

EVIDENCE FOR TRUSTED DIGITAL REPOSITORY REVIEWS:
AN ANALYSIS OF PERSPECTIVES

Jonathan David Crabtree

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Information and Library Science.

Chapel Hill
2020

Approved by:

Helen Tibbo

Christopher Lee

Nancy McGovern

Arcot Rajasekar

Katherine Skinner

© 2020
Jonathan David Crabtree
ALL RIGHTS RESERVED

ABSTRACT

Jonathan David Crabtree: Evidence for Trusted Digital Repository Reviews: An Analysis of Perspectives
(Under the direction of Dr. Helen R. Tibbo)

Building trust in our research infrastructure is important for the future of the academy. Trust in research data repositories is critical as they provide the evidence for past discoveries as well as the input for future discoveries. Archives and repositories are examining their options for trustworthy review, audit, and certification as a means to build trust within their content creator and user communities. One option these institutions have is to increase and demonstrate their trustworthiness is to apply for the CoreTrustSeal. Applicants for the CoreTrustSeal are becoming more numerous and diverse, ranging general purpose repositories, preservation infrastructure providers, and domain repositories.

This demand for certification and the subjective nature of decisions around levels of CORETrustSeal compliance drives this dissertation. It is a study of the review process and its veracity and consistency in determining the trustworthiness of applicant repositories. Several assumptions underlie this work. First, audits and reviews must be based on evidence supplied by the repository under scrutiny; second, and not all reviewers will approach a piece of evidence in the same fashion or give it the same weight. Third, the value and veracity of required evidence may be subject to reviewers' diverse perspectives and diverse repository community norms.

This research used a thematic qualitative analysis approach to identify similarities and differences in CoreTrustSeal reviewers' responses during semi-structured interviews in

order to better understand potential subjective differences among respondents. The participants' non-probabilistic sample represented a balance in perspectives across three anticipated categories: administrator, archivist, and technologist. Themes converged around several key concepts. Nearly all participants felt they were performing a peer review process and working to help the repository community and the research enterprise. Reviewers were questioned about the various CoreTrustSeal application requirements and which ones they felt were the most important. No clear evidence emerged to indicate that variations in perspectives affected the subjective review of application evidence. The same categories of evidence were often selected and identified as being critical across all three categories (i.e., administrator, archivist, and technologist). Many valuable suggestions from participants were recorded and can be implemented to ensure the consistency and sustainability of this trusted repository review process. These suggestions and concepts were also very evenly distributed across the three perspectives. The balance in perspectives is potentially due to participants' experience levels and their years of experience in various positions, holding many responsibilities, within the organizations they represented.

Keywords: trusted digital repository, policy, CoreTrustSeal, certification, digital preservation, digital curation, digital archive, compliance, standards

ACKNOWLEDGEMENTS

As I ponder on my many years of educational and work experiences, I realize that my Lord and Savior Jesus Christ has blessed me with so many people to thank. This starts with my wonderful parents, Frank and Alice Crabtree. I have always been encouraged to work hard and reach for my goals, but they have also been there to support me my entire life. When I needed any level of assistance, they were there for me. This was surely the case when hours of proofreading and reviewing this document were needed. I am truly thankful for having the best parents in the world. My academic journey began many years ago, and they have supported me the entire time. My family is very important to me, and my children have also supported my academic efforts. As they continue with their own academic and life experiences, I hope to be there for them as my parents have been for me.

The University of North Carolina at Chapel Hill has been more than my academic home. It has also been my professional home for more than 27 years. While working full-time at this great university, I have completed my undergraduate and graduate programs. This would not have been possible without the support of the Odum Institute for Research in Social Science and its great administration and staff. I have had the pleasure to work with many people through the years; there are too many to name, but a few stand out as not only colleagues but great friends. I would have never completed this dissertation without the support of my great friend Peter Leousis, whose guidance as a deputy director was more than I could have hoped for. Yet it is his deep caring for others, including myself, that makes him a special person. I am so lucky to call him a friend. His love and support helped me through many difficulties, and for that I am forever

grateful.

We both had the pleasure of working for great directors over the years. During my tenure, I have had the pleasure of working for Dr. John Reed, Angel Beza, Dr. Ken Bollen, Dr. Tom Carsey, Dr. Ashok Krishnamurthy, and presently Dr. Todd Bendor. Every single one of these great men deserve many thanks for encouraging me as well as other staff members to continue our education and grow as members of the Odum Institute. Dr. Bollen gave me the first chance to explore my academic potential and lead projects often only designated for staff who already held their doctorate. He allowed me to explore my potential and grow as a scholar and a person, and for that I am truly thankful.

No one is more important for the success of a student than their mentor. Even before she was my advisor, I called Dr. Helen Tibbo both a colleague and a friend. Her guidance has enabled me to be successful in both my professional and academic life. I will always be grateful for her advice and support over the many years we have been colleagues. This research would not have been possible without her intelligence and insights. I have also had the pleasure of working on grants and projects with each member of my committee. Through these great experiences, they have also become my close friends, and their insights and knowledge led me down this research path. Through the detailed perspective and insights of Dr. Cal Lee, I have grown as a scholar and a colleague. His support during this process and in my professional life at the University of North Carolina has been invaluable. From the early days of DataPASS and Educopia, I have also had the pleasure of working with Dr. Nancy McGovern and Dr. Katherine Skinner. It has always been such a pleasure to work with them, and this research experience has been no different. They both helped shape my professional beliefs and philosophies, and for that I will be forever grateful. I have also had the great pleasure of working with Arcot Rajasekar on

several projects, from DataNet to DataBridge, as well as developing training workshops for Data Science. His deep technical knowledge and foundation in data management systems have kept me focused on the correct way to leverage technology to curate and preserve research data. I have been very fortunate to benefit from great leaders in our field. For this I am very thankful.

I also reserve a large amount of gratitude for my friends, colleagues, and study participants within the CoreTrustSeal reviewer community. I had the great fortune to meet with each one personally, and I have learned so much from each and every one. The dedication to the scientific research community is unsurpassed in this group, whose members spend countless unpaid hours supporting the effort to ensure that future generations of researchers have access to quality data. I am so proud to be a part of this community. To each, I owe a great deal of thanks.

All of these great people were not accidents in my life. I have to thank God for blessing me with all these intelligent and giving people. God has given me the resources, knowledge, and faith to continue during times that were difficult. This is proof that all things are possible if you have faith in God and follow His path. To God, I owe everything.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER 1: INTRODUCTION AND PROBLEM STATEMENT	1
CHAPTER 2: LITERATURE REVIEW	5
The Road to Trusted Digital Repository Certification.....	5
Building Trust	6
Foundations for Digital Repositories	9
Critical Preservation Areas for Examination	12
Organizational Policy Built on Principles.....	13
Review Methods and Certification Opportunities for Repositories.....	18
Certification Next Steps.....	24
Tools, Services, and Applications Literature	26
Breaking Down the Options.....	28
Training and Education.....	30
Front End and Ingest Processors.....	32
Preservation as a Storage Service	33
Full Technological Service Systems	34
National Infrastructures	36

Policy-Based Approach	37
CHAPTER 3: ANALYTICAL APPROACH AND METHODS	39
Methods Introduction.....	39
Philosophical and Analytical Approach.....	40
Methods Overview.....	42
Initial Work and Tool Selection.....	42
Stage One Work.....	46
Stage Two Work	56
Study Population and Participant Selection.....	56
Design of Instrument.....	58
Interview Procedures	62
Determining End of Interview Process	63
Stage Three Work	64
CHAPTER 4: RESEARCH RESULTS	67
Reviewers' Perspectives and Perceived Roles.....	68
Exploring Potential Impacts on Evidence Review.....	72
Most Important Evidence.....	73
Best Forms of Evidence	77
Advice to New Applicants	79
Reviewers' Challenges and Perspectives on the CTS Process	81
Biggest Challenge for Reviewers.....	82
Strengths of CoreTrustSeal	84

Biggest Problem Facing CoreTrustSeal.....	86
Suggested Changes to CoreTrustSeal Process.....	90
Research Results Summary.....	94
CHAPTER 5: DISCUSSION AND CONCLUSIONS	96
Importance of Reviewer Experiences	96
Challenges of Detailed Evidence versus Concise Evidence	97
CoreTrustSeal Process Suggestions.....	99
Building Training Programs for Reviewers.....	101
Study Impact and Suggestions for Repositories	103
Limitations of the Study.....	106
Future Studies	107
APPENDIX A: REPOSITORY APPLICATIONS TARGETED	109
APPENDIX B: CORETRUSTSEAL REQUIREMENTS.....	113
APPENDIX C: RESEARCH QUESTIONS AND SEMI-STRUCTURED INTERVIEW INSTRUMENT.....	118
REFERENCES	121

LIST OF TABLES

Table 1. Reviewers' Perspectives Data.....	68
Table 2. Most Important Requirements	74
Table 3. Advice for New Applicants	80
Table 4. Biggest Challenges for Reviewers.....	82
Table 5. Biggest Problems Facing CTS.....	86
Table 6. Suggested Changes to the CTS Process.....	90
Table 7. The TRUST Principles.....	104

LIST OF FIGURES

Figure 1. Digital preservation tool grid.....	32
Figure 2. Typical CoreTrustSeal requirement	43
Figure 3. Evidence required for CoreTrustSeal application	44
Figure 4. Predefined codes.....	45
Figure 5. Evidence Codes	47
Figure 6. Requirement Codes	48
Figure 7. Word Cloud	49
Figure 8. Judgment codes	51
Figure 9. Word tree example	52
Figure 10. Code matrix example.....	53
Figure 11. Code intersections	54
Figure 12. Disciplinary differences	55
Figure 13. Provisional codes.....	60
Figure 14. Reviewer self-identified role	70
Figure 15. Code co-occurrence model for repository overview	75
Figure 16. Code co-occurrence model notice of perspective influence.....	76
Figure 17. Best kind of evidence	78
Figure 18. Strengths of CTS	84
Figure 19. Problems facing CTS.....	88
Figure 20. Potential changes to CTS.....	92

CHAPTER 1: INTRODUCTION AND PROBLEM STATEMENT

Building trust in our research infrastructure is important for the future of the academy. Trust in research data repositories is critical as they provide the evidence for past discoveries as well as the input for future discoveries. One approach to building trust in repositories and their data is performing trustworthy digital repository audits. The concept of auditing has been accepted as a required practice for hundreds of years. Signs of audit processes are evident throughout ancient history, even in Mesopotamia and Greece (RCV Academy, 2020). Formal audit and repository certification are relatively new for the archive community, but the use of audits for other critical services ranging from security to information systems are becoming common and are required by some communities (RCV Academy, 2020). There are many roles to be filled within organizations preserving digital objects for reuse by a community. Management holds the ultimate responsibility for the preservation of the objects, but systematic and independent audits can be useful in determining the details of the process and identifying potential weak points. Independent audits in other industries have shown the ability to reinforce shared responsibility for staff members participating in the processes (Kueppers & Sullivan, 2010). According to the financial industry, the external auditor's responsibility is to obtain assurances that the processes are free from misstatements and that they are trustworthy (Kueppers & Sullivan, 2010). Continuous reevaluation is an important part of managing an archival collection and applying independent reviews to these processes on a regular basis can help archives and repositories fulfill their responsibilities better.

Institutions seeking certification for trustworthy repository status face a diverse array of pressures to provide quality evidence of standards compliance and professional competencies (Ross, 2007). Research is needed to inform the review process as well as the evidence required to comply with emerging metrics. Accounting firms have practical guides for measuring the effectiveness and efficiency of audits (IIA, 2010; Pett et al., 2014). As we develop larger national digital platforms and more institutions plan to preserve our national digital resources, it is critical that we make informed decisions relating to preservation infrastructure.

Archives and repositories are interested in building trust. These organizations are examining their options for trustworthy review, audit, and certification. One option these groups have is to apply for the CoreTrustSeal (CoreTrustSeal, 2019). Applications for the CoreTrustSeal increased following the merger of the Data Seal of Approval and the World Data Seal. Repositories have also begun to seek ISO 16363 (Audit and Certification of Trustworthy Digital Repositories) certification. The Primary Trustworthy Digital Repository Authorization Body (PTAB) recently certified the first two repositories against ISO 16363. Applicants for the CoreTrustSeal have become more diverse. In addition to traditional domain repositories, they are now coming from general purpose repositories and preservation infrastructure providers.

With this expanded number of reviews taking place, reviewers and repositories face challenges with the review process, starting with repository self-audit and preparation, all the way through to the awarding of certification. The subjective nature of decisions around levels of compliance lies at the heart of this endeavor and drives this dissertation. Audits and reviews must be based on evidence supplied by the repository under scrutiny.

Further complicating matters is the fact that one cannot assume that all reviewers will approach a given piece of evidence in the same fashion or give it the same weight. The value and veracity of required evidence could be subject to the diverse perspectives of the reviewers and

diverse repository community norms. The archive community needs to understand how a wide array of reviewers with differing professional and educational backgrounds will judge such evidence. Reviewers with technology backgrounds could perhaps view technology requirements differently than those with archival backgrounds. In addition, the assigned role perceived by the reviewer could also impact the results. If they view themselves as a strict auditor their responses could be different than if they felt they were a peer reviewer.

Now that many CoreTrustSeal applications have been approved, what can we learn from the reviewers of these applications? This dissertation seeks to answer the following research questions:

RQ1: Does the reviewer's perspective impact their evaluation of trustworthy digital repository evidence? If so, what are the nature of these differences?

RQ2: Does CoreTrustSeal reviewers' perception of the application strengths vary given their perceived role in the process?

The evidence repositories provide for CoreTrustSeal reviews is guided by 16 defined requirements that range from policies describing a repository's mission to details of technological infrastructure used to preserve their content. Although such evidence comes from diverse repositories across many disciplines, each must show their level of compliance with the standard. Reviewers evaluate this evidence based on the trustworthy digital repository review instrument's guidance in addition to their educational and professional training. Certifying bodies will presumably select reviewers from the pool of experienced data archive professionals and specialists around the world. The potential exists for reviewers to have diverse professional backgrounds, such as information technologists, memory institution administrators, and digital repository archivists. Each of these groups may have differing perspectives on the evidence required for trustworthy repository audit and certification. For this study technologists are defined

as those reviewers with educational degrees in computer science or who hold positions that manage technological systems. Archivists are defined as those reviewers that hold degrees in information and library science. Administrators hold positions managing organizations and help define policy for their organizations.

Some evidence required for certification focuses on the repository's ability to articulate high-level policies that guide their institution. A well-articulated mission statement that focuses on preservation or identification of the designated collection that should be the target of the review fall in this category. Other evidence is more technical in nature and describes how the repository performs data migrations or replicates content for geographical diversity. The analysis of some pieces of evidence may prove more subjective depending on the background and training of the reviewer. Research is needed to provide guidance for both reviewers and repositories seeking certification to make the process as efficient and effective as possible.

CHAPTER 2: LITERATURE REVIEW

The Road to Trusted Digital Repository Certification

Audits and certifications are common practices for industry and business around the world. Evidence of financial and business audits has been seen throughout early history (RCV Academy, 2020). Financial audit practices have been evolving and responding to changes in business practices for many years (Byrnes et al., 2018). As transactional data have increased in volume and size, these processes must change, and the standardization and certification of auditors are required in this dynamic world. Professional organizations provide guidance and best practices based on standards (Institute of Internal Auditors, 2020). Having these controls and organizations to maintain them as best practices is important for the processes to be trustworthy. The effectiveness of these audits is measured against the achievement of the stated goals of the organization.

A financial or business audit is similar to the potential audit of archives and repositories. They both have a wide array of internal and external stakeholders (IIA, 2010). Yet does acquiring certification thorough audits benefit archives and repositories? The results are often mixed, although some research has shown that companies gain a competitive advantage over other groups that do not have certification, especially in the case of ISO 9000 certifications (Anderson et al., 2009). In some industries, it is clear that certification does gain a competitive advantage (Noriah, 2006), but for the communities that rely on archives and repositories to preserve data, it is the quality of the process that is the most important. Some research has also shown that certification does increase the quality associated with the reviewed process (Withers

& Ebrahimpour, 2000). The key to satisfying the needs of the communities served by archives and repositories is to gain their trust. Trust is a complicated concept, and building that trust is critical.

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 pointed 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.

The word *trusted* or *trust* shows up in both of these recommendations. This commission was operating during the early days of the OAIS reference model development, and its report described and defined the challenges of archiving digital materials, serving as a foundation of requirements for future audit checklists and risk assessments (“Preserving Digital Information,” 1996). It is also noteworthy for discussing the importance of stakeholder trust for successful repositories.

Building trust is a complicated issue. Some research has shown that researchers in the social sciences often relate an organization’s trustworthiness to its reputation (Yakel et al., 2013). Ross and McHugh (2006) highlighted the role of evidence in establishing a trust relationship between a community and the archive. Document management systems play a key role in providing archives’ ongoing, updated, and accurate documentation as evidence for

trustworthiness. Archives are asked to provide documentation showing that this evidence is integrated into employees' official job descriptions and, thus, the required workflows (Ross & McHugh, 2006).

This literature review focuses on the certification of trustworthy repositories and, specifically, the evidence that repositories present to attain these certifications. The review will investigate areas pertaining to the following questions:

- How is trust represented in evidence provided by repositories?
- What certification and audit methods are available?
- What are the differences between these methods?
- What are some potential technological tools that repositories could use to demonstrate trustworthiness?

The review of these questions will help repositories prepare for future reviews of trustworthy status, help build trust within their communities, and inform more research on the application of trustworthiness standards. Many stakeholders will benefit from the certification of repositories.

In 2014, the Research Data Alliance (RDA) working group on Repository Audit and Certification wrote:

Repository certification is important because it promotes trust and confidence in the usability and persistence of shared data resources. It also helps repositories improve their practices and procedures. However, the value of certification is not apparent to all communities; more work needs to be done to clarify the problem that certification solves and to demonstrate that certification is worth the associated effort. (RDA, 2018b, p. 1)

Trust in our institutions, especially repositories and archives, has always been an expectation. One of the most well-rounded conversations around the definition of “trust” is contained in Ayoung Yoon’s (2015, p.7) dissertation “Data Reuse and Users’ Trust Judgments: Toward Trusted Data Curation”:

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.

Yakel et al. (2013) categorized work on trust into three domains: stakeholder trust in organizations, structural assurances, and social factors. Although all three certainly pertain to trusted certification, the first category of organizational trust explored by Pirson and Malhotra (2011) points to a key concern for repositories seeking trustworthy status. Prieto (2009, p. 001) found 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.

Several authors working in this space (Bak, 2016; Yakel et al., 2013; Yoon, 2014) utilized the official definition of trust to begin the conversation.

‘Trust’ is defined by the Merriam Webster Dictionary as: assured reliance on the character, ability, strength, or truth of someone or something... one in which confidence is placed... a charge or duty imposed in faith or confidence or as a condition of some relationship... something committed or entrusted to one to be used or cared for in the interest of another.

Bak (2016, p.376) noted that, using this pure definition, “most cultural institutions are already trusted,” so more complete definitions of trust are needed. Prieto explained that “user communities are the most valuable component in ensuring a digital repository’s trustworthiness” (2009, p. 603). Ultimately, it is this user community that needs to feel the repository is trusted in addition to the reviewer approving the evidence.

To gain the trust of their user communities, many repositories believe that certification and audits of trustworthy status are helpful. Initial research by Donaldson et al. (2017) showed that repositories that have sought peer-reviewed certification status through the Data Seal of Approval (2018b) perceive the status as having many benefits. Noteworthy in this context are the findings related to the perceived value of building the confidence of their stakeholders, demonstrating their transparency to users, and raising awareness about the importance of digital preservation. Such efforts help build the confidence of users in the repositories they visit. In particular, the notion of transparency is critical and is directly tied to the evidence provided in trustworthy reviews. Several participants in Donaldson et al.'s (2017) study mentioned the benefit of demonstrating transparency. One of the most important components of providing evidence for any of the repository audit and review schemes is to provide public documentation of policies containing the evidence. In many cases, it is the repository's high-level policies that will point to low-level actions that ensure the "trustworthy" status. For a repository to be trusted by any auditing board, it must have a solid understanding of high-level preservation and organizational policies grounded in the foundational principles for digital repositories defined in the OAIS reference model (Consultative Committee for Space Data Systems, 2012).

Foundations for Digital Repositories

Research funders and governments are encouraging—and, in many cases, requiring—research data to be shared (American Education Research Association, 2013). Archives, libraries, and repositories are the natural recipients of these data, and they are increasing skill sets and infrastructures to help handle the increase in deposited data (Weber et al., 2012). For many years, archives and domain repositories have been stewards of our nation's research data, and some have developed standard policies to help provide interoperability as well as protect

these valuable assets better (DataPASS, 2018).

The foundations of a data archive require two critical pieces. Practical organizational curation and preservation policies are the first line of defense in explaining trustworthy status. The policies should be well-documented, inclusive of all areas of responsibility, and presented in a public manner. These policies should be grounded in community best practices and adhere to the archives community of approved standards. Repositories often find writing sound policies to be a challenging part of the trustworthy repository process and fall back on technological solutions not grounded in their mission, goals, and user needs.

This is not to say that repositories do not need technological solutions, but rather that the policies should drive the selection and implementation of the technology. In turn, the automated technological infrastructure and workflows designed to preserve the digital assets for delivery to the community at a future time should support these high-level policies. These technological tools need to be applied during all phases of archival processes, from ingest to dissemination, yet all too often repositories rely too heavily on technology to solve portions of the archival workflow, such as storage, without adequate documentation or policies that define these trustworthy efforts.

Technological components and curation workflows can be complicated to describe, especially to those outside the archive community. Work on the Open Archival Information Systems (OAIS) Reference Model (Consultative Committee for Space Data Systems, 2012) started in April 1994, when a working group was developed related to “archiving space data” (Lee, 2010). This reference model has helped repositories describe the processes, technologies, and workflows they use to curate and preserve data under their care (Crabtree, 2009). Such efforts have been 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.

The Research Libraries Groups and OCLC produced a checklist to guide the community called “Trusted Digital Repositories: Attributes and Responsibilities” (TDR; Research Libraries Group [RLG], 2002, p. 5). This document helped the digital preservation community define what TDR should do and what metrics could be used to evaluate the performance of organizations embarking on the long-term stewardship of digital information. RLG’s work was the inspiration and basis of *The Trustworthy Repositories Audit & Certification: Criteria and Checklist* (RLG/NARA Task Force on Digital Repository Certification, 2007).

The TRAC metrics are closely aligned with the OAIS reference model’s ISO 14721 standard, as was the TDR" - TRAC aligned with TDR intentionally and explicitly. The TRAC checklist

divides the functional requirements into three sections:

- Organizational infrastructure
- Digital object management
- Technologies, technical infrastructure, and security

Each of these areas is broken down in detail to their individual requirements, which can also be grouped into the previously discussed required high-level policy and technological infrastructure.

All three of these functional requirements demand high-level policies for trustworthy status.

Technological infrastructure alone or undocumented information cannot fulfill these requirements. In many cases, high-level policy must point to low-level actions performed by technological infrastructure and tools.

These requirements are based on the relevant International Standards Organizations¹ (ISO, 2018a) standards for quality assurance in an organization (ISO, 2018b), the now defunct standard for data security and information management systems (ISO, 2005), records management best practices (ISO, 2001), and finally the critical OAIS Reference Model (ISO,

¹ The TRAC standard has been updated to reflect changes in international standards.

2018c).

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 needs assessment and should be documented as evidence of compliance. For a repository or archive to declare that it is OAIS-compliant, the RLG–NARA task force expects these checklist items to be addressed. These checklist items provide the foundations for many trustworthy audit and review standards, ranging from the CoreTrustSeal² to the ISO 16363 standard (ISO, 2012). The later ISO standard required some changes to the TRAC community best practice document for alignment with ISO principles, but overall, the goal and aims of the criteria are similar to the aforementioned TRAC checklist. These will be reviewed in more detail further into our discussion. To understand these items in detail, it is necessary to look at a few of the critical components that an archive needs to define.

Critical Preservation Areas for Examination

Around the same time that TRAC was being finalized in January 2007, the CRL (2018) hosted a meeting to seek consensus on what should be the core criteria on which repositories should be judged. In addition to CRL, the invitees were the Digital Curation Center (DCC; Digital Curation Centre, 2018), Digital Preservation Europe (DPE; Digital Preservation Europe, 2018a), and Nestor (2018) from Germany. These four organizations agreed that the following 10 criteria were core to the preservation of digital information and should be used as high-level audit criteria (Center for Research Libraries, 2007).

- The repository commits to continuing maintenance of digital objects for identified community/communities.
- Demonstrates organizational fitness (including financial, staffing, and processes) to fulfill its commitment.

² Coretrustseal.org

- Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.
- Has an effective and efficient policy framework.
- Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
- Maintains/ensures the integrity, authenticity, and usability of digital objects it holds over time.
- 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.
- Fulfills requisite dissemination requirements.
- Has a strategic program for preservation planning and action.
- Has technological infrastructure adequate to continuing maintenance and security of its digital objects.

These 10 principles reflect the principles of the TRAC audit checklist and can be organized within the three main functional requirements broken down by TRAC (RLG/NARA Task Force on Digital Repository Certification, 2007) as well as other audit and review standards that followed.

Organizational Policy Built on Principles

Organizations seeking trustworthy status need to address the 10 principles within their published policies. This starts with an organizational mission statement, which is not to be underestimated as it is the first document many auditors will review when assessing a repository seeking trustworthy status. Does the repository perceive data management and its long-term preservation within its mandate? Does the repository fall under another controlling institution and, if so, does the parent organization address the commitment to the goals of the lower organization in their mission statement? Does the mission statement touch on the 10 principles above? Upon beginning the review process, a systematic reflection on the entire policy is important and should start with the mission statement. For these mission statements to be used as evidence and to promote transparency, they need to be published where the designated community and reviewers can access them.

As repositories grow, so do their financial requirements and responsibilities. Before an organization begins the certification process, it needs to be understood that a core component of the review will be an examination of the overall host organization and the financial stability of that organization. Although no organization can be without any financial risk, policies and procedures need to be defined in the event of a loss of funding. A key question to ask is what happens to the data? A post-apocalyptic management approach is required to ensure the preservation of the data. Deaccessioning policies need to be in place to state who will be the new steward of the data (DataPASS, 2018).

A central area that is often forgotten is the use of a defined records management system for the maintenance of the repository's policy actions as well as planning documents. An accurate record of any decisions related to the archive will be helpful and, in some cases, required as evidence in a trustworthy audit and review (ISO, 2018c). Good policies should be updated and changed from time to time, and these versions should be tracked. A comprehensive approach to policy, planning, and records management would be one of the most critical steps repositories can take to protect the data entrusted to their care. Part of these policies are those related to access rights, data management plans, and data deposit agreements, which are some of the most vital documents and should be preserved and curated in the same manner as other data entrusted to the repository. Policies should cover all areas in the TRAC principles and often include both higher-level policies and lower-level policies that document the actions envisioned in the higher-level policy.

Dedicated policies on the metadata curated to preserve the digital object should be complete and well-rounded. In October 2014, the PTAB (2018) hosted the first training course for potential ISO 16363 (ISO, 2012) auditors. In that course David Giaretta said, "that he would

not talk very much about metadata” (PTAB, 2014b). He felt that it could be very misleading because the definition of “metadata” varies wildly between disciplines and the type of data to be preserved. In addition, the OAIS model refers to many different things that could be considered metadata in some form or another (ISO, 2018c). He preferred to use the term *representation information* defined in the OAIS reference model (ISO, 2018c), as it is more inclusive. Lee (2010, p. 4024) pointed out that:

One of the most important insights embedded in the Reference Model is that “Content Information” to be preserved by an archive is composed not only of a set of bit sequences” (the “data object”) but also associated sufficient “Representation information” to allow the bits to be rendered, used, and understood.

A vital part of this is what the definition of *understood* is, which points back to the definition of the designated community in Core Principle 1. The Reference Model gives the repository or archive the flexibility to make this decision, but also builds a description of the preservation properties that should be present in the metadata curated and preserved with the object. This information was vital in the development of the PREMIS preservation metadata schema (PREMIS Editorial Committee, 2011) that would serve as great evidence during a trustworthy review. Many communities of practice have evolved around the development of disciplinary metadata and the documentation of the designated communities’ metadata of choice and the application of such metadata in repository workflows should be included in the reviews.

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. (Consultative Committee for Space Data Systems, 2012, p 2-4)

Mentioned in both the core principles and the OAIS model, this designated community or the

“identified community/communities” (Center for Research Libraries, 2007) are a key concept that archives and repositories need to define for their organization. They directly map to the “designated community” defined in the OAIS model (ISO, 2018c). In my opinion, this is one of the most critical areas to define for a repository. Many of the following areas of concern depend on the repository’s choice of the designated community. Repositories and archives make commitments to that community and need to maintain a connection with that group. As the requirements of a community change or the repository adds a new community, the repository needs to understand the digital information produced and used by that community. Repository staff should also understand the methodologies, software, and tools used by that new community. Domain repositories that concentrate on a particular academic discipline are best suited to make this connection. But as the demand increases for trusted repositories to provide stewardship for the ever-increasing amount of data produced, archives are expanding their holdings to new disciplines (The Dataverse Network, 2018). Domain repositories and archives have the skills needed and are a natural place for expansion; the domain repository community simply needs to understand the increase in demands as the diversity of their designated communities grows.

This designated community decision has both financial and policy implications. The required information needed for the designated community to “understand” the data object relates directly to the definition of the “significant properties” for the digital object. What is considered a significant property for each digital object can vary and is often not immediately clear, but significant properties should always flow from the needs of the designated community and are often defined in the representation information for those objects (Hockx-Yu & Knight, 2008). It is the combination of the selected representation information and the data objects that make up the “information object” to be preserved by the archive or repository. Many domain

repositories, for example those in the social science communities, have adopted standards for representation such as the Data Documentation Initiative (DDI; Data Documentation Initiative, 2018). These standards help repositories define the required representation information and promote interoperability across repositories based on the needs of the social science research community. Transparency of repository policies and procedures as noted earlier are perceived as critical in the building of organizational trust, which is a key component of trustworthy status.

The information object, if preserved alone, is not enough to ensure trustworthiness as defined by the OAIS Reference Model. As explained by Lee (2010), it needs the preservation description information as well, which consists of reference information, provenance information, context information, and fixity information. The combination of the information object and the preservation description information makes up the archive information package (AIP) (ISO, 2018c) that provides the materials for repositories to comply with the sixth core principle (Center for Research Libraries, 2007).

Before a repository proceeds with an audit or a review, the staff should have a solid understanding of the definitions and required contents of the three critical information packages described in the reference model. In addition to the previously described AIP, the submission information package (SIP) contains the digital products and materials delivered to the repository by the data producers while the dissemination information package (DIP) is used to deliver the digital objects back to the designated community (ISO, 2018c). These are critical units of analysis that will be used during the review process, so they need to be well understood. A first step is to use the OAIS functional model to describe the repository's processes in these terms (Crabtree, 2009; ISO, 2018c). This process will start to bring into focus the technological infrastructure of the organization as it relates to the OAIS model. Organizational policies should

define how the repository's technological infrastructure will be utilized. Documentation of these technological infrastructures will provide the needed evidence during the review.

The repository's technological infrastructure is vital to the long-term preservation of the digital objects and should be used in combination with practical preservation policies. Digital preservation systems must utilize a holistic approach to preservation that ensures that all ten core principles of preservation are addressed (Center for Research Libraries, 2007) to fulfill repository review criteria.

Many different concepts and techniques can be used to build the infrastructure required to ingest, manage, curate, preserve, and disseminate digital information as long as they address all the functional entities described in the OAIS Reference Model (ISO, 2018c). The Educopia Institute (Educopia, 2018) published a guide to distributed digital preservation that seeks to provide a reference for those designing such distributed systems (Skinner et al., 2010).

Some of the previously mentioned certification and audit checklist information, schemas, and standards need more discussion. The next section will explore different review and audit options that exist for repositories and how they relate to each other. It also covers how a repository should begin the audit process.

Review Methods and Certification Opportunities for Repositories

The OAIS Reference Model and the ten core principles help archivists examine repository practices to evaluate the trustworthiness of the organization against audit or review standards. A review of a repository compares the local actions and capabilities of the repository to the criteria described in the standards. The next step toward trustworthy status is the use of this review or audit to provide a certification that the repository has completed a selected audit method. Audits can be self-audits, peer reviews, or formal audits by a certifying body. Any

review must be viewed in the overarching context in which it was performed regardless of whether the review results in a certification.

Many organizations begin with the important self-audit, which allows the repository to identify areas for improvement and address those areas immediately to better ensure the preservation of their digital collections. One approach to this is the risk management approach, which is taken when using the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA; Digital Curation Centre and DigitalPreservationEurope, 2018). This tool is a web-based environment that helps the repository identify risks to the digital objects, and it documents the evidence a repository has in terms of policies and procedures that will mitigate those risks. The system ranks risks and allows the repository to prioritize areas that need immediate attention compared to areas that can be improved over time. This approach is supported by research on the preservation pressure points facing repositories (Ross, 2007). The DRAMBORA tool has been used by many organizations to begin the self-assessment process and can be quite helpful (Pejsova & Vaska, 2010). The Odum Institute Data Archive (The Odum Institute, 2018s) utilized the tool in its early phase before it was a web-based tool, and it was very helpful. The archive discovered many places where policies and procedures needed to be documented better in order to provide the evidence needed to mitigate the risks identified.

Launched at the same time as DRAMBORA, The University of Glasgow Humanities Advanced Technology and Information Institute (HATII; University of Glasgow, 2018) developed the Data Asset Framework (DAF; Data Asset Framework, 2018) to assist with repository audits. This framework breaks down the evaluation into four units:

- Planning the audit
- Identifying and classifying assets

- Assessing management of data assets
- Reporting and recommendations

They also offer an online tool that assists repositories that want to construct a registry of assets as part of their audit. The DAF is geared toward the research data asset type. These tools can be used together to help the repository prepare for and complete a self-evaluation.

Planning for the audit and review is a critical part of the process. The lack of prior planning can lengthen the review and increase the costs. DPE (2018a) created a Planning Tool for Trusted Electronic Repositories (PLATTER; DPE, 2018b) to help repositories plan for audits. Several such checklists were developed around the same time, but PLATTER aims to identify the difference in organizations and accounts for these in the audit (HATII, 2008). PLATTER classifies repositories using independent axes grouped into four categories:

- Purpose and function
- Scale
- Operation
- Implementation

The PLATTER planning process uses the strategic planning approach and a series of specific, measurable, assignable, realistic, and time-related (SMART) goals (HATII, 2008). The PLATTER process uses strategic objective plans (SOPs), which the repository identifies as critical to building trust, to plan for the audit.

Another helpful self-audit tool developed by Nancy McGovern at the Massachusetts Institute of Technology is the Internal Audit Tool (Archivematica, 2014). McGovern described the tool as:

[an] example highlight[ing] a simple yet effective Drupal-based tool for ongoing self-assessment and incremental development by repositories. The tool was initially

developed at ICPSR using the 2007 version of TRAC pending ISO approval and has now been updated at MIT Libraries to reflect the ISO 16363 version of the TRAC requirements. The tool reflects the framework developed for the Digital Preservation Management workshop (<http://dpworkshop.org>) that moved from ICPSR to MIT. (McGovern, 2013, p.1)

The tool was originally developed to comply with the TRAC audit checklist (RLG/NARA Task Force on Digital Repository Certification, 2007) and has since been modified to comply with the ISO 16363 (ISO, 2012) version of the requirements. One area the tool addresses that the TRAC documentation does not is levels of compliance. The tool can document and manage the progress an archive has made as it gets closer to a formal audit. Ongoing assessment is a requirement for any audit or certification, and this tool helps repositories and archives manage these assessments (McGovern, 2013).

If an organization uses the Lots of Copies Keep Stuff Safe (LOCKSS; LOCKSS, 2018b) to create a distributed preservation network (Skinner et al., 2010) a tool called the Safe Archive Audit Tool (SafeArchive, 2018) can help gather compliance evidence based on TRAC. This tool, developed in 2010, also allows organizations to record and preserve evidence of TRAC compliance that are policy based in addition to the automated audit process. The goal is to provide an audit system for an asymmetric preservation network that partners can utilize to ensure compliance (Altman & Crabtree, 2011). McGovern (2013, p.3) described the need for this tool as:

A gap in digital preservation practice has been the ability for humans in repositories to interact with the network responsible for replication and related functions to has produced results from several networks establish protocols and define rules as content is added, partners join and drop, and requirements evolve; to review and respond to the results of monitoring and auditing the integrity of copies; to provide documented and auditable results of replication functions as part of full TRAC audits.

The tool has not been updated to reflect the new ISO 16363 standards but is still useful for auditing against the 2007 TRAC criteria.

Many organizations have worked on the area of digital preservation and auditing to provide guidance for organizations seeking to review their policies and procedures for digital preservation. A few interesting ones include the Sustaining Heritage Access through Multivalent (SHAMAN) Project (Shaman, 2018) that sought to help organizations with digital preservation by using a lifecycle management approach. From 2008 to 2011, the project worked on developing a next-generation digital preservation framework for scientific and academic outputs. Another organization that has worked to assist organizations is the Alliance for Permanent Access (APA) (APARSEN, 2014). They encouraged organizations to promote sustainable open access to the core scientific infrastructure (APARSEN, n.d.). This project was completed in 2014. In some disciplines, organizations are tackling data quality and repository review as well. The World Meteorology Organization (WMO, 2018) supports quality and standards in climate and weather data using the ISO 9000 standard (Bureau of Meteorology, 2018). Another organization that is certifying scientific data is the World Data System (WDS), which is promoting quality scientific data especially in the areas of geosciences and earth observation (WDS, 2018b). They are an interdisciplinary body of the International Council for Science (ICSU) and work closely with the ICSU's Committee on Data for Science and Technology (CODATA; CODATA, The Committee on Data for Science and Technology, 2018).

Once organizations have performed a self-assessment or policy review using one of the methods described thus far, the next step could be an official certification for trustworthy digital repository status. There are several options from which repositories can choose, ranging from lightweight certifications, often in a peer review fashion, to full audits performed by a paid organization that acts as a certifying body. This increasingly rigorous review process can be seen as progressing from a self-assessment followed by a peer review of that assessment. This can

then be followed by a lightweight seal of approval performed by a certifying agency and could even culminate with an official international standards audit based on approved audit methodologies.

The Data Archiving and Networked Services (DANS; Data Archiving and Networked Services, 2018), based in the Netherlands, developed the Data Seal of Approval (2018b) that provided repositories with a framework and community of peer reviewers that certify organizations as qualifying for the Data Seal of Approval. The process involves an online form that leads the repository through a series of 16 guidelines based on international standards for trustworthy repositories (Data Seal of Approval, 2018a). The seal is granted for a period of 2 years, and a process of renewal is in place to ensure the long-term quality of the assessment. As part of this organized community, certified repositories agree to perform three peer reviews each year to help grow the community and make it sustainable (Dillo & De Leeuw, 2015). This seal has now merged with the World Data Seal (RDA, 2018b) and has evolved into the new CoreTrustSeal (CoreTrustSeal [CTS], 2017a) mentioned previously. I will discuss the CoreTrustSeal in detail later.

A seal similar to the Data Seal of Approval is offered by Nestor (2018) based in Germany. Nestor describes itself as “the German competence network for digital preservation,” and it provides training and standardization activities to supplement the offering of a data quality seal. The Nestor Seal is based on DIN standard 31644 (DIN, 2014). The DIN Standards Committee for Information and Documentation (DIN, 2018) is responsible for standards and practices for libraries, archives, and data publishing organizations. The Nestor Seal fits with the European Framework for Audit and Certification of Digital Repositories (Trusted Digital Repository, 2018) as well as the Data Seal of Approval. They offer what they call an “Extended

Certification” that was designed to be complementary to the Data Seal of Approval. The Nestor Seals are still in existence, but there is not much uptake outside of the German context.

CRL (2018) is an international consortium of libraries and has been active in the audit and certification of trustworthy repositories. It published the TRAC repository checklist and criteria and has been performing audits based on that checklist (CRL, 2018). It started in 2009 by performing an in-depth assessment of both Portico and HathiTrust. All of its assessments are posted online, and it aims to be transparent in the processes. It has done a total of six audits (CRL, 2018) thus far, and each provides ongoing requirements for the organizations. The cost of these reviews can be expensive. Although exact costs depend on the unique qualities of each archive, the average has been around \$60,000. This is a significant limitation for smaller organizations without a business model to support the effort. The audits follow the TRAC checklist published in 2007 (RLG/NARA Task Force on Digital Repository Certification, 2007), and CRL does not mention that it will be transitioning to the new ISO 16363 standard (ISO, 2012) that superseded that checklist. This is an evolving area, and the uptake of international standards process can be long and slow.

Certification Next Steps

With the passing of the ISO 16919 (ISO, 2014) and the certification of PTAB as the first auditing body (Giaretta, 2017b), we are now able to officially audit with ISO 16363. The first 16363 certification was granted to The National Cultural Audio Visual Archives of the Indira Gandhi National Centre for the Arts in India in 2018. (Giaretta, 2018) Why it took so many years from the publication of ISO 16363 to the first audit is tangled with the understanding of the ISO processes and procedures. Every process has an associated standard, and it is often the case that standards rely on additional standards. For an organization to become an ISO certifying body and

certify against the OAIS standard, it must also qualify as a certifying body under the ISO 17021 standard (ISO, 2011). Previously, no organizations that met this standard was trained to perform an OAIS audit. These organizations got their authorization from national standards boards, such as the American National Standards Institute (2018). Once an OAIS certifying body has been approved, it must then train auditors under the guidelines of the new ISO 16919 (ISO, 2014) to ensure that the appointed audit teams follow all higher-level standards in addition to the ISO 16363 (ISO, 2012) under which the repositories would be reviewed. Currently the PTAB group is leading training and certification around ISO 16363 and has recently certified the second repository against ISO 16363 (Giaretta, 2017a) when the US Government Publishing Office achieved certification (Giaretta, 2018).

During the time that elapsed in the approval of the ISO certification process, a working group in the Research Data Alliance began to coordinate the collaboration between the Data Seal of Approval and the World Data Seal to align their requirements and to more closely reflect the directions of ISO 16363 (RDA, 2018b). The ICSU World Data System has a history of certification and review of its members data systems. (“Certification — World Data System: Trusted Data Services for Global Science,” n.d.). This group has now merged its certification efforts into the new CoreTrustSeal (2017b) and has identified 16 core pieces of evidence that need to be presented to reviewers. The CoreTrustSeal is managed as a peer review process in which at least two peers review the written submissions for compliance. There is no on-site audit process. The seals are valid for 3 years and, although not free, they are affordable, with only a 1,000 Euro administration fee. Archives previously certified with the DSA or the WDS are now transitioning to the new requirements. Many archives and repositories across numerous disciplines are viewing the CTS as a useful credential, and the number of requests for review are

increasing rapidly. At the beginning of this project, 45 organizations had achieved the new seal (CoreTrustSeal, 2017a). CoreTrustSeal is emerging as a new community standard, with 61 WDS and 39 DSA certifications that still need to transition to the new CoreTrustSeal. The CoreTrustSeal applications are also coming from a more diverse set of applicants. It is receiving applications from national general purpose repositories, infrastructure providers, and cloud service providers, in addition to typical traditional domain repositories. This increase in applications could be due to the increased visibility from the European Open Science Cloud that has recommended the CoreTrustSeal (European Commission, 2018) and the GoFAIR initiatives (GO FAIR, 2018).

Groups such as the RDA are continuing to leverage pilot studies and working groups in an effort to bring some clarity to this process. They hope to coordinate with certifying bodies and seal of approval communities to help repositories make decisions that align with their interests and develop tools to verify compliance. Many technological tools such as iRODS (Moore, 2008) show promise by enforcing some policy-based rules at a granular level, but the audit of policies and procedures is still very much a manual, time-consuming process. This process requires higher-level policies to shape the lower-level actionable policy that technology tools can handle, which is especially difficult when an organization does not plan ahead for the audit and fails to have processes in place to document the evidence with quality organizational policies. The number of tools and services to aid with this is growing. The following section provides a snapshot of several categories of these tools that can be used to enforce actions after organizations define appropriate organizational policies.

Tools, Services, and Applications Literature

The digital preservation community is fortunate to have a wide base of support around

the world. In addition to RDA, conferences such as the International Conference on Digital Preservation (iPRES, 2018), the Digital Curation Centre (2018), and the National Digital Stewardship Alliance (Library of Congress, 2018) hosted by the Digital Library Foundation (DLF) (CLIR + DLF, 2020) are just a few of the groups working in this space today. Academic, commercial, non-profit, and governmental organizations all have interest in this area and devote countless hours and efforts to help meet the data preservation needs of the archive and repositories community.

Although this situation is encouraging, it also creates inherent problems when an array of technology developers and administrators work on similar tools and services simultaneously and without full knowledge of other programs and projects. Overlap and redundancy are common, as is varying levels of quality and support in products and services. This review highlights starting strategies as well as some of the more powerful services and tools that may be needed. A complete list of these would be impractical, so the discussion will include some links to resources designed around keeping archivists updated on new technology. This review will be broken down into five primary categories:

- Fundamental training
- Front-end and ingest processors
- Preservation as a storage service
- Full service systems
- National research infrastructure

When examining the metrics within the Trustworthy Digital Repository (TDR) audit schemes as well as other evidence required in various certification efforts, it is clear that becoming “trustworthy” requires addressing two topics. As previously discussed, the first is the

organizational policies and procedures established by the group seeking the certification. These policies and procedures are put in place to ensure long-term organizational stability, definitions of supported communities, dedication to the preservation mission, and provision of options when an organization unfortunately fails. An organization must have such policies and adhere to these policies before it can be deemed trustworthy.

The second area of focus is the policies and procedures for digital object management. Policies here can dictate the use of specific tools and include a wide range of actionable processes archivists and technologists use in typical workflows. Although also too numerous for an exhaustive list, a few are file format migration tools, compression and decompression routines, checksum calculations, media migration, automated replication tools, and scans for sensitive information. In this area of digital object management, information technology and tools can be most instrumental. These two symbiotic policy areas, where high-level policies describe lower-level workflow actions informed by each other, provide repositories with the ability to ensure preservation goals. The scope of this section will include many key components, but it is not meant to be an exhaustive list.

Breaking Down the Options

Few practitioners in the field of information and library science have adequate time to research the vast array of options and details around various preservation systems. This may be a primary driver that steers libraries and archives to seek out vendors and pay for services rather than develop tools and workflows of their own. In many cases, this is the best option if the purchasers of such services understand what the vendor is doing and promising and if the vendor can document the offering. This should result in appropriate memoranda of understanding. Seeking tools that promote transparency to help build trust is important. An example of this is

the efforts of Archivematica (2018) to promote the transparency of workflows. Fortunately, there is a project that helps sift through the numerous combinations of options to help make these decisions. The POWRR project (Digital POWRR, 2016) maintains a tools grid to help define the strengths and weaknesses of preservation tools. (Digital POWRR, 2018). This popular informational source gives practitioners a simple way to review lists of potential options. The tools are listed using an approach that matches the workflows typical of repositories and archives while also matching the standard OAIS nomenclature. (ISO, 2018c). This valuable effort was recognized by the community, especially Educopia's (2018) effort to align national approaches to digital preservation (Skinner & McGovern, 2012). The latest version of the tool grid is POWRR by COPTR (COPTR, 2018; Digital POWRR, 2018) and features a hyperlinked design that provides a simple to use interface for users. The IMLS Laura Bush 21st Century Librarian Program is currently funding this effort.³ The COPTR system allows for automated updates and enables a disseminated approach to maintain a catalog of tools, but ensuring accuracy of the tools remains difficult as software development cycles are compressed with the adoption of continuous integration techniques (Martin, 2003). At the time of this writing, I have noticed a few features missing on several popular platforms, yet the tools are helpful and a great place for practitioners to start.

Tools and services developed for use in digital preservation are numerous and diverse. Some have a very small and focused application such as the BagIt (Library of Congress, 2016) technology designed to aid in transfer technology used during migration of content to distributed and redundant storage locations. Others have more integrated and workflow style complex

³ <https://digitalpowrr.niu.edu/>

interfaces, such as Archivematica (2018) or Archivists' Toolkit (2018)⁴ that are designed to be integrated into the archivist's daily workflows and procedures (Chandler et al., 2001). Some tools cross all the boundaries of the research data lifecycle (Digital Curation Centre, 2018) while others just touch on a single process. Although technological tools are essential in digital preservation, they must be utilized appropriately in a holistic approach to managing the archive or repository. The technologies selected for digital curation workflows must flow from choices in organizational high-level guiding policies into an active environment that documents or enforces preservation actions. A simple example might be where an organizational policy may state that submissions for ingest into their archive will be migrated to sustainable formats for long-term preservation. This could inform technological workflows to generate simple ASCII file formats to be preserved along with the original file submitted.

Training and Education

Training repository staff and technologists about digital preservation and data management is critical to digital preservation and the certification of repositories. There is a real danger of repository managers selecting tools and service for digital preservation without knowledge or training necessary to assemble policies and technologies that are robust enough to be certified. Opportunities to learn about digital preservation are numerous and growing. Library and information science schools around the world have been gearing up to handle the deluge of digital objects being cascaded upon practitioners in the field. Academic institutions are developing courses and materials not only for resident students, but also for postgraduate training and distance education. Examples range from online training options such as Curating Research

⁴ Archivist Workbench has been part of ArchivesSpace (2018) since 2009 after a merger with Archon.

Assets and Data using Lifecycle Education (CRADLE, 2018) and MANTRA (2018) to professional training including the offerings at the Digital Curation Centre (2018). Conferences and workshops such as the DigCCurr (2018) workshops and Digital Preservation Management offered by McGovern (2016) have also been instrumental in training managers and preservation staff. The Institute of Museums and Library Services (IMLS, 2018) has funded many projects that support training, including POWRR (Digital POWRR, 2016) and projects managed by the Online Computer Library Center (OCLC; Connaway, 2015). These projects and many other have built a framework that should be leveraged and used as a foundation for additional training (Lee et al., 2017). The Data Curation Network (DCN) is leveraging a cross-institutional model to curation that shares the training and skill across organizations (Johnston et al., 2018). There are available models for archives and repositories to leverage in training staff, which is often a great need (Bresnahan & Johnson, 2013).

Professional organizations such as the National Digital Stewardship Alliance (NDSA), the International Federation of Data Organizations for the Social Sciences (2018), and the RDA (2018a) encourage active participation and knowledge exchange across the disciplinary boundaries that can be beneficial for those entrusted to steward digital assets. Groups such as the Open Preservation Foundation (2018) and the CRL (2018) provide tools and services that can be leveraged to learn more. For deeper understanding around unique identifiers, groups such as DataCite (2018) or ORCID (2018) can provide the tools and knowledge to educate archive teams how to properly identify digital objects and provide the quality provenance required for TDR status. Once an organization is at the point of beginning an audit, groups such as the PTAB (2018) offer training and information to start the process. Other organizations such as the Data Seal of Approval (2018), the World Data System (2018a), and the CRL have educational

opportunities based on their metrics as well. Training and education are geared toward attaining enough knowledge to make the correct decisions. The challenge is that information technology changes so rapidly that one quickly falls behind on the current state of the art. Finding available and affordable options for keeping up with digital preservation technologies and their associated strengths or weaknesses can be challenging.

Front End and Ingest Processors

This category of tools is quite broad and often integrated with, or contained within, various other repository software suites. The POWRR system defines five categories in the ingest and six software types for the processing suite (Figure 1).

Figure 1. Digital preservation tool grid (Digital POWRR, 2018).

Digital POWRR Tool Evaluation Grid	Ingest	Processing			Access	Storage			Maintenance		Other						
	Copy	File Dedupe	Auto Unique ID	Auto Metadata Creation	Auto SIP Creation	Auto DIP Creation	Auto AIP Creation	Reliable, Long-Term Bit Preservation	Redundancy	Geographically Dispersed Data Storage Model	Exit Strategy	Migration	Monitoring	Auto Recovery	Open Source	Clear Documentation	Cost
		Virus Scan	Auto Metadata Harvest	Manual Metadata	Rights Management	Package Metadata	Public Interface										
		Fixity Check															

Notable tools to apply in these areas of the curation workflow are those such as BitCurator (2018), designed around digital forensics to enable curators to make appropriate decisions for ingest and acquisitions as well as identify sensitive information prior to transfer into the systems. Meanwhile, the Audit Control Environment (ACE; Smorul, 2010) helps calculate and track fixity, which is required for TDR certification, making it important for use as evidence. This group also includes tools designed to process Internet-based content, such as Heritrix (2018) and individual tools for specific processes, such as metadata extraction, as exemplified with the Embedded Metadata Extraction Tool (EMET; Library of Congress, 2011). Two additional tools that stand out in this category are Curators Workbench (UNC-Libraries, 2018) and the Duke

DataAccessioner (2016) for their ability to integrate into the ingest workflow and ease of use.

Preservation as a Storage Service

With the rapid advances in network technologies and cloud-based systems, we have seen more providers enter the market providing offsite data preservation and replication (Erikson & Lunt, 2015). These “preservation as a service” models vary greatly in the levels of curation service and tools they offer. Many also have sophisticated bit preservation strategies and have the ability to provide evidence during a TDR audit so an organization can show it has the procedures and plans to protect the digital object across diverse technological and geographical divides. These range from commercial cloud offerings such as Amazon Glacier (Amazon Web Services, Inc., 2018) to cooperatives such as MetaArchive (2018) that form alliances to better protect their digital assets. Some use sophisticated Private LOCKSS Networks (PLN; LOCKSS, 2018a) while others use diverse technologies such as the integrated Rule-Oriented Data System (iRODS, 2018) that can apply machine actionable policies at the object level, enabling an active storage environment. Research into the use of distributed preservation and the application of the Outer OAIS-Inner OAIS Model to define roles and responsibilities each organization performs helps understand these practical implementations of distributed preservation (Zierau, 2017). Some have attempted to offer combinations of these technologies, such as the Digital Preservation Network (DPN; 2018), which also tried to address the difficult issue of forecasting preservation costs into the future.

A difficult part of TDR audits evidence is often a sustainability plan (Blue Ribbon Task Force, 2008) for the preservation technology. DPN offered a “pay up front” model, where deposits were guaranteed for 20 years included in the one-time deposit fee. Unfortunately, this effort did not stand the test of time and is in the process of sunseting. As with the Digital

Preservation Network, Duracloud offers affordable scalability and reliability by offering to replicate repositories' objects across more than one commercial cloud provider in an effort to prevent a technology lock-in failure. In addition, Chronopolis (Regents of the University of California, 2020) offers services to offset the closure of DPN. Both of these groups offer bulk discounts and enable economies of scale as the number of repositories and archives they serve grows. The curation features of these groups vary, but as these organizations mature, more full-featured services are to be expected.

Full Technological Service Systems

Organizations face a difficult challenge when selecting technology for preservation workflows. In many cases, both commercial and open-sourced systems are commonly referred to as “full” technological preservation solutions but in fact need to be combined with other tools to ensure trustworthy digital repository status. It is critical to remember that preservation starts early in the digital object lifecycle as provenance starts there. These early lifecycle curation tools are often left out of full-service models. Organizational policies need to be in place to make sure that internal workflows begin capturing required information even before these technological solutions are entrusted with the data. In addition, the organizations leveraging these technology solutions need to secure and document service level agreements with chosen service providers to provide a transparent view of the preservation workflow.

Web archive systems such as Archive-It (2018), with ties to the Internet Archive (2016), provide users with a very simple beginning to end preservation system for valuable Internet content. Other commercial offerings, including Rosetta (Ex Libris, 2016), Portico (2018), Archivematica (2018), and Preservica (2018) offer turnkey services. Documentation is key, but that being said, the CRL (2018) has certified six such services against the TRAC guidelines that

were at the heart of the new ISO 16363 standard (ISO, 2012), so these services can offer auditable technology.

Stand-out open-sourced, locally-managed and do-it-yourself-style full service applications exist along with some excellent commercial offerings. One of the commercial offerings that provides valuable services to the preservation of digital objects is Nesstar (2016), which is a suite of three products that provide a full range of curation services. A similar open-sourced offering developed at Harvard University is The Dataverse Project (2018), which was originally developed for application in the social science disciplines, but recent additions have expanded its application to many other areas. Dataverse offers many ingest tools for TDR, but as with most systems, requires additional tools and policies working in conjunction to provide end-to-end preservation. Often these full-service options are not truly full service. They focus on specific use cases and could mislead users into believing they are truly full service. Training and education are required when selecting these options to prevent requirements from being left out.

Some groups offer data management and consulting services that can provide solutions that help fill in the procedures and requirements necessary to achieve TDR status. The Odum Institute's (2018c) data management services are one such example as it offers three levels of service that includes accepting digital objects as part of their archive holdings. In addition, the Digital Scholarship & Preservation Services at Cornell University Library (2018) offer in-depth conservation services and advice on digital preservation issues. In recent years Figshare (2015) has become a player in this market, and recent partnerships with DuraCloud shows it has tools that can be leveraged to provide audit evidence of solid preservation strategies. Understanding how these distributed preservation models allocate responsibilities has an impact on the evaluation of trust. Work by Zierau and McGovern (2014) helps to define these roles and

responsibilities using an Outer OAIS-Inner OAIS model.

National Infrastructures

Many countries are building national data infrastructures and, in some cases, offering those resources as tools that can be leveraged by repositories seeking TDR status. In Europe, these national infrastructures are joining forces within the Consortium of European Social Science Data Archives (CESSDA, 2018) to build international systems. The World Data System (WDS) an interdisciplinary body of the International Council for Science (ICSU; WDS, 2018) seeks to increase international membership to achieve similar goals. These organizations can provide infrastructure as well as support services, in some cases, for those seeking TDR status. Similar efforts are underway with the Australian National Data Service (ANDS, 2016) as well as in the United States, where efforts are spread across several National Science Foundation (NSF) projects, including the National Data Service (NDS) based out of Illinois (Board of Trustees of the University of Illinois, 2016) and the collection of Big Data, DIBBS, and DataNet awards from the NFS (“Datenet,” 2015; NSF, 2016a, 2016b). These projects have many offerings that could be repurposed or leveraged to provide better repository services. For organizations handling population and spatial data, the Terra Populus (2018) project, which integrates data on population and environment, has much to offer. The Data Observation Network for Earth (DataONE) offers a suite of tools and best practices for publishing and sharing data and is built on strong metadata standards and strategies for unique identifiers. The DataNet Federation Consortium (2016) has created a network of federated data stores that span diverse disciplines from the temporal dynamics of learning and plant biology to engineering workflows and social science research. At the heart of this collaboration is the grid-based system iRODS (2018) that employs a policy-based data management system (Reagan et al., 2011; Ward et al., 2011).

Policy-Based Approach

Throughout the audit process, organizations will rely on documents that represent the expression of policies to be utilized as evidence that procedures and workflows adhere to the standards set forth within the audit metric. In many cases, these will be higher-level documents that speak to encompassing issues such as the mission statement of the organization. The higher-level policies will be broken down into actionable policies the closer one gets to the actual day-to-day workflows of archivists. To prepare for a trusted digital repository audit, organizations will have to gather evidence that each of the metrics is being addressed. The PTAB (2018) has developed a worksheet to help participants prepare for ISO16363 audits. This worksheet lists individual metrics and what evidence is required to successfully satisfy that metric (PTAB, 2014a) Other auditing bodies have similar worksheets, but the ISO 16363 metrics are far more detailed, presenting more of a challenge and more benefits than those from simpler review schemes.

Many of these organizational metrics are high-level, policy-driven, and focused on the overall organization, its governance structures, primary mission, and the stability of the business model. Yet a substantial number of metrics deal with the handling and curation of digital objects that lend themselves well to automation using information technology. As previously mentioned, these are typically the application of processes to aid data management and range from automated metadata extraction from known file formats to the calculation and monitoring of fixity information as data objects move through the preservation workflows. Reagan Moore, with the Data Intensive Cyber-Environments (DICE) group, investigated and documented which of the ISO 16363 metrics can be automated, monitored, and controlled by utilizing policy-based rule management with iRODS. He suggested that 140 preservation tasks can be managed with

policy-based rules (Moore et al., 2015). Work performed during the DataNet Federation Consortium project to develop a policy template and examples of policies demonstrate the value of this approach. Current efforts have shown that it is possible to automate a large number of tasks required for trusted digital repository status.

The objective of creating computer actionable policies for each task remains a viable approach to preservation. Generic operations can simplify the implementation of preservation tasks while policies can manipulate the multiple objects needed to execute the preservation tasks. This makes it possible to automate preservation processes. (Moore et al., 2015, p.8)

Templates and examples that can serve as models and real-world examples of these working policies have been created (Chen et al., 2015a, 2015b). Currently these policies are implemented as workflows of chained together micro-services. The iRODS Consortium (2018) has worked to provide a sustainable open-source toolkit for data management and is being leveraged across the world at institutions and corporations that have the resources to develop workflows that take advantage of this technology. As greater numbers of organizations begin preparing for trustworthy status, it will be beneficial to share automation techniques to make these audits as efficient and effective as they can be with the understanding that this must be done in conjunction with well-documented, higher-level organizational policies.

CHAPTER 3: ANALYTICAL APPROACH AND METHODS

Methods Introduction

The primary purpose in applied qualitative research is to understand a real-world problem better (Guest et al., 2013). This research seeks to understand the challenges faced by reviewers who have recently performed CTS reviews. Creating policies and gathering evidence for these applications have the potential to create stress and additional workloads for archivists. The methods selected for this research provide practical insights into the review process and its impact on the organizations by examining the perspectives of reviewers involved in these reviews.

I am building upon my more than 20 years of work in a data archive focused on information technology and preservation (Crabtree, 2013). My experiential foundation includes a solid foothold into the information technology infrastructures required to maintain a leading social science data archive (The Odum Institute, 2018b) as well as my transition to archive manager and assistant director at The Odum Institute. Along the way, I have gathered educational experiences designed to inform my decisions as archive manager and to propel Odum into a leading role representing social science data archive communities. One of the most informative ways an organization can learn about its policies and procedures is to participate in self-audits. I have been involved in self-audits of the Odum archive on several occasions, ranging from DRAMBORA (Digital Curation Centre and DigitalPreservationEurope, 2018) risk-based assessments to TRAC (RLG/NARA Task Force on Digital Repository Certification, 2007) and Data Seal of Approval (2018b) audit frameworks. My recent ISO 16363 auditor training (PTAB,

2018) will serve as the grounding framework that pulls my experiences here at Odum together.

I understand that my experiential knowledge can help ground me in the topic (Guest et al., 2013) while also introducing bias, which requires leveraging my knowledge effectively with a clear understanding of this potential bias (Maxwell, 2013). An additional threat to the validity of qualitative research is reactivity due to the researcher's influence on the settings or participants the research is examining (Maxwell, 2013). I sought to control for this reactivity by understanding and utilizing my influence, not removing it, in order to use it productively (Maxwell, 2013). I did this by leveraging the use of a positionality memo to remind myself during the research process of this potential bias. Positionality refers to a researcher's position in relation to the topic of study, the participants, and the research context. An ongoing positionality memo is a place for a researcher to track their assumptions of normality and help ensure that they do not impose these assumptions onto the data. In other words, it is a type of corrective against bias and contributes to a data-driven, rather than researcher-driven, approach (Maxwell, 2013). I used this memo⁵ as an audit process and made new journal entries every 4 weeks, on average, during the research project.

This research focuses on the perspective of the reviewers of CoreTrustSeal evidence. A semi-structured interview instrument was developed and executed with reviewers to better understand how they feel about the process and to discover any potential challenges they face. The goal of this project was to understand the reviewer's perspective and the review process as well as to evaluate how these perspectives impact the evaluations.

Philosophical and Analytical Approach

To guide me in this research, I used the philosophical paradigm of pragmatism, which

⁵ See Appendix C.

asserts that “truth may be interpreted in terms of the practical effects of what is believed and, in particular, the usefulness of these effects” (Savin-Baden & Major, 2012, p.60). This stance fits with my mission as assistant director here at Odum and with the objectives of typical projects I have worked on over the years. I think this project fits well in the classical pragmatism movement and is closely aligned with “the basic idea of this form of pragmatism is that the meaning of conceptions should be sought in their practical applications” (Savin-Baden & Major, 2012, p.25).

In alignment with pragmatism, the analytical approach for the final analysis in this project was guided by thematic analysis methods. Thematic analysis is an analytical approach that moves beyond descriptive coding to identifying higher-level concepts or themes—namely, categories that provide meaningful (and abstract) insights into participants’ values, beliefs, and practices. Its advantages include its flexible use and the analytical freedom it provides researchers across a range of data types (Nowell et al., 2017). It is a useful method for examining different perspectives from research participants (Braun & Clarke, 2006). This analytical approach fits well within the classical pragmatism philosophy with the goal to produce practical results. The iterative process of discovering themes will result in a conceptual framework—namely, a collection of condensed concepts that can be used by reviewers as well as archives.

The research workflow examined the evidence reviewers evaluate in an effort to supplement topics for semi-structured interviews with selected reviewers to better understand how they feel about the process and discover any potential challenges they faced. Although data were collected on the certification evidence as presented in various repository CoreTrustSeal applications, the goal of this project was to understand the reviewers’ perspectives and the review process as a whole.

Methods Overview

I conducted this research in three guiding stages. The first stage was designed to discover unpredicted provisional codes (Saldaña, 2016) to be used in stages two and three. These codes were developed by leveraging experiential knowledge in combination with a content analysis of the approved applications. An important decision when performing qualitative research is the choice of coding method. Saldaña (2016) broke down the various coding methods and defined some foundational principles for coding decisions. He asserted that:

Some methodologists advise that your choice of coding method(s) and even a provisional list of codes should be determined beforehand (deductive) to harmonize with your study's conceptual framework, paradigm, or research goals. But emergent, data-driven (inductive) coding choices are also legitimate. (p. 75)

The work in stage one began with the identification of predefined codes based on the researcher's knowledge and document variables. These codes were combined with emergent codes developed from a content analysis of CoreTrustSeal applications. This combination of predefined codes and emergent codes enabled the potential for unanticipated codes to be identified from the data during the first stage.

Moving into stage two, additional predefined codes and question topics designed for use during the reviewers' interviews were defined by leveraging my experiential knowledge. These were combined with those topics generated in stage one and defined the final interview instrument used to collect the data to investigate the research questions. The resulting interview data were explored using a qualitative analysis thematic approach in stage three.

Initial Work and Tool Selection

To start the first stage, I gathered and analyzed the set of public CoreTrustSeal evidence from approved applications posted on the public CoreTrustSeal website.⁶ On February 5, 2019,

⁶ www.coretrustseal.org

there were 45 approved applications publicly available for review.⁷ The CoreTrustSeal application consists of an overview of the repository and 16 comprehensive requirements.⁸ Figures 2 and 3 depict a typical requirement and supplied evidence from a single application.

Figure 2. Typical CoreTrustSeal requirement.

Context

R0. Please provide context for your repository.

– Repository Type. Select all relevant types from:

- **Domain or subject-based repository**
- **Institutional repository**
- **National repository system, including governmental**
- **Publication repository**
- **Library/Museum/Archives**
- **Research project repository**
- **Other (Please describe)**

Comments

– Brief Description of the Repository's Designated Community

⁷ See Appendix A

⁸ See Appendix B

Figure 3. Evidence required for CoreTrustSeal application.

0. Context

Applicant Entry

Self-assessment statement:

Repository Type:

Domain or subject based repository. The area of interest is Social Science Data with a broad remit as to what sorts of digital data are acquired and preserved by the Australian Data Archive. We are a digital data archive and do not preserve physical objects but will digitize and preserve relevant documents and data of interest.

Designated community:

Australian Data Archive (ADA) was established at The Australian National University (The ANU) in 1981 (under the original title of the Social Science Data Archive) with a brief to provide a national service for the collection and preservation of computer readable data relating to social, political and economic affairs and to make these data available for further analysis. The ADA acquires, documents, preserves and disseminates data online to a broad range of social science researchers in the university, government, and other sectors. The designated communities are reflected by the way the holdings are separated into a series of sub-archives reflecting the nature of the data and/or major thematic divisions. The sub-archives are: Social Science – predominantly survey or polling based quantitative social science data ; Historical – an archive of Australian census data tables from 1834 to the present day; Indigenous – A thematic archive bringing together research data about Aboriginal and Torres Strait Islanders; Longitudinal –major longitudinal cohort and panel surveys of the Australian population; Qualitative – a collection which provides specialist data archiving and access services to qualitative researchers; Crime & Justice – major collections of data in crime, law and justice, including criminal justice administrative data; International – a central point of access for links to international data sources around the world.

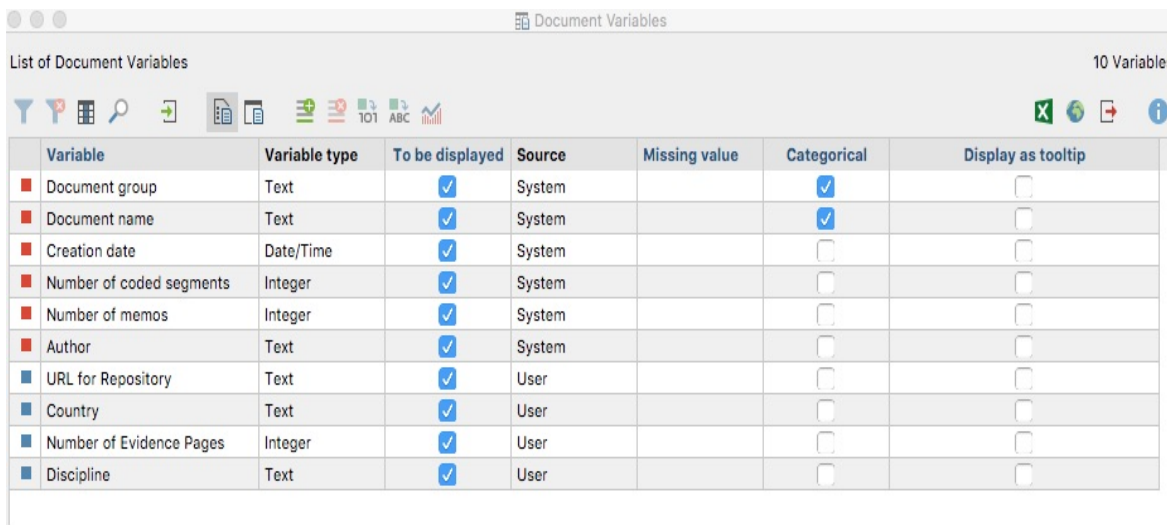
Overview of ADA in ADA User Guides, Introduction to the Australian Data Archive (pg2):
<http://ada.edu.au/ada/user-guides> PDF download IntroductionToADA.pdf (accessed 12/04/2017)

Once submitted to CoreTrustSeal for review, these applications are assigned to two anonymous reviewers who use the CoreTrustSeal guidelines to assign levels of compliance. Before final approval, the CoreTrustSeal board reviews each application. I have observed how, when the two auditors strongly disagree in their assessments, it can be confusing to the applicant.

The repositories represented within these applications are very diverse. They include 18 countries and 12 disciplines that reviewers will be asked to evaluate. These diversities could be a source of variation across applications and reviewers' evaluations could be influenced by this.

The initial predefined codes (Figure 4) are based on these variables.⁹

Figure 4. Predefined codes



The screenshot shows a software window titled "Document Variables" with a subtitle "List of Document Variables" and a count of "10 Variables". The window contains a table with the following data:

Variable	Variable type	To be displayed	Source	Missing value	Categorical	Display as tooltip
Document group	Text	<input checked="" type="checkbox"/>	System		<input checked="" type="checkbox"/>	<input type="checkbox"/>
Document name	Text	<input checked="" type="checkbox"/>	System		<input checked="" type="checkbox"/>	<input type="checkbox"/>
Creation date	Date/Time	<input checked="" type="checkbox"/>	System		<input type="checkbox"/>	<input type="checkbox"/>
Number of coded segments	Integer	<input checked="" type="checkbox"/>	System		<input type="checkbox"/>	<input type="checkbox"/>
Number of memos	Integer	<input checked="" type="checkbox"/>	System		<input type="checkbox"/>	<input type="checkbox"/>
Author	Text	<input checked="" type="checkbox"/>	System		<input type="checkbox"/>	<input type="checkbox"/>
URL for Repository	Text	<input checked="" type="checkbox"/>	User		<input type="checkbox"/>	<input type="checkbox"/>
Country	Text	<input checked="" type="checkbox"/>	User		<input type="checkbox"/>	<input type="checkbox"/>
Number of Evidence Pages	Integer	<input checked="" type="checkbox"/>	User		<input type="checkbox"/>	<input type="checkbox"/>
Discipline	Text	<input checked="" type="checkbox"/>	User		<input type="checkbox"/>	<input type="checkbox"/>

Codes for country, discipline, and number of evidence pages provided initiated the code list for the stage one analysis. In many cases, materials need some predefined initial coding (Saldaña, 2016) to structure the evaluation. The CTS applications are structured with requirement sections labeled R0–R16. These sections were coded so comparisons across applications could be made. In addition, the applications also have a section for reviewer comments; these sections were initially coded as well. Another important predefined code identifies the level of compliance each reviewer assigned the evidence.

Content analysis typically consists of four general steps. To begin the process, a decontextualization of the material allows the researcher to become familiar with the data (Bengtsson, 2016). Systematically reviewing and evaluating thousands of pages of documentation and transcripts required a structured approach. To help build that structure, I

⁹ See Appendix A

chose MAXQDA (n.d.), a content analysis and qualitative analysis software program for this task. This software tool is self-described as follows by the developer:

MAXQDA helps you collect, organize, analyze, visualize and publish your data. It supports various methodological frameworks, including grounded theory, literature reviews, exploratory market research and qualitative content analyses, giving you maximum flexibility to analyze what you want, the way you want.

MAXQDA allows a researcher to select portions of text in a document and assign various codes to these sections. These codes can be compared and examined to detect overlaps or intersections with other codes as well as variables assigned by the researcher. Codes are assigned manually by highlighting sections of the text and assigning an existing or new code. MAXQDA also allows for auto-coding of text with predefined codes. Researchers can perform lexical searches and assign codes based on the results. Once these codes have been assigned, the software has analysis tools that help compare and contrast codes. It examines the relationships the codes have with the various documents, the document variables assigned by the researcher, and other codes.

Stage One Work

To structure the analysis in stage one with initial coding (Saldaña, 2016), the documents were initially coded in MAXQDA to match repository evidence (Figure 5) with the corresponding requirement.

Figure 5. Evidence codes

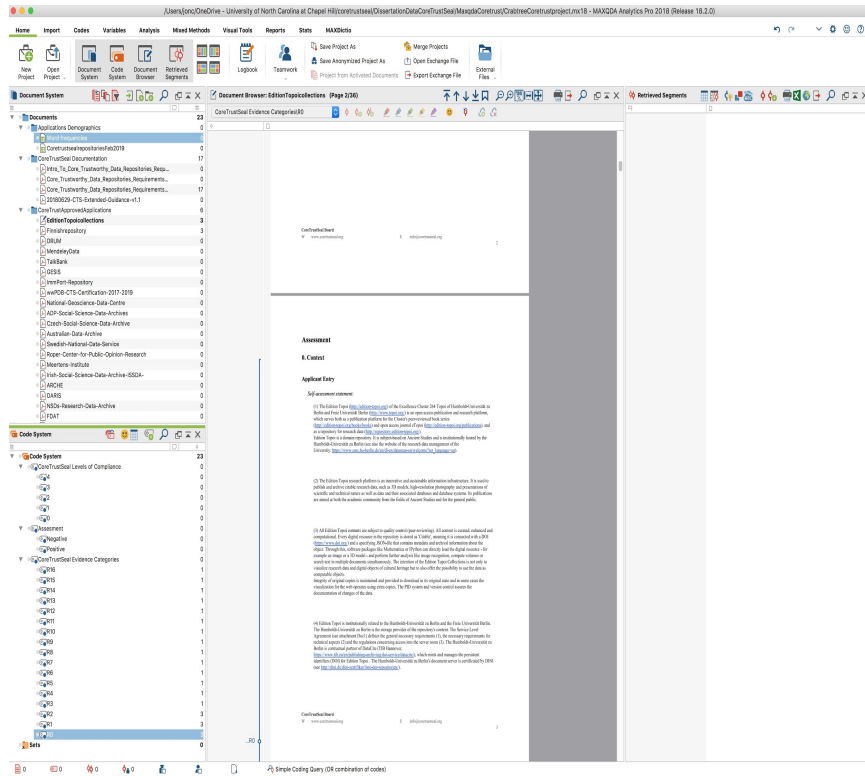


Figure 6. Requirements code

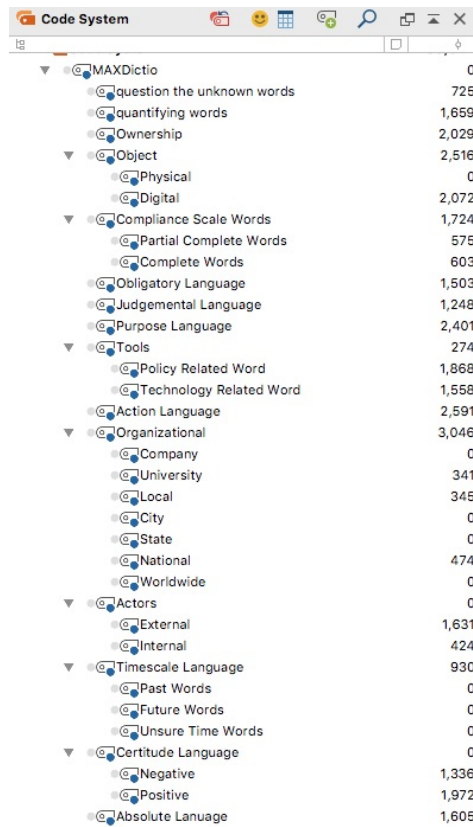
Code System	Count
Core System	23
CoreTrustSeal Levels of Compliance	0
4	0
3	0
2	0
1	0
0	0
Assesment	0
Negative	0
Positive	0
CoreTrustSeal Evidence Categories	0
R16	1
R15	1
R14	1
R13	1
R12	1
R11	1
R10	1
R9	1
R8	1
R7	1
R6	1
R5	1
R4	1
R3	1
R2	3
R1	3
R0	3

All sections of the documents were coded to identify the requirements they were intended to satisfy. To accomplish this, I manually selected each block of evidence in all documents and assigned the relevant codes, R0 through R16 (Figure 6), that correspond to their CoreTrustSeal requirement. This manual process was required due to the variable length of content in each section and proved faster than editing the applications individually. In addition, the sections of the application that contain the reviewers' comments and compliance levels were manually assigned a code to identify where the reviewers are assigning compliance levels and if they have any additional comments. These give structure for later comparisons across documents and requirements as well as reviewer comments.

After assigning predefined codes and variables to the documents, I began the process of discerning relevant topics by doing a close, attentive reading of the data and leveraged discourse analysis (Mihás, 2019). In this process, language is analyzed to identify potential areas for

group words by language categories. For example, when a person uses the word “required,” it suggests an obligation; when a person uses the word “approved,” they are assigning judgment. Language categories such as judgmental, organizational, obligatory, quantifying, certitude, timescale, and actors were used. These codes (Figure 8) are designed to explore for potential connections between applications that might identify topics to discuss with the reviewers during later interviews. This type of language codes are leveraged in discourse analysis and are focused on the language around an event and how it was used (Mihas, 2019). These categories address how decisions are made and frame the evidence for the reviewer. Also included in the applications are the comments of the individual reviewers for each requirement. These types of language codes help investigate their comments for potential information to expand the interviews beyond initial predetermined questions and codes. During this iterative process, new categories may evolve. These categories are developed in an inductive manner, where they come from the language of the text.

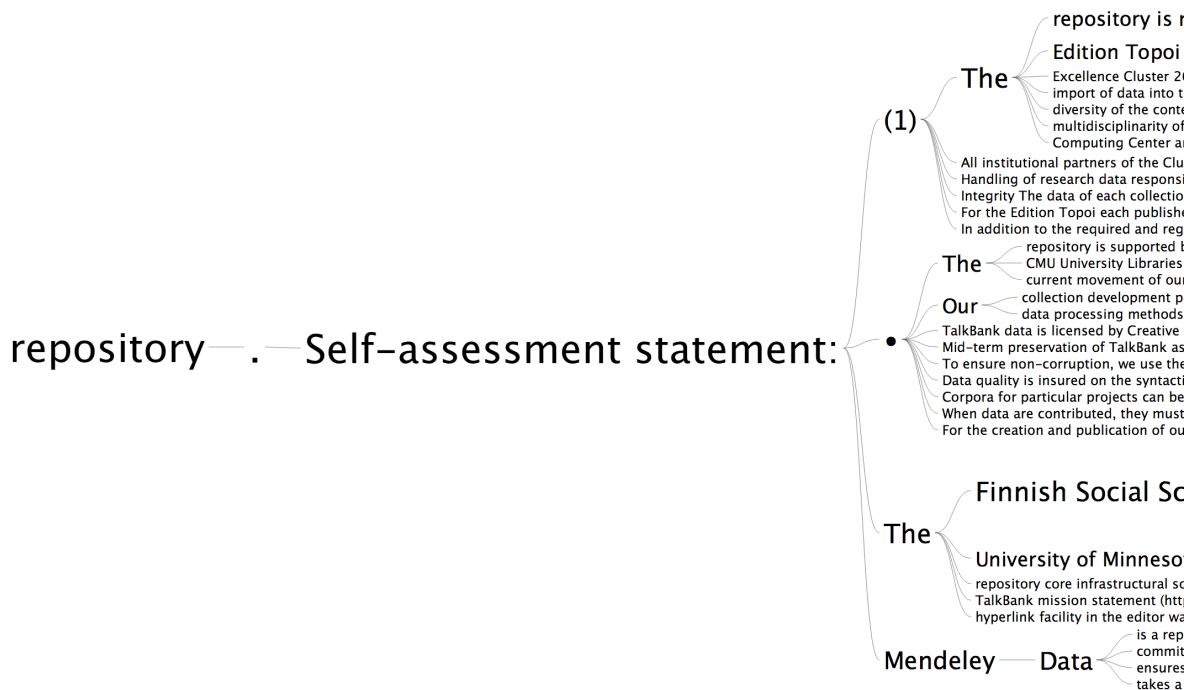
Figure 8. Judgment codes



Code System	Frequency
MAXDictio	0
question the unknown words	725
quantifying words	1,659
Ownership	2,029
Object	2,516
Physical	0
Digital	2,072
Compliance Scale Words	1,724
Partial Complete Words	575
Complete Words	603
Obligatory Language	1,503
Judgemental Language	1,248
Purpose Language	2,401
Tools	274
Policy Related Word	1,868
Technology Related Word	1,558
Action Language	2,591
Organizational	3,046
Company	0
University	341
Local	345
City	0
State	0
National	474
Worldwide	0
Actors	0
External	1,631
Internal	424
Timescale Language	930
Past Words	0
Future Words	0
Unsure Time Words	0
Certitude Language	0
Negative	1,336
Positive	1,972
Absolute Language	1,605

I selected the top 100 words from the generated frequency list and placed them in the category that most closely matched the use of the words. To help guide these decisions, I leveraged two features of MAXQDA to follow the most inductive process possible. The first was the Word Combinations feature, which looks at combinations of words across the documents. I used the 3- to 5-word combinations and, as previously done with frequency analysis, used the typical stop lists to remove words that were not useful. These word combinations helped show how the common words were used together by the applicants. I used these word combinations for a search in the second feature of MAXQDA, the Interactive Word Tree.

Figure 9. Word tree example

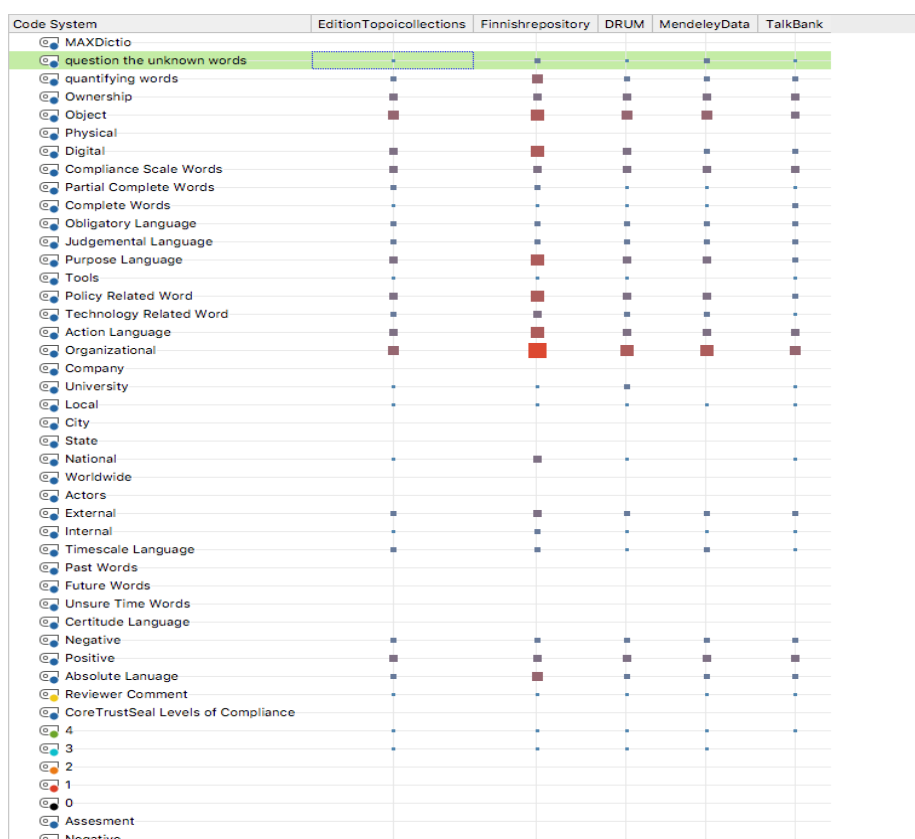


These word trees (Figure 9) allow the researcher to browse selected word combinations and see how they are connected in the text. By reviewing these trees, I was able to decide to which category each word belonged. Once I had a defined dictionary with the common words categorized, the next step was to use the Autocode with Dictionary feature in MAXQDA to autocode the words and the surrounding sentences. The codes assigned to categories associated with the words in the dictionary were automatically used. This autocoding process has identified more than 36,000 code segments to be used by tools to consolidate them into more usable categories later. This is the point where Bengtsson (2016) suggested the second step of recontextualization, where the researcher reviews the documents to ensure that all aspects have been covered in the coding approach. This requires a close attentive reading and review of the coded text. I used the MAXQDA document browser looking for gaps in the code coverage as well as leveraging the document comparison charts in MAXQDA. This close review of the

applications shows that very good coverage exists, so the data are ready to begin the third step of categorization and the final step of compilation and continued discovery analysis.

The MAXQDA code matrix tool was used for the analysis by visualizing the frequency of codes across the evidence documents, potentially suggesting critical similarities and differences (Figure 10). Figure 10 provides a sample on five documents. This process was performed on the entire collection to help explore the documents and related codes.

Figure 10. Code matrix example



To further identify similarities between applications, complex code queries were used to discover intersections and overlap across codes. To demonstrate this method, Figure 11 shows how the overlap of reviewers' comments and certitude language codes identify areas where reviewers used this language. This is a good example of an unexpected category that was helpful in shaping the stage two interviews. In this example, reviewers' comment codes overlap with

certitude language, suggesting that the reviewers requested something or had hesitations in the application. A closer look at the comments showed that some reviewers approved the applications, but with certain expectations noted for any future renewals of the certification. Provisional codes related to this example were created in stage two of the research, and new related questions were incorporated into the interview instrument. This code will be relevant, especially in RQ2 investigations into the perceived role of the reviewer.

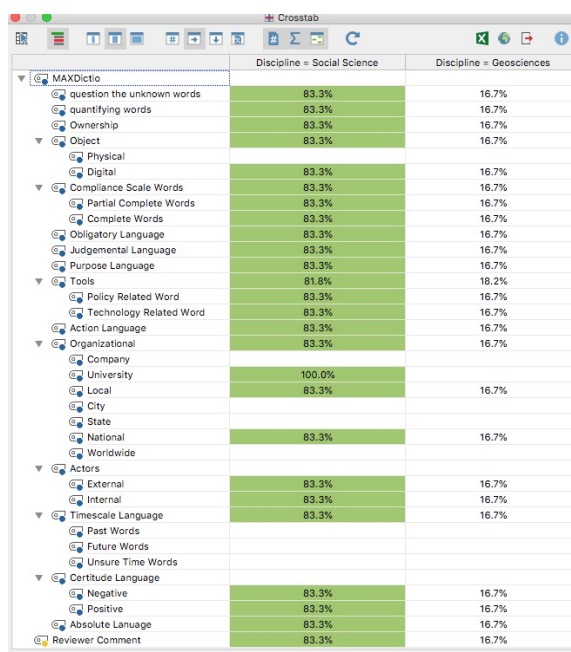
In addition to the qualitative tools available in MAXQDA described thus far, I employed a mixed methods approach by running cross-tabulations across documents' predefined variables identified earlier in this stage. Figure 12 presents an example of differences that were shown

Figure 11. Code intersections

Retrieved Segments	
<div>CoreTrustApprovedApplica...</div> <div>EditionTopolcollections</div> <div>4: 2410 - 5: 201</div>	<div>Reviewer Entry</div> <div>Accept or send back to applicant for modification:</div> <div>Accept</div> <div>Comments:</div> <div>CoreTrustSeal Board</div> <div>W www.coretrustseal.org E info@coretrustseal.org</div> <div>5</div> <div>1. Mission/Scope</div> <div>Minimum Required Statement of Compliance:</div> <div>0. N/A; Not Applicable.</div> <div>Applicant Entry</div> <div>Statement of Compliance:</div>
<div>CoreTrustApprovedApplica...</div> <div>EditionTopolcollections</div> <div>6: 74 - 7: 197</div>	<div>Reviewer Entry</div> <div>Accept or send back to applicant for modification:</div> <div>Accept</div> <div>Comments:</div> <div>CoreTrustSeal Board</div> <div>W www.coretrustseal.org E info@coretrustseal.org</div> <div>7</div> <div>2. Licenses</div> <div>Minimum Required Statement of Compliance:</div> <div>0. N/A; Not Applicable.</div> <div>Applicant Entry</div> <div>Statement of Compliance:</div>
<div>CoreTrustApprovedApplica...</div> <div>EditionTopolcollections</div> <div>10: 75 - 11: 168</div>	<div>Reviewer Entry</div> <div>Accept or send back to applicant for modification:</div> <div>Accept</div> <div>Comments:</div> <div>This is acceptable even though Edition Topol is in the implementation phase. They are moving towards "fully implemented" and have a plan to ensure the availability of data if the renewal proposal is not successful.</div> <div>CoreTrustSeal Board</div> <div>W www.coretrustseal.org E info@coretrustseal.org</div> <div>11</div> <div>4. Confidentiality/Ethics</div> <div>Minimum Required Statement of Compliance:</div> <div>0. N/A; Not Applicable.</div>
<div>CoreTrustApprovedApplica...</div> <div>EditionTopolcollections</div> <div>10: 75 - 10: 376</div>	<div>Reviewer Entry</div> <div>Accept or send back to applicant for modification:</div> <div>Accept</div> <div>Comments:</div> <div>This is acceptable even though Edition Topol is in the implementation phase. They are moving towards "fully implemented" and have a plan to ensure the availability of data if the renewal proposal is not successful.</div>

between disciplines of social science and geoscience. This is another example that will help in the second stage of interview instrument development. Differences between disciplines were anticipated, and the use of this predefined variable illustrates some differences. During a review of the documents, it became evident that some applications were more detailed than others, but none of the predetermined codes seemed to identify a pattern. This is another area in which reviewers' perspectives can be examined for more information.

Figure 12. Disciplinary differences



	Discipline = Social Science	Discipline = Geosciences
MAXDictio		
question the unknown words	83.3%	16.7%
quantifying words	83.3%	16.7%
Ownership	83.3%	16.7%
Object	83.3%	16.7%
Physical		
Digital	83.3%	16.7%
Compliance Scale Words	83.3%	16.7%
Partial Complete Words	83.3%	16.7%
Complete Words	83.3%	16.7%
Obligatory Language	83.3%	16.7%
Judgemental Language	83.3%	16.7%
Purpose Language	83.3%	16.7%
Tools	81.8%	16.2%
Policy Related Word	83.3%	16.7%
Technology Related Word	83.3%	16.7%
Action Language	83.3%	16.7%
Organizational	83.3%	16.7%
Company		
University	100.0%	
Local	83.3%	16.7%
City		
State		
National	83.3%	16.7%
Worldwide		
Actors		
External	83.3%	16.7%
Internal	83.3%	16.7%
Timescale Language	83.3%	16.7%
Past Words		
Future Words		
Unsure Time Words		
Certainty Language		
Negative	83.3%	16.7%
Positive	83.3%	16.7%
Absolute Language	83.3%	16.7%
Reviewer Comment	83.3%	16.7%

The goal of stage one was to provide some potential unexpected categories to supplement my existing knowledge and perspective on the review process. One concept was identified that helped redefine and shape RQ2 as well as contribute an additional set of questions to the interview instrument. Some reviewers used language that included future expectations for applicants. This concept led me to think about the role the reviewers view themselves as playing in the process. Conversations with colleagues further convinced me that part of this research

should investigate the perceived role with which reviewers identify while performing the reviews. The only other topic that emerged was the variation in detail across applications. Overall, the process was informative and allowed me to think more inductively about the applications. Although it did not radically change my concept or approach for the interview design, I think it was a valuable exercise and did provide information to refine my research and supplement knowledge needed in the design of the interview instrument.

Stage Two Work

The second stage of the project utilized the previous content analysis from stage one combined with my experiential knowledge as both a digital curator and a CoreTrustSeal reviewer to develop an interview instrument. This instrument was utilized in gathering the perspectives of various reviewers of CoreTrustSeal applications. My background and experience performing CoreTrustSeal reviews as well as current conversations and challenges faced by the CoreTrustSeal board were leveraged in the interview instrument design, but used in combination with the previous analysis of the applications for a hybrid approach. The concept of this hybrid design helped address potential challenges while developing the instrument. The primary reason was to help prevent bias from my experiential knowledge. I sought to control this by utilizing my influence and knowledge productively (Maxwell, 2013). This was demonstrated by assisting with interview instrument creation using inductive techniques in stage one, which helped prevent all questions coming directly from the researcher's preconceptions.

Study Population and Participant Selection

My study population was those reviewers who had performed CoreTrustSeal audits since the merger of the DSA and the WDS. I am personally very active in the archive and repository preservation community and utilized a social based recruiting technique (Guest et al., 2013) to

gather a maximum variation sample (Maxwell, 2013). Maximum variation in the subject's background related to education, training, discipline, and work experiences was considered. To operationalize this, I identified 37 reviewers who were currently participating in the CTS review process.¹⁰ An online background analysis of these reviewers was performed to identify five characteristics to use in this research: the country and discipline they represent as well as educational background, current position held, and if they held a formal archive degree. Three primary categories for their current positions were designed to standardize the position groups. The three groups were administrative roles, archivist roles, and technologist roles. For this study technologists are defined as those reviewers with educational degrees in computer science or held positions that manage technological systems. Archivists are defined as those reviewers that hold degrees in information and library science. Administrators hold positions managing organizations and help define policy for their organizations.

Thirteen initial participants were selected in this purposive of judgment sample (Kelly, 2010; "Purposive Sample," 2008), with design and thought toward what each of them may bring to the process.¹¹ Based on reviewers' professional, educational, and technical backgrounds, a diverse group was selected for initial interviews. The use of these various diverse sources provided triangulation to help in the validity testing of the results (Fielding & Fielding, 1986; Maxwell, 2013).

As CoreTrustSeal reviewers are not identified to the repository under review, I am required to maintain the anonymity of the subjects to protect that process and did not seek to identify which reviewers were assigned to each application. No linkages between these will be

¹⁰ See Appendix D

¹¹ See highlighted selections in Appendix D

constructed and any information provided in the interviews that could identify the applications reviewed was redacted.

Design of Instrument

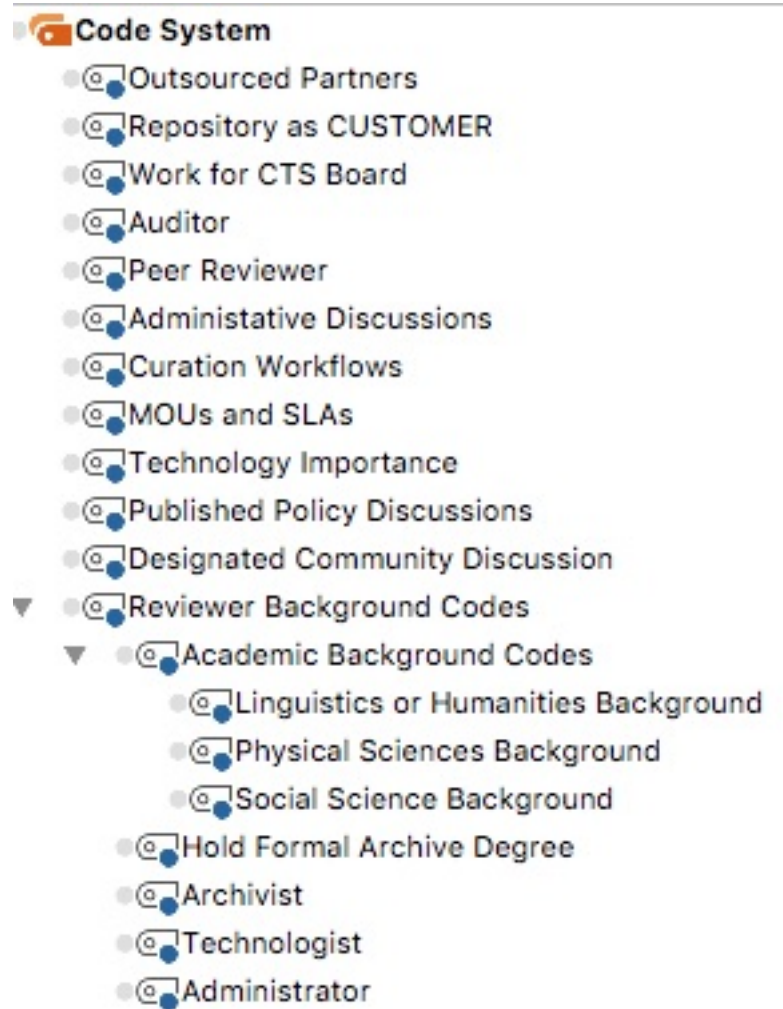
The interviews with reviewers utilized in-depth semi-structured interviews (Seidman, 2013; Stake, 1995) to understand each perspective completely enough to identify uniqueness between them (Guest et al., 2013). This open-ended approach provided a conversational look and feel for the instrument (Guest et al., 2013). The semi-structured approach used a set of defined questions as an instrument, but it must be used with caution to allow participants enough freedom to introduce new topics into the interview and the researcher to discover more using an inductive approach. This is a very common interview design and typically has six to 12 questions, with the possibility that each question could have up to four sub-questions (Rowley, 2012). The key to this approach is the flexibility during the interview to probe deeper to ensure that the main questions were covered sufficiently (Rowley, 2012). What makes the design of this interview instrument unique was the use of the stage one content analysis of the CoreTrustSeal applications to help shape the questions. As previously noted, some question topics had already evolved during exploratory analysis.

My experience and position on the CTS board gave me a unique perspective that I could leverage in developing both the interview questions and the predetermined provisional codes (Saldaña, 2016). These provisional codes built the initial code base, but with the understanding that they would expand to more categories during the interview process. Some researchers start with as few as five to six codes while others recommend 12–60 codes (Saldaña, 2016). I identified 15 provisional codes (Figure 13) in addition to the 20 initial codes. These were

described in the first draft of a codebook.¹² These codes were primarily designed to help identify differences in reviewers' perspectives needed to investigate RQ1 but were also valuable in understanding any potential variations in perspectives on their roles in the process for RQ2. These codes were generally divided into two groups. The first helped confirm the background research on the reviewers' educational and professional training. The second was used to look for levels of importance on topics like technology, administrative policy, and curation workflows.

¹² See Appendix F

Figure 13. Provisional codes



Three codes were included to help investigate RQ2 and the role of the reviewers. These were added after stage one work identified this as another area of interest. These provisional codes were used to begin the evaluation of the interviews and were expanded as new topics, as seen in the interview transcripts later.

The design of the semi-structured interview instrument was broken into six sections.¹³ The first is the introduction designed to get permission to record the interview as well as inform

¹³ See Appendix E

the participant of any review board protocols. The second section was designed to help clarify reviewers' background information. All efforts were made to describe and place the reviewers in the correct category of technologist/archivist/administrator, but some information was not available online and in many cases the decision was not so simple. Two questions targeting this information were used to validate the choice of category for each person.

The third section began with questions targeted at generating a conversation around the review process and making the reviewers feel comfortable. They also helped determine if any similarities or differences in perspectives emerged that could impact their evaluations. Questions that identified participants' views on strengths and weaknesses of CTS as well as any challenges they faced were also included in this section. The next section of questions examined the reviewers' perspectives on the importance of various requirements. As reviewers cannot be asked specific questions about individual applications due to privacy reasons, these questions focused on describing and ranking evidence categories. Reviewers were asked to rank their top three requirements and explain why they thought these were the most important. In addition, they were asked to describe the most critical components of these requirements. The intent was that differences could emerge between reviewers that could be based on various differences in their backgrounds or perspectives.

The fifth section of the interview instrument was designed to focus on RQ2 and the topics identified during the stage one work. These questions investigated how the reviewers saw themselves in the CTS process. Did they see themselves as an auditor, formal reviewer, or peer-to-peer helper? Did they see themselves as working for the CTS board or for the repository? Probing techniques were used to further identify any topics that should emerge from this line of questions. In addition, this section included a question that sought to identify any differences in

detail they may have noticed across the applications they reviewed. The final section included additional open-ended questions to probe for advice for future applicants and potential changes to the CTS requirements. These were examined to help identify any differences in perspectives that could be used to explore RQ1.

Interview Procedures

The interview instrument was tested on three pilot subjects (Kelly, 2010) to adjust for timing and the exploration of potential issues with the instrument as well as the multiple technologies that would be used to record the interviews. The interviews were recorded in three fashions for future coding and analysis to ensure complete redundancy. The first was a standard digital recorder, which provided a reliable and solid base for future annotation and coding of the interviews. The interview process was designed to utilize our most basic mode of inquiry (Seidman, 2013). I was listening for and probing into the processes and procedures used in the audit performance in the hopes that others could leverage the sharing of these experiences in the community. The second method used a Livescribe (2018) smartpen that recorded the interview and linked any interviewer's written notes to the recorded conversation. I also used the MAXQDA App to record the interviews on my iPhone for transmission to MAXQDA for analysis.

These three techniques provided redundancy as well as added unique insights into the interaction between participants and the interviewer. As I had identified some categories during the first stage of the project, I noted such categories during the participants' responses when they described that part of the conversation. A challenge for me personally was directing the conversation. One of the attributes of a good interviewer is knowing when and how to probe and direct the interview. Seidman (2013) concluded that "listening is the most important skill in

interviewing,” which is demonstrated more bluntly in the section entitled “Listen More, Talk Less” (p. 81). The goal is to utilize just enough probing to get inside that inner voice of the participant (Seidman, 2013). During the interviews, I took care to ensure that the topics were sufficiently covered, and that bias did not cause issues.

Allowing the participants to talk freely about the topic while being guided by the process required me to concentrate during the process and leverage techniques that ask additional questions when responses were not clear (e.g., Tell me more about that; Why did you feel that way?; Seidman, 2013). I avoided leading questions during the interview and was careful not to interrupt when following up. Understanding that interviewers and participants in a qualitative interview are never equal (Seidman, 2013), I paid close attention to a number of social and demographic constructs that can cause bias (Rowley, 2012; Seidman, 2013).

Determining the End of the Interview Process

Two criteria are used to determine when a researcher has enough interviews, and the interview process has ended. The first is sufficiency (Seidman, 2013). There needs to be enough coverage to reflect the range of participants that make up the population. The use of maximum variation in the purposeful sample (Rowley, 2012), as previously discussed, helps with this aspect. The other criterion is saturation, which is a point where the researcher is not learning anything new (Seidman, 2013). Research has shown that nonprobabilistic samples utilizing semi-structured interview instruments achieved thematic saturation data after 12 interviews (Guest et al., 2013). The more structured the instrument is, the faster saturation occurs. But this is typically a guide, and other articles have suggested looking for saturation after 10 interviews and continuing the interviews until it is achieved (Galvin, 2015; Hagaman & Wutich, 2017).

After each interview, I transcribed and coded it in MAXQDA based on the predetermined

provisional codes identified prior to the interviews as well as emerging codes identified during the interview and transcription process. The interviews were continuously coded after each interview and reviewed using the thematic analysis techniques that will be described next, in stage three. Sometimes two or three interviews were done on the same day. In these cases, the coding was done at the same time after they all were transcribed. When informational redundancy emerged in the interviews, it signaled a potential end to the process.

Stage Three Work

The third stage of this research used a thematic qualitative analysis approach to not only determine the similarities and differences in the CoreTrustSeal reviewers' perspectives, but also to identify the themes (Rowley, 2012) in the interviews and conceptualize them. The goal was to provide a description of the identified themes as well as name those themes and use them to build a conceptual framework of the materials. The thematic analysis process is iterative. Themes are identified in the transcripts and then coded and compared to identify commonalities, which often identifies new combined themes (Castleberry & Nolen, 2018). This comparison between the participants informs the final output, with the goal of better understanding potential subjective differences between the respondents.

In the first phase, the researcher builds a familiarization with the documents (Nowell et al., 2017). Nowell et al. explained that it is best to follow guidelines to make the thematic analysis more trustworthy. I reviewed each of the transcribed interviews for completeness and then checked for errors against the recorded interview. I also read the transcripts again to become more familiar with the content. After gaining an overarching perspective on the materials, coding the thematic analysis continued (Maxwell, 2013) to identify connections and links between the unique perspectives. This concurred with Nowell et al.'s (2017) six-phase approach to trusted

thematic analyses. The adapted step-by-step approach I used follows that concept and includes:

Phase 1: Familiarize yourself with your data

Phase 2: Generate initial codes

Phase 3: Search for themes

Phase 4: Review content for potential themes

Phase 5: Define and name themes

Phase 6: Produce the report

During phase 2, my plan was to identify the initial codes from the data in the transcripts and create specific code boundaries to enhance the existing codebook to ensure a consistent approach (Braun & Clarke, 2006). This code framework describes how codes were assigned along with a clear audit trail to provide a consistent result. A defined codebook for the codes was also used to help with a clear audit trail. The codebook—a list of codes with their definitions—helps ensure that codes are adequately described and systematically applied. Codes were applied to the text using the autocoding features of MAXQDA as well as any manual coding as needed.

Once the data had been coded, the discovery of themes in phase 3 was the next primary task. DeSantis and Ugarriza (2000), who write in the nursing field, described themes as follows:

A theme is an abstract entity that brings meaning and identity to a recurrent experience and its variant manifestations. As such, a theme captures and unifies the nature or basis of the experience into a meaningful whole. (p. 362)

Leveraging the MAXQDA tools such as advanced code search, code matrix browsers, and word trees, which were described earlier, I sought to identify themes and combinations of codes that lead to new themes (Castleberry & Nolen, 2018). This iterative process continued until saturation was achieved.

Upon saturation, phase 4 began by looking to check for areas missed by the earlier

thematic analysis. By looking at code patterns and code combinations in the MAXQDA application, I sought to identify more themes that were potentially missed in phase 3. This review checked for accuracy as well as demonstrated saturation of the themes to ensure they had proper coverage. Often during this process it becomes clear that some potential themes do not have enough data for them to be useful, so those can be collapsed during this process (Nowell et al., 2017). This quality assurance step helped make the analysis more trustworthy.

During phase 5, I reviewed the themes that had emerged, named or renamed them appropriately, and identified their relevance for the study's objectives (Braun & Clarke, 2006). I used this naming and renaming of the themes process to help move from description to concepts. This iterative process sometimes required new themes or the collapse of multiple themes into new single themes.

The final phase began once all the themes in the transcripts were identified and named and when representative quotations within each theme were reviewed to ensure that the themes made collective sense. This is where the conceptual framework was identified and described. Once the themes converged around a story, the analysis built the final conceptual framework, and the summary could be written in this final phase. The themes were tied to direct quotes from the participants and the applications reviewed to better tell the story.

CHAPTER 4: RESEARCH RESULTS

The results are organized into three main topic areas. The first section discusses reviewers' perspectives. Each participant was asked to describe their perspective and the category with which they felt it was best associated. This first section discusses how these descriptions compare to the initial selection criteria used to place them in the group of reviewers to be interviewed. The second section compares interview responses across these perspectives to identify any differences. The section focuses on determining if the subjective responses to the presented evidence varied across the different categories of perspectives. The final section describes reviewers' perspectives on the CoreTrustSeal processes and examines their opinions as to the CTS's strengths and challenges.

The reviewers' quotations and viewpoints gathered from the interviews are presented using the following standards:

- Square brackets [] are used when I supply information that is not in the original words of the reviewers. For example, if a reviewer used the CTS requirement number in the sentence to refer to a review criterion for example "R1 for Mission Statement," I might replace that with "[Mission Statement]" to enhance readers' understanding.
- Ellipses (...) indicate that I have taken out sentences from the original interviews.
- (pause) indicates a pause.

To protect the confidentiality of the reviewers and ensure the integrity of the CTS process, the respondents were assigned anonymous identifiers ranging from 13 to 25. Initial instrument testing and various recording equipment tested leveraged identifiers 1–12, so 13

became the first valid interview. These will be utilized to identify the source of the quotations included in the results. In addition, the repositories and organizations mentioned by the reviewers as well as any other identifying information were anonymized to ensure confidentiality in the study.

Reviewers' Perspectives and Perceived Roles

In total 13 interviews were conducted with individuals who had previously been involved in CTS application reviews. These interviews were all performed in person at various locations around the world. The goal of the sample was to have a balance in perspectives. Reviewers were asked to self-identify as to which perspective best described them. Seven of the 13 reviewers self-identified as expected based upon the public information found in their curriculum vitae or on published websites (see Table 1).

Table 1
Reviewers' Perspectives Data

Document name	Anticipated Perspective	Self Identified Perspective	Additional Perspectives Mentioned	Archivist	Technologist	Administrator
finalingest_190725_0018	Archivist	Administrative	ALL	1	2	1
finalingest_190725_0017	Administrator	Administrative	Archivist	1	0	1
finalingest_191025_0025	Administrator	Administrator	Archivist	1	0	1
finalingest_190726_0019	Administrator	Administrator	Technologist	0	1	1
finalingest_190529_0013	Administrator	Administrator		0	0	1
finalingest_191024_0024	Archivist	Administrator		0	0	1
finalingest_190531_0014	Administrator	Archivist	Administrator	2	0	1
finalingest_191023_0022	Technologist	Archivist	Technologist	2	3	0
finalingest_191025_0026	Technologist	Archivist	Technologist	1	1	0
finalingest_191024_0023	Archivist	Archivist		1	0	0
finalingest_190722_0015	Technologist	Technologist	Archivist	1	2	0
finalingest_190724_0016	Archivist	Technologist	ALL	2	1	1
finalingest_191022_0021	Technologist	Technologist		0	2	0

In some cases, reviewers were confident in the decision, as with reviewer 13, who said, "I'm absolutely an admin." For other reviewers, this was a hard decision. In some cases, they could not make the choice. They preferred to rank the categories in order they apply to them the most.

I think I'm doing more admin work than anything else. And I would say administrator first, uh (pause), technology may be second and archive is third, but again, I don't have any formal [archive training]. (19)

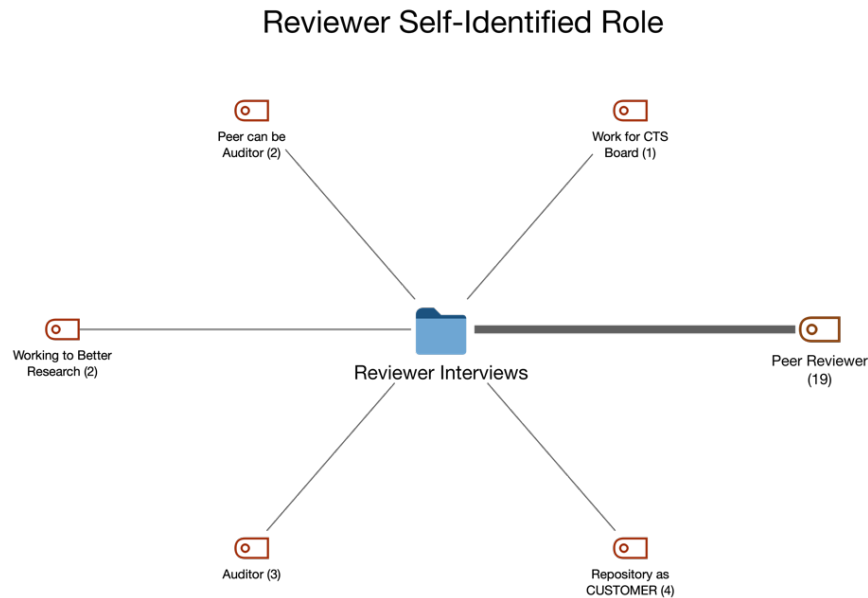
So I'm not a developer, but I, (pause) there is definitely a big focus on technology. And so (pause) I'd put (pause) I'd probably put technology over archivist and administration. (16)

In some cases, the interviewees used hedging language that indicated it was not an easy decision, such as 15, who said, “probably technologist.”

Many times, the reviewers felt they had several perspectives. All but four of the reviewers mentioned other perspectives or identified with more than one of the three choices. Two reviewers felt they filled all three roles. In only one situation was the anticipated perspective not correctly predicted, and the reviewers did not mention the anticipated perspective as being a function they performed. The final number of reviewers that self-identified as administrators was slightly higher than the other two categories, but given the difficulty some had in selecting a primary category and the balance in the resulting responses, this did not affect the ability to achieve convergence in the themes as predicted.

The perspective different reviewers gained from their various educational, training, and work experiences had the potential to impact the subjective review of evidence. But does the role these reviewers perceive they serve in this process also potentially affect this review of evidence? Each reviewer was asked if they felt they were acting as peer reviewers trying to help the applicants or auditors that would only judge and not aim to assist the applicant achieve the seal. In addition, they were asked whether they felt they worked for the CTS board, the repository as a customer, or the research community in general. All but one reviewer felt they were acting as a peer reviewer (Figure 14).

Figure 14. Reviewer self-identified role



Even the one reviewer who felt they were an auditor also thought they could be a peer reviewer with no conflict. On several occasions, the feeling of helping a peer organization was expressed.

Reviewer 22 felt the role was similar to reviewing articles:

My feeling is I'm really helping them to improve, which is similar to when I'm reviewing a paper for publication. My hope is that through my review the archive will go back and say, you know, we need to pay attention to this part. Maybe we need a procedure here.

The concept of helping a peer was strong and is a good sign as the CTS business model is built around peers helping in the review process as volunteers. But as reviewer 25 noted, we also need to uphold the standard:

I'm trying to help the peer because I think I already said that with all of this work, you know, I do it because I believe in the greater good of it all. And, but there is also, um, a reputation of, um, standards to take care of and that is the, the kind of balance that we need to find.

This role of guiding the applicant is also seen in the suggestion for future improvements during the next regularly scheduled review. In the experience of reviewer 24, they “certainly have written things along the lines of yeah, this is okay, but in future you might like to (pause).” Even

when not expressed as a category, the concept of improving the research community was identified with. Reviewer 15 mentioned that they “usually ... tried to look at it with, from the perspective of the research community.” Reviewer 16 felt strongly that our goal was to better the research community:

Yeah, I would definitely say [we are working for] the research community at large and (pause) what the repository that's applying is doing, I mean they are trying to improve the services or the quality of what they offer. (pause) because ideally, I mean they want to support research, reproducible research the community. So, I think that's what we're doing it for.

The concept of a need to balance the peer review role with the more formal audit concepts was also mentioned. One reviewer pointed out that the goal should be clear communications of why the requirements are there:

It's all about clearly communicating. I think that both parties understand each other, you know, the researchers have certain needs, but also the repository has certain needs from a technical perspective. And they don't often speak the same language. So that is it. It's also a matter of being able to explain why these things are necessary and why certain, for example, file formats might be better suited for long term preservation. (15)

The balance of perspectives by participants in the study was as expected, which provided the perspectives of administrators, archivists, and technologists that were needed to evaluate the potential subjective impact on reviews. Many of the reviewers felt they had several perspectives and could potentially see themselves as more than one of these perspectives. The most identified category was administrator. Although not investigated in this study, it could be that the seniority of the reviewers impacted this perspective. Almost all of the reviewers were senior members of the community, and many were also CTS board members.

The CoreTrustSeal and many organizations that perform repository evaluations in general are relatively new. These organizations are often formed by leaders in the digital preservation community and typically include senior repository managers. This resulted in the majority of

CoreTrustSeal reviewers having lots of experience, and they often either served on the CoreTrustSeal Board or were senior members of their home organizations. This study was limited to reviewers who had performed review of CoreTrustSeal applications. The total population to select from was limited to 23 reviewers due to the fact that a reduced number of CoreTrustSeal applications had been reviewed. As of this writing, many of these potential reviewers are senior members of the community and often have experience as administrators in addition to their original training and education. This was shown by the fact that many reviewers felt they had more than one perspective and selected more than one category.

This is subject to change as the number of organizations that achieve the CoreTrustSeal provide additional peer reviewers. The number of successful CoreTrustSeal applications doubled since the beginning of this research, and many more reviewers had become active in this community. If this trend continues, the potential for more reviewers with less experience will increase and could potentially impact the subjective review of applications.

In contrast, almost all reviewers who participated in this study felt they served the role of a peer reviewer. Too few people in the sample felt that they worked for the CTS board or were official auditors to make a strong observation of this role. As the CTS standard is driving and supported by a group of peer repositories, this is not a negative result.

Exploring Potential Impacts on Evidence Review

CoreTrustSeal applications are evaluated based on 17 requirements labeled R0–R16 (CoreTrustSeal, 2017b). These are grouped into three general categories of requirements that consist of organizational infrastructure, digital object management, and technology. To set the context for the review, the R0 requirement gives an overview of the repository, defines the designated community, describes the level of curation performed, and identifies any external

partners in the workflows. The three categories roughly align with principles and activities with which administrators, archivists, and technologists are often tasked as demonstrated in OAIS. In the organizational infrastructure category are requirements that describe the repository's mission statement, license management, continuity of access plans, ethics statements, staffing levels, and training plans. The digital object management section describes the plans for appraisal, digital object storage, preservation, data quality controls, curation workflows, and data discovery. As expected, the technology category evaluates the various technical specification and security plans provided by the repository under review.

Most Important Evidence

Reviewers were asked to identify the elements of the required evidence that they felt were most important. The codes identified during the analysis align with the CTS requirements to aid in identifying the potential impact of reviewers' perspectives on the review of this evidence. In addition to the direct questioning of the reviewer in these areas, notations were made each time a reviewer mentioned the importance of a category. The summation of these combined references to various categories was leveraged to identify the level of importance reviewers perceived for these types of evidence. The category that received the most references or selections was the repository overview and context. On several occasions, the reviewers felt it was the most critical component. Reviewer 13 felt it was critical:

unless an organization answers the contextual part, well then, the rest of the application is almost pointless ... Understanding what it is that repository is trying to achieve and for whom is absolutely vital. So, this is a difficulty that the reviewers have faced always and it's a difficulty that that means that absolutely the repositories have to answer that section.

Reviewer 16 concurred, mentioning that curation "is one of the most important [requirements] simply because ... without the information I can't, I just (pause), I can't review the application because I don't have the background information."

Throughout the interviews, reviewers mentioned the importance of this requirement 20 times (Table 2). Although the identified technologists did mention this category fewer times, it was relatively evenly mentioned across the groups.

Table 2

Most Important Requirements

Interview Codes	Archivist > 0	Administrator > 0	Technologist > 0	Total
What are most important requirements				
Policies Online		1	1	2
Curation Workflows	4	3	4	11
Documentation of Agreements	1	1		2
Mission Statement	3	1	1	5
Repository Sustainability	1	2	1	4
Repository Overview, Context and Designated Community	7	8	5	20
Technology	2	4	2	8
Transparent Processes	3	2	2	7
Preservation of Data	5	4	6	15
Blend of All	1	1		2
Funding	1	1		2
Data Integrity	2	1	1	4
SUM	30	29	23	82
N = Documents	9 (69.2%)	8 (61.5%)	7 (53.8%)	3 (100.0%)

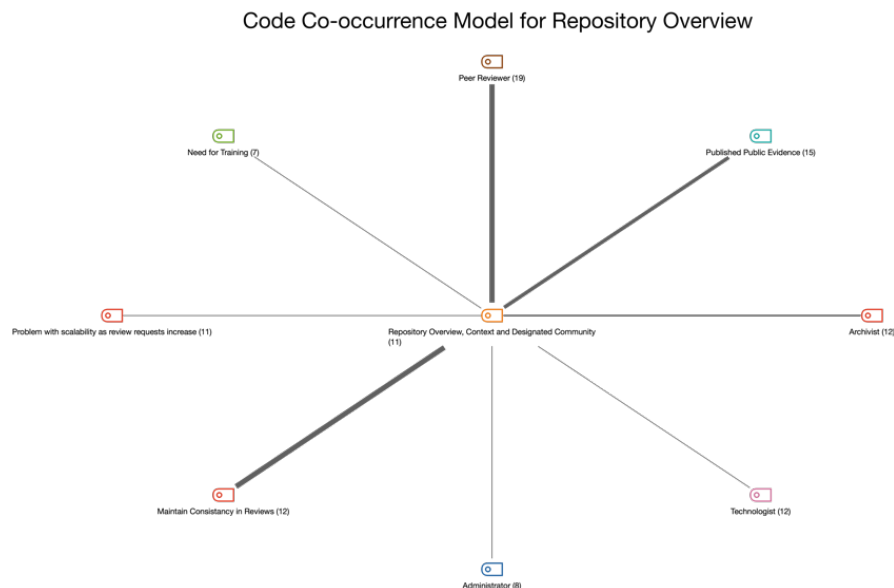
The category mentioned the second most often was the preservation of data. This was discussed 15 times during our conversations and was also very evenly distributed across the identified groups. Some reviewers ranked a couple of priorities in response to this question. Reviewer 21 felt that the most important category was the “preservation plan (pause),” but then also mentioned that the “repositories context and designated community” were the second most important. Reviewer 19 had the same perspective. They thought that the most important “requirement is the preservation plan [inaudible] and [repository context and designated community discussion] for [them] also is a key [requirement] because it really is synthesis of many things” Reviewer 19 self-identified as an administrator, and reviewer 21 identified as a technologist, yet they held the same views on the most important requirements.

The category of curation workflows was ranked as the third most discussed requirement and was also discussed by all groups evenly. Although mentioned by several reviewers, only one

reviewer felt the single most important requirement was curation workflow evidence. For them, the primary evidence that was most important was “the data curation workflows, ... (pause), the level of curation has [to be described]” (24). Reviewer 24 self-identified as a technologist but also had formal archival training and their anticipated perspective was as an archivist. As they had formal archival training, it is not surprising that they valued the curation workflow.

Those reviewers who discussed the repository context as being the most important were often concerned about maintaining consistency (Figure 15) in the review as the CoreTrustSeal increases in popularity and the number of reviews increases. They also valued the present evidence being publicly available for the designated community to review. Transparency during the peer review process is important to this community.

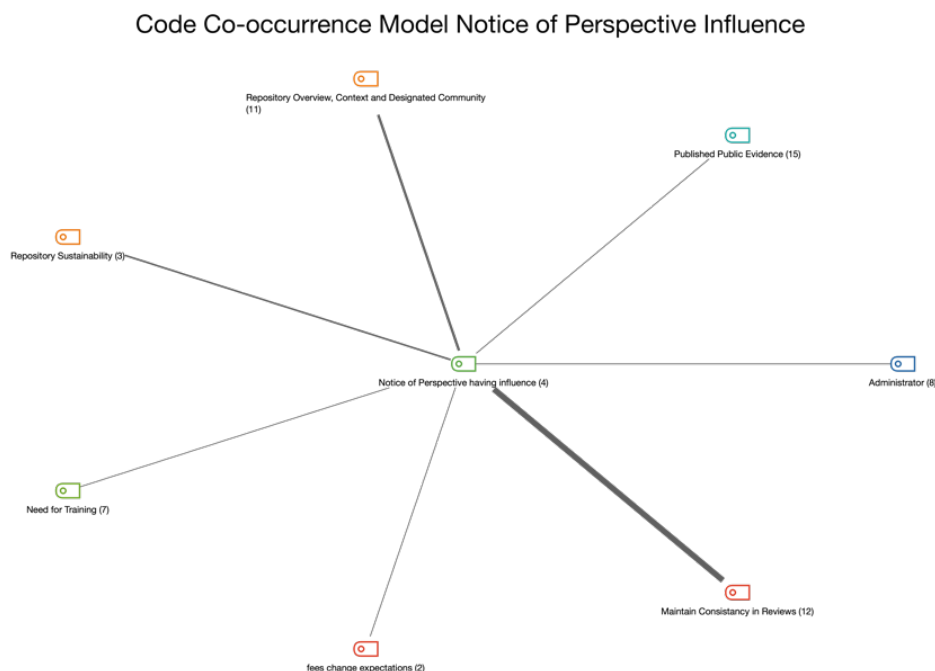
Figure 15. Code co-occurrence model for repository overview



Several reviewers noted that reviewer perspective has the potential to impact the subjective review of the presented evidence, which was the original hypothesis in this study. Some reviewers shared the concern that it could impact the consistency of reviews moving

forward as a larger number of reviewers are utilized as CTS performs more reviews (Figure 16).

Figure 16. Code co-occurrence model notice of perspective influence



The business model of the CoreTrustSeal depends on volunteer reviewers from approved repositories to review their peers. This transparent peer review process is important to the community. It is not surprising that participants mentioned this influence while answering questions about both their own perspective and their views on presented evidence. Many of these questions in combination would lead participants to think about this impact. Reviewer 13 had a particularly important perspective on this topic:

I think human nature says that we, (pause) we will judge, and we will maybe judge in a quite strong way based upon our biases. And that's something that we need to try and obviously stop. But I think people, I hope, I always hope that who are interested in being a reviewer do it partly because they have an interest in learning more themselves and how things might help their own repository and their own experiences, but also that they have the opportunity to help others.

Reviewer 25 mentioned the subjective nature of the reviews but also noted that this is

unavoidable and doubted that these reviews could be automated:

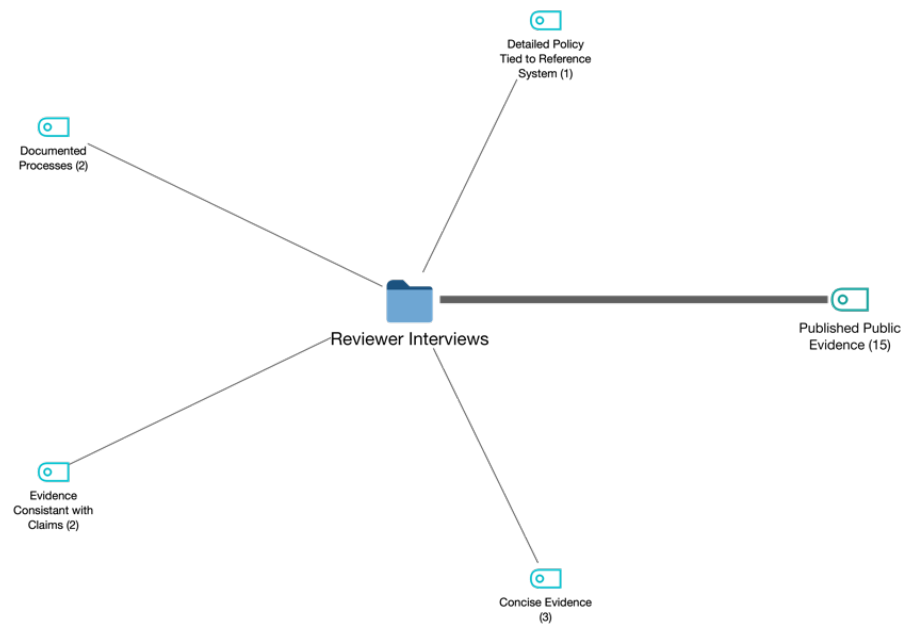
It's in a way it's subjective and you can see that because we come to other conclusions based on the same materials. So that's why I also struggle with that we can ever, (pause) ever make this [a] machine process.

It is important that reviewers understand their biases during the review process. This is encouraging because the first step in controlling the consistency of reviews could be the understanding of these biases based on perspective.

Best Forms of Evidence

When reviewers were asked what the best kinds of evidence were, there was strong agreement. They would like to see published evidence that is viewable to the public. The CoreTrustSeal application requires a summary of the repository's policies in English for the application, but typically these summaries reference external policy documents that may not be public or in English.

Figure 17. Best kind of evidence



During the interviews, all but one reviewer mentioned the need for public evidence (Figure 17) as being very important. Reviewer 13 stated that it is often the first thing they look for: “Well, I mean the first thing you look for is do they have they included a URL?” The concept of having well-documented, publicly available policies was also noted by Reviewer 19 as a sign the repository is well organized and that it takes a structured approach to the preservation of the digital objects it is entrusted to preserve:

What I like is a clear summary in the application. Not too many details and links to public documents available on the repository website, not wiki hidden in somewhere that only they can access. I really like that's what we call public evidence, right? Anyway, so yeah, formalized documents I think really is a sign of structured approach to organizing your repository. If you have those documents really referenced, labeled made available on the website, even if it's not a document in the sense of a document, if it's a website within page, I mean html pages work fine for me. As long as everything is public documented and your users can see it, access it, understand what you're doing.

Reviewer 17 mentioned that “actually the form is not [important], if it's like a, a pdf or

whatever, it doesn't really matter to [them], but [they] think it's really important also that it's publicly available.” According to Reviewer 15, “there's a large emphasis on [evidence] being public. So, [reviewers] think that it's also important that [the repository] have at least some general documentation and some specific documentation on [evidence] like preservation policy has on publicly available [sites].” The form of this evidence can vary, but it is very clear that reviewers want to see these posted in public places.

Advice to New Applicants

When new repositories embark on the process of becoming certified using any standard, the task can be demanding. CoreTrustSeal utilizes a simplified peer review process that leverages 17 broad categories. This method does reduce the requirements, but sometimes describing a system as complex as digital archiving in broad, yet concise terms can be challenging. The reviewers collectively provided some advice during these interviews. Several reviewers provided suggestions similar to other responses, yet even the less frequently discussed suggestions can provide valuable advice to new applicants.

One of the more obvious strategies that is often overlooked by new applicants is the review of applications that have been previously approved (Table 3). Once approved, all applications are published for the community to review on the CTS website. It sounds simple, but it is effective and, as Reviewer 26 mentions, a new applicant should seek “to find a similar institution that has the CoreTrustSeal and to read their application.” It does help to look for an approved repository that has a similar designated community or content type. As Reviewer 16 suggested, the new applicant can always ask the CTS leadership to suggest an exemplar application that is similar to use:

Well, first of all, ... have a look at the published applications out there. Maybe ask [CTS] before [you start] is there ... some applications ... in my domain that you would

recommend we can look at before we apply?

Table 3

Advice for New Applicants

	Archivist > 0	Administrator > 0	Technologist > 0	Total
▼ Interview Codes				
▼ Advice for new applicants				
Prepare Public Evidence	2	1	1	4
Detailed Policies the Follow Reference Model	2	1		3
Administrative Buyin Needed	3	3	1	7
Prepare Ahead		1		1
New Applicants Benifit from a Facilitator	1	1		2
Review Approved Applications Similar to their Repository	3	2	2	7
Follow CTS Checklist			1	1
Gain Overall Understanding of Requirements	1		1	2
Be Concise	2	1	3	6
Read Extended Guidance		1	1	2
Start with Team Meeting	1	1		2
Σ SUM	15	12	10	37
## N = Documents	9 (69.2%)	8 (61.5%)	7 (53.8%)	3 (100.0%)

Several reviewers talked about the need for an organization to include a wide-ranging group of people in the process and to be as inclusive as possible. This clearly requires administrative buy-in to get the resources needed to submit the application. Reviewer 17 explicitly mentioned that the process would be long and hard without full buy-in from the organization:

I [feel], approval or a high level buy in from. Yeah. Because ... you need a lot of different people and ... if you are not supported from, from top down, uh, will be very difficult for you ... to get because it's quite a long process.

To prevent this process from missing a critical input, Reviewer 25 suggested having a meeting at the repository to get organized.

My advice that I always give to people who come to me is, um, that I say, what do you need to do is organize an afternoon at your repository. So, there needs to be someone who is driving this process, um, or [go to] person. So you take that list of [CTS] requirements, you go through them and for each requirements you think, Oh, this is something for [Person A] or for [Person B] in your own institution you said you inform those people, you give them the requirement that is um, closely [related] to their job, ask them to fill out right from the top of their heads what they would answer if they were asked. And then you sit together for a day or an afternoon with all of those people in the room and you tried to assess where you are.

This process quickly gets a draft on paper and identifies holes in the application that point to the need for additional documented policies or plans. Several reviewers also mentioned the need for evidence to be concise. Reviewer 15 thought that evidence “should be substantial, but sometimes, you know, being a bit more concise makes sense. It makes it a lot better.” Reviewer 22 felt strongly that the applications needed to be concise:

The most important advice is to be concise. I think, you know, say, I mean if you really are meeting a requirement, you should be able to say, uh, in a nutshell how you're meeting that requirement and then you can expand on it maybe, but, uh, uh, you know, some really brief statement should be able to say it and then, um, uh, I don't think there's a need to go on and on for pages.

Some of the suggestions were not mentioned by a large number of participants but are still important to include in future information produced to help new applicants. One of those suggestions was to read the CTS Extended Guidance before starting. This is an often-overlooked document that provides guidance for both applicants and reviewers alike. This simple suggestion would help allow new applicants understand the requirements and get an idea for what a reviewer will be looking for.

Reviewers’ Challenges and Perspectives on the CTS Process

Reviewers were also asked to talk about the challenges they face while doing reviews as well as any concerns they had with the CTS process. The goal was to help provide input to the CTS community and governance structure to better enable a sustainable model for CTS reviews. The CTS community is growing rapidly and adding new reviewers each month. The timing of this study early in the life of CTS provided a relatively small population of reviewers that had performed reviews against the standard. The sample also included some of the original members developing the standard and several reviewers with many years of experience managing repositories. New applicants and future reviewers have much to learn from this dedicated group

of experienced reviewers.

Biggest Challenge for Reviewers

Reviewers were asked to speak about their biggest personal challenges while performing reviews (Table 4). Some responses were similar and provided some insights into issues future reviewers could face. One challenge was the translation of documents from their source language to a summary in English, as required by the application.

Table 4

Biggest Challenges for Reviewers

Interview Codes	Archivist > 0	Administrator > 0	Technologist > 0	Total
Biggest challenge for reviewers				
Insourse vs outsource confusion	1	2	1	4
Is repository in scope			1	1
granularity of evidence	1	1		2
Language Translation Issues	2	3	4	9
Takes lots of concentration to detail	2	2	2	6
Show these are required by the community and this explains rec	2		2	4
Too Verbose	1		1	2
Time Constraints		2	2	4
Reviewing External Materials		1	1	2
Consistency	1		1	2
Problems defining Preservation policies	1		1	2
Problem describing staffing	1		1	2
SUM	12	11	17	40
# N = Documents	9 (69.2%)	8 (61.5%)	7 (53.8%)	3 (100.0%)

Evidence provided in CTS applications must be in English, but most often the person or persons filling out the application are not native English speakers. Even with innovative tools used by translators (e.g., Google), the meaning of some words in many languages are not easily translated directly into English. Reviewer 19 summarized this issue well:

Another thing I think I struggle with is sometimes when applications, so this is the language problem, I guess when applications are written in poor language and call it poor English because we accept applications that only in English. ... I recognize not being a native English speaker, that this is not an easy thing to do. Um, people sometimes provide you with evidence that's not in English, so you have to run a Google translate, whatever to try and get the gist of what everyone is trying to communicate to you as a reviewer.

Reviewer 24 realized that:

if English isn't someone's first language then when reviewing those reviews ... probably the extra work I have to do to understand them is probably nothing compared to the work that they have to do to actually express what they wanted ... in a second language. Because I think nuance is everything.

So, this challenge is not only faced by the reviewer, but also the repository if their native language is not English. As Reviewer 26 noted, “Most of the applicants aren’t English born, so to say. ... it’s always difficult.” Applicants can utilize translators (both human and computerized), but technical specifications are often challenging to translate. This has the potential to be misleading at times.

Other challenges faced by reviewers range from the significant time it takes to perform these reviews at an appropriate granularity to the attention to detail required to review external references (documents such as mission statements) that are often not in English. For reviewers to understand, they need to pay close attention to details; this is critical for a peer review process to succeed over time as more diverse reviewers are added to the pool of potential reviewers.

Reviewer 16 understood this, stating:

For me personally, as a reviewer I do ... need to be very, very disciplined ..., um, I have to discipline myself so that I make sure that I really go through the application step-by-step and don't miss anything. So, I have to remind myself really to every time. So, every time I do a review, basically I go back, look at the, the requirements, I read, the text, I read the extended guidance ... so just to make sure I don't forget anything. So, it's, you know, for me it's something, I mean it's, it's not, it's not a challenge, but something that I have to, as a reviewer, make sure I do because I think that that would help. That helps the quality of the review.

This type of dedication to detail will help keep the CTS sustainable and aid in the consistency of reviews, but it also increases the time needed to complete each review. Indeed, it is easy to underestimate the impact of this on the volunteer workforce needed to perform the reviews.

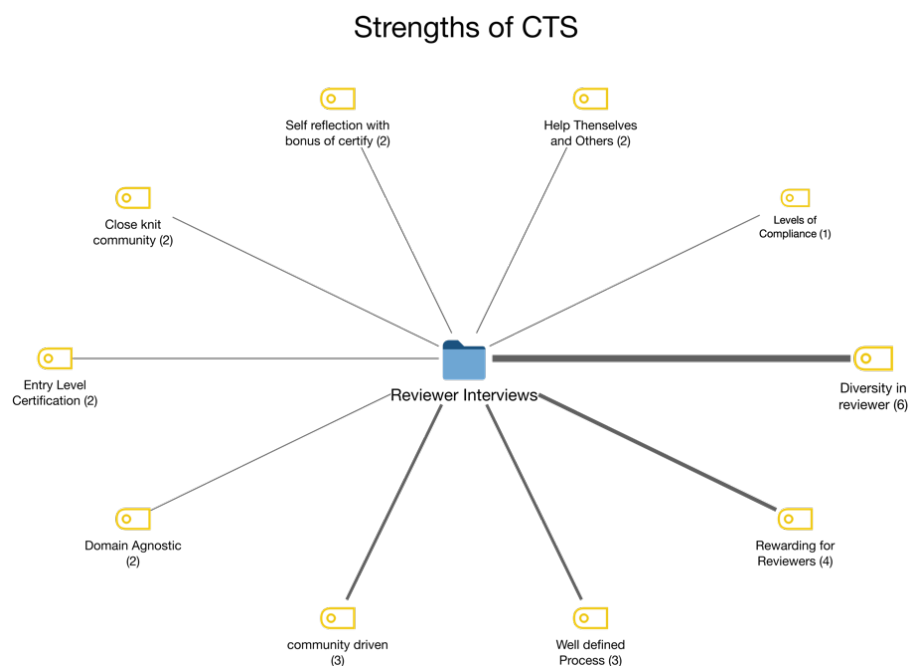
Reviewer 19 pointed out that it is “not stressful, but at the same time difficult to fit in because it's

voluntary work, sometimes quite challenging to fit that into your schedule.”

Strengths of CoreTrustSeal

In order to keep this volunteer workforce sustainable, the process needs to have some valuable strengths. The reviewers were asked to talk about the strengths that the CTS process embodied (Figure 18).

Figure 18. Strengths of CTS



The top two responses were very encouraging. Diversity in the reviewers was seen as a strength and suggests that reviewers do not need to be disciplinary specific. Participants valued not only the diversity in the reviewers, but also felt that the process was rewarding for reviewers. This is particularly important for a business model that relies on volunteer labor. Several reviewers felt like Reviewer 16, who said, “it’s definitely rewarding. ... it’s something that I enjoy doing.” The enjoyment of learning about other organizations and doing “in-depth dive into that particular application” (26) seemed to make the task worthwhile for several reviewers. The concept of job

satisfaction is very important to sustain this community effort. Reviewers seem to be willing to put in the effort if “in the end [they are] really, really happy because CoreTrustSeal, um, may make so much greater impact” (25).

An additional observation by the reviewers was that they felt an advantage was gained by having the diversity in the reviewer pool and the reviewers being disciplinary agnostic. One could say there is a tendency in the various scientific communities to create silos where only those in that community are trusted to provide professional advice. Several of the reviewers here pointed out that this may be an opportunity lost. They felt that CoreTrustSeal provides an outside viewpoint from a wide array of diverse reviewers that come from many other disciplines and perspectives. Reviewer 19 believed that this is a serious advantage:

But I really think that the fact that, (pause) it is disciplinary agnostic is serious, is really a strength because we are not really (pause) looking at those repositories from a disciplinary perspective. We're looking at them from a generic sense as we're looking at their missions from a generic sense and not from a disciplinary perspective. So, it gives us, I think more and more neutral (pause) vision and, or more [fresh], the judgment we can have is more neutral. ... with the reviewers [being] picked almost randomly and I think that's a good thing.

The CoreTrustSeal reviewers represent “a mixture of expert opinion and [disciplinary] training. These [are] a strength. It's appropriate to what we're trying to do, which is a mixture of evaluation, community building, reaching consensus on some of these points and overall improvement to our community” (18). The concept of an external entity review that can provide insights into new ways to preserve and share data for a repository’s designated community seems to be a “real strength and the key [to] CoreTrustSeal is [this] diversity of the reviewers” (28). I know from personal experience that reviewing repositories outside my trained discipline takes a little more effort because I am required to learn enough about the digital objects and the repository’s designated community to make judgments and often suggestions to better protect

these data. In such cases, the reviewer may spend more time and effort, thereby increasing the attention to detail and often providing greater incentive.

Several reviewers mentioned that the process is rewarding. The task of learning about a new community is often rewarding for many in our community. Another reviewer discussed the concept of self-reflection. In addition to being personally rewarding, reviewers can learn new techniques or approaches that can be leveraged at their home repository as well. In the end the entire community benefits from this diversity.

Biggest Problem Facing CoreTrustSeal

With any complex process, we know there can be challenges. The goal should be to identify them and work toward a solution. The CoreTrustSeal process is not immune to these typical challenges. The reviewers in this study were consistent across the three perspectives when expressing the largest concerns (Table 5). These top two issues were also closely related.

Table 5

Biggest Problems Facing CTS

	Archivist > 0	Administrator > 0	Technologist > 0	Total
Interview Codes				
Biggest Problem Facing CTS				
Dynamic of not requiring certification	1	1		2
Maintain Consistency in Reviews	8	8	6	22
Increase interest in CTS	2		2	4
Judging Science behind data		1	3	4
New distributed technologies a challenge to face	2		2	4
Problem with scalability as review requests increase	8	4	9	21
complex partnerships hard to define	1	1	1	3
cts vs wds	2	2		4
fees change expectations	1	2		3
glam decision not marginalized time sensitive	1	1	1	3
in Scope out of scope as an issue	1	1		2
problem with applicants understanding minimum compliance	2	2	1	5
too busy to stay in touch	1	1		2
SUM	30	24	25	79
# N = Documents	9 (69.2%)	8 (61.5%)	7 (53.8%)	3 (100.0%)

Reviewers were concerned with the rate of increase in new applications. They questioned whether the CoreTrustSeal processes would be able to scale appropriately and whether the community would find reviewers willing to volunteer. Every growing organization encounters

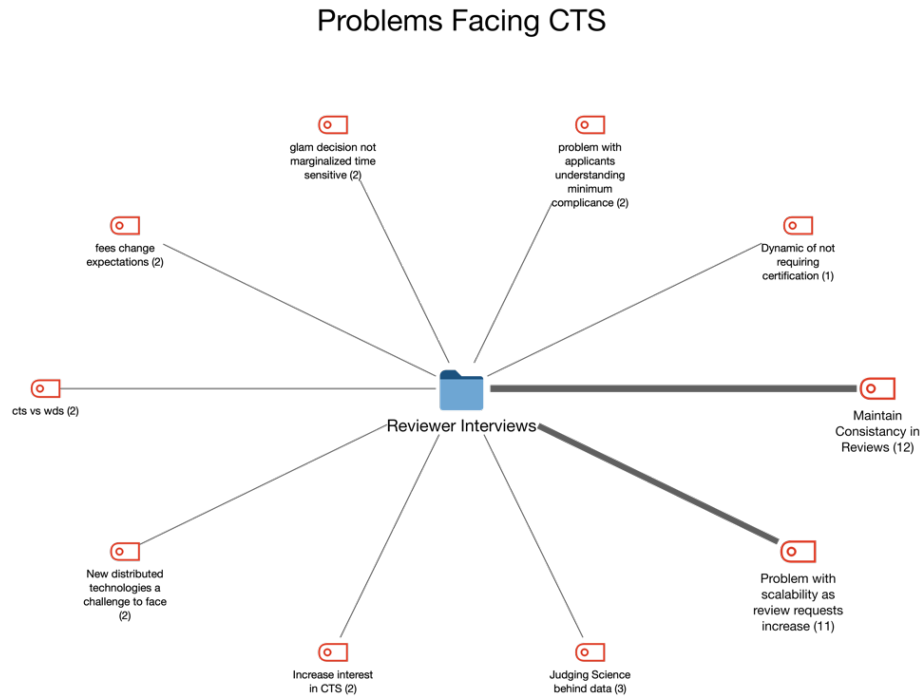
challenge scaling, but the goal is to do so without frustrating the community while maintaining quality. Reviewer 15 spoke about this challenge and pondered if the workflow that brings every application to the full CoreTrustSeal board for review will work:

I think that the turnaround time can be long. (pause) because of the way we're currently doing the reviews. So (pause) with two reviewers and then the board having to go over it (pause) that is sometimes taking them a while and especially if the reviews they not come in on time. Um, so I wonder if that's always necessary to have the full board review each and every application that might not be sustainable if we scaled up to certain point.

The CoreTrustSeal board's final review and approval are designed to help maintain consistency and follow up on any errors the reviewers make. As Reviewer 19 explained, "It's a voluntary work. Reviewers are doing this on their free time. So, this is, I think, one ... of the challenges. So, scaling up and making sure that we can [maintain consistency] as a certification body." The initial challenge will be finding enough reviewers to take on the increasing number of applications but as the number of reviewers increase, so does the challenge of consistency.

I think at some point if we have too many reviewers, I think it's the consistency [that] is going to become a problem. [inaudible] because I mean, what we're doing now is we have an experienced reviewer paired with someone maybe who doesn't do as many reviews. And I think that would, it might get difficult, at least for, for a certain period if we would [increase] the assembly of reviewers too much [too fast]. (16)

Figure 19. Problems facing CTS



With this increase in reviewers and application volume also comes the challenge of maintaining consistency in the reviews (Figure 19). The community seems to support CoreTrustSeal, as seen in the increase of applications. But as Reviewer 13 pointed out,

[The community thinks] the CoreTrustSeal is a really useful and wonderful thing to have. But unless we can think about our business model, unless we can really get, start getting towards the point where we could have someone more full time, someone who can really focus on this and insured, that professionalism, ... We have to continue to maintain the consistency. And this is, this is all big challenges.

Reviewers talked about the attention to detail required and the substantial time commitments necessary to perform consistently good reviews. But as Reviewer 15 noted, “you try to be as consistent as possible, but every assessment is so different that it's often quite hard to achieve.” The word *coherence* is defined as having the quality of being both logical and consistent. When asked about the challenges facing CoreTrustSeal moving forward, this was the word that Reviewer 17 thought about first:

I have one word, coherence. [It is] a coherence a when it comes to reviews, for example, was one of the most obvious things, but also a coherence in supplied evidences that support the self-assessments that are two things that pop into my mind.

Reviewer 18 explained that the CoreTrustSeal process is designed to be a transparent process by necessity. This means that all approved applications are placed on a public website for other repositories to review and critique.

... scaling to demand while maintaining consistency. (pause) we think we are pretty consistent but with, with our transparent outcome, which I think is incredibly important, you are open to comparison and criticism on almost every point of where you are not consistent.

These challenges are well known, and all but three reviewers talked about them being the biggest challenge for CoreTrustSeal. Reviewer 13 recognized this as well as the fact that the reviewers change each time a repository is recertified every three years. This enhances the need for new reviewers to understand what previous reviewers requested to maintain consistency. As Reviewer 13 stated, “we’re starting to build training materials. That's great. But we have to obviously, as we build, we have to maintain the level of the review quality. We have the commitment [to] maintain the consistency.” This reviewer also pointed out the difficulty with recertifications every three years as the second set of reviewers will almost never be the same due to the random selection process.

You only know what you’re looking at and, and how you're looking at now [and not the previous application]. We need to be able to kind of manage that to a point. I think you can't give people experience. They don't have, but I think we need to give the reviewers sometimes a little bit better guidance. (13)

It is encouraging that the reviewers not only recognized a potential challenge, but also offered solutions.

The reviewers all seemed to be very dedicated to the success of CoreTrustSeal as well as the preservation of digital objects for reuse by future generations of researchers. When asked

about potential changes to the CoreTrustSeal process, many great ideas were suggested and were often similar across the defined perspectives.

Suggested Changes to CoreTrustSeal Process

The reviewers made many suggestions that have the potential to improve the processes and workflows within CoreTrustSeal, although they discussed some areas more. More than half the participants discussed the need for better training of new reviewers (Table 6). The CoreTrustSeal Extended Guidance was developed to help reviewers and applicants know what to expect in the application and to learn more about the requirements, but reviewers do not receive formal training before performing reviews. New reviewers rely on their own experience when submitting an application for their repository. These were successful applications, suggesting the reviewer had some experience that would be beneficial, but often many members of an organization take part in filling out a successful CoreTrustSeal application. Reviewer 19 mentioned that there “needs to be maybe more professionalized in the sense of [they] would love to have formal training provided [and] to have to see formal training sessions offered to reviewers.”

Table 6

Suggested Changes to the CTS Process

Interview Codes	Archivist > 0	Administrator > 0	Technologist > 0	Total
▼ Potential Suggested Changes to the CTS Process				
Balance Reviewers Strengths	1		1	2
Modify Requirements			1	1
Separate Review Meeting from Board Meeting	5	4	4	13
Faster Turn Around Times	1	1	1	3
Best Practice or Exemplars	2	2		4
Pre-Registration	1	1		2
Better communications	2	2	1	5
Improve community support	2	2		4
Improve Validation of General Repositories and Technology	4	2	3	9
Administrative Technology Support	2	2	1	5
Organize Requirements on TRUST			1	1
No Need for Specialized Reviewers	3	3	3	9
Keep Volunteers			1	1
More Staffing	1	1	1	3
Need for Training	4	3	5	12
Σ SUM	28	23	23	74
# N = Documents	9 (69.2%)	8 (61.5%)	7 (53.8%)	3 (100.0%)

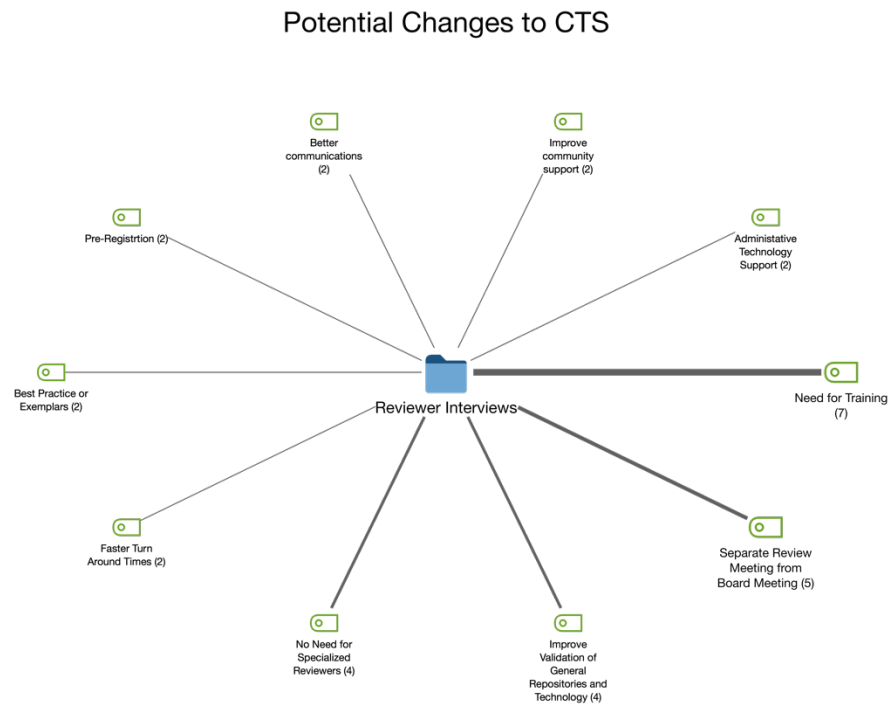
Although these new training materials would be great for increasing consistency as CoreTrustSeal expands and adds new reviewers, Reviewer 22 noted that:

it's important to have ongoing training of the reviewers and (pause) even though we're all doing reviews and sharpening our skills as we do those, I think if we could identify some areas where the reviews are perhaps [weaker] (pause) where they could be improved.

Reviewer 25 noted that, as CoreTrustSeal is a certification body, “we don't provide any [access] to workshops, trainings, (pause) programs, whatever.” This is clearly a need and on the minds of both reviewers and board members.

Another suggestion that was discussed by several participants was the need for a dedicated application review meeting by the board (Figure 20). CoreTrustSeal board members currently meet twice a month via video conferences. This meeting is planned as a comprehensive meeting with goals to accomplish all business in front of the board each time. This includes administrative duties and assigning reviewers when needed as well as the approval process for applications completed by the two anonymous reviewers. The board as a group must vote on each award of the CoreTrustSeal.

Figure 20. Potential changes to CTS



Often the discussion of the submitted reviews dominates the schedule, leaving little time to spend on administrative duties and decisions. Little if any time is available to discuss improvements or evaluations of the process. The challenge is resources and time for the volunteer board members. As Reviewer 18 noted, “if [they] had infinite resource and engagement, right. [they] would separate out the review meetings from the governance meetings.” This would allow more time for each diverse task if those time slots were available.

Reviewer 23 stated that:

maybe we could have a subgroup in the board or something that's because in the board, we have sort of this administrative stuff and marketing, keeping up the standards ... and then we have the reviews and those are quite separate things. So, we might think about having a separate body for their reviews, a separate meeting focusing only on the CoreTrustSeal acceptance.

This suggestion has lots of merit and should be investigated by the board. Any additions or changes to meeting commitments should also be balanced with the understanding of the

voluntary commitments by the board. Several reviewers mentioned that the time commitments were already substantial.

Other suggestions by the reviewers include publishing exemplar sections from approved applications for new applicants to leverage. This would help them better understand the requirements. These could be used as training materials for new reviewers as well. Additional paid staff was also encouraged to help with better communications and to decrease the time needed to complete reviews. Being primarily a volunteer organization, it is sometimes hard to predict when some members will have the available bandwidth to push various deliverables through the workflow. The current staff is already reaching its limit. Another suggestion from several of the reviewers was the creation of a plan to handle the increase in applications from generalist repositories and technological service providers.

Most early CoreTrustSeal applications were from disciplinary-specific repositories or special collections. These groups had a well-defined and narrowly scoped designated community. They also typically had a well-defined collection that the application was targeted toward. These characteristics allowed the repository to staff well-trained curators and preservationists who were knowledgeable about the needs of their designated communities. These staff also had a clear understanding of the required digital object formats and what it took to preserve them. Increasingly, applications are arriving that are more generalist in nature, often spanning many disciplines or even organizations. This makes it critical for reviewers to understand the level of curation to which the repository is committed, and it is often more difficult to define the targeted designated community. In addition, the CoreTrustSeal board has received some applications from technological service providers that only provide the hardware and software required for the preservation of the digital objects. These groups often rely on the

self-deposit method and require the depositor to be responsible for curation.

Many repositories are increasingly expanding collections to include a more diverse data collection or even becoming institutional repositories with the mandate to store all data for an institution. One of the suggestions reviewers offered in this study was the need for training materials and a plan to help reviewers understand what types of organizations are in scope for the certification. This would also help new applicants understand the requirements for being certified if they have a diverse collection and designated community. For repositories using one of the technical service provider's tools, such a planning document would help define which areas of the CoreTrustSeal application are potentially satisfied by the use of these tools. This could potentially lead to the certification of tools and services at a granular level and would potentially allow them to satisfy a subset of the CoreTrustSeal requirements. Work is ongoing to collect feedback from the community, and a draft planning document has been circulated for comment.

Research Results Summary

The selection of the sample reviewers from the population represented a balance in perspectives across the three anticipated categories of administrators, archivists, and technologists. In many cases, reviewers had difficulty selecting one category and often stated they served multiple roles. Themes in the study converged around several key concepts. Nearly all participants felt they were performing a peer review process and working to help the repository community and the research enterprise. Reviewers were asked about the various CoreTrustSeal application requirements and which ones, from their perspective, were the most important. No clear evidence emerged to indicate that variations in perspectives had an impact on the subjective review of application evidence. Often the same categories of evidence were selected and identified as critical across all three categories (i.e., administrator, archivist, and

technologist).

The results were similar when reviewers were asked to describe the most important evidence and suggestions to new CoreTrustSeal applicants. They expressed concerns pertaining to sustainability of the process as the number of applications increase. In addition, concerns about consistency of reviews were expressed as a critical issue to address going forward.

Although in this study perspective did not seem to have an impact on the subjective review of evidence, several participants noted that the potential exists as the number of reviewers increase.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

Importance of Reviewer Experiences

The participants in this study were among the first reviewers to evaluate repositories against CoreTrustSeal after the merger of the Data Seal of Approval and the World Data Systems certifications. Many were also elected as CoreTrustSeal board members. When reviewing each potential participant, a wide range of academic and educational backgrounds was evident. During the interviews, participants typically had a wide range of experiences across the three identified perspectives. Many reviewers were also senior members of their organizations and often held several positions before arriving at their current positions. These experiences across the three perspectives probably affected the outcomes and the priority each reviewer places on various pieces of CoreTrustSeal evidence. It was still a bit surprising that reviewers who self-identified as technologists did not often mention the importance of technology in the presentation of evidence.

An overview of the interview discussions as a whole showed that this group of reviewers clearly cared deeply about the research process and the value quality data contribute to scientific discovery. Every reviewer was keenly interested in learning about the needs of each community reviewed, supporting data preservation and sharing. This was shown in the holistic view each reviewer displayed when talking about the various components of trusted repository evidence presented for evaluation. The expectations and knowledge outside each of their academic backgrounds indicated that they had worked hard to learn the diverse and complex requirements that repositories must follow to ensure that the objects entrusted to them are safe. All these

reviewers also had experience with previous review standards for trustworthy digital repositories and often managed repositories of their own. Such experiences enabled reviewers to internalize the values of the CoreTrustSeal standard and, regardless of their personal backgrounds, evaluate applications evenly and fairly. This community is currently benefitting from the extensive experience of these reviewers. Many of these were the founders of repository review organizations and are keenly aware of the holistic perspectives needed to perform a good review.

The challenge moving forward will be to retain these experienced reviewers as well as grow this pool of qualified reviewers. There are many digital curation and archive professionals around the world to draw from, but the voluntary nature of this work, combined with the reduction of resources within repositories, presents a challenge. Repositories are faced with an increasing demand for digital preservation services; thus, many staff members do not have the available time to volunteer for these activities.

Challenges of Detailed Evidence versus Concise Evidence

Several of the reviewers interviewed were interested in concise evidence provided by the applicants, which can sometimes be a challenge because the complex environment in which repositories operate have many individual components. In addition, they often have many staff members contributing to the workflows during the preservation of digital objects. The CoreTrustSeal requirements were based on the detailed OAIS model for describing and managing repository workflows, yet CoreTrustSeal was purposely designed to be a lightweight instrument. Thus, a larger population of repositories can attempt certification due to the reduction in staff labor required to apply. The challenge is that the responses used to describe repository functions in CoreTrustSeal must still provide enough detail for the reviewer to make an informed decision. The reviewers must ensure that the complex processes of the repository

are trustworthy. This increases the challenge for the applicant to be concise yet provide enough information to satisfy the reviewer. It also increases the subjective nature of the reviewers' decisions as they lack full information by the nature of this exercise.

Many of the CoreTrustSeal requirements are expressed in broad terms. One example shown here is requirement nine, aimed at the repository describing their storage procedures: “The repository applies documented processes and procedures in managing archival storage of data” (CoreTrustSeal, 2017b). This requirement certainly covers a broad topic and is designed to fit many repositories' responses. The instructions for the applicant do give some guidance on the details required, but these are also very broad categories.

- How are relevant processes and procedures documented and managed?
- Does the repository have a clear understanding of all storage locations and how they are managed?
- Does the repository have a strategy for multiple copies? If so, what is it?
- Are risk management techniques used to inform the strategy?
- What checks are in place to ensure consistency across archival copies?
- How is deterioration of storage media handled and monitored?

This is in contrast to the more detailed requirements of the ISO 16363 standard that breaks down these requirements much more granularly. The following examples highlight the details requested for ISO 16363 in the area of managing AIPs.

4.4.1 The repository shall have specifications for how the AIPs are stored down to the bit level.

4.4.1.2 The repository shall actively monitor the integrity of AIPs.

4.4.2 The repository shall have contemporaneous records of actions and administration processes that are relevant to storage and preservation of the AIPs.

4.4.2.1 The repository shall have procedures for all actions taken on AIPs.

4.4.2.2 The repository shall be able to demonstrate that any actions taken on AIPs were compliant with the specification of those actions.

These are just a few of the requirements that demonstrate the details and specifics that an ISO 16363 audit would request pertaining to the preservation of the digital objects, thereby demonstrating the complexity of a preservation system and the details required to evaluate it. When repositories apply for CoreTrustSeal, the process is designed to be less burdensome and a more lightweight review, but the fact remains that the repository must appropriately describe these same complex functions, which can make it difficult for the repository to remain concise with its evidence while providing necessary information.

The challenge here is determining how concise an applicant should be while still providing enough detail to be useful. The simplicity of the CoreTrustSeal requirements is designed to make it easier for the repository to be certified compared to ISO 16363, but this can also make the process much harder for the reviewer. The answer to this challenge could be found in the suggestion of several reviewers. During the interviews, the reviewers stated that they wished to see links to public policy documents embedded in the summary provided as evidence. A good strategy for applicants would be to provide a concise overview in the application and reference appropriate detailed public documents describing the more complex components of the repository functions. This would also be a great starting point if, in the future, the repository attempts a more detailed ISO 16363 audit or wishes to provide more details to its designated community.

CoreTrustSeal Process Suggestions

During the interviews, the participants made several suggestions that had the potential to improve the CoreTrustSeal process and workflows. The reviewers have many combined years of

diverse experience, and these suggestions are important to consider for the sustainability of the CoreTrustSeal community. The suggestions range from the need for more administrative resources to the materials provided for new applicants as an aid to improve future applications. All of these have the potential to reduce turnaround times for reviews, which is a desired outcome.

Reviewers expressed that administrative resources should be improved. Some of these suggestions were for more administrative staff to assist the management of the applications and better communicate with the repositories. Administrative technical support was also suggested to aid with the management of the website as well as the review submission systems. All of these have the potential for reducing the amount of time it takes to approve an application. The CoreTrustSeal is primarily a volunteer organization, but an increase in paid staff or contracted services would help make the processes a little smoother and faster.

Another area that was seen as important was to provide some exemplar materials to aid new applicants. The number of approved applications has now reached a level that would allow for the selection of exemplars for each category that could help guide new applicants. It may also be possible to have suggestions for various disciplines and repository types. As the diversity of new repositories increases, they need to understand if they are in scope for CoreTrustSeal. There is current work to develop a plan to handle the increase in applications from generalist repositories and technical service providers—another area of concern discussed by reviewers. Actions are already being taken.

When discussing the application approval workflow, several reviewers expressed an interest in holding a meeting just for the review and approval of applications. Such an important task seems to always dominate the biweekly board meetings. In some cases, these reviews take

most of the allocated time and do not allow for time to discuss administrative business and improvements to the organization. These tasks are vital and would help increase productivity. This would also potentially allow a different subset of board members to focus on reviews, allowing some of the others to focus their volunteered time toward administrative duties. Collectively, these reviewer suggestions have the potential to increase the efficiency of the CoreTrustSeal process and provide more information to new applicants. But the single most discussed improvement was the need for reviewer training. Reviewers saw this training as the highest priority to improve the certification process and as critical to keeping the reviews consistent as the number of applications for CoreTrustSeal increase.

Building Training Programs for Reviewers

The most important need that was identified by the study participants was a training program for reviewers. Currently, the only requirement for being a reviewer is to be employed at an organization that has a trusted repository certified by CoreTrustSeal. In addition, the only instructions provided to those new reviewers are the publicly available CoreTrustSeal requirements and the accompanying extended guidance (CoreTrustSeal Standards and Certification Board, 2019). The extended guidance is helpful and was one of the suggestions for new applicants to read before submitting a certification application, but it is not a training guide. In order to maintain consistency across reviews, the study participants often suggested the need for more training.

Consistency in evaluation is often a challenge, especially when the materials are potentially subjective in nature. One method that can be effective is the use of a standardized rubric for scoring of materials. The current extended guidance contains some initial concepts of this. For example, the extended guidance for R1 states that, “If data management is not referred

to in the mission of the repository, then this Requirement cannot have a Compliance Level of 3 or higher” (CoreTrustSeal Standards and Certification Board, 2019). This is a good start, but many places in the extended guide do not have this, and the level of detail could be enhanced to give reviewers more metrics to be looking for in the provided evidence.

The educational community has experience using rubrics in developing coursework and instructional design. The technique of leveraging rubrics has been used effectively in all levels of education. “Whether we teach elementary school or graduate students, rubrics orient us toward our goals as teachers” (Andrade, 2005). The challenge is to develop a rubric that is flexible enough yet provides the consistency needed for the CoreTrustSeal reviews. In addition to developing rubrics, CoreTrustSeal must also develop clearly defined training regimes using these rubrics. As Andrade stated, “rubrics are not replacements for good instruction...rubrics can serve the purposes of teaching and learning as well as evaluation” (2005). New reviewers should leverage this new rubric for evaluation of CoreTrustSeal applications. This will also aid in consistency across reviews by guiding new applicants and train them to provide appropriate evidence as well.

Developing a rubric associated with the levels of compliance in the CoreTrustSeal framework would be an improvement to the CoreTrustSeal Extended Guidance (CoreTrustSeal Standards and Certification Board, 2019) and could provide the basis for the design of reviewer training materials. These materials could be delivered online, which would be preferred due to reviewers’ diverse locations. To enhance transparency of the CoreTrustSeal peer review process, these courses could be available to the public so new applicants could also leverage them. This would allow repositories to learn what reviewers will be looking for. These online courses could include an introduction that discusses implicit biases and would help reviewers understand them

to help identify them and mitigate their impact. This is especially important for inexperienced reviewers. Once implemented, these training courses will not only increase the consistency of the reviews, but also have the potential of increasing the quality of evidence provided in the first drafts of applications. The potential for reducing the number of resubmissions is important because this can reduce the turnaround time for certification and increase the repositories' satisfaction with the process. This also has the potential for increasing the sustainability of the community, which is vital for this business model. The ultimate goal should be for the repository or archive to improve preservation practices as a result of the review. If self-assessments and peer reviews are done in good faith, they could have a powerful effect on the longevity of the archive and its ability to preserve its digital assets. Certifications and seals of approval all play an important role in the process of an archives becoming more OAIS compliant.

Study Impact and Suggestions for Repositories

The review of the literature painted a complex picture that is still evolving in the world of audits and certification of trusted digital repositories. The first step for archives that wish to proceed down this road is to educate themselves on what “trust” in their specific communities means as research has shown that “trust” is relative to the domain or discipline and varies with the repository's status of dependability in the community (Yakel et al., 2013). Recent discussions around the conceptual areas that repositories need to examine in order to be trusted focus on the letters that combine to spell trust (Lin et al., 2020). Modeled after the FAIR principles concept (Wilkinson et al., 2016), the TRUST principles (Table 7) outline what actions repositories must demonstrate to show their designated community that they are trustworthy.

Table 7

The TRUST Principles

Principle	Guidance for repositories
Transparency	To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
Responsibility	To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
User Focus	To ensure that the data management norms and expectations of target user communities are met.
Sustainability	To sustain services and preserve data holdings for the long-term.
Technology	To provide infrastructure and capabilities to support secure, persistent, and reliable services.

These are the principles and concepts that repositories should seek to describe when applying for any seal or certification to demonstrate that they are a trusted repository. Many of these attributes were identified as important by reviewers in this study. If repositories wish to demonstrate trust to their community, following the well-seasoned advice of reviewers in this study would contribute to this goal. These concepts should be followed while providing evidence for the CoreTrustSeal application.

The first principle of transparency was mentioned several times during the study interview, and the need for evidence that points to publicly available policy documents was one of the top suggestions by reviewers. In order to make the application concise, it is also important that links to these public documents be included in the summaries provided within the CoreTrustSeal application. The evidence provided should be supported by public policies that are consistent with the claims in the summary.

One of the most important categories of evidence discussed by the reviewers was the context and background information for the repository under review. Reviewers often felt that this was critical to the review process. Repositories should describe their designated community and provide information about the needs of that community to set the framework for curation

requirements in that community. A brief description of the repository is important, but applicants should ensure that the collection or collections targeted for certification are well described and designated in this summary. Often repositories have many collections that vary widely, and some of those collections may not be the target of the certification. This can confuse the reviewer and could inadvertently expand the perceived requirements for curation.

Applicants should always approach each requirement with the same level of specifics and detail. Yet reviewers often mentioned that curation workflows and the preservation of the digital objects were some of the first things on their minds. Documented curation workflows that demonstrate to the reviewer that the repository has a defined plan for each object under its care is important. Once these AIPs are created, reviewers felt that it is very important to have a preservation plan to describe how they will be cared for and managed over time.

Some other important advice for new applicants was geared not toward the evidence requirements, but rather to help applicants with the process of applying. The first suggestion was to plan. The application often requires input from many different individuals at the repository's organization. Administrative buy-in is critical because the process will require time and effort across many areas of the organization. Repositories seeking the CoreTrustSeal certification should have the repository staff review approved applications that are published on the CoreTrustSeal website and pay attention to certified repositories that have characteristics and designated communities similar to their own. It can be helpful to assign various staff members to the task of outlining a bulleted draft for the portions of the application they will be responsible for. The reviewers also suggested holding a meeting to compare notes and identify areas that need more documentation or evidence. Collectively, as a group, it is easier to identify who might have the information needed to create this required documentation. It can also be helpful to

follow the terminology and guidance provided in the OAIS model (Giaretta, 2012).

A successful CoreTrustSeal application will require a team effort from the repository staff as well as administrative support. This ongoing process repeats every three years. A trusted repository should always reevaluate and review its policies on a regular basis to ensure it is providing adequate services to its designated community and preserving the digital objects it is entrusted to manage. This is even more critical for research data repositories, where the required formats and research methodologies are rapidly changing to keep up with scientific innovation and the numerous new data sources. The process of repository evaluation is time consuming and complex, but it provides a framework for this recurring self-evaluation that is critical to remaining a trusted digital repository.

Limitations of the Study

As with any study, there are limitations to the research and lessons learned during the process that would have benefited from changes. Now that more than 90 repositories have the CoreTrustSeal, the number of potential reviewers to interview has increased. When the sample for this study was taken, only 37 reviewers had participated in a CoreTrustSeal review. The 13 participants selected were experienced members of the repository community who often held many positions during their tenures. This excess of experience in the sample was very valuable in gaining insights into the processes, but also limited the perspectives of new and less experienced reviewers. As CoreTrustSeal expands, it is probable that more reviewers will have far less experience and they will not have the well-balanced perspectives obtained in this study. Many of the reviewers in this study self-identified as having more than one of the targeted perspectives. In addition, the study did not obtain a metric for participants' experience level, which would have helped weight the results based on experience levels in addition to the

background and training. Any future studies should create a metric for experience to include in the analysis. This community would benefit from a repeated study in five years after the community has expanded to ensure that the newer generation of reviewers remains as unbiased and well-rounded as the initial group of reviewers tended to be.

Another challenge with the study is that it relied on the reviewers' memory to describe their experiences. This memory was leveraged to obtain the requirements that the reviewers were first to remember. However, the data could also be skewed by reviewers recalling their most recent experiences. The CoreTrustSeal requirements are relatively new, which limits the number of applications reviewers had to recall, but their answers could have been skewed by more recent events.

One methodology tested during this study was the use of the Pencast system that records audio and synchronizes it with the interviewer's note writing. Personally, I felt that the struggle to use this consistently across the interviews was too much for it to be useful. I often found myself concentrating on the interview and engaging with the participant and forgetting to evenly record notations using the Pencast. The resulting information was inconsistent and not used in this study. Although it did provide a third copy of the recordings, the synchronization of the notes was not helpful.

Future Studies

Now that a greater number of CoreTrustSeal applications have been approved, future studies could be done to assist the community. The most important study would be to design a rubric for use in training reviewers and approving CoreTrustSeal applications. Given the larger number of approved applications, a content analysis of these approved evidence lists could provide the details needed to design a rubric that allows for the consistent and sustainable

approval of CoreTrustSeal applications. As the number of applications increases, this will aid in the scaling of the process. The results would be a great addition to the current CoreTrustSeal Extended Guidance and could be used to build training courses for reviewers.

A follow-up study that compares new reviewers' perspectives to the perspectives of more senior reviewers would be interesting. This was not possible during the current study, but the growth of the CoreTrustSeal peer review community could now support such a study. Although no evidence of influence due to various backgrounds and education were discovered in this study, this could have been impacted by the lack of newer reviewers with less experience. Including these new reviewers in a balanced sample could help identify and maintain strategies to prevent inconsistent evaluations and provide greater balance of reviews across various repositories.

Another interesting follow-up study would be to compare repository evaluation to the scientific peer review process. Some research into these processes has been performed (Wessely, 1998). What could be learned from these evaluations and how would they apply to the process of building trust in research repositories?

APPENDIX A: REPOSITORY APPLICATIONS TARGETED

Repository	Link	Country	Pages	Discipline
WDC—Solid Earth Physics, Moscow	http://www.wdcb.ru/sep/index.html	Russia	26	Physics
The Language Archive	https://archive.mpi.nl/	Netherlands	28	Language
DKRZ—WDC Climate	http://www.wdc-climate.de/	Germany	26	Climate
CLARIN Center IvdNT	http://www.ivdnt.org/	Netherlands	34	Language
IMS Repository	http://clarin04.ims.uni-stuttgart.de/fedora/	Germany	49	Language
Qualitative Data Repository	https://qdr.syr.edu/	USA	29	Social Science
USGS EROS Center (WDC—Earth Resources Observation and Science)	https://eros.usgs.gov/	USA	29	Geosciences
Australian Antarctic Data Centre	http://data.aad.gov.au/	Australia	150	Climate
DataFirst Data Portal	http://www.datafirst.uct.ac.za/dataportal/	South Africa	23	General
CELR META-SHARE	https://metashare.ut.ee/	N/A	42	Language
CLARIN Center BBAW	http://clarin.bbaw.de/en/	Germany	39	Language
Chinese Astronomical Data Center	http://casdc.china-vo.org/?locale=en	China	20	Astronomy
CSIRO Data Access Portal	https://data.csiro.au/	Australia	35	General

Norwegian Marine Data Centre (NMD)	https://www.hi.no/en/hi/forskning/research-groups-1/the-norwegian-marine-data-centre-nmd	Norway	27	Marine
Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC)	http://daac.ornl.gov/	USA	25	Biological Science
UC3 Merritt	https://merritt.cdlib.org/	USA	33	General
ISRIC WDC - Soils	https://www.isric.org/about/world-data-centre-soils-wdc-soils	Netherlands	43	Soils
Tilburg University Dataverse	https://www.tilburguniversity.edu/about/university-library/	Netherlands	34	General
The Language Bank of Finland	http://www.kielipankki.fi/	Finland	61	Language
LDC Catalog	https://catalog ldc.upenn.edu/	USA	26	Language
The ILC4CLARIN Centre at the Institute for Computational Linguistics	https://dspace-clarin-it.ilc.cnr.it/	Italy	29	Language
Cornell Institute for Social and Economic Research (CISER)	http://ciser.cornell.edu/	USA	51	Social Science
Digital Repository of Ireland	http://www.dri.ie/	Ireland	59	Social Science & Humanities
Scholars' Mine	http://scholarsmine.mst.edu/	USA	77	General

DANS: Electronic Archiving SYstem (EASY)	https://easy.dans.knaw.nl/	Netherlands	45	General
FDAT	http://fdat.escience.uni-tuebingen.de/portal	Germany	49	General
NSD's Research Data Archive	http://www.nsd.uib.no/nsddata/arkivering/en/index.html	Norway	48	General
DARIS	http://forscenter.ch/en/data-and-research-information-services/	Sweden	47	Social Science
ARCHE	https://arche.acdh.oeaw.ac.at/browser/	Austria	37	Humanities
Irish Social Science Data Archive (ISSDA)	http://www.ucd.ie/issda/	Ireland	53	Social Science
Meertens Institute	http://www.meertens.knaw.nl/cms/	Netherlands	51	Language & Culture
Roper Center for Public Opinion Research	http://ropercenter.uconn.edu/	USA	43	Social Science
Swedish National Data Service	http://snd.gu.se/	Sweden	37	General
Australian Data Archive	https://www.ada.edu.au/ada/home	Australia	69	Social Science
Czech Social Science Data Archive	http://archiv.soc.cas.cz/	Czechia	39	Social Science
ADP Social Science Data Archives	http://www.adp.fdv.uni-lj.si/eng/	Slovenia	47	Social Science
National Geoscience Data Centre (NGDC)	http://www.bgs.ac.uk/services/ngdc/home.html	UK	38	Geosciences
The Worldwide PDB (wwPDB)	https://www.wwpdb.org/	N/A	90	Biological Science

ImmPort Repository	http://www.immport.org/	USA	21	Biological Science
GESIS Data Archive for the Social Sciences	https://www.gesis.org/en/home/institute/departments/data-archive-for-the-social-sciences/	Germany	46	Social Science
TalkBank	http://talkbank.org/	USA	33	Language
Mendeley Data	http://data.mendeley.com/	USA	46	General
Data Repository University of Minnesota (DRUM)	https://conservancy.umn.edu/handle/11299/166578	USA	37	General
The Finnish Social Science Data Archive (FSD)	http://www.fsd.uta.fi/en/	Finland	42	Social Science
Edition Topoi Collections	http://repository.edition-topoi.org/	Germany	37	Ancient Studies
Totals:		18 countries	1950 pages	12 disciplines

APPENDIX B: CORETRUSTSEAL REQUIREMENTS

Core Trustworthy Data Repositories Requirements Background & General Guidance

The Core Trustworthy Data Repositories Requirements are intended to reflect the characteristics of trustworthy repositories. As such, all Requirements are mandatory and are equally weighted, standalone items. Although some overlap is unavoidable, duplication of evidence sought among Requirements has been kept to a minimum where possible. The choices contained in checklists (e.g., repository type and curation level) are not considered to be comprehensive, and additional space is provided in all cases for the applicant to add “other” (missing) options. This and any comments given may then be used to refine such lists in the future.

Each Requirement in the Catalogue is accompanied by guidance text to assist applicants in providing sufficient evidence that their repositories meet the Requirement, outlining the types of information that a reviewer will expect in order to perform an objective assessment.

Furthermore, the applicant must indicate a compliance level for each of the Requirements:

- 0 – Not applicable
- 1 – The repository has not considered this yet
- 2 – The repository has a theoretical concept
- 3 – The repository is in the implementation phase
- 4 – The guideline has been fully implemented in the repository

Compliance levels provide a useful part of the self-assessment process, but all applicants will be judged against statements supported by appropriate evidence, not against self-assessed compliance levels. In this regard, if the applicant believes a Requirement is not applicable, the reason for this must be documented in detail. Note also that compliance levels 1 and 2 can be valid for internal self-assessments, while certification may be granted if some guidelines are considered to be at level 3—in the implementation phase—since the Requirements include an assumption of a repository’s continuous improvement.

Responses must be in English. Although attempts will be made to match reviewers to applicants in terms of language and discipline, this is not always possible. If evidence is in another language, an English summary must be provided in the self-assessment.

Because core certification does not involve a site visit, the Requirements should be supported by links to public evidence. Nevertheless, it is understood that for reasons such as security, it may not always be possible to include all information on an organization’s website, and provisions are made within the certification process for repositories who want sensitive parts of their evidence to remain confidential.

Repositories are required to be reassessed every three years. It is recognized that while basic systems and capabilities evolve continuously according to technology and user needs, they may not undergo major changes in this timeframe. However, the Trustworthy Repository ISO standard (ISO 16363) has a five-year review cycle, and a shorter period is considered necessary for a core trust standard to allow for possible modifications and corrections. Hence, an organization with well-managed records and business processes should reasonably expect to be able to submit an application with only minimal revisions after three years, unless the

Requirements themselves have been updated within the intervening period.

Glossary of Terms

Please refer to the Core Trustworthy Data Repositories Requirements Glossary:
<https://goo.gl/rQK5RN>.

Requirements Background Information Context

R0. Please provide context for your repository.

– *Repository Type*. Select all relevant types from:

- Domain or subject-based repository
- Institutional repository
- National repository system, including governmental
- Publication repository
- Library/Museum/Archives
- Research project repository
- Other (Please describe)

Comments

– *Brief Description of the Repository's Designated Community*

– *Level of Curation Performed*. Select all relevant types from:

- A. Content distributed as deposited
- B. Basic curation – e.g., brief checking, addition of basic metadata or documentation
- C. Enhanced curation – e.g., conversion to new formats, enhancement of documentation
- D. Data-level curation – as in C above, but with additional editing of deposited data for accuracy

Comments

– *Outsource Partners*. If applicable, please list them.

– *Other Relevant Information*

Organizational Infrastructure

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).

Compliance Level:

Digital Object Management

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

Guidance:

The repository should provide evidence to show that it operates a data and metadata management system suitable for ensuring integrity and authenticity during the processes of ingest, archival storage, and data access.

Integrity ensures that changes to data and metadata are documented and can be traced to the rationale and originator of the change.

Authenticity covers the degree of reliability of the original deposited data and its provenance, including the relationship between the original data and that disseminated, and whether or not existing relationships between datasets and/or metadata are maintained.

For this Requirement, responses on data integrity should include evidence related to the following:

- Description of checks to verify that a digital object has not been altered or corrupted (i.e., fixity checks).
- Documentation of the completeness of the data and metadata.
- Details of how all changes to the data and metadata are logged.
- Description of version control strategy.
- Usage of appropriate international standards and conventions (which should be specified).

Evidence of authenticity management should relate to the follow questions:

- Does the repository have a strategy for data changes? Are data producers made aware of this strategy?
- Does the repository maintain provenance data and related audit trails?
- Does the repository maintain links to metadata and to other datasets? If so, how?
- Does the repository compare the essential properties of different versions of the same file? How?
- Does the repository check the identities of depositors?

This Requirement covers the entire data lifecycle within the repository, and thus has relationships with workflow steps included in other requirements—for example, R8 (Appraisal) for ingest, R9 (Documented storage procedures) and R10 (Preservation plan) for archival storage, and R12–R14 (Workflows, Data discovery and identification, and Data reuse) for dissemination. However, maintaining data integrity and authenticity can also be considered a mindset, and the responsibility of everyone within the repository.

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

XI. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

Guidance:

Repositories must work in concert with depositors to ensure that there is enough available information about the data such that the Designated Community can assess the substantive quality of the data. Such quality assessment becomes increasingly relevant when the Designated Community is multidisciplinary, where researchers may not have the personal experience to make an evaluation of quality from the data alone. Repositories must also be able to evaluate the technical quality of data deposits in terms of the completeness and quality of the materials provided, and the quality of the metadata.

Data, or associated metadata, may have quality issues relevant to their research value, but this does not preclude their use in science if a user can make a well-informed decision on their suitability through provided documentation.

For this Requirement, please describe:

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

Technology

XV. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

Applicant Feedback

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

APPENDIX C: RESEARCH QUESTIONS AND SEMI-STRUCTURED INTERVIEW INSTRUMENT

RQ1

Does the reviewer's perspective impact their evaluation of trustworthy digital repository evidence? If so, what are the nature of these differences?

RQ2

Does CoreTrustSeal reviewers' perception of the application strengths vary given their perceived role in the process?

Introduction:

This interview is part of my dissertation research at the University of North Carolina under the review of Helen Tibbo.

Q1: Is it OK for me to record this interview?

My Action: If OK with recording, I will make sure they are all working at this point.

Thank You.

These recordings will be deleted after 6 months, and your name and/or identifying information will be removed from the transcripts. Only myself, my advisor, and the transcriptionist will have access to these recordings prior to their deletion. All results will be de-identified and presented in summary form. At any point in the interviews you can refuse to answer any questions or end your participation at any time.

Reviewer Background Information:

The purpose of this section is to begin the conversation as well as supplement the collection of demographics information related reviewers training and experience needed for RQ1. This is designed to group reviewers into three predefined categories that fit them best: technologist versus archivist versus administrator.

My career path to archives and digital preservation has been somewhat diverse.

Q2: How would you describe your academic and professional background?

Q3: How did you become interested in archives and repositories?

Q4: Which of the following best describes you? Technologist/Archivist/Administrator. If you see yourself serving more than one of these roles, rank them as you wish.

Opening Questions on CoreTrustSeal (CTS) Experiences:

The purpose of this section is to begin with more open-ended questions to get respondents talking about their CTS experiences and begin to identify any similarities/differences in

perspectives across the three categories of reviewers for RQ1.

Q5: *Overall, how has your experience as a CTS reviewer been?*

Q6: *What do you see as the strengths of the CTS?*

Q7: *What do you see as the challenges of the CTS process?*

Q8: *As a CTS reviewer, what have been your greatest challenges?*

Examining reviewers' perspectives:

The goal of this section is to probe deeper into perspectives of the reviewers on evidence. What are the evidence components that they feel are most important? Which requirements do they feel are critical to the evaluation? These questions are generally from my experiences on the CTS board.

Q9: *In your opinion, what are the top three requirements in the CTS application and why do you think they are the most important?*

Q10: *What are the most important forms of evidence?*

Potential Probes: What role does publish policy play? MOUs/SLAs/Technical Documentation

Defining reviewers' perceived role as well as any differences they have seen across applications:

This section's goal is to understand how the reviewers see themselves in the CTS process for RQ2. Is it as an auditor, formal reviewer, peer-to-peer helper or working for CTS board and does it vary across the three identified groups? Also, this section includes questions looking at perspectives across different applications. Questions in the group came from the Stage 1 content analysis and categories identified during that process.

Q11: *Who do you feel you are serving during a CTS application review?*

Potential Probes: *Do you think applicants are formal customers or peers seeking your help? Do you think the word auditor or peer reviewer fits your role best?*

Note: If they reference auditor, use Q11-1 and skip Q11-2. If they reference peers, skip to Q11-2.

Q11-1: *If you see your role as more of an auditor, what do you see as your primary responsibility?*

Q11-2: *If you are a peer-reviewer, how do you balance the needs as a reviewer to "grade/evaluate" the application and help guide your peer repositories toward better repository management?*

Q12: *What differences, if any, have you seen in the details of applications across*

different disciplines?

Concluding open-ended questions:

This section allows reviewer to give feedback to CTS as well as possibly see differing priorities based on perspectives across the identified groups.

Q13: *What advice would you have for those applying for new CTS?*

Possible Probes: *What policy development advice would you give repositories seeking CTS? What components do you feel are required for good evidence?*

Q14: *What, if anything, do you wish were different about the CTS process?*

Q15: *What topics do you think are important that we have not covered today?*

Q16: *Would it be OK to contact you further if I have any questions pertaining to your responses?*

Possible open probes to use during the interview:

To clarify and/or get more detail

- It sounds like you are saying “. . .”. Is that a fair summary?
- Tell me more about that.
- Can you give me an example?

Get rationale or feeling

- Why does that stand out in your memory?
- Why do you think you noticed that?
- What was significant about this to you?

Redirect back to topic

- How does this issue relate to the topic we started with?
- Can you recall the associations that led you from our original topic to this one?
- I'd like to understand more about how this relates to the earlier topic we were talking about.

REFERENCES

- Altman, M., & Crabtree, J. (2011). TRAC-based auditing of LOCKSS: Using the SafeArchive system. In *Proceedings of the Archiving 2011 Conference* (Vol. 7, pp. 165–170). Society for Imaging Science and Technology.
<http://imaging.org/IST/store/epub.cfm?abstrid=44591>
- Amazon Web Services, Inc. (2018). *Amazon S3 Glacier & S3 Glacier deep archive*.
[//aws.amazon.com/glacier/](https://aws.amazon.com/glacier/)
- American Education Research Association. (2013). *OSTP announces plans to increase access to federally funded research*.
<http://www.aera.net/ResearchPolicyAdvocacy/AERAandOpenAccess/OSTPAnnouncesPlanstoIncreaseAccessstoFederal/tabid/14784/Default.aspx>
- American National Standards Institute (ANSI). (2018). *Home Page*. <http://www.ansi.org/>
- Anderson, S. W., Daly, J. D., & Johnson, M. F. (2009). Why firms seek ISO 9000 certification: regulatory compliance or competitive advantage? *Production and Operations Management*, 8(1), 28–43. <https://doi.org/10.1111/j.1937-5956.1999.tb00059.x>
- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27–31. <https://doi.org/10.3200/CTCH.53.1.27-31>
- ANDS. (2016). *Home*. <http://ands.org.au/home>
- APARSEN. (n.d.). *Report on peer review of research data in scholarly communication (Part A of D33.1)*. http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D33_1B-01-1_1_incURN.pdf
- APARSEN. (2014). *Strategic plans*.
<http://www.alliancepermanentaccess.org/index.php/about/strategic-plans/>
- Archive-It. (2018). *About Archive-It*. <https://archive-it.org/blog/learn-more/>
- Archivematica. (2014). *Internal audit tool*.
https://www.archivematica.org/wiki/Internal_audit_tool
- Archivematica. (2018). *Open-source digital preservation system*.
<https://www.archivematica.org/en/>
- ArchivesSpace. (2018). *History*. <https://archivesspace.org/about/history>
- Archivists' Toolkit. (2018). *Archivists workbench*. <http://archiviststoolkit.org/node/263>

- Bak, G. (2016). Trusted by whom? TDRs, standards culture and the nature of trust. *Archival Science*, 16(4), 373–402. <https://doi.org/10.1007/s10502-015-9257-1>
- Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, 2, 8–14. <https://doi.org/10.1016/j.npls.2016.01.001>
- BitCurator. (2018). *Home Page*. <http://www.bitcurator.net/>
- Blank, G., & Rasmussen, K. (2004). The data documentation initiative: The value and significance of a worldwide standard. *Social Science Computer Review*, (22), 307–318. <https://doi.org/10.1177/0894439304263144>
- Blue Ribbon Task Force. (2008). *Sustaining the digital investment: Issues and challenges of economically sustainable digital preservation*. https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/UCSD_US/S081215T.pdf
- Board of Trustees of the University of Illinois. (2016). *The National Data Service*. <http://www.nationaldataservice.org/>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bresnahan, M. M., & Johnson, A. M. (2013). Assessing scholarly communication and research data training needs. *Reference Services Review*, 41(3), 413–433. <https://doi.org/10.1108/RSR-01-2013-0003>
- Bureau of Meteorology. (2018). *Quality management—WMO*. http://www.bom.gov.au/wmo/quality_management.shtml
- Byrnes, P. E., Al-Awadhi, A., Gullvist, B., Brown-Liburd, H., Teeter, R., Warren, J. D., & Vasarhelyi, M. (2018). Evolution of auditing: From the traditional approach to the future audit. In D. Y. Chan, V. Chiu, & M. A. Vasarhelyi (Eds.), *Continuous auditing* (pp. 285–297). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78743-413-420181014>
- Certification — World Data System: Trusted Data Services for Global Science. (2020) <https://www.worlddatasystem.org/services/certification>
- Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, 10(6), 807–815. <https://doi.org/10.1016/j.cptl.2018.03.019>
- Center for Research Libraries. (2014). *Home Page*. <http://www.crl.edu/>
- Center for Research Libraries. (2007). *Ten principles*. <http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re>
- Center for Research Libraries. (2018). *Certification and assessment of digital repositories*. <https://www.crl.edu/archiving-preservation/digital-archives/certification-assessment>

- CESSDA. (2018). *Home*. from <http://cessda.net/>
- Chandler, R., Landis, B., & Westbrook, B. (2001). *Archivists workbench* [White paper].
- Chen, S.-Y., Conway, M., Crabtree, J., Lee, C. A., Misra, S., Moore, R., Rajasekar, A., Russell, T., Simmons, I., Stillwell, L., Tibbo, H., & Xu, H. (2015a). *Policy examples workbook—iRODS 4.2 DataNet Federation Consortium*. iRODS Consortium.
<http://datafed.org/dev/wp-content/uploads/2016/05/DFC-policy-examples.pdf>
- Chen, S.-Y., Conway, M., Crabtree, J., Lee, C. A., Misra, S., Moore, R., Rajasekar, A., Russell, T., Simmons, I., Stillwell, L., Tibbo, H., & Xu, H. (2015b). *Policy template workbook—iRODS 4.2 DataNet Federation Consortium*. iRODS Consortium.
<http://datafed.org/dev/wp-content/uploads/2016/05/DFC-policy-template.pdf>
- CLIR + DLF. (2020). *DLF + NDSA*. <https://www.diglib.org/groups/dlf-ndsa/>
- CODATA, The Committee on Data for Science and Technology. (2018). *Home Page*
<http://www.codata.org/>
- Connaway, L. (2015, October). *The library in the life of the user: Engaging with people where they live and learn*.
<http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-library-in-life-of-user.pdf>
- Consultative Committee for Space Data Systems. (2012). *Reference model for an open archival information system (OAIS)* (Magenta Book No. 650.0-M-2). National Aeronautics Space Agency. <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- COPTR. (2018). *Main page*. http://coptr.digipres.org/Main_Page
- CoreTrustSeal. (2017a). *Core certified repositories*. <https://www.coretrustseal.org/why-certification/certified-repositories/>
- CoreTrustSeal. (2017b). *Core trustworthy data repositories requirements*.
<https://doi.org/10.17026/dans-22n-gk35>
- CoreTrustSeal. (2018). *Assembly of reviewers*. <https://www.coretrustseal.org/about/assembly-of-reviewers/>
- CoreTrustSeal Standards and Certification Board. (2019). *CoreTrustSeal trustworthy data repositories requirements: Extended Guidance 2020–2022*.
<https://doi.org/10.5281/ZENODO.3632533>
- Cornell University Library. (2018). *Digital scholarship & preservation services*.
<https://www.library.cornell.edu/about/inside/departments/dsps>
- Crabtree, J. (2009). *International data technology alliance workshop*.
http://www.odum.unc.edu/content/pdf/Odum_background_paper_Final.pdf

- Crabtree, J. (2013). Building on the work of colleagues: A moment of reflection. *IASSIST Quarterly, Special Volume Honoring Sue Dodd*.
http://www.iassistdata.org/sites/default/files/iqvol371_4_crabtree.pdf
- CRADLE. (2018). *Research data management & sharing MOOC*. <http://cradle.web.unc.edu/>
- Creswell, J. W. (2009). *Research design: qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage Publications.
- Data Archiving and Networked Services. (2018). *Homepage*. <http://www.dans.knaw.nl/>
- Data Asset Framework. (2018). *Home Page*. <http://www.data-audit.eu/>
- Data Documentation Initiative. (2018). *Welcome to the Data Documentation Initiative*.
<http://www.ddalliance.org/>
- Data Seal of Approval. (2018a). *Guidelines*.
<http://www.datasealofapproval.org/en/information/guidelines/>
- Data Seal of Approval. (2018b). *Home*. <http://datasealofapproval.org/en/>
- DataAccessioner. (2016). *Home Page*. <http://dataaccessioner.org/>
- DataCite. (2018). *Home Page*. <https://www.datacite.org/>
- Datanet. (2015). In *Wikipedia, the free encyclopedia*. Retrieved from
<https://en.wikipedia.org/w/index.php?title=Datanet&oldid=666920057>
- DataNet Federation Consortium. (2016). *Home*. <http://datafed.org/>
- DataPASS. (2018). *Best practices*. <http://www.data-pass.org/best-practices.jsp>
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2005). *The SAGE handbook of qualitative research* (3rd ed). Sage Publications.
- DeSantis, L., & Ugarriza, D. N. (2000). The concept of theme as used in qualitative nursing research. *Western Journal of Nursing Research*, 22(3), 351–372.
<https://doi.org/10.1177/019394590002200308>
- DigCCurr. (2018). *About DigCCurr II: Extending an international digital curation curriculum to doctoral students and practitioners*. <http://ils.unc.edu/digccurr/aboutII.html>
- Digital Curation Centre. (2018). *DCC curation lifecycle model*.
<http://www.dcc.ac.uk/resources/curation-lifecycle-model>
- Digital Curation Centre. (2018). *DCC Home*. <http://www.dcc.ac.uk/>
- Digital Curation Centre and DigitalPreservationEurope. (2018). *Welcome to DRAMBORA interactive: Log in or register to use the toolkit*. <http://www.repositoryaudit.eu/>

- Digital POWRR. (2016). *About POWRR*. <http://digitalpowrr.niu.edu/>
- Digital POWRR. (2018). *Digital preservation tool grid*. <http://digitalpowrr.niu.edu/tool-grid/>
- Digital Preservation Europe. (2018a). *DPE Home*. <http://www.digitalpreservationeurope.eu/>
- Digital Preservation Europe. (2018b). *Publications*.
<http://www.digitalpreservationeurope.eu/platter/>
- Digital Preservation Network. (2018). *Home*. <http://dpn.org/>
- Dillo, I., & De Leeuw, L. (2015). Ten years back, five years forward: The data seal of approval. *International Journal of Digital Curation*, 10(1), 230–239.
<https://doi.org/10.2218/ijdc.v10i1.363>
- DIN. (2014). *Information and documentation—Criteria for trustworthy digital archives*.
<http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855&artid=147058907&languageid=en&bcrumblevel=3>
- DIN. (2018). *DIN standards committee for information and documentation*.
http://www.nabd.din.de/cmd;jsessionid=YFJDUC9GGTJGZ16FAAUQOMVD.4?=&languageid=en&workflowname=InitCommittee&search_committee=nabd
- Dingwall, G. (2004). Trusting archivists: The role of archival ethics codes in establishing public faith. *The American Archivist*, 67(1), 11–30.
<https://doi.org/10.17723/aarc.67.1.mw0914r2p52xx2t4>
- Donaldson, D. R., Dillo, I., Downs, R., & Ramdeen, S. (2017). The perceived value of acquiring data seals of approval. *International Journal of Digital Curation*, 12(1), 130–151.
<https://doi.org/10.2218/ijdc.v12i1.481>
- Educopia. (2018). *Home page*. <http://educopia.org/>
- Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic fieldnotes* (2nd ed.). The University of Chicago Press.
- Erikson, C., & Lunt, B. (2015). Alternatives for long-term storage of digital information. *IPRES 2015 Proceeding*. <https://phaidra.univie.ac.at/view/o:429585>
- European Commission. (2018). *Turning FAIR data into reality: final report and action plan from the European Commission expert group on FAIR data*.
<https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF>
- Ex Libris. (2016). *Overview*. <http://www.exlibrisgroup.com/category/RosettaOverview>
- Fielding, N., & Fielding, J. L. (1986). *Linking data*. Sage Publications.

- Figshare. (2015). *Figshare gets serious about storage*.
https://figshare.com/blog/figshare_gets_serious_about_storage/156
- Figshare. (2018). *Home Page*. <https://figshare.com/>
- Galvin, R. (2015). How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge? *Journal of Building Engineering*, 1, 2–12. <https://doi.org/10.1016/j.job.2014.12.001>
- Giaretta, D. (2012). *OAIS model and certification of trusted digital repositories*. Retrieved from <http://93.63.166.138:8080/dspace/bitstream/2012/117/3/Giaretta-David-CH2012-OAIS%20model%20and%20certification%20of%20trusted%20digital%20repositories.pdf>
- Giaretta, D. (2017a). *PTAB certified clients*. <http://www.iso16363.org/iso-certification/certified-clients/>
- Giaretta, D. (2017b). *PTAB first in the world to be accredited for ISO 16363 audit and certification*. <http://www.iso16363.org/ptab-accredited-for-iso-16363-audit-and-certification/>
- Giaretta, D. (2018). *First repository in the world has been awarded ISO 16363 certification*. <http://www.iso16363.org/first-repository-in-the-world-has-been-awarded-iso-16363-certification/>
- GO FAIR. (2018). *Home page*. <https://www.go-fair.org/>
- Guest, G., Namey, E. E., & Mitchell, M. L. (2013). *Collecting qualitative data: a field manual for applied research*. SAGE Publications.
- Hagaman, A. K., & Wutich, A. (2017). How many interviews are enough to identify metathemes in multisited and cross-cultural research? Another perspective on Guest, Bunce, and Johnson's (2006) landmark study. *Field Methods*, 29(1), 23–41.
<https://doi.org/10.1177/1525822X16640447>
- HATII. (2008). *D3.2 Repository planning checklist and guidance*.
<http://www.digitalpreservationeurope.eu/platter.pdf>
- Heidegger, M., & Hofstadter, A. (1988). *The basic problems of phenomenology*. Indiana University Press.
- Heritrix. (2018). *Home page*. <http://crawler.archive.org/index.html>
- Hockx-Yu, H., & Knight, G. (2008). What to preserve? Significant properties of digital objects. *International Journal of Digital Curation*, 3(1), 141–153.
<https://doi.org/10.2218/ijdc.v3i1.49>
- IIA. (2010). *Measuring internal audit effectiveness and efficiency*.

- http://iia.nl/SiteFiles/IIA_leden/Praktijkguiden/Practice%20guide%20Measuring%20Internal%20Audit%20Effectiveness.pdf
- Institute of Internal Auditors. (2020). *Standards & guidance—International Professional Practices Framework (IPPF)*. <https://na.theiia.org/standards-guidance/Pages/Standards-and-Guidance-IPPF.aspx>
- Institute of Museum and Library Services (IMLS). (2018). *Home*. <https://www.imls.gov/>
- International Federation of Data Organizations for the Social Sciences. (2018). *IFDO Home*. <http://ifdo.org/wordpress/>
- Internet Archive. (2016). *Home page*. <https://archive.org/>
- iPRES. (2018). *iPRES Home Page*. <https://ipres-conference.org/>
- iRODS. (2016). *Home*. <http://irods.org/>
- iRODS. (2018). *iRODS Data Grids...*
https://www.irods.org/index.php/IRODS:Data_Grids,_Digital_Libraries,_Persistent_Archives,_and_Real-time_Data_Systems
- ISO. (2001). *ISO 15489-1:2001—Information and documentation—Records management—Part 1: General*.
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=31908
- ISO. (2005). *ISO/IEC 17799:2005/Cor 1:2007*.
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=46381
- ISO. (2011). *ISO/IEC 17021:2011—Conformity assessment—Requirements for bodies providing audit and certification of management systems*.
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=56676
- ISO. (2012). *ISO 16363:2012—Space data and information transfer systems—Audit and certification of trustworthy digital repositories*.
http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510
- ISO. (2014). *ISO 16919:2014—Space data and information transfer systems—Requirements for bodies providing audit and certification of candidate trustworthy digital repositories*.
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57950
- ISO. (2018a). *home*. <http://www.iso.org/iso/home.htm>
- ISO. (2018b). *ISO 9000 quality management*. http://www.iso.org/iso/iso_9000
- ISO. (2018c). *ISO 14721:2012—Space data and information transfer systems—Open archival information system (OAIS)—Reference model*.
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57284

- Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., Stewart, C., Blake, M., Herndon, J., McGeary, T. M., & Hull, E. (2018). Data curation network: A cross-institutional staffing model for curating research data. *International Journal of Digital Curation*, 13(1), 125–140. <https://doi.org/10.2218/ijdc.v13i1.616>
- Kelly, S. E. (2010). Qualitative interviewing techniques and styles. In I. Bourgeault, R. Dingwall, & R. De Vries, *The SAGE handbook of qualitative methods in health research* (pp. 307–326). SAGE Publications Ltd. <https://doi.org/10.4135/9781446268247.n17>
- Kueppers, R. J., & Sullivan, K. B. (2010). How and why an independent audit matters. *International Journal of Disclosure and Governance*, 7(4), 286–293. <https://doi.org/10.1057/jdg.2010.22>
- Larrimer, N. (2016). *ANAB announce accreditation for ISO 16363*. <http://anab.org/media/59172/hu327.pdf>
- Lee, C., Allard, S., McGovern, N., & Bishop, A. (2017). Open data meets digital curation: An investigation of practices and needs. *International Journal of Digital Curation*, 11(2), 115–125. <https://doi.org/10.2218/ijdc.v11i2.403>
- Lee, C. A. (2010). Open Archival Information System (OAIS) reference model. In *Encyclopedia of Library and Information Sciences* (3rd ed., pp. 4020–4030). Taylor & Francis. <http://www.tandfonline.com/doi/book/10.1081/E-ELIS3>
- Library of Congress. (2011). *ARTstor releases Embedded Metadata Extraction Tool (EMET) and final report*. http://www.digitalpreservation.gov/news/2011/20110217news_artstor_final_report.html
- Library of Congress. (2016). *Library develops bagit specification for transferring digital content*. http://www.digitalpreservation.gov/news/2008/20080602news_article_bagit.html
- Library of Congress. (2018). *National Digital Stewardship Alliance*. <http://www.digitalpreservation.gov/ndsa/NDSAtoDLF.html>
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giarretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST principles for digital repositories. *Scientific Data*, 7(1), 144. <https://doi.org/10.1038/s41597-020-0486-7>
- Livescribe. (2018). *Store*. http://store.livescribe.com/?gclid=Cj0KEQjwrte4BRD-oYi3y5_AhZ4BEiQAzIFxn0mqc8oEzJkWBaRziIOdOpeHXLF9zsrRtXBil81vp8aArwk8P8HAQ
- LOCKSS. (2018a). *Networks*. <http://www.lockss.org/community/networks/>
- LOCKSS. (2018b). *What is LOCKSS?* <http://www.lockss.org/about/what-is-lockss/>

- MANTRA. (2018). *Mantra*. <http://datalib.edina.ac.uk/mantra/>
- Martin, R. C. (2003). *Agile software development: principles, patterns, and practices*. Prentice Hall.
- MAXQDA. (n.d.). *What is MAXQDA?* <https://www.maxqda.com/what-is-maxqda>
- Maxwell, J. A. (2013). *Qualitative research design: an interactive approach* (3rd ed.). SAGE Publications.
- McGovern, N. (2013). *Trust in repositories: Building and measuring trustworthiness using TRAC*. <http://or2013.net/sessions/trust-repositories-building-and-measuring-trustworthiness-using-trac>
- McGovern, N. (2016). *Digital preservation management (DPM) workshops*. <http://www.dpworkshop.org/workshops/fiveday.html>
- MetaArchive. (2018). *Home page*. <https://www.metaarchive.org/>
- Mihas, P. (2019). *Learn to analyze written text using discourse analysis*. SAGE Research Methods. <https://methods.sagepub.com/dataset/written-text-discourse-analysis-united-states?fromsearch=true>
- Moore, R. (2008). Towards a theory of digital preservation. *The International Journal of Digital Curation*, 3(1), 63–75.
- Moore, R., Rajasekar, A., & Xu, H. (2015). DataNet Federation consortium preservation policy toolkit. *Proceedings of the 12th International Conference on Digital Preservation*. <http://phaidra.univie.ac.at/o:429554>
- NARA. (2018). *National archives*. <https://www.archives.gov/>
- National Science Foundation. (2016a). *Big data regional innovation hubs: Establishing spokes to advance big data applications*. https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505264
- National Science Foundation. (2016b). *Data infrastructure building blocks*. http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504776
- Nesstar. (2016). *Publish data on the web*. <http://www.nesstar.com/>
- Nestor. (2018). *Nestor-Siegel*. http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html;jsessionid=E1D74DEA1B06DB494FB0FD26EDF34590.prod-worker2
- Noriah, M. A. (2006). Gaining competitive advantage in personal dosimetry services through ISO 9001 certification. *Radiation Protection Dosimetry*, 125(1–4), 101–104. <https://doi.org/10.1093/rpd/ncl381>

- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 160940691773384. <https://doi.org/10.1177/1609406917733847>
- Open Preservation Foundation. (2018). *Shared solutions for effective and efficient digital preservation*. <http://openpreservation.org/>
- Open-Data-in-a-Big-Data-World-Short. (2016). <http://www.icsu.org/science-international/accord/open-data-in-a-big-data-world-short>
- ORCID. (2018). *Home* <http://orcid.org/>
- Pejsova, P., & Vaska, M. (2010). *Audit DRAMBORA for trustworthy repositories: A study dealing with the digital repository of grey literature*. http://greynet.org/images/GL13-S3P,_Pejsova_and_Vaska.pdf
- Pett, J., Hubble, M., & Koelzer, L. (2014). *Metrics by design: A practical approach to measuring internal audit performance*. <https://www.pwc.com/us/en/risk-assurance-services/publications/assets/internal-audit-performance-metrics.pdf>
- Pirson, M., & Malhotra, D. (2011). Foundations of organizational trust: What matters to different stakeholders? *Organization Science*, 22(4), 1087–1104. <https://doi.org/10.1287/orsc.1100.0581>
- Portico. (2018). *Digital preservation*. <http://www.portico.org/digital-preservation/>
- PREMIS Editorial Committee. (2011). *PREMIS data dictionary for preservation metadata (2.1)*. <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>
- Preservica. (2018). *Home page*. <http://preservica.com/>
- Preserving Digital Information. Waters, Donald; Garrett, John (1996, May 1). <https://www.clir.org/pubs/reports/pub63/>
- Prieto, A. G. (2009, 001). From conceptual to perceptual reality: trust in digital repositories. *Library Review*, 58(8), 593–606. <https://doi.org/10.1108/00242530910987082>
- PTAB. (2014a). *Audit preparation*. <http://www.iso16363.org/preparing-for-an-audit/>
- PTAB. (2014b). *Previous courses*. <http://www.iso16363.org/courses/previous-courses/>
- PTAB. (2018). *Home Page*. <http://www.iso16363.org/>
- Purposive Sample. (2008). In P. Lavrakas, *Encyclopedia of Survey Research Methods*. Sage Publications, Inc. <https://doi.org/10.4135/9781412963947.n419>
- RCV Academy. (2020). *Auditing origin and evolution—History of auditing*. <https://www.rcvacademy.com/auditing-origin-evolution/>

- RDA. (2018a). *Home*. <https://rd-alliance.org/>
- RDA. (2018b). *Repository audit and certification DSA–WDS partnership WG*. <https://www.rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg.html>
- Reagan, M., Rajasekar, A., Schroeder, W., & Wan, M. (2011). *Policy-based data management*. <http://datafed.org/dev/wp-content/uploads/2012/04/DCC-iRODS.pdf>
- Regents of the University of California. (2020). *Chronopolis*. <https://libraries.ucsd.edu/chronopolis/>
- Research Libraries Group (RLG). (2002). *Trusted digital repositories: Attributes and responsibilities*. <http://www.oclc.org/resources/research/activities/trustedrep/repositories.pdf>
- RLD/NARA Task Force on Digital Repository Certification. 2007. *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Publisher: Chicago, CRL.
- Rickards, L., Vardigan, M., Diepenbroek, M., Dillo, I., Genova, F., L'hours, H., Hugo, W., Khodiyar, V., Li, G., Minster, J.-B., Trilsbeek, P., Wyborn, L., Edmunds, R., Mokrane, M., & Panagou, E. (2016). *DSA–WDS partnership working group catalogue of common procedures for certification*. <https://doi.org/10.15497/rda00019>
- Ross, S. (2007). *Preservation pressure points: Evaluating diverse evidence for risk management*. <http://www.repositoryaudit.eu/images/PreservationPressurePoints.pdf>
- Ross, S., & McHugh, A. (2006). The role of evidence in establishing trust in repositories. *D-Lib Magazine*, 12(7/8). <https://doi.org/10.1045/july2006-ross>
- Rowley, J. (2012). Conducting research interviews. *Management Research Review*, 35(3/4), 260–271. <https://doi.org/10.1108/01409171211210154>
- SafeArchive. (2018). *Home*. <http://www.safearchive.org/>
- Savin-Baden, M., & Major, C. H. (2012). *Qualitative research: the essential guide to theory and practice*. Routledge.
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). SAGE.
- Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (4th ed.). Teachers College Press.
- Shaman. (2018). *Home*. <http://shaman-ip.eu/>
- Skinner, K., & McGovern, N. (2012). *Aligning national approaches to digital preservation proceedings*. <https://educopia.org/publications/anadp>

- Skinner, K., Schultz, M., & MetaArchive Cooperative (U.S.). (2010). *A guide to distributed digital preservation*. Educopia Institute.
- Smorul, M. (2010). *Audit control environment*.
<http://www.digitalpreservation.gov/meetings/documents/othermeetings/Smorul-Audit.pdf>
- Speck, J. G. (2010). Protecting public trust: An archival wake-up call. *Journal of Archival Organization*, 8(1), 31–53. <https://doi.org/10.1080/15332748.2010.483389>
- Stake, R. E. (1995). *The art of case study research*. Sage Publications.
- Taylor, S. J., & Bogdan, R. (1998). *Introduction to qualitative research methods: A guidebook and resource* (3rd ed.). Wiley.
- Terra Populus. (2018). *Integrated data on population and environment*. <http://www.terrapop.org/>
- The Dataverse Network. (2018). *Home*. <http://thedata.org/>
- The Dataverse Project. (2018). *Home Page*. <http://dataverse.org/>
- The Odum Institute. (2018a). *About the archive*.
<http://www.odum.unc.edu/odum/contentSubpage.jsp?nodeid=586>
- The Odum Institute. (2018b). *Advancing social science teaching and research*.
<http://www.odum.unc.edu/odum/home2.jsp>
- The Odum Institute. (2018c). *Data management services*.
<http://www.irss.unc.edu/odum/contentSubpage.jsp?nodeid=661>
- Trusted Digital Repository. (2018). *Making the digital world more reliable*.
<http://www.trusteddigitalrepository.eu/Trusted%20Digital%20Repository.html>
- UNC-Libraries. (2018). *Curators-workbench*. <https://github.com/UNC-Libraries/Curators-Workbench>
- University of Glasgow. (2018). *Information studies*.
<http://www.gla.ac.uk/subjects/informationstudies/>
- Ward, J. (2014). *An examination of machine-level policy implementation within iRODS data grids*. <https://irods.org/wp-content/uploads/2014/06/irods-user-mtg-18Jun2014-jward.pdf>
- Ward, J. H., Wan, M., Schroeder, W., Rajasekar, A., de Torcy, A., Russell, T., Xu, H., & Moore, R. W. (2011). *The integrated Rule-Oriented Data System (iRODS 3.0) micro-service workbook*. CreateSpace.
- Weber, N. M., Palmer, C. L., & Chao, T. C. (2012). Current trends and future directions in data curation research and education. *Journal of Web Librarianship*, 6(4), 305–320.
<https://doi.org/10.1080/19322909.2012.730358>

- Wessely, S. (1998). Peer review of grant applications: What do we know? *The Lancet*, 352(9124), 301–305. [https://doi.org/10.1016/S0140-6736\(97\)11129-1](https://doi.org/10.1016/S0140-6736(97)11129-1)
- World Data System. (2018a). *Home*. <https://www.icsu-wds.org/>
- World Data System. (2018b). *Introduction*. <https://www.icsu-wds.org/organization/intro-to-wds>
- World Meteorological Organization (WMO). (2018). *Applications of meteorology programme*. http://www.wmo.int/pages/prog/amp/index_en.html
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L. B. da S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3. <https://doi.org/10.1038/sdata.2016.18>
- Withers, B., & Ebrahimpour, M. (2000). Does ISO 9000 certification affect the dimensions of quality used for competitive advantage? *European Management Journal*, 18(4), 431–443. [https://doi.org/10.1016/S0263-2373\(00\)00032-3](https://doi.org/10.1016/S0263-2373(00)00032-3)
- Yakel, E., Faniel, I., Kriesberg, A., & Yoon, A. (2013). Trust in digital repositories. *International Journal of Digital Curation*, 8(1), 143–156. <http://dx.doi.org/10.2218/ijdc.v8i1.251>
- Yoon, A. (2014). End users' trust in data repositories: definition and influences on trust development. *Archival Science*, 14(1), 17–34. <https://doi.org/10.1007/s10502-013-9207-8>
- Yoon, A. (2015). *Data reuse and users' trust judgments: toward trusted data curation*. <https://cdr.lib.unc.edu/record/uuid:2c2268b3-88cf-4397-b038-b39e88f80d83>
- Zierau, E., McGovern, N.Y. (2014). Supporting the Analysis and Auditing of Collaborative OAIS's Using an Outer OAIS-Inner OAIS (OO-IO) Model. In Proceedings of the 11th International Conference on Preservation of Digital Objects (iPRES 2014).
- Zierau, E. (2017). OAIS and Distributed Digital Preservation in Practice: An exploration of Danish and other use cases that contribute to the development of the Outer OAIS-Inner OAIS Model for Distributed Digital Preservation In Proceedings of the 14th International Conference on Preservation of Digital Objects (iPRES, 2017)