Many data warehouses are currently underutilized by managers and knowledge workers. Can high-quality end-user metadata help to increase levels of adoption and use?

**DOES DATA WAREHOUSE END-USER METADATA ADD VALUE?**

*By Neil Foshay, Avinandan Mukherjee, and Andrew Taylor*

While organizations routinely realize data is a key asset that must be exploited in order to achieve success in today’s competitive business environment, unlocking the potential business benefits of this data remains an elusive goal for many firms. In order to exploit the value of their data, many organizations have implemented data warehouses and business intelligence applications over the last decade, often at significant cost expenditure. These data warehouse environments serve to collect and integrate data, and to turn it into information that is accessible for query and analysis to produce insights that can inform and influence business decisions.
he primary consumers of the contents of data warehouse and business intelligence applications are decision makers and knowledge workers within the organization. Often, the performance of these end users is measured by how effectively they are able to use available data to make good decisions. Many of the end users are not technically oriented and need a significant amount of support to use a data warehouse effectively.

What would it mean if many of these managers and knowledge workers don’t fully understand and trust the data they are being provided? What if this lack of understanding and trust of data means end users don’t perceive their data warehouse to be easy to use or useful, and consequently don’t use it?

There is a good deal of evidence, based on the experience of data warehouse practitioners, that the situation described here is the current reality in many organizations. The primary evidence: the perception that many data warehouses are underutilized. For example, a study by Raden [9] found data warehouses and business intelligence applications have low adoption rates within organizations (compared to spreadsheets and standalone databases). While there are likely many factors involved causing this situation (poor data quality, dissatisfaction with the business intelligence tool used, and lack of data training, to name a few), it may be the case that data warehouse end users often don’t fully understand or trust their data. Consequently, these end users are not willing to risk making key decisions with data they have little insight into.

Data warehouse practitioners seem to understand this situation is common in organizations. However, there may be a perceptual gap between what end users want from the data warehouse and what data warehouse practitioners think end users want. A good deal has been written regarding one method for supporting end users: metadata.

Metadata has been described, generically, as “data about data.” This definition is not particularly helpful in understanding the value of metadata in an information systems context. Dempsey and Heery offer this more extensive definition: “Metadata is data associated with objects which relieves their potential users of having full advance knowledge of their existence or characteristics. It supports a variety of operations” [3].

Additional understanding of metadata is offered by Tannenbaum [10] who states that metadata serves to answer five important questions regarding information in an organization:

- What do I have?
- What does it mean?
- Where is it?
- How did it get there?
- How do I get it?

Metadata helps data warehouse end users to understand the various types of information resources avail-
able from a data warehouse/business intelligence environment. These resources can take many forms including data elements, queries, reports, and published documents. Gardner [6] states that metadata is critical for all aspects of a data warehouse and provides a roadmap or blueprint for many warehouse functions. Haley and Watson describe the value of end-user metadata as follows: “We have found that users without this [metadata] refrain from using the data warehouse, spend inordinate amounts of time developing and testing queries, or ask someone more skilled to write their queries” [7].

That metadata provides benefits to end users seems to be well understood. However, there is a perception in the data warehouse industry that many data warehouses today do not provide “good” metadata to end users. Further, there are many types of metadata and research into perceptions of end-user metadata is lacking. In light of the apparent paucity of research on use of metadata, the study that is the basis for this article was conceived with the following objectives:

- To propose a taxonomy for classifying end-user metadata;
- To gain an understanding of what types of metadata are provided today to data warehouse end users;
- To understand how data warehouse practitioners and the end users they support perceive different types of metadata; and
- To develop and test an empirical model, for end users, that links perceptions of metadata quality and use with user attitudes toward data, perceived usefulness and ease of use of the data warehouse and, ultimately, data warehouse use.

**End-User Metadata Taxonomy**

To address the first three objectives of the study, we propose a standard taxonomy for classifying end-user metadata. Data warehouse practitioners recognize there are numerous types of metadata that serve the needs of different types of users. Metadata serves the needs of technical stakeholders (data warehouse practitioners responsible for developing and maintaining the data warehouse) as well as business stakeholders (decision makers and knowledge workers who consume the information generated by a data warehouse). Thus there is metadata for IT users and metadata for business users. This research focuses exclusively on the latter. Laney [8] suggests there are three types of metadata: definitional, navigational, and administrative. Laney’s framework was intended to be a generalized categorization scheme for metadata. Based on Laney’s definition, and refinements made through interviews with data warehousing experts, the authors developed the end-user metadata taxonomy presented in Table 1. The proposed taxonomy expands on Laney’s definition and is specific to the metadata requirements of data warehouse end users.

**Research Model**

To operationalize the final objective of the study, we developed and tested a composite research model, presented in Figure 1, which explores the relationship between end-user perceptions of metadata quality, their attitudes toward the data available from the data warehouse, their overall satisfaction with the data warehouse, and their level of use of the warehouse. The model was developed through consultation with a panel of data warehouse and metadata experts and by leveraging the Technology Acceptance Model (TAM) developed by Davis et al. [2] as well as the work on individual attitudes and beliefs pioneered by Fishbein and Ajzen [4].

![Figure 1. End-user metadata conceptual model.](image-url)
The final level in the model is “data warehouse use.” This variable seeks to measure the extent to which users currently access the data warehouse and their plans for future use.

The study also evaluated the influence of other factors on user attitudes and data warehouse use. These include end-user perceptions of:

- The level of quality of the data in the warehouse;
- The usefulness of the business intelligence tool used to access the data warehouse; and
- The quality of training received regarding the content of the data warehouse (data training).

These factors have been recognized within the data warehouse industry as being important to data warehouse success. As such, they have been included in our research model. The core hypotheses for the study are:

H1: End-user metadata quality and use influence end-user attitudes toward the data in their data warehouse.

H2: User attitudes toward data influence user perceptions of both the usefulness and ease of use of the data warehouse.

H3: User perceptions of ease of use and usefulness of the data warehouse influence the level of use of the warehouse.

The hypotheses were tested using multivariate statistical methods.

**FINDINGS**

**Respondent Profile.** The study, which was conducted from April 2005 to October 2005, had the primary focus on the end user. However, the study gathered valuable insights from data warehouse practitioners as well. Data was collected through two online surveys—one aimed at data warehouse practitioners and the other at data warehouse end users. Overall, responses were received from 268 data warehouse practitioners from 266 organizations and 621 end-user respondents from 104 organizations representing a subset of the 266 organizations. Thus, we had 104 organizations from which we obtained matched-pair or dyadic responses, that is, responses from a data warehouse practitioner and at least one end user working in the same organization.

Generally, technical respondents played a technomanagerial role in supporting one or more data ware-
Types of End-User Metadata Provided. The second objective of the study is to explore the types of metadata provided to end users. Approximately 88% of the data warehouses studied provide one or more types of metadata to end users—12% of data warehouses in the study provide no metadata to end users. Not surprisingly, the most commonly provided metadata is Definitional—77.7% of warehouses provide basic data definitions. Approximately half (51%) of the warehouses provide Data Quality metadata. Lineage (36.3%) and Navigational (31.9%) are the least commonly provided. The most common combination of categories provided are Definitional and Quality (42.6%); 13.9% of data warehouses in the study provide all four categories of end-user metadata.

Perceived Usefulness of/Satisfaction With End-User Metadata. To address the third objective, both technical and user respondents were asked their opinions on the level of usefulness and satisfaction with various categories of metadata to end users. End users were asked about their own perceptions regarding usefulness of and level of satisfaction with the metadata they are provided, while data warehouse practitioners were asked to speculate about the perceptions of the end users they support. Table 2 presents usefulness and satisfaction scores provided by both respondent groups.

As highlighted in Table 2, there is consensus between the technical and end-user stakeholders that end-user metadata is useful. However, the views of the two groups diverge on the relative usefulness of the types of metadata. For example, data warehouse practitioners believe Definitional metadata to be the most useful to end users. However, end users indicate Definitional metadata as one of the least useful of the four categories, and consider Data Quality metadata to be the most useful.

While metadata is deemed useful, end-user responses indicate they are only slightly satisfied with the metadata available to them (overall average satisfaction score—5.5 on a 9-point scale). Users are most satisfied with Data Quality metadata and least satisfied with Lineage metadata. Interestingly, technical respondents, who play a role in providing metadata to end users, feel that end users are most satisfied with Definitional metadata. They, however, do realize that their end users are somewhat ambivalent about the metadata they use (overall average satisfaction score—5.4). The results presented in Table 2 indicate end users are looking for ‘better’ metadata and that many of the data warehouse practitioners who responded to the survey know this.

The Influence of Metadata on User Attitudes. Here, we address the final objective of the study and provide an overview of some of the most significant findings related to the study’s core hypotheses and on the effect of moderating variables. A series of regression analyses were performed to test the study hypotheses (summarized earlier as H1–H3). Overall, the three study hypotheses were supported and were statistically significant at a p<0.05 level. Figure 3a presents a summary of the results of the analysis.

Based on the multivariate statistical analysis of the data collected during the study, all hypotheses are accepted at the desired significance level. The following observations can be made from the results of the analysis:

- End-user metadata factors have a moderately strong influence on user attitudes toward data in the warehouse. Their influence is roughly the same as that of “Other factors”: user perceptions of the quality of training provided, business intelligence tool usefulness, and data quality. Metadata factors combined with these other factors seem to have a relatively strong influence on attitudes. This finding is very significant. It suggests that metadata is
an important factor in influencing whether or not a user will develop a positive attitude toward a data warehouse.

• The direct influence of “Other factors” included in the study on data warehouse use appears to be quite weak. It seems other factors operate in a manner similar to the metadata factors— their influence on use is indirect. Of the other factors included, perceived data quality was the most influential on user attitudes toward data.

• The influence of user attitudes toward the data available from the warehouse on both perceived usefulness and perceived ease of use of the data warehouse is relatively strong.

• The influence of perceived usefulness and perceived ease of use of the data warehouse on the level of use of the warehouse is moderately strong. As such, it appears that factors other than perceived usefulness and ease of use are involved in determining the extent of data warehouse use.

• User experience with using a data warehouse has a very important influence on the degree to which end-user metadata quality and use influence user attitudes. As highlighted in Figure 3b, the strength of the relationship (as measured by R² value) between metadata factors and user attitudes toward data becomes less significant to the user over time. When users are relatively inexperienced, end-user metadata has a stronger influence on their attitudes toward data.

**Implications for Practice**

This study offers a number of useful insights to data warehouse practitioners. Here, we describe key implications from the study that may be of particular interest.

The end-user metadata categories proposed by the study are a good starting point for practitioners. Based on survey responses, the basic categories of metadata suggested in this study seem to make sense to the majority of data warehouse end users. The categories (Definitional, Data Quality, Navigational, and Lineage) offer a useful framework for data warehouse practitioners in developing and communicating their metadata strategy to various stakeholders and for prioritizing metadata development projects. More importantly, the categories provided offer a means for defining specific end-user metadata requirements.

End users and data warehouse practitioners agree that metadata is important and useful, but... The results of this study point to the fact that end users and data warehouse practitioners are in agreement regarding the usefulness of end-user metadata. They also agree that all four types of metadata are useful to end users. However, practitioners and users don’t agree on what the most useful category is.

Based on satisfaction scores computed from the survey responses, it seems that end users are not really satisfied with the metadata they are currently provided. Our findings also point to the types of metadata that end users find less useful or satisfactory. The data warehouse practitioners who participated in the study seem to be aware that the metadata they provide today is not adequate; perhaps due to the fact that it’s often too static and does not provide the informational depth end users seek. Awareness on the part of practitioners that end users are generally not satisfied with metadata is important as it leads them to understand what users need and how to provide users with what they need. For example, excessive attention of practitioners to Definitional metadata at the expense of other types can be overkill.

Metadata can influence user attitudes. The empirical evidence from the study suggests that end-user metadata quality and use positively influence user attitudes about data and (indirectly) the level of use of the data warehouse. Its influence on attitudes and use is roughly equivalent to the influence of user perceptions of data quality, the business intelligence tool in use, and data warehouse content (data) training. This does not suggest that metadata is in any way a substitute for these factors. Rather, metadata should be thought of as complementary to data quality and training programs—metadata can play a role in such programs in addition to having intrinsic value.

Quality is king. Data Quality emerged during the study as the most influential category of metadata on end-user attitudes. Again, it is important to note that Data Quality metadata is in no way a substitute for the actual quality of the data in the warehouse. Rather, it may serve to inform users regarding data quality so
that they can make appropriate use of the data that is available to them. This means users can employ data of good quality with confidence and make informed decisions regarding whether or not they should use data of questionable quality for specific purposes. Additionally, metadata can be used to facilitate communication between users and the data warehouse team regarding data quality issues. The importance of data quality and data quality metadata has been identified in other studies, including [1] and [5].

End-user experience is a key factor. The strength of the relationship between end-user metadata and user attitudes weakens as users gain experience with using a specific data warehouse. This indicates that over time, end-user reliance on metadata diminishes as they come to understand their data. When users are inexperienced, the relationship between metadata and attitudes toward data is quite strong. However, for data warehouse practitioners, this probably doesn’t diminish the need for Data Quality metadata over time. Of course, when changes occur to the data warehouse, all forms of metadata are important.

The influence of metadata factors on the attitudes of new data warehouse users presents a tremendous opportunity for practitioners to potentially improve data warehouse adoption rates. If new users of a warehouse can be provided with good Data Quality metadata in conjunction with effective training, there is a great likelihood they will develop positive attitudes toward their data (other factors, such as actual data quality, being equal). This, in turn, should encourage higher rates of adoption and use of the data warehouse—a significant goal for many IS departments.

CONCLUSION
This study reported here was aimed at developing a better understanding of the role played by end-user metadata in a data warehouse environment. The results of the study support the three main hypotheses. First, end-user metadata quality and use have a moderate influence on user attitudes toward data [H1]. The impact of these factors is approximately the same as the collective influence of user perceptions of data quality, the effectiveness of the business intelligence tool used, and the quality of training received. Collectively, metadata and these other factors have a relatively strong influence on user attitudes toward data.

The results of the study indicate there is a relatively strong relationship between user attitudes toward data and their overall satisfaction with the data warehouse, as measured by perceived usefulness and ease of use [H2]. Finally, user perceptions of ease of use and usefulness of the data warehouse influence the level of use of the warehouse [H3]. These factors have a moderate effect on use. More research is required to identify other factors that might influence the level of use of a data warehouse.

The implications of the results are significant for data warehouse practitioners. The study indicates users appreciate the value of metadata and suggests there is a tremendous opportunity to encourage increased levels of adoption and use of their data warehouses, particularly among new data warehouse users, by implementing effective metadata and integrating it into data quality and training programs. Enhancing the value derived from adequate usage of a data warehouse can provide any organization with a sustainable competitive advantage by allowing managers and knowledge workers to make better, more informed decisions. Metadata is a key factor in deriving this value.

REFERENCES

Neil Foshay (nfoshay@stfx.ca) is an assistant professor of information systems in the Gerald Schwartz School of Business and Information Systems at St. Francis Xavier University in Antigonish, Nova Scotia, Canada. Avinandan Mukherjee (mukherjeeav@mail.montclair.edu) is a professor of marketing at Montclair State University in Montclair, NJ. Andrew Taylor (w.a.taylor@bradford.ac.uk) is an associate dean of research and a professor of information systems at the Bradford School of Management in Bradford, England.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.