Xpansion

Xpansion Whitepaper

Managing Data Assets: The Importance of Data Discovery and Metadata Management

December 2022

Contents

The Importance of Data Discovery and Metadata Management

In this paper, we look at some of the problems that firms face with finding, classifying and exploiting their datasets. We then look at data discovery as a solution and we present Xplore, our platform for data discovery and metadata management.

Data Management Challenges

In today's digital world, firms generate and consume increasing amounts of data that originate from and go to a multitude of complexly linked internal and external data systems. With this abundance of data, various data management challenges arise.

Time Lost Looking for Data

Employees spend valuable time looking for the data they need. A 2022 report by Coveo[1], a provider of an AI-powered relevance platform, found that the average employee in a digital workplace spends 3.6 hours every day looking for information. This time could be better invested in using the data, analysing it and drawing insights to drive the business forward instead.

Duplicated Data

When data is not properly catalogued or inventoried, often spread across silos and departments, it is very hard for data consumers to find and leverage existing datasets. Subsequently, they end up duplicating data by recreating it or purchasing it from vendors multiple times. A 2018 report by IDC[2], a global market intelligence firm, estimated that data professionals on average waste 20% of their time every week building information assets that already exist.

Siloed and Departmental Data

Data is typically held in silos and is not easily or entirely available to other departments or teams. This impedes data sharing and collaboration across the firm. In turn, such a data culture limits the ability of

[1] Workplace Relevance Report 2022 - Are Employees Driven by Information, or Stalled? | Coveo AI
[2] State of Data Discovery and Cataloguing

employees to exploit enterprise-wide knowledge and hinders the top managers from grasping a clear overview of the firm's global data landscape and the data assets contained within the organisation.

Poor Data Quality

Siloed data, together with the data duplication that usually results from it, give rise to inconsistencies in the data which compromise its quality, integrity and trustworthiness. Data then becomes stale and unreliable, causing the waste of valuable time and money for a lot of organisations.

Non-Compliant Data Access

Even with the best intentions to do the right thing, firms face challenges when trying to comply with all the regulations around data access and usage entitlements. Besides substantial fines and losses arising from non-compliant use of data, considerable resources and time are wasted when gathering the information required by regulators performing routine audits or investigating incidents.

How do firms tackle these challenges?

The responses of firms to these data management challenges span a wide, varied spectrum. Some companies have fully implemented third-party solutions (including data catalogues, data governance and data discovery platforms) and others have resorted to in-house developed data management and data permissioning tools. Some organisations have adopted external systems partially and tried to complement them with internal solutions. A growing number of firms are devising plans about how to best tackle these problems, yet others have not taken any measures and have no plans to do so in the foreseeable future.

How successful have all these measures been and what is hindering firms who have not taken any action?

Many solutions to the data management challenges have been provided by different vendors and they usually offer some form of one or more of these functionalities and capabilities:

- Data discovery
- Data cataloguing
- Data classification
- Metadata management
- Data lineage
- Data privacy and compliance

- Data governance
- Business glossary

Companies that have adopted these systems report different degrees to which their expectations were met.

Generally, data management tools that are rolled out satisfactorily and that result in a high ROI are characterised by:

- System simplicity and ease to use
- Seamless integration with other systems
- Scalability
- Support of a multitude of data sources

For a successful deployment, these characteristics are accompanied by an awareness of the importance of coupling the technological solutions with a cultural evolution in the firm. This includes instilling and fostering a sharing and collaboration mindset and incentivising the use of the tools to gain user buy-in.

On the other hand, system complexity, limited integration and scalability, limited support of data sources, along with the lack of a sharing and collaborating data culture compromise the ROI of the solutions and are also the reasons behind the inability or unwillingness of some firms to tackle the data management issues they face.

How to increase the value of a Data Discovery Platform?

Low user adoption can be a major impediment to the success of a data discovery platform. The value of such a tool depends highly on the metadata management capabilities it offers, in particular the user-added metadata. In many cases, low user adoption is caused by a rather paradoxical situation. Users will rely on the data discovery platform if it offers them more information (especially in the form of user-generated metadata) about the datasets than they already know. However, this highly sought-after information gets populated in the platform by users utilising it. Thus, it is critical at all deployment stages of such a solution to keep its functionality practical, simple to use and relevant to the business.

Here are some ways a data discovery platform can help mitigate and address some of the data management challenges while remaining practical, relevant and easy to use.

Data Discovery

Invaluable time and resources are spent looking for data and quite often data gets duplicated. A data discovery tool allows users to register data sources and crawl them to discover their datasets. This brings to the table valuable data assets that would otherwise remain hidden and untapped. For data discovery to be impactful, it must operate intelligently in the background and must be a combination of automated and manual discovery.

Data Classification and Taxonomy

When data is properly and meaningfully classified, it becomes easier to eliminate redundancies and reconcile inconsistencies. Moreover, searching for and browsing available datasets becomes simpler and more efficient. A data discovery platform allows building a customised data classification where datasets can be contained in a meaningful way that is contextual to the type of data and its consumers. It also promotes the use of tagging to improve categorisation and make data searches more effective.

Data Access and Data Usage Reporting

A crucial problem that data discovery solutions help with is the management of permissions and access rights to datasets spread across various data sources of a firm. Through the data discovery platform, users browse discovered datasets, utilise those they are authorised to use and finally request access to any interesting datasets they encounter for which they do not have permissions. Data discovery solutions also define the capacity in which a dataset can be compliantly used. For example, some data might be only available for internal use while other data might be available for external publishing or even re-distribution. This information about the compliant usage of the data is flagged at the point of consumption to remind users of permitted usage policies.

By managing permissions, the platform can also gather all the data about the usage of datasets, allowing it to become a reporting hub capable of answering questions around who accessed different datasets, the frequency at which certain datasets were accessed, which datasets were most popular and the like.

Data Integrity and Reliability

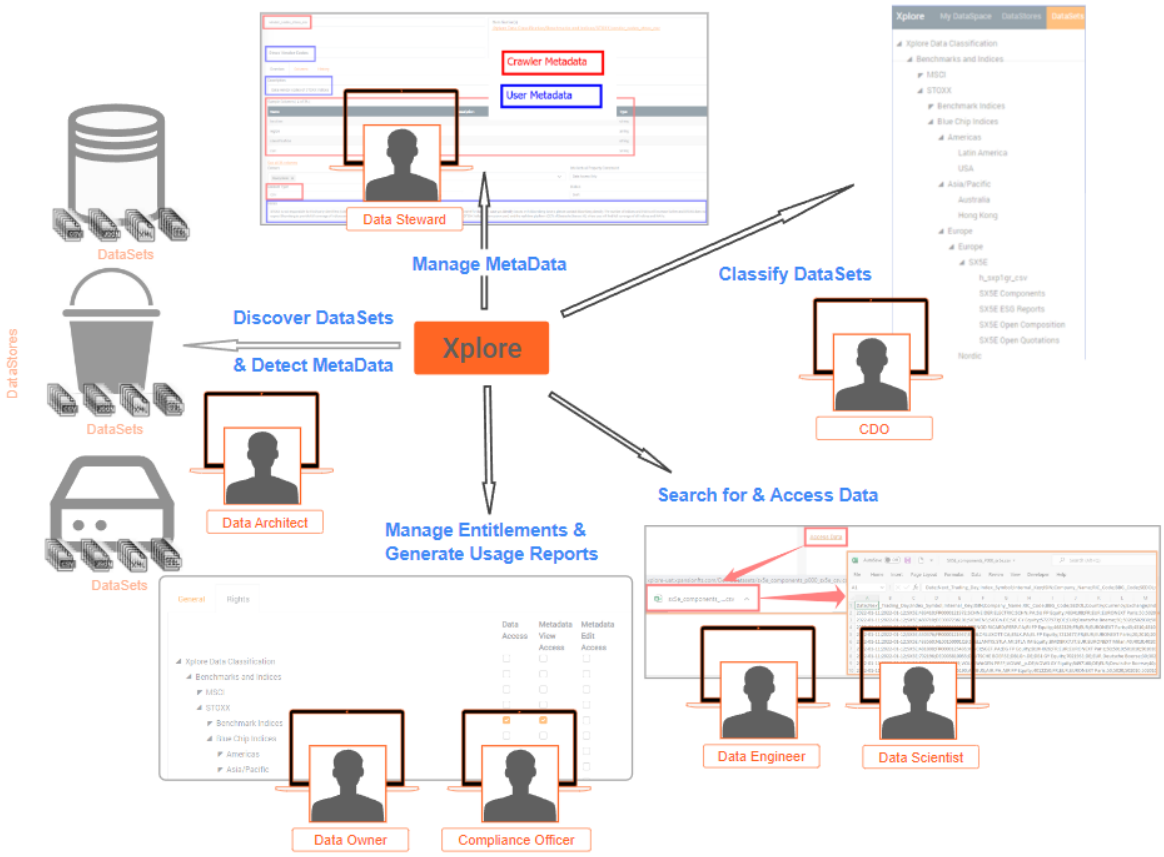
Inaccuracies and inconsistencies in data waste valuable time and money. They also compromise the trustworthiness and reliability of the data, which can sometimes be hard to restore. The data discovery platform detects and flags anomalies and changes in datasets, such as new or missing columns in a daily csv file or a change in a data type. This prevents costly effects in the downstream pipeline by allowing data owners to take proactive corrective measures before data is ingested and processed.

Data Location

A data discovery tool provides a framework that links data sources and data repositories without the need to migrate or move the data from its location. A data discovery solution allows data to remain in its original place while providing a layer of discovery and metadata management capabilities. This is critical to avoid complex, risky and lengthy data migration projects.

It is very important to note that, while all these functionalities benefit from an automated aspect, it is instrumental to combine automation with diligent human effort and validation. Automation offers speed, accuracy and repeatability but cannot take a data discovery tool too far without validation by data stewards and enrichment by data users.

Xplore: A Practical, Simple and Powerful Data Discovery Platform



Data Discovery

Xplore indexes data sources and discovers their datasets, thus unveiling valuable data assets. Firms can easily map their data sources to Xplore DataStores and launch an automated crawling mechanism that returns all the datasets with their technical metadata. Indexing is automated and can be scheduled to run automatically in the background to detect new datasets.

Metadata Management

Xplore stores both technical and user metadata, creating a bridge between data and its meaning. Users can enrich the metadata of datasets with an extra layer of user-defined metadata capturing the human knowledge that would otherwise remain undocumented. Data stewards can also deploy custom metadata fields to describe datasets with more relevant metadata that would in turn make searching and filtering more efficient.

Inventory and Classification

Xplore enables firms to build a Data Classification tailored to their business context, providing meaningful organisation and navigation of their datasets. A flexible, user-friendly tool allows data stewards to construct a meaningful and contextual inventory for their datasets.

Access Rights Management and Usage Reporting

Xplore allows firms to control users' access rights to datasets and tracks and reports on data usage, ensuring compliance with data governance policies. The management of permissions can be done at any level of the data classification allowing rights to be granted and revoked in a very granular way.

Data Search

Users of Xplore can use natural language to search the metadata of datasets, including the technical metadata such as column names and user-generated metadata such as notes, descriptions and tags.

Concluding Remarks

In an increasingly digital data world, managing data as an asset is a must and poses real challenges for data operations teams. Data is the lifeblood of any organisation and flows within a complex link of systems, users and sources. The digital nature of data and complexities around usage, licensing and permissions make managing data assets a difficult task. Efficient organisations have implemented data architectures that incorporate solutions for data classification, data usage transparency and data collaboration.

Xplore is a data discovery platform which provides firms with excellent visibility of their data landscape. With Xplore, users can register their data sources, crawl them to discover their datasets, classify these uncovered digital assets into a bespoke relevant data classification, manage their automatically generated metadata and enrich it with valuable human-generated metadata, as well as manage user entitlements to datasets, flag usage restrictions to ensure compliant consumption and provide useful usage reporting. Xplore offers a pragmatic approach to data discovery and metadata management, combining automated and manual functionality to help solve practical, real-life use cases. Built by a team experienced in reference and market data, Xplore is designed with a special focus on the financial services industry and the regulations around it.

If you would like to enquire about Xplore or request a product demo, contact us or visit our website at https://www.xpansionfts.com/.

Xpansion

Audley House
13, Palace Street
London SW1E 5HX

+44 (0) 203 740 2951