

The International Journal of Digital Curation

Volume 8, Issue 1 | 2013

Trust in Digital Repositories

Elizabeth Yakel,
School of Information,
University of Michigan

Ixchel Faniel,
OCLC Research

Adam Kriesberg,
School of Information,
University of Michigan

Ayoung Yoon,
School of Information and Library Science,
University of North Carolina

Abstract

ISO 16363:2012, Space Data and Information Transfer Systems - Audit and Certification of Trustworthy Digital Repositories (ISO TRAC), outlines actions a repository can take to be considered trustworthy, but research examining whether the repository's designated community of users associates such actions with trustworthiness has been limited. Drawing from this ISO document and the management and information systems literatures, this paper discusses findings from interviews with 66 archaeologists and quantitative social scientists. We found similarities and differences across the disciplines and among the social scientists. Both disciplinary communities associated trust with a repository's transparency. However, archaeologists mentioned guarantees of preservation and sustainability more frequently than the social scientists, who talked about institutional reputation. Repository processes were also linked to trust, with archaeologists more frequently citing metadata issues and social scientists discussing data selection and cleaning processes. Among the social scientists, novices mentioned the influence of colleagues on their trust in repositories almost twice as much as the experts. We discuss the implications our findings have for identifying trustworthy repositories and how they extend the models presented in the management and information systems literatures.



Introduction

Kelton, Fleischmann and Wallace (2008) call for increased focus on the theoretical and empirical research into user trust in content. At the same time, the data sharing/reuse literature has also begun to examine this issue (e.g. Van House, 2002; Zimmerman, 2008; Faniel & Jacobsen, 2010). However, there has been less study of user trust in digital repositories (Ross & McHugh, 2006) and particularly how ISO 16363:2012: Space Data and Information Transfer Systems - Audit and Certification of Trustworthy Digital Repositories (hereafter ISO TRAC) engenders user trust.

Van House (2002) linked trust in repositories to sharing knowledge and scholarship. In her study of biodiversity data and epistemological trust, Van House acknowledged that trust played a role in data sharing in digital repositories. Trust was an implicit factor in sharing information, and the epistemic community served as the major source for determining trust. More recently, Prieto (2009) examined users' trust in digital repositories and underscored the roles of stakeholders – users or producers – and the significance of their trust. He argued that the roles of the repository's stakeholders cannot be disregarded in the process of determining trustworthy status, as the repository's goal is to serve the user (or designated) communities. He viewed “the digital repository as a trusted system,” noting “user communities and their perceptions of trust” as key. Our research builds on Prieto and examines how data reusers construct trust around digital repositories.

Drawing on ISO TRAC and the management and information systems literatures, we identify two components of trust: trustworthy actions by repositories and trust by external stakeholders. We argue that both of these components are necessary for a repository to be considered trustworthy. We define trust as:

‘A psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another.’ (Rousseau, Sitkin, Burt & Camerer, 1998)

In digital repositories, for example, data reusers may face reputational harm if they unintentionally misuse data due to insufficient contextual information.

Our study is based on in-depth interviews with 66 data reusers from two disciplinary communities: quantitative social science and archaeology. We are interested in how they conceptualize trust in data repositories. In particular, we ask the following research questions:

1. How do data reusers construct/conceive of trust in repositories?
2. How do data reusers associate repository actions with trustworthiness?

Our findings indicate that designated communities do associate repository actions with trust. We found similarities and differences across the two disciplines. Both disciplinary communities associated trust with a repository's transparency. However, archaeologists mentioned guarantees of preservation and sustainability more frequently than social scientists who talked about the influence of colleagues and institutional reputation.



Literature Review

Constructing Stakeholder Trust in Digital Repositories

Early documents associated with the digital repository audit and certification process acknowledged that trusted status is only partially achieved through an audit.

‘Certification for digital repositories will involve far more than the documentation of criteria ... It must recognize standards and best practices relevant to the community of the repository, as well as those of the information management and security industries as a whole. In other words, audit and certification of trusted digital repositories cannot exist in a vacuum.’ (CRL & OCLC, [2007](#))

Thus we examine how users perceive trustworthy activities and then formulate a decision to trust a repository. We begin this literature review with a discussion of ISO TRAC. Then we summarize key aspects of the management and information systems literatures pertaining to trust in repositories. Throughout the review, we discuss parallels between ISO TRAC and these literatures.

ISO TRAC and trust

ISO TRAC presents a set of functions for repositories to enact in order to be considered trustworthy (i.e. selection, data processing/cleaning, preservation). The document also references designated communities as having an active role in the attainment of these criteria and thus the construction of trust. For example, Section 3.3.2 notes that: “The preservation policy might then include information about the expected level of understandability by the repository’s Designated Community for each Archival Information Package.” Similarly, Section 4.2.5.2 is also predicated on a reaction by the designated community. While ISO TRAC does not dictate exactly how to satisfy a particular audit requirement, it does provide suggestions for the types of evidence it views as acceptable to meet the stated criteria; however, these suggestions range from very specific to very vague. For instance, Section 3.1.3 calls for a collection policy and proceeds to identify a collection policy both as the action and the evidence required to meet this criterion. In contrast, Section 4.1.1 lists a variety of potential types of evidence to demonstrate the identification of Information Properties, ranging from mission statements to workflow and Preservation Policy documents. These evidentiary materials pertain to multiple repository functions at different levels (administrative, operational, etc.). While ISO TRAC is ostensibly about repository actions, in many cases it requires a designated community to recognize trustworthy actions as well as to acknowledge repository principles, such as transparency, taken on their behalf.

Management and information systems literatures and trust

The management and information systems research on trust and organizations can be divided into two main areas: employee trust and external stakeholder trust in the organization. We are concerned with external stakeholder trust. We approached this body of literature from two perspectives: organizational trust and technology acceptance. The management literature examines organizational trust (e.g. Pirson &



Structural assurance

A second concept in repository trust is structural assurance. Structural assurance “refers to one’s sense of security from guarantees, safety nets, or other impersonal structures inherent in a specific context” (Gefen, Karahanna & Straub, [2003](#)). We focus on three aspects of structural assurance: third party endorsement, guarantees, and reputation.

A third party endorsement occurs when an organization submits to judgment by an external assessor’s standards. This often results in some type of visible notification, which leads to consumers’ validation of the endorsement (Kimery & McCord, [2002](#)). Third party endorsement has been shown to affect trust in organizations conducting online business (Gefen, Karahanna & Straub, [2003](#)). Examples of third party endorsements for a repository are ISO TRAC Certification or the Data Seal of Approval. Here, the management and information systems literatures agree with the ISO TRAC statement that: “It is important to acknowledge that there is real value in knowing whether an institution is certified to related standards or meets other controls that would be relevant to an audit” (ISO TRAC, [2012](#)).


Guarantees are actions on the part of the organization that stakeholders perceive as mitigating risk. Gefen, Karahanna and Straub ([2003](#)) identify typical types of guarantees in the eCommerce literature, such as not conveying inaccurate information, not making statements against violations of privacy, unauthorized use of credit card information, and unauthorized tracking of transactions. In the repository setting, we interpret guarantees to be statements displayed on repository websites concerning data preservation and/or sustainability of the organization. While selection, metadata and data processing are repository functions, we considered preservation and sustainability as mechanisms of structural assurance. In a repository context, these fit into Gefen, Karahanna and Straub’s definition by adding a sense of security and forming safety nets. These functions are also central to ISO TRAC.

Reputation had been shown to affect initial trust formation as well (e.g. Jarvenpaa & Tractinsky, [1999](#); McKnight, Cummings & Chervany, [1998](#)). Institutional reputation is key because it is built over time and reflects stakeholders’ recognition of specific, cumulative behavior on the part of an organization.

Social factors

Social factors represent “the individual’s internalization of the reference group’s subjective culture” (Thompson, Higgins & Howell, [1991](#)). Derived from Triandis ([1977](#)), social influence is a major cultural factor that has been shown to affect trust in organizations. Therefore, we identified three major types of social influence: peers, mentors or senior colleagues, and institutions (Venkatesh, Morris, Davis & Davis, [2003](#)), and analyzed our data to see how our respondents spoke about these in terms of influencing decisions to trust a repository.

We seek to bridge concepts in the management and information systems literatures with those from digital curation by adapting them and examining how they pertain to digital repositories. Through our literature review, we were able to isolate factors which we believe affect stakeholder trust in repositories. Some of these factors depend on organizational actions (benevolence, integrity, identification, and transparency). Others, such as social factors, social influence and aspects of structural assurance



(third-party endorsement and reputation) depend on external acknowledgement. Also of note, two of the direct trust factors – identification and transparency – appear to be closely aligned with ISO TRAC. Furthermore, we expect that aspects of structural assurance (third party endorsement and the guarantees that we define as preservation and sustainability), which pertain to ISO TRAC and the certification process may influence trust in digital repositories. We take a closer look in this paper.

Methods

Our findings are drawn from data collected during three rounds of interviews conducted between June 2011 and April 2012. In total, we spoke with 66 participants: 22 novice social science researchers, 22 expert social science researchers, and 22 archaeologists. We selected archaeologists and quantitative social scientists for two primary reasons. First, these disciplines work with very different types of digital data and have different scholarly traditions around data sharing and reuse. Quantitative social scientists work with structured data and codebooks, often controlled by standards. Archaeologists use heterogeneous data, often triangulating data from multiple sources created using local practices and a variety of de facto standards. Second, repositories in these two disciplines are at different stages of maturity, with those for social scientists more mature and more widely known and supported.

In our series of semi-structured, hour-long interviews, we asked respondents to discuss their experiences of reusing data in their particular field of research. Topics of inquiry included how respondents discovered and evaluated data for reuse, and their experiences and thoughts about digital data repositories. We used convenience and snowball sampling to recruit participants. We began with personal contacts in each field, then we recruited additional participants at workshops and conferences. Finally, we asked interviewees to nominate colleagues for us to interview. Through our selection process, we attempted to recruit a range of researchers in each field in terms of topics studied, research methods used, the centrality of data reuse to their research agendas, and level of expertise. Interviewees were paid \$25 US dollars for their participation in the study.

All interviews were audio recorded and transcribed. Transcripts were then coded using NVivo – a qualitative data analysis software tool. Prior to analyzing each group of transcripts, we developed a code set based primarily on the themes we addressed in our interview protocols. Top-level categories included context, data reuse, data sharing and repository codes. We also remained open to emergent themes that arose from the data. While slight differences in the protocols led to minimal code differences (i.e. addition of an “Archaeological Ethics” code), we made an effort to keep the three code sets as similar as possible, in order to facilitate comparison across groups of participants. Two members of our project team coded each group of transcripts. We calculated inter-rater reliability using Scott’s Pi. The coders achieved scores of 0.88, 0.77, and 0.73 for the novice social scientists, expert social scientists and archaeologists’ transcripts respectively.



Findings

We have organized our findings into three sections. First, we examine the degree to which data reusers' recognized trustworthy actions on the part of repositories. Second, we present data on reusers' construction of trust in repositories, utilizing dimensions from the management and information systems literatures. Third, we highlight some of the disciplinary differences and similarities between archaeologists and quantitative social scientists, as well as between novice and expert social scientists.

Recognizing Trustworthy Actions by Repositories

The management and information systems literatures do not focus on organizational functions around the technologies, but we found that data reusers did link trust to repository processes (see Table 1). In particular, 18 interviewees (27.3%) mentioned data processing, metadata or data selection in conjunction with a trust decision. For example, when asked about trusting a repository, CBU10 described a range of functions and policy information she found important:

'The staff, the operations, the existence, the process of how you make data available, what are the restrictions...Just the detailed information about the repository.'

In terms of metadata, CCU02 clearly differentiated between trusting the repository to fulfil its role and ensuring accuracy in the data. At the same time he asserted that when a repository fulfilled its role, it was easier to discern inaccuracies in the data:

'They're very keen on producing the comprehensive metadata. And it's not that I trust each research [datum]... but I trust that the metadata is there for me to go back and check...on my own. I don't give [the archaeological repository] a sort of blanket trust that all the data in there is correct...they provide enough metadata for me to check that on my own...I sort of trust going there because I know that I can find the information I need to validate it.'

CBU14 linked selection to data quality, which she saw as a marker of trustworthiness of the repository:

'I mean, I wouldn't use a scale from a very overtly conservative or overtly liberal organization that was involved in other kinds of political activities outside of collecting data because that would make you question what the goal is in collecting that data. So that would, I think, affect sort of the trustworthiness of repositories, at least in my field.'

In short, reusers did recognize trustworthy actions by repositories; however, these actions alone were only part of the trust decision. Next, we examine dimensions of stakeholder trust to see how it influences the construction of trust in repositories.



Trust in Repositories

Our interviewees demonstrated classic attributes associated with the three core dimensions of stakeholder trust. Of those interviewed, 22.7%, 57.6%, 12.1% discussed trust in the repository, structural assurance and social factors respectively (see Table 2). As previously stated, four factors comprise trust in a repository: identification, benevolence, integrity and transparency. While our interviewees discussed repository trust in terms of these factors, they most frequently discussed transparency (15%) followed by identification.

Identification, the internalization of stakeholder interests, has an important parallel with ISO TRAC. Identification could be seen as a metric for how well the designated community perceives that the repository understands its needs. Six interviewees discussed identification. Two made statements about the importance of identification specifically when discussing trust (see Table 2); another four talked about the importance of identification when discussing the types of added value repositories provide. For example, CCU21 asserted:

‘Data migration is critical...I believe that a good repository has to be field-centric. That is to say, if you’re going to put archaeological data into a repository, that repository has to understand archaeology because when the data must be migrated, they need to be able to look at it and to understand whether or not the migration is correct. It’s one thing to say we got all the bits moved, it’s another thing to say it still makes sense for archaeological data.’

Ten of the archaeologists and quantitative social scientists spoke about the importance of transparency in similar terms and as a direct measure of trust. Archaeologists’ focus on transparency is interesting considering their lack of a scholarly tradition of data sharing and the recent emergence of digital repositories in that field. As one noted:

‘There was a relationship already between the museum and the university. And having to be related to a famous museum that has a reputation, it does make the source more reliable...So knowing that, they developed the work and that they were backing up the information. Also knowing that I have access to the collection itself if I wanted to and that they are explicit about everything that they did. They tell you all the methods that they use. They tell you every single person who wrote down anything. They tell you all the updates that they did with the material. So having that explicit and having that personal relationship with them between my university and the museum.’

A quantitative social scientist, CBU38, characterized data repositories it this way:

‘They’re valuable to me as a researcher because it’s a central place to go. The ones we’ve spoken about are considered to be high quality, so I do trust that because I don’t have to worry

about issues. There's...a certain sense of transparency with what's going on at those places, what their missions are.'

Structural Assurance

Interviewees mentioned all aspects of structural assurance: third party endorsements, preservation and sustainability guarantees, and reputation; however, they emphasized guarantees and institutional reputation. Only one referred to seals (CBU27), a form of third party endorsement.

In discussing trust, approximately 15% (ten interviewees) mentioned preservation or sustainability issues, some linking the two concepts. Along with transparency, guarantees were the second most frequently discussed dimension of trust.

'Long term preservation is important so to know that it's kind of a sustainable practice and it's going to be there in the long run. I mean, one of the concerns that I know comes up with a lot of these repositories is, what happens when the NEH or the NSF funding runs out? Who's going to take care of the collection, who's going to run it? So for using a repository, I'd want to know that there was a long term plan for it. That's really important.' (CCU04)

CBU28 summarized the third aspect of structural assurance, institutional reputation:

'They're the only repository that I know around for individual investigator data. They've existed for a long time, they have incredible reputation for being able to maintain data, keep it well preserved, the issue of preservation is key, and that they go through extensive interrogation of the data to make sure that it is of high enough quality to be allowed to be part of their repository.'

CBU28 made many aspects of institutional reputation apparent, including the longitudinal aspect of observed behavior and links to preservation, and implicitly to other repository processes supporting quality data. At 41% interviewees linked institutional reputation to trust – more than any other single factor.

Social Factors

Social factors signify some type of social influence from a referent group, such as peers, colleagues or advisors. Twelve percent of our interviewees mentioned a diverse set of colleagues, both specifically and generally. Interviewees mentioned peers as well as advisors, but a number of interviewees just mentioned colleagues generically. As noted in the discussion of the social factors, influences can come from peers, mentors, senior colleagues and institutions. For example, CBU19 stated:

'It would be I want to stick with [repository] only until I find out from ... my friends or colleague or an adviser.'

A couple of interviewees did cite mentors as influential in trust decisions, but our interviewees tended to refer to colleagues more in terms of a disciplinary community and general practices.

‘Trust, that would be part of a decade or so there of my own experience with using the data and then the organization’s long history, and then within the profession, it’s very well spoken of. So, largely, informal mechanisms are why I trust [repository name].’ (CBU32)

This expression of social influence was also interesting because of its impersonality. Direct connection or knowing a specific person did not seem necessary in helping reusers to form a repository trust judgment.

Differences and Similarities in Discipline and Levels of Expertise

Archaeologists and quantitative social scientists both talked about the connection between repository functions and trust in repositories. Yet, their discussions had different tenors. Archaeologists were more likely to discuss metadata; quantitative social scientists discussed data processing and the importance of selection (see Table 1).

Repository Functions	Archaeologists (n=22)	Quantitative Social Scientists (n=44)	All (n=66)
Data processing	0	5 (11.36%)	5 (7.58%)
Metadata	4 (18.18%)	0	4 (6.06%)
Provenance	0	1 (2.27%)	1 (1.52%)
Selection	1 (4.55%)	8 (18.18%)	9 (13.64%)
Data processing	0	5 (11.36%)	5 (7.58%)

Table 1. Frequency interviewees linked repository functions and trust.

In Table 2, we compare dimensions of trust from the management and information systems literatures to statements made by the archaeologists and quantitative social scientists. Disciplinary differences emerge here, too. Quantitative social scientists and archaeologists both cited repository transparency as a trust factor, but the archaeologists were twice as likely to do so. Both disciplines also focused on elements of structural assurance; however they identified different aspects. Almost half of the archaeologists (40.91%) talked about preservation or sustainability as major trust inducers. Just over half of the quantitative social scientists (52.27%) mentioned institutional reputation. Social influence in the form of specific colleagues, or more generally the disciplinary community, was more of a factor for quantitative social scientists. Taking a closer look, we saw distinct differences between novice and expert quantitative social scientists. At 65.22% versus 36.36%, novices were more likely than experts to discuss institutional reputation as a trust factor.

Concepts	Archaeologists (n=22)	Quantitative Social Scientists (n=44)	All (n=66)
Stakeholder Trust in the Organisation			
Benevolence	0	1 (2.27%)	1 (1.52%)
Identification	1 (4.55%)	1 (2.27%)	2 (3.03%)
Integrity	1 (4.55%)	1 (2.27%)	2 (3.03%)
Transparency	5 (22.73%)	5 (11.36%)	10 (15.15%)
Social Factors			
Colleagues	1 (4.55%)	7 (15.91%)	8 (12.12%)
Structural Assurance Guarantees			
Preservation /Sustainability	9 (40.91%)	1 (2.27%)	10 (15.15%)
Institutional reputation	4 (18.18%)	23 (52.27%)	27 (40.91%)
Third Party Endorsement	0	1 (2.27%)	1 (1.52%)

Table 2. Frequency interviewees mentioned trust factors.

Discussion

Our discussion focuses on four major findings:

1. Repository functions as indicators of trust,
2. Transparency as a trust factor,
3. Expanding the definition of structural assurance to include guarantees of preservation and sustainability,
4. The effects of discipline and level of expertise.

Interviewees did cite repository functions when discussing trust – a dimension absent from the management and information systems literatures. This finding aligns with Prieto's (2009) identification of several important elements relating to increasing users' perceptions of trust in digital repositories, such as repository policies, consumer services, and systematic process in the repositories that decrease users' uncertainty about the repository's authenticity, integrity and accessibility. This is good news for ISO TRAC. Data reusers appear to be noticing repository functions, particularly data processing, metadata and selection, and have expectations about how these should be handled. ISO TRAC is full of instances where repository actions affect data reusers. Our study provides some evidence that these stakeholders understand this mechanism.

As expected, the stakeholder construction of trust did exhibit different dynamics than trust in for-profit organizations. For example, we found support for retaining transparency as a factor in repository trust. Pirson and Malhotra (2011) dropped

References

- Boersma, M.F., Buckley, P.J. & Ghauri, P.N. (2003). Trust in international joint venture relationships. *Journal of Business Research*, 56, 1031-1042. [doi:10.1016/S0148-2963\(01\)00315-0](https://doi.org/10.1016/S0148-2963(01)00315-0)
- Center for Research Libraries and the Online Computer Library Center (OCLC). (2007). Trustworthy repositories audit & certification: Criteria and checklist. Dublin, OH and Chicago, IL: OCLC and CRL. Retrieved from http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
- Davis, F.D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly*, 13, 319-340. [doi:10.2307/249008](https://doi.org/10.2307/249008)
- Faniel, I.M. & Jacobsen, T.E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work (CSCW)*, 19(3-4), 355-375. [doi:10.1007/s10606-010-9117-8](https://doi.org/10.1007/s10606-010-9117-8)
- Faniel, I.M., Kriesberg, A. & Yakel, E. (2012). Data reuse and sensemaking among novice social scientists. *Proceedings of the American Society for Information Science and Technology*, 49. [doi:10.1002/meet.14504901068](https://doi.org/10.1002/meet.14504901068)
- Gefen, D., Karahanna, E. & Straub, D.W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27, 51-90. Retrieved from <http://www.jstor.org/stable/30036519>
- International Organization for Standardization. (2012). ISO 16363:2012. Space data and information transfer systems - Audit and certification of trustworthy digital repositories. Geneva, Switzerland: ISO.
- Jarvenpaa, S.L. & Tractinsky, N. (1999). Consumer trust in an Internet store: A cross-cultural validation. *Journal of Computer Mediated Communication*. 5(2), 1-36. Retrieved from <http://jcmc.indiana.edu/vol5/issue2/jarvenpaa.html>
- Kelton, K., Fleischmann, K.R. & Wallace, W.A. (2008). Trust in digital information. *Journal of the American Society for Information Science and Technology*, 59(3), 363-374. [doi:10.1002/asi.20722](https://doi.org/10.1002/asi.20722)
- Kimery, K.M. & McCord, M. (2002). Third-party assurances: The road to trust in online retailing. *Journal of Information Technology Theory and Application*, 4(2), 63-81. Retrieved from <http://aisel.aisnet.org/jitta/vol4/iss2/7/>
- Lewicki, R.J. & Bunker, B.B. (1996). Developing and maintaining trust in work relationships. In R.M. Kramer & T.R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research*. Thousand Oaks, CA: Sage Publications.
- Mayer, R.C., Davis, J.H. & Schoorman F.D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709-734. [doi:10.5465/AMR.1995.9508080335](https://doi.org/10.5465/AMR.1995.9508080335)

