What you need to know to prepare a Data Management Plan
Where do we go?

- DANS
- EC’s Open Research Data Policy
- Planning FAIR data management
- Support for data management
- Practical exercise
Institute of Dutch Academy and Research Funding Organisation (KNAW & NWO) since 2005

Mission: promote and provide permanent access to digital research information

First predecessor dates back to 1964 (Steinmetz Foundation), Historical Data Archive 1989

https://dans.knaw.nl/en
DataverseNL for short- and mid-term data storage

NARCIS: Gateway to scholarly information in the Netherlands

EASY: certified long-term Electronic Archiving System for self-deposit;
> 36k data sets, ~70% Open
The what, why and how of data management planning

From the training *Essentials 4 Data Support* by Research Data Netherlands
EC’S OPEN RESEARCH DATA POLICY
Data Management in H2020 - summary

• Research data should be as open as possible, as closed as necessary.
• Deposit the data in a research data repository for sharing and preserving the data long term.
• “Sharing” means “outside the project consortium”.
• A Data Management (DMP) is due by month 6. It is a regular project deliverable.
• A DMP is a living document: to be used, updated and shared. You can use the Horizon 2020 template in DMPonline.
• Manage and document all data FAIRly, whether they will be open or not.
EC in the Guidelines: “This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept (...) without suggesting any specific technology, standard, or implementation solution”

The EC Open Research Data policy

Key sources of information

- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020

- Guidelines on FAIR Data Management in Horizon 2020

- Annotated model grant agreement, clause 29.3

- Infographic summarising key policy points

- Open Access and Data Management
  http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm
DATA MANAGEMENT PLAN
A DMP is a brief plan to define:
• how the data will be created
• how it will be documented
• who can access it
• where it will be stored
• whether it will be shared
• where it will be preserved

DMPs are sometimes submitted as part of grant applications, sometimes afterwards, but they are useful whenever researchers are creating data.
DMPonline

A web-based tool to help researchers write DMPs

https://dmponline.dcc.ac.uk

Choose your funder to get their specific template

Choose any additional optional guidance
PLANNING
DATA MANAGEMENT
Research data means data in the form of facts, observations, images, computer program results, recordings, measurements or experiences on which an argument, theory, test or hypothesis, or another research output is based. Data may be numerical, descriptive, visual or tactile. It may be raw, cleaned or processed, and may be held in any format or media.
Making data FAIR

• **Findable**
  – Assign persistent IDs, provide rich metadata, register in a searchable resource,

• **Accessible**
  – Retrievable by their ID using a standard protocol, metadata remain accessible even if data aren’t...

• **Interoperable**
  – Use formal, broadly applicable languages, use standard vocabularies, qualified references...

• **Reusable**
  – Rich, accurate metadata, clear licences, provenance, use of community standards...

www.force11.org/group/fairgroup/fairprinciples and http://www.nature.com/articles/sdata201618
Some “F” questions

§2.1 Making data findable, including provisions for metadata

- Use metadata and **specify standards for metadata creation** (if any). If there are no standards in your discipline **describe what type of metadata will be created and how**.
- Use search keywords
- Persistent and unique identifiers such as DOI
- File and folder naming conventions: see OpenAIRE-EUDAT July 2016 webinar
- Versioning of the datasets and clear version numbers
Metadata and documentation

• Metadata and documentation is needed to locate and understand research data.
• Use relevant standards to enable interoperability.

http://rd-alliance.github.io/metadata-directory

• Check what the long-term repository supports or expects.
Documentation?

- Code book explaining the variables
- Study design
- Lab journal
- iPython or Jupyter notebook
- Statistical queries
- Software or instruments to understand or reproduce the data
- Machine configurations
- Consent information
- Data usage licence
- ...

In short: document and preserve everything that is needed to reproduce the study – ideally following the standard in your discipline
Some “A” questions

§ 2.2 Making data openly accessible:

• Explain which data can’t be shared openly, if any
• Specify how access will be provided in case of restrictions, e.g. through a data committee, a license, or arranged with the repository.
• Will methods or software tools needed to access the data (if any) be included or documented?
• Deposit the data and associated metadata, documentation and code preferably in certified repositories which support Open Access.

Data Seal of Approval
ICSU World Data System
nestor seal
ISO 16363
Where to find a repository?

1. Use an external data archive or repository already established for your research domain to preserve the data according to recognised standards in your discipline. More information for selecting a data repository.

2. If available, use an institutional research data repository, or your research group’s established data management facilities.

3. Use a cost-free data repository such as Zenodo.

4. Search for other research data repositories in Re3data.org.

More information: https://www.openaire.eu/opendatapilot-repository

Zenodo: http://www.zenodo.org  Re3data.org: http://www.re3data.org
Storing or archiving? Both!

Storing and backing up files while research is active

- Likely to be on a networked filestore or hard drive
- Easy to change or delete

Archiving or preserving data in the long-term

- Likely to be deposited in a digital repository
- Safeguarded and preserved
Raw data employed to perform the algorithm used in the scientific paper: "Kinematic reconstruction of the human arm joints in robot-aided therapies with Hermes robot"

Garcia-Aracil, Nicolas M; Bertomeu-Motos, Arturo; Moraes, Ricardo; Lledó, Luis D; Díez, Jorge A.; Catalan, Jose M.

This file contains the raw data necessary to perform the algorithm introduced in the scientific paper:

PAPER: Kinematic reconstruction of the human arm joints in robot-aided therapies with Hermes robot

Authors: Arturo Bertomeu-Motos, Ricardo Morales, Luis D. Lledó, Jorge A. Díez, Jose M. Catalan, Nicolas Garcia-Aracil.


Raw data acquired necessary to perform the algorithm introduced in this paper.

a) Robot Joints: Robot joints generated to develop the simulation, in radians (1-7 columns). This robot is referenced in the paper.
b) Direct Upper Limb Joints: Upper limb joints generated to develop the simulation, in radians (q1-q7 columns). This data is used to simulate the accelerometer value.

This work is part of the project "Adaptive Multimodal Interfaces to Assist Disabled People in Daily Activities" funded by the European Commission’s Horizon 2020 Framework Programme, Call number ICT-22-2014: Multimodal and Natural computer interaction, under Grant Agreement No 645322.
Sharing data: what is meant?

With collaborators while research is active

Data are mutable

(Open) data sharing

Data are stable, searchable, citable, clearly licensed

Dataverse
Dropbox
Before clocks were invented, people kept time using different instruments to observe the Sun’s zenith at noon. Towns and cities set clocks based on sunsets and sunrises. Time calculation became a serious problem for people travelling by train, sometimes hundreds of miles in a day. UTC is the World's Time Standard.

A440, which has a frequency of 440 Hz, is the musical pitch above middle C and serves as a general tuning stand musical pitch. Prior to the standardization on 440 Hz, countries and organizations followed the Austrian government's 1885 recommendation of 435 Hz. In the period instrument movement, a consensus has arisen around a modern baroque pitch of 415 Hz (A♭ of A440), baroque for some special church music (Chorton pitch) at 466 Hz (A♯ of A440), and classical pitch at 430 Hz.

In the aftermath of the French Revolution (1789), the traditional units of measure used in the Ancien Régime were replaced. The livre monetary unit was replaced by the decimal franc, and a new unit of length was introduced which became known as the metre. The metre gained adoption in continental Europe during the mid nineteenth century, particularly in scientific usage, and was officially established as an international measurement unit by the Metre Convention of 1875.

Medical classification is the process of transforming descriptions of medical diagnoses and procedures into universal medical code numbers. SNOMED Clinical Terms (SNOMED CT) is intended to provide a set of concepts and relationships that offers a common reference point for comparison and aggregation of data about the health care process. SNOMED-CT is designed to be managed by computer.
Some “R” questions

§ 2.4 Increase data re-use (through clarifying licences)

- License the data to permit the widest reuse possible
- Specify a data embargo, if this is needed
- How long will the data remain reusable?
- Describe data quality assurance processes

Re-use over time
Licensing research data and software

EUDAT licensing wizard helps you pick licences for data & software

You should also license Open Access data, or waive rights.

[Horizon 2020 Open Access guidelines point to:]

http://ufal.github.io/public-license-selector/
Overwhelmed?

A DMP is also a communication instrument!
RDM stakeholders

- Commercial partners
- Researchers
- Front office
- Back office
- Publishers
- Data Availability Policy
- Research funders
- Institution
- RDM policy
- Facilities
- €
- $£
- Information and awareness
- Training
- Storage and archiving
- GREECE
- MARINA ANGELOU
EUDAT B2 data service suite

https://www.eudat.eu/
OpenAIRE support materials

https://www.openaire.eu/opendatapilot
https://www.openaire.eu/support

• Briefing papers, factsheets, webinars, workshops, FAQs

• Information on:
  • Open Research Data Policy
  • Creating a data management plan
  • Selecting a data repository
  • Personal data
Related webinars

**Introductory RDM webinar**, Tony Ross-Hellauer & Marjan Grootveld, 30 May 2017:
- Reasons to manage data
- How to manage and share data (+ how to respond to concerns about sharing)
- OpenAIRE services
- Slides: [https://www.slideshare.net/OpenAIRE_eu/20170530open-research-data-in-horizon-2020](https://www.slideshare.net/OpenAIRE_eu/20170530open-research-data-in-horizon-2020)

**Open Research Data in H2020 and Zenodo**, Marjan Grootveld & Krzysztof Nowak, 26 October 2016:
- Sustainable file formats differ across domains and repositories
- Funders embrace the FAIR data principles – implications for Data Management Planning?

**How to write a DMP**, Sarah Jones & Marjan Grootveld, 7/14 July 2016:
- What is a Data Management Plan and why to write it?
- Example DMPs in different domains, with lots of links!
- Guidance, e.g. storing =/= archiving; how to find a repository; file-naming conventions

**FAIR data in Trustworthy Data Repositories**, Peter Doorn & Ingrid Dillo, 12/13 December 2016:
- Proposal for scoring datasets on Findability, Accessibility and Interoperability = Reusability levels
- Inspired by the Data Seal of Approval criteria for Trustworthy Data Repositories
PRACTICAL EXERCISE
Data organisation – 10 minutes

- Read the case description
- Design a data organisation for this project:
  1. Folder structure
  2. Naming convention
- Discuss it with your neighbour
- You have just 10 minutes - don’t loose yourself in the project details
Just checking…

• Who of you works in this way?
• Who does consider to start working in this way?
• Who thought about access rights or authorisation to (parts of) the structure?
• Who thought about informed consent forms?
• Who made a folder for information (i.e. metadata) about 600 interviews?
**Meaningful file names**

Below are tips on meaningful and consistent file names. Read more in 'Choosing a file name'.

- Make sure to use consistent file names. When you use a date in the file name, choose a notation (for instance, YYYYMMDD or yymmd).  
- Do not use strange characters like @*%^()<> in the file name.  
- Use traceable file names, such as Project_Instrument_locatieYYYYMMDD.ext.  
- Make sure to only use each file once in the folder structure. If you store a file in more than one place, several versions of the same file can unwillingly be created.  
- See also version management.

It is good practice to note the file naming and its meaning in a readme.txt.

Even if a researcher is well underway with his project consistent file naming is still an option by using a bulk file rename utility. It is important, however, to check if this bulk renamer delivers on its promises.
What makes a good DMP?

Has the researcher taken time to reflect on what to do?

- The reviewer wants to be reassured that due consideration has been given to data and that the approach seems reasonable.
  - Focus on the data – don’t describe how you’ll deliver your publications.
  - A publication that describes the data $\neq$ depositing the data.

- Is the plan appropriate?
  - Adopt relevant standards and practices that are in line with norms for your field.
  - Show that you know about support services e.g. university storage, subject repositories…

- Does the plan show proper engagement with the issues?
  - Be specific.
  - Justify decisions and any restrictions.
  - Plan when you evaluate and update your DMP.
So…

• Data management is all in a day’s work.
• Planning is more important than the plan, yet
  • Start early with an explicit plan
  • Keep it up to date
  • Involve the other stakeholders
Questions?

- www.openaire.eu
- @openaire_eu
- facebook.com/groups/openaire
- linkedin.com/groups/OpenAIRE-3893548

marjan.grootveld@dans.knaw.nl