Informatica

Data Sheet

Cloud Data Governance and Catalog

Trusted Data: The Key to Scaling Digital Business

While digital transformation has been an organizational priority for years, the focus for most enterprises has shifted towards scaling digital business by leveraging modern, AI-enabled applications. However, to drive trusted outcomes with AI, enterprises need accurate and reliable data. The need to harness large volumes of trustworthy data, manage complex data landscapes and minimize fragmentation has made this task increasingly challenging.

To effectively leverage data and AI 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, AI-powered data intelligence is essential for organizations aspiring to accelerate value from their digital transformation initiatives.

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

The Informatica® Cloud Data Governance and Catalog (CDGC), a service of the 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.

CDGC delivers predictive data intelligence powered by the Informatica **CLAIRE®** AI and ML engine.

Key Benefits

- Enhance data-driven decision making by improving data literacy with faster insights
- Advance data asset usability with business context via automation, bulk curation and crowdsourcing
- Assess and mitigate exposure risks with sensitive data discovery and automatic assignment of recommended data policies
- Empower data users with federated governance through metadata access controls
- Build trust with data through inferred data lineage for strategic decision-making
- Enable trusted use of governed AI models and their underlying data
- Deliver trustworthy data and develop a data governance framework

Organizations that drive business value from trusted data can leverage 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 through metadata insights. 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.

CDGC 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

CDGC 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

The IDMC service provides a centralized, comprehensive view of your data. It features universal metadata connectivity, supporting nearly all your data sources. Additionally, it provides a runtime option to run serverless or within your on-premises or virtual private cloud.

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 CDGC custom metadata framework, you can use simple Excel files to ingest custom metadata. You can also 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 Solutions and Capabilities	PowerCenter, data integration, multidomain master data management (MDM) 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, SAP Datasphere (Preview), Snowflake, Databricks Delta Tables, Oracle Cloud Storage, Oracle ADB, MariaDB, Microsoft Fabric Data Warehouse
On-Premises	Oracle, IBM Db2, Netezza, SQL Server, Teradata, JDBC, MySQL, SAP HANA DB, Postgres, MongoDB, Local/Shared Filesystem
Database Scripts	Apache HiveQL Script, DB2 LUW SQL, Google BigQuery SQL Script, Microsoft SQL Server SQL, Oracle SQL, Snowflake SQL, Teradata BTEQ
BI and Analytics Platforms	Tableau, Microsoft Power BI, QlikView, Qlik Sense, Qlik Sense Cloud Microsoft SSRS, Cognos, Google Looker, SAP Analytics Cloud (Preview), SAS Base Programs, TIBCO Spotfire, Oracle BI
Other ETL and Data Science Platforms	Azure Data Factory, Databricks Notebooks, Databricks Unity Catalog, Microsoft SSIS, Microsoft SSAS, Oracle Data Integrator, Talend, IBM InfoSphere DataStage
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. CDGC uses intelligent data element and entity classification for automated **metadata management** and extraction from heterogeneous sources. Data profiling and classification can be automated across data assets at the field, column and table levels.

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

CDGC can auto-tag similar fields and columns across the enterprise using rule-based automation. The IDMC service automatically associates glossary terms with data and infers relationships among datasets using AI/ML capabilities, including schema matching. Users can also extend data elements that are critical for governance but not possible to scan and establish a connection between the created data element and other technical and business assets.

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.

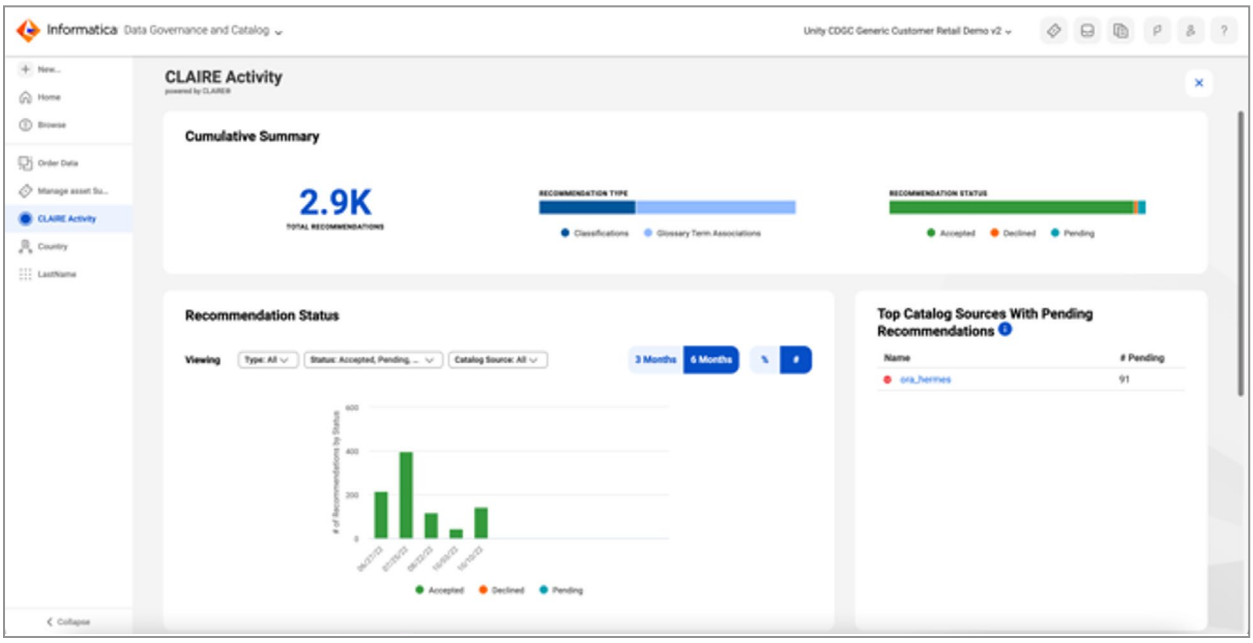

Figure 1. CLAIRE activity analytics provide a summary of intelligent glossary and classification associations across the organization.

Powerful, Intuitive Search and Browsing Capabilities

Users can perform natural language-like searches to locate critical data across business and technical domains with filtering and preview capabilities to quickly review and identify desired assets. All personas can easily explore data assets using browsable hierarchical views for context, relating technical data sources to business-curated datasets to provide a seamless experience. With asset page customization, users create custom views of objects based on their persona and find relevant information. Saved search functionality allows users to save frequent queries and perform the same search with a single click. In addition, the Informatica QuickLook browser extension allows users to search for text on any webpage and locate corresponding data assets available in the data catalog directly from their web browser.

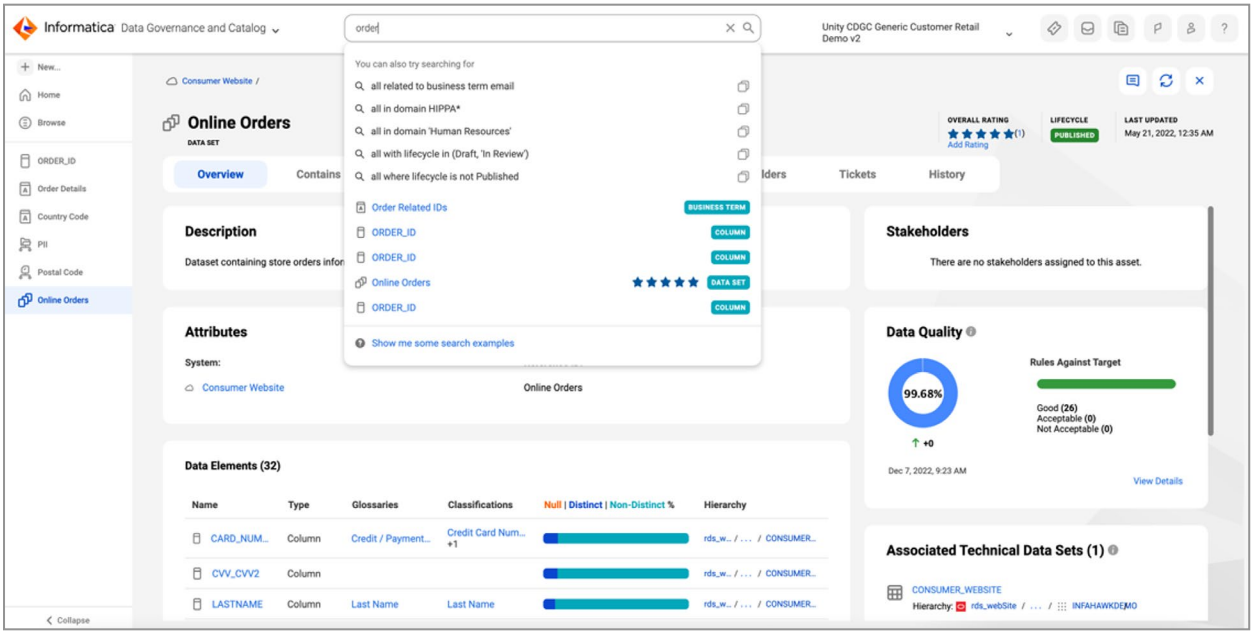

Figure 2. Quickly find data assets using powerful semantic search capabilities.

Metadata Access Control

Metadata access control offers new options for users to engage with assets in both CDGC and CDMP. As a data steward, you can control and manage detailed access policies within the Metadata Command Center. This capability helps ensure easy and efficient governance over metadata access of user roles in CDGC and CDMP. It also provides granular, attribute-based access control for any asset within the metadata platform through asset-specific roles. Federating data governance allows business domains to exercise greater authority over their data. Organizations can catalog their entire data landscape, managing asset ownership securely and at scale, while providing varied levels of access to different users.

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. Add new relationship types or modify or remove existing relationship types to accurately represent the relationships between assets as per organizational guidelines.

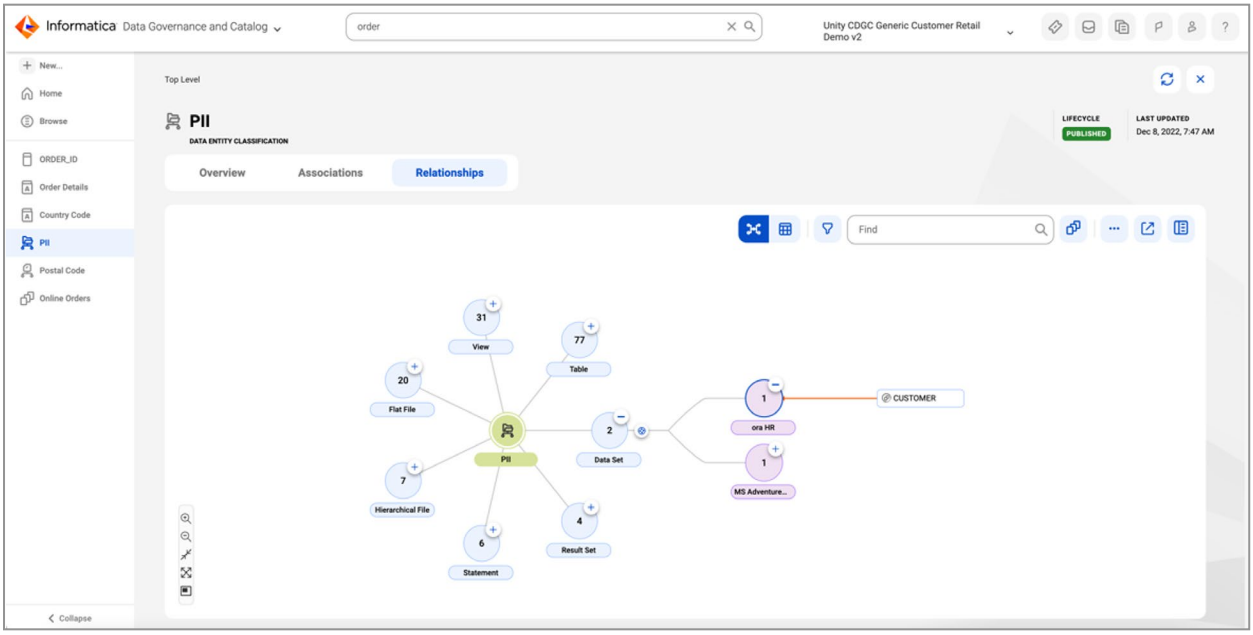

Figure 3. Interactive graphical views of data relationships help users discover and understand assets across the organization.

Data Lineage and Impact Analysis

Interactively trace data origin with data lineage views at any level. This encompasses business-friendly, system-level summarized views. These views also highlight endpoints to granular, column-level technical views that include intricate details automatically derived from parsing SQL scripts and stored procedures. Users 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 and save your lineage preference to recreate the same lineage views without re-applying the settings. Data quality overlays allow you to monitor quality scores and how they change throughout the data flow across your data estate.

Inferred Data Lineage

Due to technological limitations or security constraints, complete lineage may not be visible after metadata extraction in many enterprises. However, this gap can be bridged by employing inferred data lineage, which means that the data flow and relationships have been analyzed to make educated deductions about how data moves through processes, transformations and storage locations; for example, if a data pipeline extracts data from a source database, performs some transformations and then loads it into a target data warehouse, the inferred data lineage would show the flow from source to target, even if there isn't explicit documentation for each step. The system infers the lineage based on the observed patterns and dependencies, filling in gaps where documentation might be missing and providing a more complete picture of data movement across an organization. This is critical to ensure that your data is AI-ready.

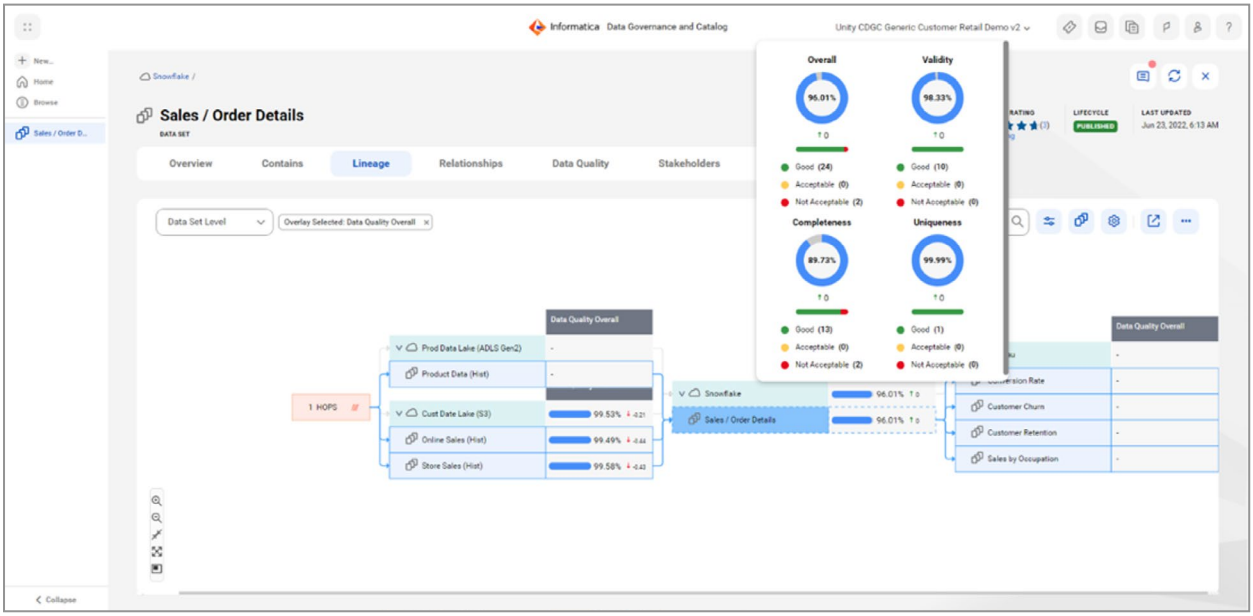

Figure 4. Explore data-element-level lineage from source to target and gain an additional understanding of your data by displaying graphical overlays, including data quality scores.

Integrated Data Quality and Observability

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. This approach significantly reduces the burden on stakeholders. Users are also offered data quality previews to help them assess the accuracy and usability of their data. Leveraging data observability capabilities, users can identify and analyze data anomalies for technical data sets, data elements and catalog sources, and prompt appropriate remedial action on data. Additionally, CDGC can automatically notify stakeholders of data quality status changes via the user interface or email, allowing them to act swiftly based on insights.

Collaboration and Social Curation

CDGC empowers data analysts and data scientists to easily find the most relevant and trusted data for analytics by utilizing AI, human expertise and collaboration. Users can determine if a data asset is added to a data collection in the **Cloud Data Marketplace** and view its details or even add a data asset to a collection on the marketplace. Data owners and subject matter experts can certify datasets. Data consumers can provide ratings and reviews for datasets, enabling 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. Saved searches and dashboard sharing allow you to share searches and dashboards with several users, user groups and user roles in your organization to simplify collaboration and accelerate adoption.

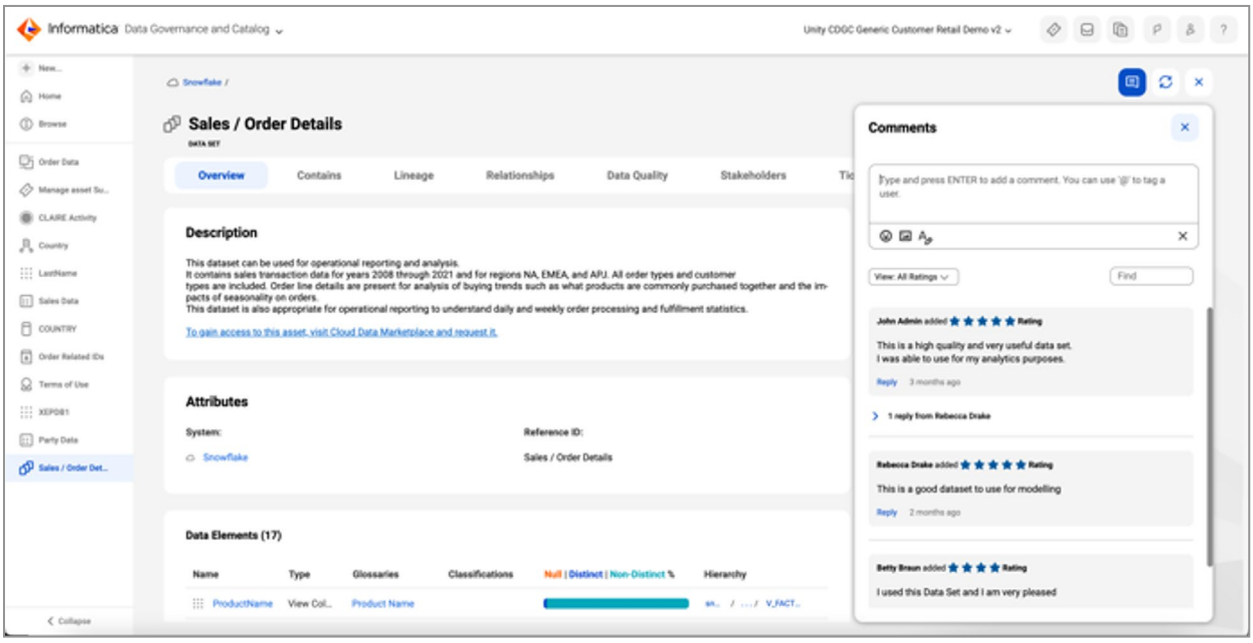

Figure 5. Collaboration capabilities include the ability to drive discussions and share comments and user ratings for data assets. Users can also certify technical assets.

Automated Customizable Workflows

Use workflows to automate processes and notifications, for reviewing and approving new data assets, and modifying existing assets. The automated workflows within CDGC help ensure that stakeholders create and modify assets in compliance with data governance principles within the organization's policies. The IDMC service offers predefined workflows for common processes. Design custom multi-step workflows based on asset types and roles to help simplify and accelerate workflow creation and implementation across broader deployments.

AI Model Governance

AI model governance advances 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. AI model governance also helps track and monitor model performance and key metrics, such as data drift, that may lead to model performance degradation and unreliable business outcomes.

Goal-Oriented Dashboards

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.

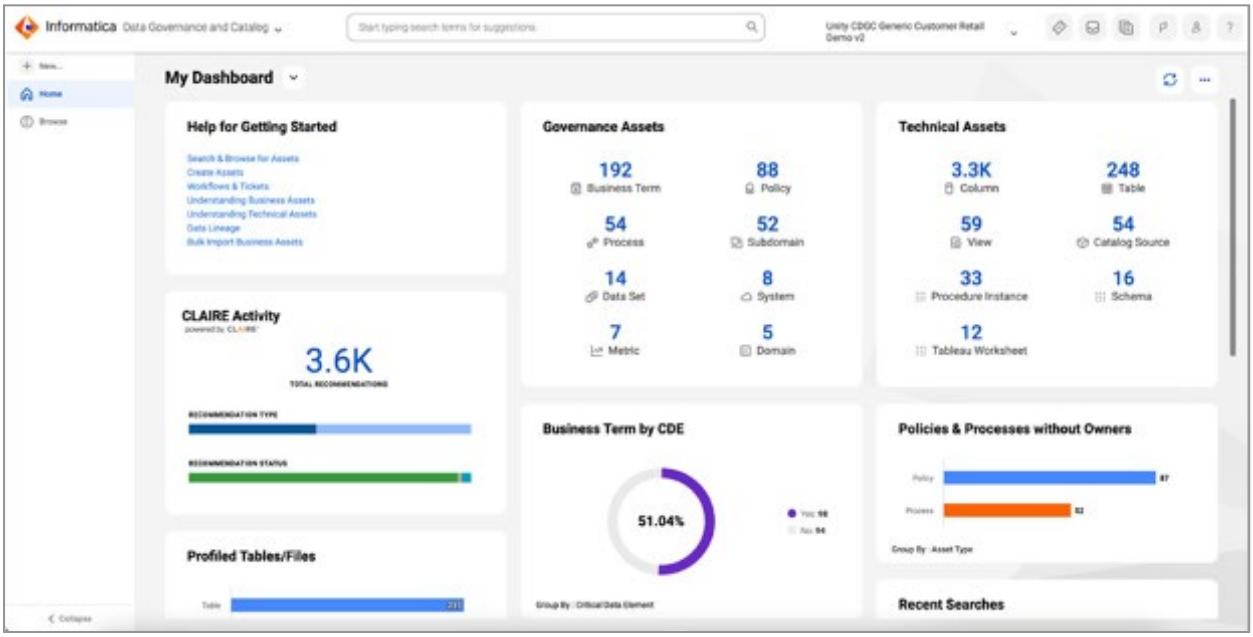

Figure 6. Manage your business, technical and governance assets from centralized, configurable, interactive dashboards.

Public APIs to Provide Seamless Integrations with Third-Party Systems

CDGC helps enhance data discoverability, enterprise-wide visibility, and interoperability through API integrations with multiple third-party systems and interfaces. Data stewards can retrieve policies, rules and conventions from APIs. This functionality allows reliable data to become more discoverable, accessible and reusable across the organization. Simultaneously, it plays a crucial role in ensuring adherence to data quality and compliance standards.

Key Benefits

Enhance Data-Driven Decision-Making by Improving Data Literacy With Faster Insights

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

Advance Data Assets Usability 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 CDGC, 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 with AI-enabled capabilities such as automatic data classification, automatic association of business glossary terms to technical data assets, and bulk data curation. CDGC captures a detailed history of changes made to the assets as well as 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.

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

CDGC includes more than 215 out-of-the-box classifications to facilitate the automatic discovery and classification of potentially sensitive data. This includes data relevant to industry-related regulatory frameworks such as GDPR, PII, PHI and PCI-related 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 risks.

With improved transparency, your data protection and data sharing plans can help ensure compliance with policies for sensitive information use. This approach helps limit customer and intellectual property information exposure and aids in averting risks from abuse and data loss.

Empowering Data Users with Federated Governance Through Metadata Access Controls

The introduction of this capability brings several business advantages. It significantly reduces the risk exposure from insider threats by enforcing policies across IDMC, including platforms such as CDGC and CDMP. Furthermore, it accelerates the time to value with minimal disruption to the existing enterprise architecture and processes. Organizations can open data-consuming applications on IDMC platforms, such as CDGC, CDMP and CLAIRE GPT, to their entire organization, fostering a more integrated and agile data environment.

Building Trust With Data Through Inferred Data Lineage for Strategic Decision-Making

Inferred data lineage as part of CDGC, can help organizations deliver the data that builds confidence in their analytics and AI models, improving customer experience programs, helping to ensure regulatory compliance with industry policies and accelerating cloud modernization initiatives. Business users can enhance governance and privacy, deepen data analytics, transition to the cloud and augment the customer experience with greater ease and assurance. Concurrently, your IT teams and data analysts can refine change management, improve operational efficiency, reinforce data security and enhance responsible AI governance.

Making strategic decisions encompassing digital transformation while remaining compliant with an array of regulations, including new AI controls, requires trustworthy data. Investing in an intelligent, enterprise-scale data catalog designed for multi-cloud and on-premises environments can empower organizations to succeed in the new age of AI systems.

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. AI model governance capabilities provide insights into AI models and the data used to train models. Insights are also provided for the outputs produced, related policies' potential impact and which models are available for reuse. This approach ensures that models that are used are relevant, their lineage is understood and policies applied are checked. It also provides visibility into data drift to check the impact on the model's prediction capability. Informatica offers a holistic solution for integrated governance of AI models and the data the models utilize.

Deliver Trustworthy Data and Develop a Data Governance Framework

CDGC can 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. The service enables automated workflows to run whenever specific events and changes happen. These capabilities make it easier to connect data consumers with trustworthy data to democratize and share data with confidence.

Next Steps

To learn more about intelligent data governance tools that can help you connect data consumers with trusted data, visit **www. informatica.com/ cloud- data-governance- and- catalog**.

Where data & AI come to

Informatica (NYSE: INFA), a leader in enterprise AI-powered cloud data management, brings data and AI to life by empowering businesses to realize the transformative power of their most critical assets. We have created a new category of software, the Informatica Intelligent Data Management Cloud™ (IDMC), powered by AI and an end-to-end data management platform that connects, manages and unifies data across virtually any multi-cloud, hybrid system, democratizing data and enabling enterprises to modernize their business strategies. Customers in approximately 100 countries and more than 80 of the Fortune 100 rely on Informatica to drive data-led digital transformation. **Informatica. Where data and AI come to life.™**

IN06-4152-0125