How can Research Data Management Help to Produce Data for Comparative Research?

Irena Vipavc Brvar *(ADP - Slovenian Social Science Data Archives)*
Peter Doorn *(DANS – Data Archiving and Networked Services, NL)*
Ellen Leenarts *(DANS – Data Archiving and Networked Services, NL)*

ESRA 2019, Zagreb, 15-19 July, 2019
Highlighting the contribution of survey research in the changing data environment

Whom we represent

- **DANS: Data Archiving and Networked Services – The Netherlands**
  - DANS archives and makes available over 5,000 social science data sets, many of which are survey data (since 1960s)
  - Provides training and expertise on data management
- **ADP: Social Science Data Archives – Slovenia**
  - National repository (academic research, public and private sector, national statistic office)
  - Archives and distribute over 600 datasers, since 1997
- **CESSDA: Consortium of Social Science Data Archives**
  - CESSDA catalogue makes available 20,000 data sets from member countries
  - Provides Data Management Expert Guide
  - ELSST and metadata model

cessda

RDA RESEARCH DATA ALLIANCE

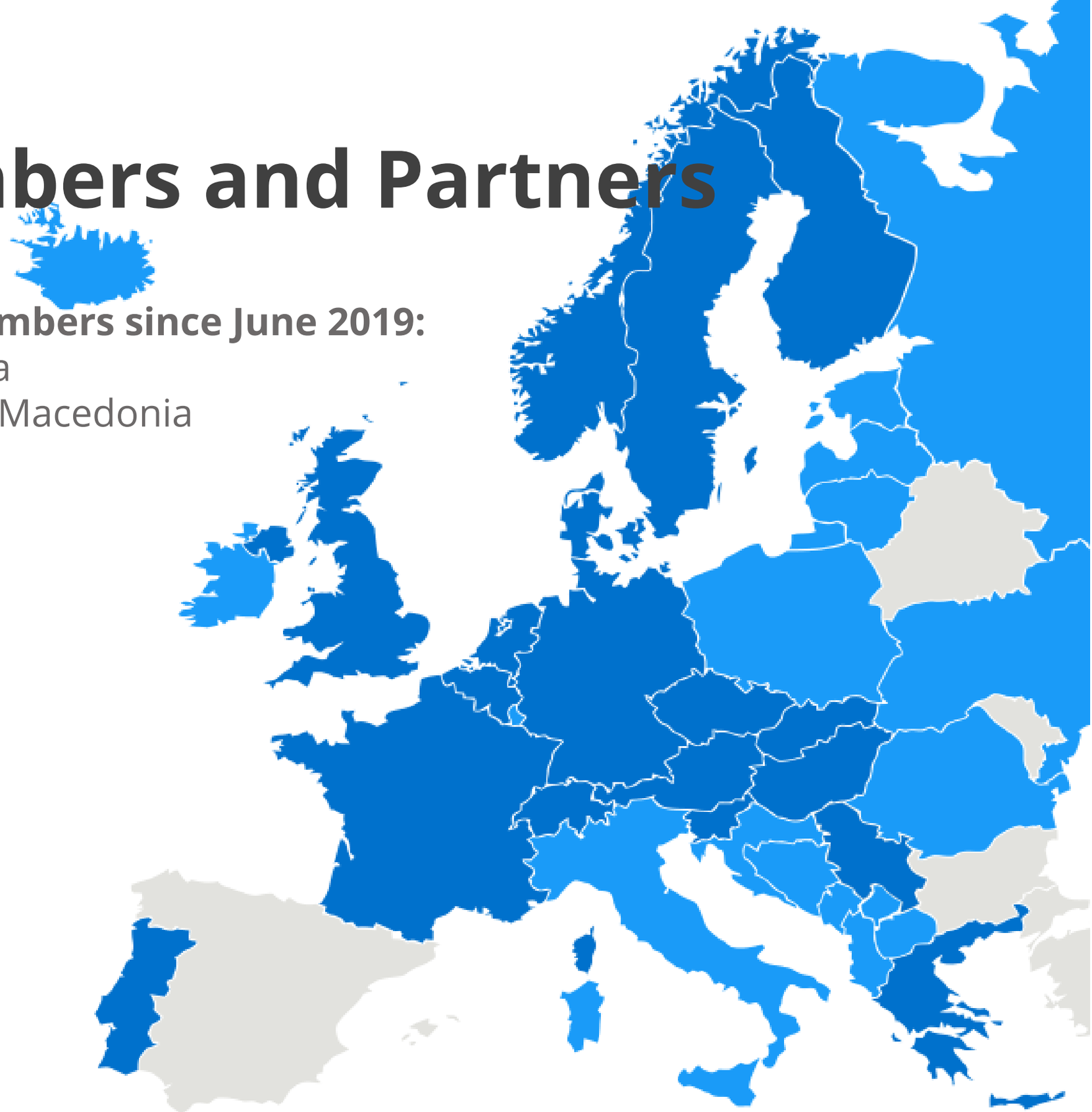
adp DANS

CESSDA Members and Partners

- Austria
- Belgium
- Czech Republic
- Denmark
- France
- Finland
- Germany
- Greece
- Hungary
- Netherlands
- Norway
- Portugal
- Serbia
- Slovakia
- Slovenia
- Sweden
- Switzerland (Observer)
- UK

New members since June 2019:

- Croatia
- North Macedonia



Research Data Management

- Increasingly demanded by research funders, universities and other academic organisations in the context of Open Science
- Many researchers view it as yet another bureaucratic hurdle
- Varying requirements by different organisations drive researchers into despair
- Nevertheless, good practices in data management make sense:
  - Precaution against data fraud and sloppiness
  - Makes research (data collection) process more transparent
  - FAIR (Findable, Accessible, Interoperable, Reusable) data makes it easier to share survey data

Open Science will become the modus operandi of Horizon Europe. It will go beyond the open access policy of Horizon 2020 and require open access to publications, data, and to research data management plans.

2021-2027

Reinforce openness

Practical challenges for researchers in data sharing

SPRINGER NATURE

Springer Nature have published the results of a survey of >7,700 researchers worldwide, looking at data sharing during publication

Main challenge to data sharing is **organising data in a presentable and useful way**

Almost half of all respondents (46%) said that **organising data** was a challenge, followed by **confusion around copyright** (37%) and **not knowing where to share data** (33%)

76%

of respondents

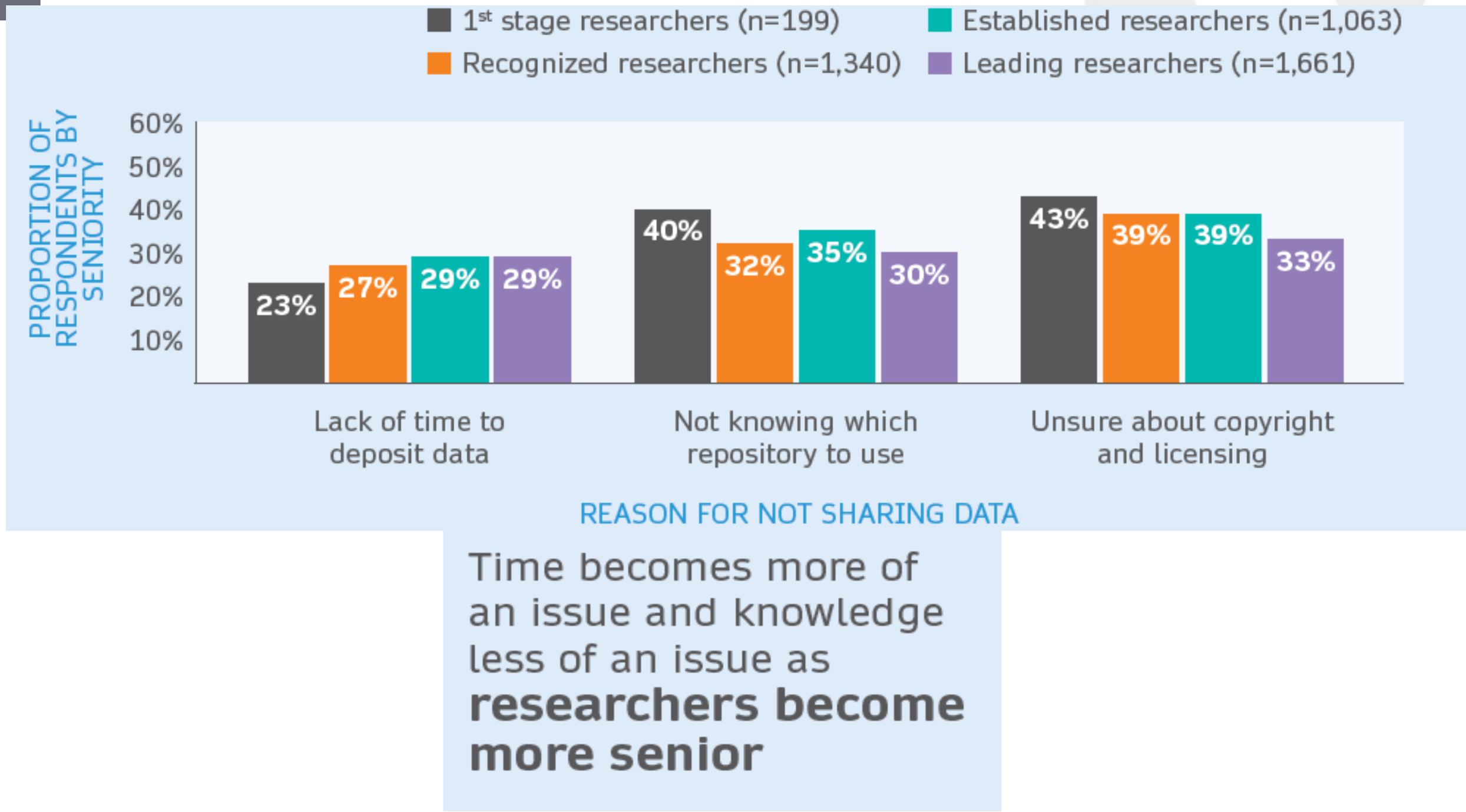
highly rate the importance of their data being discoverable:

Source: Astell, Mathias; Admin, Springer Nature (2018): Infographic - Practical challenges for researchers in data sharing. figshare. Journal contribution.
https://doi.org/10.6084/m9.figshare.5996786.v4

cessda

DANS

Lack of time vs. lack of knowledge



Source: Astell, Mathias; Admin, Springer Nature (2018): Infographic - Practical challenges for researchers in data sharing. figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.5996786.v4

cessda

DANS

What can be done to increase data sharing?

Source: Astell, Mathias; Admin, Springer Nature (2018): Infographic - Practical challenges for researchers in data sharing. figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.5996786.v4

Guide developed by CESSDA Archives

Training / Training Resources / Data Management Expert Guide

Data Management Expert Guide

This guide is designed by European experts to help social science researchers make their research data Findable, Accessible, Interoperable and Reusable (**FAIR**).

You will be guided by different European experts who are - on a daily basis - busy ensuring long-term access to valuable social science datasets, available for discovery and reuse at one of the CESSDA social science data archives.

Self-study for researchers
(15 hours of online content)

www.cessda.eu/DMEG

cessda

adp DANS

Data Management Plan - DMP

A good data management strategy takes into account:

- technical,
- organisational,
- structural,
- legal,
- ethical and
- sustainability aspects.

Easily find and understand data

Increase impact

Make research verifiable

Increase reuse potential

Comply with funder mandates

The time invested in setting up a good data management strategy pays off when the time comes to reproduce your analysis and results.

Chapters in the guide

CESSDA Training Working Group. (2017-2018). CESSDA Data Management Expert Guide Bergen, Norway: CESSDA ERIC. Retrieved from https://www.cessda.eu/DMEG

Recurring elements - DMP

cessda
Consortium of European Social Sciences Data Archives

Adapt your Data Management Plan

A list of Data Management Questions based on the Expert Tour Guide on Data Management

The CESSDA Expert Tour Guide on Data Management is available at https://www.cessda.eu/DMGuide

PLAN

Overview

Title of the project

Date of this plan

Description of the project
- What is the nature of the project?
- What is the research question?
- What is the project time line?

Origin of Data
- What kind of data will be used during the project?
- If you are reusing existing data: What is the scope, volume and format? How are different data sources integrated?
- If you are collecting new data can you clarify why this is necessary?

Principal researchers
- Who are the main researchers involved?
- What are their contact details?

Collaborating researchers (if applicable)
- What are their contact details and their roles in the project?

Funder (if applicable)
- If funding is granted, what is the reference number of the funding granted?

Data producer
- Which organisation has the administrative responsibility for the data?

Project data contact
- Who can be contacted about the project after it has finished?

Data owner(s)
- Which organisation(s) own(s) the data?
- If several organisations are involved, which organisation owns what data?

Roles
- Who is responsible for updating the DMP and making sure that it's followed?
- Do project participants have any specific roles?
- What is the project time line?

Costs
- Are there costs you need to consider to buy specific software or hardware?
- Are there costs you need to consider for storage and backup?
- Are potential expenses for (preparing the data for) archiving covered?

Adapt your DMP: Part 1

« Previous Next » Search this guide Search

The Data Management Plan (DMP) is an important tool to structure the research data management of your project. After working on each chapter you should be able to answer part of the questions which make up a DMP.

This is the first of six 'Adapt your DMP' sections in this tour guide. When you have finished the chapter on data management planning, you can start filling in the 'Overview of your research project' section. Below you can see what elements and corresponding questions are generally included in that section. You can select appropriate questions and answer them to adapt your own DMP.

For easy reference, we have put together a list of DMP-questions for all chapters in this tour guide. You can view and download it (CESSDA, 2017) and keep it as a reference while you are studying the contents of this guide.

- Title of the project
- Date and version of this plan
- Description of the project
- Origin of the data
- Principal and collaborating researchers
- Funder (if applicable)
- Data producer
- Project data contact
- Data owner(s)
- Roles
- Costs

CESSDA Training Working Group. (2017-2018). CESSDA Data Management Expert Guide Bergen, Norway: CESSDA ERIC. Retrieved from https://www.cessda.eu/DMEG

cessda

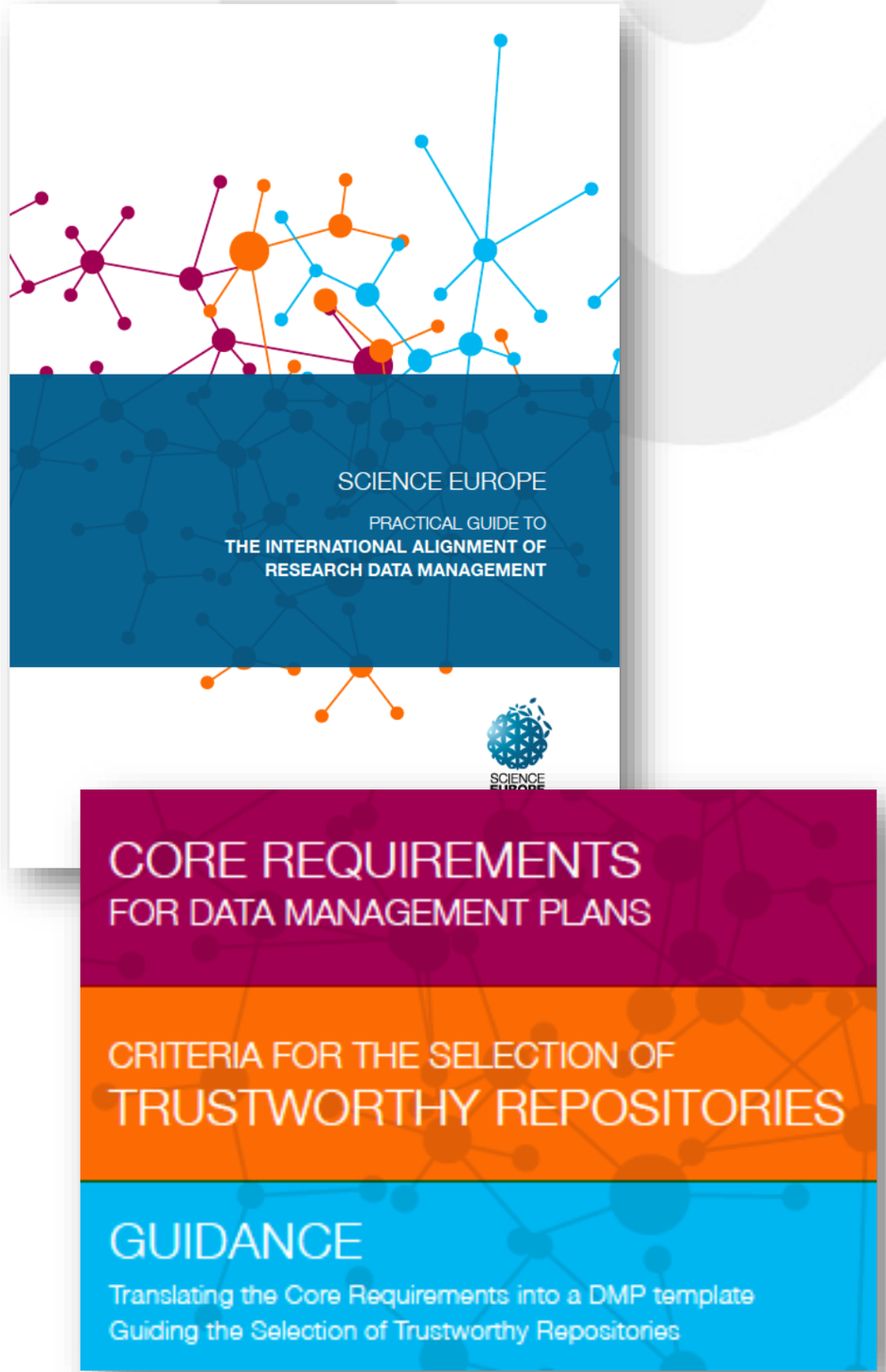
adp DANS

Topics relevant for harmonization / comparative research

- Sampling and data gathering process well documented.
- Response rate information available.
- Questionnaire / instrument + its development
- Do you need to inform participants?
- What format will you use for long time availability / interoperability issues?
- Can you save data in data archive / repository? Which (sensitive) variables will you use for future merging?

European-wide alignment of RDM requirements

- Science Europe Initiative (https://www.scienceeurope.org/policy/policy-areas/research-data/rdm-initiative/)
  - S.E. is European organization of 36 Research Funding and Research Performing Organisations.
  - Joint budget of €18 billion per annum.
  - Practical Guidelines on RDM published in November 2018: https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf

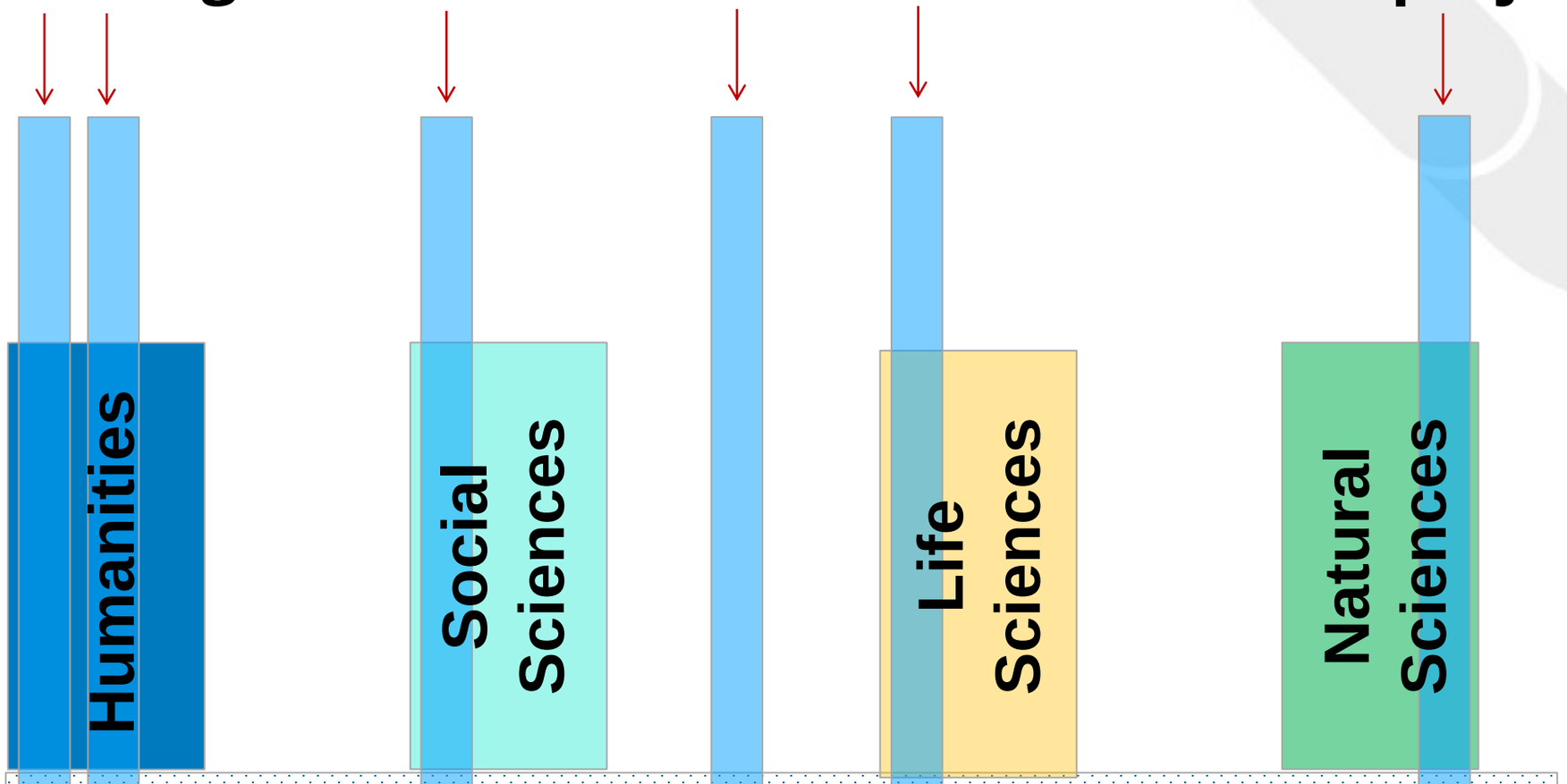


Common core and domain specific requirements for DMP's

Data Management Plans for individual research projects

Institu-tional variations?

Domain specific require-ments



Common Core RDM require-ments:

1. Data description and collection or re-use of existing data
2. Documentation and data quality
3. Storage and backup during the research process
4. Legal and ethical requirements, codes of conduct
5. Data sharing and long-term preservation
6. Data management responsibilities and resources

Science Europe DMP Core Requirements: Six Main Topics

1. Data description and collection or re-use of existing data
2. Documentation and data quality
3. Storage and backup during the research process
4. Legal and ethical requirements, codes of conduct
5. Data sharing and long-term preservation
6. Data management responsibilities and resources

- Every Topic covered by 2-4 questions; 15 questions in total
- Questions have been mapped to the FAIR data principles
- Questions are open (no tick boxes… yet?)
- Additional guidance is available
- Sample templates are available, in report and in DMP Online tool

CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS

When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

1. Data description and collection or re-use of existing data

a. How will new data be collected or produced and/or how will existing data be re-used?

b. What data (for example the kinds, formats, and volumes) will be collected or produced?

2. Documentation and data quality

a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?

b. What data quality control measures will be used?

3. Storage and backup during the research process

a. How will data and metadata be stored and backed up during the research process?

b. How will data security and protection of sensitive data be taken care of during the research?

4. Legal and ethical requirements, codes of conduct

a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

c. How will possible ethical issues be taken into account, and codes of conduct followed?

5. Data sharing and long-term preservation

a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

c. What methods or software tools will be needed to access and use the data?

d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

6. Data management responsibilities and resources

a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Example on next slide

An example

Topic

Requirement (question)

3 STORAGE AND BACKUP DURING THE RESEARCH PROCESS

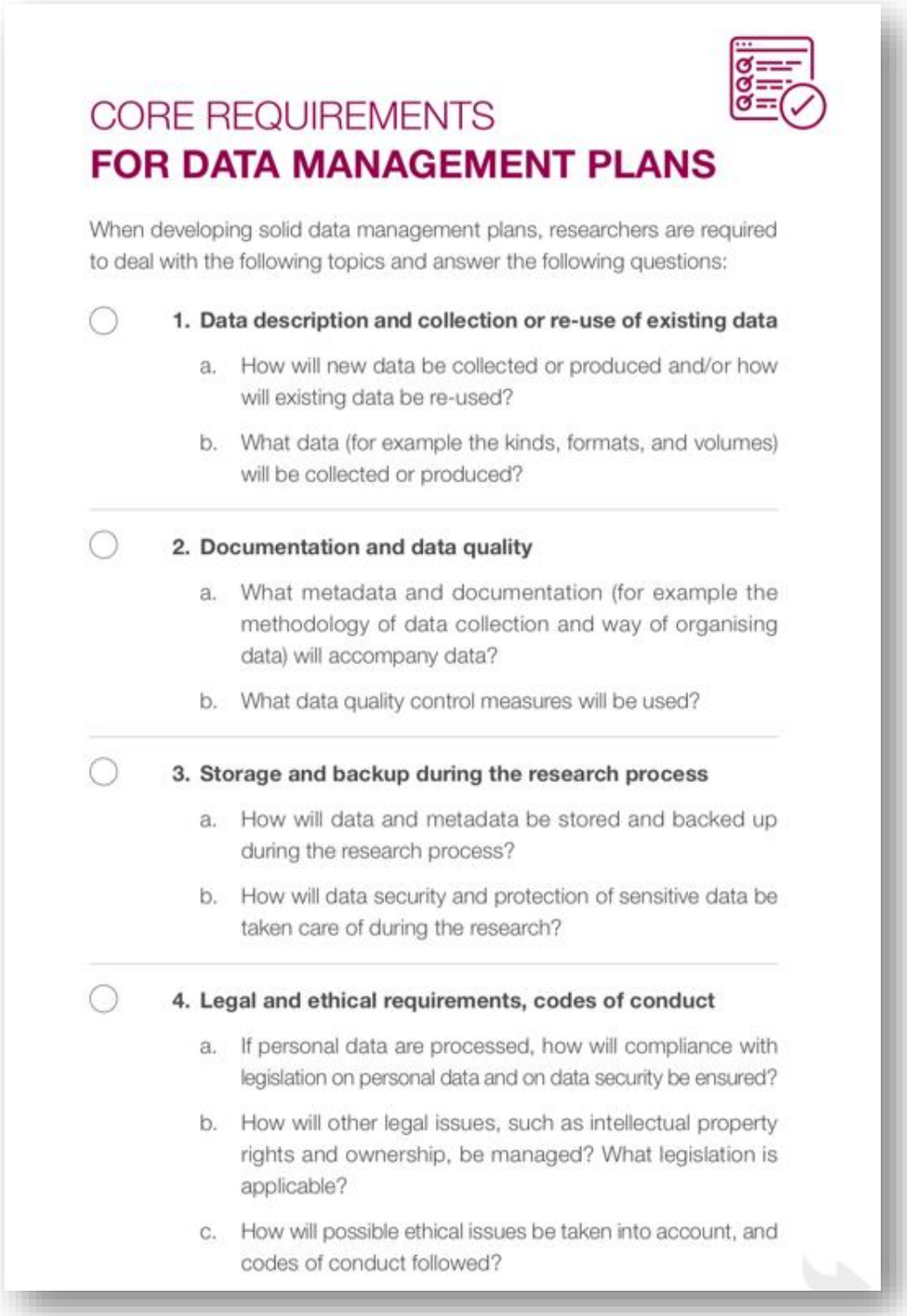
3a

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

Guidance

- Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations.
- Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.

Simplifying the RDM demands for researchers: From requirements to Domain Data Protocol



CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS

When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

1. **Data description and collection or re-use of existing data**
   a. How will new data be collected or produced and/or how will existing data be re-used?
   b. What data (for example the kinds, formats, and volumes) will be collected or produced?
2. **Documentation and data quality**
   a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?
   b. What data quality control measures will be used?
3. **Storage and backup during the research process**
   a. How will data and metadata be stored and backed up during the research process?
   b. How will data security and protection of sensitive data be taken care of during the research?
4. **Legal and ethical requirements, codes of conduct**
   a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
   b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
   c. How will possible ethical issues be taken into account, and codes of conduct followed?

Domain Data Protocol

- to be formulated/accepted by research communities
- to be endorsed by research funders
- principle: comply or explain
- reduces need for individual data management plans
- simplifies evaluation of DMPs by funders

Protocol for Social Science Data (samples)

Draft available at:
http://tiny.cc/il8i9y (Google Doc)
http://tiny.cc/3iwu9y (Google Form)

2. Documentation and data quality
a. What metadata and documentation will accompany the data?

2.a.1. The data will be described by metadata and documented in sufficient detail for other researchers in the social and behavioral sciences to identify, discover and understand the data, so that replication of the research results is possible.

- ❑ Comply
- ❑ Explain: . . . . . .

Comply or Explain

3. Storage and backup during the research process
a. How will data and metadata be stored and backed up during the research process?

3.a.1. Data storage during the research process:

- ❑ The data will be organised and stored in line with good professional practices as recommended by research communities in the social and behavioral sciences (and CESSDA) and/or with the institutional guidelines for RDM of the home institution where the data will be hosted.
- ❑ The project will check if its home institution has a backup strategy in place which meets the requirements described in this section. If yes, the project will make use of the backup policy of the home institution.
- ❑ A version control mechanism will be used and all changes to the raw data (weighting, recoding, creation of new variables, corrections for or omission of outliers) will be duly documented.
- ❑ The project will explicitly assign the responsibilities for backup administration to:
  - ❑ one or more members of the project team.
  - ❑ technical support staff of the home institution or an external service provider.

Alternatives: tick what applies

Science Europe Topic	**RDM Requirements**	Protocol Articles
1. Data description and collection or re-use of existing data	a. How will new data be collected or produced and/or how will existing data be reused?	2
	b. What data will be collected or produced?	5
2. Documentation and data quality	a. What metadata and documentation will accompany the data?	3
	b. What data quality control measures will be used?	4
3. Storage and backup during the research process	a. How will data and metadata be stored and backed up during the research process?	3
	b. How will data security and protection of sensitive data be taken care of during the research?	5
4. Legal and ethical requirements, codes of conduct	a. Compliance with legislation on personal data and data security	4
	b. Other legal issues, such as intellectual property rights and ownership	5
	c. Ethical issues and codes of conduct	5
5. Data sharing and long-term preservation	a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?	10
	b. How will data for preservation be selected, and where will data be preserved long-term?	7
	c. What methods or software tools will be needed to access and use the data?	3
	d. Persistent identifier (such as DOI)	1
6. Data management responsibilities and resources	a. Who will be responsible for data management?	3
	b. Resources (money, time) needed for data management and ensuring that data will be FAIR	1
6 Topics	15 Requirements	61

For discussion

- Comments are welcome!
- Questions:
  - Is the idea of a Domain Data Protocol useful?
  - Is the Protocol too long or too detailed? What should be shorter?
  - Is the language understandable (not too difficult, complicated, technical?)
  - Which organization(s) could be asked to endorse it?
  - Which topic needs most guidance and training?

Thank you

Contact us: training@cessda.eu

Irena.vipavc@fdv.uni-lj.si
peter.doorn@dans.knaw.nl
ellen.leenarts@dans.knaw.nl

cessda.eu @CESSDA_Data

@pkdoorn
@dansknaw