Chapter 12
Effective Measurement of DQ/IQ for BI

Arun Thotapalli Sundararaman
India/Birla Institute of Technology and Science, India

ABSTRACT

DQ/IQ measurement in general and in the specific context of BI has always been a topic of high interest for researchers. The topic of Data Quality (DQ) in the field of Information Management has been well researched, published, and studied. Despite such research advances, there has been very little understanding either from a theoretical or from a practical perspective of DQ/IQ measurement for BI. Assessing the quality of data for a BI System has been one of the major challenges for researchers as well as practitioners, leading to the need for frameworks to measure DQ for BI. The objective of this chapter is to provide an overview of the existing frameworks for measurement of DQ for BI, analyze the gaps therein, review proposed solutions, and provide a direction for future research and practice in this area.

INTRODUCTION

There has been a growing interest among researchers and industry practitioners in DQ, DQ assessment and improvement. However, the basic problem of DQ assessment and improvement still remains largely open. For example, Massachusetts Institute of Technology (MIT) has instituted Total Data Quality Management (TDQM) research

effort which is grown from industry needs for high quality data. The objective of this program is to establish a solid theoretical foundation and devise practical methods for business and industry to improve DQ. The research scope of this program covers 3 areas i.e. 1) definition of DQ that addresses issues of data quality definition, measurement, and derivations; 2) Analysis of DQ Impact on Business that addresses the value chain relationship between DQ and business outcome

and 3) improvement of DQ that addresses various methods for improving DQ.

Thus research interest in DQ remains an enduring subject, which is reiterated even by very recent literature. For example, Fehrenbacher et al., 2012 state that the importance of DQ is ever increasing and that research in this field focuses mainly on *2 aspects i.e. criteria and assessment.* This recent work observes that while researchers have developed a number of frameworks, criteria lists and approaches for assessing and measuring DQ, still research in this discipline indicates that assessing DQ remains to be challenging. This work argues that although DQ is subjective, most of the existing frameworks and assessment methodologies do not often consider the context in which the assessment is performed. Through empirical data research this cited work suggests that the perceived importance of DQ criteria has changed over the last decade.

Heinrich et. al., 2009 observe that many DQ metrics are designed on an ad hoc basis and that they are often highly subjective and propose six normative requirements in order to enable a scientific approach and an evaluation of the metrics.

Direct references to DQ in BI Systems in published literature are limited. The same may be explored through references to related types of Information Systems such as DW, DSS, EIS or data mining. Similarly, the terms DQ and IQ have been used interchangeably in literature. There has been no consensus about the distinction between DQ and IQ – DQ may be referred to technical issues or quality in sources or 'raw data' and IQ may be referred to non-technical issues or quality in processed data stores or insights presented from data for decision making.

The objectives of this chapter are as follows:

- Identify gaps that exist in measuring DQ for BI based on published frameworks;
- Analyze existing DQ/IQ measurement frameworks in light of their applicability for BI;

- Examine DQ requirements (normative requirements) applicable to DQ for BI;
- Discuss at length a proposed framework and how it seeks to address the gaps;
- Consolidate findings and directions for future work in the study of measurement of DQ for BI.

In summary, past literature shows that poor quality of data in a warehouse adversely impacts the usability of the warehouse and managing DQ in a warehouse is very important (Shankaranarayanan, 2005). The needs of DQ definition and approach for DW and/or Decision Support Systems are unique and different. However, most of the research work focused on this specialized type of information system, have approached with metadata model (integrating / extending existing metadata in a warehouse with quality-related metadata, which has associated practical problems in implementation) and have not focused on implementation aspects of the framework.

BACKGROUND

There has been no consensus about the distinction between data quality and information quality. In this chapter, no distinction has been made and term data quality has been used to refer to the full range of issues.

Given the limited research advances in definition of DQ metrics or DQ measurement for BI, it will be valuable to focus on research based on DQ metrics requirements framework. As it applies to any other measurement, it is important to identify and summarize the requirements for DQ measurement. This view is supported in the works of Heinrich et. al., 2009. Such a focus is expected to guide the principles of DQ metrics definition and/or measurement – either for individual DQ dimensions or comprehensive (i.e. covering all DQ dimensions) DQ metrics such as overall DQ score (Sundararaman, 2012).

Related Content

Exploration of Soft Computing Approaches in Itemset Mining
Jyothi Pillaiand O. P. Vyas (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications (pp. 1830-1856).*
www.irma-international.org/chapter/exploration-of-soft-computing-approaches-in-itemset-mining/142705

Strategies for Improving the Efficacy of Fusion Question Answering Systems
José Antonio Robles-Flores, Gregory Schymik, Julie Smith-Davidand Robert St. Louis (2011). *International Journal of Business Intelligence Research (pp. 46-63).*
www.irma-international.org/article/strategies-improving-efficacy-fusion-question/51558

An Overview of the Data Science Process and Data Analytics Within Organisations
Mohammad Izzuddin Mohammed Jamil (2023). *Handbook of Research on AI and Knowledge Engineering for Real-Time Business Intelligence (pp. 88-104).*
www.irma-international.org/chapter/an-overview-of-the-data-science-process-and-data-analytics-within-organisations/321488

The Roles of Knowledge Management and Organizational Innovation in Global Business
Kijpokin Kasemsap (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications (pp. 1759-1783).*
www.irma-international.org/chapter/the-roles-of-knowledge-management-and-organizational-innovation-in-global-business/142700

Integration of Data Mining and Business Intelligence in Big Data Analytics: A Research Agenda on Scholarly Publications
Atik Kulakli (2021). *Integration Challenges for Analytics, Business Intelligence, and Data Mining (pp. 13-43).*
www.irma-international.org/chapter/integration-of-data-mining-and-business-intelligence-in-big-data-analytics/267863