Ixchel M. Faniel
OCLC Research

Adam Kriesberg
University of Michigan

Elizabeth Yakel
University of Michigan

Social Scientists' Satisfaction with Data Reuse

Note: This is a preprint of an article accepted for publication in *Journal of the Association for Information Science and Technology* copyright © 2015.

Abstract:
Much of the recent research on digital data repositories has focused on assessing either the trustworthiness of the repository or quantifying the frequency of data reuse. Satisfaction with the data reuse experience, however, has not been widely studied. Drawing from the information systems and information science literatures, we develop a model to examine the relationship between data quality and data reusers' satisfaction. Based on a survey of 1,480 journal article authors who cited Inter-University Consortium for Political and Social Research (ICPSR) data in published papers from 2008 – 2012, we found several data quality attributes - completeness, accessibility ease of operation, and credibility – had significant positive associations with data reusers' satisfaction. There was also a significant positive relationship between documentation quality and data reusers' satisfaction.

Suggested citation:
Faniel, I. M., Kriesberg, A. and Yakel, E. (2015). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*. doi: 10.1002/asi.23480

Introduction

In the past decade, digital data repositories in a variety of disciplines have increased in number and scope (Hey, Tansley, & Tolle, 2009). These repositories join social science data repositories that have been in existence for decades and have created a collaborative network to preserve quantitative social science data (King, 2011). Given the emergence of digital repositories to preserve and provide access to data, data managers increasingly need to provide evidence of repository success, especially when it comes to data access and reuse. By data reuse, we mean the secondary use of data for a purpose other than the original intention of the data producer (Karasti & Baker, 2008; Zimmerman, 2008). To date, research on repository success has focused on two areas: auditing and assessing the trustworthiness of repositories and developing ways to measure data reuse frequency. Drawing on the information systems and information science literatures, we contribute to research in the area by developing a model to examine the relationship between data quality and data reusers' satisfaction. Specifically, in this paper we report on a survey conducted among social scientists that reused data housed at Inter-university Consortium for Political and Social Research (ICPSR).

The audit and assessment literature focuses on establishing the trustworthiness of repositories. The attributes and criteria used to assess trustworthiness focus on the sustainability of organizational and technical infrastructures, workflows for data curation, and to a lesser extent responding to the needs of the designated community (Consultative Committee for Space Data Systems, 2012; Data Archiving and Networked Services (DANS), 2010; Dobratz, Schoger, & Strathmann, 2007). Auditors assess whether a repository has the mechanisms in place to sustain authentic and reliable data and provide long-term access to that data in a way that is meaningful to data reusers (Consultative Committee for Space Data Systems, 2012). The assumption behind these audits is that long term access to reliable and authentic data enables long term reuse of that data. Several researchers have developed indices to measure data reuse frequency and impact. Ingwersen and Chaven (2011) developed the Data Usage Index which is calculated from server logs and indicates the level of interest in and reuse of data assessed through searches and downloads. Services such as EZID and DataCite also facilitate tracking data reuse (Brase, 2009; California Digital Library, 2014; Michener et al., 2011; Wilkinson, 2010). Yet, measuring the frequency of citations or downloads only gets us part way to assessing impact. Based on social science citation data, Fear (2013) proposed two new measures for reuse impact: secondary citation (based on citations of the article citing the data) and diversity (based on the disciplinary breadth of data reuse). She found that no datasets consistently had high-impact across her metrics, indicating that how one measures impact can be a value judgment. Even though evidence that a repository can be trusted and data are being viewed, accessed, and cited speaks to repository success, this evidence tells us nothing about data reusers' perceptions of their experiences or how to improve these interactions. Our study addresses this gap by presenting empirical evidence on social scientists' satisfaction with data reuse. In order to suggest ways repository staff might improve data reusers' satisfaction, we focus on the relationship between perceptions about data quality and satisfaction with data reuse. Drawing from Caro et al. (2008) we define data quality as "the ability of a data collection to meet user requirements" (p. 514).

Satisfaction with commercial products has been studied extensively in the marketing literature; perceived product performance is one of the key predictors of this variable (e.g. Anderson & Sullivan, 1993; Churchill & Surprenant, 1982; Oliver & DeSarbo, 1988). Oliver (2010) suggests studying a product's micro-dimensions (i.e. attributes) in order to make the necessary improvements. Like Oliver, data reuse studies have treated data like a product possessing various quality attributes, even though its receipt may not result in a fee based transaction (Wang & Strong, 1996). To date, the studies have identified important data quality attributes and their role in evaluation and selection of data for reuse (e.g. Card, Shapiro, Amarillas, Mckean, & Kuhn, 2003; Faniel, Kansa, Whitcher Kansa, Barrera-Gomez, & Yakel, 2013; Niu, 2009; Van House, 2002; Zimmerman, 2008). However, consumer products studies suggest quality attributes related to choice may not be the same as those related to satisfaction (Oliver 2010). For data reusers, the differences may be partly due to what is experienced in the early versus later periods of the reuse process. The data reuse process consists of a number of stages before and after evaluation and selection, such as discovery, access, preparation, and analysis (Faniel, Kriesberg, & Yakel, 2012; Rolland & Lee, 2013; Zimmerman, 2008; Zimmerman, 2007). In this study we consider the entire data reuse experience by surveying social scientists retrospectively and determine what data quality attributes are positively associated with their satisfaction. Using a critical incident approach, we ask them to recall an instance of data reuse resulting in a specific journal publication. Given our process-based approach, we suspect that a journal's rank also may be positively associated with data reusers' satisfaction and control for it in our model. The following research question motivates our study:

> *What data quality attributes influence data reusers' satisfaction after controlling for journal rank?*

Results from multiple regression analysis show several data quality attributes - completeness, accessibility, ease of operation, and credibility – had significant positive associations with data reusers' satisfaction. There was also a significant positive relationship between documentation quality and data reusers' satisfaction. Our study contributes to the data reuse literature by extending beyond merely highlighting what data quality attributes are important in the early stages of data reuse. By surveying social scientists retrospectively our study contributes knowledge about what attributes influence satisfaction with the entire data reuse process. Moreover we contribute a complementary perspective on repository success that is rooted in data reusers' perceptions of their whole data reuse experience and offer suggestions about how repository staff might influence perceptions of data quality and thereby increase repository success. In the following sections of the paper, we contextualize our study through a literature review, introduce our research model, and discuss our research methodology. Then, we present our findings before concluding with a discussion of the implications of our work and the potential for future research in this area.

Literature Review

In the information systems and information science literatures, the quality of digital content is presented as a multidimensional construct (e.g. Marchand, 1990; Olaisen, 1990; Taylor, 1986; Wang & Strong, 1996). One of the most comprehensive frameworks is based on a survey of people using data to make business decisions within their organizations (Wang & Strong, 1996).

Respondents in this study were asked to rate the importance of several data quality attributes, which yielded a framework of 15 attributes across four categories – intrinsic, contextual, representational, and accessibility. Intrinsic attributes describe the qualities data have in their own right, such as believability, accuracy, and source reputation. Contextual attributes include relevancy, timeliness, and completeness which were considered important for the task at hand. Representational attributes center on the presentation of data, including interpretability, consistency, and concision. Lastly, accessibility refers to concepts, such as availability and security. Drawing from Wang and Strong (1996) and studies of the web, Caro et al. (2008) present a similar framework for web portal data quality containing twice as many attributes, such as the amount and usefulness of documentation and attributes relating specifically to a web portal, such as attractiveness, customer support, and response time.

In both cases, these authors developed their conceptual frameworks by surveying respondents about the importance of a pre-defined set of data quality attributes. By contrast, empirical studies in the information science literature have tended to ask respondents to talk through the criteria they employ to decide what digital content to use. Although different names are used, findings in the information science literature show attributes similar to those presented in the Wang and Strong (1996) and Caro et al. (2008) frameworks, such as topicality (i.e. relevancy), accuracy, depth (i.e. completeness), recency or currency (i.e. timeliness), obtainability or availability (i.e. accessibility), and source reputation (Barry, 1994; Rieh, 2002; Wang & White, 1999; Wang & Soergel, 1998). Furthermore, respondents' reliance on criteria varies given the type of task and type of use decision. In Rieh's (2002) study, faculty and students deciding what bibliographic material to use mentioned quality attributes, such as goodness, usefulness, and accuracy of content, more than those related to cognitive authority (i.e. credibility of authors, publishers, document types, content). She argues that authority is less of a concern, because faculty and students selected material from known databases and library systems. Wang and White (1999) suggest the difference may be due to the type of decision. In their study, faculty and students mention attributes related to credibility and believability, such as the degree to which the material was a seminal or standard reference or the author and/or journal were reputable or authoritative, when deciding what bibliographic material to cite in research papers.

While understanding the relative importance of the aforementioned criteria and how they are employed and influence choice over the course of a task is informative, this current study examines the relationship between quality attributes and satisfaction. Specifically, we examine data reusers' satisfaction at the end of the entire reuse process. Studies that have examined the relationship between data quality and satisfaction, focus on satisfaction with the information system rather than the data (e.g. Seddon & Kiew, 1994; Seddon & Yip, 1992; Teo & Wong, 1998). Moreover, the studies have treated quality as a one dimensional construct, which runs counter to empirical findings showing individuals perceive quality as multi-dimensional (e.g. Faniel & Jacobsen, 2010; Rieh, 2002; Wang & Strong, 1996). We contend that a more fine-grained examination of the relationships individual data quality attributes have with data reuse satisfaction is warranted. We expect that this more fine-grained understanding will provide key diagnostics for data mangers to assess repository success. In the next section we discuss our research model.

The Research Model

Satisfaction

Oliver (2010) defines satisfaction as "the consumer's fulfillment response" (p. 8). In the context of our study, we apply this concept to the data reuse experience. Satisfaction is an affective response that can be measured based on a cumulative series of experiences over time. However, Oliver (2010) recommends measuring satisfaction using the critical incident approach when there is interest in obtaining diagnostic information. Given our interest in utilizing our findings to improve data reuse experiences via digital repositories and to determine whether a key data reuse outcome, journal rank is associated with data reusers' satisfaction, our measure is based on a specific reuse incident. In the paragraphs that follow, we hypothesize the relationships between several data quality attributes relating specifically to data (relevancy, completeness, accessibility, ease of operation, credibility) and data reusers' satisfaction. We also hypothesize the relationship two additional measures of data quality identified in the literature (data producer reputation, documentation quality) have on satisfaction. In this model, we consider quality to be a multidimensional construct composed of both attributes that relate directly to the data (relevancy, completeness, accessibility, ease of operation, credibility) and those provide additional insight into the data (data producer reputation, documentation quality). Lastly we include journal rank in our research model, given that publication is a desired outcome of data reuse.

Data relevancy

Data relevancy refers to the degree to which data apply to and help the task at hand (Wang & Strong, 1996). Previous research shows that it is one of the key criteria faculty and students use to select documents in their scholarly work (Rieh, 2002; Wang & Soergel, 1998) and scientific communities rate it as an important data quality attribute (Huang, Stvilia, Jörgensen, & Bass, 2012; Stvilia et al., 2014). In the data reuse literature, relevancy is considered during data reuse decisions. The criteria used to evaluate data relevancy are rooted in the reusers' research objectives (Darby et al., 2012). For instance, in the earthquake engineering community, a research objective for those running computational models is validation; they look to reuse existing experimental data that matches criteria related to their model parameters (Faniel & Jacobsen, 2010). The choice about which data to reuse also needs to align with methodological and data collection norms of one's discipline. Ecologists reusing data are concerned with justifying their sampling choices (Zimmerman, 2007), whereas social scientists seek out data that closely matches the conceptual definitions and measurements of their theoretical frameworks (Faniel et al., 2012).

Data reusers cannot always identify data that are a perfect match. For instance, the original study may have limitations inherent in the research design (Chin & Lansing, 2004; Faniel & Jacobsen, 2010). This leads to a choice: collect data, halt the research until the perfect data are found, or modify research objectives and reuse the available data. Given the time, money, and effort to collect data along with the likelihood that no data will be perfect, researchers often satisfice with the available data and shape reuse projects around what is possible given their access. The more reusers have to modify their research objectives to reuse the data, the less satisfied they are going to be with data reuse. Reusers who settle for data less relevant to research objectives are likely to have lower levels of satisfaction.

H1: Data relevancy is positively related to data reusers' satisfaction.

Data Completeness

Data completeness or comprehensiveness is the extent to which data have sufficient breadth, depth, and scope for the task at hand (Wang & Strong, 1996). Findings from the information science literature are mixed about whether completeness is an important quality attribute (Barry, 1994; Olaisen, 1990; Stvilia, Mon, & Yi, 2009; Wang & Soergel, 1998). The differences may have to do with study design since all criteria are not equally observable; just because respondents did not discuss a particular criterion does not mean that they did not consider it (Wang & Soergel, 1998). Differences also may have to do with completeness being less important than other factors. In a study of managers in the banking and insurance industries, Olaisen (1990) found that content did not need to be complete as long as it was reliable and the source credible. In the data reuse literature, findings show completeness is an important data quality attribute within scientific research communities (e.g. Huang et al., 2012; Stvilia et al., 2014). Quantitative social scientists talk about completeness with respect to missing data. For example, research participants skip questions or drop out of studies; data producers' change or drop questions over time (Faniel et al., 2012). In some human subjects' studies, research data may be missing because legal restrictions prevent sharing (Rolland & Lee, 2013). Missing data are an issue during reuse because smaller sample sizes impact the types of analyses reusers can perform and the level of confidence they have in the results. Rather than reject less complete data, reusers may opt to modify their research objectives if no better alternatives exist. Decisions to reuse less than complete data are likely to result in lower levels of satisfaction.

H2: Data completeness is positively related to data reusers' satisfaction.

Data Accessibility

Data accessibility refers to the extent to which the data are available or easily and quickly retrievable (Wang & Strong, 1996). Unlike other quality attributes, studies examining accessibility underscore the role of the technology used to access content (Caro et al., 2008; Huang et al., 2012; Olaisen, 1990; Stvilia et al., 2014, 2009; Wang & Strong, 1996). In the case of data reuse, the source can be anything from an internationally known digital data repository to a personal website. However, there are few mentions of data accessibility in digital data repository studies, because few disciplinary communities have moved beyond the early stages of repository development, use, and adoption. Most data reuse still occurs on a small, person-to-person scale with known colleagues or through data published in journal articles (Faniel et al., 2013; Faniel & Jacobsen, 2010; Zimmerman, 2007). Nevertheless, data reusers form perceptions of accessibility based on how easy or long it takes to get data; this is likely to affect data reusers' satisfaction. Reusers who request small amounts of data from many colleagues increase the likelihood that sharing will occur more readily (Zimmerman, 2007), but this also increases reusers' time and effort. The more difficult and time intensive it is to obtain data, the less productive data reuse becomes, which goes against perceptions that reusing data will positively impact productivity relative to collecting one's own data (Faniel, 2009). The less productive reusers perceive themselves during attempts to access data, the less satisfied they will be with their data reuse experience as a whole.

H3: Data accessibility is positively related to data reusers' satisfaction.

Data Ease of Operation

Data ease of operation is the extent to which data are easily managed and manipulated (i.e. updated, integrated, aggregated, reproduced) (Wang & Strong, 1996). Many disciplines cite the importance of ease of operation (e.g. Caro et al., 2008; Lee et al., 2002; Olaisen, 1990), although not all (e.g. Barry, 1994; Wang & Soergel, 1998). This is partly due to the type of content. The data reuse literature suggests that data's ease of operation influences decisions to reuse data when researchers want to integrate data from multiple datasets (Faniel et al., 2013; Faniel et al., 2012). Research shows ease of operation can be affected when data are missing, collected differently over time, or at different levels of aggregation (Faniel et al., 2012). Ease of operation can also be affected when fields have not developed disciplinary wide ontologies or controlled vocabularies that tie dispersed data elements and different concepts together seamlessly (Faniel et al., 2013). In many cases, data reusers do the additional work. This is likely to continue as data repositories move from a dataset to a variable centric model of data delivery (Vardigan, Granda, Hansen, Ionescu, & LeClere, 2010). The additional work reusers assume if data are not easy to operate acts to reduce the productivity boost expected from reusing others' data and is likely to lower satisfaction.

> *H4: Data ease of operation is positively related to data reusers' satisfaction.*

Credibility

Credibility refers to the "quality of being believed or accepted as true, real, honest" (Merriam-Webster, 2014). In this study we focus on two ways credibility concerns emerge in the data reuse process: data credibility and source credibility. The first is data credibility or the intrinsic quality data have in their own right (Wang & Strong, 1996). We stipulate data credibility because research has shown that the genre and form of the object matter (Rieh & Danielson, 2007). Similar to Wilson's (1983) notion of intrinsic plausibility, data credibility focuses on the reusers' perceptions of the data. This is evident in the research. When evaluating the credibility of content, people consider qualities such as reliability, validity, accuracy, and objectivity (Barry, 1994; Herring, 2001; Rieh, 2002; Wang & Soergel, 1998; Lee et al. 2002). In the credibility literature, two classic dimensions of credibility are believability and trust (Fogg & Tseng, 1999). We have created a broad definition of data credibility which combines these aspects of the concept.

Second, we address source credibility, particularly the perceived quality of the data producer's research work, what we term data producer reputation. This is similar to Wilson's (1983) concept of personal authority, which considers a source's reputation and accomplishments to date. Studies examining use of scholarly documents show faculty and students distinguish between the credibility of the content and the source (Barry, 1994; Rieh, 2002; Wang & Soergel, 1998). Similar findings are discussed in the data reuse literature when it comes to data credibility and data producer reputation.

Data Credibility

In a survey of researchers within the condensed matter physics and genomics communities, accuracy, believability, reliability, verifiability, precision, and authority were rated as important data quality attributes (Huang et al., 2012; Stvilia et al., 2014). Thus, our definition of data credibility encompassing many dimensions mirrors other constructs in the literature. Reusers

assess these attributes by seeking detailed depictions of research events, such as sampling methodologies, data collection procedures, measurement and coding, and instrumentation (Darby et al., 2012; Faniel et al., 2013; Faniel & Jacobsen, 2010; Faniel et al., 2012; Rolland & Lee, 2013). They do this to minimize misusing the data and increase their confidence in results (Faniel & Jacobsen, 2010; Faniel et al., 2012; Rolland & Lee, 2013). Reusers' level of concern about the results is proportional to the extent they believe the data to be true. If the data do not support hypotheses, reusers have to decide whether the data or their theoretical models are at fault. Reusers who believe the data are credible are likely to be more confident about their results. When they can rule out data as the problem, reusers are more likely to be satisfied with their data reuse experience regardless of their results.

H5: Data credibility is positively related to data reusers' satisfaction.

Data Producer Reputation

Data producer reputation is another means of assessing credibility. We define this term as the extent to which data reusers' perceive a data producer's research to be highly regarded. Reusers use a data producer's reputation as a means to evaluate the trust they have in the data. Reusers consider lineage, institutional affiliation, competence, commitment, past performance, prior experience, and shared orientations and values to inform their views of a data producer's reputation (Chin & Lansing, 2004; Faniel et al., 2013; Faniel et al., 2012; Jirotka et al., 2005; Van House, Butler, & Schiff, 1998; Van House, 2002; Zimmerman, 2008). Metzger and Flanagan (2010) also identify source reputation as one of the key heuristics of credibility assessment and note that it is a credibility cue, rather than a judgment based on deep investigation into the source.

> "When choosing between sources, people are likely to believe that a source whose name they recognize is more credible compared to unfamiliar sources. People appear to reason that name recognition is earned by positive interactions over time that are spread through social networks. The reputation heuristic may also be a subset of the 'authority' heuristic in credibility assessment (Metzger, Flanagan, & Medders, 2010, p. 426).

This relates closely to other credibility findings about the reliance on expertise in credibility judgments (Fogg et al., 2001). In this sense, data reusers are relying on others' expertise in assessing data. The implication is that reputable data producers create reputable data based on peer-reviewed publications and methodologies that are highly consistent with disciplinary norms and transparent. Strong beliefs in data producers and their research work reduce doubt about the data and thereby increase satisfaction with data reuse. A data producer's transparency also acts to lower reusers' fear of data misuse, which is likely to increase satisfaction. Lastly, reusers are likely to have an easier time publishing their results, because reputable data are more likely to be trusted than challenged during peer review which is likely to also increase satisfaction.

H6: Data producer reputation is positively related to data reusers' satisfaction.

Documentation Quality

Documentation quality refers to the degree to which written documentation about the data is suitable for use. Since research data are contextual by nature, reusers' need to have details

about data's context of production to decide whether the data are relevant, understandable, and trustworthy (e.g. Card et al., 2003; Carlson & Anderson, 2007; Faniel & Jacobsen, 2010; Niu, 2009). Based on three case studies in astronomy, social science, and anthropology, Carlson and Anderson (2007) found "data were not self-contained units that could easily be circulated, but always needed complementary external information to be understood or trusted" (p. 647). Written documentation is one piece of complementary external information and if done well it can be the primary piece of contextual information about the data. Studies show reusers get details about data's context of production in various ways, including their experience collecting similar data, verbal communication with and observation of data producers, metadata, and shared understanding about disciplinary standards, artifacts, and procedures, and published articles (e.g. Birnholtz & Bietz, 2003; Bourne, 2005; Carlson & Anderson, 2007; Faniel et al., 2012; Jirotka et al., 2005; Kriesberg, Frank, Faniel, & Yakel, 2013; Rolland & Lee, 2013; Van House, 2002; Zimmerman, 2008). Yet, documentation stands out as a key means of communicating data collection procedures even though data producers create it in very different ways (Chin & Lansing, 2004; Faniel et al., 2013; Kansa, Kansa, & Arbuckle, 2014; Zimmerman, 2008). Even within quantitative disciplines, data collection practices are highly specific (Carlson & Anderson, 2007). Making sense of how and why data are constructed are time- and effort-intensive processes that involve accessing information about the research methodology, coding procedures, problems encountered during a study and how they were resolved, missing data, instrumentation, and lab or field conditions (e.g. Carlson & Anderson, 2007; Faniel & Jacobsen, 2010; Faniel et al., 2012; Rolland & Lee, 2013). In some cases, reusers rely primarily on written documentation that has been amassed from multiple sources into one document, sometimes by repository staff, and the document acts as the hub of their search for detailed contextual information (e.g. Faniel & Jacobsen, 2010; Faniel et al., 2012; Vardigan, Heus, & Thomas, 2008). In others, they have to seek out and weave together a web of contextual details from multiple sources to create a coherent set of documented information (Faniel et al., 2013; Rolland & Lee, 2013). When documentation quality is high, there is less need to seek additional sources of information about the data, so reusers believe they are being more productive and therefore will be more satisfied with data reuse.

> *H7: Data documentation quality is positively related to data reusers' satisfaction.*

Journal Rank

Journal rank refers to the reputation of the journal where the data reuse study appears. Journal rank or impact factor can have an impact on a scholar's reputation and we were interested in seeing whether this impacted data reuse satisfaction, particularly since where an article will appear is unknown until late in the data reuse process. Typically peer reviewed, published journal articles represent acceptance of research within one's disciplinary community. Moreover these outputs are used as a means for evaluating performance for tenure and promotion. Within disciplinary communities, journals are ranked as more or less prestigious. A journal's prestige is based on objective measures, such as citation analysis and subjective measures, such as perceived importance of a journal within a disciplinary community. Research suggests journal impact factor relates to higher rates of data sharing (Piwowar & Chapman, 2010) as well as higher citations of the articles where the data originally appeared (Piwowar & Vision, 2013). We believe it also might positively relate to data reuse. Specifically for data reusers who are interested in disseminating their research through journal articles, we suspect that the more

highly ranked the journal where the article is published, the more satisfied they will be with their data reuse experience.

H8: Journal rank is positively related to data reusers' satisfaction.

In an effort to develop an alternative perspective on repository success, we proposed a model of data reusers' satisfaction that examines data quality attributes and journal rank of the article where the data reuse study was published. Before discussing our results, we provide details about our research methodology.

Research Methodology

The authors partnered with ICPSR for this study. Founded in 1962, ICPSR is a leader in the field of social science data preservation, access, and curation. Holding more than 50,000 data files, it serves diverse social science research communities. ICPSR recruits data from major studies and contracts with several survey organizations and federal agencies to obtain their data for preservation; funders also mandate data deposit for some projects. To encourage deposit, ICPSR provides data depositors with feedback about the number of downloads their dataset receives as well as citations to studies that use the dataset. The citations are listed on the homepage for each dataset as well as in the Bibliography of Data-Related Literature found on the ICPSR website.

Sample

For this study, we requested a subset of the Bibliography of Data-Related Literature from ICPSR to identify a sample of first authors of journal articles who reused data deposited in ICPSR (ICPSR, 2014). The bibliography was used for several reasons. First, the bibliography provided an identifiable list of social scientists that completed the entire data reuse process, from data discovery, access, and selection to data preparation, analysis, and study publication. Second, we used the critical incident technique Oliver (2010) suggested to focus participants on a recent instance of data reuse that resulted in a journal article publication listed in the Bibliography of Data-Related Literature. Third, we were able to collect journal rank data for the participants' published articles and examine its influence on data reusers' satisfaction.

We received a subset of the bibliography containing journal articles, conference proceedings, theses, book sections, and books published between 2006 and 2012 that cited data deposited in ICSPR. There were a total of 8,461 entries. We reduced the sample to journal articles produced between 2009 and 2012 since prior research has found that memory recall decreases with the passage of time (Tourangeau, Rips, & Rasinski, 2000). In some instances the first authors published more than one journal article within the 3 year timeframe. We kept the most recent journal publication and removed the others from our list. In some instances the first authors were also the producers of the cited data; we eliminated these authors from the list as well. Next, we conducted online searches, limited to five minutes, to gather contact information for the first authors. Individuals were removed from the list when contact information could not be found. The resulting sample size was smaller than anticipated, so we returned to the Bibliography of Data-Related Literature and identified the ten most frequently used datasets deposited in ICPSR, this time gathering first authors of journal articles published in 2008. These

names were processed as previously described. We were left with a sample size of 1,480 after additional individuals were removed from the list, due to personal requests, undeliverable email, or incorrect identifications.

Operationalization of Constructs

Our proposed research model contains the nine constructs discussed in the previous section of this paper. All items were collected from study participants and measured based on a seven-point Likert scale ranging from one (strongly disagree) to seven (strongly agree), with the exception of journal rank, which the research team collected independently[1]. Data reusers' satisfaction was comprised of four survey items adapted from Flavian et al. (2006). The data quality indicators (relevancy, completeness, accessibility, ease of operation, credibility) were adapted from Lee et al. (2002); each concept was comprised of three survey items. The data producer reputation and documentation quality constructs were newly created for this study. Each also consisted of three survey items. Lastly, the research team collected journal rank data independently. The data was collected from Scopus SCImago Journal Rank (SJR) because it had the best coverage of the journals represented in our survey data. Our decision to use Scopus aligns with Norris and Oppenhiem (2007) who concluded that Scopus was the best resource for social science citation analysis. There were 276 SJR rankings for journal articles in our survey, whereas Web of Science and Google Metrics had only 246 and 262 respectively. In addition, SJR rankings for all journals were available, which allowed us to understand where our survey data fell within the full Scopus sample.

The survey items were finalized based on the results of two pilot tests. The first pilot consisted of three cognitive walkthroughs. The cognitive walkthroughs were conducted in one-on-one sessions with three ICPSR staff members who reuse data as part of their jobs. Using the concurrent think aloud technique, participants were asked to verbalize their thoughts while answering the survey questions (Groves et al., 2009). We used participants' feedback to clarify questions. This test with subject matter experts who were representative of the main population of interest, as recommended by Anderson and Gerbing (1991), served as a test of content validity which we define as "the degree to which items in an instrument reflect the content universe to which the instrument will be generalized" Straub et al. (2004, p. 424). Through this test, we ensured that question content was able to measure the complex constructs we sought to measure (Bernard 2000). The second pilot employed a web-based survey using Qualtrics. This pilot was administered to forty four social scientists, twenty-seven of whom completed the survey. We utilized the pilot data to determine survey timing and to select final survey items.

Survey Administration

We created the final web-based survey using Qualtrics, a web-based survey administration platform. Prior to survey administration, the director of ICPSR sent an email informing potential participants about the upcoming survey invitation and encouraged participation. One week later, we sent an email to participants inviting them to complete the survey. The email was personalized; each potential participant was asked to complete the survey based on a particular journal article they authored. Three additional follow-up notices were sent in successive weeks and the survey was closed after six weeks. We received 249 usable surveys out of 1,480 for a response rate of 16.8%. We removed nine additional survey responses from our analysis because the journals in which the respondents published their articles did not have an SJR ranking.

Data Analysis and Results

To prepare to test the hypotheses, we tested for convergent validity, discriminant validity, and reliability. We also examined diagnostics related to multiple regression assumptions, multicollinearity, and identification of outliers. We identified three outliers. In the paragraphs that follow we discuss data analyses results after removing the three outliers, which resulted in n=237.

To demonstrate convergent validity, items for each survey construct were submitted to principal components factor analysis with varimax rotation. Results showed a single factor for each construct. However, some items were removed because of low factor loadings, including EASEOFOP05 and SATIS01 (Table 1). The remaining factor loadings for each construct were above the .70 threshold. The Cronbach's alpha for each variable was also around or above the .70 threshold (Nunnally, 1978). To test discriminant validity, we subjected items comprising each construct to principal components factor analysis with varimax rotation (Table 2). Items that had low factor loadings (below .50) on their own construct or loaded on another construct at or above .50 were considered for elimination. Additional items (RELEV03, REP01) were removed for cross loading on other factors

Table 1. Survey items and factor loadings

Construct	Factor Loadings	Survey Items
Data Relevancy	0.93	The data were useful for my work. (RELEV01)
	Eliminated	The data were appropriate for my work. (RELEV03)
	0.93	The data were applicable to my work. (RELEV04)
Data Completeness	0.84	The data included all necessary values for my project. (COMPLT01)
	0.84	The data were sufficiently complete for my needs. (COMPLT04)
	0.86	The data had sufficient breadth and depth for my task. (COMPLT05)
Data Accessibility	0.84	The data were easily retrievable. (ACCESS01)
	0.92	The data were easily accessible. (ACCESS02)
	0.90	The data were easily obtainable. (ACCESS03)
Data Ease of Operation	0.92	The data were easy to manipulate. (EASEOFOP01)
	0.92	The data were easy to aggregate. (EASEOFOP02)
	Eliminated	The data were easy to merge with other data. (EASEOFOP05)
Data Credibility	0.87	The data were credible. (CRED01)
	0.84	The data were reliable. (CRED02)
	0.81	The data were objectively collected. (CRED03)
Data Producer Reputation	Eliminated	The data producer has a reputation for quality scholarship. (REP01)
	0.91	The data producer has a reputation for creating good datasets. (REP02)
	0.91	The data producer has a reputation for providing accurate documentation. (REP03)
Documentation Quality	0.87	The documentation was sufficient for my use of the data. (DOC01)
	0.84	The documentation increased my understanding of the data. (DOC02)
	0.86	The documentation was presented in a way that was clear to me. (DOC03)
Data reusers' satisfaction	Eliminated	I think I made the correct decision to use the data. (SATIS01)
	0.82	The experience that I had with the data was satisfactory. (SATIS02)
	0.83	I was satisfied with the way the data were delivered to me. (SATIS03)
	0.73	I was satisfied with the tools and services provided during my data reuse. (SATIS04)

Table 2. Results of factor analysis

	Component							
	1	2	3	4	5	6	7	8
ACCESS02	**.898**	.108	.004	.079	.027	.159	.126	.023
ACCESS03	**.845**	.120	.088	.119	-.035	.149	.046	.224
ACCESS01	**.775**	.100	.028	.055	.243	.110	.122	.137
DOC01	.083	**.834**	.096	.099	.177	.089	.053	.131
DOC03	.170	**.780**	.120	.038	.232	.153	.146	.044
DOC02	.096	**.739**	.150	.230	.103	.032	.152	.158
COMPLT05	.025	.091	**.848**	.188	.038	.087	.128	-.024
COMPLT01	.010	.049	**.798**	.057	.080	.148	-.005	.316
COMPLT04	.081	.229	**.729**	.226	.232	.079	.038	.030
RELEV04	.149	.153	.239	**.800**	.122	.120	.274	.100
RELEV01	.084	.147	.255	**.760**	.241	.151	.166	.077
CRED03	.152	.183	.218	.157	**.734**	-.124	.214	.028
CRED02	-.017	.306	.072	.183	**.688**	.303	.117	.196
CRED01	.115	.359	.125	.489	**.585**	.108	.116	.052
EASEOFOP02	.182	.059	.158	.092	.103	**.847**	.183	.103
EASEOFOP01	.264	.203	.143	.171	.017	**.799**	.096	.130
REP02	.137	.103	.057	.231	.109	.111	**.846**	.092
REP03	.140	.209	.076	.124	.219	.170	**.807**	.058
SATIS03	.301	.339	.142	.223	-.009	.185	.109	**.656**
SATIS04	.293	.081	.122	-.094	.423	.125	.197	**.614**
SATIS02	.135	.233	.432	.429	.093	.125	-.039	**.545**

We also used two methods to test for common method bias given our cross-sectional survey design, Harman's single factor test and examination of the correlation matrix (Lowry & Gaskin, 2014). Based on an unrotated factor solution for all constructs, results of the Harman's single factor test produced 21 distinct factors with the largest factor explaining 37.4% of the variance. Next, correlations among all constructs were examined (Table 3). Correlations were significant among all independent variables at $p < .001$, with the exception of journal rank. Since Harman's single factor test showed that less than the majority of the variance (i.e. 50%) could be explained by a single factor and all correlations among constructs were well below the .90, we concluded that common method bias was not a problem (Lowry & Gaskin, 2014). Once we were confident that we had valid, reliable constructs and no evidence of common method bias, we created our variables by averaging the survey items that comprised each. Next, we calculated z-scores for each variable in preparation for the final multiple regression analysis. Descriptive statistics for the independent and dependent variables are shown in Table 3. We also requested demographic information from our respondents. All 237 respondents did not answer all of the questions, but enough provided answers for us to report rough approximations. Most of the respondents who answered these questions (80.5%) were tenure track faculty (n=230) and had reused data for an average of 12.67 years (n=219). As social scientists, their disciplinary areas included sociology, medical sciences, social work, psychology, education, and political science (n=230). Approximately 60% felt there was sufficient data available for reuse in their field (n=231). The respondents were split between using one dataset (58.3%) and combining variables from multiple datasets (41.7%) during data reuse (n=230). Even though nearly 67.5% of the respondents reported collecting their own data (n=231), only 11.3% indicated that they had contributed their data to ICSPR (n=230). Moreover, ICPSR was not their only data source. When it come to the specific data reuse incident for which we surveyed respondents (n=234),

only 32.1% reported they exclusively accessed the data for their article through ICPSR; 43.2% used a combination of sources including ICPSR and 20.9% indicated that all of their data came from other sources.

Table 3. Descriptives and correlations

	μ	S.D.	α	1	2	3	4	5	6	7	8
1. Journal rank	1.43	1.13	n/a								
2. Data relevancy	6.51	0.53	0.85	.044							
3. Data completeness	5.71	1.03	0.78	.004	.492**						
4. Data accessibility	6.04	1.05	0.86	-.006	.305**	.177**					
5. Data ease of operation	6.00	1.00	0.81	.010	.403**	.348**	.443**				
6. Data credibility	6.23	0.64	0.78	.039	.556**	.405**	.292**	.330**			
7. Data producer reputation	6.32	0.80	0.79	-.006	.489**	.242**	.343**	.395**	.492**		
8. Documentation quality	6.05	0.73	0.82	-.030	.444**	.348**	.325**	.358**	.575**	.395**	
9. Data reusers' satisfaction	6.19	0.68	0.69	.032	.495**	.495**	.529**	.489**	.524**	.398**	.535**

*p < .05, **p < .01, ***p < .001

We tested and found that assumptions of linearity, normality, heteroscedasticity, and independence of error terms were not violated. We also tested for multicollinearity. Small tolerance values and high variance inflation factors are indicators of multicollinearity and common thresholds have been established at 0.10 and 10,respectively (Hair, Anderson, Tatham, & Black, 1998). Results showed the lowest tolerance value as .52 and highest variance inflation factor (VIF) as 1.94, leading us to determine that multicollinearity was not a problem in our dataset.

Results from the multiple regression analysis are shown in Table 4. The R^2 of 55.5% indicated that the model effectively explained the variance in data reusers' satisfaction ($F(8,228)=35.59$, $p < .001$). Four of the five data quality attributes had significant positive associations with data reusers' satisfaction, including data completeness (B =.245, t=4.51, p<.001), data accessibility (B =.320, t=5.96, p<.001), data ease of operation (B =.134, t=2.31, p<.05), and data credibility (B =.148, t=2.38, p<.05). Surprisingly, the relationship between data relevancy and data reusers' satisfaction was not significant. Documentation quality also showed a strong positive association with data reusers' satisfaction (B =.204, t=3.44, p<.01), but data producer reputation was not significant. Interestingly journal rank, a key criterion for measuring performance in academia also was not significant. Journal rank had no bearing on data reusers' satisfaction. In

short, H2, H3, H4, H5, and H7 were supported, whereas H1, H6, and H8 were not supported. In the section that follows we discuss the implications of our findings in detail.

Table 4. Regression results

	B
Constant	-.030
Data Relevancy	.066
Data Completeness	.245***
Data Accessibility	.320***
Data Ease of Operation	.134*
Data Credibility	.148*
Documentation Quality	.204**
Data producer reputation	.008
Journal rank	.030
Model Statistics	
N	237
R^2	55.5%
Adjusted R^2	54.0%
Model F	35.59***

*p < .05, **p < .01, ***p < .001

Discussion

The major objective of this study was to introduce data reusers' satisfaction as another means to measure repository success. From the data reuse literature we know data quality attributes contribute to decisions about what data to reuse. However, in this study, we were interested in empirically testing whether any of the same factors also were positively associated with data reusers' satisfaction. We found that data completeness (H2), data accessibility (H3), data ease of operation (H4), and data credibility (H5) were significant as predicted. Support for these hypotheses suggests that data reusers' satisfaction corresponded with reusing data that were comprehensive, easy to obtain, easy to manipulate, and believable. We also found that documentation quality (H7) was significant. Higher levels of documentation quality corresponded with higher levels of data reusers' satisfaction.

Data accessibility had the strongest relationship with data reusers' satisfaction. Social scientists were more satisfied when the data they reused were easily obtainable. This may be due to the fact that most respondents relied on multiple sources to obtain data. Despite the large data collection ICPSR houses, only about 32% of the social scientists obtained data from ICSPR exclusively. Data completeness had the second strongest relationship with data reusers' satisfaction. A possible explanation for the strength of the relationship is that data completeness affects sample sizes and the power of the statistical tests social scientists were able to perform. More complete data results in larger sample sizes, a higher probability of correctly rejecting the null hypothesis or conversely correctly accepting the alternative hypotheses, and greater confidence in findings. Documentation quality had the third strongest relationship to data

reusers' satisfaction. Drawing from prior research, the importance of high documentation quality may be attributed to the fact that it facilitated an in-depth understanding of the data collection procedures and, subsequently, increased trust in the data and the results of reuse (Carlson & Anderson, 2007; Faniel & Jacobsen, 2010). Data credibility had the fourth strongest relationship with data reusers' satisfaction. Social scientists were more satisfied when they believed in the data. It may be that social scientists who believe the data truly depicted research events, were more satisfied with the results of their data analysis regardless of whether the outcome supported their hypotheses. Data ease of operation had the fifth strongest relationship with data reusers' satisfaction. Social scientists were more satisfied with data reuse when the data were easy to manipulate. This is likely due to the fact that when they obtained data from multiple sources they then had to combine variables from multiple datasets.

Not all of the hypotheses were supported. Surprisingly, data relevancy (H1) and data producer reputation (H6) were not significant in the multiple regression model. A possible explanation as to why data relevancy was not significant is that social scientists who willingly adapt research questions to available data still demanded high levels of relevancy to go forward with reuse. Reusing data that are not highly relevant would require reusers to alter their research objectives to the point of being less fruitful lines of inquiry, so they avoid these data. It may be that data producer reputation was not significant because social scientists were basing their judgments of it on name recognition rather than a through vetting of the person. Studies show that even when reputable data producers deposit data, reusers still critically examine the data to determine whether to reuse or reject (Faniel & Jacobsen, 2010; Rolland & Lee, 2013; Zimmerman, 2008).

Interestingly, journal rank (H8) also was not significant. Comparing the mean journal rank for our sample (1.45) to the mean of all journals represented in Scopus Journal Analyzer (0.61) suggests that the social scientists may have published their data reuse studies in fairly reputable journals. Moreover data reuse is a common phenomenon in the social sciences. There is no stigma associated with reusing others' data and top journals important to the social science community accept data reuse studies. Consequently social scientists judge their satisfaction with data reuse based on the quality of data alone.

Theoretical Implications

Prior studies have found significant positive relationships between data quality and satisfaction, but data quality was treated as a one-dimensional construct in this work and satisfaction with the technology used to access data was the dependent variable of interest (Seddon & Kiew, 1994; Seddon & Yip, 1992; Teo & Wong, 1998). Studies that have examined data quality as a multi-dimensional construct have only examined the importance of individual data quality attributes or their role in decisions to reuse the data (Faniel et al., 2013; Faniel & Jacobsen, 2010; Faniel et al., 2012; Huang et al., 2012; Stvilia et al., 2013; Zimmerman, 2008; Zimmerman, 2007). In this study we treated data quality as a multi-dimensional construct comprised of several different attributes with an interest in how each related to data reusers' satisfaction. We developed a model of data reusers' satisfaction in which different data quality attributes were treated independently. We also created new, valid, reliable measures for documentation quality and data producer reputation to examine additional quality attributes that provide insight to the data. We also ruled out a key competing hypotheses by controlling for journal rank. We found data quality attributes can operate independently on data reusers' satisfaction to provide a more nuanced understanding of what really matters. Data reusers were

more satisfied when data were comprehensive, easy to obtain, easy to manipulate, believable, and came with high quality documentation. Interestingly, data reusers' satisfaction was not associated with data that were applicable or deposited by reputable data producers, which showed that data quality attributes important in decisions early in the data reuse process (data selection) were not necessarily the same as those that positively relate to data reusers' satisfaction when reflecting on the entire process.

Practical Implications

Our findings have several implications for repository staff who are interested in providing services leading to reusers' satisfaction. As a first step, we suggest incorporating assessments of comprehensiveness, ease of accessibility, ease of manipulation, believability, and high quality documentation into repository staff's evaluation processes. For instance, completeness is relative, so when data managers assess data based on completeness, the question should be not only how much or what data are missing, but also how missing data may influence the power of statistical tests. When data are selected that are less than fully complete, data managers should be clear why, how, and where data are missing (e.g. what variables) so reusers can make more informed decisions more quickly. Looking across studies to determine whether data can be integrated and reused together should also be considered, because this may minimize the impact of missing data.

Second, data often have to be embargoed upon deposit or have other access restrictions. Although reusers expect to encounter these issues, they may influence satisfaction if not handled properly and equitably. When setting embargo periods, staff should consider balancing the needs of data producers and reusers given publication cycles and other disciplinary norms. Managers also should ensure that the process through which confidential data are made available (e.g. when, to whom, under what circumstances, etc.) is easy to understand and transparent.

Third, data ease of operation speaks to the ease of manipulating and integrating data from more than one study. This indicates that for some data sets the aggregated whole may be more valuable to data reusers than the individual data sets. For commonly integrated data sets, repository staff might consider providing instructions. For instance, ICPSR provides cross-walks and other kinds of guidance for some of its datasets to facilitate research drawing variables from multiple datasets.

Fourth, data credibility emphasizes the believability of the data. It requires details about the research events and is thus dependent on high quality documentation. In determining whether data credibility assessments can be made, repository staff should consider the extent and ease with which details about data collection can be reconstructed from the documentation. Research also suggests publications that provide discussions of support for and critique of the data are also used to assess credibility (Faniel et al., 2012), so repository staff should also consider tracking and listing publications that cite the data to offer additional insights.

Finally, our finding that documentation quality was a significant factor affecting data reusers' satisfaction is the most important outcome for repository managers because they can influence documentation quality. To a certain extent, repository mangers can control what information is

submitted along with the data. Through working with designated communities, these managers can also understand what is considered high quality documentation in a particular discipline. Furthermore, processing the documentation is a major way in which data managers add value to the data. In the case of social scientists, the data documentation initiative has created an expected structure for data and set norms about what types of documentation should accompany data sets. Not all data will or needs to be of perfect quality. Data reusers are willing to reuse data with limitations as long as those limitations are transparent (Faniel & Jacobsen, 2010). The importance of transparency is important for all data quality attributes as this enables data reusers to better understand the data and assess problems in their analysis. Given the strong relationship between documentation quality and data reusers' satisfaction, data managers should spend more time reviewing what is being submitted along with the data. They should also consider what makes for high quality documentation in a particular discipline and consult with the disciplinary community to develop guidelines for data producers to follow. Using the guidelines to review incoming documentation and rate its quality could be used as another means data selection.

Limitations and Future Research

In examining the relationship between data quality attributes and data reusers' satisfaction, this study considered seven data quality attributes. The data quality frameworks Wang and Strong (1996) and Caro et al. (2008) proposed had 15 and 33 attributes, respectively. While we made informed decisions about what data quality attributes to include in this study, future research should consider whether there are others that should be examined when studying data reuse within academic research communities. For instance, we did not include data quality attributes related to the information system (e.g. data repository) used to access the data, such as fast and easy navigation, secure movement of data between user and information system, on-line support, or response time, because we could not guarantee that all respondents obtained data from the same source. Future research that examines data reuse from a repository exclusively should consider drawing from Caro et al. (2008) and Wang and Strong (1996) to include these concepts. As another example we chose not to include representational consistency, which measures the extent to which data are compatible and in the same format (Wang and Strong, 1996), because we thought ease of operation encompassed it in that it is difficult to merge or aggregate data that are not compatible. We also had to drop two data quality attributes (interpretability and traceability), important for data reuse, because they were not valid, reliable constructs. Future research should consider adapting or developing new measures for these constructs. Similar to the information systems and information science literatures (e.g. Barry, 1994; Caro et al., 2008; Herring, 2001; Huang et al., 2012; Lee et al., 2002; Rieh, 2002; Stvilia et al., 2014; Wang and Soergel, 1998; Wang and Strong, 1996), we also kept our definition of data credibility broad. However, treating data credibility as a multidimensional construct, each dimension could be examined independently for differing associations with satisfaction. Future research should consider whether data credibility should be treated as a multidimensional construct where each dimension is examined independently.

We limited our study of data reusers' satisfaction to the social science community. Although this community is diverse, quantitative data are commonly reused to test statistical models. Reuse of the same data for a different purpose or in a diverse community might yield other judgments about data quality "…depending on the context of a particular use and the individual or community value structures for quality" (Stvilia, Twidale, Smith, & Gasser, 2008, p. 983).

Differences might also be due to how long data reuse has been occurring in the community. Data reuse within the social science community is quite common and has been occurring for decades, which means the norms for acceptable data quality attributes and levels are well established. Future research should consider examining the relationships data quality attributes have with data reusers' satisfaction in disciplines with less established traditions of data reuse. The type of data and the different means through which the data are collected within different disciplinary communities might also influence data quality judgments during reuse. We examined social science data that is primarily survey data. Future research should consider reuse of other types of data collected through different methodological procedures, such as sensor data collected during experiments conducted in the laboratory, through fieldwork, such as archaeological data collected during an excavation or survey, or as a by-product of some activity, such as social media data.

This cross-sectional study relied on retrospective accounts of a critical data reuse incident in order to account for the full data reuse process – discovery, access, decision to reuse, preparation, analysis, publication. However the study only measured data quality attributes at the end of the data reuse process to examine their relationship with satisfaction. Surprisingly we found that data relevancy and data producer reputation, while important when deciding whether to reuse data were not significantly related to data reusers' satisfaction at the end of the process. Future research should consider examining the relationship data quality attributes have with data reusers' satisfaction longitudinally at each step in the data reuse process. Such an approach would provide a more holistic view about which data quality attributes are important for which data reuse actions. This would provide a better understanding of the dynamics of the data reuse process, the role of data quality attributes, and how repository staff might intervene to improve the reuse experience throughout the process. While common method variance was not a problem in this study, separating the collection of the independent variables from the dependent variable would be another advantage to using this approach.

Endnotes

[1]Two additional data quality attributes were collected (interpretability, traceability), but later dropped from the study, because factor analysis indicated that they were not valid or reliable measures.

Acknowledgements

This research was funded by a National Leadership Grant from the Institute of Museum and Library Services, LG-06-10-0140-10, "Dissemination Information Packages for Information Reuse." We also thank ICPSR for assisting us in this study.

Bibliography

Anderson, E. W., & Sullivan, M. W. (1993). The Antecedents and Consequences of Customer Satisfaction for Firms. *Marketing Science*, *12*(2), 125–143.

Barry, C. L. (1994). User-Defined Relevance Criteria: An Exploratory Study. *Journal of the American Society for Information Science*, *45*(3), 149.

Birnholtz, J. P., & Bietz, M. (2003). Data at Work: Supporting Sharing in Science and Engineering. In *ACM Conference on Supporting Group Work* (pp. 339–348). Sanibel Island, FL. doi:10.1145/958160.958215

Bourne, P. E. (2005). Will a Biological Database be Different from a Biological Journal. *PLoS Computational Biology*, *1*(3), 179–181. doi:10.1371/journal.pcbi.0010034

Brase, J. (2009). DataCite - A Global Registration Agency for Research Data. In *Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology, 2009. COINFO '09* (pp. 257–261). doi:10.1109/COINFO.2009.66

California Digital Library. (2014). Why Use EZID? Retrieved May 5, 2014, from http://ezid.cdlib.org/home/why

Card, J. J., Shapiro, L., Amarillas, A., Mckean, E., & Kuhn, T. (2003). Broadening Public Access to Data Through the Development of Tools for Data Novices. *Social Science Computer Review*, *21*(3), 352–359. doi:10.1177/0894439303253983

Carlson, S., & Anderson, B. (2007). What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. *Journal of Computer-Mediated Communication*, *12*(2), 635–651. doi:10.1111/j.1083-6101.2007.00342.x

Caro, A., Calero, C., Caballero, I., & Piattini, M. (2008). A Proposal for a Set of Attributes Relevant for Web Portal Data Quality. *Software Quality Journal*, *16*(4), 513–542. doi:10.1007/s11219-008-9046-7

Chin, G., & Lansing, C. S. (2004). Capturing and Supporting Contexts for Scientific Data Sharing via the Biological Sciences Collaboratory. In *ACM Conference on Computer Supported Cooperative Work* (pp. 409–418). Chicago, Illinois, USA: ACM. doi:10.1145/1031607.1031677

Churchill, G. A., & Surprenant, C. (1982). An Investigation into the Determinants of Customer Satisfaction. *Journal of Marketing Research*, *19*(4).

Consultative Committee for Space Data Systems. (2012). *Space Data and Information Transfer Systems — Audit and Certification of Trustworthy Digital Repositories* (Standard No. ISO 16363:2012 (CCSDS 652-R-1)). Washington, D.C.: Consultative Committee for Space

Data Systems. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

Darby, R., Lambert, S., Matthews, B., Wilson, M., Gitmans, K., Dallmeier-Tiessen, S., … Suhonen, J. (2012). Enabling Scientific Data Sharing and Re-Use. In *2012 IEEE 8th International Conference on E-Science (e-Science)* (pp. 1–8). doi:10.1109/eScience.2012.6404476

Data Archiving and Networked Services (DANS). (2010). *Data Seal of Approval: Quality Guidelines for Digital Research Data*. The Hague. Retrieved from http://www.datasealofapproval.org/?q=about

Dobratz, S., Schoger, A., & Strathmann, S. (2007). The nestor Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification. *The Journal of Digital Information*, *8*(2). Retrieved from http://journals.tdl.org/jodi/index.php/jodi/article/view/199/180

Faniel, I., Kansa, E., Whitcher Kansa, S., Barrera-Gomez, J., & Yakel, E. (2013). The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 295–304). New York, NY, USA: ACM. doi:10.1145/2467696.2467712

Faniel, I. M. (2009). *Unrealized Potential: The Socio-Technical Challenges of a Large Scale Cyberinfrastructure Initiative*. National Science Foundation. Retrieved from http://hdl.handle.net/2027.42/61845

Faniel, I. M., & Jacobsen, T. E. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported Cooperative Work*, 19(3-4), 355–375. doi:10.1007/s10606-010-9117-8

Faniel, I. M., Kriesberg, A., & Yakel, E. (2012). Data Reuse and Sensemaking Among Novice Social Scientists. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1–10. doi:10.1002/meet.14504901068

Fear, K. M. (2013). *Measuring and Anticipating the Impact of Data Reuse* (Dissertation). University of Michigan, Ann Arbor, MI. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/102481

Flavián, C., Guinalíu, M., & Gurrea, R. (2006). The Role Played by Perceived Usability, Satisfaction and Consumer Trust on Website Loyalty. *Information & Management*, *43*(1), 1–14. doi:10.1016/j.im.2005.01.002

Fogg, B. J., Swani, P., Treinen, M., Marshall, J., Laraki, O., Osipovich, A., … Shon, J. (2001). What makes Web sites credible?: a report on a large quantitative study (pp. 61–68). ACM Press. doi:10.1145/365024.365037

Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. Hoboken, N.J.: Wiley.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. (1998). *Multivariate Data Analysis* (5th ed.). Upper Saddle River, N.J.: Prentice Hall.

Herring, S. D. (2001). Using the World Wide Web for Research: Are Faculty Satisfied? *The Journal of Academic Librarianship*, *27*(3), 213–219. doi:10.1016/S0099-1333(01)00183-5

Hey, A. J. G., Tansley, S., & Tolle, K. M. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, Wash.: Microsoft Research. Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm

Huang, H., Stvilia, B., Jörgensen, C., & Bass, H. W. (2012). Prioritization of Data Quality Dimensions and Skills Requirements in Genome Annotation Work. *Journal of the American Society for Information Science and Technology*, *63*(1), 195–207. doi:10.1002/asi.21652

ICPSR. (2014, May). ICPSR Bibliography of Data-Related Literature. Retrieved May 6, 2014, from http://www.icpsr.umich.edu/icpsrweb/ICPSR/citations/

Ingwersen, P., & Chavan, V. (2011). Indicators for the Data Usage Index (DUI): An Incentive for Publishing Primary Biodiversity Data Through Global Information Infrastructure. *BMC Bioinformatics*, *12*(Suppl 15), S3. doi:10.1186/1471-2105-12-S15-S3

Jirotka, M., Procter, R., Hartswood, M., Slack, R., Simpson, A., Coopmans, C., … Voss, A. (2005). Collaboration and Trust in Healthcare Innovation: The eDiaMoND CaseStudy. *Computer Supported Cooperative Work*, *14*(4), 369–398. doi:10.1007/s10606-005-9001-0

Kansa, E. C., Kansa, S. W., & Arbuckle, B. (2014). Publishing and Pushing: Mixing Models for Communicating Research Data in Archaeology. *International Journal for Digital Curation*, *9*(1), 57–70. doi:10.2218/ijdc.v9i1.301

Karasti, H., & Baker, K. (2008). Digital Data Practices and Long Term Ecological Research Program Growing Global. *The International Journal of Digital Curation*, *3*(2), 42–58. doi:10.2218/ijdc.v3i2.57

King, G. (2011). Ensuring the Data-Rich Future of the Social Sciences. *Science*, *331*(6018), 719–721. doi:10.1126/science.1197872

Kriesberg, A., Frank, R. D., Faniel, I. M., & Yakel, E. (2013). The Role of Data Reuse in the Apprenticeship Process. In *Proceedings of the 76th ASIS&T Annual Meeting* (Vol. 50). Montreal, QC, Canada. Retrieved from http://asis.org/asist2013/proceedings/submissions/papers/49paper.pdf

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A Methodology for Information Quality Assessment. *Information & Management*, *40*(2), 133–146. doi:10.1016/S0378-7206(02)00043-5

Lowry, P. B., & Gaskin, J. (2014). Partial Least Squares (PLS) Structural Equation Modeling (SEM) for Building and Testing Behavioral Causal Theory: When to Choose It and How to Use It. *IEEE Transactions on Professional Communication*, *Early Access Online*. doi:10.1109/TPC.2014.2312452

Marchand, D. (1990). Managing Information Quality. In I. Wormell (Ed.), *Information Quality: Definitions and Dimensions* (pp. 7–17). London, U.K.: Taylor Graham.

Merriam-Webster. (2014). *credibility*. Retrieved from http://www.merriam-webster.com/dictionary/credibility

Metzger, M. J., Flanagan, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, 60(3), 413–439.

Michener, W., Vision, T., Cruse, P., Vieglais, D., Kunze, J., & Janée, G. (2011). DataONE: Data Observation Network for Earth — Preserving Data and Enabling Innovation in the Biological and Environmental Sciences. *D-Lib Magazine*, *17*(1/2). doi:10.1045/january2011-michener

Niu, J. (2009). *Perceived Documentation Quality of Social Science Data*. University of Michigan. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/63871

Norris, M., & Oppenheim, C. (2007). Comparing Alternatives to the Web of Science for Coverage of the Social Sciences' Literature. *Journal of Informetrics*, *1*(2), 161–169. doi:10.1016/j.joi.2006.12.001

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York, NY: McGraw-Hill.

Olaisen, J. (1990). Information Quality Factors and the Cognitive Authority of Electronic Information. In I. Wormell (Ed.), *Information Quality: Definitions and Dimensions* (pp. 91–121). London, U.K.: Taylor Graham.

Oliver, R. L. (2010). *Satisfaction: A Behavioral Perspective on the Consumer* (2nd ed.). Armonk, NY: M.E. Sharpe.

Oliver, R. L., & DeSarbo, W. S. (1988). Response Determinants in Satisfaction Judgments. *Journal of Consumer Research*, *14*(4), 495–507.

Piwowar, H. A., & Chapman, W. W. (2010). Public Sharing of Research Datasets: A Pilot Study of Associations. *Journal of Informetrics*, *4*(2), 148–156. doi:10.1016/j.joi.2009.11.010

Piwowar, H. A., & Vision, T. J. (2013). Data Reuse and the Open Data Citation Advantage. *PeerJ PrePrints*. doi:10.7287/peerj.preprints.1

Rieh, S. Y. (2002). Judgment of Information Quality and Cognitive Authority in the Web. *Journal of the American Society for Information Science and Technology*, *53*(2), 145–161. doi:10.1002/asi.10017

Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A multidisciplinary framework. *Annual Review of Information Science and Technology*, *41*(1), 307–364. doi:10.1002/aris.2007.1440410114

Rolland, B., & Lee, C. P. (2013). Beyond Trust and Reliability: Reusing Data in Collaborative Cancer Epidemiology Research. In *Collaboration and Sharing in Scientific Work* (pp. 435–444). San Antonio, Texas: ACM.

Seddon, P. B., & Kiew, M.-Y. (1994). A Partial Test and Development of the DeLone and McLean Model of IS Success. In J. I. DeGross, S. L. Huff, & M. Munro (Eds.), *Proceedings of the Fifteenth International Conference on Information Systems, Vancouver, British Columbia, Canada, December 14-17, 1994* (pp. 99–110). Association for Information Systems. Retrieved from http://aisel.aisnet.org/icis1994/9

Seddon, P., & Yip, S.-K. (1992). An Empirical Evaluation of User Information Satisfaction (UIS) Measures for Use with General Ledger Accounting Software. *Journal of Information Systems*, *6*(1), 75–92.

Stvilia, B., Hinnant, C. C., Wu, S., Worrall, A., Lee, D. J., Burnett, K., … Marty, P. F. (2013). Studying the Data Practices of a Scientific Community. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 425–426). New York, NY, USA: ACM. doi:10.1145/2467696.2467781

Stvilia, B., Hinnant, C. C., Wu, S., Worrall, A., Lee, D. J., Burnett, K., … Marty, P. F. (2014). Research Project Tasks, Data, and Perceptions of Data Quality in a Condensed Matter Physics Community. *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23177

Stvilia, B., Mon, L., & Yi, Y. J. (2009). A Model for Online Consumer Health Information Quality. *Journal of the American Society for Information Science and Technology*, *60*(9), 1781–1791. doi:10.1002/asi.21115

Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2008). Information Quality Work Organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, *59*(6), 983–1001. doi:10.1002/asi.20813

Taylor, R. S. (1986). *Value-Added Processes in Information Systems*. Norwood, NJ: Ablex Pub. Corp.

Teo, T. S. H., & Wong, P. K. (1998). An Empirical Study of the Performance Impact of Computerization in the Retail Industry. *Omega*, *26*(5), 611–621. doi:10.1016/S0305-0483(98)00007-3

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge, U.K.; New York: Cambridge University Press.

Van House, N. (2002). Digital Libraries and Practices of Trust: Networked Biodiversity Information. *Social Epistemology: A Journal of Knowledge, Culture and Policy*, *16*(1), 99. doi:10.1080/02691720210132833

Van House, N. A., Butler, M. H., & Schiff, L. R. (1998). Cooperative Knowledge Work and Practices of Trust: Sharing Environmental Planning Data Sets. In *Proceedings of the 1998 ACM Conference On Computer Supported Cooperative Work* (pp. 335–343). Seattle, Washington: ACM. doi:10.1145/289444.289508

Vardigan, M., Granda, P., Hansen, S. E., Ionescu, S., & LeClere, F. (2010). Documenting the Research Life Cycle: One Data Model, Many Products. *IFC Bulletin*, *33*, 91–104.

Vardigan, M., Heus, P., & Thomas, W. (2008). Data Documentation Initiative: Toward a Standard for the Social Sciences. *International Journal of Digital Curation*, *3*(1), 107–113. doi:10.2218/ijdc.v3i1.45

Wang, P. L., & White, M. D. (1999). A Cognitive Model of Document Use During a Research Project. Study II. Decisions at the Reading and Citing Stages. *Journal of the American Society for Information Science*, *50*(2), 98–114.

Wang, P., & Soergel, D. (1998). A Cognitive Model of Document Use During a Research Project. Study I. Document Selection. *Journal of the American Society for Information Science*, *49*(2), 115–133. doi:10.1002/(SICI)1097-4571(199802)49:2<115::AID-ASI3>3.0.CO;2-T

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33.

Wilkinson, M. (2010). *DataCite: The International Data Citation Initiative. Datasets programme* (No. 163). Working Paper Series des Rates für Sozial- und Wirtschaftsdaten. Retrieved from http://www.econstor.eu/handle/10419/43618

Wilson, P. (1983). *Second-Hand Knowledge: An Inquiry into Cognitive Authority*. Westport, CT: Greenwood Press.

Zimmerman, A. (2007). Not by Metadata Alone: The Use of Diverse Forms of Knowledge to Locate Data for Reuse. *International Journal on Digital Libraries*, *7*(1-2), 5–16. doi:10.1007/s00799-007-0015-8

Zimmerman, A. S. (2008). New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. *Science, Technology & Human Values*, *33*(5), 631–652. doi:10.1177/0162243907306704