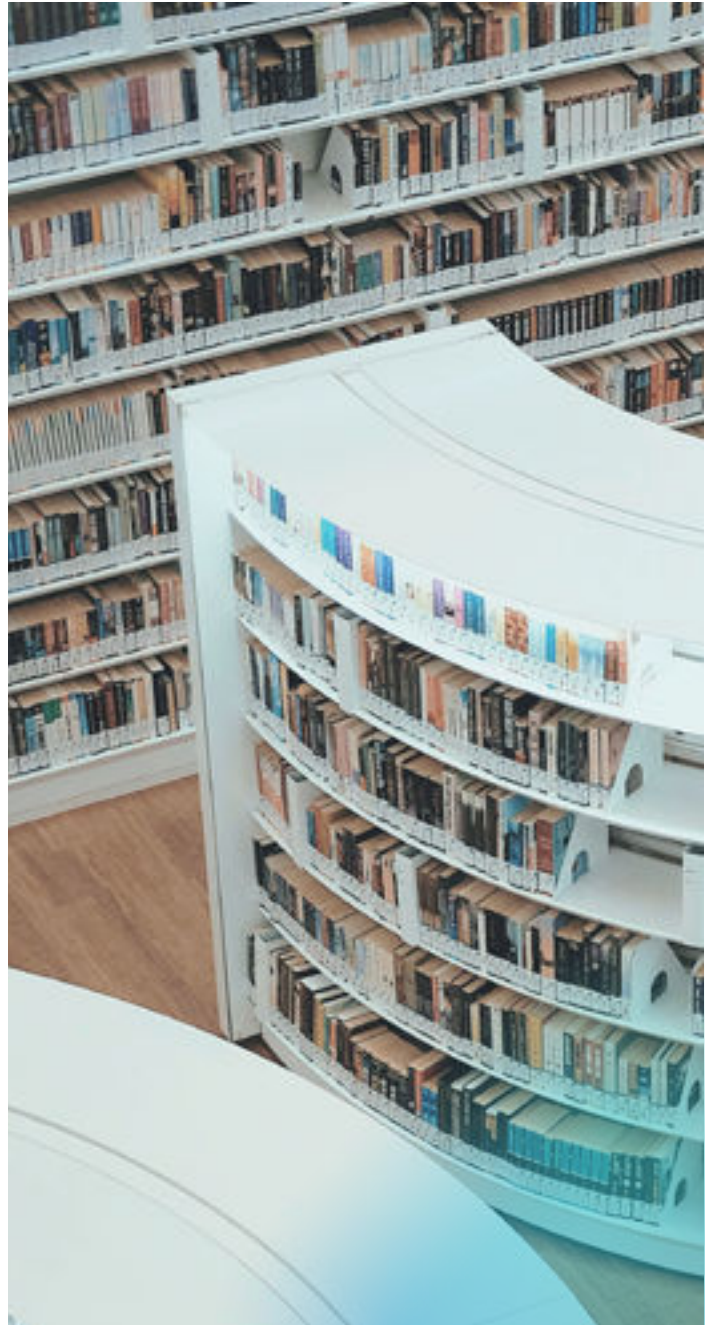# 10 Traps to Avoid for a Successful Data Catalog Project

How to build, sustain, and maximize the value of data catalogs.

Table of Contents

Executive Summary

The methodology used to implement a data catalog will change from one company to another, and many factors must be considered for the project to bear fruit. The right method is the one that takes these factors into account while at the same time avoiding potential traps to ensure project success.

Metadata management is an important component of a robust data management project, and it requires more than just the data catalog solution, no matter how connected it may be.

A data catalog will of course reduce the workload, but it won't, in and of itself, guarantee metadata management success.

This eBook is the fruit of a variety of contexts, such as industry sector, company size, and organizational culture, that impact data catalog usage. Its purpose is to highlight the pitfalls and preconceived ideas that should be avoided when rolling out an enterprise-wide data catalog project.

The 10 traps described in this eBook are articulated around four central themes that are crucial to the success of the initiative:

1. Supporting a data culture within the organization
2. Ensuring internal project sponsorship
3. Leading the project to success
4. Enabling technical integration of the data catalog

Recognizing and understanding pitfalls and traps are invaluable for a smooth roll out and execution of the project.

Supporting the Data Culture

Organizations with data as their sole product are very rare. While data is typically everywhere in a business, it is often a byproduct of the company's activities. It is therefore not surprising to find that some collaborators are not as aware as they should be of data's importance. Indeed, data culture isn't innate, and a lack of awareness of the importance of data can become a major obstacle to a successful data catalog deployment.

Let's illustrate this with 10 common preconceptions and traps to avoid:

Trap 1: Not all collaborators are sensitive to what's at stake with metadata management

This first obstacle is probably due to the lack of a detailed understanding of the initiative. That's why stakeholders must emphasize the importance of metadata management to colleagues who still misunderstand the crucial role data can play in an organization.

It's quite likely that a larger program will need to be set up that includes an awareness initiative emphasizing the stakes around enterprise data management. One of the most important elements to emphasize is the fact that data is for a common good—owners of a dataset have a duty to make it visible and understandable to all stakeholders and colleagues.

Indeed, one of the most common obstacles in a metadata management initiative is resistance to the effort needed to produce and maintain documentation. This is even more of an issue when it's felt that potential users are limited to a small group of people who already fully understand the subject. When it's understood that the target user group is in fact much larger—the entire organization and potentially all staff—it becomes obvious that this knowledge has to be recorded in a scalable manner.

The collaborator in charge of the production and maintenance of knowledge in a given field or topic will understand more readily the benefits this effort delivers. Their knowledge and skills are shown to a much larger audience, which reflects well on them. Managing metadata is also a great time saver because data product owners will have fewer direct requests for sharing information with colleagues because they can find the information on their own.

Trap 2: A data catalog doesn't do everything

An inaccurate understanding of data tools and their uses can lead to mistakes and cause suboptimal, even detrimental, decisions.

The data catalog is a central software component for metadata management, but it's likely not the only tool used. It is therefore not advisable to try to do everything using only this tool. This may sound obvious, but in practice it can be difficult to identify the limits, where it may be necessary to bring a more specialized solution into the mix.

As the keystone to documentation, the data catalog should be the entry point for any collaborator who has questions related to a concept linked to data. However, this doesn't make it "the solution" in which everything has to be found. This nuance is important because referencing or synthesizing information doesn't necessarily mean carrying this information wholesale.

There are many subjects that come up during the preparation phases of a metadata management project. They include technical or functional modeling, data habilitation management, workflows for access requests, and more. All of these topics are important, carry value, and are linked to data.

However, they are not specifically destined to be managed by the solution that documents the organization's data assets. It is therefore important to begin by identifying requirements, defining a response strategy with the most relevant tooling, and then integrating this tooling into an ecosystem that's larger than just the data catalog.

Yet in a landscape where vendors offer competing solutions with sometimes unclear boundaries, choosing the right software components for a metadata management project can be challenging. The tools selected should work together seamlessly, sharing information to help users efficiently find the answers they need. To navigate this complexity, teams must take a strategic approach—first identifying their specific requirements, then selecting the most suitable solutions, and finally integrating them into a broader data ecosystem that extends beyond the data catalog itself.

Ensuring Sponsorship for the Project

Metadata management projects will inevitably lead to multiple changes, which impact organization and collaborator responsibilities. These changes cannot be made without the initiative being supported by the upper echelons in the organization.

Trap 3: A data catalog project cannot succeed without internal management support

In a metadata management initiative, some collaborators will inherit new responsibilities and directives. The initiative is often steered by a dedicated, cross-functional team that orchestrates the project and facilitates its execution.

That said, the collaborators being asked to contribute to the project are often not actually managed by that team— they work in another service or department. Without a managerial go-between across the different teams, a common discourse, and in some cases written objectives relating to the responsibilities of each team, the initiative remains somewhat fragile. At the very first obstacle, the initiative can even be sunk because the necessary steps to success were never officially made.

The best approach will depend largely on the internal organization of each company. At the same time, it is a best practice to write down objectives to make them official, nudge the work of the contributors in the right direction, and steer the results in the desired direction.

ACTIAN™
a division of HCLSoftware

zeenea

Trap 4: A data catalog project requires an initial investment

It's common to carry out a census of all the available information at the start of a metadata management project before feeding the data catalog. This information usually comes from existing documentation, but it can also come from colleagues who have their own insights and expertise.

The first step is to centralize and secure this metadata by inputting it into the data catalog. The catalog needs to provide a simple way to centralize all this information and share it with all users. For example, the Actian Data Catalog is a connected solution that provides various mechanisms to do this. It can automatically bring up metadata from master systems, eliminating the need for contributors to handle these tasks manually.

Moreover, the connectivity serves another purpose—ensuring the catalog is kept up to date and aligned with the master systems. This applies to the metadata that's automatically synchronized but also to the metadata coming from collaborator contributions. By its nature, an information system is alive—data evolves, as does the associated documentation. An upkeep of the documentation is therefore critical to ensure its freshness.

This enrichment effort also addresses another issue—required documentation at a given time won't necessarily be the same at a later date. For example, new regulations can necessitate specific and additional documentation, and new criteria may be needed to better optimize the data. These evolutions require constant enrichment and updating. The project must be planned for the long term, with all the financial and human implications considered.

Steering the Project to Success

As with all projects, a metadata management initiative has to be properly steered to meet the objectives within the best timeframe and cost parameters. It is important that steering the project doesn't fall into the traps described here.

Trap 5: Metadata quantity should never become more important than quality

The purpose of the data catalog is to document company data assets. When the project starts, the absence of information often leads to the same tendency—to add a lot of information.

A good data catalog isn't characterized by the quantity of objects, but rather the quality and coherence of the information. These characteristics require close supervision to identify priorities both in terms of the perimeters covered and the information selected for inclusion.

Even if this process causes frustration, it will very quickly prove its effectiveness and crucial importance for the project to succeed. Users will rightly consider the data catalog as a source of truth the same way a dictionary is for language. What use would a dictionary have if all it provided was a list of words without information such as grammatical class, spelling varieties, etymology, or definition?

The same applies for the data catalog. It's better to offer selected and quality content, starting with a target audience, to deliver experiences that encourage people to come back for future searches. It's difficult to keep users interested if the first exposure is a failure.

ACTIAN™
a division of HCLSoftware

zeenea

Trap 6: A data catalog won't be filled spontaneously, even when it's open to users

Even though the data catalog is open to many users—some of whom will have knowledge about the organization's data assets—spontaneous and regular updating of data from the start of the project is exceedingly rare. Both the quality and quantity of the information has to be supervised, and it's also important to educate contributors about processes.

Managing contributions can be achieved by creating established processes that enable control by stakeholders. Inviting users to correct or enrich the data catalog also helps with contributions.

Trap 7: It's impossible to set all the project objectives at the start and not evolve them

The data catalog must meet expectations for many users and their myriad requirements. It's therefore unreasonable to assume that the complete list of expectations is available at the start of the project.

Likewise, it's naive to believe that this list is fixed and immutable from the beginning. Instead, stakeholders will need to continuously collect and analyze requirements, interpret them accurately, and prioritize and transform them into appropriate content.

Generally, requirements evolve according to different parameters that are not yet established at the onset. For instance, the level of enterprise and staff maturity with regards to data management will change over time, as will the development of use cases around data and data related regulations. All these parameters can potentially have a strong impact on the content that the data catalog will cover, both in terms of scope and the nature of the information provided for the data assets.

Integrating the Data Catalog

Integrating the data catalog into the enterprise ecosystem provides opportunities to create new and sustainable value. It's essential to consider the aspects of integration and understand the potential rewards.

Trap 8: Not all data must be entered manually

Many systems produce, aggregate, and enable metadata to be entered for localized value— metadata that's useful or relevant within a specific system, department, or organizational unit rather than across the entire enterprise.

The information should be consolidated and retrievable in the data catalog and not have to be entered multiple times in different systems, which saves time while helping to ensure data reliability and availability. The data catalog therefore presents an opportunity to consolidate this information with the knowledge of contributors in their respective fields.

However, consolidation must be thought out through a technical integration rather than a manual effort. Carrying out data imports and exports between systems through human actions is not a best practice or efficient.

One strength of a data catalog is its capacity to ingest metadata via technical integration processes, ensuring a robust synchronization between systems. The concept of a metadata platform makes complete sense when viewed through this lens.

ACTIAN™
a division of HCLSoftware

zeenea

Trap 9: The data catalog isn't an 'automagical' tool

Thinking that a data catalog can extract all types of metadata, regardless of its source or format, is misleading. The catalog should of course facilitate metadata retrieval, but some metadata will not be retrievable automatically.

The first reason for this resides in the origin of some metadata—information may simply not be present in systems because it originated solely from the knowledge of experts. In this case, the data catalog can become the master system and be eligible to receive this information.

Conversely, some information in a system can be impossible to retrieve in an automated manner because:

- An interface is lacking to enable information to be accessed in a stable manner, although this is rare with modern systems

- There is difficulty isolating relevant information for the data catalog from other information, even when resorting to retro engineering to try and isolate relevant data, like using code source analysis to derive data lineage information

The risk of producing "noise" around the information is therefore high and can lead to a degradation of the quality of the data catalog content and ultimately discourage data teams from using it.

Trap 10: The data catalog must not be connected to a unique data source

Metadata comes from numerous and varied data products. As a result, multiple and complementary sources are needed for a comprehensive understanding of the data. It is precisely the reconciliation of this information in a central solution—the data catalog—that provides the necessary elements to users.

A connected data catalog is a valuable asset because data product discovery and the associated metadata retrieval are made considerably easier as a result of automation. Connectivity can also extend to complementary systems. These systems can enable, if needed, the visualization of data lineage and can document the flow and transformation of data between systems.

These systems can also function independently. Yet their inclusion in the catalog contributes to a comprehensive and detailed mapping of the company's data assets.

Finally, because the catalog can document a wide range of asset types, various connected sources help enrich both the semantic and physical layers. These sources are integrated progressively, following a strategic, iterative approach designed to maximize value at each stage.

ACTIAN™
a division of HCLSoftware

zeenea

Key Takeaways to Drive Success

These traps illustrate the importance of optimizing metadata management and adapting a data catalog to meet the specific needs of an organization. Each element has a direct impact on the overall success of the data catalog project and can put the initiative in jeopardy if ignored—or contribute value if strategically considered.

Depending on the nature of the organization, some topics can potentially require more work than others. That said, companies can steer clear of the main pitfalls if they keep the 10 traps in mind. Stakeholders for the data catalog project must contextualize each of the traps to work out the best plan of action and demonstrate that with a sound strategy and execution, they can achieve the full value of a data catalog.

About Actian

Actian empowers enterprises to confidently manage and govern data at scale. Organizations trust Actian data management and data intelligence solutions to streamline complex data environments and accelerate the delivery of AI-ready data. Designed to be flexible, Actian solutions integrate seamlessly and perform reliably across on-premises, cloud and hybrid environments. Learn more about Actian, the data division of **HCLSoftware**, at **actian.com**.

About the Author

David Martin
Head of Customer Success at Zeenea

As a software architecture and data specialist, David had the opportunity to accompany many companies throughout their transformations when he was an IT consulting director. Today, he is the head of the Customer Success team at Zeenea, where he accompanies clients in their approach to the valorization and democratization of data.