LEAPS League of European Accelerator-based Photon Sources

LEAPS Data Strategy + Open Science

Andy Götz
ESRF + EOSC-A

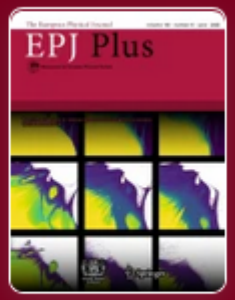
On behalf of the LEAPS WG3

18th November 2025

LEAPS data strategy

Regular Article | Open access | Published: 17 July 2023

Volume 138, article number 617, (2023) Cite this article



The European Physical Journal Plus

Andy Götz, Erwan le Gall, Uwe Konrad, George Kourousias, Oliver Knodel, Salman Matalgah, Oscar Matilla, Darren Spruce, Ana Valceril Orti, Majid Ounsy, Thomas H. Rod & Frank Schluenzen

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.

Special issue on LEAPS in European Physical Journal Plus

Home › Collection

Focus Point on Accelerator-based Photon Science Strategy, Prospects and Roadmap in Europe: a Forward View to 2030.

Participating journal: The European Physical Journal Plus

The League of European Accelerator-based Photon Sources (LEAPS) is an alliance of all synchrotron and free electron laser user facilities in Europe. These facilities are an essential element of the scientific landscape in Europe as well as worldwide, undergoing fast and profound changes of great relevance for the scientific community, the knowledge transfer to industry and the enhanced interaction with widening countries. It is the goalof LEAPS to contribute to a substantial societal impact. This focus point issue features the current status and future scientific strategy of LEAPS. The collection includes papers about most of the LEAPS members, along with two specific papers on the LEAPS general strategy and the critical aspect of data management and data driven solutions....

LEAPS data strategy

Andy Götz, Erwan le Gall ... Frank Schluenzen

Regular Article | Open access | 17 July 2023 | Article: 617

European
-based
rces

LEAPS – Fact sheet

19 facilities in 16 institutions

more than 300 operating beamlines

more than 700 Mio € operation budget

more than 24.000 users
from all disciplines and industry

free access to all academic users from all MS and worldwide

outstanding service to industry and innovation

more than 25.000 publications in the past 5 years

see ALL LEAPS documents at
https://leaps-initiative.eu/about/leaps-documents/

LEAPS League of European Accelerator-based Photon Sources

LEAPS – some figures

THE LARGEST NETWORK OF RESEARCH INFRASTRUCTURES IN EUROPE CONTRIBUTING TO EUROPEAN GLOBAL LEADERSHIP

A NETWORK OF SYNCHROTRONS AND FREE ELECTRON LASERS IN

10 countries

16 institutions

19 facilities

LEAPS actively collaborates with Analytical Research Infrastructures in Europe (ARIE).

LEAPS facilities are open science pioneers.

OUR ACTIVITIES IN NUMBERS

>300 operating Experimental stations

>30.000 user / year from all EU and beyond

>1 million hours beam time / year

>7.000 publications / year

55.000 Protein Data Bank entries supporting health industry

Materials for Energy
Life-science - Pharmacy Food
Information Technology
Environment Materials

FUNDING (reference period 2021-2027)

800M€ (yearly) operational budget

1,6B€ (over this period) investments

1,5B€ (over this period) upgrade programs (partly already funded)

LEAPS League of European Accelerator-based Photon Sources

Six Working Groups and two Strategy Groups
hundreds of people involved

Technological + Networking

- **WG1 – Beamline technology**
- Detectors
- Optics and BL Instrumentation
- Sample Environment

NEW: Photon Diagnostics

- **WG2 – Photon Sources**
- Compact Sources
- FEL Developments
- Storage Rings

- **WG3 – DATA management and software**

- **WG4 – Industry & innovation**
- **WG5 – User Services & Impact**
- **WG6 – Education, Training & Outreach**

Strategy Groups
- **SG for Synchrotrons**
- **SG for FELS**

Each WG has representatives from all LEAPS members – One/two spokespersons appointed by the WG itself – Taking lead in LEAPS-INNOV

- Develop and periodically update a roadmap for key enabling technologies and an action plan for their specific fields in consultation with the SGs;
- initiate, support implementation and monitor joint projects in their field. This includes setting up new projects, i.e. writing proposals in reply to open calls by the EU;
- exchange knowledge between LEAPS Members, Associates, Partners and to external stakeholders;
- deal with any other issues identified by the GA which requires the specific expertise of the particular WG.

SGs support the GA in its strategic work such as the roadmap process

LEAPS League of European Accelerator-based Photon Sources

LEAPS Partners and Associates

Canadian Light Source | Centre canadien de rayonnement synchrotron

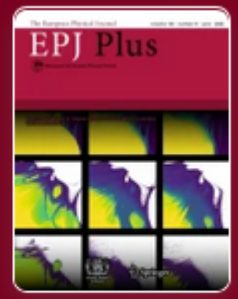
SESAME

LEAPS League of European Accelerator-based Photon Sources

LEAPS data strategy

Regular Article | Open access | Published: 17 July 2023

Volume 138, article number 617, (2023) Cite this article



The European Physical Journal Plus

Andy Götz, Erwan le Gall, Uwe Konrad, George Kourousias, Oliver Knodel, Salman Matalgah, Oscar Matilla, Darren Spruce, Ana Valceril Orti, Majid Ounsy, Thomas H. Rod & Frank Schluenzen

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.

Goals

1. Increasing efficiency of experiments

2. Open Science

3. Sustainability of solutions

LEAPS Data Strategy - summary

1. Data are essential to science, even more so for producers of huge data like the LEAPS facilities → **a strategy for (raw+processed) data is essential**
2. The Paper proposes **twenty objectives in seven areas** that the LEAPS facilities management should endorse and adopt
3. Implementing the objectives will first and foremost serve the creators of the data, i.e., **experimental teams, to deal more efficiently with data** they produce to do **better science**
4. Data should adhere to the **FAIR principles** and **Open Science** practices to share them with the scientific community and to **maximize their reuse**
5. The data strategy aims to **minimize** the amount of **data which are not published** and thereby contribute to **improving** the **efficiency** in terms of **scientific publications** and **energy consumption** of the LEAPS facilities
6. Provide high quality data for AI/ML algorithms while ensuring sovereignty

Increasing efficiency of experiments = Objectives

[3.1] LEAPS facilities to setup a working group for sharing data processing and analysis software and workflows for existing codes and new algorithms.

[3.2] LEAPS facilities to share development of new AI/ML-based data processing codes.

[3.3] LEAPS facilities to continue the development a remote analysis platform.

[3.4] LEAPS facilities to work closely with user communities to define common metadata and strategies for data formats, compression, and publication.

[3.5] LEAPS facilities to harmonize their data policies in terms of scope and embargo period of data.

Increasing efficiency of experiments = Status

[3.1] LEAPS facilities to setup a working group for sharing data processing and analysis software and workflows for existing codes and new algorithms.

[3.2] LEAPS facilities to share development of new AI/ML-based data processing codes.

[3.3] LEAPS facilities to continue the development a remote analysis platform.

[3.4] LEAPS facilities to work closely with user communities to define common metadata and strategies for data formats, compression, and publication.

[3.5] LEAPS facilities to harmonize their data policies in terms of scope and embargo period of data.

LEAPS League of European Accelerator-based Photon Sources

LEAPS + EOSC = Objectives

[4.1] LEAPS facilities to implement the FAIR principles for data → WIP

[4.2] LEAPS facilities to commit to adopting and sustaining the outcomes of PaNOSC and ExPaNDS.

[4.3] LEAPS facilities to participate in building the EOSC with the other EOSC science clusters.

[4.4] LEAPS facilities to adopt the practices of Open Science.

LEAPS – Open Science Projects

EXPANDS

Since 2019, the European Open Science Cloud (EOSC) Photon and Neutron Data Service (ExPaNDS) project including, 10 LEAPS and LENS national facilities work together to make their data more open and FAIR. Many LEAPS partners have already adopted its main outcomes such as the FAIR data policy framework and self-assessment, the common metadata schema, the OAI-PMH endpoint for data catalogues or the PaN training platform.

ExPaNDS

This project receiv
Union's Horizon 20
programme unde

PANOSC

The Photon and Neutron Open Science Cloud (PaNOSC) is a European project for making FAIR data a reality in 6 European Research Infrastructures (RIs), developing and providing services for scientific data and connecting these to the European Open Science Cloud (EOSC).

panosc

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

LEAPS + European Open Science Cloud (EOSC) –

Outcomes of PaNOSC + EXPaNDS projects

1. **FAIR data policy** and **DMPs**.
2. **FAIR assessment** and common **PID** framework.
3. **Standardized metadata** (**NeXus/HDF5**, **PaN** ontologies).
4. **Federated search** API for PaN data catalogues.
5. **Open data portal** for searching + downloading data.
6. **Community AAI** UmbrellaId → myAccessID
7. **JupyterLab notebooks** and **NeXus/HDF5** files visualization → H5Web
8. **Remote data analysis** with **VISA** + data analysis workflows → EWOKS.
9. **Simulation software** for experiments and data (Oasys, SIMEX, McStas)
10. PaN-training platform (**pan-training.eu**)

LEAPS + EOSC = Status

[4.1] LEAPS facilities to implement the FAIR principles for data → WIP

[4.2] LEAPS facilities to commit to adopting and sustaining the outcomes of PaNOSC and ExPaNDS.

[4.3] LEAPS facilities to participate in building the EOSC with the other EOSC science clusters.

[4.4] LEAPS facilities to adopt the practices of Open Science.

LEAPS League of European Accelerator-based Photon Sources

PaNOSC EOSC Node – on-going

EOSC Node | PaNOSC

Physical Sciences & Engineering

The PaNOSC (Photon and Neutron Open Science Cloud) EOSC Node federates data and services from Europe's photon and neutron research infrastructures.

It builds on the PaNOSC and ExPaNDS projects, integrating 11 facilities and their catalogues, AAI, VREs, training, and visualisation services. The Node will enhance FAIR data availability, enable AI/ML applications, and strengthen Open Science practices across domains. It connects with EOSC EU Node and other thematic/national nodes to maximise scientific and societal impact.

EOSC Node | PaNOSC homepage

Key objective: To provide FAIR, high-quality data and services from photon and neutron RIs into the EOSC Federation, increasing scientific impact, adoption of FAIR, and cross-node collaboration.

Science areas: Photon and neutron science, including X-ray crystallography, neutron scattering, CryoEM, spectroscopy, palaeontology, imaging, materials science, life sciences.

https://eosc.eu/building-the-eosc-federation/eosc-node-panosc/

PaNOSC EOSC Node – home page

Overview of PaNOSC Node Services

Content Provider

- ESRF – European Synchrotron ... 4
- Helmholtz-Zentrum Dresden-R... 3
- Paul Scherrer Institute (PSI) 3
- ESS – European Spallation Sou... 2
- MAX IV 2
- ALBA 1
- Deutsches Elektronen-Synchrot... 1
- Elettra 1
- Institut Laue-Langevin (ILL) 1
- PaNOSC 1

Keywords

- PaNOSC 17

19 results found

Search

MAX IV Open Data Repository

The MAX IV Open Data Repository provides

(Show more)

MAX IV Metadata Catalogue SciCat

MAX IV SciCat is the metadata catalogue for

(Show more)

AiiDAlab Quantum ESPRESSO app

AiiDAlab, a Jupyter-based web platform powered by the

(Show more)

PSI Data Portal

Find, access, and reuse public data acquired at

(Show more)

PaN-Finder

The PaN-Finder is a discovery portal that

(Show more)

PaN-Space

The PaNOSC User Space is an integrated online environment

(Show more)

Materials Cloud Archive

Open repository for research data relevant to

(Show more)

ESRF VISA Service

The ESRF VISA (Virtual Infrastructure for Scientific

(Show more)

Elettra Open Access Data Portal

Opendata Portal for the Elettra Sincrotrone Trieste

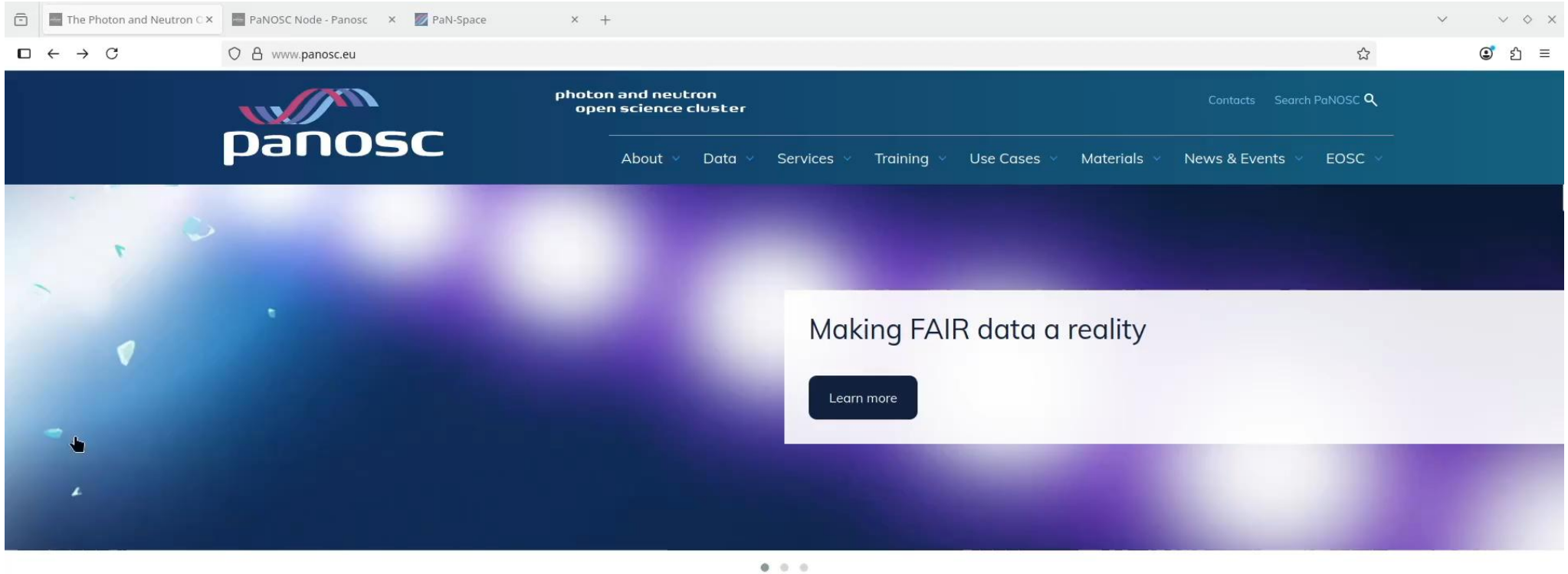
(Show more)

ALBA synchrotron Data Portal

https://eosc.panosc.eu/

LEAPS League of European Accelerator-based Photon Sources

PaNOSC EOSC Node – preview



The Photon and Neutron Open Science Cluster (PaNOSC)

The Photon and Neutron Open Science Cluster (PaNOSC) is the Science Cluster representing Photon and Neutron European Research Infrastructures (RIs), developing and providing services for its scientific community and connecting these to the European Open Science Cloud (EOSC).

LEAPS League of European Accelerator-based Photon Sources

PaN Data Commons = Objectives

[5.1] LEAPS facilities to implement the **PaN Data Commons** based on a federated solution for searching and accessing open data.

[5.2] LEAPS facilities to **link their data repositories** to the PaN Data Commons to expose open data according to the FAIR principles.

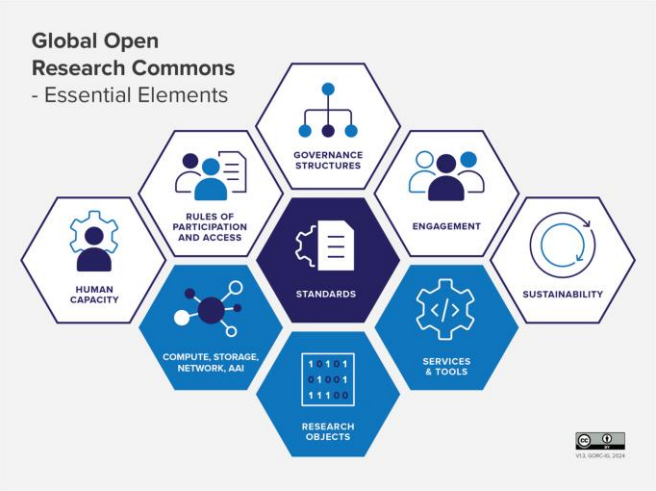
[5.3] PaN Data Commons to **monitor** the number of searches, downloads and citations of data published.

PaN Data Commons = Status

[5.1] LEAPS facilities to implement the **PaN Data Commons** based on a federated solution for searching and accessing open data.

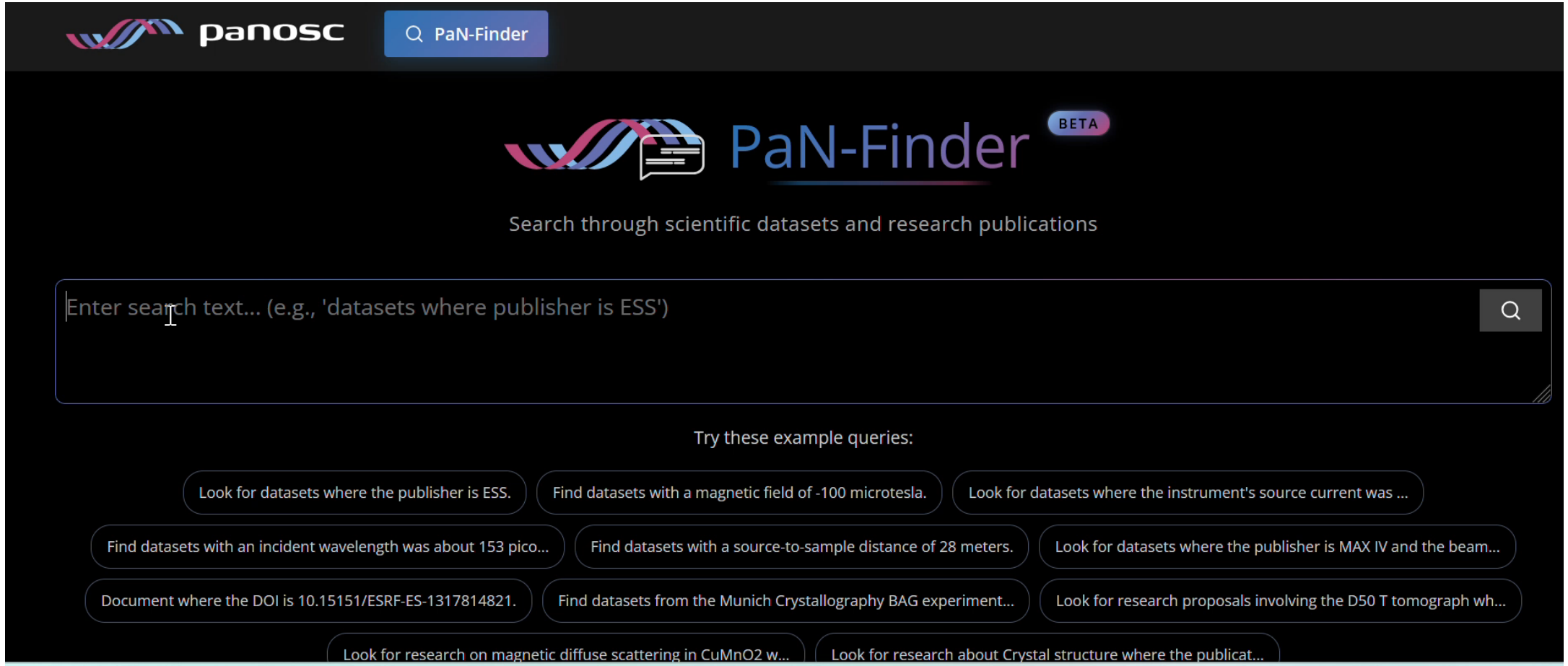
[5.2] LEAPS facilities to **link their data repositories** to the PaN Data Commons to expose open data according to the FAIR principles.

[5.3] PaN Data Commons to **monitor** the number of searches, downloads and citations of data published.



LEAPS League of European Accelerator-based Photon Sources

PaN-Finder – an AI search engine

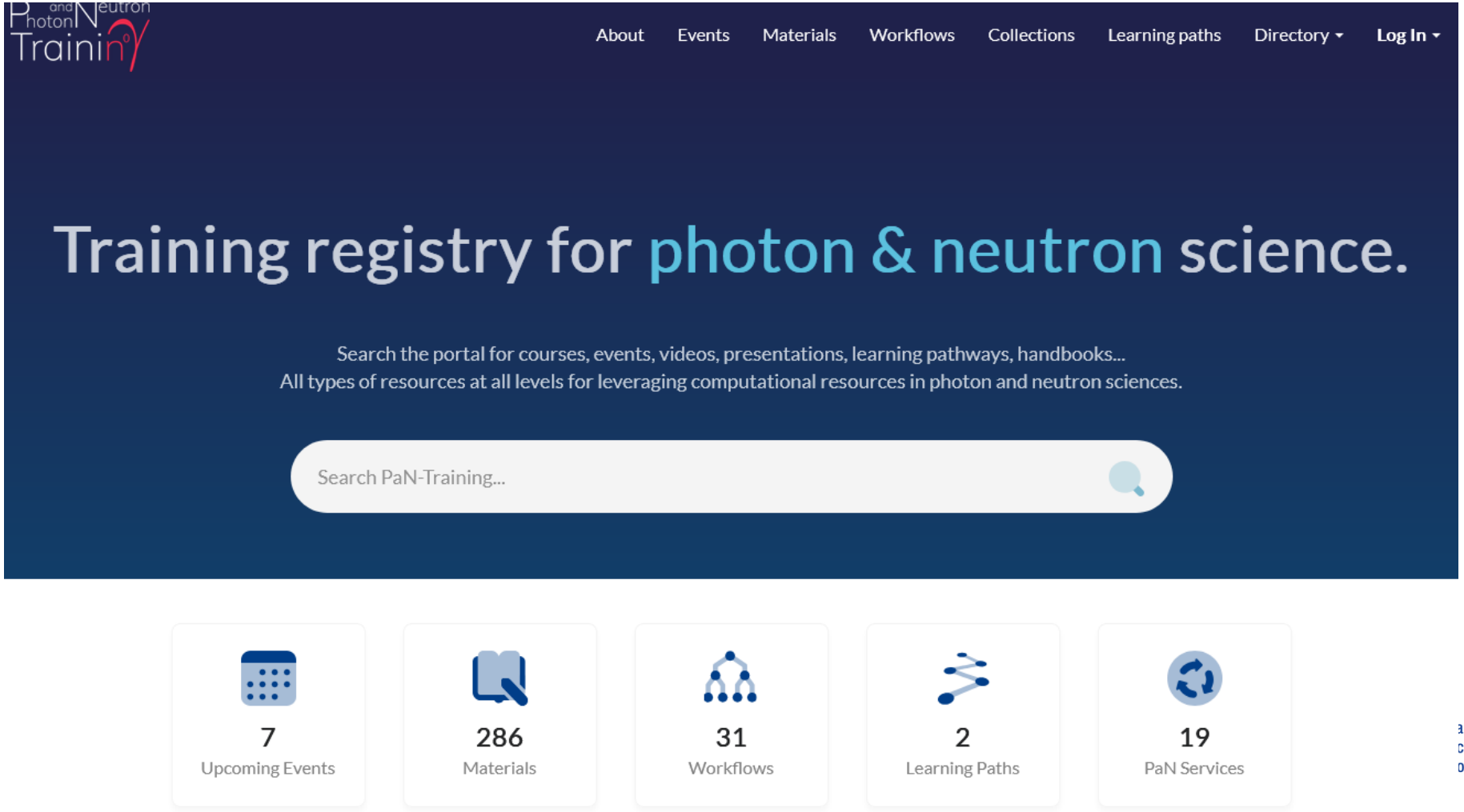


https://pan-finder.panosc.ess.eu/

LEAPS League of European Accelerator-based Photon Sources

Training + E-learning = Objectives

[6.1] LEAPS facilities should adopt the PaN-training portal as the **standard community platform** for sharing training and e-learning material on photon science → https://pan-training.tesshub.hzdr.de/



Sustainable Software = Objectives

[7.1] LEAPS facilities will continue to maintain the common software catalogue of software for PaN facilities to actively promote common software [1]

… as well as data processing methods and workflows

[7.2] LEAPS facilities will set up official collaborative projects supported by formal MoUs to foster collaboration on specific outcomes of the PanOSC and EXPaNDS projects, e.g., VISA platform, FAIR data management, search APIs, data portal, data catalogues, DMPs, etc.

[7.3] LEAPS facilities will continue the work launched in the frame of LEAPS-INNOV WP7 for data compression and reduction to develop common validated algorithms to help address the data deluge

Sustainable Software = Objectives

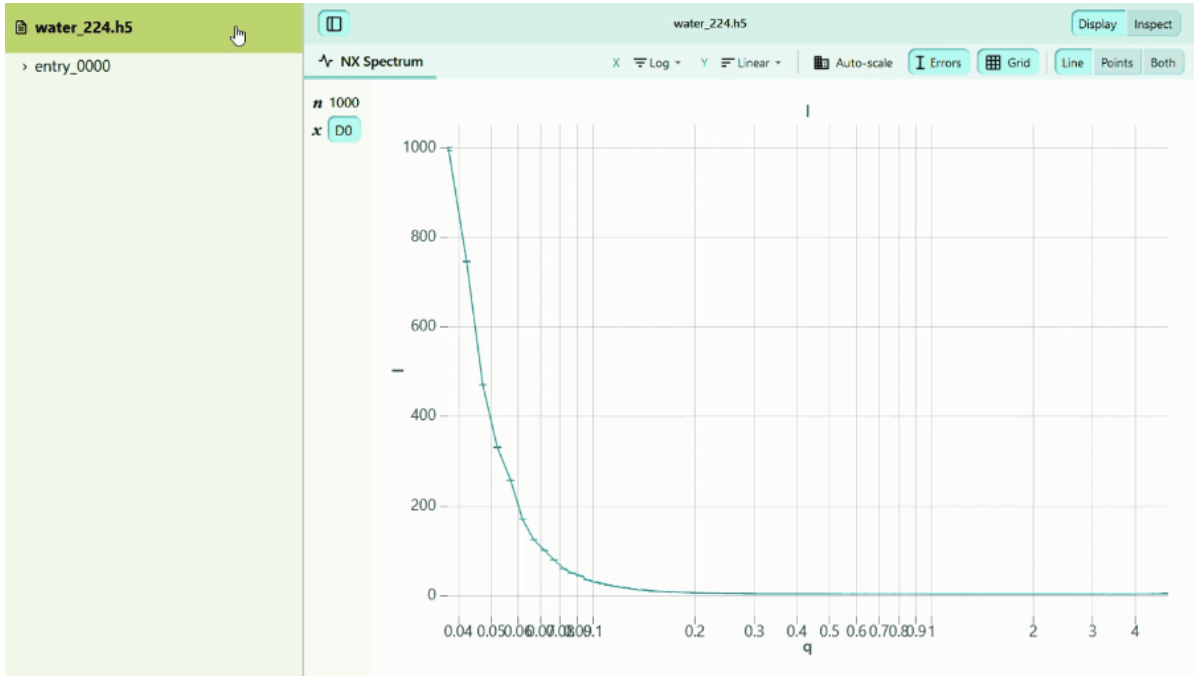
[7.1] LEAPS facilities will continue to maintain the common software catalogue of software for PaN facilities to actively promote common software [1]

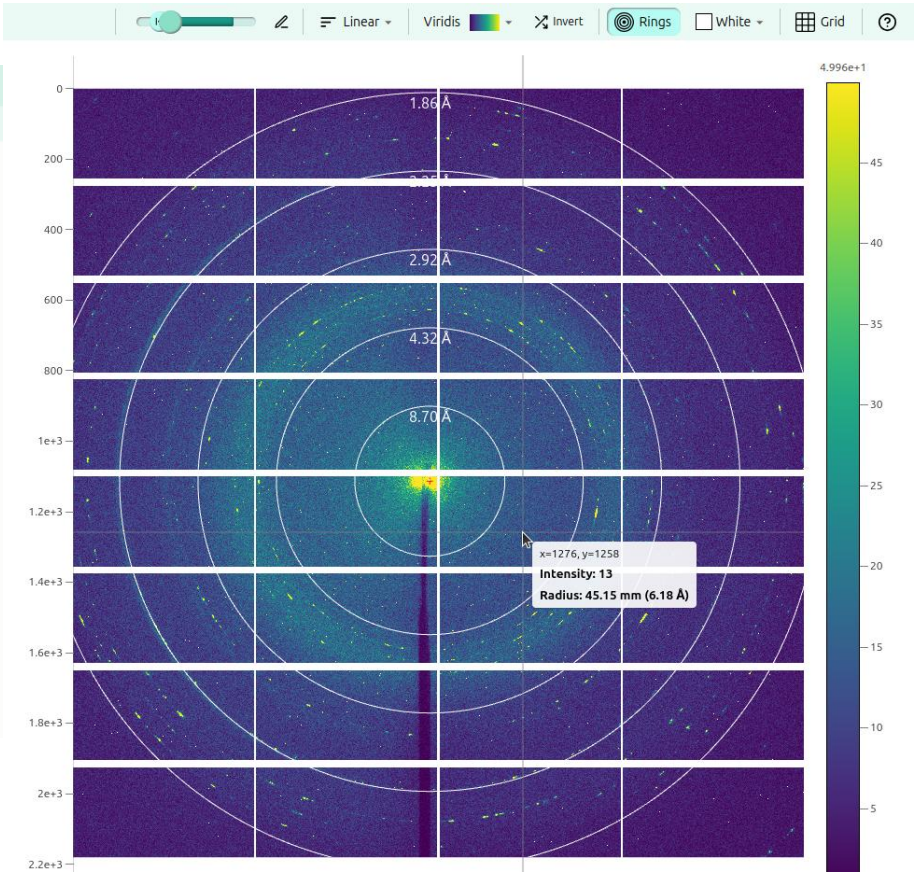
… as well as data processing methods and workflows → WorkflowHub

[7.2] LEAPS facilities will set up official collaborative projects supported by formal MoUs to foster collaboration on specific outcomes of the PanOSC and EXPaNDS projects, e.g., VISA platform, FAIR data management, search APIs, data portal, data catalogues, DMPs, etc. → LEAPS WG3 [2]

[7.3] LEAPS facilities will continue the work launched in the frame of LEAPS-INNOV WP7 for data compression and reduction to develop common validated algorithms to help address the data deluge → hdf5plugin [3]

H5Web Visualization Ecosystem



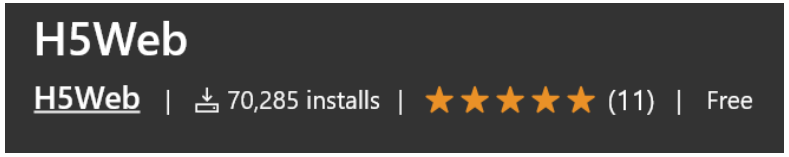


H5Web Generic HDF5 file viewer

- Integrated into **ESRF data portal (as a service)**, for viewing data in HDF5 files from experiments
- Available as **JupyterLab** and **VS Code extensions**, and soon as part of stand-alone web service, **myHDF5**, for viewing local and hosted HDF5 files

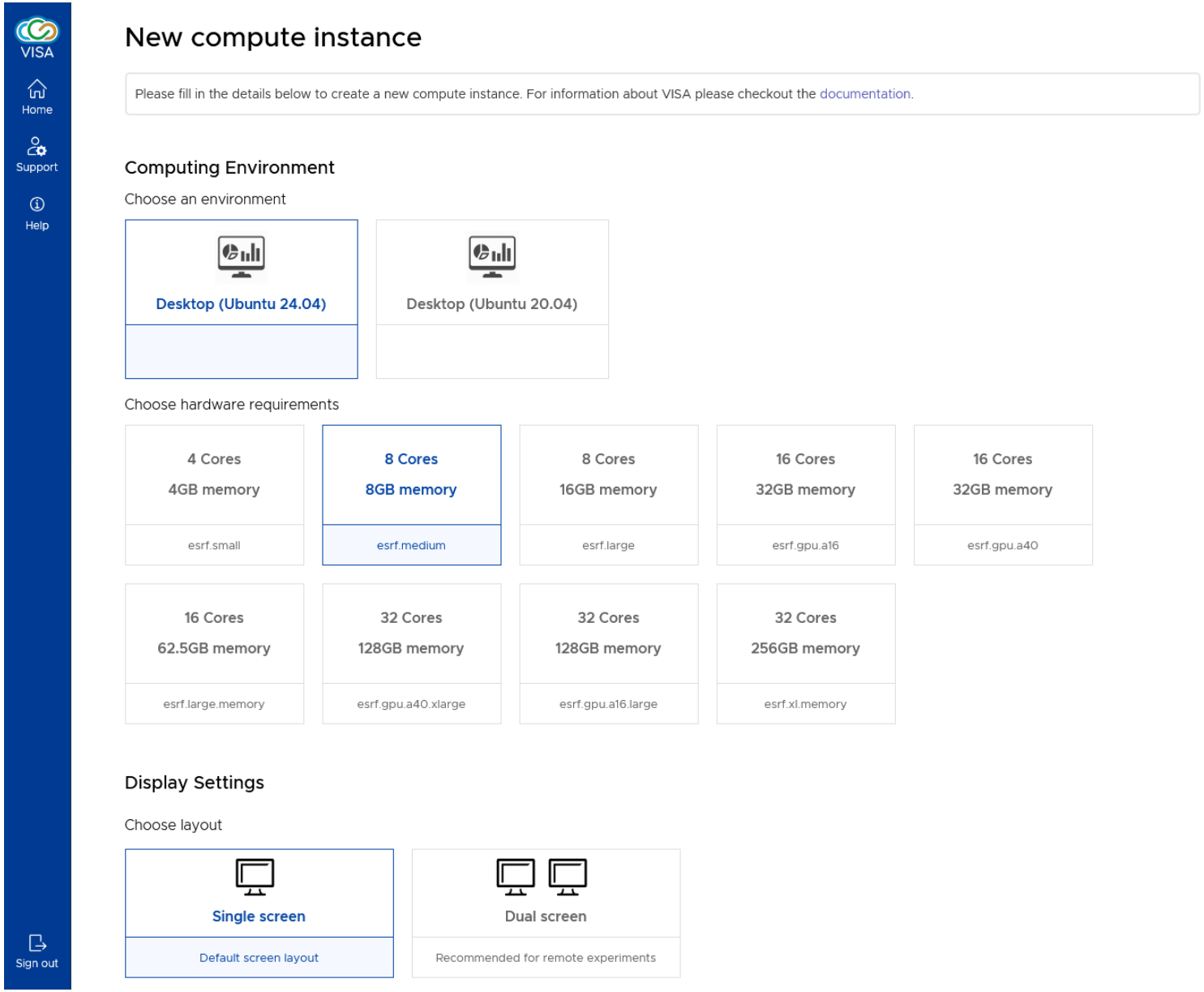
Used in various web applications at ESRF including:

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



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

VISA – Virtual Remote Environment



LEAPS League of European Accelerator-based Photon Sources

Climate Change + Green IT = Objectives

[8.1] LEAPS facilities to reduce the volumes of data using innovative data compression schemes, e.g., by continuing the work started in the LEAPS-INNOV EC funded project.

[8.2] LEAPS facilities to improve the remote operation solutions and to encourage their use to reduce travel to the facilities.

[8.3] LEAPS facilities will measure their carbon footprint equivalent and take measures to reduce it with HW and SW suppliers.

Climate Change + Green IT = Status

[8.1] LEAPS facilities to reduce the volumes of data using innovative data compression schemes, e.g., by continuing the work started in the LEAPS-INNOV EC funded project.

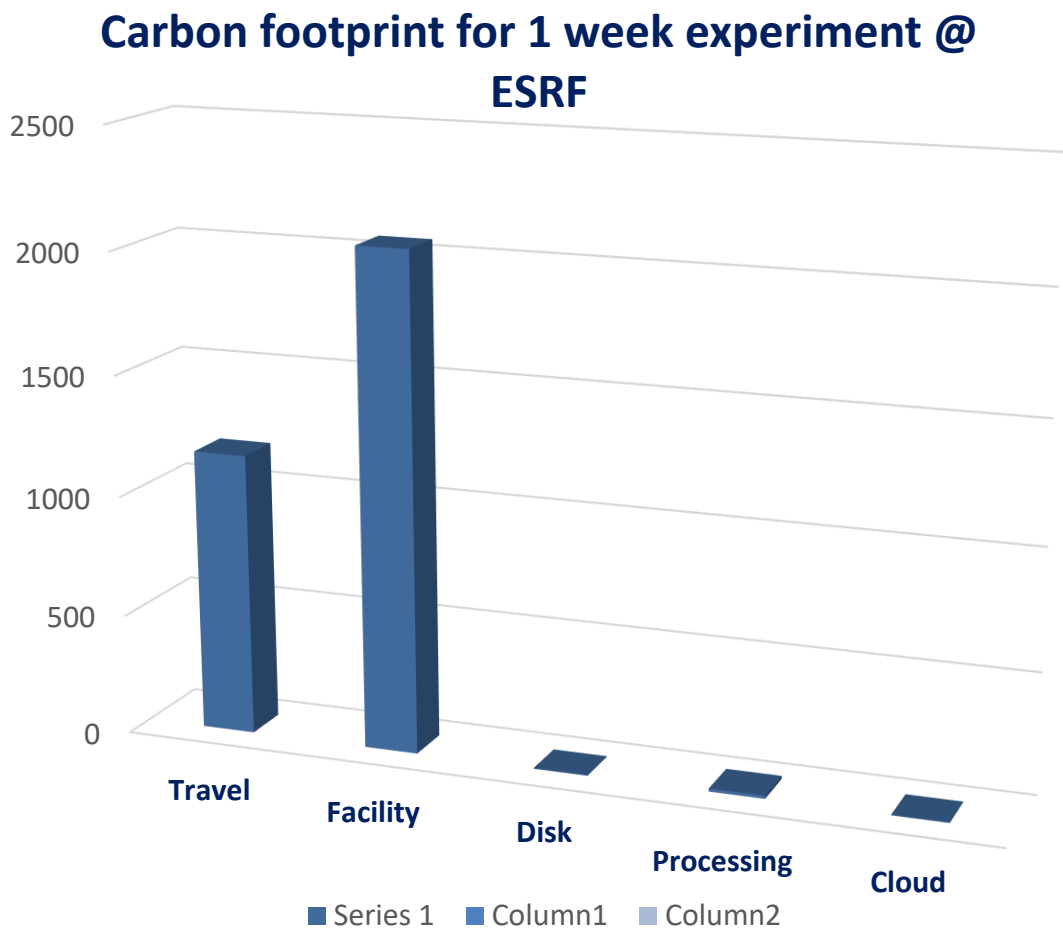
[8.2] LEAPS facilities to improve the remote operation solutions and to encourage their use to reduce travel to the facilities.

[8.3] LEAPS facilities will measure their carbon footprint equivalent and take measures to reduce it with HW and SW suppliers.

Estimated carbon footprint of experiment

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

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



TOTAL = 3.253 tons !

Sustainable Goal = 5 tons / human / yr

LEAPS League of European Accelerator-based Photon Sources

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
$4x10^{-6}$ of CO_2 equivalent needed
to acquire the raw data!

TAKEAWAY → TAKE CARE of DATA by making it FAIR!

LEAPS League of European Accelerator-based Photon Sources

Climate Change + Green IT = Indicators

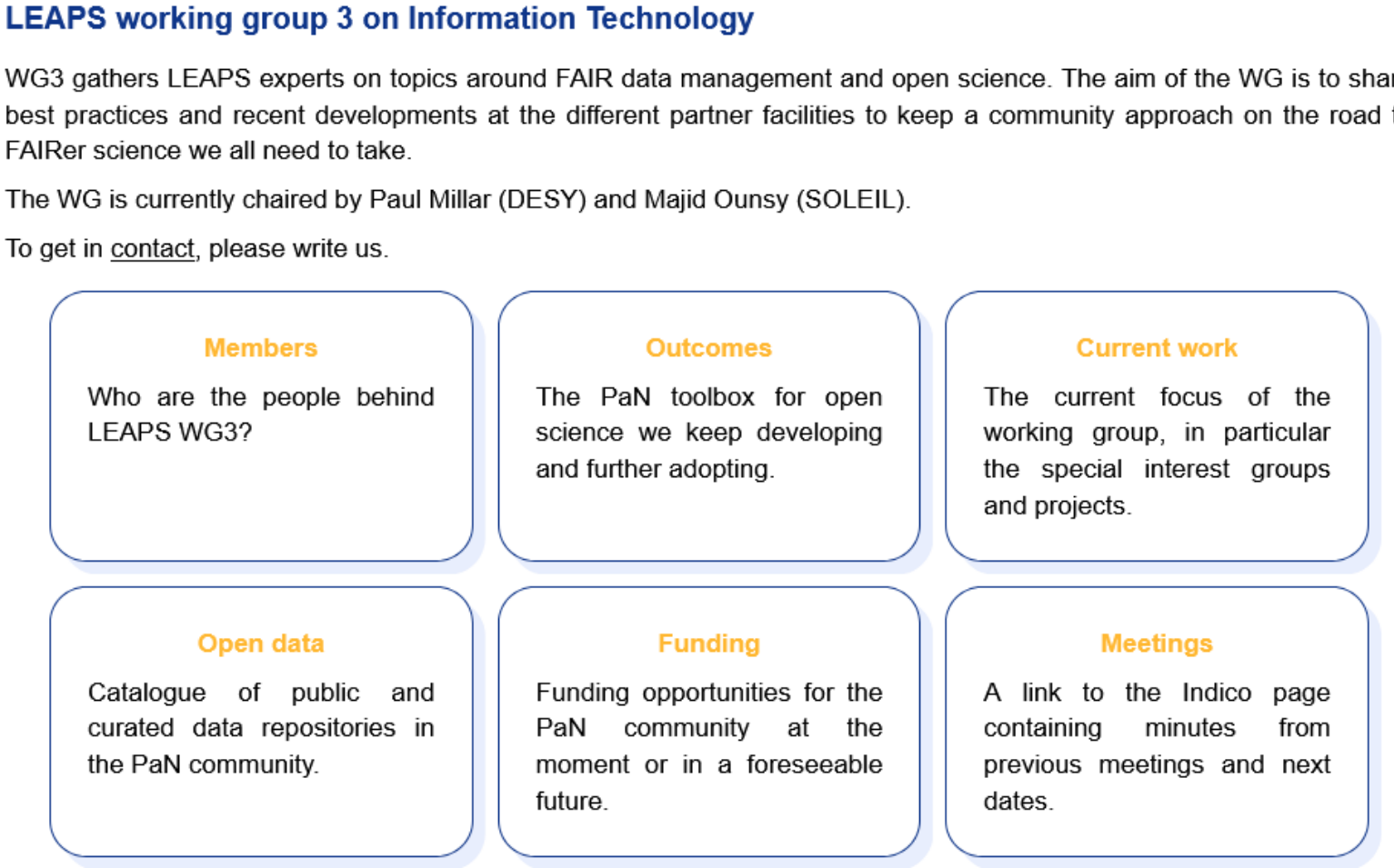
- Percentage of remote/mail-in experiments (travel footprint)
- Data center efficiency (PUE)
- Energy/carbon footprint of Data Analysis Workflows
- Energy/carbon footprint of Data Storage including data compression/decompression
- Data reuse

→ CO2 calculation to be started in next 2 years

LEAPS League of European Accelerator-based Photon Sources

Sharing know-how = Objectives

[9.1] Establishing a centralized sustained web portal to present and list all LEAPS IT collaborations → https://leaps-wg3.desy.de/



LEAPS League of European Accelerator-based Photon Sources

IT Collaborations

LEAPS First Plenary Meeting, Hamburg, Germany, 12-14 Nov. 2018

1994 · EPICS
Distributed control systems to operate particle accelerators, telescopes and other large experiments

Tango · 1999
Distributed control systems, supervisory control and data acquisition

2000 · areaDetector
Application for controlling area 2-D detectors, including CCDs, pixel array detectors, and online imaging plates

MxCube · 2005
Software platform for macromolecular crystallography experiments

2005 · LIMA
Library for image acquisition, unified control of 2D detectors

IcePAP · 2007
Motor control system for high resolution position applications

2008 · Sardana
Tango based control system for instrument control and data acquisition

ISPyB · 2009
Information System for Protein CrystallographY Beamlines

Collaborations Participants

	ESRF	ALBA	DESY	MAX IV	DIAMOND	ELETTRA	SOLEIL	HZB	SOLARIS	PSI	XFEL	HZDR
Tango	■	■	■	■		■	■		■			
MxCube	■	■	■	■		■	■	■				
Taurus	■	■	■	■		■			■			
LIMA	■	■	■	■			■		■			
Sardana	■	■	■	■					■			
EPICS			■		■			■		■		
areaDetector					■					■		
umbrellaID	■	■	■	■	■	■		■		■		■
ISPyB	■	■	■	■	■	■	■	■				
ICAT	■	■			■	■	■	■				
CalipsoPlus JRA2	■	■	■		■	■	■			■	■	■
Silx	■	■		■	■	■					■	
Software Catalog	■	■	■		■			■				
IcePAP	■	■		■					■			
#EM	■	■		■								
PanDA					■		■					

Sharing know-how = Objectives

2010 · Taurus
Tango based control system for instrument control and data acquisition

Software Catalogue · 2011
PaNdata software catalogue is a database of software used mainly for data analysis of neutron and photon experiments

2011 · ICAT
PaNdata meta-data management system for large facilities virtual labs

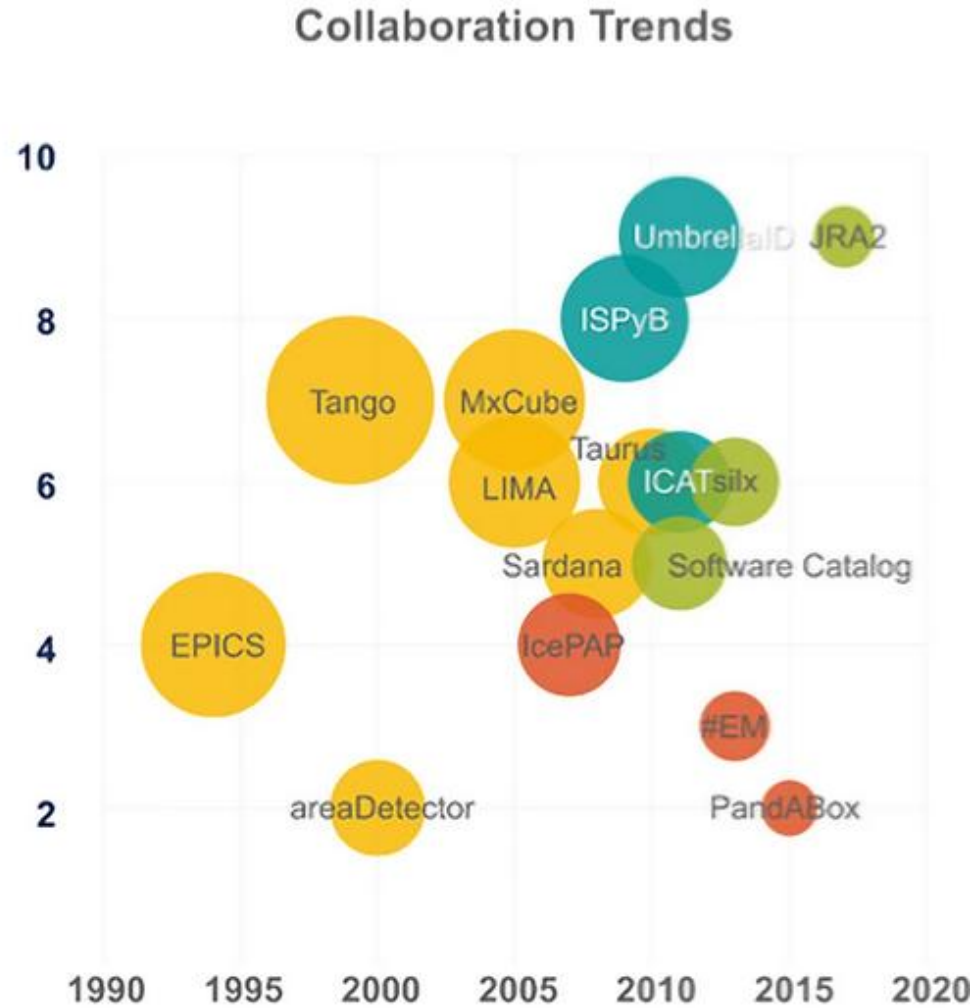
umbrellaID · 2011
PaNdata federated identity system for users of large neutron and photon facilities to access experimental information, remote control, data and in a close future centralized data analysis

2013 · silx
Python packages for data assessment, reduction and analysis applications at synchrotron radiation facilities

#EM · 2013
Fast 18 bit isolated electrometer for synchrotron applications – continuous scan, sample by sample real time calculations, etc.

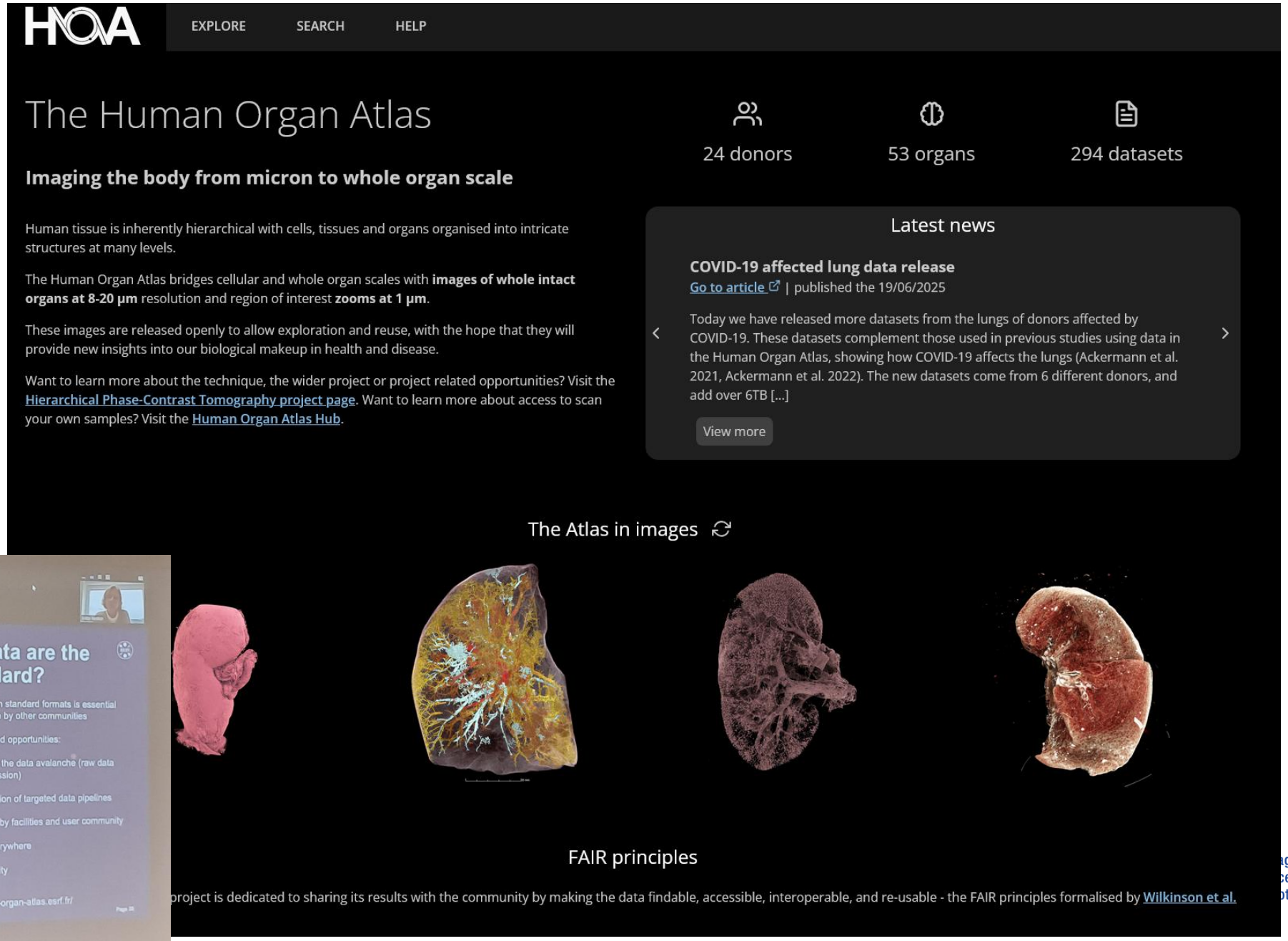
2015 · PandABox
Position and acquisition processing platform

CALIPSOplus JRA2· 2017
Demonstrator of photon science data analysis as a service (Daas)



X collaboration start
Y # participating institutions
Size years existing · # participants

Example of FAIR data – Human Organ Atlas

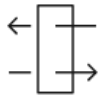


https://human-organ-atlas.esrf.fr/

What are the Advantages of a Data Strategy?

Make informed decisions - A data strategy provides a structure for using data-driven insights to inform decisions regarding business strategies, operations, planning, investments and more.

Launch AI initiatives - Applications, and especially generative AI, typically require large amounts of clean, reliable and accessible data to build, train and refine. A data strategy helps enforce data quality and data governance standards to provide trusted data for these initiatives.



Increase productivity - Data strategies can help accelerate productivity by identifying operational bottlenecks, inefficient processes, redundancies and opportunities for automating workflows.

Reduce costs - A data strategy can help reduce costs by increasing the efficiency of data storage and processing. It can also help protect data against costly breaches or regulatory compliance violations. According to the IBM Cost of a Data Breach Report, the average breach costs USD 4.88 million.

Innovate - A data strategy can yield data-driven insights into the latest trends both in and outside of the business. Organizations can use these insights to help develop innovative new products or services to take advantage of emerging market opportunities.

Outperform competitors - Data strategies help organizations harness real-time business intelligence as a strategic asset. Stakeholders can use this information to react more quickly and effectively to the latest competitive trends and tactics.

Credits: https://www.ibm.com/think/topics/data-strategy

LEAPS League of European Accelerator-based Photon Sources

Conclusions

1. **Increasing efficiency of experiments** → progressing to deal with HUGE data volumes (1 imaging experiment can produce ~1 PB / week)

2. **Open Science** → EOSC is leading the adoption of Open Science e.g. PaNOSC EOSC Node, OSCARS financing Open Science proposals

3. **Sustainability of solutions** → the adoption of the data strategy as part of the services offered by LEAPS facilities is the best (only?) way to ensure long term

Resources

1. Götz, A., le Gall, E., Konrad, U. *et al.* **LEAPS data strategy**. *Eur. Phys. J. Plus* **138**, 617 (2023). https://doi.org/10.1140/epjp/s13360-023-04189-6

2. B. M. Murphy, A. Götz, C. Gutt, C. McGuinness, H. M. Rønnow, A. Schneidewind, S. Deleddah and U. Pietschd, ***FAIR data – the photon and neutron communities move together towards open science***, in IUCrJ, 2024, DOI: https://doi.org/10.1107/S2052252524011941

3. https://www.leaps-initiative.eu/resource/leaps-technology-roadmap-2025/
4. https://leaps-wg3.desy.de/
5. https://panosc.eu

LEAPS League of European Accelerator-based Photon Sources

"The strength of LEAPS lies in its staff and users, hailing from all European countries, beyond those which host the facilities."

@leaps_initiative

@LEAPSinitiative

https://leaps-initiative.eu

Tak

Tack

Thanks

Bedankt

Danke

Dziękuję

Merci

Grazie

Gracias

شكك