

# Media and Personal Involvement in the Perceptions of Data Quality

Andrew Borchers

Kettering University, USA

## INTRODUCTION

This article introduces the concepts of data quality as described in the literature and discusses research results on how individual perceptions of data quality are influenced by media (World Wide Web versus print) and personal involvement with the topic. A search of literature on “data quality” and “media credibility” reveals that researchers in both the information systems and journalism fields have explored this topic. While these threads have developed separately, these streams of research approach similar issues of how people view the quality of information they receive from different sources.

This topic is important in that the Internet revolution has fundamentally changed how people share information. But with such access comes a challenge as stated by Gilster (as cited in Flanagin & Metzger, 2000): “When is a globe spanning information network dangerous? When people make too many assumptions about what they find on it. For while the Internet offers myriad opportunities for learning, an unconsidered view of its contents can be misleading and deceptive.”

The importance of this topic is underscored by recent research and seminars on the topic of data quality on the Web. A 2003 conference in Wadern, Germany, for example, brought researchers together to discuss topics such as “criteria and measurement for quality of Web data, representation and exchange of quality information and usage and maintenance of data quality in Web querying and data integration”. The conference included working groups that focused on quality assessment, trust, data integration, and metadata (<http://www.dagstuhl.de/03362/>).

## BACKGROUND

The importance of data quality has been echoed among information systems and journalism practitioners for many years. Research by Redman (1998) summarizes the practical implications of poor data quality. He points out the consequences of poor data quality in areas such as decision-making, organizational trust, strategic planning and implementation, and customer satisfaction. Redman

conducted detailed studies and found increased cost of 8-12% due to poor data quality. Service organizations can find increased expenses of 40-60% (Redman, 1998). Strong, Lee, Yang, and Wang (1997) support the seriousness of this issue in their study of 42 data quality projects in three organizations. Research by other authors note data quality issues in a number of settings including accounting (Xu, 2000; Kaplan, Krishnan, Padman, & Peters, 1998), airlines, health care (Strong et al., 1997), criminal justice (Laudon, 1986) and data warehousing (Ballou, 1999).

As for a formal definition of data quality, Umar, Karabatis, Ness, Horowitz, and Elmagardmid (1999) quote Redman (1998):

*A product, service, or datum X is of higher quality than product, service, or datum Y if X meets customer needs better than Y.*

Umar et al. (1999) go on to point out that this definition has been generally accepted and is consistent with author's work. The definition is somewhat incomplete, however, as it does not delve into the various dimensions of data quality.

A number of authors in the information systems field have gone further than Redman and written conceptual articles on “data quality” (Wand & Wang, 1996; Wang, Reddy, & Kon, 1995; Wang & Strong, 1996; Strong et al., 1997). This work suggests that data quality is a multidimensional concept (Wand et al., 1996) that can be viewed from a number of different perspectives. A panel discussion in 2000 (Lee, Bowen, Funk, Jarke, Madnick and Wand) found five different perspectives to discuss data quality. These included an ontological perspective (specification of a conceptualization) that include different views of reality based on actual observation versus computer influenced observations, architectural perspective (a view that focuses on system infrastructure and its influence on data quality), context mediation perspective (focusing on communication across space and time), time-based e-commerce perspective (focusing on the real-time nature of e-commerce) and an information product perspective (focused on data as a product of an organization).

In talking about “data quality”, a key beginning is to determine from the literature just what is meant by the

term. In a definitive work on the topic, Wang et al. (1996) provides a conceptual framework for data quality. In a way consistent with Redman's customer perspective, they start by defining "high-quality data as data that is fit for use by data consumers". Using a two-stage survey and sorting process, Wang develops a hierarchical framework for data quality that includes four major areas: intrinsic, contextual, representational, and accessibility.

Intrinsic data quality refers to the concept that "data have quality in their own right" (Wang et al., 1996). Intrinsic dimensions include accuracy, objectivity, believability, and reputation. Contextual data quality is based on the idea that data does not exist in a vacuum – it is driven by context. Contextual dimensions include relevancy, timeliness and appropriate amount of data. Representational data quality relates to the "format of the data (concise and consistent representation) and meaning of data (interpretability and ease of understanding)". Accessibility refers to the ease with which one can get to data (Wang et al., 1996).

Beyond the information systems literature, there is a second relevant body of literature that comes from the journalism field. Their focus is on perceptions of Internet credibility (Flanagin & Metzger, 2000; Johnson & Kaye, 1998). The major thrust of this literature is in comparing the Internet to traditional sources with respect to credibility. Note that when referring to "credibility", these authors say "the most consistent dimension of media credibility is believability, but accuracy, trustworthiness, bias and completeness of information are other dimensions commonly used by researchers." (Flanagin et al., 2000, p. 521). Hence, there is a rough correspondence of thinking about "credibility" in the journalism literature to the concept of "intrinsic" and "contextual" data quality in the information systems literature.

## **COMPARISON OF RESEARCH FINDINGS**

Research on data quality and media differences has been undertaken by a number of scholars over the years. In this section, the author will focus on work by Klein (1999, 2001), Flanagin et al. (2000) and Borchers (2003). All of these authors have examined data quality in a similar way, focusing on perceived differences based on media (print versus Internet). In addition, Flanagin et al. (2000) examine whether Internet users verify what they find. Borchers (2003) extends the discussion by examining the effect of personal involvement in the topic.

Klein (1999, 2001) has studied perceptions of data quality by surveying a sample of approximately 70 graduate business students conducting class projects. In one early study, Klein (1999) found Web-based material to be

more timely, but less believable and of lower reputation, accuracy and objectivity than printed material. In a more formal result, Klein (2001) found traditional text sources to be perceived as more accurate, objective and to have higher reputation and representational consistency. Internet sources were found to be stronger in timeliness and appropriate amount.

Flanagin et al.'s work (2000) focuses in three areas. First, they look at the perceived credibility of television, newspapers, radio and magazines compared to the Internet. The major finding, unlike Klein, is that there is little difference in credibility between media. Second, Flanagin et al. look at the extent to which Internet users verify what they receive. Here, they find that few Web users verify the information they receive. Those with limited Internet experience verify less than those with more experience. Third, and most important to this discussion, Flanagin et al. look at whether perceived credibility varies depending on the type of information being sought. Flanagin et al. cite Gunther in suggesting that "greater involvement with the message results in, first, a wider latitude of rejection."

Borchers (2003) considered the literature cited previously and examined a number of interesting questions. In keeping with Klein and Flanagin et al., he examined how people perceive Web-based material compared to printed material considering dimensions such as "timely", "believable", "reputation", "accuracy", or "objectivity". Second, Borchers studied whether individuals with personal involvement in a topic (e.g., cancer) are better discriminators of data quality than those who are not involved with a topic. Finally, given their role as health care acquirers (Bates & Gawande, 2000; Looker & Stichler, 2001), Borchers explores whether women are better discriminators of data quality than men on health related topics such as cancer.

Figure 1 demonstrates what Borchers hope to find. H0, his initial hypothesis, is that the perception of low creditable sources is significantly less than high creditable sources. Hence, the two lines for Internet-based and print-based text should have a positive slope. H1 suggests a significant gap between the lines for Internet-based sources and text-based sources on the timeliness, believability, reputation, accuracy, and objectivity dimensions. This assertion was based on prior literature by Klein (1999). H2 suggests that the slope of the lines should vary based on one's personal involvement in cancer. This is to say, that persons with high personal involvement in cancer should be better discriminators of data quality. Finally, H3 suggests that women are better able to differentiate credible from non-credible sources. Hence, the slope of the lines should vary based on gender.

Borchers studied 127 subjects on their perception of information on cancer based on exposure to Internet and print media. Subjects were drawn from mid-career students in MBA and MSIS classes at a Midwestern univer-

3 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/media-personal-involvement-perceptions-data/14537](http://www.igi-global.com/chapter/media-personal-involvement-perceptions-data/14537)

## Related Content

---

### Composition in Object-Relational Database

David Taniar, Eric Pardede and J. Wenny Rahayu (2005). *Encyclopedia of Information Science and Technology, First Edition* (pp. 488-494).

[www.irma-international.org/chapter/composition-object-relational-database/14285](http://www.irma-international.org/chapter/composition-object-relational-database/14285)

### Application of Tree-Based Solutions

David Paper and Kenneth B. Tingey (2002). *Annals of Cases on Information Technology: Volume 4* (pp. 260-271).

[www.irma-international.org/article/application-tree-based-solutions/44511](http://www.irma-international.org/article/application-tree-based-solutions/44511)

### Leveraging Knowledge Reuse and Systems Agility in the Outsourcing Era

Igor Crk, Dane Sorensen and Amit Mitra (2008). *Journal of Information Technology Research* (pp. 1-20).

[www.irma-international.org/article/leveraging-knowledge-reuse-systems-agility/3694](http://www.irma-international.org/article/leveraging-knowledge-reuse-systems-agility/3694)

### Reuse of Formal Specifications

Laura Felice and Daniel Riesco (2005). *Encyclopedia of Information Science and Technology, First Edition* (pp. 2425-2430).

[www.irma-international.org/chapter/reuse-formal-specifications/14626](http://www.irma-international.org/chapter/reuse-formal-specifications/14626)

### Design Patterns from Theory to Practice

Jing Dong, Tu Peng, Yongtao Sun, Longji Tang and Yajing Zhao (2009). *Encyclopedia of Information Science and Technology, Second Edition* (pp. 1047-1052).

[www.irma-international.org/chapter/design-patterns-theory-practice/13704](http://www.irma-international.org/chapter/design-patterns-theory-practice/13704)