

# 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



## 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

### Project budget:

€ 14,455,594.08

### Duration: 40 months

(January 2019 – 30 April 2022)

**Project website:**  
[www.SSHOpenCloud.eu](http://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

# 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

# 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

# CESSDA workplan project 2018 - DataverseEU

## 🌀 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

# 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

# Partners SSHOC task 5.2





# Development process

DataverseSSHOC project has two parallel tracks of the development:

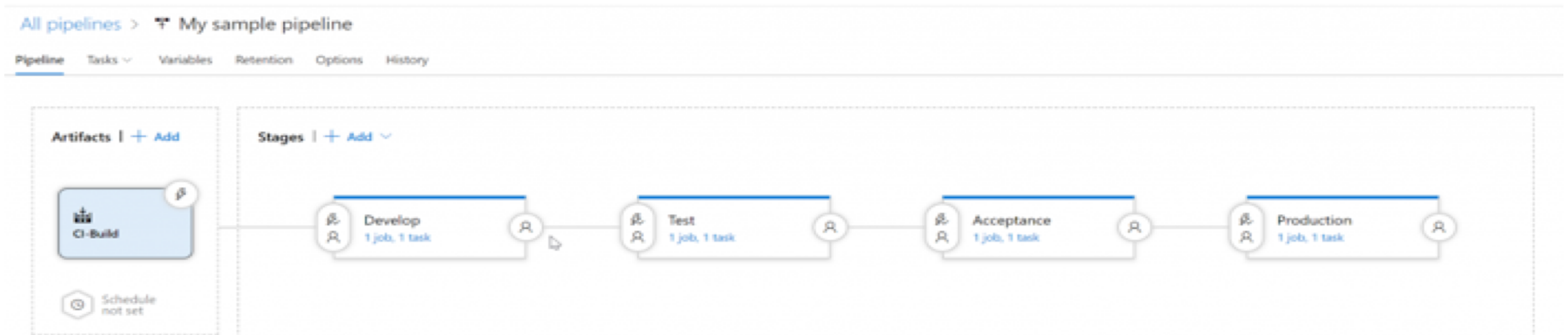
- **Core development team** is working on the modification and extension of the Dataverse core functionality.
- **The application development** team will create new or will integrate existent tools that will be published on Dataverse App Store website.

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

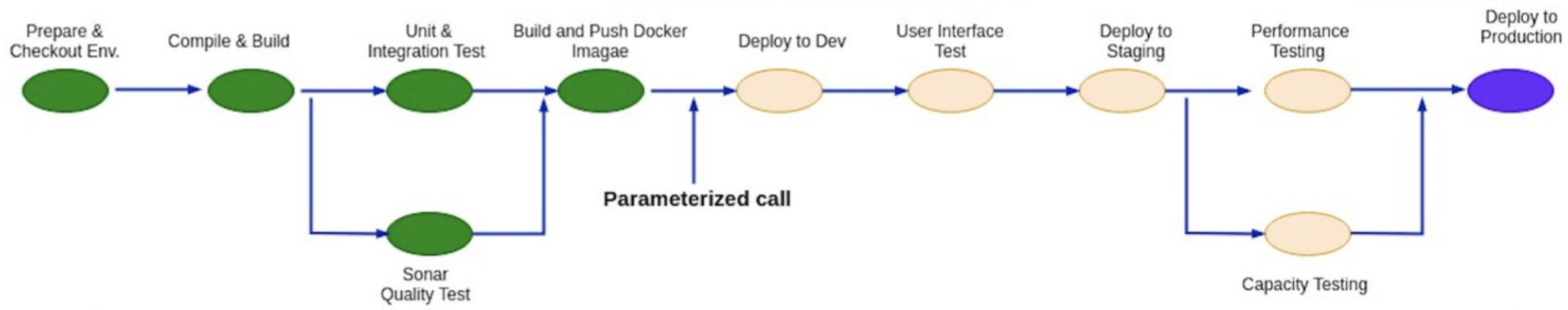
# Development methodology

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)



# DataverseEU deployment in CESSDA Cloud



Jenkins Pipeline from Component Repository

Jenkins Pipeline from Deployment Repository to Cessda Dev Project

Manual Deploy to Production to Production Project

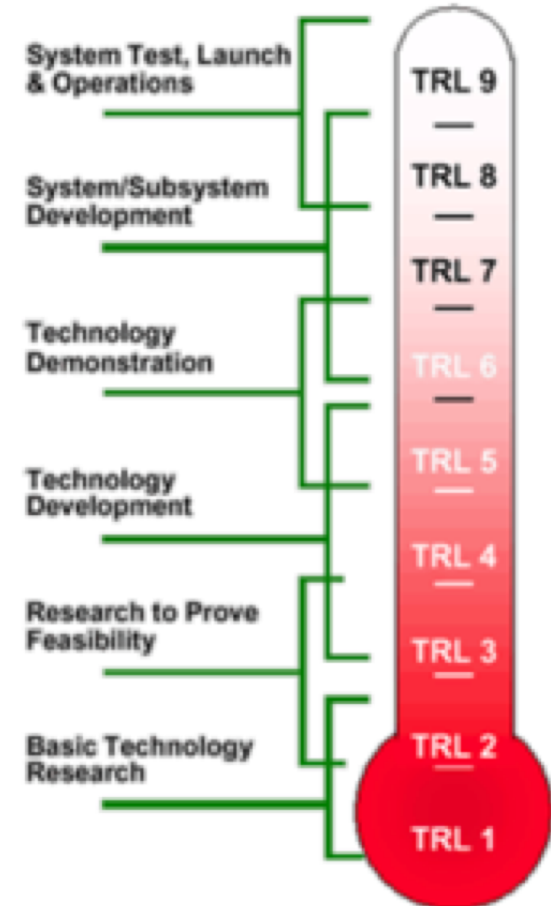
Docker Compose for the local development and testing

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



# 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





# Testing process for the core and apps

Testing process follows the CESSDA maturity model

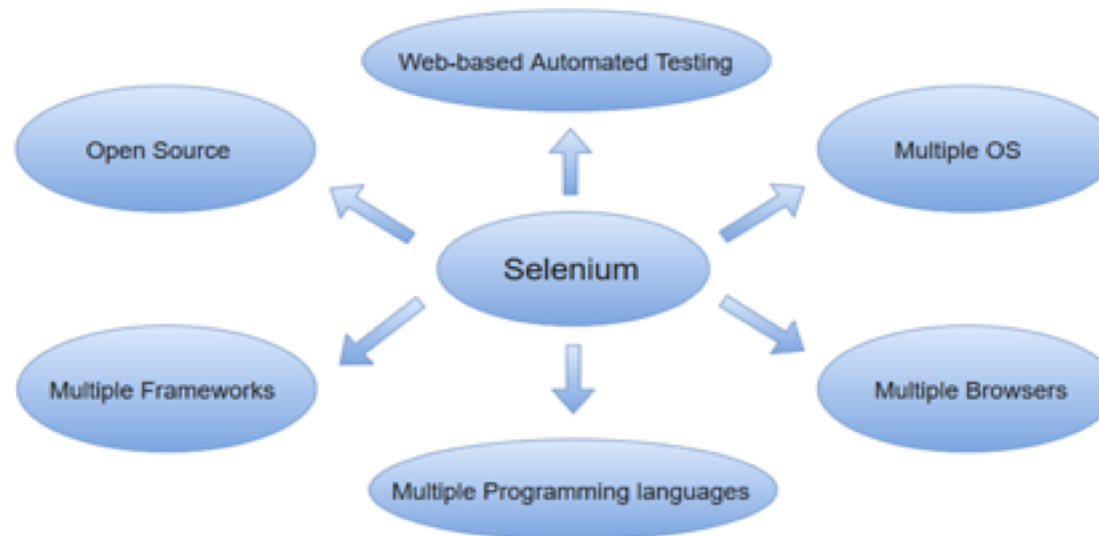
<https://zenodo.org/record/2591055#.XKR6ny2B2u5>

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

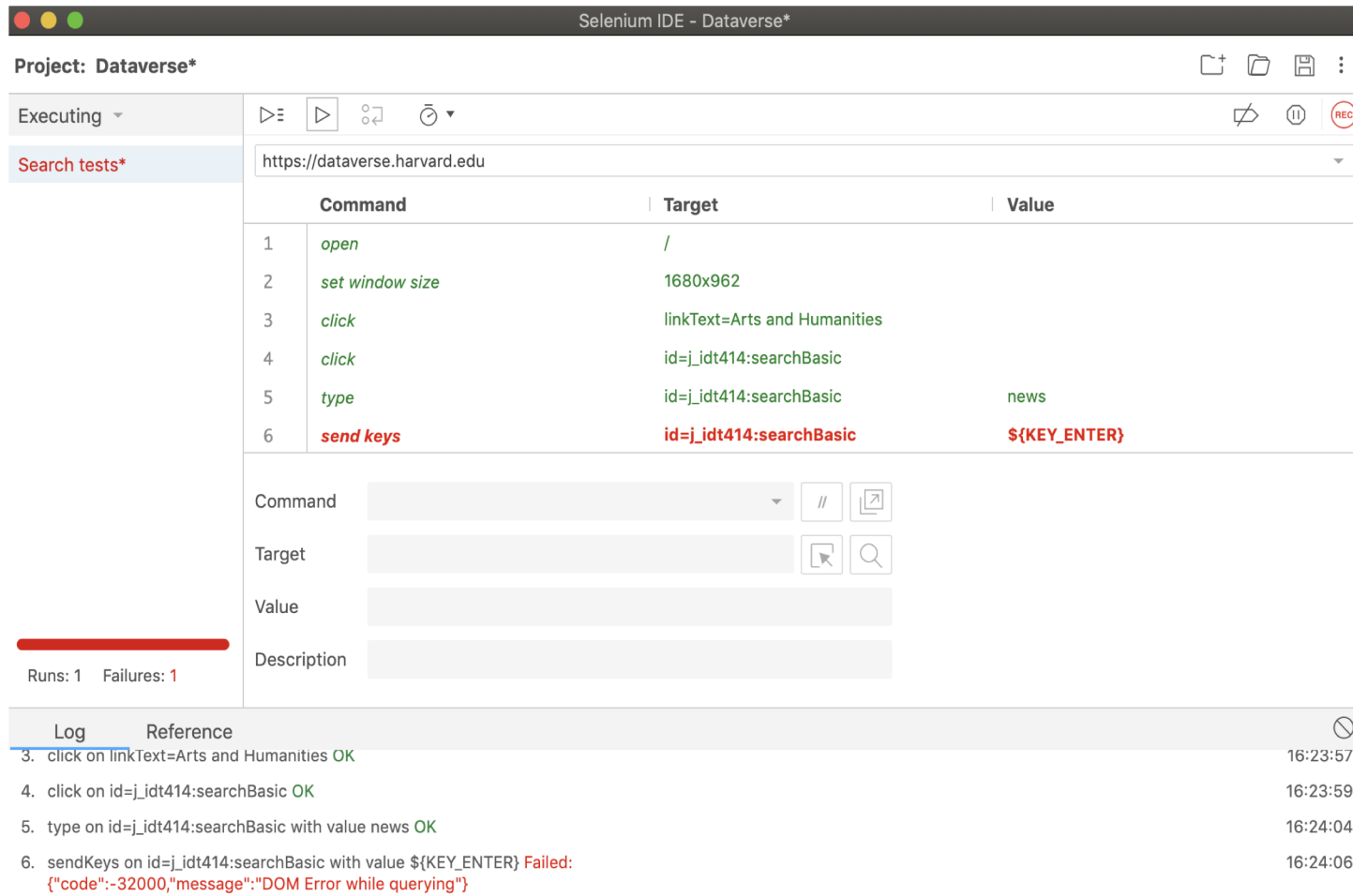
Goal: to score as high as possible according to CESSDA maturity model

# 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



# Quality Assurance (QA) as a community service



The screenshot shows the Selenium IDE interface for a project named 'Dataverse\*'. The URL is 'https://dataverse.harvard.edu'. The test script consists of six steps:

Step	Command	Target	Value
1	open	/	
2	set window size	1680x962	
3	click	linkText=Arts and Humanities	
4	click	id=j_idt414:searchBasic	
5	type	id=j_idt414:searchBasic	news
6	send keys	id=j_idt414:searchBasic	\${KEY_ENTER}

Below the script, there are input fields for Command, Target, Value, and Description. At the bottom, a log shows the execution results:

- 3. click on linkText=Arts and Humanities OK 16:23:57
- 4. click on id=j\_idt414:searchBasic OK 16:23:59
- 5. type on id=j\_idt414:searchBasic with value news OK 16:24:04
- 6. sendKeys on id=j\_idt414:searchBasic with value \${KEY\_ENTER} Failed: 16:24:06  
{"code":-32000,"message":"DOM Error while querying"}

🔗 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!

# 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**





# 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)



# 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

# 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.

# Weblate as Dataverse app

Webate Dashboard Projects Languages

DataverseEU / Bundle.properties ⚠ / bundle\_DE (generated) / translate

Q All strings ▾ 64 / 1568 ▾ Zen

### Translate

**Additional context**  
This field supports only certain .

**Source** Context: htmlAllowedMsg  
htmlAllowedMsg

**Translation** **bundle\_DE (generated)**  
Dieses Feld unterstützt nur bestimmte .

Needs editing ?

Save Suggest Skip

### Things to check

Trailing stop 1

### Glossary

**Source** **Translation**

No related strings found in the glossary.

**Add word to glossary** Add

Source

Translation

### Source information

**Screenshot context**  
No screenshot currently associated!

**Context**  
This field supports only certain .  
htmlAllowedMsg

**Flags**  
No flags currently set!

**Source string age**  
an hour ago

**Translation file**  
de\_AT/Bundle\_de.properties, string 64

Nearby strings 11 Comments Machine translation Other languages History

	Source	Translation	State
59	more	Mehr...	✓
60	less	Weniger...	✓
61	select	Auswählen...	⚠
62	selectedFiles	Ausgewählte Dateien	✓
63			✓

# Controlled Vocabularies support

## Subject \*

- Agricultural Sciences
- Arts and Humanities
- Astronomy and Astrophysics
- Business and Management
- Chemistry

## Keyword (Autocomplete)

### Term

Compulsory and pre-school education

### Vocabulary

Education.CompulsoryAndPreSchool



### Vocabulary URL

https://vocabularies.cessda.eu/TopicCla:

### Term

Family life and marriage

### Vocabulary

SocialStratificationAndGroupings.Family



### Vocabulary URL

https://vocabularies.cessda.eu/TopicCla:

### Term

edu

### Vocabulary



- Compulsory and pre-school education
- EDUCATION
- EDUCATION
- EDUCATION
- EDUCATION
- EDUCATION
- Educational policy
- Higher and further education
- Life-long/continuing education
- Vocational education and training

## Related Publication

## Api Documentation

[ Base URL: /vocabularies.cessda.eu/ ]

[View this documentation on GitHub](#)

Api Documentation

[Terms of service](#)

[Apache 2.0](#)

**basic-error-controller** Basic Error Controller

**cv-controller** CV Controller

**operation-handler** Operation Handler

**resolver-resource** Resolver Resource

**resolver-urn-resource** Resolver Urn Resource

**vocabulary-resource** Vocabulary Resource

**GET** /v1/vocabulary\_getVocabularies

**GET** /v1/Vocabulary/SKOS\_getVocabulariesRaw



# 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

# Thank you for your attention!

[marion.wittenberg@dans.knaw.nl](mailto:marion.wittenberg@dans.knaw.nl)

[vyacheslav.tykhonov@dans.knaw.nl](mailto:vyacheslav.tykhonov@dans.knaw.nl)

Join our community



<https://www.sshopencloud.eu>



[@SSHOpenCloud](https://twitter.com/SSHOpenCloud)



[info@sshopencloud.eu](mailto:info@sshopencloud.eu)



[/in/sshopencloud](https://in.sshopencloud)

