

WHITE PAPER

The Guide to Enforcing Access Control Policies with Your Data Catalog's Metadata

The Case for Data Catalogs

Research from IBM shows business leaders spend 70% of their time finding data, and only 30% utilizing it. How valuable is data if it's being used to less than a third of its potential?

As data becomes necessary to compete in today's fast-moving market, businesses are more reliant than ever on data analytics to drive strategic insights and decision-making. This includes integrated business analytics and dashboards, as well as self-service analytics. Aggregating and analyzing data can help improve business operations in every industry from healthcare to travel and leisure – and the jumping off point for effective data discovery and utilization by all users, is a well-maintained data catalog.

A data catalog is an organized inventory of data assets that allows data users to find, access, and evaluate data in a centralized location for analytical and business uses. Data catalogs leverage metadata

to help users quickly catalog data across an organization's entire data landscape, understand the data available to them, and operationalize that data for insight-driving analyses.

Metadata is particularly useful for surfacing data that might be relevant to a user's needs, even if it does not exactly match their search keywords. Additionally, since data sources and products continue to grow and evolve, curating and managing data in a centralized data catalog streamlines and enhances data usage and outcomes.

The business case for data catalogs is clear – research shows that leaders in data management drove 69% more revenue, 57% greater profits, 72% higher customer satisfaction, and 62% more product and service launches than those without a data management strategy and resources.

Data Catalog Challenges in a Modern Data Stack

Data catalog tools help users discover and leverage data for analysis, which in turn informs decisions that affect the entire enterprise. While this is a boon for efficiency and productivity, maintaining control over data access as the number of users and data sources grows can be challenging. With more people trying to access more data – including sensitive data – the surface area for risk and misuse increases substantially.

A survey of data engineers revealed that 92% are subject to one or more data use rules or regulations, and Gartner reports that 80% of organizations worldwide will have to adhere to at least one privacy-focused regulation within the next two years. This is happening against the backdrop of a growing body of data use legislation and increasingly common high-profile data leaks and breaches by

major corporations, including Amazon, T-Mobile, and WhatsApp. Organizations leveraging data catalogs to enable easy access to all business data must have a way to manage access control and audit usage, or else risk paying millions of dollars in fines and unquantifiable amounts in reputational damage.

To complicate matters further, many companies now have multiple cloud data platforms in their modern data stacks, each with different access control capabilities. Managing access and use consistently across these platforms while still enabling data discoverability and self-service access can become a substantial task on data teams' time and productivity. Overly broad policy enforcement could result in costly noncompliance, but overly restrictive policies that lock down all data defeat the purpose of investing in a data catalog.

Data Catalogs: A Tipping Point for Data Access Control

To achieve the greatest possible data catalog ROI, it goes without saying that organizations must enable self-service data use without unnecessary overhead. In short, when data teams adopt a data catalog, they should also adopt a data access control solution.

Automating data access helps ensure the right people are able to find, access, and use the right data for the right purposes, without the time-consuming manual processes that delay speed to data insights and introduce potential for human error. This complements the very function of a data catalog – to efficiently find, access, and evaluate data for analytics. Since data catalogs centralize data assets and leverage metadata from multiple sources, a data access control tool is essential to ensuring any data that might be queried is protected. The two solutions go hand-in-hand.

Modern data access control solutions like Immuta give data engineering and DataOps teams the power to leverage metadata to build and automatically enforce data access policies at query time. Many Immuta customers have adopted leading data catalogs, such as Alation, Collibra, Informatica, and others, as their enterprise-wide data catalog solution. When combined with Immuta, these customers are better able to maximize their existing data catalog investments by operationalizing metadata for safe data analytics use.

When an organization brings on a data catalog, a data access control solution should be considered simultaneously. This is the easiest way to protect data across the entire enterprise, reap the benefits of the data catalog, enable faster enterprise-wide data access, and make life easier for data architects and engineers.

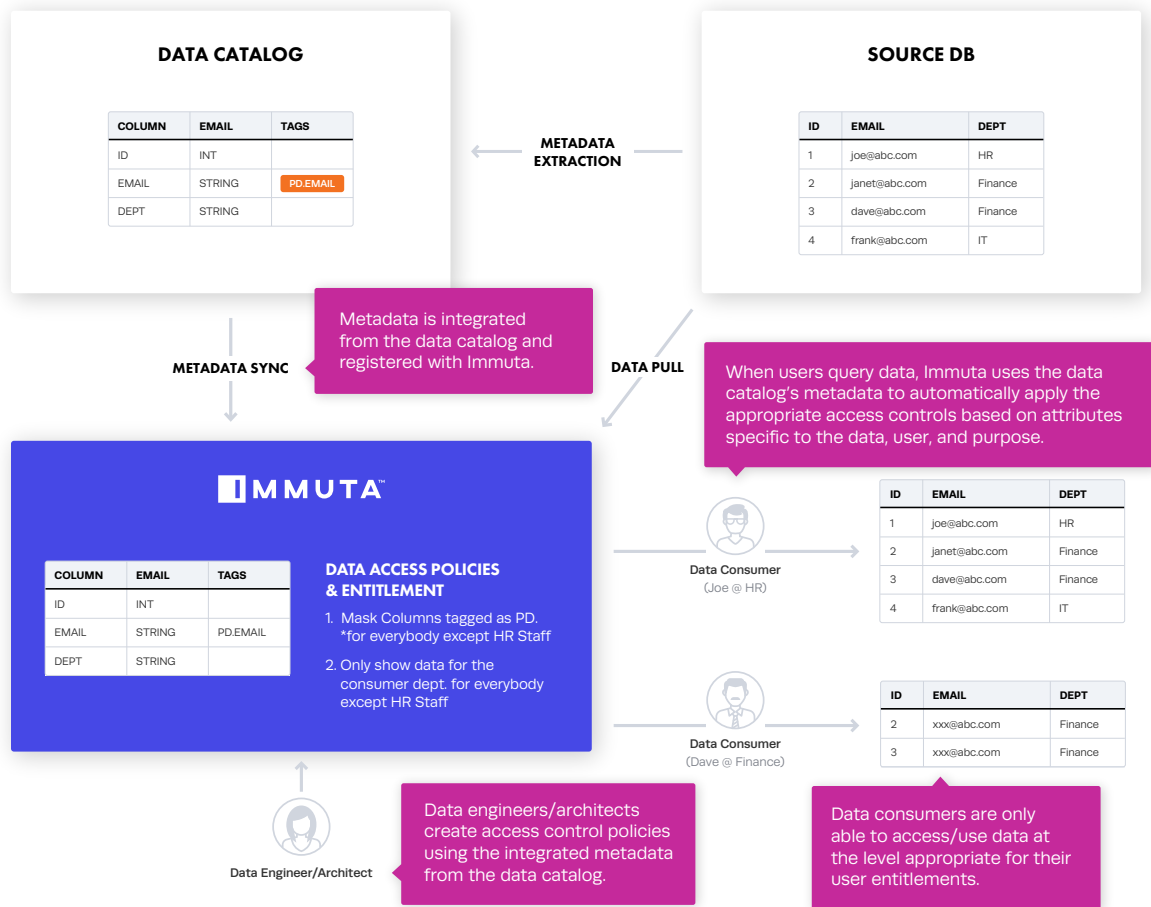
How Access Control for Data Catalogs Works

Data stewards are primarily responsible for creating and maintaining data access policies, but the onus is on data architects and engineers to ensure the policies are continuously implemented across any cloud data platform, including Databricks, Snowflake, AWS, and others. Additionally, data teams need to account for real-time changes to data, roles, purposes, or other dynamic attributes, and be able to adjust policies as necessary without creating bottlenecks.

Solutions like Immuta provide a data access control layer that can be integrated with metadata from data catalogs to provide automated security and privacy controls that enforce the enterprise policies for data analytics use. Assets that should not be accessed by certain analysts are not necessarily

fully blocked – instead, analysts are able to query a resource that provides an appropriate level of insights while still retaining, masking, or obfuscating objects that shouldn't be accessed or viewed. All of this is done automatically at query time, meaning analysts still receive quality data that is amended for their specific needs and permissions. This approach avoids the conflict between overly broad and overly restrictive data privacy controls, maximizing data's utility without compromising its security.

To see how this works in production, watch this on-demand webinar presented by the Director of Enterprise Architecture at Aon. Below is a representation from the webinar of how Aon leveraged its data catalog with Immuta.



Implementing Immuta With Your Data Catalog

How exactly does Immuta work with your data catalog? First, metadata is integrated from the data catalog and registered with Immuta, which data governance engineers and architects can then use to create access control policies. When users query data, Immuta uses the data catalog's metadata to automatically apply the appropriate access controls based on attributes specific to the data, user, and purpose. Ultimately, data consumers are only able to access and use data at the level appropriate for their user entitlements.

Data teams are able to integrate external data catalogs with Immuta with minimal overhead – in fact, the process can be done in just four steps:

1. In Immuta, click the **App Settings** icon.
2. Click the link in the **Configuration** panel.
3. Click **External Catalogs** and select your data catalog. Enter the **URL** and **API Key** to test the connection.
4. Create data access policies.

The Benefits of Data Access Control for Data Catalogs

Metadata and data usage are mutually beneficial for data catalogs: The more data consumers use the data catalog, the better its metadata becomes – and the better the metadata is, the easier it is for users to find data. Adding Immuta to that cycle helps ensure that the right data is accessible to the right people at the right time, based on metadata-informed access control policies.

Leveraging metadata to automatically enforce these policies at query time removes the burden on the DataOps team to manually monitor policy implementation. Now, data users will not be prohibited from accessing data that doesn't map to their role, but rather will see data with the appropriate level of insight based on dynamic attributes like usage purpose, data type, time, and more. This continuous enforcement on cloud data platforms applies to any data consumer, whether accessing

data from data science notebooks, analytical tools, or other sources, and it further reduces the need to use manual workflow steps to request or deny access to sensitive data.

With Immuta, data catalog users are able to:

- Provide secure, self-service data access, enabling data access in minutes instead of months
- Increase permitted use cases by 4x by avoiding overly restrictive access control policies and applying advanced privacy controls
- Maximize their data catalog investments by operationalizing metadata for secure data analytics use with full auditing capabilities

To find out how Immuta can help solve your data access control needs and optimize your data catalog use, book a capabilities briefing with our team.

[REQUEST A DEMO](#)

