



## D7 – Report on Workshop II



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra



SWISS NATIONAL SCIENCE FOUNDATION

*Deliverable Lead:* UL-ADP, CPC

*Related Work package:* WP3

*Author:* Irena Vipavc Brvar (UL-ADP)

*Dissemination level:* Public (PU)

*Submission date:* 29<sup>th</sup> March 2016

*Project Acronym:* SEEDS

*Website:* <http://www.seedsproject.ch>

*Call:* Scientific cooperation between Eastern Europe and Switzerland (SCOPES 2013-2016)

*Start date of project:* 1<sup>st</sup> May 2015

*Duration:* 24 months

## Version History

Version	Date	Changes	Modified by
1.0	March 15, 2016	Internal version	UL-ADP
2.0	March 22, 2016	Revised version	FORS

## Acknowledgments

This report has been developed within the “South-Eastern European Data Services” (SEEDS) ([www.seedsproject.ch](http://www.seedsproject.ch)) project. The participant organisations of the SEEDS project are:

Name	Short Name	Country
Centre for Monitoring and Research, Podgorica	CeMI	Montenegro
Centre for Political Courage, Pristina	CPC	Kosovo
Institute for Democracy and Mediation, Tirana	IDM	Albania
Faculty of Humanities and Social Sciences University of Zagreb	FFZG	Croatia
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 Archives, Ljubljana	UL-ADP	Slovenia

## Table of Contents

1. Report on Workshop II, Ljubljana .....	4
2. List of Participants .....	10
Annex I: Programme .....	11

# 1. Report on Workshop II, Ljubljana

The second training workshop in the SEEDS project took place between 9th and 11th February, 2016 in Ljubljana, Slovenia.

The workshop was originally scheduled to take place between months 12 and 14, but was moved to month 10 due to the availability of the computer labs with specific software needed for the training. The availability of statistical software was also the reason why the workshop was moved from Zagreb (as listed in the project plan) to Ljubljana.

In the SEEDS project description we said that:

*“The second workshop will cover data solicitation and curation of both quantitative and qualitative data. Participants will be expected to arrive with real datasets that will be processed, cleaned, and documented through hands-on training using tools of the trade. There will be presentations and discussion of best practices and standards regarding data preparation and documentation, including the international metadata standard DDI (Data Documentation Initiative).”*

When putting together the programme for the workshop we had this vision in mind and tried to do our best to fulfil it. We focused on the archival process from the beginning to the end.

Datasets used at the workshop were from the Regional Research Promotion Programme Western Balkans (RRPP). The RRPP programme is interested in long-term preservation of the survey data and other materials produced in RRPP supported projects and the cooperation with FORS and other partners in the SEEDS project was established to achieve this goal.

The workshop included (see detailed content in Annex I):

- Presentation about RRPP research programme (source of real datasets used for processing);
- Talk about communication with researchers;
- Pre-SIP and SIPs: best practice and ingest with hands-on session;
- AIP and metadata management with hands-on session on AIP;
- AIP and data management (curating qualitative and quantitative data);
- DIP: best practice in data dissemination (qualitative and quantitative data);
- Planning next steps in SEEDS project.

Janez Štebe, head of UL-ADP opened the workshop (9th of February). A short presentation of the content of the workshop followed by Brian Kleiner, Head of Data and Research Information Service at FORS, and SEEDS principle investigator.

There, an introduction of the invited speakers was made. The first speaker was **Magdalena Solska**, programme advisor of the Regional Research Promotion Programme Western Balkans (RRPP), Switzerland.

*“The RRPP is aimed at fostering and promoting social science research in the Western Balkans (Albania, Bosnia and Herzegovina, Kosovo, Macedonia, Montenegro and Serbia). Social science research aids in the understanding of the specific needs for reform and in*

*identifying the long-term implications of potential policy choices. Researchers receive support through research grants, methodological and thematic trainings, as well as opportunities for regional and international exchange and networking. The RRPP is coordinated and operated by the Interfaculty Institute for Central and Eastern Europe (IICEE) at the University of Fribourg (Switzerland). The programme is fully funded by the Swiss Agency for Development and Cooperation (SDC), Federal Department of Foreign Affairs.” <http://www.rpp-westernbalkans.net/>*

M. Solska presented the project, saying that under the umbrella topic “Political, Social and Economic Change in the Western Balkans” 71 research projects (47 regional and 24 national) were supported. Most of them are already finished, some will finish in 2016. Research outcomes range from monographs, edited volumes, and scientific papers to policy briefs, documentaries, and websites and are available on the RRPP website. Leading research institutions are from Albania (11), Macedonia (18), Serbia (23), BiH (15), Kosovo (3) and Montenegro (1). The RRPP project has three main pillars of support, namely research funding and promotion, capacity building and networking and policy dialogues.

Due to the close end of the project, the RRPP team contacted FORS in order to discuss long-term preservation issues and the availability of the gathered material. In the last months several e-mails were exchanged between FORS, RRPP, and researchers involved in the RRPP projects. The RRPP team contacted project leaders by explaining the need of depositing data and materials to FORS and SEEDS partners’ involvement in it. In this initial stage of contacting, several researchers expressed concern in depositing data and related materials; they were also saying that there are no transcriptions existing from interviews, and focus groups; they might have used secondary sources on which they do not have authorship rights (like from NSI) or are concerned about the anonymisation of the data (e.g. research on elites).

The second invited speaker was **Arja Kuula-Lummi**, from the Finnish Social Science Data Archive (FSD). Her talk was about the qualitative data service offered by FSD and communication practices with researchers. Qualitative surveys are common in Finland and in some fields even more supported than quantitative. In the last 12 years they archived around 5 to 20 qualitative surveys per year, with approximately 100 contacts made per year. Among the most common reasons for not archiving data in the archive are unclear ownership, IT problems, inadvertent misuses of data, and concerns about ethics and data protection issues. Interviews are in most cases more personal than when the respondent is answering closed-ended questions. It is important that researchers inform participants about the use of collected materials and if needed negotiate access conditions.

Hence, several guides were prepared by FSD<sup>1, 2</sup>. There is specified that it is important to respect participants’ consent and exclude some interviews from the deposited collection if consent is not granted. She also reported about experiences where consent for archiving research data was given long after the interview was carried out. FSD contacted research participants to get consent to the archiving for four major studies. 169 respondents were contacted, and only 4 did not consent to the archiving of the data. Researchers are more and more aware of the requirements of research funders and journals regarding the availability of

---

<sup>1</sup> FSD – Informing Research Participants - <http://www.fsd.uta.fi/aineistonhallinta/en/informing-research-participants.html> [4.2.2016]

<sup>2</sup> FSD – Processing Qualitative Data Files - <http://www.fsd.uta.fi/aineistonhallinta/en/processing-qualitative-data-files.html> [4.2.2016]

their research, hence they are more willing to cooperate with data archives. More about methodological and ethical dilemmas can be read in the IASSIST Quarterly 25<sup>3</sup>.

The session ended with an open discussion and SEEDS partners explaining the difficulties they have in gaining RRPP data. Dr. Solska promised to send an additional letter to the project leaders, explaining the importance of making their material available. It is clear from contacts already made that some surveys will not have data. SEEDS partners will, with additional help of FORS if needed, do their best in convincing researchers about the importance of data preservation. In cases where data will not be accessible, the study description of the project will still be made available through FORSbase=SEEDSbase. Due to the absence of the former SERSCIDA partner from BiH, UL-ADP agreed to organise the work on surveys covering BiH.

In the second part of the day, we focused on Pre-SIP and SIP, which are part of the OAIS model<sup>4</sup>. **Brian Kleiner**, from FORS, talked about the importance of the availability of proper documentation of a survey. He said that "*Documentation is any information that serves as a record of a research project and that renders data usable and meaningful*". Initial contact with researchers will include discussions about data and documentation. In order to prepare study descriptions and data as well as possible, it is important to take time for the pre-SIP and SIP stage and clear issues, such as unknown variables in a data file or missing documentation. A list of what kind of documentation a data archivist should get from researchers was provided. It was emphasised that when depositing the data, the researcher should define the conditions of access. Data archives strongly promote open access to data, but this might not be always feasible (e.g. due to protection of individuals taking part in the survey). Some special conditions about access can be declared by researchers in such circumstances. The deposit contract and end user contract from FORS were shown and explained. Researchers need to be aware of what will happen to their data when it will come to the archive, and how it will be used when it leaves the archive. Users will have to respect conditions and properly cite materials used. The researcher that has produced the data is still the owner of the data, but the data archive has the right to distribute the data on the basis of predefined conditions. In FORS, most of the surveys are distributed for non-commercial purposes to authenticated users.

The partners in the SEEDS project were asked to try to contact RRPP researchers in advance and bring data and material from certain surveys to this workshop, so they could go through the whole process of data archiving. In the first day of the workshop partners described surveys and materials they were able to gain and will work on in this workshop. Most of the partners reported difficulties in contacting leading researchers and gaining materials from them. It was concluded that more effort will be needed to gain data or even a response from researchers in RRPP projects.

The decision was made beforehand to use ownCloud for saving and sharing materials.

By agreement with FORS, their new platform FORSbase will be used for archiving and distribution of RRPP studies. FORS will create a separate instance of FORSbase for the

---

<sup>3</sup>A. Kuula: Methodological and ethical dilemmas of archiving qualitative data. IASSIST Quarterly. 2011. 25:1-2. 12-17. [http://www.iassistdata.org/sites/default/files/iq/iqvol34\\_35\\_kuula.pdf](http://www.iassistdata.org/sites/default/files/iq/iqvol34_35_kuula.pdf) [4.2.2016]

<sup>4</sup> See presentation from I Workshop for more information about OAIS [http://seedsproject.ch/?page\\_id=380](http://seedsproject.ch/?page_id=380)

SEEDS project /RRPP studies. It will be named SEEDSbase<sup>5</sup>. **Brian Kleiner** and **Sybil Krügel** presented FORSbase and its functionalities during the workshop. FORSbase is a tool that supports the entire cycle of research data and makes the archiving process easier. It is designed in a way that researchers can upload data and document surveys themselves, and prepare study descriptions for which they are most equipped. The material is then reviewed by FORS staff and published on the FORS website. The participants of the workshop were getting familiarised with uploading data and documentation into FORSbase. They learnt how to prepare a study description, review accessible material and data and prepare it for secondary use. In the hands-on sessions participants used a test version of FORSbase to create a new survey instance and then insert survey level metadata, as well as uploading and describing data files and documentation for the test RRPP surveys.

The second day (10<sup>th</sup> of February) was dedicated to the AIP and DIP part of the OAIS model – so archival and distribution materials were prepared. In the morning **Irena Vipavc Brvar**, UL-ADP, talked about the documentation that a data archivist should gain from researchers, formats in which ADP is accepting materials and how materials should be treated in the archive. She emphasised that it is important to use unique identifiers and versioning. ADP uses JIRA as a tool for project tracking. In it they describe the work that was performed on each survey. It is also important to have information about which data were already cleaned, checked, prepared and what is left open. Unfortunately, researchers are usually busy and sometimes it takes several days or even weeks for them to answer. In some cases, the data archivist might already have left the archive for another job, and the need might arise to correct something in a study description or data a few years after the original publishing. Having proper documentation on study preparation proved to be extremely helpful. DDI fields for File Description and Data Description level were elaborated and the use of the Nesstar Publisher tool for preparing content of this fields was shown and explained.

**Arja Kuula-Luumi**, FSD, presented the archiving process of qualitative data in FSD. She expressed her opinion that it was important to first review a data and documentation as a whole. In most cases qualitative surveys might be personal and intense anonymisation would be needed. But there are also cases where this is part of already publicly available information and anonymisation will not be necessary. In case of qualitative surveys, it is important to prepare unit level metadata. A data unit could be an interview, a diary, a life story or any other written text, or pictures; and elements on metadata level are usually background information of the participants, situational information and information concerning the content of data. Most data prepared for secondary use will need to be anonymised. Personal data and direct identifiers will have to be excluded. Kuula-Luumi defined personal data as any kind of data that may be used to identify a natural person and direct identifiers as information that is sufficient on its own to identify an individual. Anonymisation of unit level metadata and data was shown on few examples. Since few of the surveys in the RRPP pack are of qualitative nature, it was important to go through the presentation in details. In the hands-on session, Kuula-Luumi went through qualitative surveys that were brought to the workshop, with the help of participants, and discussed how she would propose cleaning and preparing them case by case.

**Irena Vipavc Brvar**, ADP, then added parts that were important for curation of quantitative data. Data should be cleaned, anonymised and prepared by the original researcher, and the data archive offers its expertise. Before starting cleaning and anonymising a data file, it

---

<sup>5</sup> <https://seedsdata.unil.ch/>

should be clear under which access conditions it will be available. We can distinguish between a Secure Use File (ScUF), which will most likely be accessible in a secure environment, a Scientific Use File (SUF), a Public Use File (PUF), which are most commonly distributed and a Campus Use Files (CUF), usually specifically prepared for training purposes. Similar rules for anonymisation, as discussed for qualitative interviews, apply also for quantitative data. Personal information of respondents needs to be protected. While preparing data for long-term preservation and use, it is important to prepare files and content in a way that it can be used with nonproprietary software. The use and interoperability between different statistical software is thus guaranteed.

The participants of the workshop were able to work on their materials. Since all changes made on original file need to be documented, the use of syntax command files was recommended. Guides on SPSS and STATA syntaxes was prepared and introduced<sup>6</sup>. These two statistical packages were recognised as the most common statistical software used in participants' countries. At the end of the hands-on section, participants presented their preliminary work, and some dilemmas were discussed. Since many issues will be the same in other countries (other RRPP surveys) we felt it was useful to present how someone tackles the issue, and which were proposed solutions from the lecturers, based on principles presented in the introduction.

The third day (11<sup>th</sup> of February) was used to discuss best practices on data dissemination – DIP as part of the OAIS model (given by **Sybil Krügel**, FORS). A DIP is usually derived from an AIP, yet it might differ in content and form. Not everything that is saved in an AIP is needed in a DIP. Documents needed to understand specific data should be distributed in a format most suitable to users. Ms. Krügel demonstrated how to prepare a DIP in FORSbase and how a user would access those content / material. In addition, the content and implications of the end user contract of FORS were discussed. Among other things, their users should respect Swiss federal law and norms of data protection, use documentation and data in conformity with scientific standards, cite study or materials used, and respect confidentiality. A user should destroy data after the analysis and reapply for access if needed. FORS is not responsible for misuse of data. Data should not be provided to third parties. If a study has certain specific conditions, prior authorisation of use by the data depositor might be needed.

We ended the three days event with a **SEEDS project meeting** where the next steps in the project were planned. It was agreed that a separate mailing list for the RRPP project will be created, so the communication on the RRPP project will not burden SEEDS project directly. Partners should provide names of people who will be involved. At a first stage, RRPP projects need to be evaluated in order to know how much work will be needed to prepare surveys for secondary use. FORS will prepare a more detailed overview of surveys available (either from RRPP or Balkandata website). The documentation might be in different languages so different partners might be involved in one study. A list of research projects will be available via Google docs. Each of us can edit and make comments.

Furthermore, forthcoming SEEDS deliverables were discussed, such as the report on evaluation of research and legal conditions (D3), the establishment plan (D4) and the policy and procedures document (D5). The templates prepared for the SERSCIDA project could be of some help here. For the report on integration of technical system (D6) partners need to

---

<sup>6</sup> Other Materials: [http://www.adp.fdv.uni-lj.si/seeds\\_workshop2\\_lj2016/other-materials-en.php](http://www.adp.fdv.uni-lj.si/seeds_workshop2_lj2016/other-materials-en.php)

think about tools they would like to use for their services and how they would like to use their website. Due to the absence (maternity leave) of the main technical person in the SEEDS project, all technical implementations will be postponed for a few months. In the meantime, the content of the website should be considered. The report on the present workshop (D7) will be prepared by UL-ADP, with the help of CPC, at the beginning of March 2016.

The third workshop will take place in Zagreb. The preliminary date agreed upon was the first week of November 2016. This workshop will cover the Data Seal of Approval (DSA). Each participant will evaluate the steps that will need to be taken in account for them to satisfy the requirement of the DSA. The workshop will also cover important questions concerning data security and backup, from technical and conceptual viewpoints.

Partners should also think about preparing local events (WP4) after the third workshop in Zagreb.

The official reporting period for the SEEDS project is from May 1, 2015 until April 30, 2016. FORS will review funders' requirement and send information to partners so they can start preparing the financial reports.

## 2. List of Participants

	<b>Last name, First Name</b>	<b>Organization</b>	<b>Country</b>
1	Babaunski, Klime	ISPJR - Institute for Sociological, Political and Juridical Research	Macedonia
2	Banović, Jelena	IES - Institute of Economic Sciences Belgrade	Serbia
3	Bezjak, Sonja	ADP - Social Science Data Archives	Slovenia
4	Bradić-Martinović, Aleksandra	IES - Institute of Economic Sciences Belgrade	Serbia
5	Cekik, Aneta	ISPJR - Institute for Sociological, Political and Juridical Research	Macedonia
6	Demiri, Lirika	CPC - Centre for Political Courage	Kosovo
7	Gjipali, Ardit	Agency for Research, Technology and Innovation	Albania
8	Glavica, Marijana	FFZG - Faculty of Humanities and Social Sciences, University of Zagreb	Croatia
9	Gorišek, Žiga	ADP - Social Science Data Archives	Slovenia
10	Haxhiu, Genc	CPC - Centre for Political Courage	Kosovo
11	Jelovica, Vatroslav	FFZG - Faculty of Humanities and Social Sciences, University of Zagreb	Croatia
12	Kleiner, Brian	FORS - Swiss Centre of Expertise in the Social Sciences	Switzerland
13	Kranjec, Irena	FFZG - Faculty of Humanities and Social Sciences, University of Zagreb	Croatia
14	Krügel, Sybil	FORS - Swiss Centre of Expertise in the Social Sciences	Switzerland
15	Kuula-Luumi, Arja	FSD - Finnish Social Science Data Archive	Finland
16	Malić, Larisa	IES - Institute of Economic Sciences Belgrade	Serbia
17	Masten, Sergeja	ADP - Social Science Data Archives	Slovenia
18	Mitrushi, Artela	IDM - Institute for Democracy and Mediation	Albania
19	Ojsteršek, Maja	ADP - Social Science Data Archives	Slovenia
20	Serafimovska, Eleonora	ISPJR - Institute for Sociological, Political and Juridical Research	Macedonia
21	Solska, Magdalena	RRPP - Regional Research Promotion Programme, University of Fribourg	Switzerland
22	Štebe, Janez	ADP - Social Science Data Archives	Slovenia
23	Tomović, Nikoleta	CeMI - Center for Monitoring and Research	Montenegro
24	Vipavc Brvar, Irena	ADP - Social Science Data Archives	Slovenia
25	Vodopijavec, Alen	FFZG - Faculty of Humanities and Social Sciences, University of Zagreb	Croatia
26	Vujovic, Ivana	CeMI - Center for Monitoring and Research	Montenegro
27	Xhemajli, Valmir	CPC - Centre for Political Courage	Kosovo
28	Zabijakin Chatleska, Vesna	ISPJR - Institute for Sociological, Political and Juridical Research	Macedonia
29	Zdravković, Aleksandar	IES - Institute of Economic Sciences Belgrade	Serbia

## Annex I: Content of the Workshop



# Ljubljana Workshop

(9-11 February 2016)

Faculty of Social Sciences, Kardeljeva ploščad 5, Ljubljana, Slovenia

Computer lab 25

### Tuesday, February 9, 2016

09.00 to 09.15	<b>Welcome and opening</b> (Janez Štebe, ADP and Valentina Hlebec, Vice-dean for Research, FDV)
09.15 to 09.30	<b>Programme and objectives of workshop</b> (Brian Kleiner, FORS)
09.30 to 11.00	<b>Invited speakers</b>  About the RRPP data (Magdalena Solska) Communicating with researchers (Arja Kuula, FSD)
11.00 to 11.15	<i>Coffee and tea break</i>
11.15 to 12.30	<b>Pre-SIP and SIPs: best practice in ingest</b> (Brian Kleiner, Sybil Krügel FORS; Arja Kuula, FSD) Getting the data and documentation from researchers Metadata at the study level Negotiating with researchers, including contracts How to create SIPs
12.30 to 13.30	Lunch
13.30 to 15.30	<b>Hands- on: ingesting the RRPP data</b> (Brian Kleiner, Sybil Krügel FORS)
15.30 to 15.45	<i>Coffee and tea break</i>
15.45 to 17.00	<b>Hands- on: ingesting the RRPP data (continued)</b> (Brian Kleiner, Sybil Krügel FORS)

## Wednesday, February 10, 2016

09.00 to 10.30	<b>AIPs and data management: best practice in curation</b> (Irena Vipavc Brvar, ADP; Sybil Krügel FORS) <ul style="list-style-type: none"><li>• How to create AIPs</li><li>• Metadata at the data and file level</li></ul>
10.30 to 10:45	<i>Coffee and tea break</i>
10.45 to 12.30	<b>AIPs and data management: best practice in curation</b> <ul style="list-style-type: none"><li>• Curating qualitative data (Arja Kuula, FSD)</li><li>• Curating quantitative data (Irena Vipavc Brvar, ADP; Sybil Krügel FORS)</li></ul>
12.30 to 13.30	<i>Lunch</i>
13.30 to 15.00	<b>Hands- on: : curating the RRPP quantitative data</b> (Irena Vipavc Brvar, ADP; Brian Kleiner, Sybil Krügel FORS)
15.00 to 15.15	<i>Coffee and tea break</i>
15.15 to 16.30	<b>Hands- on curating the RRPP qualitative data</b> (Arja Kuula, FSD)
16.30 to 17.00	<b>DIPs for qualitative data: best practice in data dissemination</b> (Arja Kuula, FSD)
17.30	<i>Dinner – <a href="#">Restaurant Favola</a>, Dunajska street 129 (on your expense)</i>

## Thursday, February 11, 2016

09.00 to 10.00	<b>DIPs: best practice in data dissemination</b> (Brian Kleiner, Sybil Krügel FORS) <ul style="list-style-type: none"><li>• Determining access conditions</li><li>• Creating user contracts</li><li>• How to create DIPs</li></ul>
10.00 to 10.15	<i>Coffee and tea break</i>
10.15 to 12.30	<b>Hands-on: preparing the RRPP data for dissemination</b> (Brian Kleiner, Sybil Krügel FORS)
12.30 to 13.30	<i>Lunch</i>
13.30 to 15.00	<b>Planning next steps for SEEDS</b> (Brian Kleiner, FORS)

---

*End of the Workshop*