Data Strategy - the future of the ExPaNDS and PaNOSC projects
PART II

LEAPS Plenary- 21/10/2021

Presenters: **Andy Götz** (ESRF) + **Patrick Fuhrmann** (DESY)

Contributors: Sophie Servan (DESY), Jordi Bodera (ESRF)
on behalf of PaNOSC + ExPaNDS partners

We have One Common **GOAL**

Our most important goal is to provide FAIR Data

- One of the **main objectives** of PaNOSC and ExPaNDS is to avoid this:

Data availability

The data supporting this study can be made available from the corresponding author upon request.

Article | Open Access | Published: 07 February 2020

4D imaging of lithium-batteries using correlative neutron and X-ray tomography with a virtual unrolling technique

Ralf F. Ziesche, Tobias Arlt, Donal P. Finegan, Thomas M. M. Heenan, Alessandro Tengattini, Daniel Baum, Nikolay Kardjilov, Henning Markötter, Ingo Manke, Winfried Kockelmann, Dan J. L. Brett & Paul R. Shearing ✉

Nature Communications **11**, Article number: 777 (2020) | Cite this article

9747 Accesses | **34** Citations | **223** Altmetric | Metrics

What Users need

- Good (meta)data + logbooks

- Performant Download services

- Digital Object Identifiers for Data

- Remote data analysis

- Access to Open Data

- Credit for Data re-use

- Data Management

What PaNOSC+ExPaNDS provide

- Metadata catalogs + e-logbooks

- Download service (e.g. Globus)

- Digital Object Identifiers for Data

- VISA Remote Analysis portal

- Open Data Portal

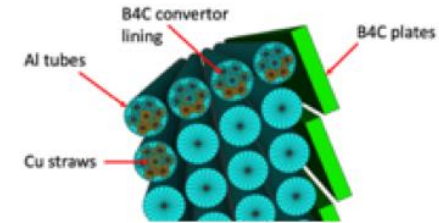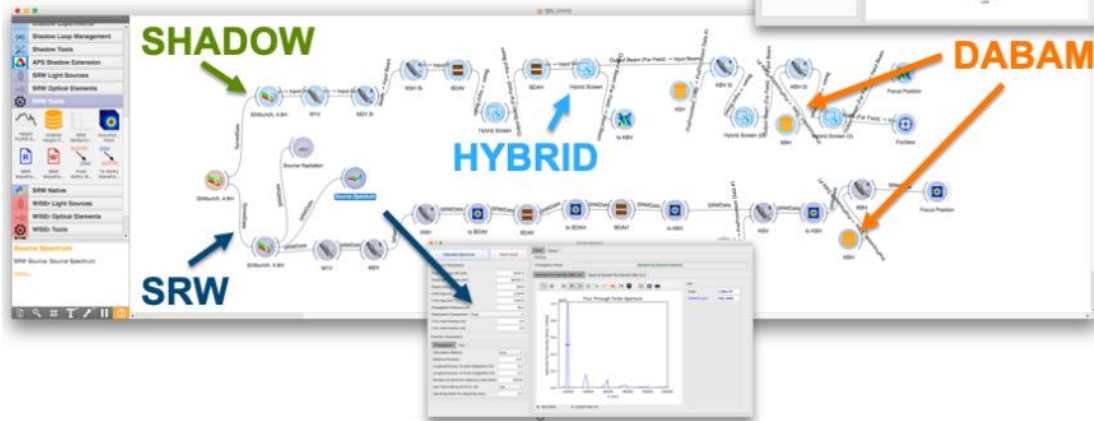- Metrics about Data Re-Use

- Data Management Plans

10 Primary Outcomes of PaNOSC and ExPaNDS

1. **FAIR data policy** and **DMPs**
2. **FAIR assessment** and common **PID** framework
3. Standardised 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**
7. **JupyterLab notebooks** and **Nexus/HDF5** files visualisation
8. **Remote data analysis** with VISA + data analysis pipelines
9. **Simulation** software for simulating experimental data (SIMEX)
10. **PaN-learning** platform (pan-learning.org + pan-training.org)

Simulation in PaNOSC provides support to OASYS + McStas

PaNOSC continued simulation activities started in EUCALL with SIMEX, lasers + neutrons

LEAPS Data should be getting more attention

- Why it's important
  - attribution of published data to our facilities: **impact**, visibility
  - globalisation of research
- What PaNOSC+ExPaNDS do
  - **FAIR data policy** framework for PaN
  - support for implementation: metadata catalogues, active **DMPs**, **PID** infrastructures...
  - open data harvested and searchable in **EOSC**

☞ cf. demo at ExPaNDS mid-term review

- What we need from LEAPS
  - commitment to FAIR data management
  - means + resources to implement data policies
  - → **recommend updating of policies and hiring data managers**

Example of open data publishing:

Human Organ Atlas

refer to PaNOSC Use Case 23 for more info

Human Organ Atlas   EXPLORE   SEARCH

Welcome to the Human Organ Atlas

The Human Organ Atlas uses **Hierarchical Phase-Contrast Tomography** (HiP-CT) 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.

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**

Aknowledgements

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

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

Users need remote data analysis

- Why it's important
  - next generation of data analysis in globalised research
  - **remote access** to facilities
- What we do
  - develop and deploy **VISA** platform
  - make **Jupyter** notebooks available at all sites
  - make data analysis pipelines **interoperable**
- What we need from LEAPS
  - commitment to VISA
  - endorse the **shift to cloud computing**

https://bit.ly/VISA-video
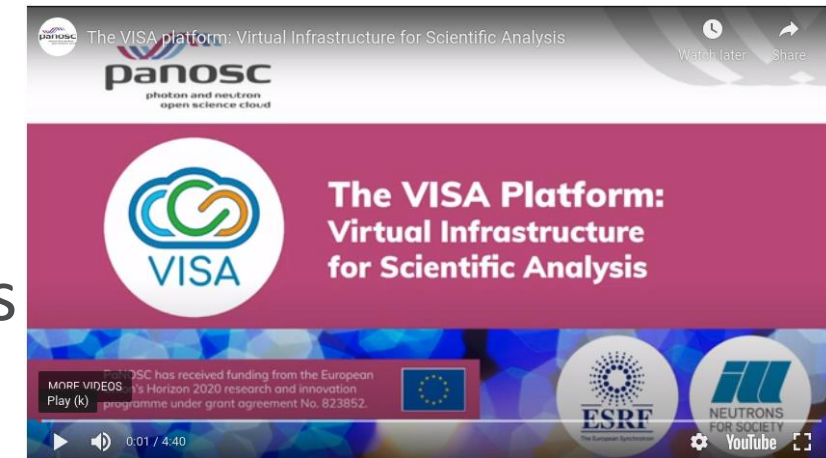
VISA remote analysis platform
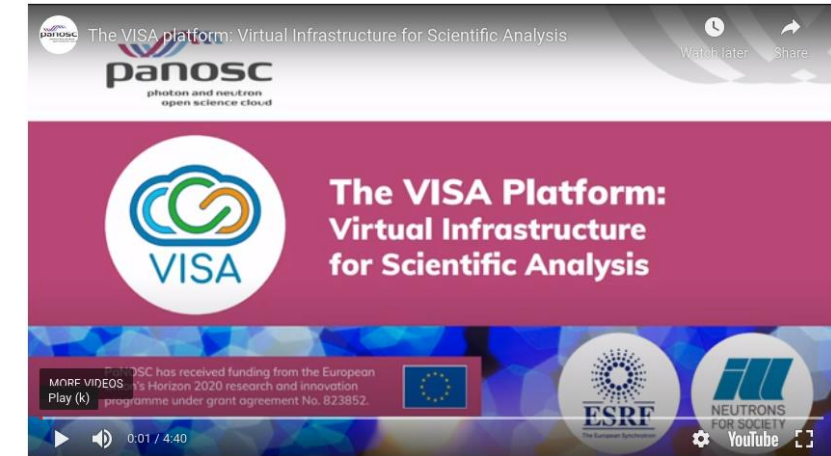
Remote analysis + visualization with Jupyter notebooks

- **JupyterLab has been widely adopted as remote analysis tool**

- **PaNOSC provides**
  - Jupyter-slurm
  - Nexus/HDF5 visualization

- **PaNOSC + ExPaNDS**
  - Developed for Use Cases

Home > All Use Cases

https://www.panosc.eu/all-use-cases/

All Use Cases

Submit your use case
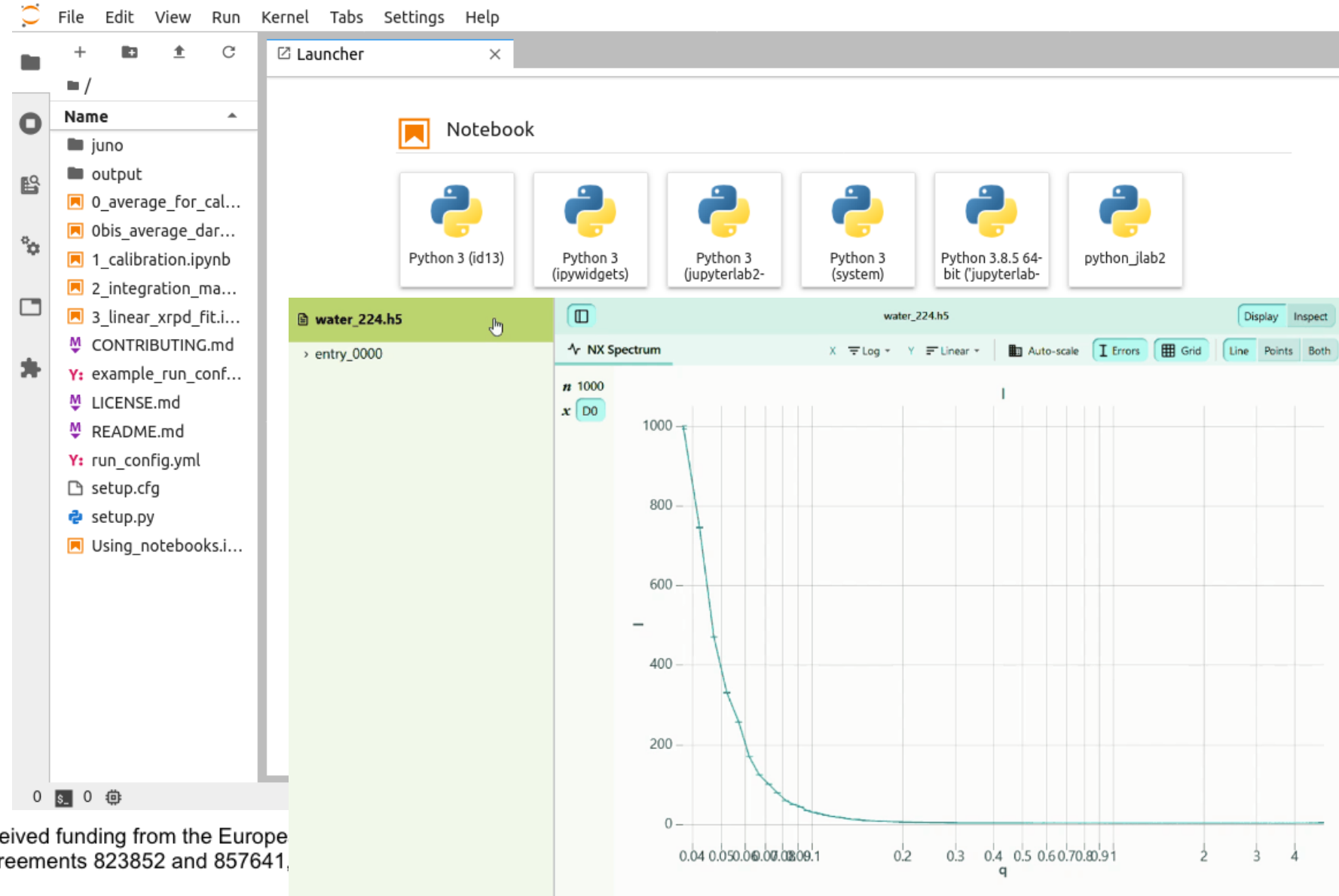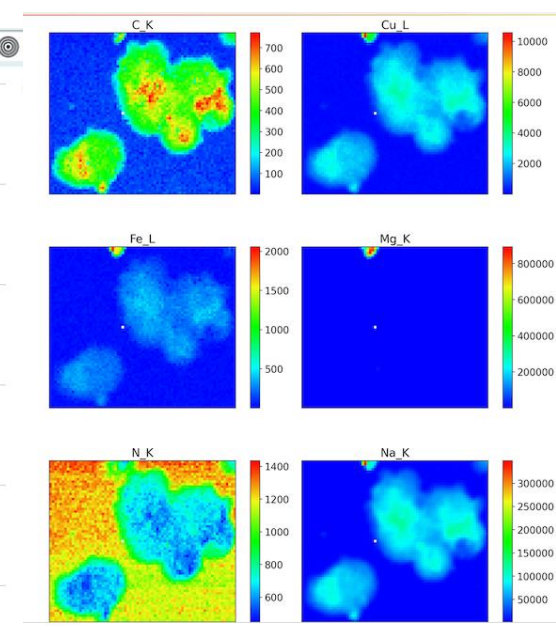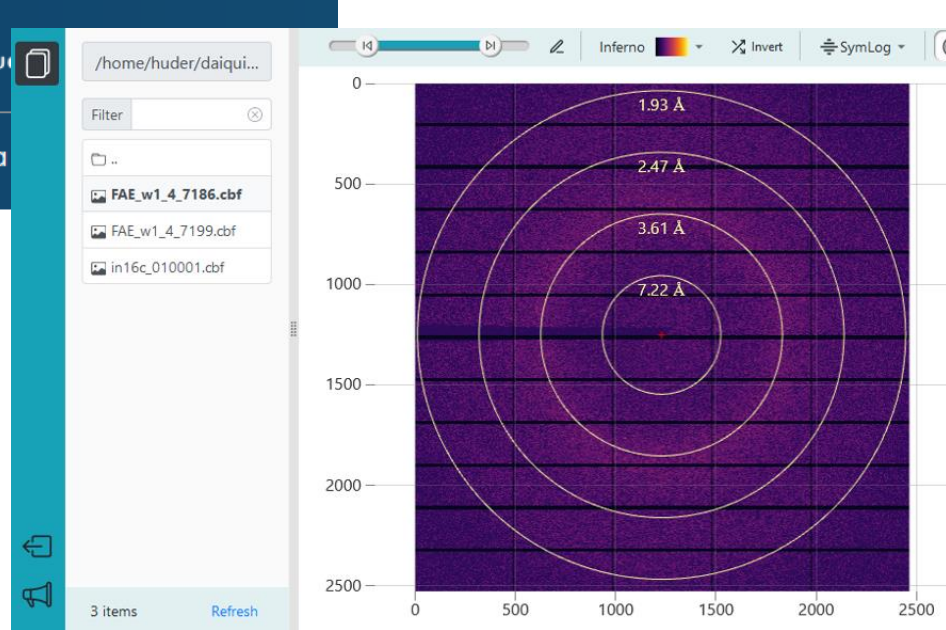
- **Use Case 22** – <u>BRAGGY diffraction image viewer</u> (ESRF; PaNOSC-related WPs: <u>WP3</u>, <u>WP4</u>, <u>WP6</u>)

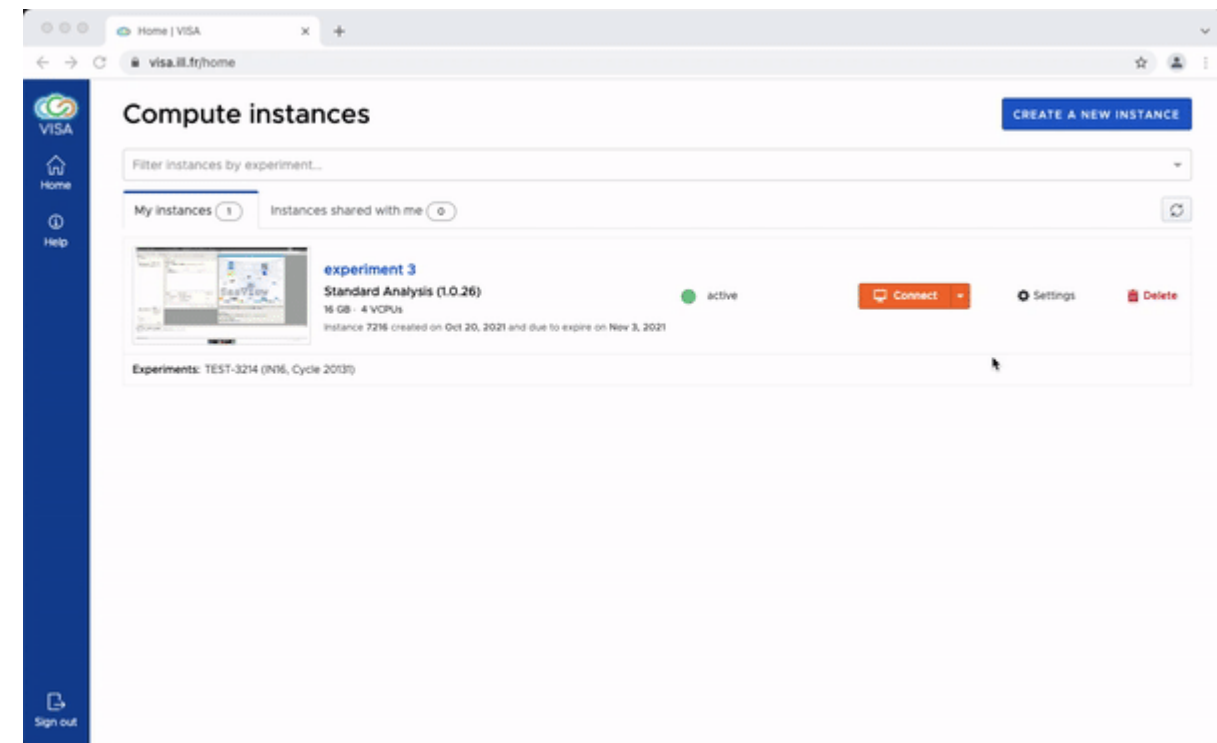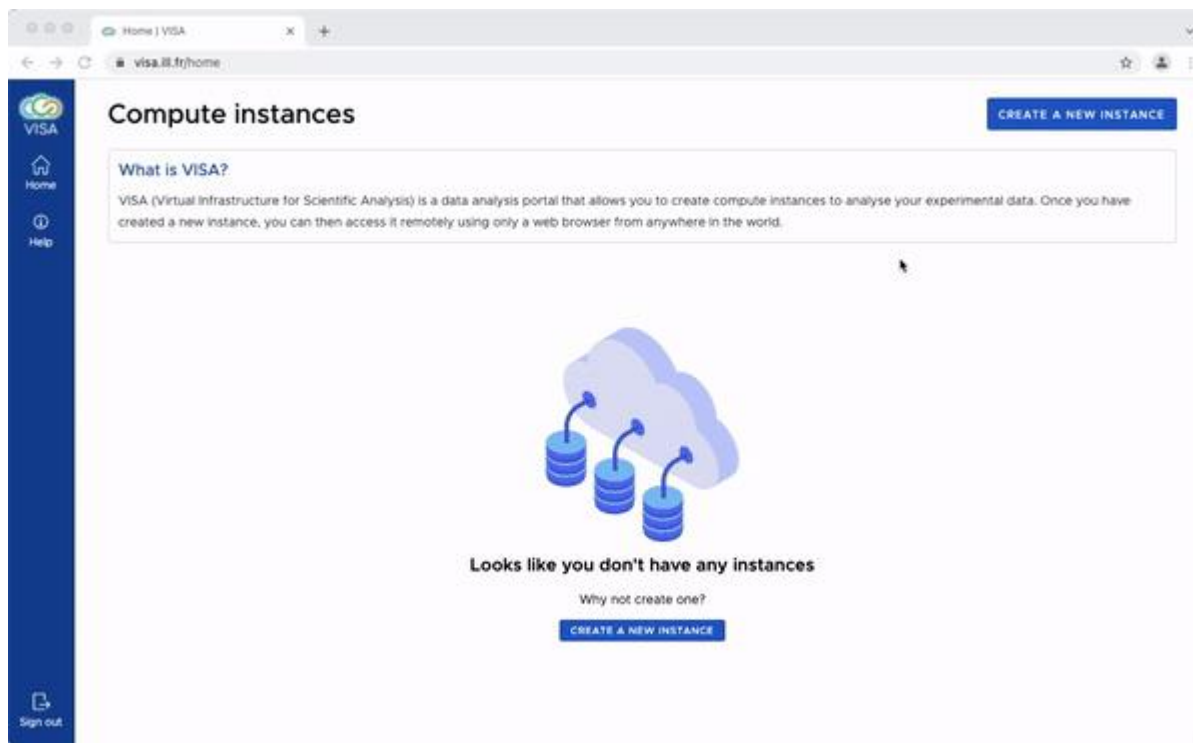- **Use Case 21** – <u>Online visualization and analysis of HDF5 PyMCA Output Files</u> (CERIC-ERIC; PaNOSC-related WP: <u>WP4</u>)

- **Use Case 20** – <u>Using an e-learning platform to support presentations</u> (European Spallation Source – ERIC, ESS-ERIC; PaNOSC-related WP: <u>WP8</u>)

- **Use Case 19** – <u>Online visualization and analysis of Bruker NMR output</u> (CERIC-ERIC; PaNOS related WP: <u>WP4</u>)

- **Use Case 18** – <u>In-silico Neutron diffraction from Boro-carbon systems: precise and reliable t for exact structural analysis and defect detection</u> (Extreme Light Infrastructure, ELI-ERIC; PaNOSC-related WP: <u>WP5</u>)

- **Use Case 17** – <u>Reusing data for validation of force fields</u> (European Spallation Source – ERIC ESS-ERIC; PaNOSC-related WPs: <u>WP3</u>, <u>WP4</u>, <u>WP5</u>)
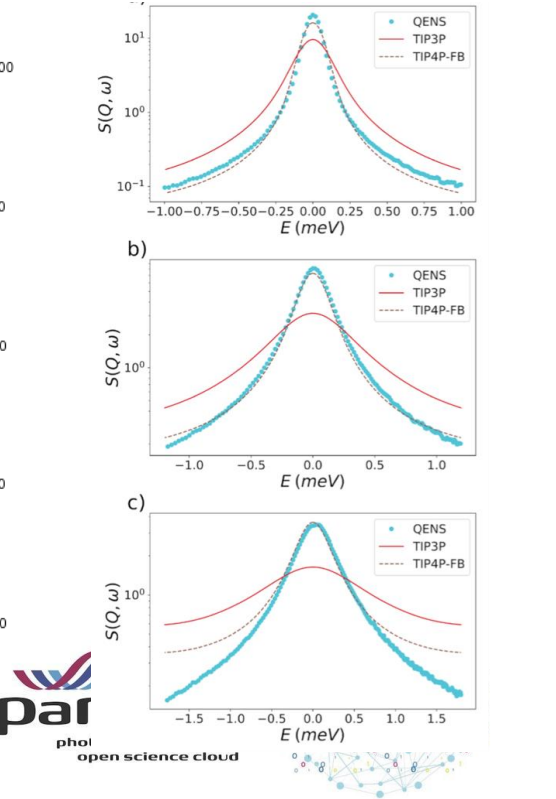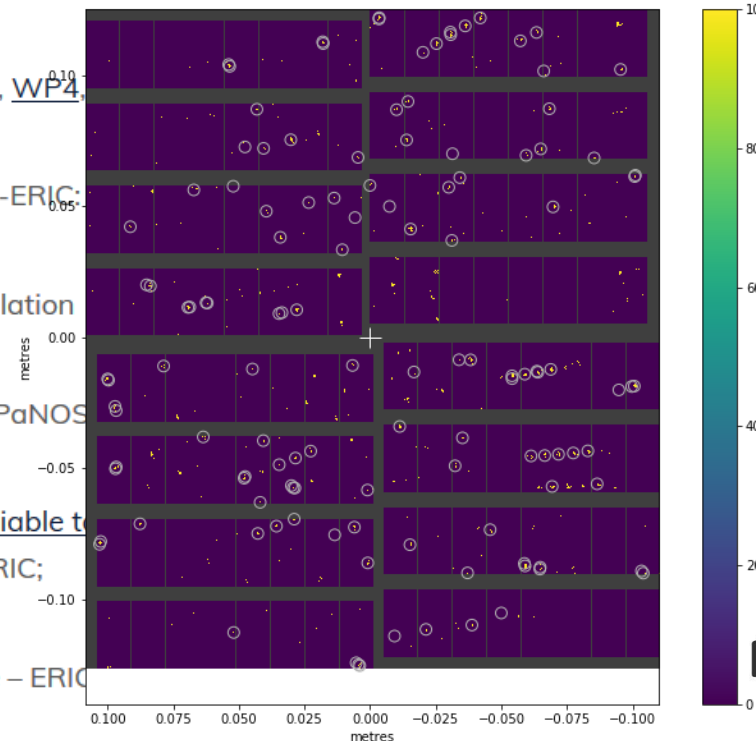
PaN Open Science Commons - Concept

- **Vision** – create a common space for PaNOSC and ExPaNDS facilities where petabytes of PaN FAIR data, analysis software, notebooks, workflows, and training material can be **F**ound, **A**ccessed (downloaded and/or executed), **R**e-Used + **I**mproved i.e. **FAIR**

- **Remote access** – the PaN commons will be accessible remotely while being executed locally (close to the data) or via the EOSC (data needs to be moved)

- **Remote users** – the PaN commons will enable and encourage remote users and experiments (urgently required in the post-COVID-19 phase)

Sustain Published Data ➡ Through PaN Repositories

PaN Open Data Portal

PaN Open Data Repository

PaN Search API

Local repositories

image: http://blogs.nature.com/scientificdata/2013/07/23/scientific-data-to-complement-and-promote-public-data-repositories/
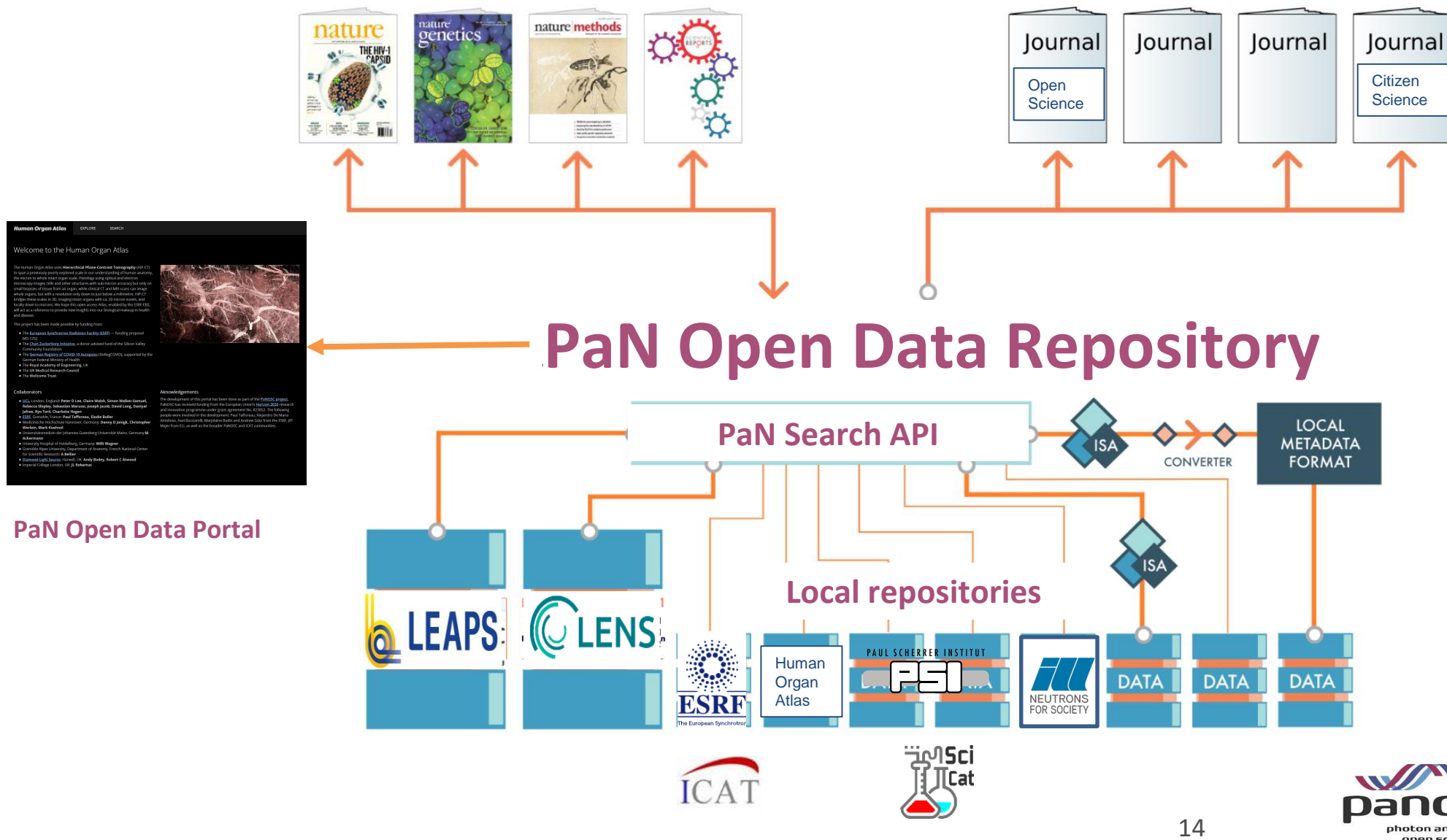
Examples from other clusters in EOSC

- **EOSC-Life**
  - EU COVID-19 portal
- **ESCAPE**
  - High Energy Physics – CERN Open Data
  - Virtual Observatory
- **ENVRI-FAIR**
  - ICOS, DISSCO, …
- **SSHOC**
  - CLARIN, DARIAH, …
- **PaNOSC**+**ExPaNDS**
  - ISIS, ILL, ESRF, PSI, …
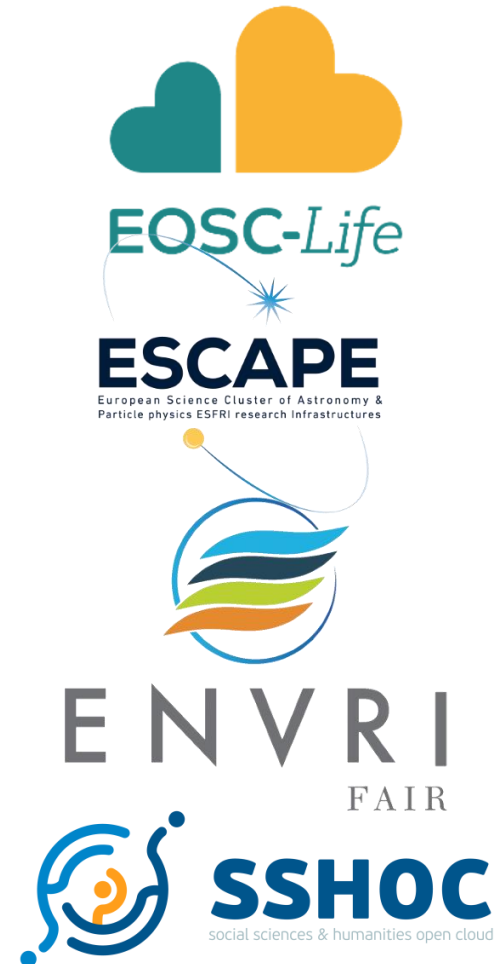
Sustaining the PaN Open Science Commons

- Option 1 - **Local implementation ($$$)** – **all sites** implement a local data repository and the PaNOSC API which supports federated searching of Open Data

- Option 2 - **Centralised implementation ($) -** all sites contribute data (and money) to **one site** which implements a PaN data repository for Open Data + Open Science

- Option 3 - **Hybrid implementation ($$)** - **some sites** implement a local data repository and make Open Data available via the PaNOSC search API, sites without a data repository contribute Open Data (and money) to **a centralised site**

Questionnaire sent to LEAPS partners

1. How many of the PaNOSC + ExPaNDS outcomes will your facility adopt?

1. Do you share the vision of a PaN Open Science Commons?

1. Should we publish a LEAPS strategy paper on Data + Open Science Strategy?

panOSC
photon and neutron
open science cloud

ExPaNDS
European Open Science Cloud Photon
and Neutron Data Services

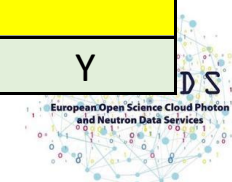# Q1: LEAPS facilities uptake of PaNOSC+EXPaNDS outcomes

Yes, already adopted (Y)
Not Planning to be adopted (N)
In progress of being adopted (WIP)
Planned to be adopted (P)
Under evaluation (U)

FACILITY	FAIR data policy	DMPs	DOIs	Nexus HDF5	Search API	Open Data Portal	AAI	JupyterLab	VISA	SIMEX	Pan-learning/ training
ALBA	P	P	WIP	WIP	P	P	U	P	U	U	U
DESY	WIP	P	P	Y	WIP	P	WIP	Y	U	N	WIP
DIAMOND											
ELETTRA	Y	WIP	Y	Y	WIP	WIP	Y	Y	WIP	Y	WIP
ESRF	Y	WIP	Y	Y	WIP	WIP	Y	Y	WIP	Y	WIP
EuXFEL	WIP	WIP	Y	WIP	WIP	WIP	WIP	Y	WIP	Y	WIP
FELIX	Y	P	WIP	U	U	WIP	U	U	N	N	U
HZB	Y, N, P [1]	P	WIP [2]	Y	P	Y	P	U	U	U	U
HZDR	WIP	WIP	Y	N	N	WIP	WIP	WIP	P	N	Y
INFN	U	U	U	U	U	U	U	U	U	U	U
ISA*	U	U	U	U	U	U	U	U	U	U	
MAX IV											
PSI	WIP	WIP	Y	WIP	Y	Y	WIP	WIP	N	N	N
PTB	Y	WIP	Y	WIP	N	Y	N	N	N	N	N
SOLARIS #											
SOLEIL	U	P	WIP	Y	P	P	WIP	U	WIP	U	Y

panosc
photon and neutron
open science cloud

European Open Science Cloud Photon
and Neutron Data Services

Q1: LEAPS facilities uptake of PaNOSC+EXPaNDS outcomes in the future

FACILITY	FAIR data policy	DMPs	DOIs	Nexus HDF5	Search API	Open Data Portal	AAI	JupyterLab	VISA	SIMEX	Pan-learning/ training
ALBA	Y	Y	Y	Y	Y	Y	U	Y	U	U	U
DESY	Y	Y	Y	Y	Y	Y	Y	Y	U	N	Y
ELETTRA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
ESRF	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
EuXFEL	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
FELIX	Y	Y	Y	U	U	Y	U	U	N	N	U
HZB	y	Y	Y	Y	Y	Y	Y	U	U	U	U
HZDR	Y	Y	Y	N	N	Y	Y	Y	Y	N	Y
PSI	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N
PTB	Y	Y	Y	Y	N	Y	N	N	N	N	N
SOLEIL	U	Y	Y	Y	Y	Y	Y	U	Y	U	Y
INFN	U	U	U	U	U	U	U	U	U	U	U
ISA*	U	U	U	U	U	U	U	U	U	U	
DIAMOND	* Will fill-in after Oct 25										
MAX IV											
SOLARIS #											

panosc
photon and neutron
open science cloud

ExPaNDS
European Open Science Cloud Photon
and Neutron Data Services

Results

- **Q2: Does your facility share the view of a PaN Data Commons?**
  Answer = 12 Yes + 4 No

- **Q2b: Which scenario do you prefer for a PaN Data Commons?**
  Answer = 7 Scenario 1 + 5 Scenario 3

- **Q3: Do you see a need for a common Open Science and Open Data strategy paper for LEAPS**
  Answer = 8 Yes + 2 No + 6 Undecided

	Q2	Q2b	Q3
ALBA	Y	3	Y
DESY	Y	1	N
DIAMOND			
ELETTRA	Y	1	Y
ESRF	Y	3	Y
EuXFEL	Y	1	Y
FELIX	Y	3	Y
HZB	Y	1	Y
HZDR	Y	1	Y
INFN	Y	3	N
ISA	Y	3	N/A
MAX IV			
PSI	Y	1	N/A
PTB			N/A
SOLARIS			
SOLEIL	Y	1	Y

Business Models for PaN Open Data Commons

1. EU / National Project funding → **Not for Operation**

2. Collaboration Contracts between Ris → **Operation**

3. Agreements including other funding agencies

4. New legal entity → e.g. an ERIC ?

Conclusion

1. **Primary outcomes of PaNOSC + ExPaNDS will be adopted at the majority of sites**

2. **LEAPS facilities share the vision of a PaN Data Commons**

3. **LEAPS facilities will make FAIR data reality**

4. **Sustainability to be addressed with internal funding + EOSC**

Thank you to all PaNOSC + ExPaNDS contributors!