

Data Mastering Holds Promise as Underpinning Technology for Data Mesh

The 451 Take

There is some ambiguity around the term “data mesh” as organizations try to optimize their data management strategies, gain more consistent business value and become more data-driven. Data mesh is largely a socio-technological construct rather than a specific technology. What a data mesh approach seeks to achieve is a decentralized data architecture that groups and curates data by a specific business domain.

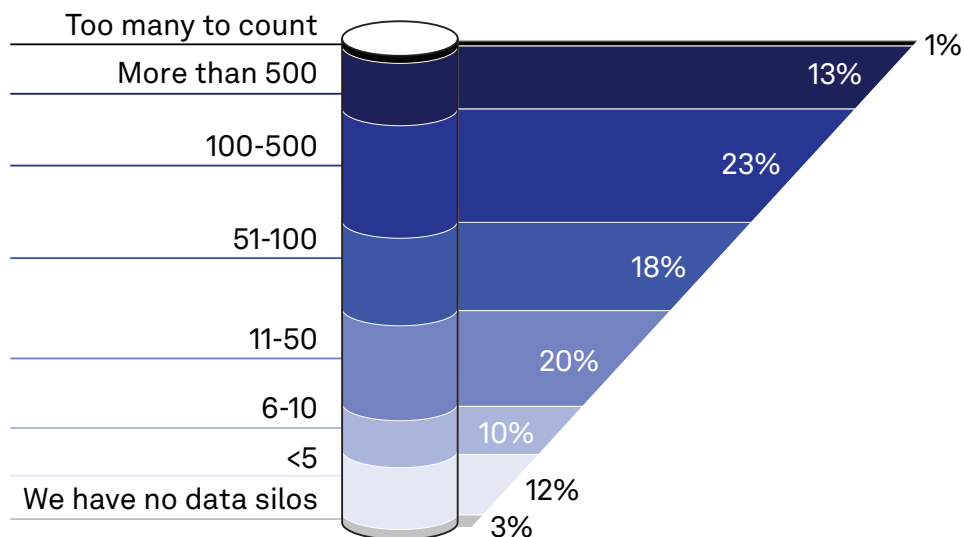
Such an approach gives more autonomy to individual “data owners” and subject-matter experts to help productize and operationalize the datasets that self-service data consumers depend on for insight. However, the decentralized data architecture associated with data mesh is not mutually exclusive with the goal to achieve a more consistent view of enterprise data resources.

The federated method associated with data mesh seeks to break down functional data silos. Challenges remain, however. If data has not been mastered in some way to create universal and persistent IDs for key entities, the federated queries associated with data mesh will likely be very difficult, or will create even more silos. Having a data mastering solution — the process by which unmastered data records are linked to master records across domains — in this sense can be highly supportive to achieving a successful data mesh strategy.

Modern, distributed IT architecture can pose a barrier to organizations trying to gain a more unified view of enterprise data resources. Hybrid, cloud and multicloud architectures — as well as countless scattered SaaS applications — can confound efforts toward data integration and analysis.

Nearly four out of ten respondents (37%) in 451 Research’s Voice of the Enterprise: Data & Analytics, Data Management & Analytics 2021 survey estimate that their organization has over 100 distinct departmental data silos, which creates data variety and complexity. Larger organizations, and organizations that base a higher proportion of their strategic decisions on data, are even more likely to report a high number of data silos.

Estimated Number of Data Silos Across Organizations



Q. How many data silos would you estimate exist across your organization?

Base: Data management respondents (n=193)

Source: 451 Research’s Voice of the Enterprise: Data & Analytics, Data Management & Analytics 2021

Given this complexity, many organizations are looking at data mastering technology to supplement their data mesh approach and gain a more integrated and consistent view of data resources. Data mastering can serve as both a complement and augmentor to these efforts, by providing standardized keys for data that can be understood across systems and domains. Data mastering can act as a foundational way to create useful mappings between data identifiers across the organization, aiding in data mesh strategy.

Business Impact

Data mesh is a socio-technological concept, but it can benefit from consistent implementation of data mastering.

Enterprise data needs to be cleansed and standardized for the data mesh concept to work at its full potential. Waiting to standardize and cleanse data very near the point of data consumption likely burdens talent and technical resources that are already stretched thin. Consistent data mastering can form a foundational layer to provide standardized keys for data so that federated query is possible.

Enablement of end users should be a key focus for data mesh strategy, and federated query can help accelerate outcomes.

Assessing desired business outcomes is an important part of data mesh strategy, and most organizations want to enable their business users to be more productive with data. Federated query, enabled by data mastering, helps ensure that data is validated, cleansed and standardized before the point of consumption by end users, thus accelerating insight.

Multicloud and hybrid architecture intensifies the need for a unified view of data. In 451 Research's Voice of the Enterprise: Data & Analytics, Data Platforms 2022 survey, 40% percent of organizations report using multicloud deployments, while 41% report using hybrid cloud. Diversified cloud environments can scatter and silo data resources. Technological mechanisms to give a "lingua franca" to these sources, such as data mastering, can help buffer against data silos and skewed insights.

Methods such as data mastering can help overcome the major challenges in gaining a more unified view of data. As organizations try to gain a more unified view of data amid data mesh efforts, challenges remain. Data quality becomes more complex with more data sources, and data security/privacy requirements bear down with proliferating regulations. Data mastering can provide the standardized keys that make handling requirements across environments, such as data lakes and data warehouses, simpler and more consistent.

Looking Ahead

Data mesh, at face value, may seem contradictory: a socially decentralized construct to ultimately get a more unified view of enterprise data resources. However, as organizations face exponentially growing volumes of data, there is merit in giving more autonomous control to individual business domains and subject-matter experts to help steward data appropriately for specific business use cases.

Supporting technology, including automation and machine learning, will be necessary here. While there is enormous value in expertise, humans alone cannot consistently curate data sets at the scale necessary for a large or data-intensive organization. If the enterprise wants to federate data consistently, there should be a universal key for data. Data mastering, particularly if it includes elements of human-improved automation and machine learning, can help achieve this.

Flexibility and compatibility are critical in a data mesh approach. Data needs to integrate with existing pipelines and feed into applications such as visualization environments, so APIs and open architecture are table stakes. Hybrid, cloud and multicloud compatibility is a pressing reality for technologies that support data mesh approaches. Data mastering, as an underpinning technology for data mesh, can extend to all the enterprise's data-driven functions. Not just confined to customer 360 or CX use cases, data mastering can feed into programs for advanced analytics as well as data science and AI/ML model development. Given current concerns around data privacy, data security and compliance, data mastering can provide an additional layer of control and audit necessary for today's proliferating requirements.



Tamr is a highly scalable next-generation data mastering platform that allows customers to use data product templates to consolidate messy source data into clean, curated, analytics-ready datasets. Organizations benefit from Tamr Mastering, the industry's first suite of data product templates that combine human curation, patented machine learning, mastering rules and enrichment with first- and third-party data to accelerate business outcomes and deliver business-changing insights. Tamr's cloud-native and SaaS solutions enable industry leaders such as Toyota, Western Union and GSK get ahead and stay ahead in a rapidly changing competitor environment. Tamr's work in the public sector includes the U.S. Air Force, U.S. Army, Department of Homeland Security and various other federal and local government agencies. www.tamr.com