



E x P a N D S

European Open Science Cloud Photon
and Neutron Data Services

Europe's Photon and Neutron Open Science Cloud for Raw and Processed Data: Aims and Achievements to Date @ ORSO 2024

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. Achievements of ExPaNDS + PaNOSC**
- 2. The Photon and Neutron Open Science Cloud**
- 3. Towards the European Open Science Cloud**
- 4. 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.

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.

The goal of ExPaNDS and PaNOSC was to 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

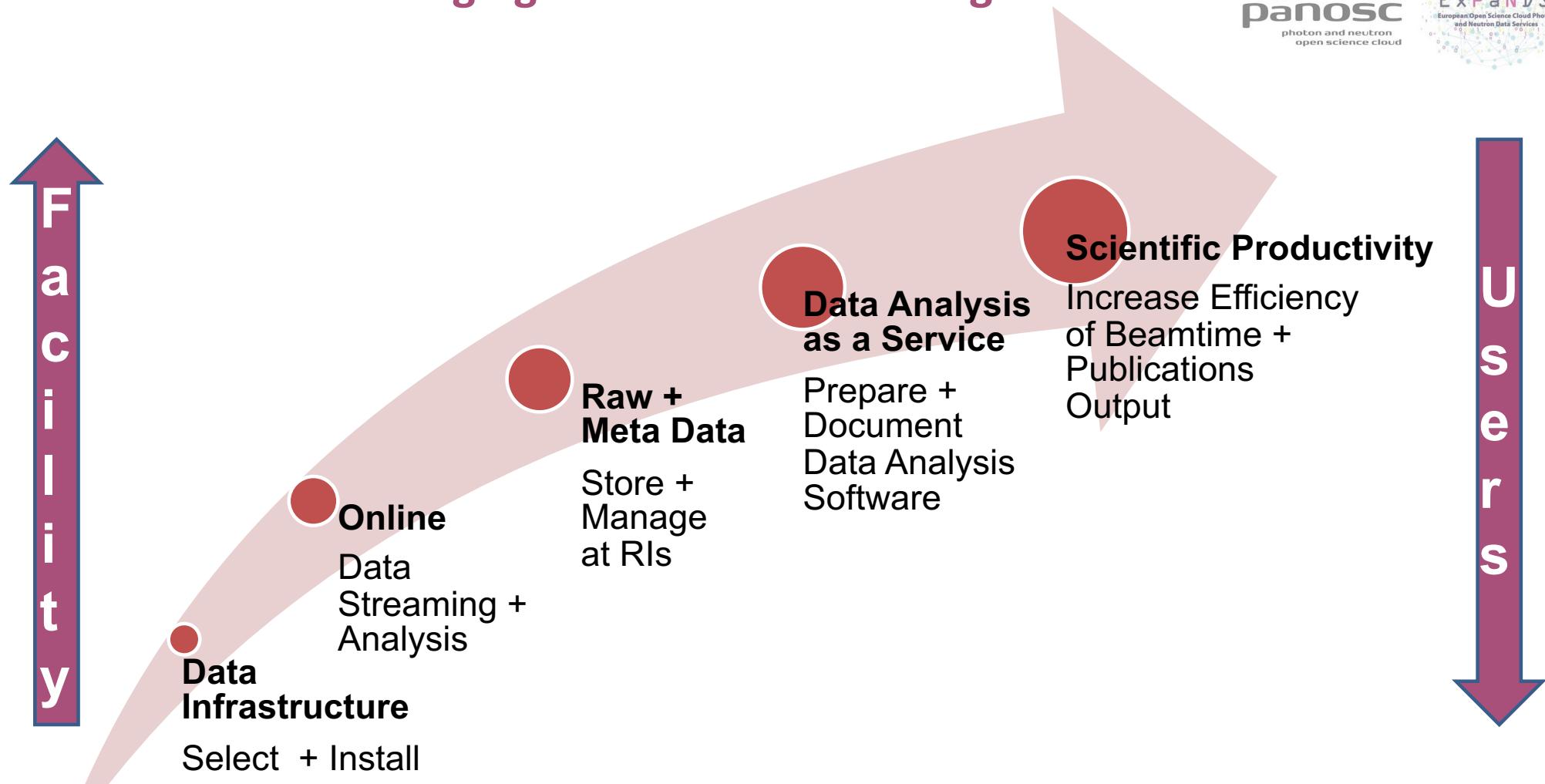
Facilities

- **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.

European support for the PaN community



2010

2015

2018

European Open Science Cloud

2019

2020

2021

EOSC Future

2023



Policies
Analysis
AAI
Training

Common data policy

FAIR data policy

Data Management Plans

Software Catalogue

Remote analysis

Jupyter

UmbrellaID

AARC Blueprint

eduTeams

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



Photon (LEAPS)



Neutron (LENS,



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.

PaNOSC + ExPaNDS - Main Achievements

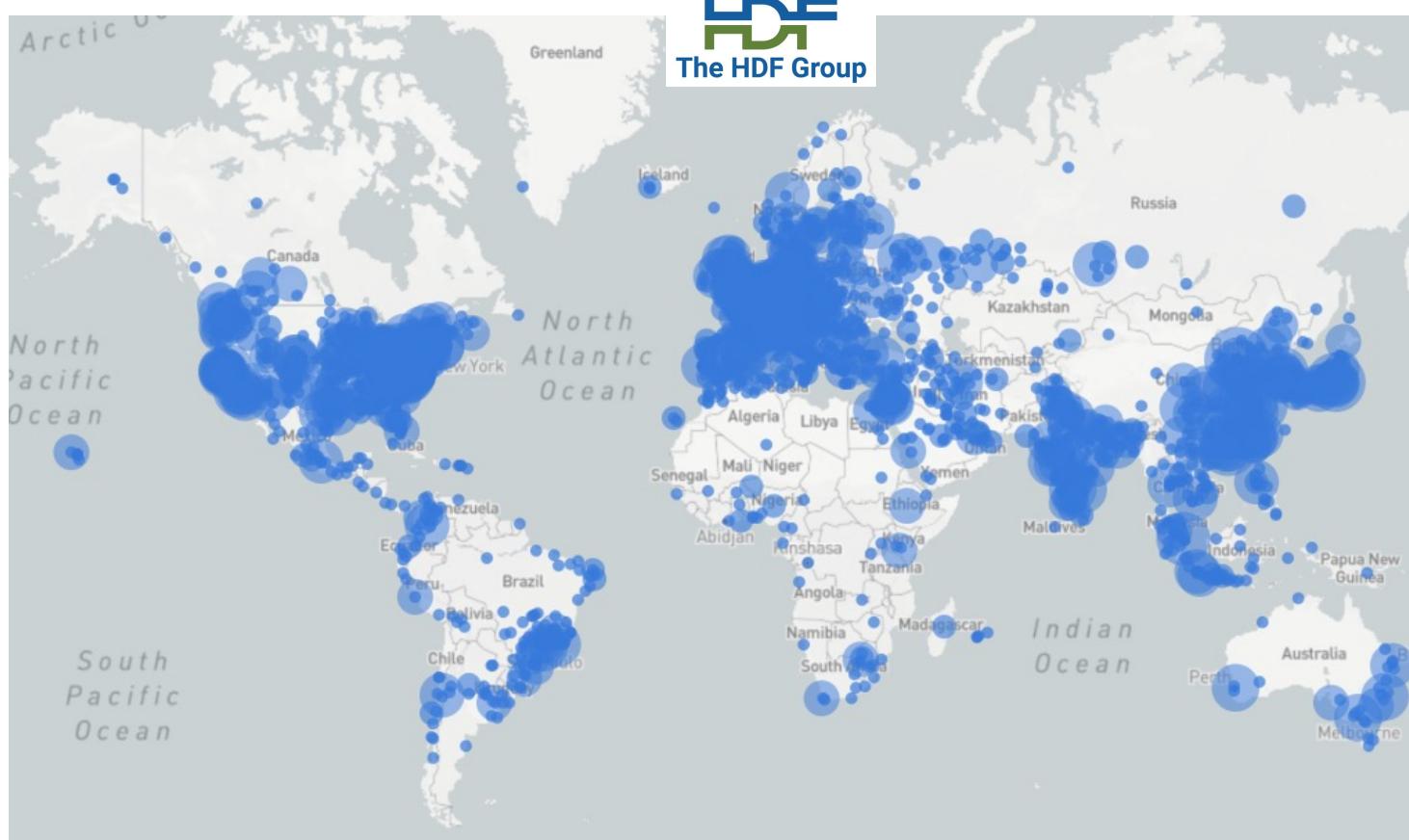


1. **FAIR data policy** and **DMPs**
2. Standardised metadata **Nexus/HDF5** + **PaNET** → **Wout's talk**
3. Guidelines for **FAIR data** self-evaluation
4. **Federated search API** for PaN data catalogues
5. **Open Data portal** for searching + downloading data
6. Community **AAI UmbrellaId** → **EduGAIN AAI**
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-training + learning** platform (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.

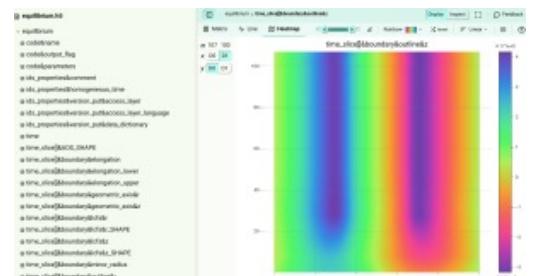
HDF5 is needed to acquire, process and store large datasets and has been adopted as de facto standard across most facilities (with NeXus conventions for metadata)



New Features

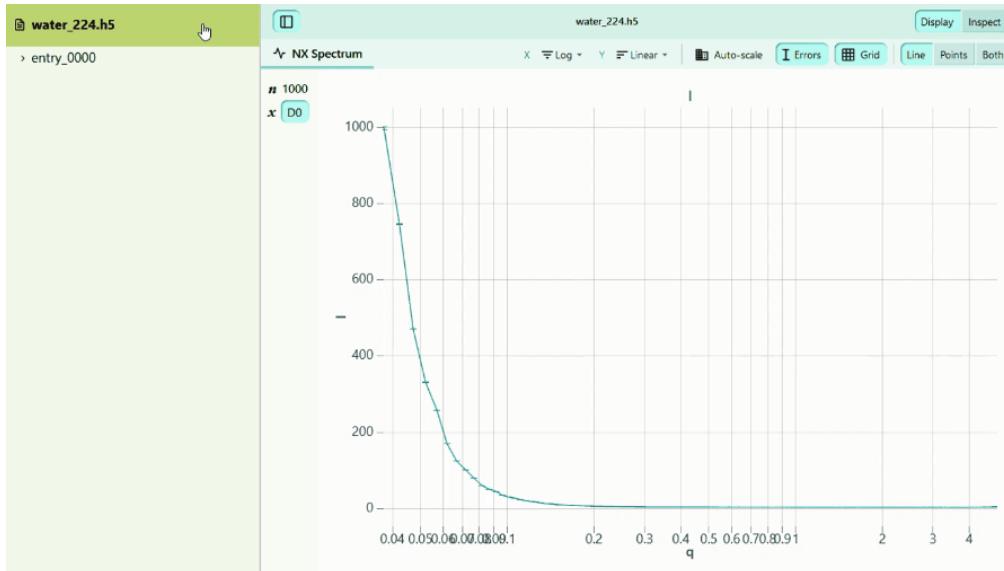
- Multi-dataset I/O
- Selection and vector I/O
- Subfiling VFD
- Multi-Threaded
- Concurrency (WIP)
- Sparse data storage (WIP)

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.

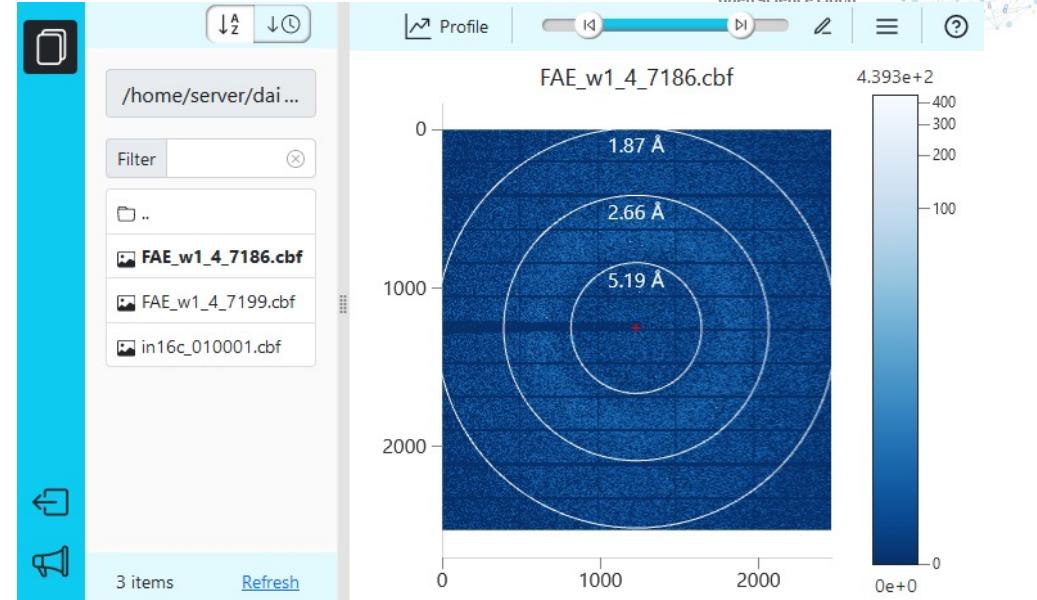
H5Web Visualization Ecosystem



Generic HDF5 file viewer

- 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>



A huge thanks to **Brian Maranville** for **h5wasm** !



H5Web

H5Web | ⚡ 28,187 installs | ★★★★★ (9) | Free

Visualize HDF5 files in VS Code



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/>



myHDF5

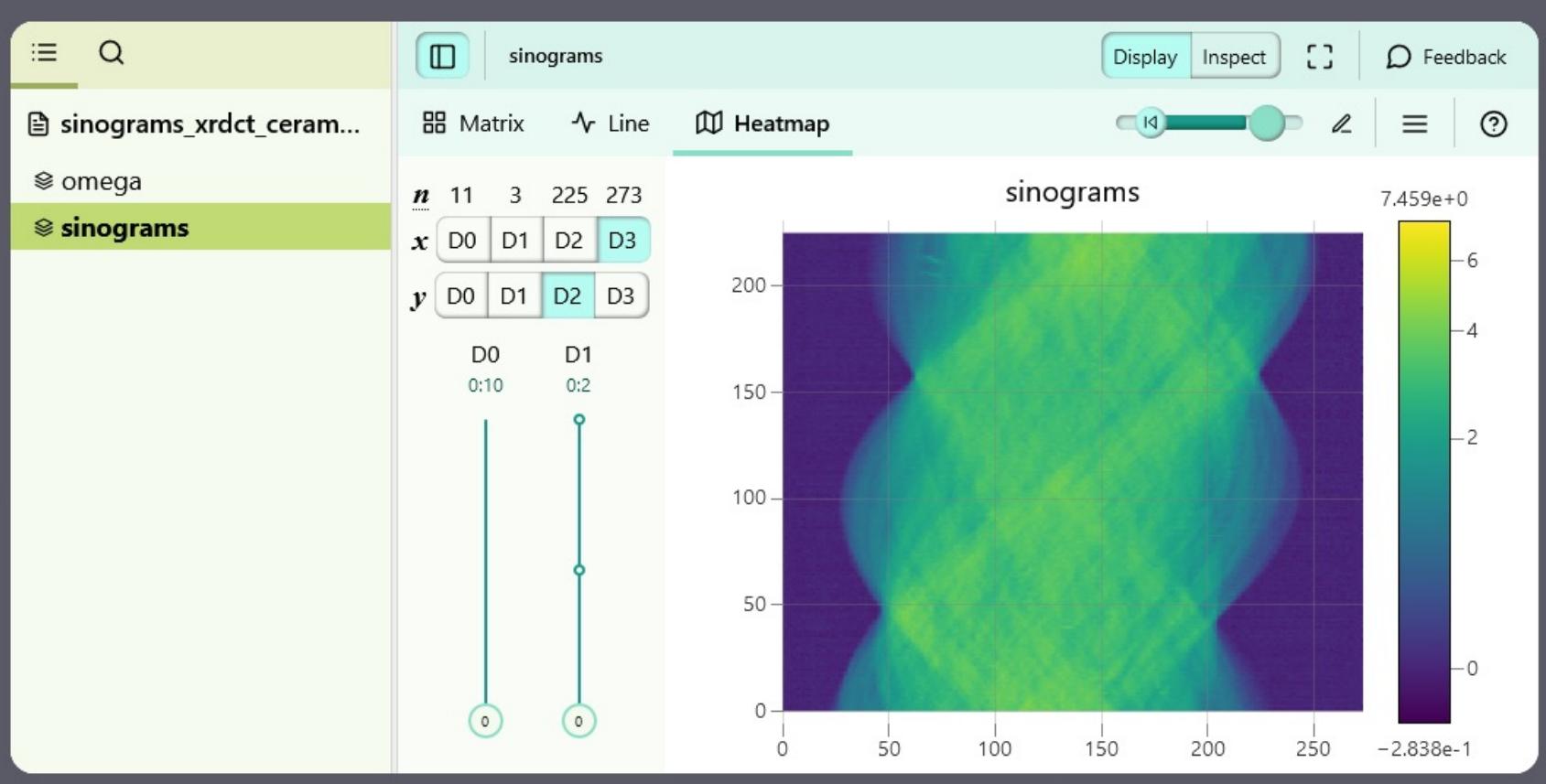
⊕ Open HDF5

⊖ Help

Opened files

Z sinograms_xrdct_ceramic...

« Made by PaNOSC at ESRF



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 626562 and 654704, respectively.

VISA - Remote Data Processing/analyses



New compute instance

Please fill in the details below to create a new compute

Experiments

Select the experiments you wish to associate with your compute instance.

Instance not associated to any specific experiments

Computing Environment

Choose an environment

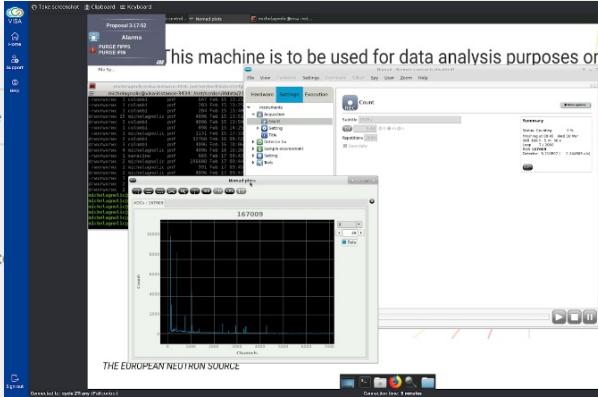
Desktop staging

Desktop

Bliss

Choose hardware requirements

4 Cores 4GB memory esrf.medium	8 Cores 16GB memory esrf.large	16 Cores 32GB memory esrf.gpu.a40	32 Cores 128GB memory esrf.gpu.a40.xlarge
--	--	---	---



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

The VISA platform: Virtual Infrastructure for Scientific Analysis

The VISA Platform: Virtual Infrastructure for Scientific Analysis

PoNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852 and 857641, respectively.

Watch on YouTube

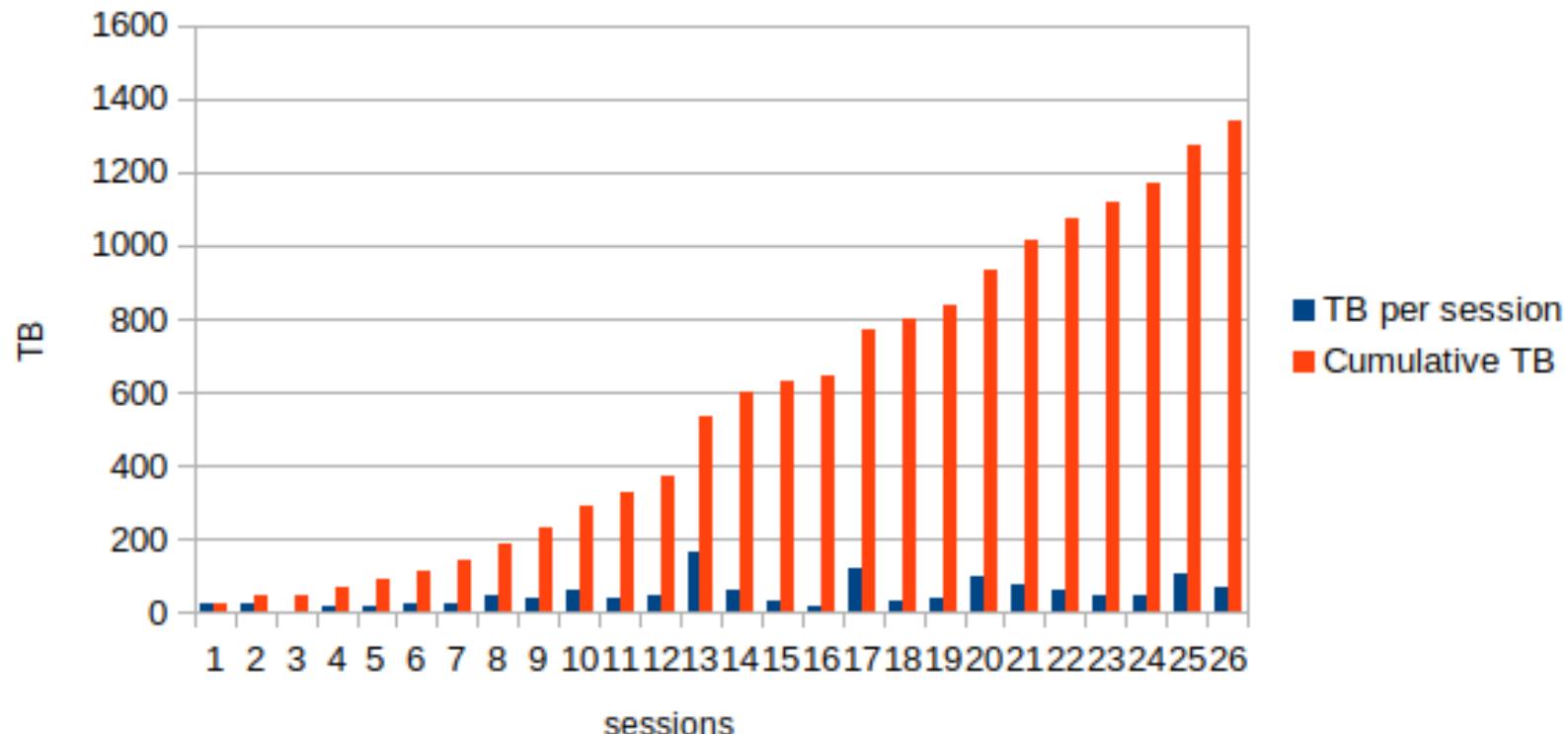


PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

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



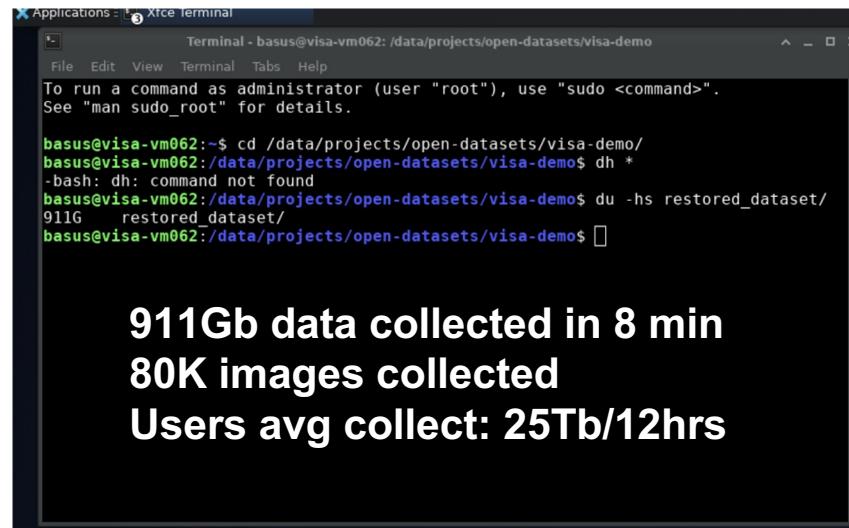
TB per session vs cumulative TB



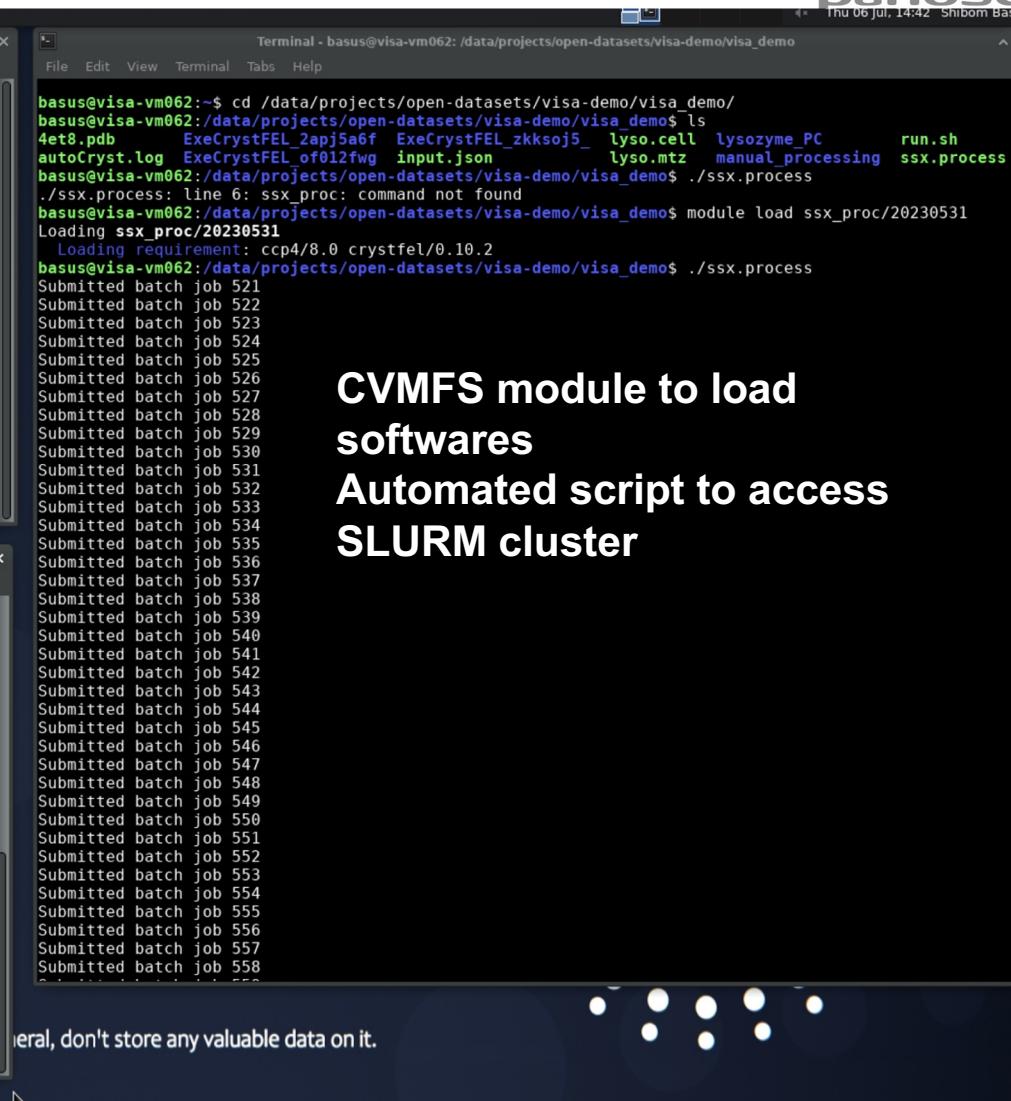
PanOSC is a
and innovation programme under grant agreements 643002 and 643004 respectively under grant agreement no 643002.



VISA for ID29 SSX - demo dataset and processing



**911Gb data collected in 8 min
80K images collected
Users avg collect: 25Tb/12hrs**

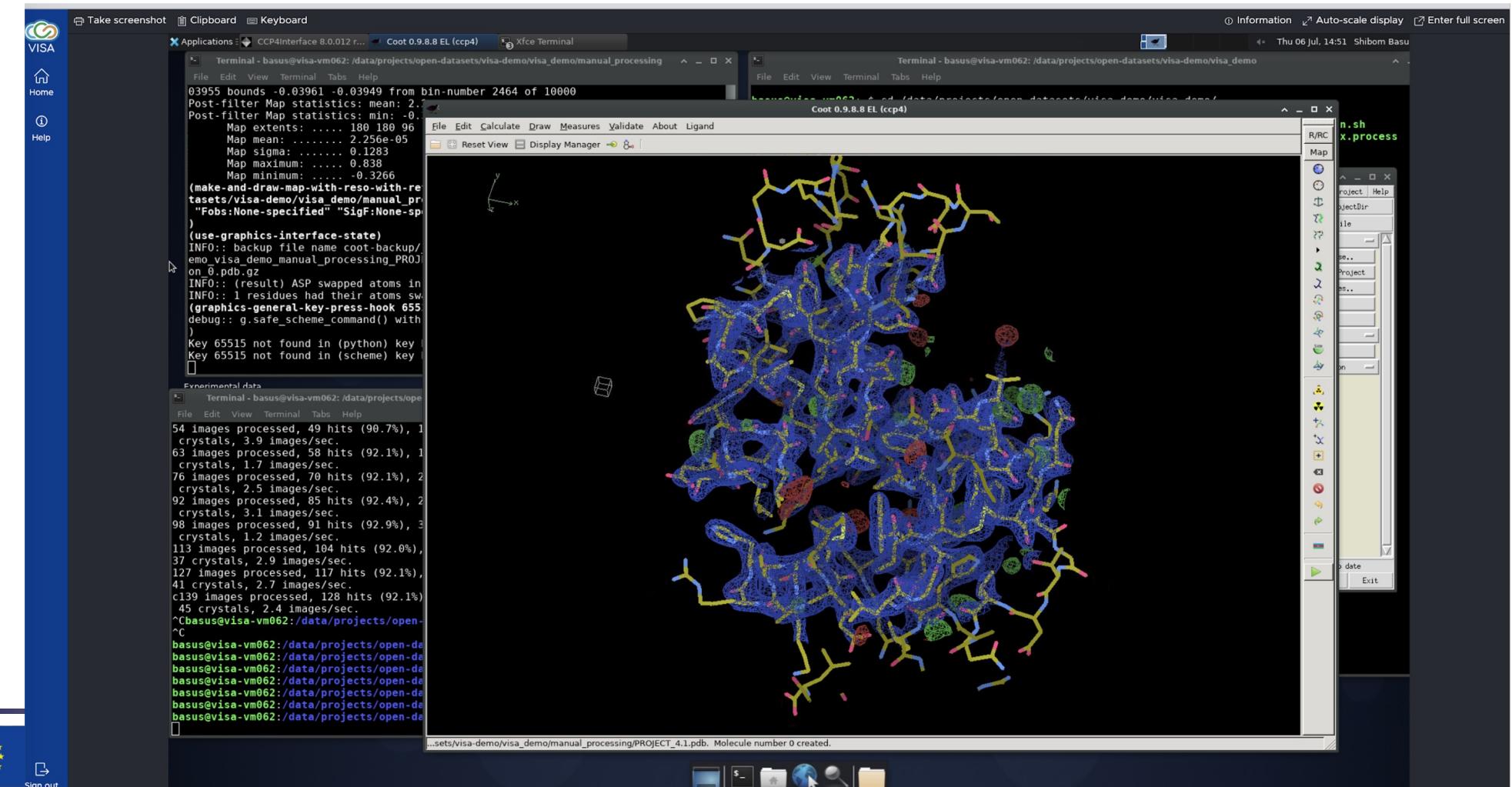


**CVMFS module to load
softwares**

**Automated script to access
SLURM cluster**



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



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

The screenshot shows the homepage of the Pan-training.org website. At the top, there is a blue header bar with the text "Share training material, videos, events, etc. on Pan-training.org". Below this, the main content area has a dark blue background. On the left, there is a logo consisting of the letters "PaN" in white and red, with a Greek letter "gamma" symbol integrated into the design. To the right of the logo are navigation links: "Catalogue ▾", "E-learning", "Events", and "About ▾". In the top right corner, there is a link "Log in the catalogue to upload content" and a small graphic of a network of nodes. The central part of the page features the text "Training for photon & neutron science. Find educational material from institutes around Europe." Below this, there is a section titled "We have:" with a bulleted list:

- 175 materials:
 - 48 PaN E-learning courses and
 - 127 other materials
- 261 events and
- 14 workflows

At the bottom left, there is a search bar with the placeholder text "Search PaN training...". To the right of the search bar is a magnifying glass icon. On the far right, there is a logo for "elixir TeSS" featuring the elixir logo and the text "TeSS".

<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.



Open Training - courses on Pan-learning.org

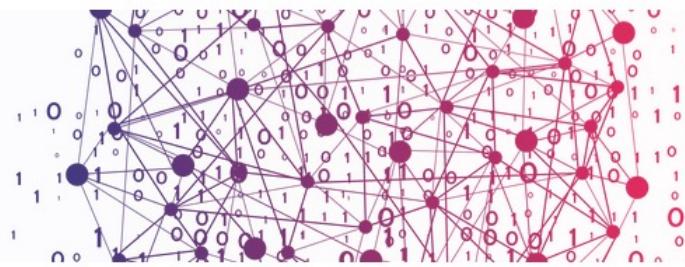
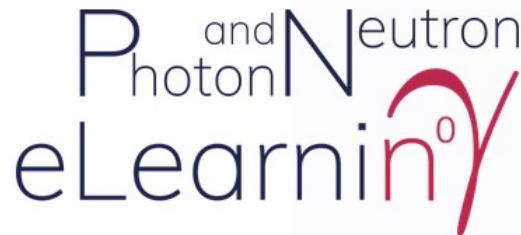
photon and neutron
open science cloud



Achieving 100% Open Educational Resources:

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

The screenshot shows the top navigation bar of the Pan-learning.org website. It includes a menu icon (three horizontal lines), a logo consisting of a stylized 'n' with a Greek letter, and a dropdown menu with options: Catalogue, e-Learning, Events, and About. To the right, it says "You are not logged in. (Log in)".



Welcome to the e-Learning platform

This e-Learning platform hosts free education and training for scientists and students. Below you will find courses on both the theory of photon and neutron scattering and how to use python code or software for data reduction and modelling.

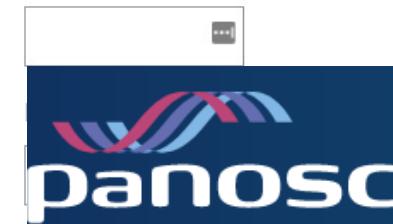


PaNOSC and E
and innovation p

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

Login

Username



FAIR data + Open Science lecture at Hercules School 2024



“The role of the scientist in making data FAIR for reproducible science”

Talk given first time as part of Hercules School in 2024

Rated: 40% Very good, 28% Good, 12% Average



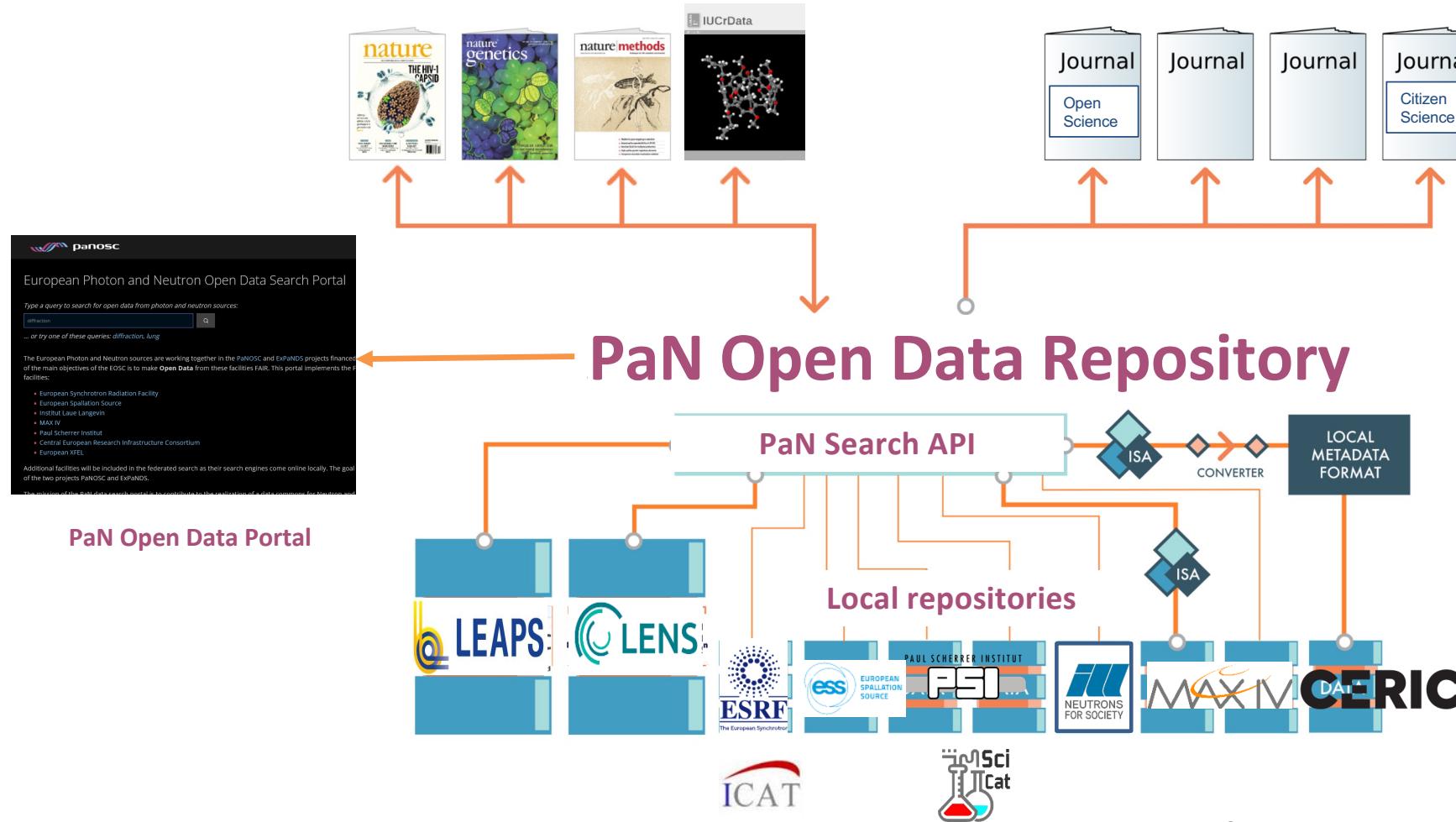
- “An interesting talk with engaging discussion. Great inclusion in the program”
- “Happy this lecture was included.”
- “**This teaching is absolutely necessary for our generation.** Although the lecture was severely lacking in nuance and very moralistic (this should definitely be improved for next time), the comments made by the students and the discussions that this sparked were really interesting and made it very enjoyable and reflective about our everyday works. Such deep discussion are much more efficient to change the way of working than “you should do this” type of lecture.”
- “This lecture was truly interesting and turn into a debate rapidly even if it was not planned. Maybe change the format to a lecture and debate could be nice for the future.”
- “Please push some of the questions/arguments offline earlier to focus on the core topic.”



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 Data from federated PaN Repositories



19

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>



Facility

Technique

Chemical Formula

Incident Wavelength

Incident Photon Energy

Temperature

Search bar: diffraction

Results: 50+ documents found

Document 1: 10.16907/e8effd03-b358-473c-9f66-fa5660b7ffb2

Advances in long-wavelength native phasing at X-ray free-electron lasers

Long-wavelength pulses from the Swiss X-ray free-electron laser (XFEL) have been used for de novo protein structure determination by native single-wavelength anomalous diffraction (native-SAD) phasing of serial femtosecond crystallography (SFX) data. In this work,...

> Details, services ... Released by **PSI** on January 1st 2020

Score: 0.900

Document 2: 10.22003/XFEL.EU-DATA-700000-00

Example Data

The European XFEL (EuXFEL) example data proposal contains experimental datasets from various original beam-times, currently covering the techniques of serial femtosecond crystallography (SFX), coherent diffraction imaging (single particle imaging, SPI), X-ray powder...

> Details, services ... Released by **EuXFEL** on January 1st 2018

Score: 0.900

Document 3: 10.5291/ILL-DATA.INTER-368

Measurement of 41K's coherent scattering length using powder diffraction

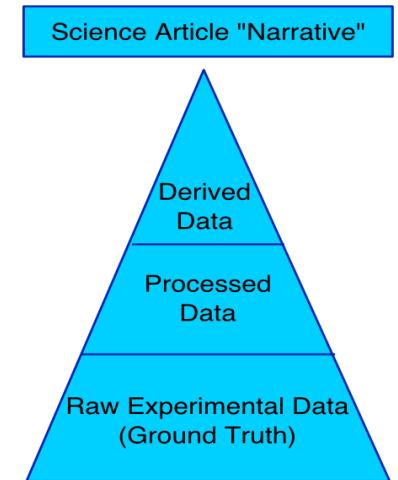
Score: 0.900

Text: 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](#), which 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.

2024 – PaNOSC continues as a community activity



PaNOSC Partners are members of IFAI LENS



Central European
Research Infrastructure
Consortium

Photon and Neutron
Open Science Cluster
→ EU Node

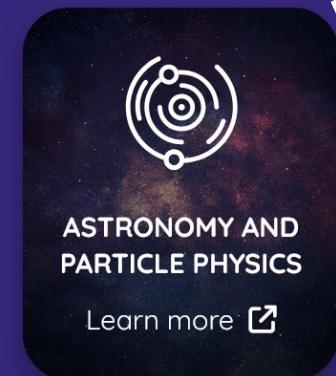


PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Cross-Domain Research Infrastructure Collaboration for Open Science

Research Infrastructures and Collaborative Initiatives

The science clusters have grown out of five collaborative projects from world-class Research Infrastructures (RIs) to the European Strategy Forum on Research Infrastructures (ESFRI). Other outcomes of the projects are cornerstones of the European Open Science Cloud (EOSC). The clusters are multidisciplinary initiatives with harmonised approaches in their specific scientific domain.

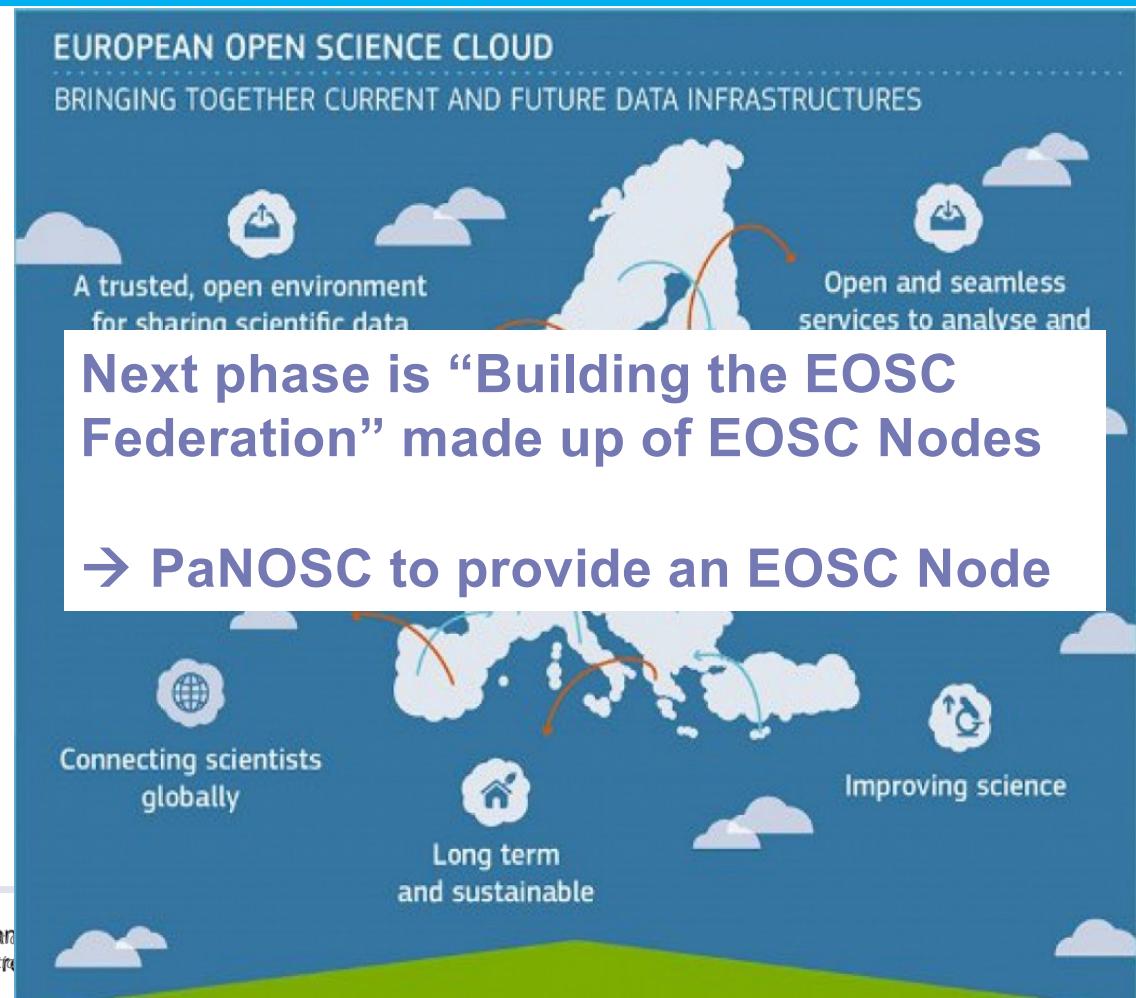


European Open Science Cloud

- 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.

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



Raw data table generated from the CIF

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 CD9_{EC2}, the structure of which was described previously [Oosterheert *et al.* (2020). *Life Sci. Alliance*, **3**, e202000883]. CD9_{EC2} crystallized in space group *P1* 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: PASS
Testing presence of archive:

Testing: All archives are accessible: PASS

Running checks with downloaded images

Testing image 4: Timage type and dimensions: PASS

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

Where PaNOSC can help ORSO activities

6th ORSO Workshop: *The output is expected to be a joint paper on Metadata and the setting up of a reflectometry data base for shared data.*

1. **FAIR data policy** – ORSO database policy issues
2. **Standardised metadata** – adoption of **Nexus/HDF5 + PaNET**
3. **JupyterLab notebooks** – provide examples + training
4. **ORSO Database** – to be included in PaNOSC EOSC Node
5. **Remote data analysis** – provide VISA service linked to processing pipelines in PaNOSC EOSC Node
6. **Reflectometry training** – register training material pan-training.eu
7. ORSO to identify a **PaNOSC contact person**



Conclusion



- 1. The Photon and Neutron Open Science Cloud (PaNOSC) is implementing a Data Commons of FAIR data in EOSC**
- 2. Facilities have the role of managing data for Users**
- 3. Users role is to enrich metadata, publish and cite data**
- 4. ORSO must continue its good work on metadata, databases + training material!**



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. ESUO, ENSA, LEAPS and LENS facilities
3. IUCr Committee on Data
4. Users for use cases, data and feedback
5. European Commission for funding PaNOSC + ExPaNDS
6. ESRF colleagues Jean-Francois Perrin, 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.

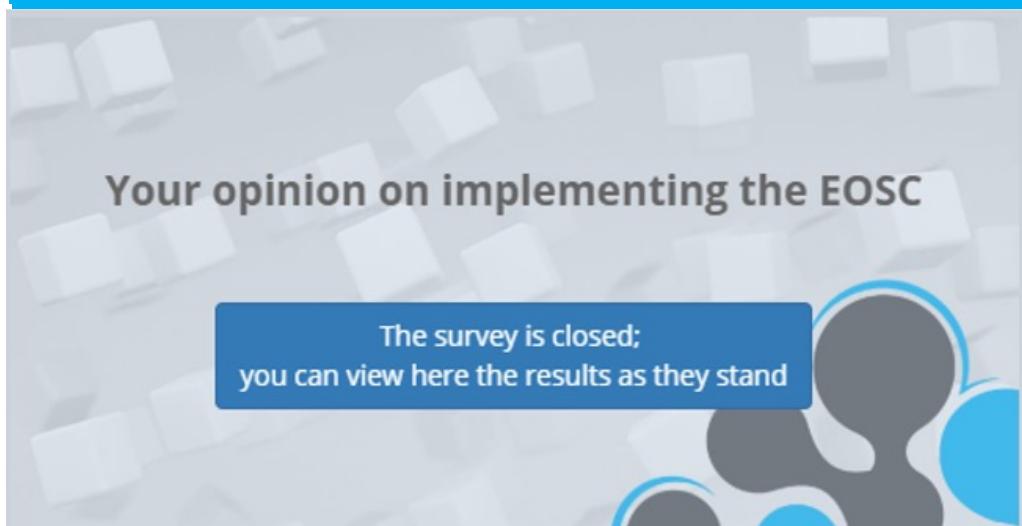
How can the EC help solve the data issues?

- 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.

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



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



Open Science with Jupyter notebooks

- Notebooks document
- If used app
- For example
- Notebooks
- Currently, I 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.



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
	P	P	WIP	WIP	WIP	WIP	P	WIP	Y	WIP	N
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



PaNOSC and ExPaNDS projects and innovation programme under

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

VINYL/
OASYS/
McStas



Pan-
learning/
training

European Photon Source
and Neutron Data Services

panosc

photon and neutron
open science cloud

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

Regular Article

THE EUROPEAN
PHYSICAL JOURNAL PLUS



LEAPS data strategy

Andy Götz^{1,a}, Erwan le Gall^{2,b}, Uwe Konrad³, George Kourousias^{4,c}, Oliver Knodel^{3,d}, Salman Matalgah^{5,e}, Oscar Matilla⁶, Darren Spruce^{7,f}, Ana Valceril Ortí⁸, Majid Ounsy⁹, Thomas H. Rod⁹, Frank Schluenzen¹⁰

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 paper 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.



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.

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

