Building Trust in Digital Preservation

NIH Trustworthy Data Repository
Workshop April 8-9 2019

Jonathan Crabtree

Building Trust

In 1995, the Commission on Preservation and Access and the Research Libraries Group commissioned a taskforce on the archiving of digital information. Two of the recommendations of that report point to the need for trusted repositories and a way to certify them ("Preserving Digital Information," 1996):

> *A critical component of the digital archiving infrastructure is the existence of a sufficient number of trusted organizations capable of storing, migrating and providing access to digital collections.*

> *A process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information.*

Trust Can Be Complicated

- Some research has shown that researchers in the social sciences often relate an organization's trustworthiness to its reputation (Yakel, Faniel, Kriesberg, & Yoon, 2013a).
- Ross and McHugh (2006) highlight the role of evidence in establishing a trust relationship between a community and the archive.

How to Build Trust?

- *How is trust represented in evidence provided by repositories?*
- *On what foundational models should repositories base their policies?*
- *What certification and audit methods are available?*
- *Using CoreTrustSeal to demonstrate trust!*

Perception of Trust

- (A Yoon, 2015). Yoon notes that:

"Trust is not a new concept in the field of archives, which traditionally is responsible for the curation of information. (Speck, 2010) said the concept of trust has been considered an integral component in the existence of archives, which made people expect a large volume of scholarly literature to be produced on the subject. However, Speck (2010) argued that discussions of trust have been limited either to discussions related to the ethics of the archival professions (Dingwall, 2004) or to the notion of "trusted" digital information and repositories. While archival and curation communities have understood the term trust as a synonym for "reliable" and "authentic" in relation to curation activities (RLG/OCLC, 2002, p. 8), little research exists on how (potential) users perceive the concept of trust"

Three Domains of Trust

- Yakel et al (Yakel, Faniel, Kriesberg, & Yoon, 2013b) break down work on trust into three domains: Stakeholder trust in organizations, structural assurances, and social factors.

Organizational Trust

- In the work by Adolfo Prieto (Prieto, 2009) he finds that:

– While digital repositories may be trustworthy because of adherence to technological standards, accepted practices, and mechanisms for authenticating the authorship and accuracy of their content, it is ultimately their respective stakeholders – both those who deposit and use content – whose perceptions play a central role in ensuring a digital repository's trustworthiness.

Critical User Community

- As noted by Yoon (Ayoung Yoon, 2014), (Prieto, 2009) states: ‘‘User communities are the most valuable component in ensuring a digital repository’s trustworthiness’’ (p. 603). In the end it is this user community that needs to feel the repository is trusted in addition to the reviewer approving the evidence.

Gaining Users Trust

- Initial research by Donaldson et al (Donaldson, Dillo, Downs, & Ramdeen, 2017) show that repositories that have sought peer reviewed certification status through the Data Seal of Approval (“Home | Data Seal of Approval,” 2014) perceive the status as having many benefits.
- A critical component was transparency

Great Common Language

- The OAIS reference model has helped repositories describe the processes, technologies, and workflows they use to curate and preserve data under their care (Crabtree, 2009). This has been very valuable to the repository community, but the model was designed to do more (Giaretta, 2012).
- The ultimate goal was to protect these digital assets and to ensure preservation for future generations of researchers.

Defining Designated Community

"The OAIS should then make a decision between maintaining the minimum Representation Information needed for its Designated Community or maintaining a larger amount of Representation Information that may allow understanding by a larger Consumer community with a less specialized Knowledge Base, which would be the equivalent of extending the definition of the Designated Community. Over time, evolution of the Designated Community's Knowledge Base may require updates to the Representation Information to ensure continued understanding." (CCSDS, 2002, p 2.4)

Defining Trusted Digital Repository

- RLG-NARA Task Force in 2002
- This work by RLG was the inspiration and basis of *The Trustworthy Repositories Audit & Certification: Criteria and Checklist* (TRAC) (Center for Research Libraries (CRL), 2007). The TRAC metrics are closely aligned with the OAIS reference model/ ISO14721 standard.
- The TRAC checklist is the foundation of many repository audit and review processes and specifies the policies and procedures in an archive or repository that need assessment and that should be documented as evidence of compliance.

Building Consensus

Center for Research Libraries - Ten Principles, 2007

1) *The repository commits to continuing maintenance of digital objects for identified community/communities.*
2) *Demonstrates organizational fitness (including financial, staffing, and processes) to fulfill its commitment.*
3) *Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.*
4) *Has an effective and efficient policy framework.*
5) *Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.*
6) *Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.*
7) *Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.*
8) *Fulfills requisite dissemination requirements.*
9) *Has a strategic program for preservation planning and action.*
10) *Has technical infrastructure adequate to continuing maintenance and security of its digital objects.*

Transparency Builds Trust in Preservation

- Preservation planning is required
  - Technology watch
  - Designated community changes
  - Ongoing risk analysis
- Technology is required
  - Bit level preservation
  - Persistent identifiers
  - Fixity Checks
  - Format migration
  - Distributed storage
  - Automated metadata generation
  - Defined metadata standards
  - Information security systems
  - .....
- Digital preservation can not be trusted without transparent policies

Thank You