

# Informatica Enterprise Data Catalog

## Benefits

- Automatically catalog and classify all types of data across the enterprise using an AI-powered catalog
- Identify domains and entities with intelligent curation
- Enrich data assets with governed and crowdsourced annotations, ratings and reviews
- Find data assets through powerful Google-like semantic search
- Discover and understand your data assets with a holistic view including lineage, relationship views, and data profiling and quality stats
- Get a complete picture of your data environment
- Open APIs to integrate into your environment and expose intelligent metadata anywhere
- Automatically associate business glossary terms to technical data assets

## Unleash the Power of Data With an Intelligent Data Catalog

Data is the lifeblood of our economy, and data-driven companies turn their data assets into revenue and profits. The first step in any data-driven digital transformation initiative is to manage your data as an enterprise asset: take inventory of it, assess its value, and maximize its use—just like you do with other significant capital and operational investments.

Data is diverse and distributed across many different departments, applications, and data warehouses (some on-premises, others in the cloud), making it a challenge to know exactly what data you have and where. In the world of big data this becomes even more complex.

Informatica® Enterprise Data Catalog is an AI-powered data catalog that provides a machine-learning-based discovery engine to scan and catalog data assets across the enterprise—across cloud and on-premises, and big data anywhere. The intelligence in Enterprise Data Catalog is provided by the CLAIRE™ engine, which provides intelligence by leveraging metadata to deliver intelligent recommendations, suggestions, and automation of data management tasks. This enables IT users to be more productive and business users to be full partners in the management and use of data.

Informatica Enterprise Data Catalog provides data analysts and IT users with powerful semantic search and dynamic facets to filter search results, data lineage, profiling statistics, holistic relationship views, data similarity recommendations, and an integrated business glossary. Collaboration capabilities leverage subject matter expertise and social curation combined with the power of AI to guide user experience and automate data curation. You can now easily and efficiently manage enterprise data assets to maximize their value throughout the company. Users can quickly find data and easily manage the life cycle of business terms, definitions, reference data, and more.

Informatica Enterprise Data Catalog is an AI-powered data catalog that provides a machine-learning-based discovery engine to scan and catalog data assets across the enterprise—across cloud and on-premises, and big data anywhere.

## Key Features

### **Metadata APIs to Integrate Into Your Environment**

Enterprise Data Catalog includes REST-based APIs that enable you to integrate it into your environment and consume catalog content anywhere. Organizations can share any intelligent metadata—applications, BI reports, and dashboards—with business users.

### **EDC Tableau Extension for Governed Self-Service Analytics**

The Tableau extension for Informatica Enterprise Data Catalog delivers agile, self-service analytics with governed data. It enables Tableau users to access the full resources of Enterprise Data Catalog from within the native Tableau user interface. Business users who are consumers of Tableau reports can leverage an intelligent search bar within Tableau to find data assets certified by data stewards, as well as access business and technical context, data quality, profiling statistics, and peer recommendations. Users can also collaborate with their peers on the data assets via reviews, ratings, and Q&A, all without leaving the Tableau interface.

### **Semantic Search With Intelligent Facets**

Find and discover the most relevant datasets for your analysis using powerful semantic search with intelligent facets. Advanced keyword search with token matching finds the most relevant data assets in the catalog. Semantic search is even applied to inferred data domains so no data asset is left undiscovered. Intelligent facets, based on the search results, allow users to alter the search to the datasets of interest.

### **Data Lineage and Impact Analysis**

Interactively trace data origin through business-friendly summarized lineage views that highlight the end points and not all the complex details in between. A drill-down lineage view expands any lineage path to show columns and lineage diagram metrics. Users can perform detailed impact analysis on upstream and downstream data assets.

### **Holistic Relationship Discovery**

Get a holistic view of data in a knowledge graph that lets you quickly search, discover, and understand enterprise data and meaningful data relationships. Automatically discover related datasets, technical, business, semantic and usage-based relationships. The holistic data view shows related datasets, tables, views, data domains, reports, and users. This aids in progressive discovery of other datasets of interest.

### **Automated Classifications With Intelligent Domain and Entity Recognition**

Automatically classify and identify domains and entities such as customer, product, order etc. across all structured and unstructured data assets at the field, column, and table level. This is a crucial step in the ability for companies to catalog, govern, and extract value from their data assets. This classified data enables better search, filtering of search results, and business glossary recommendations. Informatica provides over 60 packaged data domains such as email, credit card number, social security number, country, city, URL, and company name. Users can add their own custom domains too. Data assets can be classified using data rules (i.e., columns with data that matches specific logic defined in the rule) or column name rules (i.e., finds columns that match column name logic defined in the rule).

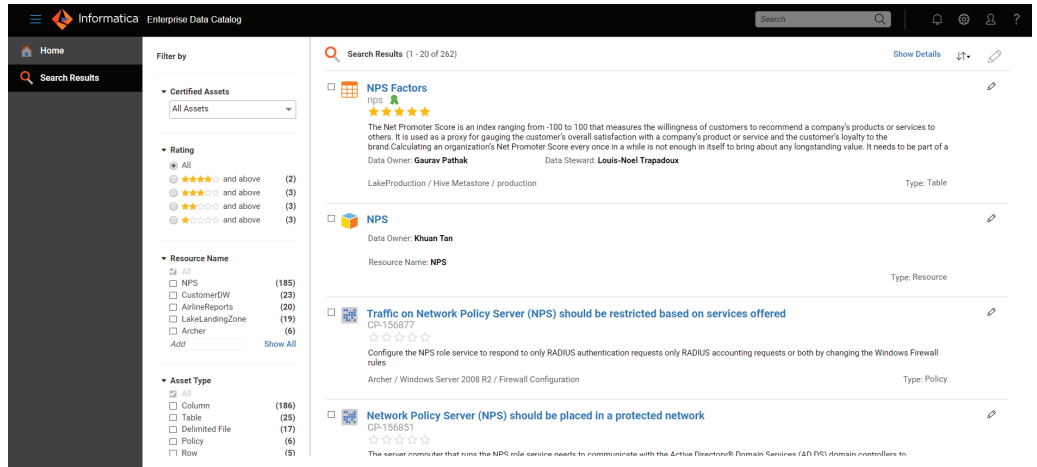


Figure 1: Quickly find datasets with smart semantic search and dynamic facets. View ratings and certified datasets.

## Collaboration and Social Curation

Informatica Enterprise Data Catalog empowers data analysts and data scientists to easily find the most relevant and trusted data for analytics by harnessing the combined power of AI and human expertise and collaboration. Data owners and subject matter experts can certify datasets. Data consumers can provide ratings and reviews for datasets enabling social curation of data. Users can follow datasets of interest and get notified of changes, and a Q&A platform allows subject matter experts to answer common questions from users.

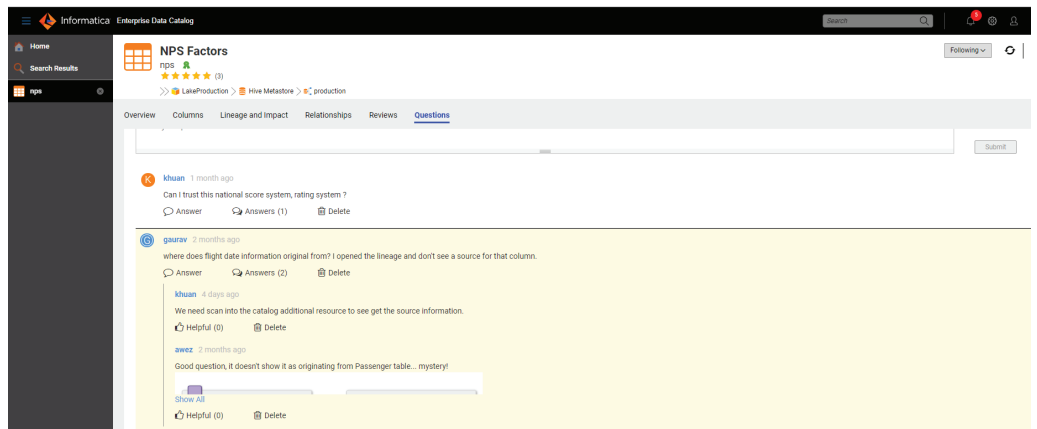


Figure 2: Enable collaboration with Q&A capabilities.

## Integrated Data Quality Statistics

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, and data type and data domain inference.

### **Automatic Association of Business Glossary Terms**

Informatica Enterprise Data Catalog allows for easy import of business glossary assets such as terms, policies, and classifications from Informatica Axon™. Add rich business context to the data by associating business terms with the right technical metadata. Informatica Enterprise Data Catalog enables automatic association of business terms with the technical metadata, which helps eliminate a tedious manual process for data governance. This allows business and IT stewards to collaboratively manage business metadata that includes efficient human workflow automation. Informatica Enterprise Data Catalog also supports import of business glossary assets from Informatica Business Glossary and third-party tools.

### **Intelligent Data Similarity**

Advanced statistical and machine learning algorithms identify similar data and subsets of data. This powerful capability helps users find the most relevant and trusted data they need. For example, a telecom analyst interested in customer churn analysis might query data containing pre-paid customer activity for the current quarter. Informatica Enterprise Data Catalog can recommend a cleaner version of the data (substitute data), data containing customer activity for the previous quarter (union-able data), and a customer detail table to enrich the dataset (joinable data).

### **Universal Metadata Connectivity**

Extract metadata from any type of data sources across the enterprise such as databases, data warehouses, applications, cloud data stores, BI tools, Hadoop, NoSQL, and more. Below are some examples of data sources supported for metadata extraction:

- Databases: Oracle, MS SQL Server, Sybase ASE, IBM DB2, IBM Netezza, Teradata, JDBC, Amazon Redshift, Azure SQL DB, Azure SQL DW, Google BigQuery
- Hadoop: Cloudera Navigator, Hive (Cloudera/Hortonworks/MapR/IBM BigInsights/EMR), HDFS, Hortonworks Atlas
- Mainframes: DB2 z/OS, DB2 i5/OS
- BI: SAP BusinessObjects, Tableau, Cognos, MicroStrategy, OBIEE, QlikView
- ETL: Informatica PowerCenter®, Informatica Big Data Management®, Informatica Cloud®, Informatica Data Integration Hub, Microsoft SSIS, Oracle Data Integrator
- Business Glossary: Informatica Axon, Informatica Business Glossary
- Data Modeling: Erwin
- Applications: Salesforce, SAP, Oracle, Workday, Informatica MDM
- File Systems: HDFS, Amazon S3, Azure Blob, Azure Data Lake Store Gen 2, Microsoft SharePoint, Microsoft OneDrive, Windows/Linux Filesystems
- File Formats: MS Excel, MS Word, MS PowerPoint, Adobe PDF, Flat Files, CSV, Delimited, XML, JSON, Avro, Parquet

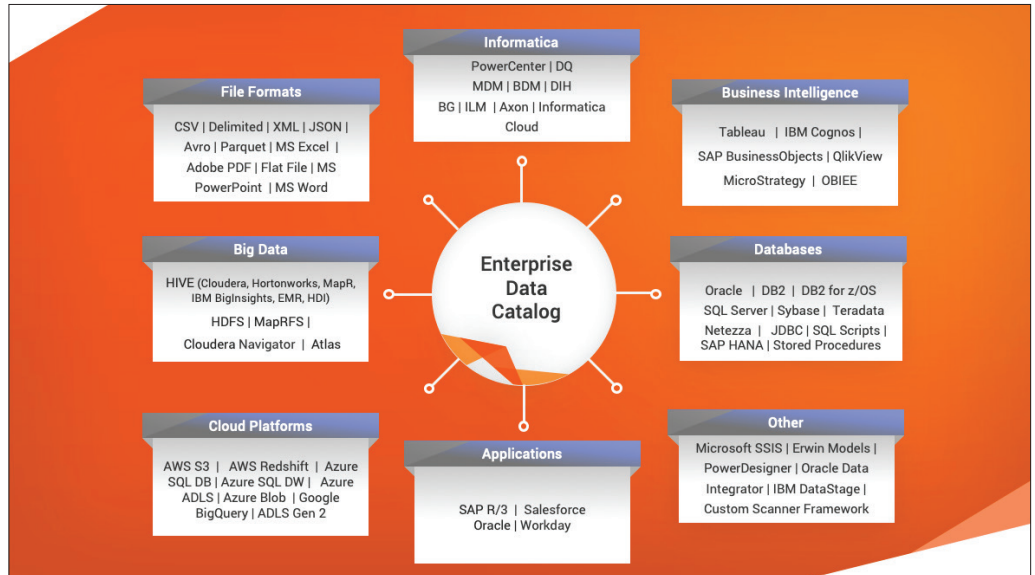


Figure 3: Informatica Enterprise Data Catalog supports universal metadata connectivity.

### Custom Attributes With Business Classifications

Enrich datasets by crowdsourced or expert classifications, comments, and other attributes available to anyone with appropriate security permissions. Assigning custom attributes and annotations to datasets including business glossary terms enhances business-IT collaboration and search results.

### Resource-Level Security

Grant user and group read/write permissions at the resource level to allow users to view or edit custom attributes, perform domain curation, and associate business glossary terms.

### Big Data Scale Deployments

Enterprise Data Catalog is built for big data scale deployments that can be deployed on Hadoop clusters. It supports parallel metadata ingestion and high-speed distributed indexing to quickly update catalog content and deliver unmatched search performance and provides fault tolerant high availability for 24x7 implementations.

### Unified Administration

Manage and monitor the catalog resources, metadata extract schedules, profiling runs, and more from one unified admin console. A job control dashboard provides widgets for task monitoring and resource views. Email alerts assist administrators in proactively responding to catalog issues.



Figure 4: Understand your data with holistic data relationship views.

## Benefits

### Intelligently Catalog All Types of Data Across the Enterprise

Informatica Enterprise Data Catalog intelligently discovers many types of data and their relationships across the enterprise. Prebuilt scanners collect metadata from databases, data warehouses, applications, cloud data stores, BI tools, Hadoop, NoSQL, and more. All the metadata is indexed and cataloged in a highly scalable graph database architected for fast updates, smart search, and fast queries. As more and more data is created and propagated throughout the enterprise, similar and duplicate datasets inevitably arise. Informatica Enterprise Data Catalog leverages advanced statistical and machine learning algorithms to discover similar data and subsets of data, helping users find the most relevant and trusted data they need.

### Find Data Assets Quickly Through Powerful, Google-Like Semantic Search

Trying to find the data you need across hundreds of enterprise systems may sometimes seem futile. Only through powerful semantic search built on comprehensive metadata services and a scalable infrastructure can one even hope to find relevant data. Informatica Enterprise Data Catalog delivers semantic search with intelligent facets to further refine search results. Because Informatica uniquely associates business, technical, and operational metadata, business users can search with business terms to find their data and then browse holistic relationship views to find related data assets.

## 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.

## Discover and Understand Your Data Assets With Holistic Relationship Views and Lineage

The classic saying, “You can't manage what you can't measure” is true when it comes to managing data assets. To get the most value from data, you need to understand what you have, where it came from, how it has changed, and what level of trust you have in the data. Informatica Enterprise Data Catalog answers all these questions and more with complete end-to-end summary and detail lineage, profiling statistics, and holistic relationship views, providing a clear picture of your data.

## Enrich Data Assets With Business Context Through Governed and Crowdsourced Annotations

Informatica Enterprise Data Catalog maximizes the reuse and value of data by automatically classifying enterprise data assets down to the field/column level. To further increase the value of data, Informatica Enterprise Data Catalog captures the context of who is using the data and for what purpose, along with crowdsourced tags, annotations, ratings, and reviews. This “wisdom of crowds” helps to enrich and curate data, making it even more valuable throughout the enterprise. Informatica Enterprise Data Catalog integrates with Informatica Axon for easy import of business glossary assets such as business terms, definitions, and policies from Axon. This business metadata is automatically associated with technical metadata and operational metadata so that business analysts, data stewards, and other users can quickly find, understand, and collaborate on data assets.

## Learn More

To learn more about Informatica Enterprise Data Catalog, please visit <https://www.informatica.com/products/big-data/enterprise-data-catalog.html>.



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

IN06\_0619\_03238

© Copyright Informatica LLC 2019. Informatica, the Informatica logo, CLAIRE, Axon, Informatica Cloud, Big Data Management, and PowerCenter 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.