



The first data lake management platform built for the enterprise.

Why does metadata in the data marketplace matter?

11 SPECIFIC EXAMPLES OF WHY METADATA HELPS MAKE DATA IN A DATA MARKETPLACE USEFUL AND HOW

While there is a lot of talk about the importance of having good metadata in the data marketplace, it isn't always clear exactly what that really means. What specific impact, for example, does the presence of particular types of metadata in the data marketplace have? What are the particular types of metadata and associated metadata processes that make that impact possible? And, finally, who are the people who benefit most from the presence of those specific types of metadata in the lake?

Moreover, while good metadata has certain characteristics which drives value to the business, it is equally important how that metadata is created and maintained. To be really useful metadata needs to be accurate, comprehensive, up to data, understandable, accessible, shareable and more. The extent to which metadata in a data marketplace meets these requirements has everything to do with how that metadata is created and managed – within the marketplace and in connection to other systems in the wider data management and IT landscape.

We are going to look at both of these topics – why metadata matters and what is the right way to create and manage metadata in a data marketplace – in two articles. In this article, we'll focus on why metadata matters in a data marketplace and provide 11 specific real world examples of how having metadata in the lake helps. In a separate article, we'll focus on the ways to ensure that your metadata is created and maintained properly.

- 1. Metadata allows self-service on-demand access to data in the lake by non-technical users, like data analysts or data-savvy business people.** A key benefit of data marketplaces is the ability for business users to get the data they need on a self-service, on-demand basis through a robust search functionality. In the same way that a shopper on Amazon.com considers product descriptions, reviews, and other data points as relevant factors when making a purchasing decision, metadata gives users insight into structure, content, quality and nature of each specific data set.
- 2. Metadata helps you understand and track the quality of data in the lake.** It is a widely recognized that much of the data in enterprise systems is dirty. This is especially true with legacy and mainframe systems. With that in mind, it's important that the people doing that loading have the ability to examine the quality of that data – a task which is done through metadata. To deploy a data marketplace at scale in a large company in a reasonable timeframe, it cannot take weeks or months to assess the quality of each new data source. And once that assessment is done, the findings need to be captured and exposed to users so they can use that insight easily. Metadata enables both of these things to happen – during ingest and while looking at subsequent load logs.
- 3. Metadata can provide a complete profile and detailed insight into each data set.** There's little use in having data that users are unable to work with. So when new data is loaded into the lake, it's imperative that a statistical profile is generated. With this set of metadata, users can see detailed insights into the content, organization, and nature of each set of data that comes into the lake – as well as if it's worth keeping in the lake. Ultimately, this enables users to spend less time sorting through and judging the value of data, and more time utilizing it.

4. **Metadata allows for easy and consistent protection of sensitive data.** Protecting sensitive data from unintended exposure is essential – both to ensure regulatory compliance and the integrity and privacy of customers and others. Metadata lets you tag PII or other types of sensitive data as such. That makes it easier to keep track of that data in the lake and protect it from inappropriate use. You still might need to do other things to protect the data – like encrypt it or restrict access to it – but all that is much easier when you can find the sensitive data which has been identified as such with metadata.
5. **Metadata helps companies document how they create data which they provide to regulatory bodies.** In highly regulated industries, like financial services or healthcare, compliance with regulations often requires that companies document and maintain an audit path relative to reported data. By documenting and exposing every step taken to create a reported figure – starting with the original source of the data and continuing through all access to the data by any users and all data preparation and transformation steps along the way – metadata provides an audit path and improves regulatory compliance.
6. **Metadata reduces duplicative ETL efforts.** Metadata documents the process of how people have built specific data sets and preserves data sets as each point along that process. The presence of this metadata makes it easy for other people to review that work, reuse it or create new data preparation process that build on another person's existing efforts. By documenting data preparation flows and making that information searchable, metadata can also help organizations reduce redundant ETL efforts. Once a particular data source has been prepared and delivered out to business users, later requests for the same data can be met by tapping into the existing data preparation, rather than duplicating the effort.
7. **Metadata allows people to collaborate more effectively around shared data.** By providing a common language and view into a set of data, metadata enables better collaboration between individuals and teams. Data governance and stewardship teams for example can use the same platform of metadata to examine the data, add comments, make enhancements, pass data from one team to the next and finally deliver data to users. Using a shared platform of data makes this process much more efficient so more data can be delivered out to users faster.
8. **Metadata makes it easier to control and document who has access to what data.** As marketplaces allow more data analysts and other data savvy business people to surf through vast collections of enterprise data and build their own data sets on a self-service on-demand basis, it's important for IT, security, and data governance leaders to be able to maintain order and keep track of who sees and utilizes the data. Metadata makes it easier to administer and control data access, building on open-source tools like Sentry and Ranger, and enables administrators to set, maintain, and document access.
9. **Metadata makes data transparent so users understand it better and trust it.** By giving users visibility into everything that's happened to a particular set of data – from the moment it was on-boarded into the lake through every step of data preparation- metadata enables people to understand what the data means, where it came from, who has modified it, and how it has been enhanced. As a result, it ensures that the data remains transparent, trustworthy, and – ultimately – useful.
10. **Metadata makes data that was previously “dark” to be accessible to users.** Many organizations have large reservoirs of legacy and mainframe data systems, which are not accessible by most business users for use in reporting or analytics. Worse yet, much of this data contains little to no insight into the intended structure and content. Yet despite this plight, metadata provides a solution. By documenting its structure and contents from the moment it enters the marketplace, as well as

providing access to the data to users that didn't have any way of utilizing it before, metadata enables this "dark" data to come into the light and become useful again.

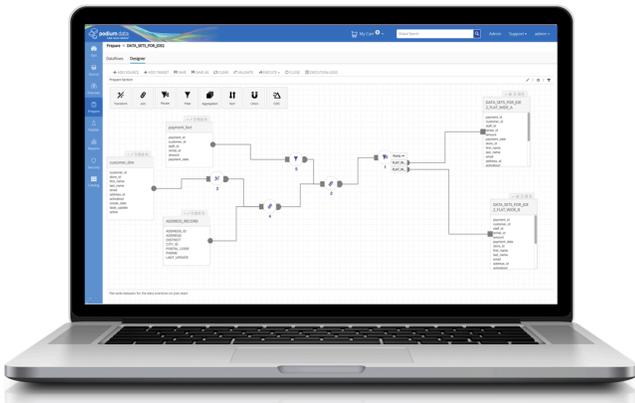
11. **Metadata makes data that was previously "dark" to be accessible to users.**

Many organizations have large reservoirs of potentially high value enterprise data in legacy and mainframe systems, which is very difficult to expose for use in analytics or reporting because the data is very complex, dirty, and poorly documented. By moving this data into a data marketplace and documenting its structure and contents, and by making that information available to data analysts and data savvy business users, metadata can move this "dark data" into the light. Users can begin tapping into this data and incorporating it into new analytic and reporting projects, making previously inaccessible high value data sets available to drive the business.

Ultimately, metadata isn't what most business users who seek value-driving insights set out looking for. What it does do, though, is make the process of finding those insights much easier for the entire business. From documenting data type and quality to providing tagging capabilities and ensuring regulatory compliance, metadata provides significant benefits to business users and IT leaders alike, which ultimately benefits the entire organization.

Learn More

Learn more about Podium's metadata management capabilities within the data lake at www.podiumdata.com



“We are using Podium to reinvent IT and our business.”

– CHIEF ARCHITECT, TOP 10 BANK

Podium Data is redefining enterprise data management with its practical, high-performance data lake management platform. Headquartered in Lowell, Massachusetts, Podium's seasoned team of entrepreneurs has deep expertise in analytics, big data, and enterprise data management. A growing number of Fortune 500 companies use Podium to provide secure self-service access to enterprise data.

For more information, go to www.podiumdata.com and follow Podium on Twitter at @PodiumData.

