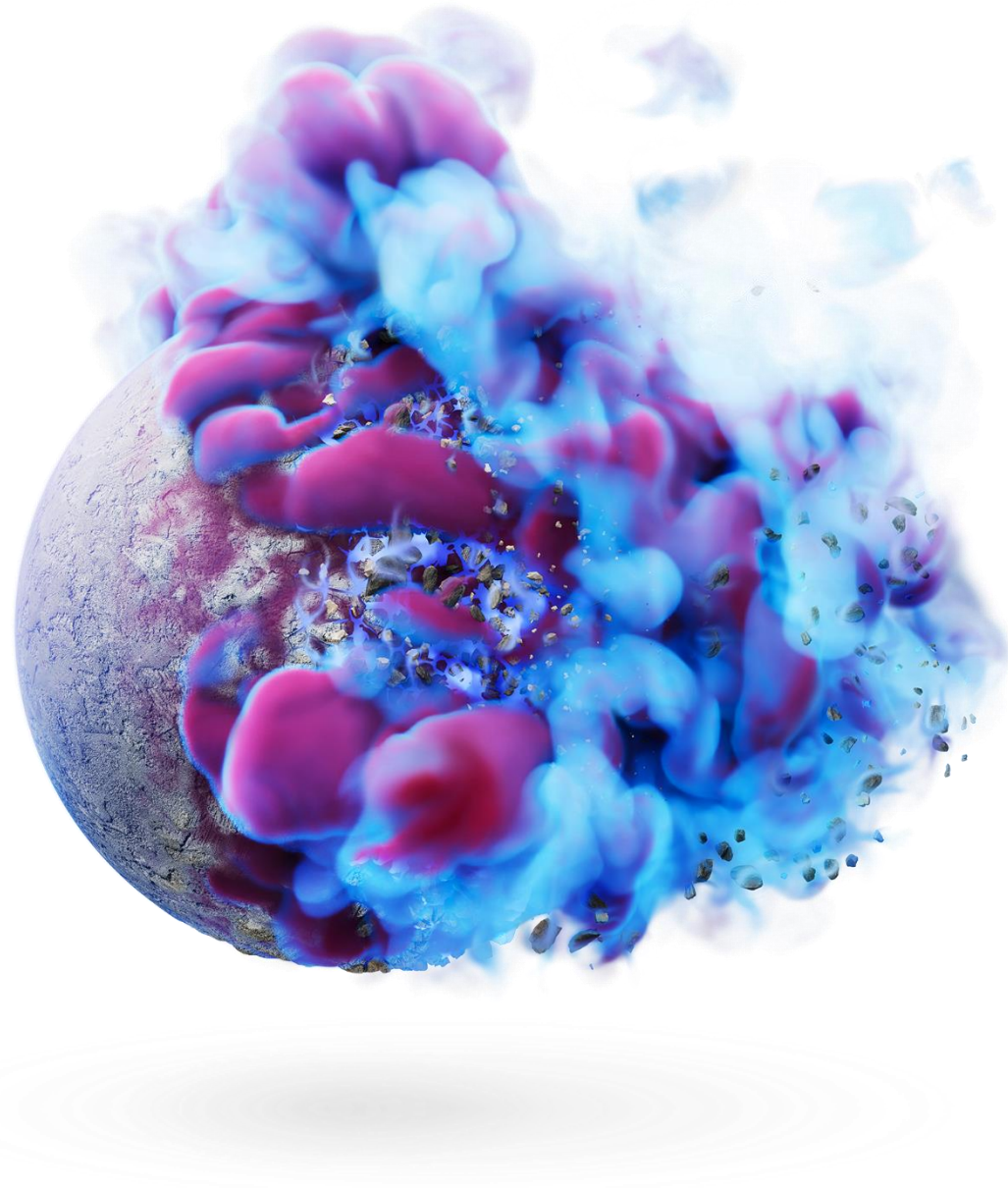
New Times: From Data for AI to AI for Data

Rutger Rienks



Databricks Says AI Agents Now Build 80% Of Enterprise Databases

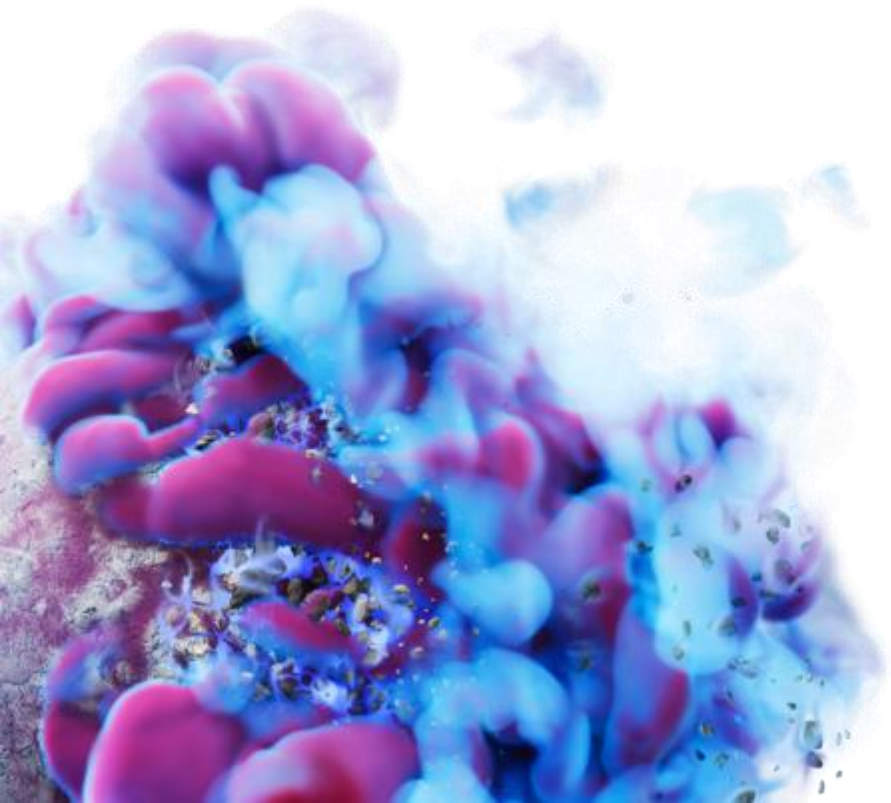
By Victor Dey, Contributor. ⓘ Victor Dey is an analyst and writer covering AI a...

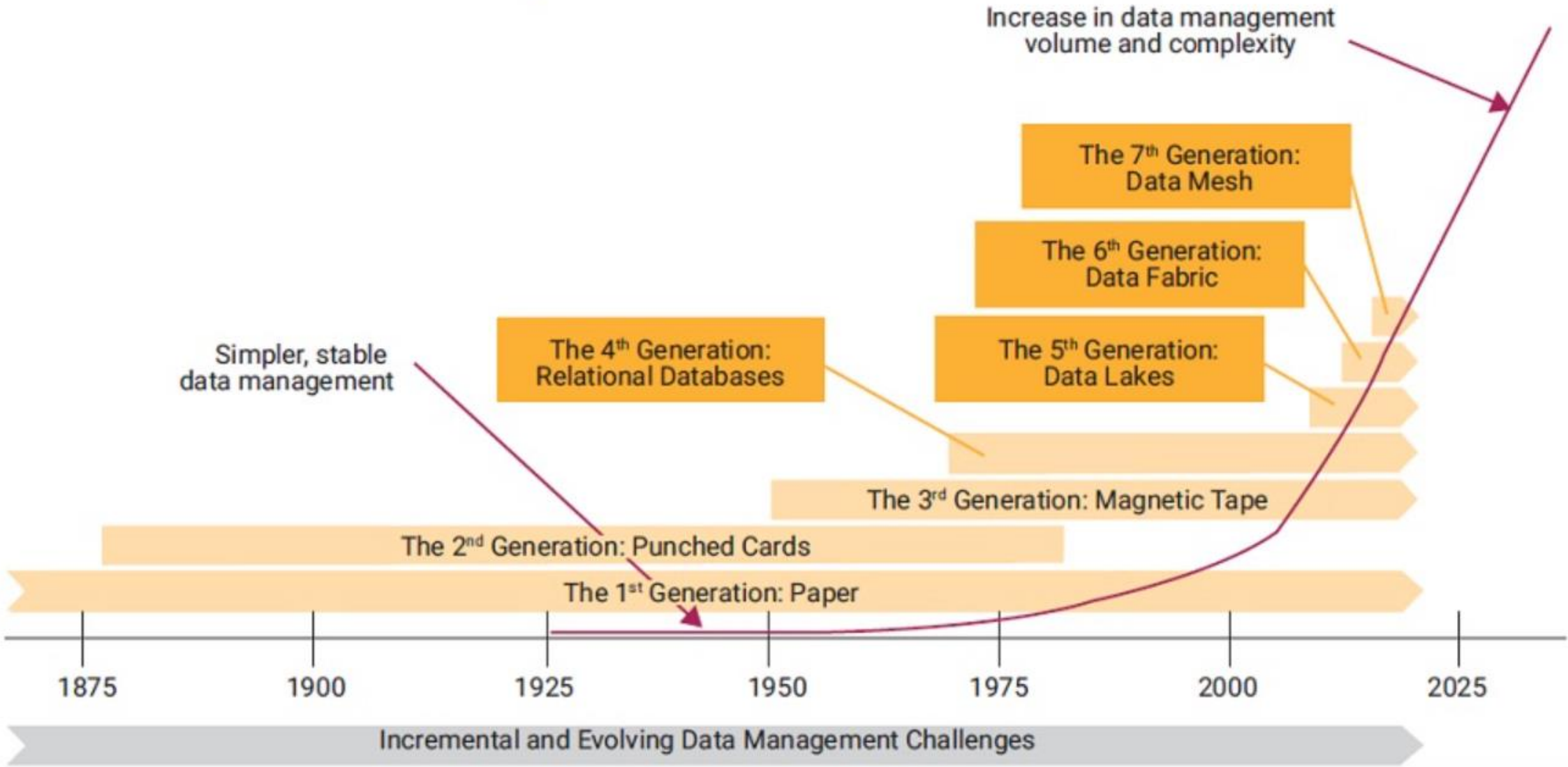
Follow Author

Published Feb 04, 2026, 10:22am EST, Updated Feb 04, 2026, 10:40pm EST

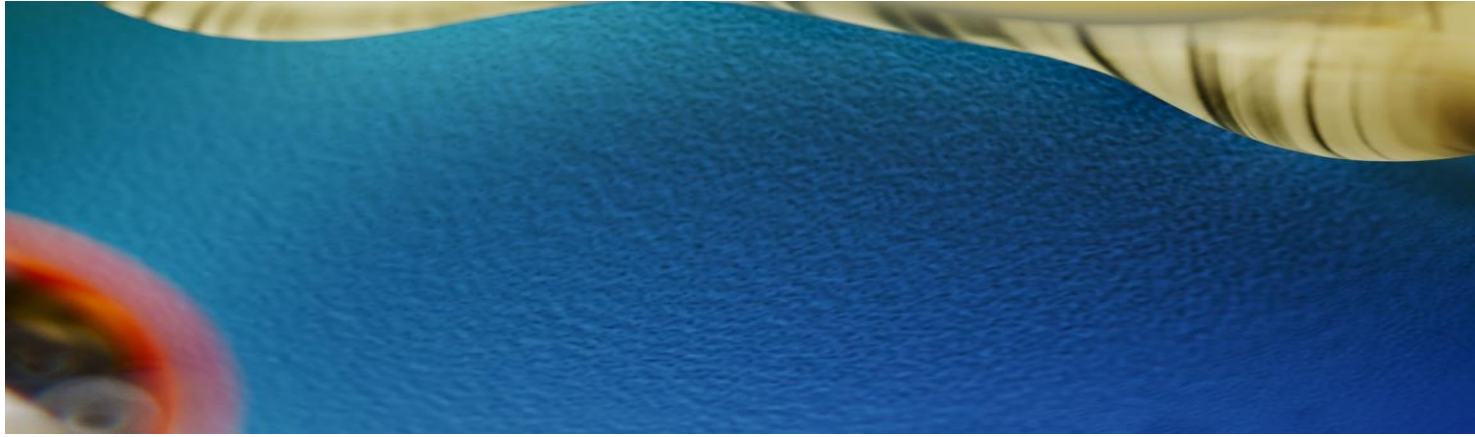
Contents

This presentation outlines the transition from preparing data for AI to using AI to manage data.




Increase in data management volume and complexity
The 7th Generation: Data Mesh
The 6th Generation: Data Fabric
The 5th Generation: Data Lakes
The 4th Generation: Relational Databases
Simpler, stable data management
The 3rd Generation: Magnetic Tape
The 2nd Generation: Punched Cards
The 1st Generation: Paper
1875
1900
1925
1950
1975
2000
2025
Incremental and Evolving Data Management Challenges

Data Cataloguing, Lineage, and Governance Challenges

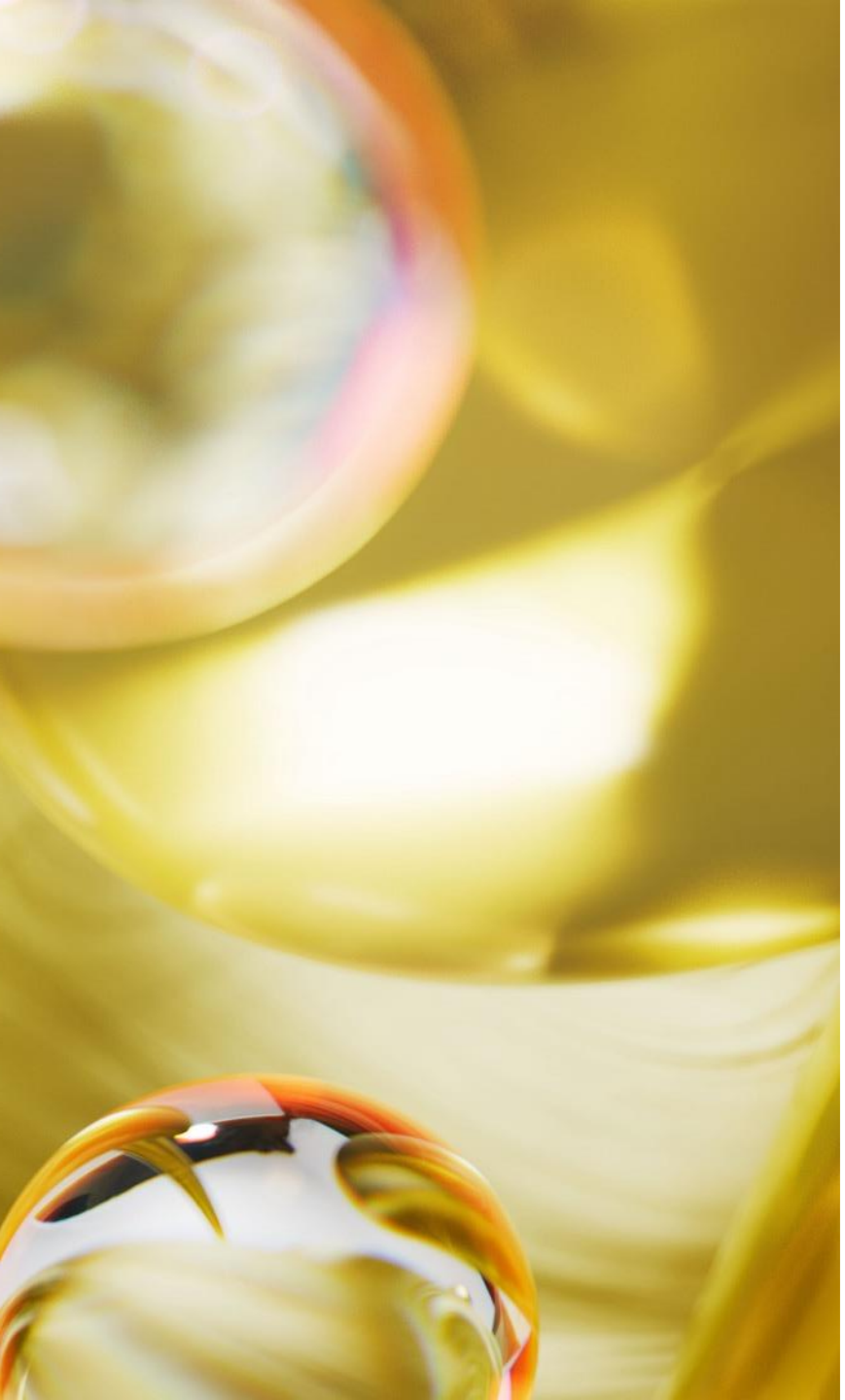


- Data volume and complexity is increasing at a pace that exceeds data governance resource bandwidth to support and govern
- On-demand business analytics and decision-making needs and appetite for high-quality data are increasing
- New technologies and capabilities, such as ML, AI, and GenAI, are generating more data while demanding more high-quality data
- Data stewards who report to business functions are typically not in full-time roles and prioritize business activities over steward activities.
- Data governance organizations are struggling to scale to demand and typically do not have the budget to increase headcount

Organizations should fully embrace automation to scale to demand and expedite Data Cataloguing, Lineage, and governance capabilities

Organization should move towards centralized self-service data & analytics

Organization should build skills in AI/GenAI to integrate these technologies and capabilities into their data governance program



GENERATIVE AI

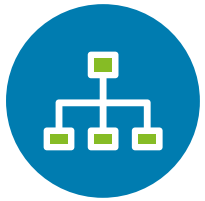
Traditional Generative AI and Large Language Models (LLMs) are **primarily focused on generating and interpreting content**. They excel in tasks such as writing texts, summarizing documents, translating languages, and answering questions based on the context provided. These models respond to specific prompts and perform a single, defined task. Their functionality is reactive; They wait for an input from the user and then produce an output.

AGENTIC AI

An AI agent can **act autonomously to achieve a goal**. This includes creating plans, calling external tools (such as APIs or databases), performing multi-step tasks, and learning from the results. Characteristic of Agentic AI is the use of a memory to retain context over longer interactions and the ability to make decisions independently. Instead of just responding, an agent can proactively initiate a series of actions to solve a complex problem, such as fully managing a data quality process.

Where does AI intersect with Data Management

The integration of AI and Gen AI into data cataloging, lineage, and governance processes offers numerous advantages, including improved accuracy, efficiency, transparency, and compliance. By leveraging these technologies, organizations can enhance their data management practices, making it easier to discover, understand, and govern their data assets effectively.



Data Governance

Includes data classification and data security; supports scalability, data governance, and access control, data encryption, and continuous monitoring, making data more manageable, secure, and useful

Data Quality

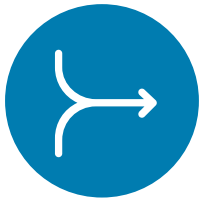
Ensures data accuracy, consistency, and reliability, which are crucial for effective decision-making and operational efficiency.

Data Lineage

Provides a comprehensive and accurate view of data flows and transformations, enabling better decision-making and risk management.

Metadata Management

Allows organizations to achieve more efficient, accurate, and scalable metadata management, ultimately leading to better data utilization and decision-making.

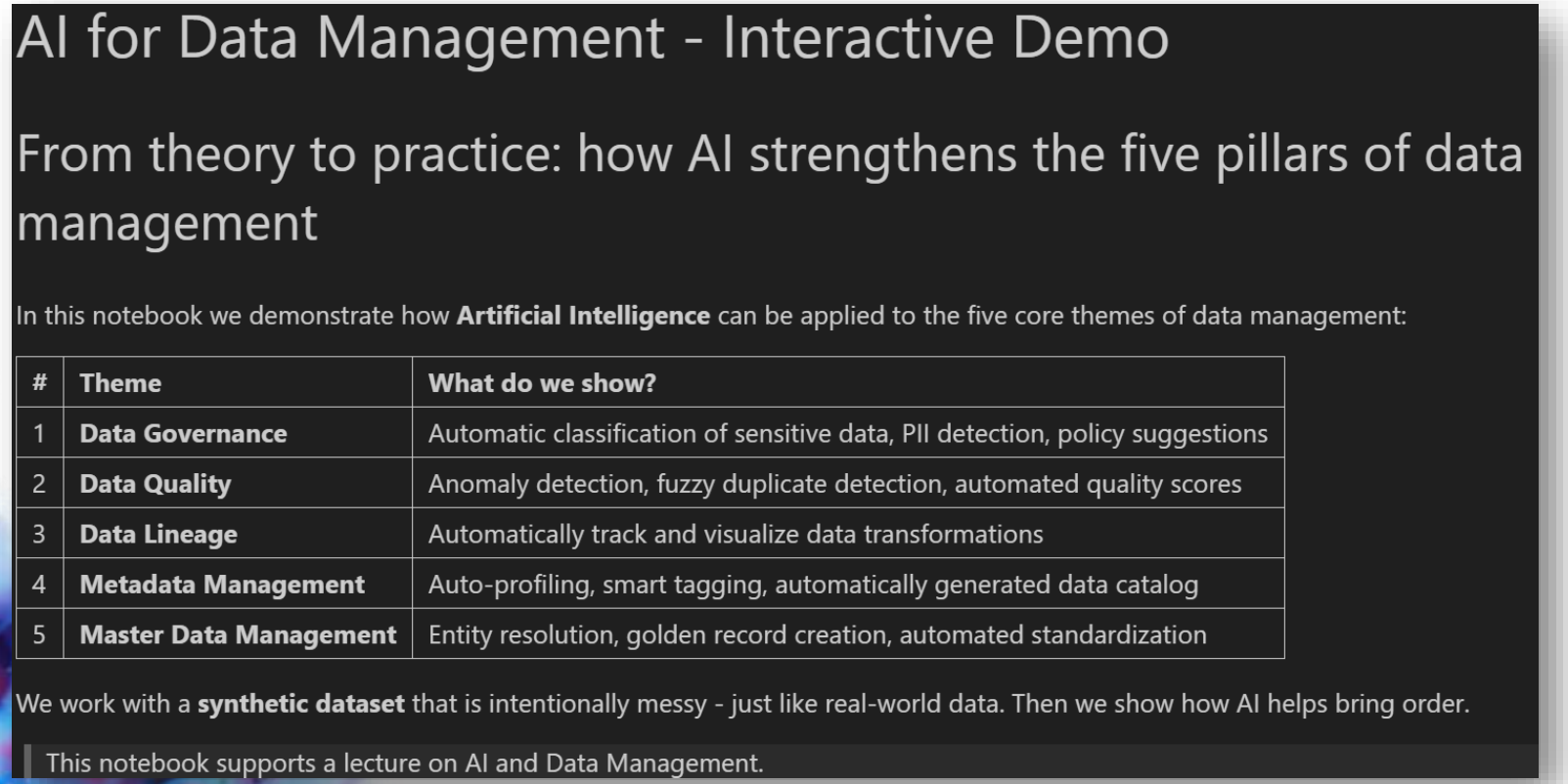


Master Data Management

Supports data enrichment, identifies duplicates, enforces policies, increasing operational efficiency and ensuring accurate and consistent master data

Let’s make it Interactive!

Download the notebooks to follow along the presentation



AI for Data Management - Interactive Demo

From theory to practice: how AI strengthens the five pillars of data management

In this notebook we demonstrate how **Artificial Intelligence** can be applied to the five core themes of data management:

#	Theme	What do we show?
1	**Data Governance**	Automatic classification of sensitive data, PII detection, policy suggestions
2	**Data Quality**	Anomaly detection, fuzzy duplicate detection, automated quality scores
3	**Data Lineage**	Automatically track and visualize data transformations
4	**Metadata Management**	Auto-profiling, smart tagging, automatically generated data catalog
5	**Master Data Management**	Entity resolution, golden record creation, automated standardization

We work with a **synthetic dataset** that is intentionally messy - just like real-world data. Then we show how AI helps bring order.

> This notebook supports a lecture on AI and Data Management.

https://github.com/rutgerrienks/Datamanagement

Data Governance

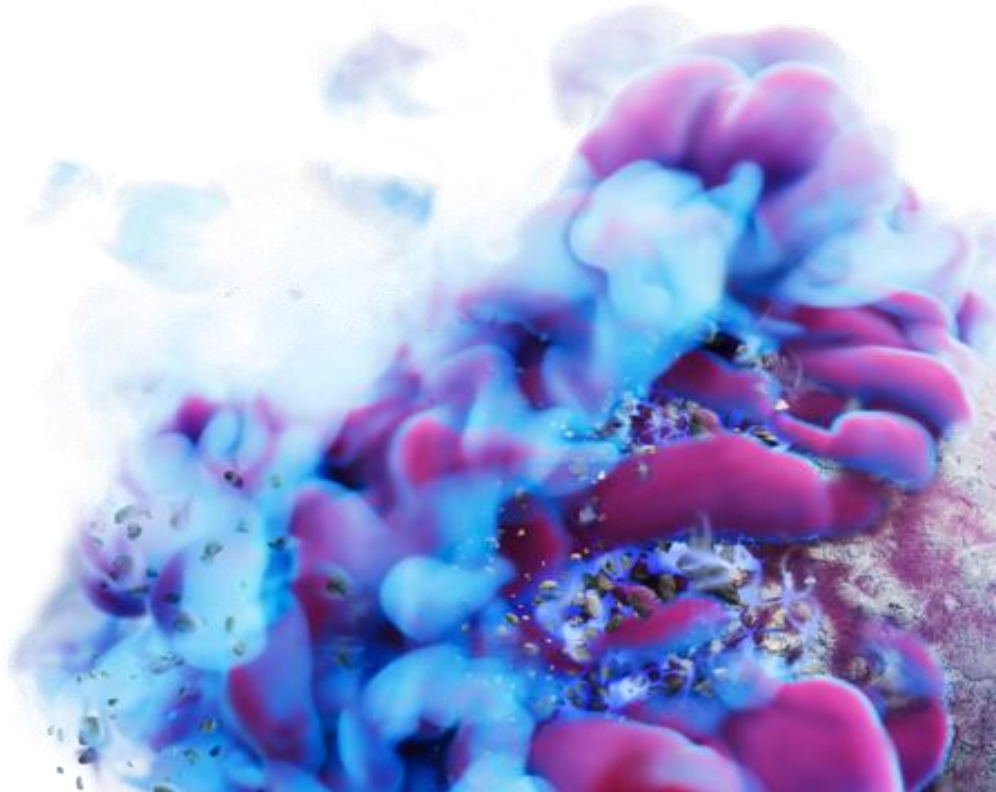
The framework for managing data assets within an organization to ensure data quality, security, integrity, and proper classification, thereby enabling effective data usage, compliance, and protection against unauthorized access.

Key Data Governance Capabilities:

- **Data Classification Standards:** Tag data with labels such as Public, Confidential, and Personal Information.
- **Access Control:** Ensuring users have data on a need-to-know basis and verifying identity and permissions before granting access
- **Data Stewardship:** Responsibly managing and overseeing an organization's data assets to ensure data quality, integrity, security, and compliance

How AI/GenAI Enhances Data Governance:

- **Automates Data Classification:** AI/GenAI can classify data and apply labels based on predefined criteria, reducing the time and effort required for manual classification and monitoring and classifying new data in real time
- **Assists Data Stewards:** AI/GenAI can automate data management workflows and processes and enable proactive monitoring and compliance through advanced analytics



Data Lineage

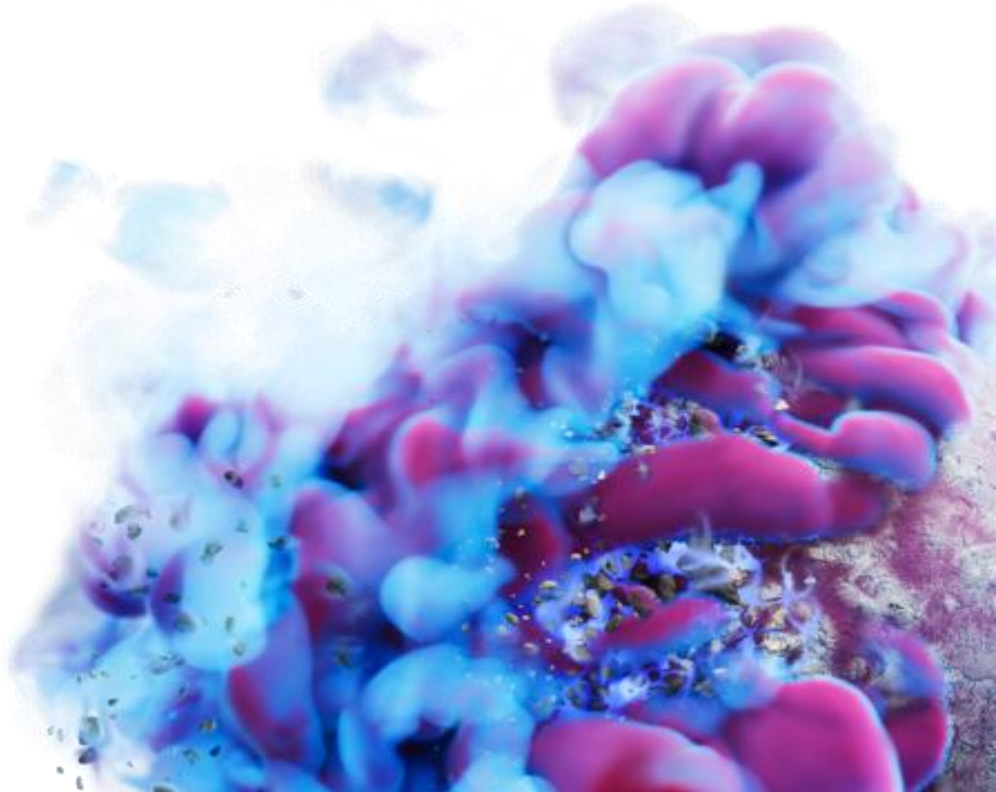
The lifecycle of data, encompassing its origins, movements, transformations, and ultimate destination. It provides a detailed map of how data flows through an organization, from its initial creation or acquisition to its final use

Key Data Lineage Capabilities:

- **Data Lineage:** Ability to trace data from its origin to its destruction
- **Data Compliance & Anomaly Detection:** Ensuring data meets regulatory compliance at each step of the data lifecycle, with AI-driven anomaly detection to mitigate risk and operational disruption.

How AI/GenAI Enhances Data Lineage:

- **Contextual Data Lineage Visualization:** AI-powered tools offer customizable, interactive visualizations of data lineage, enhancing understanding and decision-making by illustrating data flows and transformations across the organization.
- **Advanced Anomaly Detection:** AI algorithms can detect anomalies and outliers in data lineage, such as unexpected data transformations or unauthorized data access.



Metadata Management

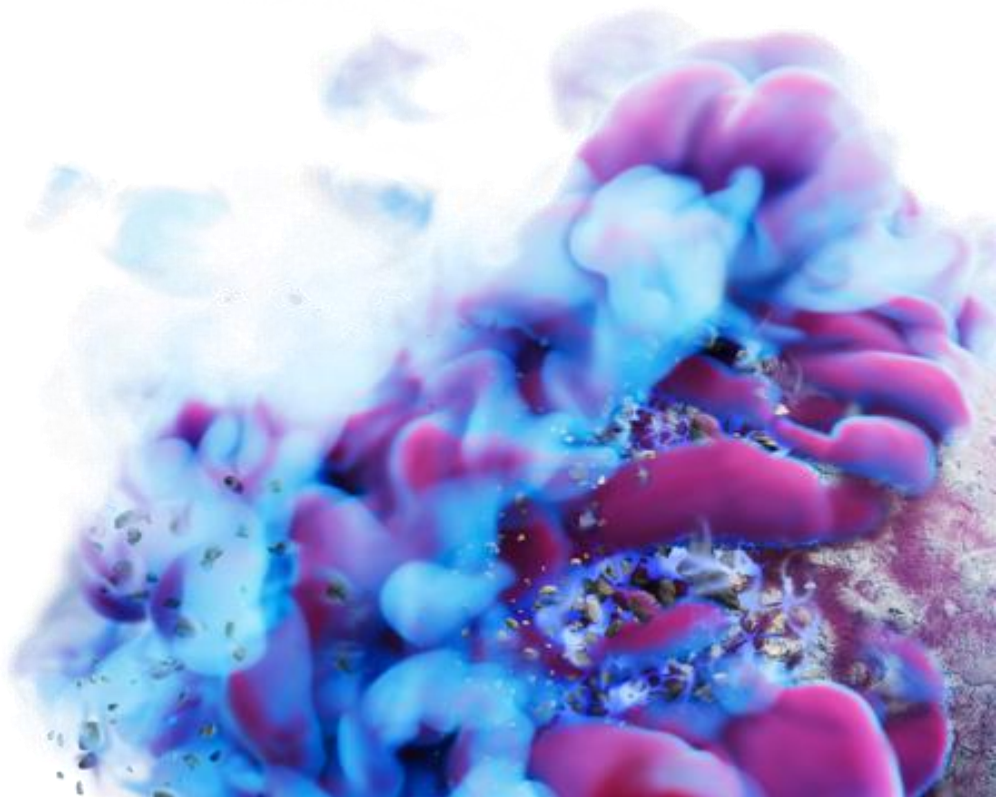
The capability to proactively manage the diverse data standards of the enterprise. It involves creating, controlling, enhancing, and maintaining consistency, quality, and usability of metadata

Key Metadata Management Capabilities:

- **Data Discovery**: Facilitates the business' ability to quickly find data assets based on metadata attributes
- **Data Collaboration**: Improves communication and understanding by providing context about data and establishing standards across data areas
- **Impact Analysis:** Asses the potential effects of changes in data definitions or structures on downstream processes and systems

How AI/GenAI Enhances Metadata Management:

- **Ease of Search:** AI supports search capabilities using natural language processing (NLP), giving users a simple way to locate data assets
- **Predictive Impact Analysis:** AI can predict the effects of changes in metadata on downstream processes by analyzing historical data and identifying patterns



Data Quality

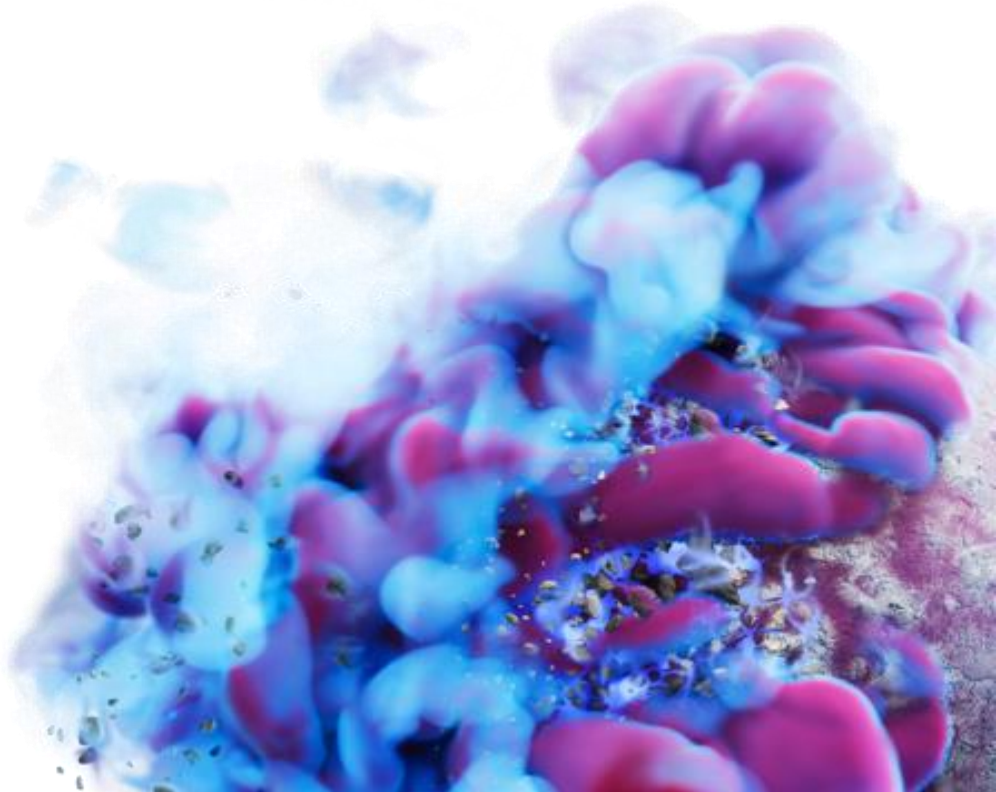
Data Quality (DQ) refers both to the characteristics of high-quality data and to the processes used to measure or improve the quality of data

Key Data Quality Capabilities:

- **Data Profiling:** Identify potential data quality issues and develop solutions for remediation
- **Data Cleansing:** Correct or remove inaccurate records from a dataset
- **Data Quality Tracking:** Track the success and opportunities with metrics reflecting the progress of data quality efforts

How AI/GenAI Enhances Data Quality:

- **Automate Data Cleansing and Data Preparation:** AI/GenAI can identify and correct errors in datasets. This could include duplicate detection/removal, error correction, and general standardization
- **Data Quality Rule Creation:** Generative AI can help users write DQ rules so organizations can deploy and monitor data in a faster timeframe



Master Data Management

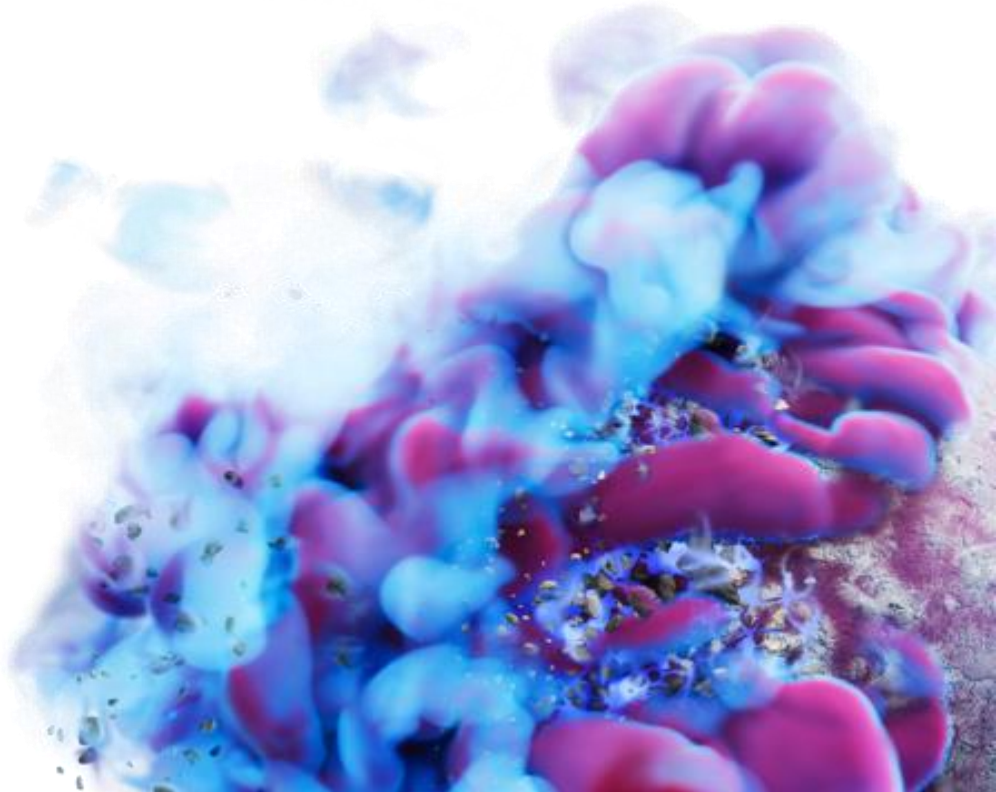
Master Data Management (MDM) is the comprehensive process of managing an organization's critical data to ensure consistency, accuracy, and accountability across the enterprise and creating a single source of truth.

Key Master Data Management Capabilities:

- **Data Quality Standardization**: MDM facilitates the creation of a centralized and consistent definition of attributes, ensuring all data elements adhere to a consistent set of standards and dimensions (parsing, match/merge, profling, enrichment)
- **Master Data Lineage**: Ability to trace and map the origins, movement, transformations and usages of Master Data

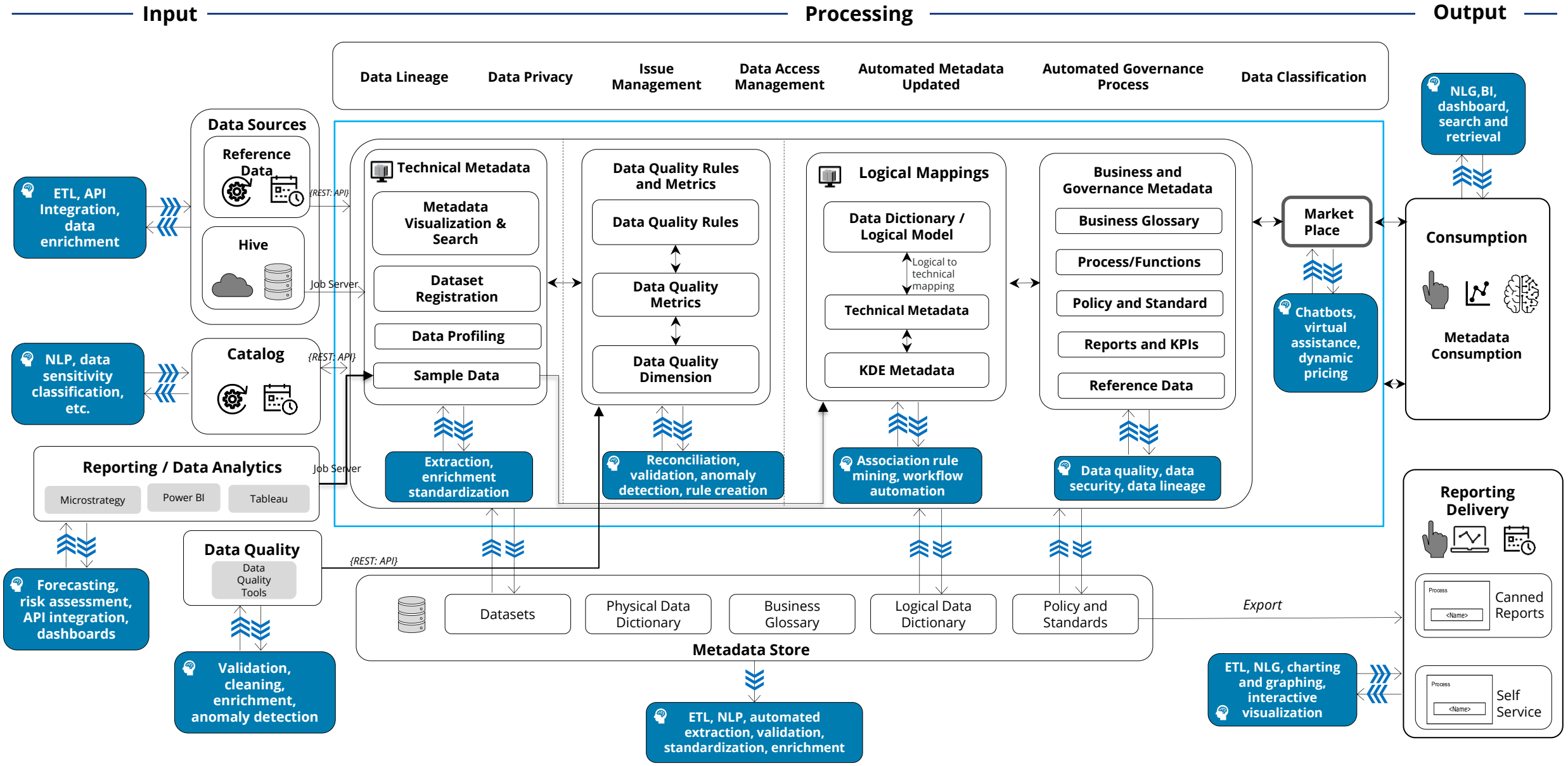
How AI/GenAI Enhances Master Data Management:

- **Automated DQ Operations:** Built in AI can detect and remove duplicates, augment and enrich hierarchies with third-party data, check for consistency and correct missing data to prevent manual intervention.
- **GenAI Powered Data Operations:** Gen/AI capabilities can help automate data tasks and create semantic drill downs of data for exploration and analytics



AI Agent Technical Conceptual Architecture

Indicates where the AI agent interacts with a specific part of the architecture.



NOTE: Blue boxes do not represent a 1:1 mapping; multiple AI agents can be implemented within each capability or area.

Tool Overview

Different data governance tools have similar overall functions but offer various capabilities and features.

The commonly used tools listed below offer a comprehensive and collaborative platform for data governance. They help automate tasks such as data lineage, data cataloging, and data discovery. They are ideal and popular choices for organizations looking for a complete data management suite that ensures strong data quality, effective metadata management, and scalability.

Collibra

Collibra's data governance platform offers automated data cataloging, end-to-end data lineage, robust data quality management, and policy enforcement. It supports role-based access, stewardship workflows, and provides collaboration tools and a centralized business glossary. With customizable dashboards and integration capabilities, Collibra ensures data accuracy, compliance, and efficient decision-making across the organization.

Alation

Alation's data governance capabilities include automated data cataloging, data lineage tracking, and data stewardship workflows. It provides robust metadata management, policy enforcement, and role-based access control. Alation also offers collaboration tools, a business glossary, and data quality management features to ensure data accuracy, compliance, and effective decision-making.

Informatica

Informatica's data governance capabilities include automated data discovery, comprehensive metadata management, and robust data quality tools. It offers end-to-end data lineage, policy enforcement, and role-based access control. The platform supports collaboration through a business glossary and stewardship workflows, ensuring data accuracy, compliance, and effective decision-making across the organization.

●●●/●●○/●○○ = High / Medium / Low Maturity

Tool Overview

Use Case	Collibra	Alation	Informatica	AI Capabilities for Each Tool
Metadata Management	●●○	●●○	●●●	• **Collibra**: Supports automated data cataloging, automated asset curation, and NLP search capabilities. • **Alation**: Uses Behavior I/O, an AI-powered engine to automatically identify meaningful patterns about how data Us being used, then giving users recommendations. • **Informatica**: Uses active metadata management to create predictive recommendations, supports automated data cataloging and NLP search.
Data Quality	●●○	●○○	●●●	• **Collibra**: Has AI/GenAi-generated AdaptiveRules, supporting automated DQ tracking, and GenAI supported DQ rule creation using NLP. • **Alation**: Integrates Snowflake's Data Quality Metric Functions (DMFs) with its data quality framework, creating high-quality data for AI model development. • **Informatica**: Has CLAIRE AI copilot, which supports automated data profiling, quality analysis, and can generate data quality rules.
Data Lineage	●●●	●●●	●●●	• **Collibra**: Has intelligent filtering through lineage diagrams and automated end-to-end data lineage tracking with Collibra Lineage. • **Alation**: Leverages machine learning algorithms analyze user behavior, data usage patterns, and metadata to identify relevant data assets. • **Informatica**: Supports automated end-to-end data lineage tracking, including automated scanning for data transformation details.
Data Classification	●●●	●●●	●●●	• **Collibra**: Can automatically classify data with industry-specific context • **Alation**: Supports automated mapping and categorization of data, including tagging • **Informatica**: CLAIRE supports intelligent similarity detection and domain discovery with tags, streamlining classification
Data Security	●●○	●●○	●●●	• **Collibra**: Automatically discovers and classifies sensitive data across existing organizations' AI systems, allowing owners to manage access controls • **Alation**: Alation AI and data governance framework flags non-compliant datasets and ensures AI models are built on reliable, governed data • **Informatica**: CLAIRE has data privacy solutions that enable sensitive data tracking, detect anomalies, and simulate risk
Master Data Management			●●●	• **Collibra**: N/A • **Alation**: N/A • **Informatica**: Automatically identifies master data, traces source lineage, and finds mappings between attributes in data models

The three main conclusions

KEY MESSAGES SUMMARIZED

The focus is shifting from manual data preparation to leveraging AI for autonomous data management

Agentic AI significantly increases efficiency by deploying autonomous agents to enrich, validate, and integrate data.

Strong governance and Human-in-the-Loop controls are essential to manage risk and ensure the reliability of the AI.