South-Eastern European Data Services

D8 – Report on data collection and preparation: Serbia

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

SWISS NATIONAL SCIENCE FOUNDATION

Deliverable Lead:	UL (ADP), IES
Related Work package:	WP2

Author(s):	Aleksandra Bradic Martinovic (IES)
	Larisa Malic (IES)

Dissemination level:	Public (PU)
Submission date:	30th March, 2017
Project Acronym:	SEEDS
Website:	http://www.seedsproject.ch
Call:	Scientific cooperation between Eastern Europe and Switzerland (SCOPES 2013-2016)
Start date of project:	1[st]May 2015
Duration:	24 months

Version History

Version	Date	Changes	Modified by
1.0	29th March 2017	Released Version	IES

Acknowledgments

Name	Short Name	Country
Faculty of Humanities and Social Sciences University of Zagreb	FFZG	Croatia
Centre for Monitoring and Research, Podgorica	CeMI	Montenegro
Centre for Political Courage, Pristina	CPC	Kosovo
Institute for Democracy and Mediation, Tirana	IDM	Albania
Institute of Economic Sciences, Belgrade	IES	Serbia
Saints Cyril and Methodius University, Institute for Sociological, Political and Juridical Research, Skopje	ISPJR	Macedonia
Swiss Foundation for Research in Social Sciences, Lausanne	FORS	Switzerland
University of Ljubljana, Social Science Data Archive, Ljubljana	UL	Slovenia

SEEDS: D8 – Report on data collection and preparation: Serbia

Table of Contents

1. Introduction and history

The first steps towards the development of the Serbian Data Center in Social Sciences (SER.DAC) were within FP7: SERSCIDA project (2012-2014). During that period the Serbian project team made publicly available the first dataset *Public Opinion: Election Study 2000* on the Nesstar platform (not available anymore) as an example and prototype of a new data service.

The most valuable experience SER.DAC stuff gained through the RRPP Data Rescue project, a SEEDS sub-project. During the project Serbian team curated 23 studies with 18 quantitative data files and over 500 qualitative data files on the SEEDSbase platform (https://seedsdata.unil.ch/).

Finally, the SER.DAC team installed its own platform for data curation and dissemination - Dataverse Serbia (http://dataverse-serbia.ien.bg.ac.rs/). The most valuable datasets from eight RRPP research studies and two additional studies with data sets has been archived in Dataverse.

All processes described in this report are in line with the Policy and Procedure Document adopted by the SER.DAC management and OAIS methodology, described in details in SEEDS D6 deliverable.

2. Data collection process

For first contacts and negotiations with the data depositors SER.DAC has used the email service from the Institute of Economic Sciences (ser.dac@ien.bg.ac.rs). First, mails were sent to potential depositors in January 2016 with descriptions about the project and our mission. From March 2016 SER.DAC team obtained sufficient control of the information needed to ensure long term preservation of the data collections. We used email channel for sending and receiving study description documents. Second, we sent to Principal investigators (PIs) an empty study description document. The most PIs replied positively, but in cases where we did not have a reply, our stuff filled in the form. Regarding the data files, depositors had a choice to use email messages, *OwnCloud* platform or personal delivery (USB stick, CD or similar devices) for the data transfer, depending of the sensitivity of the data.

SER.DAC was faced with several problems during the collection process. Probably the main problem was the lack of communication with researchers/depositors, because they were not familiar enough about our activities. We put a lot of effort to give answers to crucial questions such as data anonymization, data access, putting data under embargo etc. In one case we could not conduct anonymization with a satisfactory quality and we decided to put this data set under restriction. Above all, we had a huge problem with Dataverse installation and we resolved it with external support of DANS (Netherland) within the CESSDA SaW pilot project.

3. Conditions of use

Datasets and documentation stored in our data repository will be available free of charge. Some limitations and controlled access may be applied, depending on the sensitivity of the data and specific conditions determined by the depositor.

There are two levels of access available:

- Research data available without any restrictions (open access)
- Research data available only with special permissions given by the Archive staff or depositor

4. Collected datasets

During the SEEDS project we collected and published 10 datasets:

Dataset No. 1
Title: Images of Educational Changes in Serbia: Reflecting the Past, Envisioning the Future
SRB: Predstave o obrazovnim promenama u Srbiji: refleksije o prošlosti, vizije budućnosti
Principal investigator: Nikoleta Gutvajn
Institution: Institute for Educational Research
Dataset: Survey with primary school principals, primary school teachers of 1st grade, primary school 5th-grade homeroom teachers, primary school teachers who were members of the school development planning team, primary school counsellors, education experts and education authorities (194 variable and 1976 observations)
Access rights: Open Access
Private URL: dataverse-serbia.ien.bg.ac.rs/privateurl.xhtml?token=8ddd3e16-8fdd-42c0-a549-8ff41d790e78

Dataset No. 2
Title: Public Participation in Environmental Decision Making: the Case of Bor and Pancevo
SRB: Javno učešće u donošenju odluka o zaštiti okoline: slučaj Bora i Pančeva - Izazovi participativnog pristupa uprave za zaštitu okoline
Principal investigator: Mina Petrović
Institution: University of Belgrade, Faculty of Philosophy
Dataset: Survey with citizens of Bor and Pančevo, with the representative sample in Bor and in Pancevo (169 variables and 799 observations)
Access rights: Open Access
Private URL: dataverse-serbia.ien.bg.ac.rs/privateurl.xhtml?token=9327fb9e-f014-4bd8-9450-e0e7969cf084

Dataset No. 3
Title: Media Discourse of Poverty and Social Exclusion
SRB: Medijski diskurs o siromaštvu i socijalnoj isključenosti
Principal investigator: Dubravka Valić Nedeljković
Institution: University of Novi Sad, Faculty of Philosophy
Dataset: Focus group was conducted with 8 journalists and editors from Novi Sad (of 10 invited). An interview guide was designed for purpose of this study. Interview was semi-structured.
Access rights: Restricted

Dataset No. 4
Title: Youth - Actor of Social Change
SRB: Mladi – naša sadašnjost: Istraživanje socijalnih biografija mladih u Srbiji
Principal investigator: Smiljka Tomanović
Institution: University of Belgrade, Faculty of Philosophy, Institute of Sociology and Social Research
Dataset: Survey with young people ages 19 to 35, residents of Republic of Serbia (442 variables and 1627 observations)
Access rights: Open Access
Private URL: dataverse-serbia.ien.bg.ac.rs/privateurl.xhtml?token=ecdd3d05-5143-4c90-83c6-f932f5052716

Dataset No. 5
Title: Gender Perspectives in Family Socialisation
SRB: Rodne perspektive u porodičnoj socijalizaciji
Principal investigator: Sanja Milutinović Bojanić
Institution: Center for Ethics, Law and Applied Philosophy - CELAP
Dataset: Survey with households in Serbia and Montenegro that met following criteria - Parents (father and mother) with children of both genders (boy and girl), aged between 2-12 years old, married and belonging to the middle class (34 variables, 102 observations)
Access rights: Open Access
Private URL: dataverse-serbia.ien.bg.ac.rs/privateurl.xhtml?token=8e75634c-4ae2-4bd4-8dc6-

Dataset No. 6

Title: Resistance to Socio-Economic Changes in Western Balkan Societies. Testing Two Theories of Social Development

SRB: Otpornost na društveno-ekonomske promene u društvima Zapadnog Balkana. Testiranje dve teorije socijalnog razvoja

Principal investigator: Predrag Cvetičanin

Institution: University of Niš, Faculty of Arts

Dataset: Survey with the complete population of Serbia, Kosovo, Bosnia and Herzegovina and Macedonia, age 18 years or older, those which at the time of research (beginning of 2013) were permanent residents at these territories

Access rights: Open access

Private URL:

Dataset No. 7

Title: Gender that Matters: Poverty and Social Inclusion – Social Protection Status in Rural Kosovo and Serbia

SRB: Važnost roda: Siromaštvo i socijalna uključenost – Stanje socijalne zaštite u ruralnim oblastima Kosova* i Srbije

Principal investigator: Sanja Bojanić

Institution: Center for Ethics, Law and Applied Philosophy - CELAP

Dataset: Standardised interviews - Residents of rural areas of Serbia and Kosovo* in need for social

protection services, low-income households and also qualitative interviews - 60 respondents from rural area and 20 interviews with representatives of organizations in Serbia and in Kosovo 40 people and also 20 organizations

Access rights: Open access

Private URL: dataverse-serbia.ien.bg.ac.rs/privateurl.xhtml?token=f3002443-64d6-41df-8297-e3467960f6a8

Dataset No. 8

Title: Performance Audit and Policy Evaluation: On the Same or Parallel Tracks?

SRB: Revizija poslovanja i evaluacija politika: Da li smo na istim ili parelenim kolosecima?

Principal investigator: Milena Lazarević

Institution: European Policy Center

Dataset: Interviews with staff from state audit institutions that have engaged in performance audit; staff from the institutions at the centre of government, in charge of horizontal overview of the policy making process (Public Policy Secretariat in Serbia); staff in policy analysis/evaluation departments; staff in the institutions which were subject to audit; international staff that has acted as expert support through projects to both SAIs and policy evaluation departments; external evaluators.

Access rights: Restricted

Private URL: dataverse-serbia.ien.bg.ac.rs/privateurl.xhtml?token=8456a4a2-1548-4b52-a3ea-fb8af47b32b4

Dataset No. 9

Title: Pre-election Study - Presidential Elections Serbia 2004

SRB: Predizborno istraživanje javnog mnjenja – predsednički izbori u Srbiji 2004

Principal investigator: Institut društvenih nauka

Institution: Institute of Economic Sciences

Dataset:

Access rights: Open Access

Remark: The version 1.0 of this dataset had been published at ZACAT, we added both version 1.0 and 2.0

Private URL: dataverse-serbia.ien.bg.ac.rs/privateurl.xhtml?token=7da937c8-6cf8-40dc-9b05-65c001744c9e

Dataset No. 10

Title: Attitudes of the Unemployed Youth toward Work Conditions

SRB: Očekivanja nezaposlene omladine o uslovima rada

Principal investigator: Aleksandar Zdravković

Institution: Institute of Economic Sciences

Dataset: Survey with unemployed young people (16 to 30 years old) in Serbia

Access rights: Open Access

Private URL: dataverse-serbia.ien.bg.ac.rs/privateurl.xhtml?token=f67cd8ef-331c-4124-8dfe-025b07030176

Link to Dataverse Serbia catalogue: http://dataverse-serbia.ien.bg.ac.rs/

Link to SEEDSbase catalogue: https://seedsdata.unil.ch/

5. Data management workflow

The process of receiving datasets from researchers, preparing them for archiving and publishing is one of the crucial parts of a data archiving system. It has to be ensured that all necessary steps in data management are accomplished and that is why a data management workflow was developed. This was done in cooperation with our experienced partners - ADP and FORS. Starting point was a proposed detailed workflow produced by ADP. Sources from Ljubljana workshop were also used[1] together with materials prepared for RRPP Data Rescue project[2].

There were several versions of the workflow, developed through testing the workflow with real examples, and refined to cover the use of specific tools. The final version of the workflow to be used for managing collected datasets is a list of tasks organised into the following categories:

- communication
- general
- materials
- data
- study description
- Nesstar (optional)
- before publishing
- publishing

Communication: The tasks in this category are related to communication with a data provider. This include communication about receiving files - inviting researchers to submit their datasets to the archive, discussing how to prepare data (anonymisation and cleaning), ways of transfer, and similar. Another important issue to discuss is about access condition and agreements about deposit contracts.

General: Steps related to storing original documentation and datasets are covered in this category. Also, here we want to make sure that the study is properly identified, e.g. that the basic information for the study is determined (title, year of data collection, author). We may add study ID and a study acronym manually, or this can be done through the software systems. In this category, we also have a separate task for defining data access policies and this is linked to communication with a data provider where the access rights are discussed.

---

[1] D7 - Report on Workshop II_Ljubljana. http://seedsproject.ch/wp-content/uploads/2015/06/Report-on-Workshop-II_Ljubljana.pdf

[2] RRPP Data Rescue project. http://seedsproject.ch/?page_id=618

Materials: Here we deal with preparation of materials (documentation) for ingest. All materials has to be reviewed and it has to be decided which are going to be archived. Documents have to be stored in formats appropriate for AIP (see D6).

Data: Data ingest is in a separate category because the procedures for data ingest are different than documentation ingest procedures and the work can be done by different staff members. Here we process the data and check if anonymisation is done properly and if the submitted files are clean and properly labelled. Conversion from originally submitted file formats (usually SPSS, Stata and Excel) has to be done to the formats acceptable for AIP (reference to D6).

Study description. Here we put information only about the study description, more precisely, information about the creation of the study.

Nesstar (optional)

Before publishing: The tasks in this category are related to access rights for files in Dataverse (this is related to data the access policy), and to resending the study description document to authors for authorization.

Publishing: In this category, we have to put information about publishing datasets and newsletters about the new dataset on the website.

6. Workflow implementation in Redmine

SER.DAC used the Redmine platform (https://pm.serscida.eu/redmine/) for managing tasks, issues, and requests submitted by a community of users. In the near future we have plans to test Request tracker for this purpose. We put 10 projects in Redmine by now. There is one main frame: SEEDS-RS, where we put our 10 projects. Each project has been formed as a *new subproject*. Within each issue we had tracking lists with the following trackers:

- Communication
- General
- Materials
- Data
- Study description
- Nesstar
- Before publishing
- Publishing

Through these trackers we used entities to describe actions in more details, through the answers on questions such as title of the research, authors of the research, but also we described the process of anonymization, we defined the data access policy, and more.

During the project implementation we decided not to use Nesstar, because Nesstar is not an archiving

tool in the broader sense.

7. Repository building with Dataverse

Dataverse is an open source web application to share, preserve, cite, explore, and analyse research data[3]. It was selected during the evaluation process reported in D9 for the SEEDS project[4] as an archiving and dissemination tool for the future data archive.

According to the SEEDS project's plan, the Serbian team had to archive 10 datasets. We used Dataverse for that task (http://dataverse-serbia.ien.bg.ac.rs/dataverse.xhtml). We had external support from DANS (Netherland) within the CESSDA Saw pilot project. The Serbian team had an administrator account for this platform and we did have full access to all phases in our curation process. We filled in most of the required fields in Dataverse and we uploaded files with data (in spss format for quantitative data and pdf format for qualitative data). We did not include all received documentation in pdf, except questionnaires and focus groups.

We made two dataverses:

- Institut ekonomskih nauka Dataverse and
- RRPP Serbia Dataverse

---

[3] http://dataverse.org/about

[4] D9 - Report on tool evaluation and selection - http://seedsproject.ch/wp-content/uploads/2015/06/D9_FINAL.pdf

The first dataverse has the name of the institution and contains the project *Očekivanja nezaposlene omladine o uslovima rada (Youth Expectations in Job Search in Serbia).* The other one (RRPP Serbia Dataverse) has eight datasets with metadata and files. The tenth project, *Predizborno istraživanje javnog mnjenja – predsednički izbori u Srbiji 2004 (Pre-Election Study - Presidential Elections Serbia 2004)* is uploaded on Root Dataverse.

Link to D6: http://seedsproject.ch/wp-content/uploads/2015/06/SerbiaD6-Reportonintegrationoftechnicalsystemfinal.pdf

In conclusion, the new data service SER.DAC will be able to integrate and manage new datasets, safely store and protect data, as well as disseminate data and documentation to users. Their technical systems will function according to international norms and best practices, even if some parts of the archiving workflow will need to be handled manually. Future work should expand the technical development to accommodate for a greater volume and variety of data, automate more everyday practices, and enhance communication potential and exchange with data producers and users. This work will continue for many years to come, and will build on experience, further training, and funding. Like any others, the new data service will have to adapt technically to the ever-changing research and policy environments.