Open in app ↗   Sign up   Sign In

◐❙

tds   Published in Towards Data Science

This is your **last** free member-only story this month.
Sign up for Medium and get an extra one

Eric Broda   (Follow)
Oct 20, 2022 · 12 min read · ✦ · ▶ Listen

⬡ Save   🐦   f   in   🔗
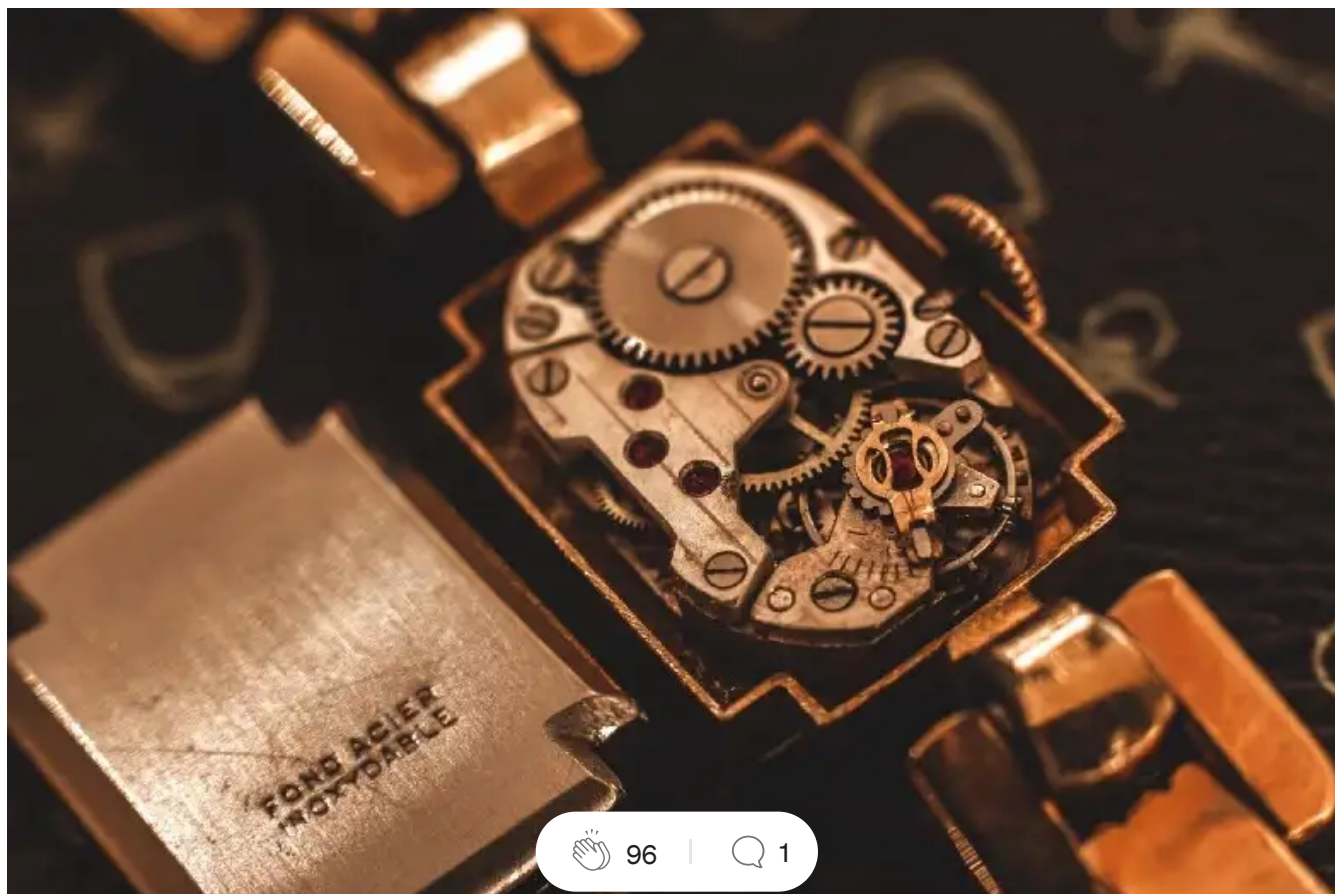
An Operating Model for Data Products

Your operating model is the primary indicator of the speed and success of your data
mesh journey. Here is what you need to know to setup an operating model for data
products in your enterprise data mesh.

👏 96 | 💬 1

Towards an Effective Data Product Operating Model

The organizational changes required to implement a data mesh are harder to address than the technology. Plain and simple, it is harder to change one person's behaviour, let alone a group, or an enterprise than it is to introduce new technology.

In this article I introduce key elements of an operating model for a data product team, and then extend it to propose an operating model for an ecosystem of data products, also known as a data mesh. Along these lines, I will address several considerations that are help you establish your data product and data mesh operating model, and, hopefully, make your organization's data mesh journey easier and a bit quicker:

- Types of teams and key interactions in a data product microcosm (small group of data products),

- Operating model for a data product,

- Operating model for a broader ecosystem of data products (ie. data mesh), and,

- Implications of operating model choices that will govern the likely evolution of data mesh within an enterprise.

I close the article with a few observations on the implications of Conway's Law on the evolution of data mesh in large enterprises:

- Data Mesh will start with local implementation, then evolve to regional implementations, probably within common time-zones (2+/-).

- Regional data mesh implementations will prevail for the foreseeable future.

- Local data product owners will define their own technology footprint supported by common regional technical platforms

- Enterprise data governance mandates will become "light-weight" with bulk of data mesh governance prioritized based upon local and regional concerns

Topology of a Data Product Team

In a previous article I summarized the distinct data product team structures advocated by visionary authors Skelton, Pais, and Dehghani. To summarize, there are several distinct teams:

- **Stream-Aligned Team** has end-to-end accountability for delivery of a software product or service.

- **Platform Teams** make it easier for stream-aligned teams to do their job by providing useful tools, utilities, and technical services.

- **Enabling Teams** act as "consultants" that help stream-aligned teams to overcome obstacles.

- **Complicated Subsystem Teams** have the deep expertise required to help stream-aligned teams consume or integration with complicated enterprise systems.
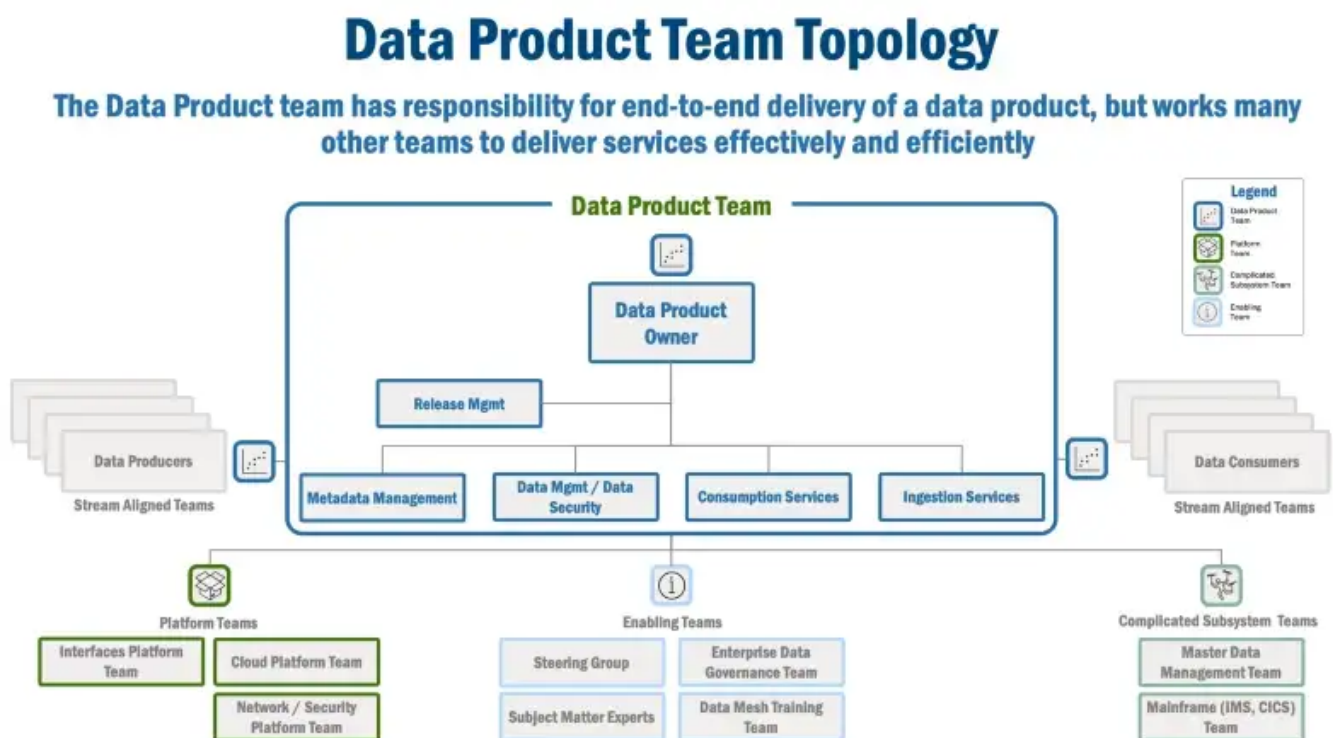

Figure 1, Data Product Team Topology

In most modern enterprises, a data product team is a Stream-Aligned Team. This team has the accountability, authority, and skills to implement a data product. The team identifies data managed by the data product by establishing clear data boundaries. The data product team is run by an empowered owner with local autonomy, funding, and decision making authority to create and operate a data product. And, the team collaborates with producers and consumers to establishing

the contract and SLAs for the data product as well as to ensure the data product delivers value.

Each data product team is typically supported by one or more "Platform Teams" whose objective is to make it easier for data product teams to consume common technology capabilities within the organization. Examples of platform teams include:

- **Cloud Platform Team**, which provides enterprise-compliant cloud Platform-as-a-Service (PaaS) that makes it easy to consume cloud capabilities in a secure, operable, and observable way mandated by the enterprise; The complexity of the cloud, especially as enterprises consider a "multi-cloud" environment (and the wide variety of skills and staff required to perform this capability effectively) makes it impractical to house in a single data product team.

- **Interfaces Platform Team**, that provide standard templates frameworks, and run-time environments that make it easy for the data product to build interoperable interfaces required to support data product ingestion, consumption, discovery, observability, and operability requirements. A large enterprise typically has teams to standards and simplify (and make X-as-a-Service) API, event steaming, data pipeline, and federated query techniques.

- **Network and Security Platform Team**, that makes it easier to interact with other systems (and people) on an enterprise network, while also providing services to make data in a data product easier to secure (both at-rest and in-motion).

Each data product is also typically supported one or more "enabling teams" that provide consulting and advice to help data product teams. There are several types or enabling teams that I have seen in enterprises:

- **Steering Groups**, that provides executive-level oversight and funding approvals on an as required basis; I find this team invaluable in influencing senior executives, addressing organizational inertia, and building momentum for data mesh.

- **Subject Matter Experts**, that has the deep experience in a specific topic that may be require by the data product team; In many cases, this is a not formal team, but is based upon a loose network of collogues that data product team members may have.

- **Enterprise Data Governance Teams**, which provides the guidance required for the data product team to remain compliant with enterprise policies and mandates.

- **Training Teams**, that help educate and build awareness of data mesh and data products within the enterprise; Initially, I see training to be provided by the first set of data product teams but eventually this team scales into an independent team as the number of data product grows and training needs grow.

And while "complicated subsystem teams" are not as common as the other teams, they also play a very important role providing expert skills within the data product ecosystem. Below are several examples of this team typically seen in large enterprises:

- **Master Data Management Teams**, that have the deep skills required to, for example, perform data matching required to ensure the accuracy, uniformity, and semantic consistency of important data assets.

- **Mainframe Teams**, such as those that manage IMS or CICS mainframe database technology.

Data Product Team Operating Model

Wikipedia defines an Operating Model as a "visual representation (model) of how an organization runs itself." There are many components in an operating model but for our purposes, an operating model describes "who does what in an organization".

Let's now elaborate on our previous data product team structure to identify the core interaction patterns prevalent in a typical enterprise (Figure 2).
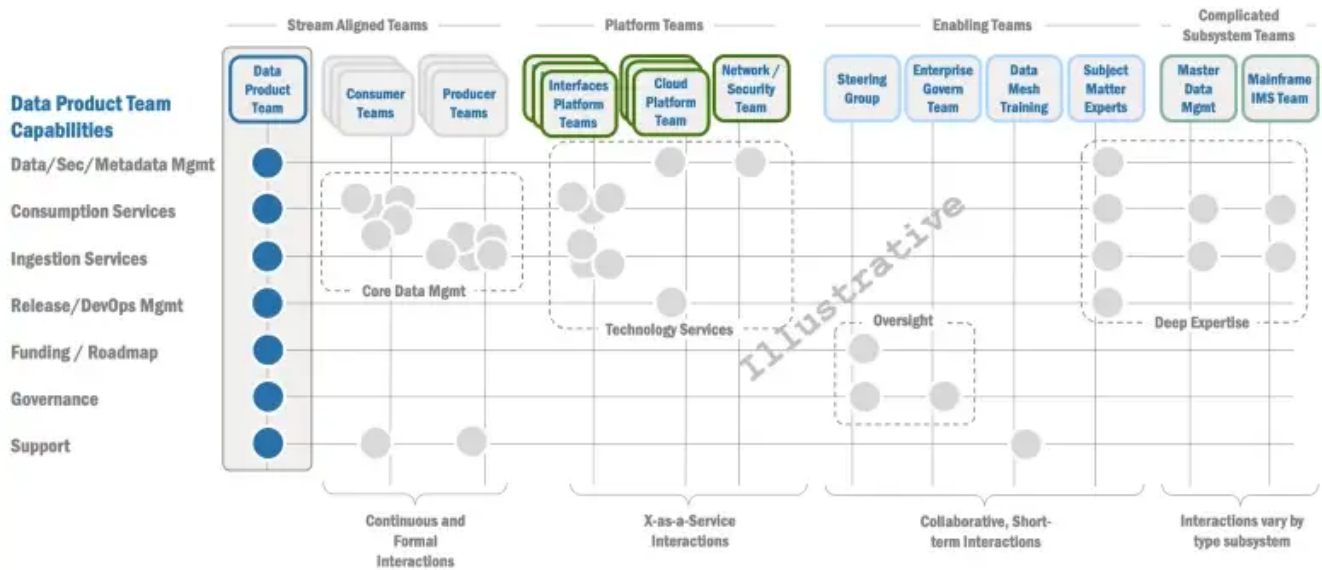
Figure 2, Data Product Team Operating Model

In our example enterprise (Figure 2) we have several types of teams, including Stream-Aligned, Platform, Enabling, and Complicated Subsystem teams and each use continuous formal long-term, "X-as-a-Service", Collaborative short-term, and subsystem dependent, interaction models, respectively.

Each dark blue dot represents a primary responsibility each of which, not surprisingly, is the responsibility of the data product team. Each light gray dot represents a material interaction between the data product team and another team.

As you can see, there are several interaction clusters:

- **Core Data Management**, showing the extensive interaction required between the data product team and other teams — source system teams, analytics teams, or other data product teams — required to safely, securely, and reliably allow the data product team to ingest and consume data.

- **Technology Services**, that form the foundational services to support the core data management capabilities.

- **Deep Expertise**, provided by both subject matter teams and our complicated subsystem teams such as Master Data Management and mainframe IMS teams, both common in large enterprises.

- **Oversight**, with our Steering Group providing guidance, strategic advice, and funding assistance, as well as an Enterprise Data Governance team that provides privacy, regulatory, and internal enterprise mandates.

Data Mesh Operating Model

But the key to establishing a high-performance operating model is to understand where important interactions take place and to use this information to optimize and streamline those interactions. So, how do we do that?

The first point to understand is somewhat counter-intuitive.

While data flows — and hence the interactions between data products in a data mesh — dictate the *technical landscape* for data mesh, it is behavioural dynamics and decision-making hierarchy dictate the actual *organizational structure* for the data mesh.

This means that a data mesh operating model is not the haphazard amalgam of data products. **Rather, the data mesh operating model — the structure of teams as well data products in an enterprise mesh — will coalesce around the organizational dynamics within an enterprise.**

This is stated best by <u>Conway's Law</u>, paraphrased as "your systems and data will follow your organization structure". Conway's Law is "based on the reasoning that in order for a product to function, the authors and designers of its component parts must communicate with each other in order to ensure compatibility between the components."

In other words, systems — and in our case data products — will be structured within defined organizational boundaries where it facilitates the simplest communications, the clearest accountabilities, greatest funding efficiency, and most rapid decision making.

Simply put, each data product team, and its enabling, platform, and complicated subsystem team partners, will naturally form — largely on its own — a somewhat self-sufficient microcosm (a small ecosystem) bound together by communication patterns and interactions around a common business goal. In other words, data product clusters will form within the boundaries of individual groups within an enterprise.

And the interactions between these various data product team microcosms form a larger ecosystem of business units or regional groups that are bound communication patterns and interactions to dictated by higher-level organizations units (geographic regions or line-of-business).

Data Mesh — Loosely Coupled Regional Ecosystems

Simply put, these interdependent ecosystems are bound by practical organizational dynamics. The most common of which is for large enterprises to be structured into country and regional groups to accommodate the distinct customer preferences or regulatory constraints in each market.
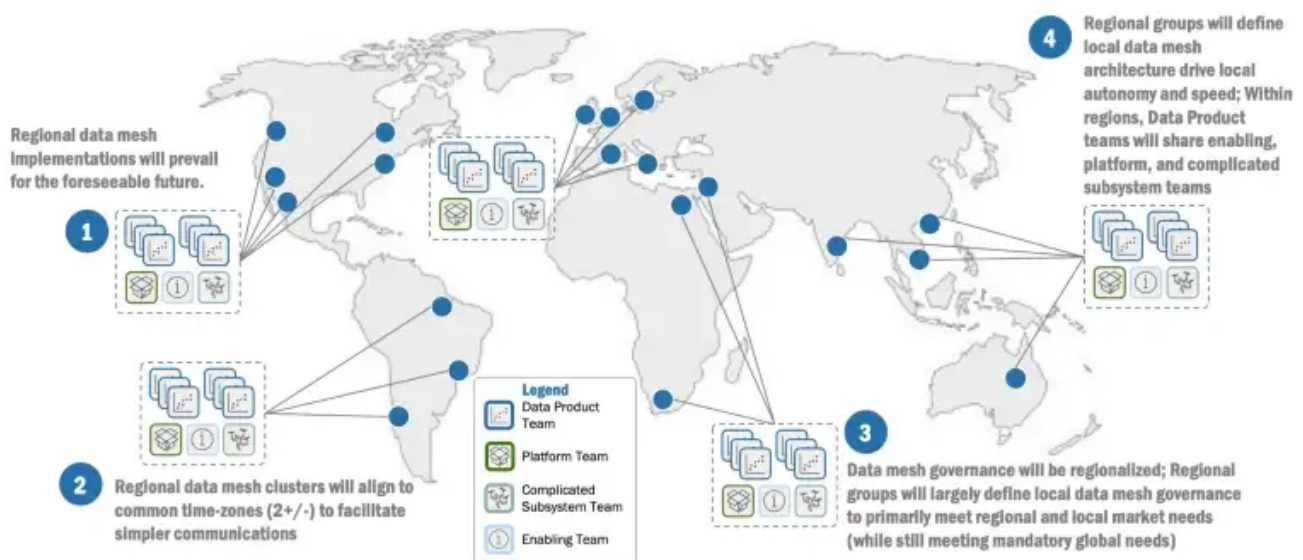

Figure 3, Data Mesh — Loosely Coupled Regional Ecosystems

For example, many large organizations have North American, European, and Asia groups that reflect each region's unique legal and market needs. And even in smaller organizations within a single country it is common to have groups or lines-of-business that have local (provincial, state, etc) business units. So, to summarize, for the foreseeable future it appears that Data Mesh will likely be composed of loosely coupled regional ecosystems. But this has profound implications.

Implication #1 — Regional data mesh implementations will prevail for the foreseeable future.

Conway's Law alone strongly suggests that a pure enterprise data mesh is probably not practical. I think most interactions (people and data flows) occur with entities in "close" (from an organizational perspective. proximity. I think practical decision making and funding is typically localized into line of business, or geographical units (after all, that is why orgs have a hierarchy) and this dynamic naturally creates clusters of data products (aka. data mesh).

Also, in my experience, I have not seen many true enterprise initiatives. In fact, other than major crisis situations (pandemic, acquisitions, etc.) there are few circumstances that have a global benefit, outcome, or imperative sufficient to warrant broader global coordination. And even then, when I have been through them, things typically start small and grow. I suspect that even in these situations, the first data mesh will likely be a regional set of interacting data products.

In fact, I suspect that anyone who has tried to facilitate decision making across regional boundaries (for example, have a data mesh spanning, say, North America and European business units) let alone group boundaries, knows that decision making is difficult and arduous, and at times near impossible.

But other circumstances conspire to support this conclusion. The regulatory environment, especially relating to security and privacy, vary from one region to another, imposing constraints that are contradictory to sharing data between regions.

As do different business priorities. A rapidly growing region may have an imperative to rapidly introduce new products, where a mature region may focus on safeguarding existing market share, each of which drive different investments and decision making which may not be conducive to driving consistency amongst data products in different regions.

And technical maturity and pace of change also matters. Some regions may be saddled with older systems that are difficult to evolve making it difficult to establish new data products, where other regions with newer technology making it easy to rapidly create data products.

Instead, data product microcosms and regional data mesh ecosystems will become the primary mechanism of data mesh growth within an enterprise.

Implication #2 — Data Mesh will start with local implementation, then evolve to regional implementations, probably within common time-zones (2+/-). True enterprise-wide data mesh implementations are probably an aspirational target for the foreseeable future.

It is a simple fact that smaller teams (5–10 people, max 15) collaborate better, move quicker, and more effectively establish the trust needed to create software. The obvious implication is that enterprises will take advantage of this dynamic and smaller data mesh implementations composed of a small number of agile and nimble data product teams. And to simplify communications it highly likely that the first data mesh implementations will be localized to similar/close time zones.

Now, it is safe to say that remote work culture and tools may provide opportunities to establish robust communications across time-zones. But even this has practical limitations. And unless people are truly prepared to encourage working at all hours of the day, most data products will be built into regional data mesh clusters within 2 +/- time zones that foster communications across common business hours.

This implies that enterprise-wide homogeneous data mesh implementations will probably not be the norm. Rather it is much more likely that in the foreseeable future enterprises will likely adopt independent regional data mesh implementations composed of fewer related data products.

Implication #3 — Enterprise data governance mandates will become "light-weight" with bulk of data mesh governance prioritized based upon local and regional concerns

As stated earlier, regulatory environments vary from one region to another. Consider privacy: The European Union has GDPR (General Data Protection Regulation) where privacy regulations are more relaxed in North America. Similarly, health care data regulations vary markedly from region to region, country-to-country, and in many cases even from state-to-state or province-to-province.

This does, however, lead to a quandary. Traditional data governance does not necessarily take this regional disparity into account. It is perhaps best stated on the Data Mesh Radio podcast by Mohammad Syed (lead strategist at Caruthers and Jackson) and Scott Hirleman (Podcast interviewer), that far too often current

enterprise data governance practices "address macro concerns at the expense of regional or local market micro needs".

Syed and Hirleman go on to emphasize a very important point: Data governance, and a core element related to data quality, is "contextualized" to the needs of a local market. Hence it is fair to expect that data product governance will also recognize the context of the local market it operates in.

Simply put, there is no one-size-fits-all. Which also implies that data product governance (if not broader data governance) must embrace a much higher degree of localization. So, how can data governance become localized?

First, data product owners, largely operating in local markets, must be empowered to respond to their local governance needs. They must be agile and nimble, and given the latitude to select not only their technology platform but also their governance approach.

Second, enterprise data governance will evolve to have a much lighter touch. Other than the few mandatory enterprise governance imperatives, governance will likely be led, guided, and verified, by regional data product owners.

Implication #4 — Local data product owners will define their own technology footprint supported by common regional technical platforms

The modern "real-time" enterprise is fueled by data which now plays a central role in driving enterprise growth. In the past, enterprise IT groups have largely optimized for cost. Today, agility and speed drive priorities.

In other words, the agility created by local autonomy trumps centralized command-and-control structures. Monolithic architectures become distributed. One-size-fits-all governance becomes federated and localized.

And, to address these changes data product owners will define their own technical landscape. This maximizes local autonomy, decision making, and responsiveness to local market.

So, it is highly likely that technology footprints will be dictated by local or regional data product owners. And, yes, obvious enterprise platforms should be leverage where appropriate, but they should not be mandated unless absolutely necessary.

Concluding Thoughts

Hopefully, after reading this article, you are now more fully aware of the organizational considerations required to implement data products in a data mesh ecosystem. And now hopefully you have gained the insight to begin to build the data product teams and data mesh operating model required to accelerate your data mesh journey.

***

This article assumes that you have a high-level understanding of data mesh. If you need some background information on data products, there are a number of articles available here (data products), here (data mesh patterns), here (data mesh architecture), here (data mesh principles) and here (lessons learned). For interested readers, a full set of data mesh patterns are available here and here.

***

All images in this document except where otherwise noted have been created by Eric Broda (the author of this article). All icons used in the images are stock PowerPoint icons and/or are free from copyrights.

The opinions expressed in this article are mine alone and do not necessarily reflect the views of my clients.

Data Mesh   Data Science   Data Governance   Data Architecture

Organizational Culture

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. Take a look.

By signing up, you will create a Medium account if you don't already have one. Review our Privacy Policy for more information about our privacy practices.

✉⁺ Get this newsletter

Get the Medium app