DUTCH TECHCENTRE FOR LIFE SCIENCES

proEVlifecycle session FAIR & data stewardship

October 14, 2020

Mijke Jetten, Celia van Gelder
mijke.jetten@dtls.nl, celia.van.gelder@dtls.nl

Session content

1. Welcome and introduction
2. FAIR data stewardship and the FAIR data principles
3. Data life cycle
4. Ten practices: ‘prepare well to prevent data disaster’
5. Summary of group efforts and FAIR data stewardship
6. Summary slides (per topic of the data life cycle)

Welcome and introduction

Your needs

Via the assignment, you, as a team, have provided us with the content of today's session.
We invite you to participate actively, ask questions, use examples etc. so we can even more tailor the next one-and-half hour to your needs.
We, as trainers, are very flexible: this session is a success if you feel that it is time well spent!

Objectives

- By the end of this session, you will be able to recognise the basics of FAIR data stewardship in the various stages of the data life cycle
- And you will be able to evaluate what actions need to be taken to solve FAIR data stewardship issues in the proEVlifecycle research project

Structure

- Based on the assignment, we will indicate your project team's strengths and challenges
- We provide short summary slides with tips and tricks

Do you have questions? Feel free to ask them, any time!

FAIR data stewardship

Data stewardship

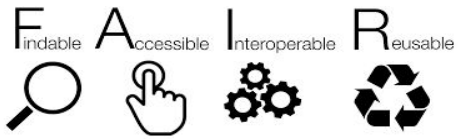
Responsible planning and executing of all actions on digital data before, during and after a research project, with the aim of optimizing the usability, reusability and reproducibility of the resulting data

Long term view

Data stewardship focuses on the long term, stimulating researchers together with their data support staff (data stewards) to plan data in such a way that is maximises the reuse of data

Data stewardship benefits the researcher

- Data is output! Data stewardship helps you to make conscious decisions about the data in your project
- Data stewardship prevents unauthorised access, avoids data loss, and facilitates the documentation and reuse of data
- Data stewardship stimulates efficiency and helps you comply to the conditions set to data by universities, funders and journals
- Data stewardship stimulates open science, FAIR data and helps meeting privacy- and security standards
- However, rule of thumb: *as open as possible, as closed as necessary*



FAIR data principles

- **FAIR for human beings, for machines and taking into account legal requirements (such as privacy)** FAIR refers to data (or any digital object), metadata (information about that digital object), and infrastructure

- The first step in (re)using data is to making data findable. Metadata and data should be easy to find for both humans and computers. For that, rich and machine-readable metadata and persistent identifier are essential

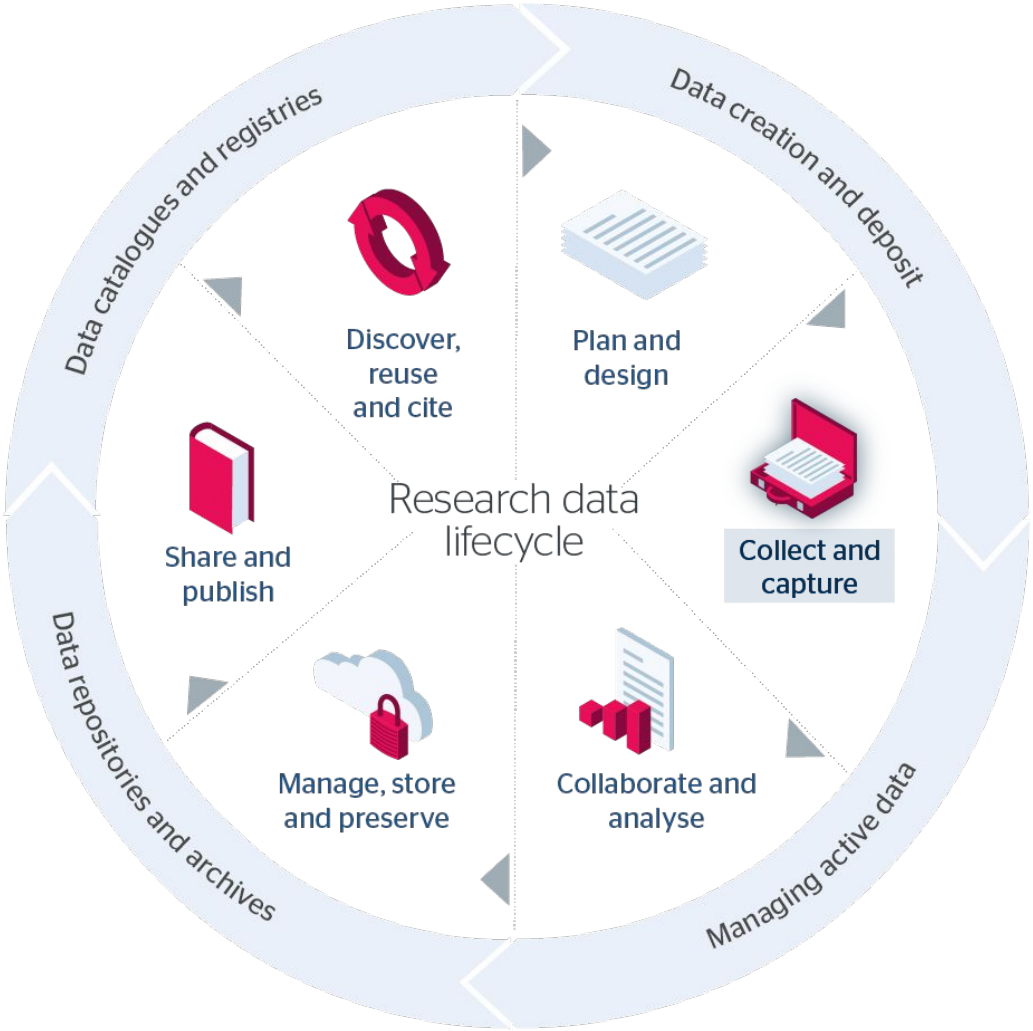
- The data must be accessible. A description should exist of how the data can be obtained. And there should be guarantees that this will still work after years.

- The data should be interoperable. It should be structured in well described way, with standardised metadata and vocabularies. This way, it can be integrated with other data, and can operates with applications or workflows for analysis, storage, and processing.

- The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described, including a clear license, so that they can be replicated and/or combined in different settings.

https://www.go-fair.org/fair-principles/
https://www.go-fair.org/how-to-go-fair/

Data life cycle



https://rdmtoolkit.jisc.ac.uk/research-data-lifecycle
https://ukdataservice.ac.uk/manage-data/lifecycle.aspx

Data life cycle: planning data

Design research; plan data management; plan consent for sharing; plan data collection, processing protocols and templates; explore existing data resources

Topics

- Writing a data management plan
- Informed consent procedures

You'll find a summary slide with basic information for each of these topics at the end of this presentation

Data life cycle: collecting data

Collect data; capture data with metadata; acquire existing third party data

Topics

- Finding data
- Collecting data
- Preregistration
- Data security
- Personal data

You'll find a summary slide with basic information for each of these topics at the end of this presentation

Data life cycle: processing and analysing data

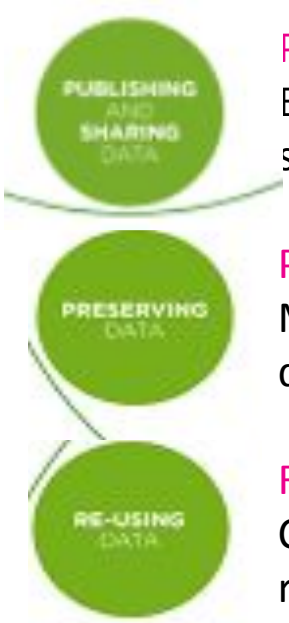
Enter, digitize, transcribe and translate data; check, validate, clean, anonymize; derive data; describe; manage and store data; analyse and interpret data; produce research output; cite data sources

Topics

- Data storage
- Data organisation
- Data versioning
- Data documentation

You'll find a summary slide with basic information for each of these topics at the end of this presentation

Data life cycle: publishing, preserving & reusing data



Publishing and sharing data

Establish copyright; create user documentation; create discovery metadata; select appropriate access to data; publish/share data; promote data

Preserving data

Migrate data to best format/media; store and backup data; create preservation documentation; preserve and curate data

Re-using data

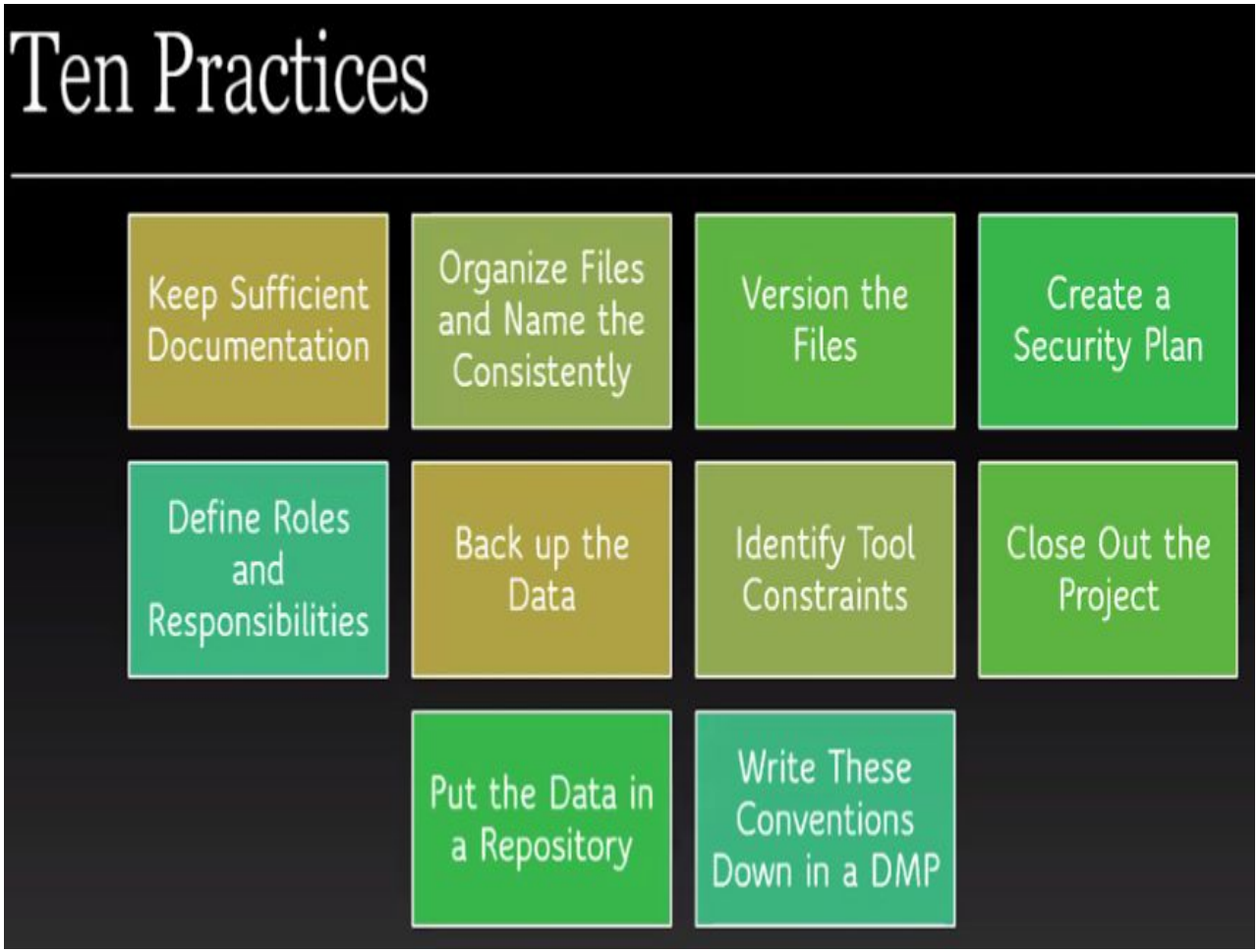
Conduct secondary analyses; undertake follow-up research; conduct research reviews; scrutinize fundings; use data for teaching and learning

Topics

- Data archiving
- FAIR software
- Data rights

You'll find a summary slide with basic information for each of these topics at the end of this presentation

Ten practices: ‘prepare well to prevent data disaster’



Briney KA, Coates H, Goben A (2020) Foundational Practices of Research Data Management. Research Ideas and Outcomes 6: e56508. https://doi.org/10.3897/rio.6.e56508

Practice 1. Keep sufficient documentation

What are you already putting into practice?

- Multiple locations for (meta)data storage
- Record of activities (preferably digital via Electronic Lab Notebooks: Jupyter, Labguru, Labfolder, OSF)
- Record of unusual observations
- Annotated code and documented scripts
- Analysis reports
- Documenting scripts
- Record of methods, workflows and protocols (including deviations!)
- Record of materials used
- Record of experimental conditions
- Record of papers citations (reference management: EndNote, Mendeley)
- Record of ‘who dunnit’
- On daily basis, in detail and backup of documentation
- Regular summaries

Practice 1. Keep sufficient documentation

What may be improved in your current practices?

Use the best practices in the previous slide!

And …

- Create documentation standards/templates to ensure recording of the same information (group effort)
- Take the ‘outsider’ perspective: can others understand my data activities? (group effort)
- Document so that research is reproducible (document more and more)
- Don’t wait: document your data right away

Practice 2. Organise files and name them consistently

What are you already putting into practice?

- International date convention (YYYY-MM-DD)
- Organise folders and subfolders (from the ‘search’ perspective; per project)
- Self-explanatory naming and abbreviations
- Quick overview, easy to scan
- Make use of standards and include version number
  Person_project_experiment_date
  Date_name_keywords
  Project_person_data type_experiment_date
- Create tables to easily find data (number experiment, date, description, and observations)

Practice 2. Organise files and name them consistently

What may be improved in your current practices?

Use the best practices in the previous slide!
“Folder management could be done better”

And …

- Record the structure you choose, and create standards/templates to ensure organising in the same way (group effort)
- Separate raw, analysed, processed data
- Separate ongoing from closed work (milestone versions)
- If you use abbreviations, make sure to explain them
- Keeping folder and file names as short as possible
- Apply the same to physical data (samples)
- Don’t wait: organise your data right away

Practice 3. Version the files

What are you already putting into practice?

- Keeping an untouched copy of the original file or raw data that won't be overwritten
- Periodically save new versions of a file
- Keep master of raw/start data, modified copy and its metadata
- Add the version in the file name (v1, v2, v3 etc.)
- Save all the important documents in at least two versions
- Each new analysis of the same experiment or revision is being numbered consistently

Practice 3. Version the files

What may be improved in your current practices?

Use the best practices in the previous slide!
“I usually - only? - safe the last draft”

And …

- Use a version control system such as Git (mentioned multiple times) (group effort)
- Alternatively, table at the beginning of a document with version, changes done and who (in parallel to naming the versions)
- Having a common group-platform for corrections or annotations (group effort)
- Especially when working with others: add date to version (to prevent all those multiple ‘final’ versions)
- More consistent version names, standardise versioning (group effort)

Practice 4. Create a security plan

What are you already putting into practice?

- Security controls: who has access to the data during the project (logging systems)
- Following company security guidelines and platforms (such as the DRE) for patient data, with limited access
- Safeguard your/others intellectual property data/protocols
- Use encryption or password protection
- Use pseudonyms, instead of real names, in case sharing of information can’t be avoided (this is still personal data!)
- Delink patient data and samples (good key list management)

Practice 4. Create a security plan

What may be improved in your current practices?

Use the best practices in the previous slide!
“I don’t work with personal or confidential data”

And …

- Regularly review the security plan, together with the institution’s security officer
- Be aware of your own responsibilities in the plan
- Explore your institution’s storage solutions, such as the DRE (https://www.researchenvironment.org/) (group effort)
- Discuss what happens with the data after your leave (group effort)
- Make sure not to lose passwords, encryption keys or key files

Practice 5. Define roles and responsibilities

What are you already putting into practice?

- Compliance training and safety protocols in the lab
- Training on the specific projects and each person's contributions
- Organise regular (lab) meetings to define, distinguish and update roles
- Transparent communication with lab members and other colleagues (including the PI)

Practice 5. Define roles and responsibilities

What may be improved in your current practices?

Use the best practices in the previous slide!

And …

- Formally appoint roles and responsibilities (group effort)
- Document responsibilities
- Standardise procedures (SOPs) before the project (group effort)
- Document change management during the project (group effort)
- Define good data management practices (use this workshop tips)
- Knowledge exchange and training (mentors) (group effort)
- Draft out ‘best practices’ (for instance for onboarding and exit processes)
- Realise you have multiple roles: as project member, as institute member, etc.

Practice 6. Back up the data

What are you already putting into practice?

- Different locations (also geographically)
- Different types of storage locations
- Backed up by the company/collaborators
- Backed up on the university's shared network
- Daily backups
- Backup on external hard drives

Practice 6. Back up the data

What may be improved in your current practices?

Use the best practices in the previous slide!

And …

- Commit to the (automated) backup procedures of your institution (group effort)
- Periodically confirm that backups are functioning properly to avoid unpleasant surprises when trying to recover data in a crisis
- Learn how to restore files (condition: good data organisation, including clear file names etc.)
- Implement the 3:2:1 rule: have at least three copies of your data; store the copies on two different media; keep it safe with one backup copy offsite
- Back up on laptop: don’t loose your laptop

Practice 7. Identify tool constraints

What are you already putting into practice?

- Research IT support and information security offices are available
- Data routinely stored in spreadsheets or databases
- Tools in place in handling and sharing large data (Laboratory Information Management System)
- Make a brief assessment of the available tools before including it in the pipeline
- Learning to use new tools for data management
- Some devices do not give us the wished output, in this case just try and optimise the procedure to get wanted data
- We always export the data using compatible formats to analyse with the implemented tools chosen in the lab
- Speed limitation for transferring huge data from storage hard/cloud to re-analyse the data when needed

Practice 7. Identify tool constraints

What may be improved in your current practices?

Use the best practices in the previous slide!
"A short course for learning new software tools that helps in data management would be helpful" (group effort)

And …

- Evaluating in the security plan both individual software tools and any potential security gaps between tools
- Also include these discussions in the documentation of the project
- Better organising of the data once collected
- Find Python packages or R packages for conversion or data visualisation
- To develop a secure and feasible research workflow (group effort)
- Discussing a more rapid technology for data transfer (group effort)

Practice 8. Close out the project

What are you already putting into practice?

- Snapshotting while working on the project
- Creating master copies of all the key-files in order to improve future retrieval and reduce efforts in case of data sharing or talks/posters presentations
- All datafiles produced are indexed in a database which makes it easier to use in future works (documentation!)
- Or, create your own overview of locations
- Write down the decisions in the project
- If I think something is unnecessary, I delete the file immediately
- Collecting protocols, different data in a separate, organised files
- Keep all the data in open formats
- Master copies containing the outcome of data analysis (i. e. graphs and tables) are stored every week on cloud and secure folder

Practice 8. Close out the project

What may be improved in your current practices?

Use the best practices in the previous slide!

And …

- Storage hardware should be updated before the hardware format falls out of regular usage and becomes difficult to read
- Regularly structure and organise outputs (group efforts)
- Make a master copy or specific copies with raw data and final data utilised for a specific project
- Implement the snapshotting at key points of the project (group effort)

Practice 9. Put the data in a repository

What are you already putting into practice?

- Ensure that the selected data is clean, error free and finalised, so it can be archived
- Created a data repository with accompanying log files for previous consortium
- We have a personal data repository in Anaxomics
- Use FTP via our own platform to make data public
- Currently I do not have any data that would be eligible for repository

Practice 9. Put the data in a repository

What may be improved in your current practices?

Use the best practices in the previous slide!

And …

- Discuss with group to using external repositories, once applicable (group effort)
- Make a plan when to put data in a repository (group effort)
- Decide about the selection of data
- Training about data management practices for data archiving would be helpful (group effort)
- Make data publically available for reproducibility and reuse (group effort)
- Checking the understandability of final data
- Discussing about the responsible person (i.e. funder, PI, ESR) for general data sharing
- Check whether the DRE operator can provide access to the data to others at final stage of the project (group effort)

Practice 10. Write these conventions down in a DMP

What are you already putting into practice?

- Acknowledge the importance of having a DMP
- DMP of the consortium

Practice 10. Write these conventions down in a DMP

What may be improved in your current practices?

Use the best practices in the previous slide!

And …

- Creating a DMP if you didn't do so yet, per project or as joint effort (group effort)
- For shared projects discuss DMP setup with others involved
- Have your DMP checked by your local RDM support office
- Periodically discuss and update the DMP with the group and with your PI (group effort)

Summary of group efforts: first steps

- Create **standards** and **templates** (for documentation, organisation/structures of files, file names, versioning, SOPs, change management)
- Use a **version control system** such as Git (and also use it for corrections and annotations)
- Explore your institution's **storage solutions**, such as the DRE, and commit to the (automated) backup procedures of your institution
- Develop a secure and feasible **research workflow** (including working towards future data integration)
- Regularly **structure** and **organise outputs** (including snapshotting at key points of the project)
- Discuss with group to using external **repositories** (including making plans: when, what and how, aimed at reproducibility and reusability)
- Create a **DMP** if you didn't do so yet, per project or as joint effort, and periodically discuss and update it
- Formally appoint **roles** and **responsibilities**
- **Knowledge exchange** and training (including exploring tools together, learn how to prepare for data archiving, learn from best practices, and check the understandability of each others data activities)

Second step: FAIR data

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Core FDP Metadata

	FDP
1	Title
2	Description
3	Publisher
4	Version
5	Language
6	License
7	Start date
8	Last update
9	Institution
	Catalog metadata
1	Title
2	Description
3	Publisher
4	Version
5	Language
	Dataset metadata
1	Title
2	Description
3	Publisher
4	Version
5	Language
6	License
7	Issued
8	Modified
9	Keywords
10	Theme
11	Contact point
12	Landing page
	Distribution Metadata
1	Title
2	Description
3	License
4	Issued
5	Modified
6	Download URL
7	Access URL
8	Media Type
9	Format
10	Byte Size

Metadata: the information we create, store, and share to describe things, allows us to interact with these things to obtain the knowledge we need
https://www.niso.org/publications/understanding-metadata-2017

FAIR for human beings, for machines and taking into account legal requirements (such as privacy)

After the first steps (see previous page), a second step could be a Bring Your Own Data Event in which we together work on creating FAIR data

Metadata for Machines Workshop

M4M

To assist domain experts who want to produce FAIR resources, in creating machine-actionable metadata that **satisfies** and **adheres** to the FAIR Principles.

Summary slides (per topic of the data life cycle)

Data life cycle: planning data

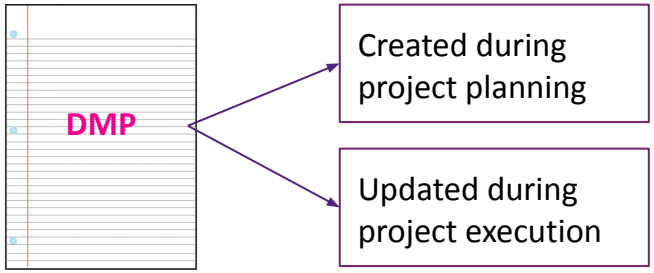
Design research; plan data management; plan consent for sharing; plan data collection, processing protocols and templates; explore existing data resources

Topics

- Writing a data management plan
- Informed consent procedures

Summary slide: writing a data management plan

A data management plan (DMP) is a document which outlines how research data will be managed over the course of a research project



Organisation; roles; costs; existing data; collection process; informed consent; ethics committee; privacy; security; data types; tools for storing, sharing; structuring data; metadata and documentation; preserving; giving access to data etc.

Nice to watch: the data management planning knowledge clip by Oncode, including extra materials

Benefits (vs.'ticking boxes')

- Makes research more efficient; think and decide timely about RDM issues
- Use it as a dynamic document; use it as a discussion document
- Useful in meetings for monitoring progress of your research

Policies & DMPs

- DMPs are used to check the 'local feasibility' of a study, i.e. a quality assurance (safety, quality, expertise) of trials executed, to acquire approval from the Executive Board
- Most funders require researchers to submit a DMP in the context of a grant application, in addition to a Data Management Paragraph

Summary slide: informed consent procedures

Agreement between the researcher and the data subject (i.e. participant), including:

- The data subject is **informed**: provide information that is received and understood (information brochure)
- The data subject gives **consent**: you need an explicit statement that the data subject freely agrees to participation in the research project

Legal perspective

As a legal base for collecting personal data (GDPR)

Ethical perspective

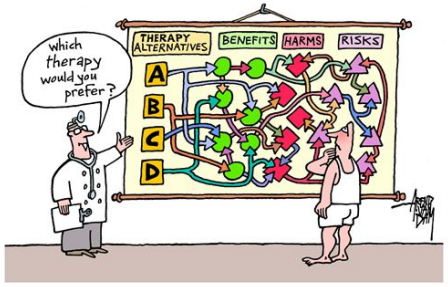
Participant is informed and thus enabled to make a voluntary decision about accepting or declining participation in research

Written informed consent (often: obliged for WMO research)

Oral informed consent

Online informed consent

Always check the informed consent templates from your institute/ethical committee



informed consent

Data life cycle: collecting data

Collect data; capture data with metadata; acquire existing third party data

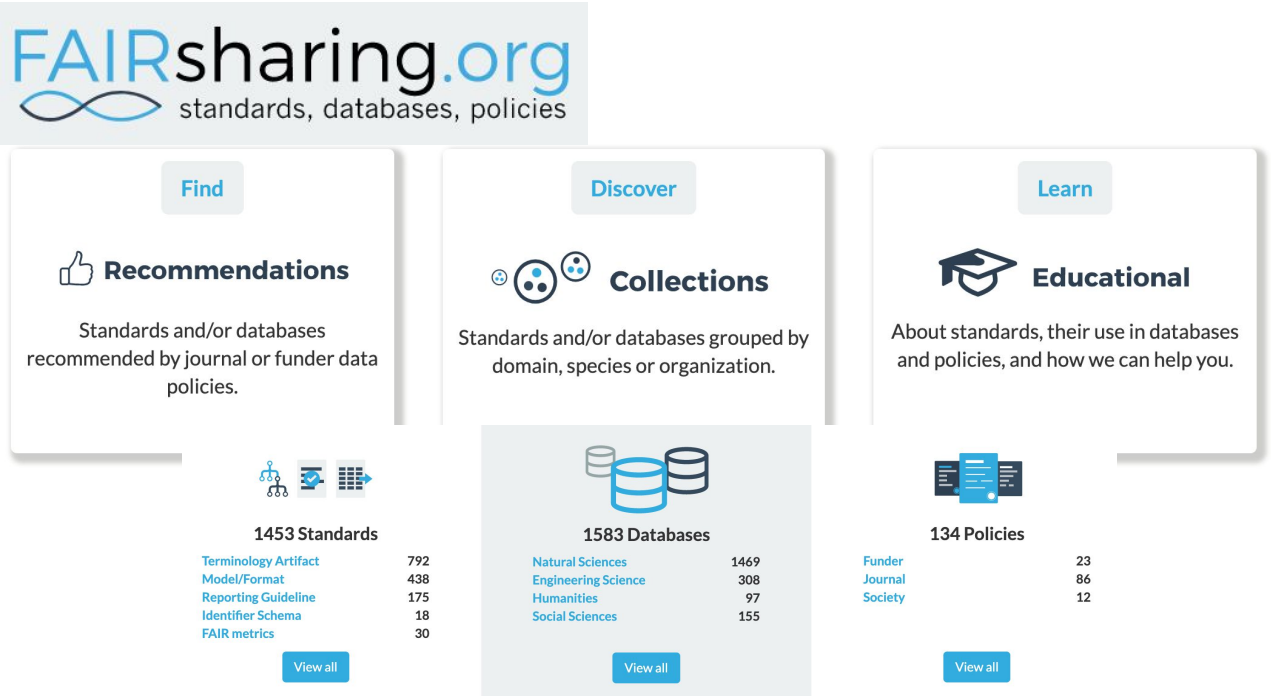
Topics

- Finding data
- Collecting data
- Preregistration
- Data security
- Personal data

Summary slide: finding data

If you intend to reuse existing data instead of collecting it yourself, there are good sources for potentially relevant existing data. The following (directories of) archives may be relevant sources for finding data.

- Open Access Directory: Data repositories
- Re3data.org
- Recommended Data Repositories from Nature
- Fairsharing



Summary slide: collecting data

https://libguides.vu.nl/rdm/data-collection

Data collection

Data collection may consist of the re-use of existing data and/or the generation of new data. You can find more specific information on the re-use of existing data on the Finding Existing Data page in this LibGuide

For data to be considered valid and reliable, data collection should occur consistently and systematically throughout the course of the research project. Data collection guidelines and established methodologies should be used to gather data. Some disciplines make use of codebooks, whereas others use protocols for data gathering. These procedures help researchers collect data according to conventional methodological steps. If a research project involves multiple partners (in a consortium) it should be clear who is responsible for the collection of what (part of the) data. Important aspects of data collection include:

- Standardisation: codebooks & protocols
- Structure / organisation of the data
- Data quality assurance methods
- Documentation & metadata
- Storage & protection

This relates to the Reproducibility of your research according to the FAIR-data principles.

Nice to watch: the collecting data clip by the VU

Summary slide: preregistration

Preregistration allows the research community to get information about upcoming research, and provide feedback on the research plans, including the protocols, methods, etc. As a result, research plans can learn from and align with other research.

- Makes your science better by increasing the credibility of your results
- Allows you to stake your claim to your ideas earlier
- It's an easy way to plan for better research
- Increase transparency
- Avoid unnecessary duplication of animal studies
- Reduce reporting bias, such as publication bias and bias induced by selective outcome reporting, p-hacking and HARKing
- Increase data sharing, by
  - allowing fellow researchers and reviewers to access information on the study design, which is often lacking in publications
  - provide a platform to share details and data of otherwise unpublished animal studies
- Create opportunities for collaborative research

OSF
COS
CENTER FOR OPEN SCIENCE
PREREGISTERED

https://www.cos.io/our-services/prereg

protocols.io

Making it easy to share method details **before**, **during**, and **after publication**.

Summary slide: data security

Location of data

- Hard-copy personal and confidential data has to be kept in a locker
- In case of personal and confidential data, the university drive is often the only location allowed
- Public (free) storage or transfer services are no option for personal or confidential data

Encrypt your devices

- Use Filevault (Mac), Bitlocker (Windows) or the open source VeraCrypt to encrypt your device

Encrypt your data

- If it is necessary to keep your data at an untrusted storage location or you need to transport it over a network, you can encrypt your data using for instance VeraCrypt or Axcrypt

Use VPN to transfer data

- The best way to ensure safe transport between the university network drive and your home computer is to use the VPN connection of the university

Wipe storage devices after use

- There are simple tools that are able to retrieve information after you deleted it. To secure your device, utilities as Permanent Eraser and CCleaner can be used to destroy the data
- To erase all data from your device you can use for instance Dban

https://www.ru.nl/rdm/collecting-data/data-security

Summary slide: personal data

Anonymisation

Re-identification of the anonymised data combined with any other population data is impossible

Pseudonymisation

Personal data can no longer be attributed to a specific data subject without the use of additional information (pseudonymisation key)

Short movie about privacy by design

- Data minimisation
- Data quality
- Goal setting
- Minimisation of use
- Security measures
- Transparency
- Rights of data subjects
- Liability

Make sure to check your local RDM policy, Standard Operating Procedures (SOPs), GDPR proof templates, and contact your legal office and/or privacy officer in case of questions and if you work with third parties or third party tools

Data life cycle: processing and analysing data

Enter, digitize, transcribe and translate data; check, validate, clean, anonymize; derive data; describe; manage and store data; analyse and interpret data; produce research output; cite data sources

Topics

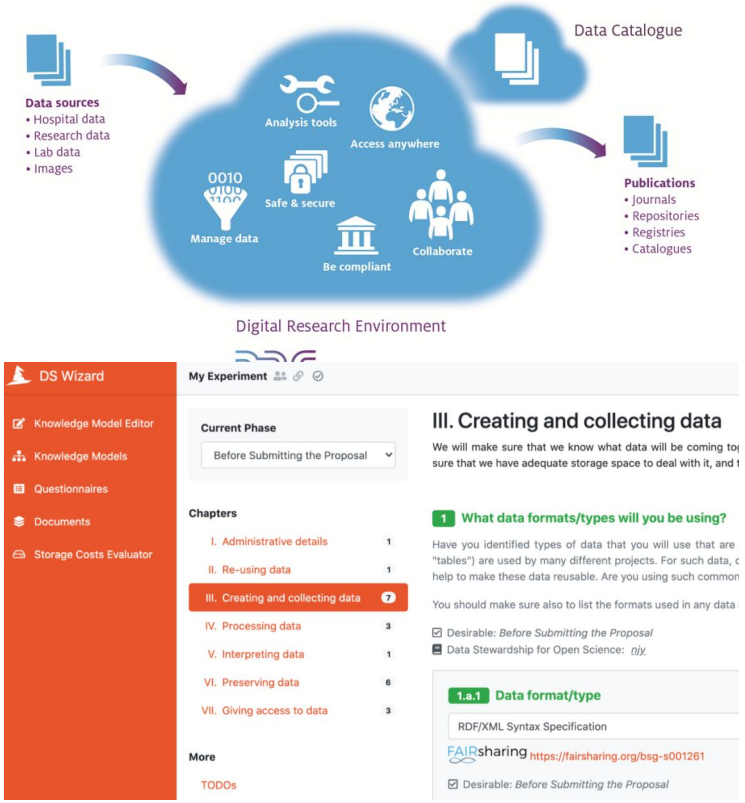
- Data storage
- Data organisation
- Data versioning
- Data documentation

Summary slide: data storage

The ideal tools allow you to do much more than only storing ...

- Import data sources, organise data, store data securely
- Collaborate on data
- Access anywhere, any time, on any device
- Data processing and analysis tools
- Fully scalable data quantity and compute
- High-performance computing
- Controlled access and logging
- Use, manage, combine and re-use data
- Use the research tools and applications you want
- Handle multiple types of data: clinical, images, omics, etc.
- Generate virtual workspaces for researchers
- Pseudonymization software

Such as … virtual/digital research environments, secured university storage, SURF services, or: ask the data steward wizard (https://ds-wizard.org/)

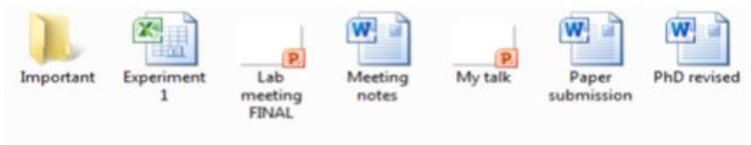


SURFdrive: store and share your files securely in the cloud

Research Drive: securely and easily store and share research data

SURFfilesender: send large files securely and encrypted

Summary slide: data organisation



In 3 years time would you know what these are?

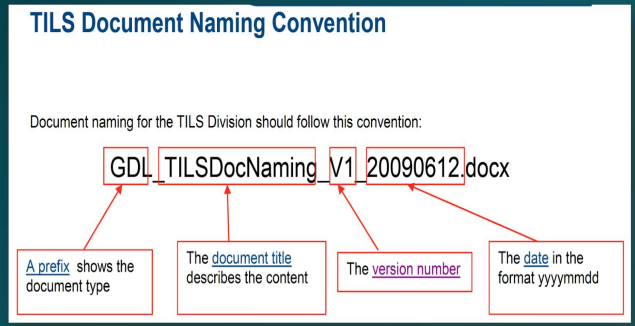
Original slide by Marta Teperek

File Naming Conventions

20190527_HelisAcademy

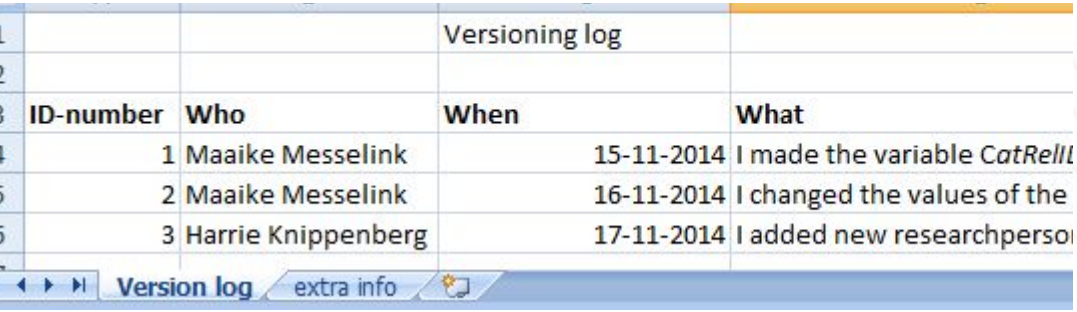
- Date or date range of experiment: YYYYMMDD
- File type
- Researcher name/initials
- Version number of file
- Don't make file names too long
- Avoid special characters and spaces
- Include a README.txt file to explain the naming convention

File naming convention:



Summary slide: data versioning

- Use a 'revision' numbering system. Example: v03_01
- Add information to identify the individual who has made the amendments & a date stamp. Example: 20190508_datav01_SJ
- Decide how many versions you want to save, which versions to keep and for how long
- Identify milestone versions and a raw data version, which can never be altered or deleted. When working with others on data, maintain a master file
- Record the changes that are made in a new version by using a version log
  Version control can also be maintained by version-control facilities in the software you are using or in special versioning software



Summary slide: data documentation

Producing high-quality documentation in the course of your research ensures that your data can be:

- Properly interpreted as relevant context is available
- Verifiable and reproducible
- Reusable (by you or by others)

3 types of data documentation

- Files that explain the **content** of the dataset, at the data level (codebook)
- Files that explain the **context** of the dataset, and how the research was done (methodology section)
- Files that explain the **structure** of the dataset (readme.txt file with the structure of the dataset)

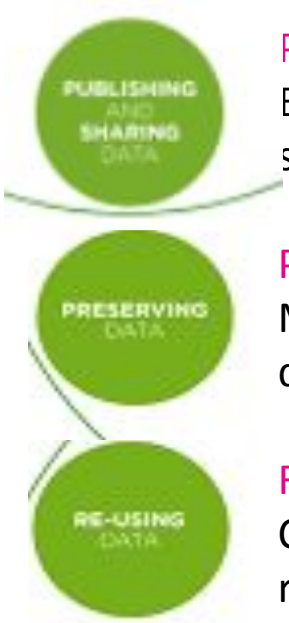
Embedded documentation

- Code, field and label descriptions
- Descriptive headers or summaries

Additional documentation

- Codebook (depending on the program)
- Readme file
- Methodology file
- Questionnaires or interview guides
- Working papers or lab notebooks

Data life cycle: publishing, preserving & reusing data



Publishing and sharing data

Establish copyright; create user documentation; create discovery metadata; select appropriate access to data; publish/share data; promote data

Preserving data

Migrate data to best format/media; store and backup data; create preservation documentation; preserve and curate data

Re-using data

Conduct secondary analyses; undertake follow-up research; conduct research reviews; scrutinize fundings; use data for teaching and learning

Topics

- Data archiving
- FAIR software
- Data rights

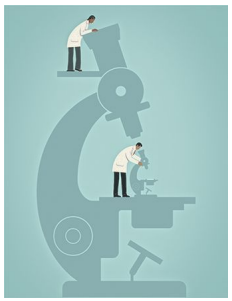
Summary slide: data archiving

Generally, RDM policies state something like this

- Data are stored at the time of publication of the research (including dissertations) at the latest, together with at least all the information necessary for potential reuse of data (metadata)
- The retention period for research data is a minimum of ten years

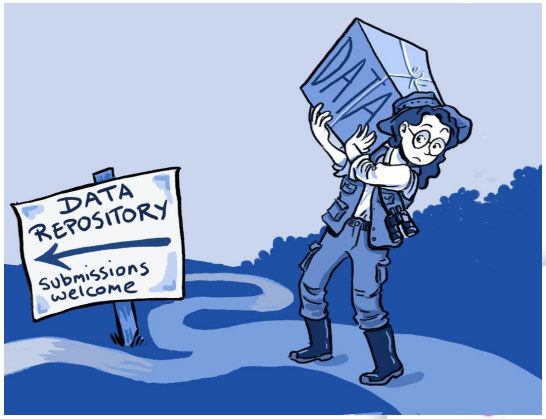
Two perspectives

- Archiving data for scientific integrity
- Archiving data for reuse
- Rule of thumb: *as open as possible, as closed as necessary*



For consideration

- Are there ethical and legal reasons not to share my data?
- Must all data be shared?
- Where is my data safe?
- Is my data in an easy to use format?
- Will my data be accessible in the long term?
- Do I have sufficient documentation and metadata?



Summary slide: data archiving

Repositories

ENA European Nucleotide Archive

DANS

DataverseNL

DRYAD

4TU.Centre for Research Data

European Genome-phenome Archive

PCDDB

figshare

zenodo

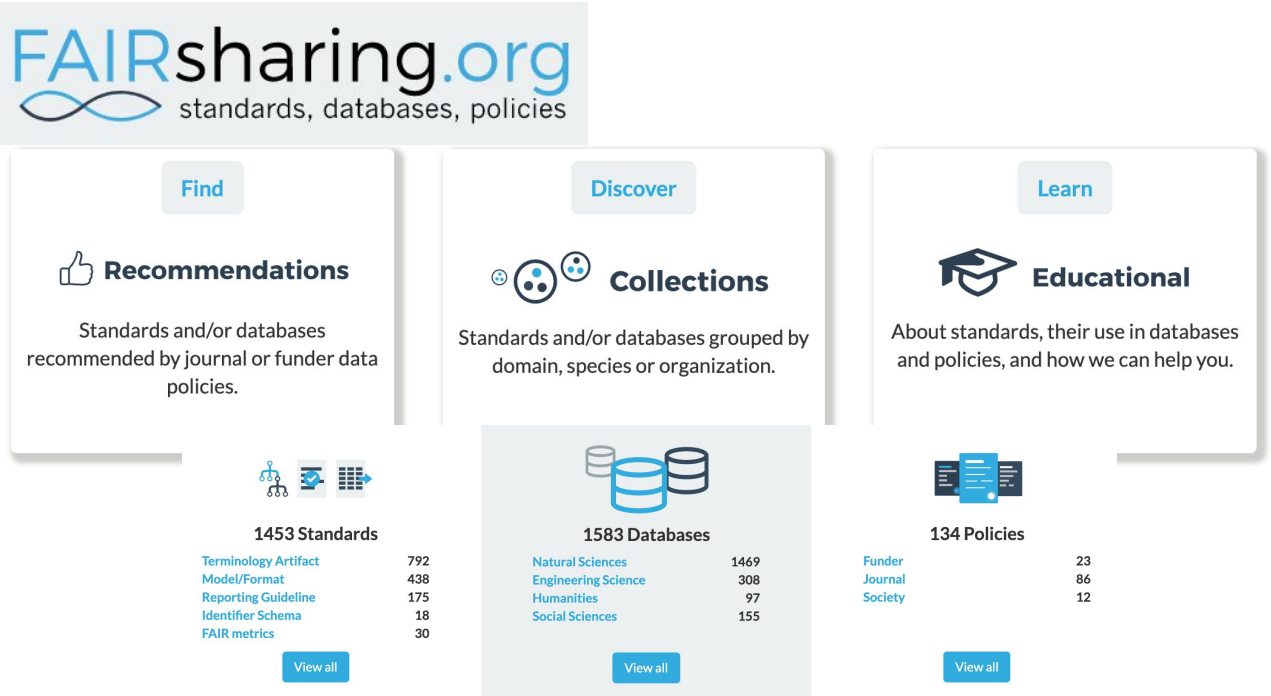
UniProt

IDR Image Data Resource

Summary slide: data archiving

If you intend to reuse existing data instead of collecting it yourself, there are good sources for potentially relevant existing data. The following (directories of) archives may be relevant sources for finding data.

- Open Access Directory: Data repositories
- Re3data.org
- Recommended Data Repositories from Nature
- Fairsharing



Summary slide: FAIR software

https://fair-software.nl/

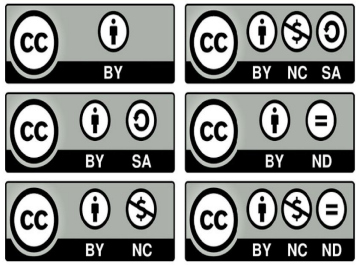
Summary slide: data rights

"No, it's MY data!"

Data ownership

Quite often … a funder, a university that pays your research, and not: you! Check your institution's policy, the funder's agreements, and the agreements made in the COSMIC project

Creative Commons licenses are also applicable to data



There are **existing licenses** that specifically apply to data. These are the so-called Open Data Commons, which can be divided into three licenses:

- Public Domain Dedication and License (PDDL)
- Attribution License (ODC-By)
- Open Database License (ODC-ODbL)

Make sure that you do not hand over any author rights belonging to your data via for instance a CC0 license. In that case, your work is dedicated to the public domain by waiving all the (copy)rights. Anyone can copy, modify and distribute your data, even for commercial purposes, all without asking permission or reference to your dataset.

Summary slide: data rights

Data use agreements

Data use agreements – also known as data transfer or exchange agreements – are contracts used for the transfer of data which are non-public or otherwise subject to restrictions

Agreement between the data owner and a recipient. Composing a data use agreement may be particular relevant in research that involves privacy-sensitive data.
Dutch examples approved by the University Medical Centers

Consider at least the following aspects:

- Legal aspects – including the General Data Protection Legislation (*Algemene Verordening Gegevensbescherming*, AVG) and the Medical Research involving Human Subjects Act (*Wet Medisch-Wetenschappelijk Onderzoek met Mensen*, *WMO*)
- Ownership of the data
- Privacy/anonymisation of human data
- Use of data by third parties
- Embargo period
- Citations and/or co-authorship
- Specific scientific purpose

Credits

Special thanks to

- Esther Plomp, Santosh Ilamparuthi & Marta Teperek from TU Delft
- Among other, Cees Hof (DANS), Jasmin Böhmer (UMCU), Christine Staiger (WUR) from the Helis DAS data stewardship training program
- Inge Slouwerhof, Maaike Messelink & Tina Reiling from Radboud University
- Mateusz Kuzak, Netherlands eScience Center

for reusing their slides