Association for Information Systems
AIS Electronic Library (AISeL)

December 2003

Assessments of Information Quality: Does the Medium Matter?

Barbara Klein
University of Michigan-Dearborn

Follow this and additional works at: http://aisel.aisnet.org/amcis2003

ASSESSMENT OF INFORMATION QUALITY:
DOES THE MEDIUM MATTER?

Barbara D. Klein
University of Michigan–Dearborn
bdklein@umd.umich.edu

Abstract

As the amount of information available through the Internet has exploded, users have come to depend on the Internet as a source of information. This study will improve our understanding of how users perceive the quality of this information. The study addresses a limitation of prior research by examining users' evaluations of the information quality of specific documents available on the web and in print. A repeated measures design is used to examine the question of whether the publication medium (print-based versus web-based) affects users' perceptions of the quality of information. The study also examines differences in users' evaluations as author credibility varies.   The findings of the study will have implications for decisions about media choice.

Keywords: Information quality, data quality, Internet, World Wide Web

Introduction

Users have come to rely on information published on the Internet for help in completing many kinds of tasks. People rely on web-based information when they read newspapers on the web, track mutual funds on the web, and so forth. In some cases, the source of this material is readily identifiable and fairly reputable. In other cases, the source of web-based material is less clear and questions about the quality of this information may arise. These problems have been noted in the literature on information accessed through the Internet. For example, since editorial and peer review processes are sometimes missing when information is published on the Internet, the information may not be reliable (Pack 1999). The trade-off between accuracy and timeliness has also been noted. The lack of review processes and ease of publication mean that information published on the Internet may be more timely but less accurate than information published in other media such as books and newspapers (Hawkins 1999).

The study proposed here is part of a research stream examining users' perceptions of the information quality of information published on the Internet. Previous studies in the research stream have compared users' perceptions of Internet-based information and information published in traditional text sources (i.e., books, journal, magazines, and newspapers) in a very broad way. The study proposed here examines users' perceptions of the information quality of specific web sites and traditional text sources. The effects of the medium (web-based versus print-based) and author credibility are examined. The results of the study will improve our understanding of the use of web-based information in the support of decision making. The remaining sections of this paper review the literature that informs the design of the study; discuss the findings and limitations of prior research conducted in this research stream; and explain the research questions, hypotheses, and methodology of the study proposed in this paper.

Literature Review

Dimensions of Data Quality

Data quality is a multi-dimensional concept. A great deal of research has sought to understand the dimensions of data quality. Definitions and frameworks for understanding data quality have been proposed by Davis and Olson (1985), Huh et al. (1990), Fox et al. (1993), Zmud (1978), Madnick and Wang (1992), and Wand and Wang (1996). Although the methodologies of these

studies differ, the dimensions of accuracy, completeness, consistency, and timeliness emerge from several of these studies. With the exception of the Zmud (1978) framework, these taxonomies of data quality have stopped short of suggesting measures with which to measure data quality.

In contrast, Wang and Strong (1996) developed a taxonomy of data quality which provides a powerful instrument for assessing users' perceptions of data quality. The instrument was developed by viewing data quality from the perspective of data consumers. Well-accepted methodologies used to develop instruments for measuring consumers' perceptions of products and services were used to develop the instrument. First, two surveys of data consumers were conducted to generate a comprehensive list of data attributes. In the first survey, data consumers were asked to list attributes of data quality. 118 attributes were generated. In the second survey, data consumers rated the importance of these 118 data attributes and an exploratory factor analysis of their responses was performed. Twenty dimensions of data quality were extracted. A second study was then performed in which subjects were asked to sort these twenty dimensions into four conceptually-derived categories (accuracy, relevancy, representation, and accessibility). Fifteen dimensions (encompassing 50 data attributes) emerged from the sorting study. The dimensions are believability, accuracy, objectivity, reputation, value-added, relevancy, timeliness, completeness, appropriate amount of data, interpretability, ease of understanding, representational consistency, concise representation, accessibility, and access security.

Data Quality and the Internet

The early literature on data quality and the Internet focused on warning users about the potential dangers of using Internet-based information and advocated the use of prescriptive guidelines for evaluating this information. For example, Hawkins (1999) recommends that users apply fourteen criteria to their evaluations of Internet-based information.

Prior findings on the use of web-based information are mixed. Rich and Belkin (1998) found that users perceive web-based information as less authoritative and credible than other kinds of information systems. However, Borchers (2002) found that, in some circumstances, users do not perceive strong differences between web-based and traditional sources of information. Additionally, Graham and Metaxas (2003) found that most college students do not double-check information found on the Internet.

Findings and Limitations of Prior Research

Findings

In an earlier study conducted in this research stream, graduate students taking an MBA course completed a survey about their perceptions of Internet and traditional text sources (i.e., books, journals, magazine, and newspapers) following the completion of a course project requiring the use of resources from both the Internet and traditional text sources. The survey asked two kinds of questions. First, questions about the extent to which the fifty data attributes identified by Wang and Strong (1996) describe information from Internet sources used for the course project were asked. Second, questions about the extent to which the same fifty data attributes describe information from traditional text sources (books, journals, magazines, and newspapers) used for the course project were asked (Klein 2001).

Scores for the data attributes for each of the dimensions of data quality were averaged to give a score for each of the fifteen data quality dimensions. Results for each dimension are shown in Table 1. Responses for each dimension could range from one to seven, with higher scores reflecting more favorable perceptions of data quality.

Timeliness and accessibility are among the highest rated data quality dimensions for Internet sources. These ratings reflect the speed with which information can be published on the Internet and the ease with which information can be accessed on the Internet. The accuracy and objectivity of Internet sources are given low ratings. Users of Internet-based information seem to be aware that errors may be present in this information.

Believability, reputation, interpretability, relevancy, and ease of understanding are the highest rated data quality dimensions for traditional text sources. Users recognize that traditional text sources are generally published by reputable sources, and they find this material relevant and easy to use. The timeliness and accessibility of traditional text sources were given low ratings. Users seem to be aware of the publication delays associated with traditional text sources and have problems accessing material stored in libraries and bookstores.

A comparison of the results for the Internet and traditional text sources shows that, the traditional text sources were given higher ratings than Internet sources on four data quality dimensions: accuracy, objectivity, reputation, and representational consistency. Internet sources were given higher ratings than traditional text sources on two data quality dimensions: timeliness and appropriate amount.

Table 1. Mean Scores for Data Quality Dimensions

Dimension of Data Quality	Perception of Internet Sources	Perception of Traditional Text Sources	Significant Difference (at $p < .05$)
Believability	5.60	5.89	No
Accuracy	4.17	4.64	Yes
Objectivity	4.03	4.67	Yes
Completeness	4.70	4.69	No
Reputation	4.61	5.29	Yes
Value-Added	4.80	4.24	No
Relevancy	5.38	5.07	No
Timeliness	5.51	4.11	Yes
Appropriate Amount	5.35	4.22	Yes
Interpretability	5.24	5.20	No
Ease of Understanding	5.04	5.06	No
Representational Consistency	3.87	4.70	Yes
Concise Reputation	4.39	4.64	No
Accessibility	5.37	4.15	No
Access Security	2.61	2.59	No

Limitations

The findings of this study indicate that at least some users are aware of the relative strengths and weaknesses of information published on the Internet and information published in traditional text sources such as books, journals, magazines, and newspapers. However, a limitation of the study is that respondents were asked questions about the quality of Internet and traditional text sources in general rather than being asked questions about specific Internet sites and text sources. The study proposed in this paper addresses this limitation by surveying users about their perceptions of the data quality of specific web sites and text sources.

Research Questions and Hypotheses

The study proposed here addresses the question of whether the publication medium (web-based versus print-based) affects users' perceptions of the quality of information. The study also examines whether differences in perceptions of web-based and print-based information hold when perceptions of the credibility of the author of the material vary.

Three hypotheses will be examined for each of the fifteen dimensions of data quality identified by Wang and Strong (1996). The three hypotheses for the accuracy dimension are stated below in the null form. The hypotheses for the other fourteen dimensions also follow this format.

$H1_0$: There is no difference in users' perceptions of the accuracy of web-based and print-based information.

$H2_0$: There is no difference in users' perceptions of the accuracy of information written by high-credibility and low-credibility authors.

$H3_0$: There is no interaction between the medium (web-based and print-based) and the credibility of the author.

Research Methodology

Research Design and Procedures

A two-factor repeated measures design (Neter et al. 1990) will be used to test the hypotheses (see Figure 1). The treatments will divide subjects into two levels for a "medium" treatment and two levels for an "author credibility" treatment.

		Medium	
		Web-Based	Print-Based
Author	High		
Credibility	Low		

Figure 1. Research Design

Medium

Two levels of the "medium" treatment are used: web-based and print-based. Documents published both in print and on the web will be used in the study. Within a level of the "author credibility" treatment, the information content of the document will be held constant, only the medium will vary. The web-based documents will be read online and will be formatted using <HTML> tags. The print-based documents will be printed on paper and will be formatted as a brochure. These documents may be PDF versions of information that is also available in a web-based format. The print-based documents will be selected so that they do not appear to have been accessed through the web.

Author Credibility

Two levels of the "author credibility" treatment will be used: high credibility and low credibility. A pretest will be conducted in order to determine perceptions of the credibility of the authors of a set of documents that are available in both web-based and print-based media. In the pretest, a group of one hundred subjects will be asked to read a collection of documents and assess the credibility of the authors of the documents. Documents that are judged to have high "author credibility" and documents judged to have low "author credibility" in the pretest will be selected for use in the study. The subjects who participate in the pretest will not participate in the main study.

Experimental Procedures

Each subject will complete four tasks, one for each cell of the research design shown in Figure 1. For each task, the subject will read a document and complete a survey measuring the subject's assessment of the quality of the information read. The survey has fifty items, one for each of the fifty data attributes identified by Wang and Strong (1996).

In order to control for order effects, a repeated measures design (Neter et al. 1990) is used in which the order of the tasks is varied among subjects (see Figure 2).

		Treatment Order			
		1	2	3	4
Subject	1	Web/Low	Print/Low	Web/High	Print/High
	2	Print/High	Web/Low	Print/Low	Web/High
	3	Print/Low	Web/High	Print/High	Web/Low
	4	Web/High	Print/High	Web/Low	Print/Low

Figure 2. Repeated Measures Design

Two different documents will be used in each cell of the research design (Figure 1) so that each subject reads different information content for each of the four tasks. Half the subjects will read and evaluate one of the documents, and the other half will read and evaluate the other document.

Subjects

A total of 400 subjects will participate in the study. Subjects will be recruited from graduate-level management courses. It is expected that the majority of subjects will have several years of work experience. Subjects will be randomly assigned to the experimental treatments.

Conclusion

As the amount of information available through the Internet has exploded, users have come to depend on the Internet as an important resource used to support decision making. This study will improve our understanding of how users perceive the quality of this information. The findings of the study will have implications for decisions about media choice and will shed light on the conditions in which conveying information about authorship is desirable. Future research will examine users' perceptions of the information quality of quantitative information and qualitative information published on the web and in print.

References

Borchers, A.S. "Intrinsic and Contextual Data Quality: The Effect of Media and Personal Involvement," Proceedings of the 2002 Information Resources Management Association International Conference, Seattle, WA, 2002, pp. 923-924.

Davis, G.B., and Olson, M.H. *Management Information Systems: Conceptual Foundations, Structure, and Development*, New York, NY: McGraw-Hill Book Company, 1985.

Fox, C., Levitin, A., and Redman, T. "The Notion of Data and Its Quality Dimensions," *Information Processing & Management* (30), 1993, pp. 9-19.

Graham, L., and Metaxas, P.T. "Of Course It's True; I Saw It on the Internet! Critical Thinking in the Internet Era," *Communications of the ACM* (46:5), 2003, pp. 71-75.

Hawkins, D.T. "What is Credible Information?," *Online* (23:5), 1999, pp. 86-89.

Huh, Y.U., Keller, F.R., Redman, T.C., and Watkins, A.R. "Data Quality," *Information and Software Technology* (32), 1990, pp. 559-565.

Klein, B.D. "User Perceptions of Data Quality: Internet and Traditional Text Sources," *Journal of Computer Information Systems* (41:4), 2001, pp. 5-15.

Madnick, S.E., and Wang, R.Y. "Introduction to the TDQM Research Program," Total Data Quality Management Research Program Working Paper #92-01, 1992.

Neter, J., Wasserman, W., and Kutner, M. H. *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Design*, Homewood, IL: Irwin, 1990.

Pack, T. "Can You Trust Internet Information?," *Link-up* (16:6), 1999, p. 24.

Rieh, S.Y., and N.J. Belkin. "Understanding Judgment of Information Quality and Cognitive Authority in the WWW," *Journal of the American Society for Information Science* (35), 1998, pp. 279-289.

Wand, Y., and Wang, R.Y. "Anchoring Data Quality Dimensions in Ontological Foundations," *Communications of the ACM* (39:11), 1996, pp. 86-95.

Wang, R.Y., and Strong, D.M. "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems* (12:4), 1996, pp. 5-34.

Zmud, R.W. "An Empirical Investigation of the Dimensionality of the Concept of Information," *Decision Sciences* (9), 1978, pp. 187-195.