WHITEPAPER

# How a Semantic Layer Brings Data Mesh to Life

ATSCALE

# Introduction

A data mesh seeks to democratize data and empower self-service business intelligence. The principles of a data mesh approach make it simple for domain owners — regardless of their level of data expertise — to use, modify, and store data in a way that makes sense for their specific business needs. It does this by enabling these domain experts to build their own data products with pre-established building blocks. Data mesh transforms untapped, disparate data stores into a connective tissue that's accessible and usable for the entire organization. And it enables this autonomy while also upkeeping universal governance. This ensures that the data gets used in the correct ways.

This white paper is broken down into four chapters, covering:

1. The Definition of Data Mesh: What Is It and Why Do I Need One?
2. Data Mesh Success Depends on Business-Ready Data
3. How to Implement a Composable Analytics Strategy that Supports your Data Mesh
4. Building Guardrails into your Data Mesh with Analytics Governance

# Chapter 1: The Definition of Data Mesh: What Is It and Why Do I Need One?

Today's leading businesses prioritize data mesh. The modern data mesh craze ultimately came about to streamline data democratization – that is, making the right data available to the right people at the right time. Democratization is a huge departure from the data management practices of the past, which relied solely on data professionals to manage and analyze data — not its users.

But adopting modern data democratization requires modern processes and technologies. That's why so many businesses have started considering a data mesh approach. Data mesh promises a road to data democratization through increased flexibility and agility, enhanced data governance, and the abstraction of technical complexities.

Data mesh is still a relatively new concept. While its beginnings can be tracked to 2019, its popularity has especially grown over the past two years. Google Trends reported around a 300% growth in search volume over the last 18 months. As the pressure to better leverage data assets in order to stay competitive, this widespread interest in data mesh will only continue to grow.

To better understand how can organizations embrace and unlock the data-democratizing power of data mesh let's first address the nuts and bolts of what it is.

## The Definition of "Data Mesh"

As a relatively new term and concept, it's hard to find a generally agreed-upon definition of data mesh. For starters, it's not the same as data fabric. "Data fabric" and "data mesh" often compete for mindshare, but Gartner defined the two as distinctly different in late 2021, stating that:

"A data fabric is the utilization of multiple existing technologies in combination to enable a metadata-driven implementation and augmented orchestration design. A data mesh is a solution architecture that can guide design within a technology-agnostic framework."

To make matters more confusing, we see different segments of the data and analytics market each putting its own spin on the concept of data mesh. Since there isn't a singular definition yet, let's cover a few of the different ways that organizations choose to describe and operationalize the data mesh.

## Varying definitions of data mesh from different layers of the modern data stack:

- **Data Virtualization** companies tend to define data mesh as a virtualization process, as in creating something that looks similar to data federation. In practice, this means different business units build analytics programs on physically disparate data sources and then centrally manage a virtualization solution that enables queries across hybrid and multiple cloud infrastructures.
- **Data Governance** companies, including data catalog solution providers, focus on the governance perspective. This definition of data mesh emphasizes the need to facilitate discoverability and interoperability. It's all about giving ownership to decentralized workgroups and empowering each of them to leverage data and analytics building blocks. This way, each business unit can build its own data products on an as-needed basis.
- **Data Integration and Transformation** providers focus on transformation. This definition emphasizes techniques for managing the decentralization of data and analytics engineering. The goal here is to enable decentralized workgroups to manage data movement and preparation for their own analytics products.

## AtScale's Definition of Data Mesh

While the various flavors of data mesh definition are broadly consistent, they all tend to emphasize elements justifying the need for their particular solutions. So it's our turn, as the leading provider of an independent semantic layer for data and analytics.

We see the semantic layer as an enabling technology for a data mesh strategy. Here are three of the key enabling capabilities of a semantic layer that are also fundamental to data mesh success:

1. It manages the translation of analytics-ready to business-ready data by enabling your data to speak the language of your business.

2. It simplifies the creation of new business-ready views with pre-built, composable building blocks.

3. It is the logical place to apply governance policies that form the guardrails on data usage, ensuring consistency, compliance, and trust.

AtScale specifically ascribes to the definition of data mesh that emphasizes a hub-and-spoke analytics program: centralized governance of data assets, infrastructure, and access with decentralized data product design and creation.

Using this definition, the goals of a data mesh become:

- Defining data domains and aligning with business domains
- Combining data domains with business context to create data products
- Registering data products and making them available for re-use based on business needs
- Creating the data mesh tissue by connecting the data domains via conform dimensions

## How Does an Organization Adopt Data Mesh?

The level of change that an organization must make, in order to adopt data mesh, depends on its relative level of data and analytics maturity. In many cases, the organizational and infrastructure groundwork for data mesh has already been laid. For instance, centralizing data assets on a modern cloud platform is a prerequisite. Likewise, the capability to transform and model data assets for delivering analytics-ready data is fundamental. For organizations that already have a robust data and analytics infrastructure, there are four basic steps needed to adopt data mesh.

## 1. Aligning data domains to business domains

Data mesh relies on the notion that there is a specific business group (i.e. business domain) that's the logical "owner" of a set of data assets (i.e. a data domain). Aligning data domains to business domains sets up the basic rules of which business unit bears responsibility for curating and augmenting which data sets. Without this alignment, an organization will inevitably see conflicts between groups as they assign different significance to the same data. Without clear ownership, no one is responsible for defining the proper usage of a given data set.

## 2. Treating data as a product

Data as a product closely aligns with the definition of data mesh. The idea is that the business owners of a given data domain take responsibility for augmenting raw data with business context, in order to make the data useful to the broadest set of users. Business context may include basic transformations of data to make it more workable for an analytics use case. It may also include augmenting historical data with aggregation logic and hierarchical dimensional logic to support drill-up / drill-down analytics. Data products need to be created with user-friendliness in mind, laser-focused on business users' needs. So data as a product must include the creation of business-oriented views of data that support self-service analysis.

## 3. Embracing Composable Analytics and Shareability

Treating data as a product obviously means positioning data assets for direct analysis by data consumers. But it also means positioning assets for reuse in more complex applications. Domain owners should be registering data products within a catalog and making them available for reuse by other groups in the organization. Some of the most valuable data analytics inherently require blending data across multiple domains. A few examples include...

- Profitability analysis blending revenue data from a CRM system with cost data from financials
- Employee productivity blending HR data with financials
- Prediction models blending sales data with 3rd party data on the economy.

Most importantly, data products need to be discoverable and accessible by different work groups around the organization.

### 4. Building a Connective Tissue

Data mesh relies on the notion that there is a specific business group (i.e. business domain) that's the logical "owner" of a set of data assets (i.e. a data domain). Aligning data domains to business domains sets up the basic rules of which business unit bears responsibility for curating and augmenting which data sets. Without this alignment, an organization will inevitably see conflicts between groups as they assign different significance to the same data. Without clear ownership, no one is responsible for defining the proper usage of a given data set.

## What Does Data Mesh Have to Do With a Semantic Layer?

The balance between maintaining centralized governance while enabling the decentralized creation of data products is easier said than done. That's where AtScale's semantic layer strategy comes in, establishing a "common language" for all business units.

One of the most important components to data mesh success is making your data business-ready. To find out more about what that transformation looks like, check out the next post in our data mesh series.

# Chapter 2: How Data Mesh Success Depends on Business-Ready Data

The modern data-driven organization creates, captures, and manages a massive amount of data from many sources. These data assets hold massive potential to generate business value: reducing costs, gaining operational efficiencies, finding new sources of revenue, and outperforming competition. Realizing this potential depends on transforming raw data assets into something that business domain experts can work with. The concept of a data mesh relies on the ability to efficiently manage this transformation so that disparate work groups can interact with data and build data products that deliver business value. Business-ready data is exactly what it sounds like: data assets that are ready for direct interaction with business domain experts.

Embracing a data mesh approach lets organizations decentralize data product creation to support intelligent decision-making across various business units. But, many make the mistake of supporting their data mesh with complex and loosely governed data pipelines, rather than a holistic strategy for delivering business-ready data to work groups. The rise of cloud data platforms and availability of open source ELT tools like dbt and Apache Airflow empowers data engineers to build and maintain pipelines to deliver subsets of data to workgroups. On the surface, this rise in open source tools seems like it would support data mesh. But, this proliferation of pipelines in practice is complex to maintain, impossible to govern, and constrained by the availability of data engineering resources. To achieve true data mesh success and put the power of creating relevant data products into the hands of your data users, your organization needs a scalable approach to delivering business-ready data.

## What is Business-Ready Data?

We can think of “business-ready data” as the last mile of data transformation before it’s put in the hands of data consumers. Data engineering pipelines transform raw operational data — which is structured to support application logic and transaction capture — into a format that’s more amenable for using query logic to ask questions of the data. This transformed data is often referred to as analytics-ready.

While it's possible to let business analysts, data scientists, and SQL-savvy business managers interact with this analytics-ready data, to do so accurately requires sophisticated data skills, understanding of the business, and the ability to map data concepts to business significance. Interacting with analytics-ready data is not easy and can be easily misinterpreted. Business-ready data, by contrast, is much easier for every business department to use. It removes ambiguity of how to represent important business concepts, and encourages exploration by business domain experts.

## The Rise of Analytics Data Pipelines

The proliferation of data pipelines that support analytics programs coincided with the shift from ETL (Extract-Transform-Load), to ELT (Extract-Load-Transform). ETL and ELT both perform the same function: transforming operational data that's intended for application logic into a structure that's intended to support analytics use cases including business intelligence, exploratory data analysis, and data sciences. They simply perform this function in two different orders. This switch from ETL to ELT happened because of the shift from on-premise to cloud data storage. Today's flexible cloud environments make it possible for data to get transformed more easily, with less coordination with infrastructure teams. Data engineers are able to script and automate transformations with tools like dbt or Airflow, then orchestrate and automate transformations directly in their Snowflake, Databricks, or BigQuery environment.

This capability has created a new generation of data engineers who can support business requests for data sets (or data views), customized for domain specific use. As the business identifies the need or opportunity for a new data product (e.g. report, dashboard, exploratory analysis, AI/ML model, etc.), they request supporting data from their data engineers. Whether these data engineers are embedded within business units or operate from a centrally-managed team, they do their best to translate business requests into transforms that deliver a materialized view of the transformed data. Ensuring that the data is presented in a form that analytics consumers can work with, using their preferred BI or AI/ML tool, is the job of the data engineer. As data consumers articulate new data requirements, data engineers modify existing or create new pipelines delivering new views of data.

The most common type of transformations implemented by data engineers to bring raw data to the point it can be more easily analyzed include:

- De-normalizing tables
- Filtering extraneous data to make the size of data assets more useable
- Establishing an appropriate level of transactional aggregation
- Reducing overall size and complexity of joins
- Blending multiple large data sets with conformed dimensions

# The Challenge of Creating Business-Ready Data

While cloud based ELT has enhanced the agility of data teams, the proliferation of data pipelines still poses a challenge. Even with careful implementation of shareable code repositories and standards, the reliance on a team of data engineers to satisfy the needs of data consumers can become challenging. In particular, it's hard for data engineers to build business-ready data sets that are ready for analysis without further manipulation.

Building on the common data engineering transforms mentioned above, creating business ready data may include:

- Building a business-oriented view of data for data consumers.
- Creating a fully-vetted definition for each metric. They will then serve as a single source of truth for key business metrics like "revenue."
- Establishing controlled, conformed dimensions. This leaves no chance of misrepresenting an important concept like a fiscal quarter or sales territory.
- Making blended data sets by augmenting historical data with third-party data or AI-generated predicted data sets.

Business-ready data provides data consumers with specific, rich data that can be directly used to answer their questions. It includes the right context, format, naming, and other clarifying factors to enable business users to work with data assets directly. But, it can't be put into motion by the analytics or engineering teams alone. That would just slow down the process — and still make it difficult for data to reflect the exact business moment at the right time. Instead, this transformation process has to be flexible and agile. The best way to do this is by giving the responsibility of translating analytics-ready data into business-ready data to the users themselves — the people who live and breathe the changing condition of the business on a daily basis.

# How a Data Mesh Approach and Business-Ready Data Work Together

But, empowering so many users to translate analytics-ready data into business-ready data, in a scalable fashion, can be complex. This is where the concept of a data mesh comes into play. The data mesh approach facilitates a decentralized analytics architecture where business domains are responsible for their data. A big part of this approach revolves around creating data as a product, owned and designed by business units. Ideally, distributed teams directly interact with data assets to build their own data products — BI assets like dashboards, ad hoc assets like a pivot table or excel model, or AI assets managed in a notebook.

But this autonomy also has to come with some form of central standardization. So, a successful data mesh practice requires a balance between centralized governance and autonomy.

## The Value of Managing Business-Ready Data in a Semantic Layer

How do organizations produce business-ready data like this — somehow balancing this need for agility and autonomy with governance? This answer is a semantic layer.

A semantic layer empowers business analysts to uphold organization-wide governance by defining metrics such as revenue, and inventory with conformed dimensions such as fiscal year, product and customer. Then, the data consumers can use this pre-established framework to respond to each business moment as needed. The semantic layer then enables teams to easily create data products. Once the business-ready data is available in the semantic layer platform, it can immediately be consumed by the analytics and AI consumers via the BI tool of their choice.

# Chapter 3: How to Implement a Composable Analytics Strategy that Supports your Data Mesh

Organizations establish data mesh in order to create more widespread, flexible data access across the entire business. And in order to achieve this, a data mesh approach must enable domain experts to build their own data products. This idea of self-service data products is fundamental to data mesh.

The best way to achieve this is by using "composable analytics" — a library of centrally-governed building blocks for composing new data products. This standard library of assets needs to be created with both shareability and reusability in mind. In today's blog post, we'll be covering why this process matters to a data mesh approach and how to accomplish it with a universal semantic layer.

## How Composable Analytics Relate to Data Mesh

Gartner notes this dilemma:

"The rise of business technologists means greater demand for self-service capabilities and faster delivery of analytics solutions. Fixation on self-service has created governance issues, such as a change of roles and processes toward greater collaboration between decentralized data and analytics (D&A) and IT. Siloed self-service analytics fail to effectively reuse the value created within accumulated information assets. It is unfortunately typical that many similar analytics outputs have to be built from scratch, wasting a lot of "reinventing-the-wheel" efforts."

The solution? Gartner recommends modular, analytics building blocks — also known as "composable analytics." It's all about establishing standard processes, and, in Gartner's words, "incubating popular and reusable steps in the self-service analytics process and registering them in an analytics catalog for future composition."

But all of this can be challenging to achieve. It takes organizational-wide standardization to make these "building blocks" interoperable and flexible enough. And they have to be versatile — able to be shared and reused across siloed departments.

# The Common Building Blocks of a Composable Analytics Strategy

In our first two chapters, we covered AtScale's perspective of data mesh. We mentioned how a semantic layer enables the creation of business-ready data, creating a balance between centralized governance and flexible, independent data usage through a "common language."

It turns out that a semantic layer can solve this other data mesh challenge: the challenge of creating composable analytics that enables organizations to create and share common "building blocks." A semantic layer enables the creation, maintenance, and governance of three shareable and reusable components: Metrics, Conformed Dimensions, and Models. These three objects serve as the base "building blocks" for all teams to use, leading to a true data mesh approach. Here's how work groups use each of these objects to compose new data products within their domain-specific perspectives:

### 1. Pre-built, Reusable, Governed Metrics

A semantic layer enables consistent versions of primary business KPIs, such as ARR, ship quantity, etc. This nomenclature eliminates the chance of different groups presenting different versions of the truth. It also provides business-vetted calculated metrics, created by the right subject matter experts within the organization. An example could be gross margin, as defined by finance. Once defined, these metrics are searchable within the semantic layer, able to be consumed and implicitly trusted by others. A semantic layer enables organizations to create time-relative metrics as well, eliminating the need to rebuild complex calculations such as quarter-on-quarter growth.

### 2. Pre-built, Reusable, Conformed Dimensions

A conformed dimension has the same meaning across various data contexts - bringing standardization to important dimensions common to multiple data formats (i.e. from different application sources). They allow measures to be categorized and described in the same way across multiple facts — a "master dimension", in other words. Their content has been agreed upon across the organization so they can be used to prove business-wide compliance. A few examples of conformed dimensions include time, product, geography, and customer. These conformed dimensions allow reusable aggregation paths for measures across multiple fact tables.

Conformed dimensions are ultimately the "connective tissue" that enables data modelers to blend disparate data sets. This means that everyone in the organization has the same way to drill down in the data and with a single version of truth.

As an example, a business that's analyzing time-based outputs could drill down from annual-level to quarterly-level data, or drill up from quarterly to annually, or anything in between. Dimensions provide the flexibility that teams need in order to connect data sets and contextualize data and analysis.

### 3. Pre-built, Reusable Models

A semantic model is a logical definition of a new view of data, based on data assets within a centralized data platform. Models enable businesses to further define analytics-ready views of raw data assets, simplifying data interaction for data consumers.

This analytics-ready data can simplify highly normalized data sets. It can also create a view of blended data assets, such as a composite of CRM and finance data or a view of marketing and economic data sourced by a third party. A reusable reference model, defining how two or more data sets get blended, establishes a formulaic approach to blending data sets. It can then be used across multiple end data products, saving time and effort on a normally-complex process. An intelligent semantic layer can help with models as well. As users compose different objects (measures, dimensions) to create a model, the semantic layer automatically provides suggestions on how to connect these objects to create the full data model.

## Shareability and Reusability: Key Tenets of Data Mesh

A data mesh approach is all about shareability and reusability. This is why composable analytics is an important aspect of setting up a data mesh within your organization. It encourages the sharing and reuse of analytics building blocks across work groups. Composable analytics also enable a hub-and-spoke model: striking a balance between centralized governance and independent data product creation for each business unit.

To create standard governance guardrails, centralized teams maintain fundamental building blocks, such as conformed dimensions and basic views of analysis-ready data. But at the same time, work groups are allowed to build and govern domain-specific elements of models, minimizing the amount of duplicative work and the risk of mistakes. And by taking a data mesh approach with composable analytics, your business analysts get to spend more of their time analyzing data, rather than waiting on data specialists to create a dimensional data model. With composable analytics, they can easily find the components to build a business ready view of the data, then start analyzing it right away with trust.

# Semantic Layer Platform

A semantic layer platform like AtScale snaps seamlessly into your data tech stack, transforming all types of data into standardized, composable analytics. We enable organizations to manage shareable elements centrally while simultaneously empowering data modelers embedded with work groups to create new data products independently.

# Chapter 4: Building Guardrails into your Data Mesh with Analytics Governance

A successful data mesh approach empowers domain experts to build their own data products with centrally-governed building blocks. And to enable this model, organizations need to facilitate the transformation of analytics-ready data into business-ready data. Then, they need to make this data standardized, sharable, and reusable for a variety of different data products — like a decision support dashboard in a BI tool like PowerBI or Tableau, an exploratory analysis in Excel or Python, an AI/ML model, etc.

But there's another piece to the data mesh puzzle: analytics governance. Governance strikes a balance by empowering self-service data product creation while maintaining a consistent level of security and quality. After all, there are correct and incorrect ways to interact with data. And a data mesh approach focuses on empowering non-data experts to interact with data correctly. Because of this, many users need a way to answer questions like, "am I using the right data?" or "is the data I'm using quality data?" Governance plays this role, empowering all data users with varied levels of expertise to confidently build and maintain high-quality data products, and leading to other innovations such as self-service BI. Analytics governance falls into a few different categories:

- Governance of Data Sources
- Governance of Metrics
- Governance of Dimensions
- Governance of Access Control
- Governance of Cloud Resource Consumption and Analytics Performance

## Governance of Data Sources

Data source governance means setting standards for your application data, 3rd party data, and data lineage. It sets clear rules for how this data gets used, modified, and stored within your organization. This entails cleaning up duplicate/unnecessary data and giving each data source a detailed metadata description that's visible to users.

Many organizations achieve data governance by implementing active governance, meaning that policies get baked into everyday workflows. As an example, active governance makes data lineage metadata accessible from each team's day-to-day business applications. This accessibility empowers team members to consume data with trust as they get their various jobs done. By putting individual data users in control, active data governance builds a community of business experts committed to data literacy. This method of data source governance also ensures that all data is high-quality, trusted, and compliant from the start. It's more effective than the passive approaches of the past, which focused on ingesting a high volume of data, then applying rigid business rules to it retroactively.

The semantic layer facilitates active data source governance. It empowers business owners with the right data domain context, enabling them to continuously improve data definitions in an agile and flexible way.

## Governance of Metrics

The governance of metrics — whether simple like revenue, cost, and quantities or complex like Annual Recurring Revenue (ARR) and Customer Acquisition Cost (CAC) — ensures a single version of the truth. Consistent metrics serve as the connecting language between the analytics layer and business: the everyday, concrete terms used to explain business-essential concepts across the organization.

When an organization doesn't set up standard definitions for metrics, each business unit will inevitably create separate definitions and calculation logic for the same metrics. This discrepancy happens because different BI and analytics tools with different calculation pipelines get adopted across decentralized teams.

The problem of inconsistent metrics explains why we've seen the rise of metric stores in recent years. Businesses see the need for a stand-alone layer that sits between data warehouses and downstream data consumers, ensuring consistency in metrics usage. A semantic layer creates the foundation for a metrics store by laying out a "common language" and definition for all metrics to follow.

## Governance of Dimensions

Dimension governance also plays an important role in facilitating composability for a successful data mesh approach. This means standardizing conformed analysis dimensions such as time, geography, and product.

These dimensions serve as the building blocks for data products and the common medium for blending disparate data sets into usable products.

Similarly to metrics governance, the semantic layer can also provide a “common language” for different domain experts to use. These conformed dimensions serve as the connective tissue for stitching together different data sources: the heart of a true data mesh approach.

## Governance of Access Control

Access governance is a security function as well as a tool for facilitating quality data products. It ensures that only the appropriate users get granted access to a critical asset. As the medium between data sources and stores, the semantic layer is the natural place to enforce access control policies, ensuring that the right user has access to the right data assets at the right time. It should provide row and column-level security and it integrates with the source data access controls.

## Governance of Cloud Resource Consumption and Analytics Performance

Most (if not all) enterprises are rushing to move data to the cloud. But with all of its benefits, the cloud also brings new challenges. Most recently, one of the biggest hurdles to cloud growth is a push for cost governance. Because cloud storage gets priced as a service, rather than bought once as a physical asset (as with on-premise storage), mounting cloud costs can catch an organization off-guard.

A new study uncovered that 81% of IT leaders have been directed to reduce or take on no additional cloud spending. So, it only makes sense that financial governance comes into play, alongside other forms of analytics governance.

Because all analytics consumption naturally passes through the semantic layer, it can provide greater visibility into usage and can be a platform for implementing usage controls. A semantic layer can also optimize consumption. A universal, stand-alone layer can monitor all of the analytics queries, detect patterns in the most commonly-asked business questions, then create aggregated data for these questions by keeping the data at its source.

Then if a similar question gets asked, the semantic engine leverages the aggregated data, improving the query performance and reducing execution time. This added efficiency saves time and resources, leading to overall lower cloud costs for the entire organization.

# The Power of an Analytics Governance to Support Data Mesh

Each level of analytics governance — data sources, metrics, dimensions, and cloud resource consumption/analytics performance — plays a significant role in enabling an effective data mesh. These guardrails lead to a "building block assembly line" of sorts, empowering users across the entire organization to create high-quality data products. And a well-built data mesh approach sets the stage for a variety of new innovations, from fully-realized data democratization, to cost and resource optimization, to stronger data-driven decision-making, and so much more.

Elif Tutuk, Global Head of Product at AtScale

The career of **Elif Tutuk** includes more than 15 years of experience in Business Intelligence (BI) and analytics. Her innovations have led to patents for search and conversational analytics, data analysis, data management, and more. Her research and technology development for augmented intelligence (a combination of data science and AI) has led to the rise of third-generation analytics. This is where BI, AI, and machine learning come together to enhance analytics across the data lifecycle – from how data is prepared to how analysis is performed and insights are delivered.

Her success as a product leader has been quite diversified; she's worked in product management, technical product marketing, design, product development, and research.

In her recent role before joining AtScale, as Qlik's Vice President of Innovation and Design, she oversaw a global team of UX designers, product designers, and engineers in planning and executing an innovation roadmap of cloud data integration and analytics products.

Elif's work in data and analytics world has empowered people in virtually any job to get value from data, allowing them to ask questions and automatically generate insights in an easy, conversational manner. The technology she led and innovated has allowed thousands of users to understand and use data more effectively and enabled data-driven decisions without bias, enabling businesses to deploy ethical strategies that deliver results without sacrificing integrity. From healthcare and supply chain challenges to climate control and beyond, Elif has worked tirelessly to build solutions that allow people and their organizations to do more with data.

# Introduction

A data mesh seeks to democratize data and empower self-service business intelligence. The principles of a data mesh approach make it simple for domain owners — regardless of their level of data expertise — to use, modify, and store data in a way that makes sense for their specific business needs. It does this by enabling these domain experts to build their own data products with pre-established building blocks. Data mesh transforms untapped, disparate data stores into a connective tissue that's accessible and usable for the entire organization. And it enables this autonomy while also upkeeping universal governance. This ensures that the data gets used in the correct ways.

This white paper is broken down into four chapters, covering:

1. The Definition of Data Mesh: What Is It and Why Do I Need One?
2. Data Mesh Success Depends on Business-Ready Data
3. How to Implement a Composable Analytics Strategy that Supports your Data Mesh
4. Building Guardrails into your Data Mesh with Analytics Governance

# Chapter 1: The Definition of Data Mesh: What Is It and Why Do I Need One?

Today's leading businesses prioritize data mesh. The modern data mesh craze ultimately came about to streamline data democratization – that is, making the right data available to the right people at the right time. Democratization is a huge departure from the data management practices of the past, which relied solely on data professionals to manage and analyze data — not its users.

But adopting modern data democratization requires modern processes and technologies. That's why so many businesses have started considering a data mesh approach. Data mesh promises a road to data democratization through increased flexibility and agility, enhanced data governance, and the abstraction of technical complexities.

Data mesh is still a relatively new concept. While its beginnings can be tracked to 2019, its popularity has especially grown over the past two years. Google Trends reported around a 300% growth in search volume over the last 18 months. As the pressure to better leverage data assets in order to stay competitive, this widespread interest in data mesh will only continue to grow.

To better understand how can organizations embrace and unlock the data-democratizing power of data mesh let's first address the nuts and bolts of what it is.

## The Definition of "Data Mesh"

As a relatively new term and concept, it's hard to find a generally agreed-upon definition of data mesh. For starters, it's not the same as data fabric. "Data fabric" and "data mesh" often compete for mindshare, but Gartner defined the two as distinctly different in late 2021, stating that:

"A data fabric is the utilization of multiple existing technologies in combination to enable a metadata-driven implementation and augmented orchestration design. A data mesh is a solution architecture that can guide design within a technology-agnostic framework."

To make matters more confusing, we see different segments of the data and analytics market each putting its own spin on the concept of data mesh. Since there isn't a singular definition yet, let's cover a few of the different ways that organizations choose to describe and operationalize the data mesh.

## Varying definitions of data mesh from different layers of the modern data stack:

- **Data Virtualization** companies tend to define data mesh as a virtualization process, as in creating something that looks similar to data federation. In practice, this means different business units build analytics programs on physically disparate data sources and then centrally manage a virtualization solution that enables queries across hybrid and multiple cloud infrastructures.
- **Data Governance** companies, including data catalog solution providers, focus on the governance perspective. This definition of data mesh emphasizes the need to facilitate discoverability and interoperability. It's all about giving ownership to decentralized workgroups and empowering each of them to leverage data and analytics building blocks. This way, each business unit can build its own data products on an as-needed basis.
- **Data Integration and Transformation** providers focus on transformation. This definition emphasizes techniques for managing the decentralization of data and analytics engineering. The goal here is to enable decentralized workgroups to manage data movement and preparation for their own analytics products.

## AtScale's Definition of Data Mesh

While the various flavors of data mesh definition are broadly consistent, they all tend to emphasize elements justifying the need for their particular solutions. So it's our turn, as the leading provider of an independent semantic layer for data and analytics.

We see the semantic layer as an enabling technology for a data mesh strategy. Here are three of the key enabling capabilities of a semantic layer that are also fundamental to data mesh success:

1. It manages the translation of analytics-ready to business-ready data by enabling your data to speak the language of your business.

2. It simplifies the creation of new business-ready views with pre-built, composable building blocks.

3. It is the logical place to apply governance policies that form the guardrails on data usage, ensuring consistency, compliance, and trust.

AtScale specifically ascribes to the definition of data mesh that emphasizes a hub-and-spoke analytics program: centralized governance of data assets, infrastructure, and access with decentralized data product design and creation.

Using this definition, the goals of a data mesh become:

- Defining data domains and aligning with business domains
- Combining data domains with business context to create data products
- Registering data products and making them available for re-use based on business needs
- Creating the data mesh tissue by connecting the data domains via conform dimensions

## How Does an Organization Adopt Data Mesh?

The level of change that an organization must make, in order to adopt data mesh, depends on its relative level of data and analytics maturity. In many cases, the organizational and infrastructure groundwork for data mesh has already been laid. For instance, centralizing data assets on a modern cloud platform is a prerequisite. Likewise, the capability to transform and model data assets for delivering analytics-ready data is fundamental. For organizations that already have a robust data and analytics infrastructure, there are four basic steps needed to adopt data mesh.

## 1. Aligning data domains to business domains

Data mesh relies on the notion that there is a specific business group (i.e. business domain) that's the logical "owner" of a set of data assets (i.e. a data domain). Aligning data domains to business domains sets up the basic rules of which business unit bears responsibility for curating and augmenting which data sets. Without this alignment, an organization will inevitably see conflicts between groups as they assign different significance to the same data. Without clear ownership, no one is responsible for defining the proper usage of a given data set.

## 2. Treating data as a product

Data as a product closely aligns with the definition of data mesh. The idea is that the business owners of a given data domain take responsibility for augmenting raw data with business context, in order to make the data useful to the broadest set of users. Business context may include basic transformations of data to make it more workable for an analytics use case. It may also include augmenting historical data with aggregation logic and hierarchical dimensional logic to support drill-up / drill-down analytics. Data products need to be created with user-friendliness in mind, laser-focused on business users' needs. So data as a product must include the creation of business-oriented views of data that support self-service analysis.

## 3. Embracing Composable Analytics and Shareability

Treating data as a product obviously means positioning data assets for direct analysis by data consumers. But it also means positioning assets for reuse in more complex applications. Domain owners should be registering data products within a catalog and making them available for reuse by other groups in the organization. Some of the most valuable data analytics inherently require blending data across multiple domains. A few examples include...

- Profitability analysis blending revenue data from a CRM system with cost data from financials
- Employee productivity blending HR data with financials
- Prediction models blending sales data with 3rd party data on the economy.

Most importantly, data products need to be discoverable and accessible by different work groups around the organization.

### 4. Building a Connective Tissue

Data mesh relies on the notion that there is a specific business group (i.e. business domain) that's the logical "owner" of a set of data assets (i.e. a data domain). Aligning data domains to business domains sets up the basic rules of which business unit bears responsibility for curating and augmenting which data sets. Without this alignment, an organization will inevitably see conflicts between groups as they assign different significance to the same data. Without clear ownership, no one is responsible for defining the proper usage of a given data set.

## What Does Data Mesh Have to Do With a Semantic Layer?

The balance between maintaining centralized governance while enabling the decentralized creation of data products is easier said than done. That's where AtScale's semantic layer strategy comes in, establishing a "common language" for all business units.

One of the most important components to data mesh success is making your data business-ready. To find out more about what that transformation looks like, check out the next post in our data mesh series.

# Chapter 2: How Data Mesh Success Depends on Business-Ready Data

The modern data-driven organization creates, captures, and manages a massive amount of data from many sources. These data assets hold massive potential to generate business value: reducing costs, gaining operational efficiencies, finding new sources of revenue, and outperforming competition. Realizing this potential depends on transforming raw data assets into something that business domain experts can work with. The concept of a data mesh relies on the ability to efficiently manage this transformation so that disparate work groups can interact with data and build data products that deliver business value. Business-ready data is exactly what it sounds like: data assets that are ready for direct interaction with business domain experts.

Embracing a data mesh approach lets organizations decentralize data product creation to support intelligent decision-making across various business units. But, many make the mistake of supporting their data mesh with complex and loosely governed data pipelines, rather than a holistic strategy for delivering business-ready data to work groups. The rise of cloud data platforms and availability of open source ELT tools like dbt and Apache Airflow empowers data engineers to build and maintain pipelines to deliver subsets of data to workgroups. On the surface, this rise in open source tools seems like it would support data mesh. But, this proliferation of pipelines in practice is complex to maintain, impossible to govern, and constrained by the availability of data engineering resources. To achieve true data mesh success and put the power of creating relevant data products into the hands of your data users, your organization needs a scalable approach to delivering business-ready data.

## What is Business-Ready Data?

We can think of "business-ready data" as the last mile of data transformation before it's put in the hands of data consumers. Data engineering pipelines transform raw operational data — which is structured to support application logic and transaction capture — into a format that's more amenable for using query logic to ask questions of the data. This transformed data is often referred to as analytics-ready.

While it's possible to let business analysts, data scientists, and SQL-savvy business managers interact with this analytics-ready data, to do so accurately requires sophisticated data skills, understanding of the business, and the ability to map data concepts to business significance. Interacting with analytics-ready data is not easy and can be easily misinterpreted. Business-ready data, by contrast, is much easier for every business department to use. It removes ambiguity of how to represent important business concepts, and encourages exploration by business domain experts.

## The Rise of Analytics Data Pipelines

The proliferation of data pipelines that support analytics programs coincided with the shift from ETL (Extract-Transform-Load), to ELT (Extract-Load-Transform). ETL and ELT both perform the same function: transforming operational data that's intended for application logic into a structure that's intended to support analytics use cases including business intelligence, exploratory data analysis, and data sciences. They simply perform this function in two different orders. This switch from ETL to ELT happened because of the shift from on-premise to cloud data storage. Today's flexible cloud environments make it possible for data to get transformed more easily, with less coordination with infrastructure teams. Data engineers are able to script and automate transformations with tools like dbt or Airflow, then orchestrate and automate transformations directly in their Snowflake, Databricks, or BigQuery environment.

This capability has created a new generation of data engineers who can support business requests for data sets (or data views), customized for domain specific use. As the business identifies the need or opportunity for a new data product (e.g. report, dashboard, exploratory analysis, AI/ML model, etc.), they request supporting data from their data engineers. Whether these data engineers are embedded within business units or operate from a centrally-managed team, they do their best to translate business requests into transforms that deliver a materialized view of the transformed data. Ensuring that the data is presented in a form that analytics consumers can work with, using their preferred BI or AI/ML tool, is the job of the data engineer. As data consumers articulate new data requirements, data engineers modify existing or create new pipelines delivering new views of data.

The most common type of transformations implemented by data engineers to bring raw data to the point it can be more easily analyzed include:

- De-normalizing tables
- Filtering extraneous data to make the size of data assets more useable
- Establishing an appropriate level of transactional aggregation
- Reducing overall size and complexity of joins
- Blending multiple large data sets with conformed dimensions

# The Challenge of Creating Business-Ready Data

While cloud based ELT has enhanced the agility of data teams, the proliferation of data pipelines still poses a challenge. Even with careful implementation of shareable code repositories and standards, the reliance on a team of data engineers to satisfy the needs of data consumers can become challenging. In particular, it's hard for data engineers to build business-ready data sets that are ready for analysis without further manipulation.

Building on the common data engineering transforms mentioned above, creating business ready data may include:

- Building a business-oriented view of data for data consumers.
- Creating a fully-vetted definition for each metric. They will then serve as a single source of truth for key business metrics like "revenue."
- Establishing controlled, conformed dimensions. This leaves no chance of misrepresenting an important concept like a fiscal quarter or sales territory.
- Making blended data sets by augmenting historical data with third-party data or AI-generated predicted data sets.

Business-ready data provides data consumers with specific, rich data that can be directly used to answer their questions. It includes the right context, format, naming, and other clarifying factors to enable business users to work with data assets directly. But, it can't be put into motion by the analytics or engineering teams alone. That would just slow down the process — and still make it difficult for data to reflect the exact business moment at the right time. Instead, this transformation process has to be flexible and agile. The best way to do this is by giving the responsibility of translating analytics-ready data into business-ready data to the users themselves — the people who live and breathe the changing condition of the business on a daily basis.

# How a Data Mesh Approach and Business-Ready Data Work Together

But, empowering so many users to translate analytics-ready data into business-ready data, in a scalable fashion, can be complex. This is where the concept of a data mesh comes into play. The data mesh approach facilitates a decentralized analytics architecture where business domains are responsible for their data. A big part of this approach revolves around creating data as a product, owned and designed by business units. Ideally, distributed teams directly interact with data assets to build their own data products — BI assets like dashboards, ad hoc assets like a pivot table or excel model, or AI assets managed in a notebook.

But this autonomy also has to come with some form of central standardization. So, a successful data mesh practice requires a balance between centralized governance and autonomy.

## The Value of Managing Business-Ready Data in a Semantic Layer

How do organizations produce business-ready data like this — somehow balancing this need for agility and autonomy with governance? This answer is a semantic layer.

A semantic layer empowers business analysts to uphold organization-wide governance by defining metrics such as revenue, and inventory with conformed dimensions such as fiscal year, product and customer. Then, the data consumers can use this pre-established framework to respond to each business moment as needed. The semantic layer then enables teams to easily create data products. Once the business-ready data is available in the semantic layer platform, it can immediately be consumed by the analytics and AI consumers via the BI tool of their choice.

# Chapter 3: How to Implement a Composable Analytics Strategy that Supports your Data Mesh

Organizations establish data mesh in order to create more widespread, flexible data access across the entire business. And in order to achieve this, a data mesh approach must enable domain experts to build their own data products. This idea of self-service data products is fundamental to data mesh.

The best way to achieve this is by using “composable analytics” — a library of centrally-governed building blocks for composing new data products. This standard library of assets needs to be created with both shareability and reusability in mind. In today’s blog post, we’ll be covering why this process matters to a data mesh approach and how to accomplish it with a universal semantic layer.

## How Composable Analytics Relate to Data Mesh

Gartner notes this dilemma:

“The rise of business technologists means greater demand for self-service capabilities and faster delivery of analytics solutions. Fixation on self-service has created governance issues, such as a change of roles and processes toward greater collaboration between decentralized data and analytics (D&A) and IT. Siloed self-service analytics fail to effectively reuse the value created within accumulated information assets. It is unfortunately typical that many similar analytics outputs have to be built from scratch, wasting a lot of “reinventing-the-wheel” efforts.”

The solution? Gartner recommends modular, analytics building blocks — also known as “composable analytics.” It’s all about establishing standard processes, and, in Gartner’s words, “incubating popular and reusable steps in the self-service analytics process and registering them in an analytics catalog for future composition.”

But all of this can be challenging to achieve. It takes organizational-wide standardization to make these “building blocks” interoperable and flexible enough. And they have to be versatile — able to be shared and reused across siloed departments.

# The Common Building Blocks of a Composable Analytics Strategy

In our first two chapters, we covered AtScale's perspective of data mesh. We mentioned how a semantic layer enables the creation of business-ready data, creating a balance between centralized governance and flexible, independent data usage through a "common language."

It turns out that a semantic layer can solve this other data mesh challenge: the challenge of creating composable analytics that enables organizations to create and share common "building blocks." A semantic layer enables the creation, maintenance, and governance of three shareable and reusable components: Metrics, Conformed Dimensions, and Models. These three objects serve as the base "building blocks" for all teams to use, leading to a true data mesh approach. Here's how work groups use each of these objects to compose new data products within their domain-specific perspectives:

## 1. Pre-built, Reusable, Governed Metrics

A semantic layer enables consistent versions of primary business KPIs, such as ARR, ship quantity, etc. This nomenclature eliminates the chance of different groups presenting different versions of the truth. It also provides business-vetted calculated metrics, created by the right subject matter experts within the organization. An example could be gross margin, as defined by finance. Once defined, these metrics are searchable within the semantic layer, able to be consumed and implicitly trusted by others. A semantic layer enables organizations to create time-relative metrics as well, eliminating the need to rebuild complex calculations such as quarter-on-quarter growth.

## 2. Pre-built, Reusable, Conformed Dimensions

A conformed dimension has the same meaning across various data contexts - bringing standardization to important dimensions common to multiple data formats (i.e. from different application sources). They allow measures to be categorized and described in the same way across multiple facts — a "master dimension", in other words. Their content has been agreed upon across the organization so they can be used to prove business-wide compliance. A few examples of conformed dimensions include time, product, geography, and customer. These conformed dimensions allow reusable aggregation paths for measures across multiple fact tables.

Conformed dimensions are ultimately the "connective tissue" that enables data modelers to blend disparate data sets. This means that everyone in the organization has the same way to drill down in the data and with a single version of truth.

As an example, a business that's analyzing time-based outputs could drill down from annual-level to quarterly-level data, or drill up from quarterly to annually, or anything in between. Dimensions provide the flexibility that teams need in order to connect data sets and contextualize data and analysis.

### 3. Pre-built, Reusable Models

A semantic model is a logical definition of a new view of data, based on data assets within a centralized data platform. Models enable businesses to further define analytics-ready views of raw data assets, simplifying data interaction for data consumers.

This analytics-ready data can simplify highly normalized data sets. It can also create a view of blended data assets, such as a composite of CRM and finance data or a view of marketing and economic data sourced by a third party. A reusable reference model, defining how two or more data sets get blended, establishes a formulaic approach to blending data sets. It can then be used across multiple end data products, saving time and effort on a normally-complex process. An intelligent semantic layer can help with models as well. As users compose different objects (measures, dimensions) to create a model, the semantic layer automatically provides suggestions on how to connect these objects to create the full data model.

## Shareability and Reusability: Key Tenets of Data Mesh

A data mesh approach is all about shareability and reusability. This is why composable analytics is an important aspect of setting up a data mesh within your organization. It encourages the sharing and reuse of analytics building blocks across work groups. Composable analytics also enable a hub-and-spoke model: striking a balance between centralized governance and independent data product creation for each business unit.

To create standard governance guardrails, centralized teams maintain fundamental building blocks, such as conformed dimensions and basic views of analysis-ready data. But at the same time, work groups are allowed to build and govern domain-specific elements of models, minimizing the amount of duplicative work and the risk of mistakes. And by taking a data mesh approach with composable analytics, your business analysts get to spend more of their time analyzing data, rather than waiting on data specialists to create a dimensional data model. With composable analytics, they can easily find the components to build a business ready view of the data, then start analyzing it right away with trust.

# Semantic Layer Platform

A semantic layer platform like AtScale snaps seamlessly into your data tech stack, transforming all types of data into standardized, composable analytics. We enable organizations to manage shareable elements centrally while simultaneously empowering data modelers embedded with work groups to create new data products independently.

# Chapter 4: Building Guardrails into your Data Mesh with Analytics Governance

A successful data mesh approach empowers domain experts to build their own data products with centrally-governed building blocks. And to enable this model, organizations need to facilitate the transformation of analytics-ready data into business-ready data. Then, they need to make this data standardized, sharable, and reusable for a variety of different data products — like a decision support dashboard in a BI tool like PowerBI or Tableau, an exploratory analysis in Excel or Python, an AI/ML model, etc.

But there's another piece to the data mesh puzzle: analytics governance. Governance strikes a balance by empowering self-service data product creation while maintaining a consistent level of security and quality. After all, there are correct and incorrect ways to interact with data. And a data mesh approach focuses on empowering non-data experts to interact with data correctly. Because of this, many users need a way to answer questions like, "am I using the right data?" or "is the data I'm using quality data?" Governance plays this role, empowering all data users with varied levels of expertise to confidently build and maintain high-quality data products, and leading to other innovations such as self-service BI. Analytics governance falls into a few different categories:

- Governance of Data Sources
- Governance of Metrics
- Governance of Dimensions
- Governance of Access Control
- Governance of Cloud Resource Consumption and Analytics Performance

## Governance of Data Sources

Data source governance means setting standards for your application data, 3rd party data, and data lineage. It sets clear rules for how this data gets used, modified, and stored within your organization. This entails cleaning up duplicate/unnecessary data and giving each data source a detailed metadata description that's visible to users.

Many organizations achieve data governance by implementing active governance, meaning that policies get baked into everyday workflows. As an example, active governance makes data lineage metadata accessible from each team's day-to-day business applications. This accessibility empowers team members to consume data with trust as they get their various jobs done. By putting individual data users in control, active data governance builds a community of business experts committed to data literacy. This method of data source governance also ensures that all data is high-quality, trusted, and compliant from the start. It's more effective than the passive approaches of the past, which focused on ingesting a high volume of data, then applying rigid business rules to it retroactively.

The semantic layer facilitates active data source governance. It empowers business owners with the right data domain context, enabling them to continuously improve data definitions in an agile and flexible way.

## Governance of Metrics

The governance of metrics — whether simple like revenue, cost, and quantities or complex like Annual Recurring Revenue (ARR) and Customer Acquisition Cost (CAC) — ensures a single version of the truth. Consistent metrics serve as the connecting language between the analytics layer and business: the everyday, concrete terms used to explain business-essential concepts across the organization.

When an organization doesn't set up standard definitions for metrics, each business unit will inevitably create separate definitions and calculation logic for the same metrics. This discrepancy happens because different BI and analytics tools with different calculation pipelines get adopted across decentralized teams.

The problem of inconsistent metrics explains why we've seen the rise of metric stores in recent years. Businesses see the need for a stand-alone layer that sits between data warehouses and downstream data consumers, ensuring consistency in metrics usage. A semantic layer creates the foundation for a metrics store by laying out a "common language" and definition for all metrics to follow.

## Governance of Dimensions

Dimension governance also plays an important role in facilitating composability for a successful data mesh approach. This means standardizing conformed analysis dimensions such as time, geography, and product.

These dimensions serve as the building blocks for data products and the common medium for blending disparate data sets into usable products.

Similarly to metrics governance, the semantic layer can also provide a “common language” for different domain experts to use. These conformed dimensions serve as the connective tissue for stitching together different data sources: the heart of a true data mesh approach.

## Governance of Access Control

Access governance is a security function as well as a tool for facilitating quality data products. It ensures that only the appropriate users get granted access to a critical asset. As the medium between data sources and stores, the semantic layer is the natural place to enforce access control policies, ensuring that the right user has access to the right data assets at the right time. It should provide row and column-level security and it integrates with the source data access controls.

## Governance of Cloud Resource Consumption and Analytics Performance

Most (if not all) enterprises are rushing to move data to the cloud. But with all of its benefits, the cloud also brings new challenges. Most recently, one of the biggest hurdles to cloud growth is a push for cost governance. Because cloud storage gets priced as a service, rather than bought once as a physical asset (as with on-premise storage), mounting cloud costs can catch an organization off-guard.

A new study uncovered that 81% of IT leaders have been directed to reduce or take on no additional cloud spending. So, it only makes sense that financial governance comes into play, alongside other forms of analytics governance.

Because all analytics consumption naturally passes through the semantic layer, it can provide greater visibility into usage and can be a platform for implementing usage controls. A semantic layer can also optimize consumption. A universal, stand-alone layer can monitor all of the analytics queries, detect patterns in the most commonly-asked business questions, then create aggregated data for these questions by keeping the data at its source.

Then if a similar question gets asked, the semantic engine leverages the aggregated data, improving the query performance and reducing execution time. This added efficiency saves time and resources, leading to overall lower cloud costs for the entire organization.

# The Power of an Analytics Governance to Support Data Mesh

Each level of analytics governance — data sources, metrics, dimensions, and cloud resource consumption/analytics performance — plays a significant role in enabling an effective data mesh. These guardrails lead to a “building block assembly line” of sorts, empowering users across the entire organization to create high-quality data products. And a well-built data mesh approach sets the stage for a variety of new innovations, from fully-realized data democratization, to cost and resource optimization, to stronger data-driven decision-making, and so much more.

Elif Tutuk, Global Head of Product at AtScale

The career of **Elif Tutuk** includes more than 15 years of experience in Business Intelligence (BI) and analytics. Her innovations have led to patents for search and conversational analytics, data analysis, data management, and more. Her research and technology development for augmented intelligence (a combination of data science and AI) has led to the rise of third-generation analytics. This is where BI, AI, and machine learning come together to enhance analytics across the data lifecycle – from how data is prepared to how analysis is performed and insights are delivered.

Her success as a product leader has been quite diversified; she's worked in product management, technical product marketing, design, product development, and research.

In her recent role before joining AtScale, as Qlik's Vice President of Innovation and Design, she oversaw a global team of UX designers, product designers, and engineers in planning and executing an innovation roadmap of cloud data integration and analytics products.

Elif's work in data and analytics world has empowered people in virtually any job to get value from data, allowing them to ask questions and automatically generate insights in an easy, conversational manner. The technology she led and innovated has allowed thousands of users to understand and use data more effectively and enabled data-driven decisions without bias, enabling businesses to deploy ethical strategies that deliver results without sacrificing integrity. From healthcare and supply chain challenges to climate control and beyond, Elif has worked tirelessly to build solutions that allow people and their organizations to do more with data.