panosc
photon and neutron
open science cloud

ESRF

7th EIROforum School
on Instrumentation

EIRO forum
serving european science

7-11 June 2021 | Virtual event | Register by 30 May 2021

What you need to know about Research Data Management

9th June, 2021

Author: Andy Götz (ESRF)

Role: ESRF Data Policy implementor and PaNOSC coordinator

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Outline of Talk

This talk will address the topic of Research Data Management for scientists doing research in order to answer the following questions:

- What is Research Data Management (RDM)?

Why do this ?

What to know ?

What to do ?

What to expect ?

What to try ?

What to learn ?

My background in research data management

- **Radio astronomer** many years ago
- **Co-author** of the **ESRF data policy**
- Leading the ESRF data policy implementation
- **Coordinator** of the [PaNOSC](https://panosc.eu) project for making data from photon and neutron sources FAIR https://panosc.eu

Kilobytes

to

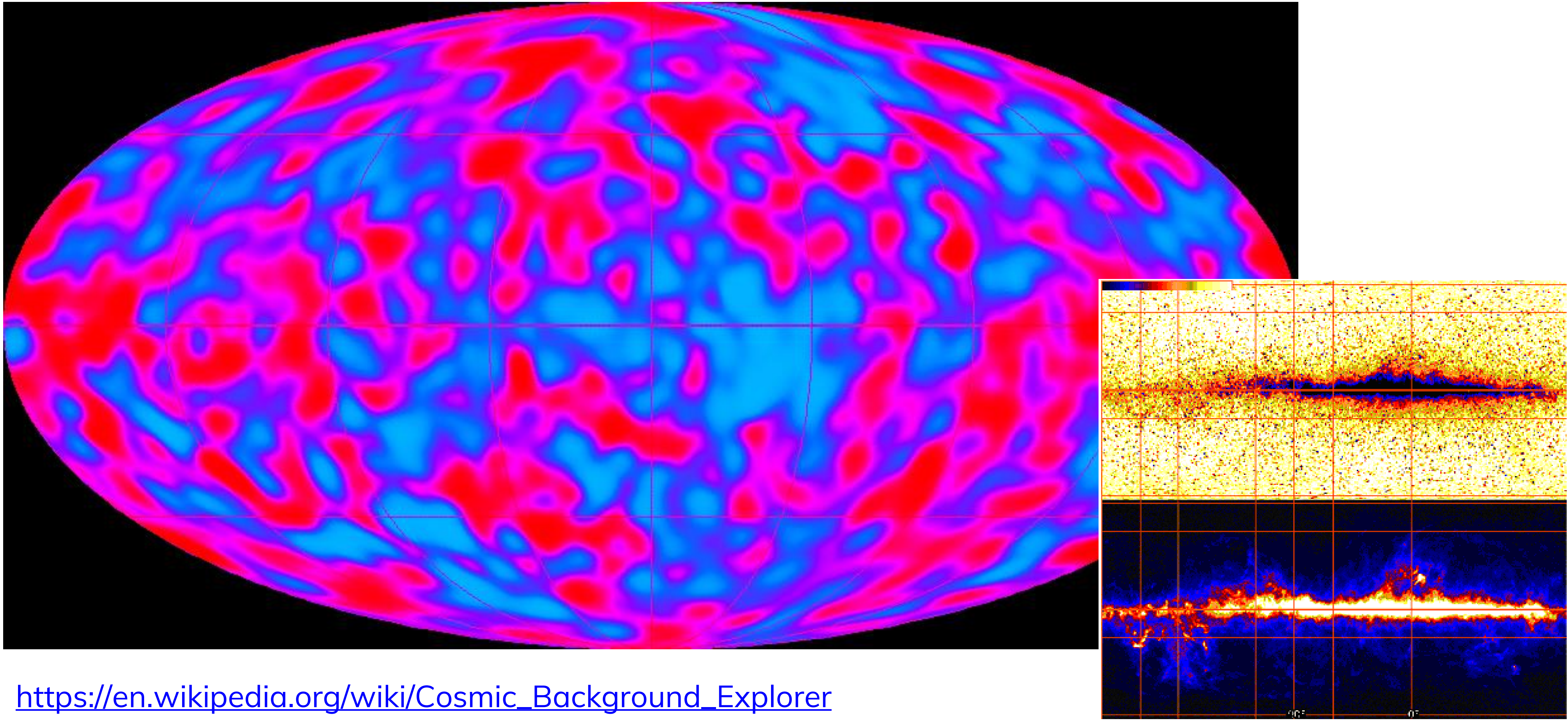
Petabytes

in

40 years

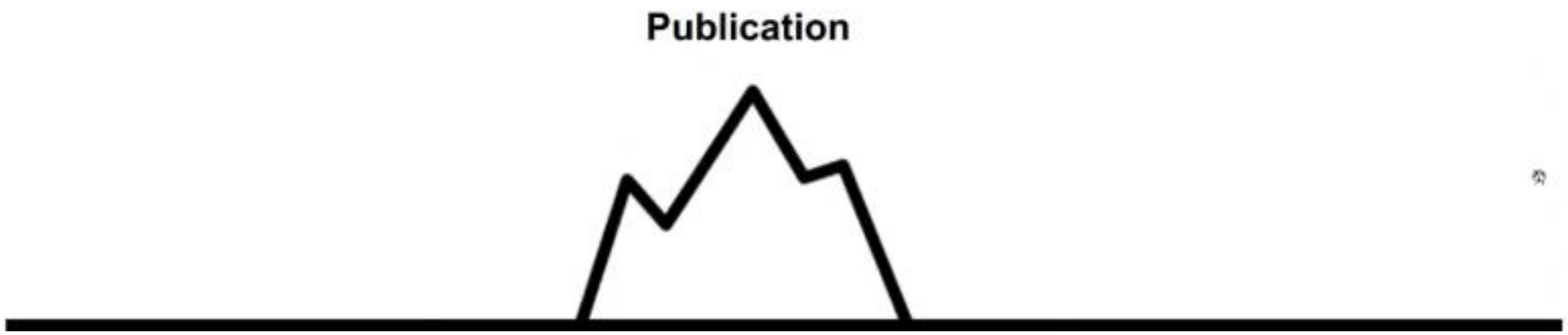
panosc

CMB anisotropy map formed from data taken by the COBE spacecraft



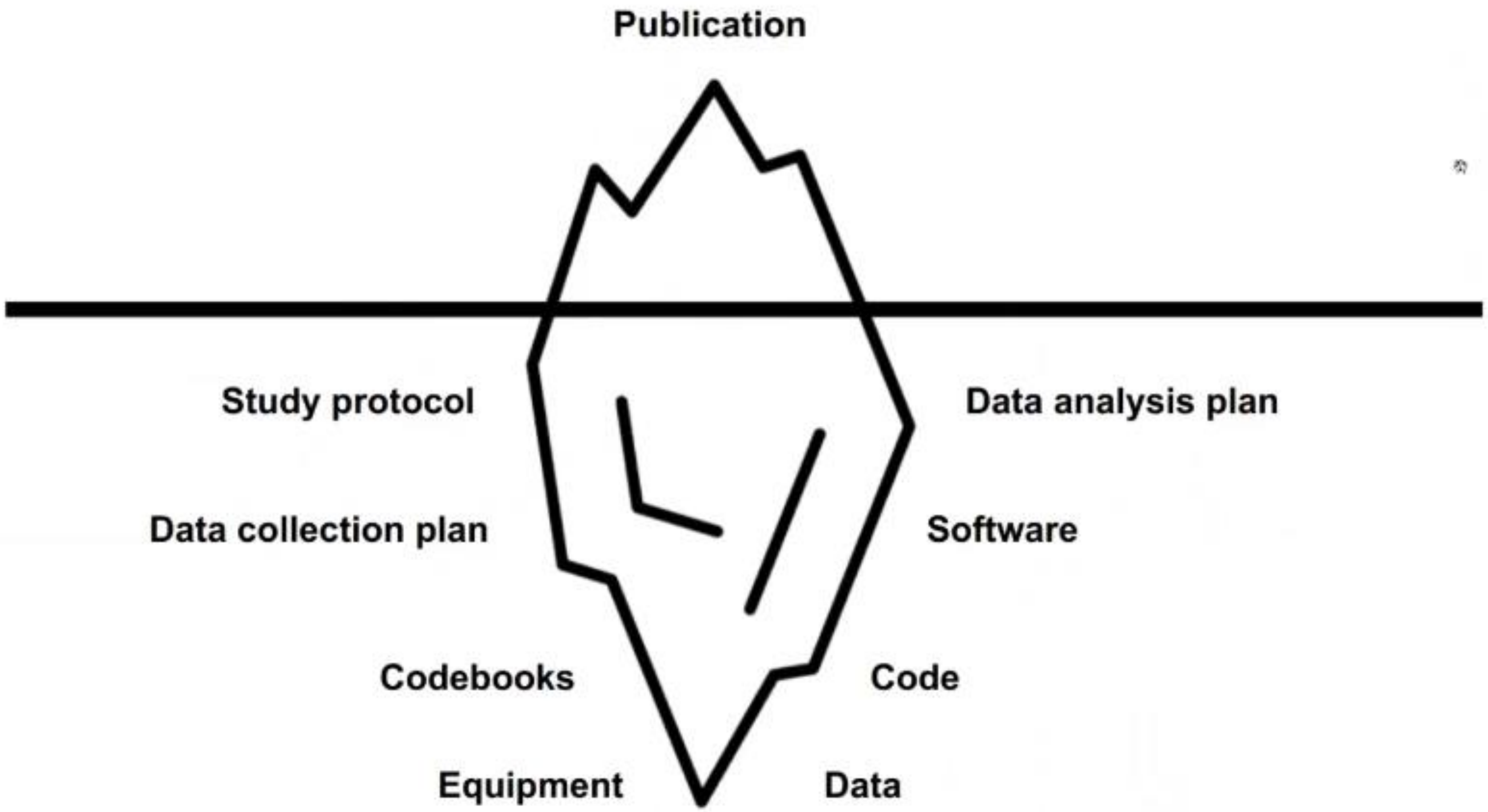
https://en.wikipedia.org/wiki/Cosmic_Background_Explorer

Science produces Publications



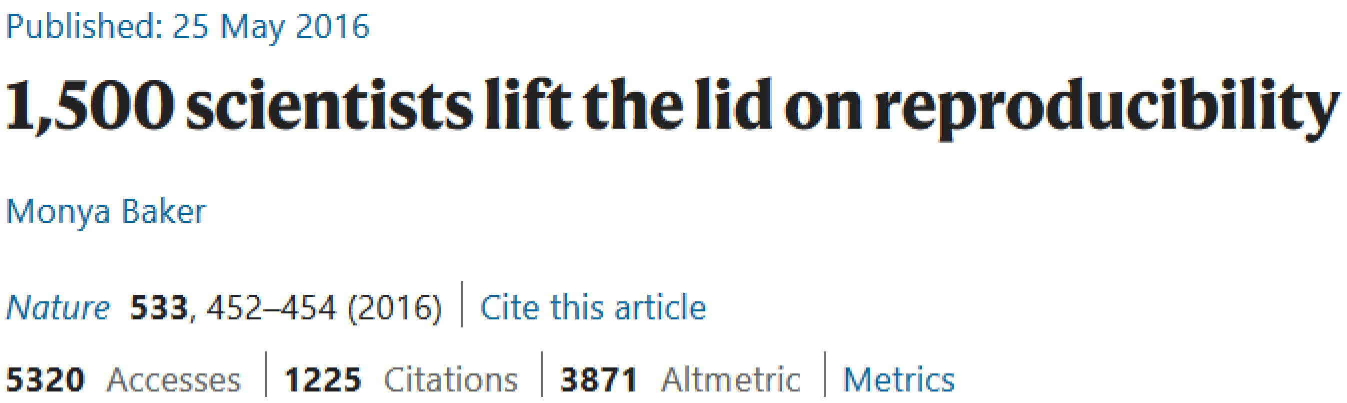
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Science produces much more than Publications



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825092

Reproducibility and Replicability



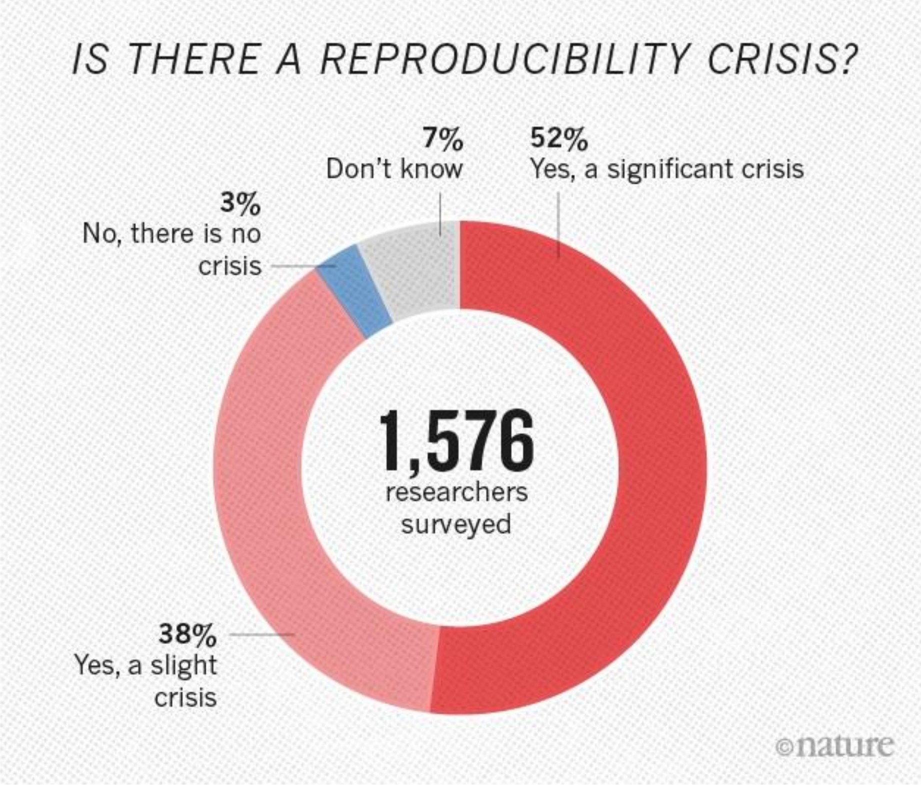
Published: 25 May 2016

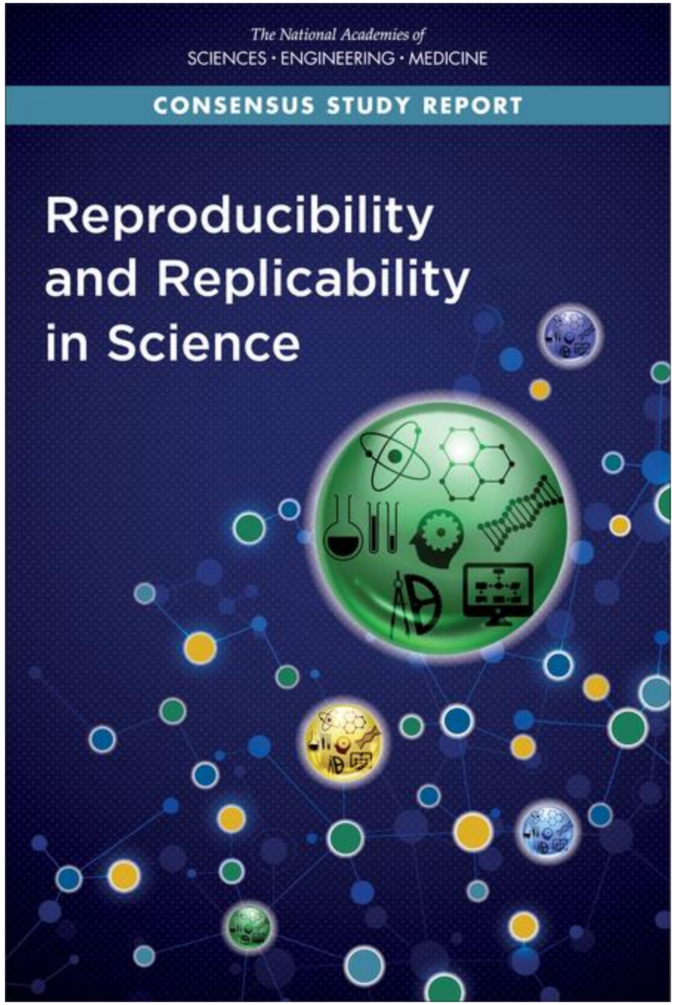
1,500 scientists lift the lid on reproducibility

Monya Baker

Nature **533**, 452–454 (2016) | Cite this article

5320 Accesses | **1225** Citations | **3871** Altmetric | Metrics




The National Academies of
SCIENCES · ENGINEERING · MEDICINE
CONSENSUS STUDY REPORT
Reproducibility and Replicability in Science

Further reading:

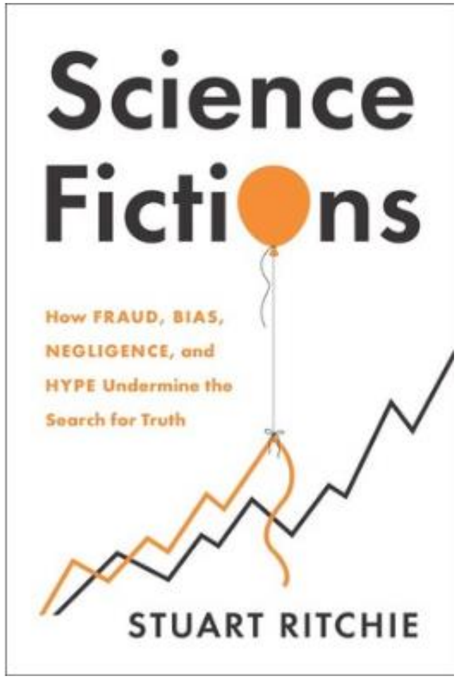
- Replication crisis – Wikipedia
- https://phys.org/news/2017-03-science-crisis.html

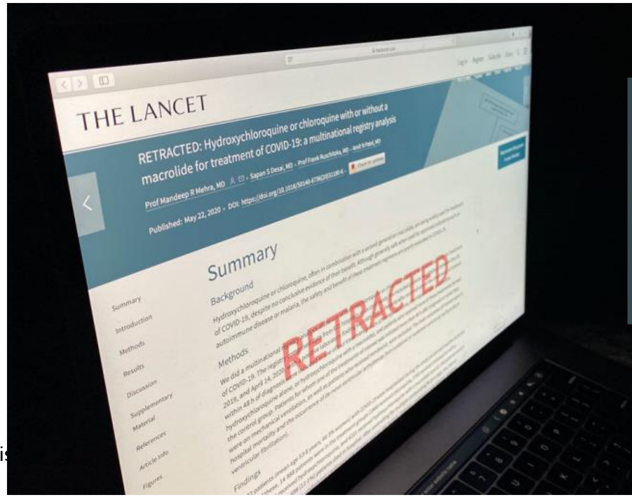
arch and innovation programme under grant agreement No. 823852

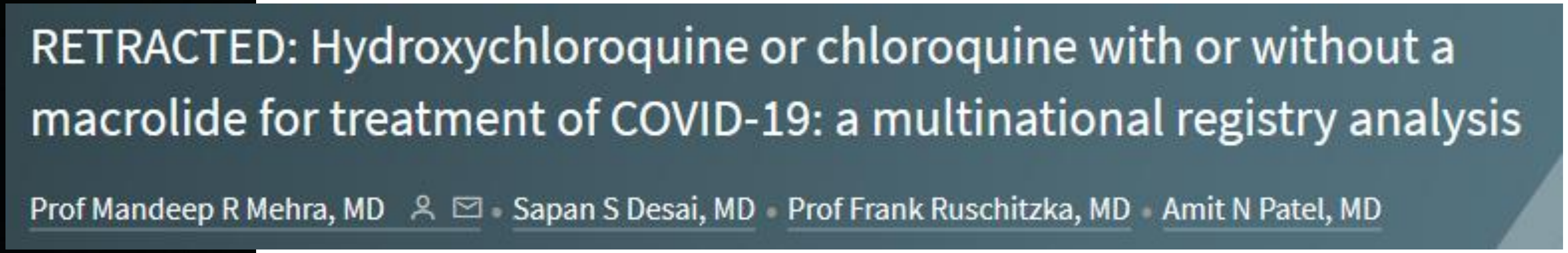
European Conduct of Scientific Integrity

Integrity, scientific method, open science

- Recommend to follow the EU Code of Integrity
  - https://allea.org/code-of-conduct/
- To AVOID having your papers RETRACTED
  - https://retractionwatch.com/


Science Fictions
How FRAUD, BIAS, NEGLIGENCE, and HYPE Undermine the Search for Truth
STUART RITCHIE


THE LANCET
RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis
Summary
RETRACTED

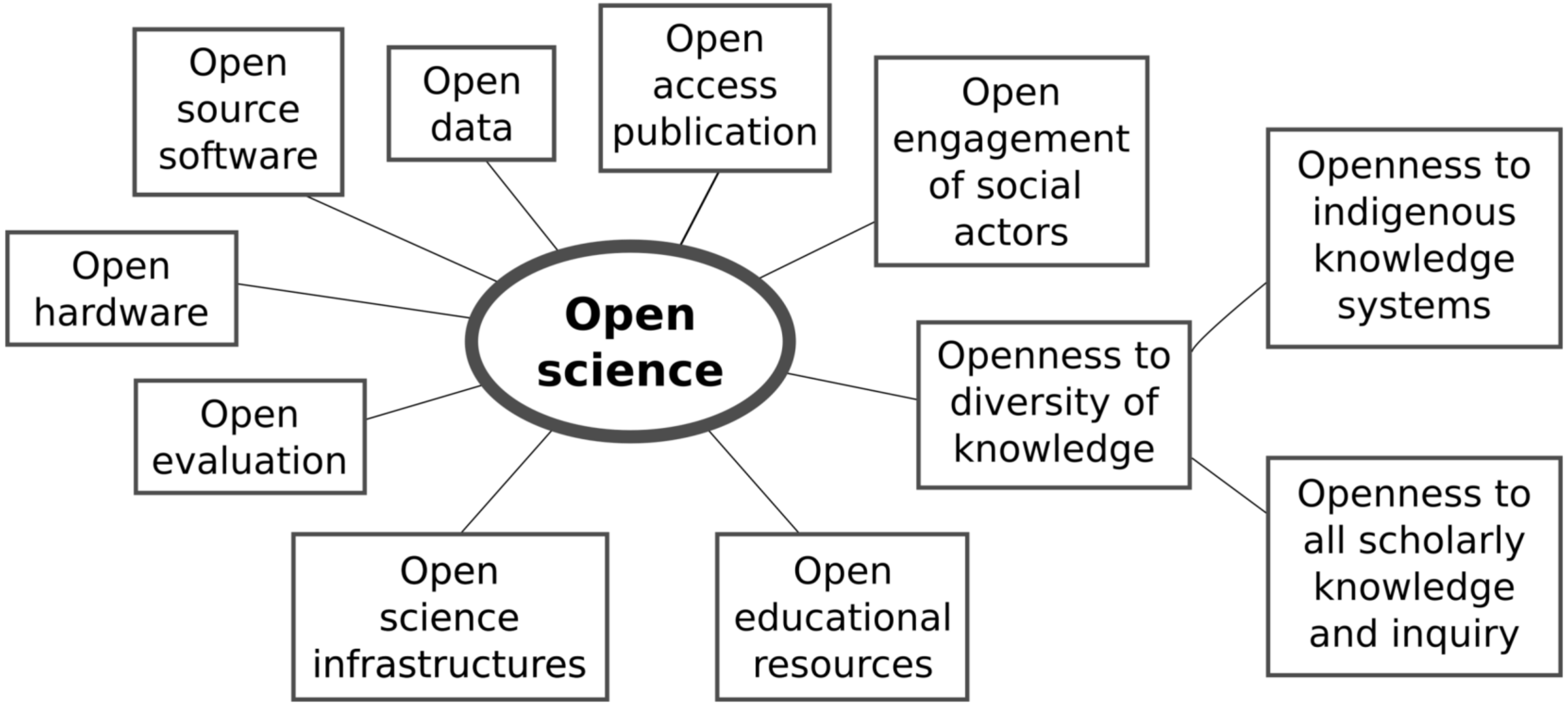

RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis

Prof Mandeep R Mehra, MD • Sapan S Desai, MD • Prof Frank Ruschitzka, MD • Amit N Patel, MD

This ... vation programme under grant agreement No. 823852

Open Science - origin

"Open Science can be seen as a continuation of, rather than a revolution in, practices begun in the 17th century with the advent of the academic journal, when the societal demand for access to scientific knowledge reached a point at which it became necessary for groups of scientists to share resources with each other" - https://en.wikipedia.org/wiki/Open_science

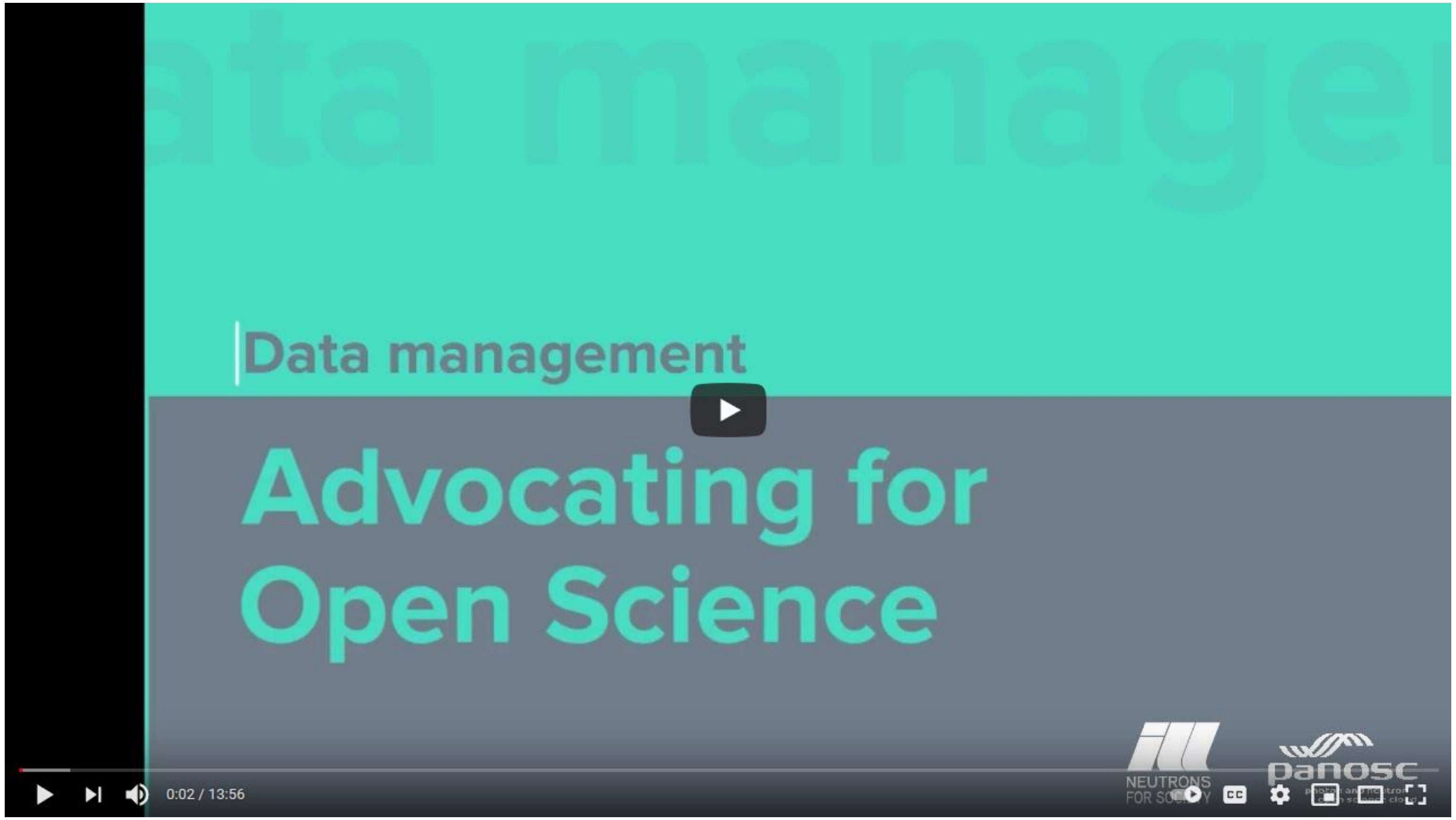

Open science
Open source software
Open data
Open access publication
Open engagement of social actors
Openness to indigenous knowledge systems
Open hardware
Openness to diversity of knowledge
Open evaluation
Openness to all scholarly knowledge and inquiry
Open science infrastructures
Open educational resources

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

By RobbieIanMorrison - Own work, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=100144897

Advocating for Open Science

Watch this interview of Petr Čermák, a strong advocate of open on the advantages of Open Science for neutrons and science in general



https://youtu.be/QKAc1y6HZNk

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Further reading – Open Science

Many resources are available on Open Science, here are some used for this talk

- Phys.org
  - Five questions about open science answered
  - Data sharing can offer help in science's reproducibility crisis
- UNESCO
  - Recommendation on Open Science

Open Science, open issues

Sarita Albagli
Maria Lucia Maciel
Alexandre Hannud Abdo
Editors

ibict

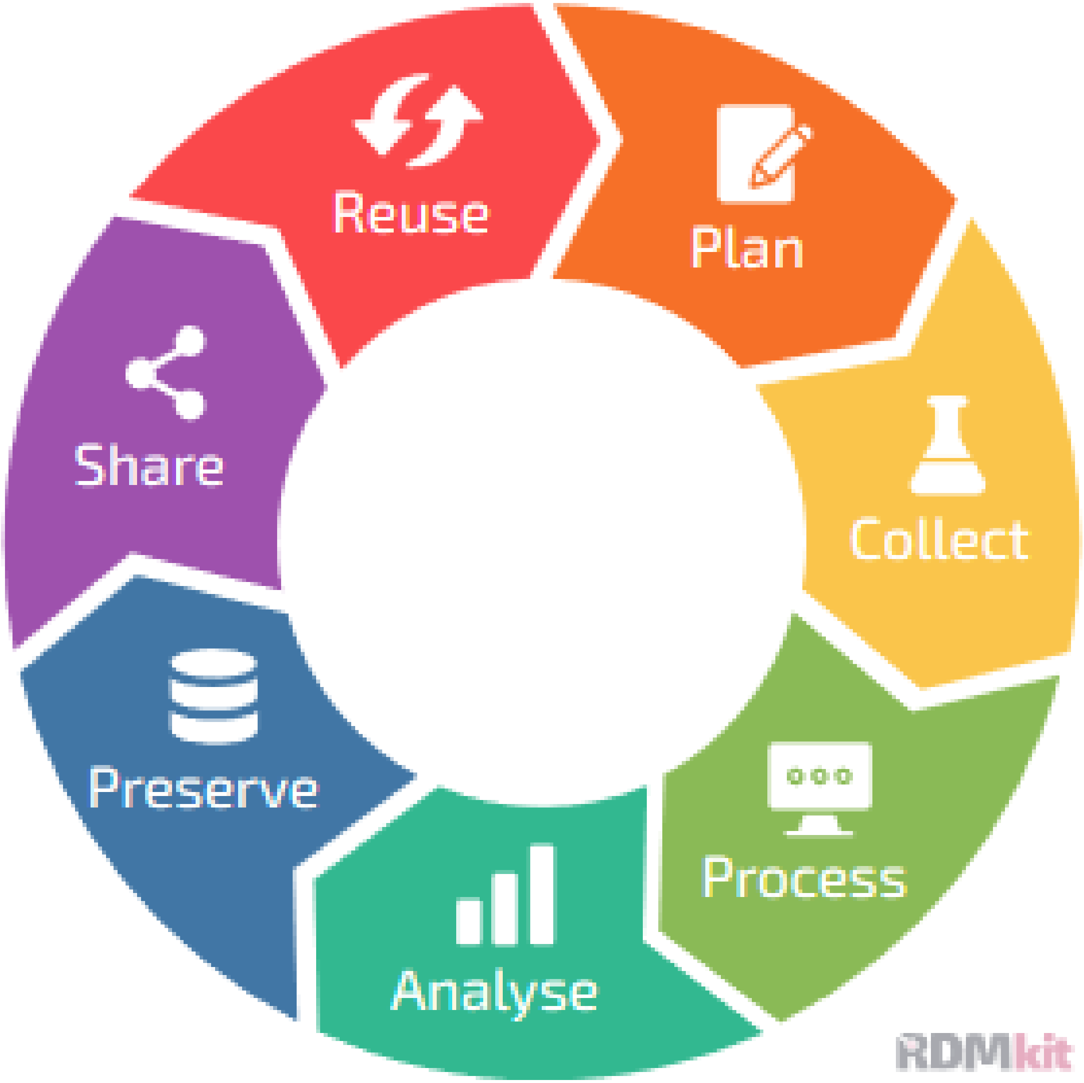
UNIRIO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

What is Research Data Management (RDM)?

RDM is the **collection of practices** required to **plan, collect, process, analyse, preserve, share** and make **data re-usable.**

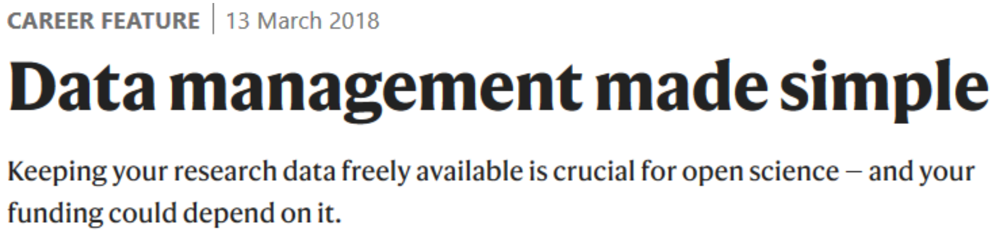
- Data management planning
- Collecting raw data and metadata
- Processing to produce new data
- Analysing data to produce results
- Curating data for the long term
- Sharing data and making it Findable, Accessible, Interoperable and Reusable



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Data management made simple

Quirin Schiermeier in Nature (2018)



CAREER FEATURE | 13 March 2018

Data management made simple

Keeping your research data freely available is crucial for open science – and your funding could depend on it.

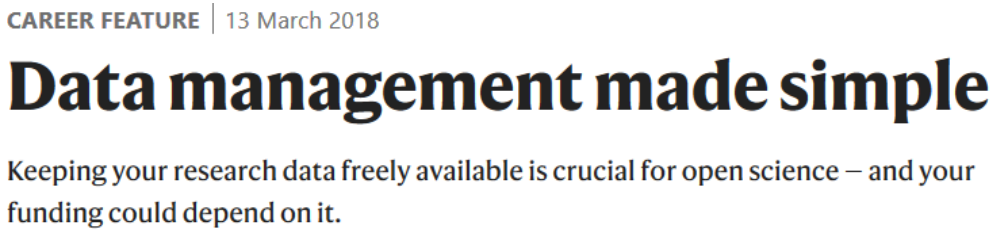
https://doi.org/10.1038/d41586-018-03071-1

1. Check the research-data requirements of your funding agency and field of research.
2. Go online for help in developing a data-management plan. A useful guide outlining UK funder expectations can be found at go.nature.com/2tnohla.
3. List the various types of data and research outputs that you expect to produce.
4. Decide what data and research materials require archiving and determine how much storage space you will need.
5. Define appropriate data file formats (see https://fairsharing.org/ for formats).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Data management made simple

Quirin Schiermeier in Nature (2018)



CAREER FEATURE | 13 March 2018

Data management made simple

Keeping your research data freely available is crucial for open science – and your funding could depend on it.

https://doi.org/10.1038/d41586-018-03071-1

6. Look for data repositories used by your research community or your host institution (see www.re3data.org for examples).
7. Check what data format and structure the chosen archive might request.
8. Provide metadata that allows others to understand, cite and reuse your data files.
9. Make clear how and when your data can be shared with scientists outside your group.
10. If your research involves sensitive data, explain any legal and ethical restrictions on data access and reuse.
11. Assign responsibility for long-term data curation to a suitable office.
12. Revisit your plan frequently and update it if necessary.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Data policies

1. Check the research-data requirements of your funding agency and field of research.

A Data policy defines the rules of access and usage to the data produced. Research Institutes like the EIROforum ones all have data policies in place now.

- You are required to accept the data policy when requesting access
- **Data is not considered as property but has a usage licence**
- Data are under **embargo** (varying from 1 yr, 3 yr, 5 yr) for use by the original creators for a limited amount of time **before being made open.**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

EIROforum member Data Policies

- **CERN** – open data policy for LHC (since 2020)
- **EMBL** – open access policy (since 2015)
- **ESA** – open data policy for most data (since 2010)
- **ESO** – open data policy (updated in 2016)
- **ESRF** – open data policy (since 2015)
- **EUROfusion** – proposal for open data policy (in progress since 2018)
- **EuXFEL** – open data policy (since 2017)
- **ILL** – open data policy (since 2012)
- Others
  - **CERIC-ERIC** – open data policy (since 2021)
  - ....

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

CERN announces new open data policy in support of open science

A new open data policy for scientific experiments at the Large Hadron Collider (LHC) will make scientific research more reproducible, accessible, and collaborative

11 December 2020.

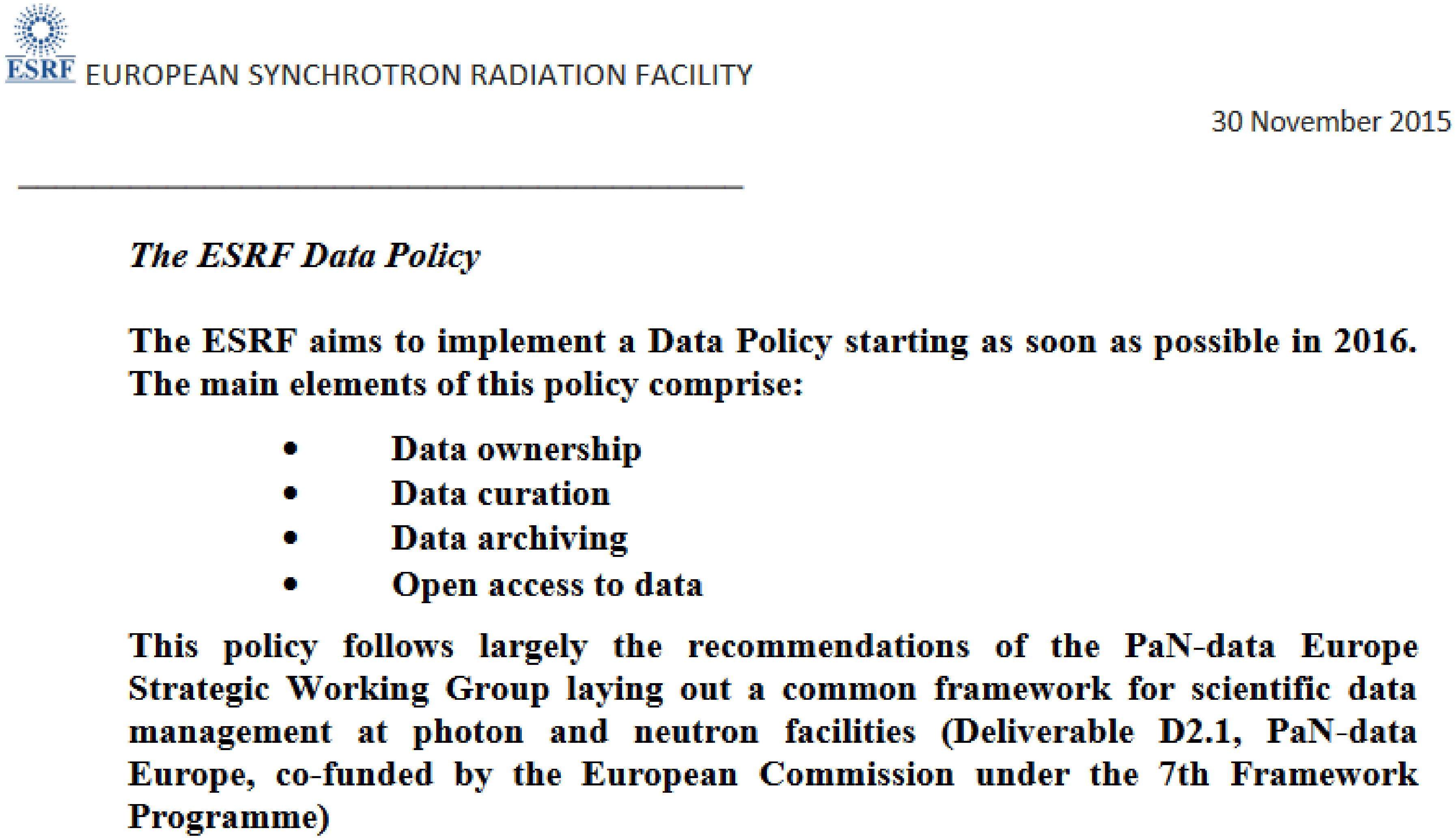
- *The four main LHC collaborations (ALICE, ATLAS, CMS and LHCb) have unanimously endorsed a new* ***open data policy*** *for scientific experiments at the* ***Large Hadron Collider*** *(LHC), which was presented to the CERN Council today. The policy commits to* ***publicly releasing*** *so-called* ***level 3 scientific data****, the type required to make scientific studies, collected by the LHC experiments. Data will start to be released approximately* ***five years after collection****, and the aim is for the* ***full dataset*** *to be publicly available by the close of the experiment concerned. The policy addresses the growing movement of* ***open science****, which* ***aims to make scientific research more reproducible, accessible, and collaborative****.*

https://home.cern/news/press-release/knowledge-sharing/cern-announces-new-open-data-policy-support-open-science

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

ESRF Data Policy

https://www.esrf.fr/datapolicy



ESRF EUROPEAN SYNCHROTRON RADIATION FACILITY

30 November 2015

The ESRF Data Policy

The ESRF aims to implement a Data Policy starting as soon as possible in 2016. The main elements of this policy comprise:

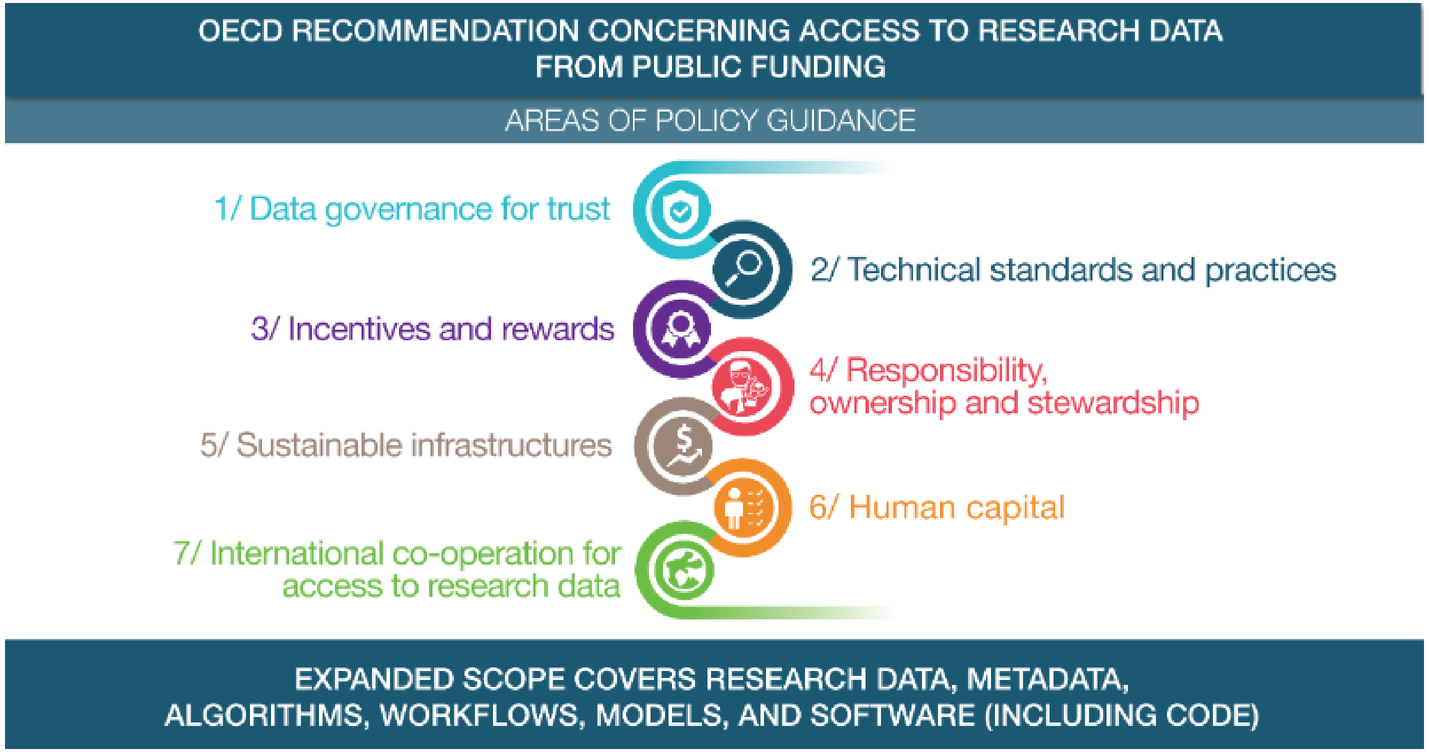
- **Data ownership**
- **Data curation**
- **Data archiving**
- **Open access to data**

This policy follows largely the recommendations of the PaN-data Europe Strategic Working Group laying out a common framework for scientific data management at photon and neutron facilities (Deliverable D2.1, PaN-data Europe, co-funded by the European Commission under the 7th Framework Programme)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Open data for publicly funded research

- The **OECD** recommendation in **2006** had a big impact on data policies
- The recommendation was updated in **2021** (https://www.oecd.org/sti/recommendation-access-to-research-data-from-public-funding.htm)



OECD RECOMMENDATION CONCERNING ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

AREAS OF POLICY GUIDANCE

1/ Data governance for trust
2/ Technical standards and practices
3/ Incentives and rewards
4/ Responsibility, ownership and stewardship
5/ Sustainable infrastructures
6/ Human capital
7/ International co-operation for access to research data

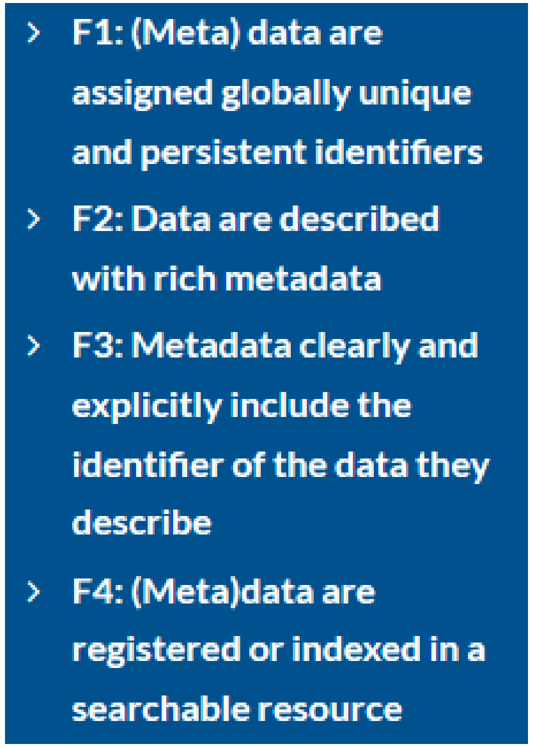
EXPANDED SCOPE COVERS RESEARCH DATA, METADATA, ALGORITHMS, WORKFLOWS, MODELS, AND SOFTWARE (INCLUDING CODE)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

FAIR Principles

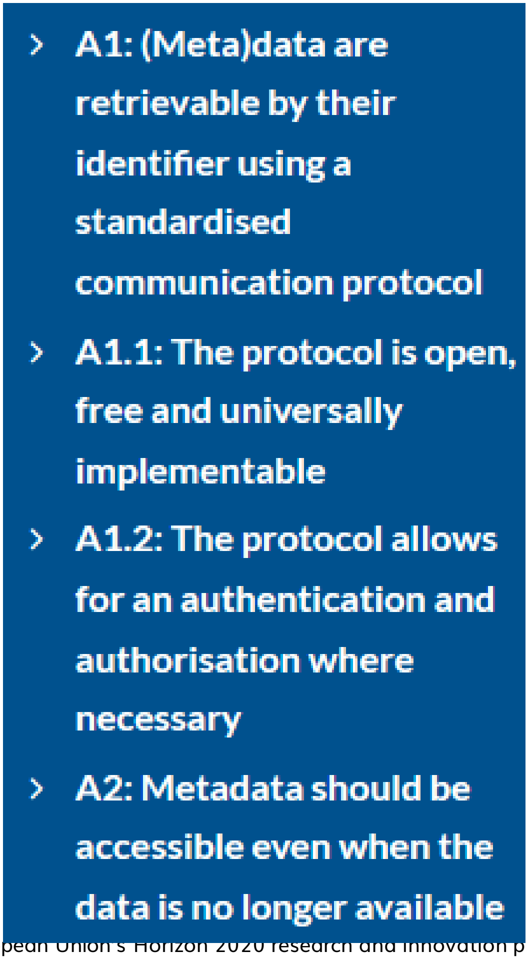
https://www.go-fair.org/fair-principles/

Findable



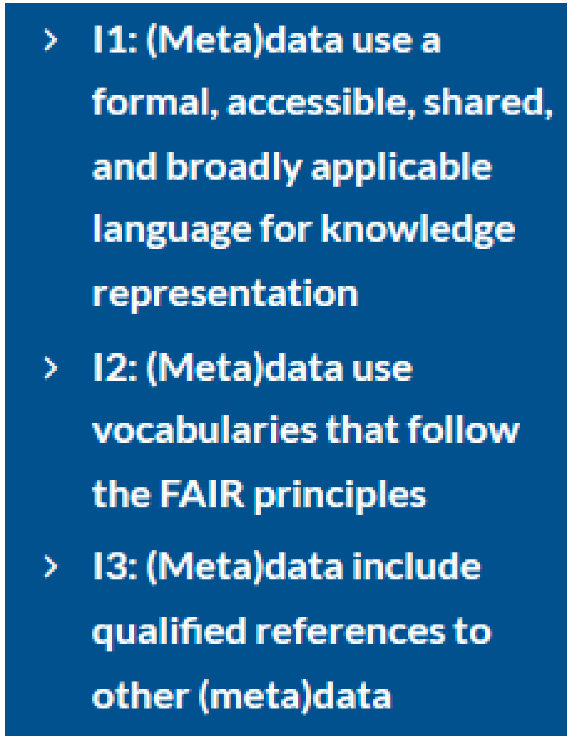
- F1: (Meta) data are assigned globally unique and persistent identifiers
- F2: Data are described with rich metadata
- F3: Metadata clearly and explicitly include the identifier of the data they describe
- F4: (Meta)data are registered or indexed in a searchable resource

Accessible



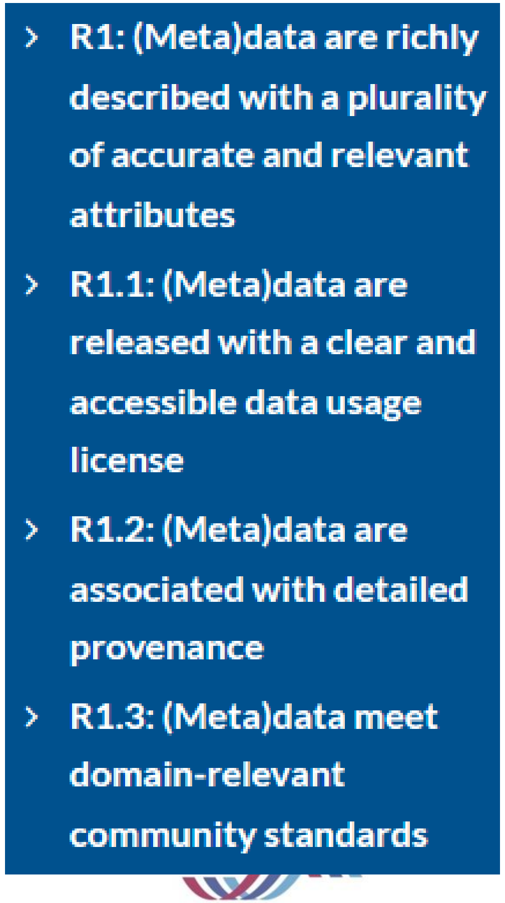
- A1: (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1: The protocol is open, free and universally implementable
- A1.2: The protocol allows for an authentication and authorisation where necessary
- A2: Metadata should be accessible even when the data is no longer available

Interoperable



- I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2: (Meta)data use vocabularies that follow the FAIR principles
- I3: (Meta)data include qualified references to other (meta)data

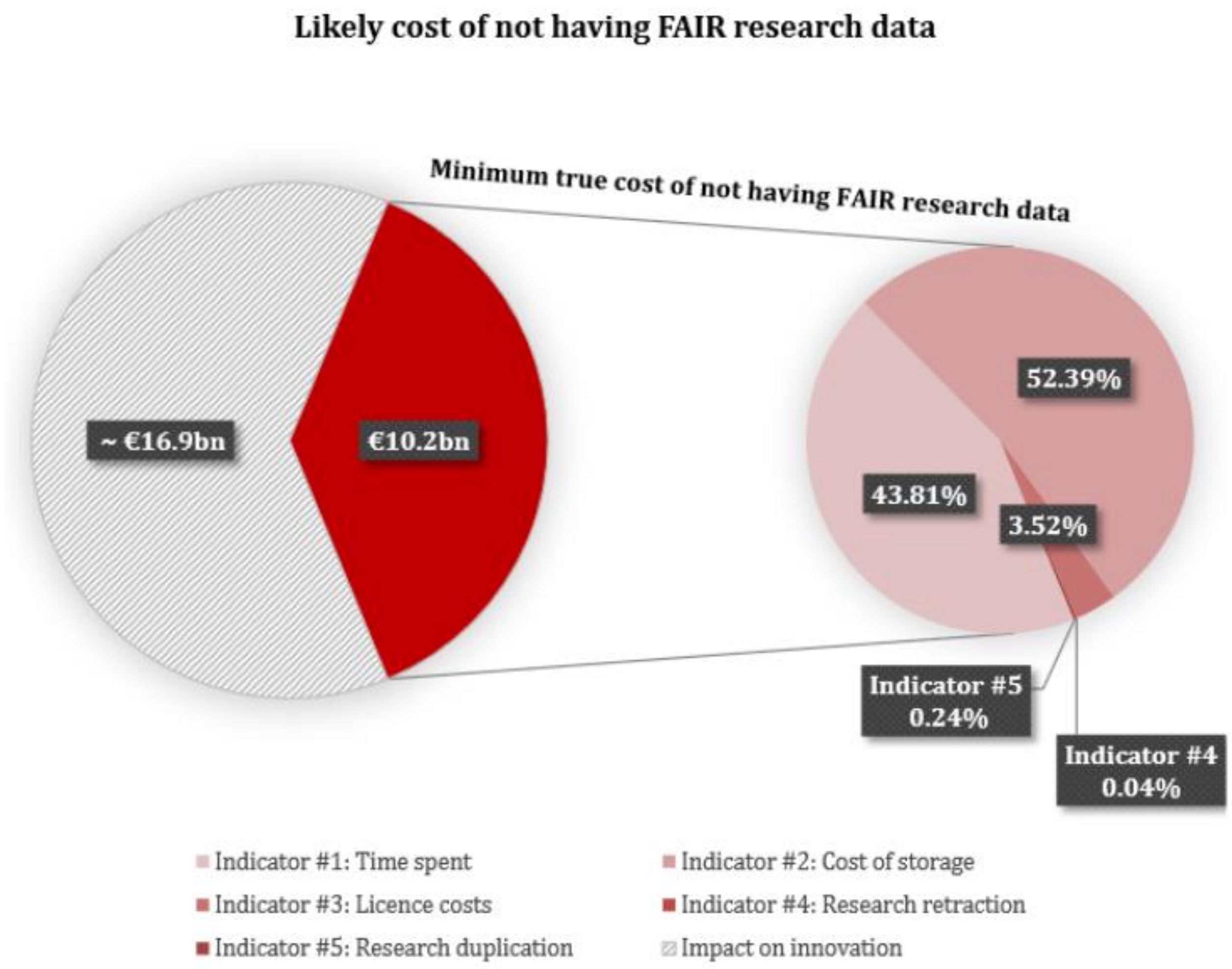
Reusable



- R1: (Meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1: (Meta)data are released with a clear and accessible data usage license
- R1.2: (Meta)data are associated with detailed provenance
- R1.3: (Meta)data meet domain-relevant community standards

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

The cost of not having FAIR data = estimated €10.2bn / year



“Cost-benefit analysis for FAIR research data “ (https://op.europa.eu/s/pevt)

Findable Accessible Interoperable Reusable

panosc

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Data and research outputs

3. **List the various types of data and research outputs that you expect to produce.**

- Output from your research is everything you produced to come up with your findings including :
  - o Raw data
  - o Metadata
  - o Processed data
  - o Analysis workflows
  - o Logbooks
  - o Software
  - o Etc.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Metadata and Why it is important

8. Provide metadata that allows others to understand, cite and reuse your data files.

Documentation or information about a data set.

https://data.research.cornell.edu/content/writing-metadata

- Metadata is all additional data you need to understand your data

- Examples range from file name, time, to experiment condition, energy, sample name, sample parameters, …

- Use the standard vocabularies defined for your domain e.g. Nexus, FITS, …

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Metadata vocabularies

Many standard vocabularies exist for processed data. There are fewer vocabularies for raw data but they do exist. Check the existing standards for your domain.

- Don’t invent a new vocabulary until you are sure none exists

- Databases of standard vocabularies:
  - https://fairsharing.org/ - FAIRsharing as a community approach to standards, repositories and policies
  - https://www.dcc.ac.uk/guidance/standards/metadata/list - list of Metadata standards

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Metadata – Take away messages

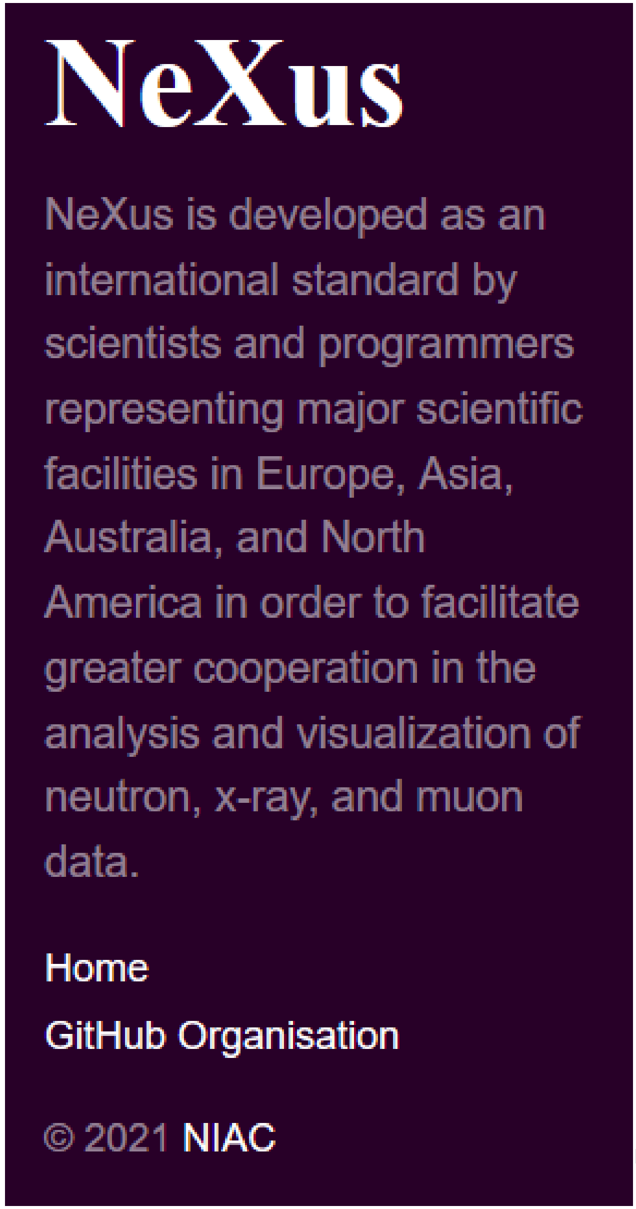
Metadata have a tendency to get treated as 2nd class data.

Whatever you do TAKE YOUR METADATA SERIOUSLY !
The quality of your data depends on it!

- RECORD them DIGITALLY
- STORE them with your DATA
- FOLLOW the STANDARD(s)
- ENSURE others can UNDERSTAND your (meta)data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Example vocabulary – Nexus for photon and neutron sources



NeXus

NeXus is developed as an international standard by scientists and programmers representing major scientific facilities in Europe, Asia, Australia, and North America in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

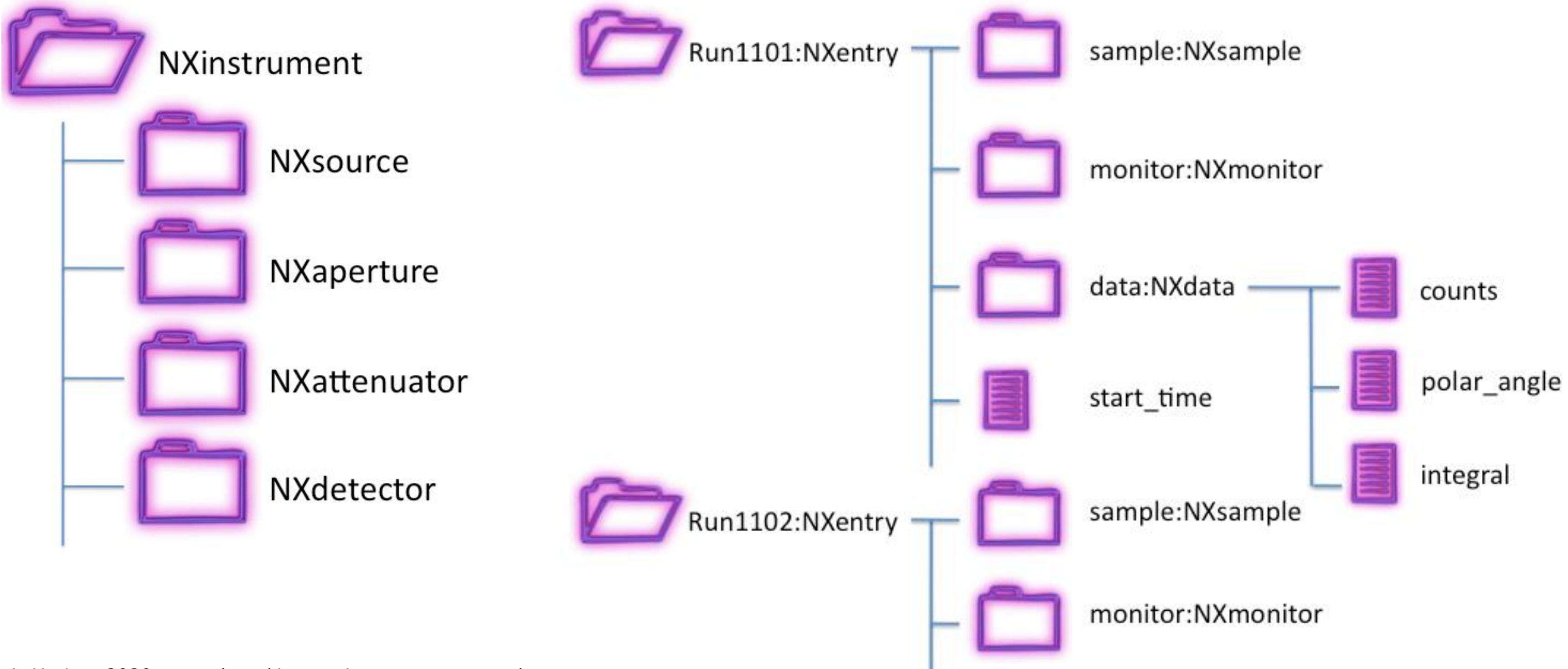
Home

GitHub Organisation

© 2021 NIAC

https://www.nexusformat.org/

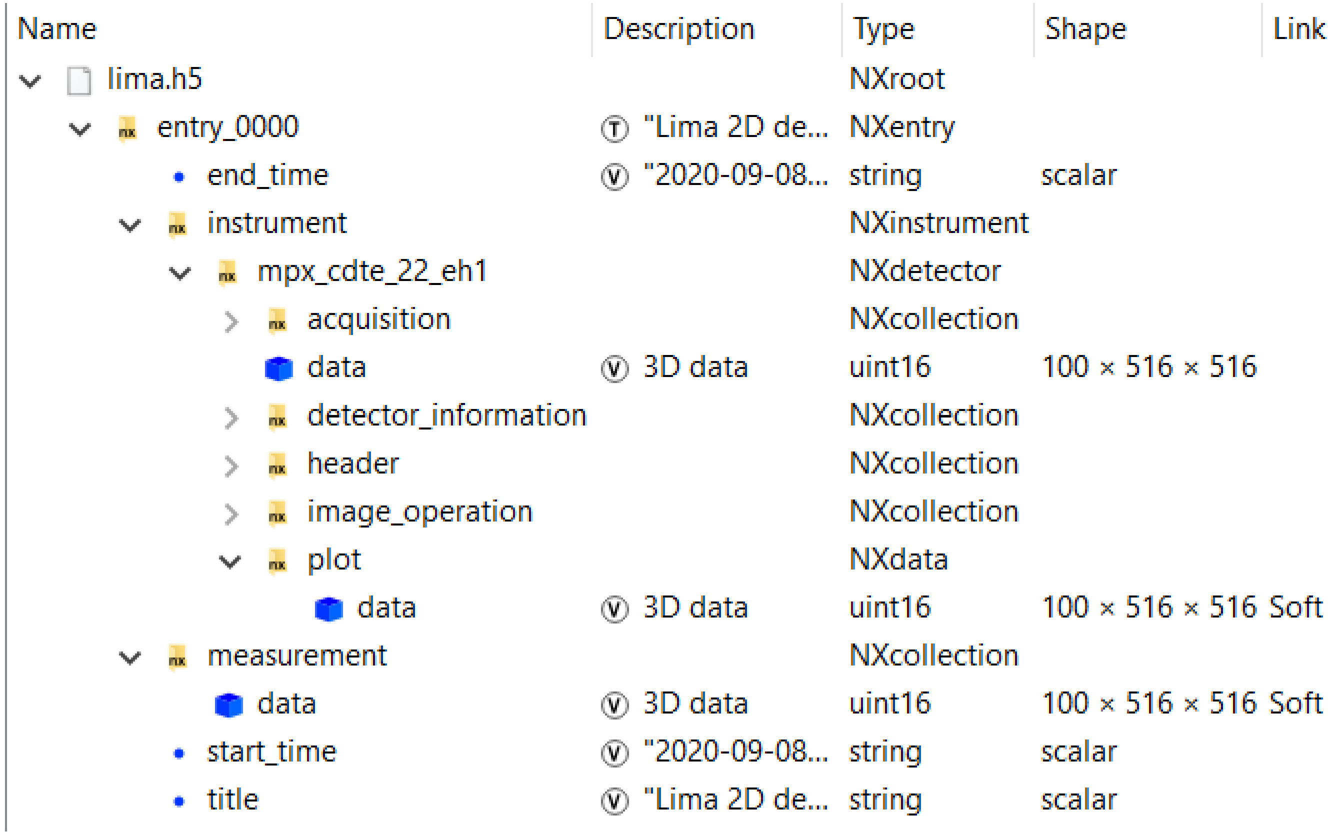
Nexus provides a standard vocabulary for:



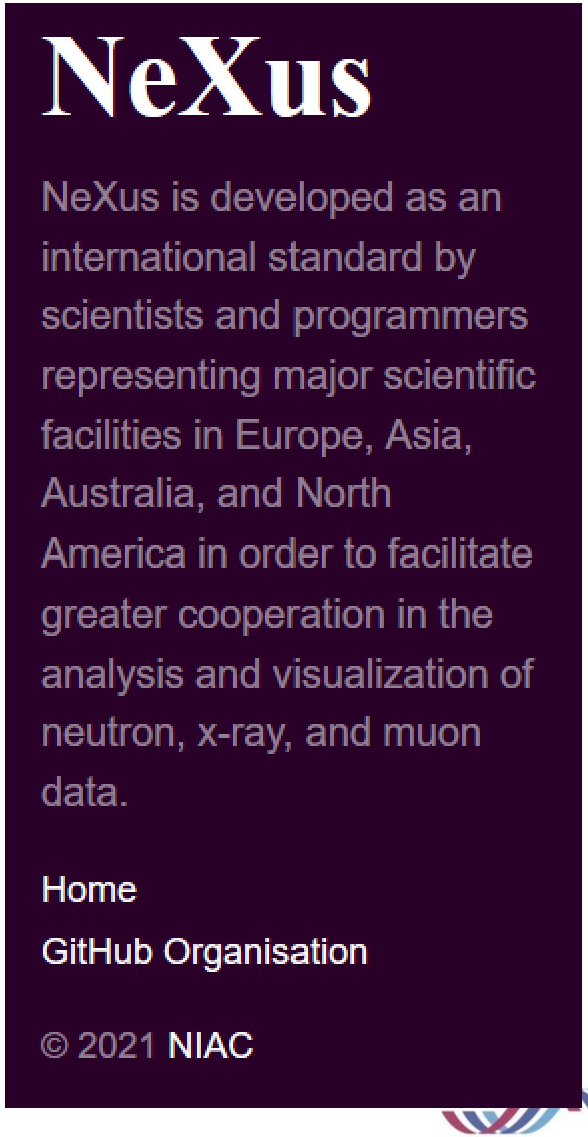
European Union's Horizon 2020 research and innovation programme under g

Example vocabulary – Nexus for photon and neutron sources

Example of structure of data file from ESRF:



Name	Description	Type	Shape	Link
lima.h5		NXroot		
entry_0000	Ⓣ "Lima 2D de...	NXentry		
end_time	Ⓥ "2020-09-08...	string	scalar	
instrument		NXinstrument		
mpx_cdte_22_eh1		NXdetector		
acquisition		NXcollection		
data	Ⓥ 3D data	uint16	100 × 516 × 516	
detector_information		NXcollection		
header		NXcollection		
image_operation		NXcollection		
plot		NXdata		
data	Ⓥ 3D data	uint16	100 × 516 × 516	Soft
measurement		NXcollection		
data	Ⓥ 3D data	uint16	100 × 516 × 516	Soft
start_time	Ⓥ "2020-09-08...	string	scalar	
title	Ⓥ "Lima 2D de...	string	scalar	



NeXus

NeXus is developed as an international standard by scientists and programmers representing major scientific facilities in Europe, Asia, Australia, and North America in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

Home

GitHub Organisation

© 2021 NIAC

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Data formats

5. Define appropriate data file formats (see https://fairsharing.org/ for formats).

7. Check what data format and structure the chosen archive might request.

Data formats refer to how the bytes in a file are interpreted. Not the data vocabularies. Data formats must be readable over the long term (for archiving). Data formats must be efficient

- Example data formats:
  - CSV (Comma Separated Values)
  - TIFF for images
  - HDF5 as container
- USE the STANDARD(s) for your community

Further reading: ETD Guidance Brief File Formats

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

E-logbooks

Provide metadata that allows others to understand your experiment.

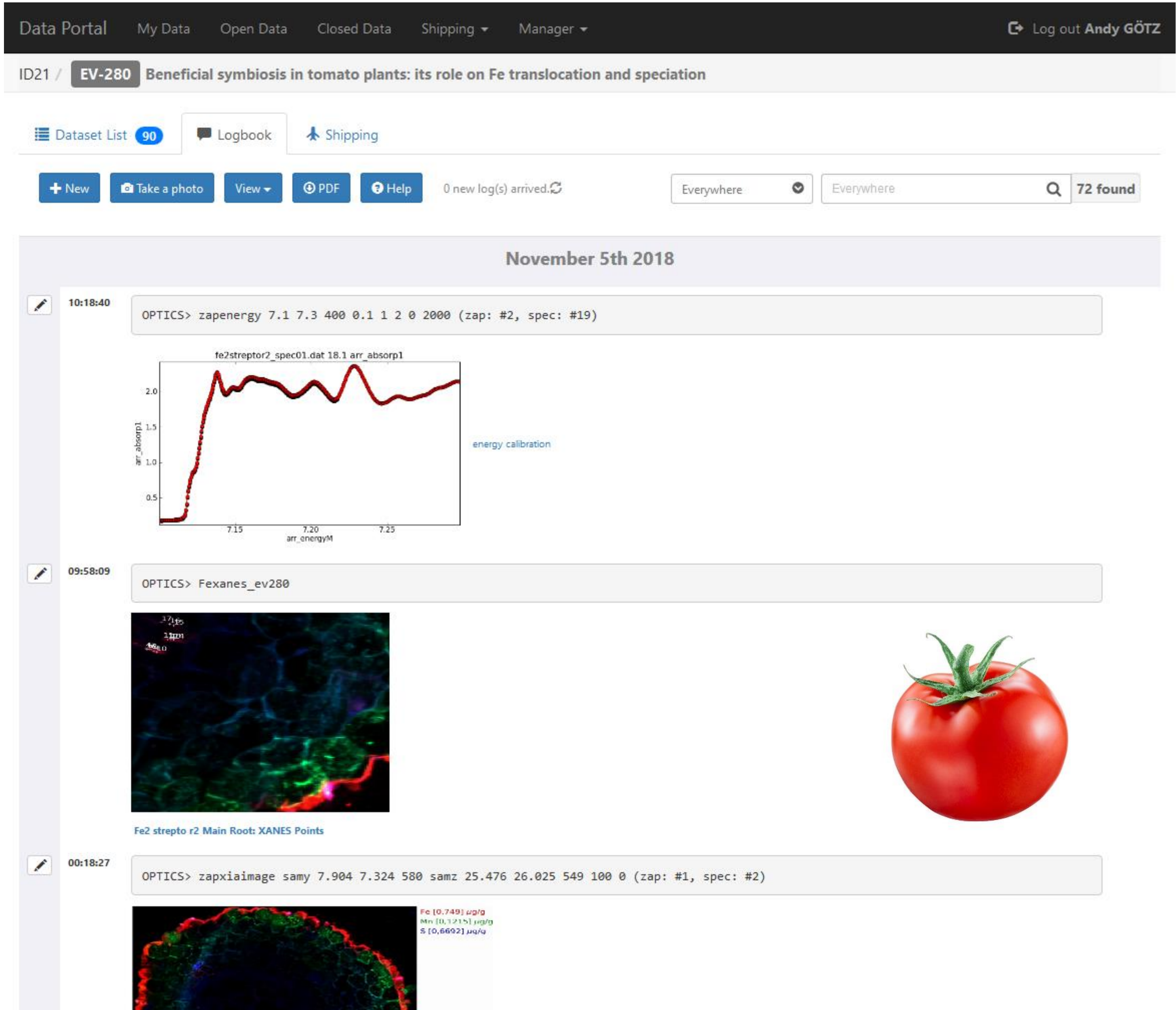
Logbooks are an essential part of the scientific method. All scientists should keep a logbook. E-logbooks replace paper logbooks.

- **E-logbook advantages**
  - Shared editing online
  - Powerful search facilities
  - Access rules during embargo period
  - Allows others to understand what you did during the experiment
- **E-logbook is metadata** and will be part of the open data

Further reading: https://guides.library.oregonstate.edu/research-data-services/data-management-lab-notebooks

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

ESRF e-logbook example – ID21 / EV-280



ID21 / EV-280 Beneficial symbiosis in tomato plants: its role on Fe translocation and speciation

0 new log(s) arrived.

Everywhere | Everywhere | 72 found

November 5th 2018

10:18:40

```
OPTICS> zapenergy 7.1 7.3 400 0.1 1 2 0 2000 (zap: #2, spec: #19)
```

energy calibration

09:58:09

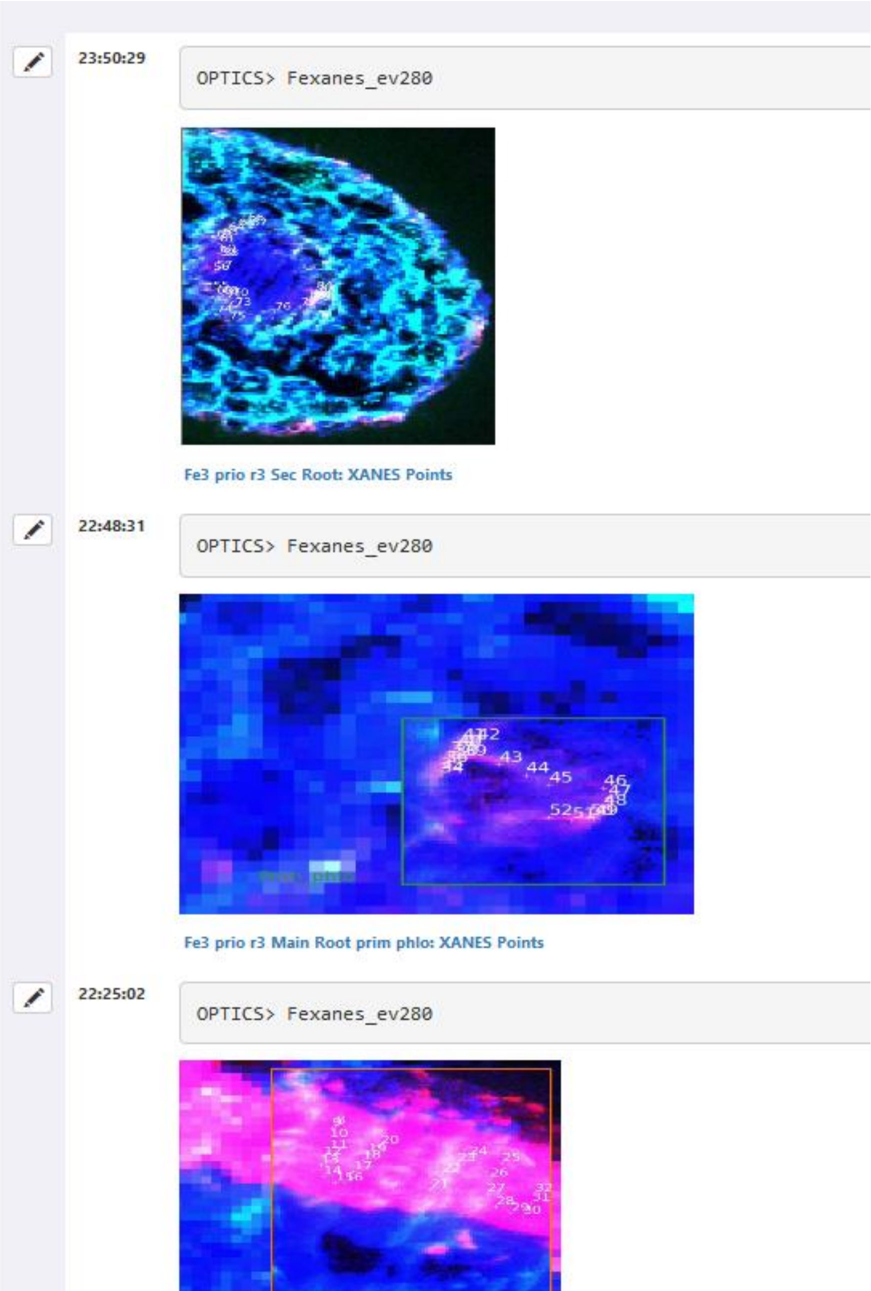
```
OPTICS> Fexanes_ev280
```

Fe2 strepto r2 Main Root: XANES Points

00:18:27

```
OPTICS> zapxiaimage samy 7.904 7.324 580 samz 25.476 26.025 549 100 0 (zap: #1, spec: #2)
```

ent No. 8



23:50:29

```
OPTICS> Fexanes_ev280
```

Fe3 prio r3 Sec Root: XANES Points

22:48:31

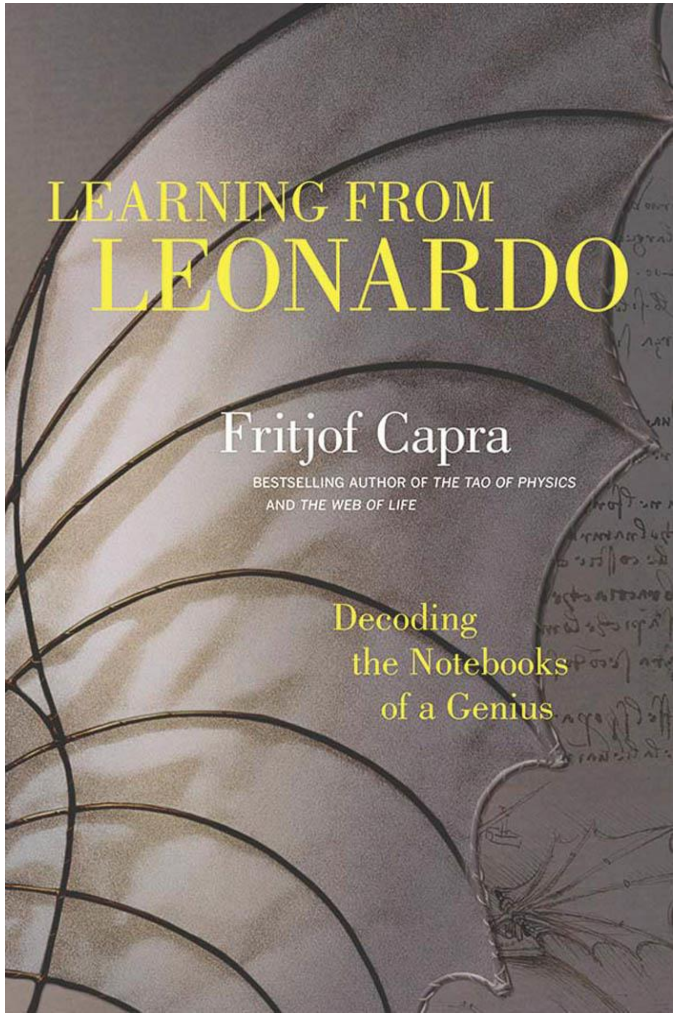
```
OPTICS> Fexanes_ev280
```

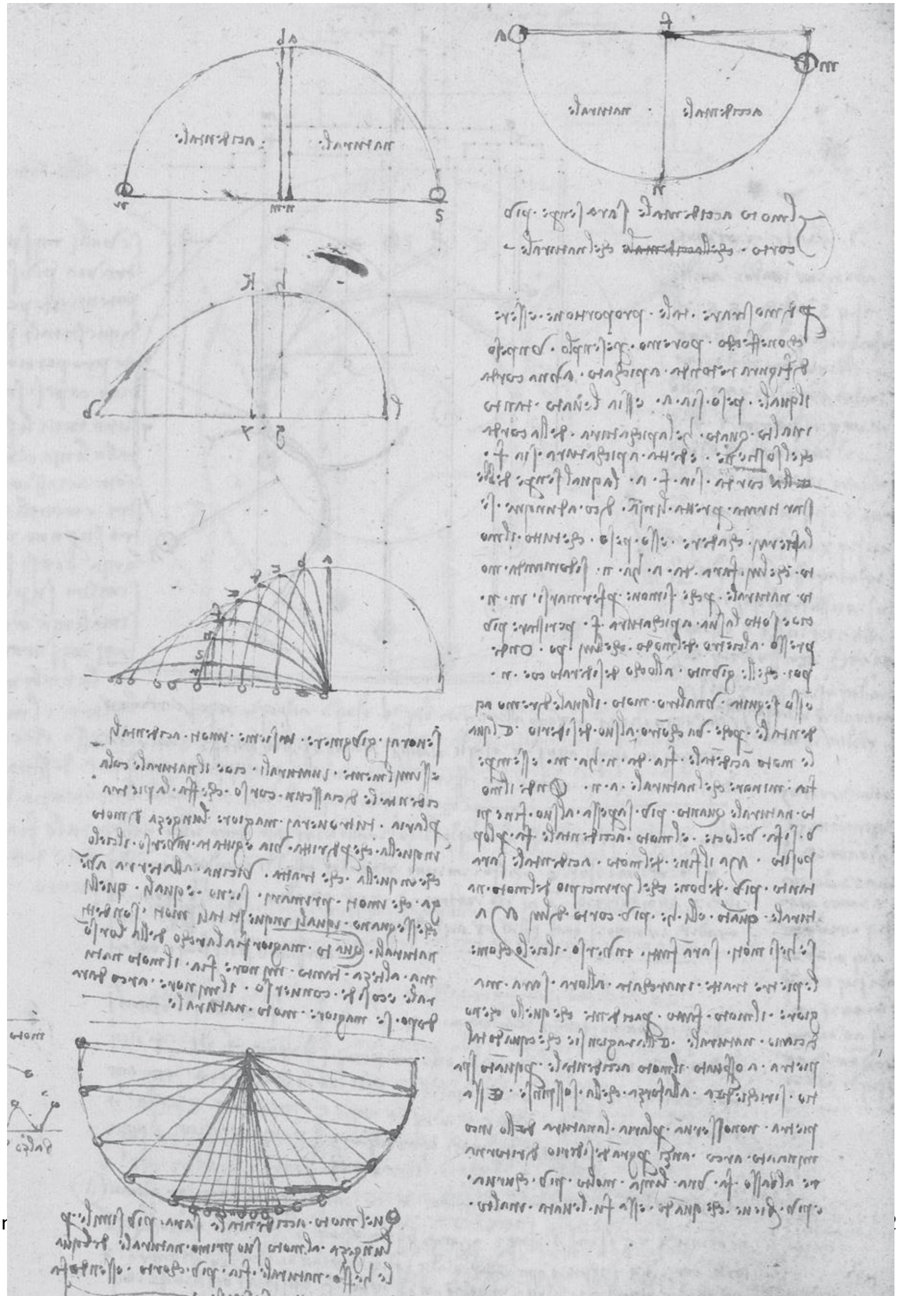
Fe3 prio r3 Main Root prim phlo: XANES Points

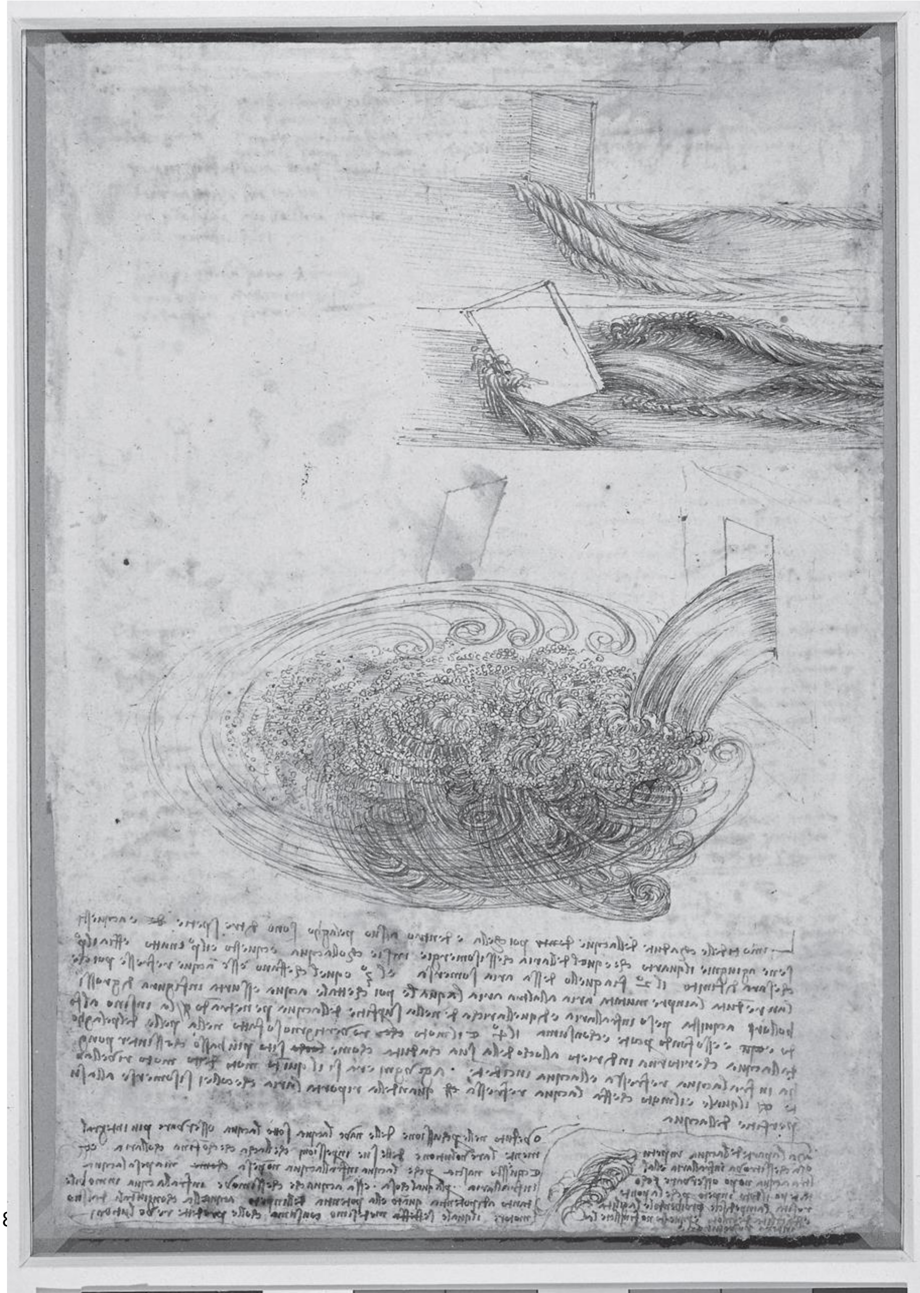
22:25:02

```
OPTICS> Fexanes_ev280
```

Notebooks can inspire Logbooks e.g. **Leonardo da vinci's** notebooks







notebooks can be very useful for posterity…

This project has received funding from the European

Open Source Software

Software is an essential part of a scientists toolset. Many scientists have learned to program so they can analyse their data. The resulting software is part of the outcomes of the research.

- Wherever possible **use Open Source software**

- When **writing software** :
  - Follow **best practices** for software
  - Publish it under an **Open Source license**
  - Store it in an **open (Git) repository** with **version control**
- **Cite your software** in your publications

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

E-Life author guide

https://reviewer.elifesciences.org/author-guide/full

eLife

- Source Code:
  - *Relevant software or source code should be deposited in an open software archive. Where appropriate, authors can upload source code files to the submission system (for example, MATLAB, R, Python, C, C++, Java).* ***Any code provided should be properly documented, in line with these instructions*** *(courtesy of PLOS). Please also refer to our Software sharing policy.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

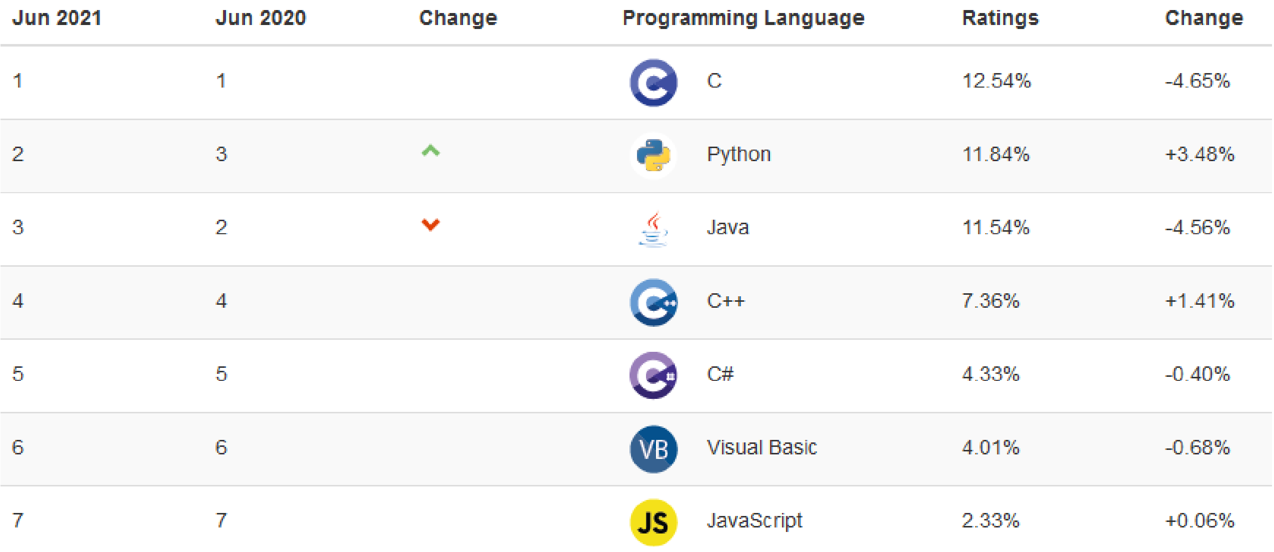
Software tools

Many specific and generic tools exist. One common tool which is being adopted widely is JupyterLab and the Python language.

- **Python** has become the de facto programming language in science

- **Jupyter** notebooks enable reproducible publications https://jupyter.org

- **Binder** service can preserve and run the software for an analysis - https://mybinder.org/



Jun 2021	Jun 2020	Change	Programming Language	Ratings	Change
1	1		C	12.54%	-4.65%
2	3	⌃	Python	11.84%	+3.48%
3	2	⌄	Java	11.54%	-4.56%
4	4		C++	7.36%	+1.41%
5	5		C#	4.33%	-0.40%
6	6		Visual Basic	4.01%	-0.68%
7	7		JavaScript	2.33%	+0.06%

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Data Management Plans (DMP)

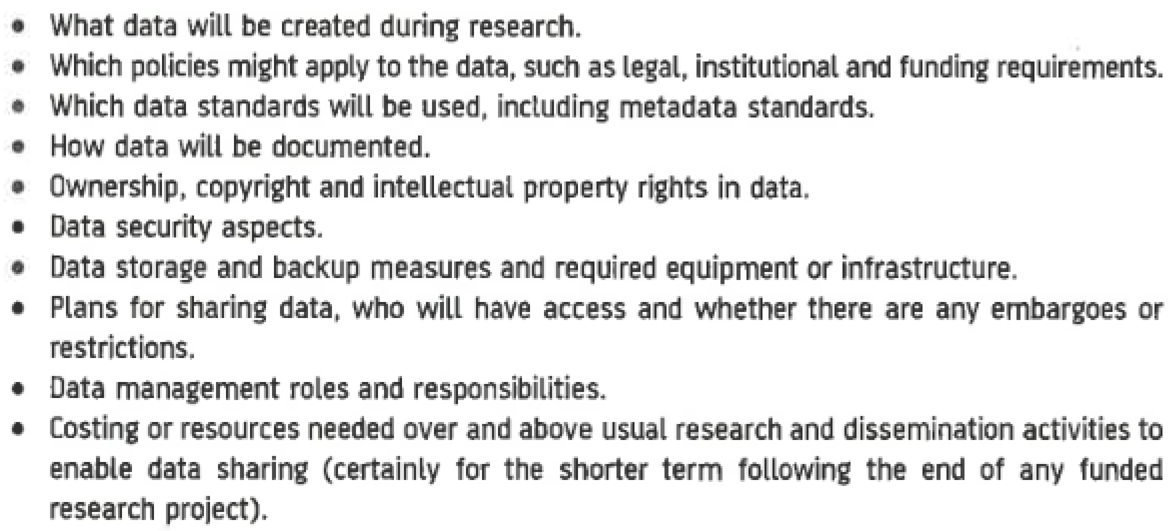
2. **Go online for help in developing a data-management plan. A useful guide outlining UK funder expectations can be found at go.nature.com/2tnohla.**

12. **Revisit your plan frequently and update it if necessary.**

- DMP document the data management steps in a more formal manner
- Funders are requiring DMPs to ensure RDM is planned
- Facilities will require DMPs more and more to be sure Users can deal with the research data
- DMPs are living documents which need to be updated throughout the project
- Examples of DMPs can be found on DMPonline

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Typical questions to be answered by the DMP



- What data will be created during research.
- Which policies might apply to the data, such as legal, institutional and funding requirements.
- Which data standards will be used, including metadata standards.
- How data will be documented.
- Ownership, copyright and intellectual property rights in data.
- Data security aspects.
- Data storage and backup measures and required equipment or infrastructure.
- Plans for sharing data, who will have access and whether there are any embargoes or restrictions.
- Data management roles and responsibilities.
- Costing or resources needed over and above usual research and dissemination activities to enable data sharing (certainly for the shorter term following the end of any funded research project).

“**Managing and Sharing Research Data: A Guide to Good Practice**” by Louise Corti et al

https://study.sagepub.com/corti2e

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Data repositories

6. Look for data repositories used by your research community or your host institution (see www.re3data.org for examples).

A data repository stores data for citing, accessing and archiving data over the long term. Repositories can be provided by facilities or community based. Choose the right repository with the service you expect

- Facilities offer repositories for raw and (sometimes) processed data e.g. https://data.esrf.fr
- Choose repository which is certified e.g. http://go.nature.com/2eLHBFP)
- Use an institute or community archive which is sustainable

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Data archiving

9. **Make clear how and when your data can be shared with scientists outside your group.**
10. **If your research involves sensitive data, explain any legal and ethical restrictions on data access and reuse.**
11. **Assign responsibility for long-term data curation to a suitable office.**

- Data need to be archived for long term future use
- You don’t know when and how your data could turn out to be useful
- The meaning of long term depends on the data e.g. is 10 years enough?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Digital Object Identifier (DOI)

A DOI or Digital Object Identifier, is a string of numbers, letters and symbols used to permanently identify any object and link it to the web.

DOIs were originally used for publications and are now used for many things including movies, samples, instruments and scientific DATA.

- A DOI is one implementation of a PID (Persistent Identifier)
- A web address (url) is not a PID because it is not guaranteed
- Make sure the data you want to cite has a DOI
- Cite the instrument, samples etc. you used

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Journal require datasets accessible

More and more journals require datasets used in the publication to be cited and accessible. For example eLife, Nature, Plos, Science, …

- eLife – https://reviewer.elifesciences.org/author-guide/full

All datasets used in a publication should be cited in the text and listed in the reference section and/or data availability statement. References for data sets and program code should include a persistent identifier, for example a Digital Object Identifier (DOI) or accession number.

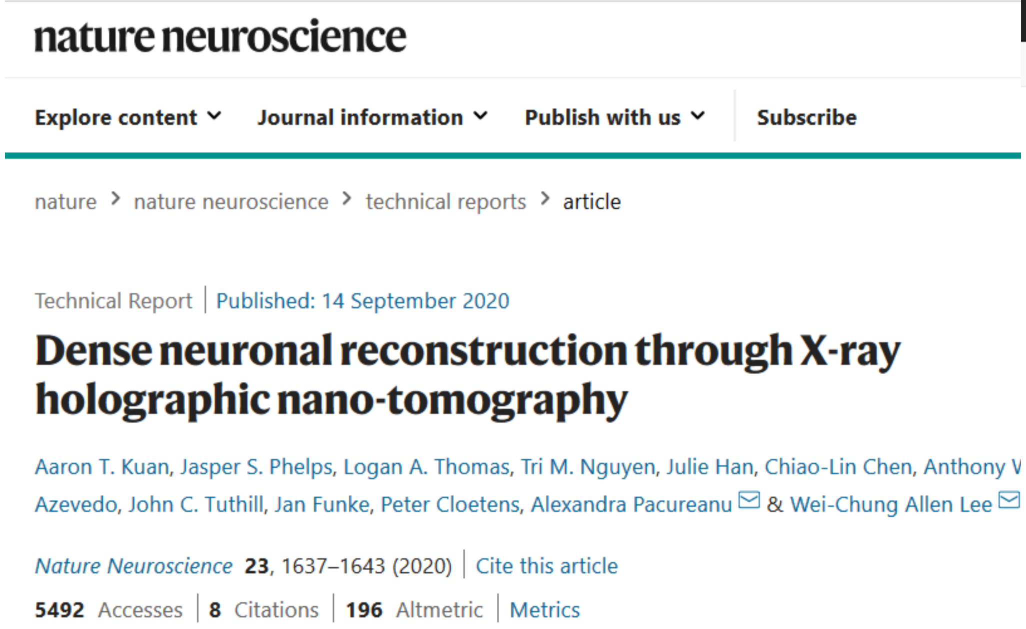
…

Relevant software or source code should be deposited in an open software archive.

…

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Example of article citing data



nature neuroscience

Explore content ˅ Journal information ˅ Publish with us ˅ Subscribe

nature › nature neuroscience › technical reports › article

Technical Report | Published: 14 September 2020

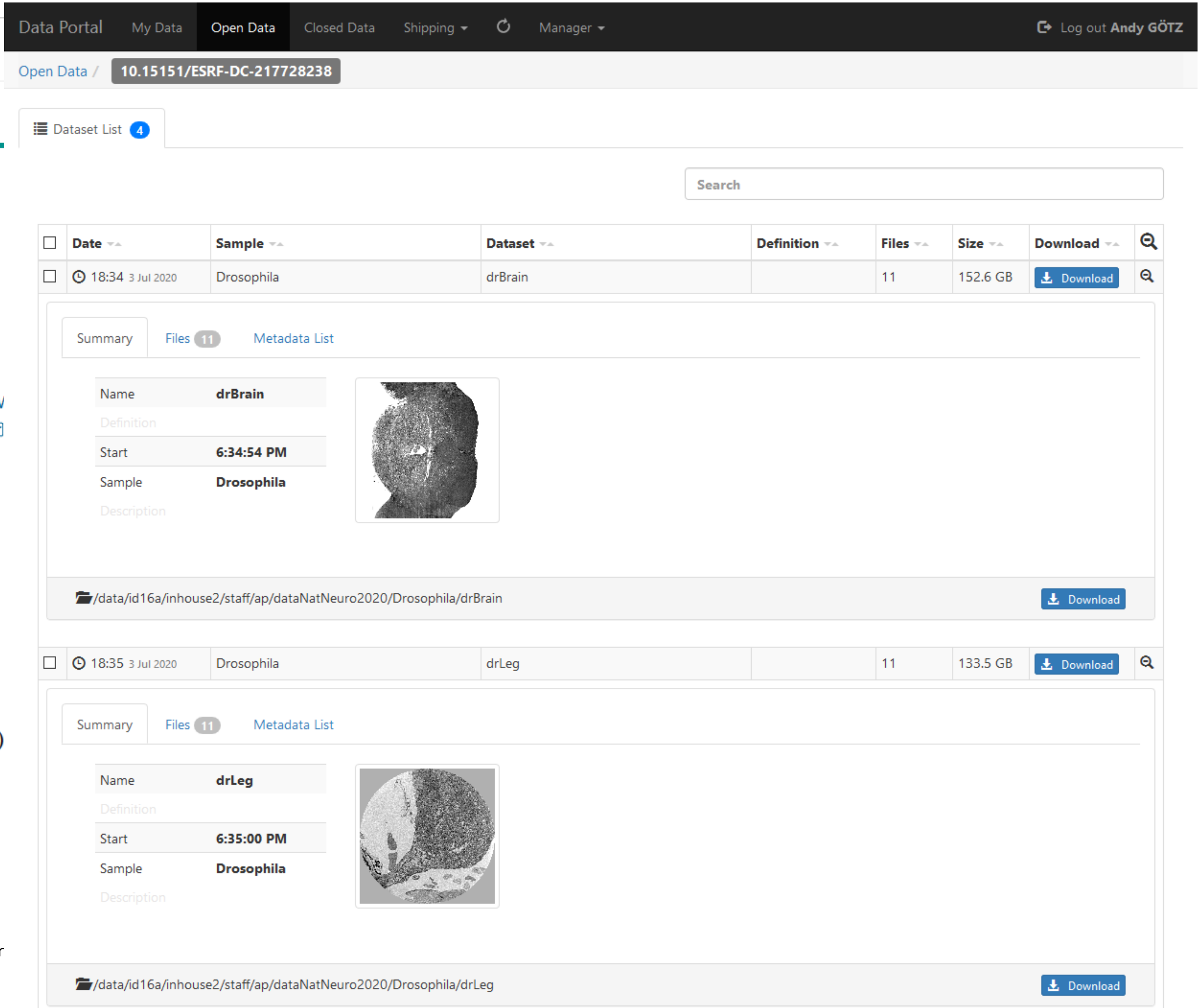
Dense neuronal reconstruction through X-ray holographic nano-tomography

Aaron T. Kuan, Jasper S. Phelps, Logan A. Thomas, Tri M. Nguyen, Julie Han, Chiao-Lin Chen, Anthony W Azevedo, John C. Tuthill, Jan Funke, Peter Cloetens, Alexandra Pacureanu ✉ & Wei-Chung Allen Lee ✉

Nature Neuroscience **23**, 1637–1643 (2020) | Cite this article

5492 Accesses | **8** Citations | **196** Altmetric | Metrics

https://doi.org/10.1038/s41593-020-0704-9



Data Portal My Data Open Data Closed Data Shipping ▾ ↻ Manager ▾ Log out **Andy GÖTZ**

Open Data / **10.15151/ESRF-DC-217728238**

Dataset List 4

Search

□	Date	Sample	Dataset	Definition	Files	Size	Download	🔍
□	18:34 3 Jul 2020	Drosophila	drBrain		11	152.6 GB	Download	🔍

Summary Files 11 Metadata List

Name	**drBrain**
Definition	
Start	**6:34:54 PM**
Sample	**Drosophila**
Description	

/data/id16a/inhouse2/staff/ap/dataNatNeuro2020/Drosophila/drBrain Download

□	Date	Sample	Dataset	Definition	Files	Size	Download	🔍
□	18:35 3 Jul 2020	Drosophila	drLeg		11	133.5 GB	Download	🔍

Summary Files 11 Metadata List

Name	**drLeg**
Definition	
Start	**6:35:00 PM**
Sample	**Drosophila**
Description	

/data/id16a/inhouse2/staff/ap/dataNatNeuro2020/Drosophila/drLeg Download

3. ESRF (https://data.esrf.fr/public/10.15151/ESRF-DC-217728238) (anonymous login)

DOI: doi.esrf.fr/10.15151/ESRF-DC-217728238

This project has received funding from the European Union's Horizon 2020 research and ir

Data storage

4. **Decide what data and research materials require archiving and determine how much storage space you will need.**

- Data volumes are constantly increasing (up to Petabytes)
- You could be faced with more data than you can store locally
- Research facilities provide services to keep raw data at the facility
- Access to remote data is via remote data services (similar to cloud)
- Commercial cloud offer practically unlimited resources at a cost
- Data stored on commercial cloud disappear when you stop paying

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

File naming conventions

3. List the various types of data and research outputs that you expect to produce.

Adopt a directory and file naming convention which will allow you to know what the file contains.

- For example:

  Proposal/Beamline/Sample_name_Scan_type.ext

  MA1234/ID56/Gold_50_nm_ptycho_scan.h5

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Own your identity in the digital world

In a digital world you need to control your identity and not give it away to the corporate world to exploit. It is highly recommended to create your own identity using ORCID – a free non-commercial service

- Benefits of an ORCID identity:
  - You will be distinguished from every other researcher, even researchers who share your same name,
  - Your research outputs and activities will be correctly attributed to you,
  - Your contributions and affiliations will be reliably and easily connected to you,
  - You will save time when filling out forms, (leaving more time for research!),
  - You will enjoy improved discoverability and recognition,
  - You will be able to connect your record to a growing number of institutions, funders, and publishers,
  - Your ORCID record is yours, for free, forever.

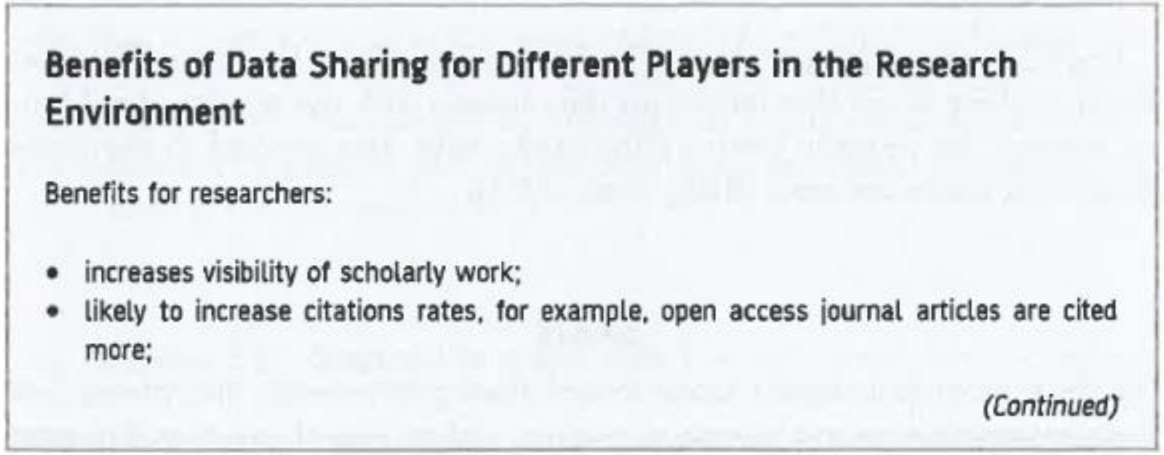
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

What are the advantages of doing RDM?

- Better data and metadata means better science
- Saves you time and improves your results
- Allows you to use standard data services
    - Remote data analysis
    - Data archiving
    - DOI
- Publications with open data are cited more often
- You get more credit for your work
- Science is more reproducible and replicable

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Benefits of data sharing

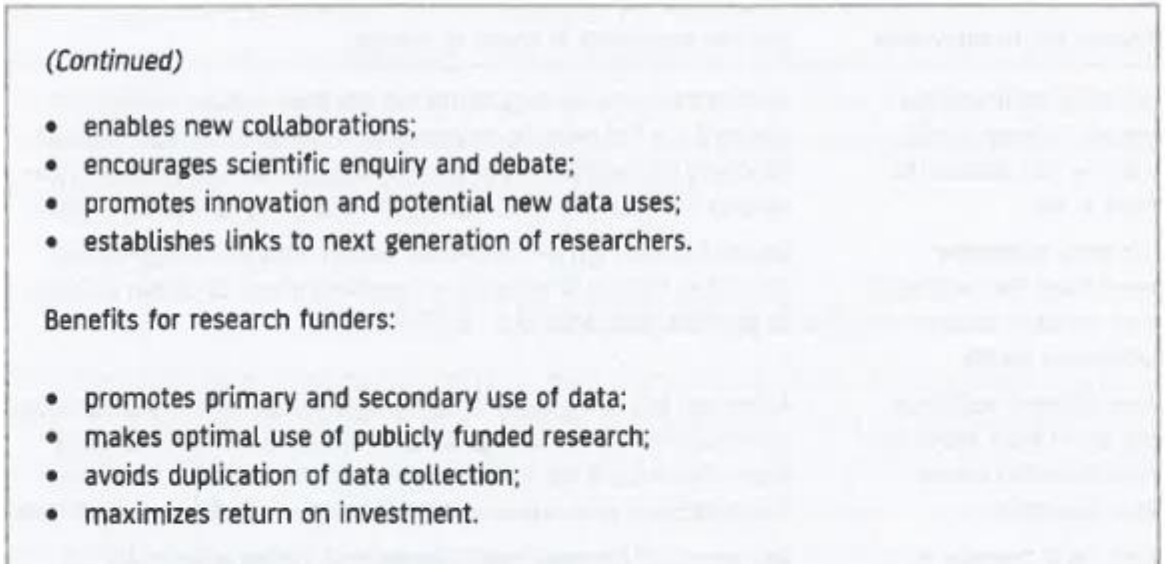


Benefits of Data Sharing for Different Players in the Research Environment

Benefits for researchers:

- increases visibility of scholarly work;
- likely to increase citations rates, for example, open access journal articles are cited more;

(Continued)

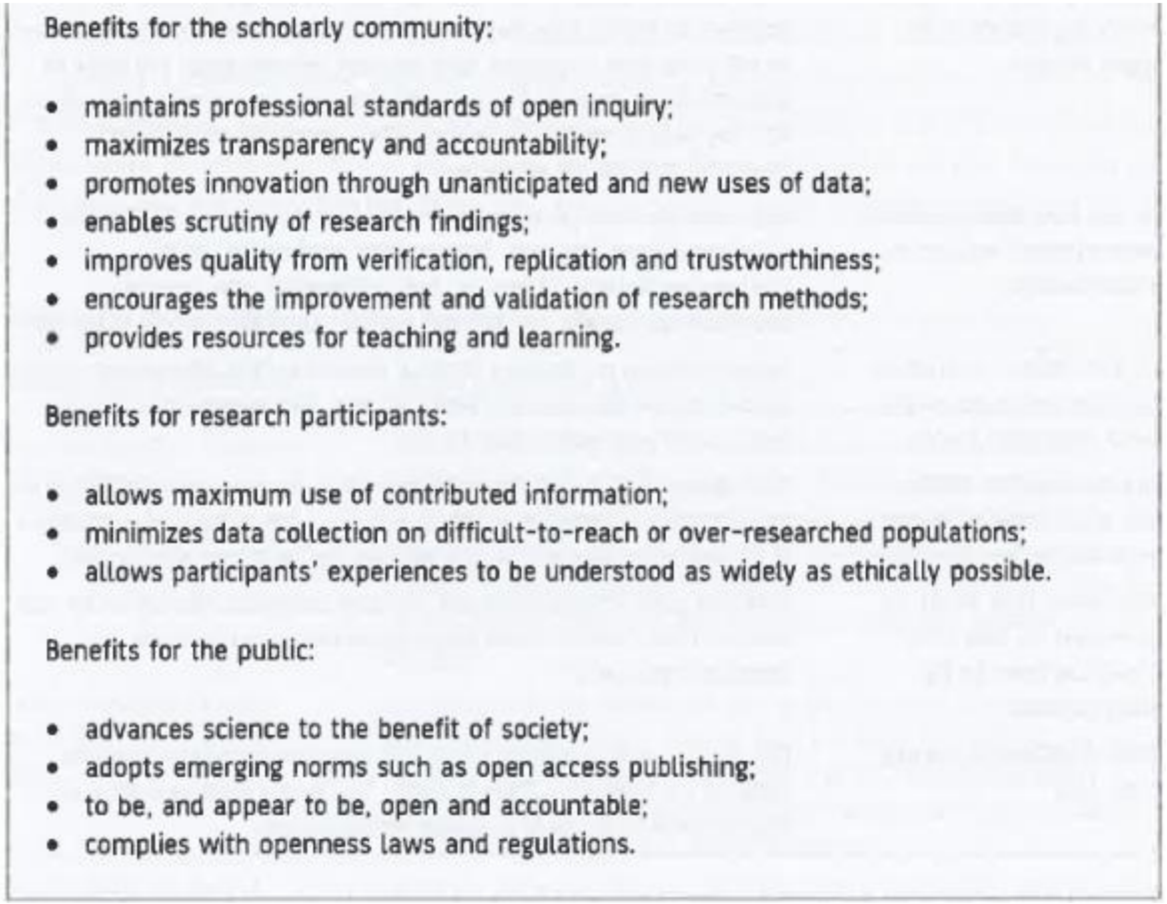


(Continued)

- enables new collaborations;
- encourages scientific enquiry and debate;
- promotes innovation and potential new data uses;
- establishes links to next generation of researchers.

Benefits for research funders:

- promotes primary and secondary use of data;
- makes optimal use of publicly funded research;
- avoids duplication of data collection;
- maximizes return on investment.



Benefits for the scholarly community:

- maintains professional standards of open inquiry;
- maximizes transparency and accountability;
- promotes innovation through unanticipated and new uses of data;
- enables scrutiny of research findings;
- improves quality from verification, replication and trustworthiness;
- encourages the improvement and validation of research methods;
- provides resources for teaching and learning.

Benefits for research participants:

- allows maximum use of contributed information;
- minimizes data collection on difficult-to-reach or over-researched populations;
- allows participants' experiences to be understood as widely as ethically possible.

Benefits for the public:

- advances science to the benefit of society;
- adopts emerging norms such as open access publishing;
- to be, and appear to be, open and accountable;
- complies with openness laws and regulations.

"**Managing and Sharing Research Data: A Guide to Good Practice**" by Louise Corti etc al

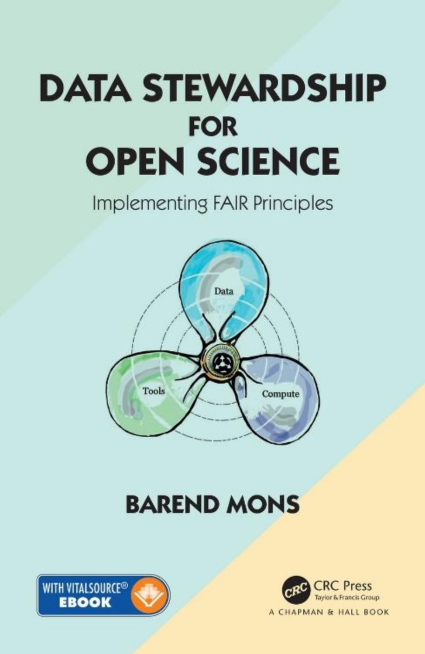
https://study.sagepub.com/corti2e

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Learning more about RDM

- RDMKit - https://rdmkit.elixir-europe.org/index.html
    - Provides a rich set of resources for all aspects of RDM mainly for researchers working in the Life Sciences but also for other Sciences. Very comprehensive overview, pragmatic approach, up-to-date. An excellent place to start and/or find information.

- Recommended reading:



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Tools to help you manage your research

A non-exhaustive list of tools to explore

- Open science framework – osf.io
- Protocols.io
- Fairsharing.org
- Jupyter.org notebooks

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Conclusion

Adopting best practices for Research Data Management has many benefits especially helping MAKE BETTER SCIENCE

- Follow a **checklist** similar to the one described in this talk which covers the following topics
  - Data Management Plan, Data Policy, Data Outputs, File types, File Formats, Software, Workflows, e-Logbooks, Data Storage, Data Archiving, Data DOI
  - Apply the **FAIR principles** – ask yourself if you or someone else will be able to use or understand your data
  - Make your Data **FAIR**
- The **digital tools** exist for treating your data seriously
- There is a lot more to science than just text publications …

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Acknowledgements

- RDMKit Elixir online guide
- University of Saskatchewan
  - https://library.usask.ca/studentlearning/workshops/grad-research.php#panel-section-3-ResearchDataManagementWhatYouNeedtoKnow
- Nature magazine
  - https://nature.com
- PaNOSC, ExPaNDS, EOSC H2020 projects
- Wikipedia, Internet

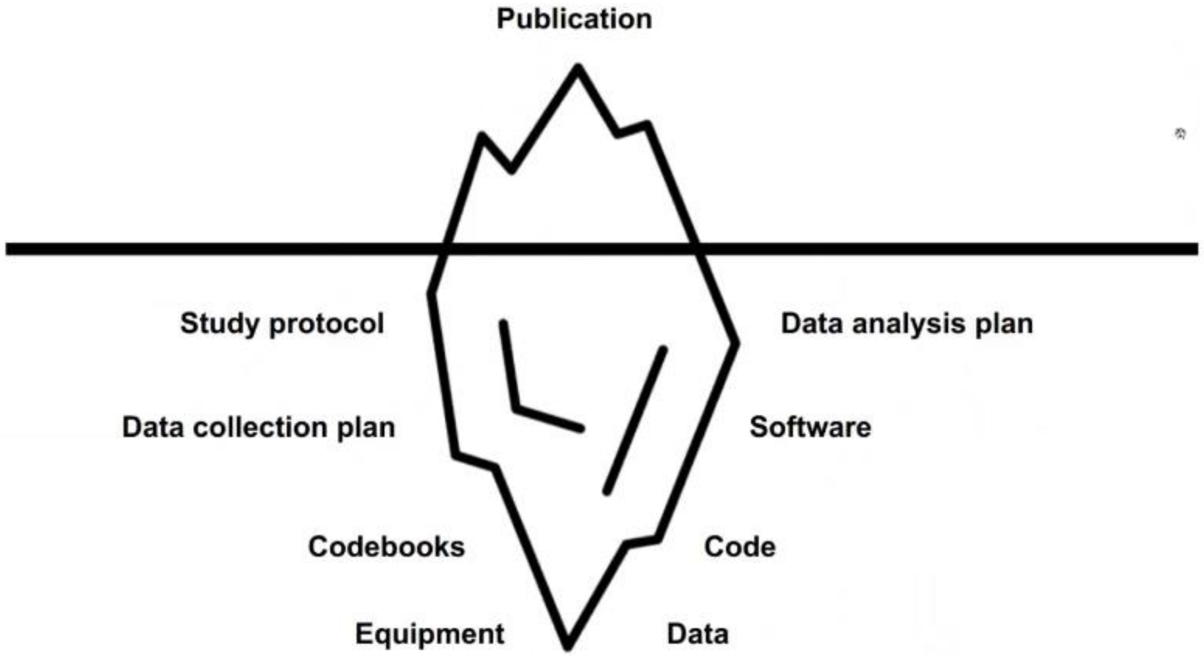
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

panosc

photon and neutron open science cloud

Thank you

andy.gotz@esrf.fr



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852