SSHOC

Dataverse in the European Open Science Cloud

Marion Wittenberg (DANS) task leader
Slava Tykhonov (DANS) lead developer

European Dataverse Workshop 2020
23-24 January 2020
Tromsø, UiT The Arctic University of Norway

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Project: SSHOC

social sciences & humanities open cloud

Horizon 2020
European Union Funding
for Research & Innovation

Type of action & funding:

Research and Innovation action
(INFRAEOSC-04-2018)

Partners: 47

(20 beneficiaries + 27 LTPs)

SSH ESFRI Landmarks and Projects
& international SSH data infrastructures

Duration: 40 months

(January 2019 – 30 April 2022)

Project budget:

€ 14,455,594.08

Project website:

www.SSHOpenCloud.eu

Objectives:

- creating the social sciences and humanities **(SSH)** part of European Open Science Cloud **(EOSC)**
- maximising **re-use** 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

Task 5.2 Hosting and sharing data repositories

Objective

Development of a research data repository service on EOSC, for SSH institutions currently without such a facility for their designated communities

Deliverables

After 38 months: Data repository service running on EOSC
After 40 months: Report on principles of governance and sustainability of the data repository service

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

History of task 5.2

To investigate how Dataverse can be used as a Research Data Management service for CESSDA Service Providers

- CESSDA SaW project
- CESSDA DataverseEU project

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

CESSDA SaW project 2016-2017 (EU Horizon 2020)

- Project for Strengthening and Widening CESSDA (Consortium of Social Science Data Archives)
- WP4 – Strengthening and widening through knowledge exchange
- Task about how CESSDA Service Providers could help each other to become more mature
- DANS offered a test environment of Dataverse to run experiments
  - Pipeline to migrate metadata from NESSTAR to Dataverse
  - Localisation of Dataverse (translation and adjustment of metadata)
  - Connection from Dataverse to Islandore
  - Archive in a box

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

CESSDA workplan project 2018 - Dataverse*EU*

- POC how to make Dataverse CESSDA compliant
  - Creating Docker images for installation on CESSDA Google Cloud
  - Experiments with Kubernetes infrastructure
  - Translations by partners
  - Adaptation to support CESSDA metadata Model (CMM)
  - DA|RA PiD plugin development
  - Support of Controlled vocabularies (CESSDA CV Service)

- CESSDA Partners: DANS, AUSSDA, ADP, SND, TARKI, GESIS

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

SSHOC task 5.2 Hosting and sharing data repositories

- Makes use of Dataverse software
- 4 ERICs: DARIAH, CLARIN, EHRIS and CESSDA
- Building mature infrastructure based on requirements of involved communities
- Investigating sustainable governance models
- Training Service Providers and institutes how to use Dataverse as a service

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Partners SSHOC task 5.2

cessda

DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

CLARIN
Common Language Resources and Technology Infrastructure

E-RIHS
EUROPEAN RESEARCH INFRASTRUCTURE FOR HERITAGE SCIENCE

Data Archiving and Networked Services
DANS

PSNC

UNIVERSITY OF TROMSØ
UiT / THE ARCTIC UNIVERSITY OF NORWAY

National Research Council of Italy

AUSSDA
AUSTRIAN SOCIAL SCIENCE DATA ARCHIVE

SUB

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Development process

DataverseSSHOC project has two parallel tracks of the development:

- **Core development team** is working on the modification and extension of the Dataverse core functionality.
- **The application development** team will create new or will integrate existent tools that will be published on Dataverse App Store website.

Our goal is to build the distributed and mature data infrastructure based on sustainable microservices.

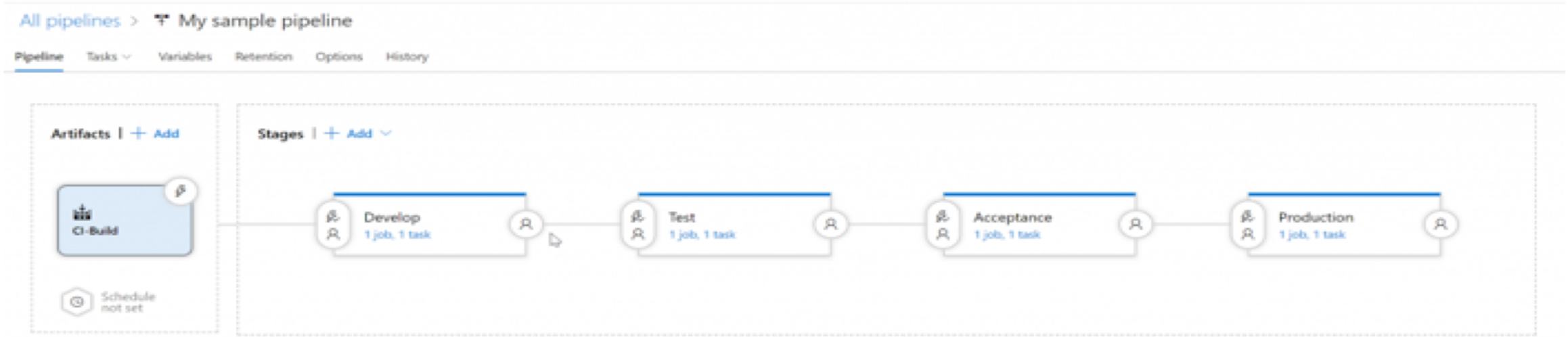
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Development methodology

We follow SCRUM with quick but small updates, all new functionality will go through DTAP (development, testing, acceptance and production) pipeline:

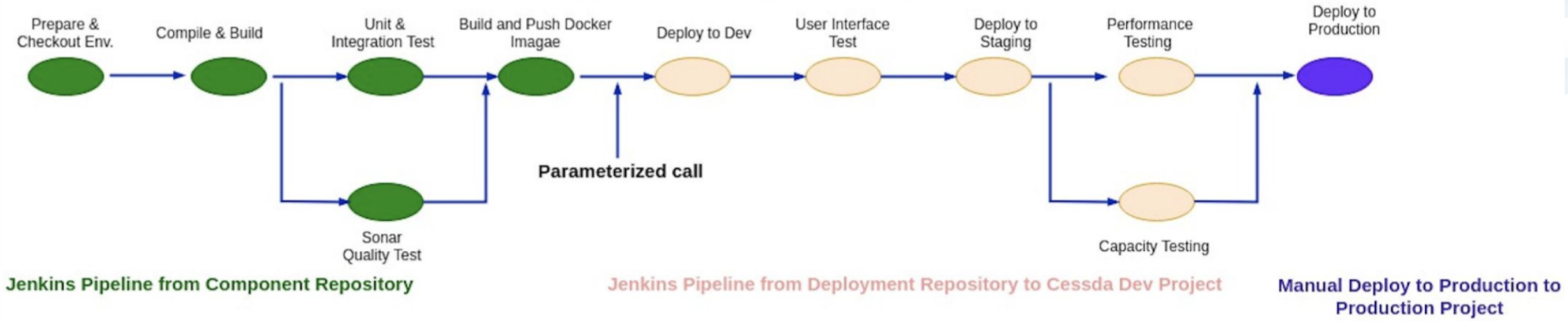
local (own PC) – test (cloud) – acceptance (cloud) – production (cloud)



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

DataverseEU deployment in CESSDA Cloud



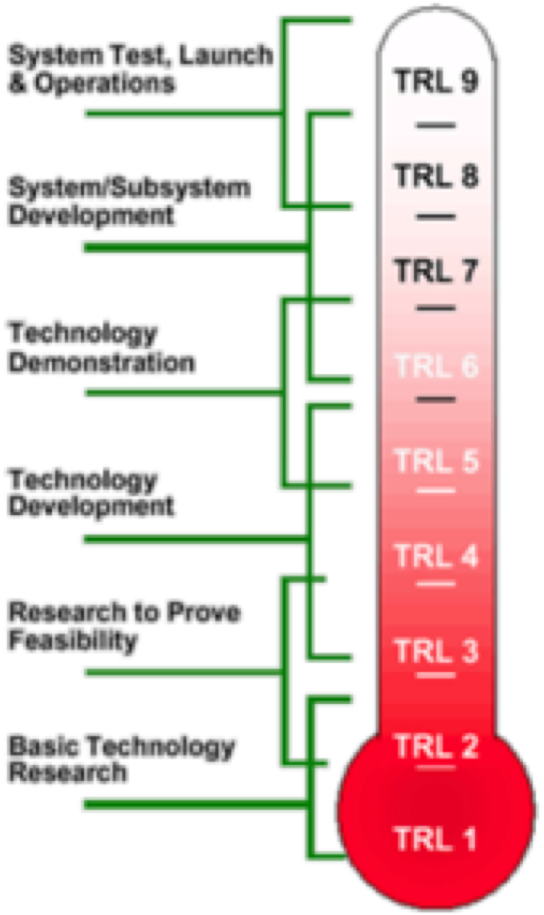
Docker Compose for the local development and testing

Kubernetes (K8s) for the deployment of services in the Cloud (CI/CD pipeline with Jenkins and Helm)

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Services in European Open Science Cloud (EOSC)

- EOSC requires the level 8 of maturity (at least)
- we need the highest quality of software to be accepted as a service
- clear and transparent evaluation of services is essential
- the evidence of technical maturity is the key to success
- the limited warranty will allow to stop out-of-warranty services



SSHOC
social sciences & humanities open cloud

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Testing process for the core and apps

Testing process follows the CESSDA maturity model
https://zenodo.org/record/2591055#.XKR6ny2B2u5

Important: every change of Dataverse functionality should be supplied with unit tests, changes of external functionality should get Selenium scenarios.

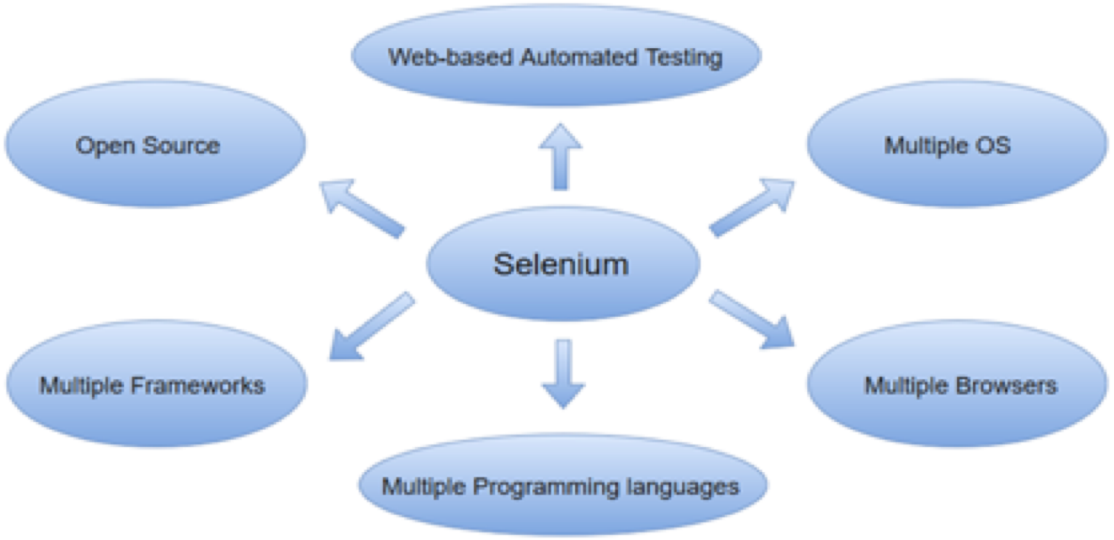
Goal: to score as high as possible according to CESSDA maturity model

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

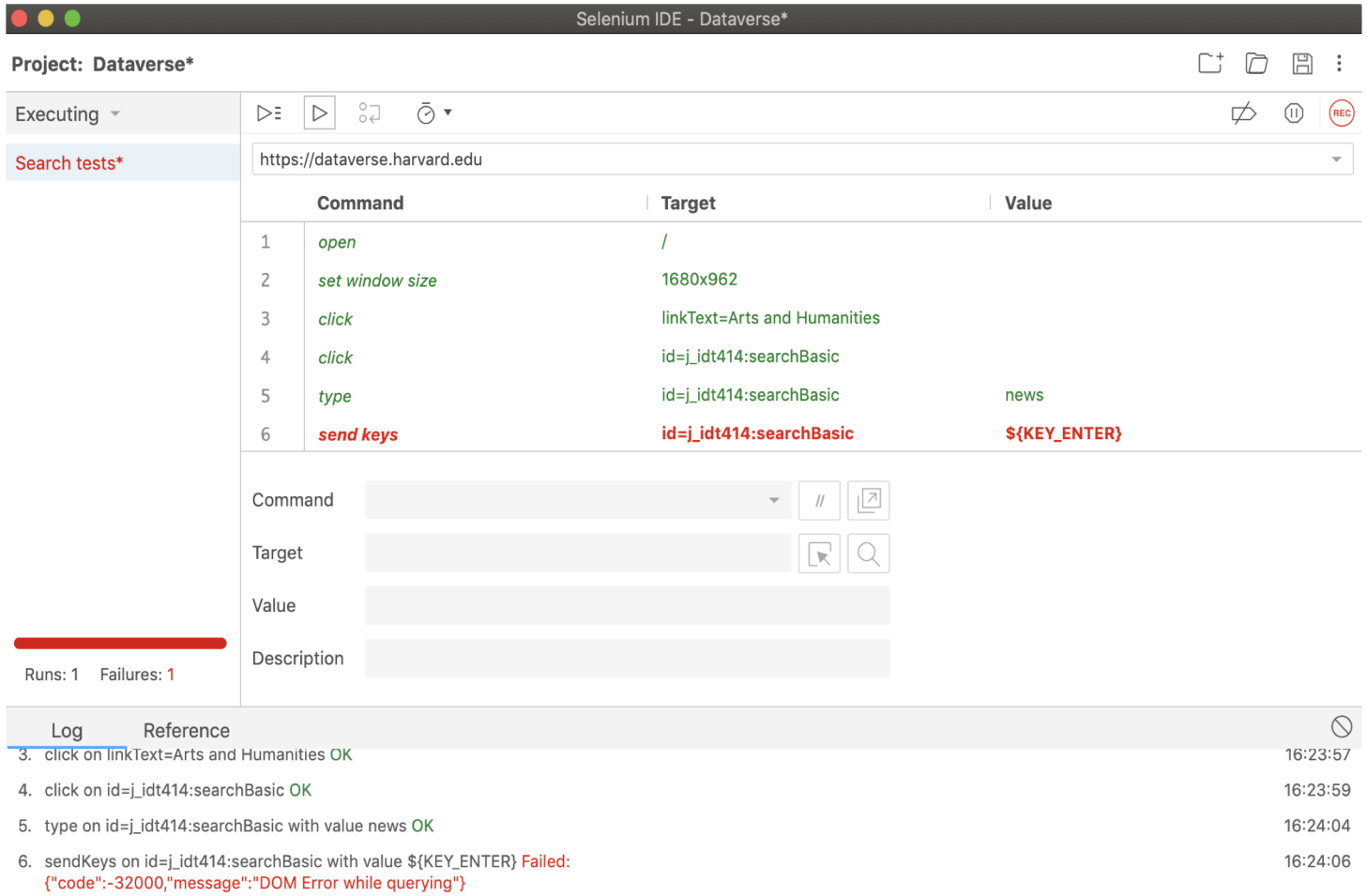
Testing procedures

- Writing unit tests (Java) for core development team
- Integration tests (Selenium) for app development team
- Performance/stress tests (Apache JMeter) before Dataverse will go to Acceptance/Production



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Quality Assurance (QA) as a community service



- Selenium IDE allows to create and replay all UI tests in your browser
- Shared tests can be reused by Dataverse CI/CD pipeline
- Let's work together on it!

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Community-driven QA plan

- Download test-suite with all tests provided by community
- Install Selenium IDE in Google Chrome
- Create new project in Selenium and upload tests
- Run all tests and provide feedback back to the community
- Everybody can create a new tests specific to your requirements, record and share with others
- more contributors will bring even more maturity of Dataverse services with CI/CD pipeline
- It means Dataverse maintenance cost will start to drop down

maturity of the community = maturity of services

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Dataverse App Store

Let’s build different services out of tools!

Data preview: DDI Explorer, Spreadsheet/CSV, PDF, Text files, HTML, Images, video render, audio, JSON, GeoJSON/Shapefiles/Map, XML

Interoperability: external controlled vocabularies (CESSDA CV Manager)

Data processing: NESSTAR DDI migration tool

Linked Data: RDF compliance including SPARQL endpoint

Federated login as a service (OAuth/Shibboleth in the same installation)

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Applications maturity level

Every software package should follow the same CESSDA Maturity Model to be accepted as a service.

Must have: k8s infrastructure with upstream Docker images, warranty statement, documentation, unit tests, Selenium tests, jenkins pipeline

Running demonstration service will allow to create the connection to your own Dataverse

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Multilingual support

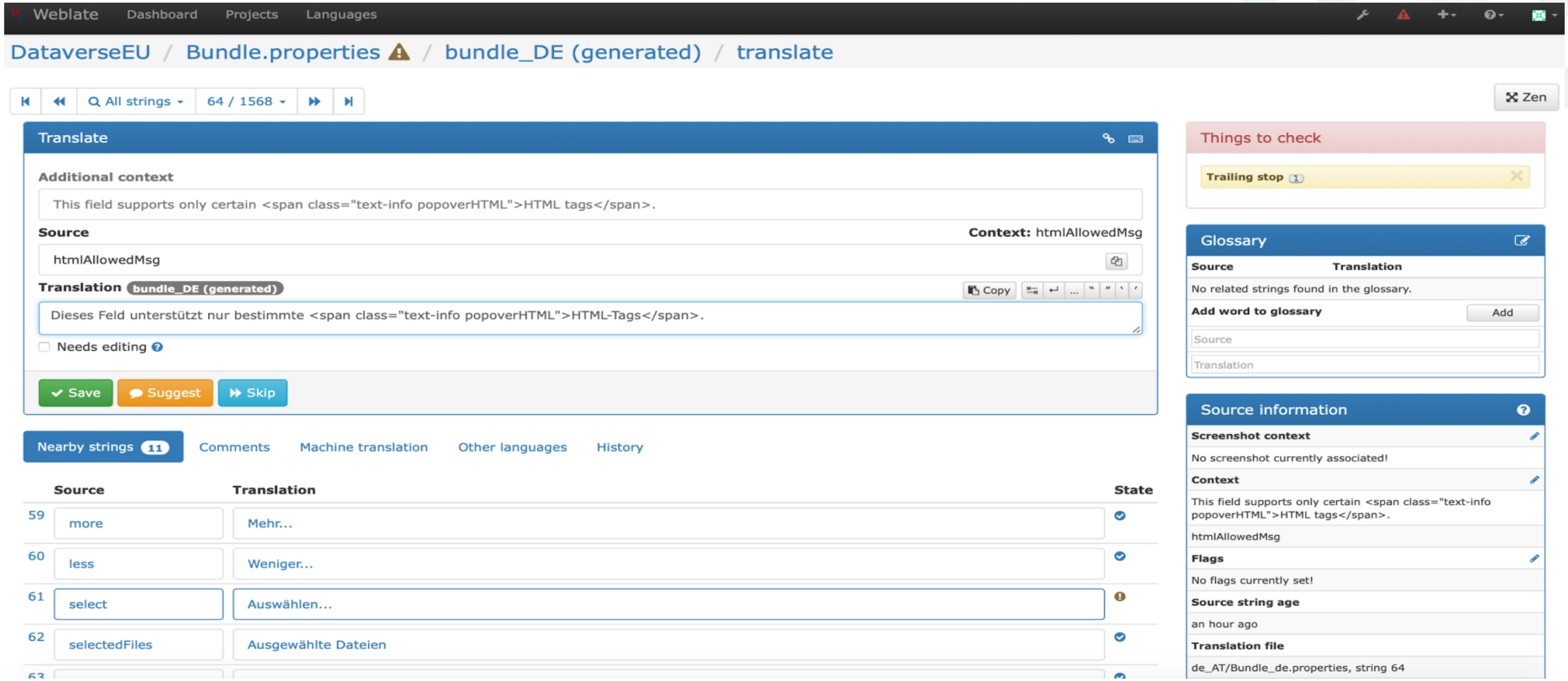
DataverseEU will run Weblate as a service for the user interface, metadata schema and SOLR translation.

We’ve developed an experimental but adjustable pipeline for multilingual support that allows to download and synchronize all translations available in Dataverse Consortium github and provides easy access for translators to keep all properties up-to-date.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

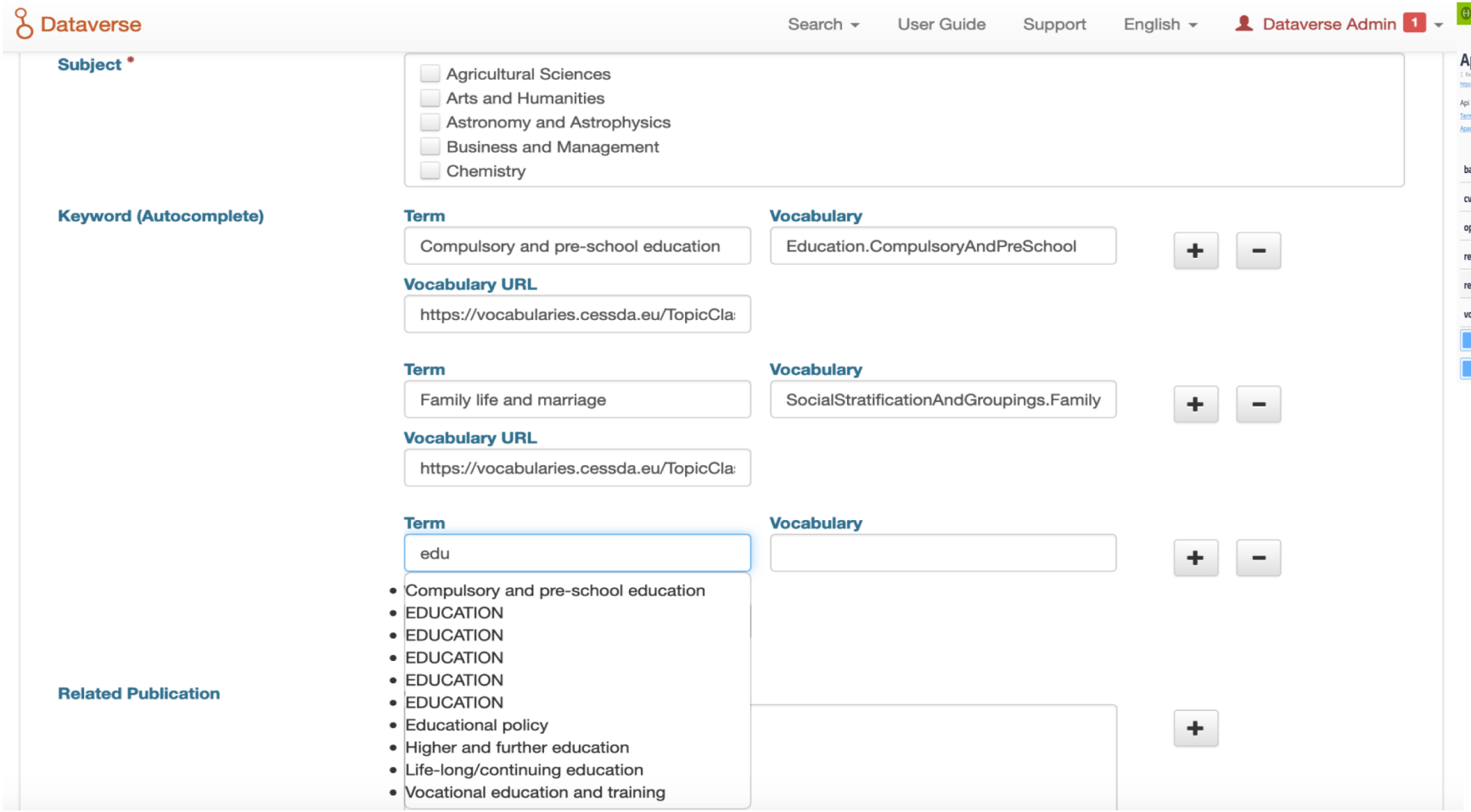
Weblate as Dataverse app

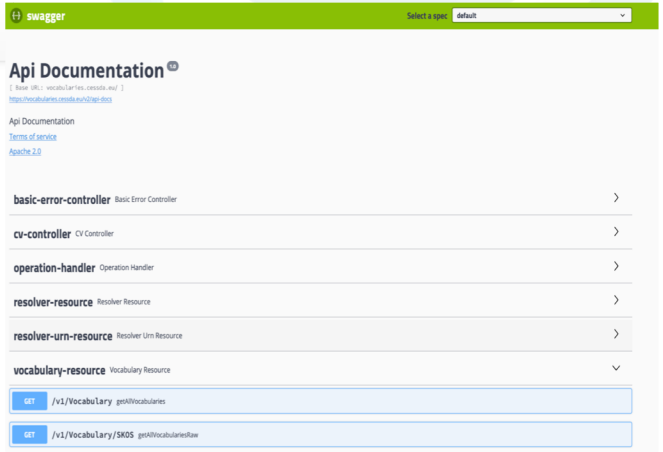


This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Controlled Vocabularies support





This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Who is going to benefit from SSHOC?

- (European) (SSH) institutes and researchers will be offered a Dataverse installation on the cloud
- (European) (SSH) institutes will be offered a Dataverse archive in a box solution for their own purposes
- Many of the features to be developed in SSHOC will benefit also other Dataverse installations / communities

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Thank you for your attention!

marion.wittenberg@dans.knaw.nl

vyacheslav.tykhonov@dans.knaw.nl

Join our community

https://www.sshopencloud.eu

@SSHOpenCloud

info@sshopencloud.eu

/in/sshopencloud

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud