Open Data and Open Science

Spotlight on Publishing and Open Science

15th November, 2021

by Andy Götz (ESRF)

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 Data for scientists doing research with an emphasis on data from life sciences @ ESRF

- What you should learn from this talk:

What is the Global, European, Local view?

What are Open and FAIR Data?

What do you as scientists get?

What should you as scientists do?

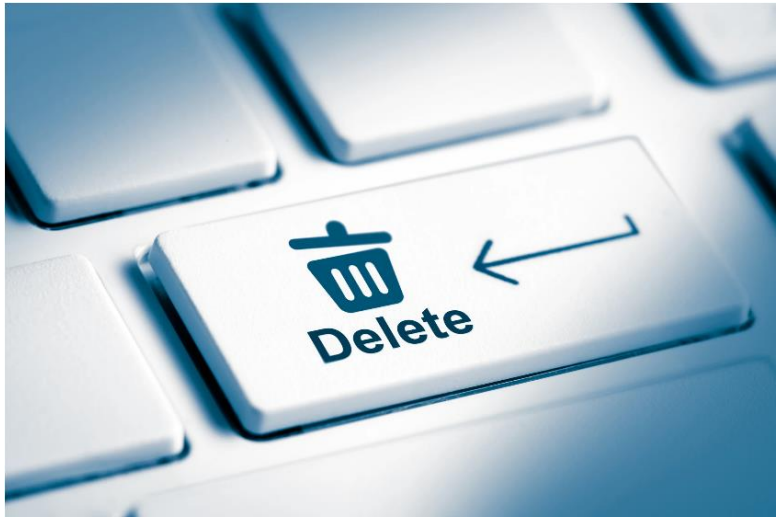
What services are available?

What services are planned

panosc

My background in Open Data

- Software developer and manager @ ESRF since 30+ years
- **Co-author** of the **ESRF data policy** + leading the ESRF data policy implementation
- **Coordinator** of the PaNOSC project for making data from photon and neutron sources FAIR https://panosc.eu
- Member of the **IUCr Committee on Data**



ESRF has gone from deleting all data to preserving all raw data in 30 years

panosc

Definition of Open Data

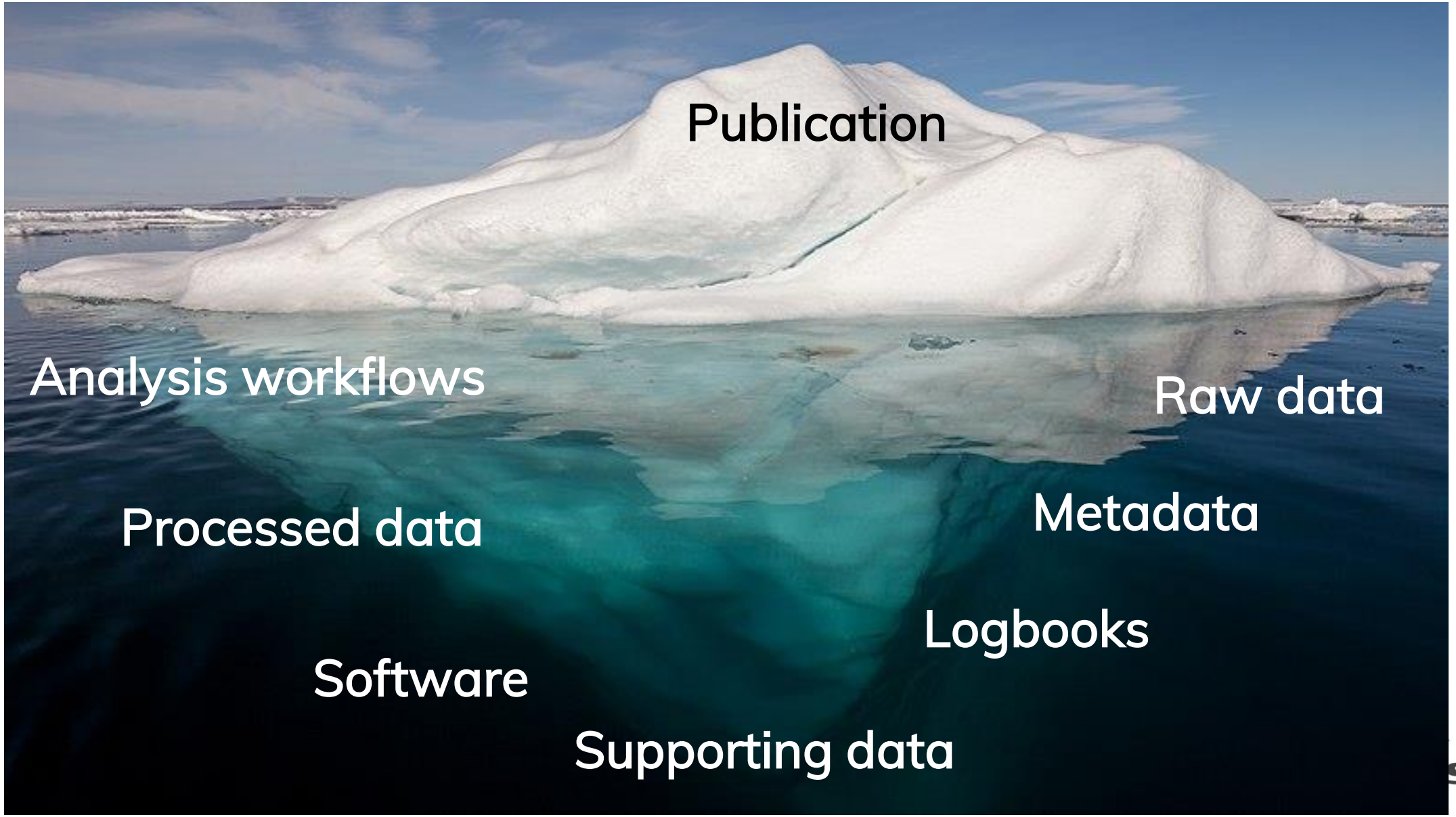
Data that are freely available to all for re-analysis and re-use

panosc

Some questions raised by Open Data

- How to access to Open Data?
- How to attribute Open Data to give credit?
- How to ensure Open Data are preserved?
- How to ensure Open Data are useful?
- Who owns Open Data?

Science produces much more than Publications



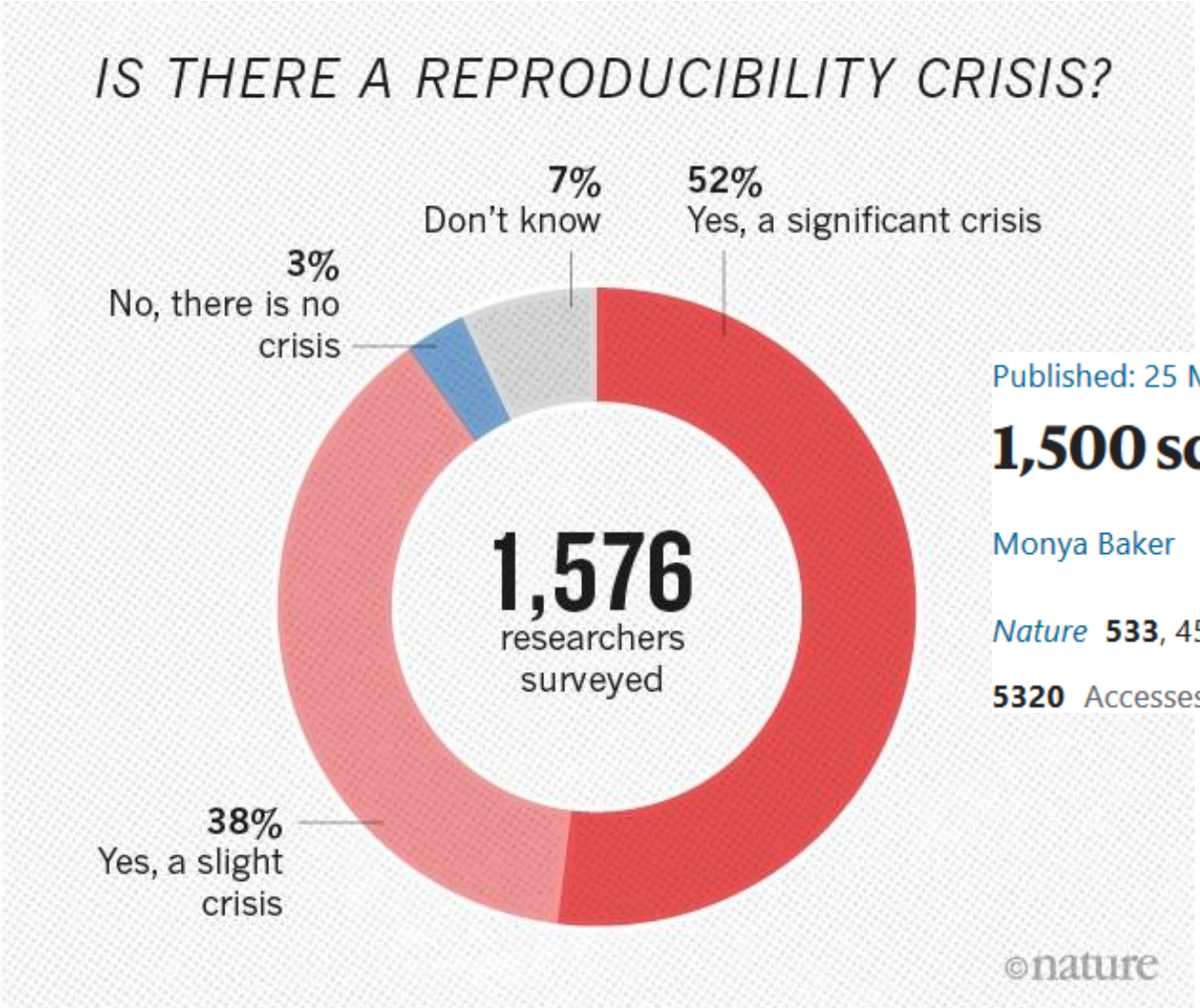
https://en.wikipedia.org/wiki/File:Iceberg_in_the_Arctic_with_its_underside_exposed.jpg

Why make data open?

1. Improve the reproducibility of scientific publications
2. Improve the quality of scientific data by curation
3. Save time for scientists by managing data
4. Managing data enable data services to be developed
5. Publicly funded research should be public
6. Allow data to be re-used by other scientists
7. Preserving data prevents data loss
8. Data can be used for developing new software
9. AI/ML algorithms can be trained on open data
10. Difficult to interpret data can get help from community

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

Open Data are a fundamental part of addressing the Reproducibility Crisis



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

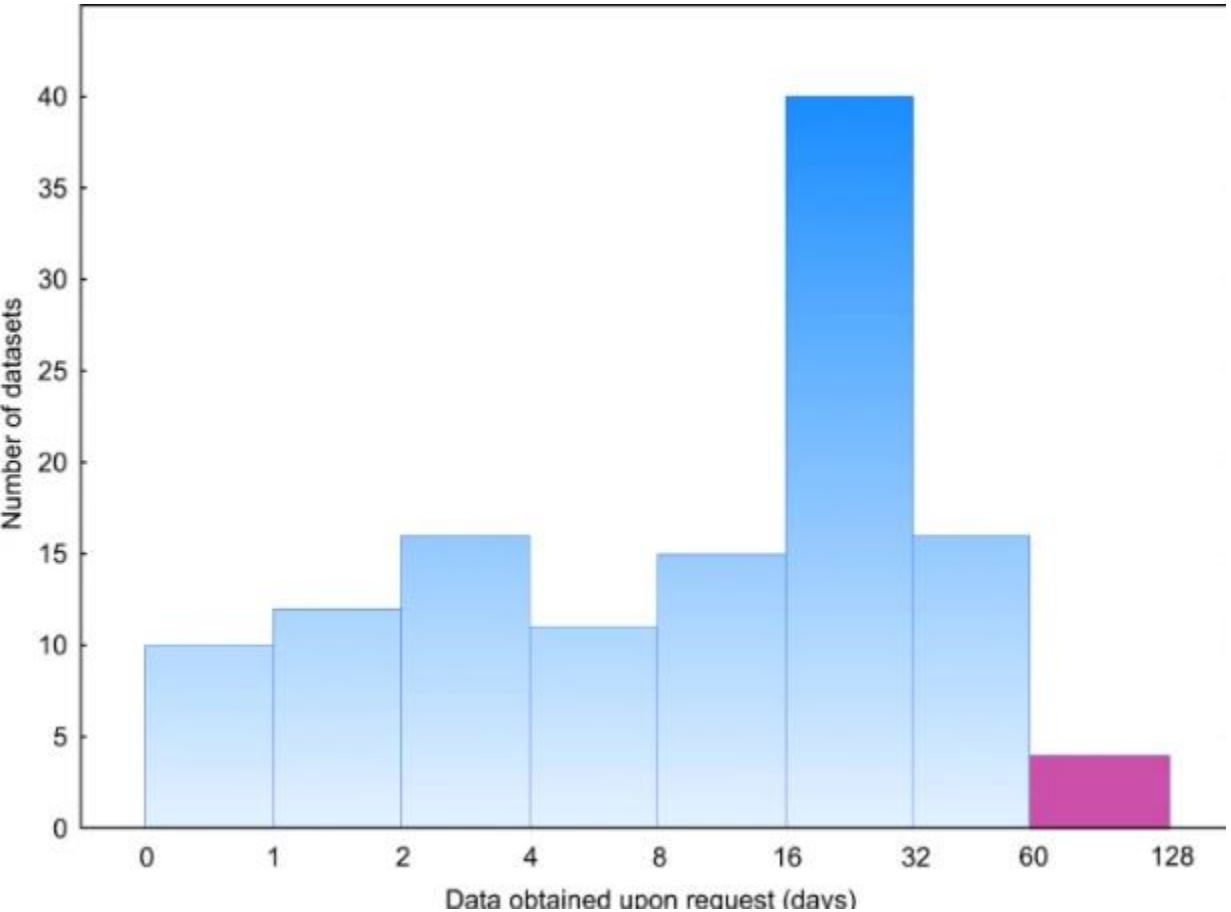
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

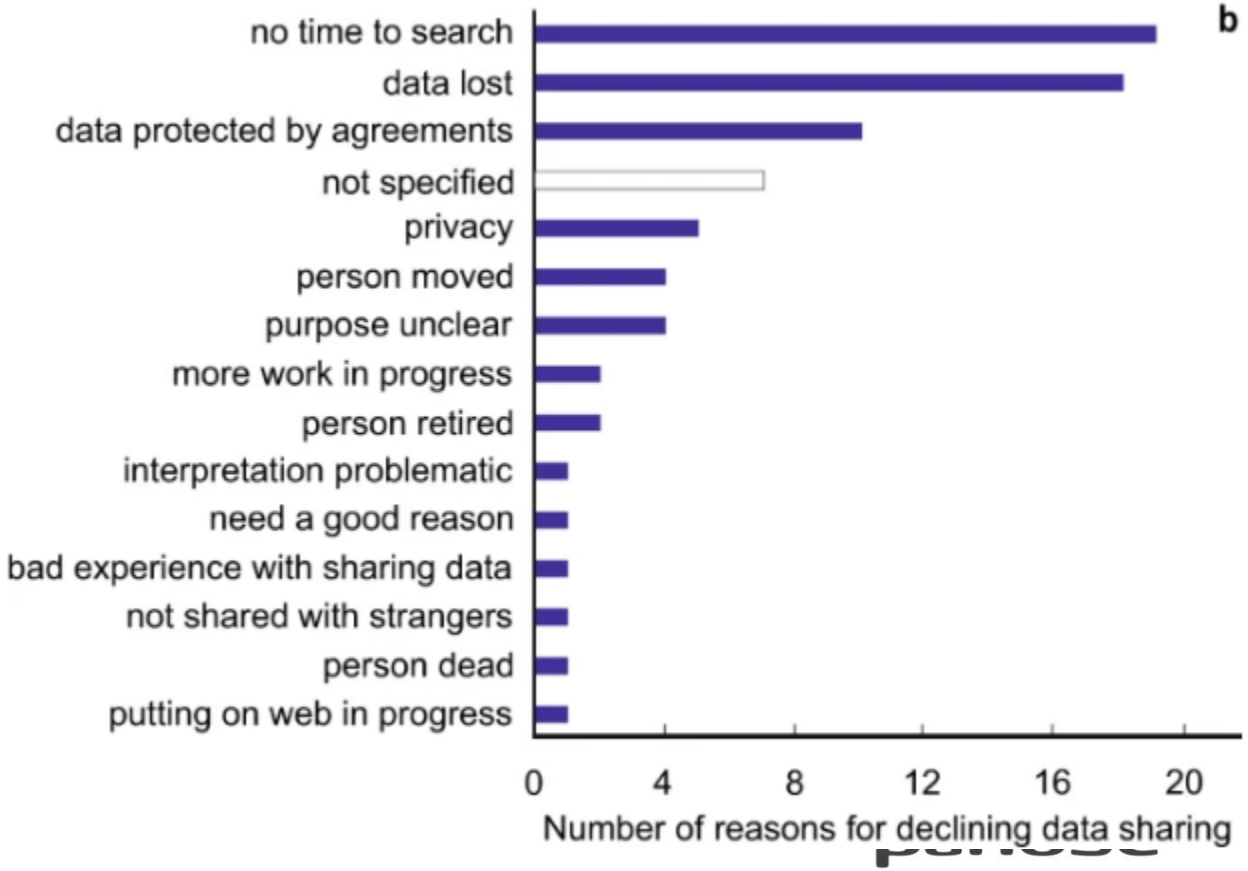
Analysis | Open Access | Published: 27 July 2021

Data sharing practices and data availability upon request differ across scientific disciplines

Leho Tedersoo ✉, Rainer Küngas, Ester Oras, Kajar Köster, Helen Eenmaa, Äli Leijen, Margus Pedaste, Marju Raju, Anastasiya Astapova, Heli Lukner, Karin Kogermann & Tuul Sepp

Scientific Data **8**, Article number: 192 (2021) | Cite this article





Recent study on Data availability in Science + Nature

... critical data are still unavailable for re-analysis or meta-analysis for more than half of the papers published in Nature and Science in the last decade.

While the majority of data are eventually available, it is alarming that less than a half of the data clearly stated to be available upon request could be effectively obtained

Our experience shows that receiving data typically required long email exchanges with the authors

https://doi.org/10.1038/s41597-021-00981-0

panosc

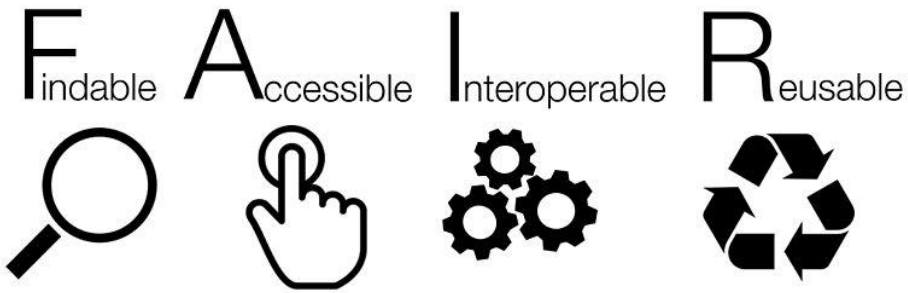
FAIR guiding principles

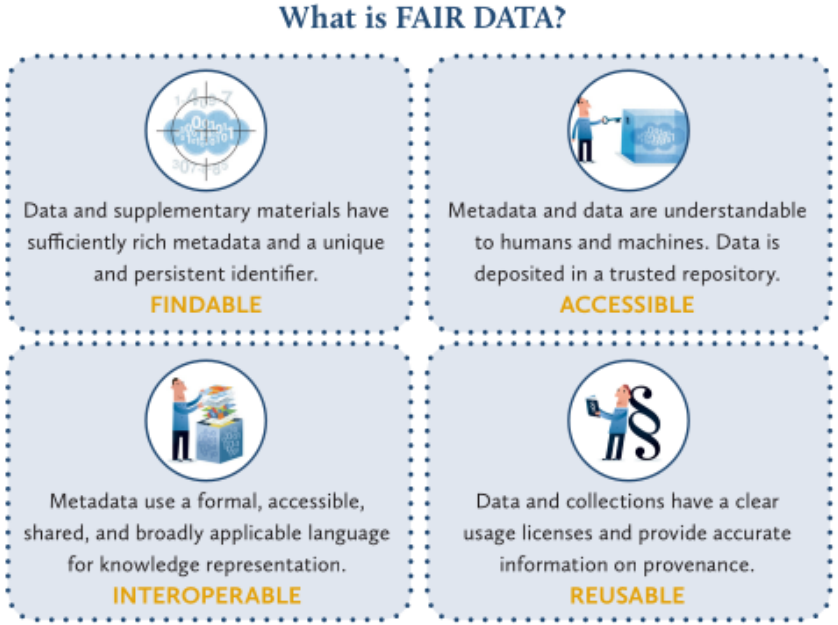
Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

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





This project has received funding from th

panosc

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

https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0

European Open Science Cloud – a cloud of FAIR data

- EC has spent 260 million euros on the EOSC so far
- The main aim of the EOSC is to make FAIR data a reality
- 4 of the 7 goals of the EOSC concern data
  - Opening up scientific data: The implementation of data management plans to make research data findable, accessible, interoperable and re-usable (FAIR principles).
  - Federation of existing scientific data infrastructures with new provisioning schemes, such as the cloud or specialised facilities, and the development of nodes to link existing national data centres, European e-infrastructures, external providers and research infrastructures.
  - Development of specifications for application interoperability (APIs), data portability and data sharing: To enable data to be shared across disciplines and infrastructures, more standardisation of meta-data and, perhaps, the actual data itself will be needed.
  - Creation of search tools: New software will be required to enable scientists to search, browse and access research data.

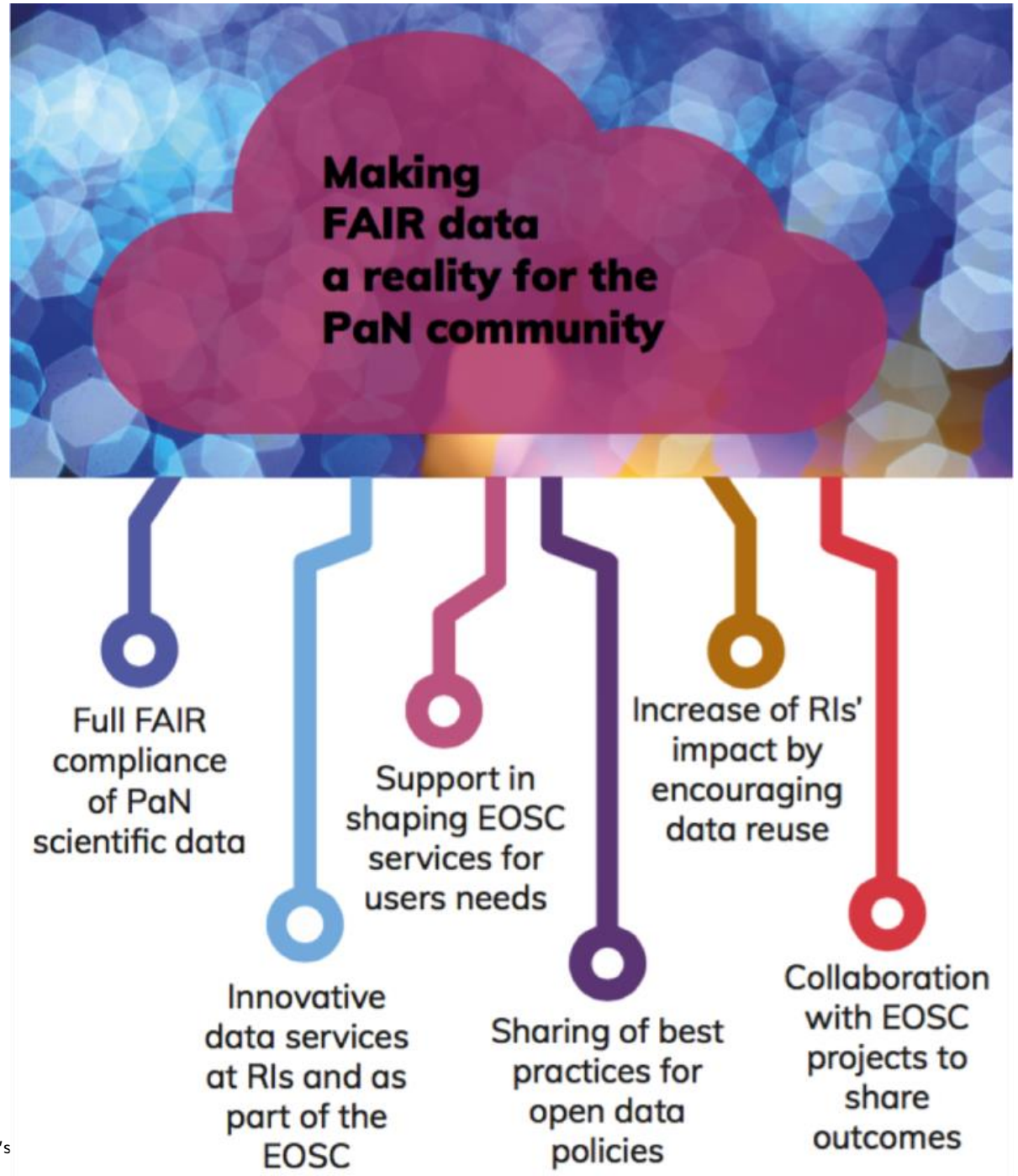
https://sciencebusiness.net/report/european-open-science-cloud-who-pays-what

H2020 financed projects - PaNOSC and ExPaNDS

EU Call	HORIZON 2020 INFRA-EOSC-04	HORIZON 2020 INFRA-EOSC-5B
Description	Cluster of ESFRI PaN Sources	EOSC PaN Data Services
Partners	ESRF, ILL, ESS, EU-XFEL, CERIC-ERIC, ELI-DC, EGI	DESY, ALBA, DLS, ELETTRA, EGI, HZB, HZDDR, Max IV, PSI, Soleil, UKRI
Observers	GEANT EU-DAT National RI's	
Linked 3rd Party	DESY STFC CESNET	
Start – End (Duration)	2018-12-01 – 2022-11-30 [4 Years]	2019-09-01 – 2023-02-28 [3 ½ Years]
Coordinators	A. Götz, G. Bodera	P. Fuhrmann, S. Servan
Budget	12 M Euros	6 M Euros
Home Page	https://panosc.eu	https://expands.eu
Twitter	@PaNOSC_eu #PaNOSC	@ExPaNDS_eu #ExPaNDS
GitHub	github.com/panosc-eu	github.com/expands-eu

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

PaNOSC objectives



This project has received funding from the European Union's

FAIR == AI-ready

No-one really argues against the idea that data, as well as the accompanying workflows and services should be findable, accessible under well-defined conditions, interoperable without data munging, and thus optimally reusable.

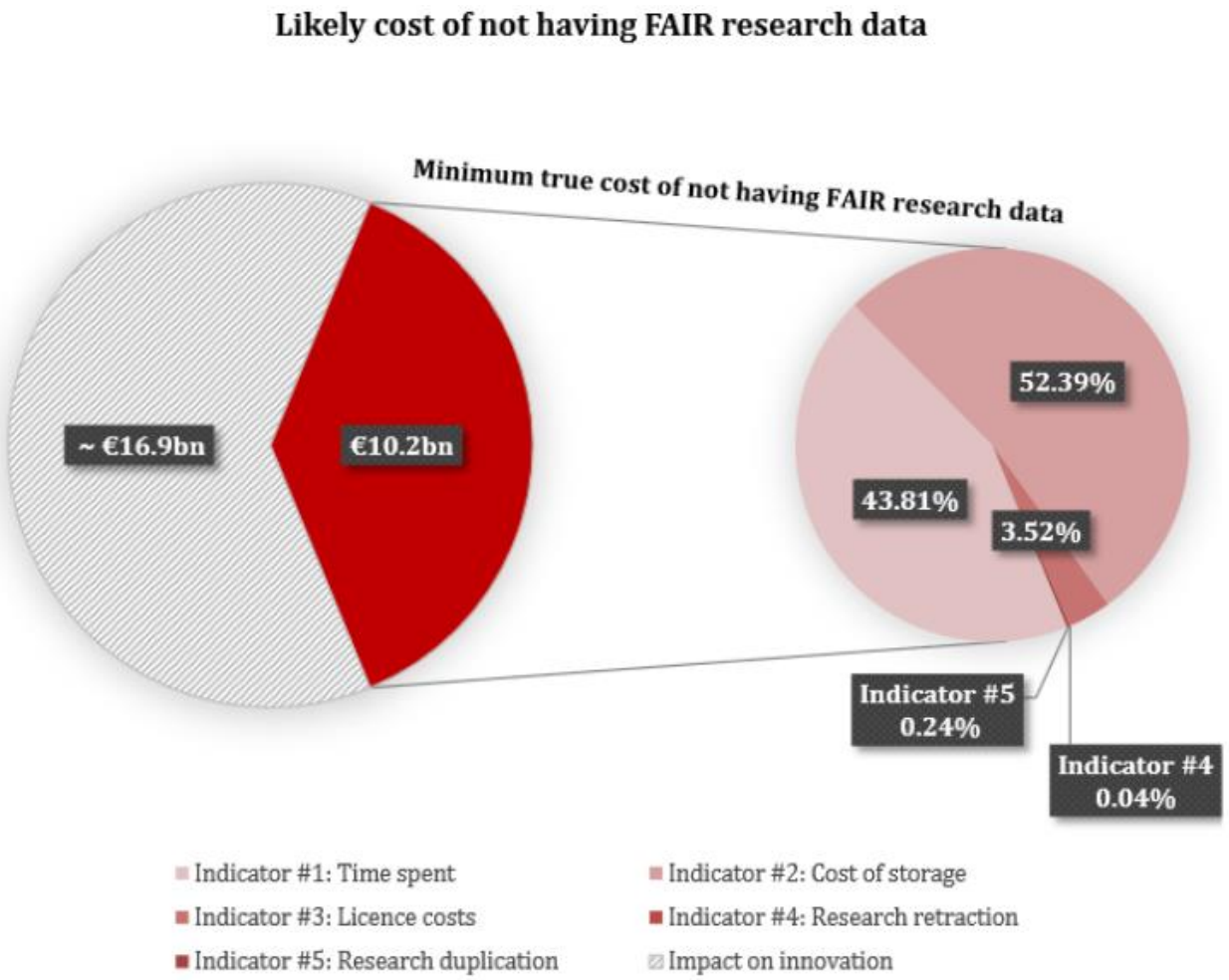
Being FAIR is not a goal in itself; FAIR Data and Services are needed to enable data intensive research and innovation and (thus) have to be "AI-ready①" (= future proof for machines to optimally assist us).

However, the fact that science and innovation becomes increasingly "machine-assisted" and hence the central role of machines, is still overlooked in some cases when people claim to implement FAIR.

https://doi.org/10.1162/dint_e_00023 The FAIR Principles, **B.Mons** et al

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

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

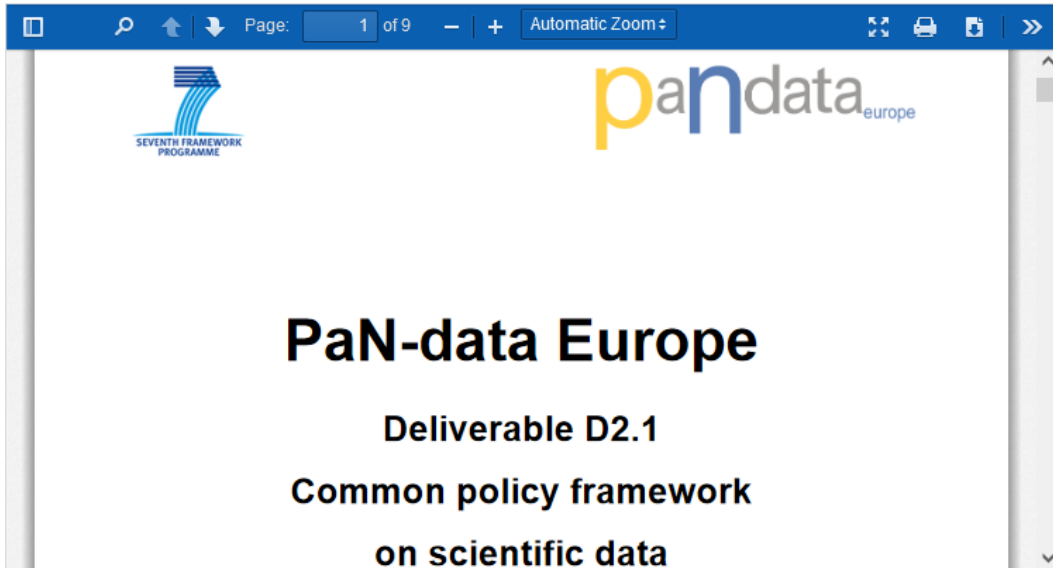


"Cost-benefit analysis for FAIR research data " (https://op.europa.eu/s/pevt)

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

Photon and Neutron Sources – Open Data policies

- **2010** - drafted the so-called **PaNdata Data Policy Framework**
- The framework was used for **ALL PaN open data policies**
- **2020** - **updated** by **PaNOSC**
- **2021 - ExPaNDS + PaNOSC reviewed and** documented **Lessons Learned** and **Best Practices**



https://doi.org/10.5281/zenodo.3738497

Implementing the data policies at PaNs remains challenging

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

panosc

EU Research facilities Data Policies

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

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

ESRF Data Policy

https://www.esrf.fr/datapolicy

ESRF EUROPEAN SYNCHROTRON RADIATION FACILITY

30 November 2015

The ESRF Data Policy

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

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

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

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

panosc

Main elements of ESRF Data Policy

- Data are **under embargo for 3 years** but can be released earlier by the experimental team
- All data have a **DOI assigned automatically at the session level** with appropriate high-level metadata from proposal
- **Experimental team can register a DOI** for a subset or superset of the data on provision of high-level metadata e.g. abstract
- PI can request an **extension** to the **embargo** period
- **DOI** must be **cited** when data are re-used
- **Processed** data can be uploaded to data portal
- Data can be downloaded from https://data.esrf.fr
- After embargo period and/or on creation of bespoke DOI Data are available under **CC-BY-4.0 licence** as **Open Data**

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

https://doi.org/10.1107/S2052252519005918

IUCrJ
ISSN 2052-2525
BIOLOGY | MEDICINE

Findable Accessible Interoperable Re-usable (FAIR) diffraction data are coming to protein crystallography

John R. Helliwell,[a]‡ Wladek Minor,[b]§ Manfred S. Weiss,[c]¶ Elspeth F. Garman,[d]‡‡ Randy J. Read,[e]‡‡ Janet Newman,[f]§§ Mark J. van Raaij,[g]§§ Janos Hajdu[h,i]¶¶ and Edward N. Baker[j]‡‡‡

The structural biology community has always been at the forefront of sharing processed, *i.e.* analysed, results (the PDB was started in 1971)

(i) Authors should provide a permanent and prominent link from their article to the raw data sets

(ii) A registered Digital Object Identifier (doi) should be the persistent identifier of choice

panosc

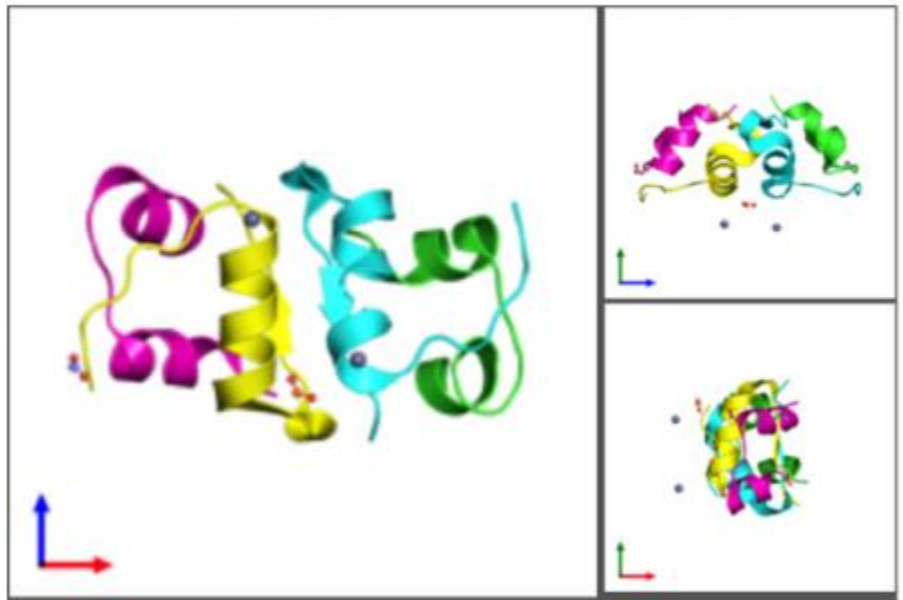
PaNOSC Use Cases - https://www.panosc.eu/all-use-cases/

Use Case 10 – Linking raw data to the Protein Data Bank in Europe (PDBe)

The OneDep deposition system enables depositors to provide the location of raw datasets as a 'digital object identifier' (doi), within their mmCIF file in the PDB archive and this DOI is now directly linked from an entry page at PDBe

1145 PDB depositions from data taken at ESRF since 2020

PDB 6gv0 coloured by chain and viewed from the front



Experimental raw data

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

Raw experimental data related to PDB entry 6gv0:

Data DOI: 10.5281/zenodo.4456817

Dataset type: diffraction image data

Making data FAIR

- **The ESRF Data Policy implementation implements**
  - Findable by ensuring all data have a DOI
  - Accessible by making data open and downloadable

- **Scientists help is needed to ensure**
  - Interoperability by providing complete metadata
  - Reusability by providing complete e-logbooks, auxiliary information on the sample

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

panosc

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

Digital Object Identifiers (DOIs)

- All data must be referenced by a DOI
- **First DOI** minted on request of a scientist (**Matt Bowler**) was for ID30A1
- **Create DOIs** for data which you reference in publications
- **Make your data count!**

ESRF

DOI > 10.15151/ESRF-DC-142893590

Data collection | Dataset | Open access

STRUCTURAL EVIDENCE FOR A ROLE OF THE MULTI-FUNCTIONAL HUMAN GLYCOPROTEIN AFAMIN IN WNT TRANSPORT

Andreas Naschberger ; Bernhard Rupp.

DOI 10.15151/ESRF-DC-142893590

Licence (for files)
Creative Commons Attribution 4.0

Contributors
Matthew W. Bowler.

Abstract
Afamin, a human plasma glycoprotein and putative transporter of hydrophobic molecules, has been shown to act as extracellular chaperone for poorly soluble, acylated Wnt proteins, forming a stable, soluble complex with functioning Wnt proteins. The 2.1-Å crystal structure of glycosylated human afamin reveals an almost exclusively hydrophobic binding cleft capable of harboring large hydrophobic moieties. Lipid analysis confirms the presence of lipids, and density in the primary binding pocket of afamin was modeled as palmitoleic acid, presenting the native O-acylation on serine 209 in human Wnt3a. The modeled complex between the experimental afamin structure and a Wnt3a homology model based on the XWnt8-Fz8-CRD fragment complex crystal structure is compelling, with favorable interactions comparable with the crystal structure complex. Afamin readily accommodates the conserved palmitoylated serine 209 of Wnt3a, providing a structural basis how afamin solubilizes hydrophobic and poorly soluble Wnt proteins.

Proposals	Beamlines
OPID-1	ID30A1

Publication year
2018

Experimental report

The filename list is not available.

Experimental data

The data can be accessed by clicking on the link below

Access data

Reference
Below is the recommended format for citing this work in a research publication.

Naschberger A., Rupp B. (2018). Structural Evidence for a Role of the Multi-functional Human Glycoprotein Afamin in Wnt Transport. European Synchrotron Radiation Facility (ESRF). doi:10.15151/ESRF-DC-142893590

European Synchrotron Radiation Facility

Access to data is governed by the ESRF data policy .

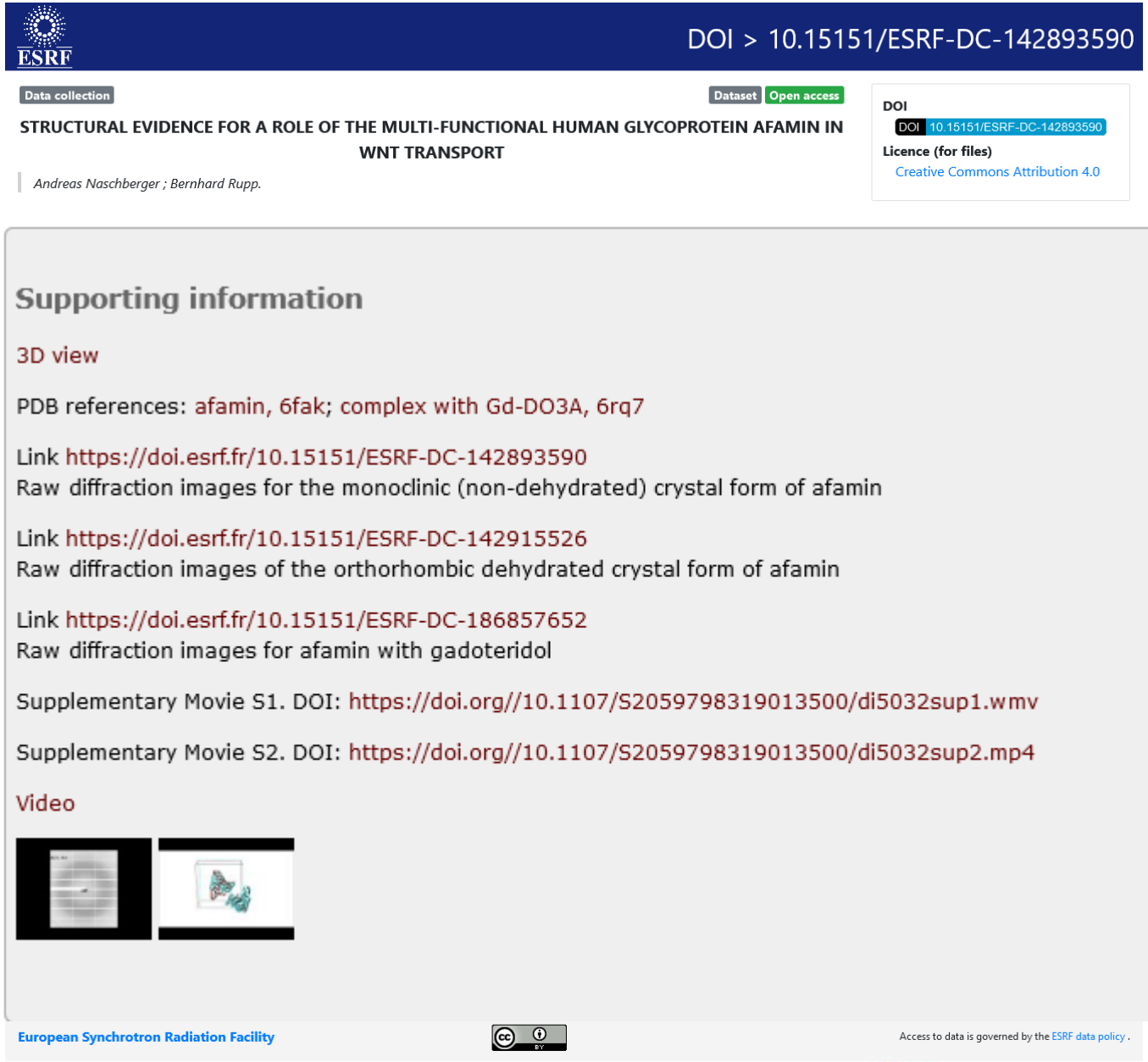
https://doi.esrf.fr/10.15151/ESRF-DC-142893590

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

panosc

Digital Object Identifiers (DOIs)

- **First DOI** minted by a scientist (**Matt Bowler**) cited in the article →



ESRF

DOI > 10.15151/ESRF-DC-142893590

Data collection | Dataset | Open access

STRUCTURAL EVIDENCE FOR A ROLE OF THE MULTI-FUNCTIONAL HUMAN GLYCOPROTEIN AFAMIN IN WNT TRANSPORT

Andreas Naschberger ; Bernhard Rupp.

DOI
DOI 10.15151/ESRF-DC-142893590

Licence (for files)
Creative Commons Attribution 4.0

Supporting information

3D view

PDB references: afamin, 6fak; complex with Gd-DO3A, 6rq7

Link https://doi.esrf.fr/10.15151/ESRF-DC-142893590
Raw diffraction images for the monoclinic (non-dehydrated) crystal form of afamin

Link https://doi.esrf.fr/10.15151/ESRF-DC-142915526
Raw diffraction images of the orthorhombic dehydrated crystal form of afamin

Link https://doi.esrf.fr/10.15151/ESRF-DC-186857652
Raw diffraction images for afamin with gadoteridol

Supplementary Movie S1. DOI: https://doi.org//10.1107/S2059798319013500/di5032sup1.wmv

Supplementary Movie S2. DOI: https://doi.org//10.1107/S2059798319013500/di5032sup2.mp4

Video

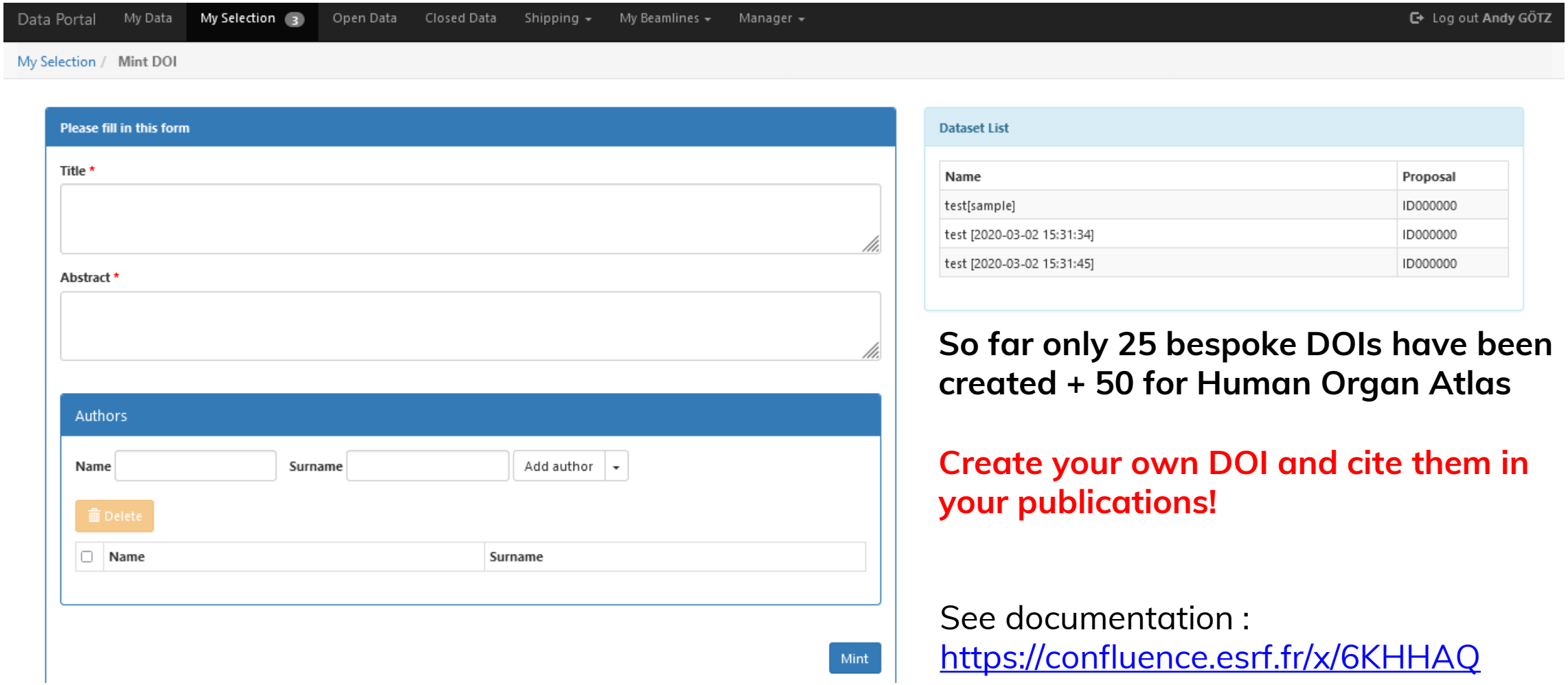
European Synchrotron Radiation Facility

Access to data is governed by the ESRF data policy .

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

panosc

Creating your own DOIs on data.esrf.fr



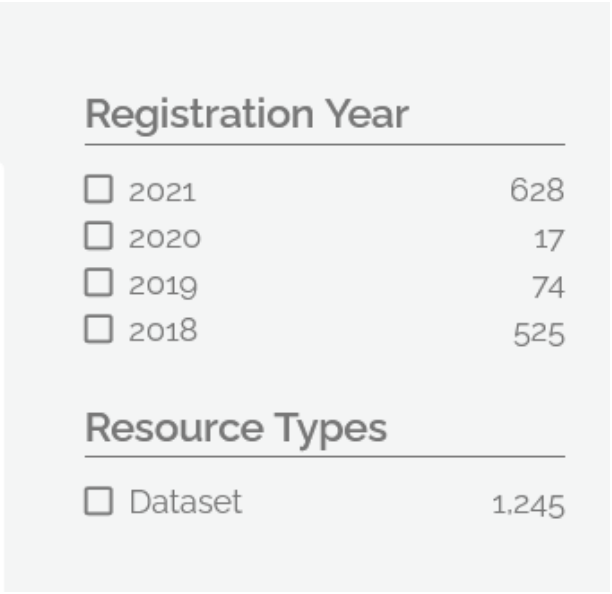
So far only 25 bespoke DOIs have been created + 50 for Human Organ Atlas

Create your own DOI and cite them in your publications!

See documentation :
https://confluence.esrf.fr/x/6KHHAQ

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

Data DOIs minted et ESRF



- Total of 1245 DOIs created so far (1170 automatically)
- 294 proposals are available as Open Data i.e. not under embargo

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 goal is to 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

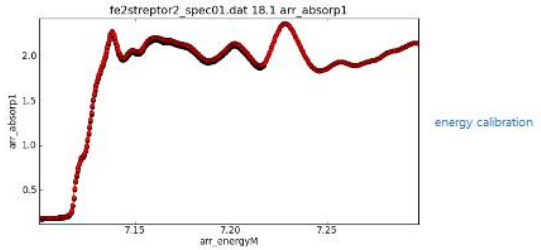
ID21 / EV-280 Beneficial symbiosis in tomato plants: its role on Fe translocation and speciation

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)

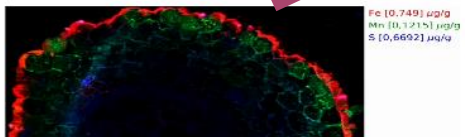


energy calibration

09:58:09 OPTICS> Fexanes_ev280

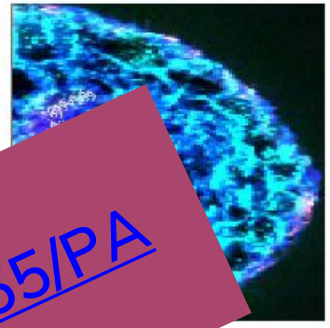
Fe2 strepto r2 Main Root: XANES Poi

00:18:27 OPTICS> zapxiaimage samy 7. samz 25.476 26.025 549 100 0 (zap: #1, spec: #2)



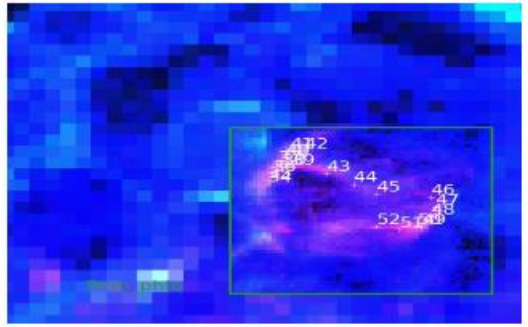
https://indico.eli-laser.eu/event/3/contributions/16/attachments/17/55/PA NOSC%20ESRF%20LOGBOOK.pdf

23:50:29 OPTICS> Fexanes_ev280



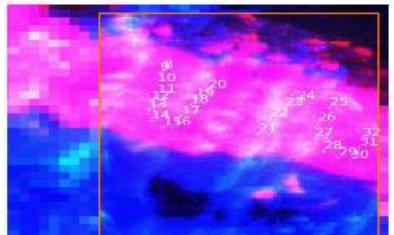
XANES Points

OPTICS> Fexanes_ev280



Fe3 prio r3 Main Root prim phlo: XANES Points

22:25:02 OPTICS> Fexanes_ev280



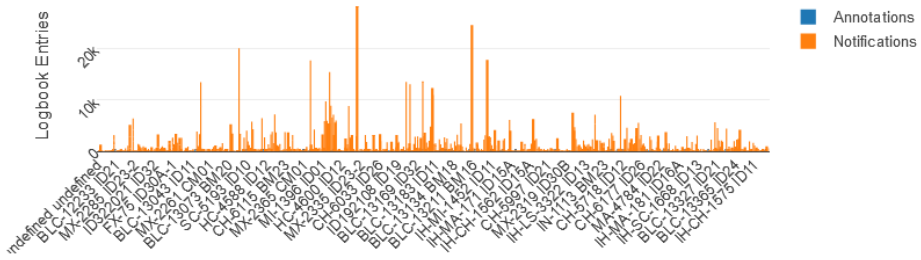
nent No. 8

Electronic notebook usage statistics

- E-Notebooks are still not used extensively
- SB sessions not making use of e-notebooks at all
- Ideas for future
  - Link e-notebooks to LIMS (ISPyB)
  - Train scientists

Logbook Usage for period **2021-08-01** and **2021-11-15**.
Change url parameters, startDate and endDate to change manually the range of dates. Example: /manager/starts/logbook?startDate=2020-08-01&endDate=2020-09-01

Number of entries in the logbook per investigation



General

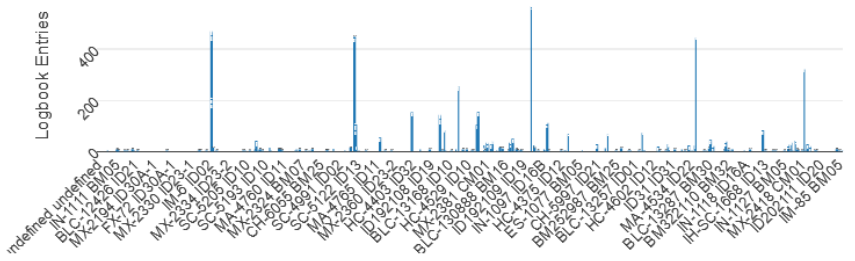
Instruments	**43**
Investigations	**803**
Average Events/Investigations	**1167**
Max. Annotation/investigation	**566**
Max. Notifications/investigation	**28219**
Total Events	**937037**

Event types

Annotations	**4878 (0.52%)**
Notifications	**862935 (92.09%)**

This plot shows the number of entries in the logbook for each investigation.It can give a hint about the verbosity (number of logs automatically sent) of each beamline.

Annotations (hand made entries) per investigation



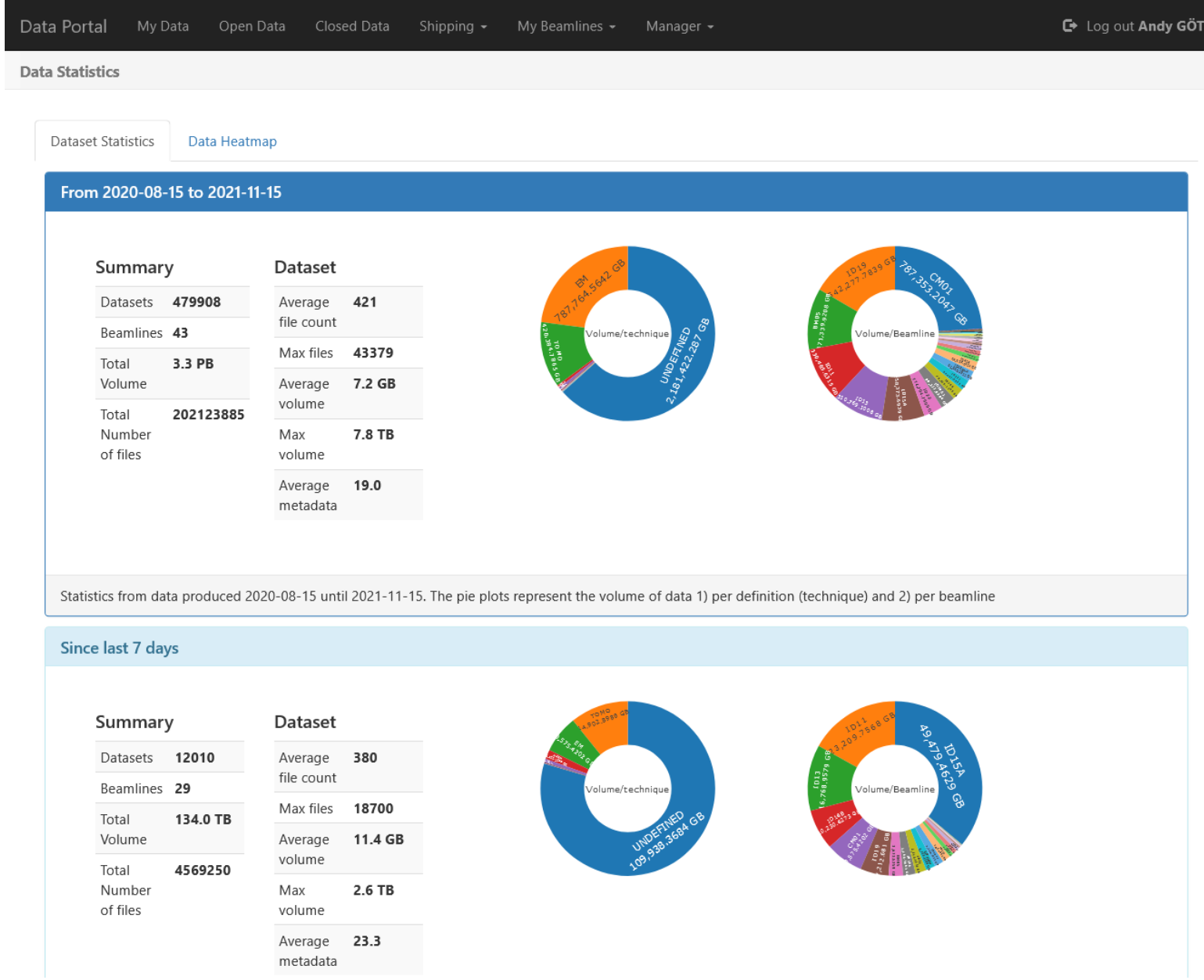
Annotations

Investigation with Annotations	**160**
Total Annotations	**4878**
Max. Annotations/investigation	**566**

This project has received funding from the European Union's Horizon 2020 r

Data archiving statistics since 15/8/2020 (EBS)

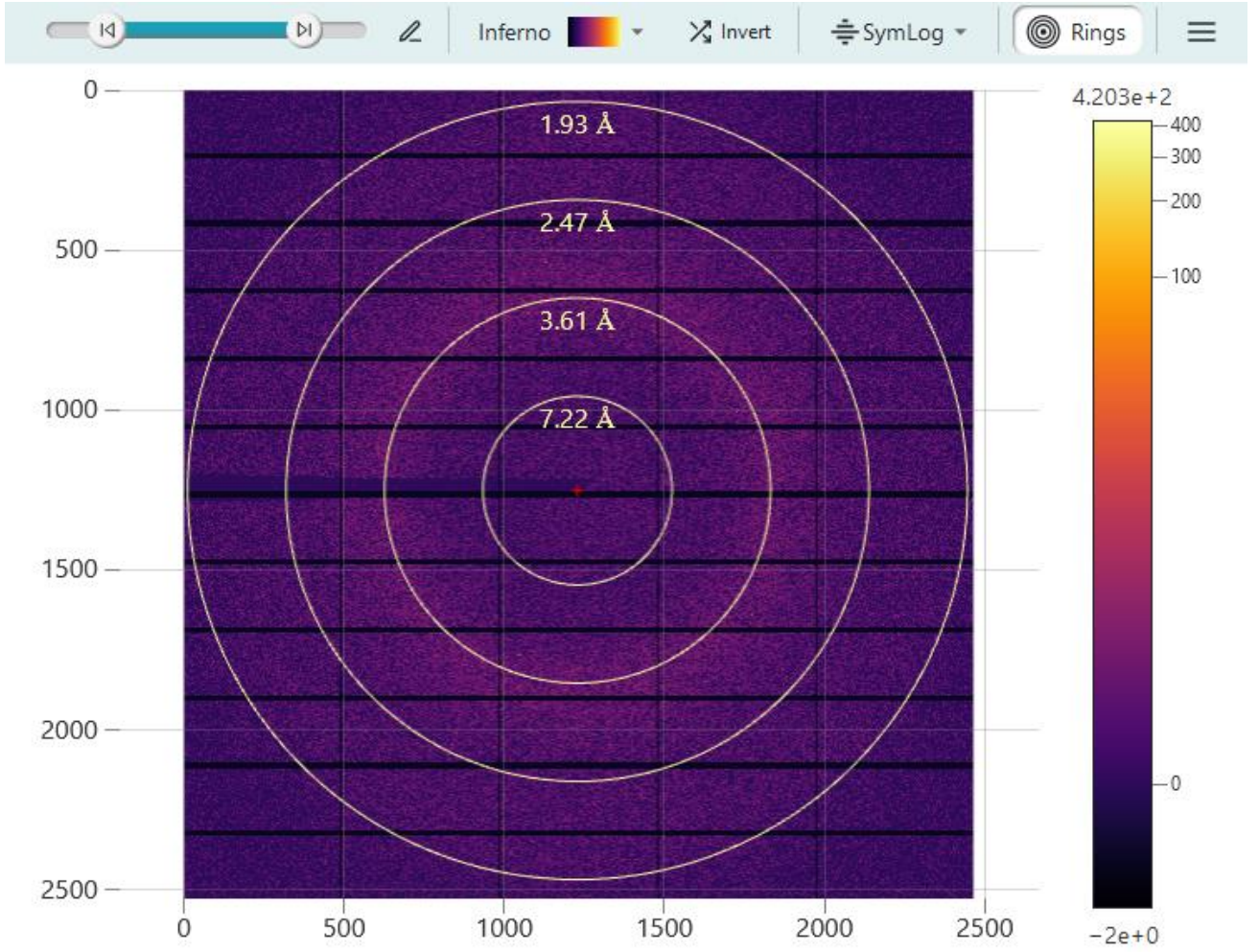
- Total of 3.3 PB
- CryoEM 22%
- Tomography 12%
- Other 63%
- Some beamlines are still missing



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

Data services planned for the future

1. Braggy diffraction image viewer on the web
2. Linking raw data to PDB
3. Remote analysis desktop (VISA)
4. Gold standard data format
5. Link e-logbook to LIMS
6. Analysis workflows
7. Improved metadata



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

Human Organ Atlas

- FAIR data i.e. can be re-used
- 50 Datasets already Open
- Opened immediately after processing (<18 months)

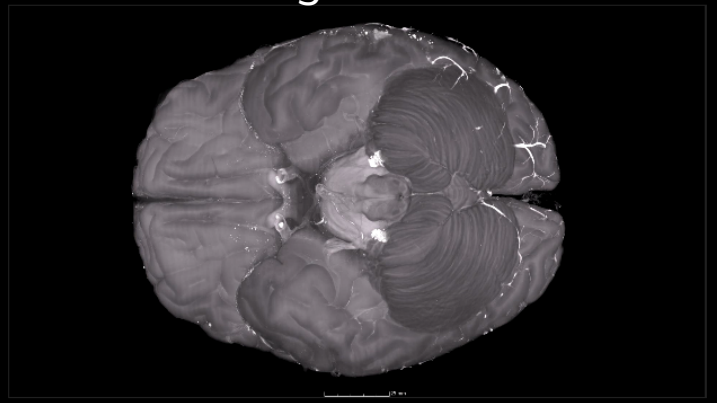
Goal is to create open data for an atlas of the entire human body at micron scale

This project has received funding from the European Union's Horizon 2020 re

Human Organ Atlas EXPLORE SEARCH HELP

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

Welcome to the Human Organ Atlas



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

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

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

Funding

This project has been made possible by funding from:

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

Collaborators

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

Reference

Walsh, C.L., Tafforeau, P., Wagner, W.L. *et al.* Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography. *Nat Methods* (2021). https://doi.org/10.1038/s41592-021-01317-x

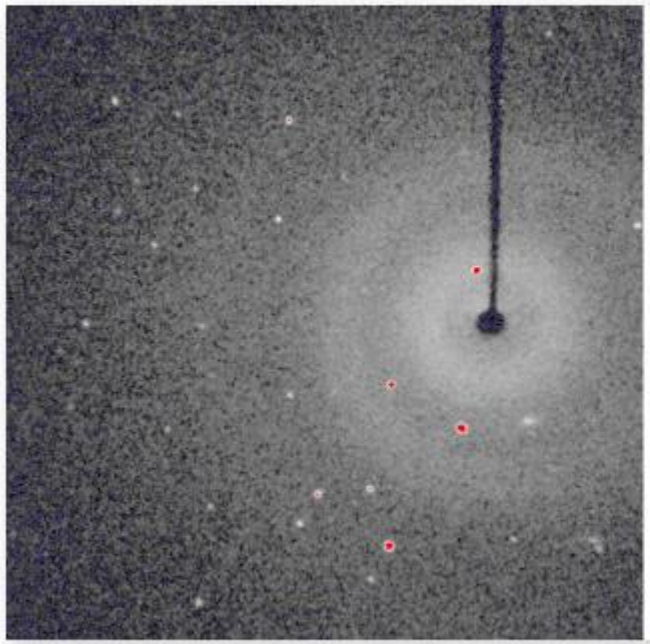
Aknowledgements

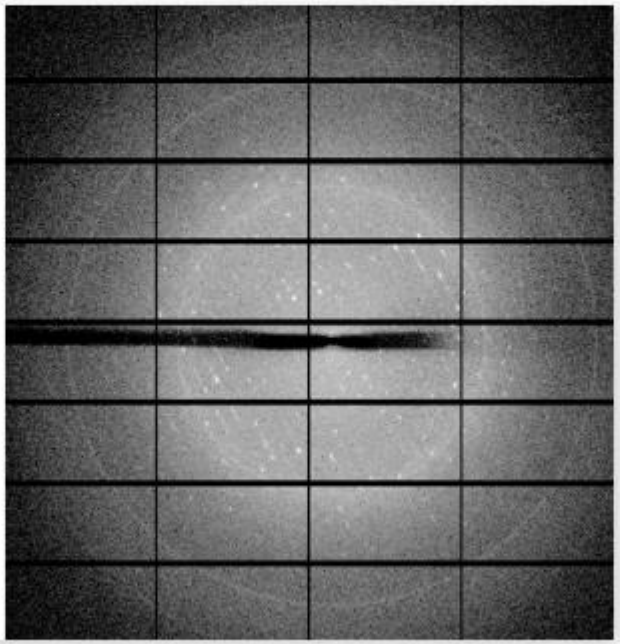
The development of this portal has been done as part of the PaNOSC project. PaNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852. The following people were involved in the development: Paul Tafforeau, Alejandro De Maria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

IUCrData to publish Raw Data Letters in 2022 – another motivation for Open Data with good metadata

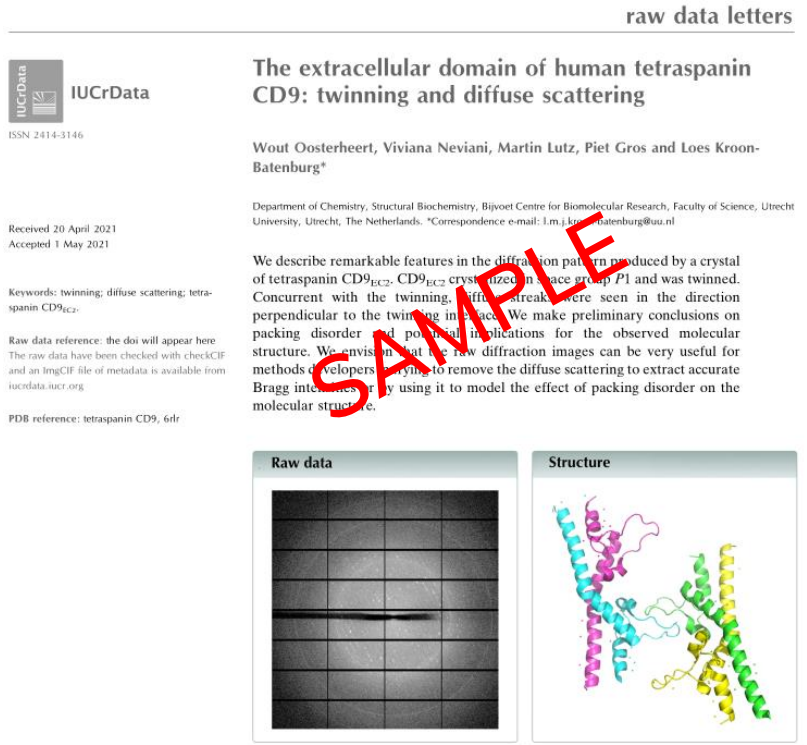
IUCrData to publish Raw Data Letters

IUCrData is launching a new section for authors to describe their unprocessed or `raw' diffraction images. The new section will publish short descriptions of crystallographic raw data sets in the biological, chemical or materials science fields and provide a persistent link to the location of the raw data.





Read more



raw data letters

IUCrData

ISSN 2414-3146

The extracellular domain of human tetraspanin CD9: twinning and diffuse scattering

Wout Oosterheert, Viviana Neviani, Martin Lutz, 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 $CD9_{EC2}$.

Raw data reference: the doi will appear here
The raw data have been checked with checkCIF and an ImgCIF file of metadata is available from iucrdata.iucr.org

PDB reference: tetraspanin CD9, 6rlr

We describe remarkable features in the diffraction pattern produced by a crystal of tetraspanin $CD9_{EC2}$. $CD9_{EC2}$ 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. We make preliminary conclusions on packing disorder and potential implications for the observed molecular structure. We envision that the raw diffraction images can be very useful for methods developers applying 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.

SAMPLE

Raw data

Structure

https://iucrdata.iucr.org/x/

panosc

What you need to do to CARE for your DATA – 10 rules

- Rule 1. Love Your Data, and Help Others Love It, Too
- Rule 2. Share Your Data Online, with a Permanent Identifier
- Rule 3. Conduct Science with a Particular Level of Reuse in Mind
- Rule 4. Publish Workflow as Context
- Rule 5. Link Your Data to Your Publications as Often as Possible
- Rule 6. Publish Your Code (Even the Small Bits)
- Rule 7. State How You Want to Get Credit
- Rule 8. Foster and Use Data Repositories
- Rule 9. Reward Colleagues Who Share Their Data Properly
- Rule 10. Be a Booster for Data Science

https://doi.org/10.1371/journal.pcbi.1003542

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

Ten Simple Rules for the Care and Feeding of Scientific Data

Alyssa Goodman[1], Alberto Pepe[1]*, Alexander W. Blocker[1], Christine L. Borgman[2], Kyle Cranmer[3], Merce Crosas[1], Rosanne Di Stefano[1], Yolanda Gil[4], Paul Groth[5], Margaret Hedstrom[6], David W. Hogg[3], Vinay Kashyap[1], Ashish Mahabal[7], Aneta Siemiginowska[1], Aleksandra Slavkovic[8]

Conclusion

Adopting best practices for Open Data has many benefits but especially helping MAKE BETTER SCIENCE

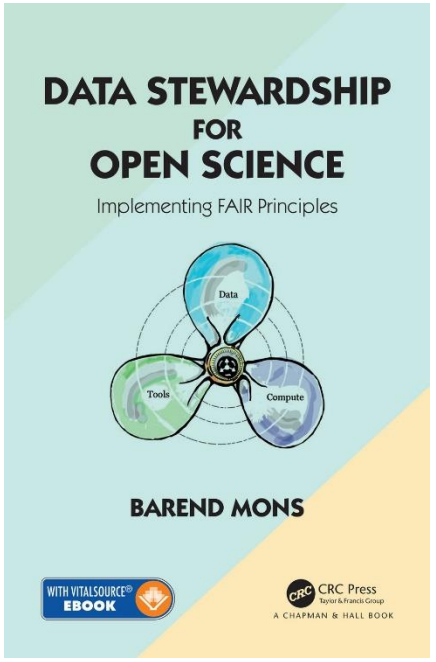
- Some of the elements for generating Open Data:
  - Data Management Plan, Data Policy, Data Outputs, File types, File Formats, Software, Workflows, e-Logbooks, notebooks, Data Storage, Data Archiving, Data DOI
- Apply the **FAIR principles** – ask yourself if you or someone else will be able to use or understand your data
- Make your Data count – **cite your data DOIs** (raw and processed data)
- Open Data is not a **goal** on its own, **reproducibility** and **data re-use** are

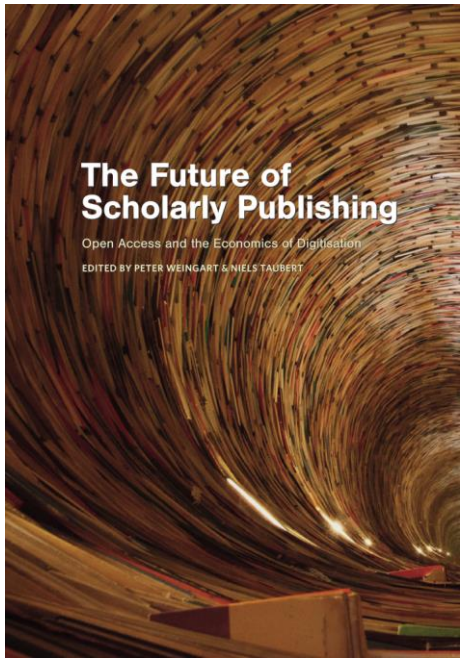
- The Open Data movement has a lot to offer scientists and science

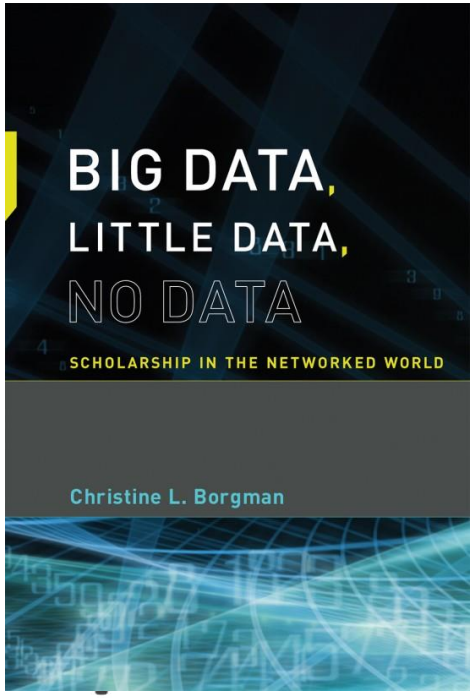
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Learning more about Research Data Management

- 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.







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

Acknowledgements

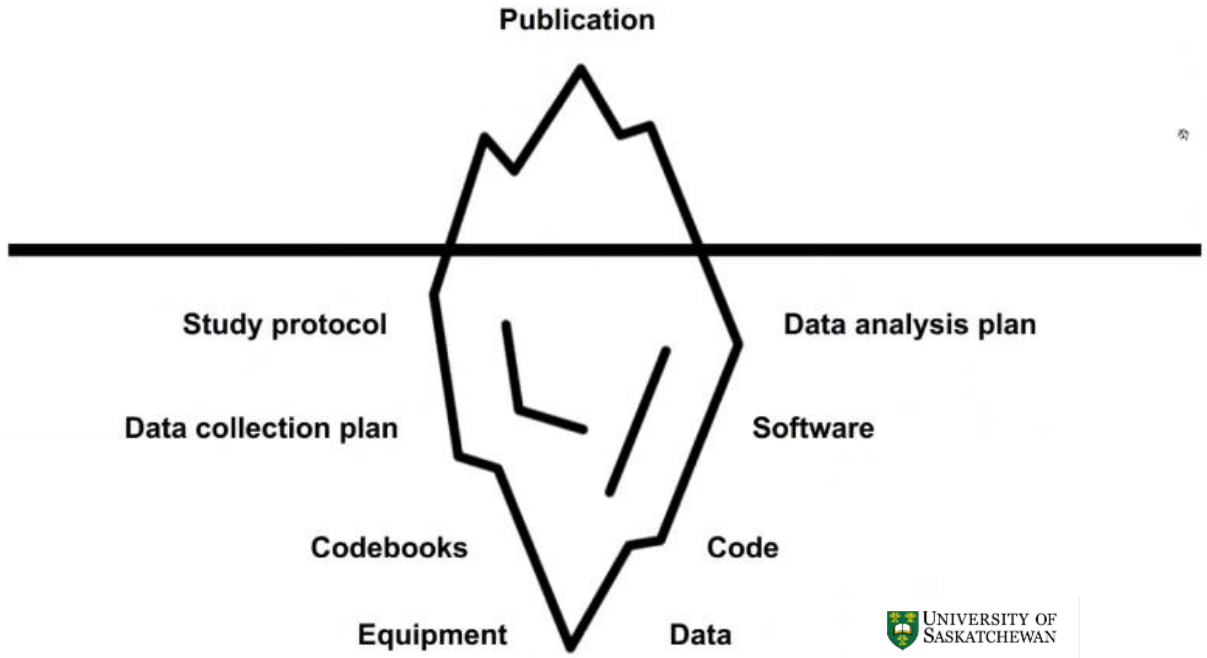
- ESRF data stewards - Alex + Marjolaine + colleagues
- PaNOSC, ExPaNDS, EOSC H2020 projects
- IUCr Committee on Data
- Research Data Alliance
- ICAT collaboration
- Wikipedia
- Internet

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

panosc

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