Marjan Grootveld

Data Archiving and Networked Services (DANS)

What you need to know to succeed with your   data management plan

OpenAIRE: uno strumento per la Scienza Aperta in Europa
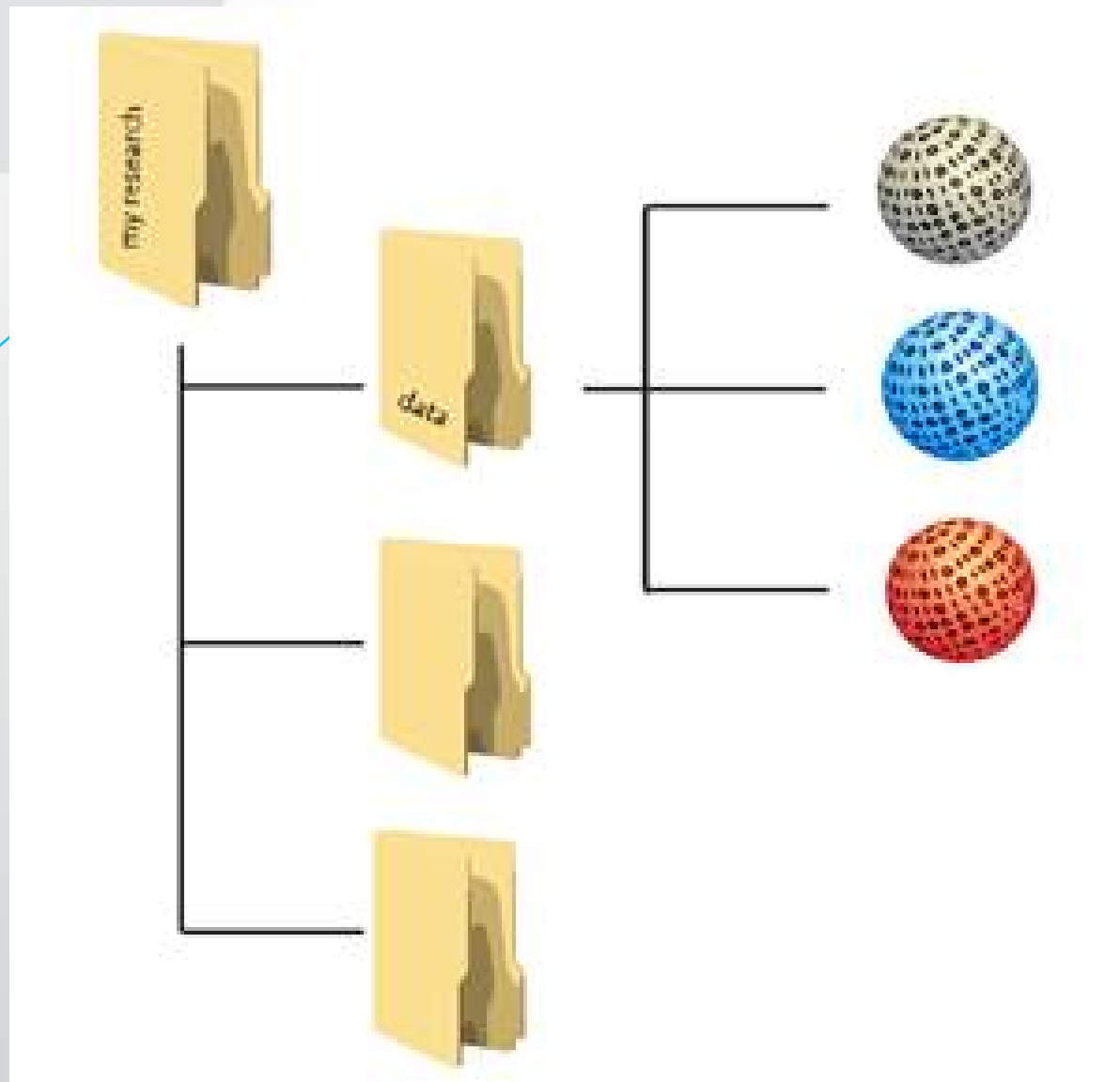**Rome, May 30 2016**

DATA MANAGEMENT PLAN (PART 2)

The what, why and how of data management planning

From the training *Essentials 4 Data Support* by Research Data Netherlands

Data organisation – 10 minutes

- **Read the case description**
- **Design a data organisation for this project:**
  1. Folder structure
  2. Naming convention
- **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 m_____) about 600 intervi_____**

Recall from the EC's DMP template: Briefly specify
- how data will be captured/created
- how it will be documented
- according to what standards
- who will be able to access it
- where it will be stored
- how it will be backed up, and
- where and how it will be shared and preserved long-term

Possible folder structure

- **Raw** mpeg4 files (N = 600) – **access limited** to PI and project team
  - File name includes unique ID of interviewee & ID of interviewer & "raw"

- **Processed** mpeg4 files, after anonymisation etc. (N depends on the content)
  - File name includes unique ID of interviewee & ID of interviewer & version number

- **Transcripts** txt files (N = 600)
  - File name: raw data file name & version number (i.e. version of the transcript)

- **Informed consent** PDF/A (N = 600) - **access limited** to PI and interviewers
  - Consent given for OA
  - Consent given for OA after embargo period
  - Consent given for RA
  - Consent as yet undecided
  - List of all interviewee IDs + contact information

- **Documentation**
  - Project plan
  - DMP
  - Structured interview questions & other "interviewer alignment" documentation
  - Subfolder Metadata about the interviewees (N = 600) - **access limited**!
  - Communication with interviewees – general
  - Progress document listing the currently released data (using interviewee IDs)

Meaningful file names

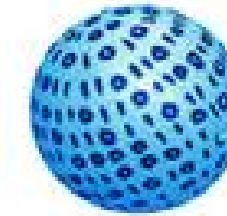Below are tips on meaningful and consistent file names. Read more in *'Choosing a file name'*. [2]

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

white_data_20140708.csv

blue_data_20140708.docx

red_data_20140708.R

red_data_20140708_v02.R

File naming and version management

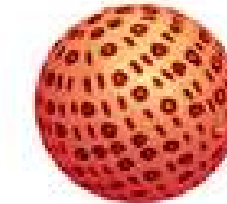It is good practice to note the file naming and its meaning in a readme.txt.

Even if a researcher is well underway with his project consistent file naming is still an option by using a *bulk file rename utility*. [3] It is important, however, to check if this bulk renamer delivers on its promises.
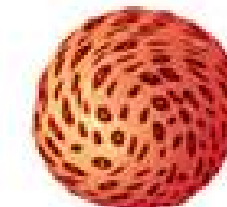
ing-data/

Roles and responsibilities

Commercial partners

Publishers
Data Availability Policy

Researchers

Front office

Back office
data centers

• • ▶ Information and awareness
• • ▶ Training
• • ▶ Storage
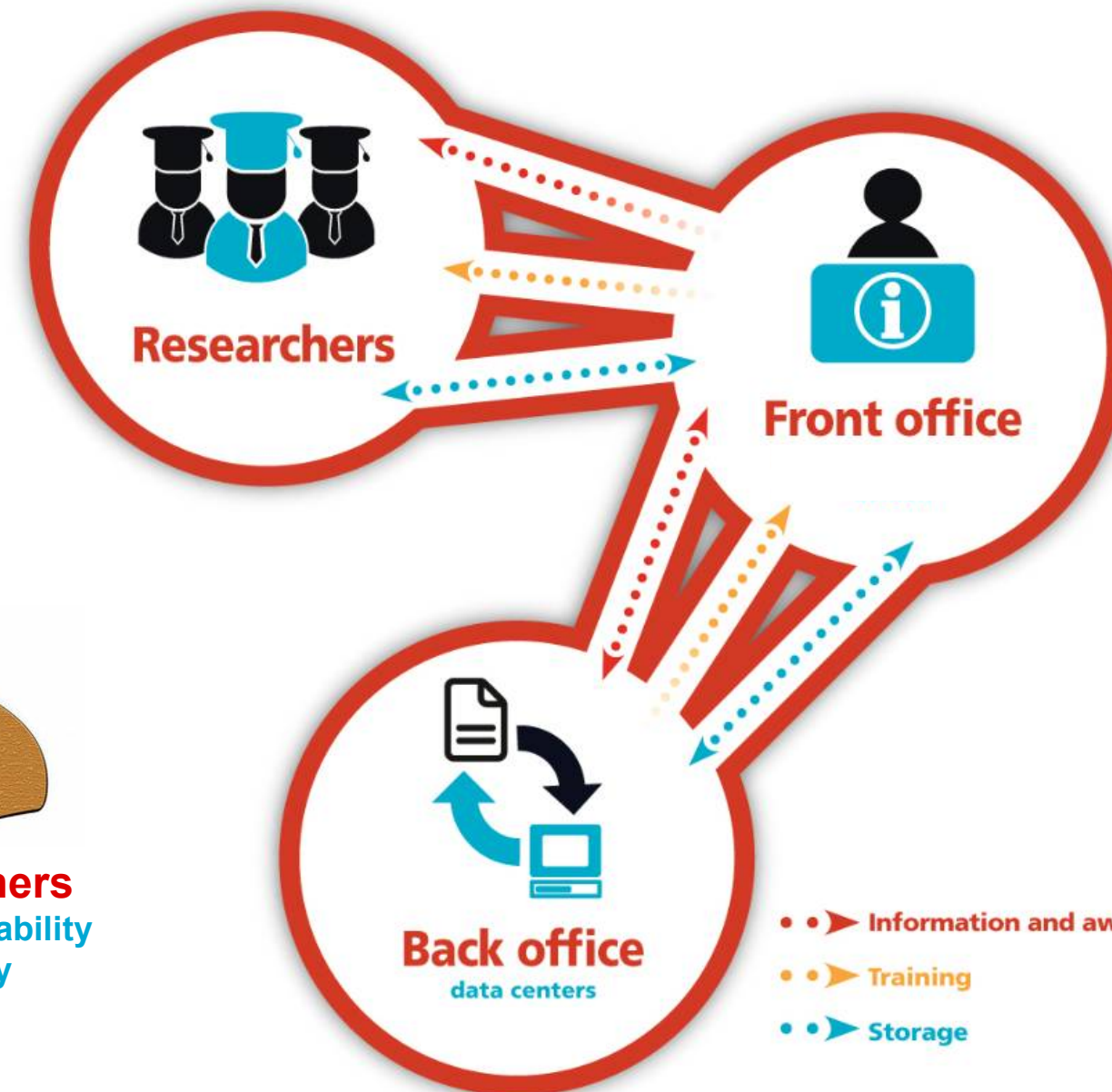
Institution
RDM policy
Facilities
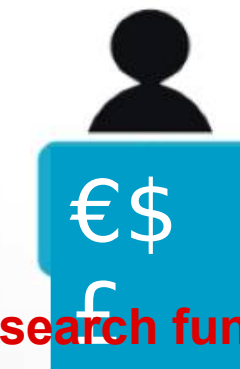
Research funders

A DMP is about 'keeping' data

- **Storing data < > archiving data**
- **Archived data < > findable data**
- **Findable < > accessible**
- **Accessible < > understandable**
- **Understandable < > usable**

is not safe
- **Figshare is not a Trustworthy Digital Repository**
- **a persistent identifier is essential but no guarantee for usability**
- **Data in a proprietary format are not sustainable**

… and about sharing data

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**

13

Good luck with your data management!

Any questions?

👆 www.openaire.eu

🐦 @openaire_eu

f facebook.com/groups/openaire

in linkedin.com/groups/OpenAIRE-3893548

✉ **marjan.grootveld@dans.knaw.nl**