Implementing Standardised Metadata in a Self-Archiving System Based on DataVerse

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

Session: *With or Without You - Standardised Metadata in Survey Data Management*
ESRA 2019, 15-19 July 2019, Zagreb (Croatia)

Univerza *v Ljubljani*

cessda eric

CORE TRUST SEAL

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

CORE TRUST SEAL ✓

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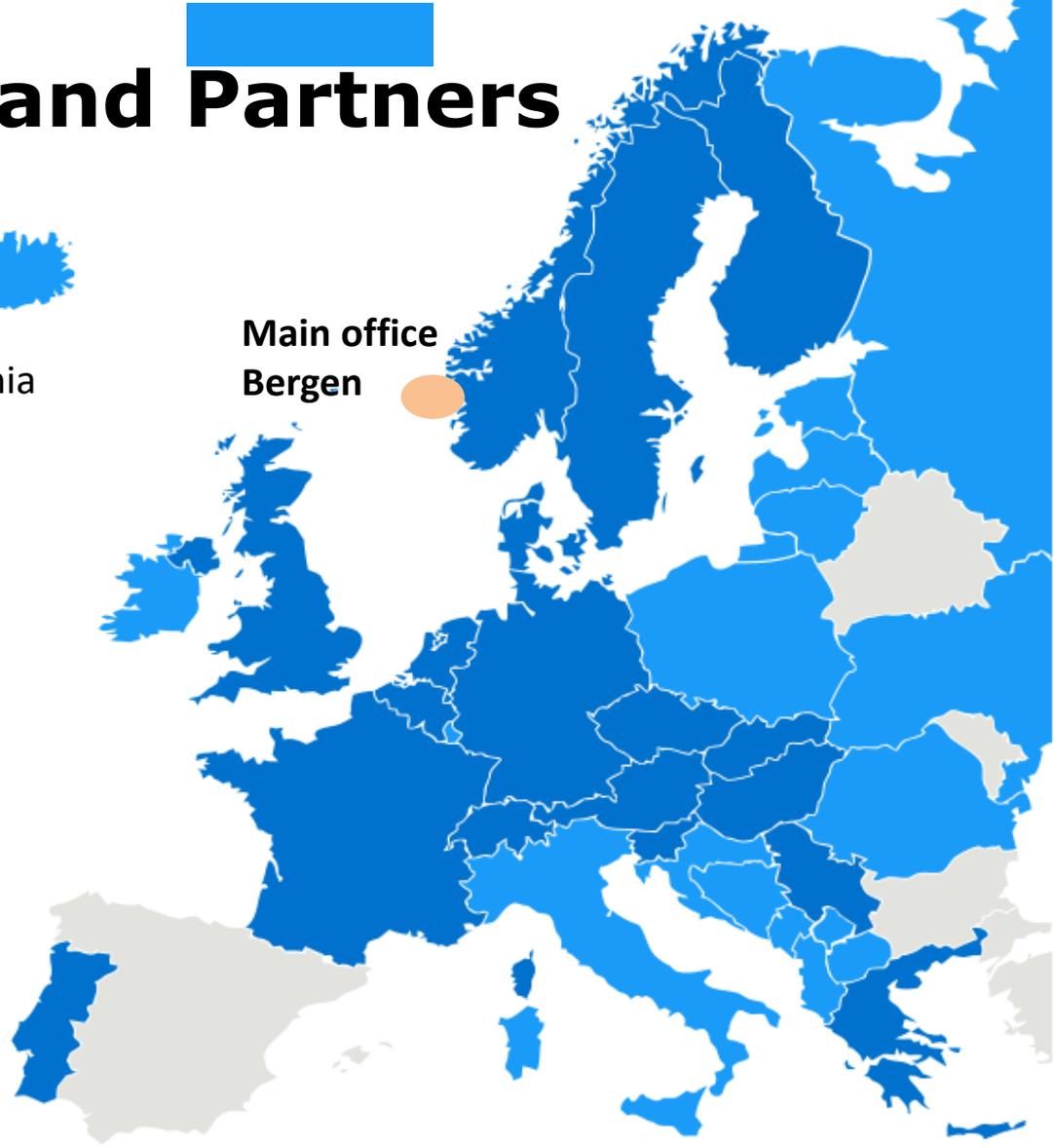http://www.adp.fdv.uni-lj.si/eng/
CTS Certification 2017-2019

CESSDA Members and Partners

- Austria
- Belgium
- Czech Republic
- Denmark
- France
- Finland
- Germany
- Greece
- Hungary
- Netherlands
- Norway
- Portugal
- Serbia
- Slovakia
- Slovenia
- Sweden
- Switzerland (Observer)
- UK

Members to be:
- Croatia
- North Macedonia

Main office Bergen

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.
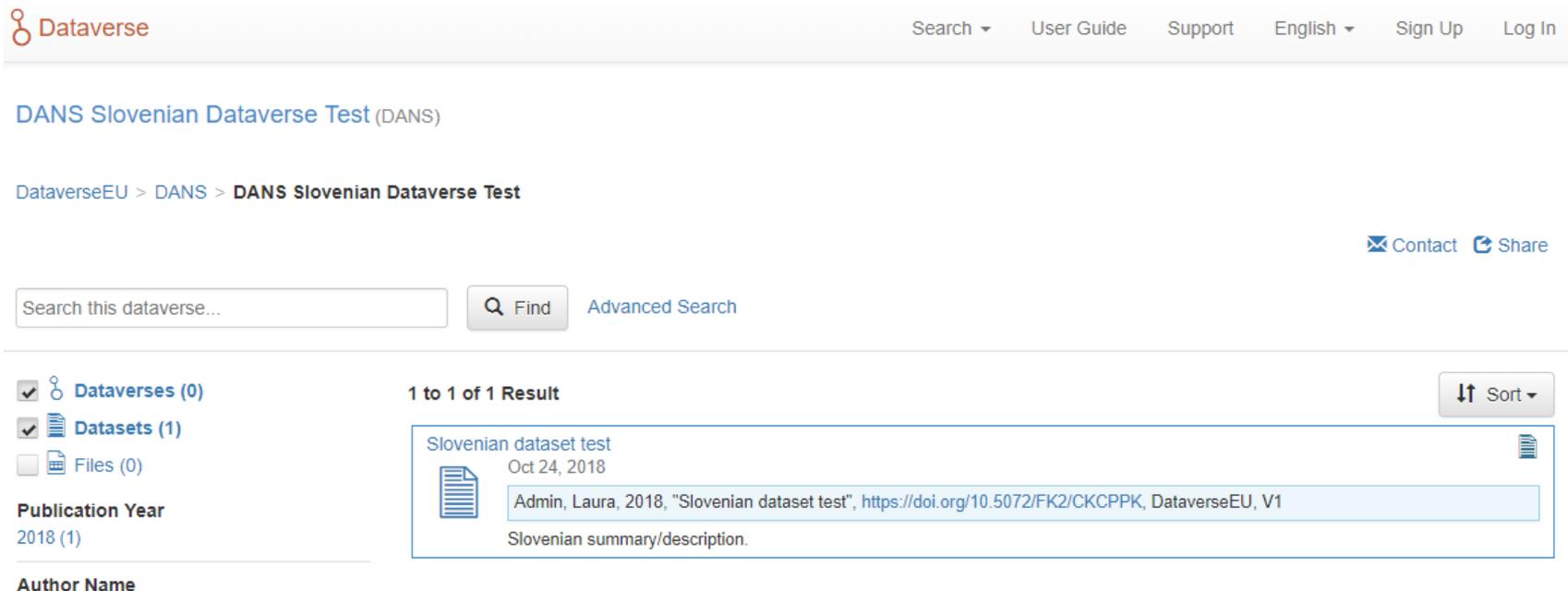
Needed features:

- 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

Adjusting DataVerse within DataVerseEU 2018 project

- creating an <u>easy installation</u> **DataverseEU Docker module** and distributing it via Github

  - **https://github.com/IQSS/dataverse-docker**

- creating a <u>multilingual user-interface</u>

- connecting with a <u>chosen Persistent Identifier</u> (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 <u>a real challenge</u>:

  - 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

- 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 need for FAIR metadata

FINDABILITY

F3. (meta)data are registered or indexed in a searchable resource.
F2. data are described with rich metadata.
F4. metadata specify the data identifier.
F1. (meta)data are assigned a globally unique and eternally persistent identifier.

ACCESSIBILITY

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

INTEROPERABILITY

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

REUSABILITY

R1.3. (meta)data meet domain-relevant community standards.
R1. (meta)data have a plurality of accurate and relevant attributes.
R1.2. (meta)data are associated with their provenance.
R1.1. (meta)data are released with a clear and accessible data usage license.

ISSUES and Workarounds

- The metadata fields were adapted to follow the **CESSDA Metadata Model**.

- 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.

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 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	❶ 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

CMM doi: 10.5281/zenodo.3238175

Current and future development

- Looking into how to incorporate controlled vocabularies into the otherwise free text fields (+ 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
  - and incorporating ADP CVs to the existing adaptation of the metadata fields.
  - We need to create user guidelines (for self-depositors and final users)
  - And connect with PID of choice – in addition to DOI also URN (National Library of Slovenia)?

DataverseEU > DANS > DANS Slovenian Dataverse Test

| Metrics | 0 Downloads |

cessda
VS Vocabulary Service

Slovenian dataset test Version 1.0

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

ADP

Analiziraj podatke! Deli raziskavo! Prispevaj k znanosti!

ADP katalog

ADP slovenski katalog

ADP English Dataverse

Search this dataverse...   🔍 Find   Advanced

Search

☑ ⅄ **Dataverses (2)**
☑ 📄 **Datasets (0)**
☐ 📄 Files (0)

Publication Year
2019 (2)

1 to 2 of 2 Results

⇅ Sort ▾

ADP English Dataverse
Jun 11, 2019

ADP slovenski katalog
Jun 11, 2019

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 financers 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**

✉ **arhiv.podatkov@fdv.uni-lj.si**

f **Arhiv.Druzboslovnih.Podatkov**

🐦 **@ArhivPodatkov**

Univerza *v Ljubljani*

cessda eric

CORE TRUST SEAL

seriss SYNERGIES FOR EUROPE'S RESEARCH INFRASTRUCTURES IN THE SOCIAL SCIENCES