

How CLAIRE AI Engine Can Help You Automate Data Governance

Boost Productivity and Accelerate Data-Driven Decision-Making

Many successful organizations are relying more on data to drive their business. By using data effectively, companies can achieve improved productivity, lower costs and increased revenue growth. Although many organizations recognize the importance of data, they often require assistance in providing their end-users with reliable and relevant data to support informed decision-making. It is crucial to ensure that the data is trustworthy and appropriate for use.

According to Anaconda's 2022 State of Data Science report, data professionals spend a significant amount of their time (38%) on data preparation and cleaning instead of more advanced tasks such as selecting, training and deploying models. The report also reveals that 63% of commercial respondents are worried about the lack of data professionals.¹ These findings underscore the importance of organizations making the most of their data experts' time.

Organizations can save time and effort by simplifying tedious tasks like finding, selecting, cleaning and preparing data. By doing so, data experts can concentrate on more critical work, such as generating fresh insights.

Organizations need to help their people find, understand, trust and access data to achieve their business goals in today's digital world. An excellent way to do this is to use a robust data governance solution that uses automation and artificial intelligence (AI) to give context and transparency to data.

Key Benefits

- Enable faster, data-driven results that boost productivity and reduce costs
- Accelerate time-to-value for analytics initiatives with improved data understanding
- Enhance regulatory compliance and mitigate risk exposure with better protection for sensitive data

¹ Anaconda, 2022 State of Data Science Report

Improve Efficiency and Productivity with Informatica CLAIRE AI Engine as Your Copilot

The **Informatica Intelligent Data Management Cloud™ (IDMC)** offers a unified, AI-powered solution that uses predictive data intelligence to deliver reliable data for intelligent decision-making. This cloud-native solution harnesses the power of Informatica's cross-platform, metadata-driven AI engine **CLAIRE®** to automate data management across different environments, including multi-cloud, on-premises and hybrid.

With AI copiloting capabilities, IDMC automates various aspects of data management – such as **data cataloging**, **data quality**, data observability, **master data management (MDM)**, data governance, data privacy and data sharing – to boost efficiency and productivity throughout the enterprise.

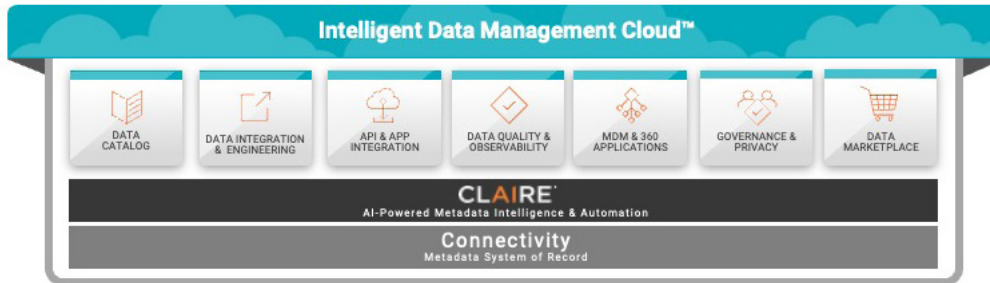


Figure 1. Informatica Intelligent Data Management Cloud provides a single, unified, AI-powered solution.

With IDMC and CLAIRE as a copilot, organizations can improve performance by simplifying processes, increasing efficiency and delivering data faster to their teams. This helps to support the business goals and stay ahead of competitors.

Key Capabilities

Metadata Scanning and Knowledge Graph Creation

With CLAIRE, you can quickly scan and extract **metadata** from a variety of sources such as code, SQL scripts, stored procedures, big data platforms, ETL tools, data integration platforms, business intelligence platforms and various file formats.

The AI engine then uses this information to create a helpful knowledge graph for businesses. This graph reveals the connections and pathways through which data flows within the organization. This is especially helpful because, in most cases, data is stored in separate applications and department-level repositories. With this knowledge graph, data professionals can quickly discover and understand how data is related across the organization. This saves time and makes it easier to manage complex data landscapes.

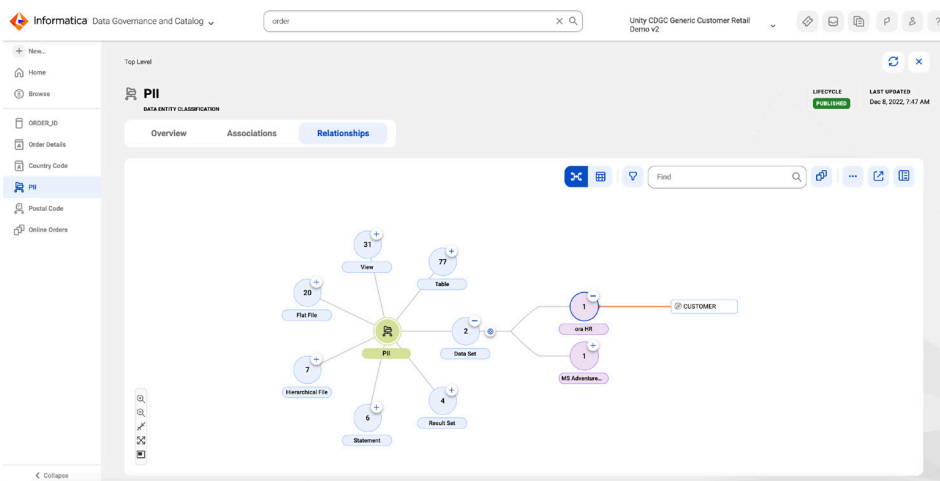


Figure 2. Interactive graphical views of data relationships help organizations discover and understand assets.

Automated Data Classification

CLAIRE can help organizations classify and organize data automatically by identifying data element types, entity types and similar columns. The AI engine also learns from manual associations to classify similar fields, columns and tables across the enterprise. This automatic classification helps ensure that your data is labeled and organized accurately, reducing the risk of errors and improving overall data quality and findability. It also saves time and effort on manual tasks, resulting in significant cost savings.

Intelligent Glossary Associations

With CLAIRE, business glossary terms can be automatically linked to physical data elements, giving data assets more meaning and making them more understandable for users. By using auto domain discovery and natural language processing (NLP), CLAIRE streamlines the process of glossary associations, significantly reducing the time required for this task. For example, one healthcare organization that originally spent 2.5 months manually associating thousands of glossary terms to columns could perform the same task in less than 10 minutes using CLAIRE.

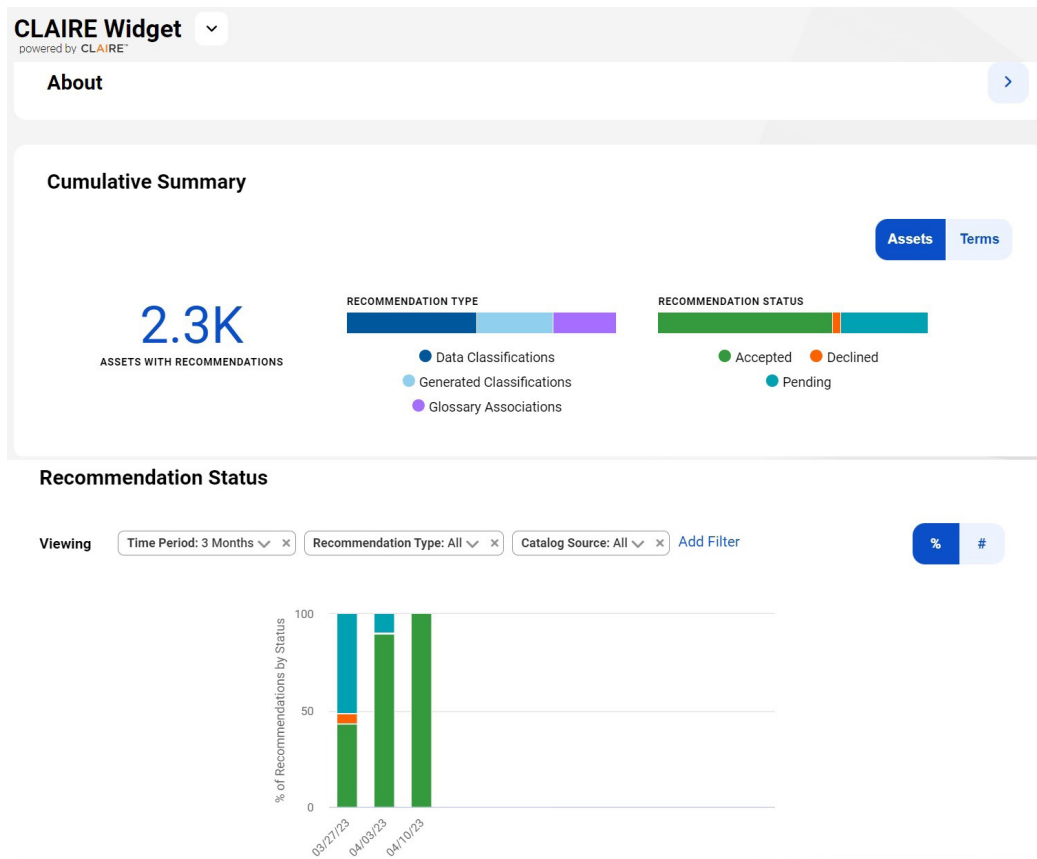


Figure 3. CLAIRE activity analytics summarizes intelligent glossary and classification associations across the organization.

Data Quality Rule Generation and Automated Application

CLAIRE uses NLP to improve data quality. It creates rules from users' natural language sentences and automates them for organizations to use across their data estate. This allows enterprises to apply data quality rules based on business glossary association or data classification. It helps businesses define and execute rules automatically and consistently.



Figure 4. In Cloud Data Governance and Catalog, users can describe data quality rules using natural language. CLAIRE can translate those user-generated descriptions and offer recommendations for rules.

Automated Data Profiling and Insights

CLAIRE can detect anomalies and offer essential insights by analyzing data profiling results. These insights are crucial in creating and enforcing data quality rules to monitor the profiled data continuously. CLAIRE automatically can suggest and generate rules related to completeness, uniqueness and validity based on the outcomes of its analysis.

Additionally, the AI engine can automatically assign these rules upon approval, making monitoring data more efficient. By constantly monitoring profiling results, CLAIRE can refine and update its recommendations, quickly detecting critical issues that need immediate attention.

Insight Statement	Score	Insight Type	Columns	Status
<input type="checkbox"/> Numeric values found outside the 95% standard deviation range.	High	Number Value Distribution	_ID	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	Address	
<input type="checkbox"/> The majority of the data values in the column are unique.	High	Uniqueness Check	Address	Approved
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	Address1	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="refresh"/>
<input type="checkbox"/> Data appears incomplete. The column includes one or more null, blank, or empty values or v...	High	Completeness Check	City	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	City	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	Company_Name	
<input type="checkbox"/> The majority of the data values in the column are unique.	High	Uniqueness Check	Company_Name	Disapproved
<input type="checkbox"/> Data appears incomplete. The column includes one or more null, blank, or empty values or v...	Low	Completeness Check	Contact_Name	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	Contact_Name	
<input type="checkbox"/> Data appears incomplete. The column includes one or more null, blank, or empty values or v...	Low	Completeness Check	Country	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	Country	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	DUNS_Number	
<input type="checkbox"/> Numeric values found outside the 95% standard deviation range.	High	Number Value Distribution	DUNS_Number	
<input type="checkbox"/> One or more date values do not match the locale format.	High	Date-Locale Check	DUNS_Number	
<input type="checkbox"/> One or more dates do not comply with a valid date pattern.	High	Date Validity Check	DUNS_Number	

Figure 5. CLAIRE can detect anomalies and present insights based on profiling results. User acceptance of insights automatically creates and assigns data quality rules for profiling and monitoring.

Inferred Relationships

Finding relevant datasets can be time-consuming due to the often-siloed nature of data across the enterprise. CLAIRE can help by automatically providing a comprehensive view of data. It **discovers** relationships using machine learning (ML) techniques, identifies primary and unique keys and infers joins across structured datasets. This reduces potential months of documentation effort to just minutes for data stewards.

Data analysts and scientists can also better understand relevant and potentially complementary datasets with this information. To continuously improve its ability to identify relationships, CLAIRE includes humans in the data curation. Users can accept or reject inferred relationships, allowing CLAIRE to learn from these actions.

Sensitive Data Management

CLAIRE offers more than 225 built-in classifications to help automatically find and categorize data, including data related to the General Data Protection Regulation (GDPR), the Payment Card Industry Data Security Standard (PCI DSS), personally identifiable information (PII), electronically protected health information (ePHI) and intellectual property (IP). Users can customize these classifications or make entirely new ones.

Additionally, users can create policies that apply to specific data assets, either directly or through classifications. These policies specify which data assets are considered sensitive and where these rules should be applied.

Data Protection Policy Management and Enforcement

As mentioned earlier, CLAIRE can automatically help organizations link policies to data assets by using AI/ML-based data classification upon discovery. This provides an understanding of the policies that govern data usage. Organizations can enforce these policies using IDMC data protection masking and obfuscation capabilities. Alternatively, they can use third-party or platforms-as-a-service offerings that host the data to manage policy-compliant data access.

Key Benefits

Boost Productivity to Reduce Costs and Enable Faster Data-Driven Results

Enable data stewards, curators and data governance teams to be more productive and efficient with CLAIRE AI copilot capabilities. This will save organizations time and money. Additionally, data stewards can simplify and speed up their daily tasks with intelligent glossary associations, relationship graphs, automated lineage and other helpful features.

Improve Data Understanding to Accelerate Time-to-Value for Analytics Initiatives

With CLAIRE, data users can more easily locate and access the data they need for analytics and data science initiatives. It provides semantic search, automated data classifications, an integrated business glossary and end-to-end data lineage to assist in data discovery, understanding and quality. When data consumers have the necessary context and transparency to understand and trust the data, they can use it with confidence to make better decisions. This accelerates time-to-value for data-intensive projects.

Enhance Regulatory Compliance and Mitigate Risk Exposure by Protecting Sensitive Data

Using CLAIRE as an AI copilot helps organizations establish and enforce data privacy and security policies. With CLAIRE, organizations can automatically identify, classify and mask sensitive data, associate relevant policies with data assets, and assist with monitoring data access and usage through data lineage and dashboards. This helps to reduce the risk of non-compliance with regulations and internal policies.

Where data & AI come to



Worldwide Headquarters

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

Informatica (NYSE: INFA) brings data and AI 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.

IN17-4591-0923

© 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.

Next Steps

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](#), [Cloud Data Quality](#) and [Cloud Data Marketplace](#) services pages. Or check out our self-guided, interactive demos in our [Experience Lounge](#).