

IRI

PRODUCE DATA
PRODUCTS WITH
VERACITY WITH
VORACITY

BDOQ
BIG DATA QUARTERLY

Data Mesh Adoption: Challenges and Opportunities

Best Practices Series

Adopting a Data Mesh Approach

THERE IS A GROWING DISCONNECT BETWEEN ENTERPRISES seeking greater data-driven capabilities and the actual data that is on the ground of their business units. That's because bottlenecks, silos, over-centralization, and organizational layers are hindering the access and capabilities needed by rapidly expanding userbases. With a growing reliance on data for decision making, there needs to be greater accountability of the data that is moving through these organizations.

Don't look to the cloud to help resolve these issues. Although the cloud may provide increased storage and faster processing capabilities, it cannot unravel the issues faced by today's organizations. Instead, many companies see data mesh as the most effective way to deliver the data power enterprises require—and move toward long-sought data democracy. The time has come to empower business domains with their own accountability, their own pipelines, and their own governance.

Data mesh was an academic and theoretical concept that is now becoming a reality. Almost all data executives, 92%, agree it's more viable to have a decentralized data architecture that empowers individual business users, a recent study out of BARC Research found. Currently, 19% report they have data mesh at their sites, and 35% are planning to move in this direction.

"To become a more data-driven enterprise, you need a shift towards more decentralized data ownership in the business domains," the BARC study's authors stated. "A decentralized model ensures scalability and thus flexibility and agility. In addition, domain expertise creates even better data products that meet business requirements. In this scenario, the central data teams

take on a new, very important role. They support the business domains as advisors and ensure the availability and stability of the data and analytics platform."

Data mesh is often conjoined with data fabric, but the two represent different architectural approaches. Data fabric provides a single view of data from all points in the enterprise and beyond. Data mesh, which focuses on decentralization, is intended to provide localized control over these data assets. The two architectures, combined or separate, provide decision makers and applications faster access to essential data, as well as ensuring the quality of that data.

The following are best practices associated with successful data mesh design and development:

- **Be ready to transform.** Data mesh addresses many concerns about the viability of enterprise data assets scattered across enterprises and beyond. Within the more traditional model of centralized data environments, data managers and analysts may not be fully in touch with the needs of business units or their customers. Transforming to a data mesh means greater availability of data—and key applications such as data analytics or AI—to end users who need it when they need it. In turn, this translates to increased flexibility for the business and the ability to react faster to changes with customers or markets.

The key word here is "transformation." That is, data mesh implementation—and ensuring business value—may not be realized in the early stages. To deliver on its potential, data mesh requires cultural change as well as technological



Best Practices Series

evolution. It serves as an integration opportunity but requires a high level of collaboration between business teams that need to share data with one another. This means greater autonomy within more decentralized business units that will be charged with handling their own data and contributing to the mesh.

- **Determine if mesh is a good fit.** Importantly, data mesh is not an ideal fit for every enterprise—particularly small to medium-size ones. A well-functioning data mesh is based on decentralization across disparate business units, which may not be part of the architecture of a smaller organization with limited resources. A large enterprise, however, may have numerous data pipelines across various business units, with more localized data management requirements. A data mesh architecture may help strengthen those more distributed elements.
- **Build a business case.** Of course, as with any technology investment and rollout, there needs to be a viable business case for data mesh. In the long run, implementing data mesh is more of a business than a technical initiative. Empowering disparate data sources while simultaneously weaving them into an enterprise whole means rethinking many aspects of the business and the way information is handled and shared. The restructuring of data management oversight and responsibilities that data mesh brings means change management is required. With such a federated approach to data management, power shifts back to the individual business units, whose

To deliver on its potential, data mesh requires cultural change as well as technological evolution.

managers best understand the value of their data assets. This calls for a reorientation of how data is organized and leveraged across the enterprise. It's a long-term effort, not an overnight sprint, so the value of a data mesh development and proliferation effort will not be fully evident at first. Data mesh is a step-by-step proposition.

- **Implement an enterprise-wide culture change.** This makes organizational culture a key deciding factor in the viability and success of data mesh. The organization must be ready to leverage the benefits of data mesh before implementation is underway. It's challenging to transition from more centralized data management to a decentralized environment.
- **Put good governance in place.** Governance concerns, including security and data compliance, are also issues that must be addressed as data mesh evolves. Security becomes even more of a priority in a highly decentralized data mesh environment. In essence, data mesh reinforces the notion that data security is everyone's responsibility now, not relegated to a security professional or team at the center of the enterprise.
- **Refresh skills for decentralized environments.** The distributed data teams that will undertake data mesh development and maintenance need to have the skills, as well as the roles, to manage such initiatives. No longer can enterprises rely on core data management teams to handle all the challenges that come down the data pipeline. Training is essential, as is keeping the rest of the organization up-to-date with the new data culture that is evolving. Organizations may even need to develop a team with the skills to focus on implementing the architecture—one that consists of data owners, product managers, data engineers, data scientists, and DBAs.

Data mesh will not magically deliver overnight results, as it requires ongoing engagement from across the organization. It is not an end in itself. Data mesh takes data environments one step closer to the ideal of data democratization. In an era in which data analytics means competitive differentiation, it's critical that all decision makers—no matter how technically proficient—have access to the data they need when they need it. Data mesh is opening up data when and where it is needed.

—Joe McKendrick

Produce Data Products with Veracity with Voracity



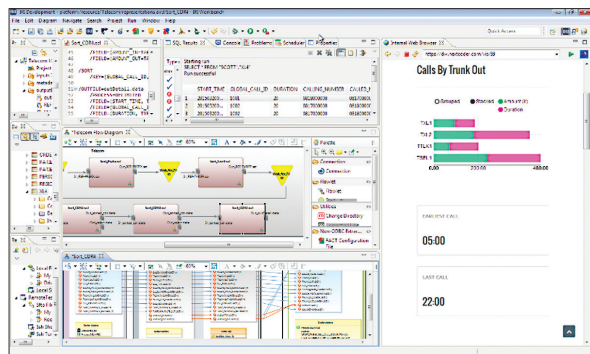
IRI Voracity
An Insatiable Appetite for Data

THE **IRI VORACITY** data management platform has been featured and described in multiple best practice and big data-related issues of *DBTA Magazine*. You may remember it for its ability to combine data discovery, integration, migration, governance, and analytics. Because Voracity maps data between multiple source and target formats and silos, it's a platform you may wish to consider using at the center of your data mesh.

The data mesh is essentially a more recent paradigmatic manifestation of other Voracity-supported data integration architectures. Voracity can be used in ways not too dissimilar from a **Logical Data Warehouse** or **Production Analytic Platform**. What distinguishes a data mesh primarily are its characteristics of: 1) Domain-driven ownership/design, 2) Data as a product, 3) Self-service data infrastructure, and 4) Federated computational governance.

As described in previous *DBTA* best practices issues around Data Integration and Governance, Voracity can serve as a Data Infrastructure *and* Federated Governance platform which supports domain-driven design. If you review our **Big Data Packaging** Page, you'll see that the core idea of Voracity as a data mesh platform is the same—to integrate and govern data from disparate sources to make informational output, or data products, fit for data science use (and re-use).

In a data mesh, you wind up with something akin to data marts, and in Voracity, you can—without a data warehouse team—produce immediate data products or reports, too. Integration, cleansing, masking, reporting are all same-job, same-pass affairs. You can quickly wrangle your own data products that contain trustworthy, business-formatted subsets (or formatted reports) with clean, anonymous, and de-duplicated data inside.

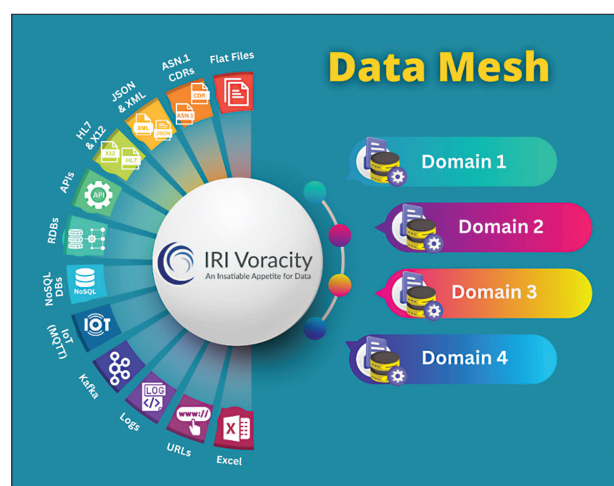


Voracity follows “A Data Mesh Approach to Data Warehousing,” as described in this 2021 Alexis Mackenzie article in *Towards Data Science*—from FAIR principles supported by SortCL data definition language files cataloging structured source **metadata** to using non-competing, real-time read replicas supported by the IRI Ripcurrent **module**.

Voracity delivers several **Federated Governance** platform features recognized in data mesh architecture, including:

data search, profiling, classification, masking and cleansing (quality), plus basic cataloging and lineage. A number of data observability and access control capabilities are also included.

Voracity **Data Infrastructure** platform features include: support for polyglot source and target data stores, pipeline orchestration, CI/CD integration, data product versioning (via Git), PII and access credential encryption, and big data transformation performance that scales vertically or horizontally. A number of graphical self-service data engineering and democratization features are also available through the **DataSwitch** web front-end.



Beyond the functionality Voracity embeds, another key consideration is how it addresses a common problem with data meshes: a lack of clarity around data ownership. When different people are looking at different data products to find answers to the same quantitative question, they will get different answers, and usually after data has been replicated too many times.

A properly designed data mesh can clarify data ownership and thus an authority within the company for engineering and business teams to ask about: common data definitions, data discrepancies, inter-domain integration points, and reducing unnecessary data normalization steps.

The aforementioned FAIR metadata in Voracity can help align data producers and consumers and standardize data formatting. Data sources and accesses are registered in the **IRI Workbench** design IDE and readily identified in job scripts. Voracity data-product projects with those artifacts can be shared, compared, and discussed from a common repository like **Git**.

This framework in Voracity is flexible and easy enough to create small test projects quickly. By conducting a variety of POCs, your company can find the best ways to guild governed, efficient data products.

IRI VORACITY

www.iri.com/products/voracity