



Marjan Grootveld, DANS

@MarjanGrootveld  
@openaire\_eu

# What you need to know to prepare a Data Management Plan



EKT and OpenAIRE Greek National Event,  
June 15 2017



# 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



**DataverseNL Dataverse Network**

Wednesday May 9, between 20.00 and 21.00 CET, the service will be offline, because of maintenance.

Utrecht University Dataverse | Erasmus University Rotterdam Dataverse | Avans Hogeschool Dataverse | Windesheim Hogeschool Windesheim Dataverse

Search this dataverse...   [Advanced Search](#)

**Dataverses (270)**  
**Datasets (441)**  
**Files (1,504)**

**Dataverse Category**  
 Organization or Institution (85)  
 Research Group (41)  
 Researcher (14)  
 Research Project (9)

**Publication Date**

1 to 10 of 711 Results

The self and others in the experience of pride [Dataset]  
 May 1, 2017 - Department of Social Psychology Dataverse  
 Oech, Y. van; Zeelenberg, M.; Breugelmans, S.M., 2017, "The self and others in the experience of pride [Dataset]", doi:10.4119/6L005C, DataverseNL Dataverse, V1

Dataset for: The self and others in the experience of pride

Biologically driven vegetation mosaics in grazing ecosystems  
 Apr 28, 2017 - Groningen Institute for Evolutionary Life Sciences Dataverse  
 Meentemeyer, R.A.; van der Knaap, J.; Smit, C., 2017, "Biologically driven vegetation mosaics in grazing ecosystems", V1

Biologically driven vegetation mosaics in grazing ecosystems: the battle between bioturbation and succession model resulting in figures 2 and 3, designed and written by Johan van de Ko...

DataverseNL for short- and mid-term data storage

**DANS** Data Archiving and Networked Services

**EASY**

Get exposure and credit for your data: write a data paper for the new peer reviewed, online-only open access Research Data Journal

For more info: [brill.com/rdj](http://brill.com/rdj)




EASY offers sustainable archiving of research data and access to thousands of datasets.

Search...  [Search help](#)

[Advanced search](#) [Browse](#)

**DEPOSIT YOUR DATA**

Instructions in English or Nederlands (PDF).

[Instruction summary](#)

EASY: certified long-term Electronic Archiving System for self-deposit; > 36k data sets, ~70% Open

**NARCIS**

NARCIS: Gateway to scholarly information in the Netherlands

The gateway to scholarly information in the Netherlands

[Submit Content to NARCIS](#)

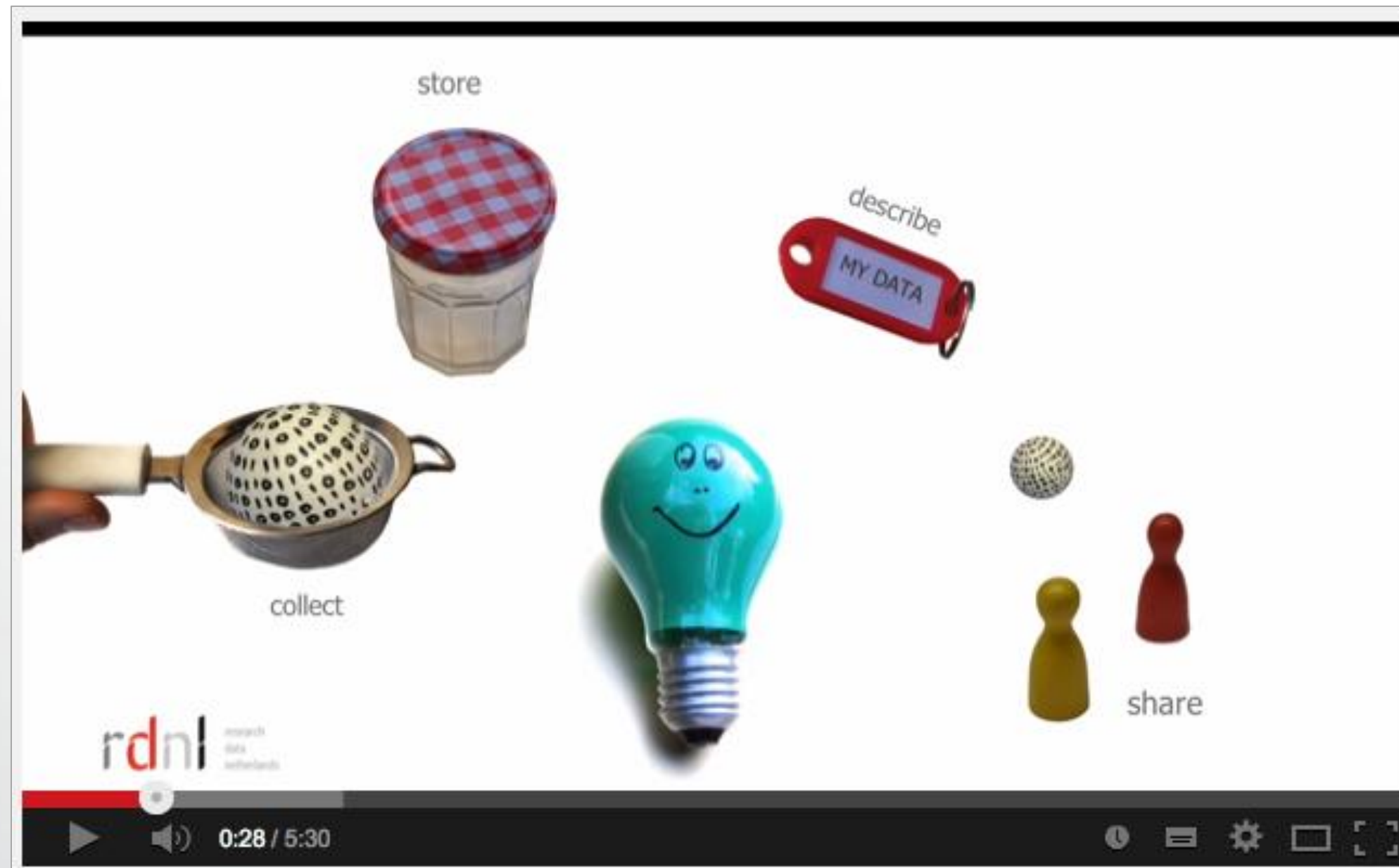
Search...

1,331,807	165,834	65,261	54,332	2,954
PUBLICATIONS	DATA SETS	RESEARCH	PEOPLE	ORGANISATIONS

Benefit from our knowledge on research data management by our **training sessions, consultancy** and information material.

 <b>TRAINING</b> DANS supports researchers (indirectly) in data management by providing training sessions. »	 <b>CONSULTANCY</b> DANS assists in developing data management policy and obtaining certification. »	 <b>INFORMATION MATERIAL</b> Watch the video 'Why share data' or download other information material. »
--	--	---

# 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.



# RESEARCH DATA – OPEN BY DEFAULT



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”

FAIR Data Management





# The EC Open Research Data policy

## Key sources of information

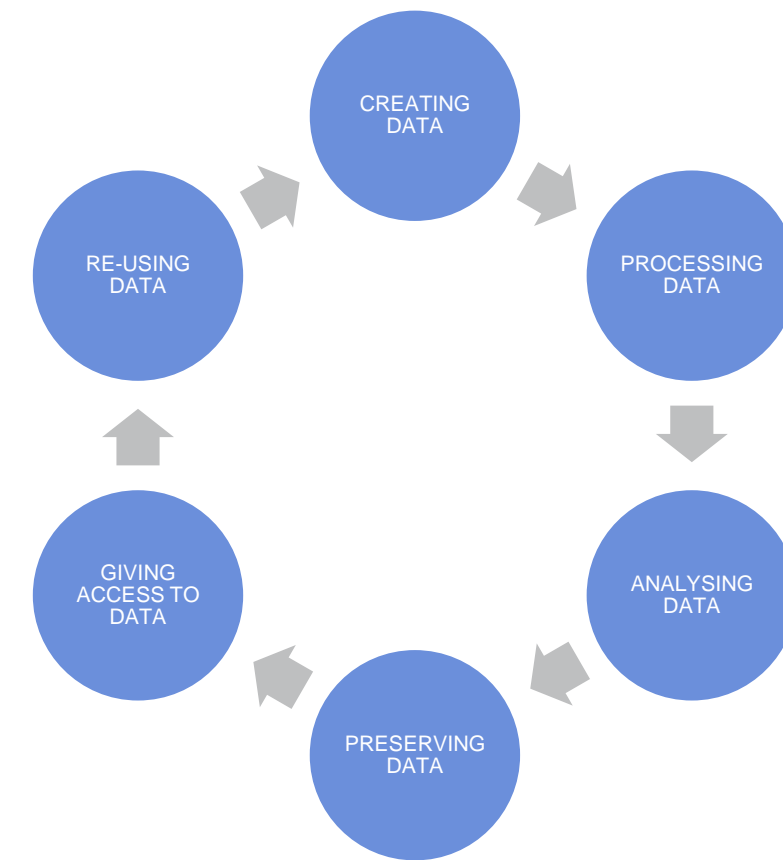
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- Guidelines on FAIR Data Management in Horizon 2020  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- Annotated model grant agreement, clause 29.3  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/amga/h2020-amga\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf)
- Infographic summarising key policy points  
[http://ec.europa.eu/research/press/2016/pdf/opendata-infographic\\_072016.pdf](http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf)
- Open Access and Data Management [http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm)

# DATA MANAGEMENT PLAN

# Data Management Plans

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>

**Create a new plan**

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.  
If you aren't responding to specific requirements from a funder or an institution, [select here to write a generic DMP](#) based on the most common themes.

**If applying for funding, select your research funder.**  
Otherwise leave blank.

European Commission (Horizon 2020) [Not applicable/not listed.](#)

**To see institutional questions and/or guidance, select your organisation.**  
You may leave blank or select a different organisation to your own.

University of Glasgow [Not applicable/not listed.](#)

**Tick to select any other sources of guidance you wish to see.**

- DCC guidance
- EUDAT
- School of Humanities
- Computing

[Create plan](#)

Choose your funder to get their specific template

Choose any additional optional guidance





# PLANNING DATA MANAGEMENT

# Research data



**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...

# 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
- 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?



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

Zenodo: <http://www.zenodo.org>

Re3data.org: <http://www.re3data.org>



# Storing or archiving? Both!

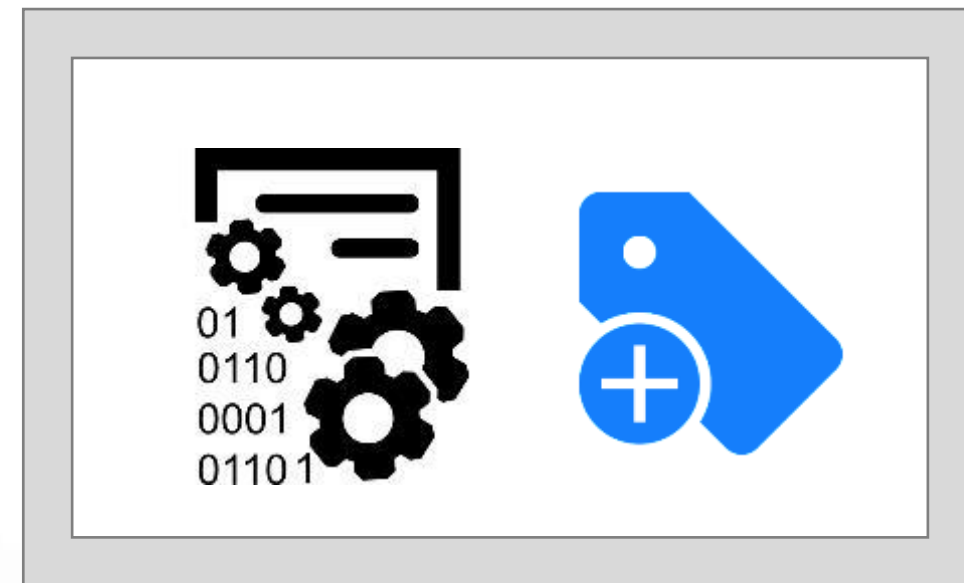
Storing and backing up files while research is active



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

April 25, 2016

Dataset

Open Access

# 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; Morales, 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.

Conference: EMBC 2015, IEEE 37th International Conference in Medicine and Biology Society, August 2015.

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

a) Robot Joints: Robot joints generated to develop the simulation, in radians (j1-j7 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.

Preview

**Publication date:**

April 25, 2016

**DOI:**

DOI 10.5281/zenodo.50439

**Keyword(s):**

algorithm generation

reconstruction algorithm

Kinematic reconstruction

**Published in:**

2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC): pp. 1190 - 1193.

**Grants:**

European Commission:

- AIDE - Adaptive Multimodal Interfaces to Assist Disabled People in Daily Activities (645322)

**Meeting:**

2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2015), Milan, 25-29 August 2015

**Related identifiers:**

Part of:

10.1109/EMBC.2015.7318579

**Communities:**

European Commission Funded Research (OpenAIRE)

**License (for files):**

Creative Commons Zero - CC0 1.0

# Sharing data: what is meant?

With collaborators while research is active



Data are mutable



(Open) data sharing



Data  
Repository



Data are stable, searchable,  
citable, clearly licensed

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**.



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

A440, which has a frequency of 440 Hz, is the musical mid nineteenth century, particularly in scientific usage, and was above middle C and serves as a **general tuning stand** officially established as an international measurement unit by the Metre Convention of 1875.

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 $\flat$  of A440), *baroque* for some special church music (*Chorton pitch*) at 466 Hz (A $\sharp$  of A440), and *classical* pitch at 430 Hz.

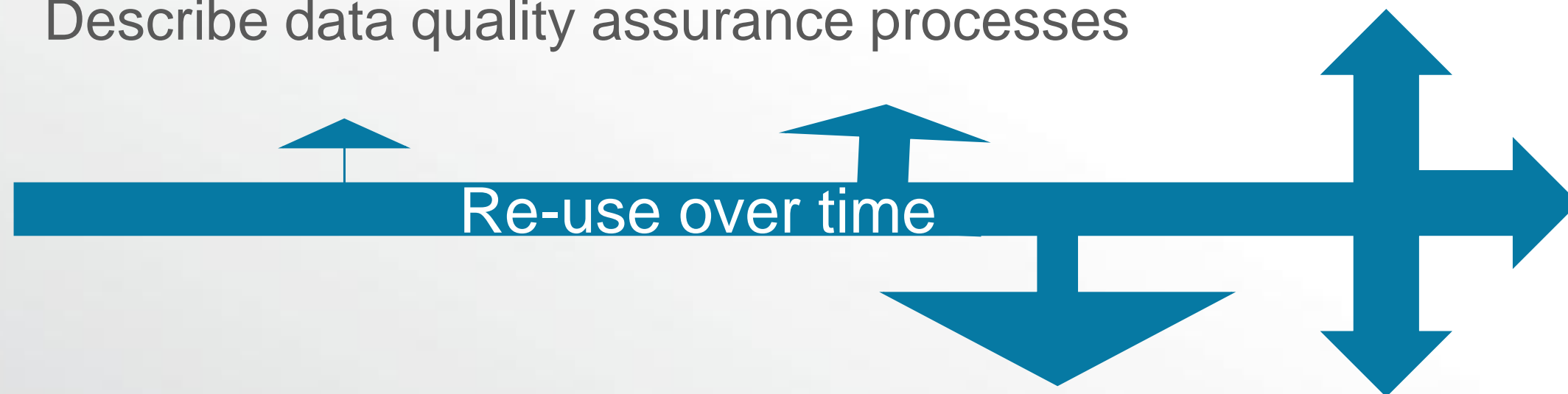
Medical classification is the process of transforming descriptions of medical diagnoses and procedures into universal medical code numbers. SNOMED Clinical Terms (SNOMED CT) is intended to provide a set of concepts and relationships that offers a **common reference point for comparison and aggregation 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





# Licensing research data and software

EUDAT licensing wizard helps you pick licences for data & software



Do you own copyright and similar rights in your dataset and all its constitutive parts?

Do you allow others to make commercial use of you data?

**Creative Commons Attribution (CC-BY)**  
This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

**Public Domain Dedication (CC Zero)**  
CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

Horizon 2020 Open Access guidelines point to:



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

# Overwhelmed?

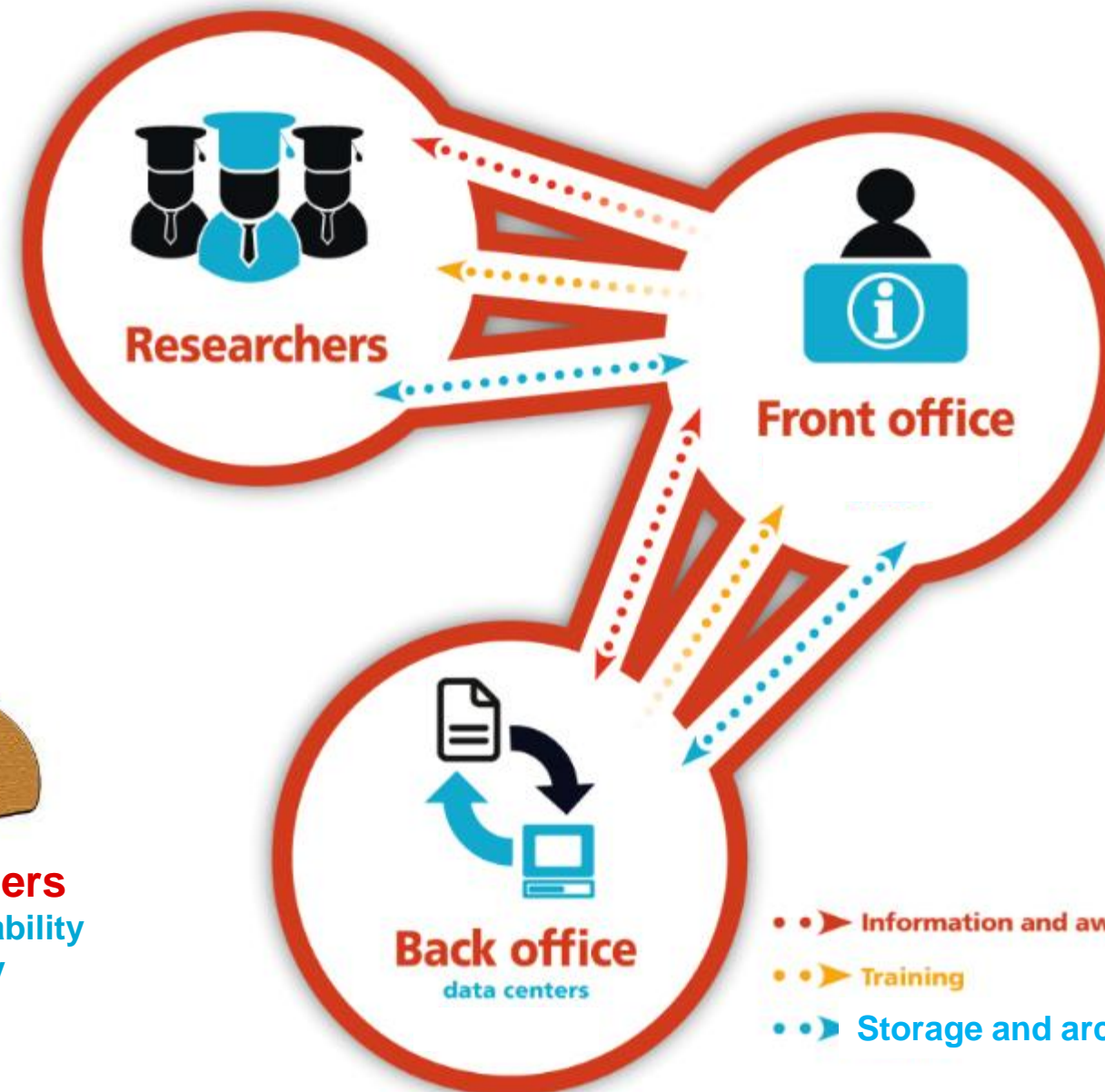


A DMP is also a communication instrument!

# RDM stakeholders



**Commercial partners**



**Institution**  
RDM policy  
Facilities



GREECE  
MARINA ANGELAKI

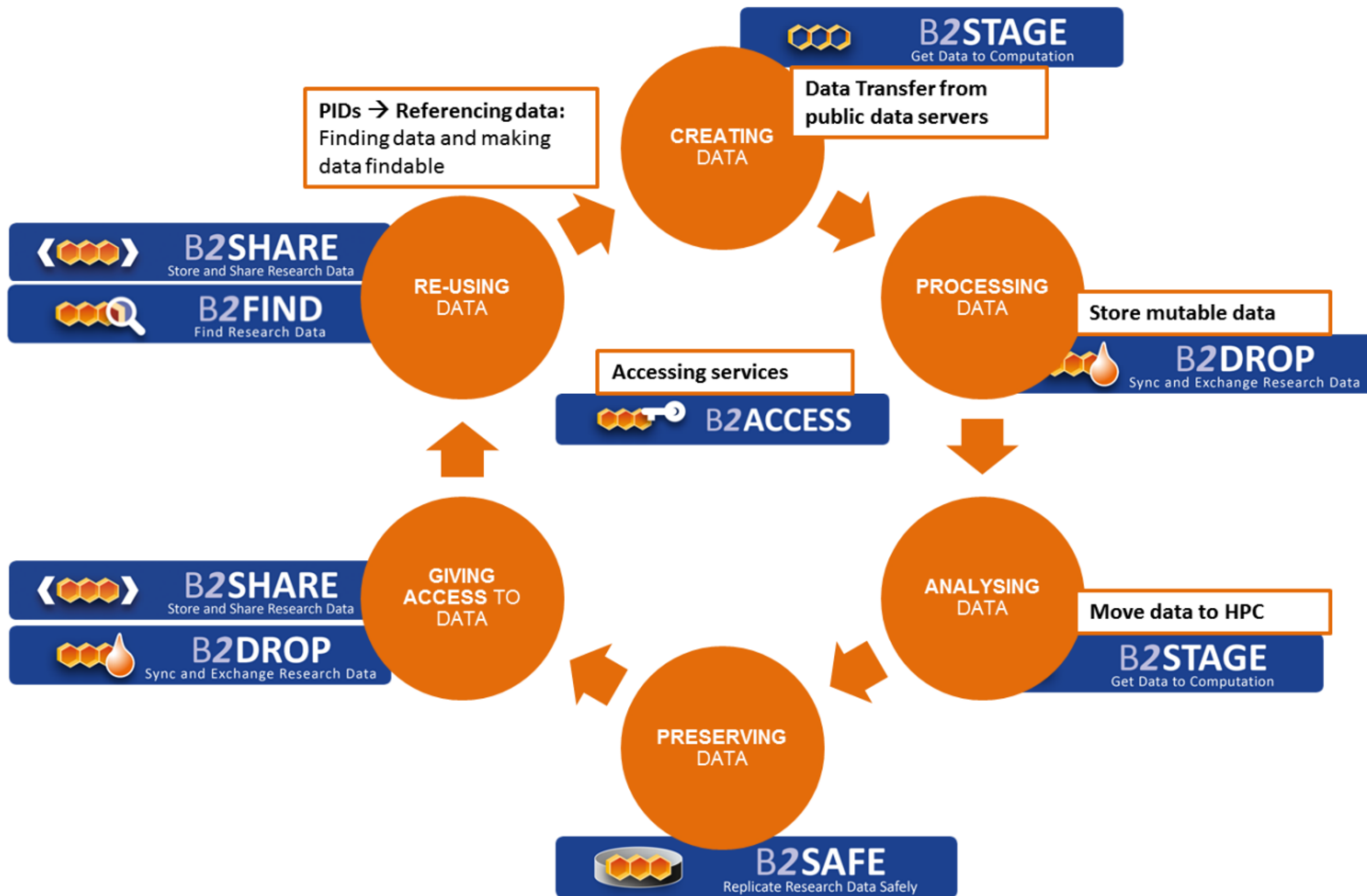


**Publishers**  
Data Availability  
Policy



**Research funders**

# EUDAT B2 data service suite





# 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?
- Slides: [www.slideshare.net/OpenAIRE\\_eu/openaire-webinar-on-open-research-data-in-h2020-oaw2016](http://www.slideshare.net/OpenAIRE_eu/openaire-webinar-on-open-research-data-in-h2020-oaw2016)
- Q&A document: <https://www.openaire.eu/public-documents?id=843&task=document.viewdoc>

How to write a DMP, Sarah Jones & Marjan Grootveld, 7/14 July 2016:

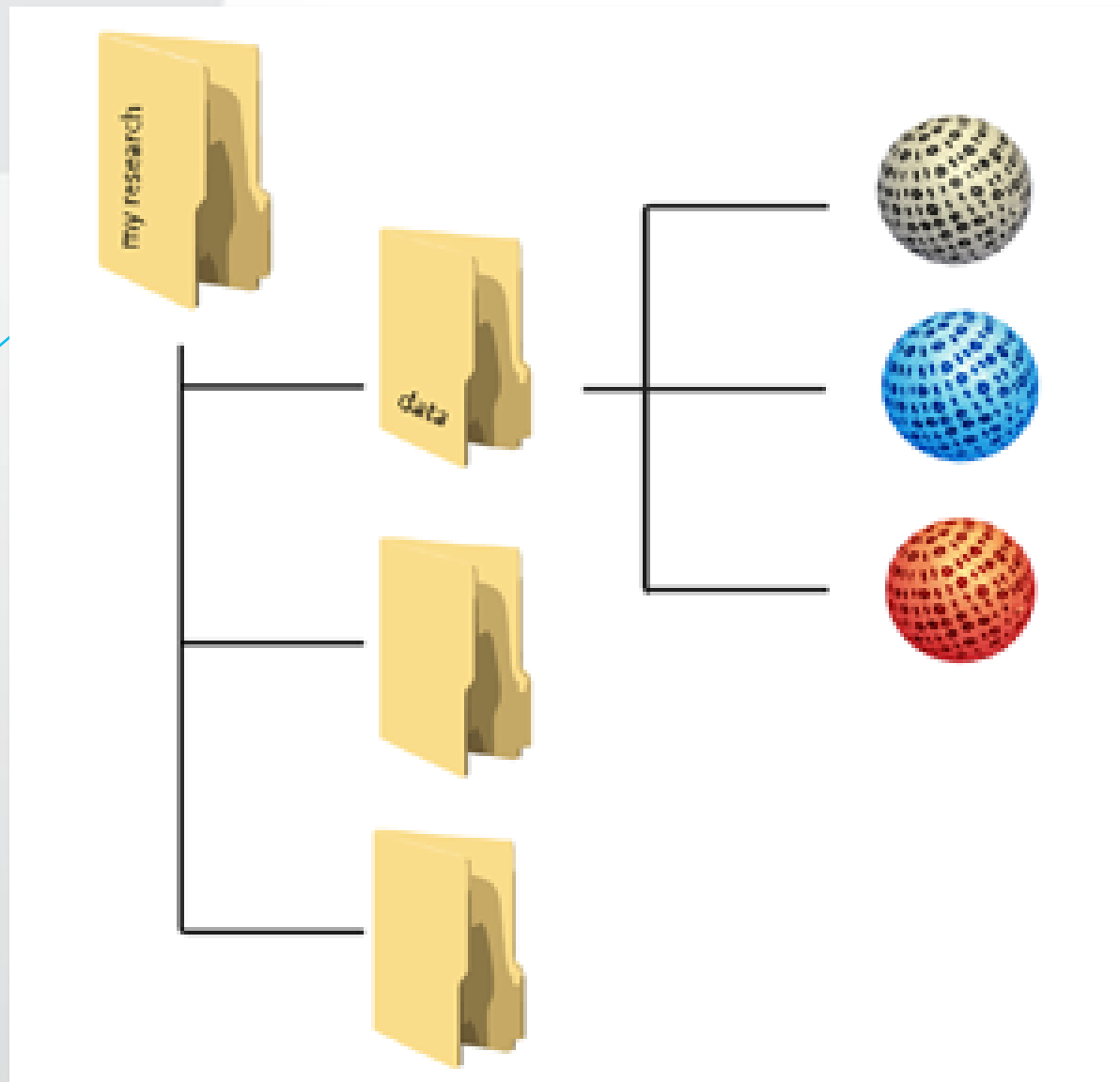
- What is a Data Management Plan and why to write it?
- Example DMPs in different domains, with lots of links!
- Guidance, e.g. storing  $\neq$  archiving; how to find a repository; file-naming conventions

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
- Slides: <http://www.slideshare.net/EUDAT/fair-data-in-trustworthy-data-repositories-webinar-1213-december-2016>

# 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](#)'.<sup>(2)</sup>

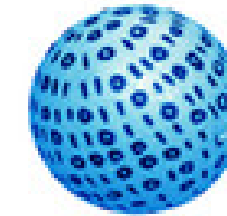
- ❑ Make sure to use consistent file names. When you use a date in the file name, choose a notation (for instance, YYYYMMDD or yymmdd).
- ❑ Do not use strange characters like ?!\*%{[<> in the file name.
- ❑ Use traceable file names, such as Project\_Instrument\_location\_YYYYMMDD.ext.
- ❑ Make sure to 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.

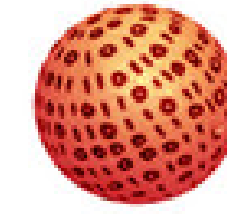
Even if a researcher is well underway with his project consistent file naming is still an option by using a [bulk file rename utility](#).<sup>(3)</sup> It is important, however, to check if this bulk renamer delivers on its promises.



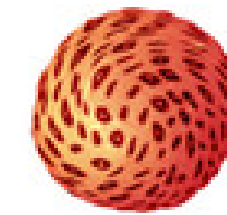
white\_data\_20140708.csv



blue\_data\_20140708.docx



red\_data\_20140708.R



red\_data\_20140708\_v02.R

*File naming and version management*



# 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](http://www.openaire.eu)

 @openaire\_eu

 facebook.com/groups/openaire

 linkedin.com/groups/OpenAIRE-3893548

 [marjan.grootveld@dans.knaw.nl](mailto:marjan.grootveld@dans.knaw.nl)