

# AI-Powered Data Cataloging and Governance for Evidence-Based Policymaking Act

## Key Benefits

- Intelligently catalog and govern all your data and metadata at scale across on-premises and cloud environments
- Speed data discovery and empower non-technical and IT users as well as data custodians with powerful, Google-like semantic search
- Automatically discover, classify, and recognize patterns in your data at scale with advanced machine learning algorithms
- Dynamically visualize and trace data with in-depth data lineage and impact analysis
- Accelerate insights and transparency with comprehensive data governance
- Improve auditability and regulatory compliance with user-based access control

On January 14, 2019, the U.S. Foundations for Evidence-Based Policymaking Act went into effect, delivering the most comprehensive data reform strategy that the public sector has seen in decades. At its core, the act aims to empower federal agencies, government officials, and constituents with better access to high-quality data driving government policies, spending, and decisions.

The act requires agencies to develop and implement a comprehensive data cataloging and governance strategy that accounts for all data assets and metadata created by, collected by, under the control of, and maintained by each agency. The act also stipulates enforcement of stringent guidelines for protecting and securing sensitive data while making non-classified information readily available to the public.

The benefits of the act may be far reaching once fully implemented. According to the McKinsey Global Institute, better management and unlocking of government data could potentially result in over \$3 trillion in economic benefits, enabling agencies to improve efficiencies, reduce waste and fraud, prevent data leakage, better serve constituents, and turn data into their most valuable asset.<sup>1</sup>

## Overcoming Data Challenges in Government

The data landscape in the public sector is highly complex, with petabytes of structured and unstructured data and metadata that is growing at an unprecedented rate residing across hundreds of disparate applications, data lakes, data warehouses, and siloed legacy systems across on-premises and multi-cloud environments. Furthermore, agencies are increasingly challenged with the lack of end-to-end visibility or understanding of their data as well as a host of data accessibility, integrity, and quality issues.

<sup>1</sup> Open data: Unlocking innovation and performance with liquid information, Mckinsey Global Institute, October 2013

Given the scope of the act, implementing its stipulations can fast become a daunting undertaking as it involves:

- Discovering what data an agency has and where it resides
- Validating each dataset across multiple systems at scale
- Determining whether the data comes from a trusted source and whether it's classified or available for public use
- Creating and maintaining a trusted, up-to-date inventory of prioritized data and metadata at all times for intra-agency and public use
- Enforcing a holistic data governance and privacy program with comprehensive audit trails

### **Accelerate Adherence to the Act With Informatica AI-Powered Enterprise Data Catalog**

Creating an intelligent inventory of an agency's data and implementing a single source of reference at scale is the first step when operationalizing the provisions of the act. Informatica® Enterprise Data Catalog delivers advanced features and capabilities designed for rapid discovery, curation, and sharing of data and metadata at scale, allowing agencies to obtain end-to-end visibility into their data regardless of where it resides. It is a foundational pillar for enforcing a holistic data governance strategy as well as enabling a multitude of use cases that support government programs, regulatory compliance, and analytics and artificial intelligence (AI)-driven initiatives.

## **Key Features**

### **Rapid Data Discovery With Powerful Google-Like Semantic Search**

Enterprise Data Catalog enables rapid discovery of data with powerful Google-like semantic search empowering non-technical and IT users—including chief data officers, data scientists, data stewards, data engineers, and various custodians of data—with the ability to easily discover and profile data, identify its location, and obtain other key attributes about the datasets at scale. Semantic search is also applied to inferred data domains including synonyms and concept matching, so that no data asset is left undiscovered across on-premises and cloud environments.

### **Automatically Discover, Classify, and Recognize Patterns in Data With Sophisticated Machine Learning Algorithms**

Enterprise Data Catalog provides automated capabilities using advanced statistical and metadata-driven machine learning algorithms designed to tackle computational complexity inherent in big data when discovering, tagging, clustering, and identifying similarities and patterns in data, enabling you to intelligently catalog all types of data at scale.

For instance, automatically classifying and identifying domains and entities such as Social Security number (SSN), date of birth (DOB), gender, credit card, country, city, and so on across all structured and unstructured data at the field, column, and table level. This enables better search and filtering of search results while providing automated glossary recommendations. Agencies can also add their own custom domains. This crucial capability will allow you to rapidly catalog and govern your data and metadata while extracting value from your data assets.

## Dynamically Visualize and Obtain a Holistic Relationship View of Your Data With In-Depth Data Lineage and Impact Analysis

You can interactively trace data origin through business-friendly, summarized lineage views across data sources regardless of where the data resides—from on-premises to cloud environments. A drill-down lineage view expands any lineage path to show columns and lineage diagram metrics. Non-technical and IT users, as well as various custodians of data, can perform detailed impact analysis on upstream and downstream data assets to gain in-depth insights that further support adherence to regulatory stipulations and provisions of the act.

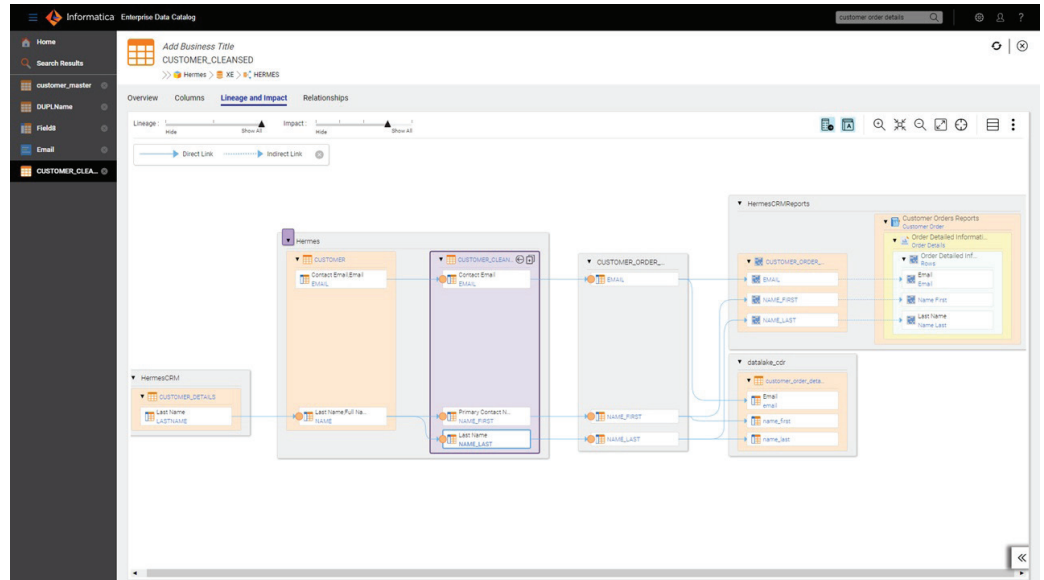


Figure 1: Dynamically visualize your data with in-depth data lineage and impact analysis.

## Speed Data Governance With Out-of-the-Box Data Domains and Business Glossary Terms With Rich Business Context

With Enterprise Data Catalog, you can enable automatic association of business terms with the physical datasets and technical metadata, which helps eliminate a tedious manual process for data governance. This allows non-technical and IT users to collaboratively manage an agency's metadata. It also supports import of business glossary assets from third-party tools.

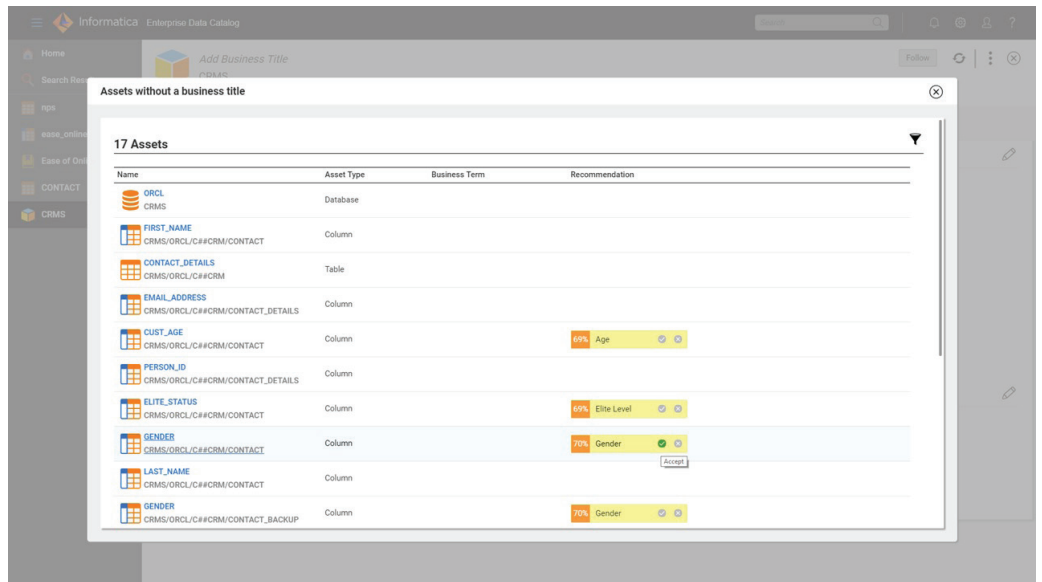


Figure 2: Speed data governance with automatic association of business terms.

### Share Data and Enable Social Curation With Intelligent Crowdsourcing and Annotations That Facilitate Intra-Agency Collaboration

Enterprise Data Catalog empowers non-technical and IT users as well as custodians of data to easily find the most relevant and trusted data for analysis by harnessing the combined power of sophisticated machine learning algorithms, human expertise, and collaboration. Data owners and subject matter experts can certify datasets and provide ratings and reviews, enabling social curation of data. Users can follow datasets of interest and get notified of changes. With this capability, agencies can increase understanding of various datasets supporting government programs. Moreover, a Q&A platform enables subject matter experts to answer common questions from users.

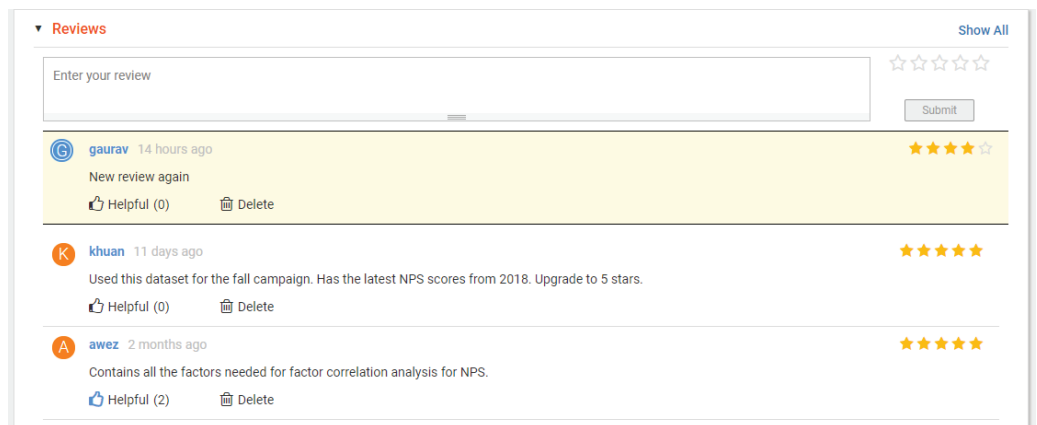


Figure 3: Intelligent crowdsourcing and annotations facilitate intra-agency collaboration.

### **Scan Metadata from Multitudes of Data Sources Across On-Premises and Cloud Environments**

Enterprise Data Catalog offers an expansive set of APIs and scanners that are regularly updated and designed to integrate in your environment. They consume data catalog content and metadata from a host of enterprise applications, Hadoop and Spark clusters, databases, and data warehouses, among others, enabling you to build a comprehensive data catalog that can scale across your enterprise supporting a multitude of use cases.

### **Analyze Data With Confidence Using Your Business Intelligence (BI) Tool of Choice**

Users can easily access the full resources of Enterprise Data Catalog when creating and delivering data visualizations using their BI tool of choice. For instance, an extension from Tableau enables users to quickly obtain a better understanding of the data under analysis, including the ability to easily identify data sources and determine whether the data assets are authenticated and prioritized.

### **Obtain Rich Data Quality Statistics**

Users can view data profiling statistics alongside technical metadata to understand the quality of data assets before using data for analysis. Profiling statistics include value distributions, patterns, data types, and data domain inferences.

### **Apply Comprehensive Resource-Level Security to Support Regulatory Stipulations**

Grant user and group read/write permissions at the resource level to allow users to view or edit custom attributes or perform domain curation that support auditing and regulatory compliance requirements.

### **Enforce a Holistic Data Governance and Privacy Program With Seamless Integration**

Enterprise Data Catalog is designed to seamlessly integrate with Informatica Axon Data Governance, Informatica Big Data Quality, and Informatica Data Masking software along with third-party tools that are in your environment enabling you to streamline the process of configuring, deploying, and maintaining a holistic data governance and data privacy program.

For instance, with the Axon Data Governance and Big Data Quality solution, custodians of data can quickly identify data quality issues and visualize them in real time. They can monitor, manage, and cleanse data across on-premises and cloud environments with granular, role-based access control. To protect classified information, Informatica Dynamic Data Masking is used to ensure that sensitive data is dynamically discovered and protected from unauthorized access. It applies data privacy and security actions in real time to mask, hide, and alert to unauthorized access. With this capability, agencies can enforce stringent data privacy and protection policies that support the act while ensuring transparency and auditability.

## 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% 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.

## Next Steps

To learn more about the Informatica AI-Powered Enterprise Data Catalog and Data Governance solution for the public sector, please visit:

[Informatica Solutions for the Public Sector](#)

[Informatica Enterprise Data Catalog](#)



**Worldwide Headquarters** 2100 Seaport Blvd., Redwood City, CA 94063, USA Phone: 650.385.5000, Toll-free in the US: 1.800.653.3871

IN17\_1120\_03699

© Copyright Informatica LLC 2019. Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and other countries. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners. The information in this documentation is subject to change without notice and provided "AS IS" without warranty of any kind, express or implied.