Responsible Open Science

Fundamentals of Research Data Management for Social Scientists

Francesca Morselli & Ricarda Braukmann, DANS

14 and 17 June 2021

ODISSEI

cessda

Data Archiving and Networked Services
DANS

House Rules

- Keep mic muted during the presentations
- Raise your hand if you want to speak or write questions in the chat
- Don't take screenshots without the permission of the participants
- Presentations are recorded | Breakout sessions are not

Veertly Platform

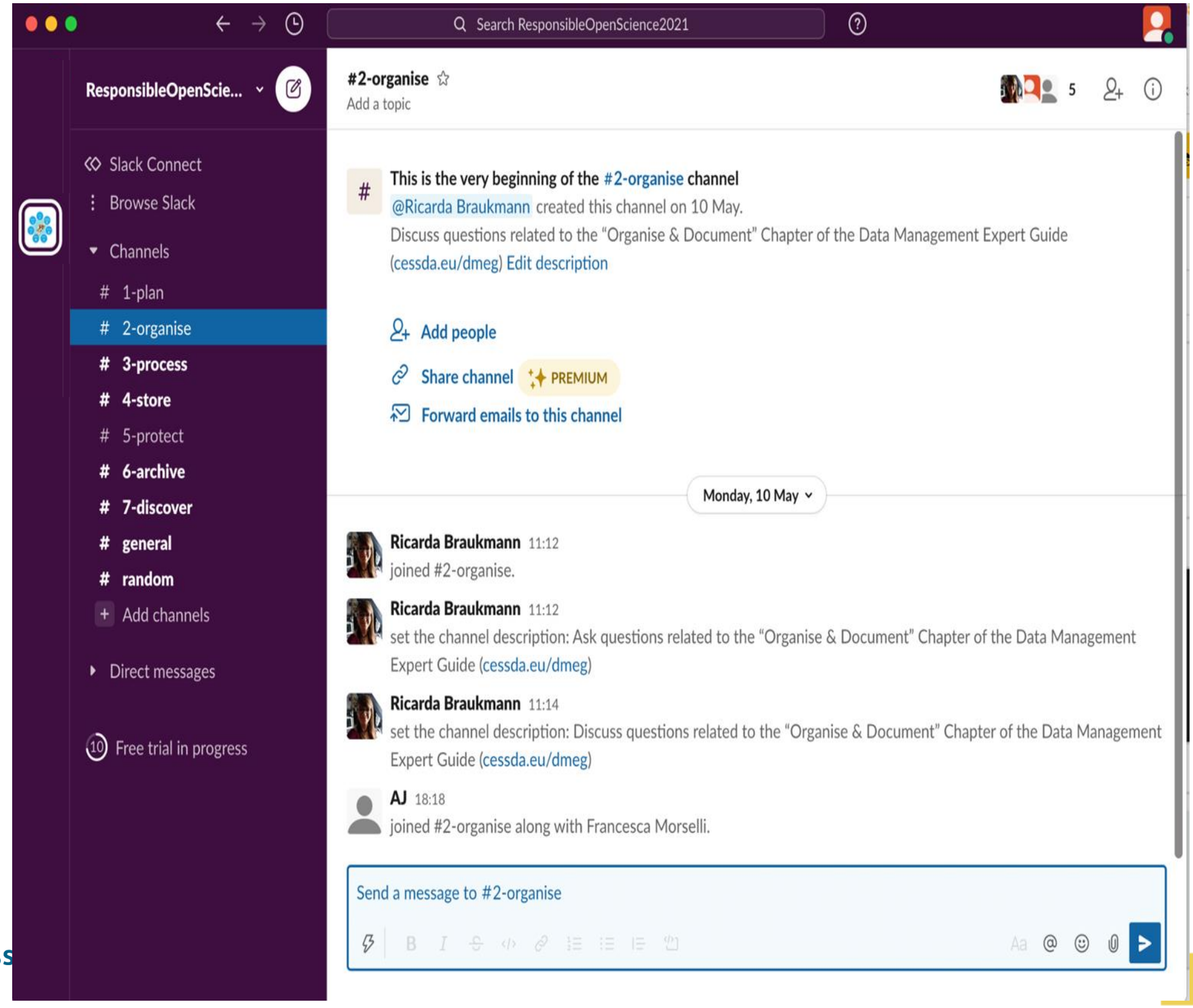
- Google Chrome gives best performance
- Set language Veertly to English
- Agenda
- Main stage
- Break-out rooms (2nd day)
- Chat

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Slack Channel



Open Data Infrastructure for Social Sciences and Economic Innovations

www.odissei-data.nl

ODISSEI

cessda

Data Archiving and Networked Services
DANS

https://odissei-data.nl/nl/

- Open Data Infrastructure for Social Science and Economic Innovations
- Dutch national infrastructure
- Coordinates data collections (i.e. surveys and panels) operated by members
- Supports researchers to access to CBS microdata
- Supports researchers to use the LISS*-panel
- Provides funding (e.g. microdata access grants)
- Super computing facilities (hosted by SURF)

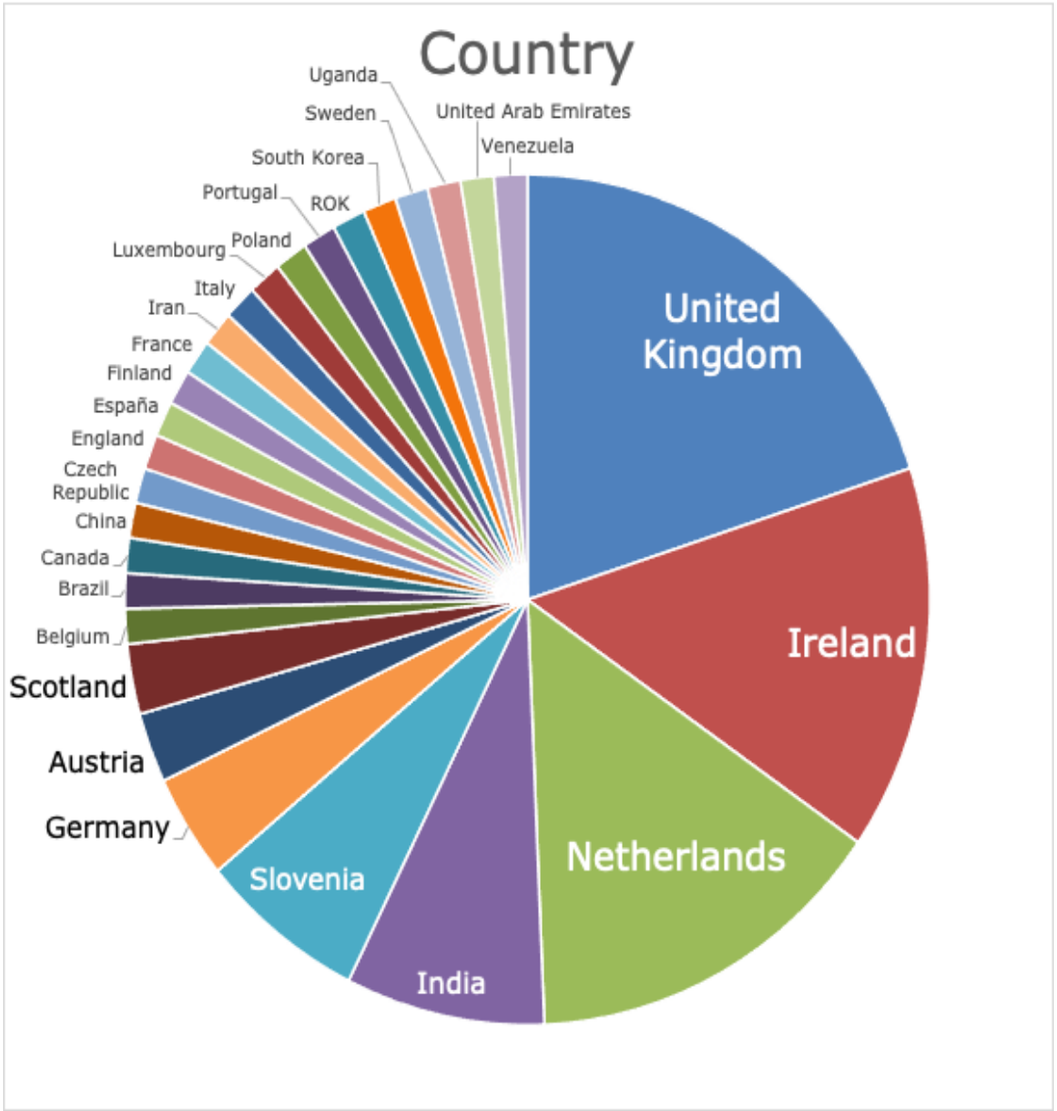
*Longitudinal Internet studies for the Social Sciences

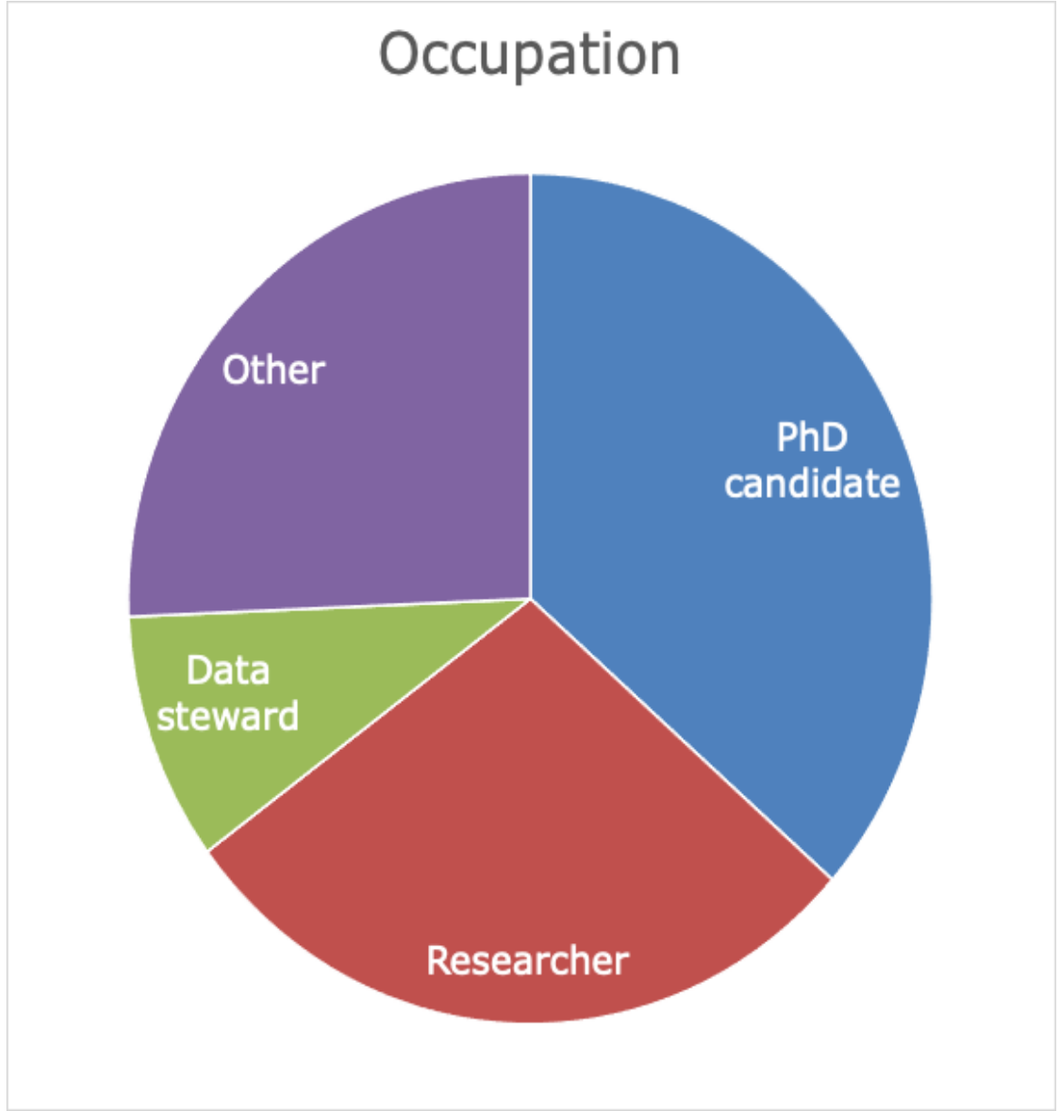
https://www.cessda.eu/

- Consortium European Social Science Data Archives
- European Research Infrastructure Consortium
    - DANS is the Dutch national service provider

- Providers data services to the social sciences
    - CESSDA Data Catalogue
    - CESSDA Training → Data Management Expert Guide

A bit about yourself

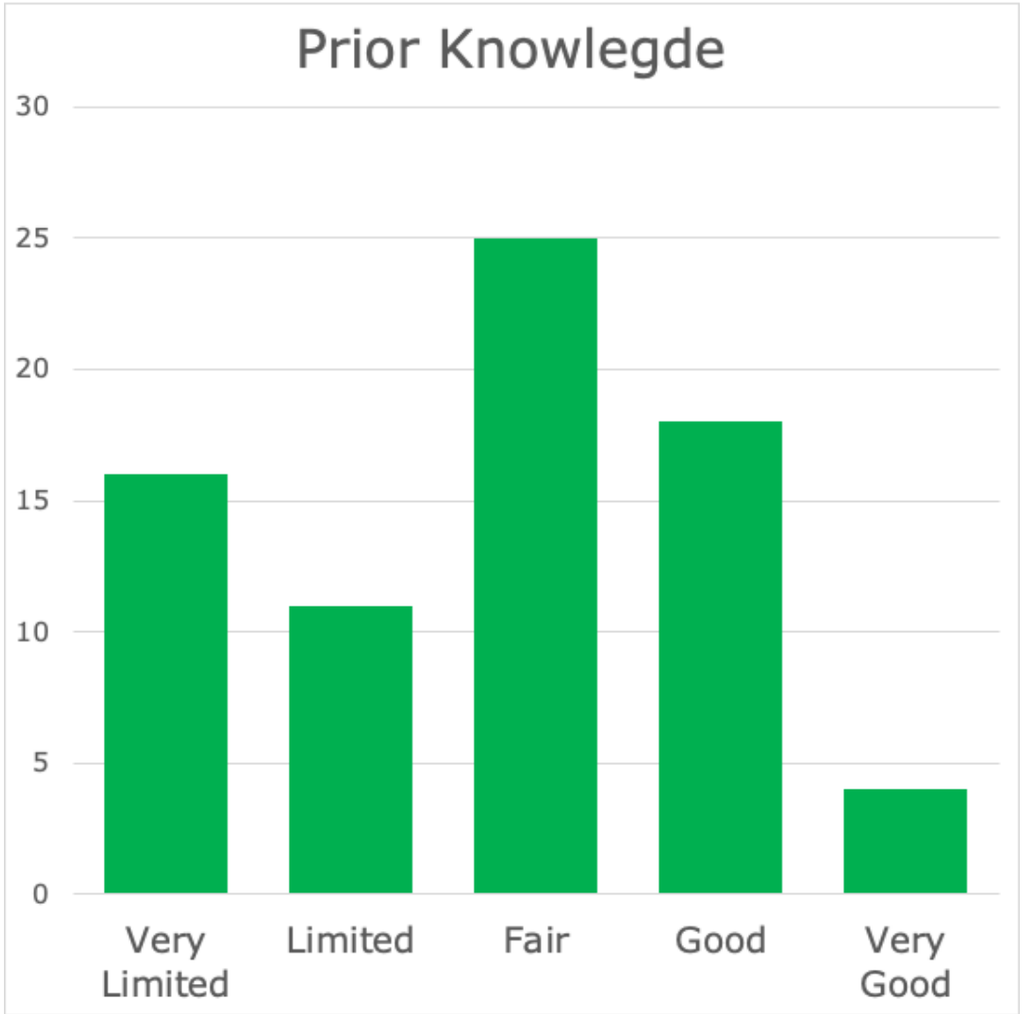


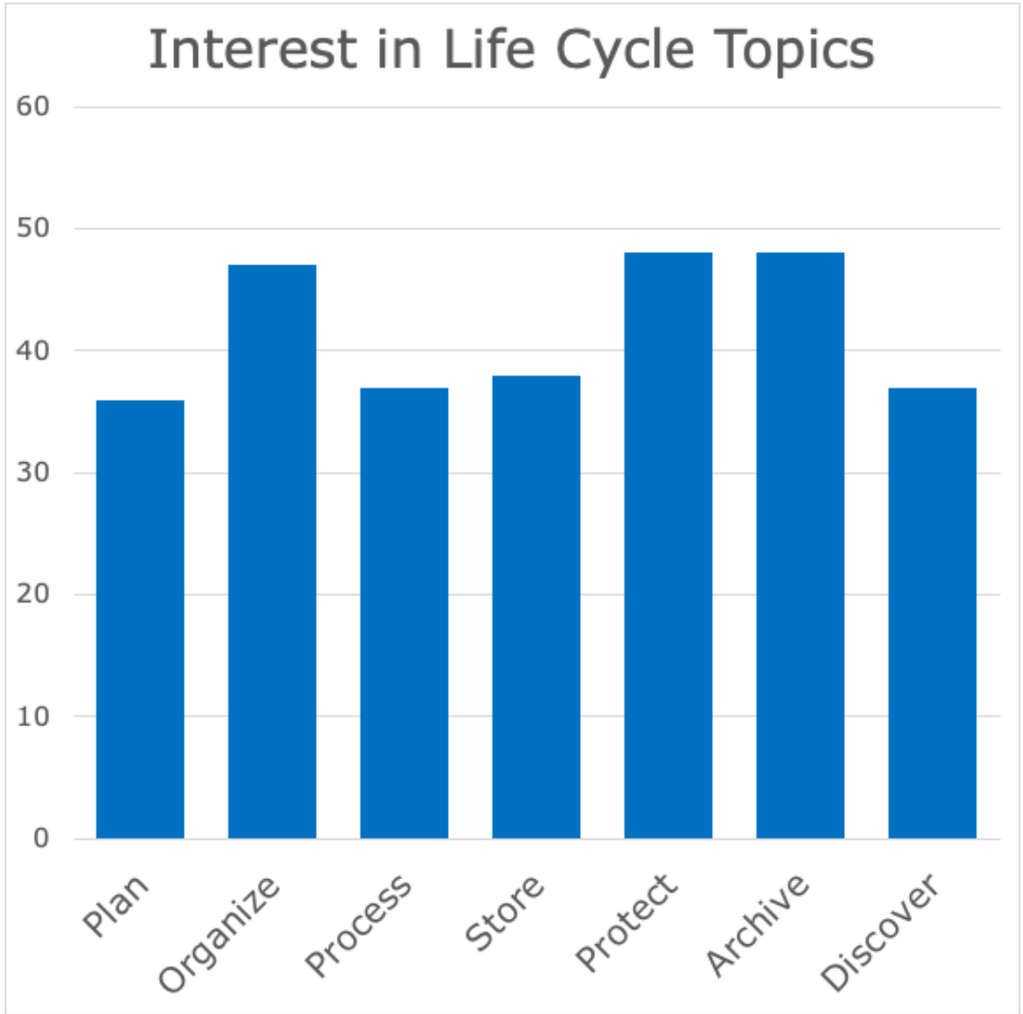


ODISSEI

cessda

Data Archiving and Networked Services
DANS



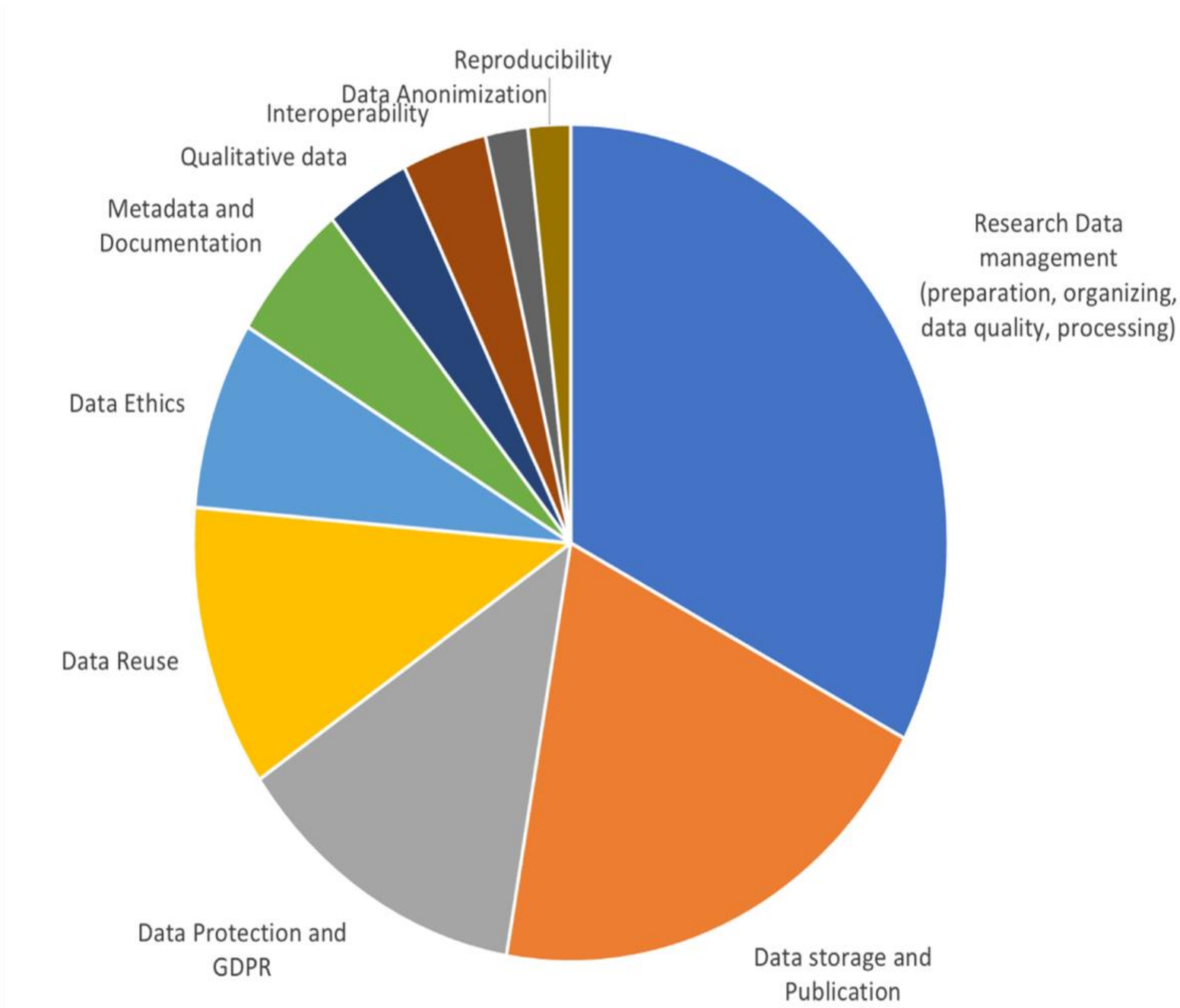


ODISSEI

cessda

Data Archiving and Networked Services
DANS

What are the challenges in RDM?

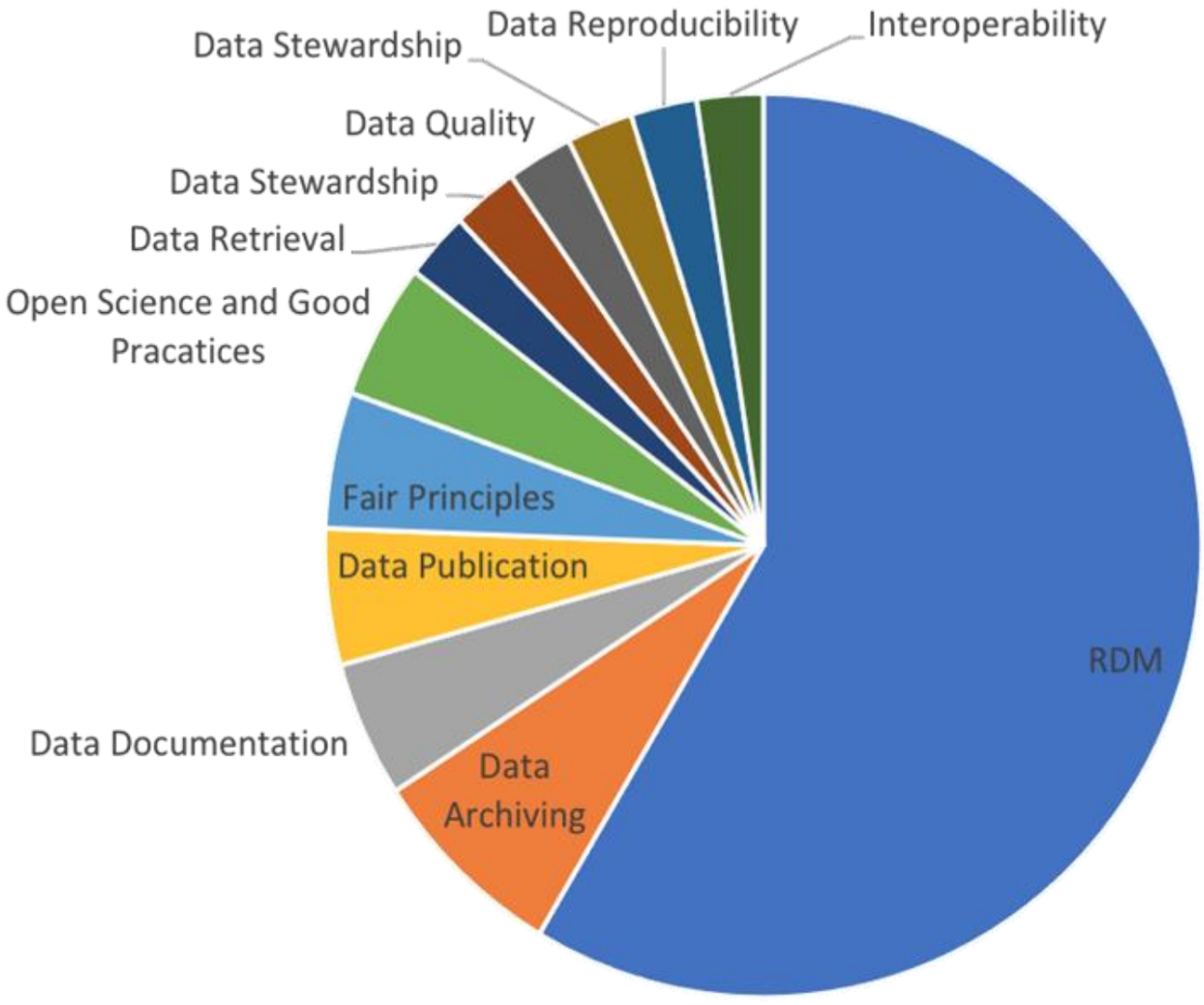


ODISSEI

cessda

Data Archiving and Networked Services
DANS

What would you like to learn?



ODISSEI

cessda

Data Archiving and Networked Services
DANS

What we won't cover

- Qualitative data (DMEG covers this aspect)
- The content of the data themselves (e.g. data completeness)

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Workshop's programme

Monday 14th June 10:00-12:30 CEST

10:00 - 10:15 Welcome
10:15 - 11:00 An Introduction to Data Management: Organizing, processing and storing research data - Ricarda Braukmann (DANS)
11:00 - 11:15 Coffee break
11:15 - 12:00 Data protection and GDPR - Emilie Kraaijkamp & Ellen Leenarts (DANS)
12:00 - 12:15 Q&A
12:15 - 12:30 Introduction to the homework - Francesca Morselli (DANS)

In between, we will ask you to work on the homework assignment.

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Workshop's programme

Thursday 17th June 10:00-12:30 CEST

10:00 - 10:05 Welcome
10:05 - 10:35 Working with administrative data: an ODISSEI case study - Tom Emery & Kasia Karpinska (ODISSEI)
10:35 - 10:45 Coffee break
10:45 - 11:45 Group work in breakout sessions
11:45 - 12:15 An Introduction to Data Management: Archiving and discovering data - Ricarda Braukmann (DANS)
12:15 - 12:30 Closing

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Introduction to Data Management

Organizing, processing and storing data

Ricarda Braukmann

(DANS)

14th June 2021

Why Research Data Management?

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Why Research Data Management?

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Why Research Data Management?

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Why Research Data Management?

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Why Research Data Management?

Easily find and understand data

Increase impact

Make research reproducible

Increase reuse potential

Comply with funder mandates

PUBLICATIONS AND DATA

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Why Research Data Management?

Findable **A**ccesible **I**nteroperable **R**eusable

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Why Research Data Management?

Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accesible

Interoperable

Reusable

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Why Research Data Management?

Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accesible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable

Reusable

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Why Research Data Management?

Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accesible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

Reusable

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Why Research Data Management?

Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accesible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

Reusable

(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.

ODISSEI

cessda

Data Archiving and Networked Services
DANS

The Data Management Expert Guide

ODISSEI

cessda

Data Archiving and Networked Services
DANS

The Data Management Expert Guide

- Guide on Research Data Management (RDM)
  - For early career researchers (in the social sciences)
- Provides useful information on RDM in one central place
- Created by CESSDA training team 2017-2020
- Free to use at **www.cessda.eu/dmeg**

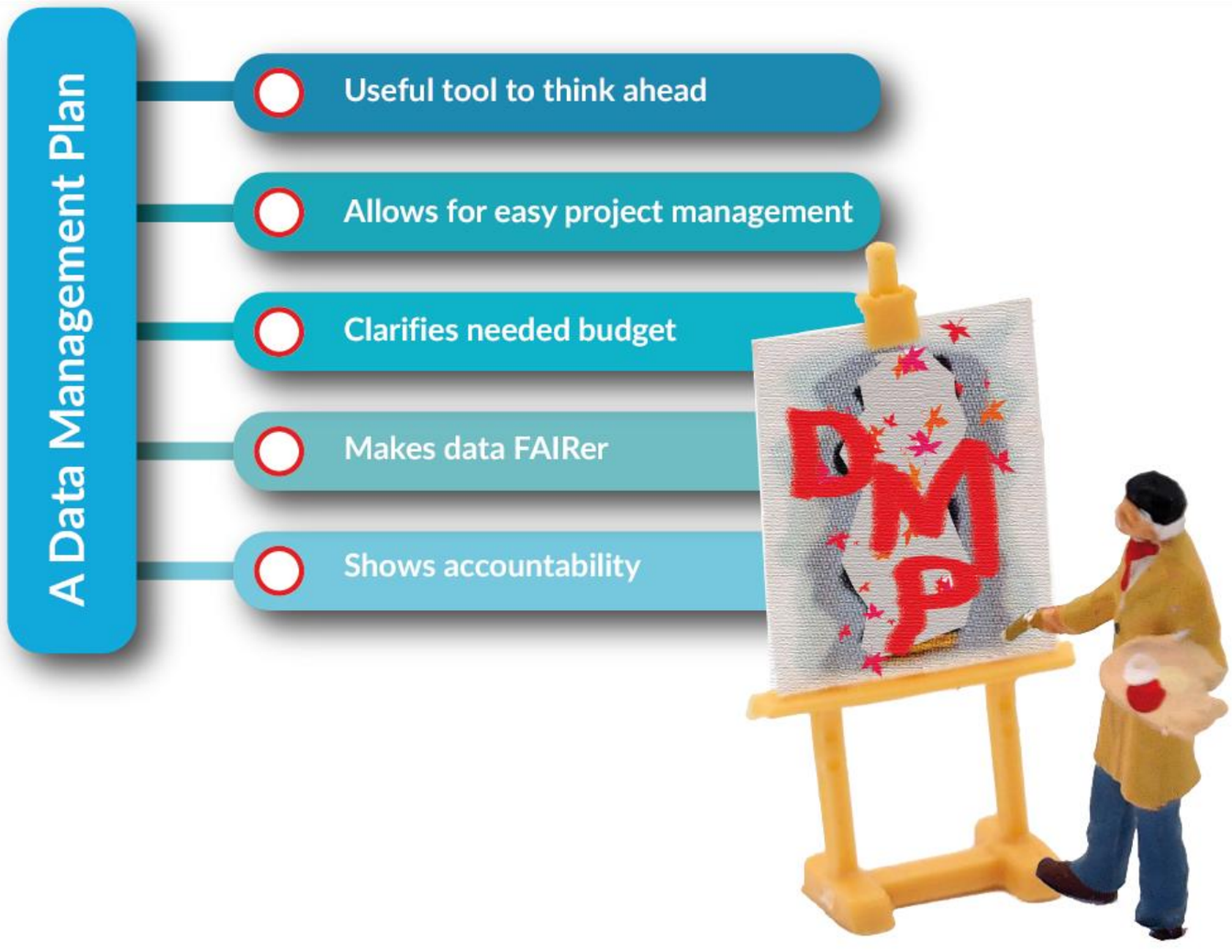
ODISSEI

cessda

Data Archiving and Networked Services
DANS

Helps you to create a DMP

- Data Management Plan checklist to answer questions for your own study

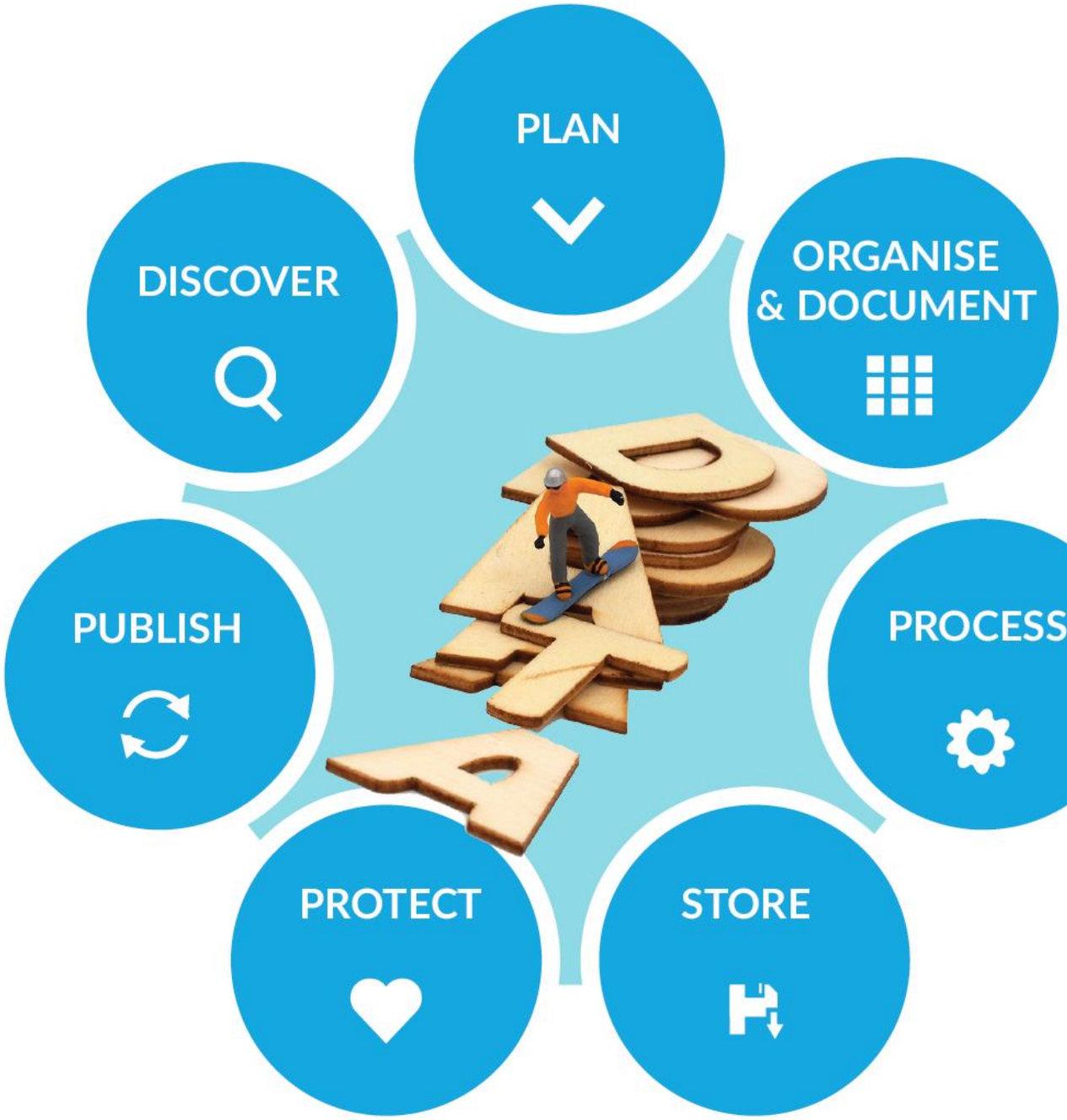


ODISSEI

cessda

Data Archiving and Networked Services
DANS

Following the Life Cycle



ODISSEI

cessda

Data Archiving and Networked Services
DANS

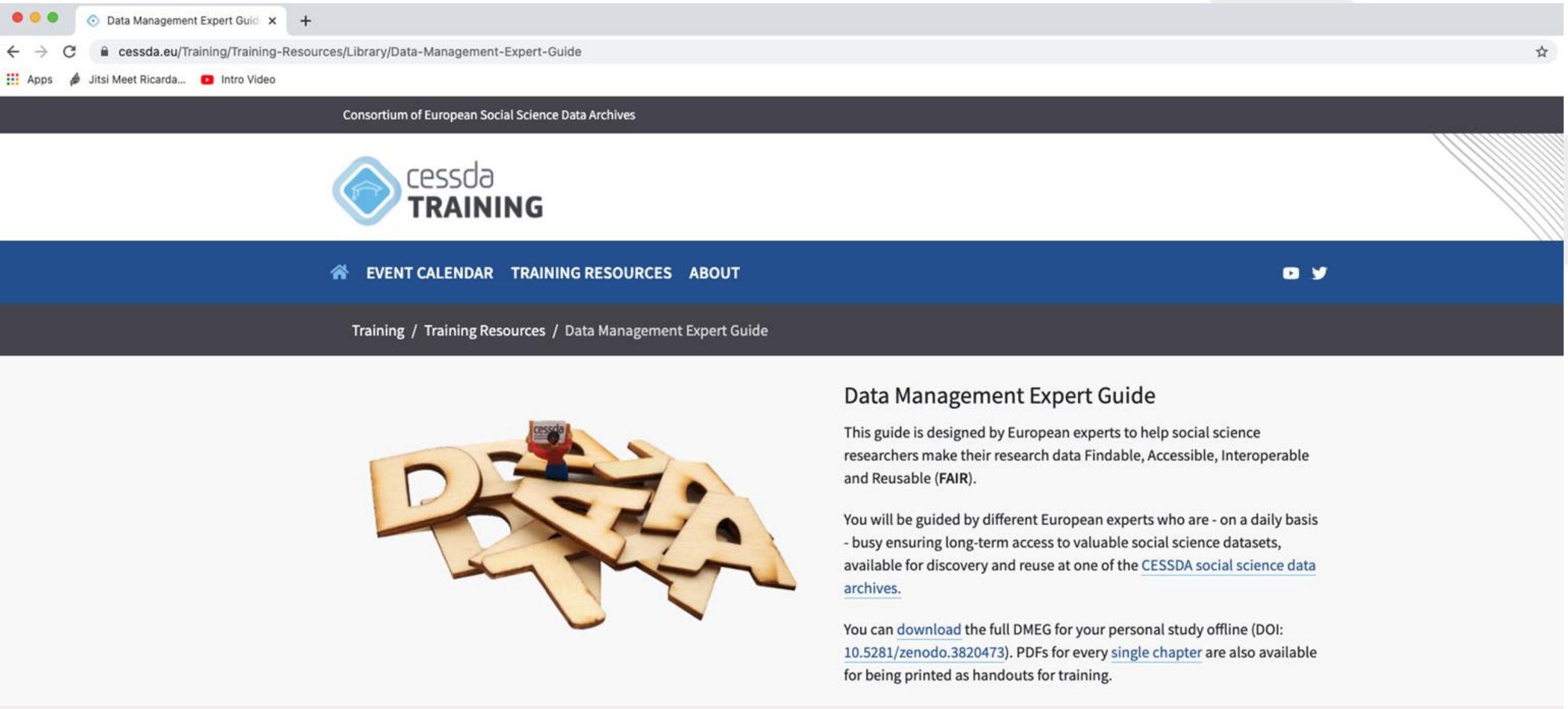
A closer look at the DMEG

ODISSEI

cessda

Data Archiving and Networked Services
DANS

A closer look at the DMEG



ODISSEI

cessda

Data Archiving and Networked Services
DANS

A closer look at the DMEG

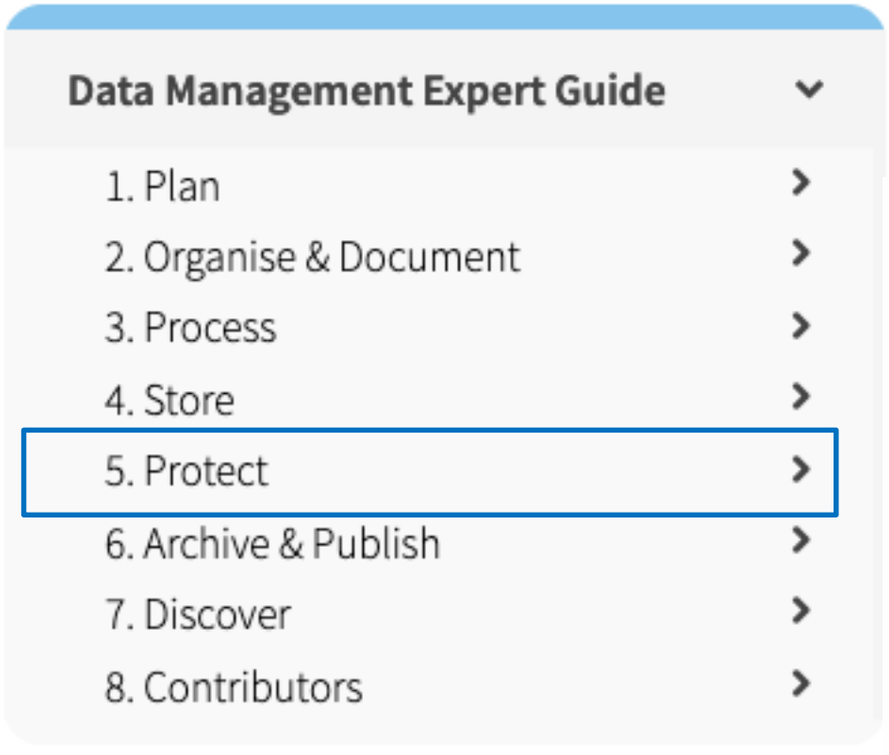
This presentation

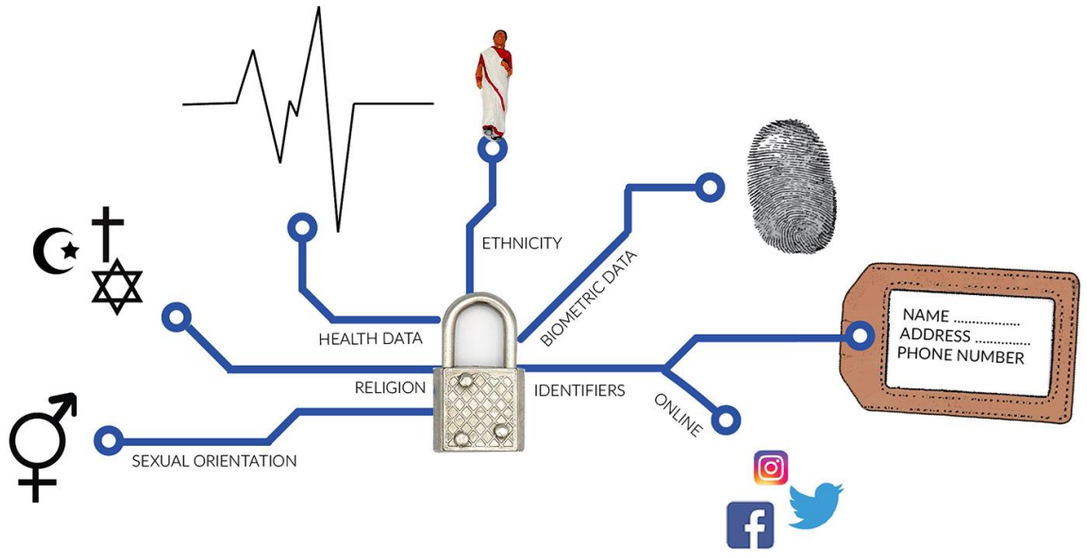
ODISSEI

cessda

Data Archiving and Networked Services
DANS

A closer look at the DMEG





ODISSEI

cessda

Data Archiving and Networked Services
DANS

A closer look at the DMEG

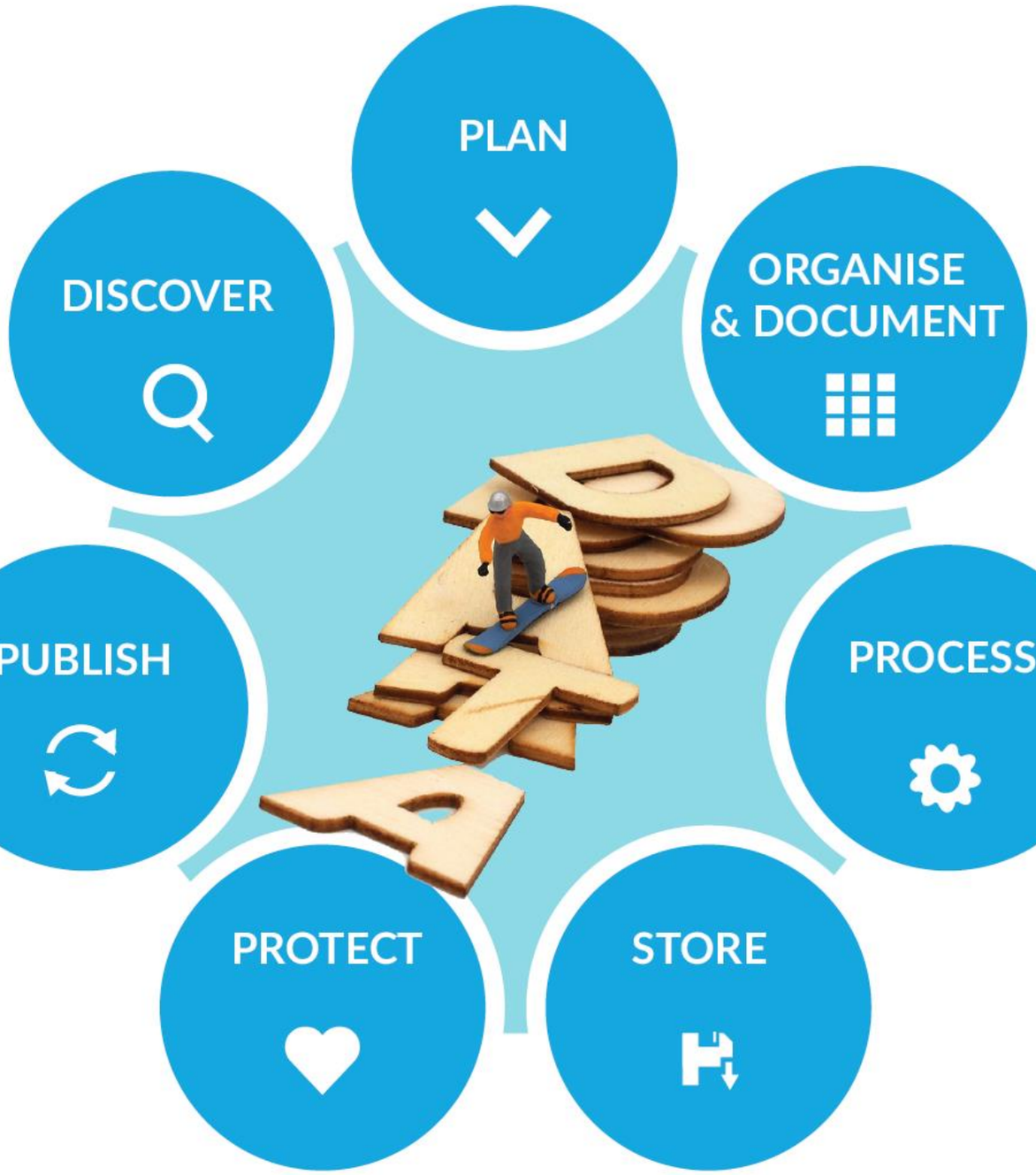
Thursday's presentation

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Plan



ODISSEI

cessda

Data Archiving and Networked Services
DANS

Plan

- Describe the project and the data you will collect
- Collect all basic information
- Contact details + ORCIDs* of involved researchers
- Grant information

**ORCID is a persistent identifier (PID) for researchers. Learn more about ORCIDs in this video and PIDs in general in this video*

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Plan

- Collect all basic information
  - Roles and responsibilities

⊖ Project data contact

Who can be contacted about the project during and 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?

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Plan

- Collect all basic information
  - Think about the costs for data storage and management

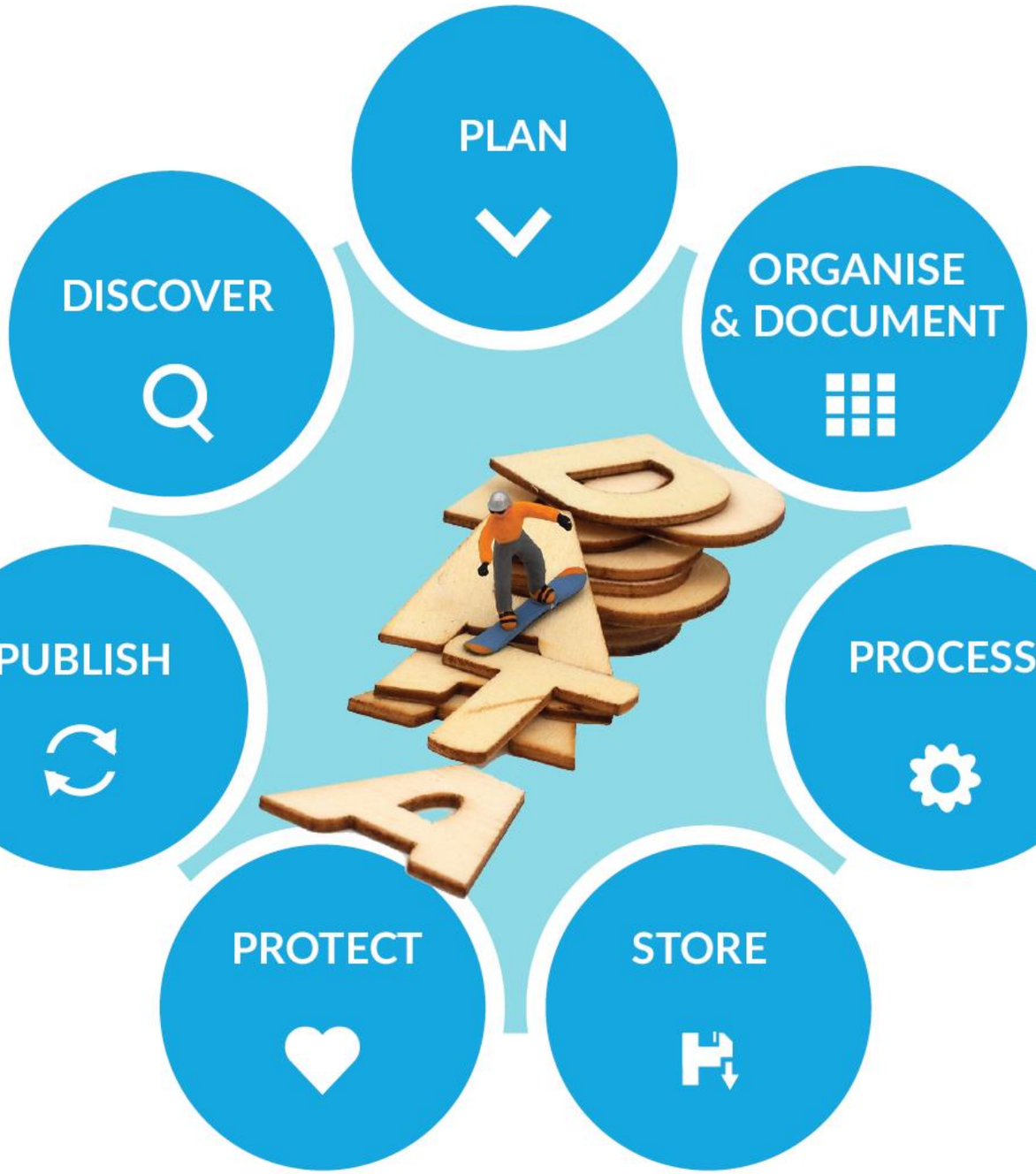
> ⊖ Costs and Resources
>
> - Are there costs you need to consider to buy specific software or hardware?
> - Are there costs you need to consider for storage and backup?
> - Are potential expenses and resources for (preparing the data for) archiving covered?
> - What resources will be dedicated to data management ensuring that data will be FAIR?

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Organise & Document



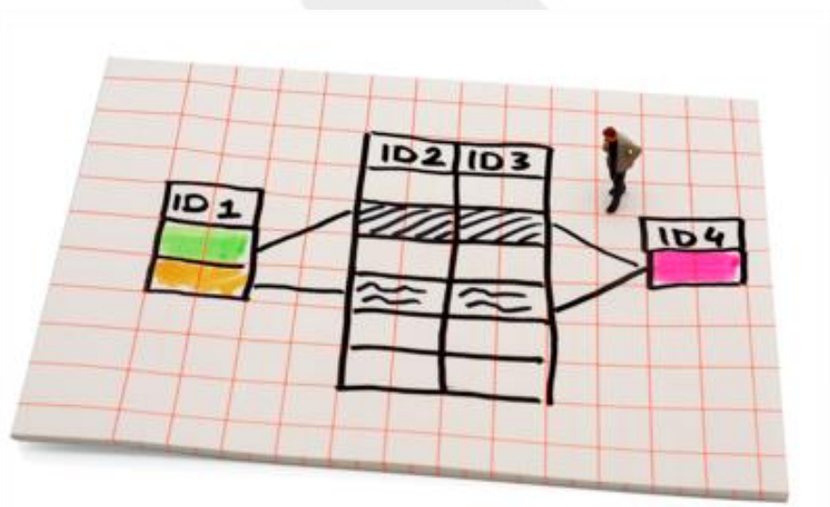
ODISSEI

cessda

Data Archiving and Networked Services
DANS

Organize

- Think about file structures
  - Depends on
    - type of data
    - analysis software
    - Linking with other datasets



- DMEG gives examples for qualitative and quantitative data

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Organize

- File naming

Best practice

According to the UK Data Archive (UK Data Service, 2017b), a best practice in naming files is to:

- Create meaningful but brief names;
- Use file names to classify types of files;
- Avoid using spaces, dots and special characters (& or ? or !);
- Use hyphens (-) or underscores (_) to separate elements in a file name;
- Avoid very long file names;
- Reserve the 3-letter file extension for application-specific codes of file format (e.g. .doc, .xls, .mov, .tif);
- Include versioning of file names where appropriate.

Documenting data file conventions

An example of how to document the data file conventions you use:

<date><type><ID1><gender><age><municipality><datatype><ID2>

where:

- <date>is the date on which the data were collected (date format should be YYYY-MM-DD);
- <type> specifies the type of event/data material;
- <ID1> is the ID of the collection event;
- <gender> is the gender of the interviewee;
- <age> is the age of the interviewee;
- <municipality> is the municipality of residence of the interviewee;
- <datatype> specifies the type of data the file contains, for instance, "trans" means transcription, "audio" means audio recording, and "image" means photograph;
- <ID2> is the ID number used to separate the images connected to the collection event.

Organize

- File naming

Best practice

According to the UK Data Archive (UK Data Service, 2017b), a best practice in naming files is to:

- Create meaningful but brief names;
- Use file names to classify types of files;
- Avoid using spaces, dots and special characters (& or ? or !);
- Use hyphens (-) or underscores (_) to separate elements in a file name;
- Avoid very long file names;
- Reserve the 3-letter file extension for application-specific codes of file format (e.g. .doc, .xls, .mov, .tif);
- Include versioning of fi here appropriate.

Documenting data file conventions

An example of how to document the data file conve

<date><type><ID1><gender><age><mu

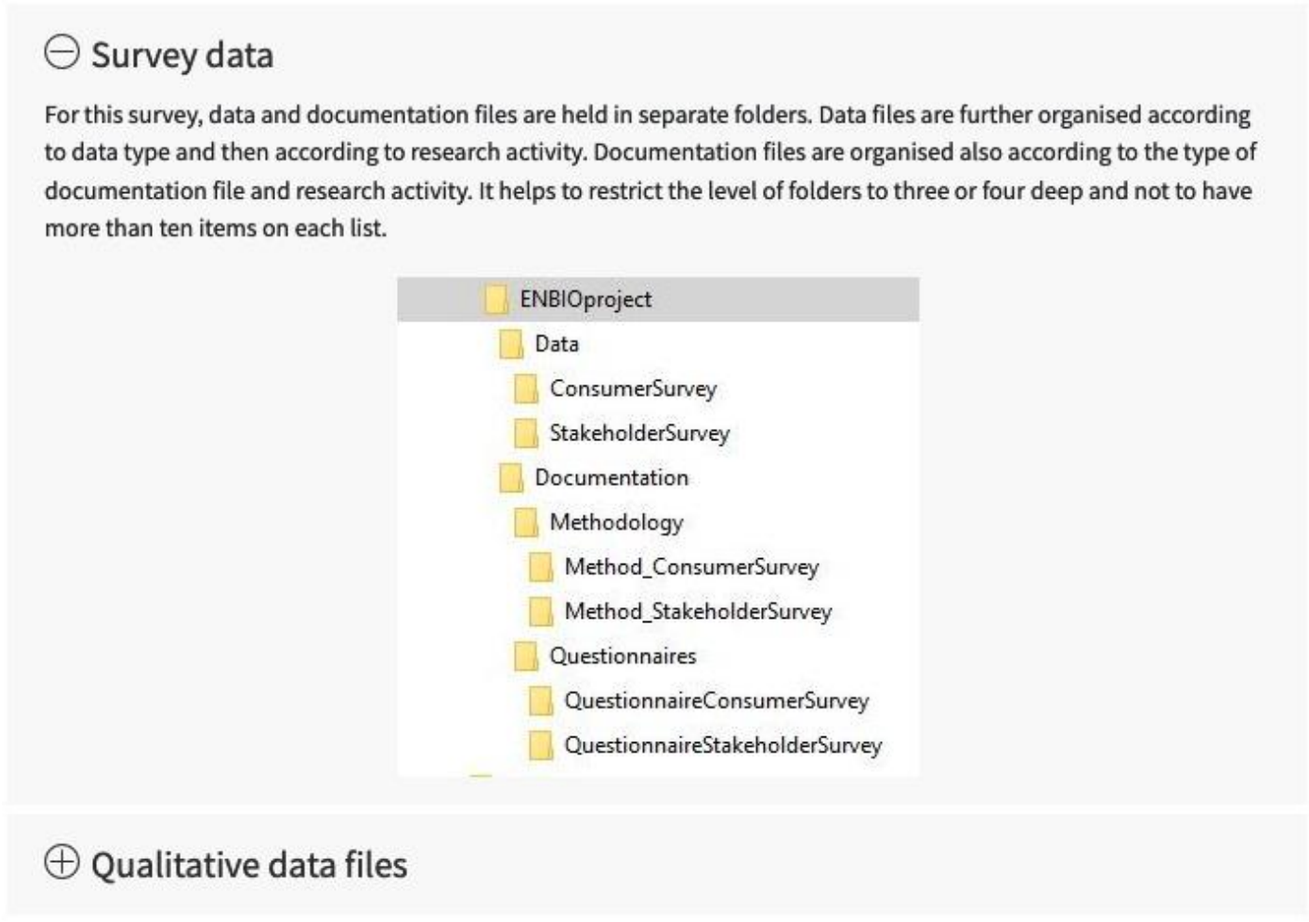
where:

- <date>is the date on which the dat were collected (date format should be YYYY-MM-DD);
- <type> specifies the type of event/data material;
- <ID1> is the ID of the collection event;
- <gender> is the gender of the interviewee;
- <age> is the age of the interviewee;
- <municipality> is the municipality of residence of the interviewee;
- <datatype> specifies the type of data the file contains, for instance, "trans" means transcription, "audio" means audio recording, and "image" means photograph;
- <ID2> is the ID number used to separate the images connected to the collection event.

Do not include personal information in filenames!

Organize

- Folder structures
  - Depends on
    - Project structure
    - Data collected
    - Collaborations



- DMEG gives examples for qualitative and quantitative data

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Organize

Metadata is defined as the data providing information about one or more aspects of the data

- Metadata and documentation

Project level details

⊕ 1. For what purpose was the data created

⊕ 2. What does the dataset contain?

⊕ 3. How was data collected?

⊕ 4. Who collected the data and when?

⊕ 5. How was the data processed?

⊕ 6. What possible manipulations were done to the data?

⊕ 7. What were the quality assurance procedures?

⊕ 8. How can the data be accessed?

Data level details

- **Information about the data file**
  Data type, file type, and format, size, data processing scripts.
- **Information about the variables in the file**
  The names, labels and descriptions of variables, their values, a description of derived variables or, if applicable, frequencies, basic contingencies etc. The exact original wording of the question should also be available. Variable labels should:
  - Be brief with a maximum of 80 characters;
  - Indicate the unit of measurement, where applicable;
  - Reference the question number of a survey or questionnaire, where applicable.

→ DMEG gives examples for qualitative and quantitative data

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Organize

- Metadata and documentation → Crucial for FAIR

Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accesible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

Reusable

(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Process

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Process

- Minimize errors in data processing → DMEG examples
  - Information on coding
  - Information on weights

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 of manual data entry
- ⊕ Minimise the number of steps
- ⊕ Conduct data entry twice
- ⊕ Perform in-depth checks for selected records
- ⊕ Perform logical and consistency checks
- ⊕ Automate checks whenever possible

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Process

- File formats
  - Can I open this in 5 years from now?
    - → Some formats are more sustainable than others
- Data archives transform data for you!

vs

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Process

- Data authenticity and versioning
  - Know what happened to the data to ensure high quality
  - Document changes

→ DMEG gives tips on versioning and logging changes

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Store

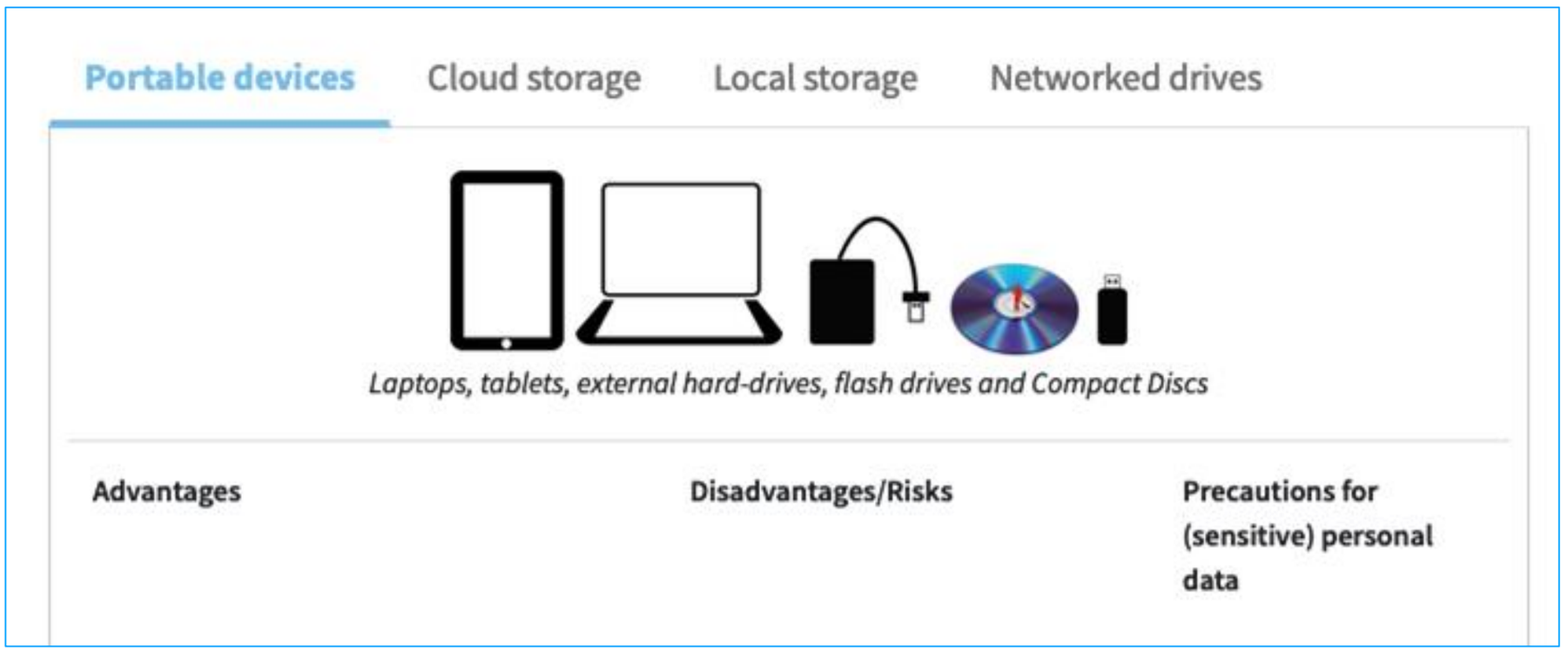
ODISSEI

cessda

Data Archiving and Networked Services
DANS

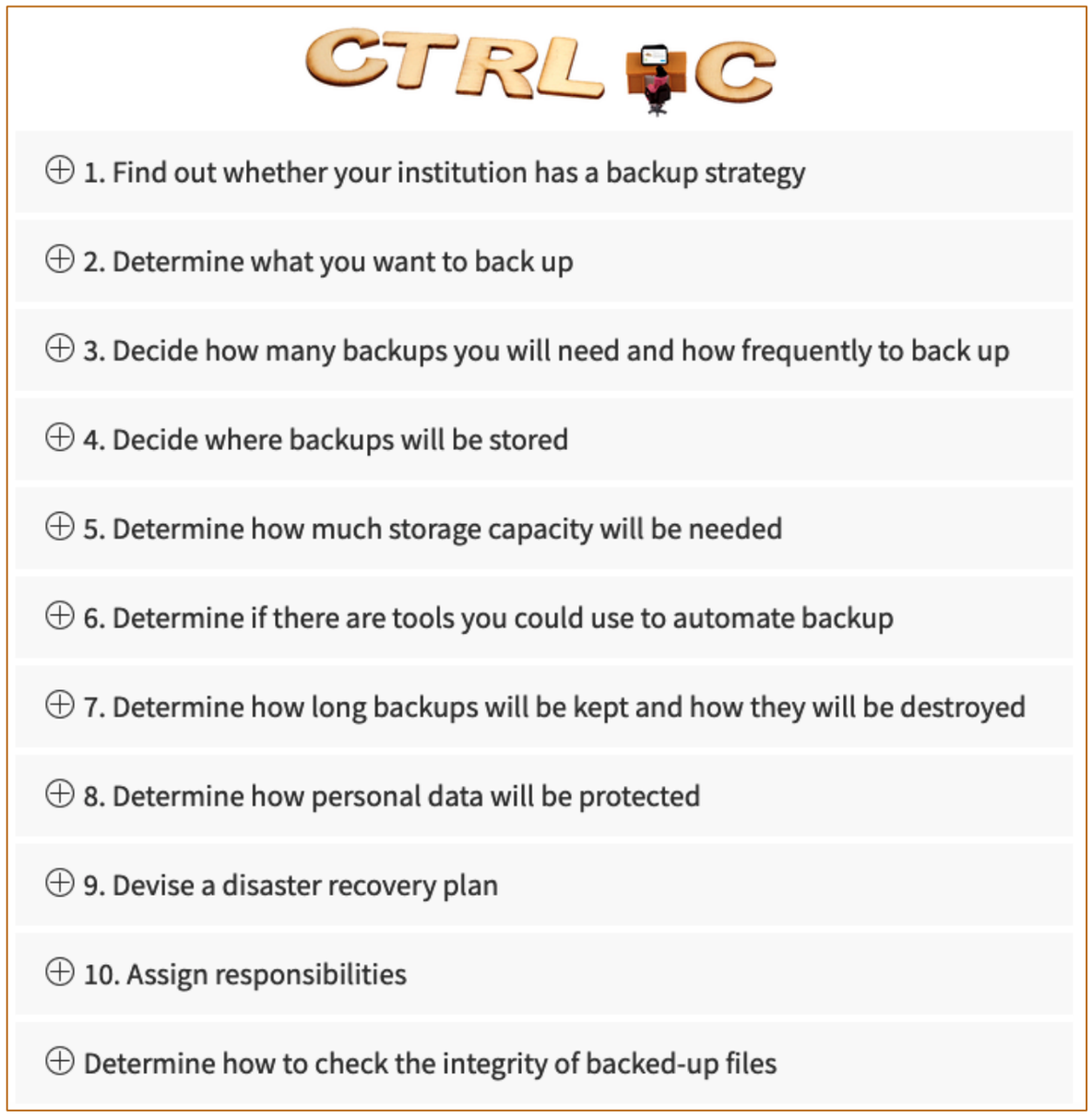
Store

- **Storage** is often organized by your institute → Find out your policies
    - How much space do you need?
    - Who needs access?



Store

- **Backup** is often organized by your institute → Find out your policies



Store

- **Security** is crucial, in particular when working with (sensitive) personal data

⊕ Passwords

⊕ Encryption

⊕ Physical, network and computer security

⊕ Secure disposal

⊕ Organisational aspects

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Store

How do you share (personal) data with your collaborators?

ODISSEI

cessda

Data Archiving and Networked Services
DANS

Thank you for your attention!

ricarda.braukmann@dans.knaw.nl

This workshop is organized by CESSDA-ERIC, ODISSEI and DANS.