CoreTrustSeal: From academic collaboration to sustainable services

Hervé L'Hours, Mari Kleemola, Lisa de Leeuw [1]

Abstract

National and international digital repositories must design and deliver sustainable services supporting a range of scientific and data management activities while reducing costs and avoiding duplication of effort. The CoreTrustSeal, launched in 2017, defines requirements and offers core level certification for Trustworthy Digital Repositories (TDR) holding data for long-term preservation. This paper traces the journey of the CoreTrustSeal through the Data Seal of Approval (DSA), ICSU World Data System (WDS), Research Data Alliance (RDA) working groups and community engagement, toward becoming a sustainable service supporting global data infrastructure. We outline the design and delivery of the service, current activities, the benefits of certification to a range of communities, and future plans and challenges. As well as providing a historical narrative and current and future perspectives, the CoreTrustSeal experience offers lessons for those involved in developing standards and best practices or seeking to develop cooperative and community-driven efforts bridging data curation activities across academic disciplines, governmental and private sectors.

Keywords

Trust, Trustworthy Digital Repositories, TDR, certification, archives, preservation

Introduction

The CoreTrustSeal[2] is a not-for-profit foundation that authors and maintains the 16 CoreTrustSeal Trustworthy Digital Repository (TDR) Requirements, and the audit procedures and process necessary to attain CoreTrustSeal TDR certification. CoreTrustSeal is governed by the CTS Board that is drawn primarily from the Assembly of Reviewers, which in turn consists of volunteer reviewers designated by CTS certified repositories.

The CoreTrustSeal provides a benchmark for those seeking assurance either for their repository or for the data they produce, own or use, to ensure the data will be actively preserved as digital assets for the long term.

This paper presents the context and process the CoreTrustSeal has taken toward providing a sustainable service. The Board acknowledges that though much has been accomplished, further work remains. This paper seeks to engage in the spirit of openness and community that has created CoreTrustSeal by sharing its experience so others may benefit from our experience in developing common standards, products and services for our data communities. Beyond the formalisation of its requirements, processes and governance bodies, the CoreTrustSeal remains a community-created, community-driven entity with all the complexity, collaboration, promise and compromise that entails.

 L'Hours, Hervé; Kleemola, Mari; de Leeuw, Lisa (2019) CoreTrustSeal: From academic collaboration to sustainable services, IASSIST Quarterly 43 (1), pp. 1-17. DOI: https://doi.org/10.29173/iq936

Standards and Preservation

TDR Certification and the OAIS Model

The word preservation may not always be clearly defined, being sometimes confused with digitisation, or mistaken for good storage practice (Harrower and Cassidy, 2017). For the CoreTrustSeal and related TDR standards, the underlying concepts which define an organisation capable of delivering long-term digital preservation are derived from their common reference: the Open Archival Information System (OAIS) model[3].

The OAIS model makes it clear that a repository shall:

- Obtain sufficient control of the information provided to the level needed to ensure Long Term Preservation.
- Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided, thereby defining its Knowledge Base.
- Ensure that the information being preserved is Independently Understandable to the Designated Community. In particular, the Designated Community should be able to understand the information without needing special resources such as the assistance of the experts who produced the information. (OAIS, 2012).

The *Consultative Committee for Space Data Systems* (CCSDS) originally developed the OAIS reference model (CCSDS 650.0-M-2 with a parallel standard process as ISO14721[4]) as part of its suite of standards for space data systems, but it became the de facto standard for a wider range of disciplines. At the time of its original publication in the late 1990s, OAIS documentation noted the need for some form of compliance certification.

Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist of the Research Library Group (RLG)[5] was developed as part of a wide consortium brought together by the National Archives and Records Administration (NARA) and RLG (Giaretta 2011, 463). CSSDS then took this forward as the *Audit and Certification of Trustworthy Digital Repositories Standard* (ISO16363)[6]. In Germany, the *Network of Expertise in Long-Term Storage of Digital Resources* awards the nestor Seal against their *Criteria for Trustworthy Digital Archives (*Kriterienkatalog vertrauenswürdige digitale Langzeitarchive)[7] (DIN31644). Together the TDR standards TRAC/ISO16363, nestor Seal, CoreTrustSeal (previously Data Seal of Approval) all derive their key concepts from the OAIS' responsibilities to a defined Designated Community.

Designated Community

It is not sufficient to ensure that data are stored in (and made available from) environments which ensure bit-level integrity checks, multi-copy/multi-site redundancy and low-risk disaster recovery methods. Though these functions are vital nodes in our (research) data networks, Trustworthy Digital Repositories are defined by ensuring the availability of data which is *also* understandable and usable by their Designated Community for the long term, ensuring the full value of data assets is assured over time.

The Designated Community is defined, in part, as an "identified group of potential Consumers who should be able to understand a particular set of information" (OAIS, 2012). These requirements make it clear that even if repositories are open to the public, their Designated Community must almost certainly be more clearly and narrowly defined.

The composition and needs of the Designated Community will change over time. New scientific discoveries, or changes to the common software tools they use, may necessitate a change to data provision through emulation or file format migration. Any effective data steward must respond to changes in the knowledge base or technical requirements of their users. The Trustworthy Digital Repository is designed to respond to these changes for the long term, ensuring data, metadata and documentation remain fit for purpose through each round of change.

To obtain the necessary expertise to fulfil this mission, applicants for TDR are likely to be disciplinary repositories or other organizations focused on a defined collection, with a specific topical area, theme, or type of data.

Sharing Expertise and Effort

There is one essential aspect of digital preservation that the OAIS model does not address directly: Preservation of digital assets requires a lot of resources. This is captured by Hedstrom (1998, 190) who defines digital preservation as "the planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable". Giaretta (2011, 8) puts it more bluntly: "the really foolproof solution for digital preservation: money… enough of it, and for an indefinite period."

In a world with limited resources, one way to reduce costs is to share the expertise and the effort of preservation. Since CoreTrustSeal certifications are public, they provide a growing knowledge base of repository practice. Certification also helps to identify the repositories one can trust to have expertise in long term preservation of digital assets and with whom one might wish to collaborate and thus share efforts and costs.

The Data Seal of Approval (DSA)

Sixteen Guidelines for Social Science and Humanities Data

When the Netherlands' *Data Archiving and Network Services* (DANS) was established by the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO), they assigned it the task of developing a Seal of Approval for data, to ensure that archived data can still be found, understood and used in the future. In 2008 the first edition of the Data Seal of Approval, written by Laurents Sesink, René van Horik and Henk Harmsen, was presented in the International Conference on Preservation of Digital Objects (Harmsen, 2008).

The criteria for the Data Seal of Approval were aligned with national and international guidelines for digital data archiving such as nestor, TRAC, and Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)[8] published by the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE). *Foundations of Modern Language Resource Archives of the Max Planck Institute*[9] and *Stewardship of Digital Research Data: A Framework of Principles and Guidelines* published by the *Research Information Network*[10] were also taken into account. In distilling the

minimum set of requirements from these sources the Data Seal of Approval sought to ensure high quality and reliable management of data for the future without requiring the implementation of new standards, regulations or heavy investment..

The result was the sixteen DSA Guidelines for the application and verification of quality aspects regarding the creation, storage and (re-)use of digital research data in the social sciences and humanities. These served as the basis for granting a 'Data Seal of Approval' by the Data Seal of Approval Board.

Three Stakeholder Groups

The guidelines were based on input from three stakeholder groups: Data Producers (quality of the research data content, formats, documentation), Data Repositories (storage quality, organisation of processes, technical infrastructure and assurance of availability), and Data Consumers (quality of data use in terms of access regulations, codes of conduct and licences). Together the DSA guidelines were intended to (DSA 2010):

- "Give **researchers** the assurance that their research results will be stored in a reliable manner and can be reused
- Provide **research sponsors** with the guarantee that research results will remain available for reuse
- Enable **researchers** a reliable means to assess the repository where research data are held.
- Allow **data repositories** to archive and distribute research data efficiently"

Fundamental to the guidelines was that sustainable archiving entails research data are reliable, accessible on the Internet in a usable format, can be referred to, and that relevant legislation with regard to personal information and intellectual property of the data is taken into account (DSA, 2010).

International Board

The response from the research data community made it clear that the DSA Guidelines were of relevance far beyond the Netherlands and beyond the social sciences and humanities. In January of 2009, DANS convened a workshop to transition the DSA governance to an international board for its further development. The initial international Board consisted of DANS (Henk Harmsen, Laurents Sesink, Lisa de Leeuw), ICPSR (Mary Vardigan), UK Data Archive (Matthew Woollard), CINES (Olivier Rouchon), MPI Nijmegen (Paul Trilsbeek), nestor (Natascha Schumann) and the Polar Research Centre (Hans Pfeiffenberger).

This original DSA Board was primarily European-based and with a tendency towards social science data though the community of Seal recipients as a whole remained more diverse and international. Under this Board, the second revision of the Data Seal of Approval was undertaken, leveraging lessons learned from the initial tranche of successful (and unsuccessful) applicants.

In 2015, after a number of changes to those voluntarily serving on the board, the DSA Seal holders were given the opportunity to vote for their representatives drawn from the growing DSA community. Representatives of DANS (Ingrid Dillo), Strasbourg Astronomical Center (Francoise Genova), University College Dublin (John Howard), Finnish Social Science Data Archive (Mari Kleemola), UK Data Archive

(Hervé L'Hours), CINES (Marion Massol), GESIS-Leibniz Institute for the Social Sciences (Natascha Schumann) and MPI Nijmegen (Paul Trilsbeek) took their elected position in January 2016.

Beyond Academic Research Data

The Data Seal of Approval envisaged that data demanded by the social sciences and humanities were of potentially broad provenance noting their guidelines were "of interest to researchers and institutions that create digital research files, to organizations that archive research files, and to users of research data" (DSA, 2010).

Interest from a wider academic sphere made it clear that the certification of digital archives is not only important for scientific archives of primary research data, but also for cultural heritage institutions such as public libraries, museums and archives. The big data revolution encompasses both the capacity for collection and for analysis of data at previously impossible scales. The range of data and the research projects themselves are increasingly heterogeneous. Public-private research partnerships are increasingly common, and data originally conceived for other purposes, including governmental, administrative and other data including social media data, are increasingly used within research. These changes make a narrow definition of research data increasingly difficult to justify as data are created, curated, stored and used by a wider range of actors across the data/research data lifecycle.

Trust in these actors is increasingly critical to our trust in data generally and in scientific data specifically. With these changes in mind, the DSA Board undertook a revision of the DSA Guidelines in 2010. Wording was clarified to encompass the full range of data of interest to the widest possible group of data users and guidance was extended to address the need for more specific examples. This exercise was repeated in 2013 resulting in the second version of the DSA guidelines (DSA 2014-2017).

Wider Adoption of TDR Principles

The added value of the DSA process was first recognized by the approximate 60 individual repositories seeking or achieving DSA at the time. Other adopters included the European Research Infrastructure Consortia (ERIC) as defined by the European Strategy Forum on Research Infrastructures (ESFRI)[11] within the European Research Area (ERA) and Innovation Union. They encompass: major scientific equipment, resources such as collections, archives or scientific data, e-infrastructures such as data and computing systems, and communication networks. The ERICs represent a broad range of potential applicants covering a wide range of subject, disciplinary and scope initiatives. Not all of them are directly comparable from technical or workflow perspectives, but all seek a trust relationship within and between their components (L'Hours et al, 2018). In this context Infrastructures such as CESSDA[12], CLARIN[13] and DARIAH[14] used DSA, and are using the CoreTrustSeal requirements. CLARIN has made certification mandatory for a large part of its centres. All CESSDA Service Providers must seek certification, and the DSA/CoreTrustSeal requirements are mapped directly to their common statutes where possible. DARIAH is using the requirements in their assessment of national contributions to the infrastructure. The emergence of the FAIR data principles (Findable, Accessible, Interoperable, Reusable, see Wilkinson et al. (2016)) alongside TDR standards as important for collaboration in the evolving European Open Science Cloud (EOSC) means that the role of CoreTrustSeal continues to evolve and advance.

A Stepped Framework of Rigour and Trustworthiness

To achieve greater harmonization between the different trust initiatives and their audit criteria, a Memorandum of Understanding (MoU) was signed in 2010 establishing a European Framework for Audit and Certification[15] between the Data Seal of Approval, Audit and Certification of Trustworthy Digital Repositories Standard (ISO16363) and nestor Seal Criteria for Trustworthy Digital Archives (DIN31644). The MoU acknowledged a hierarchy of Basic (now referred to as Core), Extended and Formal certification in order of complexity and audit rigour (Schumann 2012):

1. Basic Certification is granted by obtaining the DSA.
2. Extended Certification requires completing the DSA and an externally reviewed self-audit based either on ISO 16363 or DIN 31644.
3. Formal Certification requires completing the DSA and a full external certification based either on ISO 16363 or DIN 31644.

The memorandum clarified the position of the DSA, and by extension, the CoreTrustSeal, within the wider context of TDR standards. The mission of the CoreTrustSeal remains to offer a basic/core certification, which is low barrier to enter, community driven, and integrated into a stepped framework of improvements for those organisations seeking a more rigorous path.

Standards Integration through the Research Data Alliance

When a Research Data Alliance (RDA) Interest Group on the Certification of Digital Repositories[16] was created, there was recognition of the value of core certification but also a concern that a proliferation of such efforts may be unhelpful. Both the Data Seal of Approval and the membership criteria of the International Council for Science's World Data System (ICSU-WDS) were identified as core efforts with aligned goals. Both held a multidisciplinary remit, though for historical reasons their primary applicants differed with ICSU-WDS coming from the Earth and Space Sciences, so the partnership also offered an opportunity to serve a wider community.

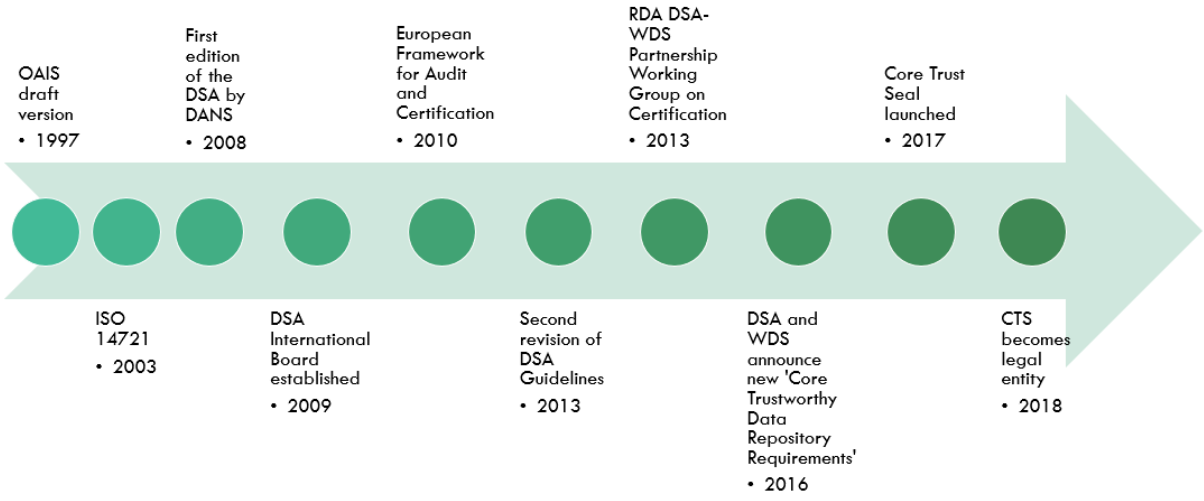
In 2013, the Repository Audit and Certification working group[17] was proposed, with a vision of realizing efficiencies, simplifying assessment options, stimulating more certifications, and increasing impact on the community (Rickards et al, 2016b). The central focus was a DSA-WDS partnership with representatives of both communities involved. In keeping with the transparency principles of the RDA, the interest and working group efforts were open to the full RDA membership for both participation and communication.

The working group undertook an analysis and comparison of the two sets of procedures and criteria with a view to creating a single set supporting the goals of both sources. The process and governance structures for core certification were aligned. The coverage, content and wording of the requirements were reviewed and revised and the resultant common requirements and procedures were run through a test bed process involving current WDS members and DSA recipients. These results were published, and their findings integrated into a second revision of the working group outputs. Process and requirements, including an introduction on the benefits of certification, background information, guidance text, and a glossary were made publicly available on the RDA website for comment during the lifetime of the working group.

With an agreed set of procedural and criteria references, the WDS and DSA began negotiations to merge the Data Seal of Approval and the WDS membership process into a single independent entity. An alignment plan was developed to define the initial governance entities and relationships of both parties. A joint position on the future of core certification was agreed upon, including short term alignment with cooperative, parallel activities, and longer-term alignment towards a single entity.

At this point the WDS Board and DSA Board[18] began applying the common requirements to new applications and renewals.

Figure 1: DSA timeline.



CoreTrustSeal Current Activities

Community-based non-profit organization

Key factors in the alignment plan were the formation of an interim board[19], the development of a common branding plan, board statutes and business plan, as well as the creation of additional guidance to support reviewer training including an online tool to support the application and administration processes.

On the 11th of September 2017 the new CoreTrustSeal organisation was announced as a "community-based non-profit organization promoting sustainable and trustworthy data infrastructures, [] governed by a Standards and Certification Board consisting of members drawn from the Assembly of Reviewers (by election) and the wider repositories stakeholders (appointed)".

From 2018, the CoreTrustSeal is a legal foundation entity under Dutch law governed by a Standards and Certification Board composed of 12 elected members representing the Assembly of Reviewers. The first Board elections were held in July 2018 and the Board 2018-2021 consists of:

Directors

- Chair—Jonas Recker (GESIS-Leibniz Institute for the Social Sciences, Germany)
- Vice-chair—Hervé L'Hours (UK Data Archive, United Kingdom)
- Secretary—Mari Kleemola (Finnish Social Science Data Archive, Finland)
- Treasurer—Ingrid Dillo (Data Archiving and Networked Services, The Netherlands)

Members

- Jonathan Crabtree (Odum Institute Data Archive, USA)
- Robert R. Downs (CIESIN-SEDAC, University of Columbia, USA)
- John Faundeen (USGS EROS Centre, USA)
- Wim Hugo (South African Environmental Observation Network, South Africa)
- Reyna Jenkins (Ocean Networks Canada)
- Dawei Lin (ImmPort Repository, DAIT-NIAID-NIH, USA)
- Mustapha Mokrane (World Data System, France)
- Paul Trilsbeek (Max Planck Institute for Psycholinguistics, The Netherlands)

Ex officio

- Rorie Edmunds (World Data System)
- Ilona von Stein (Data Archiving and Networked Systems)

The range of existing WDS, DSA and DSA-WDS certifications have been integrated into the application and certification process of the CoreTrustSeal as part of the certification renewal process. Within the ever growing CoreTrustSeal community new volunteers are being sought on an ongoing basis for the pool of reviewers.

An introduction to the CoreTrustSeal, the 16 requirements, extended guidance and a supporting glossary are all available[20].

CoreTrustSeal Application Process in Brief

Organisations with data expertise for a defined collection may seek certification against the 16 CoreTrustSeal requirements. The first step is to create an account in the CoreTrustSeal Application Management Tool. The actual application process begins when the applicant submits their self-assessment, i.e. their statements for each requirement via the Tool. Each self-assessment statement against each requirement needs to be supported by evidence. The CoreTrustSeal Board then assigns two independent peer reviewers taken from the community of CoreTrustSeal holders. By undertaking this responsibility peer reviewers become eligible for election to the CoreTrustSeal Board. Members of the CoreTrustSeal community are asked to volunteer to join the reviewer pool.

The comments and feedback from the two peer reviewers are assessed by the Board and a CoreTrustSeal is either granted for a period of three years, or the application is returned to the applicant for further work. The self-assessments and reviewers' final comments are published online once the CoreTrustSeal is awarded. Each applicant pays a fee of 1000 Euros to cover the cost of the operation, maintenance and development of the certification service.

Certification Fee

One of the drivers behind the effort to create a single core certification was an acknowledgement of the human and financial investment required to deliver such services.

The provision of a fee for CoreTrustSeal certification is a key element of the business model selected to ensure sustainability of the requirements, procedure and the service. The Board understood that any cost implication would present an issue for some members of the community. In order to ensure longevity and confidence in the criteria, however, an approach was required which went beyond a reliance on periodic project funding and in-kind contributions from participating organisations.

The Board has been careful to communicate[21] that the fee is for administrative purposes and is in line with the not-for-profit foundation status of CoreTrustSeal. Since the certification is valid for three years, the cost averages approximately 300 euros per year of certification. The fee-based model ensures that the cost of operation can be met for the maintenance of the standard, supporting procedures, associated tools, ongoing training of reviewers and engagement with the community.

The cost of core certification activities remains entirely met by the volunteer efforts of those joining the pool of reviewers including those elected and appointed to the Board. Such activities include the dual peer review of self-assessment statements and associated supporting evidence, followed by review, feedback and decisions from the Board.

Overview of the CoreTrustSeal Requirements

CoreTrustSeal take a 'whole organisation' perspective to reviewing data repositories. It starts off by asking for contextualizing background information and then focuses on the organisational infrastructure (mission, licences, continuity of access, sustainability, confidentiality/ethics, skills and guidance), digital object management (integrity, authenticity, appraisal, storage, preservation, quality, workflows, discovery, identifiers, re-use) and technology (technical infrastructure and security).

The 16 CoreTrustSeal requirements[22]:

1. The repository has an explicit mission to provide access to, and preserve data, in its domain.
2. The repository maintains all applicable licenses covering data access and use and monitors compliance.
3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.
4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.
5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.
6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse, or external, including scientific guidance, if relevant).
7. The repository guarantees the integrity and authenticity of the data.
8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.
9. The repository applies documented processes and procedures in managing archival storage of the data.
10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

11. 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.
12. Archiving takes place according to defined workflows from ingest to dissemination.
13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.
14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.
15. 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.
16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Guidance for Self-Assessment and Review

As the CoreTrustSeal has evolved there has been an increased need to provide more elaborate training and guidance for new reviewers to ensure consistency across the applications. In 2017, supported by a small project financed by the Research Data Alliance (RDA), extended guidance was created by the CoreTrustSeal Board.

The extended guidance gives a more detailed perspective on what reviewers should expect as evidence against each of the requirements. As the reviewer pool grows and more repositories are certified, the reviewers' experiences and observations are captured by the CoreTrustSeal Board to ensure the extended guidance remains in line with the latest developments. As a publicly available document, the extended guidance is also a useful tool for applicants[23]

Benefits of Certification

Over time CoreTrustSeal and other certification efforts have sought to demonstrate the benefits of certification to a variety of stakeholders. For TDR these lie not only in the certification itself but in the process by which different repository actors communicate, share and document their knowledge as they prepare and manage the relevant evidence.

In a review of their own Data Seal of Approval process, the Finnish Social Science Data Archive[24] noted:

> The use of models and metrics to assess our procedures and policies have raised our awareness about the challenges of digital preservation, revealed existing and possible problems and weaknesses as well as strengths, steered and initiated minor and major changes in our operations, and resulted in improved documentation. As a consequence, many of our processes are now better and more efficient or, they will be better – some of the bigger changes will take time to implement. We are also able to better manage risks, provide more trustworthy services for the research community, and demonstrate FSD's trustworthiness to our stakeholders. (Kleemola 2015.)

During the years Data Seal of Approval and World Data System have been conducting their certifications, user experiences have been collected via case studies, presentations at conferences and/ or engagement with the community. Furthermore the Dutch Network Digital Heritage (NDE) conducted a survey on the benefits of the Data Seal of Approval (Waterman and Sierman, 2016). The following conclusions/benefits can be drawn from their experiences and remain valid for the CoreTrustSeal.

- Performing a self-assessment does not take much time; on average, two to four days. It mainly depends on the level of existing documentation and its disclosure.
- Although most documentation is intended to be publicly accessible, an exception can be made for documentation containing privacy-sensitive and confidential information, such as a long-term vision.
- The certification process is very useful as an evaluation of internal procedures, which can be reviewed and updated where necessary. The current state of affairs, which can also serve for future accreditation, is made visible. Additionally, the procedures and documentation are evaluated, tested and approved by an external professional and the CoreTrustSeal is very helpful in determining strengths and weaknesses.
- The CoreTrustSeal reaffirms the necessity and usefulness of succession/long-term planning and helps to get these issues higher on the agenda of management.
- The CoreTrustSeal contributes to a reliable image. It can be used to improve reputation, but also as a benchmark for comparison. It clarifies what constitutes a digital repository and its business, and it creates transparency for the community in the area of sustainability.
- The CoreTrustSeal increases the confidence of users: it shows that standards are being used, just like the ones being used by traditional museums or repositories.
- The CoreTrustSeal helps to build a community: 'We' all work according to the same standards.
- The CoreTrustSeal emphasizes the need to conform towards the OAIS standards.
- Interaction with the peer reviewer is perceived as significant.
- The requirements are sufficiently generic to be applied to scientific data as well as publications.
- Because of its general approach the CoreTrustSeal is perceived as a less 'threatening', detailed and time-consuming procedure than more comprehensive standards, such as ISO or TRAC. The focus is on increasing awareness and transparency; CoreTrustSeal takes a community's and peer reviewers point of view rather than a top-down approach.
- The CoreTrustSeal is a solid foundation for applying for DIN 31644 certification.
- By renewing the CoreTrustSeal, the data repository will show its progress.
  (Waterman and Sierman, 2016)

A recent study by Donaldson et al (2017) validates these claims. Their findings demonstrate that DSA certification has allowed the repositories to:

- Build stakeholder confidence of their stakeholders in them,
- Improve their documentation,
- Gain assurance that they are following best practice,
- Demonstrate their transparency,

- Improve their processes,
- Raise awareness about the importance of digital preservation,
- Spend less time on audit and certification as compared to audit and certification through other programs,
- Improve communication among staff members, and
- Join a community of repositories who have demonstrated their commitment to digital preservation and following best practice."
  Donaldson et al (2017)

Future Plans

Certification as a Service for Complex Partnerships

Though neither the OAIS model nor the extant TDR standards preclude the idea of a TDR being a complex partnership of organisations, or a mixture of in-house and third party resource, neither do they address it in detail. The CoreTrustSeal acknowledges the rapid expansion of data management partnerships by asking for context about the organisation structure and acknowledges the increased provision of third party services by asking about outsourcing. Both of these questions speak to the scope of the repository in terms of control of and responsibility for data and acknowledge the possibility of shared responsibility for the CoreTrustSeal requirements. In an ideal world, each outsource partnership would be to a similarly certified trustworthy entity, but not all partners will seek such certification and in some cases appropriate certification may not yet exist.

Complex partnerships provide a challenge for applying the CoreTrustSeal. While outsourcing may provide cost savings and access to systems, services and expertise at scale not otherwise available to the applicant, it also introduces a more complex range of relationships and dependencies which can increase bureaucracy and risk. The CoreTrustSeal Board is actively investigating how best to define the acceptable scope of outsourcing including which requirements it might apply to and the level of control (and supporting evidence) applicants must provide to support assurances of trustworthiness. The provision of clear service level agreements is one key element of trust in complex partnership and outsourcing models but the CoreTrustSeal must keep up with the evolving nature and technical realities of modern repositories.

Widening the Evidence and Certification Community

The heritage of TDR standards leads to an inevitable focus on OAIS-defined repositories undertaking active data preservation by data/disciplinary experts for Designated Community having a defined knowledge base. Not all data assets are held in such repositories, however, and the CoreTrustSeal is increasingly receiving queries about how more general purpose repositories or institutional repositories with a broad disciplinary remit might be better supported. Many data assets may be stored and managed in such environments for some part of their lifecycle, thus forming a portion of the data provenance critical to ensuring an unbroken chain of trust from data creation/collection to use.

A general purpose institutional repository with appropriate disciplinary expertise to define and support preservation for a Designated Community could apply for the CoreTrustSeal but would need

to provide evidence related to a particular part of the data collection. The notion of providing clearer collection profiles to support better certification is already on the CoreTrustSeal radar as repositories undertake a range of curation levels (from storage of untouched deposited data to active participation in data quality improvement) and a range of responsibilities for digital objects (from harvesting metadata for resource discovery to active training and management of access to sensitive data through secure remote environments or safe rooms). The challenge is to set clear, common criteria for defining the data collections while retaining the 'core', low barrier to entry mission of CoreTrustSeal.

From a full data lifecycle perspective, there is a strong relationship between different repository types and the increased tendency for complex partnerships and outsourcing. Both present challenges for our trust in data across a range of data stewards over time. Repository host institutions, providers of metadata entry systems, data storage, and discovery and access systems may all contribute (through paid and unpaid relationships) to the overall infrastructure of people, processes and technologies necessary to ensure valuable digital assets are maintained. One potential solution is for CoreTrustSeal to engage with a wider variety of actors, including product and service vendors, to identify how they could support their clients with standardised evidence to support one or more aspects of the CoreTrustSeal requirements. This would lower the barrier to entry of certification; support standardised, transparent evidence of practice; and provide an additional 'ready for TDR' incentive to potential partners of participating product and service providers.

Conclusions

The CoreTrustSeal has grown from two complementary approaches to a single set of guidelines ensuring that data repositories can be trusted as stewards for the long term. It has grown and adapted to changing circumstances and continues to do so.

The notion of a 'Trustworthy Digital Repository' stems from the need to move beyond de facto trust in partner organisations to act as responsible stewards of data, towards de jure assertions of their trustworthiness. Standardisation, audit and certification are partly the kind of natural progression towards professionalization experienced by all mature service models. Certification also provides clear labelling of trustworthy 'nodes' in the (research) data lifecycle where outsourcing, third party relationships, and complex partnerships make the overall technical and human infrastructure of repository services more opaque.

Despite a membership and history which is predominantly academic, the CoreTrustSeal aspires to a generalised assurance of data preservation across disciplinary and specialist boundaries to ensure that digital data remain accessible to, and understandable by, those interested in seeking to use them for analysis, policy or profit.

The notion of trust is critical across the data lifecycle. The 'repository' or 'archive' model has historically defined itself as a distinct part of the lifecycle, but increasingly some repository standards and best practice have been adopted into more general research data management guidance.

Organisations which consider themselves as repositories are increasingly 'full lifecycle' actors as they are engaged with data producers pre-deposit and with researchers during the data use phase.

The deluge of data is a defining change to society and the CoreTrustSeal acknowledges these changes with a broad remit for certification of trustworthy digital repositories.

CoreTrustSeal and the CoreTrustSeal community are growing and thriving. The DSA-WDS collaboration and aligning of the two certification procedures has proved successful and the CoreTrustSeal has become an independent certification organisation that supports a variety of repositories. Today (2018), over 130 Seals have been awarded and more are in process. Certification standards like the CoreTrustSeal are also playing their part in the European Open Science Cloud (EOSC)[25] and collaborative research infrastructure developments. In addition, CoreTrustSeal certification is instrumental in helping data repositories adhere to the FAIR principles[26] and CoreTrustSeal is currently working through the EU ICT Standardisation[27] process which will permit it to be referenced for procurement purposes.

The formalisation of the CoreTrustSeal requirements, processes and governing bodies must sit alongside a flexible and responsive community-driven approach to change, if it is to continue to adapt to the rapidly evolving needs of the research data community and their collaborative partners.

Sources

[All links accessed 18 September 2018.]

Donaldson, Dillo, Downs and Ramdeen (2017). The perceived value of acquiring data seals of approval. http://dx.doi.org/10.2218/ijdc.v12i1.481

DSA (2010). Quality guidelines for digital research data ("DSA Booklet"). https://assessment.datasealofapproval.org/sitemedia/files/DSA_booklets/DSA-booklet_2010.pdf

Giaretta, David (2011). Advanced digital preservation. Heidelberg: Springer-Verlag Berlin.

Harmsen, Henk (2008). Data seal of approval - assessment and review of the quality of operations for research data repositories. International Conference on Preservation of Digital Objects (iPRES 2008), 29-30 September 2008, London. http://www.bl.uk/ipres2008/presentations_day2/34_Harmsen.pdf

Harrower, Natalie & Cassidy, Kathryn (2017). Why storage is not preservation: a conversation, surrounded by conservation. DRI Blog. http://dri.ie/why-storage-not-preservation-conversation-surrounded-conservation

Hedstrom, Margaret (1998). Digital preservation: a time bomb for digital libraries. Computers and the humanities 31(3): 189-202. DOI 10.1023/A:1000676723815

Kleemola, Mari (2015). Improving the quality of digital preservation using metrics. IASSIST Quarterly 2015. www.iassistdata.org/sites/default/files/iqvol_39_2_kleemola.pdf

OAIS (2012). Reference model for an open archival information system (OAIS). The Consultative Committee for Space Data Systems (CCSDS). https://public.ccsds.org/pubs/650x0m2.pdf

Rickards et al. (2016a) Developments in the certification of data centres, services and repositories through an RDA/WDS/DSA partnership
http://www.vliz.be/imisdocs/publications/296599.pdf#page=15

Rickards, Lesley; Mary Vardigan; Ingrid Dillo; Françoise Genova; Hervé L'Hours; Jean-Bernard Minster; Rorie Edmunds; Mustapha Mokrane (2016b). DSA–WDS partnership: streamlining the landscape of data repository certification. SciDataCon 2016, Denver, CO., 2016, 11–13 September 2016 (Session Auditing of Trustworthy Data Repositories). https://doi.org/10.5281/zenodo.252417

Schumann, Natascha (2012). Tried and trusted. Experiences with certification processes at the GESIS data archive. Iassist Quarterly Fall Winter 2012, 23-27.
http://www.iassistdata.org/sites/default/files/iqvol36_34_schumann.pdf

Waterman, Kees, & Sierman, Barbara. (2016). Survey on DSA-certified digital repositories. Report on the findings in a survey of all DSA-certified digital repositories on investments in and benefits of acquiring the Data Seal of Approval (DSA). https://doi.org/10.5281/zenodo.1188256

Wilkinson, Mark et al. (2016). The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3, Article number 160018. https://doi.org/10.1038/sdata.2016.18

---

[1] Hervé L'Hours, UK Data Archive, UK Data Service, University of Essex, United Kingdom; Mari Kleemola, Finnish Social Science Data Archive, University of Tampere, Finland; Lisa de Leeuw, Data Archiving and Networked Services, Netherlands.

[2] CoreTrustSeal website: https://www.coretrustseal.org/

[3] Reference Model for an Open Archival Information System.
https://public.ccsds.org/pubs/650x0m2.pdf

[4] ISO 14721:2012 (CCSDS 650.0-P-1.1). https://www.iso.org/standard/57284.html

[5] Trustworthy Repositories Audit & Certification: Criteria and Checklist.
http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories

[6] ISO 16363:2012 (CCSDS 652.0-R-1). https://www.iso.org/standard/56510.html

[7] Zertifizierung, *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive*. http://dx.doi.org/10.18452/1523

[8] DRAMBORA. http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora

[9] Wittenburg, P., Broeder, D., Klein, W., Levinson, S. C., & Romary, L. (2006). Foundations of modern language resource archives. In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006) (pp. 625-628). http://www.mpi.nl/publications/escidoc-58934

[10] Stewardship of Digital Research Data: A Framework of Principles and Guidelines (2008). http://www.rin.ac.uk/system/files/attachments/Stewardship-data-guidelines.pdf

[11] https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric-landscape

[12] Consortium of European Social Science Data Archives CESSDA. https://www.cessda.eu/

[13] European Research Infrastructure for Language Resources and Technology CLARIN. https://www.clarin.eu/

[14] European Research Infrastructure Consortium for the Arts and Humanities DARIAH. https://www.dariah.eu/

[15] MoU to support working on standards for Trusted Digital Repositories. http://www.trusteddigitalrepository.eu/Memorandum%20of%20Understanding.html

[16] RDA/WDS Certification of Digital Repositories IG. https://www.rd-alliance.org/groups/rdawds-certification-digital-repositories-ig.html

[17] Repository Audit and Certification DSA–WDS Partnership WG. https://rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg.html

[18] The members of these Boards were: CIESIN-SEDAC (Alex de Sherbinin), CINES (Marion Massol), DANS (Ingrid Dillo, Lisa de Leeuw), Finnish Social Science Data Archive (Mari Kleemola), GESIS-Leibniz Institute for the Social Sciences (Natascha Schumann), Institute of Remote Sensing and Digital Earth, China (Guoqing LI), International Service of Geomagnetic Indices (Aude Chambodut), MPI (Paul Trilsbeek), South African Environmental Observation Network (Wim Hugo), Strasbourg Astronomical Data Center (Françoise Genova), UKDA (Hervé L'Hours), University College Dublin (John Howard), WDC Geomagnetism, University of Kyoto (Toshihiko Iyemori), WDS-IPO (Mustapha Mokrane, Rorie Edmunds), World Glacier Monitoring Service, University of Zurich (Isabelle Gartner-Roer).

[19] The interim CoreTrustSeal board consisted of the following members: CIESIN-SEDAC, University of Columbia (Robert R. Downs), DANS (Ingrid Dillo and Ilona von Stein), Finnish Data Archive (Mari Kleemola), GESIS-Leibniz Institute for the Social Sciences (Jonas Recker), MPI (Paul Trilsbeek), Ocean Networks Canada (Reyna Jenkyns), South African Environmental Observation Network (Wim Hugo), UKDA (Hervé L'Hours), University College Dublin (John Howard), USGS EROS Centre (John Faundeen), WDC Geomagnetism, University of Kyoto (Toshihiko Iyemori), WDS-IPO (Mustapha Mokrane, Rorie Edmunds).

[20] CoreTrustSeal Data Repositories Requirements. https://www.coretrustseal.org/why-certification/requirements/

[21] CoreTrustSeal Administrative fee. https://www.coretrustseal.org/apply/administrative-fee/

[22] https://www.coretrustseal.org/why-certification/requirements/

[23] CoreTrustSeal Extended Guidance. https://www.coretrustseal.org/wp-content/uploads/2017/01/20171026-CTS-Extended-Guidance-v1.0.pdf

[24] FSD: http://www.fsd.uta.fi/en/

[25] The European Open Science Cloud. https://eoscpilot.eu/eosc

[26] FORCE11. The FAIR Principles. https://www.force11.org/group/fairgroup/fairprinciples

[27] European Comisssion. ICT standardisation. https://ec.europa.eu/growth/industry/policy/ict-standardisation_en