Preface

Although Business Intelligence (BI) applications have been dominating the technology priority list of many organizations, these efforts often fail to accomplish their objectives. In many cases, the reason for the failure is directly attributable to the poor quality of the underlying information. Achieving and sustaining high-levels of information and data quality draws on knowledge and skills across a number of disciplines such as quality management, information technology, project management, and change management. At the same time, a typical business intelligence endeavor has to address a growing number of data and information quality conundrums including cross-functional information integration of data from disparate source, the emergence of “big data” with its increasing volume, velocity, and the lack of effective information quality and governance strategies and policies.

Information (IQ) is a body of knowledge and practice around realizing the maximum value from an organization’s information assets by assuring that the information products produced from that information will meet the expectations of (“are fit for use by”) the people and system using them. Of these information products, perhaps none are more critical than business intelligence applications. BI application must deliver the right information, in the right form, to the right people at the right time for more informed decision-making. The quality of BI output is directly related to the quality of the input data. Successful information quality and governance programs are fundamental to the success of any business intelligence initiative, especially those at the enterprise level. Unfortunately, the information quality and governance capabilities in many organizations are not up to the levels needed to support their BI applications. Therefore, the mission of this book is to advance research and practice in the field of information quality and business intelligence. The book is a peer-reviewed publication dedicated to the exchange of the latest research on all aspects of practicing and managing information quality and governance for business intelligence in organizations. This book contains 21 chapters and is organized as follows.

The 1st chapter by Foshay, Taylor, and Mukherjee, “A Conceptual Model of Metadata’s Role in BI Success,” describes how organizations can increase the levels of use of their BI systems by providing the right metadata to users. A conceptual model is proposed to describe how metadata contributes the level of BI system use by creating positive attitudes toward the information available.

In “Understanding the Influence of Business Intelligence Systems on Information Quality: The Importance of Business Knowledge,” Popovic and Jaklic show how Business Intelligence System (BIS) maturity affects IQ dimensions, as well as the role that business knowledge may exert in mobilizing this link. The hypotheses are tested across 181 medium and large organizations.

The 3rd chapter, “Subjective Information Quality in Data Integration: Evaluation and Principles,” authored by AbuHalimeh and Tudoreanu, focuses on the science of human perception of information quality, and describes a subset of information quality dimensions, which are termed Subjective Information Quality (SIQ). The chapter explores SIQ while considering information obtained from multiple sources.
Authored by Decker, Liu, Talburt, Wang, and Wu, the 4th chapter, “A Case Study on Data Quality, Privacy, and Entity Resolution,” presents ongoing research conducted through collaboration between the University of Arkansas at Little Rock (UALR) and the Arkansas Department of Education to develop an entity resolution and identity management system for a business intelligence system. The research is the first known of its kind to evaluate privacy-enhancing entity resolution rule sets in a state education agency.

Chapter 5 is titled “Business Intelligence for Healthcare: A Prescription for Better Managing Costs and Medical Outcomes.” Cook and Neely use an interpretive case study approach to describe data quality problems in the healthcare supply chain and their integration with a business intelligence system in two companies.

Authored by Schmidt and Prado, the 6th chapter, “IT Architecture and Information Quality in Business Intelligence Environments,” presents a research model to identify the relations between components of data warehouse/business intelligence architecture and information quality dimensions. The model has been applied in a case study.

Chapter 7, “Information Quality Assessment for Asset Management Systems,” authored by Lee and Haider, presents the research which employs productive perspective to information, provides a method for assessing correlation of information quality dimensions, and applies Six-Sigma methodology to assess its quality in information systems utilized to manage engineering asset lifecycle and control quality of information.

In “Trends and Research of Wikis’ Quality and Governance Based on Bibliometric and Content Analyses,” Zhu, Huang, Liu, and Du explore the current research status and trends of wikis’ quality and governance by using bibliometric analysis and content analysis.

The 9th chapter is titled “Social Media Tools for Quality Business Information.” Maravilhas presents the work of how to extract quality information, unbiased, valuable for business, from social media tools, and sharing with the interested parties some ways of using it for their profit and competitive sustainability.

Chapter 10, “Improving Spatial Data Quality through Spatial ETL Processes,” authored by Molinowski and Campos, studies different problems that may exist in handling spatial data and shows several examples of how these problems can be detected and solved using spatial ETL tools.

Chapter 11 is titled “Principled Reference Data Management for Business Intelligence.” Milman, Oberhofer, Pandit, and Zhou discuss a principled approach towards management, stewardship, and governance of reference data to ensure quality and operational excellence across BI systems in the domain of typical reference data management concepts and features, leading to a comprehensive solution architecture.

Authored by Sundararaman, the 12th chapter, “Effective Measurement of DQ/IQ for BI,” provides an overview of the existing frameworks for measurement of DQ for BI, analyzes the gaps therein, and reviews proposed solutions and provide a direction for future research and practice in this area.

The 13th chapter, “Data Profiling and Data Quality Metric Measurement as a Proactive Input into the Operation of Business Intelligence Systems,” authored by Delaney proposes the use of data profiling techniques as a means of early discovery of issues and changes within the source system data and examines how this knowledge can be applied to guard against reductions in the decision-making capability and effectiveness of an organization.

Chapter 14 is titled “Agile Information Management Governance: Can You Scale it to the Enterprise?” Bates researches the possibility of scaling success of small projects to enterprise level projects by delivering projects in small, incremental chunks while maintaining business involvement and support, and setting clear, achievable objectives throughout the program.
In “Challenges of Structure and Organization in Medium-Sized Content,” Chakrabarty presents an approach to create metadata for content pieces by proposing a three structure metadata consisting of a static part, a dynamic part, and a conceptual part that links to other pieces of related information.

Chapter 16, “A Case Study to Improve Data Vendor Selection,” authored by McGraw presents a project of improving the accuracy of multi-source income data used in credit card marketing applications. The project evaluates the accuracy of third-party data providers by measuring the data across six dimensions.

The 17th chapter by Zhou and Talburt, “Strategies for Large-Scale Entity Resolution Based on Inverted Index Data Partitioning,” presents a data partitioning strategy to perform entity resolution on large datasets and discusses the importance of index-to-rule alignment, pre-resolution index closure, post-resolution link closure, etc. in a distributed processing environment.

The 18th chapter, “A Dual-Database Trusted Broker System for Resolving, Protecting, and Utilizing Multi-Sourced Data,” authored by Gibson and Holland, presents a system to allow for the sharing of data between government agencies while upholding the strictest interpretations of rules and regulations protecting individual privacy and confidentiality.

Chapter 19 is titled “Business Intelligence Architecture in Support of Data Quality.” Breur suggests an innovation to data warehouse architecture to improve data integration in BI projects. Two suggestions are made to deliver information products in the smallest possible chunks and BI teams need to provide better stewardship when they facilitate discussions between departments.

The 20th chapter is titled “The Value of Data Quality.” Hillard introduces a trading approach to information that changes the mindset of the managers of information who become motivated to maximise the value of the content they are responsible for. The approach also creates the opportunity to move the focus of data governance away from what is often perceived to be compliance activities to helping to establish pricing and the market as a whole.

The last chapter in this book deals with “Data Protection and BI: A Quality Perspective.” The author, O’Brien, examines the relationship between Data Protection (DP), Information Quality (IQ), and Data Governance (DG). This chapter provides an overview of how techniques and practices from IQ and DG can ensure that BI projects are grounded on appropriate privacy controls.

We hope that this book will stimulate further studies in information quality and governance for business intelligence.

William Yeoh
Deakin University, Australia

John Talburt
University of Arkansas – Little Rock, USA

Yinle Zhou
IBM Corporation, USA