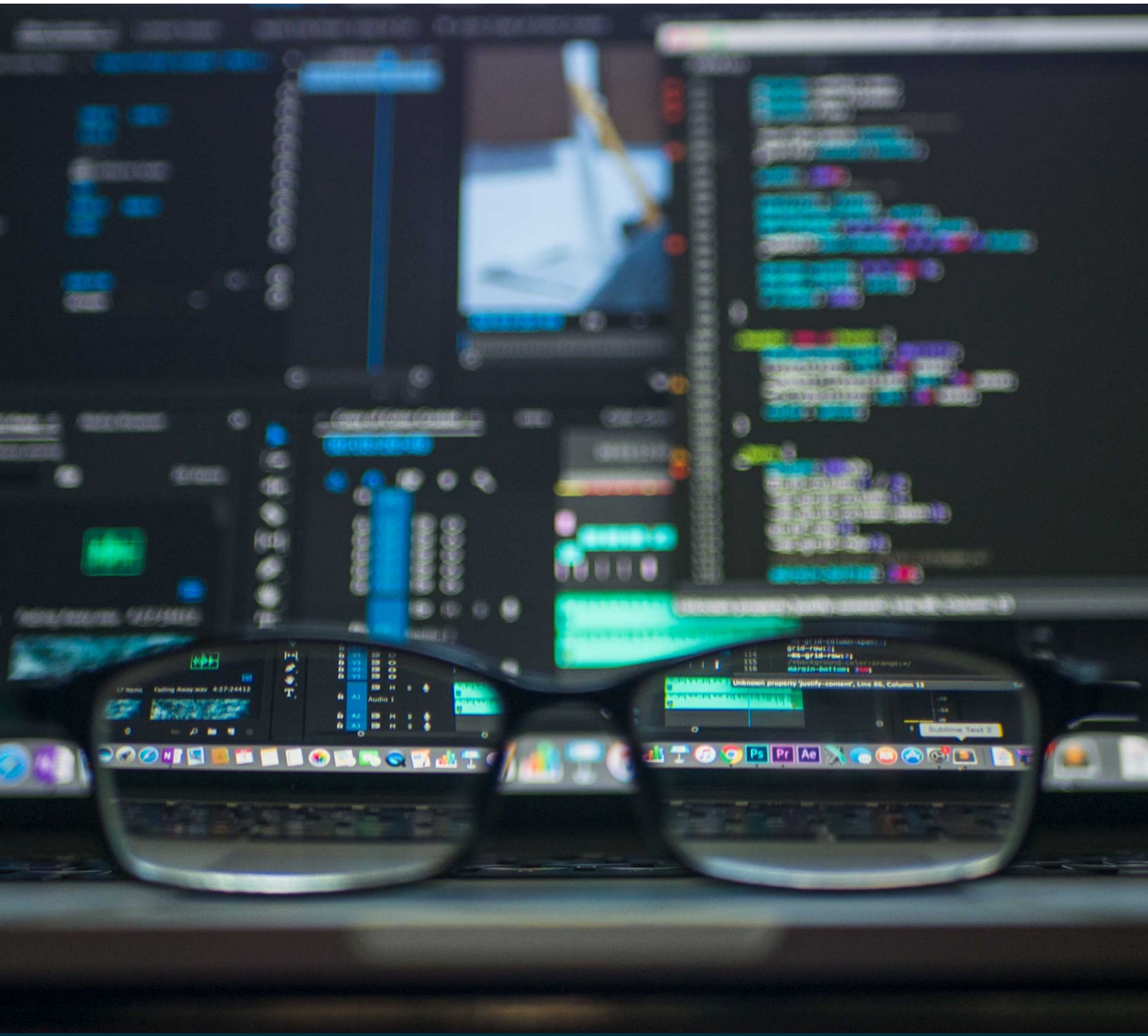
THREEBRIDGE SOLUTIONS

WWW.THREEBRIDGE.COM



8 WAYS A DATA CATALOG CHANGES EVERYTHING

SYSTEMIZING YOUR COMPANY'S MOST VALUABLE ASSET

WHAT IS A DATA CATALOG?

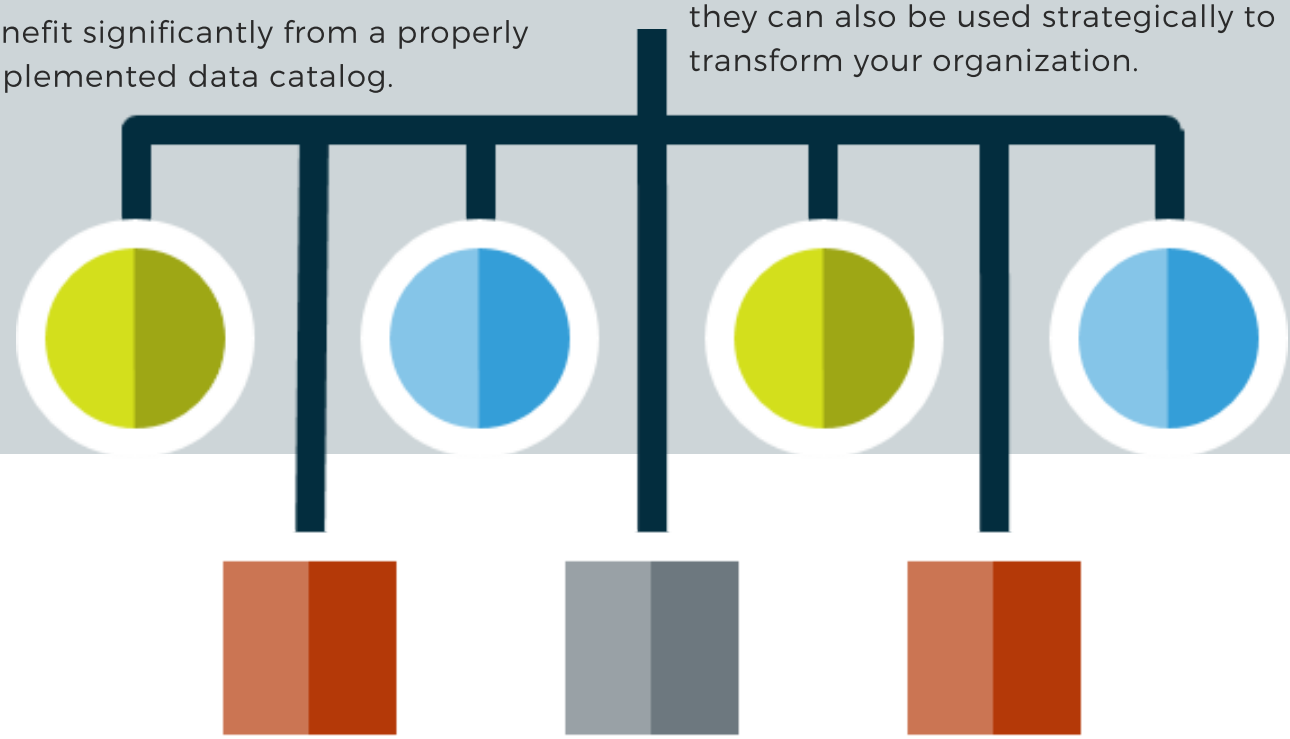
A data catalog is a service or system that profiles and organizes metadata on a company's data sources and assets.

A catalog far surpasses a simple inventory or dictionary - in a catalog, data sources and assets are automatically categorized and easily searchable, annotated and curated. Data catalogs are an essential foundation of a broader data management platform. Areas such as data governance, data quality and master data management play a key role in successful data management and a data catalog can provide the necessary foundation for managing data. Organizations that are in the process of becoming either more data-centric or data-focused will benefit significantly from a properly implemented data catalog.

Oftentimes, the need for a data catalog can be determined by the problems being posed by an organization, for example:

- "We don't even know what data we have."
- "Our users spend most of their time searching for, discovering, or managing data instead of actually using it. This needs to change."
- "We have strict regulatory requirements regarding storing and managing data."

Data catalogs not only provide a solution to these all-too-common problems, but they can also be used strategically to transform your organization.



8 WAYS A DATA CATALOG CHANGES EVERYTHING

1. COMPREHENSIVE VISIBILITY

Provide comprehensive visibility into all data sources from a central location

A data catalog centralizes all metadata and information on data sources into a single accessible location. This location can either be cloud-based or on-premises, regardless of the locations of the covered data sources.

2. SEARCH, DISCOVER, MANAGE

Allow data users to easily search, discover, and manage data sets and sources

While the data cataloging system should automatically discover and organize, self-service is a key aspect of the automation. Data users are able to find and retrieve information on data sources and assets without having to rely on subject matter experts or data owners. This feature immediately makes the tool more attractive to those users without deep expertise.

3. REDUCE TIME AND COST

Automate and reduce time and cost to find and search existing data

Data catalogs reduce the amount of time users spend finding or understanding data sources. This frees up the time of analysts, allowing them to generate insights and actually use the data, rather than searching through irrelevant data or trying to understand complex results. A good data catalog will automatically update itself, either by machine learning, custom built integrations or other automation methods.

4. PRESERVE KNOWLEDGE

Preserve institutional and tribal knowledge of data

It is not uncommon for institutional knowledge of an organization's data sources to reside with a small number of users. What's more problematic is that this knowledge is often times either poorly documented or not documented at all, residing only in the memories of individuals. This presents a massive risk in that such knowledge can be inefficient to share across the organization or lost altogether.

IN THE YEAR 2025,

the world will create 160,000,000,000 terabytes of data. 40% of that will need to be stored. How are you preparing?

5. COLLECT DESCRIPTIVE METADATA

Collect and store descriptive metadata, annotations, and comments to enrich cataloged data

One key aspect of a data catalog is the information added to enrich or improve metadata that may have been programmatically or automatically discovered by the catalog. This additional information is provided by all users, not just the users who may manage, govern or control such data sources. Enriched metadata often reflects institutional knowledge within the organization, and must be centrally stored and managed in order to provide the most value. Catalogs provide that haven for centralized storage, allowing users to eliminate multiple steps in the process of searching for information particular to other areas of the organization.

6. CATALOGING VIA MACHINE LEARNING

Enable smart cataloging through the use of machine learning

Complex and/or numerous data sources can generate thousands and thousands of new data assets that exceed the ability of manual cataloging. Additionally, the number and complexity of relationships can be unmanageable with any manual process. Newly-developed machine learning capabilities can not only handle the ever-increasing number of data sources, but can also identify trends and data usage patterns so that users don't have to identify them manually.

7. CLASSIFICATION AND SECURITY POLICIES

Implement classification and security policies around metadata

With organizations being subject to more and more data policies both internally and externally, a data catalog can provide the necessary controls to ensure compliance. External policies such as GDPR, HIPAA, SOX, etc. require appropriate technical and control measures in place to safeguard, preserve and manage organizational data. Internal policies may also require similar measures to be put in place for risk management, compliance or security. Such policies may also include a lineage component in which a data asset must be completely traceable back to its origins. Data catalogs provide the ability to thoroughly document and classify all data sources within an organization and apply any necessary policies or procedures.

8. CONTINUOUS AND AUTOMATIC UPDATING

Provide continuous population, updating, and scalability

In order for a data catalog to remain relevant, its contents must be continually updated and corrected. A data catalog with missing, incorrect or outdated information rapidly loses value and can result in mistrust or lack of use among frequent users. Integrations with other platforms and systems is essential and often implemented via published APIs. As an organization's data assets grow, the data catalog must be able to scale up and out as needed to manage these assets. Complex and/or numerous data sources can generate thousands and thousands of new data assets that exceed the ability of manual cataloging.

IT Infrastructure, including data catalog architecture, is one of the many capabilities we provide within our Smart Data Solution. For more information, visit www.threebridge.com.