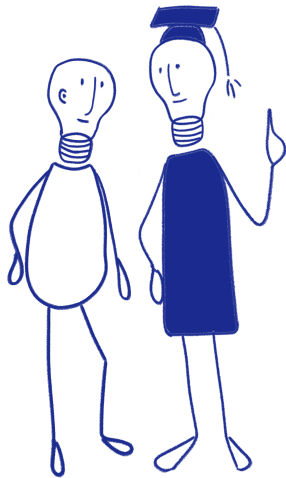


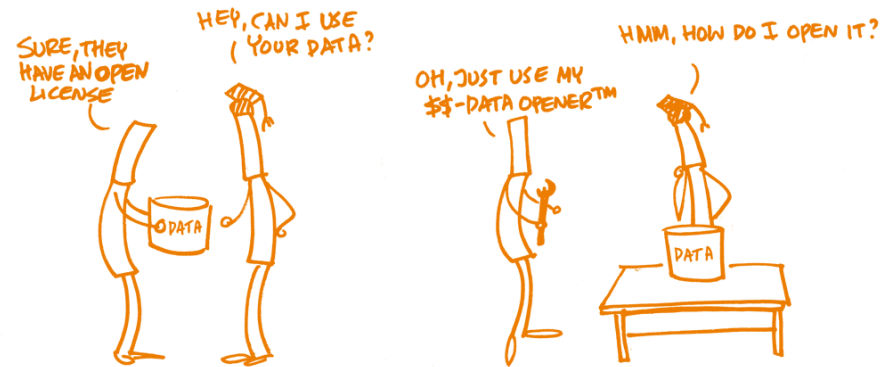
Workshop for Sustainable Research Data Management



WiSo Graduate School
Iris Vogel, Marie Ryan
12.02.2020

Topics

- Introductions
- (Sustainable) data management
- DMP Exercise
- Data Management Plan
- RDMO Exercise
- Break
- Data organisation and workflows
- Storage and publication of research data



<https://legacy.gitbook.com/book/open-science-training-handbook/book/details>

Who are you?

- Your research project?
- Your data?
- Your expectations of the workshop
-

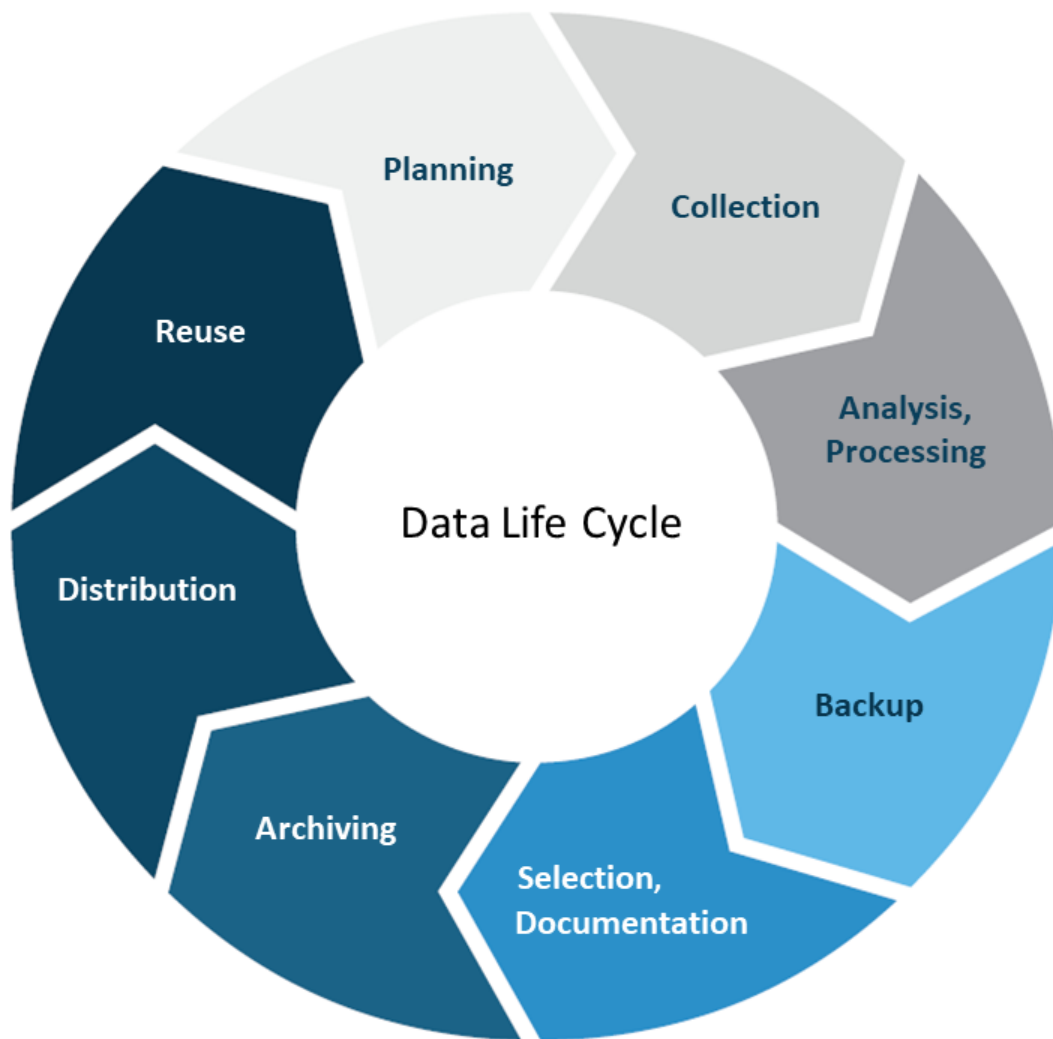
Sustainable Research Data Management

What is RDM?

Research Data Management (RDM) is a term that describes the organisation and control of the work processes involved in generating and managing research data as efficiently as possible. In this way, RDM accompanies research from the initial planning stage through to archiving, subsequent use or deletion of the data.

Aspects of RDM:

- Order and structure
- Planning
- Storage and Backup
- Security and access rights
- Long-term archiving
- Publication of research data
- Legal Aspects



Why RDM?

Good RDM makes usage easier

- Find data faster (e.g. by name, structure)
- Prevent data loss (Data is independent from people)
- Enable (semi-) automatic processing (Metadata)
- Long-term understanding of results (Documentation!)
- Using data in the long term
- Optimal use of resources (through potential reuse)
- Fulfil the requirements of money/third-party funding providers



“Sustainable” RDM

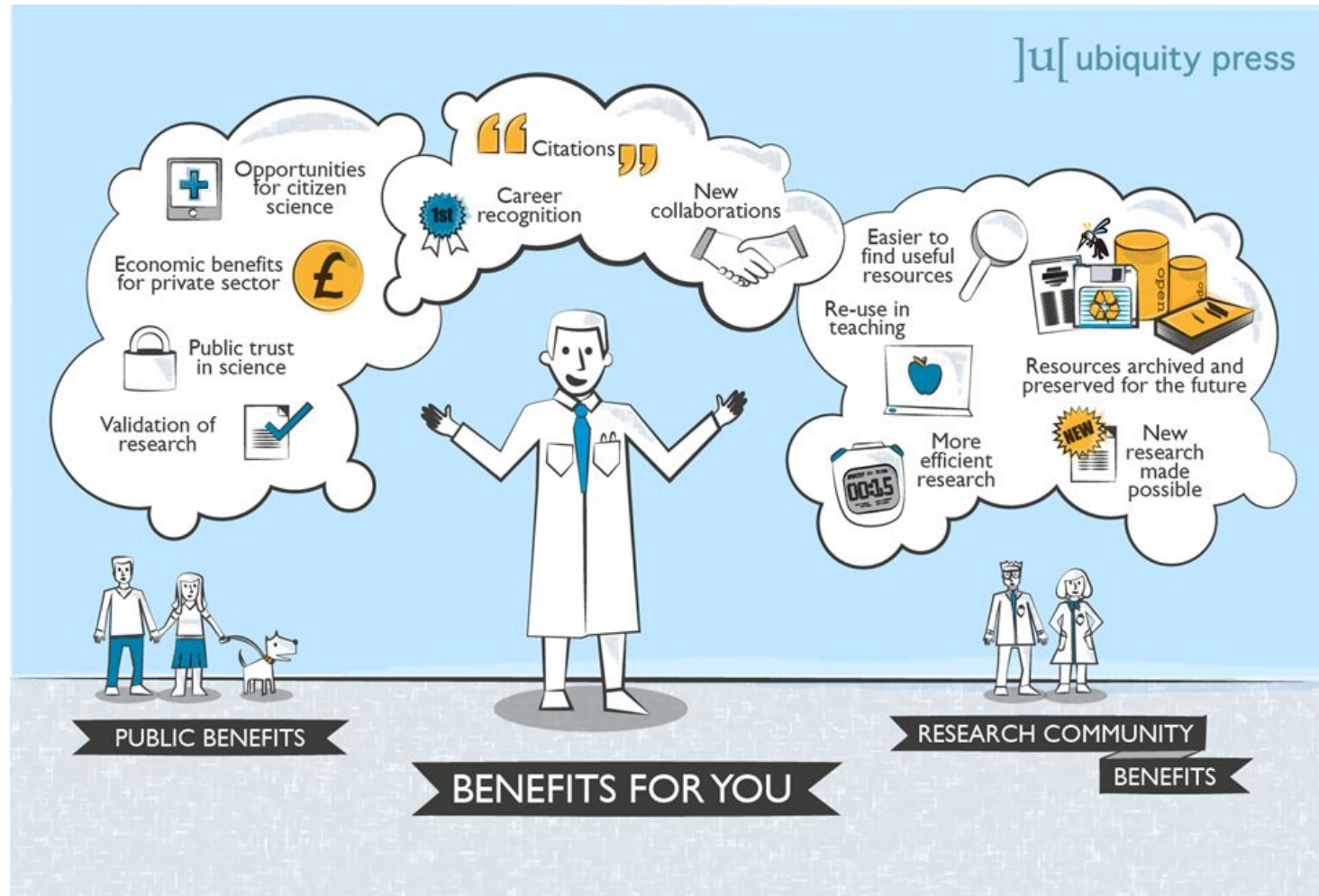
Open Data - benefits

- Makes data usable, instead of just using
- Freedom that goes beyond free of charge
- Allows for adaptation and updating of RD
- Enables collaborative work
- Strengthening communities
- Creates diversity of ideas



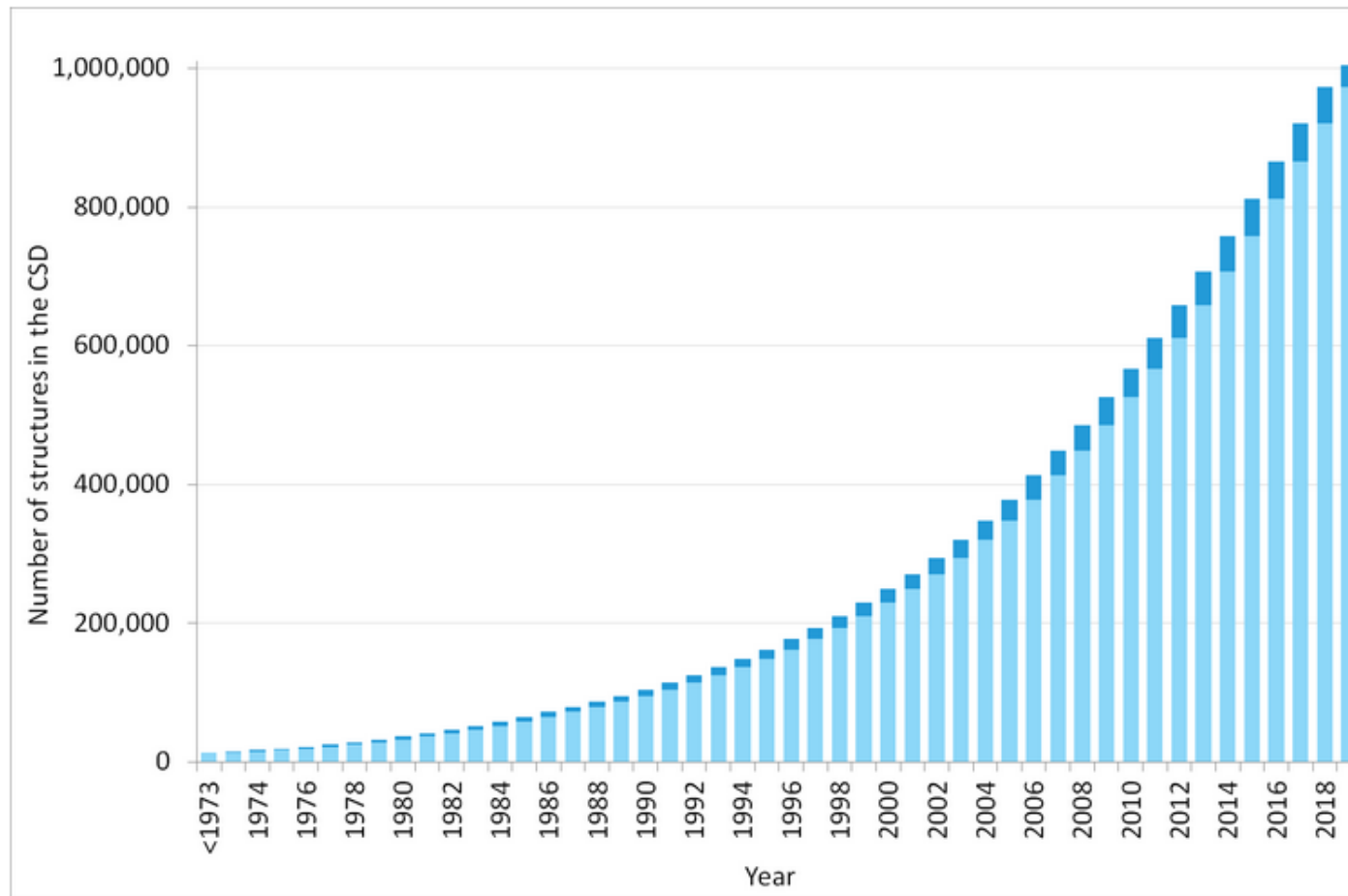
Why should I share my data at all?

Why should I share my data at all?



Slide by Brian Hole and Ubiquity Press

Growth of Data Repositories



Growth of the CSD since 1972, the red bar shows structures added annually.

Cambridge Structural Database

Findable



Accessible



Interoperable

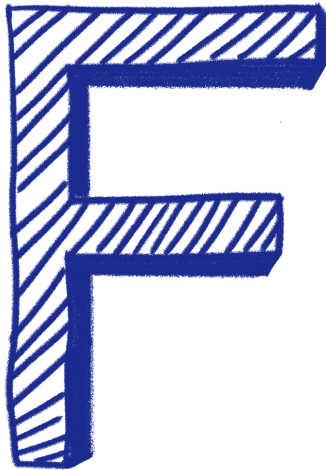


Reusable



Findable

Research data must be findable



- Description through metadata (machine readable)
- Citation capability with the aid of persistent identification (PIDs)
- Visibility (MD registration in detection systems and catalogues)

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Home Search Browse Suggest FAQ About

Browse by subject

Graphical Text

A. Humanities and Social Sciences

a. Humanities

I. Ancient Cultures

1. Prehistory
2. Classical Philology
3. Ancient History
4. Classical Archaeology
5. Egyptology and Ancient Near Eastern Studies

II. History

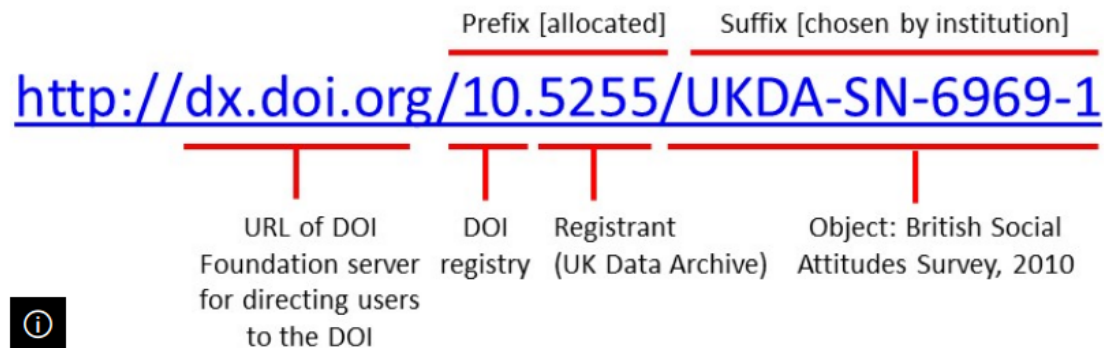
1. Medieval History
2. Early Modern History
3. Modern and Current History
4. History of Science

III. Fine Arts, Music, Theatre and Media Studies

What are Persistent Identifiers (PIDs)?

- Permanent digital identifier, consisting of digits and/or alphanumeric characters, which is assigned to a digital object and refers directly to it
- For research data:
 - DOI (Digital Object Identifier)
 - URN (Uniform Resource Name)
 - Handle
 - ORCID (People)
- URLs are problematic because they do not refer to content, but to a location
- Ideally, nothing changes in the PID when the location of the digital object changes
- If the object is deleted, a landing page with the corresponding metadata is retained

PIDs ensure that a digital object can be permanently indexed, retrieved and thus quoted.



Accessible

Research Data must be (long term) accessible



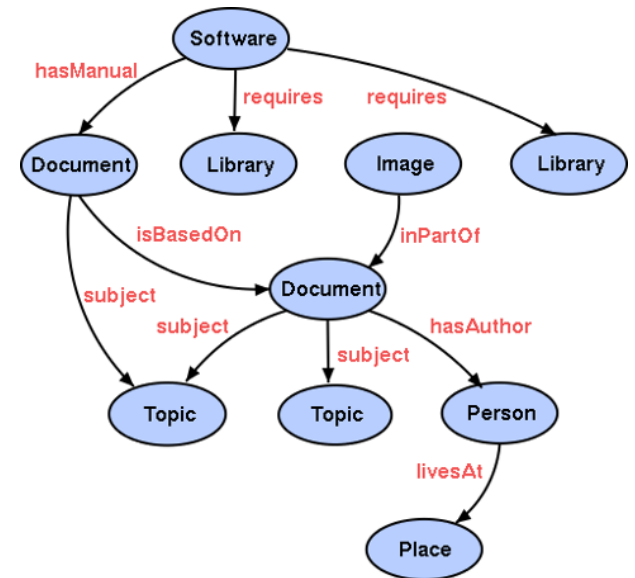
- RD are accessible via their PIDs using standardised, open protocols (HTTP(S), FTP, SMTP)
- Where necessary, the protocol supports authentication and rights management
- Secure long-term archiving
- Metadata remain available even if resources have been deleted



Interoperable RD must be interoperable

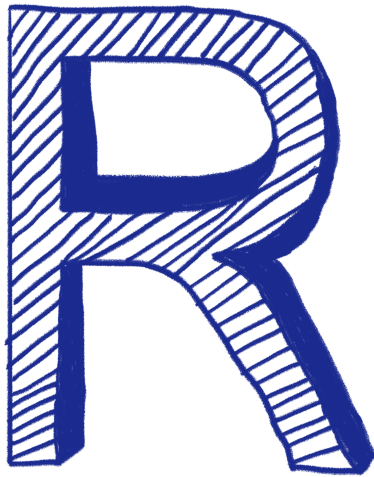


- Allow data exchange between researchers, organisations etc (adhering to format standards)
- Standardised, machine-readable metadata
- Controlled vocabularies, value ranges, thesauri and ontologies
- Standardised references to other MD/ RD (e.g. 'is new version of' , 'is supplement to' , 'relates to' ,)
- Use of open, documented formats



Reusable

RD must be reusable



- Precise, standardised, detailed description
- Unambiguous, accessible user licenses
- Clear indication of origin (Provenance information)
- Research data and metadata comply with the relevant standards for the field

What data to keep?

- Do the data have to be kept for legal reasons, and if so, for how long?
- Can the data be preserved without loss of information?
- What does the loss of information mean for the data, is it still usable?
- Is there interest in the data?
- Are the data part of a larger collection (e.g. time series)?
- Can the data be collected again, if so, with what effort?
- Is a subsequent use of the data to be expected?
- What type of subsequent use can be expected?
- Are the data of cultural / social value?
- Are the data of sufficient quality for subsequent use?
- Can the data be used for teaching purposes?

Which data should I keep?

Decision should depend on potential reuse:

1. Further publication
→ referenced (=processed) data with additional documentation
2. Verification
→ referenced data including analysis steps
3. Further analyses
→ all original data plus software used to collect it
4. Learning and Teaching
→ samples of original data and compiled data including analysis steps



Which data should I keep?

Research data must remain analysable and
interpretable

- Preservation of presentation and analysis environments
- Documentation of annotation and analysis processes



What can be removed?

- Empty data records / tables
- Duplicates
- Incomplete data records
- Versions (which?)

„Poor data quality can be worse than missing data because it can waste resources and lead to faulty ideas and solutions, or at minimum challenges trust in the results and implications drawn from the data. Improvement in data quality can thus have significant benefits. (Curating for Quality P5)“

Risk factors for research data

- The departure of those responsible
- Data and backups get "lost"
- Security issues
 - non-updated software versions
 - insufficiently secured accesses (passwords)
- Technological development
 - file formats
 - functionality
- Change of policies or services of the university
 - elimination of service offers

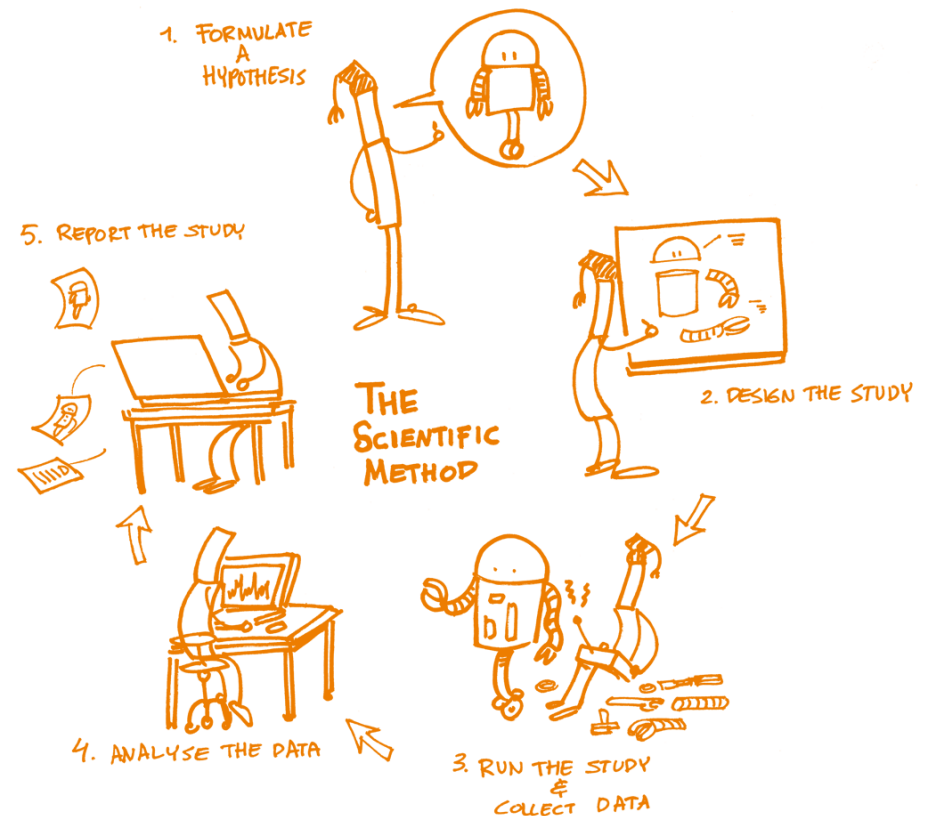


Data management

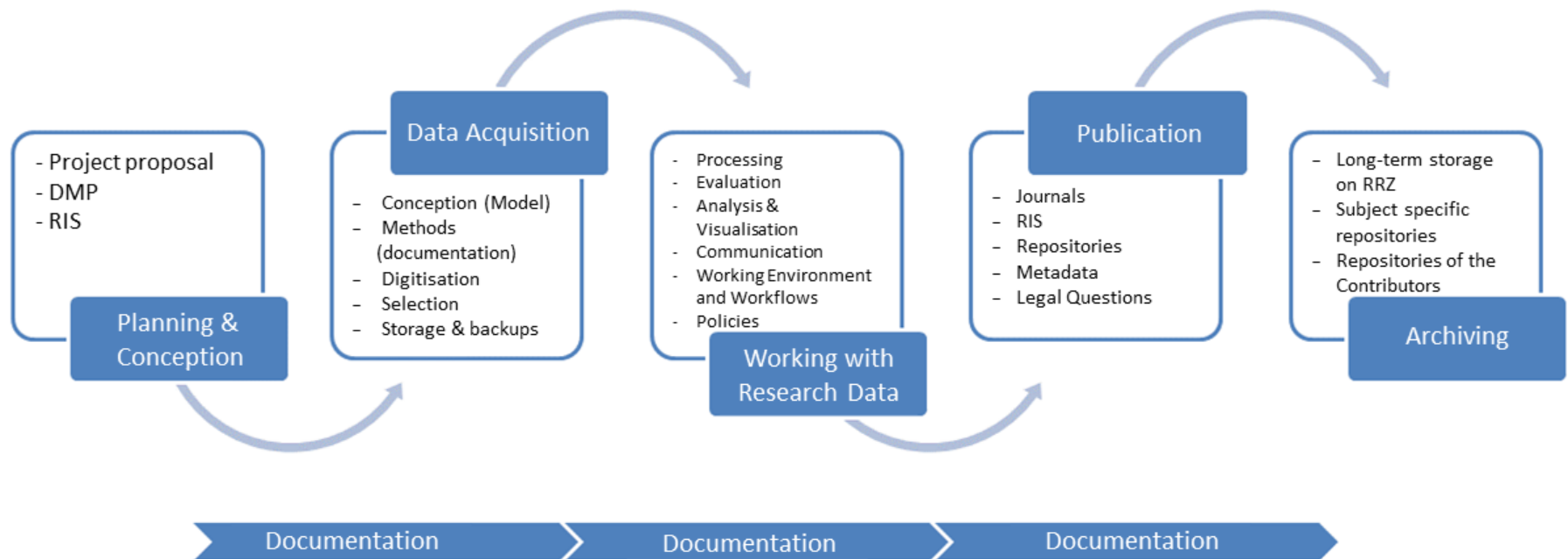
- What tools do you and your colleagues use?
- How did you come in contact with data management plans?
- Are there any instructions from supervisors?

Exercise DMP without assistance

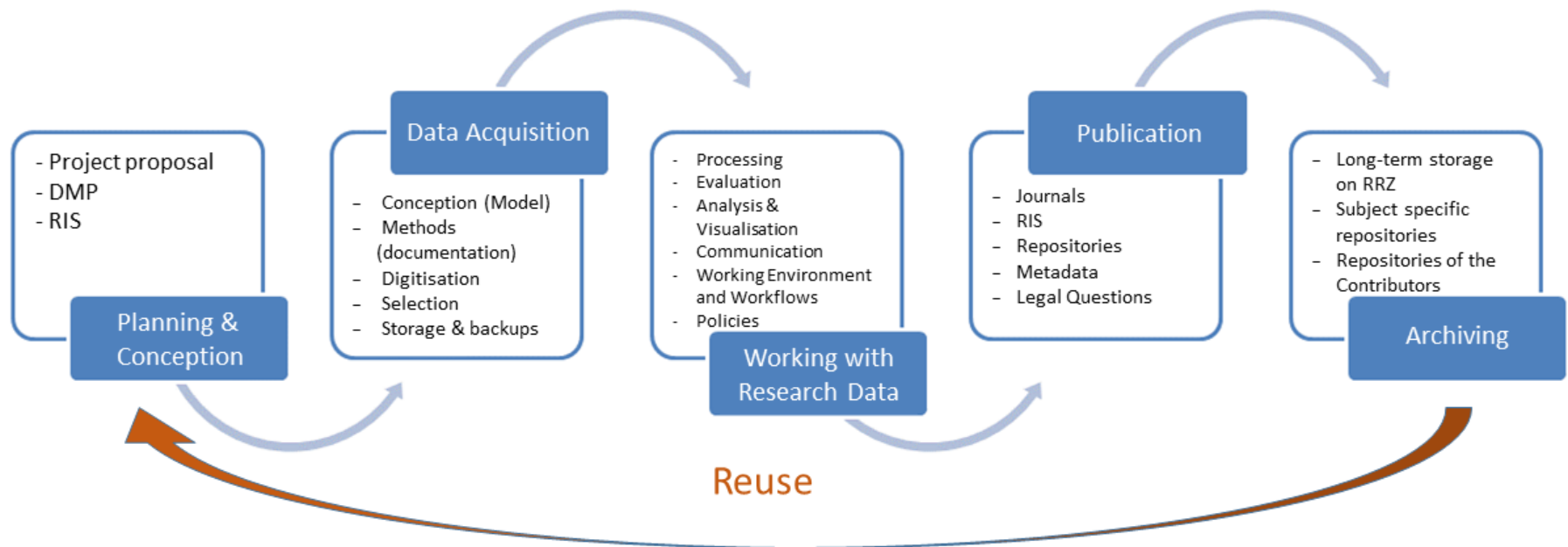
- In the group, create a data management plan for a (fictitious) project
- Consider and note for the individual phases of your project which steps and tools are necessary in data management (20min)
- Introduce your poster briefly to the other seminar participants (5min)



Phases of Research Data Management



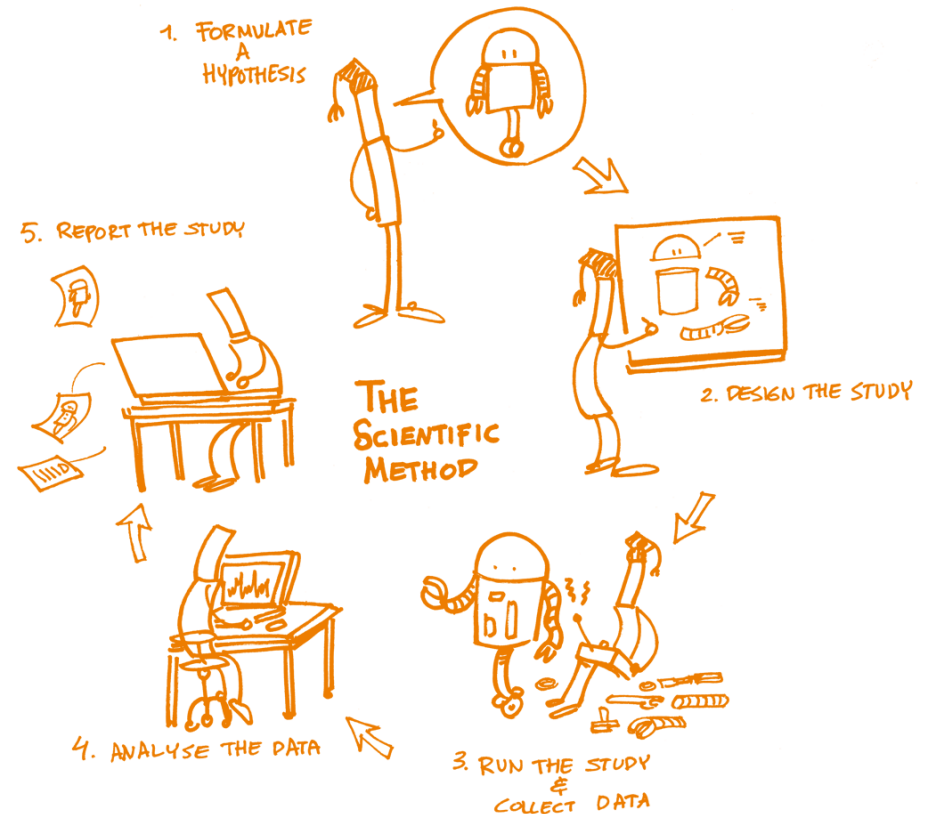
Phases of Research Data Management



Data Management Plan

Data Management plan – Definition

A data management plan (DMP) is a document describing the life cycle of data (in particular research data) from collection to archiving, including all measures taken to ensure that the data remains available, usable and traceable.





Source: Cessda Training

Elements of a DMP



1. Project description

- General information on the project such as objectives, promoter and duration
- Relevant policies

Source:[https://www.uni-bielefeld.de/\(en\)/forschungsdaten/angebot/dmp/](https://www.uni-bielefeld.de/(en)/forschungsdaten/angebot/dmp/)

Elements of a DMP



2. Existing types of data

- Description of the existing data that can be reused for the project and their integration into the project.
- Reason, if no data is reused

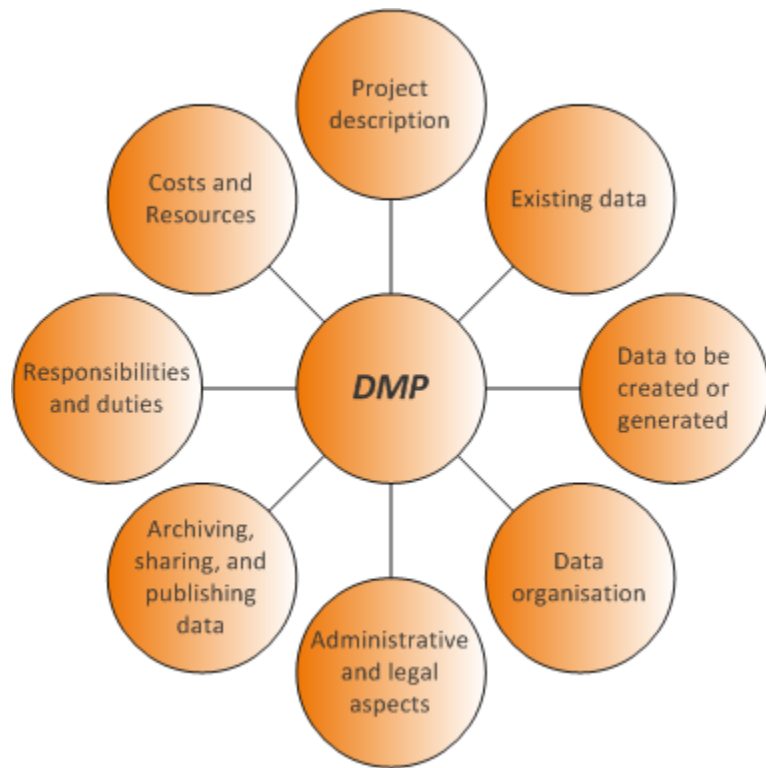
Elements of a DMP



3. Types of data to be generated

- Information on data types and formats
- Estimated data volume
- Information on the process of data generation and quality assurance (e.g. multiple sampling, validation, data peer review)
- Relevance of reuse by others

Elements of a DMP



4. Data organisation

- Information on the uniform regulation of data management within the project e.g. for:
 - Data storage
 - Data naming, folder structure
 - Synchronisation, Versioning
 - Collaborative working
 - Documentation
 - Creation of metadata

Elements of a DMP



5. Administrative and legal aspects

- Funding and legal requirements
- Copyright / Owner of data rights
- Access and use
- Data protection - for sensitive and personal data
- Data backup/ encryption and backups

Elements of a DMP



6. Archiving, data exchange and publication

- Which data types are published?
- Which repository/archive?
- Access options (free, restricted, etc.), license
- Time and financial framework of data preparation, provision or archiving
- Reasons if no publication is planned
- Information on the procedure after the end of the retention period

Elements of a DMP



7. Responsibilities and duties

How are the responsibilities for data management defined and distributed within the project?

Elements of a DMP



8. Costs and resources

Information on the costs and personnel expenditure for maintaining the data management plan, as well as ongoing costs for data curation, creation of metadata, archiving, etc.

Planning an RDM Budget

- Personnel costs (data manager, data processor, data curator)
- Infrastructure (equipment, software, etc.)
- Service costs (data, long-term archiving, publication, training)



Data management plans according to RDMO

- Optimisation of research data management before project start
- Guidelines for handling research data over the entire course of the project
- Planning the whereabouts of the research data after the end of the project
 - publication
 - long-term archiving
- Increasing the quality and efficiency of scientific work
- More than just a request from the funding agency
- In the ideal case, profit for the researcher and the institution

Exercise - RDMO (20min)

Research Data Management Organiser is a tool for creating data management plans with various templates

<https://dmp.fdm.uni-hamburg.de/>

- Create a DMP with your neighbour for your project
- Answer the questions of the RDMO catalogue
- Add your group members to the project as members
- Distribute tasks
- Export your DMP in the template for Horizon 2020

