# Metadata implementation for a Business Intelligence environment

Yuriy Verbitskiy, University of South Australia, Adelaide, Australia, yuraverb@hotmail.com
William Yeoh, University Tunku Abdul Rahman, Perak, Malaysia, yeohgs@utar.edu.my
Andy Koronios, University of South Australia, Adelaide, Australia, andy.koronios@unisa.edu.au

## Abstract

Providing easily accessible interactive information for managers, business analysts and decision makers is a vital issue for any modern university. Implementing a Business Intelligence environment to ease the decision making process is becoming more widespread in universities. But with increasing complexity of IT environments and business processes, it is becoming even harder to make the right decisions.

Metadata for business users plays a critical role in understanding data and the BI environment. A good metadata mechanism is able to provide confidence to business users during the data analysis process. However, there has been little empirical research about metadata implementation in the BI environment. In particular, such critical questions as 'Why does BI need metadata?', 'What metadata does BI need?', 'What are the requirements for a metadata project?' and 'How to implement metadata in BI?' have not been investigated in depth.

Answers to these questions are explored in this study through an action research undertaken at a large university in Australia. The research findings in conjunction with the conclusion of the study can help BI stakeholders in identifying the critical metadata needs required for a more optimised BI deployment. It also may assist BI vendors in improving the business-side metadata which are crucial for a successful BI endeavour.

## Introduction

Business Intelligence (BI) applications have been dominating the technology priority lists of many CIOs (Gartner, 2007, 2008, 2009). Gartner Research predicts that the BI market will be in strong growth till 2011 (Richardson, Schlegel, Hostmann, & McMurchy, 2008). According to (Hancock & Toren, 2006), "Business Intelligence is a set of concepts, methods, and technologies for turning separated data in an organization into useful information in order to improve business performance". The main element of a typical BI environment is an integration process, which encompasses moving data from different sources into one integrated place, storing the data there, analysing data and presenting the data to end users. In other words, BI is an approach that allows business users to leverage the data for making informed business decisions, which is a modern example of decision support systems (Lawton, 2006).

Being a powerful approach for working with diverse data within a large organization, the BI system is a complex and resourceful undertaking. For instance, Gartner formulated 12 key technical capabilities that should be delivered by the BI environment and advocates that the BI vendors implement at least 8 of these 12 capabilities (Richardson et al., 2008). Ultimately, the main indicator of success in implementing a BI system is the level of end user satisfaction (Foshay, Mukherjee, & Taylor, 2007). Understanding of the complex BI environment by business users is critical issue because "a lack of skills in using information, tools and applications presents a big barrier to the success of BI" (Schlegel & Rayner, 2009). The difficulties in understanding BI are associated with inadequate training (Sheina, 2007) and disconnect between business and technical users (Sherman, 2005). Thus, understanding of the BI environment by business users is a barrier for the whole process of using BI.

Another problem is the understanding of the data that business users work with in the BI environment. Allowing end users to handle sets of data from different data sources creates a problem because most users are not familiar with the data and data sources. So it appears that, on the one hand, BI is an effective technique that allows the delivery of required data from various sources to the end users. On the other hand, business users have problems with understanding the BI environment and the data provided by BI applications. Without solving this issue BI will not help users to make decisions effectively. That is, the understanding of the data by business users is a pressing issue in BI environment. Making decisions based on the results of BI tools is not surprisingly the biggest challenge for end users (Lawton, 2006; Xu, Zeng, Shi, He, & Wang, 2007).

Finally, focus on the technological issues without balancing with business orientation is a frequent reason for unsuccessful outcome in BI endeavours (Sheina, 2007). It is vital to address the needs of business users since they represent the major part of BI users who require additional support in using BI. While technical users understand the BI environment because it is one of their primary work objectives, business users concentrate on business processes and need an instrument that would allow them to be confident in using data and BI tools in general.

In response to this, metadata serves as a mechanism that provides the context about the data of the BI environment (Tvrdíková, 2007) and the context about *objects* of the BI environment. It allows addressing the business orientation during the BI endeavour. Without metadata the data in an enterprise cannot be understood properly (Inmon, O'Neil, & Fryman, 2008). Gartner Research argues that metadata management is one of the most important functionalities that the BI environment should deliver (Richardson et al., 2008).

Therefore, the purpose of this research is to explore the metadata implementation in the BI environment. Most of the existing BI literature does not focus on practical reasons why BI needs metadata. There are several studies that discuss the different types of metadata, but none has addressed the actual metadata required for systems implementation. There has been little empirical research about metadata implementation in the BI environment. In particular, such critical questions as ‘Why does BI need metadata?’, ‘What metadata does BI need?’, ‘What are the requirements for a metadata project?’ and ‘How to implement metadata in BI?’ have not been investigated in depth. This research is going to overcome the problem by exploring the answers to these four questions through an action research in a large university in Australia.

The next section discusses the research approach before elaborating on data collection methods and case background. The third section presents findings for the first two research questions together with a metadata model. The following section describes the requirements for the metadata implementation. Finally, the metadata implementation process is presented with the steps taken and issues found on the way.

**Research methodology and research environment**

Despite daunting complexities in implementing BI systems, there has been little empirical research about the metadata specifically for the BI environment. This study investigates the whole process of metadata implementation in the BI environment of a large Australian university. It identifies the specific metadata needs of the organisation, and subsequently proposes and implements a solution by developing a metadata prototype. When a process of change is the subject of research, the most appropriate methodology is an action research (Benbasat, Goldstein, & Mead, 1987), which was deemed to be the most suitable approach for this project.

The research is conducted in collaboration with the university's Business Intelligence team. Having more than thirty thousand students, it is crucial that the institution exploits a powerful data analysis and performance management tool. That is why the university acknowledges the importance of the BI technology and strives to deliver it to the wider organisational community. It also admits the fact that it does not have a consistent metadata implementation which benefits business users of the BI environment. These factors make the university an ideal environment for investigating the metadata implementation in the BI environment.

The paper provides the discussion and findings regarding the four research questions that have been investigated during the action research study:

1. Why does BI need metadata?
2. What metadata does BI need?
3. What are the requirements for a metadata project?
4. How to implement metadata in BI?

The design of the whole action research is provided in Figure 1. To consider the action research process more closely, the process of conducting action research can be described as a continuous cycle of four main elements: diagnosing the environment to define the problem, action planning, action and evaluation of the action (Cherry, 2002). These elements help to develop an action research design diagram with the additional research questions where the required research questions are supported by a literature review, an action research process or both.

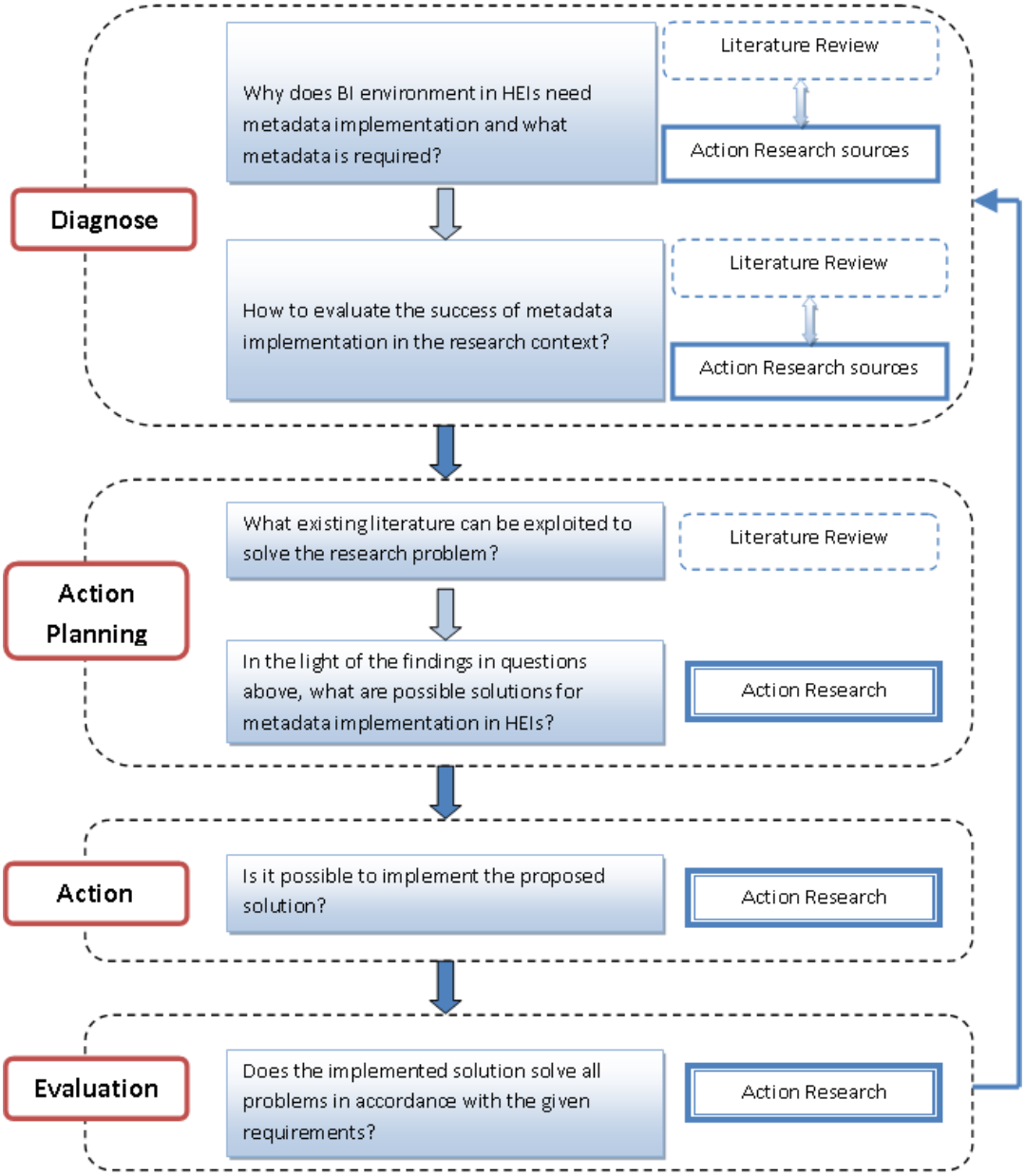


**Figure 1** The design of action research

The main data collection methods of the study are semi-structured interviews and a document analysis. A semi-structured interview allows some latitude in getting detailed information from interviewees during the interview, but at the same time to have general questions so that all participants are treated in a consistent manner (McMurray, Pace, & Scott, 2004). The interview questions posed can be divided into two main parts. The initial part of general questions focus on discovering various situations where metadata can be needed. The second part relates to more concrete questions that try to determine the exact elements of metadata users would like to have. The interview was conducted with 9 research participants from different departments and with different roles in the BI environment of the university: 6 BI team members, 2 business analysts, and 1 business user.

Document analysis allows investigating the research environment and the status of the metadata in the case organisation. At the time of the study, the organisation was migrating to Cognos BI environment version 8.4 from the previous Cognos version 7. Within Cognos version 7, some metadata techniques have been put in place but they are in immature condition. Technical users use Excel spreadsheets with the technical information regarding the data in the BI environment. Business users exploit structured descriptions for cubes and reports in the front-end of the BI environment. Also, they are able to use 'Glossary'

application which represents a web page with a list of terms and descriptions, as illustrated in Figure 2. While this metadata provides some helpful information for the business users, the BI team realises that they require a more consistent, attractive and user-friendly solution for metadata implementation.

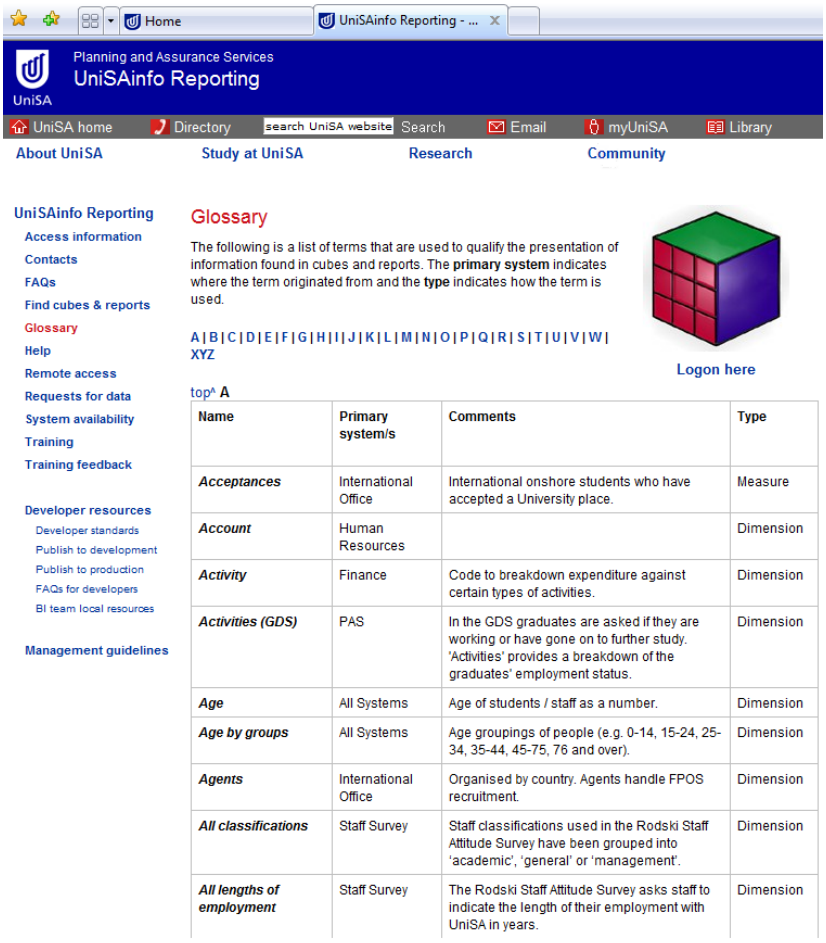


**Figure 2** Glossary website

## Metadata need and metadata model

The result of the interview analysis demonstrated that the following reasons drive the need for metadata in the BI environment:

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

The result of the data analysis indicates that the metadata solution ought to consist mainly of metadata for business users. The interview results clearly showed that the majority of participants considered business users as important, and they were more likely to provide detailed elements of metadata for the business users' BI environment. Although the remaining metadata is useful for various purposes it is not as crucial as metadata for business people. This proves that business orientation is paramount in the case organisation.

Metadata for business users is known also as business metadata. Sometimes business metadata overlaps with technical metadata – metadata that contains technical information and used primarily by technical users. But during this project the focus was primarily on business metadata that could be identified as important for business users. Usually technical metadata is not relevant for business users therefore it was not deeply investigated in this research.

The resulting metadata model (or structure of metadata) takes the existing parts of the business metadata as a basis. The existing metadata consists of two types of metadata: metadata that provides additional information about reports, cubes, documents (such metadata has been named *object* metadata) and metadata that provides the description about terms or data elements used in the BI environment (*element* metadata).

Figure 3 depicts the proposed metadata for the *object*. The enhanced *object* metadata has a number of new metadata fields which were mentioned by the interview participants. Below are descriptions of several fields of the *object* metadata. 'Type' field defines whether the *object* is a report or cube. 'Scope' field defines what is included and what is excluded from the data, e.g. whether a report contains all students or only domestic students. 'Usage' field describes how or for what purpose the data in the *object* should be used. 'Source systems' field shows from which source applications or external sources the data were taken. 'Primary audience' field describes who are supposed to be the key users of the reporting *object*. 'History' field allows finding the previous versions of the *object*.

| |
|---|
| 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 |

**Figure 3** Proposed object metadata

The proposed metadata for the *element* has 5 new metadata fields: 'Business acronym', 'Places of use', 'History', and 'Owner'. 'Business acronym' shows the business acronym for the element if it exists. 'Places of use' metadata field shows where the element is commonly used across the BI environment. 'History' field provides the reference to the previous version of the *element*. 'Owner' field represents the owner of that particular *element*, it may be the same as in 'Contact person' field for the parent *object* but it may also be different.

The current *element* metadata fields have been transformed to the following fields: 'Name of the element', 'Primary system', 'Description' and 'Type'. 'Primary system' metadata field indicates where the *element* originated from. 'Type' indicates whether the term is used as a measure or a dimension.

| Name of the element |
|---|
| Business acronym |
| Primary system |
| Description |
| Type |
| Places of use |
| History |
| Owner |

**Figure 4** Proposed element metadata

**Requirements for the metadata implementation**

The requirements for the metadata implementation allow evaluating the success of the metadata implementation and show possible implementation directions. After the analysis of the literature the following requirements, categorised into several general areas, were developed (Table 1). "Presentation of metadata" area defines the requirements that relate to the user's interface of the metadata solution. "Metadata repository" area presents the requirements regarding the technical metadata repository issues and metadata model. "Metadata infrastructure" area provides a list of technical requirements related to the metadata solution in general. "Metadata management" area offers the requirements regarding the continuing management and support of the metadata solution.

**Table 1** Requirements for the metadata implementation

| Area | Requirement | Priority |
|---|---|---|
| **Presentation of metadata** | | |
| | Layered presentation of metadata | MEDIUM |
| | Providing names and emails of contact persons | HIGH |
| | Browsing, Searching, Facets, Key words, Filters | HIGH |
| | | |
| **Metadata repository** | | |
| | Easy customization of metadata structure in the future | HIGH |
| | Hierarchic metadata classification | HIGH |
| | Metadata structure is shown in metadata model to help users | HIGH |
| | Refreshing of metadata from various sources on a regular basis | HIGH |
| | Import/Export functionality to/from Microsoft Excel | HIGH |
| | | |
| **Metadata infrastructure** | | |
| | Accessibility from multiple places, uniform access mechanism | MEDIUM |
| | Integration with existing BI environment, context-sensitivity | HIGH |
| | Interchangeable metadata format | MEDIUM |
| | API for access by other software applications | MEDIUM |
| | | |
| **Metadata management** | | |
| | Easy to support and change | HIGH |
| | Metadata stewardship | HIGH |
| | Access control | HIGH |
| | Metadata change technique | HIGH |
| | Metadata version management strategy | LOW |
| | Notification mechanism | LOW |
| | Metadata quality | HIGH |

A group meeting with the BI team aimed to identify the requirements that were the most critical for the metadata implementation. For that each requirement was explained to the participants and they were asked to set priority for the requirements according to three levels ("High", "Medium" and "Low"). Usually the members of the BI team suggested similar priorities for the requirements. When there was a disagreement between the members, they worked towards generating a group accord.

The requirements with high priority basically define the success of metadata implementation. If these requirements are met by the metadata solution, it would mean that the solution covers the most important aspects of the metadata implementation process.

**Implementation of metadata solution**

Implementation process consists of the following general steps:

1. Integration with the BI environment;
2. Metadata prototype (web interface and database structure);
3. Automatic metadata import;
4. Metadata change management;
5. Metadata interface and database structure improvement.

At the beginning of the implementation process it was crucial to investigate and implement the mechanism that would allow to integrate external application into the existing BI environment from user's point of view. For Cognos BI 8.4 it was achieved by modifying the Cognos JavaScript files that are responsible for handling user's interactions in Cognos Viewer. The modifications allow users to select any data element on the report and run the metadata application. JavaScript modifications also transmit the value of the selected data element and string identifier of the current report to the metadata application. In this case the metadata application means a web application that is responsible for providing user interface.

After the integration mechanism had been implemented, the next step was to develop a metadata prototype that consists of a basic web interface part and basic database part. It allowed to see what can be shown to users and what functionality can be provided in the metadata application. The discussion of the metadata interface revealed that the application should automatically import not only technical information about *objects* and *elements* from Cognos BI, but also relations between *objects* and *elements* since they could be changed any time. The metadata prototype also allowed to understand better the requirements for metadata application, their relevance and importance.

The additional application has been developed for an automatic metadata import. That application accessed and processed different parts of Cognos BI environment in order to import all required technical metadata. The application is scheduled to run overnight and update only technical metadata in the metadata database with the latest changes in the data model and reporting environment.

During the implementation of the automatic metadata import, it was found that the metadata application has to manage the changes in the data model and reporting environment. For this a metadata change management technique has been developed as part of the web application to allow a data administrator to define previous versions for the *objects* and *elements* in the metadata application. It will implement version control and also allow transferring business metadata from the previous version of the *object* or element to the latest version.

The implementation of the automatic metadata import and metadata change management required changes in the metadata database and metadata interface. Hence these processes were usually implemented at the same time. Furthermore, the overall metadata database and interface improvement has been implemented at the end because of a number of improvements and proposals from the different types of stakeholders.

The general architecture with the main elements of the metadata application is presented in Figure 5.

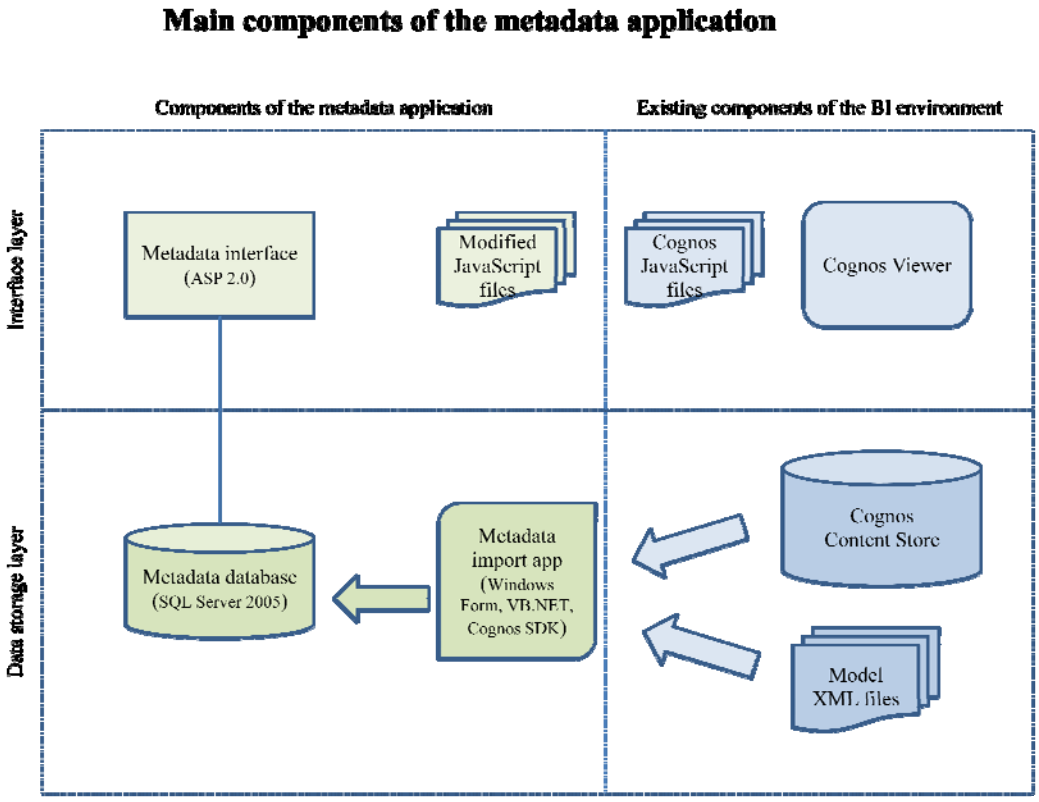


**Figure 5** Architecture of the metadata application

The figure shows the existing elements of the BI environment and the elements of the metadata application. The interface layer consists of the modified JavaScript files and web application that represents a metadata interface. The database layer consists of the metadata database and metadata import application. The metadata import application updates technical part of metadata in the metadata database while business metadata can be updated using the web interface metadata application.

Thus, in general, the metadata update process consists of two steps: automatic for technical metadata and manual for business metadata. However, business metadata are loaded automatically only once, and after that the users are able to edit or update business metadata manually. Metadata change management is also related to the metadata update process since it allows manual transfers of business metadata from the previous versions of metadata items to the latest metadata items. In this case, the business users will not need to update the business metadata for new versions of metadata *elements* and *objects*; it will be done by the data administrator.

Metadata interface requires discussion in detail since it is a key part from the user's point of view (Figure 6). Main sections of the metadata interface are: "Action Pane" area, metadata section and search results.

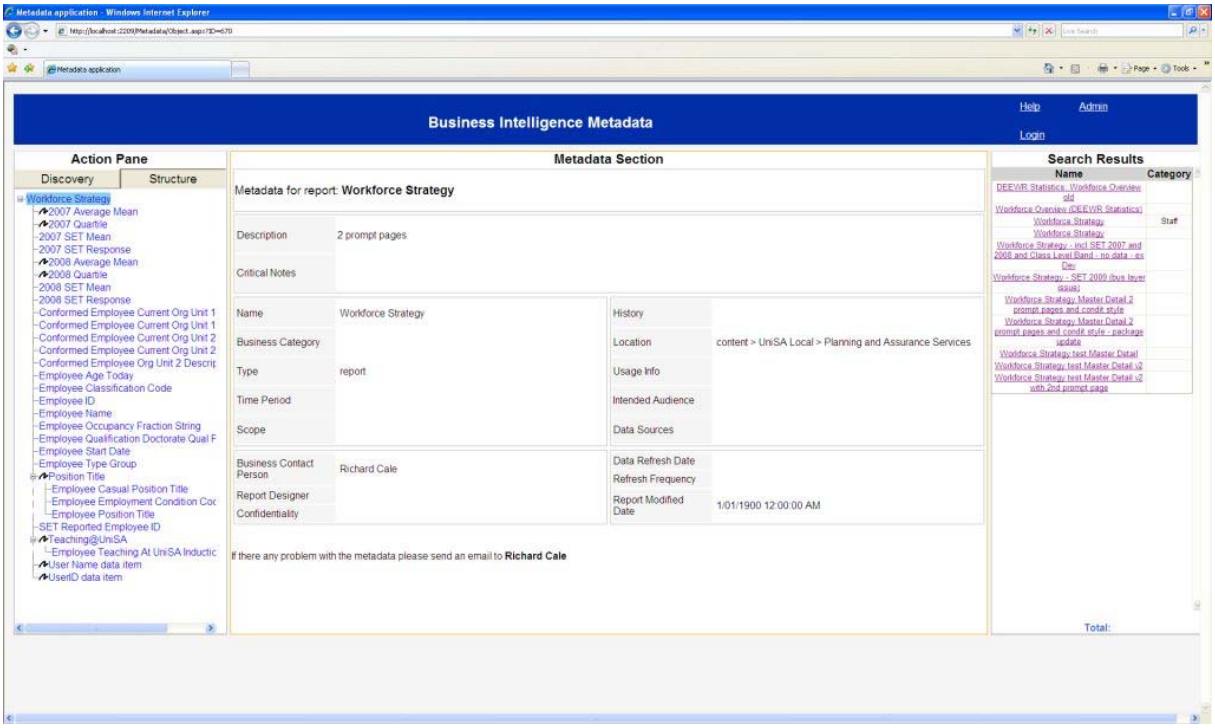


**Figure 6** Metadata interface

The “Action Pane” section includes two different views – “Structure view” and “Discovery view”. The first view shows all related metadata items in a structured tree view element. The tree view element shows current metadata *object* as a main parent node (“Report”) and all metadata elements that this *object* includes as children nodes. The tree view shows not only metadata elements that are stored in the metadata repository but also all unidentified data elements and calculations in the corresponding *object*.

The “Discovery view” provides the user with search and browse functionality. The user can choose what type of metadata they are looking for: *object* or *element*. Based on the selection of a metadata type, a list of browse categories with the most common values would be presented to the user. The user can select any of the given values and the resulting list will be filtered based on the selected values. User can also write the key words in the search bar to get the corresponding search results.

The main metadata section shows metadata about the current metadata item that can be *object* or *element*. The current metadata item can be changed by selecting it in a tree view or search results.

The “Search Results” section provides quick access to the latest search results through a list of names of corresponding metadata items.

Several key issues were identified during the metadata implementation process. One issue relates to the extraction of metadata from the BI environment. The main problem relating to the extraction is that the BI environment is not designed for the extraction of extensive metadata. That is why this process can be time consuming or not achieved in full. Another issue for the metadata implementation is the lack of functionality in the BI environment that would allow integrating seamlessly the metadata application with the BI environment. The integration mechanism should also transmit the current context from the BI environment to the metadata application. The good integration with the BI environment means that the business user would be able to get the metadata from within the BI environment.

As discovered during the implementation process these two issues – extraction of metadata from the BI environment, and integration with the BI environment – are the key requirements for the successful metadata implementation and the most challenging ones. Thus, the ability of the BI environment to support them should be investigated at the early stages.

**Conclusion**

A metadata project, same as the whole Business Intelligence endeavour, should be driven by business requirements and needs that are prone to change and development along the way. That is why the metadata implementation is a long-term process that requires the comprehensive implementation methodology. This paper presents the results of the empirical research about metadata implementation and describes the general technique that was used during the project. The paper provides a real practical example of such implementation with the answers to four key questions: 'Why does BI need metadata?', 'What metadata does BI need?', 'What are the requirements for a metadata project?' and 'How to implement metadata in BI?' Particularly, the implementation process has been described in detail with the explanation of the main implementation steps, general architecture, interface components and critical issues.

The research findings may be interesting for various BI stakeholders and metadata specialists in general, who would benefit from the description of a practical metadata implementation case. It can also help business users, who work with complex data sets and undertake comprehensive analyses in BI environment, to understand what metadata can offer to them.

**References**


Benbasat, I., Goldstein, D., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems [Electronic Version], *11*, 369-386,

Cherry, N. (2002). *Action Research. A Pathway to Action, Knowledge and Learning.* Melbourne: RMIT Publishing.

Foshay, N., Mukherjee, A., & Taylor, A. (2007). Does Data Warehouse end-user Metadata add value? *Communications of the ACM, 50*(11).

Gartner. (2007). Gartner EXP Survey of More than 1,400 CIOs Shows CIOs Must Create Leverage to Remain Relevant to the Business. Retrieved 01/04/2009, from http://www.gartner.com/it/page.jsp?id=501189

Gartner. (2008). Gartner EXP Worldwide Survey of 1,500 CIOs Shows 85 Percent of CIOs Expect "Significant Change" Over Next Three Years. Retrieved 01/04/2009, from http://www.gartner.com/it/page.jsp?id=587309

Gartner. (2009). Gartner EXP Worldwide Survey of More than 1,500 CIOs Shows IT Spending to Be Flat in 2009. Retrieved 01/04/2009, from http://www.gartner.com/it/page.jsp?id=855612

Hancock, J., & Toren, R. (2006). *Practical Business Intelligence with SQL Server 2005*: Addison Wesley Professional.

Inmon, W., O'Neil, B., & Fryman, L. (2008). *Business Metadata, Capturing Enterprise Knowledge*: Elsevier.

Lawton, G. (2006, September 2006). Making Business Intelligence More Useful. *Computer, 39*, 14-16.

McMurray, A., Pace, W., & Scott, D. (2004). Research: a commonsense approach. In (pp. 260-271): Thomson/Social Science Press.

Richardson, J., Schlegel, K., Hostmann, B., & McMurchy, N. (2008). Magic Quadrant for Business Intelligence Platforms: Gartner.

Schlegel, K., & Rayner, N. (2009). Key Issues for Business Intelligence and Performance Management Initiatives: Gartner.

Sheina, M. (2007). What went wrong with business intelligence? Retrieved 04/04/2009, from http://www.cbronline.com/article_cbr.asp?guid=BE8BE7D0-CD23-48CC-9FD4-42EF5486A846

Sherman, R. (2005). Business Intelligence Goes Back to the Future, Part 2: Couples Therapy for IT and Business Users [Electronic Version]. *Information Management Online*. Retrieved 10/10/2009, from http://www.information-management.com/news/1034667-1.html

Tvrdíková, M. (2007). *Support of Decision Making by Business Intelligence Tools.* Paper presented at the 6th International Conference on Computer Information Systems and Industrial Management Applications (CISIM'07).

Xu, L., Zeng, L., Shi, Z., He, Q., & Wang, M. (2007, 7-10 Oct. 2007). *Research on Business Intelligence in enterprise computing environment.* Paper presented at the IEEE International Conference on Systems, Man and Cybernetics, 2007. ISIC., Montreal, QC, Canada.