

# **TDK-dolgozat**

**2025.**

Csatári Levente  
Budapesti Gazdasági Egyetem  
Pénzügyi és Számviteli Kar

**A DATA MESH ÉS DATA GOVERNANCE BEVEZETÉSE EGY  
KÖZEPES MÉRETŰ VÁLLALKOZÁS ÉLETÉBE**

**THE IMPLEMENTATION OF DATA MESH AND DATA  
GOVERNANCE FOR A MIDDLE-SIZED COMPANY**

Dr. Kovács Endre  
Lénárt Marcell

Kézirat lezárásának dátuma: 2025.11.05

## Tartalomjegyzék

<b>1. BEVEZETÉS</b> .....	<b>1</b>
1.1. A kutatás célja .....	1
1.2. Megválaszolendő kérdések.....	1
1.3. Hipotézisek.....	2
<b>2. SZAKIRODALOM</b> .....	<b>3</b>
2.1 A Data Mesh alapjai .....	3
2.1.1. Data Mesh és a Data Mesh Architektúra.....	3
2.1.2. Domain-orientált adatfelelőség .....	3
2.1.3. Adat, mint termék.....	4
2.1.4. Önkiszolgáló adatplatform .....	4
2.1.5. Federált Data Governance .....	5
2.2. A Data Mesh gyakorlati megvalósítása.....	5
2.2.1. Data Platform csapat.....	6
2.2.1.1. Tárolás és feldolgozás .....	7
2.2.1.2. Ingestion és pipeline menedzsment.....	8
2.2.1.3. Önkiszolgáló hozzáférés biztosítása.....	9
2.2.2. Governance csapat.....	10
2.2.2.1. Policy-k és szabványok kialakítása .....	10
2.2.2.2. Metaadat- és Adaszerződés menedzsment .....	11
2.2.2.3. Minőségbiztosítás és monitorozás.....	11
2.2.2.4. Hozzáférés-kezelés.....	12
2.2.2.5. Governance szerepkörök .....	13
2.2.2.6. Adatbiztonság.....	14
2.2.3. Enabling csapat.....	15
<b>3. A KUTATÁS MÓDSZERTANA ÉS MÓDSZEREI</b> .....	<b>17</b>
3.1. A kutatás típusa .....	17
3.2. Módszertani lépések .....	17
3.3. Esettanulmány módszerei.....	18

3.3.1. Adatprofilozás .....	18
3.3.2. Adatminőségi tesztesetek és incidenskezelés .....	19
3.3.3. Klasszifikáció .....	19
3.3.4. Üzleti szótárak és címkézés .....	20
3.3.5. Domain kategóriák behatárolása .....	21
3.3.6. Adatöröklés .....	21
3.3.7. Adatszerződések .....	21
3.4. A kutatás értékelése .....	22
<b>4. ESETTANULMÁNY .....</b>	<b>23</b>
4.1. A célcég jellemzői .....	23
4.2. Adatforrások .....	23
4.3. Governance eszköz .....	24
4.4. Egyéb felhasznált eszközök .....	24
4.5. Szervezet .....	25
4.5.1. IT .....	25
4.5.2. Értékesítés .....	26
4.6. Data Mesh bevezetés .....	27
4.6.1. Szervezeti átalakítások .....	27
4.6.2. Adatplatform megtervezése .....	28
4.6.3. Céges policy és hozzáférés-kezelés .....	30
4.8. Használati esetek .....	30
4.8.1. Használati eset 1 .....	30
4.8.1.1. Felvetett probléma üzleti áttekintése .....	31
4.8.1.2. Adatbetöltés és profilozás .....	31
4.8.1.3. Adatminőségi tesztesetek .....	32
4.8.1.4. Használati Eset 1 konklúzió .....	33
4.8.2. Használati eset 2 .....	33
4.8.2.1. Szervezeti átalakítás .....	34
4.8.2.2. Adatplatform integráció és Domain modellezés .....	35

4.8.2.3. Adatvagyon metaadat-alapú kontextualizálása .....	36
4.8.2.4. Üzleti szótárak és címkék létrehozása .....	37
4.8.2.5. Autoklasszifikáció .....	38
4.8.2.6. Használati Eset 2 konklúzió .....	40
4.8.3. Használati eset 3 .....	41
4.8.3.1. Adatöröklődés és Adatszerződés létrehozása .....	42
4.8.3.2. Használati Eset 3 konklúzió .....	43
<b>5. EREDMÉNYEK BEMUTATÁSA.....</b>	<b>45</b>
5.1. A modell bevezetése.....	47
5.2. A Data Mesh megtérülése .....	47
5.3. Hipotézisvizsgálat .....	48
5.3.1. H1 Hipotézis.....	48
5.3.2. H1 Konklúzió .....	50
5.3.3. H2 Hipotézis.....	50
5.3.4. H2 Konklúzió .....	51
<b>6. TOVÁBBI LEHETŐSÉGEK.....</b>	<b>53</b>
<b>7. ÖSSZEFOGLALÁS .....</b>	<b>54</b>

## Táblázatjegyzék

<b>1. táblázat:</b> Az OpenMetadata autoklasszifikációs algoritmusának eredménye különböző konfidenciaszinteken.....	39
--	----

## Ábrajegyzék

<b>1. ábra:</b> A Data Mesh Architektúra. Forrás: Christ, Visengeriyeva és Harrier.....	6
<b>2. ábra:</b> Az esettanulmányi módszerek közötti kapcsolatrendszer. Forrás: Saját Szerkesztés.....	18
<b>3. ábra:</b> Ügyfélszolgálati csapat szervezeti ábrája. Forrás: Saját szerkesztés.....	25
<b>4. ábra:</b> Ügyfélértékesítés csapat szervezeti ábrája. Forrás: Saját szerkesztés.....	26
<b>5. ábra:</b> Ticketing táblán lefuttatott tesztesetek eredménye. Forrás: Saját készítés az OpenMetadata adatai alapján.....	33
<b>6. ábra:</b> Átalakított ügyfélértékesítés domain. Forrás: Saját szerkesztés.....	35
<b>7. ábra:</b> OpenMetadata-ba betöltött tábla klasszifikálás előtt. Forrás: Saját készítés az OpenMetadata adatai alapján.....	36
<b>8. ábra:</b> Klasszifikált üzleti tábla. Forrás: Saját készítés az OpenMetadata adatai alapján.....	41
<b>9. ábra:</b> Ügyfélértékesítési adatvagyon adatainak folyama. Forrás: Saját készítés az OpenMetadata adatai alapján.....	42
<b>10. ábra:</b> Átalakított architektúra. Forrás: Saját szerkesztés.....	45
<b>11. ábra:</b> Lehetséges Domain stratégiák. Forrás: Saját szerkesztés.....	49

## 1. BEVEZETÉS

Jelenleg negyedéves Gazdaságinformatikus hallgatóként tanulok Üzleti Adatelemző Informatikus specializáción, emellett egy informatikai tanácsadó cégnél dolgozom adatelemzőként. Munkám során különböző adatkezelési megoldásokkal foglalkozom, ahol egyre nagyobb hangsúlyt kap a Data Governance, azaz az adatok strukturált és szabályozott kezelése. A kutatásom megírásának nagy inspirációja volt a vállalati életben is mélyebb kutatást végezni a témában és elősegíteni cégem belépését az adat világába. A Data Governance és az ehhez kapcsolódó Data Mesh és annak architektúrája egy kiváló ötvözetét képezi tanulmányaim gazdasági vonatkoztatásának és az informatikának is. Bár ezek a fogalmak elsőre komplexnek tűnhetnek, célom, hogy a dolgozat során közérthető módon mutassam be őket, akár olyan olvasók számára is, akik még nem ismerik az adatmenedzsment alapjait.

### 1.1. A kutatás célja

A kutatás célja, hogy feltárja a Data Governance szerepét a Data Mesh szervezeti modellen keresztül, különös tekintettel annak vállalati gyakorlatban történő alkalmazására közepes méretű vállalkozások esetében. A dolgozat bemutatja a vállalaton belüli kollaboratív és decentralizált adatkezelés előnyeit, amelyet a Data Mesh modell képvisel, ahol minden szervezeti egység aktív szerepet vállal az adatkezelésben. Az elméleti alapok ismertetését követően a gyakorlatban is szemléltetem az egyes komponensek működését egy Governance eszköz segítségével, különös tekintettel arra, hogy ez miként járul hozzá a leggyakoribb adatkezelési problémák kezeléséhez.

### 1.2. Megválaszolandó kérdések

A kutatásom fő kérdése: Milyen szervezeti és technológiai előnyöket kínál a Data Mesh és a Data Governance középvállalatok számára, és miként járulhat hozzá ezek hatékony bevezetése az üzleti működéshez?

Ezek mellett pedig három használati esetre keresem a megoldást egy fiktív cégen keresztül:

1. Az IT ügyfélszolgálati divízióján belül a ticketing rendszerből származó adatok hiányosan és hibásan kerülnek átadásra, aminek következtében számos rekord felhasználhatatlanná válik. Ez súlyos fennakadásokat okoz más analitikai rendszerekben és üzleti folyamatokban, valamint az ügyfélélményt is rontja, akár ügyfélvesztést vagy a vállalati hírnév sérülését eredményezve, mivel kritikus bejelentések elveszhetnek. A vállalati vizsgálatok kizárták az emberi mulasztás lehetőségét: a probléma gyökere az architektúrában keresendő. Hogyan lehet olyan megoldást kialakítani, amely biztosítja

az adatok minőségének folyamatos monitorozását, és időben figyelmeztet a hibás adatok átadására?

2. A vállalat értékesítési szervezetében, az ügyfélértékesítés részlegén az adatkezelés nem integrált a központi rendszerekkel: a dolgozói információk különálló fájlokban “szétszórva” keletkeznek, amelyek gyakran elvesznek vagy elavulnak, és senki sem vállal felelősséget értük. Ennek eredményeként értékes elemzési lehetőségek maradnak kihasználatlanul, az értékesítés csapat pedig munkaideje nagy részét az adatok utánkeresésére és kézi frissítésére fordítja. Ezt a problémát silósodásnak nevezzük. Hogyan vonhatjuk be hatékonyan az ügyfélértékesítés csapatot a vállalati adatkezelésbe úgy, hogy közben ne terheljük őket bonyolult informatikai feladatokkal, és milyen megoldásokkal tehetjük mindezt egyszerűbbé és átláthatóbbá?
3. Egy korábbi ügyfél kéri adatainak teljes törlését. Az érintett adatok több, egymástól független rendszerben és eltérő azonosítók alatt szerepelnek. A jelenlegi folyamat manuális, időigényes, és különböző divíziók együttműködését igényli, gyakran olyan egységeket is bevonva, amelyek normál esetben nem dolgoznak szorosan együtt. Ez felvet olyan problémákat, mint a gyors szervezeti egységek közti kommunikáció hiánya. Hogyan tudnám ezt úgy megoldani, hogy az ilyen speciális esetekben is gyorsan megtaláljam az összes ilyen személyes adatot és bármely másik részleggel is, aktív együttműködéstől függetlenül, tudjak hibamentesen együtt dolgozni?

### 1.3. Hipotézisek

A kutatásom fő kérdésének megválaszolásához néhány hipotézist is felállítottam. A használati esetek és hipotézisek eredményének ötvözete segít nekem egy objektív képet generálni a fő kérdésem megválaszolásához.

- H1: Mélyebb al-domain struktúrák kialakítása elősegíti a szervezeti felelősségvállalás és elszámoltathatóság erősödését.
- H2: Egy megfelelően bevezetett Data Governance eszköz kulcsszerepet játszik a Data Mesh sikeres alkalmazásában.

## 2. SZAKIRODALOM

### 2.1 A Data Mesh alapjai

Ebben a fejezetben bemutatom a Data Mesh elméleti alapjait és annak fő pilléreit, emellett részletesen kitérek arra is, hogyan különbözik maga a Data Mesh mint működési modell a gyakran félreértelmezett Data Mesh architektúrától. Ezek a fogalmak alapvető jelentőséggel bírnak a Data Mesh és architektúrája megértésében, és nélkülözhetetlenek a későbbi, gyakorlati példák és technológiai megvalósítások, valamint az esettanulmányom értelmezéséhez.

#### 2.1.1. Data Mesh és a Data Mesh Architektúra

A Data Mesh és a Data Mesh architektúra kifejezéseket gyakran, még professzionális környezetben is, szinonimaként használják, azonban a két fogalom lényegében különbözik egymástól. A Data Mesh önmagában egy elméleti működési modell az adatszervezésre. Ennek részei az olyan alapelvek, mint az „adat,mint termék”, a decentralizált adatfelelősség és a federált Data Governance (Kinza, 2024). Ezzel szemben a Data Mesh Architektúra ezen elvek gyakorlati megvalósítását jelenti. Ide tartozik egy Data Governance eszköz alkalmazása, a közös adatplatform kiépítése, valamint az adatfolyamok megfelelő kialakítása például pipeline rendszerek segítségével.

A „Data Mesh architektúra” elnevezés sok esetben megtévesztő lehet, mivel technikai elvként pozicionálja magát, holott a valóságban a modell sokkal inkább szervezeti és működésbeli alapokon nyugszik. A Hyperight egyik cikke kifejezetten óva inti a vállalatokat attól, hogy objektív, technikai megoldást keressenek ott, ahol valójában szemléletváltásra van szükség. Bár a fogalomhasználat körüli megosztottság érthető, úgy vélem, hogy a kutatások sokszor nem fordítanak elég figyelmet a Data Mesh és annak technikai implementációja közötti különbségtételre. (Hyperight, 2022)

Véleményem szerint a Data Mesh gyakorlati alkalmazása valóban tekinthető architektúrának, azonban fontos megérteni, hogy ez nem önálló technológia, hanem elméleti elvek és a cég technikai-infrastrukturális környezetének ötvözete. Ezért a gyakorlatban szinte biztosan kijelenthető, hogy nem létezik két teljesen azonos Data Mesh architektúra. Minden szervezetben a környezet, az érettség és az üzleti igények formálják a végső megvalósítást.

#### 2.1.2. Domain-orientált adatfelelőség

A Data Mesh egyik alappillére a Domain-orientált adatfelelőség. Célja, hogy megszüntesse az adatkezelési függőségeket a központi IT vagy adatsapatoktól. Egy jó Data Mesh kialakításban ez az első felmérési alap, miszerint melyik domain, azaz terület adatmozgása és kezelése

mennyire függ egy másik területtől. Sok esetben az ilyen esetek adatkésleltetéseket okoznak, nehezen készülnek el riportok időre, az alkalmazottak túlterheltek lesznek. A Data Mesh lehetőséget biztosít arra, hogy minden terület saját maga feleljen az általa előállított adatokért, azok minőségéért és karbantartásáért. Ehhez elengedhetetlen, hogy meglegyenek az adattulajdonosok (*Data Owner*), akik felelősek az adatokért, és hogy az egyes domének képesek legyenek a saját szempontjaik szerint, átlátható módon kezelni és rendszerezni információikat. Ugyanakkor az adatok megoszthatóságát biztosító infrastruktúra nem az üzleti területek feladata, hanem az erre dedikált mérnöki vagy platformcsapaté, akik a technikai hátteret biztosítják. Ez a megosztott felelősség teremti meg a hatékony együttműködés alapját. (Christ, Visengeriyeva és Harrier, 2022)

### **2.1.3. Adat, mint termék**

Egyes ágazatokban gyakori az „eldobható” adatok használata, amelyek dokumentálása és központi elérhetősége hiányos. Ez akadályozza a domének közötti adatmegosztást, mivel az adatok értelmezése sokszor csak emberi közvetítéssel lehetséges. A Data Mesh „adat, mint termék” szemlélete erre ad megoldást: a domain csapatok felelősek saját adataik előállításáért, dokumentálásáért, karbantartásáért és publikálásáért, ami átláthatóságot, gyorsabb adatáramlást és jobb analitikai felhasználhatóságot eredményez. (Christ, Visengeriyeva és Harrier, 2022).

Az adattermékek architektúrális szerepe kiemelkedő. Egy jól átgondolt domainstruktúra megelőzi a silósodást, és biztosítja, hogy az adatok könnyen hozzáférhetők és újrafelhasználhatók legyenek. Ehhez alapvető a metaadatok kezelése, vagyis az adatokhoz kapcsolódó információk (pl. létrehozási dátum, módosító neve, kapcsolódó táblák, öröklési gráf). Bár ezek sokszor háttéradatként jelennek meg, szisztematikus feltérképezésük jelentős értéket teremt az átláthatóság, a megfelelés és a menedzsment támogatásában. (Saux, 2024)

Az adattermékeket célszerű adatszerezésekbe (*Data Contract*) rendezni, amelyek részletesen dokumentálják a sémát, a frissítési gyakoriságot, a minőségi elvárásokat és az adatfelhasználási szempontokat. Ezek biztosítják, hogy az adatok termékként való kezelése fenntartható és skálázható maradjon, megfelelően az analitikai és üzleti igényeknek. (Christ és Harrier, 2024)

### **2.1.4. Önkiszolgáló adatplatform**

Az emberi hiba csökkentése, a munkavégzés megkönnyítése és az adatrend kialakítása érdekében a modern architektúrák egyik kulcseleme az automatizáció. Az önkiszolgáló adatplatform ebben jelentős szerepet játszik, támogatja a munkatársakat és a különböző domain

csapatokat azzal, hogy az adatokat központilag elérhetővé és újrahasznosíthatóvá teszi. Így a publikált adatokkal bárki könnyedén dolgozhat, miközben megszűnik az adott domain csapatok függése a központi IT vagy adats csapatoktól.

Az automatizált folyamatok bevezetése segít kiküszöbölni az emberi hibákat, elkerülhetővé teszi a helytelen adatbevitelt, valamint biztosítja, hogy az adatok strukturáltan, szisztematikusan és egységesen kerüljenek kezelésre. Például egy HR csapat képes lehet önállóan feltölteni egy új alkalmazotti adatfájlt, amelyet az önkiszolgáló platform automatikusan validál, katalogizál, és összekapcsol a meglévő adatmodellbe, így az adat szinte azonnal felhasználhatóvá válik más területek vagy riportáló eszközök számára. Ez a fajta technikai önállóság nem csak gyorsítja a folyamatokat, de csökkenti a hibalehetőségeket is, hiszen a rendszeres és automatikus ellenőrzések beépítettek a platformba. (Sala, 2024)

### **2.1.5. Federált Data Governance**

Egy vállalat adatkezelése lehet akár fejlett és jól átgondolt is, azonban nem célszerű teljesen szabályozatlanul hagyni minden divíziót saját adattermékek kialakításában. A federált Data Governance, a Data Mesh következő kulcspillére, pontosan ezt a problémát kezeli. Megteremti az egységes és mégis rugalmas adatkezelési kultúrát. Lényege, hogy miközben lehetőséget ad az egyes domain csapatoknak arra, hogy saját módszereik szerint gondoskodjanak az adataikról, meghatároz bizonyos cégen átívelő szabályozásokat, amelyeket mindenkinek be kell tartania. (Meyer, 2025)

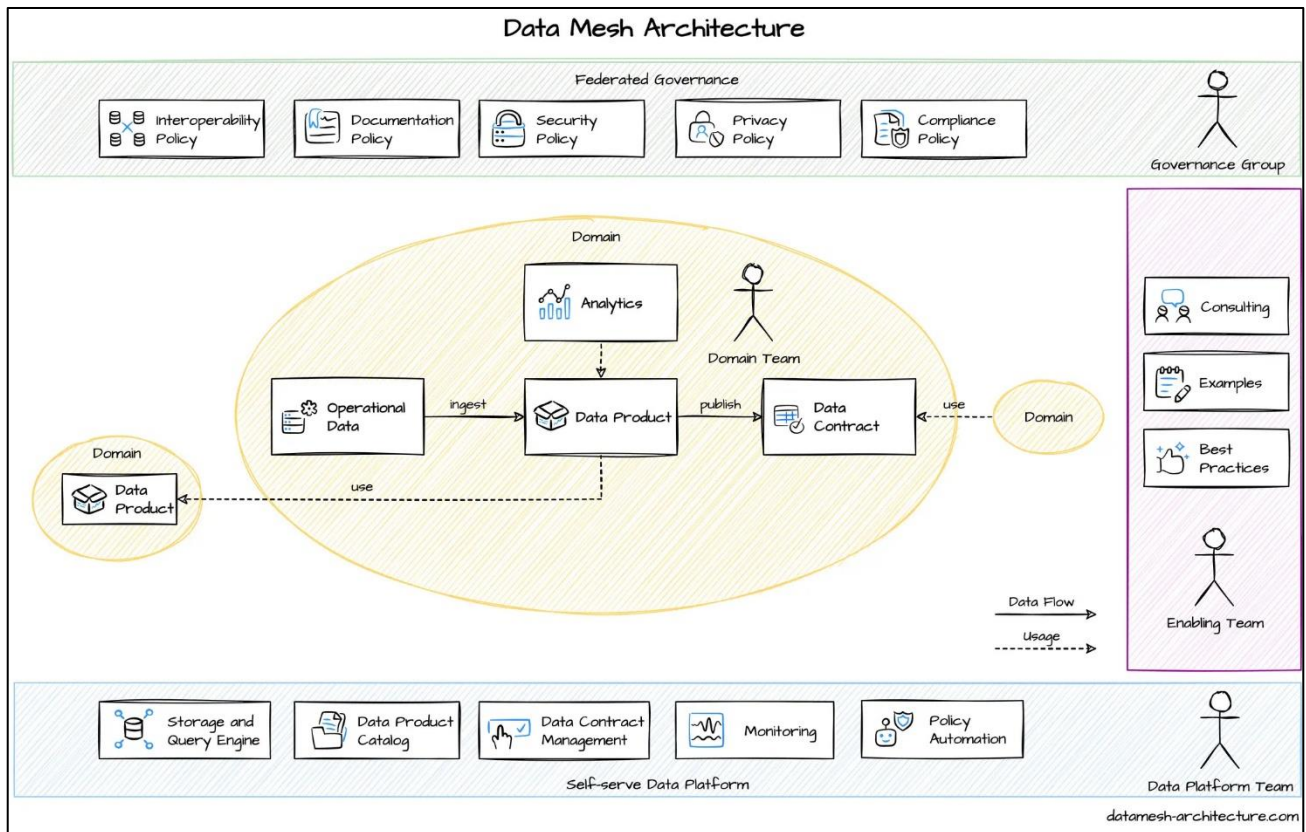
Különösen fontos ez érzékeny adatkezelés esetén, mint például a személyes adatok védelme. Egy energetikai cégnél nem döntheti el minden domain önállóan, hogy miként kezeli az ügyfélneveket, e-mail címeket vagy fogyasztási adatokat. Ezekre vonatkozóan központi elvárásokat kell lefektetni, például az adatmaszkolás vagy anonimizálás szabályait. Ezeket a szabályokat a Data Governance csapat dolgozza ki, a domain csapatok pedig a saját környezetükhöz igazítva, de kötelezően alkalmazzák. A domain adattulajdonosok és adatgazdák kulcsszerepet töltenek be ebben, hiszen ők felelnek az együttműködésért, az egységes értelmezésért és a megfelelés biztosításáért.

## **2.2. A Data Mesh gyakorlati megvalósítása**

A Data Mesh alapelveinek megértése után a következő lépés a gyakorlati megvalósítás kérdése. Ebben a fejezetben a Data Mesh architektúra felépítését mutatom be, középpontba helyezve Christ, Harrier és Visengeriyeva jól ismert referenciamodelljét. Az 1. ábrán bemutatott architektúra kiemelkedően jól szemlélteti, milyen főbb technikai és működési komponensekből

állhat egy Data Mesh rendszer, és ideális kiindulási pontot biztosít bármely szervezet számára, amely ilyen irányba kíván elmozdulni.

A bemutató során kitérek a különböző architektúrális rétegekre, azok tipikus szereplőire, valamint azokra az eszközökre, amelyek egy közepes méretű vállalat számára hatékonyan és költségtudatosan támogathatják a kialakítást.



1. ábra: A Data Mesh Architektúra. Forrás: Christ, Visengeriyeva és Harrier

### 2.2.1. Data Platform csapat

A Data Mesh architektúra sikeres működésének egyik kulcsfontosságú eleme a Data Platform egység, amely belső technológiai szervezatként biztosítja az alapvető infrastruktúrát és adatkezelési szolgáltatásokat a vállalat domain területei számára. A csapat elsődleges célja, hogy technikai komplexitástól mentesítse a domain csapatokat, így azok az üzleti problémák megoldására és adattermékek karbantartására koncentrálhatnak. A Data Platform csapat tehát egy olyan önkiszolgáló platform kialakításáért felelős, amely magában foglalja az adattárolás és feldolgozás infrastruktúráját, az adatok begyűjtését (*ingestion*), a pipeline-menedzsmentet, valamint a felhasználóbarát hozzáférés biztosítását. (Dehghani, 2022)

### 2.2.1.1. Tárolás és feldolgozás

Egy jól működő Data Mesh architektúrában elengedhetetlen az adatok megfelelő tárolási és feldolgozási rétegének kialakítása. A strukturálatlan, átdolgozandó vagy valós idejű adatok például egy adattóval (*Data Lake*); a kész, rendszerezett és tiszta adatok pedig egy adattárházzal (*Data Warehouse*) tárolva lehetnek. Ezek a platformok lehetővé teszik, hogy a különböző szervezeti egységek adatai ugyanazon keretrendszerben, de domain-specifikus logikai sémák szerint elkülönítve legyenek tárolva. Egy ilyen vállalat adattárolását jellemzően felhőalapú platform fogja össze, amely lehetővé teszi az üzleti egységek közös keretrendszerben való együttműködését bonyolult integrációk nélkül. A felhő rugalmas kapacitás- és erőforrás-kezelést biztosít, így nincs szükség saját fizikai infrastruktúra fenntartására, miközben magas szintű biztonságot nyújt. (Susnjara és Smalley, 2025)

A világ három vezető felhőalapú adatszolgáltató platformja (*Microsoft Azure*, *Amazon Web Services (AWS)* és a *Google Cloud*). Ezek a szolgáltatások kisvállalkozásoktól a nagyvállalatokig testre szabhatók, így egy közepes cég igényeihez is rugalmasan illeszthetők. A platformok nem fix díjas konstrukciót kínálnak, hanem a „pay-as-you-go” modellt alkalmazzák, a vállalatok az általuk ténylegesen felhasznált erőforrások (például adattárolás, számítási kapacitás, hálózati forgalom) után fizetnek. Ennek eredményeként a havi számla dinamikusan változik az egyes divíziók erőforrás-igényétől függően, ami egyszerre nyújt költséghatékonyságot és rugalmasságot. A vállalatoknak viszont cserébe gondosan kell megtervezniük és nyomon követniük erőforrás-beállításaikat, különben az exponenciálisan megnövekedő költségek a megtakarítás helyett többletkiadást eredményezhetnek. Megfelelő menedzsment mellett azonban a felhőalapú architektúra kényelmet, magas szintű biztonságot és költséghatékonyabb működést biztosít a hagyományos fizikai tárolási megoldásokhoz képest. (Coursera, 2025) (Yasar, Bigelow és Sullivan, 2024)

Úgy vélem, ma már egy modern vállalat számára elengedhetetlen az ilyen felhőalapú szolgáltatások igénybevétele. A saját fizikai infrastruktúra kiépítése ugyanis számos kihívást rejt magában: a komponensek skálázhatóságának korlátai, a megnövekedett fizikai és rendszervédelmi igények, valamint a folyamatos karbantartás bonyolultsága mind nehezítik a működést. Ha egy vállalat ragaszkodik a helyszíni infrastruktúrához, de emellett szeretné érvényesíteni a Data Mesh elveit, nélkülözhetetlen az infrastruktúra modernizálása a felhőszolgáltatások irányába. Ellenkező esetben a technológiai elavultság gátat szab a kritikus komponensek integrációjának, kiépítésének.

A vállalatirányítási rendszerek (*ERP*) főként nagyvállalatoknál elterjedtek, mivel több egység adatait képesek egységesen kezelni. KKV-k esetében azonban a nagyszabású megoldások (pl. *SAP*, *Microsoft Dynamics 365*) gyakran túl költségesek és összetettek, ezért inkább kisebb, specializált rendszerek alkalmazása célszerű (Kimberling, 2023). Ha egy szervezet már rendelkezik ERP-vel, annak integrálása erős tranzakciós alapot és központi adatforrást nyújt a Data Mesh számára, ám újonnan bevezetve túlzott komplexitást és költséget jelenthet.

Az adatok feldolgozása az adate mérnökök feladata. A nyers forrásokból üzletileg értékes információt kell előállítani, legyen szó strukturált vagy strukturálatlan adatokról. Közepes cégeknél egyszerre merül fel időszakos (*batch*) és valós idejű (*streaming*) feldolgozási igény (Redpanda, 2024). Az adatszerződések pontosan meghatározzák, milyen frissességgel kell a domének adatait publikálni, az infrastruktúrának pedig biztosítania kell ezek betartását. A batch feldolgozást általában *Apache Spark* vagy *Databricks*, a streaminget pedig *Apache Kafka* biztosítja, amelyek mindhárom nagy felhőszolgáltatóval integrálhatók (Microsoft, 2025)

#### **2.2.1.2. Ingestion és pipeline menedzsment**

Egy közepes méretű cég jellemzően már rendelkezik központi adatbázissal, ahol a részlegek sémákra osztva, egységesen, mégis elkülönítve tárolhatják adataikat. A Data Mesh lehetővé teszi, hogy a domain-specifikus források (pl. *Salesforce*, *Databricks*, *BambooHR*) megőrizzék saját környezetüket, miközben az egységes ingest- és pipeline-réteg biztosítja az adatok konzisztens előfeldolgozását.

A felhőalapú szolgáltatások integrációja egységes keretrendszer ad, de ha a vállalat egyetlen ökoszisztémára (pl. *SAP* vagy *Microsoft Cloud*) épít, a „vendor lock-in” kockázata megnő, a későbbi platformváltás nehéz és költséges lehet. Ezért célszerű már a tervezésnél mérlegelni a meglévő alkalmazások illesztését és a váltás lehetőségét. Megoldást jelenthet a multi-cloud stratégia, ahol több szolgáltató erőforrásai párhuzamosan használhatók, így rugalmas környezet jön létre, miközben az ingest- és pipeline-mechanismusok konzisztensen működtethetők (Cloudfare, 2020). Egy platformváltás (pl. ERP-ről felhőszolgáltatásra) jelentős adatmigrációt és a domain-csapatok technikai felkészítését igényli. Az átállás nehézségei csökkenthetők előre kidolgozott migrációs stratégiával és ütemtervvel.

Az alkalmazások közti kommunikáció és adatcsere elengedhetetlen eleme az API-k használata. Az API (*Application Programming Interface*) olyan felület, amelyen keresztül két különálló alkalmazás közvetlenül kommunikálhat, adatokat cserélhet és szolgáltatásokat hívhat

meg. Ez különösen akkor előnyös, ha a szükséges adatok előfeldolgozás nélkül, az eredeti formátumban használандók fel. (Dávid, 2023)

Data Mesh architektúra esetén az API-k kulcsszerepet játszanak, például a metaadatok lekérdezésében és továbbításában. A webes felületeken keresztüli API-hívások lehetővé teszik, hogy a felhasználók, eszközök és automatizált folyamatok egységes módon férjenek hozzá a metaadat-katalógushoz és egyéb szolgáltatásokhoz.

A REST (*Representational State Transfer*) a legelterjedtebb API-architekturális elv, amely szabványos, kliens-szerver szeparációra és stateless működésre épül. Ennek köszönhetően a REST API-k gyors, skálázható és nyelvfüggetlen kommunikációt biztosítanak. (Gupta, 2025). Ez különösen fontos a Data Mesh keretrendszerében, például egy web-alapú Governance eszköz esetében, ahol a nyelvfüggetlen, jól dokumentált és skálázható metaadat-hozzáférés alappillére az architektúrának.

### **2.2.1.3. Önkiszolgáló hozzáférés biztosítása**

A tárolási réteg és az API-k révén a vállalat önkiszolgáló felületeket alakíthat ki, ahol a munkatársak IT- vagy adatcsapat támogatása nélkül férhetnek hozzá az adatokhoz. Ezek jellemzően webes portálokon keresztül érhetők el, doménenként konszolidált nézettel és megfelelő jogosultságokkal.

Egy központi felület egyrészt egységesíti a felhasználói élményt és szabványosítja az adatlekérést, másrészt egy helyre gyűjti az összes domain adatot és a használathoz szükséges dokumentációt. Technikailag fontos a szabványos csatlakozók (*connector*) alkalmazása, hiszen a pipeline-ok kiépítése során több, egymástól független forrást kell integrálni. Ha nem saját fejlesztésű a portál (*Metabase, Redash*), egy központi dashboard készítő applikációt (*Apache Superset* vagy *Grafana*) vagy ha egy metaadat-governance eszközt (*OpenMetadata, Collibra*) használunk, akkor általában kész, központilag támogatott csatlakozók állnak rendelkezésre. Ezekkel egyszerűvé válik a külső adatforrások beemelése, és adott esetben saját csatlakozók is építhetők az eszközök által kínált keretrendszerben.

Fontos, hogy az infrastruktúra- és adatcsapatok alaposan felmérjék a rendelkezésre álló connectorok kínálatát és képességeit: ha a beépített csatlakozók nem fedik le teljeskörűen a vállalat adatforrásait, mérlegelni kell egy egyedi connector fejlesztésének erőforrás- és időigényét. Bár saját fejlesztésű, központi platform egy közepes vállalat számára sem lehetetlen, ez a megközelítés általában lassabb és költségesebb, mint egy kiforrott szoftver alkalmazása. Szerencsére sok eszköz nyílt forráskódú, így kipróbálásukhoz nincs szükség

nagyobb anyagi ráfordításra, és a testreszabhatóság is a cég kezében marad. Emellett sok esetben elérhető ezeknek az alkalmazásoknak SaaS (*Software as a service*) változata is, amely bár magasabb licenc- és szolgáltatási díjjal jár, extra funkcionalitást kínál, és a teljes üzemeltetési, skálázási feladatot átvállalja. A SaaS megoldások jellemzően nagyvállalati igényekre szabottak, ahol az infrastruktúra-menedzsment komplexitása vagy a kiegészítő szolgáltatások iránti igény indokolja a magasabb költségeket. (R. China, 2025)

### **2.2.2. Governance csapat**

A Data Mesh elméleti elveinek és a vállalati Data Governance irányelveknek megfelelő működés érdekében létrehozott Governance csapat tagjai a cég különböző domain-ágait képviselő szakértők, akik közösen dolgoznak a federált adatkezelési keretrendszer megalkotásán és fenntartásán. Feladatuk, hogy kidolgozzák és betartassák a cégen átívelő adatkezelési szabályokat, megalkossák az adatszerződések és metaadat-szabványok követelményeit, valamint biztosítják az adattermékek dokumentálását és minőségellenőrzését. Emellett a csapat felügyeli a jogszabályi megfelelést, gondoskodik az adatvédelmi előírások érvényesítéséről, és kialakítja a megfelelő jogosultsági struktúrát annak érdekében, hogy csak az arra jogosult felhasználók férhessenek hozzá érzékeny adatokhoz. (Harrier és Christ, 2022). A Governance csapat munkáját általában egy központi Data Governance eszköz segíti, amely biztosítja a feladataik központosított kezelését és átlátható menedzsmentjét.

#### **2.2.2.1. Policy-k és szabványok kialakítása**

A policy-k a vállalat központi előírásai, amelyeket a domain csapatoknak minden esetben követniük kell. A decentralizált működés keretein belül a csapatok testre szabhatják, hogyan alkalmazzák ezeket az irányelveket saját adattermékeikre, de a szabályok tartalmát vagy értelmét nem írhatják felül. (Wider, Verma és Akhtar, 2023)

A Governance csapat felelős azért, hogy a policy-k kidolgozása egységesen, közösen történjen, és a vállalat egészére vonatkozó sablon alapján készüljön el, például Harrier és Christ javasolt formátuma szerint. Bár a sablon testreszabható, a dokumentum szerkezete és alapvető elemei minden szervezetnél hasonlóak kell, hogy legyenek.

A policy-k megalkotása a releváns üzleti problémák azonosításával indul, majd a Governance csapat írásos javaslatot készít a megoldásra. A tervezés során figyelembe kell venni a technológiai feltételeket, költségeket és a domain-csapatok működésére gyakorolt hatást. Nagyobb irányelvek esetén célszerű bevonni a Data Platform csapatot az architektúrális, illetve az Enabling csapatot a bevált gyakorlatok érvényesítésére. Az irányelvek betartását

automatizált ellenőrzési mechanizmusok és pipeline-ba épített minőségellenőrző pontok biztosítják. A dokumentáció fenntartása érdekében ajánlott központi, verziókezelte policy-tár (pl. *Wiki*, *GitHub*, *Azure DevOps*) használata. (Harrier és Christ, 2022)

#### **2.2.2.2. Metaadat- és Adaszterződés menedzsment**

Az adatvagyron átlátható kezelése érdekében minden adatterületet nyilvántartásba kell venni, mivel a domének adatai eltérő formátumban és tárolási helyeken fordulnak elő. A Governance csapat katalógusba rendezi és címkézi az adattermékeket, így kulcsszavas kereséssel elérhetők a táblák, mezők, források és a transzformációk. A háttérben futó pipeline-ok biztosítják a frissítést, az adatminőség-ellenőrzést és az adatszterződések betartását, amelyek rögzítik a frissítési ütemet, a kötelező mezőket és a maszkolási szabályokat. A szerződések rendszeres felülvizsgálata csökkenti a hibás vagy hiányos adatszolgáltatásokat.

A fenti mechanizmusokat kiegészíti a központi üzleti szótár (*business glossary*), amely egységes definíciókat ad a kulcsfogalmakhoz (pl. „*beruházási költség*”, „*alkatrészazonosító*”, „*felhasználási arány*”), biztosítva, hogy a domének azonos módon értelmezzék őket. A Governance csapat a metaadat-katalógus részeként tartja karban a szótárat, rögzítve a fogalmak leírását, üzleti szabályait és adatszterződéseit. Ezáltal a felhasználók pontosan értik a mezők jelentését és az előírásokat, ami növeli az adatok megbízhatóságát és az elemzések konzisztenciáját. (Tsai, 2021)

Az üzleti szótár felépítése önmagában is jelentős feladat. A Governance csapatnak az érintett domének szakértőivel folytatott interjúk során kell összeállítania a vállalat kulcsfogalmainak objektív, egyértelmű meghatározásait. Ez magában foglalja annak eldöntését, hogy mely adattípusok esetén indokolt egy részletes glosszárrium, és melyeknél nem, valamint a fogalmak közötti kapcsolatok pontos leírását is. Bár a folyamat hosszadalmas és komoly erőforrást igényel, a megszerzett üzleti tudástár fenntartása hosszú távon megtérül, hiszen jelentősen javítja az adatok közös értelmezését és a szervezeti koherenciát.

#### **2.2.2.3. Minőségbiztosítás és monitorozás**

A Governance csapat csak akkor tudja garantálni az adatok megbízhatóságát, ha a policy-k és adatszterződések előírásait folyamatosan, automatizáltan ellenőrzi. Mivel manuálisan több száz vagy ezer adattétel felügyelete lehetetlen, a minőségbiztosítási pipeline-okkal tetszőleges szinten (tábla, oszlop, mező) definiálhatók és futtathatók validációk. Minden szabályszegésre a policy-ban előírt feltételek és az adatszterződés követelményei alapján automatikus riasztás állítható be, ám nem minden hiba azonos súlyú. Ezért érdemes minden minőségi ellenőrzést

súlyozni, és a teljes adatvagyon minőségét egy százalékos, összetett mutatóval kifejezni. Például egy-egy üres mező önmagában nem feltétlenül kritikus, de ha egy táblában a hiányzó értékek aránya meghaladja a megadott küszöböt, az jelentősen rontja az adatvagyon általános minőségét. Végül a monitoring nem csupán a hibák jelzéséről szól, hanem az ellenőrzési metrikák folyamatos gyűjtéséről és elemzéséről is, ami segít a potenciális problémák korai felismerésében és a megbízható adatszolgáltatás fenntartásában. Ezt az automatikus folyamatot adatprofilozásnak (*data profiling*) nevezzük. (Actian, 2024)

Érdemes hangsúlyozni, hogy az adatminőség-ellenőrzés technikai megvalósítása, a validációs pipeline-ok, automatizált javító szkriptek és háttér-menedzsment kiépítése sokszor inkább a Data Platform csapat feladata, mintsem pusztán Governance-szintű tevékenység. Ugyanakkor a problémák meghatározása, a szabályok finomhangolása és a prioritások kijelölése világosan a Governance csapat kompetenciájába tartozik. Véleményem szerint a leghatékonyabb megközelítés az, ha a Governance határozza meg a „mit” és a „miért” követelményeket, a Platform pedig támogatja a „hogyan” megvalósítását. A modern governance eszközök önálló tesztelést és mérést tesznek lehetővé, így a technikai és üzleti együttműködés hatékonyabban támogatja a Data Mesh működését.

#### **2.2.2.4. Hozzáférés-kezelés**

Nagy, sokszínű adatvagyon esetén alapvető, hogy pontosan szabályozzuk a hozzáféréseket, hiszen ez nemcsak az adatbiztonságot erősíti, hanem megkönnyíti a felhasználók munkáját is, mindenki csak a számára releváns adatokhoz jut hozzá. Így elkerülhető, hogy például egy gyakornok indokolatlanul érzékeny vezetői információkhoz férjen hozzá, ami jelentős biztonsági kockázatot jelenthet.

A hozzáférés-kezelésnek több keretrendszere is létezik. Ebben a fejezetben három elterjedt modellt mutatok be: a szerepkör-alapú (*RBAC*), az attribútum-alapú (*ABAC*) és a policy-alapú (*PBAC*) megközelítést. Ezek eltérő működésüktől függetlenül nem zárják ki egymást, gyakorlati környezetben gyakran hibrid modell szerűen működik.

#### **RBAC (Role-Based Access Control):**

A legelterjedtebb modell, ahol a jogosultságokat szerepkörökhöz rendelik, majd ezeket több felhasználóhoz kapcsolják. Ez egyszerűsíti az adminisztrációt, és a hierarchikus felépítés támogatja a „legkisebb jogosultság” elvét (Alturi és Ferraiolo, 2025). Hátránya, hogy nagy szervezetekben a szerepkörök száma gyorsan elszaporodik, karbantartásuk bonyolult és kockázatos lehet. (Shakarzy, 2024)

### **ABAC (Attribute-Based Access Control):**

Az RBAC-ot kiterjesztve attribútumok alapján dönt (pl. csoporttagság, adatérzékenység, időpont). A szabályok „ha–akkor” logikával adhatók meg, így kontextusfüggő és rugalmas szabályozást tesz lehetővé (L. Manor, 2024). Ugyanakkor a sok attribútum és policy átláthatatlanná válhat, ami nehezíti az auditot és a karbantartást. (Brook, 2023)

### **PBAC (Policy-Based Access Control):**

A szerepkör- és attribútumalapú megközelítést ötvözi, az üzleti irányelveket közvetlenül technikai policy-ként érvényesítve, tipikusan policy-motorokon (pl. *Open Policy Agent*) keresztül. Skálázható és jól auditálható, de bevezetése összetett, mély szakértelmet és szigorú verziókezelést igényel. (Popa, 2023)

Összességében úgy látom, hogy egy közepes méretű vállalat számára az RBAC egyszerűsége és átláthatósága biztosítja a szükséges alapot, amely ABAC-elemekkel bővíthető a dinamikus igényekhez. Speciális, erősen szabályozott ágazatokban (pl. biztosítás, e-kereskedelem, közművek) a PBAC alkalmazása indokolt lehet, hiszen a komplex compliance-követelményekhez a statikus és attribútumalapú modellek már nem elegendők. Nagyvállalati szinten a PBAC előnyei általánosan is megmutatkoznak, mivel a policy-centrikus szabályozás hosszú távon skálázhatóbb és üzletileg rugalmasabb megoldást kínál.

#### **2.2.2.5. Governance szerepkörök**

Az előző fejezetekben áttekintettük a hozzáférés-kezelés alapelveit. A hatékony működés érdekében célszerű a domain csapatokon belül világos szerepköröket kialakítani. Ezek többnyire nem új pozíciókat, hanem a meglévő munkakörök bővített felelősségeit jelentik. Kiemelt szerep az adatgazda (*Data Steward*), amelyet gyakran tévesen összekevernek az adattulajdonos (*Data Owner*) feladatkörével. Az adattulajdonos a domain teljes adatvagyonáért és a kapcsolódó szabályokért, stratégiai döntésekért felel, jellemzően vezetői szinten. Az adatgazda ezzel szemben operatív szinten dolgozik, gondoskodik az adatok minőségéről, elérhetőségéről és dokumentációjáról. Gyakran elemzők vagy más szakértők látják el ezt a feladatot. A két szerepkör így kiegészíti egymást: az adattulajdonos üzleti, az adatgazda gyakorlati szinten biztosítja a domain adatvagyonának átlátható és szabályozott kezelését. (Meganatha, 2025)

A Data Mesh szemléletben a mérnöki feladatkörök domain-orientált szerepekre bonthatók. Kiemelt szerep az adattermék-menedzser (*Data Product Manager*), aki az üzleti és

technológiai szempontokat hangolja össze, hogy a domain adatai strukturáltan, jól definiáltan és újrahasznosítható formában jelenjenek meg. Az üzleti és technikai kompetenciák szétválasztása esetén létrehozható az adattermék-fejlesztő (*Data Product Developer*) szerepkör is. Míg a menedzser az üzleti igények felméréseért és a fejlesztési irányok meghatározásáért felel, a fejlesztő technikailag valósítja meg a pipeline-okat, adatforrás-integrációt és a naprakészség fenntartását. Együttműködésük biztosítja, hogy a domain adatvagyonra mind üzleti, mind technológiai szempontból átlátható és más ágazatok számára is könnyen felhasználható legyen, erősítve a kapcsolatot a Data Platform és a Governance csapat között. (DataGalaxy, 2024)

#### **2.2.2.6. Adatbiztonság**

Üzleti szempontból kiemelten fontos az adatok biztonságos kezelése és a jogszabályoknak való megfelelés, mivel ezek elmulasztása nemcsak pénzügyi, hanem reputációs kockázatokat is hordoz. Az Európai Unió területén működő vállalatoknak különösen nagy figyelmet kell fordítaniuk a vonatkozó adatvédelmi előírások betartására, hiszen a szabályszegéseket súlyos bírságok és negatív nyilvános visszhang követheti. A GDPR (*General Data Protection Regulation*), amelyet 2016-ban fogadtak el, szigorúan szabályozza a személyes adatok kezelését. A rendelet 83. cikkelye kimondja, hogy az adatvédelmi jogsértésekért kiszabott bírságoknak „hatékonyak, arányosnak és visszatartó erejűnek” kell lenniük. A bírságok mértékét és az alkalmazott szankciókat esetről esetre bírálják el, figyelembe véve többek között a vállalat méretét, az adatvédelmi incidens súlyosságát, valamint az együttműködési hajlandóságot is. Ezért elengedhetetlen, hogy a vállalatok már a Data Governance kialakításakor gondoskodjanak a megfelelő adatbiztonsági és adatvédelmi intézkedések bevezetéséről. (Európai Unió, 2016a)

A személyes adat fogalma rendkívül tág, minden olyan információ ide tartozik, amely alapján egy természetes személy közvetlenül vagy közvetve azonosítható. A magyar gyakorlatban gyakran kiemelten védettnek tekintik az úgynevezett 4T (4 természetes) azonosító adatokat is, amelyek együtt szinte minden esetben alkalmasak egy személy egyértelmű azonosítására. Ezek a családi és utónév, a születési hely, a születési idő, valamint az édesanya születési neve. Ezen adatok kezelése különösen körültekintő adatvédelmi eljárásokat igényel. (Orbán, 2022). A 4T adatok az úgynevezett indirekt azonosítók közé tartoznak: ezek olyan információk, amelyek önmagukban nem, de más adatokkal összekapcsolva képesek egy személy egyértelmű azonosítására. Ezzel szemben a direkt azonosítók, például a személyi igazolványszám, az adóazonosító jel vagy a TAJ szám önmagukban is elegendőek egy

természetes személy azonosításához. A direkt és indirekt azonosítók összességére a szakirodalom az PII (*Personally Identifiable Information*) kifejezést használja. A GDPR 4. cikkelye szerint minden PII szerű adatot védelem alatt kell kezelni, függetlenül attól, hogy közvetlen vagy közvetett módon teszi lehetővé egy személy azonosítását. (IBM, 2022) (Európai Unió, 2016b)

A PII-nál ezen felül megkülönböztetünk érzékeny és nem érzékeny adatokat. Érzékeny adatnak tekintünk minden olyan információt, amelynek illetéktelen hozzáférése vagy kiszivárgása az érintett személy számára negatív következményekkel járhat. Ilyen lehet például bármilyen hivatalos azonosító, szexuális hovatartozás, ujjlenyomat vagy akár banki egyenleg. Az ilyen adatok védelme kiemelten fontos, ezért az ezekkel dolgozó vállalatoknak gondoskodniuk kell arról, hogy az adatok megfelelő titkosítás mellett, csak hitelesített hozzáféréssel legyenek elérhetők. Ezekhez az adatokhoz minden esetben kizárólag maga az érintett személy férhet hozzá, miután átesett a szükséges azonosítási és jogosultságellenőrzési folyamatokon. A nem érzékeny adatok ezzel szemben olyan személyes információk, amelyek ugyan alkalmasak lehetnek egy személy azonosítására, de jogosulatlan hozzáférésük általában nem okoz jelentős kárt az érintettnek. Ilyen például az e-mail cím, telefonszám, születési adatok vagy az illető vallása. Ezen adatok védelme ugyan kevésbé szigorú szabályozást igényel, mint az érzékeny adatoké, de ez nem jelenti azt, hogy nyilvánosan kezelhetők lennének, vagy ne kellene rájuk is megfelelő védelmi intézkedéseket alkalmazni. (Harmeling, 2025)

A Governance csapat felelős a Data Mesh modellben az adatbiztonsági intézkedések betartásáért. Fontos a személyes adatok kategorizálása, különös tekintettel az érzékenyekre, mivel jogosulatlan hozzáférésük súlyos következményekkel járhat. A szabályozást célszerű szervezeti szinten, belső policy-kben rögzíteni, míg a technikai megvalósítás a Platform csapat feladata. A modern Governance eszközök katalógus- és klasszifikációs funkciói automatikusan támogatják a személyes adatok azonosítását, címkézését és a kapcsolódó szabályok alkalmazását, így a Governance csapat inkább a megfelelőség felügyeletére és a policy-k fejlesztésére koncentrálhat.

### **2.2.3. Enabling csapat**

Az Enabling (Támogató) csapat feladata, hogy gördülékennyé tegye a Data Mesh bevezetését a domain- és platformcsapatok közvetlen támogatásával. Biztosítja a legjobb gyakorlatok megosztását, az eszközhasználat képzését, valamint a folyamatos visszacsatolási csatornákat. (Christ, Visengeriyeva és Harrier, 2022)

Elsődleges feladataik közé tartozik, hogy mentorálja a domain csapatokat adattermék-pipeline-ok, metaadat-katalógusok és governance policy-k kialakításában, valamint segítse a platformcsapatot új eszközök és integrációk bevezetésénél. Workshopokat, oktatásokat és Q&A üléseket tart (pl. egy Governance eszköz használatáról), miközben gyűjti a felhasználói visszajelzéseket, amelyeket a Governance és Platform csapatok felé továbbít. Így az architektúra és a szabályrendszer folyamatosan a valós igényekhez igazítható. Munkájuk révén a szervezet gyorsabban megvalósítja a Data Mesh előnyeit, és önállóan tudja skálázni adattermékeit, elkerülve a műszaki és szervezeti akadályokat.

Az Enabling csapat nem csak a domén- és platformcsapatok támogatásáért felel, hanem annak a kulturális átalakulásnak az ösztönzéséért is, amelyben a vállalat minden szintjén adatvezérelt döntéshozatal válik a normává. A csapat összegyűjti és bemutatja a Data Mesh által szolgáltatott metrikákat és ezt a felsővezetők, külső befektetők, üzleti elemzők és akár a legújabb gyakornokok számára is érthető formában prezentálja, ami segíti a tudatos finanszírozási döntéseket, valamint a modell fenntartását és skálázását. Így az Enabling csapat megteremti azt a környezetet, ahol a Data Mesh nem csupán technológiai architektúra, hanem a vállalat mindennapi működésének szerves része lesz, mindenki pontosan tudja, hogyan járulhat hozzá az adatok felelős és hatékony felhasználásához, ezzel maximalizálva a befektetés megtérülését és elkerülve a tudás elfolyását. (Owczarek, 2023)

## 3. A KUTATÁS MÓDSZERTANA ÉS MÓDSZEREI

### 3.1. A kutatás típusa

Dolgozatomban egy alkalmazott, értékelési jellegű kutatást végzek, amely egy kvalitatív esettanulmányra épül. A kutatás célja a Data Mesh működésének bemutatása Data Governance szempontból, a szervezeti átalakítástól egészen az operatív működésig. Az alkalmazott kutatás jellege abban nyilvánul meg, hogy gyakorlati problémák megoldására és a Data Mesh elveinek valószínű környezetben való vizsgálatára törekszem. Az értékelési aspektus célja, hogy feltárjam az architektúra bevezetésének előnyeit és korlátait, valamint gyakorlati tapasztalatok alapján tudjak egy saját véleményt megfogalmazni és hipotéziseimre választ adni a bevezetett rendszerrel.

### 3.2. Módszertani lépések

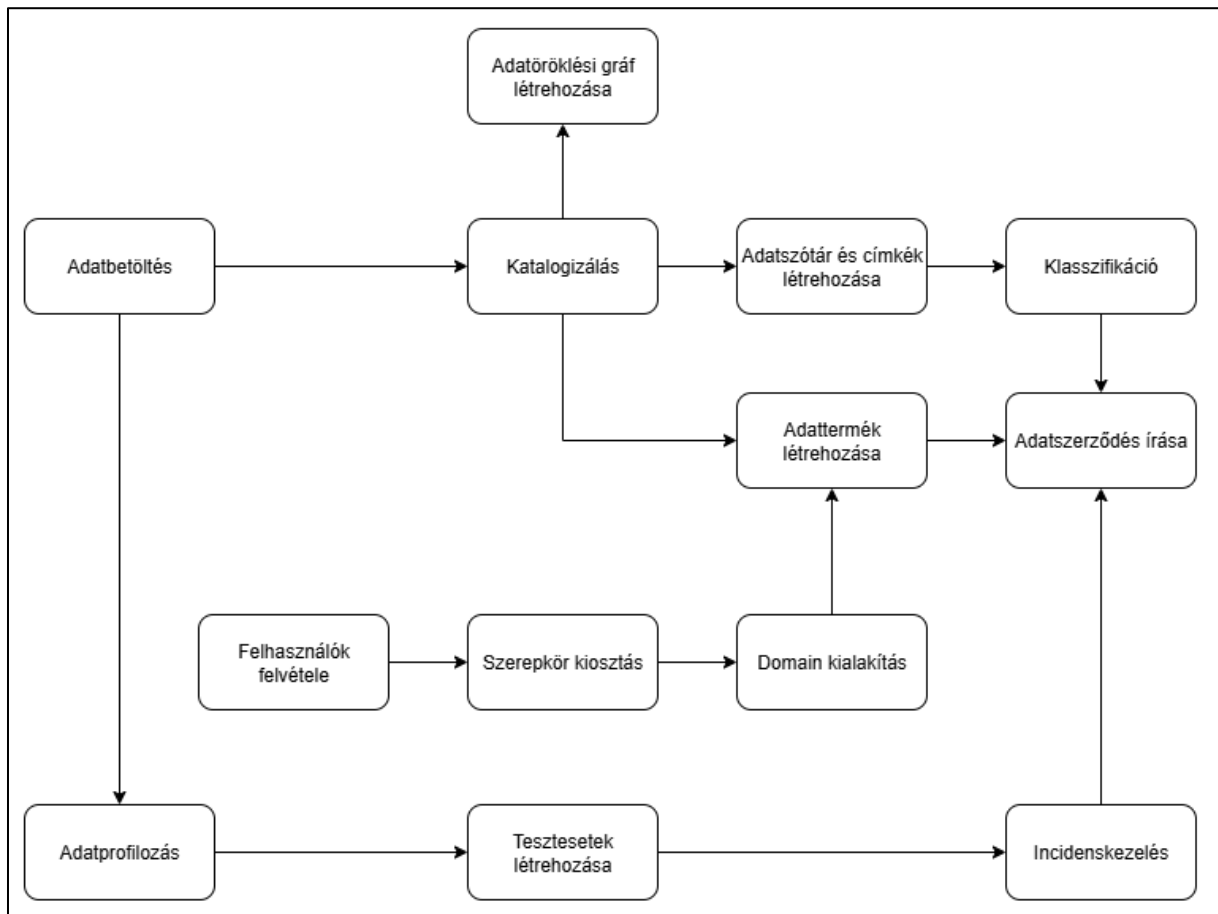
Kutatásom esettanulmányát a feltárt elméleti alapokra építve, de sajátos, alkalmazott megközelítéssel valósítottam meg. Mivel a vizsgálat központi fókusza a Data Governance és a Data Mesh elvek ötvözése volt, ezért a kutatás nem terjedt ki bizonyos technikai megvalósításokra, mint például az adatplatform alkalmazások teljes körű integrációja, API-k létrehozása vagy interfészek fejlesztése. Ezeket a Platform-típusú feladatokat a dolgozat adottnak tekinti, és feltételezi, hogy megfelelően implementált formában állnak rendelkezésre.

Az esettanulmány egy fiktív vállalat működésén keresztül szemlélteti a problémákat és a lehetséges megoldásokat. Valós vállalati adatok felhasználására két okból sem nyílt lehetőség: egyrészt a Data Mesh-t önállóan, teljes körűen alkalmazó szervezetek kis száma Magyarországon, másrészt a személyes és érzékeny vállalati adatok kezelése körüli etikai és adatvédelmi kockázatok miatt. Ennek ellenére a dolgozatban törekedtem arra, hogy a bemutatott helyzetek és megoldások a lehető legélethűbbek és legközelebb álljanak a valós vállalati gyakorlatokhoz.

A módszertan részeként a szakirodalomból feltárt jógyakorlatokat igyekeztem adaptálni, ahol az releváns volt. Ennek során külön figyelmet fordítottam arra, hogy:

- ne írjak elő irreálisan nagy szervezeti átalakításokat
- ne hozzak létre olyan folyamatokat, amelyek feleslegesen bonyolítanák a működést
- és minden esetben üzletileg helyes, ugyanakkor a szimulált vállalat sajátosságaihoz igazított megoldásokat alkalmazzak

### 3.3. Esettanulmány módszerei



**2. ábra: Az esettanulmányi módszerek közötti kapcsolatrendszer. Forrás: Saját Szerkesztés**

A 2. ábra szemlélteti az esettanulmányom által használt módszereket és azok kapcsolatait. Fontos megjegyezni, hogy a nyilak nem feltétlen technikai, hanem módszertani dependenciákat képviselnek. Ebben az alfejezetben ezen módszerek technikaibb és kevésbé intuitív részeit igyekszem bemutatni.

#### 3.3.1. Adatprofilozás

Az adatprofilozás különböző statisztikai metrikákat képes kimutatni a táblákról, ezáltal nagy segítséget nyújt az adatgazdáknak a későbbi tesztesetek kialakításában és a részletes minőségi áttekintésben. Fontos azonban, hogy a profilozás nagy táblák esetén jelentős idő- és erőforrásigénnyel járhat, ezért meg kell határozni, milyen időközönként célszerű futtatni és elkerülni a felesleges futásokat.

Az OpenMetadata ehhez támogatást nyújt mintavételezési paraméterekkel, beállítható százalékos vagy soralapú profilozás is. Százalékos mintavételnél a megadott arány alapján számolja a metrikákat, amely dinamikusan követi a táblaméret változását, ugyanakkor alacsony

arány esetén nő az anomáliák rejtve maradásának kockázata, míg magas arány esetén számolni kell a nagyobb erőforrás terheléssel. Soralapú mintavétel inkább statikusabb táblák esetén célszerű, ezért ezt a funkciót úgy látom, hogy specifikusabb eseteknél érdemes alkalmazni.

### 3.3.2. Adatminőségi tesztesetek és incidenskezelés

Az OpenMetadata low-code és no-code alapú tesztdefiniálási lehetőségeket kínál, melyekkel akár domain-szintű adattisztasági és konzisztencia-ellenőrzések is létrehozhatók. A rendszerben az ellenőrzések egy dedikált pipeline-on keresztül futtathatók, amely a meghatározott tesztek csomagként kezeli, így biztosítva az ismételhető és automatizált minőségellenőrzést. Az eszköz a hibás teszteredményeket automatikusan incidensként rögzíti, és értesítési rendszert biztosít (e-mail, Microsoft Teams, Slack vagy más platformokon), hogy a felelősök gyorsan beavatkozhatnak. Ez integrált megoldást kínál az adatminőség-felügyelet és az üzleti folyamatok összekapcsolására.

Az incidenskezelés során az új hibák alapértelmezetten *New* státuszt kapnak, majd egy éles környezetben hozzárendelhetők egy adatgazdához vagy platformcsapathoz, amely után *Assigned* státuszba kerülnek. Amint az illetékes a hibát tudomásul veszi, az állapot *Ack* értékre vált, a javítás után pedig *Resolved*-ra módosul.

### 3.3.3. Klasszifikáció

Az adatvagyon rendszerezésének egyik legfontosabb lépése az adatok egységes katalogizálása és osztályozása, azaz a klasszifikáció. Célja, hogy az adatállományokat jól értelmezhető leírásokkal, címkékkel és kategóriákkal lássa el, így azok más domain csapatok számára is könnyen azonosíthatók és használhatók legyenek előzetes szakmai ismeret nélkül. A klasszifikáció során az adatokhoz metaadatokat, például leírásokat, adatforrásokat, érzékenységi besorolásokat és üzleti fogalmakat rendelünk, ami növeli az átláthatóságot és támogatja a Data Mesh alapelvét, miszerint az adatokért való felelősség a domain-csapatok szintjén oszlik meg.

Az OpenMetadata eszköz lehetővé teszi, hogy a klasszifikáció a szervezeti struktúrával összhangban történjen. Létrehozhatók a domének, kijelölhetők az adattulajdonosok és adatgazdák, valamint a felelősségi körök. A folyamat rugalmasan igazítható a szervezeti igényekhez, de célszerű előzetesen egyeztetni a domain adattulajdonosaival a címkék, fogalmak és kategóriák meghatározásáról, hogy egységes és következetes rendszer jöjjön létre.

A klasszifikáció kiegészíthető különféle táblaszintű mutatókkal is, amelyek az adatok üzleti vagy technikai jelentőségét segítenek meghatározni. Az OpenMetadata például

lehetőséget kínál a fontossági szint (*Tier*) megadására, amely 1-től 5-ig terjedő skálán értékeli az adatvagyon kritikalitását. Emellett megadható a megőrzési időszak (*retention period*) is, amely metaadatként rögzíthető, így nyomon követhető az adatok érvényessége és lejárata. További besorolási lehetőséget jelent a tanúsítási szint (*certification*), amely arany, ezüst vagy bronz minősítéssel jelöli a táblák megbízhatósági szintjét.

### 3.3.4. Üzleti szótárak és címkézés

Az üzleti szótár (*glossary*) a Data Governance egyik alapvető eszköze, amely egységes fogalmi keretet biztosít az egész szervezet számára. Célja, hogy minden adatfogalomhoz (például *Ügyfél*, *Megrendelés*, *Termék*) pontos, konzisztens meghatározás tartozzon, ezáltal csökkenjen a félreértelmezés és a duplikált jelentések kockázata.

Az OpenMetadata natívan támogatja a szótárak létrehozását és kezelését. A rendszer lehetővé teszi a fogalmak hierarchikus rendezését (például az *Ügyfél* főfogalom alatt a *Prémium ügyfél* és *Lemondott ügyfél* alkategóriákat), amivel átlátható fogalmi struktúra alakítható ki. A platform szinonimákat is kezel, és minden fogalomhoz tulajdonos és ellenőrző (*reviewer*) rendelhető, akik a tartalom karbantartásáért és jóváhagyásáért felelnek. Az üzletileg kritikus fogalmak esetében ez a független kontroll különösen fontos, mivel biztosítja a szótár megbízhatóságát és hitelességét. Ezek a fogalmak nem egy adott táblához, hanem vállalati szintű üzleti koncepciókhoz kapcsolódnak, így a létrehozott szótárak nem korlátozódnak egyetlen adatkészletre. Ennek köszönhetően a szótárbejegyzések később más adatvagyonokhoz is hozzárendelhetők, és a szótárak használatával könnyebben feltárhatók az összefüggések az eltérő adatkészletek között, megkönnyítve a kereshetőséget. Egy jól kialakított üzleti szótár nemcsak a közös nyelv megteremtését szolgálja, hanem támogatja a klasszifikációs folyamatokat és az adatszerződések egységesítését is, ezáltal szoros kapcsolatot teremt a governance-rendszer és a domain-csapatok mindennapi működése között.

A címkék (*tag*) működése hasonlít az üzleti szótárfogalmakéhoz, azonban céljuk eltérő. Nem a fogalmak meghatározására, hanem az adatelemek gyors, kulcsszó alapú kategorizálására szolgálnak. Segítségükkel egyszerűen megjelölhetők például személyes adatot tartalmazó táblák vagy speciális figyelmet igénylő adatkörök, mint a *pénzügyi* vagy *belső használatú* adatok. A címkék egy klasszifikációba szerveződnek, amely technikai értelemben címkék gyűjteményét, governance szempontból pedig az adatok rendszerezésének folyamatát jelenti. Az OpenMetadata rendszer alapértelmezetten tartalmaz több előre definiált klasszifikációt, például *PII* vagy *személyes adat* címkéket, amelyek támogatják az adatvédelmi és biztonsági szabályok alkalmazását. Ezek kiegészíthetők egyedi, szervezeti igényekre szabott címkékkel is.

A címkék nemcsak táblákhoz és oszlopokhoz, hanem akár doménekhez és tulajdonosokhoz is hozzárendelhetők, ugyanakkor tapasztalatom szerint célszerűbb, ha globális és rugalmas struktúrában maradnak. Így biztosítható, hogy a címkék több adatelem között is szabadon használhatók legyenek. A Governance csapat feladata a címketár időszakos felülvizsgálata és finomhangolása, ami garantálja a konzisztens és naprakész használatot anélkül, hogy túlzottan kötött hierarchiához kellene alkalmazkodni.

### 3.3.5. Domain kategóriák behatárolása

Az OpenMetadata három domainkategóriát különít el:

- **forrás-közeli** (*source-aligned*): nyers, operatív adatok
- **aggregált** (*aggregate*): több forrásból származó, konzisztens adathalmazok
- **ügyfél-közeli** (*consumer-aligned*): riportálásra és döntéstámogatásra kész adatvagyon (pl. dashboardok, KPI-riportok)

A klasszifikáció mindig az elsődleges adatjellemzők alapján történik, vagyis a domain akkor is forrás- vagy ügyfél-közeli marad, ha tartalmaz néhány aggregált nézetet. A ritka kivételek nem írják felül az alapbesorolást. Ez a kategorizálás főként a domének közti transzformációk, például pipeline-ok kialakításánál válik hangsúlyossá. Az OpenMetadata eszköz lehetővé teszi a domének és csapatok összekapcsolását is, biztosítva az átlátható felelősségi és hozzáférési struktúrát.

### 3.3.6. Adatöröklés

Az adatöröklés célja, hogy egy adott adat életútját teljes körűen nyomon lehessen követni, akár több rendszeren és feldolgozási rétegen keresztül is. Ezáltal láthatóvá válik az adatok forrása, feldolgozási láncja és minden elágazási pontja, amelyen keresztül az adatok különböző célrendszerekbe eljutnak. A OpenMetadata rendszerben ez a funkció a Lineage ügynök segítségével valósítható meg, de lehetőség van más rendszerekből, például dbt projektből is importálni öröklési gráfot. Az ügynök lefuttatását követően bármely tábla lineage nézete megtekinthető. Egy ilyen ábra jelentős támogatást nyújt komplex adatfolyamatok megértésében, például az adattranszformációs rétegek holisztikus áttekintésében vagy az adatminőségi problémák forrásainak beazonosításában.

### 3.3.7. Adatszerződések

Az adatszerződések publikálása általában egyidejű az adattermék publikálásával, tehát a domain termékének teljesítésében az egyik utolsó fázist jelenti a szerződés létrehozása. Az OpenMetadata szerződés funkciója viszonylag új fejlesztés, így adattermék szinten még nem

alkalmazható teljeskörűen, azonban táblaszinten már jól használható. Egy Data Mesh rendszerben az adatszereződések publikálásáért az adatgazda és az adattermék-menedzser közösen felelnek. Általános gyakorlat szerint az adattermék-menedzseri feladatkör a platform csapat egyik tagjához tartozik. A szerződés létrehozásakor lehetőség van a séma validálására oszlopszinten, kiegészítve klasszifikációkkal. Ezáltal biztosítható, hogy a séma statikus maradjon, és az oszlopok struktúrája ne változzon drasztikusan a későbbiekben.

A szerződésekhez szabályrendszerek is hozzárendelhetők. Ezekben nemcsak a séma- és frissítési szabályok, hanem a felelősségi és szemantikai meghatározások is szerepelhetnek. Ilyen szabály lehet például, hogy egy adott adatkészlet mindig meghatározott domainhez tartozik, kijelölt adatgazdával és előre definiált adatértelmezéssel. Ez a megközelítés lehetővé teszi, hogy az adatok forrása, tulajdonosa és üzleti jelentése minden felhasználó számára egyértelműen azonosítható legyen, ezáltal támogatva az adatvagyon átláthatóságát és a konzisztens adatfelhasználást.

Végül az adatszereződéshez adatminőségi tesztesetek is hozzárendelhetők, amelyek segítségével folyamatosan ellenőrizhető, hogy a tábla technikai értelemben mennyire felel meg az elvárt minőségi követelményeknek.

### **3.4. A kutatás értékelése**

A kutatás sikerességének értékelési szempontja az, hogy az alkalmazott módszerek és az OpenMetadata Governance eszköz képesek-e a felvázolt adatkezelési problémákra módszertanilag helytálló és gyakorlati megoldást nyújtani. Elvárt, hogy a választott megközelítések legalább részben kezelni tudják a vizsgált helyzetekben megjelenő adatminőségi, szervezeti és megfeleléségi kihívásokat, ezáltal alátámasztva a Data Mesh modell valós alkalmazhatóságát.

## 4. ESETTANULMÁNY

Ebben a fejezetben az előző fejezetekben bemutatott elméleti ismereteket egy összetett gyakorlati példán keresztül szemléltetem, amelynek központi eleme a három meghatározott használati eset részletes kidolgozása lesz. Az esettanulmány célja, hogy gyakorlati tapasztalatokon keresztül mutassa be a Data Mesh működését, és hozzájáruljon a modell előnyeinek és hátrányainak későbbi, átfogó értékeléséhez. Az esettanulmány alapjául egy fiktív vállalat, a „Mesh-Electron” szolgál. Két kulcsfontosságú divízióra, az értékesítési és az IT területre helyezem a fókuszot. Ezek szervezeti modelljét Data Mesh alapokra helyezem, adataikat adattermékké alakítom, és a felvázolt használati esetek mentén dolgozom fel, bemutatva a Data Mesh elveinek gyakorlati alkalmazását a mindennapi működésben.

### 4.1. A célcég jellemzői

A kutatás központjába helyezett cég egy fiktív közüzemi szolgáltató. A cég létszáma megközelítőleg 210 fő. A kutatáshoz többféle adatforrást használtam, ami elősegíti egy cég adatstruktúráját a valósághoz legközelebbi módon szimulálni. A részlegek kiválasztásánál számításba vettem, hogy olyan területeket válasszak, amik jellemzően a való életben egy cégnél nem működnek szorosan együtt, tehát az adataik is kellőképp eltérőek. Ez fontos a kutatás szempontjából, hogy szemléltessem, hogy a Data Mesh és a Data Governance elvek milyen formában tudják mégis egyesíteni az együttműködést és az eltérő részlegek között miként alakulhat ki egy hasznos kapcsolat adatkezelés terén.

### 4.2. Adatforrások

A használati eseteim és az esettanulmány kidolgozásához hat saját generált adatforrást alkalmaztam, amelyeket a Mockaroo eszköz segítségével hoztam létre. A Mockaroo célja, hogy élethűnek tűnő, de teljes mértékben fiktív adatokat biztosítson, ami lehetővé teszi számomra, hogy valósághű szimulációkat végezzek anélkül, hogy valódi, érzékeny információkkal kellene dolgoznom. Az eszköz egyik legnagyobb előnye, hogy széleskörűen paraméterezhető és randomizálható. Pontosan olyan adathalmazt tudok generálni, amely illeszkedik az esettanulmány igényeihez, ugyanakkor a véletlenszerűség miatt biztosítja az adatok anonimitását és tényleges fiktivitását is. Ez különösen fontos, mert így a kutatás módszertanilag megalapozott marad, a példák mégis eléggé ismeretlenek, hogy ne legyen magától értetődő az esettanulmányi kidolgozásom. További előny, hogy a Mockaroo véletlenszerűen üres mezőket is képes létrehozni, ami ideálissá teszi a hibakezelési forgatókönyvek szimulálására. Így olyan helyzeteket is modellezhetek, ahol a hiányzó vagy hibás adatok nem előre megtervezett módon, hanem spontán módon fordulnak elő, hasonlóan a való életben tapasztalható problémákhoz.

### 4.3. Governance eszköz

Governance eszköznek az OpenMetadata nevű szoftver a kutatás írása időpontjában legfrissebb verzióját (1.9.7) választottam. Az alkalmazás fejlődését és funkcionalitását az elmúlt egy évben volt tapasztalatom közelről megismerni. Legalkalmasabb eszköznek a kutatásomhoz azért választottam, mivel a piacon levő többi rendszerhez képest ez egy lokálisan könnyen és ingyenesen futtatható alkalmazás és saját kategóriájában is feltörekvően tör utat magának. Az alkalmazás fejlődését közelről figyelem az 1.4.x verziók óta. Az elmúlt évben hatalmas fejlődéseken ment keresztül és a legmodernebb architektúrákkal is tartja aktívan a lépést. A ThoughtWorks technológia elemző cég már 2022-ben felfigyelt rá, mint növekvő nyíltforráskódú eszköz és a 2022 októberi Technology Radar publikációjában pozitív említést is tett róla (ThoughtWorks, 2022), míg az Atlan metadata kezelő cég 2024-ben az egyik legerősebb open-source terméknek vélte az OpenMetadata-t (Atlan, 2024). Teljes funkcionalitását a kutatásomnak nem célja lefedni, de minden amire szükségem van a használati esteim kidolgozásához rendelkezésemre áll az alkalmazáson belül és kitűnő lehetőséget biztosít a Data Mesh gyakorlati bemutatására is.

### 4.4. Egyéb felhasznált eszközök

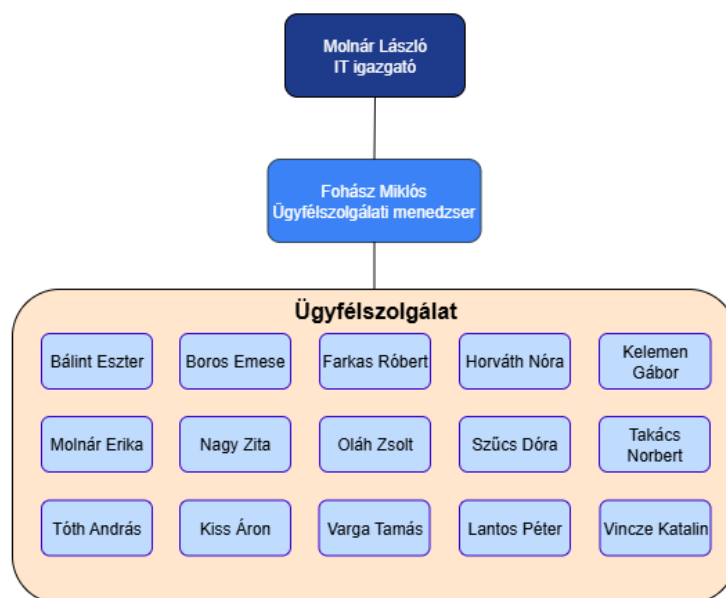
A feldolgozott adatállományokat vesszővel elválasztott .csv formátumban tároltam, amelyet *Microsoft Excel* segítségével kezeltem. Az *OpenMetadata* rendszert saját, lokális környezetben telepítettem *Windows 11* operációs rendszer alatt, *Docker* konténerizációval. A konténerek futtatásához a *WSL2 Linux* virtualizációs környezetet alkalmaztam, háttéradatbázisként pedig *MySQL*-t használtam. A Docker-környezeten belüli ütemezésekért az *Apache Airflow* felelt, amely a metaadat-kezeléshez szükséges pipeline-ok futtatását koordinálta.

Az adatforrások tárolására a *Microsoft SQL Server Management Studio 21* alkalmazást használtam, *SQL Server 2022 Express* verziójú adatbáziskezelővel. Az esettanulmány során használt ábrák és folyamatdiagramok a *draw.io* vizualizációs eszközzel készültek, amely lehetővé tette a modellek és architektúrák átlátható szemléltetését.

## 4.5. Szervezet

### 4.5.1. IT

Az esettanulmányban az IT divízió ügyfélszolgálati részlegére helyezem a hangsúlyt, amely felelős az összes megye ügyfeleinek bejelentett panaszaival kapcsolatos ügyintézésért. A részlegen összesen 15 ügyintéző dolgozik, munkájukat egy ügyfélszolgálati menedzser irányítja, aki közvetlenül az IT igazgatónak jelent. A szervezeti struktúra tehát hierarchikusan épül fel: az ügyintézők a menedzser alá, míg a menedzser az IT igazgató alá tartozik, biztosítva az egyértelmű felelősségi köröket és a hatékony információáramlást. A szervezeti struktúrát a 3. ábra vizualizálja.



**3. ábra: Ügyfélszolgálati csapat szervezeti ábrája. Forrás: Saját szerkesztés**

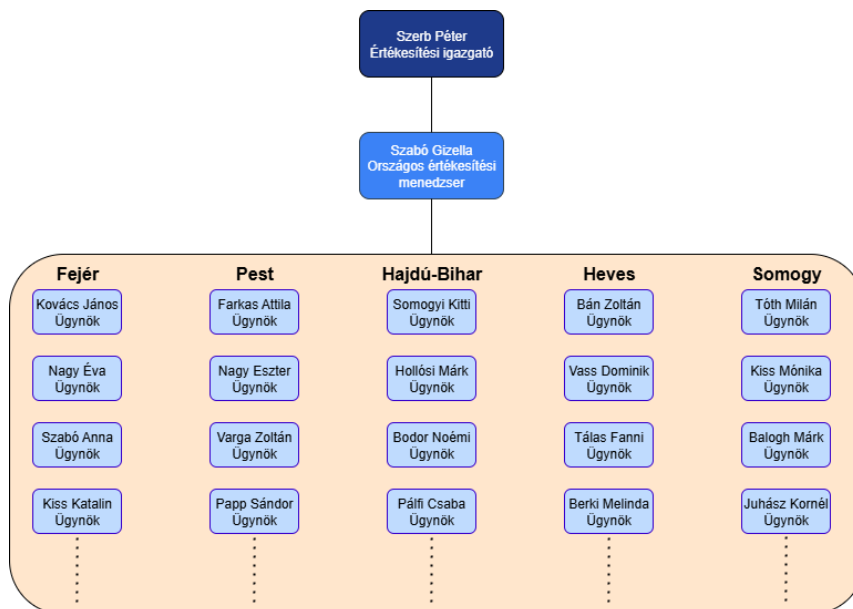
A bejelentett panaszok a vállalat meglévő ügyfeleitől érkeznek, és minden egyes panasz egy kijelölt ügyintézőhöz kerül feldolgozásra. Egy ügyfél több panaszt is benyújthat különböző témákban, és előfordulhat, hogy egy korábbi panasz ismétlődik. Minden panaszbejegyzés egyedi azonosítóval rendelkezik a pontos nyomon követhetőség érdekében. Az ügyfeleket e-mail címük alapján különítik el, mivel lehetséges, hogy több ügyfél is azonos névvel szerepel az adatbázisban.

A leggyakoribb panasztípusok közé tartozik az áramkimaradás, a számlázási hiba, az új bekötés igénylése és a mérőóra csere. A panasz felvételét követően minden bejegyzés három lehetséges státusz egyikébe kerül: Nyitott, ha a panasz beérkezett, de a megoldás még nem kezdődött meg; Feldolgozás alatt, ha a cég már dolgozik az elhárításon; valamint Lezárt, ha a probléma megoldódott. A lezárt panaszokhoz minden esetben rögzítésre kerül a lezárás dátuma is, amely pontosan jelzi a feldolgozási folyamat végét.

Az IT divízió a cég belső Microsoft SQL szerverét használja adatok tárolására, tehát a panaszok egy adatbázis táblában vannak eltárolva.

#### 4.5.2. Értékesítés

A vállalat értékesítési osztályán belül az ügyfélértékesítési csapat felelős az ügyfelekkel való kapcsolattartásért és a szolgáltatások értékesítéséért. A csapat működése területi alapon szerveződik. Az ország lefedett öt megyéje; *Fejér, Heves, Hajdú-Bihar, Pest* és *Somogy* szerint minden megyéhez 15 ügynök tartozik. Amikor egy ügyfél igénybe veszi a cég valamely szolgáltatását, a szerződéskötést mindig egy kijelölt ügynök bonyolítja le. Az ügynökök munkáját egy országos értékesítési menedzser fogja össze, aki az összesített adatokat rendszeresen továbbítja más üzleti területek felé, míg a menedzser felett az értékesítési igazgató áll. A szervezeti működést a 4. ábra szemlélteti.



**4. ábra: Ügyfélértékesítés csapat szervezeti ábrája. Forrás: Saját szerkesztés**

Egy ügyfél akár négy különböző szolgáltatást is igénybe vehet: Áram, EV-töltő, IoT és Napelem. Minden egyes szolgáltatás egy külön ügynökhöz kerül hozzárendelésre, így az ügyfél és az ügynök kapcsolata szolgáltatásonként elkülönülten nyilvántartható. A pontos ügyfél-azonosítás érdekében a szerződéskötés során az ügynökök több személyes adatot is rögzítenek: az ügyfél nevét, e-mail címét, telefonszámát, megyei és települési elhelyezkedését, valamint a lakcímhez tartozó közterület jellegét.

Továbbá minden ügyfélnél nyilvántartják azokat a szolgáltatásokat is, amelyeket korábban igénybe vett, majd lemondott. Ez azért kiemelten fontos, mert ha egy ügyfél a jövőben

újra szeretné aktiválni a korábbi szolgáltatást, a folyamat gyorsan és egyszerűen újraindítható az előző adatok felhasználásával.

Az ügynökök a saját megyei csapataikon belül Excel-fájlokban tartják nyilván az ügyfelek adatait. Ezeket a fájlokat az országos értékesítési vezető kérésére időszakosan összegyűjtik, hogy riportok készülhessenek belőlük. A fájlmegosztás jelenleg manuális módon, az ügynökök közötti többszöri átküldéssel történik.

#### **4.6. Data Mesh bevezetés**

A vállalat vezetősége felismerte, hogy a jelenlegi adatkezelési gyakorlatból hiányzik a strukturált Data Governance és a megosztott adatkezelés (*data sharing*) is csak részben valósul meg. Mivel egyre több helyen hallottak a Data Mesh koncepcióról, szakértőket vontak be annak érdekében, hogy támogassák a bevezetést és a megfelelő architektúra kialakítását. Ezek a szakértők alkotják az Enabling csapat magját, amelynek részeként én magam is szerepet vállalok az esettanulmányban: tanácsadói szerepkörben segítem a vállalatot abban, hogyan illeszthető a Data Mesh a meglévő struktúrájukhoz, és milyen lépésekben érdemes kiépíteni az architektúrát.

A modell bevezetésének első jelentős költsége a szakértők felbérléséhez kapcsolódik. Bár ez kezdetben komoly ráfordítást jelent a vállalatnak, hosszú távon költség- és időhatékonyabb megoldás, mint tapasztalat nélkül, kizárólag belső erőforrásokkal megpróbálni a Data Mesh teljes koncepcióját és technológiai hátterét kidolgozni. A külső szakértők jelenléte legalább az első élesítésig elengedhetetlen, de mivel a Data Mesh komplex modell, és az alkalmazottak számára időt igényel az elvek és folyamatok elsajátítása, célszerű a szakértői tudást hosszabb távon is integrálni a szervezetbe. Ez egy új, belső pozíció formájában valósulhat meg, ahol az Enabling csapat a vállalat szerves részévé válik, és folyamatos támogatást nyújt a Data Mesh működtetéséhez és fejlesztéséhez.

##### **4.6.1. Szervezeti átalakítások**

A megvalósítás első lépése a domain csapatok kialakítása és a megfelelő szerepkörök kiosztása, hiszen a Data Mesh alapjai a szervezeti modellre építik a hozzá kapcsolódó technológiai architektúrát. A meglévő szervezeti struktúra feltérképezése után a szervezet domain egységekre bontható. Az egyik legfontosabb stratégiai döntés, hogy a vállalat inkább nagyobb, átfogó domain egységeket határoz-e meg (például az egész értékesítési osztályt egy domainként kezelve), vagy kisebb, specializált csapatokra bontja fel a szervezetet, ahol például az ügyfélértékesítés különálló domainként működik a nagyobb értékesítési területen belül.

A gyakorlat azt mutatja, hogy célszerű kezdetben a nagyobb, tágabb domain egységek kialakításával kezdeni, majd a bevezetés előrehaladtával fokozatosan, szükség szerint létrehozni az al-domaineket. Ez a megközelítés egyszerűbbé teszi a kezdeti bevezetést, és elkerülhetővé válik, hogy a szervezetet túl korán terhelje a domain-szintű szétaprózódás. Ugyanakkor a vállalat hosszú távon számol azzal, hogy bizonyos al-szekciók (például ügyfélértékesítés vagy kampánymenedzsment) elkülönülő, érzékeny adatokat kezelnek, ezért indokolt lehet a kisebb domain-szintű szétválasztás is. Ezeket egyetlen közös domain alatt kezelni komplikációkat okozhatna, így a fokozatos bontás átgondoltabb és költséghatékonyabb megoldás. A cég vállalja, hogy a többrétegű domain-modell nagyobb governance igényt támaszt, mivel több híd kiépítése válik szükségessé a kisebb domének és a felettük álló nagyobb domain között. Ez viszont biztosítja, hogy a domain-szintek közötti együttműködés gördülékeny maradjon, és a Data Mesh hosszú távon is fenntartható és skálázható módon működjön.

Ha a domének kialakítása megtörtént, a Data Mesh bevezetését érdemes fokozatosan, csapatról csapatra végrehajtani. Ez a megközelítés lehetővé teszi, hogy a központi platformépítés párhuzamosan meginduljon, miközben a bevezetés nem ütközik a több domain egyidejű kezeléséből fakadó szervezeti és technológiai nehézségekbe. Egy adott domainen belül első lépésként ki kell osztani a Data Mesh működéséhez szükséges szerepköröket és feladatokat. Fontos felmérni, hogy a meglévő munkatársak rendelkeznek-e a szükséges kompetenciákkal, vagy bizonyos funkciók ellátásához további szakemberek bevonása válik szükségessé. A fő meghatározott szerepkörök az adattulajdonosok és adatgazdák kijelölése, emellett elengedhetetlen a domain tárolási és architektúrális igényeinek pontos felmérése is, hogy a platformcsapat ennek alapján a megfelelő technológiai környezetet tudja kialakítani. Ha egy domain csapat szervezetenként késznek érzi magát továbblépni a technológiai megvalósításra, az adattermék-menedzseri feladatkört is ki lehet osztani.

#### **4.6.2. Adatplatform megtervezése**

A második lépés a technológiai alapok kiépítése. A vállalat központi adattárolóként a meglévő Microsoft SQL Servert választotta, amely több divízió adatait kezeli, így nem igényel új beruházást, mégis lehetővé teszi a domének elkülönített, de egységes tárolását. Az esettanulmány egyszerűsítése érdekében feltételezzük, hogy az IT és az értékesítési adatok is ezen a platformon, külön adatbázisokban kerülnek elhelyezésre.

Miután felmértük a tárolási igényeket, a következő lépés a vállalati architektúra további bővítése és a teljes adatplatform kialakítása. Ehhez elengedhetetlen az adatcsapat által

kidolgozott pipeline-megoldások integrálása, amelyek biztosítják az adatok egységes feldolgozását és áramlását a platformon belül. Ez a szakasz jelentős költséget jelenthet egy olyan cég számára, amely nem rendelkezik tapasztalt adatcsapattal vagy meglévő architektúrával, mivel gyakran új szakemberek felvételét vagy tanácsadó cég bevonását igényli. Ezek a ráfordítások azonban hosszú távon megtérülnek, hiszen az adatfeldolgozás és adatkezelés alapjait teremtik meg.

Az adatfeldolgozási rétegek után szükség van egy komponensre, amely a felhasználók számára hozzáférhetővé és értelmezhetővé teszi az adatokat. Ehhez a vállalat a Microsoft-környezetbe illeszkedő, költséghatékony Power BI-t használja, amelyet már jelenleg is alkalmaz elemzési és riportálási feladatokra, így az integráció további beruházást nem igényel. A működtetést az adatcsapat elemzői végzik.

A következő lépésben a Data Mesh egyik kulcsfontosságú elemét, a Governance-réteget integráljuk a modellbe. Ehhez olyan Data Governance eszközt kerestünk, amely képes integrálódni az összes, általunk használt adatrendszerrel, támogatja az összetett adatkezelési feladatokat és az adatintegritás megőrzését, emellett biztonságos és költséghatékony megoldást kínál. Ezen szempontok alapján az OpenMetadata nyílt forráskódú platform mellett döntöttünk, mivel széles körű natív integrációs lehetőségeket biztosít, funkciókban rendkívül sokoldalú, és ingyenesen is elérhető. További előnye, hogy ha a jövőben kiterjesztett funkcionalitásra vagy vállalati szintű támogatásra lesz szükség, az OpenMetadata zökkenőmentesen migrálható a Collate nevű SaaS-szolgáltatásra, amely nagyobb méretű vállalati igényekre is optimalizált. Emellett a szoftver bevezetése és használata technikailag viszonylag egyszerű, egy infrastruktúra-üzemeltető gyorsan telepítheti, egy adatmérnök vagy adatgazda pedig könnyen integrálhatja és használhatja a mindennapi adatkezelési folyamatokban.

A végső lépés a platform bemeneti interfészeinek megtervezése. Mivel a vállalat nem várhatja el, hogy minden domain csapat mély technológiai ismeretekkel, például bonyolult SQL-lekérdezésekkel vagy adatmozgató megoldásokkal kezelje a saját adatait, szükség van felhasználóbarát bemeneti megoldásokra is. Ezek az interfészek nem csupán a munkatársak munkáját könnyítik meg, hanem biztosítják az adatok alapvető formai ellenőrzését és előfeldolgozását is, valamint visszajelzést adnak arról, hogy az adatrögzítés vagy módosítás megfelel-e a vállalati előírásoknak. Egy jól megtervezett felület tehát két célt szolgál: egyrészt csökkenti mind a felhasználói, mind a rendszeroldali hibák előfordulását, másrészt garantálja, hogy az adatok már a bevitel pillanatában megfelelő minőségben és egységes formátumban kerüljenek a rendszerbe.

### 4.6.3. Céges policy és hozzáférés-kezelés

A szervezeti és technológiai alapok kiépítése után a vállalat készen áll arra, hogy a Data Mesh architektúrát saját működéséhez igazítsa. Ebben az adattulajdonosok és az adatgazdák közösen alakítják ki az adatkezelési irányelveket, amelyeket a platformcsapat technológiai szinten valósít meg, biztosítva az automatizációt és a megfeleltetést.

A hozzáférés-kezelés meghatározó eleme a modellnek, mivel világosan rögzíteni kell, hogy az egyes adatokhoz ki, milyen feltételekkel és milyen szinten férhet hozzá. A hozzáférési jogokat a domain adattulajdonosai határozzák meg, míg az adminisztráció az adatgazdák feladata. Ez a szerepmegosztás biztosítja, hogy a szabályok egyszerre feleljenek meg az üzleti igényeknek és a biztonsági követelményeknek.

A policy-k kidolgozásával, technológiai megvalósításával és jogosultság-kezeléssel a vállalat ténylegesen bevezette a Data Mesh modellt. A következő feladat a rendszer folyamatos értékelése és fejlesztése, annak érdekében, hogy az architektúra hosszú távon is hatékonyan támogassa az adatvezérelt működést. Az Enabling csapat ebben a szakaszban is fontos szerepet játszik, mivel képzésekkel, iránymutatással és fejlesztési javaslatokkal segíti a szervezetet az új modell fenntartásában és továbbfejlesztésében.

## 4.8. Használati esetek

A használati esetek kidolgozásának módszertana ott indul, hogy az Enabling csapat feltérképezi az adott részleg jelenlegi problémáit, majd ezek alapján keresi meg a megfelelő megoldásokat. E folyamat során kialakításra kerül az átalakított szervezeti modell, amelyet később beillesztünk a Data Mesh architektúrába. A cél egy olyan működőképes, integrált adatkezelési rendszer létrehozása, amely képes a korábban azonosított problémákat kielégítően és fenntartható módon kezelni.

### 4.8.1. Használati eset 1

*„Az IT ügyfélszolgálati divízióján belül a ticketing rendszerből származó adatok hiányosan és hibásan kerülnek átadásra, aminek következtében számos rekord felhasználhatatlanná válik. Ez súlyos fennakadásokat okoz más analitikai rendszerekben és üzleti folyamatokban, valamint az ügyfélélményt is rontja, akár ügyfélvesztést vagy a vállalati hírnév sérülését eredményezve, mivel kritikus bejelentések elveszhetnek. A vállalati vizsgálatok kizárták az emberi mulasztás lehetőségét: a probléma gyökere az architektúrában keresendő. Hogyan lehet olyan megoldást kialakítani, amely biztosítja az adatok minőségének folyamatos monitorozását, és időben figyelmeztet a hibás adatok átadására?”*

#### *4.8.1.1. Felvetett probléma üzleti áttekintése*

Az ügyfélszolgálat divízió szervezeti átalakítása nem igényel jelentős módosításokat, mivel a meglévő struktúra jól illeszthető a Data Mesh szerepkör-kiosztás logikájához. A kibővített feladatkörök alapján Molnár László, az IT igazgató tölti be az IT domain adattulajdonosi szerepét, míg Fohász Miklós, az ügyfélszolgálati menedzser, az ügyfélszolgálati al-domain adatgazdjaként működik. Bár szervezeti szinten a változás nem hoz látványos átalakulást és megoldást a használati esetünkre nézve, a felelősségi körök egyértelmű kijelölése már önmagában is előrelépést jelent. Ennek köszönhetően a hibák feltárása és elhárítása egyértelműen meghatározott személyekhez rendelhető, ami a későbbi mitigációs folyamatok során jelentős hatékonyságnövekedést eredményezhet.

A felmért probléma alapján egyértelműen látható, hogy a domain adatminőségi kihívásokkal küzd. Az adatminőség kezelése kulcsfontosságú feladat minden domain működésében, hiszen ez biztosítja, hogy se emberi, se technikai mulasztás ne vezessen hibás vagy hiányos adatokhoz. A megfelelő minőség nemcsak az adatok belső használatát, hanem azok termékesített formában történő továbbadását és felhasználását is alapvetően meghatározza. A jelenlegi helyzet kezelésében a kiválasztott Governance eszköz, az OpenMetadata hatékony támogatást nyújt, mivel képes automatizált módon monitorozni, ellenőrizni és dokumentálni az adatminőségi folyamatokat, így segítve a hibák gyors feltárását és kijavítását.

#### *4.8.1.2. Adatbetöltés és profilozás*

Kezdeti lépésként az ügyfélszolgálat ticketing tábláját egy metadata ingestion pipeline segítségével importáltam az OpenMetadata katalógusába, amit a beépített MSSQL kapcsolódási ponton keresztül értem el a lokális szerverhez csatlakoztatva a vállalat adatbázisát. A biztonság érdekében az OpenMetadata számára egy dedikált felhasználót hoztam létre, amely kizárólag olvasási jogosultsággal rendelkezik.

A pipeline konfigurálása egyszerű volt, mindössze az adatbázis nevét és a létrehozott felhasználót kellett megadni, majd a futtatás után az adatok automatikusan bekerültek a katalógusba. Az adatminőségi tesztelés előkészítéséhez ezután egy profilozási pipeline-t hoztam létre az OpenMetadata profiling agent funkciójával. Mivel a vizsgált ticketing tábla mindössze 250 rekordot tartalmaz, ezért a teljes körű adatminőség-ellenőrzés érdekében 100%-os mintát alkalmaztam, így elkerülve a részleges vizsgálatból adódó torzításokat. A 4. ábrán láthatjuk is a profilozás utáni oszlopmetrikákat.

#### 4.8.1.3. Adatminőségi tesztesetek

A használati esethez kapcsolódó probléma fő okai már most jól körvonalazódnak: a bejelentés típusa és a státusz mezők üres értékeket engednek befogadni. A közvetlen megoldás átadható a platformcsapatnak, akik a forrásrendszerben vagy az integrációs rétegben elvégezhetik a szükséges módosításokat a hibák megelőzése érdekében. Ez azonban önmagában nem garantálja, hogy a jövőben ne fordulhassanak elő hasonló problémák. A tartós megoldás érdekében olyan tervre van szükség, amely képes folyamatosan monitorozni az oszlop- és táblaszintű változásokat, lehetőleg úgy, hogy a lehetséges hibák minél szélesebb körét lefedje. A minőségellenőrzési folyamat a 3.3.2 fejezetben ismertetett OpenMetadata tesztfunkcionalitásra épült.

A definiált teszteseteim a következők voltak:

- a *ticket\_id* oszlopnak egyedinek kell lennie, hogy elkerülhető legyen bármilyen azonosítási probléma a táblán belül;
- a *státusz* oszlopban kizárólag a három előre meghatározott érték jelenhet meg (*nyitott*, *lezárt*, *feldolgozás alatt*);
- Az *ügyfél*, *email*, *bejelentés dátuma* és *bejelentés típusa* mezőknek kötelezően kitöltöttnek kell lenniük, vagyis a problémafelvételnek teljesnek kell lennie.

A tesztesetek kialakítása során azonban találok bizonyos korlátokkal és dokumentációs hiányosságokkal. A felhasználói felületen elérhető low-code, no-code megoldások az egyszerűbb ellenőrzésekhez kiválóak, de összetettebb szabályok megfogalmazásakor, például annak vizsgálatára, hogy a *lezárás dátuma* mező értéke valóban későbbi legyen, mint a *bejelentés dátuma* már nehezkesebb, és a dokumentáció ezekre az esetekre nem ad részletes útmutatást. Bár a hivatalos források és gyakorlati példák inkább arra ösztönöznek, hogy az ilyen bonyolultabb adatminőség-ellenőrzéseket külön, erre specializált eszközökkel (például *dbt*) végezzük, úgy vélem, hogy ha egy cégnek mindössze néhány hasonló, specifikusabb tesztre van szüksége, akkor egy harmadik fél által kínált megoldás bevezetése indokolatlan többletterhet jelentene. Összességében a saját használati esetemben az OpenMetadata önmagában is kielégítően lefedte az igényeket, és a tesztesetek definiálása egyszerű és kényelmes folyamatnak bizonyult. Az eredmények az 5. ábrán láthatóak.

Status	Name	Column	Last Run	Incident	Actions
Failed	bejelentés típusa kitöltve	bejelentes_tipusa	September 09, 2025, 10:23 PM (UTC+02:00)	New	[Edit] [Delete]
Aborted	lezárás dátuma későbbi, mint a bejelentés dátuma	--	September 09, 2025, 10:23 PM (UTC+02:00)	--	[Edit] [Delete]
Success	ügyfél kitöltve	ugyfel	September 09, 2025, 10:23 PM (UTC+02:00)	--	[Edit] [Delete]
Success	ticket_id egyedi	ticket_id	September 09, 2025, 10:23 PM (UTC+02:00)	--	[Edit] [Delete]
Success	státusz megfelelő értékek	status	September 09, 2025, 10:23 PM (UTC+02:00)	--	[Edit] [Delete]
Success	email mező kitöltve	email	September 09, 2025, 10:23 PM (UTC+02:00)	--	[Edit] [Delete]
Success	bejelentés dátuma kitöltve	bejelentes_datuma	September 09, 2025, 10:23 PM (UTC+02:00)	--	[Edit] [Delete]

**5. ábra: Ticketing táblán lefuttatott tesztesetek eredménye. Forrás: Saját készítés az OpenMetadata adatai alapján**

Az összesített nézetből jól látható, hogy a tesztesetek többsége sikeresen lefutott, egy esetben azonban hibát kaptam. A bejelentés típusa mezőben üres értékek is szerepeltek, a definíció szerint ez nem megengedett. Egy további eset *Aborted* státuszba került, ami azt jelenti, hogy a futtatás technikai okból nem tudott lefutni, így sem sikeres, sem sikertelen eredményt nem adott vissza. Bár ez a fajta hiba sokszor oszlopmetrikákból is felismerhető lenne, a rendszeres, automatizált tesztfuttatás biztosítja, hogy a problémák ne maradjanak rejtve, és időben azonosításra kerüljenek.

#### 4.8.1.4. Használati Eset 1 konklúzió

Összességében a használati eset megoldásában az OpenMetadata egyfajta központi támogató szerepet tölt be. Az eszköz önmagában nem képes teljes körűen kezelni a felmerült problémát, azonban a Data Mesh elvekkel összekapcsolva hatékonyan segíti és előmozdítja a szükséges kollaborációt a csapatok között. A példán keresztül jól láthatóvá vált, hogyan integrálható egy domain adatvagyon a katalógusba, hogyan generálhatók rá metrikák, miként definiálhatók és futtathatók adatminőségi tesztek, valamint hogyan követhetők nyomon és kezelhetők a hibás esetek. Ezáltal a Governance eszköz nem csupán átláthatóbbá teszi az adatkezelést, hanem hozzájárul a felelősségi körök tisztázásához és az adatvagyon megbízhatóságának hosszú távú fenntartásához is.

#### 4.8.2. Használati eset 2

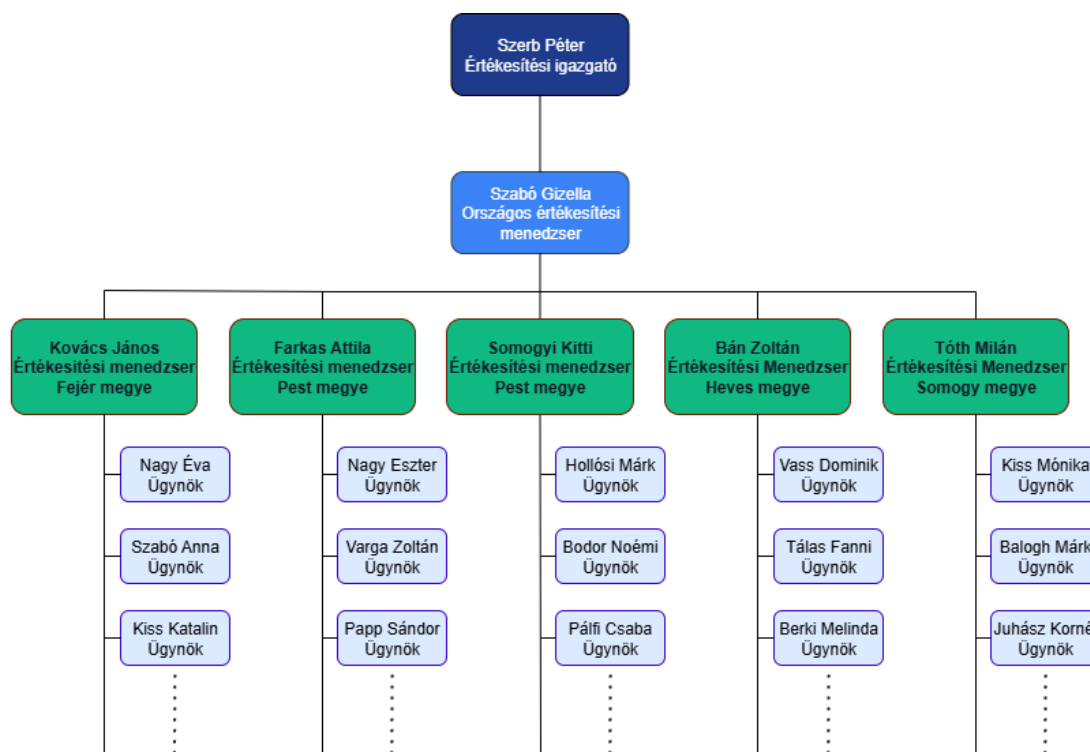
*„A vállalat értékesítési szervezetében, az ügyfélértékesítés részlegén az adatkezelés nem integrált a központi rendszerekkel: a dolgozói információk különálló fájlokban „szétszórva” keletkeznek, amelyek gyakran elvesznek vagy elavulnak, és senki sem vállal felelősséget értük. Ennek eredményeként értékes elemzési lehetőségek maradnak kihasználatlanul, az értékesítés*

*csapat pedig munkaideje nagy részét az adatok utánkeresésére és kézi frissítésére fordítja. Ezt a problémát silósodásnak nevezzük. Hogyan vonhatjuk be hatékonyan az ügyfélértékesítés csapatot a vállalati adatkezelésbe úgy, hogy közben ne terheljük őket bonyolult informatikai feladatokkal, és milyen megoldásokkal tehetjük mindezt egyszerűbbé és átláthatóbbá?*

#### **4.8.2.1. Szervezeti átalakítás**

Az Értékesítési domain jelenleg a vállalat egyik legkevésbé érett divíziója adatérettség szempontjából, ezért a Data Mesh bevezetésének első lépése itt a csapatmodell átalakítása. A jelenlegi működésben az elszórtan tárolt .csv állományok felett nem kijelölt adatgazda felel, ami hosszú távon akadályozza a strukturált és ellenőrizhető adatkezelést. A szervezeti modellre visszautalva logikusan elsőként az Országos Értékesítési Menedzsment meghatározása tűnhet a domain szintű adatgazdai szerepkör betöltésére a legegyszerűbb megoldásnak, mivel ő felel jelenleg is az országos adatkör összesítéséért, ez azonban önmagában nem bizonyul elegendőnek a megosztott felelősségű adatkezelés biztosításához.

A folyamatok átláthatósága és a későbbi ellenőrizhetőség érdekében az Enabling csapat javaslatára, a menedzsmenttel való egyeztetést követően új szerepkör került kialakításra. Minden megyei csapat élére kineveztek egy helyi vezetőt, akiket a csapatok legnagyobb tapasztalattal rendelkező tagjai közül választottak ki. Ez a döntés némileg bonyolította a szervezeti struktúrát, és fontos megjegyezni, hogy éles vállalati környezetben egy középvezetői réteg létrehozása kizárólag egy új szerepkör miatt kevésbé lenne reális megoldás. Az esettanulmányomban azonban tudatosan éltem ezzel a lépéssel, hogy nyomatékosítsam az adatgazdai feladatkörök felosztásának és szerepének jelentőségét a használati eset bemutatásában.



**6. ábra: Átalakított ügyfélértékesítés domain. Forrás: Saját szerkesztés**

A 6. ábra az ügyfélértékesítési domain új szervezeti modelljét mutatja be. A hierarchia átalakítása révén az adatkezelés mostantól többlépcsős ellenőrzési rendszerben zajlik. Minden megyei értékesítési menedzser felel a saját csapatának adataiért, azok minőségéért és helyességéért, mielőtt azok az országos értékesítési menedzserhez kerülnek. Az országos menedzser ezután az összesített adatállomány minőségét és integritását ellenőrzi, így az adatellenőrzés két szinten is megtörténik. Ez a többlépcsős kontroll egy elvárt, beépített redundanciát hoz létre, amely jelentősen növeli az adatok megbízhatóságát és csökkenti a hibás adatok továbbításának kockázatát.

#### **4.8.2.2. Adatplatform integráció és Domain modellezés**

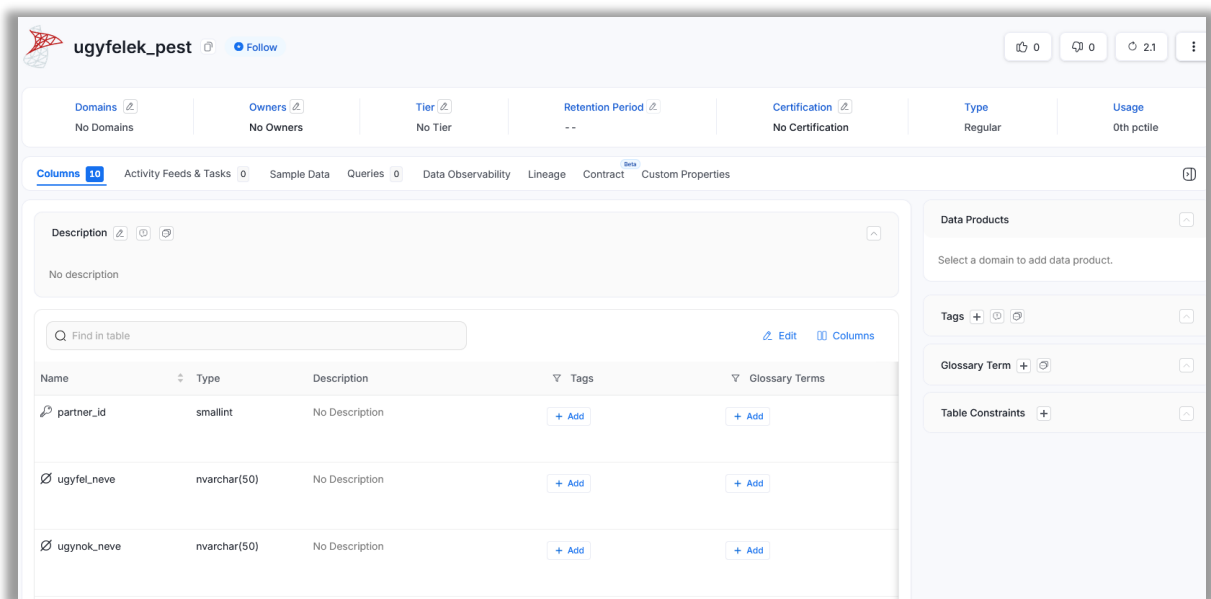
Az értékesítési domain integrációja során a korábbi Excel-alapú adattárolást a vállalat központi Microsoft SQL Server platformjára vezették át, mivel az előző megoldás nem biztosított megfelelő adatminőséget. A Data Platform csapat grafikus interfészeket készített, amelyek lehetővé teszik az ügynökök számára az adatok önkiszolgáló kezelését bonyolult SQL-lekérdezések nélkül. Az országos összesített lista nézetként (*view*) került kialakításra, amely futtatáskor dinamikusan generálja az eredményt az alapul szolgáló táblákból anélkül, hogy fizikailag tárolná az adatokat. Ez biztosítja, hogy az országos adatok mindig naprakészek legyenek, elkerülve a redundáns adattárolást és a manuális karbantartást, miközben könnyen beilleszthetők automatizált riport- és feldolgozó folyamatokba.

A technikai alapok kialakítása után megkezdtem az adatok bevonását az OpenMetadata rendszerbe, ahol a korábban ismertetett klasszifikációs elveket gyakorlati környezetben is alkalmaztam. Az esettanulmány során a cél az volt, hogy az adatok és a csapatstruktúra együttesen jelenjen meg a metaadat-katalógusban, ezzel megalapozva a domének és szerepkörök pontos modellezését.

Elsőként létrehoztam a csapatstruktúrát és a felhasználókat. Az adatgazda szerepkört a menedzserek kapták meg, biztosítva az adatok minőségéért való felelősséget. Ezután kialakítottam az értékesítési domaint és azon belül az ügyfélértékesítési al-domaint. Mivel adataink közvetlenül az ügynökök által kerülnek rögzítésre, a létrehozott domainekeket forrás-közeli kategóriába soroltam. Fontos kiemelni, hogy az aggregált nézetek, például az országos összesítés, nem változtatják meg a domain általános besorolását.

#### 4.8.2.3. Adatvagyon metaadat-alapú kontextualizálása

A következő lépésben az adatokat betöltöttem a OpenMetadata rendszerébe, hasonló módon, ahogyan az első használati eset során is tettem.



**7. ábra: OpenMetadata-ba betöltött tábla klasszifikálás előtt. Forrás: Saját készítés az OpenMetadata adatai alapján**

Az adatok betöltése után a 7. ábrán jól látható, hogy az eszköz automatikusan rögzíti a technikai metaadatokat (oszlopnevek, adattípusok), de ezek önmagukban nem elegendők egy jól használható adatvagyonhoz. Hiányoznak a leírások, címkék, felelősök és fogalomtárbeli hivatkozások, amelyek biztosítják a kereshetőséget és az üzleti kontextus megértését. Az OpenMetadata képes más rendszerekből származó metaadatok átvételére ingestion pipeline-ok

segítségével, így a már megkezdett klasszifikáció automatikusan is importálható. Az esettanulmány egyszerűsítése érdekében itt minden klasszifikáció közvetlenül az OpenMetadata-ban történik.

#### 4.8.2.4. Üzleti szótárak és címkék létrehozása

A táblákat először az ügyfélértékesítési doménhez rendeltem, majd kijelöltem a tábla tulajdonosokat, akik a tartalomért és a hozzáférési szabályokért felelnek. Ezután létrehoztam az üzleti szótárat. Fontos, hogy éles környezetben ne egyetlen adatgazda vezessen be új címkéket vagy fogalmakat, mert ez széttagolódáshoz vezethet; a jó gyakorlat szerint a Governance csapat közösen dolgozza ki őket, és jóváhagyási folyamaton keresztül vezet be.

A saját implementációmban elsőként létrehoztam az *Értékesítés* osztályhoz tartozó szótárgyűjteményt, majd ezen belül felvettem az *Ügyfélértékesítés* fogalmat, amely az értékesítés domain ügyfélkapcsolati tevékenységét írja le. A létrehozott fogalmat hozzárendeltem az adott domain tábláihoz is, így egy külső felhasználó vagy más csapat tagja is könnyen megértheti a táblák üzleti kontextusát pusztán a fogalom alapján. A másik létrehozott szótár az *Üzemeltetés* nevet kapta, amely a vállalat működtetéséhez kapcsolódó üzleti és műszaki tevékenységek fogalomrendszerét hivatott lefedni. A szótárhoz készített definíció a következőképpen hangzik:

*„Az üzemeltetés a Mesh-Electron vállalatnál az elektromos szolgáltatásokhoz kapcsolódó rendszerek, hálózatok és berendezések folyamatos működtetéséért és karbantartásáért felelős tevékenységi kör. Ide tartozik az energiaelosztó infrastruktúra, mérőórák, kapcsolóberendezések és töltőpontok napi szintű felügyelete, hibák gyors elhárítása, megelőző karbantartások tervezése és elvégzése, valamint a szolgáltatásbiztonság folyamatos biztosítása.”*

Az *Üzemeltetés* szótáron belül további fogalomként létrehoztam a *Szolgáltatáscsomag* kifejezést is, amely a vállalat által kínált főbb termék- és szolgáltatási kategóriákat definiálja. Ennek leírása a következő:

*„A vállalat által kínált termék vagy szolgáltatási csomag:*

- *Áram* - általános lakossági áramszolgáltatás
- *IoT* - okosotthonok működtetéséhez szükséges szolgáltatás (okos termékek)
- *EV töltő* - Elektromos gépjármű töltőberendezése

- *Napelem - Napelemes háztartások működtetési szolgáltatása.*”

A létrehozott szótárfogalmak az OpenMetadata rendszerében tábla- és oszlopszinten egyaránt alkalmazhatók, így lehetőség nyílik a pontos és konzisztens hozzárendelések elvégzésére. Ennek megfelelően az *ügyfélértékesítés* szótárfogalmat táblaszinten kapcsoltam az ügyfeladatokat tartalmazó táblákhoz, míg a *szolgáltatáscsomag* fogalmat célzottan a *szolgáltatás* oszlophoz rendeltem hozzá, biztosítva ezzel az üzleti kontextus egyértelmű megjelenítését a metaadat-katalógusban.

A szótárak létrehozását követően a címkék kialakítására tértem át. Az esettanulmányomhoz négy önálló klasszifikációs kategóriát alakítottam ki:

- **Üzleti címkék:** ide soroltam az *ügynök adat*, *ügyfél adat* és *földrajzi adat* címkéket, amelyek az adatok üzleti kontextusát jelölik.
- **Technikai címkék:** ebben a kategóriában hoztam létre az *azonosító* címkét, amely a rendszerben technikai kulcs- vagy azonosítómezőket jelöli.
- **Ügyfélhez kapcsolódó címkék:** ebbe a halmazba került a *kapcsolattartási adat* címke, amely az ügyfelek elérhetőségi adatait (pl. e-mail, telefonszám) fogja össze.
- **Régió:** ez a klasszifikáció a földrajzi bontást támogatja, itt minden megyéhez létrehoztam külön címkét, valamint létrehoztam egy *országos* címkét is, amely a teljes országot lefedő adathalmazokra alkalmazható.

A megőrzési idő és tanúsítási szint alkalmazását nem tartottam szükségesnek, mivel a vállalat adatai tartósan megőrzendők, és a meglévő klasszifikációk már elegendő információt nyújtanak.

#### **4.8.2.5. Autoklasszifikáció**

Az OpenMetadata a PII-adatok klasszifikációjának megkönnyítésére egy automatikus címkéző algoritmust biztosít, amely képes a megadott adatvagyon teljes körű átvizsgálására és a potenciálisan érzékeny oszlopok algoritmikus azonosítására. Ezt az úgynevezett Autoklasszifikációs ügynök végzi. Az ügynök két fő funkcióval rendelkezik: egyrészt mintarekordokat képes leképezni, vagyis a táblák fülén keresztül néhány tényleges adatbejegyzést jelenít meg, hasonlóan ahhoz, mintha közvetlenül adatbázis szinten vizsgálnánk az adatokat. Ez a funkció a második képesség, a PII-klasszifikáció alapját is képezi, mivel a rendszer ezekből a mintákból következtet az egyes oszlopok tartalmának érzékenységére. Az agent működéséhez két paraméter állítható be. A vizsgálandó minták száma, amit az alkalmazás

alapértelmezett értéként 50-re állított be, valamint az autoklasszifikációs konfidenciaszint, amely 1 és 100 közötti értéket vehet fel. Az OpenMetadata dokumentációja nem ad pontos útmutatást az optimális konfidenciaszint kiválasztására (alapértelmezés szerint 80-ra van állítva), ezért saját tesztorozatot készítettem a legpontosabb érték meghatározása érdekében. A hivatalos segédlet szerint az alacsonyabb értékek hajlamosak több fals pozitív, míg a magasabb értékek több fals negatív eredményt adni. A kísérlet során 0 és 100 közötti tartományban, 20-as lépésekben teszteltem az algoritmus pontosságát, hogy meghatározhassam a legmegbízhatóbb konfidenciaszintet. A kísérlet során a bevitt mintaadatok száma mindvégig az alapértelmezett 50 rekord volt, a vizsgálatot pedig az *ugyfelek\_pest* nevű táblán végeztem el. Az automatikus PII-klasszifikáció eredményeit az 1. táblázat foglalja össze.

**1. táblázat: Az OpenMetadata autoklasszifikációs algoritmusának eredménye különböző konfidenciaszinteken.**

<b>Konfidenciaszint (%)</b>	<b>Érzékeny adatként azonosított mezők</b>	<b>Nem érzékeny adatként azonosított mezők</b>
0	ugyfel_neve, ugnok_neve, email, telepules, kozterulet	telefonszám
20	ugyfel_neve, ugnok_neve, email, telepules, kozterulet	telefonszám
40	ugyfel_neve, ugnok_neve, email, telepules, kozterulet	telefonszám
60	ugyfel_neve, ugnok_neve, email, telepules, kozterulet	telefonszám
80	ugyfel_neve, email	-
100	-	-

**Forrás: Saját adatok alapján**

Az 1. táblázat alapján jól látható, hogy a 0 és 60 közötti konfidenciaszinteken az automatikus besorolás eredménye teljesen megegyezett, ami meglepő volt számomra. Alacsonyabb konfidenciaszinteknél arra számítottam, hogy az algoritmus közel az összes oszlopot érzékeny adatként fogja klasszifikálni. Ezzel szemben a magasabb konfidenciaszintek már a várakozásoknak megfelelő mintázatot mutattak. Egy bizonyos küszöb felett az algoritmus láthatóan nem tudja egyértelműen eldönteni, hogy egy adatmezőt érzékenynek minősítsen-e vagy sem, és ennek következtében, ahogy a 100-as értéknél is látható, egyetlen mezőt sem sorolt be. Bár elsőre a 80-as konfidenciaszint (amely az alapértelmezett érték) tűnik a legbiztonságosabbnak, ha visszatekintünk az érzékeny és nem érzékeny adatok szakirodalmi

definícióira, megállapítható, hogy egyik besorolás sem tekinthető teljesen helyesnek. Az algoritmus által érzékenynek jelölt mezők valójában nem minősülnek érzékeny adatnak, így paradox módon éppen a 0 és 60 közötti értékek közelítettek leginkább a helyes kimenetelhez, mivel egyedül a telefonszám mezőt sorolták nem érzékeny adatnak, ami valóban annak tekinthető. Összegzésképpen úgy vélem, hogy az OpenMetadata automatikus PII-klasszifikációs funkciója koncepcionálisan ígéretes, és megfelelő működés esetén nagymértékben megkönnyíthetné a Governance csapat munkáját, azonban a tapasztalt eredmények arra utalnak, hogy jelenlegi formájában nem megbízható. A konfidenciaszintekre való támaszkodás helyett sokkal inkább szükséges az emberi, kézi beavatkozás, amely biztosítja a ténylegesen pontos és jogilag is megfelelő PII-besorolást.

Az adatérzékenységi vizsgálat eredményei alapján az esettanulmányban a táblákat Tier 2 kategóriába soroltam, mivel nem tartalmazznak titkos vagy különösen érzékeny adatokat, de fontosabbak, mint egy belső segéd tábla.

#### ***4.8.2.6. Használati Eset 2 konklúzió***

A klasszifikációs és kontextus-hozzárendelési folyamat lezárásával a táblák immár teljes mértékben megfelelnek a gördülékeny kereshetőség és az egyértelmű üzleti kontextus követelményeinek. Ennek eredményeként az ügyfélértékesítés adatai a korábbi szétszórt, silószerű tárolás helyett egy integrált, könnyen kereshető, üzletileg és technológiailag is jól definiált egységet alkotnak.

Name	Type	Description	Tags	Glossary Terms
partner_id Partner azonosító	smallint	Egyedi azonosító minden ügyfélszerződéshez. Egy ügyfél többször is szerepelhet több szolgáltatás igénybevételekor.	Azonosító	--
ugyfel_neve Ügyfél neve	nvarchar(50)	Az ügyfél teljes neve, ahogy a szerződéskötéskor rögzítésre került.	Nem-érzékeny adat Személyes adat	Ügyfél adat
ugygnok_neve Ügygnök neve	nvarchar(50)	Annak az értékesítési ügygnöknek a neve, aki az adott ügyfél szerződését kezeli.	Ügygnök adat	--
szolgaltatas Szolgáltatás	nvarchar(50)	Az ügyfél általi szerződött szolgáltatás típusa	--	Szolgáltatáscsomag

## 8. ábra: Klasszifikált üzleti tábla. Forrás: Saját készítés az OpenMetadata adatai alapján

Az 8. ábrán jól látható, hogy nem minden elérhető klasszifikációs mező került kitöltésre. Ez tudatos döntés eredménye. Egy jól kereshető adattáblának nem szükséges minden oszlophoz címkét és szótárfogalmat rendelni, mivel a túlzott kontextusvesztés a tábla egyedi identitásának elmosódásához vezethet, valamint megnövelheti a találati zajt, ha a rendszer túl sok irreleváns keresésre is visszaadja ugyanazt a táblát. Ezért egy adatgazdának célszerű üzletileg és módszertanilag is előre megterveznie a klasszifikációs folyamatot, és megtalálnia az egyensúlyt a megfelelően részletes, de nem túlzott besorolás között. Bár az adatklasszifikáció elsöre egyszerű, intuitív folyamatnak tűnhet, a megfelelő üzleti és domainismeret megléte elengedhetetlen a hatékony és konzisztens eredmény eléréséhez.

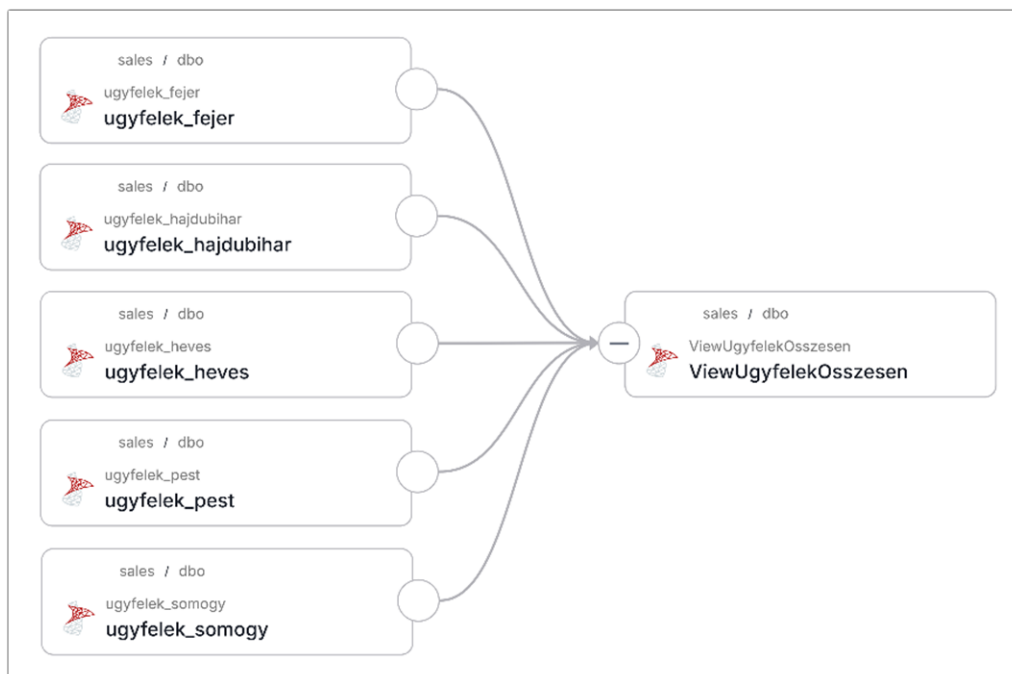
### 4.8.3. Használati eset 3

*„Egy korrábbi ügyfél kéri adatainak teljes törlését. Az érintett adatok több, egymástól független rendszerben és eltérő azonosítók alatt szerepelnek. A jelenlegi folyamat manuális, időigényes, és különböző divíziók együttműködését igényli, gyakran olyan egységeket is bevonva, amelyek normál esetben nem dolgoznak szorosan együtt. Ez felvet olyan problémákat, mint a gyors szervezeti egységek közti kommunikáció hiánya. Hogyan tudnám ezt úgy megoldani, hogy az ilyen speciális esetekben is gyorsan megtaláljam az összes ilyen személyes adatot és bármely másik részleggel is, aktív együttműködéstől függetlenül, tudjak hibamentesen együtt dolgozni?”*

A harmadik használati eset megoldása elsősorban egyszerűbbnek tűnhet az előző problémák megoldásai tekintetében, azonban valójában egy fontos, gyakran alábecsült kihívást kezel, a vállalati adatáramlás átláthatóságát. Mivel a központi adatplatform és a OpenMetadata rendszer integrációja, valamint a Data Mesh elveinek bevezetése már megtörtént, a metaadat-kezelés alapvető keretrendszere rendelkezésre áll. A feladat most az, hogy egy független vállalati szereplőként, például egy auditor vagy adattörlési kérelmet feldolgozó munkatársként képesek legyünk hatékonyan és biztosan azonosítani, hol fordulnak elő az ügyféladatok a szervezet adatvagyonában.

Első megközelítésként a szótárak és címkék alapján végzett keresés tűnik logikusnak, azonban ez önmagában nem elegendő. Bár az adatok megfelelően klasszifikáltak és a metaadat-frissítések ütemezetten zajlanak, a felhasználó nem rendelkezik ismeretekkel az adatok közötti kapcsolatok, függőségek és frissítési láncok működéséről. Ez kritikus probléma, hiszen ha egy nem naprakész vagy más forrásból újratöltődő táblából próbálunk adatot törölni vagy módosítani, a forrás frissítései később felülírhatják a módosításainkat. A Data Mesh architektúrában ezt a problémát az adatörklési (*lineage*) gráf és az adatszerződések (*data contracts*) kezelik.

#### 4.8.3.1. Adatörklődés és Adatszerződés létrehozása



9. ábra: Ügyfélértékesítési adatvagyon adatainak folyama. Forrás: Saját készítés az OpenMetadata adatai alapján

A 9. ábrán a *ViewUgyfelekOsszesen* nézethez tartozó gráf látható, amely vizualizálja, hogy ez a nézet több megyei ügyféltáblából épül fel. A gráf élei interaktívak, és lehetőség van rájuk SQL lekérdezéseket vagy szöveges leírásokat csatolni, így pontosan nyomon követhető, hogy az adat milyen formában, milyen logika mentén került egyik entitásból a másikba. Ez a funkció kulcsfontosságú a használati eset megoldásában, mivel lehetővé teszi, hogy az összes ügyféladat visszavezethető legyen az eredeti forrásukig, így biztosítva, hogy az adattörlési vagy módosítási folyamatok a megfelelő ponton induljanak el.

Az adatszerződés létrehozásakor az adott tábla adatgazdájához rendeltem a hozzá tartozó szerződést, majd részletesen feltérképeztem az adatok szemantikáját és dokumentációját is. Ez a lépés nemcsak a tábla technikai leírását egészíti ki, hanem megerősíti az üzleti stabilitását és értelmezhetőségét is, hiszen egyértelműen rögzíti, hogy az adott adatvagyon kihez tartozik, milyen célra használható, és milyen szabályok mentén kezelendő. Példaként a ticketing tábla esetében két konkrét szabályrendszert definiáltam:

- Az adatgazda minden esetben Fohász Miklós, míg a domain Ügyfélszolgálat.
- A tábla mindig ügyféladatot tartalmaz.

Ez a két egyszerű, de jól körülhatárolt szabály jelentősen megkönnyíti az adatok felderítését és konzisztens kezelését, mivel egyértelműen meghatározza a tulajdonosi és tartalmi felelősséget. Ennek köszönhetően a felhasználók már a szerződésből pontosan látják, hogy mely üzleti területhez és adatgazdához tartozik az adott adathalmaz, valamint milyen típusú adatokat tartalmaz.

További lépésként a ticketing tábla adatszerződéséhez hozzárendeltem az első használati esetben létrehozott adatminőségi teszteseteket is, ezzel biztosítva, hogy az adat nemcsak strukturálisan, hanem minőségileg is megfeleljen az előírt feltételeknek.

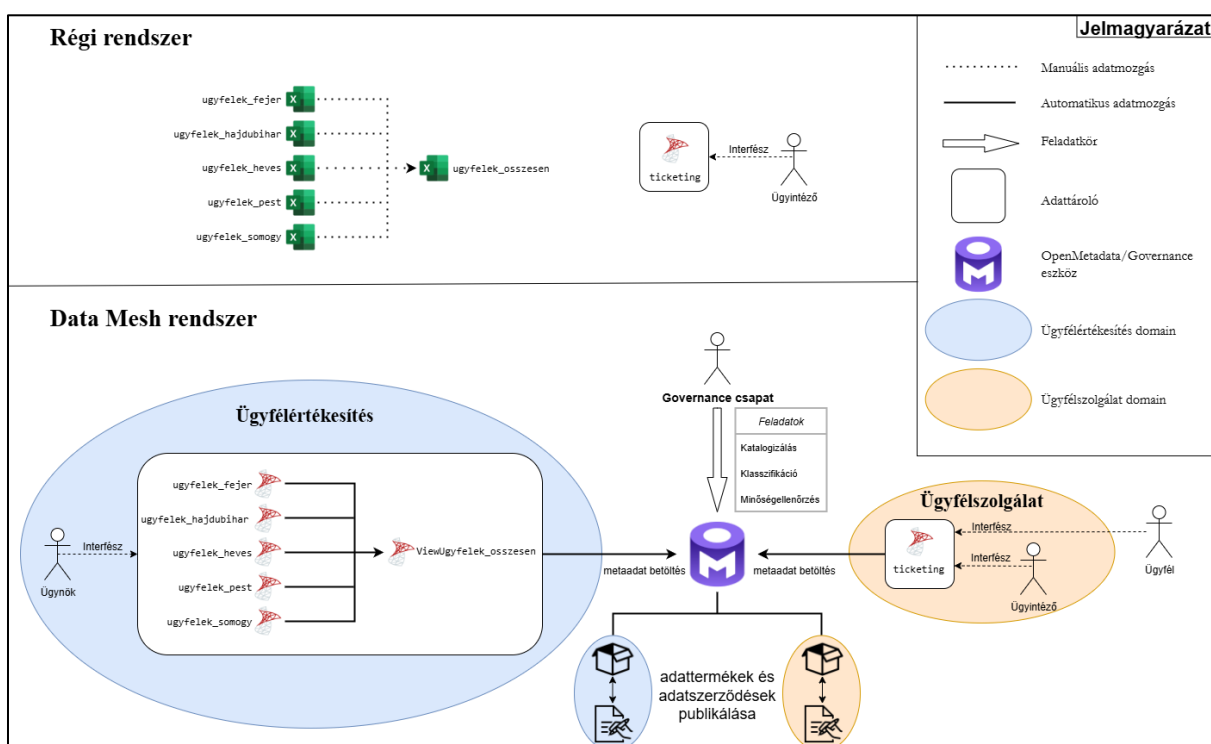
#### **4.8.3.2. *Használati Eset 3 konklúzió***

A ticketing tábla ügyfélspecifikus mezőinek klasszifikálása, az első két esettanulmány eredményei, valamint a kialakított adatöröklési gráf és adatszerződések együtt teljeskörűen lefedik a megoldáshoz szükséges lépéseket. Az így felépített struktúrában az ügyféladatok forrástól a végső adatkészletig átláthatóan követhetők és validálhatók, ami egyaránt támogatja az auditálási folyamatokat és a platformcsapat technikai feladatait. A megoldás kikerüli, hogy az auditoroknak vagy a felelős munkatársaknak több domain képviselőjével kelljen információkat egyeztetniük, hiszen az OpenMetadata által feltérképezett adattermékek önmagukban biztosítják a szükséges kontextust. Ez jelentősen gyorsítja az adatkezelési és

biztonsági folyamatokat, miközben megerősíti a Data Mesh alapelvek gyakorlati érvényesülését. Az esettanulmány során ugyanakkor az OpenMetadata adatszerződés-funkcionalitása még kiforratlannak bizonyult. Különösen hiányoltam a frissítési gyakoriság explicit megadásának lehetőségét, amely az ingestion folyamatok betöltési ütemezését kapcsolná össze a szerződéssel. Ez kritikus tényező lehetne a szerződés átfogó értelmezése szempontjából, mivel így pontos képet kaphatnánk arról, mikor és milyen rendszerességgel frissül az adott adatvagyon. Ideiglenes megoldásként a szerződések YAML formátumban exportálhatók és manuálisan kiegészíthetők ezzel az információval, azonban ez a megoldás kikerüli az OpenMetadata natív szerződéskezelési funkcióit, és hosszabb, kevésbé hatékony folyamatot eredményez.

## 5. EREDMÉNYEK BEMUTATÁSA

Esettanulmányom során részletes és gyakorlati képet adtam a Data Mesh gyakorlati bevezetésének governance vonatkozásairól bizonyítva, hogy a modell több szervezeti és technológiai szinten is kézzelfogható előnyöket eredményezhet. Bár az esettanulmány nem fedte le a teljes architektúrát, a kísérleti megvalósítás egyértelműen demonstrálta a modell skálázhatóságát és adaptálhatóságát. A bevezetés nem csak a vállalat adatkezelési problémáit oldotta meg, hanem kimutathatóan növelte a szervezet adatérettségét és technológiai kompetenciáit. A régi és új szervezeti modell összevetését a 10. ábra szemlélteti, amely áttekintést ad az átalakítások és a governance megfeleltetés fő elemeiről.



**10. ábra: Átalakított architektúra. Forrás: Saját szerkesztés**

A legmarkánsabb előrelépés az Ügyfélértékesítés egységében volt tapasztalható, ahol a szervezeti átalakítás üzleti és technikai szinten is mérhető fejlődést hozott. Az adatfelelősség delegálása és az automatizált adatfolyamok kialakítása stabil, minőségbiztosítási alapokon nyugvó működést eredményezett. Emellett a természetes és klasszifikált metaadatok kezelése révén az új domain adatai kereshetővé, átláthatóvá és teljes körű governance alá kerültek, ezáltal nagymértékben hozzájárulva a vállalat önkiszolgáló adatkezelési képességeinek fejlődéséhez. A választott Governance eszköz ebben a használati esetben bizonyult a legeredményesebbnek. Az OpenMetadata katalogizálási és klasszifikációs képességei a várakozásaimat is felülmúlták, és nagyrésztben elégedett voltam az általa kínált lehetőségekkel.

A legjelentősebb hiányosság ugyanakkor a PII adatok kezelésében mutatkozott meg. Bár az alkalmazás rendelkezik az automatikus érzékenyadat-észlelés keretrendszerével, annak működése nem érte el a szakmailag elvárható szintet. E hiányosság azonban könnyen kezelhető, amennyiben nem cél az automatizált GDPR-megfelelés teljes körű kialakítása. Sőt, bizonyos esetekben előnyös is lehet, ha a klasszifikációk meghatározása manuális munkával történik, hiszen ez növeli az ellenőrizhetőséget és csökkenti a hibás automatikus besorolások kockázatát.

Szervezeti szempontból kevésbé volt látványos az IT-Ügyfélszolgálati divízió átalakulása, ugyanakkor az adatminőségi előrelépések ezen a területen mutatkoztak meg a leginkább. Ez a divízió jó példát adott arra, hogy egy kevésbé technikai beállítottságú Governance csapat is képes hatékony szoftveres hibaellenőrzési folyamatokat kialakítani és működtetni. Bár az OpenMetadata jelenlegi funkcionalitása nem teszi lehetővé fejlesztések és minőségkorrekciók natív megvalósítását, előnyösnek bizonyult, hogy az adatgazdák vagy a kijelölt minőségellenőrök viszonylag egyszerű, low-code/no-code jellegű megközelítéssel is képesek alapvető teszteseteket lefuttatni. Ez az egyszerűség a kisebb és kevésbé komplex környezetekben elegendő lehet, azonban összetettebb rendszerekben elengedhetetlen externális minőségbiztosítási eszközök vagy robusztusabb platformtámogatás bevonása. Az SQL-alapú tesztesetek írása jelenleg még kiforratlan, különösen hátrányos, hogy hibás lefutás esetén az eszköz nem biztosít egyértelmű hibaüzenetet vagy korrekciós lehetőséget. Ez komoly korlátot jelent bonyolultabb lekérdezésekkel dolgozó tesztelési scénáriókban. Ugyanakkor feltételezhető, hogy ezen a téren további fejlesztések és bővített funkcionalitások várhatóak, amelyek hosszabb távon kiküszöbölhetik a jelenlegi hiányosságokat.

A harmadik használati eset megoldása rávilágított arra, hogy az adattermékek és adatszerződések létrehozása, bár első ránézésre egyszerű folyamatnak tűnik, jelentős mértékben épül az előző két használati eset eredményeire. A módszertan ezzel demonstrálta a Data Mesh egyik legfontosabb pillérét, az adatok termékesítését. Az OpenMetadata adatszerződés-funkcionalitása ebben a használati esetben kifejezetten hasznosnak bizonyult, ugyanakkor hiányosságai is szembetűntek. Bár a rendszer lehetőséget biztosít szemantikai szabályok, sémaellenőrzések és adatminőségi kritériumok definiálására, nem tartalmaz olyan paramétert, amellyel az ingestion folyamatok időzítése és frissítési gyakorisága közvetlenül rögzíthető lenne a szerződésben. Ez komoly hiányosság, hiszen az adatok aktualitása és életciklusa a szerződés egyik legfontosabb dimenzióját képezhetné. Alternatív megoldásként ugyan van lehetőség YAML konfiguráción keresztül manuálisan hozzárendelni a frissítési gyakoriságot, ez azonban csökkenti az OpenMetadata-n belüli egységes szerződéskezelés átláthatóságát és

hatékonyságát. Az adatszerződés publikálásáig vezető út jól mutatja, hogy egy vállalatnak mennyi egymással összefüggő folyamatot kell összehangoltan működtetnie. Míg egy érett Data Mesh környezetben ezek a lépések rutinszerűvé válhatnak, addig egy bevezetési fázisban lévő szervezet esetében a folyamat nem triviális, és számos erőforrást igényel. A tapasztalat azt mutatta, hogy a Data Mesh ambíciói olyan szintű szervezeti együttműködést várnak el, amely a gyakorlatban fokozatosan, hosszabb próbaidőszak és iterációk árán tud kialakulni, nem pedig azonnal.

### 5.1. A modell bevezetése

A Data Mesh és a Data Governance középvállalatok számára komoly szervezeti és technológiai előnyöket hozhat, de bevezetésük jelentős kihívásokkal jár. A közepes méretű cégek általában nem rendelkeznek olyan robusztus szervezeti háttérrel, mint a nagyvállalatok, így a decentralizált adatkezelés fenntartása nagy erőforrásigényű. Az esettanulmányok rámutattak, hogy az adatgazdák szerepe kritikus. Egyszerre felelnek saját adatvagyonukért és a vállalati governance illeszkedéséért, ami jelentős kapacitást igényel. További kulcstényezők a hozzáférés-kezelés és az adatplatform folyamatos fejlesztése, amelyekhez dedikált platformcsapat szükséges.

A bevezetés hatékonysága a vállalat méretétől és érettségétől függ. Egy 50-150 fős cég számára a Data Mesh nem feltétlenül térül meg, mert az adatvagyon hagyományos eszközökkel is kezelhető. 150–250 főnél viszont már sürgetővé válik a decentralizált adatkezelés és az egységes governance, mivel a komplexitás gyorsan nő. Ebben a szakaszban a Data Mesh hosszú távon javítja az adatbiztonságot, támogatja a növekedést és növeli az adat- és technológiai érettséget. A modell sikeréhez azonban belső nyitottság is kell. A vállalatnak kulturálisan és technológiailag egyaránt készen kell állnia az együttműködésre és a közös felelősségvállalásra. A legnagyobb korlátot ott jelentheti a bevezetés, ahol nincs, vagy csak minimális létszámú fejlesztői csapat áll rendelkezésre, mivel a platformcsapat a Data Mesh egyik alappillére. Konklúzióként elmondható, hogy a Data Mesh és a Data Governance középvállalatoknál akkor tudják a legnagyobb értéket teremteni, ha a szervezet már a nagyvállalati szint felé tart, képes hosszú távon fenntartani a modellhez szükséges erőforrásokat, és egyúttal megvan a nyitottság a kulturális és technológiai fejlődésre is.

### 5.2. A Data Mesh megtérülése

A Data Mesh bevezetése nemcsak technológiai, hanem stratégiai befektetés is, amelynek megtérülése (*Return on Investment - ROI*) elsősorban a hosszú távú üzleti és szervezeti hatásokban mérhető. Az architektúra kialakítása jelentős kezdeti erőforrásráfordítást igényel

mind emberi, mind technikai oldalon: új szerepkörök (adatgazda, domain-tulajdonos, adattermék-menedzser) létrehozását, governance-folyamatok definiálását, valamint az adatplatform és a metaadat-kezelő eszközök integrálását. Rövid távon a Data Mesh költségei tehát magasabbak lehetnek a hagyományos centralizált adatkezelési modellekhez képest, azonban középtávon a decentralizáció előnyei, mint az adat-hozzáférés felgyorsulása, a felelősségvállalás erősödése, valamint a domain-szintű önállóság fokozatosan ellensúlyozzák a kezdeti ráfordításokat.

A megtérülés kulcstényezői a hatékonyságnövekedés, a költségoptimalizálás és az adatvezérelt döntéshozatal minőségi javulása. Egy érett Data Mesh környezetben a domain csapatok képesek önállóan kezelni és publikálni adattermékeiket, ezáltal csökken az IT-részlegre nehezedő nyomás, valamint az adat-előállítás és -felhasználás közötti átfutási idő.

A kutatásom alapján úgy vélem, hogy a Data Mesh megtérülése elsősorban nem azonnali pénzügyi haszonban, hanem adatérettségi és szervezeti rugalmassági növekedésben realizálható. A modell valódi értéke abban rejlik, hogy skálázható, fenntartható és adaptív adatkezelési keretet biztosít, amely hosszú távon csökkenti a működési kockázatokat és elősegíti az innovációt.

### **5.3. Hipotézisvizsgálat**

Ebben a fejezetben a kutatás elején megfogalmazott hipotéziseket vizsgálom meg részletesebben. A hipotézisek célja, hogy a Data Mesh és architektúrájának egyes alapelveit alternatív szemszögből értékeljem, és ezzel rávilágítsak a modell erősségeire és esetleges gyengeségeire. Az eredmények hozzájárulnak ahhoz, hogy a Data Mesh gyakorlati alkalmazhatóságát ne csak elméleti, hanem empirikus tapasztalatok alapján is meg lehessen ítélni.

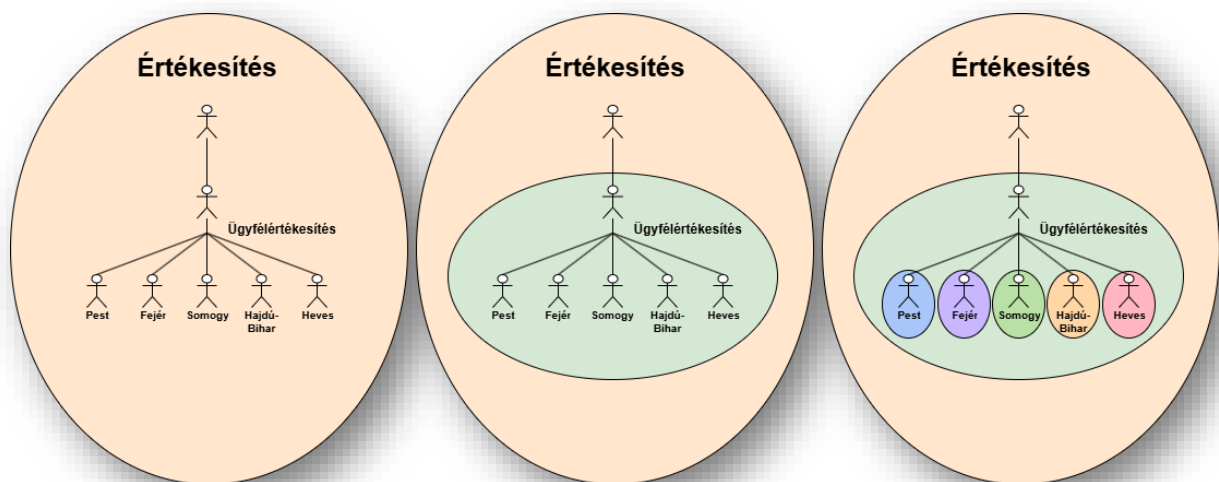
#### **5.3.1. H1 Hipotézis**

*„Mélyebb al-domain struktúrák kialakítása elősegíti a szervezeti felelősségvállalás és elszámoltathatóság erősödését”*

Az esettanulmány során bemutattam az al-domain struktúrák kialakítását, amelyek elsődleges célja, hogy a nagyobb, több alcsapattal rendelkező domainek ne váljanak túlságosan centralizálttá. A centralizáció ugyanis domain-szintű adatproblémákat generálhat, amelyek akadályozzák a Data Mesh alapelveinek érvényesülését. Az al-domain struktúrák létrehozása ezért elősegíti a domain szegmentáció gyakorlati megvalósítását, és bővíti az adatgazdai szerepkörök delegálásának lehetőségét is. Ugyanakkor a domain stratégiák tervezésekor

kritikus kérdés a szegmentáció mélységének meghatározása, vagyis hogy milyen mértékben érdemes hierarchikus bontást alkalmazni. Lényeges szempont, hogy a szülő domain milyen befolyással bír a gyermek-domain működésére és annak további leágazásaira. Ez közvetlen hatással van az adattermékek publikálásának folyamatára is. Bár az al-domain szintű felosztás lehetővé teszi, hogy a csapatok saját adattermékeiket önállóan kezeljék és publikálják, túlzott fragmentáció esetén könnyen kaotikus rendszerezéshez vezethet, különösen akkor, ha a különböző fő-domének eltérő hierarchiai megközelítést alkalmaznak.

Az általam kidolgozott ügyfélértékesítés domain a nagyobb értékesítési divízió részeként működik, azonban sajátosságai miatt önálló entitásként is értelmezhető, ezért indokoltnak tartottam az al-domain megközelítés szerinti bontását. Természetesen ez nem az egyetlen lehetséges domain-stratégia. A 11. ábra három különböző formációt szemléltet, amelyeket az esettanulmány során választhattam volna.



**11. ábra: Lehetséges Domain stratégiák. Forrás: Saját szerkesztés**

A színezett szegmensek az eltérő domain-struktúrákat mutatják. Gyakorlati szempontból a harmadik formáció is alkalmazható lett volna, hiszen az al-egységek szintjén is rendelkeztek adatgazdákkal. Ez egyúttal újabb ellenőrzőpontokat hozott volna létre az adatfolyam és az adatminőség biztosításában. Elméleti szinten azonban ez a megközelítés problémás, mert a felelősségi körök túlságosan szétszóródnak, és a megfelelő adattulajdonosi szerepkör nem tud teljes mértékben érvényesülni. Hasonló kérdés vethető fel az általam alkalmazott második modellnél is, hiszen itt sem találunk dedikált adattulajdonost. Ugyanakkor mivel az értékesítési igazgató közvetlenül felel az alárendelt egységekért, az ő szerepköre biztosítja, hogy a teljes domain és az al-domain szintű adatkezelés átlátható legyen. Emiatt nem láttam szükségét külön

al-domain adattulajdonos kijelölésének, hiszen az igazgató az egységek vezetőivel közvetlen kapcsolatban állva képes az adatkezelési folyamatokat koordinálni.

Másik lehetséges megközelítés egy lánc-szerű adattulajdonosi struktúra lenne, amely szerint egy hierarchia több mint három szintjére bontva mindig az n-1-edik szint vezetője kapná a közvetlen felelősséget a lentebbi szintek adatainak kezeléséért. Ez azonban szervezeti komplikációkhoz vezethet, lassíthatja az adatcserét, és jelentősen megnehezítheti az adattermékek rendszerezését, mivel extra nyilvántartási és erőforrásigényeket ró a szervezetre.

### **5.3.2. H1 Konklúzió**

A kutatásom során szerzett megfigyelések egyértelműen azt mutatják, hogy a kétszintű domain-struktúra bőségesen elegendő a szegmentált adatkezelés megvalósításához a domain egységeken belül. Úgy vélem, hogy azokban a szervezetekben, ahol domain szinten az ágazati elosztások gyakoriak, a bevett jógyakorlatok alapján mindenképpen érdemes két szintre bontott szegmentációt alkalmazni. Ez az elv kifejezetten előnyös a vállalat számára, hiszen egyszerre biztosítja az átláthatóságot, az elszámoltathatóságot és a felelősségi körök világos lehatárolását. A kétszintű bontás lehetővé teszi, hogy a szülő domain megfelelő kontrollt gyakoroljon, miközben az al-domainek önálló felelősséggel bírnak az adatvagyon kezelésében, ami megkönnyíti az adattulajdonosok kijelölését és az adattermékek pontosabb definiálását is.

Ezzel szemben a mélyebb, többszintű struktúrák kialakítása már komoly menedzsmenti hátrányokat vonhat maga után. Az ilyen rendszerek gyakran vezetnek széttagolt és nehezen koordinálható adatkezeléshez, amely hosszabb távon akár nem várt silósodást is eredményezhet. Ez a silósodás végső soron éppen azoknak az alapelveknek mondana ellent, amelyeket a Data Mesh hivatott támogatni: a decentralizált, de együttműködő adatkezelésnek. A gyakorlatban sokkal fenntarthatóbb és üzletileg is előnyösebb megoldásnak tartom a kétszintű domain-struktúrák alkalmazását, amelyek egyensúlyt teremtenek a kontroll és a rugalmasság között. A H1 Hipotézist elutasítottam.

### **5.3.3. H2 Hipotézis**

*„Egy megfelelően bevezetett Data Governance eszköz kulcsszerepet játszik a Data Mesh sikeres alkalmazásában.”*

Kutatásom egyik legfontosabb eszköze a gyakorlatok megvalósításában és szemléltetésében az OpenMetadata Data Governance rendszer volt. Bár az esettanulmány során kapott eredményeim azt mutatják, hogy egy Data Mesh struktúrát ilyen eszköz nélkül szinte lehetetlen hatékonyan működtetni, a legtöbb forrás és maga a Data Mesh elméleti alapja eredetileg nem

számolt ezen eszközök integrációjával a vállalati környezetbe, legfeljebb hasznos kiegészítő szolgáltatásként tekintett rájuk az architektúra kialakítása során. A korábban létrejött Governance eszközök többsége nem kifejezetten a Data Mesh támogatására készült, hanem általános adatrányítási megközelítések összegzéseként került be a szervezeti gyakorlatba. Ezek az eszközök ráadásul a Data Mesh fogalmát is megelőzik időben, így bár napjainkra fejlesztők fokozatosan Data Mesh-kompatibilissé tették őket, nem állítható, hogy alapvetően ennek az architektúrának a logikáját követték volna. Ezzel szemben az újabb generációs megoldások, például az OpenMetadata, már kifejezetten a Data Mesh elvei mentén épültek fel, és igyekeznek szélesebb funkcionalitást lefedni az architektúra gyakorlati megvalósításában. A hipotézis tehát abból a szempontból vizsgálja a Data Mesh alkalmazását, hogy milyen értékeket képesek hozzáadni a Data Governance eszközök, és mennyiben segítik elő az elvek gyakorlati érvényesülését. A megfelelő konklúzióhoz elengedhetetlen megvizsgálni, milyen változásokat eredményezne, ha egy Governance eszköz nem képezné részét az architektúrának.

A klasszifikációk és címkézések már a Governance eszközök megjelenése előtt is használt gyakorlatok voltak, azonban ha rendszerenként, elszigetelten alkalmazzák őket, az könnyen bonyolulttá és átláthatatlanná válik. Egy központosított Governance eszköz hiányában a szegmentált klasszifikációk ellenőrizhetősége és fenntartása szinte lehetetlen. Hasonló problémát jelent a minőség-ellenőrzés is. Bár sok esetben speciális, erre fejlesztett alkalmazások is bevetethők, amelyek gyakran fejlettebb funkcionalitással rendelkeznek, a kisebb volumenű vagy nem üzletkritikus adatköröknél ezek használata nem mindig indokolt. Ilyenkor a Governance eszközök integrált minőségellenőrzési funkciói gazdaságosabb és egyszerűbb megoldást jelentenek. Ugyanez érvényes a hozzáférés-kezelésre is, ahol a rendszerenként széttöredezett szabályozás jelentős kockázatot hordoz. Egy központi eszköz itt is biztosítja az egységes, átlátható és konzisztens működést, ami kulcsfontosságú a Data Mesh alapelveinek érvényesüléséhez. Ezek csak néhány funkcionalitás, amik nem lettek a Data Meshben technikai megvalósítás során definiálva, de a Governance eszközök unifikáltan és rendezetten képesek lefedni az igényeket.

#### **5.3.4. H2 Konklúzió**

A jelenlegi tapasztalatok és az eszközök fejlődési iránya alapján megalapozottnak tűnik az a feltevés, hogy idővel egy Governance eszköz alkalmazása nem csupán előnyös, hanem elvárt feltétele is lehet egy sikeres Data Mesh architektúra kialakításának.

Egy Governance eszköz szerepe nem az egyes funkciókban, hanem az egységesítésben és központosításban rejlik. A Data Mesh működésében ez biztosítja a metaadatok közös

kezelését és böngészését, amelyet a legtöbb rendszer a katalogizálási funkción keresztül nyújt. Sok eszköz azonban egy adott ökoszisztémához kötött, mint például a Databricks Unity Catalog, amely csak a Databricks környezetét támogatja. A Data Mesh elvei miatt ezért kulcsfontosságú, hogy a Governance eszköz képes legyen több, heterogén forrás egységes kezelésére és integrációjára.

A katalogizálás és a sokrétű rendszertámogatás kétségkívül elégséges feltétele annak, hogy a Data Mesh működése során egységesíteni lehessen a különböző domáinek adatait, azonban önmagában ez még nem jelenti egy Governance eszköz legnagyobb hozzáadott értékét. Az igazi előny abban rejlik, hogy az adott eszköz mennyire képes támogatni a kapcsolódó architektúrális és szervezeti elveket. Erre jó példa az OpenMetadata, amely nemcsak széles körű connector-támogatottsággal és erős katalogizációs képességekkel rendelkezik, hanem kiterjesztett funkcionalitásával lehetővé teszi, hogy egy rendszeren belül valósuljon meg a klasszifikáció, a hozzáférés-kezelés, az adattermékek és szerződések publikálása, valamint az adatminőség-ellenőrzés. Ezáltal több, különálló szoftver funkcióit ötvözi egyetlen eszközben, ami üzleti szempontból a Governance csapat munkáját, technikai oldalról a Platform csapat feladatait, szervezeti szinten pedig az Enabling csapat hatékonyságát támogatja. Ennek köszönhetően nincs szükség számos, eltérő alkalmazás párhuzamos bevezetésére és ismertetésére a vállalatban, hanem egy központi eszköz segítségével biztosítható az egységes és átlátható működés. Úgy vélem, hogy minél több Data Mesh architektúrális komponens képes egy Governance eszköz lefedni és támogatni, annál nagyobb szervezeti értékkel bír. Egy ilyen eszköz nem csupán kiegészítő megoldásként funkcionál, hanem idővel a Data Mesh működésének kritikus elemévé válhat, amely nélkülözhetetlen a modell hosszú távú fejlődésében. Az integrált funkcionalitás révén az eszköz képes összekapcsolni a különálló folyamatokat, támogatva ezzel a Data Mesh alapelveinek érvényesülését, és ezzel egyidejűleg biztosítva a vállalat számára az egységes, skálázható és fenntartható adatkezelési környezetet. A H2 Hipotézist elfogadtam.

## 6. TOVÁBBI LEHETŐSÉGEK

A Data Mesh egy viszonylag fiatal adatkezelési modell, amelynek célja a növekvő adatmennyiség és a szervezeti adatérettség hatékony kezelése. Kutatásom eredményei arra mutattak rá, hogy a modell elvei és technikai alapjai kifejezetten hasznos irányelveket adnak, azonban a gyakorlati bevezetés még mindig komoly kihívást jelent. Egy megfelelően implementált Data Mesh hosszú távon jelentős üzleti és technológiai előnyökkel járhat, ugyanakkor kérdés, hogy a költséges és összetett bevezetési folyamat után képes-e tartósan fennmaradni, vagy idővel egyszerűbb architektúrák veszik át a helyét.

Jelenleg a Data Mesh legnagyobb gyengesége az implementáció nehézsége és az ebből fakadó kockázat, hogy a modell valóban segíti-e a vállalat működését, vagy inkább újabb akadályokat teremt. Tapasztalataim szerint nemcsak maga a bevezetés nehézkes, hanem bizonyos szerepkörök, mint az adatgazdák vagy a Governance csapat skálázhatósága is problémát jelenthet egy komplex vállalati struktúrában. Ahogy a H1 hipotézisben is bemutattam, a túlzott felelősség-szegmentálás inkább gyengíti, mint erősíti a Data Mesh értékét. Nem véletlen, hogy több szakértő szkeptikusan tekint a modellre. Sokan túl ambiciózusnak tartják, és kétségbe vonják, hogy hosszú távon fenntartható lehet ekkora mértékű megfeleltetés. Én ugyanakkor úgy vélem, hogy ezek a nehézségek elsősorban az emberi kapacitás határaiból fakadnak. A jövő azonban két irányból is kínál megoldást.

Az egyik a Governance eszközök fejlődése, amelyek egyre szélesebb funkcionalitással képesek lefedni a Data Mesh architektúra igényeit. Ahogy a H2 hipotézisemben is kifejtettem, a folyamatok centralizálása és egységesítése nemcsak üzleti, hanem technikai szinten is jelentős könnyebbséget hoz, és a jövőben akár alapfeltétele is lehet a modell sikeres működésének.

A másik meghatározó tényező az AI beépülése a Data Mesh környezetbe. A jelenlegi AI-megoldások még gyerekcipőben járnak, de már most látszik, hogy képesek lesznek átalakítani a Governance feladatokat: automatizálhatják az adatgazdai teendőket, támogathatják a minőségellenőrzést, előállíthatják a policy-eket vagy kezelhetik az adatszerződéseket. Ha az AI képes átvenni az új Data Mesh szerepkörök egy részét, az radikálisan csökkentheti az emberi erőforrásra nehezedő terhet, és valóban fenntarthatóvá teheti a modellt.

Konkluzívan úgy látom, hogy a Data Mesh jövője két pilléren nyugszik: a Governance eszközök integrációján és az AI-alapú automatizmusokon. Ha ezek a technológiák együtt fejlődnek, a Data Mesh nemcsak életképes, hanem hosszú távon is meghatározó modellé válhat, amely képes az üzleti működés és az adatmenedzsment fenntartható összhangját biztosítani.

## 7. ÖSSZEFOGLALÁS

Kutatásom eredményeivel összességében nagyon elégedett vagyok, mivel a kitűzött célokat nemcsak elértem, hanem több területen is meghaladtam az előzetes várakozásaimat. Egy elsősorban egyszerűnek tűnő kutatási kérdés valójában összetett és tudományos szempontból is jelentős feladattá vált, amelynek megoldása mély szakmai és módszertani kihívásokat tartogatott. A munka során lehetőségem nyílt arra, hogy egy szimulált vállalati környezetben keresztül gyakorlati szinten demonstráljam, miként valósítható meg egy modern szervezeti adatmodell bevezetése. A kutatás folyamán strukturáltan megismertem és alkalmaztam azokat a lépéseket, amelyek egy üzleti probléma sikeres megközelítéséhez és megoldásához vezetnek. Különösen értékes tapasztalat volt számomra, hogy a munkahelyemen szerzett adatmérnöki ismereteket és valós ipari gyakorlatot ötvözhettem az egyetemen tanult üzleti és módszertani tudással. Ennek köszönhetően nem pusztán technológiai megoldásokat alkottam, hanem azokat üzleti kontextusba is sikeresen be tudtam ágyazni. Ez a kettősség, az üzleti és technológiai szemlélet együttes érvényesítése a kutatás egyik legfontosabb értékteremtő eleme volt, és egyben jól szemlélteti, hogy a modern adatmenedzsment több tudományágon átívelő megközelítést igényel.

Meggyőződésem, hogy a kutatás során elért eredmények nemcsak oktatási, hanem vállalati környezetben is hasznosíthatók, és valós értéket képviselnek az adatvezérelt működés felé törekvő szervezetek számára. Bízom benne, hogy a megszerzett tudást a jövőben tanácsadói és stratégiai szinten is kamatoztathatom, hozzájárulva a vállalatok adatérettségének fejlesztéséhez és adatstratégiáik megerősítéséhez. Emellett úgy gondolom, hogy a Data Mesh-en túlmenően az adatkezelés és az adatérettség kérdései a közeljövőben is a digitális transzformáció központi témái maradnak. Remélem, hogy kutatásom nemcsak inspirációul szolgál, hanem gyakorlati iránymutatást is nyújt mindazoknak, akik a vállalati adatmenedzsment fejlesztésén dolgoznak, és hogy az általam bemutatott eredmények valós hozzájárulást jelentenek e gyorsan fejlődő terület szakmai diskurzusához.

## Irodalomjegyzék

Action (2024). What is Data Quality Management?, *Action*.

Forrás: <https://www.action.com/what-is-data-quality-management/>

(Letöltés időpontja: 2025. augusztus 1.).

Alturi, V. és Ferraiolo, D. (2025). Role-Based Access Control, in S. Jajodia, P. Samarati, és M. Yung (szerk.) *Encyclopedia of Cryptography, Security and Privacy*. Cham: Springer Nature Switzerland, o. 2124–2127.

Forrás: [https://doi.org/10.1007/978-3-030-71522-9\\_829](https://doi.org/10.1007/978-3-030-71522-9_829).

(Letöltés időpontja: 2025. augusztus 1.).

Atlan (2024). OpenMetadata vs. DataHub: Compare Architecture, Capabilities, Integrations & More, *Atlan*.

Forrás: <https://atlan.com/openmetadata-vs-datahub/?ref=/gartner-data-catalog/>.

(Letöltés időpontja: 2025. július 8.).

Brook, C. (2023). *Attribute-Based Access Control: Pros, Cons & Use Cases | Fortra's Digital Guardian*. Forrás: <https://www.digitalguardian.com/blog/attribute-based-access-control>

(Letöltés időpontja: 2025. augusztus 4.).

Christ, J. és Harrier, S. (2024). Data Contract Specification.

Forrás: <https://datacontract.com/>.

(Letöltés időpontja: 2025. július 15.).

Christ, J., Visengeriyeva, L. és Harrier, S. (2022). Data Mesh Architecture: Data Mesh From an Engineering Perspective.

Forrás: <https://www.datamesh-architecture.com/>.

(Letöltés időpontja: 2025. július 14.).

Cloudflare (2020). What is vendor lock-in? | Vendor lock-in and cloud computing.

Forrás: <https://www.cloudflare.com/learning/cloud/what-is-vendor-lock-in/>.

(Letöltés időpontja: 2025. július 26.).

Coursera (2025). What's the Difference Between AWS vs. Azure vs. Google Cloud?

Forrás: <https://www.coursera.org/articles/aws-vs-azure-vs-google-cloud>.

(Letöltés időpontja: 2025. július 23.).

DataGalaxy (2024). Data governance roles for data mesh environments, 13 február.

Forrás: <https://www.datagalaxy.com/en/blog/data-governance-roles-in-data-mesh/>

(Letöltés időpontja: 2025. szeptember 5.).

Dávid, Á. (2023). Mi az API jelentése? Miért hasznosak az API-k?

Forrás: <https://kiszervezettmarketing.hu/weboldal-keszites/api-jelentese/>.

(Letöltés időpontja: 2025. július 26.).

Deghani, Z. (2022). *Data Mesh: Delivering Data-Driven Value at Scale*. O'Reilly Media, Sebastopol, ISBN: 978-1-4920-9239-1

Európai Unió (2016a). *Az (EU) 2016/679 rendelet 4. cikke (általános adatvédelmi rendelet - GDPR)*.

Forrás: <https://www.privacy-regulation.eu/hu/4.htm>  
(Letöltés időpontja: 2025. október 7.).

Európai Unió (2016b). *Az (EU) 2016/679 rendelet 83. cikke (általános adatvédelmi rendelet - GDPR).*

Forrás: <https://www.privacy-regulation.eu/hu/83.htm>  
(Letöltés időpontja: 2025. október 7.).

Gupta, L. (2025). What is REST?

Forrás: <https://restfulapi.net/>.  
(Letöltés időpontja: 2025. július 26.).

Harmeling, T. (2025). PII vs. PI vs. Sensitive Data: Important Key Differences, *Consent Management Platform (CMP) Usercentrics*.

Forrás: <https://usercentrics.com/knowledge-hub/personally-identifiable-information-vs-personal-data/>  
(Letöltés időpontja: 2025. szeptember 16.).

Harrier, S. és Christ, J. (2022). Data Mesh Governance by Example.

Forrás: <https://www.datamesh-governance.com/>.  
(Letöltés időpontja: 2025. július 30.).

Hyperight (2022). If Data Mesh Is Not Architecture, What Is It?

Forrás: [https://hyperight.com/if-data-mesh-is-not-architecture-what-is-it/?utm\\_source=chatgpt.com](https://hyperight.com/if-data-mesh-is-not-architecture-what-is-it/?utm_source=chatgpt.com).  
(Letöltés időpontja: 2025. július 14.).

IBM (2022). *What is Personally Identifiable Information (PII)? | IBM.*

Forrás: <https://www.ibm.com/think/topics/pii>  
(Letöltés időpontja: 2025. szeptember 16.).

Kimberling, E. (2023). Are ERP Systems a Good Fit for Small and Mid-Size Businesses?

Forrás: <https://www.thirdstage-consulting.com/are-erp-systems-a-good-fit-for-small-and-mid-size-businesses/>.  
(Letöltés időpontja: 2025. július 26.).

Kinza, Y. (2024). Definition: Data Mesh.

Forrás: [https://www.techtarget.com/searchdatamanagement/definition/data-mesh?utm\\_source=chatgpt.com](https://www.techtarget.com/searchdatamanagement/definition/data-mesh?utm_source=chatgpt.com).  
(Letöltés időpontja: 2025. július 14.).

L. Manor, G. (2024). *What Is Attribute-Based Access Control (ABAC)?*

Forrás: <https://www.permit.io/blog/what-is-abac>  
(Letöltés időpontja: 2025. augusztus 1.).

Meganatha, S. (2025). Distributed Stewardship for Data Mesh Architecture, *Immuta*.

Forrás: <https://www.immuta.com/blog/distributed-stewardship-data-mesh-architecture/>  
(Letöltés időpontja: 2025. szeptember 5.).

Meyer, M. (2025). Federated Data Governance Explained: A Framework for Scalable Control.

Forrás: <https://www.alation.com/blog/federated-data-governance-explained/>.  
(Letöltés időpontja: 2025. július 15.).

Microsoft (2025). Streamfeldolgozás az Apache Kafka és az Azure Databricks használatával.  
Forrás: <https://learn.microsoft.com/hu-hu/azure/databricks/connect/streaming/kafka>.  
(Letöltés időpontja: 2025. július 24.).

Orbán A. (2022). Természetes Személyazonosító Adat - Közzolgálati Online Lexikon, 30 november. Forrás: <https://lexikon.uni-nke.hu/szocikk/termeszetes-szemelyazonosito-adat/>  
(Letöltés időpontja: 2025. szeptember 15.).

Owczarek, D. (2023). *The Impact of Data Mesh Architecture on Data Teams*, nexocode.  
Forrás: <https://nexocode.com/blog/posts/data-mesh-and-data-teams/>  
(Letöltés időpontja: 2025. augusztus 5.).

Popa, M. (2023). *Policy-Based Access Control 101: What Is PBAC? Definition and Benefits*, Heimdal Security Blog.  
Forrás: <https://heimdalsecurity.com/blog/policy-based-access-control/>  
(Letöltés időpontja: 2025. augusztus 5.).

R. China, C. (2025). *What Is Software as a Service (SaaS)?* / IBM.  
Forrás: <https://www.ibm.com/think/topics/saas>  
(Letöltés időpontja: 2025. október 7.).

Redpanda (2024). Batch vs. streaming data processing.  
Forrás: <https://www.redpanda.com/blog/batch-vs-streaming-data-processing>.  
(Letöltés időpontja: 2025. július 24.).

Sala, F. (2024). Self-serve Data Platform.  
Forrás: <https://platform.opendatamesh.org/concepts/self-serve-data-platform/#>.  
(Letöltés időpontja: 2025. július 16.).

Saux, L. (2024). What is metadata and why is it as important as the data itself?  
Forrás: <https://www.opendatasoft.com/en/blog/what-is-metadata-and-why-is-it-important-data/>.  
(Letöltés időpontja: 2025. július 15.).

Shakarzy, A. (2024). *The Challenges of Managing Permissions with Role-Based Access Control (RBAC)*. Forrás: <https://crossidio-avgtjal3m-crossids-projects.vercel.app/academy/role-based-access-control-rbac-challenges>  
(Letöltés időpontja: 2025. augusztus 4.).

Susnjara, S. és Smalley, I. (2025). What is cloud computing?  
Forrás: <https://www.ibm.com/think/topics/cloud-computing>.  
(Letöltés időpontja: 2025. július 23.).

ThoughtWorks (2022). ThoughtWorks Technology Radar Vol. 27 October, 2022, 27, o. 26.

Tsai, P. (2021). Data Governance 101: How to Build a Business Glossary, *BigID*, 3 november.  
Forrás: <https://bigid.com/blog/what-is-a-business-glossary/>  
(Letöltés időpontja: 2025. július 31.).

Wider, A., Verma, S. és Akhtar, A. (2023). Decentralized Data Governance as Part of a Data Mesh Platform: Concepts and Approaches, in *2023 IEEE International Conference on Web Services (ICWS)*. *2023 IEEE International Conference on Web Services (ICWS)*, Chicago, IL, USA: IEEE, o. 746–754. Forrás: <https://doi.org/10.1109/ICWS60048.2023.00101>.  
(Letöltés időpontja: 2025. július 30.).

Yasar, K., Bigelow, S.J. és Sullivan, E. (2024). What is pay-as-you-go cloud computing (PAYG cloud computing)?  
Forrás: <https://www.techtarget.com/searchstorage/definition/pay-as-you-go-cloud-computing-PAYG-cloud-computing>.  
(Letöltés időpontja: 2025. július 24.).