Research and Innovation Action

Social Sciences & Humanities Open Cloud

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

Deliverable 8.2 Certification plan for SSHOC repositories

Dissemination Level	PU
Due Date of Deliverable	31/12/19 (M12)
Actual Submission Date	28/02/20
Work Package	WP8 - Governance, Sustainability, Quality Assurance
Task	T8.2 Trust & Quality Assurance
Type	Report
Approval Status	Approved by EC - 03 November 2020
Version	V1.0
Number of Pages	p.1 – p.43

Abstract:

The report provides an overview of the certification of repositories that belong to the SSHOC communities. In addition, the report outlines a certification plan for these repositories. The purpose of this document is to explore the SSHOC trust landscape in order to organise the provision of support in trust and quality work to repositories that belong to communities represented in SSHOC.

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
0.1	04/09/2019	First draft	Henri Ala-Lahti, Mari Kleemola, Tuomas J. Alaterä, Niko Koski
0.2	18/12/2019	Agreed on structure	All authors
0.3	09/01/2020	Further texts and suggestions of certification plans	All authors
0.4	14/01/2020	Adjustments and suggestions, all chapters.	Mari Kleemola, Emiliano Degl'Innocenti, Giulia Cori, Birger Jerlehag, René van Horik, Hervé L'Hours, Niko Koski
0.5	22/01/2020	Conclusion, executive summary, final version of Table 1, final edits before peer review	Tuomas J. Alaterä, Mari Kleemola, Niko Koski
0.6	22/01/2020	Peer review	Cees van der Eijk (University of Nottingham)
0.7	10/02/2020	Address peer review comments, language check, final adjustments	Mari Kleemola, Niko Koski
1.0	11/02/2020	Final version submitted to Coordinator	Niko Koski

Author List

Organisation	Name	Contact Information
CESSDA/FSD	Mari Kleemola	mari.kleemola@tuni.fi
CESSDA/FSD	Tuomas J. Alaterä	tuomas.alatera@tuni.fi
CESSDA/FSD	Niko Koski	niko.koski@tuni.fi
CESSDA/FSD	Henri Ala-Lahti	N/A
CESSDA/SND	Birger Jerlehag	birger.jerlehag@snd.gu.se
CESSDA/UKDS	Hervé L'Hours	herve@essex.ac.uk
CLARIN ERIC	Franciska de Jong	f.m.g.dejong@uu.nl
CLARIN ERIC	Dieter Van Uytvanck	dieter@clarin.eu

DARIAH/PSNC	Tomasz Parkola	tparkola@man.poznan.pl
E-RIHS/CNR	Emiliano Degl'Innocenti	emiliano.deglinnocenti@cnr.it
E-RIHS/CNR	Roberta Giacomi	roberta.giacomi@gmail.com
E-RIHS/CNR	Maurizio Sanesi	maurizio.sanesi.eng@gmail.com
KNAW/DANS	René van Horik	rene.van.horik@dans.knaw.nl

Executive Summary

This report is the first deliverable of Task 8.2 "Trust & Quality Assurance" within WP8 of the SSHOC project. The distributed character of data infrastructures within the SSHOC communities requires developing an agreed approach to assessing the trustworthiness and quality of data repositories. This deliverable provides an overview of Trusted Digital Repository (TDR) standards offering a certification framework for communities represented in the SSHOC project (CESSDA, CLARIN, DARIAH, E-RIHS). Moreover, the deliverable lays the ground for the SSHOC trust work that is needed in order to facilitate the adoption of TDR standards and the FAIR principles in SSH data repositories across the board. In this report, 'trust' refers to the landscape of issues, standards and processes related to trustworthy digital repositories. Trust between all parties in the quality of data and services is critical for research infrastructure in terms of people, processes and technologies. The level of trustworthiness can be assessed through evaluation against agreed requirements.

The SSHOC project unites 20 partner organisations and a further 25 linked third parties. When this report refers to the SSHOC repositories, it means the research data repositories within CESSDA ERIC, CLARIN ERIC, DARIAH ERIC and E-RIHS communities regardless of their participation in the SSHOC project. It is also important to note that in the context of this report, the term 'quality' refers to the technical quality of the repositories i.e. their compliance with the Trusted Digital Repository standards, not to the scientific quality of their digital assets.

In line with the aims of Task 8.2, the report specifies modes of support in building trust and helping repositories reach TDR certification. The report charts the current trust landscape within the SSHOC communities and selects the repositories that will be the main focus of the support activities provided by Task 8.2 at later stages in the project. In addition, the report outlines a certification plan for these repositories. All repositories within the SSHOC communities are potential recipients of support from Task 8.2, but the efforts must be aligned with realistic expectations of progress during the project timeframe.

CoreTrustSeal is selected as the standard TDR certification reference within the task. Due to the diversity of repositories within the SSHOC communities, a flexible yet sustainable approach to trust is needed that is adaptable to a wide variety of data infrastructures. The CoreTrustSeal provides a demonstrable approach to internal and external review, providing a means to determine the strengths and weaknesses of data stewards and a basis for comparison between them. However, certain types of organisations for which the CoreTrustSeal requirements are not applicable are also identified. Task 8.2 helps identify these cases and thus develop the CoreTrustSeal framework to better support a variety of repositories. Further work for Task 8.2 includes the provision of recommendations for sustainably maintaining trust across the SSH ERICs beyond the lifetime of the SSHOC project.

This document is relevant to the SSH ERICs and to repositories across the SSHOC communities. There are no direct dependencies with other SSHOC tasks, but Task 8.2 aligns itself as necessary with both SSHOC tasks and existing EOSC-related efforts promoting trust and the FAIR principles.

Abbreviations and Acronyms

APEF	Archives Portal Europe Foundation
CASRAI	Consortia Advancing Standards in Research Administration Information
CESSDA	Consortium of European Social Science Data Archives
CLARIN	Common Language Resources and Technology Infrastructure
CRM	Conceptual Reference Model
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DSA	Data Seal of Approval
EOSC	European Open Science Cloud
E-RIHS	European Research Infrastructure for Heritage Science
ERIC	European Research Infrastructure Consortium
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation of the European Union (2016/679)
HaS	Humanities at Scale
ICPSR	Inter-university Consortium for Political and Social Research
Iperion-CH	Integrated Platform for the European Research Infrastructure on Cultural Heritage
OAIS	Open Archival Information System
PARTHENOS	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies
RDA	Research Data Alliance
TDR	Trusted Digital Repository
WDS	World Data System

Table of Contents

List of Figures

List of Tables

1. Introduction

Purpose and scope

Research data should be managed, curated, stored and shared in a way that meets expectations regarding trustworthiness and quality, provides sustainability and preserves the investments made in these 'digital assets'. As the EOSC develops its infrastructure and portfolio of services, it will interact with a wide range of stakeholders. In this broad network of actors from different communities, disciplines, traditions and practices, trust is essential. The Trusted Digital Repository (TDR) standards which have emerged from the Open Archival Information System (OAIS) reference model offer a certification solution for repositories.[1] Adopting workflows and guidelines from TDR standards is also a way to assure that the repositories enable the FAIRness of data for the long term.

This report is the first deliverable of Task 8.2 of the Social Science and Humanities Open Cloud (SSHOC) project, and it lays the ground for the SSHOC trust work. The purpose of Task 8.2 is two-fold: firstly, to support the repositories within the SSHOC communities (or in short, "SSHOC repositories") in their work on trust and quality; and secondly, to explore the trust landscape and provide feedback and input to certification bodies from the SSHOC viewpoint. Trust between all parties in the quality of data and services is critical for research infrastructure in terms of people, processes and technologies. The level of trustworthiness can be assessed through evaluation against agreed requirements. In this report 'trust' refers to the landscape of issues, standards and processes related to trustworthy digital repositories

The SSHOC project unites 20 partner organisations and a further 25 linked third parties. When this report refers to the SSHOC repositories, it means the research data repositories within CESSDA ERIC, CLARIN ERIC, DARIAH ERIC and E-RIHS communities regardless of their participation in the SSHOC project.

It is also important to note that in the context of Task 8.2 and this report, the term 'quality' refers to the technical quality of the repositories i.e. their compliance with the Trusted Digital Repository standards, not to the scientific quality of their digital assets.

The current trust landscape within SSHOC communities and selected repositories that will be the main target of the support activities has been charted. The aim of this task is to bolster and improve the

[1] The appendix of this deliverable contains background information on the OAIS reference model and other relevant frameworks for certification.

trustworthiness of the SSHOC repositories, which will in turn ensure that research data will be accessible and FAIR to researchers, both now and in the long run.

The selected standard TDR certification reference is the CoreTrustSeal (CTS), which has already been adopted by CESSDA and CLARIN. The CoreTrustSeal consists of sixteen requirements that reflect the core characteristics of a Trusted Digital Repository. 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 stewards and a basis for comparison between them.

Relation to other tasks and activities

In the SSHOC project, trust and quality issues are included in Work Package 8 which contains 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. There is another task dealing with trust, and that is Task 4.7 *Modeling the SSHOC data life cycle* which aims at developing a common metalevel schema to improve "trust and quality of data and services". In addition, Work Package 9 on data communities aims to create a new repository of election-relevant data and information that might be a candidate for the trust support offered in Work Package 8. However, there is no overlap or direct dependencies with other SSHOC tasks.

Outside the SSHOC project, many initiatives and projects include work on trusted repositories. The most relevant ones for Task 8.2 are the initiatives of the SSHOC communities, and they are described in Chapter 3. Other major initiatives that this task team will follow are the H2020 projects building EOSC[2], and the various trust-related groups within the Research Data Alliance (RDA)[3], but describing them is beyond the scope of this report.

All the ERICs involved in SSHOC have worked with trust and quality issues before this project. This earlier work is described in Chapter 3.

2 European Open Science Cloud. EOSC Projects: https://www.eosc-portal.eu/about/eosc-projects [22 January 2020]
3 Research Data Alliance. Groups. https://www.rd-alliance.org/groups [22 January 2020]

Structure of the document

This document is organised in six sections:

- Sections two and three describe current certification frameworks for data repositories and the frameworks used by repositories belonging to the SSHOC ERICs.
- Sections four and five discuss the CoreTrustSeal, certification plans of selected repositories, and planned modes of support provided to repositories seeking certification.
- Section six presents the conclusions.

This report is addressed towards repositories belonging to the SSHOC cluster. The report is also of value to related EOSC projects such as the regional EOSC projects aimed at increasing the quality of the research infrastructure in a specific European region. The target group also contains users and developers of existing research infrastructures with no specific scientific discipline in mind. A basic level of knowledge of the key principles and standards in relation to quality assurance of repositories is assumed[4].

[4] The appendix of this deliverable provides an overview on the relevant standards for TDR certification.

2. Certification of repositories

Generally, certification refers to confirmation of the characteristics of an object, person, or organisation. In the context of the SSHOC project, certification refers to processes intended to determine whether research data repositories meet specific quality assurance standards which focus on data management, data services and digital preservation. It is a type of quality assurance aimed at digital preservation. Standards, frameworks and guidelines exist that can be used as a reference for certification of repositories.

In this report, the focus is on certification of *data repositories*. There are several ways to define a data repository. Science Europe[5] defines a data repository as a place (data storage system, archive) that holds data sets, makes data sets available to use, and organises them in a logical manner. A data repository may also be defined as an appropriate, subject-specific location where researchers can submit their data sets.[6] Re3data, a global registry of research data repositories, defines a research data repository as "a subtype of a sustainable information infrastructure which provides long-term storage and access to research data that is the basis for a scholarly publication".[7] CoreTrustSeal follows the definition in the CASRAI Dictionary[8]: "Repositories preserve, manage, and provide access to many types of digital materials in a variety of formats. Materials in online repositories are curated to enable search, discovery, and reuse. There must be sufficient control for the digital material to be authentic, reliable, accessible and usable on a continuing basis."

The team involved in Task 8.2 sees 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 the digital material is authentic, reliable, accessible and usable also for the long term.

A research data repository has a service aspect and an organisational aspect that both have their own specific quality features. On the one hand, a well-organised, stable organisation can offer low quality data repository services, and on the other hand, a high-quality data repository service cannot survive without a good organisational embedding.

5 Science Europe association website: https://www.scienceeurope.org/ [26 February 2020]
6 Science Europe Data Glossary. Data Repository. http://sedataglossary.shoutwiki.com/wiki/Data_repository [20 December 2019]
7 Registry of Research Data Repositories: www.re3data.org [22 January 2020]
8 CASRAI Repository: https://dictionary.casrai.org/Repository [22 January 2020]

The European Framework for Audit and Certification of Digital Repositories (2010)[9] consists of a sequence of three levels of certification, in increasing trustworthiness:

- Basic certification is granted to repositories that obtain Data Seal of Approval (DSA) certification. DSA was the predecessor of the CoreTrustSeal requirements. Like DSA, the CoreTrustSeal also provides basic certification.
- Extended certification is granted to basic certification repositories which in addition perform a structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644. The nestor Seal for Trustworthy Digital Archives is based on DIN 31644 and grants an extended certification.
- Formal certification is granted to repositories which in addition to basic certification obtain full external audit and certification based on ISO 16363[10] or equivalent DIN 31644.

These three certification options – CoreTrustSeal, nestor and ISO 16363 – build on the OAIS reference model, and they are relevant for social science and humanities research data repositories. They are all described in more detail in Appendix A.

Other certifications used by SSHOC repositories include, for example, the lightweight IT Service Management FitSM[11] standard which is aimed at certification of the training of IT service management auditors, and the information security standard ISO/IEC 27001.[12] These are beyond the scope of Task 8.2 and this report, but the work in Task 8.2 will monitor the impact on repositories of having to meet a range of standards or criteria in addition to those related to Trusted Digital Repositories.

The standards, assessment and certification environment for repositories and other data stewards is evolving, particularly with regard to the development of the EOSC. The FAIR principles are used to ensure that the EOSC actors cooperate to make data more Findable, Accessible, Interoperable and Reusable (see Wilkinson et al. 2016[13]). The principles themselves are being adapted to agree on a set of 'indicators' of

9 European Framework for Audit and Certification of Digital Repositories website: http://www.trusteddigitalrepository.eu/Trusted%20Digital%20Repository.html [22 January 2020]

10 The ISO standard 16363:2013 is also known as CCSDS 652.0-M-1, Audit and certification of trustworthy digital repositories.

11 FitSM - Standards for lightweight IT Service management website: https://fitsm.eu [22 January 2020]

12 International Organization for Standardization (2013). ISO/IEC 27001:2013. https://www.iso.org/standard/54534.html [22 January 2020]

13 Wilkinson, Mark D; Michel Dumontier; IJsbrand Jan Aalbersberg; Myles Axton; Arie Baak; Niklas Blomberg; … Jun Zhao (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* vol. 3, Article number: 160018. https://doi.org/10.1038/sdata.2016.18 [14 January 2020]

FAIRness (e.g. by the RDA FAIR Data Working Group[14]) but these are focussed on the 'what' of FAIR, not the 'how'. Agreed indicators of FAIRness in digital objects must then be supported by clear metrics and tests designed to manually or automatically check FAIRness by machines or humans. These factors are all in progress, and it has not yet been defined how outcomes of FAIR assessment will be used to govern the integration of data sources into the EOSC. For this reason, this certification plan does not focus on FAIR in detail. However, the developments will be monitored regularly during the SSHOC project and integrated where practical and appropriate. One area of relevant FAIR work is within a work package[15] of the FAIRsFAIR[16] project working toward certification of repositories enabling FAIR data. Work will include mapping the FAIR principles to the CoreTrustSeal Requirements and implementing a maturity model. SSHOC Task 8.2 will retain close contact with the FAIRsFAIR project.

[14] Research Data Alliance (2020). FAIR Data Maturity Model WG. https://www.rd-alliance.org/groups/fair-data-maturity-model-wg [22 January 2020]
[15] FAIRsFAIR. FAIR Certification (of Repositories) - WP4: https://www.fairsfair.eu/fair-certification [22 January 2020]
[16] FAIRsFAIR project website: https://www.fairsfair.eu/ [22 January 2020]

3. Certification status within SSHOC ERICs

CESSDA

The Consortium of European Social Science Data Archives (CESSDA) is composed of member countries, each of which assign a national Service Provider, which is usually a social science data archive or centre. These Service Providers are at the core of CESSDA and cooperate on providing services to researchers. When CESSDA formalised as an ERIC in 2017, the conditions of membership were defined in the "Annex 2 Obligations".[17] These included a reference to certification against a TDR standard.

In 2013, early introductory events for CESSDA members, led by the UK Data Service[18], included an introductory session followed by attendees undertaking a self-assessment against the Data Seal of Approval (DSA)[19] certification. A subsequent workshop reviewed these self-assessments within CESSDA. This pattern of adoption was successful and was extended as part of the H2020 CESSDA Strengthening and Widening (SaW) project in 2016–2017.[20]

During the course of the SaW project, a Research Data Alliance (RDA) Interest Group on Repository Certification identified similarities between the DSA Guidelines and the membership criteria for the World Data System's regular members. The decision was taken to cooperate on developing a single set of 'Core' TDR requirements which became less social science focussed and more generally applicable beyond academia. This cooperation created the CoreTrustSeal process and requirements and widened the community to include more members aligned with the natural sciences. CESSDA members remain active in the CoreTrustSeal 'Assembly of Reviewers' and on the CoreTrustSeal Board.

After the completion of the SaW project, the CESSDA Trust Group[21], which had undertaken the internal teaching, support and review processes, became an independent working group within CESSDA.

During 2018, it was identified that the CESSDA Service Providers were now at a range of maturity in terms of CoreTrustSeal certification. Some are just beginning the journey, some are in progress or have applied, while others have achieved CoreTrustSeal certification and are working on recertification plans. To

17 Statutes of CESSDA ERIC (C/2017/3870) (2017). *Official Journal of the European Union*, C220/1. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017Y0708(01) [14 January 2020]
18 UK Data Service website: https://www.ukdataservice.ac.uk/ [22 January 2020]
19 CoreTrustSeal website: https://www.coretrustseal.org/about/history/data-seal-of-approval/ [22 January 2020]
20 Tjalsma, Heiko; Hervé L'Hours; Mari Kleemola; Natascha Schumann & Janez Štebe (2017). Deliverable 4.4 – Report on DSA Certification for CESSDA. http://cessdasaw.eu/content/uploads/2017/11/SAW-D4.4..pdf [16 January 2020]
21 CESSDA ERIC. Trust Working Group: https://www.cessda.eu/About/Working-Groups/Trust [22 January 2020]

address this, the Trust Group developed workshops focussing on more specific aspects of the CoreTrustSeal requirements and working with a 'round table' discussion mode to encourage knowledge exchange.

In addition to ongoing support, the Trust Working Group is currently looking to identify which evidence could be developed and shared across Service Providers and to provide some oversight of the wide range of trust-related activities including FAIR and the EOSC.

Of the 20 CESSDA Service Providers, nine currently hold CoreTrustSeal certification (January 2020).

CLARIN

The Common Language Resources and Technology Infrastructure (CLARIN) is a distributed network of centres. The backbone of CLARIN is provided by technical centres, in particular Service Providing Centres, or in short: CLARIN B-Centres. These units, often a university or an academic institute, offer the scientific community access to resources, services and knowledge on a sustainable basis. In view of the sustainability there are strict criteria to become recognised as a CLARIN B-Centre: it should be based on a stable technical and institutional foundation. An assessment committee checks these requirements[22] as part of an assessment procedure.[23] One of the requirements for B-centre status is that the centre has concluded or at least initiated the CoreTrustSeal procedure. Currently there are 23 certified B-centres (January 2020)[24] and several more are aiming for this status.

Other centre types are: C-Centres (Metadata Providing Centres; their metadata are integrated with CLARIN but C-Centre status does not require the offering any further services), K-Centres (Knowledge Centres; part of the CLARIN Knowledge Sharing Infrastructure[25]), E-Centres (External Centres, offering central services without being part of any national consortium).

Up-to-date information is available on the CLARIN website.[26]

DARIAH

The Digital Research Infrastructure for the Arts and Humanities (DARIAH) aims to enhance and support digitally-enabled research and teaching across the arts and humanities. The consortium understands

22 CLARIN. Centre requirements: http://hdl.handle.net/11372/DOC-77 [22 January 2020]
23 CLARIN. Assessment procedure: https://www.clarin.eu/content/assessment-procedure [22 January 2020]
24 CLARIN. Certified centres: https://www.clarin.eu/content/certified-centres [22 January 2020]
25 CLARIN. Knowledge centres: https://www.clarin.eu/content/knowledge-centres [22 January 2020]
26 CLARIN. Overview CLARIN centres: https://www.clarin.eu/content/overview-clarin-centres [22 January 2020]

itself as a cooperative undertaking that is carried out by organisations in its Member Countries, DARIAH's Cooperating Partners and several affiliated projects. DARIAH is constituted by its Executive Bodies (General Assembly and Board of Directors), Advisory Bodies (Scientific Board), Coordinating Bodies (National Coordinators Committee, Senior Management Team), Implementation Bodies (Joint Research Committee, Virtual Competence Centres and Working Groups) and Administrative Bodies (DARIAH Coordination Office). A more detailed account is provided in the DARIAH-ERIC statutes.[27] Institutions involved in DARIAH come from various sectors and domains. However, a combination of arts/humanities researchers, computer science experts as well as content providers is clearly visible among the involved parties.

DARIAH has been involved in several activities related to trusted repositories. In 2015, a working group on Certification and Trustworthiness of Repositories was initiated and lead activities for a period of one year. The investigation within the working group was focussed on creating a trusted environment in DARIAH for archiving and disseminating data and/or publications. The concurrently running Humanities at Scale project resulted in a free online Data Deposit Recommendation Service (DDRS).[28] It is a service for humanities researchers in the EU to find suitable research data repositories using the Re3data (Registry of Research Data Repositories[29]). The DDRS recommends suitable repositories based on the researcher's country and discipline. The DDRS further applies filters to the Re3data registry so that it only recommends repositories that have an online data upload option and provide persistent identifiers to make the data citable and findable.

In late 2018, the Heritage Data Reuse Charter blog[30] was initiated to set up principles and mechanisms for improving the use and reuse of cultural heritage data issued by cultural heritage institutions and studied and enriched by researchers. Several European organisations (APEF, CLARIN, DARIAH, Europeana, E-RIHS) and European projects (Iperion-CH, PARTHENOS) are involved. The first public version of the principles provides a background for further elaboration and implementations. They include consideration on trustworthiness, including documentation, provenance and transparency.

[27] DARIAH. Documents: https://www.dariah.eu/about/documents-list/ [22 January 2020]
[28] DARIAH. Data Deposit Recommendation Service: https://ddrs-dev.dariah.eu/ddrs/ [22 January 2020]
[29] Re3data Registry of Research Data Repositories: www.re3data.org [22 January 2020]
[30] Cultural Heritage Data Reuse Charter: Mission Statement: https://datacharter.hypotheses.org/77 [22 January 2020]

There are also ongoing activities related to research data and data repositories, e.g. the Consortium for Open Research Data in the Humanities[31] that deals with CIDOC CRM modelling of humanities research data or related Digital Research Infrastructure at the Max Planck Institute for the History of Science[32].

In the context of the Humanities at Scale (HaS) project (ended 2017), DARIAH developed a new approach to assess and verify DARIAH member contributions. The approach has been implemented and is currently in use. The online contribution platform together with relevant criteria for various types of contributions is described in one of the deliverables of the HaS project[33]. Data-related services are assessed using criteria like maturity level, support level, available documentation and interoperability, which are very much aligned with the trustworthiness of repositories. However, there is no strict rule for the repositories to have a specific certification in this regard. At the time of writing, DARIAH has dozens of contributions related to data hosting services, related tools and portals. Some of them have already gained CoreTrustSeal certification (e.g. Geisteswissenschaftliches Asset Management System[34]) and some of them are investigating this possibility or are in the process of certification (e.g. NAKALA[35]).

E-RIHS

The European Research Infrastructure for Heritage Science (E-RIHS) aims to support researchers in interpreting, preserving, documenting and managing diverse cultural items spread across Europe by providing access to state-of-the-art tools and techniques for the analysis of the material reality of cultural heritage artefacts.[36]

The future E-RIHS ERIC was designed to have quality as one of its main pillars. To ensure its high level throughout the partnership, quality criteria must be met by all organisations and research groups that may state a connection with E-RIHS.

E-RIHS PP Deliverable 2.2 - *Quality Manual and KPIs* (to be delivered by February 2020) will provide a detailed description of the quality system proposed to be adopted by E-RIHS for the quality assessment

31 Consortium for Open Research Data in the Humanities website: https://www.cordh.net/ [22 January 2020]
32 Max Planck Institute for the History of Science (2017). Digital Research Infrastructure. https://www.mpiwg-berlin.mpg.de/research/projects/digital-research-infrastructure [22 January 2020]
33 de Leeuw, Lisa; Femmy Admiraal; Matej Durco; Nicolas Larrousse; Mike Mertens; Francesca Morselli; Mike Priddy; Paulin Ribbe; Carsten Thiel & Lars Wieneke (2017). D5.1 Report on Integrated Service Needs: DARIAH (in kind) contributions – Concept and Procedures. https://github.com/Dans-labs/dariah/wiki/Documents/d5-1.pdf [16 January 2020]
34 Geisteswissenschaftliches Asset Management System (GAMS) website: http://gams.uni-graz.at [22 January 2020]
35 NAKALA website: https://www.nakala.fr/ [22 January 2020]
36 E-RIHS website: http://www.e-rihs.eu/ [22 January 2020]

of prospective new partners and their services and for the quality audit of existing E-RIHS partners and their services. It will also outline the process to grant external organisations, services, projects and proposals the affiliation to E-RIHS, or its support. All such procedures will be based on a modular operation: the evaluation of the candidate's internal processes, of its scientific excellence and of the quality of its services and eventual suitability for E-RIHS.

At the present state of E-RIHS, Quality management is being designed on internal assessments to be performed by domain experts (E-RIHS being a cross-domain infrastructure) and by members of the E-RIHS Advisory Board. The evaluation by assessments will be based on a set of documents, intended to represent a guide for both the evaluatee and the evaluator. The parameters to be assessed will be varied: the candidate's internal processes, its scientific authority and the range of services it provides. A Quality and Risk Unit will also be established in the Head Office.

E-RIHS includes various partners and is divided into four platforms: FIXLAB, MOLAB, ARCHLAB and DIGILAB.[37] Of these, the DIGILAB platform is most relevant for SSHOC Task 8.2. The main goal of DIGILAB is that of maintaining and making available cultural items on a large scale. For this purpose, the platform will design and implement a highly interoperable and integrated digital ecosystem to provide researchers and experts with digital resources (tools, services, data).

DIGILAB advocates a collaborative approach to research. It aims to act as a collaborative communication hub and knowledge repository, making available information about best practices, current research, standards and protocols, policies, how-tos and instructions, including training material (i.e. webinars and guides).

To reach this status, DIGILAB intends to constitute a community fostering awareness among researchers about data sharing and reuse. The platform will set up guidelines for dataset recovery and will assist researchers and institutions willing to undertake this task; further assistance may be offered to partners in designing their local repositories.

DIGILAB will select the most suitable best practices and standards (i.e. CIDOC CRM ontology[38] – ISO 21127:2006) to manage the digital assets within the e-infrastructure. Every service provided by DIGILAB will be FAIR-oriented and compliant with EU policies and strategies concerning scientific data.

[37] E-RIHS. About: http://www.e-rihs.eu/about/about/ [27 February 2020]
[38] CIDOC CRM website: http://www.cidoc-crm.org/ [22 January 2020]

4. Certification plans

Why CoreTrustSeal for SSHOC?

There are a number of justifications in the SSHOC context for focussing on the CoreTrustSeal as the TDR certification framework for repositories. These arguments go beyond the fact that the CoreTrustSeal has been explicitly selected as a goal by CESSDA, set as a basis for CLARIN B status, and used as a reference point for Trust work throughout the DARIAH ERIC, though it should also be noted that E-RIHS is actively investigating best practice approaches and that all of the SSH ERICs have trust-related criteria which are independent of the CoreTrustSeal. Some of the key justifications behind opting for the CoreTrustSeal are that it is community- and human-driven and open to feedback, is less labour intensive compared to other standards, has a large number of active certifications, and there is openly accessible documentation available. Within the concurrent FAIRsFAIR project, work has already been done to align the CoreTrustSeal with the FAIR principles. These and further justifications are elaborated below.

The requirements which resulted in the CoreTrustSeal were fundamentally community-driven and represent a consensus on the 'core' of trust criteria as influenced by the OAIS reference model, ISO 16363 [39] and nestor Seal (see Appendix A). Together these provide a terminology base and archive/repository organisational scope. Identifying exactly what is to be certified is critical to designing a standard and applying an evaluation process.

One of the criteria set for developing the CoreTrustSeal through the RDA[40] was to develop a sustainability solution. The creation of a not-for-profit foundation with a charging model (€1000 over three years) provides that sustainability while the underlying organisational and process structure of the CoreTrustSeal remains community-driven.

The heterogeneity of the SSHOC partners and associated organisations, and the active work within CoreTrustSeal to consider the widest possible range of certification candidates are also relevant. The SSHOC support process will help recognise variation in organisation type and develop proposals of how the different types of organisations may be best supported by trust standards. The conclusions of SSHOC can be fed back into the CoreTrustSeal change management process to influence future requirements. This circle of change benefits the SSH ERICs and the wider data management community.

[39] Audit and certification of trustworthy digital repositories

[40] Research Data Alliance. RDA/WDS Certification of Digital Repositories IG: https://www.rd-alliance.org/groups/rdawds-certification-digital-repositories-ig.html [22 January 2020]

For heterogeneous data repositories and services, the CoreTrustSeal offers a flexible, human-driven assessment and evaluation process. Explicit conformity tests with binary pass or fail outcomes depend on very clear and commonly agreed criteria. Binary outcome tests cannot be based on vague notions of 'best practice' but depend on clear definitions e.g. of 'minimal' expected practice. Wherever common (ideally machine-testable) criteria can be agreed upon, they should be defined and applied, but the CoreTrustSeal model provides human-mediated supportive feedback based on the local data and organisational context of the applicant.

The review period of the CoreTrustSeal is also relevant; certification requirements are revised every three years in comparison to every five years for DIN/ISO standards. The CoreTrustSeal has been recently revised in an open call for feedback, so all interested parties have had an opportunity for input. The time scale of SSHOC fits well into contributing to the next CoreTrustSeal revision which will take place in 2022.

In common with other TDR certifications, the CoreTrustSeal offers an OAIS-based TDR approach. OAIS has also gone through an open revision process which will be published during SSHOC[41]. The changes based on feedback from the broad data archiving and repository community will be evaluated and integrated.

The number of past certifications and their availability is also an important factor in adopting a standard. By mid-January 2020, ISO 16363:2012[42] has 2 certified repositories, with 6 for TRAC (via CRL[43]) while nestor[44] has certified 4 repositories against DIN 31644[45]. At the time of writing, CoreTrustSeal has certified over 70 repositories. Furthermore, there is a large number of ongoing new certification processes, and past certification candidates from DSA and WDS may update to the CoreTrustSeal. This rate of adoption means that there is a range of prior information available to SSHOC partners and trust support candidates.

41 CCSDS The Consultative Committee for Space Data Systems (2019). Reference Model for an Open Archival Information System (OAIS). Recommended practice. CCSDS 650.0-P-3, Pink (Pre-Magenta) Book September 2019: https://cwe.ccsds.org/moims/_layouts/15/WopiFrame.aspx?sourcedoc=/moims/docs/MOIMS-DAI/Draft%20Documents/OAIS%20v3/OAIS%20final%20v3%20draft%20with%20changes%20wrt%20OAISv2%2020190924-rl.docx&action=default [22 January 2020]

42 ISO 16363. Certified clients: http://www.iso16363.org/iso-certification/certified-clients/ [22 January 2020]

43 Center for Research Libraries. Certification and Assessment of Digital Repositories: https://www.crl.edu/archiving-preservation/digital-archives/certification-assessment [22 January 2020]

44 Nestor. Evaluated Archives: https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html [22 January 2020]

45 Beuth. DIN 31644:2012-04: https://www.beuth.de/en/standard/din-31644/147058907 [22 January 2020]

The CoreTrustSeal governance and review process is openly documented and provides for a relatively accessible self-assessment, review, feedback and certification model. All CoreTrustSeal certifications are published and thus provide a growing knowledge base of Trusted Digital Repository practice.

All evaluation processes have resource implications for the applicant. Any ISO standard which allows certification has its own ISO-documented process (for ISO 16363 this is ISO 16919[46]) which must in turn align with the requirements for bodies providing audit and certification of management systems.[47] The process includes site visits and periodic internal and external audits. The ISO process is undoubtedly more rigorous and detailed than the equivalent defined by the CoreTrustSeal *Statutes and Rules of Procedure*[48] but it is, therefore, harder to replicate within a project environment like SSHOC. Donaldson et al.[49] identify candidates for the CoreTrustSeal precursor (the similarly scoped and structured Data Seal of Approval) who have taken some days or weeks to develop their self-assessment statements and associated evidence. A recent presentation from ICPSR stated that the update from their previous certification to the CoreTrustSeal took less than a week.[50] The *Survey on DSA-certified digital repositories* report[51] notes that by renewing the CoreTrustSeal the repositories are also able to demonstrate their progress. The report also summarises further benefits of certification.

Work to align the CoreTrustSeal with the FAIR principles and the needs of the European Open Science Cloud[52] is being undertaken in parallel to SSHOC, particularly through FAIRsFAIR[53] which will offer

[46] International Organization for Standardization (2014). ISO 16919:2014: https://www.iso.org/standard/57950.html [22 January 2020]

[47] International Organization for Standardization (2015). ISO 17021-1:2015: https://www.iso.org/standard/61651.html [22 January 2020]

[48] Dillo, Ingrid; Herve L'Hours & Mari Kleemola (2018). CoreTrustSeal Foundation Statutes and Rules of Procedure: https://doi.org/10.5281/zenodo.1142960 [22 January 2020]

[49] Donaldson, Devan Ray; Ingrid Dillo; Robert Downs & Sarah Ramdeen (2017). The Perceived Value of Acquiring Data Seals of Approval. International Journal of Digital Curation 12:1, 130–151. https://doi.org/10.2218/ijdc.v12i1.481 [14 January 2020]

[50] Lyle, Jared (2019). Data Repository Assessment & Certification at ICPSR: Experiences and Lessons Learned. Presentation at NIH Trustworthy Data Repositories workshop in Rockville, Maryland, 8 April, 2019. https://datascience.nih.gov/sites/default/files/20190408_NIH_repo_certification_Jared.pdf [14 January 2020]

[51] Waterman, Kees; Sierman, Barbara (2016). Survey on DSA-certified digital repositories. Report on the findings in a survey of all DSA-certified digital repositories on investments in and benefits of acquiring the Data Seal of Approval (DSA). https://doi.org/10.5281/zenodo.1188256 [22 January 2020]

[52] European Commission. European Open Science Cloud (EOSC): https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud [22 January 2020]

[53] FAIRsFAIR website: https://www.fairsfair.eu/ [22 January 2020]

associated support to applicants. This provides an opportunity for SSHOC to develop repository support in a way that aligns with future requirements for data stewards to enable FAIR data.

The *Turning FAIR into Reality* report[54] notes that the CoreTrustSeal provides an important foundational certification that ensures the quality of key responsibilities and criteria aligned with and supportive of the FAIR principles. The report proposes that "all data repositories are certified according to existing community-vetted criteria such as the CoreTrustSeal".

Selection of repositories

All repositories within the SSHOC communities are potential recipients of support from Task 8.2, but the support efforts must be aligned with realistic expectations of progress during the project timeframe. Achieving CoreTrustSeal certification depends on evidence statements (self-assessments) and the publicly available supporting documentation, both of which require resources and work from the repositories. Below it is briefly described, community by community, how the team has selected the repositories listed in Table 1. The list of repositories will be updated regularly during the SSHOC project as the landscape evolves and new repositories or infrastructures emerge.

CESSDA repositories

One of the internal working groups within CESSDA, the Trust WG, periodically monitors the certification progress and targets of CESSDA Service Providers (SPs). This information has been used to categorise the CESSDA members with a view to offering support. The CESSDA Trust Group offers confidential, internal review of draft self-assessments prior to submission; a process which will be aligned with SSHOC for the duration of the project. In the early stages, CESSDA Trust support provided generic overviews of the requirements and process. As the working group has evolved, it has charted other possible support approaches for different CESSDA Service Providers progressing at different speeds.[55] These differences typically depend on local organisational issues, repository age/maturity and the level of resources available both to offer services and to manage change. All CESSDA Service Providers are included in Table 1.

[54] Collins, Sandra; Françoise Genova; Natalie Harrower; Simon Hodson; Sarah Jones; Leif Laaksonen; Daniel Mietchen; Rūta Petrauskaitė & Peter Wittenburg (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. https://publications.europa.eu/s/kcCD [16 January 2020]

[55] L'Hours, Hervé; René van Horik; Mari Kleemola; Jonas Recker; Janez Štebe & Birger Jerlehag (2020). CESSDA Trust Group: Overview of Support Approaches. http://doi.org/10.5281/zenodo.3621378 [22 January 2020]

CLARIN repositories

It is expected from national CLARIN consortia to have at least one certified centre for each country. In principle, relatively little support from SSHOC would be required, given the role of the CLARIN Centre Committee, which is the platform where on a regular basis potential issues with regards to certification are discussed. However, there are two categories of CLARIN repositories that could benefit from such support: newcomers (i.e. countries that are not yet a member of CLARIN ERIC) and countries that have been postponing the establishment of a repository for several years. For these cases, some focussed concerted action – aimed at answering questions and even assisting in filling out the certification requirements self-assessment, could be beneficial. There certainly are candidates among the 23 currently certified CLARIN centres that could provide hands-on coaching, next to or in the context of the support to be provided by SSHOC.

DARIAH repositories

DARIAH repositories are very diverse. They differ in the context of the content, technical formats as well as standards and protocols used to make the data accessible. Based on latest statistics, there are 39 repositories providing access to resources such as corpora of textual data (e.g. French text messages) or cultural heritage content (archives, museums, libraries). They include also very specific resources such as literary bibliography, dictionaries, music database or encyclopedia. The heterogeneity makes it very challenging to apply or enforce common standards and approaches for data delivery, accessibility and reuse. Therefore, DARIAH is in an ongoing process to define strategy in this regard, with already undertaken actions mentioned in Section 3 of this report. In addition, there is an ongoing discussion in DARIAH on setting up a new working group focussed on research data management. It will help DARIAH in facilitating access to and exchange of data, identifying researchers' needs, exploring common approaches & guidelines (e.g. FAIR) as well as promoting usage of core services (e.g. persistent identifier services) in the research data management activities.

There are several DARIAH repositories that are interested in CoreTrustSeal certification: they either use the CoreTrustSeal, are in the process of obtaining it or plan to do so. Repositories that are currently not investigating CoreTrustSeal certification will be encouraged via actions raising awareness. Certification plan for the DARIAH repositories is included in Table 1.

E-RIHS repositories

E-RIHS will provide access to a wide range of cutting-edge tools, datasets and instruments for Heritage Science research, analysis and interpretation. It is therefore of fundamental importance that resources selected, created and used by scientists are managed, curated, and archived in a way that preserves their value and trustworthiness. The repository certification process will represent an important contribution to ensure the reliability of the E-RIHS resources increasing the potential for sharing data over a long period of time and their value. E-RIHS partnership joins 16 countries (15 EU Member States plus Israel), 2 ERICs and 3 institutions representing scientific communities. Each E-RIHS partner manages different

repositories. In this wide horizon the task 8.2 team selected a small cluster of repositories that are at the core of the E-RIHS digital platform (DIGILAB), both because of their scientific relevance and technical maturity level. Encouraging the selected resources (see Table 1) to join the CoreTrustSeal certification process will represent an important step towards the implementation of the DIGILAB platform as a digital ecosystem for Heritage Science, by focussing on specific aspects (e.g. data FAIRness etc.).

Certification goals for repositories

Based on the 8.2 task team's analysis on the certification landscape within the SSHOC domains, the team agreed on the following three certification goals for the repositories:

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

The certification goals as well as the list of repositories are expected to evolve during the SSHOC project[56]. For the repositories included in this report (see Table 1 below), an ambitious goal of (A) renewal of certification or (B) new certification has been set, but the outcome may also be (C) self-assessment if certification is not feasible within the timeframe of the project. Likewise, even if the certification goal were set as (C) self-assessment, the repository may be recommended to proceed to formal CoreTrustSeal certification. Another scenario is that the repository is not in scope of the CoreTrustSeal, in which case the information will be used for a gap analysis of CoreTrustSeal requirements that do not fit the candidate type.

TABLE 1. SELECTED REPOSITORIES AND THEIR CERTIFICATION GOALS. THE TABLE INCLUDES REPOSITORIES WITHIN THE SCOPE OF T8.2 FOR WHICH A CERTIFICATION GOAL CAN, AT THIS STAGE, BE ESTABLISHED.

Repository name	Country	Community	Existing Certificate	Goal
ACDH - A Resource Centre for the Humanities (ACDH-ARCHE)	Austria	CLARIN	CoreTrustSeal v2017-2019	A
ASV Leipzig	Germany	CLARIN	CoreTrustSeal v2017-2019	A

[56] Potential candidates for trust support include, for example, the new repository of election-related data and information that will be created by Task 9.3, or the forthcoming EURhisFIRM infrastructure in the field of historical economy (https://eurhisfirm.eu/index.php/the-project-2/) which has been invited to join SSHOC.

Repository name	Country	Community	Existing Certificate	Goal
Austrian Social Science Data Archive (AUSSDA)	Austria	CESSDA	Not certified	B
Bayerisches Archiv für Sprachsignale (BAS)	Germany	CLARIN	CoreTrustSeal v2017-2019	A
Berlin-Brandenburg Academy of Sciences and Humanities (BBAW)	Germany	CLARIN	CoreTrustSeal v2017-2019	A
Centre of Estonian Language Resources (CELR-EKK)	Estonia	CLARIN	CoreTrustSeal v2017-2019	A
CLARIN-PL Language Technology Centre	Poland	CLARIN	CoreTrustSeal v2017-2019	A
CLARIN.SI Language Technology Centre (CLARINSI)	Slovenia	CLARIN	DSA	B
CLARINO Bergen Centre	Norway	CLARIN	CoreTrustSeal v2017-2019	A
CMU-TalkBank (CMU)	USA	CLARIN	CoreTrustSeal v2017-2019	A
Corpus testuale OVI	Italy	E-RIHS	Not certified	B
Czech Social Science Data Archive (CSDA)	Czech Republic	CESSDA	CoreTrustSeal v2017-2019	A
Danish National Archives (DNA)	Denmark	CESSDA	Not certified	B
Data Archiving and Networked Services (DANS)	Netherlands	CESSDA	CoreTrustSeal v2017-2019	A
Data Centre Serbia for Social Sciences (DCS)	Serbia	CESSDA	Not certified	B
Digital Repository of Ireland (DRI)	Ireland	DARIAH	CoreTrustSeal v2017-2019	A
Digital Repository of Scientific Institutes (DRSI)	Poland	DARIAH	Not certified	B
Eberhard-Karls-Universität Tübingen (EKUT)	Germany	CLARIN	CoreTrustSeal v2017-2019	A
Finnish Social Science Data Archive (FSD)	Finland	CESSDA	CoreTrustSeal v2017-2019	A
Geisteswissenschaftliches Asset Management System (GAMS)	Germany	DARIAH	CoreTrustSeal v2017-2019	A
GESIS - Leibniz Institute for the Social Sciences	Germany	CESSDA	CoreTrustSeal v2017-2019	A

Repository name	Country	Community	Existing Certificate	Goal
Greek research infrastructure for the social sciences (So.Da.Net)	Greece	CESSDA	Not certified	B
Hamburger Zentrum für Sprachkorpora (HZSK)	Germany	CLARIN	CoreTrustSeal v2017-2019	A
Institut für Deutsche Sprache (IDS)	Germany	CLARIN	CoreTrustSeal v2017-2019	A
Institut für Maschinelle Sprachverarbeitung (IMS)	Germany	CLARIN	CoreTrustSeal v2017-2019	A
Instituut voor de Nederlandse Taal (IVDNT)	Dutch Language Union	CLARIN	CoreTrustSeal v2017-2019	A
LINDAT/CLARIN (LINDAT)	Czech Republic	CLARIN	CoreTrustSeal v2017-2019	A
Meertens Instituut/HUC (MI)	The Netherlands	CLARIN	CoreTrustSeal v2017-2019	A
MObile-laboratory VIsualization DAta (MOVIDA)	Italy	E-RIHS	Not certified	B
MPI for Psycholinguistics (MPI-PL)	The Netherlands	CLARIN	CoreTrustSeal v2017-2019	A
NAKALA	France	DARIAH	Not certified	B
Norwegian Centre for Research Data (NSD)	Norway	CESSDA	CoreTrustSeal v2017-2019	A
Piattaforma Lessicografica Unica del Tesoro delle Origini (PLUTO)	Italy	E-RIHS	Not certified	B
Portuguese Social Information Archive (APIS)	Portugal	CESSDA	Not certified	B
PORTULAN CLARIN	Portugal	CLARIN	CoreTrustSeal v2017-2019	A
PROGEDO Research Infrastructure	France	CESSDA	Not certified	B
Slovak Archive of Social Data (SASD)	Slovakia	CESSDA	Not certified	B
Social Science Data Archives (ADP)	Slovenia	CESSDA	CoreTrustSeal v2017-2019	A
Social Sciences and Humanities Data Archive (SOHDA)	Belgium	CESSDA	Not certified	B
Språkbanken, The Swedish language bank	Sweden	CLARIN	DSA	B

Repository name	Country	Community	Existing Certificate	Goal
Swedish National Data Service (SND)	Sweden	CESSDA	CoreTrustSeal v2017-2019	A
Swiss Centre of Expertise in the Social Sciences (FORS)	Switzerland	CESSDA	CoreTrustSeal v2017-2019	A
Tárki Data Archive	Hungary	CESSDA	Not certified	B
Tesoro della Lingua Italiana delle Origini (TLIO)	Italy	E-RIHS	Not certified	B
The CLARIN Centre at University of Copenhagen (CLARIN-DK-UCPH)	Denmark	CLARIN	CoreTrustSeal v2017-2019	A
The ILC4CLARIN Centre at the Institute for Computational Linguistics (ILC4CLARIN)	Italy	CLARIN	CoreTrustSeal v2017-2019	A
The Language Bank of Finland (FIN-CLARIN)	Finland	CLARIN	CoreTrustSeal v2017-2019	A
UK Data Service (UKDS)	United Kingdom	CESSDA	DSA	B
Universität des Saarlandes (UdS)	Germany	CLARIN	CoreTrustSeal v2017-2019	A

Modes of support

Our plan is to work on the most generally applicable issues at the higher level, community-specific at the next level, and then to apply repository-specific actions at the local level. As a rule, the task 8.2 team will also make the support materials as openly available as possible.

The following main modes of support will be provided by Task 8.2:

- Raising awareness of the relevance of certification of repositories and the role of the CoreTrustSeal in this via existing communication channels. The information will be openly available to everyone.
- Workshops and webinars on repository certification. The task 8.2 team aims to have at least one event targeted at each SSHOC community. All events will, however, be open to all repositories (but especially in the case of face-to-face events subject to availability of space).
- Review of self-assessments based on the requirements of the CoreTrustSeal. Detailed feedback will be provided on self-assessments primarily to the selected repositories (see Table 1). The self-assessments and the feedback will be confidential and not published unless otherwise agreed.

Key evidence statements may also be developed for selected groups of repositories with a view to future certification. Experiences from the CESSDA trust work have shown that starting with general overviews and discussion work well when most are new to the CoreTrustSeal and certification. This can then be followed up by more targeted topical discussions focussing, for example, on certain CoreTrustSeal requirements.

The level of SSHOC support required will depend on the starting state of the selected candidates. Organisations with clear, standardised processes and procedures or technical systems documentation will find it easier to develop and maintain an evidence base. The team will amend the modes of support according to the repositories' needs.

The highest proportion of effort will be devoted to those organisations seeking their first certification, but there is also value in working with repositories seeking updated or renewed certification. A smooth recertification processes depends on ongoing operational management of business information to support quality services. If these information and change management practices are in place to ensure service quality, the updates necessary to meet periodic recertification are minimal. The experience of organisations which have been through recertification may help inform business information systems design and management for less mature organisations.

The primary support approach will be based on iterative reviews of draft self-assessments by the Task 8.2 team members. Members not experienced in CoreTrustSeal reviews will be paired with a 'buddy' to help them learn and to ensure consistency. Direct support requests will be possible outside of the self-assessment reviews, but these represent an unpredictable burden on project resources. It is

acknowledged that until the process begins, the level of experience of repositories and their staff members is uncertain. It will be taken into consideration that large-scale amendments to technical infrastructure or data pipelines is unlikely during the project timeframe. A primary target for improvement will be to offer support in creating or updating evidence for the CoreTrustSeal requirements and in designing evidence management systems which allow repositories to retain certification over time.

5. Discussion

Challenges with chosen certification standard

The aim of SSHOC Task 8.2 is to improve the trustworthiness of the SSHOC repositories, and the selected TDR certification standard is the CoreTrustSeal which consists of 16 requirements. This approach comes with limitations and challenges of which need to be taken into consideration.

The CoreTrustSeal provides a flexible approach which can be applied to a wide range of repositories and data infrastructure service providers, but this does not ensure that the CoreTrustSeal will be perfectly fit for purpose for all SSHOC repositories and data services. The CoreTrustSeal requirements come from a domain/disciplinary repository background and focus on the idea of serving a clear (OAIS-defined) 'Designated Community'. CoreTrustSeal currently engages with potential applicants with a view to supporting a wider range of more general-purpose galleries, libraries, archives and museums, as well as complex partnership models and entities providing technical infrastructure services. Work with the heterogenous SSHOC applicants may help inform development to support the application of the CoreTrustSeal to the full range of data stewards.

In addition, repositories are not the only entities that can or should be certified. The *Turning FAIR into Reality* report identifies several other possible candidates for assessment and certification including FAIR objects, services, registries and software[57].

CoreTrustSeal, nestor Seal, and ISO 16363 all support the evaluation of organisations which claim OAIS-conformance as 'Trusted Digital Repositories'. But none of these standards provide a detailed business process/workflow model for data deposit, curation, access or storage. One outcome of SSHOC repository support may be to identify where a more granular set of standard approaches is desirable.

Engagement with other EOSC-related efforts will help ensure that the SSHOC repositories can assess against the standards which most fit their needs. Work on certification plans for SSHOC repositories can also help identify areas where organisations in scope for SSHOC do not align with the CoreTrustSeal.

[57] Collins, Sandra; Françoise Genova; Natalie Harrower; Simon Hodson; Sarah Jones; Leif Laaksonen; Daniel Mietchen; Rūta Petrauskaitė & Peter Wittenburg (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. https://publications.europa.eu/s/kcCD [16 January 2020]

Certification goals

Our repository selection process showed that while some of the data service organisations related to SSHOC are traditional CoreTrustSeal applicants, not all of them are clear candidates. Others may not have equated their current data services with TDR concepts at all. In all cases, these organisations remain critical EOSC components as either potential Trusted Digital Repositories or as service providers which support the repositories. For these circumstances, there are several possible approaches.

If the work in the task indicates that a repository/service is not eligible for the CoreTrustSeal, then the team may be able to:

- recommend some other assessment/certification method that is more appropriate, and
- make recommendations to CoreTrustSeal for changes to their requirements or scope to make the certification more inclusive.

If the CoreTrustSeal requirements or the associated assessment processes are not sufficiently rigorous for some repositories, it is also possible to use CoreTrustSeal as a first step towards ISO 16363 certification.

Wherever possible, the task 8.2 team will seek to make self-assessments (or anonymised extracts) public even if they do not reach the point of submission to CoreTrustSeal. This kind of information sharing has several benefits including the fact that some CoreTrustSeal requirements are applicable to any data stewardship organisation, and that any 'gap analysis' of assessment support indicates a possible area of EOSC partnerships which do not have supporting standards.

Aligning the CoreTrustSeal with other trust/quality expectations for other communities/domains has potential benefits for certification providers and applicants.

Alignment with wider compliance requirements

As noted elsewhere in this planning document and in 'Evolving landscape' below, the repositories within the SSHOC communities are subject to a range of other compliance expectations, both existing and in development. Beyond the CoreTrustSeal, for CESSDA there are 'Annex 2' obligations for membership, CLARIN has a number of 'centre' types, DARIAH has a wide range of members and E-RIHS is actively evolving their quality standards. One outcome of SSHOC support may be to identify gaps in or overlaps between these standard requirements.

All compliance depends on a degree of change management of processes and supporting evidence which can range from risk assessment for information security to demonstrating 'appropriate technical and

organisational measures' in line with the GDPR.[58] One goal of SSHOC is to support compliance and certification as a sustainable part of data stewardship practice. The alternative is that standards compliance becomes a periodic update to supporting evidence rather than a part of ongoing management, service delivery and improvement. SSHOC repositories' dependence on evidence shared across ERICs or dependence on third party service providers also indicates that some consideration of business information management could reduce duplication of effort and overall compliance bureaucracy.

Designing systems to manage evidence that meets multiple standards is challenging, but important for minimising the resource expenditure on standards compliance. Wherever possible, the evidence for meeting standards should align with the business information necessary to deliver high quality services.

During the implementation of the support plan, the task 8.2 team will investigate current practice at organisational and ERIC level and make recommendations for coordinated evidence management to support multiple compliance requirements.

Evolving landscape

Repositories, processes, certifications do not exist in isolation. A recent draft paper from the FAIRsFAIR project presents a vision for the FAIR ecosystem components required to ensure FAIRness across the full data lifecycle.[59]

The FAIR focus on digital objects has rapidly evolved to acknowledge that FAIRness also depends on data stewardship environments that can enable FAIR and ensure objects remain FAIR over time. In turn, this suggests that an understanding is needed of the digital objects being curated in different collections to evaluate whether their data stewards are fit for purpose.

The evolving landscape means that there is also a need to collect and share information on repositories and their practices. To avoid creating many separate registries, repositories are recommended to create Re3data entries or register in community-specific registries that automatically export to Re3data, like the CLARIN centre registry.[60] One possible outcome of Task 8.2 could be feedback on the Re3data metadata

[58] European Parliament (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [11 February 2020]
[59] L'Hours, Hervé & Ilona von Stein (2019). FAIR Ecosystem Components: Vision. http://doi.org/10.5281/zenodo.3565428 [21 January 2020]
[60] CLARIN. Centre Registry: https://centres.clarin.eu [27 February 2020]

schema, for example on unifying the Re3data and CoreTrustSeal repository types to meet the needs of those selecting repositories or searching for data.

Further work

After analysing the repositories further, it is possible, for example, to recommend proceeding to the CoreTrustSeal, or to use the information as input for updated typology of potential candidates or for gap analysis of CoreTrustSeal requirements that do not fit the candidate types.

Based on findings from forthcoming workshops, webinars, self-assessments, collaborations within SSHOC and other initiatives (like EOSC and RDA) and desk research, the task 8.2 team will continue the work and discussion in all the areas described here and address emerging issues. The aim is to provide recommendations for sustainably maintaining trust across the SSH ERICs beyond the lifetime of the SSHOC project. SSHOC Deliverable 8.3 (due in M38) will provide an update of the certification status of the SSHOC repositories and suggest certification solutions for them. For Deliverable 8.3, the relationship with any governance aspects (SSHOC Task 8.1) will also be analysed.

6. Conclusions

Data repositories are 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 the digital material is authentic, reliable, accessible and usable also for the long term. Repositories have a service aspect and an organisational aspect that both have their own specific quality features. The standards, assessment and certification environment for repositories and other data stewards is evolving, particularly with regard to the development of the EOSC.

The aim of SSHOC Task 8.2 is to improve the trustworthiness of the SSHOC repositories, and the selected certification standard is the CoreTrustSeal which consists of 16 requirements. The 8.2 task team will provide SSHOC repositories with support for achieving their certification goal which may vary from self-assessment to obtaining or renewing CoreTrustSeal certification.

The benefits of self-assessment and peer review against an agreed standard do not depend on formal certification. Undertaking the process improves internal communications and understanding as well as facilitates comparison to peer organisations and cooperation with partner organisations. The SSHOC trust work will include organisations at different levels of maturity, and the different levels of feasibility for the various goals within the project timeframe are quite diverse.

An important result of the process that has led to the overview presented here and of the steps ahead is that they help identify cases where the CoreTrustSeal requirements are ‘not applicable’ (and why), and where changes to the CoreTrustSeal framework could lead to better support for a variety of repositories.

Further work for Task 8.2 also includes providing recommendations for sustainably maintaining trust across the SSH ERICs beyond the lifetime of the SSHOC project.

7. References

Beuth. DIN 31644:2012-04: https://www.beuth.de/en/standard/din-31644/147058907 [22 January 2020]

CASRAI Repository: https://dictionary.casrai.org/Repository [22 January 2020]

CCSDS The Consultative Committee for Space Data Systems (2012). Audit and Certification of Trustworthy Digital Repositories. Recommended practice. CCSDS 652.0-M-1, Magenta Book September 2011: https://public.ccsds.org/pubs/652x0m1.pdf [22 January 2020]

CCSDS The Consultative Committee for Space Data Systems (2012). 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 [22 January 2020]

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 [22 January 2020]

CCSDS The Consultative Committee for Space Data Systems (2019). Reference Model for an Open Archival Information System (OAIS). Recommended practice. CCSDS 650.0-P-3, Pink (Pre-Magenta) Book September 2019: https://cwe.ccsds.org/moims/_layouts/15/WopiFrame.aspx?sourcedoc=/moims/docs/MOIMS-DAI/Draft%20Documents/OAIS%20v3/OAIS%20final%20v3%20draft%20with%20changes%20wrt%20OAISv2%2020190924-rl.docx&action=default [22 January 2020]

Center for Research Libraries. Certification and Assessment of Digital Repositories: https://www.crl.edu/archiving-preservation/digital-archives/certification-assessment [22 January 2020]

CESSDA ERIC. Trust Working Group: https://www.cessda.eu/About/Working-Groups/Trust [22 January 2020]

CIDOC CRM website: http://www.cidoc-crm.org/ [22 January 2020]

CLARIN. Assessment procedure: https://www.clarin.eu/content/assessment-procedure [22 January 2020]

CLARIN. Centre Registry: https://centres.clarin.eu [27 February 2020]

CLARIN. Centre requirements: http://hdl.handle.net/11372/DOC-77 [22 January 2020]

CLARIN. Certified centres: https://www.clarin.eu/content/certified-centres [22 January 2020]

CLARIN. Knowledge centres: https://www.clarin.eu/content/knowledge-centres [22 January 2020]

CLARIN. Overview CLARIN centres: https://www.clarin.eu/content/overview-clarin-centres [22 January 2020]

Collins, Sandra; Françoise Genova; Natalie Harrower; Simon Hodson; Sarah Jones; Leif Laaksonen; Daniel Mietchen; Rūta Petrauskaitė & Peter Wittenburg (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. https://publications.europa.eu/s/kcCD [16 January 2020]

Collins, Sandra; Françoise Genova; Natalie Harrower; Simon Hodson; Sarah Jones; Leif Laaksonen; Daniel Mietchen; Rūta Petrauskaitė & Peter Wittenburg (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. https://publications.europa.eu/s/kcCD [16 January 2020]

Consortium for Open Research Data in the Humanities website: https://www.cordh.net/ [22 January 2020]

CoreTrustSeal (2019). Core Trustworthy Data Repositories Requirements 2020–2022: https://zenodo.org/record/3638211 [27 January 2020]

CoreTrustSeal website: https://www.coretrustseal.org/ [22 January 2020]

CoreTrustSeal website: https://www.coretrustseal.org/about/history/data-seal-of-approval/ [22 January 2020]

Cultural Heritage Data Reuse Charter: Mission Statement: https://datacharter.hypotheses.org/77 [22 January 2020]

DARIAH. Data Deposit Recommendation Service: https://ddrs-dev.dariah.eu/ddrs/ [22 January 2020]

DARIAH. Documents: https://www.dariah.eu/about/documents-list/ [22 January 2020]

de Leeuw, Lisa; Femmy Admiraal; Matej Durco; Nicolas Larrousse; Mike Mertens; Francesca Morselli; Mike Priddy; Paulin Ribbe; Carsten Thiel & Lars Wieneke (2017). D5.1 Report on Integrated Service Needs: DARIAH (in kind) contributions – Concept and Procedures. https://github.com/Dans-labs/dariah/wiki/Documents/d5-1.pdf [16 January 2020]

Dillo, Ingrid; Herve L'Hours & Mari Kleemola (2018). CoreTrustSeal Foundation Statutes and Rules of Procedure: https://doi.org/10.5281/zenodo.1142960 [22 January 2020]

Donaldson, Devan Ray; Ingrid Dillo; Robert Downs & Sarah Ramdeen (2017). The Perceived Value of Acquiring Data Seals of Approval. International Journal of Digital Curation 12:1, 130–151. https://doi.org/10.2218/ijdc.v12i1.481 [14 January 2020]

E-RIHS website: http://www.e-rihs.eu/ [22 January 2020]

E-RIHS. About: http://www.e-rihs.eu/about/about/ [27 February 2020]

European Commission. European Open Science Cloud (EOSC): https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud [22 January 2020]

European Framework for Audit and Certification of Digital Repositories website: http://www.trusteddigitalrepository.eu/Trusted%20Digital%20Repository.html [22 January 2020]

European Open Science Cloud. EOSC Projects: https://www.eosc-portal.eu/about/eosc-projects [22 January 2020]

European Parliament (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [11 February 2020]

FAIRsFAIR project website: https://www.fairsfair.eu/ [22 January 2020]

FAIRsFAIR website: https://www.fairsfair.eu/ [22 January 2020]

FAIRsFAIR. FAIR Certification (of Repositories) - WP4: https://www.fairsfair.eu/fair-certification [22 January 2020]

FitSM - Standards for lightweight IT Service management website: https://fitsm.eu [22 January 2020]

Geisteswissenschaftliches Asset Management System (GAMS) website: http://gams.uni-graz.at [22 January 2020]

International Organization for Standardization (2013). ISO/IEC 27001:2013. https://www.iso.org/standard/54534.html [22 January 2020]

International Organization for Standardization (2014). ISO 16919:2014: https://www.iso.org/standard/57950.html [22 January 2020]

International Organization for Standardization (2015). ISO 17021-1:2015: https://www.iso.org/standard/61651.html [22 January 2020]

ISO 16363. Certified clients: http://www.iso16363.org/iso-certification/certified-clients/ [22 January 2020]

L'Hours, Herve; Mari Kleemola & Lisa de Leeuw (2019). CoreTrustSeal: From academic collaboration to sustainable services. https://doi.org/10.29173/iq936 [22 January 2020]

L'Hours, Hervé & Ilona von Stein (2019). FAIR Ecosystem Components: Vision. http://doi.org/10.5281/zenodo.3565428 [21 January 2020]

L'Hours, Hervé; René van Horik; Mari Kleemola; Jonas Recker; Janez Štebe & Birger Jerlehag (2020). CESSDA Trust Group: Overview of Support Approaches. http://doi.org/10.5281/zenodo.3621378 [22 January 2020]

Lyle, Jared (2019). Data Repository Assessment & Certification at ICPSR: Experiences and Lessons Learned. Presentation at NIH Trustworthy Data Repositories workshop in Rockville, Maryland, 8 April, 2019. https://datascience.nih.gov/sites/default/files/20190408_NIH_repo_certification_Jared.pdf [14 January 2020]

Max Planck Institute for the History of Science (2017). Digital Research Infrastructure. https://www.mpiwg-berlin.mpg.de/research/projects/digital-research-infrastructure [22 January 2020]

NAKALA website: https://www.nakala.fr/ [22 January 2020]

Nestor Working Group (2009). Catalogue of Criteria for Trusted Digital Repositories: https://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf [22 January 2020]

Nestor. Evaluated Archives: https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html [22 January 2020]

Re3data Registry of Research Data Repositories: www.re3data.org [22 January 2020]

Research Data Alliance (2020). FAIR Data Maturity Model WG. https://www.rd-alliance.org/groups/fair-data-maturity-model-wg [22 January 2020]

Research Data Alliance. Groups. https://www.rd-alliance.org/groups [22 January 2020]

Research Data Alliance. RDA/WDS Certification of Digital Repositories IG: https://www.rd-alliance.org/groups/rdawds-certification-digital-repositories-ig.html [22 January 2020]

Science Europe association website: https://www.scienceeurope.org/ [26 February 2020]

Science Europe Data Glossary. Data Repository. http://sedataglossary.shoutwiki.com/wiki/Data_repository [20 December 2019]

Statutes of CESSDA ERIC (C/2017/3870) (2017). Official Journal of the European Union, C220/1. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017Y0708(01) [14 January 2020]

Tjalsma, Heiko; Hervé L'Hours; Mari Kleemola; Natascha Schumann & Janez Štebe (2017). Deliverable 4.4 – Report on DSA Certification for CESSDA. http://cessdasaw.eu/content/uploads/2017/11/SAW-D4.4..pdf [16 January 2020]

UK Data Service website: https://www.ukdataservice.ac.uk/ [22 January 2020]

Van Horik, René (2005). Permanent pixels. Building blocks for the longevity of digital surrogates of historical photographs. DANS studies in digital archiving 1. The Hague, 2005. https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/permanent-pixels [22 January 2020]

Waterman, Kees; Sierman, Barbara (2016). Survey on DSA-certified digital repositories. Report on the findings in a survey of all DSA-certified digital repositories on investments in and benefits of acquiring the Data Seal of Approval (DSA). https://doi.org/10.5281/zenodo.1188256 [22 January 2020]

Wilkinson, Mark D; Michel Dumontier; IJsbrand Jan Aalbersberg; Myles Axton; Arie Baak; Niklas Blomberg; … Jun Zhao (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data vol. 3, Article number: 160018. https://doi.org/10.1038/sdata.2016.18 [14 January 2020]

Appendix A. Frameworks

OAIS: Reference Model for an Open Archival Information System (ISO 14721)

The OAIS reference model plays a central role in certification standards so it will be explained in more detail below.

An OAIS is defined as "an archive, consisting of an organisation of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community". A Designated Community is an 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.[61]

The model can be considered as a conceptual framework informing the design of system architectures, but it does not ensure consistency or interoperability between implementations. The OAIS reference model contains three key high-level concepts[62]:

1. The environment of an OAIS. An OAIS or archive is surrounded by 'Producers' (which provide the information to be preserved), 'Consumers' (which interact with OAIS services to find and acquire preserved information of interest), and 'Management' (who set the overall OAIS policy as one component in a broader policy domain).
2. OAIS Information. In order to understand the Data Object (this is either a physical object or a digital object) that has been archived, Representation Information is required. Representation Information is the information that maps a Data Object into more meaningful concepts. Thus, an Information Object consists of two components: the Data Object and the Representation Information. It is necessary to distinguish between an Information Package that is preserved by an OAIS and Information Packages that are submitted to, and disseminated by, an OAIS. These

[61] 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 [22 January 2020]

[62] Van Horik, René (2005). Permanent pixels. Building blocks for the longevity of digital surrogates of historical photographs. DANS studies in digital archiving 1. The Hague, 2005. https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/permanent-pixels [22 January 2020]

variants are referred to as the Archival Information Package (AIP), the Submission Information Package (SIP), and the Dissemination Information Package (DIP).

3. High-level external interactions. Producer and consumer interaction with the OAIS are based on specific Information Packages. A Producer delivers a SIP to the OAIS for use in the construction of one or more AIPs. A Consumer receives a DIP, derived from one or more AIPs, in response to a request to the OAIS.

The OAIS functional model consists of six entities:

1. Ingest. Contains the services and functions that accept the SIPs from producers, prepares the AIPs for storage, and ensures that the AIPs and their supporting Descriptive Information become established within the OAIS.
2. Archival storage. Contains the services and functions used for the storage and retrieval of the AIP.
3. Data management. Contains the services and functions for populating, maintaining and accessing a wide variety of information.
4. Administration. Contains the services and functions needed to control the operation of the other OAIS functional entities on a day-to-day basis.
5. Preservation planning. Contains services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete.
6. Access. Contains the services and functions that make the archival information holdings and related services visible to Consumers.

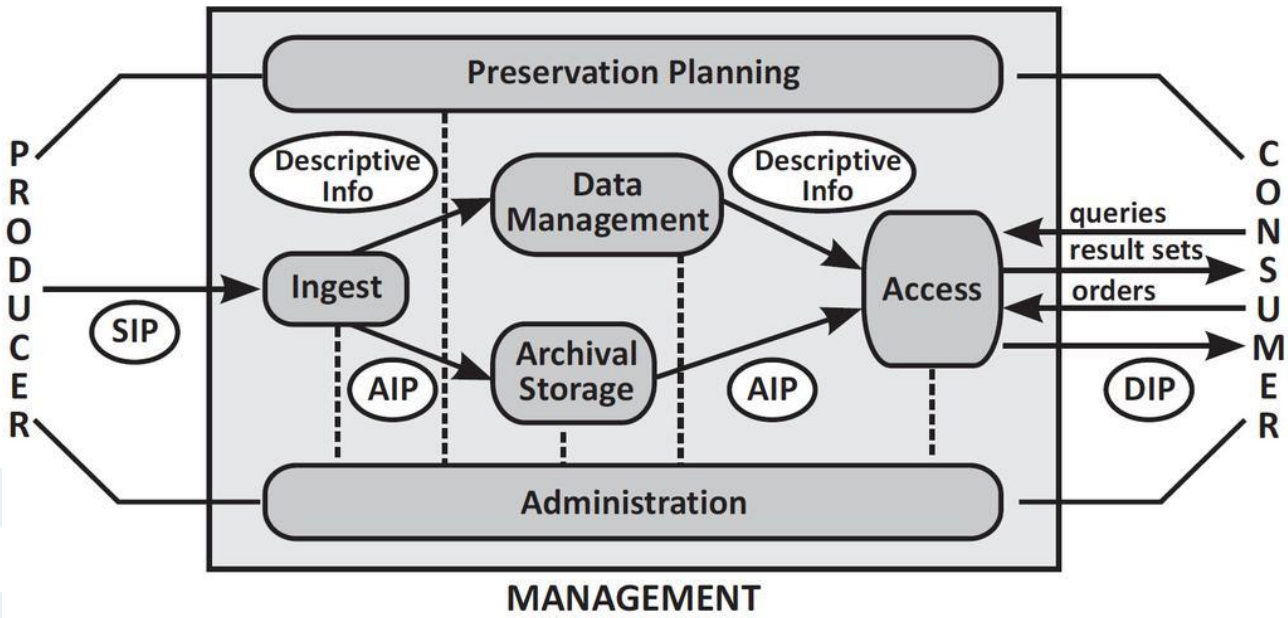

FIGURE 1. OAIS FUNCTIONAL ENTITIES. THE FIGURE CONTAINS BOTH THE THREE HIGH-LEVEL CONCEPTS AND THE SIX ENTITIES OF THE OAIS REFERENCE MODEL

ISO 16363 Audit and certification of trustworthy digital repositories

The ISO 16363 standard (or CCSDS 652.0-M-1[63]) sets out comprehensive metrics for what an archive must do, based on the Reference Model for an Open Archival Information System (OAIS). An OAIS establishes a common framework of terms and concepts relevant for the long-term archiving of digital data. More information on OAIS is given in more detail below. The ISO 16363 standard can be used as the basis for providing audit and certification of the trustworthiness of digital repositories. It provides a detailed specification of criteria by which digital repositories shall be audited. Another standard, ISO 16919:2014 also known as CCSDS 652.1-M-2[64], "Requirements for bodies providing audit and certification of candidate trustworthy digital repositories", specifies the competencies and requirements on auditing bodies. This formal approach to certify repositories is of high quality, but also expensive (as external reviewers are involved) and time consuming.

Nestor Seal (DIN 31644)

The overall aim of the nestor catalogue of criteria for trusted digital repositories[65] is to introduce stable criteria for a wide variety of long-term digital repositories and to maintain the criteria over a long period. For this reason, the 14 catalogue criteria have been formulated at an abstract level. Each criterion is enriched by detailed explanations and concrete examples and grouped into sections entitled: Organization Framework, Object Management and Infrastructure and Security. The catalogue was also designed to conform to the OAIS Reference Model. The standard is officially known as nestor Seal DIN 31644.

63 CCSDS The Consultative Committee for Space Data Systems (2012). Audit and Certification of Trustworthy Digital Repositories. Recommended practice. CCSDS 652.0-M-1, Magenta Book September 2011: https://public.ccsds.org/pubs/652x0m1.pdf [22 January 2020]

64 CCSDS The Consultative Committee for Space Data Systems (2012). 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 [22 January 2020]

65 Nestor Working Group (2009). Catalogue of Criteria for Trusted Digital Repositories: https://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf [22 January 2020]

The certification process consists of a self-assessment based on the criteria of the nestor catalogue. This assessment is reviewed by an external expert, ideally followed by an assignment. At the time of writing, four repositories had undergone a nestor certification[66].

CoreTrustSeal

The CoreTrustSeal (CTS)[67] is a community-based non-profit organisation promoting sustainable and trustworthy data infrastructures. It is governed by a Standards and Certification Board consisting of members drawn from the Assembly of Reviewers (by election) and the wider repositories stakeholders (appointed). Ultimately, CoreTrustSeal plans to collaborate on a global framework for repository certification that moves from the core to the extended (nestor Seal DIN 31644), to the formal (ISO 16363) level.

The CoreTrustSeal emerged under the auspices of the Research Data Alliance. In 2013, the Repository Audit and Certification working group was proposed, with a focus on a partnership between the Data Seal of Approval (DSA) and WDS. The working group undertook an analysis and comparison of the two sets of procedures and criteria with a view to creating a single set supporting the goals of both sources. With an agreed set of procedural and criteria references, the WDS and DSA began negotiations to merge the Data Seal of Approval and the WDS criteria into a single independent entity, which then became the CoreTrustSeal Requirements in autumn 2017. In 2018, CoreTrustSeal became a legal non-profit foundation entity under Dutch law.[68]

The application process for CoreTrustSeal certification consists of four stages: application, review, revision and approval. After submitting the self-assessment of the applicant in an online tool, the application is peer reviewed. Following feedback and revision the CoreTrustSeal Board grants a CoreTrustSeal to the applicant. The seal is valid for three years and is published online.

The CoreTrustSeal Requirements[69] are:

R0. Context for the repository.

[66] Nestor. Evaluated Archives: https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/nestor_siegel_node.html [17 December 2019]
[67] CoreTrustSeal website: https://www.coretrustseal.org/ [22 January 2020]
[68] L'Hours, Herve; Mari Kleemola & Lisa de Leeuw (2019). CoreTrustSeal: From academic collaboration to sustainable services. https://doi.org/10.29173/iq936 [22 January 2020]
[69] CoreTrustSeal (2019). Core Trustworthy Data Repositories Requirements 2020–2022: https://zenodo.org/record/3638211 [27 January 2020]

R1. The repository has an explicit mission to provide access to and preserve data in its domain.
R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.
R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.
R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.
R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.
R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).
R7. The repository guarantees the integrity and authenticity of the data.
R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.
R9. The repository applies documented processes and procedures in managing archival storage of the data.
R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.
R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.
R12. Archiving takes place according to defined workflows from ingest to dissemination.
R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.
R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.
R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

For each requirement, a compliance level is assigned. The levels are:

0) Not applicable
1) The repository has not considered this yet
2) The repository has a theoretical concept
3) The repository is in the implementation phase
4) The guideline has been fully implemented in the repository

Compliance Levels of 1 or 2 are not sufficient for a successful application. Certification may be granted if some Requirements are in the implementation phase (3).