Open Research Data & FAIR data management planning

Fieke Schoots, PhD | Belgrade, October 18th 2018

Universiteit Leiden
The Netherlands

Cartoons by Auke Herrema for RDA, Sept. 2015, Amsterdam
Illustrations:https://digitalbevaring.dk/

Program

1. Introduction
2. Open Science & Research Data
3. (Open) Research Data
4. FAIR data
5. Skills and training
6. Data management planning
7. Institutional approach: the Leiden example

1.

INTRODUCTION

Home ✓ News & Opinion

JAMA Journals Retract Six Papers by Cornell Researcher

Problems with Brian Wansink's research articles surfaced in 2017 and have now resulted in 13 retractions total.

Sep 19, 2018
ASHLEY YEAGER

Brian Wansink
WIKIMEDIA, BRIAN WANSINK

Update (September 20): Wansink says he plans to retire from Cornell in 2019, Retraction Watch *reports.*

Six papers coauthored by Cornell University consumer behavior researcher Brian Wansink have been retracted by three *JAMA* journals, according to a statement released today (September 19) by the publisher. The withdrawal follows revelations of emails suggesting misconduct, including regular data-massaging, in February along with five retractions and 13 corrections of papers in 2017.

In May, Howard Bauchner, a professor of pediatrics at Boston School of Medicine and editor-in-chief of the *JAMA* journals, raised concerns about several of Wansink's coauthored papers and the journal network asked Cornell to investigate the validity of the data.

"Cornell University has notified *JAMA* that based on its investigation they are unable to provide assurances regarding the scientific validity of the 6 studies," Bauchner writes in the statement. The university sent a statement to the journal network said it could not verify the results because it did not have access to the original data. The articles in *JAMA*, *JAMA Internal Medicine*, and *JAMA Pediatrics* have therefore been retracted.

"The university sent a statement to the journal network said it could not verify the results because it did not have access to the original data."

https://www.the-scientist.com/news-opinion/jama-journals-retract-six-papers-by-cornell-food-scientist--64828

Genome Biology

Farnham et al. Genome Biology (2017) 18:221
DOI 10.1186/s13059-017-1351-7

Open Access

COMMENT

Early career researchers want Open Science

Andrea Farnham[1], Christoph Kurz[2,3*], Mehmet Ali Öztürk[4], Monica Solbiati[5], Oona Myllyntaus[6], Jordy Meekes[7], Tra My Pham[8], Clara Paz[9], Magda Langiewicz[10], Sophie Andrews[11], Liisa Kanninen[6], Chantal Agbemabiese[12], Arzu Tugce Guler[13], Jeffrey Durieux[14], Sarah Jasim[15], Olivia Viessmann[11], Stefano Frattini[16], Danagul Yembergenova[17], Carla Marin Benito[9], Marion Porte[18], Anaïs Grangeray-Vilmint[19], Rafael Prieto Curiel[8], Carin Rehncrona[20], Tareq Malas[21], Flavia Esposito[9] and Kristina Hettne[21]

Abstract

Open Science is encouraged by the European Union and many other political and scientific institutions. However, scientific practice is proving slow to change. We propose, as early career researchers, that it is our task to change scientific research into open scientific research and commit to Open Science principles.

The Open Science situation

For 300 years, research journals have provided a stable record of the main conclusions of scientific studies, the methods, and the contact details of the scientists from whom data and materials might be obtained upon request. As the volume and complexity of research data explode, traditional research publications are failing to pay even lip service to the task of indexing data, let alone contribute meaningfully to data stewardship and preservation. Despite widespread discussion of the importance of Open Science and a growing recognition of the need for more sophisticated data stewardship practices, such as those in the FAIR (Findable, Accessible, Interoperable and Reusable) guidelines [1], the research community continues to do little to use available tools to index or share our vast datasets. The consequence is an escalation of data loss in an era when "data is the new gold" [2]. Although some in the scientific community see Open Science and data sharing as a "threat", even labeling those who use others' data as parasites [3], we believe that embracing and improving the Open Science tools already in place will facilitate better science, allowing us to harness the full potential of growing global scientific output.

Early career scientists, as relative outsiders to the scientific establishment, are often seen as dispensable, low-risk "experiments" but, working together to create interoperable systems, we have the opportunity to create change. Early career researchers have the least commitment toward professional hierarchy and are highly involved in data collection and analysis. Training young researchers to use the tools of Open Science can instigate a more reliable change in data stewardship. Here, we want to emphasize the importance of the implementation of Open Science principles across disciplines from a young researcher's point of view, and highlight the reasons why young researchers are the key for change.

As a group of early-career researchers who convened for the 2016 LERU Doctoral Summer School on Data Stewardship, we commit to: (1) the growth of an Open Science framework within which we can explicitly receive credit for and give credit to datasets with machine-readable metadata, provenance, and reproducible workflows; (2) establish training in the principles of Open Science and the relevant software and communication tools; (3) a pledge to be the first generation that will pass on the principles and mindset of Open Science to the next generation. If these three aims become a reality, we also envisage a strengthening of the credibility and reproducibility of scientific findings, and a reduction in fraudulent scientific data.

Growth of an Open Science framework

What does an Open Science framework look like? The European Commission on Open Science outlines a high-level vision of a future of science that includes Open Access, collaboration on platforms such as the European Open Science cloud, and the development of alternative metrics for measuring the impact of good science [4]. However, the roadmap to this future remains unclear to the

* Correspondence: christoph.kurz@helmholtz-muenchen.de
[2]Institute of Health Economics and Health Care Management, Helmholtz Zentrum München, 85764 Oberschleißheim, Neuherberg, Germany
[3]Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Ludwig-Maximilians-Universität München, 81 377 Munich, Germany
Full list of author information is available at the end of the article

BioMed Central

"Open Science is encouraged by the European Union and many other political and scientific institutions. However, scientific practice is proving slow to change. We propose, as early career researchers, that **it is our task to change scientific research into open scientific research and commit to Open Science principles**".

Universiteit Leiden

Search for subject or person | All

Nederlands | English

Using the Library | Search tools | Training | Research & publishing | Special Collections | About us | Support us

Home > Research & publishing > Centre for Digital Scholarship

Search the Catalogue

Books, articles, databases et

My Library account

Centre for Digital Scholarship

The Centre for Digital Scholarship collaborates closely with researchers, faculties, national and international colleagues, and other centres of expertise to facilitate and support Digital Scholarship.

The Centre for Digital Scholarship organizes meetings and workshops and it is the obvious partner for researchers to contact for questions, consultancy, and training on the following topics:

- Data management
- Text & data mining
- Open Access
- Publication advice
- Copyright
- Collaborative environments
- GIS

The CDS also offers services and advice for (research)projects that involve:

. Databases and websites
. Creating and managing digital collections
. Metadata
. Management of projects using digital research methods
. Long term preservation
. Digitisation of analogue primary sources

Research & publishing

- Centre for Digital Scholarship
  - About the CDS
  - Facilities
  - Workshops
  - Presentations and Publications
- Data management
- Text & data mining
- Open Access
- Copyright Information Office
- Geographical Information Systems
- Virtual Research Environments
- Scholarly publishing

Staff

Laurents Se... Centre Scholarship

Michelle va... Digital Scholarship Librarian

Ben Companjen
Digital Scholarship Librarian

Fieke Schoots
Digital Scholarship Librarian

Peter Verhaar
Digital Scholarship Librarian

Contact

Centre for Digital Scholarship
Universiteitsbibliotheek, BG
Witte Singel 27,
2311 BG Leiden

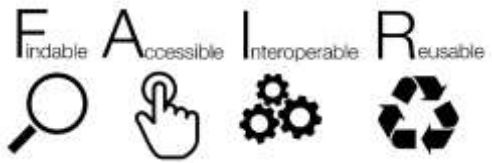
0031 71 527 2814

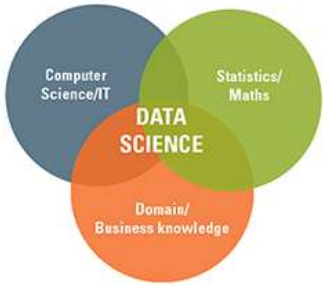
cds@library.leidenuniv.nl

Leiden University in figures

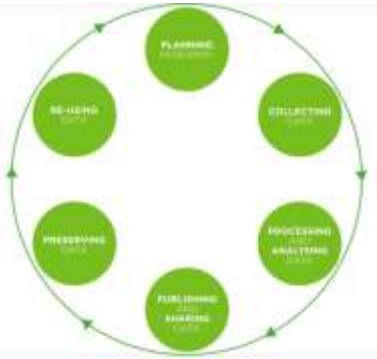
7	Faculties
16	Nobel Prizes
46	Bachelor's programmes
83	Master's programmes
120	Nationalities
28,130	Students
109,000	Alumni
2	Cities
6,700	Employees

Centre for Digital Scholarship





Transferable skills



2

OPEN SCIENCE & RESEARCH DATA

Eight pillars of Open Science for the EC

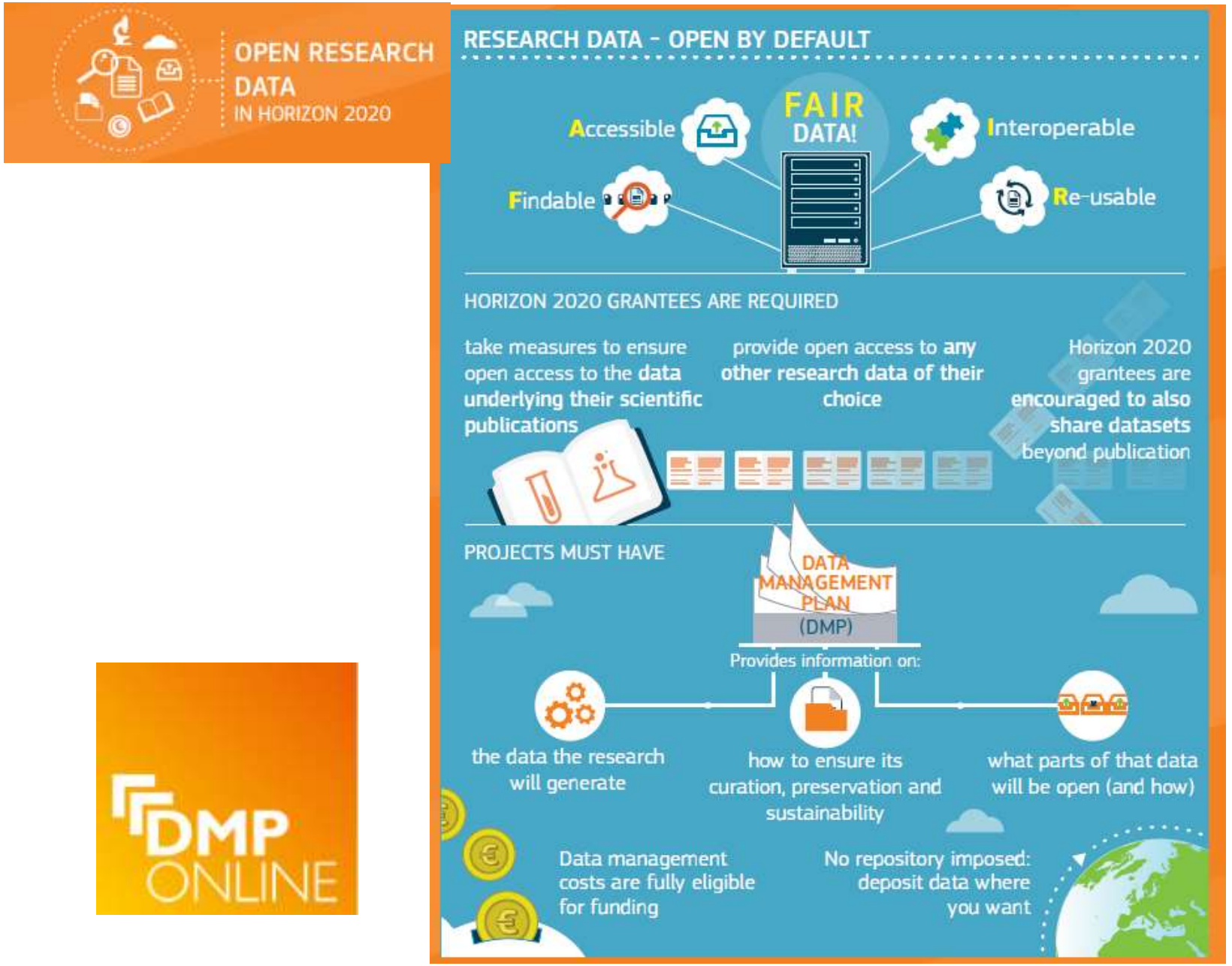
1. The future of scholarly publishing
2. **FAIR data**
3. The European Open Science Cloud
4. Education and skills
5. Rewards and incentives
6. Next-generation metrics ('Altmetrics')
7. Research integrity
8. Citizen science

“Open Science as a modus operandi”

Horizon Europe

“The principle of Open Science will become the modus operandi of the new Programme. It will go beyond the open access policy of Horizon 2020 and require open access to publications and data (with robust opt-outs for the latter), and to data management plans. The Programme will foster the widespread use of FAIR data; and activities that enhance researchers’ skills in open science and support reward systems that promote open science.”

- https://ec.europa.eu/commission/sites/beta-political/files/budget-may2018-horizon-europe-regulation_en.pdf



DMP ONLINE

https://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf
Guidelines on FAIR Data Management in Horizon 2020

Dutch National Plan Open Science

- ✓Open access to publications
- ✓Optimal reuse of research data
- ✓New evaluation and award systems for researchers

National Platform Open Science

Publishers : Data Availability Policy (DAP

Plos One : “Authors must make all data publicly available, without restriction, immediately upon publication of the article.”

http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2/

Policy Types

SPRINGER NATURE

Type 1	Type 2	Type 3	Type 4
Data sharing and data citation is encouraged	Data sharing and evidence of data sharing encouraged	Data sharing encouraged and statements of data availability required	Data sharing, evidence of data sharing and peer review of data required

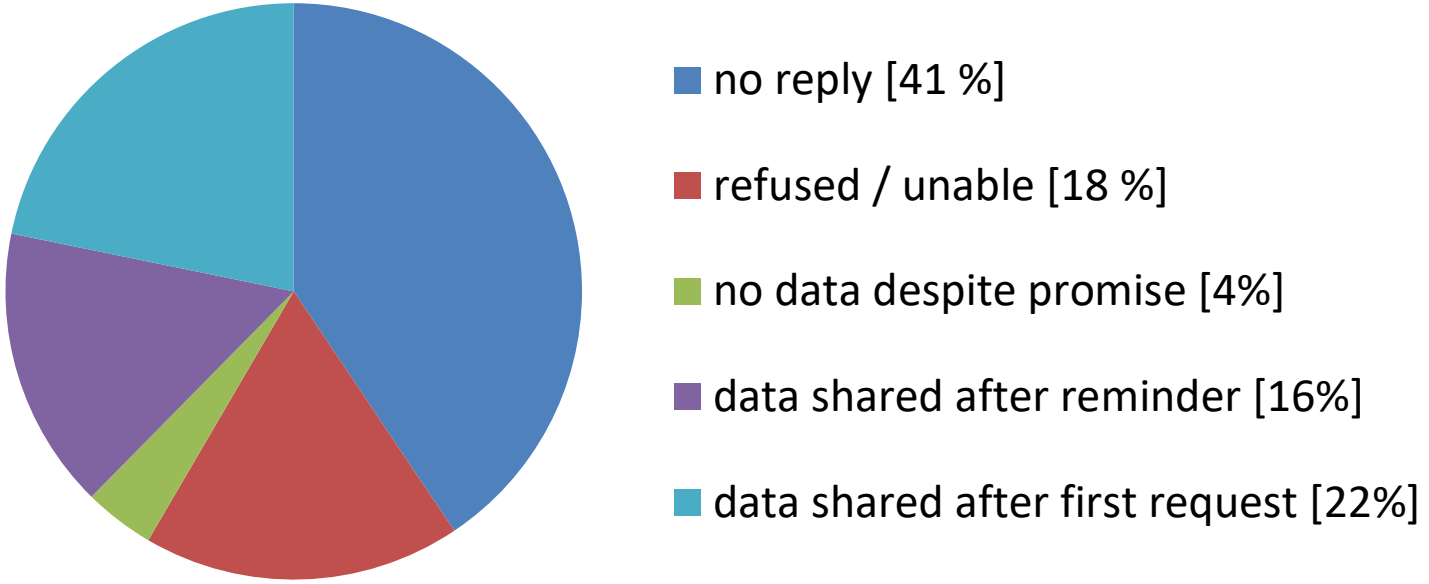
http://blogs.nature.com/ofschemesandmemes/2016/07/05/promoting-research-data-sharing-at-springer-nature

Availability of psychological research data

Vanpaemel, W. et al., (2015). Are We Wasting a Good Crisis? The Availability of Psychological Research Data after the Storm. Collabra: Psychology. 1(1), p.Art. 3.
DOI: http://doi.org/10.1525/collabra.13

APA Ethics code

"we requested data from 394 papers, published in all issues of four APA journals in 2012. We found that 38% of the researchers sent their data immediately or after reminders."



3.

(OPEN) RESEARCH DATA

Research data are…

… all data collected or generated during scientific research

Data Management: all activities to manage data in a responsible way

- ✓Organisation
- ✓Documentation
- ✓Storage
- ✓Sharing
- ✓Archiving

& Protection

Data Management Plan

Research data : stages

http://www.data-archive.ac.uk/create-manage/life-cycle

Research data: methods

remote sensing data, survey data, field recordings, sample data

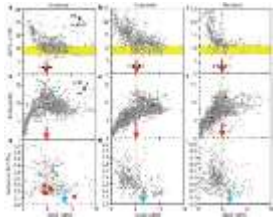
- data captured in real time that is usually unique and irreplaceable.

Observational

Experimental

- data captured from lab equipment that is often reproducible.

chromatograms, magnetic field data



text and data mining, compiled databases, 3D models

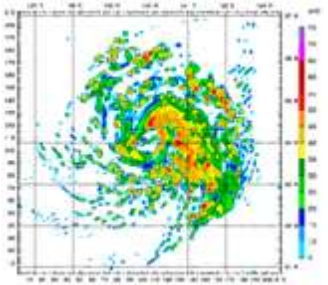
Derived or compiled

- resulting from processing or combining 'raw' data.

Models or simulation

model and metadata more

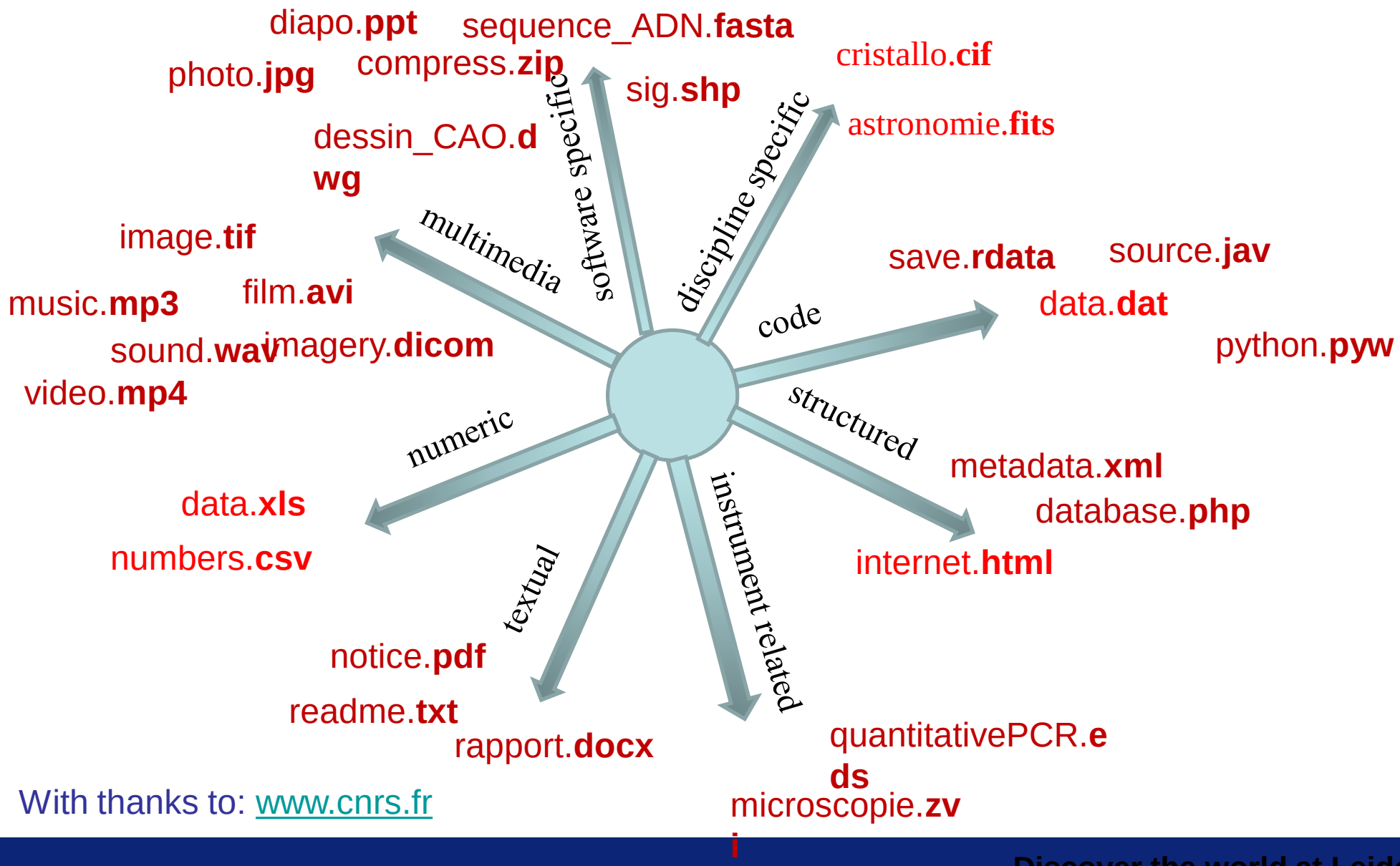
- often reproducible, if correctly documented



climate models, economic models

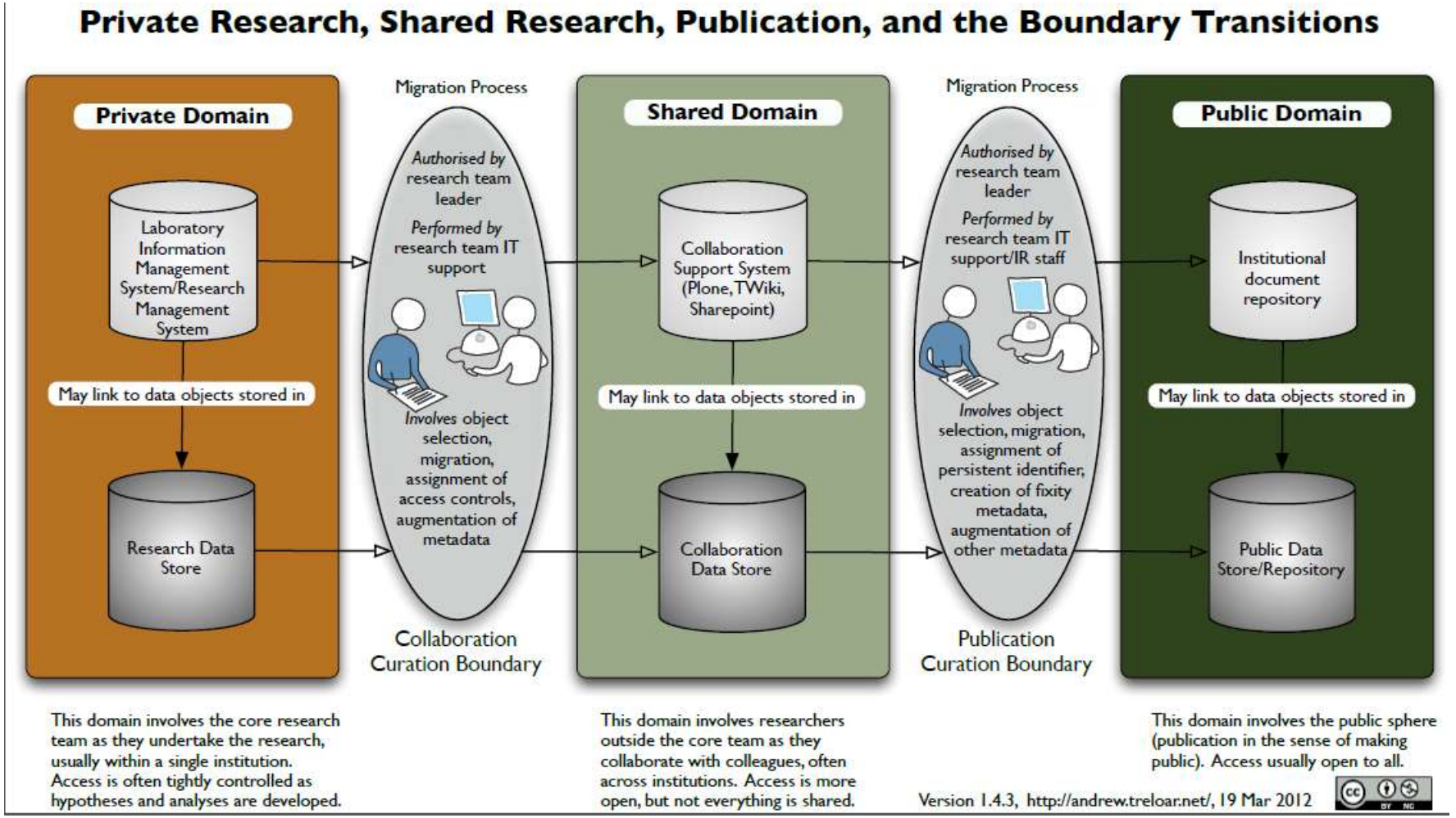
With thanks to: www.cnrs.fr

Research data: types



With thanks to: www.cnrs.fr

Research data: domain



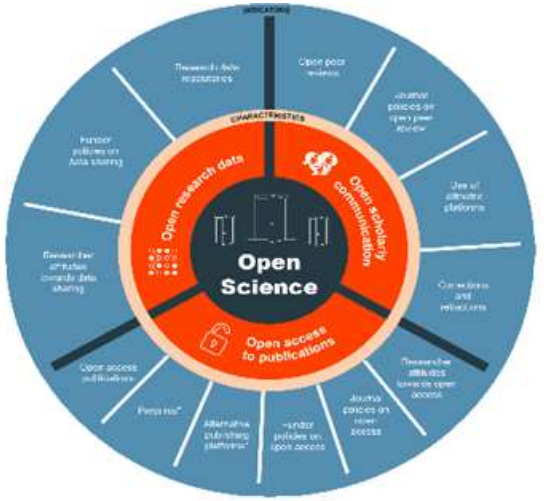
Data Curation Continuum (ANDS)

Open research data

“refers to the data underpinning scientific research results that has no restrictions on its access, enabling anyone to access it”

Indicators according to open science monitor

- ✓availability of data repositories
- ✓policies of research funders and journals
- ✓researchers' attitude towards data sharing



https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en#additional-indicators

https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/about-open-science-monitor_en

Open data: The researcher perspective

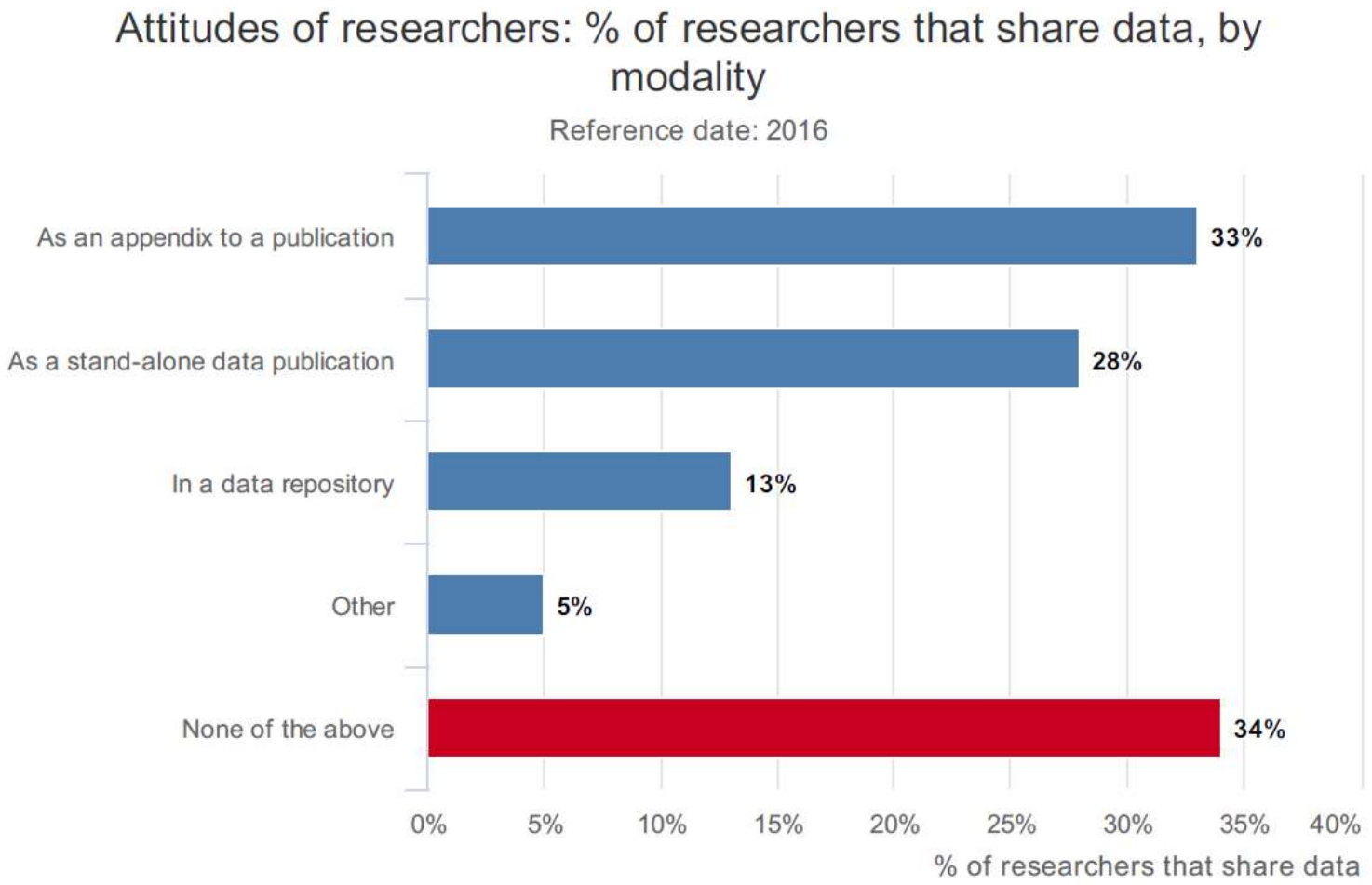
Report from CWTS / Elsevier in 2017 (n= 1162)

73% of Academics Say Access to Research Data Helps Them in Their Work; 34% Do Not Publish Their Data

https://www.cwts.nl/news?article=n-r2q244&title=73-of-academics-say-access-to-research-data-helps-them-in-their-work-34-do-not-publish-their-data

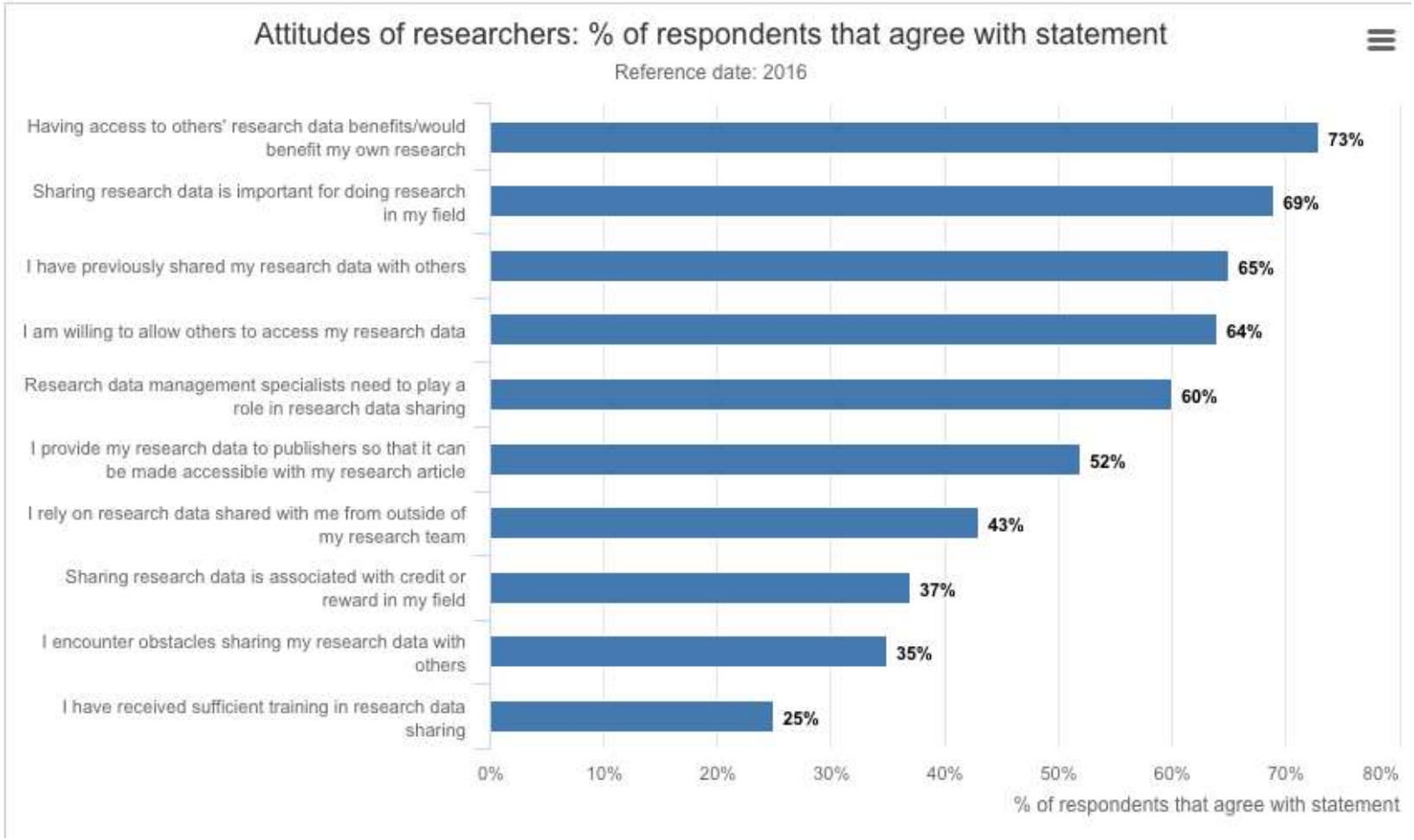
How do researchers share data?

“This indicator shows how researchers share data and their underlying attitudes, based on the survey for the 2017 report "Open data : The researcher perspective". (CWTS / Elsevier, n=1162)

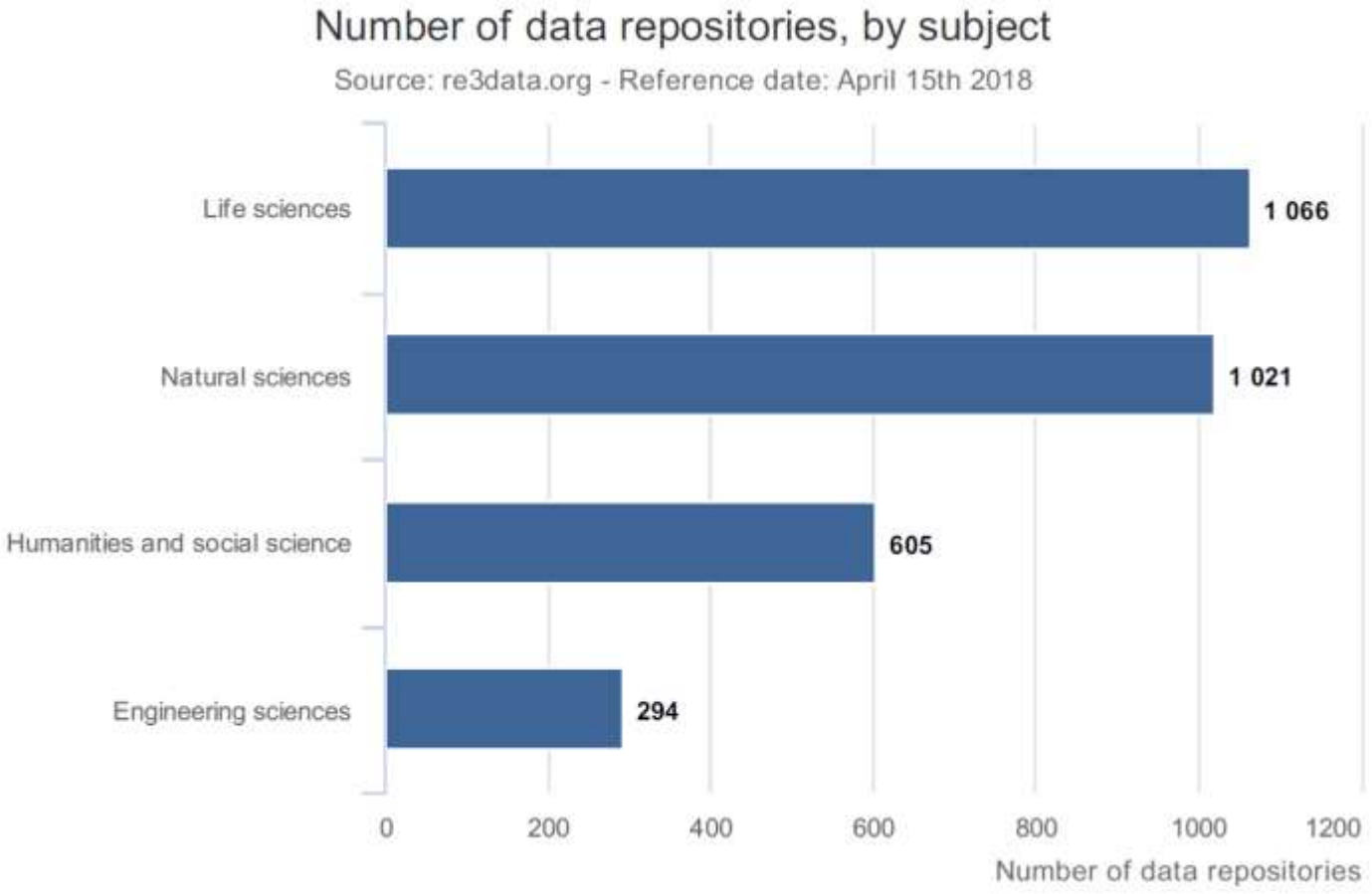


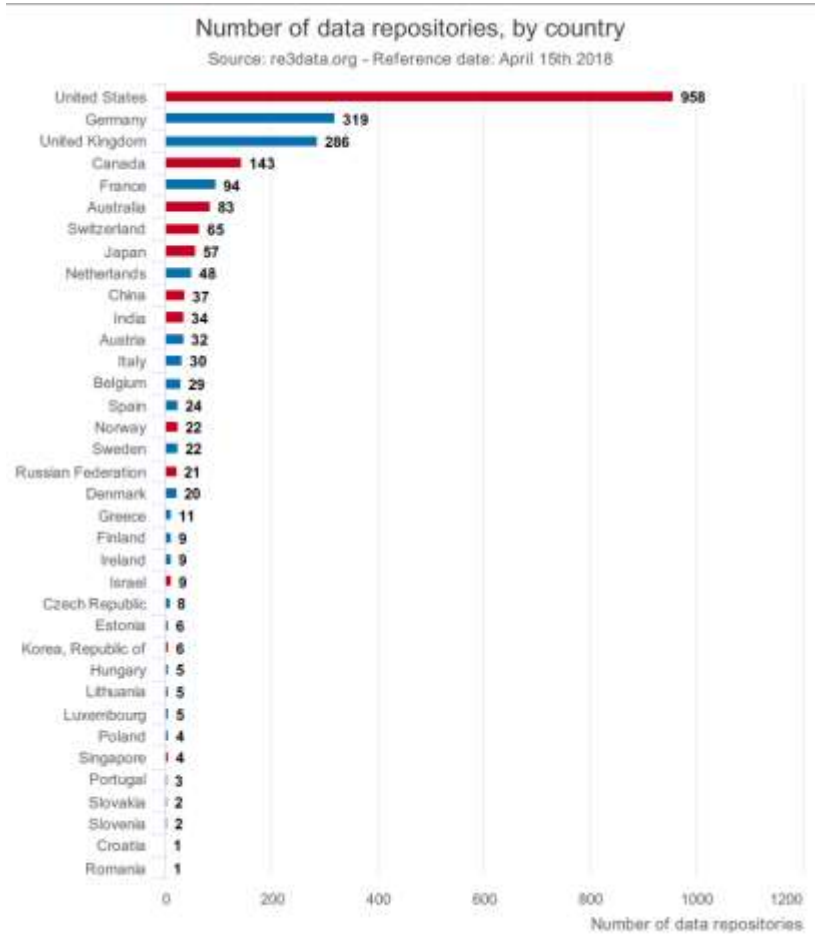
https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en#additional-indicators

- 75 % has benefit of having access to others' data
- 35 % encounter obstacles sharing research data
- 25 % has had sufficient training in research data sharing



https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en#additional-indicators





https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en#additional-indicators

Reasons not to share

- not enough training in data sharing,
- sharing data is not associated with credit or rew
- privacy issues
- proprietary aspects and ethics

“Data-sharing practices depend on the field: there is no general approach. In intensive data-sharing fields, data sharing practice is embedded into the research design and execution”.

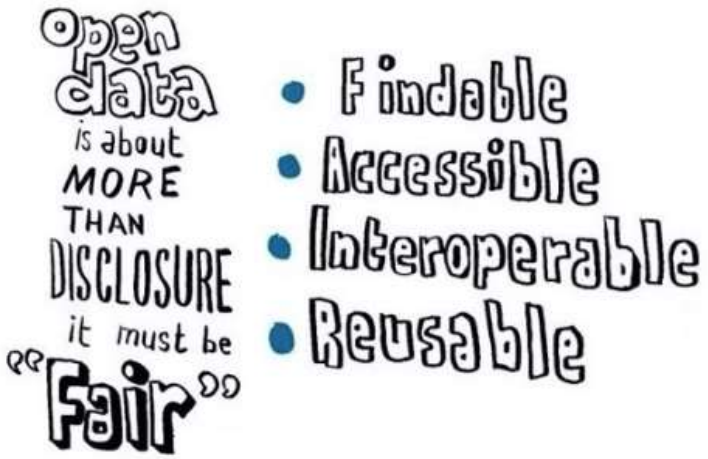
https://www.cwts.nl/news?article=n-r2q244&title=73-of-academics-say-access-to-research-data-helps-them-in-their-work-34-do-not-publish-their-data

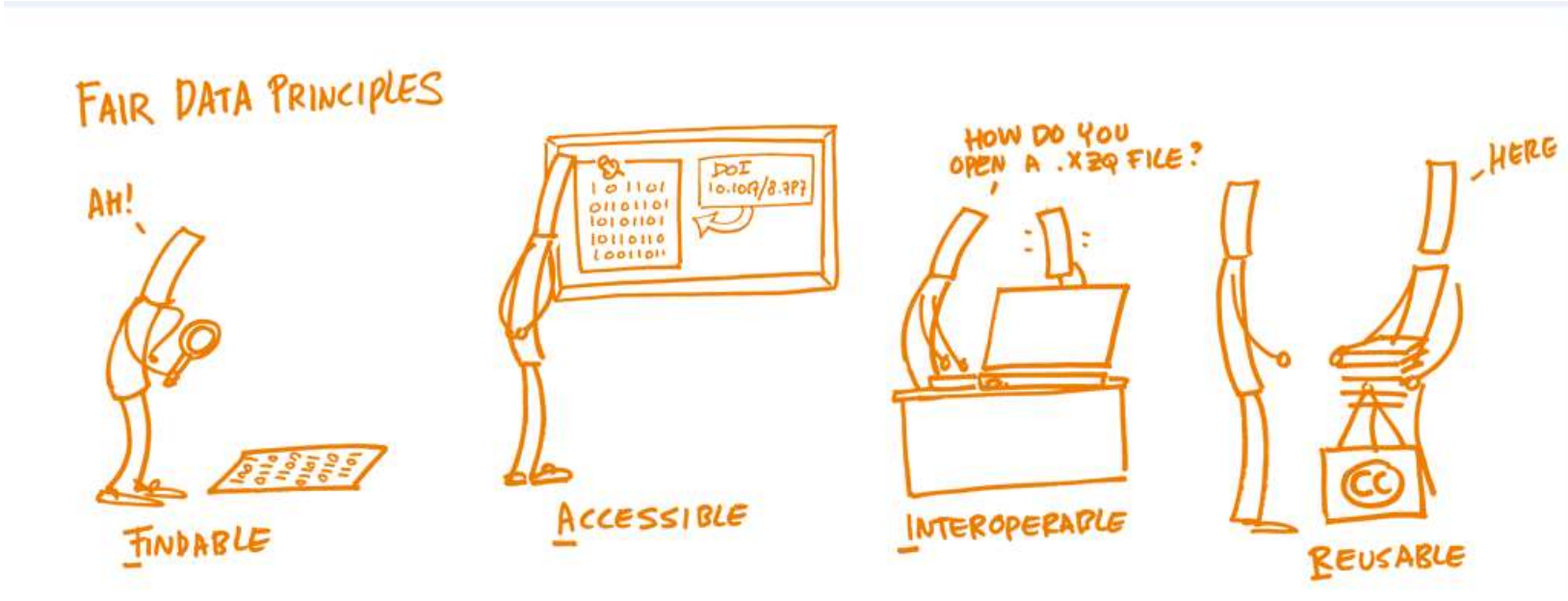
4

FAIR DATA

FAIR Data

DTL
DUTCH TECHCENTRE FOR LIFE SCIENCES





https://www.dtls.nl/fair-data/fair-data/ | https://www.force11.org/group/fairgroup

Findable data

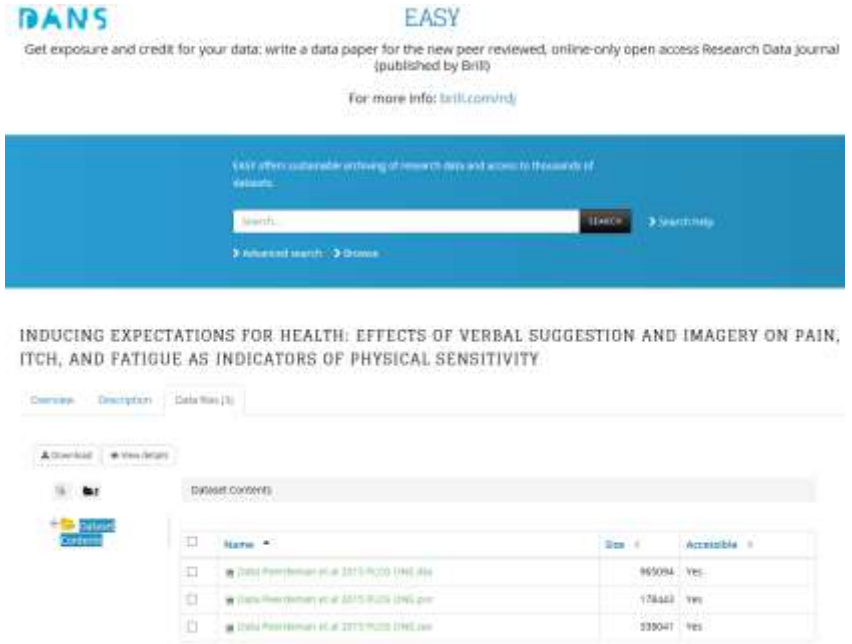
Easy to find by both humans and computer systems and based on mandatory description of the metadata that allow the discovery of interesting datasets

- ❑ Location: shared or public domain?
- ❑ Organisation
  - ✓adopt naming convention
  - ✓adopt a versioning policy
  - ✓use standards for metadata
- ❑ Persistent identifier

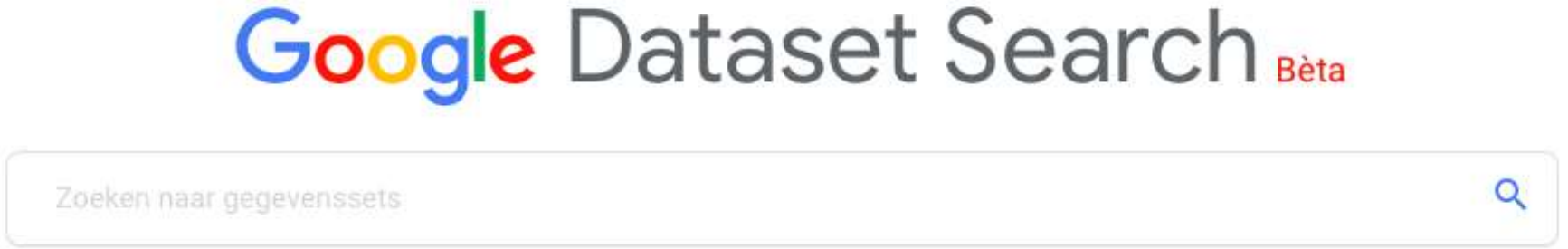
Link publication to data and vice versa

Peerdeman KJ, van Laarhoven AIM, Donders ART, Hopman MTE, Peters ML, Evers AWM (2015) Inducing Expectations for Health: Effects of Verbal Suggestion and Imagery on Pain, Itch, and Fatigue as Indicators of Physical Sensitivity. PLoS ONE 10(10): e0139563. doi:10.1371/journal.pone.0139563

Data Availability: Data have been deposited to DANS: http://dx.doi.org/10.5072/dans-zbp-8eqr.



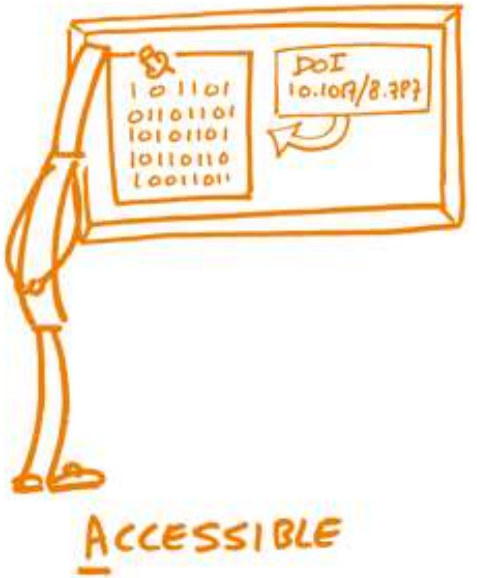
Data catalogue(s)

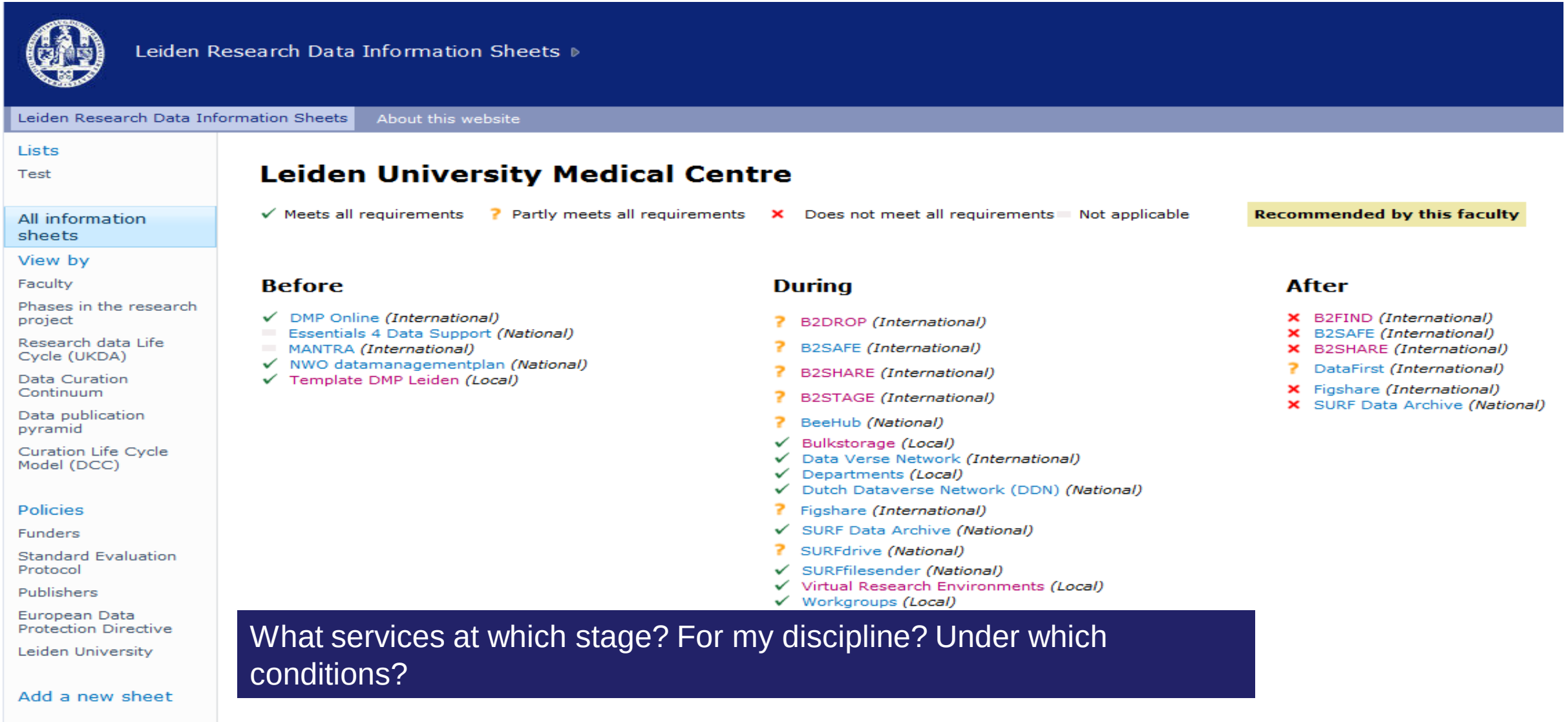


Accessible Data

Stored for long term such that they can be easily accessed and/or downloaded with well-defined license and access conditions (Open Access *when possible*), whether at the level of metadata, or at the level of the actual data content;

- Location
  - During project
  - After project: certified archive?
- Access conditions
  - authorisation processes
  - user agreements
- Software





What services at which stage? For my discipline? Under which conditions?

https://vre.leidenuniv.nl/vre/lrd/Pages/information-Sheets.aspx
https://www.edugroepen.nl/sites/RDM_platform/Dienstencatalogus/SitePages/Dienstencatalogus.aspx

Interoperable Data

- Data **formats**
  - data provided in commonly understood and preferably open formats?

- **Standards**
  - metadata
  - controlled vocabularies, keywords, thesauri or ontologies are used where possible

    "Same terminology for same things"

    https://weblog.wur.eu/openscience/vocabularies-and-the-i-in-fair-data-principles/

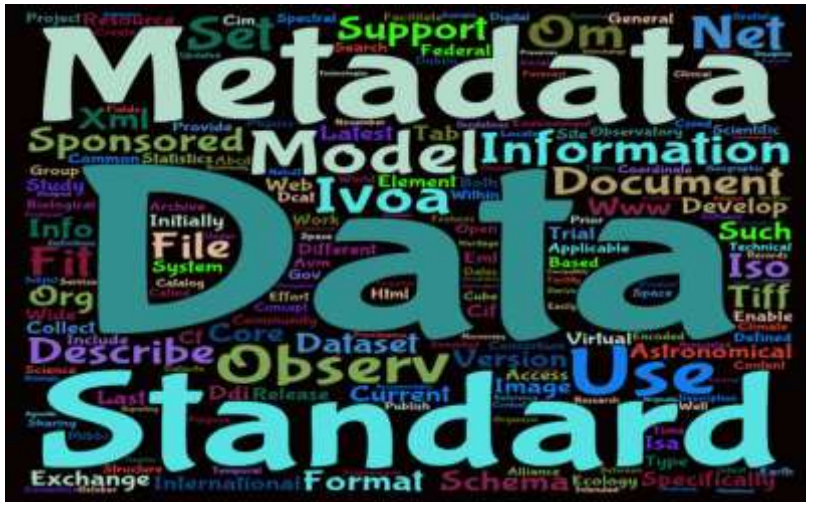
- **Documentation**
  - Are qualified references and links provided to other related data?

DANS Preferred formats

https://dans.knaw.nl/nl/deponeren/toelichting-data-deponeren/bestandsformaten

Type	Preferred format(s)	Acceptable format(s)
Tekstdocumenten	PDF/A (.pdf)	ODT (.odt) MS Word (.doc, .docx) RTF (.rtf) PDF (.pdf)
Platte tekst	Unicode text (.txt)	Non-Unicode text (.txt)
Opmaaktaal	XML (.xml) HTML (.html) gerelateerde bestanden: .css, .xslt, .js, .es	SGML (.sgml)
Spreadsheets	ODS (.ods) CSV (.csv)	MS Excel (.xls, .xlsx) PDF/A (.pdf) OOXML (.docx, .docm)
Databases	SQL (.sql) SIARD (.siard) tabellen uit DB (.csv)	MS Access (.mdb, .accdb) (versie 2000 of later) dBase (.dbf) HDF5 (.hdf5, .he5, .h5)
Statistische data	SPSS Portable (.por) SPSS (.sav) STATA (.dta) DDI (.xml) data (.csv) + setup (.txt)	SAS (.7dat; .sd2; .tpt) R (* in onderzoek)
Afbeeldingen (raster)	JPEG (.jpg, .jpeg) TIFF (.tif, .tiff) PNG (.png) JPEG 2000 (.jp2)	DICOM (.dcm) (in overleg)
Afbeeldingen (vector)	SVG (.svg)	Illustrator (.ai) EPS (.eps)
Audio	BWF (.bfw) MXF (.mxf) Matroska (.mka) FLAC (.flac)	WAVE (.wav) MP3 (.mp3) AAC (.aac, .m4a) AIFF (.aif, .aiff) OGG (.ogg)
Video	MXF (.mxf) Matroska (.mkv)	MPEG-4 (.mp4, .m4a, .m4v, ...) MPEG-2 (.mpg, .mpeg, .m2v, .mpg2, ...) AVI (.avi) QuickTime (.mov, .qt)
Computer Aided Design (CAD)	AutoCAD DXF versie R12 (.dxf)	AutoCAD andere versies (.dwg, .dxf)
Geografische informatie (GIS)	GML (.gml) MIF/MID (.mif/.mid)	ESRI Shapefiles (.shp en bijbehorende bestanden) MapInfo (.tab en bijbehorende bestanden) KML (.kml)

(metadata) standards



Work done a.o. by

- Digital Curation Centre (UK)

- Metadata standards working group

- Research Data Alliance Metadata directory

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

FAIRsharing.org
standards, databases, policies

Data Archiving and Networked Services
DANS

NARCIS

HOME OVER NARCIS INLOGGEN ENGLISH

‹ TERUG NAAR ZOEKRESULTAAT ‹ VORIGE 2 van 14 resultaten VOLGENDE ›

CentERdata

DATASET

TONIC IMMOBILITY IN RESPONSE TO TRAUMA: PREVALENCE AND CONSEQUENCES > WAVE 1 - PART 1 (2011)

Overzicht

Dublin Core metadata

Titel	Tonic immobility in response to trauma: Prevalence and consequences > Wave 1 - Part 1
Auteur	Hagenaars, M.A. Hagenaars, J.A.
Datum	2011-07-04
Trefwoord(en)	Demography and population, Health and well-being, Psychology
Uitgever	CentERdata
Samenvatting	In July 2011 a questionnaire on traumatic experiences was administered to the LISS panel.The survey investigates traumatic experiences people may have had in the past. This is part one of two questionnaires on traumatic experiences.
Rechten	2013 CentERdata
Dataset	› http://www.lissdata.nl/dataarchive/study_units/view/376
Metadata	› XML
Bron	CentERdata

Reusable Data

Ready to be used for future research and to be processed further using computational methods.

- **Provenance** information
  - how, why and by whom have the data been created and processed
- **Domain** standards
  - data and metadata meet relevant domain standards
- **License**
  - data have a clear and accessible data usage license

Indicators for FAIR Research Data

ZonMW example (Dutch Health Funder)

1. Link to the repository's search engine / (metadata) catalogue / web portal **listing** the data collection

2. **DOI code** (persistent identifier) providing a permanent link to the data collection

3. Link (or persistent identifier) to the data collection's **terms of use**
*(this **key item** can be dropped if the collection is open access)*

4. **Metadata standard** allowing the data collection to be linked with other collections

5. Link to the archive or trusted digital repository for **long-term archival**

FAIRMETRICS: http://fairmetrics.org

6

SKILLS & TRAINING

Open Science skills @ Europe

"The overarching goal is to ensure that OS skills become an integral and streamlined component of the standard education, training and career development paths of researchers, and if possible even at earlier career stages, in schools and universities."

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

Report of the Working Group on Education and Skills under Open Science, The Working Group on Education and Skills under Open Science July, 2017
doi: 10.2777/121253

Open Science Wheel: http://ec.europa.eu/research/openscience/index.cfm?pg=home&section=monitor

Open Science skills @ Europe

“The ability to re-use and re-analyse research outcomes must be considered an integral part of the basic set of competences required by researchers to conduct successful research. Data stewards and librarians also need to be up-skilled to assist researchers in properly managing, preserving and sharing data, according to clear institutional guidelines.”

Towards Open Access to Research Data. Aims and recommendations for university leaders and National Rectors’ Conferences on Research Data Management and Text and Data Mining, European University Association, 2017

https://eua.eu/component/attachments/attachments.html?id=500

Online training for RDM (supporters)

data tree

Expert Tour Guide on Data Management

cessda eric

Datatree - Data Training Engaging End-users
Home / Courses

Essentials 4 Data Support is an introductory course for those people who (want to) support researchers in storing, managing, archiving and sharing their research data.

Essentials 4 Data Support is a product of Research Data Netherlands.

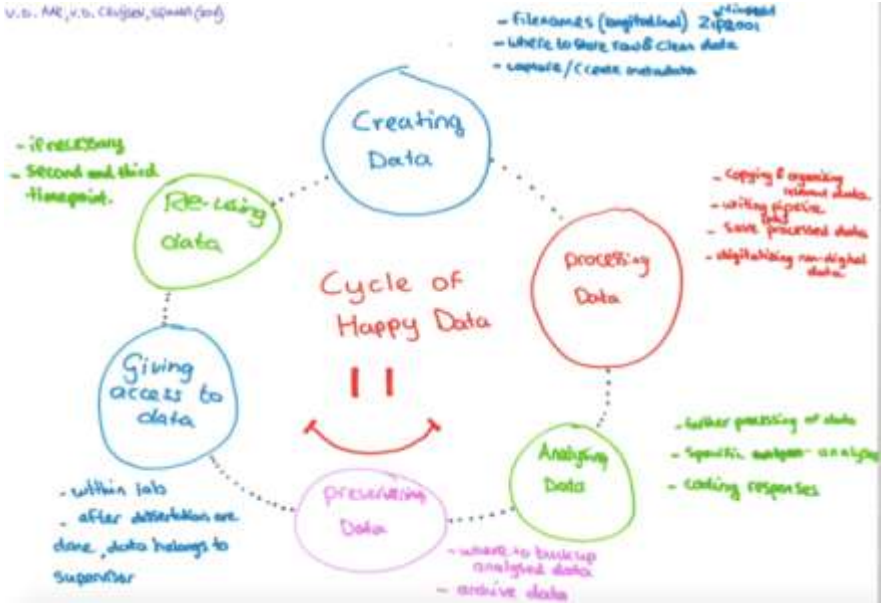
Tailor-made training for PhD students

- In collaboration with senior researcher

- Lectures on data management, DMP, data quality,
  Electronic Lab Notebooks etc. (by experts)

- Hands-on exercises
  - Design your own data life cycle
  - Find data underlying publication
  - Share data when with whom
  - Anonymising data, etc.

- Write and present DMP



Workshop ‘How to write a DMP’

Data collection & creation

Data storage and security

Documentation and metadata

Data access, sharing and reuse

Data preservation and archiving

2,5 hours
All disciplines
All templates

5

DATA MANAGEMENT PLANNING

Leiden University Data Management Plan[i]

The Research Data Management Regulations Leiden University requires researchers to write a data management plan at the start of a long-term research project[ii]. Please contact the Centre for Digital Scholarship at the University Libraries Leiden if you need help: datamanagement@library.leidenuniv.nl

Name and contact details	*Please include email address and telephone number*
Name of project and group	*Name your supervisors*
Description of your research	*Briefly describe your research to help others understand the purposes for which the data are being collected or created. Max. 50 words.*
Project duration	Start: *DD-MM-YYYY* End: *DD-MM-YYYY*
Names of people and their responsibilities for data management	*Responsibilities can be collecting, storing, documenting, sharing and archiving the data.* *Naming anyone with specific roles and responsibilities for data management is especially important for collaborative projects that involve many researchers and/or partner organisations.*
Funding body(jes)	*If applicable.*
Grant number	*If applicable. A grant number provides unique identification for the grant.*
Partner organisations	*If applicable. These may be research partners that use your data, or that you use data from.*

About this Data Management Plan

Date written	*DD-MM-YYYY*
Date last update	*DD-MM-YYYY*
Version	*A new version of the DMP should be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.* *Don't forget to include the date.*

Changes in this version of the Data Management Plan

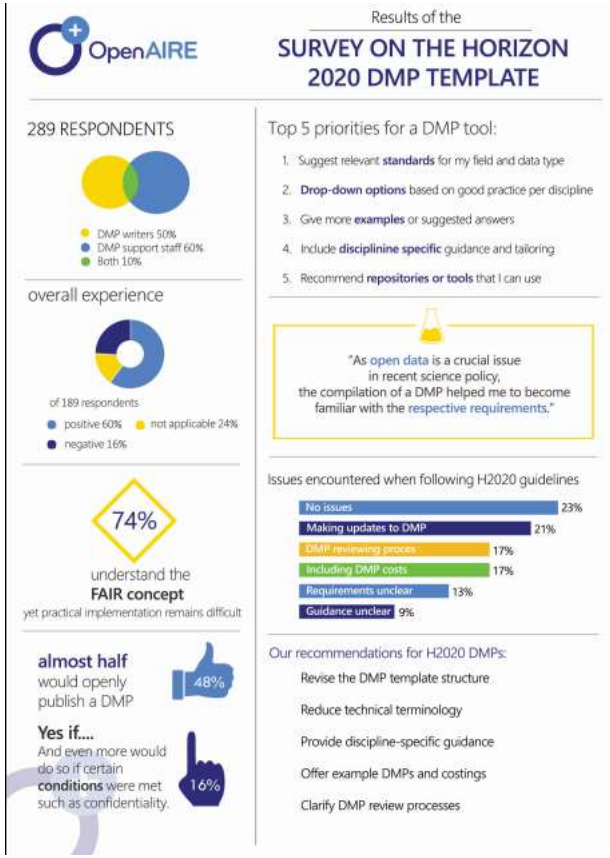
Component	**Progress / Execution** *Please describe briefly what progress you have made, any questions or issues you have encountered and want to discuss, etc.*
1. Data collection	.....
2. Data storage and back-up	.....
3. Data documentation	.....
4. Data access, sharing and reuse	.....
5. Data preservation and archiving	.....

UL Template Data Management Plan v. 2.2 / 20161129

Towards harmonisation in Europe?

1. Survey on the Horizon 2020 DMP template
   - ➢ Revise DMP template
   - ➢ Reduce technical terminology

2. Science Europe
   - ➢ Core requirements for DMP
   - ➢ Discipline approach



https://www.scienceeurope.org/policy/policy-areas/research-data/rdm-initiative/

https://zenodo.org/record/1120245#.W8eSJC-iHq0

Data Management Plan

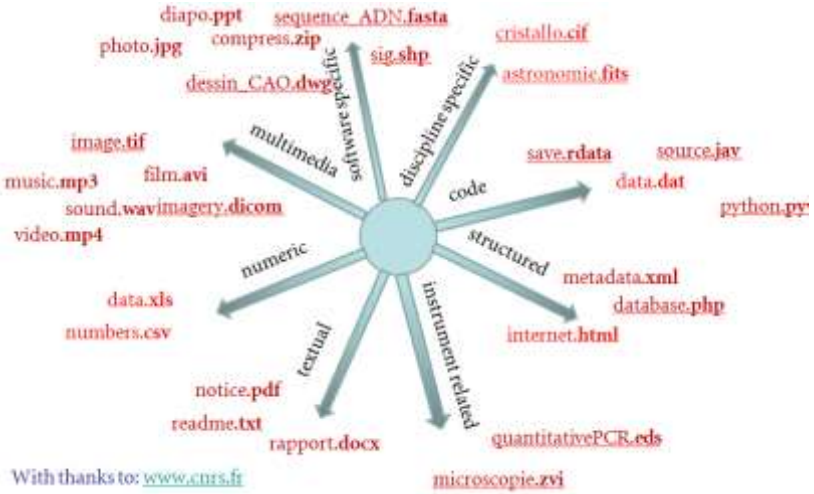
- Data collection & creation
- Data storage and security
- Documentation and metadata
- Data access, sharing and reuse
- Data preservation and archiving

It’s all about awareness and taking the right decisions from the start…

Data Collection

- ✓ **Existing data**
  - ✓Provenance
  - ✓Access
  - ✓License for reuse

- ✓**Creating data**
  - ✓Method
  - ✓Data format
  - ✓Data size

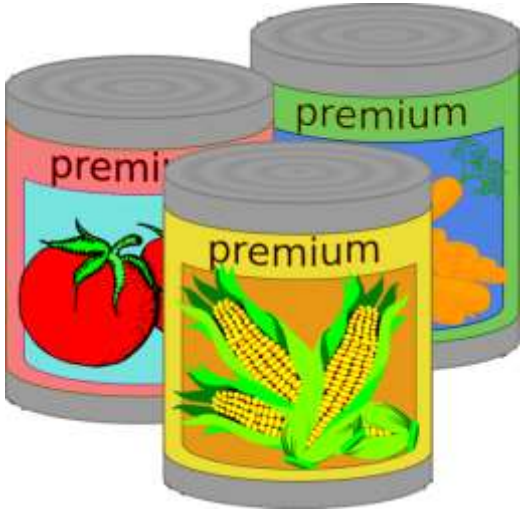


Storage and Security

- ✓ **Backup**
  - ✓ Procedure depend on circumstances, value of the data and levels of risk considered appropriate
  - ✓ 3 copies - 2 different media - 1 copy offsite
- ✓ **Protection**
  - ✓ Internal and external access policies
  - ✓ Reduce sensitivity (anonymise / pseudonymise, aggregate)
  - ✓ Passwords, encryption, firewalls, etc.
  - ✓ Secure transport and deletion!

Metadata: data about the data

- WHO created the data?
- WHAT is it?
- WHERE is it?
- WHEN were they created?
- HOW were they created?
- WHY were they created?

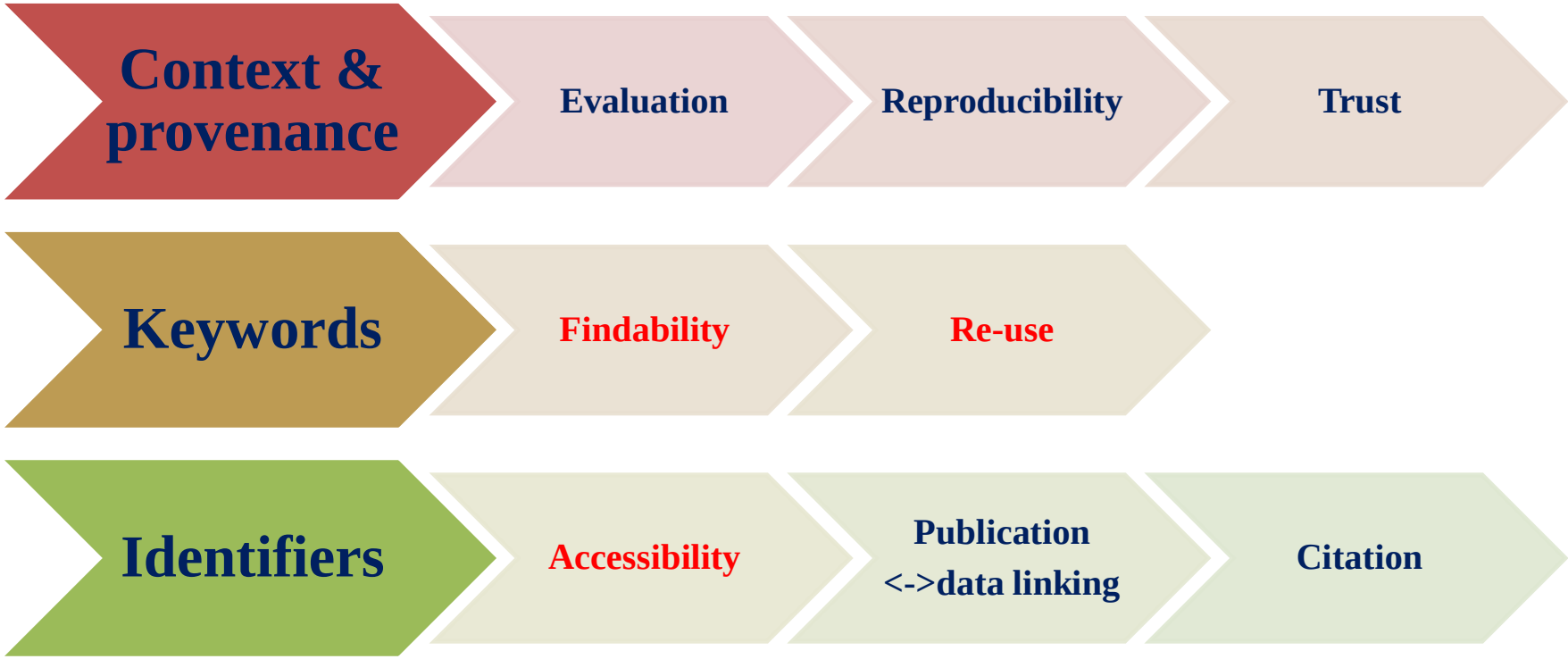


(Pixabay,CC0)

README file
https://researchdata.4tu.nl/en/publishing-research/uploading-data

Documentation

What do YOU need to find, understand en trust someone else's data?



With thanks to: www.cnrs.fr

Documentation

File name

What do we want to know from the file name?

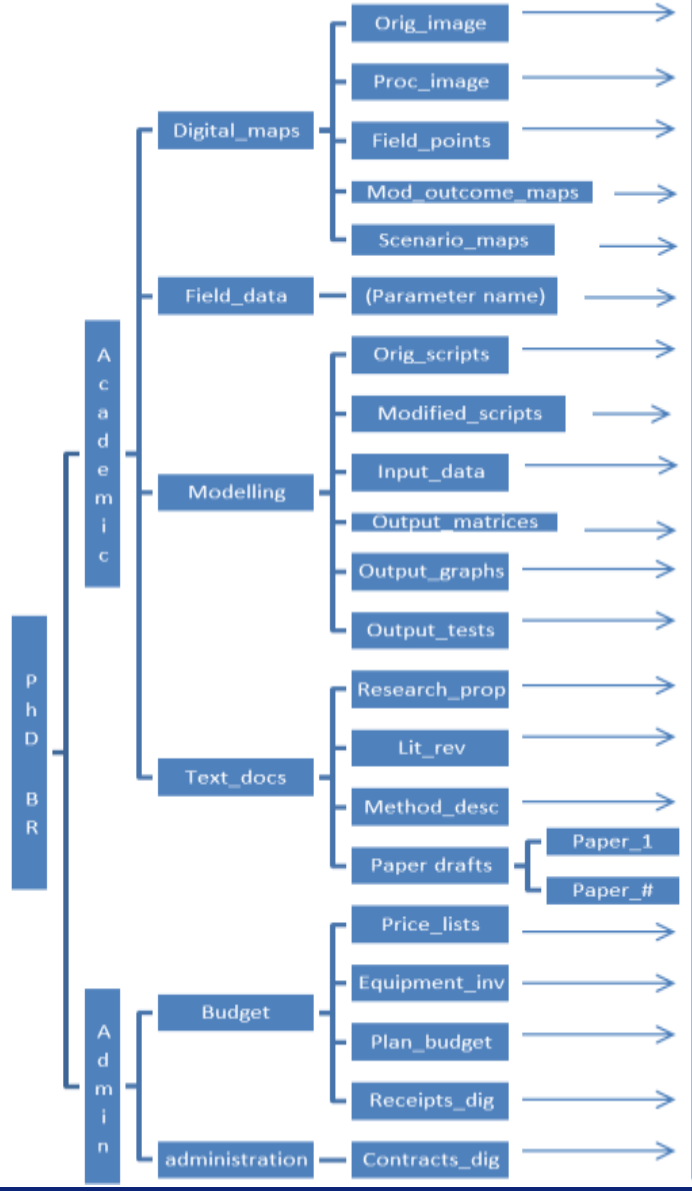
- Date (20140109)
- Project leader (WB)
- Project number (02)
- Procedure (su)
- Performer (RH)
- Type of animal (rat)
- Number of animal (04)
- Specific conditions (drug, dose, endpoint, type of measure, …)
- Other

→ We need a ‘universal’ language that everybody can understand

Example : **20131115_WB02_RHsu_rat01_CM-ST-CAN**

Documentation

Wageningen DMP Plan Example 1: The proposed folder arrangement and fil
naming strategy from the DMP by Beatriz Ramírez, Earth System Science
Research Group, Wageningen University
http://www.wageningenur.nl/nl/show/What-is-a-Data-Management-Plan.htm



Name	Example
Month_year_type	01_1989_landsat
Month_year_type_process	01_1989_landsat_orthorect
Field_points_date of actualization	Field_points_11_09_13
Month_year_outcome type	01_1989_wateryield
Year_parameter	2050_wateryield
Parameter_date of actualization	Rainfall_11_09_2013
Model_script_version	Shetran_script_v2.5
Model_script_modification_trial#	Shetran_script_cloudwater_t1
Parameter_col#_row#_trial#	Rainfall_4_250_t1
Est. parameter_col#_row#_trial#	ET_4_250_t1
Graph_x_y_trial#	Scatterplot_rainf_ET_t1
Test_parameter_trial#	Senstest_rainfall_t1
RP_trial# & _revised by	RP_t1_revLG
Topic_subtopic	Modelling_wateryields
Topic_method	Soil inf_infiltrometer
Paper code_date & revised by	TMCFvsPas_11_09_2014_LG
Company_date	Eijkelkamp_11_09_2013
Equipment list	Equipment_list
Budget_date	Budget_11_09_2013
Item_rec_date (month_year)	Mercado_rec__01_2014
Item_con_date	Houserent_con_01_2014

Data access and sharing

	Would not share with anyone	Would share with my immediate collaborators	Would share with others in my research centre or at my institution	Would share with scientists in my field	Would share with scientists outside of my field	Would share with anyone
Immediately after the data has been generated						
After the data has been normalized and/or corrected for errors						
After the data has been processed for analysis						
After the data has been analysed						
Immediately before publication						
Immediately after the findings derived from this data have been published						

Based on: Interview worksheet, Jake Carlson, Purdue University Libraries / Distributed Data Curation Center

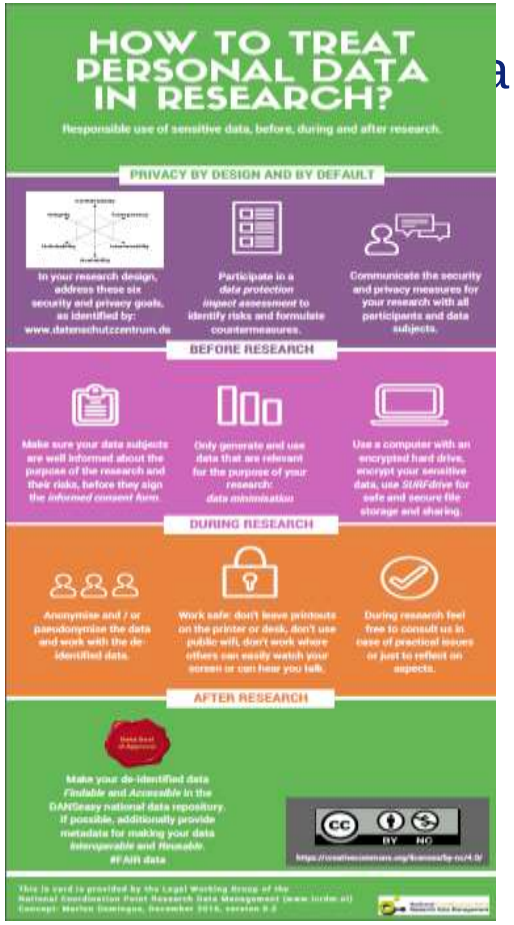
Data access and sharing

Personal data?

New European Regulation from May 25 2018!

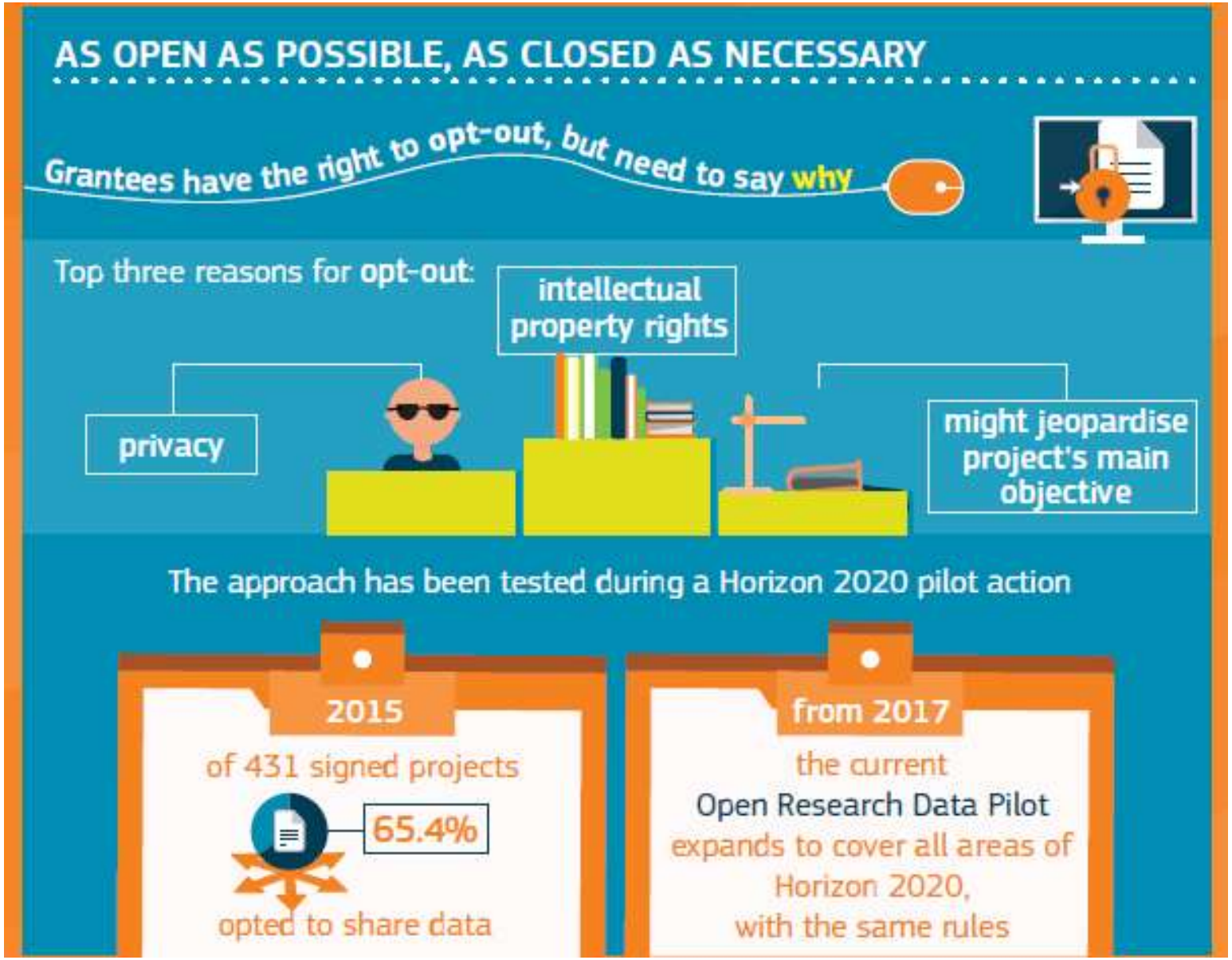
General Data Protection Regulation

University needs to register all processing of personal



https://www.staff.universiteitleiden.nl/ict/privacy-and-data-protection/general-data-protection-regulation-gdpr/general-data-protection-regulation-gdpr/service-units/leiden-university-libraries?cf=service-units&cd=leiden-university-libraries

Data access and sharing



Consent form

Do's:

- Mention the possibility of sharing in the consent form
- State conditions under which access may be granted
- Explain obligation to protect confidentiality
- Indicate how data will be anonymized
- State retention period of identifiable information
- Document consent for subsequent users of the data

Don'ts:

- Avoid terms as 'fully anonymous' or strictly confidential
- Promises to destroy data unnecessarily
- Mention expiration time period for the consent
- Promises that data will only be accessed by research team

From:
https://www.google.nl/search?q=do's+and+don'ts+informed+consent+for+sharing+data&sourceid=ie7&rls=com.microsoft:nl-NL:IE-Address&ie=&oe=&gfe_rd=cr&dcr=0&ei=zGy6WZqOCYPc8AfoyYnoDA&gws_rd=ssl

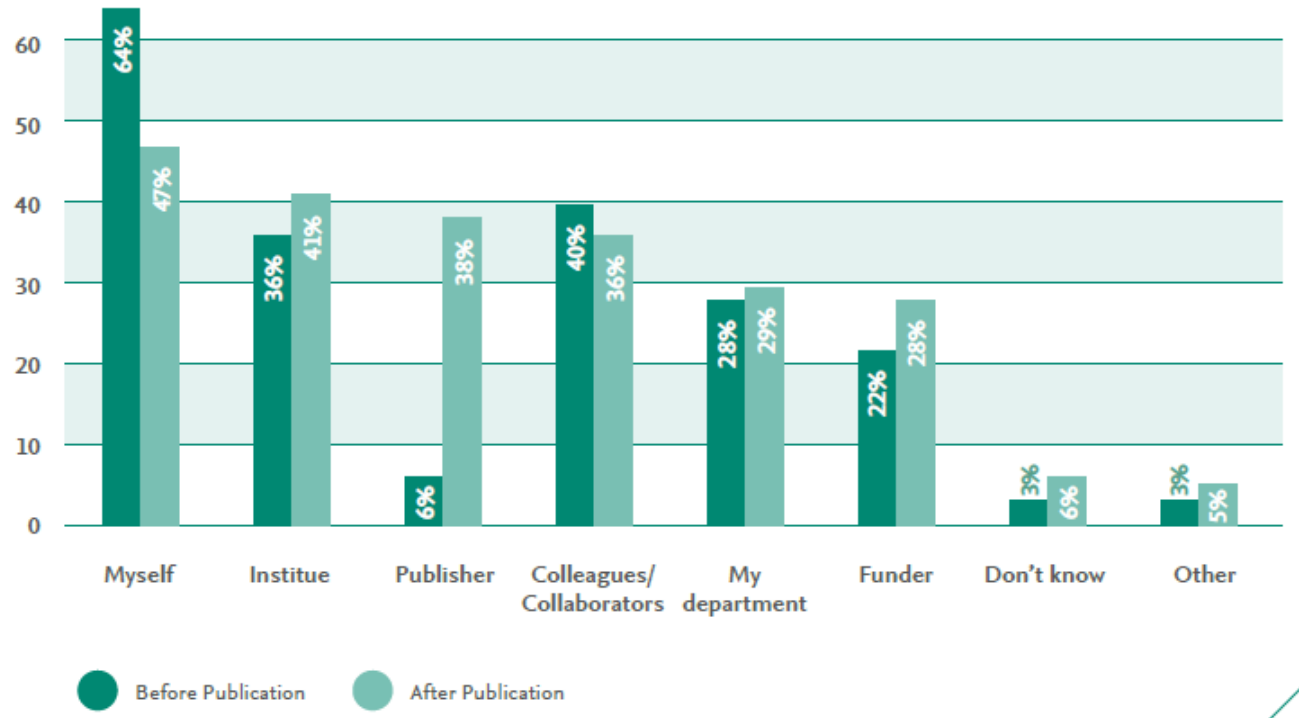
Data access and sharing

Open Data: The Researcher Perspective

Report by CWTS, Elsevier and University Leiden, April 2017

https://www.cwts.nl/news?article=n-r2q244&title=73-of-academics-say-access-to-research-data-helps-them-in-their-work-34-do-not-publish-their-datail 2017

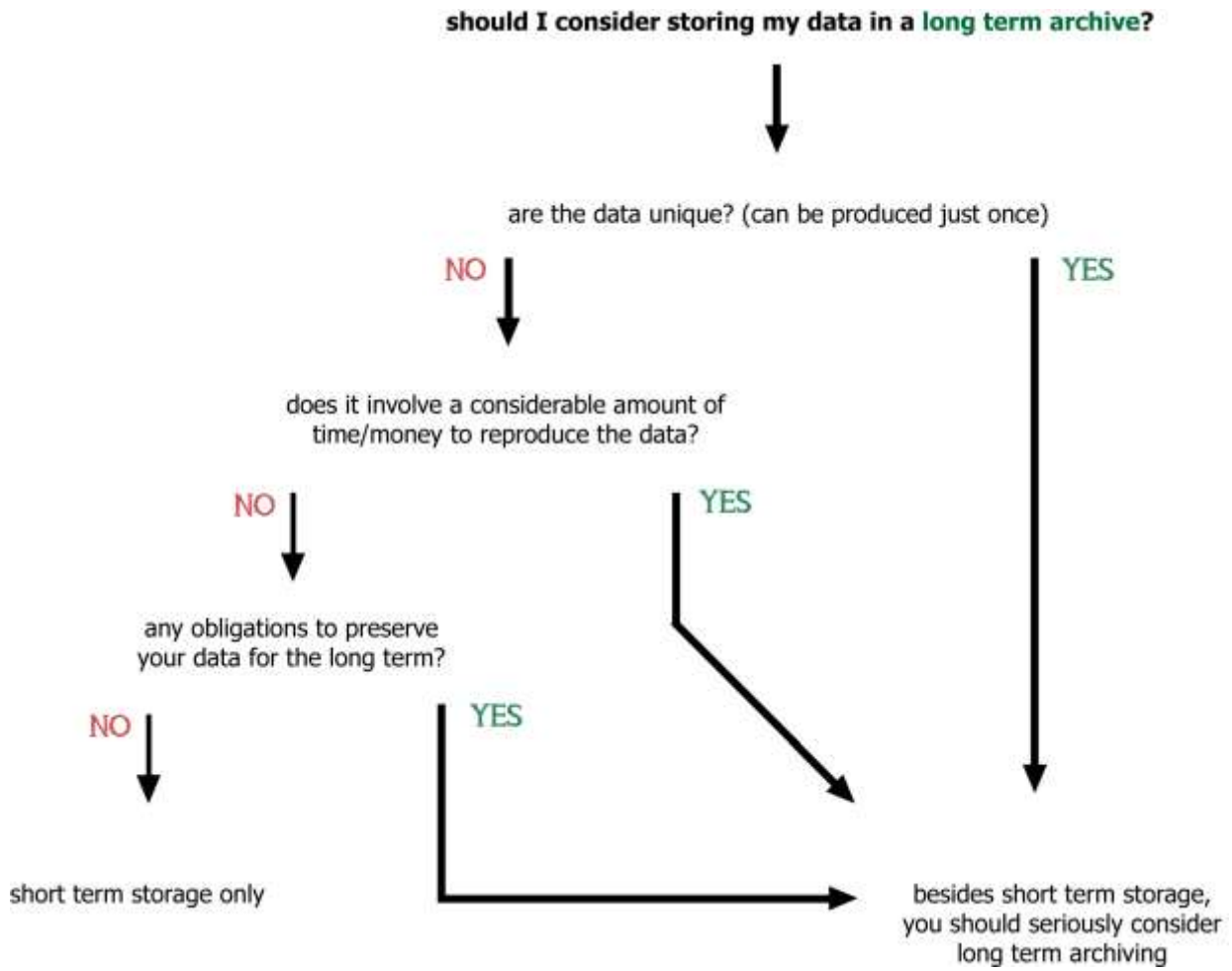
Figure 3. Research data ownership before and after publication (%, n=1162)



38 % thinks the publisher is the owner of the data after publication!

Data preservation

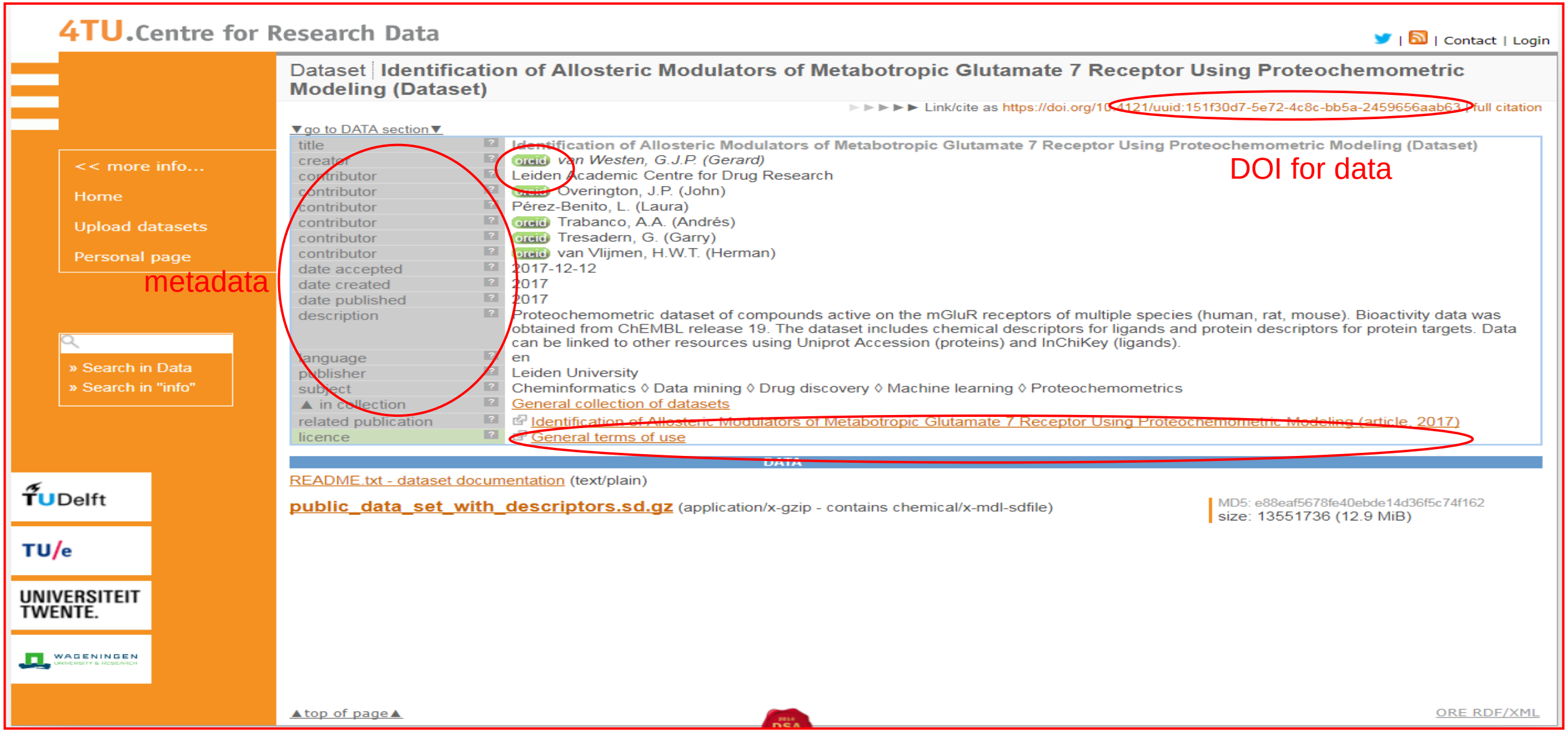
University of Wageningen : http://www.wageningenur.nl/nl/show/What-is-a-Data-Management-Plan.htm



https://repositoryfinder.datacite.org/

https://www.re3data.org/

https://www.coretrustseal.org/why-certification/certified-repositories/



https://www.library.universiteitleiden.nl/research-and-publishing/scholarly-publishing/manage-your-name-with-identifiers

7

INSTITUTIONAL APPROACH: LEIDEN EXAMPLE

Faculties

- Archaeology
- Governance and Global Affairs
- Humanities
- Law
- Medicine/LUMC
- Science
- Social and Behavioural Sciences

Universiteit Leiden
The Netherlands

Research

400

PhDs

600

Professors

40

Annual NWO subsidies

5-10

Annual ERC subsidies

RDM activities @ Leiden

Aim:

Researchers from Leiden University can manage their data in such a way that they are findable, accessible and reusable (FAIR) according to current law and regulation

Good data management is essential for university's core processes:

- Research quality
- Funding (consortia, trust)
- Valorisation, impact (Open Science, FAIR)

A roadmap for cultural change

1. Leiden Regulation for Data Management Adopted in April 2016
2. Faculties elaborate own protocols
3. Three year program to create facilities and services
4. Implementation period until end 2019
5. Training, information and advice

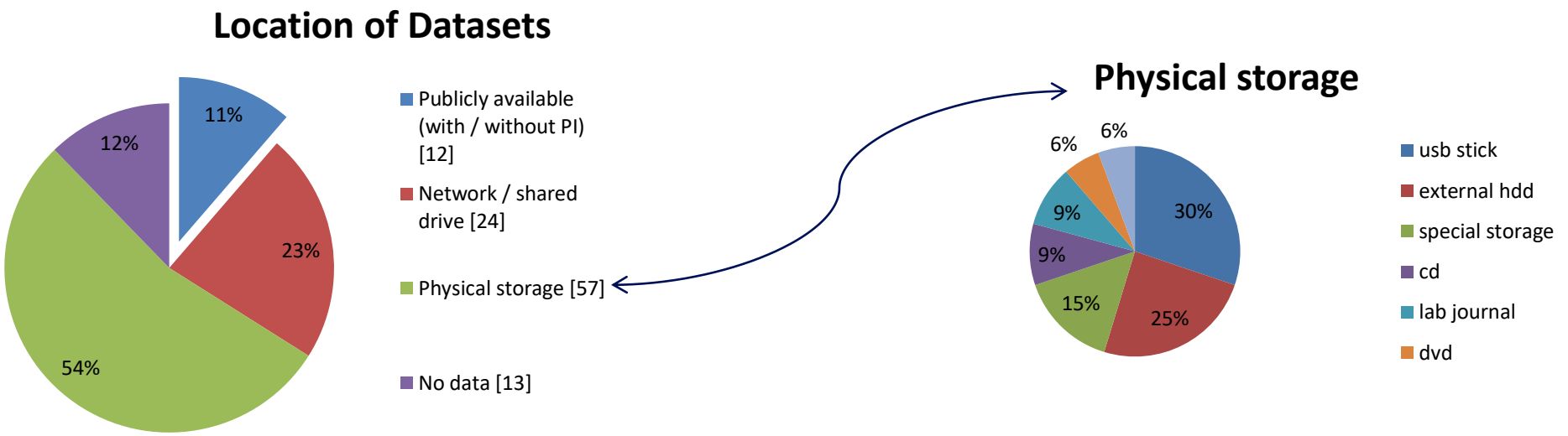
Faculty of Science protocol: data forms

PhD are obliged to state in Data Form location of research data before defence. Form signed by supervisor and scientific director

May – Dec. 2016: 102 data forms

- ✓ All data underlying theses now available to institute
- ✓ Insight in volume / types datasets
- ✓ Insight in sharing practices
- x Most of the data not ‘FAIR’ yet
- x Managing data on physical storage / shared network drives is problematic

Data forms: location of data



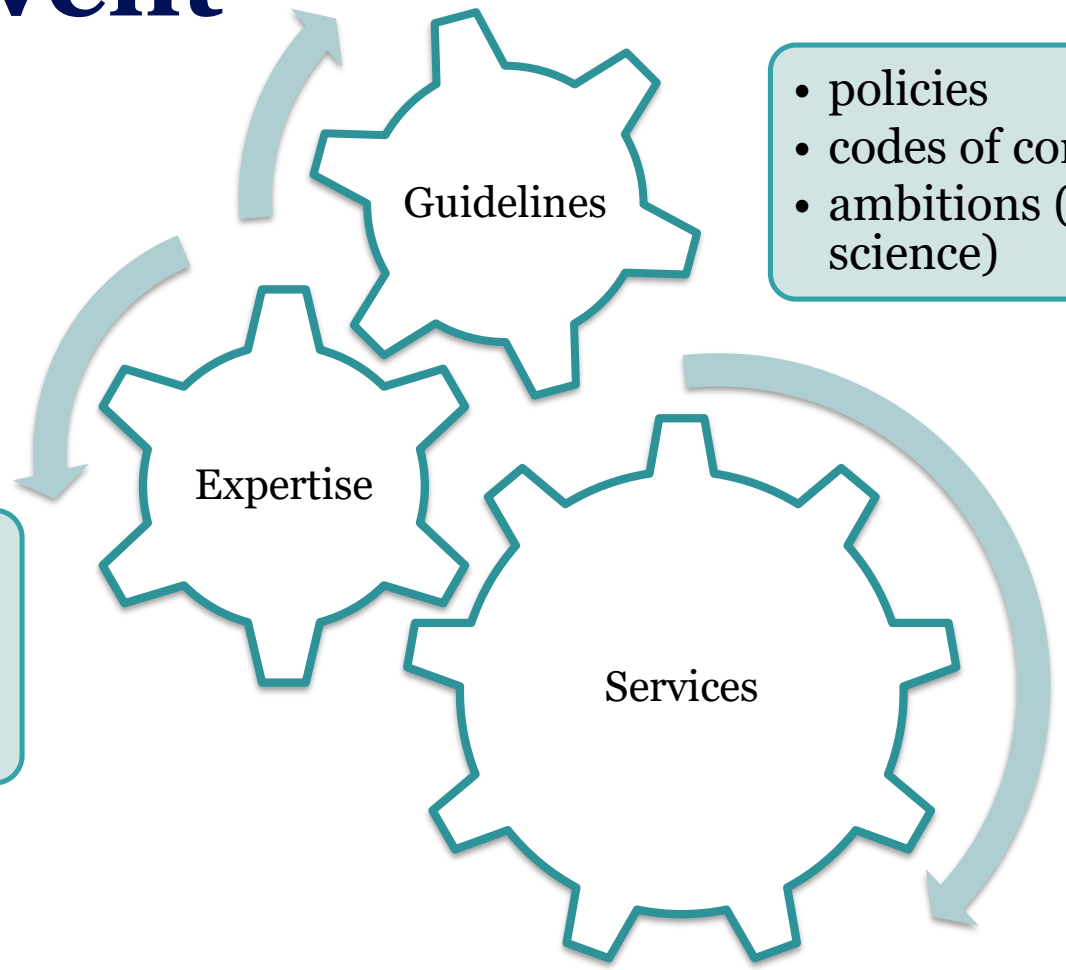
Reasons for not making data publicly available:

- Not all projects create research data
- Commercial interests / contractual obligations / third party data
- Unfamiliarity with repositories / archives

Baseline?

Might help to evaluate effect of our efforts to provide tailor-made information and (new) facilities in the next years…

“transition to Open Science is a process, not a single event”*



- policies
- codes of conduct
- ambitions (open science)

- Research ICT
- Data Management Support
- Data skills

- RDM Program for DMP tool, data set registration, storage, fault, archive
- One-stop-shop for questions
- Training

* https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf

Next steps: RDM is a means, not a goal

- **Working towards Leiden Research Support**
- **Fairification workshops**
- **Promote data re use**

  (advanced data management)

Thank you!

Universiteit Leiden
The Netherlands