



Revolutionize Investment Operations With AI-Driven Data Catalogs

SUMMARY: Are you looking to stand up or revise a data catalog? Is understanding your data one of your company's goals or principals? Understanding the full scope of your data involves a top-down understanding and a bottom-up inventory. This provides valuable insights into your current state to enable multiple efforts to be launched at your firm, but who has the time? The new methodology developed using AI technology and industry expertise reduced the time and client resources required to inventory data and stand up a catalog to a fraction of the normal amount. In this paper, we share the methodology and tools to help you achieve your firm's goals both effectively and efficiently.

WHY DO COMPANIES NEED A DATA CATALOG?

The foundation for data governance programs and data catalogs helps both IT and the business itself better understand an organization's data.

With greater understanding from data catalogs comes increased confidence in the organization's data, as well as the ability for users to find it faster. A catalog is vital in the self-service data paradigm moving across the industry as it provides standalone value and secondary value for machine learning and AI learning models.

But for all the benefits and positive outcomes, our clients often come to us lamenting the time, resources, and funds sunk into a data catalog they established. Data catalogs are slow to produce for a myriad of reasons. The benefits and outcomes fuel the forward momentum, but the volume of data inventoried, and the time and key resources needed often cause delays.

LONG LEAD TIMES PLAGUE CATALOG DEVELOPMENT

Catalogs are typically slow to produce due to the detailed work. Software designed for data catalogs requires significant manual work to prepare for them; even well-established data governance programs with strong executive sponsorship struggle to stand up a data catalog from scratch. Factors include scope definition, engaging/coercing stewards, and leveraging subject matter experts to supply inputs to the data glossary, as well as resourcing, configuring catalog platforms, and normalizing input data. It takes firms anywhere from 18 to 36 months to produce a complete enterprise data catalog. Some firms never make it to the end, and the project is canceled. Some catalogs are never rationalized, generate little value, and are scrapped.

Time to market for internal ideas is key to delivering value. The faster you can get to the first iteration, the faster you can prove concepts and show the organization what can be accomplished. Imagine delivering your firm's first iteration of a data dictionary/inventory in 90 days or less. Imagine being able to use the catalog to find data faster and build new strategies that provide an information advantage to your company. With this new approach, you don't lose momentum, you increase it.

In a recent Proof of Concept (POC), **Meradia**, the leading investment management operations consulting group, partnered with **Zengines**, a technology software vendor using the latest machine learning techniques and algorithms to accelerate data analysis and migration, to fast-track the fit-for-purpose build of a data catalog from the ground up. The use of this emerging AI technology coupled with an innovative consulting approach drastically improved time to market for the first iteration of a data catalog in 4.5 months.

In this paper, we will outline the challenges we have seen with other companies, the POC we conducted and the result. We feel this new approach will benefit many other clients.

Revolutionize Investment Operations With AI-Driven Data Catalogs

TRADITIONAL DATA CATALOG FORMATION CHALLENGES

Data catalog formations typically involve several challenges, and each derail or completely delays a data catalog implementation. We kept these challenges in mind as we launched our POC:

Balancing too little data with too much data and the need for data rationalization. Too little data will not give users the data discovery benefit. Too much data may overwhelm users with too many choices. Too much data also quickly scales up the amount of manual work required to rationalize, deduplicate, and define data, and rushing these steps often degrades the quality of the final product.

The time and manual efforts involved in data discovery, identifying all the internal data needed to form the base inventory of a catalog.

Large blocks of time are needed from key internal resources such as SMEs, Stewards, Governance teams, and the IT department. Dedicating time to standing up a catalog removes this staff from other projects while putting strain on their ability to fulfill their day-to-day responsibilities.

Time to procure, install, learn, configure, and populate vendor software. Project teams have to balance learning new systems with wrangling data to load to it, along with their daily tasks.

Time and manual efforts involved in capturing, standardizing, and loading of base data. Catalog inventories may begin with thousands of rows of data and grow to over ten thousand rows of data. Normalizing, deduplicating, and organizing the data are all enormous efforts as each piece of data must ultimately be reviewed line by line.

Competing priorities and agendas of project teams and stakeholders can often derail efforts to install data catalogs. Executives may deprioritize the data catalog effort in lieu of higher priority projects, especially when ROI is not realized quickly, leading to data catalog efforts dragging on over time.

A MODERN TAKE ON DATA CATALOG FORMATION

To overcome these challenges, our approach started with identifying the key databases, a smaller subset of data and reports used by departments. We leveraged Zengines' AI technology to quickly ingest file inputs and outputs into the key documents. Our team proposed a meaningful taxonomy to organize the data. The consulting and technology approach kept stakeholders, project team and sponsors engaged through quick wins and iterations.

THE CONSULTING APPROACH DEPLOYED BY MERADIA

Meradia's approach deployed experts with CFA, CIPM, CAIA designations, with experience in front to back offices, performance, data management and data governance. The approach focused on a specific use case for the client and utilized a rapid iteration model. By doing so, Meradia reduced the large blocks of time needed from SMEs, Stewards, Governance teams and the IT department while maintaining a high level of quality.

In other recent data catalog projects, up to 30 SMEs, Stewards, IT department staff and Data Governance staff were involved, with 15-30% of their time absorbed per week in the build and refine stages.

- The new approach removed significant burdens from the IT Department, freeing them up to maintain day-to-day operations and project work
- It also reduced SME, Data Governance, and Stewards involvement by over 50%, and the amount of their time needed per week down to 0-10%.

Meradia deployed catalog methodologies to standardize and rationalize the large amount of data. The process transformed raw data through five phases of development, with each transforming the raw data closer to the desired outcomes.

The phases consisted of:

- **Phase 1-2** developed the catalog with metadata, structure and relations
- **Phase 3** back tested the catalog with findings from earlier in the project
- **Phase 4** put the catalog in action to the user group, and consisted of user testing
- **Phase 5** aggregated the findings from the testing and back testing to make final edits until the completed final data catalog and data inventory were handed over to the client

Revolutionize Investment Operations With AI-Driven Data Catalogs

KEY POINTS

Throughout the five phases, Meradia:

- Defined the data inventory and catalog data requirements and quality standards based on best practices in the industry, and the Meradia Data Catalog Accelerator
- Utilized Meradia's Data Catalog toolkit comprised of Excel and Python to rationalize the data
- Project managed the deliverables based on a shortened schedule of initial development with minimal client involvement
- Utilized rapid iteration and feedback cycles in the last 1/3 of the timeline until completion
- Had domain and data governance experts provide feedback to the Zengines model through phases 1-2 to help further enrich the model, define the catalog to fill gaps and optimize the customization for the client
- Developed a focused taxonomy to group and organize the data in the inventory and glossary
- Used information gathered in an earlier state of project that mapped important functions to data sources to identify 'critical data' used to make decisions
- Further refined the scope for data catalog after feedback was received to ensure it was the best fit for the client

Upon completion in Phase 5, the value to the client was immediate. The client received a data inventory and data catalog, and Meradia and Zengines spotted key opportunities for improvement. Results like this provide a small and iterative approach to building an enterprise catalog or for a fit-for-purpose catalog based on use cases such as transformations, or mergers involving multiple system conversions.

For every company, time is money. Big projects rarely finish on time or on budget, and the longer they take, the more likely you are to rush and miss key opportunities.

ZENGINES' ROLE IN THE POC

The Zengines software, using its Analyzer capability, offered various benefits for creating a data catalog. From lineage analysis to automated classification, flexibility in taxonomy, and the ability to extract data from reports, Zengines offered accelerated cataloging and data analysis solutions. Its ability to derive metadata from actual data and identify unique client terms makes it a valuable tool for asset management organizations.

- A pre-seeded baseline of existing data catalog that leveraged pre-existing data catalog templates and industry standards to jump-start the process. For example, the Financial Industry Business Ontology (FIBO) is a widely recognized industry standard that defines a common language for financial data. By starting with a baseline, an organization can accelerate creating a data catalog and ensure consistency across all data assets.
- Extended lineage analysis to Excel spreadsheets and macros enable the user to track the data source and how it has been transformed or manipulated through different processing stages. Additionally, Zengines allows the user to catalog derived data and calculations, ensuring that important information is not lost or forgotten over time.
- Automated data classification significantly eliminated manual classification and saved time and effort in cataloging. Furthermore, flexibly defining or changing a taxonomy allows users to tailor the data catalog to their needs, ensuring it is a valuable and customized tool.
- Data extraction from PDF reports saved time and effort in manually transcribing information, ensuring that essential data was not overlooked or missed. Additionally, the software enables users to merge data from databases, spreadsheets and reports for analysis without altering or manipulating the actual data.
- Deriving metadata from actual data automatically identified data types and provided additional information about the data, making it easier to understand and analyze. Baseline terms, definitions, and acronyms were also included, enriched by Meradia's experience, ensuring that all users use consistent terminology.
- Time to production with reduced client work effort. Using Zengines to create a data catalog accelerated the initial catalog creation process, saving time and effort. It also does not require extensive interaction with IT, allowing users to develop and maintain the catalog independently.

Revolutionize Investment Operations With AI-Driven Data Catalogs

Traditionally, creating a data catalog has been a manual and labor-intensive process, requiring significant time and resources to identify all relevant data sources, document metadata, and establish data lineage. However, a new software-driven approach can help organizations create a data catalog more efficiently and effectively.

WHAT CLIENTS GET (VALUE PROPOSITION)

Using a professional consulting organization paired with AI-driven intelligence is a unique and valuable approach to building an enterprise data catalog, helping to manage and govern your firm's data efficiently. With this approach, you can:

- Seamlessly integrate a comprehensive data inventory into your enterprise catalog, including a preliminary taxonomy and fit-for-use data glossary
- Gain valuable insights into your data structure, management, and governance
- Enjoy the benefits of a lower-cost solution to address your governance program's first use case or integrate with your existing catalog

This approach offers considerable value that can translate to other use cases and potential outcomes and increases the chances of developing a valuable catalog with little risk of extending timelines. It offers the ability to prove the value of the catalog and analysis generated before launching into a broader enterprise campaign. Completing the data inventory and fit-for-use catalog in a fraction of the time of a typical data catalog implementation saves you money and conserves your internal resources.

This approach leverages AI and Meradia domain expertise to bridge gaps, generate value, and reduce the burden. Expanding this approach to create an enterprise data catalog would take additional time but would have similar savings as the initial project. Still, the output could be the basis for populating vendor data catalogs such as Collibra or Data.World.

Furthermore, this approach's data insights and details exceed many common attributes in these vendor products. Examples of this include providing return and analytic depth of the data, identifying where the data is part of Excel formulas, or result of formulas, and clear delineation of created data or vendor provider data. The process for creating the deliverables uncovers insights both at a data level and within the metadata. As a result, your company can enrich vendor data catalog products and benefit from a more valuable experience. Don't miss the opportunity to use this unique approach to building your enterprise data catalog.

OUR FINAL THOUGHTS

Time is money. Since projects in governance don't add to the bottom line of a company, creating quick value can help justify costs. A catalog and inventory can provide value to many companies in the industry, including yours. Understanding your data, knowing your data, and identifying the lineage are pressing needs within the investment industry. With outdated catalogs, inventories, and taxonomies starting to pervade the industry, it can be time for an initiative such as this to refresh your data inventory or catalog and provide new insights and perspectives.

My colleague, Jose Michaelraj, will be publishing the next paper in this series detailing how this approach transitioned into a transformation project and helped accelerate those efforts.

To learn more about the approach, or the technology used, contact Meradia and/or [Zengines](#). Our companies continue to work together to innovate in the investment operations sector. We believe that an eclectic mix of consulting methodologies and AI/ML driven software capabilities is a novel approach to solve existing problems.

Connect with the POC Leaders:

[Laurie Hesketh](#), Meradia, Managing Director

[Jose Michaelraj](#), Meradia, Manager

[Kevin Coffin](#), Zengines, Head of Customer Success



Andrew Jacob, CFA 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.

Kevin Coffin, CIPM is the Head of Customer Success at Zengines. With extensive experience in investment management operations transformation, Kevin is a seasoned expert in investment accounting, data governance/management, performance measurement, and reporting platforms. Throughout his 30-year career, Kevin has specialized in delivering end-to-end investment systems solutions for investment management, pension fund, and insurance companies. His hands-on expertise encompasses a wide range of areas, including alternatives, derivatives, and fixed-income processing. Contact: kcoffin@zengines.ai