IBM watsonx.data intelligence

Data Governance andCatalog, Data Quality, Data Product Hub, Manta Data Lineage

Last updated: March 17, 2026

Please Note

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice and at IBM’s sole discretion.

Information regarding potential future products is intended to outline our general product direction and should not be relied on in making a purchasing decision.

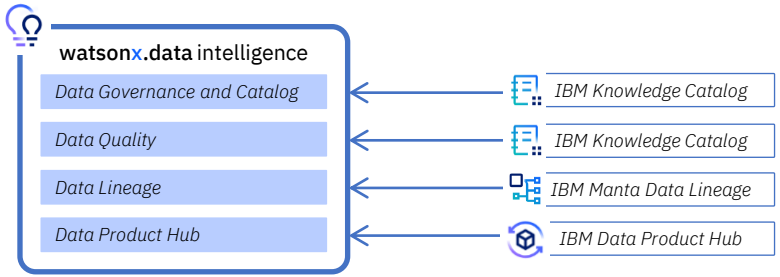
The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve similar results to those stated here.

watsonx.data intelligence!

A comprehensive stack of deeply connected capabilities covering Data Governance and Catalog, Data Quality, Data Lineage and Data Product Hub features with streamlined entitlement flexibility



for structured and unstructured data..

...built on a single part #

Intelligent data – impactful results

Data stewards...

Increase their productivity by eliminating manual error prone tasks by leveraging automated and AI assisted data governance features for data discovery, curation, quality, and data protection.

Data engineers...

Decrease time spent on optimizing data processes by automating routine processes and providing a deeper understanding of their data landscape allowing them to work smarter, not harder.

Data consumers...

Decrease time spent finding, understanding and accessing the high-quality data they need for analytical and AI use cases with trust to produce accurate and meaningful business outcomes.

3 strategic imperatives differentiate IBM in data and analytics governance

End-to-End AI-powered Data & Analytics Governance Platform

IBM delivers a fully integrated platform spanning data, model, and application layers. This end-to-end ownership enables IBM to embed **AI automation** & **metadata intelligence across the entire platform** and offer a truly unified experience.

Open

IBM partners across the ecosystem to enable **open metadata exchange** across diverse and distributed **data silos**

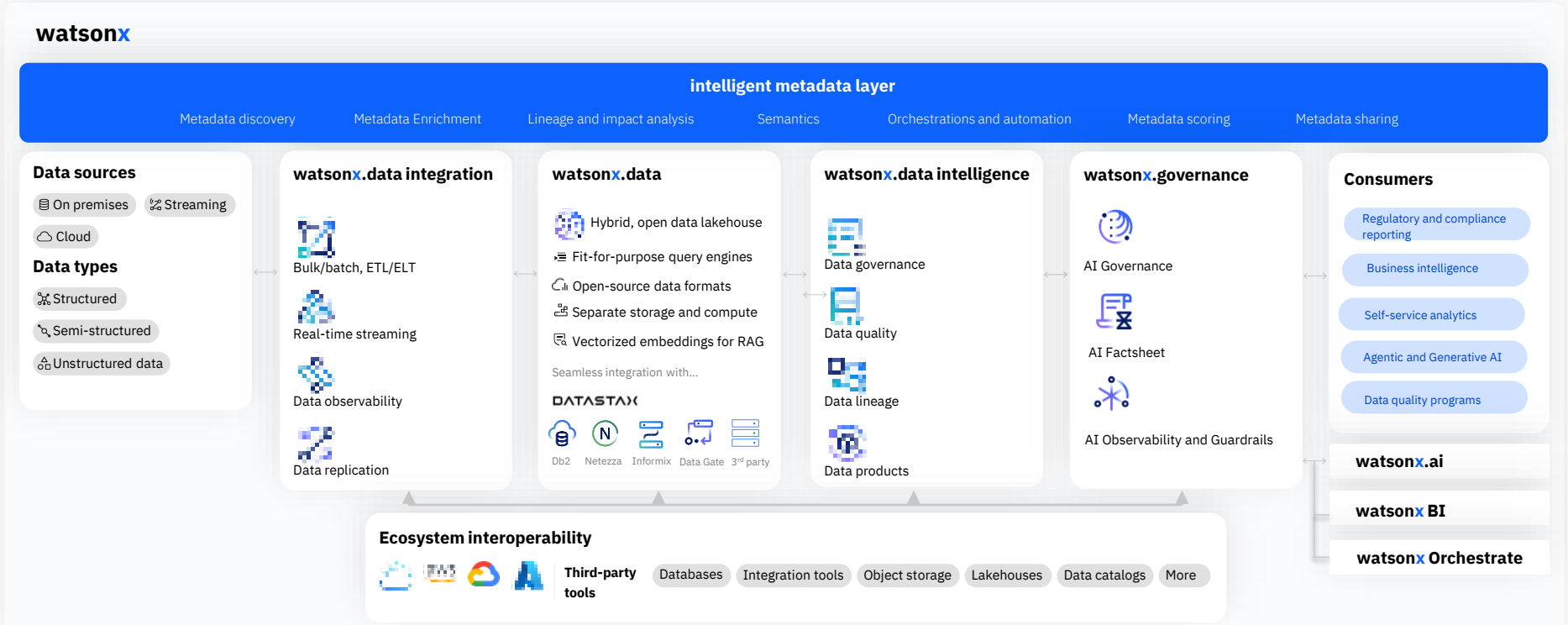
IBM's approach adapts to the ever-changing landscape of data, tools, and applications.

Hybrid

IBM is the **only vendor** that delivers robust data and analytics governance that **spans on-prem, multi-cloud, and SaaS** — enabling consistent management across hybrid entire AI and data estate.

IBM maintains a commitment to supporting enterprise clients with both **old & new technology infrastructures**.

Our Strategy: Delivering watsonx – an open, integrated Data and AI platform, built for hybrid – powered by an intelligent metadata layer



Market Recognition

IBM is recognized as a Leader in multiple Gartner® evaluations across Data Intelligence categories

2025 Gartner® Magic Quadrant™ for Metadata Management Solutions

Figure 1: Magic Quadrant for Metadata Management Solutions

2026 Gartner® Magic Quadrant™ for Data Analytics and Governance Platforms

Figure 1: Magic Quadrant for Data and Analytics Governance Platforms

2026 Gartner® Magic Quadrant™ for Augmented Data Quality Solutions

Figure 1: Magic Quadrant for Augmented Data Quality Solutions

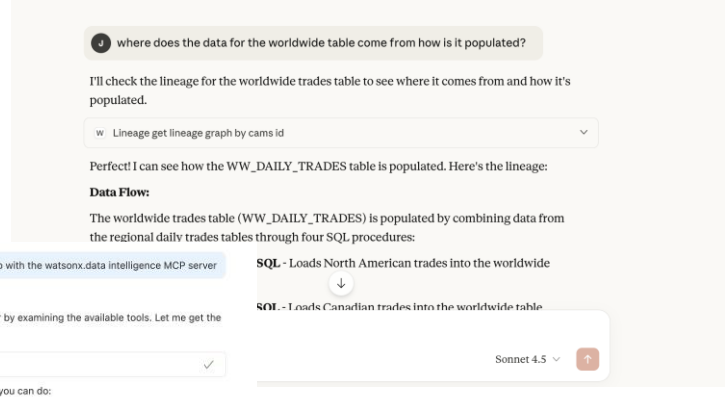
Open Source MCP Server

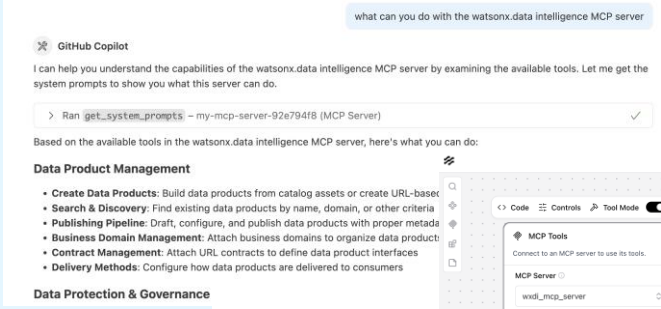
https://github.com/IBM/data-intelligence-mcp-server

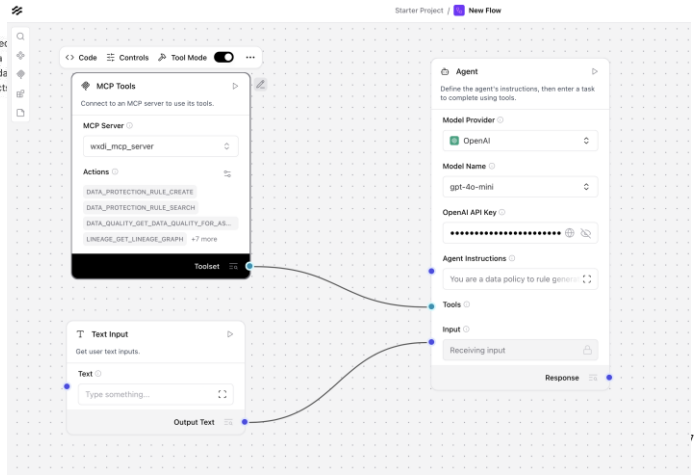
What is it?
Model Context Protocol is the open standard for connecting LLMs with external tools and data in a consistent way. This lets any assistant — whether it's watsonx Orchestrate, Claude, Microsoft Copilot, etc. — connect directly to watsonx.data intelligence capabilities like data lineage, governance, or data quality. This same openness lets clients and partners build their own AI agents.

Key Features:
- Discover assets across catalogs & data products
- Analyze lineage for root cause & impact analysis
- Transform data into views
- Create data protection rules
- … and more!

Why?
Clients demand open & extensible systems. Our MCP server allows clients to build their own chat assistants & AI agents





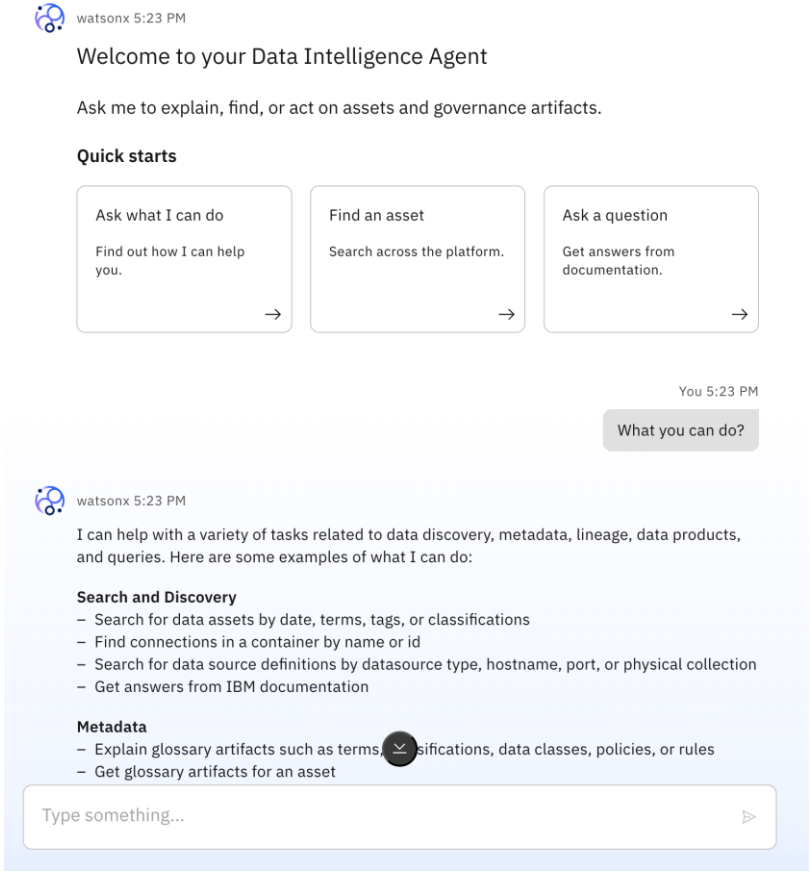


Data Intelligence Agent

What is it?
The Data Intelligence Agent helps users find, understand, and work with their data via natural language conversations.

Key Features:
- **Asset discovery & understanding**
  - Find assets from across all watsonx.data intelligence and query metadata from the catalog, reporting data mart, and lineage.
- **Lineage Analysis**
  - Conduct impact/root cause analysis and generate descriptions of data pipelines from natural language
- **Documentation Q&A**
  - Learn to use the platform and remediate errors by querying documentation
- **... and much more to come**

Why?
Customers of the ChatGPT era demand simple, declarative interactions with software. The data intelligence Agent uses the latest agentic architecture to deliver on these customer expectations



Data Quality rules in DataStage flows

What is it?

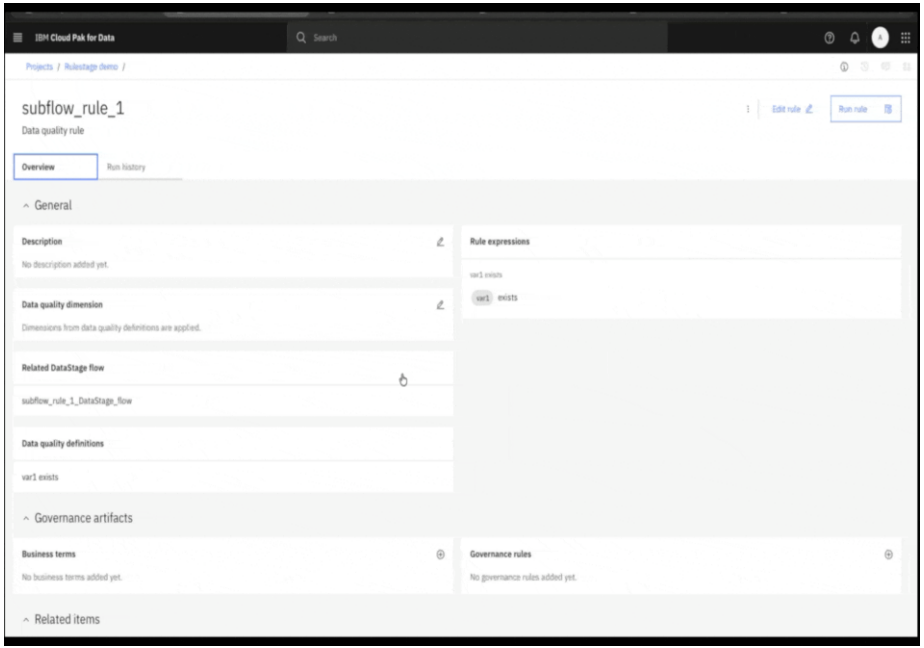
Data Quality rules in DataStage flows is a feature that enables DataStage operators to embed data quality rules directly into one or multiple DataStage flows with a one-to-many relationship.

Key Features:

- **One-to-Many Rule Integration**: Add a single data quality rule to multiple DataStage flows, eliminating the need for creating separate assets (data quality rule, DataStage flow, and subflow) for each rule instance.
- **Simplified Rule Binding**: Provides intuitive information and guidance to help DataStage operators easily integrate data quality rules into flows, including mapping input columns and defining rule variables.
- **Pre and Post-Processing Support**: Enables preprocessing of data (bringing data to the correct format) and postprocessing of results directly within DataStage flows before and after rule execution.
- **Reusable Subflows**: Data Quality Rule stages persist as reusable subflows that can be dragged from the Asset browser or Quality stages section onto the DataStage canvas.
- **Exception Management:** Provides output links to route valid records, invalid records, and detailed exception information to different targets, enabling operators to manage data quality issues within the job flow.
- **Visual Rule Stage Properties**: Displays data quality definitions, input content tables, and rule expressions with variables directly on the stage tab for easy configuration and validation

Why?

Integrating data quality rules directly into DataStage flows reduces asset duplication, boosts operational efficiency, and makes it easier to enforce consistent, high-quality data checks across multiple pipelines. This streamlines collaboration and gives operators the flexibility to validate and manage data exceptions within their workflows



Open Data Contract Standard (ODCS) for Data Quality

What is it?

A **standardized specification** that enables organizations to define **data contracts with embedded data quality rules** using the Open Data Contract Standard (ODCS). This approach allows data producers and consumers to establish clear agreements about data structure, semantics, and quality expectations in a machine-readable format. By adopting ODCS, organizations can automate data quality validation, enhance data governance, and ensure consistent data quality enforcement across the data platform.

Key Features:

- **Standardized Contract Definition**: Define data quality specifications using the industry-standard ODCS specification format.
- **Embedded Data Quality Rules**: Specify data quality rules directly within data contracts specification, including completeness, accuracy, consistency, and custom SQL-based validations.
- **API-Driven Enforcement**: Leverage Data Quality Contracts API to programmatically create, manage, and enforce data quality rules defined in contracts.
- **Version Control & Auditability**: Track contract versions and maintain clear lineage between contract definitions and enforced quality rules with the integration with Data Product Hub (DPH)
- **Multi-Dialect SQL Support:** Execute data quality validations across different SQL engines and data sources within the organization.

Why?

Traditional data quality management lacks standardization, leading to inconsistent quality definitions, manual rule creation, and poor collaboration between data producers and consumers. The Open Data Contract Standard (ODCS) addresses this by providing a common language for defining quality requirements. By embedding data quality within data contracts, organizations can automate quality enforcement, improve data trust, accelerate data product delivery, and establish clear accountability for data quality across teams.

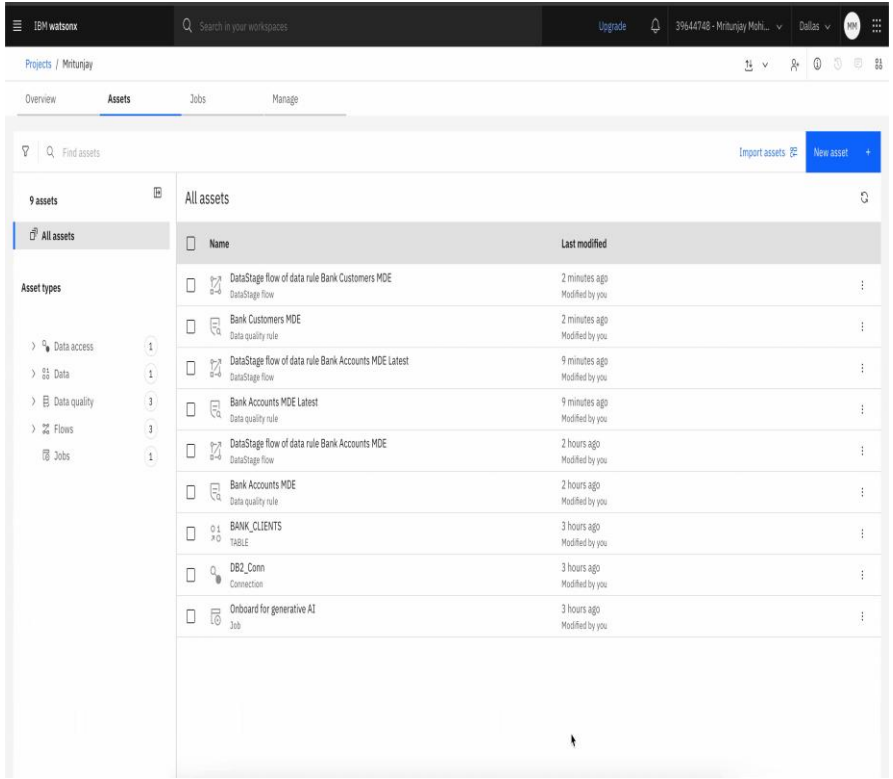
Text2SQL: Automating Complex SQL and Data Quality SQL Rule Creation

What is it?
Empower **business and technical users** to define **complex, SQL-driven data assets and data quality (DQ) rules** using plain English. By leveraging large language models (LLMs) and Text2SQL technology, this approach automatically generates executable SQL—democratizing the development of high-quality data, reducing manual effort, and enhancing governance and auditability.

Key Features:

- **Natural Language Input**: Users specify DQ rules using conversational English.
- **Automatic SQL Generation**: Text2SQL engine converts requirements into validated, actionable SQL.
- **Review & Validation Interface**: Preview, edit, and approve SQL prior to enforcement.
- **Audit & Traceability**: Maintain clear mappings and logs between business rules and generated SQL.
- **Multi-SQL Dialect Support**: Cover all major SQL engines within the organization.

Why?
Traditional data quality (DQ) rules and complex SQL-driven data asset creation require manual coding, which is a slow process. LLMs and Text2SQL technology transform this process by allowing users to define DQ rules and asset creation in plain English, automatically generating executable SQL. This shift enhances accessibility, accelerates delivery, and improves the effectiveness of data quality management.



Natural Language Explanations of Data Quality rules

What is it?

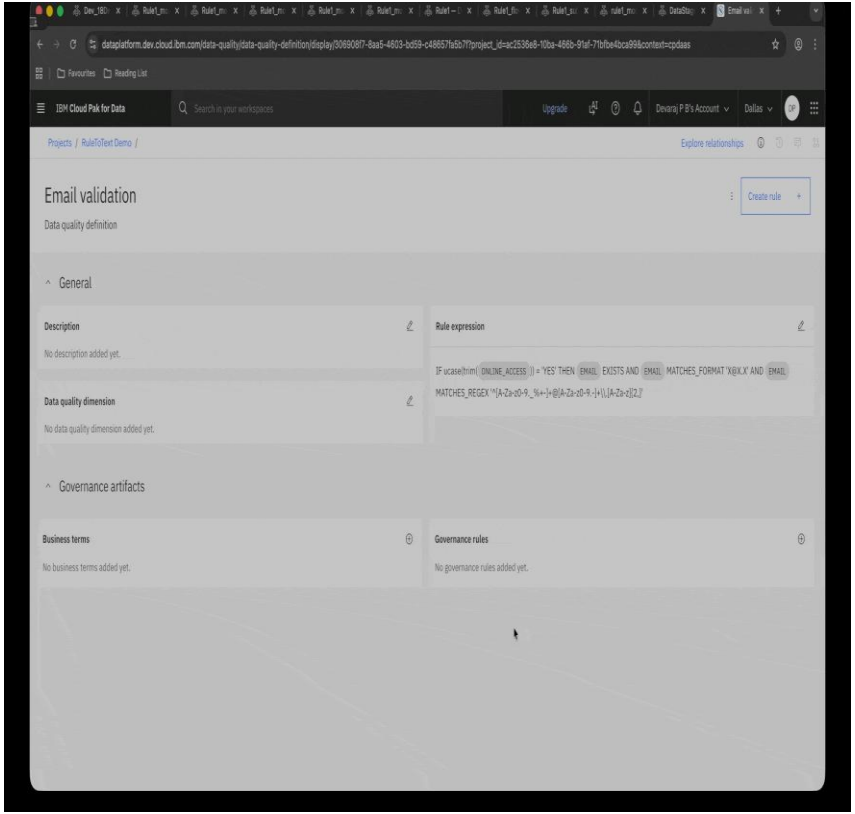
A capability within watsonx.data intelligence data quality that automatically converts technical Data Quality (DQ) rules—whether defined in SQL or business logic—into clear, plain Natural Language explanations. These explanations describe the intent and purpose of each rule in everyday language, enabling business users and non-technical stakeholders to understand, validate, and trust the data quality checks applied to their assets without needing deep technical expertise.

Key Features:

- **Automated Rule Interpretation**: Parses both definition-based and SQL-based DQ rules to extract their logic and intent.
- **Natural Language Generation**: Translates technical rule content into concise, readable descriptions suitable for business audiences.
- **Contextual Integration**: Displays generated explanations directly within DQ rule interfaces, dashboards, and reports for seamless visibility.
- **Customization & Glossaries**: Supports organization-specific terminology so generated descriptions align with internal data language and business concepts.
- **Feedback Loop**: Allows users to refine or override generated explanations, improving precision and alignment with business expectations.
- **Export & Documentation**: Provides options to export rule explanations for governance documentation, audits, and training materials.

Why?

Technical Data Quality rules, often expressed in SQL or complex logic, are hard for non-technical stakeholders to interpret—creating a communication barrier between data stewards, engineers, and business users. This capability bridges that gap by turning opaque rule logic into plain-language statements that articulate what the rule checks and why it matters.



Available only in SaaS

Unstructured Data Curation

What is it?

Unstructured Data Curation in IBM watsonx.data intelligence transforms unstructured documents like invoices, receipts, contracts, and forms using domain specifications, LLM-powered extraction, and automated logic to analyze, classify, extract, and store information as structured tables and vector store for analytics and governance

Key Features

Automated Document Profiling and Classification: Connects to sources, imports document, analyzes samples, and automatically classifies documents by document classes (e.g., invoices, receipts, forms) with statistical distribution analysis.
Predefined Domain Specifications: Includes templates for common document types with LLM extraction instructions, field definitions, and descriptions to guide structured information extraction.
LLM-Powered Extraction: Uses large language models with domain-specific prompts to intelligently extract structured fields from unstructured documents.
Customizable Extraction Schemas: Allows users to design target table schemas, specify field mappings, and define data transformations and normalization rules to fit business requirements.
Dual Ingestion Objectives: Supports both computing embeddings for vector database storage (enabling semantic search) and extracting structured information into relational tables.
Statistical Analysis and Insights: Provides detailed analytics on document characteristics, including language detection, document type distribution, and domain-specific statistics.

Why?

Most enterprise data is trapped in unstructured documents that can't be easily analyzed or governed. Unstructured Data Curation unlocks business value by transforming documents into queryable structured tables, ensures compliance through integrated quality checks and lineage tracking, and scales efficiently using LLM-powered extraction.