



How to Evaluate a Data Catalog



The Five Key Aspects for Success

Every enterprise has ambitious goals, and more often than not, data is key to reaching those goals. From customer data to application data and sensor data, the potential value of data driven decision-making to digital transformation and business impact is incredible. Creating this impact from data starts with a thriving data culture where everyone is empowered to make data-driven decisions.

But, creating a data culture is challenging. According to an [executive survey on AI and big data](#) from NewVantage Partners, “72% of businesses do not yet have a data culture despite increasing investment in big data and AI.” To overcome this challenge, more and more leading enterprises are turning to data catalogs for data search & discovery, collaboration and analysis, and curation and data governance.

[According to 451 Research](#), “An overwhelming majority of organizations see data catalogs as a potential solution” with 85% of organizations leveraging some form of data catalog.¹ Gartner calls [data catalogs a “must-have” for data and analytics leaders](#) and reports that, “Demand for data catalogs is soaring as organizations continue to struggle with finding, inventorying, and analyzing vastly distributed and diverse data assets.”²

“Demand for data catalogs is soaring as organizations continue to struggle with finding, inventorying, and analyzing vastly distributed and diverse data assets.”

Gartner

All data catalogs, however, are not created equal. With growing demand, numerous offerings have emerged. There are the data catalogs that work for a specific tool or infrastructure, and there are data catalogs built as part of command-and-control governance suites. But only a data catalog built as a platform can empower people to find, understand, and govern data, and support emerging data intelligence use cases.

¹ Doppelgangers: the many identities of the modern data catalog by Paige Bartley, January 2020

² Augmented Data Catalogs: Now an Enterprise Must-Have for Data and Analytics Leaders by Ehtisham Zaidi and Guido De Simoni, September 2019

In order for a data catalog initiative to be successful, the data catalog must address five key aspects: intelligence, collaboration, guided navigation, active data governance, and broad, deep connectivity. In this whitepaper, we will address why each of these aspects is critical to a successful data catalog implementation, and detail how the Alation Data Catalog was designed to address each.

Intelligence

Intelligence is at the core of a data catalog's ability to make data searches relevant and curation scalable. Intelligence automatically surfaces clues in the data to remove the manual effort otherwise required for discovery within the huge volume, variety, and veracity of data facing the modern enterprise.

Enterprises have struggled for decades to ensure that data consumers are leveraging relevant data sources. Manual documentation and wikis are a start but can't keep up with rapidly changing environments. Manually created labels and tags are good but are often inconsistent both in usage and taxonomy. What's needed is an intelligent system that is automated and relies on clues within the data to deliver far superior relevancy in results.

“What’s needed is an intelligent system that is automated and relies on clues within the data to deliver far superior relevancy in results.”

Collaboration

Data and analytics is a team sport, and data consumers need a place to collaborate on data, especially as remote work becomes the norm. With effective collaboration, each contributor works toward a common goal, building off of the work of others, and opening the door for greater innovation. Without collaboration, data consumers (and their knowledge) are siloed and work is needlessly recreated. To create a data culture, collaboration needs to be a seamless part of data and analytics.

Collaboration is also necessary for capturing tribal knowledge and making it available to the entire organization. Without a place to easily

capture tribal knowledge, individuals tend to keep their knowledge to themselves. Then when that person leaves for a new opportunity, retires, or is furloughed in challenging economic times, their knowledge goes with them.

Finally, when evaluating a data catalog, it's important to ask, "will people use the data catalog?"

Collaboration is critical to spurring adoption. Effective collaboration draws more people into the data catalog and with more people leveraging the data catalog, the value increases, further encouraging more people to participate — resulting in a virtuous cycle.

“Effective collaboration draws more people into the data catalog and with more people leveraging the data catalog, the value increases, further encouraging more people to participate — resulting in a virtuous cycle.”

Guided Navigation

Guided navigation is important for helping data consumers find the right data. The huge number of locations where important data can reside — from data lakes, databases, and reports to APIs and queries — makes finding data difficult even for veteran data analysts. When coupled with cryptic and inconsistent naming conventions and encoding, finding the right data can be a time-consuming, uphill battle. Guided navigation takes out the complexity.

But, finding data is only half the battle — using it is the other half. Guided navigation provides data consumers with a better understanding of how to use the data. Just by looking at a dataset, it's difficult to know whether it is right for the analysis at hand. It's not uncommon to find multiple assets with similar names containing different data. Guided navigation points data consumers to the right asset with warnings, deprecations, and endorsements and relevant related data assets and projections of the data. On top of that, guided navigation surfaces any policies or guidelines that should apply.

This process of trying to find and understand data can take up 80% of an analyst's time — and the result might still be wrong or violate a policy. Guided navigation is critical for alleviating the productivity-sink by pointing data consumers to the right data, the right context, and ensuring that the data is being used properly. Rather than giving data consumers an atlas, guided navigation provides them with turn-by-turn directions.

Active Data Governance

In this day and age, data governance is critical to the effectiveness of data and analytics. Data governance must ensure compliance and consistency and increasingly, effective data governance must be integrated into the day-to-day activities of data consumers. This requires an active approach to data governance.

Active data governance ensures compliance and consistency while also encouraging proper data uses, helping to connect policies to how data is being used. Data privacy regulations, like the [General Data Protection Regulation](#) (GDPR) in the EU, the [California Consumer Privacy Act](#) (CCPA) and more than a dozen others can result in large fines, and failing to protect customer data can cause irreparable damage to the brand. These factors make ensuring that data use adheres to policies critical.

Traditional data governance puts an emphasis on compliance. This documentation-centric approach might work for passing periodic audits but struggles to ensure that data is being used in a compliant fashion. On the other hand, many organizations employ a command-and-control approach to data governance, which imposes restrictions and controls that can have a chilling effect on data and analytics. A command-and-control approach often limits access to data, putting hurdles between data consumers and the very data that they need to move the business forward. Instead, an active approach to data governance encourages data-driven decision-making while closing the gap between guidelines, policies, and how the data is actually being used.

Broad, Deep, Connectivity

Broad, deep, connectivity is critical for overcoming data silos and the complexity of data environments. Enterprises need to reduce complexity by providing data consumers with one central location to access and

manage data. For a data catalog to effectively fill this need, it must have broad connectivity to all of the data sources that matter for data and analytics. The data catalog must also have deep connectivity, connecting to the metadata and behavioral information that provides visibility into the data and how it is being used.

The Alation Data Catalog

At Alation, we believe that addressing these five areas is critical to overcoming the challenges of modern data management and creating the foundation for data culture. Alation pioneered the data catalog space and

is leading the evolution of the catalog to become a platform for data intelligence.

Alation was designed to be strong in the five areas critical to creating a data culture: intelligence, collaboration, guided navigation, active data governance, and broad, deep connectivity. In this section, we will explain how Alation addresses each.



Intelligence

Alation provides the most intelligent data catalog for driving productive, accurate analytics, and active data governance. Alation surfaces recommendations to data consumers to help them easily find and understand data. Alation also surfaces guidelines and policies to ensure accurate, well-governed analytics, and to enable data governance teams to close the gap between policies and how data is actually being used.

At the heart of Alation’s ability to surface intelligence is a unique Behavioral Analysis Engine (BAE) that leverages machine learning-driven pattern recognition to surface insights on how data is being used. To enable teams to govern the incredible volumes of data in the enterprise, Alation automatically detects the source of each data element, conducts

impact analysis to show which data assets and people would be affected by changes to the data, and enables rules and updates to be applied across all data sets — maximizing the efforts of data governance teams. With natural language search, the intelligence within Alation is made accessible to a wide range of users, helping to raise the functional data literacy of the entire organization.

Collaboration

Alation removes the barriers to collaboration, encouraging broad community participation, and adoption. With one place for everyone to collaborate on data, tribal knowledge is also accessible to everyone, helping everyone use data better, and ensuring that knowledge isn't lost when experts move on from the organization.

“At the heart of Alation’s ability to surface intelligence is a unique Behavioral Analysis Engine (BAE) that leverages machine learning-driven pattern recognition to surface insights on how data is being used.”

To encourage participation, Alation combines human-centered design and consumer grade ease-of-use to make collaboration intuitive. Automated expert identification, pinpoints ideal stewards and makes it easy to ask questions of the right person every time. Wiki-like articles and fully searchable conversations enable knowledge to be captured from anywhere. Fully searchable, integrated conversations keep everyone connected — even across time zones — and subscriptions notifications alert data consumers when data sets, definitions, and the articles that they care about change, so that everyone is always on the same page.

Guided Navigation

Alation’s data catalog provides a guided approach to navigating the data environment. Data consumers are given intelligent suggestions as they are searching for data and using data. Whether leveraging natural language search or using the data catalog’s built-in SQL editor, data consumers are given guidance right as they are working with data — making it easier for them to find relevant data and understand how to use it. Alation’s approach

to guided navigation also makes it easier for data governance teams to ensure that data is being used properly by enabling them to surface rules and policies as the data consumer works with the data.

Alation's guided navigation points data consumers to relevant data, related data, and projections of that data. This enables business users to focus on finding the answers they need, analysts to focus on creating dashboards and reports, and data scientists to focus on building models, rather than writing queries and trying to navigate the data environment. While at the same time, data governance teams are assured that data consumers are seeing all of the context and policies that are important to accurate data use.

Active Data Governance

Alation empowers an active approach to data governance, making data governance a nearly seamless part of the activities of data consumers. With Alation, data consumers are encouraged to use data the right way, and data governance teams are given the tools to impact how data is actually being used — without stifling data use.

Alation optimizes the work of data governance teams by giving them a single place for defining enterprise data policies and workflows. Alation closes the gap between policy creation and enforcement with impact analysis to determine which assets are impacted by upstream changes. The data governance team can alert users about changes in asset status whether through the data catalog, email, or directly into BI and visualization tools like Tableau and Salesforce Einstein Analytics.

Broad, Deep Connectivity

Alation's data catalog is underpinned by broad, deep connectivity. Pre-built connectors enable the data catalog to access data from a wide range of sources from cloud data lakes to business intelligence tools. And, an Open Connector SDK enables the data catalog to connect to any source that doesn't currently have a pre-built connector. Alation not only connects to the data but also the technical, behavioral, and business metadata that fuels the data catalogs intelligence as well as query logs to provide a deeper understanding how the data is being used, including which assets are most popular and who uses them most.

Alation is the Platform for Data Intelligence

Alation pioneered the modern data catalog market, and is leading the evolution of the data catalog to become a platform for data intelligence. No one tool can serve all of an enterprise's data intelligence needs. Instead, enterprises need a platform that supports data intelligence.

According to IDC, "Rather than collecting all the diverse, distributed, and large-scale data into one common platform, intelligence about the data

"Rather than collecting all the diverse, distributed, and large-scale data into one common platform, intelligence about the data becomes the platform."

IDC

becomes the platform."³ We believe that not only is Alation the best data catalog for overcoming the challenges of modern data management and creating a data culture, but also the best platform for data intelligence use cases.

³IDC MarketScape: Worldwide Data Catalog Software 2020 Vendor Assessment, by Stewart Bond, July 2020