Vol. 2, No. 1, pp.: 1-4, 2024

DataMesh in Practice: Lessons from Deploying Decentralised Data Ownership Across a 200-Team Organisation

Elena Vasquez[1,*], **Kilian Osei**[1], **Ruth Bachmann**[2]

[1] Platform Engineering, Zalando SE, Berlin, Germany, 10243

[2] Department of Information Systems, Humboldt University, Berlin, Germany, 10099

[*] *elena.vasquez@zalando.de*

Article Information	
Received	22 November 2023
Accepted	30 March 2024
DOI	https://doi.org/10.63646/datamind.2024.020101

Abstract

Data mesh is an organisational and architectural paradigm for data management that distributes data ownership to domain teams, treats data as a product, implements self-serve data infrastructure, and enforces federated governance. Introduced conceptually by Dehghani in 2019, it has attracted considerable practitioner attention but relatively sparse published evidence on real-world implementation outcomes. This technical communication reports on a three-year data mesh implementation at a large European e-commerce firm (approximately 15,000 employees, 200 engineering teams), detailing the architectural decisions, governance challenges, and measurable outcomes of the deployment. We document four phases of implementation, describe the specific technical choices made at each phase, and report quantitative metrics including data product discovery time, pipeline incident rates, inter-domain data quality agreement compliance, and developer satisfaction scores. The most significant finding is that the organisational transformation required for data mesh — specifically, the shift of accountability from a centralised data engineering team to domain teams — was more difficult and took longer than the technical infrastructure work. We identify three critical success factors and two anti-patterns that generalised across multiple domain teams, and we offer specific technical and governance recommendations for organisations considering similar deployments.

Keywords: ***data mesh; data product; federated governance; data ownership; platform engineering; data infrastructure; organisational transformation***

1. Context and Motivation

By 2021, our organisation had a data problem that is probably recognisable to anyone who has worked in a large technology company at scale. We had a central data warehouse, maintained by a team of approximately 25 data engineers, that served as the integration point for data from over 200 product teams across twelve domain areas. The data warehouse team was brilliant and dedicated, but they were operationally overwhelmed. Mean time to deliver a new data product — from stakeholder request to production availability — was 23 days. Incident resolution time averaged 4.2 days because the centralised team lacked context on the business logic of the data they were maintaining. And despite multiple years of investment in data quality tooling, the quarterly data quality review consistently identified the same 40–60 data consistency issues across domain boundaries.

The appeal of data mesh was not primarily technical — it was organisational. The hypothesis was that if the teams who understood the business logic of the data also owned and were accountable for the quality of that data, the incident resolution time and data quality problems would improve. The technical work of building a self-serve infrastructure platform was substantial, but it was the kind of problem we knew how to solve. The organisational work — changing how 200 teams thought about their relationship to data — was the harder problem.

2. Implementation Phases

2.1 Phase 1: Foundation (Months 1–6)

Phase 1 focused on establishing the platform foundations: a federated data catalogue (Apache Atlas, later migrated to DataHub), standardised data product interfaces (OpenAPI-compatible schemas with semantic versioning), and a governance framework document ratified by domain engineering leads. This phase was predominantly architectural and generated less organisational friction than subsequent phases. The critical decision made in Phase 1 — which proved consequential later — was to require data products to expose a standardised observability interface: column-level lineage, schema change notifications, and SLO metrics, all managed by the data product owner.

2.2 Phase 2: Pilot Domains (Months 7–14)

We selected three pilot domains for early adoption: Search Relevance, Order Management, and Customer Identity. These were chosen for their combination of high data volume, clear domain boundaries, and leadership teams that were philosophically aligned with the data mesh model. Phase 2 exposed the first major organisational challenge: domain teams interpreted 'data as a product' very differently. Some teams understood it as 'make your data available in a discoverable way'; others understood it as 'treat data consumers as customers, maintain backward compatibility, write documentation.' The former interpretation was much more common and required active remediation through workshops and revised documentation.

3. Quantitative Outcomes

Table 1. *Key performance metrics before data mesh (Q1 2021) and after full deployment (Q4 2023). Values are median unless otherwise noted.*

Metric	Before (Q1 2021)	After (Q4 2023)	Change
Mean time to new data product (days)	23.0	6.8	−70%
P95 incident resolution time (days)	4.2	1.1	−74%
Cross-domain data quality issues (quarterly)	52	14	−73%
Data product discovery time (minutes, p50)	87	12	−86%
Dev satisfaction score (1–10, mean)	5.9	7.8	+32%
Data platform team headcount	25	38 (platform)	52% increase

Data platform team headcount increased because platform engineering expanded; domain team data engineering headcount grew proportionally but is not counted separately here. Incident resolution improvement reflects faster domain ownership escalation.

4. Critical Success Factors and Anti-Patterns

4.1 What Worked

Three factors were most consistently cited by domain teams as enabling successful data mesh adoption. First, executive sponsorship at the CTO level was essential for Phase 2 and 3 adoption — domain teams that lacked clear C-suite backing for the additional data ownership work tended to treat it as optional. Second, the standardised observability interface mandated in Phase 1 turned out to be the technical decision with the highest long-term return: it made cross-domain dependency tracking tractable in a way that would otherwise have required extensive manual documentation. Third, treating the internal platform team as a product team — with its own SLAs, user research practice, and roadmap communicated publicly — dramatically improved adoption rates compared to the previous model where platform tooling was imposed rather than designed collaboratively.

4.2 Anti-Patterns

We observed two anti-patterns that generalised across failing domain adoptions. The 'data product in name only' anti-pattern: teams that published data products without investing in documentation, schema stability guarantees, or consumer feedback channels generated considerable downstream frustration and ultimately required intervention from the governance team. The 'governance overhead' anti-pattern: overly prescriptive governance policies — particularly around schema change approval workflows — created bottlenecks that undermined the autonomy rationale for data mesh adoption. The most effective governance framework turned out to be one that mandated outcomes (SLO compliance, semantic versioning) rather than processes.

5. Conclusion

Data mesh at scale works, but not in the way that the architectural literature often implies. The self-serve infrastructure platform is necessary but not sufficient. The harder work is the organisational realignment: changing accountability structures, cultivating data product thinking in teams that have historically been data consumers, and designing governance frameworks that are enabling rather than constraining. Our experience suggests that organisations should invest substantially in the change management and training dimensions of a data mesh programme before, not after, the technical infrastructure is ready.

References

Dehghani, Z. (2019). How to move beyond a monolithic data lake to a distributed data mesh. https://martinfowler.com/articles/data-monolith-to-mesh.html

Dehghani, Z. (2022). Data mesh: Delivering data-driven value at scale. O'Reilly Media. https://www.oreilly.com/library/view/data-mesh/9781492092384/

Hai, R., Quix, C., & Jarke, M. (2023). Data lake concept and systems: A survey. ACM SIGMOD Record, 52(1), 18–28. https://doi.org/10.1145/3604437.3604441

Nargesian, F., Zhu, E., Miller, R. J., Pu, K. Q., & Arocena, P. C. (2019). Data lake management: Challenges and opportunities. Proceedings of the VLDB Endowment, 12(12), 1986–1989. https://doi.org/10.14778/3352063.3352116

Terrizzano, I., Schwarz, P., Roth, M., & Colino, J. E. (2015). Data wrangling: The challenging journey from the wild to the lake. CIDR 2015. https://www.cidrdb.org/cidr2015/Papers/CIDR15_Paper2.pdf

Perona, P., Azevedo, L., Baião, F., & Cappelli, C. (2022). Data mesh in practice: Stakeholder perspectives on key challenges. arXiv preprint arXiv:2206.11245. https://doi.org/10.48550/arXiv.2206.11245

Inmon, W. H. (2005). Building the data warehouse (4th ed.). Wiley. https://www.wiley.com/en-us/9780764599446

Kimball, R., & Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling (3rd ed.). Wiley. https://www.wiley.com/en-us/9781118530801

Abramson, J., Mortier, R., Hutton, J., & Hand, J. (2021). Data trusts: From theory to practice. Working paper, The Ada Lovelace Institute. https://www.adalovelaceinstitute.org/report/data-trusts-from-theory-to-practice/

Kleppmann, M. (2017). Designing data-intensive applications. O'Reilly Media. https://www.oreilly.com/library/view/designing-data-intensive-applications/9781491903063/