Special Section:
Assuring Information Quality

DONALD BALLOU, STUART MADNICK, AND RICHARD WANG,
Guest Editors

DONALD BALLOU is an Emeritus member of the faculty of the School of Business at the University at Albany. His graduate degrees are in applied mathematics, and his research has involved modeling in a variety of fields, including nonlinear fluid flow, production quality control, regional and transportation planning, information resource management, and especially data and information quality. At the 1996 Conference on Information Quality held at MIT, Harry Pazer and Dr. Ballou received the "Outstanding Research Contribution" award for their pioneering contributions to the field of information quality. Dr. Ballou's research in the area of information systems has appeared in various journals, including *Communications of the ACM, Information Systems Research, MIS Quarterly,* and *Management Science.*

STUART MADNICK has been a faculty member at MIT since 1972. He is currently John Norris Maguire Professor of Information Technology at the Sloan School of Management and Professor of Engineering Systems at the School of Engineering. He has served as the head of MIT's Information Technologies Group for more than 20 years. During that time, the group has been consistently rated number 1 in the nation among business school information technology programs (e.g., by *Business Week*). Professor Madnick has also been an affiliate member of MIT's Laboratory for Computer Science, a member of the research advisory committee of the International Financial Services Research Center, and a member of the executive committee of the Center for Information Systems Research. He has been active in industry, making significant contributions as a key designer and developer of projects such as IBM's VM/370 operating system and Lockheed's DIALOG information retrieval system. He has served as a consultant to many major corporations, such as IBM, AT&T, and Citicorp.

RICHARD WANG is Director, MIT Information Quality Program and Co-Director for MIT Total Data Quality Management Program. Dr. Wang has put the term *information quality* on the intellectual map with myriad publications and conferences. He has served as a professor at MIT, Boston University, and the University of Arizona, Tucson. In 1996, Professor Wang cofounded the premier International Conference on Information Quality, which he has served as the general conference chair, and is currently Chairman of the Board. He also cofounded the Workshop on Information Technologies and Systems (WITS) in 1991. His writings have appeared in leading journals, such as the *Communications of the ACM, Management Science, Journal of Management Information Systems, Information & Management,* and *IEEE Transactions.* Wang coauthored *Information Technology in Action: Trends and Perspectives* (Prentice Hall, 1993), *Data Quality Systems* (CMI, 1995), *Quality Information and Knowledge* (Prentice Hall, 1999), *Data Quality* (Kluwer Academic Publishers, 2001), and *Journey to Data Quality* (MIT Press, forthcoming).

ASSURING THE QUALITY OF INFORMATION is both important and difficult. But achieving high-quality information is a battle that is never really won, in part because what constitutes victory is not clear, as different parties have differing views as to the definition of success. Yet all concerned agree that striving to achieve or acquire high-quality information needs to be a high priority, as the consequences of not having it can be devastating. The very existence of the organization can be threatened by poor information quality (IQ).

The issue of IQ is not new, as, throughout history, people have benefited by, and suffered from, the quality of information made available to them. An obvious example is information made available to military commanders during battles. What is new in the past several decades is the explosion in the quantity of information and the increasing reliance of most segments of society on that information. An information economy is clearly dependent upon the quality of one of its primary building blocks.

Of necessity, industry has, from the beginning of the computer age, had to address the issue of the quality of data. When systems were primarily accounting- and financial-driven, the emphasis tended to be on the assurance of the data's accuracy. In the past dozen or so years, the increasing use of information as a strategic resource has highlighted the multifaceted nature of IQ and has increased the complexity in attempting to assure IQ. The very concept of IQ is somewhat nebulous, but an effective, widely used definition for IQ is "fitness for use." "Perfect" IQ, whatever that means, is difficult, if not impossible, to achieve, but neither is it necessary. If users of the data feel that its quality, which can be described by such attributes as accuracy, completeness, timeliness, and consistency, is sufficient for their needs, then, from their perspective, at least, the quality of the information available to them is fine. Most research, whether conceptual or field-oriented, has used as its starting point this concept that the focus of assuring IQ should be to achieve a level of quality that is sufficient from the perspective of its users.

IQ research, in the information systems world at least, by its very nature is an applied discipline. Because IQ is such an important issue to practitioners, research regarding it should be grounded in the need, broadly defined, of organizations to assure IQ and in the difficulties encountered in attempting to do so. This does not mean that IQ researchers do not engage in work of a more fundamental, conceptual nature. One IQ research stream has been the study of exactly what IQ is. Work in this area has been both conceptual and field-oriented, with one reinforcing the other, and, as a result, there is now general agreement in both academia and industry as to the context for discussing IQ. As in most research areas, both conceptual and field studies are critical. Conceptual works provide insights that guide field studies, whereas the latter provide ideas for research topics and constitute validation for conceptual frameworks.

This special section contains four papers that illustrate some of the issues of concern to IQ researchers as well as various approaches and methodologies for addressing them. All have conceptual components that are tested and explored in actual organizational settings. Leading off this issue is a paper by Yang W. Lee and Diane M. Strong entitled "Knowing-Why About Data Processes and Data Quality." This

paper analyzes the role of various parties in the data production process and offers insights into how different modes of knowledge affect data quality. For example, the authors find that data collectors with why-knowledge about the data production process contribute to a greater degree than might be anticipated to the production of better quality data. Such insights are gleaned from an extensive field study, which in turn was motivated and guided by a synthesis of various research streams.

The second paper, "The Design and Implementation of a Corporate Householding Knowledge Processor to Improve Data Quality," coauthored by Stuart Madnick, Richard Wang, and Xiang Xian, addresses an important category of data quality problems caused by data misinterpretation. For example, even a simple question, such as "How much did MIT buy from IBM this year?" has multiple answers that might be right or wrong—depending on the context. This paper outlines a technical approach to a corporate householding knowledge processor (CHKP) to solve a particularly important type of corporate householding problem, entity aggregation, and illustrates the operation of the CHKP by using a motivational example in account consolidation. The CHKP design and implementation uses and expands on the context interchange technology, developed at MIT, to manage and process corporate householding knowledge.

The next paper, "Time-Related Factors of Data Quality in Multichannel Information Systems," by Cinzia Cappiello, Chiara Francalanci, and Barbara Pernici is an excellent example of the use of modeling in the area of IQ. The authors develop expressions for the currency and completeness of data in the context of multiple channels, each supporting multiple functions within an organization. Their model allows for varying levels of integration of the organization's databases, and it can be used to study the impact of various refresh periods for online data. A validation is performed via a simulation in the context of empirical data regarding various categories of financial institutions.

The fourth paper, by Yang W. Lee, "Crafting Rules: Context-Reflective Data Quality Problem Solving," offers refreshing insights into an emerging practice, data quality problem solving and provides a novel way of understanding how practitioners reflect-in-action. It analyzes how data problems are framed, analyzed, and resolved throughout the entire information discourse. It also explains how rules on data quality practice revise the actionable dominant logic embedded in organizational work routines. For example, the author finds that the discourse on contexts of data connects otherwise separately managed data processes, that is, collection, storage, and use. These insights are gleaned from an extensive five-year longitudinal study of data quality practice in five organizations.