ACTIAN™
a division of HCLSoftware

SOLUTION GUIDE

The Essential Role of an Enterprise Data Marketplace in Modern Data Management

The evolution of data catalogs simplifies finding, accessing, and democratizing data.

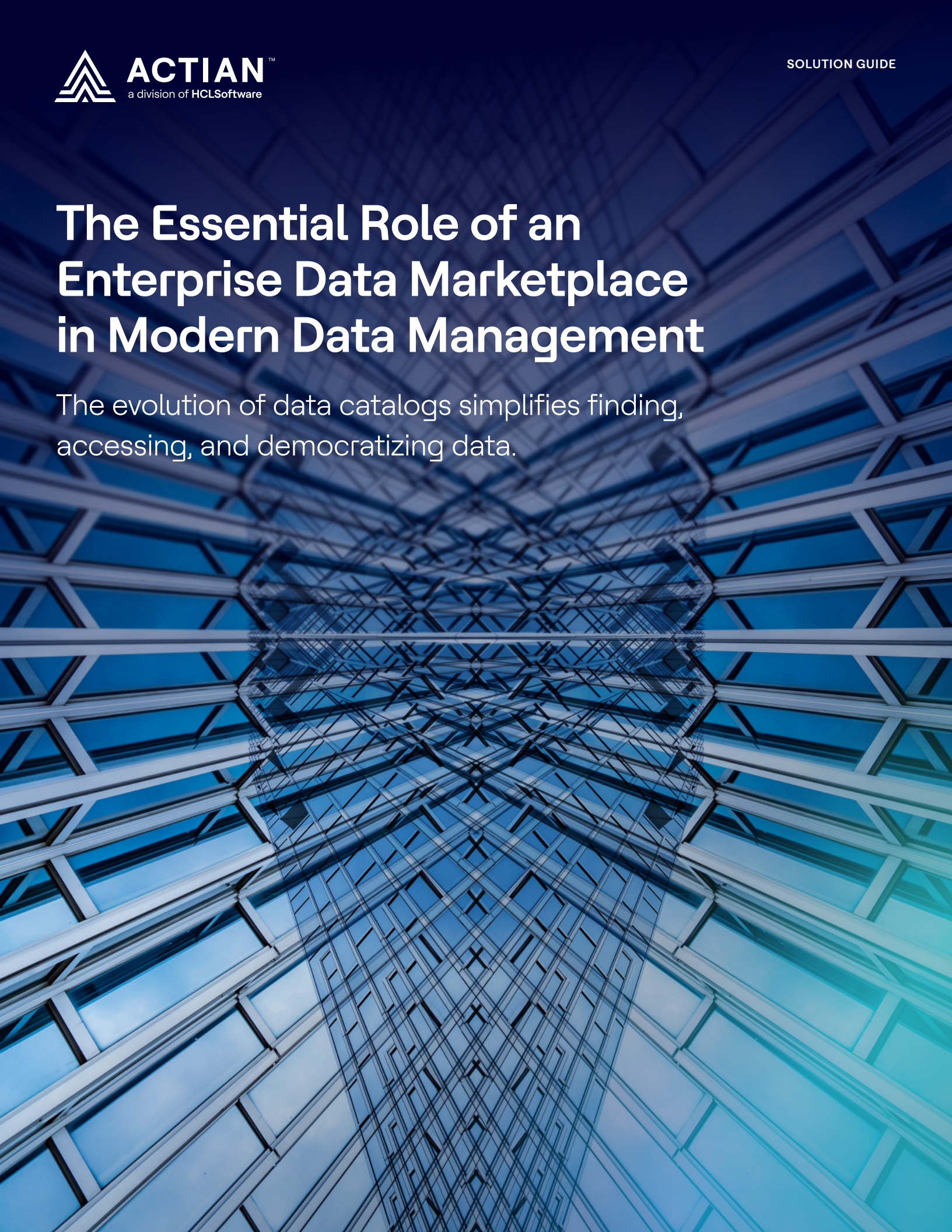


Table of Contents

What is an Enterprise Data Marketplace?

The enterprise data marketplace (EDM) is a critical pillar for democratizing data access in your organization and fully exploiting the potential of your data assets. Imagine a world where every department in your organization can easily access, understand, and leverage data from across the enterprise.

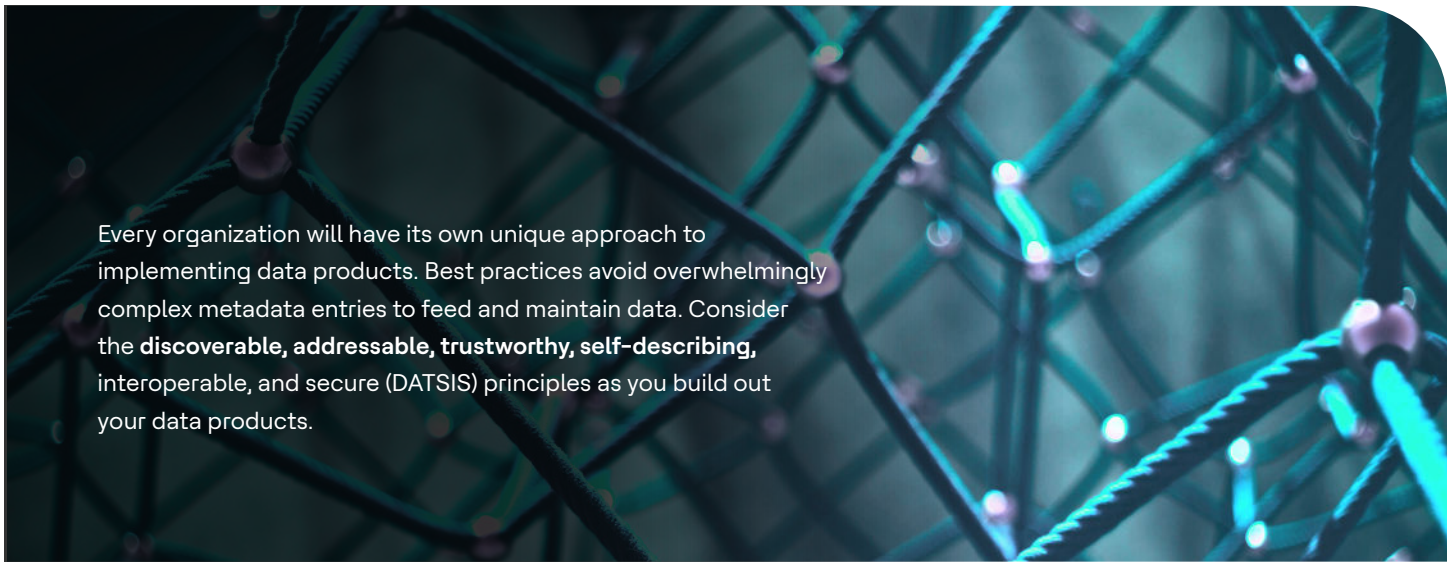
That's the promise of EDM. By breaking down traditional data silos, EDM creates a centralized hub where data becomes a shared resource, fostering collaboration and driving innovation.

> An **Enterprise Data Marketplace (EDM)** is an e-commerce-like solution that enables data producers to publish their data products while empowering data consumers to discover, understand, and consume the published data products.

Paradoxically, an EDM is an internal marketplace within the organization that relies on one or more data catalogs to share and exchange the most valuable data from different domains, which are presented as data products. A data product is defined as a set of high-value data assets specifically designed and managed to be consumed quickly and securely—all while ensuring the highest quality, availability, and compliance with regulations and internal policies.

To clarify, an EDM is not a traditional marketplace where a company makes data available to buyers for a financial transaction and the allocation of usage rights. Nor is it a public marketplace specializing in a domain where clients can buy or rent specialized data.

Instead, an EDM is an internal marketplace where your employees can search, discover, and use data from throughout the organization. The terms of access to this data are defined by your company's data governance standards and, as much as possible, supported by automation mechanisms.



ACTIAN™
a division of HCLSoftware

The Role of the Enterprise Data Marketplace in Decentralizing Metadata Management

An EDM can be seen as the experience plane of the data mesh. The EDM represents a significant evolution in our approach to data. It supports the massive decentralization phenomenon that we see in data management, particularly at the metadata management level, with the growing success of data mesh principles.

Centralized and monolithic data management anchored in a data lake or a data warehouse creates a gigantic bottleneck that inhibits innovation and limits the capacity of data teams to respond to increasingly pressing business demands (Figure 1). Faced with this observation, we are witnessing the progressive adoption of decentralized data management, notably through the data mesh.

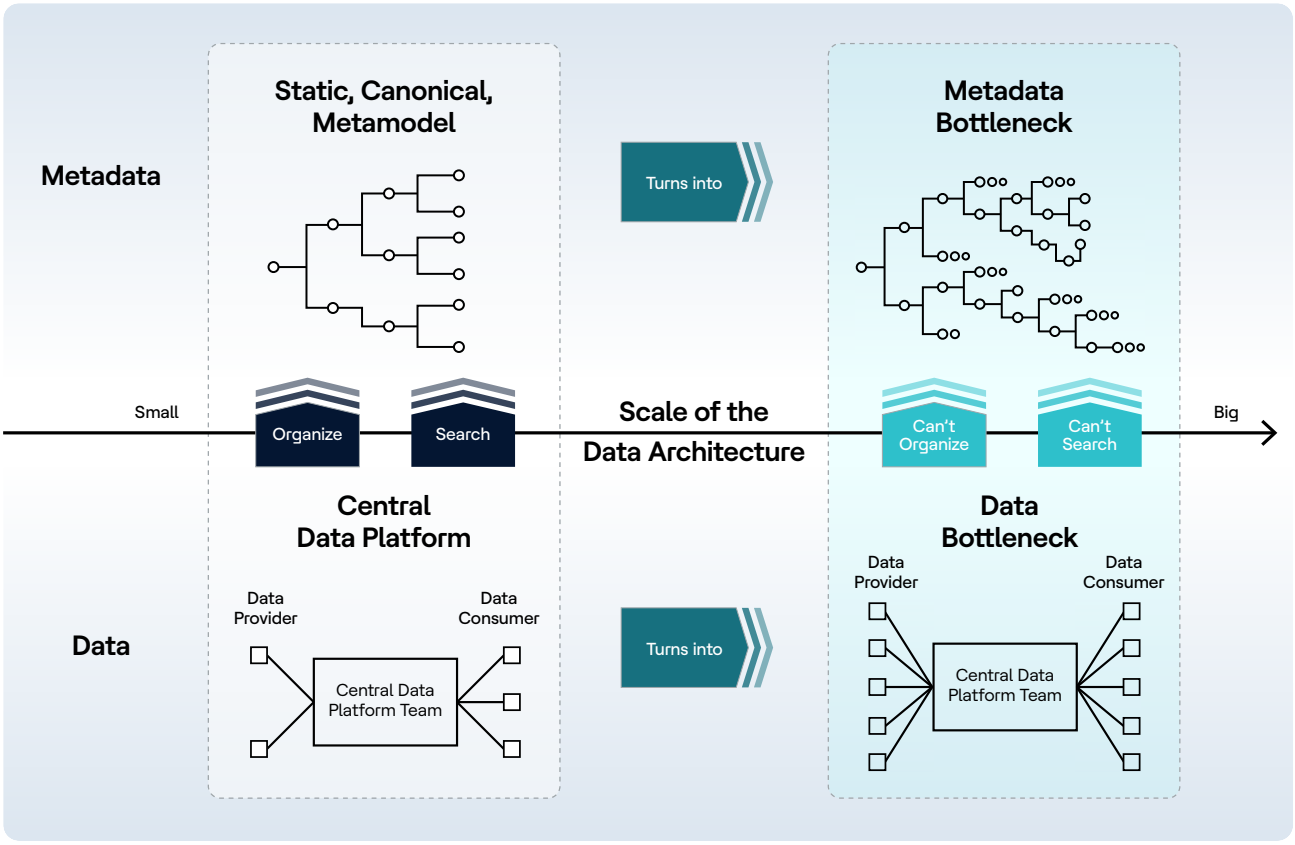


Figure 1: Bottleneck Caused by Centralized Data Management

Introduced by Zhamak Dehghani in 2019 and inspired by Amazon's transformation in the early 2000s, the concept of a data mesh is based on four fundamental principles that impact both data management and metadata management (Figure 2). These principles empower domain experts to create meaningful data products within a decentralized framework.

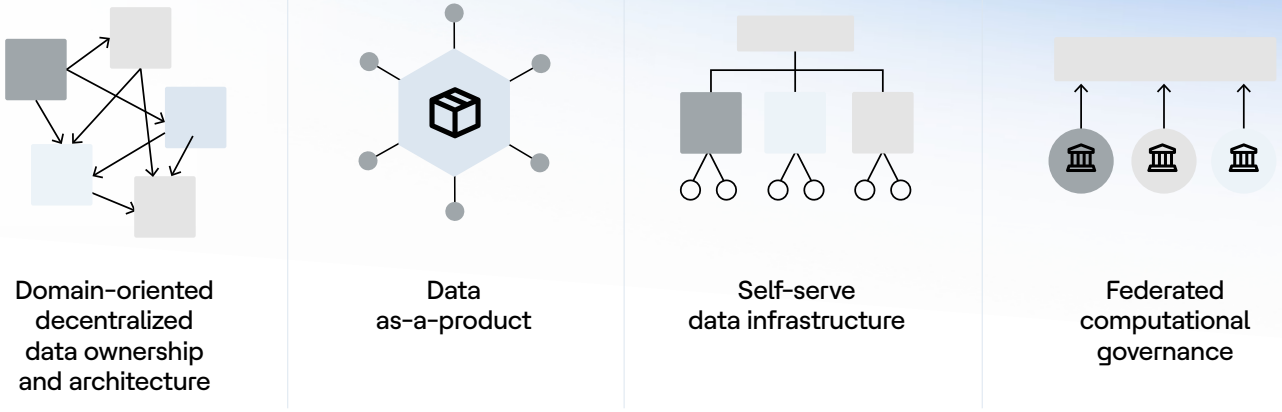


Figure 2: Four Principles of a Data Mesh

Data decentralization also requires metadata decentralization, which empowers domain teams to enrich metadata in context, increasing data accessibility, scalability, and agility. The data mesh is far from being a fad; rather, it represents a profound movement embraced by numerous companies, as supported by studies from BARC

Data Mesh is taking over the world with massive adoption by large-scale organizations

85%

of organizations recognize the relevance of the Data Mesh

-BARC Survey, 2023

54%

of companies are planning to implement or are implementing the Data Mesh

-BARC Survey, 2023

70%

of organizations have or are in the process of implementing Data Products

-BARC Survey, 2024

Although the value of the data mesh is well documented, the literature often describes an idealized final state without detailing the practical steps to achieve it. This gap raises a crucial question—how can organizations effectively transform their data management and metadata management practices to implement a data mesh? The EDM provides an answer to this question.

Key Capabilities and Characteristics of an Enterprise Data Marketplace

The EDM allows various domains of your organization to share high-value data, such as datasets, dashboards, analytical reports, and AI models, in the form of packaged data products. Below are the essential characteristics and capabilities for a functioning EDM.

Powerful Search Capabilities for Business Users

As with leveraging a data catalog, end-users—your organization's data consumers—expect their EDM to be able to discover data products simply and effectively. The internal marketplace must, therefore, rely on a powerful search engine to avoid reproducing the failures of traditional data catalogs in terms of discovery.

However, merely indexing table names, column names, descriptions, and tags is insufficient. This is evident in the nascent market for enterprise data marketplaces, where vendors rely on open-source data catalogs that need to improve in the search and discovery domain.

An EDM that references many data products must have powerful capabilities. In this sense, the knowledge graph technology on which the Actian Data Intelligence Platform data discovery platform is based is a true game changer for our ecosystem. The knowledge graph offers a breakthrough in the user experience and perfectly aligns with the decentralization movement.

A knowledge graph serves as a semantic indexing structure for the search engine and delivers a powerful approach to providing personalized search and discovery experiences. For example, Actian's knowledge graph offers a unique feature—ranking. The search engine explores the graph through lexical and semantic analysis, producing much more relevant search results.

To understand this fundamental technological difference, two objects proximate to a search query won't necessarily hold the same position in the results list. The ranking will favor the item with the highest connectivity to semantically identical or closely related items to the search query.

Finally, because a data product is a packaged set of assets, having a 360-degree view of the objects it encompasses is preferable. Once more, the knowledge graph proves to be a highly effective technology for delivering an exhaustive perspective on the objects comprising a data product.

Flexibility in Data Product Design

Data products represent a new and promising way to represent data. They are a packaged set of data assets accompanied by numerous metadata that are likely to enlighten data consumers, such as version, producing domain, quality, SLA, and output port.

Because each organization defines its unique implementation of data products and the associated metadata, flexibility is of the utmost importance. The Actian EDM, in its design, refrains from imposing a rigid definition of what constitutes a data product. Instead, it enables each organization to define it within its platform with flexibility and scalability. This approach sets Actian apart from many other vendors in the field.

Complete Information Exposure on Data Products

We must learn from past mistakes and ensure a shared and well-understood vision of the data products available within your organization. The level of information naturally depends on your definition and the quality of its maintenance over time.

It is easy to draw an analogy between the physical world and the data world. Indeed, there are many similarities between the data products available in an EDM and real products distributed in a supermarket. Just as products in a supermarket are sourced from various suppliers, data products in an EDM are supplied by business domains. Similar to large-scale distribution, these domains must adhere to specific standards outlined by federated data governance.

Much like adhering to specific regulations to have products displayed on supermarket shelves, a similar process applies in an EDM. Data products must be delivered by domains with detailed information regarding their quality, origin, and characteristics, akin to product packaging, as well as the security protocols governing their use, or access. All of this vital information is recorded within the metadata of the data products.

Actian's EDM offers a range of capabilities that cover all of these needs. Consequently, development teams are now tasked with designing and maintaining new metadata provisioning.

A Necessary Federation of Domain Data Catalogs

At Actian, our vision of the internal data marketplace relies on a fundamental notion—the federation of data catalogs.

In this architecture, each domain maintains its private data catalog, enabling it to organize its data optimally and circumvent the challenges associated with a universal metadata organization. The marketplace serves as a dedicated component, providing simplicity through which each domain can deploy the metadata and data of its data products.

Private catalogs need to be brought together with the EDM to achieve this integration. The integration aims to prevent redundant efforts in producing specific metadata, focusing on lineage, data dictionary (schema), or business definitions likely present in both systems. Additionally, federated catalogs may need to be interconnected to facilitate the sharing and synchronization of certain information, primarily the business glossary, but also certain reference data.

This is the mechanism offered by the Actian EDM (Figure 3). The EDM is an extension of federated data catalogs—a simple and powerful approach in its conception and architecture.

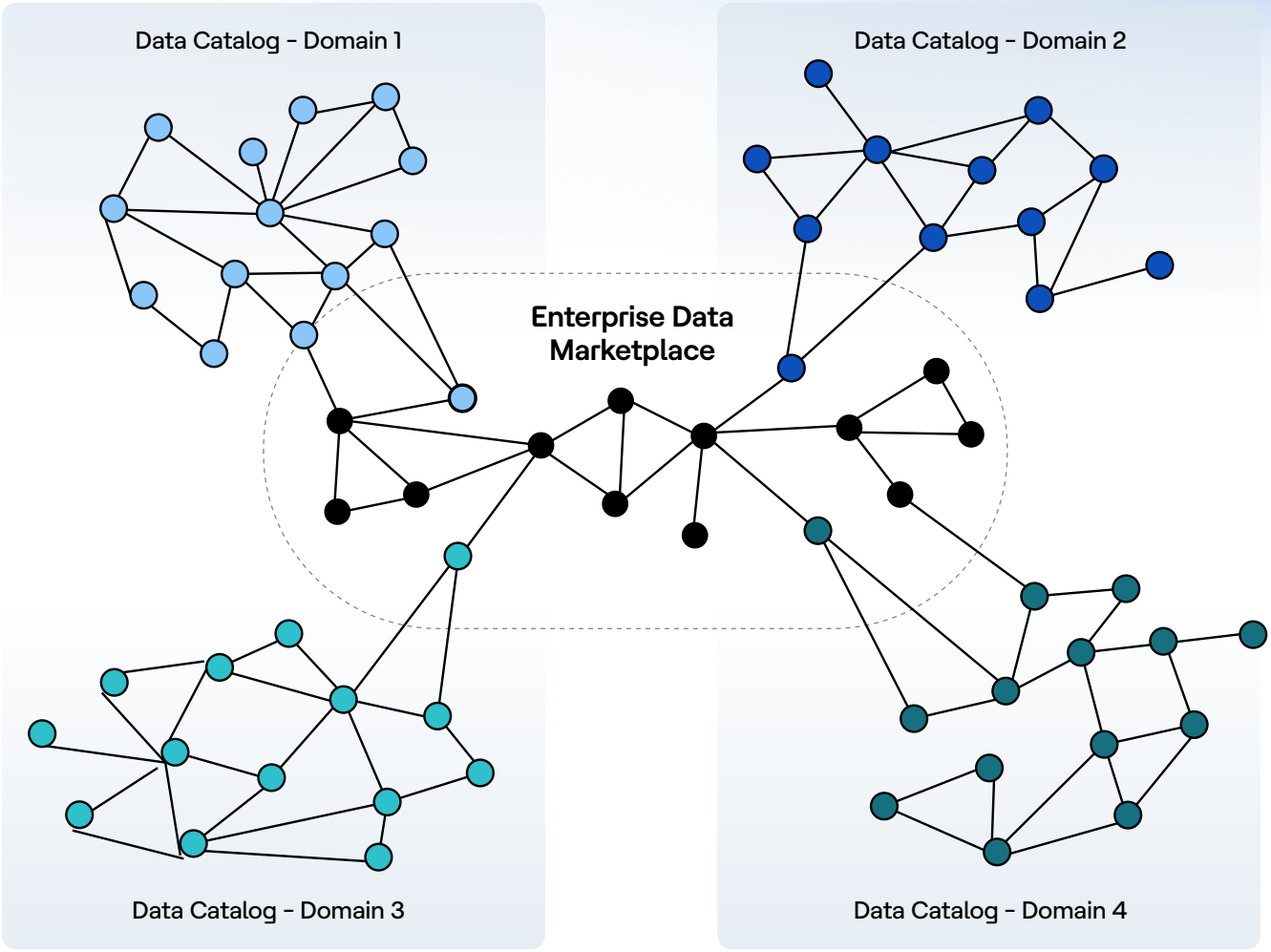


Figure 3: Federation of Domain Data Catalogs in the Actian EDM

The metadata production process undergoes a significant transformation. This aligns with one of the founding principles of the data mesh, which delegates to producers—the domains—the responsibility of supplying reliable and up-to-date metadata on the data products shared in the EDM.

In this context, the responsibility of maintaining the metadata of the shared data products no longer falls solely on the data stewards. Instead, multidisciplinary product teams, particularly the engineering teams of the domains where the data is hosted, are entrusted with designing and populating the metadata. Much like retailers that require their suppliers to adhere to a set of specifications, a significant portion of metadata management is delegated to the data engineering and development teams.

A Modern Approach to Data Governance

The decentralization of data, and consequently metadata, naturally parallels the evolution of data governance. We've observed a shift in data governance practices, which constitutes the fourth foundational principle of the data mesh approach.

Organizations started documenting their available data, but this was limited to artifacts accessible to a select few. With the advent of data catalogs, they attempted an industrial approach to reach more data consumers. However, this approach was still maintained by a central body.

EDMs address needs that are often underserved or not addressed by traditional data catalogs. Federated governance represents a restructuring of responsibilities, drawing upon domain-specific expertise to gain a deeper understanding of their significance while adhering to centralized rules. Thus, delegating certain governance rules allows each domain to be autonomous.

Domain experts can:

- Organize the data by defining a specific metadata model. Some elements of the metadata model may be shared or imposed by governance rules.
- Integrate and ensure the feeding of the catalog from the data sources it owns.
- Manage users and their permissions.
- Identify the objects it wants to share with other domains and control what information will be shared.

Centralized governance, on the other hand, is related to everything that is shared in the large-scale distribution specifications. For instance, similar to retail products, data products must furnish details on the underlying data's quality, origin, characteristics, and consumption rules aligned with company policies.

Actian's EDM adapts to your organization rather than the other way around. It offers great implementation flexibility, whether centralized, semi-centralized, or completely decentralized.

Implementing Mandatory Automation Capabilities

Automation becomes crucial given the diversity and heterogeneity of the metadata intended to be published around an EDM's data products. Without it, the metadata may suffer from poor quality. These definitions highlight the differences between technical metadata and lineage (data provenance) metadata.

Technical Metadata

Technical metadata comes in three varieties, which are those

- Automatically collected and maintained by the underlying data catalog
- Generated by data engineering teams specific to data products
- From other systems that need seamless integration with the EDM, such as data quality management platforms or potential business glossaries

Lineage Metadata

An EDM cannot overlook its responsibility to furnish the origins of data products and must be able to reconstruct lineage. Lineage diagrams should offer insights into the end-to-end transformations that are executed in order to generate consumable data products, encompassing inter-data product lineage data where dependencies exist. This functionality is often lacking or not automatically provided in traditional data marketplaces and catalogs. Conversely, an EDM built on robust data catalogs like the one from Actian effortlessly offers this capability.

Automated Data Access

In the data world, access control is a complex issue. This complexity arises from a fundamental reality—consuming data is risky. However, to create value quickly with data, you must be able to access it quickly.

In the context of decentralization, especially within the data mesh framework, the data product owner, who oversees governance and compliance, should conduct risk assessments and allocate access tokens. This entails approving access requests and specifying any necessary data transformations that are needed to align with specific use cases. The process is known as policy enforcement.

In this context, access request technology must be agnostic. Actian's technology enables the activation of various access methods, ranging from simple email requests or via conversational tools to integrate with platforms already utilized in the workflow domain, such as ServiceNow or Jira, and integrate with data policy enforcement automation platforms, such as Raito.

It is essential to highlight that many data marketplace vendors offer a data infrastructure where companies must store data to gain access. Even if done in a regulated manner, this architecture necessitates transformations for storage, which contradicts the recommended decentralization movement.

A User Experience Worthy of the Name

One of the fundamental mistakes made by all traditional data catalog vendors was assuming that a single platform could serve both data governance teams and end-users.

For instance, Amazon's marketplace provides two distinct user experiences, depending on whether you're a consumer or a seller. Similarly, in a data catalog or a data marketplace, the user experience for those designing, documenting, and exposing data products should naturally differ from those intended to consume the data products.

The Actian Data Intelligence Platform was designed as a marketplace with two specialized interfaces: Studio, the back-office application for managing EDM objects, and Explorer, exclusively dedicated to end-users for search and discovery (Figure 4).

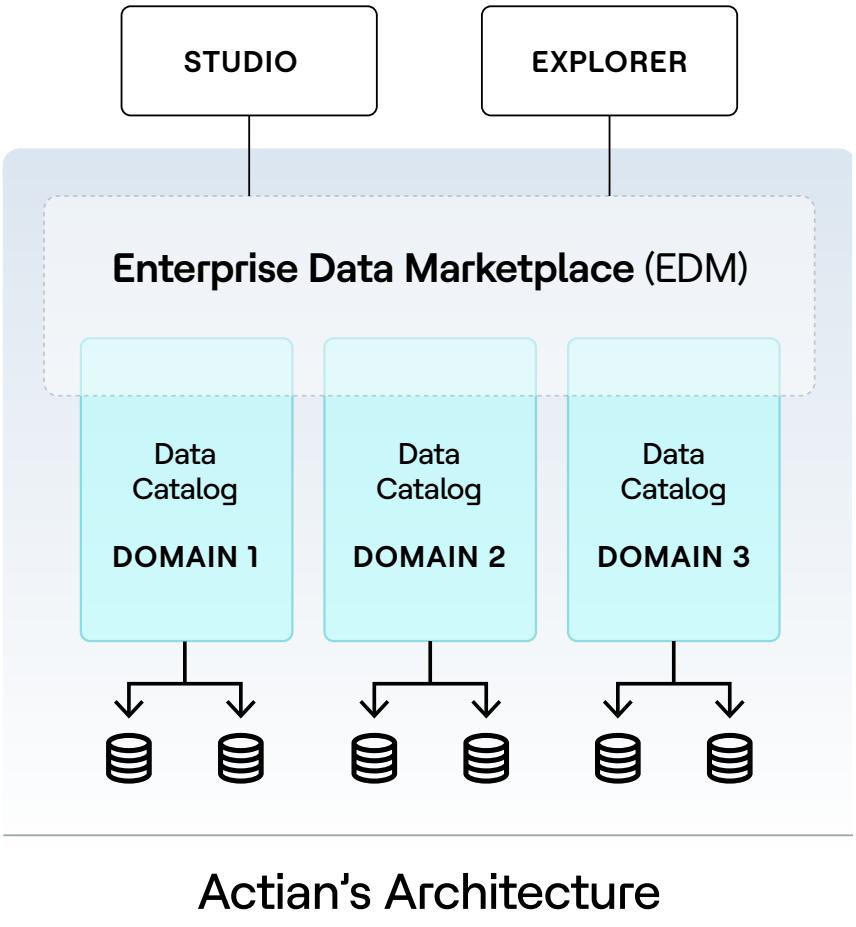


Figure 4: The Actian EDM includes two specialized interfaces: Studio and Explorer

Managing the Particular Lifecycle of Data Products

Because a data product is an evolving artifact intricately linked with a plethora of data, and other data products, the EDM must be capable of displaying its evolutions over time at any given moment. This ensures that a data consumer has visibility into the history of its capabilities.

With the increasing volume of data accessible to various consumer profiles, the quality of the underlying data in data products has become an essential aspect of data governance. Despite efforts by some data catalog vendors to address this capability, it's evident that their solutions often need to be revised and necessitate specialized and dedicated platforms for observability and quality.

An EDM should seamlessly integrate with data quality management (DQM) platforms and accommodate specific developments undertaken by data engineering teams, particularly for organizations that lack dedicated platforms. The value associated with each data product is crucial, but often challenging to quantify.

Therefore, an EDM should offer users a means to express perceived value through ranking. Additionally, data consumers should be empowered to provide suggestions for enhancing data products regarding quality, capacity, and coverage. This feedback mechanism is essential for increasing the popularity and adoption of products by a wider audience, drawing inspiration from product design principles

Leverage a Modern EDM to Securely Democratize Data

The rapid evolution of the data management landscape demands innovative and flexible solutions. Despite their initial promise, traditional data catalogs often fall short of meeting organizations' growing needs in data discovery, governance, and exploitation.

An EDM represents a significant advance by allowing you to share data products efficiently and securely. It combines the benefits of modern data catalogs with additional and essential features such as advanced search and discovery capabilities, flexibility in data product design, and federated metadata and data governance management. By adopting an EDM, you can improve data access and usage to accelerate value creation. The principles of decentralization, like those of the data mesh, are at the heart of this transformation, offering more agile governance adapted to the specific needs of different business domains.

Following best practices and drawing inspiration from product design principles is crucial to successfully implementing an EDM. You can overcome traditional obstacles and fully leverage your data assets by involving end-users, adopting agile practices, and choosing a suitable platform like the Actian Data Intelligence Platform.

Ultimately, becoming a truly data-driven organization involves reevaluating tools and approaches. A well-designed and carefully implemented EDM can transform how data is managed and used, paving the way for continuous innovation and informed decision-making.

About Actian

Actian empowers enterprises to confidently manage and govern data at scale. Organizations trust Actian data management and data intelligence solutions to streamline complex data environments and accelerate the delivery of AI-ready data. Designed to be flexible, Actian solutions integrate seamlessly and perform reliably across on-premises, cloud and hybrid environments. Learn more about Actian, the data division of **HCLSoftware**, at **actian.com**.

ACTIAN™
a division of **HCLSoftware**

Want to see Actian's EDM in action?

Contact us for a personalized demonstration

Get a demo >

Tel +1 512 231 6000 Fax +1.512.231.6010 710 Hesters Crossing Road, Suite 250, Round Rock, TX 78681 **Actian.com**

ACTIAN™
a division of HCLSoftware