

Lessons from science cluster project



Franciska de Jong*

executive director CLARIN ERIC

PaNOSC Closing Event

30 November 2022



**some slides kindly provided by CLARIN & SSHOC colleagues*

Project:



Horizon 2020
European Union Funding
for Research & Innovation

Type of action & funding:
Research and Innovation action
(INFRAEOSC-04-2018)

Partners: 53

(28 beneficiaries + 25 LTPs)

SSH ESFRI Landmarks and Projects
& international SSH data infrastructures

Project budget:

€ 14,455,594.08

Duration: 40 months

(January 2019 – 30 April 2022)

Project website:
www.SSHOpenCloud.eu



Objectives:

- creating the social sciences and humanities (**SSH**) part of European Open Science Cloud (**EOSC**)
- maximising **reuse** through **Open Science** and **FAIR** principles (standards, common catalogue, access control, semantic techniques, training)
- interconnecting existing and new infrastructures (clustered cloud infrastructure)
- establishing appropriate **governance model** for SSH-EOSC

Interconnecting existing and new open science research infrastructures



CLARIN



European *Values* Study



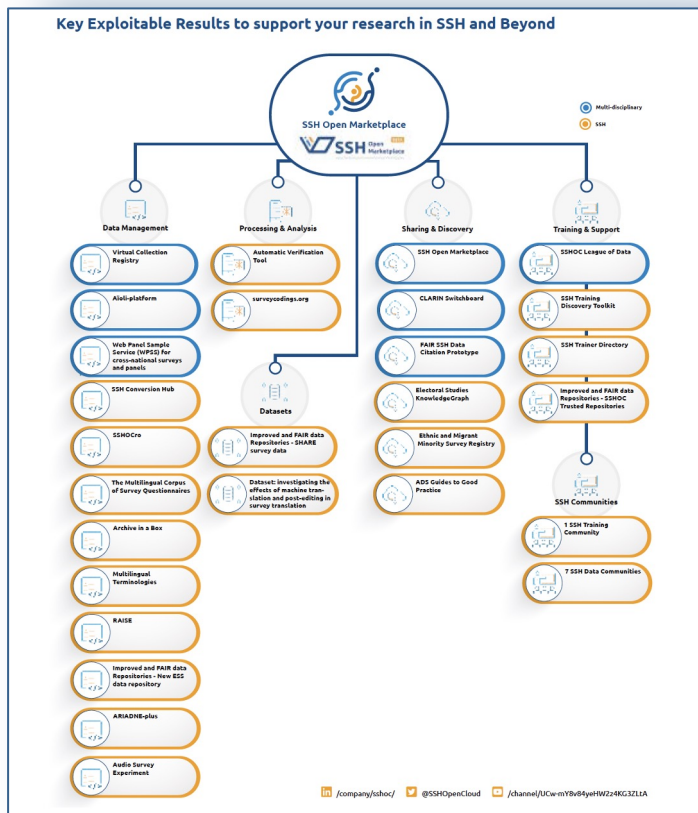
WageIndicator.org



Growing Up
In Digital Europe



Creating the social sciences and humanities (SSH) part of European Open Science Cloud (EOSC)

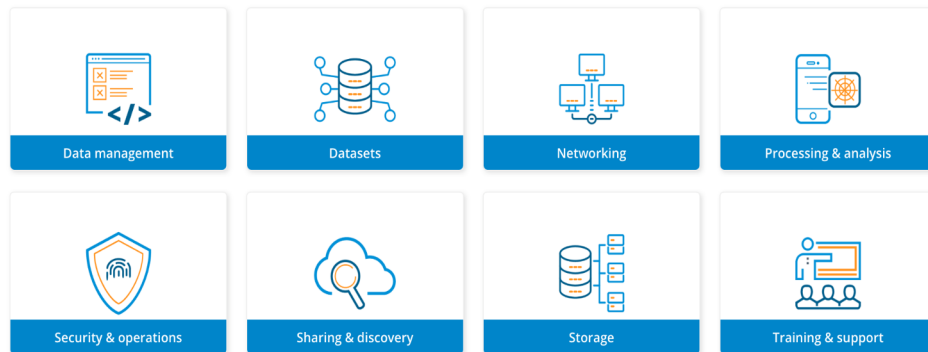


Service Catalogue

[Home](#) / Service Catalogue

Between January 2019 and April 2022 SSHOC delivers a series of services and tools for daily use by SSH researchers. All tools will be made available via the [SSH Open Marketplace](#) by April 2022.

Scroll down to browse the catalogue.



SSH-specific challenges and opportunities taken up by SSHOC and the SSH Open Cluster

- SSH's relevance for supporting the implementation of public policies
 - Responses to COVID-19
 - Wide coverage of UN Sustainable Development Goals
 - Science projects in H2020 project EOSC Future
- Interoperability as instrument to capture social and cultural dynamics
 - Diversity of data types, workflows and scholarly cultures
 - Multilinguality, multimodality, terminology, data sensitivity, ...
 - Multidisciplinarity within the cluster
- Longitudinal value of SSH data

Characteristics of SSH repository landscape

- Distributed service provision
 - Depositing services
 - Discovery platforms based on harvested metadata
- Mature model of service federation and FAIRness of resources
 - Interoperability of metadata for data and tools
 - Identity federation options
 - Federated Content Search
- > 400 SSH institutional nodes
- > 50 CTS certified repositories

Characteristics of CLARIN centre network

Centre certification

- CLARIN assessment committee
- 25 certified centres (CTS, plus CLARIN-specific requirements)
- many more centres with partial compliance

Metadata standards

- CMDI + a range of conversion workflows
- >100 profiles, > 1000 components and around 2000 concepts
- catalogue of harvested metadata:
Virtual Language Observatory: <https://vlo.clarin.eu>
- CMDI best practice guide: [link](#)



Principles behind creation of CLARIN centre network

- Compatibility at protocol and format level
- Acknowledging the unique situation of each centre (history, organization, technology, etc.)
- Focus on the strengths that a centre can contribute
- Offer for reuse of existing solutions, but no obligation to use specific software stacks
- Combination of do-it-yourself and reuse encouraged.

SSHOC Task 8.2 Trust & Quality Assurance, Impact

Purpose: Bolstering and improving trust in and quality of the SSHOC repositories by supporting them and by exploring the trust landscape and providing feedback to certification bodies.

Team:

- CESSDA/FSD
- CESSDA/SND
- CESSDA/UKDS
- KNAW/DANS
- CLARIN ERIC
- CNR/E-RIHS
- DARIAH/PNSC



cessda



FINNISH SOCIAL
SCIENCE DATA ARCHIVE



SND
Swedish National Data Service



UK Data Service

Data Archiving and Networked Services
DANS



National Research
Council of Italy

E-RIHS
EUROPEAN RESEARCH INFRASTRUCTURE
FOR HERITAGE SCIENCE



CLARIN



DARIAH-EU



PNSC



SSHOC Task 8.2 Trust & Quality Assurance, Impact

Purpose: Bolstering and improving trust in and quality of the SSHOC repositories by supporting them and by exploring the trust landscape and providing feedback to certification bodies.

Team:

- CESSDA/FSD
- CESSDA/SND
- CESSDA/UKDS
- KNAW/DANS
- CLARIN ERIC
- CNR/E-RIHS
- DARIAH/PNSC



cessda



FINNISH SOCIAL
SCIENCE DATA ARCHIVE



SND
Swedish National Data Service



UK Data Service

Data Archiving and Networked Services
DANS



National Research
Council of Italy

E-RIHS
EUROPEAN RESEARCH INFRASTRUCTURE
FOR HERITAGE SCIENCE



CLARIN



DARIAH-EU



PNSC



T8.2 Upskilling and CTS certification support

CoreTrustSeal certification support for repositories selected through an open call: 1-on-1 support, workshops, review of self-assessments

- 14 supported repositories
- Support process ran from autumn 2020 till spring 2022

Focus

- CoreTrustSeal certification aimed for by 8 out of 14 repositories
- Six repositories were mostly interested in working on the self-assessment without immediate intention to apply for CTS



T8.2 Lessons learnt

First-time certification takes up some time and resources and often requires involvement of several repository experts.

- Can be especially challenging for smaller repositories

Repositories found the support process helpful for facilitating the self-assessment and identification of things to improve.

Sustainable management of trust goes beyond certification.

- Everyday interaction with users and stakeholders needed

Diversity of organisations providing data services.

- Repositories widely differ in terms of discipline, data types and designated community
- Complex partnerships models and outsourcing can further complicate certification

T8.2 Documentation

- Kleemola Mari; Tuomas J. Alaterä; Niko Koski; Henri Ala-Lahti; Birger Jerlehag; Hervé L'Hours; Franciska de Jong; Dieter Van Uytvanck; Tomasz Parkoła; Emiliano Degl'Innocenti; Roberta Giacomi; Maurizio Sanesi & René Van Horik (2020). SSHOC D8.2 Certification plan for SSHOC repositories (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.4558303>
- Ala-Lahti, Henri; Benjamin Jacob Mathers; Hervé L'Hours; Mari Kleemola & Tuomas J. Alaterä (2022). Repositories and Beyond: Analysis of Survey for SSHOC Organisations (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.6325149>
- Ala-Lahti, Henri; Benjamin Jacob Mathers; Hervé L'Hours; Mari Kleemola & Tuomas J. Alaterä (2022). Data Repositories and Certification in a Diverse Trust Landscape: Results of SSHOC T8.2 Desk Research (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.6334025>

SSH Open Marketplace

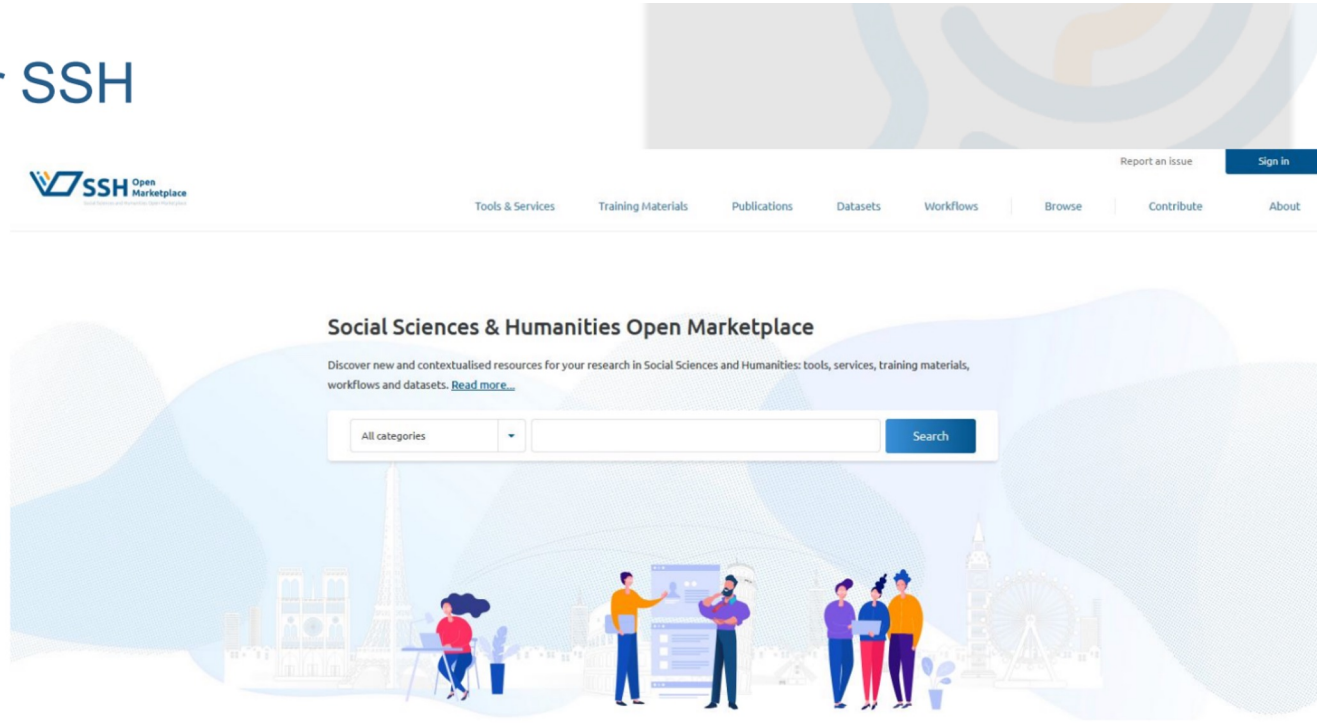
- Discovery portal for SSH resources

- Tools & services
- Training materials
- Workflows
- Datasets
- Publications

- 3 guiding principles

- Contextualisation
- Curation
- Community

- An SSH entry point to EOSC



marketplace.sshopencloud.eu



Marketplace population

- Trusted sources
 - Mass ingestion - 10+ sources
- Extraction of relations
 - Extracting tools from publications
- Prioritisation of curation tasks
 - Automatising as much as we can
 - Fine-tuning with manual curation
- Manual additions & enrichment
 - User-friendly interface
 - Datathons & hands-on session
 - Incentives & rewards

<input type="checkbox"/> Tools & Services	1699
<input type="checkbox"/> Training Materials	318
<input type="checkbox"/> Publications	2952
<input type="checkbox"/> Datasets	309
<input type="checkbox"/> Workflows	27



Marketplace Maintenance and Governance

Agreement between CESSDA, CLARIN, DARIAH

- initial duration: 1,5 years
- in-kind and cash contribution

Curation team

- Representatives from the three RIs
- Call for volunteers

Interconnecting existing and new open science research infrastructures



CLARIN



European *Values* Study



WageIndicator.org



Growing Up
In Digital Europe

