



Cloud Data Governance and Catalog

Trusted Data: The Key to Digital Transformation

Digital transformation initiatives are driving business value for organizations, propelling them forward. Unfortunately, despite sustained investments, some organizations have struggled to reap the benefits of this transformation due to the inadequate focus on one of their most valuable assets: data. Successful digital transformation heavily depends on providing trusted data for data consumers; however, complex data landscapes and fragmentation have made this task increasingly challenging.

To effectively leverage enterprise data for various use cases, including enhancing customer experience, enabling innovation, and ensuring greater compliance with regulatory authorities, data consumers must be able to trust and have visibility into the data. Comprehensive **data intelligence** is essential for organizations aspiring to accelerate their digital transformation journeys.

Cloud Data Governance and Catalog: Predictive Data Intelligence for Data and Analytics Governance

Informatica® Cloud Data Governance and Catalog, a service of Informatica Intelligent Data Management Cloud™ (IDMC), combines data governance, **data catalog** and data quality capabilities into a singular tool for automating data intelligence insights. This IDMC service is built for organizations that want to maximize their investments by deriving value from their vast data assets.

Cloud Data Governance and Catalog delivers **predictive data intelligence** powered by the Informatica CLAIRE® AI and ML engine.

Key Benefits

- Accelerate data-driven business outcomes by connecting data consumers with the data they need to build a foundation of trusted data
- Enhance data-driven decision-making by improving data literacy
- Advance usability of data assets with business context via automation, bulk curation and crowdsourcing
- Discover sensitive data and automatically assign recommended data policies to classifications to help assess and mitigate exposure risks
- Govern AI models and their underlying data to enable trusted use
- Accelerate the development of a data governance framework to deliver trustworthy data to data consumers



Organizations that want to drive business value from trusted data will appreciate its automated and recommendation-driven data classification, bulk data curation, relationship discovery and sensitive data discovery. Just as importantly, they can provide data consumers with the business context they need. The IDMC service enables efficient self-service analytics and data governance by unifying the capabilities of data discovery, data lineage, data profiling, data quality, business glossary creation, stakeholder and policy management, and the ability to document and govern AI models and their implementations.

Cloud Data Governance and Catalog integrates into your existing data landscape and scans hybrid sources, including cloud data lakes and warehouses, analytics/BI systems, databases, ETL tools, and other enterprise systems. The IDMC service is cloud-native, meaning you can deploy it into your existing infrastructure almost immediately and at the scale needed.

Key Capabilities

Broad and Deep Metadata Connectivity

Cloud Data Governance and Catalog offers broad and deep metadata connectivity that spans multi-cloud and on-premises environments. Applying wide and deep data source connectivity, it allows you to extract metadata across:

- Cloud platforms
- BI tools
- Databases
- Multi-vendor ETL
- Data science tools
- Various enterprise applications
- File formats
- SQL dialects
- Stored procedures

With universal metadata connectivity to nearly all your data sources and a runtime option to run serverless or within your on-premises or virtual private cloud, the IDMC service provides a centralized, comprehensive view of your data.



Inspect scripts, procedures and processes to fully understand logic and internal data flow. Obtain complete column-level data lineage, including an inventory of potential lineage sources with rich details. Scan static and dynamic code and perform language parsing for automated data lineage across the enterprise.

With the Cloud Data Governance and Catalog custom metadata framework, you can use simple Excel files to ingest custom metadata and derive data lineage and relationship links from critical systems where automated scanners are unavailable. Model virtually any data source or data lineage across systems.

Data sources supported include:

- **Informatica:** PowerCenter, Data Integration, Multidomain Master Data Management and Business 360 applications
- **Cloud Platforms:** Amazon Web Services (AWS) S3, AWS Redshift, AWS RDS (Oracle, MS SQL Server, PostgreSQL, MySQL), DynamoDB, Azure SQL DB, Azure Synapse, Azure ADLS Gen 2, Azure Blob, Google Cloud Storage, Google BigQuery, Snowflake, Databricks Delta Tables, Oracle Cloud Storage, Oracle ADB
- **On-Premises:** Oracle, IBM Db2, Netezza, SQL Server, Teradata, JDBC, MySQL, SAP HANA DB, Postgres, MongoDB, Local/Shared Filesystem
- **Database Scripts:** DB2 LUW SQL, Microsoft SQL Server SQL, Oracle SQL, Snowflake SQL, Teradata BTEQ
- **BI and Analytics Platforms:** Tableau, Microsoft Power BI, QlikView, Qlik Sense, Microsoft SSRS, Cognos, Google Looker
- **Other ETL and Data Science Platforms:** Azure Data Factory, Databricks Notebooks, Databricks Unity Catalog, Microsoft SSIS, Microsoft SSAS, Talend
- **Enterprise Applications:** Salesforce, Kafka, Workday, Marketo, SAP BW, SAP BW4/Hana, SAP ECC, SAP S/4Hana, SAP Business Objects, Dynamics CRM, Microsoft OneDrive, Microsoft SharePoint
- **File Formats:** CSV, Delimited, JSON, Avro, Parquet, SFTP, XMI

Contact Informatica for the most current list of supported data sources.

AI-Powered CLAIRE Engine to Drive Insights from Metadata

Automation is critical to manage and govern large data estates. Users will appreciate how Cloud Data Governance and Catalog uses intelligent data element and entity classification to help automate **metadata management** and extraction from heterogeneous sources. You can also automate **data profiling** and classification across data assets at the field, column and table levels.

The solution offers an automated approach to data discovery and classification to reduce the time and effort spent on tedious manual processes that do not scale. Data stewards can review and curate (accept/reject) from more than 215 out-of-the-box automated data classification associations recommended by CLAIRE. Users can modify and extend these classifications or add new ones as needed.

Cloud Data Governance and Catalog also learns from associations and can auto-tag similar fields and columns across the enterprise using rule-based and AI-based methodologies. The IDMC service also can automatically associate glossary terms to data and infer relationships, such as joins among datasets using AI/ML capabilities, including schema matching.

With the CLAIRE activity page, users can view analytics related to automated glossary and classification associations, including metrics on accepted, pending and declined associations. The page provides a central location to identify and act on pending curation actions. These insights help drive the usage of automated associations powered by CLAIRE and can be utilized to calculate the time saved for curation activities across the organization.

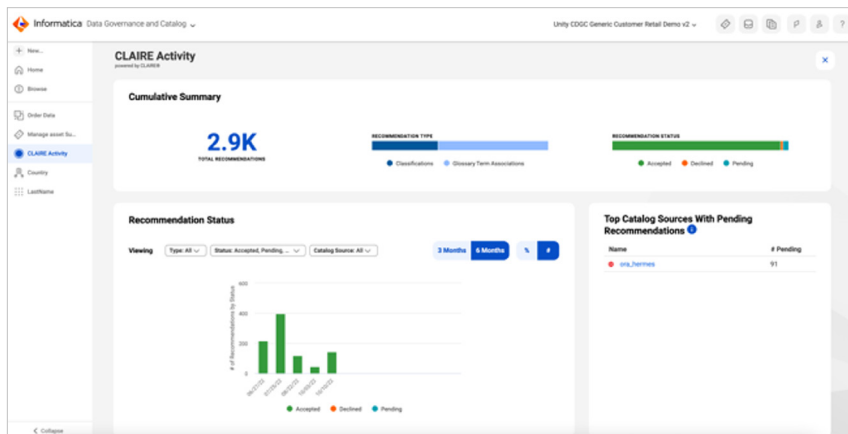


Figure 1. CLAIRE activity analytics provide a summary of intelligent glossary and classification associations across the organization.

Powerful, Intuitive Search and Browsing Capabilities

Users can perform natural language-like searches to locate critical assets across business and technical domains. They can also use filtering and preview capabilities to quickly review and identify desired assets. All personas can more easily explore data assets using browsable hierarchical views for context, relating technical data sources to business-curated datasets to provide a seamless experience.

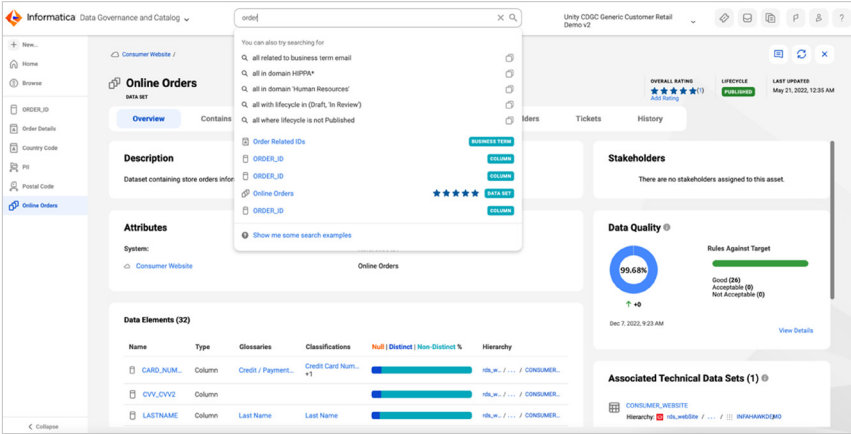


Figure 2. Quickly find data assets using powerful semantic search capabilities.

Holistic Data Relationship Views

Get a 360-degree view of data in a knowledge graph that lets you quickly search, discover and understand enterprise data impact and meaningful data relationships. Automatically discover related datasets and technical, business, semantic and usage-based relationships. The holistic data views display various asset types with direct and indirect connections to each other, providing a comprehensive view of an asset's touchpoints across other data assets. This aids in the progressive discovery of additional data assets of interest.

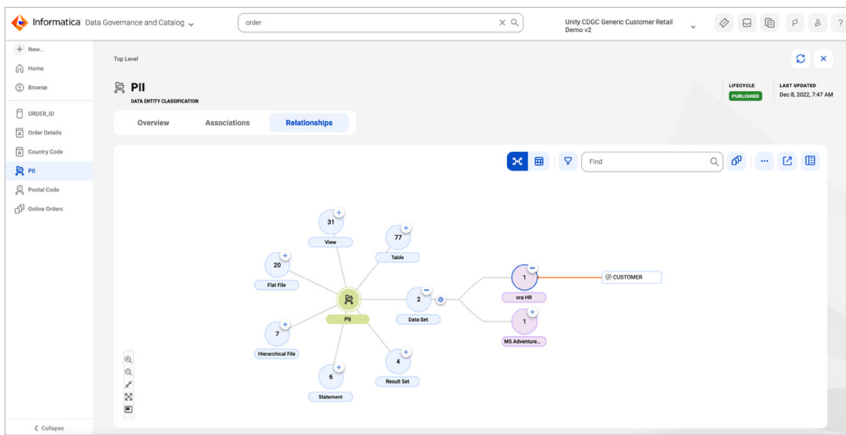


Figure 3. Interactive graphical views of data relationships help users discover and understand assets across the organization.

Data Lineage and Impact Analysis

Interactively trace data origin with data lineage views at any level, from business-friendly, system-level summarized views highlighting endpoints to granular, column-level technical views that include intricate details automatically derived from parsing SQL scripts and stored procedures. Users also can perform detailed impact analysis on upstream and downstream data assets. Conveniently visualize essential details associated with your data, such as business glossary terms, domains, policies and processes directly within data lineage views. Data quality overlays allow you to monitor quality scores and how they change throughout the data flow across your data estate.

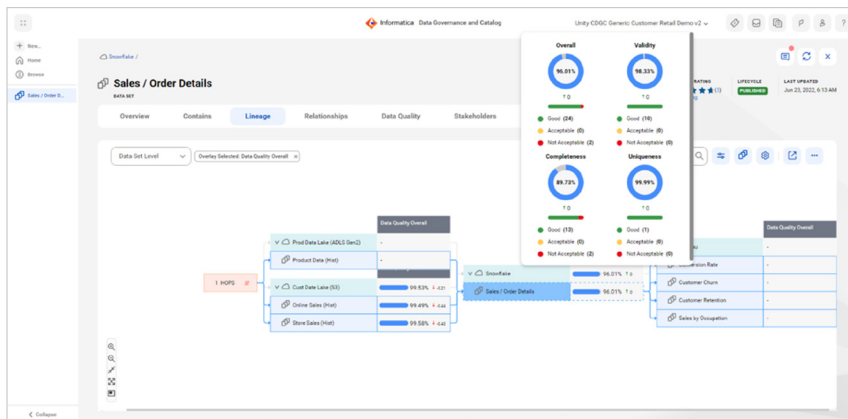


Figure 4. Explore data-element-level lineage from source to target and gain an additional understanding of your data by displaying graphical overlays, including data quality scores.

Integrated Data Quality

View data profiling statistics, rules, scorecards, and metric groups alongside technical metadata to understand the data quality of assets, an integral part of any data governance program. Profiling statistics, including value distributions, patterns, data type and data domain inferences, helps automate data quality measurement, significantly reducing the burden on stakeholders. Users are also offered data quality previews to help them assess the accuracy and usability of their data. Additionally, Cloud Data Governance and Catalog can automatically notify stakeholders of data quality status changes via the user interface or email, allowing them to act swiftly and insightfully.

Collaboration and Social Curation

Cloud Data Governance and Catalog empowers data analysts and data scientists to easily find the most relevant and trusted data for analytics by utilizing AI, human expertise and collaboration. Data owners and subject matter experts can certify datasets, and data consumers can provide ratings and reviews for datasets, enabling the social curation of data.

A Q&A platform allows subject matter experts to answer common questions from users. In addition, users can add custom attributes and annotations to datasets, further enhancing business-IT collaboration and search results to harness tribal knowledge and improve literacy.

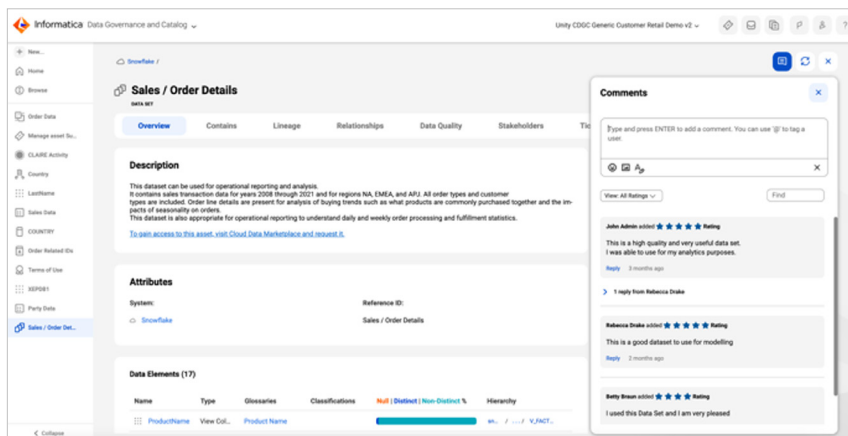


Figure 5. Collaboration capabilities include the ability to drive discussions and share comments and user ratings for data assets. Users can also certify technical assets.

Automated Workflows

Use workflows to automate processes and notifications for reviewing and approving new and modifying existing assets. The automated workflows within Cloud Data Governance and Catalog help ensure that stakeholders create and modify assets in compliance with data governance principles within your organization. The IDMC service has predefined workflows for common processes to help simplify and accelerate workflow creation and implementation.

AI Model Governance

AI model governance strives to push towards explainable AI by providing organizational visibility and transparency into models and their underlying algorithms, which is often a black box for most organizations. It details how the model was developed, the training data used for creating the models, its quality and lineage, and relevant policies. It helps track and monitor model performance and key metrics, like data drift, that may lead to model performance degradation.

Goal-Oriented Dashboards

The interactive and graphical dashboards put the user in command, providing summarized information in a visual form, including stakeholder/owner assignments and glossary metrics. Users can also monitor automated predefined workflows, check task completions and view notifications. With a variety of visualizations and drill-down capabilities, users can quickly view the summary status and explore details as needed.

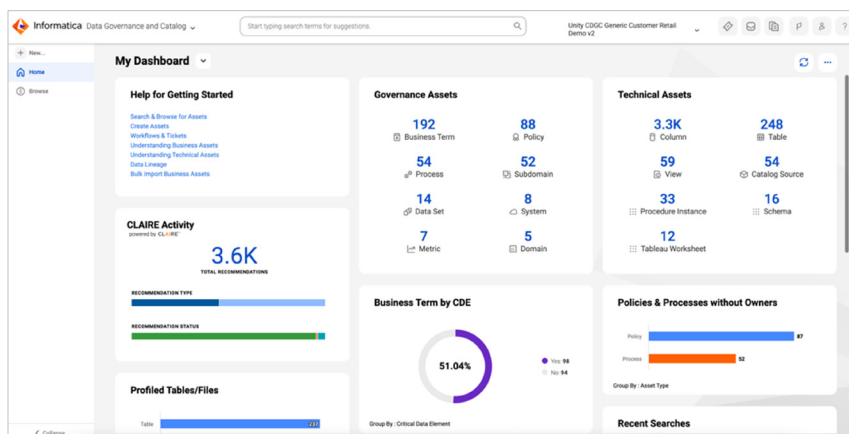


Figure 6. Manage your business, technical and governance assets from centralized, configurable, interactive dashboards.

Public APIs to Provide Seamless Integrations with Third-party Systems

Cloud Data Governance and Catalog helps enhance discoverability, enterprise-wide visibility, and data interoperability through API integrations with multiple third-party systems and interfaces. Data stewards can retrieve policies, rules and conventions from APIs, allowing reliable data to be more discoverable, accessible and reusable across the organization while helping to ensure adherence to data quality and compliance standards.



Key Benefits

Enhance Data-Driven Decision-Making by Improving Data Literacy

Organizations must thoroughly understand their data to get the most value from it. The powerful semantic search capability helps you discover the most relevant data assets. End-to-end data lineage views help you understand the full context of your data flow, including its source, transformations and usage. You can automatically associate business glossary terms with insight into quality, stakeholders, relationships, policies and classifications for rich business context. This data intelligence can help you improve data consumers' **data literacy** and confidence. You can enable **data democratization** and share accurate, complete and trustworthy data across the organization to empower data-driven decision-making at all levels.

Advance Usability of Data Assets with Business Context via Automation, Bulk Curation and Crowdsourcing

Boost productivity and maximize the usability and value of data by automating and augmenting common data management tasks at scale. With Cloud Data Governance and Catalog, you can automatically scan data and metadata across cloud and on-premises environments and streamline common data curation processes. Data professionals can spend less time on tedious tasks and focus on higher-value work thanks to capabilities such as automatic data classification, automatic association of business glossary terms to technical data assets and bulk data curation. Cloud Data Governance and Catalog also captures crowdsourced tags, annotations, ratings and reviews to further increase the value of data. This “wisdom of crowds” helps with data enrichment and curation, making it even more valuable throughout the organization while encouraging collaboration among stakeholders.

Help Assess and Mitigate Exposure Risks with Sensitive Data Discovery and Automatic Assignment of Recommended Data Policies

Cloud Data Governance and Catalog has more than 215 out-of-the-box classifications to help automatically discover and classify potentially sensitive data, including GDPR, PII, PHI and PCI data. The IDMC service can also automatically assign recommended data policies to relevant data classifications. These classifications allow data stewards to use data lineage to quickly identify datasets and sharing activity that may indicate potential privacy and similar exposure risks.

With improved transparency, your data protection and data sharing plans can help ensure compliance with policies for sensitive information use, limiting customer and intellectual property information exposure and averting risks from abuse and data loss

Enable Trusted Use of Governed AI Models and Their Underlying Data

In this age of data science, AI models are often opaque, built with poor-quality datasets and potentially noncompliant with organizational policies. The AI model governance capability provides insights into AI models and the data used to train the models. Insights are also provided for outputs produced, related policies' potential impacts and which models are available for reuse. This approach ensures that models consumed are relevant, their lineage is understood and policies applied are checked. It also provides visibility into any data drift to check the impact on the model's prediction capability. Informatica offers the only holistic solution for integrated governance of AI models and the data the models utilize.

Deliver Trustworthy Data and Develop a Data Governance Framework

A key benefit of Cloud Data Governance and Catalog is its ability to accelerate the development of a data and analytics governance framework. Its interactive dashboard helps you view, track and report the metrics required for monitoring data governance. You can define KPIs for data and analytics and create glossary hierarchies for context. The dashboard also allows you to define policy hierarchies and terms of use for data consumption and enable automated workflows to run whenever specific events and changes happen. These capabilities make it easier to connect data consumers with trustworthy data and allow you to democratize and share data with confidence.

For More Information

To learn more about intelligent data governance tools that can help you connect data consumers with trusted data, visit our [Cloud Data Governance and Catalog](#) and [Cloud Data Marketplace](#) solution pages.

Get a faster return on your investment with Informatica Professional Services

Talk to our experts in cloud data management.

[Check out our services.](#)

Informatica (NYSE: INFA) brings data to life by empowering businesses to realize the transformative power of their most critical assets. When properly unlocked, data becomes a living and trusted resource that is democratized across your organization, turning chaos into clarity. Through the Informatica Intelligent Data Management Cloud™, companies are breathing life into their data to drive bigger ideas, create improved processes, and reduce costs. Powered by CLAIRE®, our AI engine, it's the only cloud dedicated to managing data of any type, pattern, complexity, or workload across any location — all on a single platform. Informatica. Where data comes to life.

IN03-0523-4152

© Copyright Informatica LLC 2023. 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.



Informatica®