

## Editors' Comments

*ACM Journal of Data and Information Quality (JDIQ)* is alive and well!

---

We would like to report that the *ACM Journal of Data and Information Quality (JDIQ)* is alive and well! This fourth issue marks the end of our first full year of publication. At this time, we would like to summarize the status of the *JDIQ* and our plans for its future.

### 1. SUMMARY

*JDIQ* launched its inaugural issue in June 2009. Our vision is for *JDIQ* to be the premier journal and forum for the emerging field of data and information quality, attracting quality articles and readership.

### 2. EDITORIAL BOARD STRUCTURE: SENIOR ADVISORS, ASSOCIATE EDITORS, REVIEWERS

We have looked to the data quality, computer science, and MIS communities to recruit associate editors and reviewers to build the core of the review force. Additionally, we have reached out to a list of scholars who have either direct experience as EICs or experience as senior advisors. Finally, we have an in-house administrator, Jiun Hsu, for coordination and communication with the ACM headquarters and the reviewers.

We have 10 senior advisors, 24 associate editors, and over 200 reviewers so far. This year, we plan to include more associate editors, while we expect that the number of reviewers will naturally increase as more manuscripts are received.

### 3. MANUSCRIPTS REVIEW: ACCEPTANCE RATE AND PROCESS

The overall current acceptance rate of the *JDIQ* for papers that have reached a final decision (accept or reject) is approximately 22%. Each manuscript goes through a double-blind review process. Neither reviewers nor authors are known to the other party, and typically at least three reviewers are involved in reviewing each manuscript.

### 4. ISSUES OF JDIQ REVIEW CYCLE AND NOVEL "FREE FIRST ISSUE"

In reflection on the time between approval of *JDIQ* and its first publication we encountered some unexpected issues.

*4.1. Multiple Review Cycles.* In addition to the expected ramp-up process of advertising the new journal and getting submissions, we found the reviewers and AEs to be much more thorough than we had expected (which is actually

good news!). For example, among the 16 articles accepted for the first four issues, none were accepted based on its initial submission; they all required minor or major revisions. The breakdown was the following.

- Accepted after first revision: 1 manuscript
- Accepted after second revision: 5 manuscripts
- Accepted after third revision: 8 manuscripts
- Accepted after fourth revision: 1 manuscript
- Accepted after fifth revision: 1 manuscript

We have found that our reviewers and AEs are extremely motivated and hard working and take great pride in their efforts to make sure that quality is not just in the name of the *Journal on Data and Information Quality* but also in the articles that it published.

4.2. *Novel “Free First Issue”*. We are very pleased that the ACM Publications Board approved our request that the inaugural issue of *JDIQ* be made available for free from the ACM Digital Library. We think this has many advantages. (1) It provides a sample that prospective subscribers can review to help them make the decision to subscribe (2) it acts as editorial guidance to help prospective authors decide whether to submit their articles to *JDIQ*. The Table of Contents for the inaugural issue with access to all the articles is at <http://portal.acm.org/toc.cfm?id=1515693>. We encourage you to spread the word to your colleagues of this “deal of a lifetime.”

## 5. PLANS FOR THE FUTURE

5.1 *Review of Associate Editors*. Although we have been very pleased with the hard work and great efforts of our team of Associate Editors, we are now in a position to review our coverage and determine where there are opportunities to increase our depth in specific areas.

5.2 *Increase in Special Issues*. We have been very pleased with the quality of our first special issue. We were somewhat cautious and did not want to initiate too many special issues before *JDIQ* was firmly established. We have accumulated recommendations for several more special issues and are already well underway with completing the reviews for the articles in the forthcoming “Special Issue on Information Quality: The Challenges and Opportunities in Healthcare Systems and Services.” We also have the first Call for Papers in the area of Entity Resolution at the end of this issue. We welcome suggestions for other special issues.

5.3 *Increasing Readership and Prospective Authors*. To broaden the readership and attract prospective authors, we have plans in two areas. First, we will invite more special issue guest editors to address contemporary issues, such as healthcare, thus broadening the scope of topics covered by *JDIQ*, while bringing in expert scholars and reviewers with focused, yet diverse perspectives. Second, we will open the journal to quality articles from relevant conferences and workshops so that authors can submit their extended papers for a *JDIQ* review. We have had very fruitful discussions with the program committees of

various conferences and workshops, such as the International Conference on Information Quality (ICIQ), workshops associated with Very Large Databases (VLDB), Americas Conference on Information Systems (AMCIS), International Conference on Information Systems (ICIS), and other workshops and conferences that the Association of Information Systems (AIS) sigIQ supports.

## 6. IN THIS ISSUE

The four articles presented in this issue deal with different types of poor-quality data in various areas, including *missing* data (Article 2), *deceptive* data (Article 3), data flow *errors* (Article 4), and *uncertainty* in data integration (Article 5).

The articles do not claim to cover the specific deficiencies in data in a comprehensive fashion. Nevertheless, the four articles cover important aspects and topics of making the selective data deficiency visible, identifiable, and traceable. The insights from the four articles can be adapted to provide extended next steps for deepening and advancing theories by researchers and crafting useful ways for managing data and information quality by practitioners.

The first article is “Using Data Mining Techniques to Discover Bias Patterns in Missing Data” by Monica Chiarini Trembly, Kaushik Datta, and Debra VanderMeer. Lack of understanding about missing data adds unpredictable factors to the analyses results. Characterizing the missing data as exhibiting potential bias to the pool of data to be analyzed, the authors apply the association rule-mining techniques to handling missing data and propose a method for understanding systematic biases in missing data. The authors also demonstrate the use of the proposed method to characterize the patterns of the biases which can then help data consumers, such as analysts, identify and make informed decisions equipped with the understanding of the quality of the biased dataset. The method can be extended to find ways to aid decision makers in reaching informed decisions based on missing or incomplete datasets. As the authors suggest, this work can be tested with a larger dataset with more attributes and by observing domain area experts to further extend the method for application in real-world contexts.

This article was revised three times since the submission of the original manuscript.

The second article by Matthew L. Jensen, Judee K. Burgoon, and Jay F. Nunamaker, is titled “Judging the Credibility of Information Gathered from Face-to-face Interactions.” With the exception of data and information security and ethics, most data quality work has been based on the assumption that poor quality data has occurred unintentionally. Deception by information suppliers has not been studied systematically thus far, to the best of the authors’ knowledge. They provide a prototype system that examines messages gathered during direct, face-to-face information gathering. The system unobtrusively identifies kinetic and linguistic features that may indicate deception in the messages of information suppliers. The improved capability to assess credibility of information gathered face-to-face can be critical in various security-related application areas. Observers or analysts as data consumers

can decipher and assess to what extent the collected information is distorted through deception by information suppliers.

This article was revised three times since the submission of the original manuscript. The earlier shorter version of the manuscript received the best paper award at *AMCIS 2008*.

Hema S. Meda, Anup Kumar Sen, and Amitava Baqchi author the 3rd article, “On Detecting Dataflow Errors in Workflows.” Conventional research on automating workflow tends to focus on controlling the structure of information flow. The quality of data that goes through the workflow has not been the focus. Sundari et al. fill this gap and provide a new algorithm for detecting dataflow errors. Specifically, using the concept of corresponding pairs, the authors propose a graph traversal algorithm for detecting dataflow errors. Dataflow quality analysis can yield insights that can be directly applied to many critical applications such as processing large volume of data for automated workflow. Extended real-world demonstrations can suggest how this algorithm can be readily applied in different environments and contexts.

The article went through two revisions since the submission of the original manuscript.

The 4th article in this issue is “A Survey on Uncertainty Management in Data Integration” by Matteo Manani and Danilo Montesi. As we encounter increasingly more diverse sources of data from different locations, systems, processes, and disciplinary backgrounds, integrating data from those sources has become more complex, in part because the integrators do not know about all contexts of the data they intend to integrate. Motivated by the limited knowledge of the context of the data, Manani and Montesi take this often unavoidable limited knowledge as *uncertainty* about data from an integrator’s perspective. The authors lay out the current status of research on dealing with the uncertainty about data for data integration work. They identify the gap in research yet to be conducted to call for future research. Aspects of dealing with uncertain data are identified in the areas of (1) wrapping uncertain data, (2) matching data sources with uncertainty, (3) aggregation of multimatcher outcomes, (4) merging data sources based on uncertain mappings, and (5) querying uncertain mediated schemata. Subsequently, the authors identify nine key areas for future research to fill the gap, ranging from generation of uncertain measures to user involvement.

We hope that the authors and other researchers extend this research and fill some of the gaps identified. *ACM JDIQ* would particularly welcome interdisciplinary work that exploits using multiple research methods, both quantitative and qualitative, and grounds research questions in interdisciplinary theoretical basis to produce rigorous and relevant findings for dealing with uncertainty in data for various integration purposes for data consumers.

This article went through two revisions since the submission of the original manuscript.

We appreciate the work of the associate editors and the reviewers of the four articles in providing helpful recommendations throughout the multiple cycles of review. We also appreciate the authors for their fine contribution to the readers of *ACM JDIQ*.

We continue to welcome new and innovative research articles. While there is often value in addressing the same or similar questions found in the existing literature, we look forward to publishing articles that ask new and exciting questions. We also welcome research that applies theories from diverse fields to data and information quality issues, and we do not discriminate against any particular research method.

Enjoy the read, and we look forward to reviewing and providing future articles with breakthrough ideas and useful insights.

Stuart E. Madnick  
Yang W. Lee  
*Editors-in-Chief*  
<http://jdiq.acm.org>