School on Synchrotron Light Sources and their Applications

Open Science and FAIR data

15th January, 2025

Author: Andy Götz (ESRF)

Role: ESRF Data Manager, PaNOSC coordinator, EOSC Science Officer

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 Open Science and FAIR Data for scientists doing research in order to answer the following questions:

- **Open Science and FAIR Data**

What is this ?

Why do this ?

What to do ?

What to expect ?

What to learn ?

Andy Götz

1. Started as radio astronomer (**1984 – 1987)**

2. Joined **ESRF** in **1988**, to work on **accelerator controls, beamline controls, data management, manager**

3. Member of the **IUCr Commission on Data (2015 – now)**

4. Coordinator of the EU H2020 **PaNOSC** project (https://panosc.eu) on **making FAIR data reality** for Photon and Neutron sources in Europe **(2018 – 2022)**

5. **Science Officer** of the **European Open Science Cloud** (2024 – now)

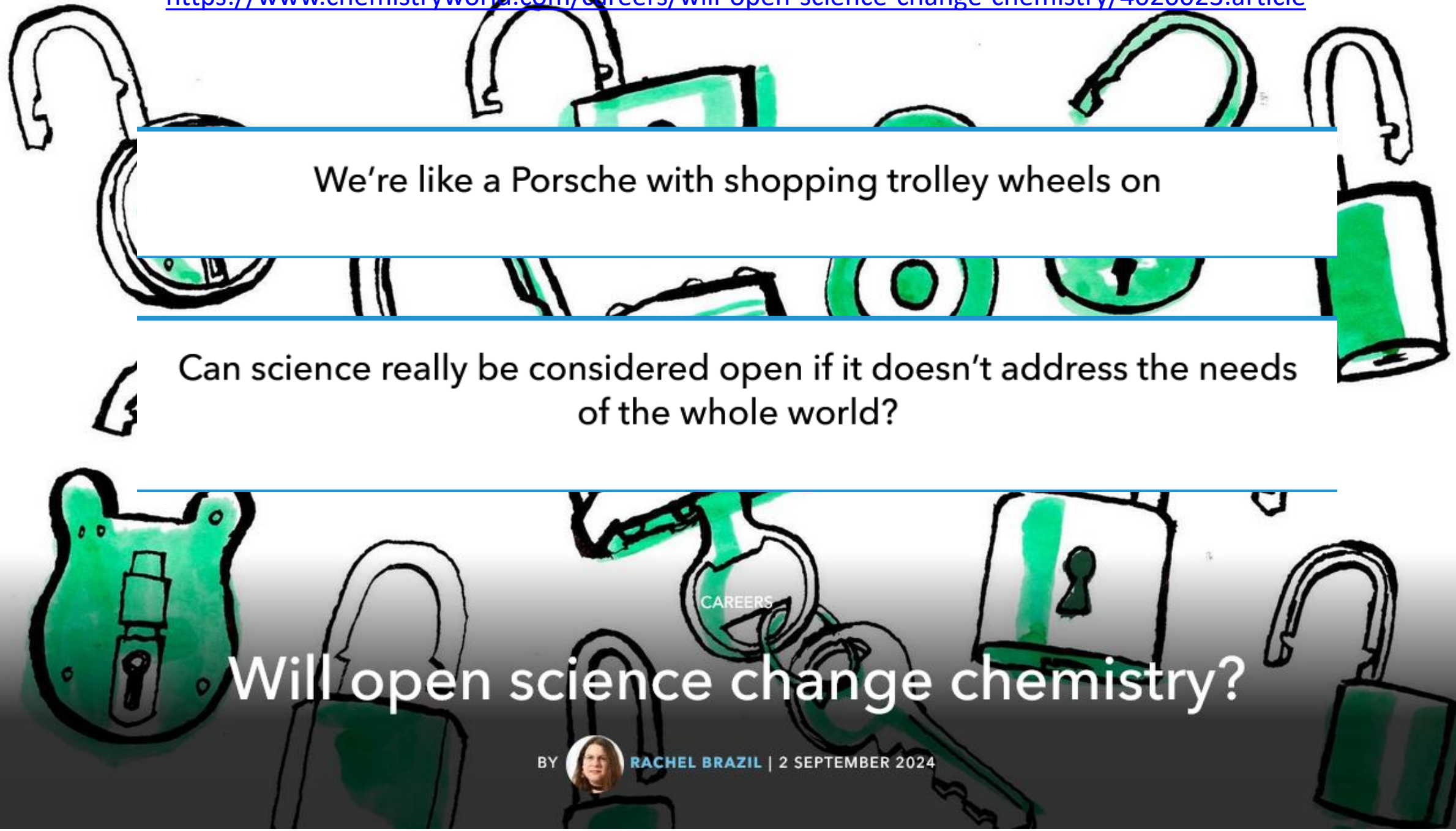
Open Science is a major movement Worldwide to change the way Science is conducted

Why a talk about Open Science + FAIR data?

FAIR data is one of the pillars of OS

panosc

https://www.chemistryworld.com/careers/will-open-science-change-chemistry/4020023.article



We're like a Porsche with shopping trolley wheels on

Can science really be considered open if it doesn't address the needs of the whole world?

CAREERS

Will open science change chemistry?

BY RACHEL BRAZIL | 2 SEPTEMBER 2024

Definition of Science by the Science Council

Science is the pursuit and application of knowledg
and social world following a systematic me

Scientific methodology includes the foll

- Objective observation: M
  mathematics as a
- Evidence
- Experiment and/or
- Induction: reasoning
- Repetition
- Critical analysis
- Verification and testing:

n from facts or examples

utiny, peer review and assessment

Science is Evidence based
the Evidence includes all data
Contributing to the Publication

https://sciencecouncil.org/about-science/our-definition-of-science/

OSCARS
Open Science Clusters' Action for Research & Society

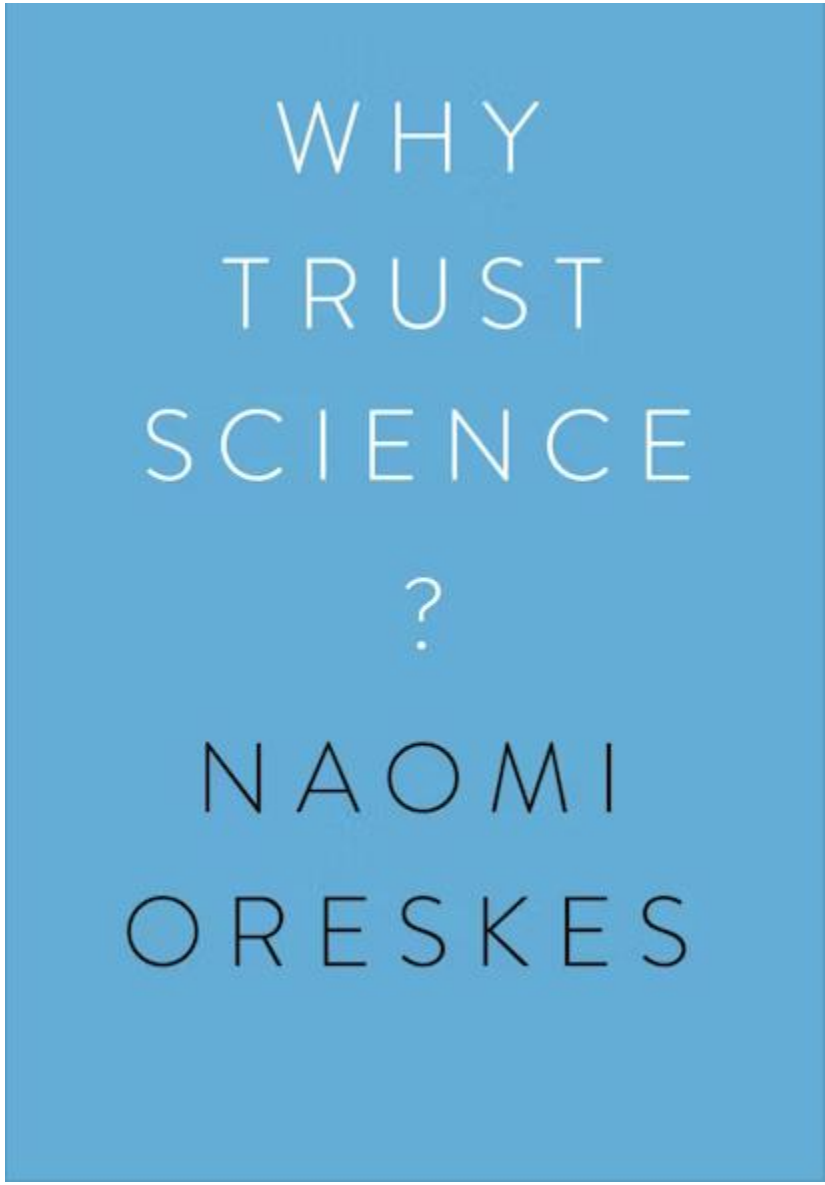
panosc

Why trust science?

- Oreskes's position is that **scientific claims are trustworthy when they reflect a consensus in scientific communities that is structured in the right way**—for example, by being diverse, open to democratic norms like transparency, and serving as a venue for the critical uptake of ideas.
- **Trust in science** for Oreskes is **rooted in science's social character.** This **trust is informed rather than blind**, as it **allows outsiders to assess claims** by judging the character of the scientific communities making the claims.

https://pmc.ncbi.nlm.nih.gov/articles/PMC9425825/

https://www.youtube.com/watch?v=t7PwqiiQmVM



Science Creates Knowledge Through a Community Effort

Shared Practices Increase the Accuracy of Scientific Findings

Scientific Thinking Is Continually Refined by New Evidence

Ensuring that Science Remains Trustworthy
Requires Constant Vigilance

Why Trust Science?

https://whytrustscience.org.uk/essay/

panosc

Why is science considered “closed” ?

Why do we need open science?

- **Results** (publications) from publicly funded research are published in journals which require a (often expensive) subscription i.e. behind a pay-wall and not available to the general public

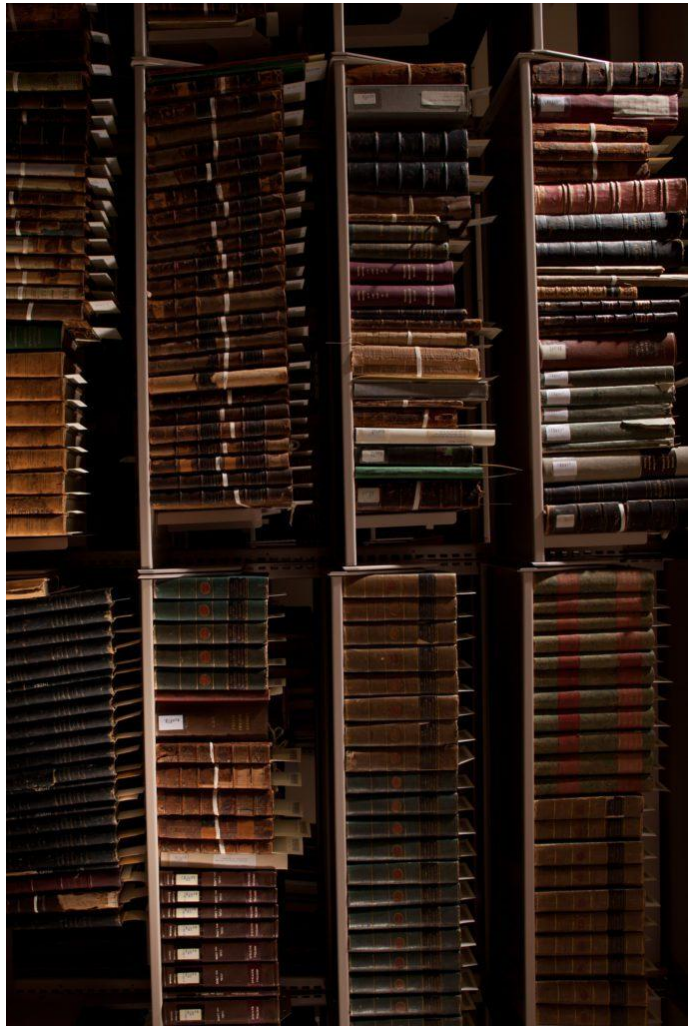
- **Evidence** (scientific data, protocols, software, hardware etc.) are not made available as part of the publication or otherwise made available

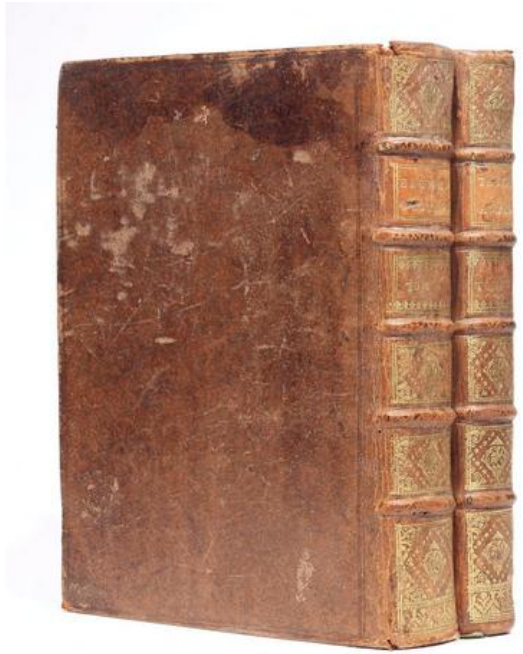
- **Community** (society) is not consulted on what topics research should be done

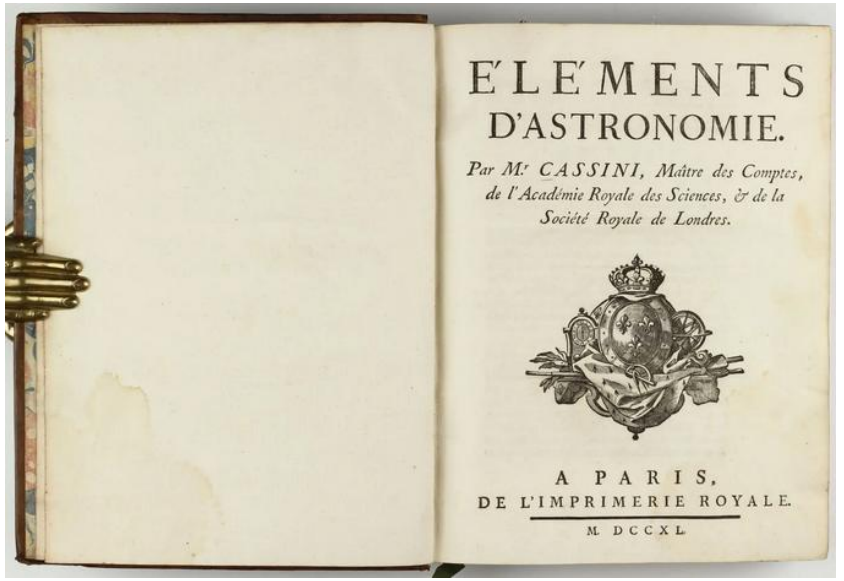
panosc

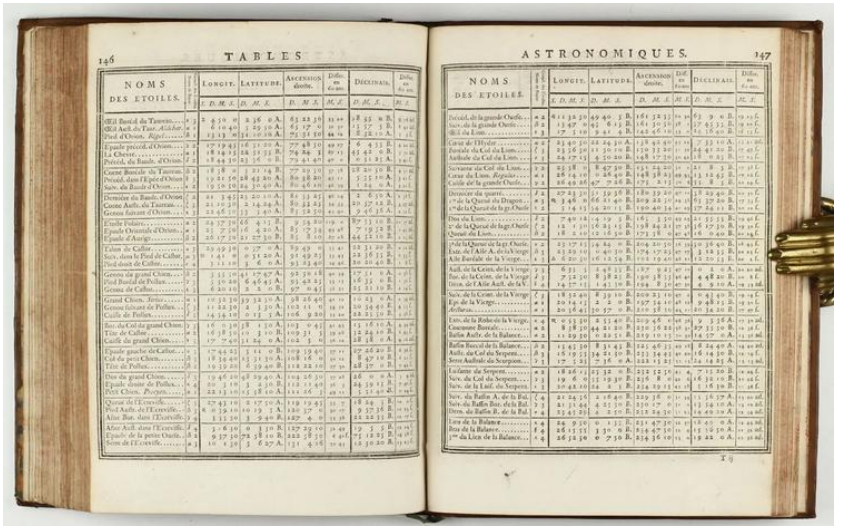
Evolution of publishing data

- In the beginning data were published as tables in the Annals of the journals …



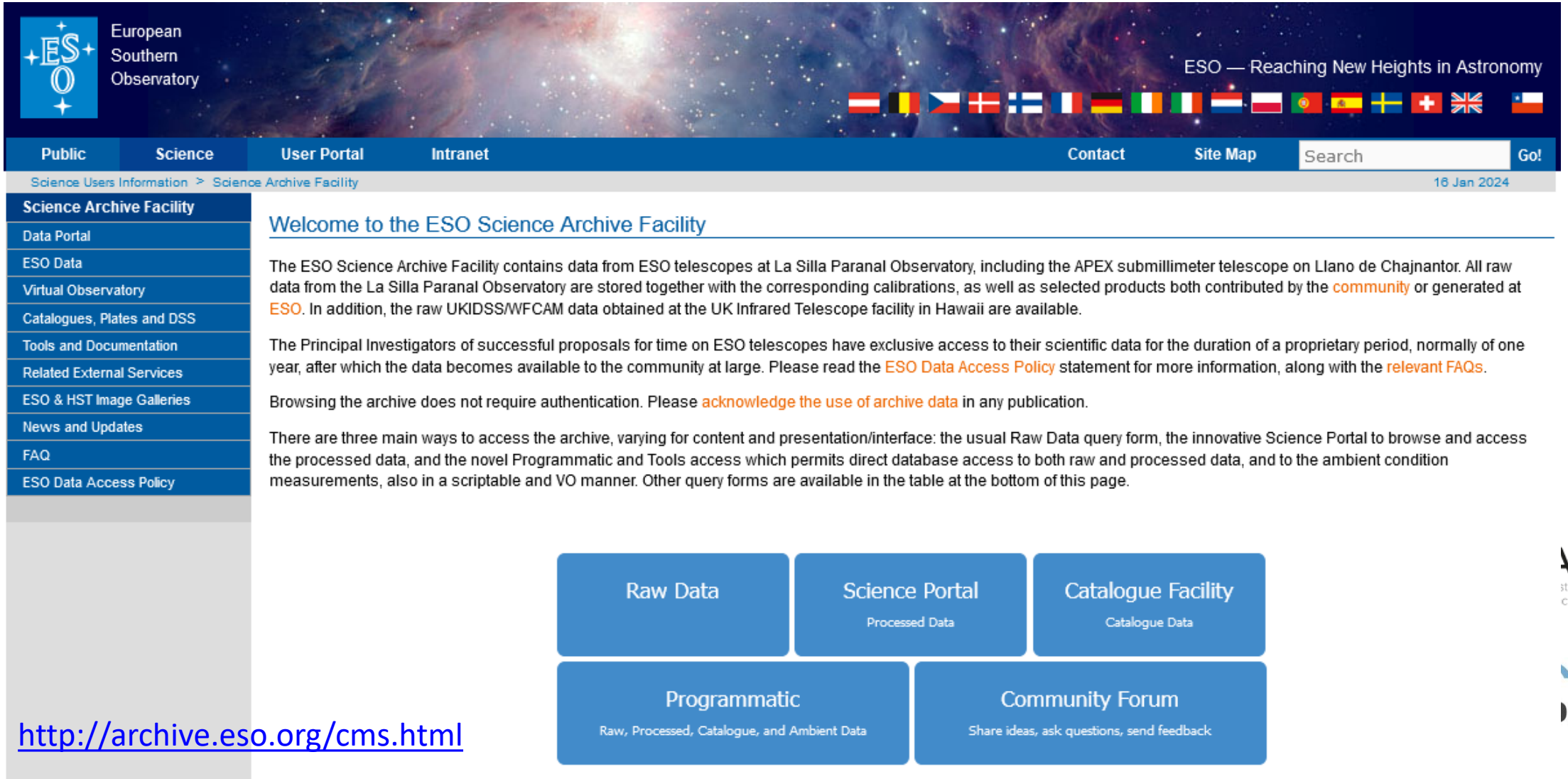






Digital technology makes publishing data easier

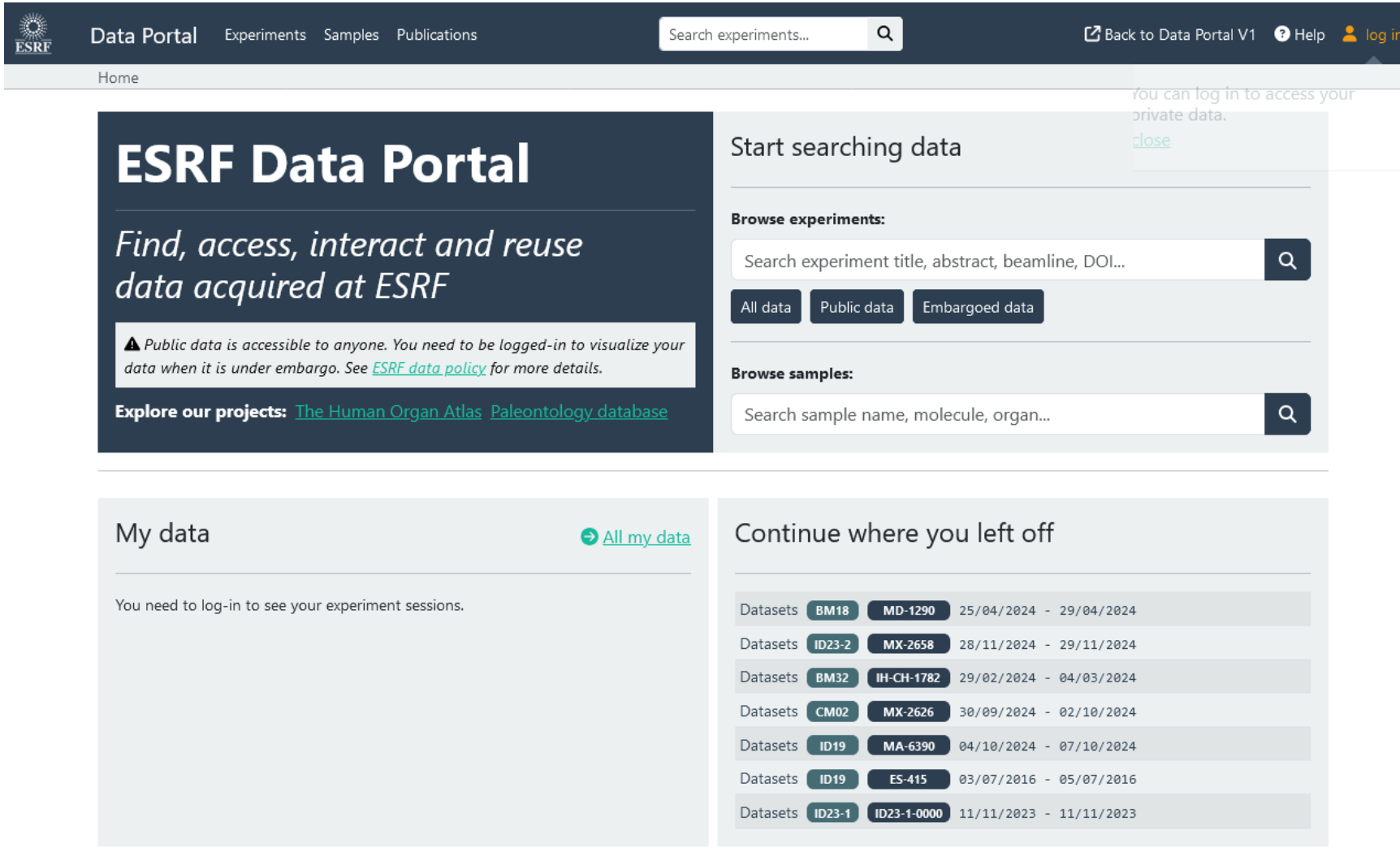
- Today they are published via databases and digital media …



http://archive.eso.org/cms.html

ESRF data portal manages petabytes of raw data and recently processed data, since 2015

Demonstrates that large volumes of data can be managed efficiently

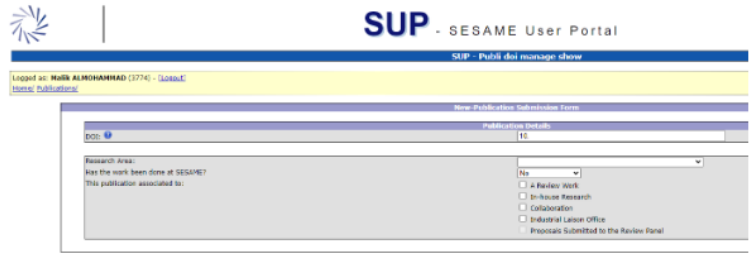


https://data.esrf.fr

Synchrotron facilities now propose data repositories

Example of SESAME – thanks to Malik AlMohammad + Salman Matalgah

SESAME Data Portal



Cite Data DOIs

Dataset List 21 | Logbook | Shipping | Samples | Proposal

For users that want to download large volume of experimental data (>2GB), ESRF users can access the Globus service, please read the documentation for proceeding: https://confluence.esrf.fr/display/SCKB/Globus

Filter by samples

Search

Date	Sample	Dataset	Definition	Files	Size	Processed	Download
16:46 13 Oct 2023		Se_EXAFS_Tatiana-20231013T091917		20	6.8 MB		Download
08:36 13 Oct 2023		Zn_EXAFS_Tatiana-20231013T071125		4	708.2 KB		Download
07:10 13 Oct 2023		Zn_EXAFS_Tatiana-20231013T070328		3	114.0 KB		Download
06:42 13 Oct 2023		Zn_EXAFS_Tatiana-20231012T203557		20	6.8 MB		Download
20:30 12 Oct 2023		Zn_EXAFS_Tatiana-20231012T202949		3	10.7 KB		Download
19:15 12 Oct 2023		Cd_EXAFS_Tatiana-20231012T115316		20	5.9 MB		Download
11:18 12 Oct 2023		Cd_EXAFS_Tatiana-20231012T031730		20	5.9 MB		Download
01:24 12 Oct 2023		Cd_EXAFS_Tatiana-20231011T180059		20	5.9 MB		Download
15:52 11 Oct 2023		Cd_EXAFS_Tatiana-20231011T145436		5	1010.8 KB		Download
14:53 11 Oct 2023		Cd_EXAFS_Tatiana-20231011T145207		3	40.6 KB		Download
14:51 11 Oct 2023		Cd_EXAFS_Tatiana-20231011T145036		3	22.9 KB		Download

Simple | LDAP

SESAME staff or long term visitors, please sign in with Your Sup Account

Session expired

Username

Password

Sign in

or sign in as anonymous

Don't have an account yet? Register now

I need further assistance

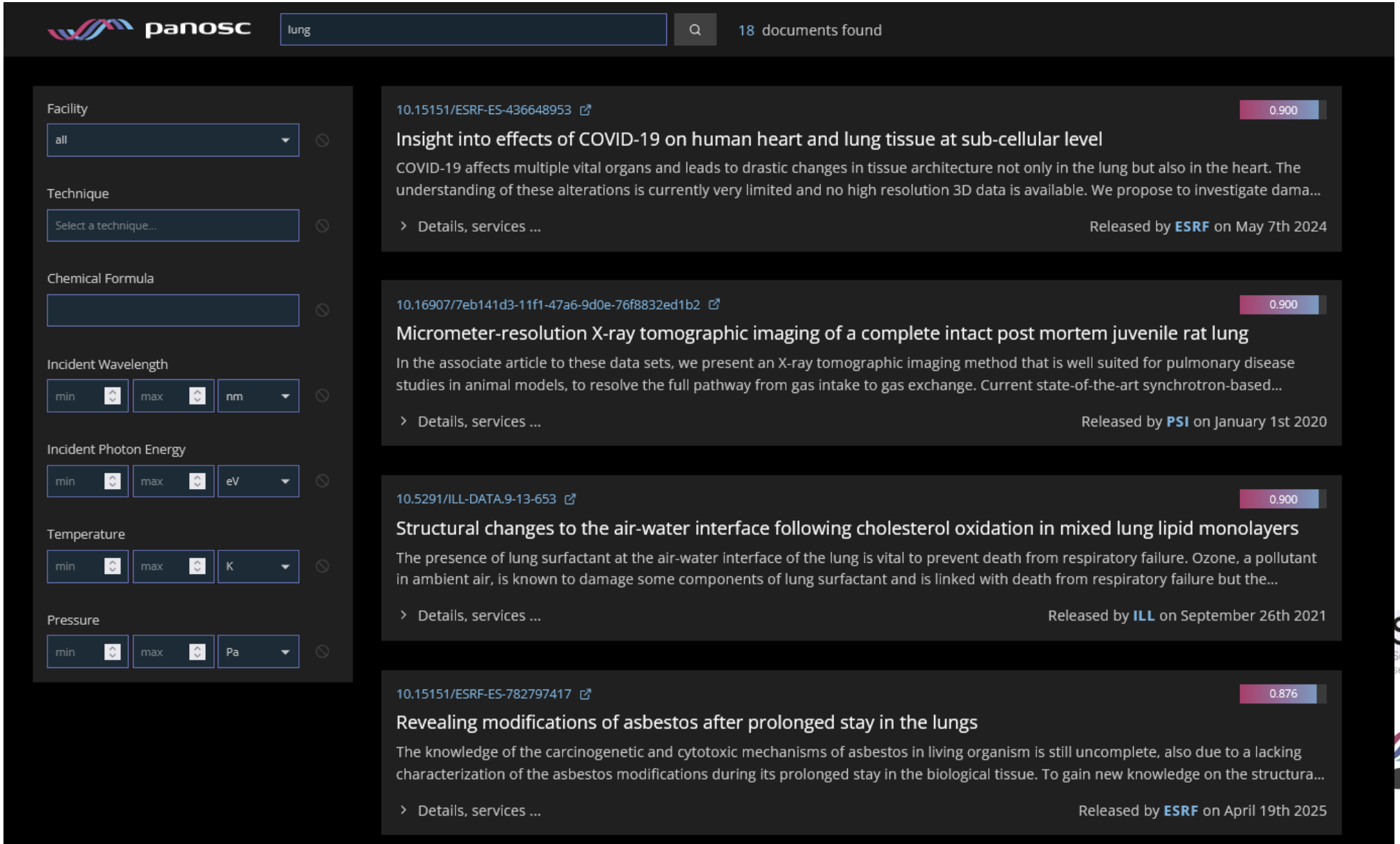
Important note

In order to login to the Data Portal of the SESAME or browse embargoed data you need to be declared a member of a proposal on the SESAME User Portal. Once this is done your account will be activated **45 days** before the first experiment starts. If you need access earlier please contact the SESAME User Office.
Anonymous login to browse open data is always possible.

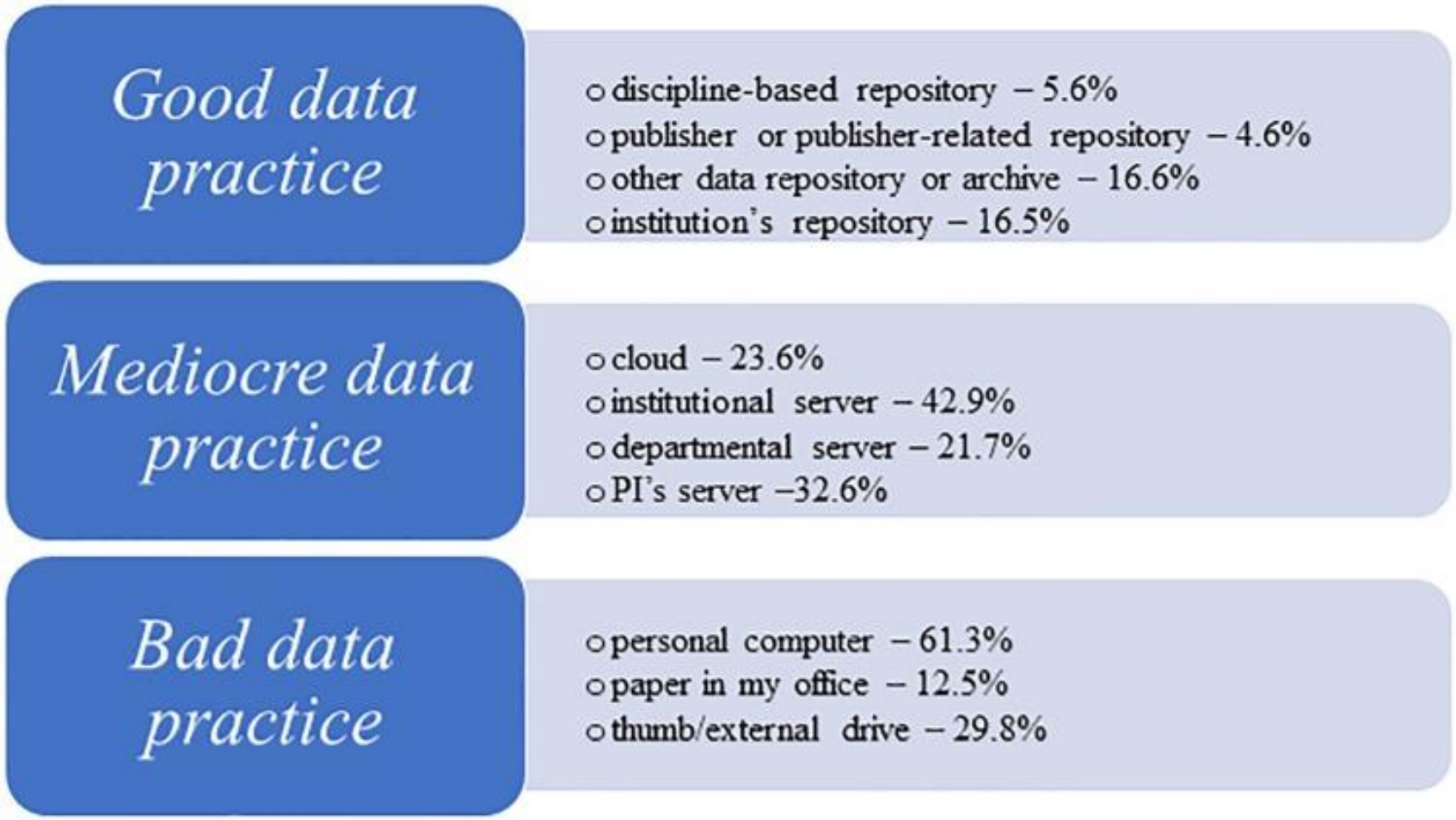
During and according to the General Data Protection Regulation, all SESAME Data portal users who did not consent to the User Portal Privacy Statement have had their account deactivated. Please contact the User Office if you wish to reactivate it.

Example of federated data search portal

https://data.panosc.eu/search/?q=lung



Adoption of good data practices are still poor



Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. PLoS ONE 15 (3): e0229003. https://doi.org/10.1371/journal.pone.0229003

panosc

Alphafold demonstrates the power of Open Data + AI

Over 2 million researchers in over 190 countries are using the AlphaFold Protein Structure Database to inform their research.

AlphaFold ... accelerated and enabled ... discoveries, including cracking the structure of the nuclear pore complex. And with this new addition of structures illuminating nearly the entire protein universe, we can expect more biological mysteries to be solved each day.

https://deepmind.google/technologies/alphafold/

The Nobel Prize in Chemistry 2024

© Nobel Prize Outreach. Photo: Clément Morin

David Baker

Prize share: 1/2

© Nobel Prize Outreach. Photo: Clément Morin

Demis Hassabis

Prize share: 1/4

© Nobel Prize Outreach. Photo: Clément Morin

John Jumper

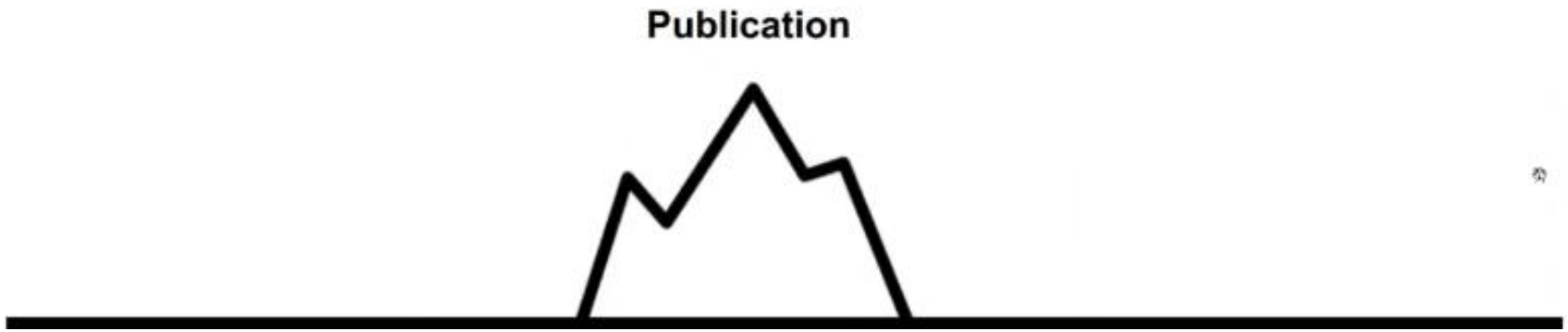
Prize share: 1/4

The Nobel Prize in Chemistry 2024 was divided, one half awarded to David Baker "for computational protein design", the other half jointly to Demis Hassabis and John Jumper "for protein structure prediction"

https://www.nobelprize.org/prizes/chemistry/2024/summary/

panosc

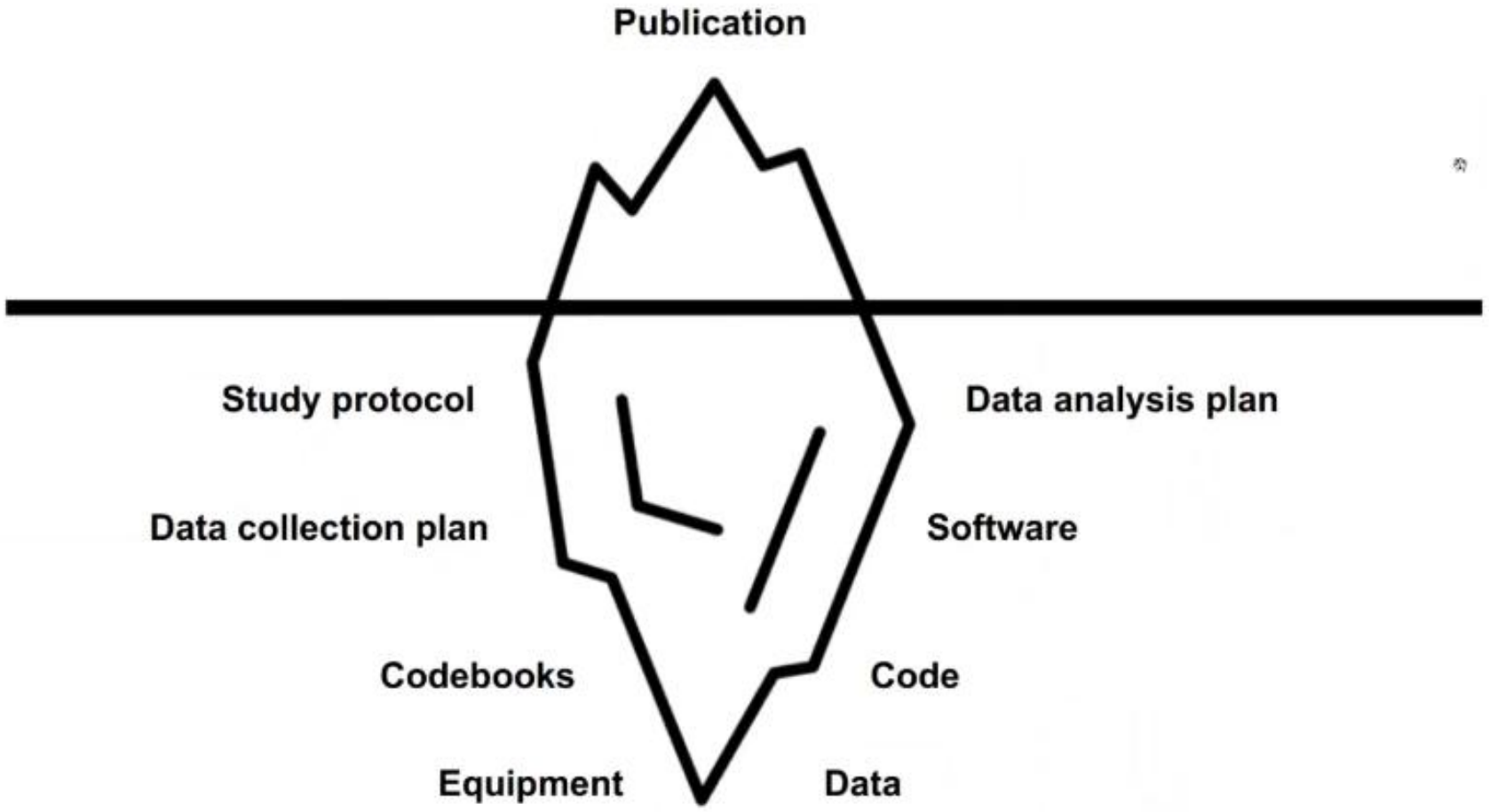
Science produces Publications



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

University of Saskatchewan

Science produces much more than Publications



https://library.usask.ca/studentlearning/workshops/grad-research.php#panel-section-3-ResearchDataManagementWhatYouNeedtoKnow

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

University of Saskatchewan

Open Science

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

Reproducibility and Replicability

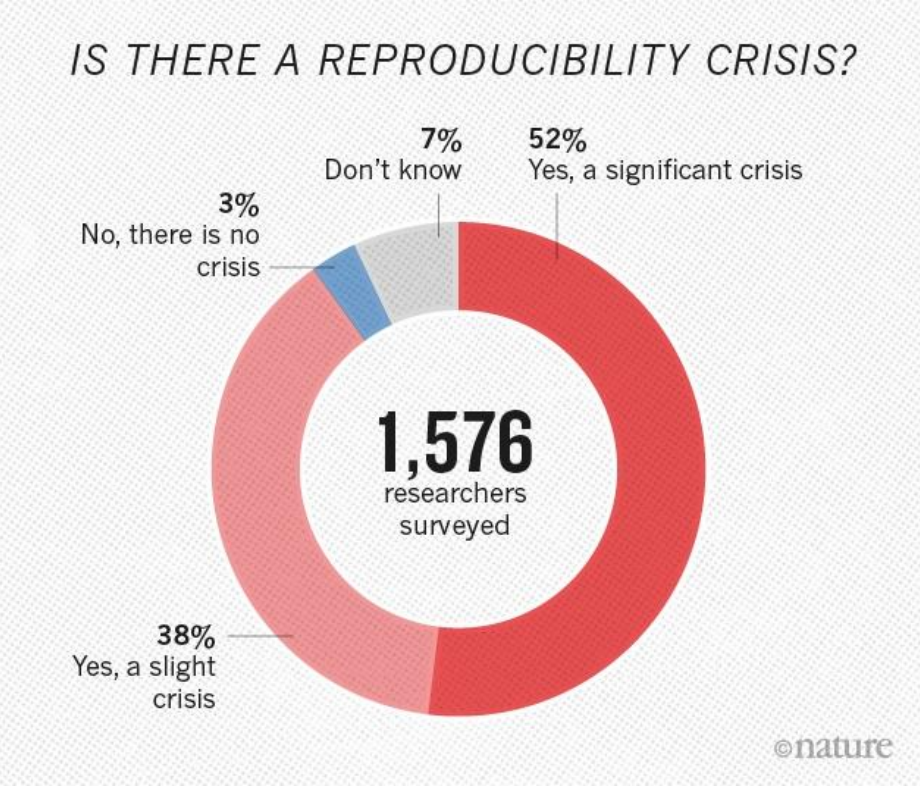
Published: 25 May 2016

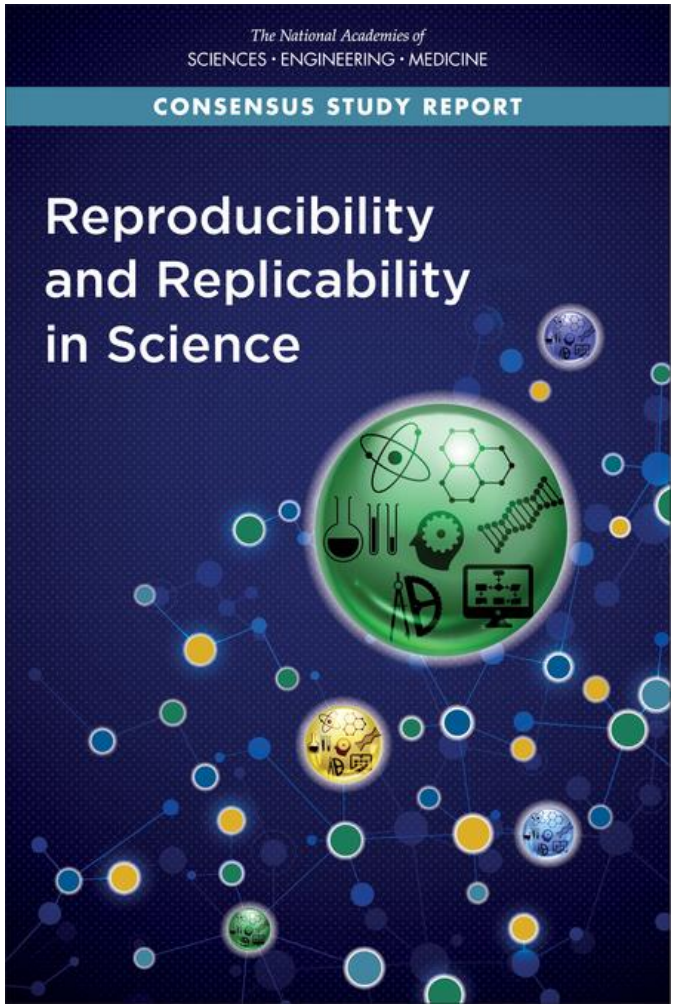
1,500 scientists lift the lid on reproducibility

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

ovation programme under grant agreement No. 823852

OSCARS Open Science Clusters' Action for Research & Society

panosc

What is Open Science?

- ✓ Open Science is: *"to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction"*

https://www.oecd.org/sti/inno/open-science.htm

"Work, Finish, Publish" → "**Work, Finish, Publish+Release**"

Michael Farady 1791-1867 Open Science

https://openscience.org/what-exactly-is-open-science/

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

Pillars of Open Science

1. **Open Access**
   - publications should be freely accessible either as Gold (journal) or Green (preprint) access
2. **Open Data**
   - data should be FAIR and freely accessible under a licence which allows re-use without restriction
3. **Open Source Software**
   - source code should be made available on a publicly accessible repository under an Open Source licence
4. **Open Hardware**
   - hardware designs should be accessible, like software, under an Open Source licence
5. **Open Educational Resources**
   - educational resources (videos, e-training courses etc.) should be made available to all
6. **Citizen Science**
   - citizens who follow the scientific method should be encouraged and facilitated and engage with scientists

OSCARS
Open Science Clusters' Action for Research & Society

panosc

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

Open Access publications – Green vs. Gold

GREEN

- Articles are free to read after an embargo period
- Bioscientifica automatically make the final published version, also known as the version of record, free
- Authors may deposit a version of their accepted manuscript in an online repository after this time
- There is no cost to authors.

GOLD

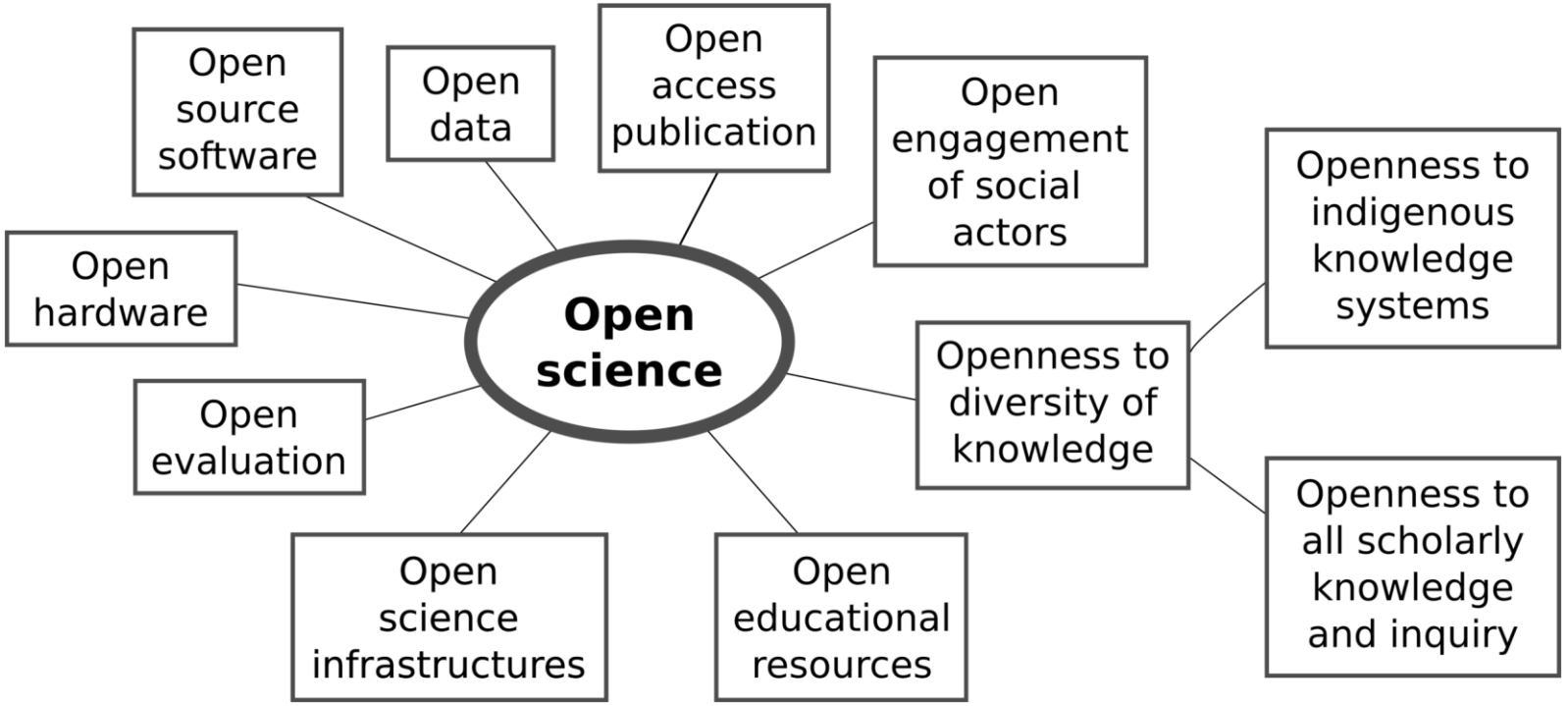
- Authors (or their funders or institutions) pay an Article Publication Charge (APC) upon acceptance
- The final published version is free immediately
- Bioscientifica deposits the article in PubMed Central
- Authors retain copyright and a range of licenses are available
- Journal could be fully open access (eg. *EDM Case Reports*) or hybrid (eg. *European Journal of Endocrinology*).

https://www.bioscientifica.com/authors/preparing-papers/publishing-open-access/

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

Unesco definition of open science

open science is defined as an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community.

United Nations
Educational, Scientific and
Cultural Organization

Open Science

OSCARS
Open Science Clusters' Action for Research & Society

panosc

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

https://unesdoc.unesco.org/ark:/48223/pf0000379949

ANNEX VI Recommendation on Open Science

This Recommendation outlines a common definition, shared values, principles and standards for open science *at the international level and proposes a set of actions conducive to a fair and equitable operationalization of open science for all at the individual, institutional, national, regional and international levels.*

https://unesdoc.unesco.org/ark:/48223/pf0000380399

Updated recommendations on the following:

1. **Scientific publications**
2. **Open research data**
3. **Open educational resources**
4. **Open source software and source code**
5. **Open hardware**
6. **Scientific knowledge**
7. **Open science infrastructures**
8. **Open engagement of societal actors**
9. **Open dialogue with other knowledge systems**
10. **Public + Private sector**

United Nations Educational, Scientific and Cultural Organization

Open Science

OSCARS Open Science Clusters' Action for Research & Society

panosc

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

European Open Science Cloud

The Vienna Declaration on the European Open Science Clou

Vienna, 23 November 2018

EU funded 22 EOSC Projects in H2020 + a further 19 EOSC Projects in Horizon Europe

https://eosc.eu/horizon-europe-projects/

African Open Science Platform

https://aosp.org.za

THE AFRICAN OPEN SCIENCE PLATFORM

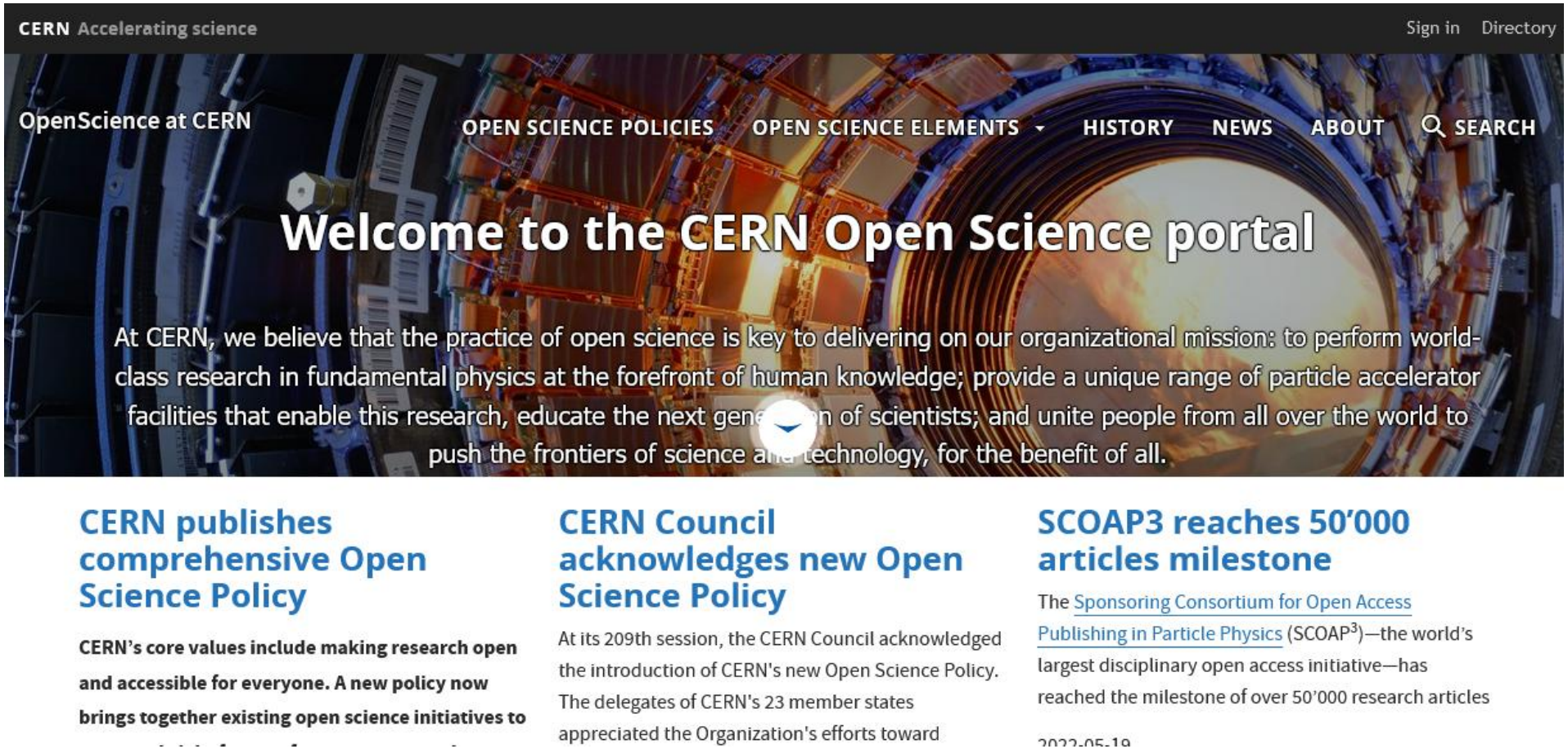
THE AFRICAN OPEN SCIENCE PLATFORM

About us | Partnerships | Membership | Initiatives | Resources | News & Events | Contact us

The Africa Open Science Platform (AOSP) was established in 2017 with an aim to position African scientists at the cutting edge of data intensive science by stimulating interactivity and creating opportunity through the development of efficiencies of scale, building critical mass through shared capacities, and amplifying impact through a commonality of purpose and voice.

CERN publishes Open Science policy

https://openscience.cern/

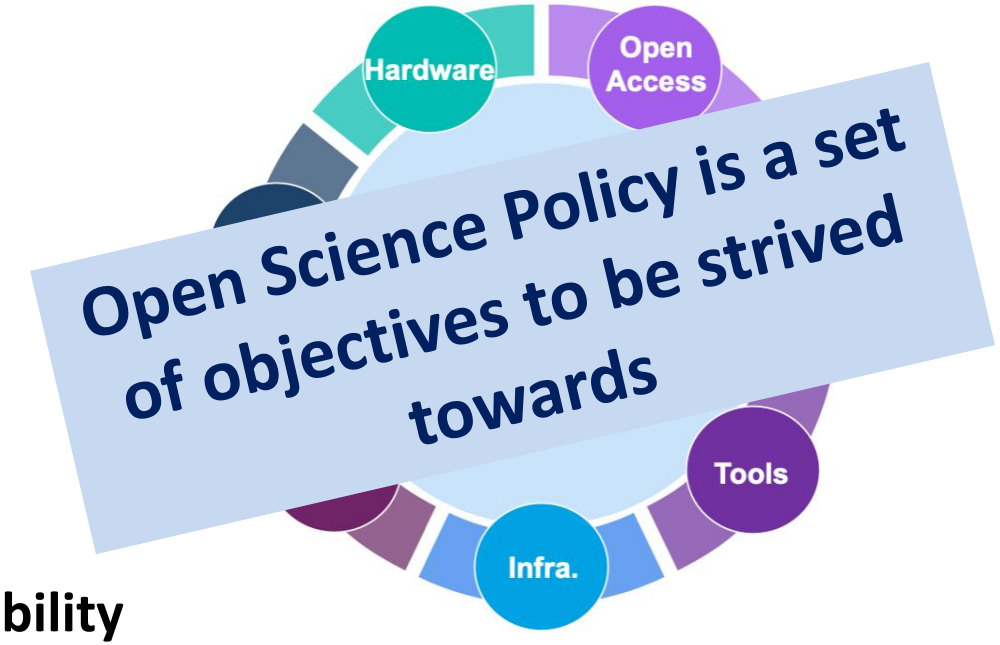


CERN Council adopted an Open Science Policy on 2022-09-29

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

CERN Open Science policy

1. **Open access to publications**
2. **Open data**
3. **Open source software**
4. **Open hardware**
5. **Research integrity, reuse and reproducibility**
6. **Infrastructure provision for open science**
7. **Research assessment and evaluation**
8. **Education, training and outreach**
9. **Citizen Science**



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

EMBL Open Science Policy

→ EMBL adopted an Open Science Policy for EMBL staff in December 2021 and is implementing it since January 2022.

→ Two main aspects:

1. Public availability of research outputs
2. Research assessment and fair attribution of credit

1. ORCID

2. DORA

EMBL Open Science Policy. Credit: Holly Joynes/EMBL

https://www.embl.org/documents/wp-content/uploads/2021/12/ip71-open-science-and-open-access-policy.pdf

This project has recei ogramme under grant agreement No. 823852

US Federally funded research to be Open Access

BRIEFING ROOM

OSTP Issues Guidance to Make Federally Funded Research Freely Available Without Delay

AUGUST 25, 2022 • PRESS RELEASES THE WHITE HOUSE

1. *Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;*

2. *Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,*

3. *Coordinate with OSTP to ensure equitable delivery of federally funded research results and data.*

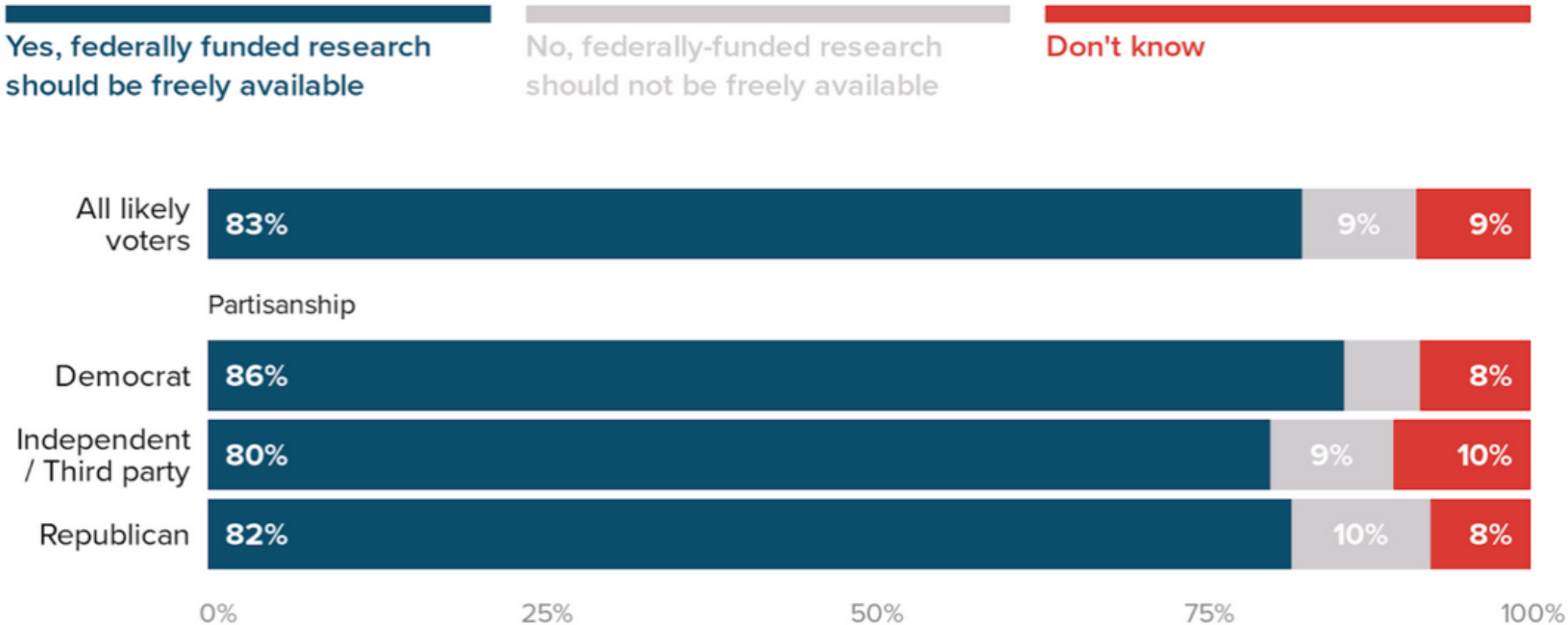
This project has rece

https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

Voters Strongly Support Free Access to Federally Funded Research

The federal government currently funds research projects at universities, non-profit institutions, and small businesses using taxpayer dollars.

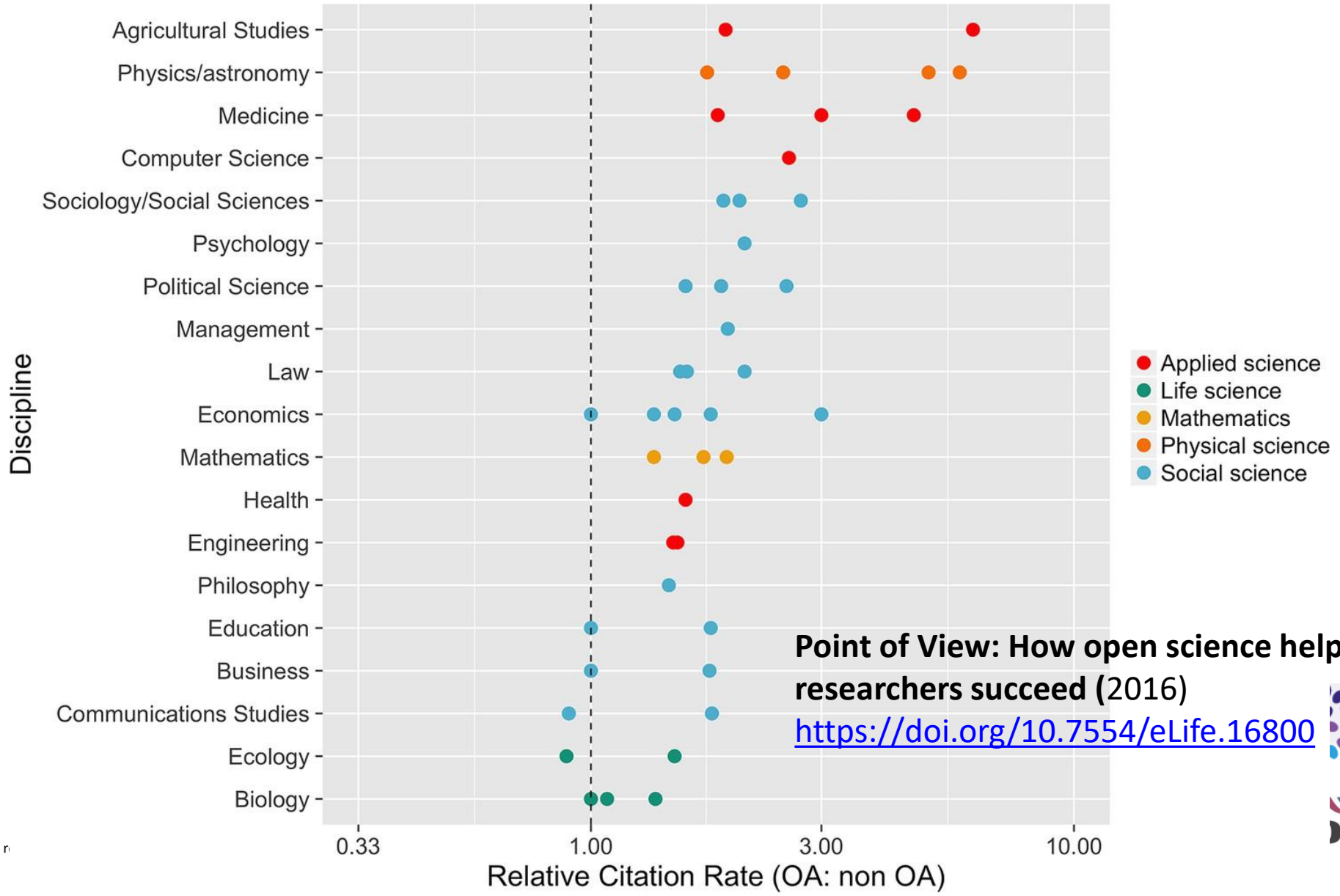
Do you think federally funded research should be freely available for taxpayers to read and access?



This project has

https://www.dataforprogress.org/blog/2022/9/8/voters-overwhelmingly-support-open-access-to-federally-funded-research

Open access leads to more citations



Point of View: How open science helps researchers succeed (2016)
https://doi.org/10.7554/eLife.16800

This project has r

Open access leads to more citations

Scientometrics (2024) 129:825–845
https://doi.org/10.1007/s11192-023-04894-0

Check for updates

Open access research outputs receive more diverse citations

Chun-Kai Huang[1] · Cameron Neylon[1] · Lucy Montgomery[1] · Richard Hosking[1] · James P. Diprose[1] · Rebecca N. Handcock[2] · Katie Wilson[3]

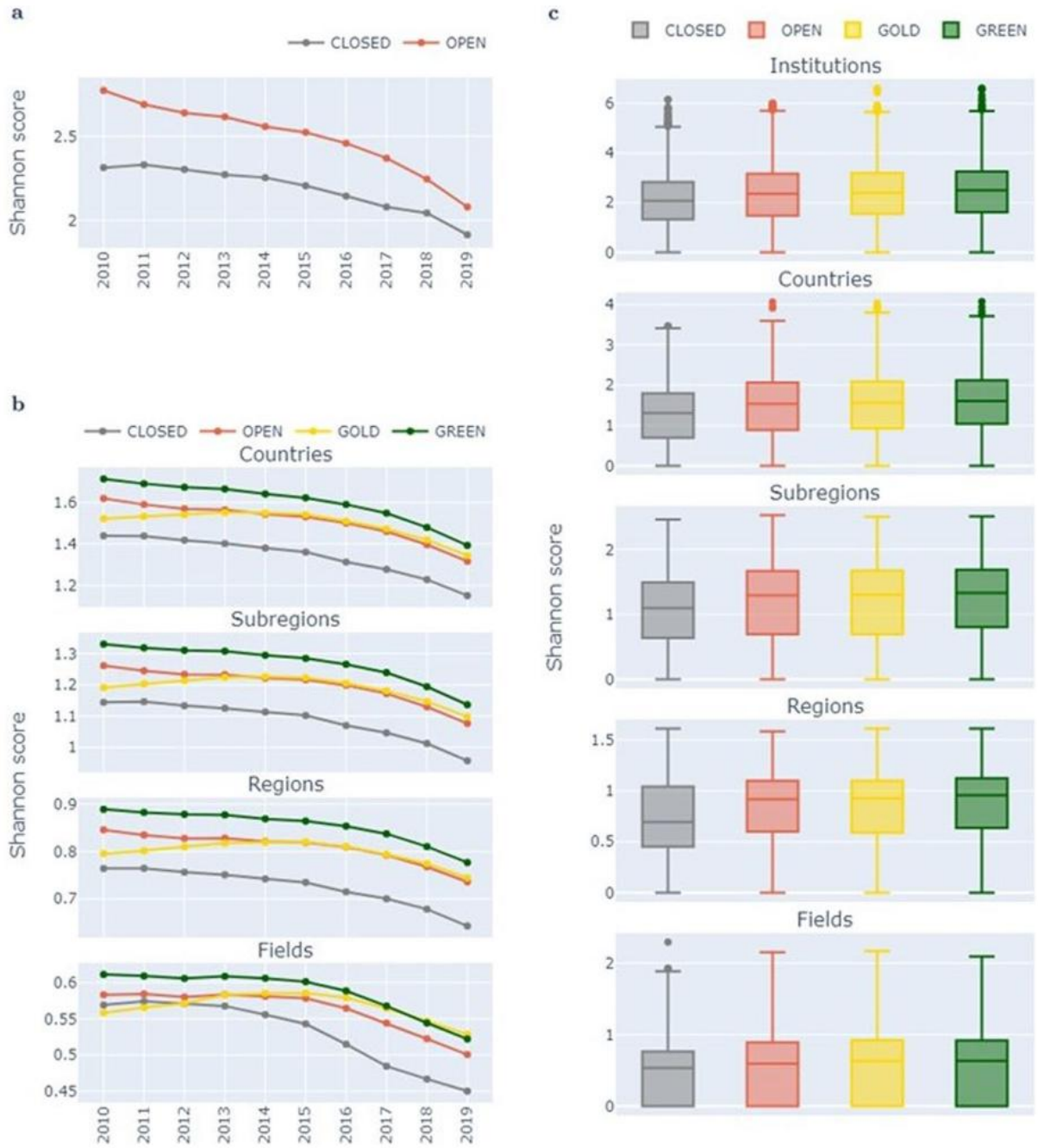


Fig. 2 Comparing citation diversity between OA categories. a The median Shannon scores by citing institutions are compared between OA and CLOSED outputs over a ten-year period. Earlier outputs receive higher scores as a result of having had more time to garner citations (hence more possibility of wider citing affiliations). However, **it is consistently observed that OA outputs perform better in the diversity of citing institutions for all years**.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grar

https://paleo.esrf.fr/

Paleontology database EXPLORE SEARCH HELP

Welcome to the Paleon database

The Paleontology data
Tomography e
communi

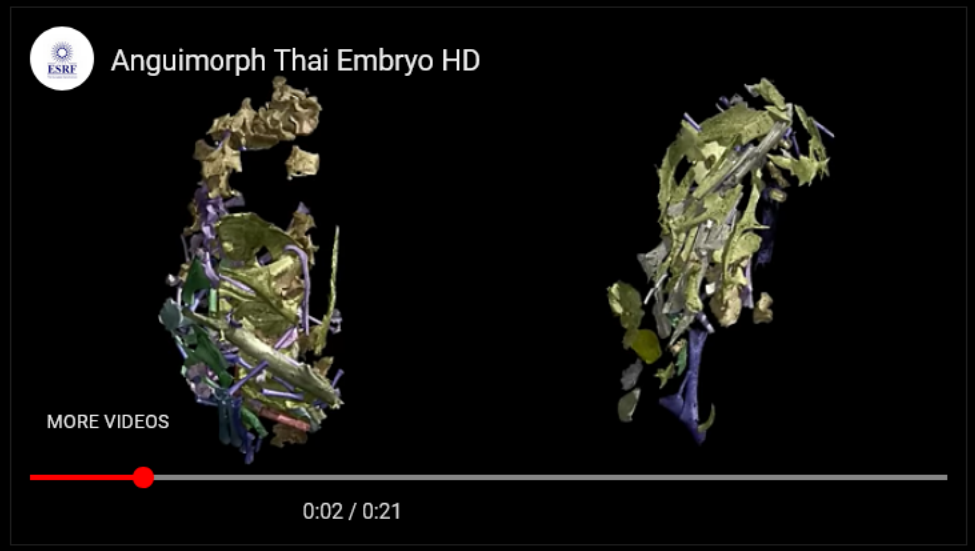
While the ESRF
processed tomo
are only deposite
curators, researche
with geographical, ge

Data hosted here cove
specific close-ups depe
JPEG2000 to help with sto
database, enabled by ESRF-EBS, will contribute to the reuse of data according to the FAIR principles.

57+ publications from users of Open Data

s access to
s, 3D meshes). Data
elevant parties (museum
each dataset is now associated
eywords to facilitate searches.

trum and can represent complete specimens or
Tomograms (i.e., the stack of virtual slices) are stored as
ce and facilitate download by users. We hope that this open-access



Anguimorph Thai Embryo HD.

See the related dataset page. More videos can be found on ESRF YouTube channel.

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

panosc

Open science is beneficial for scientists

nature > nature methods > articles > article

Article | Open Access | Published: 04 November 2021

Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography

C. L. Walsh, P. Tafforeau, W. L. Wagner, D. J. Jafree, A. Bellier, C. Werlein, M. P. Kühnel, E. Boller, S. Walker-Samuel, J. L. Robertus, D. A. Long, J. Jacob, S. Marussi, E. Brown, N. Holroyd, D. D. Jonigk, M. Ackermann & P. D. Lee

Nature Methods **18**, 1532–1541 (2021) | Cite this article

82k Accesses | **25** Citations | **2147** Altmetric | Metrics

This article is in the 99th percentile (**ranked 192nd**) of the 436,034 tracked articles of a similar age in all journals and the 98th percentile (**ranked 2nd**) of the 79 tracked articles of a similar age in *Nature Methods*

"If you don't want to share data why become a scientist?" Claire Walsh (UCL)

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

https://human-organ-atlas.esrf.eu/

Human Organ Atlas | EXPLORE | SEARCH | RECONSTRUCTIONS | HELP

Welcome to the Human Organ Atlas

Medical + Citizen science

The Human Organ Atlas us… …y to span a previously po… …man anatomy, the …

Histo… …other structures with sub-micron acc… …gan, while clinical CT and MRI scans can imag… …wn to just below a millimetre. HiP-CT bridges these scales … …20 micron voxels, and locally down to microns.

We hope … …enabled by the ESRF-EBS, will act as a reference to provide new insights int… …al makeup in health and disease. To stay up to date, follow @HiP-CT

HiP-CT imaging and 3D reconstruction of a complete brain from the body donor LADAF-2020-31. More videos can be viewed on the HiP-CT YouTube channel.

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

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**

Open science vs. science

Most of these assumptions are not new, as the tradition of openness itself is at the roots of science, but the current developments of information and communication technologies have transformed the scientific practices to a level that requires a different approach to research (FOSTER)

https://www.fosteropenscience.eu/content/what-open-science-introduction

Q: "What is the difference between Open Science and 'science'?"

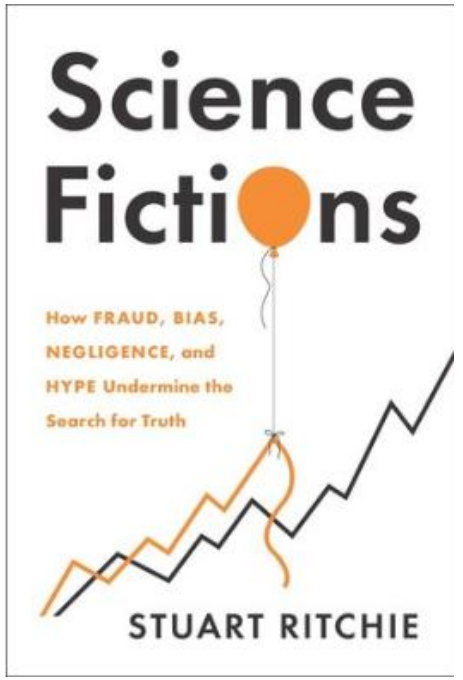
A: *Open Science refers to doing traditional science with more transparency involved at various stages, for example by openly sharing code and data. Many researchers do this already, but don't call it Open Science.*

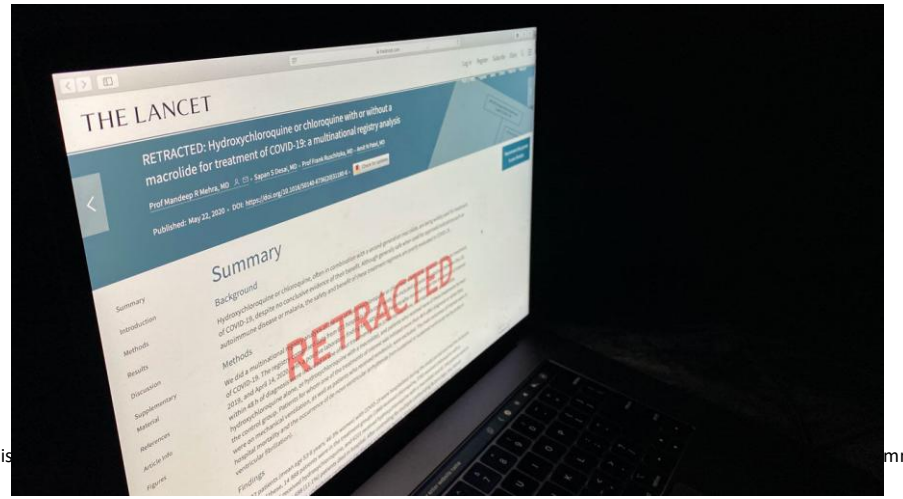
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

European Conduct of Scientific Integrity

Open Science improves integrity, scientific method

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



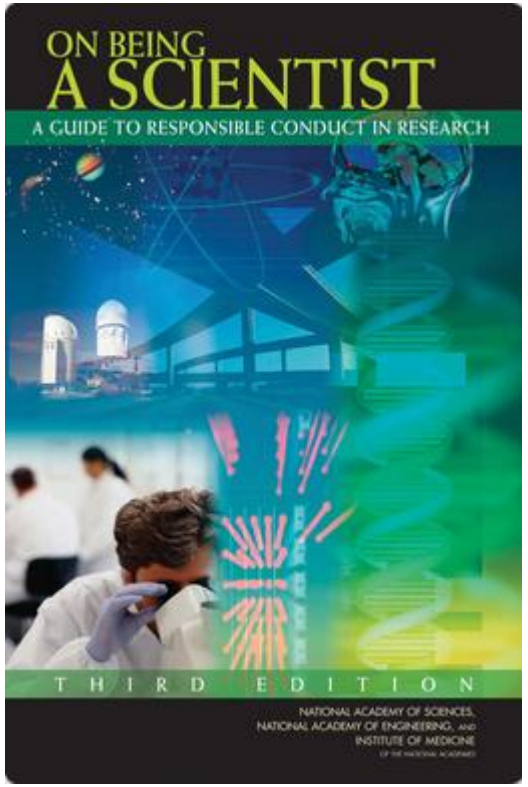


Our list of retracted or withdrawn COVID-19 papers is up to over 375. There are more than 46,000 retractions in The Retraction Watch Database — which is now part of Crossref. The Retraction Watch Hijacked Journal Checker now contains well over 200 titles. And have you seen our leaderboard of authors with the most retractions lately — or our list of top 10 most highly cited retracted papers? Or The Retraction Watch Mass Resignations List?

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**
- **EU**
  - **Progress on Open Science**



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

FAIR Data

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

The publication that started the FAIR movement

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, ... Barend Mons

+ Show authors

Scientific Data **3**, Article number: 160018 (2016) | Cite this article

Online attention

- 1207 tweeters
- 6 Google+ users
- 1 F1000
- 7808 Mendeley
- 187 blogs
- 203 news outlets
- 2 Video uploaders
- 20 Citeulike
- 22 Facebook pages
- 1 Redditors
- 27 Wikipedia page

This article is in the 99th percentile (ranked 37th) of the 317,153 tracked articles of a similar age in all journals and the 95th percentile (ranked 1st) of the 23 tracked articles of a similar age in *Scientific Data*

https://data.europa.eu/doi/10.2777/1524

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

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

Data availability – the wrong + right way

Data availabi

Data available on re nable request t e authors.

Open Research

Data Availability Statement

The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.6993871 , reference number 6993871.

OSCARS
Open Science Clusters' Action for Research & Society

panosc

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

Open Research Europe recommendations for data

My Submissions

Article Guidelines

Article Guidelines (New Versions)

Open Data, Software and Code Guidelines

Open Data and Accessible Source Materials Guidelines (HSS)

Prepublication Checks

Article Processing Charges

Finding Article Reviewers

What is required when submitting an article

1. Your dataset(s) must be deposited in an appropriate data repository.
2. Your dataset(s) must have a license applied which allows reuse by others (CC0 or CC-BY).
3. Your dataset(s) must have a persistent identifier (e.g. a DOI), allocated by a data repository.
4. You must provide a data availability statement as a section at the end of your article, including elements 1-3.
5. You must include a data citation and add a reference to data to your reference list.
6. Your dataset(s) should not contain any sensitive information, for example in relation to human research participants.
7. You should share any related software and code.
8. Your dataset(s) must be useful and reusable by others, adhere to any relevant data sharing standards in your discipline and align with the FAIR Data Principles.
9. Your dataset(s) should link back to your article, if possible.

https://open-research-europe.ec.europa.eu/for-authors/data-guidelines/

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

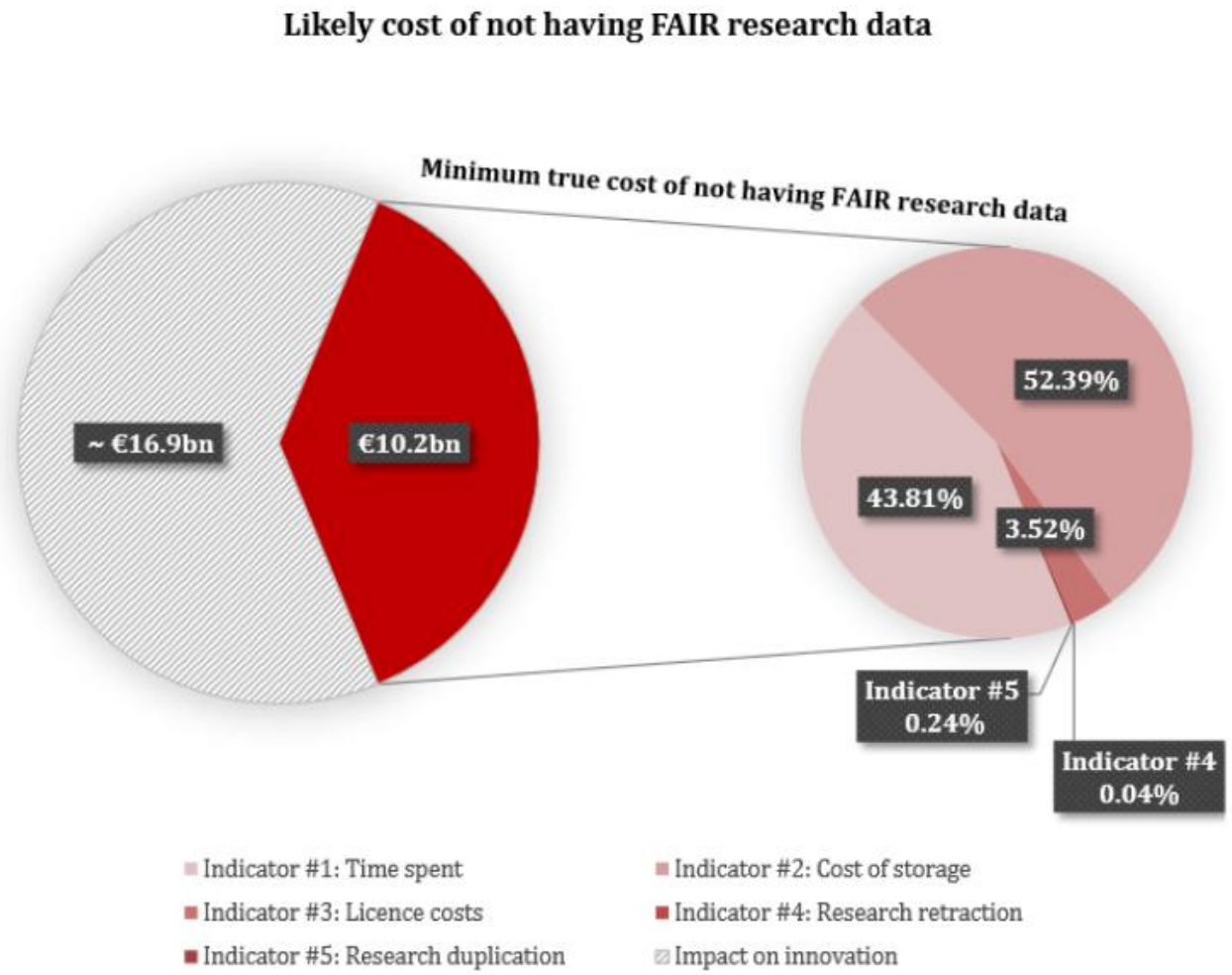
Reusable

- 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

panosc

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

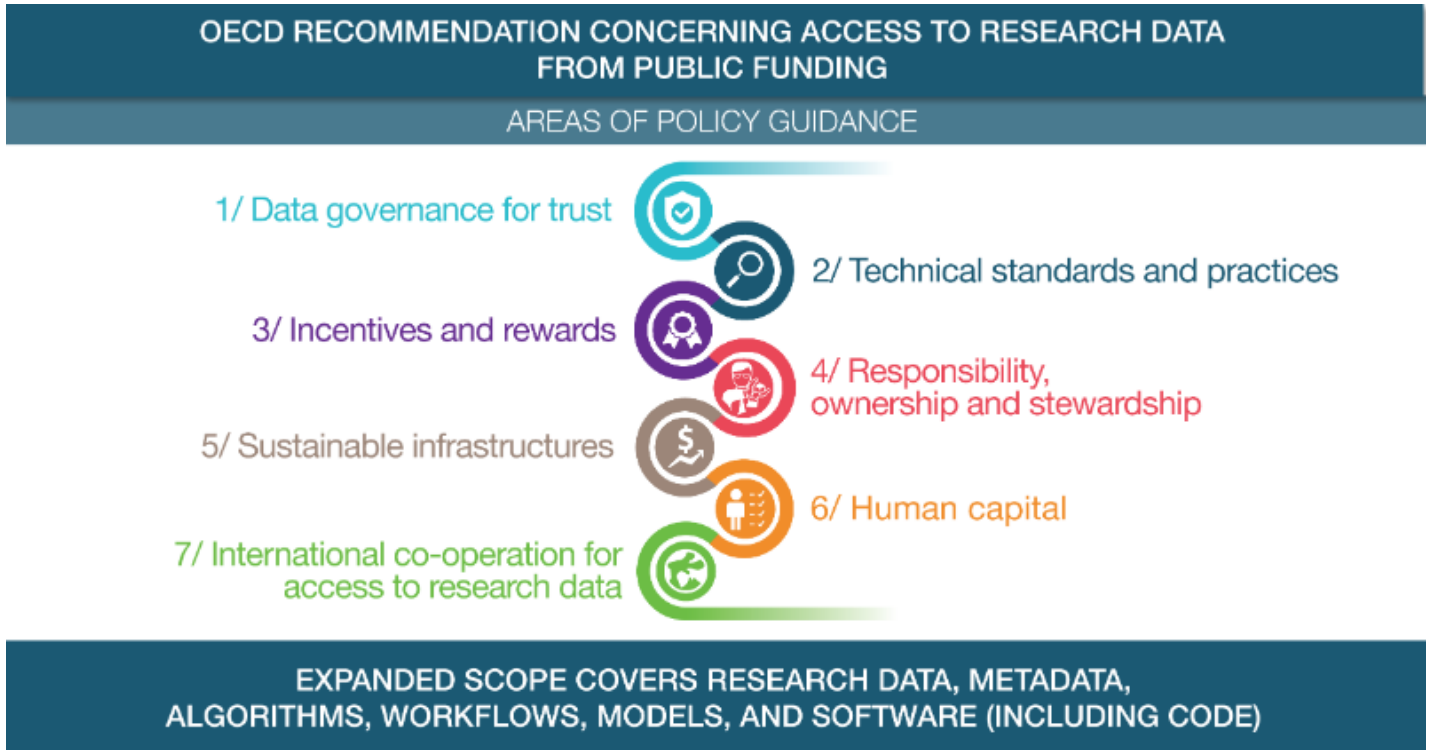


“Cost-benefit analysis for FAIR research data “ (https://op.europa.eu/s/pevt **)**

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

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)



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

Data is not adapted to being “owned”

- ➢ Data do not belong to you (or anyone) i.e. replace “my data” with “the data”
- ✓ Data policies govern who has access to Data

panosc

Research Facilities Data Policies

- **ESA** – open data policy for most data (since 2010)
- **ILL** – open data policy (since 2012)
- **ESRF** – open data policy (since 2015)
- **EMBL** – open access policy (since 2015)
- **ESO** – open data policy (updated in 2016)
- **EuXFEL** – open data policy (since 2017)
- **EUROfusion** – proposal for open data policy (in progress since 2018)
- **CERN** – open data policy for LHC (since 2020)
- **CERIC-ERIC** – open data policy (since 2021)
- **SESAME** – open data policy (since 2023)
- **PSI, ALBA, SOLEIL, ELETTRA, HZB, MAXIV, …**

OSCARS
Open Science Clusters' Action for Research & Society

panosc

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 2015

The ESRF Data Policy

The ESRF aims to implement a Data P in 2016.
The main elements of this policy c

- **Data ow**
- **Data**
- **Data a**
- **Open a**

This policy follows larg e 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)

Updated to be FAIR compliant in 2024

OSCARS Open Science Clusters' Action for Research & Society

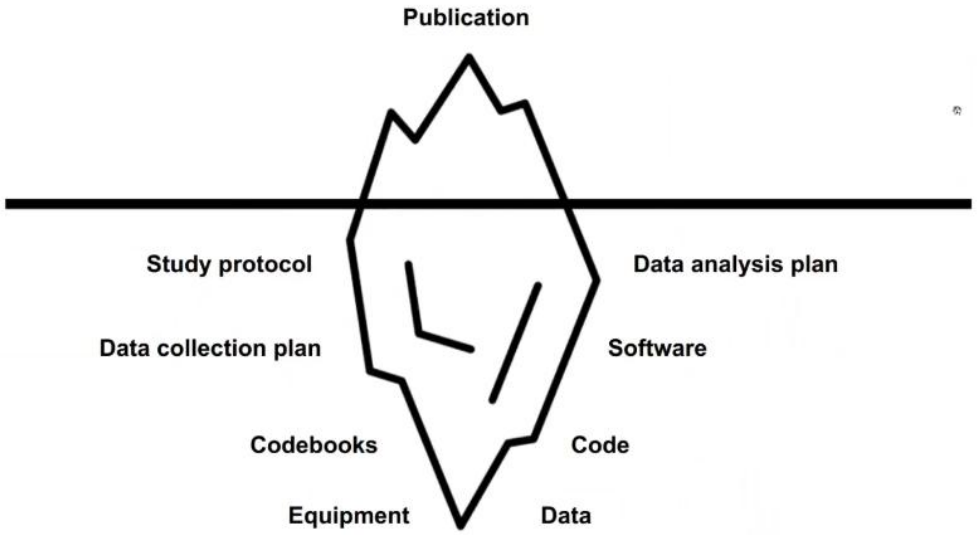
panosc

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

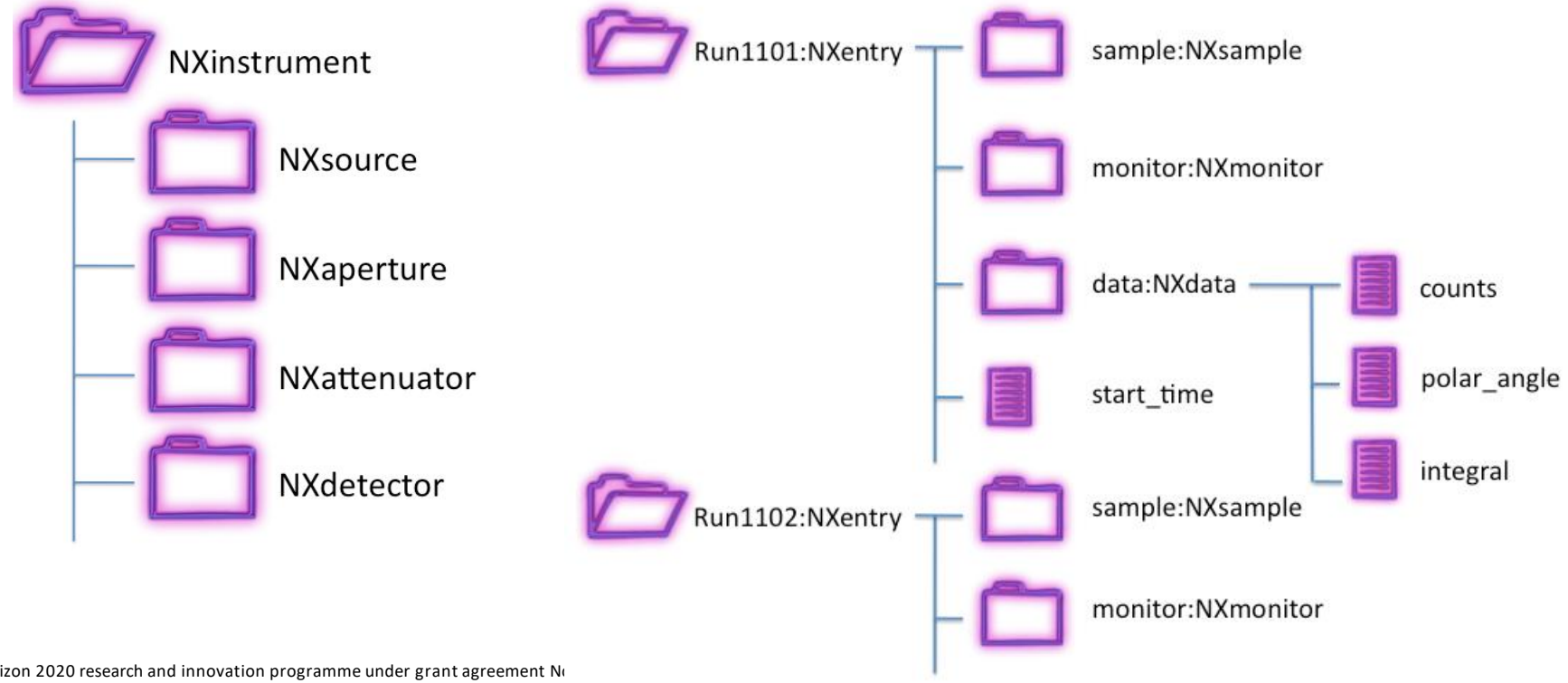
NeXus is developed as an international standard by scientists and programmers representing major scientific facilities in Europe, Asia, Australia, and North America in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

Home

GitHub Organisation

https://www.nexusformat.org/

Nexus provides a standard vocabulary for:

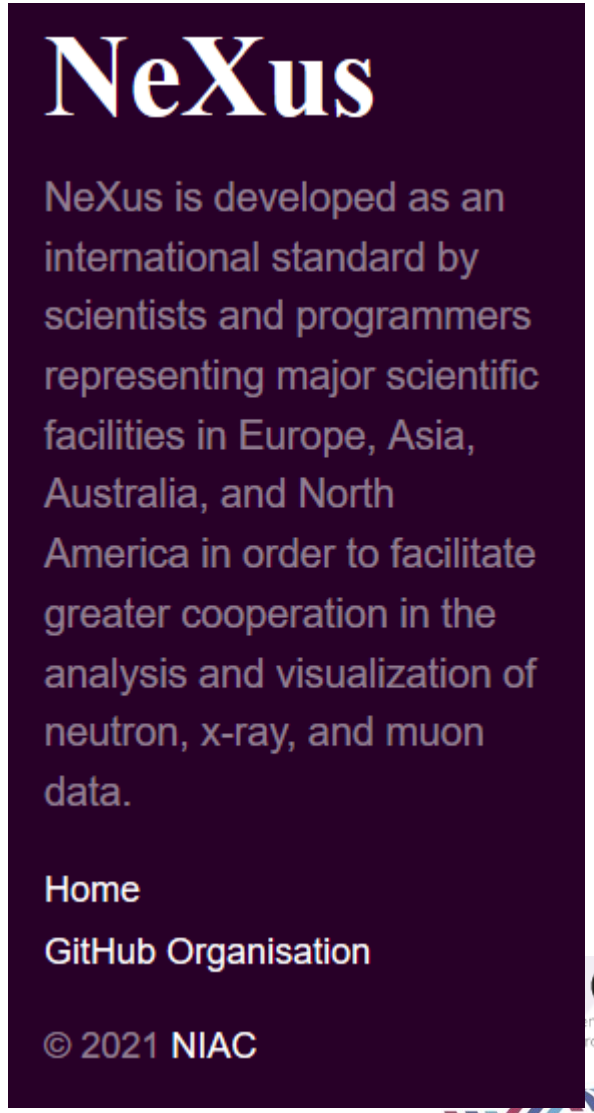


ean Union's Horizon 2020 research and innovation programme under grant agreement N

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	



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

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

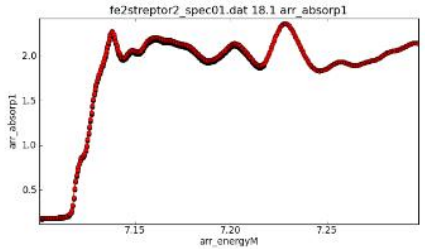
ID21 / EV-280 Beneficial symbiosis in tomato plants: its role on Fe translocation and speciation

Dataset List 90 | Logbook | Shipping

+ New | Take a photo | View | PDF | Help | 0 new log(s) arrived. | Everywhere | Everywhere | 72 found

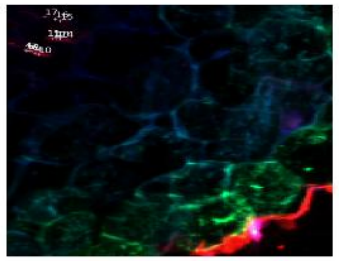
November 5th 2018

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



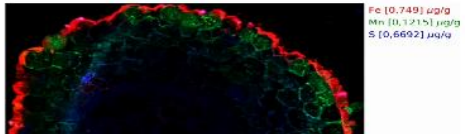
energy calibration

09:58:09 `OPTICS> Fexanes_ev280`

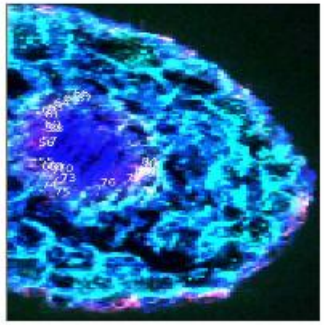


Fe2 strepto r2 Main Root: XANES Points

00:18:27 `OPTICS> zapxiaimage samy 7.904 7.324 580 samz 25.476 26.025 549 100 0 (zap: #1, spec: #2)`

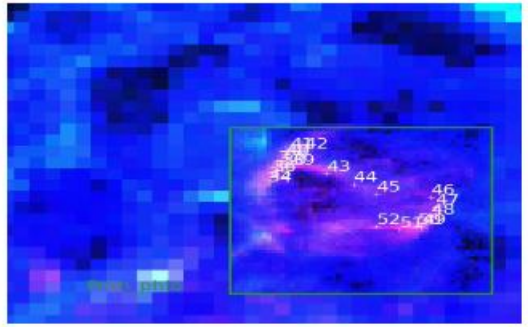


23:50:29 `OPTICS> Fexanes_ev280`



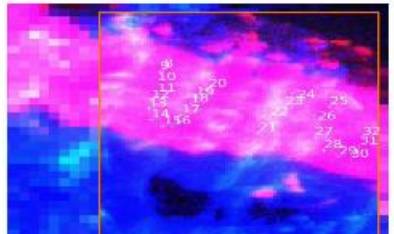
Fe3 prio r3 Sec Root: XANES Points

22:48:31 `OPTICS> Fexanes_ev280`

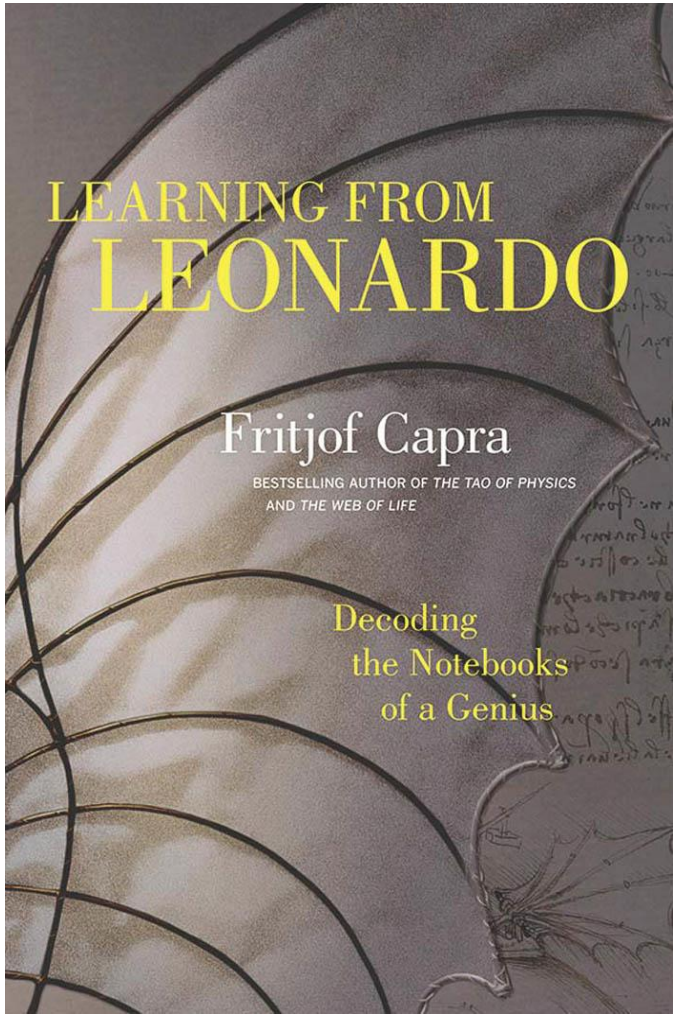


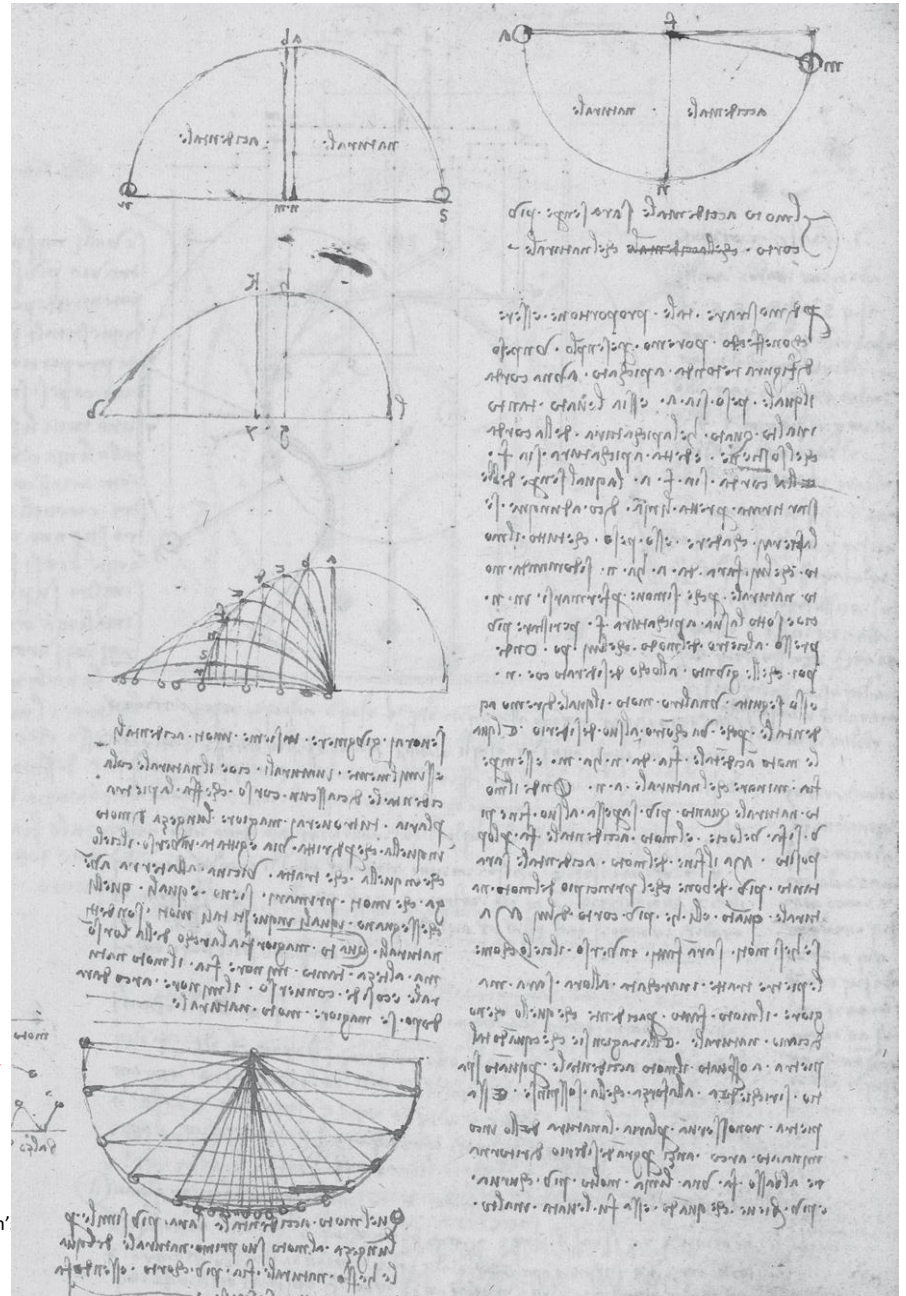
Fe3 prio r3 Main Root prim phlo: XANES Points

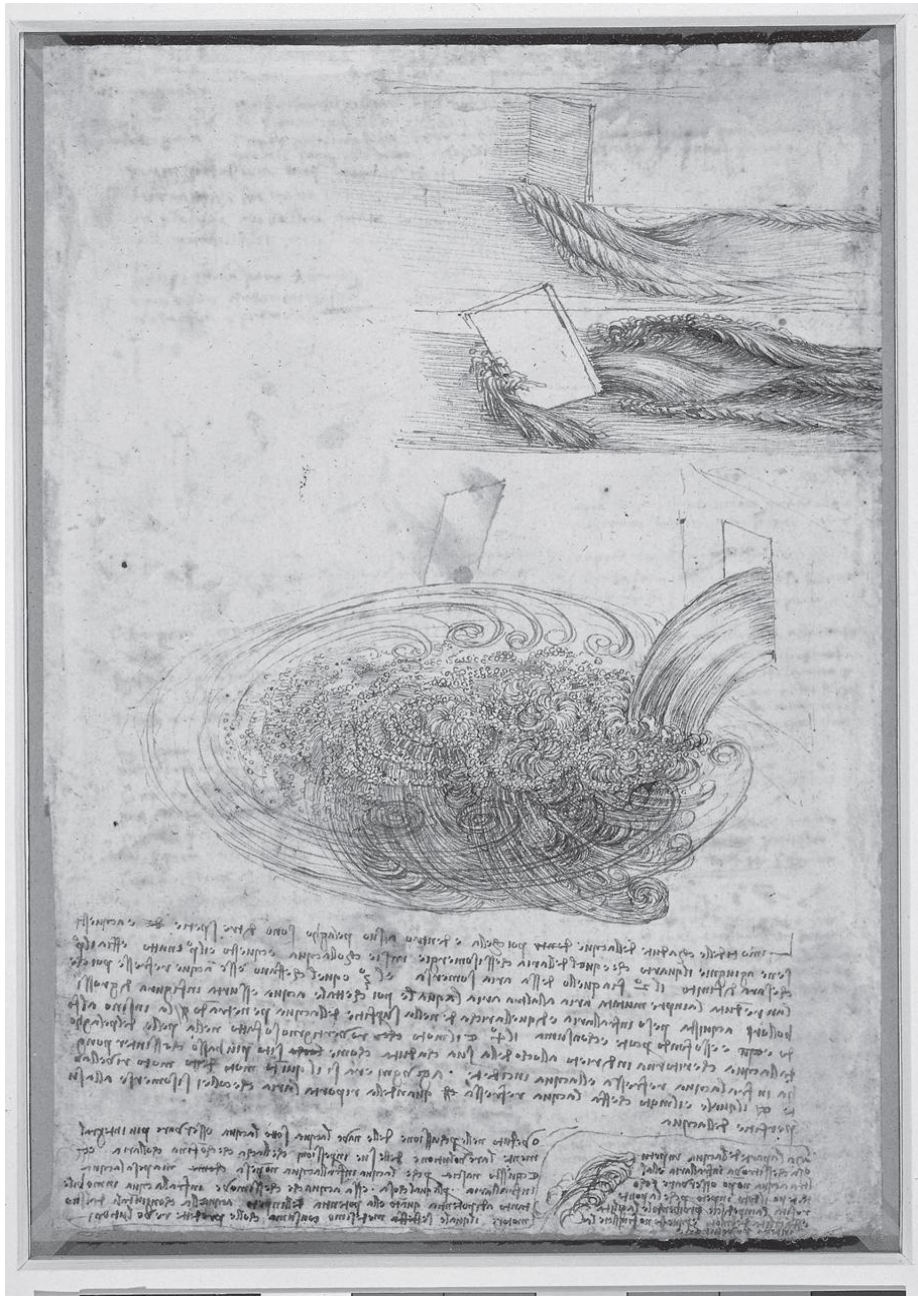
22:25:02 `OPTICS> Fexanes_ev280`



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 Union'

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

Github.com

Gitlab.com

git

OSCARS
Open Science Clusters' Action for Research & Society

panosc

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

- **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

panosc

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

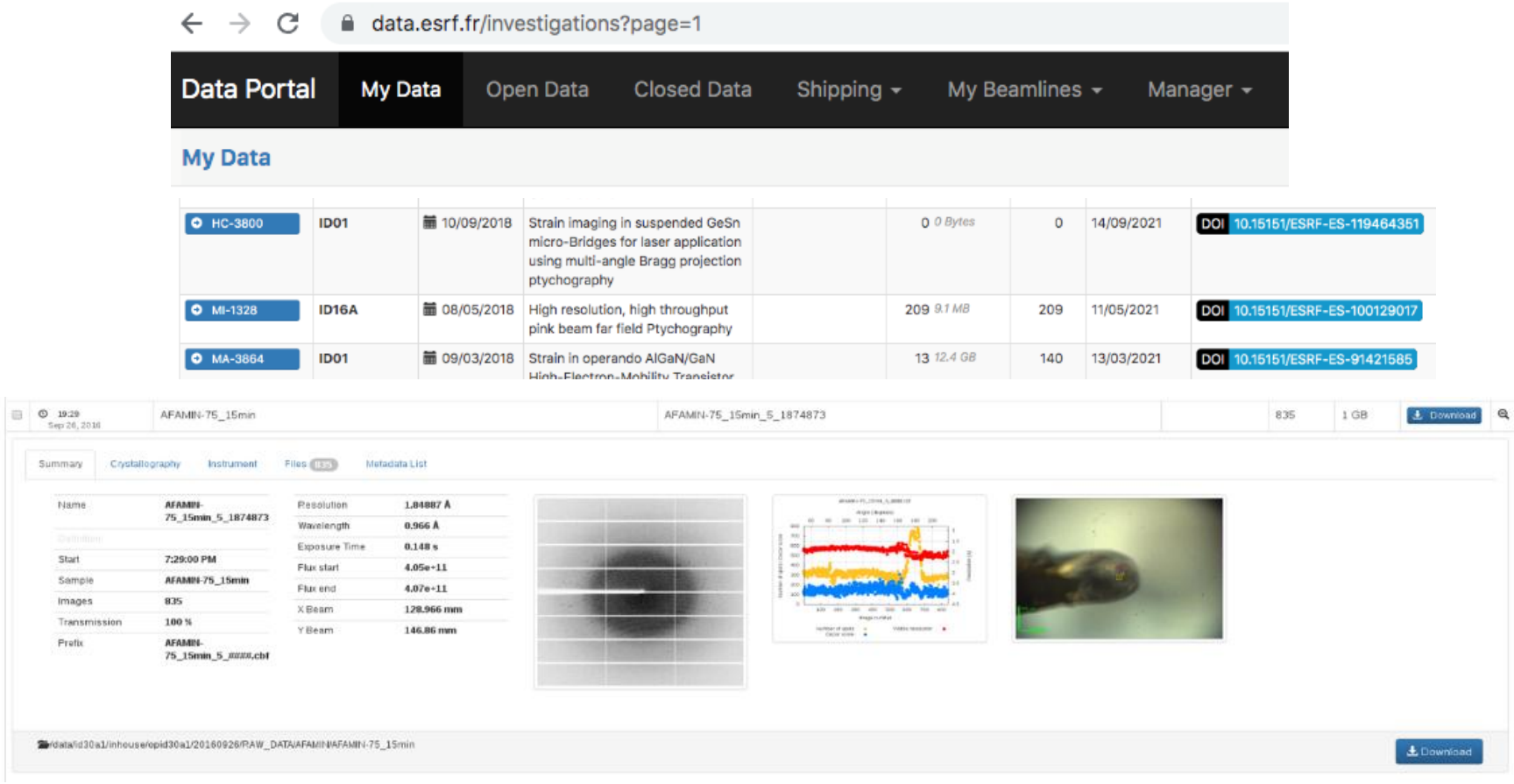
Data archiving

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

ESRF data portal - https://data.esrf.fr

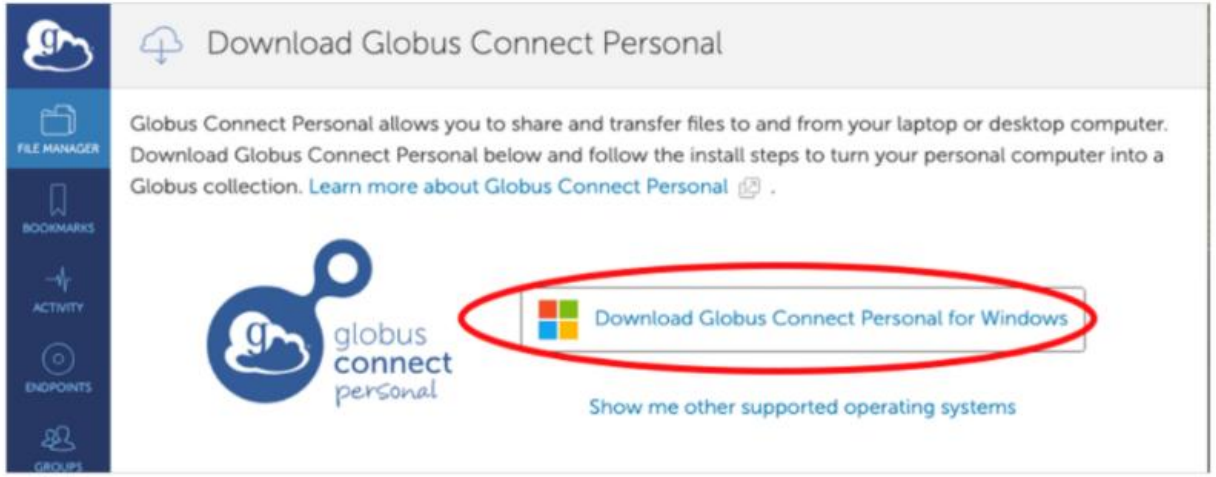


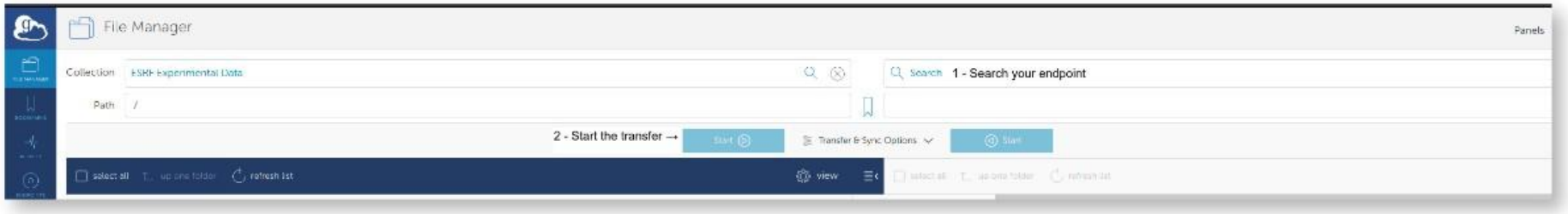
ICAT project collaboration https://github.com/icatproject

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

Downloading large data: globus online

For users that want to download large volume of experimental data **(largest transfer so far 50TB)**





The service opened in fall 2021 for all users and all data

Data access is protected using Access Control Lists (ACLs) on the storage – users cannot see others data.

OSCARS
Open Science Clusters' Action for Research & Society

panosc

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 correctly 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

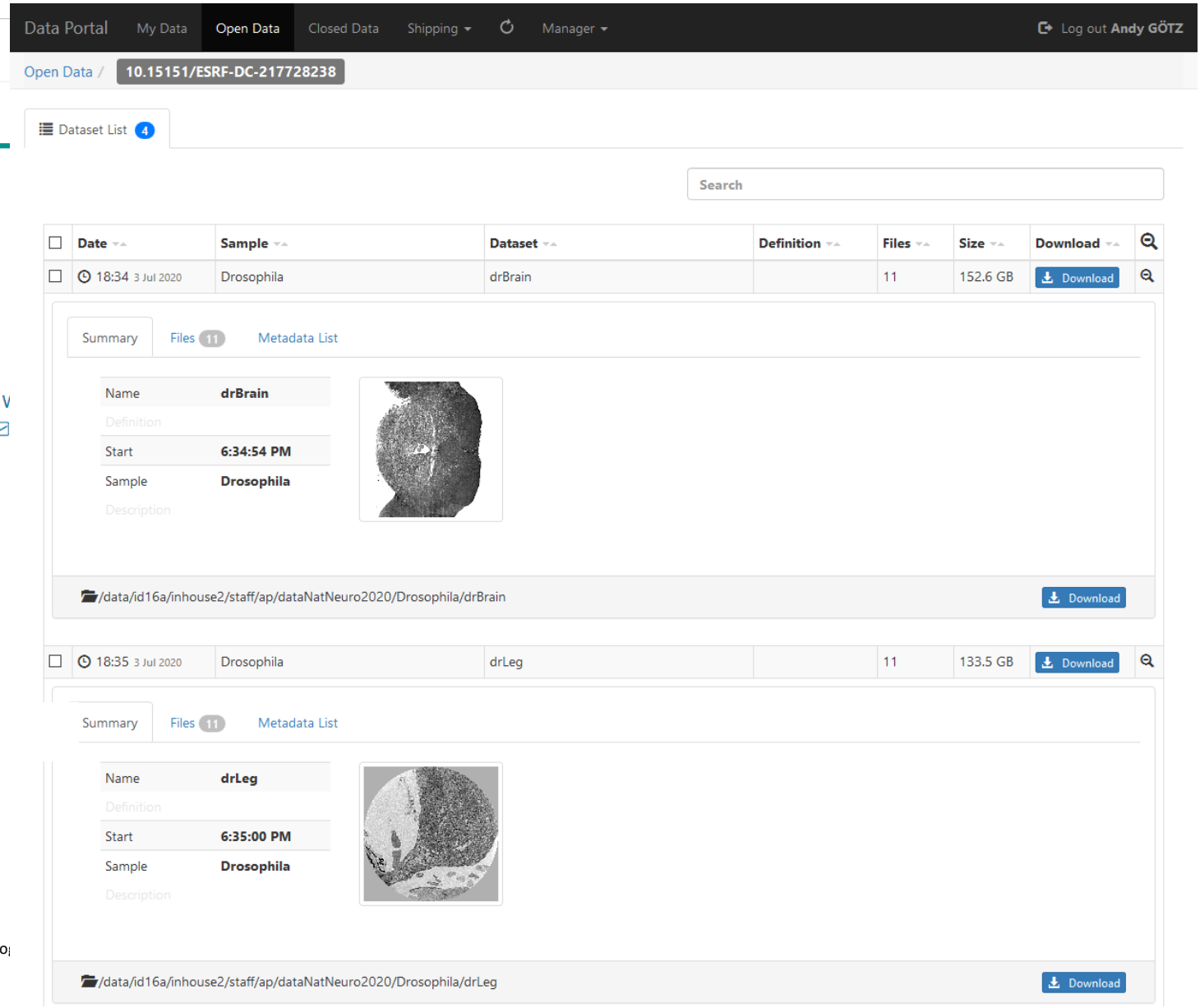
5492 Accesses | **8** Citations | **196** Altmetric | Metrics

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

https://data.esrf.fr/doi/10.15151/ESRF-DC-217728238

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

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



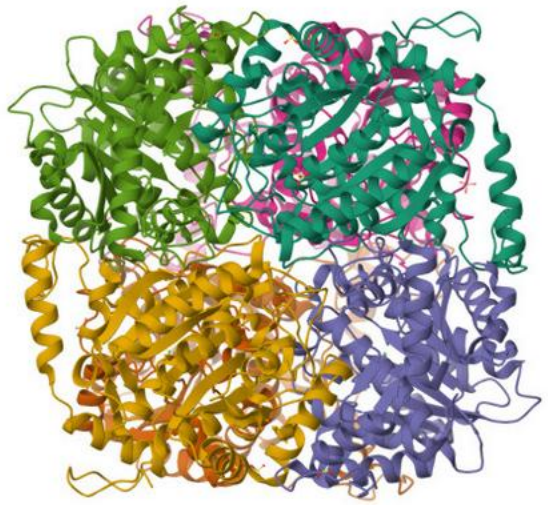
Example of PDB entry correctly citing raw data

RCSB PDB Deposit Search Visualize Analyze Download Learn About Careers COVID-19 Help Contact us MyPDB

PDB-101 wwPDB EMDataResource NAKB wwPDB Foundation PDB-IHM

Structure Summary | Structure | Annotations | Experiment | Sequence | Genome | Versions

Biological Assembly 1



Display Files Download Files Data API

9FEK

Crystal structure of guanidin

PDB DOI: https://doi.org/10.2210/pd

Classification: METAL BINDING P

Organism(s): Candidatus Nitrospira

Expression System: Escherichia c

Mutation(s): No

Deposited: 2024-05-20 **Released:**

Deposition Author(s): Puehringer,

Funding Organization(s): Vienna S

University of Vienna Research Platfo

Experimental Data & Validation

Experimental Data

Method: X-RAY DIFFRACTION
Resolution: 1.58 Å
R-Value Free: 0.214
R-Value Work: 0.200
R-Value Observed: 0.200
Space Group: P 1 2_1 1
Diffraction Data: https://doi.org/10.15151/ESRF-DC-1801440672
Diffraction Data: https://doi.org/10.15151/ESRF-ES-771426868

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

panosc

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
- Very hard for a individuals to maintain access to local storage for years
- **Research facilities provide services to keep raw data at the facility/cloud**
- Many free services exist now for scientific data e.g. Zenodo, Figshare, …
- Commercial cloud offer practically unlimited resources at a cost
- Data stored on commercial cloud disappear when you stop paying

OSCARS
Open Science Clusters' Action for Research & Society

panosc

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

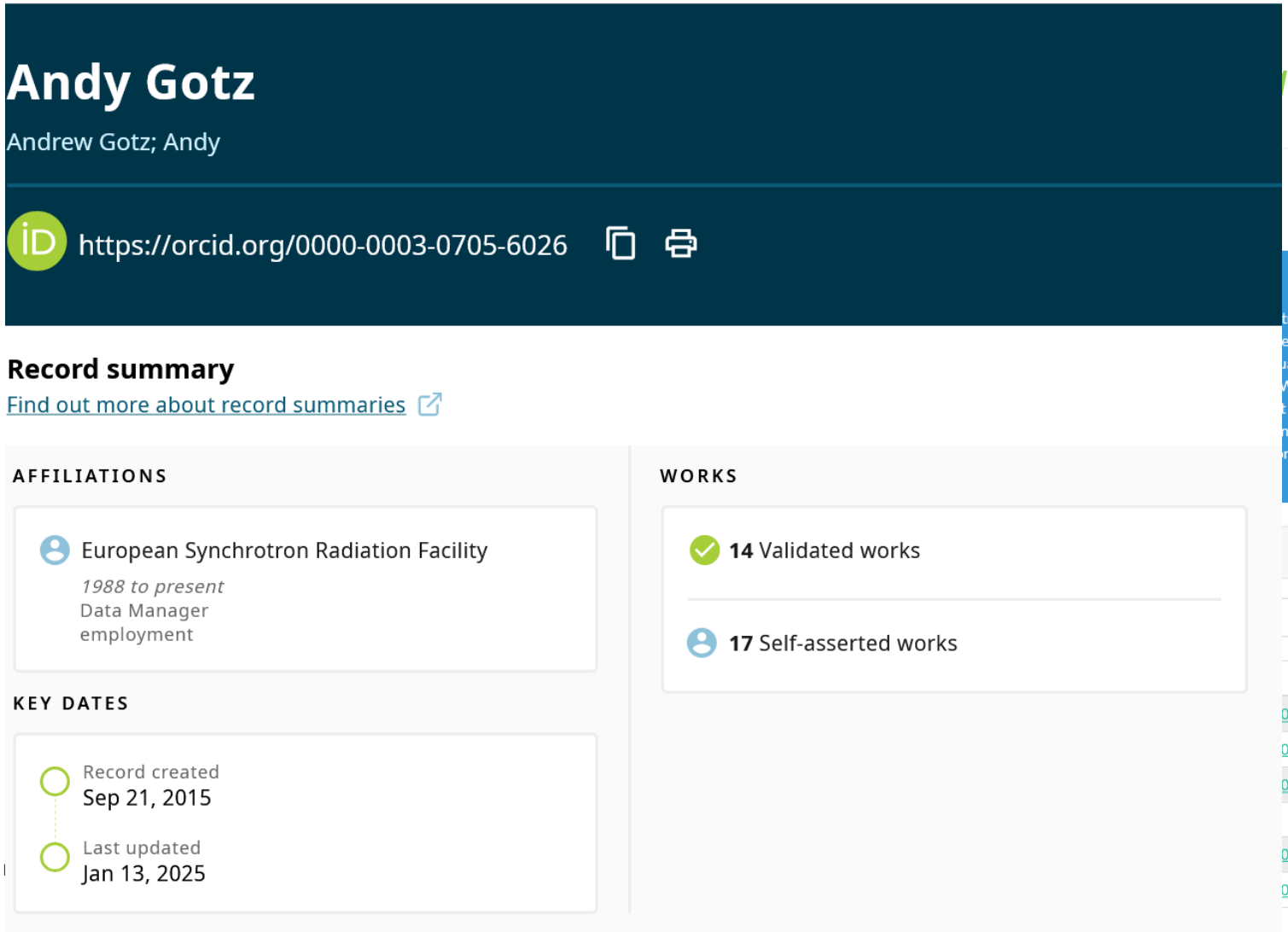
- Benefits of an ORCID identity:
  - You will be distinguished from every other researcher, even researchers who share your same name,
  - Your research outputs and activities will be correctly attributed to you,
  - Your contributions and affiliations will be reliably and easily connected to you,
  - 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

Open identifier – ORCID.org

ORCID

Connecting research and researchers



This project

Open Training – https://pan-learning.eu

Catalogue | e-Learning | Events | About | Andy Goetz

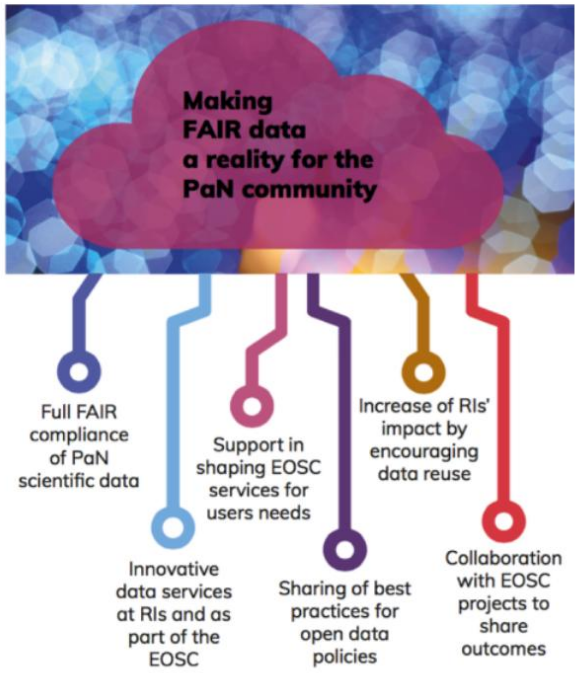
PaNOSC summerschool FAIR session

e-Learning | My courses | PaNOSC summerschool FAIR session

Turn editing on

- PaNOSC summerschool FAIR session
- Participants
- Grades
- Making FAIR data a reality
- Trust in Science
- What is Open Science
- What is EOSC
- What is PaNOSC + ExPaNDS
- Outcomes of PaNOSC + ExPaNDS
- What is Scientific Data and Metadata
- How to make data FAIR
- Your role as a researcher in Open Science
- Questions
- e-Learning
- Dashboard
- Calendar
- Content bank
- Quiz Taster

Making FAIR data a reality



What are Science, Open Science, EOSC and PaNOSC

Please answer to the questions below to start the course.

Announcements

Catalogue | e-Learning | Events | About | Andy Goetz

- PaNOSC summerschool FAIR session
- Participants
- Grades
- Making FAIR data a reality
- Trust in Science
- What is Open Science
- What is EOSC
- What is PaNOSC + ExPaNDS
- Outcomes of PaNOSC + ExPaNDS
- What is Scientific Data and Metadata
- How to make data FAIR
- Your role as a researcher in Open Science
- Questions
- e-Learning
- Dashboard
- Calendar
- Content bank
- Quiz Taster

Trust in Science

- Why trust
- Why trust science quiz

What is Open Science

- Open Science definition
- Pillars of Open Science
- Process of Openness in Research
- Open Science Schools of Thought
- Open Science Resources

What is EOSC

- The European Open Science Cloud
- EOSC Projects and Ecosystem

What is PaNOSC + ExPaNDS

- PaNOSC + ExPaNDS projects

Outcomes of PaNOSC + ExPaNDS

- PaNOSC + ExPaNDS FAIR Outcomes

What is Scientific Data and Metadata

- Scientific Data and Metadata

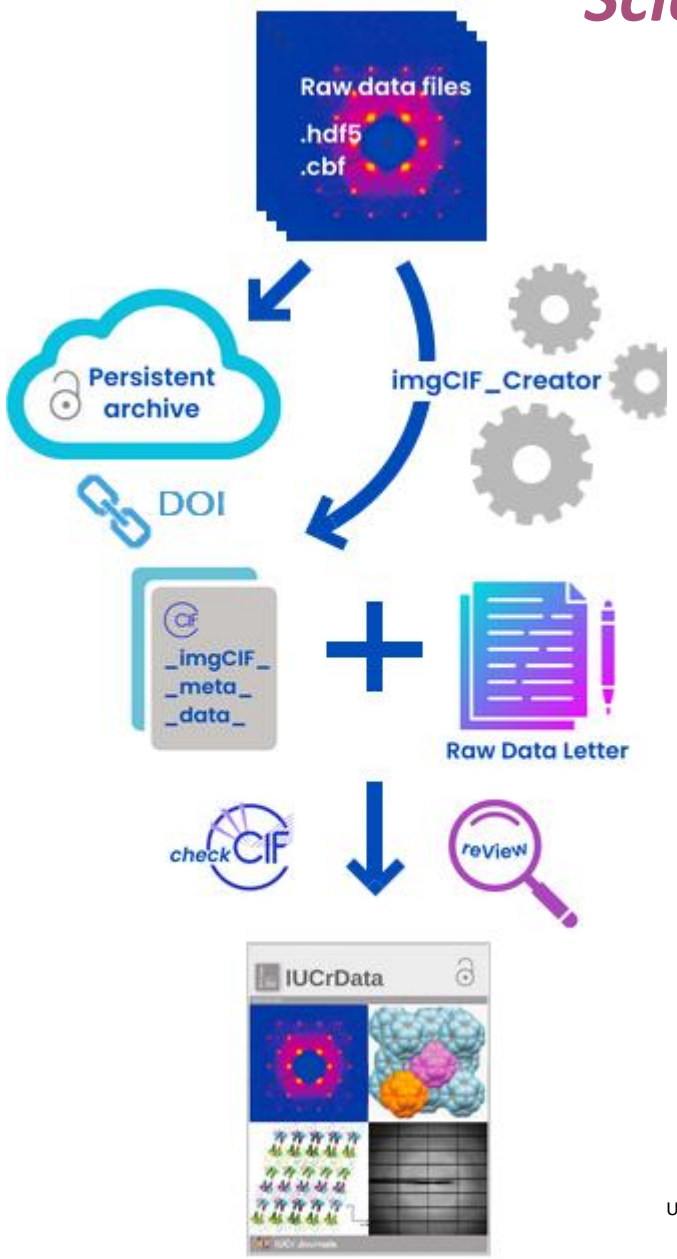
https://e-learning.pan-training.eu/moodle/course/view.php?id=114

What are the advantages of producing FAIR Data?

- 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

IUCr Journals have launched IUCrData's Raw Data Letters
Scientists are encouraged to publish raw data



IUCrData

ISSN 2414-3146

Received 20 April 2021
Accepted 1 May 2021

Keywords: twinning; diffuse scattering; tetraspanin $CD9_{EC2}$.

Crystal structure of the second extracellular domain of human tetraspanin D9: twinning and diffuse scattering

Viviana Neviani, Martin Lutz, Wout Oosterheert, Piet Gros and Loes Kroon-Batenburg*

Department of Chemistry, Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, Utrecht, The Netherlands. *Correspondence e-mail: l.m.j.kroon-batenburg@uu.nl

Remarkable features are reported in the diffraction pattern produced by a crystal of tetraspanin $CDCD9_{EC2}$, the structure of which was described previously [Oosterheert *et al.* (2020). *Life Sci. Alliance*, **3**, e202000883]. $CD9_{EC2}$ crystallized in space group *P*1 and was twinned. Concurrent with the twinning, diffuse streaks were seen in the direction perpendicular to the twinning interface. Preliminary conclusions are made on packing disorder and potential implications for the observed molecular structure. It is envisaged that the raw diffraction images could be very useful for methods developers in trying to remove the diffuse scattering to extract accurate Bragg intensities or by using it to model the effect of packing disorder on the molecular structure.

Raw data

Structure

Raw diffraction data

HDF5 data file, DOI: https://doi.org/10.5281/zenodo.1234567

Metadata ImgCIF file, DOI: https://doi.org//10.1107/S2414314622000384/me6134.cif

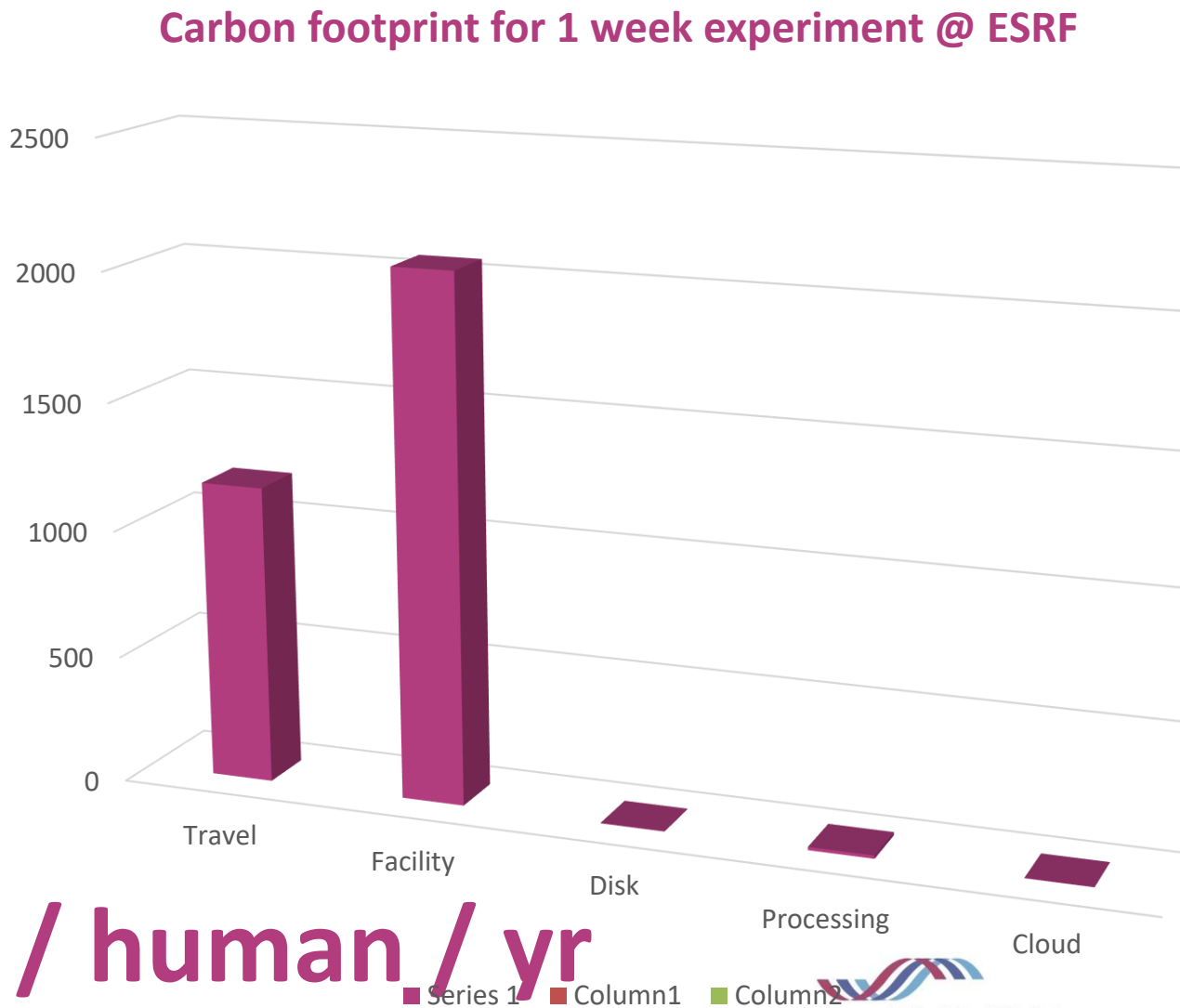
Union's Horizon 2020 research and inno

panosc

Estimated carbon footprint of experiment

- User Travel = **1170 kg**
- Beamtime energy consumption = **2056 kg**
- Data stored on disk = **1.8 kg**
- Data processing on site = **12.6 kg**
- Cloud transfer = **2.3 kg**

CO2e per kwH in France = **75 g/kWh**



TOTAL = 3.253 tons !

Sustainable Goal = 5 tons / human / yr

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

Carbon footprint of archiving data

- 200 GB Data archived on tape for 10 years
  (full tape library) ~ 13 g * 10 yrs = 130 grams

→ARCHIVING raw data for 10 years
0.000004 % of CO2eq needed
to acquire the raw data!

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

Data availability – the wrong + right way

Data availabi

Data available on re nable request t e authors.

Open Research

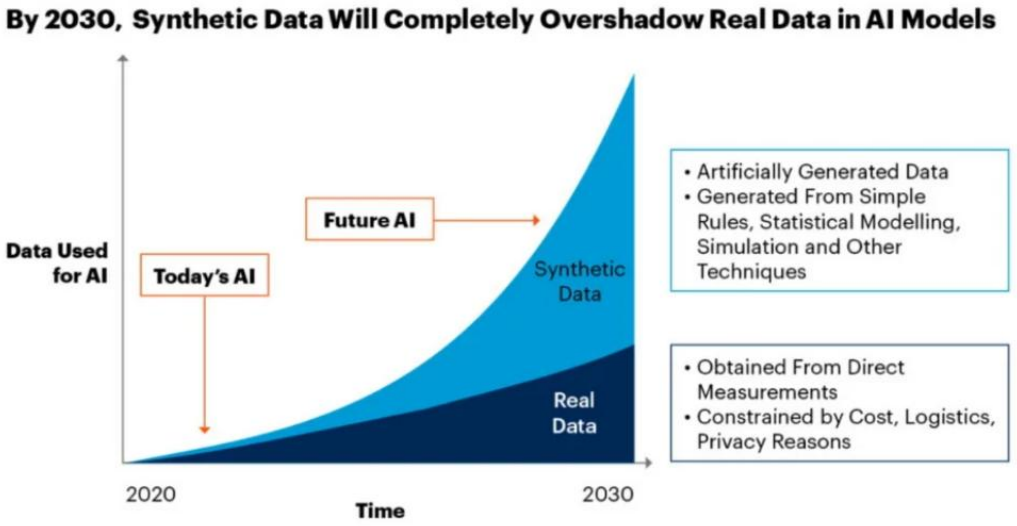
Data Availability Statement

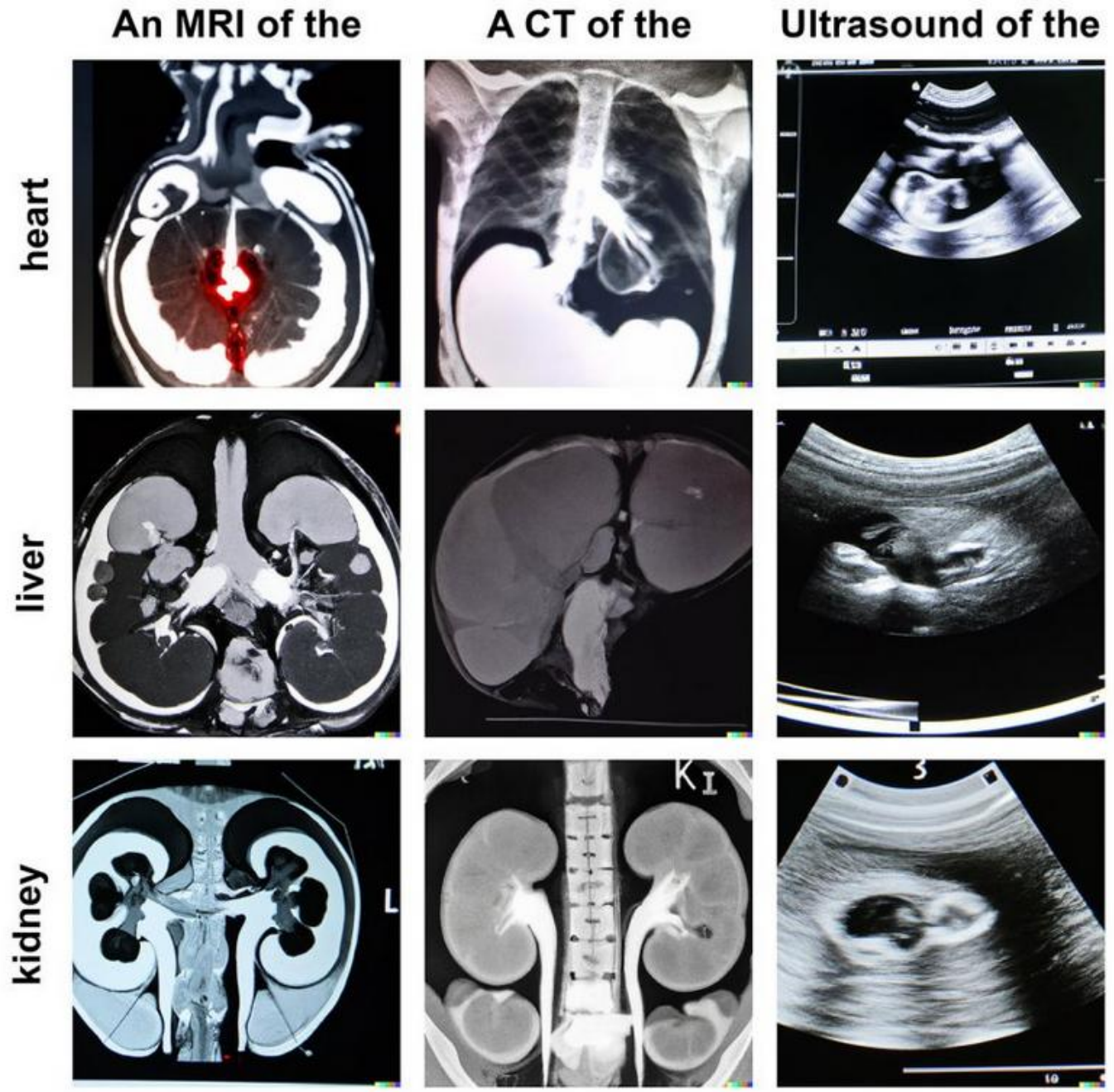
The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.6993871 , reference number 6993871.

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

Another reason for FAIR data is to distinguish from AI generated data

In the future we might need to watermark FAIR data





Examples of text-to-image–generated anatomical structures in CT, MRI, and ultrasound images created with DALL-E 2.

Image source: Adams et al., Journal of Medical Internet Research 2023 (CC BY 4.0)

Conclusions

Adopting best practices for Open Science and FAIR Data has many benefits especially helping MAKE BETTER + REPRODUCIBLE SCIENCE

- Make sure you follow a **checklist** 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
  - Spend time with your data to make it **FAIR by adding rich metadata, your orcid,** and then **releasing it, publishing it, citing the data DOI !**
- Many **digital tools** exist for treating your data seriously + publishing them
- Learn about the data you will produce before going to the synchrotron

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

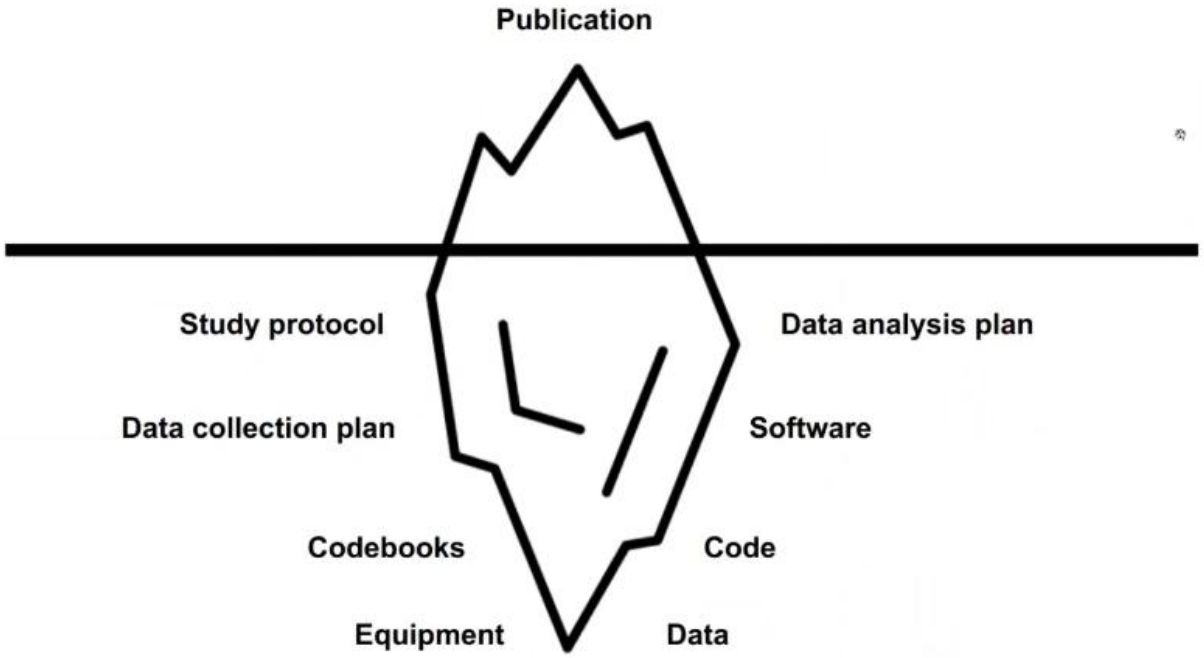
Acknowledgements

- RDMKit Elixir online guide
- University of Saskatchewan
  - https://library.usask.ca/studentlearning/workshops/grad-research.php#panel-section-3-ResearchDataManagementWhatYouNeedtoKnow
- **Nature** magazine, Scientific Data
- **PaNOSC, ExPaNDS** are **EOSC H2020** projects
- **OSCARS**, **OSTrails** are **EOSC Horizon Europe** projects
- **Wikipedia, Internet, ChatGPT**

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

Thank you

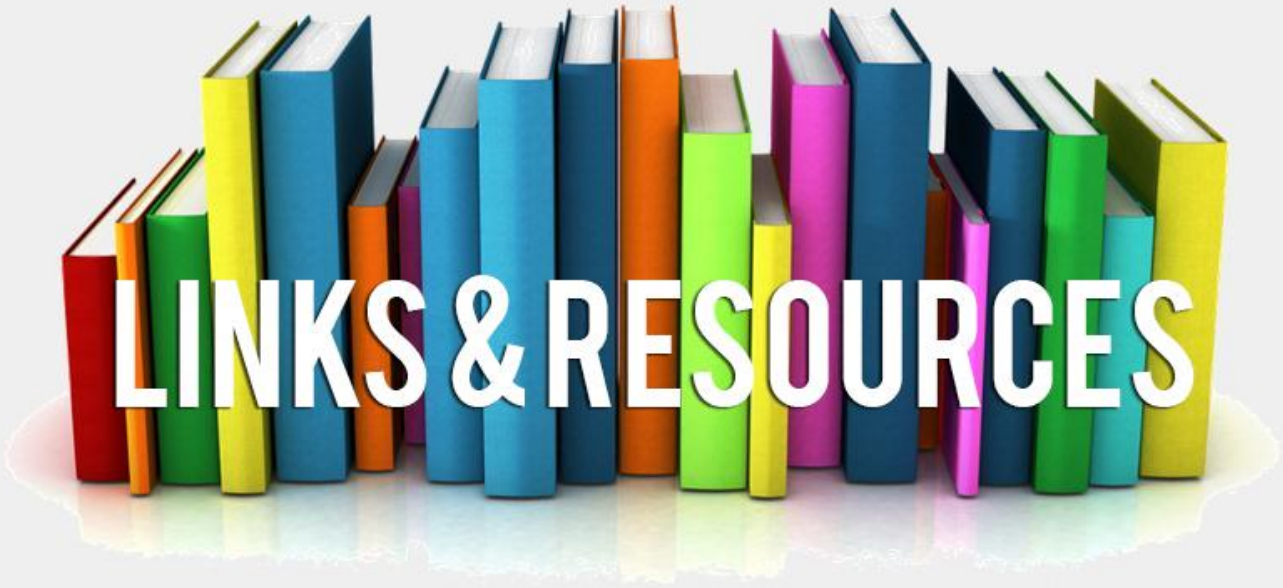
andy.gotz@esrf.fr



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

Backup slides

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

Tools to help you manage your research

A non-exhaustive list of tools to explore

- Elixir training course on “FAIR, Open Data and Open Science” https://oceantraining.eu/moodle/course/view.php?id=29
- 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

Extending FAIR data management with processed data integration in the ESRF data portal — *Marjolaine Bodin*
Hybrid event, ESRF Auditorium 09:25 - 09:40

Tiled in the Context of Data and Metadata Services — *Dan Allan*
Hybrid event, ESRF Auditorium 09:40 - 09:55

The Data Catalogue service and use case at SESAME — *Malik Almohammad*
Hybrid event, ESRF Auditorium 09:55 - 10:10

Rolling out a sample lifecycle and experimental data catalog at SOLEIL — *Gwenaëlle Abeillé*
Hybrid event, ESRF Auditorium 10:10 - 10:25

DOMAS: a data management software framework for advanced light sources — *Hao Hu*
Hybrid event, ESRF Auditorium 10:55 - 11:10

Latest news from the SciCat ecosystem: ElasticSearch integration and Job sub-system — *Massimiliano Novelli*
Hybrid event, ESRF Auditorium 11:10 - 11:25

Navigating the Data Patchwork: Strategies for Integrating Metadata Catalogs, Data Publications, and Archives
Oliver Knodel

ALBA: Towards FAIR data management principles — *Oriol Vallcorba*
Hybrid event, ESRF Auditorium 11:40 - 11:55

The data pipeline for the European Spallation Source ERIC — *Fredrik Bolmsten*
Hybrid event, ESRF Auditorium 11:55 - 12:10

Ontological definition of experimental techniques for FAIR data — *Wout De Nolf*
Hybrid event, ESRF Auditorium 14:10 - 14:25

DRACO-HELIPORT integration for metadata enhanced data-acquisition — *Mani Lokamani*
Hybrid event, ESRF Auditorium 14:25 - 14:40

NOBUGS Conference 2024 had a session on FAIR data with 11 Talks (see website for video + slides)

https://indico.esrf.fr/event/114/sessions/157/#20240926

panosc

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

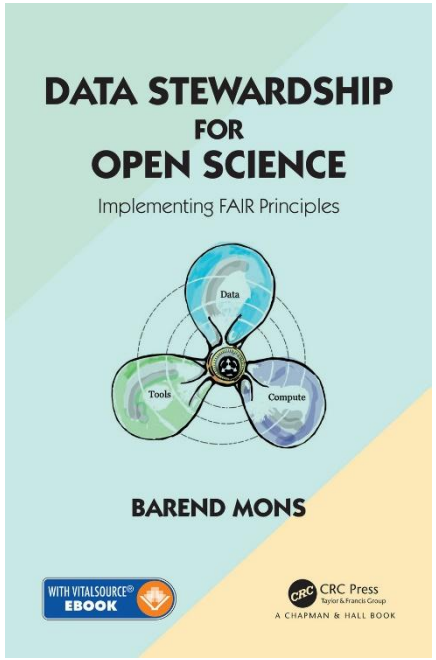
https://study.sagepub.com/corti2e

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

Learning more about FAIR RDM for data managers

- 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

OSTrails – an EOSC project to build Knowledge Graphs from Data Management Plans

Project

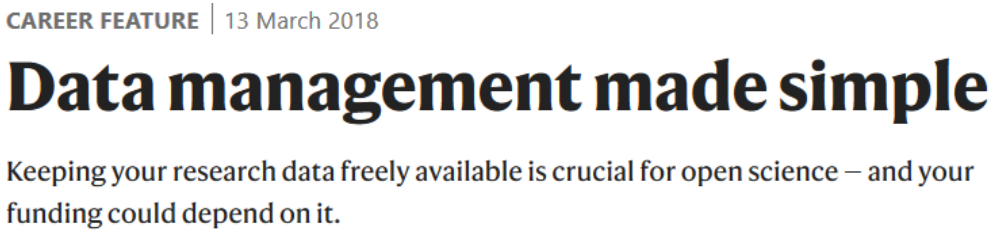
OSTrails

OSTrails aims to advance processes and instruments for Planning, Tracking, and Assessing scientific knowledge production beyond state-of-the art, working with various national and thematic contexts, improving existing infrastructure, and connecting key components. For the Plan stage, OSTrails aims to increase the efficacy of Data Management Plans, turning them from static narrative to living, interconnected "machine actionable" resources, making them the instrument of choice for improving quality of RDM. For the Track stage, OSTrails is set to establish an open, interoperable and high-quality ecosystem of Scientific Knowledge Graphs, enriching them to become evidence of communities' FAIR implementations.

rant 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

State of Open Data 2023

Support is not making its way to those who need it

Almost three-quarters of respondents have never received support with planning, managing or sharing research data.

With the global increase in policies and mandates to share data openly, who researchers are approaching for support becomes a pertinent question.

If respondents stated that they were aware of the concept of a data management plan, they were then asked if they have access to support from specialist data managers and we saw over 50% of our respondents state that they do have access to specialist research data managers in their research setting, but who else has been providing support?

Almost three quarters of respondents **had never received support** with planning, managing or sharing research data. When respondents were asked if they had ever received support with managing or making their data openly available, only 23% said they had. Of that 23%, 61% received support from informal internal sources such as colleagues or supervisors. Two other sources of support ranked highly with our respondents; their institutional libraries (31%) and their research office/in-house institutional expertise (26%).

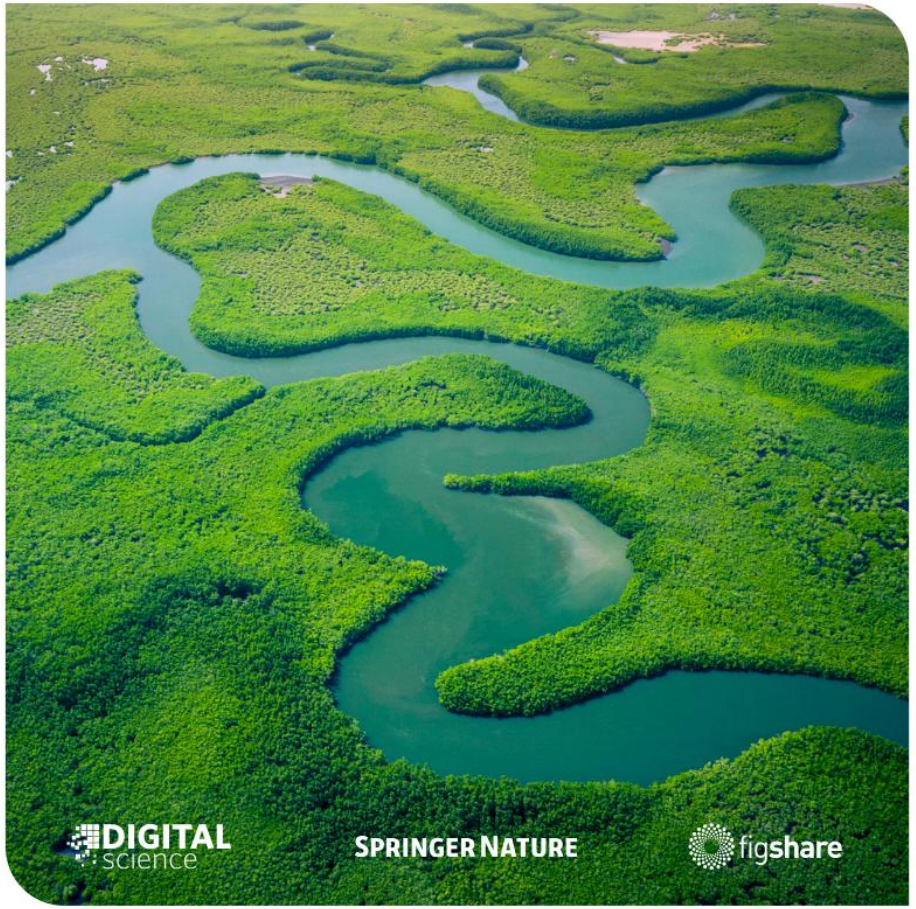
Graph showing the responses to the question *'Do you have access to support from specialist data managers?'* This question was only asked if respondents stated that they were aware of the concept of a data management plan. This graph shows the number of respondents for each answer.

https://www.digital-science.com/state-of-open-data/

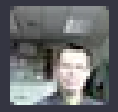
The State of Open Data 2023

The longest-running longitudinal survey and analysis on open data.

With opening remarks from Springer Nature's CPO, Harsh Jegadeesan, and Digital Science's CEO, Daniel Hook.
Authors Mark Hahnel, Graham Smith, Niki Scaplehorn, Henning Schoenenberger and Laura Day.



ChatGPT: What is Open Science?



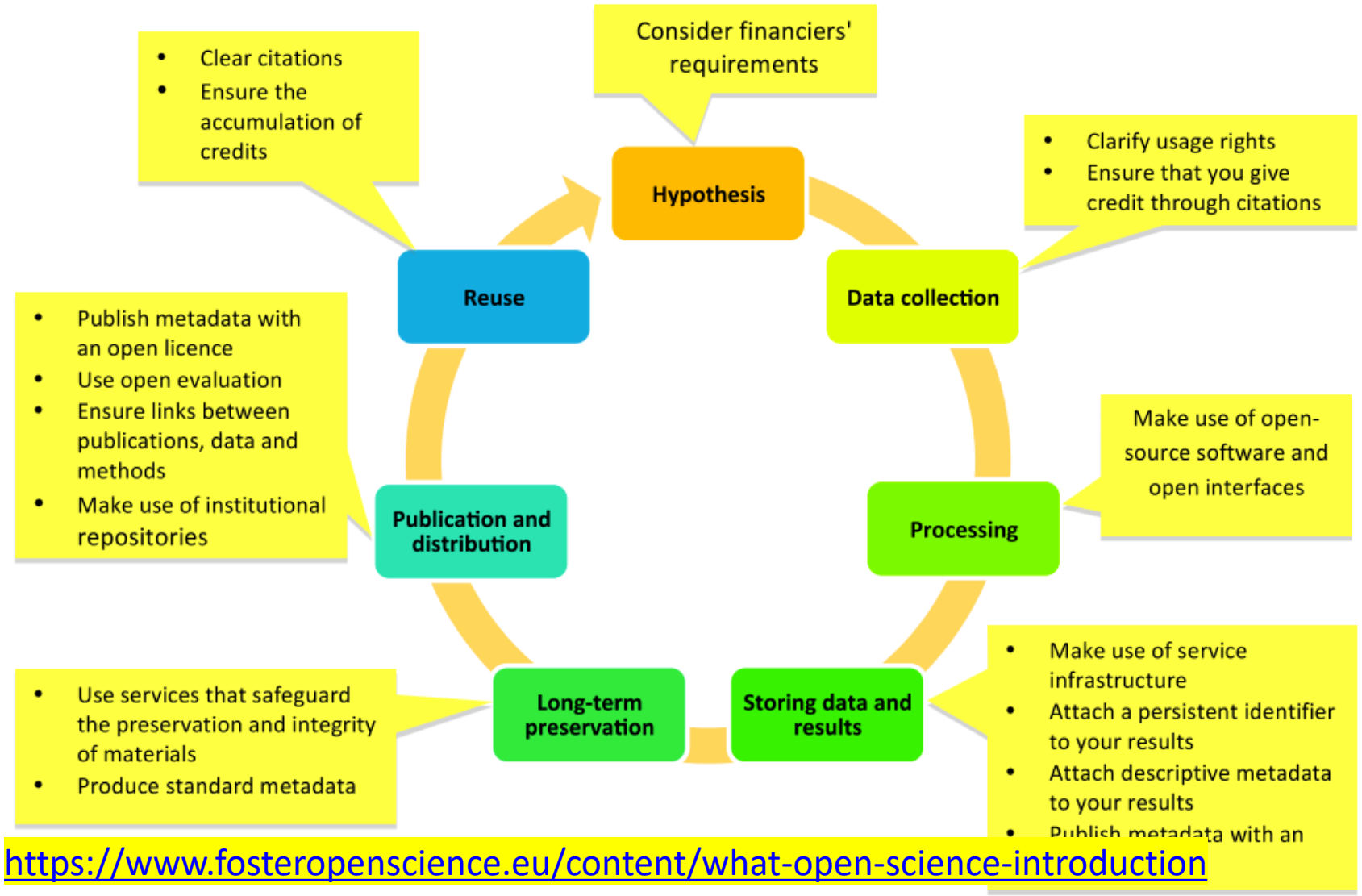
what is open science

Open science is an umbrella term that encompasses a variety of practices and principles aimed at making scientific research more transparent, accessible, and reproducible. This includes making research data, methods, and findings openly available to the public, as well as promoting collaboration and openness in the scientific process. The goal of open science is to increase the efficiency and effectiveness of scientific research, and to ensure that the knowledge generated by the scientific enterprise is widely accessible and can be used to benefit society. Open science practices can include open data sharing, open access publishing, preregistration of study designs, and the use of open-source software.

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

https://chat.openai.com/chat

Open Science is about extending the principles of openness to the whole research cycle (FOSTER)



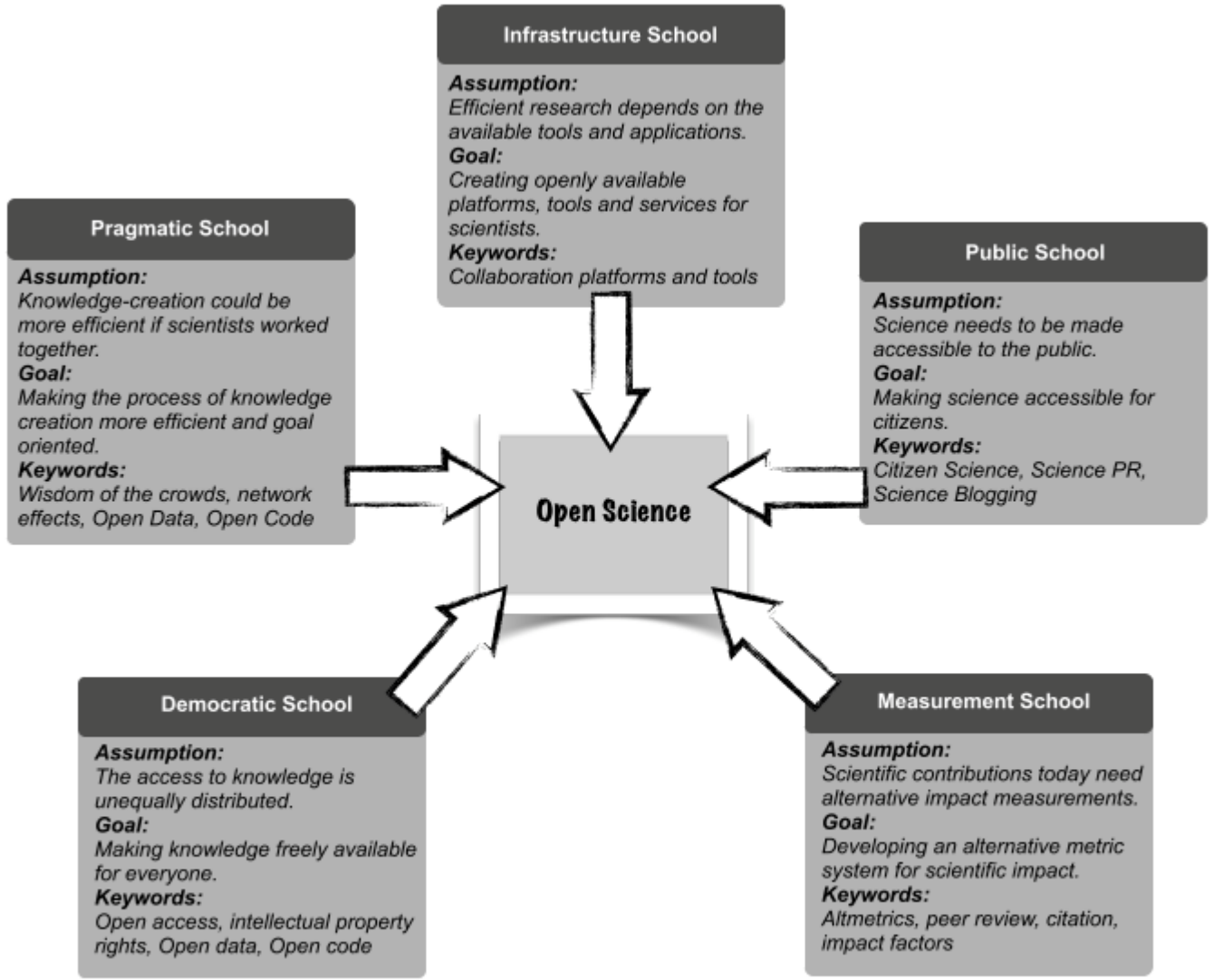
This project has rece

https://www.fosteropenscience.eu/content/what-open-science-introduction

OSCARS
Open Science Clusters' Action for Research & Society

panosc

Five schools of Thought for Open Science



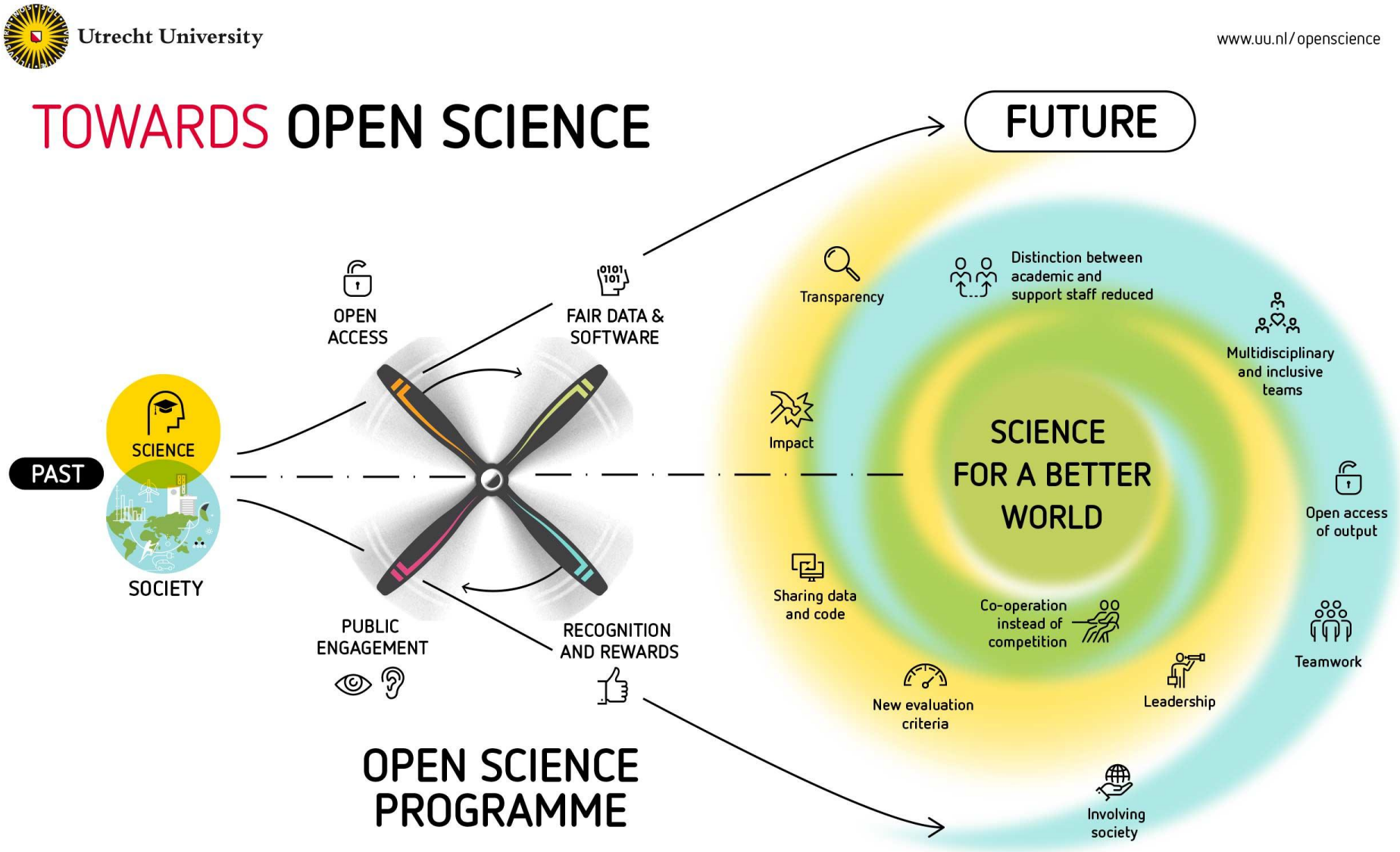
This project ha

Fecher, B. and Friesike, S. (2013). Open Science: One Term, Five Schools of Thought.

OSCARS
Open Science Clusters' Action for Research & Society

panosc

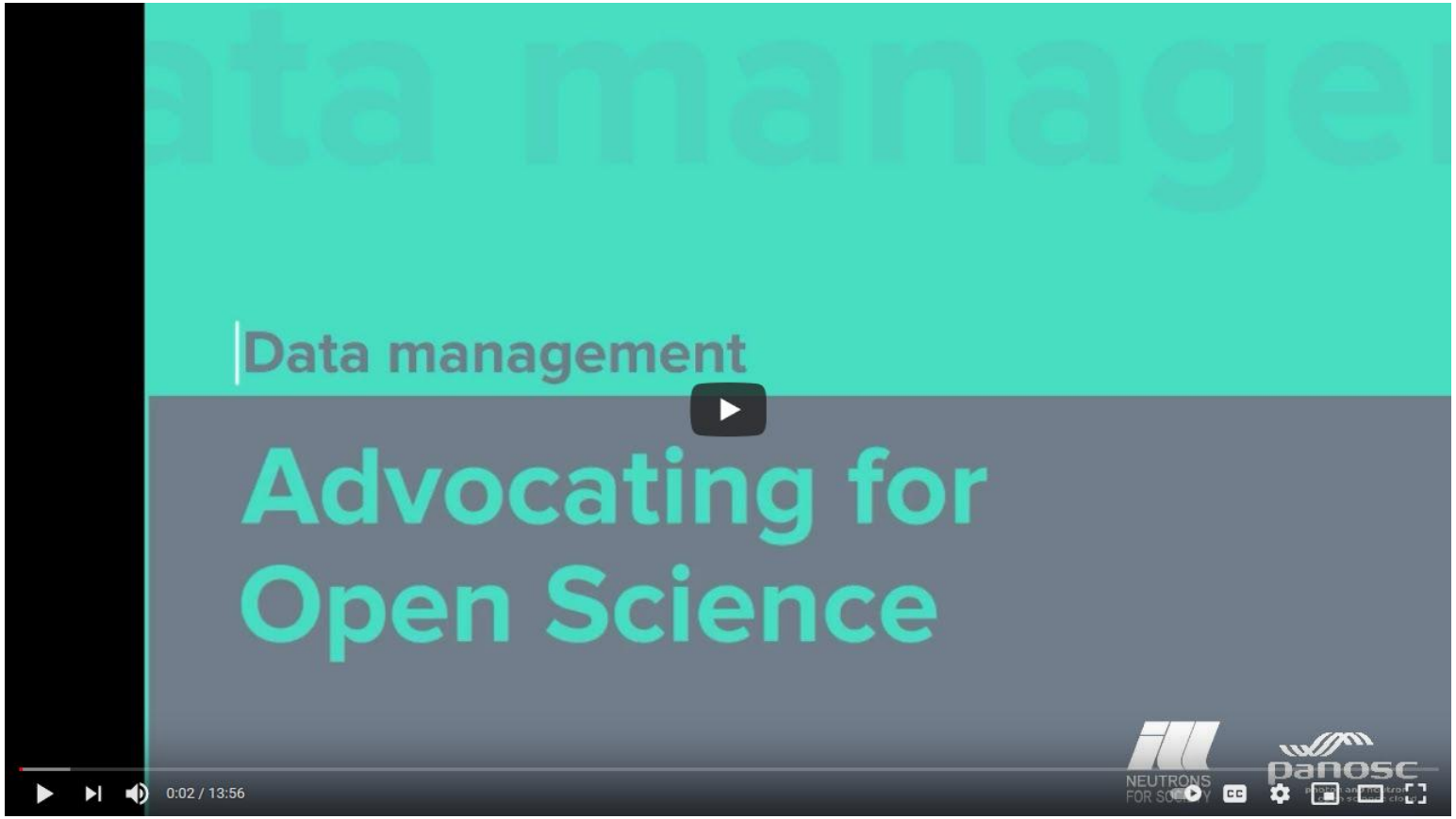
Impact of open science on the future



https://narratives.insidehighered.com/four-pillars-of-open-science/assets/rALfdOrDqh/uugraphic-2-2560x1597.jpeg

Open Science Ambassador

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

OSCARS
Open Science Clusters' Action for Research & Society

panosc

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