Europe's Photon and Neutron Open Science Cloud for Raw and Processed Data: Aims and Achievements to Date

Andy Götz (ESRF, PaNOSC coordinator)

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Talk outline

1. Acknowledgements
2. Managing Photon and Neutron data
3. The Photon and Neutron Open Science Cloud
4. The carbon footprint of raw data
5. How does this benefit Users i.e you?

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Acknowledgement of Country

We acknowledge the Traditional Owners of the lands and waters throughout Australia, and pay respect to the Elders past, present and emerging. We recognise the importance of connection to culture, land, kinship and community to the health and wellbeing of Aboriginal & Torres Strait Islander families. We acknowledge the cultural practices and traditions still carried out today and being passed down to future generations.

I'd like to begin by acknowledging the Traditional Owners of the land on which we meet today. I would also like to pay my respects to Elders past and present. I acknowledge the first custodians of this land who gathered an amazing body of knowledge over tens of thousands of years.

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

I acknowledge the **Committee on Data** of the IUCr. I thank the members of **CommDat** who have played a key role in motivating the work to make **FAIR data reality for the Photon and Neutron** community (and this talk).
A special thanks to **John Helliwell**
who led CommDat since the beginning
in 2016 up until now.

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Committee on Data (CommDat)

Membership

- J.R. Helliwell (Chair, UK; john.helliwell@manchester.ac.uk)
- B. McMahon (Secretary, UK)
- H.J. Bernstein (USA)
- A. Brink (South Africa)
- I. Bruno (UK)
- S. Coles (UK)
- K. Dziubek (Poland)
- A. Goetz (France)
- S. Kabekkodu (USA)
- L.M.J. Kroon-Batenburg (The Netherlands)
- G. Kurisu (Japan)
- W. Minor (USA)
- S. Storm (Germany)
- L. Van Meervelt (Belgium)
- J. Hester (Australia) (COMCIFS liaison)

Consultants

- S. Androulakis (Australia)
- M.P. Blakeley (France)
- G. Bricogne (UK)
- S. Grazulis (Lithuania)
- B. Matthews (UK)
- A. Sarjeant (USA)
- D. Szebenyi (USA)
- E.F. Weckert (Germany)
- J. Trewhella (Australia)

https://www.iucr.org/iucr/governance/advisory-committees/committee-on-data

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Workshop on *Raw diffraction data reuse: the good, the bad and the challenging*

Organized by

Loes Kroon-Batenburg (Netherlands), Selina Storm (Germany), John Helliwell (UK) and Brian McMahon (UK) for the IUCr Committee on Data

https://www.iucr.org/resources/data/commdat/melbourne-workshop

CIF Dictionary Workshop

Session 1

TODAY - Friday
at 1:10 pm – 3:30 pm

Room: 216	Room: 220	Room: 207
Chairs: Loes Kroon-Batenburg & James Hester	**Chairs: Nicholas Schwarz & Andrew Gotz**	**Chairs: Brinda Vallat & John Helliwell**
Session: A118 Raw Diffraction Data Reuse: Warts and All	**Session: A109** Data-Driven Science: Current Status and Outlook	**Session: A014** Databases and Data Management

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Next generation of photon sources

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

http://dx.doi.org/10.1016/j.elspec.2013.12.007

Next generation of data sources

1. Current photon sources produce petabytes of data / year

2. Next generation will produce 10-100s of petabytes / year

3. Users + facilities are challenged by the data volumes

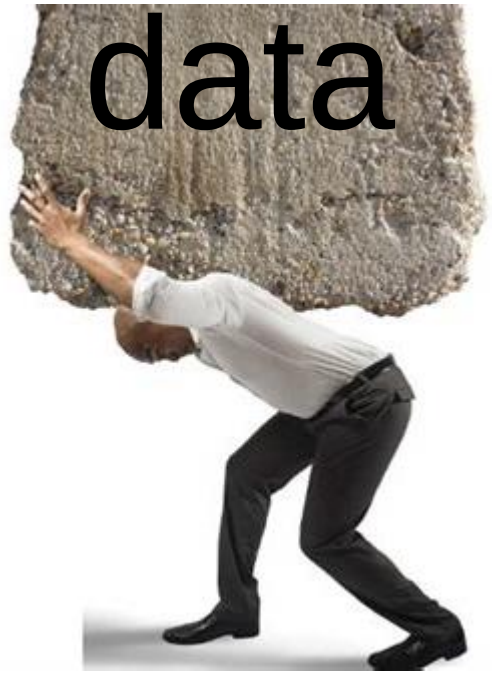
4. Reminder: 1 petabyte = 32 year long MP3
   Avatar 1 = 1 petabyte,
   Avatar 2 = 18 petabytes

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Users + Facilities feel the weight of data

Users

- Huge data volumes
  Tera → Petabytes
- Sample metadata
- Raw data quality
- Data processing
- Data exporting
- FAIR data



Facilities

- Huge data volumes
  Peta → Exabytes
- Data acquisition
- Metadata collection
- Data curation
- Data archiving
- FAIR data

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

This talk is about reducing reduce the burden on Users to manage and make data FAIR

Users

- **Data scientists**
- Provide metadata
- Keep logbooks
- Data management checklist / plans
- Publish data
- Cite data

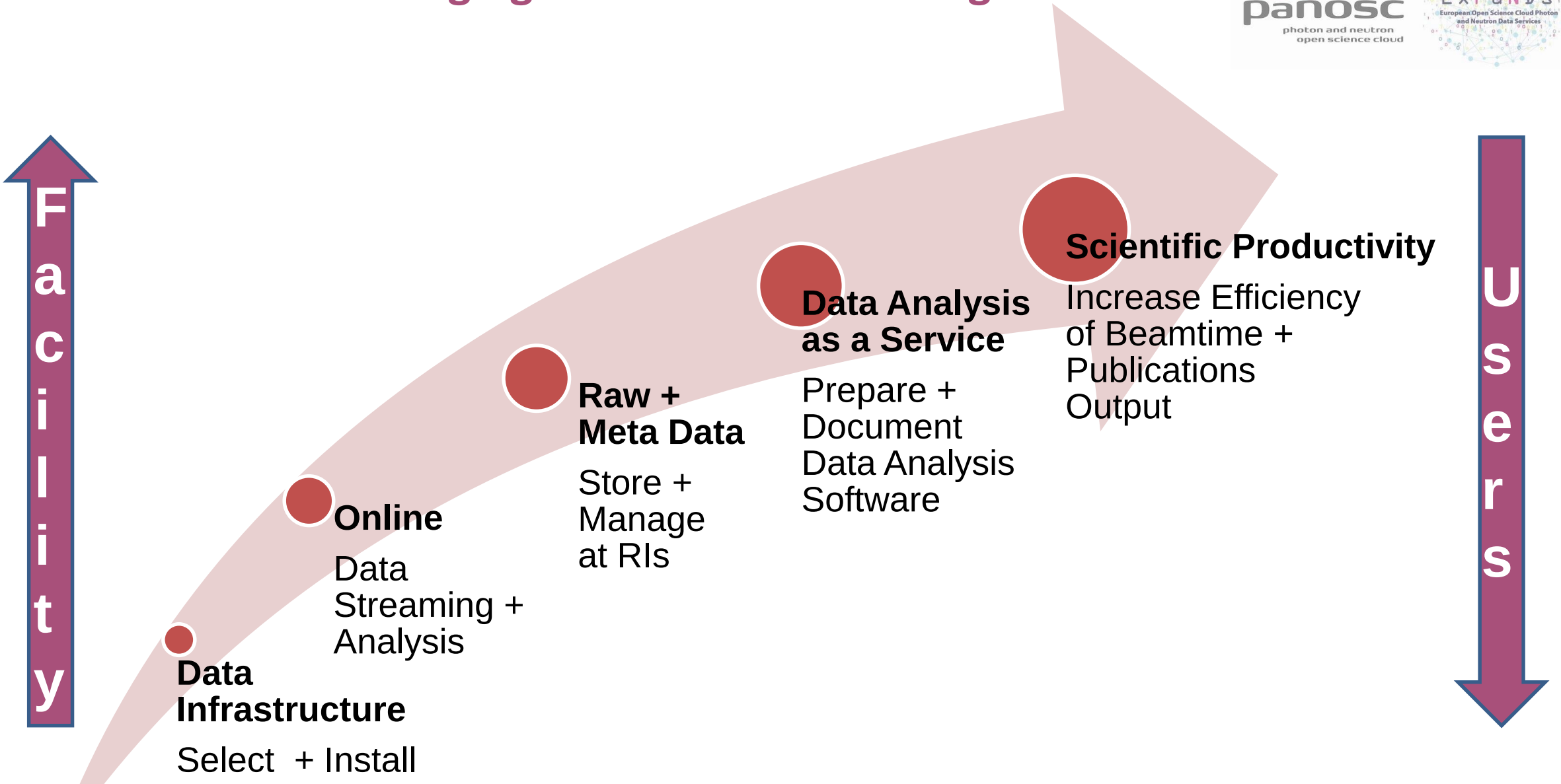
data

Facilities

- **Data managers**
- Curate raw data
- Online processing
- Metadata catalogues
- Data repositories
- Remote analysis
- Data portals

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Facilities and Users converging needs for data management



PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

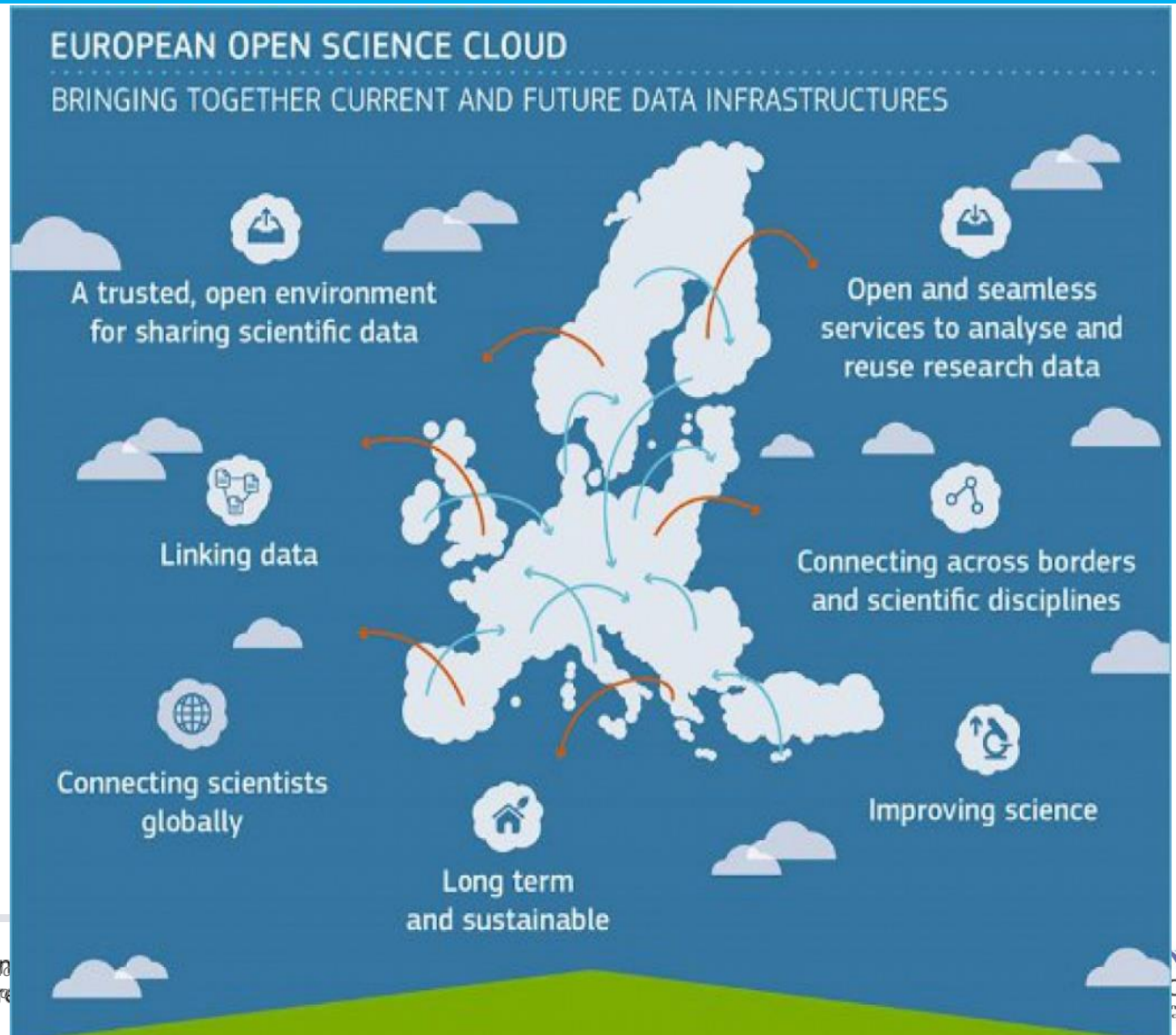
How can the EC help solve the data issues?

1. **European Commission funds common activities**
2. One of the activities of **European Data Strategy** is the **European Open Science Cloud (EOSC)**
3. **Photon and Neutron sources** in **Europe** are seen as part of the **EOSC** as **data providers**

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

European Open Science Cloud
- The vision -

- Bridging todays fragmented and ad-hoc solutions; towards a **federation of data infrastructures**
- **FAIR data and services** for data storage, management, analysis and re-use **across borders and disciplines**
- Added value for **data-driven science**, reproducible science, interdisciplinary research, digital innovation (EU DSM)



PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, re

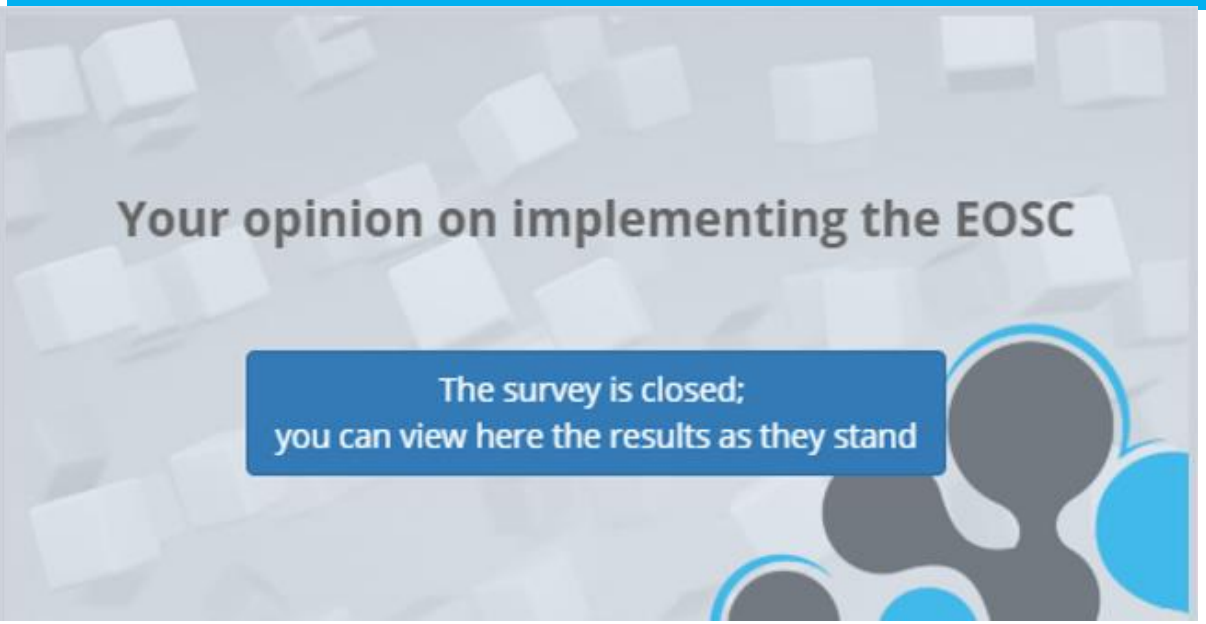
Service dimension of the EOSC

The EOSC will provide ~~two million EU~~ researchers with:

- ✓ A catalogue of European research data funded with public money;
- ✓ A catalogue of services to re-use these data;
- ✓ Tools to make their own data open and FAIR;
- ✓ Advanced tools to merge and analyse the data in a secure environment;
- ✓ A simple access gateway to these services (EOSC Portal).

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.
The ExPaNDS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.
The PaNOSC project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852.

Not a cloud from Brussels, but a research Data Commons driven by the stakeholders



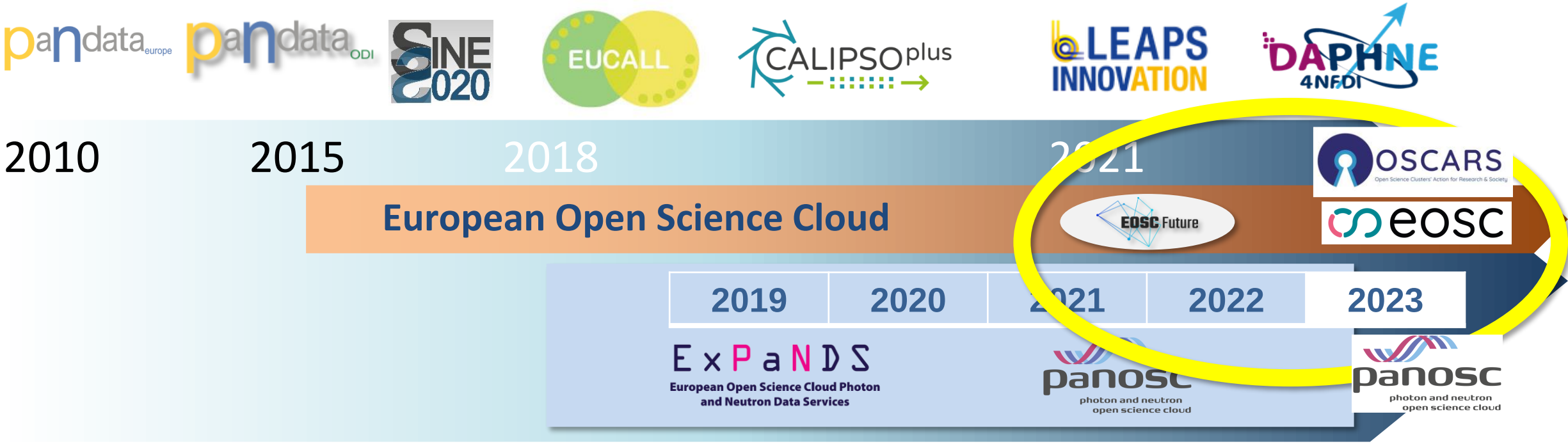
EOSC declaration 2017

Vienna Declaration 2018

EOSC Association 2020

Together, we can create a European Research Data Commons that will not only benefit Europe but help tackle the societal challenges worldwide. EOSC-A

European support for the PaN community



Policies	Common data policy	**FAIR** data policy	Data Management Plans
Analysis	Software Catalogue	Remote analysis	Jupyter
AAI	UmbrellaID	AARC Blueprint	eduTeams
Training	e-neutron	Training platform	

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Two EOSC projects for PaN : PaNOSC and ExPaNDS

PaNOSC

ExPaNDS

Photon (LEAPS)

EGI

eli delivery consortium

CERIC Central European Research Infrastructure Consortium

Neutron (LENS)

DESY
CESNET
STFC

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Cross-Domain Research Infrastructure Collaboration for Open Science

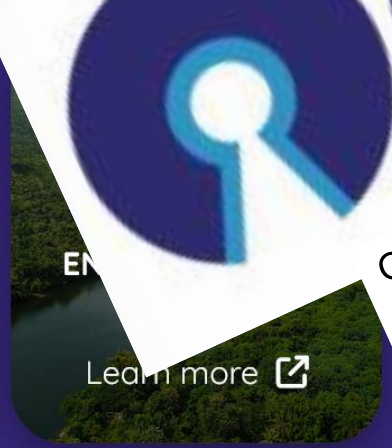
SCIENCE-CLUSTERS.eu
Research Infrastructures for Open Science

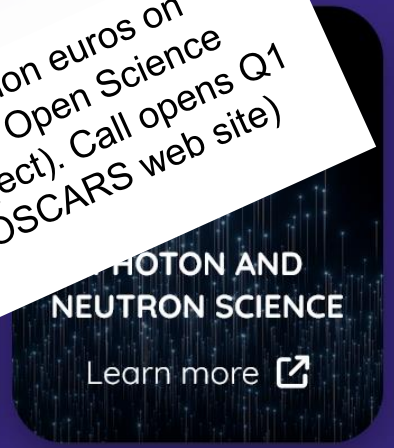
Research Infrastructures and Co…es

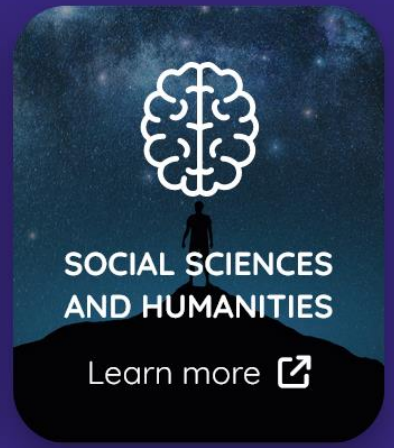
The science clusters have grown out of five collaborative projects f… to link ESFRI and other world-class Research Infrastructures (RIs) to the European C… eloped by the clusters and other outcomes of the projects are cornerstones of t… y communities and multidisciplinary initiatives with harmonised m… uster unites multiple RIs in their specific scientific domain.

OSCARS
Open Science Clusters' Action for Research & Society

OSCARS will spend 18 million euros on projects on FAIR data and Open Science (150k – 250k euros / project). Call opens Q1 2024 (see PaNOSC or OSCARS web site)







PaNOSC + ExPaNDS - Main Achievements

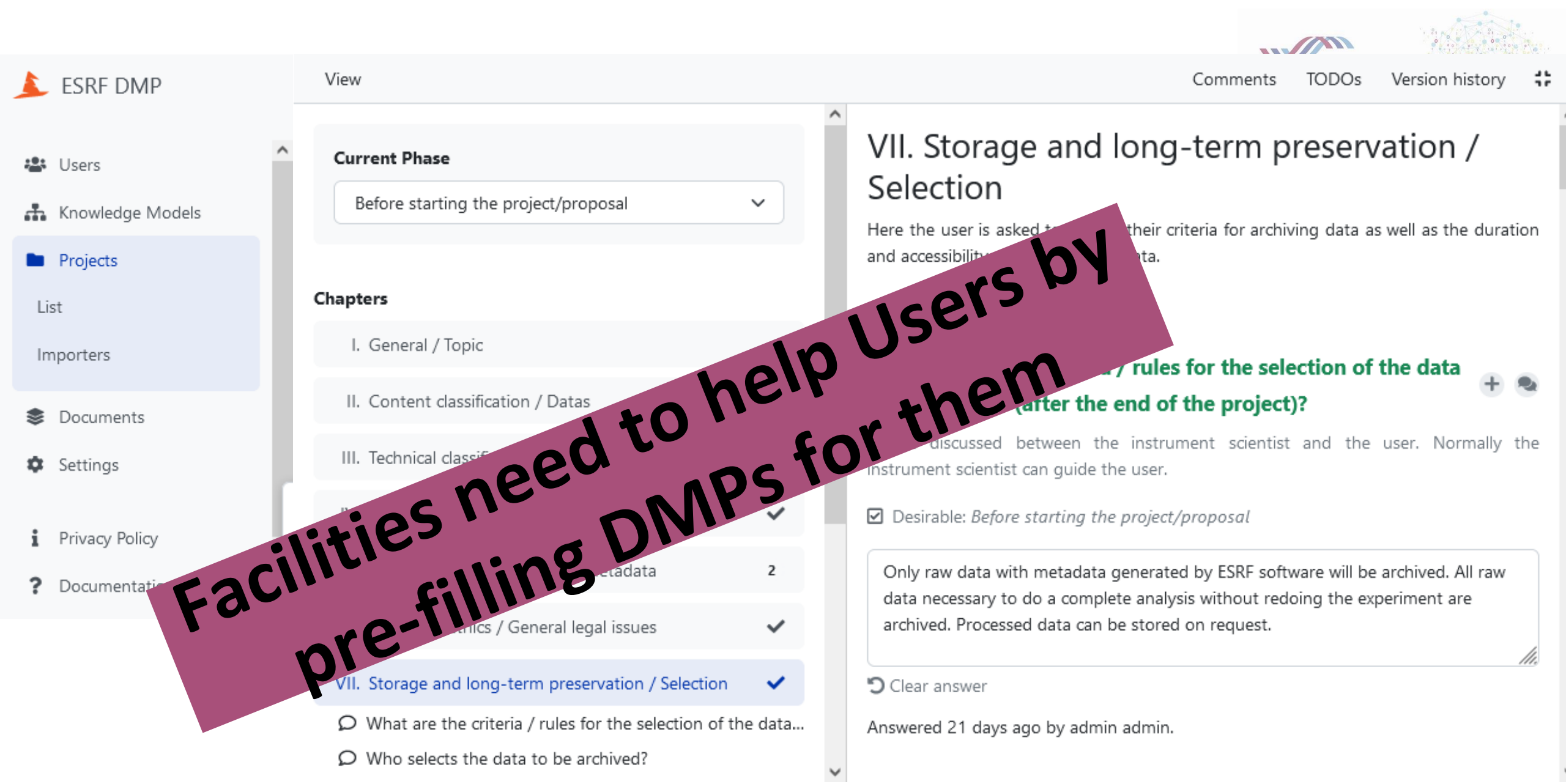
1. **FAIR data policy** and **DMPs**
2. Standardised metadata (**Nexus/HDF5**)
3. Guidelines for **FAIR data** self-evaluation
4. **Federated search API** for PaN data catalogues
5. **Open Data portal** for searching + downloading data
6. Community **AAI UmbrellaId**
7. **JupyterLab notebooks** and **Nexus/HDF5** files visualisation
8. **Remote data analysis** with VISA + data analysis pipelines
9. **Simulation** software for simulating experimental data (ViNYL)
10. **PaN-learning** platform (pan-learning.org)

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Active Data Management Plans

1. **ExPaNDS and PaNOSC have adopted active DMPs**
2. **Active DMPs are updated at different phases of the project**
3. **ESS and ESRF have chosen to use DS Wizard developed by Elixir**
4. **Example of implementation @ ESRF**
    1. Automatically generates a DMP automatically for every proposal
    2. 50 out of 82 questions are automatically filled in from DP/User/Data Portals
    3. DMPs offer a structured way to communicate information
    4. Users can use the DMP for satisfy funders requirements
    5. Next step is to use the DMPs to ensure users can manage their **data**

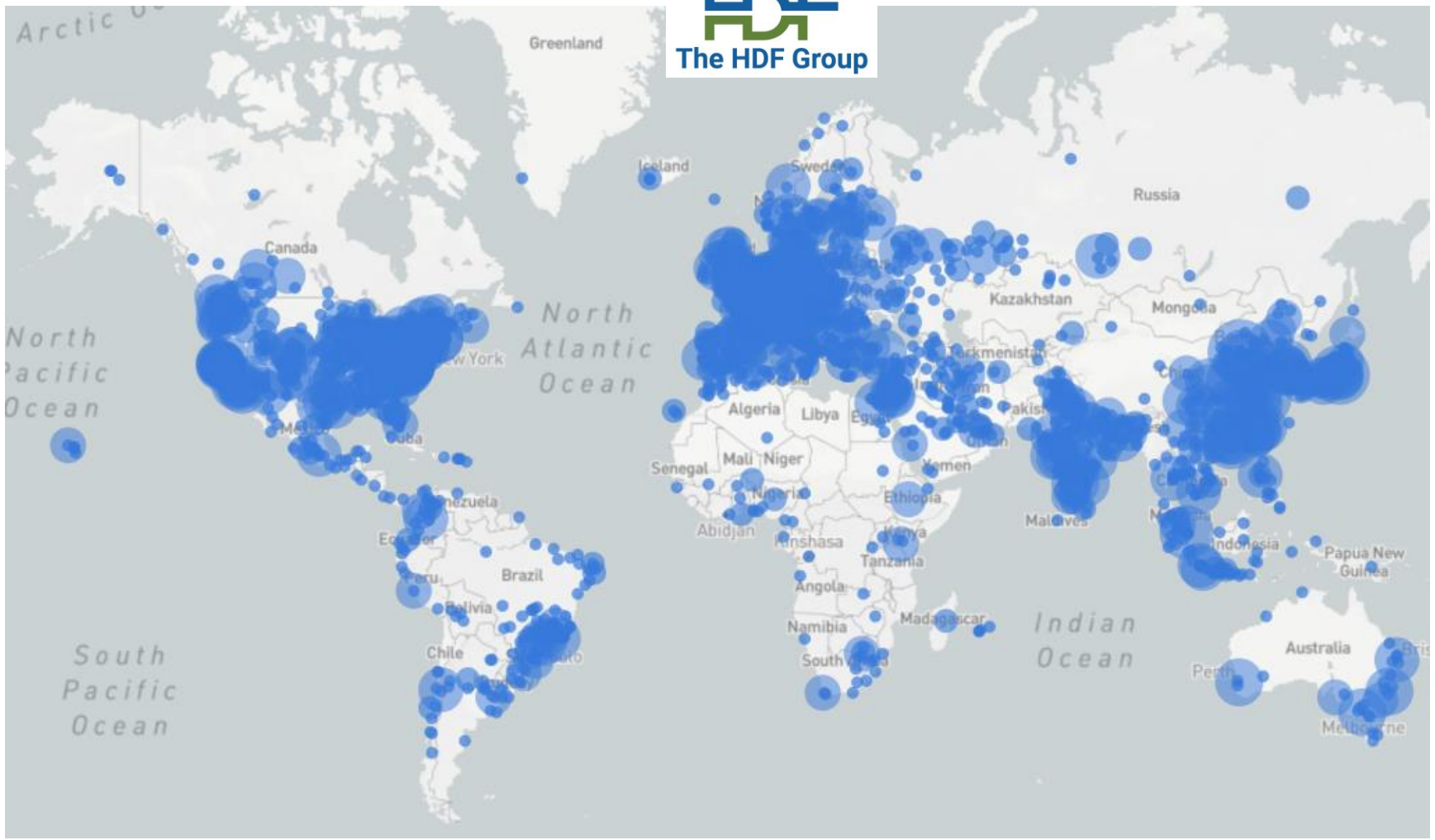
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.



Facilities need to help Users by pre-filling DMPs for them

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

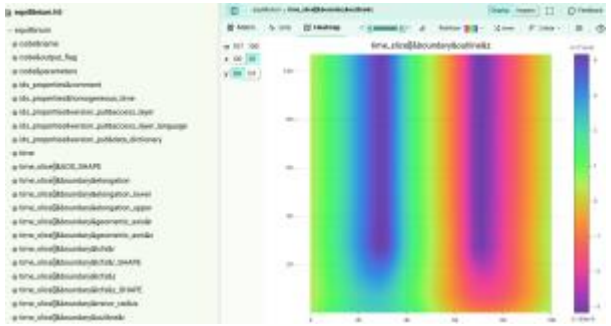
HDF5 is needed to acquire, process and store large datasets and has been adopted as de facto standard across most facilities (with NeXus conventions for metadata)



New Features

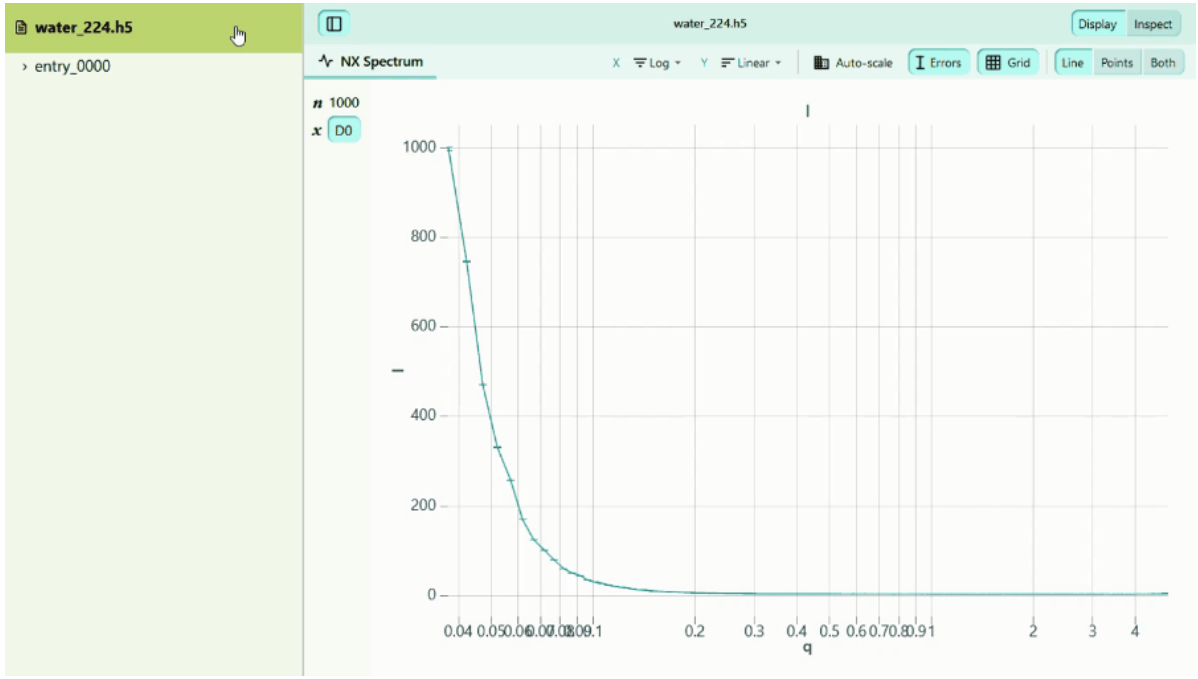
- Multi-dataset I/O
- Selection and vector I/O
- Subfiling VFD
- Multi-Threaded Concurrency (WIP)
- Sparse data storage (WIP)

PaNOSC developed H5Web to make HDF5 even easier

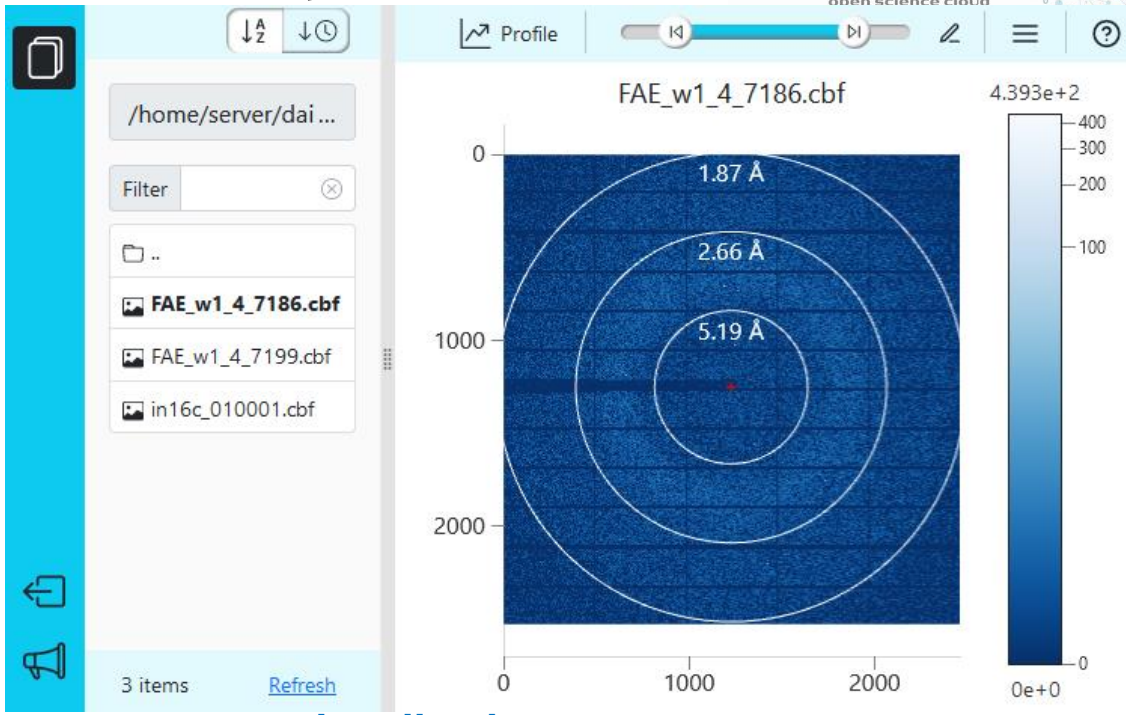


PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

H5Web Visualization Ecosystem



Generic HDF5 file viewer



Visualization components

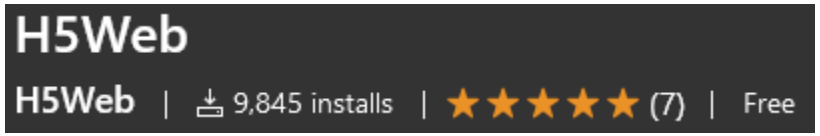
- Integrated into **data portals + web apps (ESRF, EuXFEL, DLS, AS, …)**, for viewing files generated during experiments
- Available as **JupyterLab** and **VS Code extensions**, and as stand-alone web service, **myHDF5**, for viewing local and remote HDF5 files

https://github.com/silx-kit/h5web

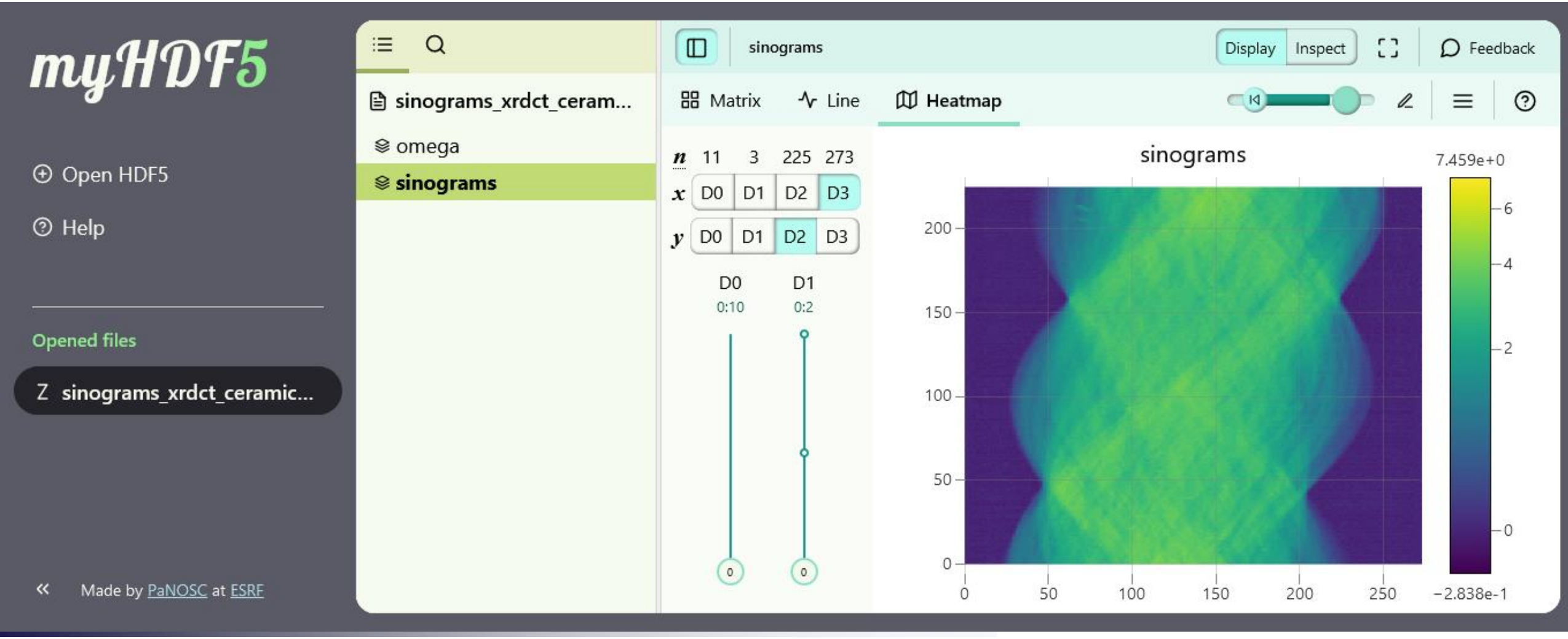
Used in various web applications at ESRF including:

- **Braggy**, diffraction image viewer (screenshot above)
- **Daiquiri**, beamline control and data acquisition software

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.



HDF5 file viewing on the web made easy
https://myhdf5.hdfgroup.org/



PaNO https://myhdf5.hdfgroup.org/view?url=https://zenodo.org/record/4751241/files/sinograms_xrdct_ceramic.h5?download=1
and innovation programme under grant agreements 823852 and 857641, respectively.

Data compression is essential to deal with the “too much” data being produced by detectors – register for the workshop here https://indico.desy.de/event/39343/

2023 European HDF User Group (HUG) plugins and data compression summit

Sep 19 – 21, 2023
DESY
Europe/Berlin timezone

Enter your search term

- Overview
- Call for Abstracts
- Registration
- Venue
- Accommodation
- Organising committee and sponsorship

Local organiser

david.pennicard@desy.de

On September 19-21, the 2023 European HDF5 Users Group (HUG) summit is going to be held both in-person and virtually at DESY in Hamburg, Germany.

The HDF5 library and toolset have been widely used to simplify management and access to scientific and engineering data with ubiquitous data solutions. This meeting will have a focus on plugins and da compression, which become an increasingly important topic as data volumes grow, and is being hoste in collaboration with the LEAPS-INNOV EU project: https://www.leaps-innov.eu/

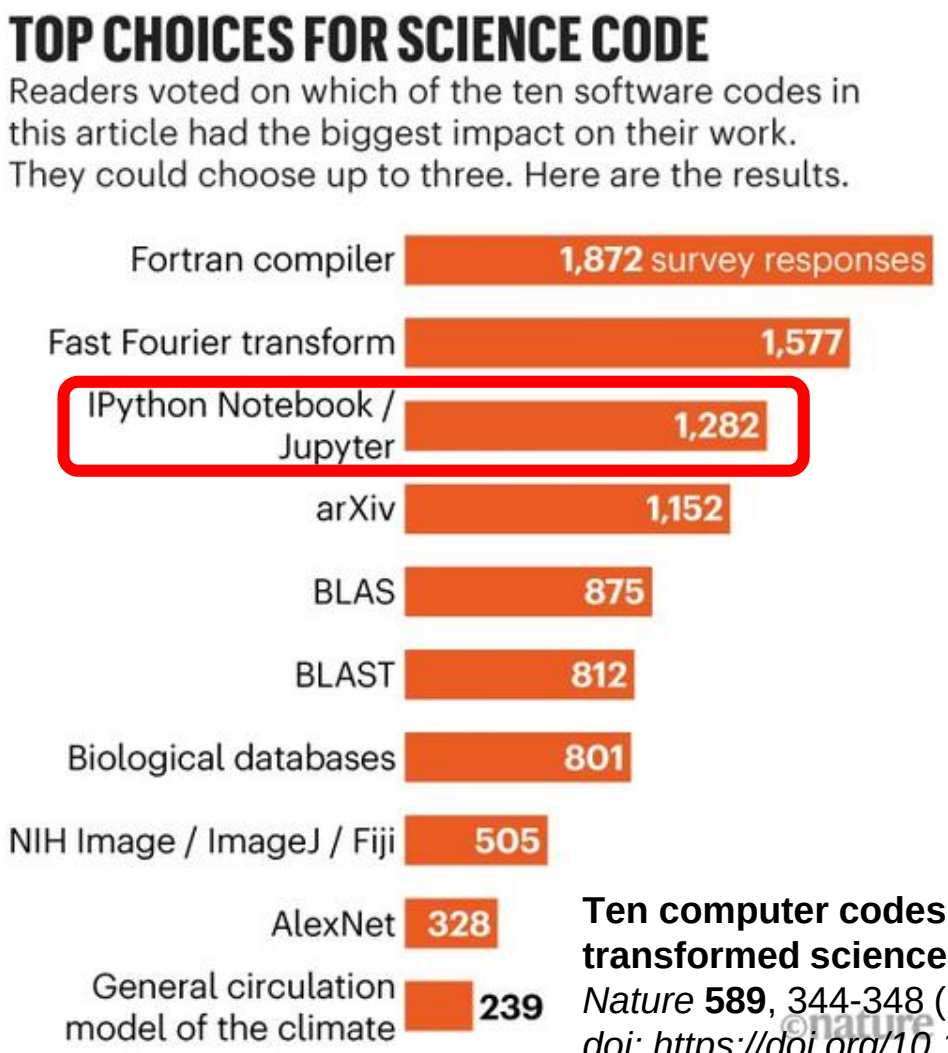
The HDF Group will give an update on the latest developments within HDF5 and present the roadmap the future. We will have invited talks and contributed presentations from HDF5 users and developers, discussions on future requirements and developments, and events such as tutorials.

Please use the links in the side panel to register and/or submit your presentation proposal. As well as taking abstracts for presentations on your work, we are keen to get feedback and discussion on next steps with HDF. So, if there are topics or problems you wish to discuss, you are also welcome to let us know via the abstract system, and we could then arrange a discussion slot.

Topics of interest include, but are not limited to:

Support for Jupyter notebooks

- **Jupyter service** now available at all PaNOSC and most ExPaNDS sites
- **Jupyter on Slurm** service developed: https://github.com/silx-kit/jupyterhub_moss/
- **H5Web** Jupyterlab plugin developed
- **VISA** provides Jupyter service
- **PaN e-learning** platform provides Jupyter as a service
- **PaNOSC summer school** trained participants to program in Python using Jupyter
- **EGI** provided Jupyter and Binder as a service



Ten computer codes that transformed science
Nature **589**, 344-348 (2021)
doi: https://doi.org/10.1038/d41586-021-00075-2

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Open Science with Jupyter notebooks

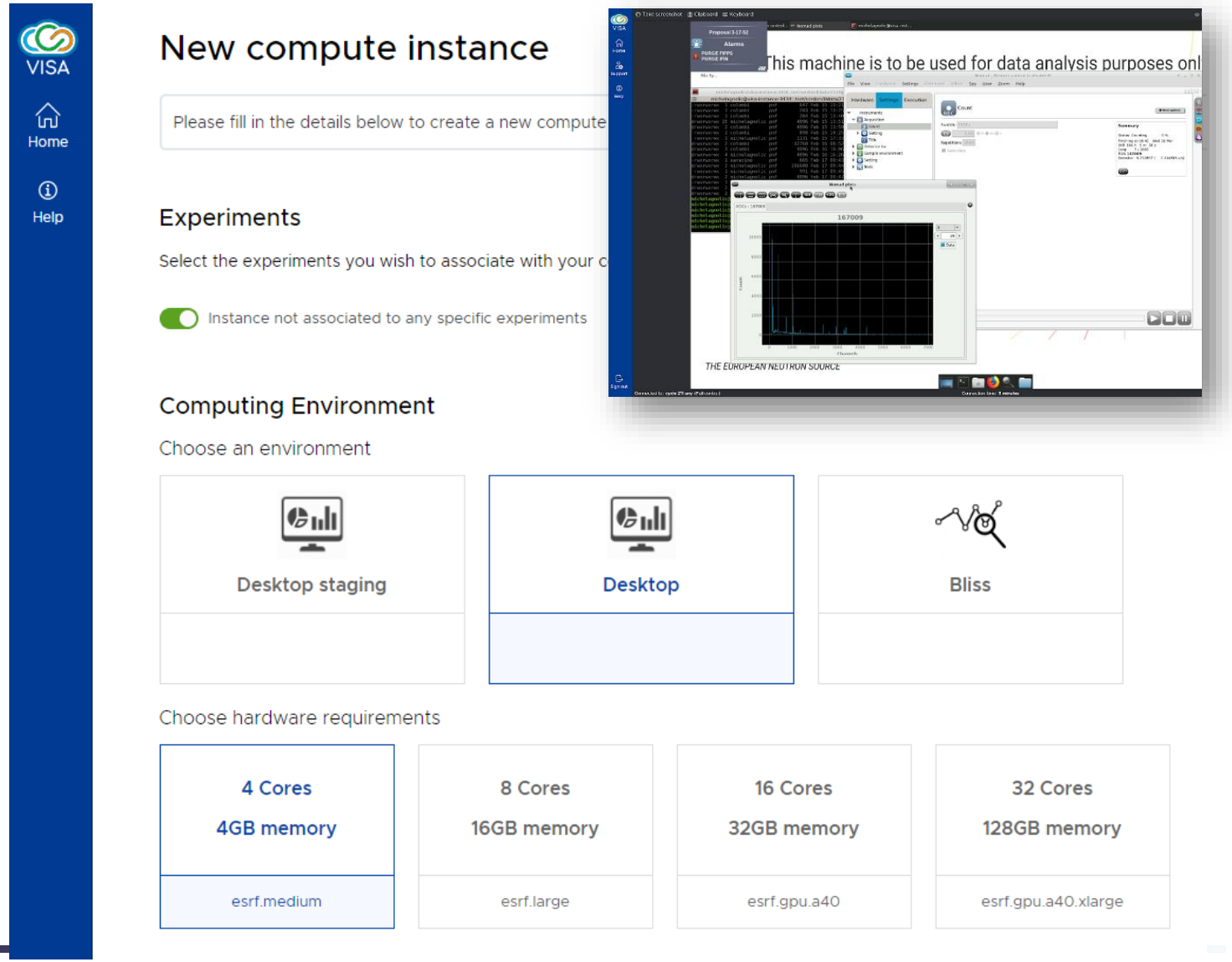
- Notebooks … n in one document
- If used app …
- For exampl …
- Notebooks … e-usable
- Currently, I … work of others, before they …

PaNOSC provides training on making reproducible publications for FAIR data

https://youtu.be/vStbMUDI_jU

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

VISA - Remote Data Processing/analyses

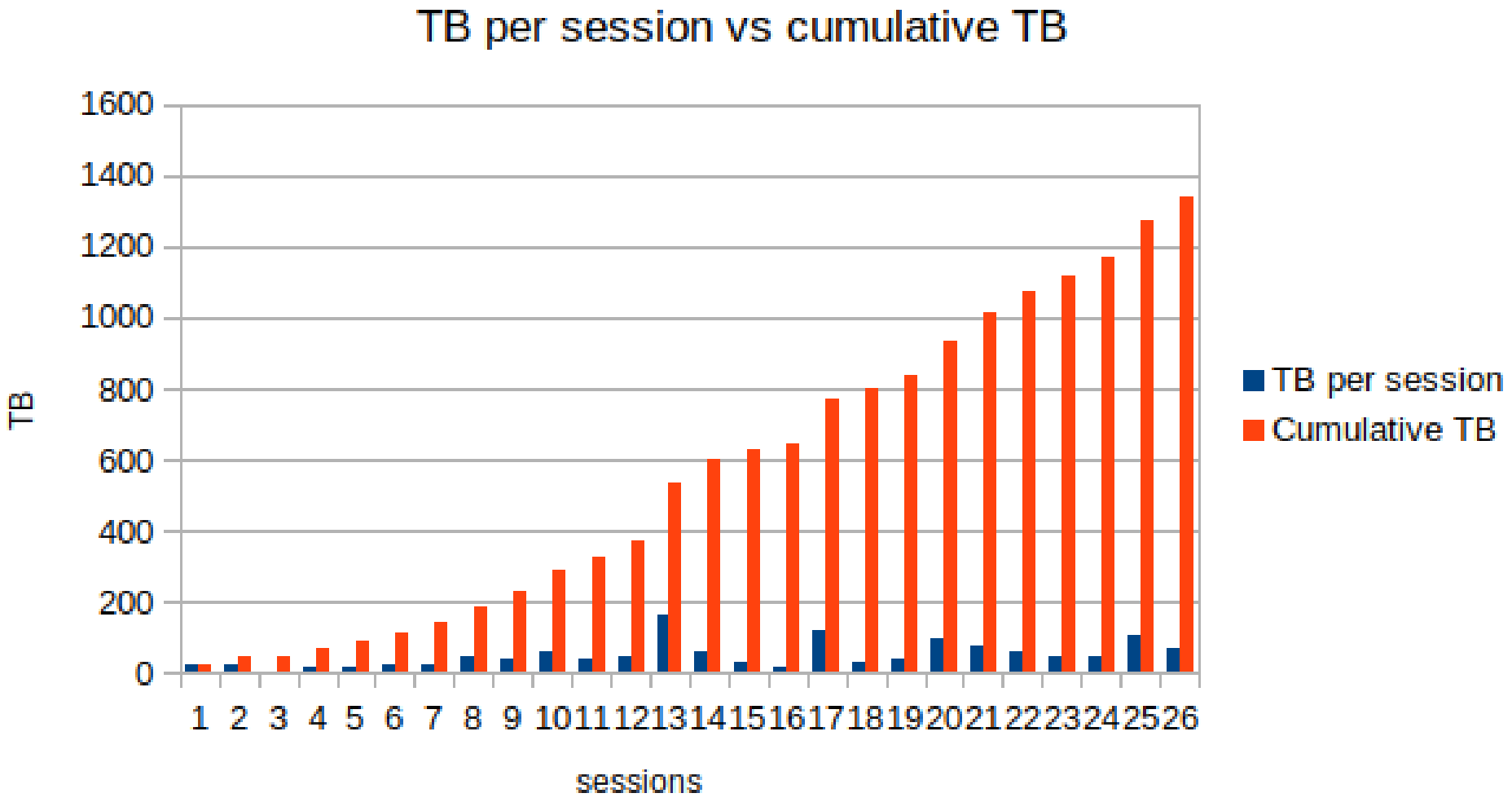


- Infrastructure for remote data processing / analysis
- Users dedicated VM
- Access to data
- Access to Provisioning of scientific SW using CVMFS and Containers
- Access to the GPUs, HPC cluster
- Infrastructure based on OpenStack
- Development led by ILL in the scope of the PaNOSC project

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

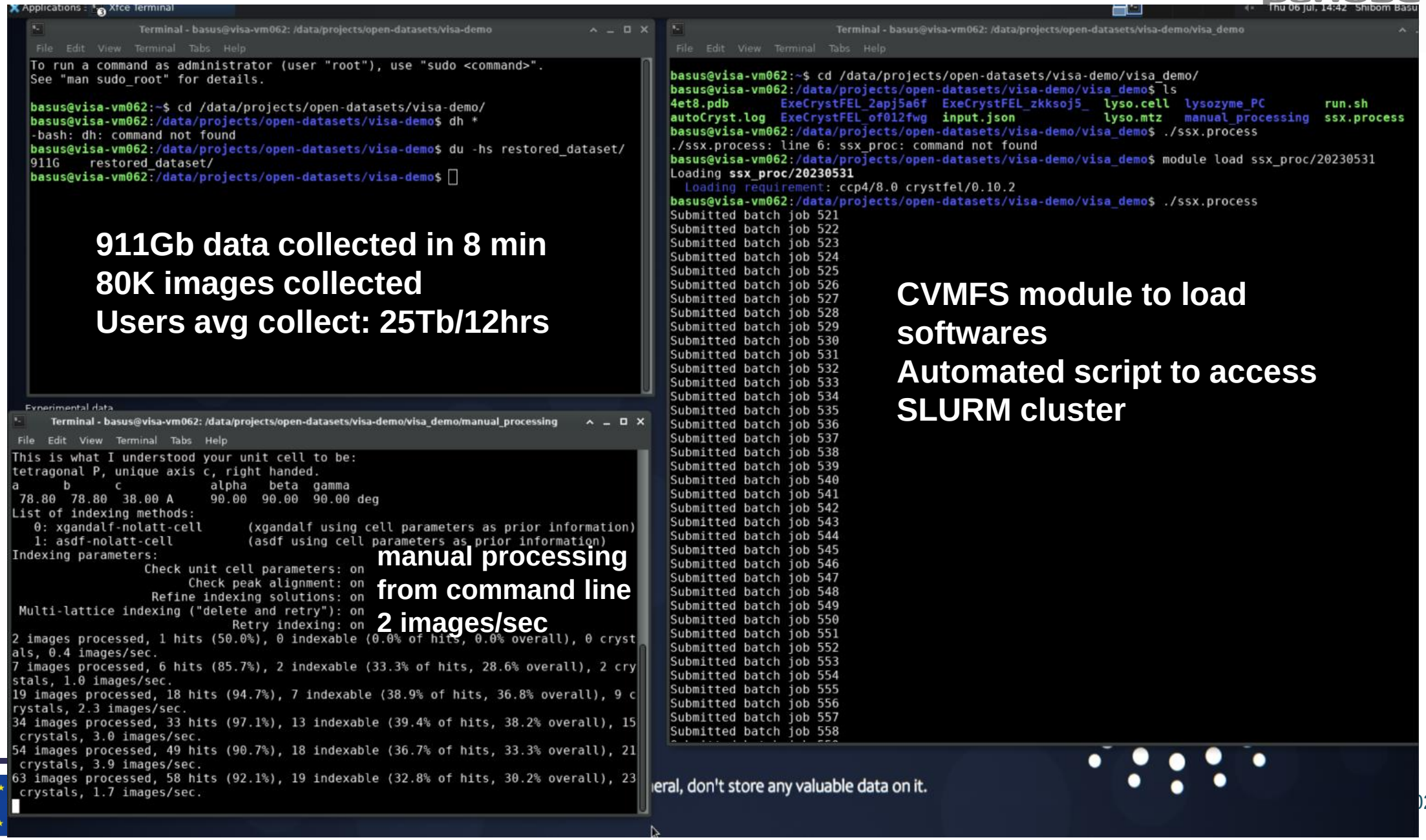
The ExPaNDS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.
The PaNOSC project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852.

ESRF-EBS serial crystallography beamline (ID29) produced > 1.2 PB in first 6 months

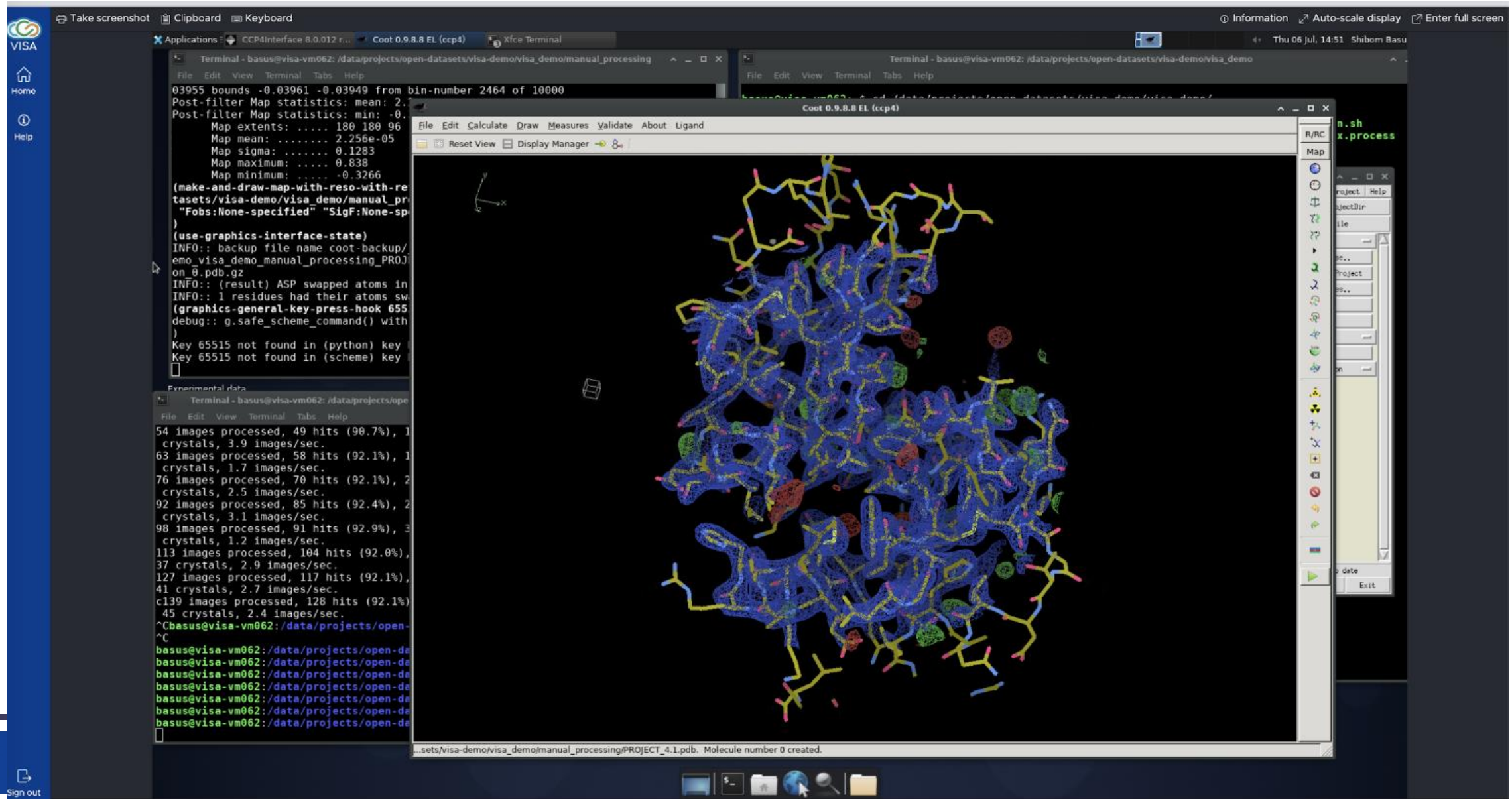


PaNOSC a
and innovation programme under grant agreements 823852 and 857641, respectively.

VISA for ID29 SSX - demo dataset and processing



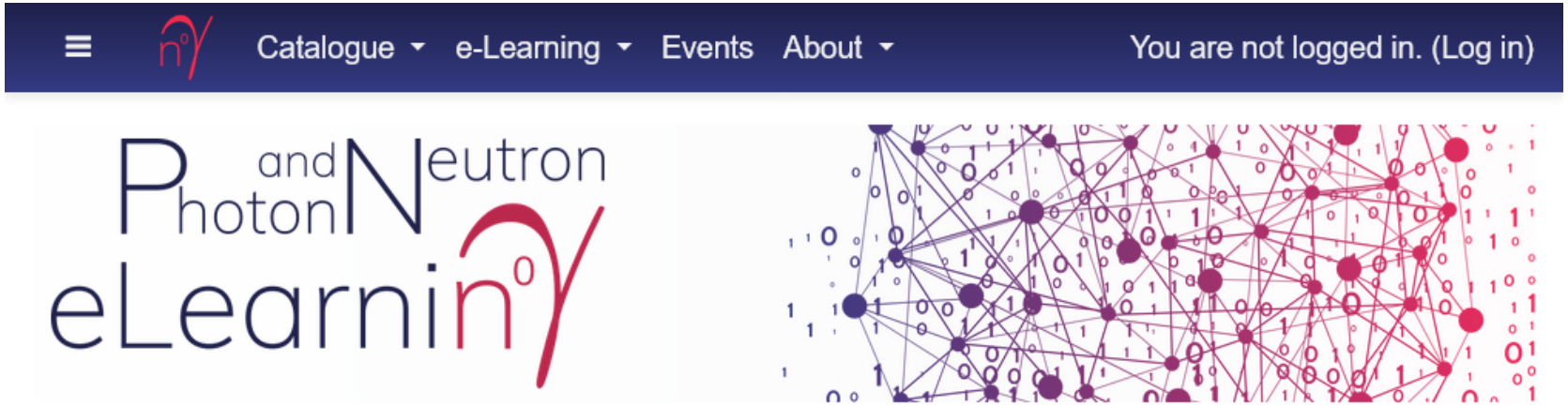
VISA for ID29 SSX – graphical viewers performance acceptable in a VISA virtual machine via browser



Open Training - courses on Pan-learning.org

Achieving100% Open Educational Resources:

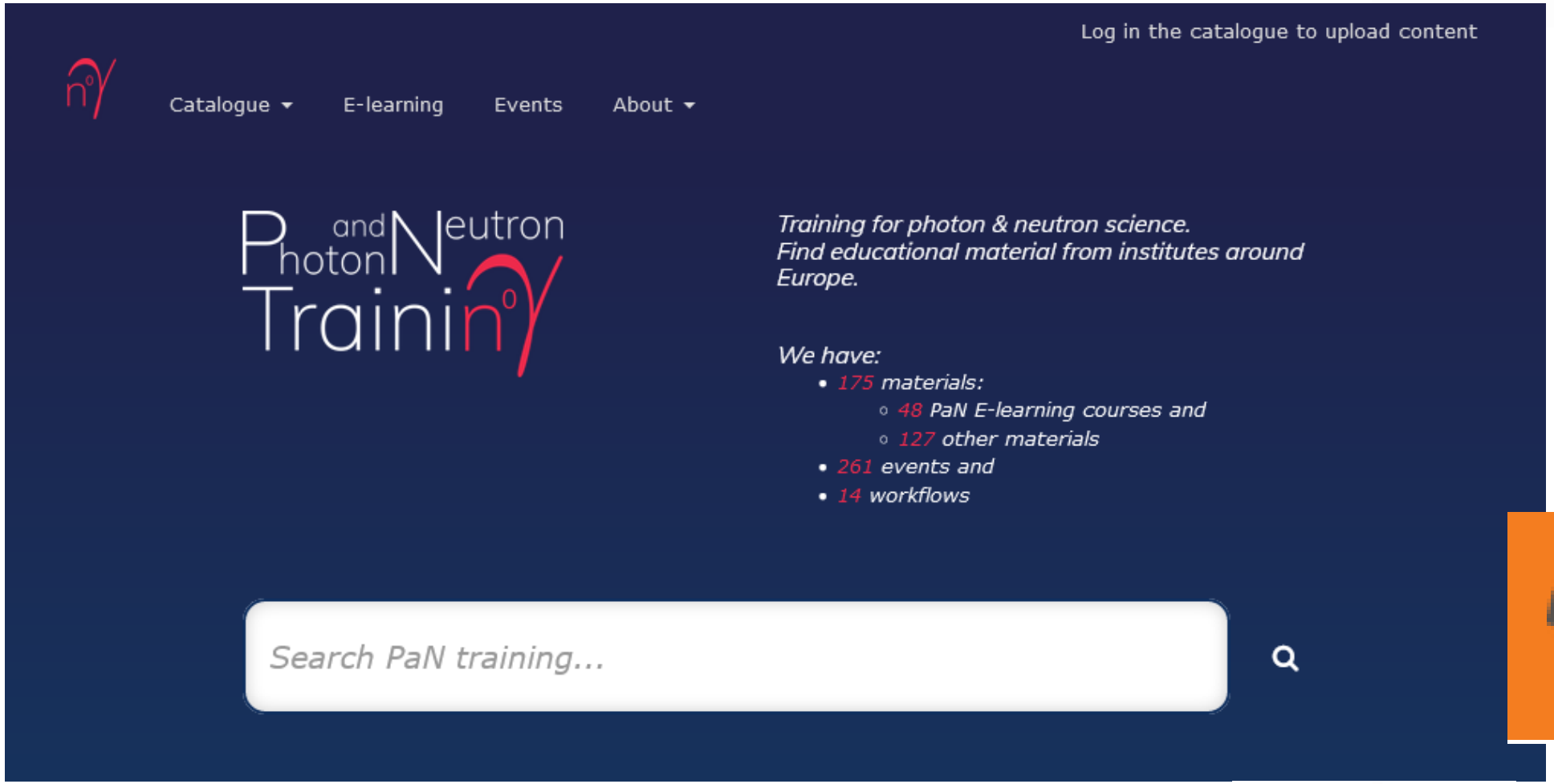
1. *Publish training material on pan-training.org*
2. *Develop learning material on pan-learning.org*



https://e-learning.pan-training.eu/moodle/

PaNOSC and E
and innovation p

Share training material, videos, events, etc. on Pan-training.org

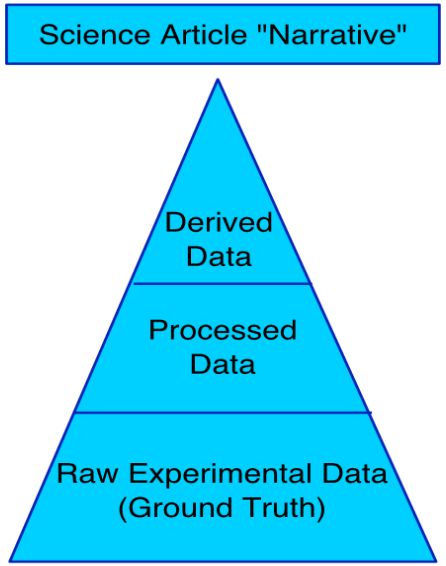


https://pan-training.eu/

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PaN Data Commons - What , Why and How

1. Create new kind of User community – **PaN Open Data User Community**
2. Open Data are **findable** via one **portal**
3. Users can analyse the Open Data on their own or request "**Data-Time**" for help
4. Users can **train, analyse, verify, publish**



"**The vital role of primary experimental data for ensuring trust in (Photon & Neutron) science**"
https://doi.org/10.5281/zenodo.5155882

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

What is a Data Commons?

Building a public data commons

The "data sharing for public good" narratives can be traced at least back to 2011, when the United Nations popularized the concept of "data commons": using privately-owned big data for sustainable development and humanitarian action.[16] The concept of the data commons is crucial, as it defines both values and institutional setups necessary for valuing access and freedom to operate, over the power to appropriate.[17]

https://openfuture.eu/publication/public-data-commons/

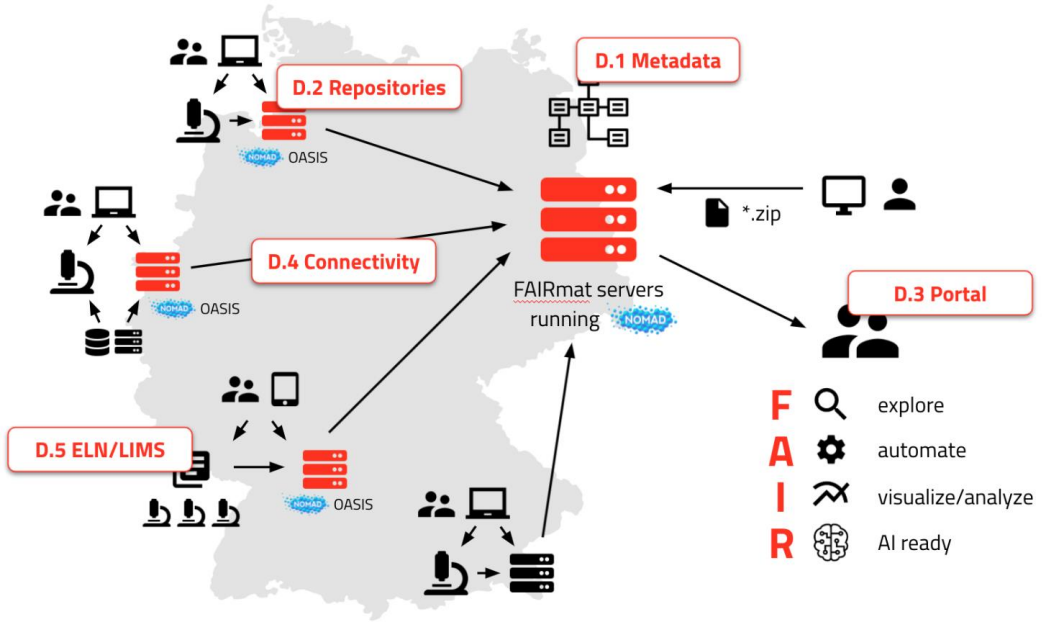
The overarching goal of the NIH Data Commons was to accelerate new biomedical discoveries by developing and testing a cloud-based platform where investigators could store, share, access, and interact with digital objects (data, software, etc.) generated from biomedical and behavioral research.

https://commonfund.nih.gov/commons

Data Commons aggregates data from a wide range of sources into a unified database to make it more accessible and useful. More on why we are building Data Commons.

https://datacommons.org/

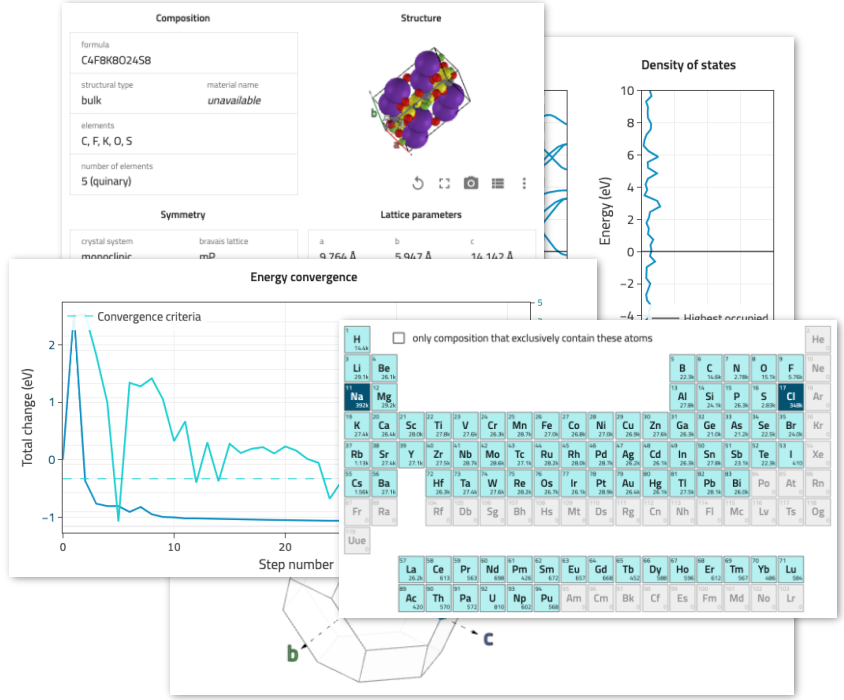
Example of a Data Commons



NOMAD: Publishing research data

More than 12 million of simulations (22 billion quantities) from over 500 authors world-wide

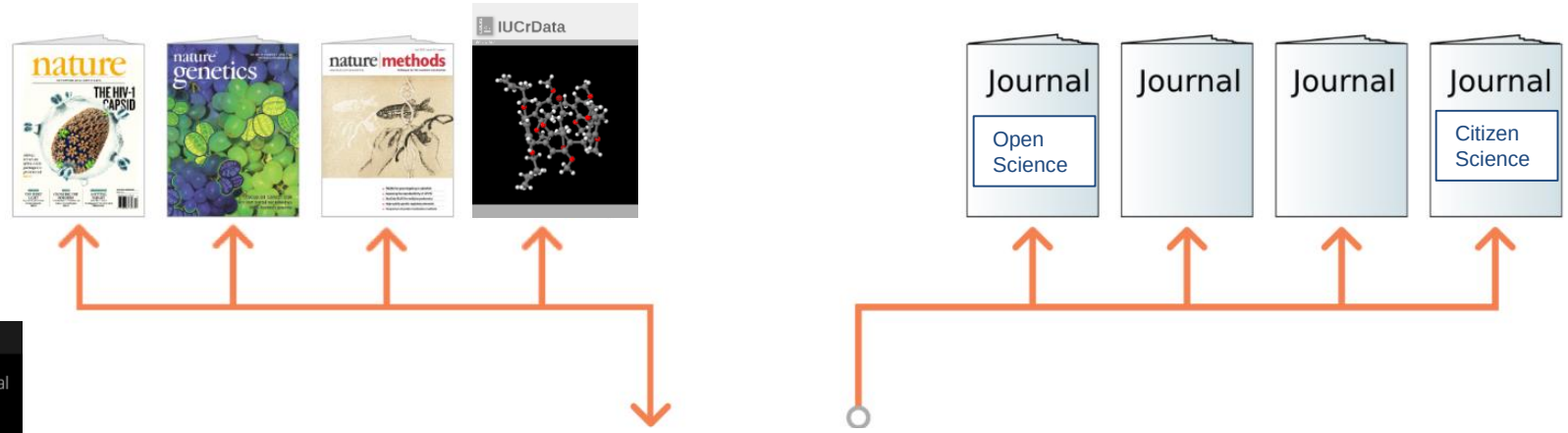
- ➔ Free publication and sharing data of data
- ➔ Extracts **rich metadata** for more than **50 codes**
- ➔ All data in a **raw** and a common **machine readable** from
- ➔ Use integrated tools to **explore**, **visualize**, and **analyze**

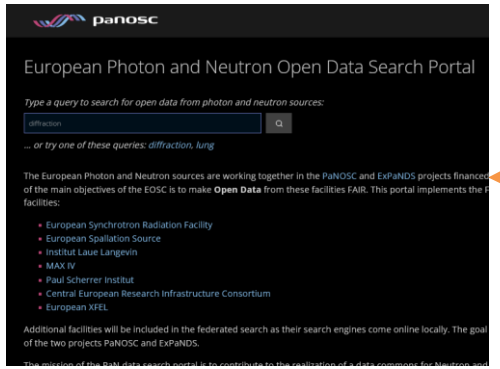


PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Slide courtesy of Markus Scheidgen

Open Data from federated PaN Repositories





PaN Open Data Portal

PaN Open Data Repository

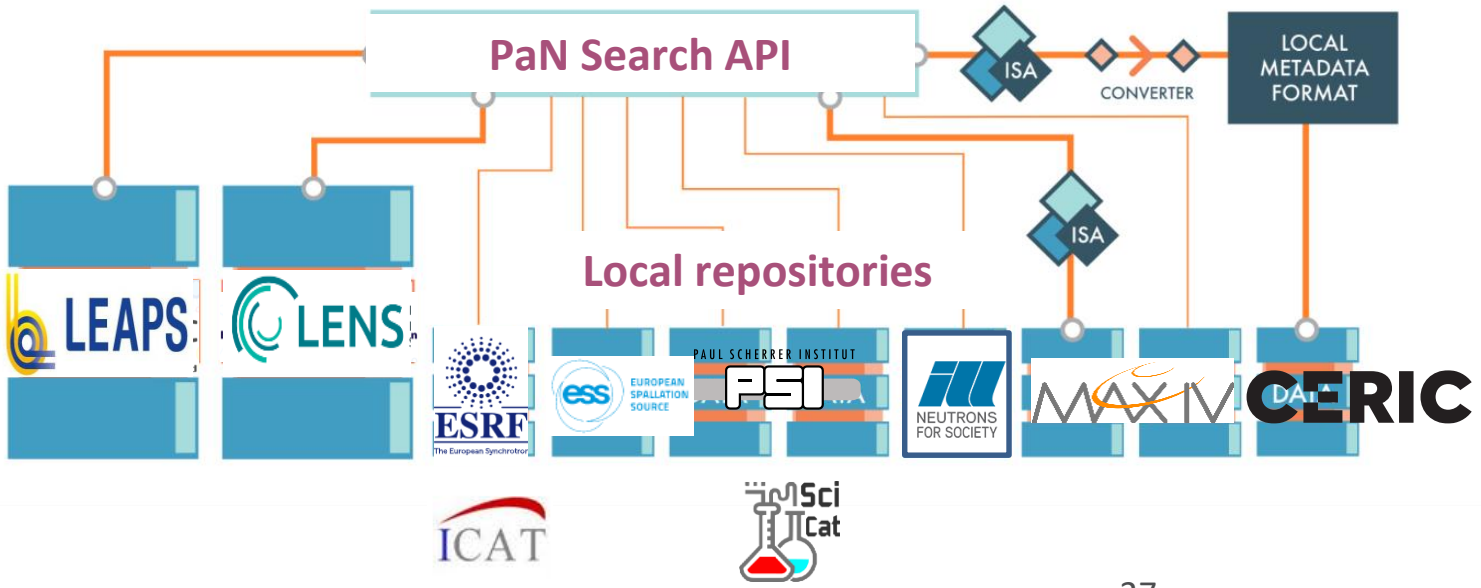
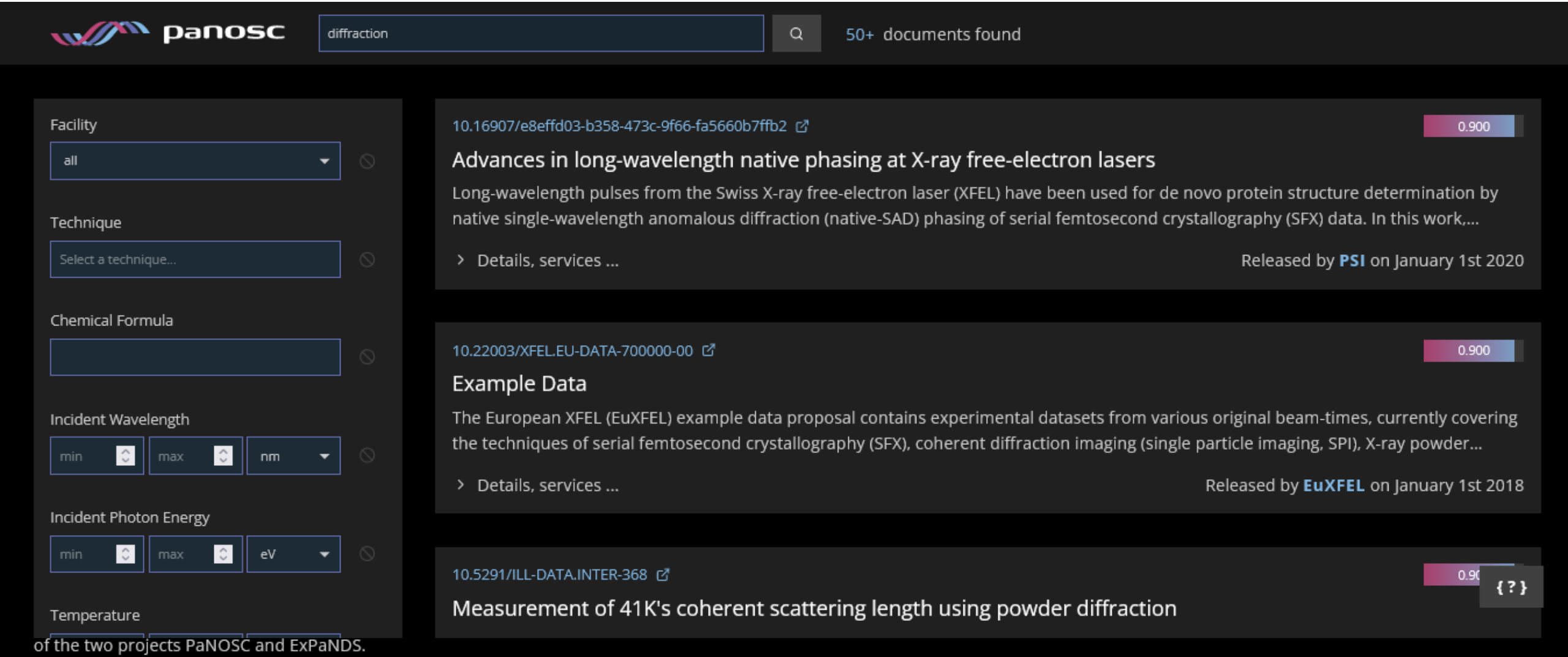


image: http://blogs.nature.com/scientificdata/2013/07/23/scientific-data-to-complement-and-promote-public-data-repositories/

PaN Data Portal - https://data.panosc.eu



of the two projects PaNOSC and ExPaNDS.

The mission of the PaN data search portal is to contribute to the realization of a data commons for Neutron and Photon science. The search results provide a link to the landing page of the

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

The ExPaNDS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.
The PaNOSC project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852.

Human Organ Atlas - https://human-organ-atlas.esrf.eu

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

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

Funding

This project has been made possible by funding from:

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

Join us at this workshop on leveraging Open Data
https://indico.synchrotron-soleil.fr/event/67/

During the last decade, most European Photon and Neutron (PaN) facilities have adopted **open data policies**, making data available for the benefit of the entire scientific community. At the same time, **machine learning** (ML) is seen as an essential tool to address the exponential growth of data volumes from PaN facilities.

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Adoption (in October 2022) of PaNOSC+ExPaNDS outcomes

FACILITY	FAIR data policy	DMPs	DOIs	Nexus HDF5	Search API	Open Data Portal	AAI	Jupyter Lab	VISA	VINYL/ OASYS/ McStas	Pan-learning/ training
ALBA	P	P	WIP	WIP	WIP	WIP	P	Y	WIP	N	U
DESY	WIP	WIP	WIP	Y	WIP	P	WIP	Y	U	Y	WIP
CERIC-ERIC	Y	WIP	Y	WIP	Y	Y	Y	Y	Y	Y	Y
DIAMOND											
ELETTRA	Y	WIP	Y	Y	Y	Y	Y	Y	Y	Y	Y
ESRF	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
ELI-ERIC	Y	Y	P	Y	Y	Y	WIP	Y	Y	Y	Y
ESS	Y	Y	Y	Y	Y	Y	Y	WIP	WIP	Y	Y
EuXFEL	Y	WIP	Y	WIP	Y	Y	WIP	Y	WIP	Y	Y
FELIX	Y	P	WIP	U	U	WIP	U	U	N	N	U
HZB	Y	P	WIP	Y	P	Y	P	U	U	U	U
HZDR	Y	WIP	Y	N	U	Y	Y	Y	P	WIP	Y
ILL	Y	WIP	Y	Y	WIP	Y	Y	Y	Y	Y	WIP
MAX-IV	WIP	U	Y	Y	Y	Y	Y	Y	U	U	U
PSI	Y	WIP	Y	WIP	Y	Y	WIP	WIP	N	N	N
PTB	Y	WIP	Y	WIP	N	Y	N	N	N	N	N
SOLARIS											
SOLEIL	Y	WIP	WIP	Y	WIP	WIP	Y	WIP	WIP	U	Y
SESAME	Y	U	P	Y	P	WIP	P	P	N	Y	N

Not Planning to be adopted (N)
Under evaluation (U)
In progress of being adopted (WIP)

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 823852 and No 871072 respectively.

2023 – PaNOSC continues as a community activity

PaNOSC Partners are members of LEAPS + LENS

ESRF The European Synchrotron

DESY.

PAUL SCHERRER INSTITUT PSI

NEUTRONS FOR SOCIETY

HZB Helmholtz Zentrum Berlin

European XFEL

diamond

Elettra Sincrotrone Trieste

ess EUROPEAN SPALLATION SOURCE

ALBA

SOLEIL SYNCHROTRON

MAX IV

INFN Istituto Nazionale di Fisica Nucleare

UKRI Science and Technology Facilities Council

ISA

CERIC Central European Research Infrastructure Consortium

eli

HZDR

SOLARIS NATIONAL SYNCHROTRON RADIATION CENTRE

PTB

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Vision of Senior Level Engagement

The Extreme Light Infrastructure ERIC (ELI)

We are increasingly seeing in some communities, the **recognition that by sharing the data** on an appropriate timescale, ideally as soon as possible, there are some **real benefits** to be had. So, I think the challenge, the cultural challenge is to demonstrate to the science community at large that actually the **benefits greatly overwhelm the risks**.

To get optimum value out of having open data ... we need to be inclusive, we need to actually involve as many different facilities and research establishments as possible and that is a **really big coordination job**. What the **ExPaNDS** and **PaNOSC** grants provided is an **excellent basis** for continuing this work on open data and being able to share data.

Data collection is not enough, advances only **come through the interpretation of data.**

There was always an understandable sense of data ownership from the scientists who conducted the experiments but the **interdisciplinary research of today requires a new way of thinking**.

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

LEAPS Data Strategy

Home > The European Physical Journal Plus > Article

LEAPS data strategy

Regular Article | Open Access | Published: 17 July 2023 | 138, Article number: 61

Download PDF

You have full access to this open access article

Eur. Phys. J. Plus (2023) 138:617
https://doi.org/10.1140/epjp/s13360-023-04189-6

THE EUROPEAN PHYSICAL JOURNAL PLUS

Regular Article

Check for updates

LEAPS data strategy

Andy Götz[1,a], **Erwan le Gall**[2,b], **Uwe Konrad**[3], **George Kourousias**[4,c], **Oliver Knodel**[3,d], **Salman Matalgah**[5,e], **Oscar Matilla**[6], **Darren Spruce**[7,f], **Ana Valceril Orti**[8], **Majid Ounsy**[9], **Thomas H. Rod**[9], **Frank Schluenzen**[10]

[1] European Synchrotron Radiation Facility, Grenoble, France
[2] Institut Laue Langevin, Grenoble, France
[3] Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany
[4] Elettra Sincrotrone Trieste, Basovizza, Italy
[5] SESAME, Allan, Jordan
[6] ALBA, Cerdanyola del Vallès, Spain
[7] MAXIV Laboratory, Lund University, Lund, Sweden
[8] SOLEIL, Saint-Aubin, France
[9] European Spallation Source ERIC, Copenhagen, Denmark
[10] DESY, Hamburg, Germany

Received: 17 December 2022 / Accepted: 14 June 2023
© The Author(s) 2023

Abstract The continuous evolution of photon sources and their instrumentation enables more and new scientific endeavors at ever increasing pace. This technological evolution is accompanied by an exponential growth of data volumes of increasing complexity, which must be addressed by maximizing efficiency of scientific experiments and automation of workflows covering the entire data lifecycle, aiming to reduce data volumes while producing FAIR and open data of highest reliability. This papers briefly outlines the strategy of the league of European accelerator-based photon sources user facilities to achieve these goals collaboratively in an efficient and sustainable way which will ultimately lead to an increase in the number of publications.

https://doi.org/10.1140/epjp/s13360-023-04189-6

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PUMA – Publication and User experiment Metadata Analyser

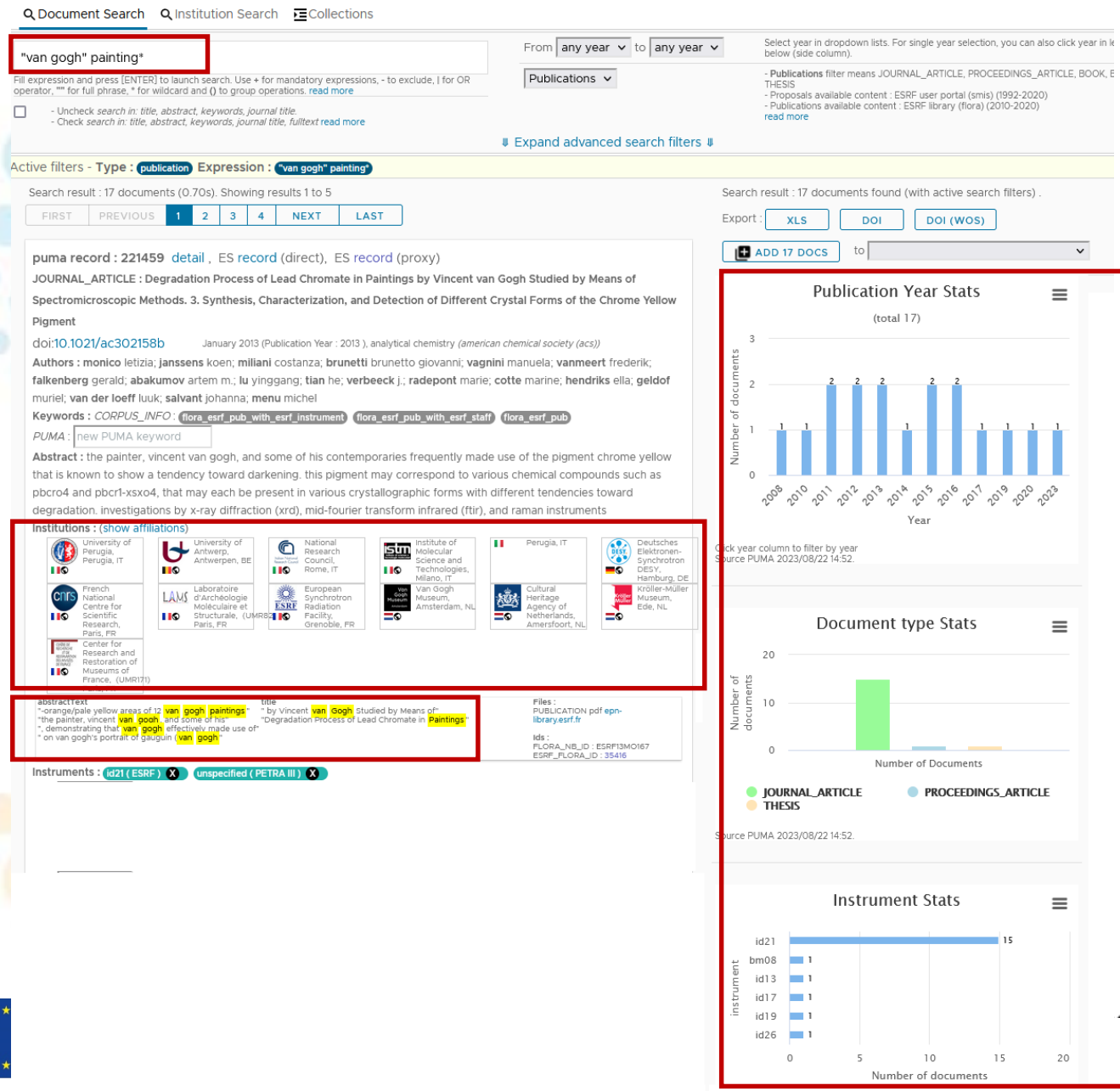
- The data that **PUMA** can access at ESRF includes information about the instruments and beamlines, techniques, scientific areas, authors, member countries, publications (>40,000), public proposals submitted (>46,000) and accepted (>20,000), and industrial proposals (>1,300).
-
- Additional publication metadata is loaded into PUMA using open data web services (citations, keywords, abstract, authors...)

- PUMA provides a quantifiable view of the data, not only enabling conclusions about the science currently being carried out at the ESRF, but also helping to identify trends.

- It facilitates reporting for the ESRF management and supports strategic planning for the facility.

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PUMA – explore corpus



Explore corpus using advanced search

- Proposal & publications
- Full text search (title/abstract/pdf text)
- Instruments, institutions, authors, journals.

PUMA – custom searches

- Topic collections (proposal/publications) can be created within PUMA for custom reporting.
- **Example : “sustainable energy research”, “neurodegenerative diseases research” “human organ atlas”**
- Dashboard based on your custom collection is then available to report on various metrics (Country, institutions, instrument used, open access, journals, citations...)
- In production at **ESRF** + **ILL**, soon at **SOLEIL** (as part of an EU project)
- Contact Renaud Duyme @ ESRF if you are interested

PaNOSC and ExPaNDS projects have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PUMA – comparison with commercial tools

Clarivates : Web Of Science & Incites.

Positive

- Extended publication corpus (contains all scientific publications)
- Advanced publication categorisation (“citation topics”)
- Advanced KPI indicators (Incites)

Negative

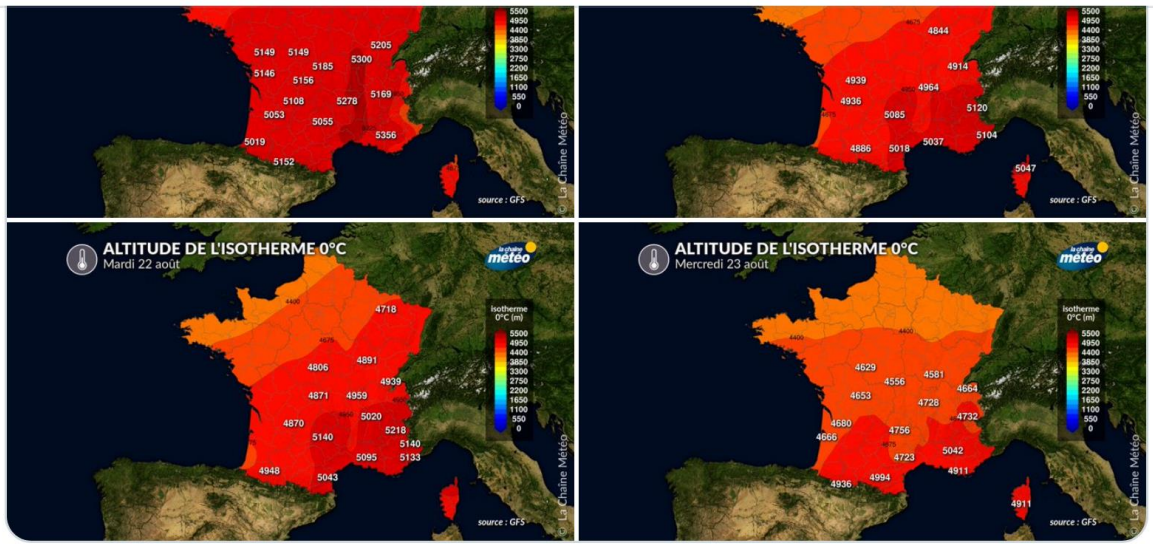
- Generic tool not built for Instrument Facilities
- No direct integration with Facility publication corpus
- No Proposal/Experiment/Instruments stats
- Searches only on title/abstract (full publication content not available)
- No Open Data stats

PaNOSC and ExPaNDS projects have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

La Chaîne Météo
@lachainemeteo · **Suivre**

🌡️ Jusqu'à mercredi, l'isotherme 0°C se situera aux alentours de 5000 mètres d'altitude dans le sud de la France (il dégèlera au #Mont_Blanc !). Prudence si vous avez prévu de randonner en haute montagne ⛰️ près des glaciers alpins qui seront fragilisés par ces températures… Voir plus

Le Parisien



11:50 PM · 20 août 2023

74 Répondre Copier le lien

Lire 4 réponses

Horizon 2020 research
y.

ESRF IT devices electrical consumption

ESRF Electrical consumption e.g. 2018

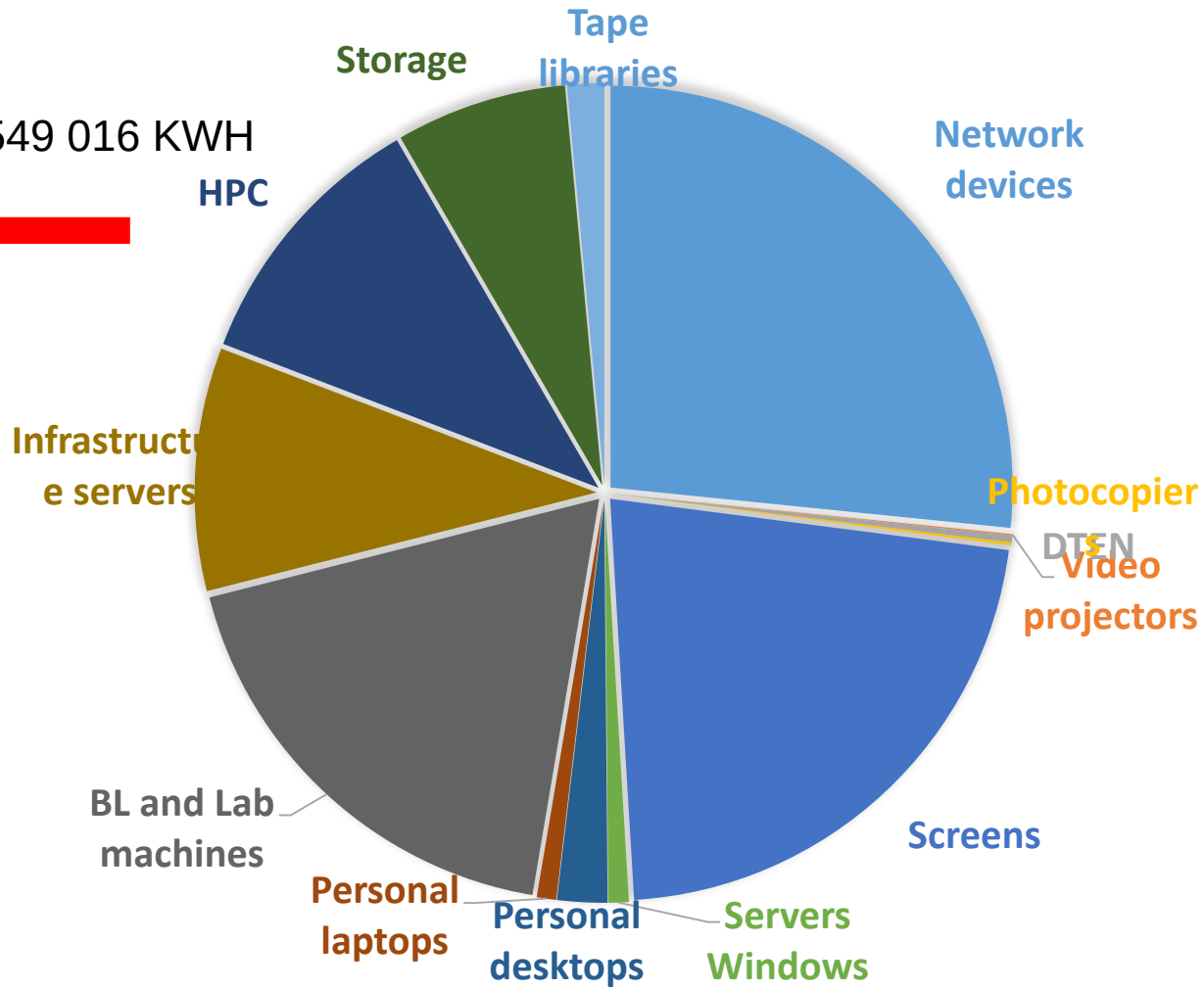
Poste	Consommation	
MACHINE	35 200 000	kWh
HQPS	3 056 000	kWh
CTA (hors central et common)	4 610 000	kWh*
SRE	3 745 000	kWh
Lignes et Laboratoires	3 439 500	kWh
Eclairage	2 080 000	kWh*
Eclairage lignes labos	1 529 000	kWh*
SEG	2 675 000	kWh
Data center central	1 146 500	kWh*
Common building	953 000	kWh
Pertes transformateurs	814 000	kWh*
Extracteurs (hors central, common et GH)	836 500	kWh*
SEB	854 500	kWh
Data center CTRM	600 200	kWh*
Central building	239 500	kWh
SEC	322 000	kWh
SEI	437 000	kWh
SAP	405 000	kWh
SRX	322 000	kWh
Guest houses	284 000	kWh
LOB	221 500	kWh
Station pompage	39 000	kWh
Autres	1 575 000	kWh*
Total	**65 385**	**MWh**

* *Estimations d'après puissances installées et profils de fonctionnement*

Source TID BIG 2018

ESRF Electrical consumption – IT Only

2 549 016 KWH



Source TID-ITS 2021

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

IT – impact on climate in 2021

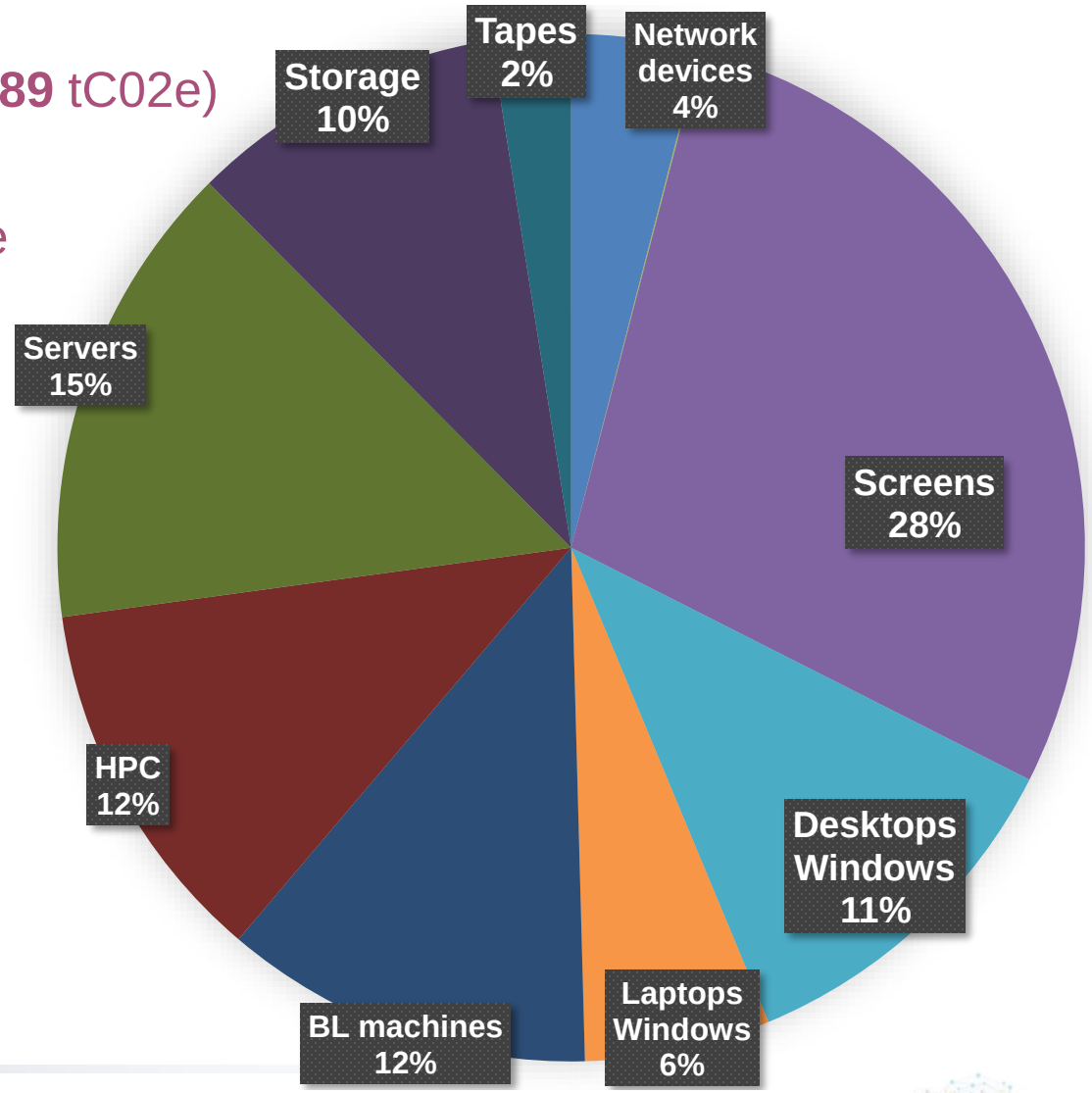
Elec consumption **2021** (**2 549 016 kWh x 35 gC02** = **89** tC02e)

+

(Manufacturing + Transport) / Equipment life time

=

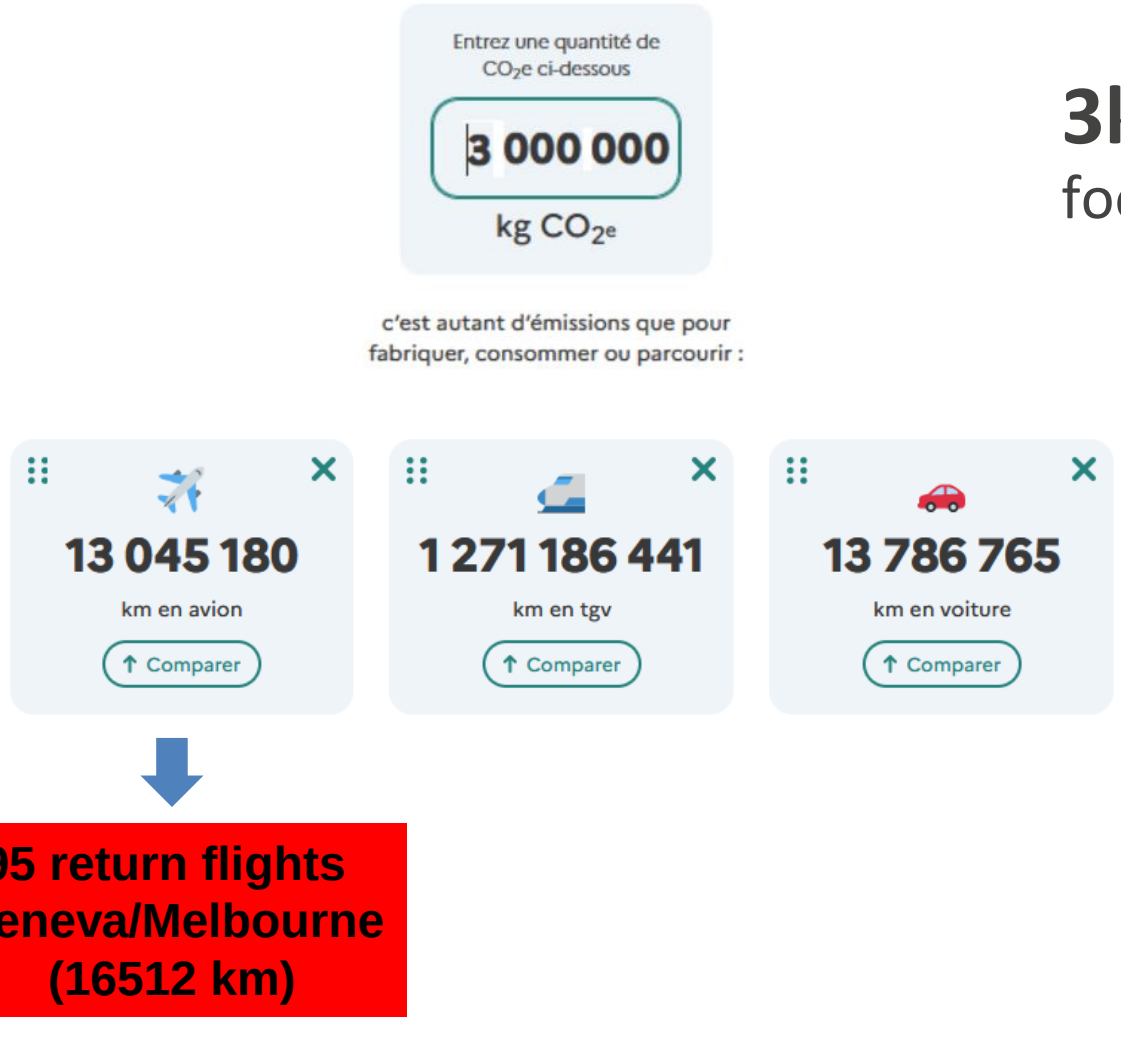
3.08 kt C02equ



Source
- IT Elec. consumtion 2021 : TID ITS.
- Manufacturing & transport : Suppliers (Dell, Lenovo…)

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.
The ExPaNDS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.
The PaNOSC project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852.

Equivalent – CO2e



395 return flights Geneva/Melbourne (16512 km)

3kt C02e IT carbon footprint at ESRF for 1 year

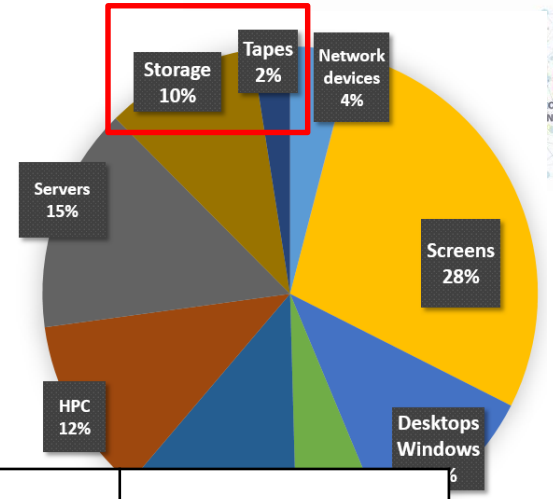
ADEME Tool to calculcate CO2eq equivalent for different objects : https://impactco2.fr/convertisseur

ADEME : **In France average emission is 10t CO2 per year per person, should be reduced to 2t.**

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.
The ExPaNDS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.
The PaNOSC project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852.

Disk vs Tape storage for a year

Disk storage vs Tape over a year (source IBM, FUJI, LENOVO)



Tape library with 1000PB (LTO8)

	KgC02		#time life (years)	Total KgCO2 (manufacturing + transporta)	Electrical consumption	Total KgC02	Volume PB
Elements	8000	10	15	5333.333333			
Robots	836	2	15	111.4666667			
drives	45	33	6	247.5			
tapes	7	51200	6	59731.78			9
Total				65424	1123	66547	**66 KgCO2/PB**

Tape library with 220 PB (LTO8)

	KgC02		#time life (years)	Total manufacturing KgCO2	Electrical consumption	Total KgC02	Volume PB
Elements	8000	10	15	5333.333333			
Robots	836	2	15	111.4666667			
drives	45	33	6	247.5			
tapes	7	11000	6	12833.33333			220
Total				18526	1123	19648	**89 KgCO2/PB**

Disk storage (Spinning disks – GSS Lenovo)

	KgC02	#	time life (years)	Total manufacturing KgCO2	Electrical consumption	Total KgC02	Volume PB
Storage Servers (GSS Lenovo)	54000	4	5	43200	10 443	53 643	
							20
Total				43200	10443	53643	**2682 KgCO2/PB**

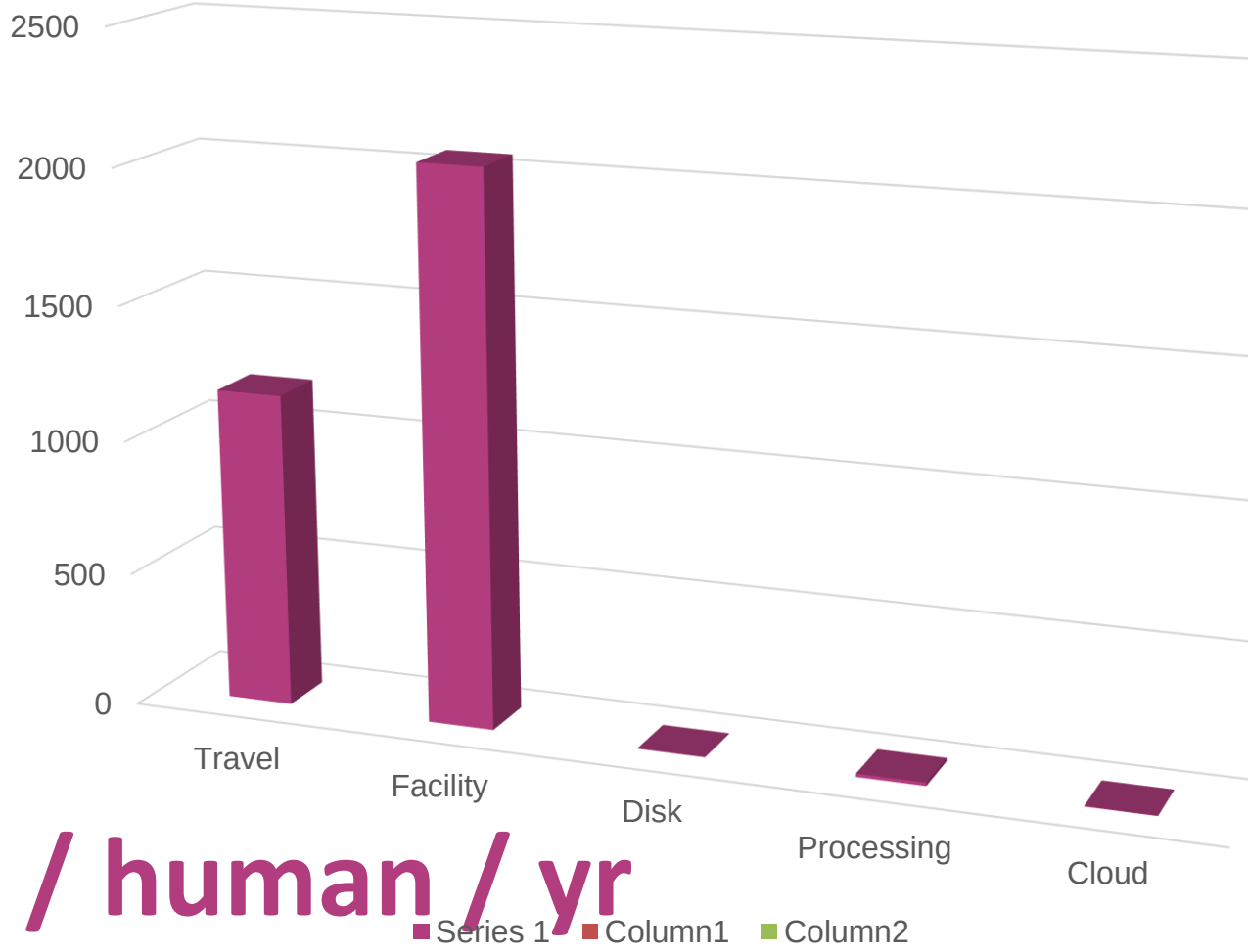
and innovation programme under grant agreements 823852 and 857641, respectively.

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**

Carbon footprint for 1 week experiment @ ESRF



TOTAL = 3.253 tons !

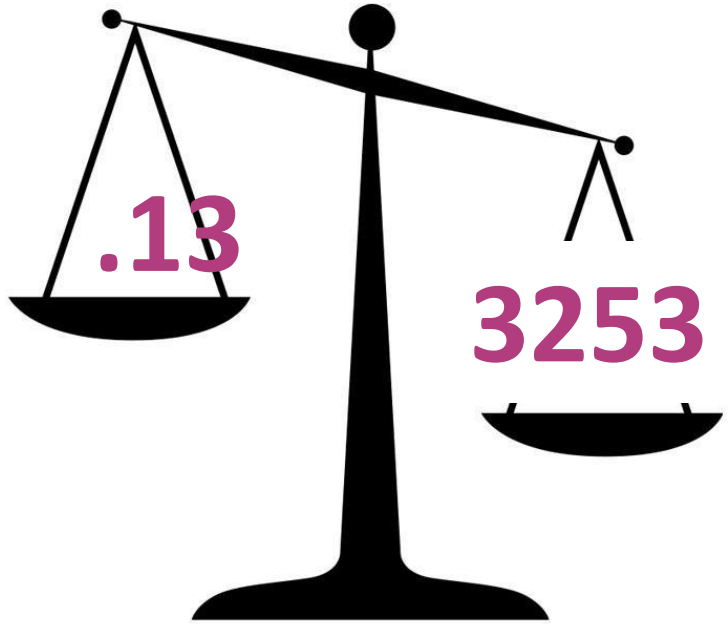
Sustainable Goal = 5 tons / human / yr

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Carbon footprint of archiving data

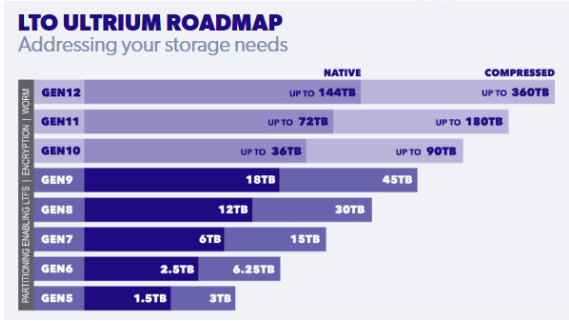
- **200 GB Data archived on tape for 10 years (full tape library)**
  ~ 13 g * 10 yrs = 130 grams

- ARCHIVING raw data for 10 years
  4e-6 % of CO2eq of beam time
  to acquire the raw data!

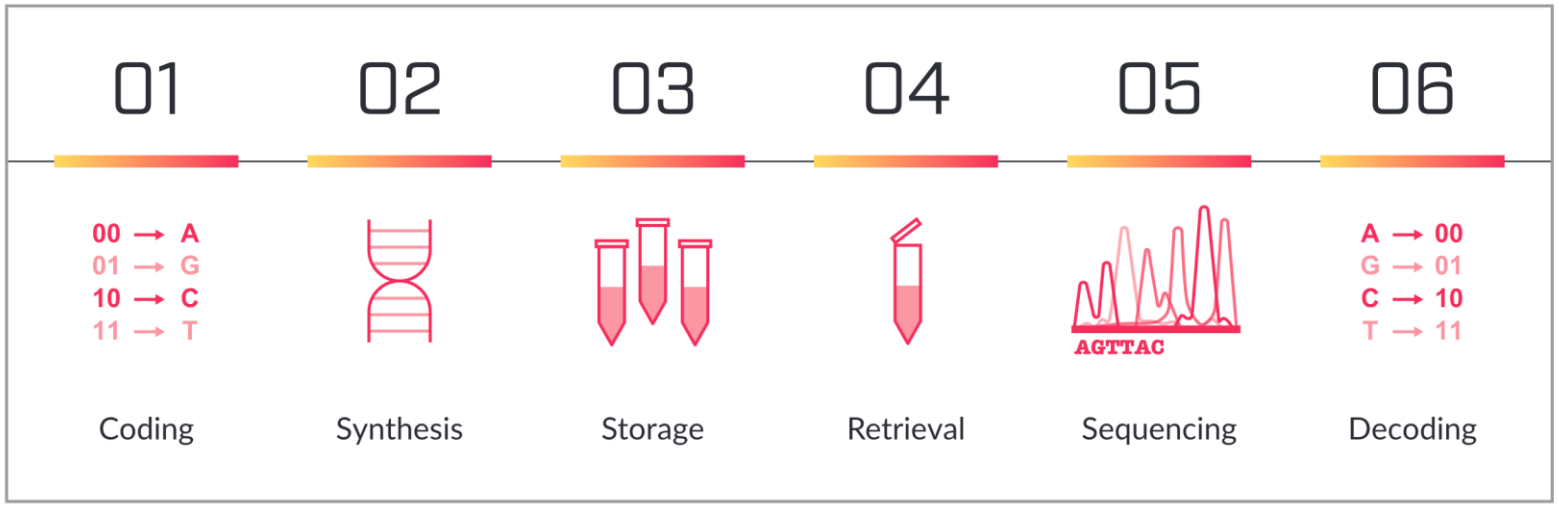


PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

DNA – the ultimate storage medium for raw data?



*"Experiments have confirmed the high theoretical information density of nearly **455 billion GB** of data per gram i.e. **455 petabytes / gram, ~6 orders of magnitude greater** than even the most advanced **magnetic tape** storage systems"*



How DNA storage works

DNA DATA STORAGE ALLIANCE
A SNIA TECHNOLOGY AFFILIATE

Matange, K., Tuck, J.M. & Keung, A.J. DNA stability: a central design consideration for DNA data storage systems. *Nat Commun* **12**, 1358 (2021). https://doi.org/10.1038/s41467-021-21587-5

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

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

IUCrData
ISSN 2414-3146

Received 20 April 2021
Accepted 1 May 2021

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

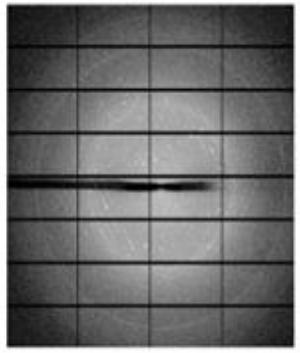
Crystal structure of the second extracellular domain of human tetraspanin D9: twinning and diffuse scattering

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

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

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

Raw data



Structure

Raw diffraction data

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

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

[CheckCif for Raw Data]

checkImgCIF report

```
ImgCIF checker version 2022-07-16

Checking block 5886687 in he4557img.cif

Running checks (no image download)
======================================

Testing: Required items: PASS

Testing: Data source: PASS

Testing: Axes defined: PASS

Testing: Our limitations: PASS

Testing: Detector translation: PASS

Testing: Scan range: PASS

Testing: All frames present: PASS
   All frames present and correct for SCAN1

Testing: Detector surface axes used properly: PASS

Testing: Pixel size and origin described correctly:

Testing: Check calculated beam centre: PASS

Testing: Check principal axis is aligned with X: PAS
Testing presence of archive:

Testing: All archives are accessible: PASS

Running checks with downloaded images
======================================


Testing image 4: Image type and dimensions: PASS
```

Raw data table generated from the CIF

Raw data	
DOI	https://doi.org/10.5281/zenodo.5886687
Data archive	Zenodo
Data format	HDF5
Data collection	
Beamline	Diamond I04
Detector	
Temperature (K)	
Radiation type	Synchrotron X-ray source
Wavelength (Å)	0.979491
Beam centre (mm)	−166.874, 172.497
Detector axis	−Z
Detector distance (mm)	-287.22
Swing angle (°)	
Pixel size (mm)	0.075 × 0.075
No. of pixels	4148 × 4362
No. of scans	1
Exposure time per frame (s)	
Scan axis	ω, X
Start angle, increment per frame (°)	0.0, 0.1
Scan range (°)	360.0
No. of frames	3600

“If you don't want to share data why become a scientist?”

PaNOSC
@Panosc_eu

← **Thread**

Research Fellow at @UCL & @esrfsynchrotron,and part of the team who developed (w PaNOSC contribution) the #HumanOrganAtlas, an online data portal of human organs scanned w @hip_ct technology at 20µm per voxel
🎥bit.ly/3Mk3yc5

PaNOSC and ExPaNDS projects have received funding from the E
and innovation programme under grant agreements 823852 and 85

Conclusion

1. The Photon and Neutron Open Science Cloud is implementing a Data Commons of FAIR data and will provide data to the scientific community

2. Facilities have the role of managing data for Users

3. Users role is to enrich metadata, publish and cite data

4. IUCr CommDat must continue its good work!

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Thank you

1. **PaNOSC** + **ExPaNDS** collaborators
2. **LEAPS** and **LENS** facilities
3. **IUCr Committee on Data**
4. **Users** for use cases, data and feedback
5. **European Commission** for funding EOSC
6. **ESRF** colleagues **Jean-Francois Perrin**, **Renaud Duyme**, **Daniele de Sanctis**, **Gerd Heber** (HDF Group) and the **EC** for slides

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.