• *ARTICLE*

The Importance of Metadata - Database Trends and Applications (dbta.com)

• *SUMMARY*

The author, in his article, highlights an important aspect on conducting business successfully in the modern era, the use of metadata management. While it is notorious throughout industries that data must be gathered, stored, transformed, and analyzed, often its meaning is left behind causing excessive backlogs for companies that try to extract meaningfulness out of it. It seems that using metadata, a company can describe and define data answering basic questions such as who, what, where, or how. Also, metadata makes data useful by embellishing it with details such as data type, length, or text descriptions.

To support his thesis, the author draws an example from a tv show called *Antiques Roadshow*. "In this show people bring items to professional antique dealers to have them examined and evaluated. The participants hope to learn that their items are treasures of immense value. The antique dealers spend a lot of time talking to the owner about their items. They ask questions, where did you get this item? or what can you tell me about its history?" It is clear, the answer to those questions provides details about the authenticity and nature of the item creating metadata.

It is assured that poor data quality is an issue for companies, especially when there's a regulatory compliance in place. Incorrect, fraudulent, redundant data may shrink their bottom lines at the end of the year, if not properly managed.

• *ANALYSIS*

Metadata makes it possible for employees within the organization to find the data. I think it can be safely said that most organizations have only a bare inkling about what data they have collected, let alone where it is located and who manages it. It provides value because it reduces duplication of effort and resources throughout the company, it makes easier to track who is responsible for the quality and timeliness of that data, it identifies potentially marketable content, and it identifies holes in the data strategy that can be filled with investment in quality data sources.

It seems that getting a handle on the metadata, significantly reduces the overall software bill. Of course, we live in an era where most data is reduced to four primary categories: HTML, JSON, CSV, and RDF, with the last one being an abstraction layer that can manage the encoding of information to the other formats.

Also, I think the end-goal of any data transformation initiative is to make the organization more internally transparent by making it easier for different automated systems to speak a common language that can in turn be reflected back to the humans who manage the company. Most managers would love the capability of asking questions of their data systems such as how much money was made last quarters without being reliant upon a team of programmer to ask that.

However, the article left me with a few question that it might require more research to find an answer. In essence, while I completely agree that data must be meaningful but how to you quantify meaningfulness? How do you measure it? With what criteria data is clustered in relevant or irrelevant?