

Marjan Grootveld and Elly Dijk

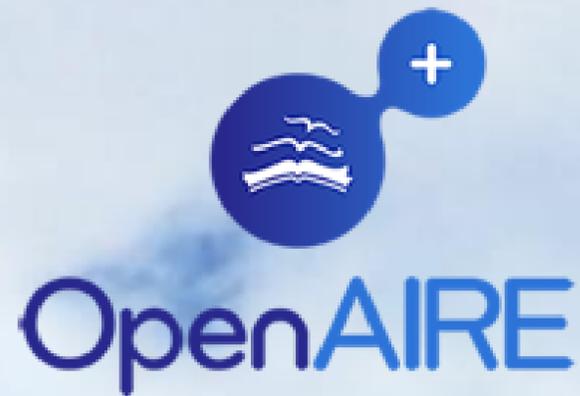
Data Archiving and Networked Services - DANS
The Netherlands



@openaire_eu
@MarjanGrootveld
@DANSKNAW

Open Research Data and Data Management

Atvērtie pētniecības dati un datu pārvaldība



**Workshop, 27 October 2017, Riga,
Latvia**



WHAT IS DANS?

What is DANS?



Mission:
promote and
provide
permanent
access to digital
research
resources

Institute of
Dutch Academy
and Research
Funding
Organisation
(KNAW & NWO)
since 2005

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 (95)
 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
 Osch, Y. van; Zeelenberg, M.; Breugelmans, S.M., 2017, "The self and others in the experience of pride [Dataset]", hdl:10411/KLOX8C, DataverseNL Dataverse, V1
 Dataset for: The self and others in the experience of pride

Biotically driven vegetation mosaics in grazing ecosystems
 Apr 28, 2017 - Groningen Institute for Evolutionary Life Sciences Dataverse
 Howison, R.A.; Cliff, H.; van de Koppel, J.; Smit, C., 2017, "Biotically driven vegetation mosaics in grazing ecosystems", V1
 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

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

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 PUBLICATIONS | 165,834 DATA SETS | 65,261 RESEARCH | 54,332 PEOPLE | 2,954 ORGANISATIONS

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

TRAINING
 DANS supports researchers (indirectly) in data management by providing training sessions. »»

CONSULTANCY
 DANS assists in developing data management policy and obtaining certification. »»

INFORMATION MATERIAL
 Watch the video 'Why share data' or download other information material. »»

WHAT IS OPENAIRE?



OpenAIRE 2020



THE NETHERLANDS

JUST DE LEEUWE



THE NETHERLANDS

ELLY DIJK

- **Open Access Infrastructure for Research in Europe**
- **Funded by Horizon2020 to develop and maintain the infrastructure to support OA policy of the EU**
- **Supports H2020 OA mandates**
 - 100% OA on scientific publications
 - Open Research Data Pilot
- **In every country a National Open Access Desk**
- **DANS is regional coordinator Research Data Management**



LATVIA

GITA ROZENBERGA

- **2018 – 2020: OpenAIRE Advance**

OpenAIRE Network: www.openaire.eu



[BLOG](#)

[NEWSLETTER](#)

[SIGN IN](#) | [REGISTER](#)



[PARTICIPATE](#)

[SEARCH](#)

[MONITOR](#)

[SUPPORT](#)

[OPEN ACCESS](#)



Search in 22,036,305 publications 571,373 datasets from 2,711 repositories and OA journals

G7 Science Ministers
endorse Open Science



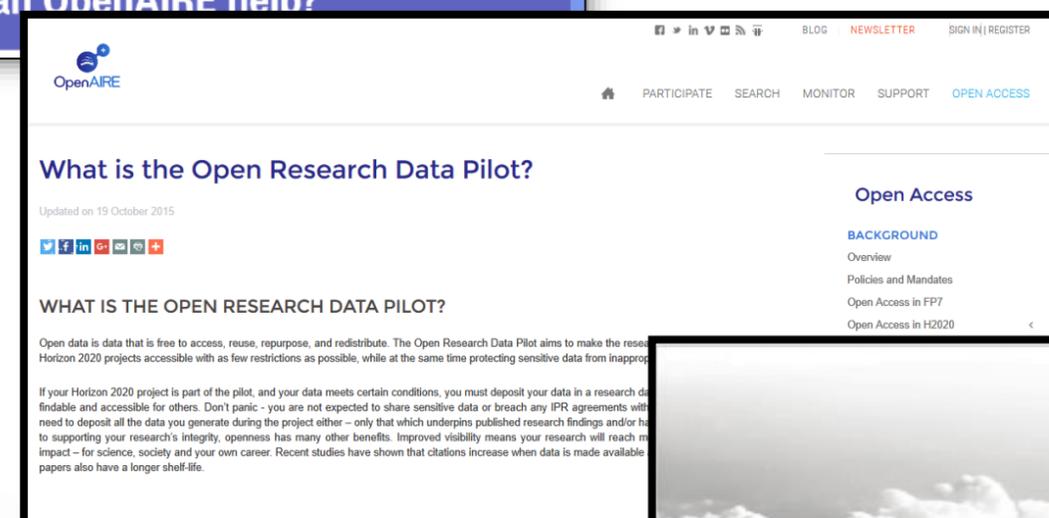
OpenAIRE support materials

- Briefing papers, factsheets, webinars, workshops, FAQs
- Information on:
 - Open Research Data Pilot
 - Creating a data management plan
 - Selecting a data repository
 - Personal data

<https://www.openaire.eu/what-is-the-open-research-data-pilot>

<https://www.openaire.eu/support>

<http://hdl.handle.net/10084/111436>



Programme workshop

- **12:00 – 12.30** Introduction Open Data
- **12.30 – 12.45** **Hands-on:** data organisation
- **12.45 – 13.15** RDM and FAIR
- **13.15 – 14.00** **Hands-on:** data management planning

- **14.00 – 14.30** Break

- **14.30 – 14.45** Storing and preserving data
- **14.45 – 15.15** **Hands-on:** finding a data repository
- **15:15 – 15:30** Wrap-up

INTRODUCTION OPEN DATA

Why are research data a topic now?

- Data deluge: more and more complex data
- Growing recognition of the value of data
- Open research data: data sharing becomes important
- Cases of data fraud



nature

International weekly journal of science

News

Report finds massive fraud at Dutch universities

Investigation claims dozens of social-psychology papers contain faked data.



Science Inside

Breaking news and analysis

Report: Dutch 'Loose Data' Found in Dozens of Studies (UPDATED)

by Gretchen Vogel on 31 October 2011 5:35 PM | 34 Comments

the guardian

Public sector can combat fraud with data sharing

Outsourcing is not the only thing to blame for procurement fraud, says **Graham Kemp**, and the public sector needs to view data less as a security risk but knowledge to be shared

The New York Times

Psychology Seen as a Red Flag for Psychology

By **DICT CAREY**
Published: November 2, 2011

A well-known psychologist in the Netherlands whose work has been published in professional journals falsified data and made up experiments, an investigation has committed. Experts say the case exemplifies flaws in the way psychology is done in a field, psychology's only a fragile respectability.



September 2011: Diederik Stapel, Social Psychology

November 2011: Don Poldermans, Cardiovascular Medicine

June 2012: Dirk Smeesters, Psychology

October 2012: Mart Bax, Cultural Anthropology

September 2015: Science: only 39% psychological studies (2008) could be reproduced



How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data

Report: Diederik Stapel
Psychologe gesteht
Fälschungen

Increased awareness of need for Data Policies

- Former EU Vice-President Neelie Kroes: “**Data is the oil for science**” (Riding the Wave report, 2010).
- **Science Europe**: Research Data Working Group recommends members to develop open data policies.
- **EU research funding programme Horizon 2020**: DMP and open data access standard for H2020 projects. Opt-out is possible.
- **European Open Science Cloud** will offer 1.7 million European researchers and 70 million professionals in science and technology a virtual environment with open and seamless services for **storage, management, analysis and re-use of research data**, across borders and scientific disciplines by federating existing scientific data infrastructures, today scattered across disciplines and Member States.

EUROPEAN OPEN SCIENCE CLOUD

BRINGING TOGETHER CURRENT AND FUTURE DATA INFRASTRUCTURES



EUROPEAN DATA INFRASTRUCTURE

UNLOCKING THE VALUE OF BIG DATA; DIGITAL BY DEFAULT



An introduction to the basics of research data
<https://www.youtube.com/watch?v=q2aiDJzJPuw>

Examples of research data

- Text or Word documents, spreadsheets
- Statistics
- Results of experiments
- Measurements
- Observations resulting from fieldwork
- Survey results
- Interview recordings: audiotapes, videotape
- Images
- Laboratory notebooks
- Database contents
- Models, algorithms, scripts



Why is digital preservation of data important?

- **Precondition for sharing and re-use**
- **Makes research more transparent**
- **Checks on claims made in publications**
- **Promotes replication research**
- **However, data re-use for comparative studies is much more important**



Why are researchers **not** willing to share?

- **Those data are mine!**
- **Discredit my findings**
- **Still analyzing the data**
- **I cannot trust the data produced somewhere else**





Why a researcher wants to share data?

- **When a researcher wants the data to be open available: social accountability;**
- **It increases the visibility of research results;**
- **It is citable thanks to digital object identifiers;**
- **It encourages the reuse of the data for new research questions and for verification purposes.**

The data sharing advantage in astrophysics

S. B. F. Dorch, T. M. Drachen, O. Ellegaard

(Submitted on 8 Nov 2015)

We present here evidence for the existence of a citation advantage within astrophysics for papers that link to data. Using simple measures based on publication data from NASA Astrophysics Data System we find a citation advantage for papers with links to data receiving on the average significantly more citations per paper than papers without links to data. Furthermore, using INSPEC and Web of Science databases we investigate whether either papers of an experimental or theoretical nature display different citation behavior.

Comments: 4 pages, 2 figures, Conference proceedings of Focus Meeting 3 on Scholarly Publication in Astronomy, IAU GA 2015, Honolulu

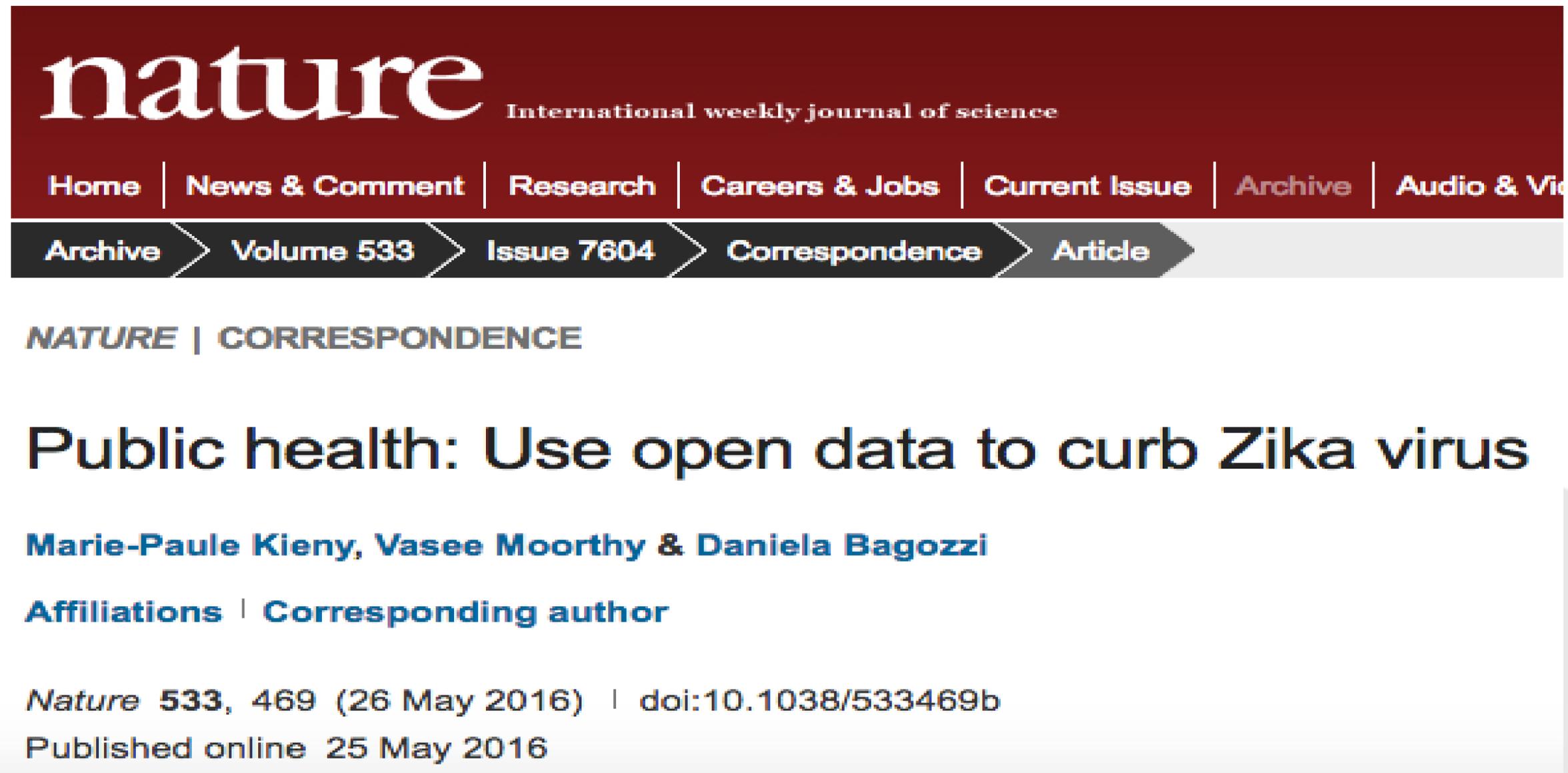
Subjects: Instrumentation and Methods for Astrophysics (astro-ph.IM); Digital Libraries (cs.DL)

Cite as: [arXiv:1511.02512](https://arxiv.org/abs/1511.02512) [astro-ph.IM]

(or [arXiv:1511.02512v1](https://arxiv.org/abs/1511.02512v1) [astro-ph.IM] for this version)

Data sharing and society

- <https://www.nature.com/nature/journal/v533/n7604/full/533469b.html>



nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video

Archive > Volume 533 > Issue 7604 > Correspondence > Article

NATURE | CORRESPONDENCE

Public health: Use open data to curb Zika virus

Marie-Paule Kieny, Vasee Moorthy & Daniela Bagozzi

Affiliations | Corresponding author

Nature **533**, 469 (26 May 2016) | doi:10.1038/533469b
Published online 25 May 2016

Surveys about open data



- **How and why researchers share data (and why they don't) (2014)**

<https://hub.wiley.com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont?referrer=exchanges>

- **Towards open Research – Practices, experiences, barriers and opportunities (2016)**

https://figshare.com/articles/Survey_of_Wellcome_researchers_and_their_attitudes_to_open_research/4055448

- **Open Data Research – a researcher perspective (2017)**

http://www.elsevier.com/_data/assets/pdf_file/0004/281920/Open-data-report.pdf

- **Providing researchers with the skills and competencies they need to practise Open Science (2017)**

https://cdn1.euraxess.org/sites/default/files/policy_library/ec-rtd_os_skills_report_final_complete_2207_1.pdf

- **The State of Open Data (2017)**

<https://figshare.com/articles/5481187>



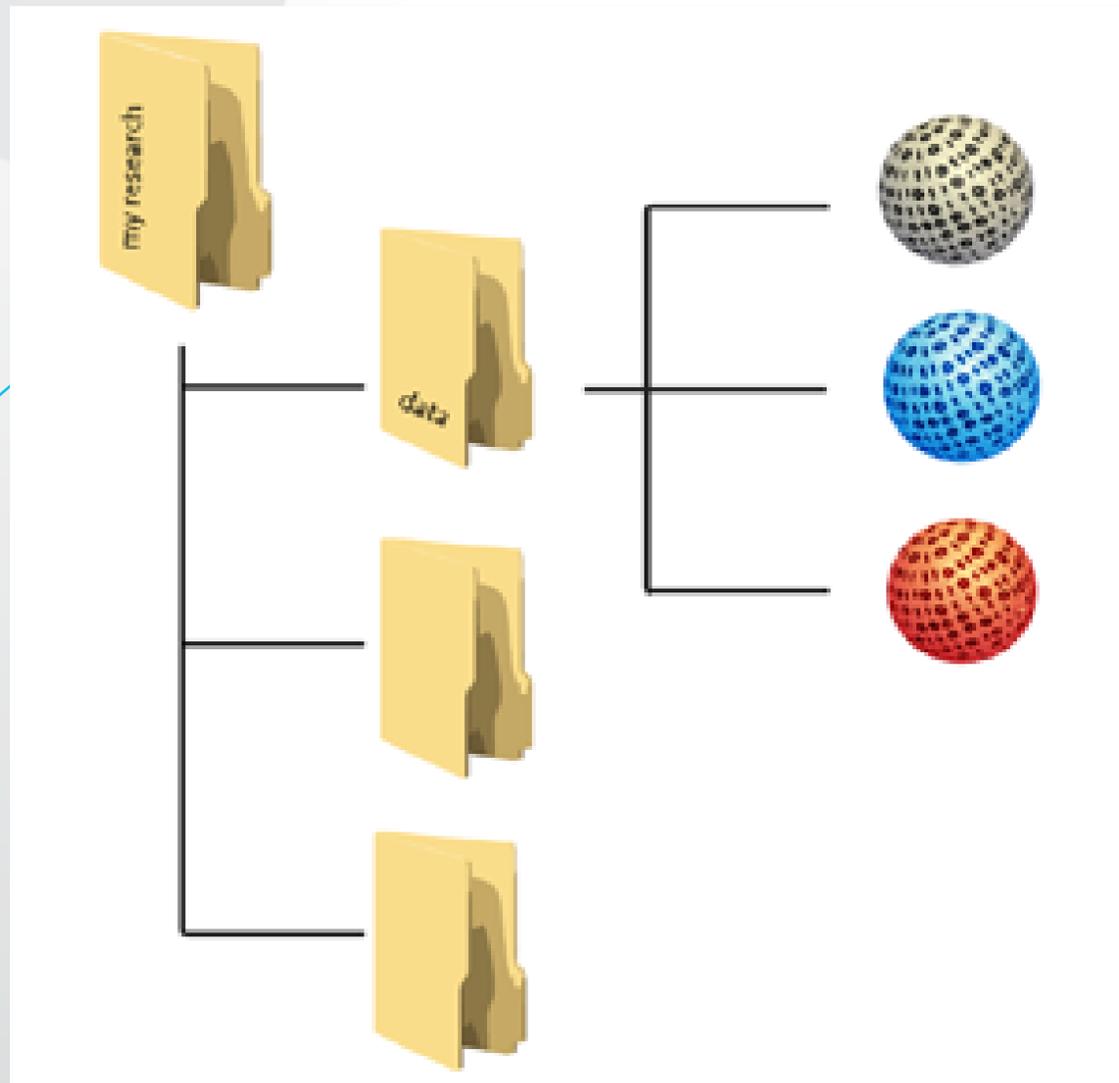
Providing researchers with the skills and competencies they need to practice Open Science

- Report by *EC Working Group on Education and Skills under Open Science*
- Survey: answered by 1,277 researchers in Europe (nearly 50% of PhD candidates).
- Open Science changes the research landscape. Research is conducted with a high degree of transparency, collegiality and research integrity.
- Training of the necessary skills and professional development of researchers;
- Three quarters of the researchers indicate that they did not participate in open access or open data training, but they would like to

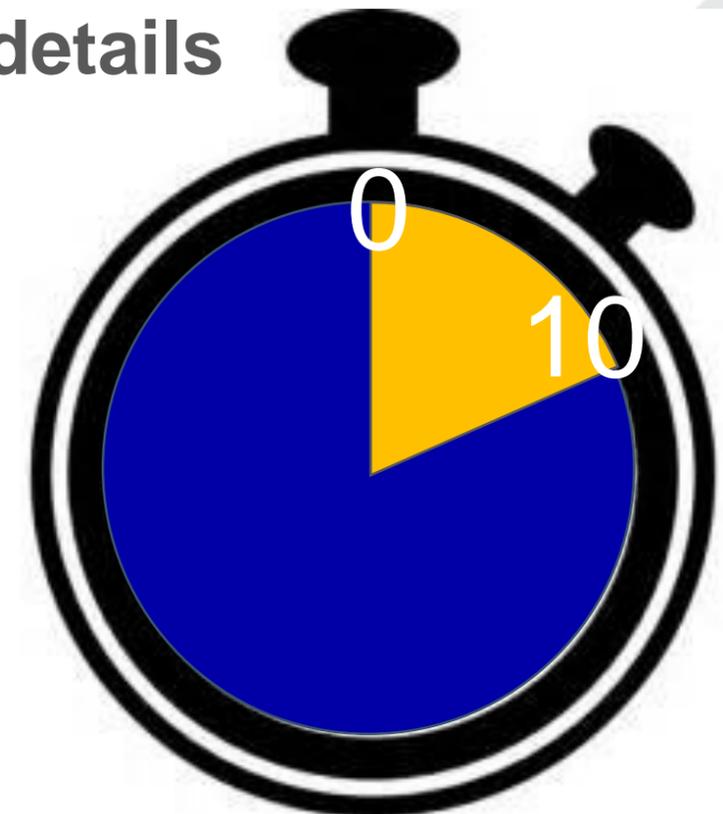
EXERCISE 1:

Data organisation

Data organisation – 10 minutes



- Read the case description
- Design a data organisation for this project:
 1. Folder structure
 2. Naming convention
- With your neighbour
- Don't drown yourself in the 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?**

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⁾

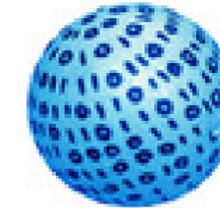
- ❑ Make sure to use consistent file names. When you use a date in the file name, choose a notation (for instance, YYYYMMDD or yymmdd).
- ❑ Do not use strange characters like ?\!@*%{[<> in the file name.
- ❑ Use traceable file names, such as Project_Instrument_location_YYYYMMDD.ext.
- ❑ Make sure to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

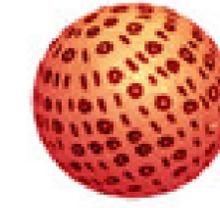
Even if a researcher is well underway with his project consistent file naming is still an option by using a [bulk file rename utility](#).⁽³⁾ It is important, however, to check if this bulk renamer delivers on its promises.



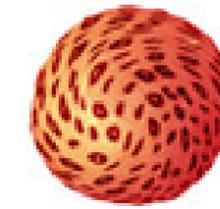
white_data_20140708.csv



blue_data_20140708.docx



red_data_20140708.R

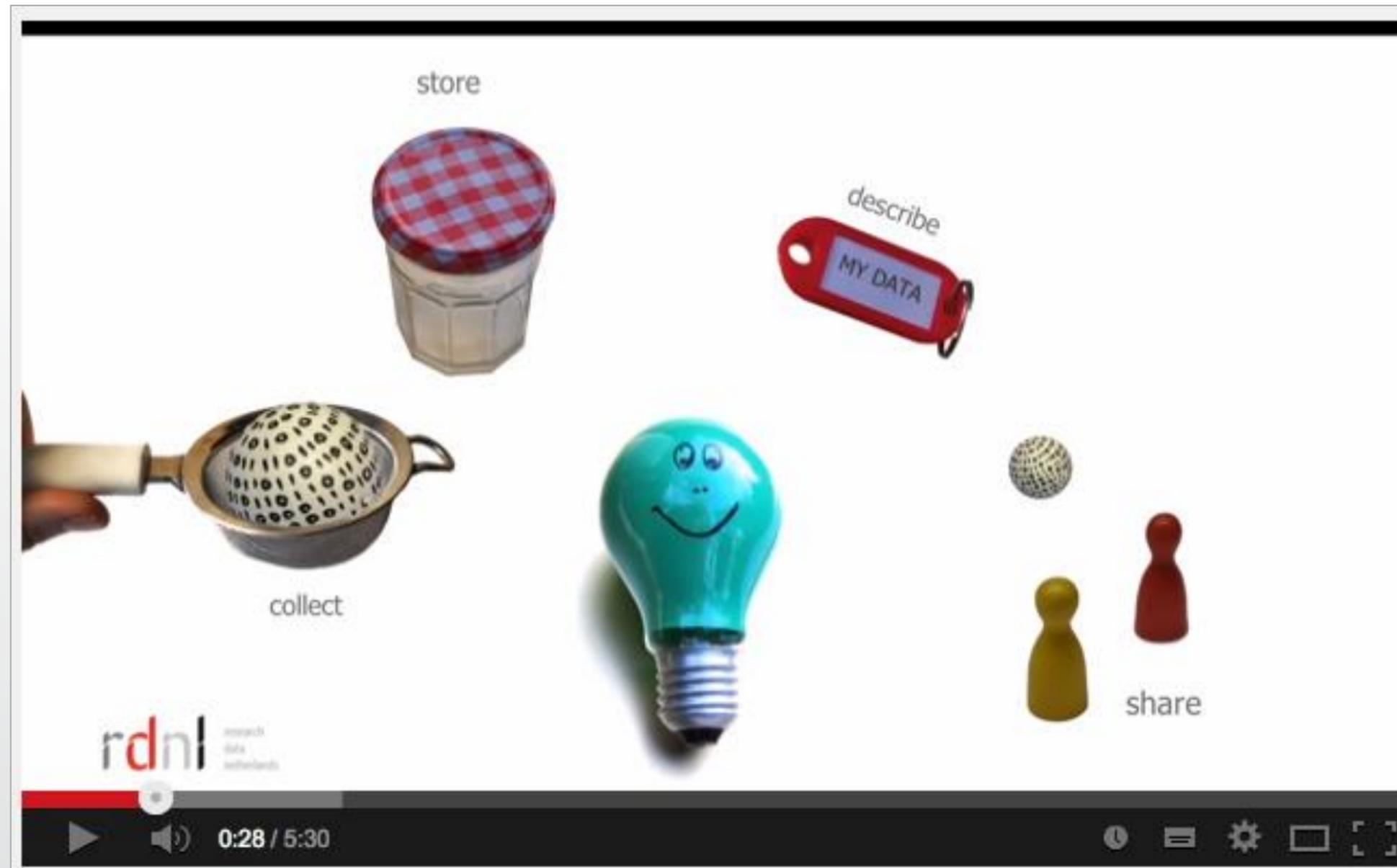


red_data_20140708_v02.R

File naming and version management

DMP, RDM, FAIR, ...

The what, why and how of data management planning

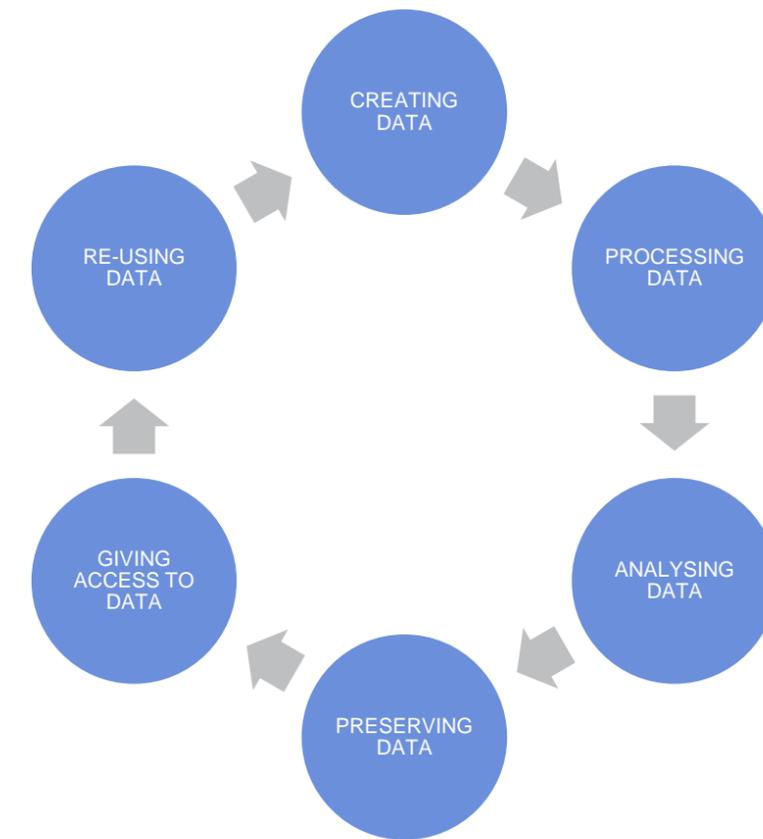


From the training *Essentials 4 Data Support* by Research Data Netherlands

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.

Open Research data

Open Research Data Pilot (2014-2016)

As of the Work Programme 2017: extended to cover all thematic areas of Horizon 2020 ('open by default')

- Projects may still opt-out at any stage (IPR, personal data protection, national security, other reasons)
- Mainly concerns data underlying publications (other data optional)
- Data Management Plan obligatory by M6 (not part of project evaluation)

Approach: as open as possible, as closed as necessary

Clarifying terminology...



In the past our policy mainly addressed the 'accessibility' part of FAIR.

- Started off with 'open access to research data'
- Moved towards open (research) data with the ORD pilot (which also covered further aspects)
- We are now seeing openness as one component of FAIR data and aim to address all of the FAIR aspects in Horizon 2020

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

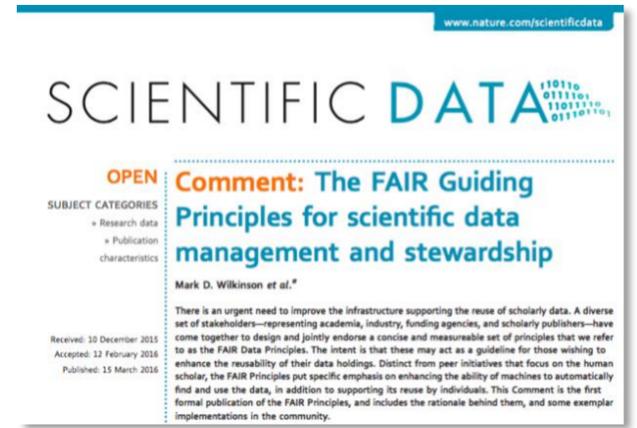


Research data



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 keywords to support searching
- Persistent and unique identifiers such as DOI
- Versioning of the datasets and clear version numbers

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



PUBLICATIONS AND DATA

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.**

CoreTrustSeal

Data Seal of Approval

ICSU World Data System

nestor seal

ISO 16363



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 (Ab of A440), *baroque* for some special church music (*Chorton pitch*) at 466 Hz (A# 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)

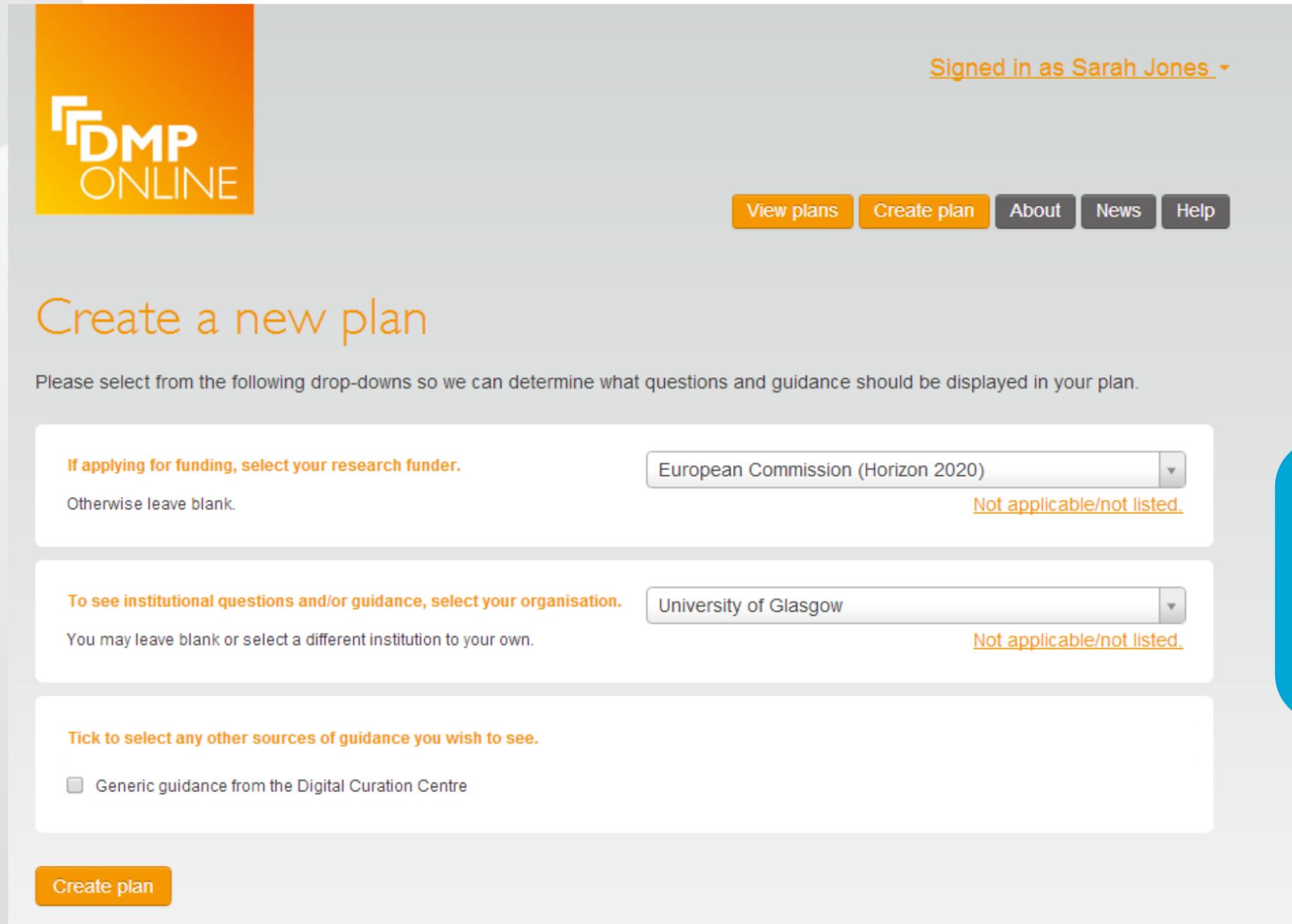
- License the data to permit the widest reuse possible
- Specify a data embargo, if this is needed
- How long will the data remain reusable?
- Describe data quality assurance processes



EXERCISE 2:

Exploring the Horizon2020 DMP template

Template in DMPonline



Signed in as Sarah Jones ▾

DMP ONLINE

[View plans](#) [Create plan](#) [About](#) [News](#) [Help](#)

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 applying for funding, select your research funder.
Otherwise leave blank. [Not applicable/not listed.](#)

European Commission (Horizon 2020) ▾

To see institutional questions and/or guidance, select your organisation.
You may leave blank or select a different institution to your own. [Not applicable/not listed.](#)

University of Glasgow ▾

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

Generic guidance from the Digital Curation Centre

[Create plan](#)

Select research funder (if any)

Select your organisation for additional questions and guidance (optional)

Select guidance (recommended)

H2020 DMP template §2.1

- Apply the six questions in section 2.1 to the Veterans case
- Discussion is allowed ;-)

2. FAIR data

2. 1. Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

What naming conventions do you follow?

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.



Your feedback



What does
it mean?



Our *suggestion*

- **Metadata provision:** we will deposit the data in repository XYZ, which supports the DDI metadata standard <add reference>. In the first year of the project we select the relevant metadata fields (D3.5).
- **Identifiability:** repository XYZ provides a URN:NBN to the dataset. This is a globally unique and persistent identifier, which can be used for citing the dataset.
- **Naming conventions:** see previous exercise
- **Keywords:** each data file gets keyword “ProjectName”. Furthermore we use existing vocabulary ABC <add reference>
- **Versioning:** we will maintain major versions of processed data and document the differences between versions. Naming convention see above.
- **Metadata standard:** Data Documentation Initiative (DDI), which is standard in social sciences

<http://rd-alliance.github.io/metadata-directory/standards/>

- **Things to do in the project team:**
 - Agree on the metadata fields
 - Agree on folder structure and file-naming convention
 - Agree on versioning convention: where, what, naming convention, who is responsible?
 - Start documenting!

Overwhelmed?

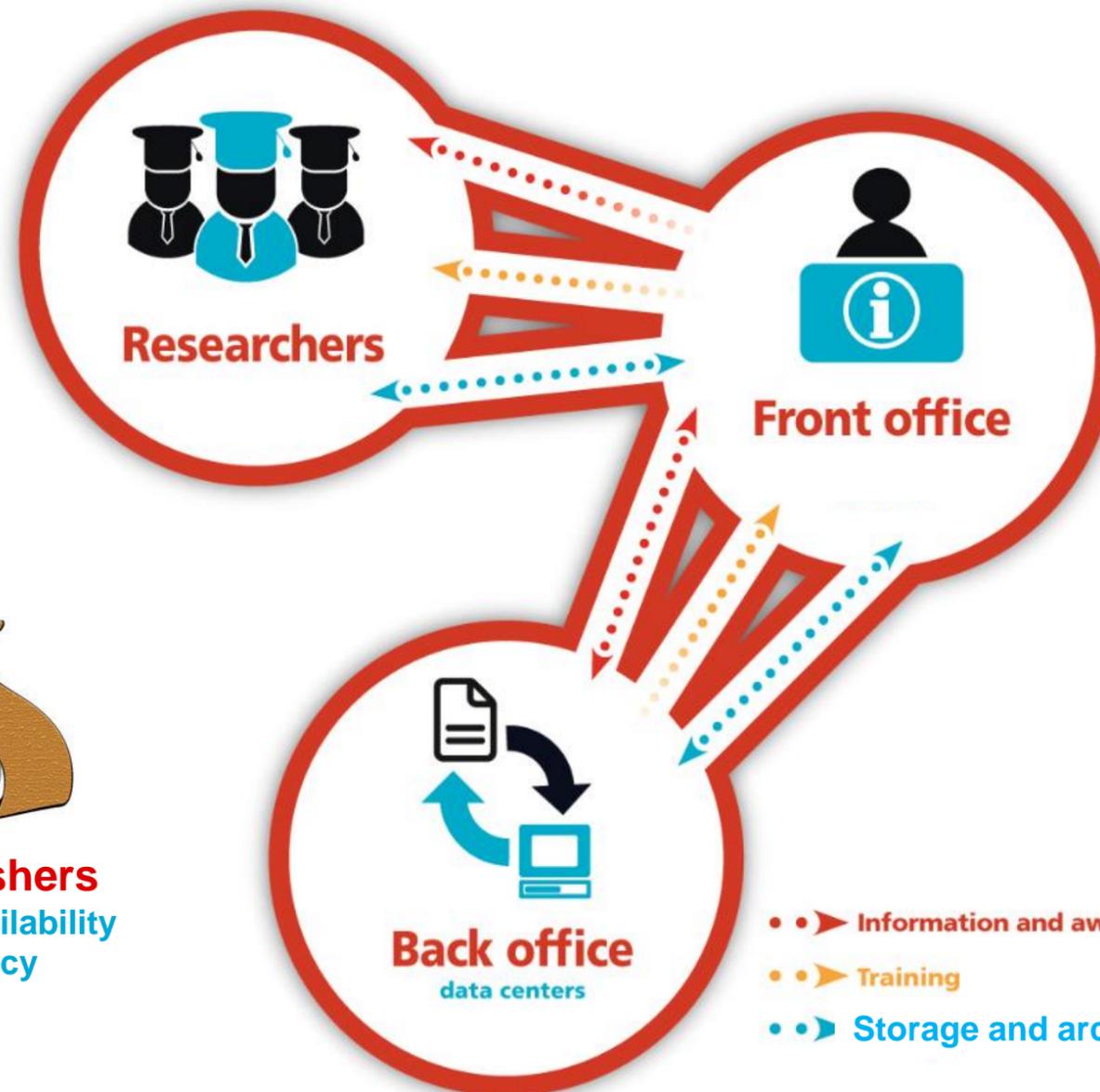


A DMP is also a communication instrument!

RDM stakeholders



Commercial partners



Institution
RDM policy
Facilities



LATVIA
GITA ROZENBERGA

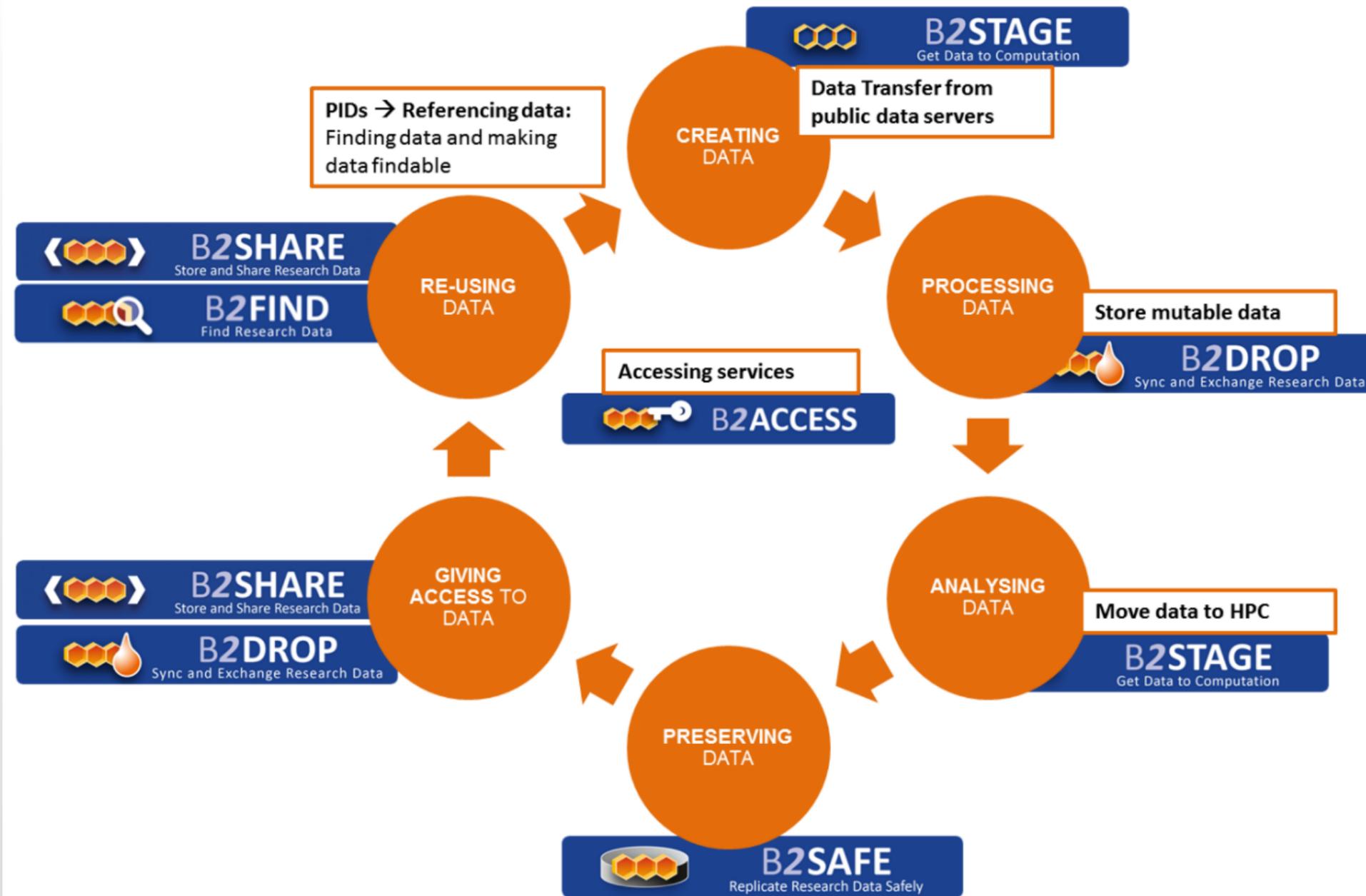


Publishers
Data Availability
Policy



Research funders

EUDAT B2 data service suite





H2020 open data pilot: to remember

Some important notions:

- Data should be made open when possible, restricted when necessary
- Metadata and other standards
- Arrangements with the identified repository
- Documentation about the software needed to access the data
- Licences to permit the widest reuse possible
- Potential value of long-term preservation

"Costs related to open access to research data in Horizon 2020 are eligible for reimbursement during the duration of the project under the conditions defined in ...":

- Only if budgeted in the proposal and granted;
- Only during the project



Recommended: Data Management Cost Guide

<https://www.uu.nl/en/files/datamanagementcostguidejune2017pdf>

BREAK

STORING AND PRESERVING DATA

Sharing and storing data **during** the research

With collaborators (also at other institutions) while research is active



Data are mutable:
easy to change
or delete

Storing and backing up files while research is active

Likely to be on a networked store or in a repository



The
Dataverse
Project



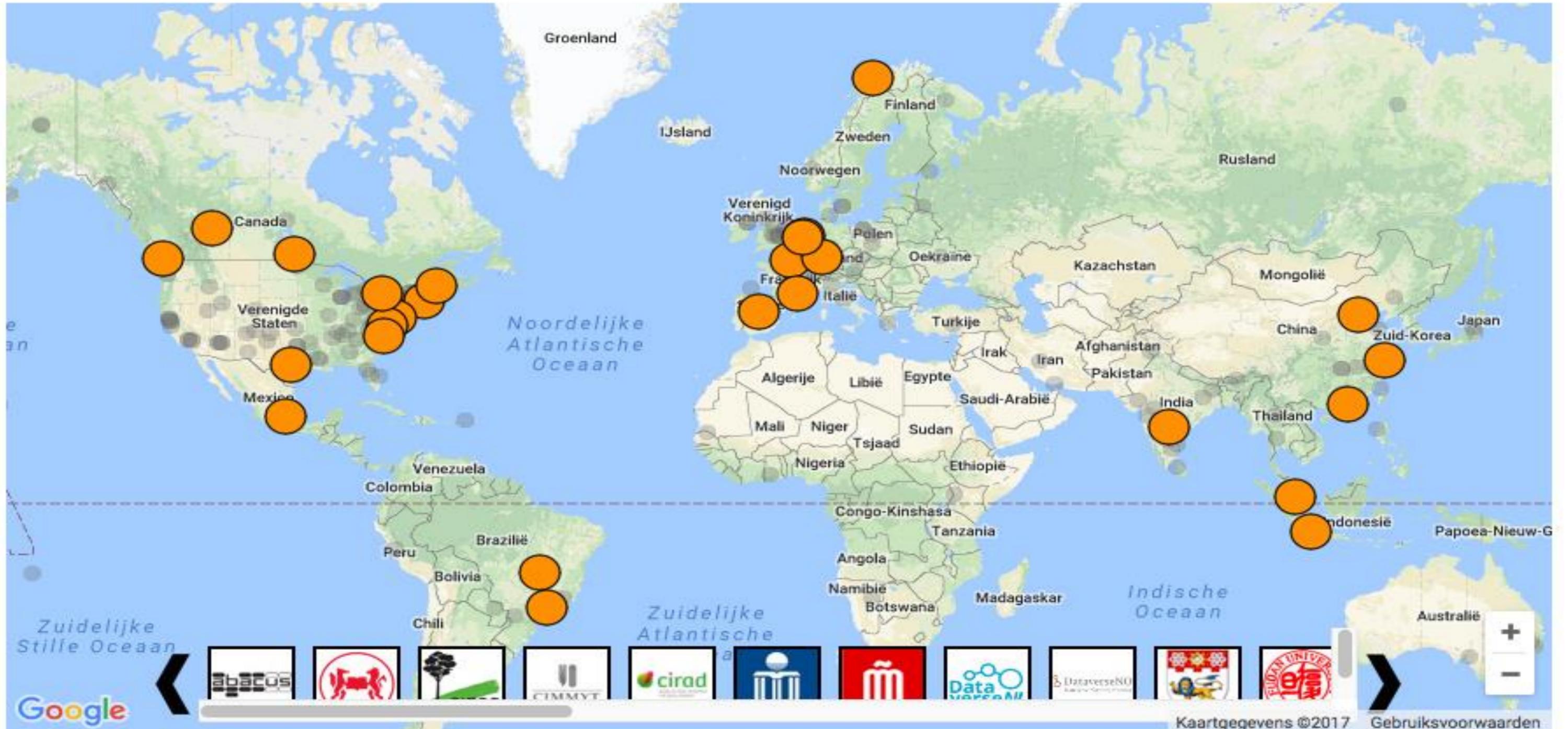
DATAVERSE REPOSITORIES - A WORLD VIEW

26 Installations

2,367 Dataverses

49,767 Datasets

2,769,967 Downloads



Stats generated: 22nd October 2017 16:07 EDT

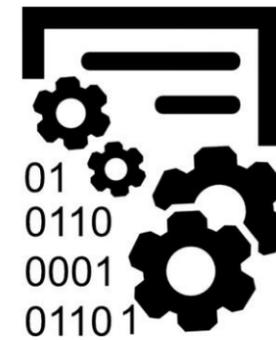
Storing data and sharing data after the research

(Open) data sharing



Data are stable, searchable, citable,
clearly licensed

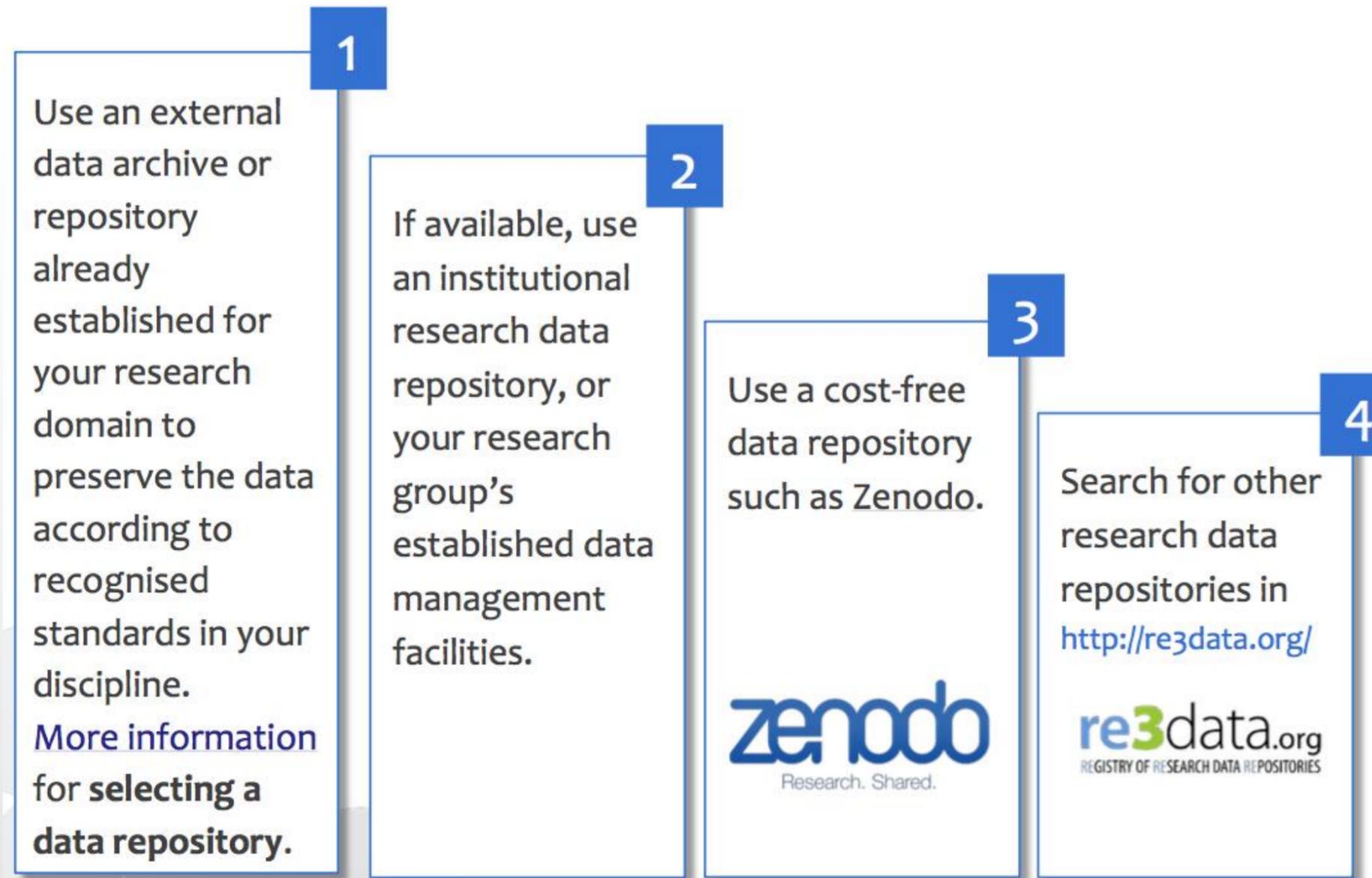
Archiving or preserving
data in the long-term



Likely to be deposited
in a digital repository

Safeguarded and
preserved

Where to find a repository?



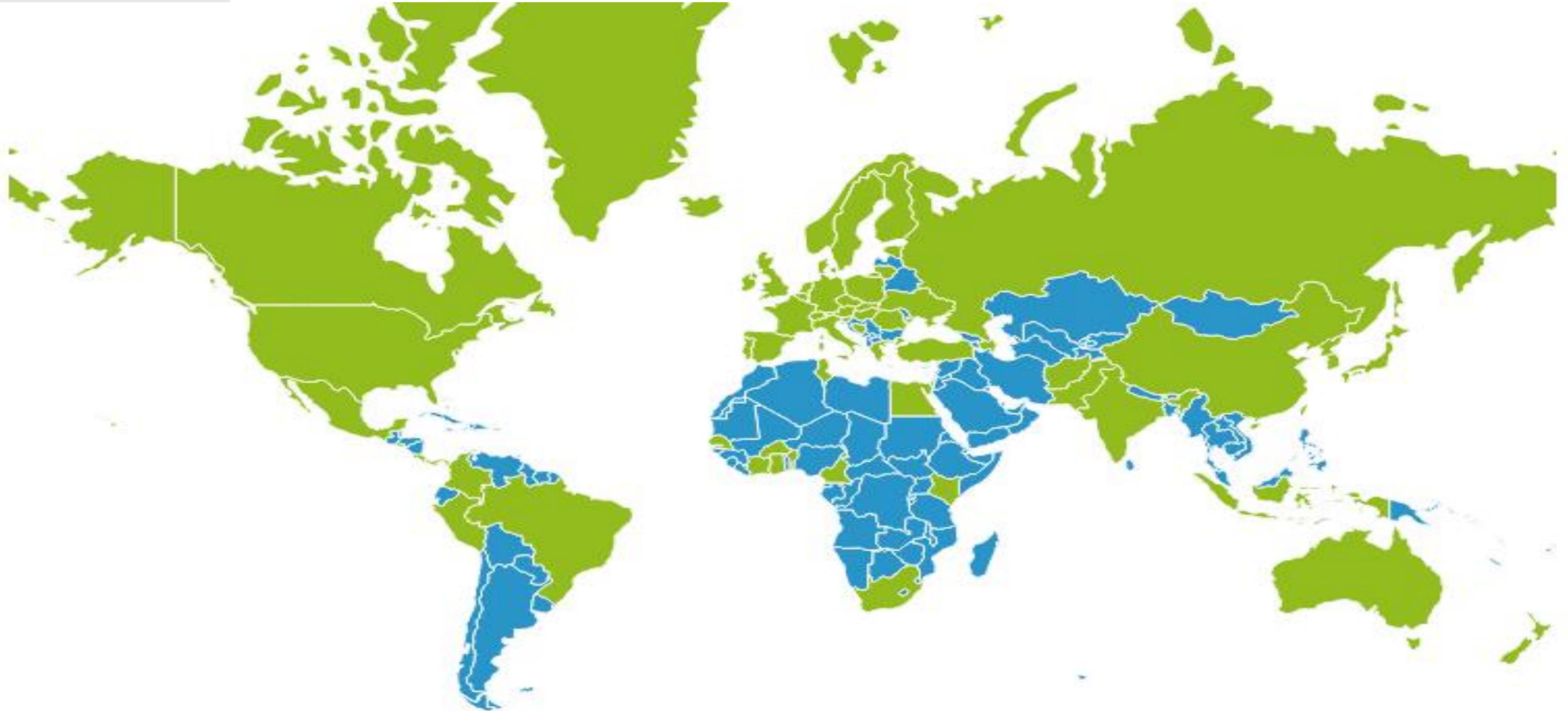
🔗 More information: <https://www.openaire.eu/opendatapilot-repository>

🔗 Zenodo: <http://www.zenodo.org>

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



re3data.org

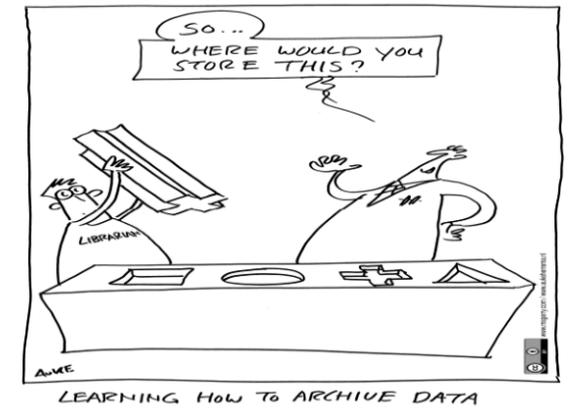


ICTWSS database

General Institutions Terms Standards

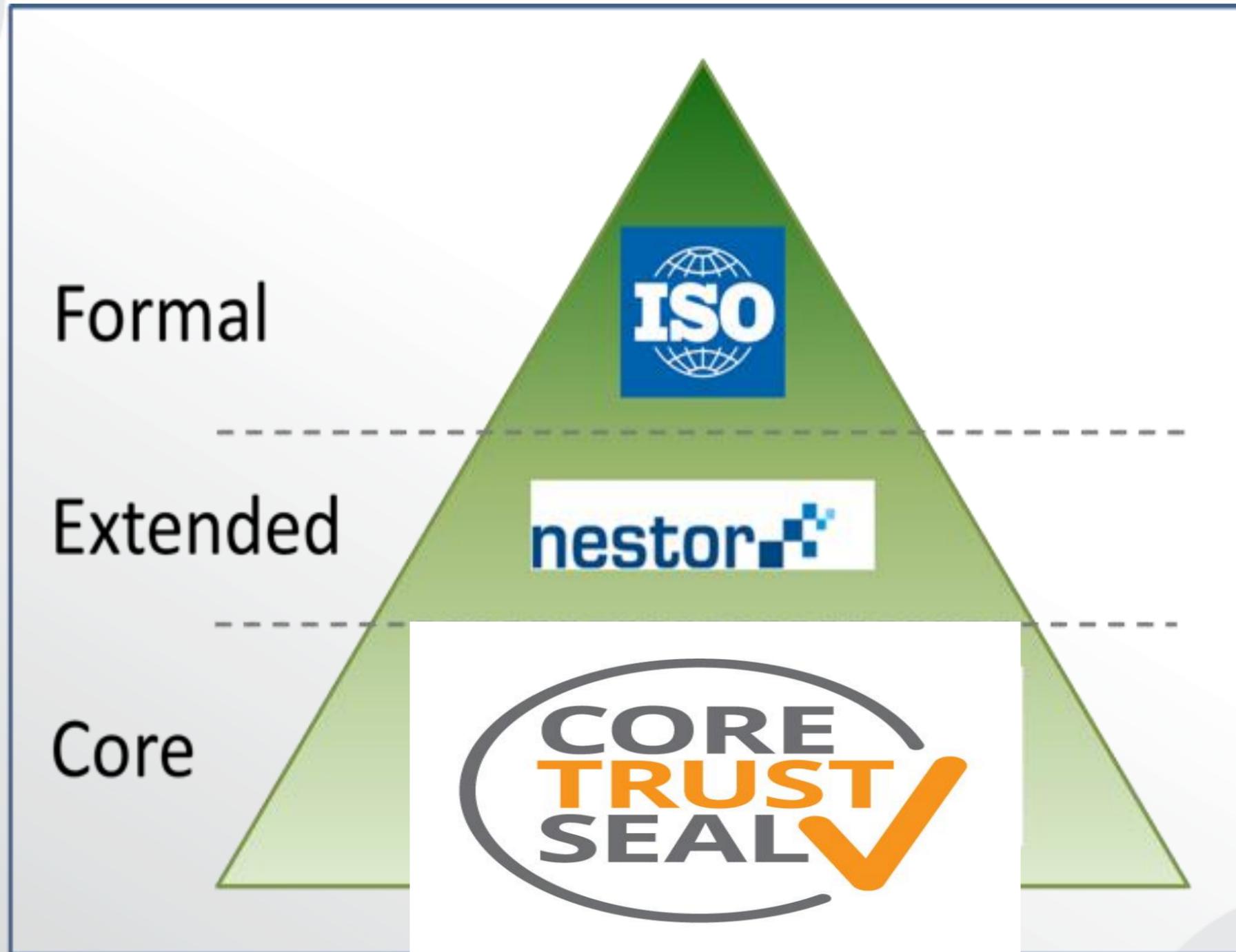
Name of repository	ICTWSS database
Additional name(s)	Database on Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts in 51 countries between 1960 and 2014
Repository URL	http://uva-aias.net/en/ictwss
Subject(s)	Economics Economic and Social Policy Social and Behavioural Sciences Humanities and Social Sciences
Description	The ICTWSS database covers four key elements of modern political economies: trade unionism, wage setting, state intervention and social pacts. The database contains annual data for all OECD and EU member states - Australia; Austria; Belgium; Bulgaria; Canada; Chile, Cyprus, the Czech Republic; Denmark; Estonia; Germany; Greece; Finland; France; Hungary; Iceland; Ireland; Israel, Italy; Japan, Korea, Latvia; Lithuania; Luxembourg; Malta; Mexico; the Netherlands; New Zealand; Norway; Poland; Portugal; Romania; Spain; Slovakia; Slovenia; Sweden; Switzerland; Turkey; the United Kingdom; and the United States – with some additional data for emerging economies Brazil; China; India; Indonesia; Russia; and South Africa; and it runs from 1960 till 2014.
Contact	Jelle.Visser@uva.nl
Content type(s)	Standard office documents Scientific and statistical data formats
Keyword(s)	trade unions wage settings state intervention social pacts
Repository type(s)	disciplinary
Mission statement for designated community	http://uva-aias.net/en/ictwss
Research data repository language(s)	eng
Data and/or service provider	dataProvider

How to select a repository?



- Certification as a '*Trustworthy Digital Repository*' with an explicit ambition to keep the data available in long term.
- Matches your particular data needs: e.g. formats accepted; mixture of open and restricted access; licenses.
- Provides guidance on how to cite the deposited data.
- The costs for e.g. depositing the data, data documentation, and support;
- Gives your submitted dataset a persistent and globally unique identifier for sustainable citations and to link back to particular researchers and grants.

Standards of Trust



Part of CTS's 16 requirements

R2. The repository maintains all applicable **licenses** covering data access and use and monitors compliance.

R3. The repository has a **continuity plan** to ensure ongoing access to and preservation of its holdings.

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with **disciplinary and ethical norms**.

R7. The repository guarantees the **integrity and authenticity** of the data.

R8. The repository accepts data and metadata based on **defined criteria to ensure relevance and understandability** for data users.

R10. The repository assumes responsibility for **long-term preservation** and manages this function in a planned and documented way.

R11. The repository has appropriate expertise to address **technical data and metadata quality** and ensures that **sufficient information** is available for end users to make quality-related evaluations.

R13. The repository enables users to **discover the data** and **refer to them in a persistent way** through proper citation.

R14. The repository enables reuse of the data over time, ensuring that **appropriate metadata** are available to support the understanding and use of the data.

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:

 or 

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

Keep everything? Forever?

Select what data you'll need and want to retain. Some selection criteria:

- Data underlying publications
- What can't be recreated, like interviews or environmental recordings
- What is potentially useful to others
- What has scientific, cultural or historical value

10 years is often stated in data policies and academic codes, but data can be valuable for ages, in climatology, sociology, health sciences, astronomy, linguistics, ...

Look beyond minimal retention periods where relevant.

RDNL Selection criteria: <http://www.researchdata.nl/en/services/data-management/selecting-research-data/>

DCC How-to guide: <http://www.dcc.ac.uk/resources/how-guides/appraise-select-data>



EXERCISE 3:

Find a research data repository

Exercise: Use re3data to find a repository

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

- Read the *Veteran tapes* project brief and identify what should be kept for the long term (5 mins)
- Search re3data.org for repositories (15mins), considering:
 1. Data type(s)
 2. Discipline
 3. Repository features
- Reporting back and questions (10 mins)





Reporting back...



What does
it mean?



What you may want to keep



- Characteristics: a) the original data cannot be recreated and b) the data may be sensitive
- **Raw** mpeg4 files (N = 600) – **access limited** to PI and project team
- **Processed** mpeg4 files, after anonymisation etc. (N depends on the content)
- **Transcripts** txt files (N = 600)
- **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 with metadata about the interviewees (N = 600) - **access limited!**
 - Communication with interviewees – general
 - Progress document listing the currently released data (using interviewee IDs)

re3data.org - examples

- “interviews” : 12 options
 - 5 with Trustworthiness certificates
- “social and behavioral sciences” : 311 options
 - 11 in the Netherlands,
 - 216 offer restricted access,
 - 70 support DDI metadata
- Browse by content type = “audiovisual data”: 367 options

Metadata standards

ABCD - Access to Biological Collection Data (2)
AVM - Astronomy Visualization Metadata (1)
CF (Climate and Forecast) Metadata Conventions (3)
DCAT - Data Catalog Vocabulary (1)
DDI - Data Documentation Initiative (23)
DIF - Directory Interchange Format (4)
Darwin Core (7)
DataCite Metadata Schema (21)
Dublin Core (70)
EML - Ecological Metadata Language (7)
FGDC/CSDGM - Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (12)
FITS - Flexible Image Transport System (2)
ISA-Tab (1)
ISO 19115 (14)
OAI-ORE - Open Archives Initiative Object Reuse and Exchange (5)
PROV (1)
RDF Data Cube Vocabulary (5)
Repository-Developed Metadata Schemas (2)
other (16)

In reality

The Netherlands: 46

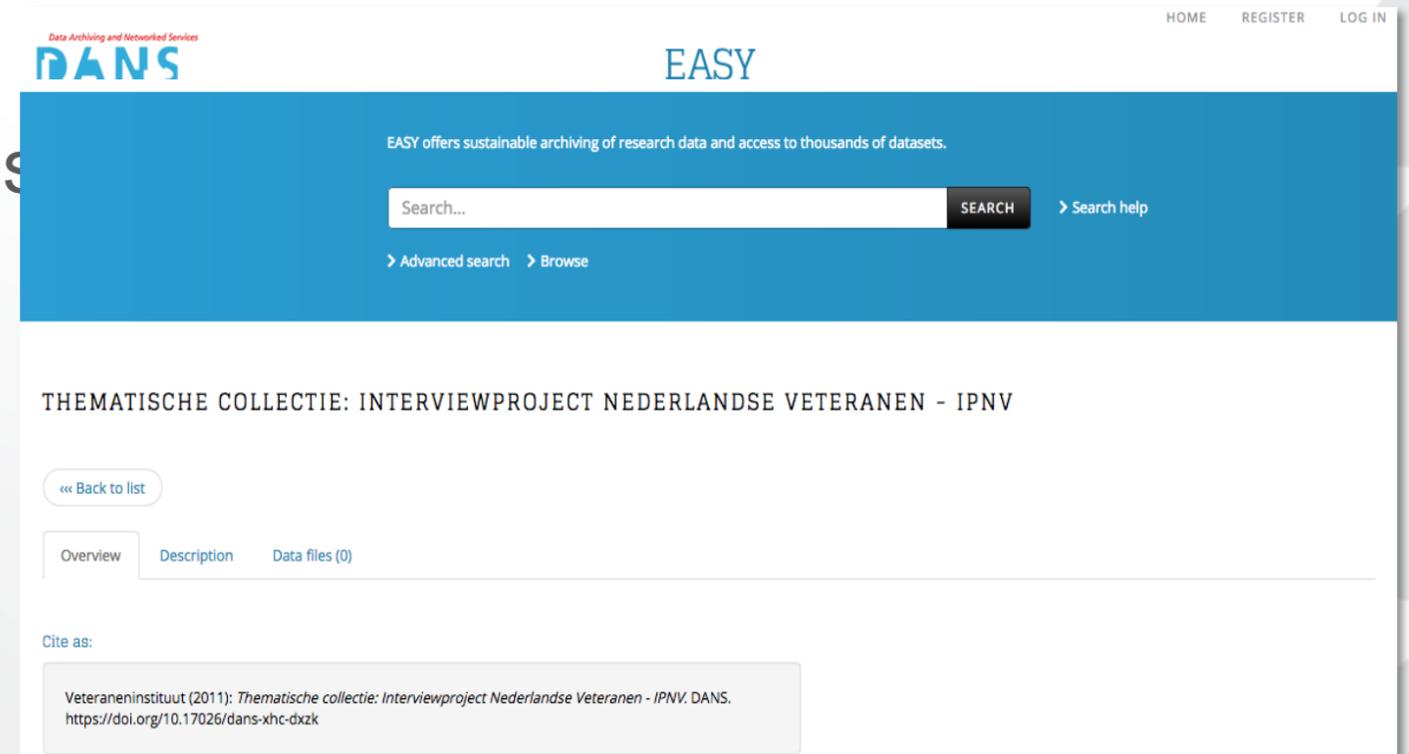
> DataSeal of Approval: 12

>> Audiovisual data: 13

>>> PID, DOI: 4. Manual inspection:

>>>> 2 repositories that accept “external” data

- 1 for science and technology
- 1 for humanities and social sciences



HOME REGISTER LOG IN

DANS EASY

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

Search... **SEARCH** > Search help

> Advanced search > Browse

THEMATISCHE COLLECTIE: INTERVIEWPROJECT NEDERLANDSE VETERANEN - IPNV

« Back to list

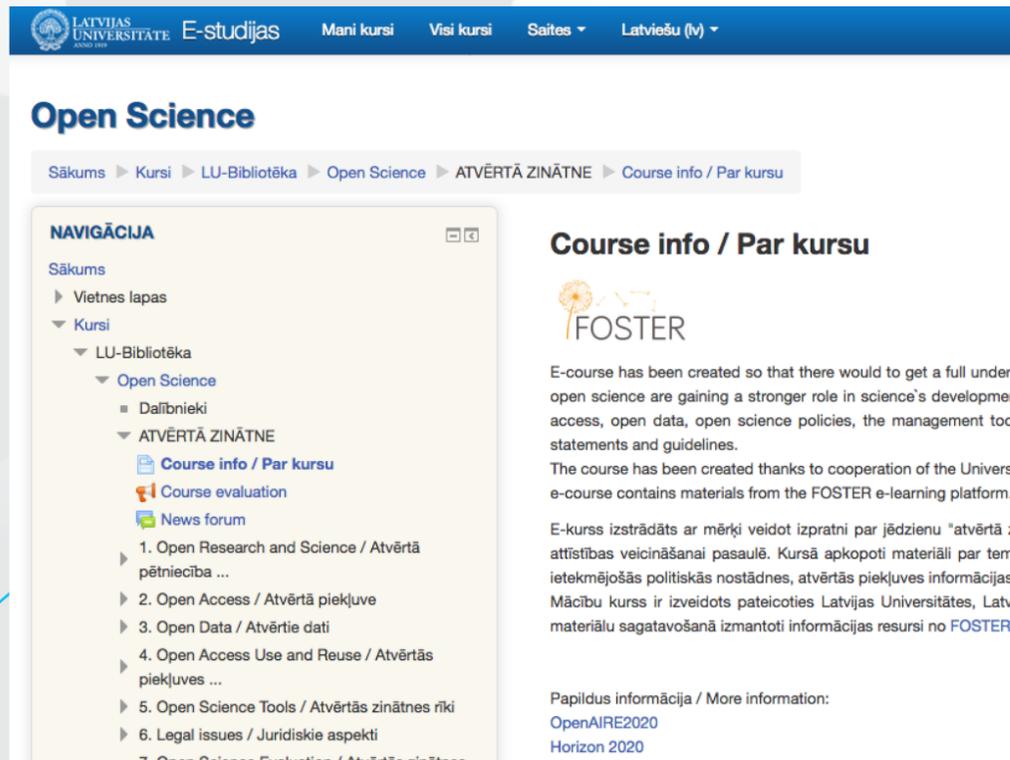
Overview Description Data files (0)

Cite as:

Veteraneninstituut (2011): *Thematische collectie: Interviewproject Nederlandse Veteranen - IPNV*. DANS. <https://doi.org/10.17026/dans-xhc-dxzk>

WRAP-UP

E-content on Open RDM



Open Science

Sākums ▶ Kursi ▶ LU-Bibliotēka ▶ Open Science ▶ ATVĒRTĀ ZINĀTNE ▶ Course info / Par kursu

NAVIGĀCIJA

- Sākums
- Vietnes lapas
- Kursi
 - LU-Bibliotēka
 - Open Science
 - Dalībnieki
 - ATVĒRTĀ ZINĀTNE
 - Course info / Par kursu
 - Course evaluation
 - News forum

Course info / Par kursu

FOSTER

E-course has been created so that there would to get a full unders open science are gaining a stronger role in science's development access, open data, open science policies, the management tools statements and guidelines.

The course has been created thanks to cooperation of the Universit e-course contains materials from the FOSTER e-learning platform.

E-kurss izstrādāts ar mērķi veidot izpratni par jēdzienu "atvārtā zi attīstības veicināšanai pasaulē. Kursā apkopoti materiāli par tema ietekmējošās politiskās nostādnes, atvārtās piekļuves informācijas Mācību kurss ir izveidots pateicoties Latvijas Universitātes, Latvij materiālu sagatavošanā izmantoti informācijas resursi no FOSTER e

Papildus informācija / More information:
[OpenAIRE2020](#)
[Horizon 2020](#)

PARTICIPATE SEARCH MONITOR SUPPORT OPEN ACCESS



DIGITAL INFRA for RESEARCH

Serving the user

30 November - 1 D

With welcome evening coo

29 November 2017 sponso

Venue: SQUARE, Brussels

- HELPDESK
 - Ask a question
 - FAQ
- RESOURCES
 - Guides
 - Copyright issues
 - H2020 Factsheets
- TRAINING
 - Workshops
 - Webinars
- BACKGROUND
 - Overview
 - Policies and Mandates
 - Open Access in FP7
 - Open Access in H2020
- IN PRACTICE
 - EU Member States
- PROJECTS
 - OpenAIRE-Connect
- PILOTS
 - FP7 Post-Grant OA Pilot
 - Open Research Data Pilot



<https://estudijas.lu.lv/mod/page/view.php?id=214107&lang=lv>

<https://www.openaire.eu/what-is-the-open-research-data-pilot>

<http://datasupport.researchdata.nl/en/> All content available under CC-BY-SA

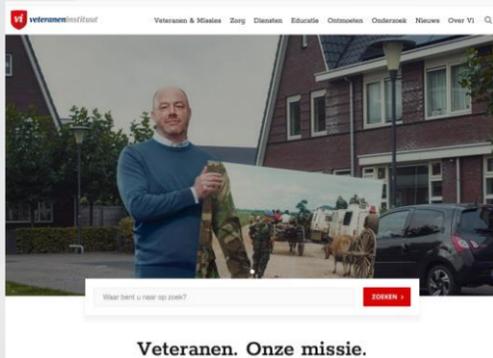
Raise your hands, please!

- I'll tell my colleagues that we need a better place to preserve research data.
- Preserving data is more important than sharing data.
- Sharing data is a must when you receive public funding.
- 80% FAIR is good enough for now.
- I'll write a DMP for my current project.
- I'll ask Gita and Ilga about Latvian regulations for data ownership.
- I'll do the course
<https://estudijas.lu.lv/mod/page/view.php?id=214107&lang=lv>
- Next year I'll teach a course on FAIR & open data.

Thank you!

Thanks to the Veteraneninstituut for data and use case.

Thanks to EUDAT and Research Data Netherlands for exercises and slides.



www.openaire.eu



[@openaire_eu](https://twitter.com/openaire_eu)



facebook.com/groups/openaire



linkedin.com/groups/OpenAIRE-3893548



elly.dijk@dans.knaw.nl

marjan.grootveld@dans.knaw.nl