Developing a Metadata Model for Business Intelligence End Users

Yuriy Verbitskiy[1], William Yeoh[2], Sharman Lichtenstein[2]

[1] Planning and Institutional Performance, University of South Australia, Adelaide, South Australia, Australia
yuriy.verbitskiy@unisa.edu.au

[2] School of Information and Business Analytics, Deakin University, Burwood, Victoria, Australia
{william.yeoh; sharman.lichtenstein}@deakin.edu.au

Abstract. The effective use of metadata can offer end users an improved understanding and greater level of assurance during the Business Intelligence (BI) report analysis process. This paper reports key findings from a case study that investigates critical end-user metadata issues in a large Australian organisation. The findings led to the development of an end-user metadata model on object (report and cube) and element (term and column) levels, which can support effective BI use and potentially increase user satisfaction at the case organisation. The adoption and use of BI applications by business stakeholders may be improved by incorporating the end-user metadata model.

Keywords: Business intelligence, metadata, end user

1 Introduction

In recent years BI applications have been consistently ranked among the top five technology priorities in a global survey of Chief Information Officers [6]. Business Intelligence (BI) refers to "a broad category of technologies, applications and processes used for gathering, storing, accessing and analysing data to help its users make better decisions" [22]. Hence meaningful information can be delivered at the right time, at the right location and in the right form [14] to assist individuals, departments or even larger units to facilitate improved decision-making. Despite the importance of BI applications, in many instances they are significantly underused, partly attributable to issues of user satisfaction [3], [4], [5], [9], [12]. End-user satisfaction is essential to improving the uptake of BI [3]. A key influence on end-user satisfaction with BI is *end-user metadata* [5], [10]. End-user metadata is needed as many end-users are not technically-oriented and require substantial support to use BI applications and fully understand the BI cube or measurement and definition of a BI term, or the meaning of a column of data originating from other systems [5].

Indeed, a potentially valuable solution to the problem of poor user understanding of BI data and reports is the use of end-user metadata [5], [21]. Metadata has been

afforded many definitions over the years. It has been defined as simply 'data about the data'. Importantly, it plays a crucial role in an effective BI environment [5], [10], [16]. Metadata serves as a mechanism that provides the context about the data and information of a BI report [18]. It addresses the how, when, why and what questions in a BI environment [7]. Gartner Research contends that metadata is one of the most important functionalities that a BI environment should deliver [2]. Inmon et al [8] and Wells and Hess [21] further assert that without metadata to support BI reports, a BI application offers little value to an organisation. While technical users understand the BI environment because it is one of their primary work objectives, business users need support that will help them feel confident about using the data and BI tools in general. Making available effective end-user metadata could provide such support.

Several studies discuss different types of metadata. Notably, metadata has been classified as business metadata (which relates to the data that is meaningful to business users) or technical metadata (which is used by information technology staff responsible for developing and administering a BI system) [1], [15], [17], [19]. In the technical domain metadata is crucial for building a data warehouse as developers need to know the data structures, source-to-target mappings, and data transformation rules during the data extraction, transformation and loading (ETL) processes [15]. In the business metadata domain, Foshay et al [5] propose an end-user metadata taxonomy of four categories: definitional, data quality, lineage and navigational metadata. Despite these early studies on metadata types, to date there is little research that explores the detailed elements of metadata requirements for business users in an enterprise-scale BI environment. This paper aims to identify the key elements of business end-user metadata by conducting an exploratory interpretive case study at a large Australian organisation. By synthesising key findings from the case study, the paper develops an end-user metadata model which may help BI practitioners implement metadata for business users.

The balance of the paper is set out as follows. First, the paper discusses the research methodology. Second, it presents and discusses related findings from the case study. Next, the paper presents an end-user metadata model emerging from the case findings. Finally, the paper draws conclusions and offers suggestions for future research directions.

2 Research Methodology

This study adopted an interpretive approach as the environment for BI and metadata effectiveness involves people whose opinions help determine success [20]. The views of BI stakeholders were sought as they could yield insights that might help to develop theory. The study used the single case method to explore end-user metadata issues leading to successful BI use within the real-life context of a large Australian higher education institution. In the past decade the higher education sector has found many uses for BI [11]. Higher education is therefore a suitable choice of industry sector for this research. The case study method was selected as it enables a deeper investigation of a phenomenon by capturing detailed "reality" and a rich picture that can reveal the highly complex and unquantifiable events associated with that phenomenon [23].

According to Miles and Huberman [13], a case study provides better explanations for the examined phenomenon.

The research project involved collaboration between the researchers and the organisation's BI unit. With more than thirty thousand students, the university believed it was crucial to exploit a powerful data analysis and performance management tool. The university acknowledged the importance of BI technology and strove to deliver it to the wider organisational community. Importantly for this study, the BI unit indicated that it did not have a consistent metadata model that benefited business users of the BI environment. Thus the university represented an ideal case organisation for investigating metadata issues in a BI environment.

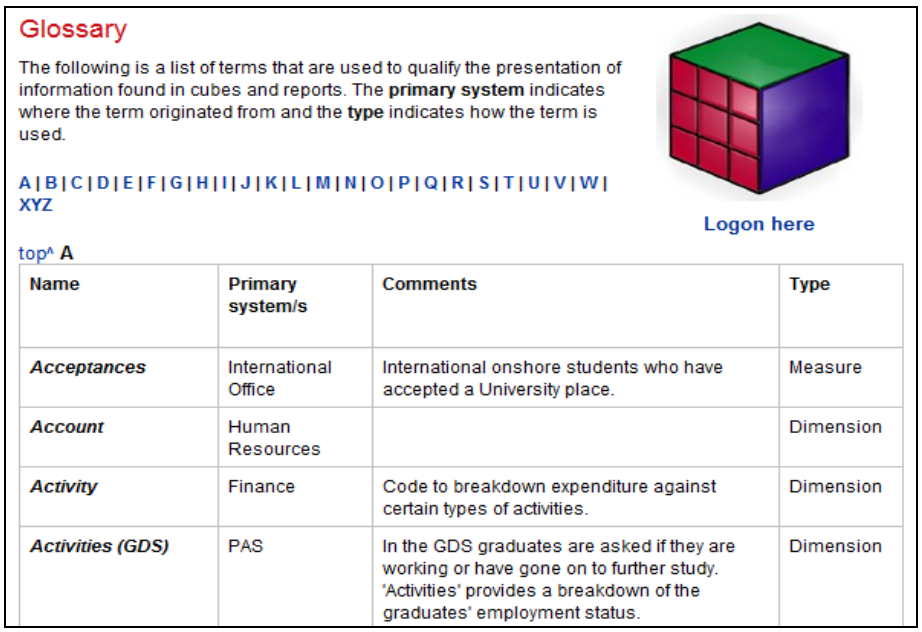
To facilitate data collection, data were gathered from multiple sources including semi-structured interviews, demonstrations of BI applications, project documentation and observations of a BI portal. A total of nine key BI stakeholders at the university were interviewed, each interview of one to one and a half hours' duration. All interviewees had been involved directly in business or IT functions of the BI initiative. The interviewees comprised a BI manager, a project officer, an enterprise data warehouse specialist, a planning officer, a BI analyst, a senior analyst, a business analyst, and two key business users. The different roles reflect the different abilities and responsibilities in the organisation's enterprise-scale BI environment. Business users can only view the reporting objects such as cubes and reports while business analysts can produce and view the reporting objects. The BI team is responsible for data modelling in BI, administration of data warehouse, data quality, implementation of ETL and business rules. In addition the BI team develops complex cubes and reports on the request of business analysts or users. Technical specialists are responsible for providing and maintaining the technical infrastructure of the BI environment, such as the data sources and data warehouse.

Importantly, the different work roles of the interviewees suggested that each user might require a particular type of metadata for work. For example, business users do not wish to view technical descriptions; rather, they seek to understand and trust the data viewed in a reporting object. At the same time, technical specialists need technical information to do their job. All such differences were probed during the interviews and identified in the later interview transcript analysis. It is important to highlight that business analysts are also business users, but with advanced skills in data manipulation. Business analysts are typically the first point of contact for business users in relation to complaints and requests for change. Therefore business analysts of the organisation could understand the complex BI problems and needs of various business users.

Each interview commenced by asking a set of predefined general questions. This procedure provided flexibility to obtain detailed information from interviewees while ensuring consistency. Selected interview questions are shown in Section 3. The interviews were transcribed and analysed by a combination of inductive and deductive content analysis with key themes thus identified [13]. The themes were synthesised to develop a metadata model for business users.

2.1 State of Metadata at the Case Organisation

At the time of study the university was migrating to a new version of the BI environment ('IBM-Cognos'). With the new version several novel solutions for metadata implementation were identified. The university was also changing its approach to using the BI environment. Earlier the front-end BI environment was encapsulated only in a certain application ('Cognos Upfront') that showed cubes, reports and related documents to users. However the new approach enabled the delivery of BI objects into an integrated staff portal so that BI would be embedded in the general user environment.



Glossary

The following is a list of terms that are used to qualify the presentation of information found in cubes and reports. The **primary system** indicates where the term originated from and the **type** indicates how the term is used.

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | XYZ

Logon here

top^ **A**

Name	Primary system/s	Comments	Type
Acceptances	International Office	International onshore students who have accepted a University place.	Measure
Account	Human Resources		Dimension
Activity	Finance	Code to breakdown expenditure against certain types of activities.	Dimension
Activities (GDS)	PAS	In the GDS graduates are asked if they are working or have gone on to further study. 'Activities' provides a breakdown of the graduates' employment status.	Dimension

Fig. 1. Metadata in Glossary application

Within the existing BI environment some metadata were already in place. However they were in an immature state. Technical specialists employed spreadsheets with links between technical and business objects in the BI environment. Business users exploited structured descriptions for cubes and reports in a 'Cognos Upfront' environment. In addition, they were able to use the 'Glossary' application which represents a simple webpage with a list of terms and descriptions as illustrated in Figure 1. While this metadata provided helpful information for the business users the BI team recognised that they required a more consistent and user-friendly metadata model. The following section presents and discusses the key research findings relating to end-user metadata needs for BI.

3 Findings

This section presents the key findings from the study organised by selected interview questions. The voices of participants are privileged by being highlighted in this presentation of findings.

Question 1: Can you tell me about the occasions when you thought that metadata or additional description would help users or yourself to understand data better?

Participant 1	I suppose the main reason we want metadata is for end users looking at reports to be able to understand where these numbers come from and what processing has done to them.
Participant 2	I just think anywhere where people use reports.
Participant 3	I think probably in all instances while I publish reports we are trying to put as much metadata around the report as we can.
Participant 4	But again it comes to interpretations of object and what does it actually represent. That is where we need some sort of clarity and consistency.
Participant 5	For developers, understanding of the data is fine. For users, it could be helpful. To clarify the rules of presentation in the business tools, the scope of the data, to resolve different definitions e.g. definition of mode for course is different from definition of mode for program.
Participant 6	So for the people who use this dimension or column, they need some description – what exactly [is] this, [and] how it was derived.
Participant 7	In general it is about data that you are looking at… I think to have some prompts around that, that says, "This is what you are looking at".
Participant 8	Knowing where the information comes from, what the information represents
Participant 9	First of all, to help people to understand what they are actually looking at

These responses highlight that staff with different roles understand the importance and role of metadata similarly. The most common view regarding metadata was that metadata should explain data on the reporting object for business users. For this to be effective, metadata should show where the data originated and information about data processing. Some participants from BI unit believed that metadata would also help with testing, impact analysis and documentation. The next question aims to identify problems with understanding data or any other elements of the BI environment.

Question 2: Are there problems with understanding data or any elements in the environment?

Participant 1	Well, yes. That example with SATAC (reports) … What we do. We manually put the metadata on the report… But often people will not see that.
Participant 2	Yes, I think so and I can give you a classic example … basically that people in higher [positions] in the uni, they don't really understand the data because there is no place where they can go to look up [a] consistent definition.
Participant 3	I think [there are] varying degrees of knowledge and understanding through all the levels.

Participant 4	Probably not. I think there is confusion about it. Most of the users understand the context around it but if you take some infrequent users [they do] not necessarily know what the difference is.
Participant 5	Yes. The user should understand the business environment that they are dealing with. Business users have different understandings of the data.
Participant 6	Yes, of course... you need a common description for each attribute or column or whatever.
Participant 7	I think there are two needs not only for business analysts who works with data but also for someone who is not a business analyst but uses the data he needs to interpret
Participant 8	That is a huge barrier in trying to identify, especially from the metadata point of view, how you define the information if it is used across the university with the same terminology but the definitions are completely different.
Participant 9	Absolutely, yes. That is one of the things that we do all the time.

Almost all participants noted this as one of their key problems. The problem arises due to technical limitations, complex business requirements in the university, the organisational structure of the university and because consistent metadata implementation had not yet occurred.

To continue the interview in the area of data understanding, the next question provided to participants an additional potential argument why data understanding could prove challenging.

Question 3: Are there problems with the interpretation of some terms which could have a different meaning for staff members with different roles?

Participant 1	Absolutely. The way we get around that is again labelling the report with metadata and a column but again it is manual metadata that ideally should be automated
Participant 2	Yes, definitely. Things like that actually can make quite a difference to the business decisions that have been made at the uni ...
Participant 3	That is another big problem and it always goes hand in hand with the question before – understanding what the data actually is and what it is used for.
Participant 4	Yes. People want to see things from a different perspective.
Participant 5	Yes, I think there are problems.
Participant 7	Someone who uses their own system in their environment still needs to understand that that terminology is for their group and that when it is a different system, it could be different.

Most participants agreed that this is a problem and that each person interprets the data from a unique perspective. The problem occurs not only between different departments within the university but also between the university and external organisations with whom they collaborate such as government agencies.

To some interviewees an explanation of the reason 'why' and 'how' terms and calculations were developed was necessary. Other participants thought such explanations were needed only in some cases but believed it more important to have a clear description of the meaning of the term or calculation. Following these questions some participants were asked to categorise all situations where they believed metadata was required. The following participant quotes illustrate the findings which follow.

"*From the start we need metadata on the report which describes what it is all about. We also need metadata on the columns again including the scope, and the business rules attached to them. I did not think about metadata on the row level, we probably do not [need that]. We also need metadata on the actual source systems and source tables. And most importantly, we need metadata for the links between them*." (Participant 1)

"*You could go into the technical side more as well, but we are mainly looking at the business side. And then complete business processes where we tend to categorise reports into a particular area like Enrolment or Student Evaluation. I will probably just start with the basic set and say, let's do that*." (Participant 2)

"*You can overwhelm people with too much information. First thing we do is we start with a small amount of information. And then you go to the people and realise that they don't use something because the description is not informative enough, which need to be improved. I do not want to go too much first. I would rather go a little and build rather than too much and people will not use it*." (Participant 8)

The above responses suggest that in the early stages of BI implementation, the BI unit was more focused on implementing metadata for business users because the satisfaction of end users was the main priority. This finding also suggests that to implement metadata it may be advisable to apply several iterations, commencing with certain business metadata then completing the entire metadata implementation process with the selected section of the metadata. The findings also reaffirm the importance of metadata for business users. Having technical metadata at the front-end interface is comparatively unimportant as such metadata are not particularly useful for end users.

4 Developing an End-User Metadata Model

The findings suggest that the university requires a single, consistent and independent metadata repository. A standardised, enterprise-level business metadata model is paramount to enhance BI use and thus increase end-user satisfaction. The findings further indicate that the current BI environment requires a comprehensive end-user metadata model in order to:

- provide consistency for descriptions and definitions of the data in a BI environment;
- provide an overall enterprise view;
- address the problem of misinterpretation of some terms which could have different meanings for staff with different roles.

Drawing on the research findings an end-user metadata model on both object and element levels was developed for the university. The metadata model consists of two main levels which relate to the end user environment – reporting objects such as cubes and reports; and elements such as terms and columns used in cubes and reports. At the university the metadata of the elements was implemented as a webpage ('Glossary' application) where the terms used in cubes and reports are provided however object metadata was not yet available for business users at the time of study.

Figure 2 depicts the object-level metadata model with critical metadata fields to enhance the understanding of a BI report/cube. 'Type' defines whether the object is a report or a cube. This metadata will allow better classification for search, sorting and presentation purposes. 'Scope' defines inclusions and exclusions from the data. For example a report includes all students or a report includes only domestic students. 'Source systems' shows the source applications or external sources for data that has been sourced from elsewhere. 'Primary audience' describes the intended key users of the reporting object. 'History' stores the previous name of the object or selected historical comments. 'Report run date' shows when the report was run and is relevant only for the report (but not relevant for a cube).

Object-level Metadata
Name of the object
Type of the object
Description
Time period
Scope
Usage
Primary audience
Critical notes
Source systems
History
Contact person
Report designer
Data Refresh date
Refresh frequency
Report run date

Fig. 2. Object-level metadata model (metadata of reports and cubes)

Other fields are briefly explained here. 'Description' provides a general overview of the object. 'Time period' explains the time period presented in the data. 'Usage' discusses the purpose of the reporting object, and how and for what purpose it should be used. 'Notes' presents critical notes regarding the object. The 'Contact/author'

field shows the main contact person and the author (or developer) of the report. For consistency this metadata field is divided into two metadata – 'Contact person' and 'Report designer'. 'Report designer' is also relevant only for the report. The 'Refresh frequency' field shows the refresh frequency of the reporting object while the 'Data Refresh date' field shows the date of the last data refresh.

Element-level Metadata
Name of the element
Business acronym
Primary system
Description
Type
Places of use
History
Owner
Refresh date

Fig. 3. Element-level metadata model (metadata of terms and columns)

As shown in Figure 3 the end-user metadata model on the element level has nine metadata fields. 'Business acronym' shows the business acronym for the element if it exists. The 'Places of use' metadata field shows where the element is commonly used across the BI environment. The 'History' metadata stores the previous name of the element or some historical comments. The 'Owner' metadata field represents the owner of that particular field. This can be the same as in 'Contact person' for the parent object however it can also be different. The 'Refresh date' shows the date when that element was last updated. The 'Primary system' metadata field indicates the origin of the term. The 'Comments' field provides comments on the meaning of the element. The 'Type' field indicates whether the term is used as a measure or dimension. The 'Description' field include the business rule that was applied to the element in business terms. In short the element-level metadata model supplies the definitions, measurements, and business rules for the terms and columns used in BI cubes and reports; whilst the overview and metadata information of a BI report is provided in object-level metadata model.

5 Conclusion

Through an in-depth case study with a large Australian organisation, the end-user metadata issue in the BI environment was examined in this study. The research findings indicate that end-user metadata play a vital role in improving the BI report analysis process and thus BI use. This finding of end-user metadata being more critical for business users confirms Foshay et al's [5] findings. The research further identified several crucial end-user metadata requirements and developed an integrated element and object-level metadata model to support BI users understanding of data and reports and therefore addressing the metadata needs of business users. The object-level metadata model provides an overview and important information about a report

or cube generated by a BI system. It contains descriptive information that has been missing in prior metadata models, aiming to enhance end-user understanding of a BI report. In the element-level metadata model, it supplies nine metadata fields to assist end users in interpreting a term or column of a BI report. This object and element-level metadata model for business users may lead to faster and better understanding of a BI report. Moreover the study has provided support for the importance of end-user metadata in the BI environment by providing empirical evidence from a case study of a large Australian university. Also the findings suggest that end-user metadata are critical for business users, especially in interpreting BI reports or cubes and in understanding the terms or business rules underlying a BI report. Of interest, the end-user metadata model was successfully implemented in the institution's BI environment.

The study offers a number of useful insights for IT practitioners responsible for BI application design and development. The end-user metadata model on object and element levels will help them facilitate the critical metadata needs of business users in supporting the BI report analysis process. Such a 'user-oriented' approach by BI practitioners would lead to greatly improved end-user satisfaction. One important limitation of the research is that the support for the findings which underpin the model is based on a single case study. However the selected case represents a large Australian institution with advanced BI applications and an enterprise-scale BI environment and is therefore an important basis for the study. Lastly, the model provides a foundation for BI stakeholders to start with their end-user metadata design and possibly customise it according to individual organisational needs. Nonetheless, the generalisability of this work should be tested in other domains and by using other research methods.

Acknowledgment. An earlier version of this paper was presented respectively at the 7th International Conference on Cooperation and Promotion of Information Resources in Science and Technology and the 16th post-ICONIP Workshop on Advances in Intelligent Computing.

References

1. Ballard, C., Farrell, D.M., Gupta, A., Mazuela, C., and Vohnik, S. (2006) Dimensional Modelling: In a Business Intelligence Environment, IBM, New York.
2. Beyer, M. A. (2007) Why Metadata Matters to Business Intelligence Initiatives, retrieved 1 June 2012 from http://www.gartner.com/DisplayDocument?doc_cd=144814&ref=g_fromdoc
3. Chen, L., Soliman, K.S., Mao, E., and Frolick, M.N. (2000) Measuring User Satisfaction with Data Warehouses: An exploratory study. Information & Management, 37(3), 103–110.
4. Clavier. P.R., Lotriet, H.H. and van Loggerenberg, J.J. (2012) Business Intelligence Challenges in the Context of Goods- and Service-Dominant Logic, Proceedings of 45th Hawaii International Conference on System Sciences.
5. Foshay, N., Mukherjee, A., and Taylor, A. (2007), Does Data Warehouse End-user Metadata Add Value? Communications of the ACM, 50(11), 70-77.

6. Gartner (2012) Gartner Executive Programs' Worldwide Survey of More Than 2,300 CIOs Shows Flat IT Budgets in 2012, but IT Organizations Must Deliver on Multiple Priorities, retrieved 1 Feb 2012 from http://www.gartner.com/it/page.jsp?id=1897514
7. Hess, T. J. and Wells, J. D (2002) Understanding How Metadata And Explanations Can Better Support Data Warehousing And Related Decision Support Systems: An Exploratory Case Study, In Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS).
8. Inmon, W., O'Neil, B. and Fryman, L. (2008) Business Metadata, Capturing Enterprise Knowledge, Elsevier.
9. Inmon, W., Strauss, D and Neushloss, G (2008) DW 2.0: The Architecture for the Next Generation of Data Warehousing, Morgan Kaufmann, Burlington, MA.
10. Little, R. G. and Gibson, M. L (2003) Perceived Influences on Implementing Data Warehousing, IEEE Transactions on Software Engineering, 29(4), 290-296.
11. Luan, J. (2002). Data Mining and Its Applications in Higher Education. New Directions for Institutional Research, Spring, no.113, 17–36.
12. Marchand, D. & Peppard J.(2008) Designed to Fail: Why IT Projects Underachieve and What To Do About It. retrieved 1 June 2012 from https://www.som.cranfield.ac.uk/som/dinamic-content/media/ISRC/Designed%20to%20Fail%20Working%20Paper.pdf
13. Miles, M. B. and Huberman, A. M. (1994) Qualitative Data Analysis: An Expanded Source Book, Sage, Thousand Oaks, CA.
14. Negash, S. (2004) Business Intelligence, Communications of the Association for Information Systems, 13, 177-195.
15. Ponniah, P. (2001), Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals, Wiley, New York.
16. Sen, A. (2002) Metadata Management: Past, Present and Future, Decision Support Systems, 37(1), 151-173.
17. Shankaranarayanan, G., and Even, A. (2006) The Metadata Enigma, Communications of the ACM, 49(2), 88-94.
18. Tvrdikova, M. (2007) Support of Decision Making by Business Intelligence Tool, In Proceedings of the 6th International Conference on Computer Information Systems and Industrial Management Applications.
19. Vaduva, A. and Dittrich, K.R. (2002) Metadata Management for Data Warehousing: Between Vision and Reality, in Proceedings of the International Symposium on Database Engineering & Applications, Grenoble , France.
20. Walsham, G. (1995) The Emergence of Interpretivism in IS Research, Information Systems Research, 6(4), 376-394.
21. Wells, J. D. and Hess, T. J. (2002) Understanding Decision-Making in Data Warehousing and related Decision Support Systems: An Explanatory Study of a Customer Relationship Management Application, Information Resources Management Journal, 15(4), 16-32.
22. Wixom, B. and Watson, H. (2010). The BI-Based Organization, International Journal of Business Intelligence Research, 1(1), 13-28.
23. Yin, R. (2009) Case Study Research: Design and Methods, 4th edition, Sage, Thousand Oaks, CA.