![SSHOC - social sciences & humanities open cloud]

Research and Innovation Action

Social Sciences & Humanities Open Cloud

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

Deliverable 5.6 - Report on principles of governance and sustainability for repository services that make use of the software delivered in D5.5

Dissemination Level	PU
Due Date of Deliverable	30/04/22, M40
Actual Submission Date	29/04/22
Work Package	WP5 – Innovation in Data Access
Task	T5.2 – Hosting and sharing data repositories
Type	Report
Approval Status	Approved by EC - 04 May 2022
Version	V1.0
Number of Pages	p.1 – p.23

Abstract: This report describes the main principles of governance and sustainability of the software delivered in *D5.5 'Archive in a box' repository software and proof of concept of centralised installation in the cloud* and the effect these characteristics have on the governance and sustainability of repository services that make use of it.

History

Version	Date	Reason	Revised by
0.0	15/11/2021	First draft/structure	Vasso Kalaitzi, KNAW/DANS
0.1	20/01/2022	Content enhancement	Marion Wittenberg, KNAW/DANS; Vasso Kalaitzi, KNAW/DANS
0.2	15/03/2022	First internal review	Marion Wittenberg, KNAW/DANS
0.3	04/04/2022	Content editing upon internal review	Marion Wittenberg, KNAW/DANS; Vasso Kalaitzi, KNAW/DANS
0.4	12/04/2022	Peer review	Daan Broeder, University of Utrecht
1.0	13/04/2022	Addressing peer review and final deliverable	Marion Wittenberg, KNAW/DANS; Vasso Kalaitzi, KNAW/DANS

Author List

Organisation	Name	Contact Information
KNAW/DANS (CESSDA)	Marion Wittenberg	marion.wittenberg@dans.knaw.nl
KNAW/DANS (CESSDA)	Vasso Kalaitzi	vasso.kalaitzi@dans.knaw.nl

Executive Summary

In the context of the Social Sciences and Humanities Open Cloud - SSHOC project realising the Social Sciences and Humanities (SSH) part of the European Open Science Cloud (EOSC), Task 5.2 Hosting and sharing data repositories (T5.2) has worked towards adjusting and enhancing the Dataverse repository software to the needs of the SSH research communities. Dataverse is an open-source, community-driven data repository software that has been tailored to serve the SSH communities. This effort led to the 'Archive in a box' solution and relevant documentation for a downloadable installation in localised platforms, that offers the opportunity for the creation of an online repository.

The current document reports on the governance and sustainability of the 'Archive in the box' repository software, exploring the effect this would have for the governance and sustainability of services using it. Further reporting on the process of investigating governance and sustainability of potential services using the software, this document provides information with regards to a) the creation of a proof of concept, b) discussions for further uptake, and c) its potential onboarding in the EOSC, in alignment with SSHOC's overall governance and sustainability planning.

The work delivered under T5.2 includes investigating the governance and sustainability for the 'Archive in a box' software and services using it. This work aspires to prove the point of a bottom-up, SSH-community led approach in supporting data repositories, as well as being the starting point for further collaboration between ERICs in this respect.

Abbreviations and Acronyms

CESSDA	Consortium of European Social Science Data Archives
CLARIN	Common Language Resources and Technology Infrastructure
CNR	National Research Council of Italy
DANS	Data Archiving and Networked Services
DARIAH	Digital Research Infrastructure for the Arts and Humanities
EOSC	European Open Science Cloud
ERIC	European Research Infrastructure Consortium
E-RIHS	European Research Infrastructure for Heritage Science
ESFRI	European Strategy Forum on Research Infrastructures
EU	European Union
GDCC	Global Dataverse Community Consortium
IQSS	Institute for Quantitative Social Science
ISTNI	Istituto di Scienza e Tecnologie dell'Informazione
KER	Key Exploitable Result
LIP	Laboratory of Instrumentation and Experimental Particle Physics
MoU	Memorandum of Understanding
PMB	Project Management Board
PSNC	Poznan Supercomputing and Networking Center
SPs	CESSDA Service Providers
SSH	Social Sciences and Humanities
SSHOC	Social Sciences and Humanities Open Cloud
WP	Work Package

Table of Contents

1. Introduction

In the framework of the SSHOC project,[1] being a community-led EU-funded project aiming at building the Social Sciences and Humanities (SSH) part of the European Open Science Cloud (EOSC),[2] Task 5.2 (T5.2) *Hosting and sharing data repositories* worked towards tailoring the data repository software *Dataverse*, and adding extra functionalities based on the needs of the SSH communities. Dataverse[3] is open-source, community-driven research data repository software developed by the Institute for Quantitative Social Science (IQSS),[4] Harvard University. Currently (February 2022), 76 repositories globally use the software as their repository software. An increasing number of developers and FAIR data experts comprising the Dataverse community[5] are contributing to improve the software together with Harvard.

SSHOC, through the T5.2 team, contributed to this end for the benefit of the European SSH communities, with regards to the sharing and publication of research data.

The SSHOC project has set up various mechanisms at different levels regarding Governance and Sustainability of its Key Exploitable Results (KERs): through a dedicated Work Package (WP8) on *Governance, Sustainability and Quality Assurance*, through internal consultations of the tasks producing the KERs with the SSHOC Project Management Board (PMB) and the SSHOC Tier1 board that includes representatives of the SSH European Research Infrastructure Consortia[6] (ERICs) and projects, and at European level in consultations with the other ESFRI[7] cluster projects[8] and relevant EOSC bodies.

T5.2 contributes to the efforts of the SSHOC project with regard to governance and sustainability, by focussing in the present report on the governance and sustainability of the software described in *D5.5 - 'Archive in a box' repository software and proof of concept of centralised installation in the cloud,[9]* after the SSHOC project has been completed, and the effect these characteristics have on the governance and

---

[1] Social Sciences and Humanities Open Cloud (SSHOC), https://sshopencloud.eu/ [04/04/2022]

[2] European Open Science Cloud, https://eosc.eu/ [04/04/2022]

[3] Dataverse Project, https://dataverse.org/about [28/03/2022]

[4] The Institute for Quantitative Social Science (IQSS), Harvard University, https://www.iq.harvard.edu/ [28/03/2022]

[5] Dataverse community on Github, https://github.com/IQSS/dataverse/graphs/contributors [28/03/2022]

[6] European Research Infrastructure Consortia, https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures/eric_en [04/04/2022]

[7] ESFRI Roadmap, https://www.esfri.eu/esfri-roadmap-2021 [04/04/2022]

[8] Gotz, Andy, Petzold, Andreas, Asmi, Ari, Blomberg, Niklas, Lamanna, Giovanni, & Dekker, Ron. (2020). ESFRI cluster projects - Position papers on expectations and planned contributions to the EOSC. Zenodo. https://doi.org/10.5281/zenodo.3675081

[9] D5.5 - 'Archive in a box' repository software and proof of concept of centralised installation in the cloud:: https://doi.org/10.5281/zenodo.6396125 [to be published]

1.1 'Archive in a box' repository software and proof of concept of centralised installation in the cloud

Within T5.2 of the SSHOC project, repository software is being developed based on Dataverse, for the sharing and publication of research data within the Social Sciences and Humanities (SSH) domain. To make it easier for institutions with lack of technical staff to install the Dataverse repository software, T5.2 developed the so-called 'Archive in a Box' software package. After downloading, this package does an automatic installation and setup of the complete repository infrastructure. Deliverable 5.5 ('Archive in a Box' repository software and proof of concept of centralised installation in the cloud) describes the details of this package. The "Archive in a Box" will be available via the SSH Open Marketplace.[10]

Additionally, to the "Archive in a Box", T5.2 worked on a proof of concept of a centralised cloud installation of the Dataverse software at the Google cloud infrastructure of CESSDA.[11] A cloud installation makes it possible to automate the installation and keep the application up and running, for instance by scaling up or down resources when needed. Another advantage of a cloud orchestrator is the ability to start a new component or part of the application if it should fail for some reason. Deliverable 5.5 also reports about the details of this proof of concept. A central installation of the Dataverse software in the cloud could be used by institutions to implement a repository service that can be used by smaller organisations.

Furthermore, T5.2 developed several additional functionalities to the Dataverse software to make the software more compliant to the needs of the SSH communities in Europe.

The following functionalities were developed:

- Workflow to translate the Graphical User Interface to languages other than English;

- Inclusion of metadata standards that are important to the European SSH research communities;

- Plugin for remote controlled vocabulary support;

- Plugin with Taverna workflows;

- PyDataverse, testing software;

---

[10] SSH Open Marketplace, https://marketplace.sshopencloud.eu/ [28/03/2022]
[11] CESSDA, https://www.cessda.eu/ [28/03/2022]

- Development of previewers;

- Integration with CLARIN[12]'s Language Resource Switchboard;[13]

- Integration with Apache Superset.

Dataverse is open-source software and all code developed in T5.2 is available from Github repositories. Since SSHOC is a project, the SSHOC Github repository is not sustainable enough. That's why the team tried as much as possible to include the code in the repository of the Global Dataverse Community Consortium (GDCC).[14] The GDCC coordinates worldwide contributions to the Dataverse project. Some code is also integrated in the master branch versions of Dataverse.

1.2 Emerging European Dataverse Community

The Dataverse software is globally used by 76 repositories and in Europe there is an emerging Dataverse community. On the map of the Dataverse project there are 29 European repositories that are currently using the Dataverse software. Among them are Sciences Po[15] in France, DataverseNO[16] in Norway, the Austrian Social Science Data Archive (AUSSDA)[17] and Göttingen Research Online[18] of the University of Goettingen in Germany. The Dutch national centre of expertise and repository for research data (DANS) is not only using Dataverse for its repository DataverseNL but is also developing 4 new Data Stations with the Dataverse software.

In a recently held survey among CESSDA Service Providers (SPs), 14 of the 23 SPs were using or planning to use the Dataverse software. Within CESSDA, these SPs are planning to collaborate and share efforts to adjust the Dataverse software to their needs.

Worthwhile to mention is that there is also a great potential for using Dataverse in communities other than SSHOC.

---

[12] CLARIN, https://www.clarin.eu/ [04/04/2022]

[13] CLARIN Language Resource Switchboard, https://switchboard.clarin.eu/ [04/04/2022]

[14] Global dataverse Community Consortium (GDCC), https://dataversecommunity.global [28/03/2022]

[15] Sciences Po, https://www.sciencespo.fr/en [04/04/2022]

[16] DataverseNO onboarded to the EOSC portal: https://marketplace.eosc-portal.eu/services/dataverseno?fromc=sharing-discovery [04/04/2022]

[17] Austrian Social Science Data Archive (AUSSDA), https://aussda.at/ [04/04/2022]

[18] Göttingen Research Online, https://www.eresearch.uni-goettingen.de/services-and-software/goettingen-research-online/ [04/04/2022]

The Global Dataverse Community Consortium (GDCC) is established to provide an international organisation to existing community efforts and collaborations regarding Dataverse. Currently the GDCC is investigating whether they can acquire a legal status in one of the European countries, so they can more easily participate in European projects and apply for project funding.

1.3 Governance and sustainability of Repository Services using the repository software delivered in D5.5

The "Archive in a Box" software package is available for downloading and usage by institutions who would like to manage a research data repository for their designated community in their own environment. The results of the proof of concept of a centralised installation in the cloud can be used at the ERIC level to provide a central service for the members of the ERIC.

The usage of the 'Archive in a Box' is efficient for institutions that would like to use the Dataverse software. The idea behind the "Archive in a Box" is simple: it does an automatic installation and setting up of the complete repository infrastructure without any extra efforts. This is especially useful for institutions with limited technical staff. This would lead to lower operating costs and minimum human resources to be committed for the institution. The software is open-source and supported by a growing global and European community. This substantial community involvement and the support of the GDCC should make the software more sustainable than commercial products or in-house developments.

As of April 1st, 2022, there are already 34 Dataverse installations making use of the source code of the "Archive in a Box". A clear indication that it's already being actively used by the Dataverse community. It was installed and is currently tested by various organisations such as the Laboratory of Instrumentation and Experimental Particle Physics (LIP)[19] in Portugal, the DataverseNO consortium from Norway, University of Paris[20] in France, Institute of Information Science and Technologies "Alessandro Faedo" - ISTI (ISTI-CNR)[21] in Italy and others.

---

[19] Laboratory of Instrumentation and Experimental Particle Physics (LIP), https://www.lip.pt/?section=home&page=homepage&lang=en [04/04/2022]

[20] University of Paris, https://u-paris.fr/en/ [04/04/2022]

[21] Institute of Information Science and Technologies "Alessandro Faedo" - ISTI (ISTI-CNR), https://www.isti.cnr.it/en/ [04/04/2022]

The idea behind the proof of concept of a centralised installation in the Cloud was that a research organisation providing research infrastructure services for others such as for instance an ERIC could offer a central service within their network to small institutions with no technical resources.

This idea originated from CESSDA, and a proof of concept was developed and tested in the CESSDA infrastructure. In 2021 an investigation was carried out to determine how many CESSDA archives or partners were interested in such a service. Unfortunately, the outcome of this survey was that such a service will not be viable within the CESSDA community. Therefore, the cloud service is not onboarded to the EOSC (See section 2).

One of the DARIAH partners within T5.2, Poznan Supercomputing and Networking Center (PSNC),[22] is currently testing whether they can offer a centralised service to partners within DARIAH. The discussions within DARIAH about this possible service are not yet finished.

---

[22] Poznan Supercomputing and Networking Center (PSNC), https://www.psnc.pl/ [04/04/2022]

2. Onboarding and connection to the EOSC

The following section describes different levels of onboarding, including connection to the EOSC, with regards to overall governance and sustainability that frames and interacts with the governance and sustainability of the repository software delivered under T5.2/D5.5, following the completion of the SSHOC project.

2.1 SSHOC consultations establishing common understanding

The SSHOC project has focused since the beginning in organising internal consultations on sustainability issues at several levels. Governance and sustainability, being of high importance for the project formally under WP8, encompasses all other SSHOC WPs, hence consultations were held at the level of the Project Management Board (PMB) and Tier1 board either on ad hoc basis depending on SSHOC developments and needs, or under the SSHOC Consortium Meetings, happening every 6 months. These meetings aim at following both SSHOC and EOSC developments, thus addressing the uncertainties existing at the beginning of the project.

In this context, the SSHOC project has created an overview of critical questions related to the SSHOC portfolio of services and tools. This overview includes all SSHOC services and tools per WP, including a short description, the beneficiaries responsible within the project, the relevant EOSC Services and Resources category, the content and technical contacts, the contact of people within the project working on governance and sustainability for each item, information on timeline of beta and production versions, as well as information whether the service or tool will be registered in the EOSC Portal Registry, the gateway to information and resources in EOSC [23].

Furthermore, a template on collecting information on the project's KERs exists, which includes the SSHOC Dataverse repository software, a SSHOC project brief to support the EC programme and policy activities, as well as updated branding guidelines attached to the output's governance and sustainability.

At the level of the SSHOC consultations, T5.2 has contributed with the work and results, as described in D5.5, i.e., with the developments and provisions of the "Archive in a Box" repository software and proof of concept of centralised installation in the cloud. The repository software, being developed is based on

---

[23] EOSC portal onboarding process: https://eosc-portal.eu/providers-documentation/eosc-portal-onboarding-process

Dataverse, is open-source and, hence, can be maintained by the European and Global Dataverse community, which already uses a lot of additional software functionalities and results achieved by T5.2.

The proof of concept of a centralised cloud installation of the Dataverse software at the moment exists in the Google cloud infrastructure used by CESSDA. In the case this proof of concept would be upgraded into an actual service, CESSDA would keep the responsibility, while the operational costs would be paid to the CESSDA member operating the service (i.e., DANS[24]), through CESSDA service providers subscriptions fees.

The Poznan Supercomputing and Network Centre (PSNC) is currently investigating whether it can deploy a centralised cloud installation of the Dataverse software in their technical environment for use by the DARIAH community.

Onboarding this proof-of-concept of SSHOC Dataverse on the SSH Open Marketplace, is not an option for the time being, as the proof of concept does not consist of a deployed production service. However, this is a point to be taken up further after the end of the SSHOC project and could be offered as a common service through a Memorandum of Understanding (MoU) among SSHOC partners.

2.2 WP8 Governance/ Sustainability/ Quality Assurance fostering further collaboration

SSHOC is one of the community-driven projects, which actively contributes to the building and implementation of the EOSC infrastructure. As such, one of its WPs (WP8) is dedicated to the development and coordination of a governance model, which will ensure sustainability of the SSH components built by the project in the framework of EOSC. The strategies produced by WP8 focus, among others, on SSH stakeholders' involvement in reducing fragmentation and supporting the sustainability of the SSH digital infrastructure landscape under the EOSC infrastructure.

Taking into consideration the unique aspects of different SSH sub-domains, a wide range of stakeholders, the requirements and needs in capacity building for a sustainable contribution, as well as the ongoing developments in the general EOSC landscape, WP8/T8.1 aimed at the establishment of an appropriate governance and sustainability model, reflected in the deliverable *D8.1 Governance & Sustainability Roadmap* (to be published).

D8.1 reports on the organisational governance and sustainability model needed for ensuring and

---

[24] Data Archiving and Network Services, https://dans.knaw.nl/en/ [04/04/2022]

safeguarding the availability, quality, and maintenance of SSHOC outputs after the end of the project, the conditions to be met for financial sustainability, as well as the actions needed for the implementation of the aforementioned structure.

D8.1 highlights the need of the research community for, as well as the value of science clusters such as SSHOC to provide a sustainable service offering. SSHOC WP8 has been monitoring the understanding of the consortium with regards to this value proposition compared to the EOSC developments, aiming at further consolidating this record of understanding before the end of the project. In this context, taking into consideration the SSHOC developments in the past three years, an MoU has been created to lay out the terms and conditions for further cooperation of consortium members following the end of the project. The ERICs involved in the project started the signing procedure in March 2022. Goal of the MoU is to establish an SSH Cluster of collaborating SSH stakeholders. Further to mutual collaboration, it aims to build upon established synergies, also encompassing the sustainability of the SSH service offering as developed in the course of the project. As the MoU is not a legally binding document, separate agreements are expected to be created to work on and achieve specific results.

This expressed intention for further collaboration, deriving from the SSHOC project and being extended through the aforementioned MoU and specific separate agreements, can provide a starting point with regards to future services using the "Archive in a box" software for the benefit of the SSH communities.

2.3 Connecting to the EOSC

Fostered by the EOSC Secretariat,[25] the EOSC Executive Board[26] and the EOSC Working Groups (2019) have produced reports with recommendations in 6 priority areas, including crucial input for the EOSC Strategic Research and Innovation Agenda (SRIA)[27] on the upcoming years of the EOSC. One of the EOSC Working Groups was dedicated to Sustainability,[28] where the SSHOC project was also represented, and delivered its report *A FAIR Lady (Iron Lady) report from the EOSC Sustainability Working Group*[29] in November 2020.

---

[25] EOSC Secretariat, https://www.eoscsecretariat.eu/ [04/04/2022]

[26] EOSC Executive Board Outputs, https://www.eoscsecretariat.eu/eosc-governance/eosc-executive-board-outputs [04/04/2022]

[27] Strategic Research and Innovation Agenda of the European Open Science Cloud, https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf [04/04/2022]

[28] EOSC Sustainability Working Group, https://www.eoscsecretariat.eu/working-groups/sustainability-working-group [04/04/2022]

[29] A FAIR Lady (olim Iron Lady) report from the EOSC Sustainability Working Group, DOI 10.2777/870770

At the same time, SSHOC participated in discussions on governance and sustainability with the other EOSC ESFRI clusters in the context of several events[30]:

- Research Data Alliance Plenary 14,[31] October 2019

- EOSC Symposium 2019: Where the EOSC makers and shakers meet,[32] November 2019

- Realising the European Open Science Cloud conference,[33] November 2020

- The ESFRI Clusters at the RDA House of Commons,[34] April 2021

- EOSC Symposium 2021[35] (Online), June 2021

- EOSC Enhance workshops[36]

Ongoing efforts are taking place with regards to service onboarding in the EOSC. The onboarding process is not needed for the repository software itself, however, its governance and sustainability are not expected to be a blocking factor in the onboarding of current and future services using it.

The Tromsø Repository of Language and Linguistics (TROLLing[37]), a global domain-specific repository for language and linguistic data, makes use of the 'Archive in a Box' software and is one of the first Dataverse repositories in Europe that is on boarded to the EOSC portal. DANS, the Dutch national centre of expertise and repository for research data, is currently in the process of investigating the onboarding to the EOSC for its DataverseNL[38] service and the domain specific Data stations based on Dataverse.

---

[30] Irena Vipavc Brvar, & Ana Inkret. (2021). D6.5 Report on Stakeholder Series events. Zenodo. https://doi.org/10.5281/zenodo.5638651 [04/04/2022]

[31] Ron Dekker, Daan Broeder, Franciska de Jong, Vasso Kalaitzi, Marieke Willems, Tracey Biller, Mingfang Wu, Marco Molinaro, Jonathan Clark, Ari Asmi, René van Horik, Mirjam van Daalen, & Ornela di Giacomo. (2019). EOSC. ESFRI Cluster Projects. RDA: Connecting commonalities and collaborative solutions for community research data services. Research Data Alliance Plenary 14 (RDA P14), Helsinki. Zenodo. https://doi.org/10.5281/zenodo.3581075 [04/04/2022]

[32] SSHOC at the EOSC Symposium 2019, https://sshopencloud.eu/eosc-symposium-2019 [04/04/2022]

[33] Realising EOSC Conference, https://sshopencloud.eu/news/realising-eosc-virtual-conference-difference [04/04/2022]

[34] The EOSC ESFRI Clusters at the RDA House of Commons report, Stephanie Parker, Ana Inkret, Irena Vipavc Brvar, Vasso Kalaitzi, Iris Buunk, & Jana Striova. (2021). The ESFRI Clusters at RDA House of Commons (v1.0). Zenodo. https://doi.org/10.5281/zenodo.5024589 [04/04/2022]

[35] Bertacchini, Veronica, Drago, Federico, Flicker, Katharina, Gebreyesus, Netsanet, Grant, Annabel, Jones, Bob, Liinamaa, Iiris, Märkälä, Anu, Marinos-Kouris, Christos, Meerman, Bert, Saurugger, Bernd, & Smith, Zachary. (2021). EOSC Symposium 2021 Report. Zenodo. https://doi.org/10.5281/zenodo.5176089 [04/04/2022]

[36] EOSC Enhance project, https://eosc-portal.eu/enhance [04/04/2022]

[37] The Tromsø Repository of Language and Linguistics (TROLLing): https://trolling.uit.no/

[38] DataverseNL, https://dataverse.nl/ [04/04/2022]

3. Governance and Sustainability Scenarios and Recommendations

3.1 Adoption, fostering and maintenance of the T5.2 software and service under by the SSH ERICs

The team of task 5.2 consists of staff members from organisations that are members of one or more of the 4 ERICs: CESSDA, DARIAH,[39] CLARIN and E-RIHS.[40] This background ensured that the team was already aware of the needs within their respective research infrastructures.

In January 2020, UiT the Arctic University of Norway[41] organised a European Dataverse Community workshop.[42] Next to a technical workshop for developers, a session was organised by SSHOC on more organisational topics regarding Dataverse for policy officers, repository managers and curators. This started the discussions about which functionalities are crucial for the European context and about sustainable business models.

Additional to this, two webinars were organised. One for the CESSDA community[43] and one for the DARIAH community.[44] As follow-up to these webinars, additional meetings with representatives of CESSDA, DARIAH, CLARIN and E-HRIS were organised.

The following paragraphs describe whether and how the different ERICs want to use Dataverse within their organisation.

---

[39] DARIAH, https://www.dariah.eu/ [04/04/2022]

[40] E-RIHS, http://www.e-rihs.eu/ [04/04/2022]

[41] UiT The Arctic University of Norway, https://en.uit.no/startsida [04/04/2022]

[42] European Dataverse Workshop: https://doi.org/10.7557/scs.2020.2 [03/04/2022]

[43] SSHOC Webinar - Discussion meeting about requirements of CESSDA Service Providers for a Dataverse repository: https://sshopencloud.eu/sshoc-webinar-cessda-service-providers-dataverse [03/04/2022]

[44] DARIAH Community Requirements for a Dataverse Repository: https://sshopencloud.eu/events/sshoc-webinar-dariah-community-requirements-dataverse-repository [03/04/2022]

CESSDA

Within two projects prior to the SSHOC project, the CESSDA SaW project[45] and the DataverseEU project, CESSDA had already looked into whether Dataverse could be a possible replacement for the Nesstar software[46] that had been used by many CESSDA Service Providers (SPs). The SSHOC project can be seen as a follow-up of both projects.

Within the SSHOC project, the team of T5.2 developed the controlled vocabulary service and the adaptation of the metadata schema to the CESSDA Metadata Model (CMM)[47] at the request of CESSDA partners. Also, a workflow was developed to translate the Graphical User Interface (GUI) of Dataverse into other languages than English.

Initially, CESSDA also thought that a centralised service in the cloud could be very useful for small (starting) CESSDA service providers. As mentioned in paragraph 1.3, there appears to be not much interest in a central service within CESSDA.

There is a lot of interest among CESSDA SPs in the 'Archive in a Box' software. In a recent inventory, 9 SPs indicate that they want to make use of this distribution. A growing group of CESSDA SPs want to see to what extent the Dataverse software can be further adapted to the needs of the CESSDA community and to what extent they can work together for this.

DANS, the Dutch CESSDA SP is partner within ODISSEI (Open Data Infrastructure for Social Science and Economic Innovations)[48]. This project is making use of the 'Archive in a Box' software for the development of its portal.

CLARIN

Team members of T5.2 and executives of CLARIN discussed the uptake of the 'Archive in a Box' and the centralised cloud installation by CLARIN or its members. CLARIN ERIC is not decisive in selecting the kind of software its members use, but they saw certain advantages in the use of Dataverse. Therefore task 5.2 developed an integration with CLARIN's Language Resource Switchboard and worked on customisation

[45] CESSDA SaW project, https://cessdasaw.eu/ [04/04/2022]

[46] Nesstar: http://www.nesstar.com/ [10/03/2022]

[47] Storviken, Silje, Hagen, Sunniva, Bockaj, Brigita, Bolko, Irena, Vipavc Brvar, Irena, Fink Kjeldgaard, Anne Sofie, Guldfeldt, Madsen, Christina, Ala-Lahti, Henri, Haanpää, Eliisa, Kleemola, Mari, Moilanen, Katja, Borschewski, Kerrin, Zenk-Möltgen, Wolfgang, Alfredsson, Iris, Lundgren, Malin, Svensson, Sara, Etheridge, Anne, & Bolton, Sharon. (2019). User Guide for the CESSDA Metadata Model (Version 0). Zenodo. https://doi.org/10.5281/zenodo.3236194

[48] ODISSEI (Open Data Infrastructure for Social Science and Economic Innovations), https://odissei-data.nl/en/en-odissei/ [04/04/2022]

of the metadata schema within Dataverse for the CLARIN Component MetaData Infrastructure (CMDI) (still in progress) of which the use is an essential requirement for CLARIN centres.

Within CLARIAH,[49] the Dutch national collaboration of CLARIN and DARIAH, the KNAW Humanities Cluster (HuC)[50] started to use the test instance of the "Archive in a Box" release. They are also working on the extension of the Dataverse metadata schema with the CMDI support and adoption of SKOSMOS[51] hosted vocabularies to the needs of its community.

There is also ongoing work to extend basic Dataverse functionality with proper CMDI metadata publication via the built-in OAI-PMH protocol. This extension is dedicated for all CLARIN community members and will allow data providers to use Dataverse as a FAIR data repository that could be integrated with various linguistic tools developed by CLARIN in the different projects.

DARIAH

After the SSHOC Dataverse webinar for DARIAH, a number of meetings with the DARIAH Coordination Office were held. These discussions should be continued. The Poznan Supercomputing and Networking Center (PSNC), one of the partners within T5.2 is interested to offer a centralised repository service to the partners within DARIAH, based on the developed proof of concept of a centralised service. Discussions about this offer are not yet finalised.

E-HRIS

A few discussions were held with representatives of E-RIHS. This hasn't resulted in concrete agreements. DANS, member of the Steering Committee of E-RIHS NL will launch at the end of April 2022 a Data Station for Archeology[52] based on the Dataverse software.

3.2 Training workshops and manual

To help organisations that will be providing a Dataverse repository service, T5.2 produced a Dataverse manual for end-users.[53] This manual aims to provide a practical, user-friendly, and concise guide

---

[49] CLARIAH https://www.clariah.nl/ [04/04/2022]
[50] KNAW Humanities Cluster (HuC): https://huc.knaw.nl/ [13/04/2022]
[51] SKOSMOS, https://skosmos.org/ [04.04.2022]
[52] DANS Data Station Archaeology, https://dans.knaw.nl/nl/data-stations/archaeology/ [04/04/2022]
[53] Heider, Veronika, Saldner, Simon, Huis in 't Veld, Laura, & Wittenberg, Marion. (2022). Dataverse End User Manual (1.0). Zenodo. https://doi.org/10.5281/zenodo.6347488

intended for typical end users of Dataverse.

In some countries there is a legal provision to offer websites in all languages spoken in the country. That means that some organisations need to translate the Graphical User Interface (GUI) in their national language(s). 5.2 has developed a specific Workflow to translate the GUI into languages other than English. Three Translation Workshops were organised for organisations which would like to translate the GUI of Dataverse. Information about these workshops is available at the SSHOC website.[54] A translation guide is also available at Zenodo[55]

4. Conclusion

T5.2 of the SSHOC project has delivered an 'Archive in a box' and a proof of concept of a centralised installation in the cloud, both based on Dataverse, open-source, community-driven research data repository software developed by the Institute for Quantitative Social Science (IQSS), Harvard University. Additional functionality was added to the needs of the SSH communities in Europe.

The 'Archive in a Box' is already in use by various European institutions. Within CESSDA and CLARIN there is a growing group of institutions and projects that use the software for their purposes. But the proof of concept has yet to be deployed at the ERIC research infrastructure level, as a broadly offered centralised service. Within CESSDA such a service does not seem viable at this moment, therefore the cloud service was not onboarded to the EOSC. Within DARIAH discussions about such a service will be continued, potentially even after the end of the SSHOC project.

The first service making use of the 'Archive in a Box' software, TROLLing - The Tromsø Repository of Language and Linguistics, is already onboarded to the EOSC.

The governance and sustainability aspects of the 'Archive in a box' software significantly affect the governance and sustainability of potential services using it, and that is in a positive way, due to continuous support and maintenance on behalf of the European and Global Dataverse communities.

Following the overall governance and sustainability work in the context of the SSHOC project, the SSH ERICs involved in the project will have the opportunity of collaborating and exploring potential further

[54] Translation Workshops Dataverse GUI: https://www.sshopencloud.eu/news/sshoc-workshop-notes-dataverse-translation-follow-event
[55] Huis in 't Veld, Laura, & Heider, Veronika. (2021). Dataverse General User Interface Translation Guide for Weblate. Zenodo. https://doi.org/10.5281/zenodo.4807371

development of services based on the aforementioned proof of concept, taking into account onboarding aspects deriving from EOSC developments.

5. References

- Social Sciences and Humanities Open Cloud (SSHOC), https://sshopencloud.eu/ [04/04/2022]

- European Open Science Cloud, https://eosc.eu/ [04/04/2022]

- ESFRI Roadmap, https://www.esfri.eu/esfri-roadmap-2021 [04/04/2022]

- Gotz, Andy, Petzold, Andreas, Asmi, Ari, Blomberg, Niklas, Lamanna, Giovanni, & Dekker, Ron. (2020). ESFRI cluster projects - Position papers on expectations and planned contributions to the EOSC. Zenodo. https://doi.org/10.5281/zenodo.3675081 [04/04/2022]

- Dataverse Project, https://dataverse.org/about [28/03/2022]

- The Institute for Quantitative Social Science (IQSS), Harvard University, https://www.iq.harvard.edu/ [28/03/2022]

- Dataverse community on Github, https://github.com/IQSS/dataverse/graphs/contributors [28/03/2022]

- Global dataverse Community Consortium (GDCC), https://dataversecommunity.global [28/03/2022]

- European Research Infrastructure Consortia, https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures/eric_en [04/04/2022]

- Sciences Po, https://www.sciencespo.fr/en [04/04/2022]

- The Tromsø Repository of Language and Linguistics (TROLLing) onboarded to the EOSC portal: https://marketplace.eosc-portal.eu/services/the-tromso-repository-of-language-and-linguistics-trolling?fromc=sharing-discovery [14/04/2022]

- Austrian Social Science Data Archive (AUSSDA), https://aussda.at/ [04/04/2022]

- Göttingen Research Online, https://www.eresearch.uni-goettingen.de/services-and-software/goettingen-research-online/ [04/04/2022]

- Laboratory of Instrumentation and Experimental Particle Physics (LIP), https://www.lip.pt/?section=home&page=homepage&lang=en [04/04/2022]

- University of Paris, https://u-paris.fr/en/ [04/04/2022]

- Institute of Information Science and Technologies "Alessandro Faedo" - ISTI (ISTI-CNR), https://www.isti.cnr.it/en/ [04/04/2022]

- Poznan Supercomputing and Networking Center (PSNC), https://www.psnc.pl/ [04/04/2022]

- Data Archiving and Network Services, https://dans.knaw.nl/en/ [04/04/2022]

- CESSDA, https://www.cessda.eu/ [04/04/2022]

- ESS Survey, https://www.europeansocialsurvey.org/ [04/04/2022]

- CLARIN, https://www.clarin.eu/ [04/04/2022]

- CLARIN Language Resource Switchboard, https://switchboard.clarin.eu/ [04/04/2022]

- KNAW Humanities Cluster (HuC): https://huc.knaw.nl/ [13/04/2022]

- SHARE, http://www.share-project.org/ [04/04/2022]

- SSH Open Marketplace, https://marketplace.sshopencloud.eu/ [28/03/2022]

- D5.5 - 'Archive in a box' repository software and proof of concept of centralised installation in the cloud:: https://doi.org/10.5281/zenodo.6396125 [to be published]

- D8.1 Governance & Sustainability Roadmap [to be published]

- EOSC Secretariat, https://www.eoscsecretariat.eu/ [04/04/2022]

- EOSC Executive Board Outputs, https://www.eoscsecretariat.eu/eosc-governance/eosc-executive-board-outputs [04/04/2022]

- Strategic Research and Innovation Agenda of the European Open Science Cloud, https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf [04/04/2022]

- EOSC Sustainability Working Group, https://www.eoscsecretariat.eu/working-groups/sustainability-working-group [04/04/2022]

- EOSC portal onboarding process: https://eosc-portal.eu/providers-documentation/eosc-portal-onboarding-process

- A FAIR Lady (olim Iron Lady) report from the EOSC Sustainability Working Group, DOI 10.2777/870770 [04/04/2022]

- Irena Vipavc Brvar, & Ana Inkret. (2021). D6.5 Report on Stakeholder Series events. Zenodo. https://doi.org/10.5281/zenodo.5638651 [04/04/2022]

- The EOSC ESFRI Clusters at the RDA House of Commons report, Stephanie Parker, Ana Inkret, Irena Vipavc Brvar, Vasso Kalaitzi, Iris Buunk, & Jana Striova. (2021). The ESFRI Clusters at RDA House of Commons (v1.0). Zenodo. https://doi.org/10.5281/zenodo.5024589 [04/04/2022]

- SSHOC at the EOSC Symposium 2019, https://sshopencloud.eu/eosc-symposium-2019 [04/04/2022]

- Realising EOSC Conference, https://sshopencloud.eu/news/realising-eosc-virtual-conference-difference [04/04/2022]

- Ron Dekker, Daan Broeder, Franciska de Jong, Vasso Kalaitzi, Marieke Willems, Tracey Biller, Mingfang Wu, Marco Molinaro, Jonathan Clark, Ari Asmi, René van Horik, Mirjam van Daalen, & Ornela di Giacomo. (2019). EOSC. ESFRI Cluster Projects. RDA: Connecting commonalities and collaborative solutions for community research data services. Research Data Alliance Plenary 14 (RDA P14), Helsinki. Zenodo. https://doi.org/10.5281/zenodo.3581075 [04/04/2022]

- EOSC Enhance project, https://eosc-portal.eu/enhance [04/04/2022]

- DataverseNL, https://dataverse.nl/ [04/04/2022]

- Bertacchini, Veronica, Drago, Federico, Flicker, Katharina, Gebreyesus, Netsanet, Grant, Annabel, Jones, Bob, Liinamaa, Iiris, Märkälä, Anu, Marinos-Kouris, Christos, Meerman, Bert, Saurugger, Bernd, & Smith, Zachary. (2021). EOSC Symposium 2021 Report. Zenodo. https://doi.org/10.5281/zenodo.5176089 [04/04/2022]

- DARIAH, https://www.dariah.eu/ [04/04/2022]

- E-RIHS, http://www.e-rihs.eu/ [04/04/2022]

- UiT The Arctic University of Norway, https://en.uit.no/startsida [04/04/2022]

- The Tromsø Repository of Language and Linguistics (TROLLing): https://trolling.uit.no/ [13/04/2022]

- European Dataverse Workshop: https://doi.org/10.7557/scs.2020.2 [03/04/2022]

- SSHOC Webinar - Discussion meeting about requirements of CESSDA Service Providers for a Dataverse repository: https://sshopencloud.eu/sshoc-webinar-cessda-service-providers-dataverse [03/04/2022]

- DARIAH Community Requirements for a Dataverse Repository: https://sshopencloud.eu/events/sshoc-webinar-dariah-community-requiremen ts-dataverse-repository [03/04/2022]

- Nesstar: http://www.nesstar.com/ [10/03/2022]

- CESSDA SaW project, https://cessdasaw.eu/ [04/04/2022]

- Storviken, Silje, Hagen, Sunniva, Bockaj, Brigita, Bolko, Irena, Vipavc Brvar, Irena, Fink Kjeldgaard, Anne Sofie, Guldfeldt, Madsen, Christina, Ala-Lahti, Henri, Haanpää, Eliisa, Kleemola, Mari, Moilanen, Katja, Borschewski, Kerrin, Zenk-Möltgen, Wolfgang, Alfredsson, Iris, Lundgren, Malin, Svensson, Sara, Etheridge, Anne, & Bolton, Sharon. (2019). User Guide for the CESSDA Metadata Model (Version 0). Zenodo. https://doi.org/10.5281/zenodo.3236194 [04/04/2022]

- ODISSEI (Open Data Infrastructure for Social Science and Economic Innovations), https://odissei-data.nl/en/en-odissei/ [04/04/2022]

- CLARIAH https://www.clariah.nl/ [04/04/2022]

- SKOSMOS, https://skosmos.org/ [04.04.2022]

- DANS Data Station Archaeology, https://dans.knaw.nl/en/data-stations/archaeology/ / [04/04/2022]

- Focus on FAIR: DANS 2021-2025 strategy: https://dans.knaw.nl/en/about/strategy/ [04/04/2022]

- Wittenberg, Marion; Heider, Veronika (2022). CESSDA Workplan 21-22 Task 4: D7 Dataverse as a CESSDA product: Business plan [not published]

- Priddy, Mike; Wittenberg, Marion; Tykhonov, Vyascheslav, et all. (2017). CESSDA SaW D4.5: Provision of development support services on the basis of identified demand. https://doi.org/10.5281/zenodo.3773121 [10/03/2022]

- Information about DataverseEU project on website AUSSDA: https://aussda.at/en/news-data/rss/detail-en/news/dataverseeu-reducing-language-barriers [10/03/2022]

- Huis in 't Veld, Laura, & Heider, Veronika. (2021). Dataverse General User Interface Translation Guide for Weblate. Zenodo. https://doi.org/10.5281/zenodo.4807371 [04/04/2022]

- Translation Workshops Dataverse GUI: https://www.sshopencloud.eu/news/sshoc-workshop-notes-dataverse-translation-follow-event [04/04/2022]

- Heider, Veronika, Saldner, Simon, Huis in 't Veld, Laura, & Wittenberg, Marion. (2022). Dataverse End User Manual (1.0). Zenodo. https://doi.org/10.5281/zenodo.6347488 [04/04/2022]