

Synechron

Getting to grips with data: The metadata challenge



Authored by:

Tim Jennings
Technical Director
London



We talked before about Data Management Deficit Syndrome, and the problems it can cause. A key enabler to addressing these issues is metadata. But... what even is metadata? Why is it important and what does it do for your organisation?

Let's start with 'why'

Data is an asset, and asset management requires that you have key information about it: e.g. what the asset is, where it is, and what condition it is in. Put simply, you cannot control what you don't understand.

Imagine being asked to take responsibility for some data in your organisation (we will talk more about data governance roles in our next blog). To be effective, you will need to understand:

- What data am I responsible for?
- What is the correct meaning/use of the data?
- Where, and in what systems is that data held?
- Who is using it, and for what?
- What data issues exist – e.g. quality, policy adherence, etc.

There are other users and stakeholders across the enterprise who will have other questions: e.g. If data needs to change, what is the impact? Where is Personally Identifiable Information stored, etc.

It is the job of your metadata (literally data about data) to help answer these questions. It maps the data in your organisation and enriches that with key information needed to control the data



Business Glossary

The Business Glossary is the common reference point for data definitions across the enterprise. This is system agnostic. This is where self-service users might come looking to see what information assets the enterprise has available, or to understand the specific meaning of data they are using.

The Business Glossary contains all relevant business data items, typically arranged in a meaningful hierarchy. For example, you may partition your data into domains, which contain entities which are themselves a collection of attributes.

Why is this necessary? In a simple example, a trade may have an execution date (when the trade was carried out) and a trade entry date (when the details of the trade were recorded in an IT system) which may not be the same. Analysts anywhere in the enterprise must understand which is which, and this clarity is contained in the Glossary.

Metadata for each data item would typically include a business definition, a business owner and a data classification according to the data security policy, and may include references to policies or business rules like retention periods, GDPR applicability, etc.

Enterprise Data Model

The Enterprise Data model imposes a standard on the representation of data entities from the Glossary. This can be done in a variety of ways. Two common approaches are:

1. A Logical Data Model showing relationships between data items in the business glossary, and usually includes the logical data types and allowable values/constraints.
2. Defined messaging protocols, so that event streams or message payloads are standardised across the enterprise

The enterprise data model is typically a reference document for data architects and IT development teams.

Working from a common model ensures that wherever a data item is surfaced in the enterprise, it means the same thing and works in the same way. This can be a key enabler to introducing messaging platforms, and for aggregating data from across the enterprise.

Data Catalogue

The Data Catalogue aligns the Glossary with physical systems to show where business data items are persisted. It will show whether each source is a Master for that glossary item, a trusted source, or otherwise.

The catalogue is obviously useful for self-service BI (e.g. analysts asking 'where can I find...'). But is it also crucial to data-set aligned data ownership. To properly understand who is using and who has access to your data, you need to know where it is.

Data Lineage

Lineage represents the linkage of data items between systems. This enables traceability from the mastering source to the downstream systems and can include the lineage into reporting datasets and reports themselves.

Regulators are increasingly asking Financial Services firms to demonstrate that they understand and are in control of the lineage for risk calculations and the regulatory reports they produce.

Lineage enables proper impact assessment of data mastering changes. This allows both pin-point focussed testing and minimises unexpected consequences arising from data changes.

Finally, 'how' can a firm obtain and manage this information?

Capturing and maintaining the metadata can be a laborious manual task, and so a rich variety of vendor tools have emerged in this space over recent years

Our experience suggests that they each started from a different component of the metadata universe, and so each has strengths in some areas over and above the others. Some are strongest in Glossary management, some in Enterprise Data Modelling, some Lineage visualisation, etc. Almost all tools offer a basic ingestion and pattern match capability to jumpstart the process of building your metadata model from the existing systems, and there are some that focus on bringing more capable AI to the classification of the data. We are seeing some evidence of 'best of breed' partnerships, where a mix and match solution is possible to achieve the best of all worlds, but naturally this has cost implications. Only the smallest organisations are likely to be able to build an effective metadata platform solely in spreadsheets.

So,

- if you are serious about managing your data then you need to be serious about metadata.
- You are likely to need to go to market for a vendor product to help in this space.
- Your choice of vendor platform will depend on the specific business drivers and objectives of your data management effort.



Authored by:



Tim is the Practice Lead for Data & Analytics, focusing on setting and executing strategic business and technology change for Financial Services firms. He has 20+ years' experience in Financial Services, with practical experience of business transformation, developing strategic IT and data architecture, and leading adoption for Front Office, Operations and Control functions.

Tim Jennings
Technical Director
London



To learn more about Tim
see his profile here

Reach out to:
tim.jennings@synechron.com

Synechron

www.synechron.com
