

white paper

# What role does CluedIn play in metadata management?

Pages: 4 | Reading time: 10min

www.cluedin.com



## What role does CluedIn play in metadata management?

#### What is Metadata?

Metadata is data that provides information about other data. The date on which a file or table was last modified, or the location of that file, are both examples of metadata. The rows within that file or table are not metadata, that is data. This distinction is important, as it helps us to define the different responsibilities of a Data Catalog and a Master Data Management platform.



#### Metadata, Data Catalogs and Data Governance

A Data Catalog serves as an inventory of available data and helps analysts and other data users to find the data that they need. It also provides information to evaluate fitness data for intended uses. It does this by storing "pointers" and a collection of metadata that relates to the data itself. It does not host a copy of this data - at most, it will allow you to sample the data. This allows a company to take a low-cost peek at what data is available, the structure of the data and to have surface-level insights without examining the data or records themselves.

The concept of Metadata Management has undergone an evolution, mainly driven by the onset and importance of the practice of Data Governance. This means that Metadata often now includes the lineage of the data, statistics on the profiling of the data, and more. In essence, the metadata around a file should give you a pretty clear idea of how ready that data is to deliver insights, without the need to actually see the data itself. Given the span of Metadata Management capabilities, we can now start to define policies around the structure of the data, so that other systems in the data landscape can enforce those policies. This is where Master Data Management systems like CluedIn come into play.



#### Metadata Management and Master Data Management

There are certain aspects of Metadata Management that can only be achieved by actually looking at and analyzing the data itself - in its entirety. This is necessary for systems like CluedIn to apply and enforce the policies that have been set at the higher Metadata level. For example, "every customer must have an ID". CluedIn is one of the many tools in the data supply chain that are required to enforce policies defined by Data Governance and Metadata Management tools. Aspects like data quality metrics, anomalies and bad data detection, duplicate record detection, data enrichment, and more can only be uncovered by analyzing the data behind the Metadata. These all end up being Metadata in the end, which can easily be written back to Metadata Management systems.

With the definitions above, we could even go so far as to say that Metadata that tells you how different assets could link and merge together could also be considered Metadata. These elements of Metadata cannot be built properly without analyzing the data at a record level.

#### Do you need Metadata Management before you embark on Master Data Management?

Yes and no. A Master Data Management system needs data to provide any value, and good Data Catalogs will scan source systems in order to report on what files, tables, and assets are available. This is a great starting point for your Master Data Management platform as it will already know what data is available for "mastering".

Having said that, it is not a requirement to have this established before you start a Master Data Management project. Augmented systems like CluedIn are designed to allow a Data Catalog to be introduced either early or late in a project, and it does not require a large amount of rework to plumb a Data Catalog into a data stack.

Every asset you have in your Data Catalog should also be in your Master Data Management platform as well. Hence, a Data Catalog provides a registry of all company data assets, and a Master Data Management platform will host and operate data that needs attention - i.e. Deduplication, Enrichment, Matching, Merging, Linking, etc.



### What do metadata management and master data management have in common?

One of the most obvious overlaps between Metadata Management and Master Data Management is that both are systems that are intended to be implemented, operated, and valued by the business. Although technology teams will benefit from using both systems, these are practices that should not be implemented or operated by IT. Why? Because both practices offer a fantastic way to assign responsibility for the ownership of data as it moves throughout the business. But they can only do so when operated by business users.

For example, it is IT's responsibility to make sure that systems can be catalogued, but it is the business's responsibility to ensure that the assets have ownership, control over how they are shared, and policies in place to address what this data can be used for and by which systems.

Similar to Master Data Management, it is IT's responsibility to make sure that data in operational systems is accessible, but the day-to-day operations and responsibility for the health of the data itself lies with the business. IT will still play a role in this, but they should be focused on activities that can be achieved automatically, such as standardizing Identifiers, whereas the business needs to have tools that help them quickly act on other data challenges without the need to involve IT.

Metadata and Master Data are closely aligned and connected. They both complement and complete each other - but they don't necessarily rely on each other. You can have Master Data without Metadata Management. You can have Metadata Management without Master Data management. However, having both will yield a valuable synergy whereby both provide cyclical value to each of the respective pillars.

An example of this synergy can be found in a simple but common example. Imagine there's a specific Power BI report that that shows the breakdown of job titles within your organization. With over 10,000 employees this report will provide a neat bird's eye view of the distribution of roles. However, on closer inspection, you see that while some people have a job title of "Software Developer", others have "Software Dev". Suddenly the accuracy and value of the report are called into question as now you have to examine the data itself to and harmonize these types of data quality issues.



Questions you are likely to ask include:

- 1. Where did this data come from? (Metadata)
- 2. What happened to these records along the way? (Metadata / Master Data / Data Engineering)
- 3. Who let this through, and who owns this data? (Metadata / Master Data / Data Governance)
- 4. How do I get this fixed? (Data Governance)
- 5. When will someone be able to address these issues? (Data Governance)
- 6. Who else is using this report? (Metadata)

This is just a glimpse of the type of questions that you might ask, with the players responsible for producing that answer in brackets. Note that three different disciplines are cited, and at least two of these questions cannot be answered by one system or set of practices alone.

#### Conclusion

Holistic management of your end-to-end data estate will always require multiple tools, processes and teams. What is important is to bring together the most powerful and automated set of technologies to help you achieve your data-driven goals in the shortest possible timeframe and with minimal disruption. In the past, integrating and managing company-wide data has been a time-consuming, costly and often manual process. It doesn't have to be that way anymore, and Metadata Management and Master Data Management are just two of the disciplines that can bring you closer to supporting greater process efficiencies and growth opportunities.

