

3

Things to Consider When Evaluating a Data Catalog



Data catalogs act as guides to the location, context, lineage, and meaning of data.

As mountains of data accumulate throughout your organization, just having a data governance plan may not be enough. An integrated data catalog can bring sense to the madness of runaway data growth.

THE AWARENESS AND READINESS GAP

Anyone with a shred of responsibility for data governance knows that accelerating growth of all forms of information—especially unstructured data from greater and more diverse devices and applications—is wreaking havoc with enterprises' ability to derive sufficient value from the data. Big data, business intelligence, data warehouses, and all forms of real-time analytics are heavily dependent upon solid data governance frameworks in order to harness the growing volume and variety of information being created, stored, and managed.

Governance policies, platforms, and tools have helped to give data stewards and "data citizens" more confidence in the accuracy, comprehensiveness, and timeliness of the data. However, a major roadblock still thwarts many organizations in their efforts to empower business stakeholders to use data for tangible benefit: finding the data and understanding what it means.

Without the proper methodology and architecture for locating relevant data and discerning its true meaning, organizations will continue to flail away in the dark in hopes of stumbling upon the right information at the right time for the right purpose.

That's where a data catalog comes in. Data catalogs act as guides to the location, context, lineage, and meaning of data. And they put data into an actionable and sensible structure that personnel ranging from the most sophisticated data analysts to everyday business users can navigate. This ability to sift through huge data sets, and a rapidly growing array of seemingly disconnected information within and outside those data sets, is often the difference between "eureka" moments and head-scratching confusion for organizations plagued with information overload.

Just as data governance gives organizations the confidence they need to trust the accuracy and confidence of the data, data catalogs give users the ability not only to find the data, but also to understand its meaning and relevance in solving business problems. For all the executive-speak about becoming a data-driven organization, no enterprise can accomplish that worthy goal without the proper data catalog.

But data catalogs vary significantly in their functionality, efficiency, and ability to work hand-in-glove with data governance platforms. Discerning among myriad marketing claims about the quality and fit of one data catalog over another requires hints about how to evaluate and select the right data platform for your needs.

Data catalogs vary significantly in their functionality, efficiency, and ability to work hand-in-glove with data governance platforms.

3 TIPS FOR SELECTING THE RIGHT DATA CATALOG

Tip 1: Believe in the power of crowdsourced metadata.

Nearly all data catalogs offer the ability to connect to data sources and read metadata, which is the lifeblood for insightful data analytics. But in order to truly trust the data in these sources, it's important to have the ability to utilize crowdsourced metadata. When data analysts and business users can see which people have accessed and used the data previously, as well as written reviews that essentially vouch for its accuracy and usefulness, the data becomes inherently more trustworthy and relevant. This open-source-like approach relies on collaboration and iterative content to build real-world ratings of the data. And this gives new and future users of the data not only the confidence in knowing the data is accurate, but also provides invaluable hints at where to access similar data to provide even richer context for the data.

Tip 2: Look for a catalog natively integrated within a data governance platform.

When the catalog is natively integrated into the data governance platform, it achieves many interoperability and data consistency benefits. One of the most important is the ability to share a common business glossary. After all, every organization—regardless of industry, geography, or business model—has a unique language with terminology that doesn't necessarily equate to what a bolt-on catalog application might use in its standard glossary. Integration from the ground up means that organizations can more easily link technical data to standardized business terms, making it easier to search for data by business definition rather than by arcane technical jargon.

Tip 3: Tracking data usage is a critical requirement—and a telling differentiator.

When business users and data analysts evaluate the quality, integrity, and appropriateness of a particular data set, it's natural for them to wonder who else in the organization has accessed that data, for what purpose, and toward which goals. This gives users the ability to easily review what their colleagues have said about, and done with, the data. Automated logging of access requests creates traceability starting at the original data source. And this helps those accessing the data understand what it can and should be used for, and where it might not be appropriate. Also, be sure that your catalog has an integrated report catalog that is linked to actual business applications. But remember: A data catalog is far more than a report repository.

IT and business executives alike need sophisticated, yet easy-to-use, catalogs to make it easier to find and use relevant data.

HOW COLLIBRA CATALOG HELPS MAKE SENSE OF IT ALL

Not every data catalog meets the three requirements discussed above—and then some. But one solution that addresses all vital data catalog functions is Collibra Catalog. The heritage of Collibra in data governance platforms is well established, and the company is recognized as a market leader for solutions that help organizations trust the accuracy, timeliness, and relevance of their data.

Its utility is underscored by a design that delivers the robust performance and extensive functionality demanded by data scientists and analysts, as well as the ease of use and self-service sought by less-technical business users. Collibra Catalog combines automated discovery tools with an orderly index of approved and trusted data sets to surface relevant data through semantic search.

Data experts will appreciate Collibra Catalog's customization and personalization capabilities, providing "insights at a glance" based on such criteria as certification level, quality, ownership, stewardship, content, and more. Line-of-business stakeholders also benefit from Collibra Catalog's use of a centralized data warehouse for one-stop-shopping of data, as well as automated access approvals to speed data identification and delivery.

Finally, Collibra Catalog is tightly integrated into the Collibra data governance platform, helping users avoid inefficiencies and inconsistencies characteristic of "bolt-on" catalogs that are not designed from the ground up to be part of the data governance framework.

CONCLUSION

As more and more enterprises appreciate the value of a robust, secure, and flexible data governance platform, IT and business executives alike need sophisticated, yet easy-to-use, catalogs to make it easier to find and use relevant data from any number of different devices, applications, and repositories.

Potential buyers of data catalogs should take their time to ensure they are evaluating different products on an apples-to-apples basis, including such requirements as crowdsourced metadata, integration with data governance platforms, and data usage tracking.

Collibra Catalog builds on the long-standing leadership of Collibra in data governance with an integrated platform that empowers both sophisticated data scientists and less-technical data citizens to locate and truly understand accurate, relevant, and timely data.

For more information about the Collibra Catalog and how it can help you democratize your data, please visit www.collibra.com/data-catalog.