatud mustrite leidmine ja andmete mõtestamine ei ole ainult töö numbritega, vaid loominguine protsess, mis nõuab abstraktset mõtlemist ja võimet näha tervikpilti erinevate elementide taga. Selles mõttes võib andmetega töötamise arengut võrrelda kujutava kunsti arenguga (joonis 10.1-2).

Ehituse areng on märkimisväärselt sarnane kujutava kunsti arenguga. Mõlemal juhul on inimkond arenenud primitiivsetelt meetoditelt keerukate visualiseerimis- ja analüüsitehnoloogiateni. Eelajaloolisel ajal kasutasid inimesed igapäevaste ülesannete lahendamiseks koopajooniseid ja primitiivseid tööriistu. Keskajal ja renessansiajal tõusis arhitektuuri ja kunsti keerukuse tase järsult. Keskaja alguseks olid ehitustööriistad arenenud lihtsast kirvest ulatuslikeks tööriistakomplektideks, mis sümboliseerivad tehniliste teadmiste kasvu.

Realismi ajastu oli esimene revolutsioon kujutavas kunstis: kunstnikud õppisid, kuidas reprodutseerida väiksemaid detaile, saavutades maksimaalse usutavuse. Ehituses oli selle perioodi vastandiks täpne inseneritehnika, üksikasjalikud joonised ja rangelt reguleeritud arvutused, mis said sajanditeks projekteerimistegevuse aluseks.

Hiljem muutis impressionism kunstilise reaalsuse tajumist: selle asemel, et sõna otseses mõttes kujutada vormi, hakkasid kunstnikud püüdma jäädvustada meeleolu, valgust ja dünaamikat, püüdes kajastada pigem üldmuljet kui absoluutset täpsust. Samamoodi liigub masinõpe ehitusanalüütikas jäikadest loogilistest mudelitest mustrituvastuse ja tõenäosuslike mustrite suunas, mis võimaldab "näha" andmetes varjatud sõltuvusi, mis on klassikalises analüüsis kättesaamatud. See lähenemisviis kajastab Bauhausi minimalismi ja funktsionaalsuse ideid, kus tähendus (funktsioon) on tähtsam kui vorm. Bauhaus püüdis eemaldada üleliigset, loobuda ornamentikast selguse, utilismi ja massihuvitavuse huvides. Asjad pidid olema arusaadavad ja kasulikud, ilma liigsetena - esteetika sündis disaini ja otstarbe loogikast.

Fotograafia tulekuga 19. sajandi lõpul sai kunst uue vahendi, mis võimaldas jäädvustada reaalsust enneolematu täpsusega ja pööras suhtumise visuaalsesse kunsti pea peale. Samamoodi on 21. sajandi tööstusrevolutsioon ehituses viinud robotikatehnoloogiate, laserite, asjade interneti, RFID ja selliste kontseptsioonide nagu Connected Construction, kus üksikute parameetrite kogumine on arenenud kogu ehitusplatsi tegelikkuse skaleeritavaks intelligentseks jäädvustamiseks, kasutusse.



Joonis 10.1-2 Era of Fine Art Evolutsioonid on kooskõlas ehitussektori andmetega töötamise lähenemisviiside .

Nii nagu visuaalkunstis toimub praegu uus mõtteviis koos tehisintellekti vahendite ja LLM tulekuga, on ehitustööstuses toimumas järjekordne kvanthüpe: tehisintellekti (AI) poolt juhitud intelligentsed süsteemid, LLM-chatid võimaldavad prognoosida, optimeerida ja genereerida lahendusi minimaalse inimsekkumisega.

Andmete roll projekteerimisel ja haldamisel on radikaalselt muutunud. Kui varem edastati teadmisi suuliselt ja empiirilisel - nii nagu reaalsust jäädvustati kuni 19. sajandini käsitsi maalitud piltidega -, siis tänapäeval keskendutakse ehitusliku "pildi" täielikule digitaalsele jäädvustamisele. Masinõppe algoritmide abil muudetakse see digitaalne pilt ehitustegelikkuse impressionistlikuks kujutiseks - mitte täpseks koopiaks, vaid üldistatud, tõenäosuslikuks arusaamaks protsessidest.

Me oleme kiiresti lähenemas ajastule, kus hoonete projekteerimise, ehitamise ja käitamise protsessid ei ole mitte ainult täiendatud, vaid suures osas juhitud tehisintellekti süsteemide poolt. Nii nagu kaasaegne digitaalne kunst luuakse ilma pintsliteta - kasutades tekstilisi juhiseid ja genereerivaid mudeleid -, kujundavad tuleviku arhitektuuri- ja insenerilahendusi kasutaja poolt seatud põhiküsimused ja parameetrid.

21. sajandil on juurdepääs andmetele, nende tõlgendamine ja analüütika kvaliteet muutumas projekti edukuse seisukohalt hädavajalikuks. Ja andmete väärtust ei määra mitte nende maht, vaid spetsialistide võime neid analüüsida, kontrollida ja muuta need tegevuseks.

Andmepõhine lähenemisviis ehituses: uus infrastruktuuritase

Inimkonna ajaloos on iga selline tehnoloogiline hüpe toonud kaasa põhimõttelisi muutusi majanduses ja ühiskonnas. Praegu oleme tunnistajaks uuele muutuste lainele, mis on oma ulatuselt võrreldav 19. sajandi tööstusrevolutsiooniga. Kuid kui sada aastat tagasi olid muutuste peamiseks tõukejõuks mehaanilised jõud ja energiatehnoloogiad, siis nüüd on selleks andmed ja tehisintellekt.

Masinõpe, LLM ja AI agendid muudavad rakenduste olemust, muutes traditsioonilised tarkvarapakid (mida käsitletakse raamatu teises osas) tarbetuks (joonis 2.2-3). Kogu andmeloogika on koondatud tehisintellekti agentidesse, mitte kõvasti kodeeritud ärieeskirjadesse (joonis 2.2-4).

Andmeajastul on traditsioonilised vaated rakendustele põhjalikult muutumas. Oleme liikumas mudeli suunas, kus mahukad, modulaarsed ettevõtlussüsteemid annavad paratamatult teed avatud, kergetele ja kohandatud lahendustele.

Tulevikus jääb alles ainult aluseks olev andmestruktuur ja kogu suhtlus sellega toimub otse andmebaasiga töötavate agentide kaudu. Ma tõesti usun, et kogu rakenduste virn kaob, sest seda pole lihtsalt vaja, kui tehisintellekt suhtleb otse aluseks oleva andmebaasiga. Ma olen kogu oma karjääri jooksul töötanud SaaS-i valdkonnas - ehitanud ettevõtteid, töötanud neis, ja kui aus olla, siis ma tõenäoliselt ei käivitaks praegu uut SaaS-i ettevõtet. Ja tõenäoliselt ei investeeriks ma praegu ka mitte SaaS-ettevõtetesse. Olukord on liiga ebakindel. See ei tähenda, et tulevikus ei ole tarkvarafirmasid, lihtsalt need näevad välja väga erinevad. Tuleviku süsteemid on andmebaasid, mille äriloogika on viidud [AI] agentidesse. Need agendid hakkavad töötama korraga mitme andmehoidlaga, mitte piirduma ühe andmebaasiga. Kogu loogika liigub tehisintellekti kihti [46].

- Matthew Berman, tegevjuht Forward Future

Uue paradigma peamine erinevus on tehnoloogilise ballasti minimeerimine. Monumentaalsete keeruliste ja suletud tarkvarasüsteemide asemel saame paindlikud, avatud ja kiiresti kohandatavad moodulid, mis sõna otseses mõttes "elavad" andmevoo sees (joonis 7.4-1 - Apache Airflow, NiFi). Tuleviku protsessijuhtimise arhitektuur näeb ette mikrorakenduste - kompaktsete, spetsiaalselt loodud vahendite - kasutamist, mis erinevad põhimõtteliselt massiivsetest ja suletud ERP-, PMIS-, CDE, CAFM-süsteemidest. Uued agendid on võimalikult kohanemisvõimelised, integreeritud ja ärispetsiifilised (nt Low-Code/No-Code joonis 7.4-6).

Kogu äriloogika läheb nendele [AI] agentidele ja need agendid teostavad CRUD [Create, Read, Update, and Delete] operatsioone mitme repositooriumiga, mis tähendab, et nad ei tee vahet, millist backend'i kasutatakse. Nad ajakohastavad mitut andmebaasi ja kogu loogika jõuab nn AI-tasandile. Ja kui kogu loogika on AI-kihis, hakkavad inimesed alustama backendide asendamist. Me näeme juba praegu üsna suurt turuosa Dynamics'i backendide ja agentide kasutamisel ning me liigume agressiivselt selles suunas, püüdes seda kõike kokku viia. Olgu see siis klienditeeninduses või muudes valdkondades, näiteks mitte ainult CRM, vaid ka meie finants- ja tegevuslahendustes. Sest inimesed tahavad rohkem tehisintellektipõhiseid ärirakendusi, kus loogikakihti saab juhtida tehisintellekti ja tehisintellekti agentide abil. [...] Üks põnevamaid asju minu jaoks on Excel koos Pythoniga, mis on võrreldav GitHubi ja Copiloti. Ma mõtlen, mida me oleme teinud: nüüd, kui teil on Excel, peaksite selle lihtsalt avama, käivitama Copiloti ja hakkama sellega mängima. See ei mõista enam lihtsalt olemasolevaid numbreid - see teeb ise plaani. Nii nagu GitHubi Copilot Workspace loob plaani ja seejärel täidab seda, on see nagu andmeanalüütik, kes kasutab Exceli tööriistana ridade ja veergude visualiseerimiseks analüüsi jaoks. Seega kasutab Copilot Exceli kui tööriista kõigi selle võimalustega, sest ta saab andmeid genereerida ja tal on Pythoni interpretaator.

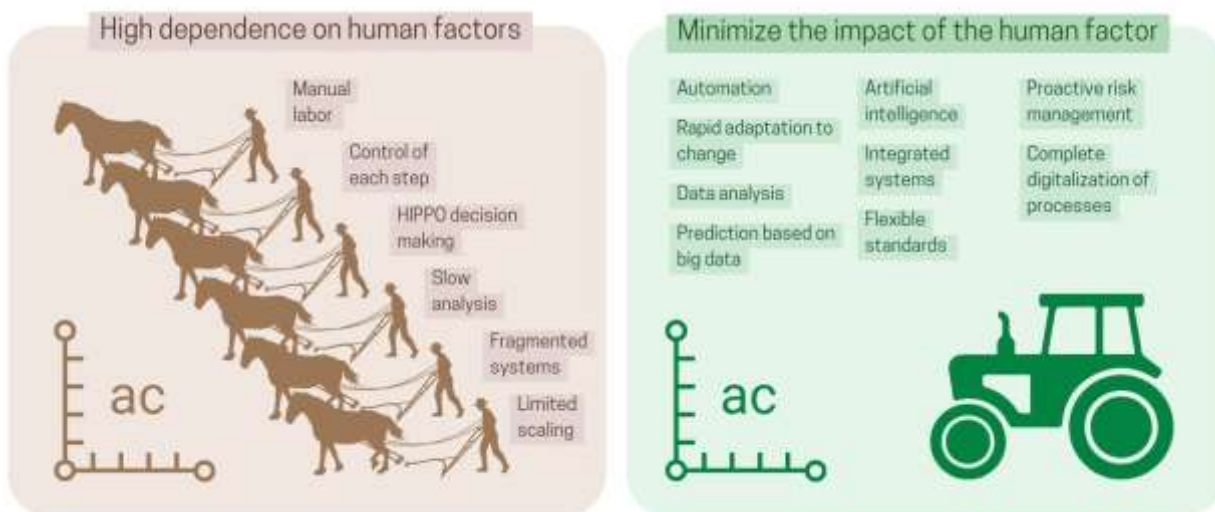
- Satya Nadella, Microsofti tegevjuht, intervjuu BG2 kanalile detsember 2024. [28]

Kontorirakenduste loogika muutumine, mida me praegu näeme - üleminek modulaarsetelt suletud süsteemidelt otse avatud andmetega töötavatele tehisintellekti agentidele - on vaid osa palju suuremast protsessist. Tegemist ei ole ainult liideste või tarkvaraarhitektuuri muutmisega: muutused

mõjutavad töökorralduse, otsuste tegemise ja ärijuhtimise aluspõhimõtteid. Ehituses viib see andmepõhise loogikani, kus andmed muutuvad protsesside keskmeks alates projekteerimisest kuni ressursside haldamise ja ehitusjärelvalveni.

Järgmise põlvkonna digitaalne kontor: kuidas tehisintellekt muudab tööruumi.

Peaaegu sajand tagasi oli inimkond juba sarnase tehnoloogilise revolutsiooni käes. Üleminek aumootoritelt elektrimootoritele kestis rohkem kui neli aastakümnet, kuid lõppkokkuvõttes katalüüsis see enneolematut tootlikkuse kasvu - peamiselt energiamahu detsentraliseerimise ja uute lahenduste paindlikkuse kaudu. See üleminek ei muutnud mitte ainult ajaloo kulgu, viies suurema osa elanikkonnast maapiirkondadest linnadesse, vaid pani ka aluse kaasaegsele majandusele. Tehnoloogia ajalugu on teekond füüsilisest tööjõust automatiseerimise ja intelligentsete süsteemide suunas. Nii nagu traktor asendas hulgaliselt mullaharijaid, asendab kaasaegne digitaaltehnoloogia traditsioonilisi kontoripõhiseid ehitusjuhtimismeetodeid (joonis 10.1-3). Veel 20. sajandi alguses töötas suurem osa maailma elanikkonnast maad käsitsi, kuni 1930. aastatel algas töö mehhaniseerimine masinate ja traktoritega.



Joonis 10.1-3 Nii nagu traktor asendas 20. sajandi alguses kümneid inimesi, asendab masinõpe 21. sajandil traditsioonilised äri- ja projekti juhtimise meetodid.

Nii nagu inimkond liikus sada aastat tagasi üksikute maatükkide harimisest primitiivsete tööriistade abil suuremahulise põlluharimise juurde, kasutades masinaid, oleme tänapäeval üleminekul teabe eraldiseisvate "silode" töötlemisest andmete massiividega töötamisele võimsate "traktorite" - ETL-pipelini ja tehisintellekti algoritmide abil.

Me oleme sarnase hüppe lävel - kuid digitaalsel tasandil: traditsioonilisest, manuaalsest ärijuhtimisest andmepõhiste mudelite juurde.

Täieliku andmepõhise arhitektuurini jõudmine nõuab aega, investeringuid ja organisatsioonilisi jõupingutusi. Kuid see tee avab tee mitte ainult järkjärgulisele paranemisele, vaid kvalitatiivsele hüppele ehitusprotsesside suurema tõhususe, läbipaistvuse ja juhitavuse suunas. Kõik see eeldab

digitaalsete vahendite süstemaatilist kasutuselevõttu ja vananenud äritavade hülgamist.

Ülesannete parameetriseerimine, ETL, LLM, asjade interneti komponendid, RFID, tokeniseerimine, suured **andmed** ja masinõpe muudavad traditsioonilise ehituse **andmepõhiseks ehituseks**, kus projekti ja ehitustegevuse iga detail kontrollitakse ja optimeeritakse andmete abil.

Varem kulus teabe analüüsimiseks tuhandeid töötunde. Nüüd täidavad neid ülesandeid algoritmid ja LLM-id, mis muudavad erinevad andmekogumid strateegilisteks allikateks, kasutades käsklusi. Tehnoloogiamaailmas toimub sama, mis juhtus põllumajanduses: me liigume kobestamiselt automatiseeritud agrotööstusele. Nii liigub ka kontoritöö ehituses - Exceli failidest ja käsitsi kokkuvõtete tegemisest - intelligentssesse süsteemi, kus andmeid kogutakse, puhastatakse, struktureeritakse ja muudetakse arusaamadeks.

Juba täna peaksid ettevõtted alustama infoväljade "kasvatamist" kvaliteetsete andmete kogumise ja teabe struktureerimise abil ning "väetama" neid puhastamis- ja normaliseerimisvahenditega, et seejärel "korjata" neid prognoosiva analüüsi ja automatiseeritud lahenduste abil. Kui kaasaegne põllumees suudab masinaga asendada sada mullaharijat, siis intelligentsed algoritmid suudavad võtta töötajatelt rutiini ja anda neile infovoogude strateegilise juhi rolli.

Siiski on oluline mõista, et tõeliselt andmepõhise -organisatsiooni loomine ei ole kiire protsess. See on pikaajaline strateegiline suund, mis sarnaneb uue kasvukoha loomisega süsteemide metsa kasvatamiseks (joonis 1.2-5), kus iga "puu" selles ökosüsteemis on eraldi protsess, pädevus või vahend, mille kasvamiseks ja arendamiseks kulub aega. Nagu tõelise metsa puhul, ei sõltu edu ainult istutusmaterjali kvaliteedist (tehnoloogia), vaid ka pinnasest (ettevõtte kultuur), kliimast (ettevõtluskeskkond) ja hooldusest (süsteemne lähenemine).

Ettevõtted ei saa enam toetuda ainult suletud, valmislahendustele. Erinevalt varasematest tehnoloogilise arengu etappidest ei toeta suured müüjad tõenäoliselt praegust üleminekut avatud andmetele, tehisintellektile ja avatud lähtekoodile, sest see ohustab otseselt nende väljakujunenud ärimudeleid ja põhilisi tuluallikaid.

Nagu selgub Harvard Business School'i uuringust [40], mida on juba käsitletud neljandat ja viiendat tehnoloogilist revolutsiooni käsitlevas peatükis, oleks kõige enam kasutatavate avatud lähtekoodiga lahenduste loomine nullist kõigi ettevõtete jaoks umbes 4,15 miljardit dollarit. Kui aga kujutame ette, et iga ettevõtte arendaks oma alternatiivid ise välja ilma juurdepääsuta olemasolevatele avatud lähtekoodiga vahenditele, mis on viimastel aastakümnetel toimunud, siis võivad ettevõtte kogukulud ulatuda kolossaalsesse 8,8 triljoni dollarini - see on irratsionaalse nõudluse hind, mida tarkvaraturul võib hinnata.

Tehnoloogia areng toob paratamatult kaasa väljakujunenud ärimudelite ümbermõtestamise. Kui varem suutsid ettevõtted teenida raha keeruliste, läbipaistmatute protsesside ja suletud andmete abil, siis tehisintellekti ja analüütika arenguga muutub selline lähenemine üha vähem elujõuliseks.

Andmetele ja vahenditele juurdepääsu demokratiseerimise tulemusena võib traditsiooniline tarkvara müügiturg oluliselt kahaneda. Samal ajal kasvab aga uus turg - digitaalse ekspertiisi, kohandamise, integreerimise ja lahenduste väljatöötamise turg. Siin ei tule väärtus mitte litsentsimüügist, vaid võimest luua paindlikke, avatud ja kohandatavaid digitaalseid protsesse. Nii nagu elektrifitseerimine ja traktorite tulek on loonud uusi tööstusharusid, avab ka suurte andmete, tehisintellekti ja elukestva õppe kasutamine ehitussektoris täiesti uusi ärivõimalusi, mis nõuavad mitte ainult tehnoloogilisi investeeringuid, vaid ka mõtteviisi, protsesside ja organisatsiooniliste struktuuride põhjalikku muutmist. Ja need ettevõtted ja spetsialistid, kes seda mõistavad ja hakkavad tegutsema juba täna, on homsed liidrid.

Maailmas, kus avatud andmed on muutumas oluliseks varaks, on teabe kättesaadavus muutnud olukorda. Investorid, kliendid ja reguleerivad asutused nõuavad üha enam läbipaistvust ning masinõppe algoritmid suudavad automaatselt tuvastada lahknevusi hinnangutes, tähtaegades ja kuludes. See loob tingimused digitaalsete muutuste uueks etapiks, mis viib meid järk-järgult ehitussektori "uberiseerimiseni".

Avatud andmed ja Uberiseerimine on oht olemasolevale ehitusettevõtluks

Ehitus on muutumas teabehaldusprotsessiks. Mida täpsemad, kvaliteetsemad ja täielikumad on andmed, seda tõhusam on ehitiste projekteerimine, arvutused, kulukalkulatsioonid, ehitamine ja kasutamine. Tulevikus ei ole põhiressurs mitte kraana, betoon ja betoonraud, vaid võime koguda, analüüsida ja kasutada teavet.

Tulevikus kasutavad ehitusettevõtete kliendid - investorid ja ehitust rahastavad kliendid - paratamatult avatud andmete ja analüütika väärtust ajaloolisi andmeid. See avab võimalused projektide tähtaegade ja kulude arvutamise automatiseerimiseks, ilma et ehitusettevõtteid tuleks kaasata kuluarvestuse küsimustesse, mis aitab kulusid kontrollida ja üleliigseid kulusid kiiremini tuvastada.

Kujutage ette ehitusplatsi, kus laserskannerid, kvadroskoopid ja fotogrammetrilised süsteemid koguvad täpseid reaalajas andmeid kasutatud betooni koguste kohta. See teave teisendatakse automaatselt lihtsateks lamedateks MESH -mudeliteks koos metaandmetega, vältides kohmakaid CAD (BIM) süsteeme, ilma et see sõltuks keerulistest geomeetristest tuumadest, ERP või PMIS. Need ehitusplatsilt kogutud andmed edastatakse tsentraalselt ühtsesse struktureeritud hoidlasse, mis on kliendile sõltumatuks analüüsiks kättesaadav, kuhu laaditakse üles erinevate ehituspoodide tegelikud hinnad ja parameetrid alates krediidi rahastamise määrast kuni dunaamiliselt muutuvate teguriteni, nagu ilmastikutingimused, ehitusmaterjalide börsikursid, logistikatariifid ja tööjõuhindade statistilised hooajalised kõikumised. Sellistes tingimustes muutuvad kõik lahknevused projekteeritud ja tegelike materjalimahtude vahel kohe ilmseks, mistõttu on võimatu manipuleerida hinnangutega nii projekteerimise etapis kui ka projekti üleandmisel. Selle tulemusena saavutatakse ehitusprotsessi läbipaistvus mitte järelevalvajate ja juhtide armee, vaid objektiivsete digitaalsete andmete abil, mis

vähendab inimtegurit ja spekulatsioonide võimalust.

Tulevikus hakkavad sellist andmekontrolli tööd tegema pigem kliendipoolsed andmehaldurid (joonis 1.2-4 CQMS manager). See kehtib eelkõige kalkulatsioonide ja projekti hinnangute kohta: seal, kus varem oli terve osakond kalkulaatoreid, on juba homme olemas masinõppe ja prognoosimisvahendid, mis seavad ehitusettevõtetele hinnapiirangud, millesse nad peavad mahtuma.

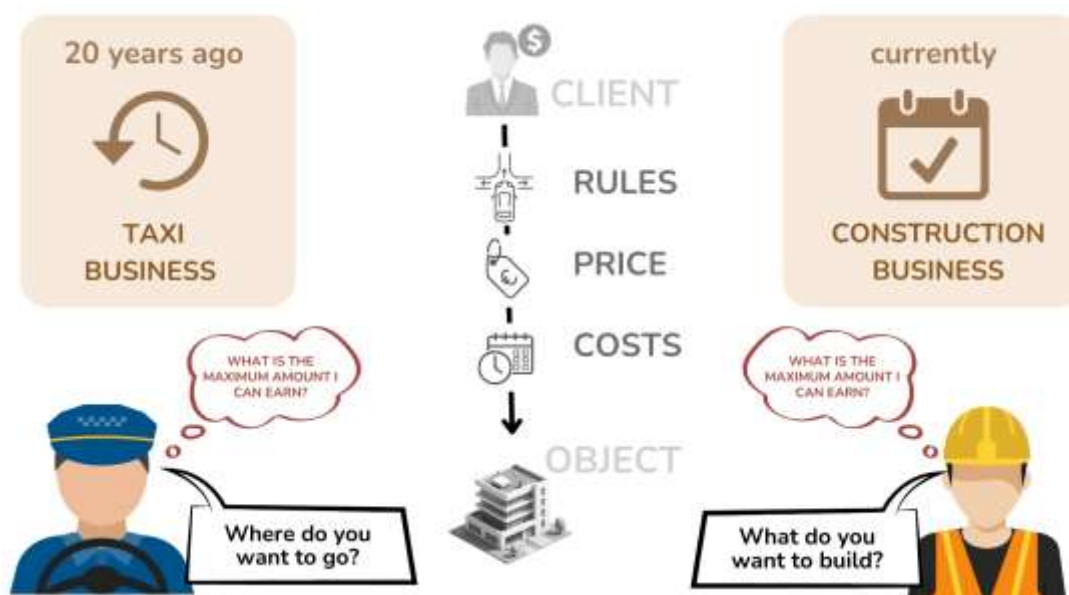
Arvestades [ehitus]tööstuse killustatust, kus enamiku süsteemidest ja allsüsteemidest tarnivad VKEd, peab digitaalstrateegia tulema kliendilt. Kliendid peavad looma tingimused ja mehhanismid, et avada tarneahela digitaalsed võimalused [20].

- Andrew Davis ja Giuliano Denicol, Accenture "Suurema väärtuse loomine kapitaliprojektide kaudu"

Selline andmete avatus ja läbipaistvus kujutab endast ohtu ehitusettevõtetele, kes on harjunud teenima raha läbipaistmatute protsesside ja segaste aruannete abil, kus spekulatsioonid ja varjatud kulud võivad olla peidetud keeruliste ja suletud vormingute ja modulaarsete konfidentsiaalsete andmeplatvormide taha. Seetõttu ei ole ehitusettevõtted, nagu ka avatud lähtekoodiga lahendusi propageerivad müüjad, tõenäoliselt huvitatud avatud andmete täielikust rakendamisest oma äriprotsessides. Kui andmed on kättesaadavad ja kliendi jaoks hõlpsasti töödeldavad, saab neid automaatselt kontrollida, mis välistab mahu ülehindamise ja kalkulatsioonide manipuleerimise võimaluse.

Maaailma Majandusfoorumi aruande "Shaping the Future of Construction" (2016) [5] kohaselt on valdkonna üheks peamiseks väljakutseks endiselt kliendi passiivne roll. Sellegipoolest peaksid just kliendid võtma suurema vastutuse projektide tulemuste eest - alates varajasest planeerimisest, jätkusuutlike suhtlusmodelite valimisest kuni tulemuslikkuse jälgimiseni. Ilma projektiomanike aktiivse osaluseta on ehitussektori süsteemne ümberkujundamine võimatu.

Kontrolli kaotamine mahu ja kulude arvutamise üle on juba viimase 20 aasta jooksul muutnud teisi tööstusharusid, võimaldades klientidel otse, ilma vahendajateta, oma eesmärgi kinni pidada. Digitaliseerimine ja andmete läbipaistvus on muutnud paljusid traditsioonilisi ärimudeleid, näiteks taksojuhid Uberi tekkimisega (joonis 10.1-4), hotelliettevõtjad Airbnb tulekuga ning jaemüüjad ja kauplused Amazoni tõusuga ning pangad neopankade ja detsentraliseeritud fintech-ökosüsteemide tekkimisega, kus otsene juurdepääs teabele ning aja- ja kuluarvestuse automatiseerimine on vahendajate rolli oluliselt vähendanud.



Joonis 10.1-4 Ehitusäri seisab silmitsi samasuguse uberiseerimisega, millega taksojuhid, hotelliettevõtjad ja jaemüüjad pidid silmitsi seisma 10 aastat tagasi.

Andmete ja nende töötlemise vahenditele juurdepääsu demokratiseerimise protsess on vältimatu ning aja jooksul muutuvad avatud andmed kõikide projektkomponentide kohta kliendi nõudeks ja uueks standardiks. Seetõttu edendavad avatud vormingute ja läbipaistvate arvutuste kasutuselevõtu küsimusi investorid, kliendid, pangad ja erakapitali investeerimisfondid (erakapitali investeerimisfondid) - need, kes on ehitatud objektide lõppkasutajad ja seejärel kasutavad objekti aastakümneid.

Suurinvestorid, kliendid ja pangad nõuavad juba praegu ehitussektoris läbipaistvust. Accenture'i uuringu "Creating more value through capital projects" (2020) [20] kohaselt on läbipaistvad ja usaldusväärsed andmed muutumas otsustavaks teguriks investeerimisotsuste tegemisel ehituses. Nagu eksperdid rõhutavad, on usaldusväärne ja tõhus projektijuhtimine võimatu ilma läbipaistvuseta, eriti kriisiaegadel. Lisaks sellele on varade omanikud ja töövõtjad üha enam liikumas lepingute suunas, mis stimuleerivad andmete jagamist ja ühist analüüsi, kajastades investore, pankade ja reguleerivate asutuste kasvavaid nõudmisi aruandekohustuse ja läbipaistvuse järele.

Investori, tellija liikumine ideest kuni valmishituseeni sarnaneb tulevikus autopiloodiga sõitmisega - ilma juhita ehitusfirma näol lubab muutuda sõltumatuks spekulatsioonidest ja ebakindlusest.

Avatud andmete ja automatiseerimise ajastu muudab paratamatult ka ehitustegevust, nagu see on juba teinud panganduses, kaubanduses, põllumajanduses ja logistikas. Nendes tööstusharudes on vahendajate roll ja traditsioonilised äritegevuse viisid loovutamas teed automatiseerimisele ja robotiseerimisele, mis ei jäta ruumi põhjendamatutele hinnalisanditele ja spekulatsioonidele.

Andmed ja protsessid kõigis inimkonna majandustegevustes ei erine sellest, millega ehitussektori spetsialistid peavad tegelema. Pikemas perspektiivis võivad ehitusettevõtted, mis täna domineerivad turul hinna ja teenuse kvaliteedistandardite kehtestamise kaudu, kaotada oma rolli peamise vahendajana kliendi ja tema ehitusprojekti vahel.

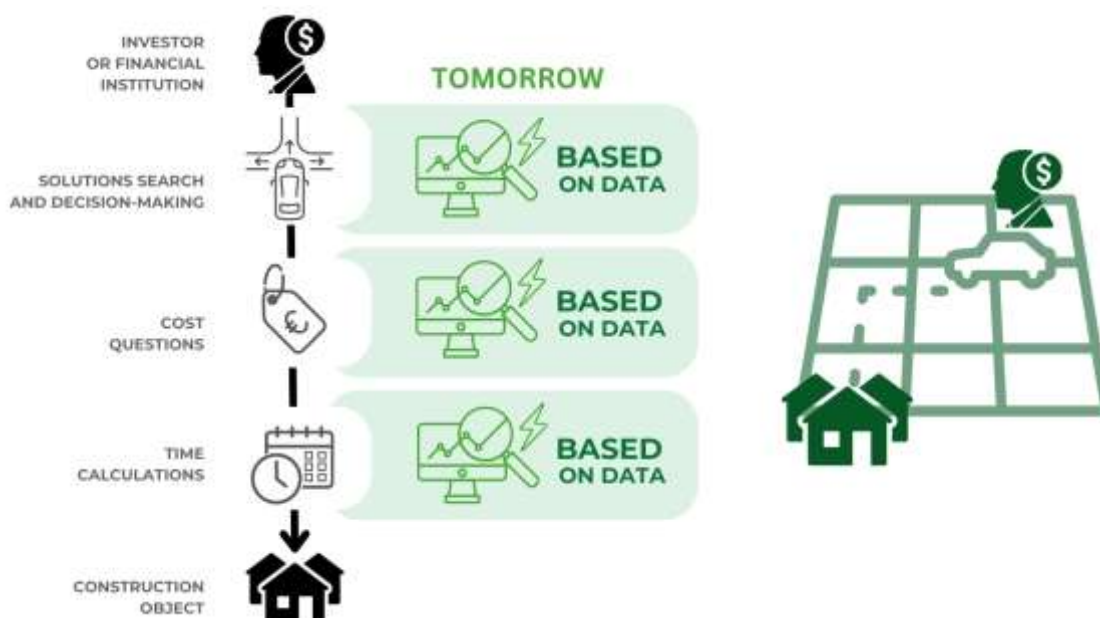
Ümberkorralduste lahendamata probleemid kui viimane võimalus kasutada aega ümberkujundamiseks

Kuid pöördume tagasi ehitussektori tegelikkuse juurde. Kuigi mõnes majandussektoris on tekkimas isesõitvad autod, detsentraliseeritud finantssüsteemid ja tehisintellektil põhinevad lahendused, on märkimisväärne osa ehitusettevõtetest endiselt paberipõhised organisatsioonid, kus olulised otsused tehakse pigem üksikute spetsialistide intuitsiooni ja kogemuste põhjal.

Selles paradigmas võib kaasaegset ehitusettevõtet võrrelda 20-aastase taksofirmaga, mis kontrollib ressursse, marsruute ja tarneaegu ning vastutab "teekonna" ajakava ja kulude eest - alates projekti ideest (logistika ja paigaldusprotsess) kuni projekti üleandmiseni. Nii nagu GPS (ehituses asjade interneti, RFID) ja masinõppe algoritmid aja/kulude arvutamisel muutsid kunagi transpordivaldkonda, on andmetel, algoritmidel ja tehisintellektil -agentidel potentsiaali muuta ehitusjuhtimist - intuiitivsetest hinnangutest prognoosivate, juhitud mudeliteni. Viimase 20 aasta jooksul on paljudes tööstusharudes - rahandus, põllumajandus, jaemüük ja logistika - järk-järgult kadunud võime spekuloida andmete läbipaistmatuse kaudu. Hinnad, tarnekulud või finantstehingud arvutatakse automaatselt ja statistiliselt usaldusväärselt - vaid mõne sekundiga digitaalsetel platvormidel.

Tulevikku vaadates peavad ehitusettevõtted tunnistama, et andmete ja nende analüüsimiseks vajalike vahendite demokratiseerimine häirib traditsioonilist lähenemisviisi projekti kulude ja ajakava hindamisele ning kaotab võimaluse spekuloida läbipaistmatute mahu- ja hinnaandmete põhjal.

Nagu reguleeritud teel sõitmine ilma juhi sekkumiseta, sarnanevad tuleviku ehitusprotsessid üha enam "Uberiseeritud" süsteemile - automatiseeritud aja- ja kuluhinnangute, läbipaistva ülesannete marsruutimise ja minimaalse sõltuvusega inimteguritest. See muudab "teekonna" olemust ideest kuni teostuseni - see muudab selle prognoositavamaks, juhitablemaks ja andmepõhisemaks.



Joonis 10.1-5 Kulud ja ehitusaegne "sõiduaeg" määratakse kindlaks masinõppe ja statistiliste vahendite abil.

Seoses uute eeskirjade ja nõuete järkjärgulise kehtestamisega peaaegu kõigis maailma riikides, mis kohustavad CAD- (BIM)-mudelite edastamist klientidele või ehitusprojekte rahastavatele pankadele, on kliendil ja tellijal võimalus tagada iseseisvalt kulude ja mahu arvutuste läbipaistvus. See on eriti oluline suurte tellijate ja investorite jaoks, kellel on piisavalt pädevust ja vahendeid mahtude kiireks analüüsimiseks ja turuhindade jälgimiseks. Suurte standardprojektide - kauplused, büroohooned, elamukompleksid - teostavate ettevõtete puhul on selline praktika muutumas standardiks.

Kuna mudelite infosisu muutub täielikumaks ja standardiseeritumaks, kaob praktiliselt võimalus manipuleerimiseks ja spekulatsiooniks. Digitaalne ümberkujundamine muudab järk-järgult ehitussektori mängureegleid ning ettevõtted, kes ei kohandu nende muutustega, võivad sattuda tõsiste probleemide ette.

Suurenenud konkurents, tehnoloogilised häired ja kahanevad kasumimarginaalid võivad mõjutada ettevõtte jätkusuutlikkust. Kuna likviidsus on piiratud, pöörduv üha rohkem tööstusharu osalisi automatiseerimise, analüüsi ja andmetehnoloogiate poole, et parandada tõhusust ja protsesside läbipaistvust. Need vahendid muutuvad oluliseks ressursiks, et jääda muutuv majanduskeskkonnas konkurentsivõimeliseks.

Võib-olla ei tasuks oodata, kuni välised asjaolud sunnivad kiiresti tegutsema - palju tõhusam on alustada ettevalmistusi juba täna, tugevdades digipädevust, rakendades kaasaegseid lahendusi ja luues andmekeskse kultuuri.

Üks viimaseid peamisi tehnoloogilisi takistusi ehitussektori ulatuslikul digitaalsel ümberkujundamisel, mis mõjutab lähiaastatel iga ettevõtet, on ehitusprojektide elementide automaatse klassifitseerimise probleem.

Ilma usaldusväärse, täpse ja skaleeritava klassifitseerimiseta on võimatu luua alust täieulatuslikuks analüütikaks, protsesside automatiseerimiseks ja elutsükli haldamiseks, kasutades tehisintellekti ja prognoosivaid mudeleid. Kuni objektide klassifitseerimine sõltub endiselt kogenud spetsialistide - töödejuhatajate, projekteerijate, kalkulaatorite - käsitsi tõlgendamisest, on ehitussektoril veel võimalus. Seda aega saab kasutada selleks, et valmistuda paratamatuteks muutusteks: kasvavad nõudmised läbipaistvuse järele, vahendite ja andmete demokratiseerimine ning automaatsete klassifitseerimissüsteemide teke, mis muudavad radikaalselt mängureegleid.

Ehitusmaailma elementide automaatse klassifitseerimise ülesanne on oma keerukuselt võrreldav objektituvastusega mehitamata juhtimissüsteemides, mis on üks peamisi väljakutseid. Kujutleme mehitamata autot, mis sõidab punktist A punkti B (joonis 10.1-5). Praegused automaatsed juhtimissüsteemid on takerdunud lidarite ja kaamerate poolt tuvastatud objektide klassifitseerimise probleemi taha. Ei piisa sellest, et auto lihtsalt "näeb" takistust või maamärki, vaid ta peab suutma ära tunda, kas tegemist on jalakäija, liiklusmärgi või prügikastiga.

Sarnane põhimõtteline väljakutse seisab silmitsi kogu ehitussektori ees. Projekti elemendid - näiteks aknad, uksed või sambad - võivad olla dokumenteeritud, kujutatud CAD-mudelites, pildistatud ehitusplatsil või tuvastatud laserskaneerimisega saadud punktpilvedes. Kuid nende visuaalsest või ligikaudsest geomeetrisest tuvastamisest ei piisa tõeliselt automatiseeritud projekti juhtimissüsteemi loomiseks. On vaja tagada, et iga element on täpselt ja järjepidevalt liigitatud tüübiks, mis on üheselt tuvastatav kõikides järgnevatel protsessides - alates kalkulasioonidest ja spetsifikatsioonidest kuni logistika, varude kontrollimise ja, mis kõige tähtsam, operatsioonideni (joonis 4.2-6).

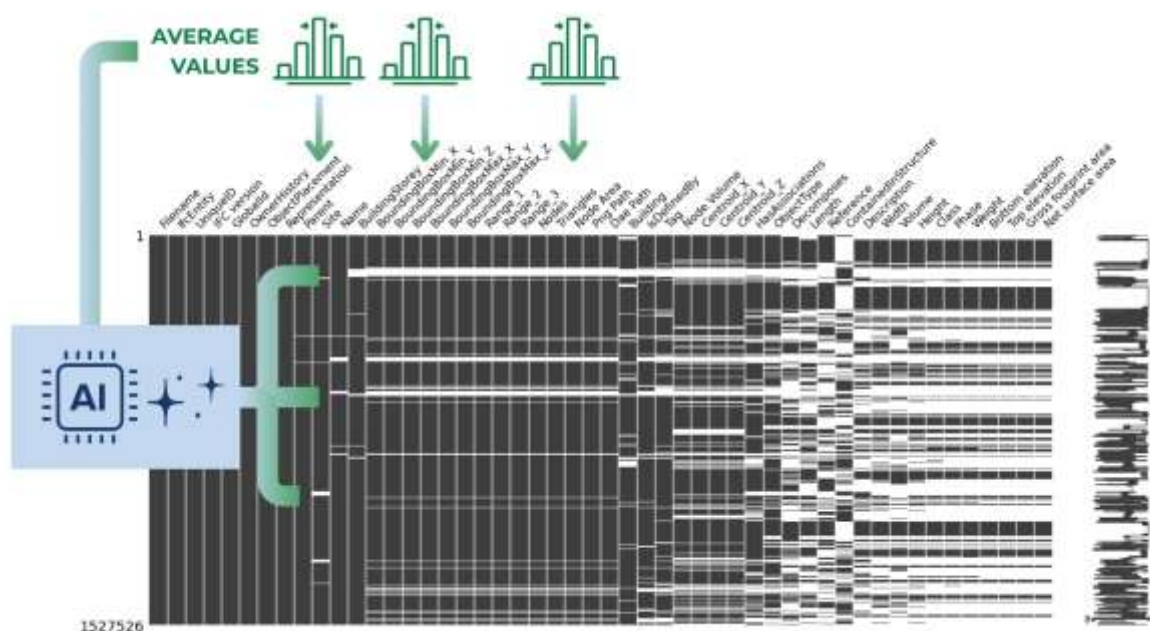
Just selles etapis - üleminekul äratundmisest sisulisele liigitusele - tekib üks peamisi takistusi. Isegi kui digitaalsüsteemid on tehniliselt võimelised eristama ja tuvastama objekte mudelites ja ehitusplatsil, seisneb peamine raskus elementide tüübi õiges ja kontekstiliselt stabiilses määratlemises erinevate tarkvarakeskkondade jaoks. Näiteks võib projekteerija märgistada ukse CAD-mudelis kategooriasse "uks" kuuluvaks elemendiks, kuid ERP- või PMIS-süsteemi ülekandmisel võib see olla valesti kirjutatud, kas projekteerija vea või süsteemide vaheliste vastuolude tõttu. Lisaks kaotab element sageli mõned olulised atribuudid või kaob andmete eksportimisel ja importimisel süsteemi raamatupidamisest. See toob kaasa lünga andmevoos ja õõnestab ehitusprotsesside läbiva digitaliseerimise põhimõtet. See tekitab kriitilise lõhe "nähtava" ja "arusaadava" semantilise tähenduse vahel, mis õõnestab andmete terviklikkust ja raskendab oluliselt protsesside automatiseerimist kogu ehitusprojekti elutsükli jooksul.

Suurandmete ja masinõppe tehnoloogiate (joonis 10.1-6) abil ehitusdetailide universaalse klassifitseerimise probleemi lahendamine on kogu tööstusharu muutuste katalüsaatoriks - ja võib-olla ootamatu avastus paljude ehitusettevõtete jaoks. Ühtne, õpitav klassifitseerimissüsteem on aluseks skaleeritavale analüütikale, digitaalsele juhtimisele ja tehisintellekti kasutuselevõtule ehitusorganisatsioonide igapäevastes tavades.

NVIDIA ja teised tehnoloogiajuhid pakuvad juba praegu teistes tööstusharudes lahendusi, mis suudavad automaatselt kategoriseerida ja struktureerida tohutul hulgal tekstilist ja visuaalset teavet.

Näiteks NVIDIA NeMo Curator'i mudel [161] on spetsialiseerunud andmete automaatsele klassifitseerimisele ja kategoriseerimisele etteantud kategooriatesse, mängides võtmerolli infotöötluspieliinide optimeerimisel, et peenhäälestada ja eelkoolitada genereerivaid tehisintellekti mudeleid. Cosmos platvormi treenitakse reaalse videote ja 3D stseenide [162] põhjal, mis annab aluse autonoomsetele süsteemidele ja digitaalsetele kaksikutele, mida NVIDIA ökosüsteemis juba ehitatakse. NVIDIA Omniverse, millest on 2025. aastaks saanud juhtiv vahend USD-vorminguga töötamiseks, mis on universaalne stseeni kirjeldus, mis võib lõpuks asendada IFC-vormingu disainiteabe edastamise protsessides. Koos robotiprotsessi simulaatoriga Isaac Sim [163] esindavad sellised lahendused nagu NeMo Curator, Cosmos ja Omniverse uut automatiseerimise taset: alates andmete puhastamisest ja filtreerimisest kuni treeningkogumite genereerimise, objektide omaduste modelleerimise ja robotite treenimiseni ehitusplatsil. Ja kõik need tööriistad on tasuta ja avatud lähtekoodiga, mis vähendab oluliselt takistusi nende kasutuselevõtuks insener- ja ehituspraktikas.

Andmete automaatne klassifitseerimine struktureeritud tabelite tasandil ei ole nii keeruline, kui see esmapilgul võib tunduda. Nagu me eelmises peatükis näitasime (joonis 9.1-10), on võimalik puuduvaid või ebaõigeid klassi väärtusi teiste elementide sarnaste parameetrite põhjal täiendada, kui on olemas kogunenud ajaloolised andmed. Kui sarnaste omadustega elemendid on juba mitmes lõpetatud projektis õigesti klassifitseeritud, võib süsteem uue või puuduliku elemendi jaoks suure tõenäosusega sobiva väärtuse välja pakkuda (joonis 10.1-6). Selline loogika, mis põhineb keskmistatud väärtustel ja kontekstianalüüsil, võib olla eriti tõhus hinnangutest, spetsifikatsioonidest või CAD-mudelitest pärinevate tabeliandmete massitöötlusel.



Joonis 10.1-6 Masinõpe aitab teil automaatselt leida täitmata (valged kastid) tabeli parameetrite keskmisi väärtusi varasemate projektide põhjal.

Masinaõppe nii kiire arengu taustal on selge: 2025. aastal on naiivne arvata, et ehitusdetailide automaatse klassifitseerimise probleem jääb kauaks lahendamata. Jah, praegused algoritmid ei ole veel täielikult välja arenenud, eriti ebatäielike või heterogeensete andmete puhul, kuid kohanemisvõimalus on kiiresti sulgumas.

Ettevõtted, kes juba investeerivad oma andmete kogumisse, puhastamisse ja korrastamisse ning võtavad kasutusele ETL-automaatika vahendid, on eelisseisundis. Ülejäänud riskivad mahajäämisega - nagu ettevõtted ei suutnud kunagi vastata transpordi- ja finantssektori digitaalse ümberkujundamise väljakutsetele

Need, kes jätkuvalt tuginevad manuaalsele andmehaldusele ja traditsioonilistele kulude ja aja arvestamise meetoditele, riskivad sattuda 2000ndate aastate taksoparkide olukorda, mis ei suuda kohaneda mobiilirakenduste ja automatiseeritud marsruudiarvutuste ajastuga 2020ndate aastate alguseks.



PEATÜKK 10.2.

PRAKTILINE JUHEND ANDMEPÕHISE LÄHENEMISVIISI RAKENDAMISEKS

Teooriast praktikasse: teekaart digitaalseks ümberkujundamiseks ehituses

Ehitustööstus on järk-järgult sisenemas uude arengufaasi, kus tuttavaid protsesse täiendavad - ja mõnikord isegi asendavad - üha enam digitaalsed platvormid ja läbipaistvad suhtlusmudelid. See esitab ettevõtetele mitte ainult väljakutseid, vaid ka märkimisväärsed võimalusi. Need organisatsioonid, kes juba täna koostavad pikaajalist digitaalstrateegiat, ei suuda mitte ainult säilitada oma positsiooni turul, vaid ka laiendada seda, pakkudes klientidele kaasaegseid lähenemisviise ja usaldusväärsed, tehnoloogiliselt toetatud lahendusi.

Oluline on mõista, et teadmised mõistetest ja tehnoloogiatest on vaid lähtepunkt. Juhtide ja spetsialistide ees seisab praktiline küsimus: kust alustada rakendamist ja kuidas muuta teoreetilised ideed tegelikuks väärtuseks. Lisaks kerkib üha enam küsimus: mille alusel hakkab äri toimima, kui traditsioonilised kuluarvestus- ja ajastusmeetodid võib klient igal ajal üle vaadata.

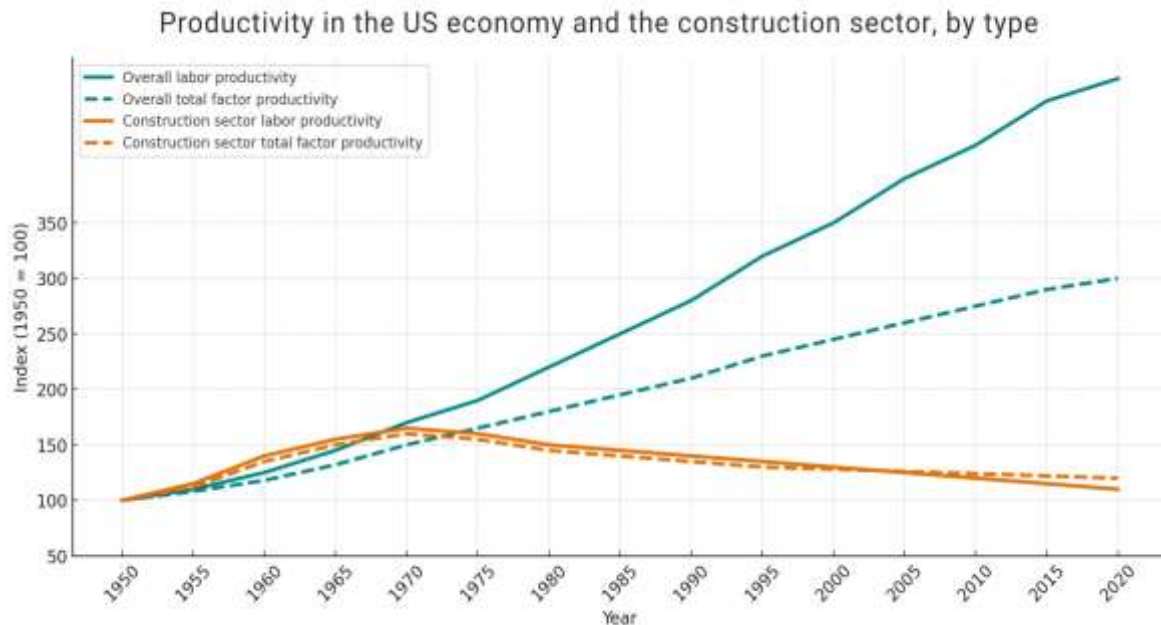
Vastus ei peitu tõenäoliselt mitte niivõrd tehnoloogias, vaid pigem uue ametialase kultuuri kujunemises, kus tööd andmetega nähakse igapäevase praktika lahutamatu osana. Just vähene tähelepanu digitehnoloogiale ja innovatsioonile on nakatanud ehitussektori tõsisesse mahajäämusse, mida on viimastel aastakümnetel täheldatud [43].

McKinsey andmetel moodustavad teadus- ja arendustegevuse kulud ehitussektoris alla 1% tuludest, samas kui autotööstuses ja lennunduses ulatub see näitaja 3,5-4,5%ni. Samamoodi jäävad IT-kulud ehituses alla 1% kogutuludest [107].

Selle tulemusena ei vähene mitte ainult automatiseerituse tase, vaid ka tööviljakus ehituses ning 2020. aastaks toodab ehitustööstus vähem kui pool sajandit tagasi (joonis 10.2-1)

Sellised tootlikkuse probleemid ehitussektoris on ühised enamikus arenenud ja arengumaades (ehitussektori tootlikkus on langenud 16 riigis 29-st OECD riigist (joonis 2.2-1)) ning viitavad mitte ainult tehnoloogia puudumisele, vaid ka vajadusele muuta süsteemselt juba iseenda lähenemist juhtimisele, koolitusele ja innovatsioonile.

Digitaalse ümberkujundamise edu ei sõltu mitte niivõrd vahendite arvust ja kättesaadavusest, vaid organisatsioonide võimest vaadata üle oma protsessid ja arendada muutustele avatud kultuuri. Võtmetähtsusega ei ole mitte tehnoloogia ise, vaid inimesed ja protsessid, mis tagavad selle tõhusa kasutamise, toetavad pidevat õppimist ja soodustavad uute ideede vastuvõtmist.



Joonis 10.2-1 Tööjõu tootlikkuse ja ressursside kogutootlikkuse paradoks USA majanduses ja ehitussektoris (1950-2020) (põhineb [43]).

Raamatu alguses võrreldi ettevõtluskeskkonna mudelit metsa ökosüsteemiga (joonis 2.1-2, joonis 1.2-4, joonis 1.3-2). Terves metsas mängivad perioodilised tulekahjud kogu oma hävitavale võimsusele vaatamata võtmerolli pikaajalises uuenemises. Nad puhastavad pinnase vanast taimestikust, toovad tagasi talletatud toitained ja loovad ruumi uuele elule. Mõned taimeliigid on isegi arenenud nii, et nende seemned avanevad ainult siis, kui nad puutuvad kokku kõrge tulekahju temperatuuriga - see on loomulik mehhanism, mis tagab ideaalse aja idanemiseks.

Samamoodi võivad kriisid toimida äris "kontrollitud läbipõlemisena", soodustades uute lähenemisviiside ja ettevõtete tekkimist, mis ei ole seotud vananenud süsteemidega. Sellised perioodid sunnivad loobuma ebaefektiivsetest tavadest, vabastades ressursse uuenduste jaoks. Nii nagu mets pärast tulekahju algab pioneeritaimedega, kujundab ka ettevõtte pärast kriisi uusi, paindlikke protsesse, mis saavad aluseks küpsele teabekeskonnale.

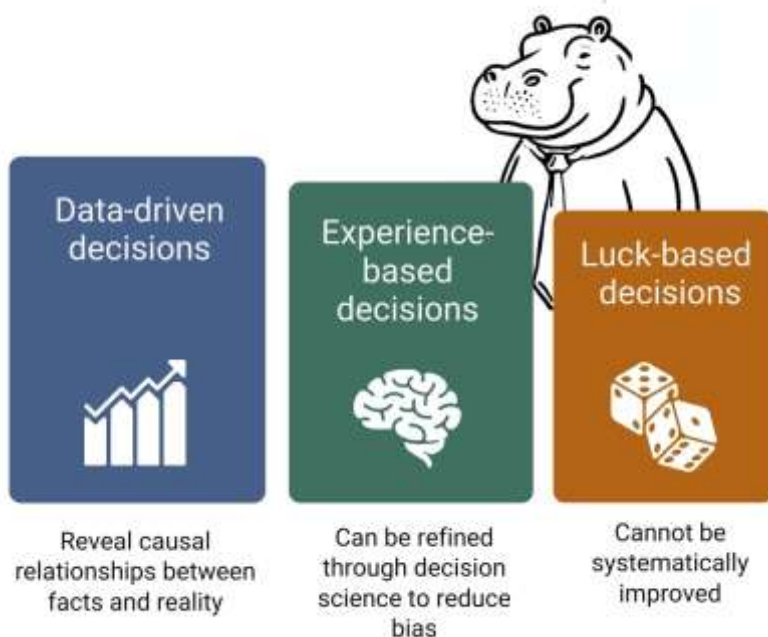
Ettevõtted, kes suudavad neid "signaalitulekahjusid" õigesti tõlgendada ja nende häirivat energiat konstruktiivseks muutuseks muuta, saavutavad uue tulemuslikkuse taseme - läbipaistvamate, kohanemisvõimeliste andmeprotsessidega, mis suurendavad organisatsiooni loomulikku võimet uueneda ja kasvada.

Tehisintellekti ja masinõppe kasvav mõju ärikeskkonnale ei ole enam kahtluse all. See ei ole lihtsalt ajutine trend, vaid strateegiline vajadus. Ettevõtted, kes ignoreerivad tehisintellekti, riskivad konkurentsivõime kaotamisega turul, mis julgustab üha enam innovatsiooni ja paindlikkust.

Tulevik kuulub neile, kes näevad AI-d mitte ainult vahendina, vaid võimalusena mõelda ümber oma äri iga aspekt - alates protsesside optimeerimisest kuni juhtimisotsuste tegemiseni.

Digitaalse vundamendi rajamine: 1-5 sammu digitaalse küpsuse saavutamiseks

Selles peatükis vaatleme digitaalse ümberkujundamise teekaarti ja selgitame välja peamised sammud, mis on vajalikud andmepõhise lähenemisviisi rakendamiseks, mis aitab muuta nii ettevõtte kultuuri kui ka ettevõtte teabeökosüsteemi.



Joonis 10.2-2 Kontrollitud ajakohastamine ja strateegia valik: juhtum, kogemus või andmed.

McKinsey uuringu "Why Digital Strategies Fail" (2018) kohaselt on vähemalt viis põhjust [164], miks ettevõtted ei suuda saavutada digitaalse ümberkujundamise eesmärgi

- **Ebaselged määratlused:** juhtidel ja juhtidel on erinev arusaam sellest, mida tähendab "digitaalne", mis põhjustab arusaamatusi ja ebakõlasid.
- **Digimajanduse vääritimõistmine:** paljud ettevõtted alahindavad digitaliseerimise poolt ärimudelitele ja tööstuse dünaamikale kaasa toodud muutuste ulatust (joonis 10.1-6).
- **Ökosüsteemide eiramine:** ettevõtted keskenduvad üksikutele tehnoloogilistele lahendustele (andmesilodele), jättes tähelepanuta vajaduse integreeruda laiematesse digitaalsetesse ökosüsteemidesse (joonis 2.2-2, joonis 4.1-12).
- **Konkurentide digitaliseerimise alahindamine:** juhid ei võta arvesse, et ka konkurendid võtavad aktiivselt kasutusele digitaal tehnoloogiaid, mis võib kaasa tuua konkurentsieelise kaotuse.
- **Digitaliseerimise duaalsuse eiramine:** tegevjuhid delegeerivad vastutuse digitaalse ümberkujundamise eest teistele juhtidele, mis muudab kontrolli bürokraatlikuks ja aeglustab muutuste protsessi.

Nende probleemidega tegelemine nõuab selget arusaamist ja digitaalstrateegiate kooskõlastamist

organisatsiooni kõikidel tasanditel. Enne digitaalstrateegia koostamist on oluline mõista lähtekohta. Paljud organisatsioonid kipuvad uusi vahendeid ja platvorme kasutusele võtma, ilma et neil oleks täielik ülevaade hetkeseisust.

1. samm: viige läbi oma praeguste süsteemide ja andmete audit.

Enne protsesside muutmist on oluline mõista, mis on juba olemas. Audit läbiviimine võimaldab tuvastada andmehalduse nõrkused ja mõista, milliseid ressursse saab kasutada. Audit on teie äriprotsesside omamoodi "röntgenülevõtte". See võimaldab teil tuvastada riskivaldkonnad ja teha kindlaks, millised andmed on teie projekti või ettevõtte jaoks kriitilise tähtsusega ja millised teisejärgulised.

Peamised tegevused:

- Kaardistage oma IT-keskkond (Draw.io, Lucidchart, Miro, Visio või Canva abil). Loetlege oma protsessides kasutatavad süsteemid (ERP, CAD, CAFM, CPM, SCM ja muud), mida me käsitlesime peatükis "Tehnoloogiad ja juhtimissüsteemid kaasaegses ehituses" (joonis 1.2-4)
- Hinnake iga süsteemi puhul andmete kvaliteediprobleeme seoses dubleerimiste sageduse, võimalike puuduvate väärtuste ja vorminguliste vastuolude esinemisega igas süsteemis.
- Määrake kindlaks "valupunktid" - kohad, kus protsessid võivad katkeda või nõuavad sageli käsitsi sekkumist - import, eksport ja täiendavad valideerimisprotsessid.

Kui soovite, et meeskond usaldaks aruandeid, peate tagama, et andmed on algusest peale õiged.

Kvaliteediandmete audit näitab teile, milliseid andmeid:

- Vajab edasist arendamist (tuleb luua automaatsed puhastusprotsessid või täiendavad transformatsioonid).
- Need on "prügi", mis ainult ummistab süsteemid ja millest saab vabaneda, kui neid enam protsessides ei kasutata.

Sellist auditit on võimalik ka ise läbi viia. Kuid mõnikord on kasulik kaasata väline konsultant - eriti teistest tööstusharudest: värske vaatenurk ja sõltumatus ehituslike "eripärade" suhtes aitab hinnata hetkeolukorda kainelt ja vältida tüüpilisi lõkse, mis tulenevad eelarvamustest teatud lahenduste ja tehnoloogiate suhtes.

2. samm: andmete ühtlustamise põhistandardite kindlakstegemine.

Pärast auditit on vaja luua ühised eeskirjad andmetega töötamiseks. Nagu me arutasime peatükis "Standardid: juhuslikest failidest aruka andmemudelini", aitab see kõrvaldada siilutud andmevood.

Ilma ühtse standardita jätkab iga meeskond "omal moel" töötamist ja te säilitate integratsioonide "loomaia", kus andmed lähevad iga konverteerimisega kaduma.

Peamised tegevused:

- Valige andmestandardid süsteemide vaheliseks teabevahetuseks:
 - ☐ Tabelite kujuliste andmete puhul võib see olla struktureeritud vorming nagu CSV, XLSX või tõhusamad vormingud nagu Parquet.
 - ☐ Lahtiselt struktureeritud andmete ja dokumentide vahetamiseks: JSON või XML.
- Andmemudelitega töötamise meister:
 - ☐ Alustage ülesannete parameetriseerimisest kontseptuaalse andmemudeli tasandil - nagu on kirjeldatud peatükis "Andmete modelleerimine: kontseptuaalne, loogiline ja füüsiline mudel" (joonis 4.3-2).
 - ☐ Kui te süvenete äriprotsessi loogikasse, liigute nõuete formaliseerimise juurde, kasutades parameetreid loogilises ja füüsilises mudelis (joonis 4.3-6).
 - ☐ Identifitseerida peamised üksused, nende atribuudid ja seosed protsessides ning visualiseerida need seosed - nii üksuste kui ka parameetrite vahel (joonis 4.3-7).
- Kasutage regulaaravaldisi (RegEx) andmete valideerimiseks ja standardiseerimiseks (joonis 4.4-7), nagu me arutasime peatükis "Struktureeritud nõuded ja regulaaravaldised RegEx". RegEx ei ole keeruline, kuid äärmiselt oluline teema füüsiliste andmemudelite tasemel nõuete loomise töös.

Ilma andmetasandi standardite ja protsesside visualiseerimiseta on võimatu pakkuda ühtset ja skaleeritavat digitaalkeskkonda. Pidage meeles, et "halvad andmed on kallid". Ja vigade hind suureneb, mida keerulisemaks muutub projekt või organisatsioon. Formaatide ühtlustamine, nimetamise, struktuuri ja valideerimiseeskirjade määratlemine on investering tulevaste lahenduste stabiilsusesse ja skaleeritavusse.

Samm 3. Rakendage DataOps ja automatiseerige protsessid.

Ilma hästi määratletud arhitektuurita seisavad ettevõtted paratamatult silmitsi ebaühtlase andmestikuga, mis sisaldub siilutud infosüsteemides. Andmed on integreerimata, dubleerivad mitmes kohas ja nende haldamine on kulukas.

Kujutage ette, et andmed on vesi ja andmearhitektuur on keeruline torustike süsteem, mis transpordib vett selle ladustamisallikast selle kasutuskohani. Andmearhitektuur määrab, kuidas teavet kogutakse, säilitatakse, teisendatakse, analüüsitakse ja edastatakse lõppkasutajatele või rakendustele.

DataOps (Data Operations) on metoodika, mis integreerib andmete kogumise, puhastamise, valideerimise ja kasutamise ühtseks automatiseeritud protsessivooks, nagu me raamatu 8. osas

üksikasjalikult käsitlesime.

Peamised tegevused:

- Looge ja kohandage ETL -konveierid protsesside automatiseerimiseks:
 - ☐ Extract: korraldage automaatne andmete kogumine PDF dokumentidest (joonis 4.1-2, joonis 4.1-5, joonis 4.1-7), Exceli tabelitest, CAD -mudelitest (joonis 7.2-4), ERP -süsteemidest ja muudest allikatest, millega te töötate.
 - ☐ Transform: seadistada automaatsed protsessid andmete teisendamiseks ühtsesse struktureeritud vormingusse ja automatiseerida arvutused, mis toimuvad väljaspool suletud rakendusi (joonis 7.2-8).
 - ☐ Laadige: proovige luua automaatne andmete üleslaadimine koondtabelitesse, dokumentidesse või tsentraliseeritud hoidlatesse (joonis 7.2-9, joonis 7.2-13, joonis 7.2-16).
- Automatiseerige arvutused ja QTO (Quantity Take-Off) protsessid, nagu me arutasime peatükis "QTO Quantity Take-Off: projektiandmete rühmitamine atribuutide järgi":
 - ☐ Seadistage automaatne mahtude väljavõtmine CAD -mudelitest, kasutades APIsid, lisaseadmeid või pöördprojekteerimise vahendeid (joonis 5.2-5).
 - ☐ Luua reeglid erinevate klasside elementide rühmitamiseks atribuutide järgi tabelite kujul (joonis 5.2-12).
 - ☐ Püüdke automatiseerida sageli korduvaid mahu- ja kuluarvutusi väljaspool moodulitest suletud süsteeme (joonis 5.2-15).
- Alustage andmete töötlemiseks Pythoni ja Pandas kasutamist, nagu me arutasime peatükis "Python Pandas: asendamatu tööriist andmetega töötamiseks":
 - Rakendage DataFrame, et töötada XLSX-failidega ja automatiseerida tabeliandmete töötlemist (joonis 3.4-6).
 - Automatiseerida teabe koondamist ja ümberkujundamist erinevate Pythoni raamatukogude abil
 - Kasutage LLM, et lihtsustada valmis koodiplokkide ja tervete torujuhtmete kirjutamist (joonis 7.2-18).
 - Proovige luua Pythonis , mis leiab vead või näeb kõrvalekaldeid ja saadab vastutava isiku (nt projektijuhi) teavitamise (joonis 7.4-2).

DataOps'i põhimõtetel põhinev automatiseerimine võimaldab teil liikuda käsitsi ja killustatud andmetöötluselt jätkusuutlike ja korratavate protsesside juurde. See mitte ainult ei vähenda töötajate koormust, kes tegelevad iga päev samade transformatsioonidega, vaid suurendab ka oluliselt kogu infosüsteemi usaldusväärsust, skaleeritavust ja läbipaistvust.

4. samm: luua avatud andmete haldamise ökosüsteem.

Vaatamata suletud moodulsüsteemide arengule ja nende integreerimisele uute vahenditega, seisavad

ettevõtted silmitsi tõsise probleemiga - selliste süsteemide kasvav keerukus ületab nende kasulikkuse. Esialgne idee luua ühtne, kõiki äriprotsesse hõlmav platvorm on viinud liigse tsentraliseerimiseni, kus igasugused muudatused nõuavad märkimisväärsed ressursse ja aega kohandamiseks.

Nagu me arutasime peatükis "Ettevõtte mütseel: kuidas andmed ühendavad äriprotsesse", nõuab tõhus andmehaldus avatud ja ühtset ökosüsteemi, mis ühendab kõik teabeallikad.

Ökosüsteemi põhielemendid:

■ Valige sobiv andmekogu:

- ☐ Tabelite ja arvutuste jaoks kasutage andmebaase - näiteks PostgreSQL või MySQL (joonis 3.1-7).
- ☐ Dokumentide ja aruannete jaoks võivad sobida pilvemälu (Google Drive, OneDrive) või süsteemid, mis toetavad JSON-vormingut.
- ☐ Tutvuge andmelao, andmejärvede ja muude vahendite võimalustega suurte andmehulkade tsentraliseeritud säilitamiseks ja analüüsimiseks (joonis 8.1-8).

■ Rakendada lahendusi juurdepääsuks konfidentsiaalsetele andmetele:

- ☐ Kui te kasutate omaloodud süsteeme, seadistage juurdepääs neile API või SDK kaudu, et saada andmeid välise töötlemise jaoks (joonis 4.1-2).
- ☐ Tutvuge CAD formaatide pöördprojekteerimisvahendite võimalustega (joonis 4.1-13).
- ☐ Seadistage ETL-pipeliinid, mis koguvad perioodiliselt andmeid rakendustest või serveritest, teisendavad need avatud struktureeritud vormingutesse ja salvestavad need hoidlatesse (joonis 7.2-3).
- ☐ Arutelu meeskonnas, kuidas tagada juurdepääs andmetele ilma patenteeritud tarkvara kasutamiseta.
- ☐ Pidage meeles: andmed on olulisemad kui liidesed. Pikaajalist väärtust annab teabe struktuur ja kättesaadavus, mitte konkreetsed kasutajaliidese vahendid.

■ Mõelge andmete tippkeskuse (CoE) loomisele, nagu me arutasime peatükis "Andmete modelleerimise tippkeskus (CoE)", või sellele, kuidas te saate pakkuda andmete ekspertiisi muul viisil (joonis 4.3-9).

Andmehalduse ökosüsteem loob ühtse teaberuumi, kus kõik projektis osalejad töötavad järjepideva, ajakohase ja kontrollitud teabega. See on aluseks skaleeritavatele, paindlikele ja usaldusväärsetele digitaalsetele protsessidele

Andmete potentsiaali avamine: 5-10 sammu digitaalse küpsuse saavutamiseni

Lisaks tehnilisele integratsioonile on digitaalsete lahenduste eduka rakendamise oluline tegur nende vastuvõtmine lõppkasutajate poolt. Klientide või kasutajate kaasamine tulemuslikkuse mõõtmisse on nii kasutajakogemuse parandamise kui ka muudatuste juhtimise väljakutse ettevõttes. Kui lahendus ei sobi tuttavasse töökorraldusse või ei lahenda tegelikke kasutaja- või kliendiprobleeme, siis seda ei kasutata ning seda ei paranda ükski lisameede ja stiimul.

Ümberkujundamine on iteratiivne protsess, mis põhineb kasutajate suhtlusandmete analüüsil koos uute protsessidega, sagedaste testimistsüklitega, pideva tagasiside ja täiustustega.

5. samm: Andmekultuuri loomine, töötajate koolitamine ja tagasiside kogumine

Isegi kõige arenenum süsteem ei toimi ilma töötajate kaasamiseta. Peate looma keskkonna, kus andmeid kasutatakse igapäevaselt ja kus meeskond mõistab nende väärtust.

Ühendkuningriigi valitsuse avaldatud aruandes 2024 "Data Analytics and AI in Government Project Delivery" [83] märgitakse, et andmeanalüüsi ja tehisintellekti edukaks rakendamiseks on kriitilise tähtsusega spetsialistide koolitamine, kellel on vajalikud pädevused andmete töötlemiseks ja tõlgendamiseks.

Andmeanalüüsi alaste teadmiste puudumine on üks peamisi probleeme, mis piirab digitaalset ümberkujundamist. Juhid on harjunud väljakujunenud rutiiniga: kvartalitsüklid, prioriteetsed algatused ja traditsioonilised viisid projektide edendamiseks. Muutused nõuavad eristuvat juhti - piisavalt kõrget positsiooni, et omada mõju, kuid mitte nii kõrget, et tal oleks aega ja motivatsiooni pikaajalise ümberkujundamisprojekti juhtimiseks.

Peamised tegevused:

- Tunnistades vajadust minna üle subjektiivsetelt, kõrgepalgalise töötaja arvamusel põhinevatelt otsustelt (HiPPO) faktidel ja andmetel põhinevale otsustuskultuurile, nagu on käsitletud peatükis "HiPPO ehk arvamuse oht otsustamisel" (joonis 2.1-9).
- Korraldage süstemaatiline koolitus:
 - ☐ korraldada koolitusi struktureeritud andmete kasutamise kohta ja kutsuda eksperte teistest tööstusharudest, kes ei ole eelarvamuslikult meelestatud tänapäeval ehitussektoris populaarsete toodete ja kontseptsioonide suhtes.
 - ☐ arutleda kolleegidega andmeanalüüsi lähenemisviiside ja vahendite üle ning õppida iseseisvalt praktilist tööd selliste vahenditega nagu Python, pandas ja LLM (joonis 4.1-3, joonis 4.1-6).
 - ☐ Luua õpperaamatukogu (soovitavalt lühivideodega) andmete struktureerimise (joonis 3.2-15) ja andmemudelite loomise teemal (joonis 4.3-6, joonis 4.3-7).
- Kasutage kaasaegseid õppetehnoloogiaid:
 - ☐ Keelemudelite (LLM) kasutamine koodi- ja andmehalduse toetamiseks, sealhulgas koodi genereerimiseks, refaktooremiseks ja analüüsiks, samuti tabeliandmete töötlemiseks ja tõlgendamiseks (joonis 3.4-1).
 - ☐ Uurida, kuidas LLMi abil loodud koodi saab kohandada ja integreerida terviklikuks torujuhtme lahenduseks, kui töötatakse võrguühenduseta arenduskeskkonnas (IDE) (joonis 4.4-14, joonis 5.2-13).

Kui juht jätkab otsuste tegemist "vanaviisi", ei veena ükski koolitus inimesi analüütikat tõsiselt võtma.

Andmekultuuri loomine on võimatu ilma pideva tagasiside andmiseta. Tagasiside võimaldab teil tuvastada protsesside, vahendite ja strateegiate puudujäägid, mida ei ole võimalik avastada sisearuannete või ametlike KPIde meetrikate abil. Teie lahenduste kasutajate komplimentaarsed kommentaarid ei anna praktilist väärtust. Kriitiline tagasiside on väärtuslik, eriti kui see põhineb konkreetsetel tähelepanekutel ja faktidel. Kuid sellise teabe saamine nõuab jõupingutusi: peate looma protsessid, mille käigus osalejad - nii sise- kui ka välised - saavad jagada kommentaare (võib olla mõistlik teha seda anonüümselt) ilma moonutusteta ja kartmata, et nende arvamused võivad mõjutada nende enda tööd. Oluline on, et nad teeksid seda moonutamata ja kartmata negatiivseid tagajärgi enda jaoks.

Igasugune õppimine on lõppkokkuvõttes iseõppimine [165].

- Milton Friedman, Ameerika majandusteadlane ja statistik

Analüütiliste vahendite rakendamisega peaks kaasnema nende tõhususe korrapärane kontrollimine praktikas (tasuvus, tulemuslikkuse näitajad), mida saab saavutada ainult töötajate, klientide ja partnerite struktureeritud tagasiside kaudu. See võimaldab ettevõtetel mitte ainult vältida vigade kordamist, vaid ka kiiremini kohaneda keskkonna muutustega. Tagasiside kogumise ja analüüsimise mehhanismi olemasolu on üks organisatsiooni küpsuse tunnuseid, mis näitab, et organisatsioon liigub juhuslikest digitaalsetest algatustest pideva täiustamise jätkusuutliku mudeli suunas (joonis 2.2-5).

6. samm: Katseprojektidest kuni suurendamiseni

Valige piisavalt suured lahingud, et need oleksid olulised ja piisavalt väikesed, et neid võita.

- Jonathan Kozol

Digitaalse ümberkujundamise käivitamine "kõik korraga ja kõikjal" on äärmiselt riskantne. Tõhusam lähenemisviis on alustada pilootprojektidega ja järk-järgult laiendada edukaid tavasid.

Peamised tegevused:

■ Valige pilootprojekti jaoks õige projekt:

☐ Määratleda konkreetne ärieesmärk või protsess koos mõõdetavate tulemustega (KPI,

ROI) (joonis 7.1-5).

- ☐ Valige ETL-i automatiseerimisprotsess, näiteks andmete automaatne valideerimine või töökoormuse arvutamine (QTO), kasutades Python ja Pandas (joonis 5.2-10)
- ☐ Kehtestada selged mõõdikud edu saavutamiseks (nt vähendada kontrolli spetsifikatsioonide või andmete valideerimisaruannete koostamise aega nädalast päevani).

■ Kasutage iteratiivseid lähenemisviise:

- ☐ Alustage lihtsatest andmekonversiooniprotsessidest ja looge mitmeformaadiliste andmete voogkonversioone oma protsesside jaoks vajalikesse vormingutesse (joonis 4.1-2, joonis 4.1-5).
- ☐ Suurendage järk-järgult ülesannete keerukust ja laiendage protsesside automatiseerimist, moodustades IDEs dokumenteeritud koodiplokkidel põhineva täieliku Pipeline'i (joonis 4.1-7, joonis 7.2-18).
- ☐ dokumenteerida ja salvestada (soovitavalt lühivideodega) edukaid lahendusi ning jagada neid kolleegidega või erialastes kogukondades.

■ Töötage välja mallid ja kaasnev dokumentatsioon selliste lahenduste kordamiseks, et teie kolleegid (või erialase kogukonna liikmed, sealhulgas sotsiaalmeedia kasutajad) saaksid neid tõhusalt kasutada.

Samm-sammult "ülesrullimine" võimaldab säilitada muudatuste kõrge kvaliteedi ja mitte langeda paralleelsete rakenduste kaosesse. "Väikestest suurteni" strateegia minimeerib riske ja võimaldab õppida väikestest vigadest, laskmata neil kasvada kriitilisteks probleemideks.

Üleminek projektipõhiselt lähenemisviisilt, kus töötajad on vaid osaliselt kaasatud, alaliste meeskondade (nt kompetentsikeskuste) moodustamisele aitab tagada toote jätkusuutliku arengu ka pärast selle esimese versiooni väljalaskmist. Sellised meeskonnad mitte ainult ei toeta olemasolevaid lahendusi, vaid jätkavad ka nende täiustamist.

See vähendab sõltuvust pikkadest heakskiitmistest: meeskonnaliikmetel on õigus teha otsuseid oma vastutusala piires. Selle tulemusel vabanevad juhid mikromanööre ja meeskonnad saavad keskenduda tegeliku väärtuse loomisele.

Uute lahenduste väljatöötamine ei ole sprint, vaid maraton. Selles on edukad need, kes esialgu püüavad teha pikaajalist ja järjepidevat tööd.

Oluline on mõista, et tehnoloogia vajab pidevat arengut. Investeeringud tehnoloogiliste lahenduste pikaajalisesse arendamisse on eduka tegevuse alus.

Samm 7: Kasutage avatud andmeformaate ja lahendusi

Nagu me arutasime moodulplatvorme (ERP, PMIS, CAFM, CDE jne) käsitlevates peatükkides, on oluline keskenduda avatud ja universaalsetele andmeformaatidele, mis tagavad sõltumatuse müüja

lahendustest ja suurendavad teabe kättesaadavust kõigi sidusrühmade jaoks.

Peamised tegevused:

■ Üleminek suletud vormingutelt avatud vormingutele:

- ☐ Kasutage avatud formaate patenteeritud formaatide asemel või leidke viis automaatse üleslaadimise või suletud formaatide avatud formaatidesse konverteerimise seadistamiseks (joonis 3.2-15).
- ☐ Rakendada vahendeid, mis töötavad koos Parquet, CSV, JSON, XLSX, mis on enamiku kaasaegsete süsteemide vahelised andmevahetusstandardid (joonis 8.1-2).
- ☐ Kui 3D-geomeetriaga töötamine mängib teie protsessides olulist rolli, kaaluge avatud vormingute, näiteks USD, glTF, DAE või OBJ kasutamist (joonis 3.1-14).

■ Kasutage vektorandmebaase andmebaasid, et tõhusalt analüüsida ja otsida teavet:

- ☐ Kasutage Bounding Box ja muid meetodeid 3D -geomeetria lihtsustamiseks (joonis 8.2-1).
- ☐ Mõelge, kus saate rakendada andmete vektoriseerimist - tekstide, objektide või dokumentide teisendamist arvuliseks esituseks (joonis 8.2-2).

■ Rakendage suurandmete analüüsi vahendeid:

- ☐ Korraldada kogutud ajalooliste andmete (nt PDF, XLSX, CAD) salvestamine analüüsiks sobivatesse vormingutesse (Apache Parquet, CSV, ORC) (joonis 8.1-2).
- ☐ Alustada põhiliste statistiliste meetodite rakendamist ja tööd representatiivsete valimitega - või vähemalt tutvuda statistika aluspõhimõtetega (joonis 9.2-5).
- ☐ Rakendada ja õppida vahendeid andmete ja nende vaheliste seoste visualiseerimiseks, et visualiseerida analüüsitulemusi. Ilma hea visualiseerimiseta on võimatu täielikult mõista andmeid ennast või nendel põhinevaid protsesse (joonis 7.1-4).

Üleminek avatud andmeformaatile ja teabe analüüsi-, salvestus- ja visualiseerimisvahendite kasutuselevõtmine loob aluse jätkusuutlikule ja sõltumatule digitaalsele haldusele. See mitte ainult ei vähenda sõltuvust müüjatest, vaid tagab ka kõigile sidusrühmadele võrdse juurdepääsu andmetele.

Samm 8. Alustage masinõppe rakendamist prognoosimiseks

Paljud ettevõtted on kogunud tohutuid andmehulki - omamoodi "teabegeisreid", mis on veel kasutamata. Need andmed on kogutud sadade või tuhandete projektide käigus, kuid sageli on neid edasistes protsessides kasutatud ainult üks kord või üldse mitte. Suletud formaatides ja süsteemides salvestatud dokumente ja mudeleid peetakse sageli vananenud ja kasutuks ballastiks. Tegelikult on need aga väärtuslik ressurs - alus tehtud vigade analüüsimiseks, rutiinsete toimingute automatiseerimiseks ja uuenduslike lahenduste väljatöötamiseks automaatse klassifitseerimise ja tunnuste tuvastamise jaoks tulevastes projektides.

Peamine väljakutse on õppida, kuidas neid andmeid ekstraheerida ja muuta need kasutatavateks

järeldusteks. Nagu on käsitletud peatükis "Masinõpe ja prognoosid", on masinõppe tehnikatel võimalik oluliselt parandada hinnangute ja prognooside täpsust mitmesuguste ehitusega seotud protsesside puhul. Kogutud andmete täielik kasutamine avab tee tõhususe parandamisele, riskide vähendamisele ja jätkusuutlike digitaalsete protsesside loomisele.

Peamised tegevused:

- Alustage lihtsate algoritmidega:
 - ☐ Proovige rakendada lineaarset regressiooni - kasutades LLM-i vihjeid -, et ennustada korduvat tulemuslikkust andmekogumites, kus sõltuvus paljudest teguritest puudub või on minimaalne (joonis 9.3-4).
 - ☐ Mõelge, millistes protsessides etappides võiks teoreetiliselt rakendada k-naabrite (k-NN) algoritmi - näiteks klassifitseerimisülesannete, objektide sarnasuse hindamise või ajaloolistel analoogidel põhineva ennustamise puhul (joonis 9.3-5).
- Andmete kogumine ja struktureerimine mudelite treenimiseks:
 - ☐ Koguda ajaloolisi projektandmeid ühte kohta ja ühtses formaadis (joonis 9.1-10).
 - ☐ Töö koostamisvalimite kvaliteedi ja representatiivsuse kallal, automatiseeritud ETLi kaudu (joonis 9.2-8).
 - ☐ Õppige andmed eraldama treening- ja testkogumiteks, nagu me tegime Titanicu andmekogumi näite puhul (joonis 9.2-6, joonis 9.2-7).
- kaaluda masinõppe meetodite kohaldamise laiendamist mitmesugustele ülesannetele, alates projekti tähtaegade prognoosimisest kuni logistika optimeerimise, ressursside juhtimise ja võimalike probleemide varajase tuvastamiseni.

Masinõpe on vahend, mille abil saab arhiveeritud andmed muuta väärtuslikuks varaks prognoosimiseks, optimeerimiseks ja teadlike otsuste tegemiseks. Alustage väikeste andmekogumite (joonis 9.2-5) ja lihtsate mudelitega, suurendades järk-järgult nende keerukust.

Samm 9. Integreeri asjade interneti ja kaasaegsed andmekogumistehnoloogiad.

Ehitusmaailm on kiiresti muutumas digitaalseks: iga ehitusfoto, iga Teams'i postitus on juba osa suuremast reaalsuse parameetriseerimise ja sümboliseerimise protsessist. Nii nagu GPS muutis kunagi logistikat, muudavad asjade internet, RFID ja automaatsed andmekogumistehnoloogiad ehitustööstust. Nagu on käsitletud peatükis "Asjade internet ja arukad lepingud", on digitaalne ehitusplats koos sensorite ja automaatse jälgimisega tööstuse tulevik.

Peamised tegevused:

- Rakendage asjade Interneti -seadmed, RFID -märgised ja kirjeldage üksikasjalikult nendega seotud protsesse:
 - ☐ hinnata, millistes valdkondades või projekti etappides on investeeringu tasuvus (ROI) kõige suurem - näiteks temperatuuri, vibratsiooni, niiskuse või liikumise jälgimiseks.
 - ☐ Kaaluge RFID kasutamist materjalide, tööriistade ja seadmete jälgimiseks kogu

tarneahela ulatuses.

- ☐ Kaaluge, kuidas kogutud andmeid saab integreerida ühtsesse infosüsteemi, näiteks Apache NiFi, automatiseeritud reaalajas töötlemiseks ja analüüsiks (joonis 7.4-5).

■ Luua reaalajas jälgimissüsteem:

- ☐ Arendage armatuurlauad, et jälgida peamisi protsessi- või projektinäitajaid, kasutades visualiseerimisvahendeid, nagu Streamlit, Flask või Power BI).
- ☐ Seadistage automaatsed teavitused, et anda märku kriitilistest kõrvalekalletest plaanist või normidest (joonis 7.4-2).
- ☐ Hinnata kogutud andmete ja tuvastatud mustrite põhjal seadmete ennetava hoolduse potentsiaali (joonis 9.3-6).

■ Kombineerida eri allikatest saadud andmeid:

- ☐ Alustage andmemudeli visualiseerimisest füüsilisel tasandil - kajastage infovoogude struktuuri ja CAD-süsteemidest, asjade interneti seadmetest ja ERP-platvormidest pärinevaid põhiparameetreid (joonis 4.3-1).
- ☐ Alustage andmete analüüsi ja juhtimisotsuste toetamise ühtse platvormi umbkaudse visandi loomisest. Märkige peamised funktsioonid, andmeallikad, kasutajad ja eeldatavad rakendusstsenaariumid (joonis 4.3-7).

Mida kiiremini hakkate tegelikke protsesse digitaalsesse maailma ühendama, seda kiiremini saate neid andmete abil hallata - tõhusalt, läbipaistvalt ja reaalajas.

10. samm. Valmistuge tööstuse tulevasteks muutusteks

Ehitusettevõtted on pidevalt väliskeskkonna surve all: majanduskriisid, tehnoloogilised hüpped, regulatiivsed muudatused. Nagu mets, mis peab vastu pidama vihmale, lumele, põuale ja kõrvetavale päikesele, elavad ettevõtted pideva kohanemise tingimustes. Ja nii nagu puud muutuvad sügavate juurestike abil vastupidavaks külma ja põua suhtes, jäävad elujõuliseks ja konkurentsivõimeliseks ainult need organisatsioonid, millel on kindel automatiseeritud protsesside alus, võime muutusi ette näha ja strateegiaid paindlikult kohandada.

Nagu mainitud peatükis "Ellujäämisstrateegiad: konkurentsieeliste loomine", on ehitussektor jõudmas radikaalsete muutuste faasi. Tellija ja töövõtja vaheline suhtlus on liikumas veenismudeli suunas, kus läbipaistvus, prognoositavus ja digitaalsed vahendid asendavad traditsioonilised lähenemisviisid. Selles uues reaalsuses ei võida mitte suurimad, vaid kõige paindlikumad ja tehnoloogiliselt küpsemad.

Peamised tegevused:

■ Analüüsige äritegevuse haavatavusi avatud andmete kontekstis:

- ☐ Hinnake, kuidas andmete demokratiseerimine Uberiseerimise raames võib avaldada laastavat mõju teie konkurentsieelsele ja äritegevusele (joonis 10.1-5).
- ☐ Mõelge strateegiale, mille eesmärk on liikuda läbipaistmatutest ja siloorsetest protsessidest avatud lahendustel, süsteemide koostalitlusvõimel ja andmete

läbipaistvusel põhinevate ärimudelite suunas (joonis 2.2-5).

■ Töötage välja pikaajaline digitaalstrateegia:

- ☐ Määrake kindlaks, kas soovite olla innovatsiooniliider või eelistate "järelajõudmise" stsenaariumi, mille puhul säästate oma ressursse.
- ☐ Kirjeldada etappe: lühiajaline (protsesside automatiseerimine, andmete struktureerimine), keskpikk perspektiiv (LLM ja ETL rakendamine), pikaajaline (digitaalsed ökosüsteemid, tsentraliseeritud hoidlad).

■ Mõelge oma teenuste portfelli laiendamisele:

- ☐ Kaaluda uute teenuste pakkumist (keskendudes energiatõhususele, ESG-le, andmesideteenustele). Uutest ärimudelitest räägime lähemalt järgmises peatükis.
- ☐ Püüdke positsioneerida end usaldusväärse tehnoloogiapartnerina, kes toetab rajatise kogu elutsüklit, alates projekteerimisest kuni käitamiseni. Usaldus teie vastu peaks põhinema süstemaatilisel lähenemisviisil, läbipaistvatel protsessidel ja võimel pakkuda jätkusuutlikke tehnoloogilisi lahendusi.

Ümberkujundavas keskkonnas võidavad mitte need, kes lihtsalt reageerivad muutustele, vaid need, kes on ennetavad. Paindlikkus, avatus ja digitaalne küpsus on jätkusuutlikkuse alus homse tuleviku ehitamisel.

Ümberkujundamise tegevuskava: kaosest andmepõhiseks ettevõtteks

Järgnev kava võib olla esialgne võrdlusalus - lähtepunkt teie enda andmepõhise digitaalse ümberkujundamise strateegia kujundamiseks:

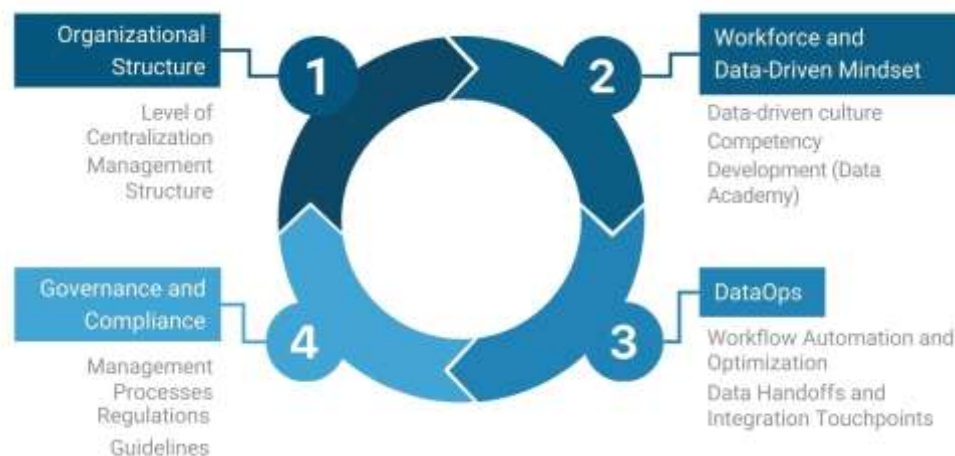
- **Audit ja standardid:** analüüsida praegust olukorda, ühtlustada andmeid.
- **Andmete struktureerimine ja klassifitseerimine:** struktureerimata ja lahtiselt struktureeritud andmete automatiseeritud ümberkujundamine.
- **Grupeeringute, arvutuste ja arvutuste automatiseerimine:** kasutage automatiseerimiseks avatud lähtekoodiga tööriistu ja raamatukogusid.
- **Ökosüsteem ja COE:** luua sisemine meeskond, et moodustada ettevõttes ühtne andmete ökosüsteem.
- **Kultuur ja õppimine:** üleminek HiPPO -lahendustelt andmepõhistele lahendustele.
- **Pilootprojektid, tagasiside ja skaleerimine:** tegutsege iteratiivselt: katsetage uusi meetodeid piiratud mahu, koguge kehtivat tagasisidet ja suurendage järk-järgult lahendusi.
- **Avatud vormingud:** kasutage universaalseid ja avatud vorminguid, et olla sõltumatu tarkvaratootjatest.
- **Masinõpe:** ML-algoritmide integreerimine protsessidesse prognoosimiseks ja optimeerimiseks.
- **Asjade interneti aadress ja digitaalne ehitusplats:** kaasaegsete andmekogumistehnoloogiate

integreerimine protsessidesse

- **Strateegiline kohanemine:** valmistuda tulevasteks tööstuse muutusteks

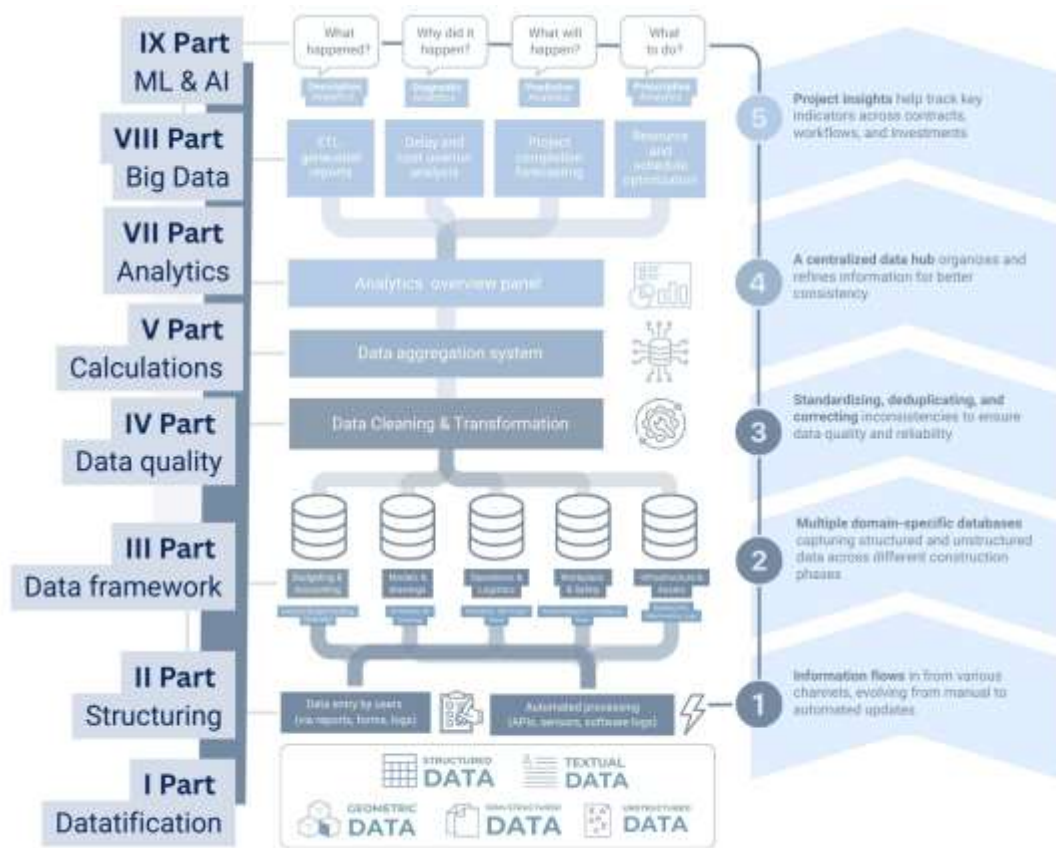
Kõige tähtsam on meeles pidada, et "andmed üksi ei muuda ettevõtet: ettevõtet muudavad inimesed, kes oskavad nende andmetega töötada". Rõhutage kultuuri, läbipaistvaid protsesse ja pühendumust pidevale täiustamisele.

Süsteemne lähenemine võimaldab üleminekut eraldiseisvatelt digitaalsetelt algatustelt täieõiguslikule andmepõhisele juhtimismudelile, kus otsused ei põhine mitte intuitsioonil või oletustel, vaid andmetel, faktidel ja matemaatiliselt arvatud tõenäosustel. Ehitustööstuse digitaalne ümberkujundamine ei tähenda mitte ainult tehnoloogia kasutuselevõttu, vaid sellise äriökosüsteemi kujundamist, kus projektiteavet edastatakse sujuvalt ja iteratiivselt erinevate süsteemide vahel. Seejuures pakuvad masinõppe algoritmid automaatset, pidevat analüüsi, prognoosimist ja protsesside optimeerimist. Sellises keskkonnas muutuvad spekulatsioonid ja varjatud andmed ebaoluliseks - alles jäävad vaid tõestatud mudelid, läbipaistvad arvutused ja prognoositavad tulemused.



Joonis 10.2-3 Ettevõtte tasandi eduka andmehalduse põhielemendid.

Raamatu iga osa vastab ehitusprojektide andmetöötamise ja -analüüsi konkreetsele etapile (joonis 2.2-5). Kui soovite pöörduda tagasi ühe varem käsitletud teema juurde ja vaadata seda andmete kasutamise voo terviklikust käsitlemisest lähtuvalt, võite viidata osade pealkirjadele joonisel 10.2-4.



Joonis 10.2-4 Raamatu osad andmetötluse torujuhtme kontekstis (joonis 2.2-5): teabe digitaliseerimisest analüütika ja tehisintellekteni.

Sõltumata teie organisatsiooni suurusest, tehnoloogilise küpsuse tasemest või eelarvest, saate juba täna alustada üleminekut andmepõhisele lähenemisviisile. Isegi väikesed sammud õiges suunas annavad aja jooksul tulemusi.

Andmepõhine ümberkujundamine ei ole ühekordne projekt, vaid pidev, iteratiivne parendusprotsess, mis hõlmab uute vahendite kasutuselevõttu, protsesside ümberkujundamist ja andmepõhise otsustuskultuuri arendamist.

Ehitus tööstuses 5.0: kuidas teenida raha, kui sa ei saa enam varjata

Pikka aega on ehitusettevõtted teeninud raha protsesside läbipaistmatusega. Peamine ärimudel oli spekulatsioon - materjalide maksumuse, tööde mahu ja protsentuaalsete hinnalisandite ülehindamine suletud ERP-süsteemis - ja PMISis - süsteemides, mis on välisaudititele kättesaamatud. Tellijate ja nende usaldusisikute piiratud juurdepääs projekti algsetele andmetele lõi pinnase skeemidele, kus arvutuste usaldusväärsuse kontrollimine muutus peaaegu võimatuks.

See mudel on aga kiiresti kaotamas oma tähtsust. Andmetele juurdepääsu demokratiseerimise, elukestva õppejõu tekkimise, avatud andmete ja ETLi automatiseerimisvahendite tuleku tõttu liigub tööstus uue tööstandardi suunas.

Selle tulemusena ei ole läbipaistmatuse enam konkurentsieelis - varsti muutub see koormaks, millest on raske loobuda. Läbipaistvus muutub valikuvõimalusest turul püsimise eelduseks.

Kellega töötavad kliendid - pangad, investorid, füüsilised kliendid, erakapitali investeerimisfondid, valitsuskliendid - uues digitaalses reaalsuses? Vastus on ilmselge: nendega, kes suudavad pakkuda mitte ainult tulemust, vaid ka iga sammu põhjendust teel selleni. Kuna avatud andmete maht kasvab, valivad partnerid ja kliendid ettevõtted, kes tagavad läbipaistvuse, täpsuse ja tulemuste prognoositavuse.

Selle taustal on tekkimas uued ärimudelid, mis põhinevad pigem andmehaldusel ja usaldusel kui spekulatsioonil:

- **Protsesside müümine ruutmeetrite asemel:** peamine vara on usaldus ja tõhusus, mitte konkreetsed lepingud, mis on allahinnatud. Peamine väärtus on usaldusväärsusel ja kontrollitud andmetel põhineva tulemuse prognoositavus. Kaasaegsed ettevõtted ei müü ehitusobjekti kui sellist, vaid:
 - täpsed tähtajad ja läbipaistvad tööplaanid;
 - mõistlikud hinnangud, mida toetavad arvutused;
 - täielik digitaalne jälgitavus ja kontroll projekti kõikides etappides.
- **Tehnika ja analüütika kui teenus:** "Data-as-a-Service " mudel (võimalus pakkuda valmis andmeid kasutajatele interneti kaudu teenusena), kus iga projekt muutub digitaalse andmeahela osaks ja äriiline väärtus seisneb selle ahela haldamise võimekuses. Ettevõtted on muutumas intelligentseteks platvormideks, mis pakuvad automatiseerimisel ja analüüsil põhinevaid lahendusi:
 - Hinnangute ja plaanide automatiseeritud ja läbipaistev koostamine;
 - masinõppe algoritmidele põhinev riski ja ajastuse hindamine;
 - keskkonnanäitajate (ESG, CO₂, energiatõhusus) arvutamine;
 - aruannete koostamine auditeeritud avatud allikatest.

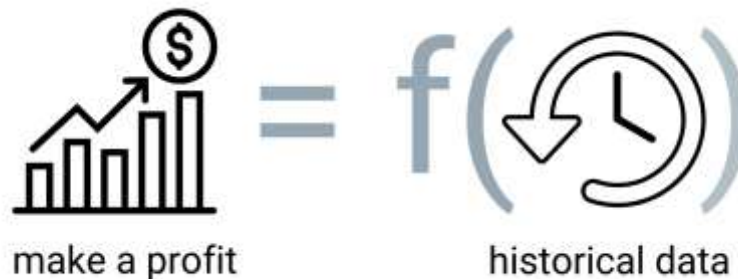
- **inseneriteadmiste tootestamine:** ettevõtte arendusi saab kasutada korduvalt ettevõttesiseselt ja levitada eraldi tootena - moodustades digitaalsete teenuste kaudu täiendava sissetulekuallika. Uues keskkonnas ei loo ettevõtted mitte ainult projekte, vaid ka digitaalseid varasid:
 - komponentide raamatukogud ja kalkulatsioonimallid;
 - automatiseeritud kontrollimise moodulid;
 - Avatud lähtekoodiga pistikprogrammid ja skriptid (nõustamise müük) andmetega töötamiseks.
- **Uut tüüpi ettevõtte: andmepõhine integraator:** turuosaline, kes ei sõltu konkreetsetest tarkvaratootjatest ega moodulsüsteemidest ega ole "lukustatud" ühte tarkvaraliidesesse. Ta tegutseb vabalt andmetega - ja ehitab oma konkurentsivõime nende põhjal. Tuleviku ehitusettevõtte ei ole lihtsalt töövõtja, vaid teabeintegraator, kes suudab kliendi jaoks täita järgmisi funktsioone:
 - Ühendage eri allikatest saadud andmeid ja tehke analüütikat;
 - Tagada protsesside läbipaistvus ja usaldusvärsus;
 - anda nõu äriprotsesside optimeerimiseks;
 - arendada vahendeid, mis töötavad avatud andmete ökosüsteemis, LLM, ETL ja Pipelines.

Tööstus 5.0 (joonis 2.1-12) tähistab "käsitsi keskmiste suhtarvude ajastu" ja õhtuste koosolekute lõppu tegevjuhtide ning kalkulatsiooni- ja raamatupidamisosakonna vahel. Kõik, mis varem oli varjatud - arvutused, hinnangud, mahud - muutub avatuks, kontrollitavaks ja arusaadavaks ka mitteeekspertidele. Need, kes esimesena ümber orienteeruvad, on võitjad. Kõik teised jäävad ehitussektori uuest digitaalsest majandusest välja.

KOKKUVÕTE

Ehitustööstus on sisenemas fundamentaalsete muutuste ajastusse. Alates esimestest savitahvlitele kantud dokumentidest kuni projektiserveritest ja ehitusplatsidel liikuvate tohutute digitaalsete andmemahitudeni on ehitusalase teabehalduse ajalugu alati peegeldanud oma aja tehnoloogia küpsust. Tänapäeval seisab ehitussektor automatiseerimise, avatud vormingute ja intelligentsete analüüsisüsteemide tulekuga silmitsi mitte järkjärgulise arenguga, vaid kiire digitaalse muutusega.

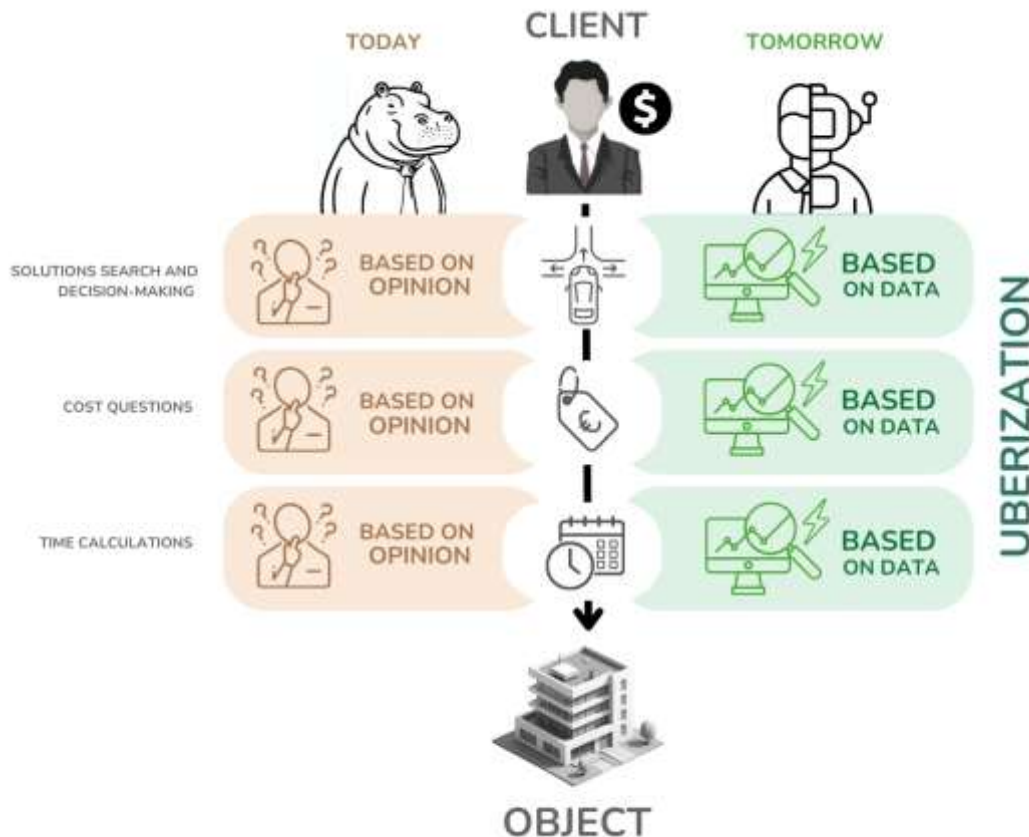
Nagu teistes majandussektorites, tuleb ka ehituses ümber mõelda mitte ainult töövahendid, vaid ka tööpõhimõtted. Varem turgu dikteerinud ja kliendi ja projekti vahel peamise vahendajana tegutsenud ettevõtted on kaotamas oma ainulaadset positsiooni. Usaldus ja võime töötada andmetega - alates andmete kogumisest ja struktureerimisest kuni nende analüüsimise, prognoosimise ja otsuste automatiseerimiseni - tõusevad esiplaanile.



Joonis 10.2-1 Struktureeritud ajaloolised andmed on tõhusa ja juhitava ettevõtte kütus.

Selles raamatus käsitletakse üksikasjalikult andmehalduse põhiprintsiipe ehitustööstuses - alates auditeerimisest ja standardimisest kuni protsesside automatiseerimiseni, visualiseerimisvahendite kasutamiseni ja intelligentsete algoritmide rakendamiseni. Vaatlesime, kuidas isegi piiratud ressursside juures saab luua toimiva andmearhitektuuri ja hakata otsuseid tegema pigem kontrollitavate faktide kui intuitsiooni põhjal. Andmetega töötamine ei ole enam ainult IT-osakonna ülesanne - sellest saab juhtimiskultuuri alus, mis määrab ettevõtte paindlikkuse, kohanemisvõime ja pikaajalise jätkusuutlikkuse.

Masinõppe tehnoloogiate, automaatsete töötlemissüsteemide, digitaalsete kaksikute ja avatud formaatide rakendamine võimaldab juba täna kõrvaldada inimteguri seal, kus see oli varem kriitiline. Ehitus on liikumas autonoomia ja kontrollitavuse suunas, kus liikumist ideest projekti realiseerimiseni võib võrrelda navigatsiooniga autopiloodi režiimis: ilma sõltumiseta subjektiivsetest otsustest, ilma vajaduseta sekkuda igas etapis käsitsi, kuid täieliku digitaalse jälgitavuse ja kontrolliga (joonis 10.2-2).



Joonis 10.2-2 Üleminekut oluliste ekspertide arvamustel põhinevalt otsustamiselt (HiPPO) andmeanalüüsile soodustab eelkõige klient.

Õppides selles raamatus esitatud meetodeid, põhimõtteid ja vahendeid, saate oma ettevõttes hakata tegema pigem andmepõhiseid kui intuitiivseid otsuseid. Samuti suudate käivitada mooduliahelaid LLMides, kopeerida valmis ETL-pipeliinid oma arenduskeskkonda (IDE) ja töödelda andmeid automaatselt, et saada vajalikku teavet soovitud kujul. Tuginedes raamatu suurandmete ja masinõppe peatükkidele, saate seejärel rakendada keerulisemaid stsenaariume - kaevandada uusi teadmisi ajaloolistest andmetest ja rakendada masinõppe algoritme oma protsesside prognoosimiseks ja optimeerimiseks.

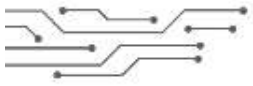
Avatud andmed ja protsessid loovad aluse projektide kulude ja tähtaegade täpsemaks hindamiseks, vältides ehitusettevõtete spekulatsioonide läbipaistmatute andmete põhjal. See on nii väljakutse kui ka võimalus, et tööstusharu mõtleks oma rolli ümber ja kohaneks uue keskkonnaga, kus läbipaistvus ja tõhusus muutuvad peamisteks eduteguriteks.

Valmisolek võtta teadmisi ja rakendada neid praktikas on digitaalse muutuse ajastul edu võti.

Ettevõtted, kes seda esimesena mõistavad, saavad uues digitaalses konkurentsies eelise. Kuid oluline on mõista, et andmed üksi ei muuda midagi. Paljud inimesed peavad muutma oma mõtteviisi ja selleks on vaja stiimuleid. Teie ettevõtte peab ümber mõtlema, kuidas ta andmeid jagab

Ettevõtet muudavad need inimesed, kes oskavad nende andmetega töötada, neid tõlgendada, optimeerida ja luua nende põhjal uue protsesside arhitektuuri.

Kui te loete neid ridu, siis olete valmis muutusteks ja olete juba ühe sammu võrra ees. Tänan teid selle tee valimise eest. Tere tulemast digitaalse transformatsiooni ajastusse!



AUTORI KOHTA

Minu nimi on Artjom Boiko. Minu teekond ehitusplatsil algas 2007. aastal - töötades kaevurina põlevkivikaevanduses minu kodulinnas, samal ajal kui ma õppisin Peterburi Mäeülikoolis, spetsialiseerudes kaevanduste ja maa-aluste rajatiste ehitamisele. Selle raamatu tagakaanel näete plahvatusohtlikku nägu, kus me kaevandasime ja lõhkusime sadu kuubikuid põlevkivi. Minu karjäär on viinud mind mitmetes erinevates suundades, alates kaevanduse ja maa-aluse ehituse töölistest kuni tööstusliku ronimise, katusetööde ja tösteseadmete paigaldamiseni. Mul on olnud au osaleda eri ulatusega projektides, alates eramajade ehitamisest kuni suurte tööstusrajatisteni maailma eri piirkondades.



Aja jooksul on minu töö nihkunud füüsilisest ehitusest infohalduse ja digitaalsete protsesside juurde. Alates 2013. aastast olen töötanud erinevatel ametikohtadel väikestes, keskmise suurusega ja suurtes ehitusettevõtetes mitmes Saksamaa piirkonnas, alates projekteerijast kuni andmehaldusjuhini. Mis puutub andmehaldusse, siis minu kogemused hõlmavad tööd andmetega erinevates ERP-süsteemides, CAD (BIM), MEP, FEM, CMS. Olen tegelenud optimeerimise, protsesside automatiseerimise, samuti analüüsi, masinõppe ja andmetöötlemisega ehitustööde planeerimise, arvutamise ja teostamise etappides tööstus-, elamu-, infrastruktuuri- ja kommunaalehitusettevõtetes.

Olen töötanud avatud lähtekoodiga tarkvara ja avatud andmetega alates 2003. aastast. Selle aja jooksul olen rakendanud mitmeid veebiprojekte - alates veebilehtedest ja veebipoodidest kuni täiemahuliste veebirakendusteni - kasutades avatud lähtekoodiga lahendusi ja avatud CMSi. Need platvormid, mis sarnanevad paljuski moodsate hoonete ERP-idega, on modulaarse ülesehituse, suure kohandatavuse ja ligipääsetavusega. See kogemus pani aluse minu professionaalsele lähenemisele - keskendumine avatud lähtekoodiga tehnoloogiatele ja koostööpõhisele arenduskultuurile. Avatud lähtekoodi ja vaba teadmiste vahetamise austamine on midagi, mida ma püüan ehitussektoris edendada. Minu töö andmete kättesaadavuse parandamiseks ehitussektoris on väljendunud mitme sotsiaalmeediakogukonna loomises, et arutada andmete avatust ja avatud lähtekoodi kasutamist ehituses, ning mitmete idufirmade käivitamises, mis arendavad lahendusi, mis võimaldavad juurdepääsu erinevatest suletud süsteemidest ja platvormidest pärit andmetele.

Minu panus erialasele kogukonnale on osalemine ettekandjana konverentsidel, mis käsitlevad CAD koostalitlusvõimet (BIM), ERP, 4D-5D, LLM Machine Learning ja Artificial Intelligence, samuti Euroopa ehitussektori väljaannetes avaldatud artiklid. Üks minu märkimisväärsed saavutusi on "BIM History" loomine. [111], mis on põhjalik kaart ehitussektori andmehalduse oluliste tarkvaralahenduste kohta. Minu seitsmeosaline artikliseeria "BIM Development and Lobbying Games", mis on tõlgitud mitmesse keelde, on leidnud laialdast tunnustust kui katse valgustada digitaalsete standardite varjatud dünaamikat.

Nii jõudsin ma kivi kaevandamisest ehitusandmete kaevandamise ja süstematiseerimiseni. Olen alati avatud professionaalsele dialoogile, uutele ideedele ja ühisprojektidele. Võtan tänulikult vastu igasuguse tagasiside ja ootan teie sõnumeid või teie nägemist minu sotsiaalmeedia jälgijate hulgas. Suur tänu, et lugesite selle raamatu lõpuni! Mul oleks hea meel, kui see raamat aitab teil paremini mõista andmete teemat ehitussektoris.

EDASINE SUHE

Lugejate arvamused mängivad olulist rolli väljaannete edasisel arendamisel ja prioriteetsete teemade valikul. Eriti väärtuslikud on kommentaarid selle kohta, millised ideed on osutunud kasulikuks ja millised on tekitanud kahtlusi ning vajavad täiendavaid selgitusi või allikate tsiteerimist. Raamat sisaldab laiaulatuslikku materjali ja analüüsi, millest mõned võivad tunduda vastuolulised või subjektiivsed. Kui leiate lugemise käigus ebatäpsusi, valesi viidatud allikaid, loogilisi vastuolusid või kirjavigu, oleksin tänulik teie kommentaaride, mõtete või kriitika eest, mille võite saata aadressil: boikoartem@gmail.com. Või sõnumite kaudu LinkedInis: [linkedin.com/in/boikoartem](https://www.linkedin.com/in/boikoartem).

Oleksin tänulik, kui raamatut "Data-Driven Construction" mainitaks sotsiaalmeedias - lugemiskogemuste jagamine aitab levitada teavet avatud andmete ja tööriistade kohta ning toetab minu tööd.

TÕLKEKOMMENTAAR

See raamat on tõlgitud tehisintellekti tehnoloogia abil. See on muutnud tõlkeprotsessi palju kiiremaks. Kuid nagu iga tehnoloogilise toimingu puhul, võivad esineda vead või ebatäpsused. Kui märkate midagi, mis tundub olevat valesti või valesti tõlgitud, palun saatke mulle e-kiri. Teie märkused aitavad parandada tõlke kvaliteeti.

ANDMEPÕHISED EHITUSKOGUKONNAD

See on koht, kus saate vabalt esitada küsimusi ning jagada oma probleeme ja lahendusi:

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>

MUUD OSKUSED JA MÕISTED

Lisaks ehitustööstuses andmetega töötamise põhiprintsiipidele käsitletakse DataDrivenConstruction'is mitmeid täiendavaid kontseptsioone, programme ja oskusi, mis on andmepõhisele spetsialistile hädavajalikud. Mõned neist on esitatud ainult ülevaatlilikult, kuid on praktika jaoks kriitilise tähtsusega.

Huvitatud lugeja võib külastada DataDrivenConstruction.io veebisaiti, kus on lingid lisamaterjalidele põhioskuste kohta. Need materjalid hõlmavad tööd Pythoni ja Pandasiga, ETL-i -protsesside ehitamist, näiteid andmetöötlustest ehitus-CAD-projektides, suurandmete süsteeme ning moodsaid lähenemisviise ehitusandmete visualiseerimisele ja analüüsile.

Raamatu "DataDrivenConstruction" ja kõigi juhtumiuuringute koostamisel on kasutatud mitmeid avatud lähtekoodiga vahendeid ja tarkvara. Autor soovib tänada järgmiste lahenduste arendajaid ja kaasautoreid:

- Python ja Pandas - andmete manipuleerimise ja automatiseerimise selgroog
- Scipy, NumPy, Matplotlib ja Scikit-Learn - raamatukogud andmeanalüüsi ja masinõppe jaoks.
- SQL ja Apache Parquet - vahendid suurte ehitusandmete koguste salvestamiseks ja töötlemiseks.
- Avatud lähtekoodiga CAD (BIM) avatud andmetöötlusvahendid avatud vormingutes
- N8n, Apache Airflow, Apache NiFi - orkestreerimise ja töövoogude automatiseerimise süsteemid.
- DeepSeek, LLaMa, Mistral - avatud lähtekoodiga LLM

Eriline tänu kõigile avatud andmete ja töövahendite teemat käsitlevates aruteludes osalenutele, kelle kriitika, kommentaarid ja ideed aitasid parandada käesoleva raamatu sisu ja struktuuri.

Jälgi projekti arengut DataDrivenConstruction.io veebilehel, kus avaldatakse mitte ainult raamatu uuendused ja parandused, vaid ka uued peatükid, õpetused ja praktilised näited kirjeldatud tehnikate rakendamise kohta.

MAKSIMAALNE MUGAVUS TRÜKITUD VERSIOONIGA

Sul on käes tasuta digitaalne versioon raamatust **Andmepõhine ehitus**. Mugavamaks tööks ja kiireks juurdepääsuks materjalidele soovitame pöörata tähelepanu **trükitud väljaandele**:



■ **Alati käepärast:** trükitud raamatust saab usaldusväärne töövahend, mis võimaldab teil igas töösituatsioonis kiiresti leida ja kasutada õigeid visualiseeringuid ja skeeme.

■ **Kvaliteetsed illustratsioonid:** kõik trükiväljaande pildid ja graafika on esitatud maksimaalse kvaliteediga.

■ **Kiire juurdepääs teabele:** lihtne navigeerimine, võimalus teha märkmeid, järjehoidjaid ja töötada raamatuga igal pool.

Ostes raamatu trükitud täisversiooni, saate mugava

abivahendi mugavaks ja tõhusaks tööks teabega: võimaluse kasutada visuaalseid materjale kiiresti igapäevastes ülesannetes, leida kiiresti vajalikud skeemid ja teha märkmeid. Lisaks toetab teie ostuga avatud teadmiste levitamist.

Tellige raamatu trükitud versioon aadressil: datadrivenconstruction.io/books



AINULAADNE VÕIMALUS STRATEEGILINE POSITSIONEERIMINE

Pakume teile võimalust paigutada reklaammaterjale DataDrivenConstructioni tasuta versioonis. Väljaande tasuline versioon on esimese aasta jooksul pärast ilmumist ärritanud tähelepanu spetsialistide seas rohkem kui 50 maailma riigist - Ladina-Ameerikast kuni Aasia-Vaikse ookeani piirkonnani. Individuaalsete koostöötingimuste arutamiseks ja üksikasjaliku teabe saamiseks paigutusvõimaluste kohta täitke ametliku portaali datadrivenconstruction.io tagasisidevorm või kirjutage raamatu lõpus loetletud kontaktidele.



RAAMATU PEATÜKID ON KÄTTESAADAVAD AADRESSIL DATADRIVENCONSTRUCTION.IO

Andmuspõhise ehituse peatükke saate lugeda Andmuspõhise ehituse veebisaidil, kus raamatu peatükid avaldatakse järk-järgult, et saaksite vajaliku teabe kiiresti üles leida ja seda oma töös kasutada. Samuti leiate sealt palju muid sarnastel teemadel avaldatud trükiseid ning näiteid rakenduste ja lahenduste kohta, mis aitavad teil arendada oma oskusi ja rakendada andmeid ehituses.



RAAMATU VIIMASED VERSIOONID ALLALAADIMINE AMETLIKUST VEEBISAIDIST

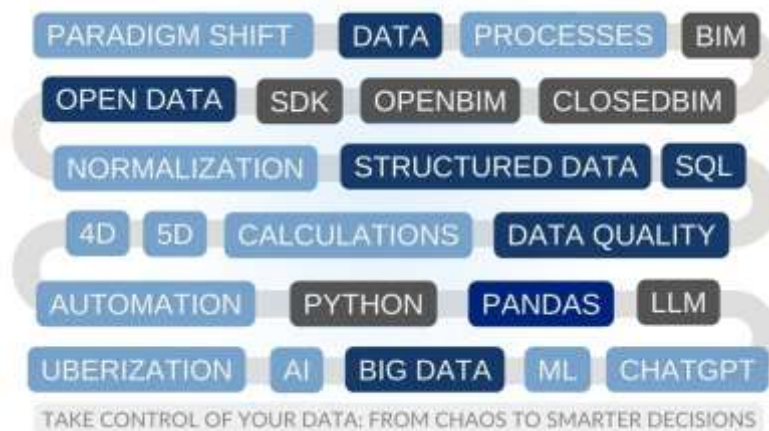
DataDrivenConstruction raamatu praegune ja uusim versioon on allalaadimiseks saadaval aadressil datadrivenconstruction.io. Kui soovite saada uuendusi raamatu uute peatükkide, praktiliste nõuannete või uute rakenduste ülevaadete kohta, tellige uudiskiri:

- Saate esimesena tutvuda raamatu uute osadega
- Saate praktilisi juhtumiuuringuid ja nõuandeid analüütika ja automatiseerimise kohta ehitustööstuses.
- Hoidke end kursis trendide, väljaannete ja rakendusnäidetega.

Mine datadrivenconstruction.io, et tellida!

ANDMEPÕHINE EHITUS: KONSULTATSIOONID, ÕPITOAD JA KOOLITUSED

DataDrivenConstructioni koolitusprogrammid ja nõustamine on aidanud kümnetel juhtival ehitusettevõtetel üle maailma suurendada tõhusust, vähendada kulusid ja parandada lahenduste kvaliteeti. DataDrivenConstructioni klientide hulka kuuluvad mõned miljardeid eurosid maksva turu suurimad tegijad, sealhulgas ehitus-, konsultatsiooni- ja IT-ettevõtted.



Miks valida meid?

- **Asjakohasus:** räägime tööstuse peamistest suundumustest ja arusaamadest.
- **Praktiline:** aitab spetsialistidel lahendada igapäevaseid ülesandeid tõhusalt PoCi abil.
- **Individuaalne lähenemine:** kohandatud teie ettevõttele, maksimeerides koolituse ja nõustamise eeliseid.

DataDrivenConstructioni meeskonna peamised valdkonnad:

- **Andmekvaliteedi juhtimine:** abi ülesannete parameetriseerimisel, nõuete kogumisel, andmete valideerimisel ja ettevalmistamisel automatiseeritud töötlemiseks.
- **Andmete kaevandamine - andmete väljavõtmine ja struktureerimine:** ETL-protsesside loomine ja andmete väljavõtmine e-kirjadest, PDF, Excel, piltidest ja muudest allikatest.
- **BIM ja CAD analüütika:** koguda, struktureerida ja analüüsida teavet RVT-failidest, IFC, DWG ja muudest CAD (BIM) formaatidest.
- **Analüütika ja andmete ümberkujundamine:** ebaühtlase teabe muutmine struktureeritud andmeteks, analüüsiks, arusaamadeks ja otsusteks.
- **Andmete integreerimine ja protsesside automatiseerimine:** alates dokumentide automaatsest loomisest kuni sisesüsteemide ja väliste andmebaasidega integreerimiseni.

Võtke ühendust DataDrivenConstruction.io'ga, et teada saada, kuidas automatiseerimise kasutamine võib aidata teie ettevõttel saavutada käegakatsutavaid äritulemusi.

GLOSSARIUM

AI (Artificial Intelligence) - tehisintellekt; arvutisüsteemide võime täita ülesandeid, mis tavaliselt nõuavad inimese intelligentsust, näiteks mustrite tuvastamine, õppimine ja otsuste tegemine.

Apache Airflow on avatud lähtekoodiga töövoogude orkestreerimisplatvorm, mis võimaldab programmeeritult luua, planeerida ja jälgida töövooge ja ETL-e, kasutades DAG-e (suunatud atsüklilised graafid).

Apache NiFi on süsteemidevaheliste andmevoogude automatiseerimise vahend, mis on spetsialiseerunud andmete marsruutimisele ja teisendamisele.

Apache Parquet on tõhus failivorming veeruslike andmete salvestamiseks, mis on optimeeritud kasutamiseks suurandmete analüüsisüsteemides. See pakub märkimisväärset pakkimist ja kiiret töötlemist.

API (Application Programming Interface) - ametlik liides, mis võimaldab ühel programmil suhelda teise programmiga ilma juurdepääsuta lähtekoodile, vahetades andmeid ja funktsioone standardiseeritud päringute ja vastuste kaudu.

Atribuut - objekti tunnus või omadus, mis kirjeldab selle omadusi (nt pindala, maht, maksumus, materjal).

Andmebaasid on organiseeritud struktuurid teabe säilitamiseks, haldamiseks ja juurdepääsuks, mida kasutatakse andmete tõhusaks otsimiseks ja töötlemiseks.

BEP (BIM Execution Plan) - ehitusinformatsiooni modelleerimise rakendusplaan, mis määratleb eesmärgid, meetodid ja protsessid BIMi rakendamiseks projektis.

Suurandmed - märkimisväärse mahu, mitmekesisuse ja ajakohastamise kiirusega teabe massiivid, mille töötlemiseks ja analüüsiks on vaja spetsiaalset tehnoloogiat.

BI (Business Intelligence) - äriteabe; protsessid, tehnoloogiad ja vahendid andmete muutmiseks sisuliseks teabeks otsuste tegemiseks.

BIM (Building Information Modeling) - ehitusinfo modelleerimine; ehitusprojektide füüsiliste ja funktsionaalsete omaduste digitaalsete esituste loomise ja haldamise protsess, mis hõlmab mitte ainult 3D-mudeleid, vaid ka teavet toimivuse, materjalide, ajakava ja kulude kohta.

BlackBox/WhiteBox - lähenemisviisid süsteemi mõistmiseks: esimesel juhul on sisemine loogika peidetud, nähtavad on ainult sisendid ja väljundid; teisel juhul on töötlemine läbipaistev ja analüüsiks kättesaadav.

Bounding Box on geomeetriline konstruktsioon, mis kirjeldab objekti piire kolmemõõtmelises ruumis minimaalsete ja maksimaalsete X-, Y- ja Z-koordinaatide abil, luues objekti ümber "kasti".

BREP (Boundary Representation) on objektide geomeetriline kujutamine, mis määratleb neid pindade piiride kaudu.

CAD (Computer-Aided Design) on arvutipõhine projekteerimissüsteem, mida kasutatakse täpsete jooniste ja 3D-mudelite loomiseks, redigeerimiseks ja analüüsimiseks arhitektuuris, ehituses, inseneriteaduses ja muudes tööstusharudes.

CAFM (Computer-Aided Facility Management) on kinnisvara ja infrastruktuuri haldamise tarkvara, mis hõlmab ruumide planeerimist, varade haldamist, hooldust ja kulude jälgimist.

CDE (Common Data Environment) - tsentraliseeritud digitaalne ruum projektiandmete haldamiseks, säilitamiseks, jagamiseks ja koostööks rajatise elutsükli kõikides etappides.

Tippkeskus on spetsialiseerunud struktuur organisatsioonis, mis vastutab konkreetse teadmisvaldkonna arendamise, standardite ja parimate tavade väljatöötamise, töötajate koolitamise ja uuenduste kasutuselevõtu toetamise eest.

CoClass on kaasaegne, kolmanda põlvkonna ehituselementide klassifitseerimissüsteem.

Kontseptuaalne andmemudel on andmebaasi projekteerimise algaasis kasutatav kõrgetasemeline esitus põhiobjektidest ja nende seostest ilma atribuutide üksikasjadeta.

CRM (Customer Relationship Management) on kliendisuhete haldamise süsteem, mida kasutatakse müügi- ja teenindusprotsesside automatiseerimiseks.

DAG (Directed Acyclic Graph) on suunatud atsükliline graaf, mida kasutatakse andmete orkestreerimissüsteemides (Airflow, NiFi) ülesannete järjestuse ja sõltuvuste määramiseks.

Dash on Pythoni raamistik interaktiivsete veebipõhiste andmete visualiseerimise loomiseks.

Armatuurlaud - armatuurlaud, mis esitab visuaalselt peamised tulemusnäitajad ja mõõdikud reaaliajaks.

Andmekeskne lähenemisviis on meetodika, mis seab andmed rakenduste või tarkvarakoodi asemel esikohale, muutes andmed organisatsiooni keskseks varaks.

Andmehaldus - tavad, protsessid ja põhimõtted, mis tagavad andmete asjakohase ja tõhusa kasutamise organisatsioonis, sealhulgas juurdepääsu-, kvaliteedi- ja turvakontrolli.

Andmejärv on hoidla, mis on mõeldud suurte koguste töötlemata andmete säilitamiseks nende algsel kujul kuni nende kasutamiseni.

Data Lakehouse on arhitektuuriline lähenemisviis, mis ühendab andmelaikade (Data Lake) paindlikkuse ja skaleeritavuse andmelao haldatavuse ja jõudlusega (DWH).

Andmepõhine ehitus on strateegiline lähenemisviis, mille puhul rajatise elutsükli iga etappi - alates projekteerimisest kuni käitamiseni - toetavad automatiseeritud, omavahel ühendatud süsteemid. Selline lähenemisviis tagab pideva, faktidel põhineva õppimise, vähendab ebakindlust ja võimaldab ettevõtetel saavutada jätkusuutlik tööstuse juhtpositsiooni.

Andmepõhine integraator - ettevõtte, mis on spetsialiseerunud erinevatest allikatest pärit andmete ühendamisele ja analüüsimisele, et teha juhtimisotsuseid.

Andmepõhine lähenemisviis - meetodika, mille puhul andmeid käsitletakse strateegilise väärtusena ja otsused tehakse pigem objektiivse teabe analüüsi kui subjektiivsete arvamuste põhjal.

Andmeminimalism - lähenemisviis andmete vähendamiseks kõige väärtuslikumale ja sisukamale, mis võimaldab lihtsustatud teabe töötlemist ja analüüsi.

Andmesoo - Struktureerimata andmete hajutatud mass, mis tekib siis, kui teavet kogutakse ja säilitatakse kontrollimatult, ilma nõuetekohase organiseerimiseta.

DataOps on meetodika, mis ühendab DevOps-põhimõtted, andmed ja analüütika, keskendudes koostöö parandamisele, integreerimisele ja andmevoogude automatiseerimisele.

Teabe digitaliseerimine on protsess, mille käigus kõik ehitustegevuse aspektid teisendatakse digitaalsesse vormi, mis sobib analüüsiks, tõlgendamiseks ja automatiseerimiseks.

DataFrame - Kahemõõtmeline tabeli kujuline andmestruktuur Pandase raamatukogus, kus read kujutavad üksikuid kirjeid või objekte ja veerud nende omadusi või atribuute.

Kirjeldav analüüs - ajalooliste andmete analüüsimine, et mõista, mis juhtus minevikus.

Diagnostiline analüüs - andmete analüüsimine, et teha kindlaks, miks midagi juhtus.

Gantti diagramm on projekti planeerimise vahend, mis kujutab ülesandeid horisontaalsete ribadena ajateljel, võimaldades visualiseerida tööde järjekorda ja kestust.

DWH (Data Warehouse) on tsentraliseeritud andmelaosüsteem, mis koondab mitmest allikast pärit teabe, struktureerib selle ja teeb selle kättesaadavaks analüüsiks ja aruandluseks.

ESG (Environmental, Social, Governance) - kriteeriumide kogum ettevõtte või projekti keskkonna-, sotsiaal- ja juhtimismõjude hindamiseks.

ELT (Extract, Load, Transform) on protsess, mille käigus andmed eraldatakse esmalt allikatest ja laaditakse repositooriumi ning seejärel muudetakse analüütilistel eesmärkidel.

ETL (Extract, Transform, Load) on protsess, mille käigus andmed ekstraheeritakse erinevatest allikatest, teisendatakse soovitud vormingusse ja laaditakse sihtmällu analüüsiks.

ER-diagramm (Entity-Relationship) - visuaalne diagramm, mis näitab üksusi, nende atribuute ja nende vahelisi seoseid, mida kasutatakse andmete modelleerimisel.

ERP (Enterprise Resource Planning) on terviklik modulaarne ettevõtte ressursside planeerimise süsteem, mida kasutatakse ehitusprotsessi erinevate aspektide haldamiseks ja optimeerimiseks.

Tunnused - masinõppes sõltumatud muutujad või atribuudid, mida kasutatakse mudeli sisendina.

Füüsiline andmemudel - andmebaasi struktuuri üksikasjalik esitus, sealhulgas tabelid, veerud, andmetüübid, võtmed ja indeksid, mis on optimeeritud konkreetse DBMSi jaoks.

FPDF on Pythoni raamatukogu PDF-dokumentide loomiseks.

Geometric Core on tarkvarakomponent, mis pakub põhilisi algoritme geomeetriliste objektide loomiseks, redigeerimiseks ja analüüsimiseks CAD-, BIM- ja muudes insenerirakendustes.

HiPPO (Highest Paid Person's Opinion) - lähenemine otsuste tegemisele, mis põhineb pigem organisatsiooni kõrgeima palgaga isiku arvamusel kui objektiivsetel andmetel.

IDE (Integrated Development Environment) - integreeritud arenduskeskkond, terviklik vahend koodi kirjutamiseks, testimiseks ja silumiseks (nt PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) on teabe esitamise spetsifikatsioon, mis määratleb andmenõuded projekti eri etappides.

IFC (Industry Foundation Classes) on BIM-andmevahetusformaad, mis tagab koostalitlusvõime erinevate tarkvaralahenduste vahel.

Tööstus 5.0 on tööstuslik arengukontseptsioon, mis ühendab digitaliseerimise, automatiseerimise ja tehisintellekti võimalused inimpotentsiaali ja keskkonnasäästlikkuse.

Andmete integreerimine on protsess, mille käigus ühendatakse eri allikatest pärinevad andmed ühtseks, sidusaks süsteemiks, et anda ühtne ülevaade teabest.

Teabesilod on isoleeritud andmesalvestussüsteemid, mis ei jaga teavet teiste süsteemidega, mis takistab andmete tõhusat kasutamist.

Asjade internet (Internet of Things) on füüsiliste objektide ühendamise interneti, et koguda, töödelda ja edastada andmeid.

k-NN (k-Nearest Neighbors) on masinõppe algoritm, mis klassifitseerib objekte sarnasuse alusel lähimate naabritega treeningvalimis.

Kaggle on andmeanalüüsi ja masinõppe võistluste platvorm.

Kalkulatsioon - ehitustööde või -protsesside maksumuse arvutamine teatud mõõtühiku kohta (nt 1 m² kipsplaadist seina, 1 m³ betooni).

KPId (Key Performance Indicators) on peamised tulemusnäitajad, mõõdetavad mõõdikud, mida kasutatakse ettevõtte või konkreetse projekti edukuse hindamiseks.

Märgised - masinõppe puhul sihtmuutujad või atribuudid, mida mudel peaks ennustama.

Õppimisalgoritm - protsess, mille käigus leitakse parim hüpotees mudelis, mis vastab sihtfunktsioonile, kasutades treeninguandmete kogumit.

Lineaarne regressioon - statistiline meetod sõltuva muutuja ja ühe või mitme sõltumatu muutuja vahelise seose modelleerimiseks.

LLM (Large Language Model) - suur keelemudel, tehisintellekt, mis on koolitatud teksti mõistmiseks ja genereerimiseks suurtest andmemahtudest, mis on võimeline analüüsima konteksti ja kirjutama programmeerimiskoodi.

LOD (Level of Detail/Development) - mudeli detailsuse tase, mis määrab geomeetrilise täpsuse ja infosisu.

Loogiline andmemudel on üksuste, atribuutide, võtmete ja seoste üksikasjalik kirjeldus, mis kajastab äriteavet ja -reegleid ning on kontseptuaalse ja füüsilise mudeli vahepealne etapp.

Masinõpe - tehisintellekti meetodite klass, mis võimaldab arvutisüsteemidel õppida ja teha ennustusi andmete põhjal ilma selgesõnalise programmeerimiseta.

Masterformaad on esimese põlvkonna klassifitseerimissüsteem, mida kasutatakse ehitusspetsifikaatide struktureerimiseks osadeks ja erialadeks.

MEP (Mechanical, Electrical, Plumbing) - hoonete tehnosüsteemid, mis sisaldavad mehaanilisi, elektrilisi ja sanitaarkomponente.

Võrguvõrk on 3D-objektide võrgusilma kujutis, mis koosneb tippudest, servadest ja külgedest.

Mudel - masinõppes erinevate hüpoteeside kogum, millest üks lähendab ennustatavat või lähendatavat sihtfunktsiooni.

Andmete modelleerimine on protsess, mille käigus luuakse andmete ja nende seoste struktureeritud esitus infosüsteemides rakendamiseks, sealhulgas kontseptuaalsel, loogilisel ja füüsilisel tasandil.

n8n on avatud lähtekoodiga tööriist töövoogude automatiseerimiseks ja rakenduste integreerimiseks madala koodiga.

Normaliseerimine - masinõppes erinevate arvandmete viimine ühtsesse skaalasse, et hõlbustada töötlemist ja analüüsi.

Reverse engineering - seadme, toimimise ja valmistustehnoloogia uurimine, analüüsides eseme struktuuri, funktsioone ja toimimist. Andmete kontekstis - teabe ekstraheerimine kaitstud vormingutest, et seda saaks kasutada avatud süsteemides.

OCR (Optical Character Recognition) on **optiline** tähemärgituvastustehnoloogia, mis teisendab tekstipildid (skaneeritud dokumendid, fotod) masinloetavaks tekstivorminguks.

OmniClass on teise põlvkonna rahvusvaheline klassifikatsioonistandard ehitusteabe haldamiseks.

Ontoloogia - mõistete omavaheliste seoste süsteem, mis vormistab konkreetse teadmishalduskonna.

Avatud lähtekood - mudel avatud lähtekoodiga tarkvara arendamiseks ja levitamiseks, mis on vabaks kasutamiseks, uurimiseks ja muutmiseks kättesaadav.

Avatud BIM on avatud BIMi kontseptsioon, mis hõlmab avatud standardite ja vormingute kasutamist erinevate tarkvaralahenduste vaheliseks andmevahetuseks.

Avatud standardid - avalikult kättesaadavad spetsifikatsioonid konkreetse ülesande saavutamiseks, mis võimaldavad erinevatel süsteemidel suhelda ja andmeid vahetada.

Pandas on avatud lähtekoodiga Pythoni andmetöötlus- ja analüüsiraamatukogu, mis pakub DataFrame ja Series andmestruktuure tabeliandmete tõhusaks töötlemiseks.

Avatud andmete paradigma on lähenemine andmetöötusele, mille puhul teave on vabalt kättesaadav kasutamiseks, taaskasutamiseks ja levitamiseks kõigile.

Parameetiline meetod on ehitusprojekti hindamismeetod, mis kasutab statistilisi mudeleid, et hinnata väärtust projekti parameetrite alusel.

PIMS (Project Information Model) on digitaalne süsteem, mis on loodud kogu projektiteabe korraldamiseks, säilitamiseks ja jagamiseks.

Pipeline - andmetöötusprotsesside jada, alates andmete ekstraheerimisest ja teisendamisest kuni analüüsi ja visualiseerimiseni.

PMIS (Project Information Management System) on projektijuhtimissüsteem, mis on ette nähtud ülesannete üksikasjalikuks kontrollimiseks üksikute ehitusprojektide tasandil.

Ennustav analüütika on analüütika osa, mis kasutab statistilisi meetodeid ja masinõpet, et ennustada tulevasi tulemusi varasemate andmete põhjal.

Prescriptive Analytics - analüütika osa, mis mitte ainult ei ennusta tulevasi tulemusi, vaid soovib ka optimaalseid tegevusi soovitud tulemuste saavutamiseks.

Proprietäärsed vormingud on suletud andmevormingud, mida kontrollib konkreetne ettevõtte, mis piiravad teabe jagamise võimalust ja suurendavad sõltuvust konkreetsest tarkvarast.

QTO (Quantity Take-Off) on elementide koguste väljavõtmine projekteerimisdokumentidest, et arvutada projekti realiseerimiseks vajalike materjalide kogused.

Kvaliteedijuhtimissüsteem - kvaliteedijuhtimissüsteem, mis tagab, et protsessid ja tulemused vastavad kehtestatud nõuetele.

RAG (Retrieval-Augmented Generation) on meetod, mis ühendab keelemudelite genereerimisvõime ja asjakohase teabe väljavõtte ettevõtete andmebaasidest, parandades vastuste täpsust ja asjakohasust.

RDBMS (Relational Database Management System) on relatsiooniline andmebaasi haldussüsteem, mis korraldab teavet omavahel seotud tabelite kujul.

Regex (Regular Expressions) on formaliseeritud keel stringide otsimiseks ja töötlemiseks, mis võimaldab määrata mallid tekstiandmete kontrollimiseks teatud kriteeriumidele vastavuse suhtes.

Regressioon on statistiline meetod muutujate vaheliste seoste analüüsimiseks.

CO₂-arvutused on meetod, mille abil hinnatakse ehitusmaterjalide ja -protsesside tootmise ja kasutamisega seotud süsinikdioksiidi heitkoguseid.

Ressursimeetod on meetod, mille abil koostatakse hinnanguid, mis põhinevad kõigi ehitustööde teostamiseks vajalike ressursside (materjalid, tööjõud, seadmed) üksikasjalikul analüüsil.

RFID (Radio Frequency Identification) on tehnoloogia, mis võimaldab automaatselt tuvastada objekte raadiosignaali abil ja mida kasutatakse materjalide, masinate ja töötajate jälgimiseks.

ROI (Return on Investment) on näitaja, mis kajastab kasumi ja investeeritud vahendite suhet ja mida kasutatakse investeeringute tõhususe hindamiseks.

SaaS (Software as a Service) on tarkvara kui teenuse mudel, kus rakendusi majutab teenusepakkuja ja teeb need kasutajatele kättesaadavaks Interneti kaudu.

SCM (Supply Chain Management) - tarneahela juhtimine, mis hõlmab kõigi protsesside koordineerimist ja optimeerimist alates materjalide hankimisest kuni valmistoodete tarnimiseni.

Andmesilod on organisatsioonisiseseid isoleeritud teabehoidlaid, mis ei ole integreeritud teiste süsteemidega, mistõttu on andmete jagamine keeruline ja ebaefektiivne.

SQL (Structured Query **Language**) on struktureeritud päringukeel, mida kasutatakse relatsiooniliste andmebaasidega töötamiseks.

SQLite on kergekaaluline, sisseehitatav, platvormideülene DBMS, mis ei vaja eraldi serverit ja toetab põhilisi SQL-funktsioone, mida kasutatakse laialdaselt mobiilirakendustes ja sisseehitatud süsteemides.

Struktureeritud andmed - teave, mis on korraldatud kindlas formaadis ja selge struktuuriga, näiteks relatsiooniliste andmebaaside või tabelite kujul.

Lahtiselt struktureeritud andmed - osalise korralduse ja paindliku struktuuriga teave, näiteks JSON või XML, mille eri elemendid võivad sisaldada erinevaid atribuutide kogumeid.

Entiteet on tegeliku maailma konkreetne või abstraktne objekt, mida saab üheselt identifitseerida, kirjeldada ja esitada andmete kujul.

Juhendatud õppimine - masinõppe tüüp, mille puhul algoritm treenitakse jaotatud andmetel, kus soovitud tulemus on iga näite puhul teada.

Taksonoomia on hierarhiline klassifitseerimissüsteem, mida kasutatakse elementide süstemaatiliseks liigitamiseks ühiste tunnuste alusel.

Titanic Dataset on populaarne andmekogum masinõppe mudelite treenimiseks ja testimiseks.

Koolitus - protsess, mille käigus masinõppe algoritm analüüsib andmeid, et tuvastada mustreid ja moodustada mudel.

Ülekandeõpe on masinõppe tehnika, mille puhul ühe ülesande jaoks koolitatud mudelit kasutatakse teise ülesande lähtepunktina.

Transformeerimine - andmete vormingu, struktuuri või sisu muutmine hilisemaks kasutamiseks.

Andmenõuded - formaliseeritud kriteeriumid, millega määratakse kindlaks äriprotsesside toetamiseks vajaliku teabe struktuur, vorming, täielikkus ja kvaliteet.

Ehitustööstuse Uberiseerimine on traditsiooniliste ärimudelite ümberkujundamise protsess ehituses digitaalsete platvormide mõjul, mis võimaldavad otsest suhtlemist klientide ja töövõtjate vahel ilma vahendajateta.

Uniclass on Ühendkuningriigis laialdaselt kasutatav teise ja kolmanda põlvkonna ehituselementide klassifitseerimissüsteem.

USD (Universal Scene Description) on arvutigraafika jaoks välja töötatud andmeformaad, mis on tänu oma lihtsale struktuurile ja sõltumatusele geomeetristest südamikutest leidnud rakendust tehnilistes süsteemides.

Andmete valideerimine on protsess, mille käigus kontrollitakse teavet kehtestatud kriteeriumide ja nõuete alusel, et tagada andmete täpsus, täielikkus ja järjepidevus.

Vektorandmebaas - Spetsiaalset tüüpi andmebaas, mis salvestab andmeid mitmemõõtmeliste vektoritena tõhusa semantilise otsingu ja objektide võrdlemise eesmärgil.

Vektorite kujutamine (embedding) on meetod andmete teisendamiseks mitmemõõtmelisteks arvvektoriteks, mis võimaldab masinalgoritmidel teavet tõhusalt töödelda ja analüüsida.

VectorOps on metoodika, mis keskendub mitmemõõtmeliste vektorandmete töötlemisele, salvestamisele ja analüüsile, mis on eriti oluline sellistes valdkondades nagu digitaalsed kaksikud ja semantiline otsing.

Visualiseerimine - andmete graafiline kujutamine teabe paremaks tajumiseks ja analüüsimiseks.

Terminid liigitati tähestikuliselt nende ingliskeelsete nimede järgi.

VIIDETE JA VEEBIMATERJALIDE LOETELU

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics," 12. detsember 2016. [Online]. Kättesaadav: <https://www.gartner.com/en/documents/3530919>. [Kasutamise kuupäev: 1. märts 2025].
- [2] KPMG, "Tuttavad väljakutsed - uued lähenemisviisid. 2023. aasta ülemaailmne ehitusuuring," 1. jaanuar 2023. [Online]. Kättesaadav: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Kasutatud: 5. märts 2025].
- [3] F. R. Barnard, "Pilt on rohkem kui tuhat sõna", 10 Mapi 1927. [Online]. Kättesaadav: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [viide: 15. märts 2025].
- [4] M. Bastian, "Microsofti tegevjuht Satya Nadella ütleb, et AGI enesereklaamimine on "mõttetu võrdlushäkkimine"," 21. veebr. 2025. [Online]. Kättesaadav: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/>. [Kasutamise kuupäev: 15. märts 2025].
- [5] W. E. Foorum, "Foorum "Ehituse tuleviku kujundamine - muutuv maastik:", 1. jaanuar 2016. [Online]. Kättesaadav: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Kasutamise kuupäev: 2. märts 2025].
- [6] C. D. Gillespie, "Savi: Maa põimumine saviajastul", 2024. [Online]. Kättesaadav: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Papüürus 3. sajandist eKr. keel on kreeka keel," 2024. [Online]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Järelevalve: olemasolevate vahendite kasutamine", 1980. [Online]. Kättesaadav: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Aadressi kuupäev: 15. märts 2025].
- [9] PWC, "Andmepõhised andmed Mida õpilased vajavad, et olla edukas kiiresti muutuv as ärimaailmas", 15. veebruar 2015. [Online]. Kättesaadav: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Kasutamise kuupäev: 15. märts 2025].
- [10] Skanska USA, "Sügisese ehitusturu trendid", 2. november 2023. [Online]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.
- [11] "Oxford Essential Quotations (4. trükk)," Oxford University Press, 2016. [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro->

- ed4-00006236. [Kasutamise kuupäev: 1. märts 2025].
- [12] "Tsitaat: Sondergaard andmeanalüüsi kohta," [Online]. Kättesaadav: <https://www.causeweb.org/cause/resources/library/r2493>. [Kasutamise kuupäev: 15. märts 2025].
- [13] "Kuidas ülemaailmne tehisintellekti huvi suurendab andmehaldusturgu," 28. mai 2024. [Online]. Kättesaadav: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Kasutamise kuupäev: 15. märts 2025].
- [14] И. McCue, "ERP ajalugu", 2024. [Online]. Kättesaadav: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "ERP hind: Kui palju maksab ERP-süsteem?," [Online]. Kättesaadav: <https://erpscout.de/en/erp-costs/>. [Kasutamise kuupäev: 15. märts 2025].
- [16] softwarepath, "Mida 1384 ERP-projekti meile ERP valiku kohta ütlevad (2022 ERP-raport)," 18.01.2022. [Online]. Kättesaadav: <https://softwarepath.com/guides/erp-report>. [Kasutamise kuupäev: 15. märts 2025].
- [17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16. detsember 2016. [Online]. Kättesaadav: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Aadressi kuupäev: 1. märts 2025].
- [18] Mckinsey, "The data-driven enterprise of 2025", 28. jaanuar 2022. [Online]. Kättesaadav: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [viitamise kuupäev: 22. mai 2024].
- [19] Wikipedia, "Moore'i seadus," [Online]. Kättesaadav: https://en.wikipedia.org/wiki/Moore%27s_law. [Kasutamise kuupäev: 15. märts 2025].
- [20] Accenture, "Building More Value With Capital Projects", 1. jaanuar 2020. [Online]. Kättesaadav: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf>. [Kasutamise kuupäev: 3. märts 2024].
- [21] Б. Marr, "Kui palju andmeid me iga päev loome? The Mind-Blowing Stats Everyone Should Read," 2018. [Online]. Kättesaadav: <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] "Kui palju andmeid toodetakse iga päev?", 2024. [Online]. Kättesaadav:

- <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]. Kättesaadav: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "Lääne-Euroopa eri piirkondades aastatel 1454-1800 igal poolsajandil toodetud trükitud raamatute koguarv" [Internet]. [Online]. Kättesaadav: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [viitamise kuupäev: 1. märts 2025].
- [25] "Hinnakujunduse näited", 2024. [Online]. Kättesaadav: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Ettevõtted tellivad andmete säilitamise keerukuse kasvades sisse," 10. mai 2024. [Online]. Kättesaadav: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Kasutamise kuupäev: 15. märts 2025].
- [27] JETSOFTPRO, "SaaS on surnud? Microsoft CEO's Shocking Prediction Explained," 13. jaanuar 2025. [Online]. Kättesaadav: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12. detsember 2024. [Online]. Kättesaadav: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Kasutamise kuupäev: 15. märts 2025].
- [29] GoodReads, "Tim Berners-Lee," [Online]. Kättesaadav: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Kasutamise kuupäev: 15. märts 2025].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1. jaanuar 2023. [Online]. Kättesaadav: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Kasutatud: 5. märts 2025].
- [31] И. Deininger, B. Koch, R. Bauknecht ja M. Langhans, "Digitaalsete mudelite kasutamine tootmiskoha dekarboniseerimiseks: näide ehitusmudeli, tootmismudeli ja energiamudeli ühendamisest", 2024. [Online]. Kättesaadav: https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Study_of_Connecting_the_Building_Model_Production_Model_and_Energy_Model.
- [32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY," 1. veebruar 2017. [Online]. Kättesaadav: <https://www.mckinsey.com/~media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi->

reinventing-construction-a-route-to-higher-productivity-full-report.pdf.

- [33] Ehituse töörühm asepeaministrile, "Ehituse ümbermõtestamine", 1. oktoober 2014. [Online]. Kättesaadav: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.
- [34] Forbes, "Without An Opinion, You're Just Another Person With Data," 15. märts 2016. [Online]. Kättesaadav: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Kasutamise kuupäev: 15. märts 2025].
- [35] Wikiquote, "Charles Babbage," [Online]. Kättesaadav: https://en.wikiquote.org/wiki/Charles_Babbage. [Kasutamise kuupäev: 15. märts 2025].
- [36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves," 12. märts 2025. [Online]. Kättesaadav: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Kasutamise kuupäev: 15. märts 2025].
- [37] Kanada Ehitusliit ja KPMG Kanadas, 2021, "Ehitus digitaalses maailmas", 1. mai 2021. [Online]. Kättesaadav: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Kasutamise kuupäev: 5. märts 2025].
- [38] ZCS, "Viienda tööstusrevolutsiooni dekodeerimine". [Online]. Kättesaadav: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Kasutamise kuupäev: 15. märts 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, "Avatud lähtekoodiga tarkvara väärtus", 24. jaanuar 2024. [Online]. Kättesaadav: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Kasutamise kuupäev: 15. märts 2025].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1. september 2008. [Online]. Kättesaadav: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] McKinsey, "Improving construction productivity," [Online]. Kättesaadav: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Aadressi kuupäev: 15. märts 2025].
- [43] A. G. a. C. Syverson, "USA ehitussektori tootlikkuse kummaline ja kohutav tee", 19.01.2023. [Online]. Kättesaadav: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and->

- awful-path-of-productivity-in-the-us-construction-sector/. [Kasutamise kuupäev: 1. märts 2025].
- [44] McKinsey, "Ehituse tootlikkuse saavutamine ei ole enam vabatahtlik," 9. august 2024. [Online]. Kättesaadav: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Kasutamise kuupäev: 5. märts 2025].
- [45] ING Group, "Lagging productivity in construction is driving up construction costs," 12. detsember 2022. [Online]. Kättesaadav: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Aadressi kuupäev: 15. märts 2025].
- [46] M. Berman, "Microsofti tegevjuhi šokeeriv ennustus: "Agendid asendavad kogu tarkvara"," 19. detsember 2024. [Online]. Kättesaadav: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Kasutamise kuupäev: 15. märts 2025].
- [47] Business Insider, "Anthropic's CEO says that in 3 to 6 months, AI will be writing 90% of the code software developers were in charge of," 15 Mar 2025. [Online]. Kättesaadav: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Kasutamise kuupäev: 30. märts 2025].
- [48] Statista, "Andmebaasi haldussüsteemide (DBMS) populaarsuse võrdlus maailmas juuni 2024 seisuga, kategooriate kaupa," juuni 2024. [Online]. Kättesaadav: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Aadressi kuupäev: 15. märts 2025].
- [49] DB-Engines, "DB-Engines Ranking," [Online]. Kättesaadav: <https://db-engines.com/en/ranking>. [viitamise kuupäev: 15. märts 2025].
- [50] "Stack Overflow Developer Survey 2023," 2024. [Online]. Kättesaadav: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [Online]. Kättesaadav: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Struktureeritud ja struktureerimata andmed: mis vahe on?", 2024. [Online]. Kättesaadav: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "Ehitusprojektide andmevormingute PDF-võrdlus," 23. apr. 2024. [Online]. Kättesaadav: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Building Information Modeling Whitepaper site," 2003. [Online]. Kättesaadav: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.

- [55] A. Boyko, "Lobitöö sõjad ja BIMi arendamine. Osa 5: BlackRock on kõigi tehnoloogiate peremees. Kuidas korporatsioonid kontrollivad avatud lähtekoodi," 2024. [Online]. Kättesaadav: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] D. Ushakov, "Otsene modelleerimine - kellele ja milleks seda vaja on? Ülevaade konkureerivatest tehnoloogiatest", 14.11.2011. [Online]. Kättesaadav: https://isicad.net/articles.php?article_num=14805. [Aadressi kuupäev: 02. märts 2025].
- [57] C. Eastman ja A. Cthers, "Eastman, Charles; And Cthers," september 1974. [Online]. Kättesaadav: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Kasutamise kuupäev: 15. märts 2025].
- [58] D. Ushakov, "Otsene modelleerimine - kellele ja milleks seda vaja on? Ülevaade konkureerivatest tehnoloogiatest", 11. november 2011. [Online]. Kättesaadav: https://isicad.net/articles.php?article_num=14805. [viide: 15. märts 2025].
- [59] D. Weisberg, "CAD-i ajalugu", 12. detsember 2022. [Online]. Kättesaadav: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [viide: 15. märts 2025].
- [60] ADSK, "Valge raamat ehitusinfo modelleerimine", 2002. [Online]. Kättesaadav: https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [viide: 15. märts 2025].
- [61] ADSK, "Valge raamat "Ehitusinfo modelleerimine praktikas"," [Online]. Kättesaadav: https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Kasutamise kuupäev: 15. märts 2025].
- [62] A. Boyko, "Lobitöö sõjad ja BIMi arendamine. Osa 2: avatud BIM VS suletud BIM. Euroopa VS ülejäänud maailm," 2024. [Online]. Kättesaadav: <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]. Kättesaadav: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.
- [64] ADSK, "Whitepaper BIM", 2002. [Online]. Kättesaadav: https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Aadressi kuupäev: 15. märts 2025].
- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [Online]. Kättesaadav: https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_B

- enefits_of_MAI.pdf. [Kasutamise kuupäev: 15. märts 2025].
- [66] M. Shacklett, "Struktureeritud ja struktureerimata andmed: 2024. [Online]. Kättesaadav: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Struktureerimata andmete kasvu mõtestamine", 2024. [Online]. Kättesaadav: <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, "Ebapiisava koostalitlusvõime kulude analüüs", 2004. [Online]. Kättesaadav: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf> . [Date of access: 02 2025].
- [69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Kättesaadav: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Kasutamise kuupäev: 15. märts 2025].
- [70] Analyticsindiamag, "6 kõige aeganõudvamat ülesannet andmeteadlastele", 15. mai 2019. [Online]. Kättesaadav: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] BizReport, "Report: Data scientists spend bulk of time cleaning up," 06. juuli 2015. [Online]. Kättesaadav: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Kasutamise kuupäev: 5. märts 2025].
- [72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27. juuli 2015. [Online]. Kättesaadav: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_a_ma_answers/. [Kasutamise kuupäev: 15. märts 2025].
- [73] B. Cyphers ja K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [Online]. Kättesaadav: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information," 1. oktoober 2013. [Online]. Kättesaadav: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Kasutamise kuupäev: 15. märts 2025].
- [75] A. Boyko, "Võitlus avatud andmete eest ehitussektoris. AUTOLISPi, intelliCADi, openDWG, ODA ja openCASCADE ajalugu", 15 05 2024. [Online]. Kättesaadav: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [Aadressi kuupäev: 16 02 2025].
- [76] Vikipeedia, "Microsoft ja avatud lähtekood" [Online]. Kättesaadav:

- https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Kasutamise kuupäev: 15. märts 2025].
- [77] TIME, "The Gap Between Open and Closed AI Models Might Be Shrinking. Here's Why That Matters," 5. november 2024. [Online]. Kättesaadav: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Kasutamise kuupäev: 15. märts 2025].
- [78] The Verge, "More than a quarter of new code at Google is generated by AI," 29. okt. 2024. [Online]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Kasutamise kuupäev: 15. märts 2025].
- [79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing," 15. detsember 2020. [Online]. Kättesaadav: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-töötlemine>. [Kasutatud: 15. märts 2025].
- [80] PWC, "PwC Open Source Monitor 2019", 2019. [Online]. Kättesaadav: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Kasutatud 15 Mar. 2025].
- [81] Travers Smith, "Avatud saladus: avatud lähtekoodiga tarkvara", 2024. [Online]. Kättesaadav: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Kasutamise kuupäev: 15. märts 2025].
- [82] Deloitte, "Andmete edastamise protsess ettevõtete ümberkujundamisel", 2021. [Online]. Kättesaadav: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Viidatud kuupäev: 15. märts 2025].
- [83] gov.uk, "Data Analytics and AI in Government Project Delivery", 20. märts 2024. [Online]. Kättesaadav: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [Kasutamise kuupäev: 5. märts 2025].
- [84] "Quote Origin: Everything Should Be Made As Simple As Possible As Possible, But Not Simpler," 13. mai 2011. [Online]. Kättesaadav: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Kasutamise kuupäev: 15. märts 2025].
- [85] "Transformer (süvaõppe arhitektuur)," [Online]. Kättesaadav: [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [Kasutamise kuupäev: 15. märts 2025].
- [86] "Pythoni pakettide allalaadimise statistika," 2024. [Online]. Kättesaadav: <https://www.pepy.tech/projects/pandas>.

- [87] Intervjuu Bit, "Top 10 Pythoni raamatukogu", 2023. [Online]. Kättesaadav: [https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more](https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more.). [Kasutamise kuupäev: 30. märts 2025].
- [88] "NVIDIA ja HP suurendavad andmetöötlus- ja genereerivat tehisintellekti tööjaamades," 7. märts 2025. [Online]. Kättesaadav: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Kasutamise kuupäev: 15. märts 2025].
- [89] P. Orac, "Kuidas töödelda miljonite ridadega DataFrame'i sekundite jooksul", 2024. [Online]. Kättesaadav: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Mis on Kaggle?", 2024. [Online]. Kättesaadav: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "NVIDIA tegevjuht Jensen Huang Keynote at COMPUTEX 2024," 2. juuni 2024. [Online]. Kättesaadav: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Kasutatud 15. märts 2025].
- [92] "Liikmed: asutajad ja korporatiivsed liikmed", 2024. [Online]. Kättesaadav: <https://www.opendesign.com/member-showcase>.
- [93] A. Boyko, "Muutuste ajastu: IFC on minevik ehk miks ADSK ja teised CAD-müüjad on valmis loobuma IFC-st USA dollari eest, 14 põhifakti", 24. november 2024. [Online]. Kättesaadav: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a>. [Kasutatud: 23. veebruar 2025].
- [94] A. Boyko, "BIM-järgne maailm. Üleminek andmetele ja protsessidele ning kas ehitussektor vajab semantikat, formaate ja koostalitlusvõimet," 20. dets. 2024. [Online]. Kättesaadav: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Aadressi kuupäev: 23. veebruar 2025].
- [95] N. I. o. Health, "NIH STRATEEGILINE PLAAN ANDMETEADUSELE," 2016. [Online]. Kättesaadav: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Kasutamise kuupäev: 23. veebruar 2025].
- [96] Harvard Business Review, "Bad Data Costs the U.S. \$3 Trillion Per Year," 22. september 2016. [Online]. Kättesaadav: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Andmekvaliteedi mõju", 1. jaanuar 2025. [Online]. Kättesaadav: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance

- Considerations During the Design Phase of Construction Projects," [Online]. Kättesaadav: <https://www.wbdg.org/resources/design-for-maintainability>. [Kasutamise kuupäev: 15. märts 2025].
- [99] O. o. D. C. P. a. Oversight, "Sõjaliste süsteemide ja seadmete korrosiooni ennetamise ja kontrolli planeerimise juhend," aprill 2014. [Online]. Kättesaadav: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guide%20book%204%20Feb%2014.pdf>. [Kasutamise kuupäev: 15. märts 2025].
- [100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1. jaanuar 2025. [Online]. Kättesaadav: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "Küünte puudumisel". [Online]. Kättesaadav: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Kasutamise kuupäev: 15. märts 2025].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information," oktoober 2013. [Online]. Kättesaadav: 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. [Kasutamise kuupäev: 15. märts 2025].
- [103] EY, "Tee süsinikdioksiidi neutraalsuse suunas", 10. märts 2023. [Online]. Kättesaadav: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Kasutamise kuupäev: 15. märts 2025].
- [104] PWC, "ESG teadlikkus", 1. juuli 2024. [Online]. Kättesaadav: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Kasutamise kuupäev: 15. märts 2025].
- [105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [Online]. Kättesaadav: <https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO2_arvutamine," 2024. [Online]. Kättesaadav: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.
- [107] McKinsey, "Imagining Construction's Digital Future", 24. juuni 2016. [Online]. Kättesaadav: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [Kasutamise kuupäev: 25. veebruar 2025].
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10. oktoober 2024. [Online]. Kättesaadav: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [viitamise kuupäev: 15. märts 2025].

- [109] SAS, "Andmejärv ja andmeladu - tunne erinevust", [Online]. Kättesaadav: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [viide: 15. märts 2025].
- [110] ADSK, "Ehitusinfo modelleerimine", 2002. [Online]. Kättesaadav: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Aadressi kuupäev: 15. märts 2025].
- [111] A. Boyko, "BIM ajaloo kaart", 2024. [Online]. Kättesaadav: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "BIMi määratlused organisatsioonide ja standardite kaupa", 27. detsember 2023. [Online]. Kättesaadav: <https://encyclopedia.pub/entry/53149>. [Aadressi kuupäev: 5. märts 2025].
- [113] CAD Vendor, "OPEN BIM programm", 2012. [Online]. Kättesaadav: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [kuupäev: 30. märts 2025].
- [114] Vikipeedia, "Tööstuse põhiklassid", [Online]. Kättesaadav: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Kasutamise kuupäev: 15. märts 2025].
- [115] Vikipeedia, "IGES", [Online]. Kättesaadav: <https://en.wikipedia.org/wiki/IGES>. [Viidete kuupäev: 30. märts 2025].
- [116] A. Boyko, "CAD (BIM) ajalugu", 15. detsember 2021. [Online]. Kättesaadav: https://miro.com/app/board/o9J_laML2cs=/. [Aadressi kuupäev: 24. veebruar 2025].
- [117] T. K. K. A. O. F. B. C. E. L. H. 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, "IFC-tarkvara toetuse võrdlusuuring: GeoBIM benchmark 2019 - I osa", 8. jaanuar 2021. [Online]. Kättesaadav: <https://arxiv.org/pdf/2007.10951>. [Kasutamise kuupäev: 5. märts 2025].
- [118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM", 13. aprill 2021. [Online]. Kättesaadav: <https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be>. [Aadressi kuupäev: 5. märts 2025].
- [119] D. Ares, "RETS kinnisvaras: miks see on tõhususe ja kasvu jaoks ülioluline", 17. detsember 2024. [Online]. Kättesaadav: <https://www.realalpha.com/blog/rets-importance-in-real-estate-explained>. [Viidatud kuupäev: 5. märts 2025].
- [120] "Flex tokeni maksumus", 2024. [Online]. Kättesaadav: <https://www.autodesk.com/buying/flex?term=1-YEAR&tab=flex>.

- [121] A. Boyko, "Forget BIM and democratise access to data (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024. [Online]. Kättesaadav: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Bauprozessmanagement-und-Bay-Bauindustrie.pdf>.
- [122] D. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh ja F. Plaschke, "Solving the construction industry productivity puzzle," 2015. [Online]. Kättesaadav: <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. jaanuar 2018. [Online]. Kättesaadav: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Date of access: 2 Mar 2025].
- [124] Apple.com, "Pixar, Adobe, Apple ja NVIDIA moodustavad OpenUSD alliansi, et edendada 3D-sisu avatud standardeid," 1. august 2023. [Online]. Kättesaadav: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Kasutatud: 2. märts 2025].
- [125] AECmag, "ADSK granuleeritud andmete strateegia", 25. juuli 2024. [Online]. Kättesaadav: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [Kasutamise kuupäev: 15. märts 2025].
- [126] A. Boyko, "The Age of Change: IFC on minevik ehk miks ADSK ja teised CAD-müüjad on valmis loobuma IFC-st USD eest 14 põhifakti," 24 11 2024. [Online]. Kättesaadav: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a>. [Aadressi kuupäev: 23 fevryall 2025].
- [127] A. Boyko, "ENG BIM Cluster 2024 | Võitlus andmete ja LLM ja ChatGPT kohaldamise eest ehituses", 7. august 2024. [Online]. Kättesaadav: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [Kasutatud: 15. märts 2025].
- [128] "Jeffrey Zeldman esitleb", 6. mai 2008. [Online]. Kättesaadav: <https://zeldman.com/2008/05/06/content-precides-design/>. [Kasutamise kuupäev: 15. märts 2025].
- [129] A. Boyko, "DWG Analyse with ChatGPT | DataDrivenConstruction," 5. märts 2024. [Online]. Kättesaadav: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [viide: 15. märts 2025].
- [130] McKinsey, "The McKinsey guide to outcompeting in the age of digital and AI," 2023. [Online]. Kättesaadav: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Aadressi kuupäev: 30. märts 2025].

- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs," 31. märts 2016. [Online]. Kättesaadav: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Kasutamise kuupäev: 15. märts 2025].
- [132] J. Bertin, "Graafika ja graafiline infotöötlus", 8. september 2011. [Online]. Kättesaadav: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=cqX_xnm4tcC&redir_esc=y. [Kasutamise kuupäev: 15. märts 2025].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking," [Online]. Kättesaadav: <https://www.causeweb.org/cause/resources/library/r1266>. [Kasutamise kuupäev: 15. märts 2025].
- [134] Ministrymagazine, "Kuidas teadus avastas loomise", jaanuar 1986. [Online]. Kättesaadav: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [viide: 15. märts 2025].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now," 23. mai 2017. [Online]. Kättesaadav: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Kasutatud 15. mai 2024].
- [136] "Kuidas luua andmete arhitektuur, et edendada innovatsiooni - täna ja homme", 3. juuni 2020. [Online]. Kättesaadav: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Aadressi kuupäev: 15. märts 2025].
- [137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630>. [Aadressi kuupäev: 15. märts 2025].
- [138] "Convertors", 2024. [Online]. Kättesaadav: <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. jaanuar 2017. [Online]. Kättesaadav: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Kasutamise kuupäev: 18. veebruar 2025].
- [140] "Torustik ehituses", 2024. [Online]. Kättesaadav: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Wikipedia, "Apache NiFi," 1. jaanuar 2025. [Online]. Kättesaadav: https://de.wikipedia.org/wiki/Apache_NiFi. [Kasutamise kuupäev: 5. märts 2025].
- [142] n8n, "Gmail AI Auto-Responder: Luua vastuste eelnõud sissetulevatele e-kirjadele," 1. mai 2024.

- [Online]. Kättesaadav: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [Kasutamise kuupäev: 15. märts 2025].
- [143] n8n, "Kinnisvara päevapakkumiste automatiseerimine Zillow API, Google Sheets ja Gmaili abil", 1. märts 2025. [Online]. Kättesaadav: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [Date of access: 15 Mar 2025].
- [144] B. T. O'Neill, "Analüütika, tehisintellekti ja suurte andmeprojektide ebaõnnestumise määr = 85% - jubedalt!", 1. jaanuar 2025. [Online]. Kättesaadav: <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. jaanuar 2025. [Online]. Kättesaadav: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Kasutamise kuupäev: 1. mai 2024].
- [147] T. Landsall-Welfair, Prognoosides rahva praegust meeleolu, Significance, 2012.
- [148] A. Boyko, "San Francisco. Ehitussektor 1980-2019", 2024. [Online]. Kättesaadav: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC failid 5000 projekti," 2024. [Online]. Kättesaadav: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Tuleviku ettevalmistamine ühendatud ehitusega," [Online]. Kättesaadav: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [Kasutatud 15. märts 2025].
- [151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing," 22. mai 2017. [Online]. Kättesaadav: <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] "Taimsete fossiilide säilitamiseks vajalikud tingimused", 2024. [Online]. Kättesaadav: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "BlackRocki Fink võlakirjadest, ühinemistest ja ülevõtmistest, USA majanduslangusest ja valimistest: Täielik intervjuu," 2023. [Online]. Kättesaadav: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.

- [154] cio, "12 kuulsat AI-katastroofi", 02. oktoober 2024. [Online]. Kättesaadav: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [Kasutamise kuupäev: 15. märts 2025].
- [155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.
- [156] Wikipedia, "Kaggle," 1. jaanuar 2025. [Online]. Kättesaadav: <https://en.wikipedia.org/wiki/Kaggle>. [Kasutamise kuupäev: 15. märts 2025].
- [157] Kaggle, "Titanic - masinõpe katastroofist", 1. jaanuar 2025. [Online]. Kättesaadav: <https://www.kaggle.com/competitions/titanic/overview>. [Kasutatud 10. märts 2025].
- [158] Ш. Johri, "Creating ChatGPT: andmetest dialoogini", 2024. [Online]. Kättesaadav: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Domingos, "Mõned kasulikud asjad masinõppe kohta", 2024. [Online]. Kättesaadav: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Quotable Quote," [Online]. Kättesaadav: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Kasutamise kuupäev: 17. märts 2025].
- [161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models," 19. detsember 2024. [Online]. Kättesaadav: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Kasutatud 25. märts 2025].
- [162] "NVIDIA annab teada Cosmos World Foundationi mudelite ja füüsilise tehisintellekti andmetööriistade suuremast versioonist," 18. märts 2025. [Online]. Kättesaadav: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Kasutamise kuupäev: 25. märts 2025].
- [163] NVIDIA, "NVIDIA Isaac Sim," [Internetis]. Kättesaadav: <https://developer.nvidia.com/isaac/sim>. [Kasutatud 25. märts 2025].
- [164] M. Quarterly, "Miks digistrateegiad ebaõnnestuvad", 25. jaanuar 2018. [Online]. Kättesaadav: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Kasutamise kuupäev: 15. märts 2025].
- [165] M. J. Perry, "My Favorite Milton Friedman Quotes," 17. november 2006. [Online]. Kättesaadav: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Kasutamise kuupäev: 1. märts 2025].
- [166] J. A. Wheeler, "Informatsioon, füüsika, kvant: seoste otsimine", 1990.

- [169] A. Boyko, "Lobitöö sõjad ja BIMi arendamine. 5. osa: BlackRock on kõigi tehnoloogiate peremees. Kuidas korporatsioonid kontrollivad avatud lähtekoodi," 2024. [Online]. Kättesaadav: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen ja J. Beetz, "A SPARQL query engine for binary-formatted IFC building models," *Advanced Engineering Informatics*, 2024.
- [171] "Ehitussektori ettevõtete arv Ühendkuningriigis 2021. aastal ettevõtte suuruse järgi," 2024. [Online]. Kättesaadav: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5000 IFC&RVT projekti," 2024. [Online]. Kättesaadav: <https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io>.
- [173] M. Popova, "It from Bit: Pioneeri füüsik John Archibald Wheeler informatsioonist, reaalsuse olemusest ja sellest, miks me elame osalusuniversumis", 2008. [Online]. Kättesaadav: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Viitedokumentide kuupäev: veebruar 2025].
- [174] *Lobisõjad andmete üle ehituses | Tehno-Feudalism ja BIMi varjatud mineviku ajalugu*. [Film]. Saksamaa: Artjom Boiko, 2023.
- [175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Dokumentide ja andmete automaatne otsimine projektidest", 16. nov. 2023. [Online]. Kättesaadav: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Kasutamise kuupäev: 2. märts 2025].
- [176] M. & Company, "Kolm uut mandaati digitaalse ümberkujundamise täieliku väärtuse haaramiseks," 22. jaanuar 2022. [Online]. Kättesaadav: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Kasutamise kuupäev: 15. veebruar 2025].
- [177] KPMG, "Ehitus digitaalses maailmas", 1. mai 2021. [Online]. Kättesaadav: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Kasutamise kuupäev: 5. aprill 2024].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time," 17. märts 2023. [Online]. Kättesaadav: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Kasutamise kuupäev: 15. veebruar 2025].
- [179] O. Business, "Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models'," 10 Jan. 2025. [Online]. Kättesaadav: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Kasutamise kuupäev: 15. märts 2025].

- [180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," 23. märts 2016. [Online]. Kättesaadav: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [Kasutamise kuupäev: 15. märts 2025].
- [181] Ühendkuningriigi Välis-, Rahvaste Ühenduse ja Arenguamet, "Digitaalse arengu strateegia 2024-2030", 18. märts 2024. [Online]. Kättesaadav: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [Kasutamise kuupäev: 15. märts 2025].
- [182] "Visioon ja strateegia ehitusdisainitööstuses", 7. november 2003. [Online]. Kättesaadav: <https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Aadressi kuupäev: 5. märts 2025].
- [183] M. Bocharov, "Informatsiooni modelleerimine", märts 2025. [Online]. Kättesaadav: <https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5>. [Aadressi kuupäev: 15. märts 2025].
- [184] "Integreeritud projekteerimine läbi tootmise: eelised ja põhjendused", 2000. [Online]. Kättesaadav: https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Viitedokumentide kuupäev: 25. märts 2025].
- [185] CAD Vendor, "Open BIM programm on turunduskampaania", 12. märts 2012. [Online]. Kättesaadav: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [Aadressi kuupäev: 30. märts 2025].

AINETE INDEKS

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477.

AIA, 289

AIM, 3, 289

AMS, 14, 84, 151, 153

Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494.

Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494

Apache ORC, 62, 330, 378

Apache Parkett, 62, 67, 380, 381, 494

API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

B

BDS, 258, 259

Big Data, 9, 67, 245, 383

BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 389, 413, 416, 448, 462, 466,

475, 476, 492, 494, 497

BlackBox, 240, 242, 243.

BMS, 8

Bokeh, 320, 337

BOM, 76, 77, 79, 257, 263

Bounding Box, 234, 373, 392, 393, 394, 414, 415, 480.

BREP, 142, 234, 263, 264, 276, 283, 284.

C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

CAE, 16, 283

CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473.

CAM, 16, 78

CAPEX, 14, 82

CDE, 84, 175, 388, 389, 390

ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421.

Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443.

CO₂, 72

CO₂, 229, 234, 235, 236, 237, 238, 239, 243

COBie, 156, 289, 292.

Euroopa Nõukogu, 56, 168, 169, 170, 477

Kopiloot, 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.
 Andmehaldus, 373, 395, 396, 398, 400, 401
 Andmejärv, 214, 373, 376, 384, 385, 386, 387, 388, 389,
 390, 400
 Data Lakehouse, 67, 373, 386, 387, 388
 Andmeminimalism, 373, 395, 396, 400, 401
 Andmesoo, 373, 395, 397, 401
 Andmeladu, 382, 383, 400
 Andmed kui teenus, 487
 andmepõhine, 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.
 Väljavõte, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330,
 345, 361, 362, 363, 365, 383, 384, 450, 475.

F

Sulg, 62, 123, 330, 378
 FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
 GIS, 58
 GLTF, 143, 278
 Google'i lehed, 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, 273, 276, 277, 278, 279, 280,
 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497.
 IGES, 262, 263, 276.
 iLOD, 289
 IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419,
 455, 460, 465, 482, 484.
 ISO 19650, 388

J

JavaScript, 320, 369, 378
 JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329,
 330, 333, 378, 474, 480
 Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346,
 417, 425.

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408,
 415, 417, 425, 426, 430, 431, 433
 k-NN, 392, 393, 442, 445, 446, 447, 448
 KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389,
 478, 479.

L

LEED, 235, 236, 238.
 LLaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494

LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 166, 187, 219, 220, 220, 221, 222, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494

Load, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475.

LOD, 287, 289

LOI, 287

LOMD, 287

Low-Code, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494

MCAD, 77, 78, 257, 284

MEP, 14, 175, 492

VÕRGUSILMAD, 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.

avatud BIM, 142, 216, 219, 256, 261, 278, 291

Avatud lähtekood, 45, 55, 97, 98, 107, 108, 273, 275, 494

ÖÖKULL, 267, 268, 269.

P

Pandad, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494.

Parkett, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480.

PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497

PDM, 16

PHP, 63, 339

Torustik, 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

erakapital, 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, 479

RPM, 14, 84, 151, 331, 333, 334

RVt, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329, 357, 376, 414, 415, 417, 497.

S

SaaS, 24, 50, 51, 52, 458
ULATUS, 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.
SMM, 261, 262, 263, 266, 272, 276, 277, 292
Streamlit, 336, 337
SVF, 142, 276, 285

T

Transform, 128, 193, 311, 323, 325, 326, 330, 331, 333,
334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367,
383, 384, 450, 475.

U

Uniklass, 154, 155, 156.
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285,
286, 292, 296

V

VectorOps, 373, 398, 400, 401
VR, 84, 271, 285

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296,
299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291,
299, 329, 378, 414, 474.