

Deloitte.



Untangle your mess and knit your mesh

A cross-company point of view

Data Mesh, Q1 2022, Deloitte, Collibra and Snowflake



Table of Content

1.	Data teams in all sectors are victims of their own success	2
	Your organisation is not the only one to experience this challenge	2
	Centralised data teams have to face an increasing demand from the business	2
	Prioritising a wide-ranging list of requirements is difficult	2
	Lack of full end-to-end data ownership	2
	Domain knowledge gets lost somewhere on the way through the central hub.....	3
2.	Governed decentralisation is the next step for increasing operationalisation	4
	Decentralised data structure	4
	The data mesh concept solves the challenges of centralisation and also brings the agility and faster time to market by democratising the data management at the business domains level.	4
	Data responsibility is decentralised	5
	Empower individual business domains to create data products that can be consumed by other teams.....	6
	Data as a product.....	7
3.	Your data mesh organisation requires strong pillars to make an impact that matters	8
	Key principles to consolidate data products	8
	Key considerations for a successful transformation	9
	The transition from the current state to data mesh will be gradual.	9
4.	Your organisation needs to consider a self-service infrastructure as a platform	10
	A decentralisation accelerator	10
	Data products as architectural quantum.....	10
	High level implementation	11
	Implementation considerations.....	12
	A federated governance model: consistency, compliance, and quality	12
	Federated data governance	13
	With a federated approach data governance facilitates data literacy throughout the organisation.....	13
	An intelligent approach to data: Embedding data governance	13
	Taking advantage of data governance.....	14
5.	How can Deloitte support your organisation in its data mesh journey?	15
	Drawing up the plan	15

Are you satisfied with the way your organisation creates impact through the intelligent use of data?

1. Data teams in all sectors are victims of their own success

In the age of digital, every organisation strives to be data-driven. To realise this great ambition, organisations must collect and manage data from everywhere (internal and external sources). Additionally, the amount of data generated every day is exploding. In this context, the number of data-related initiatives requested by the business is continuously increasing. As a result, centralised data teams struggle to deliver the needed business agility and speed of change required within the organisation.

Your organisation is not the only one to experience this challenge

In today's fast-changing business environment, organisations are facing pressure to increase their agility and connectivity. Indeed, resilient organisations have the ability to quickly react by offering innovative (data driven) services to adapt to its customers needs, market changes, macro forces... Organisations with a too centralised data landscape lack the level of agility needed to manage data at scale and generate the expected value from their data assets.

Centralised data teams have to face an increasing demand from the business

We cannot ignore the advantages of a centralised data team: standardisation, common methodology, centralized data skillset and expertise, synergies across initiatives, This team tries to build a common data asset that has to cope with very diverse cross-departmental specificities, priorities and visions.

But centralisation has its limit: Volume, velocity and diversity of data (and data sources) are exploding, Data demands from the business follow the same trends. Nowadays, centralised teams struggle to keep up the pace and to properly deliver data as well as integrating new data products.

Prioritising a wide-ranging list of requirements is difficult

When all data demands from the organisation are addressed to a central team, it complexifies the prioritisation exercise. All demands converge on the same pool of people. The interdependencies of parallel projects complexify the change management and the release management. Prioritising these requests is complex and can lead to bottlenecks. Such complexities can generate frustrations and stress.

Lack of full end-to-end data ownership

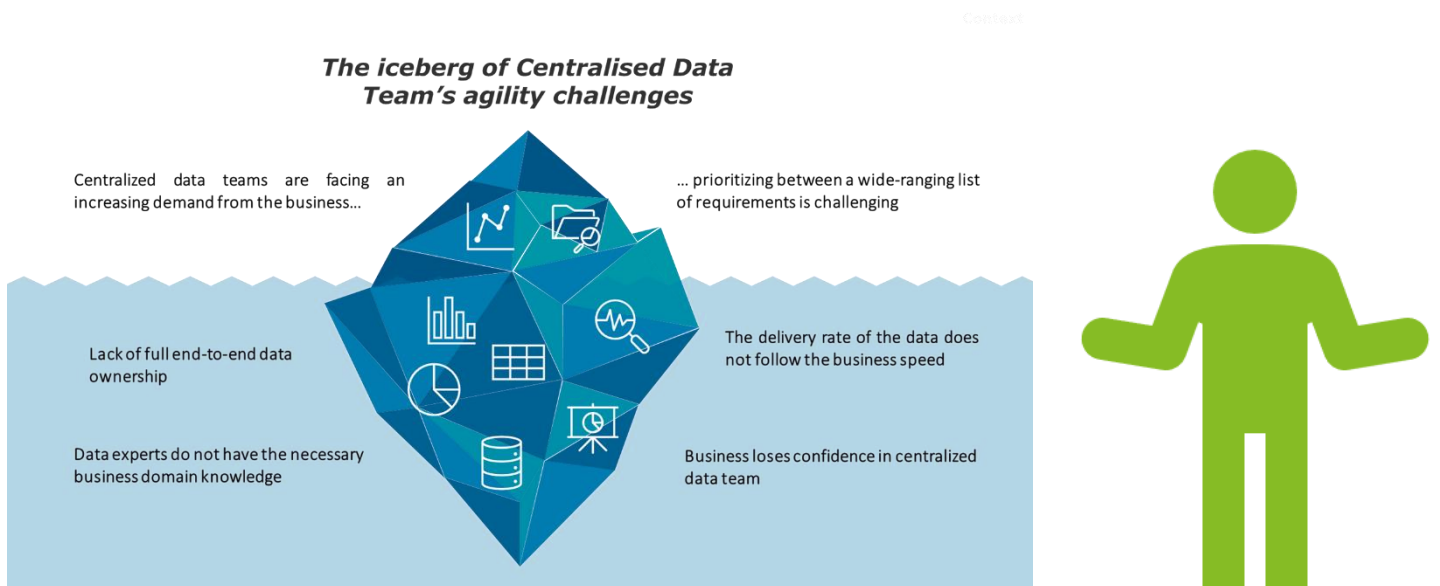
Since the data ownership lies in the business and hands-on capabilities reside in a centralised data team, there is a risk of discontinuity in the accountability of data and generated insights, putting at risk project's value generation. Indeed, people tend to put the responsibility on others, especially when things go wrong, even more if frontiers of accountability are vague.

Data ownership is an underrated responsibility in organisations. It is sometimes, wrongly, assumed by a member of the data team. And of course, the data team members have neither the time nor the ability to understand all the background and context of every data product of the organisation.

Domain knowledge gets lost somewhere on the way through the central hub

Business knowledge reasonably resides within the business, while central data team members have their own specific data expertise. But, being able to translate a business problem/opportunity into an insight solution, or understand the right business question to ask of the data, through to interrogation and modelling, and presenting the insight in a compelling way are important skills for an organisation. People combining both expertise are crucial to the success of a data driven organisation but pretty rare on the market.

A centralised data team tends to ensure connectivity and synergies between data assets, but it can lower the agility to address all diverse data needs within the organisation due to highly complex interdependencies.



2. Governed decentralisation is the next step for increasing operationalisation

In May 2019, Zhamak Dehghani published an article on [martinfowler.com](https://martinfowler.com/articles/data-monolith-to-mesh.html) called “*How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh*” (Dehghani, 2019)¹. This article lays the foundations of the data mesh concept.

A data mesh organisation is a decentralised data structure whereby a part of the data responsibility is decentralised to the business domains. The business becomes accountable to manage data as a product.

Decentralised data structure

It is a true paradigm shift from what we have seen across industries over the past few years.

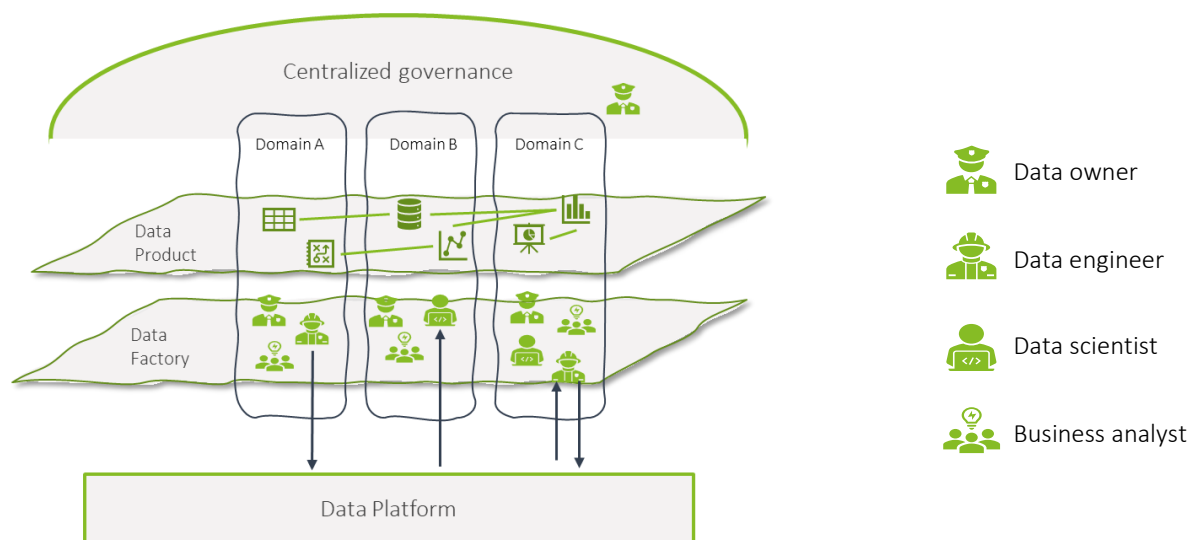
Until around five years ago, in many organisations the process of obtaining data or consumable insights was typically performed by IT. However, as demand increased it often led to siloed business initiatives inside different business domains (what is called shadow IT) resulting in a mix of intertwined, inconsistent, inefficient, and ungoverned data solutions.

Then, organisations decided to centralise all data and data initiatives into data warehouses by creating centralised data teams to manage them... These efforts of centralisation were aiming for the same goal: a large central delivery team focused solely on addressing transversal demand for business intelligence and analytics, ensuring consistency and reliability of the data across the organisation. All the various data domains were centrally managed by the data teams. The central team had to integrate all data and capture several business logics. A drawback of these big monolithic data structures is that they were controlled by IT or data teams. They started to become bottlenecks to providing data to the business domains as the demand for data exploded. Moreover, the data teams could not understand all the business aspects of the specific business domains. So, the local business teams started to set up their own processes to reach the speed of delivery required to cope with competition, resulting in disorganisation.

The data mesh concept solves the challenges of centralisation and also brings the agility and faster time to market by democratising the data management at the business domains level.

This domain-driven design is the primary means to structure the data usage and exchange as well as the full end-to-end ownership. Each domain has its own data factory, which is made up of several data profiles (e.g., data engineers, data analysts, reporting analysts, data scientists,...) who manage the data of a specific domain. The current data team is actually distributed over the business domains. They support them to get insights from the different domains of the organisation.

¹ *Mesh*. Retrieved from [martinFlower.com: https://martinfowler.com/articles/data-monolith-to-mesh.html#DomainDataAsAProduct](https://martinfowler.com/articles/data-monolith-to-mesh.html#DomainDataAsAProduct)



Data responsibility is decentralised

Currently, the business does not feel the final accountability of their data quality since they count on the centralized data team. By making them owner of their own data product, a sense of responsibility is enforced. They can rely on their own data factory to ensure the data management.

Inside each domain team, a data factory is supervised by one or multiple data owners who are responsible for their own business areas, including the domain's master data definitions, data quality, and policy compliance.

The data owners must follow the guidelines issued by the centralised governance, ensuring standardisation in the whole organisation.

Customer data, a coveted resource

Every department in the organisation wants to use customer information: Marketing wants to better understand customer's preferences, Sales needs to know how many products have been sold to a customer, Market Research desires to analyse customer behaviours, ...

Customer data is a shared asset for your organisation. The customer domain is indeed very big and needs to be contextualised based on business needs, as departments will approach it from different angles. For instance, Marketing will look at customer data differently than Supply Chain or Telephony. Customer behaviour can be used to predict demand and optimise supply chain production, stocks. Customer segmentation allows to identify the need to launch new products. In all cases mentioned, the common point is the customer ID that ensures departments talk about the same customer but from their own point of view.

How will your organization manage customer data ?

- Golden record will remain under the ownership of centralized data team ?
- Specific department will own customer data product ?

Some cross-domain master data assets, like customer data, could remain under the supervision of a centralised data team to ensure their management and quality even if parts of the master data is owned by business teams. The centralized team would then manage the golden record that will be used to uniquely identify a customer, ensuring that this golden information can be leveraged by all business teams. The business teams will then use the golden record as a unique identifier to plug new data and insights created by the department.

Empower individual business domains to create data products that can be consumed by other teams.

The data mesh structure allows to centralise expertise and rules related to data architecture, data management, data quality, and governance. These capabilities are domain-transversal whereas data specialties like data science, data engineering analytics delivery, and business intelligence are more productive when they acquire the business-specific knowledge.

In comparison with the centralised operating team, this knowledge resides only within the business and the business collaborator needs to explain it to someone who does not fully know the business and its challenges. This is also why it is very difficult for a centralised team to deliver on business needs. It is impossible to be an expert in every aspect of the business, and data engineers and data scientists are changing scope too often to have time to deep dive into them. In this centralised structure we rely a lot on the business to define the rules and express needs to the engineers who don't have the knowledge and don't have the focus on one business aspect.

In the data mesh organisation, technical profiles lie in the business, enabling better insights and faster delivery.

Data as a product

Each domain is both producer and consumer of data products (tables, reports, data bases,...) in the data mesh organisation. To ensure that all business domains can trust each-other's data products, they need to adopt the *"data as a product"* principle. Domain teams must apply product thinking to data, considering their data assets as a product and the rest of the organisation (data team and other business teams) as an end user.

This notion should not be underestimated. It is a cultural switch that must be supported with trainings and a specific data culture programme.

The main challenges to unlocking the real value of data and analytics are related not to tools and technology, but to culture and people.

" Building a data culture is not just an option. It's business-critical."

CDO from top BENELUX Insurance Company

What does Data Mesh mean for the actual centralized Data team?

Data mesh does not mean the end of the centralized data team but a shift of activities to ensure a consistent methodology and common data practices for the whole organisation.

Governance is also a capability that resides in the centralized data team. Processes, policies, and standards are set up at a centralised level.

A centralised data team can be the data centre of excellence for the whole organisation. By implementing a data culture programme with trainings and certifications, this team is the "go-to" department when collaborators have to cope with data questions or challenges.

Also, in order to have a robust, harmonised and future-proof architecture, solution architects need to sit within the centralized data team.

Depending on the maturity of each department, speed of adoption can slightly differ ; the centralized data team could continue to support data product development for those business team.

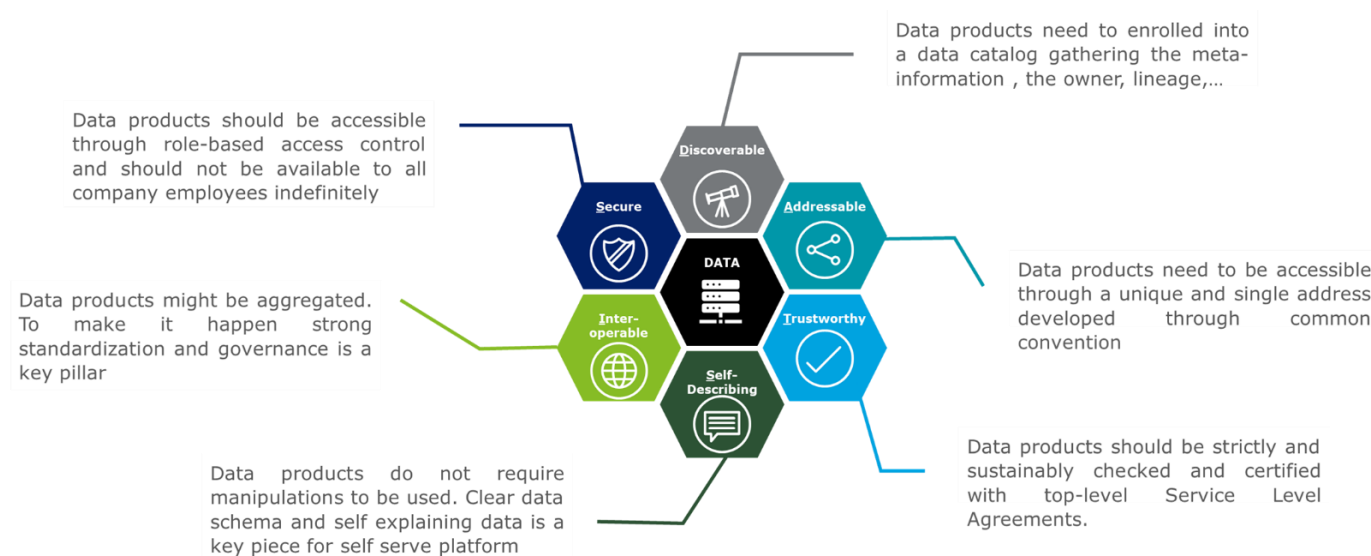
3. Your data mesh organisation requires strong pillars to make an impact that matters

Implementing a data mesh structure requires effort and adopting certain principles.

Key principles to consolidate data products

As mentioned earlier, your data product needs to be as perfect as your business product, which requires following some guidelines. The DATSIS (Discoverable, Addressable, Trustworthy, Self-Describing, Interoperable, Secure) principle should become your credo.

Figure 2 DATSIS acronym explained



The business domains are responsible for developing and maintaining the data product following the DATSIS guidelines. Nevertheless, a headquarters centralized data team is responsible for initiating and running the global governance. Standards are defined and controlled to make sure data products are secured and comply with regulations.

A second fundamental ingredient is the self-serve data platform to allow domains to develop their data products and reuse others.

Finally, a support pillar encompassing:

- **Data community:** Facilitating knowledge sharing between resources
- **Expert data support:** Providing data expertise to the business, attracting the right resources
- **Innovation and R&D:** Developing innovative use-cases

- **Data scouting:** Looking for data sets to add business value

... will facilitate the data mesh adoption.

Key considerations for a successful transformation

Reorganising an organisation is always a challenge. A data mesh transition is no exception.

As mentioned earlier, the data culture is key. In order to fully trust data, business collaborators must understand the data and their implications at all levels of the organisation. Data programmes can be put in place to share appropriate guidance around data usage as well as frameworks and standards to leverage good practices. Specific trainings can be provided to highlight processes and methods on data exchange, data quality, data governance, ...

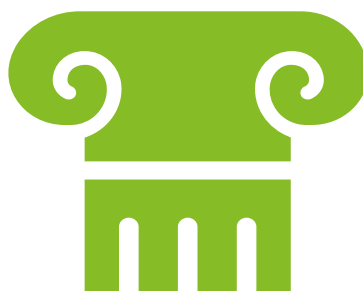
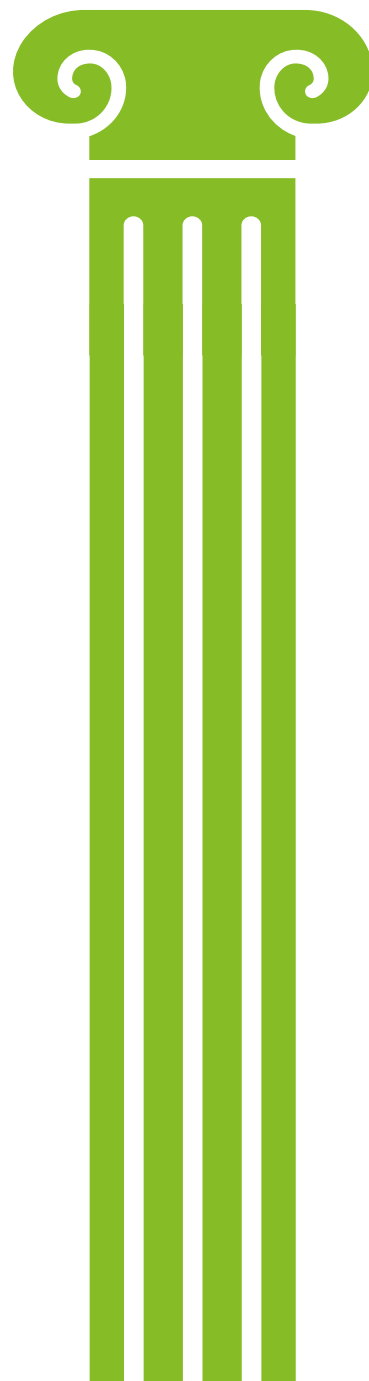
In addition, IT and business teams should have full confidence in each other to avoid constant data remediation. Certifications issued by IT or another central instance should be delivered to business users. The level of certification can depend on the roles, responsibilities and the level of data manipulation required to reassure IT on the capabilities of business and data factory members.

The transition from the current state to data mesh will be gradual.

Business domains with higher maturity (good data literacy, data profiles available, etc.), and which do not heavily impact the IT department, should be prioritised.

As the demand for data profiles is growing, it is harder to find skilled individuals. An agreement between business domains is mandatory to guide the early stage of the transition. These agreements can specify the resource allocation and can also describe the interdependencies that can exist between the business domains (e.g., application development that requires data from other domains with specific refreshing,...). This part is mandatory to increase transparency and avoid friction between business domains.

If the hiring issue persists, partnerships with external providers can be considered.



4. Your organisation needs to consider a self-service infrastructure as a platform

“A self-serve data platform must create tooling that supports a domain data product developer’s workflow of creating, maintaining and running data products with less specialised knowledge ...” ¹

One of the key principles of data mesh is the self-serve infrastructure as a platform. The capabilities described by Zhamak and needed for a self-service infrastructure as a platform are based on ease of use and a common toolset for the domain teams. In a data mesh set-up, niche technologies will not be sustainable due to the specific skill set as well as the additional work needed to follow the DATSIS principles (*Figure 2*). Snowflake’s Data Cloud provides an extensible foundation for domain teams to support a governed and decentralised data mesh architecture. It enables domains to go beyond creating and sharing data as products, but also processing logic as products.

A decentralisation accelerator

It is critical that domain teams can access the resources and tools they need on demand to support them at every stage of the data product lifecycle—from accessing the right data, to processing and preparing it, to analysing, creating models and serving data products to other domains.

The platform needs to provide an elastic performance engine, so domain teams can power large-scale pipelines, ad hoc exploration, BI reporting, feature engineering, interactive applications, and much more, all with a single engine. This allows organisations to simplify their architectures without sacrificing speed or usability. Regardless of whether the teams work in SQL, code (such as Java, Scala, or Python), or a mix, the performance engine supports them all the same. Furthermore, with elastic scalability and isolated multi-cluster compute, each domain team gets access to the dedicated resources they need without impacting performance or concurrency for other teams. This can accelerate the transition toward a decentralised but governed architecture.

With the explosion of variety, volume and velocity of data, domain teams will need a platform capable of ingesting vast amounts of data in very different formats. Data may come from different sources and can be consumed from other domains as products. The platform should be open in order to consume data from other domains at the same time as providing data to other domains, avoiding any vendor-locking. Being open means being able to interact with the ecosystem and ingest and expose standard formats and should not be confused with open-source technologies.

Data products as architectural quantum

Domains can easily design new products with the given platform capabilities to create a product pipeline for ingesting, serving and processing data products. The following diagram represents how new data products can be created from existing data domains. In the example, a new ‘design to cost’ data product can be created using the data provided by Parts, Finance, Supplier, and Bill of Materials domains. Using Snowflake Data Exchange or a third-

party data catalog such as Collibra, those data domains can advertise and expose their data products that will be immediately available to be consumed and processed to provide new data products.

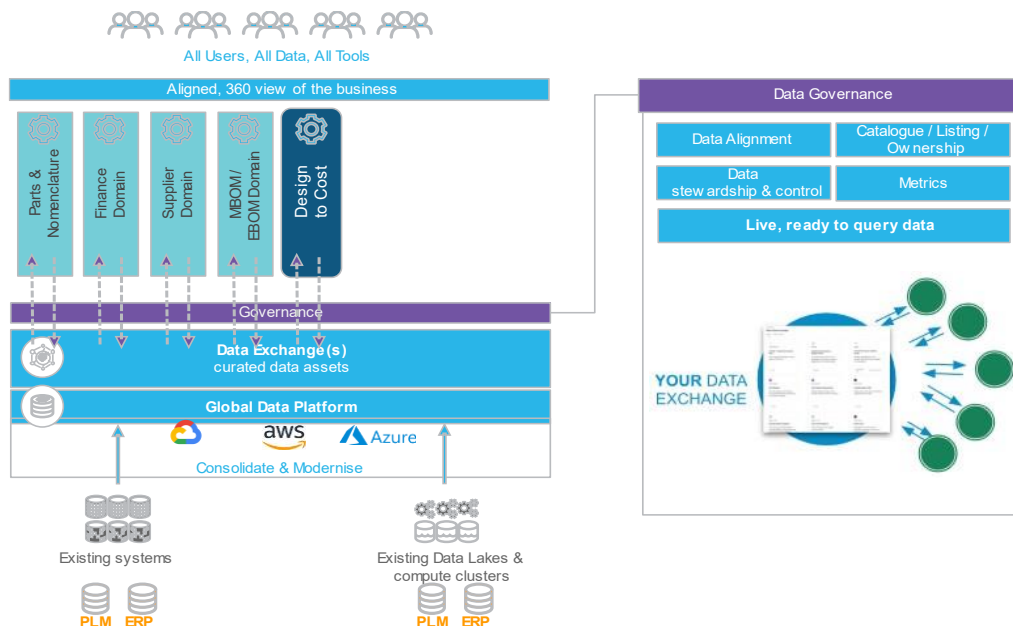


Figure 3 Standard centralise data architecture for a self-serve data platform

Connecting organisations and data teams to the most relevant data when they need it, without silos or complexity, is what the Snowflake Data Cloud is designed to do.

To deliver on that, the Snowflake Data Cloud is backed by Snowflake’s platform, which is uniquely built for performance at scale, ease of use, and governed data sharing and collaboration; and it’s well suited to support both the centralised standards and the decentralised ownership necessary for a successful data mesh deployment.

Snowflake’s platform gives these teams a one-stop shop, while supporting a wide range of skills.

High level implementation

Implementation of data mesh in Snowflake can be based on different topologies:

- Domains can be based on accounts and leverage Snowflake’s Secure Data Sharing capabilities to break down silos, cross region and cross cloud and work on a single copy of the data.
- Domains could also be based on database or schema and leverage a third-party catalog such as Collibra to make products discoverable and accessible.

In any case, Snowflake can allocate independent resources for the domains to load, process and prepare their data products leveraging the Snowflake virtual warehouses and then list the data products on a third-party catalog. Consumer oriented domains can then use these products through data sharing or intra account cross database access depending on the selected topology.

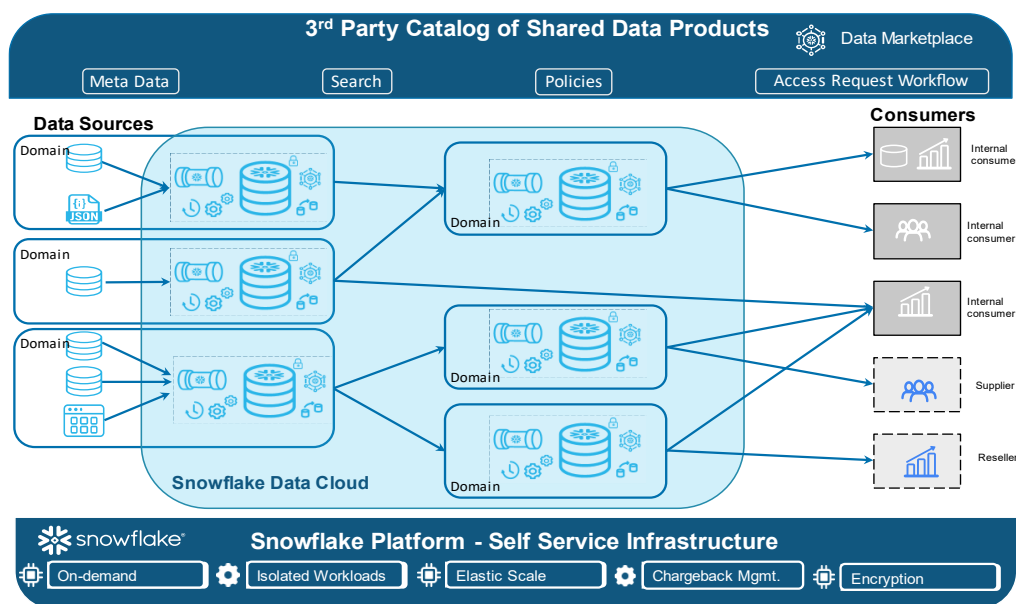


Figure 4 Example of data flow

Implementation considerations

Not choosing the right tools and infrastructure can limit the benefits of a data mesh, adding extra complexity and increasing the time to value and cost. Zhamak proposes a data infrastructure provisioning plane that only advanced data product developers use directly as it is low level infrastructure lifecycle management. SaaS platforms like Snowflake remove that complexity and dependency of low-level expertise. Snowflake resources deployment and management can be completely automated with infrastructure as code in a self-serve manner with highest level of security and governance and across any public cloud.

The next level is being able to abstract the complexities of data workflows. By using all Snowflake capabilities, data workflows can be automated using 'data as a product' approaches and integrating with different tools.

A data mesh architecture should also be complemented with other tools such as ingestion and automation, machine learning, and other adjacent technologies.

A federated governance model: consistency, compliance, and quality

There are many advantages of shifting toward a business-domain focused architecture. Nevertheless, a strong and federated data governance model is required to connect it all.

Federated data governance

As organisations transition to a data mesh design, evolving their perspective from data as an asset toward data as a product, they must also look at data governance through a slightly different lens.

Data governance focuses on data definition, quality, and the rules and processes to extract insights which enable the organisation to leverage these insights into business outcomes. Data governance should align with managing and organising data and processes to enable collaboration and compliant access across the enterprise. Ideally it provides the mechanisms for teams to discover, understand, access, and use the data to deliver insights while also adhering to regulatory, security and privacy requirements. The federated model applies necessary central enterprise oversight and support to the business.

Like domain ownership of data and data products, governance is also in the responsible hands of the business. An effective partnership between the business and governance functions, along with IT, Privacy, and Security ensures efficient, compliant, and safe access to data.

Data governance is about structuring the data environment for optimised use and access for all users.

With a federated approach data governance facilitates data literacy throughout the organisation.

The federated model is intentional in driving collaboration across domains to ensure common understanding of terminology and policy, fostering both enterprise-wide authority as well as domain-specific contextual definition that suits all needs.

Federated governance as part of a data mesh architecture must strike the right balance of centralising control and reducing risk, together with enabling the business to leverage the data to achieve business initiatives. The ideal data governance model will clearly define roles and responsibilities around data assets and provide a system of accountability. Concurrently it will support distributed collaboration with easy access to a shared space where all data stakeholders can use data and communicate with one another. By incorporating formal decision-making rules and manual processes, organisations can track data usage and improve data quality and accuracy. Self-service analytics and automation provide agility to allow the line of business to understand the data context and use it productively. At its core, data governance mitigates risk by ensuring compliance with industry and regional regulations and policies to avoid fines and maintain customer trust.

An intelligent approach to data: Embedding data governance

Organisations need the right tools and capabilities to embed data governance within their data mesh architecture in order to enable all users to create value out of compliant data. These capabilities will allow organisations to focus on embracing data collaboration while providing compliance, security, and privacy.

Collibra Data Intelligence Cloud provides the necessary capabilities for companies to embed data governance within their organisations.

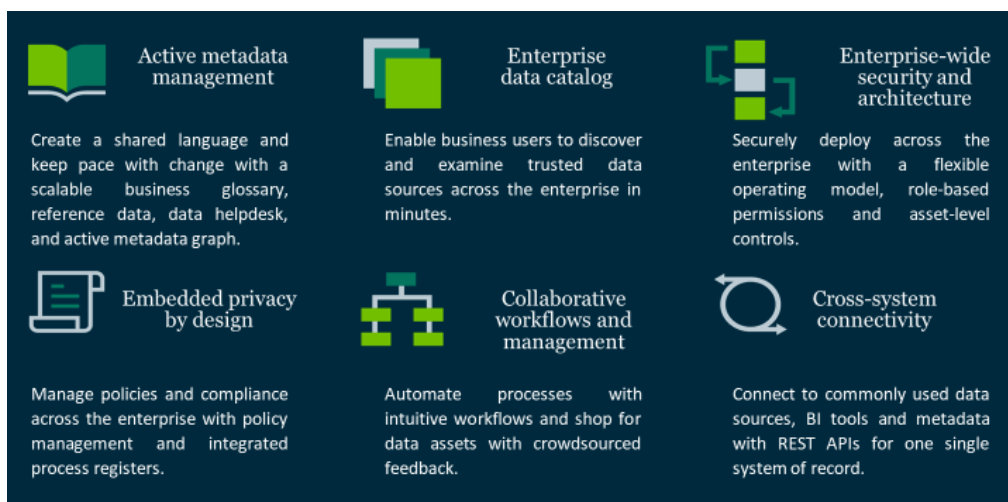


Figure 5 Data Governance Capabilities

Taking advantage of data governance

Built on a foundation of data governance, organisations can now embark on their data mesh journey with a single source of truth for data across distributed environments.

With full visibility into data across the organisation, business domains can make their data discoverable. Data are catalogued with relevant business context, including business definitions, ownership, policies, and usage. Data consumers can easily find the data they're looking for across all data sources, business applications, data science and BI tools. Users can search and understand data using common business terms and get visibility to see how data transforms and flows from system to system.

A decentralised data management model creates enormous opportunities to improve data quality and streamline data access, but seizing this opportunity goes beyond the data itself. Organisations must gain a complete understanding about the relationships and how the data are connected. Actively managing metadata creates context-rich connections across all distributed business and data domains. The ability to repurpose data, collaborate and drive better decisions with data intelligence is the end-result of connecting the right data, insights and algorithms to all users.

A flexible operating model allows configuration of any type of entity (business term, data type, business rule, quality rule, process), any reference data, and any type of relationship. This means organisations can model their desired data strategy to fit their needs and put the data in the hands of the people who need it, using the right hierarchical organisational structure. A role-based access and permissioning model provides the granularity necessary to support security and compliance requirements and deploy enterprise wide. Embedded governance capabilities, within the data catalog, ensure that access to trusted data can be granted enterprise-wide without risk.

The right data governance approach can help deal with the volume of heterogeneous systems and requirements by integrating all data sources, BI and analytic tools into a single platform, and eliminates the need for multiple point solutions. A centralised platform provides the tools necessary to reduce the effort related to manual rule writing and can help save time and resources in identifying and migrating data to the cloud. Organisations can also avoid compliance fines with centralised security controls compliance policies.

5. How can Deloitte support your organisation in its data mesh journey?

Several organisational topics must be discussed, a self-serve platform should be shaped, and a strong cross-domain governance is expected. All these themes need to be addressed in a holistic approach.

Drawing up the plan

A full data mesh set-up is not a simple structure that can be realised in a single online meeting. Before starting to implement a platform, to select a governance tool, and to rearrange teams, a first analysis is needed where the stakeholders will **explore** the range of data meshification they need, **select** the best fit for the organisation, and finally, establish the **plan** to make it happen.

As the transition decision-maker, you will not implement data mesh transformation without hoping it will last and enable great value. Before diving into planning, you might need to gauge whether or not data mesh is optimal for your organisation. If yes, determine which level of data meshification is required to drive the most value.

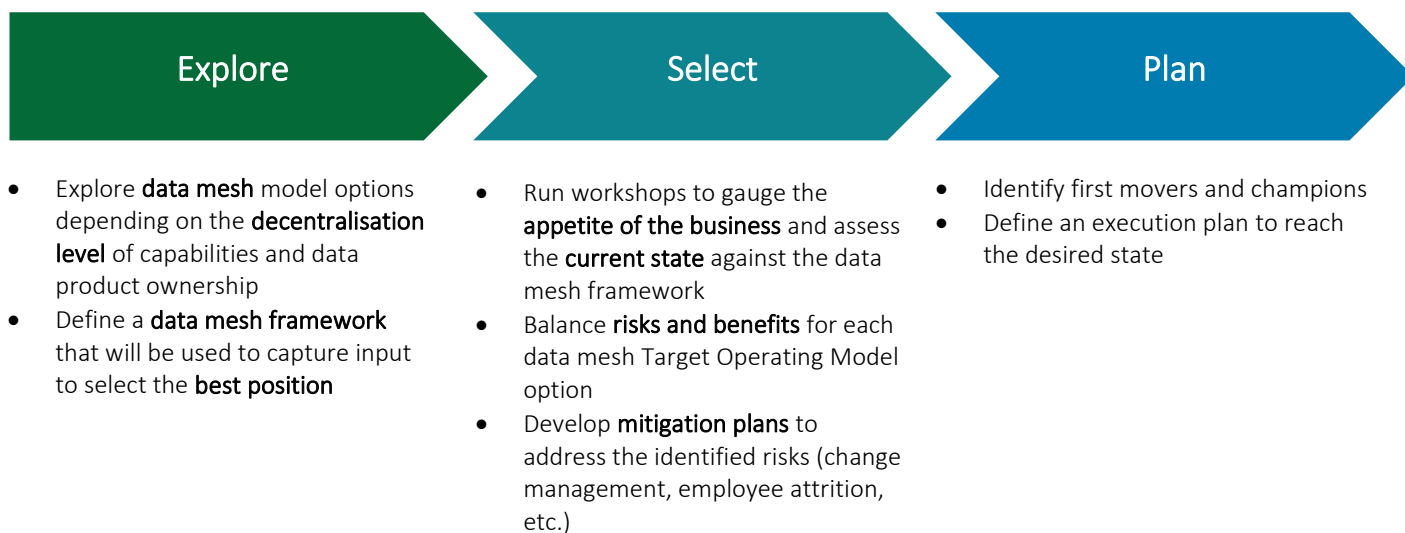


Figure 6 : First steps for a data mesh transformation

Once the plan is ready, it needs to be activated taking into consideration all business implications such as:

- How to adapt the operating model to fit with the new structure
- How to acquire talents and where (upskilling recruitment,...)
- How to redesign the governance processes
- Which tool will best fit your needs
- How to adapt your data architecture
- Do you need to adopt a DevOps model

As a data executive,

Your role is essential to steer the transformation by choosing the best data mesh model in the range of possibilities. Some capabilities are subject to discussion and may depend on the level of data maturity of your organisation, data maturity of your business departments, available resources,...

And once the transformation is complete, your role is still fundamental. In actual facts, all departments might not embrace the meshification. A centralized team might still be on duty to support.

The centralized data team office can support the business with talent acquisition. According to their needs in terms of data, the centralized data office can advise business teams to select the best candidates.

As data meshification is a long journey, the technical debt still needs to be managed for the time of transition as well as the innovation. As Chief Data Officer will own all data topics from innovations and implementation to drive the future of your organization.

From lessons learned, we advise you to **think big but start small** with a focus on specific use cases that generate rapid and tangible impact. Starting small includes a study on current foundations and how robust they are to support the transformation. A gap-fit analysis is recommended to ensure scalability growth. Organisations need to train and invest in talent to develop their skills and to ensure that they can operate the new platform to perform their tasks.

Ultimately, once the data mesh concept and implications are understood, once data (governance and platform) and people (talented profiles) aspects are robust enough, the organisation can activate the plan and start the transition of the current data team to the identified and prioritised data domains.

At Deloitte, we usually suggest several sessions to facilitate discussions between the different internal stakeholders and accelerate decision making. During the meetings, our team explains the data mesh principles to give your team the opportunity to discuss the benefits, potential needs, and impacts on operations.

Deloitte has collaborated closely with Snowflake and Collibra in order to elaborate a full business view of the data mesh.

For more information, please contact our data mesh expert team. If more details are required on Collibra or Snowflake, our experts will be happy to redirect you to our preferred partners.



Christophe Hallard

Partner in Strategy, Analytics and M&A

challard@deloitte.com



Magali Thesias

Director in Strategy, Analytics and M&A

mthesias@DELOITTE.com



Gilles Dekeyser

Senior Consultant in Strategy, Analytics and M&A

gdekeyser@deloitte.com



Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited (“DTTL”), its global network of member firms, and their related entities (collectively, the “Deloitte organization”). DTTL (also referred to as “Deloitte Global”) and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see www.deloitte.com/about to learn more.

© 2022. For information, contact Deloitte Touche Tohmatsu Limited.