



ADP - SOCIAL SCIENCE DATA ARCHIVES

Analyze data! Deposit study! Promote science!

Customizing Dataverse to workflows of ADP

Maja Dolinar and Gregor Žibert

Slovenian Social Science Data Archives (UL-ADP), 2021

*EDDI21, the 13th Annual European DDI User Conference, November 30 –
December 1, 2021*



Slovenian Social Science Data Archives (ADP – Arhiv družboslovnih podatkov)



- Founded in 1997
- Slovenian **national research data centre** for social sciences
- **Member of CESSDA ERIC since 2017**
- Status of a **trustworthy archive** (CoreTrustSeal since 2018) – certification currently in revision
- Involved in EU and national projects

Objective and Purpose

Situation: Dataverse identified as the best new option for distributing the curated catalog and to provide additional services of self-archiving.

Problem: default version of Dataverse lacks many essential features important for our workflow

Aim: present our experience in customising Dataverse to our needs, and identify issues and workarounds that repositories should be aware of

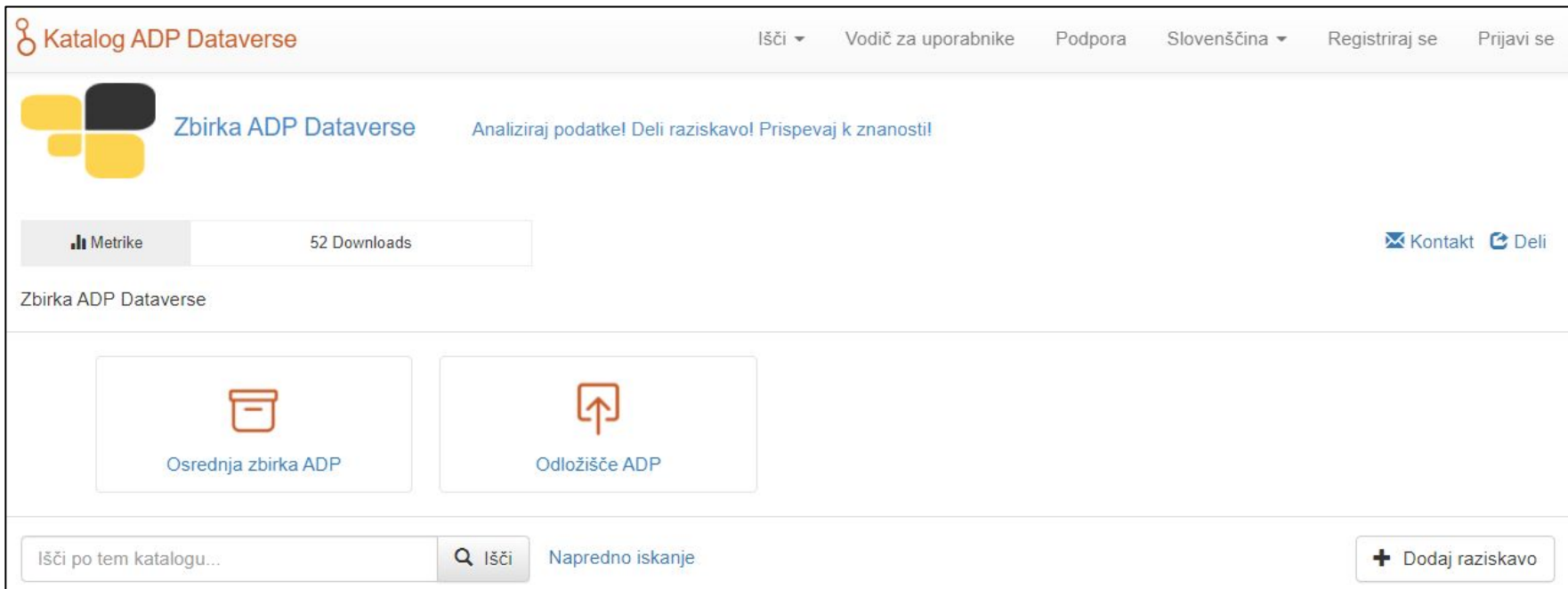
Needed features:

- multi-language user interface
- implementation of CESSDA Metadata model and ADP custom metadata fields
 - DDI Controlled Vocabularies
 - ELSST Keywords
- additional metadata fields on the level of files
- online deposit agreements and terms of use
- option of doing online analyses
- option of easy download/preview

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 'Katalog ADP Dataverse' interface. At the top, there are navigation links: 'Išči', 'Vodič za uporabnike', 'Podpora', 'Slovenščina', 'Registriraj se', and 'Prijava se'. Below the navigation is the main header for 'Zbirka ADP Dataverse' with the tagline 'Analiziraj podatke! Deli raziskavo! Prispevaj k znanosti!'. A 'Metrike' section shows '52 Downloads'. There are two main action buttons: 'Osrednja zbirka ADP' (represented by a folder icon) and 'Odložišče ADP' (represented by an upload icon). At the bottom, there is a search bar with the text 'Išči po tem katalogu...', a search button 'Išči', a link to 'Napredno iskanje', and a button to '+ Dodaj raziskavo'.

ISSUES and Workarounds

- online deposit agreements

IZJAVA O IZROČITVI GRADIV RAZISKAVE

Please confirm and/or complete the information needed below in order to continue.

Dajalec:

Ime in priimek

Maja Dolinar

Institucija

University of Ljubljana

Naslov

E-naslov

Maja.Dolinar@fdv.uni-lj.si

DOGOVORJENI POGOJI IZROČITVE

Dajalec izroča gradiva raziskave kot edini avtor oz. v dogovoru s soavtorji ali po pooblastilu nosilca pravice razpolaganja z gradivi iz raziskave (npr. naročnik) prevzemniku oz. njegovemu zakonitemu nasledniku v hrambo in za nadaljnje razširjanje uporabnikom, skladno s tem dogovorom. Dajalec obdrži vse pravice nad gradivi. Dajalec ima shranjeno kopijo izročениh gradiv. Dajalec pooblašča prevzemnika, da v njegovem imenu izročeno gradivo pregleda in ovrednoti njegov potencial za arhiviranje in objavo v sistemu aktivnega digitalnega skrbništva ter ga pripravi za dostop za namene druge rabe, skladno s poslanstvom ADP. Če potencial ne upraviči aktivnega skrbništva, se gradivo uvrsti v sistem brez aktivnega skrbništva (t. j. samo-arhiviranje). Prevzemnik potrjuje, da je prevzel izročena gradiva iz raziskave za namen hrambe in razširjanja uporabnikom pod pogoji, dogovorjenimi s to izjavo. Prevzemnik ima pravico prevzeto gradivo skladno s pogoji določenimi v tej izjavi objaviti v Zbirki ADP Dataverse in

Priloga:

[Splošni pogoji](#)
[Seznam možnosti glede avtorskih pravic](#)

Please, read and agree with above text to be able to click Accept

Strinjam se s pogoji izročitve

Dajalec:

Prevzemnik:



ISSUES and Workarounds

- multi-language user interface (using Weblate for translations - issue: translations need to be revised and updated for Dataverse version 5.2)
- the metadata fields were adapted to follow the **CESSDA Metadata Model** + incorporation of DDI CVs and ELSST keywords + ADP fields

Keyword ELSST ?	<p>(elsst) ELDERLY https://elsst.ukdataservice.ac.uk/</p> <p>(elsst) MEDIA LITERACY https://elsst.ukdataservice.ac.uk/</p> <p>(elsst) INFORMATION SOCIETY https://elsst.ukdataservice.ac.uk/</p> <p>(elsst) MASS MEDIA EXPOSURE https://elsst.ukdataservice.ac.uk/</p> <p>(elsst) MASS MEDIA USE https://elsst.ukdataservice.ac.uk/</p>
Keyword ADP ?	<p>(ADP) media https://www.adp.fdv.uni-lj.si/</p> <p>(ADP) Internet and classic media https://www.adp.fdv.uni-lj.si/</p> <p>(ADP) elderly people https://www.adp.fdv.uni-lj.si/</p> <p>(ADP) use of information and communication means https://www.adp.fdv.uni-lj.si/</p> <p>(ADP) mass media https://www.adp.fdv.uni-lj.si/</p>
CESSDA Topic Classification ?	(CESSDA Topic Classification) Information society urn:ddi:int.cessda.cv:TopicCl
CERIF Topic Classification ?	(CERIF) Press and communication sciences http://sicris.izum.si/
ADP Topic Classification ?	<p>(ADP Druga poglavja) Daily use of media https://www.adp.fdv.uni-lj.si/</p> <p>(ADP Druga poglavja) Social networks https://www.adp.fdv.uni-lj.si/</p> <p>(ADP Druga poglavja) Access to internet https://www.adp.fdv.uni-lj.si/</p> <p>(ADP Druga poglavja) Phone usage https://www.adp.fdv.uni-lj.si/</p> <p>(ADP Druga poglavja) Demography https://www.adp.fdv.uni-lj.si/</p>

ISSUES and Workarounds

- First idea: use CESSDA Vocabulary API service
 - testing mid-2020 - not working properly
- Waiting for new CESSDA Skosmos REST API
- In the meantime, we developed our own Middleware (using Docker) - not depending on external services
- Middleware obtains information from external vocabularies for frontend plugin of Dataverse, which shows the topics/keywords from the vocabularies to the users

The screenshot shows a web interface with three main sections for configuration:

- Unit of Analysis**:
 - Vocabulary: DDI Analysis Unit
 - VocabularyURL: urn:ddi:int.ddi.cv:AnalysisUnit:2.1
- Universe**:
 - Included: (empty text box)
- Time Method**:
 - Vocabulary: DDI Time Method
 - VocabularyURL: urn:ddi:int.ddi.cv:TimeMethod:1.2

A dropdown menu titled "Term" is open, showing a list of terms with checkboxes:

- Select...
- Individual
- Organization/Institution
- Family
- Family: Household family
- Household
- Housing unit
- Event/Process/Activity
- Geographic unit

ISSUES and Workarounds

- We use a dump of CESSDA and ELSST vocabularies in json format
- The script reads dump files of CESSDA and ELSST vocabularies and provides autocomplete functionality (Python)
- Resolved some issues with Bootstrap layout
- Problem with big vocabularies (e.g. ELSST) - if an user types a symbol in a field - complete list of terms for small vocabularies, for big vocabularies: developed functionality - scrollable list of items.
- CESSDA Skosmos REST API service launched - Dataverse was updated to use new service for ELSST keywords (late-2020)
- no API to update database and Solr schema (in development) - database and Solr have to be updated manually
- procedure to create/update metadata blocks/fields as we require

ISSUES and Workarounds

- Licenses and terms of use customizations for files
 - For each individual uploaded file, we can select one of the 3 different CC licences, or restricted access
 - Option to define access conditions
 - Depending on licence and access conditions, individual file is either open or restricted
 - Restricted > request access

License

For selected file(s) set a license to

CC BY - Creative Commons Attribution 4.0 ▼

CC0 - Creative Commons Zero 1.0

CC BY - Creative Commons Attribution 4.0

CC BY-NC - Creative Commons Attribution - NonCommercial License 4.0

RES0 - Restricted Access

Access conditions

Access to file subject to additional consent under following conditions

PUF - Public Use File ▼

Select...

PUF - Public Use File

SUF - Scientific Use File

ScUF - Secure Use File

ISSUES and Workarounds

- Additional metadata fields on the level of files (Title of Material, Year, Author(s)):
 - adding new metadata fields on file level required more work than on the dataset level
 - additional fields in the database
 - new code in Java and xhtml files
 - create corresponding Solr indexes

The screenshot shows a web interface for editing file metadata. The form is titled "1 File" and contains the following fields and options:

- File Name:** mpstr18_vp1_sl_v1_r1.pdf
- Title of Material:** MPSTR18 - Medijske navade starejših (6)
- Author(s):** Mateja Rek (with a "+" button to add more authors)
- Year:** 2021
- File Path:** (empty field with a help icon)
- Description:** Add file description... (text area)
- License:** For selected file(s) set a license to: CC0 - Creative Commons Zero 1.0 (dropdown menu)
- Notes:** (empty text area with a help icon)

At the bottom, there are two buttons: "CC0" and "Documentation".

Evaluation of available Plugins - curation









- **Data curation tool:** allows data owners and curators to create and edit variable-level metadata for any tabular file in a dataset
 - replaces Nesstar Publisher for editing variable level information (a tabular datafile needs to be uploaded for it to work)
 - allows to view summary statistics about data, add variable information (e.g. 'Interviewer Instructions' or 'Notes'), create variable groups, and indicate weighting variables
 - once edits have been completed and saved back to Dataverse, these changes can then be downloaded as an XML file or exported to a codebook.

test1 English ▾

Drones-Canada.tab

Admin, Dataverse, 2021, "test1", https://doi.org/10.17898/ADP_DV_LJ2P5Y ⓘ, Zbirka ADP Dataverse, DRAFT VERSION, UNF:6:5q/DbE5M7z6XI/HdLK9RvA== [fileUNF]

< Hide Groups Download Save to Dataverse

Add Group +	Search 			Items per page 25 ▾	26 - 50 of 85	<	>
All Variables	<input type="checkbox"/>	ID	Name	Label	Weight	View	
	<input type="checkbox"/>	v1081	Q1_General_Awareness_5	Q1_General_Awareness_5			
	<input type="checkbox"/>	v1035	Q1_General_Awareness_6	Q1_General_Awareness_6			
	<input type="checkbox"/>	v1016	Q1_General_Awareness_7	Q1_General_Awareness_7			

Evaluation of available Plugins - data sharing

- **Data Explorer:** lists the variables in a tabular data file allowing searching, charting and cross tabulation analysis
 - has some similar functionalities as Nesstar online analysis tool

Drones-Canada.tab
Admin, Dataverse, 2021, "test1", https://doi.org/10.17898/ADP_DV_LJ2P5Y, Zbirka ADP Dataverse, DRAFT VERSION, UNF:6:5q/DbeSM7z6XI/HdLK9RvA== [fileUNF]

Q **85** Results

Chart View

ID	Name	Label	Categories	Valid Cases
1015	PanelistIdQuestion	PanelistIdQuestion		3045
1025	gender	gender	2	3045
1057	Q5_Federal_Vote_2011	Q5_Federal_Vote_2011	2	3045
1028	Q6_Federal_PartyChoice_2011	Q6_Federal_PartyChoice_2011	6	2565
1012	Age_Rollup_broad	Age_Rollup_broad	3	3045
1019	Income	Income	4	3045

Variable HH_Income: HH_Income

Category	Value
1	342
2	269
3	487
4	560
5	434
6	230
7	296
8	427

Records Per Page 10

- Evaluated also **TwoRavens:** a web application for tabular data exploration and statistical analysis with Zelig
 - evaluated as too complicated for our users to use (able to run statistical models, view summary statistics, create and download subsets of variable vectors and more)

Evaluation of available Plugins - data sharing

PDF Preview

Filename: [Political_Social_Correlates_Covid-19_Mortality.pdf](#)
 In *test1* (version 4.2), by Admin, Dataverse

[Download File](#) [Close Preview](#)

[Previous](#) [Next](#) Page: 1 / 53

Political and Social Correlates of Covid-19 Mortality








Constantin Manuel Bosancianu* Kim Yi Dionne[†] Hanno Hilbig*[‡]
 Macartan Humphreys*[§] Sampada KC* Nils Lieber*[¶] Alexandra Scacco*


11 June, 2020. Most recent version [here](#)

Abstract

Do political and social features of states help explain the evolving distribution of reported Covid-19 deaths? We identify national-level political and social characteristics that past research suggests may help explain variation in a society's ability to respond to adverse shocks. We highlight four sets of arguments—focusing on (1) state capacity, (2) political institutions, (3) political priorities, and (4) social structures—and report on their evolving association with cumulative Covid-19 deaths. After accounting for a simple set of Lasso-chosen controls, we find that measures of government effectiveness, international and institutional trust, bureaucratic corruption and ethnic fragmentation are strongly

- **File Previewers:** A set of tools that display the content of files - including audio, html, Hypothes.is annotations, images, PDF, text, video, tabular data, and spreadsheets - allowing them to be viewed without downloading.

<input type="checkbox"/>		Political_Social_Correlates_Covid-19_Mortality.pdf Adobe PDF - 542.3 KB - Sep 3, 2021 - 1 Download MD5: 4015f59153ff0447f937155ed0e93a38 License: CC0 - Creative Commons Zero 1.0	  
<input type="checkbox"/>		Thinking-of-getting-a-cat.png PNG Image - 1.2 MB - Sep 3, 2021 - 4 Downloads MD5: e08f8b00704d429c90e151bad272dc28 License: CC0 - Creative Commons Zero 1.0	  

Explore Options 
[View Image](#)



Current and future development

- transfer of curated catalog to the Dataverse
- self-archiving service enabled for our users
- integration with our existing data curation tools and services (testing of pyDataverse)
- integration with our workflows

Future developments:

- full multilinguality - dependent on outside future developments and support
- OAI-PMH for CESSDA Data Catalog
- updated XML DDI 2.5 export
- optimization
- evaluation and testing of additional plugins (e.g. Whole Tale, Amnesia, Avgidea Data Search etc.)

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



Arhiv.Druzboslovnih.Podatkov



@ArhivPodatkov

