Love data management

Webinar on the Expert tour guide on Data Management

Love Data Week
16 Feb 2018

Veerle Van Den Eynden (UKDS)
Ellen Leenarts (DANS)
Ulf Jakobsson (SND)
Gunn Inger Lyse (NSD)

www.cessda.eu/DMGuide
Overview

- Introduction to CESSDA-ERIC, the CESSDA Training Working Group and speakers – Veerle Van den Eynden (UKDS)
- Expert tour guide, the project and future - Ellen Leenarts (DANS)
- Data management planning and how the module can help – Ulf Jakobsson (SND)
- Using the Expert tour guide for self-study and training - Gunn Inger Lyse (NSD)
- Q&A – Veerle Van den Eynden (UKDS)
CESSDA-ERIC

CESSDA = Consortium of European Social Science Data Archives
ERIC = European Research Infrastructure Consortium

- Large-scale, integrated and sustainable data services to the social sciences
- Network of social science data archives across Europe
- Sustainable research infrastructure enabling the research community to conduct high-quality research in the social sciences contributing to the production of effective solutions to the major challenges facing society today and to facilitate teaching and learning in the social sciences.
CESSDA Training Working Group

Three pillars:

• Technology: products, services and tools for social science research based on data reuse
• Trust: CESSDA Service Providers as ‘trusted repositories’ to ensure quality of data and safe and secure access
• Training: for researchers and to ‘train the trainers’

• Training Working Group started 2016, experts across all CESSDA archives, focus on
  • Data Discovery: finding and accessing data across Europe
  • Research Data Management: good practices in making data findable, understandable, sustainably accessible and reusable
Expert tour guide on Data Management

Ellen Leenarts (DANS)

About this expert tour guide

This tour guide by CESSDA ERIC (the Consortium of European Social Science Data Archives European Infrastructure Consortium) aims to put social scientists like yourself at the heart of making their research data findable, understandable, sustainably accessible and reusable.

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 17 CESSDA social science data archives. With this guide and the training events being held across Europe, we want to accompany and inspire you in your journey through the research data life cycle.
Why a CESSDA online module on RDM?

- Local face to face workshops on RDM in various countries by CESSDA partners
- Share good training practices, examples and content
- Growing demand for DMPs by funders (Horizon2020 and local funders)
- Domain specific support on RDM, see also:

Online module:
- Create opportunity for individual early career researchers for self-study
- Blended learning opportunity for local workshops
- The guide can be centrally updated and enriched
Timeline of the ‘one year’- project

- Dec 2016: Kick off
- January – April 2017: Developing content ideas, decision Course or Guide
- May 2017: Content workshop, feedback on the provisional content by target audience (early career researchers)
- May - Sept 2017: Hard work on the content, editing, styling and implementation online
- Oct 2017: Feedback period for researchers, DM specialists and trainers
- Nov – Dec 2017: Improve content, starter package for trainers, launch of the website
product statement

for social scientists who are in an early stage of practicing RDM

the CESSDA expert tour guide to data management

an openly licensed learning tool/learning bouquet

provides discipline specific, hands-on guidance from a European perspective with local expertise

unlike MANTRA or Essentials 4 Data Support

which have a more general/different audience and lack international perspective

the CESSDA expert tour guide to data management excels in balancing simplicity (short, clear, practical) with richness. It is appealing because of its fun factor and freshness (fluffyness)

learning is designed as an online tour guide (based on the research data lifecycle) which is customisable for local use/training
Content of the chapters

- Authors per chapter: ADP, CSDA, DANS, FSD, GESIS, SND, and UKDS
- Overall coordination, editing, styling and online implementation: DANS & Verbeeldingskr8
- Feedback and testing: ADP, AUSSDA, CSDA, DANS, FORS, NSD, So.Da.Net and UKDS
Recurring elements in the chapters

Recurring elements:

• Expert tips
• European diversity
• Qualitative vs. Quantitative data
• Adapt your DMP
Expert tips

**Expert tip**

How FAIR are your data?
Want to know how FAIR your data are? Have a look at the checklist by Jones and Grootveld (2017).

**Expert tips**

Any researcher who wishes to become proficient at doing qualitative analysis must learn to code well and easily. The excellence of the research rests in large part on the excellence of the coding | Strauss (1987).

- Tip 1: Document the meaning of codes
- Tip 2: Prevent coder variance

**Expert tips**

- TIP 1. Documenting consent
- TIP 2. Delivering informed consent in the best way possible
- TIP 3. Consent for surveys
- TIP 4. Research without consent
European diversity

Storage of raw research data for at least 10 years

For research conducted in the Netherlands, the raw research data are required to be stored for at least ten years. Additionally, this data must also be made available to other academic practitioners upon request (unless legal provisions dictate otherwise). Researchers who receive a Netherlands Organisation for Scientific Research (NWO) grant are required to disclose data even after ten years.

It is therefore important for researchers working on research projects in the Netherlands or collaborative projects which include research within the Netherlands to consider this in the Data Management Plan (DMP) and their project preparations, so as to ensure that they have a system in place to store the research data for at least ten years.

More information can be found in the Netherlands Code of Conduct for Academic Practice (Association of Universities in the Netherlands, 2014) and Research Data Netherlands (n.d.) can provide further guidance and advice on this requirement.

Data management requirements in Europe

There are many different local, national and international DMP templates and tools that you can use to create a DMP for your own research project. At this stage, it might be good for you to check for templates or tools that best fit your own specific situation. You can ask at your university or department whether they have their own DMP template. Or maybe your research funder requires a DMP in a specific format.

In the accordion below we sum up European diversity in funder requirements on Data Management Planning and link to DMP templates if they are available.

- EU
  - Belgium
  - Czech Republic
  - Finland
  - Germany
  - Netherlands
  - Norway
  - Slovenia
  - Sweden
  - Switzerland
  - UK
Quantitative vs. Qualitative data

Minimising errors in survey data entry

In the accordion below a summary of recommendations on minimising errors in survey data entry is given (UK Data Service, 2017a; ICPSR, 2012; Groves et al., 2004).

- Check the completeness of records
- Reduce burden at manual data entry
- Minimise the number of steps
- Conduct data entry twice
- Perform in-depth checks for selected records
- Perform logical and consistency checks

Designing qualitative data files

Qualitative data files emerge from many different types of research material. Such data files are texts (transcribed interviews or focus group sessions, various types of written texts, such as newspaper and magazine material, diaries etc.) or photographs, audio files (recordings of speech) or video files. Unlike quantitative data, qualitative data are not presented in form of variables, numbers, data matrices etc. Alike, they must be organized and stored in an exact precise manner so they are easily managed and ready for use.

Usually, individual data collection events will be structured into individual files, e.g. one interview transcript, one image, one audio recording each time makes a single file. These single files are then organized into folders of similar files. Sometimes, qualitative information may also be organized into matrix structures, e.g. textual extracts from newspaper articles or diaries may be placed into a rectangular matrix, whereby further metadata and coding can be added alongside each entry.

Designing a qualitative data structure comes down to:
- Thinking of ways to categorise data (see “Qualitative coding”);
- Developing a file naming strategy (see “File naming and folder structure”);
- Designing a comprehensive folder structure (see “File naming and folder structure”).

Designing quantitative data files

In quantitative research, the content of the data often results from numerical coding in standardised questionnaires (see “Quantitative coding”). In addition, full-text answers or textual codes can be recorded into specific types of variables in quantitative data files. Quantitative researchers may also store other material, i.e. administrative data, data from social media or various texts. However in this chapter, when we speak about quantitative data, we usually mean survey data.

Considerations in making high-quality transcriptions of qualitative data

The most common formats of qualitative data are written texts, interview data and focus group discussion data. In most cases, interview and discussion data are first digitally recorded and then transcribed. Transcription is a translation between forms of qualitative data, most commonly a conversion of audio or video recordings into text. If you intend to share your data with other researchers, you should prepare a full transcription of your recordings (Bucholtz, 2000).

There are several basic rules and steps in the process of making and checking a high-quality transcript from audio/video (Kuckartz, 2014).
User Feedback

• First version was given to partners for feedback
• 30 people provided feedback

I love the interactive tables.

It is extremely informative and goes into a lot of detail.

I want to complement you with the website. It looks fresh and modern!

Excellent overview of DMP
Clear, concise, excellent visuals

Very detailed information for a wide range of topics, step-by-step guidance.

This tool is great – it is not only useful for researchers but also for data managers or young archives.
Starter package for trainers

- Outlines for two possible workshops
  - 1-day general: Introduction to RDM
  - 1-day content-specific: Ethical and legal considerations in RDM
- Both use the content of the module + existing materials from partners
- Future updates: additional exercises, slides and explanation per chapter, overall background and structure of the module, checklist DMP, chapters in pdf, folder with all visualisations, evaluation form, etc.
- Contact ellen.teenarts@dans.knaw.nl or training@cessda.net to receive outlines and future additions to the starter package for trainers
Launch of the module in December 2017
Activities 2018

Content:
- Add discovery chapter
- Add to the Starter package for trainers: Refining the two outlines, additional exercises, slides and explanation per chapter, overall background and structure of the module, checklist DMP, chapters in pdf, folder with all visualisations, evaluation form
- Adapt content (corrections English, chapter leads adding content based on feedback local workshops)

Events:
- CESSDA Train de trainer workshop - April 2018
- Local workshops
Expert tour guide on Data Management – DMP

About this expert tour guide

This tour guide by CESSDA ERIC (the Consortium of European Social Science Data Archives European Infrastructure Consortium) aims to put social scientists like yourself at the heart of making their research data findable, understandable, sustainably accessible and reusable.

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 17 CESSDA social science data archives. With this guide and the training events being held across Europe, we want to accompany and inspire you in your journey through the research data life cycle.
Benefits of data management

The concept of Data Management implies

» How to handle, organize, structure and store research data
» Takes into account technical, organizational, structural, legislative and sustainability aspects
» Clear structure of how data is going to be managed
» Might involve some additional work at an early stage
The Data Management Plan (DMP)

» Is an important tool that will aid you as a researcher to structure the data management within your project.

» Can be seen as a formal document that outlines the frames for how to handle the data during and after the project.

» Is designed in accordance with the specific project
Added Value

» Discover possible problems at an early stage
» All information in one place
» Calculating cost for data management
» Allows early preparations
» Serious data management
Why write a DMP?

» Easier for others to understand the material
» Enables further research after the project has ended
» Research results can be verified
» Prevents unnecessary data collection
Writing your own DMP

Start with the DMP checklist

• Adapt your DMP section at the end of every chapter
• Corresponding questions to each chapter
• Reference list of DMP questions in first chapter
Adapt your Data Management Plan

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

This CESSDA set of Data Management Questions (2017) is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

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

### 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 timeline?

- **Origin of Data**
  - What kind of data will be used during the project?
  - If you are using 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 researcher(s)**
  - 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 is followed?
  - Do project participants have any specific roles?
  - What is the project timeline?

- **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?
Expert tour guide on Data Management – How to use it

About this expert tour guide

This tour guide by CESSDA ERIC (the Consortium of European Social Science Data Archives European Infrastructure Consortium) aims to put social scientists like yourself at the heart of making their research data findable, understandable, sustainably accessible and reusable.

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 17 CESSDA social science data archives. With this guide and the training events being held across Europe, we want to accompany and inspire you in your journey through the research data life cycle.
Expert tour guide on Data Management – Chapter overview

[Diagram showing the stages of data management: Plan, Discover, Organise & Document, Publish, Process, Store, Protect]
Expert tour guide on Data Management – Chapter overview

1 Designing a data file structure
Expert tour guide on Data Management – Chapter overview

1. Designing a data file structure
2. File naming and folder structure

Qualitative data files

In this example, the data contain audiotapes of the interviews, interview transcripts, stimulation material shown to the research subjects, and photographs taken by the subjects. Data files are files connected to the same interview event conducted on the 22nd of January 2013. The latter part of the name reveals the specifics of the file. In this case, “audio” means audio tape and “trans” a transcription of the audio tape. However, background information must never be stored in the file name only.
Expert tour guide on Data Management – Chapter overview

1. Designing a data file structure
2. File naming and folder structure
3. Documentation and metadata

**Project-level documentation**

Project-level documentation explains the aims of the study, what the research questions/hypotheses are, what methodologies were being used, what instruments and measures were being used, etc. In the accordion the questions which your project-level documentation should answer are stated in more detail:

1. For what purpose was data created
2. What does the dataset contain
3. How was data collected
4. Who collected the data and when
Expert tour guide on Data Management – Chapter overview

Data quality:
- How to prepare your data files for analysis and data sharing
- Strategies to minimise errors
- Manage the integrity and authenticity of your data

Coding recommendations
In the accordion below you find coding recommendations which are inspired by ICPSR (2012).

- Include identification variables
- Make code categories exclusive and coherent throughout the database
- Preserve original information
- Document the coding schemes
- Check verbatim text data for data disclosure risk
Expert tour guide on Data Management – Chapter overview

1. Storage
2. Backup
3. Security

- Passwords
- Encryption
- Physical, network and computer security
- Secure disposal
- Organisational aspects
Expert tour guide on Data Management – Chapter overview

- Legal requirements of Member States
- Impact of General Data Protection Regulation (GDPR)
- How to share personal data?
- Copyright: who owns your research data?

**QUANTITATIVE DATA | QUALITATIVE DATA**

*Best practices for anonymising qualitative data*

- Using pseudonyms or generic descriptors to edit identifying information, rather than blanking-out that information;
- Plan anonymisation at the time of transcription or initial write-up, (longitudinal studies may be an exception if relationships between waves of interviews need special attention for*
Expert tour guide on Data Management – Chapter overview

- Which data to select for publication?
- Where to archive?
- Licences and access levels

Choosing a data repository

There are hundreds of repositories worldwide. Some cater for a specific research domain, while others are general-purpose repositories. They may be called something other than a repository, for example, a data centre or archive | Whyte (2015).
Expert tour guide on Data Management – Chapter overview

Data discovery chapter to appear in 2018
Questions?

training@cessda.net

The Expert Tour Guide on Data Management by CESSDA ERIC is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. All material under this licence can be freely used, as long as CESSDA ERIC is credited as the author.”