Arch Sci (2014) 14:17–34
DOI 10.1007/s10502-013-9207-8

ORIGINAL PAPER

End users' trust in data repositories: definition and influences on trust development

Ayoung Yoon

Published online: 10 July 2013

Abstract While repositories' efforts to build trustworthy digital repositories (TDRs) led to the establishment of ISO standards, much less research has been done regarding the user's side, despite calls for an understanding of users' trust of TDRs. In order to learn about users' perspectives on trust in digital repositories, the present study investigated users' definitions of trust and factors that influence users' trust development, particularly addressing the users of three data repositories in the United States. A total of 19 participants were interviewed in this study. The results of this study indicate that users' definition of *trust* is largely based on a lack of deception, when it comes down to the specific context of data repositories. Regarding factors influencing the development of users' trust in repositories, organizational attributes, user communities (recommendations and frequent use), past experiences, repository processes (documentation, data cleaning, and quality checking), and users' perception of the repository roles were identified.

Keywords Trust · Data repository · Trusted digital repository

Introduction

Historically, cultural institutions that have been responsible for preserving paper records and physical artifacts have already developed considerable trust within the communities they serve. Libraries, archives, and repositories are trusted to store materials valuable for cultural and scholarly purposes and provide access to them to disseminate knowledge and to preserve them for future generations (Research Libraries Group/Online Computing Library Center [RLG/OCLC] 2002). The flood

A. Yoon (✉)
School of Information and Library Science, University of North Carolina at Chapel Hill, 216 Lenoir Drive CB #3360, 100 Manning Hall, Chapel Hill, NC 27599-3360, USA
e-mail: ayhounet@gmail.com; ayyoon@email.unc.edu

of digital information, however, has created new challenges in curating and archiving information. Whereas physical objects or documented reliable surrogates are available to patrons as "proof" of an institution's capability to collect and preserve for the long term, digital information is less tangible and much more mutable than other materials, and trust and reliability are considered more difficult to establish (RLG/OCLC 2002, p. 8).

Thus, new questions and new solutions in the field of archives and preservation are raised concerning whether and how the accumulated trust derived from traditional services can be transferred to the repositories of digital information (Jantz and Giarlo 2007). Ross and McHugh (2005) noted that digital information holders or service providers might already be regarded as trustworthy based on their reputations earned in the paper-based information environment. Institutions are likely to retain at least some trust from the public based on past successes (RLG/OCLC 2002). Others have made the point that digital resources are much more vulnerable than traditional paper-based information; this fact makes people and organizations insecure about digital information usage and ways of guaranteeing the authenticity and longevity of digital objects in institutions' collections (Electronic Resource Preservation and Access Network (ERPANET) 2004).

Whether or not trust derived from traditional services can be transferred to digital repositories, the concept of trust remains central in digital environments. The Commission on Preservation and Access/Research Libraries Group (CPA/RLG) Task Force on Archiving of Digital Information (1996) pointed out, "for assuring the longevity of information, the most important role in the operation of a digital archive is managing the identity, integrity and quality of the archives itself as a trusted source" (p. 23). Lynch (2000) also observed that "virtually all determination of authenticity or integrity in the digital environment ultimately depends on trust. […] Trust plays a central role, yet it is elusive". Thus, as early as the late 1990s, the discussion on trusted digital repositories (TDRs) spread and addressed how the "trusted" information can be preserved.

While trust in repositories has been much discussed, questions remain regarding whether end users will accept a repository with a solid record as "trusted". Ross and McHugh (2005) presented a broad range of trust-related issues surrounding digital repositories and argued that users' expectations (with expectations of depositors, aspirations of service providers, and management concerns) must be addressed (p. 2). Understanding users' perspectives is particularly significant because it is directly related to the fundamental missions of repositories, which is to serve a particular user group or designated community. As RLG/OCLC (2002) claimed, a trusted digital repository is "one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future" (p. 5). An empirical study to measure users' perceptions of trust has also been called for because "trusted digital repositories can be classified as 'trusted' primarily because they meet or exceed the expectations and needs of the user communities for which they are designed" (Prieto 2009, p. 603).

In response to this gap, this study attempts to investigate how users define "trust" in relation to digital repositories, and which factors influence users in building trust and/or maintaining it. Particularly, the focus of this study is in data repositories

where (digital) research data are stored and managed for reuse. Although most previous studies on TDRs attempted to answer how digital repositories can be trusted and provided criteria for developing TDRs, this study provides an in-depth understanding of users' perspectives on trust and contributes to broadening the understanding of questions such as "what is a trusted repository?" and "how can current TDRs meet users' expectations for being a trusted digital repository?" Lastly, this study provides implications for building more trusted repositories in the future.

Literature review

The first part of the literature review covers the efforts to build TDRs and the development processes and standards for Trustworthy Repositories Audit and Certification (ISO 16363). The second part of the literature review discusses definitions, dimensions, preconditions, and attributes of trust, as they have been investigated across disciplines.

Efforts to build trusted digital repositories: TRAC/ISO 16363

The concept of trust is known to be difficult to define or measure (Rousseau et al. 1998) because it is a vague term with an elusive definition (Gambetta 1988). How to define trust in a digital repository is also subjective depending on context of use. The archival and computer professions have been using the term as a synonym with "reliable", "responsible", "trustworthy", and "authentic", in relation to archival functions such as creating, managing, and using digital objects (RLG/OCLC 2002, p. 8). The RLG/OCLC report (2002) noted that no collective agreement exists, as of yet, on a more exact definition of "trusted archives (or repositories)", possibly because of the subjectivity and abstractness of the concept. This remains true today. Consequently, a "trusted" or trustworthy organization is unable to identify themselves as trusted (CPA/RLG Task Force on Archiving of Digital Information 1996).

Ross and McHugh (2005) pointed out that this situation leads the public (as well as the repositories themselves) to accept digital repositories as being trusted if they can demonstrate that they have the properties of trustworthiness. Thus, the most important question is how to verify trustworthiness and how a repository can assert its own status as "trusted" (Ross and McHugh 2005). Therefore, the efforts to identify requirements for being a "trusted" repository were initiated, and certification for digital archives being declared as "trusted" was needed. These efforts included constructing a robust audit and certification program for digital repositories to enable these institutions to maintain the authenticity, integrity, and accessibility of digital materials over the long term.

In 1996, the CPA/RLG Task Force on Archiving of Digital Information (hereafter, Task Force) argued that to be trusted, digital archives have to demonstrate that they could preserve information authentically for the long term. The Task Force emphasized the capabilities of "trusted" organizations, which

include being able to store, migrate, and provide access to digital collections because these capabilities are critical components of digital archiving infrastructure (CPA/RLG Task Force on Archiving of Digital Information 1996).

In 2002, a report by RLG/OCLC (2002) provided a starting point for describing a framework of attributes and responsibilities for TDRs. In the report, the concept of a TDR is defined as the following: the repository has associated policies, standards, and technology infrastructure that provides the framework for digital preservation; and the repository has a trusted system, such as a system of software and hardware that can be relied on to follow certain rules. The ERPANET workshop report (2003) emphasized the role of an audit in this process. An audit itself does not directly improve uncertain situations with respect to being trusted because it only assesses these situations, but such assessments can certainly be intriguing efforts to analyze and improve situations (p. 6). Quality standards for creating digital resources, actions for capture, methods and procedures for storage and repositories, and technologies were discussed as ways to assess and improve situations.

In response to these works and calls for audit and certification programs, in 2003, the Research Libraries Group (RLG) and the National Archives and Records Administration (NARA) created a joint task force to specifically address digital repository certification. First, they pointed out that institutions often declare themselves as "OAIS (Reference Model for an Open Archival Information System)-compliant" to underscore their trustworthiness, but no established or agreed understanding existed for the meaning of "OAIS-compliant" (p. 1). OAIS was designed to provide a conceptual framework for building appropriate environments, functional components, and information objects for long-term preservation. Even before it became an ISO standard in 2002, because institutions had no other developed criteria, they used OAIS to declare themselves trusted (p. 1). Thus, RLG/NARAs (2005) research focused on building criteria for measuring this compliance by providing definitions of TDRs and components that should be considered TDRs. The metrics developed in the task force were tested in 2005 through the Certification of Digital Archives Project by the Center for Research Libraries (CRL). Funded by the Andrew W. Mellon Foundation, this project conducted actual audits of three digital archives: the National Library of the Netherlands–Koninklijke Bibliotheek (KB), Portico (Ithaka Harbors, Inc.), and Inter-university Consortium for Political and Social Research (ICPSR); and one archiving system, LOCKSS; and provided methodologies for auditing and certification with corresponding costs (CRL n.d.).

Meanwhile, European researchers also responded to the call for audit and certification programs. The Network of Expertise in Long-term STOrage of Digital Resources (nestor) project published the *Catalogue of Criteria for Trusted Digital Repositories* in 2006. Initiated in Germany, the main focus of nestor was to form "a web of trustworthiness in which digital repositories can function as long-term digital archives within various environments" (Dobratz et al. 2007, para 3). Focusing on trusted repositories certification, nestor attempted to identify criteria that would facilitate the evaluation of digital repository trustworthiness, both at the organizational and the technical level (nestor Working Group on Trusted repositories Certification 2006).

All of those efforts and contributions to build solid audit, assessment, and certification programs were combined in *Trustworthy Repositories Audit and Certification: Criteria and Checklist* (TRAC) in 2007 (CRL/OCLC 2007). TRAC became the basis for *Audit and Certification of TDRs*, prepared by the Consultative Committee for Space Data Systems (CCSDS 2011). It presented three categories of criteria: (1) organizational infrastructures that include governance and organizational visibility, organizational structure and staffing, procedural accountability and preservation, financial sustainability, contracts, licenses, and liabilities; (2) digital object management that includes ingest (acquisition and creation of archival information package [AIP]), preservation planning, AIP preservation, and information and access management; and (3) infrastructure and security risk management that addresses technical infrastructure and security risk management.

In 2012, TRAC became a new ISO standard, ISO 16363: Audit and Certification of trustworthy digital repositories, with ISO/DIS 16919: Requirements for bodies providing audit and certification of candidate trustworthy digital repositories, which is waiting to be approved. The two standards complement each other in relation to accessing and building TDRs; largely based in TRAC, ISO 16363 provides a list of criteria for being a TDR and ISO/DIS 16919 provides requirements for organizations that conduct audits and certifications (National Archives 2011). The creation of ISO 16363 reflects consensus within the digital preservation community regarding best practices for digital repositories. Although ISO standards are known as "voluntary" standards rather than "must-do", they provide an influential guideline for organizations attempting to build TDRs.

These previous efforts acknowledged the role of the user community in building TDRs and suggested a way of engaging users in the process. For instance, a repository should allow users to audit/validate that the repository is taking the necessary steps to ensure the long-term integrity of digital objects and record and act upon problem reports about errors in data so that users can consider the repository as trustworthy sources (CCSDS 2011). However, much less research has been done regarding the user side despite the call for an understanding of users' trust of TDRs. Prieto (2009) reviewed the concept of trust in online environments and TDRs and underscored the significance of user communities' perceptions of trust. He argued that "user communities are the most valuable component in ensuring a digital repository's trustworthiness" (p. 603) and called for empirical research measuring users' perceptions of trust as a factor contributing to TDRs. Most recently, the dissemination information packages for information reuse project in the University of Michigan Ann Arbor and OCLC research investigated trust from two user communities, quantitative social scientists, and archeologists (Yakel et al. 2013). The findings of this project identified four indicators of trust: repository functions, transparency, structural assurance to include guarantees of preservation and sustainability, and the effects of discipline and level of expertise. More empirical research of how users perceive TDRs and of factors that influence the building of users' trust would have practical implications for repositories' ability to prove themselves as "trustworthy" to user communities.

Trust: definition, precondition, dimensions, and attributes

Trust definition

The concept of trust has been widely studied in various disciplines, such as psychology, organizational behavior, and economics. However, as researchers from different fields take varying approaches to understanding the concept of trust through their own disciplinary lenses and filters, full consensus on the definition of trust has not yet been reached. Researchers have also argued about the difficulty of defining and measuring trust (Rousseau et al. 1998) since it is a vague term with an elusive definition (Gambetta 1988). However, several efforts to derive a definition of trust from different disciplines have been made.

Mayer et al. (1995) saw trust as a relationship between a trusting party (*trustor*) and a party to be trusted (*trustee*) and defined trust as "willingness to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor" (p. 712). Similarly, Doney and Cannon (1997) saw trust as "willingness to rely on another", and Lewicki and Bunker (1995) defined trust as a "confident, positive expectation". Later, in their study of a multidisciplinary view of trust, Rousseau et al. (1998) reported that "confident expectations" and "willingness to be vulnerable" are critical components of all definitions of trust regardless of discipline and defined trust as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (p. 395).

Precondition of trust

As pre-conditions for the development of trust, two components, risk and interdependence, have been mentioned across disciplines (Rousseau et al. 1998). Defined as the perceived probability of loss, risk was considered an essential component of the pre-conditions for trust (Rousseau et al. 1998, p. 395; Rotter 1967; Sheppard and Sherman 1998). Risk was also discussed as the higher-level concepts of uncertainty (Doney and Cannon 1997; Gambetta 1988; Lewicki and Bunker 1995), which can result from a lack of information (Giddens 1990) and vulnerability (Blomqvist 1997; Rousseau et al. 1998), as discussed by Mayer et al. (1995). Interdependence (or dependence) means that a trustee holds the potential to satisfy a trustor's needs; thus, it occurs "where the interests of one party (trustor) cannot be achieved without reliance upon another (trustee)" (Rousseau et al. 1998, p. 395).

Dimensions of trust

Different types of trust can emerge from different factors. Trust can emerge based on a trustee's rational choice when a trustor perceives that the trustee will perform beneficial actions. Rousseau et al. (1998) referred to this as *calculus-based trust*, borrowing from Barber's (1983) argument that this type of trust can be derived from credible information regarding the intention of another, which may be provided by reputation or certification. Trust can also be derived from repeated interaction over

time between the trustee and trustor, which is classified as *relational trust* (Rousseau et al. 1998). Reliability and dependability in previous interactions create and increase a trustor's positive expectations or beliefs about a trustee's intention (Rousseau et al. 1998; Lewicki and Bunker 1995). Finally, there can be *institution-based trust*, a trustor's feeling of security about situations because of structure assurance, such as guarantees, regulations, or the legal system (e.g., a contract or promise) (McKnight et al. 1998; Rousseau et al. 1998).

Trust attributes

From their review of literature, Mayer et al. (1995) suggested three attributes of perceived trustworthiness: ability, benevolence, and integrity. Ability refers to the skills, competence, and characteristics of a trustee that are influential in a specific domain; benevolence is the belief of a trustor that a trustee wants to do good work for a trustor; and integrity is a trustor's perception that a trustee will adhere to principles acceptable to the trustor. Mayer et al. (1995) argued that these attributes are universally relevant and have been adopted by many researchers as a starting point to develop their own framework. Pirson and Malhotra (2011) slightly modified this framework in the context of stakeholders' trust in organizations and provided a new framework with four attributes: identification, integrity, benevolence, and transparency. Integrity is the belief that an organization will act fairly and ethically; benevolence is the belief that an organization is concerned with the stakeholders' well-being; identification refers to stakeholders' understanding of an organization's intention or interests based on shared values and commitment (Lewicki and Bunker 1995; and transparency refers to perceived willingness to share trust-related information with stakeholders. Though transparency did not appear to predict trust in the results of this study, it is worth noting that several scholars (e.g., Mishra 1996; Tschannen-Moran 2001) have argued for transparency as one attribute of trustworthiness.

While the trust relationship and the trust model investigated in previous studies are not exactly the same as the trust relationship between users and repositories, previous studies have provided useful insights into factors that may influence trust and how it can be built. Thus, this study employed related concepts developed in previous studies, adopting an integrated approach from organizational studies, sociology, and social psychology to enhance our understanding of trust in repositories from the users' point of view.

Methods

I conducted semi-structured interviews to gain more in-depth understanding of users' perceptions. Among various types of digital repositories, I limited the scope to users of data repositories, because of both the significance of data in research and the increasing attention that is being paid to data sharing and reuse (e.g., Faniel and Zimmerman 2011). Potential subjects were identified from users of three major social science data repositories in the US: Odum Institute for Research in Social

Science at the University of North Carolina, Roper Center for Public Opinion Research at the University of Connecticut, and the ICPSR at the University of Michigan. These three repositories are participating in the Data Preservation Alliance for the Social Sciences (Data-PASS). Data-PASS is a voluntary partnership of organizations created to archive, catalog, and preserve data used for social science research, and their data include opinion polls, voting records, surveys on family growth and income, social network data, government statistics and indices, and GIS data measuring human activity. Other organizations participating in Data-PASS were excluded from this study due to differences in the nature of the repositories (a government organization) or due to the difficulty of tracking their users.

I used data citation tracking to identify people who have used data from these repositories for their research. Currently, as no consistent standard for citing data has yet been established (Mooney 2011), tracking data citations is a challenging process. In addition, perhaps because a number of articles use data without citing them (Mooney 2011), the use of citation tracking to identify users might be a limitation of this sampling plan. Even though data citation tracking has limitations, it is still the most effective way to identify users of datasets from each repository. Among the three repositories, ICPSR and Roper Center provide lists of publications that have used their data,[1] and potential study participants were identified from these lists. Users of the Odum data archives were identified by searching Google Scholar, as the search results generated from the search term "Odum" provide the lists of publications that mentioned Odum as their data source.

To minimize the potential imbalance in the sample, created by the use of different repositories, a quota sampling technique was employed for deciding user numbers for each repository, based on the number of studies that have provided data to the repository (as reported by the repository or discovered through searches of the DataVerse network[2]). Thus, this study initially aimed to recruit about eight participants from ICPSR, about three participants from Odum Institute, and about ten participants from Roper Center.

Potential participants were identified through searches for publications that cite their use of a dataset from one of the three repositories. These searches were limited to journal publications and conference proceedings published since 2000. Users from the most recent years were included in the sample first, and this process was continued until a sufficient number of participants had been identified. For articles that have been written by multiple authors, either the corresponding author or the first author was contacted first.

The interview data were collected from February to July 2012. An email invitation for this study was sent to 213 potential participants identified through the process described previously. Twenty-five people who received the email invitation volunteered for interviews, but only 19 ended up participating in this study. Six volunteers were excluded because they had not been heavily involved in the process of acquiring and using data, as their co-authors or research assistants handled these

[1] ICPSR: http://www.icpsr.umich.edu/icpsrweb/ICPSR/biblio/resources?collection=DATA; Roper Center: www.ropercenter.uconn.edu/research/research_bibliography.html#.TxyOx0qQ07A.

[2] The DataVerse network: http://thedata.org/book/about-project.

activities. Among 19 participants, ten were identified as ICPSR users, three were Odum users, and six were Roper Center users. Some participants used multiple repositories. Those participants first talked about their most recent experience within the repository they identified themselves as using, but they also talked about their experience with other repositories if they felt it necessary, including repositories other than Odum, Roper Center, and ICPSR. All interviews were conducted by phone, and the data were fully transcribed. Transcribed data were inductively coded for analysis using TAMS analyzer, a software program that facilitates qualitative analysis. The codes developed for initial analysis were reviewed by one peer checking the validity of codes.

Findings

Participant information

All 19 participants were either academic or corporate researchers, which can be seen as a part of the "designated community" of each repository. Seventeen were university faculty members (including junior and senior levels), and two were classified as research associates at either a university or a research corporation. Eight participants were male and 11 were female. Most participants' ages ranged from 30 to 50 years (one in their 20 s, seven in their 30 s, five in their 40 s, and six in their 50 s). While all the participants used the repositories to acquire data for their research, a few of them said that they also used it for teaching. Two participants also had experience with depositing, which might influence their perception of repositories, but the deposit experience was not investigated extensively since it is outside of the scope of this research. The level of use varied among participants. Participants had different ways of expressing their level of experience; for instance, by number of datasets used, by number of years using the repository, or by number of times datasets were used for research projects; and it was especially difficult for users who have used the repository for a long time to express this. For example, descriptions included "Half of my publications over the last 15 years, […] double-digits number of papers (PB12)". However, most participants had used data repositories more than five times and used more than five datasets. Five of them said they had used data repositories more than 100 times, and three of them had used repositories for more than 10 years. Only three participants had used repositories fewer than three times and had used fewer than three datasets.

Defining trust: what does trust mean to users?

Participants were first asked what they think trust is, and how they define trust in the context of data repositories. Similar to the preconditions discussed in previous literature, the interviews showed that trust became relevant to a particular situation when the trustor was uncertain about something (uncertainty) and when the trustor can depend upon the trustee (dependability). Trust can arise or become necessary under uncertain circumstances because it is sometimes necessary to "place

confidence in things that you don't know" (PA05). Trust was also related to dependability. Participants expressed that "to trust" means to believe or count on someone or something (PA01, PB10, PB12), and PB10 defined trust as "being able to count on organizations or products or whatever". Dependability signified the trustor's expectation that the trustee would consistently satisfy the needs of the trustor. PA14 expressed this understanding as "[Trust is] your ability to have faith that someone is going to fulfill some kind of expectation that you have".

Participants' sense that they could rely on someone or something was highly associated with truthfulness, which is the lack of deception. In the context of data repositories, lack of deception had two components: data validity and repositories' integrity. Even though participants were asked to define trust in repositories, discussion of data validity inevitably emerged, as the needs for data came before other considerations because this was what participants actually used for their research. Belief in the integrity of repositories was another component of trust, due to the repositories' role of managing data.

Whether the data are presented accurately (*validity*) constituted the most important component of the definition of trust in the context of repositories. PA08 said, "The trust I have is […] the data in a way that accurately indicates what's there and really what was collected. So that's what I would be basing my trust on". PA03 remarked, "To trust them I believe that they are accurately representing what they say they are, which includes telling me the limitations of something and not just presenting all the good parts, but presenting the bad parts". Data should reflect exactly what it is, and accuracy in this context has nothing to do with evaluating how good or bad the data is. This dimension is also highly related to integrity, which is discussed in the next paragraph.

The *integrity* of repositories was mentioned by most of the participants. Participants' belief that organizations will be honest rather than deceitful comprises trust. As PB16 noted, "[repositories] are in fact doing what they're saying that they're doing, and they're not trying to intentionally, I guess, mislead people". PA07 echoed, "I really think about your believing that they represent themselves as true and honest […] or do what they say they are going to do and [that] they have respect for you and so on, et cetera".

Building trust: where does end users' trust originate, and how do users develop trust?

Study participants discussed a number of characteristics that contribute to the development of their trust in repositories. Regardless of the repository, participants' trust seemed to be based on five broad components: organizational attributes, the internal repository process, user (or designated) communities, their own past experiences, and their perceptions of the roles of repositories.

Organizational attributes

About half of the participants showed strong belief in the integrity of repositories. Here, integrity means that users believe repositories are honest and do not deceive or mislead their users. PA01 said, "Well, I don't think that the people or the

organizations that managed the data were trying to mislead anyone", and PB16 added, "There's no reason to think that [the repository] would be doing anything to the data to affect its integrity. They're all about just making data available to the research community".

This strong belief in the repositories' integrity is also based on users' understanding of the repositories' mission and commitment to society; as PB16 stated above, repositories are "all about just making data available to the research community". PA04 echoed this view:

> They had provisions in place for data use agreements that they try to make it so that people can diffuse the data that they were using it for research purposes that would further knowledge.

PA14 acknowledged the value and mission of the repositories, noting that "the value of having things like that at [the repository], is that they are concerned with long-term preservation". Similarly, PB12 stated, "If they stopped operating or were no longer able to archive as much data as they did in the past, then, well, the data would be lost". Understanding repositories' commitment to society named identification by Pirson and Malhotra (2011), PB16, PA04, PA14, and PB12 expressed their faith in the repositories' integrity.

Participants' belief in the staff was a third organizational attribute influencing trust. This trust is closely related to the reputation of the repositories because it can arise from the reputation of the repositories, as well as help to build a good reputation. However, it is worth noting that some of the participants' trust was directed particularly toward the staff. Participants believed that the staff "were well trained in this area" (PA01), "have expertise" (PB10), and "are the best possible people working on it" (PA03). One participant (PA09) also stated that knowing about staff helps to build his trust because "it just makes [repositories] more visible, rather than just being these mysterious sites where people put datasets up that aren't available for anyone to download". Interestingly, even though several participants expressed a strong belief in the repository staff's expertise, other participants did not know what the staff members do with data, which will be discussed in a later section.

Perceptions of and use by designated communities

Another component that emerged from the interviews was trust transferred from other users who are close to participants or trust based on others' use. PA02 noted, for example, that "It's not like I just stumbled upon it myself; I worked with other researchers who were working with [the repository], and they are the ones who told me [to use it]". If users hear about a repository from sources with more authority, they tend to trust it more:

PA07 […] maybe one of my professors said positive things about [the repository] so that I consider that professor a reliable source of information. So, to me, that is my… It's someone I trusted, that if this was a trustworthy organization, it made me feel… To trust it as well.

Another participant commented that the frequency of others' use of the sources in the repository can be one measure of trust, even though it is not same as hearing this directly from other users.

PA09 I would trust the [repository]. It's been used a lot. And so I can't even imagine how many master's theses and dissertations and research articles have been published based on [the repository] data.

The reputation of repositories was another attribute mentioned by a number of participants. Reputation plays a significant role in the trust users have in repositories and is sometimes a consideration when choosing a dataset:

PA05 I mean the institute that provides the dataset that I talk about on [the repository], that's a reputable institution, you know what they have, the data that they have are good quality data. […] So people who are distributing the data, their reputations are important.

PC19 I would trust [the repository] because I've already heard so much about it. Reputation is important and it has a great reputation, so if I say I wanted to work on a different dataset I would go there versus some other small university somewhere else. I wouldn't know about them, whether they have all of these means of data collection and data security.

PA13 First of all, they've both been in… this kind of business for a long time. They're world-renowned for being data repositories and for leading the field in terms of data preservation and data access.

Users' own past experience

In addition to recommendations from other users, users' own experiences with repositories are important factors in building trust. The majority of participants in this study had used some repositories multiple times and related that having positive experiences with repositories over time helps to enhance their trust. For example, one participant said:

PC19 I guess I did not had any problems in the past and I haven't heard of other people having problems, and the data that I accessed through the repository, everything seems to be helping and there wasn't anything suspicious or missing from it. So I guess I've had a good experience, so I have no reason to distrust them.

Repository processes: documentation, data cleaning, and quality checking

What repositories do with datasets is closely related to users' trust. Almost every participant discussed documentation (e.g., codebooks, original questionnaires, or methodology) of datasets, arguing the significance of having good documentation since it is "only possible to understand [the dataset] by the reading documentation (PA02)". Good documentation is one factor influencing the user's trust; as expressed by PA03, "Well, [I trust them] because they have really detailed and

rigorous documentation describing their methodology". Another participant compared two different repositories and each one's documentation, describing having more trust in the one with more rigorous documentation:

PA14 Yeah, [repository A] tends to have more thorough complete documentation. [Repository Z] and [repository Y] are more whatever they get from the original investigator. So a lot of times with [repository Z], you'll actually get an extant copy of the original questionnaire which sometimes there's things crossed out by hand, different codes written in again by hand because of the last-minute changes. Whereas [the repository], you're always gonna get a kind of nice, clearly organized thing without corrections and crossed out, so it's kind of like getting somebody's notes versus getting a finished manuscript. [Repositories Y and Z were not included in the study; Repository A was included.]

Several participants were also aware of the internal data cleaning process of repositories and expressed their trust from this process:

PA09 Well, there's no trust issue about it. I mean if there's an error, I think that [the repository] will do their best to make sure that it's corrected and they'll be very responsive. So that goes a long way to continuously building trust.

Whether it is true or not, a few of the participants believed that the repositories would perform quality checks and appraisals of datasets. PA09 assumed that the repositories would meet some appraisal criteria for data quality, stating, "I really don't know, but I'm guessing they would have. I don't think they would just put something out without sort of reviewing it or ensuring data quality". Two of the others (PA01 and PA02) had a different view, stating, "I don't think [the repository] is requiring each individual project or dataset that's placed on their site to pass some criteria… You can't just say, well, it's in [the repository], so it must be great" (PA01). Accordingly, appraisal was one factor that influenced at least some of the participants' trust, as some of them believed the repositories would check the quality of their data.

Users' perceptions of the roles of repositories

Participants perceived repositories in a variety of ways, particularly regarding their roles. These perceptions turned out to influence the users' trust in data repositories. Since this study does not aim to quantitatively test the correlation among factors influencing trust, it is not possible to argue for a consistent relationship between user perceptions and trust. However, the interview data indicated that some participants (PB06, PA08, and PC11) who perceived the repositories' roles as somewhat limited did not consider the repositories to be trustworthy.

For instance, PB06 defined the role of repositories as very limited, which led PB06 not to associate repositories with trust. PB06 perceived repositories' functions as below:

PB06 I make the assumption that the repository has very little to do with the data because I know that… They don't do much more than manage the files that are provided to them by the organization. They're not in the business of doing much other than putting things in storage. So [the repository] is almost irrelevant from the point of view of trust.

This case was in sharp contrast to the views of PA01, PB12, and PA14, who thought repositories did data cleaning and trust this process (see the section, "Repository processes").

Such perceptions of the roles of repositories as limited made users see repositories as more of a "library" (PC11) or "warehouse" (PA13), where there are not many jobs or processes involving data. This view diminished users' concern about the trust issue in repositories, and often made users question why trust mattered in this context.

PC11 I don't think this question makes any sense. This is like asking whether I have any concerns about using non-fiction books in the library. Some books will be shown to have mistakes; others will become influential. Each book must be judged on its own merits. But librarians cannot know this before placing an order for a book and placing it on the shelves. Librarians can have general rules of thumb in terms of ordering non-fiction books. But just because a book is in the library means almost nothing. Unless the library is run by some extremist group, I judge the book, not the library. Similarly, just because a dataset is in a repository means nothing. A repository should have very tight standards for documentation. Then the user can make informed decisions about each dataset or data series.

Discussion

This study did not account for possible personality-based aspects of trust (e.g., intrinsic trust), which can be considered a possible limitation. Even though, the findings of this study present how users perceived and defined *trust*, as well as what factors influenced their trust development. At a higher level, users perceived trust as their willingness to depend on or rely on something or someone (including an organization) in uncertain circumstances. These elements—dependability and uncertainty—align with pre-conditions of trust discussed in the previous literature across disciplines (Doney and Cannon 1997; Giddens 1990; Lewicki and Bunker 1995; Rousseau et al. 1998).

When it comes down to the specific context of data repositories, users' *trust* definition is largely based on lack of deception. In particular, a lack of deception can be achieved in two different ways: by determining data validity (or accuracy) and by assessing the integrity of repositories. Outcomes of repositories—meaning datasets deposited and processed in the repositories—should accurately represent the original dataset. Repositories should be honest and not intentionally mislead anyone. This strong presence of truth and honesty in users' definition of *trust* in data

repositories reflects the significance of data integrity and validity in their research. Acquiring valid, accurate data is the first step for any type of research that reuses data produced by others. The issue of integrity of repositories also relates to data integrity and validity as trusted sources for data. A number of factors, such as organizational attributes and repository processes, that influence users' development of trust eventually relate to the data integrity and validity issue.

Organizational attributes, user communities (recommendations and frequent use), past experiences, repository processes (documentation, data cleaning, and quality checking), and users' perception of the repository roles were identified as influencing the development of users' trust in repositories. These findings also reflect several types of trust discussed in previous literature. As Rousseau et al. (1998) and Barber (1983) argued, users could develop their calculus-based trust based on their rational choices, knowing the good intentions of repositories. This can be influenced not only by good reputation (Barber 1983) but also by others' recommendations and frequent use of repositories by user communities. Relational trust also appeared when users develop their trust based on repeated positive experiences with repositories over time; such experiences help users develop positive expectations about the repositories. Knowing the staff of repositories also helped to develop relational trust; as an example, users expressed their trust of the staff when they met the staff in a conference or knew them personally because, in one way, this contact gave users an impression of repositories as "real" and "visible", rather than being "mysterious sites" (PA09). Users might also develop institution-based trust from reputation. As Rousseau et al. (1998) found, institution-based trust can convert to formulate calculus-based and relational trust; therefore, reputation could play an important role in users' development of trust.

One finding distinctive to users' trust in data repositories is users' perception of repository roles. Each study participant had a different level of understanding of repositories, and the level of understanding was sometimes, surprisingly, not related to the level of a user's experience with repositories, since most participants of this study have frequently used repositories. Different levels of understanding—for instance, how much users know about repositories' functions or the roles of staff/repositories—are not the result of the possible differences among the three repositories in this study because these differences were also apparent in users of the same repository. The level of users' understanding of repositories might be relevant to their level of trust, as it can be seen from the factors of repository process and users' perception of the repository roles.

Participants who knew a repository's internal processes expressed their trust based on their belief in that process; however, others who did not know much about repository processes had different thoughts. In particular, a couple of participants viewed repositories as much less active players in maintaining data integrity than they are, although it is true that there might be differences among repositories in this study regarding processes and other functions. For them, because repositories are the same as a "warehouse" (PA08) or "library" (PC11), not much room existed for trust for those participants. These findings suggest that users' awareness of repositories' roles or functions can be one factor for developing users' trust. In addition, if this is a factor, a new question is raised: Should the roles or functions of

repositories be more visible to users to gain more trust? Some might argue that, if users do not know much about what repositories do, such lack of knowledge shows that repositories have been successfully performing their job because users have not experienced serious problems. However, knowing the missions and functions of repositories can be a way to decrease users' uncertainty, which can positively influence their trust, as can be seen in this study. Furthermore, building a better understanding of repositories would be important by giving users a better understanding.

Conclusion

As Prieto (2009) argued, "User communities are the most valuable component in ensuring a digital repository's trustworthiness" (p. 603). Gaining users' trust in repositories is important because one of the core missions of repositories is to serve their distinctive user communities. By understanding users' perspectives on "trusted" repositories, repositories can enhance their "trusted" status because users' perceptions of trust are often (though not always) related to repositories' practices.

If understanding users' trust is the first step, the next step entails developing a metric to measure users' trust in repositories. Trust is a complex concept to measure, but having a standardized way of measuring users' trust can help to demonstrate how repositories have effectively gained users' trust and have been perceived as trusted sources of information. In addition, trust in data itself plays a distinctive and important role for users to reuse data, which may or may not be related to the trust in repositories. Although it is not the scope of this study, findings also suggested that users' trust in data is another important area to be investigated further.

Acknowledgments I would like to give special thanks to Professor Barbara Wildemuth at the University of North Carolina at Chapel Hill, School of Information and Library Science, for her assistance with this study.

References

Barber B (1983) The logic and limits of trust. Rutgers University Press, New Brunswick

Blomqvist K (1997) The many faces of trust. Scand J Mag 13(3):271–286

Center for Research Libraries (CRL) (n.d) Certification of digital archives project. http://www.crl.edu/archiving-preservation/digital-archives/past-projects/cda. Accessed 30 Aug 2012

Center for Research Libraries/Online Computer Library Center (CRL/OCLC) (2007) Trustworthy repositories audit & certification: criteria and checklist (TRAC. Version 1.0)

Commission on Preservation and Access/Research Libraries Group (CPA/RLG) Task Force on Archiving of Digital Information (1996) Preserving digital information: report of the task force on archiving of digital information

Consultative Committee for Space Data Systems (CCSDS) (2011) Requirements for bodies providing audit and certification of candidate trustworthy digital repositories (No. CCSDS 652.1-M-1). The consultative committee for space data systems

Dobratz S, Schoger A, Strathmann S (2007) The nestor catalogue of criteria for trusted digital repository evaluation and certification. J Digit Inf 8(2). http://journals.tdl.org/jodi/article/view/199/180. Accessed 13 Nov 2011

Doney PM, Cannon JP (1997) An examination of the nature of trust in buyer–seller relationships. J Mark 61:35–51
Electronic Resource Preservation and Access Network (ERPANET) (2004) The role of audit and certification in digital preservation. Stadsarchief Antwerpen, Belgium
Electronic Resource Preservation and Access Network (ERPANET) Workshop Report (2003) Trusted digital repositories for cultural heritage. Accademia nazionale dei Lincei, Rome
Faniel I, Zimmerman A (2011) Beyond the data deluge: a research agenda for large-scale data sharing and reuse. Int J Digit Curation 6(1):58–69
Gambetta D (1988) Can we trust trust? In: Gambetta D (ed) Trust: making and breaking cooperative relations. Basil Blackwell, New York, pp 213–237
Giddens A (1990) The consequences of modernity. Stanford University Press, Stanford
ISO 14721 (2002) Reference model for an open archival information system (OAIS)
ISO/DIS 16363 (2012) Audit and certification of trustworthy digital repositories
ISO/DIS 16919 (under development) Requirements for bodies providing audit and certification of candidate trustworthy digital repositories
Jantz R, Giarlo M (2007) Digital archiving and preservation: technologies and processes for a trusted repository. J Arch Organ 4:193–213. doi:10.1300/J201v04n01_10
Lewicki RJ, Bunker BB (1995) Developing and maintaining trust in work relationships. In: Kramer RM, Tyler TR (eds) Trust in organizations: frontiers of theory and research. Sage Publications, CA, pp 114–139
Lynch C (2000) Authenticity and integrity in the digital environment: an exploratory analysis of the central role of trust. Authenticity in a digital environment pub92. http://www.clir.org/pubs/reports/pub92/lynch.html. Accessed 7 Nov 2011
Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. Acad Manag Rev 20(3):709–734
McKnight DH, Cummings LL, Chervany NL (1998) Initial trust formation in new organizational relationship. Acad Manag Rev 23(3):473–490
Mishra AK (1996) Organizational responses to crisis: the centrality of trust. In: Kramer RM, Tyler TR (eds) Trust in organizations: frontiers of theory and research. Sage Publications, CA, pp 261–287
Mooney H (2011) Citing data sources in the social sciences: do authors do it? Learn Publ 24(2):99–108. doi:10.1087/20110204
National Archives (2011) NARAtions: the blog of the United States National Archives (2011, March 15) ISO standards for certifying trustworthy digital repositories. http://blogs.archives.gov/online-public-access/?p=4697. Accessed 8 Sep 2011
Nestor Working Group on Trusted repositories Certification (2006) Catalogue of criteria for trusted digital repositories: version 1. http://files.d-nb.de/nestor/materialien/nestor_mat_08-eng.pdf. Accessed 20 Dec 2011
Pirson M, Malhotra D (2011) Foundations of organizational trust: what matters to different stakeholders? Organ Sci 22:1087–1104. doi:10.1287/orsc.1100.0581
Prieto AG (2009) From conceptual to perceptual reality: trust in digital repositories. Libr Rev 58(8):593–606. doi:10.1108/00242530910987082
Research Libraries Group/National Archives and Records Administration (RLG/NARA) Task Force on Digital Repository Certification (2005) Audit checklist for certifying digital repositories. http://web.archive.org/web/20050922170446/http://www.rlg.org/en/page.php?Page_ID=20769. Accessed 30 Aug 2011
Research Libraries Group/Online Computing Library Center (RLG/OCLC) Working Group on Digital Archive Attributes (2002) Trusted digital repositories: attributes and responsibilities. http://www.rlg.org/longterm/repositories.pdf. Accessed 1 Jul 2011
Ross S, McHugh MA (2005) Audit and certification of digital repositories: creating a mandate for the Digital Curation Centre (DCC). RLG DigiNews 9(5). http://eprints.erpanet.org/105/01/Ross_McHugh_auditandcertification_RLG_DigiNews.pdf. Accessed 11 Nov 2011
Rotter JB (1967) A new scale for the measurement of interpersonal trust. J Pers 35:615–665
Rousseau DM, Sitkin SB, Burt RS, Camerer C (1998) Not so different after all: across-discipline view of trust. Acad Manag Rev 23(3):393–404
Sheppard BH, Sherman DM (1998) The grammars of trust: a model and general implications. Acad Manag Rev 23(3):422–437
Tschannen-Moran M (2001) Collaboration and the need for trust. J Educ Adm 39(4):308–331

Yakel E, Faniel I, Kriesberg A, Yoon A (2013). Trust in digital repositories. Int J Digit Curation 8(1):143–156. doi:10.2218/ijdc.v8i1.251

Author Biography

Ayoung Yoon is a third-year doctoral student at the School of Information and Library Science, University of North Carolina at Chapel Hill. Her research interests include users' trust in data and data repositories, data curation, and personal digital archiving on the Web. She has an MSI in both preservation and archives and record management from the University of Michigan School of Information, and BA in history from Ewha Womans University, South Korea. She is currently a Carolina Digital Curation (DigCCur) Doctoral Fellow.