



The importance of Research Data Management

Better Data for Better Science –
Research Data Management Workshop
27th October, 2021
by Andy Götz (ESRF)
PaNOSC coordinator



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

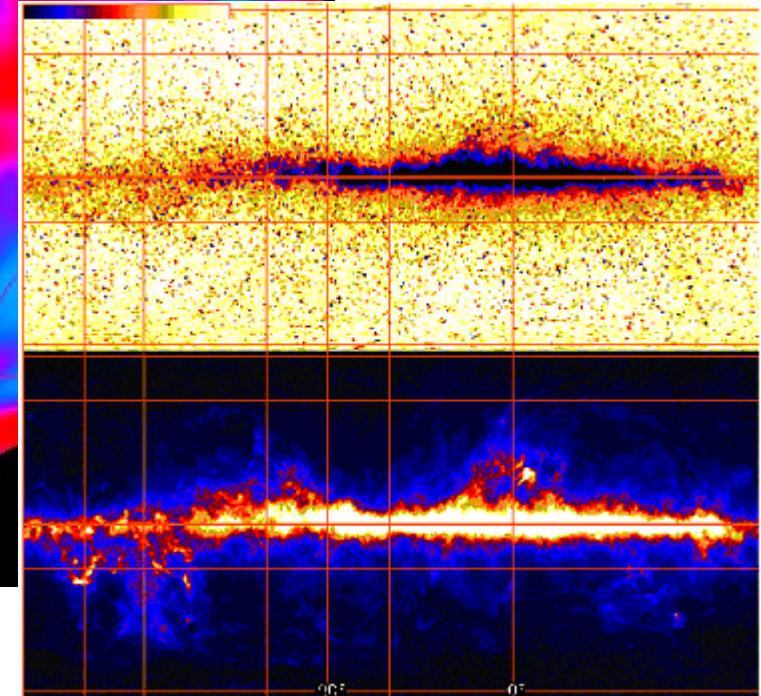
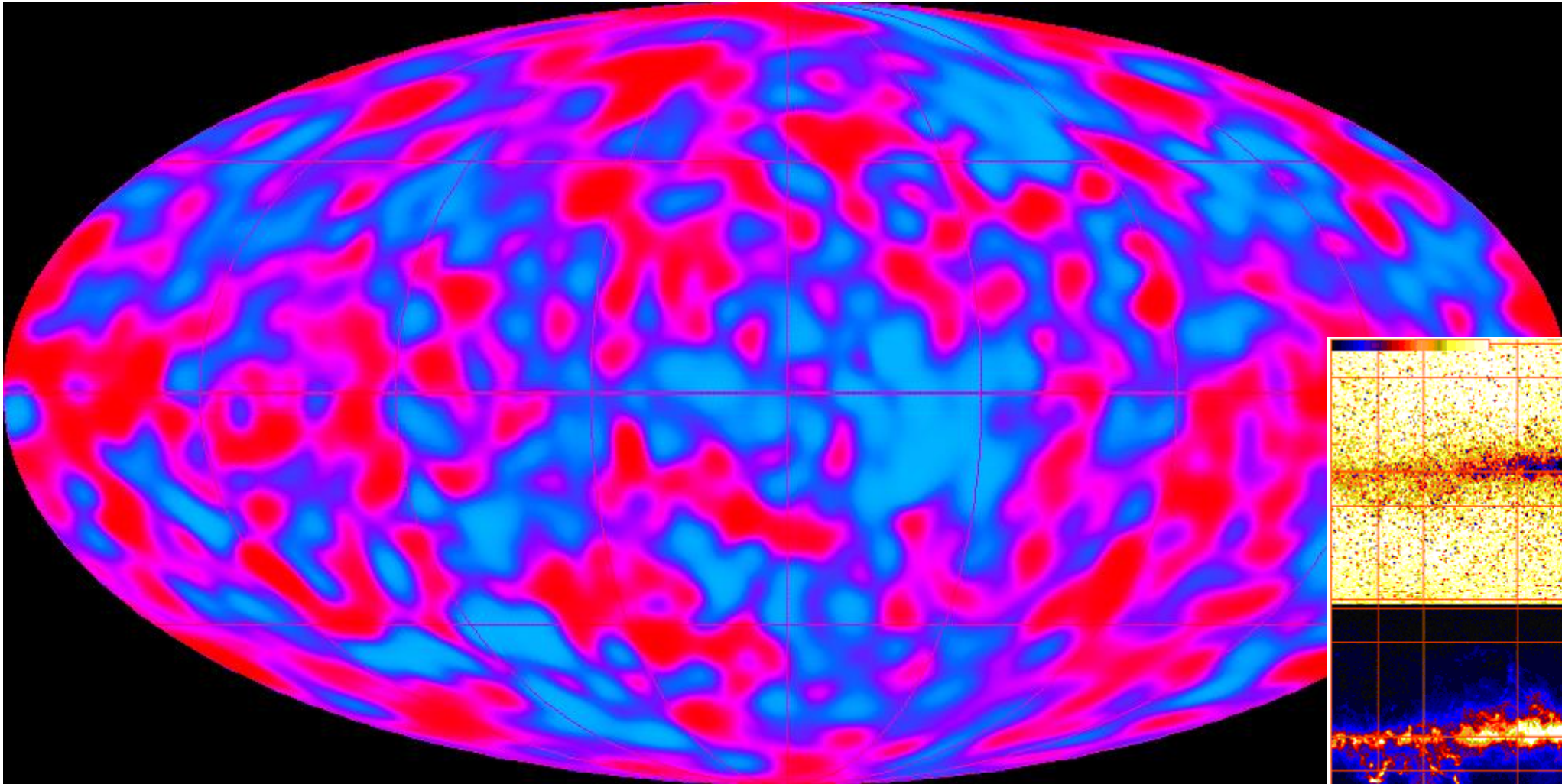
Petabytes

in

40 years



CMB anisotropy map formed from data taken by the COBE spacecraft



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

<https://researchdata.ru.ac.za/>



RHODES UNIVERSITY
Where leaders learn

Discover research and open educational
resources from **Rhodes University**



+ Follow

40,456 views

13,189 downloads

[more stats...](#)

ALL

CATEGORIES

GROUPS

SEARCH

sort by: Posted date



Hide footer

[About](#) | [Features](#) | [Tools](#) | [Blog](#) | [Ambassadors](#) | [Contact](#) | [FAQs](#) | [Privacy Policy](#) | [Cookie Policy](#) | [T&Cs](#) | [Disclaimer](#) | [Sitemap](#)

figshare. credit for all your research.

<https://opendata.cern.ch/>

Explore more than **two petabytes**
of open data from particle physics!

Start typing...

Search

search examples: [collision datasets](#), [keywords:education](#), [energy:7TeV](#)

Explore

[datasets](#)
[software](#)
[environments](#)
[documentation](#)

Focus on

[ATLAS](#)
[ALICE](#)
[CMS](#)
[LHCb](#)
[OPERA](#)
[QUEST](#)

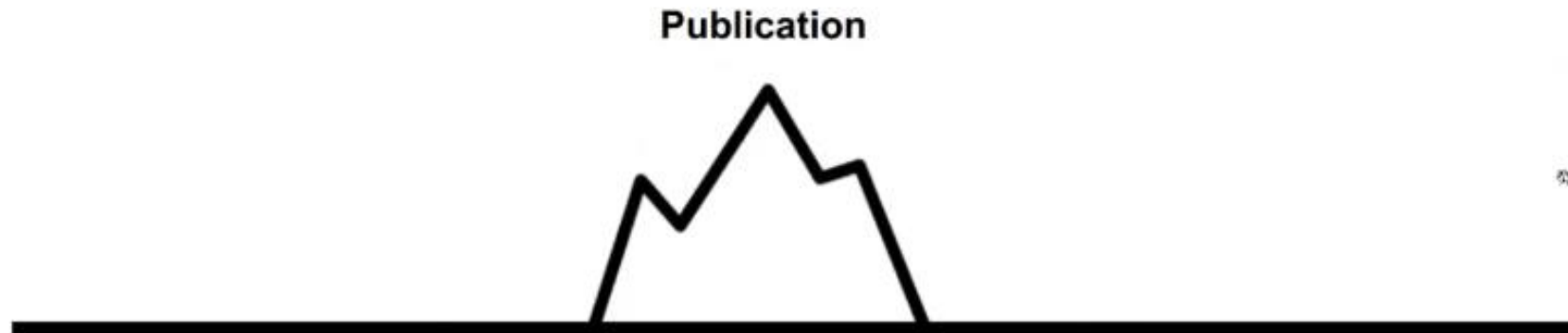
<https://rodare.hzdr.de>



Welcome to RODARE

On April 25 the HZDR data centre initially launched the new data publication platform at HZDR. We call it RODARE - **RO**ssendorf **DA**ta **RE**pository. RODARE is based on the Invenio Digital Library framework and powered by Open Source technologies and frameworks.

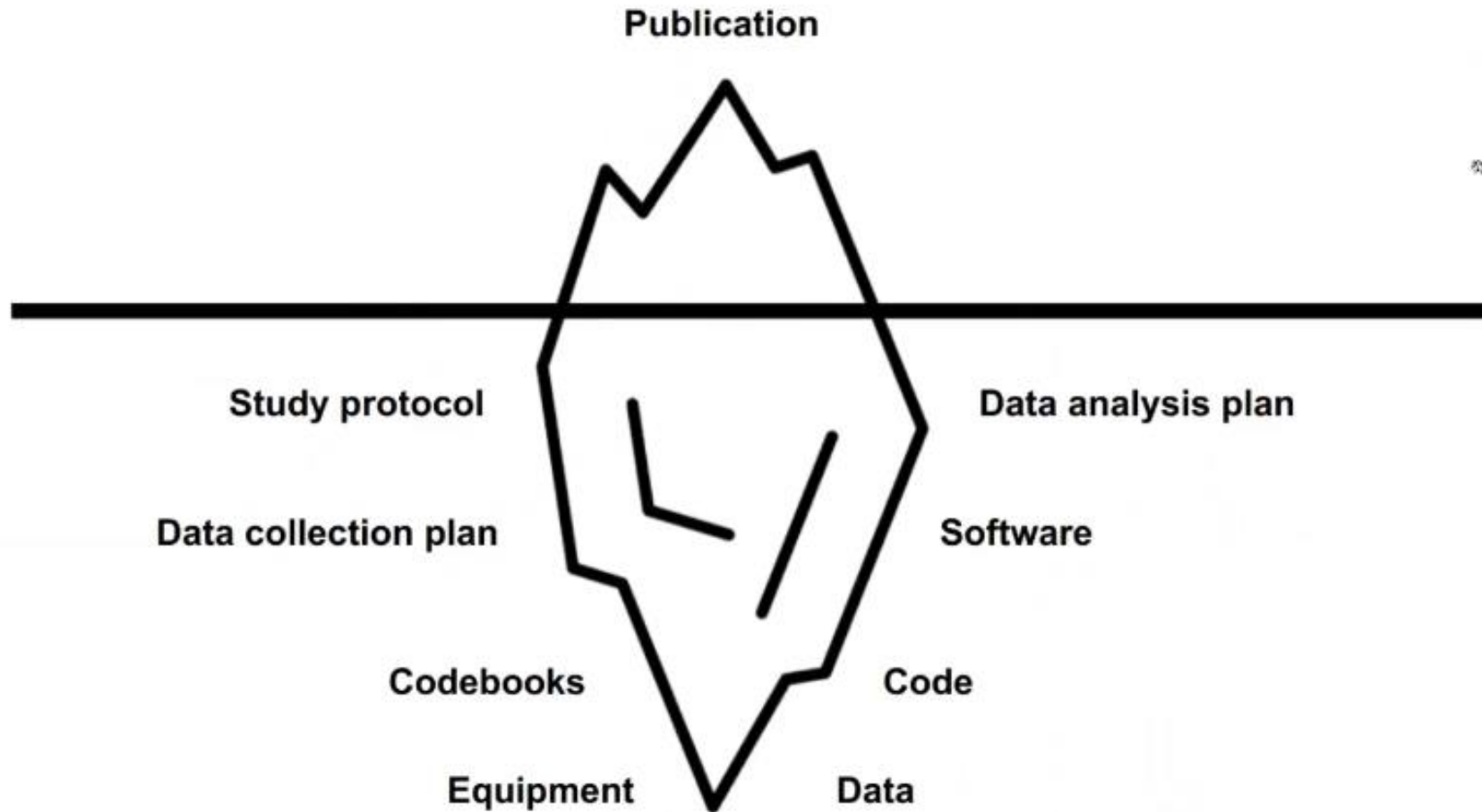
Science produces Publications



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



Reproducibility and Replicability

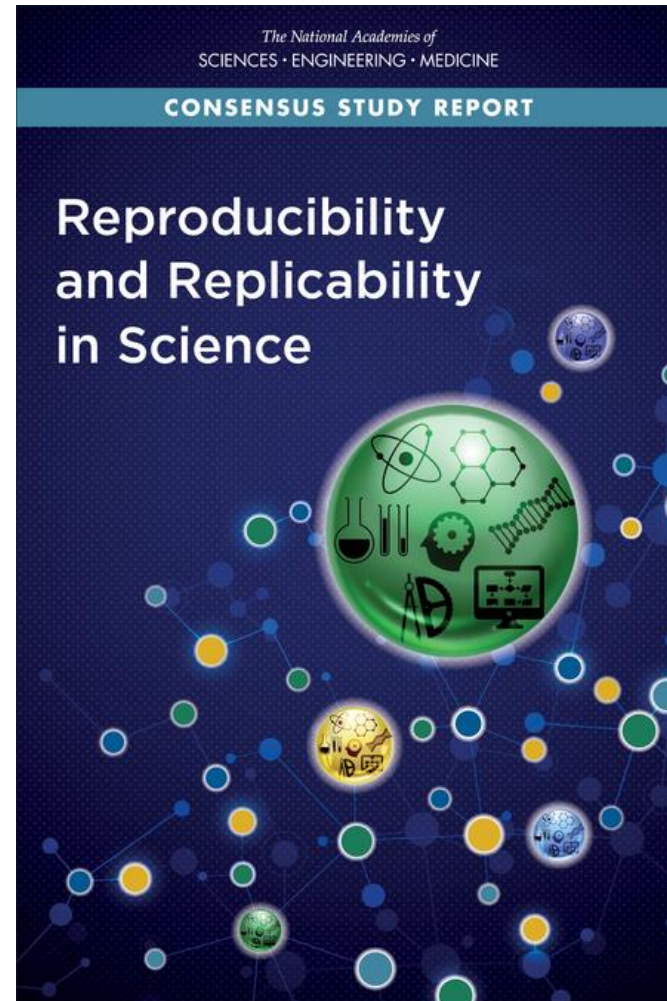
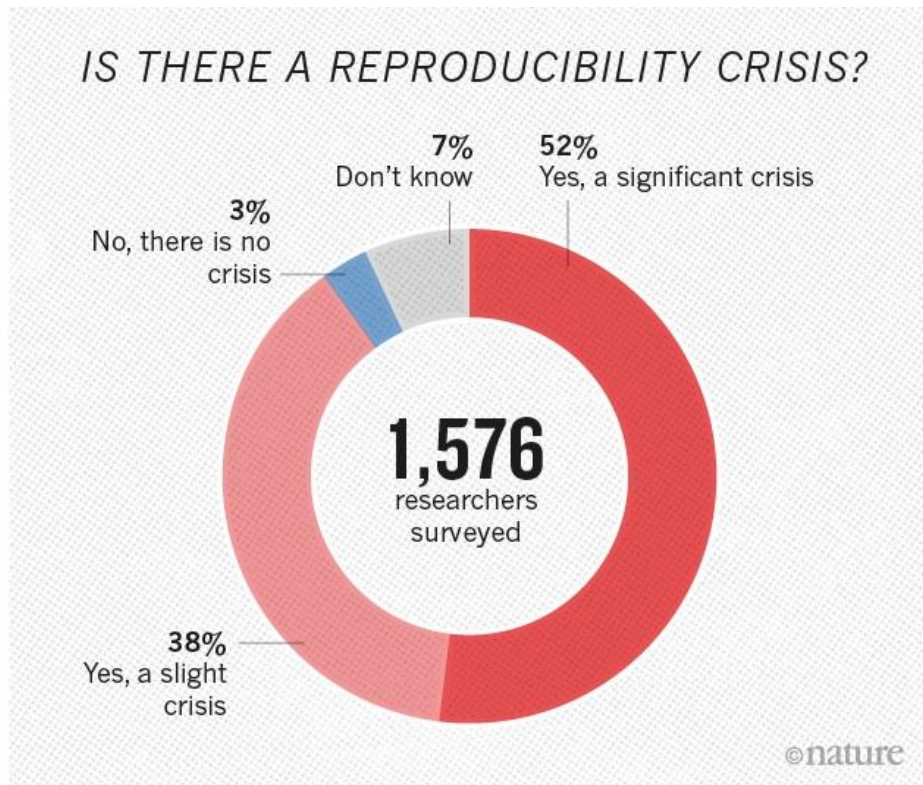
Published: 25 May 2016

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](#)



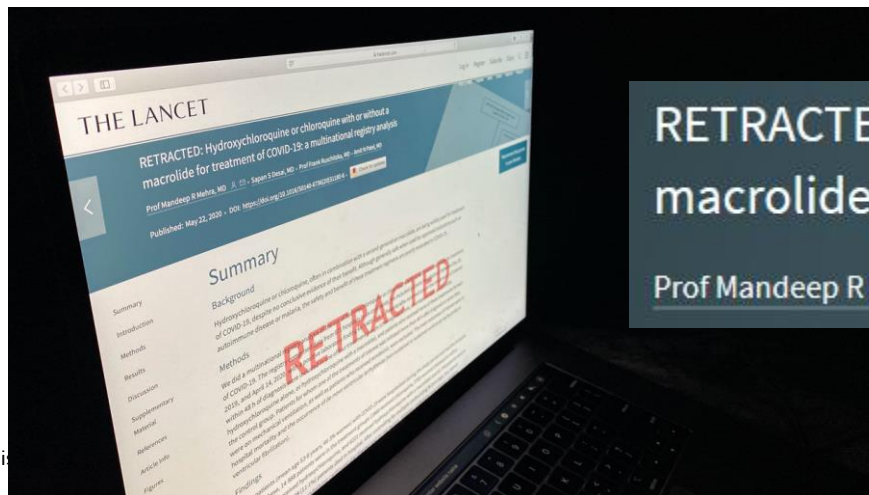
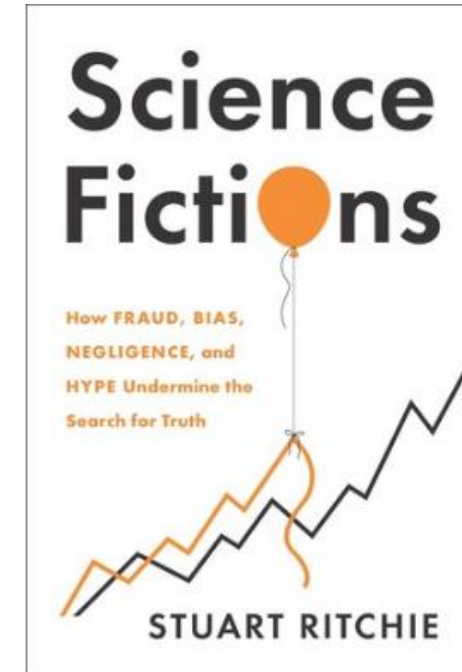
Further reading:

- [Replication crisis – Wikipedia](#)
- <https://phys.org/news/2017-03-science-crisis.html>

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/>



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



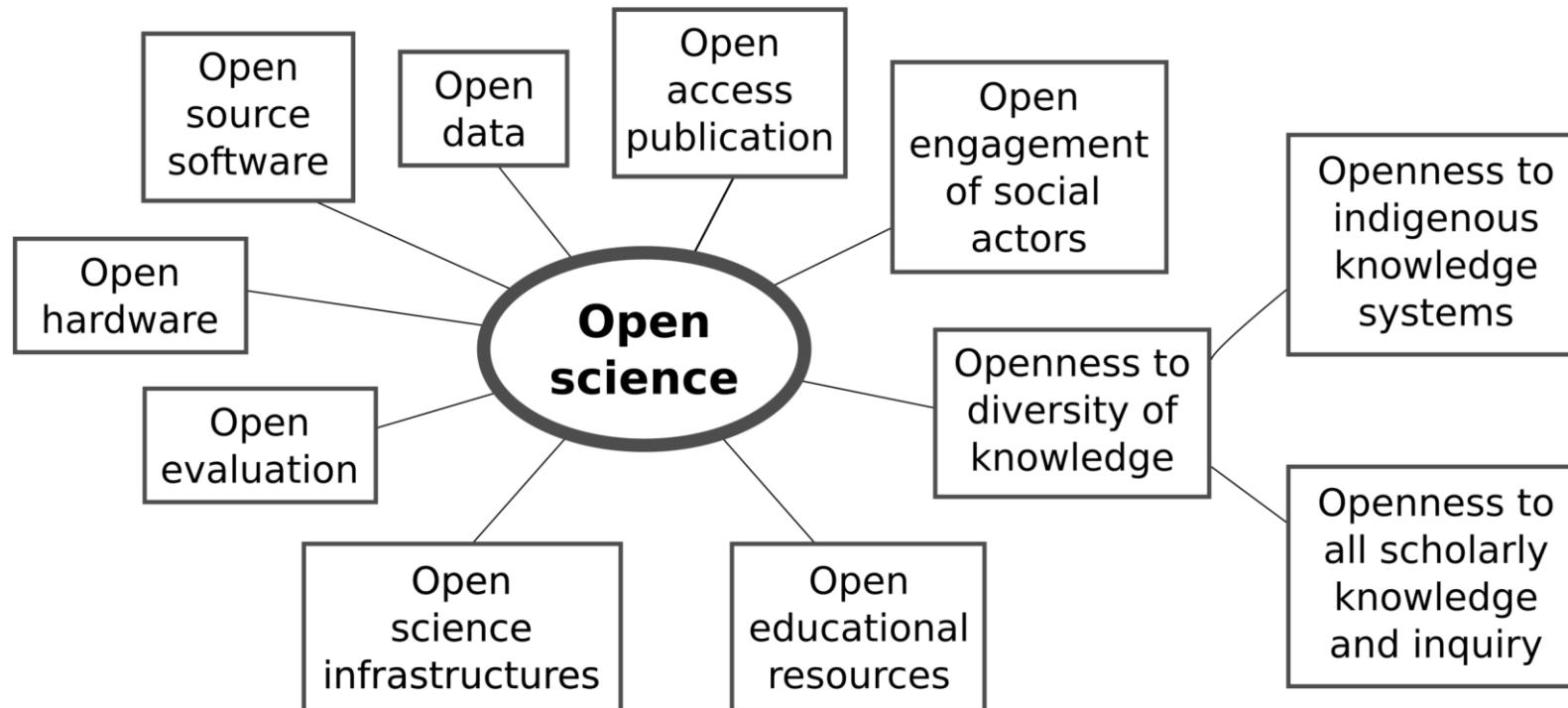
This

novation 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



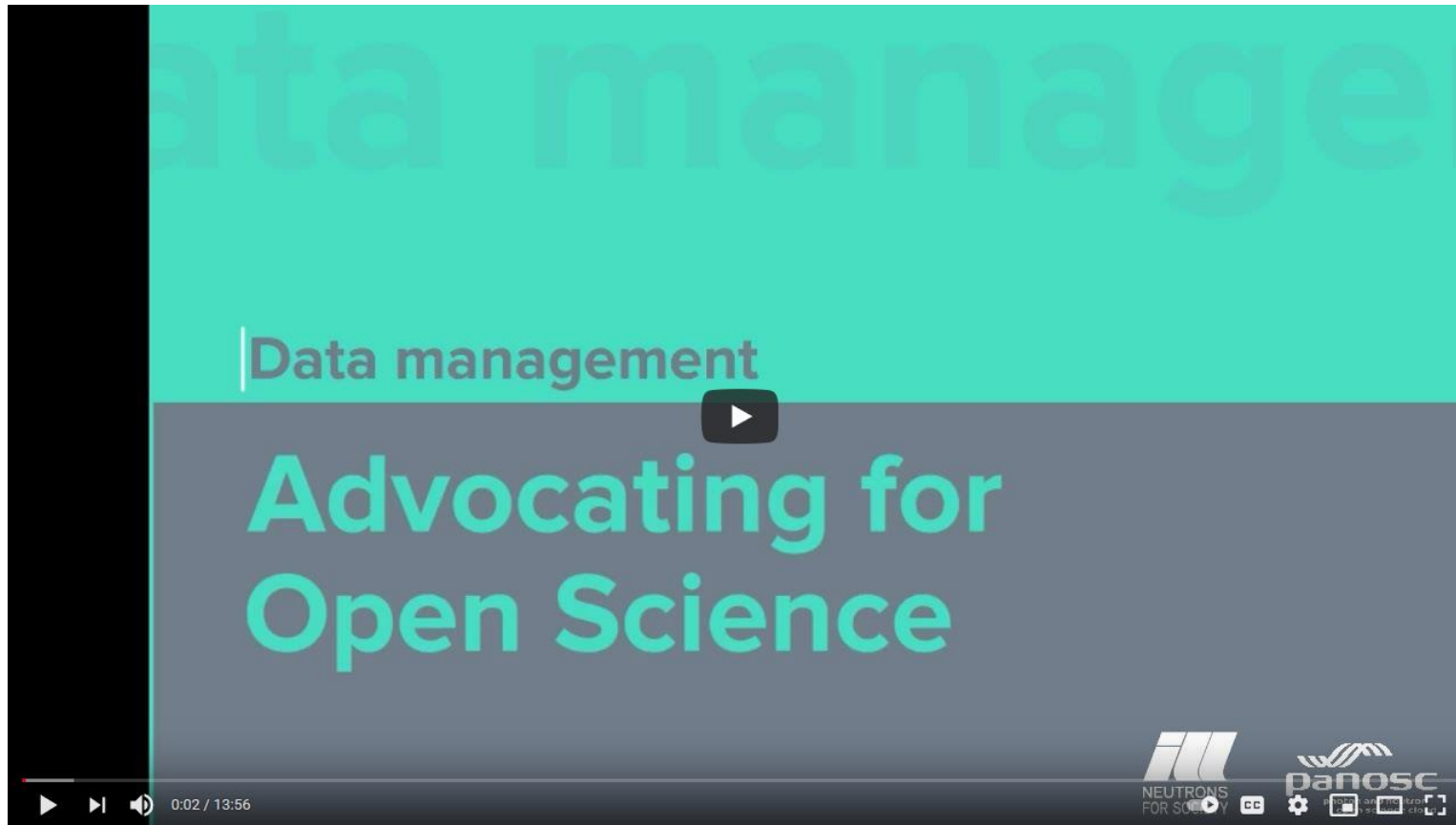
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

By RobbielanMorrison - 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](https://phys.org)
 - [Five questions about open science answered](#)
 - [Data sharing can offer help in science's reproducibility crisis](#)
- UNESCO
 - [Recommendation on Open Science](#)



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



Data management made simple

Data management made simple

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

[Quirin Schiermeier](#) in Nature (2018)

<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).



Data management made simple

Data management made simple

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

[Quirin Schiermeier](#) in Nature (2018)

<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.



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**.



EU Research facilities Data Policies

- **ESA** – open data policy for most data (since 2010)
- **ILL** – open data policy (since 2012)
- **EMBL** – open access policy (since 2015)
- **ESRF** – open data policy (since 2015)
- **ESO** – open data policy (updated in 2016)
- **EuXFEL** – open data policy (since 2017)
- **CERN** – open data policy for LHC (since 2020)
- **Laser labs**
 - **HZDR** – open data policy (since 2018)
 - **ELI-ERIC** – open data policy (December 2021)
 -



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)

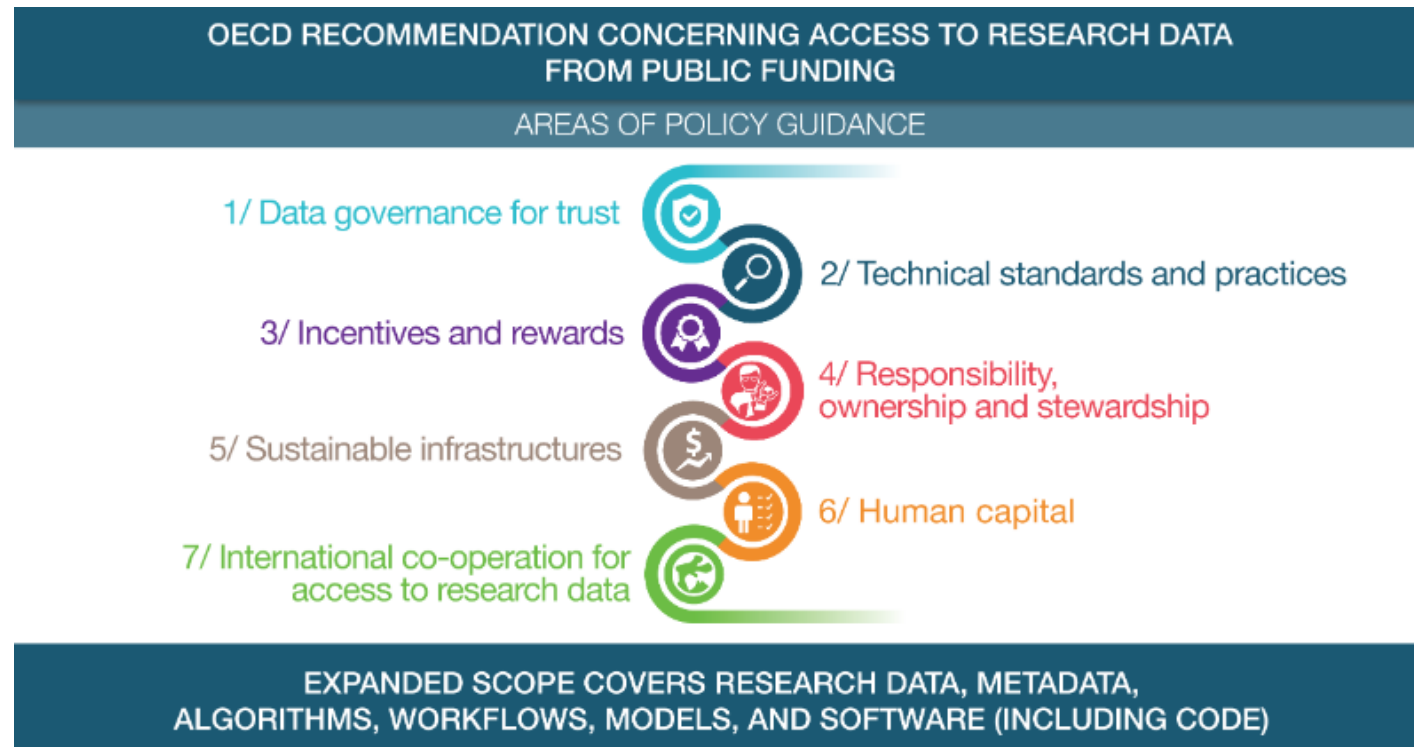


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852



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>)



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

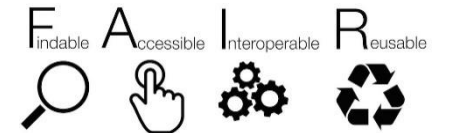
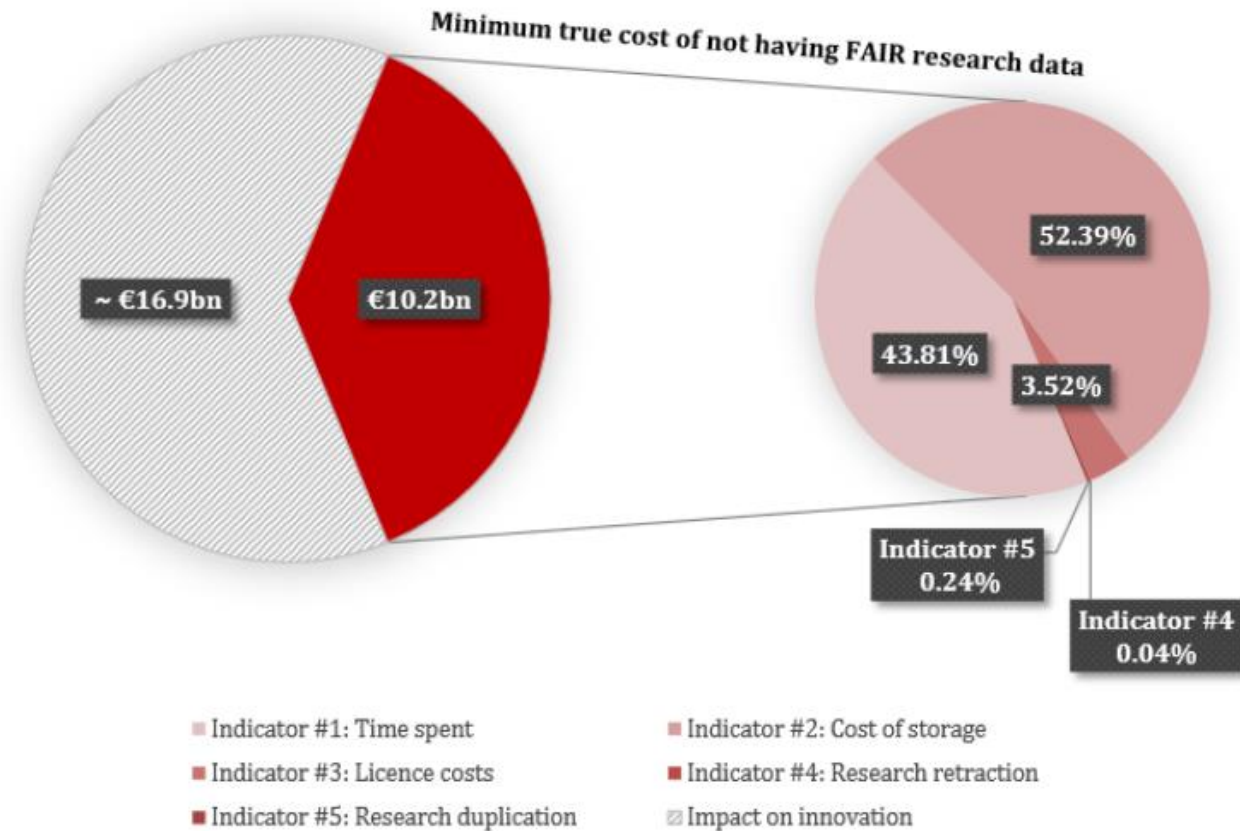
Reusable

- › 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



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

Likely cost of not having FAIR research data



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



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 :
 - Raw data
 - Metadata
 - Processed data
 - Analysis workflows
 - Logbooks
 - Software
 - Etc.



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](#), ...



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



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



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.

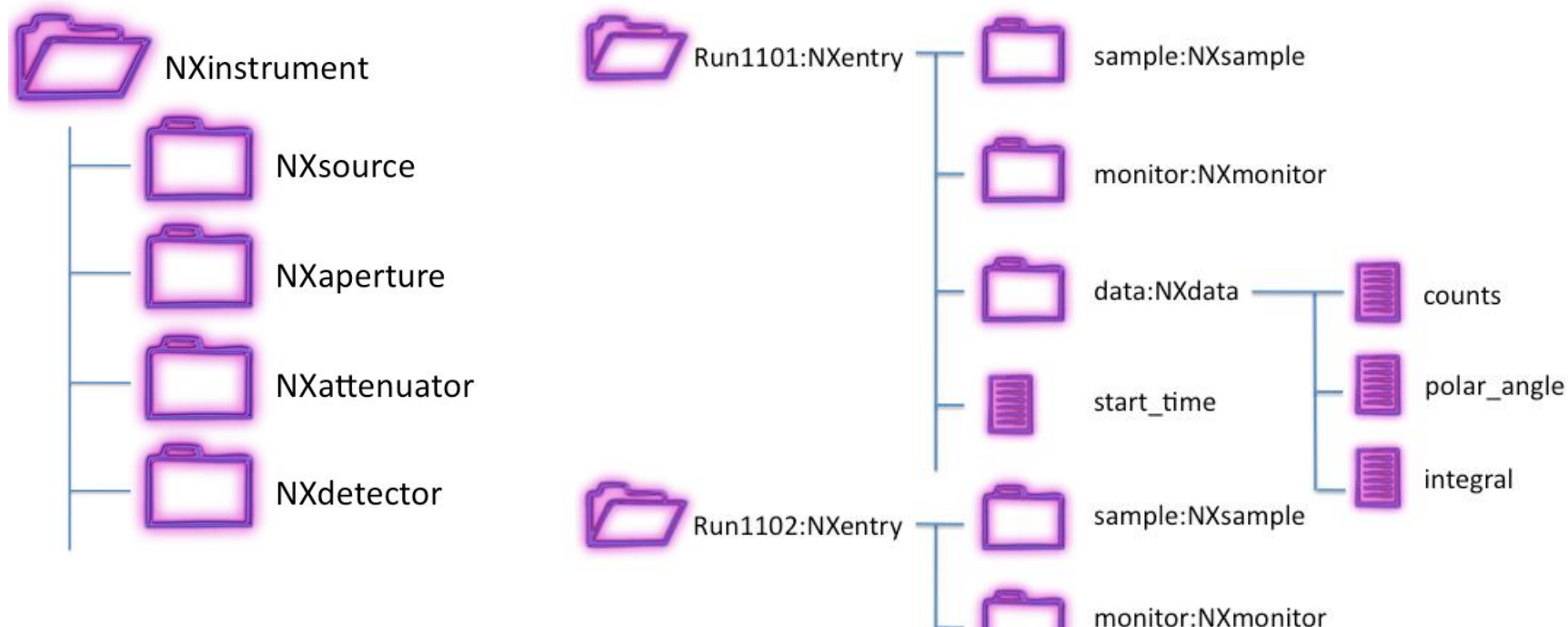
Home

GitHub Organisation

© 2021 NIAC

<https://www.nexusformat.org/>

Nexus provides a standard vocabulary for:



European Un

openPMD is another example <https://www.openpmd.org>

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				
data	3D data		100 × 516 × 516	
detector_information				
header		NXcollection		
images		NXcollection		
raw_data	3D data	NXdata		
soft_data	3D data	uint16	100 × 516 × 516 Soft	
start_time	"2020-09-08...	string	scalar	
title	"Lima 2D de...	string	scalar	

Adopt NeXus for Lasers?

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



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](#)



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>



<https://elog.psi.ch/elog/>

The ELOG Home Page

Site map : [Home] [User's Guide] [Administrator's Guide] [FAQ] [Forum] [Wishlist]

* this section : [What is ELOG ?] [Screen shots] [Use cases] [License]

Home of the *Electronic Logbook* package by [Stefan Ritt](#)

Current version is : 3.1.4

What is ELOG ?

ELOG is part of a family of applications known as *weblogs* . Their general purpose is :

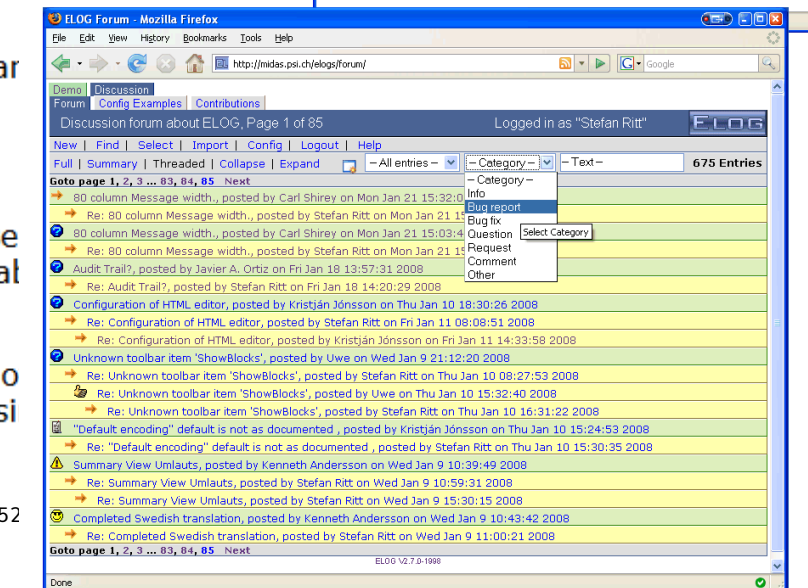
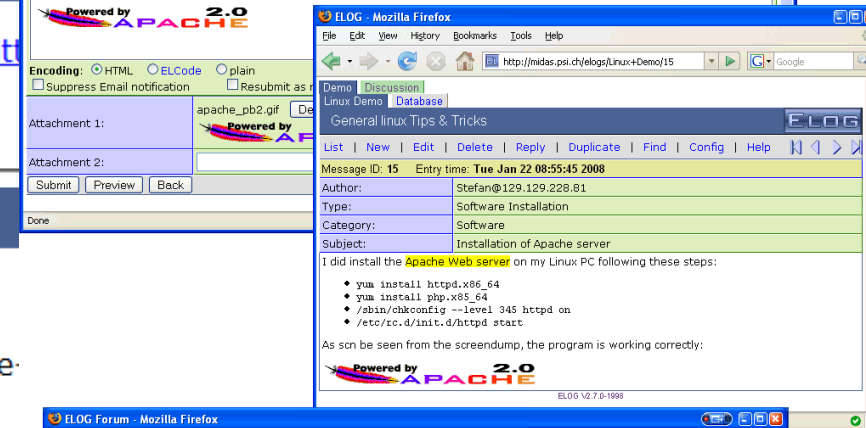
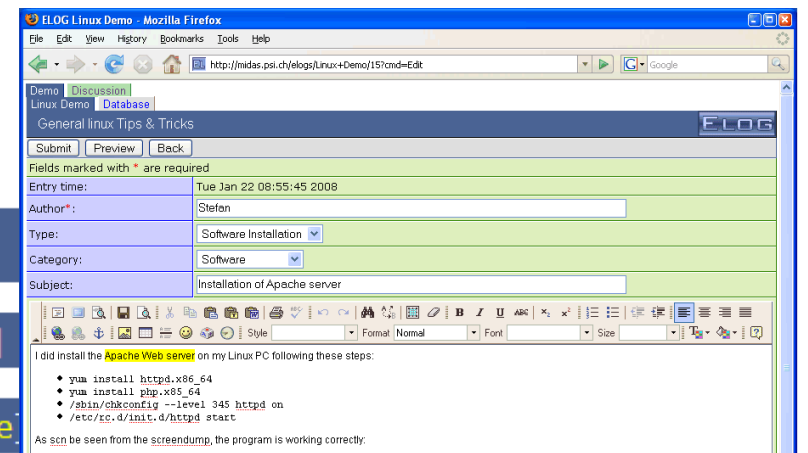
1. to make it easy for people to put information online in a chronological fashion, in the form of short, time-markup for presentation, and optional file attachments (images, archives, etc.)
2. to make it easy for other people to access this information through a Web interface, browse entries, search or comment on entries.

ELOG is a remarkable implementation of a *weblog* in at least two respects :

- its simplicity of use : you don't need to be a seasoned server operator and/or an experimented database (under Unix or Windows), a simple configuration text file, and it works. No Web server or relational data to the appropriate language for your users.
- its versatility : through its single configuration file, **ELOG** can be made to display an infinity of variants or display, how to display it, what commands are available and to whom, access control, etc. Moreover, a site can be totally different from the rest.



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

Dataset List 90

Logbook

Shipping

+ New

Take a photo

View

PDF

Help

0 new log(s) arrived

Everywhere

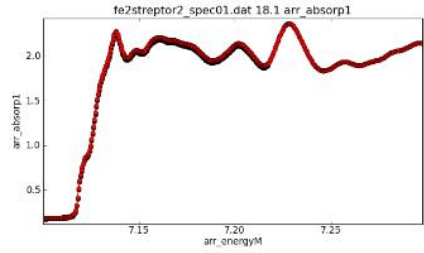
Everywhere

Q

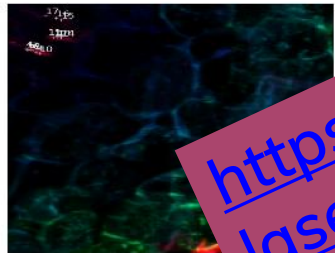
72 found

November 5th 2018

10:18:40 OPTICS> zapenergy 7.1 7.3 400 0.1 1 2 0 2000 (zap: #2, spec: #19)

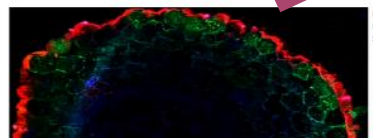


09:58:09 OPTICS> Fexanes_ev280



Fe2 strepto r2 Main Root: XANES Points

00:18:27 OPTICS> zapxiaimage samy 7.1 7.3 400 0.1 1 2 0 2000 (zap: #1, spec: #2)

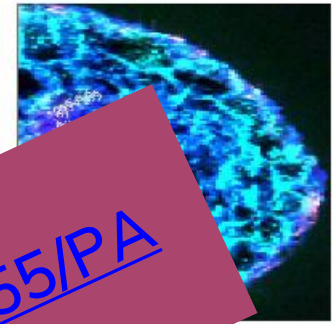


Fe [0.749] µg/g
Mn [0.1215] µg/g
S [0.6692] µg/g



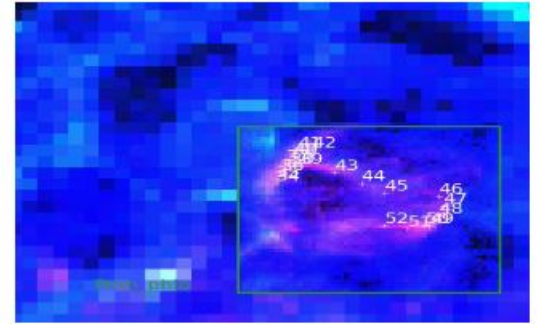
<https://indico.european-synchrotron-radiation-source.org/event/3/contributions/16/attachments/17/55/PANOSC%20ESRF%20LOGBOOK.pdf>

23:50:29 OPTICS> Fexanes_ev280



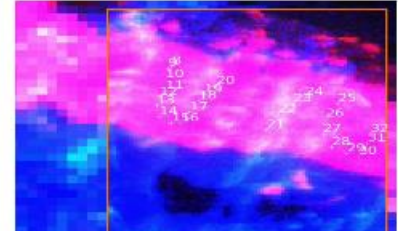
Fe3 prio r3 Main Root prim phlo: XANES Points

OPTICS> Fexanes_ev280



Fe3 prio r3 Main Root prim phlo: XANES Points

22:25:02 OPTICS> Fexanes_ev280



ment No. {

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



E-Life author guide



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

- 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.*



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 *de facto* the most popular 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%



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](#)



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 requirements.
- Plans for sharing data, who will have access, and any embargoes or restrictions.
- Data management roles.
- Costing or resource requirements for research and dissemination activities to enable data management in the shorter term following the end of any funded project.

Why don't we have more DMPs?

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

<https://study.sagepub.com/corti2e>



PaNOSC + ExPaNDS approach to DMPS

1. Keep it Simple (KISS)!
2. Ask as few questions as possible
3. Provide as many answers as possible
4. Use an existing tool to generate the DMP and manage the Knowledge Models
5. Integrate the DMPs into the user workflow
6. Help users make the most of their data

→ Avoid experiments unable to use the data



PaNOSC + ExPaNDS approach to DMPS

- Ask as few questions as possible **< 5**
- Provide as many answers as possible **> 50**
- Use an existing tool to generate the DMP
= DS Wizard
- Integrate the DMPs into the user workflow
= Proposal and Data portals
- Help users make the most of their data
= Providing RDM + Tools to process data
- Avoid experiments unable to use the data
= less than 10%



PaNOSC + ExPaNDS approach to DMPS

- Ask as few questions as possible < 5
 - Can you download, store, and treat the data you will receive during your experiment locally
 - Is the dataset reproducible in the sense that it could be created / collected anew in case it got lost?
 - Would you need additional processing infrastructure?
 - Does this dataset contain personal data?
 - Does this dataset contain sensitive data other than personal data? If yes, please describe the non-personal sensitive data used in the project.





Current Phase

Before starting the experiment

- Chapters**
- I. General / Topic ✓
 - II. Content classification / Datasets ✓
 - III. Technical classification / Date collection 2
 - IV. Data usage / Usage scenarios 5
 - V. Metadata and referencing / Metadata 3
 - VI. Legal and ethics / General legal issues 2
 - VII. Storage and long-term preservation / Selection 1

I. General / Topic

Chapter text

1 What is the scientific motivation for the experiment? +

Horizon 2020 DMP

Describe briefly the project and its aims. Provide a summary of the experiment.

Desirable: *Before starting the experiment*

Spatially-resolved local atomic fluctuations in bulk metallic glasses

Answered 15 minutes ago by dmpIngestor Ingestor.

2 Please give some keywords describing the research question. +

Horizon 2020 DMP

Desirable: *Before starting the experiment*

XRD-PDF-CT approach, Local Atomic Fluctuations, Bulk Metallic Glasses. Length-scale

Answered 14 minutes ago by dmpIngestor Ingestor.

3 Which research field(s) does this project belong to? +

Desirable: *Before starting the experiment*

- a. LS - Life Sciences
- b. CH - Chemistry
- c. ME - Engineering
- d. MA - Applied Materials Science
- e. MI - Methods and Instrumentations
- f. SC - Soft Condensed Matter Science
- g. HC - Hard Condensed Matter Science
- h. ES - Earth Sciences
- i. EM - Environment

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 preserving data over the long term. Repositories can be provided by institutions or community based. Choose the right repository which provides the service you expect

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

Setup a common repository for Lasers?



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?



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



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.



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 V 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>

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

DOI: [doi.esrf.fr/10.15151/ESRF-DC-217728238](https://doi.org/10.15151/ESRF-DC-217728238)



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

Data Portal My Data Open Data Closed Data Shipping ↕ Manager ↕ Log out Andy GÖTZ

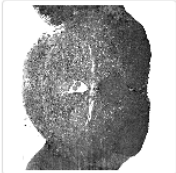
Open Data / 10.15151/ESRF-DC-217728238

Dataset List 4

Search

<input type="checkbox"/>	Date ↕	Sample ↕	Dataset ↕	Definition ↕	Files ↕	Size ↕	Download ↕	🔍
<input type="checkbox"/>	🕒 18:34 3 Jul 2020	Drosophila	drBrain		11	152.6 GB	Download	🔍

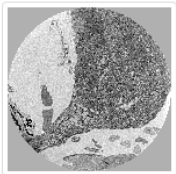
Summary Files 11 Metadata List

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

[/data/id16a/inhouse2/staff/ap/dataNatNeuro2020/Drosophila/drBrain](#) [Download](#)

<input type="checkbox"/>	🕒 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](#)

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



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



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](https://orcid.org/) identity:
 - You will be distinguished from other researchers who share your same name,
 - Your research outputs will be correctly attributed to you,
 - Your contributions will be reliably and easily connected to you,
 - You will save time by filling out forms, (leaving more time for research!),
 - You will enjoy increased 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.

Check out Umbrellaid too
<https://umbrellaid.org/>



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



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 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



Example of open data publishing:

Human Organ Atlas

Reference: “Multiscale three-dimensional imaging of intact human organs down to the cellular scale using hierarchical phase-contrast tomography” by Walsh, C.L., Tafforeau, P., Wagner, W.L., Jafree, D.J., Bellier, A., Werlein, C., Kühnel, M.P., Boller, E., Walker-Samuel, S., Robertus, J-L., Long, D.A., Jacob, J., Marussi, S., Brown, E., Holroyd, N., Jonigk, D.D., Ackermann, M., Lee, P.D. Nat Methods (2021) Accepted

Human Organ Atlas EXPLORE SEARCH

Welcome to the Human Organ Atlas

The Human Organ Atlas uses **Hierarchical Phase-Contrast Tomography** to span a previously poorly explored scale in our understanding of human anatomy, the micron to whole intact organ scale.

Histology using optical and electron microscopy images cells and other structures with sub-micron accuracy but only on small biopsies of tissue from an organ, while clinical CT and MRI scans can image whole organs, but with a resolution only down to just below a millimetre. HiP-CT bridges these scales in 3D, imaging intact organs with ca. 20 micron voxels, and locally down to microns.

We hope this open access Atlas, enabled by the ESRF-EBS, will act as a reference to provide new insights into our biological makeup in health and disease. To stay up to date, follow [@HiP-CT](#)

Funding

This project has been made possible by funding from:

- The [European Synchrotron Radiation Facility \(ESRF\)](#) — funding proposal MD-1252
- The [Chan Zuckerberg Initiative](#), a donor-advised fund of the Silicon Valley Community Foundation
- The [German Registry of COVID-19 Autopsies \(DeRegCOVID\)](#), supported by the German Federal Ministry of Health
- The Royal Academy of Engineering, UK
- The UK Medical Research Council
- The Wellcome Trust

Collaborators

- [UCL](#), London, England: **Peter D Lee, Claire Walsh, Simon Walker-Samuel, Rebecca Shipley, Sebastian Marussi, Joseph Jacob, David Long, Daniyal Jafree, Ryo Torii, Charlotte Hagen**
- [ESRF](#), Grenoble, France: **Paul Tafforeau, Elodie Boller**
- Medizinische Hochschule Hannover, Germany: **Danny D Jonigk, Christopher Werlein, Mark Kuehnel**
- Universitätsmedizin der Johannes Gutenberg-Universität Mainz, Germany: **M Ackermann**
- University Hospital of Heidelberg, Germany: **Willi Wagner**
- Grenoble Alpes University, Department of Anatomy, French National Center for Scientific Research: **A Bellier**
- [Diamond Light Source](#), Harwell, UK: **Andy Bodey, Robert C Atwood**
- Imperial College London, UK: **JL Robertus**

Reference

* Walsh, C.L., * Tafforeau, P., * Wagner, W.L., Jafree, D.J., Bellier, A., Werlein, C., Kühnel, M.P., Boller, E., Walker-Samuel, S., Robertus, J-L., Long, D.A., Jacob, J., Marussi, S., Brown, E., Holroyd, N., Jonigk, D.D., Ackermann, M., Lee, P.D. Imaging intact human organs locally resolving cellular structures using hierarchical phase-contrast tomography. Nat Methods (2021) Accepted

Acknowledgements

The development of this portal has been done as part of the [PaNOSC project](#). PaNOSC has received funding from the European Union's [Horizon 2020](#) research and innovation programme under grant agreement No. 823852. The following people were involved in the development: Paul Tafforeau, Alejandro De Maria Antolinos, Axel Bocciarelli, Marjolaine Bodin and Andrew Götz from the ESRF, Jiří Majer from UCL as well as the broader PaNOSC and ICAT communities.

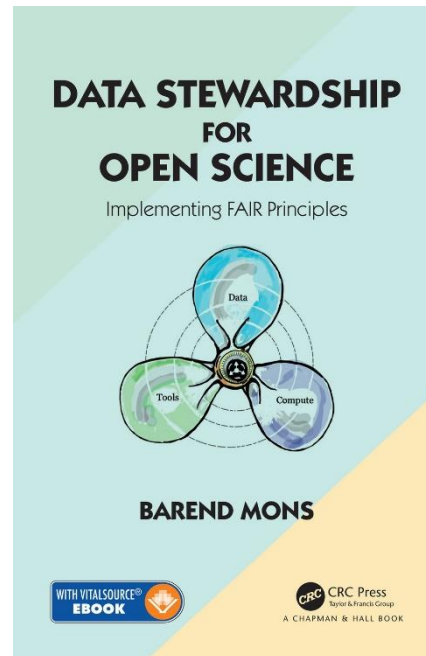
<https://human-organ-atlas.esrf.fr/>

53 PaNOSC

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:



Tools to help you manage your research

A non-exhaustive list of tools to explore

- Open science framework – osf.io
- [Protocols.io](https://protocols.io)
- [Fairsharing.org](https://fairsharing.org)
- [Jupyter.org](https://jupyter.org) notebooks




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 scientific output than text publications ...



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



Thank you

andy.gotz@esrf.fr

