



# DATA MANAGEMENT PLAN FOR A GOOD RESEARCH DATA MANAGEMENT

Katarzyna Biernacka

# Welcome at the CSES Research Seminar

- Presentation of the CSES Team and their research interests:

[https://box.hu-berlin.de/lib/696147d0-6f63-4217-b833-789641f4ba75/file/CSES/CSES\\_Research\\_Summer2020.pptx](https://box.hu-berlin.de/lib/696147d0-6f63-4217-b833-789641f4ba75/file/CSES/CSES_Research_Summer2020.pptx)

- Homepage of the research seminar:

[https://www.informatik.hu-berlin.de/de/forschung/gebiete/cses/for\\_students/teaching/Forschungsseminar](https://www.informatik.hu-berlin.de/de/forschung/gebiete/cses/for_students/teaching/Forschungsseminar)

# Agenda

- Part 1:

Theory: Basics of Research Data Management

- Part 2:

Practice: Writing a Data Management Plan

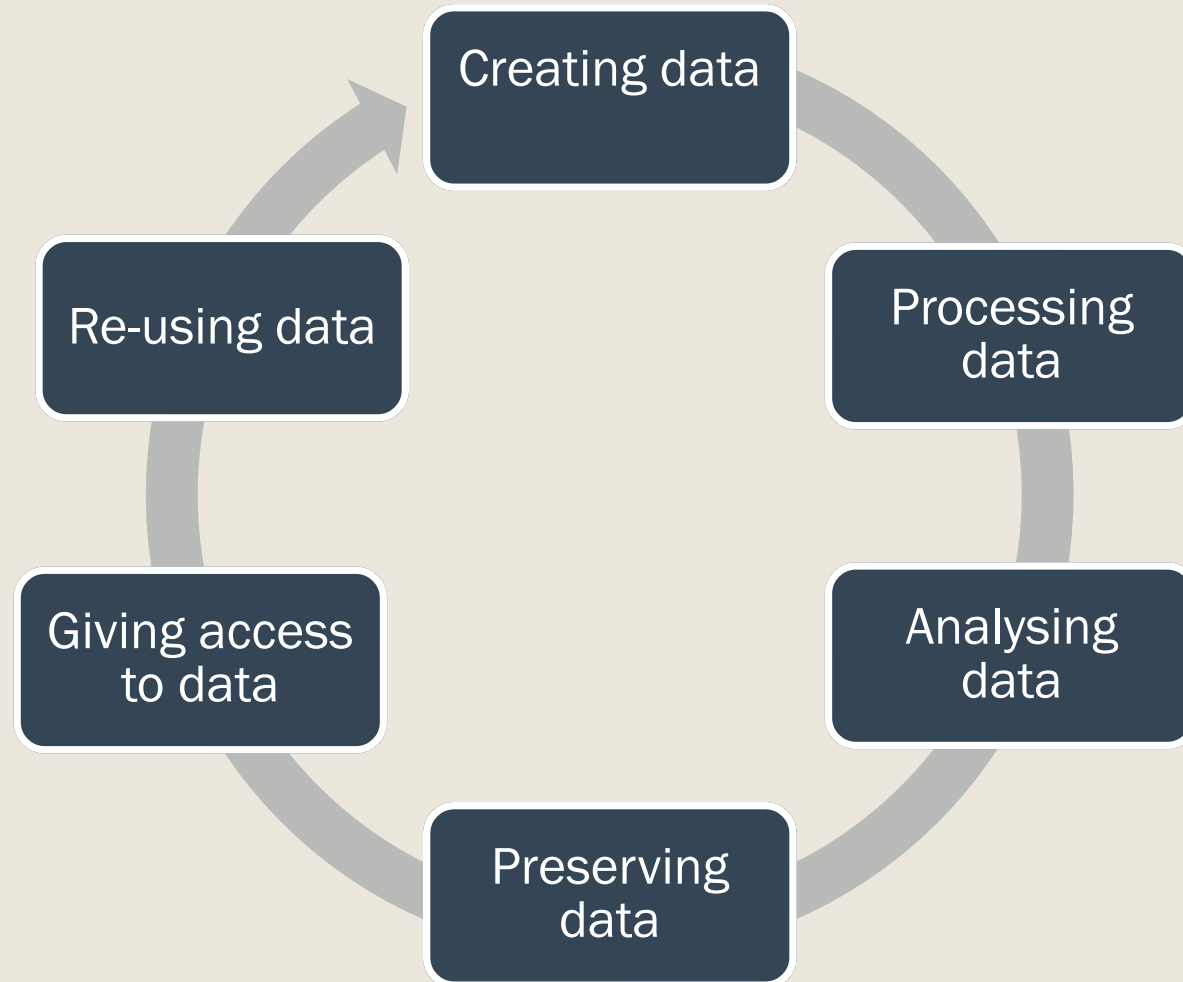
# It's a Workshop!

- Please DO ask questions!
- Rise your hand if you want to ask or comment something
- Don't interrupt others
- Ask questions in the chat too
- Use the feedback buttons where appropriate

# Research Data

„Research data refers to all information (regardless of form or presentation) needed to support or validate the development, results, observations or findings of research work, including contextual information. Research data includes all materials created in the course of academic work, e.g. through digitisation, records, source research, experiments, measurements, surveys and interviews. This includes software and code.“

# Research Data Lifecycle



Source: Research data lifecycle of the UK Data Archive (Accessed: June 2018). In the meantime, a newer version of the life cycle is presented there.

# Research Data Management

Research data management (RDM) includes all activities associated with

- processing,
- storage,
- preserving and
- publication

of research data.

RDM accompanies the research process from the initial planning to the archiving, re-use or deletion of the data.

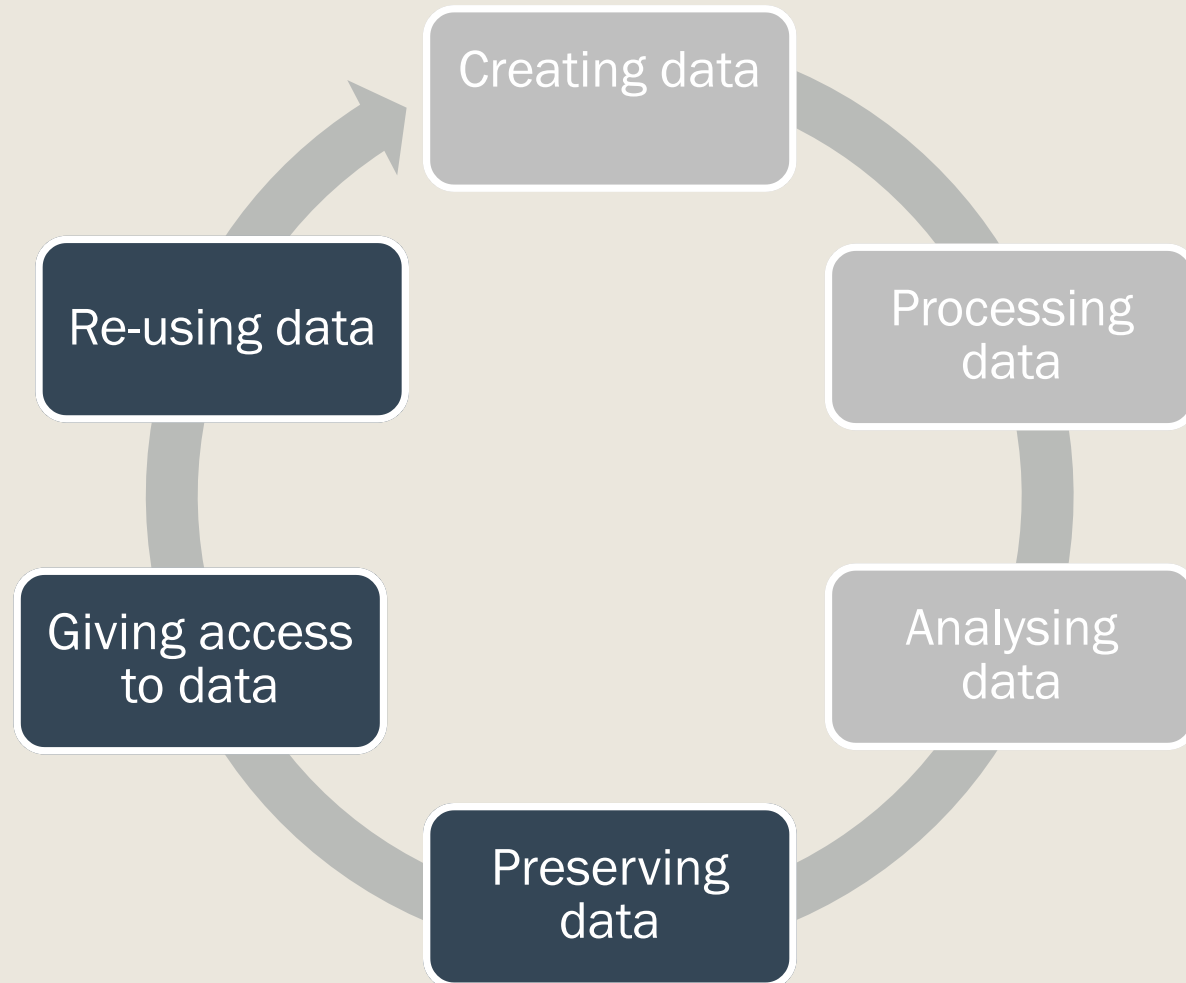
# Why is RDM important?

It allows you to:

- follow the best research practices
- save time for research
- validate and replicate findings
- avoid risk of data loss
- ensure transparency and reproducibility
- increase data visibility and number of citations
- fulfil funders' requirements (e.g. Data Management Plans)
- enable new collaborations



# Research Data Lifecycle



Source: Research data lifecycle of the UK Data Archive (Accessed: June 2018). In the meantime, a newer version of the life cycle is presented there.

# Documentation

- For what purpose was the data created?
- What does the data set contain?
- How was the data collected?
- Who collected the data?
- When was the data collected?
- How was the data processed?
- What data cleansing processes were carried out?
- How was the quality of the data ensured?
- In which formats is the data available?
- How can the data be accessed?



# Metadata – data about data

## Content metadata:

- title
- description
- author
- copyright holder
- contact details
- license details
- keywords

## Technical metadata:

- recording date
- focal length
- orifice
- exposure time
- geo data
- and many more

# Metadata standards

An overview of discipline-specific and interdisciplinary metadata standards can be found below:

<http://www.dcc.ac.uk/resources/metadata-standards>

<http://rd-alliance.github.io/metadata-directory/subjects>

Examples:

- Interdisciplinary:
  - Dublin Core
  - MARC21
- Geosciences:
  - ISO 19115
- Biodiversity:
  - Darwin Core
- Humanities:
  - Text Encoding Initiative (TEI)
- Natural sciences:
  - ICAT Schema
  - Crystallographic Information Framework
- Social sciences:
  - Data Documentation Initiative

# Authority Files, Classifications, Thesauri and Controlled Vocabularies

## Examples:

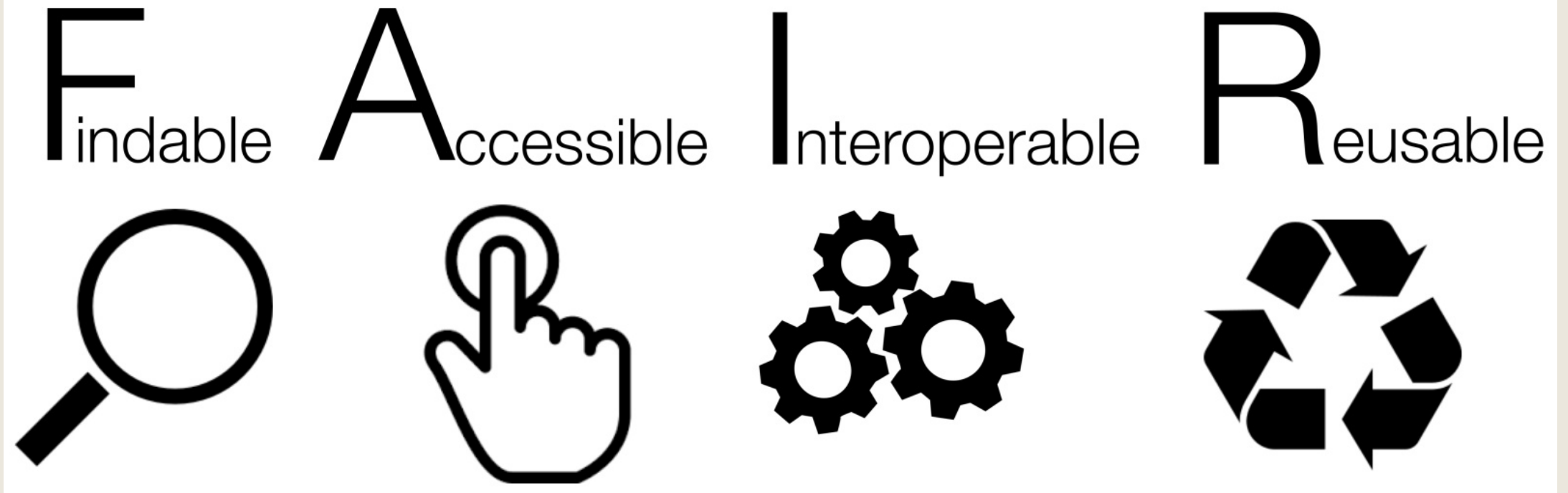
- standards for the unambiguous identification of people, e.g. International Standard Name Identifier (ISNI, ISO 27729)
- general and interdisciplinary classification systems, e.g. DDC or LCC
- intradisciplinary classifications, e.g. Mathematics Subject Classification (MSC)
- discipline-specific thesauri, e.g. TheSoz, STW Thesaurus for Economics

An overview on different systems can be found here:

Basel Register of Thesauri, Ontologies & Classifications (BARTOC) ([bartoc.org](http://bartoc.org))

Taxonomy Warehouse (<http://www.taxonomywarehouse.com/>)

# The FAIR Principles



Source:

Pundir, Sangya. [https://commons.wikimedia.org/wiki/File:FAIR\\_data\\_principles.jpg](https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg) [last access: 27.09.2018], CC-BY-SA-4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.en>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand J. Aalberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.

# Data Publication

Different ways of research data publication:

- as a supplement to a peer-reviewed article ("enhanced publication")
- in data journals
- as independent asset in a research data repository



# Data Paper in a Data Journal

a data paper documents and describes research data to facilitate dissemination and re-use since traditional articles don't provide enough space for the important and valuable description of data. The interpretations are not published in data papers.

Examples of Data Journals:

GigaScience

Earth System Science Data

Journal of Chemical and Engineering Data

# Research Data Repository

Publication of research data as an independent digital asset in:

- discipline-specific repositories, e.g. Datorium, Pangaea
- cross-disciplinary repositories, e.g. zenodo, figshare
- institutional repositories, e.g. edoc server (Humboldt-Universität zu Berlin)

# Research Data Repository

Repositories can be found at [re3data.org](https://re3data.org)

Criteria you should pay attention to when choosing a repository:

- Certification, e.g. Data Seal of Approval
- Assignment of persistent identifiers, e.g. DOI, URN
- Access to data: open, restricted or closed
- Clear terms of use for data authors and users, e.g. fees, embargo periods

# Persistent Identifier

## Digital Object Identifier (DOI)

- most frequently used for research data
- citable
- nationally and internationally recognized
- only 5 mandatory metadata fields to ensure citation

## Open Researcher and Contributor ID (ORCID) - [orcid.org](https://orcid.org)

- unique scientific identity
- is run by a non-profit initiative
- connected to Web of Science, zenodo, DataCite, etc.

# Licensing

Creative Commons Licences

Open Data Commons

Software Licences:

- MIT-License
- GNU General Public License (GPL)
- GNU Lesser General Public License (LGPL)
- Apache License

# Creative Commons



CC0 (Public Domain)



CC BY (Attribution)



CC BY-SA (Attribution – Share Alike)



CC BY-ND (Attribution – No Derivatives)



CC BY-NC (Attribution – Non-Commercial)



CC BY-NC-SA (Attribution - Non-Commercial – Share Alike)



CC BY-NC-ND (Attribution - Non-Commercial – No Derivatives)

# Re-Using Research Data

What is the result of combining two datasets with these licenses?



# Data Management Plan

All information that describes and documents sufficiently the collection, processing, storage, archiving and publication of research data within a research project.

- administrative information (project name, data creator, other contributors, contact, funding programme, etc.)
- project and dataset description
- information on data types, formats and scope
- information on metadata and their standards
- information on data publication and availability
- information on archiving and backing up the data
- responsibilities
- expenses



# Practice

Please go to

<https://dmponline.dcc.ac.uk>

# Summary

1. Don't panic!
2. Start as early as possible.
3. Just write down your research practices.
4. Update your DMP regularly!

# Katarzyna Biernacka

Mail: [biernack@hu-berlin.de](mailto:biernack@hu-berlin.de)

ORCID: 0000-0002-6363-0064

Twitter: @ICTKasia

UNLESS OTHERWISE NOTED, THE CONTENT ON THESE SLIDES IS LICENSED UNDER A CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE (CC-BY-4.0).



SOURCE:

BIERNACKA, KATARZYNA, DOLZYCKA, DOMINIKA, HELBIG, KERSTIN, & BUCHHOLZ, PETRA. (2018). TRAIN-THE-TRAINER KONZEPT ZUM THEMA FORSCHUNGSDATENMANAGEMENT (VERSION 1.0). ZENODO. [HTTP://DOI.ORG/10.5281/ZENODO.1215377](http://doi.org/10.5281/ZENODO.1215377)