

Semantična podatkovna plast: manjkajoči temelj analitike nove generacije

Jure Jeraj,¹ Stevanče Nikoloski^{1,2}

¹ DataQ d.o.o., Ljubljana, Slovenija

jure.jeraj@dataq.si, stevance.nikoloski@dataq.si

² Univerza v Novem mestu, Fakulteta za ekonomijo in informatiko, Novo mesto, Slovenija

Organizacije nimajo več le preveč podatkov, temveč pogosto tudi preveč dashboardov in poročil, med katerimi je težko prepoznati, kaj zares prinaša vrednost ali opozarja na problem. Prispevek obravnava semantično podatkovno plast kot skupno pomensko osnovo nad preverjenimi podatki: upravljani sloj metrik, dimenzij, pojmov, odnosov in pravil, ki ga lahko dosledno uporabljajo BI-orodja, aplikacije in sistemi umetne inteligence. Prikažemo razvoj od podatkovnega skladišča in analitičnih modelov do ponovno uporabne semantike ter pojasnimo razliko med zvezdno shemo, UML-modelom, ontologijo in grafom znanja. Referenčno arhitekturo ponazorimo s primerom analize prodajne marže. Na tej podlagi predstavimo podatkovnega asistenta, ki pregleda več analitičnih pogledov, izlušči pomemben signal in predlaga fokus, vendar odločitve ne izvede sam. Prispevek predstavlja praktičen arhitekturni okvir in ne empirične validacije konkretnega produkta.

Ključne besede:

semantična podatkovna plast

poslovna inteligenca

ontologija

graf znanja

podatkovni asistent

1 Ko preveč podatkov postane preveč dashboardov

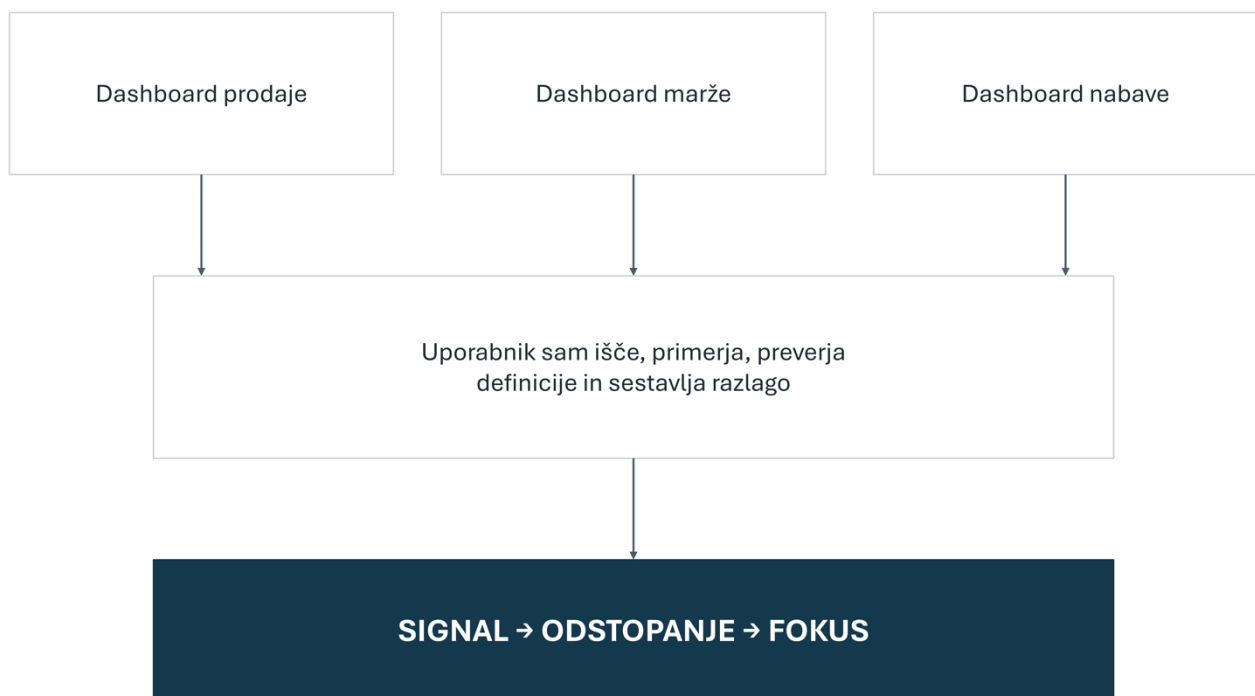
V zadnjih desetletjih so organizacije zgradile obsežne podatkovne in analitične ekosisteme, ki vključujejo podatkovna skladišča, podatkovna jezera, analitične modele, poročila in nadzorne plošče. Z razmahom samopostrežne analitike problem ni več le velika količina podatkov. Gartner opozarja na množenje dashboardov, njihovo omejeno uporabo in omejen poslovni učinek ter na podvajanje dashboardov, poročil in metrik v decentraliziranih analitičnih okoljih [1, 2].

Osnovna poslovna vprašanja se pri tem niso spremenila. Koliko je vredna dobra odločitev? Koliko nas stane napačna odločitev zaradi napačno razumljene metrike? Koliko izgubimo, če pomemben signal opazimo prepozno ali svojega dejanskega stanja sploh ne razumemo? Analitika ustvarja vrednost šele, ko uporabniku pomaga pravočasno prepoznati problem, razumeti njegov obseg in usmeriti nadaljnjo presojo.

Uporabnik zato pogosto nima težave z dostopom do podatkov, temveč z ugotavljanjem, kateri od številnih podatkov in analitičnih pogledov mu v danem trenutku prinašajo vrednost oziroma ga opozarjajo na problem. Dashboard je vnaprej pripravljen pogled na podatke. Ne ve pa, kateri od teh pogledov je trenutno najpomembnejši, in sam po sebi ne pove, koliko nas bo stalo, če pomemben signal spregledamo.

Naslednji korak analitike ni še en dashboard. Najprej potrebujemo skupen in upravljan pomen podatkov.

Podatki sami ne govorijo poslovnega jezika, poslovni uporabniki pa praviloma ne razmišljajo v tabelah, ključih in poizvedbah. Med obema svetovoma zato potrebujemo skupen jezik. Semantična podatkovna plast formalizira in poveže tisto, kar v organizacijah že obstaja: metrike, dimenzije, hierarhije, poslovne pojme, pravila, izjeme in odnose. Ko je ta pomen upravljan in ponovno uporaben tudi zunaj posameznega poročila ali BI-orodja, postane temelj analitike nove generacije.



Slika 1: Od množice analitičnih pogledov do poslovnega fokusa.

2 Semantika ni nova, vendar je postala pomembnejša

2.1. Od skupne podatkovne osnove do analitičnega modela

Podatkovno skladišče je organizacijam omogočilo, da iste očiščene in povezane podatke z ohranjeno zgodovino uporabljajo v več poročilih in analizah. Njegova vrednost ni le v tabelah, temveč v skupni podatkovni osnovi in disciplinah, kot so določanje granulacije, ločevanje dejstev in dimenzij ter dosledna uporaba ključev in zgodovine [3]. DWH zato ne postavljamo nasproti semantični plasti, temveč ga obravnavamo kot njen pomemben podatkovni temelj.

Tudi poslovna semantika v analitiki ni nova. Večdimenzionalne kocke, SQL Server Analysis Services in tabularni modeli že dolgo opredeljujejo mere, dimenzije, hierarhije, relacije in kazalnike. Microsoftova dokumentacija tabularne modele danes izrecno opisuje kot okolje za ustvarjanje objektov semantičnega modela, kot so tabele, relacije, hierarhije, mere in KPI-ji [4]. Power BI oziroma Microsoft Fabric uporablja izraz semantični model (semantic model) za logični opis analitične domene z metrikami, poslovno razumljivimi poimenovanji, dejstvi in dimenzijami [5].

Zato DWH in BI ne moremo obravnavati kot okolji brez semantike. Dobro zasnovan dimenzijski model že veliko pove o poslovnem dogodku, analitični model pa formalizira pomemben del mer in pravil. Današnji problem je predvsem v tem, da ta pomen pogosto ostane vezan na posamezen model, poročilo ali tehnološko okolje.

2.2. Kje obstoječi pristop začne omejevati

Razliko med razpršeno in upravljanjo semantiko bomo v prispevku ponazarjali na primeru izračuna in analize prodajne marže. Gre za dovolj znan poslovni kazalnik, ki hitro pokaže omejitve obstoječega pristopa: organizacija mora določiti, katero vrsto marže uporablja, katere postavke vključuje, kateri datum uporabi za analizo in kako jo pravilno agregira. Isti primer bomo v nadaljevanju uporabili tudi pri razlagi ontologije, grafa znanja in podatkovnega asistenta.

Del logike za izračun marže je lahko centraliziran že v DWH. V njem so lahko pripravljene neto prihodki, nabavna vrednost, pretvorbe valut, zgodovinska klasifikacija izdelkov in druge preverjene komponente izračuna. DWH zato zagotavlja skupno, centralizirano osnovo, česar lokalna SQL-poizvedba ali izračun v posameznem poročilu ne moreta. Kljub temu celotna poslovna definicija kazalnika pogosto ostane razpršena: končna formula je lahko zapisana v DAX-meri, posebni filtri v SQL-poizvedbi, izjeme pa v dokumentaciji ali znanju analitika. Posamezni oddelki lahko isto metriko opredelijo različno, aplikacija zunaj BI-orodja pa teh pravil morda sploh ne more ponovno uporabiti.

Problem torej ni nujno odsotnost centraliziranih podatkov, temveč odsotnost enotne, upravljanje in ponovno uporabne definicije kazalnika. Sistem umetne inteligence lahko vidi fizično shemo, vendar iz tehničnih imen tabel in stolpcev ne more zanesljivo razbrati vseh poslovnih pravil in izjem.

Premik, ki ga obravnavamo, zato ni prehod od sveta brez pomena v svet s pomenom. Gre za prehod od razpršene, deloma implicitne in posameznemu orodju prilagojene semantike k skupni, upravljeni in ponovno uporabni pomenski osnovi. Tudi novejši arhitekturni pristopi semantično plast obravnavajo kot samostojen povezovalni sloj med podatkovnimi strukturami in različnimi načini uporabe podatkov [6].

3 Semantična podatkovna plast: skupna pomenska osnova

3.1. Delovna definicija

V prispevku semantično podatkovno plast opredelimo kot upravljeni logični sloj poslovnih metrik, dimenzij, pojmov, odnosov in pravil, ki različnim porabnikom omogoča enako razumevanje in uporabo istih podatkov. Gre za avtorski referenčni okvir. Ne pomeni nujno enega samega produkta, repozitorija ali poizvedovalnega jezika. Posamezne komponente so lahko fizično porazdeljene, vendar morajo skupaj delovati kot jasno upravljan vir poslovnega pomena.

Na analitični ravni takšna plast opredeljuje predvsem metrike, dimenzije, hierarhije, agregacijska in časovna pravila, dovoljene povezave ter omejitve dostopa. Na širši domenski ravni lahko vključuje tudi poslovne pojme, njihove odnose, sinonime, pravila in povezave do konkretnih entitet ali dokumentov. Teh dveh ravni ne predstavljamo kot splošno sprejetega industrijskega standarda, temveč kot praktičen okvir za razumevanje različnih arhitekturnih vlog.

3.2. Kaj pomeni semantika na primeru prodajne marže

Na primeru marže lahko konkretno pokažemo, kaj semantika pomeni v praksi. Pojem marža je na prvi pogled preprost, vendar organizacija potrebuje bistveno več kot ime stolpca. Semantična definicija vključuje formulo, časovno pravilo, valuto, dovoljene dimenzije, pravilo agregiranja, izjeme, lastnika in obdobje veljavnosti.

V tabeli 1 je prikazana ponazoritvena definicija prodajne marže. Njen namen ni določiti univerzalne formule, saj se ta med organizacijami razlikuje, temveč pokazati vrste pomena in pravil, ki jih mora semantična plast upravljati.

Tabela 1: Ponazoritvena semantična definicija prodajne marže.

Element definicije	Ponazoritvena vrednost
Poslovno ime	Prodajna marža [%]
Poslovni namen	Izračun razlike med neto prihodki in izbranimi neposrednimi oziroma spremenljivimi stroški.
Formula za znesek	Neto prihodek - nabavna vrednost - spremenljivi logistični stroški - prodajni rabati.
Stopnja marže	Prodajna marža / neto prihodek.
Časovno pravilo	Uporabi se datum knjiženja.
Valuta	Vrednosti se pred agregiranjem preračunajo v skupno poročevalsko valuto.
Dovoljene dimenzije	Čas, izdelek, kupec, segment in prodajni kanal.
Agregacijsko pravilo	Zneski se seštevajo; odstotek se na vsaki ravni ponovno izračuna.
Poslovni lastnik	Kontroling oziroma druga imenovana poslovna funkcija.
Obdobje veljavnosti	Definicija ima datum začetka in po potrebi datum konca veljavnosti.
Izjeme	Posebna pravila za vračila, pozne rabate ali izključene vrste prodaje.

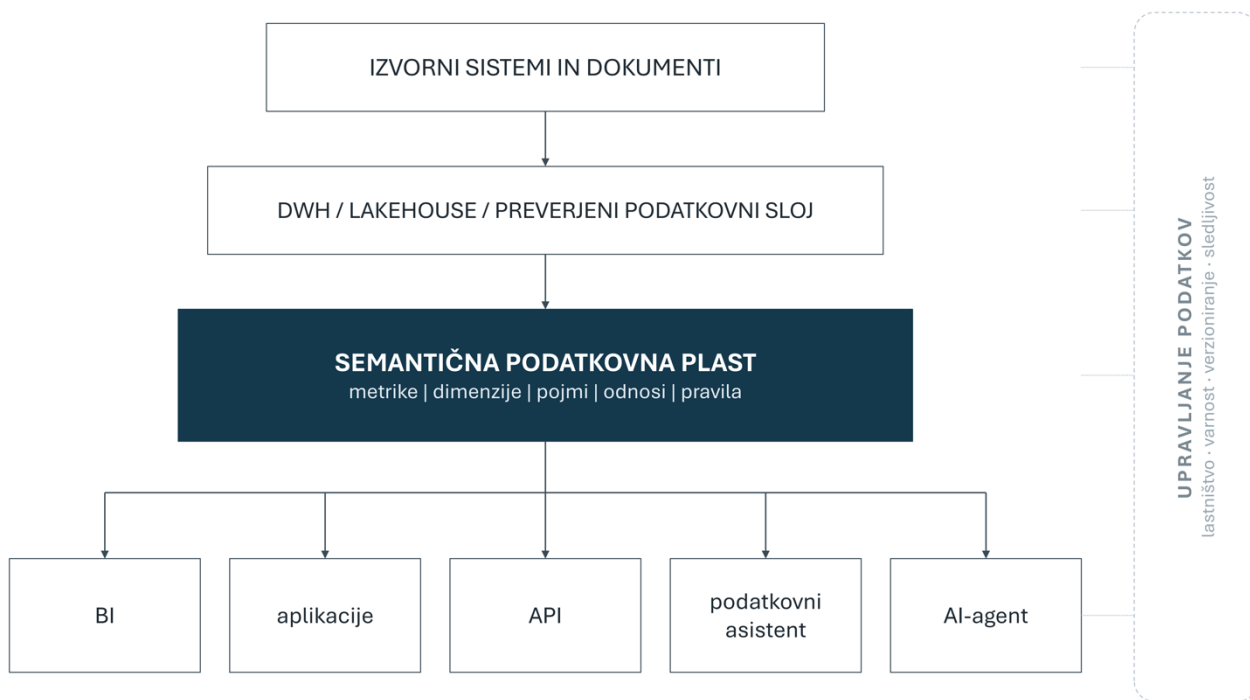
Agregacijsko pravilo je posebej pomembno. Odstotkov marže ne smemo preprosto seštevati ali povprečiti brez upoštevanja prihodkov. Če ima prvi segment 20 enot marže pri 100 enotah prihodka, je njegova marža 20 %, drugi pa 10 enot marže pri 200 enotah prihodka oziroma 5 %. Skupna marža zato ni 25 % (20 % + 5 %) in tudi ne 12,5 % kot preprosto povprečje obeh odstotkov. Pravilno jo izračunamo kot vsoto zneskov marže, deljeno z vsoto zneskov prihodka: $(20 + 10) / (100 + 200) = 10 \%$.

3.3. Razlika med DWH in semantično plastjo

Tabela 2: DWH in semantična podatkovna plast – različni, vendar povezani vlogi.

DWH oziroma lakehouse	Semantična podatkovna plast
Integrira in čisti podatke ter ohranja njihovo zgodovino.	Formalizira njihovo dogovorjeno poslovno uporabo.
Vsebuje dejstva, dimenzije, attribute in povezave.	Vsebuje metrike, pojme, hierarhije, pravila in izjeme.
Zagotavlja predvsem skupno podatkovno osnovo .	Zagotavlja predvsem skupno pomensko osnovo .
Optimiziran je za hranjenje, obdelavo in analitične poizvedbe.	Optimizirana je za konsistentno interpretacijo več porabnikov.
Poslovni pomen pogosto izraža posredno s podatkovnim modelom in transformacijami.	Poslovni pomen naredi jasen, upravljan in ponovno uporaben.

Meja med DWH in semantično plastjo ni absolutna. Semantična plast ne popravi slabega modeliranja in ne nadomesti preverjenih podatkov. Njena vrednost je v tem, da dogovorjeni pomen loči od posameznega dashboarda in ga ponudi različnim porabnikom: BI-orodjem, analitičnim aplikacijam, programskim vmesnikom in sistemom umetne inteligence.



Slika 2: Semantična plast kot skupna pomenska osnova za več porabnikov (referenčni okvir avtorjev).

4 Od zvezdne sheme do ontologije in grafa znanja

4.1. Štirje modeli, štiri različna vprašanja

Zvezdna shema, UML-model, ontologija in graf znanja lahko opisujejo isto poslovno domeno, vendar niso zamenljivi. Zvezdna shema je zasnovana za analizo dejstev po dimenzijah. UML je standardni jezik za modeliranje strukture, vedenja in arhitekture sistemov ter lahko opisuje tudi poslovne procese in podatkovne strukture [7]. Ontologija formalizira besednjak domene, pojme in njihove odnose; standard OWL na primer opredeljuje razrede, lastnosti, posameznike in podatkovne vrednosti s formalno določenim pomenom [8]. Graf znanja pa poveže konkretne entitete in odnose, pogosto iz več virov, pri čemer se lahko opira na strogo ontologijo ali ohlapnejšo shemo [9].

Tabela 3: Primerjava zvezdne sheme, UML-modela, ontologije in grafa znanja.

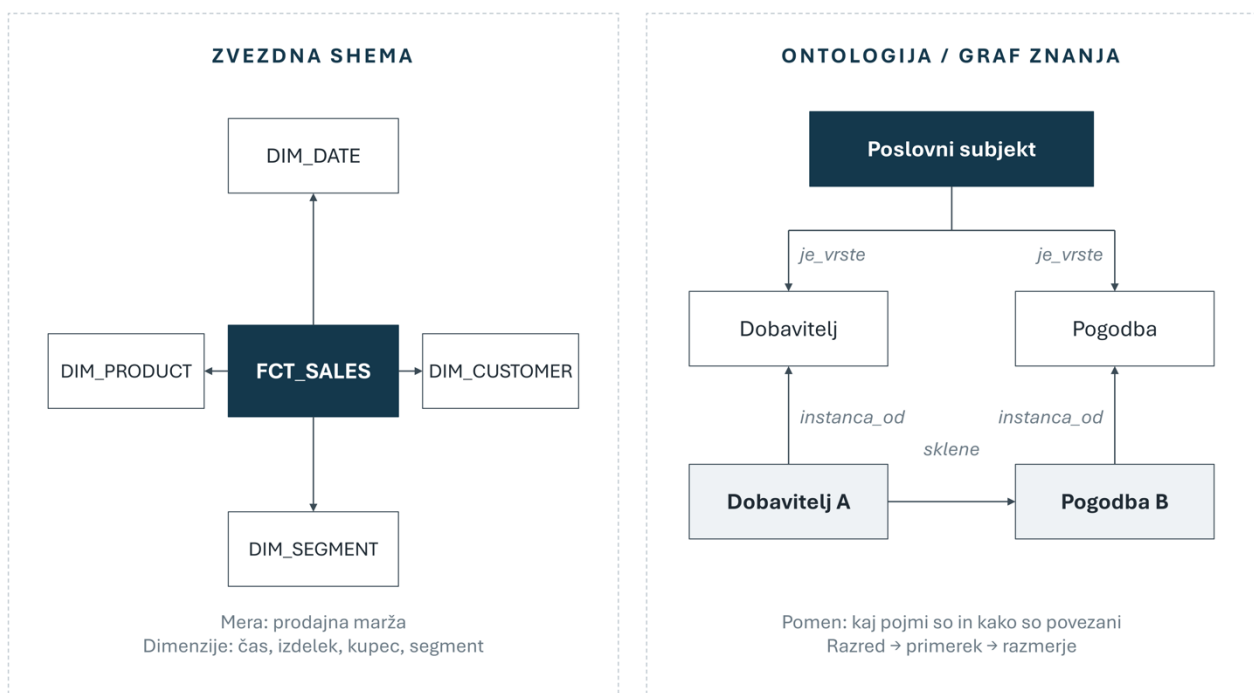
Koncept	Glavno vprašanje	Kaj dobro opiše	Primer
Zvezdna shema	Kako učinkovito analiziramo dogodek?	Dejstva, dimenzije, granulacijo, hierarhije in agregacije.	Prodaja po času, izdelku, kupcu in segmentu.
UML-model	Kako je domena ali sistem strukturiran?	Razrede, attribute, povezave, kardinalnosti in vedenje.	Kupec odda naročilo; naročilo vsebuje izdelek.
Ontologija	Kateri pojmi in odnosi veljajo ter kaj pomenijo?	Formalni besednjak, taksonomije, relacije, omejitve in pravila.	Pogodba določa cenik; cenik velja za skupino izdelkov.
Graf znanja	Katere konkretne entitete so povezane?	Instance pojmov, dejanske povezave, identiteto, kontekst in izvor.	Dobavitelj A ima pogodbo B, ki uporablja cenik C.

4.2. Zakaj relacijska shema ni isto kot ontologija

Tuji ključ v relacijski bazi pove, da sta zapisa povezana. Praviloma pa ne pojasni celotnega poslovnega pomena povezave, njenega namena, sinonimov, dovoljenih kombinacij ali pravil sklepanja. Tudi UML lahko zelo dobro prikaže razrede in relacije, vendar je pogosto predvsem dokumentacijski ali načrtovalski model. Ontologija gre korak dlje, kadar želimo pomen predstaviti v formalizirani obliki, ki jo lahko različni sistemi neposredno uporabljajo in preverjajo z mehanizmi za avtomatsko sklepanje ali validacijskimi pravili.

To ne pomeni, da mora vsaka organizacija zgraditi obsežno formalno ontologijo. Za dosledne KPI-je in standardno poročanje je semantični model metrik pogosto dovolj. Ontologija in graf znanja postaneta smiselna predvsem v kompleksnih domenah, kjer je treba povezovati več vrst entitet, časovno veljavne odnose, dokumente, pravila in dogodke.

4.3. Konkretni primer



Slika 3: Analitični pogled in širši domenski kontekst istega primera (konceptualna skica avtorjev).

Analitični pogled na levi dobro pokaže, kje in pri katerih izdelkih se je marža spremenila. Domenski pogled na desni pomaga povezati odstopanje z dobaviteljem, pogodbo, cenikom in drugim poslovnim kontekstom. Semantična podatkovna plast lahko uporabi oba pogleda: metrični model za zanesljiv izračun in domenski model za razlago odnosov.

5 Referenčna arhitektura in tehnološki vzorci

5.1. Semantična plast ni en sam produkt

Semantična plast se v praksi pojavlja v več arhitekturnih oblikah. Nekatere so tesno vgrajene v BI-platforno, druge živijo v podatkovni platformi, tretje delujejo kot samostojne programske dostopne storitve, četrte pa semantične definicije povezujejo s transformacijskimi modeli. Rešitve se deloma prekrivajo, vendar poudarjajo različne porabnike, načine upravljanja in obseg operativnega delovanja.

Pri primerjavi implementacij so zato pomembna predvsem štiri vprašanja: kje je model definiran in upravljan; kdo ga uporablja – BI, aplikacije, sistemi umetne inteligence ali vsi skupaj; ali rešitev ponuja samo definicije ali tudi izvajalno okolje (runtime); ter ali model zgolj opisuje podatke ali podpira tudi poslovna dejanja.

5.2. BI-vgrajeni in platformno vgrajeni modeli

Analysis Services in Power BI ponazarjata razvoj analitičnega semantičnega modela blizu poslovnega poročanja. Power BI Semantic Model je samostojen analitični objekt, v katerem so opredeljeni odnosi, mere, hierarhije, varnost in predstavitev podatkov za poročila ter druge odjemalce [4, 5]. To je močan in zrel vzorec, zlasti kadar je glavni način uporabe znotraj Microsoftovega analitičnega okolja.

Snowflake Semantic Views ponazarjajo drug vzorec: poslovni koncepti, metrike, entitete in relacije so shranjeni kot objekti na ravni podatkovne sheme. Snowflake jih uporablja za dosledne definicije v aplikacijah, za poizvedovanje z ukazom SELECT in kot kontekst za Cortex Analyst pri vprašanjih v naravnem jeziku [10]. Ta primer pokaže, da sposobnost generiranja SQL-a sama po sebi ni dovolj; rešitev potrebuje tudi poslovno semantiko nad fizičnimi tabelami.

5.3. Semantika kot samostojna storitev: headless, API-first in code-first

Headless semantična plast nima lastnega osrednjega uporabniškega oziroma predstavitvenega sloja. Njena naloga je, da metrike, dimenzije in pravila ponudi drugim porabnikom, kot so BI-orodja, aplikacije in sistemi umetne inteligence.

Pristop code-first pomeni, da so semantični modeli zapisani kot koda ali konfiguracija, na primer v YAML-u. Zato je mogoče upravljati njihove različice ter jih pregledovati, testirati in nameščati podobno kot druge programske artefakte.

Pristop API-first poudarja, da semantika ni namenjena samo interaktivnemu poročilu, temveč je programsko dostopna tudi drugim sistemom.

Platforma Cube oz. Cube.dev ponazarja kombinacijo pristopov headless, API-first in code-first. Deluje kot samostojna semantična storitev, ki centralizira metrike, dimenzije, povezave, pravila dostopa in optimizacijo poizvedb ter jih prek standardnih vmesnikov ponudi BI-orodjem, aplikacijam in sistemom umetne inteligence [11]. Gre za sodobno nadaljevanje idej kock in analitičnih modelov, vendar v obliki ponovno uporabne izvajalne storitve zunaj predstavitvenega sloja posameznega poročilnega orodja.

dbt Semantic Layer in MetricFlow ponazarjata pristop code-first v tesni povezavi s transformacijskimi modeli. Metrike in semantični modeli so definirani v YAML-u, MetricFlow pa na podlagi semantičnega grafa konstruira SQL in izbira poti med modeli [12]. Pristop je naraven predvsem v okoljih, kjer se podatkovni modeli, testi, sledljivost izvora (lineage) in dokumentacija že upravljajo v projektu dbt.

5.4. Operativna ontologija in grafovne platforme

Palantir Ontology predstavlja širši, operativni pristop. Palantirjeva dokumentacija opisuje Ontology kot digitalni dvojček organizacije, ki podatkovne vire in modele preslika v tipe objektov, lastnosti, povezave in vrste akcij. Akcije lahko spreminjajo objekte, lastnosti in povezave ter sprožajo dodatne učinke [13]. To je že korak od analitičnega razumevanja proti operativnemu modelu, na katerem lahko delujejo aplikacije in agenti.

Grafovne platforme, kot so Neo4j, Stardog, platforme RDF ali platforme na osnovi modela property graph, lahko služijo kot tehnološka osnova grafa znanja. Vendar grafovna baza sama po sebi še ni graf znanja in tudi ne zagotavlja dogovorjene poslovne semantike. Vrednost nastane z modeliranjem identitete, odnosov, konteksta, kakovosti in upravljanja, kar poudarja tudi sodobni pregled področja grafov znanja [9].

Primeri v tabeli ne pomenijo primerjave ali priporočila posameznega izdelka, temveč ponazarjajo različne arhitekturne vzorce in njihove značilne vloge.

Tabela 4: Tehnološki vzorci semantične plasti in sorodnih komponent.

Arhitekturni vzorec	Kaj ponazarja	Primeri
V BI vgrajeni semantični model	Mere, dimenzije in varnost blizu poročanju in poslovnim uporabnikom.	SSAS, Power BI Semantic Model
Platformno vgrajena semantika	Poslovni koncepti neposredno ob DWH ter AI- in SQL-storitvah.	Snowflake Semantic Views
Headless in API-first semantična plast	Ponovna uporaba metrik in pravil prek več odjemalcev.	Cube
Semantika blizu transformacij	Metrike kot koda ob podatkovnih modelih in transformacijah.	dbt Semantic Layer, MetricFlow
Operativna ontologija	Objekti in povezave, povezani tudi z akcijami in procesi.	Palantir Ontology
Platforma za graf znanja	Hranjenje in poizvedovanje po povezanih entitetah ter kontekstu.	Neo4j, Stardog, RDF-platforme

5.5. NL2SQL je orodje, ne semantika

Pretvorba vprašanj v naravnem jeziku v SQL (NL2SQL oziroma text-to-SQL) je uporabna tehnična sposobnost: sistem poskuša vprašanje uporabnika povezati s tabelami, stolpci, filtri in agregacijami ter iz njega ustvariti poizvedbo SQL. Raziskovalno področje je bistveno starejše od razmaha velikih jezikovnih modelov; ti so izboljšali prevajanje, niso pa odpravili problema povezovanja poslovnega jezika s shemo in pomenom podatkov [14].

NL2SQL rešuje, kako vprašanje izraziti v SQL. Semantična plast rešuje, kaj je pravilno vprašati in kako je treba rezultat interpretirati.

NL2SQL sam po sebi ne ve, katera definicija marže je v organizaciji uradna. Lahko ugiba na podlagi imen stolpcev ali opisov, lahko pa uporabi upravljano metriko iz semantične plasti. Zato ga v arhitekturi ne obravnavamo kot naslednjo razvojno stopnjo semantike, temveč kot enega od porabnikov semantičnega modela.

6 Od semantične plasti do podatkovnega asistenta in agenta

6.1. Podatkovni asistent kot izvedljiv naslednji korak

Šele na podlagi predhodno opisane arhitekture lahko natančneje uvedemo podatkovnega asistenta. Izraz še nima enotne in splošno sprejete tehnične definicije. V tem prispevku z njim označujemo AI-podprto programsko rešitev, ki z uporabo semantične podatkovne plasti analizira strukturirane podatke in z njimi povezani poslovni kontekst ter uporabniku pripravi sledljiv odgovor, ustrezen analitični pogled ali predlog fokusa. Ne gre za osebo niti za sinonim semantične plasti; asistent je njen porabnik.

Izvedljiv cilj ni, da sistem takoj samostojno sprejema poslovne odločitve. Za številna okolja bi bil pomemben napredek že asistent, ki zna iz kombinacij mer in dimenzij sproti pripraviti ustrezen analitični pogled oziroma ad hoc dashboard, namesto da bi moral uporabnik pravi pogled iskati med obstoječimi poročili. Tako mu pomaga bolje razumeti informacije, ki jih nosijo te kombinacije.

6.2. Analiza na zahtevo in proaktivni pregled

Pri analizi na zahtevo uporabnik že ve, kaj ga zanima, in vpraša: »Zakaj se je marža zmanjšala?« Analitik pregleda pripravljene dashboarde, po potrebi izvede dodatno ad hoc poizvedbo, primerja relevantna obdobja in dimenzije ter pripravi odgovor. Kakovost in hitrost analize sta pri tem odvisni od njegovega poznavanja podatkovnega modela, poslovnih definicij in razpoložljivih virov.

Pri proaktivnem pregledu uporabnik še ne ve, kateri pojav zahteva njegovo pozornost. Asistent v vnaprej določenem obsegu pregleda dogovorjene metrike, zazna pomembnejša odstopanja in predlaga, kam naj se analiza

usmeri. Prav ta način izraža sporočilo »ne še en dashboard«: rešitev ne čaka samo, da uporabnik odpre pravi pogled, temveč mu pomaga ugotoviti, kaj je pomembno.

6.3. Kaj podatkovni asistent dejansko naredi

Ne glede na to, ali analizo sproži uporabnik ali vnaprej določen pregled, mora podatkovni asistent izvesti več povezanih korakov:

- določi ustrezne metrike, obdobja in primerjalni kontekst;
- iz semantične plasti pridobi uradne definicije, dovoljene dimenzije in pravila;
- pripravi ali izvede potrebne analitične poizvedbe;
- primerja rezultate po smiselnih kombinacijah mer in dimenzij;
- loči manj pomembna nihanja od odstopanj, ki zahtevajo pozornost;
- pripravi ustrezen analitični pogled oziroma ad hoc dashboard;
- rezultat pojasni skupaj z uporabljenimi viri, omejitvami in predlogom, kam usmeriti nadaljnjo analizo.

6.4. Zakaj isto osnovo potrebujejo tudi agenti

Semantična podatkovna plast ni pomembna samo za podatkovnega asistenta. Njena vloga postane še pomembnejša pri agentu, ki naj bi več analitičnih korakov načrtoval in izvedel z manj neposrednega usmerjanja človeka. Če mora sistem pri vsakem koraku uporabnika vprašati, kateri dashboard naj odpre, katera definicija metrike velja in po kateri dimenziji naj nadaljuje analizo, še ne deluje kot agent z večjo stopnjo samostojnosti. Manj ko je podrobnih človeških navodil, jasneje morajo biti opredeljeni poslovni pomen, dovoljene povezave, pravila dostopa in meje delovanja.

To ne pomeni, da so organizacije že pripravljene na avtonomno poslovno odločanje. V podatkovni stroki se dobro zavedamo, da smo od povsem samodejnega odločanja še daleč. Podatkovni asistent je zato izvedljiv naslednji korak: pripravi ustrezen pogled, opozori na signal in pomaga pri razumevanju, človek pa ohrani odločitev in odgovornost. Agent predstavlja nadaljnjo stopnjo, na kateri lahko sistem poleg analize v okviru jasnega pooblastila sproži tudi dejanje.

Prehod od branja in analize k zapisovanju v poslovne sisteme bistveno poveča zahteve glede pooblastil, nadzora, validacije in odgovornosti. Profil NIST za generativno umetno inteligenco zato poudarja upravljanje tveganj skozi celoten življenjski cikel sistema, ne le kakovost končnega uporabniškega vmesnika [15].

Tabela 5: Praktična razlika med podatkovnim asistentom in AI-agentom.

Podatkovni asistent	AI-agent
Pripravi analizo, razlago in pogled.	Načrtuje tudi nadaljnje korake.
Opozori in predlaga fokus.	Lahko sproži proces ali delegirano dejanje.
Primarno bere podatke in kontekst.	Lahko tudi zapisuje v operativne sisteme.
Človek odloči.	Del odločanja ali izvedbe je delegiran.
Potrebuje sledljivo semantiko.	Zaradi večje avtonomije jo potrebuje še bolj.
Nižje operativno tveganje.	Višje tveganje in strožje zahteve po nadzoru.

Semantična plast ni posledica agentov. Je pogoj, da lahko podatkovni asistent ali agent deluje brez stalnega človeškega prevajanja poslovnega pomena.

Arhitekturna meja je zato jasna: preverjeni podatki in semantična plast podpirajo podatkovnega asistenta; asistent pripravi signal, razlago in fokus; človek odloči; šele ob dodatnih pooblastilih in varovalih lahko agent izvede ukrep.

7 Praktični primer: od metrike marže do poslovnega fokusa

7.1. Klasična pot skozi poročila

Predpostavimo, da se je marža v enem od segmentov zmanjšala. V klasičnem pristopu uporabnik odpre dashboard prodaje, opazi spremembo, nastavi obdobje, filtrira segment, preveri izdelke in kupce, nato pa odpre še poročilo nabave ali poišče cenik. Podatki morda obstajajo, toda analitično pot še vedno sestavi človek. Rezultat je odvisen od njegovega poznavanja modelov, definicij in mest, kjer so informacije shranjene.

7.2. Kaj doda semantična plast

Semantična plast zagotovi uradno definicijo prodajne marže iz tabele 1, ustrezen časovni kontekst, dovoljene dimenzije, pravilo agregiranja in povezavo do preverjenih podatkov. Modul za analizo ali podatkovni asistent zato ne išče poljubnega stolpca z imenom »margin«, temveč uporablja upravljano metriko. S tem vzroka še ne pojasni, zagotovi pa poslovno dosledno izhodišče.

7.3. Kaj doda graf znanja

Graf znanja lahko odstopanje poveže s konkretnimi entitetami in dogodki: segmentom, skupino izdelkov, dobaviteljem, pogodbo, cenikom ali prodajno akcijo. Dokumenti niso samodejno dokaz vzroka, ponujajo pa relevanten kontekst za nadaljnjo razlago. Pri preprostejših primerih je lahko dovolj dober indeks dokumentov; graf znanja je smiseln takrat, ko so odnosi kompleksni in jih želimo ponovno uporabljati v več analitičnih ali operativnih scenarijih.

7.4. Kaj naredi podatkovni asistent

Podatkovni asistent lahko najprej zazna, da je sprememba večja od dogovorjenega praga ali pomembno odstopa od primerjalnega vzorca. Nato sistematično preveri dovoljene dimenzije ter spremembo marže razčleni na učinke prodajne cene, nabavne cene, rabatov, prodajnega miksa in valute. Izpostavi največje kvantificirane prispevke in jih poveže z relevantnimi poslovnimi dogodki.

Primer sledljivega odgovora bi lahko bil: »Prodajna marža se je najbolj zmanjšala v segmentu X. Največja kvantificirana prispevka sta bila rast nabavnih cen skupine izdelkov Y in večji delež prodaje z akcijskim rabatom. V istem obdobju je začel veljati nov cenik dobavitelja A. Predlagana področja nadaljnjega pregleda so vpliv pogodbe B, sprememba cenika C in podaljšanje akcije D.« Gre za ponazoritveni arhitekturni scenarij, ne za rezultat empirično preverjenega prototipa.

Pomembna je tudi previdnost pri jeziku odgovora. Sistem lahko pokaže kvantificirane prispevke, časovno povezanost in verjetne razlage. Brez dodatne analize ne sme trditi, da je posamezen poslovni dogodek dokazano povzročil spremembo. Sledljiv in razločljiv odgovor mora jasno ločiti meritev, povezavo in interpretacijo.

7.5. Kje bi se začela vloga agenta

Podatkovni asistent se po pripravi fokusa ustavi. Ne spremeni cenika, ne prekine prodajne akcije in ne odpre pogajanj z dobaviteljem. Agent bi lahko v naslednjem koraku, ob ustreznih pooblastilih, odprl nalogo odgovorni osebi, pripravil predlog ukrepa ali sprožil potrditveni proces. Prav zato mora biti razlika med analitično pomočjo in operativnim ukrepanjem jasno določena že v arhitekturi.

8 Omejitve in pogoji za dejansko uporabo

8.1. Semantična plast ne odloči, katera definicija je pravilna

Semantična plast lahko definicijo formalizira, upravlja njene različice in jo dosledno uporablja, ne more pa sama razrešiti spora med prodajo, financami in kontrolingom. Metrike potrebujejo poslovnega lastnika, postopek potrjevanja in jasno veljavnost. Napačno opredeljena metrika ostane napačna, le da je napaka zdaj bolj vidna in ponovljiva.

8.2. Slabi podatki ostanejo slabi podatki

Manjkajoči, prepozni ali napačni podatki se z dodajanjem semantike ne popravijo. Še vedno potrebujemo zagotavljanje kakovosti podatkov, testiranje transformacij, nadzor izvorov in poznavanje granulacije. Semantična plast lahko zmanjša nevarnost uporabe napačne tabele ali napačnega pravila, ne more pa nadomestiti podatkovnega inženirstva.

8.3. Spremembe, dostop in sledljivost

Poslovne definicije se spreminjajo, zato morajo biti upravljane po različicah in časovno sledljive. Enako velja za pravice dostopa: uporabnik omejitev iz BI-orodja ne sme obiti prek asistenta ali agenta. Odgovor mora pokazati uporabljeno metriko, filtre, vire in dokumente; sicer je lahko prepričljiv, vendar ga ni mogoče preveriti.

8.4. Dodatna kompleksnost in meja razlage

Ontologija in graf znanja zahtevata modeliranje identitete, odnosov in časovne veljavnosti, zato nista obvezen del vsake semantične plasti. Smiselna sta tam, kjer širši kontekst prinaša dovolj vrednosti. Hkrati povezava med dogodkom in spremembo metrike še ni dokaz vzročnosti. Sistem mora ločiti izračunane prispevke, korelacije, dokumentirane dogodke in interpretacijo.

8.5. Podatkovni asistent je pogosto primernejši prvi korak kot agent

Organizacije pogosto govorijo o avtonomnih agentih, še preden imajo enotno definicijo osnovnih metrik ali urejeno lastništvo podatkov. V takem okolju avtonomija ne odpravi negotovosti, temveč jo avtomatizira. Podatkovni asistent lahko prinese pomembno vrednost z manjšim operativnim tveganjem: sistem pregleda, opozori in pojasni, človek pa ohrani odločitev in odgovornost. Prehod k agentskemu delovanju naj sledi šele takrat, ko so semantika, pravice dostopa, nadzorne točke in odgovornost ustrezno urejeni in dovolj zreli [15].

9 Razprava in zaključek

Prispevek semantične podatkovne plasti ne postavlja kot zamenjavo za DWH, dimenzijsko modeliranje ali semantične modele BI, temveč kot njihovo nadaljevanje. DWH omogoča ponovno uporabo preverjenih podatkov, semantična plast pa ponovno uporabo njihovega dogovorjenega pomena. Zvezdna shema, UML, ontologija in graf znanja pri tem rešujejo različna vprašanja, zato jih je treba povezovati glede na dejanski poslovni problem.

Praktični rezultat skupne semantike je lahko podatkovni asistent, ki ne izdela samo še enega pogleda, temveč pregleda več pogledov, izlušči signal in človeku predlaga fokus. NL2SQL je pri tem orodje za pretvorbo vprašanja v SQL, graf znanja je vir širšega konteksta, semantična plast pa zagotavlja iste potrjene definicije vsem porabnikom.

Naslednji korak analitike ni nujno še en dashboard – niti takoj avtonomni agent. Naslednji korak je zaupanja vredna semantična osnova.

Na tej osnovi lahko asistent pokaže, katere metrike in vire uporablja ter kje se mora ustaviti. Semantična podatkovna plast je zato manjkajoči temelj analitike nove generacije: ne zato, ker bi bila povsem nova, temveč zato, ker mora pomen podatkov postati skupen ljudem, analitičnim orodjem, aplikacijam in sistemom umetne inteligence.

Literatura

- [1] MACARI Edgar, »Improve Business Outcomes by Turning Dashboards Into Analytics Products«, Gartner, 11. september 2025, <https://www.gartner.com/en/documents/6943066>, obiskano 27. 7. 2026.
- [2] GANESHAN Anirudh, »3 Actions D&A Leaders Should Take to Reinvigorate Self-Service Analytics«, Gartner, 21. maj 2024, <https://www.gartner.com/en/documents/5448363>, obiskano 27. 7. 2026.
- [3] KIMBALL Ralph, ROSS Margy, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3. izdaja, Wiley, 2013.
- [4] MICROSOFT, Tabular modeling overview - Analysis Services, Microsoft Learn, 2024, <https://learn.microsoft.com/en-us/analysis-services/tabular-models/tabular-models-ssas>.
- [5] MICROSOFT, Power BI semantic models in Microsoft Fabric, Microsoft Learn, 2025, <https://learn.microsoft.com/en-us/fabric/data-warehouse/semantic-models>.
- [6] STIREWALT R. E. K., BÚR Márton, »The RAI Way: A Technical Analysis and Design Method for Building Enterprise Semantic Layers«, Advanced Information Systems Engineering Workshops, LNBIP 482, Springer, 2023, str. 74-79, https://doi.org/10.1007/978-3-031-34985-0_8.
- [7] OBJECT MANAGEMENT GROUP, Unified Modeling Language (UML), različica 2.5.1, <https://www.omg.org/spec/UML/2.5.1/>.
- [8] W3C OWL WORKING GROUP, OWL 2 Web Ontology Language Document Overview, Second Edition, W3C Recommendation, 2012, <https://www.w3.org/TR/owl2-overview/>.
- [9] HOGAN Aidan et al., »Knowledge Graphs«, ACM Computing Surveys, letnik 54, št. 4, članek 71, 2021, <https://doi.org/10.1145/3447772>.
- [10] SNOWFLAKE, Overview of semantic views, Snowflake Documentation, <https://docs.snowflake.com/en/user-guide/views-semantic/overview>.
- [11] CUBE DEV, Introduction and Data Modeling, Cube Documentation, <https://docs.cube.dev/docs/introduction>.
- [12] DBT LABS, About MetricFlow, dbt Developer Hub, <https://docs.getdbt.com/docs/build/about-metricflow>.
- [13] PALANTIR TECHNOLOGIES, Ontology core concepts, Palantir Foundry Documentation, <https://www.palantir.com/docs/foundry/ontology/core-concepts>.
- [14] KATSOGIANNIS-MEIMARAKIS George, KOUTRIKA Georgia, »A survey on deep learning approaches for text-to-SQL«, The VLDB Journal, letnik 32, 2023, str. 905-936, <https://doi.org/10.1007/s00778-022-00776-8>.
- [15] AUTIO Chloe et al., Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1, National Institute of Standards and Technology, 2024, <https://doi.org/10.6028/NIST.AI.600-1>.