

PLAN

CESSDA Expert Tour Guide

Gry Henriksen

14.05.19 Athens

www.cessda.eu/DMEG

 cessda.eu  @CESSDA_Data

Plan Research Data Management

cessda eric

[About](#) - [Consortium](#) - [Projects](#) - [Research Infrastructure](#) - [Contact](#)



[Home](#) / [CESSDA Training](#) / [Expert tour guide on Data Management](#)

Expert tour guide on Data Management



About this expert tour guide

This tour guide aims to put social scientists like yourself at the heart of making their research data findable, sustainably accessible and (re)usable.

You will be guided by 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 [15 CESSDA social science data archives](#). With this guide and training events throughout Europe, we want to accompany and inspire you in your travels through the research data lifecycle.

[Full table of contents](#) -

[CESSDA Training](#) ↑

First a few questions:

How many of you have:

- Have read chapter one of the tour?
- Have read other chapters than chapter 1?
- Have not read any part of the expert tour?
- Have looked at the DMP template?
- Have produced a DMP?

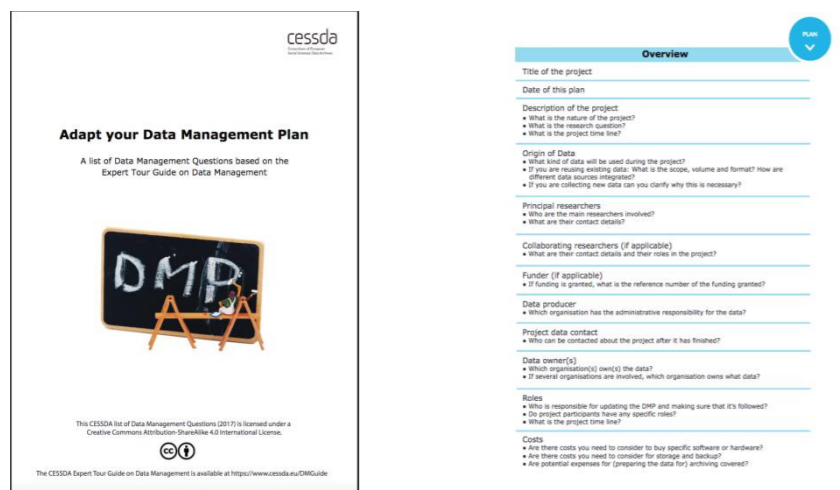
Data Management Plan

- Important tool – handle, organize, structure and store research data
- Can be a formal document that outlines how to handle data during and after a project
- Is designed in accordance with the specific project



DMP checklist

- Adapt your DMP section at the end of every chapter
- Corresponding questions to each chapter
- The DMP checklist is downloadable



Adapt your DMP: Part 1

[« Previous](#) [Next »](#)

Search this guide



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

Organise & Document

RDM Expert Tour Guide

Gry Henriksen, NSD

14.05.19 Athens

www.cessda.eu/DMEG

 cessda.eu  @CESSDA_Data

Designing a data file structure

Huge impact – processed and analysed

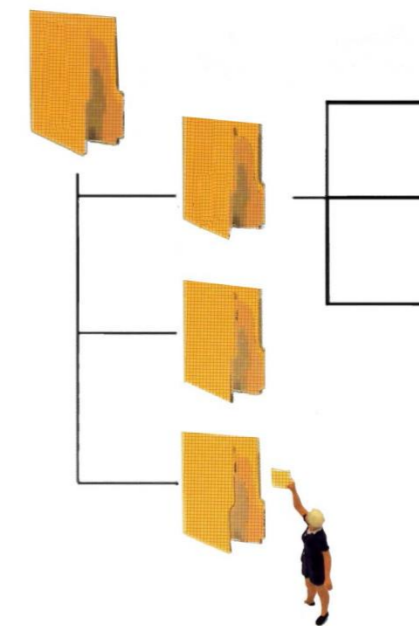
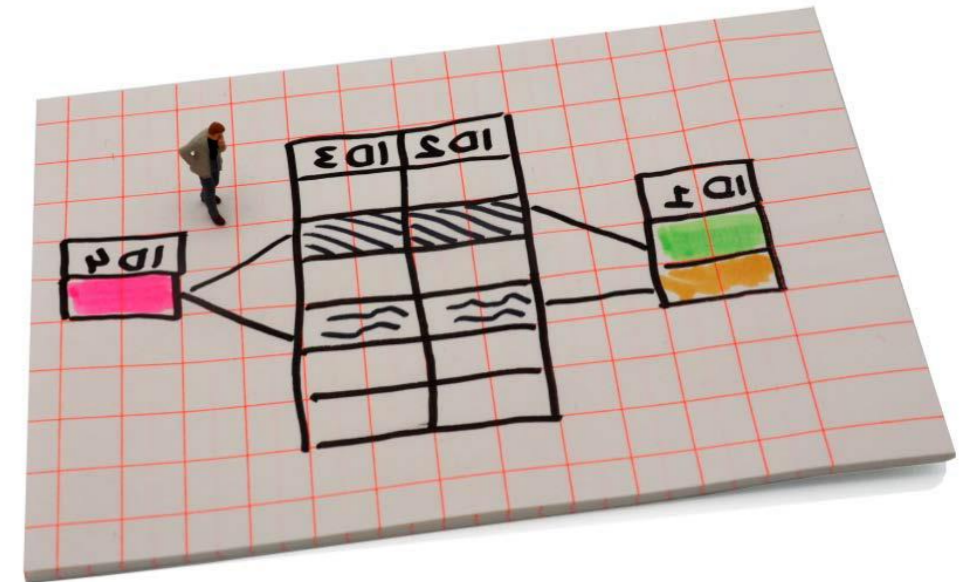
Qualitative data:

- Coding,
- File naming
- Folder structure

Quantitative data, survey data:

Flat(rectangular) data files, hierarchical files, relational database

- Standard coding
- File naming
- Folder structure



Survey data

Variable names and labels:

- Contributes to structuring
- Allows researchers to integrate documentation into the data file
- Makes the data file more FAIR compliant
- Follow basic rules and have a strategy

Documentation and metadata

What is metadata?

- Information about your data, everything that can make data reusable – for you and others
- Interviewguide, questionnaire (survey), instrument information etc

Why document data?

- Make your data Findable, Accessable, Interoperable and Reusable = FAIR
- Easier to share your data
- Gives you as a researcher more relevance in your research field – and beyond
- Good metadata documentation give your data added value

How to start?

- Use a Data Management Plan
- Start early
- Do it thru out the project
- Data archives use standards, adopt the one 'your' archive uses
- Document your data both in English and your language

Process

CESSDA Expert Tour Guide

Gry Henriksen, NSD

14.05.19 Athens

www.cessda.eu/DMEG

 cessda.eu  @CESSDA_Data

What is 'Process'

Data operations needed to prepare your data files for analysis and data sharing.

- Both by you and others in the project, and in the data archive.

Crucial to:

- Maintain the authenticity
- Prevent loss or deterioration

Data entry

Data entry changed over time:

Quantitative data:

- Automation – prevents some mistakes – produces others
- Coding
- Anonymise
- File format for long time preservation

Qualitative data:

- Transcription – record and transcribe
- Coding
- Anonymise
- File format for long time preservation

Data integrity

- Assurance of the accuracy, consistency and completeness of original information contained in the data.
- Integrity of a data file is based on its structure and on links between data and integrated elements of documentation.
- From the moment data is being entered, data integrity is at stake.

Data authenticity

- Preserve original research information/file
- Versions control
- Use best practice rules

Best practices for quality assurance, version control and authenticity

Version and edition management will help to:

1. Clearly distinguish between individual versions and editions and keep track of their differences;
2. Prevent unauthorised modification of files and loss of information, thereby preserving data authenticity.

Best practices

The best practice rules (UK Data Service, 2017a; Krejčí, 2014) may be summarised as follows:

- Establish the terms and conditions of data use and make them known to team members and other users;
- Create a 'master file' and take measures to preserve its authenticity, i.e. place it in an adequate location and define access rights and responsibilities – who is authorised to make what kind of changes;
- Distinguish between versions shared by researchers and working versions of individuals;
- Decide how many versions of a file to keep, which versions to keep (e.g. major versions rather than minor versions (keep version 02-00 but not 02-01)), for how long and how to organise versions;
- Introduce clear and systematic naming of data file versions and editions;
- Record relationships between items where needed, for example between code and the data file it is run against, between data file and related documentation or metadata or between multiple files;
- Document which changes were made in any version;
- Keep original versions of data files, or keep documentation that allows the reconstruction of original files;
- Track the location of files if they are stored in a variety of locations;
- Regularly synchronise files in different locations, such as using [MS SyncToy](#) (2016).

Data quality

- are best judged not by its size, scope, or prominence, but by how much attention is given to (preventing, measuring and) dealing with the many important problems that can arise. (American Association for Public Opinion Research (2015) (AAPOR) (CESSDA expert guide – ch3)
- defined as fitness to use

STORE

CESSDA Expert Tour Guide

Gry Henriksen, NSD

14.05.19 Athens

www.cessda.eu/DMEG

 cessda.eu  @CESSDA_Data

Storage?

- Definition: 'the action or method of storing something for future use'
- Data storage: "the retention of retrievable data on a computer or other electronic system (Oxford Dictionary)
- Storage does not means to simply push the storage button, i.e. to put something, somehow, somewhere for future use
- Storage is a systematic task

Why is it important?

- various storage solutions
- storage strategy:
 - what is stored and how
 - backup and disaster recovery
 - protect against unauthorized (mis-)use
- part of a systematic data management plan
- Closely connected to other RDM activities, e.g.
 - organization (and documentation)
 - data protection
 - publication (and long-term preservation)
- Requires systematic planning (early task)

Storage strategy

- A Storage strategy contains:
 - storage solutions and media
 - backup strategy and disaster recovery
 - data protection
- And it should be systematically implemented in a data management plan



Storage solutions:

Many different solutions:

➤ Local storage



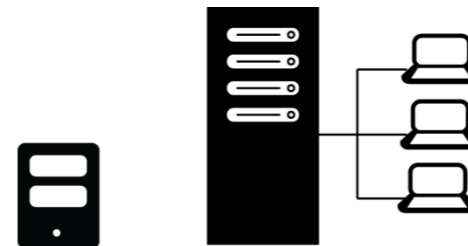
➤ Cloud storage



➤ Portable devices



➤ Network drives



Storage and media

- Optical (CD, DVD)
- Magnetic (Harddrive)
- Portable flash drive (Memory stick)
- Build in flash drive (Harddrive)

Recommendations: use at least two types of storage media, replace storage media(after 2-5 years) and carry out integrity checks,e.g. by checksum tool

Backup

Various reasons for data loss, e.g.

- hardware failure
- software malfunction
- malware or hacking
- human error
- theft, natural disaster or fire
- degradation of storage media
- etc.



Developing a backup strategy

1. Institutional backup strategy
 - How does it work?
2. What has to be back upped
 - What needs to be copied?
3. When is it back upped and how often
 - Frequency and number of copies?
4. Where is it back upped
 - Storage solutions for copies?
5. Storage capacity needed for backups
 - Memory capacity needed for copies?
6. Tools that can be used to automate backups
 - Automate backup processes of e.g. cloud services?
7. How long is it back upped and how will it be destroyed
 - Storage period and destruction of irrelevant copies?
8. How will personal data be protected
 - Data protection strategy for copies?
9. Disaster recovery plan
 - How to access and (re-)use copies
10. Responsibilities
 - Who is responsible for backups?

Data Protection and Data Security

The Worst 10 Passwords of 2017

1. *123456*
2. *password*
3. *12345678*
4. *qwerty*
5. *12345*
6. *123456789*
7. *letmein*
8. *1234567*
9. *football*
10. *iloveyou*

Password
Encryption

Time Magazine (2018): The Worst 25 Passwords of 2017.

Available at:

[http://tim.com/5071176/worst-passwords-2017/.](http://tim.com/5071176/worst-passwords-2017/)

Storage and the DMP

- Storage strategy is an important part of the DMP
- Adapt to both short and long term storage

Archive and publish

CESSDA Expert Tour Guide

Gry Henriksen, NSD

14.05.19 Athens

www.cessda.eu/DMEG

 cessda.eu  @CESSDA_Data

Cita as: Henriksen (2019), Bezjak (2018) Athens TTT workshop 2019, Train the Trainers event April 2018, Ljubljana, Slovenia

Why archive research data?

- *Archiving and publishing your data properly will enable both your future self as well as future others to get the most out of your data.*

Archiving

Archiving data

- store your data in a suitable file format, with adequate documentation and keep your data safe on along term
- make sure you can read and access the data later on
- allow access to others for verification purposes

The FAIR principle

F – findable

A – accessible

I – interoperable

R – reusable

Publishing data

Publicly disclosed your research data

To make the data findable

To make them accessible, at least metadata

To make them reusable

To be in compliance with the FAIR principle

What's in it for them

- **Career benefits**

- increased visibility, reuse and citation and therefore recognition of scholarly work

- **Scientific progress**

- enabling new collaborations, new data uses and links to the next generation of researchers

- **Norms**

- openness of research data is at the heart of scientific ethics

- **External drivers**

- Funders and publishers requirements

CESSDA archives

- **(Trusted) domain specific data repositories**
 - For high-quality data with a potential for reuse, we recommend you to assure long-term access by publishing them with a trusted repository, like many of the CESSDA archives.
- **Advantage of having expert help within reach**
 - help you to increase the comprehensibility, visibility, findability,

Thank you – Questions?

Gry.Henriksen@nsd.no

Contact us:

training@cessda.eu

Train the Trainers package:

www.cessda.eu/DMEG/TTT



cessda.eu



@CESSDA_Data