

## Final report “RRPP Data Rescue”

*Author(s):* Renate Kunz (FORS)  
Brian Kleiner (FORS)

*Submission date:* 28 February 2017  
*Project:* RRPP Data Rescue  
*Start date of project:* 1 April 2016  
*Duration:* 10 months  
*Reporting period:* 1 October 2016 to 31 January 2017

**Table of Contents**

- 1. General Information ..... 3
  - 1.1 Name of co-ordinator and partner teams..... 3
- 2. Overview of activities ..... 3
  - 2.1 Work carried out ..... 4
  - 2.2 Collaboration ..... 7
  - 2.3 List of involved individuals ..... 8
- 3. Conclusion ..... 9
- 4. Annex..... 10
  - 4.1 RRPP Workflow..... 10

# 1. General Information

## 1.1 Name of co-ordinator and partner teams

Name	Country	Role
Swiss Foundation for Research in Social Sciences, Lausanne (FORS)	Switzerland	Co-ordinator
Foundation Analitika – Center for Social Research	Bosnia and Herzegovina	Partner
Centre for Political Courage, Pristina (CPC)	Kosovo	Partner
Institute for Democracy and Mediation, Tirana (IDM)	Albania	Partner
Institute of Economic Sciences, Belgrade (IES)	Serbia	Partner
Saints Cyril and Methodius University, Institute for Sociological, Political and Juridical Research, Skopje (ISPJR) <sup>1</sup>	Macedonia	Partner

In addition to the partners listed above, in November of 2016 FORS decided to mandate its SEEDS partner institution [ADP](#) (Social Science Data Archive) from Slovenia to assist with the data curation phase of the project. ADP was paid from available remaining funds from the SCOPES/SEEDS project (in which they participate). The SEEDS project includes tasks in data collection and archiving for all partners, with the aim of building up collections of data within each partner country. ADP was brought in to help with the RRPP “data rescue” project because it became apparent that FORS alone could not carry out all the data curation work in the short time remaining until the end of the project. Also, ADP had the linguistic skills needed to work with the qualitative data produced by some of the RRPP projects. Their work began in November 2016 and ended in January 2017.

## 2. Overview of activities

The “RRPP Data Rescue” project was established to ensure the long-term preservation of data produced within the framework of the Regional Research Promotion Programme (RRPP). The aim was to make these data available for secondary use, to the largest extent possible. This is consistent with the goal of the RRPP to promote and strengthen research capacity in the Western Balkans.

The aim was to archive the data in a safe environment, and to make it usable by a large scientific community. Most of the RRPP data have additional analytic value that could be of use to other researchers to address new research questions. The safeguarding of the data was one of the most important achievements of the project and is guaranteed with SEEDSbase for the years to come.

As it became evident already at the beginning of the project, not all the data from every RRPP project could be obtained. RRPP partners have encountered difficulties obtaining not only data but even information on the study from some of the researchers. In a few cases, proper communication and cooperation was not possible and did fail. The reasons may be manifold, yet FORS together with their partners tried to convince researchers of the positive effects and benefits that depositing data can

---

<sup>1</sup> Staff from ISPJR participate as individuals.

bring and offered their help with data curation issues. In the end partners managed to obtain data for half of the RRPP studies.

## 2.1 Work carried out

As already stated in the progress report “RRPP Data Rescue”, the data evaluation and data deposit activities were more time-consuming than initially anticipated. However, partners had already start entering data and metadata in our system [SEEDSbase](#), which was set up as an additional instance of [FORSbase](#) at the beginning of the project “RRPP Data Rescue”. It is a virtual workspace for researchers in the social sciences from countries from South-eastern Europe that have no established data services. So far, SEEDSbase contains studies from RRPP only but is generally open to research projects from other origins than RRPP. SEEDSbase is designed for researchers to enter and update their studies, to deposit their data and to search for and download data from other research projects within the [SEEDSbase catalogue](#).

Displayed below are the log-in page and the catalogue of SEEDSbase:

The screenshot shows the SEEDSbase login page. At the top left is the SEEDSbase logo. To the right of the logo are language selection links: FR, EN, DE. Below these are two tabs: 'Catalogue' and 'Home'. In the top right corner, there is a 'Login' link. The main area on the left contains a login form with 'Username' and 'Password' input fields, a 'Login' button, and links for 'Forgot your username or password?' and 'Sign up'. On the right side, there is a 'Useful Resources' section with a list of quickguides: 'FORSbase: overview', 'Explore the catalogue', 'Get data', 'Register your study', and 'Deposit your data'. Below this are links for 'FAQs', 'Other resources', and 'Need help?'. Further down are logos for 'FORS' and 'seeds', and a link to 'Browse data on Nesstar'.

**Project and data catalogue**

Free text search:  Has archived data:

Advanced search

Show  entries

Ref study	Study title	End date	Data
2	Društveni i kulturni kapital u Srbiji	30.06.2012	No
10	Making work pay in Western Balkan countries: the case of Serbia and Macedonia	30.06.2013	No
16	(In)equality in Social Protection: Multi-level Analysis of Intersectionality in Social Assistance Provision- A Comparative Study	15.03.2016	No
17	Izgradnja kulturalnog identiteta i ključni razvoji unutar suvremenih nacionalnih projekata u Bosni i Hercegovini: multidisciplinarnе perspektive iz medija i kulturalne produkcije	01.01.2011	No
18	Ethnic Tensions, Institutional and Economic Performance in Bosnia and Herzegovina: What are (not) the Causal Links?	01.01.2014	No
19	U okrilju nacije: Etnički i državni identitet kod mladih u Bosni i Hercegovini	01.09.2011	No
20	Izgradnja države ili nacije: vizije, konverze i perspektive političke tranzicije u Bosni i Hercegovini	15.10.2010	No
21	Development of Functional Media Institutions in Western Balkans – A Comparative Study	01.01.2014	No
22	Feminisms in Post-Socialist Muslim Contexts	01.04.2012	No
23	Linking Rural Entrepreneurs and Diaspora in Albania and Bosnia-Herzegovina	01.04.2013	No

Showing 1 to 10 of 22 entries

It became evident at the end of 2016 that due to the above mentioned reasons the project “RRPP Data Rescue” needed to be prolonged in order to have an outcome with a reasonable quality. That is why the project leader FORS asked for a one-month extension of the project. RRPP Fribourg kindly accepted the proposal.

The remaining time of the project was dedicated to data curation activities and study and data entry tasks in SEEDSbase. Partners were asked to fill in the document “Detailed reporting for RRPP” (see Annex 4.1) so that FORS could have an overview of all studies and the remaining work. At this point in time (October) all projects were terminated and therefore ready to be archived. However, it became evident that not all RRPP studies would appear in SEEDSbase with an entry.

In addition to the “Detailed reporting for RRPP”, a contact list was created in order to oversee the process of data evaluation, deposit and curation until the moment that the data would be entered into SEEDSbase.<sup>2</sup>

At the [third SEEDS training workshop](#), which took place from November 2-4, 2016 in Zagreb, Croatia, RRPP partners had the opportunity to discuss practical issues in data cleaning and anonymization strategies within the RRPP data rescue project.

<sup>2</sup> The list “Contact list\_RRPP all projects” can be consulted on demand.

Activity	Description	Outcome
0 Management tasks, supporting functions, and quality control	Progress, final narrative and financial reports; quality control checks in SEEDSbase	Progressive narrative and financial report Final narrative and financial report
1 Data evaluation	Matching projects, contact researchers, initial assessment of work	
2 Data deposit	Negotiate deposit and access conditions, data transfer	Deposit contract and access policy
3 Data curation	Data cleaning, data anonymisation, file format migration, documentation and metadata preparation	Data ready to put in SEEDSbase
4 Studies and data in SEEDSbase	Entry of studies and data	Collection of 80 projects in SEEDSbase

There were a total of 80 projects to be curated in the “RRPP Data Rescue” project. Each project was attributed to the country of the principle investigator. The aim was to have at minimum all study descriptions in SEEDSbase at the end of the project. Exceptions are the so called “State Capacities” projects. They ran from September 2015 to September 2016 and were designed as individual projects, not linked to any institutions. One project was cancelled and four projects only have a policy brief that are not sufficient enough to have quality study descriptions that would serve future users. In addition to that, the researchers responsible for those projects were not responsive. This is why we have currently 74 study description publicly available in the catalogue in SEEDSbase.

	IES	IDM	CPC	ISPJR	Analitika	Total
Number of RRPP projects	26	11	3	18	22	80
Projects dealt within “RRPP Data Rescue” project	23	10	3	18	20	74
Number of multinational projects	3	0	1	11	5	13
Projects with data obtained...	13	1	2	10	14	40
...out of which could be put in SEEDSbase	13	0	1	10	13	36
Opting out requests	0	0	1	1 <sup>3</sup>	1	3
Number of datasets in SEEDSbase <sup>4</sup>	32	0	0	1	14	47
Number of data files	99 (41)	0	0	1 (1)	58 (49)	158 (91)

<sup>3</sup> There are two other opting out requests for projects from Macedonia, but those are only partially for some qualitative data.

<sup>4</sup> A dataset in SEEDSbase consists of a data file, or several data files, and data-related documentation. There can be several datasets linked to a study if they are different in terms of their topic, population, method (i.e. use of quantitative and qualitative methods) and/or moment of the data collection (i.e. waves with large time intervals).

There are 13 projects that were worked on in collaboration by the partners. Most work consisted of anonymization of interview texts that were led in several countries. From half of the projects, data could be obtained. With some few exceptions, related to uncleaned and unlabelled data and three opting out requests, those data can be accessed online in the catalogue in SEEDSbase. The opting out requests can be found in the annex 4.2. In most opting out cases, the future data user has the option to contact the data depositor directly and ask for accessing the corresponding data.

The rights of data depositors and data users are handled according to the data and research information service (DARIS) regulations at FORS. The data producer's rights and interests are protected in several ways: The data depositor maintains the copyright on the data that they have provided and the data user is bound by the user contract to cite the data source in all publications referring to the data used. The rights and conditions are formalised in a deposit contract between the data depositor and DARIS.<sup>5</sup>

As a general rule, access to data within SEEDSbase will only be given to data that have been anonymised by removal of variables that could allow direct identification of individual participants, such as name and address. Access will be provided only to individuals affiliated with a research institution (university or university of applied science) and after accepting a user contract. This implies citing the data in a proper fashion, for scientific research and education within an academic framework, and following scientific ethical norms of conduct.<sup>6</sup>

Several datasets within SEEDSbase were put under restricted access, this means that the data producer (the PI of a RRPP project in most cases) is informed before the data is accessed and needs to give the required permission for accessing the data. This is mainly the case with qualitative data that are of sensitive nature and/or where anonymization could not be fully accomplished.

The data curation and data entering work in SEEDSbase is still ongoing while writing this report. Several studies with data are in the process of being entered in the system. There might be even more data from studies where we could not obtain data in the first place. We are confident that there will be more data from other research projects from the Balkans being deposited in SEEDSbase, especially for those institutions that do not have access to a sustainable archiving system.

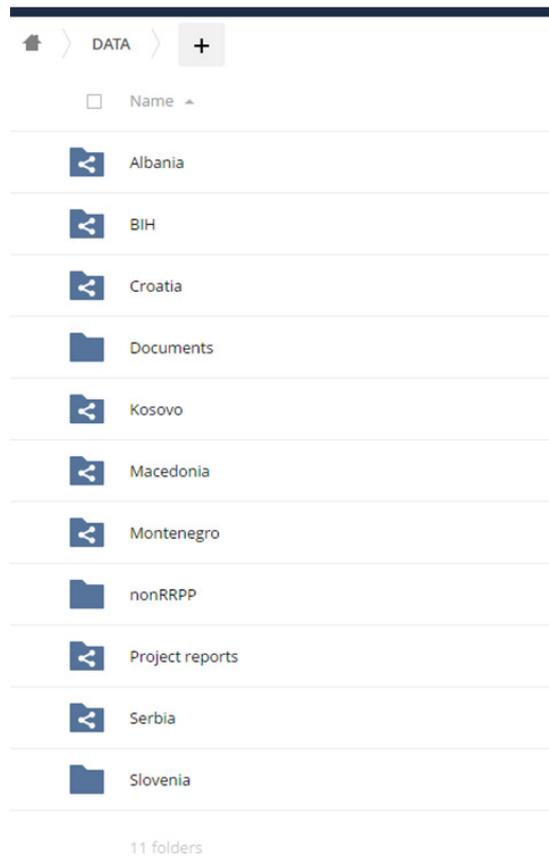
## 2.2 Collaboration

In addition to GoogleDrive, which was used to share common documents, [OwnCloud](#), an open source, self-hosted platform was used in order to be able to synchronise and share files on the one side, as well as to put certain files under restricted access/save environment. Each partner received a user name and log-in for the web access.

---

<sup>5</sup> An example of a generic deposit contract can be found here: [https://seedsdata.unil.ch/media/general\\_documentation/en/deposit\\_contract\\_FORS\\_en.pdf](https://seedsdata.unil.ch/media/general_documentation/en/deposit_contract_FORS_en.pdf). As soon as there is a data service established in one of the partner countries, the data will be exported to the applicable data service and deposit and user contracts need to be renegotiated.

<sup>6</sup> An example of a generic user contract can be found here: [https://seedsdata.unil.ch/media/general\\_documentation/en/download\\_contract\\_en.pdf](https://seedsdata.unil.ch/media/general_documentation/en/download_contract_en.pdf)



The monthly (and at the end even biweekly) Skype meetings, organised by the coordinating team at FORS, were continued and used to discuss open questions related to data curation issues and technical aspects about SEEDSbase.

The collaboration during the last months of the project has been very satisfactory, as assessed by the main applicant.

### 2.3 List of involved individuals

Name	Country	Sex	Role	Remarks
Kunz, Renate	Switzerland	f	Project coordinator	
Kleiner, Brian	Switzerland	m	Principle investigator	
Tasic, Bojana	Switzerland	f	Technical coordinator	
Bradic-Martinovic, Aleksandra	Serbia	f	Coordinator for Serbia	
Zdravkovic, Aleksandar	Serbia	m	Team member	
Malic, Larisa	Serbia	f	Team member	
Banovic, Jelena	Serbia	f	Team member	
Cekik, Aneta	Macedonia	f	Coordinator for Macedonia	
Babaunski, Klime	Macedonia	m	Team member	
Zabijakin Chatleska, Vesna	Macedonia	f	Team member	
Eleonora Serafimovska	Macedonia	f	Team member	

Mitrushi, Artela	Albania	f	Coordinator for Albania	
Bylykbashi Deliu, Nitë	Kosovo	f	Coordinator for Kosovo	
Demiri, Lirika	Kosovo	f	Team member	
Hajrullahu, Arben	Kosovo	m	Advisor	
Kelmendi, Pëllumb	Kosovo	m	Advisor	
Haxhiu, Genc	Kosovo	m	Data Specialist	
Xhemajli, Valmir	Kosovo	m	Data Specialist	
Korlat, Selma	Bosnia	f	Junior researcher	

### 3. Conclusion

To conclude, approximately half of the data from RRPP will be safely preserved for the long-term in SEEDSbase. All projects from RRPP are recorded at study-level; they have been published in SEEDSbase and can be discovered around the world by way of the [SEEDSbase catalogue](#) and are openly available for secondary use.

There are several challenges that RRPP partners had to cope with in the course of the project. It was hard to convince researchers of the positive effects of data management tasks and depositing data. Even more when their data were of sensitive nature and could harm individuals. A combined strategy of directed guidance from our RRPP partners as well as repeated but regulated communication did have an impact on the final result which is, obtaining data for half of the RRPP studies.

Currently, FORS is responsible for the maintenance of SEEDSbase and is hosting the domain in parallel and with the same accuracy as their own platform FORSbase. SEEDSbase will serve as a technical online solution for countries in the Balkans that do not have a sustainable national archiving system, until the time that individual countries establish data services that could take over the data and metadata from SEEDSbase.

## 4. Annex

### 4.1 RRPP Workflow

#### Communication with researcher

##### Contact researchers and obtain data and documentation

- Contact researcher, introduce project, and address researcher concerns ([http://www.adp.fdv.unilj.si/seeds\\_workshop2\\_lj2016/presentations/Model%20for%20initial%20contact%20with%20RRPP%20researchers.pdf](http://www.adp.fdv.unilj.si/seeds_workshop2_lj2016/presentations/Model%20for%20initial%20contact%20with%20RRPP%20researchers.pdf)).
- Explain conditions for data and documentation submission and archiving: point researcher to FORS guidelines for deposit at: <http://forscenter.ch/en/data-and-research-information-services/2221-2/deposit-data/guidelines-for-deposit/>.
- Discuss access conditions, data protection, and deposit contract; help at: <http://forscenter.ch/en/data-and-research-information-services/2221-2/deposit-data/data-protection/>.

You can negotiate

- a) "Data to be published on" - embargo date (preferably not more than 1 year - until January 2018);
- b) "Special permission" - option "With prior agreement of author";
- c) "Restrictions" - "Academic research only".

However, it is advised to have as minimal (or none) restrictions as possible.

- Communicate or negotiate the means of transfer of data and documentation, preferably using ownCloud (<https://www.serscida.eu/RRPP-Data/>). If researcher feels really strongly about not putting data on the cloud, he might get access to SEEDSbase and can upload them directly.
- Have the researcher transmit the relevant data and documentation, and confirm receipt.
- Have researcher sign deposit agreement. [https://seedsdata.unil.ch/media/general\\_documentation/en/deposit\\_contract\\_FORS\\_en.pdf](https://seedsdata.unil.ch/media/general_documentation/en/deposit_contract_FORS_en.pdf) (deposit contract);

Inform them about the user agreement.

[https://seedsdata.unil.ch/media/general\\_documentation/en/download\\_contract\\_en.pdf](https://seedsdata.unil.ch/media/general_documentation/en/download_contract_en.pdf)

##### Use of OwnCloud

All studies should have a separate folder in the Cloud. If you are missing yours you should add it. Please see a guide from Alen on how to use OwnCloud for the RRPP project.

The primary material that we had at the start of the project (downloaded from Balkandata.org) were added in the folders.

**This is the working space you should use for your work.**

Please create 3 folders inside

- SIP (*Submission information package - this will be a folder for material that you received from researchers or other public sources*)
- WD (*Working directory – for everything in between. Study description, intermediate versions of the files etc.*)
- DIP (*Distribution information package - this will be a folder for material that you will prepare for "public" distribution and will be uploaded to SEEDSbase. From uploaded files FORS will create actual AIP and DIP with proper identifiers.*)

## Use of Reporting document

On GoogleDrive in RRPP folder you will find a document Contact list RRPP all projects.xlsx. Please try to fill in as much information possible.

- All communication with researcher or anyone else in relation to the study;
- Contact information for everyone that you contacted;
- Add information about work done and the amount of time devoted for this;
- What you received / or are in the process of receiving / you will not receive and why.

## Preparing materials in-house

*Put working documents in WD folder; really final documents in DIP folder.*

Ingest and data preparation

- Review data file(s)
- Get clarification from researchers, if needed
- Get assistance from researchers, if needed (e.g., for anonymisation)
- Review documentation
- Get clarification or additional documentation from researchers, if needed
- Conduct additional cleaning of data, if needed:

[http://www.data-archive.ac.uk/media/54767/cd093-qualitivedatacollectioningestprocessingprocedures\\_08\\_00w.pdf](http://www.data-archive.ac.uk/media/54767/cd093-qualitivedatacollectioningestprocessingprocedures_08_00w.pdf)

[http://www.data-archive.ac.uk/media/54770/cd081-quantitivedataingestprocessingprocedures\\_08\\_00w.pdf](http://www.data-archive.ac.uk/media/54770/cd081-quantitivedataingestprocessingprocedures_08_00w.pdf)

**All changes in data need to be documented.**

- Organise and prepare documentation (e.g., selection of relevant files):

[http://www.data-archive.ac.uk/media/54785/cd078-documentationingestprocessingprocedures\\_08\\_00w.pdf](http://www.data-archive.ac.uk/media/54785/cd078-documentationingestprocessingprocedures_08_00w.pdf)

Some more clarification for Qualitative materials:

- Save data in \*.rtf format and perhaps also in PDF/A if you have possibility (make sure it is searchable).
  - See list of proposed formats also by UKDA (<http://data-archive.ac.uk/create-manage/format/formats-table>)
  - In the structure of directories in FSD you can also see \*.rtf (<http://www.fsd.uta.fi/aineistonhallinta/en/processing-qualitative-data-files.html>)
- If formatting is important for understanding of the context we would propose leaving it as it is.
- Details on background information depends on content of the survey itself. It is difficult to say in general what should be included. Perhaps: Name of Interviewed person / organisation, age, gender, occupation, location of interview, date of the interview etc.
- As of how to organize files – good examples are provided on FSD page. It is best if the file has some logical information in the title – like if it is questionnaire / see part “Naming data files”

<http://www.fsd.uta.fi/aineistonhallinta/en/processing-qualitative-data-files.html>



## Preparing description in SEEDSbase

You might have asked researcher to fill in some sort of study description form. Or you got most of the information from available publications. *Put working documents in WD folder; final documents in DIP folder.*

## New persons in SEEDSbase

You will be entering information in SEEDSbase on behalf of PI of the Study. In order to do that, the names of all collaborators need to be entered in the database separately, before you start describing study.

You will find the document in our GoogleDrive / RRPP folder with a name New persons in SEEDSbase.



## Adding Study description in SEEDSbase

Start importing information in SEEDSbase.

- 1) Help yourself with a quick guide - [Quick steps to depositing your data with FORSbase](#).
- 2) There are no detailed instructions on how information needs to be added in SEEDSbase, but:

- **TITLE of the dataset:** please have in mind that this will be a title everyone will use later on for citing. So, it might be that this will not be equal to the title of the project. Or it might be a bit shorter. If you chose to change the title please document this in Reporting documents so a reviewer will know that it is the same thing!
- Please choose following writing **"Title: Subtitle, Year"**. See examples <https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies?q=drug>

Title of the study and the dataset can be different. Title of the dataset is the one that will be cited later on (see for example <https://forsbase.unil.ch/project/study-public-overview/13866/0/>).

- **English titles**  
The rules on writing titles in English are summarised according to The Chicago Manual of Style, 16th edition (<http://www.chicagomanualofstyle.org/16/contents.html>):
  - The title segment, which also applies to unpublished doctoral dissertations, is capitalised in English ("headline style");
  - The first and the last word of the title or subtitle are always capitalised, regardless of the word class (e.g. A River Runs Through It; Generating Alternative Worlds: The Indigenous Protest Poetry of Romaine Moreton);
  - Additionally, all of the following word classes are capitalised: nouns, pronouns, verbs, adjectives, adverbs, and subordinating conjunctions;
  - The articles "the," "a," and "an" (e.g. Mammals of the Asian Steppes), are not capitalised, unless they are the first word of a title or subtitle (e.g. The Human Being);
  - Prepositions are also not capitalised (e.g. Catalogue of Watermarks in Italian Printed Maps), unless they represent the first or last word of a title or subtitle (e.g. On the Concept of Popular Culture), or are a part of a Latin expression (e.g. De Facto; In Vitro);
  - The conjunctives "and," "but," "for," "or," and "nor," and the words "to" and "as" are not capitalised, unless they represent the first or last word of a title or subtitle.

- **Dot at the end** – if what you wrote is not a complete sentence it should not have a dot at the end (like in year field)
- **Year:** when you have a period please write e.g. “2011–2013” - without spaces and with “en dash” in between (see explanation <http://www.thepunctuationguide.com/en-dash.html> )
- **Year2:** Please write data in the form YYYY-MM-DD (example: 2015-04-05 or 2015-11-15 till 2015-12-20)
- **Publication:** Please use Chicago 16<sup>th</sup> ed. Style ([http://www.chicagomanualofstyle.org/tools\\_citationguide.html](http://www.chicagomanualofstyle.org/tools_citationguide.html) )
- **When data is not available**  
Describe all possible collection instruments in Abstract, Reporting, and chose all instruments / methods and describe it in the note field (all specifies, why only part of collection is available etc.).  
Additionally to this, if a study or just data or part of collection from certain study will not be available (either researcher asked for opting out, or you could not get the data from them / or worse you were not able to contact them) add in the »Other significant comment« window (in Methods (instrument)):  
“NOTE: Data collected with this instrument are not available thru the system. For further information please contact primary researcher”.

3) **FINANCED BY:** chose option “Other” and in the Details part writing “Regional Research Promotion Programme – Western Balkans (RRPP)” as it is on official fliers.

4) When you are finished contact FORS for final review.



### Adding related documentation

In the study description window you have a button “Add document”. Please add all that is related (and is not data file) in here. Mark appropriately if this is supposed to be public / or you want internal treatment and other fields.

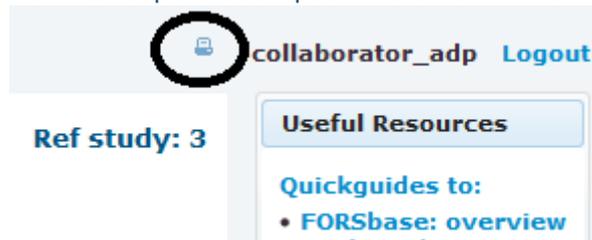
<b>Documentation type *</b>	-----
<b>Origin of document *</b>	-----
<b>Remarks</b>	<ul style="list-style-type: none"> <li>Communications to respondents</li> <li>Instructions to interviewers</li> <li><b>Collection instruments</b></li> <li>Supplementary collection materials (showcards, etc.)</li> <li>Codebook</li> <li>Variable documentation</li> <li>Coding instructions</li> <li>Syntax file</li> <li>Technical report</li> <li>Additional documentation concerning the data</li> <li>Publications (final report, articles)</li> <li>FORS documentation</li> <li>Other</li> </ul>

## Getting authorisation from the author

Since you are entering the description and not a PI, it is important that you receive authorization of the documents from PI or related research, before you publish.



How to export description: Click on Print icon in the above right corner. See picture.



You will need to print each sheet separately (Study information, Study description, Affiliated datasets and Affiliated files). Unfortunately this is in PDF and PI cannot edit it. They will need to send you comments either in text / mail, or scan their remarks they might made on the printed document.

Important: These communication (e-mail) with researcher and especially a sentence saying they authorise publishing should be exported from your tool (Outlook?) and saved in the DIP folder in the cloud.

*At the point of publishing name of SEEDS partner - collaborator will be visible in the citation. There is a way to change the order of the authors and hence, to leave an author out from the citation (with a button on dataset level). However, in order to do that, the study needs to be submitted first.*

## Depositing data and documentation to SEEDSbase

You have all materials prepared, with appropriate names of the file and you checked necessary format add them in SEEDSbase.

