



HOW DO WE GOVERN
DATA PRODUCTS?
**MOVING TOWARDS
FEDERATED DATA
GOVERNANCE**

01

Introduction

The **economic and social** potential of using data is enormous, from enabling the creation of new products and services based on innovative technologies, to increasing production efficiency, to providing us with tools to fight society's problems. To realise this potential, data must be available, reliably shared and technically easy to reuse.

Achieving an exchange between data producers and consumers in a **properly managed environment** is undoubtedly one of the most significant outcomes of a data governance initiative, and at the same time one of the most difficult to achieve.

Its absolutely vital to democratise data (both internally and externally) to be able to use them holistically to make business decisions, but there are many aspects that need to be addressed when implementing a data governance framework, and data ownership and management are two that are especially important.

In the paper, "**Governance 2.0: from governing data to governing data products**", we point out that one of the characteristics of the evolution of traditional data governance models is to **govern the data lifecycle**, and therefore expand from a "focus on data" to a "focus on data as a product and on the data products".





Our vision of [comprehensive data governance](#) involves an approach that combines global principles and policies for the most important aspects with decentralised federated management by data domains.

This approach to a distributed data strategy may run counter to what is considered normal in other aspects of life. We've always been taught that we have to learn to walk before we can run, and to be able to manage, we must first operate. But when it comes to data, we believe that the first step is for the business areas to assume responsibility, and then learn how to [generate value from the data](#).

Treating data as products and transferring ownership to the producing teams follows the theory that ownership is better aligned with data creation, making the data more [visible and better quality](#). This shift in responsibilities addresses the issues arising from the explosion of data that has overwhelmed the capabilities of the centralised data teams that are responsible for cleaning, harmonising and integrating the data.

Delegation of responsibility for specific data sets means that organisations are divided into data domains, with each domain assuming responsibility for serving data, maintaining the associated metadata, improving data quality, implementing lifecycle management, controlling code, etc.

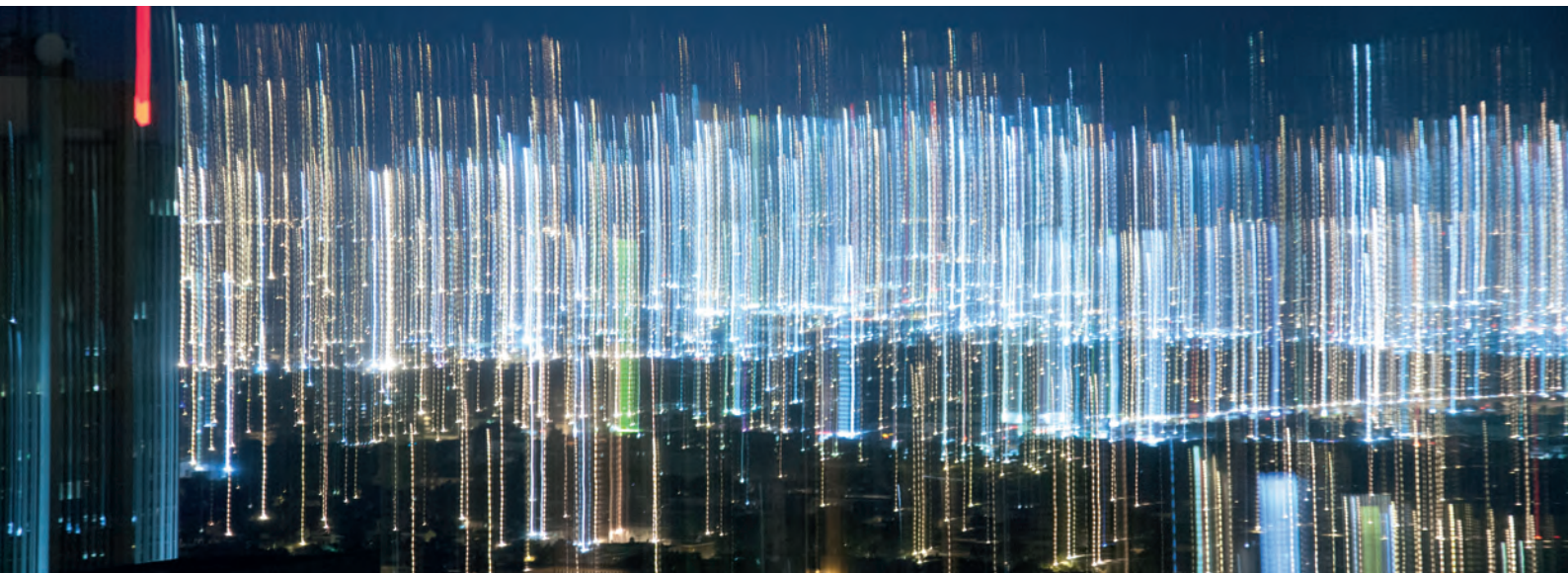
Today, finding the [balance](#) between the degree of centralisation and distribution of data governance and infrastructure is a major issue in all organisations. To decide, it's important to consider criteria such as: the size and complexity of the business, volume, regulatory aspects, types of data and degree of maturity and the data management talent of the different business units. Whether or not these distributed capacities already exist, and the time it may take to create them are fundamental criteria to be taken into account in the decision.

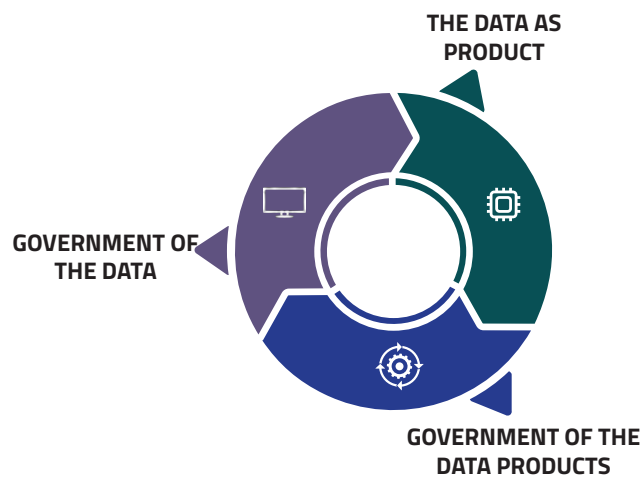
02

The role of the Central Office (Data Management Office) in a federated environment

In this federated environment of generating value from data products, the data office is a central unit that acts as a control tower and provides the different business areas and data domains with:

- A series of **global data governance principles** that include a common framework for data access, metadata, lineage, and security.
- The principles that define a data product and the **rules for creating them**. For example, the different federated domains are considered to be data products and are responsible for the content, quality and interoperability of the data they manage. The different information solutions, analytics, websites, apps are also considered to be products and consist of three components: code, data and metadata, and infrastructure.
- The **description and responsibilities** of the different participants that play a role in distributed Data Governance: Domain business owner, IT level owner, DevOps team responsibility, etc.
 - ◇ For data to be considered as products in an organisation, the data of the domains and also the data products of those domains must have a Product Owner. A person who is responsible for making the data discoverable, or in other words making the data accessible and able to be directed towards consumers. The data must also be self-descriptive, that is, we have to understand their meaning so that they makes sense for the business.
 - ◇ It's also important to manage who can access and publish data, and guide data consumption with metadata, such as use cases and sample queries, and of course to monitor internal consumption to evaluate the return on investment.





We also have to consider the creation of **data sharing agreements** (DSAs) that lay out the principles of **responsibility, quality and security** between the producer/consumer. DSAs define:

- Quality standards and tolerances.
- Responsibility for data and the related roles.
- Acceptable use of data: The purposes for which we accept the use, processing and exploitation of the data.
- Infringement protocols.

Using DSAs enables:

- Traceability of the assignment and the proper management of identities.
- Application of encryption/anonymisation mechanisms.
- Reuse of data.





03

Data infrastructure or platform in a federated environment

Having a data infrastructure or platform **equipped with technical accelerators** that are used by the entire organisation is essential to increase the efficiency and productivity of product teams. This means that it's essential to have a sufficient and optimally managed budget to build the platform. Normally, the different components are deployed in parallel with the construction of the products, minimising the "technical debt" that is generated.

In addition to the components commonly used by the different businesses, there are reference architectures for products that normally serve as preferred architectures, but that are not mandatory.

This data-as-a-service platform or infrastructure is designed and maintained by a core team of engineers and data architects. **The components used most commonly** are:

- Data Governance tool (multi-cloud/hybrid).
- Marketplace for the data and data products generated by the organisation and made available to the different organisational units.
- *Knowledge store: Documentation of the solutions; code, architecture...*
- Catalogue of IT assets that can be reused for data products.
- Security and identity identification.
- Help Desk Zone to manage incidents and maintenance.

Each organisation has its own data ecosystem and its own challenges. Not all organisations are ready for large-scale decentralisation. However, a balance can be found by choosing a reference topology that best suits our particular needs.

04

Data domains

One of the key aspects in a federated organisation is the definition of the domains, and we recommend:

- Establishing **data domains with clear boundaries**, each with an assigned DataOps team. For example, a data domain might be a functional or business area of the organisation. It needs to make sense for the organisation; if not, establishing clear ownership is difficult, and each domain must have a Product Owner.
- Each domain is responsible for **debugging and setting the quality controls for its data**. It's responsible end-to-end for having data ready for consumption and delivery to the data products that need them, whether from its domain or other domains. Treating data as products refers precisely to being clear about the standards that the data must meet and the application of the standards by the Product Owners of each domain.
- In each domain, it's important to **separate the data from the data products** so that the data can be reused by multiple products. Otherwise, if data already incorporate certain transformations or business logics, they cannot be reused for other uses.
- All domains share the same **interoperability standards** that enable them to share data. In some cases, schemas need to be defined when data are shared by multiple domains. There are several solutions such as setting up data provider-client domains.

At the same time we are defining the different domains, federated data governance must ensure that responsibilities and assets are distributed between core elements and domains in such a way that they maximise the value we get from the data. It's important to note that:

- Although we talk about a federated data governance and architecture model, it's essential to have a **global perspective and common elements** for the model to be efficient and for the assets that it generates to be reusable.
- **Each domain uses common accelerators from the global platform**, but may require its own specific set of technology.
- **The domains operate their own applications and products** and adhere to global principles.
- Even if there are several domains, there must also be a data catalogue that is common to all of the domains. In other words, there is a **central data management hub** that maintains a common catalogue and performs global operational and economic monitoring. This hub also provides services for automation, data discovery and shares best practices to reuse the assets that are generated in all of the domains.

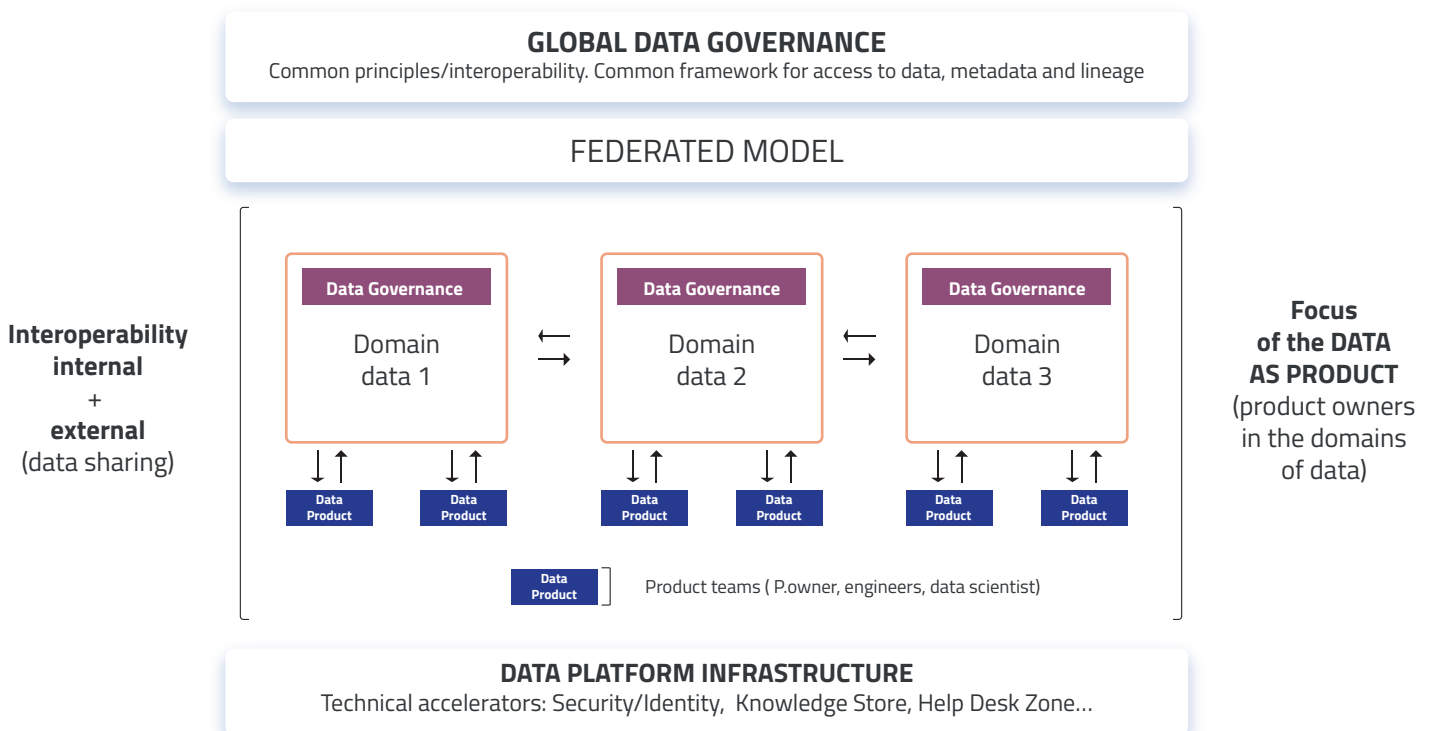


The transition to providing value through data

Organisations are trying to use data to drive **differentiated customer experiences**, create platforms and services, generate knowledge and redefine their operations through smart work processes by using technologies such as Artificial Intelligence.

Data is the most important intangible asset of organisations, but it's often distributed in multiple transactional applications, analytical environments, and Cloud environments inside and outside the organisation, which makes access and internal consumption difficult and leads to the appearance of information silos and duplication of work.

The evolution of data governance to expand its scope from managing raw data to managing data as products and the products generated from them significantly increases the impact of these initiatives in terms of the value that they add to the business.



Some of our **recommendations** for transitioning to data governance 2.0 include:

- Put in time and effort to understand the capabilities and internal dedication that you'll need both at the business level (for each domain) and at the data and IT level to set up a new Governance 2.0.

- Don't improvise when determining which components of the data platform are central and which are left to the management of each data domain. The current technological dispersion doesn't help, but it needs to be very clear at the conceptual level.
- Implement the principles of governance, interoperability and security in the Data Governance architecture and tools. The times of generating regulations and paper that no one reads or consults are long gone.
- Even though we're talking about federated or distributed models, the Central Data Management Hub still plays a vital role because of its global governance and monitoring function. It must have sufficient human and economic resources, among other reasons because it will manage the overall adoption of the model.

Lastly, it's important to note **that extending governance to data products helps respond to the strategic imperatives of organisations:**

- Be able to govern the analytical models of data exploitation by defining performance metrics, creating catalogues, etc.
- Measure the real contribution of initiatives through the economic valuation of data and data products, and generating internal and external monetisation models.
- Promote the adoption of the new culture, knowledge, processes and policies, incorporating aspects of change management and data literacy.
- Data ethics refers to best practices in terms of how data are collected, shared and used. These practices are especially important when data activities could potentially have a negative impact on individuals and society.
- Support compliance with data protection regulations, such as the GDPR, and others that are coming in the short term and will affect the data strategies of organisations: Data Governance Act, Data Act, and the Artificial Intelligence Act.

Let's talk.

Do you also believe in this evolution of Data Governance?

We'd love to talk to you!

Contact us at **t**

