Informatica

Data Sheet

Cloud Data Governance and Catalog

Benefits

- Accelerate business outcomes by providing data consumers with connections to easily find, understand, trust and access data
- Improve data literacy to increase data-driven decision making that drives business value
- Enrich data assets with business context via crowdsourcing to advance usability
- Discover sensitive data 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

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 given to one of their most valuable assets: data. Successful digital transformation heavily depends on delivering trusted data to data consumers; however, complex data landscapes and fragmentation have made this task increasingly challenging.

Data consumers need to be able to find, understand, trust and access relevant data for various use cases including enhancing customer experience, enabling innovation and ensuring greater compliance with regulatory authorities. Data intelligence coupled with seamless data delivery is essential for organizations aspiring to accelerate their digital transformation journeys.

Cloud Data Governance and Catalog: A Cloud-Native, Comprehensive Data Intelligence Solution

Informatica® Cloud Data Governance and Catalog is a cloud-native solution that enables customers to find, understand, trust and access their data. It brings together data governance, data catalog and data quality capabilities into a singular tool for automating data intelligence insights. This multi-tenant SaaS solution is built for organizations that want to maximize their investments by deriving value from their vast data assets in the cloud.

The solution delivers predictive data intelligence — automated and recommendation-driven data classification, data curation and relationship and sensitive data discovery powered by CLAIRE™ AI and ML to accelerate deep and broad metadata connectivity — allowing organizations to quickly drive business value from trusted data. Cloud Data Governance and Catalog enables efficient self-service analytics and data governance by unifying the capabilities of data discovery, lineage, profiling, business glossary creation and stakeholder and policy management — as well as the ability to document and manage 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. Since the solution is cloud native, infrastructure is available almost immediately and at the scale needed by the organization.

Key Capabilities

AI-powered CLAIRE™ Engine to Drive Insights from Metadata

Automation is critical to manage and govern large data estates. Cloud Data Governance and Catalog helps automate metadata management and extraction from heterogeneous sources. With intelligent domain and entity recognition, automated data profiling and classification across structured and unstructured data assets at the field, column and table level. The solution can also automatically associate glossary terms to data and infer relationships such as joins among datasets using AI/ML capabilities including schema matching.

Data Lineage and Impact Analysis

Interactively trace data origin with data lineage views at any level, from business-friendly, system-level summarized views that highlight endpoints to granular, column-level technical views that include all the intricate details in between. Users can efficiently derive data lineage and data transformation logic with the ability to parse SQL scripts and stored procedures in databases. Users also have the ability to perform detailed impact analysis on upstream and downstream data assets. The use of overlays allows tracking by business and technical attributes such as Stakeholders, Policy Impacts, Glossary Term definitions and more.

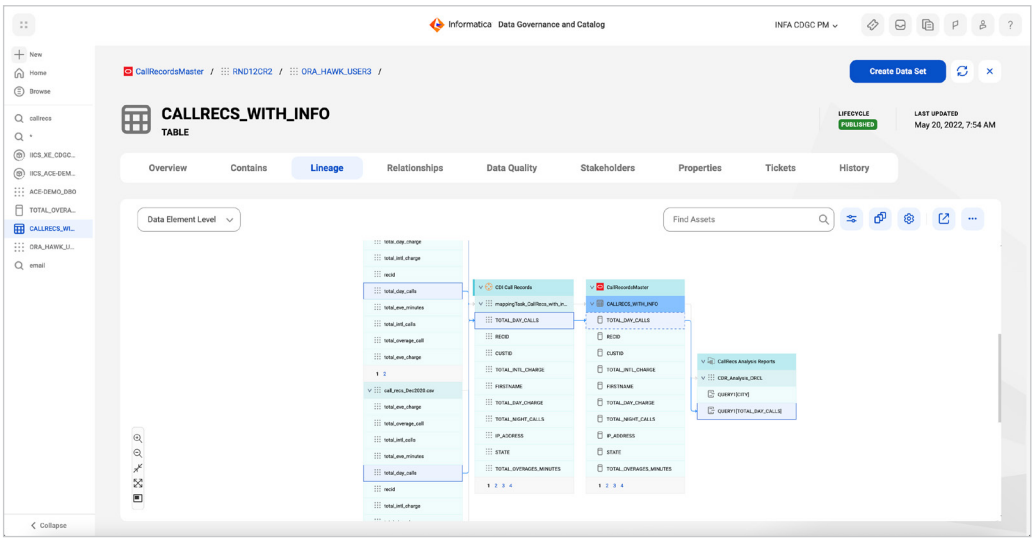

Figure 1: Explore data element-level data lineage from source to target.

AI Model Governance

AI model governance strives to push towards explainable AI by providing organizational visibility and transparency into models and their underlying algorithms that often becomes a black box for most in the organization. It provides details on how the model was developed, the training data used for creating the models along with 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.

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 and data type and data domain inference help automate the measurement of data quality, significantly reducing the burden on stakeholders.

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 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.

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.

Powerful, Intuitive Search and Browsing Capabilities

Users can perform natural language-like searches to locate critical assets across business and technical domains and make use of 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.

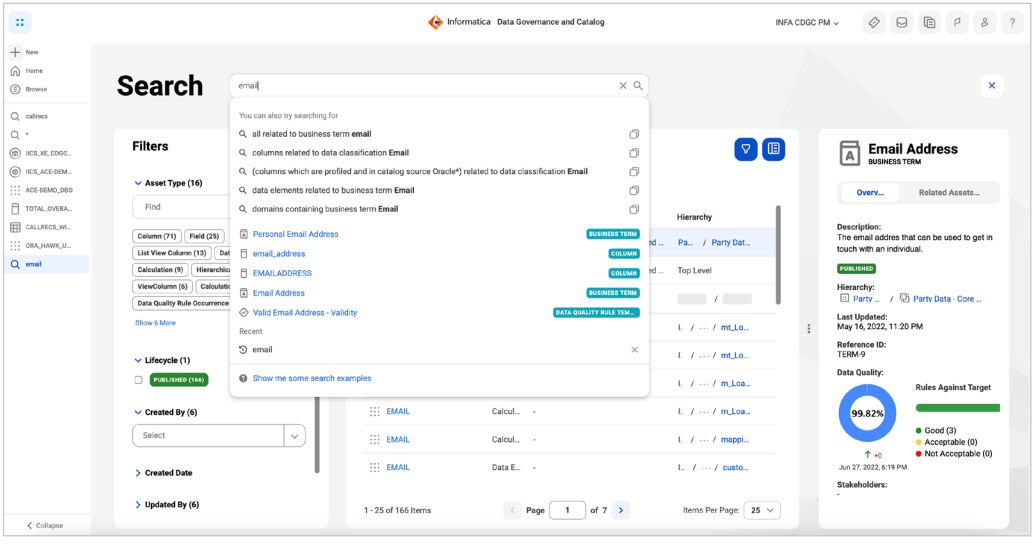

Figure 2: Quickly find data assets using powerful, semantic search capabilities.

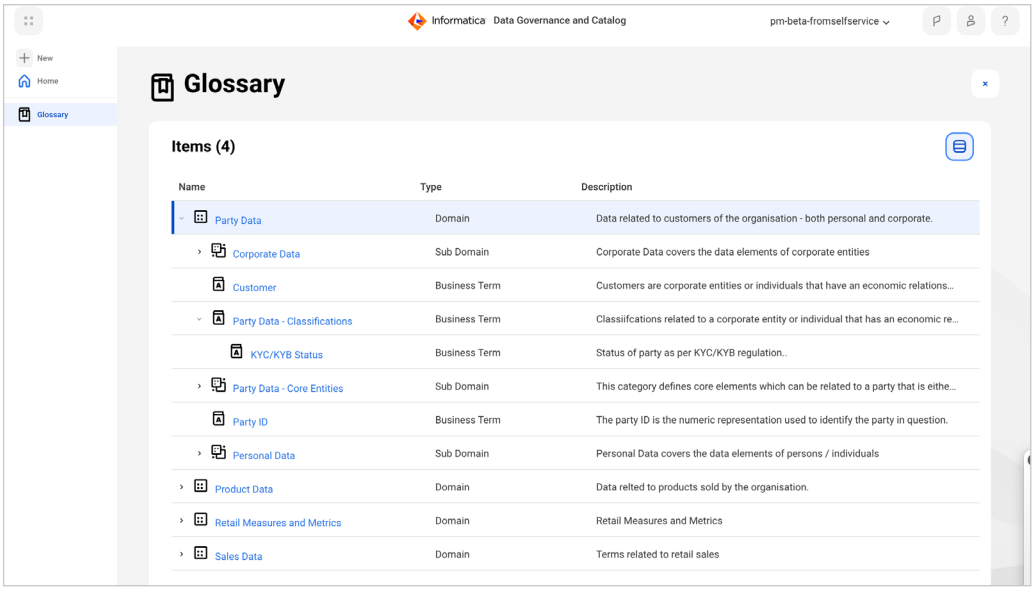

Figure 3: Create browsable hierarchies for glossary concepts.

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 pre-defined workflows, check task completions and view notifications. With a variety of visualizations and drill-down capabilities, users can quickly view summary status and explore details as needed.

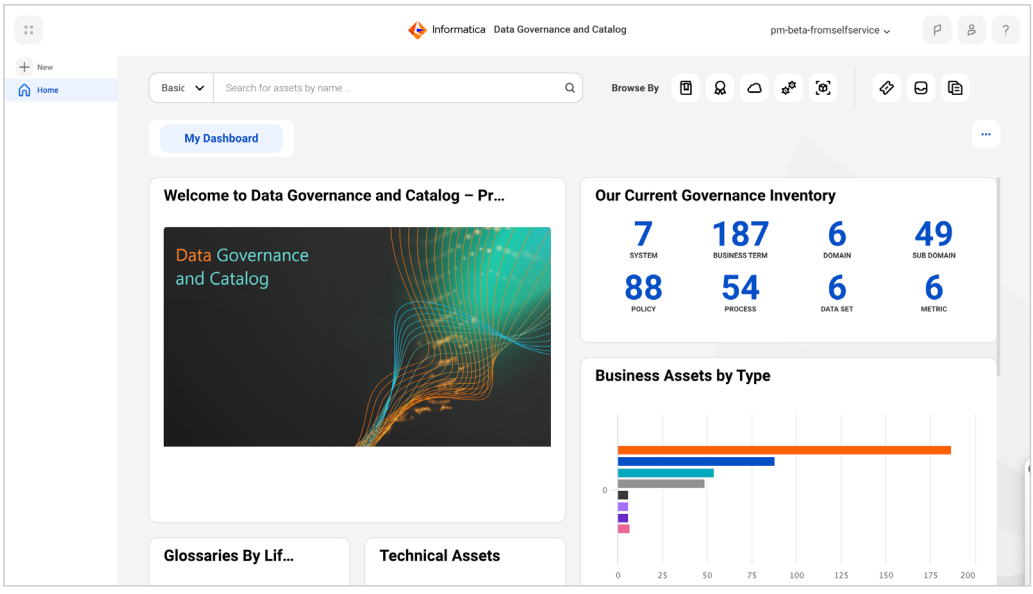

Figure 4: Manage your business, technical and governance assets from centralized, configurable, interactive dashboards.

Automated Workflows

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

Broad and Deep Metadata Connectivity

Cloud Data Governance and Catalog offers broad and deep metadata connectivity that spans multi-cloud and on-premises and environments. By applying broad and deep data source connectivity, extract metadata across cloud platforms, BI tools, databases, multi-vendor ETL and data science tools, various enterprise applications and file formats, SQL dialects, and stored procedures.

Visually inspect scripts, procedures and processes to fully understand logic and internal data flow. Obtain complete column-level data lineage, including a full inventory of all potential lineage sources with rich details. Scan both static and dynamic code, as well as perform language parsing for automated data lineage.

With the Cloud Data Governance and Catalog Custom Metadata Framework, use simple Excel files to ingest custom metadata and derive data lineage and relationship links from key systems where automated scanners are not available. Model virtually any data source or data lineage across systems.

Data sources supported include:

- **Informatica:** PowerCenter, Cloud Data Integration, Cloud Master Data Management
- **Cloud Platforms:** AWS S3, AWS Redshift, AWS RDS Oracle, AWS RDS SQL Server, AWS RDS Postgres, AWS RDS MySQL, Azure SQL DB, Azure Synapse, Azure ADLS Gen 2, Azure Blob, Google Cloud Storage, Google BigQuery, Snowflake, Databricks Delta Lake
- **Databases/Data Warehouses:** Oracle, DB2, SQL Server, Teradata, JDBC, MySQL, Sybase ASE, SAP HANA DB
- **BI and Analytics Platforms:** SAP BusinessObjects, Tableau, Microsoft Power BI, MicroStrategy, QlikView, Qlik Sense, Microsoft SSRS, SSAS
- **Other ETL and Data Science Platforms:** Azure Data Factory, Databricks Notebooks, Microsoft SSIS
- **Enterprise Applications:** Salesforce, Kafka, Workday, SAP BW, SAP BW4/Hana, SAP ECC, SAP S/4 Hana
- **File Formats:** CSV, Delimited, XML, JSON, Avro, Parquet
- **Advanced SQL Parsing:** Oracle, SQL Server, Snowflake, Azure Synapse, AWS Redshift, Google Big Query

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

Key Benefits

Improve Data Literacy to Increase Data-driven Decision Making that Drives Business Value

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 understand the full context of your data flow including its source, transformations and usage. The ability to automatically associate business glossary terms combined with insight into quality, stakeholders, relationships, policies and classifications, provides rich business context for users. With this data intelligence, improve the data literacy and confidence of data consumers, allowing them to easily find, understand, trust and access relevant data. Enable data democratization and share accurate, complete and trustworthy data across the organization to empower data-driven decision making at all levels.

Enrich Data Assets with Business Context via Crowdsourcing to Advance Usability

Data Governance and Catalog maximizes the reuse and value of data by automatically classifying data assets down to the field/column level. To further increase the value of data, the solution captures crowdsourced tags, annotations, ratings and reviews. This "wisdom of crowds" helps with data enrichment and curation, making it even more valuable throughout the organization, while encouraging collaboration among stakeholders.

Discover Sensitive Data Sets to Help Assess and Mitigate Exposure Risks

Enable data stewards to quickly identify datasets and sharing activity using data lineage that may indicate potential privacy and similar exposure risks to resolve. 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.

Govern AI Models and Their Underlying Data to Enable Trusted Use

In this age of data science, AI models are often opaque, built with poor quality datasets and potentially non-compliant 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, potential impacts of related policies and which models are available for reuse. This approach ensures models consumed are relevant, that their lineage is understood and that any applicable policies 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 as well as the data utilized by the models

About Informatica

At Informatica (NYSE: INFA), we believe data is the soul of business transformation. That's why we help you transform it from simply binary information to extraordinary innovation with our Informatica Intelligent Data Management Cloud.™ Powered by AI, it's the only cloud dedicated to managing data of any type, pattern, complexity or workload across any location — all on a single platform. Whether you're driving next-gen analytics, delivering perfectly timed customer experiences or ensuring governance and privacy, you can always know your data is accurate, your insights are actionable and your possibilities are limitless. Informatica. Cloud First. Data Always.™

Accelerate the Development of a Data Governance Framework to Deliver Trustworthy Data

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 view, track and report the metrics required for monitoring data governance — and delivers the ability to define KPIs for data and analytics, create glossary hierarchies for context, define policy hierarchies and terms of use for data consumption and enable automated workflows that can be triggered by events and changes. These capabilities help connect data consumers with trustworthy data, enabling organizations to democratize and share their data with confidence.

Learn More

Visit our webpages for Cloud Data Governance and Catalog and Cloud Data Marketplace to learn more about intelligent data governance tools from Informatica that can help you connect data consumers with trusted data. Or join us for a live product demo.

Informatica

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

IN06_0822_04152

© Copyright Informatica LLC 2022. Informatica, the Informatica logo and CLAIRE 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.