



4 Mistakes When Implementing a Data Catalog

Building and implementing a data glossary are two common goals for many investment firms in their annual strategy setting. But what happens when you rush into it without proper planning? Four common mistakes can occur which can hamper your development or make the final product irrelevant. This paper will discuss these missteps and offer prevention strategies.

MISTAKE 1: DESIGNING A GLOSSARY UNFIT FOR INTERNAL BUSINESS CLIENTS

When creating a data glossary, the goal is to incorporate business terms the business typically uses. Loading vendor data dictionaries to your glossary may be ideal for developers, but the business has its own vernacular that it uses to identify data needs. Not taking the time to develop a glossary in the business' own words is a common mistake and can ultimately translate into low usage of a data glossary.

MISTAKE 2: NEGLECTING STRUCTURE AND METADATA FOR CONSUMERS

Simply relying on a search function, or a single hierarchy, in your data glossary is another common mistake. Look to retail websites for how they allow customers to view products in multiple paths, such as a primary hierarchy, 'seasonal' groupings, or via product spotlights, and other hierarchies of a client need. An investment firm should consider how the internal users think about their data, group it for their various purposes and use specific concepts or categories for searching through their data. These inputs are crucial to help identify the structures and metadata needed to optimize the data presentation for the internal clients. One common best practice is to mirror your data governance structure along your data architecture for an initial view and use that as a base hierarchy, then expand to business-specific hierarchy overlays.

MISTAKE 3: OMITTING GUIDANCE ON ORGANIZATIONAL PREFERRED DATA AND WHERE TO GET THE DATA

Investment managers frequently report that their glossary has overwhelming amounts of data, and it's too hard to identify the right data to use. We need to ensure people access the correct data, not the data that's easiest to access. The top two questions of clients typically are: "What is the preferred data that I should use?" and "Where can I get it?"

It is common that firms do not identify gold-standard, or trusted, data and label it as such in their data catalog. A best practice is to create preferred/managed/gold standard/trusted data source tags for the consumers and implement them. To answer the second question of clients "Where can I get it?", lineage must be created. Data fields should be traced through systems and normalization back to the vendor origin, or the ultimate vendor origin. An example might be benchmark data, pulled through an aggregator. Establishing lineage is often done manually and becomes a single point in time documentation. This grows stale with time if not updated, and the data glossary becomes inaccurate, leading to distrust and less usage. The best practice is to link the data fields and business terms back to the original vendor or internal origins and identify policies and processes to either update the lineage regularly in a manual fashion or implement technology to dynamically track and update lineage.

MISTAKE 4: LACKING THE PROPER RESOURCING FOR DATA GLOSSARY IMPLEMENTATION

Firms can often underestimate the time and work needed to form a data catalog and implement it across a division or company. The effort involved in forming a glossary demands time from the data office including the data owners, stewards, and subject matter experts.

4 Mistakes When Implementing a Data Catalog

It also demands time from technology departments, including roles such as technology stewards, IT data architects, and senior leaders. A full data glossary implementation and development takes time. A minimum viable product can be achieved in year one by cataloging the data and building a base lineage. If that is accelerated, then base hierarchies could be layered in the glossary, metadata to assist in searches could be created, and trusted data tagged. Year two deliverables could include additional hierarchies for finding data, implementing lineage technologies, or developing layers of data governance technologies.

Full data catalog implementations are rarely successful within a short timeframe due to competing business priorities. In these cases, results are often partially completed due to a lack of funding in successive years. The best practice is to establish a vision for what the business wants from a data catalog, then build a multi-year roadmap to achieve it. This includes assigning it to a durable team to see it through, seeking funding for each year, and setting deliverables in the roadmap.

HOW MERADIA CAN HELP

To successfully deliver the best value for your company in a data glossary implementation, Meradia can serve as your partner to help you:

- Develop a data, performance, and reporting strategy.
- Identify appropriate vendor partners, tools, and platforms.
- Understand data catalog implementation best practices to avoid common pitfalls.
- Integrate and transition to production.



Andrew R. Jacob, CFA

Andrew Jacob brings a well-rounded data management perspective to Meradia's global portfolio of clients. Throughout his career, Andrew has supported constituents across the investment management industry, including retail investors, advisors, wholesale teams and institutional investors. He is a seasoned investment professional with diverse experience in business and entrepreneurship, possessing strong acumen of front office operations, data governance and management, and retail client journeys. Andrew's passion for helping people and driving outcomes aligns with Meradia's principles of Perspective, Passion and Impact.

Meradia
119 North High Street
West Chester, PA 19380
Phone: 610-738-7787

Info@Meradia.com
www.Meradia.com



© 2022 Meradia. All rights reserved.