



Journal of eScience Librarianship

putting the pieces together: theory and practice

Building a Trustworthy Data Repository: CoreTrustSeal Certification as a Lens for Service Improvements

Cara Key, Oregon State University, Corvallis, OR, USA, cara.key@oregonstate.edu 

Clara Llebot, Oregon State University, Corvallis, OR, USA 

Michael Boock, Oregon State University, Corvallis, OR, USA 

Abstract

Objective: The university library aims to provide university researchers with a trustworthy institutional repository for sharing data. The library sought CoreTrustSeal certification in order to measure the quality of data services in the institutional repository, and to promote researchers' confidence when depositing their work.

Methods: The authors served on a small team of library staff who collaborated to compose the certification application. They describe the self-assessment process, as they iterated through cycles of compiling information and responding to reviewer feedback.

Results: The application team gained understanding of data repository best practices, shared knowledge about the institutional repository, and identified areas of service improvements necessary to meet certification requirements. Based on the application and feedback, the team took measures to enhance preservation strategies, governance, and public-facing policies and documentation for the repository.

Conclusions: The university library gained a better understanding of top-notch data services and measurably improved these services by pursuing and obtaining CoreTrustSeal certification.

Received: July 17, 2023 **Accepted:** November 14, 2023 **Published:** December 20, 2023

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0), which permits unrestricted use, distribution, and reproduction in any medium non-commercially, provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc/4.0>.

∞ OPEN ACCESS

Introduction

The emphasis on open sharing of data is expanding in the research community. Increasingly, researchers face a need to evaluate the quality of data repositories in which to share their work. The National Science and Technology Council's *Desirable Characteristics of Data Repositories for Federally Funded Research* (2022) is an example of the developing standards for measuring repository quality. In the U.S., federal agencies such as the National Institutes for Health (NIH) (2020) are working to issue policies and guidance for selecting a repository for grant-funded research data. The number of such policies is expected to increase over the next two years, pursuant to the White House Office of Science and Technology Policy (OSTP) "Nelson Memo" (Nelson 2022).

Having accepted research data deposits in the institutional repository for several years, Oregon State University Libraries and Press resolved to pursue trustworthy data repository certification through the CoreTrustSeal. Certification was intended to bolster OSU researchers' confidence in entrusting their data to the institutional repository. The process provided for a thorough and valuable self-assessment of the repository's alignment with best practices for data repositories and revealed opportunities for service improvements.

Background

Oregon State University is classified as an R1: Doctoral University with "very high research activity" by the Carnegie Classification of Institutions of Higher Education. Fiscal year 2022 (\$471.5 million) is the fourth time in six years that Oregon State's research funding surpassed \$400 million. OSU Libraries established ScholarsArchive@OSU as the university's institutional repository in 2004. The repository contains over 78,000 assets and is primarily focused on traditional university scholarship and publications such as theses and dissertations, research articles, and technical reports. It also provides an open access and long-term storage option for research datasets produced by the university community, holding 200 published datasets at the time of this writing. The rate of dataset deposits has increased in the last three years (OSULP 2018a). The library's research data services program supports data deposits by performing curatorial review of all datasets to ensure they are thoroughly described, organized, and formatted before publication.

Since 2017, ScholarsArchive@OSU has used the open-source repository solution Hyrax, built by the Samvera community. The OSU instance, while supported by a small team of library developers to address local needs, remains close overall to the Hyrax codebase. It has minimal custom software features impacting functionality for research data deposits. The features in place mainly involve workflows, messaging, and licenses that diverge from management of other resource types (OSULP 2023a).

Trustworthy Repository Certification

The library began the process of seeking data repository certification in 2019 with two central goals. The first was to ensure that the data services being offered through the institutional repository met recognized

standards. At that time, OSULP librarians were in active discussion of the future of library research data services; librarians had recently completed a study on the effectiveness of dataset curation practices (Llebot and Van Tuyl 2019) and were engaged in conversations with regional partners about data preservation. The certification process offered an opportunity to systematically evaluate repository practices to similarly inform a strategic direction. The second motivation for pursuing certification was to better serve the university researcher community. A certification would be an unambiguous indicator of quality to a researcher evaluating repository options for their data. It could also support justifications for the use of the institutional repository in researchers' data management plans.

The global repository certification landscape is continuously changing and adapting (Downs 2021). See Table 1 for a summary of major certifications available for data repositories. At the time the library decided to pursue repository certification, two options seemed applicable: CoreTrustSeal, a certification based on a self-evaluation and designed specifically for data repositories of all disciplines; and ISO certification 16363, an extended audit for all types of repositories. Certification from CoreTrustSeal was determined to be an appropriate certification to pursue. The CoreTrustSeal was designated as the Core or Basic level in a three-tiered system, as outlined by the European Framework for Audit and Certification of Digital Repositories (DSA, RAC, and DIN 2010). Given that it was the first certification pursued for ScholarsArchive@OSU, the basic level of certification was seen as the most attainable. Additionally, the scope of the CoreTrustSeal certification was a good fit with the library's goals, and it was accepted and recognized by repository stakeholders and within the community of practice.

CoreTrustSeal, an international non-profit organization whose primary mission is to “[promote] sustainable and trustworthy data infrastructures” (CoreTrustSeal, n.d., a), offers core level certification to any interested data repository meeting the criteria of the Core Trustworthy Data Repositories Requirements. CoreTrustSeal's core certification emerged from two earlier standards, the Data Seal of Approval and the World Data System Regular Member certification, in cooperation with the Research Data Alliance (CoreTrustSeal, n.d., d). Applicants for core certification prepare a description of their compliance with each of sixteen requirements for trustworthy data repositories, in the categories *Organizational Infrastructure, Digital Object Management, and Information Technology and Security*. It should be noted that the CoreTrustSeal requirements have been updated for 2023-2025 (CoreTrustSeal Standards and Certification Board 2022), but the library sought certification using the 2020-2022 requirements (CoreTrustSeal Standards and Certification Board 2019). Reviewers from the CoreTrustSeal Board and from certified organizations evaluate and provide feedback on the application. Applicants can respond to feedback and resubmit the application up to four times. To be approved, applicants must either have fully met the criteria for each requirement or be in the process with a clear and realistic plan for meeting the criteria. The certification has a fee and must be renewed every three years (CoreTrustSeal, n.d., c).

Table 1: Major repository certifications

Name	Description	Year	Current
Open Archival Information System (OAIS). ISO 14721:2012	A framework, not a certification. Most certifications, current or not, are based on this one. Created by the Consultative Committee for Space Data Systems.	2012; Out-dated version from 2003.	Yes
NESTOR, DIN 31644	Adapted to the German context and used mainly in Germany.	2007	Yes
Trustworthy Repositories Audit and Certification (TRAC)	Published by the Center for Research Libraries and the Research Libraries Group (RLG), based on AOIS. Superseded by ISO 16363.	2007	No
Data Seal of Approval (DSA)	Developed by the Dutch Data Archiving and Networked Services (DANS) for data repositories. Oriented towards the Social Sciences and Humanities. Merged with the WDS Certification to become the CoreTrustSeal.	2008	No
ISO 16363	Related to other ISO standards that regulate certification bodies, such as the ISO 16919. Involves a full external audit of the repository. Valid for all types of repositories, not only data repositories. An update, ISO/CD 16363, is currently under development.	2012	Yes
WDS Certification	Developed by the World Data System (WDS) for data repositories. Oriented towards Earth and Space Sciences. Merged with DSA to become the CoreTrustSeal.	2011	No
CoreTrustSeal	For data repositories in all disciplines. Merged the WDS certification and the Data Seal of Approval. Based on a self-evaluation that is reviewed by peers and relies on publicly available evidence.	2017	Yes

ScholarsArchive@OSU's application was compiled by a small team of library faculty including the service manager for the institutional repository, the lead for the research data services program, an expert on digital scholarship, and the head of the library technology department. Additional library staff with expertise in metadata, software, and infrastructure were consulted on application responses for specific requirements. The application team also used completed applications from similar repositories, shared publicly on the CoreTrustSeal website (CoreTrustSeal, n.d., b), as references. ScholarsArchive@OSU's application process required the maximum of five submission attempts, ultimately resulting in successful certification in 2022. The process was prolonged due to the additional work needed to meet the requirements, described in the next section, as well as the intervals between submission and response and the workload of the application team.

Service Improvements

Although certification was not guaranteed, the certification application process itself delivered appreciable benefits for data services in the institutional repository. By performing a thorough, guided comparison of local services against an authoritative standard of practice, the application team members gained a fuller understanding of the characteristics of a high quality, trustworthy data repository. Composing responses and receiving feedback both demonstrated gaps in knowledge and illuminated opportunities for improvement. Furthermore, reviewing examples of successful applications by peer institutions yielded potential models for emulation in aspects where ScholarsArchive@OSU was deficient. Notable improvements were made in the areas of preservation, policies and documentation, and governance, as well as information sharing between library units.

A well-articulated preservation strategy is a central element for a trustworthy data repository. OSU Libraries and Press is committed to digital preservation, including emphasizing digital preservation in the library's most recent two strategic plans (OSULP 2018b). The preservation policy for the institutional repository (prior to the application process) guaranteed bit-level preservation for all repository objects and that "stored digital content will remain both viable and accessible into the indefinite future" (OSULP 2023b). The overarching library-wide digital preservation policy describes the library's mission "to make accessible and hold in trust for future use" its digital resources (OSULP 2022). CoreTrustSeal reviewer feedback stipulated expanding the existing approach to incorporate active preservation strategies, stating that "without a clear commitment to active preservation," the institutional repository could "not be considered in scope for the CoreTrustSeal Trustworthy Digital Repository status."¹ In response, repository managers added several elements to the preservation plan for ScholarsArchive@OSU. Most substantially, OSULP made an explicit guarantee that objects submitted to the institutional repository using a defined set of recommended file formats would remain usable indefinitely. To support this guarantee, the library declared its intent to perform format migration as indicated by evolving standards (OSULP 2023) and published a research guide

1 Anonymous CoreTrustSeal reviewer quoted in email message to authors, March 22, 2022.

for preferred file formats for ScholarsArchive@OSU deposits (OSUL 2023). Repository managers outlined an annual preservation assessment, which includes an update to file format recommendations, an inventory of the extant formats within the repository, a holistic review of fixity check reports, and a preservation policy review (OSULP 2023).

The CoreTrustSeal reviewers' evaluation of an application relies on evidence in the form of publicly available documentation and policies. While composing the application, team members considered the representation of ScholarsArchive@OSU through the lens of publicly available policies and user guides. Several revisions to policy details and documentation followed, including:

- Assuring future custodial responsibility for digital assets in the event that the institutional repository is discontinued.
- Incorporating the library's teaching mission into the collection policy and stating explicit learning goals for data depositors.
- Reflecting the changes to the preservation plan, described above, within the repository-level preservation policy, and clarifying the relationship between the repository-level preservation policy and the library-wide preservation policy (OSULP 2023).

The certification process served as a catalyst to bring together groups of people in the library with common interests. An inactive library digital preservation interest group was reconvened to respond to the preservation strategy requirements, and a cross-departmental repository user group was initiated to advise on decisions regarding institutional repository services. Additionally, the team compiling the application collected and compared distributed documentation, which contributed to knowledge sharing across library departments and allowed for resolving inconsistencies between resources. Post-certification, the compiled application is itself an authoritative information resource about data services in the institutional repository.

Reflections

As a relatively typical institutional repository for which research data sharing specifically is not the core mission, ScholarsArchive@OSU is probably not a standard applicant for CoreTrustSeal certification. The requirements and reviewer feedback indicate that the certification is designed for discipline-specific, dedicated data repositories. During the application process, the ScholarsArchive@OSU team often encountered confusion about the scope of responses—specifically whether to describe practices or features about only datasets, or to cover all resource types throughout the repository with their varying considerations. This experience raised doubts about the appropriateness of CoreTrustSeal for the ScholarsArchive@OSU use case. Ultimately, the team composed application responses that focused on research data but did account for the full range of resource types, and this approach resulted in successful certification. The list of certified trustworthy data repositories includes other institutional repositories as well (CoreTrustSeal, n.d.,

b), underscoring the argument that IRs can be in scope for certification; still, they may face challenges in the process. CoreTrustSeal may see the institutional repositories of Oregon State University and other R1 Carnegie-classified research institutions as in scope given their strong commitments to research.

Independent of the certification seal, the self-assessment experience strengthened the overall quality of not only data repository services, but repository-wide services. Library staff supporting research data in ScholarsArchive@OSU learned a great deal about best practices and about the institutional repository, enabling better service delivery for users. The process highlighted the importance of public-facing documentation, leading to better public policies; the takeaway lesson for the OSULP team is that while good practice is important, so is documentation of good practice. Furthermore, though it has not been emphasized above, the application team demonstrated that the data repository services in place were already sufficient to meet the majority of the core requirements. The combined effect has been a significant increase in confidence in the quality of the program.

Conclusion

Through pursuing and obtaining trustworthy data repository certification, the library was able to assess and confirm the quality of the data repository services program. The public availability of the approved application along with the visible CoreTrustSeal logo on the ScholarsArchive@OSU home page serve to convey that quality to researchers. Based on this experience, the authors strongly recommend any data repository, including institutional repositories, to use the CoreTrustSeal requirements for trustworthy data repositories as a self-assessment tool, even if not considering applying for certification.

Following this work, the repository's first annual preservation assessment was completed in June 2023, with a report available on the library website (OSUL, n.d.). Improvements to the institutional repository and the research data services program are ongoing, including workflow efficiencies, software feature improvements, and investigating preservation of datasets with MetaArchive. The repository team has subsequently composed a report of ScholarsArchive@OSU's compliance with standards for evaluating quality data repositories, which includes as its criteria the CoreTrustSeal Requirements as well as FAIR principles and the 2022 *Desirable Characteristics for Data Repositories for Federally Funded Research* (Key and Llebot 2023). The library aims to increase compliance with requirements rated less than fully implemented by the time ScholarsArchive@OSU is due to renew its certification in 2025.

ScholarsArchive@OSU's final, approved application for CoreTrustSeal Trustworthy Data Repository certification is available at the CoreTrustSeal website (n.d., b) and in the OSU institutional repository (Boock et al. 2022).

Acknowledgements

The content of this article is based on the presentation entitled “Building a trustworthy data repository: Leveraging CoreTrustSeal certification as a lens for service improvements” originally presented at **RDAP Summit 2023**, available from Open Science Framework: <https://osf.io/h4b6w>.

Competing Interests

The authors declare that they have no competing interests.

References

- Boock, Michael, Cara Key, Clara Llebot, Margaret Mellinger, and Steve Van Tuyl. 2022. “ScholarsArchive@OSU Repository Core Trust Seal Self-Assessment.” https://ir.library.oregonstate.edu/concern/technical_reports/9593v342f.
- CoreTrustSeal. n.d., a. “About.” Accessed July 14, 2023. <https://www.coretrustseal.org/about>.
- CoreTrustSeal. n.d., b. “CoreTrustSeal certified data repositories.” Accessed July 14, 2023. <https://amt.coretrustseal.org/certificates>.
- CoreTrustSeal. n.d., c. “Frequently Asked Questions.” Accessed July 14, 2023. <https://www.coretrustseal.org/why-certification/frequently-asked-questions>.
- CoreTrustSeal. n.d., d. “History.” Accessed July 14, 2023. <https://www.coretrustseal.org/about/history>.
- CoreTrustSeal Standards and Certification Board. 2019. “CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022.” <https://doi.org/10.5281/zenodo.3638211>.
- CoreTrustSeal Standards and Certification Board. 2022. “CoreTrustSeal Requirements 2023-2025.” <https://doi.org/10.5281/zenodo.7051012>.
- DSA, RAC, and DIN (Data Seal of Approval, CCSDS/ISO Repository Audit and Certification Working Group, and DIN Working Group “Trustworthy Archives – Certification”). 2010. “Memorandum of Understanding to create a European Framework for Audit and Certification of Digital Repositories.” <http://www.trusteddigitalrepository.eu/Memorandum%20of%20Understanding.html>.
- Downs, Robert R. 2021. “Improving Opportunities for New Value of Open Data: Assessing and Certifying Research Data Repositories.” *Data Science Journal* 20(1): 1–11. <https://doi.org/10.5334/dsj-2021-001>.
- Key, Cara and Clara Llebot. 2023. “Compliance of ScholarsArchive@OSU with requirements for quality research data repositories.” https://ir.library.oregonstate.edu/concern/technical_reports/vh53x4476.
- Llebot, Clara and Steve Van Tuyl. 2019. “Peer Review of Research Data Submissions to ScholarsArchive@OSU: How can we improve the curation of research datasets to enhance reusability?” *Journal of eScience Librarianship* 8(2): e1166. <https://doi.org/10.7191/jeslib.2019.1166>.
- NIH (National Institutes of Health). 2020. “Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research.” Notice number NOT-OD-21-016. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-016.html>.

- NSTC (The National Science and Technology Council). 2022. "Desirable Characteristics of Data Repositories for Federally Funded Research." <https://doi.org/10.5479/10088/113528>.
- Nelson, Alondra. 2022. "Ensuring free, immediate, and equitable access to federally funded research." Office of Science and Technology Policy memorandum. <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>.
- OSUL (Oregon State University Libraries). n.d. "Assessment Reports." Accessed July 14, 2023. <https://library.oregonstate.edu/assessment-reports-0>.
- OSUL (Oregon State University Libraries). 2023. "ScholarsArchive@OSU User Guide: Preferred File Formats." Last modified June 29, 2023. <https://guides.library.oregonstate.edu/Scholars-Archive/PreferredFileFormats>.
- OSULP (Oregon State University Libraries and Press). 2018a. *ScholarsArchive@OSU*. <https://ir.library.oregonstate.edu>.
- OSULP (Oregon State University Libraries and Press). 2018b. "Strategic Plan 2018-23." <https://library.oregonstate.edu/strategic-plan>.
- OSULP (Oregon State University Libraries and Press). 2022. "OSU Libraries Digital Preservation Policy." Last modified May 11, 2022. <https://wiki.library.oregonstate.edu/confluence/x/eSvWAw>.
- OSULP (Oregon State University Libraries and Press). 2023a. *Scholars-Archive*. GitHub repository. Last modified June 27, 2023. <https://github.com/osulp/Scholars-Archive>.
- OSULP (Oregon State University Libraries and Press). 2023b. "ScholarsArchive@OSU Policies." Last modified June 21, 2023. <https://wiki.library.oregonstate.edu/confluence/x/bivWAw>.