

# Next-Generation Data Fabrics Need A Metadata Strategy

July 7, 2021

By Michele Goetz with Gene Leganza, Noel Yuhanna, Boris Evelson, Gabrielle Raymond

FORRESTER®

## Summary

Enterprises want to use all of their data. Bringing it all together and delivering data is the challenge. Extensive moving and copying of data creates discordant context. And the proliferation of data sources for bespoke use cases creates administrative and management chaos for database administrators, stewards, and data engineers. The pendulum has swung toward dynamic data sourcing and delivery, and increased self-service and role-based consumption. Data technology trends have ushered in data as a service (DaaS), but these architectures can only work with a modern strategy and architecture for the metadata. In this report, we explain how to become metadata driven to succeed with DaaS.

# Become Metadata Driven To Scale With Data As A Service

Data fabrics are evolving with cloud-native capabilities to deploy data as a service. DaaS platforms create headless and decoupled architectures for data at scale and enable adaptability that traditional data integration and delivery lack. And as enterprises are buying in, big data integration is the most important component for data and analytics. Emerging DaaS platforms such as GoodData, Riverty, Stardog, and Stibo provide a foundation to ease data integration challenges with no-code/low-code environments. And a DaaS variation, feature stores such as Molecula and Tecton, orients the data for distributed model deployment. Metadata is at the heart of making these systems work: knowing what the data is, where it came from, and how to manage and use it.

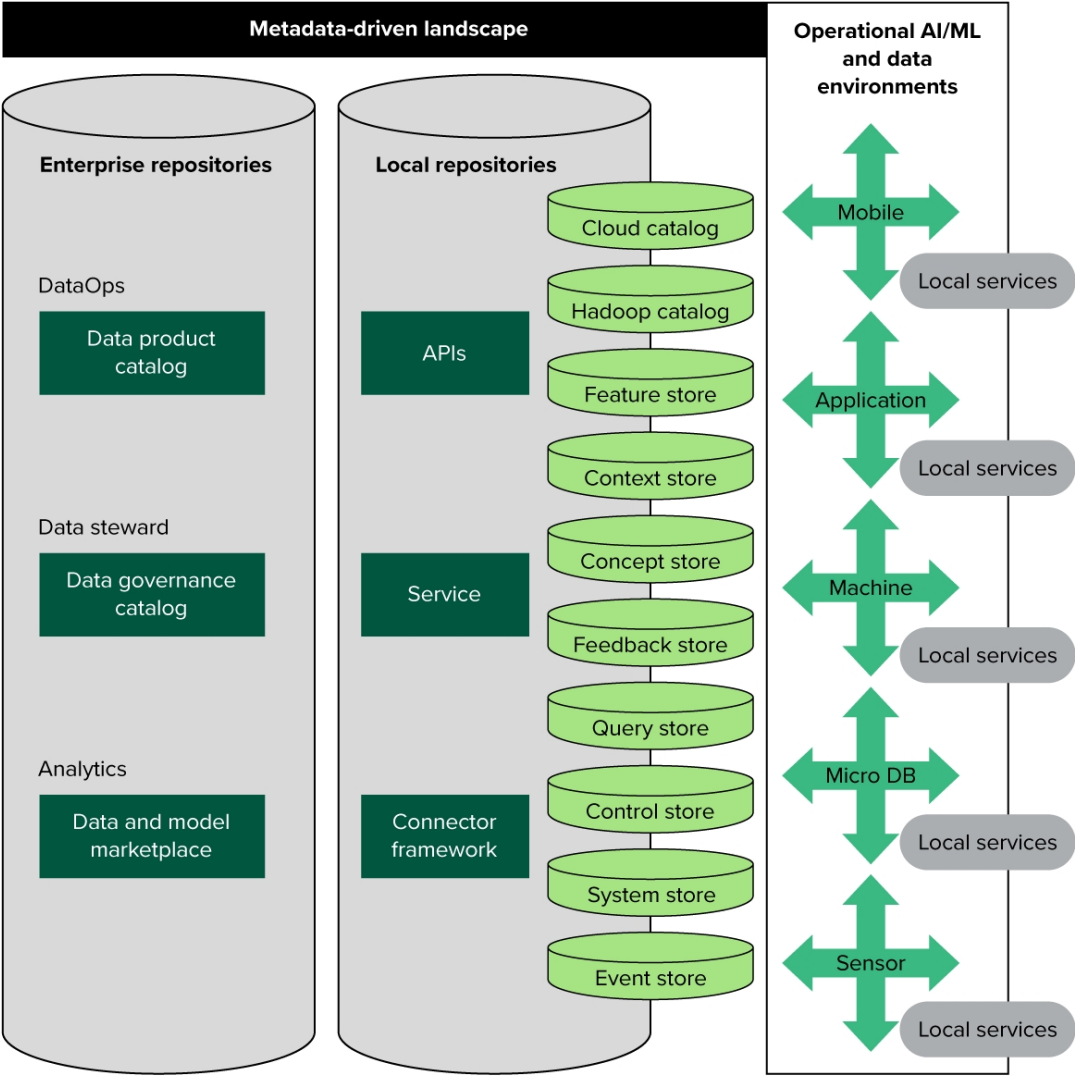
Here are the five things you need to know about metadata to succeed with DaaS:

- **What does it mean to be metadata driven?** The typical “data about data” definition oversimplifies metadata. Today’s data and data science approaches create an extensive semantic, logical, and physical footprint that platforms exploit to automate and scale intelligence. This metadata is the data used by operational systems and products to reach decisions and outcomes. For gas line safety, capturing metadata shapes the historical view of assets, condition, maintenance, and compliance that a national utility company uses to meet safety regulations and contractual agreements with municipalities and customers.
- **What is metadata-driven architecture?** Metadata-driven architecture must enable interoperability, composability, and portability to manage and enable a distributed and decoupled data landscape. Capabilities are deployed to capture, compose, and deliver data about systems, services, queries, controls, events, process, outcomes, and users. A leading global cruise line designs its data architecture to be as functional for data as it is for leveraging the attributes for personalized experiences that allow cruise passengers to choose their excursions while ambiently meeting privacy, customs, and tax rules.
- **What can metadata-driven architecture help me do today?** Metadata-driven architecture brings system-centric and data-centric views together. Data warehouse modernization is easier and faster to deploy and iterate to new use cases and domains because the metadata architecture is flexible. Automated machine learning (AutoML) and data science platforms become more effective. One chief analytics officer found that data cataloging and metadata management shortened the time to

build and deploy models from six to two months. The metadata architecture describes the data, instructs the integration and transformation, and provides the foundation to interpret model performance.

- **How is metadata-driven architecture different than a data catalog?** Today's data catalogs and metadata management tools concentrate on the data's source, lineage, and condition. That is only step one. Metadata-driven architecture creates a knowledge graph of the ecosystem that includes attributes of all data, system, and analytic artifacts. Today, no single platform can spider, scan, and describe a process, event, or experience. Thus, metadata-driven architecture becomes a connected landscape — a metadata model — of data and metadata repositories (see Figure 1).
- **How do I extend metadata-driven architecture to new use cases?** New use cases can sit on top of an existing metadata foundation, or data engineers can use domain-centric data modeling to indicate the categories, data, and systems to add and integrate. In most cases, metadata architecture begins with the analytic use case that transitions insights into operational and automation use cases. Additionally, modern data profiling and lineage tools such as DataKitchen, Monte Carlo Data, and Octopai help organizations understand the complexity, impact, and risks of extending or changing metadata structures and pipelines.

**Figure 1**  
The Metadata Landscape Is The Data Foundation Backbone



Source: Forrester Research, Inc. Unauthorized reproduction, citation, or distribution prohibited.

# Use Three Approaches Simultaneously And Iteratively To Get Started

There is no field of dreams when it comes to data. Today’s methods:

- **Follow new use cases for continuous capture, extract, and publish.** Piggy-back metadata creation and capture with data preparation and data pipeline building that support data products or solution use cases. Grow iteratively as analytics and models are extended or new ones emerge. A global energy company has a strategy to reach into existing techniques and algorithms to extract metadata that it can reuse for other models.
- **Scan existing use cases to iteratively populate catalogs with enterprise knowledge.** Application and business intelligence modernization creates an opportunity for mass scanning and metadata extraction across the enterprise. Cloud data management platforms are rapidly adding robust data catalogs to ensure metadata is captured and managed to keep environments usable, secure, and governed. For example, Google Cloud provides its data catalog as part of its modern data warehouse. And in December 2020 Microsoft released its data catalog, [Azure Purview](#), along with significant improvements to Synapse for end-to-end integration and delivery support.
- **Build the metadata foundation for metadata-driven emerging tech.** Edge computing, AI, blockchain, 5G, and new data technologies such as data mesh all run on metadata at the experience, application, network, and storage layers. That is where development gets started. A global investor firm scans trading systems and extracts metadata to align with FIBO standards for trading using blockchain.

## Supplemental Material

### Companies We Interviewed For This Report

We would like to thank the individuals from the following companies who generously gave their time during the research for this report.

Alation  
Collibra  
Data.World  
GoodData  
Google  
Microsoft  
Molecula  
Promethium  
Rivery  
Stardog  
Tecton

**Not Licensed For Distribution.**

© 2021 Forrester Research, Inc. All trademarks are property of their respective owners.  
For more information, see the [Citation Policy](#), contact [citations@forrester.com](mailto:citations@forrester.com), or call +1 866-367-7378.

ThoughtWorks



# We help business and technology leaders use customer obsession to accelerate growth.

FORRESTER.COM

## Obsessed With Customer Obsession

At Forrester, customer obsession is at the core of everything we do. We're on your side and by your side to help you become more customer obsessed.

### Research

Accelerate your impact on the market with a proven path to growth.

- Customer and market dynamics
- Curated tools and frameworks
- Objective advice
- Hands-on guidance

[Learn more.](#)

### Consulting

Implement modern strategies that align and empower teams.

- Customer and market dynamics
- Curated tools and frameworks
- Objective advice
- Hands-on guidance

[Learn more.](#)

### Events

Develop fresh perspectives, draw inspiration from leaders, and network with peers.

- Thought leadership, frameworks, and models
- One-on-ones with peers and analysts
- In-person and virtual experiences

[Learn more.](#)

FOLLOW FORRESTER



## Contact Us

Contact Forrester at [www.forrester.com/contactus](http://www.forrester.com/contactus). For information on hard-copy or electronic reprints, please contact your Account Team or [reprints@forrester.com](mailto:reprints@forrester.com). We offer quantity discounts and special pricing for academic and nonprofit institutions.

Forrester Research, Inc., 60 Acorn Park Drive, Cambridge, MA 02140 USA  
Tel: +1 617-613-6000 | Fax: +1 617-613-5000 | [forrester.com](http://forrester.com)