Preface

The impact of data quality on the information chain has been widely recognized since the onset of large-scale data processing. Furthermore, recent years have seen a remarkable change in the nature and usage of data itself due to the sheer volume of data, high accessibility leading to unprecedented distribution and sharing of data, and lack of match between the intention of data creation and its subsequent usage, to name a few. The importance of the understanding and management of data quality for individuals, groups, organizations, and government has thus increased multifold.

The data (and information) quality domain is supported by several decades of high quality research contributions and commercial innovations. Research and practice in data and information quality is characterized by methodological as well as topical diversities. The cross-disciplinary nature of data quality problems as well as a strong focus on solutions based on the fitness for use principle has further diversified the related body of knowledge. Although research pluralism is highly warranted, there is evidence that substantial developments in the past have been isolationist. As data quality increases in importance and complexity, there is a need to motivate exploitation of synergies across diverse research communities.

The above factors warrant a multipronged approach to the study of data quality management spanning: organizational aspects, i.e. strategies to establish people, processes, policies, and standards required to manage data quality objectives; architectural aspects, i.e. the technology landscape required to deploy developed processes, standards, and policies; and computational aspects which relate to effective and efficient tools and techniques for data quality.

Despite a significant body of knowledge on data quality, the community is lacking a resource that provides a consolidated coverage of data quality over the three different aspects. This gap motivated me to assemble a point of reference that reflects the full scope of data quality research and practice.

In the first chapter of this handbook, I provide a detailed analysis of the data quality body of knowledge and present the rationale and approach for this handbook, particularly highlighting the need for cross-fertilization within and across research and practitioner communities. This handbook is then accordingly structured into three parts representing contributions on organizational, architectural,

and computational aspects. There is also a fourth part, devoted to case studies of successful data quality initiatives that highlight the various aspects of data quality in action. This handbook concludes with a chapter that outlines the emerging data quality profession, which is particularly important in light of new developments such as big data, advanced analytics, and data science.

The preparation of this handbook was undertaken in three steps. Firstly, a number of global thought leaders in the area of data quality research and practice were approached to join the initiative as part of the advisory panel. The panel members contributed significantly to the refinement of this handbook structure and identification of suitable chapter authors and also supported the review process that followed chapter submissions. The identified chapter authors were then invited to provide contributions on the relevant topics. Finally, all chapter contributions were reviewed by at least two experts. To ensure that the quality of the final chapters was not compromised in any way, some contributions were rejected or substantially revised over two or three review cycles. However, I am most grateful for the time devoted by all authors to produce high quality contributions and especially for the responsiveness of the authors towards making the required changes.

I would like to take this opportunity to thank all the authors and advisors for their valuable contributions. A special thanks to Xiaofang Zhou, Divesh Srivastava, Felix Naumann, and Carlo Batini for their guidance and inspiration in the preparation of this handbook. Thanks to all the expert reviewers of the chapters, with a special thanks to Mohamed Sharaf for constant encouragement and advice. Last but not least, a big thanks to Kathleen Willamson, Yang Yang and Vimukthi Jayawardene for an enormous help in the editing and formatting work required for the preparation of this handbook.

I hope that this Handbook of Data Quality will provide an appreciation of the full scope and diversity of the data quality body of knowledge and will continue to serve as a point of reference for students, researchers, practitioners, and professionals in this exciting and evolving area.

Brisbane, Australia   Shazia Sadiq
April 2013