THE INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH AND THE DATA SEAL OF APPROVAL: ACCREDITATION EXPERIENCES, CHALLENGES, AND OPPORTUNITIES

M Vardigan[1] and J Lyle[1]*

[1]Inter-university Consortium for Political and Social Research (ICPSR), P.O. Box 1248, Ann Arbor, MI 48106, USA
**Email:*lyle@umich.edu
Email: vardigan@umich.edu

ABSTRACT

The Inter-university Consortium for Political and Social Research (ICPSR), a domain repository with a 50-year track record of archiving social and behavioural science data, applied for—and acquired—the Data Seal of Approval (DSA) in 2010. DSA is a non-intrusive, straightforward approach to assessing organizational, technical, and operational infrastructure, and signifies a basic level of accreditation. DSA assessment helped ICPSR become more transparent, monitor and improve archival processes and procedures, and raise awareness within the organization and beyond about best practices for repositories. We relate our experiences with the DSA process, and describe challenges and opportunities associated with DSA assessment.

Keywords: Assessment, Certification, Data repository, Trusted repository, Data Seal of Approval

1 INTRODUCTION

As data repositories and dissemination platforms proliferate, assessment of repository quality and trustworthiness grows in importance. Assessment promotes trust that data will be available for the long term, provides a transparent view into the workings of the repository, and improves processes and procedures through measurement against a community standard.

Common elements of assessment include review of the organizational framework (e.g., governance, staffing, policies, and finances of the repository), technical infrastructure (e.g., system design and security), and treatment of data (e.g., access, integrity, process, and preservation). This brief paper outlines the experiences of the Inter-university Consortium for Political and Social Research (ICPSR), a domain repository with a 50-year track record of archiving social and behavioural science data, in applying for accreditation under the Data Seal of Approval (DSA). DSA is a relatively lightweight but increasingly recognized accreditation system for scientific data repositories. The DSA assessment process is first described, followed by a discussion of how ICPSR approached and conducted the assessment, and finally, concluding remarks are given about the benefits and limitations of the DSA process and repository accreditation more generally.

2 DATA SEAL OF APPROVAL

The Data Seal of Approval was initiated in 2009 by the Data Archiving and Networked Services, an institute of the Royal Netherlands Academy of Arts and Sciences and the Netherlands Organization for Scientific Research, 'to safeguard data to ensure high quality and to guide reliable management of research data for the future without requiring the implementation of new standards, regulations, or high costs' (Data Seal of Approval, 2013). There are 16 guidelines to the DSA assessment—three target the data producer; three, the data consumer; and ten, the data repository (Table 1). These guidelines operationalize specific fundamental requirements: 'the data can be found on the Internet, the data are accessible (clear rights and licenses), the data are in a usable format, the data are reliable, and the data are identified in a unique and persistent way so they can be referred to' (Data Seal of Approval, 2013).

Table 1. Data Seal of Approval Guidelines (Version 2)

1	The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.
2	The data producer provides the data in formats recommended by the data repository.
3	The data producer provides the data together with the metadata requested by the data repository.
4	The data repository has an explicit mission in the area of digital archiving and promulgates it.
5	The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.
6	The data repository applies documented processes and procedures for managing data storage.
7	The data repository has a plan for long-term preservation of its digital assets.
8	Archiving takes place according to explicit work flows across the data life cycle.
9	The data repository assumes responsibility from the data producers for access and availability of the digital objects.
10	The data repository enables the users to discover and use the data and refer to them in a persistent way.
11	The data repository ensures the integrity of the digital objects and the metadata.
12	The data repository ensures the authenticity of the digital objects and the metadata.
13	The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.
14	The data consumer complies with access regulations set by the data repository.
15	The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.
16	The data consumer respects the applicable licenses of the data repository regarding the use of the data.

Self-assessments against these guidelines are completed online. The repository supplies written evidence statements describing how it complies with each guideline, along with web addresses to key resources that demonstrate compliance. The repository also applies numeric ratings indicating compliance levels for each guideline:

0 = 'N/A: Not Applicable'
1 = 'No: We have not considered this yet'
2 = 'Theoretical: We have a theoretical concept' (i.e., conceptually agreed but as yet unimplemented)
3 = 'In progress: We are in the implementation phase'
4 = 'Implemented: This guideline has been fully implemented for the needs of our repository'

Self-assessments are then peer-reviewed by a DSA Board member, who reviews the evidence and ratings provided and requests additional information if required.

Approximately 22 repositories have been granted the DSA since 2010. Seals 'for a given period can be displayed indefinitely but will need to be updated periodically if the repository wants to stay compliant with newly released standards and receive the latest DSA logo' (Data Seal of Approval, 2013).

DSA has minimal requirements in comparison to other assessments, such as the Trustworthy Repositories Audit and Certification (TRAC, 2007), the Trusted Digital Repository Checklist (2011, also known as International Organization for Standardization (ISO) Draft International Standard 16363), the Nestor Criteria for Trustworthy Digital Archives (2013, also known as Deutsches Institut für Normung 31644), and the Digital Repository Audit Method Based on Risk Assessment (2013). The ISO standard, for example, has close to 100 requirements compared with the 16 of the DSA and involves an onsite audit of the repository.

3 ICPSR's DSA EXPERIENCE

ICPSR applied for and acquired the DSA in 2010. Having previously served as a test audit subject for the TRAC checklist in 2006 (Center for Research Libraries Auditing and Certification of Digital Archives Project, 2006) and having made several improvements to procedures based on that review, ICPSR was in a good position to evaluate and certify against this more lightweight standard. We were interested in finding a basic certification standard that smaller repositories with fewer resources could use to assert their trustworthiness. Staff found the DSA process to be a non-intrusive, straightforward approach to assessing the repository. DSA is less labour- and time-intensive than other assessments; completion of its accreditation documentation took two experienced senior staff members only a few days and did not require extensive assistance from others within the organization.

The DSA assessment process helped ICPSR improve transparency with respect to repository functions and procedures, monitor high-level archival processes, and raise awareness about certification. Since ICPSR previously had undertaken a detailed TRAC test audit, the DSA assessment did not uncover significant flaws in the system, but it did help the organization continue to sharpen processes and procedures. For instance, in documenting our responses to the DSA guidelines, ICPSR staff recognized the need to make policies more public, including posting past versions of Terms of Use agreements. The DSA process also reinforced the need for succession planning for stewardship of ICPSR's digital assets and underscored the importance of improving alignment with the Open Archival Information System (OAIS) model, a framework for archives preserving information for the long term. ICPSR has since established formal succession plans with the Data Preservation Alliance for the Social Sciences (Data-PASS), a 'voluntary partnership of organizations created to archive, catalogue, and preserve data used for social science research' (Data-PASS, 2013). ICPSR is transitioning to a system that will explicitly align ICPSR's archival functions with OAIS. Currently, ICPSR identifies Submission Information Packages and is working on a system whereby data managers will make exact choices about the content of the Archival Information Packages and Dissemination Information Packages (ICPSR 2011–2012 Annual Report, 2012). This new system will enable ICPSR to manage resources more efficiently at the file level and better manage 'nontraditional' content such as video and qualitative data.

The experience of applying for DSA was overall a positive one. Displaying the DSA logo on the ICPSR website is a visible sign that the repository has met the DSA criteria and has achieved trusted status.

4 CHALLENGES AND OPPORTUNITIES

While attaining the Data Seal of Approval proved to be an inexpensive, relatively quick, and user-friendly assessment process for ICPSR, some challenges remain in promoting wider community uptake of the DSA assessment, including:

1. Limited resources at some organizations to devote to assessment,
2. Reliance on trust in user-provided web addresses,
3. Minimal focus on an organization's overall viability and stability,
4. Limited mapping to other assessment processes.

We discuss these challenges and attempt to reframe them as opportunities.

4.1 Resources

Many digital repositories are squeezed for resources, especially smaller organizations. Can repositories justify allocating precious resources to certify that they are trusted, especially if funding agencies and peers already thoroughly inspect and assess an organization's viability and compliance with community standards? This is a common view in the United States of America, where certification has not had the uptake that it has had in Europe. However, as funders continue to mandate open and continuing access to data, it is entirely possible that repositories may be called upon to more openly demonstrate their capacities to preserve data for the long term. The transparency afforded by the DSA process is a good way to accomplish this.

4.2 Trust

DSA is built primarily on the trust that information supplied in resources at repository-provided web addresses is operationalized as outlined. Reviewers are advised to review resources at given links, but external validation is limited since there are no site visits. This imposes risk that an organization could publish policies without adhering to them. We see this risk as relatively low, however, and think that providing evidence of compliance with the DSA guidelines through web addresses makes good sense given the high cost of external audits.

4.3 Organizational viability

While the Data Seal of Approval addresses issues of long-term sustainability of content (Guideline 7: 'The data repository has a plan for long-term preservation of its digital assets') and mission (Guideline 4: 'The data

repository has an explicit mission in the area of digital archiving and promulgates it'), it does not directly address the long-term viability and sustainability of the repository itself. Without an enduring organizational backbone, the long-term preservation of digital assets is limited. We recommend that the DSA consider augmenting the existing DSA guidelines to elicit more information about repository sustainability.

4.4 Mapping

Significant portions of the DSA assessment criteria map to those used by other trusted digital repository certifications, such as the Trusted Digital Repository Checklist (2011). It would be very useful for repositories wishing to acquire certification to understand the differences and similarities between the various available certification processes, in particular where they are complementary. This is being discussed in various forums, including the new Research Data Alliance (2013). The DSA guidelines have many commonalities with those of the World Data System (2013), which developed independently in the Earth and Space Sciences. Mapping these basic certification catalogues would seem to be a good start at a broader mapping across certification standards.

5 CONCLUSION

The Data Seal of Approval provided a relatively lightweight and straightforward assessment protocol against which ICPSR could evaluate and benchmark its performance as a repository. The results of the DSA process helped ICPSR to continue to refine processes and procedures. DSA offers a low entry barrier for repositories to certify that they are trustworthy while helping them to improve their own systems. Although not without cost, the Seal carries meaning that is easily recognized. We expect that as more funding agencies recognize the importance of data creators depositing their data in trusted digital repositories, greater emphasis will be placed on the DSA and other trusted repository assessment processes, with the potential even to form a tiered gradation of accreditation based on size and scope of long-term repositories.

6 ACKNOWLEDGEMENTS

We would like to thank the participants of the International Forum on Polar Data Activities in Global Data Systems, held 15-16 October 2013 in Tokyo, for their comments and suggestions.

7 REFERENCES

Center for Research Libraries Auditing and Certification of Digital Archives Project (2006) *ICPSR Audit Report*. Retrieved November 26, 2013 from the World Wide Web: http://www.crl.edu/sites/default/files/attachments/pages/ICPSR_final.pdf

Data-PASS (2013) Retrieved November 26, 2013 from the World Wide Web: http://data-pass.org/

Data Seal of Approval (2013) Retrieved November 26, 2013 from the World Wide Web: http://datasealofapproval.org/en/information/about/

Digital Repository Audit Method Based on Risk Assessment (2013) Retrieved November 26, 2013 from the World Wide Web: http://www.repositoryaudit.eu/

ICPSR 2011–2012 Annual Report (2012) *File-Level Archive Management Engine (FLAME)* Retrieved November 26, 2013 from the World Wide Web: http://www.icpsr.umich.edu/files/membership/or/annualreport/2011-2012.pdf

Nestor Criteria for Trustworthy Digital Archives (2013) Retrieved November 26, 2013 from the World Wide Web: http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.htmltml

Research Data Alliance (2013) Retrieved December 2, 2013 from the World Wide Web: https://rd-alliance.org

TRAC (2007) Retrieved November 26, 2013 from the World Wide Web: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Trusted Digital Repository Checklist (2011) Retrieved November 26, 2013 from the World Wide Web: http://public.ccsds.org/publications/archive/652x0m1.pdf

World Data System (2013) Retrieved December 2, 2013 from the World Wide Web: http://www.icsu-wds.org/

(Article history:Available online 17 October 2014)