CORETRUSTSEAL

by Ingrid Dillo & Lisa de Leeuw

Abstract: *Open data and data management policies that call for the long-term storage and accessibility of data are becoming more and more commonplace in the research community. With it the need for trustworthy data repositories to store and disseminate data is growing. CoreTrustSeal, a community based and non-profit organisation, offers data repositories a core level certification based on the DSA-WDS Core Trustworthy Data Repositories Requirements catalogue and procedures. This universal catalogue of requirements reflects the core characteristics of trustworthy data repositories. Core certification involves an uncomplicated process whereby data repositories supply evidence that they are sustainable and trustworthy. A repository first conducts an internal self-assessment, which is then reviewed by community peers. Once the self-assessment is found adequate the CoreTrustSeal board certifies the repository with a CoreTrustSeal. The Seal is valid for a period of three years. Being a certified repository has several external and internal benefits. It for instance improves the quality and transparency of internal processes, increases awareness of and compliance with established standards, builds stakeholder confidence, enhances the reputation of the repository, and demonstrates that the repository is following good practices. It is also offering a benchmark for comparison and helps to determine the strengths and weaknesses of a repository. In the future we foresee a larger uptake through different domains, not in the least because within the European Open Science Cloud, the FAIR principles and therefore also the certification of trustworthy digital repositories holding data is becoming increasingly important. Next to that the CoreTrustSeal requirements will most probably become a European Technical standard which can be used in procurement (under review by the European Commission).*

Keywords: *CoreTrustSeal; certification; repository; trusted digital repository; requirements; standard; policy; process; service; self-assessment; review; digital preservation*

DOI: *https://doi.org/10.31263/voebm.v71i1.1981*

Content

1. Introduction
2. Organisation
3. Requirements and process
4. Experiences of applicant organizations and benefits
5. Future developments

1. Introduction

If we want to share data, the long-term storage of this data in a trustworthy digital repository is a sine qua non. Data created and used by scientists should be managed, curated and archived in order to preserve the initial investment in collecting them. Researchers must be certain that the data provided by the repositories remain useful and meaningful, even in the long term. In addition, the repositories should have sustainable business models themselves.

The concept of sustainability involves a variety of challenging aspects in numerous areas: organizational, technical, financial, legal, etc. Certification can contribute significantly towards ensuring the reliability and durability of digital repositories and hence the possibilities for sharing data over a long period.

Why is increasing importance being attached to research data? First, sharing data makes science more transparent. It facilitates replication and validation of research. This will enhance research quality. Another benefit of sharing data is the possibility of reuse of data by researchers who did not generate those data themselves. This reuse will lead to greater efficiency in research. It offers researchers the ability to combine datasets and use them across disciplines. Furthermore, open data can also be used for economic and social interests beyond science. Eventually, data sharing will lead to a higher return on the initial investment. However, sharing data is still threatening to some researchers. As producers of the data they have a strong sense of ownership and do not want to share for that reason. Sometimes there is fear that their findings will be discredited by others. Another argument is that data generated elsewhere might not be reliable. These arguments have everything to do with trust. To refute this argument, we must ensure that we build the element of trust into the digital archiving services.

Trust is the basis for storing and sharing data. That trust must be present in various stakeholders. The data depositors want the assurance that their data in the digital repository are safe and will remain accessible, us-

able and meaningful. Data users have questions like: have the data been well kept, have they retained their authenticity and integrity, are the data of good quality, do the identifiers refer to the appropriate objects? The funders have other concerns. They want to be certain that their investment in data production yields optimum returns, i.e. that the data will remain available for long-term reuse.

What characteristics make digital repositories reliable? First, a digital repository's mission should be to give reliable long-term access to the digital data under their care, now and in the future. Second, there should be permanent monitoring, planning and maintenance. The threats and risks within their systems must be understood. Finally, there should be a regular audit and certification cycle in place. Reliability is not something you achieve once and can then take for granted.

Certification can make an important contribution to the confidence of various stakeholders. The CoreTrustSeal[1] has sixteen guidelines for data repositories and enables basic certification.

2. Organisation

Since September 2017 CoreTrustSeal offers data repositories a core level certification based on the DSA-WDS Core Trustworthy Data Repositories Requirements catalogue and procedures[2]. This universal catalogue of requirements reflects the core characteristics of trustworthy data repositories and is the culmination of a cooperative effort between DSA and WDS under the umbrella of the Research Data Alliance[3] to merge their data repositories certifications. CoreTrustSeal Data Repository certification replaces the DSA certification (founded in 2009) and WDS Regular Members certification (started in 2011).

CoreTrustSeal is a community based non-profit organization promoting sustainable and trustworthy data infrastructures. It is driven by the commitment to offer professional certification tools and services to data repositories and to support volunteer qualified reviewers in conducting audits in/under optimum conditions.

CoreTrustSeal is governed by a Standards and Certification Board[4]. After the elections, which will be organised late 2018, the board will be composed of elected members representing the Assembly of Reviewers[5] and appointed members representing the wider data repositories stakeholders. For the moment the founding Board established by WDS and DSA is assuming its responsibilities. After the elections the board will appoint an

Advisory Committee to assist the CoreTrustSeal Board in fulfilling its mission and objectives.

The CoreTrustSeal certification of trustworthy data repositories is envisioned as the first step in a global framework for repository certification which includes the extended level certification (nestor-Seal DIN 31644[6]) and the formal level certification (ISO 16363[7]). Ultimately, CoreTrustSeal will also endeavor to provide core level certification for other research entities such as data services and software.

3. Requirements and process

The objectives of the CoreTrustSeal are to safeguard data, to ensure high quality and to guide reliable management of data for the future without requiring the implementation of new standards, regulations or heavy investments.

CoreTrustSeal repository certification:

- Gives data producers the assurance that their data and associated materials will be stored in a reliable manner and can be reused;
- Provides funding bodies with the confidence that data will remain available for reuse;
- Enables data consumers to assess the repositories where data are held;
- Supports data repositories in the efficient archiving and distribution of data.

The CoreTrustSeal requirements

The CoreTrustSeal contains 16 requirements[8] for the application and verification of quality aspects with regard to the creation, storage and (re-) use of digital data. They have been designed with a focus on scientific and scholarly materials but may be applied to all types of digital information. These guidelines serve as a basis for granting a 'CoreTrustSeal' by the CoreTrustSeal Board.

The criteria for awarding the CoreTrustSeal to data repositories are in accordance with national and international guidelines for digital data archiving such as the Kriterienkatalog vertrauenswürdige digitale Langzeitarchive developed by NESTOR[9], the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)[10] published by the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE), and Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist of

the Research Library Group (RLG)[11]. The following publications have also been taken into account: Foundations of Modern Language Resource Archives by the Max Planck Institute[12], and Stewardship of Digital Research Data: A Framework of Principles and Guidelines[13] by the Research Information Network. The CoreTrustSeal requirements can be seen as a minimum set distilled from the above proposals.

Fundamental to the requirements are five criteria that together determine whether or not the digital data may be considered as sustainably archived:

- The data can be found on the Internet.
- The data are accessible, while taking into account relevant legislation with regard to personal information and intellectual property.
- The data are available in a usable format.
- The data are reliable.
- The data can be referred to (persistent identifiers).

The above criteria have a strong connection to the FAIR Guiding Principles for scientific data management and stewardship[14] which aim to make data Findable, Accessible, Interoperable, and Re-usable.

Self-assessment and peer-review process

Starting point for obtaining the CoreTrustSeal is the website https://www.coretrustseal.org/, where an application form can be submitted via an online tool. The self-assessment is meant to supply evidence that the applicant repository adheres to the 16 requirements and meets the relevant compliance levels. Next to that some background information on the context of the repository is also required.

Since the CoreTrustSeal is used in an international environment, the language of communication is English in order to increase transparency.

After submission of a self-assessment by a repository and the payment of an administrative fee[15], the CoreTrustSeal Board appoints two peer reviewers to evaluate the self-assessment. The evaluation takes a maximum of two months, after which the peer reviewers will either confirm the evidence or require additional information depending on the adherence to the requirements and compliance levels. Resubmission of the modified application and requests for additional information by the reviewers will continue until they are satisfied with the evidence and award the CoreTrustSeal, but note that a maximum of 5 revised self-assessments can be reviewed as part of the same paid application leading to a certification

valid for 3 years. Additional revisions can be accepted at the discretion of the CoreTrustSeal Board. In the event of a dispute, the applicant repository can contact the CoreTrustSeal Board.

During the application process, a self-assessment will not be made public. The self-assessment, including all evidence, will only be published on the websites of the CoreTrustSeal and the applicant repository after the CoreTrustSeal has been awarded.

Because applications awarded with a CoreTrustSeal, including all evidence and peer review comments, are publicly available on the website[16], they can be used as references or samples.

After the CoreTrustSeal is awarded by the Board, the CoreTrustSeal logo may be displayed on the repository's website only. The Board will provide a widget with a link to the approved assessment and logo for implementation on the applicant's website. At the same time, the Board will post the approved assessment of the new repository on the website, using the name of the specific repository.

A CoreTrustSeal for a given period cannot be displayed indefinitely but will need to be updated periodically if the repository wants to stay compliant to the CoreTrustSeal requirements. CoreTrustSeal-certified repositories will be contacted automatically when an update is imminent.

4. Experiences of applicant organizations and benefits

During the years Data Seal of Approval (DSA) and World Data System (WDS) have been conducting their certifications a lot of user experience has been collected either via case studies[17] [18], presentations at conferences and/ or engagement with the community. Furthermore, the Dutch Network Digital Heritage (NDE)[19] have conducted a survey[20] on the benefits of the Data Seal of Approval. The following conclusions/benefits can be drawn from their experiences and remain valid for the CoreTrustSeal.

Benefits

- 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[21].
- Interaction with the peer reviewer is perceived as significant when working on the application's supporting evidence.
- 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 reviewer's 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.

5. Future developments

CoreTrustSeal is doing well. The community is growing and thriving. Since the launch of the CoreTrustSeal in September 2017, 22 CoreTrustSeals

have been awarded in the following domains: generic (4), Social Sciences and Humanities (15), Earth Sciences (1) and Life Sciences (2). We expect that most of the 110 legacy Seals, awarded either by DSA, WDS or both, will renew their certification to stay compliant with the latest version of the CoreTrustSeal requirements. Which will be accompanied by many new applications from the international repository community either because they are part of infrastructures that use the CoreTrustSeal or want to show their compliance in regards to the FAIR principles.

The added value of the CoreTrustSeal process is not only recognized by individual repositories. Within the European research infrastructures, building confidence in the services offered is considered increasingly important. In this context infrastructures such as CESSDA[22], CLARIN[23] and DARIAH[24] are using the CoreTrustSeal requirements. CLARIN has even made CoreTrustSeal certification mandatory for a large part of its centers. CESSDA is working to integrate the CoreTrustSeal requirements with their own infrastructure and DARIAH is using the requirements in their assessment of national contributions to the infrastructure.

Certification standards like the CoreTrustSeal are also playing their part in the European Open Science Cloud (EOSC)[25]. CoreTrustSeal certification could very well be used as a rule of engagement for the participation of data repositories in the EOSC. Next to that CoreTrustSeal certification is instrumental in helping data repositories adhere to the FAIR principles[26].

Furthermore, the CoreTrustSeal requirements are under review by the European Commission at the moment (April 2018), as the CoreTrustSeal requirements will most probably become a European Technical standard[27] which can be used in procurement.

Ingrid Dillo Ph.D.
ORCID: http://orcid.org/0000-0001-5654-2392
DANS – Data Archiving and Networked Services
E-Mail: ingrid.dillo@dans.knaw.nl

Lisa de Leeuw
DANS – Data Archiving and Networked Services
E-Mail: lisa.de.leeuw@dans.knaw.nl

* All links accessed April 6, 2018.

1 https://www.coretrustseal.org/
2 https://www.coretrustseal.org/why-certification/requirements/
3 https://rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg.html
4 https://www.coretrustseal.org/about/standards-and-certification-board/
5 https://www.coretrustseal.org/about/assembly-of-reviewers/
6 https://www.din.de/en/getting-involved/standards-committees/nid/wdc-beuth:din21:147058907
7 https://www.iso.org/standard/56510.html
8 https://www.coretrustseal.org/why-certification/requirements/
9 https://edoc.hu-berlin.de/handle/18452/2175
10 http://www.repositoryaudit.eu/about/
11 https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac
12 http://pubman.mpdl.mpg.de/pubman/item/escidoc:58934:4/component/escidoc:58935/Wittenburg_2006_foundations.pdf
13 http://www.rin.ac.uk/system/files/attachments/Stewardship-data-guidelines.pdf
14 https://doi.org/10.1038/sdata.2016.18
15 https://www.coretrustseal.org/apply/administrative-fee/
16 https://www.coretrustseal.org/why-certification/certified-repositories/
17 http://www.dcc.ac.uk/resources/case-studies/ads-dsa
18 http://www.iassistdata.org/sites/default/files/vol_40-3_6_13.pdf
19 http://digitalpreservation.nl/seeds/the-benefits-of-certification/
20 http://www.ncdd.nl/wp-content/uploads/2016/10/201611_DE_Houdbaar_Report_DSA-survey_2016.pdf
21 https://public.ccsds.org/pubs/650x0m2.pdf
22 https://www.cessda.eu/
23 https://www.clarin.eu/
24 https://www.dariah.eu/
25 https://eoscpilot.eu/eosc
26 https://www.force11.org/group/fairgroup/fairprinciples
27 https://ec.europa.eu/europeaid/sectors/economic-growth/trade/technical-standards_en