ontotext

Data mesh is still in its infancy, and data personas and organizations are craving clarity and specificity. It is critical to be aware of the “why” and “what” and fully understand the role that knowledge graphs play when considering adopting a data mesh strategy.

The debate on what constitutes a data mesh rages on. Most agree that it is not a platform or a commercial off-the-shelf product and that it should make data more accessible, secure, discoverable, and interoperable. But beyond the confusion of what it is, there are also the challenges of understanding the processes and data connections needed to deliver on a data mesh effort. Clarifying what data mesh is and how it works is important, but even more so is starting with the why.

Why Organizations Need A Data Mesh

The data mesh paradigm aims to address some of the biggest pain points for most organizations. One of these is scaling enterprise data infrastructure to incorporate different data types from various sources. This becomes particularly problematic when dealing with veracity of data. Data mesh solves this by promoting data autonomy, allowing users to make decisions about domains without a centralized gatekeeper. It also improves development velocity with better data governance and access with improved data quality aligned with business needs.

The inability to establish a domain-centric sense of data ownership that is often far removed from the business understanding of data is another common challenge, making it hard for

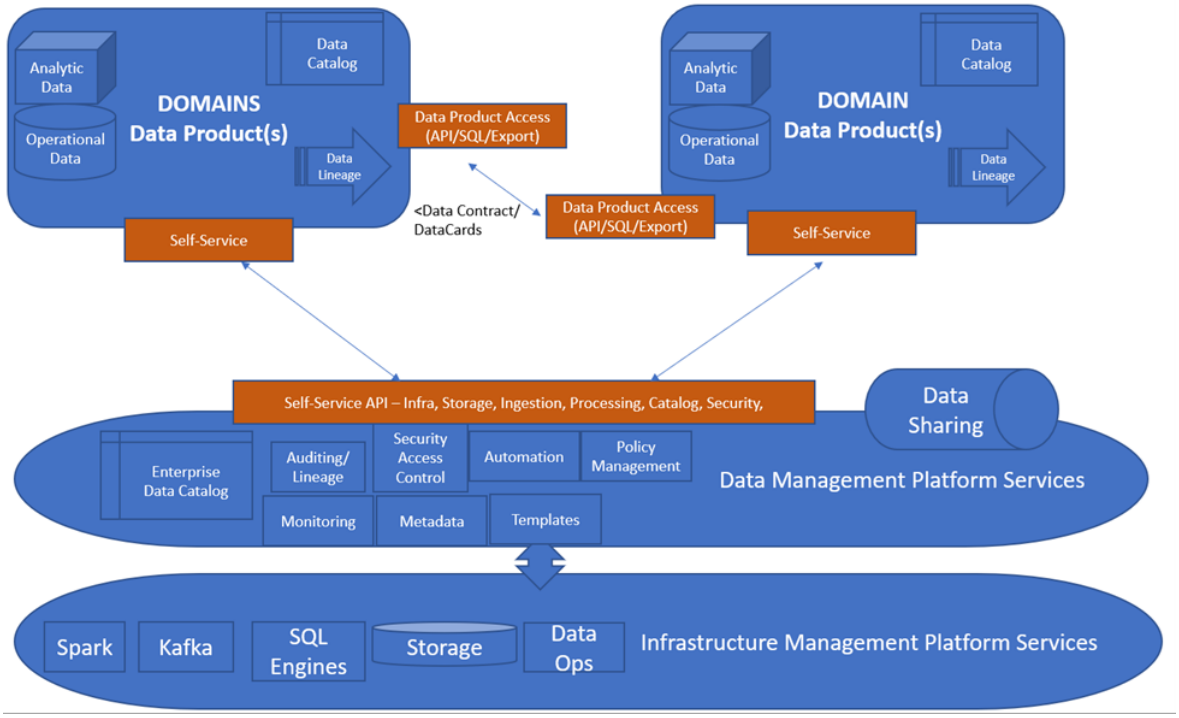
Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

organizations to leverage their data as a strategic asset. A mesh approach helps distribute data ownership and reduce dependencies between services, creating an environment that promotes data-driven thinking.

Coordinating and communicating effectively across cross-functional teams is another challenge, and often the root of failures as the gap between data and the business widens. By allowing organizational autonomy between teams, data mesh eliminates central bottlenecks and delivers value from data.

Figure 1 Shows the overall idea of a data mesh with the major components:



What Is a Data Mesh and How Does It Work?

Think of data mesh as an operational mode for organizations with a domain-driven, decentralized data architecture. It's a combination of implementation, organizational patterns, and a technology-agnostic set of principles. By pushing the ownership to the product owners of the data, it can better serve the consumers, changing the way data

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

projects are managed within organizations.

According to Zhamak Dehghani, who pioneered the paradigm, key principles of data mesh include:

- Treating data as a product – Empower business units to own and manage their data and provide it to others as a product itself.
- Domain-driven ownership of data – Data is owned by those who create it, understand its nuances, and have the expertise to manage it. Users gain the ability to find, explore, create, and enrich new data sources based on use cases with centralized governance for security and privacy.
- Self-serve data platforms – Allows domain teams to provide self-serve capabilities that reduce the complexity of data creation and consumption of data products.
- Federate trust and computational platform – Serves as an ecosystem where users get value from aggregating and correlating independent data products as the data mesh is created based on interoperability standards.

To understand the principles of data mesh better, and how to best enable it, we need to first discuss the key components including domain, data product, data contracts, and data sharing.

What Is a Domain?

The term domain in a data mesh context is a logical grouping of organizational units fulfilling a functional condition. It's a set of tasks that the domain is assigned to perform—the reason the domain exists within organizational constraints. In a data mesh, domains are represented by a node, which can be an operational data store (ODS), a data warehouse, or a data lake tailored to the domain's requirements. Domains can ingest operational data, build analytical data models as data products, and publish them with data contracts to serve other domains' data needs. Mesh emerges when teams use other domains' data products and the domains communicate with others in a governed manner.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

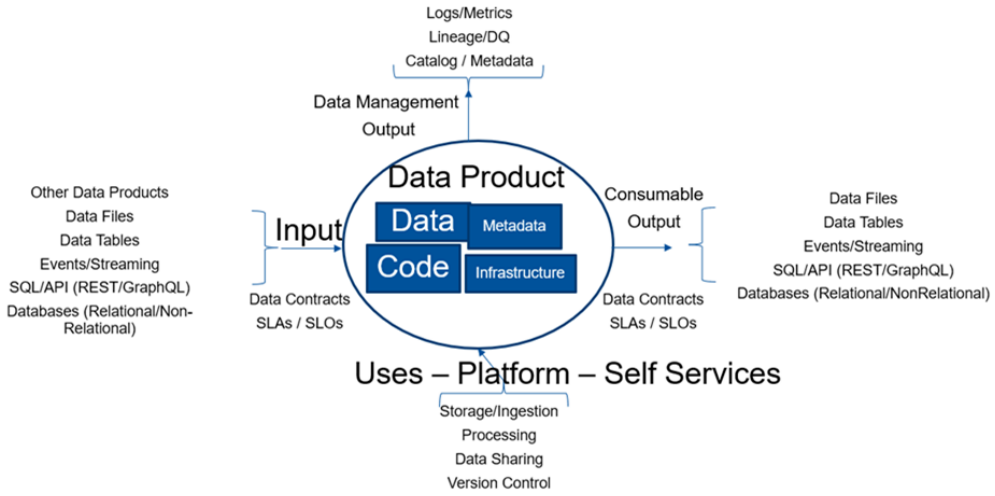
ontotext

What Is a Data Product and Who Owns Them?

A data product is the node on the mesh that encapsulates code, data, metadata, and infrastructure. They are crafted, curated, and presented to consumers as self-service, providing a reliable and trustworthy source for sharing data across the organization.

Domain-oriented teams own data products and are accountable for managing their service level agreements (SLAs), data quality, and governance. The data product owner is responsible for establishing the mechanisms for allowing data producers and data consumers to interact and transact safely and reliably. The owner also provides infrastructure and mechanisms to permit data producers and consumers to interact.

Figure 2 Shows the concept of a data product:



What Is a Data Contract?

Data contracts let domain developers build products with specifications. They guarantee interface compatibility and include terms of service and an SLA. Contracts cover how data is used and the required quality of data. Their goal is to provide transparency for data usage and dependencies and to define terms of service and SLAs. However, this requires a cultural shift and users need time to become familiar with them and understand the importance of

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

data ownership. Data contracts should also include information schema, semantics, and lineage.

What is Data Sharing?

Data sharing enables domain teams to connect and share data products without copying. Ideally, data should not be copied, reducing silo proliferation and keeping ownership with domain owners. Centralized data governance should be used to share datasets securely between producers and consumers, best done through metadata linking.

Don’t Forget Team Structure

Team and organizational structure are an important aspect to consider for data mesh. It is typical to organize teams around selected domains rather than have a centralized team. Domain teams are responsible for all processes—data collection, transformations, cleaning, enrichment, and modeling. Within a domain, teams are organized vertically and consist of roles required to deliver data such as DataOps engineers, data engineers, data scientists, data analysts, and domain experts.

Knowledge Graphs and Data Mesh

The semantic and context-driven fundamentals of knowledge graphs make it ideal to support enterprise data mesh and data fabric-based development. Knowledge graphs can be leveraged to ensure data contracts are standardized, uniform, consistent, semantically correct, and aligned with data sets. They power data sharing platforms to connect data between users, systems, and applications in a consistent, unambiguous manner. This enables compliance with data contracts, ensuring data types, schema, entities, and their inter-relationships across data products are semantically valid.

Domain-centric and enterprise data catalogs can leverage a knowledge graph for storing the semantics with metadata. Knowledge graphs help in automatic metadata extraction, generation and gatekeeping of data quality, and certifying data assets based on semantic rules and validation criteria. Leveraging knowledge graphs with data mesh can give rise to a semantic data mesh, providing the data across different domains in the mesh with context

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

and meaning. This facilitates semantic data discoverability, data interoperability and integration for data augmentation, enrichment and providing explainability for AI and ML use cases.

Data mesh is first and foremost a culture, processes, and people shift, and these things are hard to change quickly in larger organizations. It is not something that can easily be adopted and implemented like a data architecture. Some organizations focus on specific aspects of data mesh or implement a simpler version of the architecture.

For organizations, data mesh is not a matter of yes or no, rather, it should be an exercise in identifying what is preventing them from delivering value to the business in a timely, effective manner. The ability to connect, share, and access data effectively across the enterprise is a very likely answer and is one that a data mesh supported by knowledge graphs is set up to deliver.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com