

Research Data Management Planning: problems and solutions

Workshop

Irena Vipavc Brvar, Sonja Bezjak, ADP

12th Applied Statistics 2015, Ribno (Bled), 20th September 2015

- Open Science **Principles**
- **Open Access** and **Open Data**
- **Data Management Planning**
 1. Data Collection
 2. Documentation and metadata
 3. Ethics and legal compliance
 4. Storage and Backup
 5. Selection and Preservation
 6. Data SharingResponsibilities and Resources
- HANDS-ON – Preparing DMP for my research

"Open Data Excuse" Bingo

#openDataExcuses

Lawyers want a custom License	Data Protection	Poor Quality	People may misinterpret the data
It's too big	There's no API	It's too complicated	Thieves will use it
It's not very interesting.	There's already a project to...	We will get too many enquiries	What if we want to sell it later
We'll get spam	We might want to use it in a paper	I don't mind, but someone else might	Terrorists will use it

For open data teams; print out a copy and put it on your office wall. Cross out each excuse people give you. There are no prizes, but you can tweet "bingo! #openDataExcuses" if you think it might make you feel better*.

* it won't

Generate your own bingo grids at <http://data.dev8d.org/devbingo/>

Open Science Principles

Research publications and research data needs to be openly accessible:

- to be able to **check the validity** of published findings / better monitoring, assesment and evaluation of research
- support **multiple findings** (using different theoretical starting points, different analytical approaches) based on the existing knowledge.

Open science includes:

- Open Access to Literature from Funded Research
- Access to Research Tools from Funded Research
- Data from Funded Research in the Public Domain
- Invest in Open Cyberinfrastructure

(Science Commons, <http://sciencecommons.org/resources/readingroom/principles-for-open-science/>)

Open Access - definition

„Open Access is the immediate, online, free availability of research outputs without restrictions on use commonly imposed by publisher copyright agreements. Open Access includes the outputs that scholars normally give away for free for publication; it includes peer-reviewed journal articles, conference papers and datasets of various kinds.“

Some advantages of OA:

- Access can be greatly improved (basis for the transfer of knowledge: teaching/research/civil society)
- Increased visibility and higher citation rates (up to 3 times higher citation rates)
- Free Access to information (enabling people from poorer countries to access).

(source: <https://www.openaire.eu/oa-overview>)



Why sharing data?

„Research data are a valuable resource, usually requiring much time and money to be produced. Many data have a significant value beyond usage for the original research.“

- encourages scientific **enquiry and debate**
- promotes innovation and potential **new data uses**
- leads to **new collaborations** between data users and data creators
- maximises **transparency and accountability**
- enables scrutiny of research findings
- encourages the improvement and validation of research methods
- **reduces the cost** of duplicating data collection
- **increases the impact** and visibility of research
- promotes the research that created the data and its outcomes
- can **provide a direct credit** to the researcher as a research output in its own right
- provides **important resources** for education and training.

Documents supporting OA

OECD:

[OECD Principles and Guidelines for Access to Research Data from Public Funding](#) (2007)

EU Council:

[Council Conclusions on scientific information in the digital age: access, dissemination and preservation](#) (2007)

EU Commission:

[Commission Recommendation of 17. 7. 2012 on access to and preservation of scientific information](#) (17.7.2012)

Slovenia:

[National strategy of open access to scientific publications and research data in Slovenia 2015-2020](#) (3.9.2015)

Open Access and Open Data Policies

European Union level

“**Member states** are invited to:

- Harmonise access and usage policies for research and education-related public e-infrastructures ...

Research stakeholder organisations are invited to:

- Adopt and implement open access measures for publications and data resulting from publicly funded research”

(Reinforced European Research Area Partnership for Excellence and Growth COM(2012) 392 final http://ec.europa.eu/euraxess/pdf/research_policies/era-communication_en.pdf)

National level: Slovenian case

[National strategy of open access to scientific publications and research data in Slovenia 2015-2020](#), adopted on the 3rd September 2015

OA documents adopted by several EU Member States

- Spain (the law): [Recommendations for the implementation of Article 37 of the Spanish Science, Technology and Innovation Act: Open Access Dissemination](#),
- Belgium: [Brussels declaration on open access to Belgian publicly funded research](#),
- Ireland: [Ireland: the transition to open access](#),
- Portugal (national policies): [Portugal open access policy landscape](#),
- Denmark: [Denmark's national strategy for Open Access](#),
- Sweden: [Proposal for national guidelines for open access to scientific information](#),
- Austria: [New Policy for Open Access and Publication Costs](#),
- Norway: [Education, research and open access in Norway](#)

OA Policies: Funders

- **UK:** Six of the seven RCUK funders require data management plans, or equivalent, at the application stage, as do Wellcome & CRUK
- **USA:** White House Office of Science and Technology Policy requirement for DMPs announced March 2013 (programmes awarding >\$100m annually)
- **Argentina:** all institutions belonging to the National Science and Technology System that receive public funds (partly or entirely) shall create free and open access institutional digital repositories where all the scientific and technological publications (which includes journal articles, technical and scientific papers, theses, etc.) and research data must be available (from Nov 2013).

OA Policies: Journals

“An inherent principle of publication is that others should be able to **replicate and build upon** the authors' published claims. A condition of publication in a Nature journal is that authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications.” [Nature](#)

„PLOS journals require authors **to make all data underlying the findings described in their manuscript fully available** without restriction, with rare exception.

When submitting a manuscript online, authors must provide a *Data Availability Statement* describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the final article.

Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection.

[Plos](#); / [Recomended Repositories](#)

Citation: Cameron TL, Bell KM, Gresshoff IL, Sampum L, Mullan L, Ermann J, et al. (2015) XBP1-Independent UPR Pathways Suppress C/EBP- β Mediated Chondrocyte Differentiation in ER-Stress Related Skeletal Disease. PLoS Genet 11(9): e1005505. doi:10.1371/journal.pgen.1005505

Editor: Gregory S. Barsh, Stanford University School of Medicine, United States of America

Received: April 8, 2015

Accepted: August 14, 2015

Published: September 15, 2015

Copyright: © 2015 Cameron et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The full datasets are available from the NCBI Gene Expression Omnibus (GEO), accession number GSE72261 (<http://www.ncbi.nlm.nih.gov/geo/>).

Funding: This work was funded by the National Health and Medical Research Council of Australia grant #607398 (JFB), the Victorian Government's

Data Availability Statement: The full datasets are available from the NCBI Gene Expression Omnibus (GEO), accession number GSE72261 (<http://www.ncbi.nlm.nih.gov/geo/>).

Data Availability Statement: Data are available in the LSHTM Data Compass repository, <http://dx.doi.org/10.17037/DATA.4>.

Open Access

DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS

10,529 Journals
6,439 searchable at Article level
134 Countries
2,084,119 Articles



OpenDOAR

The Directory of Open Access Repositories - *OpenDOAR*

- Deposit your research data at the same time as publication
- Repositories might be institutional or subject based
- They curate data and provide identifiers (DOI, URN) → data can be cited

Horizon 2020

- Mandatory immediate open access to all peer-reviewed **publications** (up to 12 months delay for SSH)
- Mandatory open access **to research data** for some areas (Open Access Pilot for 2014-2015)-Data Management Plans

Open Access activities in H2020

- [Fact sheet: Open Access in Horizon 2020](#), 9. dec. 2013
- [Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020](#), 11. dec. 2013
- [Guidelines on Data Management in Horizon 2020](#), 11. dec. 2013

Pilot on Open Research Data (H2020)

Areas of the 2014-2015 Work Programme participating in the Open Research Data Pilot are:

- Future and Emerging Technologies
- Research infrastructures – part e-Infrastructures
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Societal Challenge: Secure, Clean and Efficient Energy – part Smart cities and communities
- Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – except raw materials
- Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies
- Science with and for Society

Projects in other areas can participate on a voluntary basis.

What is Research data management plan?

„A further new element in Horizon 2020 is the use of **Data Management Plans** (DMPs) detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. „ (

Guidelines on Data Management in Horizon 2020, 2013, p. 3)

H2020 DMP should address the following points:

- Data set reference and name
- Data set description
- Standards and metadata
- Data sharing
- Archiving and preservation (including storage and backup)

(Guidelines on Data Management in Horizon 2020, p. 5)

But data archiving needs more information ☺

Research Data Management definitions

„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.“

(Štebe J. et al: Preparing research data for open access : guide for data producers, 2015, p. 3)

“An explicit process, covering the creation and stewardship of research materials to enable their use for as long as they retain value” (DCC)

„The activities of data policies, data planning, data element standardization, information management control, data synchronization, data sharing, and database development, including practices and projects that acquire, control, protect, deliver and enhance the value of data and information.“

(http://dictionary.casrai.org/Data_management)

Data management in H2020 (EU funder)

Data Management Plans (DMPs) **mandatory for all projects participating in the pilot** (deliverable within the first six months)

Other projects invited to submit a DMP if relevant for their planned research

DMP questions:

- What data will be collected / generated?
- What standards will be used / how will metadata be generated?
- What data will be exploited? What data will be shared/made open?
- How will data be curated and preserved?

([Guidelines on Data Management in Horizon 2020](#))

“ All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication. ”

— University of Edinburgh Research Data Management Policy

WHERE DO YOU STORE
YOUR RESEARCH DATA?

- ☐ USB DRIVE
- ☐ DROPBOX
- ☒ RESEARCH DATA REPOSITORY

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Data centres / repositories policies

When choosing an appropriate data centre check its:

- acquisition policy
- general and special requirements
- advantages of depositing data



Search for Repositories (0 Reviewed Repositories)

The image shows the search interface of re3data.org. At the top is a long search bar with a blue 'Search' button on the right. Below the search bar is a filter section with three columns: 'Subject', 'Content Type', and 'Country (of the responsible institutions)'. Each column has a dropdown menu with the text 'Add subjects', 'Add content types', and 'Add countries' respectively. Below these dropdowns is a blue button labeled 'Raw data'. At the bottom of the filter section, there are three checkboxes: 'Certificates' (checked), 'Open Access' (checked), and 'Persistent Identifier' (checked). A red 'Reset filter' button is located at the bottom right of the filter section.

Social Science Data Archives, UL

- from 1997 / national data repository for social sciences
- 600 social science surveys
- depositors from all 4 (3 public) universities, private research centres, Statistical Office of Slovenia (8-10 research centres per year)
- cca. 700 users yearly (90 % education, 10 % scientific/research purpose)
- Using standard metadata description format (DDI) -> makes interoperability with EU data archives / repositories possible

ESSDA

of Social Sciences working abroad, 2013
DLG - Democracy and Local Governance
[DRPREH81]² Social catering, 1981 - 1983
[DRV76]¹ Longterm development of university education in SR Slovenia : Generation 1976 Graduates
DZRS - National Assembly of the Republic of Slovenia
EB - Eurobarometer
EES - European Election Study
[EGOOMR00]² Quality of measurement of egocentered social networks
[EKOMIG10]³ Survey on Economic Migrations and Migration Workers
[ELCZSE96]¹ 24 hours

DESCRIPTION TABULATION ANALYSIS

Dataset: [EUVET12]³ 7 EU VET - Study on vocational education in seven european countries

Abstract

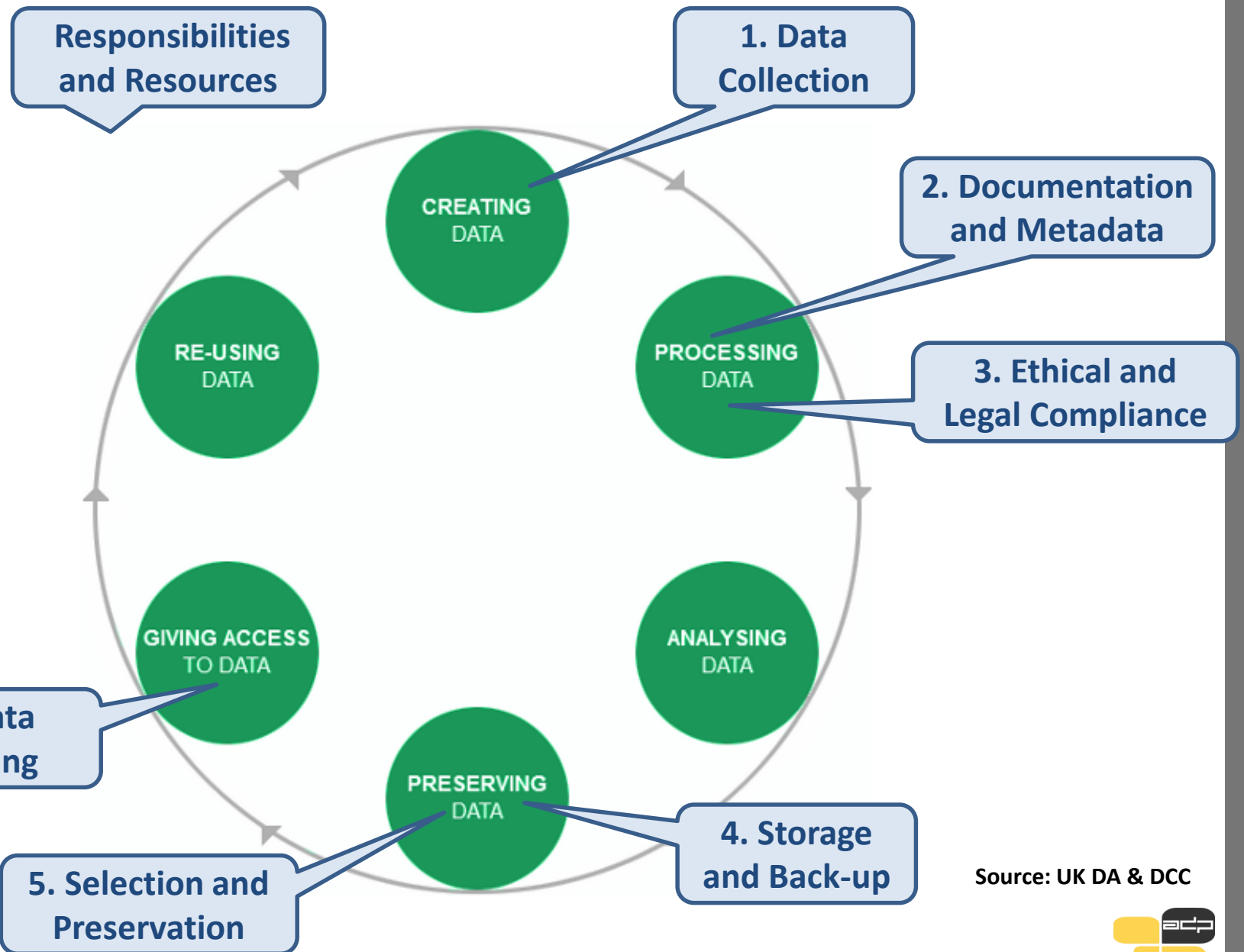
The 7EU - VET project - Detailed Methodological Approach to Understanding the VET Education - is a research study on vocational education and training which builds on theoretical backgrounds and secondary analyses of the existing documentation as well as on national and EU data in order to conduct quantitative and qualitative studies and derive empirical results. The project is built upon one of the goals of the Lisbon strategy, which is the promotion and the quality of vocational education and training. The project addresses the following scientific and research objectives: Are educational and training systems flexible enough to adequately respond to different and changing needs? How efficient and successful are the systems (for advising and informing) and can we set up a successful model which could be implemented across the countries? What are the perceptions of the VET systems by the young people and what is their view of their future possibilities for career building and mobility? In what way are the ICT solutions implemented in vocational education programs? For international comparison use data of 17 and 18 year old pupils.

LITERAL QUESTION

D1. How far do you agree with following statements concerning your professional and life goals? (Please tick one box in each row) I strive towards ... Receiving a high income

Values	Categories	N	
1	Not at all	511	2.9%
2	Slightly	1321	7.6%
3	Fairly	3590	20.7%
4	Quite	5515	31.8%
5	Completely	6418	37.0%
-88	not answered	272	
-77	not applicable	0	

Research Data Lifecycle



Source: UK DA & DCC

1. Data Collection

Researcher:

- Start research planning
- Check data collections
- Locate existing data
- Check policies: institutional, state, EU, disciplinary
- Check Funder's requirements
- Design research
- Develop / adjust RDMP
- Develop consent form
- Collect data
- Capture metadata

Support:

- check ADP's collection of questionnaires and other materials useful for new surveys (export in DDI possible – direct use in survey tools – Blaise, *1ka*),
- contact disciplinary repository, attend workshops...
- consult with research office at your institution
- attend workshops on RDM; on consent form and other ethical obligations / Open Data.

1. Data Collection: relevant questions

1.1 What data will you collect or create?

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

1.2 How will the data be collected or created?

- What standards or methodologies will you use?
- How will you structure and name your folders and files?
- How will you handle versioning?
- What quality assurance processes will you adopt?

2. Documentation and metadata

Researcher:

- Enter data: digitize, transcribe, translate...
- Check, validate, clean
- Anonymize
- Describe
- Store

Support:

- Provide guidelines on Data file (formats, organizations, storage)
- Guidance on Anonymisation and tools
- Which accompanying materials to save and how
- Temporary university repository (UL, MB, UP)
- Inclusion in study process – lectures about processing of data.

2. Documentation and metadata: relevant questions

What documentation and metadata will accompany the data?

- What information is needed for the data to be to be read and interpreted in the future?
- How will you capture / create this documentation and metadata?
- What metadata standards will you use and why?



The Licenses



Attribution CC BY

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

[View License Deed](#) | [View Legal Code](#)



Attribution-ShareAlike CC BY-SA

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

[View License Deed](#) | [View Legal Code](#)



Attribution-NonCommercial CC BY-NC

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

[View License Deed](#) | [View Legal Code](#)



Attribution-NonCommercial CC BY-NC

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

[View License Deed](#) | [View Legal Code](#)



Attribution-NonCommercial-ShareAlike CC BY-NC-SA

This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.



Attribution-NonCommercial-NoDerivs CC BY-NC-ND

This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

3. Ethical and Legal Compliance

- 1) **personal data** – data that allows living individuals to be identified legal aspect: data protection / privacy legislation (data about living individuals)
- 2) **confidential information** - information given in confidence, agreed to be kept confidential (secret) between two parties
legal: duty of confidentiality

Before collecting data:

- prepare informed consent: information about research, data sharing and preservation

After collecting data:

- protect identities: anonymisation, avoid collecting personal data if not needed
- regulate access where needed: by group, users, time period...

3. Ethical and Legal Compliance

Researcher:

- Inform participants how personal data and confidential information will be used, stored...
- Need consent from participant to share such data
- Store securely, avoid disclosure

Support:

- Comitee for ethical issues
- Information comissioner
- Data center

3. Ethical and Legal Compliance: relevant questions

3.1 How will you manage any ethical issues?

- Have you gained consent for data preservation and sharing?
- How will you protect the identity of participants if required? e.g. via anonymisation
- How will sensitive data be handled to ensure it is stored and transferred securely?

3.2 How will you manage copyright and Intellectual Property Rights (IPR) issues?

- Who owns the data?
- How will the data be licensed for reuse?
- Are there any restrictions on the reuse of third-party data?
- Will data sharing be postponed / restricted e.g. to publish or seek patents?

4. Storage and Back-up

Researcher:

- Appropriate format
- Suitable medium
- Back-up and store
- Assure metadata and documentation
- Deposit

Support:

- Protocols for data deposit and archiving at data center.
- Correct metadata (documentation and data saved in proper format, necessary metadata documentation – use of standards (DDI) and tools.
- Workshops for depositors.
- ADP being designated repository for social science data (defined by Slovene Research Agency)



4. Storage and Back-up

4.1 How will the data be stored and backed up during the research?

- Do you have sufficient storage or will you need to include charges for additional services?
- How will the data be backed up?
- Who will be responsible for backup and recovery?
- How will the data be recovered in the event of an incident?



4.2 How will you manage access and security?

- What are the risks to data security and how will these be managed?
- How will you control access to keep the data secure?
- How will you ensure that collaborators can access your data securely?
- If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems?

5. Selection and Preservation

Researcher:

- Check Quality of data
- Type of access
- Interesting data
- Rare data
- Important data

Support:

- Get involved with trusted repository or journal

5. Selection and Preservation: relevant questions

5.1 Which data are of long-term value and should be retained and/or preserved?

- What data must be retained/destroyed for contractual, legal, or regulatory purposes?
- How will you decide what other data to keep?
- What are the foreseeable research uses for the data?
- How long will the data be retained and preserved?

5.2 What is the long-term preservation plan for the dataset?

- Where e.g. in which repository or archive will the data be held?
- What costs if any will your selected data repository or archive charge?
- Have you costed in time and effort to prepare the data for sharing / preservation?

6. Data Sharing

Researcher:

- Regulation access
- Copy-right
- Promotions

Support:

- Protocols for accessing data.
- Licences.
- Making data available – repository – open data rules – H2020.
- Help repository in promotional activities: lectures, hands-on, report about your publications...



lukasbenc, <https://www.flickr.com/photos/lukasbenc/3493808772> (CC BY-NC-SA)

6. Data Sharing: relevant questions

6.1 How will you share the data?

- How will potential users find out about your data?
- With whom will you share the data, and under what conditions?
- Will you share data via a repository, handle requests directly or use another mechanism?
- When will you make the data available?
- Will you pursue getting a persistent identifier for your data?

6.2 Are any restrictions on data sharing required?

- What action will you take to overcome or minimise restrictions?
- For how long do you need exclusive use of the data and why?
- Will a data sharing agreement (or equivalent) be required?

Responsibilities and Resources

Who will be responsible for data management?

- Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?
- Who will be responsible for each data management activity?
- How will responsibilities be split across partner sites in collaborative research projects?
- Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners?

What resources will you require to deliver your plan?

- Is additional specialist expertise (or training for existing staff) required?
- Do you require hardware or software which is additional or exceptional to existing institutional provision?
- Will charges be applied by data repositories?

For each minute of planning at the beginning of a project, you will save 10 minutes of headache later.



Materials used at the workshop

Štebe, Janez, Bezjak, Sonja, Vipavc Brvar, Irena. **Preparing research data for open access : guide for data producers.** Faculty of Social Sciences, Založba FDV, 2015.

<http://www.dlib.si/?URN=URN:NBN:SI:DOC-G0DPXMZ1>

Horton Laurence: [Data Management Plan \(hands-on\)](#). ADP course for doctoral students: Research data management and open data, 25 July 2015. European Consortium for Political Research (ECPR), Ljubljana, Slovenia, 2015 .

Horton Laurence: [Research Data Management Planning 3](#). ADP course for doctoral students: Research data management and open data, 25 July 2015. European Consortium for Political Research (ECPR), Ljubljana, Slovenia, 2015 .

Contact

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



www.adp.fdv.uni-lj.si



arhiv.podatkov@fdv.uni-lj.si



Arhiv.Druzboslovnih.Podatkov



@ArhivPodatkov

Univerza v Ljubljani



ICPSR

cessda

