Users' Trust in Trusted Digital Repository Content

Devan Ray Donaldson
University of Michigan
School of Information
3339A North Quad, 105 S. State St.
Ann Arbor, MI 48109-1285
devand@umich.edu

ABSTRACT
Scholars who study trust in digital archives have largely focused their attention on the power of certification by third-party audit as a way to communicate trustworthiness to end-users. In doing so, they assume that the establishment of a network of trusted digital archives will create a climate of trust. But certification at the repository level also assumes the trustworthiness of digital objects within a repository; specifically that digital repository objects are authentic and reliable. This paper proposes the use of document-level seals of approval as a means of communicating to end-users about the trustworthiness of digital objects that is commensurate with specific user interaction. Implications of this proposed research stress the importance of assessing the 'real-world' impact of trust signals on users.

Categories and Subject Descriptors
H.1.2 [**Information Systems**]: User/Machine Systems – *human factors, human information processing.*

General Terms
Reliability, Experimentation, Human Factors, Verification.

Keywords
Authenticity, End-Users, Integrity, Trust, Trusted Digital Repositories.

1. INTRODUCTION
Archival scholars state that the trustworthiness (i.e., authenticity and reliability) of digital objects is important to users [5]. Criteria for repository certification include requirements for document level authenticity (i.e., integrity and identity) to ensure that users can be confident that they are interacting with authentic digital objects [14, 15]. Prior empirical research suggests that authenticity is important to end-users [4, 16]. Given that archival scholars, repository certification criteria, and prior empirical research all stress the importance of the trustworthiness of digital objects for end-users, it is surprising that research on how to communicate with end-users about archival trustworthiness is scant. End-users, *those not involved in the creation and preservation of the digital objects they use*, presumably know the least about the creation and maintenance of the digital objects they use, as compared to other classes of users such as creators or preservers. End-users have the greatest amount of uncertainty regarding whether or not a given digital object is authentic and reliable. Digital archivists must somehow provide end-users with information about their authenticity and reliability.

There are two potential ways to communicate with end-users about the trustworthiness of digital objects, specifically by: 1) exposing preservation metadata related to the authenticity and reliability of digital objects to end-users, or 2) using cues or symbols to denote the authenticity and reliability of digital objects for end-users. As a record, preservation metadata can be quite complex, sometimes providing more extensive data than the digital objects for which the preservation metadata were created. Given this, cues or symbols attesting to authenticity and reliability may be a more effective way of communicating to end-users about the trustworthiness of digital objects than exposing end-users to preservation metadata. This paper proposes seals of approval at the document level as one possible way to address this issue.

iPRES2011, Nov. 1–4, 2011, Singapore.

2. BACKGROUND
2.1 Archival Literature, Research and Users' Trust in Digital Objects
According to Duranti [5], archival trust involves two components: authenticity and reliability. Authenticity refers to the idea that a document is what it claims to be. Reliability refers to the idea that a record "can be treated as the fact of which it is evidence" [p. 7], and depends upon the form and procedure of creation for a record. Duranti wrote that both authenticity and reliability are important to users. Essentially, users need to know that a record [pp. 8-9]:

- is the same that was placed in the file by the creator of the file itself, and that it has been preserved in its integrity,
- is the same as the one that was transmitted to its addressee, and has not been manipulated or substituted in the course of the transmission,
- was made under controlled circumstances as part of the regular workflow,
- was made within a reasonable time after the occurrence of the facts it is about, and
- was generated by somebody who was competent to make that specific record, with either duty or the direct interest to make it accurate.

Empirical research on trust in digital objects has focused more on creators and preservers than end-users. MacNeil [9, p. 56] conducted case studies to ascertain which specific elements creators considered essential for verifying a record's authenticity. She also found out about the kinds of procedural controls exercised over systems and the records contained within them which, in the creators' view, support a presumption of authenticity. Donaldson and Conway [3] and Foscarini [7] found that preservers use preservation metadata to validate claims of authenticity for digital objects. Preservation metadata are "the information a repository uses to support the digital preservation process," and typically include some combination of descriptive, structural, technical and/or administrative metadata [12]. Little research has been done to assess whether or not preservation metadata could have trust value for end-users as they do for preservers in validating claims of authenticity for digital objects. This is important to consider because prior empirical research suggests that *end-users do have concerns about authenticity*. In Duff et al.'s [4] study, historians complained about copying errors, stating that such mistakes not only undermined belief in the continuing authenticity of a specific source, but also compromised the credibility of copies of other sources. Zhou [16] found that users of digitized archival materials were more likely to think those materials had been altered and were less confident in their own authenticity assessments than those who interacted with non-digital archival materials. If end-users have concerns about authenticity, how should archivists go about clarifying these concerns? How should archivists attest to the authenticity of the digital objects they preserve and make accessible for end-users? Should preservers provide end-users with preservation metadata because preservation metadata are what preservers use to validate document level authenticity claims? Or should preservers use symbols or cues such as seals of approval to denote the archival trustworthiness of digital objects?

2.2 Criteria for Repository Certification and Users' Trust in Digital Objects

In 2002, the RLG/OCLC Working Group on Digital Archive Attributes (WGDAA) [15] wrote the groundbreaking report entitled *Trusted Digital Repositories: Attributes and Responsibilities*. The working group defined a Trusted Digital Repository (TDR) as "one whose mission is to provide reliable, long-term access [of] managed digital resources to its designated community, now and in the future" [p. i]. The WGDAA also specified three levels of trust to apply to the establishment of TDRs, including [p. 9]: 1) How cultural institutions earn the trust of their designated communities, 2) How cultural institutions trust third-party providers, and *3) How users trust the documents provided to them by a repository*. Regarding the third identified trust level, the WGDAA wrote that users must be certain that a document received is the one requested and that a retrieved document can be verified to be the exact item deposited into the digital repository in the past. The working group recommended message authentication codes signed by trusted institutions and public key encryption systems as ways of addressing these concerns. While prior research suggests that preservers use checksums to establish the authenticity of digital objects [3, 7], research on the impact of such mechanisms on end-users' trust is limited in the literature.

Other closely-related means of establishing the trustworthiness of digital documents include certification of archives. The Archival Workshop Program Committee [1] characterized certification of archives as "[a] method by which an [a]rchive's customers could gain confidence in the authenticity, quality, and usefulness of digitally archived materials" [n. p.]. Subsequent certification standards endow a preservation repository with responsibility to ensure the authenticity of its digital objects through explicit criteria for repository level certification. For example, Trusted Repositories Audit and Certification (TRAC) [14] states in Section B6.10 that any repository that gains trusted status must enable the dissemination of authentic copies of the original or objects traceable to originals. TRAC explicitly states that, "[a] repository's users must be confident that they have an authentic copy of the original object, or that it is traceable in some auditable way to the original object" [p. 41]**.** Section A3.8 [p. 15] specifies that a repository must commit to defining, collecting, tracking, and providing, on demand, its information integrity measurements. Examples of mechanisms designed to address the integrity of digital documents include use of checksums at ingest and throughout the preservation process as well as keeping an explicit, complete, correct, and current record of the chain of custody for all digital content from the point of deposit forward (i.e., provenance). The criteria outlined in Sections A3.8 and B6.10 underscore the idea that part of repository level certification involves establishing the trustworthiness of digital documents, and establishing and maintaining trust in digital documents is accomplished using metadata. Given the importance of the association between repository level certification and document level authenticity and reliability outlined in standards for repository certification, more research needs to be done on how to effectively communicate with end-users about authenticity and reliability of digital objects.

The information needed to address Sections A3.8 and B6.10 of the TRAC criteria for repository certification would be best characterized as preservation metadata. Yet, as a record, preservation metadata can be quite extensive, sometimes more complex than the digital objects for which the preservation metadata were created. Cues or symbols attesting to authenticity and reliability such as seals of approval may be a more effective way of communicating to end-users about the trustworthiness of digital objects than exposing end-users to preservation metadata. Of course, seals of approval should only be granted to digital objects that have certain preservation metadata that can attest to their authenticity and reliability, even if those metadata are not exposed to end-users.

2.3 Research on the Effect of Repository Certification on Users

Little research has been conducted to understand the extent to which third-party audit and certification affect users' perceptions of trustworthiness. The CASPAR Consortium [2] conducted a study asking creators, curators and users of curated digital objects about the most important factors when determining whether to trust a repository. Among the most important factors, according to the study subjects, were: the track record of the repository's ability to curate objects; the repository's preservation of the audit trail for digital objects in its custody; and control of integrity within the repository. The findings are interesting because they indicate three important factors regarding users' trust in repositories that are interrelated and involve the authenticity and reliability of digital objects: how repositories curate digital objects, the metadata repositories collect for their digital objects, and control of integrity for digital objects.

2.4 Seals of Approval

While third-party certification checklists specify that TDRs be transparent in communicating audit results to the public, specific means of conveying information about the authenticity and reliability of digital objects is up to TDRs to decide. Research has shown that many users rely on cues and defer to heuristic rather than systematic processing when making trust judgments of digital objects found on the web [13]. As such, use of cues or signals to denote third-party certification may be an effective way to communicate this type of information and thereby build trust in digital objects with end-users.

Findings from empirical research in Human-Computer Interaction and E-Commerce support the idea that third-party seals of approval enhance users' trust. Fogg et al. [6] conducted a study with 2,500 participants and found that a website won credibility with users by showing seals of approval from known companies. Miyazaki and Krishnamurthy [11] conducted experiments designed to ascertain how online firm participation in Internet seal of approval programs affected consumers. They found that the presence of an Internet seal of approval logo resulted in higher levels of information disclosure and anticipated website patronage for consumers who experience relatively high levels of online shopping risk. Findings from these studies could be used to suggest the need for empirical research regarding the impact that document-level seals of approval could have on users' assessments of digital object trustworthiness.

Harmsen [8] describes a Data Seal of Approval program in which repositories complete an assessment document, undergo audit by a member of the international Data Seal of Approval Assessment Group, and publish the results of this assessment. Afterwards, repositories are allowed to use the logo of the data seal on their websites. To date, research on the Data Seal of Approval is very limited. Mitcham and Hardman [10] conducted a case study in which they outlined issues the Archaeology Data Service (ADS) faced in undertaking the repository certification process that precedes approved use of the seal. They also presented the potential benefits of Data Seal of Approval self-certification. One of the benefits of the Data Seal of Approval, the authors wrote, is enhancing the trust of their users. The effect of the Data Seal of Approval on ADS users was not examined in the case study. Since one of the perceived benefits of seals of approval is to positively influence end-users' trust in digital repositories, research ought to be done to examine the impact of seals of approval on end-users. Further, repository level certification says something specific about the trustworthiness of digital objects within a repository; specifically that digital objects are authentic and reliable. Document level seals of approval may be an appropriate way to communicate with end-users about the authenticity and reliability of digital objects.

3. RESEARCH DESIGN

To address the research question (How does a document-level seal of approval affect users' perceptions of trustworthiness of TDR content?), this paper proposes an exploratory experiment to investigate this phenomenon. The following proposed experiment focuses on digitized books as examples of TDR content.

3.1 Proposed Experiment

3.1.1 Hypothesis

Based upon prior research on seals of approval, this paper hypothesizes that participants will rate digitized books with seals of approval as more trustworthy than books without seals.

3.1.2 Design

This paper proposes use of an experimental design (see Table 1), selecting digitized books (B_n) that either have a seal of approval (denoted by the * symbol in Table 1) or do not. Participants will only see one version of each book. Book information content will be held constant for all conditions, ensuring that any effects would be due to the seals. All books used in this experiment will be randomly selected from a TDR. Seals will be assigned to books from the randomly-selected pool of TDR digitized books.

Treatment		Control
$*B_{1\text{-}10}B_{11\text{-}20}$ n=30	$B_{1\text{-}10}*B_{11\text{-}20}$ n=30	$B_{1\text{-}20}$ N=30

Table 1. Experimental design for assessing impact of document-level seals of approval on users' perceptions of trustworthiness of TDR content.

3.1.3 Participants and Procedure

Who to recruit for an experiment involving users of a TDR depends upon its designated community. Some designated communities are narrowly defined while others are loosely defined. Large-scale repositories that are not discipline-specific typically have very loosely-defined designated communities. This proposed experiment focuses on recruiting a sample of intended users of a TRAC-certified TDR - HathiTrust (HT) (http//:www.hathitrust.org). HathiTrust is based out of the University of Michigan but has over 50 institutional partners. The designated community for this TDR includes not only the students, faculty, and staff of all of its partners, but extends to include anyone with an Internet connection. A good place to start in terms of recruiting subjects for this proposed experiment would be undergraduate and graduate students at one of HT's partner institutions.

Each participant will be randomly assigned to a treatment or control group. To control for order effects, treatment and control groups will be subdivided. Thirty participants (n=30) will be recruited per subgroup to account for the law of large numbers. Participants will be asked to think about conducting a research task in which certain questions would need to be answered regarding eighteenth-century English literature. To simulate the seamless nature of cyberinfrastructure in which TDR content can be found, participants will be told to use the search engine provided to find books that could help them answer a series of questions. Participants will be able to type whatever search terms they choose, but every participant will be provided with the same set of search results (just in a different order). Half of the treatment group will see books with seals of approval added to their search result listing (odd-numbered) and the other half of the treatment group will see seals accompanying search result listings for even-numbered books. Each participant will assign a

trustworthiness rating (e.g., on a 5-point likert scale with 1 being not trustworthy at all and 5 being completely trustworthy) for each of the books they select.

4. EXPECTED OUTCOMES

Archival scholars, repository certification criteria, and prior empirical research suggest that end-users care about the archival trustworthiness of digital objects. So the question then becomes how to communicate with end-users about the trustworthiness of digital objects. This paper has argued for research to explore the impact of document-level seals of approval on users' perceptions of trustworthiness of TDR content. Empirical results that support the hypothesis that document-level seals of approval increase users' trust in digital objects would suggest that seals aid users in the way in which third-party certification was intended. Empirical results that fail to support this hypothesis would suggest that document-level seals of approval do not aid users in making trust judgments for digital objects and would need to be reexamined.

In an aggregated search environment, TDR content, which by definition has been upheld to best practices for authenticity, is listed alongside content in search results from other sources, which may or may not be upheld by the same standards. TDR administrators and designers need to develop effective ways of communicating with users about the trustworthiness of TDR content. This is a challenge, but if addressed, it could be of great benefit for users.

5. ACKNOWLEDGMENTS

I would like to acknowledge Kathleen Fear, Paul Conway, Paul Resnick, Eric Cook, Maciej Kos, Tracy Liu, Ann Zimmerman and the Archives Research Group at the School of Information for their comments and suggestions on previous drafts of this paper.

6. REFERENCES

[1] Archival workshop on ingest, identification, and certification standards. 1999. National Archives and Records Administration, http://nssdc.gsfc.nasa.gov/nost/isoas/awiics/ (accessed 10 August 2011).

[2] CASPAR Consortium. 2009. Report on Trusted Digital Repositories. Technical Report.

[3] Donaldson, D. R., and Conway, P. 2010. Implementing PREMIS: A Case Study of the Florida Digital Archive, *Library Hi Tech* 28(2): 273-289.

[4] Duff, W., Craig, B., and Cherry, J. 2004. Historians' Use of Archival Sources: Promises and Pitfalls of the Digital Age, *The Public Historian* 26(2): 7-22.

[5] Duranti, L. 1995. Authenticity and Reliability: The Concepts and their implications, *Archivaria* 39: 5-10.

[6] Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., and Tauber, E. R. 2003. How do users evaluate the credibility of web sites?: A study with over 2,500 participants. Paper presented at Proceedings of the 2003 conference on Designing for user experiences, San Francisco, California.

[7] Foscarini, F. 2008. "Cultures of Trust: Legal, Technical, and Archival Perspectives on the Use of Digital Signature Technologies," *Lecture Notes in Informatics (LNI),* vol. P-133: 37-47.

[8] Harmsen, H. 2008. Data seal of approval - assessment and review of the quality of operations for research data repositories. Paper presented at iPres, The British Library, http://www.bl.uk/ipres2008/presentations_day2/34_Harmsen.pdf (accessed 10 October 2010).

[9] MacNeil, H. 2000. "Providing Grounds for Trust: Developing Conceptual Requirements for the Long-term Preservation of Authentic Electronic Records," *Archivaria* 50: 52-78.

[10] Mitcham, J. and Hardman, C. 2010. ADS and the Data Seal of Approval – case study for the DCC, Digital Curation Centre, http://www.dcc.ac.uk/resources/case-studies/ads-dsa (accessed 29 September 2011).

[11] Miyazaki, A. D., and Krishnamurthy, S. 2002. Internet Seals of Approval: Effects on Online Privacy Policies and Consumer Perceptions. *Journal of Consumer Affairs* 36 (1): 28-49.

[12] PREMIS Editorial Committee. 2011. PREMIS data dictionary for preservation metadata *version 2.1.* Washington, DC: Library of Congress, http://www.loc.gov/standards/premis/v2/premis-2-1.pdf (accessed 30 August 2011).

[13] Rieh, S. Y. 2002. Judgment of information quality and cognitive authority in the web. *Journal of the American Society for Information Science and Technology* 53 (2): 145-61.

[14] RLG-NARA Digital Repository Certification Task Force. 2007. *Trustworthy repositories audit and certification: Criteria and checklist.* OCLC and CRL, http://www.crl.edu./sites/default/files/attachments/pages/trac_0.pdf (accessed 13 October 2010).

[15] RLG/OCLC Working Group on Digital Archive Attributes. 2002. *Trusted digital repositories: Attributes and responsibilities.* Mountain View, CA: RLG, http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf (accessed 13 October 2010).

[16] Zhou, X. 2005. A Comparison of Users' Response to Digital versus Physical Archival Material, Paper presented at the Society of American Archivists Annual Meeting, New Orleans, LA: 1-11.