

ARE YOU GIVING METADATA MANAGEMENT THE RIGHT ATTENTION?

A METADATA MANAGEMENT ILLUSTRATIVE STUDY

WRITTEN BY:

Brian Gager, Manager, Software Development



**WE BUILD SOFTWARE
THAT HELPS OUR CLIENTS GROW**

Tallan is a nationwide professional services firm that provides business-critical customized technology solutions for clients in a variety of industries. Through the combined expertise of our business technology and creative practices, Tallan provides full life cycle software development services for building web and mobile applications, solving integration challenges and leveraging data for business intelligence solutions. Since 1985, we've combined the best talent with the best technologies to help our clients not only achieve, but exceed, their business goals.

We welcome your comments or questions. Email us at bi@tallan.com or call 800-677-3693.



TABLE OF CONTENTS

ARE YOU GIVING METADATA MANAGEMENT THE RIGHT ATTENTION?	1
AUDIENCE	2
INTRODUCTION	2
THE PROJECT	2
BUSINESS REQUIREMENTS	3
PROBLEMS DEVELOP	3
FINALLY IN PRODUCTION!	4
WHAT WENT WRONG?	5
CONFLICTING/UNCLEAR DEFINITIONS OF BUSINESS TERMINOLOGY	5
POORLY DEFINED DATA GOVERNANCE STRATEGY	6
LACK OF ADEQUATE BUSINESS METADATA DOCUMENTATION	6
CONCLUSION	7

AUDIENCE

There is no shortage of definitions available on the internet for Metadata and Metadata Management. However, it is often difficult in the context of day to day operations to see how ignoring this important subject affects your business. Few subjects straddle the Business and IT more than metadata, which makes it that much more difficult for all concerned to get a clear understanding of the topic and how it impacts them. This paper is intended to provide a different perspective to both business stakeholders and information technology workers and provide some insight beyond the usual definitions of metadata. Whether you are a CIO, business unit manager, business analyst, data warehouse manager, or a data consumer, metadata matters to you.

INTRODUCTION

Metadata management is a critical but often ignored component in a successful data warehouse implementation. It is a path to establishing common understanding between the business and IT, but not often well understood by either side of the organization. Often, the problems that result from a lack of metadata strategy are not obvious, but their results are no less costly. Recognizing that common problems related to your warehouse can be traced to lack of metadata management is a step towards a solution.

The following illustrative study recounts how a lack of foundational knowledge of Metadata Management concepts cost a company well over \$150,000 in unnecessary effort spent and doubled the projected length of a reporting project. It will demonstrate how the lack of a corporate glossary, a data warehouse dictionary, and poor data governance caused numerous roadblocks and project overruns. It will then explain how these pitfalls could have been avoided had these tools and processes existed. The savings that can be made in effectively managing your metadata are quantifiable and significant. You just have to know where to look!

THE PROJECT

GNR Industries is a large retailer struggling to reduce costs and improve productivity in a very challenging market. After a thorough internal review of the company's operations, GNR embarked on a series of sweeping initiatives to improve profitability. A critical component would be an effort to create and publish several new enterprise reports to the company portal. The goal was to create the most comprehensive corporate dashboard to date by combining data that already existed in various applications and reports throughout the company. A number of other initiatives were scheduled to coincide with the release of the new dashboards, including incentive programs built around the report metrics.

BUSINESS REQUIREMENTS

Jeff Brown, the Sales Group's business analyst was assigned to gather requirements for the project. Jeff began preparing the report mockups and documentation. His analysis coincided with the views of the project sponsors that the information on these reports was really just a repackaging of data points that were already available in other reports. Data definitions were understood by everyone in the group. He expected this to be a quick and easy project, and had even mentioned to his supervisor that it looked like a "slam dunk" to deliver on time. Because of these assumptions, definitions and business rules were not provided in the requirements to save some time. Besides, since the company did not have a **corporate glossary**, there was not a quick and easy way to verify that all of the terms in the proposed reports existed or were correctly defined. A corporate glossary would have provided detailed business definitions and rules for each term, but given the aggressive timeline, Jeff went along with the assumptions that all of this data was "common knowledge". After all, everyone knew what terms like "Gross Sales" meant.

Now that requirements were approved, Tami Allen, the project manager who would oversee development was assigned to the project and Bob Jones, the data warehouse manager was also brought in to review the requirements. Bob had joined the team fairly recently and was still becoming familiar with the enterprise data warehouse (EDW).

One of the things that concerned Bob was the lack of a **data dictionary** for the EDW. Traditionally at GNR, development of the EDW was tied very closely to reporting projects. These projects always went through the company's software development process and normally had very detailed requirements. At the end of every project, however, the documentation was archived to various locations, depending on the sponsoring manager. Bob was surprised that there had never been an effort to capture these project artifacts into a data warehouse dictionary. Consequently, whenever questions would arise regarding what was in the EDW, developers or analysts would be dispatched to mine the archived documentation for answers.

PROBLEMS DEVELOP

Bob tasked Dirk, one of his more seasoned developers, to review the report requirements and verify that all of the data was available in the EDW. After several days of working to match requirements with existing data warehouse elements, Dirk came back to Bob with a number of questions.

First, there were several terms in the report requirements that had similar names to data warehouse elements. Without having detailed business rules, Dirk wasn't sure if the elements were an exact match. Since there was no data warehouse dictionary, Dirk wasn't able to tell exactly what the rules were that made up EDW data without a lot more time reviewing code. Second, he found quite a few elements that didn't appear to be in the data warehouse at all.

Bob and Tami presented these issues to Jeff, who went back to his group to clarify the questions that were raised. Getting this detail proved to be very challenging and time consuming. As Jeff began to have conversations within his group, it became clear that many of the data elements considered to be

common knowledge were sparking a great deal of debate within his own team. Apparently there were many interpretations for a number of these terms, depending on the application or team that used them. No thought had ever really been given to ensuring that data element definitions were uniquely named and defined. However, these differences really came to light for the whole group as Jeff attempted to reconcile these terms into an enterprise view of the data. This debate continued for several weeks as operational and strategic implications to creating the correct business rules were discussed and resolved.

Finally, after several weeks of meetings and countless developer and analyst resources were spent researching old documentation and applications, Jeff was able to provide very detailed answers to Bob and Tami. Bob reviewed the revised requirements and realized there were a significant number of data elements that were either not in the warehouse or had different meaning than the new requirements. Bob estimated there would be another five weeks of ETL work (Extract, Transfer, and Load) and validation before all of the data would be in the warehouse. Furthermore, resources would have to be pulled from other data warehouse projects to complete this unscheduled work.

Naturally, this caused problems for Tami. Developers were lined up and ready to code weeks ago but had been idled or re-assigned to other tasks. Now that two months had passed working through the requirements and getting the new data elements into the warehouse, Tami was finally able to get developer resources assigned to complete the reports. When report development finally began, it went smoothly and the reports were released to production without much incident but now over three months late and very much over budget.

FINALLY IN PRODUCTION!

More problems surfaced after the reports were moved to production and exposed to several thousand employees from various business units on the company portal. While training documents were created for a segment of these new users, no thought was given as to how to make this information available to *all* business users. Predictably, a great deal of confusion was unleashed to the business, as users began to digest the new reports and compare them to other data available to the business.

This was particularly acute when teams from other functional areas began looking at the data. Since the business rules were different for similarly named elements between different groups, people naturally began to question the data. And since documentation was not publicly available for both the new and old reports, it was difficult to explain these variances. The team never considered that, although the reports were developed for a specific group of people, they would generate great interest beyond that group. This is not uncommon in a successful data warehouse environment, where the user base often grows beyond the original sponsors.

The difficulty and confusion of this production release prompted a complete review of existing reports on the portal. GNR decided that it was time to take a better look at how they were managing their metadata, though a lot of time, money, and confidence had already been wasted.

WHAT WENT WRONG?

The project was eventually delivered over three months late and at least \$150,000 over budget. The late delivery of the reports caused the company to roll out scheduled productivity incentives after a critical selling season. It was difficult to quantify the dollars in increased productivity that might have been recouped had the project rolled out on time. The number of meetings caused by confusion about data terminology, resources spent digging through old documents and code, delays that resulted in lost productivity all added up to the budget overrun.

Without a careful analysis of how these dollars were spent, it might appear on the surface that the project was just slow and not managed well. However, the obstacles imposed by poor metadata management insidiously added time and money to the project at every junction. So what were the key metadata-related issues that caused this project to go astray?

CONFLICTING/UNCLEAR DEFINITIONS OF BUSINESS TERMINOLOGY

While the terms **corporate glossary** and **data dictionary** are often used interchangeably, there are subtle differences between the two. Consider the corporate glossary to be the enterprise view of all business terms and definitions. Each term should be uniquely identifiable. For instance, the Finance team and the Sales team should not have two terms called “Net Profit” that have different business definitions. They should be resolved to either have the same definition or be distinguished in a way that makes them unique, such as “Sales Net Profit and Financial Net Profit.” A **data dictionary** is a catalog of data available in the data warehouse or other database, with sufficient descriptions of each item to remove any ambiguities. The corporate Glossary and the data dictionary could be linked, where a term in the data dictionary was described by a term in the corporate glossary. Generally, the data dictionary can be a subset of the items available in a corporate glossary.

At GNR, the lack of both the corporate glossary and the data dictionary caused significant problems to the business and the IT side of the project. The business did not have clearly defined, documented, and unambiguous terminology. The business analyst and project stakeholders believed that the terminology required for the reports was “common knowledge”. Unfortunately, as the project progressed and questions began arising about the data definitions, it became clear that there was *significant* ambiguity in terminology. This difference was difficult to resolve because there was little or no documentation available. When there was documentation, it was seldom easily accessible or well known. Had GNR used a corporate glossary, at least the business users would have been able to clearly articulate requirements.

Nearly as troubling was the fact that there was no data dictionary that described the data warehouse. Several years before this project, the data warehouse was created to support an initial group of corporate reports. Requirements were developed for the reports and subsequently for the data warehouse. While at the time the requirements were fairly thorough, the documentation that was produced during this project was archived to a folder on a network file share. Many of the principals involved in that project had left the company or moved on to other projects. Consequently, the

artifacts produced from that project were unknown to the principals in this current effort. Knowledge was lost!

Many organizations fail to take advantage of work that is already done. This is especially true with data dictionaries and corporate glossaries. While it initially seems daunting to document an entire business or even a database, over time the work that is already done throughout the normal software development lifecycle can be leveraged to populate a living repository that is available to all users who need the information. Over time, this saves countless hours of research and correcting wrong decisions that are made when assumptions are drawn about the meaning of corporate data.

POORLY DEFINED DATA GOVERNANCE STRATEGY

A lack of understanding existed about the level of detail needed to effectively communicate business requirements to the data warehouse team. No formal process existed for introducing new business terms to the organization. A key component of data governance includes having a defined process for introducing new terms into the corporate vocabulary. New terms are submitted for inclusion into the corporate dictionary and reviewed by the data governance team for clarity. Once the term is deemed necessary and unambiguous, it becomes part of the corporate glossary.

A solution at GNR would have been to create a Data Governance team that meet bi-weekly to review new terminology submissions as well as identify and resolve ambiguities with current terminology. The makeup of the team would include Jeff, Tami, one representative from each corporate business unit, and a VP from IT and the business side to facilitate the team. Over time, this process would lead to common understanding and documentation of terminology between business and IT and within individual business groups. The work to define a term is done once, correctly.

LACK OF ADEQUATE BUSINESS METADATA DOCUMENTATION

A great deal of effort was expended to accurately define the business rules for all report elements once the knowledge gaps were realized. However, little thought was given to how these report definitions and business rules should be communicated to users, given the fact that the reports would be widely available throughout the organization. After the project was completed, all documentation regarding requirements, data element definitions, and business rules was archived into the project folders. Report definitions were not published in a manner that was accessible to users who were actually viewing the reports. This was consistent for all reports that were already available on the portal. There was no online documentation to explain how these reports were used, the meaning of report terms, or any history of terminology changes. Naturally, confusion resulted when users would compare reports that appeared to have very similar terms and find discrepancies, not knowing what the actual business rules were. Not only did this cause confusion, it instilled a lack of trust in the data, which took time and resources to overcome.

To extract the most value from business metadata, it should be accessible to all who need it. While having access to a corporate glossary and/or data dictionary is useful, metadata should be integrated into end user applications, such as reports, documentation on the corporate portal, spreadsheets, and business applications. Some of the ways GNR could have made metadata accessible:

- Enterprise Reports: Create links to report glossaries that tie into the data dictionary, ensuring that up to date definitions are readily accessible for all reports and all report elements.
- Excel: For the business analysts who browse cubes through Microsoft Excel, creating a link in the context menu to view definitions of metrics and other metadata would be very beneficial. Analysts are often forced browse cube data and attempt to infer definitions simply by the term names.
- Company Portal: Documents on the portal that use business terms could have tool tips or links that tie directly into the metadata repository
- Corporate Glossary/ Data Dictionary- Publish on the corporate intranet. Make the dictionary easily accessible for all users.

CONCLUSION

The company did ultimately learn from this painful experience and chose to aggressively start managing their metadata. The first step was to ensure that all the work already performed on this project was not wasted. GNR created a metadata repository to house the business and data definitions that were verified during the project. This allowed subsequent projects to build on this effort, and ensure that future users asking the same question would not have to repeat the same research again and again. The company made a serious effort to expose the business metadata to end users, and eventually improved communication and understanding of corporate information.

It was not long after the repository was initiated that the benefits became clear. GNR had initially budgeted an expense increase of 10% to every data-related project estimate to cover the cost of maintaining the metadata. However, within months they quickly realized that database related projects began coming in 10-30% under forecast as the analysis time spent on an average project dropped dramatically. Furthermore, GNR discovered ways to automate the synchronization of project documentation with the metadata repository, so the initial 10% of project expense to maintain metadata was virtually eliminated.

Organizations are struggling to understand what Metadata is, how it affects the organization and how to effectively manage it. The true costs of this failure to grasp metadata management concepts are often hidden in the inefficiencies and project overruns that were described in this paper. Once you realize that you are actually spending money by not pursuing a metadata management strategy, it will be easier to justify investing resources in maturing your understanding and utilization of metadata. It's never too late!