

# SSHOC webinar

## Dataverse, requirements of CESSDA Service Providers

18 March 2020, 11.00 -12.30 CET



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



Laura Huis in 't Veld (DANS)  
functional manager



Slava Tykhonov (DANS)  
lead developer



Marion Wittenberg (DANS)  
task leader



# Content of this Presentation

- 🔗 Background of SSHOC task 5.2
- 🔗 What is Dataverse?
- 🔗 Quality and maturity of software with SSHOC task 5.2
- 🔗 Data previewers
- 🔗 CESSDA metadata and controlled vocabularies
- 🔗 Translations using weblate

Project:



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 (February 2022): Data repository service running on EOSC

After 40 months (April 2022): Report on principles of governance and sustainability of the data repository service

# 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



# History of task 5.2

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

## CESSDA SaW project

-  DANS offered a test environment of Dataverse to run experiments

## CESSDA DataverseEU project (2018)

-  POC how to make Dataverse CESSDA compliant

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



# Why this webinar

- 🌀 To demonstrate our work
- 🌀 To discuss with you your ideas on:
  - 🌀 Dataverse installation on your own technical environment
  - 🌀 Dataverse running on the CESSDA Google cloud
  - 🌀 What kind of extra functionality, like data viewers
  - 🌀 Translations
  - 🌀 Metadata, controlled vocabularies



# What is Dataverse?

- Repository software for sharing and publishing datasets
  - Metadata and files (all kind off file formats)
  - Roles and permissions for users to control access
- Open Source Software developed by IQSS Harvard
- Lots of other organisations contribute to the development of the software (Global Dataverse Community Consortium ([GDCC](#)))
- A few examples of the 55 dataverse services:
  - <https://dataverse.nl> - Netherlands
  - <https://data.aussda.at> - Austria
  - <https://dataverse.harvard.edu> - Harvard USA
  - <https://dataverse.unc.edu> - ODUM Institute USA
  - <https://dataverse.adu.edu.au> - Australia
  - <https://dataverse.scholarsportal.info> - Canada

# Welcome to DataverseNL



Utrecht University

Utrecht University



Maastricht University



Tilburg University

rijksuniversiteit  
 groningen

University of Groningen



Search this dataverse...

Find

Advanced Search

☒ Datasets (347)☒ Files (1,338)☐ Files (7,587)

## Dataverse Category

Organization or Institution (108)

Research Group (59)

Research Project (57)

Department (22)

Researcher (13)

More...

## Publication Year

2019 (651)

2016 (223)

2018 (202)

2013 (163)

2017 (156)

1 to 10 of 1,685 Results

Sort

### OCISS: Effect of spectral contrast enhancement on speech-on-speech intelligibility and voice cue sensitivity in cochlear implant users

Mar 17, 2020 - Ear Nose Throat UMCG Dataverse



El Boghdady, Nawal; Langner, Florian; Gaudrain, Etienne; Bagkent, Deniz; Nogueira, Waldo, 2020, "OCISS: Effect of spectral contrast enhancement on speech-on-speech intelligibility and voice cue sensitivity in cochlear implant users", <https://hdl.handle.net/10411/GPSSP1>, DataverseNL, V1

JND\_data.csv: Just-noticeable-differences for German CI participants for F0 and VTL. SoS\_Intelligibility\_data.csv: Speech-on-speech intelligibility data for German CI participants for F0 and VTL manipulations of the masker and for 2 coding strategies. SoS\_Comprehension\_data.csv: S...

### OCISS: Effect of Channel Interaction on Vocal Cue Perception in Cochlear Implant Users

Mar 13, 2020 - Ear Nose Throat UMCG Dataverse



Nogueira, Waldo; El Boghdady, Nawal; Langner, Florian; Gaudrain, Etienne; Baskent, Deniz, 2020, "OCISS: Effect of Channel Interaction on Vocal Cue Perception in Cochlear Implant Users", <https://hdl.handle.net/10411/TYWSXR>, DataverseNL, V1

JND\_data.csv: Just-noticeable-differences for German CI participants for F0 and VTL. SoS\_Intelligibility\_data.csv: Speech-on-speech intelligibility data for German CI participants for F0 and VTL manipulations of the masker and for three simulation patterns. SoS\_Comprehension\_data...

### A Process-oriented Dataset of Revisions during Writing

Mar 12, 2020 - Department Cognitive Science and Artificial Intelligence





## Increases of correct memories and spontaneous false memories due to eye movements when memories are retrieved after a time delay

**Version 1.0**

Houben, Sanne T.L.; Otgaar, Henry; Roelofs, Jeffrey; Smeets, Tom; Merckelbach, Harald, 2020, "Increases of correct memories and spontaneous false memories due to eye movements when memories are retrieved after a time delay", <https://hdl.handle.net/10411/1XDVH1>, DataverseNL, V1

Cite Dataset ▾

Learn about [Data Citation Standards](#).

### Dataset Metrics ⓘ

0 Downloads ⓘ

### Description ⓘ

Eye Movement Desensitization and Reprocessing (EMDR) is an effective treatment for post-traumatic stress disorder. However, literature on possible adverse memory effects of EMDR is scarce. Using the Deese/Roediger- McDermott (DRM) false memory paradigm, we examined the susceptibility to spontaneous false memories after performing eye movements, as used in EMDR. In Experiment 1, 72 undergraduates received word lists containing negative and neutral associated words and immediately after this they were given a free recall and recognition test. In Experiment 2, 68 undergraduates underwent the free recall and recognition test 48 h later. During the free recall phase in both experiments, participants either performed eye movements or not (control condition). In Experiment 1, the two conditions did not differ statistically with regard to correct and false recall/ recognition. In Experiment 2, correct memory rates were higher in the eye movement than in the control condition and this was accompanied by an increase in spontaneous false memories on both free recall and recognition. Although our experimental approach is far removed from clinical practice, our findings suggest that eye movements as used in EMDR might amplify both correct and false memory rates.

### Subject ⓘ

Social Sciences

### Keyword ⓘ

Eye movements, EMDR, False memory, DRM, Therapy

Files

Metadata

Terms

Versions

Search this dataset...

Find

Filter by

File Type: All ▾ Access: All ▾

Sort ▾

1 to 3 of 3 Files

Request Access



#### S1\_FMEM\_LIBERAL.sav

SPSS Binary - 118.3 KB - Mar 10, 2020 - 0 Downloads  
MD5: ac32b7afc06d56e9dc0301b20fdde1c9

Request Access



#### S1\_FMEM\_LIBERAL\_FINAL\_EXP1.sav

SPSS Binary - 80.1 KB - Mar 10, 2020 - 0 Downloads  
MD5: 1e3ccdab74860f57bc6a43f8efbf8cd1

Request Access

## Permissions ^

Current access configuration to your dataverse.

Select if all users or only certain users are able to add to this dataverse, by clicking the Edit Access button.

 Edit Access**Who can add to this dataverse?**


Anyone adding to this dataverse needs to be given access

**When a user adds a new dataset to this dataverse, which role should be automatically assigned to them on that dataset?**

Contributor - Edit metadata, upload files, and edit files, edit Terms, Guestbook, Submit datasets for review

## Users/Groups ^


All the users and groups that have access to your dataverse.

 Assign Roles to Users/Groups**1 User/Group**

User/Group Name (Affiliation) ^	ID ^	Role ^	Action
Dataverse Admin (DANS)	@dataverseAdmin	Admin	 Remove Assigned Role

## Roles ^

All the roles set up in your dataverse, that you can assign to users and groups.

 Add New Role

**Admin** - A person who has all permissions for dataverses, datasets, and files. 

[AddDataverse](#) [AddDataset](#) [ViewUnpublishedDataverse](#) [ViewUnpublishedDataset](#) [DownloadFile](#) [EditDataverse](#) [EditDataset](#) [ManageDataversePermissions](#)  
[ManageDatasetPermissions](#) [PublishDataverse](#) [PublishDataset](#) [DeleteDataverse](#) [DeleteDatasetDraft](#)

**Contributor** - For datasets, a person who can edit License + Terms, and then submit them for review. 

[ViewUnpublishedDataset](#) [DownloadFile](#) [EditDataset](#) [DeleteDatasetDraft](#)

# Would you like to 'play' with dataverse?

Please use the demo site of DANS:  
<https://demo.dataverse.nl/dataverse/general>

Note: User permissions for production dataverses are usually more restrictive



# Main differences compared to NESSTAR

- 🌀 Dataverse is suitable for all kinds of data, not only survey data
  - 🌀 Dataverse has a persistent identifier (study level, and possible for file level)
  - 🌀 Dataverse can keep track of versions of the dataset
  - 🌀 Federated login
  - 🌀 Dataverse has a variety of user permissions
- 
- 🌀 Dataverse is not very suitable for publishing survey data in a very granular way (but we are working on data viewers)

# Questions for this first part?







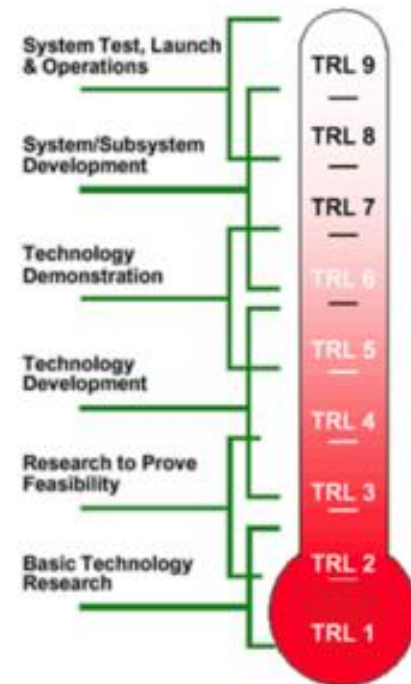
---

# SSHOC Dataverse

Overview of functionality we would like to develop

# Quality of services in EOSC (CESSDA Cloud)

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



# Dataverse applications maturity

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

Must have: Kubernetes 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, for example, you can connect PDF viewer deployed on CESSDA Cloud to your own Dataverse instance.

# Dataverse App Store

Weblate as a service: multilingual translations (September 2020)

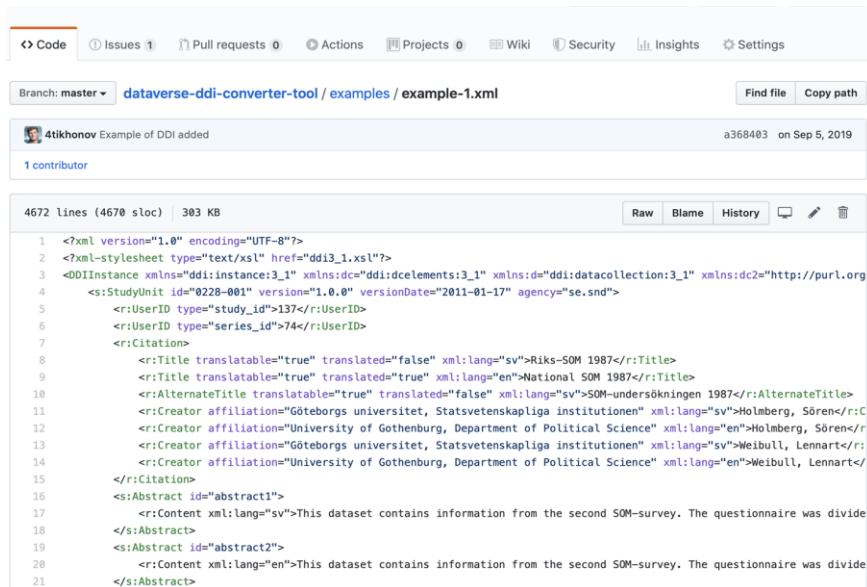
Interoperability: external controlled vocabularies (CESSDA Vocabulary service, December 2020)

Data previewers: DDI Explorer, Spreadsheet/CSV, PDF, Text files, HTML, Images, video render, audio, JSON, GeoJSON/Shapefiles/Map, XML (June 2021 in CESSDA Cloud)

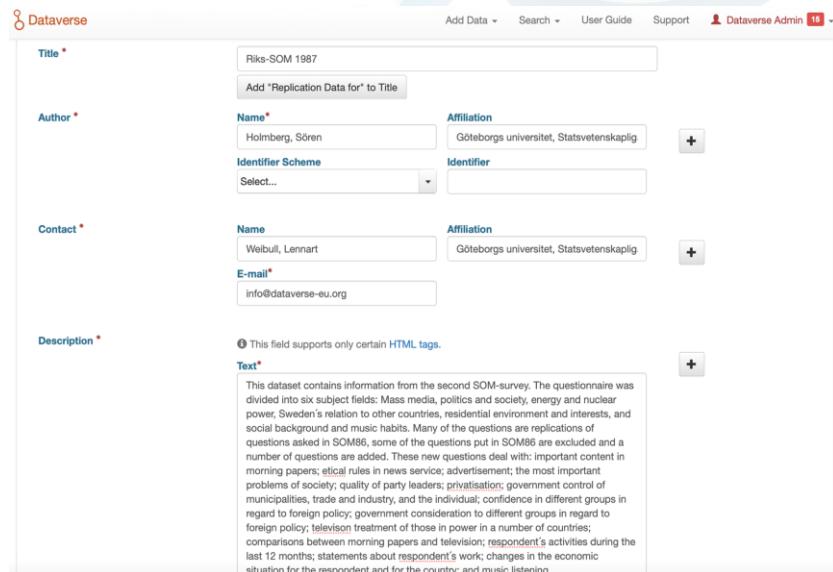
Data processing: NESSTAR DDI migration tool (June 2021 in CESSDA Cloud)

# NESSTAR DDI migration tool

Pipeline: NESSTAR XML results as metadata and files stored in Dataverse



The screenshot shows the GitHub interface for the repository 'dataverse-ddi-converter-tool'. The 'Code' tab is selected, displaying the XML content of 'example-1.xml'. The XML is a DDI (Data Documentation Initiative) instance for a study unit. It includes metadata such as the title 'Riks-SOM 1987', author 'Holmberg, Sören', and affiliation 'Göteborgs universitet, Statsvetenskapliga institutionen'. The XML also contains abstracts in Swedish and English. The repository has 4672 lines of code, 4670 SLOC, and 383 KB. It was last updated on September 5, 2019, by user '4tkhonov'.



The screenshot shows the Dataverse interface for the dataset 'Riks-SOM 1987'. The 'Title' field is filled with 'Riks-SOM 1987'. The 'Author' field is filled with 'Holmberg, Sören' and the 'Affiliation' field is filled with 'Göteborgs universitet, Statsvetenskaplig'. The 'Identifier Scheme' is set to 'Select...'. The 'Contact' field is filled with 'Weibull, Lennart' and the 'Affiliation' field is filled with 'Göteborgs universitet, Statsvetenskaplig'. The 'E-mail' field is filled with 'info@dataverse-eu.org'. The 'Description' field is filled with a detailed description of the dataset, which is a replication of the Riks-SOM 1987 survey. The description mentions that the dataset contains information from the second SOM-survey and that the questionnaire was divided into six subject fields: Mass media, politics and society, energy and nuclear power, Sweden's relation to other countries, residential environment and interests, and social background and music habits. The description also mentions that many of the questions are replications of questions asked in SOM86, some of which are excluded and a number of new questions are added. These new questions deal with: important content in morning papers; ethical rules in news service; advertisement; the most important problems of society; quality of party leaders; privatisation; government control of municipalities, trade and industry, and the individual; confidence in different groups in regard to foreign policy; government consideration to different groups in regard to foreign policy; television treatment of those in power in a number of countries; comparisons between morning papers and television; respondent's activities during the last 12 months; statements about respondent's work; changes in the economic situation for the respondent and for the country; and music listening.

NESSTAR study unit is metadata, variables will be saved as tab file that can be opened in DDI preview.

More information: <https://github.com/IQSS/dataverse-ddi-converter-tool/>

# Dataverse previewers

All Dataverse previewers are maintained by Global Dataverse Community Consortium (GDCC).


Most of them developed by institute QDR (USA), Dataverse SSHOC contributed Spreadsheet viewer for tabular data.

DDI Explorer was developed by Scholars Portal (Canada) and integrated to Kubernetes by SSHOC project.

Available viewers: Spreadsheet/CSV, PDF, Text files, HTML, images, video render, audio preview, STATA, DDI Explorer


All previewers are available only for Open Access Data.

# Example: PDF previewer

 Dataverse Add Data ▾ Search ▾ User Guide Support Sign Up Log In

This file is part of "Bureauonderzoek Archeologie Centrumplan Eelde".


**Dataset Citation**  
Mol, K.J.J.N., 2018, "Bureauonderzoek Archeologie Centrumplan Eelde", <https://doi.org/10.17026/dans-2c6-w58x>, EASY, V1

 Cite Dataset ▾ [Learn about Data Citation Standards.](#)

Preview Metadata Versions

Open Read Document

Previous Next Page: 1 / 27



**BUREAUONDERZOEK ARCHEOLOGIE  
CENTRUMPLAN EELDE**  
Arcadis Archeologische Rapporten [2017 - 116]

15 FEBRUARI 2017

Built-in Dataverse viewer



## PDF Preview


Filename: AAR 116 Centrumplan Eelde.pdf  
In *Bureauonderzoek Archeologie Centrumplan Eelde* (version 1.0), by Mol, K.J.J.N.

Download File Close Preview

File uploaded on 2020-02-19

Previous Next Page: 2 / 27

BUREAUONDERZOEK ARCHEOLOGIE CENTRUMPLAN EELDE



**Contactpersonen**

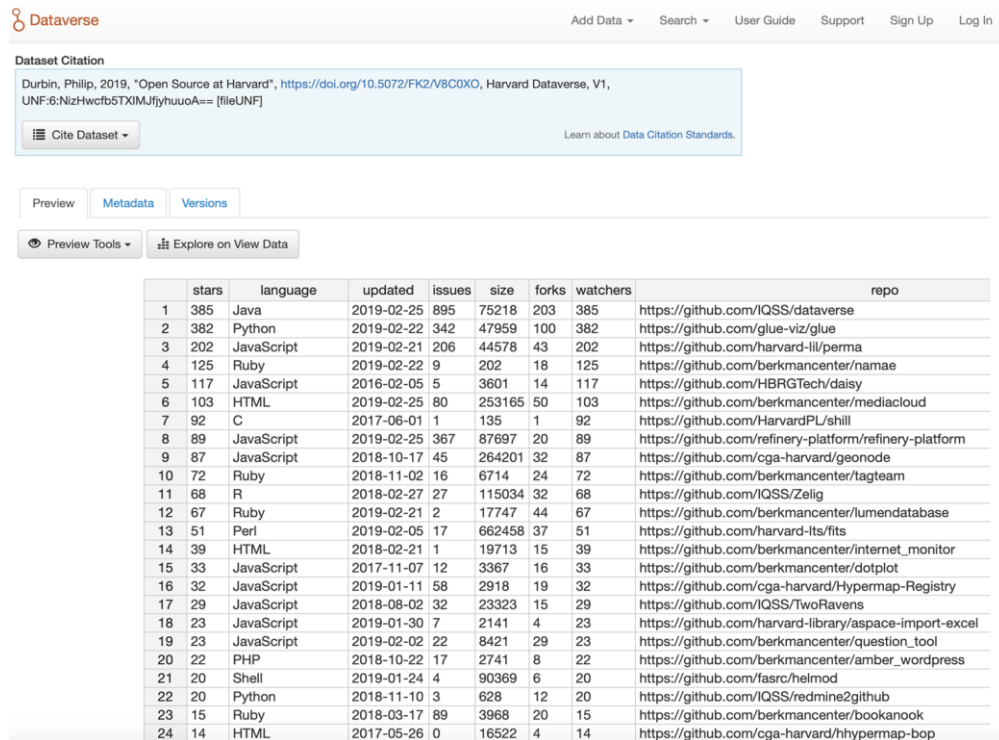
**I. BENJAMINS**  
Senior-Adviseur Archeologie en Cultuurhistorie  
E: ingrid.benjamins@arcadis.com

**K. MOL**  
Adviseur Archeologie en Cultuurhistorie  
E: kees.mol@arcadis.com

Arcadis Nederland B.V.  
Postbus 264  
3800 AD Arnhem  
Nederland

Preview as a widget

# Example: Spreadsheet previewer



**Dataverse** Add Data Search User Guide Support Sign Up Log In

**Dataset Citation**

Durbin, Philip, 2019, "Open Source at Harvard", <https://doi.org/10.5072/FK2/V8C0XO>, Harvard Dataverse, V1, UNF:6:NizHwcfbSTXIMJfyhuuoA== [fileUNF]

[Cite Dataset](#) [Learn about Data Citation Standards.](#)

Preview Metadata Versions

[Preview Tools](#) [Explore on View Data](#)


	stars	language	updated	issues	size	forks	watchers	repo
1	385	Java	2019-02-25	895	75218	203	385	<a href="https://github.com/IQSS/dataverse">https://github.com/IQSS/dataverse</a>
2	382	Python	2019-02-22	342	47959	100	382	<a href="https://github.com/glue-viz/glue">https://github.com/glue-viz/glue</a>
3	202	JavaScript	2019-02-21	206	44578	43	202	<a href="https://github.com/harvard-ill/perma">https://github.com/harvard-ill/perma</a>
4	125	Ruby	2019-02-22	9	202	18	125	<a href="https://github.com/berkmancenter/namea">https://github.com/berkmancenter/namea</a>
5	117	JavaScript	2016-02-05	5	3601	14	117	<a href="https://github.com/HBRGTech/daisy">https://github.com/HBRGTech/daisy</a>
6	103	HTML	2019-02-25	80	253165	50	103	<a href="https://github.com/berkmancenter/mediacloud">https://github.com/berkmancenter/mediacloud</a>
7	92	C	2017-06-01	1	135	1	92	<a href="https://github.com/HarvardPL/shill">https://github.com/HarvardPL/shill</a>
8	89	JavaScript	2019-02-25	367	87697	20	89	<a href="https://github.com/refinery-platform/refinery-platform">https://github.com/refinery-platform/refinery-platform</a>
9	87	JavaScript	2018-10-17	45	264201	32	87	<a href="https://github.com/cga-harvard/geonode">https://github.com/cga-harvard/geonode</a>
10	72	Ruby	2018-11-02	16	6714	24	72	<a href="https://github.com/berkmancenter/tagteam">https://github.com/berkmancenter/tagteam</a>
11	68	R	2018-02-27	27	115034	32	68	<a href="https://github.com/IQSS/Zelig">https://github.com/IQSS/Zelig</a>
12	67	Ruby	2019-02-21	2	17747	44	67	<a href="https://github.com/berkmancenter/lumendatabase">https://github.com/berkmancenter/lumendatabase</a>
13	51	Perl	2019-02-05	17	662458	37	51	<a href="https://github.com/harvard-lts/lts">https://github.com/harvard-lts/lts</a>
14	39	HTML	2018-02-21	1	19713	15	39	<a href="https://github.com/berkmancenter/internet_monitor">https://github.com/berkmancenter/internet_monitor</a>
15	33	JavaScript	2017-11-07	12	3367	16	33	<a href="https://github.com/berkmancenter/dotplot">https://github.com/berkmancenter/dotplot</a>
16	32	JavaScript	2019-01-11	58	2918	19	32	<a href="https://github.com/cga-harvard/Hypermap-Registry">https://github.com/cga-harvard/Hypermap-Registry</a>
17	29	JavaScript	2018-08-02	32	23323	15	29	<a href="https://github.com/IQSS/TwoRavens">https://github.com/IQSS/TwoRavens</a>
18	23	JavaScript	2019-01-30	7	2141	4	23	<a href="https://github.com/harvard-library/aspace-import-excel">https://github.com/harvard-library/aspace-import-excel</a>
19	23	JavaScript	2019-02-02	22	8421	29	23	<a href="https://github.com/berkmancenter/question_tool">https://github.com/berkmancenter/question_tool</a>
20	22	PHP	2018-10-22	17	2741	8	22	<a href="https://github.com/berkmancenter/amber_wordpress">https://github.com/berkmancenter/amber_wordpress</a>
21	20	Shell	2019-01-24	4	90369	6	20	<a href="https://github.com/fasrc/helmod">https://github.com/fasrc/helmod</a>
22	20	Python	2018-11-10	3	628	12	20	<a href="https://github.com/IQSS/redmine2github">https://github.com/IQSS/redmine2github</a>
23	15	Ruby	2018-03-17	89	3968	20	15	<a href="https://github.com/berkmancenter/bookanook">https://github.com/berkmancenter/bookanook</a>
24	14	HTML	2017-05-26	0	16522	4	14	<a href="https://github.com/cga-harvard/hypermap-bop">https://github.com/cga-harvard/hypermap-bop</a>

Some features:

- Can be sorted by values in columns
- select functionality to copy and paste rows/columns/values
- values can be modified but not saved



# Image previewer


 Dataaverse

[Add Data](#) [Search](#) [User Guide](#) [Support](#) [Sign Up](#) [Log In](#)

[Export Metadata](#)

[File Metadata](#)

**Preview**



**SHA1** 1849eca39d4f8bc8d660c2dc1b475d93f3907cb

**Publication Date** 2020-02-24

**Size** 757.3 KB

**Type** JPEG Image

**File Path** images/

**Deposit Date** 2020-02-19

# Questions concerning part 2?



# Configuring Dataverse for CESSDA

## - CESSDA Mandatory Metadata fields (CMM)

DataverseEU : Comparison made between standard Dataverse metadata fields and CMM.

SSHOC: Implement findings from the DataverseEU project

# Dataverse controlled vocabularies plugin

- Developed by DANS in CESSDA Dataverse project, extended as a plugin in SSHOC Dataverse
- Implemented as simple xhtml extension (javascript) of Dataverse templates
- can be enabled by providing CV middleware API endpoint (internal/external microservice)
- can be connected to any external controlled vocabulary exposed through API endpoint

# Current situation (default dataverse)



Topic Classification ?

Term ?

Vocabulary ?



Vocabulary URL ?

Enter full URL, starting with http://

# Desired situation

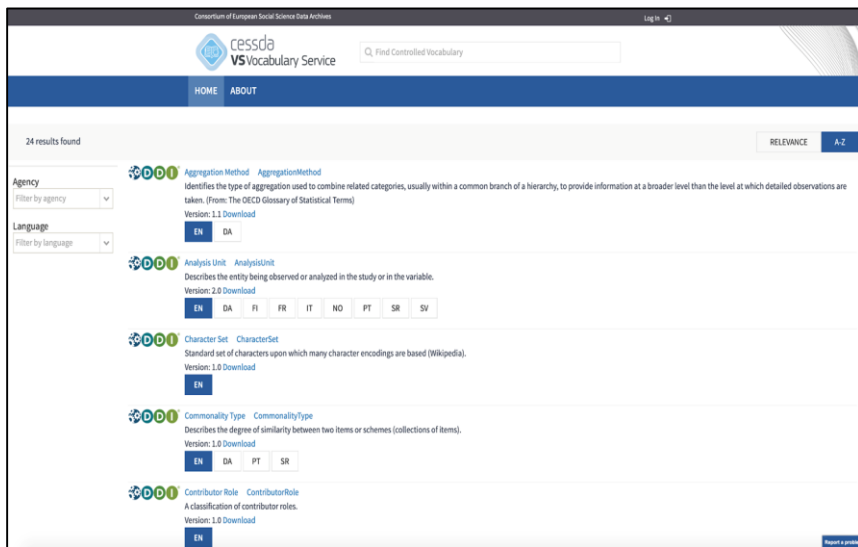
- 🔗 For certain fields, users are required to use a specific field
- 🔗 When typing, the system will return search results from this CV only

## Mapping:

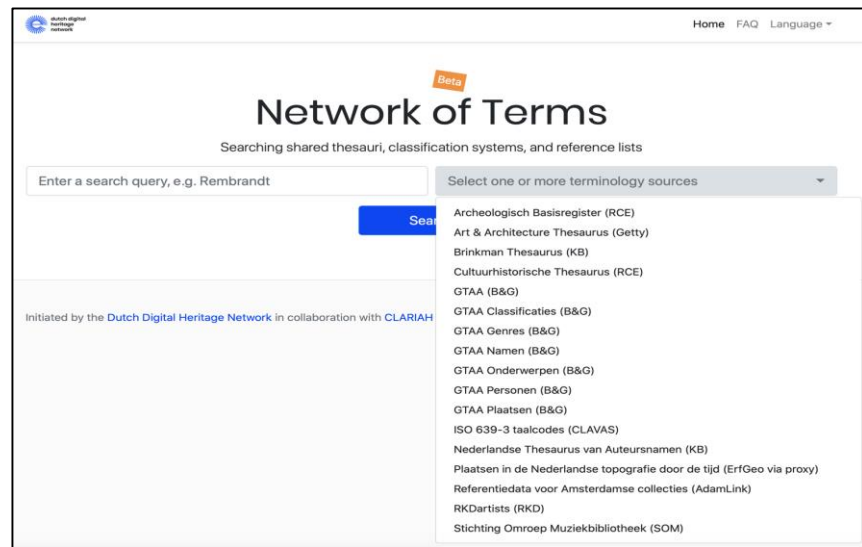
Dataverse metadata field	Controlled Vocabulary
Keyword	ELSST
Topic Classification	CESSDA Topic Classification
Kind of Data	<a href="#"><u>DDI Controlled Vocabulary for General Data Format</u></a>
Unit of Analysis	<a href="#"><u>DDI Controlled Vocabulary for Analysis Unit</u></a>
Time Method	<a href="#"><u>DDI Controlled Vocabulary for Time Method</u></a>
Sampling Procedure	<a href="#"><u>DDI Controlled Vocabulary for Sampling Procedure</u></a>

# External controlled vocabularies support

Dataverse SSHOC project developing a plugin that allows to connect repository to API endpoints delivering controlled vocabularies:

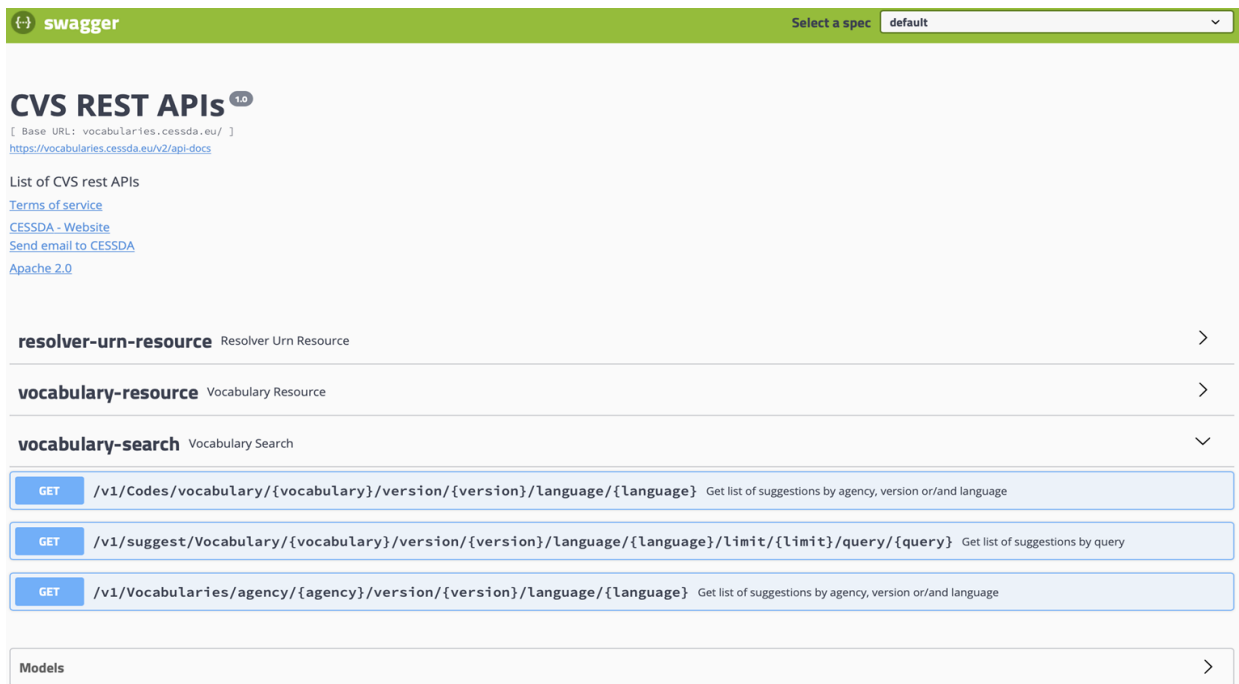


CESSDA Vocabulary service



NDE network

# CESSDA CVs Open API



The image shows the Swagger UI for the CESSDA CVs Open API. At the top, there is a green bar with the Swagger logo and a dropdown menu labeled "Select a spec" with "default" selected. Below this, the title "CVS REST APIs" is displayed with a "1.0" version indicator. A small text block provides the base URL: "[ Base URL: vocabularies.cessda.eu/ ]" and a link to the API docs: "[https://vocabularies.cessda.eu/v2/api-docs](\"https://vocabularies.cessda.eu/v2/api-docs\")". A section titled "List of CVS rest APIs" contains links for "Terms of service", "CESSDA - Website", "Send email to CESSDA", and "Apache 2.0". Below this, three API endpoints are listed: "resolver-urn-resource" (Resolver Urn Resource), "vocabulary-resource" (Vocabulary Resource), and "vocabulary-search" (Vocabulary Search). Each endpoint has a "GET" method and a description of its function. At the bottom, there is a section for "Models" with a right arrow.

swagger Select a spec default

## CVS REST APIs <sup>1.0</sup>

[ Base URL: vocabularies.cessda.eu/ ]  
<https://vocabularies.cessda.eu/v2/api-docs>

List of CVS rest APIs

[Terms of service](#)  
[CESSDA - Website](#)  
[Send email to CESSDA](#)  
[Apache 2.0](#)

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

**vocabulary-resource** Vocabulary Resource >

**vocabulary-search** Vocabulary Search >

**GET** /v1/Codes/vocabulary/{vocabulary}/version/{version}/language/{language} Get list of suggestions by agency, version or/and language

**GET** /v1/suggest/Vocabulary/{vocabulary}/version/{version}/language/{language}/limit/{limit}/query/{query} Get list of suggestions by query

**GET** /v1/Vocabularies/agency/{agency}/version/{version}/language/{language} Get list of suggestions by agency, version or/and language

**Models** >

- Developed by GESIS in Germany
- Maintained by CESSDA
- Deployed on CESSDA Cloud k8s
- Has multilingual support
- Shows different versions of CVs
- Elasticsearch syntax to query



# Example of term details in CESSDA Vocabulary service

## Family: Household Family in English

Family.HouseholdFamily	Family: Household family	A more specific term that refers only to people who are related through family ties (see definition for 'Family') and live in the same household at a point in time. If not known whether the analysis unit is 'Family' or 'Household family', use 'Family'.
Household	Household	A person or a group of persons who share the same dwelling unit and common living arrangements. These common living arrangements may include pooling some, or all, of their income and wealth, and consuming certain types of goods and services collectively, mainly housing and food (Eurostat).
HousingUnit	Housing unit	U.S. Census: A housing unit is a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied (or if vacant, is intended for occupancy) as separate living quarters. Separate living quarters are those in which the occupants live and eat separately from any other persons in the building and which have direct access from the outside of the building or through a common hall.

## Family: Household Family in French (Famille: ménage)

Family.HouseholdFamily	Famille : ménage	Un terme plus spécifique, se réfère aux personnes qui sont liées par des liens familiaux (voir la définition pour 'Famille') et qui vivent dans le même ménage à un moment donné. Si vous ne savez pas si l'unité d'analyse est 'Famille' ou 'Famille : ménage', utilisez 'Famille'.
Household	Ménage	Une personne ou un groupe de personnes qui partagent la même unité d'habitation et les mêmes conditions de vie. Ces conditions de vie peuvent inclure la mise en commun d'une partie, ou de la totalité, des revenus et des richesses, et la consommation de certains types de biens et de services, principalement le logement et la nourriture (Eurostat).
HousingUnit	Logement	Recensement US : une unité d'habitation est une maison, un appartement, un mobile home, un groupe de pièces, ou une pièce unique qui est occupée (ou si vacante, est destinée à l'occupation) comme lieu de vie séparé. Les lieux de vie séparés sont ceux dans lesquels les occupants vivent et mangent séparément de toute autre personne dans le bâtiment et qui ont un accès direct depuis l'extérieur du bâtiment ou par un hall commun.

Keywords are language dependent:

if Dataverse interface will be switched to another language (for example, French), it will query CESSDA vocabulary service to find keywords in the selected language.

Keyword linked to the same code, doesn't matter in which language!

# Questions about part 3?



# Weblate introduction

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

New releases of dataverse software can induce changes in User Interface, e.g. adding or deleted buttons. Issue is: how to keep track?

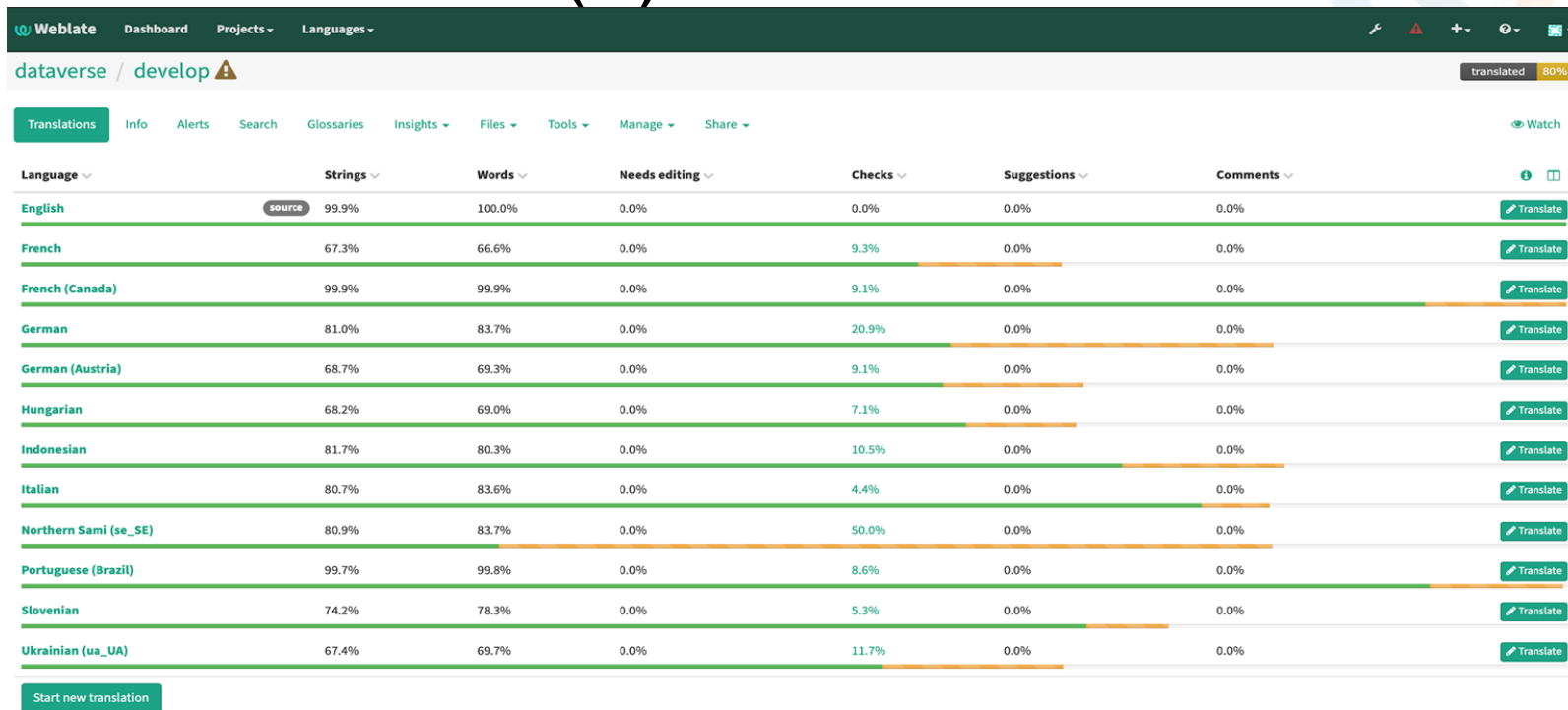
We've developed an experimental but adjustable pipeline for multilingual support that allows to download and synchronize all translations and provides easy access for translators to keep all properties up-to-date.

All translations are available in Dataverse Consortium github.

# Weblate as community service

- All translators will get account and permissions
- after Dataverse release will get new properties, all properties will be synchronized and available in Weblate
- translators will receive a notifications with assigned tasks and will start a collaborative work
- Weblate GUI indicates all missing translations and provides statistics to admin with information what's missing
- After all translations are ready, updated properties will be published as language packages in GDCC github repo:  
<http://github.com/GlobalDataverseCommunityConsortium/dataverse-language-packs>
- Every Dataverse version has own branch in github and handled separately by Weblate

# Weblate demo (1) - available translations

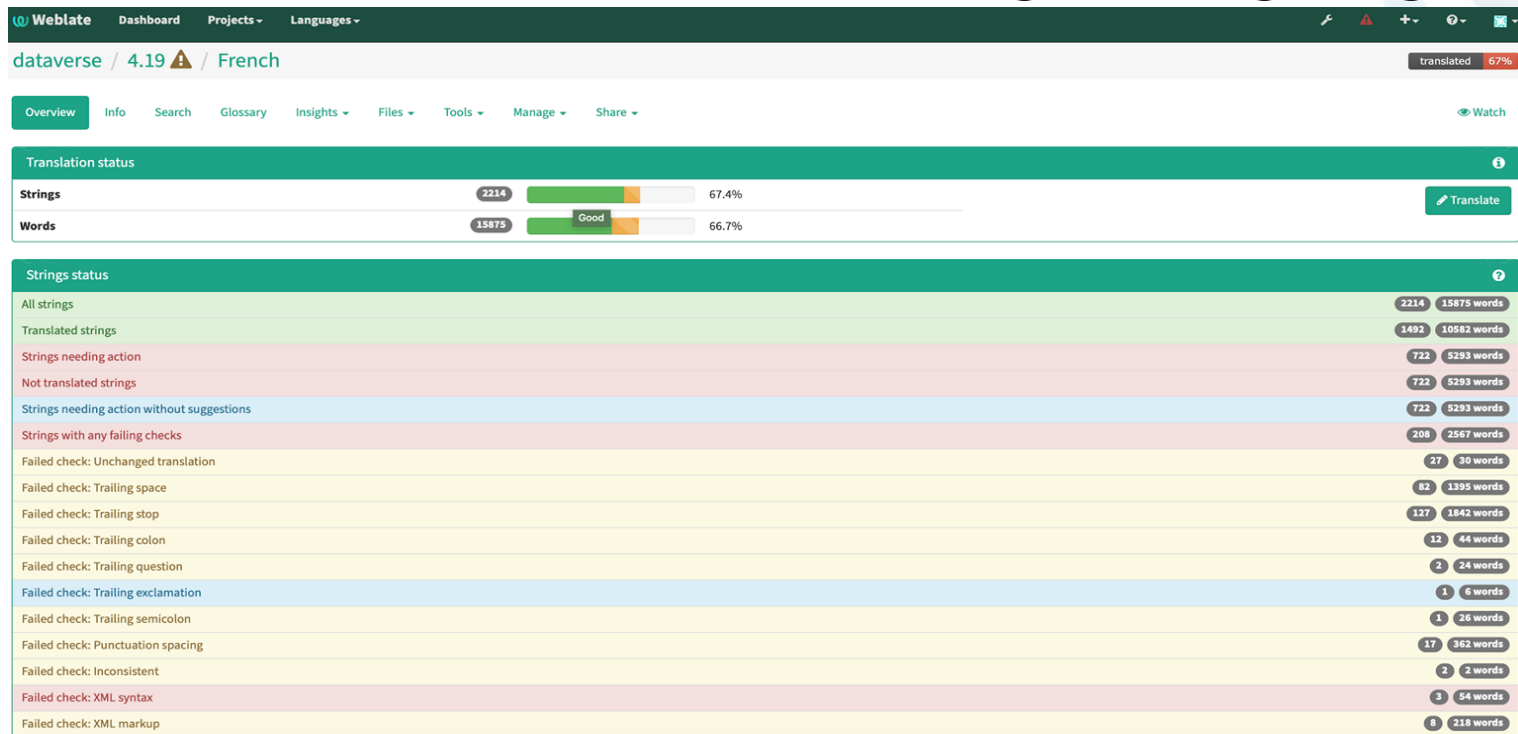


Language	Strings	Words	Needs editing	Checks	Suggestions	Comments	
English	source 99.9%	100.0%	0.0%	0.0%	0.0%	0.0%	<a href="#">Translate</a>
French	67.3%	66.6%	0.0%	9.3%	0.0%	0.0%	<a href="#">Translate</a>
French (Canada)	99.9%	99.9%	0.0%	9.1%	0.0%	0.0%	<a href="#">Translate</a>
German	81.0%	83.7%	0.0%	20.9%	0.0%	0.0%	<a href="#">Translate</a>
German (Austria)	68.7%	69.3%	0.0%	9.1%	0.0%	0.0%	<a href="#">Translate</a>
Hungarian	68.2%	69.0%	0.0%	7.1%	0.0%	0.0%	<a href="#">Translate</a>
Indonesian	81.7%	80.3%	0.0%	10.5%	0.0%	0.0%	<a href="#">Translate</a>
Italian	80.7%	83.6%	0.0%	4.4%	0.0%	0.0%	<a href="#">Translate</a>
Northern Sami (se_SE)	80.9%	83.7%	0.0%	50.0%	0.0%	0.0%	<a href="#">Translate</a>
Portuguese (Brazil)	99.7%	99.8%	0.0%	8.6%	0.0%	0.0%	<a href="#">Translate</a>
Slovenian	74.2%	78.3%	0.0%	5.3%	0.0%	0.0%	<a href="#">Translate</a>
Ukrainian (ua_UA)	67.4%	69.7%	0.0%	11.7%	0.0%	0.0%	<a href="#">Translate</a>

[Start new translation](#)

Source: [http://weblate.dataverse.org/projects/dataverse/4\\_19main/](http://weblate.dataverse.org/projects/dataverse/4_19main/)

# Weblate demo (2) - manage language



# Weblate demo (3) - properties to translate

Weblate Dashboard Projects Languages

dataverse / 4.19 / French / translate translated 67%

1 / 722

Translate

Source: Original Format Context: downloadOriginal

Translation: French

Needs editing

Save Suggest Skip

Glossary

Source Translation

No related strings found in the glossary.

Add word to glossary

Source

Translation

Source information

Screenshot context

No screenshot currently associated!

Context

downloadOriginal

Flags

No flags currently set!

Source string age

4 minutes ago

Translation file

fr\_FR/Bundle\_fr.properties, string 46

Nearby strings 11 Other occurrences 2 Comments Machine translation Translation memory Other languages History

	Source	Translation	State
41	Contributor	Contributeur	✓
42	Manager	Gestionnaire	✓
43	Curator	Conservateur	✓
44	Explore	Explorer	✓
45	Download	Télécharger	✓
46	Original Format		✗
47	Archival Format (.tab)		✗
48	Deaccession	Retrait	✓

# Weblate demo (4) - info and repository to push

**Project website** <https://dataverse.org>

**Translation process**

- Translations can be made directly.
- Translation suggestions can be made.
- Any authenticated user can contribute.
- The translation uses monolingual files.
- The translation base language is editable.

**Repository** <https://github.com/GlobalDataverseCommunityConsortium/dataverse-language-packs/>

**Repository branch** develop

**Last remote commit** 4.19 in develop branch [d8d6404](#)  
Vyacheslav Tykhonov authored 37 minutes ago

**Repository containing Weblate translations** <http://example.com/git/dataverse/develop/>

**Filemask** `*/Bundle_*.properties`

**Monolingual base language file** `en_US/Bundle.properties`

<b>Number of strings</b>	28782
<b>Number of words</b>	206375
<b>Number of characters</b>	1308125
<b>Number of languages</b>	13
<b>Number of source strings</b>	2214
<b>Number of source words</b>	15875
<b>Number of source characters</b>	100625



# Weblate demo (5) - publish translations in new branch

GlobalDataVerseCommunityConsortium / dataverse-language-packs

Unwatch 14 Star 10 Fork 17

Code Issues 7 Pull requests 6 Actions Projects 0 Wiki Security Insights

Repository for language files associated with Dataverse

58 commits 14 branches 0 packages 0 releases 6 contributors

Branch: develop New pull request Create new file Upload files Find file Clone or download

Switch branches/tags

dataverse-v4.20

Branches Tags

Create branch: dataverse-v4.20 from 'develop'

fr_CA	Bundle properties added for compatibility with 4.19	6 days ago
fr_FR	"-" replaced with "-" in language folder names	10 months ago
hu_HU	Bundle properties added for compatibility with 4.14	9 months ago
id_ID	Indonesia Language second update	8 months ago
it_IT	Bundle properties added for compatibility with 4.14	9 months ago
pt_BR	Bundle properties added for compatibility with 4.19	13 days ago
se_SE	Bundle properties added for compatibility with 4.14	9 months ago
sl_SI	Bundle properties added for compatibility with 4.14	9 months ago
ua_UA	Bundle properties added for compatibility with 4.14	9 months ago
README.md	Bundle properties added for compatibility with 4.19	13 days ago

Suggestions for the translation pipeline:

- “develop” branch always contains all available languages with their latest versions of translations
- Weblate will use “develop” as a monolingual base to create translations for new Dataverse versions like 4.20
- new branch (dataverse-v4.20) should be created to synchronize all Weblate translations with github repository
- Weblate admin should check the consistency of all available translations and decide when language package is ready to push

# Questions concerning part 4 and the overall webinar



# Thank you for participating in this webinar

If you have questions, please contact us via email

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

[laura.huisintveld@dans.knaw.nl](mailto:laura.huisintveld@dans.knaw.nl)

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

