

PREPARING RESEARCH DATA FOR OPEN ACCESS

Guide for data producers

Janez Štebe, Sonja Bezjak, Irena Vipavc Brvar



Arhiv družboslovnih podatkov

Janez Štebe, Sonja Bezjak, Irena Vipavc Brvar
Preparing research data for open access.

Guide for data producers

Publisher: Faculty of Social Sciences, Založba FDV

Ljubljana 2015

Copyright ©

This work of the following authors Janez Štebe, Sonja Bezjak, Irena Vipavc Brvar is licensed under a Creative Commons Attribution 4.0 International License.

Prepared by: Arhiv družboslovnih podatkov
(Social Science Data Archives)

Language editing: Stuart Macdonald

Book editing: Medium Žirovnica

Access mode (URL): <http://knjigarna.fdv.si>

URN:NBN:SI:doc-GoDPXMZ1

CIP - Kataložni zapis o publikaciji

Narodna in univerzitetna knjižnica, Ljubljana

001.891:004.6(0.034.2)

ŠTEBE, Janez

Preparing research data for open access [Elektronski vir] : guide for data producers / Janez Štebe, Sonja Bezjak, Irena Vipavc Brvar. - El. knjiga. - Ljubljana : Faculty of Social Sciences, Založba FDV, 2015

ISBN 978-961-235-722-1 (pdf)

1. Bezjak, Sonja 2. Vipavc Brvar, Irena

277831680

Releasing this guide was financially supported by the FOSTER project.



TABLE OF CONTENT

1 Reasons for open access	1
1.1 Principles	1
1.2 Benefits	2
2 How to participate in the open access	3
2.1 Managing research data	
at the planning and creation stage	3
2.1.1 Quality of study.	4
2.1.2 Ethical obligations.	4
2.1.3 Digital curation.	6
2.1.4 Supporting research	
data management planning	7
2.2 Depositing data with a data centre for preservation	8
2.2.1 WHERE: selection of data centre	8
2.2.2 WHY choose ADP?	9
2.2.3 HOW to deposit research data.	10
2.2.4 Research data acquisition stage.	11
2.3 Data publication	12
2.3.1 Access to research data	12
2.3.2 Advantages of acknowledging	
data as a scientific publication.	13
2.3.3 How to participate in promoting data use	15
3 Further support	16
Appendix: The DSA Guidelines 2014-2015	17

1 REASONS FOR OPEN ACCESS

1.1 Principles

Open access to research data means that by default¹ data can be accessed on equal terms and used with minimal restrictions for academic research and teaching purposes and beyond. Research data constitute primary sources that underpin scientific research and enable derivation of theoretical or applied findings. Data should be prepared for open access in a form which would enable future users to assess and understand them for the purposes of reuse.²

There are a number of benefits for supporting open access to research data, the main one being their value to other researchers beyond the original purpose of creation.

¹ In exceptional cases e.g. to protect the identity of human subjects, special or more restricted access conditions are set.

² Štebe, Janez, Sonja Bežjak and Sanja Lužar (2013): Odrpti podatki. Načrt za vzpostavitev sistema odprtega dostopa do raziskovalnih podatkov v Sloveniji [OPEN DATA – Action Plan for the Establishment of a System of Open Access to Publicly Funded Research Data in Slovenia]. Ljubljana: Fakulteta za družbene vede, Založba FDV. [http://www.adp.fdv.uni-lj.si/ohivuu/publikacije/odprti_podatki_zakljucno_porocilo/], p. 20.

³ It is difficult to predict all potential motives for future reuse of data. Taking into account the whole spectrum of possible research problems, use of data could include anything from practical, applied to theoretical or indeed serendipitous research.

Value and importance of data could be evaluated by the following criteria:

- Are the data unique? Can the study or experiment be repeated? Does it represent an observation in time that cannot be recollected and reproduced, or does it have historical or cultural value?
- Is creating data costly or time consuming? Are procedures, knowledge and instruments used of limited availability or do they represent a rare implementation and therefore have greater re-usability for future research?
- Are data accompanied by research publications and are thus important to verify and reproduce the research process and methodologies?

Establishing whether data are useful for somebody else means evaluating their applicability for solving a certain problem. This recognition represents the practical basis of information literacy and as such we ask the following:

- Is there a potential problem or set of problems which could be solved by using the research data? Is there any data demand expected?³
- Could data be located and accessed in a way that a certain problem could be solved?
- Are data properly documented so that users are able to use them appropriately?
- Are the data available in a preservation format that will enable reuse by software and processing tools?

1.2 Benefits

The benefits of open access to research data can be divided into the following:

Benefits for data creators:	<ul style="list-style-type: none"> · Future reuse of data by data creators themselves is made easier if data are properly documented and preserved. · Indirect benefits for data creator's academic reputation and career development include: increased citation counts and wider dissemination of their research findings; data may be formally assessed for scientific excellence and form part of researcher's bibliography thus adding points to the score of scientific excellence.
Benefits for the scientific community:	<ul style="list-style-type: none"> · New research findings which can only be achieved by using and combining available data from multiple sources, e.g. studies of trends or international comparative research, studies of rare or small populations, meta-analysis. · More robust research findings can be derived as a result of validating and through further analysis based on data related to publication. · Increasing the number of findings by enhancing the number of users of the same data, with efficiencies gained through removal of duplication of effort and/or the need to repeat raw data collection due to a shortage of funds and time. · Training of future scientists, using research data for teaching purposes.
Benefits for the public:	<ul style="list-style-type: none"> · Research findings based on existing data on certain topics can improve quality of life; in Social Sciences gaining a better understanding of the community is important, including building collective identity and eliminating prejudice, discrimination, inequality etc. · Open access to data also introduces savings relating to public spending on research and education.

An increasing number of research funders, representatives of the scientific community, and the public recognise the importance and benefits of open access to data. They have in common the principle that research data produced by publicly-funded research are a public good.⁴

When we think about depositing data for open access, the aforementioned benefits should be taken into account. It is also important to consider whether open access is required by the research funder⁵, or if there are ethical obligations, disciplinary codes of practice, institutional policies or a requirement of journal publishers to make available those data associated with the published article.⁶

4 OECD (2007): OECD Principles and Guidelines for Access to Research Data from Public Funding. [<http://www.oecd.org/sti/scienceandtechnologypolicy/oecdprinciplesandguidelinesforaccessstoresearchdatafrompublicfunding.htm>].

Research Councils UK: RCUK Common Principles on Data Policy. [<http://www.rcuk.ac.uk/research/datapolicy/>, 21. 1. 2015].

LERU Research Data Working Group (2013): LERU Roadmap for Research Data. Advice paper, no. 14. [http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf].

Science Europe: Working Group on Research Data. [<http://www.scienceeurope.org/policy/working-groups/Research-Data>, 21. 1. 2015].

5 Usually one of the incentives for open access is also respecting funder's policies.

6 Scientific journals increasingly require open access to data. It has value as a version of data, cited by the article – for the purposes of verification or replication of results.

2 HOW TO PARTICIPATE IN THE OPEN ACCESS

In terms of facilitating open access from the beginning of a project, the basic strategy of research data management planning and actual research data construction is to actively address two objectives. Firstly, to assure and enhance the usability of data for the purposes of the original project, while having in mind the perspective of further reuse, and secondly attending to the enhanced usability for a wider set of purposes and possible users. A study could be of high quality, yet the further reuse potential of resulting data may be low.⁷ In the case of quantitative data, a dataset may include a narrow set of variables which would limit a theoretical set of possible analysis. A multi-thematic study for example with a more diverse conceptualization and with an emphasis on high quality would be better value and have greater research potential for more users.

For the purpose of providing access to high quality data to the wider research community, it is necessary to start proper research data management planning early to avoid significant problems later in the data creation phase. This includes care for quality of conceptualization, sampling and establishing a protocol for data collection, care during collection and creation of data, and finally the preparation of data for analysis. It is important to take care with on-going and final contextual documentation of data as well as preparation of metadata for deposition in a domain data centre, archive or data repository. Furthermore, ethical obligations, selection and appraisal of data with long-term value for preservation should be taken into account. Such efforts result in an enhanced data publication, available in a data centre, citeable as a scientific reference by users in future papers and publications.

2.1 Managing research data at the planning and creation stage

A basic assurance of data quality and future usability is planning and creating research data according to disciplinary standards and good practices, as well as respecting ethical and legal obligations. Research data creators have to ask themselves about what can be done to assure usability of the data even after the end of the project.

⁷ Corti, Louise (2012): Evaluating Research Data. On: Managing the Material: Tackling Visual Arts as Research Data. [http://www.data-archive.ac.uk/media/369163/managing_research_data-14sept2012b.pdf].

2.1.1 Quality of study

One of the main objectives of the researcher should be the generation of high quality data throughout the entire research process.⁸ This includes:

Care for quality of conceptualization, measurement, sampling and creation of a data collection protocol	Using well-accepted disciplinary concepts and methodologies with references in publications, a standardised approach to measuring individual concepts, applying higher number of indicators and a wide set of socio-demographic questions. Introduction of random sampling covering the whole population, attention to sample size, adoption of measures to achieve high response rates. Use of existing survey questions, standard classifications, and a good practice to pretest a questionnaire. ⁹
Care for collection, creation and preparation of data for analysis	Using procedures for data collection, follow instructions and forms, e.g. computer assisted interviewing with conditional routes, range of values and logical coherency checks. Verification and cleaning of data, error correction, adding variable descriptions and value labels which are coherent with the questionnaire, verification of qualitative interview transcript entries etc. Writing notes and field work control, collecting information about reasons for non-response, calculating response rates, collecting paradata. ¹⁰
Care for documentation of data and preparation of metadata for deposition in a data centre, archive or repository	Preparing metadata according to the DDI ¹¹ standard or appropriate structured documentation which would enable transformation to DDI Codebook - including descriptions of the study, variables, files and related materials such as questionnaires or qualitative interview protocols, data entry forms, codebooks, study reports etc.

2.1.2 Ethical obligations

Social Sciences deal primarily with data about individuals and other potentially sensitive or disclosive data. As part of responsible research practice we need to think about how to protect the privacy of research subjects throughout the whole research lifecycle, starting with planning and ending with publishing data in an open access environment. Researchers are obliged to work in accordance with the Personal Data Protection Act,¹² as well as respecting research community codes.¹³ Ethical codes recommend to protect participants against possible harm or distress, assuring participation by obtaining informed consent, and respecting disciplinary methodological standards.

8 See e.g. Arnež, Marta et. all (2012): Smernice za zagotavljanje kakovosti [Quality Guidelines for Official Statistics]. Metodološki priročniki, no. 2. Ljubljana: Statistični urad Republike Slovenije. [<http://www.stat.si/doc/pub/Smernice.pdf>].

9 Hoffmeyer-Zlotnik, Jürgen H. P. and Uwe Warner (2012): Demographic Standards for Surveys and Polls in Germany and Poland: National and European Dimension. Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. [<http://nbn-resolving.de/urn:nbn:de:0168-ss0ar-371202>].

10 UKDA: Create and Manage Data: Formatting Your Data: Quality assurance. [<http://www.data-archive.ac.uk/create-manage/format/quality>, 20. 1. 2015].

DASISH: Publications and presentations. [<http://dasish.eu/publications/presentations/>, 20. 1. 2015].

11 DDI is a metadata specification for Social Sciences. See: DDI Alliance [<http://www.ddialliance.org/>, 20. 1. 2015].

12 Ministry of Justice of the Republic of Slovenia (2013): Personal Data protection act of the Republic of Slovenia. [https://www.ip-rs.si/fileadmin/user_upload/doc/ZVOP-1_in_ZVOP-1a__English_/Personal_Data_Protection_Act_of_Slovenia_status_2013_final_eng.doc].

13 →

Before the start of any re- search project the following need to be considered:	How the collected data will be used?
	How the results will be presented?
	Who will have the access to the data?
	How the research data will be managed after the project is complete?

Any participant in a research project has to agree to participate. Consent could be given in a written or a verbal form. Special attention should be paid to the preparation of the informed consent form and accompanying information sheet when conducting qualitative interviews, creating audio- and video content, or when research involves sensitive subjects and groups e.g. children, medical studies, crime and workplace studies.

A participant has to be informed about the fol- lowing:	General information about the project.
	Participation is voluntary and that the participant may withdraw their consent at any time.
	Measures taken to maintain confidentiality including use of pseudonyms, indirect identifiers, data anonymization, disclosure control etc.
	Potential uses of any collected data.

It is important from the perspective of open access to data that the participants are informed about the planned dissemination, storage and sharing of data.¹⁴

We could use phrases such as:	I agree that the data deposited will be preserved and made accessible in the Social Science Data Archives.
	I understand that other (secondary) users could use the data in their analysis and publications, if they agreed to preserve the confidentiality of individuals and institutions recorded in the materials.

Efficient and carefully prepared information about responsible management of data throughout the whole data lifecycle could enhance the trust of participants. They would perceive the research having rigour and integrity and perhaps be more motivated to participate.

13 Univerza v Ljubljani (2014): Etični kodeks za raziskovalce Univerze v Ljubljani [Code of Ethics for Researchers at University of Ljubljana]. [http://www.uni-lj.si/mma/Etični_kodeks_za_raziskovalce_UL/2014121104120/].

European Science Foundation, ALLEA (2011): The European Code of Conduct for Research Integrity [http://www.esf.org/fileadmin/Public_documents/Publications/Code_Conduct_ResearchIntegrity.pdf].

Slovensko sociološko društvo (1992): Kodeks profesionalne etike SSD [Code of Professional Ethics of Slovene Sociological Association]. [http://www.sociolosko-drustvo.si/wp-content/uploads/2012/09/Kodeks_profionalne_etike_SSD-19921.pdf].

Statistično društvo Slovenije (1991): Deklaracija poklicne etike Statističnega društva Slovenije [Declaration on Professional Ethics of Statistical Society of Slovenia]. [<http://www.stat-d.si/images/files/etika.doc>].

14 ICPSR: Recommended Informed Consent Language for Data Sharing. [<http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confidentiality/conf-language.html>], 21. 1. 2015].

In general, to decrease the disclosure risk, data should be processed in the following ways:	Direct identifiers should be removed (e.g. names, addresses, telephone numbers and other personal identifiers such as personal identification number).
	Indirect identifiers should be statistically protected (such as detailed geographic location, detailed description of work place, precise dates).

By following the aforementioned safeguards both user and participant can be confident that data is not disclose while still remaining useful for the majority of research purposes. If needed, the researcher could prepare two data files: one with a lower level of data protection to be used for advanced scientific use and available to registered researchers under more strict use conditions; and the second one with a higher level of protection, suitable for a wider range of users.

2.1.3 Digital curation

Digital curation represents a challenge for data creators to preserve the authenticity and provenance of research data during the project and after it ends. That is to assure permanent usability of data and transparency of the data generation process. Some practical recommendations for responsible data management throughout the project are presented below:

Thinking about appropriate organization of materials:	Systematic and intuitive naming and re-naming of files and variables.
	Data formats and software used.
	File transfers, file sharing and remote access.
	File version control.
Actions:	Manage backups.
	Prepare contextual documentation and metadata.
	Define access conditions for digital materials.
	Make appropriate provisions to prevent unauthorized access. ¹⁵

¹⁵ UK Data Service (2014): Benefits of managing and sharing research data. Colchester: University of Essex. [<http://ukdataservice.ac.uk/media/440285/whisharedata.pdf>].

2.1.4 Supporting research data management planning

Planning and preparing data for open access consists of a range of research data management processes as detailed above to maximise data quality, working in accordance to ethical requirements and effective digital curation procedures. Data management plans (DMP) are fast becoming established as the formal obligation in response to funder or research institution requirements. They aim to deal with all aspects in relation to research data throughout the project and their eventual deposition in open access solutions at the end of the project. Support services are available to help design the DMP and to implement it. The primary reference point for support would be the domain data centre, archive or data repository, supplemented with support from a range of stakeholders including:

Research institution:	<ul style="list-style-type: none">• Assures and manages internal open access policies and designs procedures to help research projects with the preparation and implementation of DMPs.• Manages expert training for researchers and support staff.• Provides infrastructure in the form of technological and advisory services to provide digital preservation for data throughout the lifecycle (research offices, libraries, CSIC,¹⁶ SICRIS,¹⁷ operating data centre and service networks).• Provides common services and tools to support research groups.
Library:	<ul style="list-style-type: none">• Provides information about the availability of existing data sources.• Provides information about options to deposit data in a data centre or data archive.• Helps select an appropriate or recommended data centre or data archive.• Provides information about open access conditions and advantages.• Supports preparation of formal DMPs.• Provides support with preparation of basic study metadata and documentation, author's rights, and explains other deposition requirements.
Funder:	<ul style="list-style-type: none">• Manages national/disciplinary policies which require the preparation of DMPs as part of the research project application process, based on the principles of open access to research data financed by public funds, perceived as being a public good.¹⁸• Acknowledges costs for open access data and metadata preparation and provides funds to cover them.• Monitors implementation of open access obligations.

16 Central specialised information centres: Research infrastructure. [<https://www.arrs.gov.si/en/infra/osaic/predstavitev.asp>, 21. 1. 2015].

17 Slovenian Current Research Information System. [<http://www.sicris.si/public/jqm/cris.aspx?lang=eng&opdescr=home>, 21. 1. 2015].

18 DCC: Overview of funders' data policies. [<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>, 21. 1. 2015].

European Commission (2013): Guidelines on Data Management in Horizon 2020. [http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-datamgt_en.pdf].

2.2 Depositing data with a data centre for preservation

2.2.1 WHERE: selection of data centre

The decision about choosing which data centre or archive to use for the deposit and future availability of research data should be made in the planning phase. It makes sense to contact a data centre at the beginning of the project and to become familiar with the requirements that need to be met for acquisition. Formal DMP forms, required by most funders, include an item about the place of deposit. Some funders have the recommended place of deposit already set in their policies. Usually this is the national disciplinary data centre.¹⁹ Projects funded by Horizon 2020 also encourage participants of the Open Research Data Pilot to offer their data to an established research data centre.²⁰

When choosing an appropriate data centre the acquisition policy, requirements of the chosen data service provider, and the advantages of depositing data should be understood. Attention should be paid to the following detail:

- How is the mission of the data centre specified? How is the user community specified?
- What kind of data are acquired?
- Is the centre actively involved in digital curation and does it provide data access to a specified user community?
- What are the criteria for selection and how is the selection and acquisition process specified? What are the technical characteristic requirements?
- How do they solve legal and ethical issues related to privacy protection and Intellectual property rights?

¹⁹ E.g. ESRC from Great Britain recommends UK DS, which is a partner organization to the Slovenian ADP.

Economic & Social Research Council (2013): ESRC Research Data Policy. [http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf], p. 6.

²⁰ E.g. catalogues of the following data centres could be of some help: Registry of Research Data Repositories [<http://www.re3data.org/>, 21. 1. 2015] and Databib [<http://databib.org/>, 21. 1. 2015].

2.2.2 WHY choose ADP?

The disciplinary data centre services for Social Sciences are in Slovenia provided by the Social Science Data Archives (ADP—Arhiv družboslovnih podatkov). The advantages of depositing data to the ADP are, among others, the following:

ADP evaluates the importance of research data for science and their long-term usability.	Acquired data are acknowledged as a scientific publication in their own right and represent a basis for quantitative assessments of scientific excellence in accordance to the criteria of the Slovenian Research Agency. ²¹
Adopts an acknowledged digital curation approach.	ADP follows the approach of data centres in the field of long-term preservation of data, in accordance with the international reference model OAIS ²² and principles of DSA. ²³
Provides access to data and enables searching and browsing through standard data descriptions for the purposes of discovery.	ADP provides advice and training to their users. Data made available in the catalogue represent a basis for citation in reference lists. Registered users are obliged to make reference to the data in their publication. Users are instructed about how to cite data at the point of access.
Offers support with data management planning and preparation of data for open access.	ADP is actively involved both nationally and internationally in the field of opening up research data. Based on the experience and demands, ADP offers RDM counselling and assists users with preparing data for open access.

21 ARRS (2014): Rules on the Procedures of the (co)financing and Monitoring of Research Activities Implementation. [<https://www.arrs.gov.si/en/akti/prav-sof-ocen-sprem-razisk-dej-sept-11.asp>].

22 OAIS is an ISO-standard which provides a functional frame for preserving digital objects in data centres. See: OAIS-Based Processes [<http://www.icpsr.umich.edu/icpsr-web/content/datamanagement/lifecycle/oais.html>], 21. 1. 2015].

23 Data Seal of Approval: The Guidelines 2014 – 2015. [<http://www.datasealofapproval.org/en/information/guidelines/>], 21. 1. 2015].

2.2.3 HOW to deposit research data

Depositing research data with ADP is not necessarily a singular activity, but normally involves the mutual exchange of information through negotiation about those details required to fulfil the data acquisition process. ADP provides counselling, instructions and tools which make data preparation and deposition processes easier. Researchers who want to store their research data in the ADP should follow the steps listed below:

<p>1) Make sure that the data satisfy the criteria for acquisition.²⁴</p>	<ul style="list-style-type: none"> · The richness of content in terms of adequacy of conceptualization, suitability and potential for use, and thematic broadening of the ADP catalogue should be evaluated. · It is important that methodologies are transparent and of the highest quality, accompanied by complete and valid data and documentation to enable further analysis. · The depositor, as a copyright owner, is willing to deposit data with ADP for dissemination. 	<p>24 ADP: ADP study classification by relevance. [http://www.adp.fdv.uni-lj.si/eng/za_uporabnike/o_arhiviranju/klasifikacija_adp_pomembnosti/, 21. 1. 2015].</p>
<p>2) Researchers can propose data for deposition using the »Acquisition Proposal«.²⁵</p>	<p>The entry form is available for short descriptions of study content and its methodological characteristics. The information provided enables an assessment of the quality of data and their suitability for the acquisition.</p>	<p>25 ADP: Acquisition Proposal. [http://www.adp.fdv.uni-lj.si/eng/evidentiranje/, 21. 1. 2015].</p> <p>26 ADP: License Agreement. [http://www.adp.fdv.uni-lj.si/eng/za_dajalce/izjava_o_izrocitvi/, 21. 1. 2015].</p>
<p>3) The depositor fills in the »Licence Agreement«,²⁶ in which deposited materials are listed and data access conditions specified.</p>	<ul style="list-style-type: none"> · The researcher who has author's rights to the data approves the deposition with the Archives, devolving future responsible management of data and their onward distribution. · He/she guarantees the authenticity and the legal rights of the data (i.e. intellectual property rights, privacy protection). · By default ADP gives the depositor an option to choose between two licenses: Creative Commons Attribution alone (CC BY) and Creative Commons Attribution + Non-commercial (CC BY-NC). · The depositor specifies any exceptions in access conditions (certain research data could have limited or controlled access, e.g. sensitive data, restricting access for a certain reasonable period of time). 	<p>27 Study description could be prepared in a common text editor or using a special tool. For that purpose ADP suggests using Nesstar Publisher tool using the pre-prepared template.</p> <p>ADP: Study Description Form. [http://www.adp.fdv.uni-lj.si/eng/o_arhivu/interaktivno/obrazec_opis_raziskave/, 21. 1. 2015].</p> <p>Nesstar Publisher [http://www.nesstar.com/software/publisher.html, 21. 1. 2015].</p>
<p>4) The researchers fill in the »Study Description Form«²⁷ with the following information:</p>	<ul style="list-style-type: none"> · Data creation and data collector. · Funder of the project (public or private, including project or grant ID number). · Detailed description of the research including methodology (sample, population, data collection time frame, response rate etc.). 	<p>ADP: ADP's template for Nesstar Publisher. Training. [http://www.adp.fdv.uni-lj.si/za_uporabnike/uposabljanje/, 21. 1. 2015].</p>

5) The researcher prepares clear and concise documentation and any other ancillary materials about the data for deposition in the Archive.

- The data should be provided in electronic form accompanied by descriptive information such as file size, format, number of variables and units, code definitions etc.²⁸
- The materials should be prepared respecting archival instructions and addressing recommendations about formats.
- The depositor should guarantee provisions have been taken to protect personal data. All direct identifiers should be removed.
- Materials which help understand and verify the content of data should also be attached: e.g. codebooks, frequency files, instructions for interviewers, information about the implementation of the study, links to study reports or their copies if not being publically available. Both electronic and printed versions of the questionnaire in the original form should be attached.

2.2.4 Research data acquisition stage

Based on information provided and a review of offered materials ADP will evaluate and assess if data are suitable for acquisition. Emphasis will be placed on their scientific value as well as the future usability for different purposes and potential users.

The following is of crucial importance:

- Topical relevance, completeness of data, high number of variables, quality of methodology.
- Study as a part of series, international comparability, potential of linking with other data and the existence of harmonized standard demographic variables.
- In addition, both depth and exhaustiveness of interviews are important in the case of qualitative data.

²⁸ Lužar, Sanja, Maja Ojsteršek in Irena Vipavc Brvar (2012): Priporočila za urejanje podatkovne datoteke [Guidelines for editing data file]. ADP. [<http://www.adp.fdv.uni-lj.si/blog/wp-content/uploads/2012/11/PriporocilaZaPodatkovnoDatoteko2.pdf>].

²⁹ Compare with: NERC Data Value Checklist. [<http://www.nerc.ac.uk/research/sites/data/policy/data-value-checklist.pdf>, 21. 1. 2015].

No single criterion is decisive. The evaluation of fitness for further use is crucial and is also based on the rarity of either data type or topic in the existing catalogue of a data centre. Formal criteria such as formats, appropriate data documentation etc. are obvious preconditions for data re-usage, without which data could be useless for future users. Completeness of data, supporting documentation and provenance, and any recorded changes are important for continuity of digital preservation which is taken over by an archive.²⁹

In the case of a positive evaluation, ADP and the depositor agree on the acquisition and then make necessary provisions for distribution. This includes editing documentation and data in accordance with international standards, and preparing the final version of the dataset and the study description. As a final step, the researcher approves that the data and accompanying material is valid before it is published in the catalogue.

2.3 Data publication

2.3.1 Access to research data

Data centres and archives process deposited research data, prepare them for long-term preservation, publish them in the catalogue together with metadata and documentation, thus making the data both findable and accessible. Researchers and other users can familiarize with the content of data and their generation process through the “Study description” document which accompanies data publication. They are also provided with the information about the access conditions.

Users are provided with the following services while accessing data:

Options to search and familiarize	
ADP publishes data in a list of studies ³⁰ on the website, where it is possible to:	<ul style="list-style-type: none">· Browse through topics, series, authors etc.· Access shorter study descriptions, datasets, question texts and related materials.· Find a link to access data.
Parallel to the publication on the website, information about the data as well as the microdata themselves are available in the Nesstar catalogue of ADP. Additionally, it provides:	<ul style="list-style-type: none">· Full study descriptions using the DDI metadata standard.· Simple or advanced searching through individual elements of study descriptions including question texts, and searching for comparable data across studies for the purpose of longitudinal research.· Online analysis of data.· Access to microdata for download and further analysis.
Metadata about studies are included in catalogue aggregators:	<ul style="list-style-type: none">· Information about data publication can be entered into the shared catalogue of Slovenian libraries – COBIB.SI and the Open Science national portal.³¹· When a study is entered into the COBIB.SI system, it is displayed in the researcher's bibliography and thus contributing points to the scoring of scientific excellence as a recognised scientific data publication.· It is also possible to search the CESSDA Data Catalogue, which is a common catalogue of European Social Science Data Archives.³²

30 ADP: The list of studies by study's ID [<http://www.adp.fdv.uni-lj.si/opi-si/>, 21. 1. 2015].

31 Open Science Slovenia [<http://openscience.si/>, 21. 1. 2015].

32 CESSDA: The CESSDA Catalogue. [<http://www.cessda.net/catalogue/>, 21. 1. 2015].

Respecting access conditions	
As a precondition of access each user is required to register and define the purpose for which the data will be used:	<ul style="list-style-type: none"> · The archive distinguishes between using data for academic and using data for public (including commercial) purposes. · Among academic users there's further distinction between registered researchers (based on SICRIS data) and other users who use data mainly for educational purposes. · Additionally, access can be controlled by exceptions defined by the depositor at the time of deposition to the ADP (e.g. access with a special regime of confidentiality, only with author's agreement for a certain period of time).
Reporting data use	
Data centres analyse access statistics for the purposes of reporting and for planning of promotional and educational activities:	<ul style="list-style-type: none"> · Website traffic (e.g. measured via Google analytics). · Access to metadata and related materials. · Online analysis and data downloads.

2.3.2 Advantages of acknowledging data as a scientific publication

Published data are appraised in the same way as any other scientific publication. They represent a »first class information object within scholarly information systems« and the following advantages originate from that:

Publication is uniquely identified	Published data (and all ensuing versions) have a unique identifier which is designated by the data centre. At an international level, similarly to regular monographic publications and articles in journals, different forms of permanent identifiers have been introduced for microdata, such as: DOI, ³³ URN etc. Permanent identifiers, if explicitly and appropriately listed, represent a discrete connection between publication and underlying data.
Citing data used	Upon registration, in published study descriptions, in data workshop materials and in user guides, all users are reminded to cite research data and related materials. ³⁴ It is recommended that data be cited in journal article reference lists with a full reference to the author, the title, the place of data access, the permanent identifier, according to citation rules as detailed by each scientific journal.

³³ Ball, Alex and Monica Duke (2012): How to Cite Datasets and Link to Publications, Edinburgh: DCC. [http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_to_Cite_Link.pdf].

³⁴ ADP: Predavanja in predstavitve ADP. [http://www.adp.fdv.uni-lj.si/publikacije_adp/predavanja/], 21. 1. 2015].

Data are acknowledged as a publication and are therefore evaluated and evidenced in the bibliography of researcher's institutions	The author or a group of authors could receive points to the scoring of scientific excellence based on the publication of data in the ADP catalogue. Deposited data of scientific significance and of potential interest for further use could be entered into COBIB.SI ³⁵ system after being published by the ADP. That could also contribute to the scoring for the evaluation of project effectiveness.
Citations of other data-related publications are increased	Open access publication of data increases the number of citations and results in the favourable reception of articles written by the original authors of the research. ³⁶ Data publication is accompanied by authors' own publications and therefore increases the probability of other users who may consider using the data for their own analysis. At the same time, an article in which analysed data are referenced becomes more reliable through verification and promotes further analysis of the same data.
Citation indexes	Citing data used in reference lists increases traceability of use and visibility of existing data sources. Similar to citation indexes for traditional publications, data citation indexes for research data will become important for measuring efficiency and impact. Those effects will be realised through a consistent approach to citing research data use (e.g. by using permanent identifiers). Citing data in reference lists of publications also facilitates verification of the long-term importance of data for science. ³⁷

35 The entry is classified under the category 2.20 *Complete Scientific Database or Corpus*, defined as: »An electronic data collection, the scientific relevance of which is demonstrated by the use for the purpose of researching a wide range of theoretical and applied problems. The data collection must be the outcome of an accomplished research and comply with high quality standards. The quality is assessed on the basis of the detailed accompanying documentation. The data collection must be publicly available in the national or international scientific data archives. The collection must be documented and available in a form that allows the repetition of published scientific findings made on its basis.« See: IZUM (2013): Typology of documents/works for bibliography management in COBISS. Maribor: IZUM. [http://home.izum.si/COBISS/bibliografije/Tipologija_eng.pdf].

36 Costas, Rodrigo, Ingeborg Meijer, Zohreh Zahedi and Paul Wouters (2013): The Value of Research Data - Metrics for datasets from a cultural and technical point of view. Copenhagen: Knowledge Exchange. [www.knowledge-exchange.info/datametrics].

37 See discussion on: Whyte, Angus and Andrew Wilson (2010): How to Appraise and Select Research Data for Curation. DCC. Research Data. Advice paper, no. 14. [<http://www.dcc.ac.uk/resources/how-guides/appraise-select-data#sthash.4WjHGwZ1.dpuf>].

2.3.3 How to participate in promoting data use

Along with data centres and archives, the following stakeholders can participate in promoting and exploiting access to research data:

Researcher:	<ul style="list-style-type: none">· Includes information about a data publication in researcher's own bibliographic reports, and lists these as evidence of scientific integrity.· When publishing reports, the researcher cites his or her data with a full reference in bibliographic reference lists.· Participates in promoting data usage through training when data are complex and difficult to use (e.g. organization of workshops, seminars or conferences covering aspects of working with data).
Research institution:	<ul style="list-style-type: none">· Reports statistics about their own data, which are available in data centres, and their use to add into the excellence profile.³⁸· Offers research data management support, stimulates data deposition by acknowledging it as part of scientific career progression.· Promotes data reuse.
Library:	<ul style="list-style-type: none">· Provides information about accessioning datasets into a bibliographic system for the purpose of scientific evaluation and resource discovery.· Provides data citation guidance.· Provides information on datasets, modes of access and possible use at information literacy training.· Provides information about linking data with publications and research projects.
Funder:	<ul style="list-style-type: none">· Monitors if data are openly available as specified in the research project contract and advises about possible consequences if not available.· Assures that properly published research data are acknowledged as a contribution to science in the framework of a national system of scientific evaluation.· Stimulates the creation of high quality data and their deposition in open access solutions, introduces sanctions for controlling access.· Provides conditions for sustainable data infrastructure and services.· Promotes reuse of data in their policies.
Common tasks:	<ul style="list-style-type: none">· Coordinating development of supporting infrastructure and services aimed at helping researchers, and to decrease the burden of adhering to additional open access requirements.³⁹· Training current and future researchers and supporting personnel, data librarians, data scientists etc.⁴⁰

³⁸ LERU Research Data Working Group (2013): LERU Roadmap for Research Data. Advice paper, no. 14. [http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf], p. 21.

³⁹ Ibid.

⁴⁰ FOSTER [<https://www.fostero-penscience.eu/>], 21. 1. 2015].

3 FURTHER SUPPORT

Dealing with research data throughout the data lifecycle involves being familiar with a wide spectrum of standards and good practices. Recently, in order to fulfil funders requirements researchers are increasingly being asked to complete formal data management.

To further support planning and working with data, the following are available:	<ul style="list-style-type: none">· Guides and tools, prepared by expert organisations, such as DCC.⁴¹· Online tools such as DMPonline⁴² or DMP Editor.⁴³· Online research data management training resources, e.g. MANTRA.⁴⁴
Data centres also provide guidance about:	<ul style="list-style-type: none">· Methodology – about exploiting existing data and about professional standards related to data collection methods.· Technical aspects – data and metadata formats, normalisation and migration, documenting, versioning, digital curation.· Legal perspectives – agreements, privacy protection, intellectual property rights and licenses, ethics.

ADP provides detailed recommendations about depositing and archiving procedures on their website section “About archiving”.⁴⁵ We recommend using guidelines from similar data centres as well.⁴⁶ Familiarization with DSA guidelines⁴⁷ is also recommended in order to achieve a coordinated and responsible approach to research data management with stakeholders involved throughout the whole data lifecycle, from planning and creating, digital preservation, access in data centres and to the final use.

41 Digital Curation Centre: Resources for digital curators. [<http://www.dcc.ac.uk/resources>, 20. 1. 2015].

42 Digital Curation Centre: DMPonline. [<https://dmponline.dcc.ac.uk/>, 20. 1. 2015].

43 OpenMetadata.org: Data Management Plan (DMP) Editor. OpenMetadata.org. [http://www.openmetadata.org/site/?page_id=373, 21. 1. 2015].

44 EDINA (2014): MANTRA – Research Data Management Training. The University of Edinburgh. [<http://datalib.edina.ac.uk/mantra/>].

45 ADP: About archiving. [http://www.adp.fdv.uni-lj.si/za_uporabnike/o_arhiviranju/, 21. 1. 2015].

46 UK Data Service: Prepare and manage data. [<http://ukdataservice.ac.uk/manage-data.aspx>, 21. 1. 2015].

ICPSR: Data Management & Curation. [<http://www.icpsr.umich.edu/icpsrweb/content/data-management/index.html>, 21. 1. 2015].

Tjalsma, Heiko and Jeroen Rombouts (2011): Selection of research data, Guidelines for appraising and selecting research data. Haag in Delft: Stichting SURF, Data Archiving and Networked Services (DANS), 3TU.Datacentrum. [<http://act.dans.knaw.nl/nl/over/organisatie-beleid/Publicaties/DANSselectionofresearchdata.pdf>].

ICPSR and DANS (2010): Preparing data for sharing; Guide to Social Science data archiving. Amsterdam: Pallas Publications. [<http://act.dans.knaw.nl/nl/over/organisatie-beleid/Publicaties/DANSpreparingdataforsharing.pdf>].

47 Data Seal of Approval: The Guidelines 2014-2015. [<http://www.datasealofapproval.org/en/information/guidelines/>, 21. 1. 2015].

Appendix: The DSA Guidelines 2014-2015

Guidelines Relating to Data Producers:

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.
2. The data producer provides the data in formats recommended by the data repository.
3. The data producer provides the data together with the metadata requested by the data repository.

Guidelines Related to Repositories:

4. The data repository has an explicit mission in the area of digital archiving and promulgates it.
5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.
6. The data repository applies documented processes and procedures for managing data storage.
7. The data repository has a plan for long-term preservation of its digital assets.
8. Archiving takes place according to explicit work flows across the data life cycle.
9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.
10. The data repository enables the users to discover and use the data and refer to them in a persistent way.
11. The data repository ensures the integrity of the digital objects and the metadata.
12. The data repository ensures the authenticity of the digital objects and the metadata.
13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

Guidelines Related to Data Consumers:

14. The data consumer complies with access regulations set by the data repository.
15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.
16. The data consumer respects the applicable licences of the data repository regarding the use of the data.⁴⁸

⁴⁸ Data Seal of Approval: The Guidelines 2014-2015. [<http://www.data-sealofapproval.org/en/information/guidelines/>, 21. 1. 2015].

DEPOSITING RESEARCH DATA WITH ADP

ADP accepts quality research data with high reuse potential that is well-organised, properly prepared and documented for further analysis. The process of depositing research data consists of the following steps:

1. Check the ADP criteria for the acquisition
2. Fill in the Acquisition Proposal Form
3. Fill in the Study Description Form
4. Edit and document data, prepare accompanying materials
5. Fill in and sign the Licence Agreement Form.

www.adp.fdv.uni-lj.si/eng/za_dajalce/

ADP provides the following services to data depositors:

- support with data management planning and preparing data for open access,
- verifying and evaluating the importance of research data for science and its long-term usability,
- introducing digital curation processes and workflows,
- enabling discovery of and access to data, searching and browsing through study descriptions,
- promotion of data use and, formal training about how to work with data in collaboration with partners.

Univerza v Ljubljani



CONTACT:

University of Ljubljana
Faculty of Social Sciences
Social Science Data Archives
Kardeljeva ploščad 5
SI-1000 Ljubljana

www.adp.fdv.uni-lj.si 

arhiv.podatkov@fdv.uni-lj.si 

Arhiv.Druzboslovnih.Podatkov 

@ArhivPodatkov 