



# Choosing a Data Catalogue

Thinking about getting a Data Catalogue?

Confused about the aspects you should be considering?

Learn more about how to choose the best one for your needs...



If you want to use your data effectively, it's essential to know what it means, where it is, and who owns it. Whether you need to ensure you're complying with regulation or wanting to exploit the latest data innovations, a Data Catalogue can provide a central 'one-stop-shop' for the discovery, understanding and enabling of data assets across your enterprise.

But given the broad variety and functionality of Data Catalogue products available, how do you choose the best solution for your commercial and operational needs? We've produced this article to share our experience of delivering a successful Data Catalogue procurement process.

If you have any questions about our approach or would like to discuss how we could support your Data Catalogue selection and implementation, please contact us.

## What is a Data Catalogue?

A Data Catalogue is a centralised repository of information about the data in an organisation, like its definition, format, location, source, quality, usage, and who is responsible for that data from a business and technical standpoint. This information is also known as metadata.

Metadata can be used by teams to quickly find, explore, and understand data. It sounds so obvious, but it's often overlooked - we've worked with organisations where for example, two different definitions of "customer" have caused endless confusion and wasted time.

## Why you might want a Data Catalogue

Data Catalogues offer several benefits if you're looking to make the most of your data:

- **Data Discovery:** Data Catalogues make it easier to discover and explore data assets, helping you unlock the potential of your data.
- **Data Comprehension:** Data Catalogues contain well-formed descriptions, commentary and ratings of data, to better understand how data is relevant to the business.
- **Data Quality:** Data Catalogues can show the quality of the data, helping you identify and rectify issues, so data can be accurate, up-to-date and reliable.
- **Data Governance:** Data Catalogues can help ensure your data is secure, and compliant with regulations, for example by recording the provenance of data and its usage, helping you check it isn't being used for unauthorised purposes.
- **Data Efficiency:** Data Catalogues can help you identify and eliminate duplicate or unused data, reducing both storage costs and the time wasted managing it.

## Features to consider in a Data Catalogue

Data Catalogues are more than just a dictionary of business definitions and technical metadata. Most tools in the marketplace include a number of useful features:



**Search capability:** The central function of a Data Catalogue is to store information about your data, so you must be able to easily search and retrieve it.



**System compatibility:** Don't enter all of your information manually! The ability to connect to systems and automatically harvest metadata is crucial to building a comprehensive and meaningful catalogue.



**Team collaboration:** Your Data Catalogue will be a much more powerful tool if you can encourage your teams to share information, so it's important that people can comment and rate data assets, and add aliases and usage information.



**Flexibility & scalability:** The right Data Catalogue needs to remain fit for purpose as the size of your business, the complexity of your technology stack and the demands of your metadata model change over time.



**System integration:** Typically, a Data Catalogue is part of a larger effort to enable data in an organisation, so make sure you can integrate it with your other applications, like reporting and analytics tools.



**Security:** Data Catalogues will probably contain information about sensitive data, so you'll need user-level permissions for access control, encryption and other security measures.

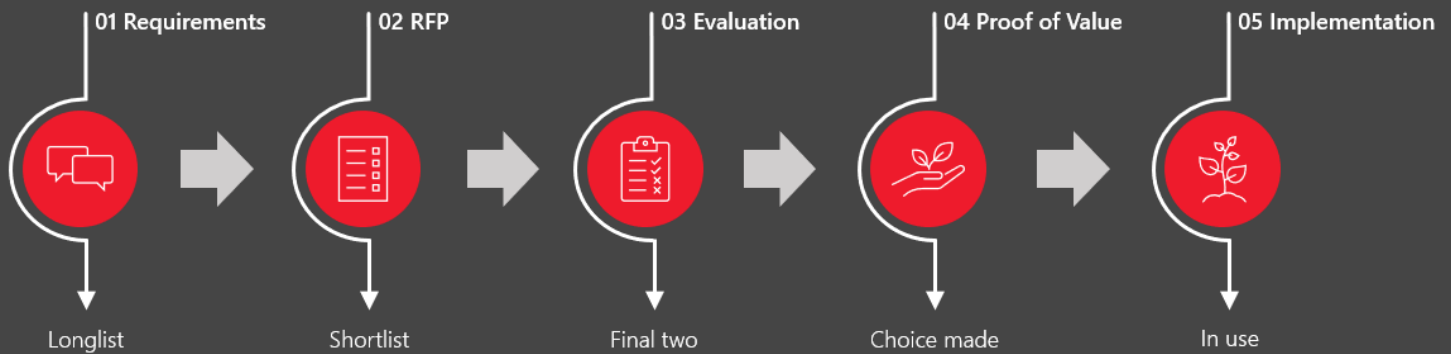


**Integrated Data Governance:** Cataloguing your data is only part of a data management journey. A Data Catalogue may have data profiling, data quality management and data governance tools embedded within the application.

Data Catalogues  
should enable people  
to enable data

## Getting started

Our typical starting point is to develop a Data Catalogue RFP (Request For Proposal) process. But make sure you have support for the initiative from senior stakeholders and an agreed business case before you begin.



We'll help identify stakeholders and gather requirements to understand the required budget, functionality and integrations, and identify half a dozen tools to send an RFP to. We find that some of the vendors will drop out at this point, as it's quite a lot of work for them to describe how their tool works for each of the requirements.

We'll have a set of measurement criteria and weightings to evaluate the submissions and deliver a balanced and objective assessment of the available solutions. This will narrow it down to the top two tools, and a Proof of Value will result in a winner. Then it's just a question of implementing the chosen tool and embedding new ways of working into your organisation.

**TOP TIP:** Look beyond industry analyst reports and opinions - a wealth of factors go into determining the right Data Catalogue for your current and future needs.

## Aspects to consider

### Vendor

Index evaluations, such as the Gartner magic quadrant, are a useful resource to show the market leading data catalogue solutions but there are other very capable catalogues out there that may better suit your requirements. Make sure you consider the vendor experience, support, and availability.

Ask questions like:

- Does the vendor have a good reputation and experience in the Data Catalogue space?
- What level of support do they offer (e.g. 24 hour customer support? Configuration and customisation support? API support? Cloud native storage?)
- How reliable are their systems? (This is especially relevant for cloud-based tools)

Understanding the vendor experience can help you avoid buyer's remorse and select the right supplier for your organisation.

## Pricing

The big one! Some Data Catalogues can be prohibitively expensive, and this may not always be obvious. Try to get a price up front and check that it includes ancillary services such as IT support, Cloud storage (if required), service subscription fees, optional extras, API (Application Programming Interface) support and other considerations.

## Cloud Services

The cloud makes data much more accessible and can provide real-time access to Data Catalogues for all your data enablement needs. When selecting a cloud Data Catalogue, make sure the location of the cloud suits your business needs. Data Protection regulation is tightening up all over the world and a cloud-based Data Catalogue should be able to provide you with the right protections for your data.

**TOP TIP:** If you're working with data in the cloud, consider a Cloud Data Management Readiness Assessment – it examines all facets of business readiness and can indicate areas where small changes can lead to big rewards.

## Consider Potential Data Risks

Data Catalogues can help organisations identify and manage data risks, such as data breaches, privacy violations, and data loss. Still, it's important to consider that they can also *create* data risks. Data Catalogues that exist in the cloud or which are accessed by multiple entities should be carefully considered and vetted by data privacy and security even before being selected as a candidate, since even a Proof of Value exercise could expose important data.

## Proof of Value pilot

The goal of the selection process is obviously to determine the optimal solution for your commercial and operational needs. In our experience, once you've reduced your list to 2 or 3 tools, it's incredibly useful to conduct a live pilot evaluation to allow you to gauge more accurately how the technology behaves with your data and aspirational ways of working. A Data Catalogue can be a big investment and it's essential that it works well within your business environment.

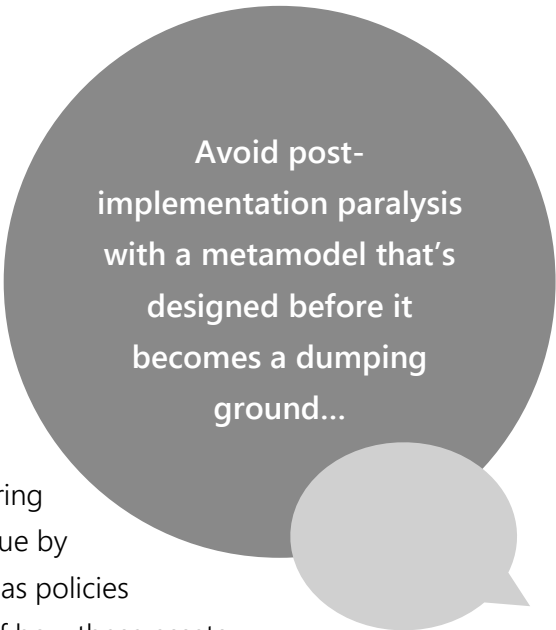
You may have heard of a Proof of Concept, well a Proof of Value (PoV) goes one better because it tests not only the essential idea of a Data Catalogue, but also the value and effectiveness of the features of the tools under consideration. Plus, it allows the wider stakeholder group to see the software in action and business users to start visualising how to incorporate the new set of capabilities into their processes.

## Implementation

Vendors are usually responsible for the installation and configuration of the tool, but we're often asked to help define and embed the adoption of new processes and practices that the Data Catalogue enables. There's no point having a shiny new tool if your metadata isn't accurate and your people aren't confident using it.

One of the big mistakes we find companies make after procuring their Data Catalogue is to try and demonstrate immediate value by using the product as a repository for lists of data assets, such as policies and technical metadata, before thinking through the model of how these assets will be organised. This unfettered storage approach leads to a 'dumping ground' mentality where vast swathes of metadata are ingested, stored and subsequently ignored.

We've successfully helped companies avoid this by helping build better business cases, ensuring details such as how the data will be ingested and modelled within the system are considered before they start.



Avoid post-implementation paralysis with a metamodel that's designed before it becomes a dumping ground...

## Next steps

Data Catalogues are an essential component of an organisation's data strategy and, as we have seen, require consideration and evaluation before selecting the right tool.

Need help? Ortechcha helps businesses select, implement, and adopt Data Catalogue solutions every day. To find out more please contact David, our Data Management Tooling expert.



**David Lee-Smith** is a Partner at Ortechcha, the specialist consultancy dedicated to helping companies enable and exploit data, serving clients globally from offices in the US and UK. He leads the Data Management Tooling Practice which operates a tool agnostic approach focusing on the best tooling outcome for clients. David has 25+ years of experience in database design, data architecture and data management.

Contact him at: [david.lee-smith@ortechcha.com](mailto:david.lee-smith@ortechcha.com)



the Data Management experts

[www.ortechcha.com](http://www.ortechcha.com)