Factors of Trust in Data Reuse

Ayoung Yoon, Yoo Young Lee

Abstract

Purpose: This study aims to quantitatively examine factors of trust in data reuse from the reusers' perspectives.

Design/methodology/approach: This study utilized a survey method to test the proposed hypotheses and to empirically evaluate the research model, which was developed to examine the relationship each factor of trust has with reusers' actual trust during data reuse.

Findings: This study found that the Data Producer (H1) and Data Quality (H3) were significant, as predicted, while Scholarly Community (H3) and Data Intermediary (H4) were not significantly related to reusers' trust in data.

Research limitations/implications: Further disciplinary specific examinations should be conducted to complement the study findings and fully generalize the study findings.

Practical implications: The study finding presents the need for engaging data producers in the process of data curation, preferably beginning in the early stages and encouraging them to work with curation professionals to ensure data management quality. The study finding also suggests the need for re-defining the boundaries of current curation work or collaborating with other professionals who can perform data quality assessment that is related to scientific and methodological rigor.

Originality/value: By analyzing theoretical concepts in empirical research and validating the factors of trust, this study fills this gap in the data reuse literature.

Keyword: data reuse, data curation, trust, data service, scholarly communication, research data

Introduction

The data curation community has been concerned about the issue of trust in data, most commonly in relation to curation activities performed by data repositories. Trustworthy Repositories Audit and Certification: Criteria and Checklist (TRAC)/ISO 16363 and Data Seal of Approval (DSA) are some well-known efforts to preserve and provide access to trusted content through repository certification. Since Prieto (2009) argued the need to understand trust from users' perspectives, several research studies have also demonstrated how a repository's intermediary role contributes to data reusers' trust (Donaldson and Conway, 2015; Frank *et al*., 2017; Yakel *et al*., 2013; Yoon, 2014a). While past studies underscore the significance of users' trust in repository and curation activities, they also present the need to investigate users' trust in the larger context of data reuse. The landscape of data sharing and reuse is very dynamic, and

This is the author's manuscript of the article published in final edited form as:
Yoon, A., & Lee, Y. Y. (2019). Factors of trust in data reuse. Online Information Review, 43(7), 1245–1262.
https://doi.org/10.1108/OIR-01-2019-0014

data exchange for reuse often takes place without any intermediary, such as in peer-to-peer exchanges.

Trust plays a fundamental role in data reuse. Past trust studies have demonstrated that trust mediates and enhances knowledge sharing (e.g., Ho *et al*., 2010; Renzl, 2008) because trust is fundamental in society and in human relationships (e.g., Gambetta, 1988; Weber *et al*., 2004). Van House (2002) discussed the role of trust in a repository context related to sharing knowledge and scholarship, but the role of trust is also notable outside of a repository context, where a less intermediary role is involved in data exchange. Data reuse involves various types of relationships and communication among different stakeholders, such as data producers, data curators, data reusers, and other scholarly communities. Yoon's (2017) study on data reusers' trust development presents this dynamic relationship with various stakeholders, as well as the social perceptions embedded in reusers' trust judgments.

There has been growing attention toward the concept of trust in data reuse, and past studies have explored specific aspects of data to be trusted—such as data integrity, quality, and provenance—both in and out of a repository context (Donaldson and Fear, 2011; Lemieux, 2014; Mayernik *et al*., 2008; Yoon 2016b). While these studies contribute to the understanding of the nature of trust, as well as trust factors, few studies solely focus on understanding and identifying factors of trust in data reuse. Recently Wolski, Howard, and Richardson (2017) discussed a trust framework for online data services, but the model was theoretical and without empirical support. Built on previous studies exploring different trust factors during data reuse, this study aims to quantitively examine factors of trust in data reuse from the reusers' perspectives. This study contributes to the field of trust research in data sharing, reuse, and curation by examining trust factors and providing implications to improve current data reuse and curation practices.

Literature review

Not many studies have formally defined the term *reus*e. van de Sandt, Dallmeier-Tiessen, Lavasa, & Petras (2019) argued that the term reuse is a complicated concept and no common definition is proposed across the disciplines yet. Despite the difficulties of proposing agreed definition across the disciplines, researchers generally understand it to indicate the use of data by someone who did not collect it. Therefore, reuse refers to a secondary use of data that is not defined by their original purpose but is intended to address new problems (Karasti & Baker, 2008; Zimmerman, 2008; Yoon, 2017). Broadly, reuse includes the reproduction or replication of prior study results as it contributes to the existing knowledge (King, 1995). Recently, the concept of repurposing has been added to the discussion of data reuse. In this context, data reuse has been defined as the use of data more than once for the same purpose, while data repurposing has been described as the use of data for a completely different purpose (Data Governance and Quality, 2012). Faniel and Jacobsen (2010) pointed out that the absence of a reuse definition causes major challenges in providing reusable data, even though other studies have demonstrated that data reuse can be beneficial to researchers.

Recent literature argued the key benefit of reusing data for the wider research communities (van de Sandt, et al, 2019). Birnholtz and Bietz (2003) and Borgman (2011) argued shared data can be used not only to validate existing results but also to generate new findings built on the work of others. Re-analyzed data or data combined with new data can also help to verify published results or arrive at new conclusions (National Academy of Science, 2009). Thus, research data must be available for use beyond the purposes for which they were initially collected to enable others to ask new questions of extant data, advance solutions for complex human problems and the state of science, reproduce research, and expand the instruments and products of research to new communities (Borgman, 2010; Borgman, 2011; Hey & Trefethen, 2003; Hey, Tansley, & Tolle, 2009).

Previous research has demonstrated the relationship between data curation and data reuse and has suggested that well-curated data is an integral part of data reuse. Coates (2014) argued that, because data are a key piece of the scholarly record, the management of data has an impact on the integrity of the scholarly record and on the potential for data sharing and reuse. Steinhart et al. (2008) argued that a well-developed data curation infrastructure, by exposing data for reuse, would enable new discoveries and ensure access to and preservation of scholarly outputs. The Digital Curation Centre (DCC) (n.d.) also argued that good practices of data curation can support data reuse in multiple ways; they ensure that the appropriate steps are taken to make data available in the first place (i.e., by presenting data and their associated descriptions in forms that are accessible and understandable to reusers); they prevent the unauthorized use of data (i.e., by maintaining legal constraints and usage rights); they provide the means of assuring data integrity and authenticity; and they enable reusers to be able to access high-quality data they can trust.

These previous literatures suggested that understanding and meeting reusers' needs and expectations is important to enhance data reusability because curators decide what information to collect, provide, and preserve based on reusers' needs and expectations. Trust is a useful concept to understand users' expectations and needs, as the concept of trust is woven into the lifecycle of data—from the creation, preparation, and management of data to their sharing and reuse to their preservation—and into the relations with parties involved in this lifecycle.

Due to the significance of data reuse in scholarly communication, research community and data curation practices, data reuse literature has been emerging recently. A number of studies have examined data reuse practices in various contexts, such as social science (Niu, 2009; Yoon, 2014b; Yoon and Kim, 2017), ecology (Zimmerman, 2008), archeology (Faniel *et al*., 2013), and science and engineering (Birnholtz and Bietz, 2003; Carlson *et al*., 2011; Carlson and Anderson, 2007; Kim and Yoon, 2017). Many studies have suggested that data reuse is not an easy, one-step process, and finding data that researchers can trust is the most important first step toward data reuse (Yoon, 2017).

Previous studies have also discussed several elements that possibly contribute to reusers' trust formation. One major source of trust is Communities of Practice (CoPs) (Van House *et al*., 1998). Defined as the "groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis" (Wenger *et al*., 2002, 4), CoPs help to share knowledge through trust-based relations, which a consensual knowledge base and shared identity enhance (Hislop, 2004). Different types of CoPs can exist, such as the data reuser group itself (e.g., within laboratories or disciplines). Whether they are physically connected or geographically distributed, CoPs share practices, experiences, understandings, technology, and languages. Data reusers form trust within their CoPs, and they can judge trust if data producers from their CoPs generated the data they intend to use (Van House *et al*., 1998).

Epistemic communities (ECs) are another major source of trust. An epistemic community is "a network of professionals with recognized expertise and competence in a particular domain [with] an authoritative claim" (Hass, 1992, 3). ECs differ from broader scientific communities that share a set of causal approaches and knowledge and are formed from their principled approaches to the issues at hand (e.g., economists as a disciplinary community vs. Keynesians as an epistemic community; Haas, 1992, 19). Trust is related to ECs' members taking strong or weak social views on the role of communities as holders of knowledge (Faulkner, 2010; Poutanen, 2001). In a data reuse context, ECs have assessment mechanisms and demonstrate competence, honesty, and shared understanding, which helps members decide who is, and what is, trustworthy (Jirotka *et al*., 2005; Van House, 2002). ECs are even more helpful when it is difficult to judge unknown data producers' skills based on the data itself, as first-hand knowledge of the skills or values of other researchers affects the assessment of trust and reuse decisions (Zimmerman, 2008).

In addition to seeing data producers as part of CoPs, data producers can be used to judge trustworthiness and make decisions regarding reuse. When ecologists assess the trustworthiness of data sources, their first focus is on several aspects of data producers, such as competence, commitment, and reputation, although they may not automatically accept data (Zimmerman, 2008). While information about the data producer cannot solely provide total trust of the data, the information helps lessen reusers' concerns about data quality (Zimmerman, 2008).

Trust also stems from factors in the data itself, such as collection methods, measurements, or variables. Wallis *et al*. (2007) found that habitat biologists asked how data-collection instruments were chosen and how data producers calibrated the instruments before reusing them. Zimmerman (2008) remarked that what is being observed is sometimes the primary source of trust, which is justified by how it is collected. Faniel and Jacobsen (2010) also found that earthquake data reusers' understanding of how data producers collected and measured data increased their trust in data reliability. Knowing how problems were resolved during the collection (or experiment) processes also helped reusers know that the data were properly processed (Faniel and Jacobsen, 2010).

Because assessing trust of data inevitably requires an in-depth understanding of their context (Jirotka *et al*., 2005), information about data, delivered by any means, is important. Information can be obtained through reusers' previous knowledge, their familiarity with artifacts and processes, their perceptions of the competence or honesty of data producers, or their direct interactions with colleagues, experts, or data producers (Birnholtz and Bietz, 2003; Van House, 2002). Documentation can also deliver information, which is probably the ideal situation because information contained within documentation is stable. Faniel and Jacobsen (2010) found that earthquake engineering data reusers developed their trust by reviewing documentation.

Reusers' trust can also be closely related to data repositories. Previous studies have reported that it is important for reusers to know how the data are processed (Carlson and Anderson, 2007; Yoon, 2014a). In Carlson and Anderson's (2007) work, reusers wanted to know how the data were "cooked," and they did not trust the data if they could not find out. Other organizational attributes, such as the integrity of repositories, transparency, reputation, and the structural assurance that guarantees preservation and sustainability also influence reusers' trust (Yakel *et al*., 2013; Yoon, 2014). Reusers' perceptions of the roles of repositories, even though not always correct, still influence their trust (Yoon, 2014).

Finally, reusers' personal knowledge, skills, and experiences play a role in trust assessments. Because reusers often work with data that they did not create, knowledge and skills from reusers' own data-collection experiences help them not only understand the data, but also judge the data's quality and trustworthiness (Borgman, 2007; Zimmerman, 2008). Data reusers judge the competence and commitment of data producers based on their own perceptions and personal knowledge, and the reusers' understanding of errors that potentially occur during data collection is key to their ability to judge data quality (Zimmerman, 2008).

This review of previous studies reveals trust's important role in data reuse. Reusers' trust, or distrust, of the data forms and eventually leads to data reuse behaviors (Yoon, 2017). Although the literature on trust in the data sharing and reuse context is continually growing, few attempts have been made to explain and validate the relationship between factors of trust and reusers' trust, moving from a theoretical understanding and using a quantitative approach. By analyzing theoretical concepts in empirical research and validating the factors of trust, this study fills this gap in the data reuse literature.

Research Model and Hypothesis Development

A research model was developed to examine the relationship each factor of trust has with reusers' actual trust during data reuse. The following section will discuss each research construct and its related hypothesis. From previous research, this study identified several trust factors at the following levels: data, data producer, community, and intermediary.

Data producer

Due to the data producers' roles in generating quality data, they can influence reusers' formation of trust in that data. Previous literature has suggested that data producers can be used to judge the trustworthiness of data (Zimmerman, 2018). Yoon (2015) further explored three dimensions of data producers: ability, ethics, and commitment. Data producers' ability refers to their research competence and expertise, such as (in this study) skills in quantitative methodology, which directly influence one data factor—scientific rigor. Ethics refers to data that comes from research conducted in an ethical manner—not just in compliance with relevant ethical regulations (e.g., IRB), but also with the data collection being driven by ethical motivations. Data producers' commitments made to the data during the process of data creation, preparation, and/or management also influence the development of reusers' trust.

H1. Data producers are positively related to data reusers' trust in data.

Scholarly community

Previous studies have suggested that scholarly community can be a major source of trust in data reuse and can help reusers' make trust judgments about the data (Jirotka *et al.*, 2005; Van House *et al.*, 1998; Van House, 2002; Zimmerman, 2008). Two sub-factors, social acknowledgment and community reliance, were identified in Yoon's (2015) study. Social acknowledgement, including both formal and informal acknowledgement, refers to positive social recognition of the data based on different types of peer evaluations. Community reliance refers to feelings of reliance on the data based on other people's experiences. Yoon (2015) argued that a shared experience with other data reusers is another useful source of reusers' trust.

H2. The scholarly community to which a data producer belongs is positively related to data reusers' trust in data.

Data quality

Data quality is another important factor influencing reusers' trust judgments. Many previous studies pointed out the importance of collection methods, measurements, or variables (Faniel and Jacobsen, 2010; Wallis *et al*., 2007). Several studies also distinguished factors related to the intrinsic quality of data (e.g., validity and scientific rigor), as well as the quality of the research product, such as the conditions of data to be reused, which are related to management and curation actions (Peer *et al*., 2014; Yoon, 2016). From Yoon's (2015) research, this study identified four sub-factors of data quality: scientific rigor, preparedness, comprehensiveness, and transparency. Scientific rigor (validity) refers to an objective quality of the data. Preparedness is reusers' perceptions about the degree to which the data have been accurately prepared and are ready to be reused, and comprehensiveness desires all aspects of the data to be understandable. Lastly, transparency means that information about the data has been documented or provided transparently.

H3. Data quality is positively related to data reusers' trust in data.

Data intermediary

Several pieces of research have discussed the role of data repositories in enhancing reusers' trust in data to support data reuse (Carlson and Anderson, 2007; Yoon, 2014a). Yoon (2015) pointed out the role of repositories in data preparation or curatorial activities and how the repository staff's professionalism, knowledge, expertise, and commitment regarding the data helped to build data reusers' trust in the data.

H4. Data intermediaries are positively related to data reusers' trust in data.

Trust

While trust is a complex concept that has been widely studied in various contexts, this study operationalized the definition of trust as the mental status of data reusers leading to the actions of reusing data (Yoon, 2017). Grounded in theories of trust, this study understands trust as both a psychological state and a behavioral indicator (e.g., Giffin, 1967; Good, 1988; Lewis and Weigert, 1985).

Figure 1. Research Model

Research Method

This study utilized a survey method to test the proposed hypotheses and to empirically evaluate the research model.

Sample
Our study aimed to address the actual data reusers to, in turn, accurately address the relationship between trust factors and reusers' trust. To identify researchers who have had experience reusing data, the project team utilized data citation tracking from the ProQuest database. Despite the recent development in data citation standards, data citation is not yet fully implemented as an academic practice, and it is also only applied to the most recent research (Altman and King, 2007; Gray *et al*., 2002; Mooney, 2011; Yoon, 2017). Tracking data citation may result in some limitations, such as the exclusion of research that does not indicate data reuse. However, it is still an effective method, as previous research has implemented it as a way of identifying data reusers (Faniel *et al*., 2015). We employed keywords—"data reuse" or "secondary data"—in our search, which, by themselves, resulted in 439,447 articles, respectively. To address researchers with recent data reuse experiences, we refined our search results to include only research published since 2010. We also excluded articles published outside of North America. We manually examined those articles to ensure the published research reused existing data and thus appropriate for our study context. After excluding irrelevant articles (e.g., articles that were not generated from the secondary data and mentioned the term "data reuse" or "secondary data" in other contexts), a total of 1,464 potential participants were identified for our study.

Trust construct development
Our proposed research model contained four trust factors with the ten sub-trust constructs discussed in the previous section. The survey items for measuring each construct were developed from previous studies, and each sub-construct had two to five measurement items. All items were collected from study participants and measured based on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The scales were refined and validated through the process of instrument development (e.g., subject matter experts' reviews for content validation, pre-tests to clarify questions, pilot-tests in Survey Monkey, a web-based survey administration platform, the determination of survey timing [approximately 15 minutes], and the selection of final survey items). See Table 1 for the measurement items for each trust construct.

Survey administration
An online survey was distributed to the potential participants through Survey Monkey using institutional accounts. The personalized initial invitation to the survey was sent to 1,464 researchers in November 2016. Each potential participant was asked to complete the survey based on a particular research article he or she had authored. Three reminders about participating in the survey were sent to the potential participants once every three-week period. The survey was closed in January 2017. We received 184 responses, which made an 8% response rate. We removed 16 survey responses in which respondents did not complete all survey items. We also performed casewise deletion due to missing values (> 5% per indicator), which led to the elimination of an additional 23 survey responses (Hair, 2017). Therefore, 145 responses were used for the final data analysis.

Data Analysis and Results

The partial least squares path modeling (PLS-PM) method was applied to test the hypotheses using the plspm R package, version 0.4.9 (Sanchez, 2013). The PLS-PM approach is a two-step process in the structural equation modeling (SEM) method. It allows the estimation of a factor model for constructing measurements, as well as relationships between constructs for structural models (Henseler *et al*., 2016). We chose the partial least squares path modeling (PLS-PM) as a statistical method for this study due to data characteristics and research model (Hair, 2017; Ravand & Baghaei, 2016). Since the PLS-PM method is variance-based approach to SEM (Structural Equation Modeling) while traditional SEM approaches are covariance-based, there are no distributional assumptions and no issue with small sample sizes (Chin, 2010). However, the PLS-PM method is well suited to perform both measurement and structural models and it is also considered as the primary approach to measure the formative construct in the measurement model (Hair, 2017). The reliability and validity of each construct were first verified in the measurement model, and then the structural model was conducted to examine this study's research hypotheses.

Participants

Participants were 145 (39.3% male, 55.2% female, and 5.5% unspecified) researchers whose publications revealed that they had some experience in data reuse. They reported that they frequently (41.4%) or occasionally (46.2%) reused other researchers' data. The majority of the participants (82.1%) primarily reused quantitative data sets. To obtain these data sets, they checked multiple places, but most of the participants obtained data directly from data producers (77.9%), followed by data repositories (42.8%).

Although we did not limit our search to certain disciplines, participants were mainly in the field of either health sciences (53.8%) or social sciences (38.6%). This may be due to the long history of having data reuse practice in the field of health and social science (Clubb, Austin, Geda, & Traugott, 1985). Also, as the previous literature suggested (van de Sandt, et al, 2019), because the term referred data reuse were varied across the disciplines, other disciplines may not use the term "data reuse" or "secondary data" even when reusing data. Most of them (80.7%) were Caucasian. Most had a Ph.D. degree (86.2%). More than half of them were tenured or tenured track faculty (70.3%) and were professors, from assistant to emeritus (70.3%). Their age varied from 25 to 65+.

Measurement model

The overall quality of the measurement model was checked first by measuring convergent validity with values of factor loadings (> 0.7) and average variance extracted (AVE) (> 0.5) (Ravand and Baghaei, 2016). Some items under each construct manifested low factor loadings; therefore, they were removed from the initial model. The eliminated variables were: ABILITY01, ABILITY02, ETHICS01, ETHICS02, COMMIT01, RAPPORT01, RAPPORT02, SOACK04, COMREL01, COMREL02, COMREL03, COMREL04, TRANS01, TRANS02, PREPAR01, PREPAR02, SCIRIG03. (See Table 1. Survey items and factor loadings.) The revised model, without these unqualified variables, was reexamined. The values of factor loadings for each item ranged from 0.707 (TRANS03) to 0.943 (REPOSI03), and AVEs for each construct ranged from 0.631 (Data) to 0.809 (Data Intermediary) (See Tables 2 and 3).

Table 1. Survey items and factor loadings.

Table 2. Loadings and cross-loadings for the measurement model.

Table 3. Unidimensionality and AVE.

The unidimensionality of the revised model was assessed through Cronbach's alpha (> 0.7) (Cortina, 1993), composite reliability with Dillon-Goldstein's rho (> 0.7) (Chin, 1998), first eigenvalue (> 1.0), and second eigenvalue (< 1.0) (Ravand and Baghaei, 2016). The results demonstrated that the values of Cronbach's alpha were between 0.702 (Data Producer) and 0.921 (Data Intermediary), and Dillon-Goldstein's rhos were from 0.835 (Data Producer) to 0.945 (Data Intermediary), which were above the acceptance values. In addition, the first eigenvalues of all constructs were significantly larger than one, while their second eigenvalues were smaller than one (Table 3).

To verify if there were any traitor indicators, the discriminant validity was inspected via cross-loadings and the Fornell-Larcker criterion, which compares the square root of the AVE values between latent variables (Hair, 2017). The given values of cross-loadings (bolded in Table 2) in each construct proved that they were greater than any other loadings across the row. The square root of AVEs for each construct was also greater than any other inter-construct correlations. (See Table 4.). Therefore, we can conclude that both the reliability and the validity of the measurement model are acceptable.

Table 4. Correlation matrix, square roots of AVEs.

Structural model

In the next step, the structural model was examined to answer our research questions. We first evaluated the results of the regression equations to identify which factors influenced data reusers' trust. Then, the overall quality of the structural model was assessed using the R^2 determination coefficient. The results highlighted that Data Producers and Data Quality positively influenced reusers' Trust in data that they reused, while effects of Scholarly Community and Data Intermediary on Trust were not statistically significant. (See Figure 2). The Data Producers' ability, ethics, and commitment positively influenced data reusers' Trust ($\beta = 0.1918$, $p < 0.05$), and Data (e.g., transparency, preparedness, comprehensiveness, and scientific rigor) had a strong, positive relationship with Trust ($\beta = 0.5065$, $p < 0.001$). However, the Community to which researchers belonged ($\beta = 0.0388$, $p > 0.05$) and the Data Intermediary, such as data repositories ($\beta = 0.0615$, $p > 0.05$), did not have a significant effect on Trust. In general, this model accounts for 52% of the variance of Trust by its independent latent variables ($R^2 = 0.52$). (See Table 5 for the summary of hypothesis testing results).

Figure 2. Hypotheses testing results.

Table 5. Summary of hypothesis testing results.

Unlike traditional modeling techniques, there are no distributional assumptions and no sample size limitations in the PLS-PM approach; therefore, re-sampling procedures, like bootstrapping, are recommended to assess the variability of the parameter estimates (Chin, 2010; Sanchez, 2013). In this study, the bootstrapping method with the 5,000 re-sample was applied to estimate the significance of path coefficients (Hair, 2017). The results (Table 6) confirmed that the coefficient values were significantly different from zero at the 5% confidence interval.

Table 6. Path coefficient and bootstrap standard errors.

Discussion

The major objective of this study was to investigate the factors that influence data reusers' trust in data and to empirically test whether the trust factors identified in previous studies were positively associated with data reusers' trust. We found that the Data Producer (H1) and Data Quality (H3) were significant, as predicted.

The role of data producers in forming reusers' trust is important in this study in two ways: 1) their ability and ethics contribute to producing quality research products (data), and 2) thus, their role is related to data quality, particularly the intrinsic quality of data (e.g., scientific rigor). As the quality of research cannot be separated from quality production, it is not surprising that data producers are positively related to reusers' trust. Another contribution of data producers in reusers' trust is their commitment, not only to data creation, but also to data preparation and management, which are important for data sharing. Previous studies have already argued for the importance of the roles of data producers and their contributions to data curation for supporting reuse (e.g., Lyon, 2007; National Science Board, 2005). In the data curation lifecycle, the roles of data producers are essential for ensuring further curation activities performed by data curators; the activities that should be performed by data producers are the starting point of curation activities in the DCC curation lifecycle model through the "Conceptualize" phase, although the model does not visibly specify the engagement of the original investigators. This study adds additional importance to producers' roles, as the results have shown that their work directly impacted reusers' trust.

As many previous studies argued for the importance of data quality, data quality was also positively related to reusers' trust in this study. While the intrinsic quality of data (e.g., scientific rigor associated with good quality research) was one sub-construct of data quality in this study, the research also suggests other quality dimensions related to data management—such as preparedness, comprehensiveness, and transparency—are also directly related to reusers' trust in the data with which they work. These are not entirely new considerations in data management, and the data management community has been creating best practices for data documentation, file organization and naming, ethical guidelines for preparing datasets (e.g., ICPSR, 2009; UK Data Archive, 2011), and tools for supporting data management (e.g., DMP tools). However, the depth of implementation of these best practices in real settings may vary.

This study also found some contradictory results to previous studies regarding Scholarly Community and Data Intermediary, as both constructs were not significantly related to reusers' trust in data. Despite the results, it is too early to declare that there is no relationship between these two factors and reusers' trust, as the participants of this study were heavily from the social and health science fields, and the sample size was relatively small for fully generalizing the findings. However, these findings do allow this study to say that external factors are less important in determining trust in data reuse. Perhaps, the results indicate data reusers' own assessment of data is more important than external views, such as the scholarly community's evaluation and overall reputation. Further, previous qualitative study reported that sometimes data reusers find problems and errors, even in well-known and reputable data (Yoon, 2017), which may influence the role of external evaluation on data in this survey.

While several previous studies argued the role of data repositories as the major source for users' trust (Bak, 2015; Frank *et al*., 2017; Yakel *et al*., 2013; Yoon, 2014a), Data Intermediary was not a factor influencing reusers' trust in this study. One possible explanation is that only about 40% of survey participants had experience reusing data gained from repositories, which could have influenced the survey result. In addition, it is also known that some repository functions—mostly related to their contribution to data curation, such as data cleaning, preparation, and packaging—are not always recognized by repository users (Yoon, 2014a). Further, Frank *et al*. (2017) suggested that deeper interaction with repositories, and internal loci of relationships to repositories, can help reusers build their trust. As this study did not have information about the nature of relationships between study participants and the repositories they used, this should be further investigated to statistically claim the relationship between data repositories and reusers' trust.

This study's findings suggest several practical implications. First, the research presents the need for engaging data producers in the process of data curation, preferably beginning in the early stages and encouraging them to work with curation professionals to ensure data management quality. The survey results indicate that many data exchanges still happen at the personal level, between peers. Reusers' experiences working with data directly from data producers (whether individual researchers or research institutions) may be different than working with data from data repositories, where professionals perform intermediary roles and have expertise in data curation. In circumstances where data sharing and reuse are conducted among individuals without intermediary parties, data producers' understanding and experiences of data management are significant for ensuring the reusers' trust in the data. Even when data producers deposit their data in either domain or institutional repositories, their initial management practices greatly influence the intermediaries' curatorial work. Growing numbers of libraries and repositories provide educational sessions for researchers to teach data management planning and practices, but many of them focus on data management planning as part of funders' requirements—a one-time interaction rather than on-going, shared management efforts.

While scientific rigor is a part of defining data quality in this study and also supports reusers' trust in data, whether rigor of research and data can be or should be curated is controversial, as scientific and methodological rigor in research is mostly the responsibility of data producers. Repositories' data assessment usually regards usability (e.g., file organization, documentation) and long-term preservation (e.g., file format), particularly because the research quality assessment requires a high level of domain expertise concerning research methods, measurements, experiments, and potential novelties and impacts that directly relate to content. If the assessment of intrinsic data quality does not belong within the domain of traditional curation, perhaps it is worth thinking about how this assessment should be integrated into data curation, either by re-defining the boundaries of curation or by collaborating with other professionals who can perform this role. Because good data curation practices encourage the involvement of, and collaboration with, different parties, the roles and responsibilities of data curation do not belong solely to data curators or data producers, but to many other professionals as well. Emerging discussion and practice of data peer review may suggest one possible method for building reusers' trust in data by sharing the responsibility of curating the intrinsic quality of data. Mayernik et al. (2014) argued that it is essential to divide data and separately review two portions of it: the technical (e.g., metadata, documentation, file formats), reviewed by data curators, and the scientific (e.g., appropriate collection methods, validity, reliability), reviewed by experts in the scientific community. They asserted that such a division is necessary due to the different forms of expertise required by each data portion. This division of responsibilities would be an initial step for data to be properly managed by data producers and to be trusted by potential reusers.

Conclusion

This study empirically examined the factors of trust in data reuse with practical implications. While increasing numbers of funders require data management and sharing, if reusers do not trust data, these data will be dead by sitting on the servers. Although, in many cases, data sharing and reuse happen through peer-to-peer exchange, the number of reuse cases through intermediaries, such as libraries and repositories, will increase. Supporting data reuse is critical for promoting more data reuse and facilitating the process. From data reusers' experiences, this study's findings provide several implications for moving forward to integrate reusers' trust into data curation and management practices.

While this study contributes to the understanding of reusers' perspectives on data and trust, it also has some limitations. As already noted, this study utilized a relatively small sample size. This sample size was methodologically acceptable but, given the diverse nature of data reuse in various disciplinary specific contexts, the findings may not be fully generalizable. Also the participants of this study were dominantly from the social and health sciences, which are known to be unique in their data reuse culture (Curty *et al*., 2016). Disciplinary specific examinations should be conducted to complement the study findings.

In addition, while this study solely focused on defining and examining trust constructs in data reuse, data reuse itself is complex and influenced by many other factors. Having an extended model of data reuse, combining a trust factor with other individual and institutional factors that have been previously examined (e.g., social norms, reuser concerns, efforts to reuse, etc.), will provide a more fruitful understanding of data reuse.

Acknowledgement
TBA

References

Altman, M., and King, G. (2007), “A proposed standard for the scholarly citation of quantitative data”, *D-Lib Magazine* Vol. 13 No. 3/4 available at: http://www.dlib.org/dlib/ march07/altman/03altman.html (accessed 11 November 2018).

Bak, G. (2015), “Trusted by whom? TDRs, standards culture and the nature of trust”, *Archival Science*, Vol. 16 No. 4, pp. 373-402, available at: https://doi.org/10.1007/s10502-015-9257-1 (accessed 23 September 2018).

Birnholtz, J., and Bietz, M. (2003), “Data at work: Supporting sharing in science and engineering”, in Pendergast, M., Schmidt, K., Simone, C., and Tremaine, M. (Eds.) *GROUP'03: Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work*, ACM SIGGROUP: Sanibel Island, FL, pp. 339-348.

Borgman, C. (2007), *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*, MIT Press: Cambridge, MA.

Borgman, C. (2010), "Research data: Who will share what, with whom, when, and why?" Presented at the *China-North America Library Conference*, Beijing. available at: http://works.bepress.com/borgman/238 (accessed 11 November 2018).

Borgman, C. (2011), The conundrum of sharing research data. *SSRN eLibrary*. available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155 (accessed 11 November 2018)

Carlson, J., Fosmire, M., Miller, C. C., and Nelson, M. S. (2011), “Determining data information literacy needs: A study of students and research faculty”, *Portal: Libraries and the Academy* Vol. 11 No. 2, pp. 629-657, available at: https://doi.org/10.1353/ pla.2011.0022 (accessed 11 November 2018).

Carlson, S., and Anderson, B. (2007), “What are data? The many kinds of data and their implications for data re-use”, *Journal of Computer-Mediated Communication* Vol. 12 No. 2, pp. 635-651, available at: http://jcmc.indiana.edu/vol12/ issue2/carlson.html (accessed 11 November 2018).

Chin, W. (1998), “The partial least squares approach for structural equation modeling”, in Marcoulides, G. A. (Ed.), *Methodology for Business and Management: Modern Methods for Business Research*, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 295-336.

Chin, W. (2010), “Bootstrap cross-validation indices for PLS path model assessment”, in edited Vinzi, V., Chin, W., Henseler, J., and Wang, H. (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications*, Springer Berlin Heidelberg: Berlin, pp. 83-97, available at: https://doi.org/10.1007/978-3-540-32827-8_4.

Clubb, J. M., Austin, E. W., Geda, C. L., & Traugott, M. W. (1985), "Sharing research data in the social science." In S.E. Fienberg et al. (Eds.), *Commission on Behavioral and Social Sciences and Education* (pp. 39-88). Washington, D.C.: National Academy Press.

Cortina, J. M. (1993), “What is coefficient alpha? An examination of theory and applications”, *Journal of Applied Psychology* Vol. 78 No. 1, pp. 98-104, available at: https://doi.org/10.1037/0021-9010.78.1.98 (accessed 25 October 2018).

Curty, R., Yoon, A., Jeng, W., and Qin, J. (2016), “Untangling data sharing and reuse in social sciences”, *Proceedings of the Association for Information Science and Technology* Vol. 53 No. 1, pp. 1-5, available at: https://doi.org/10.1002/ pra2.2016.14505301025 (accessed 23 September 2018).

Data Governance and Quality: Data Reuse vs. Data Repurposing (2012). available at: http://dataqualitybook.com/?p=349#more-349 (accessed 23 September 2018).

Donaldson, D., and Fear, K. (2011), “Provenance, end-user trust, and reuse: An empirical investigation”, paper presented at the 3rd USENIX Workshop on the Theory and Practice of Provenance (TaPP’2011), June, Heraklion, Crete, Greece, available at: http://www.usenix.org/event/tapp11/tech/final_files/Donaldson.pdf (accessed 11 November 2018).

Donaldson, D., and Conway, P. (2015), “User conceptions of trustworthiness for digital archival documents”, *Journal of the Association for Information Science and Technology* Vol. 66 No. 12, pp. 2427–2444, available at: https://doi.org/ 10.1002/asi.23330 (accessed 23 September 2018).

Faniel, I., Kriesberg, A., and Yakel, E. (2015), “Social scientists’ satisfaction with data reuse”, *Journal of the Association for Information Science and Technology* Vol. 67 No. 6, pp. 1404-1416, available at: https://doi.org/10.1002/asi.23480 (accessed 11 November 2018).

Faniel, I., and Jacobsen. T. E. (2010), “Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues’ data”, *Computer Supported Cooperative Work (CSCW)* Vol. 19 No. 3/4, pp. 355-375, available at: https://doi.org/10.1007/s10606-010-9117-8 (accessed 11 November 2018).

Faniel, I., Kansa, E., Whitcher Kansa, S., Barrera-Gomez, J., and 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*, ACM: New York, pp. 295-304, available at: https://doi.org/10.1145/ 2467696.2467712 (accessed 11 November 2018).

Frank, R. D., Chen, Z., Crawford, E., Suzuka, K., and Yakel, E. (2017), “Trust in qualitative data repositories”, *Proceedings of the Association for Information Science and Technology* Vol. 54 No. 1, pp. 102-111, available at: https://doi.org/10.1002/ pra2.2017.14505401012 (accessed 11 November 2018).

Gambetta, D. (1988), “Can we trust trust?” in Diego, G. (Ed.), *Trust: Making and Breaking Cooperative Relations*, Basil Blackwell: Oxford, UK, pp. 213- 237, available at: http://citeseerx.ist.psu.edu/viewdoc/ summary?doi=10.1.1.24.5695 (accessed 23 September 2018).

Gray, J., Szalay, A. S., Thakar, A. R., Stoughton, C., and vandenBerg, J. (2002), “Online scientific data curation, publication, and archiving”, *Proceedings of SPIE* 4846, pp. 103-107, available at: doi:10.1117/12.461524 (accessed November 11, 2018).

Hair, J. F. ed. (2017), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd ed., Sage: Los Angeles.

Henseler, J., Hubona, G., and Ray, P. A. (2016), “Using PLS path modeling in new technology research: Updated guidelines”, *Industrial Management & Data Systems* Vol. 116 No. 1, pp. 2-20, available at: https://doi.org/10.1108/IMDS-09-2015-0382 (accessed 26 October 2018).

Hey, T., Tansely, S. & Tolle, K. (Eds.). (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft. available at: http://research.microsoft.com/en-us/collaboration/fourthparadigm/4thparadigm_science.pdf. (accessed November 11, 2018)

Hey, T., Trefethen, A. (2003). "The data deluge: An e-science perspective." In F. Berman, G.C. Fox, & T. Hey, (Eds.), *Grid computing: Making the global infrastructure a reality*. New York, NY: Wiley.

Ho, L. A., Kuo, K. T., Lin, C., and Lin, B. (2010), “The mediate effect of trust on organizational online knowledge sharing: An empirical study”, *International Journal of Information Technology & Decision Making* Vol. 9 No. 4, pp. 625-44, available at: doi:10.1142/S0219622010003981 (accessed 23 September 2018).

Inter-University Consortium for Political and Social Research (ICPSR) (2009), “Principles and good practice for preserving data”, IHSN Working Paper No. 003, ICPSR, available at: http://www.ihsn.org/home/sites/default/files/resources/IHSN-WP003.pdf.

Jirotka, M., Procter, R., Hartswood, M., Slack, R., Simpson, A., Coopmans, C., Hinds, C., and Voss, A. (2005), “Collaboration and trust in healthcare innovation: The eDiaMoND case study”, *Computer Supported Cooperative Work* Vol. 14, pp. 369-398.

Karasti, H., & Baker, K. S. (2008), "Digital data practices and the long term ecological research program growing global", *International Journal of Digital Curation*, *3*(2), 42–58. doi:10.2218/ijdc.v3i2.57

King, G. (1995). "Replication, Replication", *PS: Political Science and Politics, 28*(3), 444–452. doi:10.2307/420301

Kim, Y., and Yoon, A. (2017), "Scientists' data reuse behaviors: A multilevel analysis", *Journal of the Association for Information Science and Technology* Vol. 68 No. 12, pp. 2709-2719, available at: https://doi.org/10.1002/asi.23892 (accessed 11 November 2018).

Lemieux, V. (2014, October 14), "Why we're failing to get the most out of open data", GE Reports, available at: http://www.ideaslaboratory.com/post/99984665378/why-were-failing-to-get-the-most-out-of-open-data (accessed 26 February 2015).

Lyon, L. (2007), "Dealing with data: Roles, rights, responsibilities, and relationships", UKOLN, University of Bath: Bath, available at: http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/dealingwithdata#downloads (accessed 11 November 2018).

Mayernik, M. S., Callaghan, S., Leigh, R., Tedds, J., and Worley, S. (2015), "Peer review of datasets: When, why, and how", *Bulletin of American Meteorological Society* Vol. 96 No. 2, pp. 191-201, available at: http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-13-00083.1 (accessed 11 November 2018).

Mayernik, M. S., Wallis, J. C., Pepe, A., and Borgman, C. L. (2008), "Whose data do you trust? Integrity issues in the preservation of scientific data", paper presented at the IConference, February 28-March 1, Los Angeles, CA.

Mooney, H. (2011), "Citing data sources in the social sciences: Do authors do it?" *Learned Publishing* Vol. 24 No. 2, pp. 99-108, available at: doi:10.1087/20110204 (accessed 11 November 2018).

National Academy of Science. (2009). *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age.* Washington, DC: NAS. available at: http://www.nap.edu/catalog.php?record_id=12615 (accessed 11 November 2018).

National Science Board. (2005), "Long-lived digital data collections enabling research and education in the 21st century", No. NSB-05-40, National Science Foundation, available at: http://www.nsf.gov/pubs/2005/nsb0540/ (accessed 11 November 2018).

Niu, J. (2009), "Overcoming inadequate documentation", paper presented at the Annual Meeting of the American Society for Information Science & Technology (ASIS&T), November, Vancouver, British Columbia, Canada.

Peer, L., Green, A., and Stephenson, E. (2014), "Committing to data quality review", *International Journal of Digital Curation* Vol. 9 No. 1, pp. 263-291, available at: https://doi.org/10.2218/ijdc.v9i1.317 (accessed 11 November 2018).

Prieto, A. G. (2009), “From conceptual to perceptual reality: Trust in digital repositories”, *Library Review* Vol. 58 No. 8, pp. 593-606, available at: https://doi.org/10.1108/00242530910987082 (accessed 11 November 2018).

Ravand, H., and Baghaei, P. (2016), “Partial least squares structural equation modeling with R”, *Practical Assessment Research & Evaluation* Vol. 21 No. 11, available at: http://pareonline.net/getvn.asp?v=21&n=11.

Renzl, B. (2008), “Trust in management and knowledge sharing: The mediating effects of fear and knowledge documentation”, *Omega* Vol. 36 No. 2, pp. 206-220.

Sanchez, G. (2013), “PLS path modeling with R,” Trowchez Editions: Berkeley, available at: https://www.gastonsanchez.com/PLS_Path_Modeling_with_R.pdf (accessed 18 September 2017).

UK Data Archive. (2011), “Data management recommendations: For research centres and programmes”, UK Data Archive, available at: http://www.data-archive.ac.uk/media/257765/ukda_datamanagementrecommendations_centresprogrammes.pdf.

van de Sandt, S., Dallmeier-Tiessen, S., Lavasa, A., & Petras, V. (2019). "The definitions of reuse", *Data Science Journal* Vol. 18 No. 1, DOI: http://doi.org/10.5334/dsj-2019-022

Van House, N. A. (2002), “Digital libraries and practices of trust: Networked biodiversity information”, *Social Epistemology* Vol. 16 No. 1, pp. 99-114.

Van House, N. A., Butler, M. H., and Schiff, L. R. (1998), “Cooperative knowledge work and practices of trust: Sharing environmental planning data sets”, in *Proceedings of the ACM Conference On Computer Supported Cooperative Work*, A \CM: Seattle, pp. 335-343.

Wallis, J. C., Borgman, C. L., Mayernik, M. S., Pepe, A., Ramanathan, N., and Hansen, M. (2007), “Know thy sensor: Trust, data quality, and data integrity in scientific digital libraries”, paper presented at the European Conference on Research and Advanced Technology for Digital Libraries, September, Budapest, Hungary.

Weber, J. M., Malhotra, D., and Murnighan, J. K. (2004), “Normal acts of irrational trust: Motivated attributions and the trust development process”, *Research in Organizational Behavior* Vol. 26, pp. 75-101, available at: https://doi.org/10.1016/S0191-3085(04)26003-8 (accessed 23 September 2018).

Wenger, E., McDermott, R. A., and Snyder, W. (2002), *Cultivating Communities of Practice*. Harvard Business Press: Brighton, MA.

Wolski, M., Howard, L., and Richardson, J. (2017), “A trust framework for online research data services”, *Publications* Vol. 5 No. 2, p. 14, available at: https://doi.org/10.3390/publications5020014 (accessed 11 November 2018).

Yakel, E., Faniel, I., Kriesberg, A., and Yoon, A. (2013), “Trust in digital repositories”, *International Journal of Digital Curation* Vol. 8 No. 1, pp. 143-156, available at: https://doi.org/10.2218/ijdc.v8i1.251 (accessed 11 November 2018).

Yoon, A. (2014a), “End users’ trust in data repositories: Definition and influences on trust development”, *Archival Science* Vol. 14 No. 1, pp. 17-34, available at: https://doi.org/10.1007/s10502-013-9207-8 (accessed 11 November 2018).

Yoon, A. (2014b), “‘Making a square fit into a circle’: Researchers’ experiences reusing qualitative data”, *Proceedings of the American Society for Information Science and Technology* Vol. 51 No. 1, pp. 1-4, available at: https://doi.org/10.1002/meet.2014.14505101140 (accessed 11 November 2018).

Yoon, A. (2016), “Visible evidence of invisible quality dimensions and the role of data management”, paper presented at *i*Conference 2016, Philadelphia, PA, March 2016, available at: https://doi.org/10.9776/16123 (accessed 11 November 2018).

Yoon, A. (2017), “Data reusers’ trust development”, *Journal of the Association for Information Science and Technology* Vol. 64 No. 8, pp. 946-956, available at: https://doi.org/10.1002/asi.23730 (accessed 11 November 2018).

Yoon, A., and Kim, Y. (2017), “Social scientists’ data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories”, *Library & Information Science Research* Vol. 39 No. 3, pp. 224-233, available at: https://doi.org/10.1016/j.lisr.2017.07.008 (accessed 11 November 2018).

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* Vol. 33 No. 5, pp. 631-652, available at: https://doi.org/10.1177/0162243907306704.

Table 1. Survey items and factor loadings.

Construct	Sub-construct	Factor loadings	Survey items
Data producer	Ability	Eliminated	The producers of the data are the experts in the domain of this research (that generated the data). (ABILITY01)
		Eliminated	The producers of the data are knowledgeable in the methodology used to produce the data. (ABILITY02)
		0.801	The producers of the data are capable to produce good quality of data. (ABILITY03)
	Ethics	Eliminated	The dataset that I reused is compliant with ethical regulation (e.g., IRB). (ETHICS01)
		Eliminated	The data producers have moral motivations to conduct research. (ETHICS02)
		0.761	The data have been created in an honest manner, which lacked deception and distortion. (ETHICS03)
	Commitment	Eliminated	The producers of data spend a lot of time preparing data to make them available to others. (COMMIT01)
		0.701	The producers of data are committed to data creation, preparation, management, and dissemination. (COMMIT02)
	Rapport	Eliminated	I have a close relationship with the data producers. (RAPPORT01)
		Eliminated	I have known the data producers in-person for a long time. (RAPPORT02)
Scholarly community	Social acknowledgement	0.765	The dataset that I reused is well known in the area of research. (SOACK01)
		0.857	The dataset that I reused has a good reputation. (SOACK02)
		0.763	The dataset that I reused is from researchers (or organizations) with a good reputation. (SOACK03)
		Eliminated	The dataset that I reused is from funded research or published journals. (SOACK04)
	Community reliance	Eliminated	The data are used a lot by many other researchers. (COMREL01)
		Eliminated	The data are from a reuser group (e.g., list serve, workshop, conference meeting). (COMREL02)
		Eliminated	I have interacted with other researchers regarding my data reuse experiences. (COMREL03)
		Eliminated	There are people who help and support my data reuse process. (COMREL04)
Data quality	Transparency	Eliminated	The dataset that I reused includes information or documentation about history of the data

			(e.g., collection, manipulation, and management process). (TRANS01)
		Eliminated	The dataset that I reused includes information or documentation about errors about the data. (TRANS02)
		0.711	The dataset that I reused is transparent in terms of the collection, manipulation, and management process. (TRANS03)
	Preparedness	Eliminated	The dataset that I reused is well organized. (PREPAR01)
		Eliminated	The dataset that I reused is easy to understand and process. (PREPAR02)
		0.797	The dataset that I reused includes accurate information about the data without careless errors. (PREPAR03)
		0.758	The dataset that I reused is well prepared for other people to reuse. (PREPAR04)
	Comprehensiveness	0.791	The dataset that I reused provides all information that I need to reuse the data. (COMPRE01)
		0.854	The dataset that I reused provides comprehensive and good quality documentation. (COMPRE02)
	Scientific rigor	0.774	The data that I reused are accurate. (SCIRIG01)
		0.787	The data that I reused use adequate measures, methodology or study design. (SCIRIG02)
		Eliminated	The data that I reused have good validity. (SCIRIG03)
Data intermediary	Data repositories	0.796	The data are prepared by data professionals beyond the data producers (e.g., repository staff, data managers, etc.) (REPOSI01)
		0.909	The data professionals are knowledgeable and have expertise regarding the data. (REPOSI02)
		0.943	The data curated by the professionals are well managed. (REPOSI03)
		0.941	The data curated by the professionals are of good quality. (REPOSI04)

Table 2. Loadings and cross-loadings for the measurement model.

		Data producer	Scholarly community	Data quality	Data intermediary	Trust
Data Producer	ABILITY03	**0.855**	0.463	0.636	0.517	0.567
	ETHICS03	**0.778**	0.402	0.371	0.322	0.413
	COMMIT02	**0.734**	0.476	0.587	0.462	0.400
Scholarly Community	SOACK01	0.350	**0.795**	0.430	0.558	0.267
	SOACK02	0.499	**0.904**	0.465	0.619	0.392

	SOACK03	0.535	**0.836**	0.494	0.493	0.473
Data Quality	TRANS03	0.521	0.480	**0.707**	0.500	0.533
	PREPAR03	0.515	0.430	**0.814**	0.641	0.524
	PREPAR04	0.462	0.460	**0.753**	0.535	0.435
	COMPRE01	0.527	0.447	**0.808**	0.469	0.513
	COMPRE02	0.585	0.480	**0.875**	0.574	0.567
	SCIRIG01	0.601	0.383	**0.792**	0.596	0.653
	SCIRIG02	0.546	0.407	**0.804**	0.491	0.617
Data Intermediary	REPOSI01	0.323	0.500	0.431	**0.796**	0.328
	REPOSI02	0.507	0.519	0.602	**0.909**	0.509
	REPOSI03	0.540	0.630	0.685	**0.943**	0.500
	REPOSI04	0.576	0.670	0.697	**0.941**	0.561
Trust	TRUST	0.593	0.466	0.701	0.540	**1.000**

Table 3. Unidimensionality and AVE.

Latent variable	MVs	Cronbach's alpha	Dillon-Goldstein's rho	Eig. 1st	Eig. 2nd	AVE
Data Producer	3	0.702	0.835	1.880	0.612	0.625
Scholarly Community	3	0.809	0.889	2.180	0.621	0.716
Data	7	0.902	0.923	4.430	0.751	0.631
Data Intermediary	4	0.921	0.945	3.240	0.419	0.809
Trust	1	1.000	1.000	1.000	0.000	1.000

Table 4. Correlation matrix, square roots of AVEs.

	Data producer	Scholarly community	Data quality	Data intermediary	Trust
Data Producer	**0.791**				
Scholarly Community	0.563	**0.846**			
Data Quality	0.681	0.552	**0.794**		
Data Intermediary	0.556	0.649	0.685	**0.899**	
Trust	0.593	0.466	0.701	0.540	**1.000**

Table 5. Summary of hypothesis testing results.

Hs	Statements	Results	Beta (p)
H1	Data producers are positively related to data reusers' trust in data.	Supported	0.1918 *
H2	The scholarly community to which a data producer belong is positively related to data reusers' trust in data.	NOT supported	0.0388
H3	Data quality is positively related to data reusers' trust in data.	Supported	0.5065 ***
H4	Data intermediaries are positively related to data reusers' trust in data.	NOT supported	0.0615

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$

Table 6. Path coefficient and bootstrap standard errors.

	Original	Mean. Boot	Std. Error	perc. 0.025	perc. 0.975
Data Producer → Trust	0.1918	0.1892	0.0906	0.00828	0.362
Scholarly Community → Trust	0.0388	0.0413	0.0773	-0.10487	0.201
Data Quality → Trust	0.5065	0.5107	0.0992	0.31641	0.705
Data Intermediary → Trust	0.0615	0.0602	0.0978	-0.13697	0.249

Figure 1. Research Model

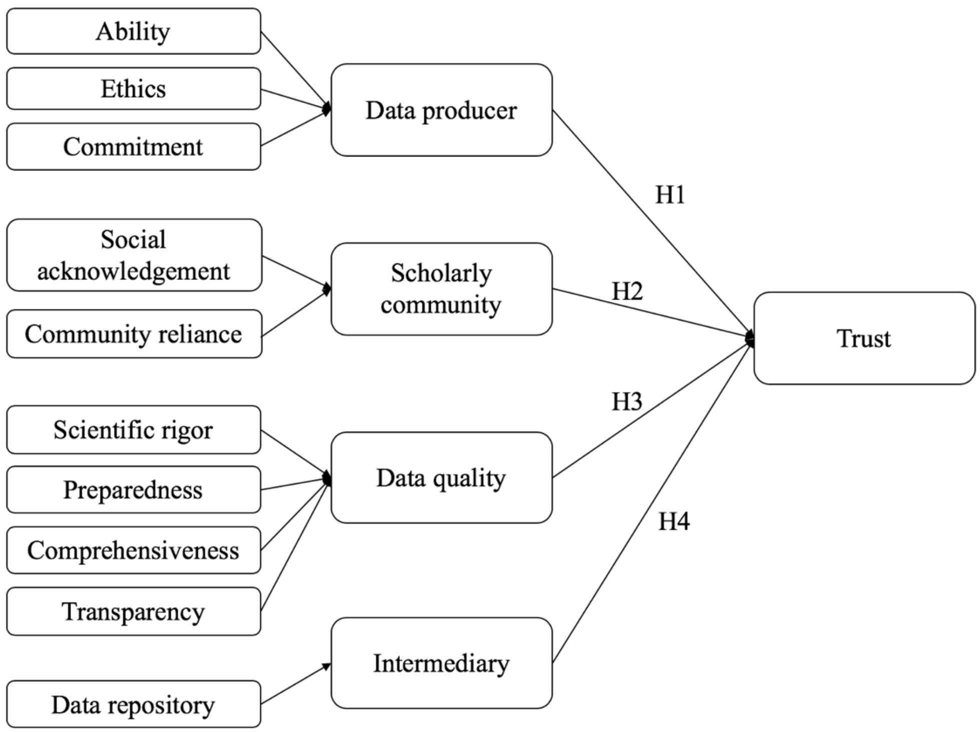

Figure 2. Hypotheses testing results.