

Key Questions for your Data Catalog RFP

Navigate data catalog selection with precision and foresight



Table of Contents

Introduction

<u>Scalability</u>

Extensibility

Unification

Common Language

<u>Al-Ready</u>

<u>Security</u>

Future-Proof

Implementation and Support

Cost and Value

<u>Vendor Stability</u>

Introduction

In today's data-driven landscape, organizations face unprecedented challenges in managing, understanding, and extracting value from their growing data assets. A robust data catalog has become essential infrastructure for modern enterprises seeking to transform raw information into actionable intelligence while maintaining proper governance and compliance.

However, not all data catalogs are created equal. As the market has evolved, significant differences have emerged in capabilities, approaches, and underlying architectures. What might have been sufficient just a few years ago may now leave critical gaps in your data management strategy.

This buyers' guide aims to equip you with the precise questions needed to evaluate data catalog vendors thoroughly. By diving deep into key areas—from scalability and extensibility to security and Al-readiness—you'll be able to identify solutions that not only meet your current requirements but can also evolve alongside your organization's data journey.

Whether you're replacing an existing catalog that's reaching its limits, implementing your first formal data catalog solution, or simply reassessing your current vendor relationship, the questions in this guide will help you navigate vendor conversations with confidence and clarity.

Use this guide to cut through marketing promises and technical jargon to determine which solutions can truly deliver on the promise of unified data intelligence for your organization. Your choice of data catalog will fundamentally shape how your teams discover, understand, trust, and leverage data for years to come—making this evaluation process a critical investment in your data future.

1. Scalability Questions

Comprehensive Source Coverage

- → "What data sources do you support out-of-the-box? List by type (databases, warehouses, lakes, BI, ETL, etc.)."
- → How do you handle metadata extraction from legacy systems and proprietary platforms?
- → What is your roadmap for adding new connectors, and how do you prioritize which sources to support?
- → How frequently are connectors updated to accommodate changes in source systems?

Connector Framework

- → Do you offer a framework for building custom connectors? How is it documented?
- → What technical expertise is required to develop custom connectors?
- → Can custom connectors be maintained through platform upgrades without redevelopment?
- → Do you offer professional services to develop custom connectors? At what cost?

Metadata Volume

- → What is the maximum volume of metadata your platform can handle? Please provide specific metrics (number of tables, columns, lineage relationships, etc.).
- → How does performance degrade as metadata volume increases?
- → What architectural features ensure performance at scale?
- → Can you provide reference customers with metadata volumes similar to our environment?

Metadata Velocity

- → What is the minimum latency between a change in a source system and its reflection in your platform?
- → Do you support real-time metadata updates? How are these implemented?
- → How do you handle high-frequency metadata changes from sources like streaming platforms?
- → What is the performance impact of enabling real-time updates?

2. Extensibility Questions

Metadata Model Flexibility

- → How can we extend your base metadata model to accommodate our organization's unique requirements?
- → Can we add custom attributes to standard entity types? Does this require coding or configuration?
- → How are custom extensions preserved during platform upgrades?
- → Can we define custom relationship types between entities?
- → How do custom extensions affect search, lineage, and other platform capabilities?

APIs and SDKs

- → Please describe your API architecture and coverage. Are all platform functions accessible via API?
- → Which programming languages do you provide SDKs for?
- → How are your APIs versioned, and what is your deprecation policy?
- → Do you have rate limits or other restrictions on API usage?
- → Can you provide examples of customers who have built significant extensions using your APIs?
- → Is your API documentation publicly available? Can we review it before purchase?

3. Unification Questions

Integrated Capabilities

- → How does your platform unify discovery, observability, and governance capabilities?
- → Are these capabilities truly integrated or separate modules with different user experiences?
- → Do all capabilities operate on the same underlying metadata model?
- → How does information flow between different functional areas of the platform?
- → Can you demonstrate end-to-end workflows that cross traditional boundaries (e.g., from discovery to governance)?

Metadata Graph

- → How does your platform model and store relationship information between data assets?
- → What types of lineage do you capture (technical, business, process, etc.)?
- → How granular is your lineage tracking (column-level, table-level, etc.)?
- → Can your lineage visualization handle complex graphs with thousands of nodes?
- → How do you maintain lineage accuracy when source systems change?
- → Can we extend the lineage model with custom relationship types?

4. Common Language Questions

Persona Support

- → Which specific user personas does your platform support?
- → How does the user experience differ for technical versus business users?
- → Can we customize the experience for different roles within our organization?
- → How do you handle terminology differences between technical and business contexts?
- → Can you demonstrate the platform from the perspective of each key persona?

Collaboration Features

- → What capabilities support collaboration between different user personas?
- → How do you facilitate communication between data owners and data consumers?
- → What workflows exist for activities that cross persona boundaries (access requests, issue reporting, etc.)?
- → Do you provide notification mechanisms when relevant assets change?
- → Can users provide feedback and enrichment to improve metadata quality?

5. Al-Ready Questions

AI/ML Asset Support

- → How does your platform catalog ML models, features, and experiments?
- → What specialized metadata do you capture for AI/ML assets?
- → How do you track lineage between training data, models, and model outputs?
- → Do you integrate with common ML platforms (Databricks, Sagemaker, etc.)?
- → How do you address governance requirements specific to AI systems?

AI-Enhanced Metadata Management

- → How do you apply AI to improve the platform itself?
- → What automated classification and tagging capabilities do you offer?
- → How accurate are your AI-powered metadata enrichment features?
- → Do you use AI for anomaly detection or quality monitoring?
- → What is your roadmap for expanding AI capabilities within the platform?

Al Integration

- → How can AI systems and agents interact with your platform?
- → Do you support retrieval-augmented generation (RAG) patterns for LLMs?
- → What mechanisms exist for AI systems to update metadata?
- → How do you ensure appropriate governance when AI systems access metadata?
- → Can you describe any customers using your platform as part of an AI strategy?

6. Security Questions

Metadata Ingestion Security

- → How does your architecture ensure secure metadata collection?
- → What deployment options exist for high-security environments?
- → How are credentials for source systems managed and protected?
- → Do metadata collectors require full access to source systems or read-only access?
- → How is sensitive metadata protected during collection and transport?

Deployment Options

- → What deployment models do you support (SaaS, private cloud, on-premises)?
- → For cloud deployments, which cloud providers are supported?
- → Can we deploy in our own VPC/VNET with complete network isolation?
- → How do you handle hybrid deployment scenarios?
- → What is the performance impact of different deployment models?

Security Standards

- → What security certifications does your platform hold?
- → How frequently do you conduct security assessments and penetration testing?
- → What encryption standards do you employ for data at rest and in transit?
- → How do you handle security vulnerabilities and patches?
- → Can we review your security documentation and audits?

7. Future-Proof Questions

Innovation Track Record

- → How frequently do you release major platform updates?
- → What significant innovations have you introduced in the past 12 months?
- → How do you gather customer input for roadmap planning?
- → What percentage of your revenue is invested in R&D?
- → How do you balance innovation with platform stability?

AI Transformation Vision

- → What is your vision for how AI will transform metadata management?
- → How are you positioning your platform to leverage emerging AI capabilities?
- → What autonomous metadata management capabilities are on your roadmap?
- → How will you help customers transition to AI-driven data management?
- → Are you exploring generative AI integration for metadata creation and enrichment?

Standards Adoption

- → Which industry metadata standards do you support?
- → How do you participate in standards development?
- → How quickly do you adopt new standards after they emerge?
- → Do you support metadata interchange with other platforms?
- → What is your position on open metadata initiatives?

8. Implementation and Support Questions

Implementation Approach

- → What is your typical implementation timeline for an organization of our size?
- → How do you structure implementation phases to deliver early value?
- → What resources will we need to allocate for implementation?
- → How do you handle data migration from existing catalogs or documentation?
- → What is your approach to change management and user adoption?

Ongoing Support

- → What support tiers do you offer, and what are the response times for each?
- → Do you provide dedicated customer success managers? At what cost?
- → How do you handle feature requests and product feedback?
- → What training resources are available for different user roles?
- → Can you describe your user community and how customers share best practices?

9. Cost and Value Questions

Pricing Structure

- → How is your platform priced (by users, data sources, metadata volume, etc.)?
- → What are the cost implications as we scale our usage?
- → Are all features included in the base price, or are some modules additional?
- → What costs should we anticipate beyond the platform license?
- → Can you provide a comprehensive TCO analysis including implementation, training, and ongoing management?

Value Realization

- → How do customers typically measure ROI from your platform?
- → Can you provide case studies with quantified business benefits?
- → What metrics should we track to assess the value delivered?
- → How quickly do customers typically realize initial value?
- → What are the most common business outcomes your customers achieve?

10. Vendor Stability and Vision

Company Background

- → What is your company's financial position and funding status?
- → How long have you been operating in the metadata management space?
- → What is your current customer count and retention rate?
- → Who are your largest customers, and can we speak with them?
- → How does metadata management fit into your broader product strategy?

Strategic Vision

- \rightarrow What is your 3-5 year vision for the metadata management space?
- → How are you positioning your platform for emerging technologies and practices?
- → What strategic partnerships have you formed to enhance your offerings?
- → How do you view the competitive landscape, and what differentiates your approach?
- → How will your platform evolve to address changing regulatory requirements?