



# Data Cataloging for Data Governance: 5 Essential Capabilities

How to drive your enterprise data  
governance program forward



# The Opportunity

**In today's innovation-driven economy, your ability to effectively leverage enterprise data determines success or failure.**

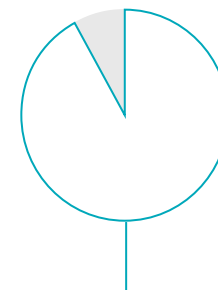
A global auto manufacturer wants to reimagine its customer outreach, so it looks for ways to improve its analytics insights—and realizes it needs a complete and governed view of all its enterprise data. A global financial services firm, seeking to more rapidly launch new services, automates its data governance program to expand employee access to key information.

As these examples show, successful enterprises realize that trusted, governed data represents a key business asset. By finding new ways to leverage data, organizations are able to improve the bottom line: identifying new product opportunities, optimizing pricing, reducing costs and risks, and improving the customer experience—to name just a few business initiatives.

For any business seeking competitive advantage in data, enterprise data governance has become a strategic priority. Data governance ensures that data used to make critical business decisions is trustworthy. It enables you to measure and monitor the quality of your data, so that everything from analytics to customer experience initiatives are more successful. And at the same time, an enterprise data governance program promotes regulatory compliance, helps reduce risk, and protects personal and sensitive data.

A successful data governance program provides holistic, end-to-end governance of critical enterprise data. It ensures that enterprise data is trusted, governed, and protected. And it enables collaboration between your business and IT stakeholders, so that data isn't just IT's responsibility—it's something to be used and governed across your organization.

None of these benefits, however, is possible until you know what data you have, what the quality is, and what systems and processes use it. This is where an enterprise data catalog makes the difference. And not just any data catalog but one that enables agile data governance at scale, with the ability to provide broad visibility across all enterprise-wide data assets, as well as deep understanding of those assets.



**92%** of executives say data governance is critically important to big data business adoption.<sup>1</sup>

<sup>1</sup> New Vantage Partners, Big Data Executive Survey 2017



# You can't govern what you don't know about or can't classify

A successful data governance program needs to include two very critical pieces: discovering and understanding relevant data. Your governance initiatives won't get far if they don't include these two essential processes.

However, the volume and variety of data that modern enterprises must govern pose a challenge. Relevant data may reside on-premises, in the cloud, or in multi-cloud environments. And while data often exists in silos (walled off by boundaries such as business function, geography, or source), organizations must increasingly govern data holistically—whether to seize new opportunities or mitigate risk. For example, data-centric regulations such as the EU General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) require an enterprise-wide view of customer personal data.

The degree to which you are successful at discovery and inventory, therefore, underpins the success of your data governance program. Let's look at these two processes more closely.

## Discovering your data

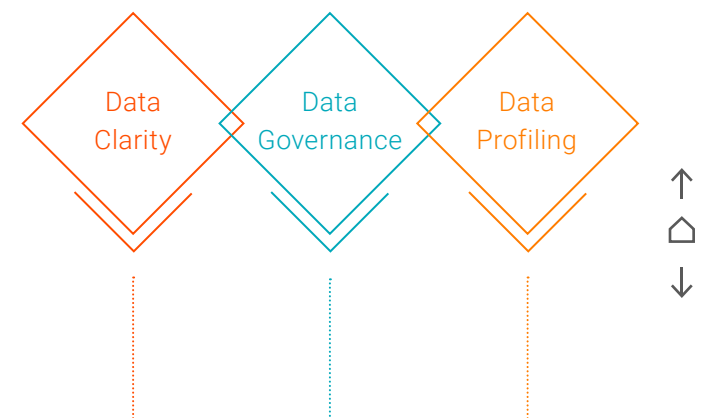
Any data governance initiative needs to answer some simple questions:

- What data do I have?
- Where does it reside?
- What processes does it touch?
- What is its quality?
- Who's responsible for it?
- Where does it come from?
- What data is related to? And so on

A data catalog solution provides answers. It gives you visibility into data wherever it resides, so you can understand what data to bring into your enterprise data governance program.

But given the sheer volume of data that enterprises deal with today, a data catalog must deliver these answers at scale. The only way to do this is by leveraging artificial intelligence (AI) and machine learning (ML). Manual efforts simply aren't adequate to the task.

Top 3 drivers of investment in a data catalog:<sup>2</sup>



<sup>2</sup>"The Intelligent Data Catalog: A Foundation for Analytical Excellence," Aberdeen Group, 2019

Your data catalog should provide visibility into data wherever it resides—in applications, infrastructure systems, on-premises, or in the cloud. It should also put data in context, so users can see where data has come from, who's been using it, and how it's been used in the past. Such context helps users make informed decisions when using data so that ultimately it delivers greater value to the business.

### **Understanding your data**

Data must be defined and classified for it to be searchable and usable across the enterprise, and this step goes beyond the data itself to involve policies, processes, and stakeholder identification. Many organizations start by defining policies and processes, then proceed to discovery.

While AI and ML can be employed to great advantage here, it's important that business and IT stakeholders collaborate at this stage. For instance, in defining data you will want to have a glossary of agreed-upon business terms. This vocabulary provides clear meaning and context

that link to underlying technical metadata. A glossary should answer questions about business terms such as:

- How is this term defined?
- Who is the owner of this term?
- Where does the data related to this term originate?
- If figures are associated with this term, how are they calculated?

The benefits of having a glossary include improving company-wide communication, removing conflict, reducing the time to find data, reducing the number of iterations required for project delivery, and providing a solid foundation for data governance, regulatory compliance, and data stewardship initiatives.

A key capability here involves associating business glossary terms, definitions, and policies to technical data assets. You'll have greatest success if you're able to automatically make these associations at scale leveraging AI and ML.

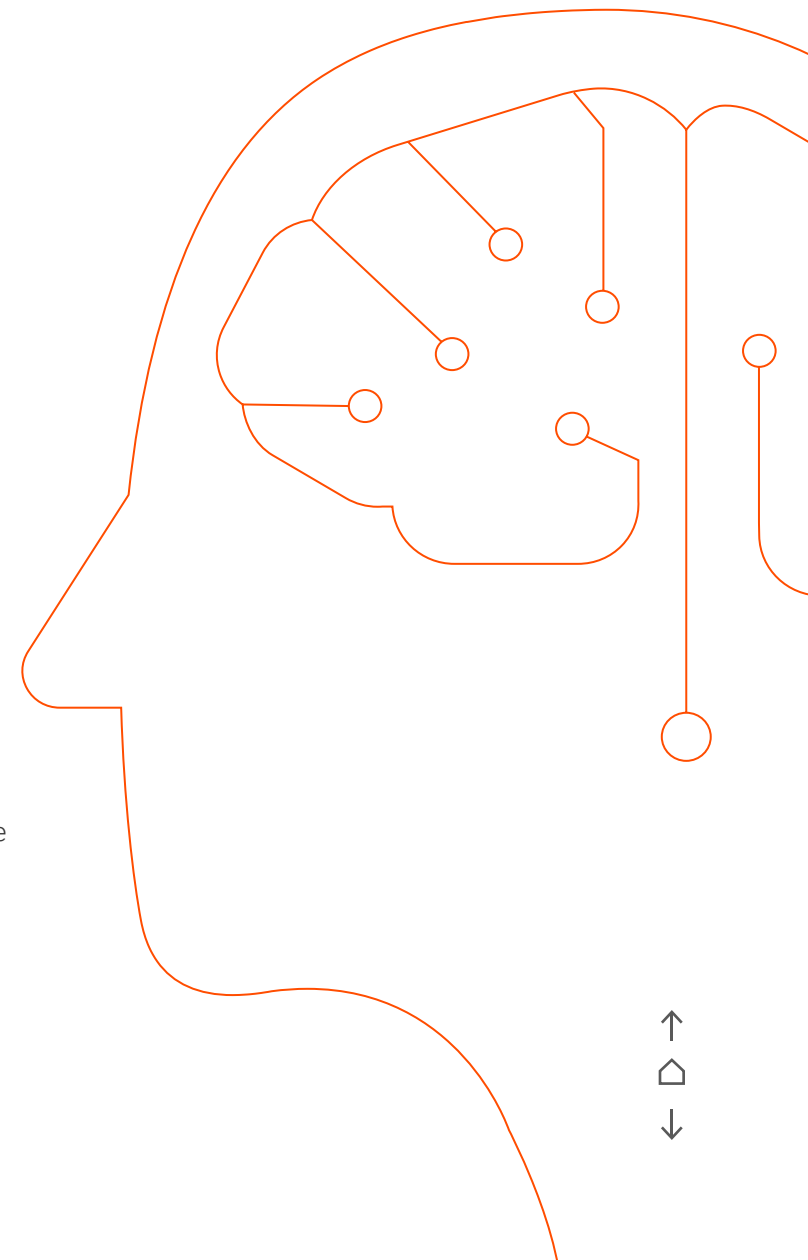
**The Role of AI**

Data governance should be designed to reduce the time-consuming manual processes that make it difficult or impossible to scale with the rise in the volume and variety of data. It should also accommodate the expansion in the number of end-users who expect easy, on-demand access to democratized data. By leveraging AI to automate and streamline processes like data discovery and cataloging, and adding rich business context to the data, data governance frees up talented personnel to focus on deeper analyses and higher-value processes.

An AI-powered data catalog provides a machine-learning-based discovery engine to scan and catalog data assets across the enterprise, whether on-premises, in the cloud, or big data anywhere. Look for AI-powered auto discovery, enrichment, and curation capabilities that:

- Automatically identify and classify entities in structured and unstructured data
- Automatically detect similarity between datasets, and understand relationships and dependencies
- Automatically associate business terms and definitions to physical datasets

While intelligent data catalogs have the ability to automatically scan and catalog assets across the enterprise, it's not enough to just provide visibility into all data. Your catalog should also guide users to the most relevant and trusted data for their business needs. By leveraging machine intelligence, an intelligent data catalog can provide the business context for data assets at scale, driving success for your data governance and compliance initiatives.



## Democratizing data across the business at PayPal.



### **Driving greater business agility at PayPal.**

For global payments platform PayPal, financial data management is a critical business priority. With a growing volume of data to manage, PayPal looked to automate previously manual data-related controls and gain greater efficiencies. With an Informatica data governance solution including an enterprise data catalog, PayPal was able to automate its data governance program across the enterprise, giving non-technical users access to information, rules, and policies about how that information is managed, audited, and maintained.

The new tools capture business definitions and metrics, centralize policies, and allow users to document business workflows all through one interface. End users can finally understand how data is being used—and what best practice looks like. This has enabled PayPal to:

- Democratize data across the business
- Establish repeatable global processes
- Increase time spent creating business value

As a result, PayPal stakeholders have the data they need at their fingertips. And PayPal stays ahead of the competition in launching new payments products.

[READ MORE](#)



# 5 Essential Capabilities

**To support data governance programs that deliver trusted, governed, and protected data, make sure your data catalog solution provides these capabilities.**

An enterprise data catalog supports data governance, but also provides a foundation for any initiative that depends upon trusted, governed, and protected data. Here's how an intelligent data catalog enables success.

## **1. Broad and deep metadata connectivity**

Successful data governance depends upon the ability to discover data by scanning and cataloging data assets across the enterprise. The ability to work seamlessly with data from a broad range of sources – traditional on-premises data, cloud data, big data, business intelligence tools and other apps, structured and unstructured data – is essential. It's not just the breadth of connectivity. It's also the ability to extract different types of metadata that can be leveraged for a deeper understanding of the data and to automate downstream processes.

## **2. Lineage and impact analysis**

To govern data effectively, you need to understand where data is located, who owns it, where it's coming from, and how it's being used. An effective data catalog will be able to provide a business-friendly view of the end-to-end lineage, so business users can understand the provenance of critical data. And by tying certain data (such as, for example, sensitive credit card

numbers) to business terms and policies, it provides guidance for conducting impact analysis related to internal rules and external regulations.

## **3. Data quality monitoring**

At the end of the day, your data needs to be consistently trustworthy. Having visibility into the quality of your data is a must for any successful enterprise data governance program. Your data catalog solution should provide visibility into data profiling statistics (such as value distributions and other patterns) and technical metadata that facilitate a deeper understanding of the data quality. By guiding you to higher quality data, the end results will be better suited for compliance reporting and advanced analytics, as well as for providing meaningful insights to end users across the organization, right up to the CIO and CEO level.



#### 4. AI-powered scalability

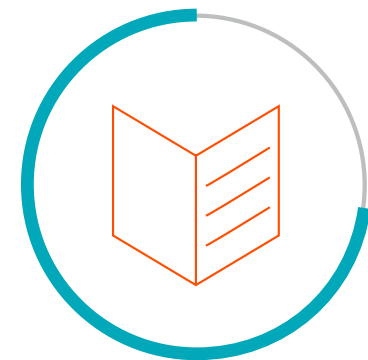
Achieving enterprise scale is impractical without an AI-driven platform that can streamline processes, such as automated curation, tagging and classification, automatic detection of similar data, and automatic association of business terms to technical data assets. With thousands of datasets each containing hundreds of columns and thousands of business terms, manual identification and tagging would be an impractical alternative. If it takes weeks or months to complete a data discovery process or to generate a report, the results may have little or no value.

To streamline the otherwise tedious process of associating technical metadata with business terms and policies, an AI-powered data catalog solution can automatically associate specific business terms and policies with the relevant datasets, a necessity for achieving and maintaining enterprise scale. Data stewards can focus on higher-value analyses of AI-curated datasets to deliver far more value.

#### 5. Collaboration

In addition to harnessing the power of AI for scalability, a data catalog should harness the collective knowledge and subject matter expertise of all stakeholders in contact with the data. This includes IT architects, data owners, data stewards, and data consumers.

Collaboration across all of these stakeholders is critical to a successful data governance program. By providing a holistic view of your data, along with rich business context, the data catalog should enable IT to better support business requests and requirements. Business users should be able to tag, comment, rate, approve workflows, and so on. With tight integration between policies and the data under governance controls, an intelligent data catalog enables both business and IT stakeholders to use data effectively and responsibly.



Of companies using a data catalog, **73%** are more likely to be satisfied with their ability to share data compared to non-catalog users.<sup>3</sup>

<sup>3</sup>"The Intelligent Data Catalog: A Foundation for Analytical Excellence," Aberdeen Group, 2019



## Cross-organizational alignment at Bank of Ireland, courtesy of data governance.



### Data Governance and Bank of Ireland

The Bank of Ireland set out to reengineer its business in line with its digital ambitions and shifting customer focus—all while staying compliant with industry regulations, such as the GDPR. It needed clear rules around ownership of data, and a more in-depth understanding of how the business operated. Different teams needed answers to different questions, underscoring the need for a data catalog: What data do I have? How do I identify critical data? Is my data secure? Which process does my data affect?

To answer these questions, the bank created a framework to address these issues pragmatically. The framework covers an end-to-end lifecycle, including: identifying critical data, defining data ownership on an individual basis, monitoring data handling to ensure consistent data quality, optimizing processes and identifying data-related problems within them, and fixing those problems with a unified, enterprise-wide strategy.

Learn more about enterprise data governance in this solution brief.

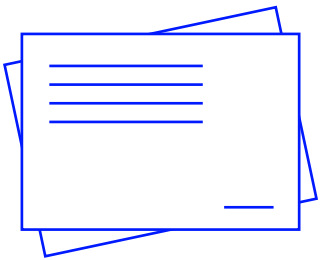
[READ MORE](#)



# Further Reading

## Informatica Enterprise Data Catalog

Read the Enterprise Data Catalog data sheet for a deeper look at the capabilities and benefits of the Informatica data catalog solution.



[READ MORE](#)

# About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100 percent focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

## Worldwide Headquarters

2100 Seaport Blvd, Redwood City, CA 94063, USA

Phone: 650.385.5000

Fax: 650.385.5500

Toll-free in the US: 1.800.653.3871

[informatica.com](http://informatica.com)

[linkedin.com/company/informatica](https://www.linkedin.com/company/informatica)

[twitter.com/Informatica](https://twitter.com/Informatica)

[CONTACT US](#)

IN18-0319-3590

© Copyright Informatica LLC 2019. Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and other countries.



Informatica®