# Describing Data Quality Problem through a Metadata Framework

William Yeoh
*Information Systems, Deakin University, Burwood, VIC, Australia.*, gsyeoh@gmail.com

Te-Wei Wang
*Management Information Systems, University of Illinois at Springfield, Springfield, IL, United States.*, twang22@uis.edu

Yuriy Verbitskiy
*Planning and Institutional Performance, University of South Australia, Adelaide, SA, Australia.*, Yuriy.Verbitsky@unisa.edu.au

# DESCRIBING DATA QUALITY PROBLEM THROUGH A METADATA FRAMEWORK

**William Yeoh**
Deakin University
william.yeoh@deakin.edu.au

**Te-Wei Wang**
University of Illinois at Springfield
twang22@uis.edu

**Yuriy Verbitskiy**
University of South Australia
yuriy.verbitskiy@unisa.edu.au

## ABSTRACT

Data quality management is a pressing issue in the adoption and implementation of Business Intelligence system in every organisation. We believe that an extensive metadata infrastructure is the primary technical solution for management of data quality in Business Intelligence. We derived a set of data quality dimensions by carefully examining the data quality management principles and current BI environment. Then, a high-level metadata framework is proposed.

### Key Words

Metadata, Business Intelligence, Data Quality, Framework

## INTRODUCTION

Recently Business Intelligence (BI) market has experienced high growth and BI technologies have consistently received attention by many Chief Information Officers (Gartner, 2012). BI is "*a broad category of technologies, applications, and processes used for gathering, storing, accessing, and analysing data to help its users make better decisions*". In a typical BI environment, data from different sources are extracted, transformed and loaded (ETL) into an Enterprise Data Warehouse (EDW) and, from there, are used for reporting across the organisation (see Figure 1). During this process, data quality plays a critical role in BI success since poor data quality can hinder business decisions at all levels of the organisation (Daniel et al. 2008; Khatri and Brown, 2010). Being very complex at the same time, data quality is an issue that costs billions of dollars (Khatri and Brown, 2010; Vassiliadis, 2000) because data used in the BI environment are used for decision making at every organisational levels and various business processes (Ballou and Tayi, 1999).

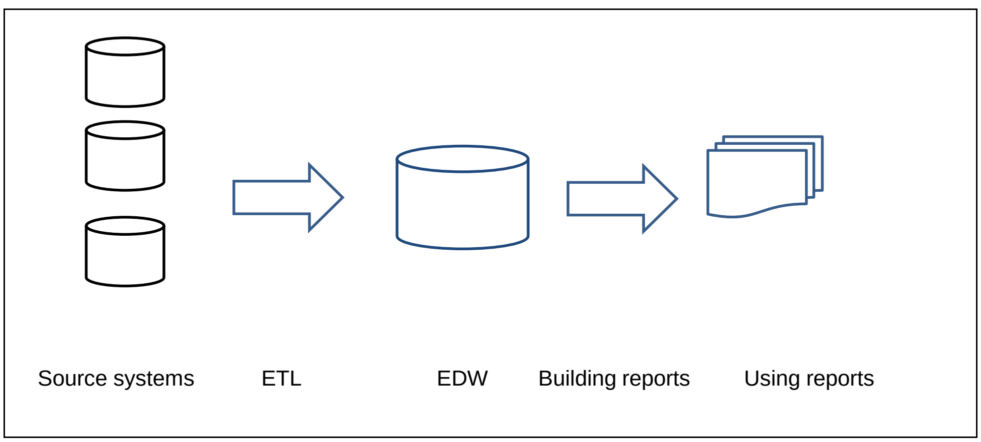


**Figure 1. Main Business Intelligence stages**

In general, there are two main sources of data quality issues involved in BI projects. The first is rooted in the source systems. There are always data quality issues where data do not conform to business requirements even after extensive testing by business users and IT developers. The second source of data quality problem resides on the entire BI processes. By integrating data from the source systems BI projects create new requirements for the existing data. Now the data should conform not only to the original source system requirements but also to new business requirements gathered for the BI project (Ballou and Tayi, 1999). It is a major task of the BI processes to drive more data exploration and analysis, thereby exposing all existing data quality issues, even those which were not considered before. BI processes may themselves become the source of data quality issues because they involve very complex operations with the data during all BI stages.

In a typical BI process such as the one showing in Figure 1, the ETL represents the first key challenge for any BI project, "*70% of the risk and effort in the DW/BI project comes from this step*" (Kimball et al., 2008). The main reason for this is because during this step data with a different structure and designed for a different operational purpose are brought into one place and transformed in such a way as to enable them to be integrated and used together (Ballou and Tayi, 1999). It is challenging to address varying levels of data quality across source systems or even within one source system (Daniel, 2008).

The second risk for data quality is the database designed for the EDW. Database design for an EDW mainly depends on the business requirements for decision making, the structure of the available data in the source systems and quality levels of the source data. The database design process is extremely complex in its own right involving understanding of the source data and related business processes, analysing business requirements, and ensuring naming conventions and data integration through using conforming dimensions and facts.

The next step is report building. Production of reports may not seem risky in terms of data quality if the data in the EDW conform to business requirements, but this BI process has its own reasons for concern. Full understanding of original business requirements for the EDW data set, EDW data set structure, related business processes and specific report requirements are the main factors for a successful report development process. Merely to understand the data structure in the EDW can be a challenge since it usually stores multidimensional, integrated data that were designed to be used for different purposes. Presenting the data in the simplest way is another challenge. It requires solid knowledge of the available BI Reporting tools.

After the report development process is finished with the reports being approved by the corresponding business unit, there comes the last BI stage – using the reports. Even during this stage there is a risk of data quality issues appearing, especially when the report is used widely across the organisation. Some users do not understand the objective of the report or some of its content. Some users might find that the report functionality does not meet their needs and requires modification. Some users might claim that the data in the report are wrong. These issues are related to business users' understanding of the data and their experience with the Reporting tools. Adequate training as well as specific information related to the report is required.

For BI projects to succeed it is important to ascertain where the data quality issues originate in order to respond to them appropriately and also to inform all stakeholders involved in the BI projects about their sources. BI will expose and amplify data quality issues along the way but will not show BI as a source of all data quality issues. For data quality issues to be fixed in source systems requires appropriate business processes, tools and staff. It is useful to discuss responsibilities and possible actions at the beginning of the BI project in case critical issues are found later.

Data quality represents a huge risk for BI implementation. It is essential for BI to have comprehensive data quality management, to show the status of data quality across all BI stages starting from source systems and ending with the use of reports. Otherwise BI implementation will always have issues with business confidence in data quality and common understanding within the organisation of data quality issues and their sources.

There is a need for a framework that will assist with general BI issues, and address data quality in particular. This paper advocates that gathering, managing and providing BI metadata regarding data quality to all relevant types of users - technical and business users - is the best possible technical solution. Metadata serves as a mechanism that provides the context about the data within the BI environment (Tvrdíková, 2007) and the context about different elements of the BI environment. Without metadata the data and the BI environment itself cannot be understood properly (Inmon et al., 2008). In fact, Gartner argues that metadata management is one of the most important functionalities that the BI environment should deliver (Richardson et al., 2008).

In this paper we focus on data quality issues that originate from BI processes. We also discuss general principles of data quality management in the BI environment, present the key metadata elements and show how they can be applied to support data quality management in BI. In brief, this paper presents discussion and findings regarding the following research questions:

1. What data quality issues does Business Intelligence face?
2. What are the general data quality management principles for BI?
3. How can BI Metadata infrastructure support data quality management?

We address question 1 and 2 above through literature review. Then, an exemplar metadata model is presented to illustrate the feasibility of using a metadata model for quality management purpose. The remainder of this paper has been structured as follows. The next section provides a review of related work before elaborating on the data quality issues in the BI environment. The fourth section then outlines the data quality management principles. In the fifth section the authors present and discuss the metadata framework addressing data quality issues in the BI environment. The final section provides a summary of the research.

## RELATED WORK

There have been several studies addressing data quality issues in the BI environment. It is common to use some metadata for dealing with data quality issues in BI (Daniel et al., 2008; Vassiliadis, 2000; Vassiliadis et al., 1999; Foshay et al., 2007; Rodic and Baranovic, 2009; Li and Osei-Bryson, 2010; Farinha et al., 2009; Gao et al., 2006). Some researchers acknowledge the characteristics of the BI environment which require them to focus on specific data quality dimensions (Ballou and Tayi, 1999; Vassiliadis et al., 1999; Foshay et al., 2007; Li and Osei-Bryson, 2010). However, not all previous work focused on dealing with data quality issues as such; some work focused on other topics in the data quality area like prioritisation of data quality enhancement in the data warehouse (Ballou and Tayi, 1999) and overall understanding of data quality across the organisations in specific industries (Gao et al., 2006).

There are literatures linking metadata model with data quality. Jarke et al. (1999) described a tool that presents metadata quality information and, therefore, allows analysing the status of metadata information stored in the metadata repository with ease. In the work of Vassiliadis et al. (1999) and Jarke et al. (1999), the authors proposed a wide-ranging metadata model with an added quality metamodel. This metadata model covers information about the data warehouse in three perspectives: conceptual, logical and physical. The extensive metadata model was enhanced by adding quality information about each object of the metadata model. The quality model also allows the user to distinguish between subjective quality goals and objective quality factors.

In addition, the idea of providing quality metadata about the data from which reports are generated is investigated in Daniel et al. (2008). Focusing on three quality dimensions (completeness, consistency and confidence) the authors proposed the use of quality-aware reports which allow users to get the quality status of the report, to change the report output by tuning quality indicators and to share quality related feedback. Also, implementation of data quality processes based on metadata and integrated into the ETL process is described in Rodic and Baranovic (2009). Data quality processes are proposed to be implemented by using data quality rules applied in two certain parts of the BI environment (Staging area of ETL and EDW). This approach is suggested for situations where data quality tools are not available or there is difficulty with data quality tool integration. To address the challenge of constant changes in the quality of BI components (source systems, processes, etc.) and in the users' requirements, Li and Osei-Bryson (2010) proposed a mechanism to capture and present quality changes to relevant end-users. For this they suggested two new services in the data warehouse architecture – Quality Factory, that stores quality information, and Quality Notification Service, that is responsible for notifying interested end-users. However, there is little literature that has explicitly addressed the data quality issues in the entire BI manufacturing process from source systems to report generation. In other words, a comprehensive management mechanism for both quality factory and quality notification system has not been addressed.

## DATA QUALITY ISSUES IN THE BI ENVIRONMENT

Often, data quality is referred to as information quality with the main goals of "*meeting business requirements*" (Richardson et al. 2008) or "*satisfying data usage requirements*" (Khatri and Brown, 2010). This definition conforms to a widely accepted understanding of data quality as being multidimensional (Gao et al., 2006) since there are many aspects of meeting business requirements. Data quality issues may appear at any stage of the BI cycle (Figure 2). If during the implementation of any BI stage the requirements are not fully understood or not implemented appropriately, the result will not meet the business requirements. In the end it does not matter what the exact reason is for a wrong number in the end user's report, but in order to manage data quality we certainly need to understand all data quality sources. Finally, if the user of the report does not have full confidence in the data presented, does not understand what it means and does not have any tools to further investigate their concerns, than he might think that there is a data quality issue even when this might not the case. Derived from Figure1, Figure 2 shows all BI stages as sources of data quality issues. Based on Figure 2, the detailed explanation of data quality issues at each stage of BI manufacturing process are as follows.

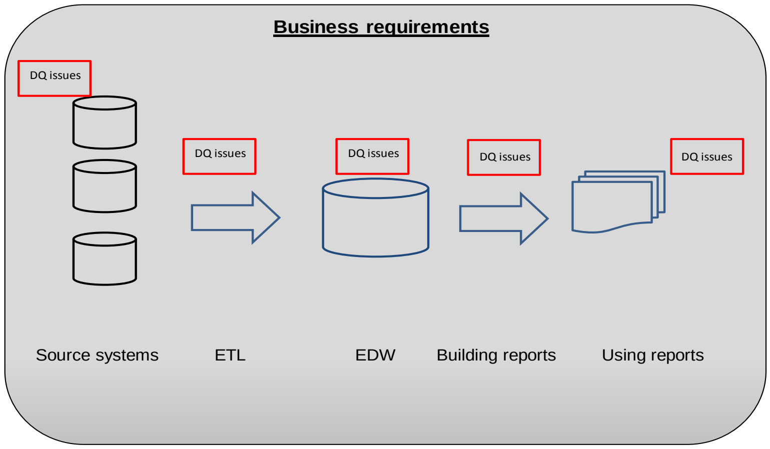


**Figure 2. Main Business Intelligence stages as data quality sources**

### ETL development

Before starting to build the next ETL implementation technical staff needs to audit the source data (perform data profiling). Following this they are able to present the results of data profiling to business users in order to validate the business requirements in terms of data quality, enabling them to decide how to implement the ETL in the most effective way. The challenge here is that technical staff must have regular data profiling results for every significant ETL stage and as a result to have profiling results before and after the ETL process as regularly as required (e.g. daily). This allows the data quality to be constantly monitored during the ETL process and to be communicated directly to business users if required.

Having data profiling statistics for ETL greatly assists with implementing comprehensive data quality processes. It is proposed to have a centralized dimensional data model called Error Event Schema to record all errors during the ETL process (Kimball et al., 2010). Similarly we believe that it is necessary to have *overall ETL statistics* which stores not only errors but also all major profiling results during the ETL process to allow monitoring of data quality through time and to pick up new issues that arise in source systems or during the ETL process.

Another required data quality statistic is related to relationships between the fact table and the dimension. It is a quite common situation, especially immediately after building a new fact table, to find that a number of records in the fact table do not have corresponding values in the dimension. This usually happens because of a mismatch between data used for populating the fact table and data used for populating the dimension. It means that there are data quality issues in one of the sources. In this situation the predefined foreign key (e.g. '-1') in the fact table is assigned and it relates to a specific "Unknown" record in the dimension. Getting the total number of unknown records in the fact table related to the specific dimension can be implemented simply by counting records with the foreign key value equal to '-1'. This *"Unknown" statistic* is an essential candidate to be included in the overall ETL statistics. Ideally there should not be any unknown records in the fact table but it is a long-term goal that cannot be achieved in a short term and that is why it requires constant monitoring.

It is also possible to have records in the fact table whose foreign keys are not associated with any primary keys in the dimension due to an error in ETL. This happens when the dimension gets rebuilt (with new primary keys) without rebuilding (changing foreign keys in the fact table to match new primary keys in the dimension) the fact tables. This type of issue signals critical problems in the ETL implementation that may require a review of all ETL development processes. This is a crucial statistic to have as part of the overall ETL statistics.

If there are data quality issues with values that form the dimension then it would be reasonable to have statistics about the number of records existing in the fact table and associated with every (or top 10/100) value in the dimension. This would allow a focus on the key values in the dimension for the data quality testing. Except for overall ETL statistics there should be a clear automatic presentation of relationships between all data elements involved in ETL – ETL relationship diagram. This diagram would enable precise impact analysis for changes in the operational sources and ETL itself. It would also allow the ETL result – EDW data column – to be confirmed with business requirements or, in other words, to get a lineage for the data element to confirm the source and transformations along the way. Providing lineage and dependency information on data is one of the essential requirements for any ETL system (Kimball et al., 2008). In situations where the ETL is implemented via scripting, it is hard to deliver the relationship diagram automatically and therefore the whole ETL process becomes a black box without any means of quality assurance. Using the manually updated documentation to resolve this issue usually results in out-of-date, inconsistent documents to support, which slows down the whole ETL process.

### EDW design

During this stage EDW modellers identify and analyse the corresponding business process, determine the grain of the fact table and identify candidate dimensions and fact measures (Kimball et al., 2008). In addition they have to implement data integration strategy using conformed dimensions and conformed facts.

Each fact table is designed to assist with a particular business process. Identifying and understanding the business process can be challenging in a large organisation. Designing the new fact table is an even more complex task. It is important to store information about each business process (and associated business requirements) implemented in the EDW and information about each related fact table (e.g. description of the main goal of the fact table and for what purposes it can be used). This helps to clarify business requirements behind each fact table and gives a picture of how each business process is covered in the EDW. This and other EDW specific information can be stored in overall EDW statistics.

Fact table grain is the level of detail in the fact table and is the key to understanding the fact table. In order to properly determine the grain for the future fact table it is necessary to not only analyse the business requirements but also to check what data is available in the source systems. The general rule here is to design the fact table with the lowest grain possible (Kimball et al., 2008). There are many factors involved which may lead the DW team to a different decision (e.g. performance, business requirements, simplicity or existing grain in other similar fact tables). Therefore it is very important to provide information about the grain of the fact table in the overall EDW statistics. Each fact table has a number of joined dimensions and fact measures. This information together with the grain defines the fact table. Having that information in the overall EDW statistics allows users to see how business requirements associated with particular business processes were implemented in the EDW.

### Report development

The next source of data quality issues is the report development process. If a report developer does not fully understand business requirements or is not confident about the relationship between data in the EDW then it is highly likely that the resulting report will not meet expectations. Understanding business requirements is not only about having sufficient documentation. It is also about knowing what data is available in the EDW, its data structure, and main features. For instance, the grain of the fact table is the key information for the report developer when he builds the report based on the corresponding fact table. The report developer might think that he groups the data at some level when in fact he groups the data at another level simply because he does not know the grain of the related fact table. At the same time determining and describing the grain for complex data sets of the report is also important.

Describing the report content becomes increasingly difficult when more data are available in the EDW. Powerful reporting tools allow working with complex data using complex aggregations and presenting them in different ways (e.g. lists, crosstabs, charts, maps) but they do not help with selecting the most transparent presentation method. Even describing the data sets in the report can be challenging depending on the complexity of data involved and filters applied.

### Report use

At the last stage when report users are using reports data quality issues may still arise. Mainly these relate to situations where report users misunderstand the main goal of the report, and misuse the information presented in the reports.

One way of dealing with this issue is to provide additional information or metadata required for the report user to understand the report. This metadata should contain the context of the report (e.g. report purpose, audience, and usage information), information about related business processes and data quality information. Another way is to provide a mechanism to allow the user to give feedback about the report (similar to the approach in Daniel et al., 2008). That feedback is not the real indicator since it represents only the user's opinion about his understanding of the report. The user might think that he does not understand the report correctly when in fact he does but is not completely confident, or, he might think that he understands the report completely when in reality he does not.

In brief, data quality issues, originating from the BI processes, indicate weak data quality management in the BI environment. They show that within BI processes there are problems with understanding the business requirements or with ensuring that business requirements are implemented correctly during each major BI stage. In order to deal with data quality issues it is important to assist all related users with understanding business requirements and their implementation during the BI processes.

Some data quality issues can be easily identified by having the necessary statistics automatically gathered from BI processes (e.g. overall ETL statistics) and some data quality issues cannot be identified easily even with gathering sufficient statistics (e.g. overall EDW statistics). Even if the statistics do not show the data quality issues explicitly, it is still critical to have those statistics in order to assist with general awareness of the BI environment and data quality status. Having all data quality information in one place would show where data quality issues exist and where there are no such issues and should make clear who is responsible for fixing data quality issues. It will also assure all involved stakeholders that data quality is constantly monitored and therefore they can be confident of the BI system. Consequently, the possibility of adopting BI in the organisation would be increased.

## DATA QUALITY MANAGEMENT AND ITS PRINCIPLES

Data quality issues need to be addressed through practicing quality management principles. By data quality management we mean establishment of tools, information and business processes that facilitate data quality management. Data quality management must have tools that deliver *integrated* data quality information in an *open* way to all participants involved in the BI processes. Data quality management should have *consistency* in business processes across different BI stages as well as in *delivery* and *integration* principles of data quality. Data quality management processes must have *consistent responsibilities* principles and embed strong *assurance* principles. By principles we mean high level requirements for the data quality management process. Each of these principles defines the requirements for the corresponding data quality tools, information or business processes.

### Integration

Comprehensive data quality management is possible only with having all data quality related information incorporated in one place. It allows engagement of different types of users in the same data area, to see others' progress and have confidence in data quality. Integrated data quality allows us to automate or semi-automate data quality processes, which can be helpful between BI stages and/or between different teams involved.

Integrated metadata describes elements of the BI environment at different BI stages and represents a foundation for data quality integration. It connects different elements of the BI environment and hence provides lineage and dependency capabilities for data in the BI environment. Data quality information is a part of the overall integrated BI metadata. In other words, data quality management requires integrated data quality information which in itself requires sound integrated BI metadata. All this demands tight integration with BI databases, which store data as they move though ETL stage, come into the EDW and appear on the end-user report.

### Responsibility

Acknowledging the fact that responsibility for data quality is dispersed across the organisation, there is a need for a common rule for sharing this responsibility. We propose a principle of consistency in responsibility for data quality across the organisation – responsibility rests with the unit which has oversight of the area that is a source (upstream to the BI process) of data quality issues.

### Openness/Transparency

Data quality results from any BI stage should be easily accessible by anyone in the organisation to provide confidence to all involved users and, especially, to business users, and to increase awareness of existing data quality sources and processes of managing data quality in the BI environment.

### Consistency

To make data quality more comprehensible by all types of users the data quality information should be based on a consistent data structure across the BI environment. There needs to be a consistent way of storing, and delivering data quality information and a clear way of analysing data quality information.

### Delivery (One stop shop)

In terms of delivery data quality information should be easily accessible from the BI environment. On opening a BI report the user wants to see the overall status of the report in terms of data quality and if required to get quality details about particular data elements presented in the report.

### Assurance

Assurance practices needs to be embedded into each data quality management process. The whole data quality management process should be based on best practice and independent quality assurance ideally performed by an external party. Quality assurance can be performed by an internal party depending on the cost-benefit of using the external party.

## METADATA FRAMEWORK TO SUPPORT DATA QUALITY MANAGEMENT

Earlier we defined metadata as a mechanism for describing elements of the BI environment and data. Having metadata infrastructure that gathers processes and furnishes relevant metadata information to all types of users allows them to efficiently manage and share the knowledge about the BI environment. It seems reasonable to have a broad BI metadata system that will, in addition to other metadata information, provide data quality information. Thus a data quality initiative would be a part of overall metadata strategy. Below we describe major metadata elements that represent a basis for data quality metadata in the BI environment.

As we mentioned earlier, data quality has many dimensions. These dimensions can be represented as the attributes for describing data quality. Examples of these dimensions include consistency, completeness, confidence, freshness, coherence and accuracy (Daniel et al., 2008; Ballou and Tayi, 1999; Vassiliadis et al., 1999; Foshay et al., 2007; Jarke et al., 1999).

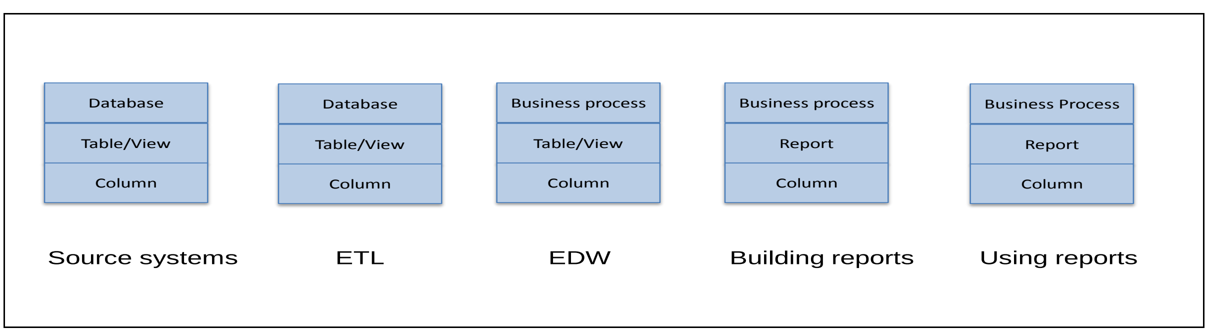


**Figure 3. Main metadata elements for data quality model**

Based on data quality issues that we identified earlier, we propose the following main metadata elements for the metadata data quality model. Each BI stage has three core metadata elements that should be used for designing a data quality model of the related BI stage. Some of the metadata elements appear to be the same in different BI stages. It means that the same metadata element may have different data quality attributes depending on the BI stage e.g. a table element in source systems would have different data quality attributes from a table element in the EDW.

If design of the metadata data quality model is based on a dimensional approach then each BI stage would represent a business process while main metadata elements would represent compulsory dimensions for each fact table that represents the corresponding business process (BI stage). Below is an example of how an ***ETL relationship diagram*** can be implemented using proposed core metadata elements (Figure 4). Each relationship within the ETL processes would have a number of source data elements and a number of destination data elements. All data elements would be uniquely identified by a value in the database, table and column dimensions. The relationship dimension will have attributes about a business rule (if applicable) and about transformation implemented along the way. Relationship fact table will allow the definition of many source and destination data elements for each particular relationship. The relationship Role dimension would store a role of the data element in the relationship, for instance, source data element or destination data element. The “Count” measure would have a constant value of ‘1’ in order to allow counting data elements involved in the relationships.

As well as describing particular BI stage, proposed metadata elements can be used to describe relationships between elements of the BI environment existing in different BI stages. For example, consider a connection between the EDW BI stage and the Building reports BI stage which is implemented by storing the relationship between Column element on the report and source Column elements from the EDW (Figure 5). Each Column element in the report is a transformation of one or many Column elements from the EDW. Information about transformation and corresponding business rule is stored in the Relationship dimension. This model provides a lot of insightful information e.g. a list of all EDW tables used for the report, a list of all reports dependent on a particular EDW table and how business processes in the reporting area are correlated with the business process in the EDW.

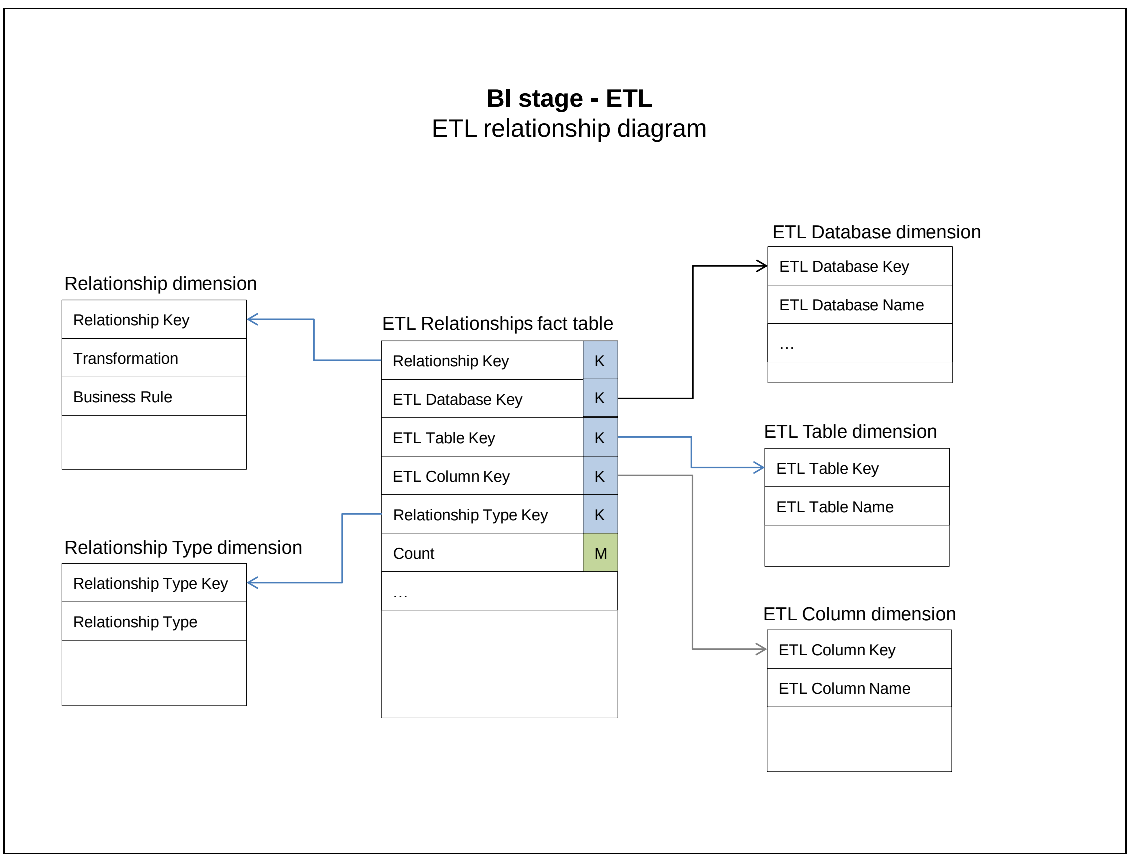


**Figure 4. Example of ETL Relationship Diagram**

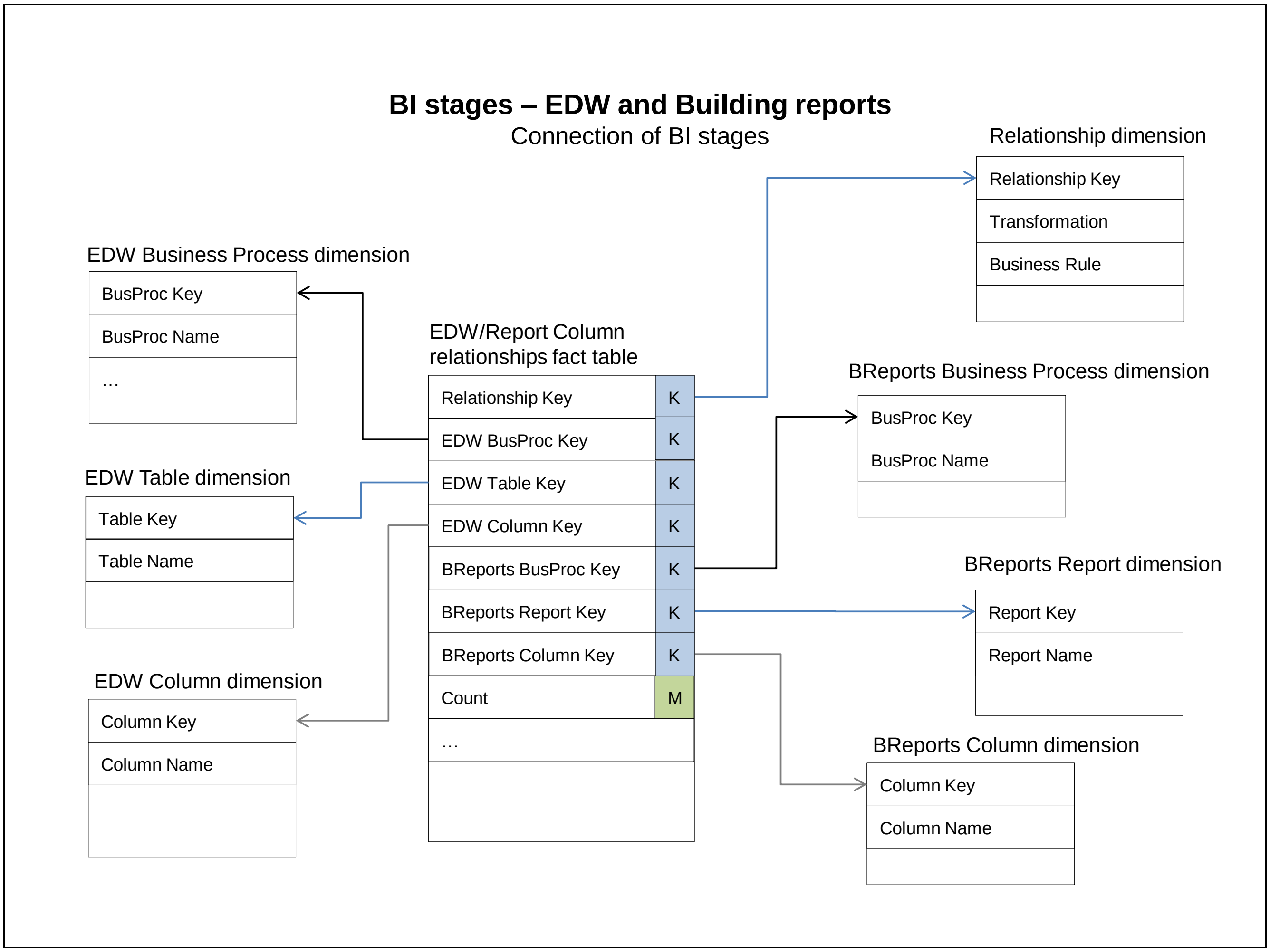


**Figure 5. Example of Connection between BI stages**

## SUMMARY

In summary, data quality represents a massive risk for the BI effort. BI processes have to incorporate steps for dealing with data quality issues originating from source systems and assure data quality within BI processes. Comprehensive data quality management is a necessity for BI success in the organisation. But data quality management is not only about monitoring and fixing data quality issues. It is also about educating all types of BI users about data quality, data and the BI environment.

In this paper we have discussed data quality as being a main indicator for meeting business requirements. We discussed major sources of data quality issues, presented particular examples and outlined the information required to deal with given data quality issues. Furthermore, we recommended key principles to assist with the development of a comprehensive data quality management process. In addition, use of a metadata framework has been proposed as a main technical solution for data quality management with key metadata elements. Accordingly, we presented examples of the proposed metadata elements for describing data quality information across the BI environment. Combined, these ideas and proposals represent a methodology for comprehensive management of data quality in the BI environment.

## REFERENCES


1. Ballou, D.P. and Tayi, G.K. (1999) Enhancing Data Quality in Data Warehouse Environments. *Communications of the ACM,* 42, 1, 73-78.
2. Daniel, F., Casati, F., Palpanas, T. and Chayka, O (2008) Managing Data Quality in Business Intelligence Applications. *Proceedings of the VLDB Endowment,* Auckland, New Zealand.
3. Farinha, J., Trigueiros, M.J. and Belo, O. (2009) Using inheritance in a metadata based approach to data quality assessment, *Proceedings of the first international workshop on Model driven service engineering and data quality and security*. ACM, Hong Kong, 1-8.
4. Foshay, N., Mukherjee, A. and Taylor, A. (2007) Does data warehouse end-user metadata add value? *Communications of the ACM,* 50, 11, 70-77.
5. Gao, J., Lin, S. and Koronios, A. (2006) Data Quality in Engineering Asset Management Organisations – Current Picture in Australia. *Proceedings of 1st World Congress on Engineering Asset Management*, Gold Coast.
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. Inmon, W., O'Neil, B. and Fryman, L. (2008) *Business Metadata - Capturing Enterprise Knowledg*e. Morgan Kaufmann, Burlington, VT.
8. Jarke, M., Jeusfeld, M.A., Quix, C., and Vassiliadis, P. (1999) Architecture and quality in data warehouses: An extended repository approach. *Information Systems*, 24, 3, 229-253.
9. Khatri, V. and Brown, C.V. (2010) Designing data governance. *Communications of the ACM*, 53, 1, 148-152.
10. Kimball, R., Ross, M, Thornthwaite, W., Mundy, J. and Becker, B. (2008) *The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems*, 2$^{nd}$ edition , John Wiley & Sons.
11. Li, Y. and Osei-Bryson, K.-M. (2010) Quality factory and quality notification service in data warehouse, *Proceedings of the 3rd workshop on information and knowledge management*. ACM, Toronto, 25-32.
12. Nichols, D.M., Chan, C., Bainbridge, D., et al. (2008) A lightweight metadata quality tool, *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*. Pittsburgh PA, 385-388.
13. Richardson, J., Schlegel, K., Hostmann, B., and McMurchy, N. (2008) Magic Quadrant for Business Intelligence Platforms, Gartner.
14. Rodic, J. and Baranovic, M. (2009) Generating data quality rules and integration into ETL process, *Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP*. ACM: Hong Kong, 65-72.
15. Tvrdíková, M. (2007) Support of Decision Making by Business Intelligence Tools. *Proceedings of the 6th International Conference on Computer Information Systems and Industrial Management Applications.*
16. Vassiliadis, P. (2000) Data Warehouse Modeling and Quality Issues, in Department of Electrical and Computer Engineering, National Technical University of Athens: Athens, Greece.
17. Vassiliadis, P., Bouzeghoub, M. and Quix, C (1999) Towards Quality-Oriented Data Warehouse Usage and Evolution, in *Advanced Information Systems Engineering*, M. Jarke and A. Oberweis (Eds), Springer Heidelberg. 164-179.
18. Wixom, B. and Watson, H. (2010). The BI-Based Organization, *International Journal of Business Intelligence Research*, 1, 1, 13-28.