Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782 Start Date of Project: 01/01/2019 Duration: 40 months

Deliverable 8.3

Trustworthy Digital Repository status update and certification solutions for SSHOC repositories

Dissemination Level	PU
Due Date of Deliverable	28/02/2022 (M38)
Actual Submission Date	15/03/2022
Work Package	WP8 - Governance/ Sustainability/ Quality Assurance
Task	Task 8.2 Trust & Quality Assurance, Impact
Type	Report
Approval Status	Approved by EC - 04 May 2022
Version	V1.0
Number of Pages	p.1 – p.46

Abstract:

The report provides a status update on the certification of repositories belonging to the SSHOC communities. In addition, the report considers certification solutions for SSH repositories and outlines suggestions for the sustainable management of trust beyond the lifetime of the project.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.

SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782

History

Version	Date	Reason	Revised by
0.0	02/11/2020	First draft structure (section 3)	Niko Koski (CESSDA ERIC/TAU-FSD)
0.1	14/01/2022	First draft content	Henri Ala-Lahti, Mari Kleemola (CESSDA ERIC/TAU-FSD)
0.2	28/01/2022	All content ready	All authors
0.3	02/02/2022	Version for peer review	Henri Ala-Lahti, Mari Kleemola, Tuomas J. Alaterä (CESSDA ERIC/TAU-FSD)
0.4	06/02/2022	Peer review	Franciska de Jong (CLARIN ERIC)
0.5	11/02/2022	Address peer review comments	Henri Ala-Lahti, Mari Kleemola (CESSDA ERIC/TAU-FSD), Ben Mathers, Hervé L'Hours (CESSDA ERIC/UKDS), Daan Broeder (CLARIN ERIC)
0.6	14/02/2022	Final review by WP leader	Emiliano Degl'Innocenti (CNR)
0.7	14/02/2022	Revision after WP leader review, submitted to Coordinator	Henri Ala-Lahti, Mari Kleemola (CESSDA ERIC/TAU-FSD)
1.0	28/02/2022	Final version for submission	Henri Ala-Lahti

Author List

Organisation	Name	Contact Information
CESSDA ERIC/TAU-FSD	Mari Kleemola	mari.kleemola@tuni.fi
CESSDA ERIC/TAU-FSD	Henri Ala-Lahti	henri.ala-lahti@tuni.fi
CESSDA ERIC/TAU-FSD	Tuomas Alaterä	tuomas.alatera@tuni.fi
CESSDA ERIC/UKDS	Hervé L'Hours	herve@essex.ac.uk
CESSDA ERIC/UKDS	Benjamin Jacob Mathers	bm21346@essex.ac.uk
CLARIN ERIC	Daan Broeder	d.g.broeder@uu.nl
KNAW/DANS	René van Horik	rene.van.horik@dans.knaw.nl
CESSDA ERIC/SND	Birger Jerlehag	birger.jerlehag@snd.gu.se

CNR	Emiliano Degl'Innocenti	emiliano.deglinnocenti@cnr.it
CNR	Maurizio Sanesi	maurizio.sanesi@cnr.it
CESSDA ERIC/TAU-FSD	Niko Koski	N/A

Executive Summary

Research data should be managed, curated, stored and shared in a way that lives up to the expectations regarding trustworthiness and quality, provides sustainability and preserves the investments. The Trustworthy Digital Repository standards which have emerged from the Open Archival Information System (OAIS) reference model offer a certification solution for repositories. CoreTrustSeal (CTS) offers baseline certification and supports the concept of outsourcing. Adopting workflows and guidelines from CoreTrustSeal is also a way to assure that the data published by the repository follow the FAIR principles. Even outside of the formal certification framework the CoreTrustSeal criteria provide a demonstrable approach to internal and external review, supporting a benchmark for comparison and a means to determine the strengths and weaknesses of data repositories.

This deliverable is the final deliverable of the SSHOC Task 8.2 Trust & Quality Assurance. It will describe the certification standards of Trustworthy Digital Repositories (TDRs) from the perspective of the SSH domain and summarise the certification support activities provided to the SSHOC community. The experiences gained from the support process will be considered in addition to the results of the examination of the trust in the domain of Social Sciences and Humanities (SSH) and the certification landscape. The support activities are based on the earlier work of T8.2 outlined in the *Deliverable 8.1 Certification plan for SSHOC repositories*, which laid the ground for the SSHOC trust work to facilitate the adoption of TDR standards and the FAIR principles in SSH data repositories. In this deliverable, 'trust' refers to the myriad of issues, standards and processes related to the level of trustworthiness of digital repositories.

The deliverable will also discuss possible certification solutions for SSH repositories, consider the complex partnership models of TDRs and outsourcing of their services, and examine how trust can be sustainably managed after the SSHOC project.

The experiences and feedback gained from the trust support work demonstrate that the support process has been beneficial for the repositories involved and allowed them to improve their procedures. While certification can be resource-intensive for certain repositories, there are few alternatives for a lighter certification beyond the core certification. The diversity of the SSH repository and data service landscape means that there is no certification solution suitable for all. Complex partnership models and outsourcing of services should also be considered when seeking certification. In some cases, organisations may opt for an assessment instead of formal certification. This has proven beneficial and useful for certain data services in improving their practices.

Ensuring the sustainable management of trust is not solely dependent on assessment or certification, as trust goes beyond the technical aspects of repositories and also involves people. Therefore, future endeavours to manage trust should make use of the existing and planned networks of trustworthy

repositories that can share both expertise and responsibility, while recognising the need for more enduring sources of funding for managing trust sustainably.

Abbreviations and Acronyms

CESSDA	Consortium of European Social Science Data Archives
CLARIN	Common Language Resources and Technology Infrastructure
CTS	CoreTrustSeal
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DIN	Deutsches Institut für Normung
EOSC	European Open Science Cloud
ERIC	European Research Infrastructure Consortium
E-RIHS	European Research Infrastructure for Heritage Science
ESS	European Social Survey
FAIR	Findable, Accessible, Interoperable, Reusable
FTE	Full-time equivalent
ISO	International Organization for Standardization
LAM	Libraries, archives and museums
OAIS	Open Archival Information System
RDA	Research Data Alliance
SHARE	Survey of Health, Ageing and Retirement in Europe
SSH	Social Sciences and Humanities
TDR	Trusted Digital Repository
TRUST	Transparency, Responsibility, User focus, Sustainability and Technology

Table of Contents

1. Introduction

The complex partnership models in the EOSC and in the SSH domain require trust in the quality of data and services among all parties: data repositories, stakeholders, data users and outsourcing partners. Research data should be managed, curated, stored and shared in a way that meets expectations regarding FAIR[1], trustworthiness and quality, provides sustainability and preserves the investments made to generate these 'digital assets'.

1.1 Purpose and scope

This report is the second and final deliverable of Task 8.2 of the Social Sciences and Humanities Open Cloud (SSHOC) project. The purpose of T8.2 is to bolster and improve trust in and quality of the repositories belonging to the SSHOC communities in two ways: firstly, by supporting the repositories within the SSHOC communities (or in short, "SSHOC repositories") in their work on trust and quality, and secondly, by exploring the trust landscape and providing feedback and input to certification bodies from the SSH viewpoint. The first deliverable[2] laid ground for the SSHOC trust work. During the course of the project, T8.2 has provided support to data repositories pursuing self-assessment and certification. This report delivers an overview of the outcomes of the support activities, the certification status of repositories, lessons learned, criteria by which trusted services might be outsourced within the Trustworthy Digital Repositories (TDR) model to support complex partnership models, and recommendations for the sustainable assurance of trust after the SSHOC project ends.

In this report, 'trust' refers widely to the landscape of issues, standards and processes related to TDR. The level of trustworthiness can be demonstrated through transparent evaluation of evidence of practice that meets agreed standards. The SSHOC trust work focuses on CoreTrustSeal as the TDR certification framework for repositories. The CoreTrustSeal framework distinguishes sixteen requirements that reflect the core characteristics of a TDR. Even outside of the formal certification framework, the CoreTrustSeal criteria provide a demonstrable approach to internal and external review, providing a means to determine the strengths and weaknesses of data repositories.

It is important to note that in the context of T8.2 and this report, the term 'quality' refers to the technical quality of the repositories, i.e., their compliance with TDR standards, not to the scientific quality of the digital assets managed by the repositories.

[1] FAIR = Findable, Accessible, Interoperable, Reusable

[2] Kleemola et al. *SSHOC D8.2 Certification plan for SSHOC repositories*. https://doi.org/10.5281/zenodo.3725868

1.2 Relation to other tasks and activities

In the SSHOC project, trust and quality issues are included in Work Package 8 which includes four tasks that complement each other: Task 8.1 on governance and sustainability, Task 8.2 on trust and quality, Task 8.3 on legal and ethical issues, and Task 8.4 on overarching clusters. Outside the SSHOC project, many initiatives and projects have included work on trusted repositories. Task 8.2 team has collaborated closely with two Horizon 2020 projects providing similar trust support to repositories: FAIRsFAIR WP4[3] and EOSC Nordic WP4[4]. Other major initiatives that the task team have followed are the projects aimed at building EOSC[5], and the various trust-related groups within the Research Data Alliance (RDA)[6], but describing them is beyond the scope of this report. An overview of activities being undertaken by projects involved in building a FAIR ecosystem for the European Open Science Cloud has been provided by the FAIRsFAIR Synchronisation Force.[7]

1.3 Structure of the document

This deliverable is organised into five sections:

- Section 2 outlines the certification of Trustworthy Digital Repositories.
- Section 3 describes the support activities and outcomes of the support provided by Task 8.2 as well as the results of the survey, desk research and stakeholder infrastructure questionnaire conducted to examine the trust landscape of organisations related to the SSHOC project.
- Section 4 provides a discussion on suggested evaluation and assessment solutions for SSHOC repositories and TDR partnership models and outsourcing
- Section 5 considers the sustainable management of trust beyond the timeframe of the SSHOC project.
- Section 6 presents the conclusions.

[3] FAIRsFAIR. FAIR Certification (of Repositories) – WP4: https://www.fairsfair.eu/fair-certification [31 January 2022]
[4] EOSC Nordic Organisation. WP4: FAIR data (NeIC): https://www.eosc-nordic.eu/organisation/ [31 January 2022]
[5] European Open Science Cloud. EOSC Projects: https://eosc-portal.eu/about/eosc-projects [31 January 2022]
[6] Research Data Alliance. Groups: https://www.rd-alliance.org/groups [31 January 2022]
[7] Grootveld et al., *D5.6 Report 3 of the Synchronisation Force* (V1.0_DRAFT). https://doi.org/10.5281/zenodo.5336658

2. Trustworthy Digital Repository certification

2.1 Certification based on the OAIS model

A wide variety of institutions or partnerships providing data storage and management, including galleries, archives, libraries, museums and records management systems, describe themselves as 'repositories'. The team involved in Task 8.2 broadly defines repositories as "organisations that preserve, manage, and provide access to digital research data in a variety of formats. A repository must have sufficient control and rights to ensure that the digital material is authentic, reliable, accessible and usable also for the long term."[8]

In the context of the SSH domain, assessment, including self-assessment, refers to processes intended to determine whether research data repositories meet specific quality assurance standards which focus on data management, data services and long term preservation of digital objects. The framework used for reference is the CoreTrustSeal Trustworthy Digital Repositories (TDR) certification that is based on the OAIS Reference Model (CCSDS 2012[9]; OAIS= Open Archival Information System). The justification for the selection of CoreTrustSeal as the TDR certification framework as well as motivations and benefits of CoreTrustSeal are discussed in more detail in D8.2.[10]

The mandatory responsibilities of an OAIS are (CCSDS 2012):

- Negotiate for and accept appropriate information from information Producers.
- 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 to be 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.
- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the Archive, ensuring that it is never deleted unless allowed as part of an approved strategy. There should be no ad-hoc deletions.

[8] Kleemola et al., *SSHOC D8.2 Certification plan for SSHOC repositories*. https://doi.org/10.5281/zenodo.3725868

[9] OAIS reference model is defined by recommendation CCSDS 650.0-B-2 of the Consultative Committee for Space Data Systems. The text is identical to ISO 14721:2012.

[10] Kleemola et al., *SSHOC D8.2 Certification plan for SSHOC repositories*. https://doi.org/10.5281/zenodo.3725868

- Make the preserved information available to the Designated Community and enable the information to be disseminated as copies of, or as traceable to, the original submitted Data Objects with evidence supporting its Authenticity.

A designated community is defined in the OAIS Reference Model as "[a]n identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities."[11]

2.2 CoreTrustSeal

CoreTrustSeal is a core-level certification framework for data repositories.[12] It is maintained by a community-based non-profit organisation. To obtain CoreTrustSeal certification, a repository conducts a self-assessment based on 16 CoreTrustSeal requirements which is then reviewed by two volunteer peers from the CoreTrustSeal Assembly of reviewers. The reviews are based on the evidence supplied publicly[13]. Passing the review will result in certification that is valid for three years and published on the CoreTrustSeal website. Providing a public and formal certification outcome of the assessment process means that the CoreTrustSeal increases confidence that the certified entity will be able to meet its obligations. In the data repository context, this essentially means that the data producers and research funders know that their data are preserved reliably and curated to enable reuse.

Even in case of a self-assessment without the intention of applying for certification immediately, CoreTrustSeal serves as a frame of reference which helps to identify the most obvious areas of development and strengths of a repository.

The CoreTrustSeal requirements are updated every three years through a community review process. The CoreTrustSeal has been widely adopted. There are 120 certified data repositories worldwide (February 2022).[14] Out of these, about half are from the SSH communities. CoreTrustSeal was established by a Research Data Alliance Working Group[15] led by, and leading to the replacement of, the Data Seal of

[11] CCSDS The Consultative Committee for Space Data Systems, *The Reference Model for an Open Archival Information System (OAIS). Recommended practice.* http://public.ccsds.org/publications/archive/650x0m2.pdf [31 January 2022]

[12] 'Core' refers to the minimum key criteria that all trustworthy repositories should meet as agreed through a community-developed standard. The goal is to be generally applicable and have a low barrier to entry. CoreTrustSeal website: https://www.coretrustseal.org/ [31 January 2022]

[13] Sensitive evidence can be shared directly with reviewers, but CoreTrustSeal prefers all evidence to be publicly available.

[14] CoreTrustSeal - Core Certified Repositories: https://www.coretrustseal.org/why-certification/certified-repositories/ [7 February 2022]

[15] Repository Audit and Certification DSA–WDS Partnership WG: https://www.rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg.html [31 January 2022]

Approval (DSA) and World Data System (WDS) certifications. A number of organisations still have these earlier certificates but are expected to migrate to the CoreTrustSeal.

2.3 Other repository certification standards

Repositories have seen a wide range of documented best practices, recommendations and standards (formal and de facto) developed to guide their work. The CoreTrustSeal is one of a number of related formal standards with associated assessment processes and certification outcomes that result in TDR status. There are two other formal

ISO 16363 defines the expectations of a trustworthy repository while ISO 16919: 2014 defines the necessary requirements for bodies providing audit and certification of TDRs.[16] Both are administered by PTAB (Primary Trustworthy Digital Repository Authorisation body).[17]

The nestorSeal grew out of DIN 31644[18], which is a national TDR standard developed by the German DIN Committee[19]. A key difference between nestorSeal, ISO 16363 and the CoreTrustSeal is the capability-maturity levels used during the self-assessment and formal assessment stages. Unlike the latter two TDR standards, each nestorSeal requirement is scored by using the classifications: "Not yet actioned"; "Planned"; "Planned in detail"; "Implemented".[20]

Further differences between the CoreTrustSeal, nestorSeal and ISO 16363 concern the degree of detail, and the specificity of the individual requirements that make up each standard. The CoreTrustSeal addresses the core functions necessary for a repository to function as a trusted digital repository, thus making it the most 'lightweight' of the three. The nestorSeal can be viewed as the next 'step-up' in terms of the complexity and specificity of the necessary requirements. Finally, ISO 16363 is the most complex and detail-specific of the three. CoreTrustSeal has 16 Requirements, NestorSeal has 34 'Criteria' and ISO16363 has 108 'metrics'.

After the development of these three formal TDR standards, the TRUST principles were established in 2020 by several stakeholders in the digital repository community.[21] TRUST stands for: Transparency,

[16] CCSDS The Consultative Committee for Space Data Systems, *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories. Recommended Practice*. https://public.ccsds.org/Pubs/652x1m2.pdf

[17] PTAB - Primary Trustworthy Digital Repository Authorisation Body Ltd: http://www.iso16363.org/ptab/ [31 January 2022]

[18] DIN 31644:2012-04: https://www.beuth.de/en/standard/din-31644/147058907 [31 January 2022]

[19] DIN, Standards Committees: https://www.din.de/en/getting-involved/standards-committees [31 January 2022]

[20] nestor, Explanatory notes on the nestor Seal for Trustworthy Digital Archives: English version: https://d-nb.info/1047613859/34 [31 January 2022]

[21] Lin et al. "The TRUST Principles for digital repositories". https://doi.org/10.1038/s41597-020-0486-7

Responsibility, User focus, Sustainability and Technology. The principles "…provide a common framework to facilitate discussion and implementation of best practice in digital preservation by all stakeholders."[22] They were developed in an attempt to provide alignment across the aforementioned TDR standards, albeit in very high-level terms.[23] The TRUST principles also provide a set of broad guidance statements that may be referred to by data services for which there is no formal standard, governing body, assessment process or certification solution at this time. Although the TRUST principles are not a formal standard, repositories can publicly endorse them.[24]

2.4 Certification plans for SSHOC repositories

In deliverable D8.2[25], the following three certification trails were established for the repositories that would become involved in the SSHOC support program:

- (A) renewal of existing CoreTrustSeal certification,
- (B) new, initial CoreTrustSeal certification, and
- (C) self-assessment using the CoreTrustSeal requirements.

In this deliverable, "SSHOC repositories" refers to the research data repositories within CESSDA ERIC, CLARIN ERIC, DARIAH ERIC and E-RIHS community nodes regardless of their participation in the SSHOC project. Two other ERICs participating in the SSHOC project, ESS and SHARE, are not included because they differ from the other four in their main focus, which is on conducting international surveys in partnership with various data organisations rather than representing the repositories of the same discipline.

Appendix 1 shows the repositories that were initially selected as the primary target of T8.2 monitoring and/or support activities. The ambitious goal of (A) renewal of pre-existing certification or (B) new certification was set for the repositories on the list, but in some cases the outcome was ultimately (C) self-assessment, since for some repositories full CoreTrustSeal certification was not feasible within the timeframe of the project (unlike initially presumed in D8.2). Likewise, even if the certification plan was set as (C) self-assessment, the task team may have recommended the repository to proceed to formal CoreTrustSeal certification. Repositories that were deemed not in scope of the CoreTrustSeal were offered support from SSHOC with the goal of improving practices. In return they provided input for the discussion on gaps in existing certification frameworks.

[22] Ibid.
[23] Ibid.
[24] The TRUST Principles: An RDA Community Effort: https://www.rd-alliance.org/rda-community-effort-trust-principles-digital-repositories
[25] Kleemola et al. *SSHOC D8.2 Certification plan for SSHOC repositories*. https://doi.org/10.5281/zenodo.3725868

The repositories that have attained the CoreTrustSeal certificate since are highlighted in Appendix 1 (situation as of February 2022). It should be noted that several applications were received from repositories not included in Appendix 1. The likely reasons for this are: a certification process being already underway or perceived redundancy of the support activities in the case of certification renewal (see section 3.4 for more details).

3. Certification support activities

The T8.2 support activities included three primary modes of support to SSH repositories; awareness raising and communication, events and, most importantly, one-on-one support provided to selected repositories. This section describes these activities and their outcomes as well as the results of the efforts to examine the SSH repository and trust landscape.

3.1 Dissemination activities and raising awareness

Part of the Task 8.2 efforts was meant to disseminate existing CoreTrustSeal certification support materials via communications channels of the SSHOC project. To this end, a web page was created in the SSHOC Service Catalogue with the assistance of WP2.[26] This page was used to provide links to helpful resources developed by the SSHOC project and other actors to aid SSHOC repositories in pursuing CoreTrustSeal certification or self-assessment. In addition, awareness about certification and the activities of T8.2 was raised by communicating about them through the SSHOC Certification mailing list, within the SSHOC network and informal contacts with other relevant networks, including the FAIRsFAIR[27] and EOSC Nordic[28] projects that had similar certification support programs.

Furthermore, as CoreTrustSeal is a community-driven certification framework developed to serve a wide range of repositories, T8.2 task team will communicate the outcomes of the support work and feedback to the CoreTrustSeal Board for the development of CoreTrustSeal certification after the end of the SSHOC project.

[26] Improved and FAIR data Repositories – SSHOC Trusted Repositories: https://sshopencloud.eu/sshoc-certification-support [31 January 2022]
[27] FAIRsFAIR project: https://www.fairsfair.eu/ [31 January 2022]
[28] EOSC Nordic project: https://www.eosc-nordic.eu/ [31 January 2022]

3.2 Workshops and webinars

The plans for face-to-face workshops were hampered by the COVID-19 pandemic, and all events were ultimately organised as virtual webinars. Task 8.2 participated in or organised the following events:

- SSHOC Webinar: *How to improve the quality of your repository? SSHOC and certification of repositories*[29], 24 April 2020. 61 participants.
- DARIAH Virtual Annual Event: *Building trustworthy repositories: Introduction to CoreTrustSeal certification*[30] (face-to-face workshop converted into an online one due to travel restrictions), 14 October 2020. 20 participants.
- An invitation-only webinar for the 14 supported repositories[31], 28 January 2021. 29 participants.
- FAIRsFAIR, SSHOC and EOSC-Nordic Workshop: *Towards a network of FAIR-enabling trustworthy digital repositories (TDRs)*. 13 January 2022.

The feedback received from participants indicates that the events were considered useful and well-organised. They were seen particularly useful in providing general overviews of certification and giving the opportunity to ask questions directly from community members experienced in certification. In addition, they were regarded as useful in networking and finding answers to common challenges.

The activities of T8.2 were also showcased and mentioned in several other webinars and events, such as the RDA Global Adoption week - Identify, Store and Preserve[32] (17 June 2020) as well as meetings with other stakeholders and projects including FAIRsFAIR and EOSC-Nordic.

[29] SSHOC WEBINAR: How to improve the quality of your repository? SSHOC and certification of repositories: https://sshopencloud.eu/sshoc-webinar-repositories-quality-certification [31 January 2022]
News item, *SSHOC webinar on certification of data repositories*: https://www.sshopencloud.eu/news/sshoc-webinar-certification-data-repositories [31 January 2022]

[30] Building trustworthy repositories: Introduction to CoreTrustSeal certification: https://www.sshopencloud.eu/building-trustworthy-repositories-introduction-coretrustseal-certification [31 January 2022]

[31] News item, *Update from SSHOC Certification Support*: https://www.sshopencloud.eu/news/update-sshoc-certification-support [31 January 2022]

[32] Adoption of RDA Recommendations Focus on Identity, Store and Preserve 2: https://www.youtube.com/watch?v=Mhtud8RXVXE [31 January 2022]

3.3 One-on-one support for repositories

In June 2020, T8.2 launched an open call[33] for repositories seeking support in attaining CoreTrustSeal certification. The application process helped gauge interest and align resources for reviewing self-assessments and other support activities. An application form was published on the SSHOC website and disseminated widely to European SSH data repositories with the help of WP2. All institutions self-identifying as repositories were able to apply.

Nine applications were received before the original deadline (of 31 July) which was extended by one month; five additional repositories applied during the extension period. In total, 14 repositories applied; six of them were included in the repositories selected as the primary recipients of SSHOC certification support in D8.2 (see Appendix 1). The eight other applicants were also deemed to be in scope of SSHOC and accepted in the support program. The repositories came from 13 different European countries (Figure 1).

[33] Call for Applications: https://www.sshopencloud.eu/news/call-applications-sshoc-repository-certification-support-0 [31 January 2022]

Figure 1. Repositories supported by Task 8.2.

The following repositories received certification support:

- **Center for Socio-Political Data (CDSP) / Sciences Po**. Established in 2006, the Center for Socio-Political Data (CDSP), a joint service unit of Sciences Po and the French National Centre for Scientific Research (CNRS), offers the scientific community services related to social science data and coordinates and participates in major projects in the field.
- **CLARIN-LV**. The Latvian centre of CLARIN, the Common Language Resources and Technology Infrastructure, offers access services to language data, tools and expertise.
- **Corpus OVI dell'italiano antico.** The first historical dictionary of ancient Italian to be born directly on the web, the Textual Corpus OVI (Opera del Vocabolario Italiano - CNR) is the largest database available today concerning the Italian language prior to 1400.

- **Croatian Social Science Data Archive (CROSSDA)**. A national infrastructure public service whose role is to ensure the long-term preservation and dissemination of social science research data. CROSSDA became a member of CESSDA in 2019.
- **DARIAH-DE Repository**. The German member of the Digital Research Infrastructure for the Arts and Humanities (DARIAH-EU) that supports the humanities and cultural sciences working with digital resources and methods in research and teaching.
- **Digital library of University of Maribor**. The institutionalised repository of the University of Maribor, Slovenia supporting open access to scientific, research and professional works, and research data resulting from research and education at the University.
- **Digital Repository of Scientific Institutes**. Polish initiative whose mission is digitisation and maximum dissemination, as well as providing permanent access and long-term digital preservation of scientific resources, in particular literature, scientific objects and data.
- **Historic Graves**. A community-focused grassroots heritage project surveying historic graveyards primarily in Ireland.
- **Lithuanian Data Archive for Social Sciences and Humanities (LiDA)**. A virtual digital infrastructure for data acquisition, long-term preservation and dissemination providing data access through Dataverse. LiDa is an aspiring CESSDA Service Provider.
- **mdwRepository**. mdwRepository supports capturing and preserving the intellectual outputs of the University of Music and Performing Arts, Vienna by ensuring and promoting sustainable services of ingest, storage and access to media objects.
- **NAKALA**. A repository of the French Very Large Research Infrastructure Huma-Num dedicated to social sciences and humanities hosting data from all types of projects in the field and accepting all types of data (text files, audio, video, images).
- **Publications of the Serbian Academy of Sciences and Arts**. A multi-institutional repository serving as the institutional repository of the Serbian Academy of Sciences and Arts (SASA) and covering several disciplines and archiving various content types including publications, theses, working papers, datasets, conference presentations etc.
- **SBX/CLARIN Repository (Språkbanken Text CLARIN Repository)**. A Swedish member of CLARIN providing access to linguistic data and tools.
- **Slovak Archive of Social Data (SASD)**. SASD accesses, processes, documents, stores and curates data files from social science research projects and promotes their dissemination for secondary use in academic research and for educational purposes. SASD is a member of CESSDA.

The applicants provided background information on their repository in the application form. A preliminary analysis of the information provided by the repositories revealed a broad range of academic domains, geographic coverage as well as levels of maturity. Personnel available to the repositories ranged from 0.5 FTE to 100 FTE. Nine repositories indicated their repository type[34] as "domain- or subject-based repository" and five selected multiple types. The repositories' level of curation varied between "basic curation" (e.g., brief checking, addition of basic metadata or documentation; 4 repositories),

[34] CoreTrustSeal types include Domain or subject-based repository; Institutional repository; National repository system, including governmental; Publication repository; Library/Museum/Archives; Research project repository; and Other.

"enhanced curation" (e.g., conversion to new formats, enhancement of documentation; 6 repositories) and "data-level curation" (as in enhanced curation but with additional editing of deposited data for accuracy; 4 repositories). The applicants also provided their reasons for pursuing certification. The reasons indicated by the repositories were broad-ranging and included, for instance, "showcasing trust", "obligation set by the respective ERICs", "adherence to the FAIR and TRUST principles", as well as "enhancing and improving practices and quality". One applicant had attained a prior repository certification (Data Seal of Approval, a predecessor of CoreTrustSeal).

The main mode of one-one-one support that T8.2 provided to the applicants was reviewing drafts of CoreTrustSeal self-assessments to assist them to identify gaps in their practices and documentation and help with the process of writing self-assessments and providing appropriate evidence. Each repository was assigned two supporters from the task team. A folder was created for each repository in a Google Drive with restricted access for the supporters and representatives of the repository. The restricted folder contained a template of the CoreTrustSeal requirements for the repository to complete with their draft self-assessment statements and links to relevant evidence. The supporters and other members of the task team were able to view the document and add comments.

The two supporters set up virtual meetings with the repository. The repositories were first requested to formulate concise self-assessment statements to four CoreTrustSeal requirements (suggested requirements were R0: Context, R1: Mission/Scope, R5: Organizational infrastructure and R13: Data discovery and identification) in order to get acquainted with self-assessment and gather supporting evidence. The repositories were given feedback on their statements in virtual meetings and assigned further Requirements to work with. The purpose was to continue the iterative process until the repository had formulated self-assessment statements for all 16 requirements.

Based on the self-assessment statements, the task team made recommendations on whether/how the repositories should continue the self-assessment process and if they were deemed ready to apply for CoreTrustSeal certification. Even if it turned out that the repository was not eligible for CoreTrustSeal certification, support was provided since the exercise was useful in determining and improving the quality of practices and services. In addition, non-eligible repositories allowed identifying cases of repositories that could benefit from some form of self-assessment or from other assessment frameworks.

Because the repositories varied in their level of maturity, curation practices, organisational infrastructure and documentation practices, common deadlines to all repositories were not established. Rather, the support was provided largely on the repository's terms. However, as a webinar for peer support was organised in January 2021, the repositories were highly encouraged to finish some self-assessment statements prior to the webinar in order to have a fruitful discussion on concerns regarding self-assessment and certification.

T8.2 supported the repositories in individual meetings between the assigned supporters and repository representatives throughout the year 2021. Regular meetings were organised with nearly all of the participating repositories roughly once a month. The support progress of each repository was followed in the monthly meetings of the task.

Depending on the goals of the repository and the progress they had made, the support process ended in January 2022 or will last until the end of the SSHOC project (April 2022). The repositories supported until the end of the project are those that are projected to be able to submit a CoreTrustSeal application.

The supported repositories were at different stages of maturity and organisational practices. Their goals also varied from aligning practices with the CoreTrustSeal requirements to full CoreTrustSeal certification. Some repositories required more support as well as more time for gathering documentation to be used as evidence than others. The repositories also differed in terms of the resources available for the certification support. Eight repositories had the goal of formal certification (Table 1), while six of them opted for self-assessment against the CoreTrustSeal requirements without an intention to apply within the project timeline.

At the time of writing this deliverable, six out of eight repositories with CoreTrustSeal as their goal have submitted or are close to submitting their application. These six and one additional repository that is potentially able to apply for CoreTrustSeal certification in the spring will continue to be supported until the end of the SSHOC project, April 2022.

Three repositories can be considered as having dropped out of certification support. In two cases this was because they could not commit resources to the process and in one case the repository decided to complete the application on its own and contact T8.2 in case of questions.

Formal CoreTrustSeal certification as goal	CoreTrustSeal Application submitted or close to being submitted	Support continued until April 2022	Dropped out	Interest in a wider network of existing and aspiring TDRs
8	6	7	3	7

Table 1. Number of supported repositories that had CoreTrustSeal certification as their goal, have submitted or are about to submit their application, will continue to be supported until the end of the SSHOC project, dropped out of the certification support process, and were interested in a trust and certification network after SSHOC.

The supported repositories were also asked about their interest in a post-project network of existing and aspiring trustworthy digital repositories. Seven of them responded and said yes. The network is being discussed with relevant stakeholders and a webinar on the network was organised by FAIRsFAIR, SSHOC and EOSC-Nordic on 13 January 2022, but at the time of writing this, there are no concrete timeline or plans yet.

The certification support process has allowed the participating repositories to assess their practices, organisational infrastructure and documentation. Some participants noted that completing the self-assessment for the first time takes up considerable time and resources, particularly in the case of smaller repositories, but they nonetheless acknowledged the benefits of the self-assessment and the support process. Even the repositories that did not have the formal CoreTrustSeal certification as their goal have benefited from the assessment, as the process has helped them to improve the information available to their users and stakeholders, and to assess their policies, procedures and data management processes against the CoreTrustSeal requirements. The repositories that aim for submitting a CoreTrustSeal application found the support helpful in facilitating the self-assessment and easing the effort required by having experienced experts answering questions and providing comments on the self-assessments. An example is the OVI, one of the constituent institutes of the CNR scientific network, that noticed the lack of public visibility of their documentation. This led to a notable development of self-awareness of digital content shared on the network, which directly influenced new online publications of documents and sometimes also of creation or modification of existing manuals, significantly increasing the usability of the information that their users can find online.

For some newer repositories the support process has been an opportunity to receive peer support in setting up their repository and ask for views on various aspects related to data services. In these cases, the support sometimes went beyond the confines of the CoreTrustSeal application and extended to technical and customer service aspects of data repositories. The process has deepened cooperation within the SSH community and provided the participating repositories with contacts they can turn to with questions about certification and issues related to trust.

The benefits of the certification support have been communicated to the SSHOC community and other interested parties with a success story describing the process of one participating repository and their thoughts on the support.[35]

3.4 SSH repository landscape

As explained in chapter 3.3, the task team identified 49 repositories as the main target group of candidates for certification monitoring and/or support (see also Appendix 1). The certification status of the listed repositories in January 2020 and after two years is summarised in Table 2. Nine out of the 49 repositories received or renewed their CoreTrustSeal. The number of repositories that have never been certified dropped by two but at the same time 15 repositories let their certification expire. Although some of them are in the process of renewing, the share of expired certifications is significant. This section and Discussion (Chapter 4) provide some insights as to why the share of repositories with valid certification has dropped.

Certification status in January 2020			**Certification status in January 2022**	
CoreTrustSeal v2017-2019	31		CoreTrustSeal v2020-2022	6
DSA	3		CoreTrustSeal v2017-2019	15
Not certified	15		Expired	15
Total	*49*		Not certified	13
			Total	*49*

Table 2. Certification status of listed repositories in January 2020 and January 2022.

The representatives of many of the repositories on the list participated in the events organised by T8.2, but only six applied for the one-on-one support. There can be several reasons for this, one of them being the lack of resources and commitment required for the self-assessment process. Another reason could be that some repositories felt that they did not require any support with their applications. In addition, certification may not have been relevant for some of the more recent organisations that had only just started setting up their services. Many of these incipient repositories attended the webinars, so they have the required information to apply when they have the means to do so.

Task 8.2 examined the trust landscape of SSHOC repositories by conducting a survey targeted at SSH organisations offering research data and metadata services, by examining various repositories belonging to four SSHOC infrastructures through desk research, and by inquiring SSHOC stakeholder views on certification with a questionnaire.

[35] News item *SSHOC Champion: Enabling access and reuse of our rich European heritage*: https://www.sshopencloud.eu/news/sshoc-champion-enabling-access-and-reuse-our-rich-european-heritage

The survey yielded responses from 14 SSHOC organisations comprising 11 data repositories and three other organisations offering data services. The results of the survey indicated that certification is linked to improved documentation on an organisation's processes. Certified organisations also provided more basic information on their activities to users and stakeholders than non-certified organisations. However, there is a great deal of variation in terminology, typology of services and essential information provided by data-holding organisations. This points to the need for further work in reaching community-agreed definitions and minimum information that data-holding organisations should provide. The results also demonstrated that due to the different service types and the fact that CoreTrustSeal is designed for repositories that preserve data in the long term, there are data services in the SSH community for which certification alternatives are limited.[36]

The desk research was conducted by collecting information from the websites of 93 data repositories on a list collected by the task during the planning of certification support. The list contained repositories from four SSHOC infrastructures: CESSDA ERIC, CLARIN ERIC, DARIAH ERIC and E-RIHS. The results demonstrated the diversity of SSHOC data repositories, particularly in terms of designated communities, disciplines and types of data. The analysis also supported the findings of the survey, as certified repositories provided more information on all of their activities. Certification was strongly connected to infrastructure membership, with CESSDA and CLARIN repositories being mostly certified and DARIAH and E-RIHS repositories significantly less so. The results did not allow confirming reasons for the notable differences in the certification status between infrastructures, but they may include the certification requirements in place by CESSDA and CLARIN for their members, lack of awareness of the benefits of certification, lack of resources for applying for certification, perceived difficulty of seeking certification, and specificity of data types for certain repositories.[37]

The results of the stakeholder infrastructure questionnaire showed that all of the stakeholders at least recommend, if not formally require, CoreTrustSeal certification to their affiliated organisations. Most of them (with the exception of CLARIN, which has its own internal B-centre certification) do not currently require or recommend any other certifications. In addition, none of them had plans to require or recommend further certifications or frameworks in the future. Two of the infrastructures provided targeted support to their affiliated organisations in seeking compliance and certification.

The infrastructures' and their members' experiences of utilising the CoreTrustSeal were mainly positive, but some criticism was also mentioned although the benefits of certification were acknowledged. The criticism included the length of the certification (review) procedure, the effort and time that applying for

[36] Ala-Lahti, Henri, Mathers, Benjamin Jacob, L'Hours, Hervé, Kleemola, Mari, & Alaterä, Tuomas J. (2022). *Repositories and Beyond: Analysis of Survey for SSHOC Organisations* (v1.0). Zenodo. https://doi.org/10.5281/zenodo.6325149

[37] Ala-Lahti, Henri, Mathers, Benjamin Jacob, L'Hours, Hervé, Kleemola, Mari, & Alaterä, Tuomas J. (2022). *Data Repositories and Certification in a Diverse Trust Landscape: Results of SSHOC T8.2 Desk Research (v1.0).* Zenodo. https://doi.org/10.5281/zenodo.6334025

certification takes (including the amount of communication required with various staff members), and the researchers' and the public's lack of awareness about the CoreTrustSeal.

The infrastructures were also asked to provide their opinions on the most and least relevant CoreTrustSeal requirements. All the requirements were mentioned as being relevant by at least one respondent, but five were mentioned by more than one. These were R5: Organizational Infrastructure, R6: Expert Guidance, R10: Preservation Policy, R12: Workflows, and R16: Security. Nearly all respondents decided not to select the least relevant requirements and noted that all of them are necessary and important for a TDR. Finally, the infrastructures were asked their interest in participating in a support network for increasing the uptake of certification. Three answered in the affirmative, although with some reservations depending on what form the network would eventually take. The establishment of such a network is discussed in Chapter 5.

The findings of the survey, desk research and stakeholder infrastructure questionnaire demonstrate the need for and advocacy of certification, CoreTrustSeal certification in particular, as the means to enhance repository practices and ensure trustworthy digital preservation. At the same time, the challenges of certification, such as the time and resources required and the diversity of the repository landscape, may prove an obstacle to certification for some repositories and data-holding organisations. Repositories planning CoreTrustSeal re-certification should also consider the duration of the certification process, which may span several months, in order to avoid the expiration of the certification. Further work and cooperation is required in reaching community-agreed definitions and standards for data-holding organisations as well as examining the diverse trust landscape of data repositories.

4. Discussion

This section considers evaluation and assessment solutions for SSH repositories as well as TDR partnership models and outsourcing of services.

4.1 Suggested evaluation and assessment solutions for SSH repositories

The experiences from the SSHOC certification support process and the examination of the trust landscape indicate that the benefits of core TDR certification like CoreTrustSeal are widely recognised. Even in cases when a repository is not in scope for CoreTrustSeal certification, self-assessment against the Requirements or an appropriate subset of Requirements is deemed useful.

During the support process, some repositories reported that writing the self-assessment statements against CoreTrustSeal requirements is resource-intensive and time-consuming. However, it can be argued that running a repository on a sustainable level requires good and well-managed documentation which is a relatively time-consuming task that needs to be appropriately resourced.

CoreTrustSeal, as a core certification, comprises the very minimum criteria for trustworthy repositories. Meeting these criteria involves providing evidence of the repository practices and procedures, all of which should be diligently documented to enable the running of the repository. Documentation should also be sufficiently accessible to the designated community. If the process and systems documentation and appropriate agreements with outsourcing or insourcing partners are in place and the responsibilities clear, applying for certification should not require a great deal of extra effort. CoreTrustSeal does not require producing a lot of new content for the reviewers but rather linking to the documentation already available.

This is backed up by the experience that first-time certification usually takes longer because typically it includes upgrading, creating or publishing documentation. Re-certification (or first-time certification if public documentation is in place) tends to be much quicker[38]. However, from the long-term preservation perspective, and for running any repository efficiently and reliably, documentation is what must be in place regardless of certification ambitions.

The amount of time and resources depends heavily on the repository's starting level, the staff and the funding available. But if certification is required or considered a valuable asset for running the repository, it is also easier to justify the allocation of resources required. It may be advisable to connect the certification to other development projects or separate funding. Based on the WP8 support team's

[38] The resources implications of evaluation processes were discussed in Kleemola et al., *SSHOC D8.2 Certification plan for SSHOC*, 20. https://doi.org/10.5281/zenodo.4558303

experiences, provision of peer support can also be beneficial to repositories at a wide range of maturity and size. Virtual meetings and feedback from subject experts can provide a great deal of progress to even less well-resourced repositories in relatively little time.

The idea of creating a 'lighter' certification scheme has come up several times in SSHOC trust discussions. Closer examination revealed that the idea of a 'lighter' approach was based on questions related to long-term preservation mission and curation activities. Some repositories did not have an explicit long-term preservation mission; instead, long-term preservation and usability of data was assumed as a self-evident outcome. This resulted in difficulties with CoreTrustSeal requirements related to preservation that can be solved with better guidance and clarification of what is expected for the requirements. A more common issue was that not all SSH repositories are in scope of CoreTrustSeal; for example, a repository that focuses on publication and access of data and does not have a long-term preservation mission, is out of scope. The scope of CoreTrustSeal certification has been discussed also by the wider CoreTrustSeal community (CoreTrustSeal Board 2021[39]). If a repository is not in scope of CoreTrustSeal due to lack of long-term preservation mission, the solution cannot be a 'lighter' TDR certification since CoreTrustSeal is already the community-reached minima of trustworthiness. However, there is a demand for assessment, certification, and recognition for SSH repositories and services without long-term preservation mission.

In all cases, repositories need to be clear on their responsibilities and priorities. This will enable their users to make informed decisions and them to decide which assessment frameworks suit them best. For repositories with a long-term preservation mission, CoreTrustSeal is the core certification. If, for example, data findability and accessibility are important, FAIR evaluation[40] is useful; if security aspects are important, there are ISO standards (like ISO 27001[41]) for security; and if IT service management is essential, FitSM[42] defines a baseline of IT service management effectiveness.

It should also be noted that not all organisations, repositories or services need or want certification. In the diverse SSH landscape, there are repositories that belong to the sphere of libraries, archives, and museums (LAM) that not only preserve digital data but also physical objects. For these, the digital data, while necessary and important, may nonetheless be of secondary importance compared to the actual physical objects. This could be one of the reasons for the lower number of certified repositories in the field of heritage science compared to fields where the data is born digitally.[43] But in case a LAM repository

[39] CoreTrustSeal Standards and Certification Board, *CoreTrustSeal: Specialists, Generalists, and Repository & Data Service Providers*. https://doi.org/10.5281/zenodo.4568875

[40] For example, F-UJI tool to assess FAIRness of research data objects based on metrics developed by the FAIRsFAIR project: https://www.f-uji.net/ [31 January 2022]

[41] ISO 27001 standard: https://www.iso.org/isoiec-27001-information-security.html [31 January 2022]

[42] FitSM: https://www.fitsm.eu/ [31 January 2022]

[43] Ala-Lahti, Henri, Mathers, Benjamin Jacob, L'Hours, Hervé, Kleemola, Mari, & Alaterä, Tuomas J. (2022). *Data Repositories and Certification in a Diverse Trust Landscape: Results of SSHOC T8.2 Desk Research (v1.0).* Zenodo. https://doi.org/10.5281/zenodo.6334025

seeks CoreTrustSeal certification, they would be certified for their trustworthy digital aspects, not for their physical collections.

Based on the SSHOC support experiences, there is no apparent demand for anything above CoreTrustSeal at the moment. CoreTrustSeal certification should not be extended by introducing requirements beyond the core criteria expected of TDRs.

In the repository field, there are several organisations that provide data-related services but do not curate or preserve data in the long-term. Not being repositories according to the definitions given above, they cannot seek to be TDRs and are not in scope of CoreTrustSeal as such. However, the experiences gained in the support process demonstrate that such organisations also benefit from assessing their procedures against the CoreTrustSeal requirements, even if they do not seek formal certification. Demonstrating the efforts to professionalise the services, ensure quality, and serve the designated communities / customers could be done by publishing the self-assessment results on the repository's website. The necessity of creating and maintaining documentation describing the services applies to these organisations as well. A peer support process could build on the CoreTrustSeal requirements.

The task team found it interesting to consider what certification strategies could be adopted by or supported for a specific domain or infrastructure. It would seem always sensible for any organisation to think about what parts and aspects of CoreTrustSeal requirements are most important for its functioning and why. SSH infrastructures could, for instance, select a subset of the CoreTrustSeal requirements to focus on and where they might provide shared evidence. Infrastructures may also require additional information pertaining to topics addressed by the CoreTrustSeal requirements, the examples of which include CESSDA's Annex II obligations about metadata and CLARIN's proposal for a centre's recommended format registration with the CLARIN standards committee list. Also, the usefulness of self-assessment of a limited subset of CoreTrustSeal requirements can be helpful, for instance in the case of CLARIN C-centres that only publish metadata but have no long-term preservation obligation, as discussed within the CLARIN Center Assessment Committee.

Work on CoreTrustSeal should thus be seen as part of the organisation's wider evaluation and assessment framework. Based on the SSHOC trust support experiences, the CoreTrustSeal generalist level 'core' of expectations is appropriate to SSH and no clashes between CoreTrustSeal and more specific SSH requirements were found.

Based on the feedback received from the supported repositories, CoreTrustSeal guidance would benefit from clarifications. Many participants found it challenging to discern where to provide which information, and how extensive this information should be and how much details about procedures and processes are expected. Shared evidence like the Dataverse Software Guide for CTS certification[44] was found useful and the SSH ERICs should consider providing shared evidence for their members whenever possible. The

[44] Dataverse Software Guide for CTS certification: https://dataverse.org/cts-guide [31 January 2022]

CoreTrustSeal requirements should also cover rights, access and licences better (R2). The requirement of formal, written agreement between the repository and an organisation that would guarantee to take over in case of service discontinuity (R3) was problematic for repositories e.g. in case that their host organisation is responsible by default so no written agreements were deemed necessary or even possible. It would be good to clarify that R11 (Data Quality) is not about scientific quality of the data. Data citations are an important part of data discovery and identification (R13) so the requirements on citation could be more specific. Whenever possible, information should be collected in a more structured way instead of prose narrative.

The task team will continue collecting feedback about the CoreTrustSeal Requirements and provide detailed feedback for CoreTrustSeal Board's consideration during the 2022 review of requirements.

Certification and self-assessment contribute to ensuring sustainability of data and they are useful for improving various repository practises and showcasing trustworthiness. Trust, however, is not contingent on certificates alone but is earned through interaction with and accountability to the designated community. This is why cooperation through peer networks can be an important route to growth and sustainability. Such networks could be starting points for shared responsibility for sustainable data and enable ensuring continuity of repositories. However, the current models of trust support that largely rely on project-based funding are not favourable to ensuring sustainability, which is why more enduring solutions are required.

4.2 TDR Partnership Models and Outsourcing

Entities offering trustworthy digital repository services (and that are therefore candidates for CoreTrustSeal certification[45]) may be a discrete organisation, consortia, or hosted as part of a larger organisation. In each of these organisational structures some elements of the functions and/or processes that make up the services may be outsourced to a third party. Repositories may also depend on host or partner organisations for some part of their services; CoreTrustSeal refers to this as 'insourcing'. When evaluating trustworthiness it is important for certification bodies to be able to identify what activity is being undertaken by the applicant and when it is managed through some other outsource or insource relationship. To this end, the boundary between the repository and the host institution needs to be clearly defined so there can be a clear delineation as to when an insource relationship is in place. For applicants it is important to be able to define what processes are undertaken by others and how they are managed to an agreed level of quality, efficiency etc. Examples of such functions may include, but

[45] CoreTrustSeal, Data Repositories Requirements: https://www.coretrustseal.org/why-certification/requirements/ [31 January 2022]

are not limited to, technological infrastructure, software development, systems maintenance and service desk management.

Outsourced services such as consolidated storage (bit-level integrity) can be offered in the same way across disciplines. Outsourcing that provides SSH-specific services will need to support the range of metadata schemes and data formats required by the SSH community.[46] There are few community-agreed minima on what information should be made available about outsourced services at this time. Ideally such information would include specification of which data service functions they support (deposit, appraisal, curation, preservation, access etc.), the nature of the relationship (host organisation, service level agreement, legal-contractual) and the specifics of the disciplinary support for data and metadata.

For the purposes of this document the following broad definitions are used:

Outsourcing

> *The use of an external resource (third-party provider), to procure some (or all) of the functions required for the repository to function.* [47]

Insourcing

> *The (sometimes mandatory) use of services or resources managed by a host institution.*[48]

In either case the activity may be paid for or provided on a not-for-profit basis. A repository utilising a third-party software package (whether open source or purchased), but having no other service relationship in place, would not be classed as outsourcing in a CoreTrustSeal application.

The nature of an insourcing relationship may mean that the repository/applicant has less opportunity to mutually agree a level or service or define expectations in a formal contract. Otherwise the issues of managing insource and outsource relationships (and their impact on a CoreTrustSeal application) are broadly similar.

Examples of the different repository functions that can be outsourced would be:

> **Acquisition and Appraisal** - Collections development (e.g. data selected for acquisition, selection of data for long term preservation); assessment of data quality and relevance

[46] For more information about metadata and data formats, see Broeder et al., *SSHOC D3.1 Report on SSHOC (meta)data interoperability problems*. https://doi.org/10.5281/zenodo.3569868

[47] This definition is a modified version of Webopedia's definition of *IT Outsourcing:* https://www.webopedia.com/definitions/it-outsourcing/ [31 January 2022].

[48] This definition was formulated based on internal discussion within the SSHOC T8.2 team, thus no reference is available.

Negotiation and administration - Negotiation with depositors (e.g. drafting of contractual agreements, negotiation of the contents of the submission information package); legal/regulatory compliance checks (particularly for sensitive data); user support pre-ingest (e.g. via a service desk).

Ingest and curation - Processing of submission information packages; cataloguing of digital objects and/or collections; generation of metadata (and supporting documentation)

Archival storage - Generation of archival information package(s); data backup (e.g. cloud storage); data transfer and integrity assessments (prior to or when preparing for storage)

Resource discovery - Persistent identifiers; catalogue management (internal); metadata management (registry listings [repository]); third-party catalogue listings [digital objects].

Access management - Creation and management of licences; access portal (design / maintenance); user access support (e.g via a service desk)[49]

Long term preservation - parts of the preservation function may be outsourced including technology watch functions (e.g. potential file and/or software obsolescence); audits and file maintenance (e.g. periodic reviews of data integrity, file migration in the event of obsolescence); software emulation (for legacy media). The limitations on outsourcing related to LTDP are described below.

The above provides a partial list of examples of different functions that repositories can choose to either outsource to third-party providers or insource from a host institution.[50]

Repositories can opt to outsource either technical systems, staff activity or a combination of both. Such partial arrangements are not uncommon and are often underpinned by hybrid outsourcing strategies that combine a plurality of relational and contractual elements.[51] For example, a repository may choose to outsource only some of the staff roles within a given team, creating a mix of subcontractors and internal employees. To add a further layer of complexity, the repository may then be outsourcing one or more of its technical services to another third-party provider, which both the subcontracted and internal employees utilise in their roles. This example illustrates the complexity of such arrangements and the

[49] Access type (and the controls required) can influence a repository's decision regarding the outsourcing of some (or all) of their access management functions. The low cost/simplicity of open access compared to the cost and complexity of managing access to sensitive data may lead some repositories to outsource the latter. Decisions to do so are dependent on facilities available and/or access to appropriately trained/able staff. Due to the costs of sensitive storage, ongoing engagement with researchers who re-use the data is required to drive future preservation decisions.

[50] Not all areas covered in the above list would necessarily be suitable for insourcing. File maintenance, data backup and software emulation are examples of ones that potentially could (amongst others).

[51] Rai et al., "Hybrid Relational-Contractual Governance for Business Process Outsourcing." https://doi.org/10.2753/MIS0742-1222290208

need for careful consideration as to whether outsourcing/insourcing are suitable options. In all cases where outsourcing is partial, the repository is still responsible for providing evidence for the processes they continue to undertake themselves.

Questions that digital repositories may wish to ask themselves before implementing an outsource solution include:

- Can a third-party maintain the level of community trust that the digital repository (particularly publicly funded digital repositories) has earned?
- Will any decisions made by the third-party company concerning the technical infrastructure / architecture underlying the services provided be in the best interest of the repository and its community, even if those decisions reduce or do not contribute to the third-party company's profit?
- Does the relationship between the repository and its designated community change when services are outsourced?
- Does the responsibility of the digital repository change when services are outsourced?[52]

Such questions hint at the potential 'pros and cons' of outsourcing repository functions; there are a number of potential trust-related risks when an organisation decides to outsource a particular service/service function. Outsourcing arrangements can be opaque, at least from a user perspective. This lack of clarity negatively impacts on the levels of trust between a repository and a designated community/community of end-users, which can be further compounded should the organisation fail to clearly distinguish between the service(s) that it provides and those provided by a third-party.[53]

Information security is another challenge when outsourcing, particularly for repositories that store sensitive digital objects and/or significant amounts of personal data (e.g., data concerning producers, repositories, researchers using the datasets etc.). A breach of information security (even when the repository is not responsible) can result in the repository sustaining significant reputational damage and possible legal consequences.

The decision to insource or outsource a service that cannot otherwise be provided may be motivated by a perceived increase in cost-saving and/or operational efficiency. Realising the potential benefits of outsourcing is strongly dependent on the service function being outsourced and the organisational structure and infrastructure in place e.g. generalist or specialist status; organisational structure (independent repositories versus those based within a host institution); existing technical and organisational infrastructure.

[52] The hypothetical questions posed were adapted from those provided by Jerrard et al., *Privatizing Libraries*.

[53] A common example of this being when organisations outsource their storage functionality to a third-party provider whilst maintaining their own internal system and/or user-facing service front, but without explicitly highlighting such arrangements to end-users.

Most respondents to the SSHOC repository landscape survey appeared to have at least some form of outsourcing arrangement with a third-party provider. This was particularly apparent for non-repository data services.[54] The SSHOC desk research focused solely on certified and non-certified repositories rather than wider-service providers and also provided evidence of insourcing and outsourcing.[55] Some of the CoreTrustSeal certified repositories (and a smaller number of the non-certified repositories) made reference to third-party agreements with external providers. Arrangements were typically referenced in either their Website terms and conditions, Data Protection policies and/or other documents that provided Legal Information to End Users. The disparate and sporadic provision of information concerning third-party agreements meant that it was not feasible to judge the specific repository functions, the formality or the level of SSH outsourcing. The survey and the desk research demonstrate the challenges in obtaining this information, along with the need for community consensus on the appropriate, and minimum levels of information pertaining to insourcing and outsourcing arrangements.

For applicants seeking to achieve CoreTrustSeal certification it is also important to consider their impact on self-assessment statements and supporting evidence. Almost any part of the CoreTrustSeal requirements could be undertaken by insourcing/outsourcing so long as the repository can demonstrate that it has an appropriate relationship with the provider and that they are able to maintain the services to be outsourced. In their report, the CoreTrustSeal Board note:

> "The applicant may outsource to third parties. Outsourcing roles and relationships should be well defined, and all parties must provide evidence related to all of the functions or processes they help undertake."[56]

The exceptions are that CoreTrustSeal applicants must have a mission to (see CoreTrustSeal requirement R01), and take responsibility for deciding on what preservation actions must be taken to preserve data and metadata (for the designated community). This includes the maintained usability and understandability of metadata (e.g. by updating to more modern schemas and/or ontologies) and data (e.g. by migrating to more modern formats or offering emulation solutions). The challenge for the CoreTrustSeal process (applicant and reviewer) is in defining when evidence provided for an insource/outsource relationship is sufficient.[57]

---

[54] Ala-Lahti, Henri, Mathers, Benjamin Jacob, L'Hours, Hervé, Kleemola, Mari, & Alaterä, Tuomas J. (2022). *Repositories and Beyond: Analysis of Survey for SSHOC Organisations* (v1.0). Zenodo. https://doi.org/10.5281/zenodo.6325149

[55] Ala-Lahti, Henri, Mathers, Benjamin Jacob, L'Hours, Hervé, Kleemola, Mari, & Alaterä, Tuomas J. (2022). *Data Repositories and Certification in a Diverse Trust Landscape: Results of SSHOC T8.2 Desk Research (v1.0).* Zenodo. https://doi.org/10.5281/zenodo.6334025

[56] CoreTrustSeal Standards and Certification Board, *CoreTrustSeal: Specialists, Generalists, and Repository & Data Service Providers*, 1. https://doi.org/10.5281/zenodo.4568875

[57] Exemplary evidence includes the nature of agreements between parties and documented assurance that the third party is able to deliver the contracted functions and/or services.

Some insource/outsource arrangements may be underpinned by the notion of mutual trust rather than rigorous, contractually specific legal agreements). Even with more formalised relationships the ability to provide evidence of such outsourcing arrangements can be problematic as the terms of the arrangement may be commercially sensitive and therefore not available as 'evidence'. There may be significant differences in the level of expectation between a paid for outsource partner (with or without rigorous contractual expectations in place) compared to an insource relationship with a host organisation, where even basic operational level agreements can be hard to develop. Difficulties with forming such agreements may be due to internal resistance to making such specific guarantees, but also because host organisations often do not have existing mechanisms in place to provide such agreements/guarantees.

Difficulties in defining the acceptable types (and quantities) of supporting evidence for an insource/outsource relationship present challenges for both applicants and reviewers of the CoreTrustSeal certification, yet they are not unique to the CoreTrustSeal, or to data repositories or the wider data industry. All elements of federated infrastructures (including federated infrastructures such as EOSC[58]) depend on a variety of insource and outsource partnerships that must deliver performant and trustworthy services. Further exploratory work is therefore needed to identify potential solutions to these issues. The feasibility of any proposed solution would also need to be addressed, particularly given the broad nature of CoreTrustSeal and the diverse nature of the certified repositories it represents.

The range of non-repository data services and their functions may also become more clearly defined as the sector matures, in turn making it easier to specify the minimum levels of contractual agreement between parties more clearly (along with other types of formal agreement[59]). Evidence of contractual agreements is required to ensure that trust in the applicant can be appropriately extended to trust in an associated insource or out-source partners. Thus, it is important for applicants to clearly, and explicitly outline which functions they outsource, to whom they are outsourced to, and how the agreements between the parties are defined in as clear a way as possible so as to maximise their chances of receiving CoreTrustSeal certification.

[58] EOSC Portal: https://eosc-portal.eu/about/eosc [31 January 2022]

[59] Other types of formal agreement include Operational Level Agreements agreed between the repository and the host institution when insourcing a particular function, amongst others.

5. Sustainable management of trust after SSHOC

The SSHOC project has provided means to assess and increase the trustworthiness of repositories in the social sciences and humanities by supporting repositories on their journey to enhanced trustworthiness and specifically on the application of the CoreTrustSeal certification. These supporting actions have raised awareness of policies, guidelines, standards and best practices that strengthen the long-term management of digital assets and the FAIR ecosystem. Information on the SSHOC website, participation in workshops and feedback on self-assessments against the CoreTrustSeal requirements have raised trust management to a higher level. In the communication with the repositories, the T8.2 team has stressed that trust management requires active attention also after the end of the SSHOC project. This is illustrated by, for example, that the CoreTrustSeal certification is granted for a limited period of time and that the CoreTrustSeal requirements themselves are subject of a review process (next review in 2022).

Some individual ERICs, such as CESSDA and CLARIN, provide targeted support for their members on issues related to trust and seeking compliance. While this support is useful in allowing repositories to address trust issues with peers from the same field with similar challenges, there are also plans for a higher-level trust network. The long-term aspect of data management is taken into account by the SSHOC, FAIRsFAIR and EOSC-Nordic projects in several ways. The projects initiated preliminary discussion towards the development of a European network of FAIR enabling trustworthy digital repositories.[60] This network

> "could communicate inputs from and promote cooperation between existing partnerships. A higher level network would help achieve this (compared to the current status quo) by having a narrower focus, in turn permitting a focus on coordinating communications and outcomes across the various regional, national and international networks (and the standards employed by them) that are currently in existence. As technologies, methods and user communities evolve, there is a need to update existing practices and create new ones in order to maintain FAIR data in trustworthy repositories."[61]

The article "FAIR + Time: Preservation for a Designated Community" contains valuable building blocks for the formulation of a long-term policy for digital objects. The article suggests building a long-term access strategy around the OAIS Reference Model (ISO14721).[62] The article states that

[60] See von Stein et al., *D4.4 Coordination Plan for a sustainable network of FAIR-enabling Trustworthy Digital Repositories*. https://doi.org/10.5281/zenodo.5726691
[61] Ibid., 8.
[62] For OAIS Reference Model, see: CCSDS The Consultative Committee for Space Data Systems, *The Reference Model for an Open Archival Information System (OAIS). Recommended practice.* http://public.ccsds.org/publications/archive/650x0m2.pdf

> "[c]oncepts [related to the OAIS Reference Model] inform a number of mandatory responsibilities that include the provision of active long-term preservation sufficient to ensure that digital objects (data and metadata) become and remain independently understandable to a designated community of users that have a defined knowledge base. Repository preservation policies, procedures and actions are defined in light of both cultural and technological change. [...] Only an organisation that meets clear criteria, including the provision of active long-term preservation measures for a designated community, can be termed a 'trustworthy digital repository'. As indicators and tests emerge for the assessment of data and metadata as FAIR it will become possible to identify whether a TDR is also enabling FAIR data."[63]

The future efforts to manage trust in a sustainable manner within SSH should align with the approaches outlined in the SSHOC Task 8.1's Deliverable *Governance and Sustainability Roadmap*[64] as well as the objectives of the SSHOC *Memorandum of Understanding for the establishment of the SSH Open Cluster*[65]. Future directions based on the experience of a wide range of EOSC projects can be found in the recommendations of the Synchronisation Force White Paper.[66] The EOSC Association will be progressing through a range of advisory groups[67] and associated task forces, including one focussed on long term data preservation[68] that will provide recommendations on the vision and sustainable implementation of long-term data preservation policies and practices, as well as suggestions to later strategy execution.

The EOSC Sustainability Working Group (SWG) has commissioned a study to assess and make recommendations about the role of digital preservation capacity within the EOSC community, and to make recommendations about the role of digital preservation within the emerging EOSC vision.[69] The report is relevant for the sustainable management of trust after the ending of the SSHOC project as it recommends contributing to "ongoing CoreTrustSeal+FAIR preservation work for alignment of repository certification schemas with FAIR".[70]

Modern repositories are often developed through partnerships and with a wide range of insource and outsource options. Trust between these different data service actors is essential for the data ecosystem,

[63] L'Hours et al., *FAIR + Time: Preservation for a Designated Community*. https://doi.org/10.5281/zenodo.5797776
[64] Forthcoming.
[65] Document not publicly available.
[66] Dillo et al., *D5.7 Recommendations for a FAIR EOSC - White Paper FAIRsFAIR Synchronisation Force 2021*. https://doi.org/10.5281/zenodo.5793105
[67] EOSC Association Advisory Groups: https://eosc.eu/advisory-groups [31 January 2022]
[68] EOSC Association Task Force on Long-Term Data Preservation: https://www.eosc.eu/advisory-groups/long-term-data-preservation [31 January 2022]
[69] Currie and Kilbride, *FAIR Forever? Long Term Data Preservation Roles and Responsibilities, Final Report.* https://doi.org/10.5281/zenodo.4574234
[70] Ibid., 46.

and future assessment and certification options and networks should cover data service providers and the issue of enabling FAIR data.

Repository certification is a key part of building a trusted FAIR data ecosystem. It needs to be applied in a way that acknowledges the differences in goals, practices and maturity of data repositories and other service providers. Interoperability, standards, automation and technology are all parts of the solution, but reusability of data and long term preservation of understandability is ultimately dependent on domain and disciplinary expertise. SSH repositories are critical to the future delivery of the wider EOSC.

6. Conclusion

The experiences from the trust support work indicate that the benefits of repository certification are broadly acknowledged, and the repositories involved found the support process useful. Although some participants opted for self-assessment without intention to apply for formal certification due to resources available or not being in scope for CoreTrustSeal, the process enabled them to improve their documentation, policies and procedures. Feedback on the CoreTrustSeal certification pointed to the need to clarify its guidance and to define in even more detail the necessity to include evidence and documentation rather than spell out the procedures in the assessment.

Examination of the SSHOC trust landscape demonstrated the diversity of organisations and data service types. This diversity combined with the fact that certification, particularly for the first time, demands time from the repositories points to the difficulty of finding certification solutions that are suitable for all organisations. Furthermore, complex partnerships models and outsourcing pose their own challenges for certification. While outsourcing is not an obstacle to certification as such, it requires repositories outsourcing their functions to explicitly state these functions and related documentation.

Although the certification options are limited for organisations out of scope of formal certification schemes, such as CoreTrustSeal, there are some alternatives for them, as mentioned in section 4.1. These organisations, as well as repositories seeking formal certification, should consider their needs and select the framework that suits their requirements. In addition, some organisations may find it useful to conduct a self-assessment against CoreTrustSeal requirements even if they do not plan to seek certification. This has proved beneficial for certain data services in improving their practices.

Overall, the CoreTrustSeal expectations are appropriate for SSH repositories seeking TDR status and no clashes between CoreTrustSeal and more specific SSH requirements were found. There is no apparent demand for anything more detailed and extensive than CoreTrustSeal at the moment. CoreTrustSeal certification should not be extended by introducing requirements beyond the core criteria expected of repositories.

Sustainable management of trust goes beyond assessment, evaluation and certification. While these are valuable in demonstrating trustworthiness to users and stakeholders, earning trust and maintaining it in a sustainable manner also requires cooperation, accountability and shared responsibility. Managing trust in the future entails commitment to the common goals agreed on by the community and peer collaboration through support programmes and networks, while ensuring solid resources for the upkeep of these collaborative efforts.

7. References

Ala-Lahti, Henri, Mathers, Benjamin Jacob, L'Hours, Hervé, Kleemola, Mari, & Alaterä, Tuomas J. (2022). *Repositories and Beyond: Analysis of Survey for SSHOC Organisations* (v1.0). Zenodo. https://doi.org/10.5281/zenodo.6325149

Ala-Lahti, Henri, Mathers, Benjamin Jacob, L'Hours, Hervé, Kleemola, Mari, & Alaterä, Tuomas J. (2022). *Data Repositories and Certification in a Diverse Trust Landscape: Results of SSHOC T8.2 Desk Research (v1.0).* Zenodo. https://doi.org/10.5281/zenodo.6334025

Broeder, Daan, Thorsten Trippel, Emiliano Degl'Innocenti, Roberta Giacomi, Maurizio Sanesi, Mari Kleemola, Katja Moilanen, Henri Ala-Lahti, Caspar Jordan, Iris Alfredsson, Hervé L'Hours and Matej Ďurčo. 2019. *SSHOC D3.1 Report on SSHOC (meta)data interoperability problems* (v1.0). Zenodo. https://doi.org/10.5281/zenodo.3569868

CCSDS The Consultative Committee for Space Data Systems. 2012. *The Reference Model for an Open Archival Information System (OAIS). Recommended practice*. CCSDS 650.0-M-2, Magenta Book, June 2012. http://public.ccsds.org/publications/archive/650x0m2.pdf

CCSDS The Consultative Committee for Space Data Systems. 2014. *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories. Recommended Practice*. CCSDS 652.1-M-2, Magenta Book, March 2014. https://public.ccsds.org/Pubs/652x1m2.pdf

CoreTrustSeal Standards and Certification Board. 2021. *CoreTrustSeal: Specialists, Generalists, and Repository & Data Service Providers* (v02.00). Zenodo. https://doi.org/10.5281/zenodo.4568875

Currie, Amy and William Kilbride. 2021. *FAIR Forever? Long Term Data Preservation Roles and Responsibilities, Final Report* (Version 7). Zenodo. https://doi.org/10.5281/zenodo.4574234

Dillo Ingrid, Simon Hodson, Sara Pittonet Gaiarin and Marjan Grootveld. 2021. *D5.7 Recommendations for a FAIR EOSC - White Paper FAIRsFAIR Synchronisation Force 2021* (Version 1.0 DRAFT). Zenodo. https://doi.org/10.5281/zenodo.5793105

Grootveld, Marjan, Simon Hodson, Sara Pittonet Gaiarin, Joy Davidson and Ingrid Dillo. 2021. *D5.6 Report 3 of the Synchronisation Force* (V1.0_DRAFT). Zenodo. https://doi.org/10.5281/zenodo.5336658

Jerrard, Jane, Nancy Bolt and Karen Strege. 2012. *Privatizing Libraries*. ALA Editions: Special Reports.

Kleemola, Mari, Tuomas J. Alaterä, Niko Koski, Henri Ala-Lahti, Birger Jerlehag, Hervé L'Hours, Franciska De Jong, Dieter Van Uytvanck, Tomasz Parkola, Emiliano Degl'Innocenti, Maurizio Sanesi and René van Horik. 2020. *SSHOC D8.2 Certification plan for SSHOC repositories* (v1.0). Zenodo. https://doi.org/10.5281/zenodo.4558303

L'Hours, Hervé, Mari Kleemola, Ilona von Stein, René van Horik, Patricia Herterich, Joy Davidson, Olivier Rouchon, Mustapha Mokrane and Robert Huber. 2022. *FAIR + Time: Preservation for a Designated Community* (02.00). Zenodo. https://doi.org/10.5281/zenodo.5797776

Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, Hervé L'Hours, Wim Hugo, Reyna Jenkyns, Varsha Khodiyar, Maryann E. Martone, Mustapha Mokrane, Vivek Navale, Jonathan Petters, Barbara Sierman, Dina V. Sokolova, Martina Stockhause and John Westbrook. 2020. "The TRUST Principles for digital repositories." *Scientific Data* 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

Rai, Arun, Mark Keil, Rob Hornyak and Kim Wüllenweber. 2012. "Hybrid Relational-Contractual Governance for Business Process Outsourcing." *Journal of Management Information Systems* 29, 2: 213–256. https://doi.org/10.2753/MIS0742-1222290208

von Stein, Ilona, Hervé L'Hours, Linas Cepinskas, Benjamin Mathers, Ingrid Dillo, Maaike Verburg, Mustapha Mokrane, Patricia Herterich and Olivier Rouchon. 2021. *D4.4 Coordination Plan for a sustainable network of FAIR-enabling Trustworthy Digital Repositories* (1.0_DRAFT). Zenodo. https://doi.org/10.5281/zenodo.572669

8. Appendixes

Appendix 1. Update of certification status

The table lists the 49 repositories selected as candidates for certification support in SSHOC D8.2 Certification plan for SSHOC repositories (January 2020). The repositories that received one-on-one support from task 8.2 are highlighted in blue. The repositories that have achieved the estimated goal by 31.1.2022 are highlighted in green (A = renewal of existing CoreTrustSeal certification, B = new CoreTrustSeal certification). In addition, six repositories not on this list were also supported by task 8.2. It should be noted that, as CoreTrustSeal is valid for three years from the certification date, the certificate in accordance with 2017-2019 requirements was still valid for 12 repositories (31.1.2022), so starting the renewal process during SSHOC was probably not on their agenda.

Repository name	Country	Community	Certification in Jan 2020 (D8.2)	Goal (D8.2)	Status 31.1.2022	Certification in Jan 2022	Certification date	Certification valid until	One-on-one support from SSHOC
Corpus testuale OVI	Italy	E-RIHS	Not certified	B	Not certified	NA	NA	NA	yes
Digital Repository of Scientific Institutes (DRSI)	Poland	DARIAH	Not certified	B	Not certified	NA	NA	NA	yes
NAKALA	France	DARIAH	Not certified	B	Not certified	NA	NA	NA	yes
PROGEDO Research Infrastructure	France	CESSDA	Not certified	B	Not certified	NA	NA	NA	yes

Slovak Archive of Social Data (SASD)	Slovakia	CESSDA	Not certified	B	Not certified	NA	NA	NA	yes
Språkbanken, The Swedish language bank	Sweden	CLARIN	DSA	B	Expired	Expired	11.11.2016	11.11.2019	yes
ACDH - A Resource Centre for the HumanitiEs (ACDH-ARCHE)	Austria	CLARIN	CoreTrustSeal v2017-2019	A	Renewed	CoreTrustSeal v2020-2022	9.7.2021	9.7.2024	no
CMU-TalkBank (CMU)	USA	CLARIN	CoreTrustSeal v2017-2019	A	Renewed	CoreTrustSeal v2020-2022	12.10.2021	12.10.2024	no
Data Archiving and Networked Services (DANS)	Netherlands	CESSDA	CoreTrustSeal v2017-2019	A	Renewed	CoreTrustSeal v2020-2022	6.9.2021	6.9.2024	no
Digital Repository of Ireland (DRI)	Ireland	DARIAH	CoreTrustSeal v2017-2019	A	Renewed	CoreTrustSeal v2020-2022	28.10.2021	28.10.2024	no
Finnish Social Science Data Archive (FSD)	Finland	CESSDA	CoreTrustSeal v2017-2019	A	Renewed	CoreTrustSeal v2020-2022	6.11.2020	6.11.2023	no
Austrian Social Science Data Archive (AUSSDA)	Austria	CESSDA	Not certified	B	Certified	CoreTrustSeal v2017-2019	28.7.2020	28.7.2023	no
CLARIN.SI Language Technology Centre (CLARINSI)	Slovenia	CLARIN	DSA	B	Certified	CoreTrustSeal v2017-2019	30.9.2020	30.9.2023	no
Portuguese Social Information Archive (APIS)	Portugal	CESSDA	Not certified	B	Certified	CoreTrustSeal v2020-2022	1.9.2021	1.9.2024	no
UK Data Service (UKDS)	United Kingdom	CESSDA	DSA	B	Certified	CoreTrustSeal v2017-2019	18.5.2020	18.5.2023	no

ASV Leipzig	Germany	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	19.2.2019	19.2.2022	no
Bayerisches Archiv für Sprachsignale (BAS)	Germany	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	17.5.2019	17.5.2022	no
CLARINO Bergen Center	Norway	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	18.11.2019	18.11.2022	no
CLARIN-PL Language Technology Centre	Poland	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	17.12.2019	17.12.2022	no
Eberhard Karls Universität Tübingen (EKUT)	Germany	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	27.3.2019	27.3.2022	no
Geisteswissenschaftliches Asset Management System (GAMS)	Germany	DARIAH	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	18.4.2019	18.4.2022	no
Hamburger Zentrum für Sprachkorpora (HZSK)	Germany	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	13.2.2019	13.2.2022	no
Institut für Deutsche Sprache (IDS)	Germany	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	18.3.2019	18.3.2022	no
LINDAT/CLARIN (LINDAT)	Czech Republic	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	28.8.2019	28.8.2022	no
PORTULAN CLARIN	Portugal	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	17.12.2019	17.12.2022	no
The CLARIN Centre at University of Copenhagen (CLARIN-DK-UCPH)	Denmark	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	27.7.2019	27.7.2022	no

Universität des Saarlandes (UdS)	Germany	CLARIN	CoreTrustSeal v2017-2019	A	Certified	CoreTrustSeal v2017-2019	15.2.2019	15.2.2022	no
Berlin-Brandenburg Academy of Sciences and Humanities (BBAW)	Germany	CLARIN	CoreTrustSeal v2017-2019	A	Expired	Expired	25.10.2018	25.10.2021	no
Center of Estonian Language Resources (CELR-EKK)	Estonia	CLARIN	CoreTrustSeal v2017-2019	A	Expired	Expired	2.11.2018	2.11.2021	no
Czech Social Science Data Archive (CSDA)	Czech Republic	CESSDA	CoreTrustSeal v2017-2019	A	Expired	Expired	23.1.2018	23.1.2021	no
GESIS - Leibniz Institute for the Social Sciences	Germany	CESSDA	CoreTrustSeal v2017-2019	A	Expired	Expired	15.9.2017	15.9.2020	no
Institut für Maschinelle Sprachverarbeitung (IMS)	Germany	CLARIN	CoreTrustSeal v2017-2019	A	Expired	Expired	2.12.2018	2.12.2021	no
Instituut voor de Nederlandse Taal (IVDNT)	Netherlands	CLARIN	CoreTrustSeal v2017-2019	A	Expired	Expired	10.12.2018	10.12.2021	no
Meertens Instituut/HUC (MI)	Netherlands	CLARIN	CoreTrustSeal v2017-2019	A	Expired	Expired	5.3.2018	5.3.2021	no
MPI for Psycholinguistics (MPI-PL)	Netherlands	CLARIN	CoreTrustSeal v2017-2019	A	Expired	Expired	18.1.2019	18.1.2022	no
Norwegian Centre for Research Data (NSD)	Norway	CESSDA	CoreTrustSeal v2017-2019	A	Expired	Expired	19.3.2018	19.3.2021	no
Social Science Data Archives (ADP)	Slovenia	CESSDA	CoreTrustSeal v2017-2019	A	Expired	Expired	23.1.2018	23.1.2021	no

Swedish National Data Service (SND)	Sweden	CESSDA	CoreTrustSeal v2017-2019	A	Expired	Expired	7.2.2018	7.2.2021	no
Swiss Centre of Expertise in the Social Sciences (FORS)	Switzerland	CESSDA	CoreTrustSeal v2017-2019	A	Expired	Expired	20.3.2018	20.3.2021	no
The ILC4CLARIN Centre at the Institute for Computational Linguistics (ILC4CLARIN)	Italy	CLARIN	CoreTrustSeal v2017-2019	A	Expired	Expired	18.4.2018	18.4.2021	no
The Language Bank of Finland (FIN-CLARIN)	Finland	CLARIN	CoreTrustSeal v2017-2019	A	Expired	Expired	23.4.2018	23.4.2021	no
Danish National Archives (DNA)	Denmark	CESSDA	Not certified	B	Not certified	NA	NA	NA	no
Data Center Serbia for Social Sciences (DCS)	Serbia	CESSDA	Not certified	B	Not certified	NA	NA	NA	no
Greek research infrastructure for the social sciences (So.Da.Net)	Greece	CESSDA	Not certified	B	Not certified	NA	NA	NA	no
MObile-laboratory VIsualization DAta (MOVIDA)	Italy	E-RIHS	Not certified	B	Not certified	NA	NA	NA	no
Piattaforma Lessicografica Unica del Tesoro delle Origini (PLUTO)	Italy	E-RIHS	Not certified	B	Not certified	NA	NA	NA	no
Social Sciences and Humanities Data Archive (SOHDA)	Belgium	CESSDA	Not certified	B	Not certified	NA	NA	NA	no

Tárki Data Archive	Hungary	CESSDA	Not certified	B	Not certified	NA	NA	NA	no
Tesoro della Lingua Italiana delle Origini (TLIO)	Italy	E-RIHS	Not certified	B	Not certified	NA	NA	NA	no

List of Figures

List of Tables