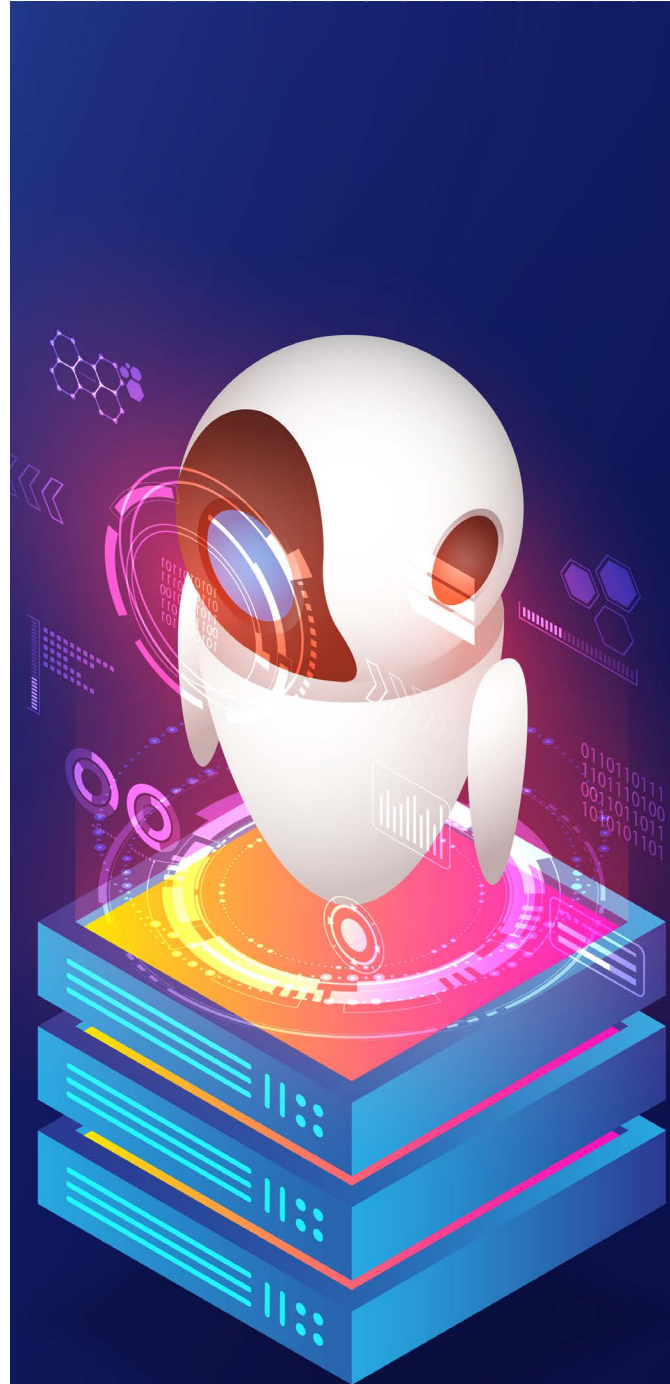




2025 WHITE PAPER

Why a Metadata Foundation Matters in Data Analytics and AI

Metadata plays a central role in enterprise data organizations, both in traditional data analytics and modern AI initiatives. In analytics, technical metadata is how organizations map their data estate for business and data science teams. Metadata also provides the supporting context needed to derive other critical artifacts such as data lineage, quality, and ownership.



2025 White Paper

Why a Metadata Foundation Matters in Data Analytics and AI

2025 White Paper

Why a Metadata Foundation Matters in Data Analytics and AI

As organizations scale their AI efforts, metadata becomes even more essential. The success of AI depends on the quality and structure of the data it consumes. Most of this data is unstructured, complex, and difficult to manage without a strong metadata foundation.

Traditional Challenges in Metadata Collection, Management, and Governance

Historically, metadata collection, management, and governance have been manual and process-heavy.

Technical metadata, such as column names, was typically harvested from structured sources like SQL databases and cloud data warehouses including Snowflake. This was done through manual onboarding, and the metadata was then stored in catalogs for search and analysis.

Data stewards were responsible for mapping this metadata to business definitions in a glossary, assigning sensitivity levels, flagging quality issues, and managing access controls. These tasks were performed through manual review processes and required rigid approval workflows.

This top-down approach made it difficult to scale and maintain consistency. Human reviewers had to interpret metadata manually, which introduced delays and inconsistencies. The process was inefficient and could not meet the needs of modern data environments.

AI introduces a new approach by serving as both an assistant and an accelerator. It brings machine-driven verification, intelligent inference, and scalable automation to transform metadata management across analytics and AI initiatives.

How AI Transforms Metadata Collection, Management, and Governance

AI plays a key role in every stage of metadata management, from collection and cataloging to governance and discovery. This is especially important for unstructured data, which is the primary data type used in many AI workloads.

1. AI-Powered Metadata Collection

AI enhances metadata harvesting in several important ways:

- **Schema-Agnostic Onboarding:** AI allows for the automatic discovery of data sources, whether structured or unstructured, without requiring rigid schemas. BigID supports fully automated discovery and detects changes in data sources continuously.
- **Automatic Contextualization:** AI can translate technical field names into business-friendly terms that are easier to understand across teams. BigID uses AI to classify data and generate friendly names to improve usability.
- **Glossary Mapping Automation:** Mapping technical terms to business glossaries is often inaccurate and time-consuming when done manually. AI improves both speed and accuracy. BigID is the first vendor to provide this capability at scale.

2. AI-Driven Cataloging and Discovery

Cataloging metadata is not new, but AI improves how metadata is stored, accessed, and scaled:

- **Scalability:** Traditional metadata catalogs are often limited to a few million objects. BigID's AI-driven catalog scales to tens of billions of objects.
- **Semantic Search:** AI vectorization enables search based on meaning rather than just keywords. This allows users to find relevant data even when metadata is incomplete or mislabeled. BigID is the first metadata catalog to combine vectorized semantic search with enterprise-grade scalability.

3. AI-Enhanced Governance and Stewardship

AI supports data stewards by automating tasks and providing deeper insights:

- **Ownership and Provenance Inference:** AI can determine data ownership, detect its source and creation date, and identify the business domain associated with the data.
- **Data Summarization and Quality Insights:** AI can summarize datasets, surface high-quality alternatives, and detect issues that would be difficult or time-consuming for human reviewers to identify. BigID integrates these capabilities into its platform to support smarter governance.

4. Metadata Management for Unstructured Data

Unstructured data includes content such as code, chat messages, emails, and documents. These data types are difficult to manage manually due to their size and complexity. AI makes metadata management for unstructured data possible by:

- Labeling and tagging content
- Categorizing data by business taxonomies
- Classifying by sensitivity and topic
- Identifying bias, defects, or anomalies
- Generating summaries
- Finding related or alternative data sets for training AI models

BigID is the first vendor to offer full metadata cataloging for unstructured data and provide AI assistants that support governance across the entire lifecycle from collection to AI preparation.

5. Inference for Lineage and Data Quality

AI is highly effective at identifying patterns and making inferences across large datasets. This makes it ideal for:

- **Data Lineage:** Understanding how data moves and transforms across systems.
- **Data Quality:** Comparing datasets to identify errors, anomalies, or areas for improvement.

BigID makes it practical to capture and use this metadata insight at enterprise scale.

The BigID AI Advantage in Metadata Management

- BigID is redefining how organizations collect, manage, and govern metadata across all types of data. The platform delivers:
- Automated, schema-agnostic metadata discovery
- Business-contextualized metadata through AI-generated naming
- Scalability to tens of billions of metadata objects

- Semantic search powered by vectorization
- Integrated AI assistants for governance and stewardship
- Full metadata coverage for unstructured data
- AI-powered insights into lineage and data quality

Summary: Metadata and AI Are Foundational to Modern Data Strategy

As organizations invest in analytics and AI, metadata is emerging as a strategic requirement. Traditional metadata management approaches cannot keep up with the scale, complexity, and speed of modern data environments.

BigID provides a modern foundation for metadata management. By using AI to automate discovery, classification, governance, and enrichment, BigID empowers organizations to gain deeper insights, improve AI outcomes, and take control of their data at scale.

About BigID

BigID enables security, compliance, privacy, & AI data management for all data, everywhere. BigID is enterprise-ready and built to scale: enabling a data-centric approach to comprehensive cloud data security & DSPM, accelerating compliance, automating privacy, and streamlining overnance. Customers deploy BigID to proactively discover, manage, protect, and get more value from their regulated, sensitive, and personal data across their data landscape.

BigID has been recognized for innovation as a World Economic Forum Technology Pioneer; named to the Forbes Cloud 100; the Inc 5000 for 4 consecutive years; the Deloitte 500 for 4 consecutive years; Market Leader in Data Security Posture Management (DSPM); Leader in Privacy Management in the Forrester Wave; and an RSA Innovation Sandbox winner.

Find out more at <https://bigid.com>.

Know Your Data, Control Your Data.

Data Security • Compliance • Privacy • AI Data Management

Reduce risk, accelerate time to insight, and get data visibility and control across all your data - everywhere.

“Tools like BigID are the future.

Organizations should be leveraging these tools to remove the manual processes from data discovery, provide better visibility, and help with prioritization of controls.



Ryan O'Leary
Future of Trust: Battling Data Discovery Confusion