



Vaasan yliopisto
UNIVERSITY OF VAASA

Väinö van der Meer

Datan käytettävyys ja yhteentoimivuus data mesh - arkkitehtuurissa

Kirjallisuuskatsaus data meshin vaikutuksista datan käytettävyyteen ja yhteentoimivuuteen sekä näitä edistävästä ja estävästä tekijöistä organisaatioissa

Tekniikan ja innovaatiojohtamisen akateeminen yksikkö
Data-arkkitehtuurin kandidaatintutkielma
Kestävä energia ja älykäs teknologia

Vaasa 2026

VAASAN YLIOPISTO**Tekniikan ja innovaatiojohtamisen akateeminen yksikkö**

Tekijä:	Väinö van der Meer		
Tutkielman nimi:	Datan käytettävyys ja yhteentoimivuus data mesh -arkkitehtuurissa: Kirjallisuuskatsaus data meshin vaikutuksista datan käytettävyyteen ja yhteentoimivuuteen sekä näitä edistävästä ja estävästä tekijöistä organisaatioissa		
Tutkinto:	Tekniikan kandidaatti, Energia- ja informaatiotekniikan ohjelma		
Oppiaine:	Data-arkkitehtuuri		
Työn ohjaaja:	Maarit Välisuo		
Valmistumisvuosi:	2026	Sivumäärä:	84

TIIVISTELMÄ:

Tutkielmassa tarkastellaan data mesh -arkkitehtuurin vaikutuksia datan käytettävyyteen ja yhteentoimivuuteen. Lähtökohtana on havainto, että perinteiset keskitetyt data-arkkitehtuurit voivat aiheuttaa organisatorisia ja teknisiä pullonkauloja ja heikentää datan hyödyntämistä organisaatioissa. Data mesh esitetään hajautettuna lähestymistapana, jossa datan omistajuus siirtyy liiketoimintadomaineille ja dataa hallitaan datatuotteina.

Tutkimuksen tavoitteena on selvittää, miten domain-lähtöinen malli vaikuttaa datan käytettävyyteen ja yhteentoimivuuteen sekä mitkä tekijät edistävät tai estävät niiden toteutumista data mesh -ympäristössä. Lisäksi tarkastellaan käytännön ratkaisuja ja arvioidaan, millaisiin organisaatioihin data mesh soveltuu. Tutkimus toteutetaan kirjallisuuskatsauksena, joka perustuu vertaisarvioituihin tutkimuksiin ja teollisuuslähteisiin.

Tulokset osoittavat, että data mesh voi parantaa datan käytettävyyttä erityisesti silloin, kun datatuotteet toteuttavat keskeisiä käytettävyyssominaisuuksia, kuten löydettävyyttä, ymmärrettävyyttä, osoitettavuutta, natiivia saavutettavuutta ja riittävää laatua. Käytettävyyttä tukevat itsepalvelualustat, standardoitu metadata, datasopimukset sekä turvallinen ja yhdenmukainen pääsynhallinta. Yhteentoimivuutta tarkastellaan teknisestä, semanttisesta, organisatorisesta ja hallinnollisesta näkökulmasta, joiden perusteella standardoidut rajapinnat, yhteiset skeemat, metadata ja federoitu hallinta voivat tukea datan yhdistämistä ja uudelleenkäyttöä yli domain-rajojen. Tulokset osoittavat myös, että ilman yhteisiä standardeja, riittävää hallintaa ja semanttista yhdenmukaisuutta hajautettu malli voi johtaa metadatan laadun vaihteluun, semanttiseen hajautumiseen ja heikentyneeseen yhteentoimivuuteen.

Tulosten perusteella data mesh soveltuu parhaiten organisaatioihin, joissa datan määrä ja kompleksisuus ovat suuria, domain-rakenne on selkeä sekä datan hallintakäytännöt, vastuunjako ja tekninen kyvykyys ovat riittävän kypsiä. Vastaavasti organisaatioissa, joissa nämä edellytykset eivät täyty, hajautettu malli voi heikentää datan käytettävyyttä sekä yhteentoimivuutta. Johtopäätöksenä data mesh ei automaattisesti paranna datan käytettävyyttä tai yhteentoimivuutta.

AVAINSANAT: data mesh -arkkitehtuuri, datan käytettävyys, datan yhteentoimivuus, domain-lähtöinen omistajuus, datatuotteet, federoitu hallinta, socio-tekninen muutos

Sisällys

1	Johdanto	7
1.1	Tausta	7
1.2	Tutkimusaukko	7
1.3	Tutkimuskysymykset	8
1.4	Tutkimusmenetelmä	9
1.5	Rajaukset	9
1.6	Keskeiset käsitteet	9
1.6.1	Data-arkkitehtuuri	9
1.6.2	Keskeiset data-arkkitehtuurimallit	10
1.6.3	Datan käytettävyys	11
1.6.4	Datan yhteentoimivuus	11
1.6.5	Käytettävyyden ja yhteentoimivuuden ero	11
1.6.6	Muut käsitteet	11
2	Data mesh -arkkitehtuuri ja sen keskeiset periaatteet	13
2.1	Data mesh -arkkitehtuurin neljä peruseriaatetta	14
2.1.1	Domain-lähtöinen omistajuus	14
2.1.2	Tuotteen ajattelutapa datalle	15
2.1.3	Itsepalveludata-alusta	19
2.1.4	Hajautettu laskennallinen hallinto	20
2.2	Organisaation muutos ja vastuunjako data mesh -arkkitehtuurissa	21
2.3	Datan elinkaari data mesh -arkkitehtuurissa	24
2.3.1	Datan kerääminen	25
2.3.2	Datan analysointi	26
2.3.3	Datan hallinnointi ja jakaminen	26
2.4	Data mesh -arkkitehtuurin soveltuvuuden yleiset edellytykset	27
3	Datan käytettävyys data mesh -arkkitehtuurissa	30
3.1	Datan käytettävyyden osa-alueet	32
3.1.1	Löydettävyys ja osoitettavuus	32

3.1.2	Ymmärrettävyys ja itsekuvailevuus	34
3.1.3	Laatu, luotettavuus ja eheys	35
3.1.4	Datatuote ja sen itsenäinen arvo	38
3.1.5	Natiivi saavutettavuus ja tekninen käytettävyys	39
3.1.6	Turvallisuus ja yhteentoimivuus käytettävyyden ehtoina	40
3.2	Datan käytettävyyden käyttäjä- ja sidosryhmälähtöinen tarkastelu	40
3.2.1	Data-analyttikot	42
3.2.2	Data scientistit	42
3.2.3	Kehittäjät	43
3.2.4	Domain-omistajat	44
3.2.5	Alustatiimi	45
3.2.6	Liiketoimintakäyttäjät ja päätöksentekijät	45
3.2.7	Hallinto- ja tietoturvatyöntekijät	46
3.3	Tietoturva ja kyberturvallisuus käytettävyyden näkökulmasta	46
3.3.1	Tietoturva osana datan käytettävyyttä	47
3.3.2	Hajautettu tietoturva ja hallintamallit	47
3.3.3	Kyberturvallisuusriskit ja käytettävyyden välinen tasapaino	49
3.4	Datan käytettävyyden edistävät ja estävät tekijät	50
3.4.1	Edistävät tekijät	50
3.4.2	Estävät tekijät	53
3.4.3	Tekoäly käytettävyyden tekijänä	55
3.5	Datan käytettävyyden arviointi mittareiden avulla	55
4	Datan yhteentoimivuus data mesh –arkkitehtuurissa	59
4.1	Yhteentoimivuuden tasot data mesh -arkkitehtuurissa	60
4.1.1	Semanttinen yhteentoimivuus	61
4.1.2	Organisatorinen yhteentoimivuus	61
4.1.3	Tekninen yhteentoimivuus	62
4.1.4	Hallinnollinen yhteentoimivuus	63
4.2	Yhteentoimivuutta edistävät tekijät ja ratkaisut data mesh -arkkitehtuurissa	63
4.3	Yhteentoimivuuden haasteet domainien välillä	67

4.4	Empiiriset havainnot ja tapausesimerkit data mesh -toteutuksissa	68
4.4.1	Onnistuneet toteutukset ja havaitut hyödyt	69
4.4.2	Epäonnistuneet toteutukset, tulosten tulkinta ja rajoitteet	71
4.5	Datan yhteentoimivuuden mittarit	72
5	Johtopäätökset	75
5.1	Vastaukset tutkimuskysymyksiin	75
5.2	Tutkimustulosten arviointi	78
5.3	Tutkimustulosten tulkinnan rajoitukset	79
5.4	Ehdotukset jatkotutkimukselle	79
	Lähteet	80

Kuvat

Kuva 1. Data mesh -arkkitehtuurin vastuunjako ja hallinta.	23
---	----

Taulukot

Taulukko 1. Muut käsitteet.	11
Taulukko 2. Vastuunjako perinteisessä data-arkkitehtuurissa ja data mesh -mallissa.	22
Taulukko 3. Datan käytettävyyden määrittelyn näkökulmat ja arviointitavat.	31
Taulukko 4. Datan käytettävyyden mittareita data mesh -ympäristössä.	55
Taulukko 5. Datan yhteentoimivuuden näkökulmat, viitekehukset ja arviointitavat.	59
Taulukko 6. Datan yhteentoimivuuden mittarit.	72

Algoritmit

Algoritmi 1. Esimerkkimuotoinen datasopimus.	17
Algoritmi 2. Esimerkkimuotoinen datatuotteen metadata- ja katalogimäärittely.	18

Lyhenteet

API	Application Programming Interface
BI	Business Intelligence
BLOB	Binary Large Object
EIF	European Interoperability Framework
ETL	Extract, Transform, Load
IaC	Infrastructure as Code
IoT	Internet of Things
IT	Information Technology
MCP	Model Context Protocol
SLA	Service Level Agreement
SLO	Service Level Objective
SQL	Structured Query Language
UX	User Experience

1 Johdanto

1.1 Tausta

Datan määrä ja monimuotoisuus ovat kasvaneet merkittävästi (Bode ja muut, 2024, s. 1; Goedegebuure ja muut, 2024, 11:2). Serra (2024, s. 3) liittää kasvun esimerkiksi IoT-laitteiden, sosiaalisen median, pilvipalveluiden ja muiden digitaalisten järjestelmien yleistymiseen. Kasvava datamäärä on samalla lisännyt organisaatioiden tarvetta kehittää skaalautuvampia data-arkkitehtuuria (SAP, 2026; Bodapati, 2025, s. 770). Perinteiset keskitetyt ratkaisut, kuten data warehouse ja data lake, voivat aiheuttaa organisatorisia pulonkaloja ja pitkiä vasteaikoja datapyynnöille (Dehghani, 2022, s. 112; Bode ja muut, 2024, s. 1). Dehghani (2022, s. 130) korostaa lisäksi, että keskitetyt datatiimit voivat eriytyä liiketoimintadomaineista, mikä heikentää liiketoimintakontekstin ymmärrystä. Dehghani (2022, s. 27, 33, 107) sekä Goedegebuure ja muut (2024, 11:7) kuvaavat data mesh -arkkitehtuurin vastauksena näihin ongelmiin. Mallissa datan omistajuus hajautetaan liiketoimintadomaineille, ja sitä tuetaan federoidulla hallintamallilla.

Tässä työssä datan käytettävyyttä tarkastellaan luvussa 3 ja datan yhteentoimivuutta luvussa 4, jotta molempia ilmiöitä voidaan analysoida selkeästi omista näkökulmistaan. Tätä ennen luvussa 1 määritellään keskeiset käsitteet. Luvussa 2 käsitellään data mesh -arkkitehtuurin keskeiset periaatteet, organisaation vastuunjaon muutos, datan elinkaari sekä arkkitehtuurin soveltuvuus organisaatioille, jotka muodostavat perustan myöhemmälle analyysille.

1.2 Tutkimusaukko

Goedegebuure ja muut (2024, 11:2–11:3) kertovat, että vaikka data mesh on herättänyt kiinnostusta teollisuudessa, akateemisia yrityksiä määritellä tai kehittää konseptia on ollut vähän. Tätä täydentävät Bode ja muut (2024, s. 1–2) sekä Wider ja muut (2023, s. 746), joiden mukaan data mesh -konseptin vaikutuksista, toteutusstrategioista ja käytännön haasteista on edelleen rajallisesti tutkimustietoa.

Havainnot viittaavat siihen, että data mesh voi parantaa datan saavutettavuutta, nopeuttaa tiedon hyödyntämistä sekä lisätä datan laatua ja luotettavuutta hajautetun vastuunjaon kautta (Bode ja muut, 2024, s. 14–15).

Toisaalta tutkimustulokset ovat kuitenkin osittain kaksijakoisia. Bodapati (2025, s. 771) toteaa organisaatioiden kamppailevan keskitetyn hallinnon ja domain-kohtaisen autonomian tasapainottamisen kanssa, minkä lisäksi datatuotteiden standardien määrittely yli organisaatorajojen koetaan haastavaksi. Dolhopolov ja muut (2024, s. 6) puolestaan toteavat, että hallitsematon hajauttaminen voi johtaa yhteensopimattomiin datatuotteisiin ja noudattamatta jääviin säädöksiin. Näitä havaintoja tukevat myös Bode ja muut (2024, s. 10–11), joiden mukaan käytännön toteutuksissa metadatan heikko laatu ja vastuun siirtäminen domaineille ovat aiheuttaneet haasteita datan tulkinnalle ja datatuotteiden hallinnalle. Nämä tekijät voivat heikentää myös datan käytettävyyttä hajautetussa ympäristössä.

1.3 Tutkimuskysymykset

- (1) Miten domain-lähtöinen data mesh -arkkitehtuuri vaikuttaa datan käytettävyyteen?
- (2) Miten domain-lähtöinen data mesh -arkkitehtuuri vaikuttaa datan yhteentoimivuuteen?
- (3) Mitkä tekijät edistävät ja estävät datan käytettävyyttä ja yhteentoimivuutta?
- (4) Millaisissa organisaatioissa data mesh tukee parhaiten datan käytettävyyttä ja yhteentoimivuutta?

1.4 Tutkimusmenetelmä

Tutkimus perustuu kirjallisuuskatsaukseen, eikä se sisällä omaa empiiristä aineistoa tai kvantitatiivista analyysiä. Johtopäätökset muodostetaan aiemman tutkimuksen sekä ta-paus-, teollisuus- ja asiantuntijalähteiden perusteella.

1.5 Rajaukset

Tutkimus rajautuu tarkastelemaan data mesh -arkkitehtuuria erityisesti datan käytettävyyden ja yhteentoimivuuden näkökulmista. Työ ei pyri kattamaan kaikkia data meshiin liittyviä organisatorisia, teknisiä tai liiketoiminnallisia ulottuvuuksia. Työssä kuvataan mesh -arkkitehtuurin keskeiset periaatteet ja rakenne yleisellä tasolla, jotta datan käytettävyyttä ja yhteentoimivuutta voidaan tarkastella kontekstissaan.

Työssä ei käsitellä kaikkien yksittäisten teknologioiden tai työkalujen yksityiskohtaista implementointia. Lisäksi tutkimus keskittyy organisaation sisäiseen data mesh -arkkitehtuuriin eikä tarkastele syvällisesti organisaatioiden välistä datan jakamista.

Data mesh -arkkitehtuurin soveltuvuutta eri organisaatioihin tarkastellaan työssä rajautusti, enimmäkseen datan käytettävyyden ja yhteentoimivuuden näkökulmasta.

1.6 Keskeiset käsitteet

1.6.1 Data-arkkitehtuuri

Data-arkkitehtuuri on datan hallinnan kokonaissuunnitelma, joka vastaa luonteeltaan rakennusarkkitehdin piirustusta (Tupper, 2011, s. 11, 25). Se määrittelee, miten organisaation tietovarannot on järjestetty, miten ne liittyvät toisiinsa ja miten niitä hyödynnetään (Tupper, 2011, s. 42, 58).

Arkkitehtuuri toimii tulkinnallisena kerroksena, joka jäsentää dataa ja tukee informaation hyödyntämistä organisaation toiminnassa (Tupper, 2011, s. 12, 23). Samalla data-arkkitehtuuri tarjoaa rakenteellisen kehyksen dataympäristön hallintaan ja tukee datan tarkoituksenmukaista hyödyntämistä (Dehghani, 2022, s. 16, 44; Tupper, 2011, s. 12).

1.6.2 Keskeiset data-arkkitehtuurimallit

Data-arkkitehtuurit ovat kehittyneet ratkaisemaan edeltäjiensä rajoitteita. Data warehouse keskittyy rakenteelliseen dataan, kun taas data lake mahdollistaa suurten raaka-data-aineistojen tallentamisen. Data lakehouse yhdistää näiden molempien ominaisuuksia (Dolhopolov ja muut, 2024, s. 2; Bode ja muut, 2024, s. 3–4). Käytännössä organisaatiot hyödyntävät usein useita ratkaisuja samanaikaisesti, jolloin data lakehouse toimii teknisenä perustana ja data mesh sen päälle rakentuvana hallintomallina (Dolhopolov ja muut, 2024, s. 2; Serra, 2024, s. 182).

Data mesh voidaan nähdä arkkitehtuurina, operatiivisena mallina tai periaatteiden koelmana, mutta Dehghani (2022, s. 6) määrittelee sen sosiotekniseksi paradigmaksi. Sitä kuvataan myös jatkumona aiemmille malleille, kuten data lakehouse -ratkaisuille (Dolhopolov ja muut, 2024, s. 2; Bode ja muut, 2024, s. 16–17).

Blohm ja muut (2024, s. 646) sekä Bode ja muut (2024, s. 3) kuvaavat data fabricin teknologiavetoisena arkkitehtuurina, joka hyödyntää metadataa, integraatiopalveluita ja dataputkia hajautetuista lähteistä peräisin olevan datan yhdistämiseen ja hallintaan. Data mesh ja data fabric eivät ole toisiaan poissulkevia, vaan täydentävät toisiaan käytännössä (Serra, 2024, s. 182; Bode ja muut, 2024, s. 4). James Serran (2024, s. 183) mukaan ne täydentävät toisiaan.

Serra (2024, s. 202–203) korostaa lisäksi, että organisaatiot hyödyntävät käytännössä usein useiden data-arkkitehtuurien yhdistelmiä sen sijaan, että ne perustuisivat vain yhteen malliin.

1.6.3 Datan käytettävyys

Datan käytettävyys tarkoittaa sitä, kuinka hyvin data tukee käyttäjän tavoitteita ja käyttötarkoitusta. Käytettävyyttä voidaan tarkastella eri näkökulmista, kuten data mesh -ominaisuuksien, teknisen laatu näkökulman, käyttäjä- ja UX-näkökulman sekä sosioteknisen näkökulman kautta. Näitä näkökulmia käsitellään tarkemmin luvussa 3.

1.6.4 Datan yhteentoimivuus

Datan yhteentoimivuus tarkoittaa datan kykyä siirtyä, yhdistyä ja säilyttää merkityksensä eri järjestelmien välillä. Yhteentoimivuutta voidaan tarkastella muun muassa teknisestä, semanttisesta, organisatorisesta ja hallinnollisesta näkökulmasta sekä data mesh -arkkitehtuurin viitekehyyksessä. Näitä näkökulmia käsitellään tarkemmin luvussa 4.

1.6.5 Käytettävyyden ja yhteentoimivuuden ero

Käytettävyys ja yhteentoimivuus ovat toisiaan täydentäviä mutta erillisiä käsitteitä. Data voi olla käytettävää mutta ei yhteentoimivaa, jos sitä ei voida jakaa tai yhdistää muihin järjestelmiin. Vastaavasti data voi olla yhteentoimivaa mutta ei käytettävää, jos sen sisältö on heikkolaatuista tai vaikeasti tulkittavaa.

1.6.6 Muut käsitteet

Taulukossa 1 määritellään tärkeimmät käsitteet tutkimuksessa. Määritelmät on koottu ja sovellettu useista tässä työssä käytetyistä tutkimus- ja käytännön lähteistä.

Taulukko 1. Muut käsitteet.

Käsite	Määritelmä
Alustatiimi (Platform team)	Tiimi, joka kehittää ja ylläpitää yhteistä data-alustaa ja sen itsepalvelukyvykkyyksiä
Continuous Integration / Continuous Delivery (CI/CD)	Tarkoittaa ohjelmistokehityksen toimintamallia, jossa muutokset integroidaan, testataan ja julkaistaan automatisoidusti jatkuvana prosessina. Data mesh -ympäristössä sitä

	voidaan hyödyntää esimerkiksi datatuotteiden, datasopimusten ja datan laadun automaattiseen validointiin ja julkaisuun.
Datasopimus (Data contract)	Muodollinen sopimus, joka sisältää muun muassa SLO:t, datan skeeman, päivitystiheyden ja sallitut käyttötarkoitukset
Datan alkuperä (Data lineage)	Kuvaus datan historiallisesta matkasta ja muunnoksista, joka tarjotaan rajapintojen kautta.
Data-alusta (Data platform)	Infrastruktuuri, joka tarjoaa työkalut datan hakuun, muuntamiseen, käsittelyyn ja tarjoiluun jokaiselle domainille
Datakatalogi (Data catalog)	Rekisteri, jota käytetään datatuotteiden metadatan tallentamiseen ja niiden löydettävyyden mahdollistamiseen
Datan seurattavuus (Data observability)	Kyky seurata datan jakamiseen liittyviä lokitietoja, alkuperää ja mittareita rajapintojen kautta
Dataputki (Data pipeline)	Dataan kapseloitu sisäinen logiikka, joka vastaa datan puhdistamisesta ja rikastamisesta ennen sen tarjoamista
Datatuote (Data product)	Arkkitehtuurinen kvantti (pienin itsenäinen yksikkö), joka sisältää koodin, datan, metadatan ja tarvittavan infrastruktuurin.
Federoitu hallinta (Federated governance)	Keskitettyjen sääntöjen ja hajautetun toteutuksen yhdistelmä
Infrastructure as Code (IaC)	Lähestymistapa, jossa infrastruktuuri määritellään koodin avulla datatuotteiden tarjoamisen automatisoimiseksi
Itsepalvelualusta (Self-service platform)	Alusta, joka tarjoaa palveluita datatuotteiden elinkaaren hallintaan ilman keskitetyn tiimin väliintuloa
Keskitetty IT	Malli, jossa yksi tiimi vastaa datan toimituksesta, mikä aiheuttaa usein pullonkauloja toimitusajoissa
Liiketoiminta-alue (Domain)	Organisaatiokohtainen raja, joka määrittelee tiimin vastualueen ja osaamisen tiettyyn dataan
Metadata	Tietoa datan rakenteesta, merkityksestä ja omistajuudesta
Ontologia (Ontology)	Muodollinen malli käsitteistä, niiden merkityksistä ja suhteista, jonka avulla eri järjestelmät voivat tulkita dataa yhdenmukaisesti
Policy as Code	Lähestymistapa, jossa hallinta-, tietoturva- ja käyttöoikeussäännöt määritellään koneellisesti tulkittavana koodina automaattisen valvonnan ja täytäntöönpanon mahdollistamiseksi
Semantiikka	Semantiikka viittaa datan merkitykseen ja tulkintaan.
Semanttinen yhteentoimivuus	Menetelmät ja dokumentaatio, joilla varmistetaan, että muut voivat yhdistää ja käyttää saatavilla olevaa tietoa
SLA (Service Level Agreement) ja SLO (Service Level Objective)	Mitattavia tavoitteita (kuten päivitysväli ja täydellisyys), jotka poistavat dataan liittyvää epävarmuutta

2 Data mesh -arkkitehtuuri ja sen keskeiset periaatteet

Tässä luvussa tarkastellaan data mesh -arkkitehtuurin teoreettista perustaa, sen neljää ydinperiaatetta, organisaation vastuunjaon muutosta sekä datan elinkaaren toteutusta hajautetussa ympäristössä. Luku muodostaa kokonaiskuvan data mesh -toimintamallista ja luo perustan luvuille 3 ja 4. Luvun lopussa tarkastellaan edellytyksiä sille, millaisille organisaatioille malli soveltuu.

Data mesh on hajautettu data-arkkitehtuuri, jonka tavoitteena on vähentää keskitettyihin data-alustoihin liittyviä koordinointi-, omistajuus- ja ketteryysoongelmia jakamalla datan omistajuus liiketoimintadomaineille (Dehghani, 2022, s. 4, 105; Goedegebuure ja muut, 2024, 11:2). Perinteisissä keskistetyissä ratkaisuissa, kuten data warehouse- ja data lake -malleissa, data on tyypillisesti IT-tiimin hallinnassa (Goedegebuure ja muut, 2024, 11:15). Keskitetty toimintamalli voi johtaa organisatorisiin pullonkauloihin ja heikentää ketteryyttä tilanteissa, joissa sama tiimi vastaa kasvavasta määrästä datalähteitä ja käyttötapauksia (Dehghani, 2022, s. 112; Bode ja muut, 2024, s. 2; Goedegebuure ja muut, 2024, 11:15). Goedegebuure ja muut (2024, 11:15) korostavat, että datan tarjoajien ja hyödyntäjien asiantuntemuksen eriytyminen voi vaikeuttaa liiketoimintakontekstin huomioimista datan hallinnassa. Vastaavasti SAP (2026) liittyy keskitettyihin toimintamalleihin myös omistajuusongelmia ja heikkoa domain-ymmärrystä. Serra (2024, s. 171) kuitenkin huomauttaa, että hajauttamista perustellaan usein teknisellä skaalautuvuudella, vaikka perinteiset keskitetyt mallit kykenevät hallitsemaan myös hyvin suuria datamääriä. Tämän vuoksi data mesh -hankkeiden haasteet liittyvät usein enemmän prosesseihin, hallintaan ja organisaation toimintatapoihin kuin pelkästään arkkitehtuurin teknisiin rajoitteisiin.

Vaikka data mesh -konseptin keskeiset periaatteet esiintyvät useissa toteutuksissa, teollisuudessa ei ole vielä saavutettu yhtenäistä teknistä tai teknologista konsensusta siitä, miten arkkitehtuuri tulisi konkreettisesti toteuttaa (Araújo Machado ja muut, 2022, s. 270).

2.1 Data mesh -arkkitehtuurin neljä peruseriaatetta

Dehghani (2022, s. 15, 29, 47, 67) mukaan data mesh -arkkitehtuurin peruseriaatteet ovat:

- (1) Domain-lähtöinen omistajuus
- (2) Data tuotteena
- (3) Itsepalveludata-alusta
- (4) Hajautettu laskennallinen hallinto

2.1.1 Domain-lähtöinen omistajuus

Domain-lähtöinen omistajuus tarkoittaa, että dataa hallitaan liiketoimialueen mukaan, eikä keskitetysti IT-organisaatiossa. Goedegebuure ja muut (2024, 11:7–11:10) sekä Dehghani (2022, s. 4, 15–16) kuvaavat mallia, jossa kukin domain vastaa oman datansa tuotannosta, laadusta, ylläpidosta ja jakamisesta.

Goedegebuure ja muut (2024, 11:10) sekä Dehghani (2022, s. 66) korostavat, että domain-tiimeillä on syvälinen ymmärrys oman liiketoiminta-alueensa datasta, sen käyttötarkoituksista sekä datan laadun ylläpitämiseen liittyvistä prosesseista. Dehghani (2022, s. 16, 27) liittyy datan jakamisvastuun niille domaineille, jotka ovat lähimpänä datan alkuperää ja sen käyttöä. Tämän vuoksi domain-lähtöinen omistajuus voi parantaa datan laatua, merkityksellisyyttä ja ymmärrettävyyttä. Lisäksi tietomallit voidaan suunnitella domainin tarpeiden mukaan (Dehghani, 2022, s. 115; Goedegebuure ja muut, 2024, 11:12 SAP, 2026).

Toisaalta hajautettu omistajuus voi heikentää käytettävyyttä, sillä se lisää riskiä epäyhteisille tietomalleille, päällekkäiselle datan tuotannolle sekä ristiriitaisille liiketoimintamääritelmille (Bode ja muut, 2024, s. 10–11; Dolhopolov ja muut, 2024, s. 6). Tämä tekee hallinnasta ja standardoinnista kriittisen osan data mesh -mallia. Bode ja muut (2024, s. 10–11) huomauttavat lisäksi, että domainien tuottamien kuvausten ja datan kuluttajien tarpeiden välillä voi esiintyä ristiriitoja, jotka vaikeuttavat datan oikeaa tulkintaa.

Machado ja muut (2022, s. 266) sekä Dehghani (2022, s. 68, 71) jakavat domainit lähdesuuntautuneisiin ja kuluttajasuuntautuneisiin domaineihin. Lähdesuuntautuneet domainit tuottavat raakaa dataa sen syntypisteessä, kun taas kuluttajasuuntautuneet domainit muodostavat uusia datatuotteita yhdistelemällä ja jalostamalla muiden domainien dataa.

2.1.2 Tuotteen ajattelutapa datalle

Data meshissä dataa käsitellään tuotteena, mikä tarkoittaa, että data suunnitellaan, ylläpidetään ja tarjotaan käyttäjille ohjelmistotuotteiden tavoin sisältäen esimerkiksi metatiedot, rajapinnat ja jatkuvan ylläpidon (Dehghani, 2022, s. 29–31; Bode ja muut, 2024, s. 4). Dehghani (2022, s. 29–30) sekä Goedegebuure ja muut (2024, 11:7) kuvaavat data mesh -mallin korostavan käyttäjälähtöisyyttä, jossa datan kuluttajia tarkastellaan datatuotteiden asiakkaina. Dehghani (2022, s. 33–40) kuvaa datatuotteiden keskeisiksi käytettävyyssominaisuuksiksi löydettävyyden, osoitettavuuden, ymmärrettävyyden, luotettavuuden, turvallisuuden, yhteentoimivuuden, natiivin saavutettavuuden sekä datatuotteen itsenäisen arvon.

Datatuotteille määritellään keskeiset rakenteelliset ja hallinnalliset elementit, kuten rajapinnat, metatiedot, käyttöehdot, palvelutasot (SLA:t ja SLO:t) sekä datasopimukset (Dehghani, 2022, 2022, s. 151–152; Dolhopolov ja muut, 2024, s. 6). Datasopimukset eivät ole pelkkiä teknisiä rajapintakuvauksia, vaan kattavia sopimuksia, jotka määrittelevät datan rakenteen, semantiikan, muunnokset sekä käyttöehdot eri domainien välillä (Serra, 2024, s. 174; Goedegebuure ja muut, 2024, s. 11:8–11:10). Näiden elementtien avulla datan käyttöä voidaan standardoida ja hallita systemaattisesti hajautetussa ympäristössä (Dehghani, 2022, s. 83). Dehghani (2022, s. 83–84) sekä Wider ja muut (2023, s. 749) korostavat, että datatuotteiden ohjelmallinen hyödynnettävyys perustuu selkeästi määriteltyihin rajapintoihin, skeemoihin ja standardeihin. Datatuotteet voivat tarjota dataa esimerkiksi API-, SQL-, suoratoisto- tai BLOB-rajapintojen kautta.

Algoritmissa 1 esitetään YAML-muotoinen esimerkki datasopimuksesta, joka sisältää datatuotteen metadatan, omistajan, käyttöehdot, skeemamäärittelyn sekä laadunvalvontasäännöt. Esimerkki havainnollistaa, miten datatuotteen rakenne, semantiikka, käyttörajoitteet ja laatuvaatimukset voidaan kuvata ohjelmallisesti ja yhtenäisessä muodossa hajautetussa data mesh -ympäristössä. Datasopimuksia voidaan hyödyntää myös automaattisessa validoinnissa, laadunvalvonnassa ja hallintapolitiikkojen toteuttamisessa.

```
# Datasopimuksen versio
dataContractSpecification: 0.9.0

# Datatuotteen yksilöivä tunnistus
id: orders-latest-npii

info:
  # Datatuotteen perustiedot
  title: Orders Latest NPII
  version: 1.0.0
  description: Successful customer orders in the webshop. All orders
  since 2020-01-01.
  owner: Checkout Team

terms:
  # Käyttöehdot ja rajoitukset
  usage: Data can be used for reports, analytics and machine learning
  use cases.
  limitations: Not suitable for real-time use cases.
  billing: 5000 USD per month
  noticePeriod: P3M

schema:
  # Datamallin rakenne
  type: dbt
  specification:
    models:
      - name: orders
        columns:
          - name: order_id
            type: string
            description: Primary key of the orders table

          - name: order_timestamp
            type: timestampz
            description: The business timestamp in UTC when the order
            was successfully registered.

          - name: order_total
            data_type: integer
            description: Total amount of the order in the smallest mone-
            tary unit (e.g., cents).

quality:
```

```

# Datun laadun automaattiset tarkistukset
type: SodaCL
specification:
  checks for orders:
    - row_count between 100000 and 300000

```

Algoritmi 1. Esimerkkinuotoinen datasopimus (Christ ja muut, n.d.).

Datasopimus määrittelee datan rakenteen, käyttöehdot ja laatuvaatimukset, kun taas datatuotteen metadata- ja katalogimäärittely tukee datatuotteen löydettävyyttä, omistajuuden hallintaa ja ohjelmallista rekisteröintiä data-alustalla. Algoritmissa 2 esitetään DataHubin esimerkin pohjalta YAML-muotoinen datatuotteen metadata-määrittely, jossa datatuote liitetään domainiin, omistajiin, datavarantoihin, ulostulorajapintaan (output port), tunnisteisiin ja sanastotermeihin. DataHubin dokumentaatiossa YAML-pohjaista data product -määrittelyä käytetään datatuotteiden määrittelyyn ja hallintaan koodina.

```

# Datatuotteen yksilöivä tunniste
id: pet_of_the_week

# Domain, johon datatuote kuuluu
domain: Marketing

# Datatuotteen katalogissa näytettävä nimi
display_name: Pet of the Week Campaign

# Lyhyt kuvaus datatuotteen sisällöstä
description: |-
  This campaign includes Pet of the Week data.

# Datatuotteeseen kuuluvat data-assetit, kuten datasetti, dashboard ja
dataputki
assets:
  - urn:li:dataset:(urn:li:dataPlatform:snowflake,long_tail_compan
ions.analytics.pet_details,PROD)
  - urn:li:dashboard:(looker,dashboards.19)
  - urn:li:dataFlow:(airflow,snowflake_load,prod)

# Datatuotteen tarjoama output port eli kuluttajille julkaistava ra-
japinta
output_ports:
  - urn:li:dataset:(urn:li:dataPlatform:snowflake,long_tail_compan-
ions.analytics.pet_details,PROD)

# Datatuotteen omistajat ja vastuut
owners:
  - id: urn:li:corpuser:jdoe

```

```

    type: BUSINESS_OWNER
  - id: urn:li:corpuser:fbar
    type: urn:li:ownershipType:architect

# Datatuotteeseen liitetyt tunnisteet löydettävyyden tukemiseksi
tags:
  - urn:li:tag:adoption

# Datatuotteeseen liitetyt sanastotermit semanttisen ymmärrettävyyden
tukemiseksi
terms:
  - urn:li:glossaryTerm:ClientsAndAccounts.AccountBalance

# Lisädokumentaatio tai organisaation sisäinen tieto datatuotteesta
institutional_memory:
  elements:
    - title: URL for campaign
      description: |-
        Go here to see the campaign.
      url: https://example.com/pet_of_the_week

# Datatuotteen omat lisäominaisuudet, kuten elinkaaren tila ja palvelu-
taso
properties:
  lifecycle: production
  sla: 7am every day

```

Algoritmi 2. Esimerkkimuotoinen datatuotteen metadata- ja katalogimäärittely (DataHubin esi-
merkin pohjalta suomennettu ja mukailtu, DataHub, n.d.).

Vaikka datan käsittely tuotteena parantaa käytettävyyttä ja hallittavuutta, se asettaa myös uusia vaatimuksia organisaatioille (Dehghani, 2022, s. 30). Data mesh -kirjallisuudessa datatuotteiden hallintaan liitetään esimerkiksi roolit kuten domainin datatuotteen omistaja (domain data product owner) ja datatuotteen kehittäjä (data product developer), jotka vastaavat datatuotteiden kehittämisestä, ylläpidosta ja käytettävyydestä (Dehghani, 2022, s. 42; Araújo Machado ja muut, 2022, s. 267). Dehghani (2022, s. 42, 57) sekä Goedegebuure ja muut (2024, 11:14) kuvaavat datatuotteiden ylläpidon edellyttävän osaamista esimerkiksi dokumentaatiosta, versionhallinnasta, testauksesta sekä palveluiden ja dataputkien ylläpidosta.

2.1.3 Itsepalveludata-alusta

Itsepalveludata-alusta tarjoaa domain-tiimeille työkalut datan käsittelyyn, kuten orkestrointipalveluita ja Infrastructure as Code -malleja, joiden avulla dataputkia ja infrastruktuurin provisiointia voidaan automatisoida ilman, että domain-tiimien tarvitsee hallita infrastruktuurin matalan tason toteutuksia itse (Goedegebuure ja muut, 2024, 11:13-11:14; Serra, 2024, s. 175).

Bodapati (2025, s. 774), Bode ja muut (2024, s. 4) sekä Dehghani (2022, s. 8, 66, 116) kuvaavat itsepalveludata-alustan keskeisenä tavoitteena teknisen kompleksisuuden abstrahointia siten, että domain-tiimit voivat keskittyä liiketoimintalähtöiseen datan hyödyntämiseen infrastruktuurin hallinnan sijaan. Tämä tukee hajautetun mallin autonomiaa vähentämällä domain-tiimien tarvetta hallita monimutkaista infrastruktuuria itse ja mahdollistamalla itsenäisemmän työskentelyn (Bode ja muut, 2024, s. 4; Bodapati, 2025, s. 774; Dehghani, 2022, s. 55, 66).

Alustan keskeisiä toimintoja ovat datan tallennus ja prosessointi, datatuotteiden julkaisu, käyttöoikeuksien hallinta sekä metadatan hallinta (Goedegebuure ja muut, 2024, 11:14). Goedegebuure ja muut (2024, 11:8, 11:14) kuvaavat datakatalogin keskeiseksi osaksi itsepalvelualustaa, koska se mahdollistaa datatuotteiden löydettävyyden, metadatan hallinnan ja omistajuuden läpinäkyvyyden.

Gahle ja Barbarino (2026, s. 2–4) kuvaavat käytännön toteutuksissa hajautettua data-alustaa, joka hyödyntää pilvipohjaisia ratkaisuja, kuten Snowflakea datan käsittelyyn ja Azure Blob Storagea datan tallennukseen. Heidän mukaansa alusta tukee datan federointia tarjoamalla yhtenäisen pääsypisteen hajautettuihin tietovarantoihin ilman, että data kulkee keskitetyn infrastruktuurin kautta. Toteutuksessa hyödynnetään lisäksi metadata-ohjattuja ingestio-putkia (metadataohjattuja datan latausputkia), API-rajapintoja sekä CI/CD-käytäntöjä datan hallinnan ja integraation tukena. Vaikka kyseessä ei ole varsinainen data mesh -toteutus, ratkaisu sisältää data mesh -arkkitehtuurille tyypillisiä piirteitä, kuten hajautettuja tietolähteitä, federoitua datan käyttöä sekä ohjelmallisesti

hallittuja rajapintoja. Data mesh ei määrittele tiettyä teknologiaa, vaan toimintamallin ja arkkitehtuuriperiaatteet.

Data mesh ei välttämättä tarkoita aiemmista teknologioista luopumista, vaan olemassa olevaa data lake -arkkitehtuuria voidaan hyödyntää osana itsepalvelualustaa esimerkiksi mesh-solmujen (yksittäinen domain tai datatuote + sen infrastruktuuri) tallennusratkaisuna (Araújo Machado ja muut, 2022, s. 270).

Goedegebuure ja muut (2024, 11:16-11:18, 11:26) korostavat, että itsepalvelualustan kehittäminen voi olla haastavaa ja edellyttää kypsää data-alustaa sekä osaavia data-alustainsinöörejä. Ilman riittävän kypsää itsepalvelualustaa domain-tiimit eivät kykene tehokkaasti rakentamaan ja hallitsemaan datatuotteita itsenäisesti, mikä voi vaikeuttaa data mesh -arkkitehtuurin käytännön toteutusta.

2.1.4 Hajautettu laskennallinen hallinto

Hajautettu laskennallinen hallinto (federated computational governance) yhdistää hajautetun omistajuuden ja yhteiset standardit data mesh -arkkitehtuurissa. Mallissa domain-tiimit vastaavat operatiivisista päätöksistä ja omien datatuotteidensa hallinnasta, kun taas organisaatiotason hallinto määrittelee yhteiset periaatteet, standardit ja politiikat (Dehghani, 2022, s. 68, 77).

Federoitu hallinta pyrkii tasapainottamaan domain-tiimien autonomian ja koko organisaation yhteentoimivuuden. Dehghani (2022, s. 71) sekä Dolhopolov ja muut (2024, s. 4) kuvaavat tätä 'dynaamiseksi tasapainoksi', jossa hallintasäännöt ja politiikat mahdollistavat yhteisen toimintamallin ilman, että domain-tiimien ketteryyttä rajoitetaan liikaa.

Hallinto toteutetaan usein automatisoidusti esimerkiksi datan validointisääntöjen, pääsynhallintapolitiikkojen sekä skeema- ja API-standardien avulla, jotka varmistavat datatuotteiden yhdenmukaisuuden ja sääntöjenmukaisuuden hajautetussa ympäristössä (Dolhopolov ja muut, 2024, s. 5–6; Goedegebuure ja muut, 2024, 11:13–11:14). Widerja

muut (2023, s. 749) sekä Dolhopolov ja muut (2024, s. 5, 14) korostavat lisäksi datatuotteiden määriteltyjä syöttö- ja ulostuloportteja, joiden kautta dataa voidaan hyödyntää ja yhdistellä muiden datatuotteiden kanssa. Dehghani (2022, s. 80, 84) liittyy yhteiset standardit ja rajapinnat keskeiseksi osaksi datatuotteiden yhteentoimivuutta, sillä ilman niitä hajautettu malli voi johtaa datan sirpaloitumiseen ja vaikeuksiin datatuotteiden yhdisteltävyydessä.

Dolhopolov ja muut (2024, s. 7–8) tarkentavat 3 erillistä tapaa järjestää metadatan hallintaa data mesh -ympäristössä. Ensimmäisessä mallissa metadata tallennetaan keskitettyyn metadataravastoon (Type I), mikä mahdollistaa vahvan kontrollin mutta rajoittaa domain-tiimien autonomiaa. Toisessa mallissa metadata on federoitu (Type II), jolloin jokaisella domainilla on oma metadatasolmunsa, mutta solmut synkronoidaan keskenään. Tämä vastaa parhaiten data mesh -arkkitehtuurin tavoitetta tasapainottaa autonomia ja yhteentoimivuus. Kolmannessa mallissa metadata on täysin hajautettu (Type III), mikä maksimoi domain-autonomian, mutta tekee yhteentoimivuuden ja sääntöjen valvonnan huomattavasti vaikeampaa (Dolhopolov ja muut, 2024, s. 7–8).

Federoitu hallintamalli on vaikea toteuttaa käytännössä, koska se edellyttää samanaikaisesti organisaatiokulttuurin muutosta, selkeää vastuunjakoja sekä riittävää teknistä automaatiota, joiden yhteensovittaminen on osoittautunut haastavaksi (Bode ja muut, 2024, s. 10, 19).

2.2 Organisaation muutos ja vastuunjako data mesh -arkkitehtuurissa

Kuten luvussa 2.1.1 todettiin, data mesh siirtää datan omistajuutta liiketoimintadomaineille. Tässä luvussa tarkastellaan tarkemmin, miten tämä muutos vaikuttaa organisaation vastuunjakoon. Malli ei kuitenkaan poista keskitettyä roolia kokonaan, vaan keskitetty IT vastaa edelleen alustasta, standardeista ja tietoturvasta, mikä tekee arkkitehtuurista käytännössä hybridin. Taulukossa 2 esitetään keskeiset erot perinteisen data-arkkitehtuurin ja data mesh -mallin vastuunjaossa eri päätösalueilla.

Taulukko 2 osoittaa, että vastuu siirtyy erityisesti datan laadun, mallinnuksen ja sisällön osalta domaineille, kun taas keskitetty IT säilyttää keskeisen roolin yhteisten standardien ja infrastruktuurin määrittelyssä. Tämä muutos vaikuttaa suoraan siihen, miten dataa tuotetaan, hallitaan ja hyödynnetään organisaatiossa.

Taulukko 2. Vastuunjako perinteisessä data-arkkitehtuurissa ja data mesh -mallissa (Dehghani, 2022, s. 48–151).

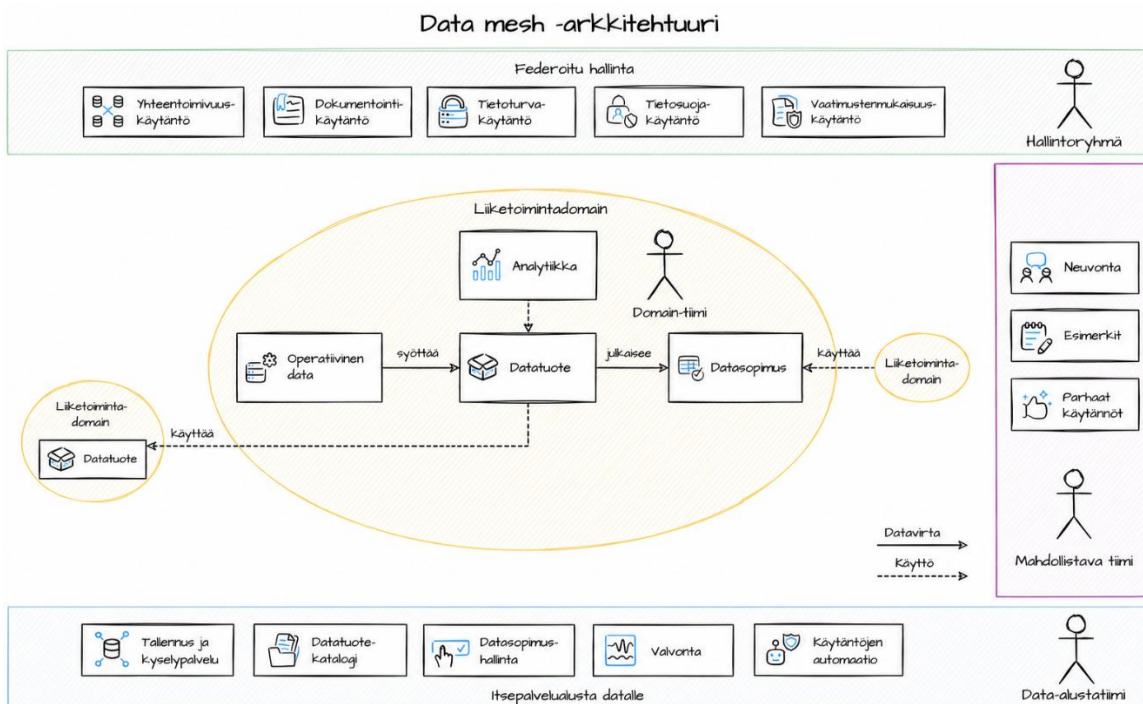
Päätösalue	Perinteinen omistaja	Data mesh -omistaja
Data-arkkitehtuuri	Keskitetty IT	Jaettu: keskitetty IT määrittää standardit, domainit suunnittelevat datatuotteet
Datan laatu	Keskitetty IT	Domainit
Datan saatavuus	Keskitetty IT	Jaettu hallintamalli
Data-alusta	Keskitetty IT	Keskitetty alustatiimi
Datan mallinnus	Keskitetty IT	Domainit

Data mesh -mallissa keskitetyn IT-osaston rooli muuttuu perinteisestä datan omistajasta mahdollistavaksi alustatiimiksi, jonka tehtävänä on tarjota yhteinen itsepalvelualusta ja varmistaa tekninen sekä hallinnollinen yhtenäisyys (Bode ja muut, 2024, s. 4; Goedegebuure ja muut, 2024, 11:13; Dehghani, 2022, s. 48). Goedegebuure ja muut (2024, 11:13) sekä Serra (2024, s. 175, 210) kuvaavat käytännössä tämän tarkoittavan infrastruktuurin, tietoturvan ja standardien automatisointia esimerkiksi Infrastructure as Code -mallien avulla, jotta domain-tiimit voivat kehittää datatuotteita ilman syvää infrastruktuurin hallinnan osaamista. Bode ja muut (2024, s. 4), Goedegebuure ja muut (2024, 11:18) sekä Dolhopolov ja muut (2024, s. 7) korostavat, että ilman yhteisiä alustapalveluita ja standardeja hajautettu malli voi johtaa tekniseen päällekkäisyyteen, yhteensopimattomiin ratkaisuihin ja heikentyneeseen yhteentoimivuuteen.

Dehghani (2022, s. 27), Serra (2024, s. 172, 176) sekä Bode ja muut (2024, s. 11) toteavat samalla, että päätösvalta datan sisällöstä ja käytöstä siirtyy enemmän domaineille, kun taas keskitetty IT ohjaa toimintaa epäsuorasti alustan, standardien ja politiikkojen kautta. Samalla domain-tiimien rooli laajenee merkittävästi. Ne eivät enää pelkästään tuota dataa, vaan vastaavat myös sen laadusta, mallinnuksesta ja jakamisesta datatuotteina

muille (Dehghani, 2022, s. 27, 85; Goedegebuure ja muut, 2024, 11:10; Bode ja muut, 2024, s. 11). Tämä lisää domainien autonomiaa, mutta myös vastuuta siitä, että data on hyödynnettävää organisaation laajuisesti (Dehghani, 2022, s. 55; Bode ja muut, 2024, s. 11; Serra, 2024, s. 173–175).

Kuvassa 1 havainnollistetaan data mesh -arkkitehtuurin vastuunjako ja keskeisiä toimijoita. Domain-tiimit vastaavat operatiivisen datan jalostamisesta datatuotteiksi ja niiden julkaisemisesta datasopimusten kautta, kun taas keskitetty alustatiimi tarjoaa yhteisen itsepalvelualustan sekä automatisoidut hallinta- ja tietoturvamekanismit. Ylimpänä federoitu hallintamalli määrittää organisaation yhteiset politiikat, standardit ja yhteentoimivuuseriaatteet. Kuva havainnollistaa, miten data mesh yhdistää domain-kohtaisen autonomian ja keskitetyn hallinnan hajautetussa ympäristössä.



Kuva 1. Data mesh -arkkitehtuurin vastuunjako ja hallinta. Kuva perustuu lähteeseen Christ ja muut (n.d.), ja sen teksti on suomennettu tekoälyn avulla ChatGPT:tä hyödyntäen. Kehote: 'Käännä kuvan englanninkieliset tekstit suomeksi säilyttäen alkuperäinen rakenne, visuaalinen asettelu, symbolit, värit ja sisältö mahdollisimman muuttumattomina. Säilytä alkuperäisten termien määrä ja merkitys ilman lisätyjä tulkintoja.'

Goedegebuure ja muut (2024, 11:17, 11:20), Bodapati (2025, s. 773) sekä Serra (2024, s. 195–196) korostavat, että data mesh edellyttää domain-tiimeiltä uudenlaista osaamista ja kykyä kehittää sekä ylläpitää datatuotteita itsenäisesti. Tämä edellyttää teknisiä taitoja, domain-asiantuntemusta sekä tuotemaisia toimintamalleja. Kaikilla tiimeillä ei kuitenkaan ole tällaisia resursseja tai osaamista, mikä voi säilyttää riippuvuutta keskitetystä tu-esta ja vaikeuttaa hajautetun toimintamallin käytännön toteutusta (Goedegebuure ja muut, 2024, 11:17; Serra, 2024, s. 198; Bode ja muut, 2024, s. 12).

Näiden tekijöiden seurauksena data mesh -arkkitehtuuri luo jatkuvan jännitteen do-main-autonomian ja keskitetyn ohjauksen välille (Dehghani, 2022, s. 71, 80; Goedegebuure ja muut, 2024, 11:11). Liiallinen keskittäminen voi palauttaa mallin kohti perinteistä arkkitehtuuria, kun taas liian vähäinen ohjaus voi johtaa hajanaisiin käytäntöihin ja heikentyneeseen yhteentoimivuuteen (Dehghani, 2022, s. 80; Serra, 2024, s. 191; Bode ja muut, 2024, s. 10). Dolhopolov ja muut (2024, s. 4) sekä Dehghani (2022, s. 71) kuvaavat onnistuneen toteutuksen edellyttävän dynaamista tasapainoa näiden välillä.

2.3 Datan elinkaari data mesh -arkkitehtuurissa

Serra (2024, s. 148) kuvaa modernin data-arkkitehtuurin datan elinkaarena, jossa data etenee ingestoinnista ja tallennuksesta kohti jalostettua, analyytiikka- ja päätöksentekokäyttöön soveltuvaa muotoa.

Data mesh ei määrittele vain rakennetta, vaan myös sen, miten dataa kerätään, hallitaan ja hyödynnetään hajautetussa ympäristössä. Datan käsittely toteutuu dataputkien kautta, joiden rooli muuttuu merkittävästi data mesh -arkkitehtuurissa. Perinteisissä ark-kitehtuureissa dataputket ovat usein keskitetysti hallittuja ja näkyviä arkkitehtuurikomponentteja, mutta data meshissä ne muuttuvat datatuotteen sisäisiksi toteutusyksityiskohdiksi (Dehghani, 2022, s. 26, 28). Dehghani (2022, s. 11, 153) sekä Goedegebuure ja muut (2024, 11:7) kuvaavat, että datatuotteet yhdistävät datan, sitä käsittelevän koodin sekä infrastruktuurin yhdeksi autonomiseksi kokonaisuudeksi. Tällöin dataputket ja

transformaatiologiikka kapseloidaan datatuotteen sisäiseksi toiminnallisuudeksi, joka vastaa datan ingestoinnista, käsittelystä ja tarjoamisesta kuluttajille. Lisäksi dataputkiin integroidaan usein automaattisia validointi- ja auditointimekanismeja, jotka varmistavat datan laadun ennen sen julkaisemista datatuotteena (Machado ja muut, 2022, s. 269; Serra, 2024, s. 46).

Data mesh haastaa perinteisen keskitetyn ETL-ajattelun siirtämällä dataputket ja transformaatiologiikan domain-kohtaisten datatuotteiden sisäiseksi toteutusyksityiskohdiksi (Dehghani, 2022, s. 26, 28, 153). Samalla datatuotteiden rajapinnat, skeemat ja hallintapolitiikat pyritään standardoimaan ohjelmallisesti, jotta dataa voidaan hyödyntää johdonmukaisesti eri domainien välillä (Dehghani, 2022, s. 84–85). Itsepalvelualusta tukee tätä tarjoamalla standardoituja työkaluja (luvussa 2.1.3).

2.3.1 Datan kerääminen

Koska datan omistajuus säilyy domaineilla, kuten luvussa 2.1.1 kuvattiin, datan kerääminen tapahtuu domain-kohtaisesti operatiivisista järjestelmistä (Dehghani, 2022, s. 21–22; Goedegebuure ja muut, 2024, 11:8). Dehghani (2022, s. 27) sekä Bode ja muut (2024, s. 4) korostavat, että toisin kuin perinteisissä malleissa, dataa ei siirretä keskitettyyn hallintaan, vaan omistajuus säilyy niillä tiimeillä, jotka tuntevat datan parhaiten. Goedegebuure ja muut (2024, 11:8) kuvaavat lähdesuuntautuneiksi (source-aligned) datatuotteiksi esimerkiksi tilaus-, lasku- ja tuotetietoja, jotka perustuvat operatiivisiin tietolähteisiin. Serra (2024, s. 178) antaa esimerkkeinä lähdesuuntautuneesta domain-datasta liiketoimintadomainien alkuperäisiä operatiivisia tietoja, kuten maksuihin ja asiakkuuksiin liittyvää dataa. Dolhopolov ja muut (2024, s. 3–4) puolestaan kuvaavat lähdesuuntautuneeksi domain-dataksi toimitusketjuun, logistiikkaan ja varastotilanteisiin liittyviä tietoja hajautetussa data mesh -ympäristössä.

Datan keräämisessä on myös tärkeää varmistaa datan käytettävyys ('fit for use') (Strong ja muut, 1997, s. 103–104). Datan käytettävyyteen ja yhteentoimivuuteen vaikuttavat keskeisesti myös sen laatuominaisuudet, kuten tarkkuus ja eheys (ISO/IEC 25012, 2008,

s. 7–10). Keräysvaiheessa datan laatu ja validointi vaikuttavat siihen, voidaanko data myöhemmin julkaista luotettavana datatuotteena.

2.3.2 Datan analysointi

Dehghani (2022, s. 11) sekä Goedegebuure ja muut (2024, 11:2–11:3) kuvaavat analysoinnin tapahtuvan hajautetusti domainien tuottamien datatuotteiden pohjalta, mikä mahdollistaa ristiinanalyysin (cross-domain analysis) yhdistämällä eri domaineista saatavia tietoja ilman datan siirtämistä keskitettyyn hallintaan. Goedegebuure ja muut (2024, 11:19, 11:24) sekä Dehghani (2022, s. 35) kuvaavat itsepalvelualustan tukevan tätä tarjoamalla työkaluja, kuten federoidun kyselymoottorin ja BI-työkaluja, joiden avulla liiketoimintakäyttäjät voivat luoda oivalluksia itsenäisesti.

Analyysin onnistuminen riippuu kuitenkin vahvasti datan ymmärrettävyydestä ja meta-datan laadusta. Bode ja muut (2024, s. 11) sekä Goedegebuure ja muut (2024, 11:9) korostavat, että ilman riittävää kontekstia dataa on vaikea tulkita oikein. Ilman riittävää dokumentaatiota, semanttista yhtenäisyyttä ja datan alkuperän (data lineage) näkyvyyttä analyytikot eivät voi luottaa dataan, mikä johtaa datasoihin (data swamps) ja hidastaa päätöksentekoa (Goedegebuure ja muut, 2024, 11:11, 11:12; Dolhopolov ja muut, 2024, s. 2). Vaikutuksia eri käyttäjä- ja sidosryhmiin, kuten data-analyytikoihin, data scientistis-teihin ja liiketoimintakäyttäjiin, tarkastellaan tarkemmin luvussa 3.2.

2.3.3 Datan hallinnointi ja jakaminen

Dehghani (2022, s. 76, 92), Araújo Machado ja muut (2022, s. 267) sekä Goedegebuure ja muut (2024, 11:10) korostavat, että datatuotteiden hallinta edellyttää uusia domainkohtaisia vastuita ja rooleja, kuten datatuotteen omistaja- ja datatuotteen kehittäjä -rooleja, jotka vastaavat datatuotteiden kehittämisestä, laadusta, käytettävyydestä ja teknisestä ylläpidosta koko niiden elinkaaren ajan. Dehghani (2022, s. 88) sekä Goedegebuure ja muut (2024, 11:10, 11:12) kuvaavat datatuotteen omistajan vastaavan erityisesti datatuotteen liiketoiminta-arvosta, kuluttajien tarpeista ja palvelutasotavoitteista. Wider

ja muut (2023, s. 748) sekä Dehghani (2022, s. 88, 92) kuvaavat datatuotteen kehittäjän vastaavan datatuotteen teknisestä toteutuksesta, rajapinnoista ja datan jalostamisesta analysoitavaan muotoon. Data mesh -arkkitehtuurin vaikutuksia näihin rooleihin ja niiden vastuisiin tarkastellaan tarkemmin luvussa 3.2.

Data mesh -arkkitehtuurissa dataa ei siirretä keskitettyyn varastoon, vaan sitä jaetaan hajautetusti datatuotteina (Dehghani, 2022, s. 11). Kuten luvuissa 2.1.2 ja 2.1.4 todettiin, datan jakaminen perustuu standardoituihin rajapintoihin, metatietoon, datasopimuksiin sekä palvelutasolupauksiin (SLA/SLO), ja sitä ohjataan yhteisillä mutta hajautetusti toteutettavilla hallintaperiaatteilla.

Datatuote toimii arkkitehtuurisena perusyksikkönä (data product quantum), jonka kautta data tehdään muiden domainien hyödynnettäväksi (Dehghani, 2020; Araújo Machado ja muut, 2022, s. 267). Jakaminen ei perustu datan fyysiseen siirtämiseen, vaan usein nollakopiointiin (zero-copy) ja virtualisointiin, jolloin dataan tarjotaan looginen pääsy ilman tarpeetonta duplikointia (Dehghani, 2022, s. 21; Goedegebuure ja muut, 2024, 11:17; Serra, 2024, s. 82, 154).

2.4 Data mesh -arkkitehtuurin soveltuvuuden yleiset edellytykset

Tässä luvussa tarkastellaan yleisellä tasolla, millaisille organisaatioille data mesh -arkkitehtuuri soveltuu. Tarkastelu luo pohjan tutkimuskysymykseen 4 vastaamiselle, jossa arvioidaan, millaisissa organisaatioissa data mesh tukee parhaiten datan käytettävyyttä ja yhteentoimivuutta. Luvuissa 3 ja 4 arvioidaan lisäksi käytettävyyteen ja yhteentoimivuuteen liittyvien havaintojen avulla, millaisia edellytyksiä ja valmiuksia data mesh -arkkitehtuurin hyödyntäminen organisaatioilta vaatii.

Bode ja muut (2024, s. 16–17) kuvaavat data mesh -arkkitehtuurin soveltuvuuden riippuvan organisaation rakenteesta, datan määrästä ja teknisestä kyvykkyydestä. He erottavat startup- ja scaleup-yritykset sekä vakiintuneet organisaatiot toisistaan. Startup- ja

scaleup-ympäristöissä korostuvat ketteryys, nopea skaalautuvuus ja selkeä omistajuus. Käytännössä tämä voi mahdollistaa data mesh -mallin rakentamisen ikään kuin puhtaalta pöydältä. Vakiintuneissa organisaatioissa käyttöönottoa vaikeuttavat puolestaan usein legacy-järjestelmät, jäykät organisatoriset rakenteet ja keskitetyt data-arkkitehtuurit (Bode ja muut, 2024, s. 16–17).

Serra (2024, s. 137, 202) kuvaa arkkitehtuurivalinnan riippuvan datan mittakaavasta ja organisatorisesta kompleksisuudesta. Hänen mukaansa perinteinen tietovarasto voi edelleen olla riittävä ratkaisu organisaatioille, joiden datamäärät ovat suhteellisen pieniä – esimerkiksi alle 10 teratavun luokkaa – ja joiden analytiikkatarpeet ovat maltillisia. Data mesh puolestaan on suunnattu erityisesti suuriin, domain-orientoituneisiin organisaatioihin, joilla on resursseja hallita hajautettua toimintamallia. Bodapati (2025, s. 773) korostaa lisäksi, että data mesh edellyttää domain-kohtaisten tiimien, rajapintojen ja hallintamallien systemaattista suunnittelua sekä keskitetyn alustan ja hajautetun vastuun yhdistämistä.

Lisäksi data mesh edellyttää domaineilta riittävää teknistä osaamista, kuten data-analytiikan, data engineeringin (datan keräämiseen, muuntamiseen ja hallintaan liittyvän teknisen kehitystyön) ja koneoppimisen kyvykkyyksiä. Osaavan henkilöstön puute on tunnistettu merkittäväksi käyttöönottoa rajoittavaksi tekijäksi erityisesti hajautetuissa ympäristöissä (Goedegebuure ja muut, 2024, 11:17). Goedegebuure ja muut (2024, 11:18) tunnistavat skenaarioita, joissa data mesh ei ole suositeltava: tällaisia ovat muun muassa organisaatiot, joilta puuttuu domain-orientaatio tai joilla on alhainen datan hallintamallin ja itsepalvelualustan kypsyysaste. Jos hyödyllisiä analyttisiä tietolähteitä on vähän, siirtymän korkeat kustannukset ja riskit ylittävät usein saavutettavat hyödyt (Goedegebuure ja muut, 2024, 11:18). Serra (2024, s. 196) korostaa, että hajautettu toimintamalli voi kasvattaa operatiivisia kustannuksia ja organisatorista kompleksisuutta, koska jokainen domain tarvitsee omia teknisiä asiantuntijoita ja ylläpitokyvykkyyksiä keskitetyn IT-mallin sijaan. Tämän vuoksi data mesh edellyttää riittäviä organisatorisia resursseja ja osaamista.

Bodapati (2025, s. 771) raportoi, että 63 % Fortune 1000 -yrityksistä kamppailee tasapainon löytämisessä keskitetyn hallinnon ja domain-kohtaisten tarpeiden välillä. Lisäksi hajautettu arkkitehtuuri edellyttää organisaatioilta kyvykkyyttä ylläpitää yhteisiä standardeja, hallita datatuotteiden rajapintoja ja koordinoida toimintaa yli domain-rajojen (Bodapati, 2025, s. 772–773). Araújo Machado ja muut (2022, s. 265–266) kuvaavat data meshin paradigman muutoksena, jossa siirrytään keskitetystä data-arkkitehtuurista kohti tuotekeskeistä ja hajautettua vastuunjakoja. Onnistunut soveltaminen edellyttää organisaatioilta kykyä omaksua domain-lähtöinen toimintamalli ja datan tarkastelu tuotteena.

Näiden havaintojen perusteella data mesh tukee parhaiten datan käytettävyyttä ja yhteentoimivuutta organisaatioissa, joissa domain-rakenne on selkeä, datan määrä on suuri sekä hallintamallit ja itsepalvelualusta ovat riittävän kypsiä. Vastaavasti näiden puuttuessa hajautettu malli voi heikentää käytettävyyttä ja vaikeuttaa yhteentoimivuutta.

3 Datan käytettävyys data mesh -arkkitehtuurissa

Tässä luvussa analysoidaan data mesh -arkkitehtuurin vaikutuksia datan käytettävyyteen (TK1) sekä tunnistetaan käytettävyyttä edistäviä ja estäviä tekijöitä (TK3). Luvuissa 3.1–3.3 tutkitaan data mesh -arkkitehtuurin vaikutuksia datan käytettävyyteen eri näkökulmista, kuten datatuotteiden ominaisuuksien, sidosryhmien sekä tietoturvan kautta. Sidoryhmä- ja tietoturvanäkökulmia tarkastellaan erillisinä kokonaisuuksina, koska niiden vaikutukset ulottuvat yksittäisiä datatuoteominaisuuksia laajemmin hajautetun arkkitehtuurin toimintaan ja datan hyödyntämiseen. Luvussa 3.4 tarkastelu siirtyy näitä vaikutuksia selittäviin tekijöihin. Siinä tunnistetaan käytettävyyttä edistävät ja estävät tekijät. Samalla luodaan pohjaa sen arvioinnille, millaisissa organisaatioissa malli parhaiten tukee käytettävyyttä (TK4), ja luvun lopussa esiteltävät mittarit kuvaavat, mitä käytettävyyden osa-alueisiin liittyviä tekijöitä voidaan mitata ja seurata käytännössä.

Data mesh -kirjallisuudessa käytettävyys liitetään datatuotteiden niin sanottuihin 'ei-neuvoteltaviin' ominaisuuksiin, joihin kuuluvat löydettävyys (discoverability), ymmärrettävyys (understandability), luotettavuus (trustworthiness), osoitettavuus (addressability), turvallisuus (security), yhteentoimivuus (interoperability), natiivi saavutettavuus (native accessibility) sekä datatuotteen itsenäinen arvo (valuable on its own) (Dehghani, 2022, s. 33–40). Nämä ominaisuudet määrittävät, millä edellytyksillä dataa voidaan hyödyntää hajautetussa ympäristössä ja siirtävät vastuun datan käytettävyydestä suoraan datan tuottajille (Dehghani, 2022, s. 30, 45). Tässä luvussa käytettävyyden tarkastelu perustuu enimmäkseen tähän viitekehykseen.

'Ei-neuvoteltavien' ominaisuuksien lisäksi, käytettävyyttä tarkastellaan täydentävien näkökulmien kautta, jotka on koottu Taulukkoon 3. Näkökulmia ja viitekehyksiä käytetään erityisesti datan käytettävyyden eri osa-alueiden tunnistamiseen, minkä jälkeen data mesh -arkkitehtuurin vaikutuksia näihin osa-alueisiin tarkastellaan myös muiden lähteiden kautta. Nämä näkökulmat eivät ole toisistaan erillisiä, vaan ne ovat osittain päällekkäisiä 'ei-neuvoteltavien' data mesh -ominaisuuksien kanssa. Tämän vuoksi täydentävistä

näkökulmista hyödynnetään vain sellaisia osa-alueita, jotka täydentävät data mesh -kirjallisuuden omaa käytettävyyšnäkökulmaa. ISO/IEC 25012 -standardi tarkastelee datan laatua teknisten ja järjestelmäriippuvaisten ominaisuuksien kautta, kuten saatavuutta, saavutettavuutta, siirrettävyyttä, eheyttä ja luottamuksellisuutta. Wangin ja Strongin (1996) laatunäkökulma puolestaan painottaa käytettävyyden kontekstisidonnaisuutta. Näkökulmassa datan laatu muodostuu 20 dimensiosta, jotka on ryhmitelty useisiin laatuokkiin, joista vain osaa käsitellään tässä tutkimuksessa. Mallia hyödynnetään ensisijaisesti 'fit for use' -laatunäkökulmana. Käyttäjä- ja UX-näkökulma täydentää tätä korostamalla käytettävyyden subjektiivisuutta, kognitiivista kuormaa ja käyttökokemuksen sujuvuutta (Kashfi ja muut, 2016). Yhdessä nämä näkökulmat täydentävät data mesh -lähestymistapaa: ISO korostaa teknisiä edellytyksiä, laatunäkökulma sen soveltuvuutta käyttökontekstiin ja UX-näkökulma datan käytön sujuvuutta. Lisäksi käytettävyyttä voidaan tarkastella sosioteknisenä näkökulmana, jossa datan ymmärrettävyys ja hyödynnettävyys riippuvat myös dokumentaatiosta, metadatatista ja käyttökontekstista. Swartsin mukaan datan hyödyllisyys ja käytettävyys ovat toisiaan täydentäviä ilmiöitä (Swarts, 2025).

Wang ja Strong (1996, s. 14–16) korostavat, että datan laatu ja käytettävyys muodostuvat useista toisiaan täydentävistä ulottuvuuksista. Tämän perusteella voidaan päätellä, että yksittäinen mittarikaan ei riitä kuvaamaan käytettävyyttä kokonaisvaltaisesti, vaan arviointi edellyttää useiden mittareiden tarkastelua rinnakkain. Datan käytettävyyden arviointia mittareiden avulla tarkastellaan tarkemmin luvussa 3.5.

Taulukko 3. Datan käytettävyyden määrittelyn näkökulmat ja arviointitavat.

Näkökulma ja arviointitavat	Määritelmä	Fokus	Rajoitteet
Laatunäkökulma (Wang ja Strong, 1996)	Data on 'fit for use': vastaa käyttäjän tarpeita ja käyttökontekstia	Datan laatu ja käyttökonteksti (20 dimensiota, kuten relevanssi, ajantasaisuus, ymmärrettävyys)	Ei ota kantaa tekniseen toteutukseen
ISO-standardi (ISO/IEC 25012, 2008)	Käytettävyys osana datan teknistä laatua	Järjestelmäriippuvaliset tekniset ominaisuudet	Korostaa teknistä laatua, rajallinen käyttäjänäkökulma

Data meshin 'ei-neuvoteltavat' ominaisuudet (Dehghani, 2022)	Käytettävyys toteutuu datatuotteiden 'ei-neuvoteltavien' ominaisuuksien kautta	Datatuotteet hajaute- tussa ympäristössä	Ei standar- doitu, toteutus vaihtelee
Käyttäjä-/UX- ja so- siotekninen näkö- kulma (Kashfi ja muut, 2016; Swarts, 2025)	Kuinka helposti ja suju- vasti käyttäjä löytää, ym- märtää ja hyödyntää da- taan sekä miten data liite- tään käyttäjän käyttökon- tekstiin ja merkityksiin	Käyttäjäkokemus ja käy- tön sujuvuus, merkityk- sen muodostuminen ja käytön sujuvuus	Subjektiivinen ja käyttäjäkoh- tainen, vaikea mitata
Mittareihin perus- tuva arviointi	Käytettävyiden osa-aluei- siin liittyviä tekijöitä kuva- taan ja seurataan mitta- reilla	Käytettävyiden tekijöi- den mittaaminen ja seuranta	Ei yksin kuvaa käytettävyttä; osa mittareista epäsuoria

3.1 Datan käytettävyyden osa-alueet

Tässä alaluvussa tarkastellaan ja arvioidaan vaikuttavia tekijöitä datan käytettävyyteen pääasiassa data mesh -kirjallisuuden 'ei-neuvoteltavien' ominaisuuksien kautta.

3.1.1 Löydettävyys ja osoitettavuus

Data meshissä datatuotteiden tulee olla löydettävissä ilman raskaita keskitettyjä välikä- siä. Goedegebuure ja muut (2024, 11:8, 11:16) korostavat systemaattisessa katsaukses- saan, että löydettävyys on yksi datatuotteen ydinkriteereistä. Käytännössä tämä edellyt- tää datakatalogia tai muuta yhteistä rekisteriä, johon datatuotteiden metadata, omista- juus, käyttörajapinnat ja muut itsekuvailevat tiedot julkaistaan, jotta data voidaan löytää ja ymmärtää tehokkaasti yli domain-rajojen (Goedegebuure ja muut, 2024, 11:16; Serra, 2024, s. 198).

Dehghani (2022, s. 35) toteaa, että jokaisella datatuotteella tulee olla yksilöllinen ja py- syvä osoite, kuten URL tai muu ohjelmallisesti käytettävä viittausmekanismi, joka mah- dollistaa datatuotteiden johdonmukaisen käytön myös muuttuvassa ympäristössä. Goedegebuure ja muut (2024, 11:9) puolestaan korostavat, että datatuotteiden rajapin-

tojen tulee tarjota yksilöllisiä ohjelmallisesti saavutettavia päätepisteitä, kuten API-rajapintoja, tallennusrajapintoja, viestinvälitysjärjestelmien topiceja (viestivirtoja) tai SQL-päätepisteitä. Lisäksi osoitteiden ja rajapintojen tulee noudattaa yhteisiä konventioita (kuten nimeämis-, metadata- ja rajapintakäytäntöjä), jotta datatuotteet ovat löydettävissä esimerkiksi datakatalogien kautta ja hyödynnettävissä yhdenmukaisesti eri domainneissa (Dehghani, 2022, s. 35; Goedegebuure ja muut, 2024, 11:8–11:9).

Bode ja muut (2024, s. 2–3, 14) raportoivat, että organisaatiot kokivat datan saavutettavuuden parantuneen, kun dataa ei enää tarvinnut pyytää keskitetyn tiimin kautta, vaan se oli suoraan löydettävissä ja hyödynnettävissä datatuotteina. Myös dbt Labsin (2024) julkaisema Siemensin data mesh -käyttöönottoa kuvaava tapausesimerkki tukee havaintoja datan löydettävyyden ja saavutettavuuden paranemisesta. dbt Labs (2024) raportoi, että aiemmin datan käyttö edellytti oikean omistajan etsimistä ja manuaalista pyyntöä, mikä saattoi kestää viikkoja, mutta data mesh -mallissa pääsy dataan voitiin myöntää minuuteissa. Lisäksi datan latausaika lyheni jopa 93 % (6 tunnista 25 minuuttiin). Tapausesimerkin mukaan datatuotteet koottiin keskitetysti saavutettavaan Siemens Data Cloud Marketplace -ympäristöön, jossa käyttäjät voivat selata ja löytää olemassa olevia datatuotteita.

JPMorganin edustajien mukaan data mesh -arkkitehtuuri kannustaa datan jakamiseen kopioinnin sijaan ja tarjoaa näkyvyyden datan käyttöön eri domaineissa (Jain ja muut, 2021). Tämä voi parantaa saavutettavuutta ja vähentää datan päällekkäisyyttä, koska dataa hyödynnetään yhteisesti sen sijaan, että sitä kopioitaisiin useisiin järjestelmiin (Jain ja muut, 2021; Serra, 2024, s. 193). Toisaalta hajautettu arkkitehtuuri voi myös lisätä datan duplikaation riskiä, jos domainit alkavat ylläpitää omia rinnakkaisia kopioitaan ilman riittävää koordinaointia ja yhteisiä hallintakäytäntöjä.

Volkswagen Autoeuropassa Amazon DataZoneen perustuva data marketplace vähensi datan löytämiseen ja käyttöoikeuksien saamiseen kuluneen ajan päivistä minuutteihin, mikä nopeutti datan hyödyntämistä organisaatiossa (Mukherjee ja muut, 2024).

Covestro puolestaan raportoi Amazon DataZoneen ja SDLF-kehikseen perustuvan data mesh -ratkaisun mahdollistaneen 70 % vähennyksen uusien datatuotteiden time-to-marketissa sekä yli tuhannen dataputken hallinnan hajautetussa ympäristössä (Janssen ja muut, 2025).

Kuitenkin havaittiin, että monissa organisaatioissa ei ole selkeää näkyvyyttä siihen, mitä dataa on olemassa ja mistä sitä löytyy. Tämä viittaa siihen, että hajautus ei itsessään ratkaise löydettävyyttä, vaan ongelma siirtyy metadatan ja datakatalogien laatuun (Bode ja muut, 2024, s. 2–3, 10, 11).

3.1.2 Ymmärrettävyys ja itsekuvailevuus

Deghani (2022, s. 35–36) korostaa, että data meshissä datan ymmärrettävyys perustuu siihen, että datatuotteet ovat itsekuvailevia ja sisältävät dokumentaation, skeemakuvaukset, metadatan sekä semanttisen kontekstin, joiden avulla data voidaan tulkita oikein ilman erillistä manuaalista ohjausta.

Empiiriset havainnot osoittavat, että metadatan laatu on yksi keskeisistä käytännön haasteista data mesh -ympäristöissä. Puutteellinen dokumentaatio voi johtaa tilanteeseen, jossa data on teknisesti saatavilla, mutta käytännössä vaikeasti hyödynnettävissä (Bode ja muut, 2024, s. 10–11). Lisäksi metadatan laatu, vastuunjaon epäselvyydet ja seurattavuuden (observability) puutteet nousivat keskeisiksi ongelmiksi hajautetussa ympäristössä (Bode ja muut, 2024, s. 10–11). Näiden havaintojen perusteella datatuotteiden käytettävyys voi vaihdella merkittävästi domainien välillä riippuen esimerkiksi metadatan laadusta, dokumentaatiosta ja hallintakäytännöistä.

Dolhopolov ja muut (2024, s. 14) korostavat, että metadatan semanttinen rikastaminen on huomattavasti vaikeampaa kuin esimerkiksi datan indeksointi tai käyttöoikeuksien hallinta, mikä johtaa sen alhaiseen toteutusasteeseen käytännön data-alustoilla. Bode ja muut (2024, s. 11) puolestaan havaitsivat kuilun datatuotteiden tarjoajien tuottamien kuvausten ja kuluttajien tarvitsemien tietojen välillä. Lisäksi Goedegebuure ja muut

(2024, 11:12) korostavat yhteisten käsitteiden ja liiketoimintasanastojen merkitystä datatuotteiden semanttisen tulkinnan tukena yli domain-rajojen. Puutteellinen metadata, semanttisen rikastamisen haasteet ja yhteisten käsitteiden puute voivat heikentää datan ymmärrettävyyttä ja käytettävyyttä, jolloin data voi olla teknisesti saatavilla, mutta sen hyödyntäminen on vaikeaa puutteellisen kontekstiymmärryksen vuoksi (Bode ja muut, 2024, s. 11; Goedegebuure ja muut, 2024, 11:10, 11:12). Lisäksi datan ymmärrettävyyttä voivat heikentää käsitteiden erilaiset tulkinnat, yhteisen sanaston puute sekä skeemojen ja ontologioiden erot domainien välillä, mikä voi heikentää datatuotteiden semanttista yhteentoimivuutta ja vaikeuttaa niiden yhdistämistä yli domain-rajojen (Dolhopolov ja muut, 2024, s. 6; Goedegebuure ja muut, 2024, 11:12; Dehghani, 2022, s. 35).

3.1.3 Laatu, luotettavuus ja eheys

Data meshissä laadun odotetaan paranevan, koska vastuu siirtyy lähemmäs datan lähettäjä (Dehghani, 2022, s. 30, 45). Domain-tiimit ymmärtävät oman datansa merkityksen paremmin kuin keskitetty tiimi (Bode ja muut, 2024, s. 10; Dehghani, 2022, s. 107). Goedegebuure ja muut (2024, 11:8, 11:10) korostavat, että datan luotettavuuden arviointia tukevat metadata, palvelutasotavoitteet (SLO) sekä laatuominaisuudet, kuten ajantasaisuus, tarkkuus ja täydellisyys, joita voidaan kuvata ja hallita esimerkiksi datakatalogin avulla.

Wang ja Strong (1996, s. 6) määrittelevät datan laadun datan soveltuvuutena käyttötarkoitukseensa ('fitness for use'). Data mesh -kirjallisuudessa datan laatuun liitetään lisäksi ominaisuuksia, kuten tarkkuus, täydellisyys, eheys ja ajantasaisuus, joita datatuotteiden tulee ylläpitää koko elinkaarensa ajan (Dehghani, 2022, s. 37; Goedegebuure ja muut, 2024, 11:10). Luotettavuus tarkoittaa sitä, että käyttäjä voi arvioida datan soveltuvuutta päätöksentekoon (Dehghani, 2022, s. 36; Goedegebuure ja muut, 2024, 11:9–11:10). ISO/IEC 25012 -standardissa eheys määritellään datan tarkkuuden ja täydellisyyden säilymisinä, kun taas johdonmukaisuus viittaa datan ristiriidattomuuteen ja yhteensopivuuteen suhteessa muuhun dataan tietyssä käyttökontekstissa (ISO/IEC 25012, 2008, s. 10). Dolhopolov ja muut (2024, s. 6) liittävät nämä ominaisuudet hajautetussa data mesh

-ympäristössä laskennalliseen hallintaan, jonka avulla voidaan automatisoida esimerkiksi datan laadun, johdonmukaisuuden ja elinkaaren hallintaa.

Empiiriset havainnot viittaavat siihen, että hajautettu arkkitehtuuri voi parantaa myös järjestelmien luotettavuutta. Esimerkiksi Bodapati (2025, s. 771) raportoi keskimäärin noin 31 % parannuksen järjestelmien luotettavuudessa data mesh -malliin siirryttäessä. Tutkimus perustuu yhteensä 35 yrityksen tapaustutkimuksiin eri toimialoilta, kuten finanssi-, terveydenhuolto-, vähittäiskauppa- ja valmistussektoreilta. Tämä laajempi aineisto viittaa siihen, että havaitut hyödyt eivät ole täysin yksittäistapauksia.

Käytännön havainnot ovat kuitenkin kaksijakoisia. Bode ja muut (2024, s. 10–12) havaitsivat, että hajautettu omistajuus voi vahvistaa datan laatua ja saavutettavuutta, koska domain-tiimit tuntevat oman datansa käyttökontekstin paremmin. Samalla tutkimuksessa tunnistettiin haasteita liittyen metadatan laatuun, vastuunjakoon ja federoituun hallintaan, minkä seurauksena datatuotteiden laatu ja käytettävyys voivat vaihdella domainien välillä. Tällöin ongelmaksi muodostuu yksittäisten datatuotteiden laadun sijaan koko hajautetun ympäristön laadullinen epäyhtenäisyys.

Relevanssi ja ajantasaisuus ovat datan käytettävyyden toisiaan täydentäviä ulottuvuuksia (Wang ja Strong, 1996, s. 20). Data mesh -arkkitehtuurissa relevanssin voidaan katsoa korostuvan, koska vastuu datasta siirtyy domain-tiimeille, joilla on syvällisin ymmärrys datan liiketoimintakontekstista, käyttötapauksista ja kuluttajien tarpeista (Dehghani, 2022, s. 107; Bode ja muut, 2024, s. 10; Goedegebuure ja muut, 2024, 11:10–11:11). Tämän vuoksi datatuotteet voivat vastata aiempaa paremmin tiettyjen käyttäjäryhmien toistuviin tiedontarpeisiin ja tukea päätöksentekoa tarjoamalla tehtävään soveltuvaa dataa tarkoituksenmukaisella tarkkuudella ja ajantasaisuudella (Blohm ja muut, 2024, s. 643–644; Strong ja muut, 1997, s. 104; ks. tarkemmin luku 3.2).

Ajantasaisuus ei kuitenkaan ole hajautetussa ympäristössä yksiselitteinen hyöty. Eri domainit voivat päivittää ja hallita dataansa eri rytmissä sekä erilaisilla käytännöillä, mikä

voi johtaa epä johdon mukaisuuksiin hajautetussa dataekosysteemissä (Serra, 2024, s. 191; Tupper, 2011, s. 183). Tällöin eri datatuotteet voivat tarjota eri aikaan muodostettuja näkymiä samasta ilmiöstä, mikä vaikeuttaa yhtenäisen 'yhden totuuden lähteen' (single source of truth) ylläpitämistä koko organisaation tasolla (Dehghani, 2022, s. 26; Serra, 2024, s. 190, 192–193). Wang ja Strong (1996, s. 20–22) huomauttavatkin, että ajantasaisuutta tulee arvioida suhteessa käyttökontekstiin ja käyttäjän tehtävään. Tupper (2011, s. 183) kuvaa tilanteita, joissa päivitykset tapahtuvat asynkronisesti (eri aikaan, ilman yhteistä päivitysrytmiä), mikä johtaa epä johdon mukaisiin lukuihin eri järjestelmien välillä. Serra (2024, s. 195) toteaa, että useiden datatuotteiden ja niiden metadatan synkronointi federoidussa ympäristössä on teknisesti vaativaa ja voi heikentää datan eheyttä. Käytettävyyden kannalta tämä voi tarkoittaa, että data on kyllä saatavilla nopeasti, mutta vaikeuttaa datan tulkintaa ja eri lähteistä muodostettujen näkymien vertailua, erityisesti silloin kun metadata, käsitteet tai päivityskäytännöt vaihtelevat domainien välillä (Strong ja muut, 1997, s. 110; Bode ja muut, 2024, s. 11). Tämä on yksi niistä kohdista, joissa käytettävyys ja yhteentoimivuus alkavat limittyä, sillä ilman semanttista yhteensopivuutta ja yhteisiä standardeja tekninen ajantasaisuus ei takaa tiedon todellista hyödynnettävyyttä (Dolhopolov ja muut, 2024, s. 6; Goedegebuure ja muut, 2024, 11:9).

Datan laatuun liittyvät myös tarkkuus ja täydellisyys, jotka voivat hajautetussa ympäristössä vaihdella domainien välillä (Dehghani, 2022, s. 37, 154; Serra, 2024, s. 192). Virheellinen tai puutteellinen data voi heikentää datatuotteiden hyödynnettävyyttä, erityisesti silloin, kun eri lähteistä yhdistettävä tieto ei ole keskenään yhtenäistä tai noudata yhteisiä standardeja (Dehghani, 2022, s. 79–80; Goedegebuure ja muut, 2024, 10:9–11:9, 11:12; Serra, 2024, s. 191–192).

Kirjallisuudessa on raportoitu myös tilanteita, joissa data mesh -toteutukset eivät ole parantaneet datan laatua odotetusti. Esimerkiksi HelloFreshin tapauksessa hajautettu toimintamalli johti standardoinnin puutteeseen ja epäselvään omistajuuteen, mikä heikensi luottamusta dataan (Chee, 2021). Vastaavasti BlaBlaCar raportoi, että ratkaisut eivät

skaalautuneet ja vaativat jatkuvaa ylläpitoa, mikä vaikeutti datan hyödyntämistä (Guitouni, 2023).

3.1.4 Datatuote ja sen itsenäinen arvo

Dehghani (2022, s. 107) korostaa, että domain-tiimit ymmärtävät parhaiten oman datansa merkityksen ja käyttökontekstin, mikä mahdollistaa datan rakenteen ja sisällön suunnittelun käyttäjälähtöisesti. Kuten luvussa 2.1.2 todettiin, data mesh -mallissa datatuotteet suunnitellaan suoraan kulutettaviksi, mikä korostaa datan hyödyntämistä käyttäjän näkökulmasta (Dehghani, 2022, s. 29–45; Blohm ja muut, 2024, s. 644). Goedegebuure ja muut (2024, 11:7, 11:11) sekä Bode ja muut (2024, s. 10) liittävät domainlähtöisen vastuun siihen, että datatuotteet voidaan rakentaa paremmin liiketoiminnan käyttötarkoituksia ja käyttäjien tiedontarpeita vastaaviksi.

Datatuotteen käytettävyys ei riipu pelkästään datan sisällöstä tai teknisestä saatavuudesta, vaan erityisesti siitä, kuinka hyvin se on kuvattu ja kontekstualisoitu. Datatuotteen itsenäinen arvo tarkoittaa, että datatuotteen tulee olla hyödynnettävissä ilman jatkuvaa tukea sen tuottajatiimiltä (Dehghani, 2022, s. 81; Goedegebuure ja muut, 2024, 11:10). Tämä edellyttää, että datatuotteessa määritellyt elementit (kuten dokumentaatio ja konteksti, ks. luku 2.1.2) tukevat itsenäistä käyttöä, jotta kuluttaja voi ymmärtää datan semantiikan, rakenteen ja soveltuvuuden omiin tarpeisiinsa (Dehghani, 2022, s. 81, 84–85; Goedegebuure ja muut, 2024, 11:9).

Keskeinen muutos perinteisiin arkkitehtuureihin verrattuna on vastuun siirtyminen. Aiemmin data saattoi olla keskitetysti saatavilla ilman selkeää omistajaa, mutta data meshissä jokaisella datatuotteella on vastuullinen omistaja. Tämä pakottaa domain-tiimit arvioimaan, onko data aidosti hyödynnettävissä muille. Samalla vaatimustaso kasvaa merkittävästi, sillä kaikkien domainien ei ole realistista toimia täysimittaisina tuoteteiminä.

3.1.5 Natiivi saavutettavuus ja tekninen käytettävyys

Datatuotteen saavutettavuus eri käyttäjäryhmille ja erilaisille käyttötavoille on keskeinen osa datan käytettävyyttä (Dehghani, 2022, s. 38; Goedegebuure ja muut, 2024, 11:9). Natiivi saavutettavuus tarkoittaa, että datatuotteet ovat suoraan hyödynnettävissä ilman erillisiä muunnoksia tarjoamalla useita rajapintoja ja formaatteja, kuten SQL-kyselyitä, tiedostoja tai striimejä eri käyttäjäryhmien tarpeisiin (Dehghani, 2022, s. 39, 155; Goedegebuure ja muut, 2024, 11:9). Dehghani (2022, s. 30, 45) korostaa, että datatuotteiden tulisi olla suoraan kulutettavia eri käyttäjien tarpeisiin ilman raskaita välivaiheita, kun taas Bode ja muut (2024, s. 2) liittävät itsepalvelupohjaisen datanhallinnan datan nopeampaan saatavuuteen ja parempaan hyödynnettävyyteen. Tämä vähentää teknistä kitkaa ja nopeuttaa datan käyttöä hajautetussa ympäristössä. Natiivin saavutettavuuden merkitys korostuu erityisesti eri käyttäjä- ja sidosryhmien tarpeissa, joita tarkastellaan tarkemmin luvussa 3.2.

Teknisestä näkökulmasta datan käytettävyys data mesh -ympäristössä tarkoittaa sitä, kuinka nopeasti, luotettavasti ja ennakoitavasti datatuotteet ovat hyödynnettävissä. Tämä liittyy erityisesti kyselyiden vasteaikoihin, järjestelmän saatavuuteen sekä datan päivitysrytmiin ja tuoreuteen, joita hallitaan datatuotekohtaisina palvelutasotavoitteina (Dehghani, 2022, s. 37; Goedegebuure ja muut, 2024, 11:14). Wider ja muut (2023, s. 750) liittävät palvelutasotavoitteet myös datasopimuksiin ja niiden automaattiseen validointiin hajautetussa ympäristössä.

ISO/IEC 25012 -standardin näkökulmasta tekninen käytettävyys liittyy myös järjestelmäriippuvaisiin laatuominaisuuksiin, kuten saatavuuteen, saavutettavuuteen, siirrettävyyteen ja palautettavuuteen. Data mesh -ympäristössä nämä ominaisuudet korostuvat, koska dataa hyödynnetään useiden domainien, rajapintojen ja teknisten ympäristöjen välillä. Lisäksi skaalautuvuus on keskeinen tekniseen käytettävyyteen vaikuttava tekijä. Goedegebuure ja muut (2024, 11:16) toteavat, että data mesh mahdollistaa organisaation skaalautumisen horisontaalisesti (lisäämällä datalähteitä) ja vertikaalisesti (lisäämällä kuluttajia) ilman keskitettyjen tiimien muodostamia pullonkauloja.

Hajautetussa mallissa suorituskyky voi parantua, koska keskitetyt pullonkaulat poistuvat. Useita domaineja leikkaavien kyselyiden latenssin on raportoitu laskeneen keskimäärin noin 45 %. Samalla datan hyödyntämisen nopeus voi kasvaa, sillä aikaa oivallusten tuottamiseen (time-to-insight) on lyhentynyt jopa 43 % (Bodapati, 2025, s. 771).

Tulokset viittaavat siihen, että datatuoteajattelu voi nopeuttaa datan hyödyntämistä ja saatavuutta. Gartnerin vuoden 2024 kyselyyn viittaavan Oczkowski (2025) mukaan data mesh -arkkitehtuuria hyödyntävät organisaatiot raportoivat 40 % nopeamman time-to-marketin datatuotteille. Bodapati (2025, s. 771) puolestaan raportoi 37 % nopeammat kehityssykliä uusille analyyttisille tuotteille. Vaikutukset eivät kuitenkaan ole yksiselitteisiä, sillä nopeampi julkaisu ei automaattisesti tarkoita parempaa käytettävyyttä.

3.1.6 Turvallisuus ja yhteentoimivuus käytettävyyden ehtoina

Turvallisuus vaikuttaa käytettävyyteen kaksisuuntaisesti. Selkeät käyttöoikeudet, tietojen luokittelu ja automatisoitu pääsynhallinta voivat tukea datan hallittua hyödyntämistä hajautetussa ympäristössä, kun taas liian raskaat tai epäselvät tietoturvakäytännöt voivat hidastaa datan saatavuutta ja heikentää käytettävyyttä (Dehghani, 2022, s. 80, 87; Wider ja muut, 2023, s. 751). Tietoturvaa ja kyberturvallisuutta käsitellään laajemmin luvussa 3.4.

Hajautetussa data mesh -ympäristössä käytettävyyden keskeinen edellytys on yhteentoimivuus (Goedegebuure ja muut, 2024, 11:9). Yhteentoimivuutta tarkastellaan lisää luvussa 4.

3.2 Datan käytettävyyden käyttäjä- ja sidosryhmälähtöinen tarkastelu

Data ei ole käyttökelpoista, jos se ei ole riittävän tarkkaa, uskottavaa ja kontekstiin sopivaa (Wang ja Strong, 1997, s. 103–104). Konteksti määräytyy eri käyttäjien ja sidosryh-

mien mukaan, sillä eri toimijat hyödyntävät dataa erilaisissa käyttötarkoituksissa ja tilanteissa. Eri sidosryhmien huomioiminen on keskeinen datan käytettävyyden kannalta, koska käytettävyys ei ole datan staattinen ominaisuus, vaan toteutuu suhteessa sen käyttäjään ja käyttötarkoitukseen (Wang ja Strong, 1996, s. 6–9; Strong ja muut, 1997, s. 110). Käyttäjä- ja UX-näkökulmasta käytettävyys ei riipu vain datan teknisestä saatavuudesta, vaan myös siitä, kuinka vaivattomasti käyttäjä pystyy löytämään, ymmärtämään ja hyödyntämään dataa omassa työssään (Kashfi ja muut, 2016, s. 3, 26). Swartsin (2025) mukaan data koetaan aidosti käytettäväksi vasta silloin, kun käyttäjä pystyy yhdistämään sen kuvaamaan todelliseen ilmiöön. Data meshissä sidosryhmät jaetaan tyyppillisesti datan tuottajiin, kuluttajiin ja hallinnoijiin, joilla on erilaiset odotukset datatuotteiden ominaisuuksille (Goedegebuure ja muut, 2024, 11:19–11:20; Dehghani, 2022, s. 30, 77). Tässä työssä datan kuluttajiksi tarkastellaan erityisesti data-analyytikoita, data scientisteja ja liiketoimintakäyttäjiä, datan tuottajiksi kehittäjiä ja domain-omistajia sekä hallinnoijiksi alustatiimejä sekä hallinto- ja tietoturvatimijoita.

Tutkimukset osoittavat, että merkittävä este datan hyödyntämiselle on usein kuilu tuottajien tarjoaman metadatan ja kuluttajien tarvitseman kontekstiymmärryksen välillä (Bode ja muut, 2024, s. 11). Käyttäjät tarvitsevat datan tulkintaan riittävän kontekstin, kuten selkeät määritelmät, dokumentaation ja tiedon alkuperän, jotta data voidaan ymmärtää oikein ja hyödyntää tarkoituksenmukaisesti. Strong ja muut (1997) huomauttavat, että vaikka suuret tietomäärät olisivat teknisesti saatavilla, käyttäjät voivat kokea ne käytännössä saavuttamattomiksi, jos relevantin tiedon löytäminen tai tulkinta vaatii liikaa asiantuntemusta. Data mesh -arkkitehtuurissa domain-tiimien oletetaan tuntevan oman datansa käyttökontekstin paremmin kuin keskitetyn IT-tiimin, mikä voi tukea datatuotteiden kuvaamista käyttäjien tarpeita paremmin vastaavalla tavalla.

Data mesh -ympäristössä toimivilla tiimeillä on tutkimuksessa raportoitu 28 % korkeampi henkilöstön pysyvyysaste verrattuna perinteisiin keskitettyihin IT-rakenteisiin. Samassa tutkimuksessa liiketoiminnan sidosryhmien tyytyväisyyden datapalveluihin on todettu

kasvavan siirtymän myötä keskimäärin 42 % (35 yrityksen tapaustutkimus eri toimialoilta) (Bodapati, 2025, s. 771).

3.2.1 Data-analyttikot

Data-analyttikot kokevat datan käytettäväksi, kun se on helposti saavutettavissa heille tuttuun työkalujen, kuten SQL-kyselyiden, raportointiohjelmistojen tai taulukkolaskennan kautta. Dehghani (2022, s. 38) korostaa, että käytettävyys on suoraan sidoksissa siihen, kuinka vaivattomasti käyttäjä voi hyödyntää dataa omilla natiiveilla työkaluillaan. Goedegebuure ja muut (2024, 11:9) puolestaan toteavat, että datatuotteiden tulisi tukea useita rajapintoja ja tiedon esitysmuotoja, jotta eri käyttäjäryhmät voivat hyödyntää dataa omilla käyttötavoillaan.

Data mesh -toteutuksissa on raportoitu jopa 31 % parannuksia järjestelmien luotettavuudessa sekä 45 % vähennyksiä useita domaineja yhdistävien analyysikyselyiden viiveessä (Bodapati, 2025, s. 771), mikä voi nopeuttaa analyttikoiden tiedonhakuja ja analyysityötä hajautetussa ympäristössä. Lisäksi agenttipohjainen tekoäly (Agentic AI) mahdollistaa datan löytämisen ja tulkitsemisen luonnollisella kielellä keskustellen, mikä vähentää teknisen SQL-osaamisen tarvetta ja madaltaa kynnyksiä data-analyysiin (Tonnamelli ja muut, 2026, s. 1–4, 6). Tämä 'datademokraattinen' lähestymistapa lyhentää päätöksentekosyklejä ja vähentää riippuvuutta erikoistuneista teknisistä rooleista (Tonnamelli ja muut, 2026, s. 6).

3.2.2 Data scientistit

Data scientistit puolestaan painottavat käytettävyydessä pääsyä raakadataan tai kuratointeihin datamalleihin (valmiiksi käsiteltyihin ja laadunvarmistettuihin datakokonaisuuksiin), joita he voivat hyödyntää koneoppimismallien kouluttamisessa (Dehghani, 2022, s. 38–39; Goedegebuure ja muut, 2024, 11:9). Tyypillisesti lähes puolet data scientistien ajasta kuluu datan lataamiseen ja puhdistamiseen, mikä on merkittävä este tehokkuudelle (Dehghani, 2022, s. 29; Bode ja muut, 2024, s. 2). Aiemmassa luvussa käsitellyt

suorituskyky- ja luotettavuushyödyt voivat tukea myös data scientistien työskentelyä hajautetussa ympäristössä.

Data mesh -mallissa vastuu laadusta siirtyy ylävirtaan (lähde-domaineille), mikä on vastakohta perinteisille malleille, joissa keskitetty tiimi yritti korjata laatua jälkikäteen (Dehghani, 2022, s. 30; Goedegebuure ja muut, 2024, 11:7). Tämä vähentää merkittävästi data scientistien perinteisesti käyttämää aikaa datan puhdistamiseen ja 'kierrättämiseen', johon heillä kuluu usein jopa puolet työajastaan (Dehghani, 2022, s. 29). Tonarelli ja muut (2026, s. 1) toteavat, että koneellisesti suoritettavat datasopimukset voivat muuntaa datasopimusten ehdot automaattisesti teknisiksi validointi- ja hallintasäännöiksi. Näiden avulla voidaan parantaa datan rakenteellista eheyttä ja semanttista tarkkuutta esimerkiksi koneoppimismallien koulutuksessa.

3.2.3 Kehittäjät

Dehghani (2022, s. 48) sekä Goedegebuure ja muut (2024, 11:20) liittävät kehittäjien (data product developer) näkökulmasta käytettävyyden siihen, kuinka vaivattomasti datatuotteita voidaan rakentaa, testata, julkaista ja ylläpitää alustan tarjoamien automaatioiden ja standardien avulla. Alustan keskeinen tavoite on vähentää kehittäjiltä vaadittavaa teknistä erikoistumista ja poistaa rutiininomaista työtä, jolloin he voivat keskittyä liiketoiminta-arvon luomiseen (Dehghani, 2022, s. 53; Goedegebuure ja muut, 2024, 11:13).

Serra (2024, s. 175) kuvaa data mesh -mallin siirtävän kehittäjiä kohti laaja-alaisempaa generalist-roolia, jossa domain-tiimit vastaavat datatuotteiden rakentamisesta, julkaisemisesta ja ylläpidosta koko niiden elinkaaren ajan. Araújo Machado ja muut (2022, s. 267) liittävät tähän myös uudet roolit, kuten datatuotteen kehittäjä, joka vastaa datatuotteiden rakentamisesta, ylläpidosta ja tarjoilusta. Standardoidut rajapinnat ja API-standardit korvaavat satoja manuaalisia dataputkia, mikä vähentää teknistä kitkaa ja turhauttavaa rutiinityötä, sillä datatuotteet integroidaan toisiinsa suorien porttien ja koodina määriteltyjen sopimusten kautta. Bode ja muut (2024, s. 14–15) sekä Goedegebuure ja muut

(2024, 11:15) korostavat alustan roolia itsepalvelupohjaisena ympäristönä, kun taas Dolhopolov ja muut (2024, s. 6) liittävät automaattisen testauksen, sopimusvalidoinnin ja riippumattoman julkaisemisen hajautettujen datatuotteiden hallintaan. Alusta tukee tätä mahdollistamalla datatuotteiden automaattisen testauksen, validoinnin ja julkaisemisen ilman keskitettyjä pullonkauloja.

3.2.4 Domain-omistajat

Dehghani (2022, s. 45, 77) kuvaa domain-omistajat (data product owner) pitkäaikaisina datatuotteiden vastuuhenkilöinä, jotka vastaavat datatuotteiden omistajuuden jatkuvuudesta, käyttäjien tarpeiden huomioimisesta sekä datan jakamiseen liittyvästä hallinnasta. He määrittelevät datatuotteiden rakenteen, dokumentaation, metadatan ja palvelutasot, jotka mahdollistavat datan itsenäisen hyödyntämisen. Datan käytettävyyden on siten keskeinen onnistumisen mittari, jota voidaan arvioida esimerkiksi käyttäjätyytyväisyyden, datatuotteiden käyttöasteen ja Net Promoter Scoren (NPS) avulla (Dehghani, 2022, s. 45; Oczkowski, 2025). Domain-omistajien tehtävänä on lisäksi tasapainottaa eri käyttäjäryhmien ja domainien tarpeita federoidussa hallintamallissa, jossa päätöksenteko tapahtuu useiden toimijoiden yhteistyönä (Dehghani, 2022, s. 77).

Dehghani (2022, s. 29, 42) kuvaa data mesh -mallin siirtävän domain-omistajan roolia kohti tuotemaista ja käyttäjälähtöistä vastuuta, jossa datan kuluttajia kohdellaan datatuotteen asiakkaina ja tavoitteena on jopa 'hurmata' käyttäjäkokemus vähentämällä datan käytettävyyteen liittyviä kitkatekijöitä. Datatuotteen onnistumista arvioidaan sen tuottaman arvon, käyttäjätyytyväisyyden ja elinkaaren hallinnan kautta. Onnistumista mitataan esimerkiksi datan laadulla, datan kulutuksen läpimenoajalla ja käyttäjätyytyväisyyttä kuvaavalla Net Promoter Scorella (NPS) (Dehghani, 2022, s. 45). Oczkowski (2025) liittää samaan rooliin myös datatuotteen vision, tiekartan, priorisoinnin, käyttäjäpalautteen ja suorituskykymittareiden seurannan.

Empiiriset havainnot osoittavat, että datatuotteiden tarjoaminen muille domaineille voidaan kokea ylimääräiseksi työksi, jolle ei nähdä suoraa paikallista hyötyä, mikä voi heikentää dokumentaation ja metadatan laatua hajautetussa ympäristössä ja johtaa vähäisempään priorisointiin (Bode ja muut, 2024, s. 11–12). Swartsin (2025, s. 213) mukaan dokumentaatio ja metadata eivät kuitenkaan tue pelkästään datan teknistä käyttöä, vaan myös sitä, miten datan merkitys ja käyttötarkoitus välittyvät kuluttajille.

3.2.5 Alustatiimi

Dehghani (2022, s. 48, 57, 65) sekä Goedegebuure ja muut (2024, 11:13) tarkastelevat alustatiimin roolia erityisesti skaalautuvuuden, automaation ja teknisen kitkan vähentämisen näkökulmasta. Alustan tavoitteena on tarjota domain-tiimeille sellainen käyttökokemus, että datatuotteiden kehittäminen, testaaminen ja ylläpito onnistuvat ilman syvällistä infrastruktuurin hallinnan erikoisosaamista. Tämä toteutetaan abstrahoimalla matalan tason tekninen monimutkaisuus itsepalvelualustan taakse ja tarjoamalla valmiita infrastruktuuri- ja hallintapalveluita domain-tiimien käyttöön (Dehghani, 2022, s. 52, 65; Goedegebuure ja muut, 2024, 11:13). Alustatiimi ei rakenna varsinaisia datatuotteita, vaan tarjoaa yhteisen infrastruktuurin, alustapalvelut ja automatisoidut hallintamekanismit, joiden avulla domain-tiimit voivat kehittää ja ylläpitää datatuotteitaan itsenäisesti ilman merkittävää teknistä kuormaa tai keskitettyjä pullonkauloja (Goedegebuure ja muut, 2024, 11:13).

3.2.6 Liiketoimintakäyttäjät ja päätöksentekijät

Liiketoimintakäyttäjien näkökulmasta data mesh voi parantaa käytettävyyttä erityisesti nopeamman tiedonsaannin, vähäisempien pullonkaulojen ja päätöksenteon nopeutumisen kautta. SAP (2026) kuvaa data meshin vähentävän keskitettyjen datatiimien aiheuttamia pullonkauloja, nopeuttavan time-to-insight-aikaa ja tukevan päätöksentekoa. Bodapati (2025, s. 770–771) raportoi yhden globaalin pankkiorganisaation saavuttaneen 43 % lyhyemmän time-to-insight-ajan siirryttyään domain-lähtöiseen data-arkkitehtuu-

riin. Tonnarelli ja muut (2026, s. 1) liittävät luonnollisen kielen käsittelyn ja agenttipohjaisen tekoälyn datan jakamisen ja käytön helpottamiseen, koska AI-agentit voivat löytää, pyytää ja kysellä datatuotteita ilman hallintatason heikentämistä. Käyttäjien luottamusta ja käytettävyyden seuranta tukee lisäksi se, että datatuotteilla on nimetty datatuotteen omistaja, joka vastaa omistajuuden jatkuvuudesta sekä mittareista, kuten datan laadusta, datan kulutuksen läpimenoajasta ja käyttäjätyytyväisyydestä (Dehghani, 2022, s. 45; Bode ja muut, 2024, s. 4).

3.2.7 Hallinto- ja tietoturvatuojat

Hallinto- ja tietoturvatuojajille data mesh voi parantaa datan jäljitettävyyttä ja hallinnan läpinäkyvyyttä, koska datan jakaminen, käyttö ja kulutus voidaan nähdä keskitetyn katalogin kautta yli organisaation (Jain ja muut, 2021). Hajautettu malli voi kuitenkin lisätä hallinnan, koordinoinnin ja turvallisuusvastuiden monimutkaisuutta useiden itsenäisten domainien vuoksi (Bode ja muut, 2024, s. 10; Goedegebuure ja muut, 2024, 11:17). Hajautetun mallin tietoturvaan, hallintamekanismeihin ja kyberturvallisuusriskeihin liittyviä näkökulmia tarkastellaan tarkemmin seuraavassa luvussa.

3.3 Tietoturva ja kyberturvallisuus käytettävyyden näkökulmasta

ENISAn (2025, s. 8–9, 28) mukaan kalasteluhuijaus (phishing) muodosti yleisimmän tunnistetun alkuperäisen hyökkäysvektorin, kun taas haavoittuvuuksien hyväksikäyttö, tietomurrot ja datan ulosvienti olivat myös merkittäviä kyberturvallisuusuhkia. Raportissa korostuvat lisäksi kiristyshaittaohjelma -hyökkäysten (ransomware) yleisyys sekä tietomurtoihin johtaneet tunkeutumiset, sillä kiristyshaittaohjelma muodosti suurimman osan raportoiduista kyberrikollisuuteen liittyvistä tapauksista. Hajautetuissa data mesh -ympäristöissä tämä voi korostaa rajapintojen, identiteetin- ja pääsynhallinnan sekä segmentoinnin merkitystä, koska suuri määrä datatuotteita, integraatioita ja palvelurajapintoja voi kasvattaa hyökkäyspintaa ilman yhtenäisiä tietoturvakäytäntöjä.

3.3.1 Tietoturva osana datan käytettävyyttä

Data meshissä turvallisuus on yksi datatuotteiden 'ei-neuvoteltavista' ominaisuuksista ja tietoturva on keskeinen osa käytettävyyttä. Tietoturva vaikuttaa suoraan datan käytettävyyteen, koska datan tulee olla käyttäjille saavutettavissa, ymmärrettävää ja luotettavaa omassa käyttökontekstissaan (Wang ja Strong, 1996, s. 19, 32). ISO/IEC 25012 -standardissa käytettävyys kytkeytyy lisäksi järjestelmäriippuvaisiin ominaisuuksiin, kuten saataavuuteen, palautettavuuteen ja siirrettävyyteen (ISO/IEC 25012, 2008, s. 10). Wang ja Strong (1996, s. 20) korostavat, että käytettävän järjestelmän tulee olla samanaikaisesti saavutettava mutta turvallinen, jolloin käytettävyys ei tarkoita rajatonta pääsyä dataan vaan hallittua ja luotettavaa saavutettavuutta.

Data mesh -ympäristössä tietoturvan ja käytettävyyden yhteensovittaminen korostuu, sillä hajautettu omistajuus tarkoittaa, että datan käyttö riippuu useiden domainien käytännöistä, standardeista ja kontrollimekanismeista (Dehghani, 2022, s. 80, 156). Dehghani (2022, s. 156) korostaa, että pääsynhallinnan ja muiden politiikkojen määrittely voidaan tehdä keskitetysti, mutta niiden toteutus tapahtuu yksittäisten datatuotteiden tasolla policy-as-code -mallin avulla. Hajautettu malli voi kuitenkin lisätä siiloutumisen ja epäyhtenäisten hallintakäytäntöjen ja kontrollimekanismien riskiä, jos domainien käytännöt eroavat liikaa toisistaan (Dehghani, 2022, s. 80; Wider ja muut, 2023, s. 751). Strong ja muut (1997, s. 106) huomauttavat lisäksi, että tiukat käyttöoikeus- ja hyväksyntäprosessit voidaan kokea esteiksi datan saavutettavuudelle, mikä voi heikentää datan koettua arvoa ja luotettavuutta.

3.3.2 Hajautettu tietoturva ja hallintamallit

Data mesh -arkkitehtuurissa tietoturva ja hallinta perustuvat usein identiteetin- ja pääsynhallinnan, salauksen sekä automatisoitujen politiikkojen yhdistelmään, joita toteutetaan itsepalvelualustan ja hajautetun hallintamallin kautta (Wider ja muut, 2023, s. 748–

752). Wider ja muut (2023, s. 748–750) korostavat, että alustatasolla toteutettavat standardit, politiikat ja automatisoidut hallintamekanismit ovat keskeisiä hajautetun ympäristön yhdenmukaisuuden ja hallittavuuden kannalta.

Federoitu hallintamalli mahdollistaa yhteisten sääntöjen ja politiikkojen määrittelyn organisaatiotasolla, mutta niiden toteutus ja päätöksenteko tapahtuvat domain-kohtaisesti. Tämä lisää joustavuutta ja autonomiaa, mutta edellyttää yhteisiä standardeja ja hallintakäytäntöjä, jotta hajautettu ympäristö säilyy hallittavana ja yhteentoimivana (Goedegebuure ja muut, 2024, 11:9; Wider ja muut, 2023, s. 748–752). Dolhopolov ja muut (2024, s. 5–6) korostavat lisäksi käytön seurannan ja auditoinnin merkitystä hajautetussa ympäristössä. Datatuotteiden käyttöä, käyttäjäidentiteettejä, käyttöajankohtia ja käyttömalleja seuraamalla voidaan havaita poikkeavaa toimintaa ja tukea tietoturvapoliitikkojen valvontaa.

Rose ja muut (2020, s. 3–6) korostavaa Zero Trust -lähestymistavassa jatkuvaa todentamista, vähäisimmän oikeuden periaatetta ja implisiittisten luottamusalueiden pienentämistä, joita Deghani (2022, s. 13–14, 80) liittää hajautettujen data mesh -ympäristöjen ja monipilviympäristöjen turvalliseen hallintaan sekä datatuotteiden päätepestekohtaiseen pääsynhallintaan. Dolhopolov ja muut (2024, s. 16–17) kuvaavat lisäksi hienojakoisen pääsynhallinnan ja datan peittäminen (data masking) merkitystä hajautetuissa ympäristöissä. Esimerkiksi Fybrik-alusta mahdollistaa turvallisen datan käytön useissa ympäristöissä hyödyntämällä resurssikohtaista pääsynhallintaa ja arkaluonteisen tiedon maskausta.

Datasopimukset voivat lisäksi tukea turvallista datan jakamista määrittelemällä datatuotteiden rajapintoja, laatuvaatimuksia, metatietoja ja pääsynhallinnan periaatteita yhteinäisesti yli domain-rajojen (Bodapati, 2025, s. 773–774). Lisäksi automaatio, tekoäly ja tekoälyagentit voivat tukea hajautetun ympäristön hallintaa esimerkiksi valvomalla datatuotteiden käyttöä, politiikkojen noudattamista, käyttöoikeuksia ja reaaliaikaista auditointia (Tonnarelli ja muut, 2026, s. 1).

3.3.3 Kyberturvallisuusriskit ja käytettävyyden välinen tasapaino

Goedegebuure ja muut (2024, 11:17) toteavat hajautetun data mesh -arkkitehtuurin lisäävän järjestelmän kokonaiskompleksisuutta, koska se edellyttää useiden riippumattomien yksiköiden koordinoitua. Serra (2024, s. 192–193) puolestaan huomauttaa, että rajapintojen ja datatuotteiden määrän kasvu voi samalla kasvattaa järjestelmän hyökkäyspintaa. Dehghani (2022, s. 8, 157) korostaa datatuotteiden välistä eristystä ja riippumattomaa toimintaa, joiden tavoitteena on estää yksittäisten muutosten tai häiriöiden leviäminen muihin datatuotteisiin. Myös ENISA (2025, s. 63) painottaa segmentoinnin ja pääsynhallinnan merkitystä uhkien rajaamisessa, mikä voi auttaa rajaamaan uhkien leviämistä hajautetuissa ympäristöissä. Tällöin yksittäinen tietomurto ei välttämättä vaaranna koko järjestelmää, vaan vaikutukset voivat rajoittua yksittäisiin datatuotteisiin tai domeineihin. Verizonin (2025, s. 61) mukaan väärinkonfiguraatiot ja suojaamattomat tietokannat ovat lisäksi yleisiä tietomurtojen syitä, mikä korostaa yhdenmukaisten hallinta- ja tietoturvakäytäntöjen merkitystä hajautetuissa ympäristöissä.

Dehghani (2022, s. 80) toteaa, että hajautetussa ympäristössä datan saavutettavuuden, yhtenäisten käytäntöjen ja pääsynhallinnan välillä on löydettävä tasapaino, sillä globaalit standardit tukevat datatuotteiden käytettävyyttä ja yhtenäistä käyttökokemusta, mutta pääsynhallinnan toteutus tapahtuu paikallisesti yksittäisten datatuotteiden ja palveluiden tasolla. Rose ja muut (2020, s. 49) huomauttavat lisäksi, että monimutkaiset turvallisuuskäytännöt ja tunnistautumisosprosessit voivat heikentää käyttäjäkokemusta ja aiheuttaa niin kutsuttua turvallisuusväsymys -ilmiötä (security fatigue). Tämän vuoksi itsepalvelualustojen tavoitteena on abstrahointi (teknisen monimutkaisuuden piilottaminen käyttäjältä) infrastruktuurin ja tietoturvan monimutkaisuutta käyttäjiltä automatisoitujen alustapalveluiden avulla (Dehghani, 2022, s. 48, 65; Goedegebuure ja muut, 2024, 11:13–11:14).

3.4 Datan käytettävyyden edistävät ja estävät tekijät

3.4.1 Edistävät tekijät

Vastuun siirtyminen datan lähteelle voi edistää datan laatua ja käytettävyyttä, koska datan tuottavat tiimit tuntevat datan kontekstin paremmin ja pystyvät ylläpitämään sitä jatkuvasti (Dehghani, 2022, s. 30, 77). Tässä mallissa vastuu laadusta siirtyy 'ylävirtaan', jolloin ongelmia pyritään korjataamaan mahdollisimman lähellä niiden syntypaikkaa (Dehghani, 2022, s. 30). Dehghani (2022, s. 87) korostaa lisäksi automaation merkitystä virheiden ja poikkeamien havaitsemisessa jo datatuotteen rakentamisvaiheessa, mikä voi tukea datan laadun jatkuvaa hallintaa hajautetussa ympäristössä.

Wider ja muut (2023, s. 746, 750–751) mainitsevat, että hajautetussa ympäristössä hallinta ei voi perustua pelkkiin paikallisiin käytäntöihin, vaan alustalla on oltava mekanismeja, joilla datatuotteet voidaan julkaista, löytää ja hallita yhdenmukaisesti. Tämä edellyttää tasapainoa domain-autonomian ja yhteisten standardien välillä, jossa itsepalvelualusta tukee yhdenmukaisuutta tarjoamalla automatisoituja ja helppokäyttöisiä toimintamalleja domaineille (Wider ja muut, 2023, s. 750–751; Dehghani, 2022, s. 87).

Dolhopolov ja muut (2024, s. 2–3) korostavat, että metadatan hallintajärjestelmä on keskeinen osa mitä tahansa big data -alustaa ja olennainen osa federoitua hallintaa, sillä sen avulla voidaan hallita, yhdistää, ymmärtää ja kontrolloida olemassa olevia dataresursseja. Datakatalogi kokoaa yhteen datatuotteisiin liittyvää metadataa, kuten omistajuutta, käyttötarkoituksia, riippuvuuksia ja muuta tuotteisiin liittyvää tietoa, mikä tukee datan löydettävyyttä ja auttaa käyttäjiä hahmottamaan saatavilla olevia datatuotteita (Goedegebuure ja muut, 2024, 11:8, 11:14; Dolhopolov ja muut, 2024, s. 4). Metadata ei siten ole pelkkä tekninen kuvaus, vaan keskeinen mekanismi, joka tukee datatuotteiden semanttista tulkintaa, datan alkuperän jäljitettävyyttä sekä datan ymmärrettävyyden ja luotettavuuden arviointia hajautetussa ympäristössä (Dolhopolov ja muut, 2024, s. 2–5; Goedegebuure ja muut, 2024, 11:9). Dehghani (2022, s. 25, 80) sekä Dolhopolov ja muut (2024, s. 3, 5) korostavat, että semanttinen yhteentoimivuus tukee käytettävyyttä, koska

domainit voivat mallintaa samoja käsitteitä eri tavoin omiin käyttötarkoituksiinsa. Tällöin datan yhdistäminen edellyttää riittävää semanttista kuvausta, yhteisiä tunnisteita ja mallien välistä mappauksia.

Dehghani (2022, s. 35–36) sekä Goedegebuure ja muut (2024, 11:9) sanovat, että datatuotteiden käytettävyyttä tukevat selkeä dokumentaatio, käyttörajapintojen kuvaukset ja datatuotekohtaiset palvelutasotavoitteet (SLO:t), joilla määritellään esimerkiksi datan ajantasaisuus, saatavuus ja laatu. Käytettävyyden kannalta keskeistä on, että käyttäjä pystyy löytämään, ymmärtämään ja hyödyntämään datatuotetta ilman syvällistä domain-tietämystä tai jatkuvaa riippuvuutta tuottajatiimistä (Dehghani, 2022, s. 35; Wider ja muut, 2023, s. 750). Bode ja muut (2024, s. 11) sekä Bodapati (2025, s. 773) puolestaan korostavat dokumentaation, metadatan, omistajuuden ja standardoitujen rajapintojen merkitystä datatuotteiden onnistuneessa hyödyntämisessä. Lisäksi palvelutasotavoitteet tekevät datan laadusta läpinäkyvää ja ennakoitavaa, ja ne voivat toimia osana data-sopimusta (Dehghani, 2022, s. 36; Goedegebuure ja muut, 2024, 11:9).

Goedegebuure ja muut (2024, 11:10) korostavat, että datan laadun ja eheyden automaattinen valvonta voi edistää käytettävyyttä, koska virheitä ja poikkeamia voidaan havaita ennen kuin ne vaikuttavat datan hyödyntämiseen. Tätä tuetaan esimerkiksi CI/CD-palveluiden, Infrastructure as Code (IaC) -käytäntöjen sekä policy-as-code -ratkaisujen avulla, jotka mahdollistavat datatuotteiden automaattisen testaamisen, validoinnin, politiikkojen valvonnan ja julkaisemisen ilman keskitettyjä pullonkauloja (Dolhopolov ja muut, 2024, s. 6; Goedegebuure ja muut, 2024, 11:13–11:15; Tonnarelli ja muut, 2026, s. 1). Bodapati (2025, s. 770) sekä Goedegebuure ja muut (2024, 11:13) liittävät itsepalvelualustat ja automaation erityisesti datan hyödyntämisen nopeuttamiseen, manuaalisen työn vähentämiseen sekä hajautetun ympäristön yhdenmukaisuuden tukemiseen.

Dehghani (2022, s. 38–39) sekä Goedegebuure ja muut (2024, 11:9) korostavat, että teknisestä näkökulmasta käytettävyyttä edistävät datatuotteiden natiivi saavutettavuus ja monipuoliset rajapinnat. Datatuotteet voidaan tarjota useissa eri muodoissa, kuten SQL-

kyselyinä, columnar-tiedostoina tai viestijonoina (streams), mikä mahdollistaa niiden hyödyntämisen eri käyttäjäryhmien tarpeisiin ilman manuaalisia muunnoksia (Dehghani, 2022, s. 38). Datatuotteiden yksilölliset ja ohjelmallisesti saavutettavat rajapinnat mahdollistavat lisäksi datan hyödyntämisen eri järjestelmien ja käyttörajapintojen kautta (Goedegebuure ja muut, 2024, 11:9). Lisäksi suorituskykyyn liittyvät tekijät, kuten kyselyiden vasteajat, järjestelmän saatavuus ja datan ajantasaisuus (freshness), vaikuttavat suoraan siihen, kuinka nopeasti ja luotettavasti data on hyödynnettävissä (Dehghani, 2022, s. 37; Bodapati, 2025, s. 773).

Bode ja muut (2024, s. 4) sekä Dehghani (2022, s. 3) toteavat, että käytettävyys rakentuu myös organisatorisesti ja sosiaalisesti. Vaikka tekniset mekanismit mahdollistavat datan hyödyntämisen, käytännössä se edellyttää tiedon jakamista, yhteistyötä ja yhteisten käytäntöjen kehittymistä domainien välillä (Bode ja muut, 2024, s. 11). Bode ja muut (2024, s. 13) korostavat lisäksi koulutusten, yhteisöllisten tapahtumien ja hackathonien merkitystä datatuotteiden kehittämisen tukena. Kuluttajien kokemusten ja parhaiden käytäntöjen jakaminen esimerkiksi yhteisöllisten tapahtumien, kuten hackathonien tai 'pitch-illojen' kautta, voi parantaa datatuotteiden käytettävyttä ja rakentaa tarvittavaa momentumia organisaation sisällä (Bode ja muut, 2024, s. 13).

Dehghani (2022, s. 48, 85) sekä Wider ja muut (2023, s. 750) toteavat, että käytettävyttä tukevat yhdenmukaiset pääsynhallintakäytännöt, automatisoitu politiikkojen valvonta sekä itsepalvelualustat, jotka vähentävät teknistä kitkaa ja mahdollistavat turvallisen datan hyödyntämisen hajautetussa ympäristössä. Itsepalvelualusta hyödyntää informaation piilottamista (information hiding), jolloin asiantuntijat voivat saada datasta arvoa ilman syvällistä domain-tietämystä (Wider ja muut, 2023, s. 750). Laskennallinen hallintomalli, kuten 'Data Product MCP', vähentää data-analyysin teknistä taakkaa suorittamalla datasopimusten ehtoja ja sääntöjä automaattisesti ilman manuaalisia hyväksyntävaiheita (Tonnarelli ja muut, 2026, s. 1).

Bode ja muut (2024, s. 4), Blohm ja muut (2024, s. 646) sekä Serra (2024, s. 182) kertovat, että data fabric voi täydentää data mesh -arkkitehtuuria yhdistämällä hajautettuja tietolähteitä teknisen integraatiokerroksen ja virtualisoinnin avulla, mutta sen vaikutus datan käytettävyyteen on usein rajallinen, koska se ei määrittele datan hajautettua omistajuutta eikä siten ratkaise semanttiseen ymmärrettävyyteen tai laatuun liittyviä organisatorisia haasteita samalla tavalla kuin data mesh. Virtualisointi voi lisäksi helpottaa hajautettujen tietolähteiden hyödyntämistä tarjoamalla yhtenäisen pääsyn heterogeenisiin tietolähteisiin ilman erillisiä datakopioita (Serra, 2024, s. 83).

Käytännön teknologiavalinnat voivat myös tukea datatuotteiden käytettävyyttä, kun ne tarjoavat esimerkiksi skeemanhallintaa, metadatan hallintaa, versionhallintaa ja skaalautuvaa analytiikkaa. Serra (2024, s. 239–240) kuvaa esimerkiksi Databricksin ja Snowflaken tukevan data mesh -periaatteita tarjoamalla ominaisuuksia, jotka voivat vahvistaa domainien autonomiaa, datan laatua, löydettävyyttä ja käytettävyyttä.

3.4.2 Estävät tekijät

Hajautettu arkkitehtuuri voi johtaa tilanteeseen, jossa eri domainit mallintavat ja määrittelevät samoja käsitteitä eri tavoin omiin käyttötarkoituksiinsa, mikä voi heikentää datan ymmärrettävyyttä ja vaikeuttaa datan yhdistämistä eri lähteistä, jos yhteiset tunnisteet, semanttiset kuvaukset ja mallien väliset mappaukset ovat puutteellisia (Dehghani, 2022, s. 25, 80). Bode ja muut (2024, s. 11) huomauttavat lisäksi, että datatuotteiden tarjoama metadata ei aina vastaa hyödyntäjien tarpeita, mikä voi johtaa pitkittyneisiin hakuaikeihin ja epäselviin odotuksiin datatuotteiden sisällöstä ja käyttötarkoituksista.

Pelkkä itsepalvelu ei myöskään riitä parantamaan käytettävyyttä, jos datatuotteet eivät ole helposti löydettävissä ja ymmärrettävissä, jolloin tilanteena on tekninen saatavuus ilman kykyä hyödyntää dataa tehokkaasti liiketoiminta-arvon luomiseen (Goedegebuure ja muut, 2024, 11:17–11:18; Tonnarelli ja muut, 2026, s. 1). Puutteellinen metadatan hallinta ja datan alkuperän (data lineage) seuranta voivat lisäksi vaikeuttaa datan tulkintaa ja heikentää järjestelmän hallittavuutta erityisesti federoiduissa hauissa, joissa

metadataa ja riippuvuuksia joudutaan kokoamaan useista hajautetuista datatuotteista (Serra, 2024, s. 195). Dehghani (2022, s. 157) huomauttaa lisäksi, että liian yleisluontoinen metadatan käsittely voi johtaa monimutkaisiin ja keskitettyihin rakenteisiin, minkä vuoksi metadata tulisi määritellä osana datatuotteiden rajapinta- ja datasopimuksia.

Datan tarjoaminen muille domaineille koetaan usein lisätyöksi, jolle ei nähdä suoraa paikallista hyötyä, mikä voi heikentää datatuotteiden laatua, ylläpitoa ja saatavuutta (Bode ja muut, 2024, s. 11; Goedegebuure ja muut, 2024, 11:12). Ilman selkeää omistajuutta vastuu datan laadusta ja ylläpidosta voi jäädä epäselväksi, mikä lisää riskiä datan laatu- ja eheysongelmille hajautetussa ympäristössä (Bode ja muut, 2024, s. 2). Dehghani (2022, s. 45) korostaa tämän vuoksi data product owner -roolin merkitystä datan laadun, ylläpidon ja käyttäjätyytyväisyyden jatkuvuuden varmistamisessa. Lisäksi ilman asianmukaisia kannustimia domain-tiimit saattavat priorisoida operatiiviset tehtävänsä datatuotteiden kehittämisen edelle, mikä voi heikentää datatuotteiden jatkuvaa kehittämistä ja käytettävyyttä (Bode ja muut, 2024, s. 11; Goedegebuure ja muut, 2024, 11:12).

Käytettävyyttä voivat heikentää myös epäyhtenäiset käyttöoikeuskäytännöt, monimutkaiset tunnistautumisprosessit sekä liian tiukat tietoturvakontrollit, jotka lisäävät käyttäjien kognitiivista kuormaa ja hidastavat datan hyödyntämistä hajautetussa ympäristössä (Dehghani, 2022, s. 59, 80; Rose ja muut, 2020, s. 49; Verizon, 2025).

Kirjallisuudessa on raportoitu tilanteita, joissa data mesh -toteutukset eivät ole parantaneet käytettävyyttä odotetusti. Esimerkiksi HelloFreshin kokemusten mukaan hajautettu toimintamalli johti standardoinnin puutteeseen, epäselvään omistajuuteen ja jatkuvaan operatiiviseen kuormaan, mikä heikensi luottamusta dataan ja vaikeutti sen hyödyntämistä (Chee, 2021). Vastaavasti BlaBlaCar raportoi, että varhaiset ratkaisut eivät skaalautuneet, koska ne olivat liian sidottuja yksittäisiin käyttötapauksiin ja edellyttivät jatku-

vaa kontekstikohtaista ylläpitoa. Lisäksi datan hyödyntäminen hidastui, kun käyttäjät joutuivat turvautumaan tikettipohjaisiin prosesseihin eivätkä luottaneet uuteen dataan (Guitouni, 2023).

3.4.3 Tekoäly käytettävyyden tekijänä

Tekoäly on integroitumassa osaksi data meshin evoluutiota, jossa agenttipohjaiset tekoälyratkaisut ja suuret kielimallit mahdollistavat datan löytämisen ja hyödyntämisen luonnollisen kielen avulla (Tonnarelli ja muut, 2026, s. 1). Perinteiset koontinäytöt (dashboards) korvautuvat 'keskustelevalle analytiikalla', jossa agenttipohjainen tekoäly ja kielimallit mahdollistavat datan löytämisen ja tulkitsemisen luonnollisella kielellä (Tonnarelli ja muut, 2026, s. 1, 8). Tekoälyagentit asettavat kuitenkin yhteentoimivuudelle uusia vaatimuksia, sillä ne edellyttävät koneluettavia datasopimuksia ja merkitystietoista metadataa, jotta dataa voidaan tulkita automaattisesti (Tonnarelli ja muut, 2026, s. 3, 7).

Tähän liittyen esimerkiksi 'Model Context Protocol' (MCP) mahdollistaa datatuotteiden löytämisen ja hyödyntämisen keskustelunomaisesti ilman syvää teknistä osaamista (Tonnarelli ja muut, 2026, s. 1, 5). Lisäksi tekoälyagentit voivat vähentää datan hyödyntämisen teknistä kuormaa automatisoimalla datatuotteiden löytämistä, käyttöoikeuksien hallintaa ja kyselyiden muodostamista luonnollisen kielen avulla. Tonnarelli ja muut (2026, s. 1, 7) kuvaavat MCP-ratkaisua, jossa tekoälyagentit voivat hyödyntää datatuotteita ja valvoa datasopimusten noudattamista reaaliaikaisesti ilman manuaalisia hyväksyntäprosesseja. Automatisoitu datan havainnointi voi auttaa tunnistamaan datan laatuun liittyviä ongelmia varhaisessa vaiheessa (Gates, 2025). SAPin (2026) mukaan koneoppimisalgoritmien hyödyntäminen voi lisäksi parantaa datatuotteiden laatua ja hyödyllisyyttä ajan myötä.

3.5 Datan käytettävyyden arviointi mittareiden avulla

Tässä alaluvussa datan käytettävyyttä tarkastellaan mittareiden avulla, jolloin luvussa 3 esitetyt näkökulmat muunnetaan arvioitaviksi ominaisuuksiksi. Goedegebuure ja muut

(2024, 11:12) toteavat, että federoidun hallintotiimin on kerättävä mittareita arvioidakseen datatuotteiden arvoa ja koko data mesh -ympäristön tilaa. Mittareiden avulla voidaan myös tehdä päätöksiä tuotteiden kehittämisestä tai poistamisesta.

Eri mittarit voivat mitata yksittäisiä käytettävyyden osa-alueita tai useita osa-alueita samanaikaisesti. Esimerkiksi completeness- ja accuracy-mittarit kuvaavat datan laatua ja luotettavuutta, suorituskyky ja SLA-toteutuminen mittaavat teknistä käytettävyyttä, ja adoption rate sekä time-to-insight kuvaavat datatuotteiden hyödyntämistä käytännössä. Lisäksi käyttäjätyytyväisyyttä mittaavat indikaattorit, kuten NPS, kuvaavat käyttäjäkokeumusta, kun taas metadataan ja dokumentaatioon liittyvät mittarit tukevat datan käytännön hyödynnettävyyden arviointia. Lisäksi MacDonald (2024) liittyy datan ajantasaisuuden mittaamiseen esimerkiksi data freshness-, data latency- ja data accessibility -mittareita, kun taas Alation (2026) korostaa time-to-trust- ja coverage of critical data elements -mittareiden merkitystä datan hallinnan ja käytettävyyden arvioinnissa.

Taulukossa 4 esitetään käytettävyyden mittareita sekä kuvataan, mitä ne mittaavat ja mihin käytettävyyden osa-alueeseen ne ensisijaisesti liittyvät. Mittarit havainnollistavat, mitä käytettävyyden osa-alueita voidaan mitata sekä jäsentävät käytettävyyden eri osa-alueita. Ne toimivat osa-alueiden jäsentäjinä sekä mahdollisena pohjana jatkotutkimukselle. Mittarit on koottu ja sovellettu useista työssä käytetyistä lähteistä.

Taulukko 4. Datan käytettävyyden mittareita data mesh -ympäristössä.

Mittari	Osa-alue	Mitä mittaa
Datan katalogikattavuus (Data Catalog Coverage)	Löydettävyys ja osoitettavuus	Datatuotteiden löydettävyyttä ja dokumentoinnin kattavuutta
Metadatan laatuindeksi (Metadata Score)	Ymmärrettävyys ja itsekuvailevuus	Metadatan sisällöllistä laatua, luotettavuutta ja hyödyllisyyttä
Metadatan täydellisyys (Metadata Completeness)	Ymmärrettävyys ja itsekuvailevuus	Metadatan riittävyttä ja ymmärrettävyyttä
Datan alkuperän kattavuus (Data Lineage Coverage)	Luotettavuus, ymmärrettävyys ja itsekuvailevuus	Kuinka kattavasti datan alkuperä, muunnokset ja käsitteilyvaiheet on dokumentoitu
Datan laatu (Data Quality Score, Data Completeness, Data Accuracy)	Laatu, luotettavuus ja eheys	Datan laatua, täydellisyyttä ja virheettömyyttä

Datan löytämiseen kuluva aika (Time-to-Discovery)	Löydettävyyys ja saavutettavuus	Aikaa, joka kuluu relevantin datan löytämiseen
Muutosten epäonnistumisaste (Change Fail Ratio)	Datatuotteen ylläpidettävyys ja operatiivinen laatu	Kuinka suuri osa datatuotteen tehtävistä muutoksista aiheuttaa virheitä tai laadun heikkenemistä
Luottamukseen kuluva aika (Time-to-Trust)	Laatu, luotettavuus ja eheys	Aikaa, jossa uusi data muuttuu luotettavaksi ja tuotantokelpoiseksi
Kattavuus kriittiselle datalle (Coverage of Critical Data Elements, CDE)	Laatu, luotettavuus ja eheys	Kuinka suuri osa liiketoimintakriittisestä datasta on hallinnan, laadunvalvonnan ja metadatan piirissä
Datan ajantasaisuus (Data Freshness, Data access latency)	Natiivi saavutettavuus ja tekninen käytettävyys	Datan ajantasaisuutta
Palvelutasosopimuksen noudattamisaste (SLA Compliance)	Natiivi saavutettavuus ja tekninen käytettävyys	Sovittujen palvelutasojen toteutumista
Alustan automaatio (Platform Automation Coverage / Domain autonomy index)	Natiivi saavutettavuus ja datatuotteen itsenäinen arvo	Automatisoinnin tasoa
Suorituskyky ja operatiivinen toimivuus (Query Performance, Data Product Incidents, Cross-domain latency, Query success rate)	Natiivi saavutettavuus ja tekninen käytettävyys	Järjestelmän suorituskykyä ja häiriöitä
Dataputkien epäonnistumisaste (Pipeline failure rate)	Natiivi saavutettavuus ja tekninen käytettävyys	Virheiden määrä, voi viitata yhteensopivuusongelmiin
Käyttö ja uudelleenkäyttö (Adoption Rate, Cross-domain Reuse, Network Interconnections)	Käyttäjä- ja sidosryhmät	Datatuotteiden käyttöä, uudelleenkäyttöä ja datatuotteiden välisten yhteyksien määrää
Aika tehokkaaseen käyttöön (Time-to-Insight)	Käyttäjä- ja sidosryhmät	Aikaa datapyynnöstä käyttökelpoiseen tulokseen
Julkaisunopeus (Time-to-Market)	Käyttäjä- ja sidosryhmät	Aikaa ideasta datatuotteen julkaisuun
Tiimin pysyvyys (Retention Rate)	Käyttäjä-, sidosryhmä- ja sosiotekninen näkökulma	Datatiimien pysyvyyttä ja domain-osaamisen jatkuvuutta
Käyttäjäkokemus (Net Promoter Score, Stakeholder Satisfaction)	Käyttäjä- ja sidosryhmät ja sosiotekninen näkökulma	Käyttäjäytyvyyttä ja kokemusta
Innovointi (Innovation Quotient, Innovation Velocity)	Käyttäjä- ja sidosryhmät	Uusien datatuotteiden ja keksintöjen määrää
Pääsyyn kuluva aika (Time-to-Access)	Saavutettavuus ja prosessin sujuvuus	Aikaa ja vaivaa, joka kuluu käyttöoikeuksien ja valtuutusten saamiseen
Datatuotteen muutosnopeus (Lead Time to Change)	Datatuotteen ylläpidettävyys ja ketteryys	Aikaa, joka kuluu olemassa olevan datatuotteen muuttamiseen tai päivittämiseen

Kustannustehokkuus (Cost per Query)	Liiketoimintamittari	Käytön kustannustehokkuus
Arvon realisoituminen (Value Realization Index)	Liiketoimintamittari	Datatuotteiden tuottamaa liiketoiminta-arvoa

4 Datan yhteentoimivuus data mesh –arkkitehtuurissa

Luvussa 4.1 analysoidaan data mesh -arkkitehtuurin vaikutuksia datan yhteentoimivuuteen (TK2). Luvussa 4.2 käsitellään yhteentoimivuutta edistäviä mekanismeja ja ratkaisuja, kun taas luvussa 4.3 tarkastellaan yhteentoimivuutta estäviä haasteita (TK3). Luvussa 4.4 arvioidaan vaikutuksia empiiristen havaintojen ja tapausesimerkkien kautta. Luvun aikana tarkastellaan myös organisaation kypsyyden ja hallintamallin merkitystä (TK4), ja luvussa 4.5 esitetään mittareita yhteentoimivuuden arviointiin.

Datan yhteentoimivuutta voidaan tarkastella osittain samojen teoreettisten näkökulmien kautta kuin datan käytettävyyttä luvussa 3, mutta painotus siirtyy datan hyödyntämisestä sen yhdistettävyyteen ja siirrettävyyteen eri järjestelmien ja domainien välillä. Tässä luvussa ensisijaisena jäsenyydenä toimii Euroopan unionin (2017) yhteentoimivuuksiin viitekehys, jossa yhteentoimivuus jaetaan tekniseen, semanttiseen, organisatoriseen ja hallinnolliseen tasoon. Tässä tutkimuksessa EIF:n legal interoperability -tasoa tarkastellaan vain hallinnollisen yhteentoimivuuden näkökulmasta.

Dehghani (2022, s. 6) sekä Goedegebuure ja muut (2024, 11:29) tarkastelevat data meshia sosio-teknisenä lähestymistapana tai paradigmana, jossa ihmiset, organisatoriset rakenteet ja tekniset ratkaisut vaikuttavat toisiinsa. Tässä näkökulmassa yhteentoimivuus nähdään myös datatuotteen 'ei-neuvoteltavana' ominaisuutena (Dehghani, 2022, s. 33; Goedegebuure ja muut, 2024, 11:9). Lisäksi yhteentoimivuutta voidaan tarkastella laatu-, käyttäjä-/UX- ja sosioteknisestä näkökulmasta, joissa korostuvat datan tulkittavuus, käytön sujuvuus sekä merkityksen säilyminen eri järjestelmien ja käyttökontekstien välillä. Mittaripohjainen arviointi puolestaan konkretisoi nämä näkökulmat arvioitaviksi ilmiöiksi sekä myös jäsentää yhteentoimivuuden eri osa-alueita. Taulukko 5 esittää tutkimuksessa käytetyt yhteentoimivuuden keskeiset lähestymistavat, viitekehykset ja arviointitavat. Näitä näkökulmia ja viitekehyksiä käytetään luvussa 3 esitetyn lähestymistavan tavoin eri osa-alueiden jäsentämiseen.

Taulukko 5. Datan yhteentoimivuuden näkökulmat, viitekehykset ja arviointitavat.

Lähestymistapa/ viitekehys	Määritelmä	Fokus	Rajoitteet
Laatunäkökulma sekä käyttäjä-/UX- ja sosiotekninen näkökulma (Wang ja Strong, 1996; Kashfi ja muut, 2016; Swarts, 2025)	Yhteentoimivuus liittyy datan johdonmukaiseen esitystapaan, siirrettävyyteen, tulkittavuuteen ja käytön sujuvuuteen siten, että data voidaan käsitellä, ymmärtää ja hyödyntää myös alkuperäisen kontekstin ulkopuolella ja eri järjestelmien välillä	Datan esitystapa, tulkittavuus, käyttäjäkokemus ja käytön sujuvuus	Ei kuvaa tarkasti teknisiä integraatioita tai hallintamalleja, ja osa näkökulmista on käyttäjäkohtaisia sekä vaikeasti mitattavia
ISO-standardi (ISO/IEC 25012, 2008)	Yhteentoimivuus liittyy järjestelmäriippuvaiseen datan laatuun, jossa korostuvat datan siirrettävyys, saavutettavuus ja kyky tulla käsitellyksi eri järjestelmissä	Järjestelmät ja data	Ei huomioi laajemmin organisatorisia tai käyttäjäkokemukseen liittyviä asioita
Data mesh (Dehghani, 2022)	Yhteentoimivuus on datatuotteen 'ei-neuvoteltava' ominaisuus, jota tuetaan yhteisillä standardeilla ja hallintamalleilla	Datatuote hajauteudessa ympäristössä	Ei standardoitu, ei tarjoa yhtenäistä toteutusmallia
Mittareihin perustuva arviointi	Yhteentoimivuutta arvioidaan esimerkiksi ristiinkäytön määrällä, standardien noudattamisella ja integraatioiden tehokkuudella	Mitattavat vaikutukset ja toteuma	Mittaa vaikutuksia ja toteumaa, ei suoraan yhteentoimivuutta kokonaisuutena
Yhteentoimivuusviitekehys (EIF) (Euroopan unioni, 2017)	Yhteentoimivuus jäsenyy neljään tasoon: tekninen, semanttinen, organisatorinen ja hallinnollinen	Järjestelmät, organisaatio ja hallinto	Ei tarjoa suoraa arviointimittaristoa, vaan toimii jäsentävänä viitekehysenä

4.1 Yhteentoimivuuden tasot data mesh -arkkitehtuurissa

Tässä alaluvussa yhteentoimivuutta tarkastellaan ensisijaisesti Euroopan unionin (2017) yhteentoimivuusviitekehysten kautta. Sen avulla yhteentoimivuus jäsenyy neljään tasoon: tekniseen, semanttiseen, organisatoriseen ja hallinnolliseen (Euroopan unioni, 2017, s. 21). Tekninen taso keskittyy rajapintoihin ja turvalliseen tiedonsiirtoon, kun taas semanttinen taso varmistaa, että tiedon merkitys ja formaatti ymmärretään samalla ta-

voin kaikissa osapuolissa (Euroopan unioni, 2017, s. 29–30). Organisatorinen yhteentöimivuus tarkoittaa liiketoimintaprosessien ja vastuiden yhdenmukaistamista, ja hallinnollinen taso kattaa lainsäädännön sekä yhteiset hallintopolitiikat (Euroopan unioni, 2017, s. 28). Vaikka tämä nelitasoinen malli on alun perin kehitetty yleiseksi yhteentöimivuusviitekehukseksi, se tarjoaa myös data mesh -arkkitehtuurin tarkasteluun hyödyllisen jäsennyksen, koska siinä erotellaan teknisiä, semanttisia, organisatorisia ja hallinnollisia ulottuvuuksia, jotka liittyvät myös hajautettuun datan hallintaan. Lisäksi yhteentöimivuutta tarkastellaan osittain myös laatu näkökulmasta, jossa korostuvat tiedon johdonmukainen esitystapa, tulkittavuus ja saavutettavuus eri käyttäjille ja järjestelmille (Wang ja Strong, 1996, s. 6, 16).

4.1.1 Semanttinen yhteentöimivuus

Semanttinen yhteentöimivuus tarkoittaa sitä, että tiedon merkitys ja sisältö ymmärretään samalla tavalla eri järjestelmissä ja organisaatioyksiköissä. Se kattaa sekä semanttiset että syntaktiset näkökulmat, kuten sanastot, skeemat, datan väliset suhteet ja tiedon esitysmuodon (Euroopan unioni, 2017, s. 29). Data mesh -arkkitehtuurissa semanttinen yhteentöimivuus on erityisen tärkeää, koska eri domainit voivat mallintaa samoja käsitteitä omista käyttökonteksteistaan käsin (Dehghani, 2022, s. 25). Dehghani (2022, s. 80) toteaa lisäksi, että datatuotteiden semanttisten ja syntaktisten skeemojen määrittelystä tulee koko mesh-ympäristöä koskeva yhteinen kysymys. Tämän vuoksi datatuotteiden kuvaaminen standardoidun metadatan ja yhteisten liiketoimintasanastojen avulla tukee tiedon yksiselitteistä tulkintaa eri domainien välillä (Goedegebuure ja muut, 2024, 11:9, 11:11–11:12). Dolhopolov ja muut (2024, s. 2, 5) puolestaan korostavat metadatan hallinnan ja semanttisen rikastamisen merkitystä datan ymmärtämisessä, yhdistämisessä ja hyödyntämisessä hajautetussa ympäristössä.

4.1.2 Organisatorinen yhteentöimivuus

Organisatorinen yhteentöimivuus tarkoittaa liiketoimintaprosessien, vastuiden ja organisaatorakenteiden yhdenmukaistamista siten, että yhteistyö eri yksiköiden välillä on

mahdollista (Euroopan unioni, 2017, s. 28). Data mesh -arkkitehtuurissa tämä näkyy vas-
tuun siirtymisenä keskitetystä IT-organisaatiosta domain-tiimeille, jotka vastaavat omien
datatuotteidensa laadusta, ylläpidosta ja jakamisesta (Dehghani, 2022, s. 30; Wider ja
muut, 2023, s. 750). Tämä edellyttää uusia rooleja ja toimintamalleja, kuten datatuote-
omistajia, jotka vastaavat datatuotteiden tarjoamisesta, hallinnasta ja yhteentoimivuu-
desta omissa domaineissaan sekä varmistavat, että datatuotteet vastaavat käyttäjien tar-
peita (Goedegebuure ja muut, 2024, 11:10, 11:19–11:20). Lisäksi federoitu hallintatiimi
koordinoi koko mesh-ympäristöä koskevia standardeja ja hallintakäytäntöjä (Goede-
gebuure ja muut, 2024, 11:20). Bode ja muut (2024, s. 4) puolestaan tarkastelevat data
meshia sosioteknisenä ja hajautettuna lähestymistapana, jossa domainit omistavat tuot-
tamansa datan ja hyödyntävät domain-osaamistaan datan laadun parantamisessa. Bode
ja muut (2024, s. 11) sekä Goedegebuure ja muut (2024, 11:17) huomauttavat kuitenkin,
että hajautettu toimintamalli voi lisätä koordinoinnin ja hallinnan monimutkaisuutta,
mikä voi vaikeuttaa datan jakamista ja yhteistyötä domainien välillä.

4.1.3 Tekninen yhteentoimisuus

Euroopan unioni (2017, s. 30) määrittelee teknisen yhteentoimivuuden rajapintoihin,
standardeihin ja viestintäprotokolliin liittyvänä yhteentoimivuutena, joka mahdollistaa
tiedonsiirron eri järjestelmien välillä. Goedegebuure ja muut (2024, 11:9) kuvaavat tek-
nisen yhteentoimivuuden toteutuvan ohjelmallisesti saavutettavien rajapintojen kautta,
kuten API-, SQL- ja suoratoistorajapintoina, kun taas Wider ja muut (2023, s. 750) liittävät
nämä osaksi hajautetun data mesh -alustan datatuotteita ja niiden hallintaa. Itsepalve-
lualusta toimii teknisenä perustana, joka automatisoi näiden standardien toteutuksen
(standards as code) ja varmistaa, että datatuotteet ovat natiivisti saavutettavia ilman ma-
nuaalisia integraatoratkaisuja (Goedegebuure ja muut, 2024, 11:13; Dehghani, 2022, s.
84). Rose ja muut 2020, s. 48) sekä Goedegebuure ja muut (2024, 11:17) huomauttavat,
että riittämätön standardointi sekä riippuvuus suljetuista rajapinnoista voivat lisätä in-
tegraatioiden monimutkaisuutta, vaikeuttaa yhteentoimivuutta ja kasvattaa toimittaja-
lukkojen riskiä.

4.1.4 Hallinnollinen yhteentoimivuus

Hallinnollinen yhteentoimivuus liittyy yhteisten sääntöjen, politiikkojen ja hallintamallien yhteensovittamiseen siten, että eri domainit voivat toimia hajautetussa ympäristössä (Euroopan unioni, 2017, s. 27). Data mesh -arkkitehtuurissa tätä tuetaan federoitun laskennallisen hallinnan avulla, jossa yhteiset politiikat, standardit ja tavoitteet yhdistetään domain-kohtaiseen autonomiaan (Dolhopolov ja muut, 2024, s. 4; Deghani, 2022, s. 146). Deghani (2022, s. 146, 156) kuvaa, kuinka laskennalliset politiikat voidaan upottaa datatuotteisiin policy-as-code -lähestymistavan avulla siten, että esimerkiksi pääsynhallinta-, salaus-, laatu- ja sääntöjenmukaisuusvaatimuksia voidaan valvoa automaattisesti datan käsittelyn eri vaiheissa. Tämä voi tukea esimerkiksi PII-tiedon (personally identifiable information, henkilötiedot) suojaamista ja sääntelyvaatimusten noudattamista hajautetussa ympäristössä (Deghani, 2022, s. 85). Wider ja muut (2023, s. 750) puolestaan liittävät hajautetun pääsynhallinnan domain-tiimien vastuuseen, jossa domainit päättävät omien datatuotteidensa käyttöoikeuksista yhteisten politiikkojen puitteissa. Lisäksi alustatasolla voidaan tarjota automatisoituja palveluita esimerkiksi tietojen anonymisointiin, politiikkojen valvontaan ja sääntöjenmukaisuuden toteuttamiseen (Goedegebuure ja muut, 2024, 11:14).

4.2 Yhteentoimivuutta edistävät tekijät ja ratkaisut data mesh -arkkitehtuurissa

Tässä luvussa yhteentoimivuutta tarkastellaan organisatoristen, hallinnollisten, teknisten ja semanttisten mekanismien kautta. Organisatorisiin mekanismeihin kuuluvat esimerkiksi vastuunjako ja domainien välinen koordinointi, hallinnollisiin mekanismeihin data-sopimukset, federoitu hallintamalli ja yhteiset standardit, teknisiin mekanismeihin rajapinnat, integraatoratkaisut ja virtualisointi sekä semanttisiin mekanismeihin metadatan rikastaminen, tietämysgraafit ja ontologioiden kohdistaminen.

Goedegebuure ja muut (2024, 11:8–11:11) korostavat avoimien standardien, datasopimusten, itsepalvelualustojen, datakatalogien ja federoidun hallintomallin merkitystä datan yhteentoimivuuden mahdollistajina. Näiden tehtävänä on varmistaa, että datatuotteet ovat rakenteeltaan, merkitykseltään ja käyttöehdoiltaan yhteensopivia yli domainrajojen (Goedegebuure ja muut, 2024, 11:8–11:9; Tonnarelli ja muut, 2025, s. 2). Data-sopimukset ovat erityisen keskeisiä yhteentoimivuuden kannalta, koska ne määrittelevät yhtenäisesti datatuotteiden rakenteen, semantiikan, käyttöehdot ja laadulliset vaatimukset domainien välillä (Goedegebuure ja muut, 2024, 11:7, 11:10; Tonnarelli ja muut, 2025, s. 2). Tämä voi vähentää erillisten muunnosten ja manuaalisen tulkinnan tarvetta datan hyödyntämisessä (Tonnarelli ja muut, 2025, s. 1–2). Datakatalogit puolestaan tukevat datatuotteiden löydettävyyttä sekä mahdollistavat datatuotteiden metatietojen, skeemojen ja käyttörajapintojen tarkastelun, kun taas itsepalvelualusta mahdollistaa yhteisten rajapintojen ja standardien hyödyntämisen hajautetussa ympäristössä (Goedegebuure ja muut, 2024, 11:10, 11:13; Dehghani, 2022, s. 84). Wider ja muut (2023, s. 749–750) sekä Goedegebuure ja muut (2024, 11:11) korostavat lisäksi, että federoitu hallintomalli pyrkii yhdistämään yhteiset standardit ja politiikat domainien autonomiaan hajautetussa ympäristössä.

Tekoälyä voidaan hyödyntää federoidun hallinnan tukena, sillä järjestelmät, kuten Data Product MCP, mahdollistavat tekoälyagenttien suorittaman datatuotteiden löytämisen, kyselyn, semanttisen löydettävyyden sekä datasopimusten reaaliaikaisen valvonnan ilman hallinnan tason heikkenemistä (Tonnarelli ja muut, 2025, s. 1). Tekoälyagentteihin ja suuriin kielimalleihin perustuvat ratkaisut voivat lisäksi helpottaa data-analyysiä luonnollisen kielen kautta sekä tunnistaa ja estää sääntöjenvastaiset kyselyt jo ennen niiden suorittamista, mikä tukee politiikkojen ja käyttöehtojen automaattista valvontaa (Tonnarelli ja muut, 2025, s. 1, 5). Tonnarelli ja muut (2025, s. 5) arvioivat lisäksi, että tällaiset ratkaisut voivat helpottaa datatuotteiden välistä uudelleenkäyttöä ja tukea data mesh -mallin mukaista domain-lähtöistä datan hallintaa.

Teknologisella tasolla yhteentoimivuutta tukevat standardoidut rajapinnat, tapahtumavirrät, semanttiset kuvaukset sekä tietämysgraafit, joiden avulla datatuotteiden merkityksiä, suhteita ja metatietoja voidaan kuvata yhtenäisemmin (Goedegebuure ja muut, 2024, 11:9; Dolhopolov ja muut, 2024, s. 5). Lisäksi automaatio, kuten policy-as-code ja CI/CD-palvelut, mahdollistaa hallintasääntöjen, datasopimusten sekä dataputkien ja rajapintojen automaattisen validoinnin, testaamisen ja julkaisemisen suoraan datatuotteissa (Goedegebuure ja muut, 2024, 11:13–11:15; Dolhopolov ja muut, 2024, s. 6).

Goedegebuure ja muut (2024, 11:17) toteavat, että datan kopiointi voi lisätä hallinta- ja integraatiohaasteita hajautetussa ympäristössä, minkä vuoksi ratkaisuksi ehdotetaan datan virtualisointia ja nollakopiointia (zero-copy access). Serra (2024, s. 82, 155) määrittelee datan virtualisoinnin lähestymistapana, jossa dataa voidaan käyttää ja yhdistää eri lähteistä loogisen näkymän kautta ilman datan fyysistä siirtämistä tai kopiointia. Virtualisointi voi vähentää ETL-prosessien, tallennuksen ja hallinnan kustannuksia sekä tukea datan ajantasaisuutta ja turvallisuutta, koska data säilyy alkuperäisessä lähteessään (Serra, 2024, s. 83). Lisäksi yhtenäinen SQL-rajapinta voi helpottaa datan hyödyntämistä eri järjestelmistä riippumatta (Serra, 2024, s. 85). Dolhopolov ja muut (2024, s. 11) kuvaavat lisäksi Fybric-alustan mahdollistavan datan virtualisoinnin ja siirtämisen eri ympäristöjen välillä ilman datan kopiointia, mikä voi vähentää viivettä ja tukea tehokasta datanhallintaa hajautetuissa ympäristöissä.

Data fabric -ratkaisut tukevat yhteentoimivuutta erityisesti teknisellä tasolla tarjoamalla hajautettujen tietolähteiden päälle yhtenäisen integraatio- ja abstrahointikerroksen (piilottaa taustajärjestelmien tekniset erot ja tarjoaa yhtenäisen pääsyn dataan) (Blohm ja muut, 2024, s. 644; Serra, 2024, s. 154–155). Data fabric hyödyntää aktiivista metadataa, tietämysgraafeja ja semanttista virtualisointia integraatioiden ja hallinnan automatisointiin (Blohm ja muut, 2024, s. 644, 646). Virtualisointi voi lisäksi vähentää datan fyysisen siirtämisen tarvetta ja helpottaa datan yhdistämistä useista lähteistä reaaliaikaisesti (Serra, 2024, s. 154–155). Bode ja muut (2024, s. 3–4) kuitenkin korostavat, että data

fabric on ensisijaisesti tekninen arkkitehtuurimalli, kun taas data mesh sisältää myös organisatorisia ja sosiaalisia näkökulmia, kuten hajautetun omistajuuden ja domain-vas- tuut. Tämän vuoksi lähteissä data fabric ja data mesh esitetään usein toisiaan täydentä- vinä lähestymistapoina eikä kilpailevina ratkaisuinä (Blohm ja muut, 2024, s. 650; Serra, 2024, s. 182–183).

Hooshmand ja muut (2022) sekä Dolhopolov ja muut (2024, s. 5) korostavat, että onto- logioiden hyödyntäminen ja kohdistaminen tukevat semanttista yhteentoimivuutta yh- denmukaistamalla eri domainien käsitteitä ja tietomalleja. Dolhopolov ja muut (2024, s. 6) sekä Goedegebuure ja muut (2024, 11:28) toteavat lisäksi, että tietämysgraafit ja se- manttisen verkon teknologiat voivat helpottaa datan yhdistämistä yli järjestelmärajojen tukemalla suhteiden tunnistamista ja automaattista linkitystä. Tutkimukset osoittavat, että semanttinen rikastaminen on muita hallinnan osa-alueita vaativampi tehtävä, ja käy- tettyjen menetelmien välillä on eroja: osa järjestelmistä tyytyy perustason asiansanoituk- seen, kun taas edistyneemmät hyödyntävät tietämysgraafeja tai pilvipalveluita suhteiden automaattiseen löytämiseen (Dolhopolov ja muut, 2024, s. 14–15). Tätä tarkasteleva tut- kimus ontologioiden kohdistamisesta viittaa siihen, että eri menetelmien välillä on mer- kittäviä eroja sekä laadussa että tehokkuudessa, mikä on merkityksellistä hajautetuissa data-arkkitehtuureissa, joissa pyritään parantamaan järjestelmien välistä yhteentoimi- vuutta ja automatisoimaan tiedon yhdistelyä (Abd Nikooie Pour ja muut, 2025, s. 1, 28, 29).

Kokonaisuudessaan data mesh -ympäristön yhteentoimivuus rakentuu teknisten ratkai- sujen, yhteisten standardien, federoidun hallinnan sekä semanttisen yhdenmukaisuuden yhdistelmästä, jossa sekä organisatoriset että teknologiset mekanismit vaikuttavat data- tuotteiden tehokkaaseen yhdisteltävyyteen ja hyödyntämiseen yli domain-rajojen.

4.3 Yhteentoimivuuden haasteet domainien välillä

Tässä alaluvussa yhteentoimivuutta tarkastellaan erityisesti teknisiä, semanttisia, organisatorisia ja hallinnollisia haasteita, jotka voivat estää datatuotteiden tehokkaan yhdistämisen hajautetussa ympäristössä.

Hajautettu arkkitehtuuri tuo mukanaan haasteita. Semanttinen hajautuminen voi johtaa tilanteeseen, jossa domainit määrittelevät samat käsitteet eri tavoin, mikä aiheuttaa ongelmia datan yhdistämisen kanssa (Serra, 2024, s. 192; Dolhopolov ja muut, 2024, s. 6). Tekninen heterogeenisyys (erilaiset teknologiat ja toteutustavat) ja erot tiimien osaamistasoissa lisäävät integraatiokustannuksia ja voivat johtaa epäyhtenäisiin toteutuksiin, mikä vaikeuttaa datan kuluttamista (Serra, 2024, s. 192; Goedegebuure ja muut, 2024, 11:17). Serra (2024, s. 195–196) liittää näitä riskejä myös data quantum -konseptiin, jossa metadata on saatavilla vain datatuotteen sisältä, mikä voi vaikeuttaa datan alkuperän seuranta ja metadatahakuja federoidussa ympäristössä ilman monimutkaisia kyselyitä tai jatkuvaa synkronointia. Tämä voi heikentää metadatan yhtenäistä hyödyntämistä ja vaikeuttaa datatuotteiden hallintaa koko hajautetussa ympäristössä.

Hajautettu autonomia voi samalla, kun se parantaa domain-tason tehokkuutta ja paikallista päätöksentekoa, lisätä organisatorisen fragmentaation riskiä (Bode ja muut, 2024, s. 11; Goedegebuure ja muut, 2024, 11:17–11:18). Tällöin domain-tiimit optimoivat omia ratkaisujaan paikallisesti, jolloin yhteinen semantiikka, uudelleenkäytettävyys ja datatuotteiden välinen koordinaatio heikkenevät (Serra, 2024, s. 191; Dolhopolov ja muut, 2024, s. 6). Tiimit saattavat kokea muiden domainien palvelemisen ylimääräisenä työnä, jolloin tuotteiden ylläpito ja jakaminen voivat jäädä toissijaiseksi suhteessa domainin omiin tavoitteisiin (Bode ja muut, 2024, s. 11). Seurauksena datatuotteet voivat toimia hyvin oman domaininsa sisällä, mutta niiden yhdistettävyys, standardienmukaisuus ja yhteentoimivuus voivat heikentyä hajautetussa ympäristössä (Dolhopolov ja muut, 2024, s. 6; Goedegebuure ja muut, 2024, 11:17; Serra, 2024, s. 192).

Lisäksi datan kopiointi ja päällekkäiset toteutukset voivat kasvaa, jos domainit hyödyntävät dataa omiin käyttötarkoituksiinsa olemassa olevien datatuotteiden uudelleenkäyttöön sijaan (Goedegebuure ja muut, 2024, 11:17; Serra, 2024, s. 193). Metadatan heikko laatu ja puutteellinen dokumentaatio heikentävät datatuotteiden ymmärrettävyyttä ja käytettävyyttä (Bode ja muut, 2024, s. 11; Dolhopolov ja muut, 2024, s. 14), mikä voi samalla vaikeuttaa datatuotteiden välistä yhteentoimivuutta, koska dataa ei pystytä tulkitsemaan, yhdistämään tai hyödyntämään yhdenmukaisesti hajautetussa ympäristössä.

Myös datan alkuperän seuraaminen ja metadatahaut voivat hidastaa järjestelmän suorituskykyä federoidussa ympäristössä (Serra, 2024, s. 195). Näin ollen yhteentoimivuuden haasteet ovat samanaikaisesti semanttisia, teknisiä ja organisatorisia, ja ne kytkeytyvät toisiinsa hajautetussa ympäristössä, korostaen tarvetta federoidulle hallintomallille (Dehghani, 2022, s. 89; Serra, 2024, s. 192).

4.4 Empiiriset havainnot ja tapausesimerkit data mesh -toteutuksissa

Empiirisiä havaintoja ja tapausesimerkkejä tarkastellaan erityisesti siitä näkökulmasta, miten data mesh -periaatteet ovat vaikuttaneet datan yhdistettävyyteen, uudelleenkäyttöön ja hallintaan eri organisaatioissa. Havainnot osoittavat myös, ettei data meshille ole muodostunut yhtä yleispätevää teknistä toteutusmallia. Organisaatiot, kuten Zalando ja Netflix, ovat rakentaneet omiin tarpeisiinsa räätälöityjä ratkaisuja: Zalando hyödyntää 'bring your own bucket' -mallia ja standardoitua metadatalaitosta. 'Bring your own bucket' tarkoittaa, että domain-tiimit voivat käyttää omia Simple Storage Service-tallennusalueitaan mutta liittävät ne yhteiseen data mesh -infrastruktuuriin ja metadatahallintaan. Netflix taas painottaa operatiivisen monimutkaisuuden vähentämistä GraphQL- ja Apache Iceberg -teknologioilla (Araújo Machado ja muut, 2022, s. 268–269). Tämä viittaa siihen, että yhteentoimivuuden onnistuminen riippuu enemmän organisaation kypsyydestä, hallintamallista ja standardointikäytännöistä kuin yhdestä valmiista teknisestä mallista.

4.4.1 Onnistuneet toteutukset ja havaitut hyödyt

Onnistuneissa toteutuksissa data mesh parantaa datan uudelleenkäyttöä ja vähentää duplikaatiota. Esimerkiksi JPMorgan Chasen tapauksessa datan jakaminen korvasi kopiointin, mikä vähensi tallennuskustannuksia, minimoi datan epäjohdonmukaisuuksia ja paransi näkyvyyttä datan alkuperään mesh-luettelon avulla (Jain ja muut, 2021).

Hajautettu malli mahdollistaa nopeamman datan hyödyntämisen, koska domain-tiimit voivat tuottaa ja käyttää dataa ilman keskitettyjä pullonkauloja (Goedegebuure ja muut, 2024, 11:16). Aika oivallusten saavuttamiseen (time-to-insight) on raportoitu lyhentyneen merkittävästi, esimerkiksi eräs globaali pankki saavutti tässä 43 % lyhennyksen (Bodapati, 2025, s. 771). Lisäksi Siemensin data mesh -käyttöönottoa kuvaavassa tapausesimerkissä BI- ja AI-datatuotteiden kuvataan hyödyntävän yhteistä data stackia ja jaettuja datasettejä. Datatuotteita säilytettiin saavutettavissa sijainneissa, jolloin useat tiimit pystyivät hyödyntämään niitä uudelleen (dbt Labs, 2024).

AKTIN-infrastruktuurissa kyselyjen onnistumisaste oli keskimäärin 86,7 % ja kokonaisvirhetaso 1,1 %, mikä tukee kyselyiden onnistumisasteen ja operatiivisen luotettavuuden käyttöä yhteentoimivuuden epäsuorina indikaattoreina (Bienzeisler ja muut, 2025, s. 22–23). Vaikka AKTIN-infrastruktuuri ei ole suoraan data mesh -arkkitehtuuri, se perustuu federoituun toimintamalliin, jossa data säilytetään paikallisissa solmuissa ja tietoa hyödynnetään hajautetusti ilman keskitettyä tietovarastoa.

Ontologioiden automaattista yhdistämistä tarkastelevassa OAEI 2025 -tutkimuksessa arvioitiin algoritmeja, joiden tavoitteena on tunnistaa vastaavuuksia eri tietomallien ja käsitteiden välillä automaattisesti. Parhaat menetelmät saavuttivat yli 0,9 F-luvun (F-luku on tarkkuuden ja kattavuuden yhdistelmämittari), mikä viittaa siihen, että semanttinen kohdistaminen voidaan toteuttaa erittäin tarkasti myös hajautetuissa ympäristöissä (Abd Nikooie Pour ja muut, 2025). Tämä on merkityksellistä data mesh -arkkitehtuurissa, jossa eri domainit voivat käyttää toisistaan poikkeavia tietomalleja ja käsitteitä. Tulokset viit-

taavat siihen, että semanttisen yhteentoimivuuden automatisointi voi tukea datatuotteiden yhdistämistä ja vähentää manuaalisen integraatiotyön tarvetta hajautetuissa data-arkkitehtuureissa.

Data mesh -arkkitehtuuriin perustuvissa toteutuksissa on raportoitu myös konkreettisia parannuksia datan hyödyntämisessä ja skaalautuvuudessa. Esimerkiksi HelloFresh hallitsee data mesh -mallillaan noin 1,2 petatavua dataa, ja datamäärä kasvaa 7,5 % kuukausittain. Arkkitehtuurin avulla HelloFresh on kyennyt lyhentämään raakadatan muuntamisen liiketoimintatiedoksi päivistä minuutteihin (Kohoutek, 2023). Amazonin Finance Automation -ratkaisussa data mesh -arkkitehtuuri mahdollistaa noin 850 löydettävän tietoa-aineiston ja yli 300 kuratoidun datatuotteen hallinnan, mikä tukee datan löydettävyyttä ja hyödyntämistä yli domain-rajojen (Arora ja muut, 2023). Ratkaisussa datatuotteiden standardointi, itsepalvelukyvykkyudet ja keskitetty datakatalogi mahdollistavat datan hyödyntämisen eri analytiikkapalveluissa ilman vahvaa riippuvuutta keskitetystä integraatiotiimistä (Arora ja muut, 2023), mikä viittaa yhteentoimivuuden onnistuneeseen skaalautumiseen laajassa hajautetussa ympäristössä. Gartnerin vuoden 2024 kyselyyn viittaavan Oczkowski (2025) mukaan data mesh -arkkitehtuuria hyödyntävät organisaatiot raportoivat jopa 35 % vähennyksen datan hallintamalliin liittyvässä hallinnollisessa kuormassa verrattuna keskitettyihin malleihin.

Lisäksi Covestro raportoi siirtymisensä data mesh -arkkitehtuuriin johtaneen 70 % vähennykseen uusien datatuotteiden markkinoille tuloajassa. Organisaatio hyödynsi standardoituja datamalleja, Infrastructure as Code (IaC) -käytäntöjä, liiketoimintasanastoja sekä automatisoitua metadatan hallintaa, joiden tavoitteena oli helpottaa datan jakamista ja hyödyntämistä yli domain-rajojen (Janssen ja muut, 2025). Vaikka markkinoille tuloajan lyheneminen ei mittaa yhteentoimivuutta suoraan, tulos saattaa viitata siihen, että standardointi, datasopimukset ja yhteensopivat rajapinnat voivat vähentää integraatiotyötä ja helpottaa datatuotteiden yhdistämistä hajautetussa ympäristössä.

4.4.2 Epäonnistuneet toteutukset, tulosten tulkinta ja rajoitteet

Tässä alaluvussa tarkastellaan aiemmin esitettyihin onnistuneisiin toteutuksiin sekä haavittuihin haasteisiin ja epäonnistumisiin liittyviä tulkintoja ja rajoitteita yhteentoimivuuden näkökulmasta.

Tulokset eivät kuitenkaan osoita, että data mesh parantaisi yhteentoimivuutta automaattisesti. Kirjallisuudessa on raportoitu myös tapauksia, joissa data mesh -toteutusten odotetut hyödyt eivät ole toteutuneet ainakaan alkuvaiheessa. HelloFreshin kokemusten mukaan ensimmäinen yritys hajautetusta mallista, jossa vastuut olivat osittain hajautettuja, johti standardoinnin puutteeseen, epäselvään omistajuuteen ja jatkuvaan 'tulipalojen sammuttamiseen', mikä heikensi luottamusta dataan ja alustan toimivuutta (Chee, 2021).

Lisäksi Winter ja Hackl (2025, s. 1, 6, 11,) havaitsivat kymmentä suurta organisaatiota tarkastelevassa tutkimuksessaan, että vaikka organisaatiot olivat omaksuneet data mesh -periaatteita, käytännön toteutukset jäivät usein keskitetyimmiksi kuin mallin teoriassa oletetaan, eikä datan omistajuuden hajautuminen useinkaan toteutunut täysimääräisesti. Tämä viittaa siihen, että domainien autonomian toteuttaminen voi olla haastavaa säännellyissä ympäristöissä, joissa keskitetty hallinto nähdään tärkeänä standardoinnin, yhteentoimivuuden ja riskienhallinnan kannalta.

Havainnot kokonaisuudessaan viittaavat siihen, että data mesh voi tukea yhteentoimivuutta erityisesti silloin, kun datatuotteet ovat standardoituja, löydettäviä ja hallinnollisesti yhteensovitettuja. Esimerkiksi datan uudelleenkäytön lisääntyminen, datatuotteiden määrän kasvu ja nopeutunut datan hyödyntäminen viittaavat siihen, että hajautettu arkkitehtuuri voi vähentää keskitettyihin malleihin liittyviä pullonkauloja ja helpottaa datan käyttöä yli domain-rajojen. Samat piirteet, jotka lisäävät domainien autonomiaa, voivat myös heikentää yhteistä semantiikkaa, lisätä integraatioiden monimutkaisuutta ja hajauttaa hallintaa. Tämän vuoksi empiirisiä hyötyjä ei tulisi tulkita pelkästään data mesh -periaatteiden seurauksiksi.

Näin ollen kirjallisuus ja havainnot tukevat tulkintaa, jonka mukaan data mesh voi parantaa yhteentoimivuutta, mutta vain organisaatioissa, joissa hajautettu omistajuus yhdistyy riittävän vahvaan tekniseen ja organisatoriseen koordinointiin. Näiden havaintojen perusteella data mesh voi parantaa datan uudelleenkäyttöä, löydettävyyttä ja hyödyntämisen nopeutta, mutta vaikutukset kohdistuvat ensisijaisesti tekniseen ja operatiiviseen yhteentoimivuuteen. Samalla havainnot osoittavat, että malli voi lisätä semanttista hajautumista ja hallinnan haasteita.

4.5 Datan yhteentoimivuuden mittarit

Tässä luvussa datan yhteentoimivuutta tarkastellaan mittareiden avulla, jolloin aiemmin esitetyt näkökulmat muunnetaan arvioitaviksi ominaisuuksiksi. Mittareilla on sama rooli kuin luvussa 3.5, mutta tarkastelu kohdistuu käytettävyyden sijaan yhteentoimivuuteen. Empiirinen näyttö on osittain rajallista, minkä vuoksi arviointi perustuu myös epäsuoriin indikaattoreihin ja tapausesimerkkeihin. Osa mittareista on esitelty myös luvussa 3.5 (taulukko 4), koska samat mittarit voivat kuvata sekä datan käytettävyyttä että yhteentoimivuutta eri näkökulmista ja toimia epäsuorina indikaattoreina yhteentoimivuudesta.

Teknistä yhteentoimivuutta voidaan arvioida esimerkiksi rajapintojen, skeemojen ja integraatioiden toimivuuden avulla, kun taas organisatorista yhteentoimivuutta voidaan arvioida datatuotteiden ristiindomain-uudelleenkäytön avulla (Euroopan unioni, 2017, s. 30–31; SAP, 2026). Standardien, yhteisten skeemojen ja metadatakäytäntöjen noudattaminen puolestaan tukee erityisesti teknistä ja semanttista yhteentoimivuutta (Goedgebuure ja muut, 2024, 11:9). Semanttista yhteentoimivuutta voidaan arvioida muun muassa yhteisten skeemojen ja sanastojen kattavuuden sekä metadatan laadun avulla, jotka tukevat datan yhdenmukaista tulkintaa eri domaineissa ja voivat vähentää semanttisia ristiriitoja (Euroopan unioni, 2017, s. 29; Dehghani, 2022, s. 128; Bode ja muut, 2024,

s. 11). Lisäksi tehokkuusmittarit, kuten time-to-insight, voivat epäsuorasti heijastaa yhteentoimivuuden tasoa, koska sujuvat integraatiot ja standardoidut rajapinnat nopeuttavat datan yhdistämistä ja hyödyntämistä käytännössä (Oczkowski, 2025).

Taulukko 6 kokoaa data mesh -ympäristössä käytettyjä yhteentoimivuuden mittareita sekä kuvaa, mitä ne mittaavat, miten ne liittyvät yhteentoimivuuteen ja mihin yhteentoimivuuden osa-alueeseen ne ensisijaisesti kohdistuvat. Suorat mittarit kuvaavat datan siirtymistä, yhdistämistä ja merkityksen säilymistä, kun taas epäsuorat mittarit heijastavat yhteentoimivuuden vaikutuksia, kuten suorituskykyä, käyttöä ja hyödyntämisen tehokkuutta. Mittarit on koottu ja sovellettu useista tässä työssä käytetyistä tutkimus- ja käytännön lähteistä.

Taulukko 6. Datan yhteentoimivuuden mittarit.

Mittari	Mitä mittaa	Yhteys yhteentoimivuuteen	Osa-alue
Ristiindomain-uudelleenkäyttö (Cross-domain reuse)	Datatuotteiden käyttö useissa domaineissa	Osoittaa suoraan yhteensopivuuden yli domain-rajajen	Organisatorinen ja tekninen
Integraatioiden määrä ja monimutkaisuus (Integration count & complexity)	Integraatioiden määrä ja monimutkaisuus	Ei mittaa suoraan siirtymistä/yhdistämistä, vaan arkkitehtuurin tilaa	Tekninen
Standardien noudattamistasaste (Standard compliance rate)	Standardien noudattaminen	Mahdollistaa datan siirron ilman muunnoksia	Tekninen ja hallinnollinen
API-käyttö / onnistumistasaste (API usage / success rate)	Rajapintojen toimivuus	Osoittaa käytännön tiedon siirron onnistumisen	Tekninen
Datasopimusten noudattaminen (Data contract compliance)	Datasopimusten noudattaminen	Varmistaa yhteensopivuuden säilymisen	Hallinnollinen ja semanttinen
Skeeman yhteensopivuus (Schema compatibility)	Skeemojen yhteensopivuus	Mahdollistaa datan yhdistämisen ilman muunnoksia	Tekninen ja semanttinen
Jaettujen skeemojen kattavuus (Shared schema coverage)	Yhteisten skeemojen käyttö	Parantaa rakenteellista yhteensopivuutta (operatiivinen mittaaminen on käytännössä hankalaa)	Semanttinen
Liiketoimintasanaston kattavuus (Business glossary coverage)	Käsitteistön yhtenäisyys	Tukee semanttista yhteentoimivuutta (operatiivinen mittaaminen on käytännössä hankalaa)	Semanttinen
Semanttinen yhdenmukaisuus (Semantic consistency)	Käsitteiden yhtenäisyys	Estää semanttisia ristiriitoja (operatiivinen mittaaminen on käytännössä hankalaa)	Semanttinen

Metadatan laatu (Metadata quality)	Metadatan laatu	Tukee epäsuorasti yhteentoimivuutta (semanttista yhteentoimivuutta tukeva mittari)	Semanttinen
Ontologioiden yhteensovitus (Ontology alignment) (esim. F-luku)	Tietomallien yhtenäisyys	Parantaa semanttista yhteensopivuutta	Semanttinen
Domainien välinen viive (Cross-domain latency)	Kyselyiden viive	Heijastaa integraatioiden tehokkuutta	Tekninen
Aika oivallukseen (Time-to-insight)	Aika oivallukseen	Heijastaa integraation sujuvuutta	Tekninen ja organisatorinen
Käyttöönottoaste (Adoption rate)	Käyttöaste	Voi viitata hyödynnettävyyteen yli domainien	Organisatorinen
Kyselyiden onnistumisaste (Query success rate)	Onnistuneet kyselyt	Osoittaa yhdistämisen toimivuuden	Tekninen
Datan löytämisen onnistuminen (Data discovery success)	Datan löydettävyys	Tukee epäsuorasti yhdistettävyyttä	Organisatorinen ja semanttinen
Dataputkien epäonnistumisaste (Pipeline failure rate)	Virheiden määrä	Voi viitata yhteensopivuusongelmiin	Tekninen
Datan ajantasaisuus (Data freshness)	Datan ajantasaisuus	Epäsuora vaikutus integraatioiden hyödyntämiseen	Tekninen
Liitosten määrä (Interconnections / Network Effect)	Aktiivisten datatuotteiden välisten yhteyksien määrää	Aktiivisten datatuotteiden välisten yhteyksien määrää	Tekninen ja organisatorinen
Skeeman pysyvyysaste (Schema Stability / Backward Compatibility)	Taaksepäin yhteensopivien skeemamuutosten osuutta	Vähentää integraatioiden rikkoutumista muutostilanteissa	Tekninen
Metadatan periytyvyys (Metadata Propagation)	Automaattisesti siirtyvän metadatan kattavuutta integraatioketjuissa	Mahdollistaa yhteentoimivuuden skaalautumisen ilman manuaalista työtä	Tekninen ja hallinnollinen
Datasopimusten testikattavuus (Contract Test Coverage)	Automaattisten sopimustestien kattavuutta rajapinnoissa	Varmistaa tuottajien ja kuluttajien yhteensopivuuden	Tekninen

5 Johtopäätökset

Tutkimuksen keskeinen johtopäätös on, että data mesh -arkkitehtuuri voi parantaa datan käytettävyyttä ja yhteentoimivuutta, mutta vaikutukset riippuvat toteutustavasta. Hajautettu omistajuus voi muun muassa lisätä datan liiketoimintarelevanssia ja nopeuttaa sen hyödyntämistä, mutta samalla se altistaa laadun vaihtelulle, metadatan epätasaisuudelle ja semanttiselle hajautumiselle.

Keskeinen havainto on, että käytettävyys ja yhteentoimivuus kytkeytyvät tiiviisti toisiinsa.

5.1 Vastaukset tutkimuskysymyksiin

(1) Vaikutus datan käytettävyyteen

Hajautettu domain-malli voi vaikuttaa datan käytettävyyteen sekä positiivisesti että negatiivisesti. Positiivisesti se voi parantaa datan löydettävyyttä, saavutettavuutta, osoitettavuutta, ajantasaisuutta, relevanssia ja liiketoimintakontekstin ymmärrettävyyttä, koska vastuu siirtyy niille tiimeille, jotka tuntevat datan parhaiten. Lisäksi se voi nopeuttaa datan hyödyntämistä, vähentää keskitettyihin ratkaisuihin liittyviä pullonkauloja sekä mahdollistaa datan natiivin käytön ilman monimutkaisia integraatioita.

Samalla hajautettu malli lisää useita riskejä. Datan laatu, dokumentaatio ja metadatan kattavuus voivat vaihdella domainien välillä, mikä johtaa epätasaiseen käytettävyyteen. Ymmärrettävyys voi heikentyä, jos datatuotteet eivät ole riittävästi dokumentoituja tai itsekuvailevia. Lisäksi hajautus voi aiheuttaa datan duplikaatiota, synkronointiongelmia ja epäjohdonmukaisia näkymiä samasta ilmiöstä, mikä vaikeuttaa datan tulkintaa. Vaikutus ei siten ole yhtenäinen, vaan riippuu erityisesti domainien kypsyydestä, metadatan hallinnasta ja hallintamallin toteutuksesta.

Käyttäjä- ja sidosryhmänäkökulmasta keskeinen havainto on, että datan käytettävyys ei ole kaikille käyttäjille sama ominaisuus, vaan riippuu eri käyttäjien erilaisista tarpeista. Tämän vuoksi data mesh voi parantaa käytettävyttä vain, jos datatuotteet kuvataan ja tarjotaan tavalla, joka vastaa sekä tuottajien että kuluttajien käytännön tarpeita.

Lisäksi hajautettu ympäristö voi lisätä tietoturva- ja pääsynhallintahaasteita, mikäli turvallisuuspolitiikat ja käyttöoikeudet eivät toteudu yhdenmukaisesti domainien välillä. Toisaalta automatisoitu pääsynhallinta ja policy-as-code -ratkaisut voivat parantaa datan turvallista saavutettavuutta ja vähentää manuaalisia hyväksyntäprosesseja.

(2) Vaikutus datan yhteentoimivuuteen

Data mesh vaikuttaa yhteentoimivuuteen eri tasoilla eri tavoin. Se voi parantaa teknistä yhteentoimivuutta standardoitujen rajapintojen, datasopimusten, avoimien standardien ja itsepalvelualustan kautta, jotka mahdollistavat datan ohjelmallisen käytön yli domain-rajajen. Se voi myös lisätä datan uudelleenkäyttöä, vähentää datan kopiointia ja nopeuttaa datan yhdistämistä, erityisesti silloin kun datatuotteet ovat standardoituja ja löydettäviä.

Toisaalta hajautettu malli lisää merkittäviä haasteita erityisesti semanttisella ja organisatorisella tasolla. Semanttinen yhteentoimivuus voi heikentyä, jos domainit määrittelevät keskeiset käsitteet eri tavoin, mikä estää datan yhdistämisen ilman lisämuunnoksia. Lisäksi tekninen heterogeenisuus voi lisätä integraatioiden monimutkaisuutta, ja organisatorinen fragmentaatio voi heikentää datan jakamista ja koordinoitua domainien välillä. Myös metadatan puutteet, dokumentaation heikko taso ja datan duplikaatio voivat heikentää yhteentoimivuutta.

Näin ollen data mesh voi samanaikaisesti parantaa teknistä yhteentoimivuutta, mutta heikentää semanttista ja organisatorista yhteentoimivuutta, ellei sitä tueta vahvalla standardoinnilla ja federoidulla hallinnalla.

Yhteiset tietoturva- ja pääsynhallintakäytännöt voivat myös tukea turvallista yhteentoinivuutta domainien välillä. Hajautettu hallinta voi kuitenkin kasvattaa hyökkäyspintaa ja vaikeuttaa yhdenmukaisten tietoturvakäytäntöjen toteuttamista, vaikka yksittäinen tietomurto ei välttämättä vaaranna koko järjestelmää.

(3) Edistävät ja estävät tekijät

Keskeisiä edistäviä tekijöitä ovat muun muassa datakatalogit, standardoitu ja rikastettu metadata, datasopimukset, palvelutasotavoitteet ja -sopimukset (SLO/SLA), avoimet standardit, itsepalvelualusta sekä federoitu hallintamalli, jotka yhdistävät hajautetun omistajuuden ja yhteiset käytännöt. Käytettävyyden näkökulmasta keskeisiä tekijöitä ovat lisäksi datan hyvä löydettävyys ja osoitettavuus, selkeä dokumentaatio, itsekuvailevat datatuotteet, natiivi saavutettavuus eri rajapintojen kautta sekä riittävä suorituskyky ja matala tekninen kitka. Yhteentoinivuutta tukevat erityisesti standardoidut rajapinnat, semanttinen yhdenmukaisuus sekä datan uudelleenkäytettävyys yli domain-rajojen. Näitä tukevat teknologiat, kuten data fabric -ratkaisut, tekoälypohjaiset työkalut, datan seurattavuus (observability) sekä ontologioiden kohdistaminen, jotka voivat parantaa erityisesti teknistä ja semanttista yhteentoinivuutta. Käytettävyyttä ja yhteentoinivuutta tukevat lisäksi automatisoitu politiikkojen valvonta, policy-as-code -ratkaisut sekä yhdenmukaiset pääsynhallintakäytännöt.

Keskeisiä estäviä tekijöitä ovat muun muassa metadatan puutteet, dokumentaation heikko taso, domainien epätasainen kypsyyss, puutteellinen automaatio, semanttinen hajautuminen, tekninen heterogeenisyys sekä organisatorinen fragmentaatio. Käytettävyyden näkökulmasta ongelmia aiheuttavat myös heikko löydettävyys ja puutteellinen ymmärrettävyys, datan laadun ja ajantasaisuuden vaihtelu, synkronointiongelmat sekä monimutkaiset tai hitaat käyttöprosessit. Yhteentoinivuutta heikentävät lisäksi datan duplikaatio, integraatioiden monimutkaisuus sekä standardien puutteellinen noudattaminen. Nämä tekijät voivat johtaa tilanteeseen, jossa datatuotteet ovat teknisesti saatavilla ja toimivat paikallisesti, mutta ovat vaikeasti

hyödynnettävissä tai eivät ole yhteentoimivia organisaatiotasolla. Näitä voivat heikentää myös epäyhtenäiset tietoturvakäytännöt, puutteellinen pääsynhallinta ja hajautetun ympäristön lisääntynyt hyökkäyspinta.

(4) Soveltuvat organisaatiot

Data mesh soveltuu parhaiten suuriin ja kompleksisiin organisaatioihin, joissa on selkeä domain-rakenne, riittävä datakypsyys sekä kyvykkyys ylläpitää standardoitua metadataa, dokumentaatiota ja itsepalvelualustaa. Lisäksi malli edellyttää domaineilta riittävää teknistä osaamista, kuten data engineering- ja analytiikkakyvykkyksiä, joiden puute on tunnistettu yhdeksi keskeiseksi käyttöönoton esteeksi. Luvuissa 3 ja 4 esitetyt havainnot osoittavat, että pelkkä hajautettu omistajuus ei riitä, vaan käytettävyys ja yhteentoimivuus edellyttävät vahvaa federoitua hallintaa, yhteisiä standardeja ja semanttista yhteensopivuutta domainien välillä. Ilman näitä hajautettu malli johtaa helposti laadun vaihteluun, semanttiseen hajautumiseen ja heikentyneeseen yhteentoimivuuteen.

Pienemmissä organisaatioissa tai ympäristöissä, joissa datan määrä, monimuotoisuus tai organisatorinen kypsyys ovat rajallisia, data mesh -arkkitehtuurin tuoma kompleksisuus ja hallinnollinen kuorma voivat ylittää saavutettavat hyödyt. Tällöin yksinkertaisemmat keskitetyt ratkaisut voivat tarjota paremman tasapainon käytettävyyden ja hallittavuuden välillä.

5.2 Tutkimustulosten arviointi

Tulosten perusteella data mesh on ensisijaisesti organisatorinen ja sosiotekninen lähestymistapa, jonka hyödyt riippuvat toteutuksen laadusta. Keskeinen haaste liittyy teknisen ja semanttisen yhteentoimivuuden väliseen eroon. Vaikka tekninen integraatio voidaan toteuttaa, datan merkityksen yhtenäisyys ei ole automaattinen.

Osa havaituista hyödyistä ei ole yksinomaan data mesh -mallin seurausta, vaan voi toteutua myös muissa arkkitehtuureissa riittävällä hallinnalla ja standardoinnilla. Data mesh -mallin vaikutuksia koskeva empiirinen näyttö on pääosin tapauskohtaista, mikä tekee vaikeaksi arvioida, missä määrin havaitut hyödyt ovat yleistettävissä eri organisaatioihin.

5.3 Tutkimustulosten tulkinnan rajoitukset

Tutkimus perustuu kirjallisuuskatsaukseen, eikä sisällä omaa empiiristä aineistoa. Käytetyt lähteet painottuvat osittain teollisuuden tuottamiin raportteihin, joissa voi esiintyä kaupallista vinoumaa. Lisäksi monet käytetyt mittarit kuvaavat käytettävyyttä ja yhteentoimivuutta epäsuorasti, mikä vaikeuttaa ilmiöiden tarkkaa mittaamista. Data mesh -kirjallisuus on myös suhteellisen nuorta, eikä yhtenäistä akateemista konsensusta ole vielä muodostunut.

5.4 Ehdotukset jatkotutkimukselle

Jatkotutkimuksessa voisi olla hyödyllistä tarkastella erityisesti data meshin vaikutuksia datan käytettävyyteen ja yhteentoimivuuteen kvantitatiivisten mittareiden avulla, analysoida hajautetun mallin pitkäaikaisia vaikutuksia datatuotteiden laatuun ja ylläpidettävyyteen sekä tutkia data fabric -ratkaisujen, ontologioiden ja tekoälyn mahdollisuuksia semanttisen yhteentoimivuuden ja hajautetun hallinnan tukemisessa.

Lähteet

- Abd Nikooie Pour, M., Blomqvist, E., Cotovio, P. G., Coulet, A., Ferraz, L., Hertling, S., Jain, S., Jiménez-Ruiz, E., Kraus, F., Lambrix, P., Li, H., Li, Y., Liu, X., Monnin, P., Paulheim, H., Pesquita, C., Sharma, A., Shvaiko, P., Silva, M. C., . . . Zhou, L. (2025). Results of the Ontology Alignment Evaluation Initiative 2025. *OM 2025: The 20th International Workshop on Ontology Matching*. Noudettu 17.4.2026 osoitteesta <https://ceur-ws.org/Vol-4144/om2025-oaei-paper0.pdf>
- Alation. (2026, 9. helmikuuta). ROI of Data Automation: How to Calculate (and Why Most Get It Wrong). Noudettu 17.4.2026 <https://www.alation.com/blog/roi-of-data-automation/>
- Araújo Machado, I., Costa, C., & Santos, M. Y. (2022). Data mesh: Concepts and principles of a paradigm shift in data architectures. *Procedia Computer Science*, 196, 263–271. <https://doi.org/10.1016/j.procs.2021.12.013>
- Arora, N., Gaurav, K. S., Long, A., Misra, P., & Rao, R. (2023, 14. heinäkuuta). *How Amazon Finance Automation built a data mesh to support distributed data ownership and centralize governance*. AWS Big Data Blog. Noudettu 18.4.2026 osoitteesta <https://aws.amazon.com/blogs/big-data/how-amazon-finance-automation-built-a-data-mesh-to-support-distributed-data-ownership-and-centralize-governance/>
- Bienzeisler, J., Kombeiz, A., Ehrentreich, S., Otto, R., Schirrmeister, W., Pegoraro, M., Brammen, D., Puladi, B., Röhrig, R. & Majeed, R. W. (2025). Implementation report on pioneering federated data access for the German National Emergency Department Data Registry. *npj Digital Medicine*, 8, Article 94 <https://doi.org/10.1038/s41746-025-01481-w>
- Blohm, I., Wortmann, F., Legner, C., & Köbler, F. (2024). Data products, data mesh, and data fabric: New paradigm(s) for data and analytics? *Business & Information Systems Engineering*, 66(5), 643–652. <https://doi.org/10.1007/s12599-024-00876-5>

- Bodapati, R. V. P. K. (2025). Data mesh architecture for scalable business intelligence systems. *Journal of Computer Science and Technology Studies*, 7(5), 770–777. <https://doi.org/10.32996/jcsts.2025.7.5.86>
- Bode, J., Kühn, N., Kreuzberger, D., & Holtmann, C. (2024). Toward Avoiding the Data Mess: Industry Insights From Data Mesh Implementations. *IEEE Access*, 12, 95402–95416. <https://doi.org/10.1109/ACCESS.2024.3417291>
- Chee, C. W. (2021, 20. lokakuuta). *HelloFresh journey to the data mesh*. HelloTech. Noudettu 20.4.2026 osoitteesta <https://engineering.hellofresh.com/hellofresh-journey-to-the-data-mesh-7fe590f26bda>
- Christ, J., Visengeriyeva, L. & Harrer, S. (n.d.). Data mesh architecture: Data mesh from an engineering perspective. Noudettu 10.5.2026 osoitteesta <https://www.data-mesh-architecture.com/>
- DataHub. (n.d.). Data products. Noudettu 18.5.2026 osoitteesta <https://docs.data-hub.com/docs/dataproducts>
- Dehghani, Z. (2020). *Data Mesh Principles and Logical Architecture*. Martin Fowler. Noudettu 13.4.2026 osoitteesta <https://martinfowler.com/articles/data-mesh-principles.html>
- Dehghani, Z. (2022). *Data Mesh: Delivering Data-Driven Value at Scale* (1. painos). O'Reilly Media.
- Dolhopolov, A., Castellort, A. & Laurent, A. (2024). Implementing Federated Governance in Data Mesh Architecture. *Future Internet*. <https://doi.org/10.3390/fi16040115>
- ENISA. (2025). ENISA Threat Landscape 2025. European Union Agency for Cybersecurity. <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025>
- Euroopan unioni. (2017). *New European Interoperability Framework (EIF): Promoting seamless services and data flows for European public administrations*. Publications Office of the European Union. <https://doi.org/10.2799/78681>
- Gahle, D. S. & Barbarino, M. (2026). *The IAEA fusion data lake project – Accelerating AI and big data applications through open science and FAIR data*. arXiv. <https://doi.org/10.48550/arXiv.2604.01797>

- Gates, S. (2025, 28. kesäkuuta). *What is data completeness? Definition, examples, and KPIs*. Monte Carlo. Noudettu 13.4.2026 osoitteesta <https://www.montecarlodata.com/blog-what-is-data-completeness/>
- Goedegebuure, A., Kumara, I., Driessen, S., Van Den Heuvel, W.-J., Monsieur, G., Tamburri, D. A. & Di Nucci, D. (2024). Data Mesh: A Systematic Gray Literature Review. *ACM Computing Surveys*, 57(1), Article 11. <https://doi.org/10.1145/3687301>
- Gugałka, I. (2024, 25. kesäkuuta). *Measuring the Success of Data Mesh in Your Organization*. InTechHouse. Noudettu 13.4.2026 osoitteesta <https://intechhouse.com/blog/measuring-the-success-of-data-mesh-in-your-organization/>
- Guitouni, S. (2023, 10. lokakuuta). *11 lessons learned managing a Data Platform team within a data mesh*. BlaBlaCar. Noudettu 20.4.2026 osoitteesta <https://medium.com/blabla-car/11-lessons-learned-managing-a-platform-team-within-a-data-mesh-0b191b7652ce>
- Hooshmand, Y., Resch, J., Wischniewski, P., & Patil, P. (2022). From a monolithic PLM landscape to a federated domain and data mesh. *Proceedings of the Design Society*, 2, 713–722. <https://doi.org/10.1017/pds.2022.73>
- ISO/IEC 25012. (2008). *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model*. International Organization for Standardization. Noudettu 20.4.2026 osoitteesta <https://www.iso.org/standard/35736.html>
- Jain, A., Person, G., Conroy, P. & Shankar, N. (2021, 5. toukokuuta). How JPMorgan Chase built a data mesh architecture to drive significant value to enhance their enterprise data platform. AWS Big Data Blog. Noudettu 17.4.2026 osoitteesta <https://aws.amazon.com/blogs/big-data/how-jpmorgan-chase-built-a-data-mesh-architecture-to-drive-significant-value-to-enhance-their-enterprise-data-platform/>
- Janssen, J., Majhi, M., Perillo, G., & Ariza, M. (2025, 3. marraskuuta). *Scaling data governance with Amazon DataZone: Covestro success story*. AWS Big Data Blog. Noudettu 18.4.2026 osoitteesta <https://aws.amazon.com/blogs/big-data/scaling-data-governance-with-amazon-datazone-covestro-success-story/>

- Kashfi, P., Nilsson, A., & Feldt, R. (2016). Integrating user experience practices into software development processes: The implication of subjectivity and emergent nature of UX. *Journal of Systems and Software*, 111, 159–173. <https://doi.org/10.1016/j.jss.2015.09.003>
- Kohoutek, S. (2023, 15. helmikuuta). *How AWS supports HelloFresh's Data Mesh Journey*. HelloTech. Noudettu 18.4.2026 osoitteesta <https://engineering.hellofresh.com/how-aws-supports-hellofreshs-data-mesh-journey-f992f1548e97>
- MacDonald, L. (2024, 19. heinäkuuta). What is data timeliness? It's the boost your business needs. Monte Carlo. Noudettu 13.4.2026 osoitteesta <https://www.montecarlodata.com/blog-what-is-data-timeliness/>
- Mukherjee, D., Taylor, A., Mikoleizig, M., Kumar, R., Almond, S., & Sun, W. (2024, 31. lokakuuta). *How Volkswagen Autoeuropa built a data mesh to accelerate digital transformation using Amazon DataZone*. AWS Big Data Blog. Noudettu 18.4.2026 osoitteesta <https://aws.amazon.com/blogs/big-data/how-volkswagen-autoeuropa-built-a-data-mesh-to-accelerate-digital-transformation-using-amazon-datzone/>
- Oczkowski, W. (2025, 27. lokakuuta). *How to Build a Data Mesh Team: Roles and Responsibilities*. InTechHouse. Noudettu 13.4.2026 osoitteesta <https://intechhouse.com/blog/how-to-build-a-data-mesh-team-roles-and-responsibilities/>
- Roessler, K. (2026, 25. helmikuuta). Ponemon Cost of Insider Risks: Key Findings 2026. DTEX Systems. Noudettu 12.4.2026 osoitteesta <https://www.dtex.ai/blog/ponemon-cost-of-insider-risk-2026/>
- Rose, S., Borchert, O., Mitchell, S. & Connelly, S. (2020). *Zero Trust Architecture*. NIST Special Publication 800-207. <https://doi.org/10.6028/NIST.SP.800-207>
- SAP. (2026, 26. tammikuuta). *What is data mesh? Principles, architecture, comparisons*. Noudettu 13.4.2026 osoitteesta <https://www.sap.com/resources/what-is-data-mesh>
- Serra, J. (2024). *Deciphering Data Architectures: Choosing Between a Modern Data Warehouse, Data Fabric, Data Lakehouse, and Data Mesh*. (1. painos). O'Reilly Media.

- dbt Labs. (2024). *Siemens implements a data mesh architecture at scale with dbt Cloud*. Noudettu 15.4.2026 osoitteesta <https://www.getdbt.com/case-studies/siemens>
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103–110. <https://doi.org/10.1145/253769.253804>
- Swarts, J. (2025). The construction of data usability. *Technical Communication Quarterly*, 34(2), 199–217. <https://doi.org/10.1080/10572252.2024.2401359>
- Tonnarelli, M., Scaramuzza, F., Harrer, S., & Dietz, L. W. (2026). Data Product MCP: Chat with your Enterprise Data. arXiv. <https://doi.org/10.48550/arXiv.2601.08687>
- Tupper, C. D. (2011). *Data architecture: From Zen to reality*. Morgan Kaufmann. <https://doi.org/10.1016/C2010-0-66325-5>
- Verizon. (2025). 2025 Data Breach Investigations Report. <https://www.verizon.com/business/resources/Tea/reports/2025-dbir-data-breach-investigations-report.pdf>
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–34. <https://doi.org/10.1080/07421222.1996.1151809>
- Wider, A., Verma, S. & Akhtar, A. (2023). Decentralized data governance as part of a data mesh platform: Concepts and approaches. Teoksessa *Proceedings of the 2023 IEEE International Conference on Web Services (ICWS)* (s. 746–754). IEEE. <https://doi.org/10.1109/ICWS60048.2023.00101>
- Winter, R., & Hackl, T. (2025). Exploring data mesh adoption in large organizations. *Issues in Informing Science and Information Technology*, 22. <https://doi.org/10.28945/5509>