

# ACCELERATE DATA-DRIVEN TRANSFORMATION WITH DATA MESH

Author:

Giancarlo Vercellino

October 2021

Sponsored by



IDC #EUR148251021



# Accelerate Data-Driven Transformation with Data Mesh

## Ingestion

According to IDC, the amount of data created in the next three years is expected to exceed that generated in the past 30 years. This is due to the COVID-19 pandemic and the increasingly massive use of collaboration platforms. By 2024, about 143 zettabytes will be created, captured, copied, and consumed globally — 19TB for each person in the world. About 21% of the datasphere will be data from embedded systems (including IoT architectures) and metadata (which will become even more important both for data governance and for developing more sophisticated and flexible analytical workflows).

In 2006, Clive Humby coined the phrase "data is the new oil," which is actually a simplification of a more structured analysis. Humby argued that data, like oil, had to be extracted, processed, refined, and managed with great care to realize the full potential of its value.

Why has data become such an important production factor for businesses? In the past 20 years, in addition to the exponential growth of available data, the open source culture has fully established itself within the enterprise domain, with IBM's acquisition of Red Hat in 2019 a key milestone. The open data culture is now even more widespread and pervasive, creating a greater need to share data among organizations and create new value.

Data management platforms have evolved in three areas — open source, Big Data, and self-service — in terms of both technology (from RDBMS to no-SQL) and architectures (with the transition from data warehouses to data lakes and data mesh). At the same time, there has been a gradual transition between an organizational paradigm in which corporate analytics were the exclusive responsibility of a small group of data engineering and business intelligence specialists to a model in which the entire organization is redefined around and through the "democratization" of machine learning and predictive analytics: the so-called data-driven organization model.

## Bigger, Better, Faster

Building a data-driven organization requires some managerial exactness to achieve the following goals: the creation of a data-driven culture, the industrialization of algorithms, the monetization of data, and clear information on the data assets. This demand for exactness creates top management roles with very specific functions that integrate the traditional IT and technology

### AT A GLANCE

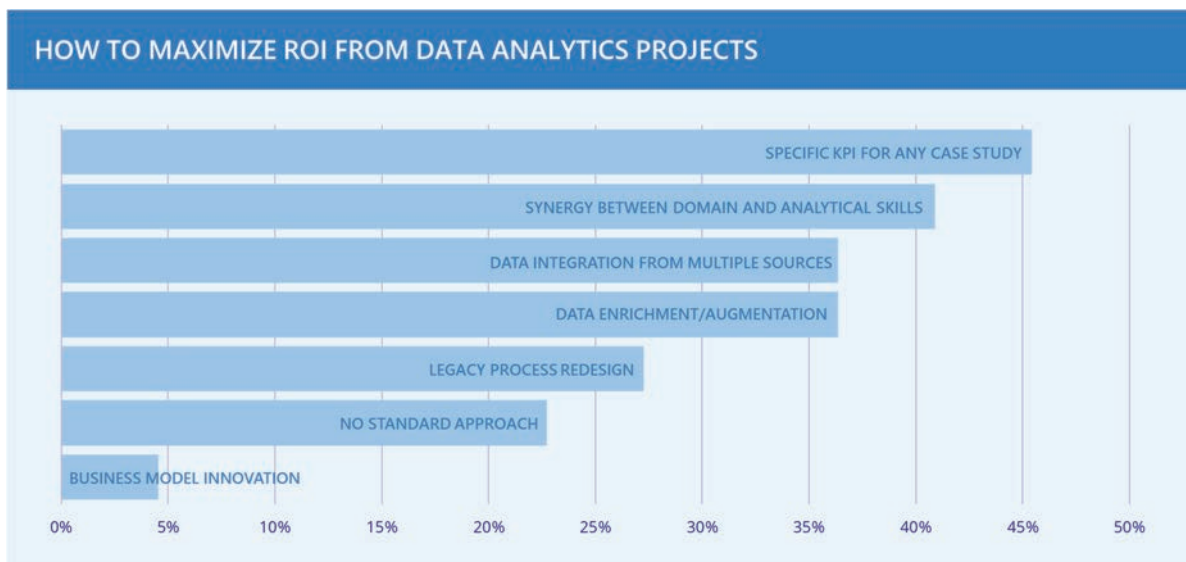
#### KEY TAKEAWAYS

- » By 2024, around 21% of the data produced globally will come from IoT architectures and production systems.
- » 23% of the enterprise segment still has to work on data governance and process augmentation.
- » In the next three years, half of the Forbes Global 2000 will have a dedicated data monetization team.
- » Data mesh architectures can foster the rapid development of a data-driven organization.

roles, from the chief data officer to chief analytics officers. It also involves, in addition to investments in technology, organization, and skills, a reconsideration of the paradigms that guide data architectures. In the transition from a data lake to a data mesh, the organization evolves from a centralized platform to a decentralized ecosystem of data products. This transition leads to a focus on data discovery rather than ETL, on "data as a product" to be sold within your organization rather than data as an entity to manage ingestion and transformation. According to IDC, by 2023 one in four companies will use operational insights deeply integrated into business operations to guide executive decisions and promote new forms of automation. Executives have already identified many use cases worth considering and investing in, from customer service to maintenance, from supply chain to marketing, from IT automation to security.

As data sources and use cases rapidly expand, it becomes very complex for companies to design, both at a technology and market level, a data-driven transformation strategy that enables them to achieve clear and well-defined targets, following an orderly and repeatable approach. When it comes to scaling operations at industrial level, centralized data engineering structures face a number of problems, from becoming an organizational bottleneck to lacking business domain understanding, and from data quality process to regulatory compliance. This can often hinder transformation projects: over 40% of European companies say about one in three projects is likely to fail or deliver results well below expectations. Failure risk tends to double when companies have not implemented solid data governance, while it is reduced to about a third when governance processes enable the joint orchestration of flexibility, performance, and standards.

FIGURE 1  
Strategies to Maximize ROI in Data-Driven Transformation Projects



Source: IDC Italy, Data Strategy Digital Forum, September 29, 2020 (web survey, n = 22)

Data mesh directly addresses these issues, offering an elegant solution to rationalize at a technical level the organizational chaos generated by a tsunami of data and a myriad of use cases, without overlooking the organizational challenges that commonly undermine transformation.



The data mesh paradigm is based on four essential principles:

- The first, and probably most important one, is "data as a product," a shift from seeing data as just an input/output of applications to defining "data products" as the smallest architectural units. Teams with IT and business knowledge can now work on one or more data products that guarantee well-defined, documented, discoverable, observable data, with "time traveling" features, multimodal access, and clear ownership and accountability.
- Data products should be organized together into "data domains," therefore decentralizing the business and technical governance with independency to remove technological roadblocks or enforcements that slow down the change management processes, with a measurable contribution to the overall organization value.
- "Self-service infrastructure" makes the data domain autonomous but keeps cost management centralized.
- "Federated governance" to provide synergy, interoperability, security standards, global observability, and data quality across all the domains.

Only through an architecture with these characteristics is it possible to promote a faster, more efficient, and more scalable data-driven transformation process.

European companies are dealing with the challenge of data-driven transformation through distinct approaches and strategies depending on industrial dynamics and organizational maturity. The development of specific vertical models for analytical purposes and the synergy between domain skills and analytical skills are the guiding principles that often inspire Italian companies. If we analyze the dominant choices related to some specific aspects such as platform rationalization, data governance standardization, data team organization, and augmentation of processes through analytical insights, we see that about 23% of the enterprise segment is still in the experimentation phase, with a level of governance and automation in preliminary development; 46% are taking significant steps to rationalize the platforms and organize the operational teams, following alternative models characterized by a greater (or lesser) decentralization of teams and infrastructures; 31% have already achieved a high level of development and automation, with significant integration of analytical workflows with common business operations.

## Data-Driven Organization = $f(\text{Data Mesh})$

---

Agile Lab's formula is clear: only the transition of data platforms to a data mesh architecture can enable real data-driven transformation. Agile Lab works on cutting-edge data platform technologies, with a specialty in data strategy (the advisory and consulting services to foster the data-driven transformation), development (end-to-end assistance services, from advisory to planning, from design to implementation), managed services (providing management services for data-intensive platforms), and open source (developing solutions with open source technology to combine innovation and time to market). Its experience in working with large customers in the banking and insurance sectors and with complex organizations in the industrial and utilities sectors has recently enabled a rapid expansion of business operations in Italy and

Europe, and has helped to highlight the scaling issues companies face when adopting traditional data warehouse and/or data lake approaches.

Collaboration with the open source ecosystem has defined Agile Lab from the start. The company has made a significant contribution to data mesh architectures by creating witboost, a composable platform that offers open source modules and enterprise features to solve specific data engineering experience problems and that can be used to build specific vertical solutions such as data mesh.

FIGURE 2  
witboost: Composable Platform by Agile Lab



Source: Agile Lab

Agile Lab's modus operandi includes six steps:

- **Strategic alignment.** In the first phase, project goals are defined and an adoption manifesto with the client's management signed, defining the new organizational blueprint and the architectural principles. This leads to the transition from a data lake architecture to a data mesh paradigm (mapping of sources, reuse of pre-existing platforms, etc.).
- **Governance and standards.** Once the buy-in is acknowledged — essential to enable the widespread adoption of the data mesh inside the organization — standards must be defined in the federated council, potentially reusing the pre-existing governance environment. Such standards will become the reference playbook for the data mesh.
- **Road map and pilot.** After the buy-in is acknowledged, it's very important to set up a flexible road map to guarantee the gradual introduction of the paradigm in the existing planning. This prepares the way for the first pilot project, identifying a single strategic data domain to start the transformation process with and to influence the wider organization.
- **Mesh supervision.** A crucial activity is defining and adopting a mesh supervision plane to enable and measure the success of the initiative. This should enable simple and transparent access to all the data products published in the organization and help

monitor and keep track of data product usage, consumer satisfaction, and the value generated by the mesh through a clear set of metrics and indicators.

- **Data product marketplace features.** At this point, it is essential to define the billing model to regulate the TCO attribution for data product consumption and the economic or reputational incentives needed to remunerate the data product teams that need to be encouraged to increase data asset quality and promote their usage within the organization.
- **Progressive integration of all domains.** In the last phase, data products belonging to the various domains mapped inside the organization are progressively integrated with the pre-existing technologies (e.g., data lakes) by migrating data ownership to the related data product owners, until producer/consumer engagement becomes fully operational on the data mesh. At that point, the operational teams can independently create or consume trusted data products with "embedded" quality, security, and compliance standards.

Once the data mesh is established, the ecosystem of data products starts to effectively become a real **data product marketplace** — this is the biggest change that comes with data mesh. At this point all the benefits are fully exploited at both a technical and economic level:

- Data producer/data consumer engagement is much faster and more seamless, **reducing process lead time** and improving resilience against change.
- Data product **marketplaces enable data discovery** and **greater trust** over data and between producers and potential consumers.
- **TCO attribution is 100% transparent.**
- Federated governance helps to **guarantee and monitor standards, technology change, and IT shadow initiatives.**
- Use cases sharing the same data assets can benefit from accelerated time to value, de-risking the implementation.

This journey summarizes how data-driven transformation is imagined, pursued, and achieved by Agile Lab. This paradigm provides perfect support to a real digital transformation process aimed at creating new revenue streams, new product bundles, and iterating rapidly on the market — the business needs requiring the greatest speed and flexibility, especially for complex multinational organizations integrating many business units.

In future, the ability to manage and enrich data will be a distinctive feature of international competition. According to IDC, by 2023, half of the companies in the Forbes Global 2000 will have operational teams fully dedicated to data monetization, both with innovation goals at the strategic level and with operating targets defined at the product and service development level. Data mesh architectures could play a key role in accelerating data-driven transformation, helping companies to tackle the new challenges of the data economy.

## MESSAGE FROM THE SPONSOR

Agile Lab is an Italian company that is growing its presence in Europe and beyond based on a combination of data engineering and data science expertise, as well as its data engineering experience. Its witboost platform is a useful toolbox to speed up the implementation of data mesh initiatives, a growing trend that is seeing greater customer interest as organizations seek to benefit from the advantages of domain-driven design and decentralized ownership that are typically associated with microservices architectures. The data mesh is seen as a new organizational paradigm that can help enterprises increase analytics agility by overcoming the technical and organizational limitations of monolithic centralized architectures associated with data warehouses and data lakes. Data mesh is also a huge opportunity for organizations to introduce transparency and clear accountability to roles that are core to data management processes. Agile Lab's practical experience can also provide the organizational blueprint for data mesh. [www.agilelab.it](http://www.agilelab.it).

## About the Analyst

### [Giancarlo Vercellino](#), Associate Director, Research and Consulting



Giancarlo Vercellino is associate director at IDC Italy. He is responsible for market research in the software and service area and is involved in consulting engagements with national and multinational clients. Before joining IDC, he was a market analyst, consultant, and business manager at a number of companies and research centers. He has an MSc in IT strategic management from the Polytechnic of Turin and a PhD in management, economics, and industrial engineering from the Polytechnic of Milan. He also took part in the UCLA Global Access Program, attending MBA courses at the Anderson School of Management.

## About IDC

---

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

### **IDC UK**

5th Floor, Ealing Cross,  
85 Uxbridge Road  
London  
W5 5TH, United Kingdom  
44.208.987.7100  
Twitter: @IDC  
idc-community.com  
[www.uk.idc.com](http://www.uk.idc.com)

### **Global Headquarters**

140 Kendrick Street,  
Building B  
Needham,  
MA 02494  
+1.508.872.8200  
[www.idc.com](http://www.idc.com)

## Copyright and Restrictions

---

Any IDC information or reference to IDC that is to be used in advertising, press releases, or promotional materials requires prior written approval from IDC. For permission requests contact the Custom Solutions information line at 508-988-7610 or [permissions@idc.com](mailto:permissions@idc.com). Translation and/or localization of this document require an additional license from IDC. For more information on IDC visit [www.idc.com](http://www.idc.com). For more information on IDC Custom Solutions, visit [http://www.idc.com/prodserve/custom\\_solutions/index.jsp](http://www.idc.com/prodserve/custom_solutions/index.jsp).

Copyright 2021 IDC. Reproduction is forbidden unless authorized. All rights reserved.