

Tenets of Successful Data Cataloguing for Life Science Organizations

Implementing a comprehensive data catalogue solution to help overcome business challenges

INDERPREET KAMBO, Principal, IM Practice, IQVIA

PHYLLIS IMPARO, Practice Lead, Data Governance and Strategy, IQVIA

JAY DZWIL, Sr Principal, Data Governance and Strategy, IQVIA

As businesses perform the increasingly complex function of collecting, storing, and securing their data and intellectual property needed to manage their day-to-day operations, data governance has evolved into a necessary internal auditor that is critical for supporting an organization's overarching data management strategy.

One key component of an effective data governance process is data cataloguing.



Data cataloguing can be defined as the library of metadata for an organization's data products (internally and externally sourced) including reports, combined with the governance, policies, workflows, and required business context necessary to facilitate the democratization of data.

The goal is to create a connection of data producers and consumers in a safe and compliant manner while providing data transparency across the organization.

A data catalogue includes three critical elements:



COMPREHENSIVE METADATA

Metadata broadly includes the actual definitions of the data products, their associated lineage, usage guidelines, and the contacts that own and manage those data products. To work as an enterprise solution, the data catalogue should be able to ingest metadata from a variety of data repositories and represent a complete view of the data used by the organization.

There are many solutions that provide a "catalogue" of the data within a particular repository, but their scope is often limited to the data that it actively manages. Having multiple catalogues for the user to find the data they are looking for does not provide the positive user experience that most organizations are seeking. When a user does not readily find what they are looking for in an easily consumable fashion that will limit their adoption and use of the catalogue solution, so the user experience needs to be considered carefully.



GOVERNANCE PROCESSES AND WORKFLOWS

Making your data widely visible in a catalogue solution requires a well-defined governance process to ensure that your data products are kept current and compliant with policies (internal and external). The definition of the processes that are applied to the data products, the policies that govern the data products, and the workflows that manage the data in your data catalogue, are all a part of the governance process.

An effective catalogue should provide a clear understanding of how data is or may be used and guide a user through the processes required for each individual data product. Keep in mind that governance is often seen as burdensome, so any opportunity to reduce the overhead or and mitigate that burden throughout the process should be considered.



USER EXPERIENCE AND INTERFACE

A primary key to ensure stakeholder adoption of a data catalogue solution is user experience (UX). It is important that a user can find what they need when searching for and navigating data or participating in workflow activities. It is critical that the user can find the data they are looking for without having to wade through a lot of irrelevant information, that they can find any data that they know exists, and that workflows do not become cumbersome for those involved in the processing.

UX also underscores the importance of considering all user populations when establishing a catalogue solution. If you have a catalogue full of great metadata and well defined processes that is well maintained but no one uses it, it is not providing the value to the organization that is expected. Too often, catalogues are built considering only a subset of the organization and then fails to meet the larger organization's needs, resulting in very low adoption rates. This outcome typically seen as a failure of the catalogue solution.

It should be noted that a data catalogue solution is an enabling tool; it does not automatically create processes and governance where none previously existed. Rather, the organization should look at the data catalogue solution as the method for automating and better

controlling processes and adjust their existing processes as they get closer to implementing a data catalogue solution. Keep in mind that there will be a transition period. Users will need support from internal advocates and time to adjust and adapt to utilizing a new tool that manages processes that may have been manual in the past.

Implementing a comprehensive data catalogue solution can help overcome some typical, but significant business challenges with data:



COMMON UNDERSTANDING OF THE DATA

It is not uncommon for there to be fundamental disconnects between and amongst different technology and/or business teams about the meaning of data within an organization. These disconnects can contribute to the creation of reports with different results for what should be the same metric, the purchase or creation of multiple, yet similar data products, or misusing data because they do not truly understand what it represents. All of these can have ripple effects across the business and waste time and/or resources that could be better used to solve other business challenges.



KNOWING HOW DATA FLOWS AND IS USED (DATA LINEAGE)

Organizations can stumble when there is a lack of transparency on how data flows within the organization. A prime example is when changes are made to a data product without a clear line of sight into every area that may have been impacted. This scenario can cause outages to applications or result in reports being produced with inaccurate information and the consequences can vary depending on the degree of change and the breadth of use of the data. It is important to have a clear process with documentation of how the data is used and to enable a proactive communication protocol to avoid these incidents from occurring. To that end, the right data catalogue solution will enable tracking and auditing of who has access to data and the usage to not only enable a better understanding of the data, but also solve critical compliance requirements.



UNDERSTANDING OF THE DATA PRODUCTS THE ORGANIZATION POSSESSES

There are several dimensions that are part of this challenge ranging from a simple “inventory” of data products, to understanding the data product and its applications, to who the expert is that stakeholders can go to with questions.

When organizations lack a central place for individuals to find out what data products exist in the organization this can lead to the same data being purchased multiple times or a creation of duplicative data products with subtle differences. Typically, the IT team will deploy a data catalogue that contains all the physical data dictionaries, and maybe some data lineage, and then consider the effort complete. While that may serve basic IT needs, it fails to meet the business user needs, and will reduce the value of the data catalogue.

Cataloguing the data dictionaries is relatively simple since most data catalogue tools have processes that automatically harvest data dictionaries from many common technology platforms but gathering the business language, context, roles and responsibilities, and business perspective means gathering information from individuals within the organization. This is frequently a manual effort, and often not prioritized or resourced as a fundamental part of the data catalogue journey. Having this information, however, builds a more complete data catalogue and will ultimately provide value to the entire organization.



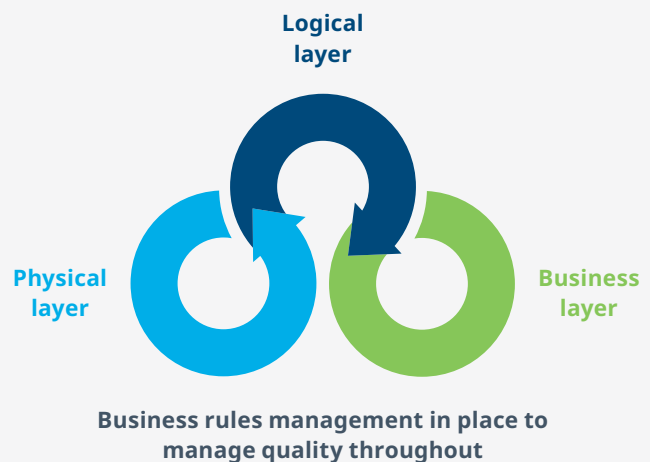
The three key characteristics of an effective data cataloguing solution for the life sciences industry

While these are not all the characteristics of an effective data catalogue solution, they represent some key characteristics that are important to organizations in the life sciences industry.

1

DATA VISUALIZATION — A COMPLETE PICTURE OF YOUR DATA ASSETS (BUSINESS AND IT)

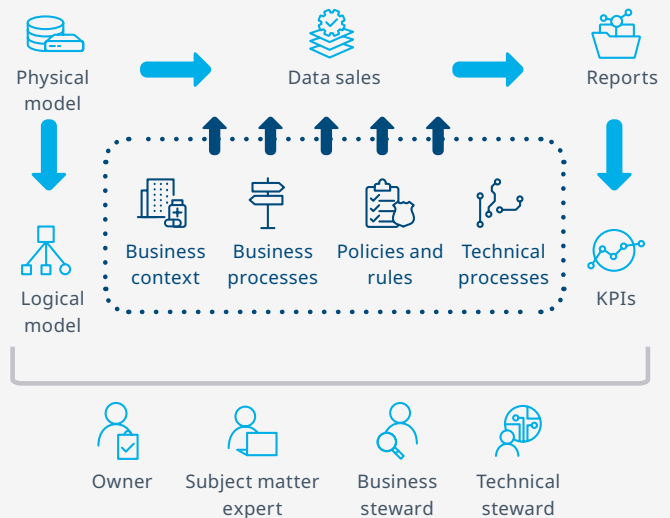
One of the core components of a successful data cataloguing solution is to seamlessly describe the physical, logical, and business relationship of various data assets. This is done most effectively when visualized in a graphic manner that can be rendered based on the user’s role. This component is essential to make data easily accessible, digestible, and available to a larger community, and to help in the recognition of value from the data catalogue solution. Documenting the physical data dictionary, logical data dictionary, business glossary, and their inter-relationships with layers to explain the business context, rules, processes, or regulations effecting those data products becomes a foundational step in an enterprise data-democratization strategy.



2

DATA QUALITY AND TRANSPARENCY

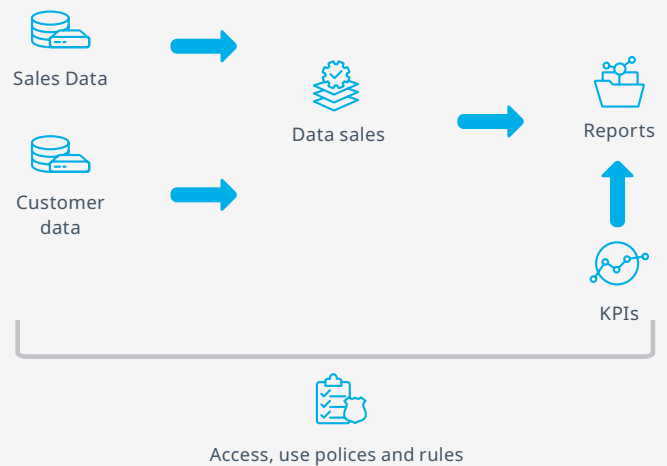
Users searching for data products want to know the quality of the data, what validations were done, who owns the data, and where else the data is used. Providing the mechanisms for a user to understand how the data is handled and who to contact with questions can increase their confidence in the data, help them select the right data for their use case, and ensure it is used consistently with its intended purpose. It can be valuable for someone trying to decide what data product to use to easily see other use cases for the data so they can be sure the data product is also the best fit for their needs. The leading data catalogue solutions will provide complete visibility of the data products to the end user, and while their initial view might be based on their role, they can navigate the relationships in the data as needed.



3

DATA CONTROLS AND MANAGEMENT

A clear line of sight into who has access to data and how data is being leveraged to drive business outcomes is imperative. A comprehensive data catalogue provides the mechanism to govern and manage data, the way to access the data, and use of data across a company ecosystem. The leading data catalogue solutions include workflow capabilities that allow for automation of processes, to reduce the dependencies on manual workflows, and provide an audit trail for future understanding of how certain assets got to their current state. A solid foundation of controls and governance is also critical to ensuring the management of data privacy and compliance for every organization.



In the next article, we will present information about how to know when you are ready to embark on a data cataloguing journey, or when you should try to relaunch the solution, you already have, along with some of the considerations for starting your implementation journey. Our experts will share helpful hints from their experiences with different implementations and how to leverage those learnings in your journey.



About IQVIA Connected Intelligence™

Connected Intelligence brings together IQVIA's unique portfolio of capabilities to create intelligent connections across its unparalleled healthcare data, advanced analytics, innovative technologies and healthcare expertise to speed the development and commercialization of innovative medicines that improve patients' lives.

Discover new insights, drive smarter decisions, and unleash new opportunities with the power of [IQVIA Connected Intelligence](#)