panosc
photon and neutron
open science cloud

School on
Synchrotron Light Sources
and their Applications

ICTP

ESRF
The European Synchrotron

The role of the [Open] Scientist in making Data FAIR for Reproducible Science

School on Synchrotron Light Sources and their Applications

15th January, 2026

Presenter: **Andy Götz** (ESRF Data Manager + PaNOSC coordinator)

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

Outline of Talk

This talk will address the topic of Open Science and FAIR Data for scientists doing research in order to answer the following questions:

- **Part 1**
  - **What is Open Science?**
  - **Why make Science Open?**
- **Part 2**
  - **What is FAIR Data ?**
  - **Why make Data FAIR ?**
- **What is your role in all of this ?**

Andy Götz – a bit about me

1. Started as radio astronomer (1984 – 1987)
2. Joined **ESRF** in **1988**, worked on **accelerator controls, beamline controls, data management, open science** (1988 – now)
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)
6. **Coordinator** of the **Photon and Neutron Open Science Cloud Node** of the **EOSC Federation** (2025 – now)

Kilobytes
to
Petabytes
in
40 years

Open Science

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

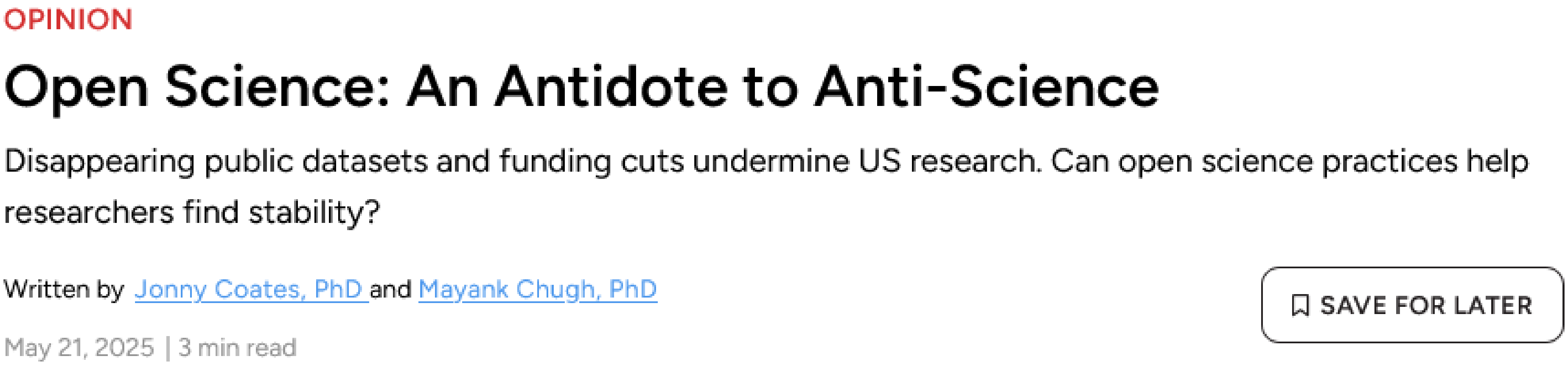
Open Science is a major movement Worldwide to change the way Science is conducted

Why a talk about Open Science + FAIR data?

FAIR data is one of the pillars of OS

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

Open Science in the News



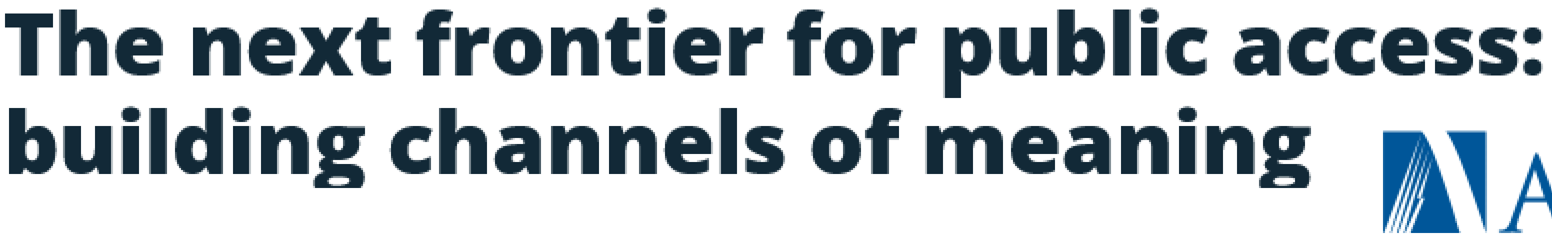
OPINION

Open Science: An Antidote to Anti-Science

Disappearing public datasets and funding cuts undermine US research. Can open science practices help researchers find stability?

Written by Jonny Coates, PhD and Mayank Chugh, PhD

May 21, 2025 | 3 min read



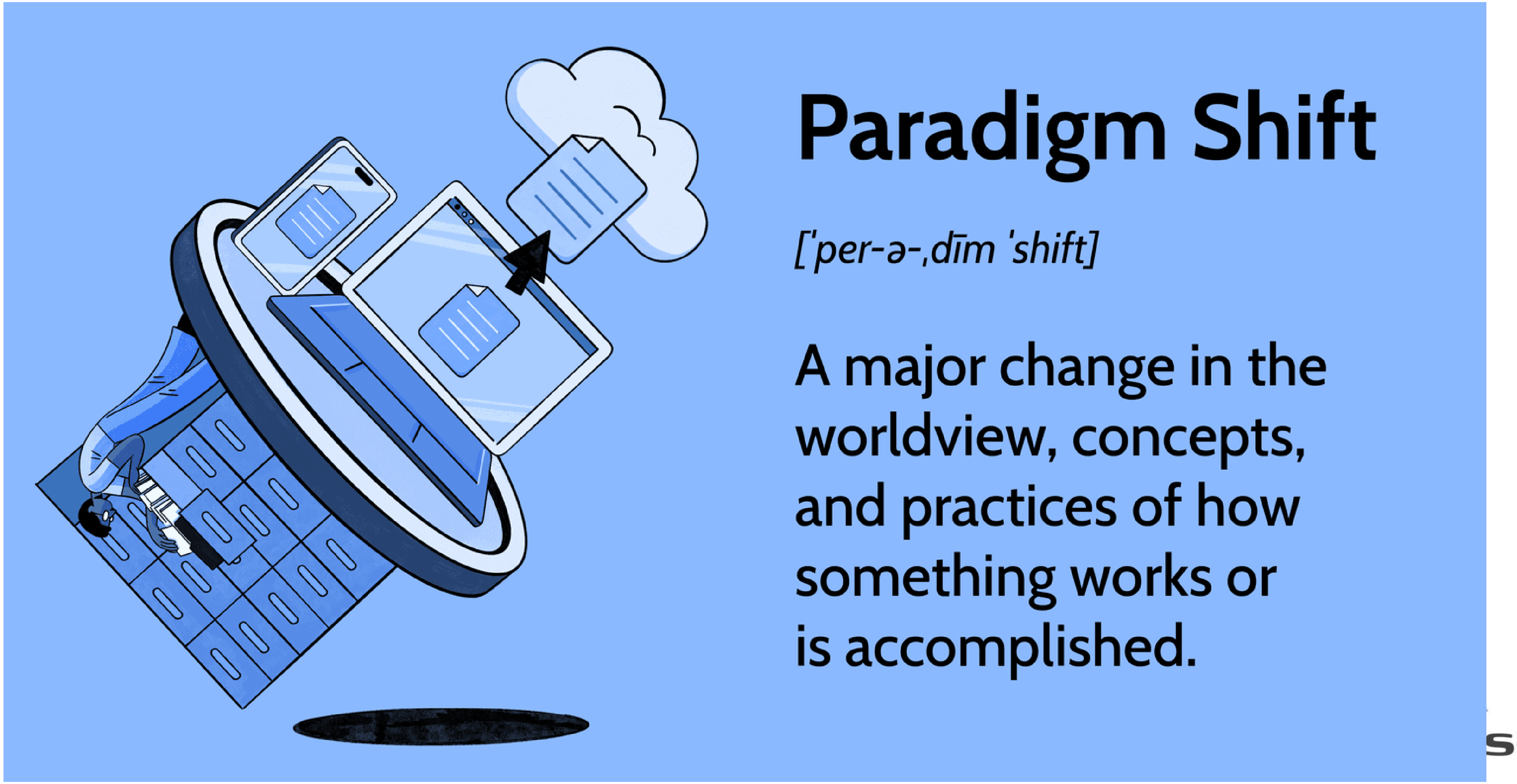
The next frontier for public access: building channels of meaning

AAAS
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

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 a Paradigm Shift in Science



Paradigm Shift

['per-ə-ˌdīm 'shift]

A major change in the worldview, concepts, and practices of how something works or is accomplished.

Open Science is a Culture Change

https://openscientist.pubpub.org

Published on Nov 22, 2020 DOI 10.21428/8bbb7f85.35a0e14b

Open Scientist Handbook

Version 2.0 with edits.

by Bruce R. Caron

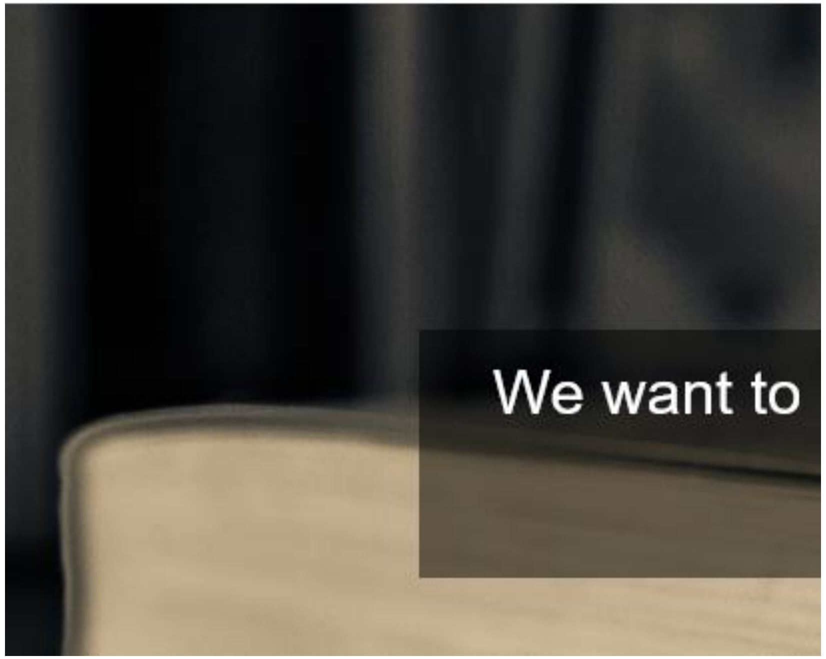
last released
2 years ago

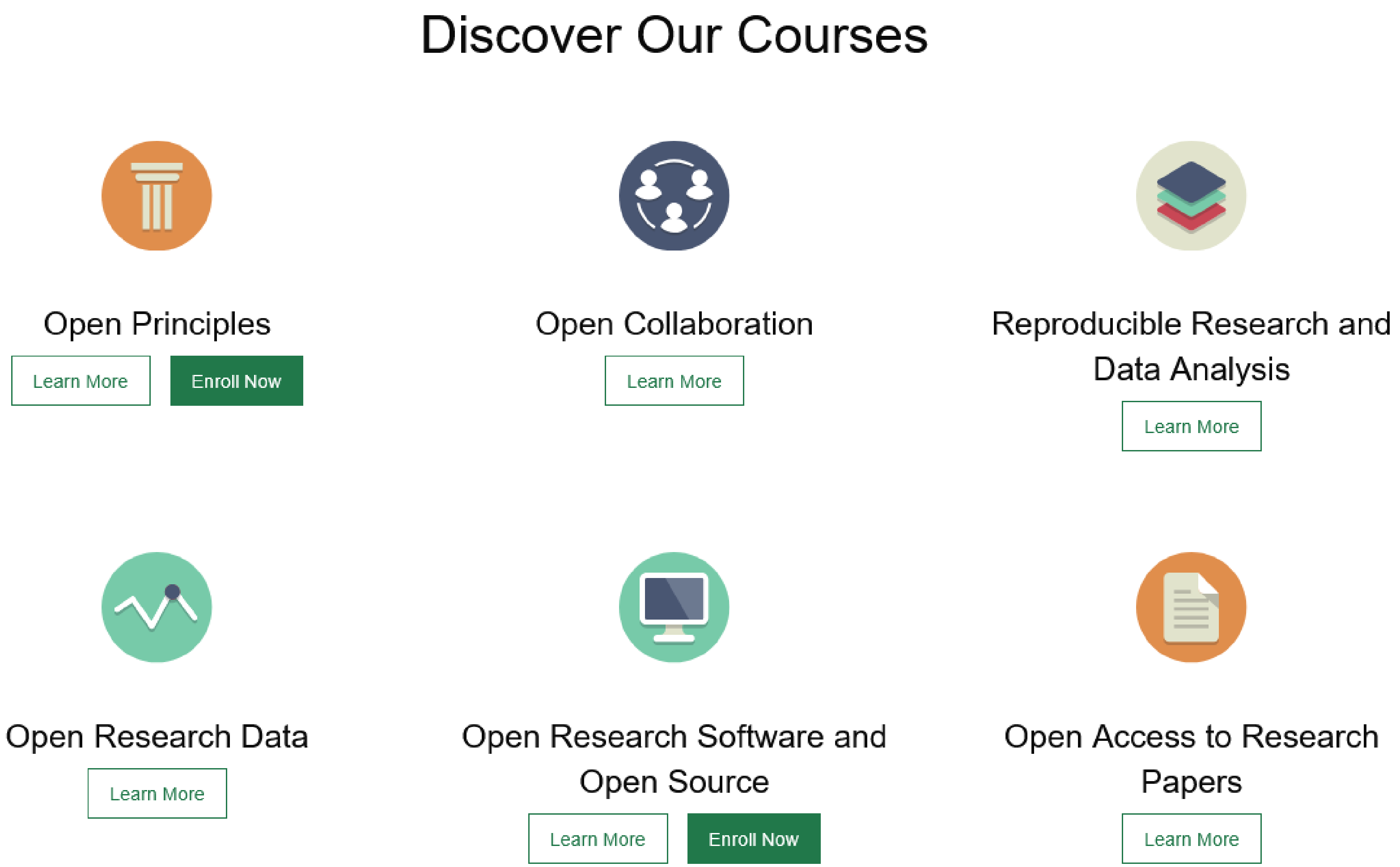
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 823852

Open Science is a Culture Change

https://opensciencemooc.eu/

OPEN SCIENCE MOOC
FREE | OPEN | LEARNING





Discover Our Courses

Open Principles

Learn More | Enroll Now

Open Collaboration

Learn More

Reproducible Research and Data Analysis

Learn More

Open Research Data

Learn More

Open Research Software and Open Source

Learn More | Enroll Now

Open Access to Research Papers

Learn More

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

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

“Publication is not finished until you publish the data”

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

Why is science considered “closed” ?

Why do we need open science?

- **Results** (publicati
  which require a (ofte
  available to the gene
- **Evidence** (scienti
  available as part of th
- **Community** (so

https://youtu.be/jLJ7ZO3wOW4

This project has received funding from the European Union’s H

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

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

Open Science is about sharing

- **Publications –** results need to be shared through text-based publications; publications need to be made accessible by everyone (Open Access)
- **Peer review –** the reviewing process should be transparent (Open Peer Review)
- **Data –** data must be made available to verify results and to allow others to derive new results from the data (Open Data)
- **Software –** making software used to derive the results available ensures traceability

- **Understanding – science is based on consensus of the scientific community;**
  it doesn't "work" if you keep your results + data to yourself;
- **Sharing results allows + encourages citizens to understand and be involved in science (often called Citizen Science)**

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

The Tradition before Open Science

- **Limited Access to Publications**:
  - Paywalled journals restrict access to scientific knowledge, preve… se in less privileged circumstances, from benefiting fr… n of new knowledge in the real world.
- **Lack of Transparency**:
  - The close… king it difficult to verify and
-
  - …d, fearing that sharing it will jeopardize their competitive edge or … funding.
- **Sc**
  - Software used to produce results are not open source making it difficult to reproduce the plots and results in publications

Scientists have been doing Open Science implicitly since centuries by sharing their ideas and results in publications

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

https://opusproject.eu/openscience-news/science-for-everyone-why-not-open-science/

Benefits of Open Science

Democratization of Knowledge: Open Science breaks down barriers, making scientific knowledge accessible to a wider audience, including students, educators, policymakers, and the general public.

Faster Innovation: By sharing data and findings openly, researchers can build upon each other’s work more rapidly, accelerating the pace of innovation and problem-solving.

Greater Transparency: Open Science promotes transparency and accountability, reducing the likelihood of scientific misconduct and enhancing the credibility of research.

Increased Collaboration: Collaborative efforts fostered by Open Science can lead to breakthroughs and solutions to complex challenges that may have been beyond the reach of individual researchers or institutions.

Engaging the Public: Open Science allows citizens to participate more actively in the scientific process, fostering a sense of ownership and trust in scientific endeavors.

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

https://opusproject.eu/openscience-news/science-for-everyone-why-not-open-science/

Many questions

https://www.ouvrirlascience.fr/join-the-debate/

In my field, we don't do open science

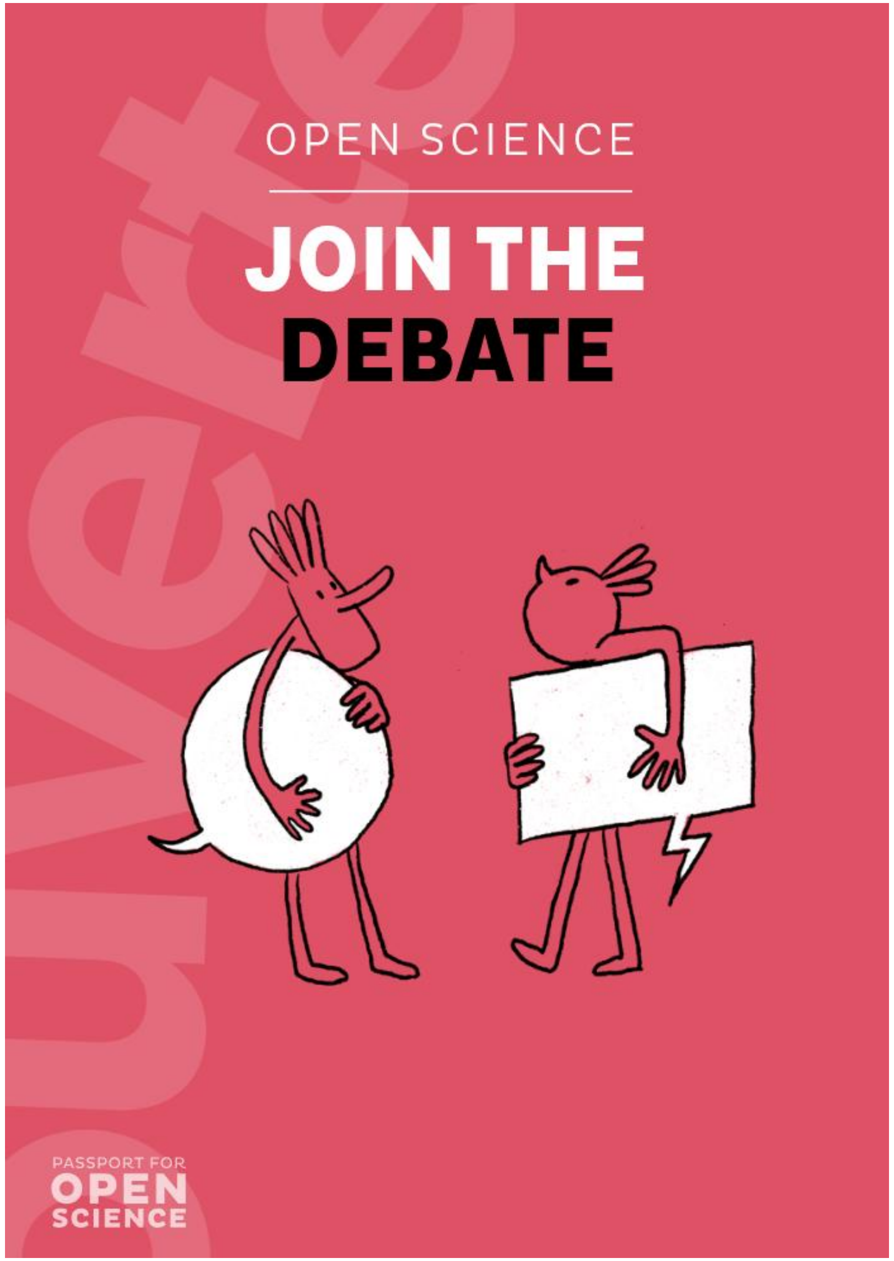
If I disseminate my scientific work in open access, everyone will be able to use it without citing me

Open access is a threat to certain publishers

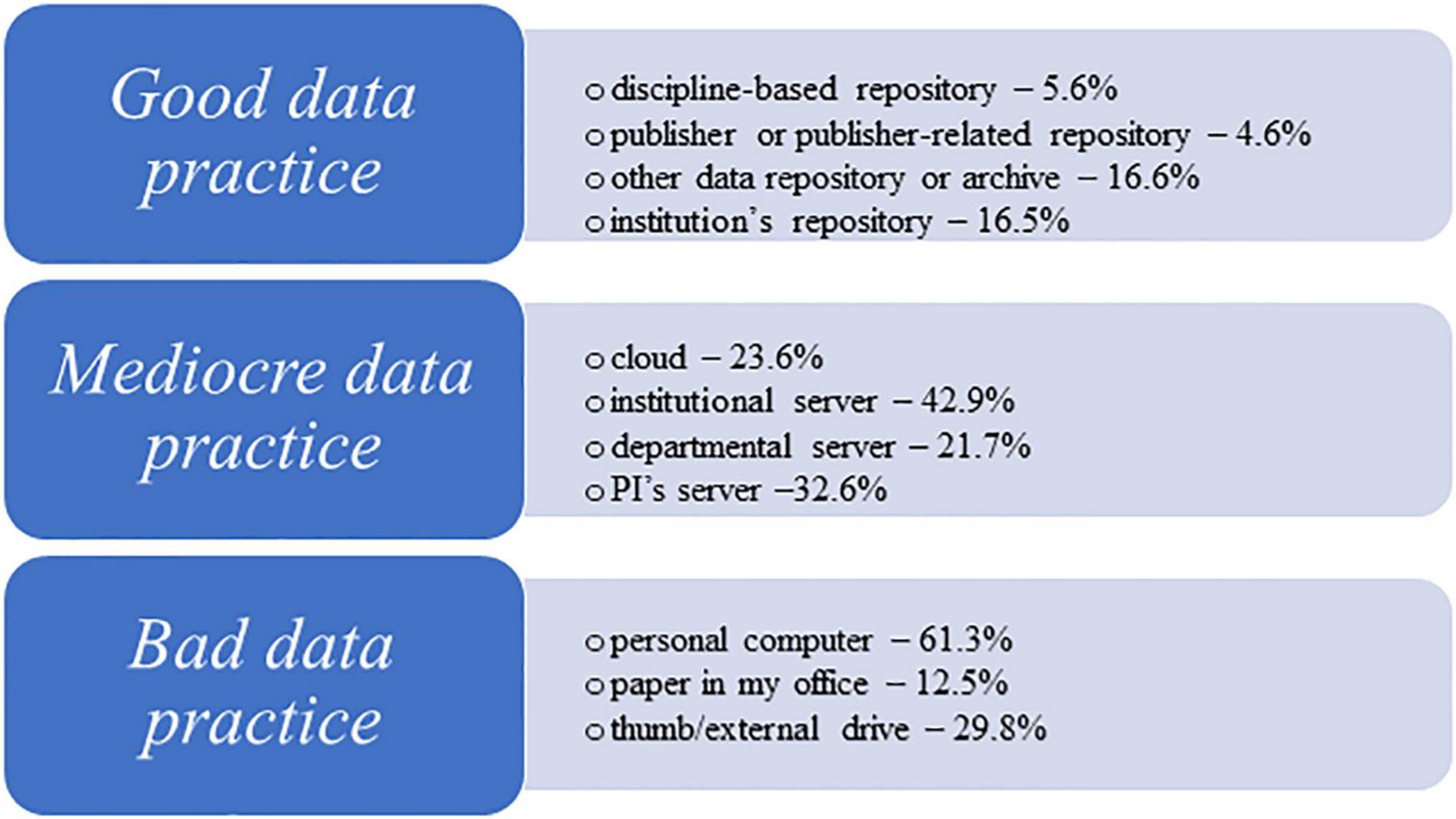
Open access publishing is too costly for my institute

In my field I have to choose a journal based only on the impact factor

A data management plan will simply increase my workload without benefiting me



Adoption of good data practices are still poor



Good data practice

- ○ discipline-based repository – 5.6%
- ○ publisher or publisher-related repository – 4.6%
- ○ other data repository or archive – 16.6%
- ○ institution's repository – 16.5%

Mediocre data practice

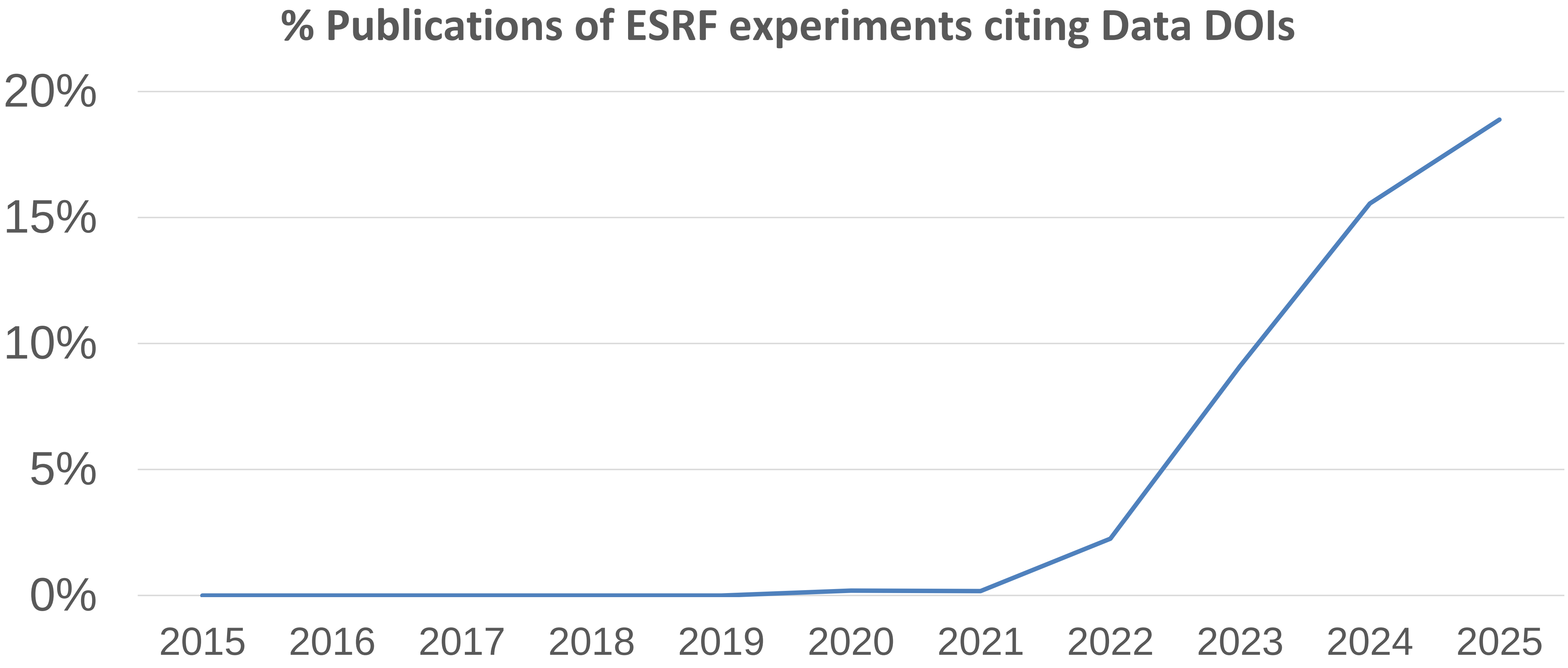
- ○ cloud – 23.6%
- ○ institutional server – 42.9%
- ○ departmental server – 21.7%
- ○ PI's server –32.6%

Bad data practice

- ○ personal computer – 61.3%
- ○ paper in my office – 12.5%
- ○ thumb/external drive – 29.8%

Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. PLoS ONE 15 (3), March 11 2020: e0229003. https://doi.org/10.1371/journal.pone.0229003

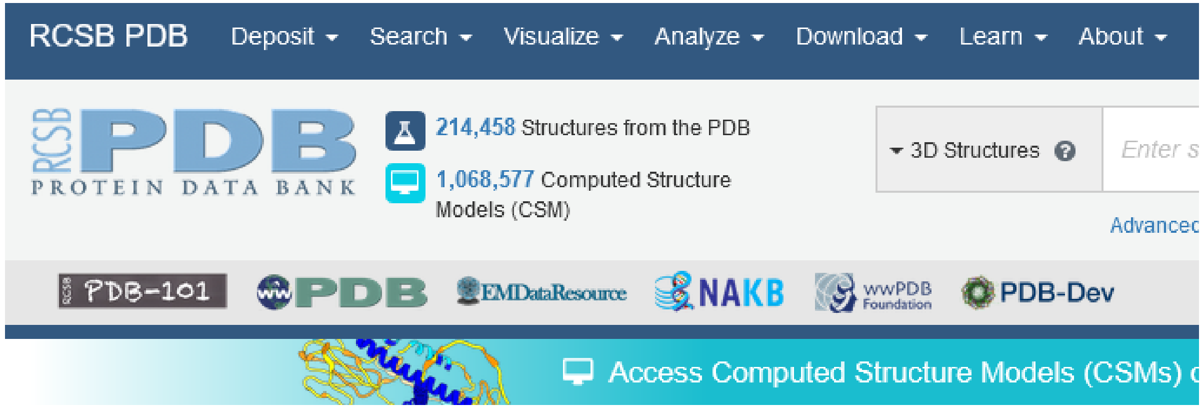
Citing of Data DOIs is increasing …



Our goal is 100% citation of Data DOIs …

Examples of Open Science from photon and neutron science

Protein Data Bank + AlphaFold : Open data since 1971; the 200k+ protein structures have been used to train an AI alphofold to predict 3 million+ protein structures



214,458 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

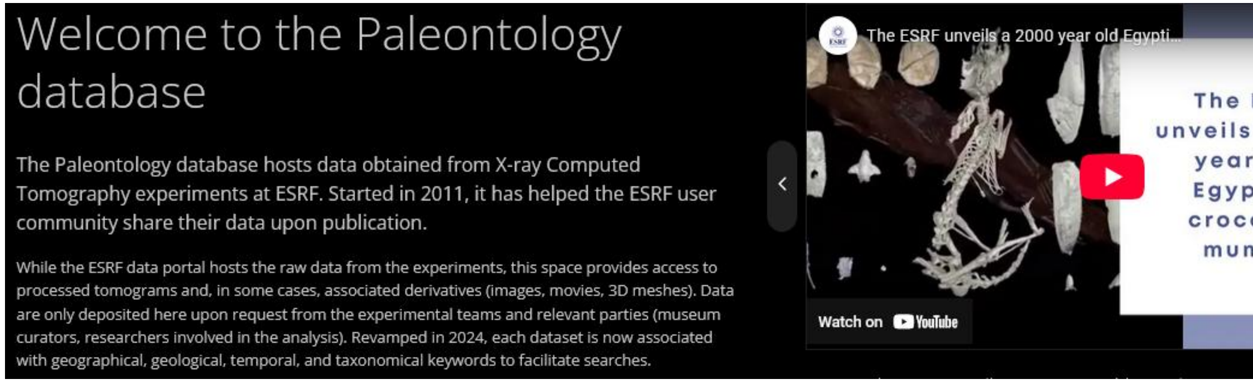
Access Computed Structure Models (CSMs) o

Paleontology Database : https://paleo.esrf.fr

Of 399 articles published:

→ 342 publications from users

→ 57 linked to open data by non-users



Welcome to the Paleontology database

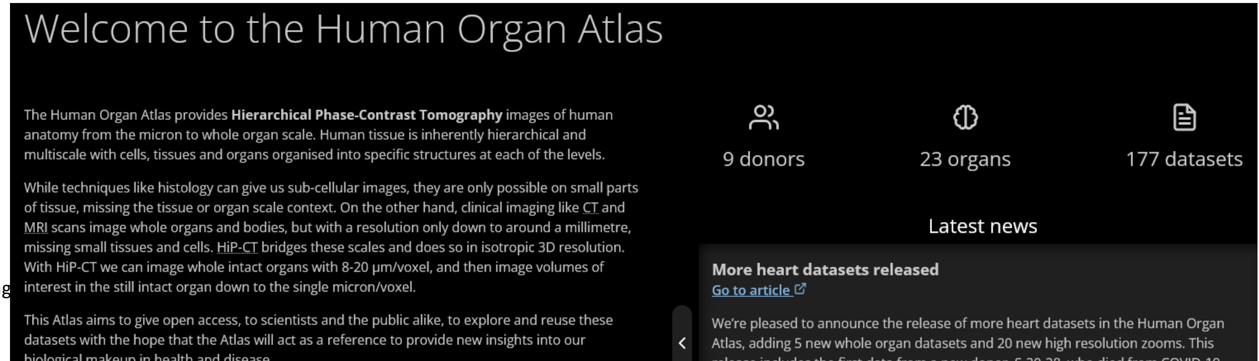
The Paleontology database hosts data obtained from X-ray Computed Tomography experiments at ESRF. Started in 2011, it has helped the ESRF user community share their data upon publication.

While the ESRF data portal hosts the raw data from the experiments, this space provides access to processed tomograms and, in some cases, associated derivatives (images, movies, 3D meshes). Data are only deposited here upon request from the experimental teams and relevant parties (museum curators, researchers involved in the analysis). Revamped in 2024, each dataset is now associated with geographical, geological, temporal, and taxonomical keywords to facilitate searches.

Human Organ Atlas :

→ 28 publications from experimental teams

→ 8 publications from open data



Welcome to the Human Organ Atlas

The Human Organ Atlas provides **Hierarchical Phase-Contrast Tomography** images of human anatomy from the micron to whole organ scale. Human tissue is inherently hierarchical and multiscale with cells, tissues and organs organised into specific structures at each of the levels.

While techniques like histology can give us sub-cellular images, they are only possible on small parts of tissue, missing the tissue or organ scale context. On the other hand, clinical imaging like CT and MRI scans image whole organs and bodies, but with a resolution only down to around a millimetre, missing small tissues and cells. HiP-CT bridges these scales and does so in isotropic 3D resolution. With HiP-CT we can image whole intact organs with 8-20 μm/voxel, and then image volumes of interest in the still intact organ down to the single micron/voxel.

This Atlas aims to give open access, to scientists and the public alike, to explore and reuse these datasets with the hope that the Atlas will act as a reference to provide new insights into our biological makeup in health and disease.

9 donors

23 organs

177 datasets

Latest news

More heart datasets released

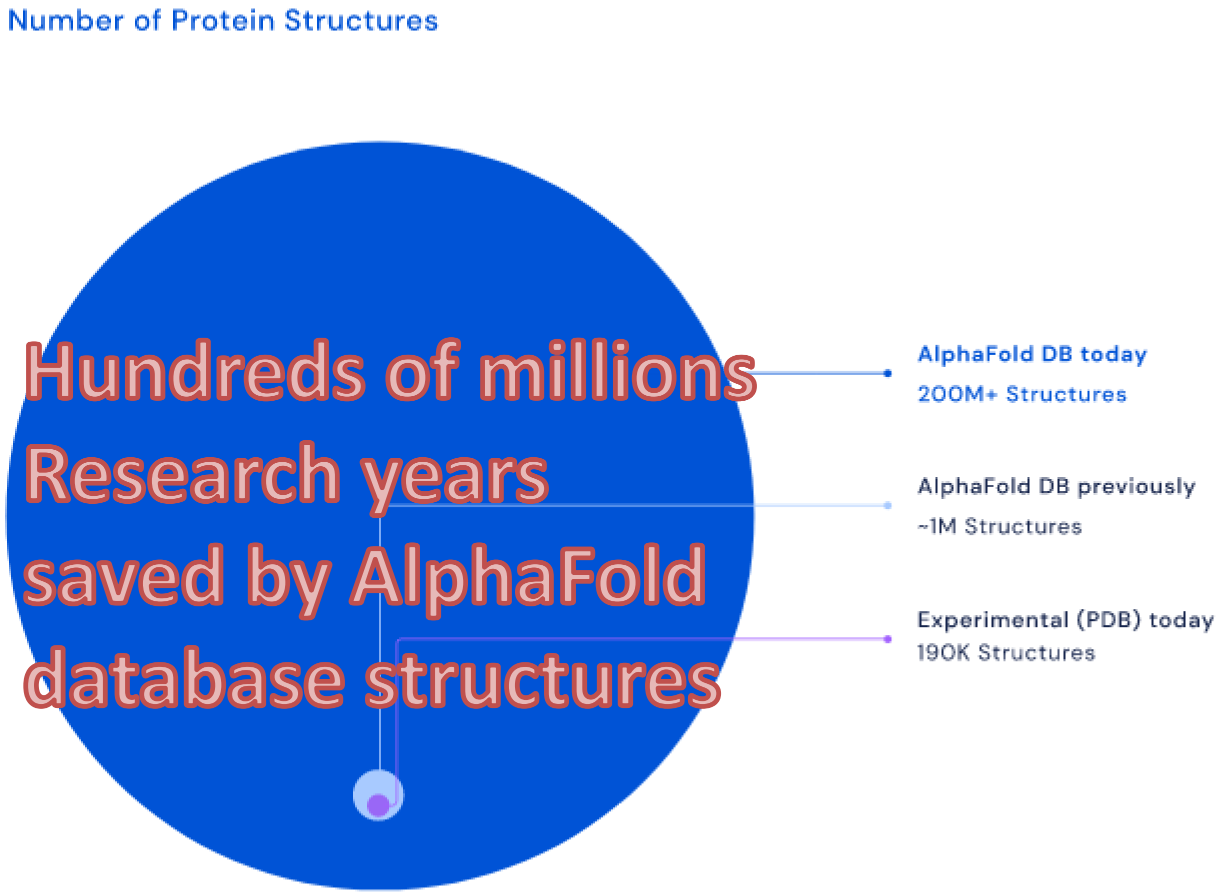
Go to article

We're pleased to announce the release of more heart datasets in the Human Organ Atlas, adding 5 new whole organ datasets and 20 new high resolution zooms. This release includes the first data from a new donor, S-20-28, who died from COVID-19

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

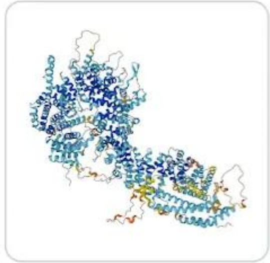
Alphafold demonstrates the power of Open Data + AI

Alphafold has basically solved the protein folding problem using AI an learning from the open data in the Protein Data Base (PDB)



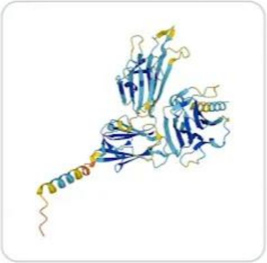
Hundreds of millions Research years saved by AlphaFold database structures

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



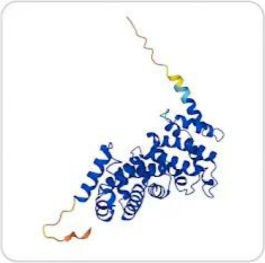
Nuclear pore complex protein Nup205

Part of a large complex that acts as a gateway in and out of the cell nucleus



Gametocyte surface protein P45/48

From the malaria parasite; a candidate protein for including in vaccines

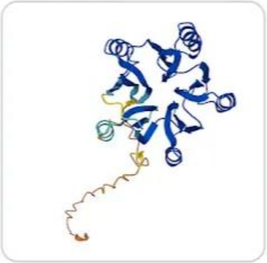


CCR4-NOT transcription complex subunit 9

Regulates an important cellular process (the rate of mRNA degradation)

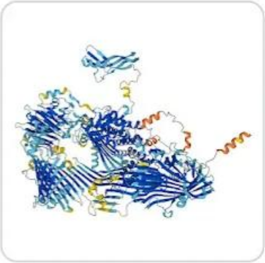
Ice nucleation protein

Bacterial protein that can trigger ice formation at relatively high temperatures, causing frost damage to plants



F20H23.2 protein

Plant protein; represents a potential new structural superfamily unlike anything seen before



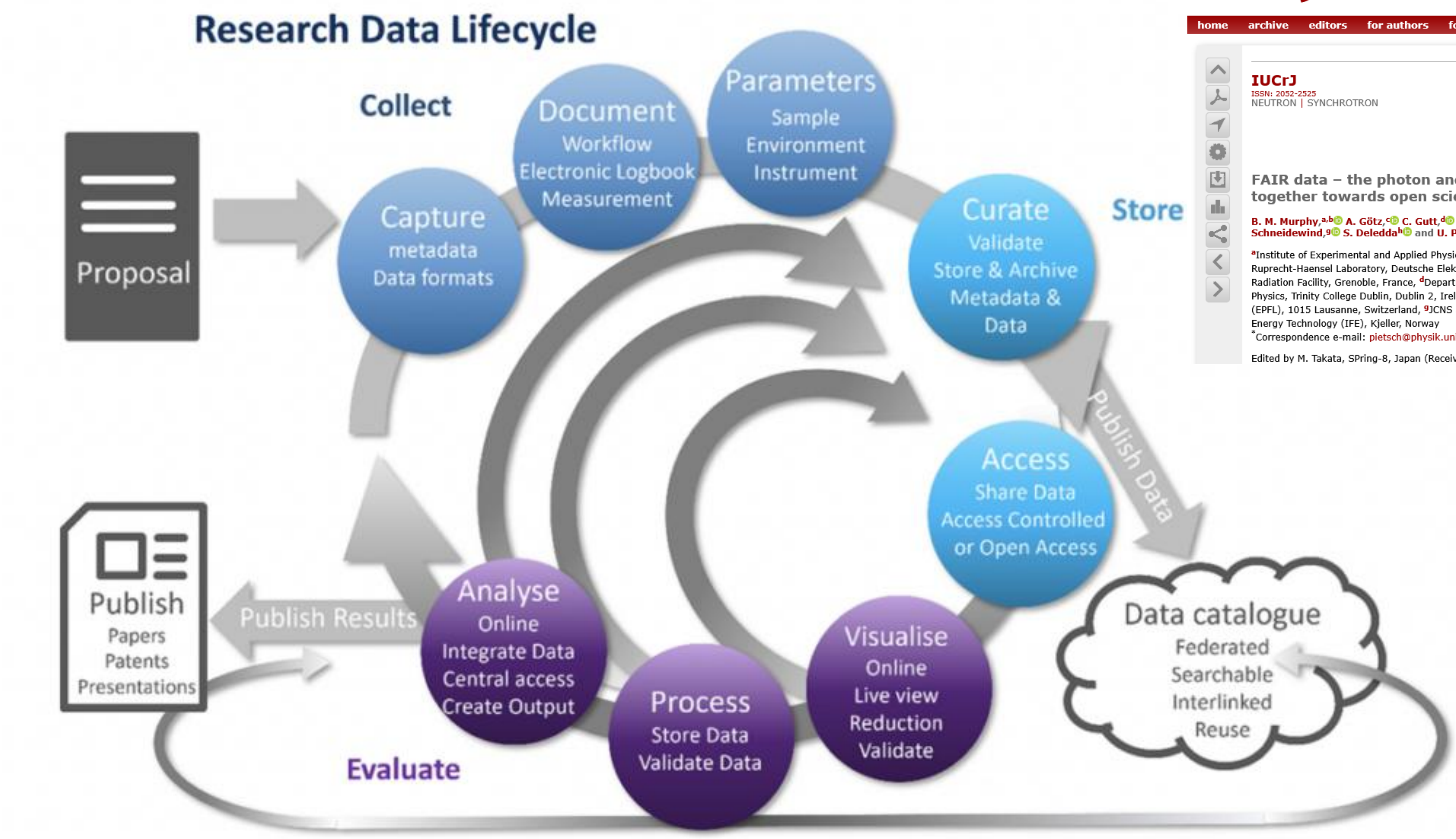
Vitellogenin

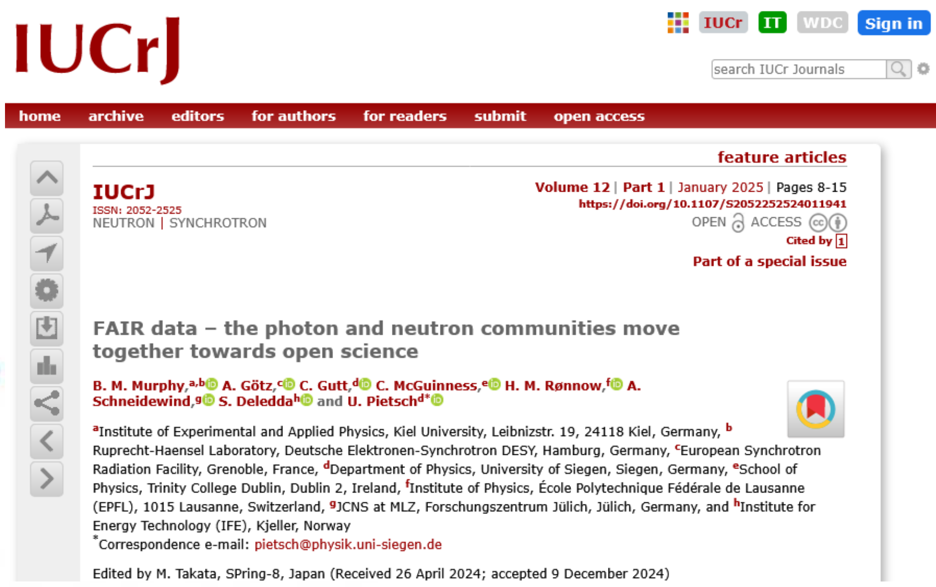
Involved in the immune system of egg-laying animals including honeybees

There is more to Scientific Publications than meets the eye!

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

There is much more to Scientific Publications than meets the eye!




IUCr IT WDC Sign in
search IUCr Journals
home archive editors for authors for readers submit open access

IUCrJ
ISSN: 2052-2525
NEUTRON | SYNCHROTRON

feature articles

Volume 12 | Part 1 | January 2025 | Pages 8-15
https://doi.org/10.1107/S2052252524011941
OPEN ACCESS
Cited by 1
Part of a special issue

FAIR data – the photon and neutron communities move together towards open science

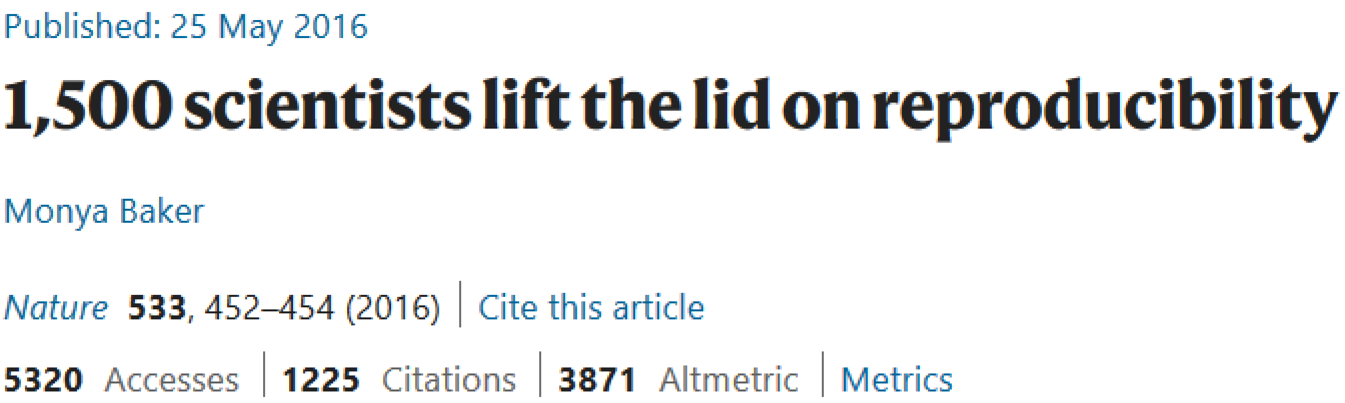
B. M. Murphy,[a,b] A. Götz,[c] C. Gutt,[d] C. McGuinness,[e] H. M. Rønnow,[f] A. Schneidewind,[g] S. Deledda[h] and U. Pietsch[d*]

[a]Institute of Experimental and Applied Physics, Kiel University, Leibnizstr. 19, 24118 Kiel, Germany, [b]Ruprecht-Haensel Laboratory, Deutsche Elektronen-Synchrotron DESY, Hamburg, Germany, [c]European Synchrotron Radiation Facility, Grenoble, France, [d]Department of Physics, University of Siegen, Siegen, Germany, [e]School of Physics, Trinity College Dublin, Dublin 2, Ireland, [f]Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland, [g]JCNS at MLZ, Forschungszentrum Jülich, Jülich, Germany, and [h]Institute for Energy Technology (IFE), Kjeller, Norway
*Correspondence e-mail: pietsch@physik.uni-siegen.de

Edited by M. Takata, SPring-8, Japan (Received 26 April 2024; accepted 9 December 2024)

B. M. Murphy, A. Götz, C. Gutt, C. McGuinness, H. M. Rønnow, A. Schneidewind, S. Deledda and U. Pietsch, *FAIR data – the photon and neutron communities move together towards open science*, in IUCrJ, 2024

Reproducibility and Replicability



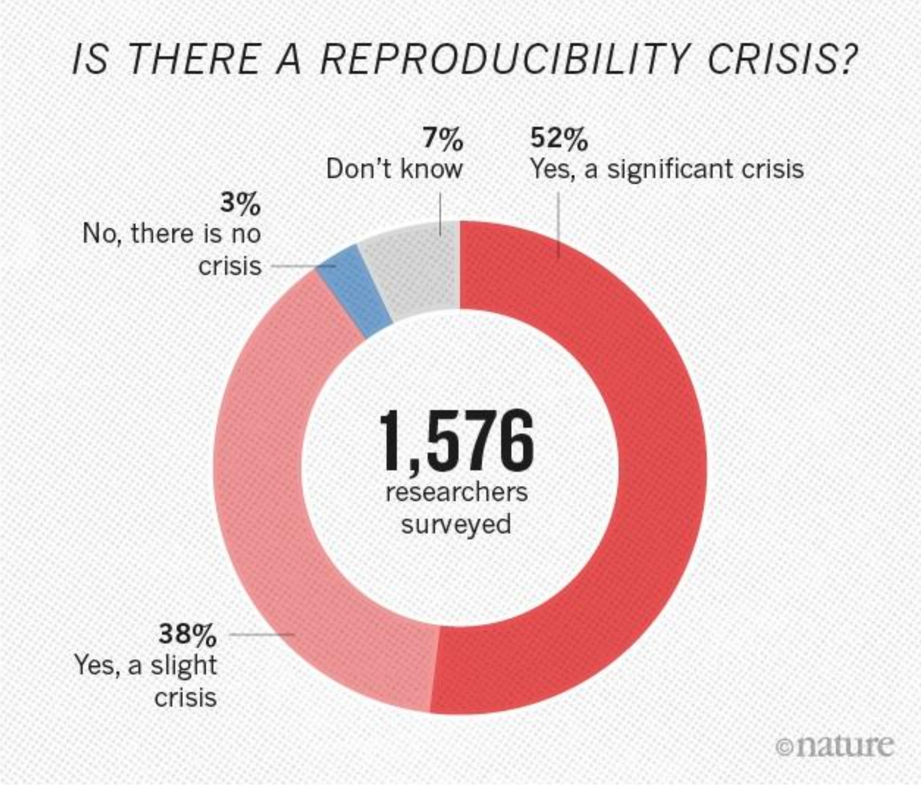
Published: 25 May 2016

1,500 scientists lift the lid on reproducibility

Monya Baker

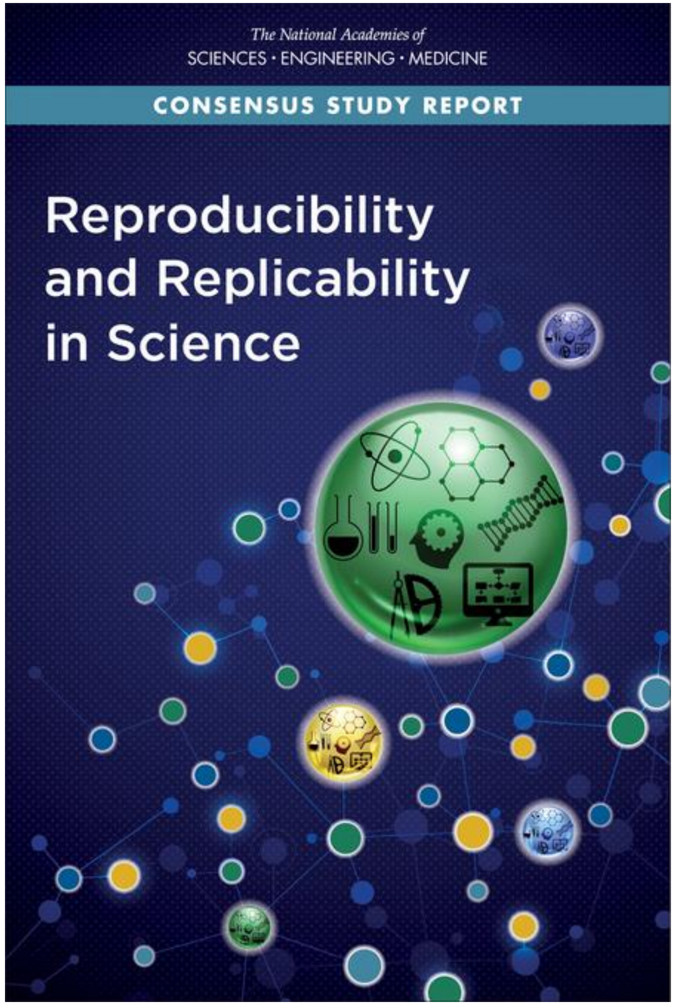
Nature **533**, 452–454 (2016) | Cite this article

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



IS THERE A REPRODUCIBILITY CRISIS?

©nature



The National Academies of
SCIENCES · ENGINEERING · MEDICINE

CONSENSUS STUDY REPORT

Reproducibility and Replicability in Science

Further reading:

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

ovation programme under grant agreement No. 823852

Open Access publications – Green vs. Gold

Your role: make sure your publications is in Open Access either in a journal or an archive e.g. national archive

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

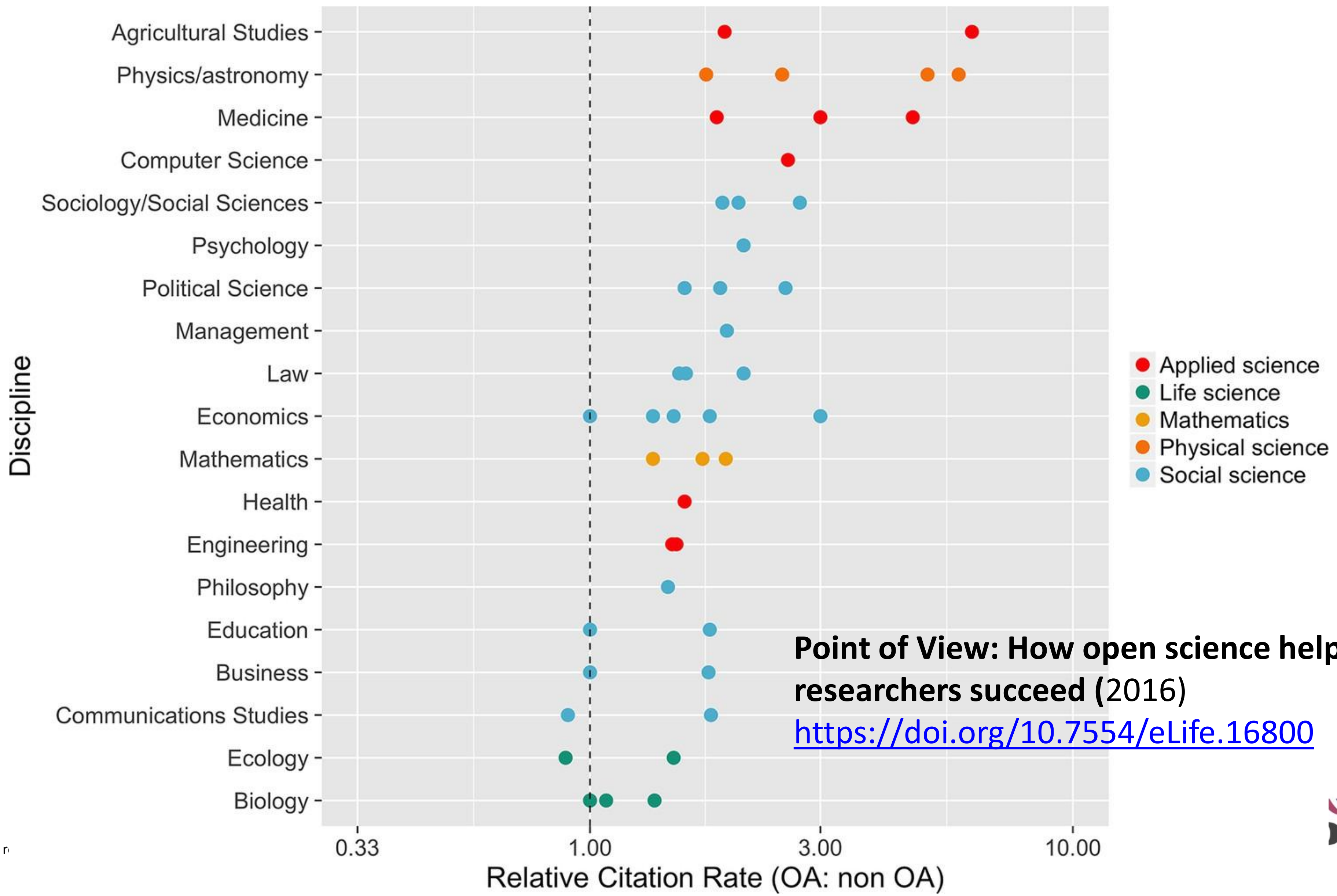
- Authors (or their funders or institutions) pay an Article Publication Charge (APC) upon acceptance
- The final published version is free immediately
- Bioscientifica deposits the article in PubMed Central
- Authors retain copyright and a range of licenses are available
- Journal could be fully open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

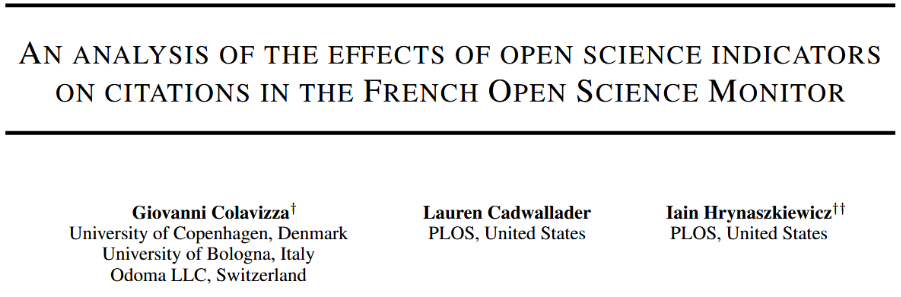
panosc

Open access leads to more citations



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

New study shows Open Science leads to more citations



An analysis of the effects of open science indicators on citations in the French Open Science Monitor

Giovanni Colavizza[†]
University of Copenhagen, Denmark
University of Bologna, Italy
Odoma LLC, Switzerland

Lauren Cadwallader
PLOS, United States

Iain Hrynaszkiewicz[††]
PLOS, United States

https://arxiv.org/abs/2508.20747

https://frenchopensciencemonitor.esr.gouv.fr

(https://barometredelascienceouverte.esr.gouv.fr/about/opendata)

65% of French publications are in Open Access

→ 46.8% substantial increase in citations if OSI adopted

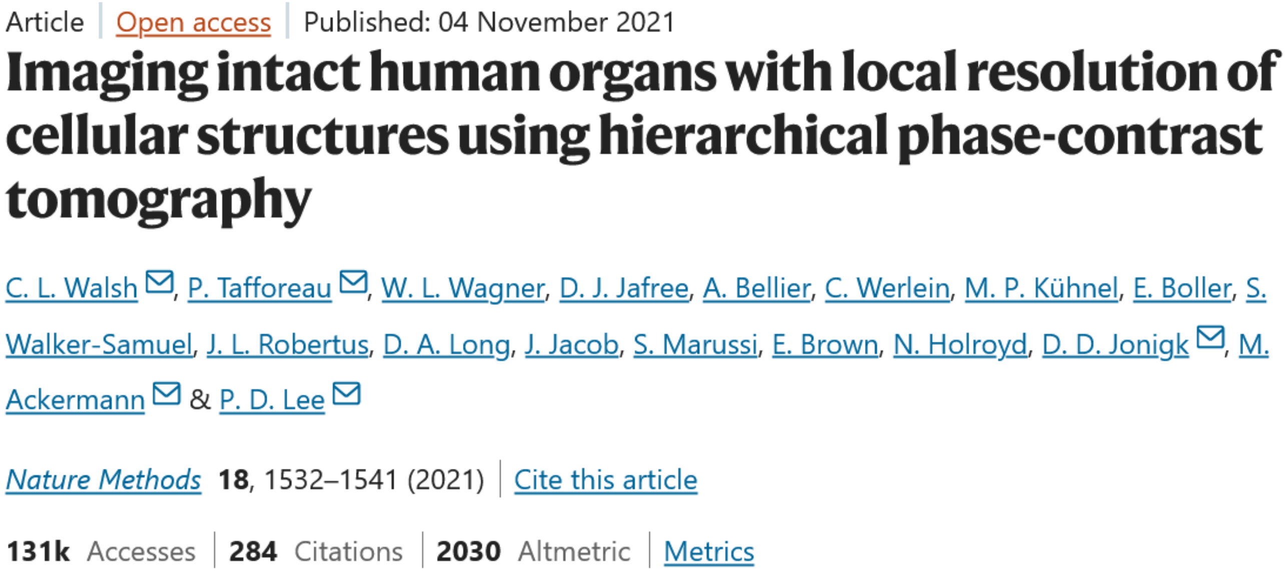
→ 19% more citations if pre-print posting

→ 14.3% more citations if data shared

→ 13.5% more citations if software shared

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

Open science is beneficial for scientists



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

131k Accesses | **284** Citations | **2030** Altmetric | Metrics

This article is in the 99th percentile (ranked 197th) of the 453,495 tracked articles of a similar age in all journals and the 98th percentile (**ranked 1st**) 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)

panosc
photon and neutron open science cloud
Interview with
Claire Walsh (UCL - ESRF)
on the Human Organ Atlas
PaNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852
ESRF
UCL

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

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'?"

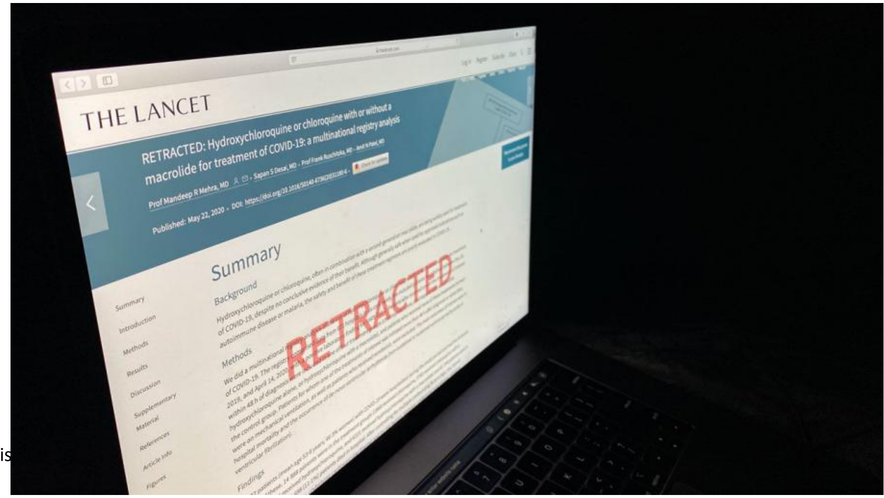
A: *Open Science refers to doing traditional science with more transparency involved at various stages, for example by openly sharing code and data. Many researchers do this already, but don't call it Open Science.*

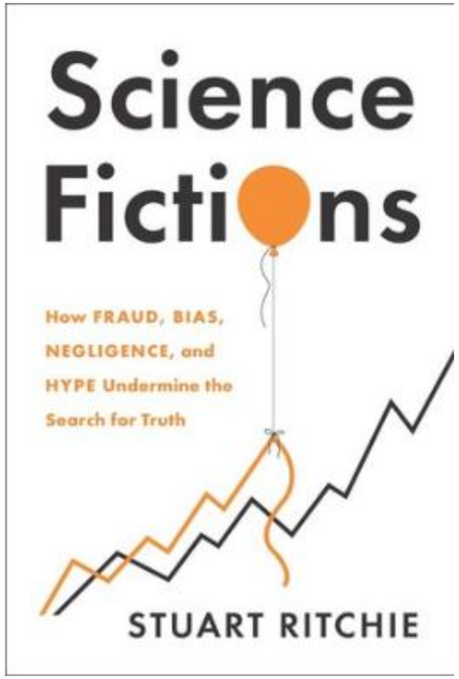
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

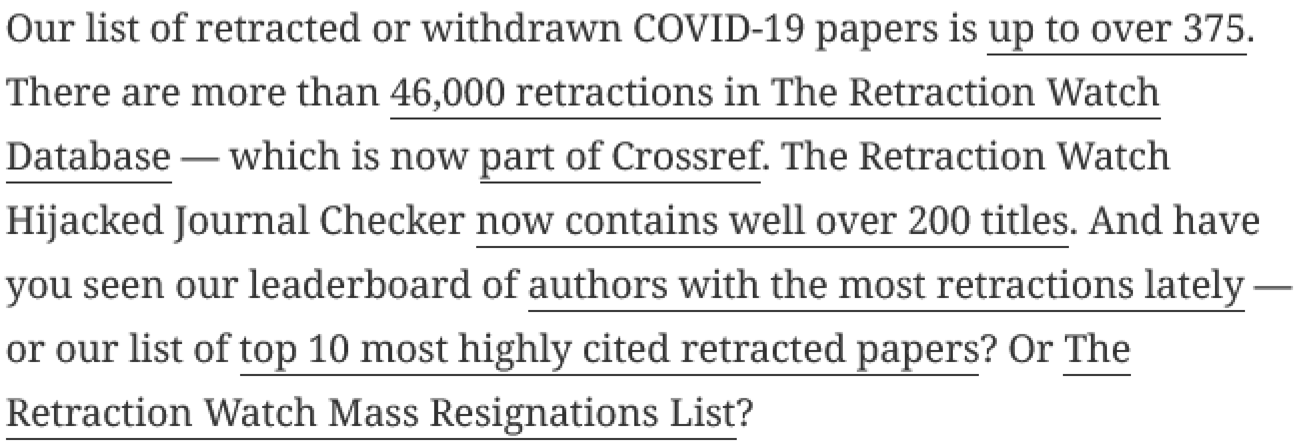
European Conduct of Scientific Integrity

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/

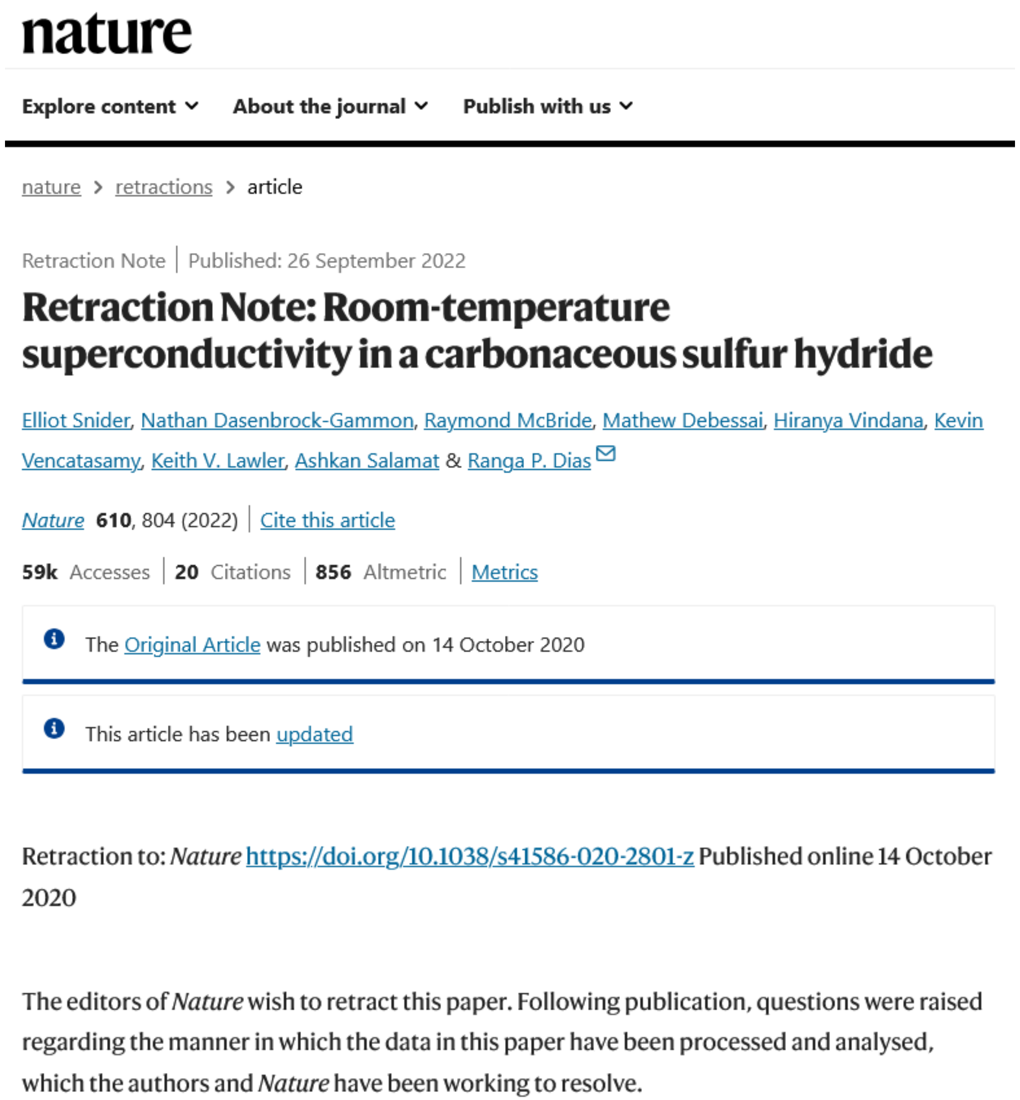

THE LANCET
RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis
Summary
RETRACTED


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



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?

Examples of data retraction



nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > retractions > article

Retraction Note | Published: 26 September 2022

Retraction Note: Room-temperature superconductivity in a carbonaceous sulfur hydride

Elliot Snider, Nathan Dasenbrock-Gammon, Raymond McBride, Mathew Debessai, Hiranya Vindana, Kevin Vencatasamy, Keith V. Lawler, Ashkan Salamat & Ranga P. Dias ✉

Nature **610**, 804 (2022) | Cite this article

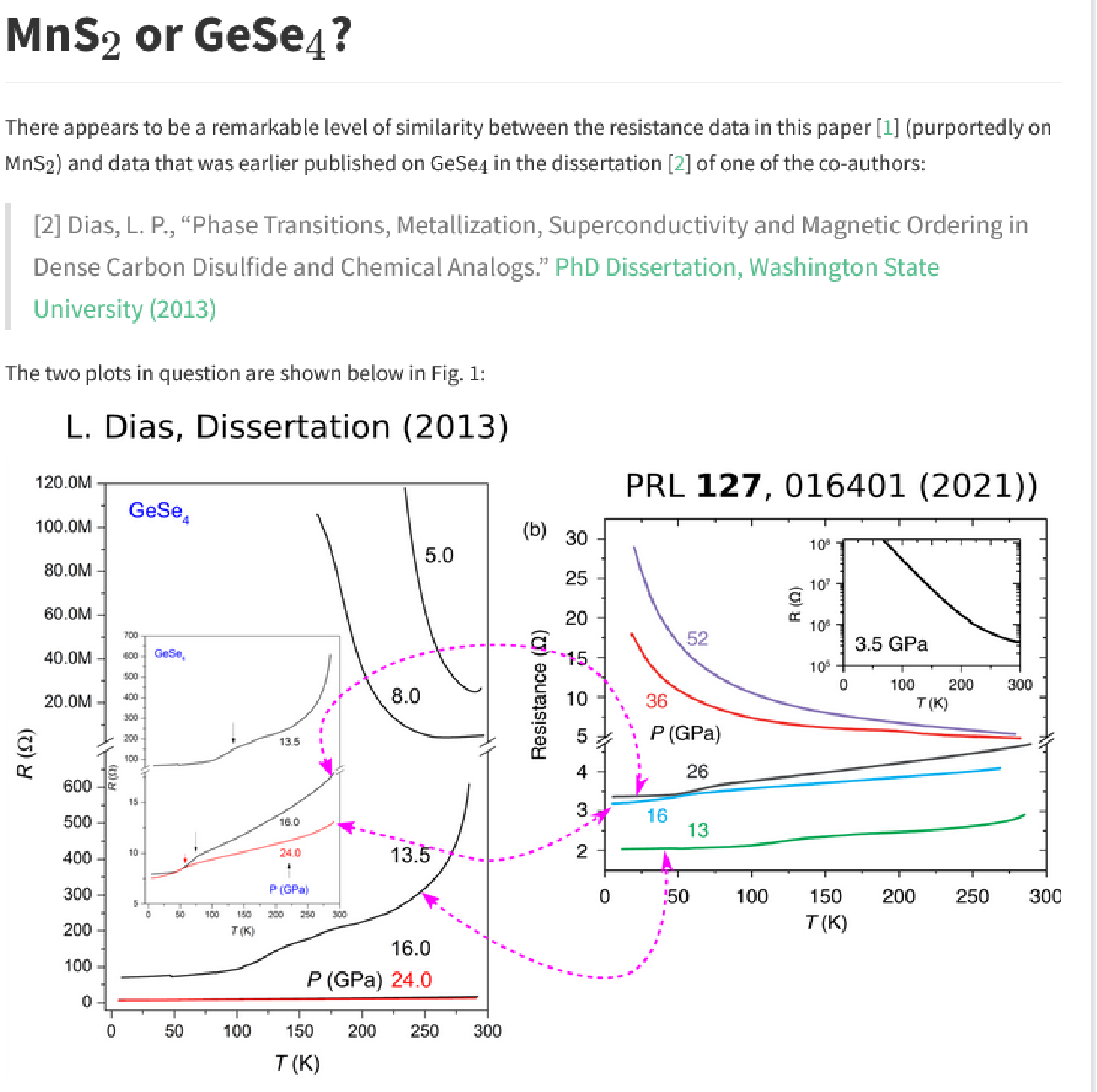
59k Accesses | **20** Citations | **856** Altmetric | Metrics

ℹ The Original Article was published on 14 October 2020

ℹ This article has been updated

Retraction to: *Nature* https://doi.org/10.1038/s41586-020-2801-z Published online 14 October 2020

The editors of *Nature* wish to retract this paper. Following publication, questions were raised regarding the manner in which the data in this paper have been processed and analysed, which the authors and *Nature* have been working to resolve.



MnS_2 or $GeSe_4$?

There appears to be a remarkable level of similarity between the resistance data in this paper [1] (purportedly on MnS_2) and data that was earlier published on $GeSe_4$ in the dissertation [2] of one of the co-authors:

> [2] Dias, L. P., "Phase Transitions, Metallization, Superconductivity and Magnetic Ordering in Dense Carbon Disulfide and Chemical Analogs." PhD Dissertation, Washington State University (2013)

The two plots in question are shown below in Fig. 1:

https://www.pubpeer.com/publications/F342DD2D2E72E5E2FD507089562B94

https://hxstem.substack.com/p/this-has-got-to-be-bullshit-personal

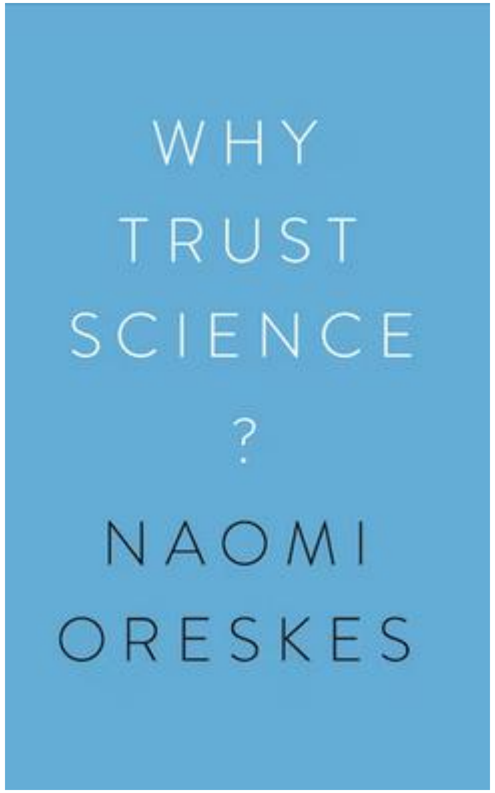
Further reading on Open Science

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
  - GO-FAIR

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

Further reading on Open Science



https://www.ouvrirlascience.fr/category/resources/

panosc

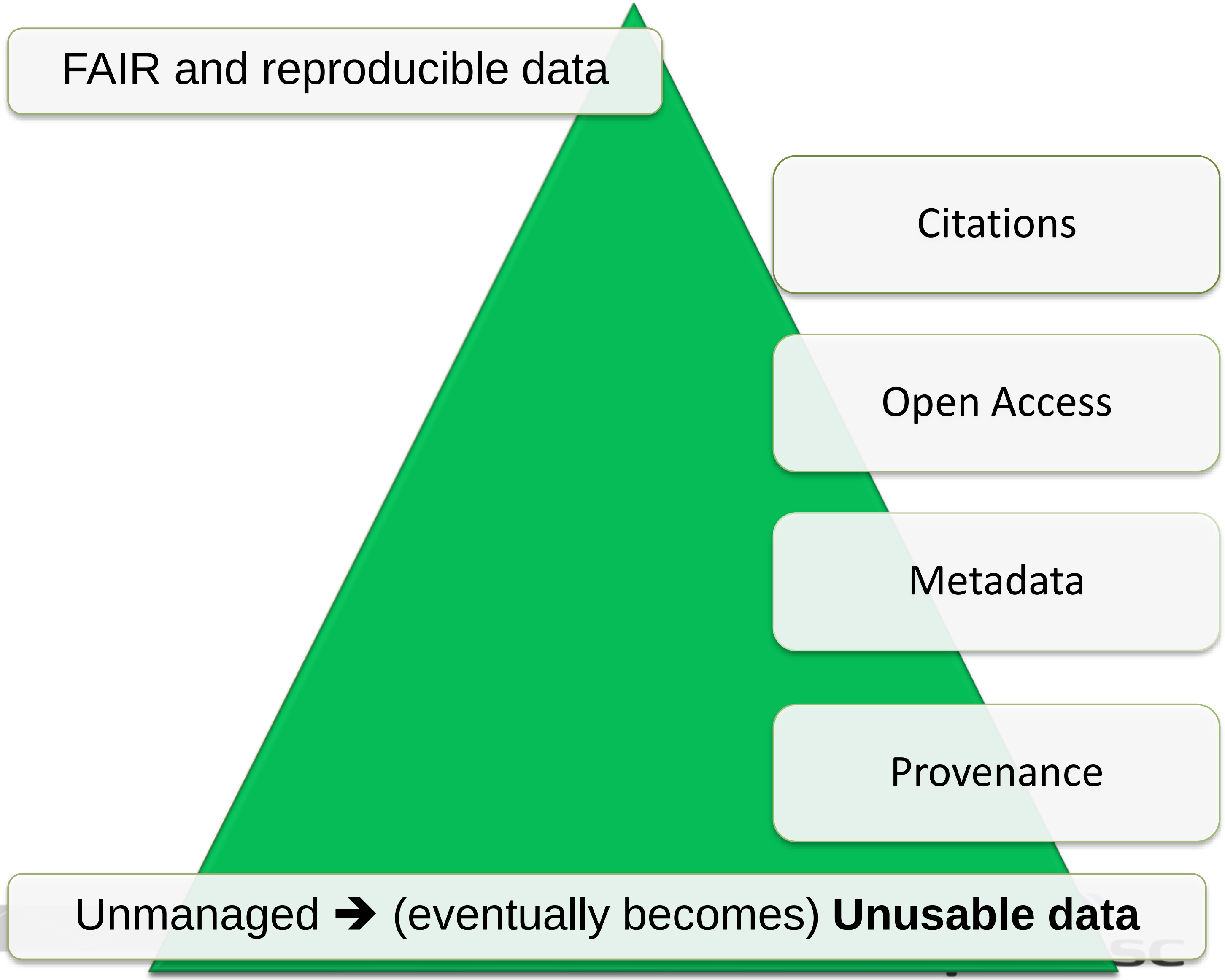
FAIR Data

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

The Twelve Labors of Hercules → 12 Labors of FAIR



FAIR and reproducible data

Citations

Open Access

Metadata

Provenance

Unmanaged ➔ (eventually becomes) **Unusable data**

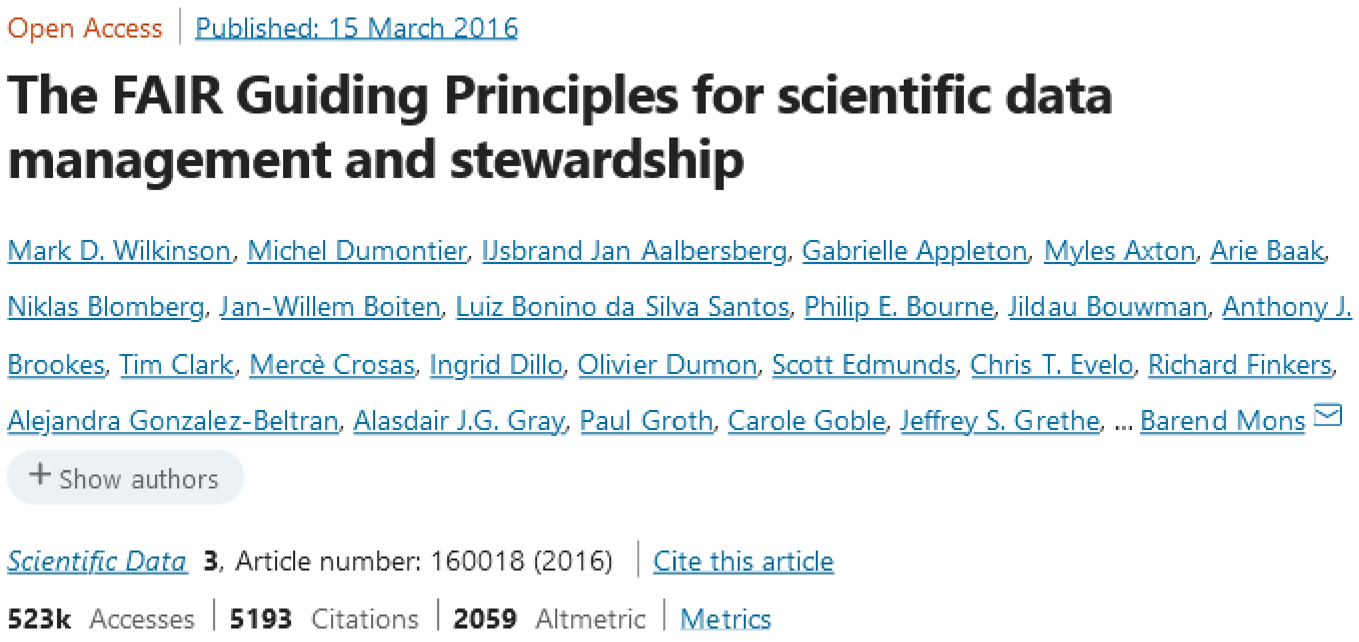
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 823852

12 Labors of FAIR Data for Scientists

1. **FAIR principles –** understand the principles
2. **Data availability –** cite the data DOIs the right way
3. **Data Policy –** understand the data policy before agreeing
4. **Data Outputs –** which your experiment will produce
5. **Metadata –** all metadata to enable data to be reused
6. **Data formats –** use open community standards
7. **E-logbooks –** write up your experiment as you go
8. **Software -** prefer open source software possible
9. **Software environment –** use open source tools if possible
10. **Data Management Plans –** generate and update
11. **Data repositories –** facility/domain one or open one
12. **Data storage –** what data should be kept

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

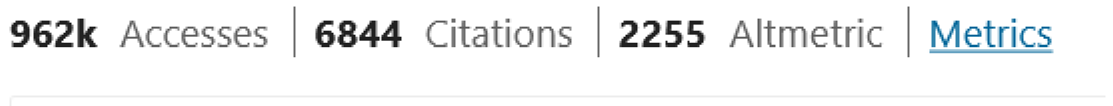
Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris 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

523k Accesses | **5193** Citations | **2059** Altmetric | Metrics

1 year



962k Accesses | **6844** Citations | **2255** Altmetric | Metrics

European Commission
Final Report and Action Plan from the European Commission Expert Group on FAIR Data
TURNING FAIR INTO REALITY
2018
Research and Innovation

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

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

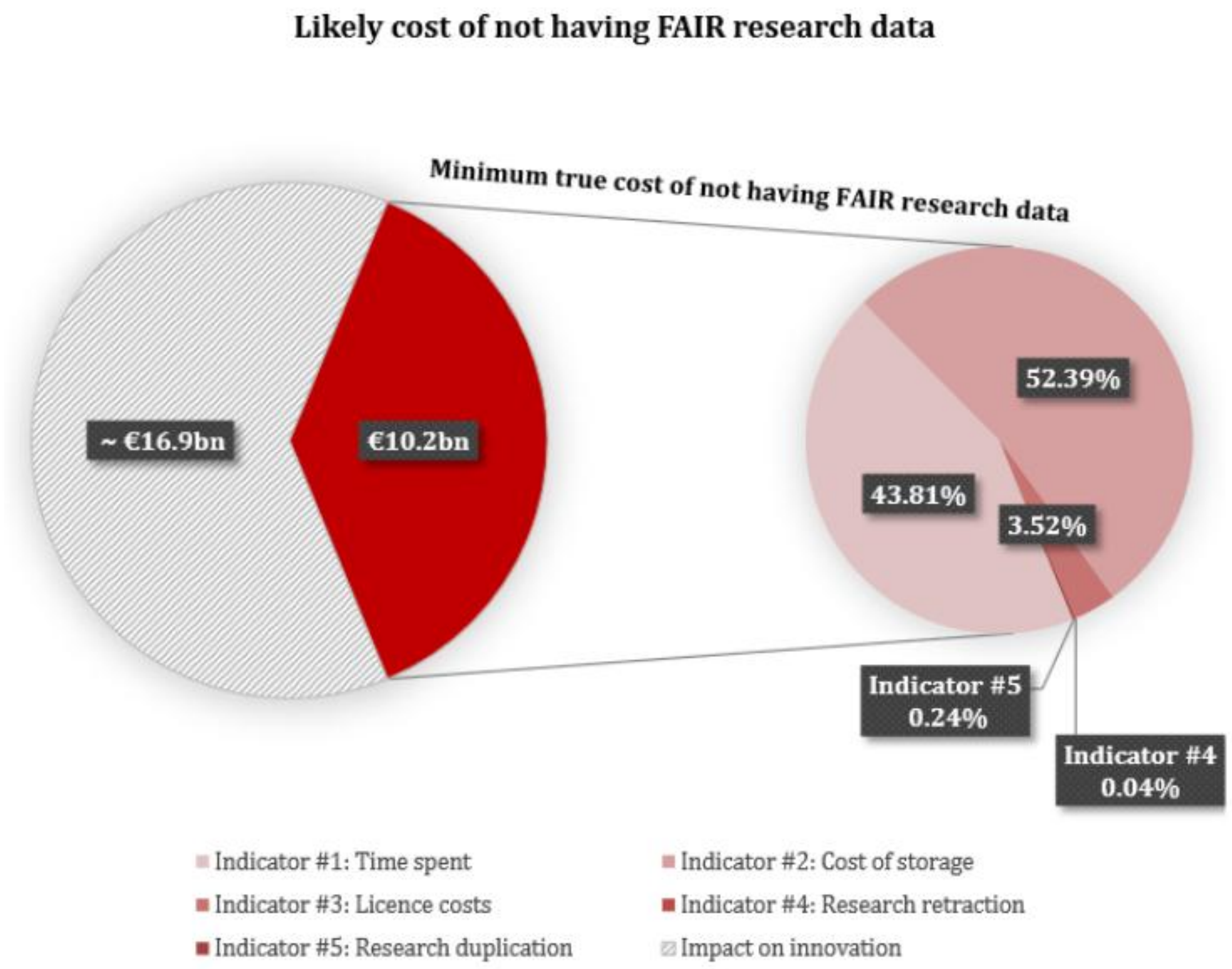
GO FAIR

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

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



Findable Accessible Interoperable Reusable

“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

OSCARS Open Science Clusters' Action for Research & Society

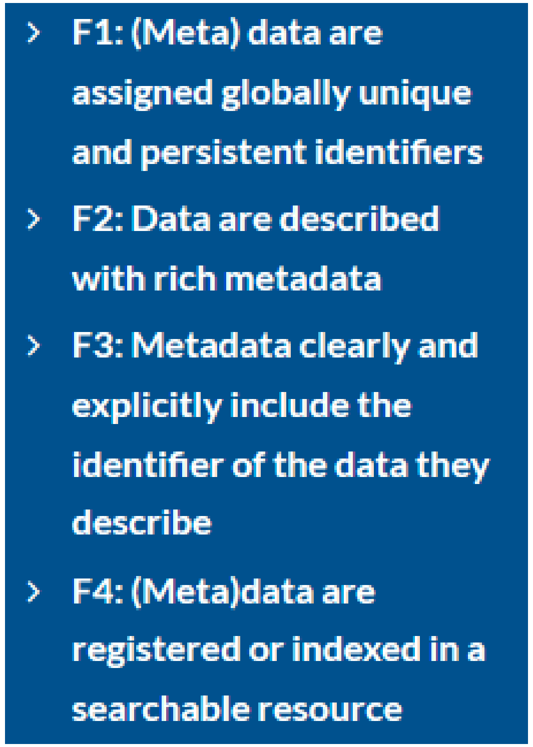
panosc

1st Labor

FAIR Principles

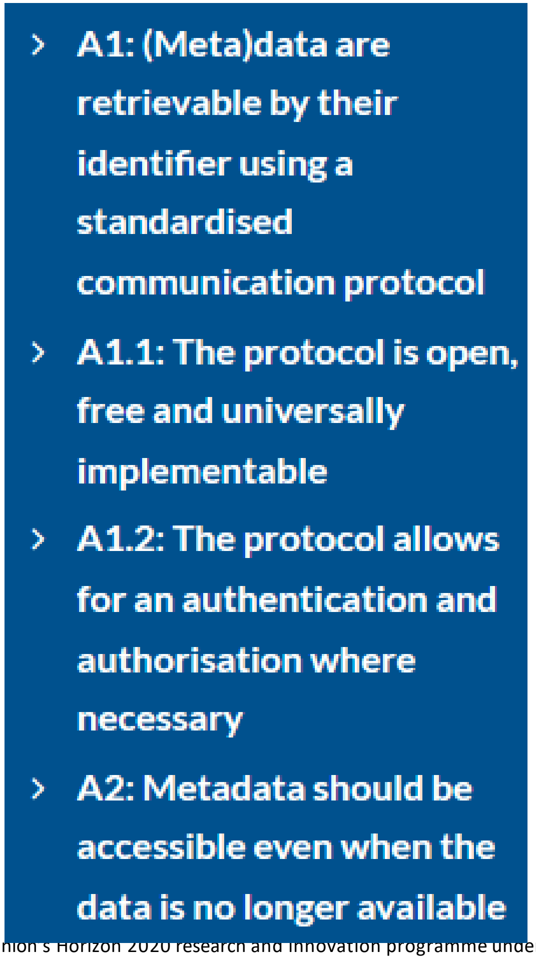
https://www.go-fair.org/fair-principles/

Findable



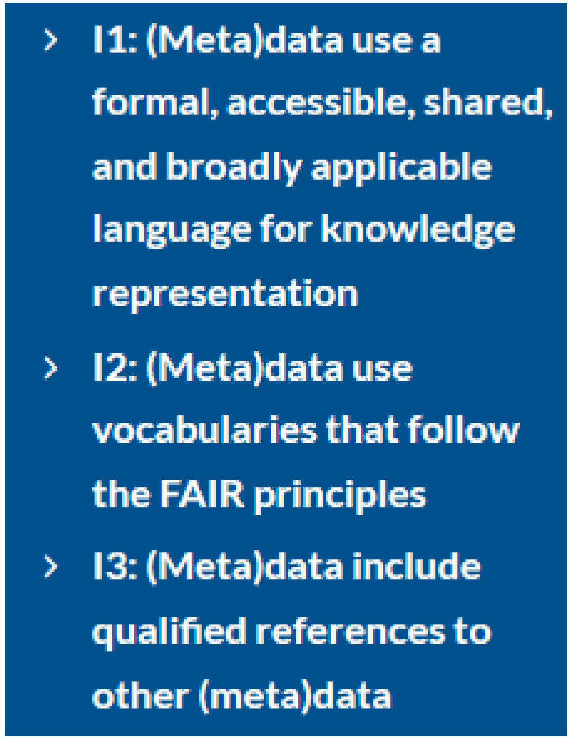
- F1: (Meta) data are assigned globally unique and persistent identifiers
- F2: Data are described with rich metadata
- F3: Metadata clearly and explicitly include the identifier of the data they describe
- F4: (Meta)data are registered or indexed in a searchable resource

Accessible



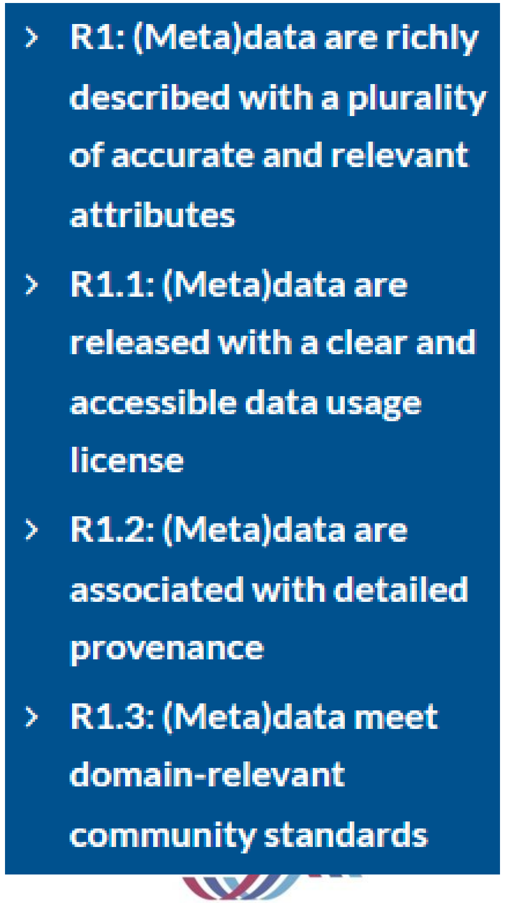
- A1: (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1: The protocol is open, free and universally implementable
- A1.2: The protocol allows for an authentication and authorisation where necessary
- A2: Metadata should be accessible even when the data 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

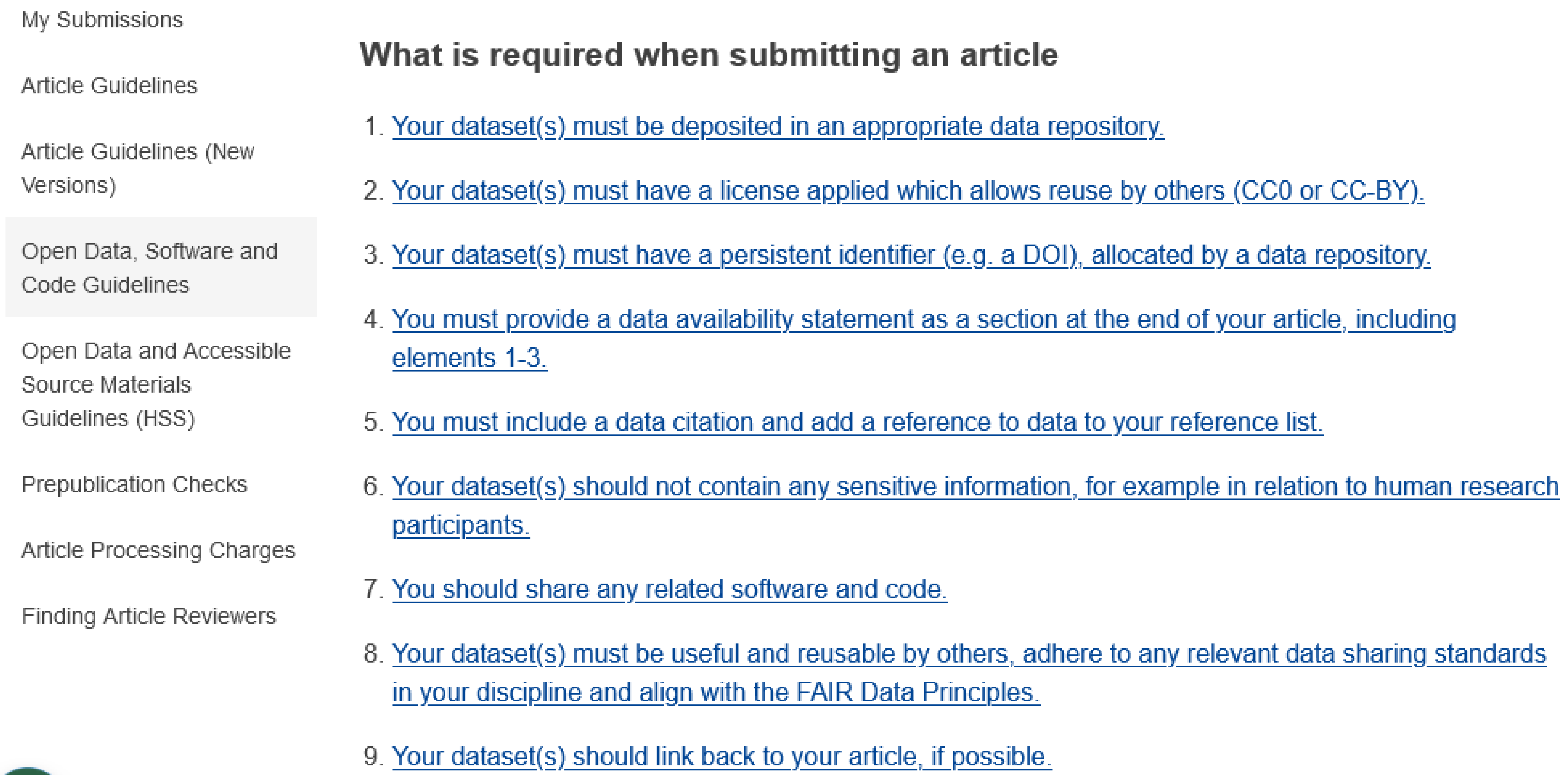
Reusable



- 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

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.

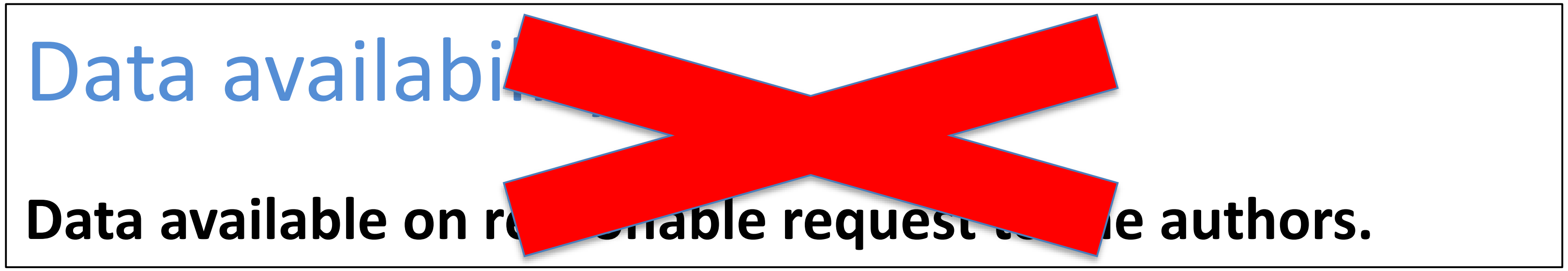
https://open-research-europe.ec.europa.eu/for-authors/data-guidelines/

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

2nd Labor

Data availability – the wrong + right way

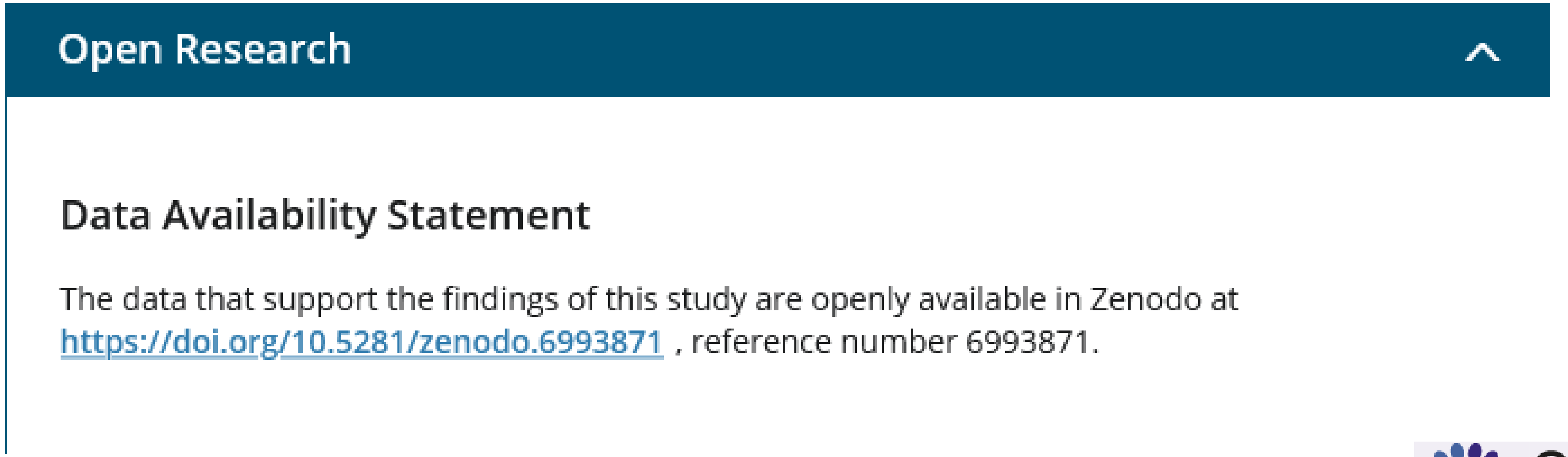
✗



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.

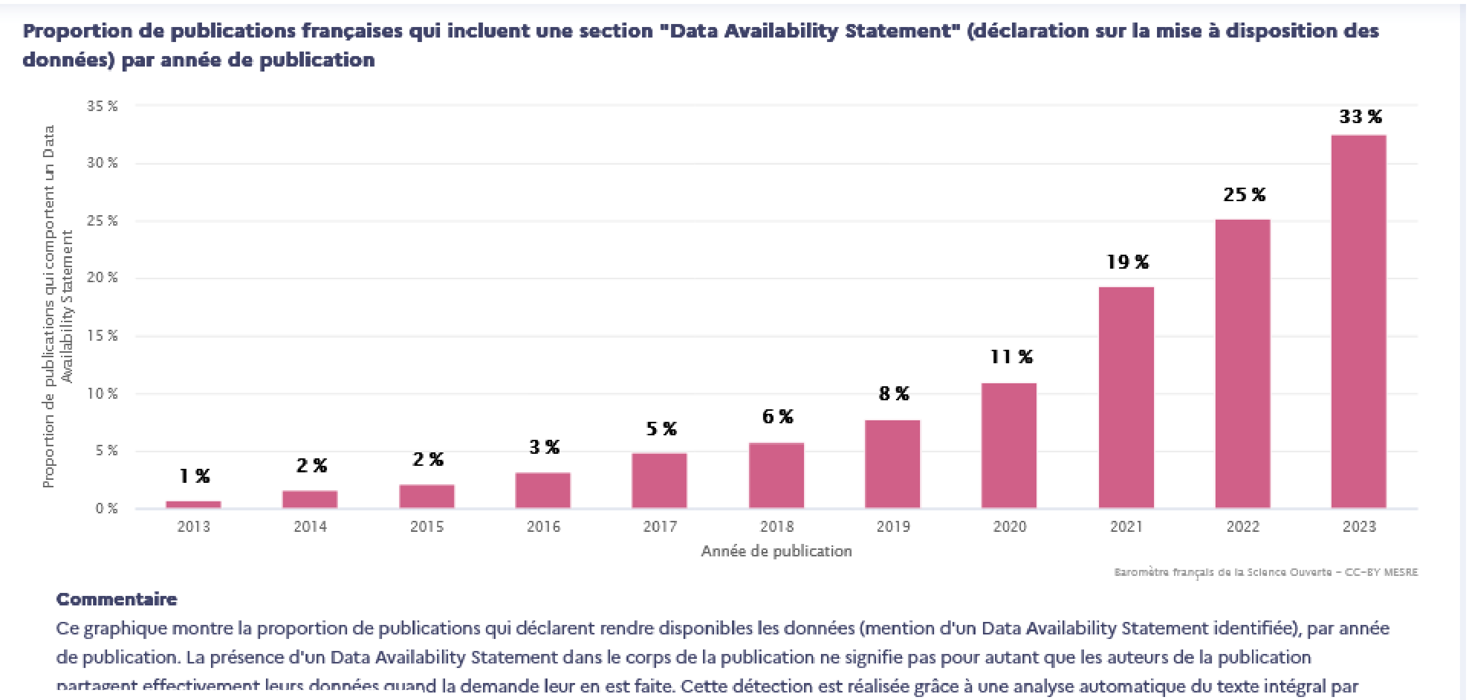
OSCARS
Open Science Clusters' Action
for Research & Society

panosc

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

Data availability statement in publications

https://barometredelascienceouverte.esr.gouv.fr/donnees-de-la-recherche/general

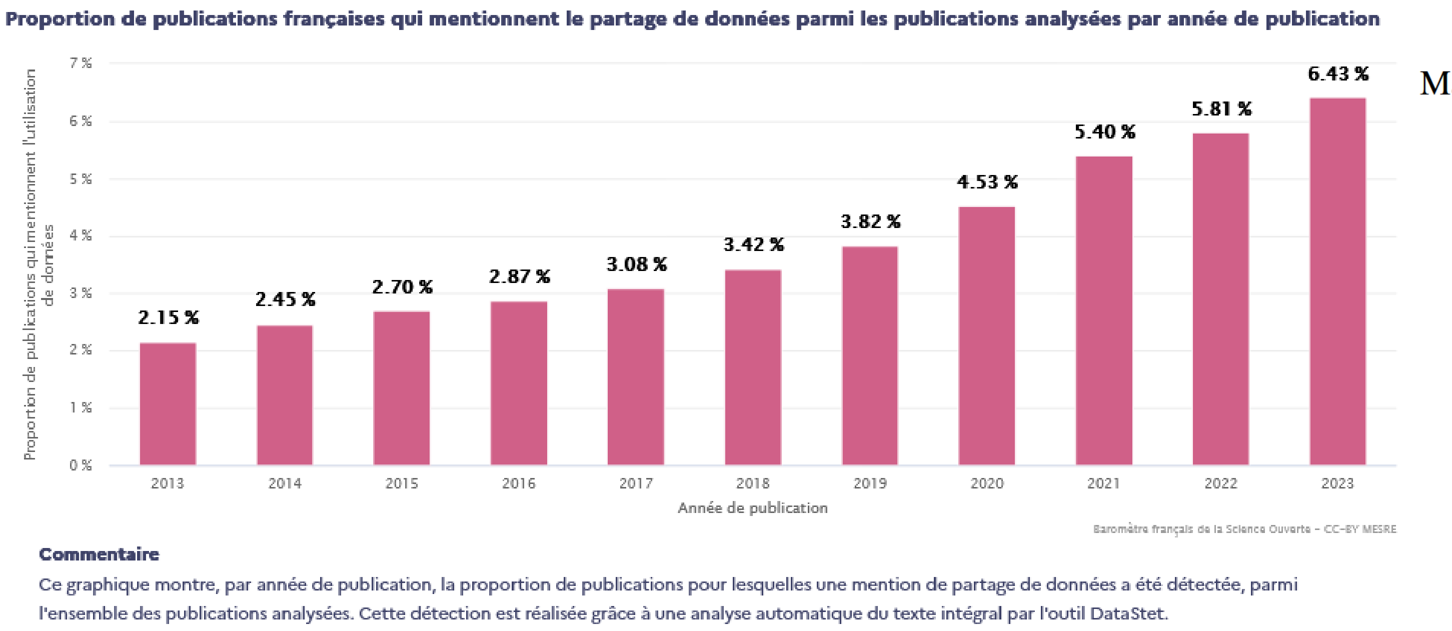


Proportion de publications françaises qui incluent une section "Data Availability Statement" (déclaration sur la mise à disposition des données) par année de publication

Baromètre français de la Science Ouverte – CC-BY MESRE

Commentaire

Ce graphique montre la proportion de publications qui déclarent rendre disponibles les données (mention d'un Data Availability Statement identifiée), par année de publication. La présence d'un Data Availability Statement dans le corps de la publication ne signifie pas pour autant que les auteurs de la publication partagent effectivement leurs données quand la demande leur en est faite. Cette détection est réalisée grâce à une analyse automatique du texte intégral par



Proportion de publications françaises qui mentionnent le partage de données parmi les publications analysées par année de publication

Baromètre français de la Science Ouverte – CC-BY MESRE

Commentaire

Ce graphique montre, par année de publication, la proportion de publications pour lesquelles une mention de partage de données a été détectée, parmi l'ensemble des publications analysées. Cette détection est réalisée grâce à une analyse automatique du texte intégral par l'outil DataStet.

Many researchers were not compliant with their published data sharing statement: a mixed-methods study

Mirko Gabelica[a], Ružica Bojčić[b], Livia Puljak[c,*]

- We analyzed 3416 articles published by BioMed Central that contained a data availability statement (DAS); the most frequent DAS category (42%) indicated that the data sets are available on reasonable request.
- **Of 1792 manuscripts in which the DAS indicated that authors are willing to share their data, only 123 (6.8%) provided the requested data.**

https://doi.org/10.1016/j.jclinepi.2022.05.019

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)

DRIVEN BY doi®

2nd Labor

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.

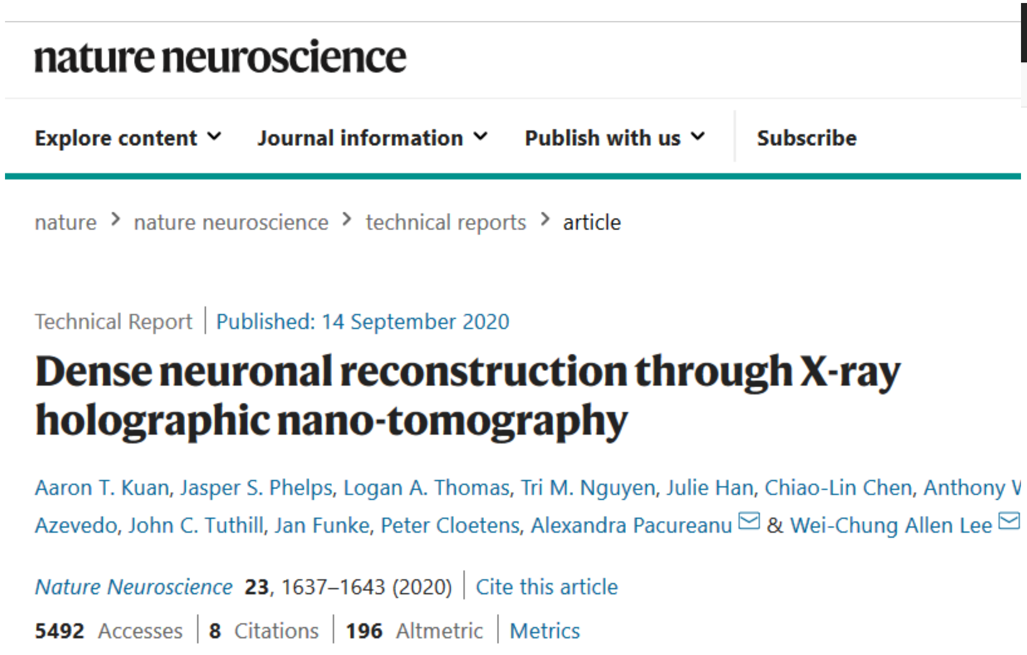
- A DOI is one implementation of a PID (Persistent Identifier)
- A web address (url) is not a PID because it is not guaranteed
- Make sure the data you want to cite has a DOI
- Cite the instrument, samples etc. you used

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

Example of article correctly citing data



nature neuroscience

Technical Report | Published: 14 September 2020

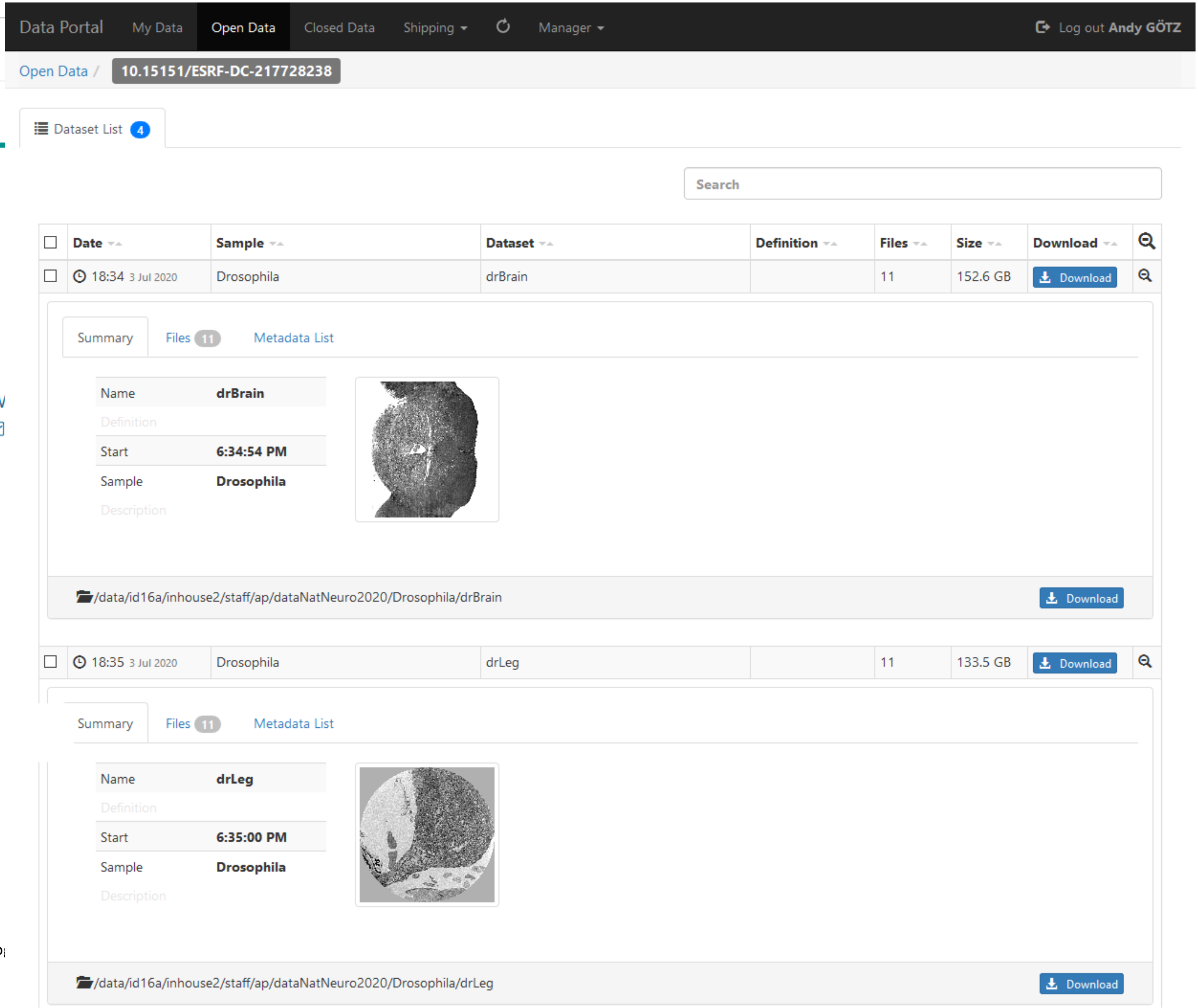
Dense neuronal reconstruction through X-ray holographic nano-tomography

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



Open Data / 10.15151/ESRF-DC-217728238

Dataset List 4

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

Summary | Files 11 | Metadata List

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

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

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

Summary | Files 11 | Metadata List

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

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

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;

Data policies

3rd Labor

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**.

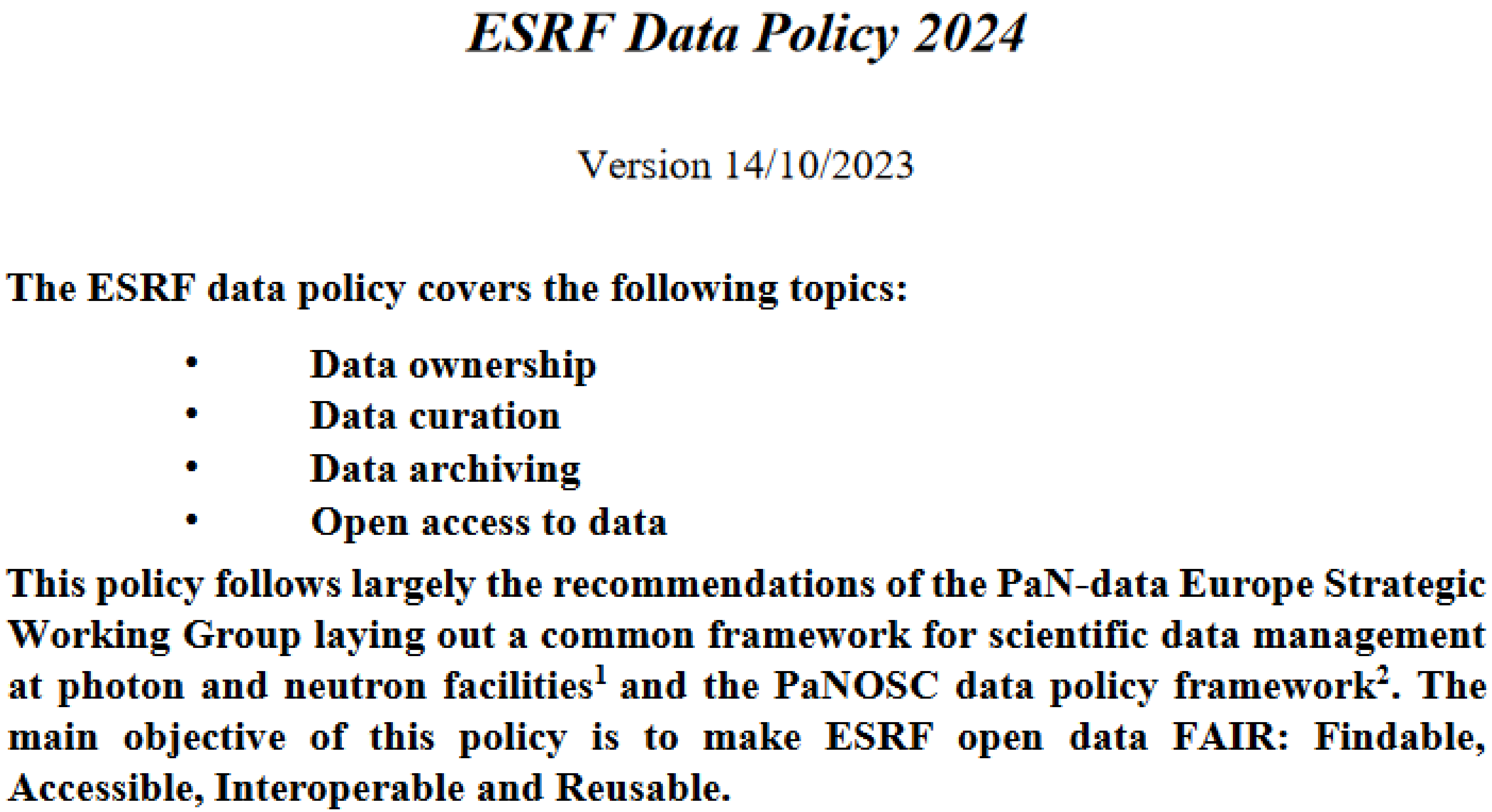
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

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, SOLEIL, ELETTRA, HZB, MAXIV**, …

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

https://www.esrf.fr/datapolicy



ESRF Data Policy 2024

Version 14/10/2023

The ESRF data policy covers the following topics:

- **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[1] and the PaNOSC data policy framework[2]. The main objective of this policy is to make ESRF open data FAIR: Findable, Accessible, Interoperable and Reusable.

Data and research outputs

4th Labor

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 needed to understand your data**

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

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

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

Metadata vocabularies

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

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

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

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

Metadata – Take away messages

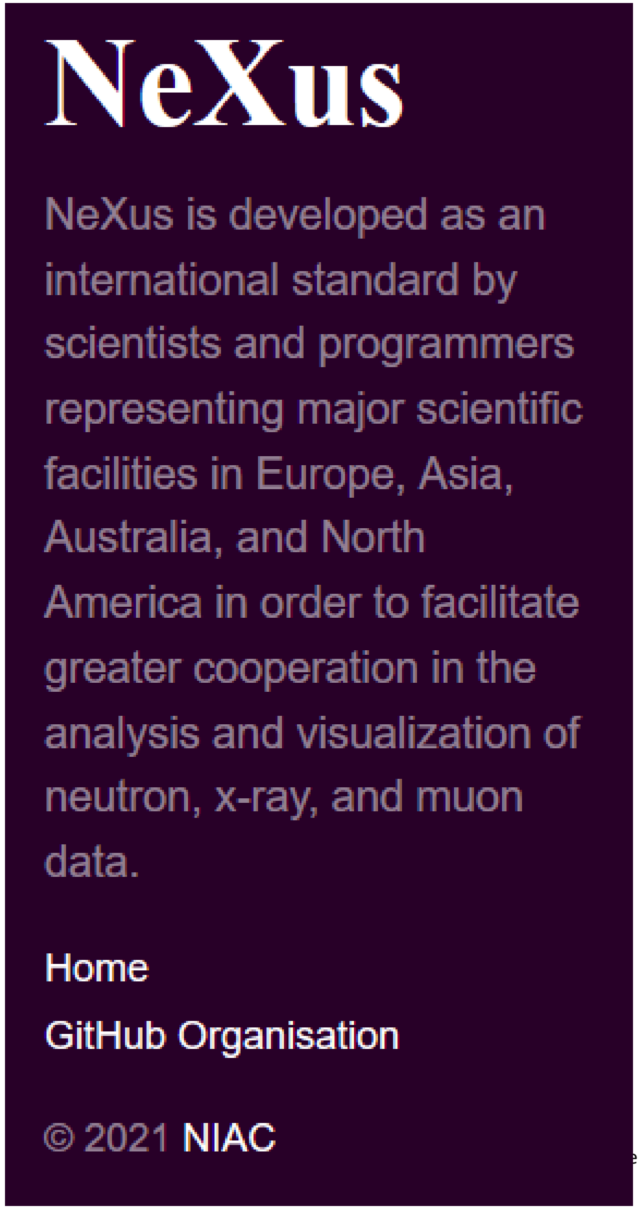
Metadata have a tendency to get treated as 2nd class data.

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

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

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

Example vocabulary – Nexus for photon and neutron sources



NeXus

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

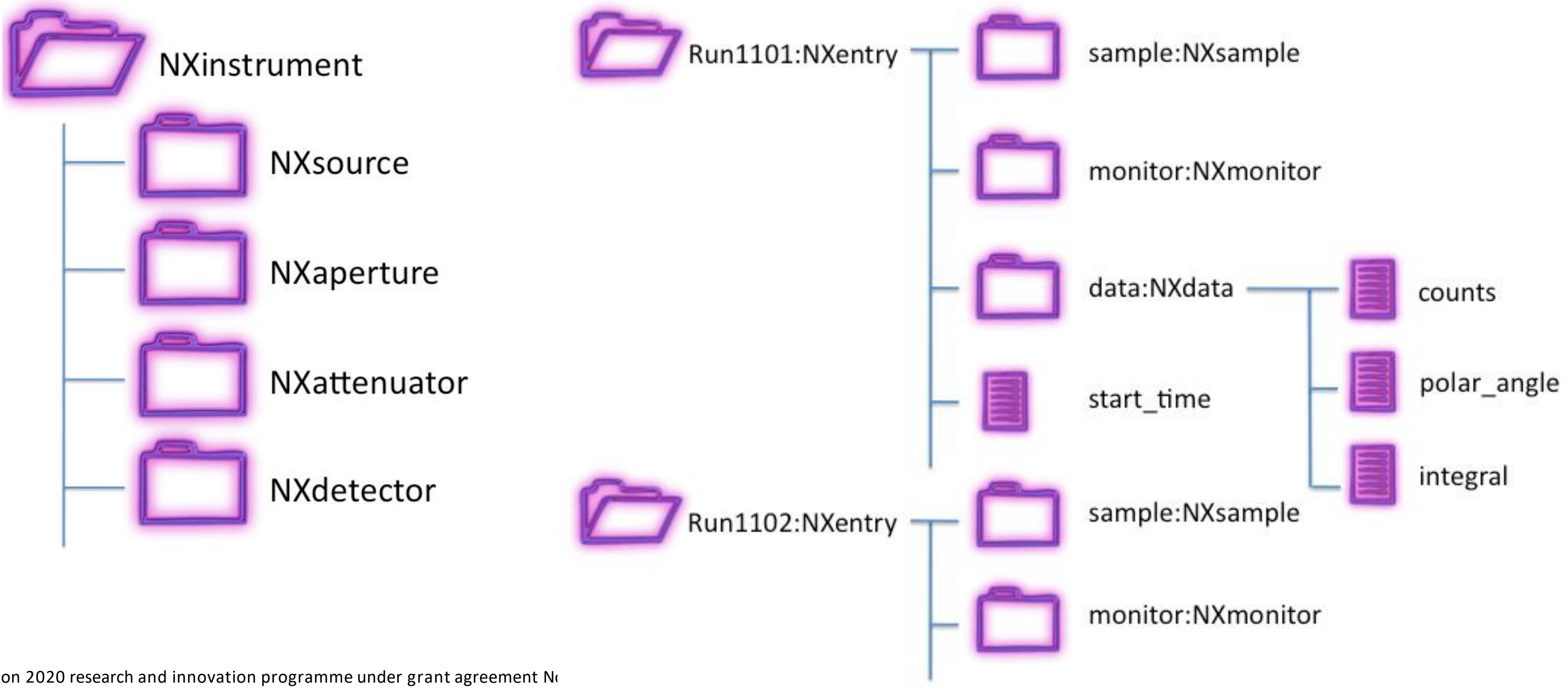
Home

GitHub Organisation

© 2021 NIAC

https://www.nexusformat.org/

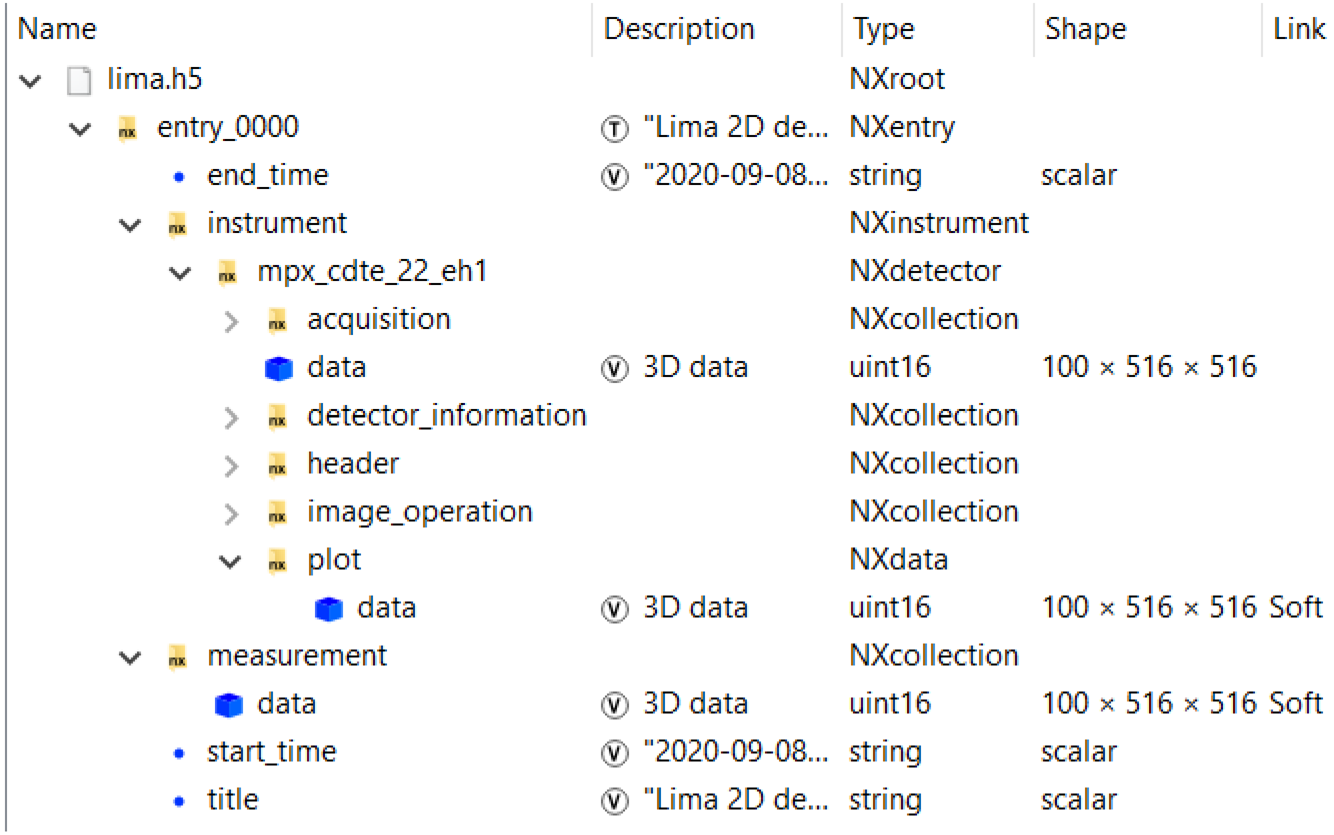
Nexus provides a standard vocabulary for:



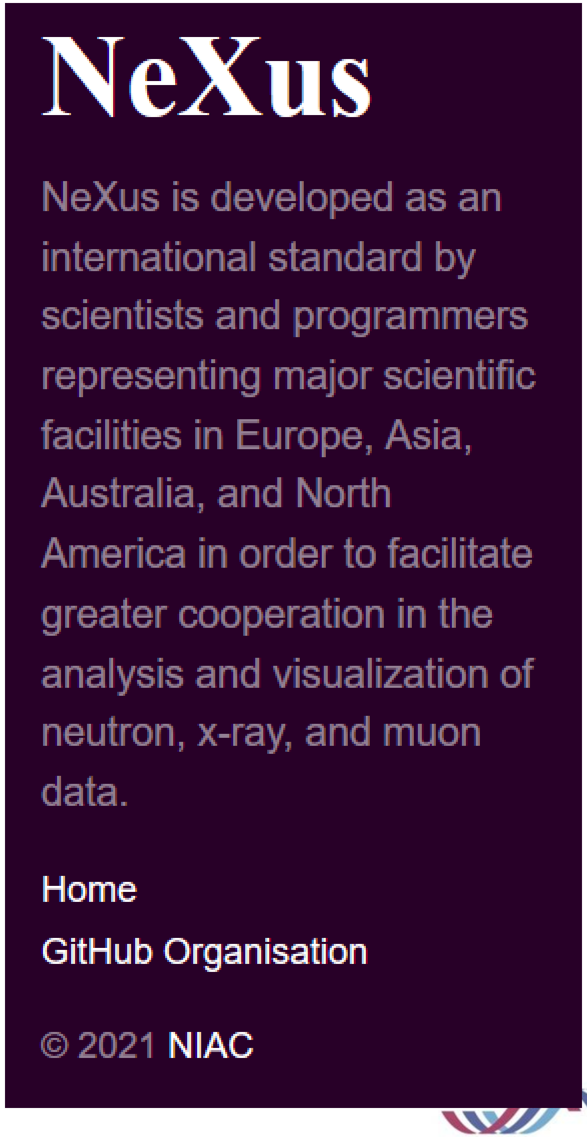
ean Union’s Horizon 2020 research and innovation programme under grant agreement N

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	



NeXus

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

Home

GitHub Organisation

© 2021 NIAC

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

Data formats

6th Labor

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

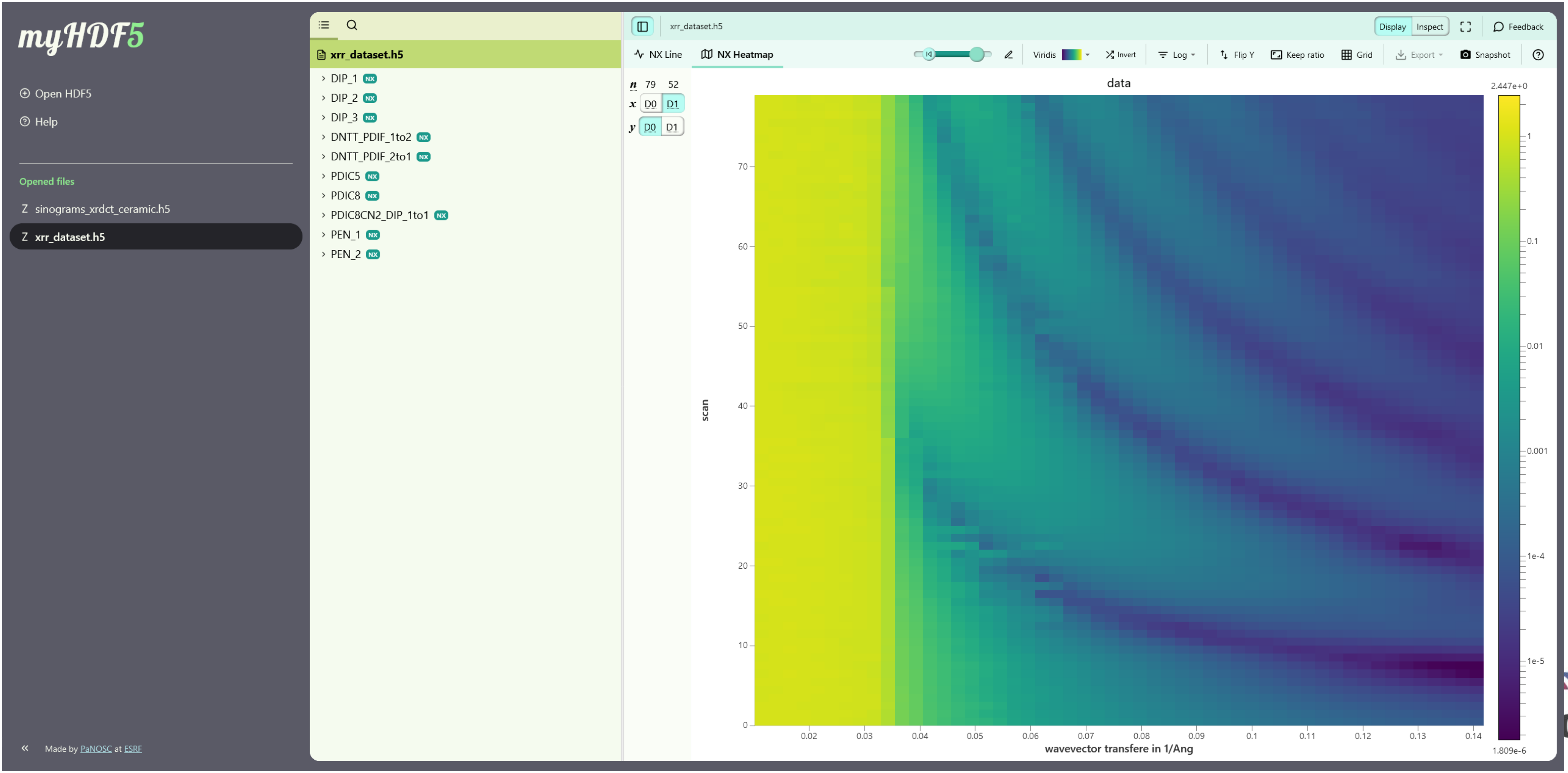
- Example data formats:
  - CSV (Comma Separated Values)
  - TIFF for images
  - HDF5 as container
- **USE** the **STANDARD**(s) for your **community**

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

Nexus/HDF5 - Plotting data + metadata online in the browser

- →https://myhdf5.hdfgroup.org/view?url=https%3A%2F%2Fzenodo.org%2Frecord%2F6497438%2Ffiles%2Fxrr_dataset.h5%3Fdownload%3D1



Thi

E-logbooks

7th Labor

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.

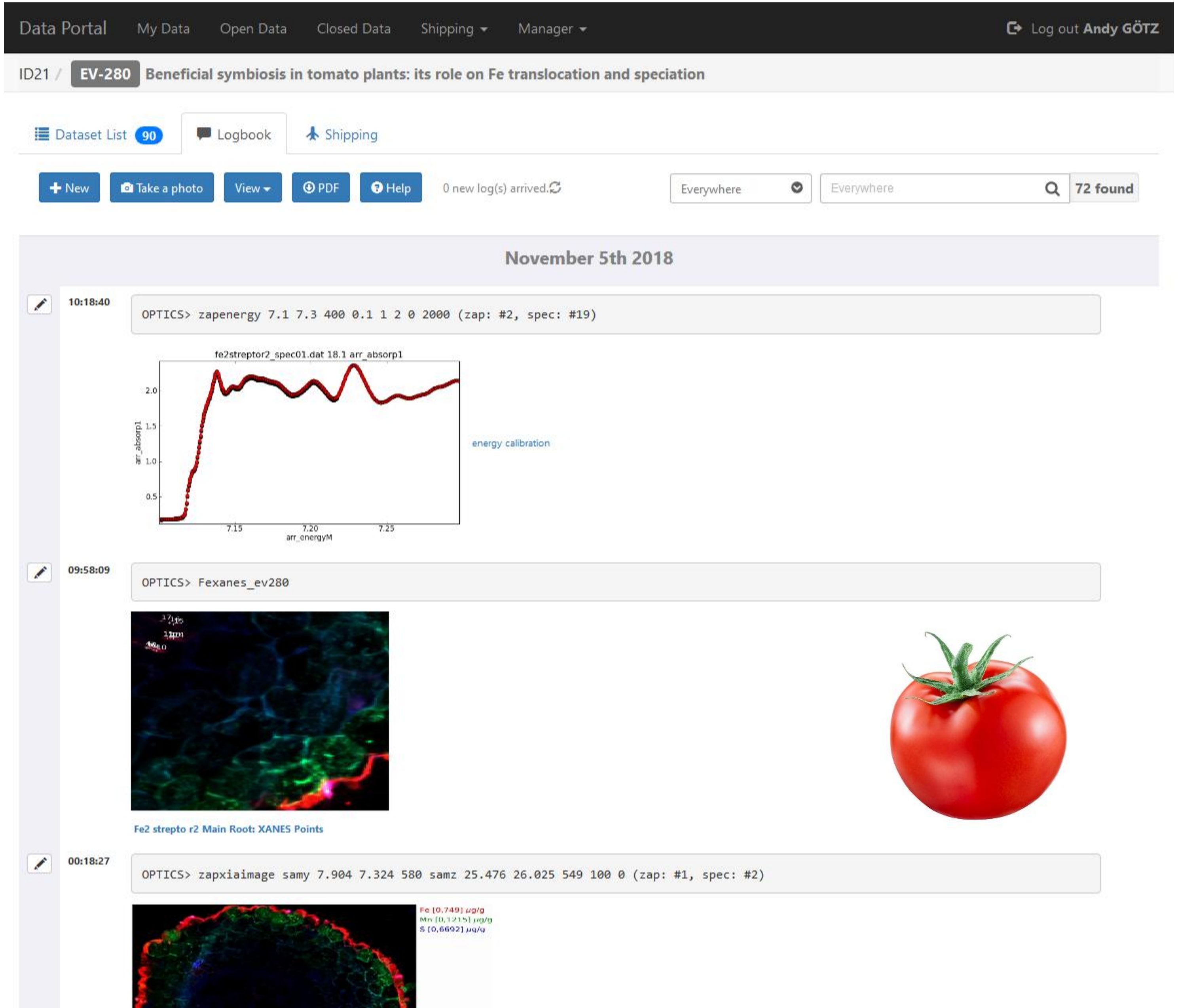
- **E-logbook advantages**
  - Shared editing online
  - Powerful search facilities
  - Access rules during embargo period
  - Allows others to understand what you did during the experiment
- **E-logbook is metadata** and will be part of the open data

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

panosc

ESRF e-logbook example – ID21 / EV-280



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

0 new log(s) arrived.

November 5th 2018

10:18:40

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

fe2streptor2_spec01.dat 18.1 arr_absorp1

energy calibration

09:58:09

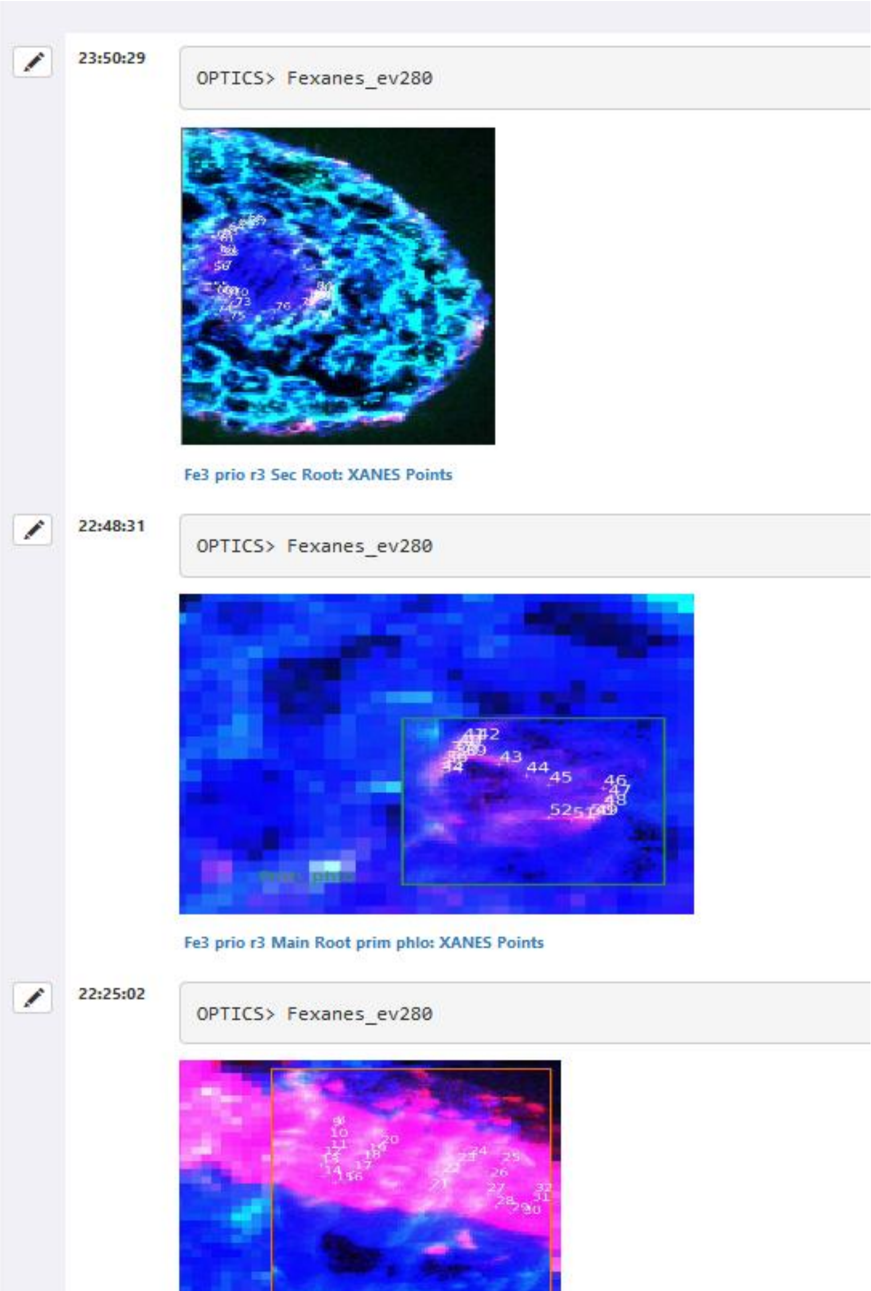
```
OPTICS> Fexanes_ev280
```

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)
```

Fe [0.749] µg/g
Mn [0,1215] µg/g
S [0,6692] µg/g



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
```

8th Labor

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

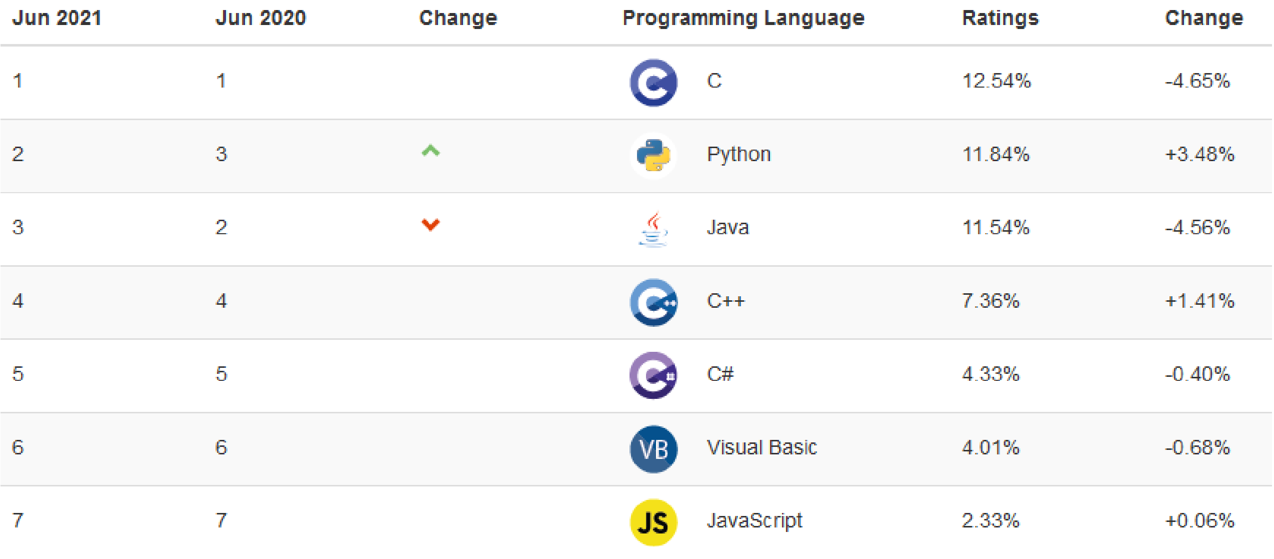
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

panosc

Software environment + tools

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

- **Python** has become the de facto programming language in science
- **Jupyter** notebooks enable reproducible publications https://jupyter.org
- **Binder** service can preserve and run the software for an analysis - https://mybinder.org/



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

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

Data Management Plans (DMP)

10th Labor

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.

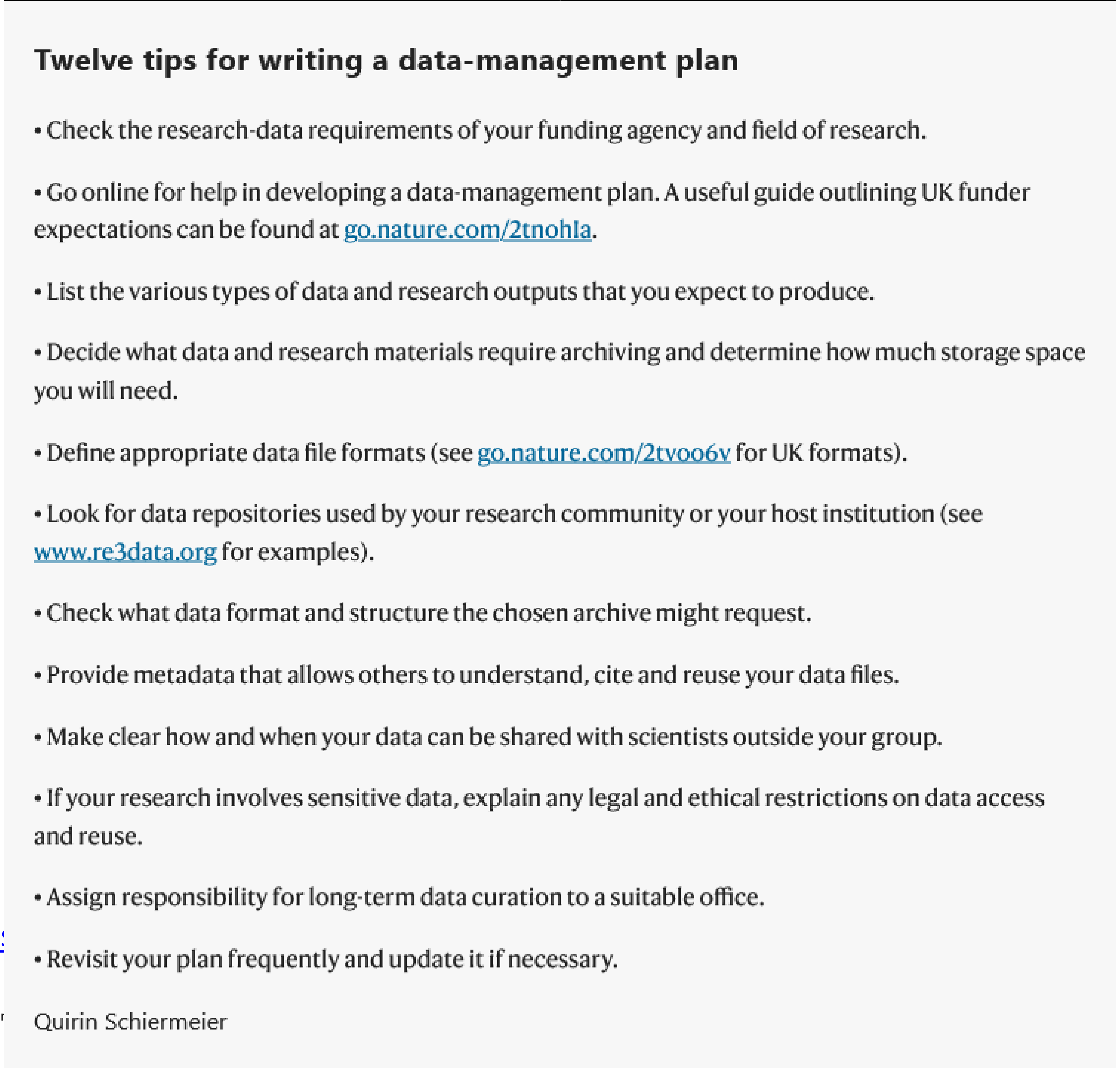
- DMP document the data management steps in a more formal manner
- Funders are requiring DMPs to ensure RDM is planned
- Facilities will require DMPs more and more to be sure Users can deal with the research data
- DMPs are living documents which need to be updated throughout the project
- Examples of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

Twelve tips for writing a data-management plan

Data Management Made Simple, Quirin Shiermeier, *Nature* **555**, 403-405 (2018), *doi: https://doi.org/10.1038/d41586-018-03071-1*

http



Twelve tips for writing a data-management plan

- Check the research-data requirements of your funding agency and field of research.
- 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.
- List the various types of data and research outputs that you expect to produce.
- Decide what data and research materials require archiving and determine how much storage space you will need.
- Define appropriate data file formats (see go.nature.com/2tvoo6v for UK formats).
- Look for data repositories used by your research community or your host institution (see www.re3data.org for examples).
- Check what data format and structure the chosen archive might request.
- Provide metadata that allows others to understand, cite and reuse your data files.
- Make clear how and when your data can be shared with scientists outside your group.
- If your research involves sensitive data, explain any legal and ethical restrictions on data access and reuse.
- Assign responsibility for long-term data curation to a suitable office.
- Revisit your plan frequently and update it if necessary.

Quirin Schiermeier

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

11th Labor

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 , https://human-organ-atlas , https://paleo.esrf.fr , …
- Choose repository which is certified e.g. http://go.nature.com/2eLHBFP
- Use an institute or community archive which is sustainable e.g. PDB, COD, …, https://www.re3data.org/

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?

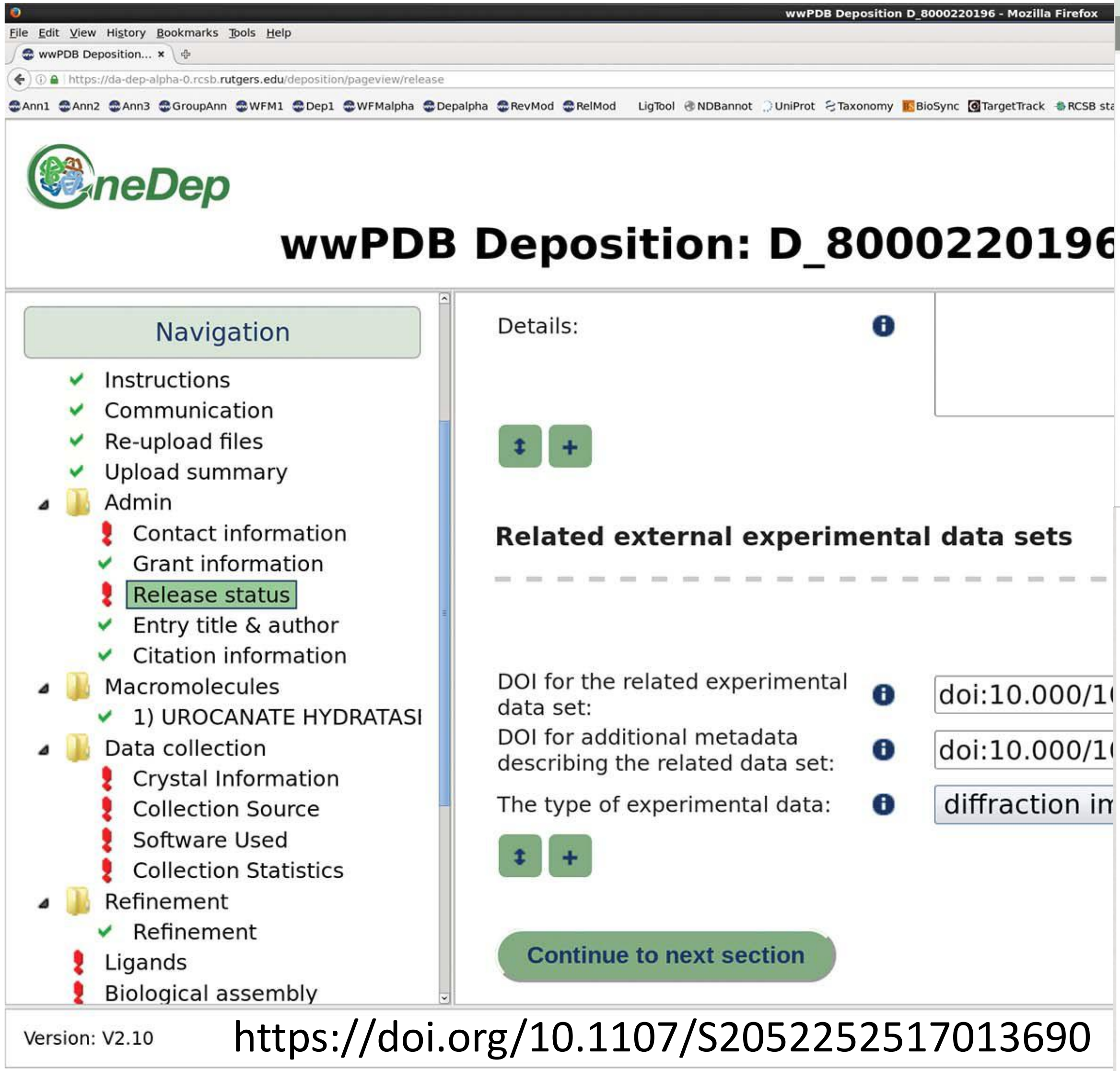
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 823852

Non-facility repositories

			# datasets
NIH National Institutes of Health Office of the Director Data Science at NIH Integrated Res Crystallograph	**Macromolecular**	https://proteindiffraction.org	**9648**
SBGrid Data Bank	Macromolecular	https://data.sbgrid.org/	811
MX-RDR Macromolecular Xtallography Raw Data Repository	Macromolecular	https://mxrdr.icm.edu.pl/	410
zenodo	General Macromolecular Crystallography community	https://zenodo.org/	238
CXIDB Coherent X-ray Imaging Data Bank	Coherent imaging/XFEL	https://cxidb.org/	218
XRD-Arc OneDep IUCr	Macromolecular linked to PDBj	https://xrda.pdbjbk1.pdbj.org	100

https://zenodo.org/doi/10.5281/zenodo.10014981

Linking raw data to the PDB


wwPDB Deposition D_8000220196 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
wwPDB Deposition... ×
https://da-dep-alpha-0.rcsb.rutgers.edu/deposition/pageview/release
Ann1 Ann2 Ann3 GroupAnn WFM1 Dep1 WFMalpha Depalpha RevMod RelMod LigTool NDBannot UniProt Taxonomy BioSync TargetTrack RCSB sta
OneDep

wwPDB Deposition: D_8000220196

Navigation

- ✓ Instructions
- ✓ Communication
- ✓ Re-upload files
- ✓ Upload summary
- Admin
  - ! Contact information
  - ✓ Grant information
  - ! Release status
  - ✓ Entry title & author
  - ✓ Citation information
- Macromolecules
  - ✓ 1) UROCANATE HYDRATASI
- Data collection
  - ! Crystal Information
  - ! Collection Source
  - ! Software Used
  - ! Collection Statistics
- Refinement
  - ✓ Refinement
- ! Ligands
- ! Biological assembly

Version: V2.10

Details:

Related external experimental data sets

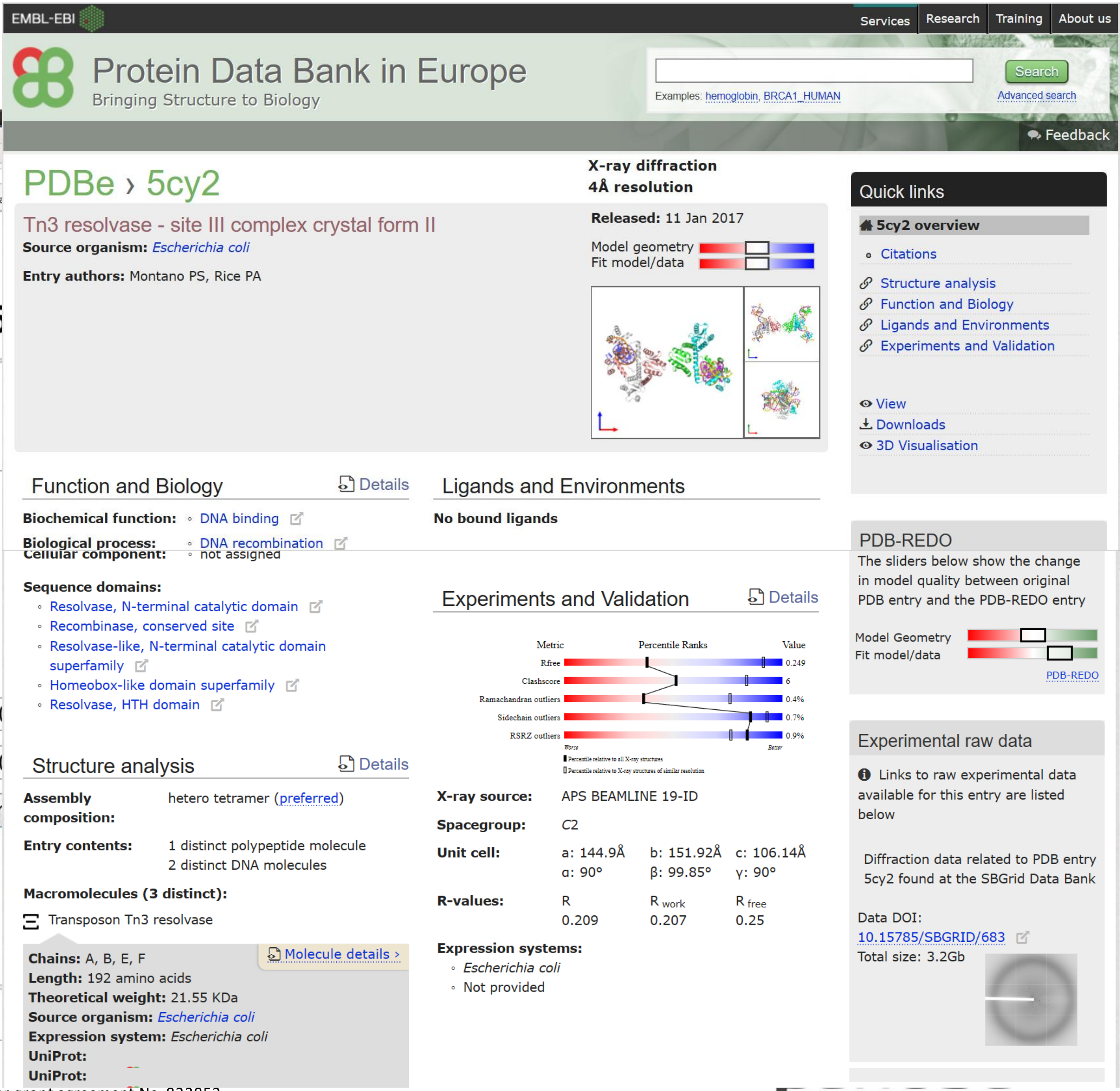
DOI for the related experimental data set: doi:10.000/1(

DOI for additional metadata describing the related data set: doi:10.000/1(

The type of experimental data: diffraction im

Continue to next section

https://doi.org/10.1107/S2052252517013690



Protein Data Bank in Europe

Bringing Structure to Biology

Examples: hemoglobin, BRCA1_HUMAN

PDBe › 5cy2

X-ray diffraction
4Å resolution

Tn3 resolvase - site III complex crystal form II

Source organism: *Escherichia coli*

Entry authors: Montano PS, Rice PA

Released: 11 Jan 2017

Model geometry
Fit model/data

Quick links

- 5cy2 overview
- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation
- View
- Downloads
- 3D Visualisation

Function and Biology

Details

Biochemical function: DNA binding

Biological process: DNA recombination

Cellular component: not assigned

Sequence domains:

- Resolvase, N-terminal catalytic domain
- Recombinase, conserved site
- Resolvase-like, N-terminal catalytic domain superfamily
- Homeobox-like domain superfamily
- Resolvase, HTH domain

Ligands and Environments

No bound ligands

Experiments and Validation

Details

Metric	Percentile Ranks	Value
Rfree		0.249
Clashscore		6
Ramachandran outliers		0.4%
Sidechain outliers		0.7%
RSRZ outliers		0.9%

▮ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

X-ray source: APS BEAMLINE 19-ID

Spacegroup: *C*2

Unit cell:	a: 144.9Å α: 90°	b: 151.92Å β: 99.85°	c: 106.14Å γ: 90°
R-values:	R 0.209	R work 0.207	R free 0.25

Expression systems:

- *Escherichia coli*
- Not provided

Structure analysis

Details

Assembly composition: hetero tetramer (preferred)

Entry contents: 1 distinct polypeptide molecule
2 distinct DNA molecules

Macromolecules (3 distinct):

Transposon Tn3 resolvase

Molecule details ›

Chains: A, B, E, F
Length: 192 amino acids
Theoretical weight: 21.55 KDa
Source organism: *Escherichia coli*
Expression system: *Escherichia coli*
UniProt:
UniProt:

PDB-REDO

The sliders below show the change in model quality between original PDB entry and the PDB-REDO entry

Model Geometry
Fit model/data

PDB-REDO

Experimental raw data

Links to raw experimental data available for this entry are listed below

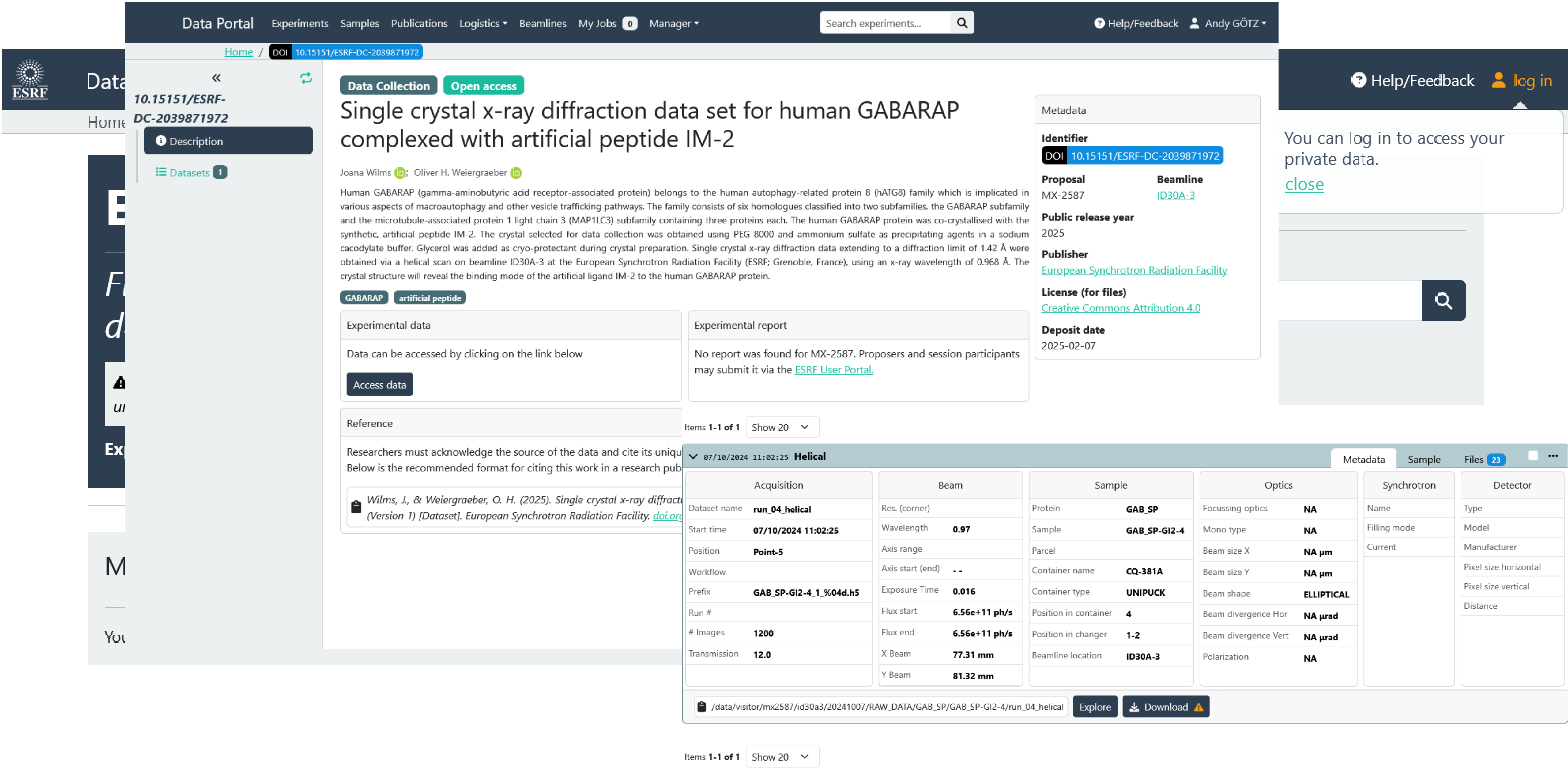
Diffraction data related to PDB entry 5cy2 found at the SBGrid Data Bank

Data DOI:
10.15785/SBGRID/683

Total size: 3.2Gb

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



You can log in to access your private data.
close

Data Collection Open access

Single crystal x-ray diffraction data set for human GABARAP complexed with artificial peptide IM-2

Joana Wilms; Oliver H. Weiergraeber

Human GABARAP (gamma-aminobutyric acid receptor-associated protein) belongs to the human autophagy-related protein 8 (hATG8) family which is implicated in various aspects of macroautophagy and other vesicle trafficking pathways. The family consists of six homologues classified into two subfamilies, the GABARAP subfamily and the microtubule-associated protein 1 light chain 3 (MAP1LC3) subfamily containing three proteins each. The human GABARAP protein was co-crystallised with the synthetic, artificial peptide IM-2. The crystal selected for data collection was obtained using PEG 8000 and ammonium sulfate as precipitating agents in a sodium cacodylate buffer. Glycerol was added as cryo-protectant during crystal preparation. Single crystal x-ray diffraction data extending to a diffraction limit of 1.42 Å were obtained via a helical scan on beamline ID30A-3 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France), using an x-ray wavelength of 0.968 Å. The crystal structure will reveal the binding mode of the artificial ligand IM-2 to the human GABARAP protein.

GABARAP artificial peptide

Experimental data

Data can be accessed by clicking on the link below

Access data

Experimental report

No report was found for MX-2587. Proposers and session participants may submit it via the ESRF User Portal.

Reference

Researchers must acknowledge the source of the data and cite its uniqu
Below is the recommended format for citing this work in a research pub

Wilms, J., & Weiergraeber, O. H. (2025). Single crystal x-ray diffracti (Version 1) [Dataset]. European Synchrotron Radiation Facility. doi.org

Metadata

Identifier
DOI 10.15151/ESRF-DC-2039871972

Proposal
MX-2587

Beamline
ID30A-3

Public release year
2025

Publisher
European Synchrotron Radiation Facility

License (for files)
Creative Commons Attribution 4.0

Deposit date
2025-02-07

Items 1-1 of 1 Show 20

07/10/2024 11:02:25 **Helical**

Acquisition	
Dataset name	run_04_helical
Start time	07/10/2024 11:02:25
Position	Point-5
Workflow	
Prefix	GAB_SP-GI2-4_1_%04d.h5
Run #	
# Images	1200
Transmission	12.0

Beam	
Res. (corner)	
Wavelength	0.97
Axis range	
Axis start (end)	- -
Exposure Time	0.016
Flux start	6.56e+11 ph/s
Flux end	6.56e+11 ph/s
X Beam	77.31 mm
Y Beam	81.32 mm

Sample	
Protein	GAB_SP
Sample	GAB_SP-GI2-4
Parcel	
Container name	CQ-381A
Container type	UNIPUCK
Position in container	4
Position in changer	1-2
Beamline location	ID30A-3

Optics	
Focussing optics	NA
Mono type	NA
Beam size X	NA µm
Beam size Y	NA µm
Beam shape	ELLIPTICAL
Beam divergence Hor	NA µrad
Beam divergence Vert	NA µrad
Polarization	NA

Synchrotron
Name
Filling mode
Current

Detector
Type
Model
Manufacturer
Pixel size horizontal
Pixel size vertical
Distance

/data/visitor/mx2587/id30a3/20241007/RAW_DATA/GAB_SP/GAB_SP-GI2-4/run_04_helical

Items 1-1 of 1 Show 20

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

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

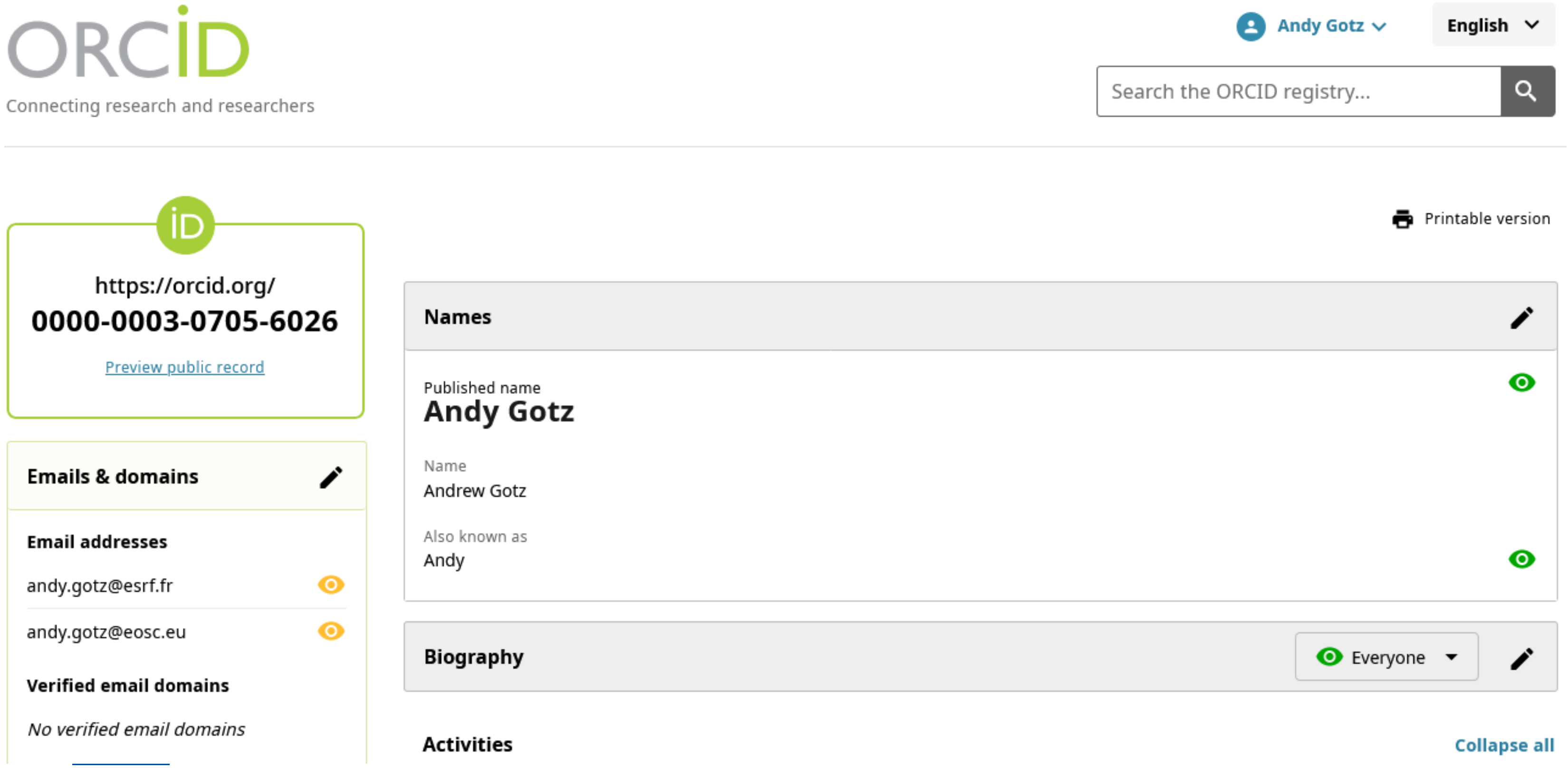
- Benefits of an ORCID identity:
  - You will be distinguished from every other researcher, even researchers who share your same name,
  - Your research outputs and activities will be correctly attributed to you,
  - Your contributions and affiliations will be reliably and easily connected to you,
 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

Achieving100% Open Identifiers:

1. *All scientists encouraged to create an ORCID*
2. *Encourage the use of ORCID for users for publications*



https://orcid.org/
0000-0003-0705-6026

Emails & domains

Email addresses

andy.gotz@esrf.fr

andy.gotz@eosc.eu

Verified email domains

No verified email domains

Names

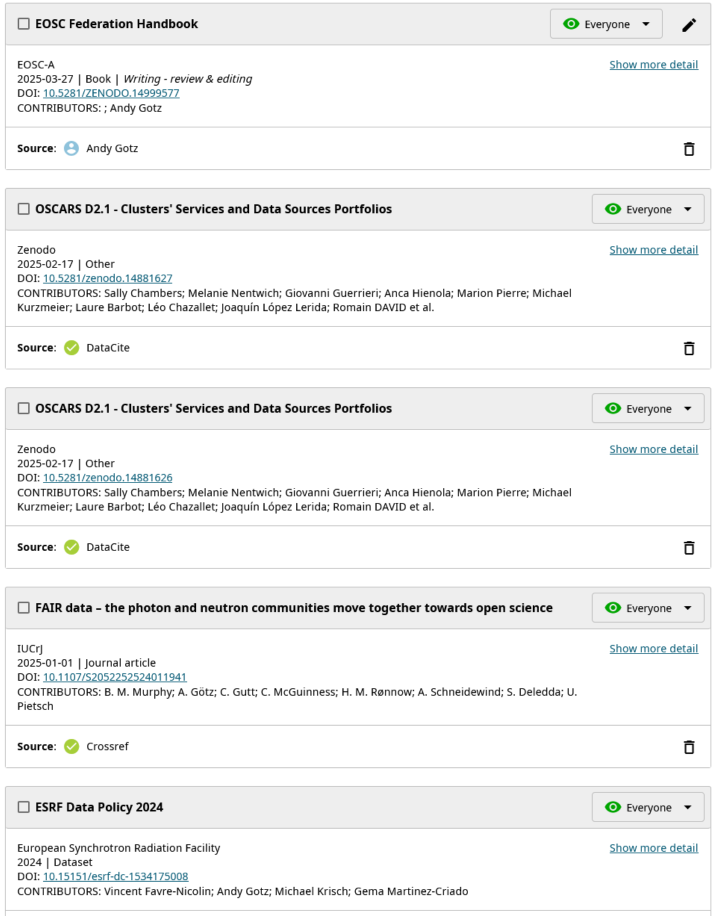
Published name
Andy Gotz

Name
Andrew Gotz

Also known as
Andy

Biography

Activities



☐ **EOSC Federation Handbook**

EOSC-A
2025-03-27 | Book | *Writing - review & editing*
DOI: 10.5281/ZENODO.14999577
CONTRIBUTORS: ; Andy Gotz

Source: Andy Gotz

☐ **OSCARS D2.1 - Clusters' Services and Data Sources Portfolios**

Zenodo
2025-02-17 | Other
DOI: 10.5281/zenodo.14881627
CONTRIBUTORS: Sally Chambers; Melanie Nentwich; Giovanni Guerrieri; Anca Hienola; Marion Pierre; Michael Kurzmeier; Laure Barbot; Léo Chazallet; Joaquín López Lerida; Romain DAVID et al.

Source: ✓ DataCite

☐ **OSCARS D2.1 - Clusters' Services and Data Sources Portfolios**

Zenodo
2025-02-17 | Other
DOI: 10.5281/zenodo.14881626
CONTRIBUTORS: Sally Chambers; Melanie Nentwich; Giovanni Guerrieri; Anca Hienola; Marion Pierre; Michael Kurzmeier; Laure Barbot; Léo Chazallet; Joaquín López Lerida; Romain DAVID et al.

Source: ✓ DataCite

☐ **FAIR data – the photon and neutron communities move together towards open science**

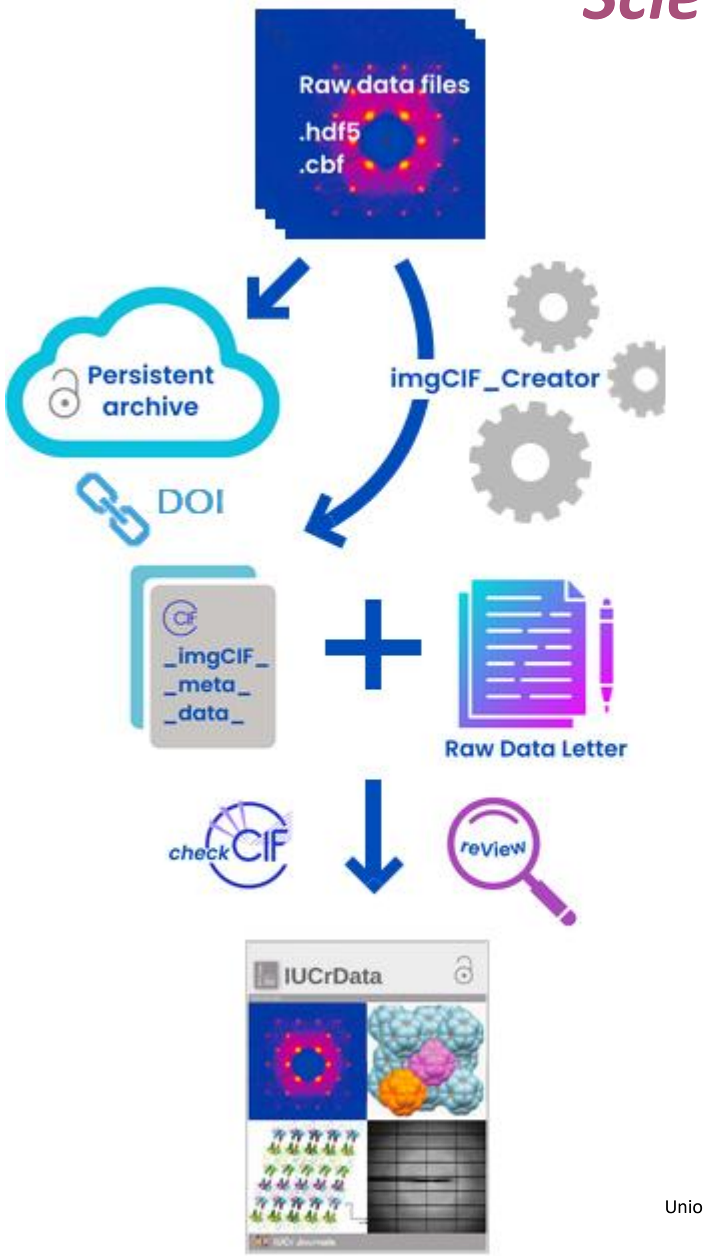
IUCrJ
2025-01-01 | Journal article
DOI: 10.1107/S2052252524011941
CONTRIBUTORS: B. M. Murphy; A. Götz; C. Gutt; C. McGuinness; H. M. Rønnow; A. Schneidewind; S. Deledda; U. Pietsch

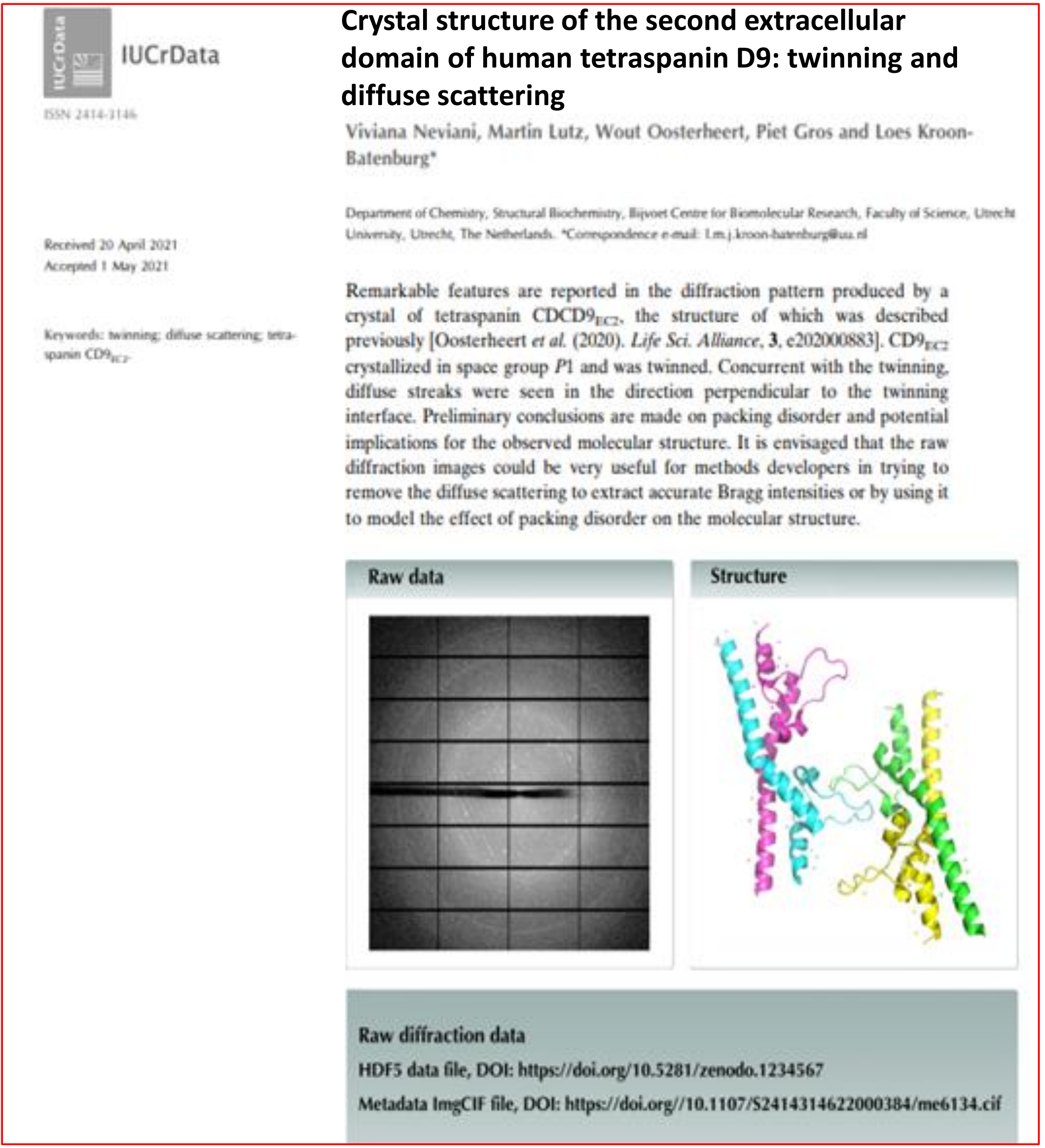
Source: ✓ Crossref

☐ **ESRF Data Policy 2024**

European Synchrotron Radiation Facility
2024 | Dataset
DOI: 10.15151/esrf-dc-1534175008
CONTRIBUTORS: Vincent Favre-Nicolin; Andy Gotz; Michael Krisch; Gema Martinez-Criado

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




IUCrData

ISSN 2414-3146

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

Received 20 April 2021
Accepted 1 May 2021

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

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

Estimated carbon footprint of experiment

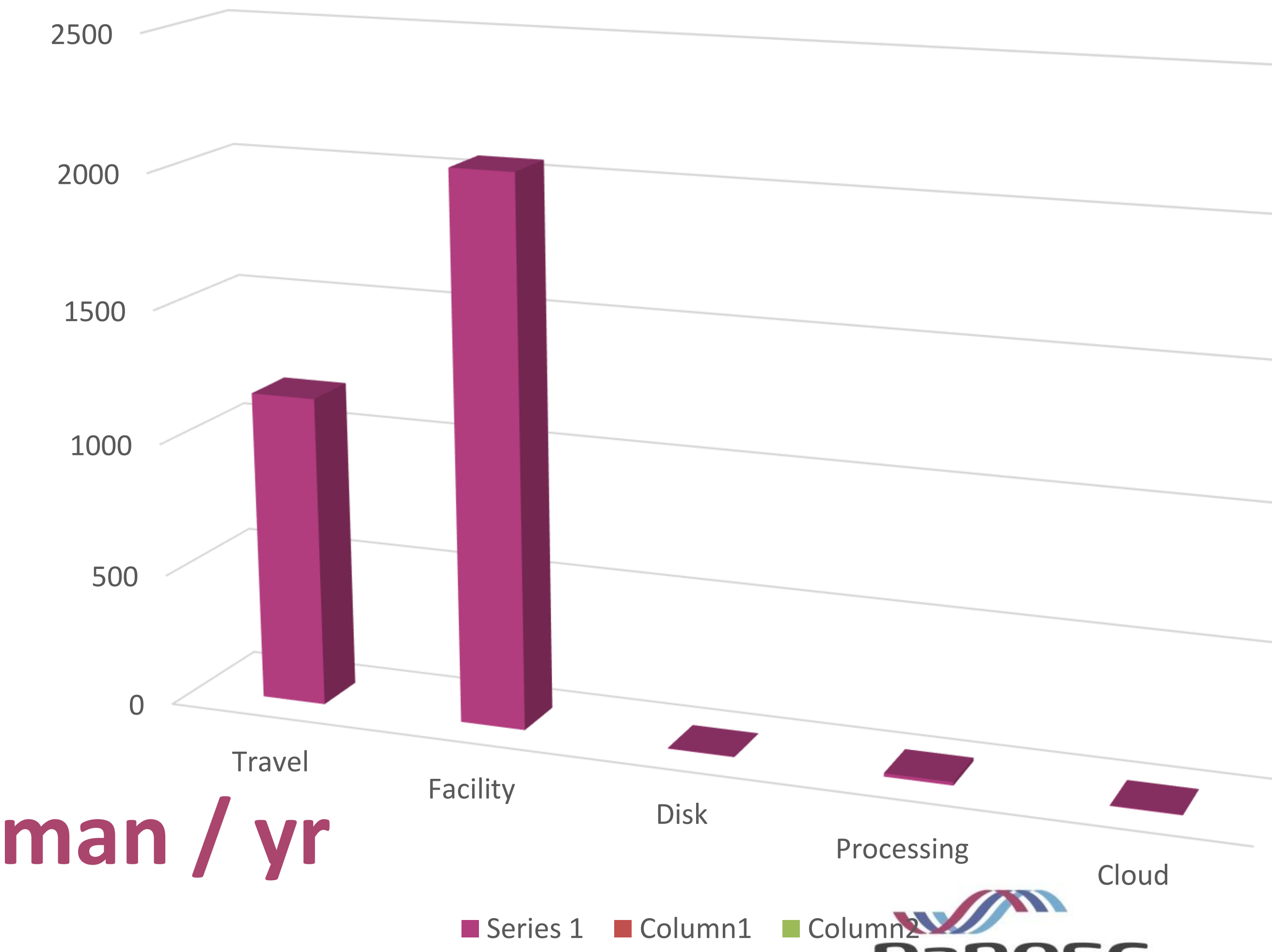
- User Travel = **1170 kg**
- Beamtime energy consumption = **2056 kg**
- Data stored on disk = **1.8 kg**
- Data processing on site = **12.6 kg**
- Cloud transfer = **2.3 kg**

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

TOTAL = 3.253 tons !

Sustainable Goal = 5 tons / human / yr

Carbon footprint for 1 week experiment @ ESRF



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 CO_2 equivalent needed to acquire the raw data!

TAKEAWAY → TAKE CARE of DATA by making it FAIR!

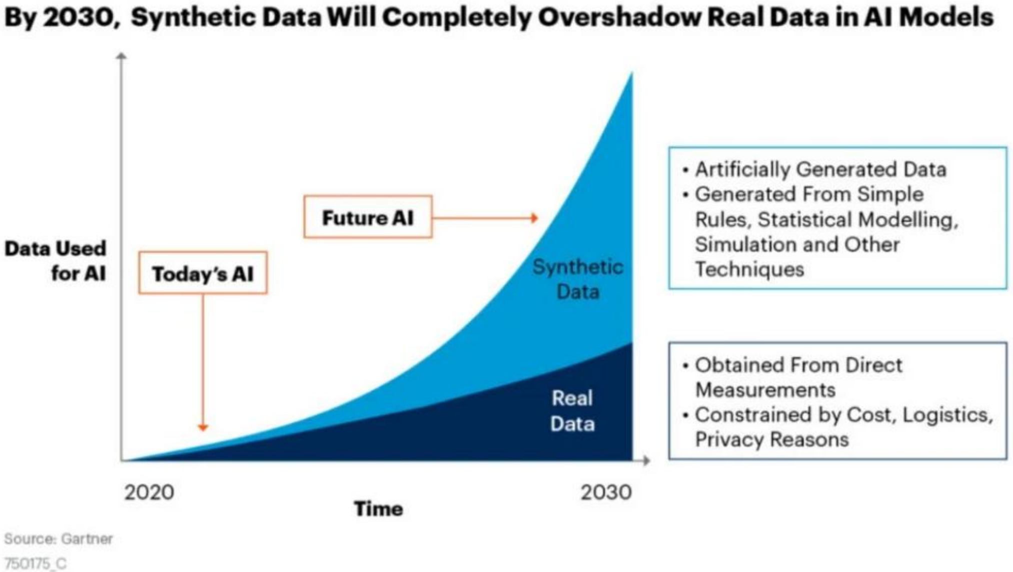
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

What are the advantages of 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
- You protect yourself against plagiarism and fraud

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

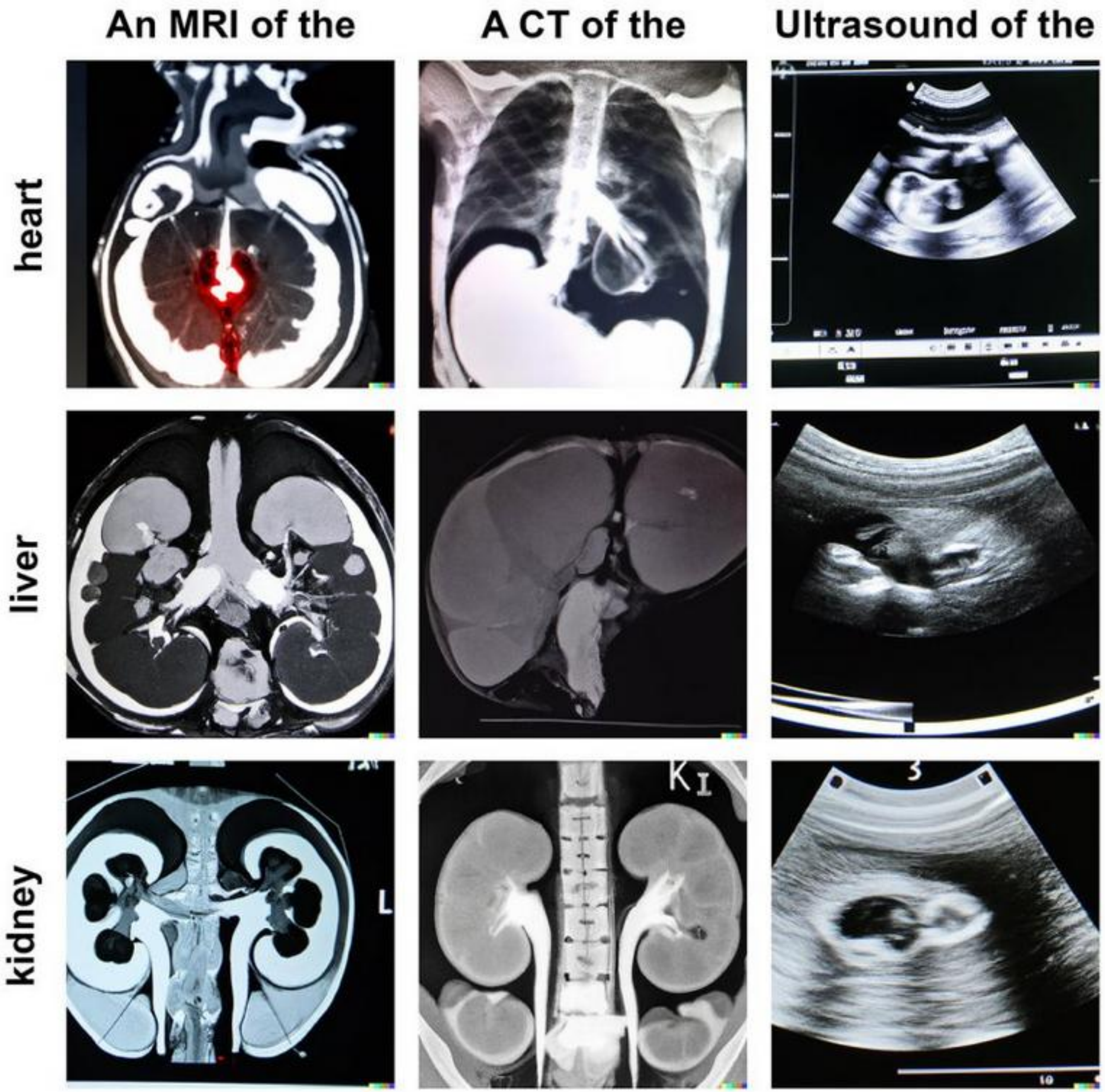


By 2030, Synthetic Data Will Completely Overshadow Real Data in AI Models

- Artificially Generated Data
- Generated From Simple Rules, Statistical Modelling, Simulation and Other Techniques

- Obtained From Direct Measurements
- Constrained by Cost, Logistics, Privacy Reasons

Source: Gartner
750175_C



gramn

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 #1

1. **Learn about the data you will produce before going to the synchrotron**
2. **Make sure you follow a checklist which covers the following topics:**
    1. Data Management Plan, Data Policy, Data Outputs, File types, File Formats, Software, Workflows, e-Logbooks, Data Storage, Data Archiving, Data DOI
3. **Spend time with your data to make it FAIR by adding rich metadata, your ORCID, releasing it, publishing it, citing the data DOI !**
4. **Many digital resources and tools exist for treating your data seriously + publishing them**

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

Conclusions #2

1. Scientific results are much more than the publications
2. Data Availability in publications – STOP using the phrase “data available on reasonable request”
3. Make sure you cite data DOIs !!!

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

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

panosc

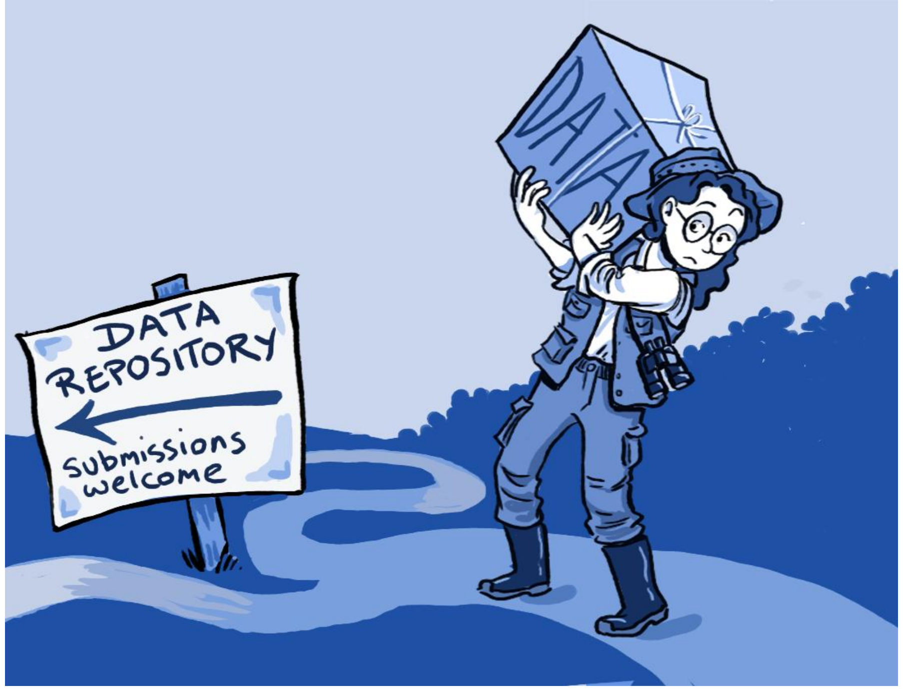
panosc
photon and neutron
open science cloud

~~Data availability~~

~~Data available on reasonable request to the authors.~~

Thank you for listening!

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

Join the debate

https://www.ouvrirlascience.fr/join-the-debate/

In my field, we don’t do open science

If I disseminate my scientific work in open access, everyone will be able to use it without citing me

My data belongs to me

Open access is a threat to certain publishers

Open access publishing is too costly for my institute

If I make my thesis open access, I won’t be able to publish it

In my field I have to choose a journal based only on the impact factor

A data management plan will simply increase my workload without benefiting me

Engaging in open science will penalise me in the evaluation process as a researcher

opean Union’s Horizon 2020 resea

Acknowledgements

- RDMKit Elixir online guide
- Data Management Made Simple, Quirin Shiermeier, *Nature* **555**, 403-405 (2018), *doi: https://doi.org/10.1038/d41586-018-03071-1*
- University of Saskatchewan
  - https://library.usask.ca/studentlearning/workshops/grad-research.php#panel-section-3-ResearchDataManagementWhatYouNeedtoKnow
- **Nature** magazine, Scientific Data
- **PaNOSC, ExPaNDS EOSC** H2020 projects
- **OSCARS**, **OSTrails EOSC** Horizon Europe projects
- **Wikipedia,**
- **Internet**

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

Resources to help make your research reproducible

A non-exhaustive list of resources to explore per country (mainly EU but also China)

- https://the-turing-way.netlify.app/index.html
- https://www.crs.uzh.ch/en/resources/CRS-Reproducibility-Notes.html
- https://www.ouvrirlascience.fr/home/
- https://www.fun-mooc.fr/fr/cours/la-science-ouverte/
- https://avointiede.fi/en
- https://open-science-future.zbw.eu/en/
- https://www.openscience.nl/en
- https://www.csic.es/en/open-science
- https://www.ciencia-aberta.pt/home
- https://gacr.cz/en/gacr-and-open-science/

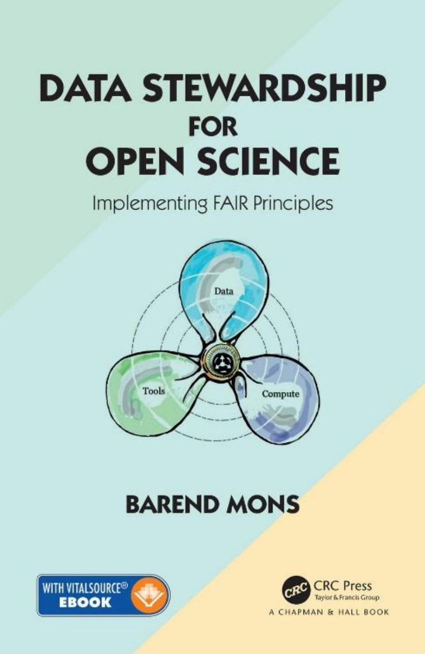
- https://eosc-austria.at/hands-on-open-science/
- https://open-science.it/english
- https://open-sci.cn/

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.

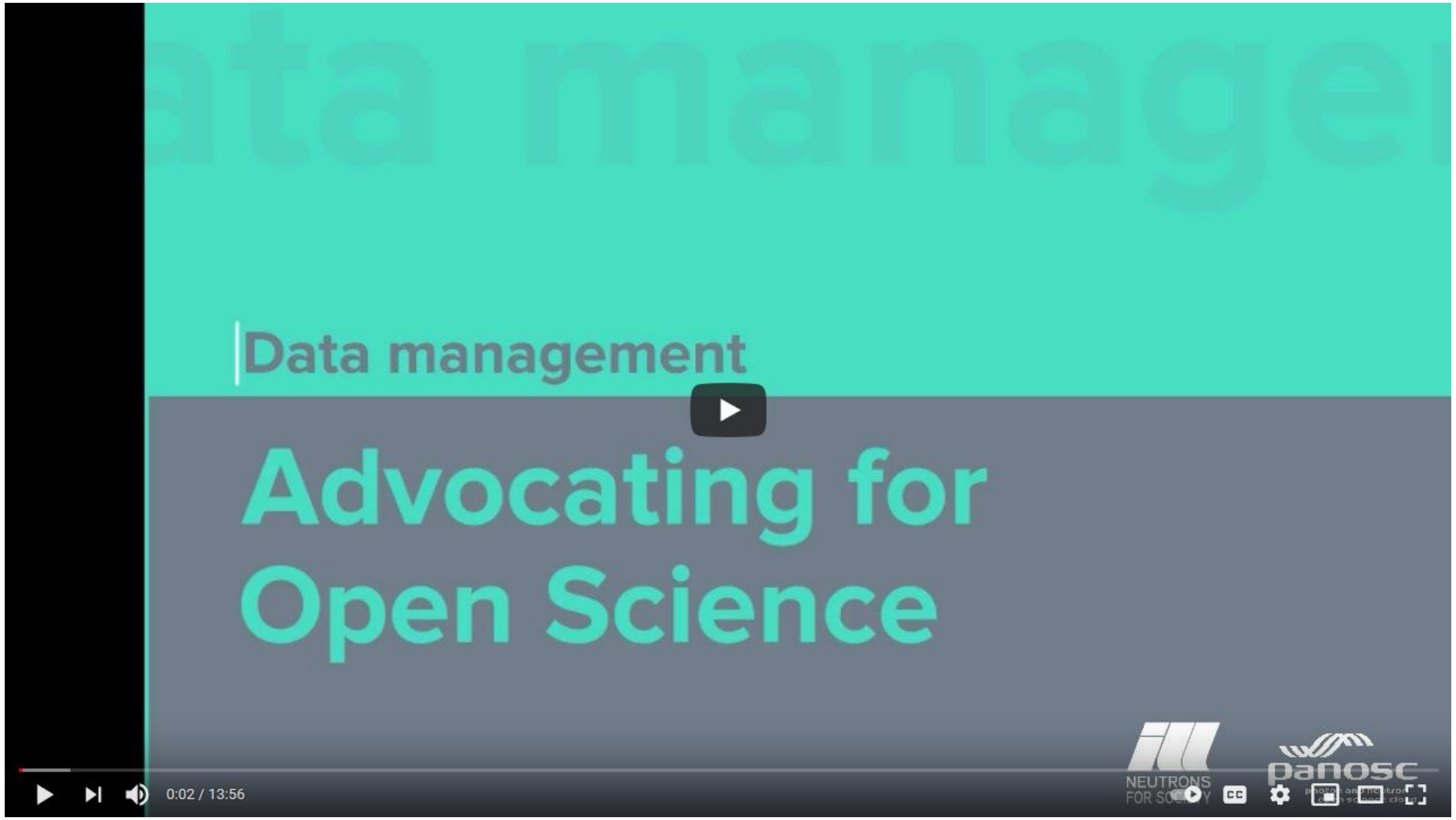
- Recommended reading:



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

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



Data management

Advocating for Open Science

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