ADP - SOCIAL SCIENCE DATA ARCHIVES

Analyze data! Deposit study! Promote science!

Research Data Management following FAIR principles

Bezjak Sonja & Vipavc Brvar Irena

Workshop for researchers and doctoral students at the Faculty of Public Administration, University of Ljubljana, 9. 5. 2023

Content

16.30 - 17.15

Introduction

About ADP & CESSDA

Data Management Planning

- what is data
- FAIR principles
- what is RD life-cycle
- data discovery

17.15-18.00

DMP Q&A

- research data management
- DMEG chapters (interactive)

Aims of this lecture

1) Participants ***understand the concepts of open science***: "open data", "FAIR principles", "research data lifecycle", "research data management", "data publication", "data citation"

2) Participants understand **the basics of data management planning** adopted for social sciences

CESSDA Training Team (2017 - 2022). *CESSDA Data Management Expert Guide.* Bergen, Norway: CESSDA ERIC. Retrieved from https://dmeg.cessda.eu/

Slovenian Social Science Data Archives (*ADP-Arhiv družboslovnih podatkov*)

- Founded in 1997 → 25th anniversary
- Slovenian **national research data centre** for social sciences
- **Member of CESSDA** ERIC since 2017
- Status of a **trust-worthy archive** (CoreTrustSeal since 2018)
- involved in EU and national projects

CORE TRUST SEAL

ADP Social Science Data Archives
http://www.adp.fdv.uni-lj.si/eng/
CTS Certification 2017-2019

ADP's mission

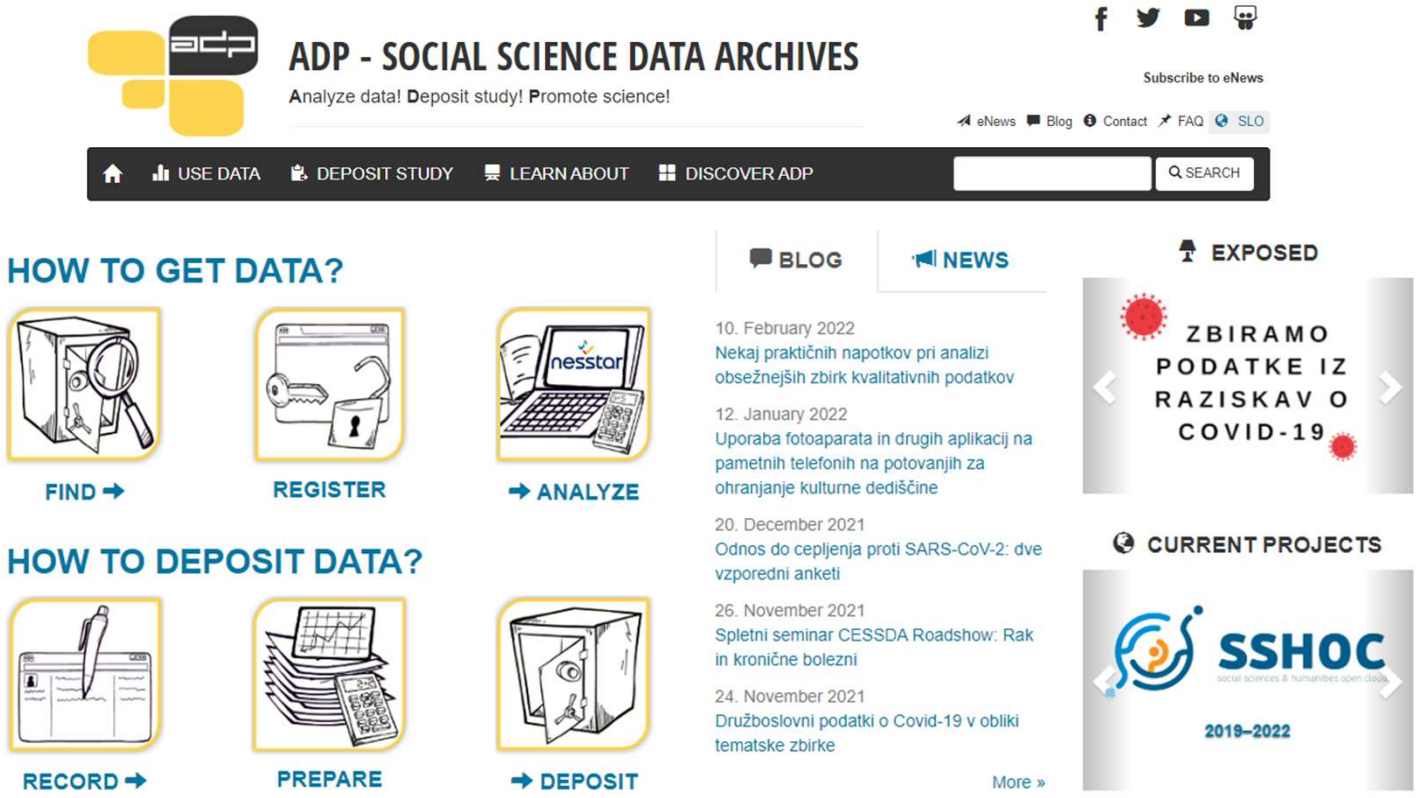
To ensure and **promote** *sustainable services* of **ingest, storage and access** to *quality research data from the field of Slovenian social sciences* and broader, with *potential for secondary analysis*.

Main services:

- **Acquiring** important research data from a wide range of social sciences
- **Appraisal** of submitted research data and their **selection** for deposit **Ingesting and processing** research data and other documentation, together with the creation of metadata
- Long-term digital **preservation** (AIP), **access** and **re-use** for scientific, educational and other purposes (DIP)
- **Training** researchers on:
    - research data management
    - re-use of research data
- **Promotion** of open data and open science (students, librarians, journals, citizens…)

https://www.adp.fdv.uni-lj.si/eng/spoznaj/adp/poslanstvo/

Slovenian national research data centre *for social sciences*

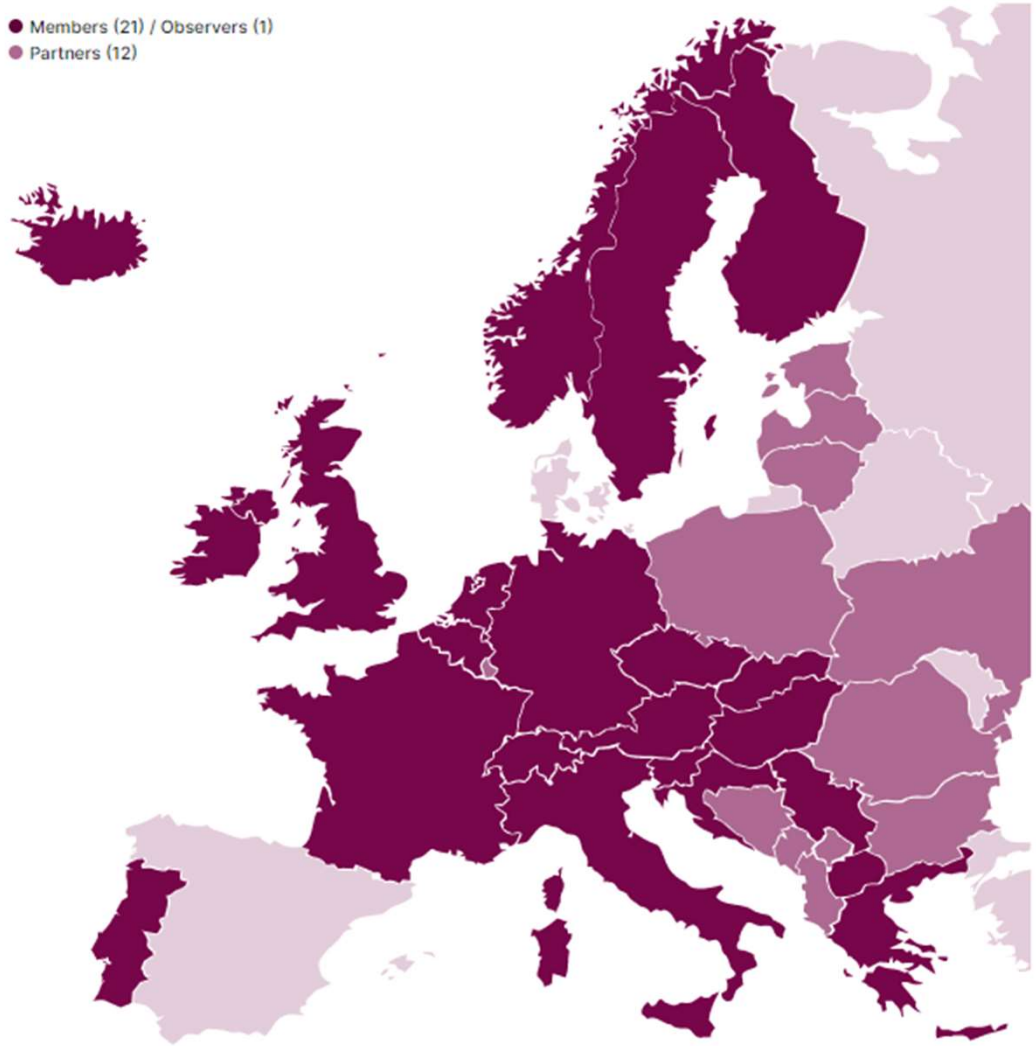


https://www.adp.fdv.uni-lj.si/eng/

QUICK FACTS

- **775 social science studies** research data accessible in a data catalogue + 150 metadata only
- **1000 users registered per year** (90 % education, 10 % scientific/research purpose)
- **500 units of research data reused** for detailed secondary-analysis per year

Consortium of European Social Science Data Archives (CESSDA)



"Member countries seek to ***increase the scientific excellence and efficacy of European research in the social sciences"***

Key tasks:

Developing **standards and best practices** around the management and archiving of social science data.
Facilitating access to important data resources
Work done by **developing tools, training and co-ordinating network**.
CESSDA data catalogue.
(https://datacatalogue.cessda.eu/)

Getting to know each other

Savvas Stavrinos

All participants, please, write in a chat

1) What is your research topic?

2) What kind of data you plan to (re)use?

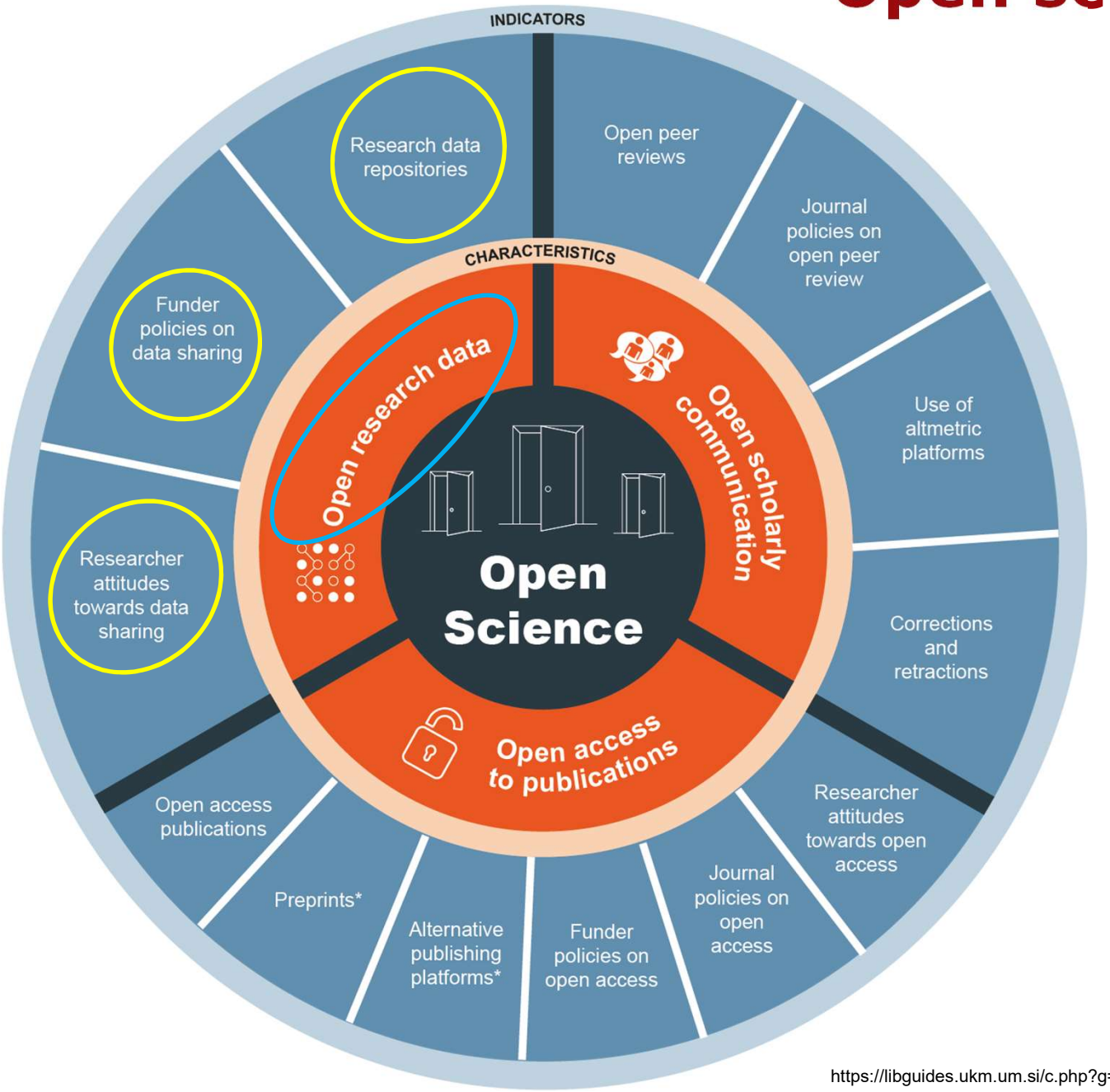
Open Science Definition

Open Science is the practice of science in such a way that others can **collaborate and contribute**, where research data, lab notes and other research processes are **freely available**, under **terms that enable reuse, redistribution and reproduction** of the **research** and its underlying **data** and **methods**.

(FOSTER Open Science)

https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition

Open science



https://libguides.ukm.um.si/c.php?g=654719&p=4601102

Open science & Open data

Benefits

- Data publication may lead to **increased visibility, reuse** and **citation** and therefore **recognition of scholarly work**.

- Be aware that *whenever you use the published data you are obliged to cite them*. For more information see the paragraph on data citation.

- **Career**
- **Scientific progress**
- **Publishers**
- **Funder requirements**
- **Organizational demands**

CESSDA Training Team (2017 - 2022). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. Retrieved from https://dmeg.cessda.eu/

Example 1: Horizon Europe

Proper Research Data Management (RDM) is mandatory for any Horizon Europe project generating or reusing research data. It is a key part of Horizon Europe's open science requirements.

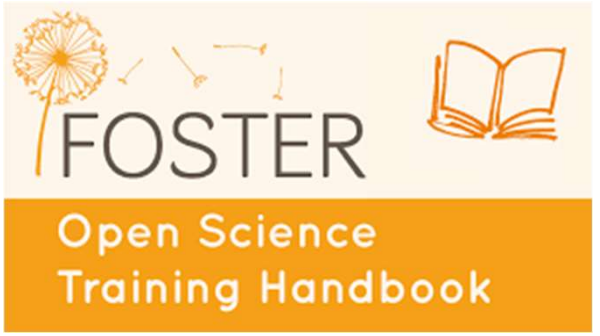
In Horizon Europe, *beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the* ***FAIR principles***, and should at least do the following:

- Prepare a Data Management Plan (DMP) and keep it updated throughout the course of the project
- Deposit data in a trusted repository and provide open access to it ('as open as possible, as closed as necessary')
- Provide information (via the same repository) about any research output or any other tools and instruments needed to re-use or validate the data

Keep in mind that 'research data' is a very broad concept and certainly not limited to numerical/tabular data.

https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm

FAIR principles



F → **FINDABLE**

It should be ***easy to find the data and the metadata*** for both humans and computers. Automatic and reliable discovery of datasets and services depends on machine-readable persistent identifiers (PIDs) and metadata.

https://www.fosteropenscience.eu/content/open-science-training-handbook

FAIR principles

A → **Accessible**

The (meta)data should be ***retrievable by their identifier using a standardized and open communications protocol***, possibly including authentication and authorisation. Also, metadata should be available even when the data are no longer available.

https://www.fosteropenscience.eu/content/open-science-training-handbook

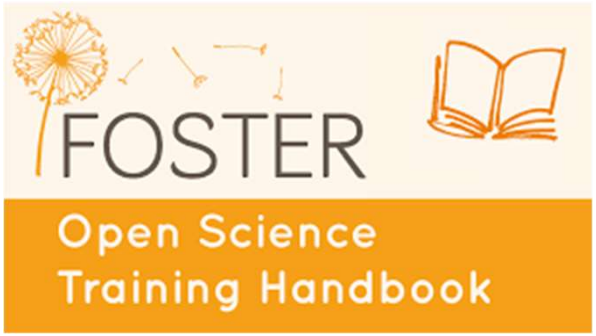
FAIR principles

I → **Interoperable**

The data should be able to be combined with and used with other data or tools. ***The format of the data should therefore be open and interpretable for various tools,*** including other data records. The concept of interoperability applies both at the data and metadata level. For instance, the (meta)data should use vocabularies that follow FAIR principles.

https://www.fosteropenscience.eu/content/open-science-training-handbook

FAIR principles



R → **Re-usable**

Ultimately, FAIR aims at optimizing the reuse of data. To achieve this, ***metadata and data should be well-described so that they can be replicated and/or combined in different settings.*** Also, the reuse of the (meta)data should be stated with (a) clear and accessible license(s).

https://www.fosteropenscience.eu/content/open-science-training-handbook

Example 2: University of Ljubljana

Univerza v Ljubljani

Doctoral School University of Ljubljana

A DMP must be submitted by generations of doctoral students enrolled in academic year 2021/2022 and thereafter.

The handling of research data is regulated in the Rules and Regulations for Doctoral Studies at UL (Article 50). (https://www.uni-lj.si/doctoral_school/rules/)

The doctoral student submits a draft DMP

(1) upon registering the doctoral dissertation proposal (see Article 36 of the Rules),

(2) an updated version of the DMP upon presentation of the research results (see Article 43 of the Rules)

(3) upon submission of the dissertation (see Article 45 of the Rules).

https://www.uni-lj.si/doctoral_school/research_data_management/questions

Where will data be stored?

Univerza v Ljubljani

Doctoral School University of Ljubljana

Rules and Regulations for Doctoral Studies at the University of Ljubljana (PDF) in force from 1 October 2021 (https://www.uni-lj.si/doctoral_school/rules/)

The doctoral student submits research data to a data repository, data centre or research data archive. Preferably, research data should be sent to the disciplinary national or international data centres intended for specific types of data. In the fields where there are (still) no disciplinary data centres, the data is submitted to a general data repository or the Repository of the University of Ljubljana. The supervisor advises the doctoral student about the most appropriate repository for their field. It is also important for the doctoral student to consult in advance with the selected data centre regarding the possibilities and conditions for data publishing, as the centre may have its own requirements that the UL DMP has not taken into account.

Big data can be stored in the data archive on the Vega supercomputer via the Repository of the University of Ljubljana. For life sciences, the Slovenian hub ELIXIR Slovenia has set up a research infrastructure that enables the storage of research data, calculations and other functionalities. **The national data centre Social Science Data Archive is available for social sciences and certain types of humanities data. Language-related disciplines can make use of CLARIN.SI – the Slovenian research infrastructure for linguistic resources and technology.**

https://www.uni-lj.si/doctoral_school/research_data_management/questions

Real life experience from ADP

Challenging situations before publishing data

1) I will finalize my thesis next week. I need to publish my data.

2) Data access agreement doesn't allow me to share the variables.

3) I promised participants to use my data only for my PhD thesis.

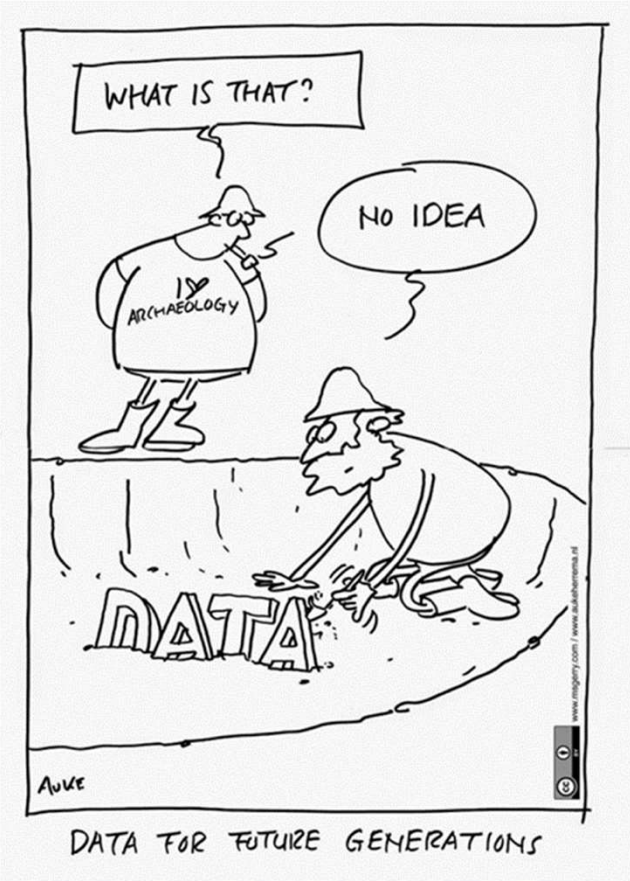
Let's check for solutions!

Iowa State University Library, DMP Guide.
https://instr.iastate.libguides.com/dmp/writingDMP

What is research data…

… primary sources that underpin scientific research and enable derivation of theoretical or applied findings.

(Preparing research data for open access : guide for data producers, 2015)



What is research data…

The tangible forms this 'material' may take are e.g. "***facts, observations, interviews, recordings, measurements, experiments, simulations, and software; numerical, descriptive and visual;*** *raw, cleaned up and processed*"

(Van Berchum & Grootveld, 2017).

Social Sciences

Methods

- Opinion polls
- *Surveys*
- Interviews
- *Mass media, social media*
- Laboratory experiments
- *Field experiments*
- Fieldwork notes
- *Demographic records*
- Census records
- *Voting records*
- Economic indicators

Sources

- Generate your own data
- Obtain it from other researchers
- Data repositories
- Existing records

Sanchez, Candela: DATA MINDFULNESS: MAKING THE MOST OF YOUR DISSERTATION. University of Edinburgh. Research Data Service.
https://libraryblogs.is.ed.ac.uk/datablog/files/2019/08/Data-Mindfulness-Making-the-Most-of-your-Dissertation-handbook-1.pdf

Arts & Humanities

Methods

- Newspapers
- *Photographs, video material*
- Letters
- *Diaries*
- Literature: books, articles
- *Church records*
- Court records
- *Maps*
- Art artefacts
- *Historic artefacts*

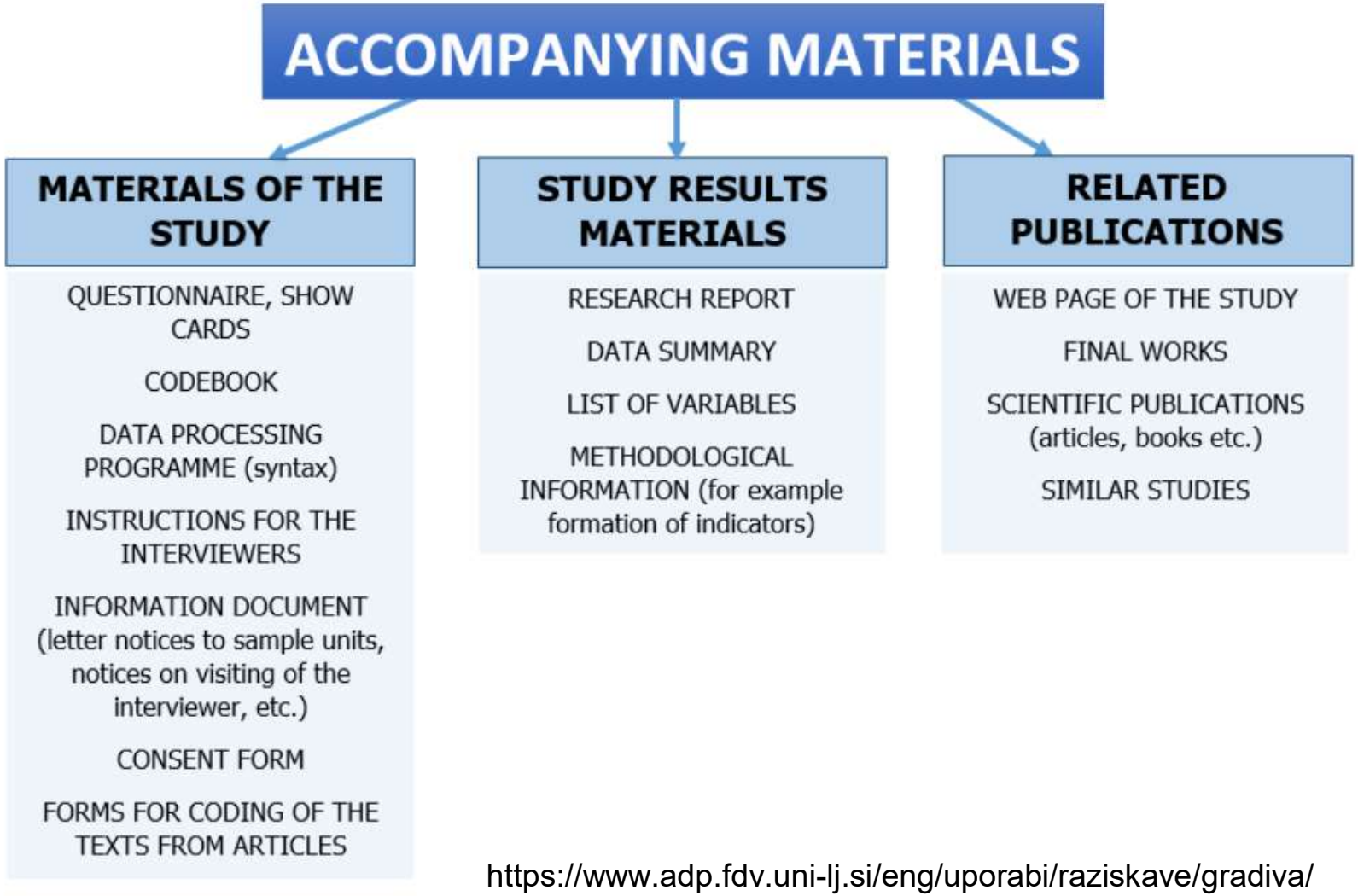
Sources

- Libraries
- Archives
- Museums
  Public/corporate/government records
- Data repositories

Sanchez, Candela: DATA MINDFULNESS: MAKING THE MOST OF YOUR DISSERTATION. University of Edinburgh. Research Data Service.
https://libraryblogs.is.ed.ac.uk/datablog/files/2019/08/Data-Mindfulness-Making-the-Most-of-your-Dissertation-handbook-1.pdf

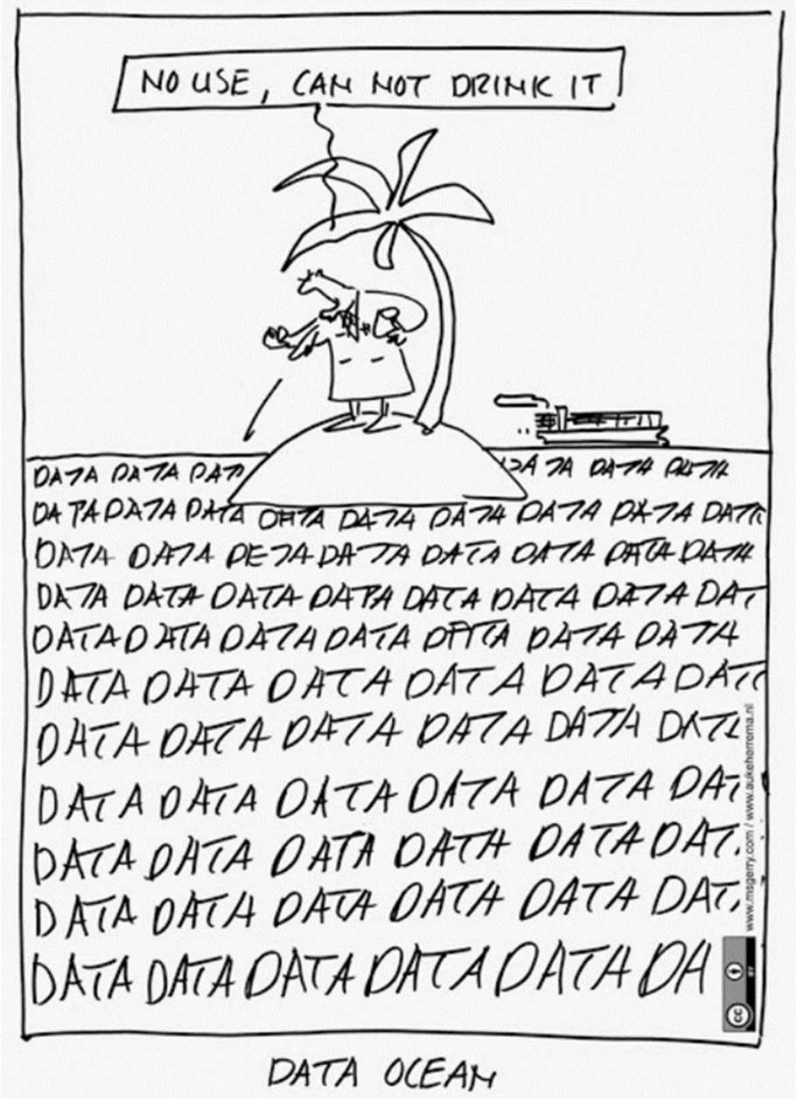
It's all about data but not only ☺

Other materials wanted

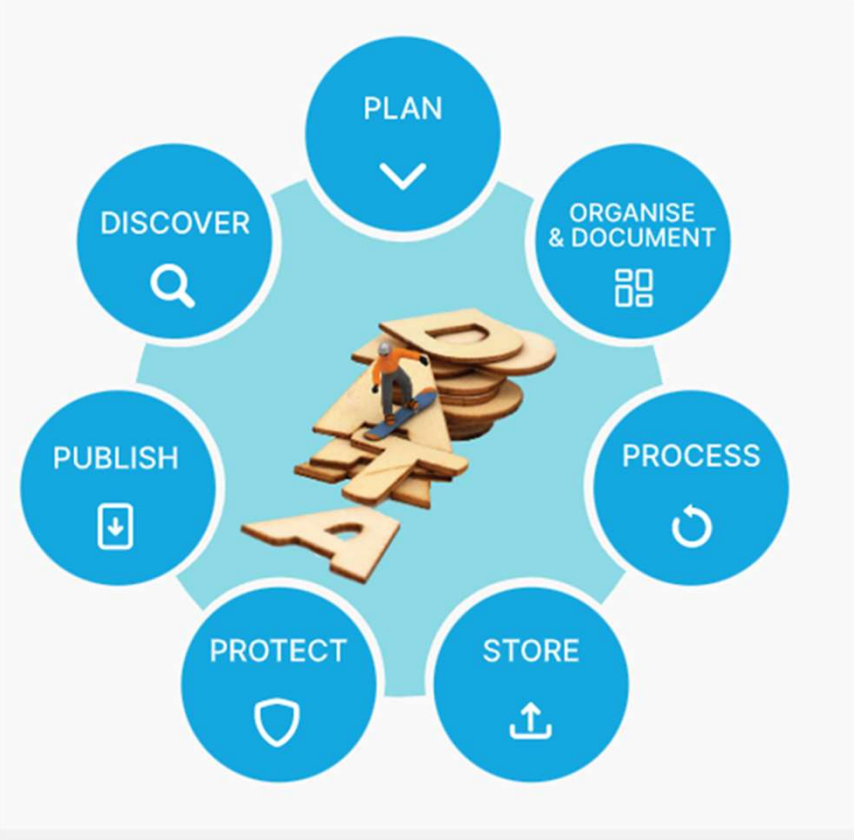


https://www.adp.fdv.uni-lj.si/eng/uporabi/raziskave/gradiva/

Where to start?



Research Data Lifecycle



The research data lifecycle is a model that illustrates the stages of data management and describes how data flow through a research project from start to finish.

(Princeton Research Data Service, https://researchdata.princeton.edu/research-lifecycle-guide/research-lifecycle-guide)

Research data management

… refers to how you ***handle, organise, and structure*** your research data throughout the research process.

… addresses also your plans for the data ***after*** the research is complete.

- A good data management strategy takes into account technical, *organisational, structural, legal, ethical and sustainability aspects*.
- Makes your research *time-efficient*, reproducible and safe as possible, if your data management is well thought through, *structured*, and *documented*.

How to write a DMP

Guide developed by CESSDA Archives

Training / Training Resources / Data Management Expert Guide

Data Management Expert Guide

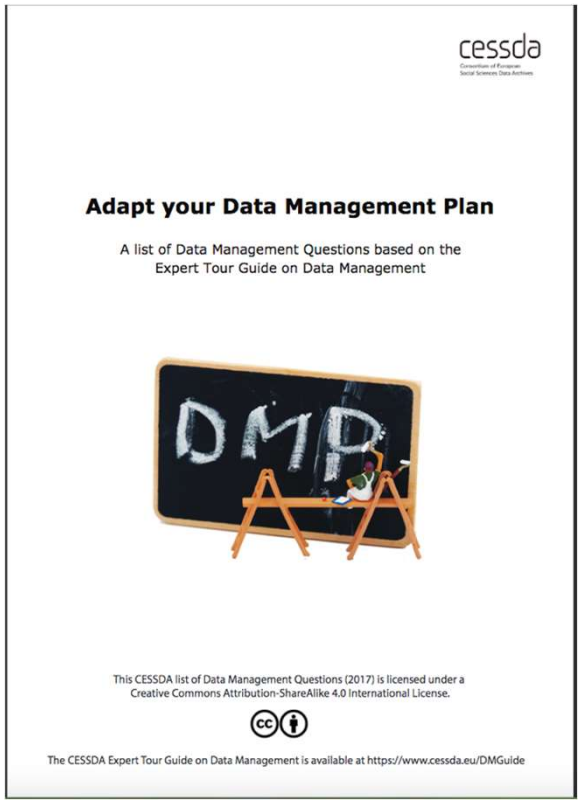
This guide is designed by European experts to help social science researchers make their research data Findable, Accessible, Interoperable and Reusable (**FAIR**).

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

DMP through chapters


cessda

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?

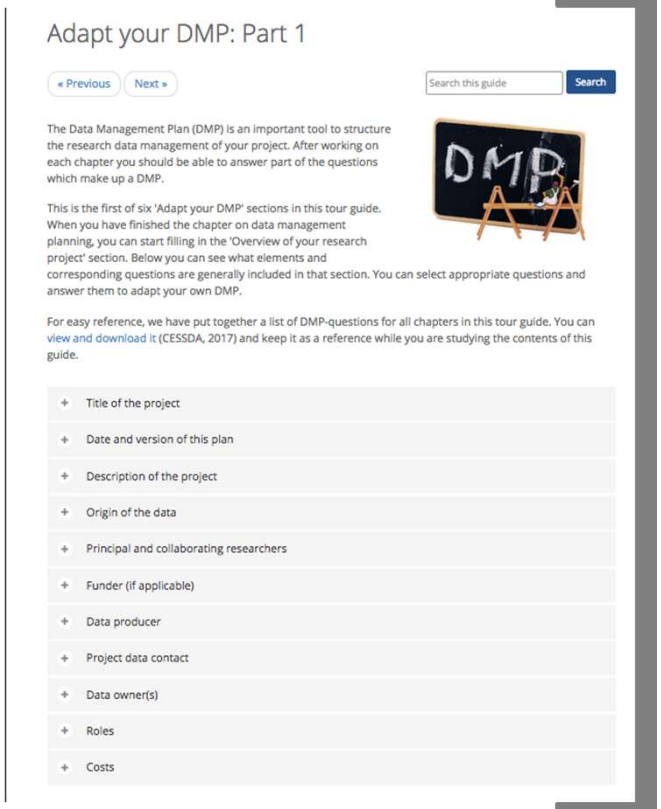
Data producer
- Which organisation has the administrative responsibility for the data?

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 Team (2017 - 2022). *CESSDA Data Management Expert Guide.* Bergen, Norway: CESSDA ERIC. Retrieved from https://www.cessda.eu/DMEG

Data life cycle

You should alwalys start here!

CESSDA Training Team (2017 - 2022). *CESSDA Data Management Expert Guide.* Bergen, Norway: CESSDA ERIC. Retrieved from https://dmeg.cessda.eu/

7. DISCOVER

- **Why**

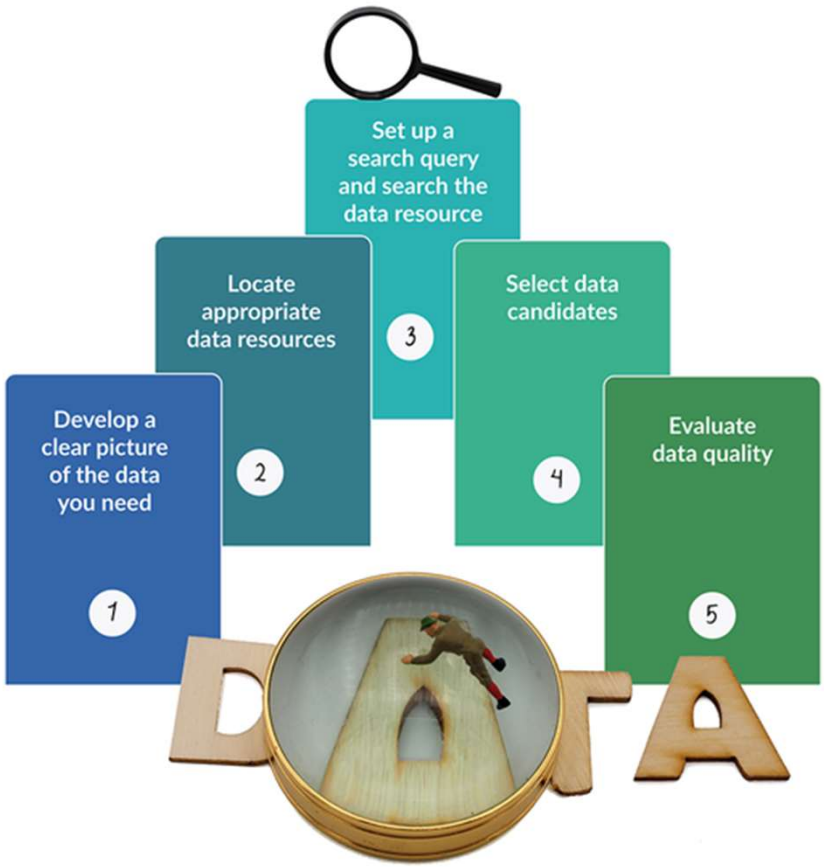
❖ Reuse data and save costs and time

❖ Compare results or make replication studies

❖ Reuse verified elements of research design

❖ Enhance data quality and foster innovation

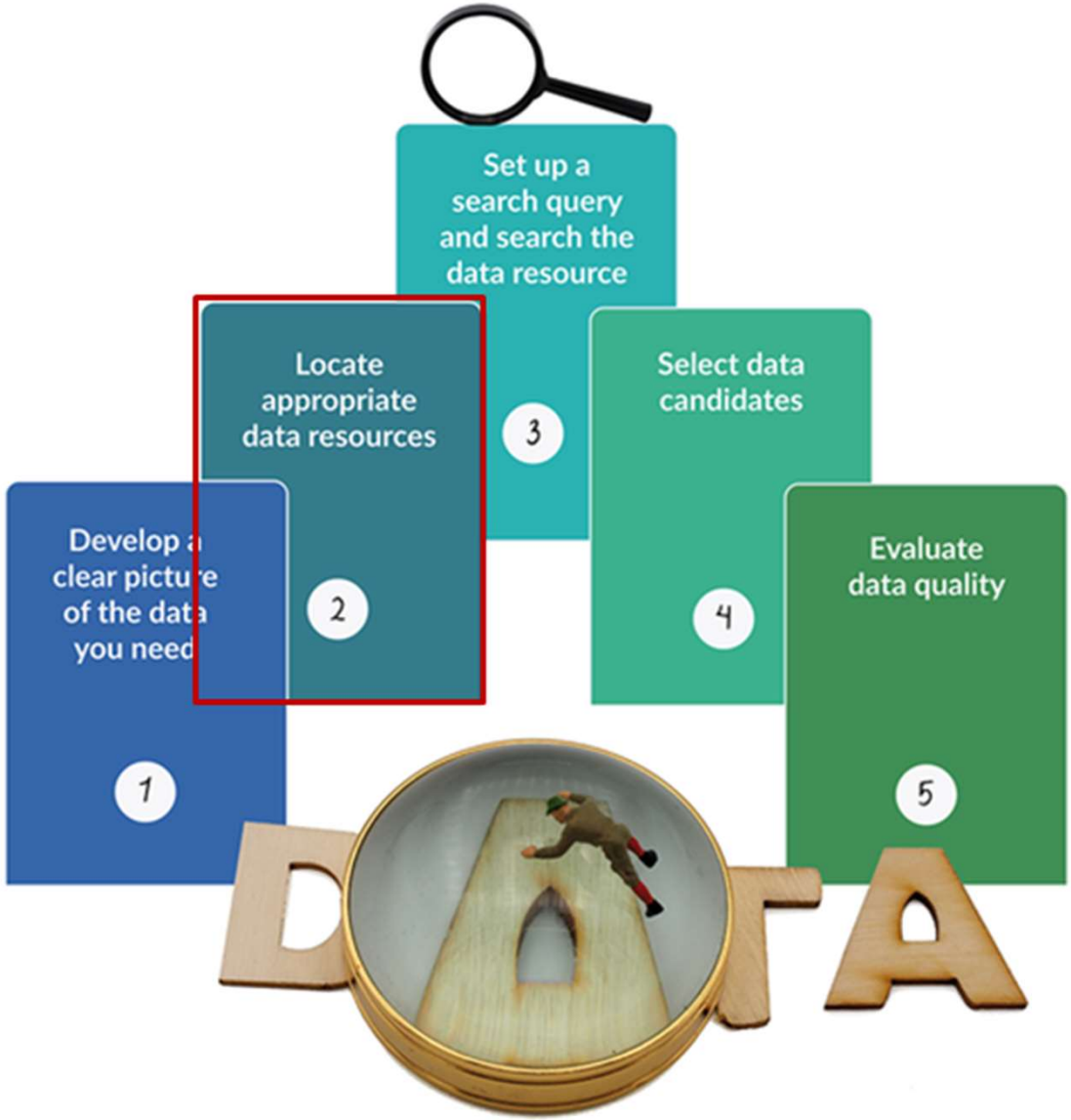
Steps in data discovery



CESSDA Training Team (2017 - 2020).

Four ways we can use archived data

- New analysis: one or multiple data sources e.g. combine micro and macro, just secondary data or secondary data combined with primary data
- Replication
- Use of study design/methodology (e.g. data collection tools (interview schedules & survey questions) or sampling strategies)
- Teaching : Subject-based or research methods,
  Datasets made for training purposes – e.g. easySHARE


Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

DATA DISCOVERY
Where do I start

Data repositories

Digital archives collecting, preserving and displaying datasets, related documentation and metadata.

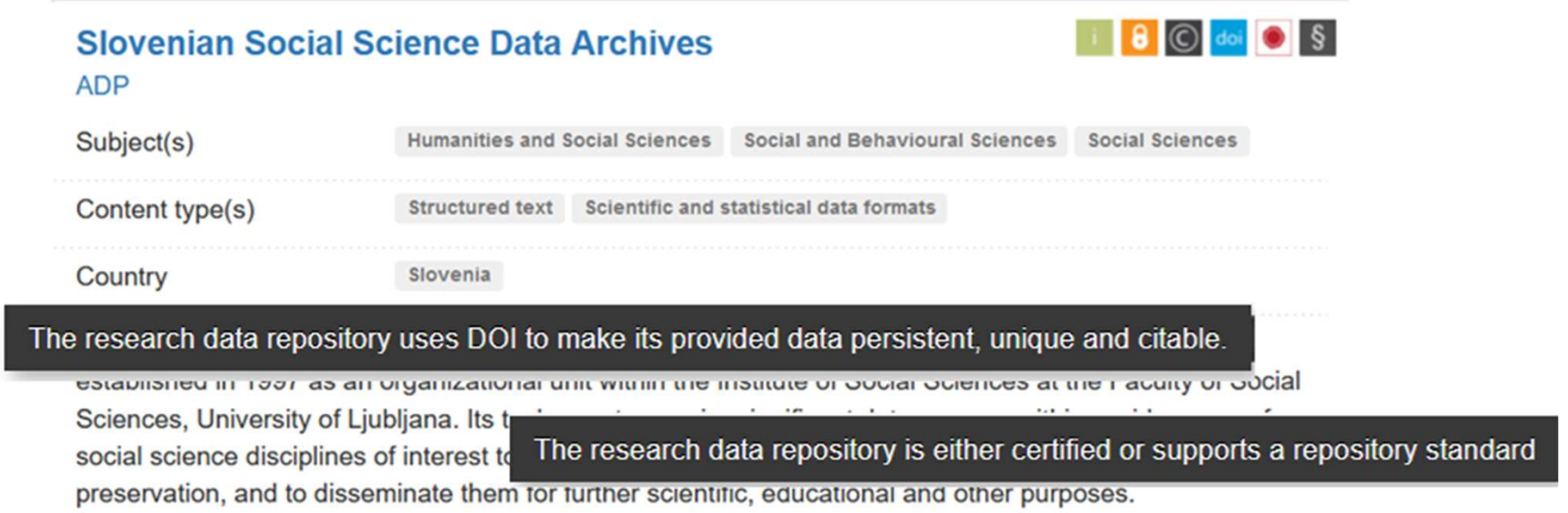
domain-specific trusted repositories (e.g. CESSDA archives) - focus on high-quality data with a potential for reuse	institutional research data repositories e.g. universities	general purpose repositories e.g. Zenodo, Figshare, Harvard Dataverse

DATA DISCOVERY
Registries

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Search: by subject, content type and country

For data archives with a certificate (a trusted repository), open access or for data sets that have a persistent identifier



European social science data archives

Data collections include:

- variation between archives
- quantitative data - major source of individual level data
- qualitative
- outputs of
  - major academic projects
  - government/policy
  - small research teams
  - individual researchers
- recent and less recent data
- different languages

FINNISH SOCIAL SCIENCE DATA ARCHIVE

UK Data Service

Data Archiving and Networked Services DANS

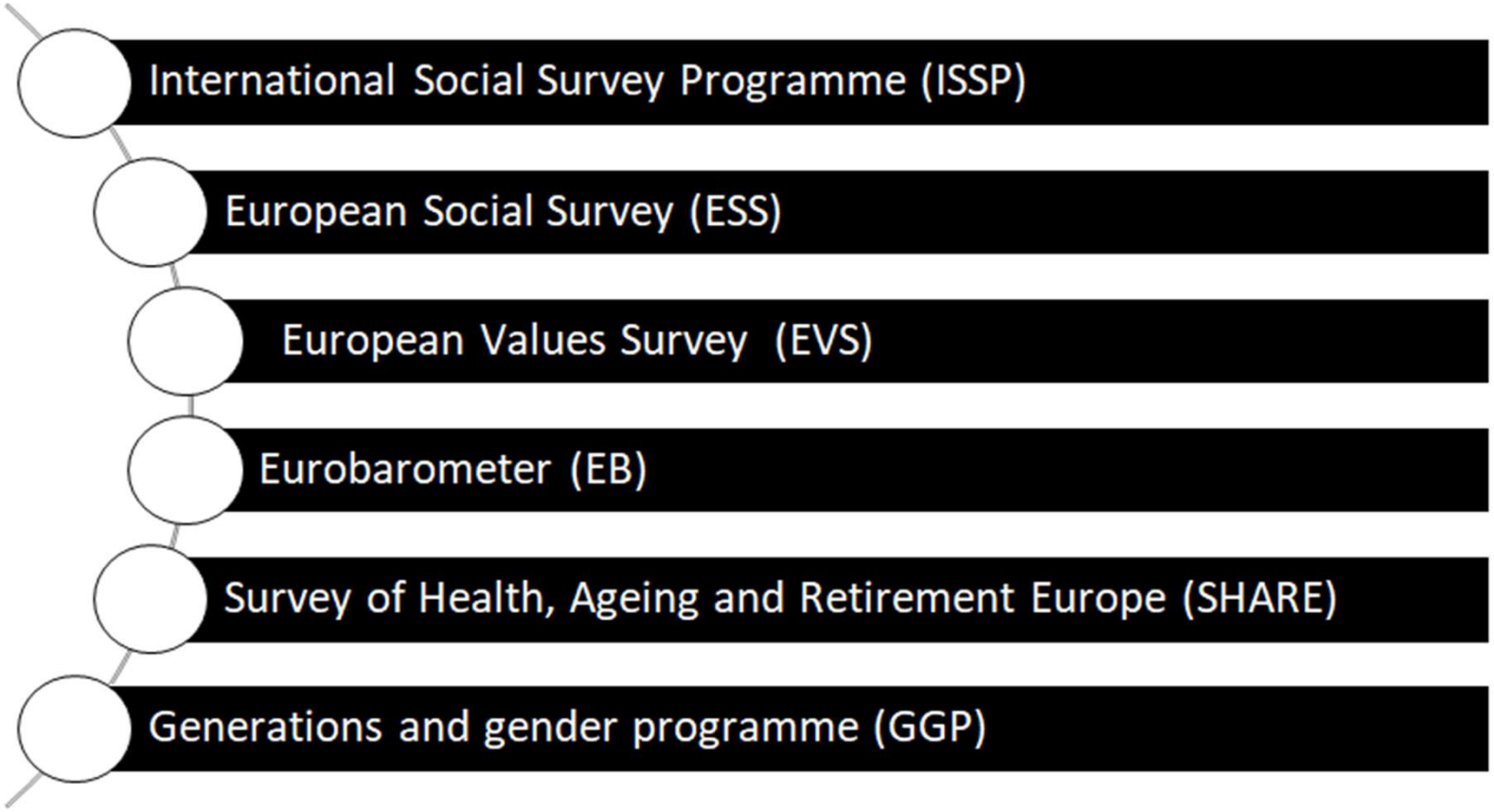
ISSDA Irish Social Science Data Archive

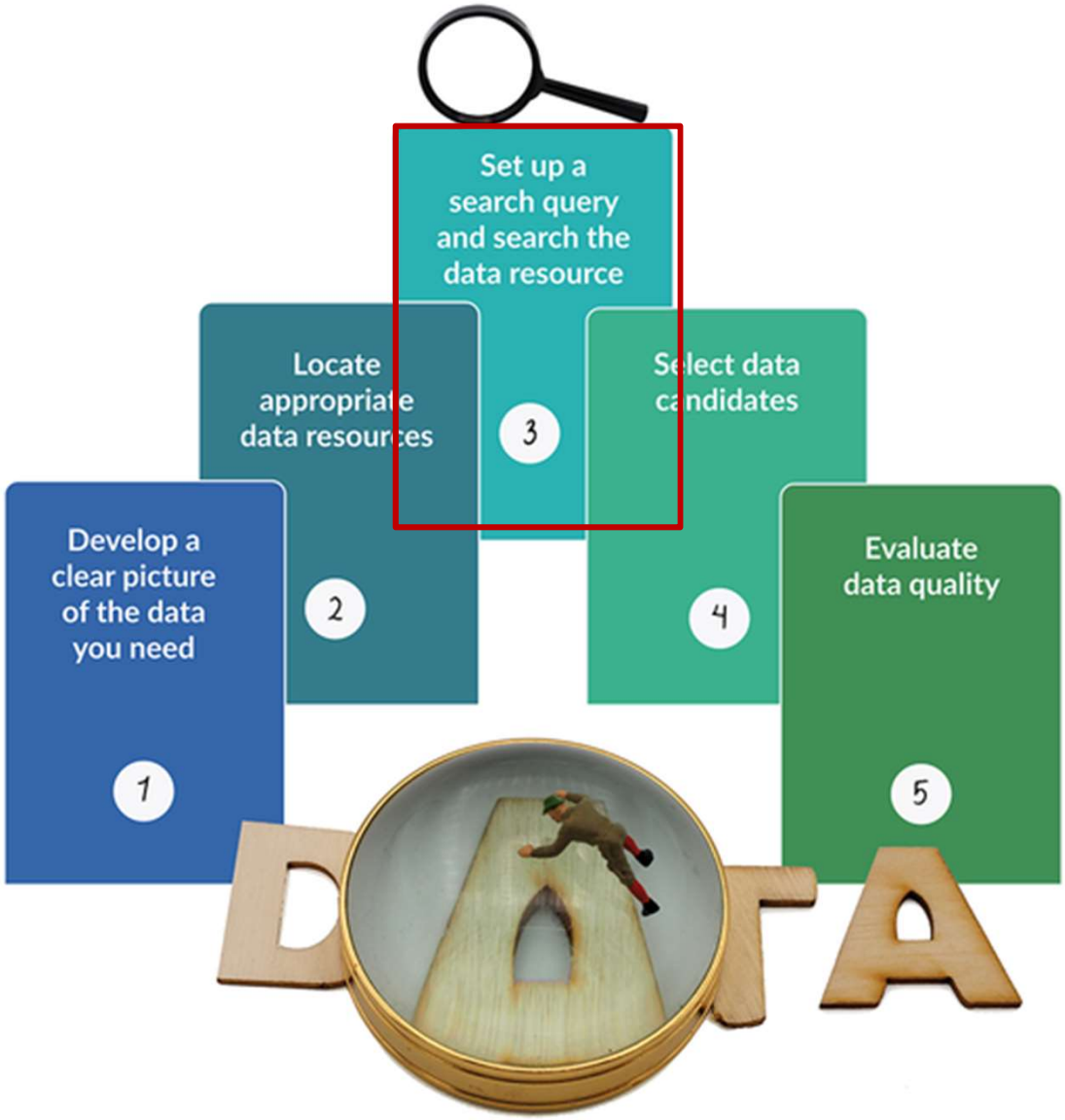
CROSSDA

gesis Leibniz Institute for the Social Sciences

Cross-national studies

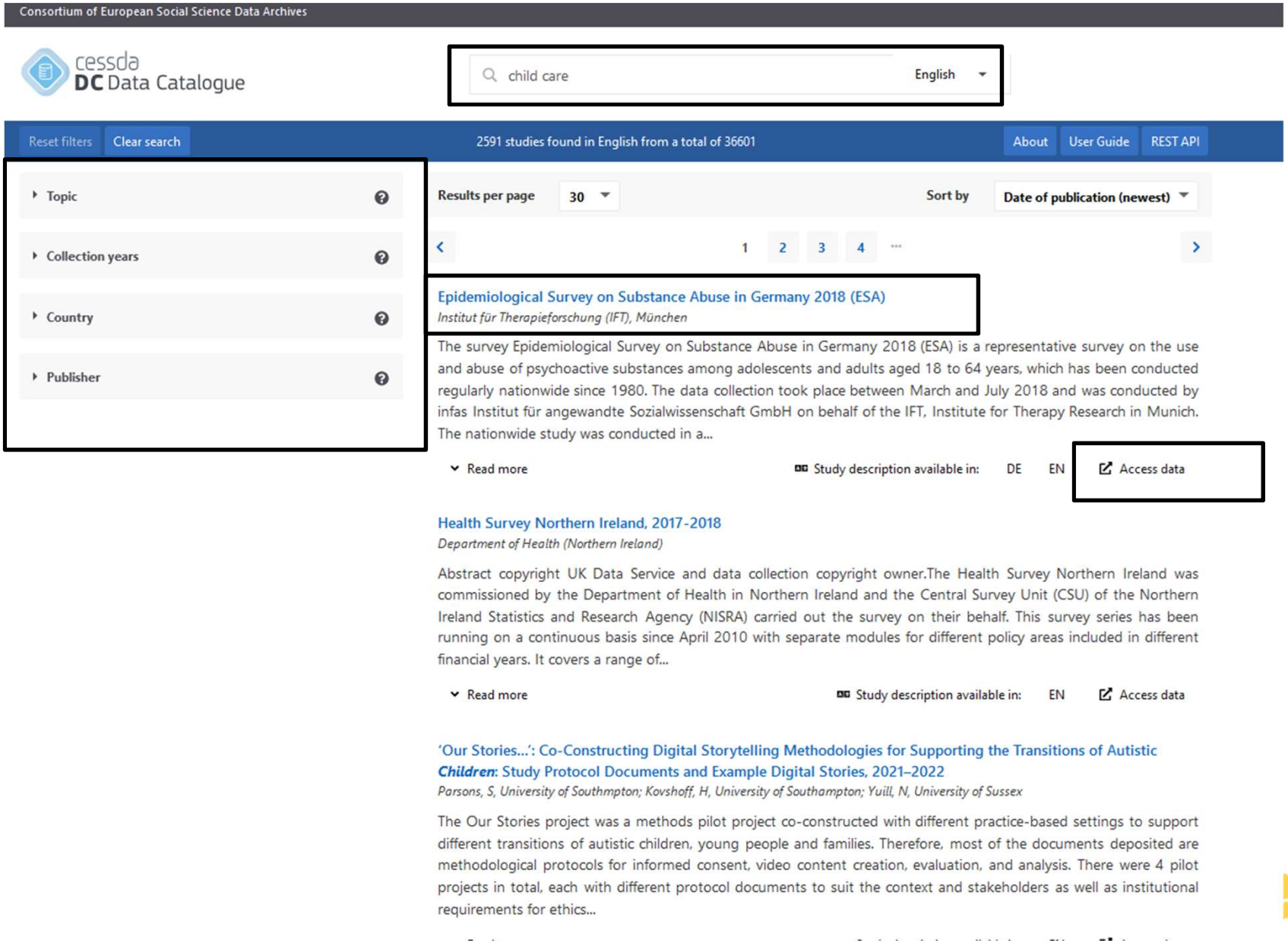
International survey research programmes include many European countries




Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

CESSDA DMEG

CESSDA Data Catalogue
(https://datacatalogue.cessda.eu/)



Survey presented in ADP

ADP Catalogue / / evara17

SURVEY ABOUT CYBERATTACK PROTECTION MOTIVATION IN HIGHER EDUCATION, 2017: RESEARCHERS AND PEDAGOGUES AT SLOVENIAN UNIVERSITIES

Study description | Data description | Accompanying Materials | Nesstar Browser

Basic Study Information

ADP - IDNo: EVARA17

DOI: https://doi.org/10.17898/ADP_EVARA17_V1

Main author(s):

Mihelič, Anže
Vrhovec, Simon

Data file producer:

Vrhovec, Simon, Univerza v Mariboru = University of Maribor, Fakulteta za varnostne vede = Faculty of Criminal Justice and Security (Ljubljana, Slovenia; 2021)

Funding agency:

Univerza v Mariboru = University of Maribor, Fakulteta za varnostne vede = Faculty of Criminal Justice and Security

Study Content

Keywords ADP: protection-motivation theory, higher education, university, cybersecurity, cyber threat, computer security, internet

Keywords ELSST:
CYBERCRIME, CYBERBULLYING, COMPUTER SECURITY, INTERNET

Topic Classification CESSDA
Conflict, security and peace
Information society
Topic Classification CERIF
Criminology
Topic Classification ADP
THREATS OF CYBERATTACKS

Delil

TERMS OF USE:

The data are unrestricted for academic purposes only and licensed under a Creative Commons Attribution + NonCommercial 4.0 International License

nesstar download data | study description

DOCUMENTATION STATUS

4 - Full Study description and XML DDI Codebook Data description with full questions text.

CLASS OF THE STUDY

7 - studies that permits theoretical generalisations or relates on a practical problem, less influential

How to CITE this study?

Mihelič, A. and Vrhovec, S. (2022). *Survey about cyberattack protection motivation in higher education, 2017: Researchers and pedagogues at Slovenian universities* [Data file]. Ljubljana: University of Ljubljana, Slovenian Social Science Data Archives. ADP - IDNo: EVARA17. https://doi.org/10.17898/ADP_EVARA17_V1

COBISS.SI

Survey presented in ADP (Nesstar Catalogue)

Arhiv družboslovnih podatkov

DESCRIPTION | TABULATION | ANALYSIS | EN | ?

- [EVARA17]³ Survey about cyberattack protection motivation in higher education, 2017 : Researchers and pedagogues at Slovenian universities
  - Metadata
    - Study Description
      - Bibliographic Citation
      - Study Scope
      - Methodology And Processing
      - Data Access
      - Other Study Description Materials
    - Data Files Description
    - Other Documentation
  - Variable Description
    - ID, GROUP
    - PERCEIVED THREATS
    - PERCEIVED SEVERITY (ORGANIZATION)
      - Perceived severity (organization) [1]: A successful cyberattack on our organization would greatly jeopardize the privacy of its confidential data.
      - Perceived severity (organization) [2]: A lot of our organizationâ€™s confidential data collected by a successful cyberattack could be misused for criminal purposes.
      - Perceived severity (organization) [3]: A lot of our organizationâ€™s confidential data collected by a successful cyberattack could be misused against it.
    - FEAR OF CYBERATTACKS
    - PERCEIVED VULNERABILITY (INDIVIDUAL)
      - Perceived vulnerability (individual) [1]: It is very likely that I will be a victim of a cyberattack in the future.
      - Perceived vulnerability (individual) [2]: My chances of becoming a victim of a cyberattack are very high.
      - Perceived vulnerability (individual) [3]: I strongly feel that I will become a victim of a cyberattack in the future.
    - PERCEIVED SEVERITY (INDIVIDUAL)
    - PERCEIVED VULNERABILITY (ORGANIZATION)
    - MANDATORINESS
    - PSYCHOLOGICAL REACTANCE
    - MEASURE EFFICACY
    - SELF-EFFICACY

Dataset: [EVARA17]³ Survey about cyberattack protection motivation in higher education, 2017 : Researchers and pedagogues at Slovenian universities

Researchers and pedagogues at Slovenian universities

Variable PVi1: Perceived vulnerability (individual) [1]: It is very likely that I will be a victim of a cyberattack in the future.

PREQUESTION TEXT

Mark your agreement with the statements about exposure to cyberattacks.

LITERAL QUESTION

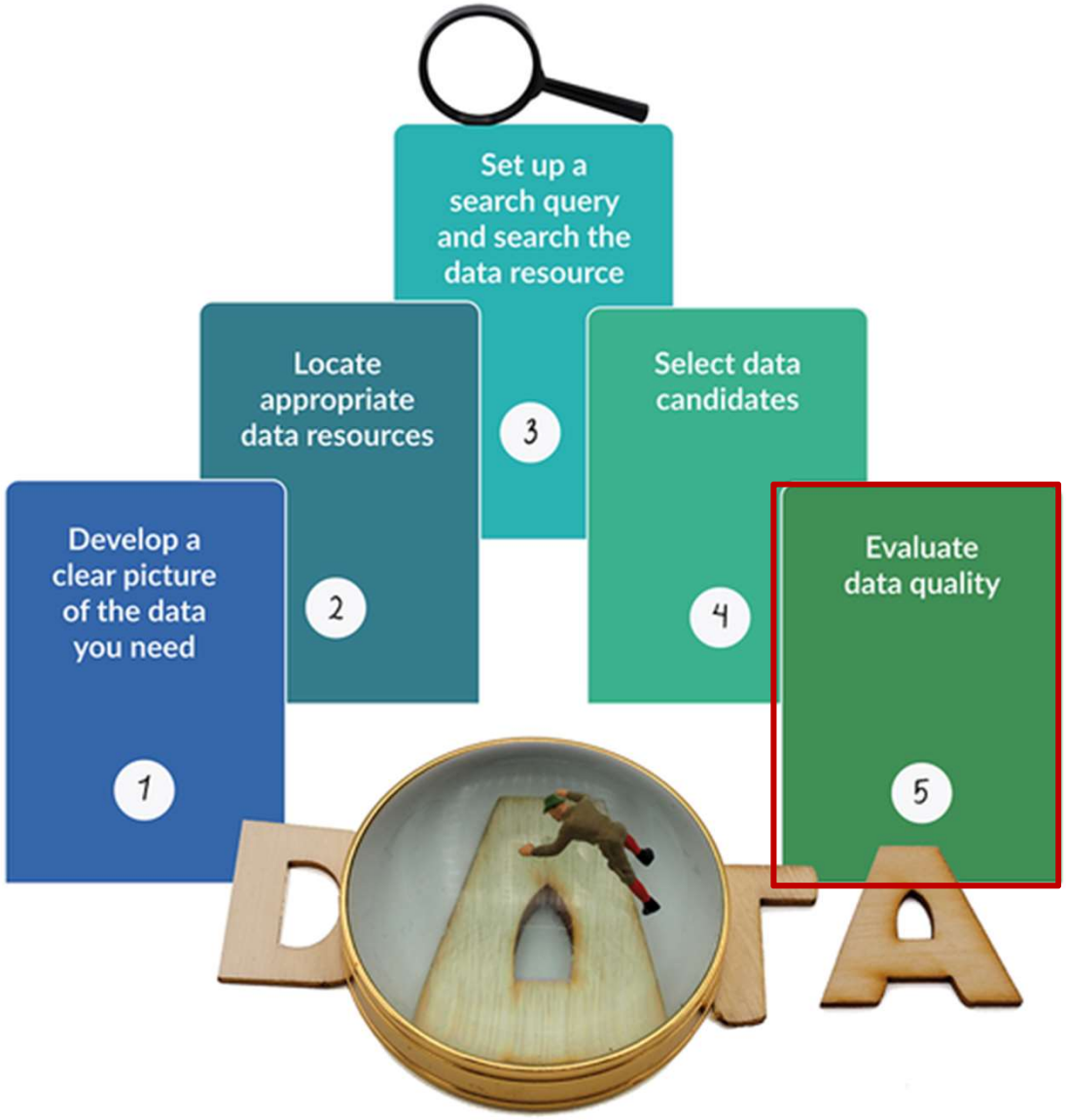
It is very likely that I will be a victim of a cyberattack in the future.

Values	Categories	N	
1	Strongly disagree	23	9.0%
2	Disagree	52	20.4%
3	Somewhat disagree	33	12.9%
4	Neutral	62	24.3%
5	Somewhat agree	51	20.0%
6	Agree	31	12.2%
7	Strongly agree	3	1.2%
-99	Missing	69	

SUMMARY STATISTICS

Valid cases	255
Missing cases	69
Minimum	1.0
Maximum	7.0
Mean	3.671
Standard deviation	1.565

This variable is numeric


Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

CESSDA DMEG

What to look for when assessing quality?

Metadata ("data about data“):

- Why the data was created?
- What the dataset contains?
- How data was collected?
- Who collected the data and when?
- How was the data processed?
- Any manipulations done to the data?
- What quality assurance procedures were used?

CESSDA Training Working Group (2017)

Data access arrangements 1

Open data

any user, no registering (acknowledge source)

Registration

- often with institutional user name and password
- may wait for user name or password
- register use of data

Terms and conditions

- not trying to identify individuals, households or organisations
- not distributing data to others
- “data is for non-commercial use only” or for “use in research or teaching” only.

Download

from catalogue (but sometimes complete a request form)

Data access arrangements 2

- Sometimes permission from the data owners required
- Sensitive or confidential data = more strict (and lengthy) process
- Some services operate a dedicated safe room or safe access service
- Access by users outside the country can be prohibited for confidential data
- Free (except for commercial use and supplementary services)

If you are unsure, ask the relevant data service for help.

And finally…remember to cite data

Why?

- It gives credit the data creators
- It makes data easier to find

How?

- Give enough information to locate the exact version of the data
- Look for recommended citation
- Use persistent identifiers (Digital Object Identifier - DOI)

ELEMENTS OF DATA CITATION

- **Author:** Name(s) of each individual or organizational entity responsible for the creation of the dataset.
- **Date of Publication:** Year the dataset was published or disseminated.
- **Title:** Complete title of the dataset, including the edition or version number, if applicable.
- **Publisher and/or Distributor:** Organizational entity that makes the dataset available by archiving, producing, publishing, and/or distributing the dataset.
- **Electronic Location or Identifier:** Web address or unique, persistent, global identifier used to locate the dataset (such as a DOI). Append the date retrieved if the title and locator are not specific to the exact instance of the data you used.

These are the minimum elements required for dataset identification and retrieval. Fewer or additional elements may be requested by author guidelines or style manuals. Be sure to include as many elements as needed to precisely identify the dataset you have used.

Source: IASSIST – Quick guide to Data Citation

ISSP Research Group (2017): International Social Survey Programme: Work Orientations IV - ISSP 2015. GESIS Data Archive, Cologne. ZA6770 Data file Version 2.1.0, doi:10.4232/1.12848

Hafner-Fink, M. and Malešič, M. (2016). Slovenian Public Opinion 2015: Work Orientation (ISSP 2015), Role of Government (ISSP 2016), Mirror of public opinion and National Security Survey [Data file]. Ljubljana: University of Ljubljana, Social Science Data Archives. ADP – IDNO: SJM15. https://doi.org/10.17898/ADP_SJM15_V1

Data life cycle

CESSDA Training Team (2017 - 2020). *CESSDA Data Management Expert Guide.* Bergen, Norway: CESSDA ERIC. Retrieved from https://www.cessda.eu/DMGuide

5. Protect → Research Ethics

- Disciplinary Code of Ethics (ASA)
- National Code of Ethics – Soc. Assoc.

- European Code of Research Integrity
- University (UNI-LJ)
- Institute

- Funder – Horizon Europe / other EC projects / grants
- Scientific Journal <-ethical committee approval before publishing

Ethics are an integral part of a research project, from the conceptual stage of the research proposal to the end of a research project.

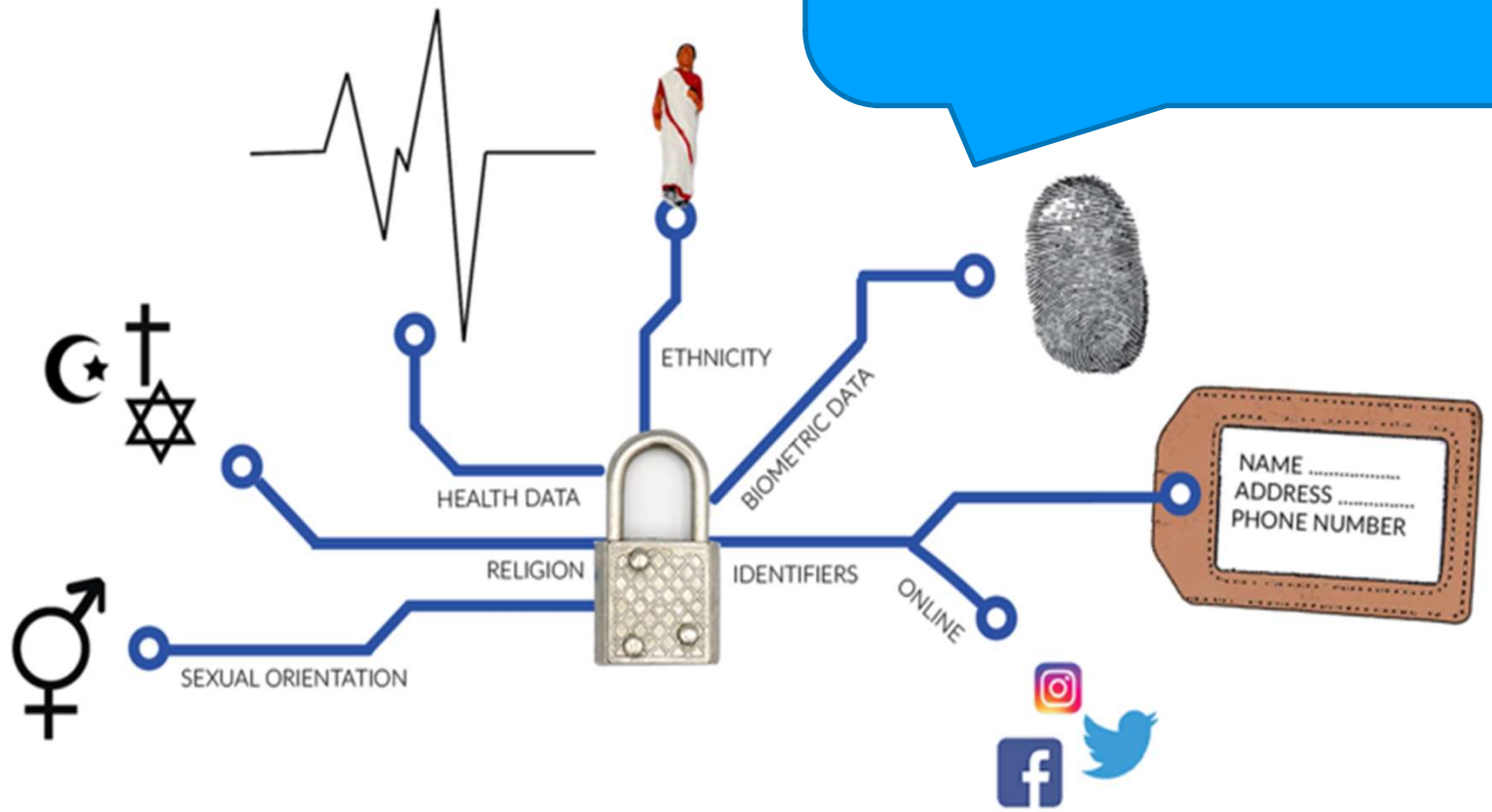
Short definition “personal data” by GDPR

Personal data is any information that may be used to identify a person directly or indirectly

- **Directly identifying personal data**
  - *through full name, personal identification number*
- **Indirectly identifying personal data**
  - *through a combination of background information*

Sensitive personal data

… „by reference to an identifier such as **a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity** of that natural person" (GDPR, Article 4)



CESSDA Training Team (2017 - 2020). *CESSDA Data Management Expert Guide.* Bergen, Norway: CESSDA ERIC. Retrieved from https://www.cessda.eu/DMGuide

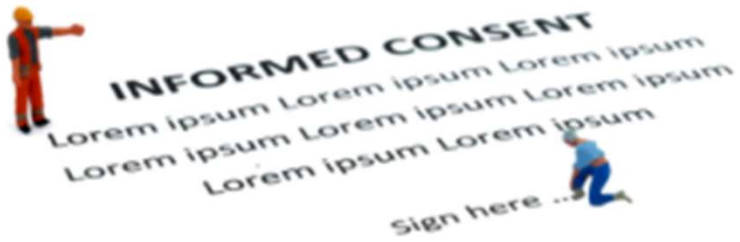
Grounds for Processing Personal Data

The General Data Protection Regulation (GDPR)

There are 6 grounds for the processing of personal data, and one of these must be present in order to process a data subject's personal data:

1. **Consent of the data subject**
2. Necessary for the performance of a contract
3. Legal obligation placed upon controller
4. Necessary to protect the vital interests of the data subject
5. Carried out in the public interest or is in the exercise of official authority
6. Legitimate interest pursued by controller

Informed consent



Informed consent is the process by which a researcher discloses appropriate information about the research so that a participant may make a voluntary, informed choice to accept or refuse to cooperate.

Consent needs to be freely given, informed, unambiguous, specific and by a clear affirmative action that signifies agreement to the processing of personal data.

⊖ Click to see examples of consent forms

- ⊕ UK Data Archive
- ⊕ MRC Cognition and Brain Sciences Unit - University of Cambridge
- ⊕ FORS (Swiss Centre of Expertise in the Social Sciences)

https://ukdataservice.ac.uk/learning-hub/research-data-management/ethical-issues/consent-for-data-sharing/
https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/5.-Protect/Informed-consent

Strategy for Sharing Data

- Obtain **informed consent**, also for data sharing and preservation or curation
- Protect identities where needed e.g. **anonymisation** and not collecting personal data if not necessary
- **Regulate access** where needed (all or part of data) e.g. by group, use or time period
- **Securely store** and protect personal and sensitive data

DPIA - Data Protection Impact Assessment

The DPIA is a written document to be formally approved by the University and DPO.

- Sensitive data
- *Consent not possible*
- Long term processing / archiving
- *Vulnerable group*
- Very identifiable data
- *Combination of the above*

More on DPIA in SI Ocena učinka v zvezi z varstvom podatkov

EDPB has set 9 criteria:

-Sensitive data or data of a highly personal nature (4)

-Data processed on a large scale(5)

-Data concerning vulnerable data subjects (7)

Data life cycle

CESSDA Training Team (2017 - 2020). *CESSDA Data Management Expert Guide.* Bergen, Norway: CESSDA ERIC. Retrieved from https://www.cessda.eu/DMGuide

6. Publish: DATA Publication (P!!!)

PUBLICATIONS AND DATA

It is expected that a Data Publication will ensure that data will potentially be considered as a first-class research output (Knowledge Exchange, 2013).

For a dataset to "count" as a publication should be:

- Properly documented with metadata;
- **Reviewed for quality;**
- Searchable and discoverable in catalogues (or databases);
- Citable in articles.

Costas, R., Meijer, I., Zahedi, Z. and Wouters, P. (2013). The Value of Research Data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report, available from www.knowledge-exchange.info/datametrics

Where to publish?

- Journal supplementary material service
- Institutional data repository — Repository of the University *of Ljubljana*
- General purpose repository — zenodo
- Domain specific data repository — DARIAH-SI
- Trusted domain specific data repository — CLARIN.SI

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Data publication with domain spec. repository

Advantages

- Offers **specialist domain knowledge** and data management expertise, e.g. to create a catalogue record and documentation;
- More likely to **accept complete datasets**;
- Provides preservation and curation to **community standards**, e.g. file formats migration;
- **Ability to control access** of (sensitive) personal data;
- May handle data re-use queries;
- May make your data visible via dissemination and promotion.

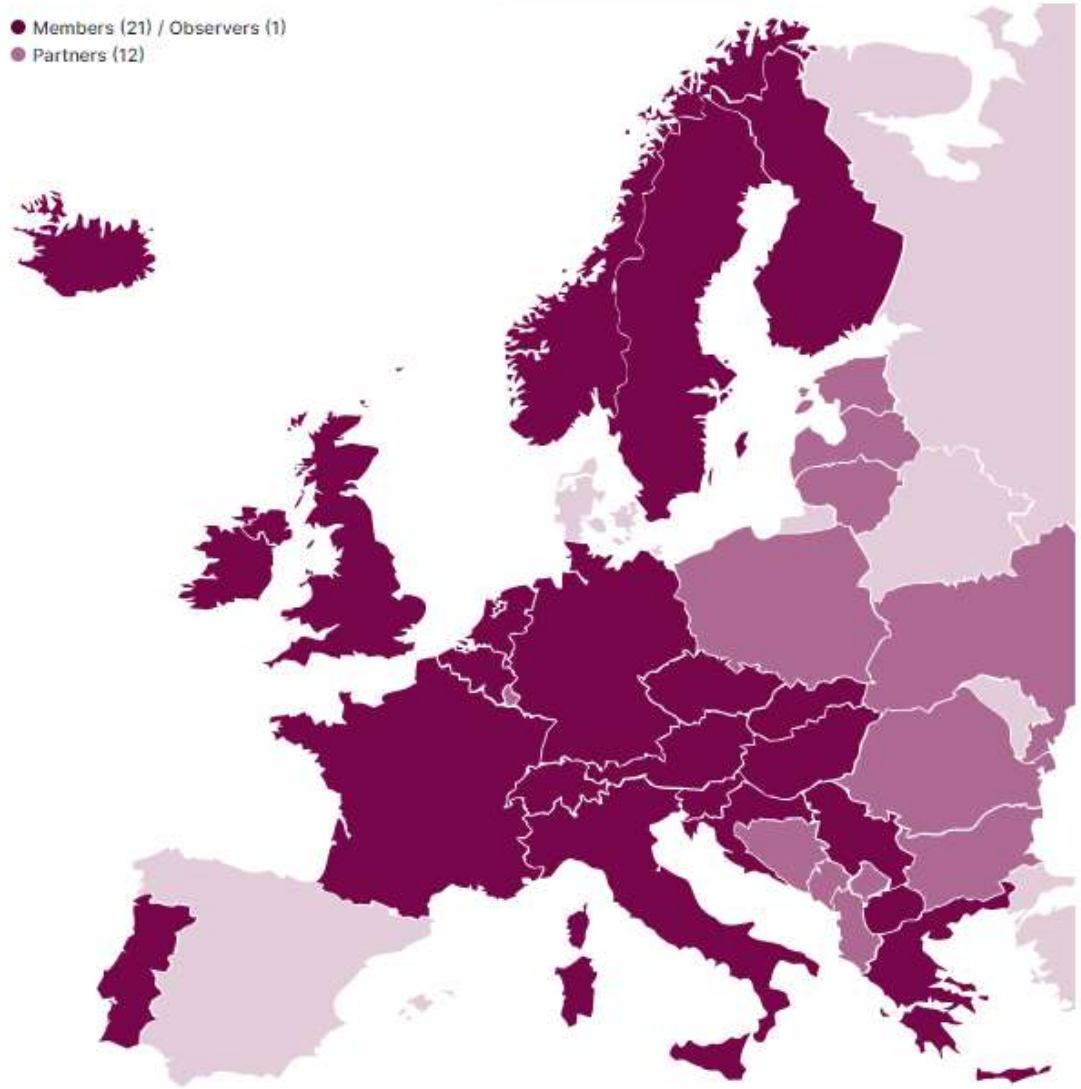
Disadvantages

- Most likely to be selective about what kind of data they accept;
- Requires advance planning of the effort needed to meet high standards for metadata and documentation.

(CESSDA DMEG, 2017-2020)

Consortium of European Social Science Data Archives

To find out more, use the map below or see the list of members and partners.



https://www.cessda.eu/About/Consortium

Archive and publish with ADP

1) Slovenian national data archive for social sciences

2) Trust-worthy

3) Get credits for publication

4) Get advise and support from data experts

5) Get training

6) Get involved with ADP‘s partnering data archives

Time to…

Present your DMP ☺

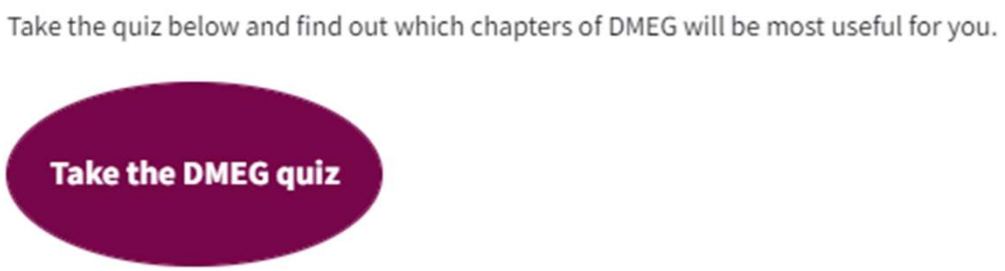
https://instr.iastate.libguides.com/dmp/writingDMP

Important take away

Challenging situations	Solutions as part of DMP
• I will finalize my thesis next week. I need to publish my data.	• Contact your data archive at least 6 months before you plan to finish your thesis.
• Data access agreement doesn't allow me to share the variables.	• Check access agreement before signing it, be sure that you are able to share your data.
• I promised participants to use my data only for my PhD thesis.	• Check with data archive if the consent form allows you publishing your data, before you collect the data.

Additionally

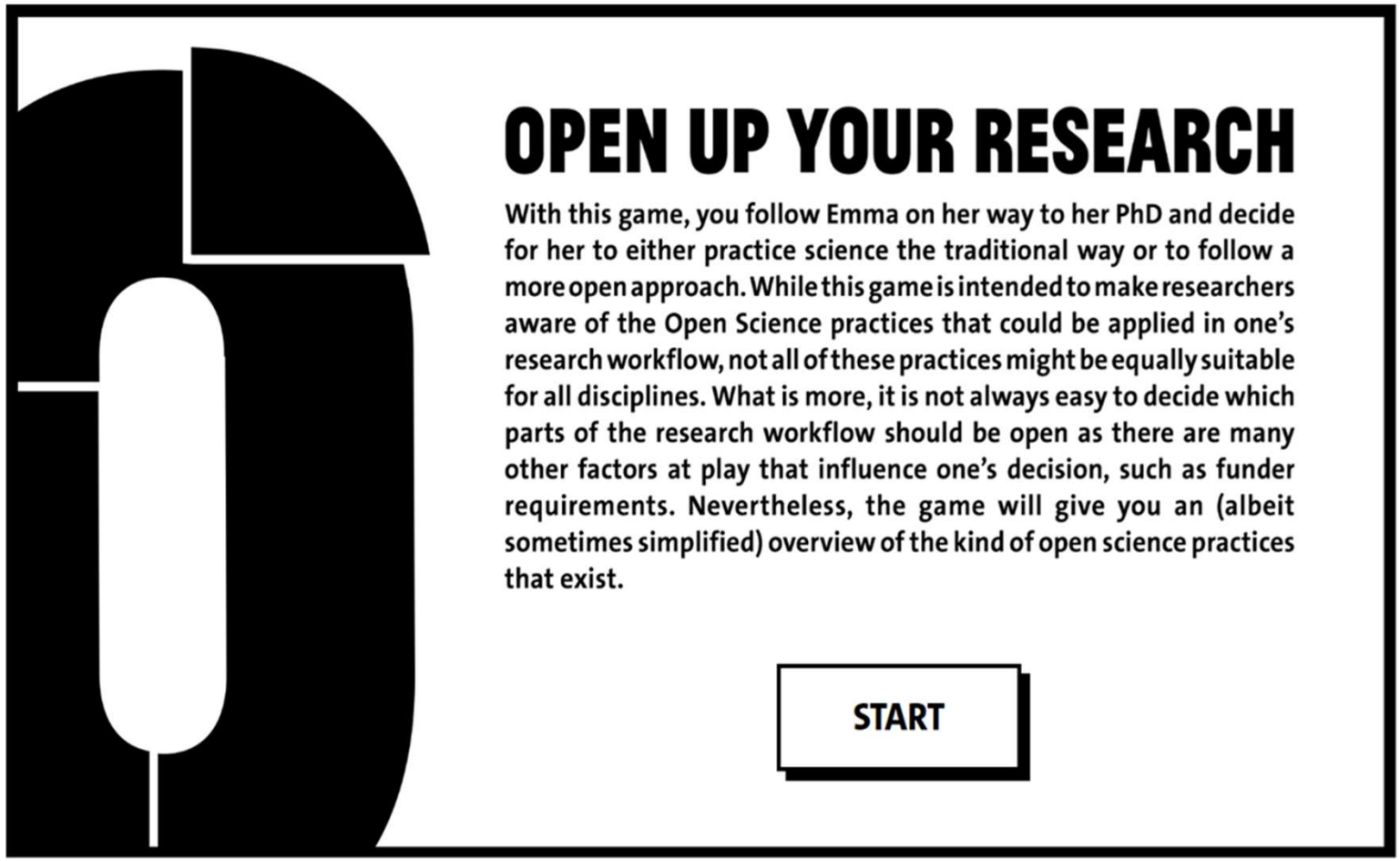
CESSDA Quiz: www.cessda.eu/dmeg



LoD Data Management Challenge game: https://lod.sshopencloud.eu

Open Science Game: Open Up Your Research

INTRODUCING EMA →



https://www.openscience.uzh.ch/en/moreopenscience/game.html

Röthlisberger, Melanie, Höfler, Manuela, Hermans, Katherine, Furrer, Eva. 2021. Open Up Your Research - An Open Science Game. University of Zurich.

IF YOU HAVE ANY FURTHER QUESTIONS…

… CONTACT ADP

University of Ljubljana

Faculty of Social Sciences

Social Science Data Archive

Kardeljeva ploščad 5

1000 Ljubljana

Slovenia

www.adp.fdv.uni-lj.si

arhiv.podatkov@fdv.uni-lj.si

Arhiv.Druzboslovnih.Podatkov

@ArhivPodatkov

FURTHER READINGS

1) PROTECT - CONSENT FORM
2) COSTING TOOL (https://ukdataservice.ac.uk/media/622368/costingtool.pdf)
3) PROTECT AND STORE YOUR DATA (https://study.sagepub.com/corti2e/student-resources/data-collection/answers-to-in-chapter-exercises/61-book-answers)
4) ANONYMIZATION OF QUALITATIVE DATA (https://study.sagepub.com/corti2e/student-resources/data-collection/answers-to-in-chapter-exercises/81-anonymisation-of)
5) FILE FORMATS (https://ukdataservice.ac.uk/media/622179/exercise_open_file_formats.pdf)
6) Template for transcribing interviews, with uniform style and layout (Model qualitative interview transcript)

How FAIR are your data?

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

.....

Let's see if your data is FAIR

https://www.cessda.eu/content/download/3845/35038/file/20170707_How_FAIR_are_your_data_Jones.pdf

Read the points and add notes to elements that still need to be resolved.

COSTING TOOL

https://www.ukdataservice.ac.uk/media/622368/costingtool.pdf

ACTIVITY	COMMENTS AND SUGGESTIONS	√	COST
Data description • Are data in a spreadsheet or database clearly marked with variable and value labels, code descriptions, missing value descriptions, etc.? • Are labels consistent? • Do textual data like interview transcripts need description of context, e.g. included as a heading page?	• if data description is carried out as part of data creation, data input or data transcription – low or no additional cost • if needed to be added afterwards – higher cost • codebooks for datasets can often be easily exported from software packages		
Data cleaning • Do quantitative data need to be cleaned, checked or verified before sharing, e.g. check validity of codes used, check for anomalous values? • Will data match documentation, e.g. same number of variables, cases, records, files? • Does textual information in data need to be spell-checked?	• if carried out as part of data entry and preparation before data analysis – low or no additional cost • if needed afterwards – higher cost		
Documentation • Do you have documentation for the data that describes the context and methodology of how data were gathered, created, processed and quality controlled?	• often essential contextual and methods documentation will be written up in publications and reports • if all data creation steps are well documented and documentation is kept well organised during research – low or no additional cost		

Why publish research data?

Data Sharing and Management Snafu in 3 Short Acts

Karen Hanson, Alisa Surkis and Karen Yacobucci (2012) NYU Health Sciences Library: https://www.youtube.com/watch?v=N2zK3sAtr-4

Sources

CESSDA Training Team (2017 - 2022). *CESSDA Data Management Expert Guide.* Bergen, Norway: CESSDA ERIC. Retrieved from https://www.cessda.eu/DMGuide

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

Some slides originate from Train the Trainers package of CESSDA DMEG. We would like to thank colleagues from CESSDA to make it possible to re-use them for events like this.

Viri v slovenskem jeziku (sources in SI)

ADP: **Življenjski krog podatkov**:
https://www.adp.fdv.uni-lj.si/usposobi/ZKG/nacrtovanje/

Načrt ravnanja z raziskovalnimi podatki – vprašalnik:
https://www.adp.fdv.uni-lj.si/publikacije_adp/publikacija/338/

Informacijski pooblaščenec:

- infografika – Podlaga za obdelavo osebnih podatkov
  https://upravljavec.si/pogosta-vprasanja/komunikacija-b2c-in-b2b/
- Ocena učinka v zvezi z varstvom osebnih podatkov

https://www.ip-rs.si/zakonodaja/reforma-evropskega-zakonodajnega-okvira-za-varstvo-osebnih-podatkov/kljucna-podrocja-uredbe/ocena-ucinka-v-zvezi-z-varstvom-podatkov/

Načrtovanje zbiranja raziskovalnih podatkov skladno z načeli FAIR *(predavanje za doktorske študente, 2019)*, https://www.adp.fdv.uni-lj.si/publikacije_adp/publikacija/324/

DCC: DMPOnline *(posnetek predstavitve uporabe, P. Čerče, ZRS Koper)*

https://www.youtube.com/watch?v=wt9Y9AJKtVs

https://seriss.eu/wp-content/uploads/2019/07/Admin-data-guidelines_quick-ref.pdf
https://seriss.eu/wp-content/uploads/2019/07/Social-media-guidelines_quick-ref.pdf