



**ADP - SOCIAL SCIENCE DATA ARCHIVES**

**Analyse data! Deposit study! Promote Science!**

# **USING DATAVERSE AS A (SELF-)REPOSITORY SYSTEM FOR DATA ARCHIVES**

**Maja Dolinar (ADP), Irena Vipavc Brvar (ADP)**

*Session: Building, managing and maintaining data and knowledge-sharing ecosystems for medium-sized research centres*

**SciDataCon-IDW 2018, 5 – 8 November 2018, Gaborone (Botswana)**



cessda eric



# Social Science Data Archives (ADP)



- established in 1997
- Slovenian national data repository for social sciences
  - 600 social science surveys with data in a data catalogue + 150 with metadata
    - approx. 1200 users registered in 2017 (90 % education, 10 % scientific/research purpose)
    - 168 survey data used for detailed secondary-analysis in 2017
- member of CESSDA ERIC
- obtained CoreTrustSeal in beginning of 2018
- involved in EU projects



 ADP Social Science Data Archives  
<http://www.adp.fdv.uni-lj.si/eng/>  
CTS Certification 2017-2019



# Objective and Purpose

In 2016 starting to look into the possibilities of establishing a

## **SELF-ARCHIVING TOOL**

for researchers and doctoral students to deposit smaller studies and studies of lesser scientific and methodological quality.

- through inclusion in various international projects, we identified Dataverse as the most appropriate tool to enable easy self-archiving to our users

cessda saw

***Strengthening And Widening  
The European Infrastructure  
For Social Science Data  
Archives***

*(2015 – 2017)*

cessda

***DataVerseEU 2018***

*Project partners: ADP (Slovenia), AUSSDA (Austria), DANS (the Netherlands), GESIS (Germany), SND (Sweden) and TARKI (Hungary).*

# Objective and Purpose

## Audience:

- **PhD students:** will be required to make data of their thesis publicly available
- **Researchers:** that produced either less quality data or data that are poorer in content and are likely to be used for limited amount of users
- **Regular users of the ADP:** who browse our online catalogue and use our research data.

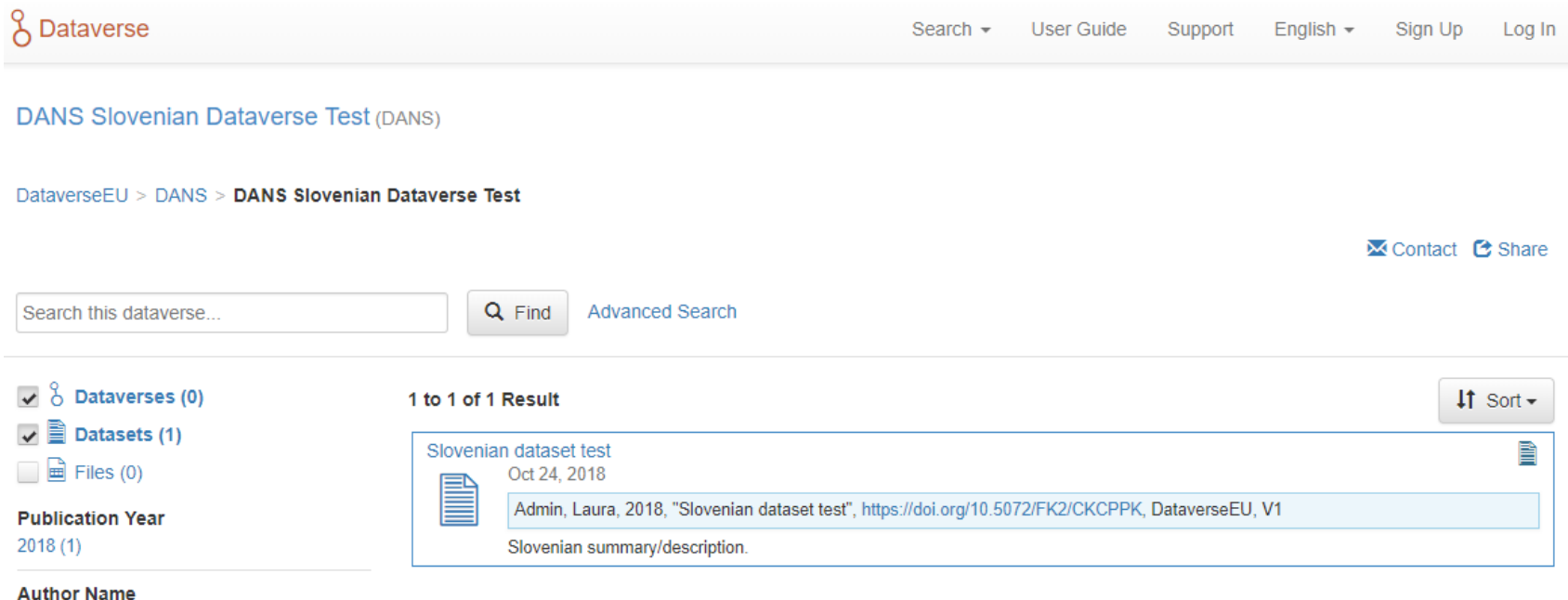
## Needed features:

- easy self-depositing system and quality check
- multilingualism & licenses
- possibility of adapting the metadata fields
- catalogue browsing for final users
- option of doing online analyses
- option of easy download

# About DataVerse

<https://dataverse.org/>

- an open source web application to share, preserve, cite, explore, and analyse research data. It facilitates making data available to others and allows you to replicate others' work more easily.
- consists of multiple virtual archives (dataverses)
- each dataverse contains datasets
- each dataset contains descriptive metadata and data files



The screenshot shows the DataVerse website interface. At the top, there is a navigation bar with the DataVerse logo, a search dropdown, and links for User Guide, Support, English, Sign Up, and Log In. Below the navigation bar, the breadcrumb trail reads "DataverseEU > DANS > DANS Slovenian Dataverse Test". There are "Contact" and "Share" buttons on the right. A search bar contains the text "Search this dataverse..." and a "Find" button. Below the search bar, there are filters for "Dataverses (0)", "Datasets (1)", and "Files (0)". The "Publication Year" filter shows "2018 (1)". The "Author Name" filter is empty. The search results section shows "1 to 1 of 1 Result" and a "Sort" dropdown. The result is for "Slovenian dataset test" dated "Oct 24, 2018". The citation text is "Admin, Laura, 2018, 'Slovenian dataset test', <https://doi.org/10.5072/FK2/CKCPPK>, DataverseEU, V1". Below the citation is a link to "Slovenian summary/description".

# Adjusting DataVerse within DataVerseEU 2018 project

- creating an easy installation ***DataVerseEU Docker module*** and distributing it via Github
  - <https://github.com/IQSS/dataverse-docker>
- creating a multilingual user-interface

The screenshot shows the DataVerse website interface. At the top left is the 'Dataverse' logo. The top navigation bar includes 'Search', 'User Guide', 'Support', 'English' (with a dropdown arrow), 'Sign Up', and 'Log In'. The main content area displays the title 'DANS Slovenian Dataverse Test (DANS)'. Below the title is a breadcrumb trail: 'DataverseEU > DANS > DANS Slovenian Dataverse Test > Slovenian dataset test'. A 'Metrics' section shows '0 Downloads'. A language dropdown menu is open, listing 'English', 'Slovenian' (highlighted in yellow), 'German', 'French', and 'Swedish'. At the bottom right of the page, there are 'Contact' and 'Share' links.

- connecting with a chosen Persistent Identifier (PID) provider
- development of an API for CESSDA Controlled Vocabularies, Topic Classification, European Language Social Science Thesaurus (ELSST)

# ISSUES and Workarounds

- default DataVerse installation very complex – solved with a docker installation that gives easy installation solution out-of-the-box
- ensuring multilingualism of user interface proved to be a real challenge:
  - DataVerse Harvard team did not think about translating the application in other languages than English
  - becomes an issue with new versions released! Currently in discussion with Harvard team on how to deal with translation management

Katalog ADP Dataverse Išči ▾ Vodič za uporabnike Podpora Slovenian ▾ Registriraj se Prijavi se

DANS Slovenian Dataverse Test (DANS)

DataverseEU > DANS > **DANS Slovenian Dataverse Test** ✉ Kontakt 📄 Deli

[Napredno iskanje](#)

---

**Katalogi (0)**
↑↓ Razvrsti ▾

**Raziskave (1)**

**Datoteke (0)**

**Publication Year**  
2018 (1)

**Author Name**

1 do 1 od 1 Result

Slovenian dataset test 📄

Oct 24, 2018

Slovenian summary/description.

- default metadata shema of DataVerse includes mostly text input – we needed to check the possibility of closed lists from which to select content of metadata fields

The image shows a screenshot of a DataVerse metadata form. On the left, there is a list of metadata fields with their values. On the right, there is a form with various input fields and dropdown menus. Two red arrows point from the text in the list to the dropdown menus in the form.

Field	Value
Citation Metadata	Production Place
Dataset Persistent ID	hdl:10695/UTI
Title	Survey about
Subtitle	KAJEN06
Alternative URL	http://www.adj
Author	Toš, Niko (CJI Research Cen
Contact	Use email t Toš, Niko (CJI Research Cen
Description	The purpose c Research field home, at work effect of smok health). (2004
Subject	Medicine, Hea
Keyword	estimation of h health smoking awareness ab passive smoki
Language	Slovene

Form fields on the right:


- Software:
- Name:
- Type:
- Grant Agency:
- Name:
- Abbreviation:
- Logo URL:
- Origin of Sources:
- Depositor:
- Deposit Date:
- Start:
- Start:



# ISSUES and Workarounds

- the metadata fields were adapted to follow the **CESSDA Metadata Portfolio**.
- Some of the default metadata fields needed to be adapted (for example name change or type of field change) or discarded and some additional fields were added to follow the CMM-Model.

At this stage, only the metadata terms fields were adapted (defined field names, field types and repeatability requirements), whereas the help texts that accompany each field, remain to be adapted in November and translated in different languages.

Files	Metadata	Terms	Versions
Citation Metadata ^			
Dataset Persistent ID	doi:10.5072/FK2/IOQ0WD		
Publication Date	2018-10-24		
Title	English dataset		
Author	Admin, Laura (DANS)		
Contact	 Use email button above to contact. Admin, Dataverse (Dataverse.org)		
Description	This is a test with an English dataset. (2018)		
Subject	Arts and Humanities		
Keyword	Keyword		
Notes	Some notes		
Depositor	Admin, Dataverse		
Deposit Date	2018-10-24		

# Current and future development

- looking into how to incorporate controlled vocabularies into the otherwise free text fields – through a developed ADP for CESSDA Vocabulary Manager (+ inclusion of ELSST keywords!)
- Local installation of DataVerse to be used as a self-archiving tool!
  - for now we were testing the application on [CESSDA Cloud](#)
    - incorporating ADP CVs to the existing adaptation of the metadata fields
  - creating user guidelines (for self-depositors and final users)
  - connecting with PID of choice – in addition to DOI also URN (National Library of Slovenia)?

[DataverseEU](#) > [DANS](#) > [DANS Slovenian Dataverse Test](#) > **Slovenian dataset test**

 Metrics

0 Downloads



**Slovenian dataset test** Version 1.0

Admin, Laura, 2018, "Slovenian dataset test", <https://doi.org/10.5072/FK2/CKCPPK>, DataverseEU, V1



# Usability of DataVerse for other small or medium-sized research centres

1. Useful for small and medium-size data repositories, who are in the process of establishing their own online data repository and are facing a low budget

- Dataverse as an open source application provides an opportunity to easily set up and quickly run their own online data repositories (easy installation via Docker!)
  - This would help them to promote the activities of their repository to possible financiers and ministries.

2. Application is interesting for repositories that are thinking of establishing a self-deposit service (or ongoing research repository) as Dataverse gives them an opportunity to easily set up such a service next to their traditional archiving solutions.

See more on: <https://github.com/IQSS/dataverse-docker> (all current translations and adaptation included and regularly updated!)



# Contact



Thank you for your attention! Questions?

University of Ljubljana  
Faculty of Social Sciences  
**Social Science Data Archive**  
Kardeljeva ploščad 5  
1000 Ljubljana  
Slovenia



[www.adp.fdv.uni-lj.si](http://www.adp.fdv.uni-lj.si)



[arhiv.podatkov@fdv.uni-lj.si](mailto:arhiv.podatkov@fdv.uni-lj.si)



[Arhiv.Druzboslovnih.Podatkov](https://www.facebook.com/Arhiv.Druzboslovnih.Podatkov)



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

Univerza v Ljubljani



cessda eric

