agilelab

The Practical Guide to Successful
Data Mesh Implementations

How to identify, avoid, and solve the most common obstacles for a successful Data Mesh initiative.

While organizations see the value and potential of Data Mesh, many struggle to overcome common obstacles that slow down time to value or even derail the project. This guide describes the common challenges and underlying causes to streamline the implementation and adoption process, so organization will minimize risks and realize the value of Data Mesh.

INTRODUCTION

More and more companies worldwide understand, year after year, that value must be harvested from an effective data strategy. Many organizations aspire to be "data-driven" and huge amounts of money have been invested in platforms, systems, and technologies. More specifically, companies look to Data Lake or Data Warehouse architectures to make these data-driven business decisions, hyper-personalize customer experiences, perform predictive analytics, and several other extremely high-valuable data-driven initiatives.

In the last decade, driven by the "Data is the new oil" motto, many data strategies shifted towards an ungovernable, unconsumable, low-quality accumulation of data supported by more and more powerful data processing engines along with cheaper and cheaper storage thanks to Cloud infrastructure offerings.
Practically, these developments led to opening the data silos and allowing this waterfall of data flow into the hands of centralized data engineering teams. These teams possess excellent technical skills, but often lack business acumen and context, and have limited control over the entire data lifecycle. This creates a cycle of change management dependencies and delays that result in bottlenecks, slow time-to-market of data initiatives, and increasing frustration among both data experts and business users.

Successful companies require complex data strategies that go beyond making data more accessible, (e.g. adding virtualization layers to provide unified access like in the Data Fabric model) -- instead, they need to embrace dynamic, lean, and solid organizational models, based on:

- decentralization and distribution of data ownership
- computational automated and federated governance
- flexibility, and scalability in the data infrastructure
- accessible, trustful, discoverable, and easily consumable data assets.

The Data Mesh paradigm is rooted in addressing and resolving these issues. Over the last few years, its pragmatic approach to solving the issues of traditional data architecture has gained traction, thanks to its technology-agnostic approach aiming to address the scalability of the whole organization floating around the data world.

Data Mesh now attracts worldwide attention, and many companies want to exploit this potential in their organizations. Many have already started the journey, while others are struggling to get started because Data Mesh can be hard to:

> understand   > communicate   > implement correctly   > scale

Like all transformations, it involves people who may struggle to embrace this change. To succeed, organizations need an easier, lower-risk path forward.

DATA MESH HIDDEN RISKS
Identifying the key risks that threaten overall Data Mesh success

Data Mesh distributes the data management practice across multiple domains (Domain Oriented Ownership principle). As shown below, the core principles of Data Mesh work together to avoid several key problems found in other architectures:

- Data silos creation , prevented by the *data as a product* principle
- Duplicated efforts to set up and maintain infrastructure, mitigated by *self-serve data platform*
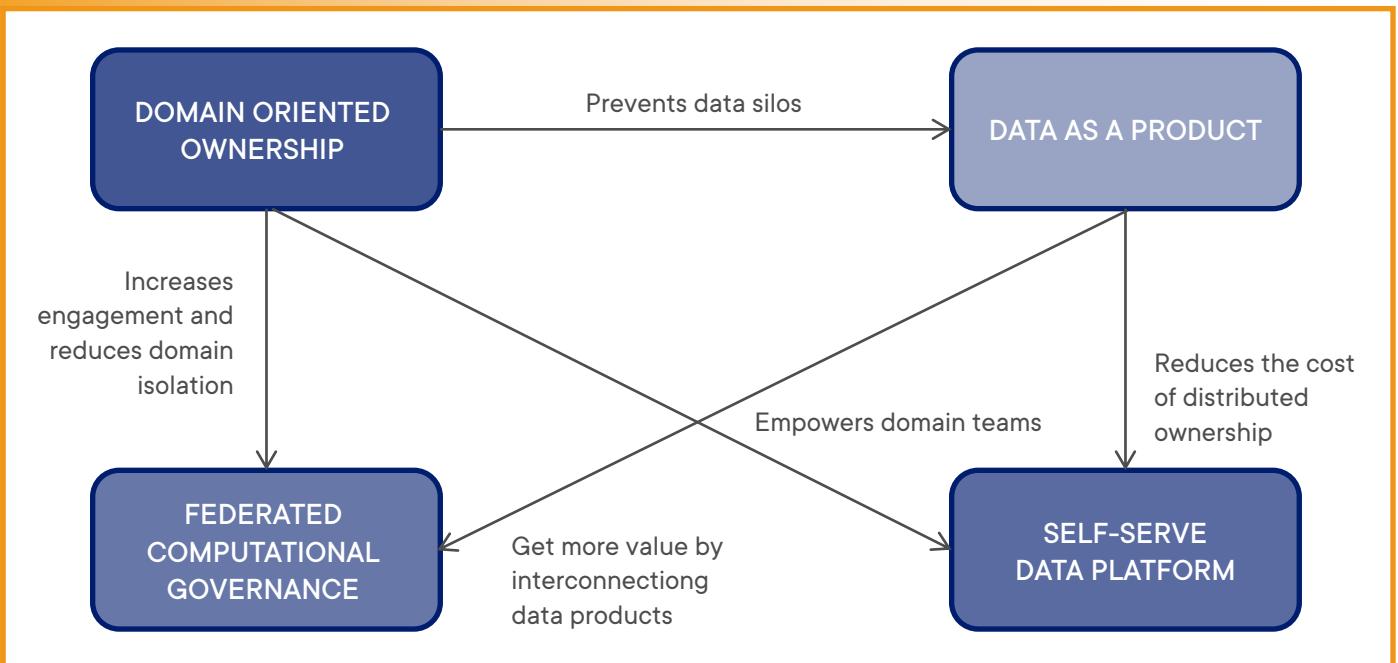- Practice divergence, mitigated by federated *computational governance*

Figure 1 - How the 4 principles of Data Mesh support and reinforce each other

The big problem is that an organization cannot implement a Data Mesh by simply defining the architecture and providing guidelines, this approach will not scale at an Enterprise level.
All the principles must be enforced to ensure the end-to-end experience that final users create and consume Data Products while enforcing a sustainable governance and maintenance effort of the data ecosystem. In the long run, many unknown problems will arise because this approach creates new organizational patterns, new team relationships, and dynamics. The 4 principles provide a stong, flexible framework to address future issues.

Data Mesh is a technology agnostic pattern and canot be solved with a technology-driven approach. Many technology vendors propose their own solutions around Data Mesh, but these appproaches only partially deliver the four principles. Specific technologies may fit into some of the aspects (e.g., interoperability and data consumer experience), but they will never allow the organization to implement a holistic Data Mesh.

Agile Lab leads successful Data Mesh initiatives with its customers since mid-2020. This hands-on experience provides a tremendous insight to what will come further down the road.

This paper highlights all the significant impediments in scaling a Data Mesh initiative, generating awareness, and supporting standard. It also covers how the Data Mesh Boost product helps an organization solve these challenges, maintain a technology-agnostic approach and embrace all the core Data Mesh principles. The result is a functional Data Mesh implementation delivered more quickly and with lower risk than other approaches.

Data Mesh Boost delivers three primary components that together manage the end-to-end lifecycle of the Data Products:

Data Product Builder

Allows building a Data Product in 5 minutes, leveraging ready-to-use templates with a technology-agnostic approach. This is an implementation on the Data Product Experience plane.

Data Product Provisioning

Implements a Utility plan that provides self-service capabilities to seamlessly deploy Data Products without any human intervention.

Data Product Marketplace

Deploys the Data Mesh Experience plane is to provide information, transparency, and trust to the Data Mesh users.
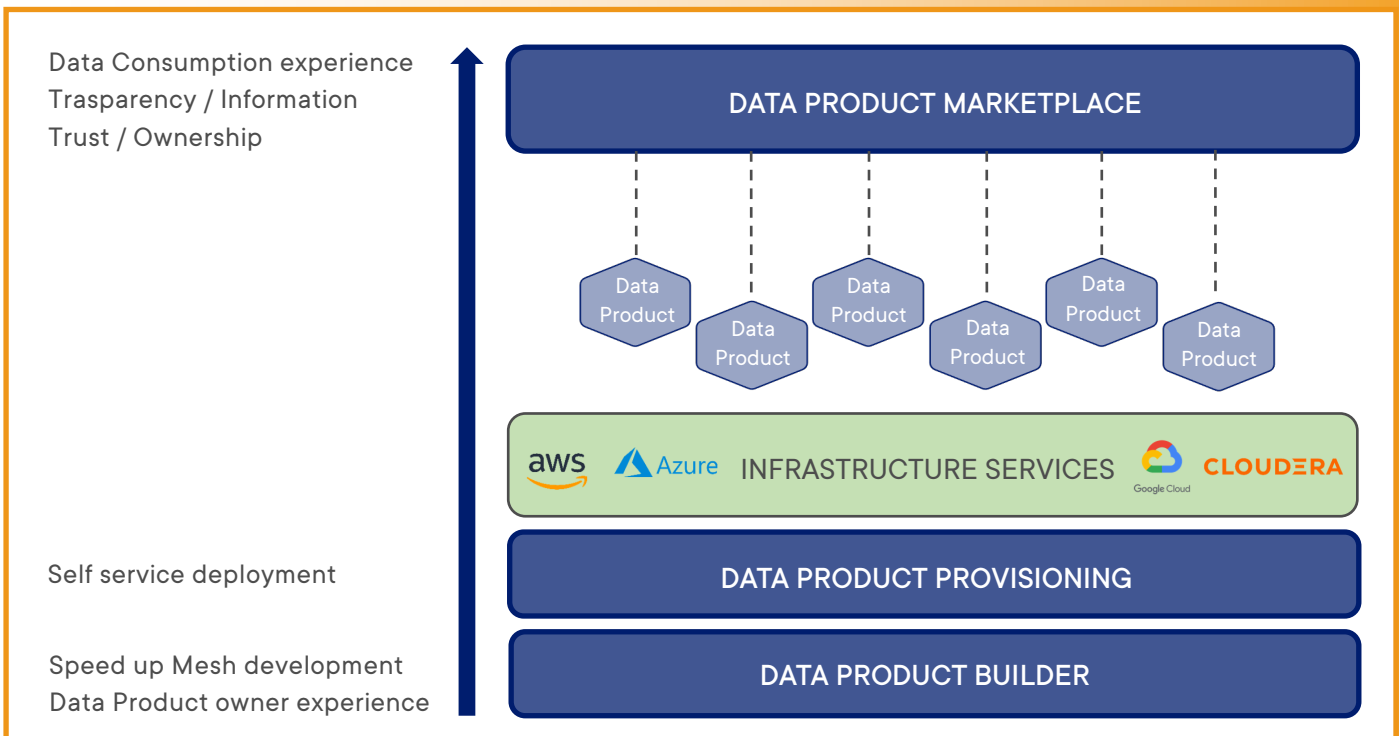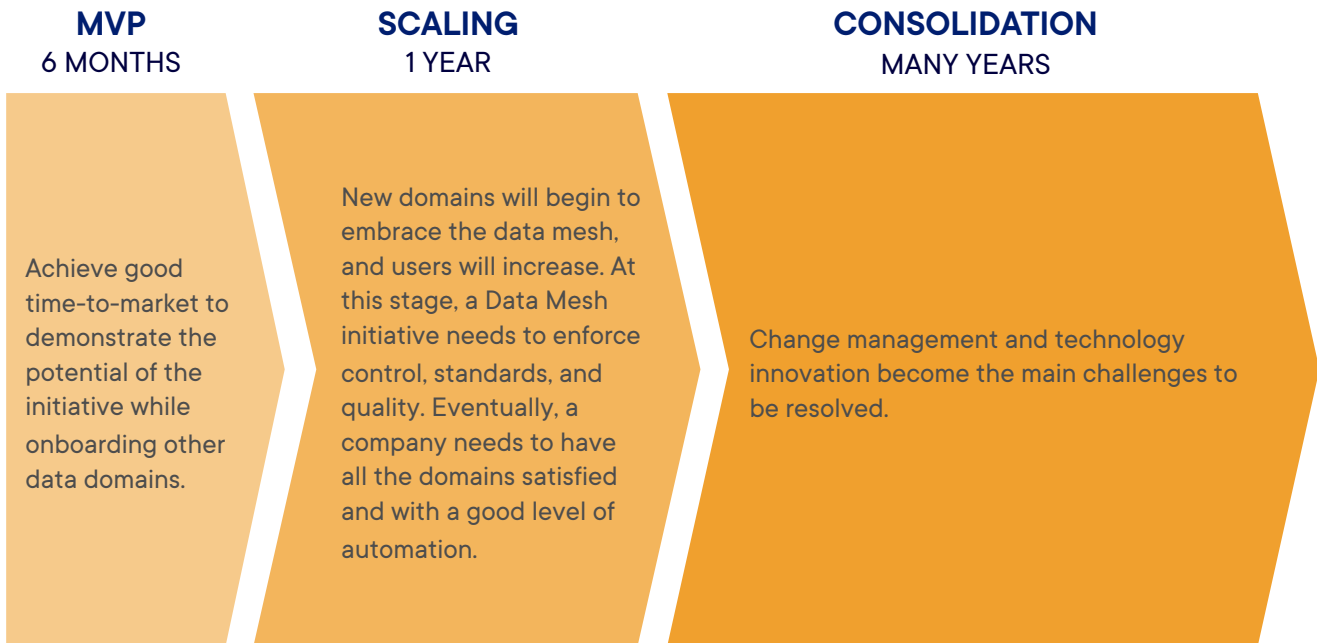

Figure 2 - The three Data Mesh Boost components work together to manage the end-to-end data product lifecycle.

For clarity, consider the challenges an organization faces with implementing Data Mesh based on a timeline. Consider T0 as the time an organization decides to start a Data Mesh implementation. Thus, an organization will experience the following three phases:

MVP
6 MONTHS

SCALING
1 YEAR

CONSOLIDATION
MANY YEARS

Achieve good time-to-market to demonstrate the potential of the initiative while onboarding other data domains.

New domains will begin to embrace the data mesh, and users will increase. At this stage, a Data Mesh initiative needs to enforce control, standards, and quality. Eventually, a company needs to have all the domains satisfied and with a good level of automation.

Change management and technology innovation become the main challenges to be resolved.

MVP PHASE
Create a foundation for success from the very beginning

How do we accelerate the time-to-market for Data Mesh?

In the beginning, some organizations attempt to develop Data Products (DP) even without the automated support of a platform. However, a Data Product Platform will become increasingly important as adoption grows. If an organization starts developing DPs without platform capabilities, it must consider rework cycles within domains to adapt to them. This can repeat across domains if platform capabilities are released iteratively, adversely impacting domain engagement. Finding the right trade-off between the Data Product's time-to-market (to demonstrate value to the rest of the organization) and platform capabilities is not easy. If a company wants to wait to build a fully-featured platform before starting DP development, it risks losing momentum and traction because it will not be clear how complex a Data Mesh platform could be in that organization. It is crucial to have a solid foundation of platform capabilities since day one to minimize domain rework.

Experience demonstrates that Implementing a fully-featured Data Mesh platform from scratch could take more than two years. Based on experience with Zalando and other companies that are actively implementing data mesh, it is clear that even the pioneers are still working to build a comprehensive experience around it.

CHALLENGES

How can organizations enable DPs development from day one?

How can organizations minimize DP team rework as platform capabilities grow?

IMPACTS

Introducing platform capabilities later means generating from three to six months of rework for each domain.

SOLUTION

Platform-thinking is the way to go. A Data Product Platform means technical complexity and misalignment to standard or best practices can be almost entirely removed from domain teams' shoulders. Instead, the Platform enables a self-service and guided DP development experience. This reduces the need for large tech teams in every domain and accelerates the time-to-market of data initiatives, for example, by providing templates of Data Products for common use cases.

THE AGILE LAB APPROACH

Data Mesh boost enables a simplified DP developing experience by providing ready-to-use capabilities to quickly develop and deploy quality and compliant Data Products from day 1, without generating rework later. It acts as a platform for developers (but also facilitates data consumption) and it is fully customizable to customer needs and standards. It is designed to evolve over time and not generate breaking changes that impact domains.

Data Mesh boost manages the end-to-end lifecycle of Data Products, empowering the development of a healthy Data Mesh without creating any lock-in in technology, standards, and practices.

The main extensibility points are:

Specific Provisioners: using infrastructure templates, an organization develops its own provisioning routines to incorporate security standards and secret management, developed in whatever language or paradigm it requires.

Data Product components: starting with use-case templates, organizations create a custom starter kit that best fits their specific use cases and technologies.

UI Plugins: organizations customize the UI by adding plugins to deliver micro-frontends from third-party platforms or applications.

Provisioning Plugins: organizations share a Data Product's metadata to other platforms (e.g., Data Catalogs)

Builder Hooks: *Data Mesh boost* creates custom hooks triggered whenever some action is performed on the Data Product Builder, for example, a post-commit hook or post-deploy hook.
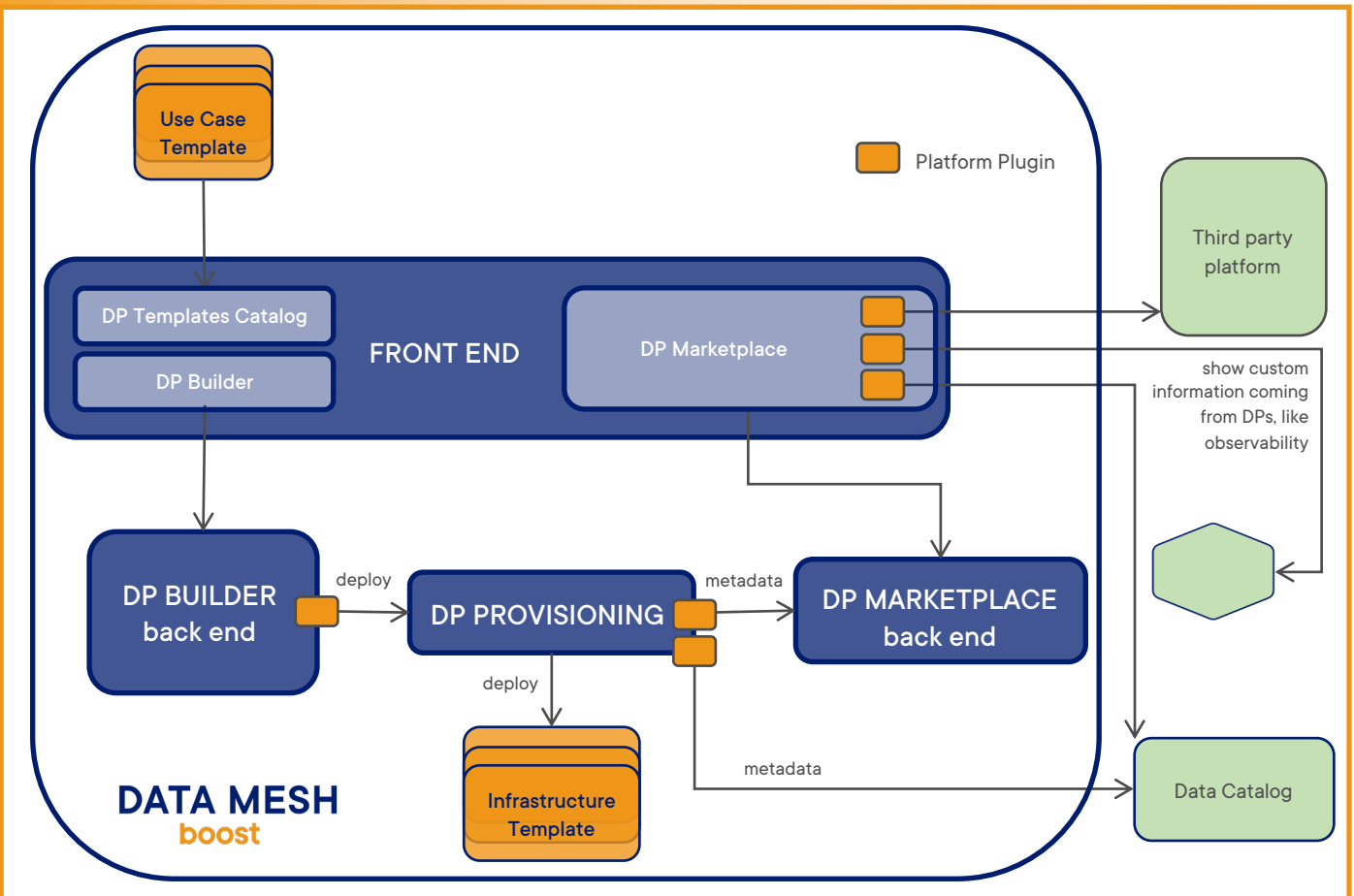

Figure 3 - The main extensibility points of *Data Mesh boost*

BENEFITS

Data Mesh boost accelerates the implementation time of the platform from 2-3 years to 2-3 months.

How do we reduce the time to create Data Products?

Data Mesh's main goal is to gain agility and speed in data initiatives. At the same time, this approach shifts many responsibilities to DP teams, particularly around data quality. Distributing ownership across domains that typically have little or no data engineering experience creates an inefficient data engineering practice. How does an organization hope to gain speed, given this scenario?

CHALLENGES

How does a company guarantee there is enough automation and fast-tracks for developers? Do they have to deal with Infrastructure as Code (IaC) or infrastructure provisioning? Will they be able to meet security standards and achieve a reasonable time-to-market with all these accountabilities on them?

> How does it enable quick wins for DP teams? Their engagement is necessary from the beginning.

How is it possible to reduce complexity for DP developers?
> A DP is made of multiple components, and developers will struggle to manage all the repositories and associated metadata. DP should be versioned and immutable, so the number of repositories will soon increase.
> DPs must be standards compliant. Do DP teams need to know the specs of all the standards to implement their DPs?

How is it possible to make DP teams' life easier?
> How painful it is to develop the first DP?
> What happens when they find out the standards are not being met and they have to rework the DP?
> How do they figure out which version of the DP is currently deployed in each environment? Also, since the DP is a set of components, are they able to easily check which elements are part of DP version X?
> How do they review a DP before deploying it? Is it easy, or do they have to navigate across multiple repositories?

IMPACTS

The time-to-market to implement Data Products is one of the primary metrics that teams will evaluate. Developing a single, simple Data Product from scratch without facilities may require more than one month. When this is multiplied across multiple data products and domains the resulting inefficiency is unsustainable.

Even the fastest developers team cannot compete with a set of well-defined, certified, compliant-by-design sets of templates and blueprints. Similar to the extremely successful approach used with microservices, where containerized semi-developed services were made available by the "platform team", most of the Data Products in a Data Mesh can be eventually factored into ready to (re)use templates, thus achieving several critical enhancements at once. Data Product templates:

SOLUTION

1. include (and hide) by-design quality/security/compliance standards;
2. enforce federated governance policies
3. reduce the friction in the adoption path for teams without strong data engineering and DevOps skills
4. limit the variety of technologies to be adopted by the various domains (no particular impediments are to be set to domains, but having a specific reference architecture dramatically reduces the data/systems integration effort)
5. extend the development resources of the "current" central data engineering team.

Templates are pieces of software that become services + documentation + computational federated governance policies. It's a complex deliverable that can require time to be implemented, but it's an effort that is well repaid by the extremely improved time-to-market of data initiatives that can align with the set of technologies and standards integrated with the template. Starting the Data Mesh Journey with a set of templates for the most common use cases, targetting the main infrastructure vendors, already available, is the best option.

Furthermore, Data Product templates should be modular so internal components have a defined implementation and lifecycle. Examples include components definition and provisioning, e.g. an AWS S3 bucket as output port, or a certain type of analytical workload like a Spark job on Azure Databricks.

THE AGILE LAB APPROACH

Data Mesh boost eliminates the burden of managing multiple repositories and complying with standards and formalisms.

Leveraging the templates, an organization can build a simple Data Product in 5 minutes, already compliant with standards and without requiring specific expertise in the Data Mesh paradigm.

The standards and formalisms are already built into the templates. Data Product teams only need to follow the wizard that is automatically created by *Data Mesh boost* based on the template definition.

Data Product Template

1 Data Product details

Data Product name *
Card Transactions
the Data Product name

Fully Qualified Name
Card Transactions
Human-readable that uniquely identifies an entity

Display Name
Card Transactions
Optional name used for display purposes

Description *
Exposing all kind of transactions that can happen with debit or credit cards
Help others understand what this Data Product is for.

Domain *
cards
the Domain of the Data Product

Owner *
Jason Miller
Owner of the Data Product

Figure 4 - The Data Mesh Boost template cloning process allows simple DPs to be built in 5 minutes

Data Mesh boost also manages code repository organization to keep all the reports under control and compliant to a standard naming convention. When a template is cloned, it automatically goes into its own Data Product "area" that reduces human errors, enables automation, and improves the developer experience.
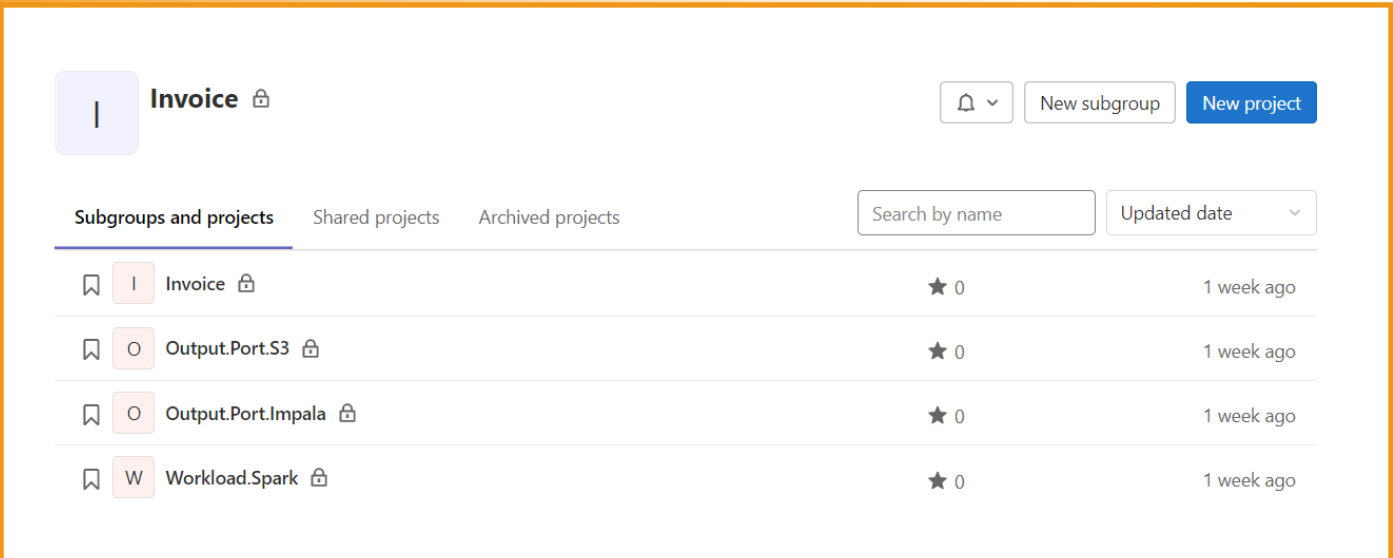

Figure 5 - The Data Product repository structure organizes the code repository to reduce human errors, enable automation, and improve the developer experience.

Data Product teams no longer need to deal with IaC because *Data Mesh boost* provides a self-service experience. When a template is cloned, there is another benefit: it is automatically linked to a specific provisioner that handles deployment, meeting all security standards and infrastructure-level requirements.
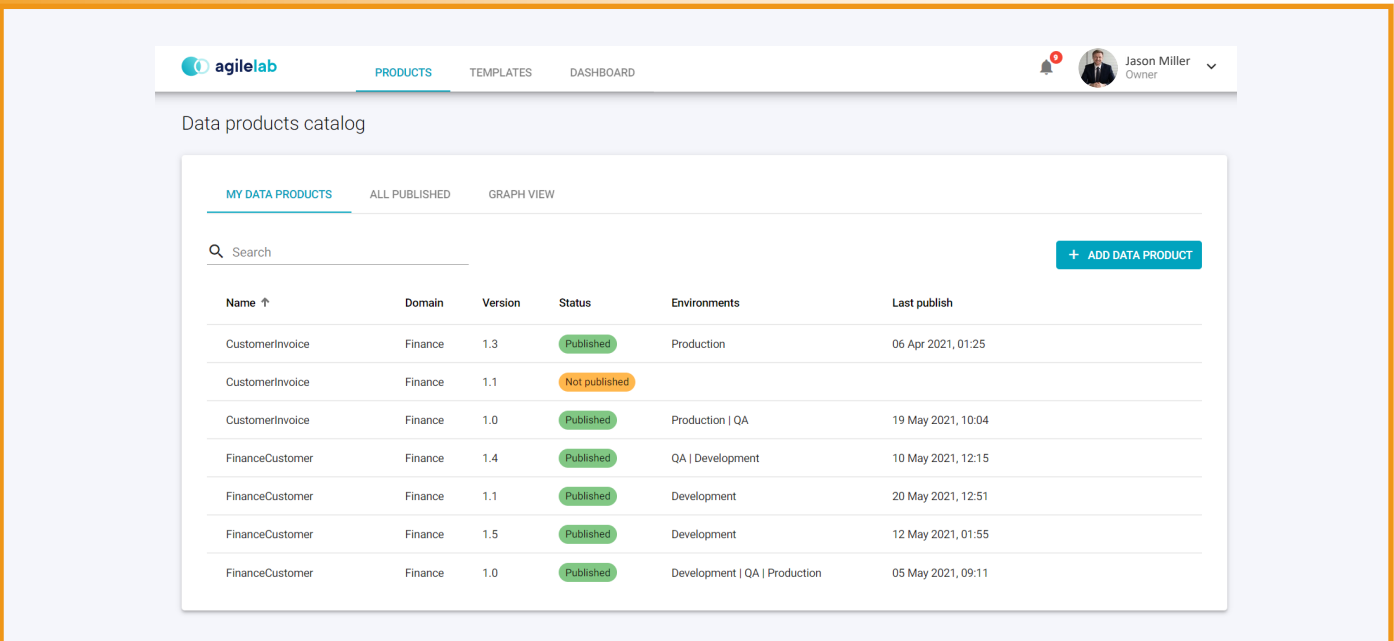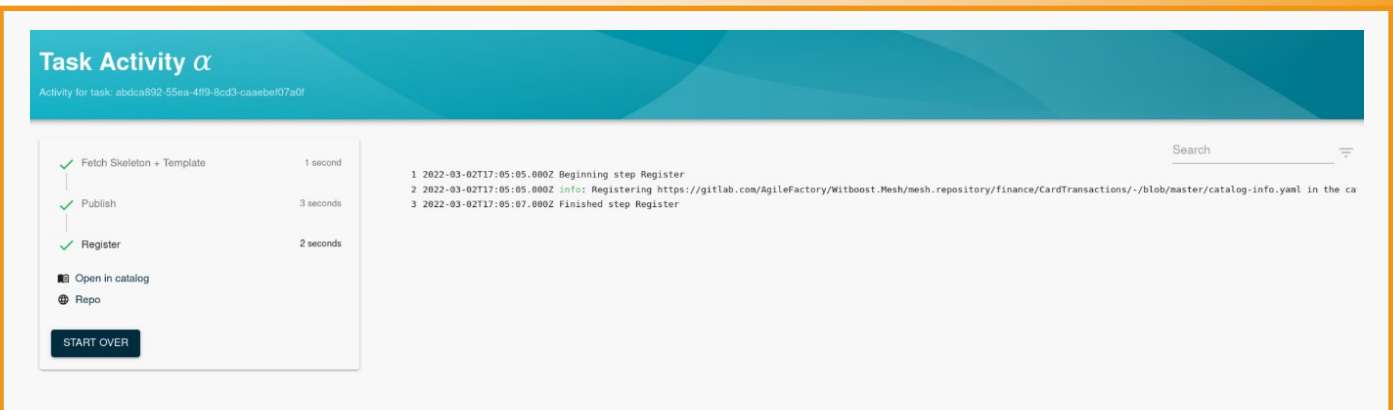

Figure 6 - Data Mesh Boost automatically handles deployment, with all security standards and infrastructure requirements.

Once Data Products have been deployed, *Data Mesh boost* will ensure control over them knowing which version is currently in use and in which environment they were deployed.

In addition, *Data Mesh boost* offers innovative capabilities to duplicate and create new versions of a Data Product. This enables and enhances a common practice, allowing the evolution of data products without risk of errors in the change process.

Figure 7 - *Data Mesh boost* automatically handles deployment, with all security standards and infrastructure requirements.

The Data Product builder UI allows users to perform all the main actions required to manage the end-to-end life cycle of a Data Product:

- Create a DP
- Test a DP
- Commit a DP
- Describe a DP
- Deploy a DP
- Create a new version of a DP

- Clone a DP
- Modify a DP
- Undeploy a DP
- Revert a version of a DP
- Delete a DP

Of these, the commit action is essential because it allows to create an immutable snapshot of a Data Product and all of its components. This snapshot is what will eventually be deployed. Without this step, it will be difficult to tell which version of each component of the Data product has been deployed to a specific environment.

In addition, DPs will be tested against all computational policies, as a build step, prior to commit. This uncovers any issues before committing and deploying a DP, saving significant time and reducing frustration. This testing provides significant details on what elements are non-compliant to increase team awareness.

Data Products can also be deployed to the Global Marketplace. All *Data Mesh boost* customers can publish their entire data products and monetize them. In addition, system integrators can publish ready-to-use Data Products to the global marketplace as a revenue-generating activity.

Data Mesh boost customers can access the Global Marketplace and download ready-to-use Data Products into their workspace. They need to adapt them to their own standards, models, and potential differences in the technology stack, but the marketplace products will dramatically speed up the adoption process.

Imagine how many companies use SalesForce as their operating system. The source-aligned Data Products that ingest data from SalesForce will look pretty much the same in all of them. Likewise, many utilities use SAP ISU to manage billing processes, etc.

Instead of talking about monetization of data, consider monetization of Data Products, including the business logic and technology stack. This is a new paradigm and a step toward fundamental composition and reuse of Data Products across companies and industries.
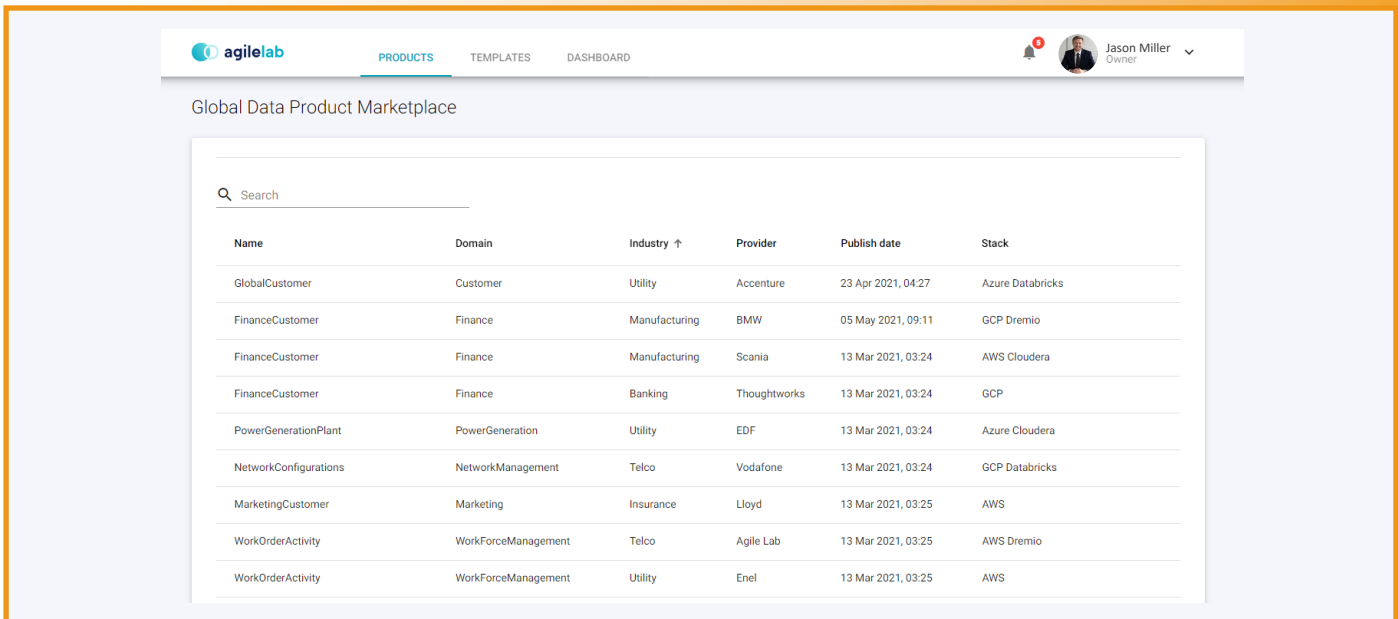

Name	Domain	Industry ↑	Provider	Publish date	Stack
GlobalCustomer	Customer	Utility	Accenture	23 Apr 2021, 04:27	Azure Databricks
FinanceCustomer	Finance	Manufacturing	BMW	05 May 2021, 09:11	GCP Dremio
FinanceCustomer	Finance	Manufacturing	Scania	13 Mar 2021, 03:24	AWS Cloudera
FinanceCustomer	Finance	Banking	Thoughtworks	13 Mar 2021, 03:24	GCP
PowerGenerationPlant	PowerGeneration	Utility	EDF	13 Mar 2021, 03:24	Azure Cloudera
NetworkConfigurations	NetworkManagement	Telco	Vodafone	13 Mar 2021, 03:24	GCP Databricks
MarketingCustomer	Marketing	Insurance	Lloyd	13 Mar 2021, 03:25	AWS
WorkOrderActivity	WorkForceManagement	Telco	Agile Lab	13 Mar 2021, 03:25	AWS Dremio
WorkOrderActivity	WorkForceManagement	Utility	Enel	13 Mar 2021, 03:25	AWS

Figure 8 - The Global Data Product Marketplace allows customers to download ready-to-use Data Products into their workspace.

BENEFITS

Data Mesh boost will reduce the implementation of simple Data Products by over 80% -- from 2-3 months to 5-10 days. Again, multiplied across multiple data products and domains results in a massive increase in efficiency that accelerates implementation signifcantly

How do we address confusion around Data Mesh and Data Products?

Data Mesh is quite conceptual and takes time to understand all the principles and how they affect an actual implementation.

Remember that most people do not understand what a Data Product is.
Not every person in a large organization understands it completely. Usually, they try to stretch DP until it fits with what they've done before or are currently doing. But misconceptions and misalignments create entropy that absorbs a lot of energy from the enablement and solutions team.
It often happens that people in the organization, after realizing that Data Mesh is gaining traction, start claiming that they are already creating Data Products and their platform is fully aligned with the Data Mesh principles.

Successful organizations set very clear boundaries on what is inside the Data Mesh perimeter, and what is not.

According with Zhamak Dehghani, the most common misconceptions are:

Domain-oriented data distribution is the same as Subject-oriented data warehouse design.

Domains are designed.

The Data Team is Implementing a Data Mesh.

Data Mesh is just about data teams.

Data Mesh is just about technology.

Event-driven Data Mesh!

Where to buy Data Mesh?

Data Product is just clean, modelled data with metadata.

CHALLENGES

How can an organization prevent people from claiming to build DPs around that don't meet Data Mesh standards?

> How can organizations define the perimeter of Data Mesh?
> How can organizations prove that a DP is/is not part of the Data Mesh? What prevents people from manually deploying a DP within the Data Mesh perimeter?

Misunderstanding on such topics can lead to several months of wasted development time and undermine the claimed autonomy of domains.

IMPACTS

SOLUTION

Platform-thinking plays a fundamental role here, as the platform provides automated support to make federated governance policies actually computational. Further, an effective data product platform will leverage metadata to create a Data Product Marketplace-like experience of fully compliant and discoverable Data Products. This requires centralization, in terms of a single point of entry for the Data Mesh: Data Product owners looking to provision the infrastructure and deployment of their Data Products must interact with an authority (the platform) that takes care of tracking the data product lifecycle, validation, and certification.

THE AGILE LAB APPROACH

With *Data Mesh boost*, it is immediately clear who is following data mesh principles, and who is not. If Data Products are available in the DP Marketplace, it means they have passed all the embedded requirements:

- Built with supported technologies and with embedded best practices and standards
- Deployed through the provisioner and pass all computational policies
- Published all the metadata in the Marketplace

The tool automatically applies the process and ensures that Data Products are not misunderstood or misimplemented.

The only way to get to the Marketplace and then prove that a Data Product is compliant and part of the Data Mesh is to follow the process proposed by *Data Mesh boost*, standardize it, and eliminate any mystification.

Figure 9 - The Data Product Marketplace removes any mystery about data product compliance.

BENEFITS

Data Mesh boost will bring clarity in your data mesh, ensuring that all the Data Products are fully compliant.

SCALING PHASE
Enforcing control, standards, and quality

How do we eliminate divergence from standards and enforce interoperability?

Data Products need to be interoperable and expose data through highly standardized interfaces (output ports). Defining such standards is a good first step, but as there are many domains involved and hundreds of DPs, it is challenging to have all the people implementing them right. Also, as the standards, it will be hard to do it smoothly because there is no way to implement the new standard across all DPs at the same time. Backward compatibility in an evolving architecture is an absolute requirement.

CHALLENGES

How can companies guarantee that nobody is implementing output ports out of standards?

How can companies introduce new standards incrementally without breaking existing data products?

IMPACTS

Non-interoperable Data Products will create data silos, destroying the Data Mesh mission and bringing the company back into a worse situation than the previous one.

SOLUTION

Quality and compliance governance standards can be cumbersome to be implemented, but even more to be managed, especially if every DP team would be free to "just follow guidelines" without any form of centralized automated authority. Also, standards change over time, like infrastructure requirements for cost management or pure GDPR compliance policies. Some of these can be invisible to users when they can be applied Data Mesh wise (e.g. integrated into Data Product templates and related components' specific provisioners), some others instead need explicit implementation by the DP team AND verification + certification but the authority.
Achieving this in a scalable way requires a central platform that manages the Data Products lifecycle, including the whole validation and certification process.

THE AGILE LAB APPROACH

Data Mesh Builder leverages an advanced template engine (powered by Backstage), which allows the governance team to define and enforce standards using prepackaged ready-to-use software components.
With these templates, engineers can spin up an output port with an organization's best practices embedded from the start.
By starting the process with these best practices, overall adoption will increase across the domain team. When the right way is also the easiest one, engineers will adopt it.
Engineers can introduce and enforce new standards through a new template that team members leverage when creating new versions of their Data Products.
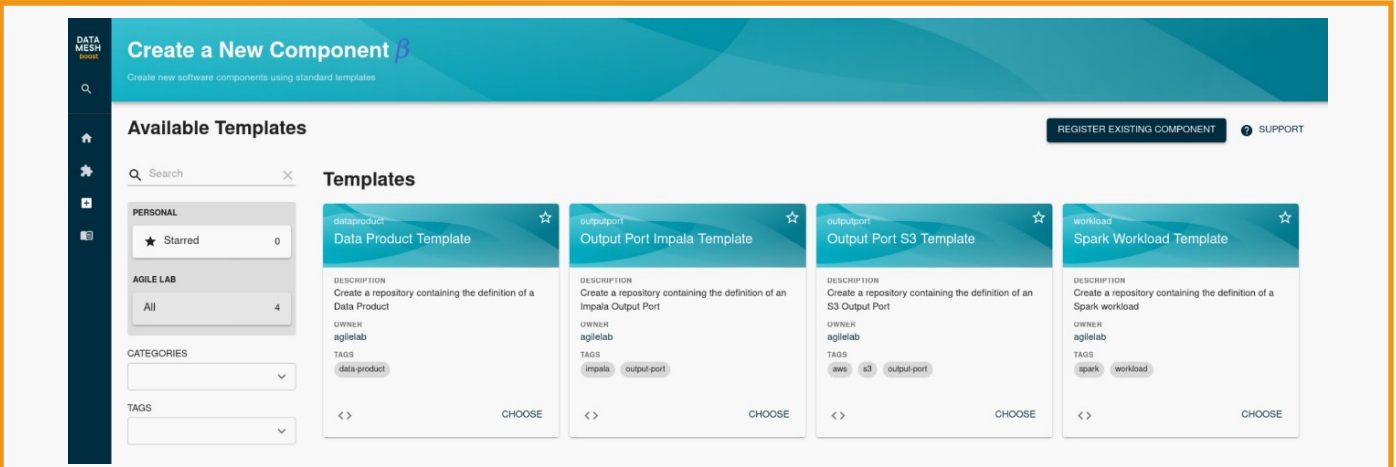
Data Mesh Builder can also track standard usage and adoption level.

For policy enforcement, *Data Mesh boost* provides a validation engine inside the Data Product Provisioning module. This leverages Open Policy standards to enforce compliance with the template and all computational policies defined by the federated governance team.
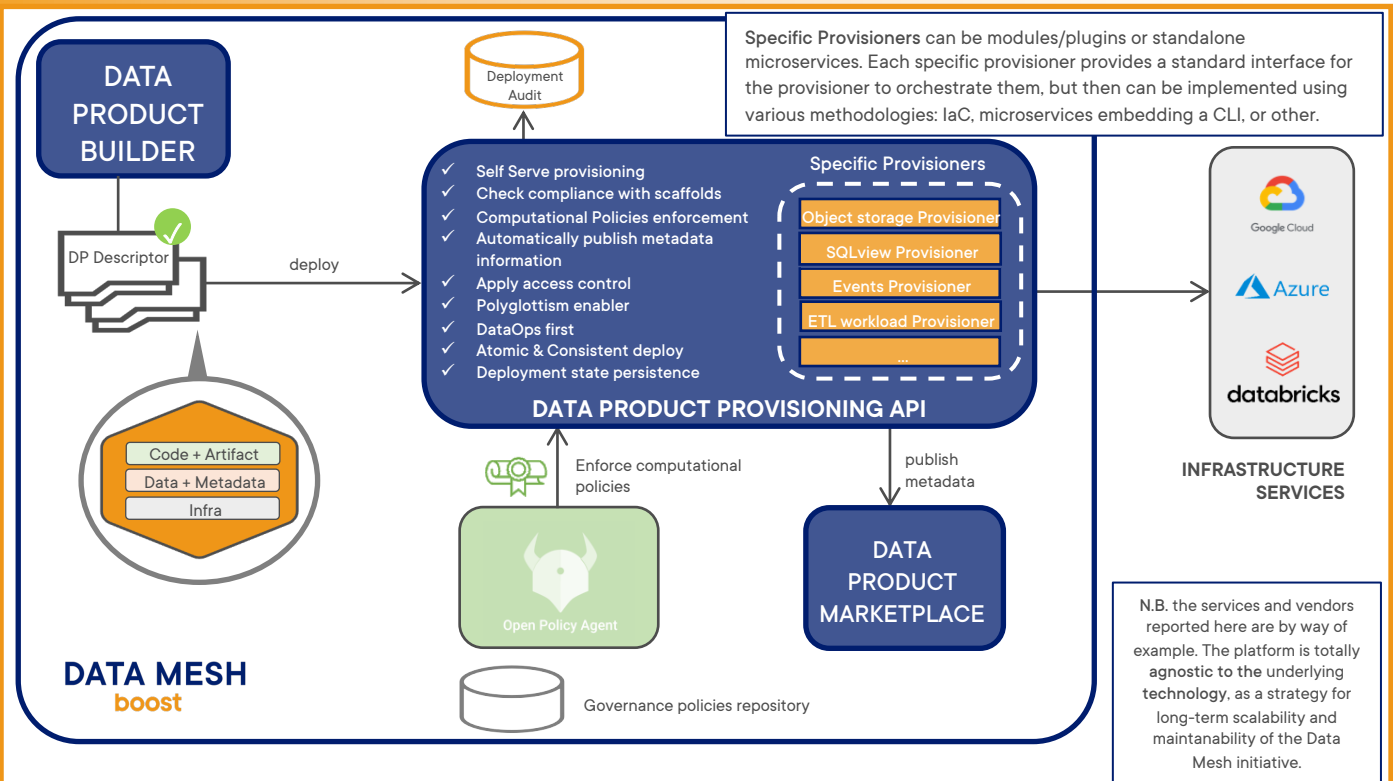

Specific Provisioners can be modules/plugins or standalone microservices. Each specific provisioner provides a standard interface for the provisioner to orchestrate them, but then can be implemented using various methodologies: IaC, microservices embedding a CLI, or other.

N.B. the services and vendors reported here are by way of example. The platform is totally agnostic to the underlying technology, as a strategy for long-term scalability and maintanability of the Data Mesh initiative.

Figure 11 - Computational Policies defined by the federated governance team, are automatically enforced.

BENEFITS

Data Mesh boost will protect the organization from silos creation.

How do we avoid divergence and decay of Data Product Quality?

In the long run, high-quality Data Products will play an essential role in the success of a Data Mesh initiative. Without the proper quality, the project cannot succeed. With distributed ownership, a downstream DP is forced to rely on the quality of upstream ones (even across different domains). It is impossible to create high-quality DPs based on low-quality ones. If even one person lowers the quality standards, it will create a domino effect that will destroy the Data Mesh usability and the mutual trust of the domains, leading to the failure of the Data Mesh Initiative.

This risk cannot be tolerated; quality standards must be clear, enforced by computational policies, and not waivable.

It is also important to consider the decay that will hit the people after the initial excitement. When the innovation agenda creates challenging deadlines and pressures the data owners, they will be tempted to drop quality for speed.

CHALLENGES

How can companies guarantee that all the data products have a minimum common quality standard?

How can companies guarantee that the innovation agenda (or just lazy DP owner) will not bypass quality checks?

How can companies incentivize people to spend time and budget to enhance their quality standards?

IMPACTS

Poor quality will turn the *mesh* into a *mess*, and domains will lose trust in each other. As domains start to duplicate data, creating data ownership misunderstandings, and the overall effort will increase.

SOLUTION

Standards and governance policies range from simple checks on DP's metadata (derived from Data Products declarative descriptors) to more complex but effective features like data quality and data observability. The key element requires these checks cannot be skippable or avoidable. Today, several data observability options exist in the market. These tools can be challenging to integrate and implement but provide some protection for data quality. A fully integrated solution is preferred.

THE AGILE LAB APPROACH

The quality of the Data Product must be a first-class priority of the entire process, and cannot be bypassed. When a DP is deployed, *Data Mesh boost* automatically builds an extended declarative descriptor that is fully inspectable and applies computational policies. *Data Mesh boost* incorporates all the standard quality controls into the distribution phase (Data Product Provisioner).

Defined rule options include:

- Data quality and data observability
- Compliance with open API definitions
- Presence of specific components or templates
- Compliance of metadata and documentation
- Regulation constrains
- Security constraints

Once rules are defined, they will be automatically verified during the deployment phase.
In *Data Mesh boost*, all the Data Products must pass through the provisioner to be published in the Marketplace. This is a step that cannot be skipped, even if the CEO or other executive requests special privileges.

This process ensures all distributed Data Products comply with defined standards and policies.
However, the quality gate is not just a go/no-go step. It creates a quality score published in the Data Product Marketplace. This incentivizes data product owners because the data quality score will be directly associated with their reputation.

Also, *Data Mesh boost* includes the plugin to integrate any data observability framework within the Data Product templates .
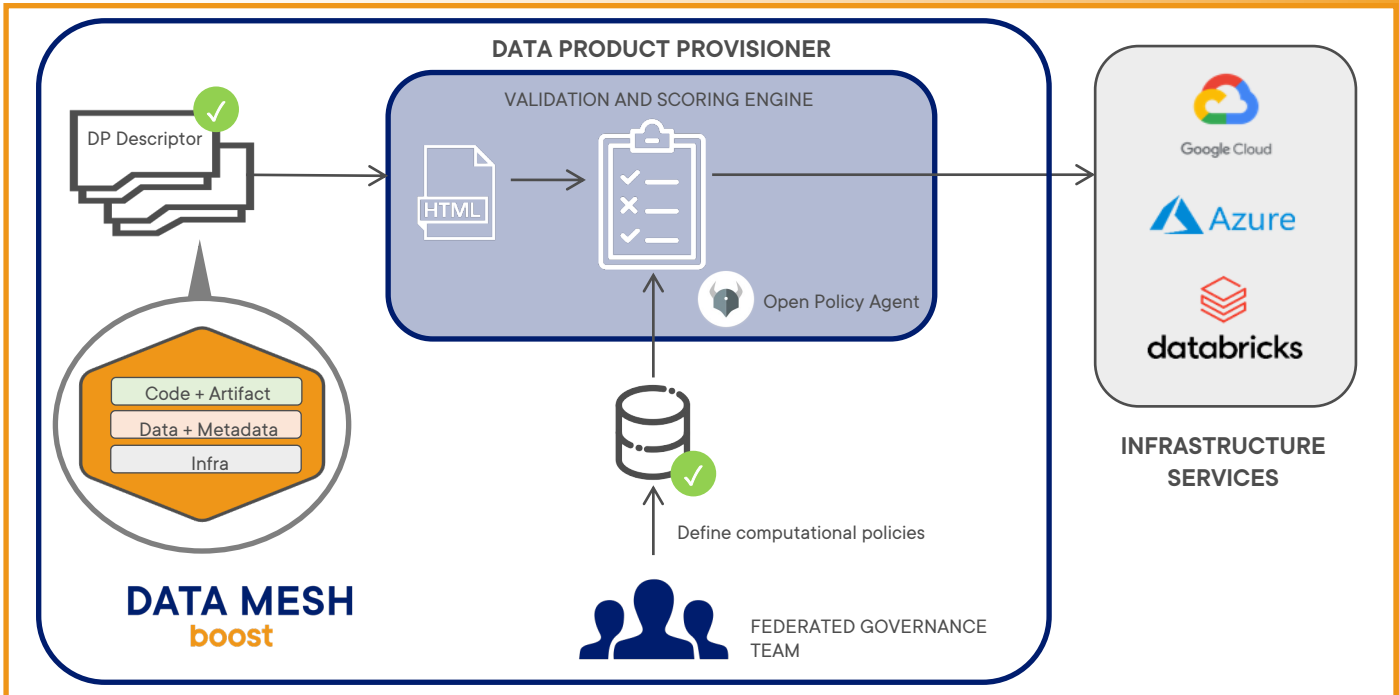
Figure 12 - Computational Policies enforcing quality standards

+

BENEFITS

Data Mesh boost will increase the quality of your Data Mesh enforcing computational policies and creating incentives

How do we scale the Data Mesh practice successfully?

At first, most organizations will start with at most a couple of domains, typically champion ones to provide a better initial understanding of the data mesh process, fewer legacy data systems, and increased ROI. Nevertheless, using only a portion of the available domains in the picture will diminish the full power of the Data Mesh, because the available value-added use cases are limited if only a few domains are available.

CHALLENGES

How can organizations demonstrate the value of a Data Mesh initiative? Which KPIs matter the most?

How can organizations demonstrate the value of a DP?

How can organizations demonstrate that no new silos are being created?

IMPACTS

Data Mesh initiatives that limit their scope to just a couple of domains will fail because data consumers will find it impossible to link and correlate data for their use cases. Fewer domains mean that some data will always be missing.

SOLUTION

To convince other domains to jump into the Data Mesh, it is necessary to demonstrate the value Data Mesh is unleashing, from the very beginning of the process. The value could be time-to-market of data initiatives, ROI, enhanced quality, operational freedom, broken silos, and much more. Relevant KPIs should be calculated (automatically) and made available for consultation to the organization.

THE AGILE LAB APPROACH

Data Mesh boost provides a customizable set of KPIs that can be extended through a plugin mechanism by the customer. These KPIs demonstrate how the mesh is breaking silos and the value an organization is creating.

The two main KPIs included are:

Consumer vs Source – aligned Data Products:

The more consumer-aligned data products are available the more value is generated, if there are only source-aligned DPs, the data mesh is neither creating new data nor extracting value.

Network Effect:

Measures how many relationships between DPs are crossing domains' boundaries. If DPs are exchanging data across domains, then a company is breaking down silos and definitely leveraging cross-domain ownership and knowledge – providing more value.
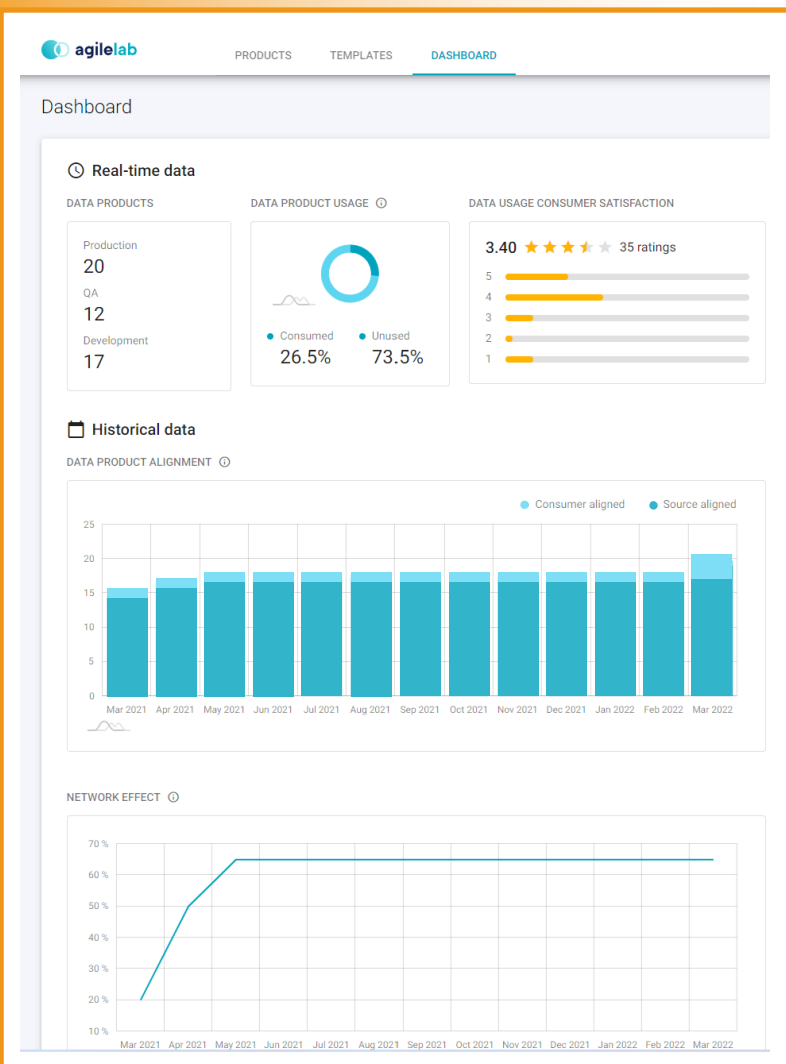
Another Data Mesh benefit is the improved time-to-market for new data initiatives. *Data Mesh boost* calculates the process lead time for each Data Product from its inception to production deployment. Managing the end-to-end process of DP creation and deployment simplifies the process to understand bottlenecks and identify corrective actions as well as opportunities for automation.

+ BENEFITS

Data Mesh boost will clearly demonstrate the ROI of the Data Mesh initiative

Figure 13 - Computational Policies defined by the federated governance team, are automatically enforced.

CONSOLIDATION PHASE
Supporting and managing change and innovation

How do we avoid increasing change management complexity?

Because of decentralization, multiple domains perform their change management independently (and it's the right way to go) without apriori waterfall planning. This means that it is possible (and likely) that a change in a domain will generate unplanned changes in another domain. This effect is called a "change storm" because it potentially generates issues, at the budget and operative levels, across multiple cascading domains which inevitably raises internal conflicts. This effect occurs most often one or two years after the first drop of data products.

CHALLENGES

How can organizations evolve data products without breaking downstream DPs?

How can organizations discover that an upstream DP is going towards the end of life?

How can organizations discover that an upstream DP is about to change?

How can organizations be sure everybody will follow the right path?

How can organizations smooth the communication between domains when a "change storm" will happen?

How can organizations make it easier to embrace change in all the domains?

How could a data product team size its support effort? (Usage information are crucial)

IMPACTS

Change storms could generate rework in several domains, wasting months of work and eroding trust and confidence in the entire initiative.

SOLUTION

An organization must be prepared for this situation to prevent skeptics from having strong arguments that may negatively impact the overall Data Mesh initiative. Shifting from a centralized waterfall change management to a decentralized agile one is challenging and requires time, training, willingness to change, and tools.

At the Data Mesh level, change management drives Data Products development and lifecycle; for this reason, facilitation and standards of interoperability and no-breaking-change policies must be adopted and "granted ", thus increasing domains' agility and mutual trust. For example, DP versioning should be mandatory to guarantee zero impact in case of schema, format, or other breaking changes a certain DP team might need to introduce. An integrated tool for change management helps in multicasting changes to consumers of a certain DP.

THE AGILE LAB APPROACH

Data Mesh boost provides a customizable set of KPIs that can be extended through a plugin mechanism by the customer. These KPIs demonstrate how the mesh is breaking silos and the value an organization is creating.

Data Mesh boost leverages the *Data Product Builder* and the *Data Product Provisioner* modules to track all the major activities across all the domains:

- Creation of a new DP
- New version of an existing DP
- Modification of an existing DP
- Deploy in different environments of a DP
- Deletion of a DP

These events are broadcasted (as notifications) to all DP teams with a focus on those directly impacted by the changes. These notifications create awareness about what changes are pending and when these changes will occur. They also provide a trigger to initiate point-to-point conversations between DP owners to further facilitate communication.

Data Mesh Builder defines deprecation policies that apply automatically when a DP team wants to remove a DP. Along with the deprecation period, the platform will send regular reminders to all the teams that will be impacted once the DP will be removed, so they can plan for their migration.
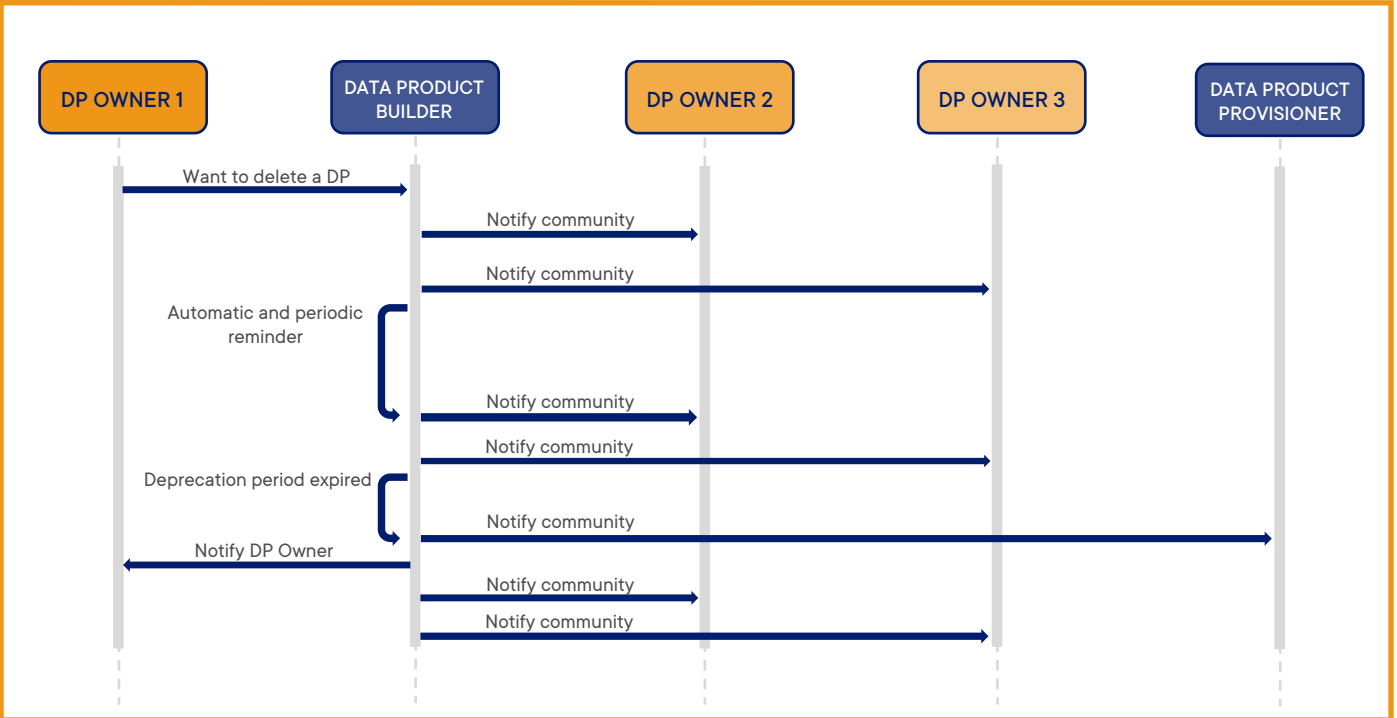
Figure 14 - The change management helper tracks and notifies users about pending changes to their DPs

BENEFITS

Data Mesh boost will smooth the collaboration and the communication across domains.

How do we remain flexible for changing technologies?

Data Mesh is technology agnostic by definition and this is the hardest part to adhere to. A company needs to provide an abstraction to decouple data-producer and data-consumer experiences from the underlying technology. It's perfectly fine to start with a specific technology, but very soon this company will discover that not all the domains agree with this decision. In the majority of cases, this prevents these domains from supporting or even joining the initiative. Often, these disaffected groups fight to introduce into the Data Mesh another technology. Unfortunately, companies that are locked into a specific technology platform will struggle to bring them on board and the entire initiative will be at risk.

In addition, new technologies and opportunities arise constantly, as do changes in strategy. For example, a company may adopt a multi-platform strategy in the future. It is better to plan countermeasures from the beginning, and a technology-agnostic platform offers the most flexibility.

CHALLENGES

How can an organization prevent people from claiming to build DPs around that don't meet Data Mesh standards?

How can organizations abstract away Developers and Consumers from the technology?

Does the platform support a multi-cloud strategy in the long period?

Does the platform allow the adoption of one or more new technologies in the next 5 years? What will be required from a platform standpoint?

Can the platform depreciate a currently used technology in the next 5 years? What will be required from a platform standpoint?

How will adopting a new technology impact an organization with thousands of PDs in the mesh?

The advent of new technologies with strong traction impossible to integrate into the paradigm could cause the whole initiative to derail, wasting all the investments made.

IMPACTS

SOLUTION

Think about standards and create templates to facilitate and speed up the adoption process, but plan for emerging requirements and related template development ahead of time. A platform that integrates templates and DP lifecycle without specific technology dependencies will significantly streamline this effort. Of course, integration concerns or potential migration activities must be considered but they depend on the tech change introduced, so flexibility is critical.

Furthermore, a solution that integrates templates available from a central platform allows proper management of templates' lifecycle including evolution, deprecation, and usage tracking.

THE AGILE LAB APPROACH

Data Mesh boost is designed to be entirely technology agnostic, allowing users to introduce and support any technology through the template mechanism. This means that *Data Mesh boost* customers easily adopt changes including a new multi-cloud strategy or adding multiple technologies to an existing Data Mesh implementation.

When a new technology is needed, two kinds of templates can help:

- Infrastructure Template (or specific provisioner
- Use cases Templates

These templates are linked, so when a Data Product team is using a template to build its own use case, *Data Mesh boost* automatically chooses the linked Infrastructure Template that completely automates the deployment tasks.

In the same way, it's very straightforward to deprecate technologies by deprecating the related templates which there are not available for the creation of any new Data Products. Because templates are including the provisioning process, existing Data Products will no longer deploy. This automatically forces people to adopt new technologies, rework their Data Products and follow the technological innovation.

The decision about introducing or deprecating a technology will be taken by the Federated Governance team. This means that decisions are inclusive, not simply top-down directives.

BENEFITS

Data Mesh boost will free the organization from technology lock-in and will improve its innovation attitude.

Figure 15 - Some of the technologies supported by Data Mesh Boost, a technology-agnostic platform

CONCLUSION

The journey towards successful Data Mesh adoption contains both challenges and rewards. Experience in this evolving subject plays a fundamental role. It's critical to understand the 4 pillars of Data Mesh are intertwined and mutually supportive -- they are all equally important to the overall success of the initiative.

With great autonomy and power comes great responsibility but, to ensure success, teams can't be expected to simply "follow the rules" without support. A technology agnostic platform provides the structure that ensures compliance, safety and trust. This security allows the whole initiative to scale quickly, on an organizational as well as on a technological level. The right platform enables and empowers the huge opportunities the Data Mesh delivers for organizations willing to become data-driven.

Data Mesh boost is the best partner for this journey.

LEARN MORE

ABOUT AGILE LAB

Agile Lab is a specialized Sofware Factory in Data Engineering that offers data platform enabling services as well as business vertical solutions mixing business specific and deep technology know-how.
We have created witboost, an open-source modular platform, that can solve many real-world Data Engineering problems.

www.agilelab.it   info@agilelab.it   Agile Lab   agilelab_official