



# PaNOSC Closing Event

Paving the way towards the PaN FAIR Data Commons

29-30 November 2022

Grenoble - France

## WP4 – outcomes, adoption, future plans

**Author:** Fabio Dall'Antonia

**Affiliation:** European XFEL

2022/11/29



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

# Some thoughts on the big picture



[FAIR Principles](#) [Implementation Networks](#) [News](#) [Events](#) [Resources](#) [About GO FAIR](#) [Q](#)

Open  
science

Open data  
FAIR data

WP4:  
Services for  
FAIR data  
analysis

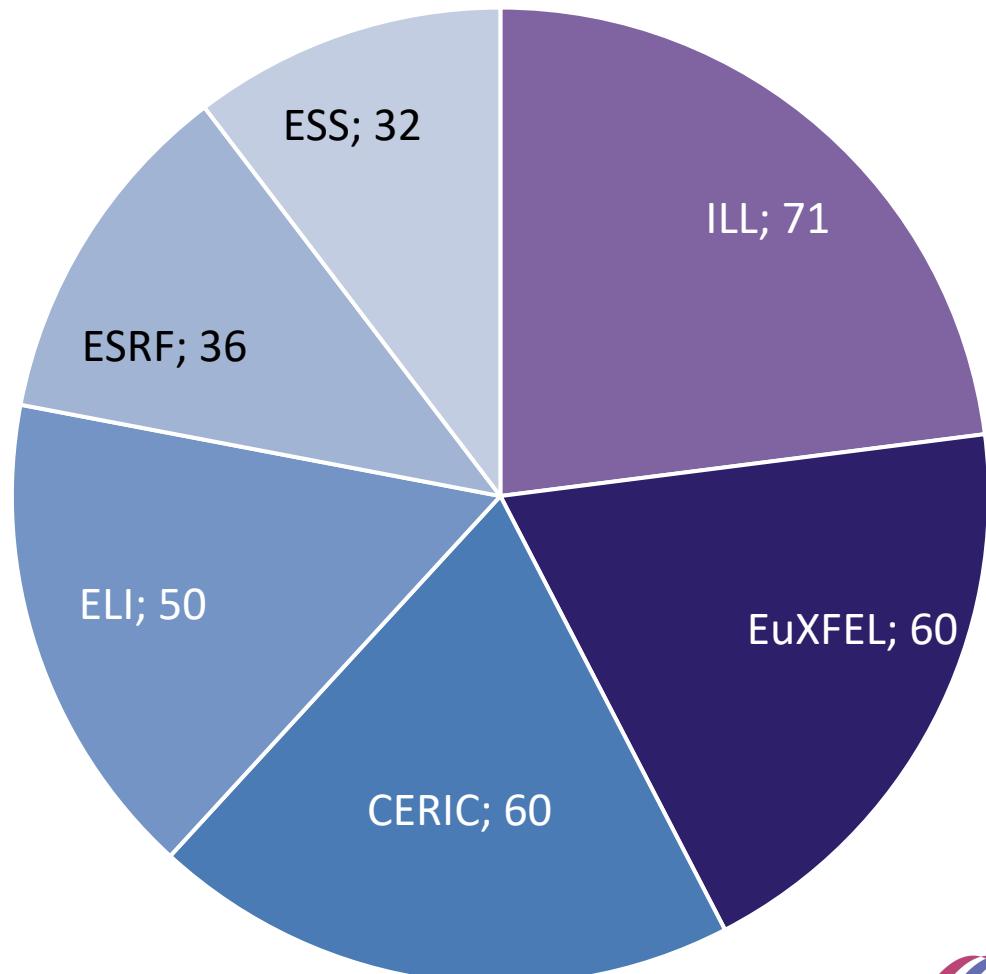


# WP4: Data Analysis Services

## Work package contributions

- Lead beneficiary: European XFEL
- ILL: remote desktop, cloud-platform (VISA)
- ESRF: HDF5 service (h5web ecosystem)
- CERIC: HDF5 service (h5nuvola)
- ELI: portal front-end
- EuXFEL: HDF5 packages/tools, Jupyter integration
- ESS: analysis SW tools, service testing

Person Months



# WP4: Data Analysis Services

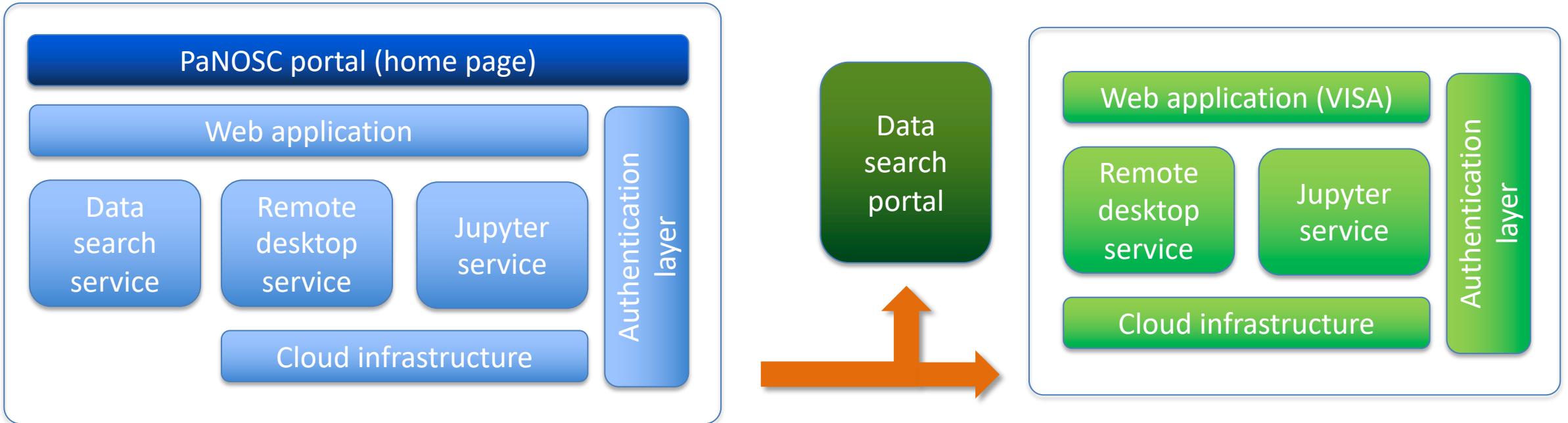
## Tasks and deliverables

<b>Task 4.1</b>	Survey data analysis requirements and solutions at the partner sites, and horizon scan other emerging tools and technologies
<b>Task 4.2</b>	Remote desktop based analysis services
<b>Task 4.3</b>	EOSC integration and common portal for remote data analysis services
<b>Task 4.4</b>	Jupyter ecosystem based data analysis services
<b>Task 4.5</b>	Deployment of remote analysis services at PaNOSC facilities
<b>Task 4.6</b>	Publicly accessible demonstrator

<b>D 4.1</b>	Report on the current technical elements of data analysis at each partner site
<b>D 4.2</b>	Prototype remote desktop and Jupyter service
<b>D 4.3</b>	Remote desktop and Jupyter service deployed at EOSC
<b>D 4.4</b>	Publicly accessible demonstrator

# WP4: Challenges and opportunities

## Data analysis portal: change of strategy/concept



### Change of work package leader

- Hans Fangohr (M1 – M23)
- Sandor Brockhauser (M24 – M28)
- Fabio Dall'Antonia (M29 – M48)

# Major WP4 outcomes



Search portal

X-ray scattering tensor tomography data for a validation sample  
X-ray scattering tensor tomography (XSTT) facilitates the investigation of microstructural organization in statistically large sample volumes. Leveraging recent developments of diffractive optics, orders of magnitude faster XSTT has been developed and demonstrated....  
Released by **PSI** on January 1st 2021

10.16907/c4fdaf60-562b-4501-a314-a7c153d676ee 0.900  
10.15151/ESRF-ES-514223675 0.900  
Released 27. July 2024  
Facility European Synchrotron Radiation Facility  
Type IH-MA  
Services Jupyter SLURM  
PanData Software Catalogue

[data.panosc.eu](http://data.panosc.eu)



Software catalogue



Node and GPU availability						
Partition	# nodes	# avail	# GPUs avail	# P100 avail	# V100 avail	# A100 avail
jhub	4	4	0	0	0	0
all	459	184	0	0	0	0
allgpu	165	99	99	30	32	37
exfel	354	153	8	8	0	0
upex	354	153	8	8	0	0

h5web

nexus\_entry  
nx8\_rgb  
uint8\_rgb  
int32\_rgb  
float32\_rgb  
nexus\_entry  
nx\_process  
nx\_data  
title  
absolute\_default\_path  
spectrum  
image  
log\_spectrum  
spectrum\_with\_aux  
complex  
complex\_spectrum  
rgb-image  
descending-axes  
scatter

Nx Heatmap  
NX Line  
NX Heatmap  
Nexus 2D  
y D0 D1  
x D0 D1  
n 20 41  
y 0 25  
x 0 100  
nexus\_entry

HSNUVOLA  
Databases  
Data sources  
VUO

Al\_K  
Mg\_K  
Scatter\_Compton000  
Scatter\_Peak000



# Adoption of WP4 outcomes

FACILITY	FAIR data policy	DMPs	DOIs	Nexus HDF5	Search API	Open Data Portal	AAI	Jupyter Lab	VISA	VINYL/O ASYS/Mc Stas	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

Yes, already adopted (Y)

Planned to be adopted (P)

Not Planning to be adopted (N)

Under evaluation (U)

In progress of being adopted (WIP)



# Plans for the future

## Task leftovers, loose ends and new ideas

- Harmonize “openness levels” and AAI workflows for data analysis services
- Facilitate open data access at RI level
- Harmonize open data management: preparation for services, facility-wise
- Ramping up VISA for open data use (for some facilities)
- Identify more use cases, promote services
- Domain-specific open data services
- Not discussed here: open data transfer for horizontal access



PaNOSC has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 823852

# Openness: authentication vs. anonymity

„As open as possible,  
as restricted as necessary“

CLOUD APPLICATIONS AT DESY

Bei Ihrem Konto anmelden

Deutsch ▾

Benutzername

Passwort

Angemeldet bleiben

Anmelden

Oder anmelden mit

GitHub

EGI Check-in

Helmholtz AAI



## WP4 levels of openness

- Anonymous: Search portal usage
- Authentication (misuse barrier): e. g. VISA
- Authorisation: moderation of access to resource usage (e. g. DESY OpenStack cluster for VISA)

Services for analysis of open data should strive for **non-exclusivity** where ever possible, but limited resources could still raise the need to introduce a selection of users

Login to Helmholtz AAI OAuth2 Authorization Server

Umbr

Lakes College West Cumbria

Northumbria University

Umbrella ID

University of Cumbria

University of Cumbria

# Making open data a reality: work to finish

10.22003/XFEL.EU-DATA-700000-00

Example Data

The European XFEL (EuXFEL) example data proposal contains experimental data from original beam-times, currently covering the techniques of serial femtosecond coherent diffraction imaging (single particle imaging, SPI), X-ray powder diffraction scattering (SAXS) and X-ray photon correlation spectroscopy (XPCS).

Released

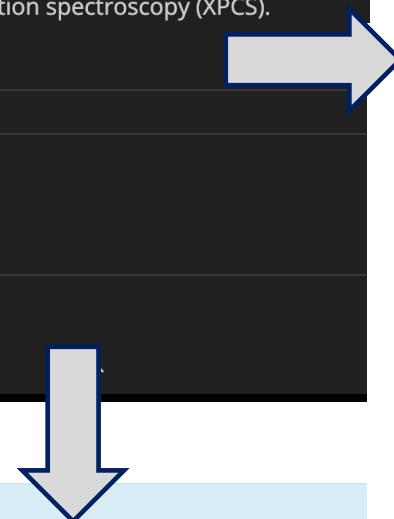
Facility

Type

Services

VISA

PaNdata Software Catalogue



## Proposal Runs

Automatically assess new runs (after being closed by DAQ) as: [To be evaluated manually](#)

Automatically start run calibration after migration: [No](#)

Run Number (alias)	Run type	Sample Name	Techniques	Start date	Run status	Data Assessment	Calibration	Run Comment	Edit
0034 (SPI on sucrose solution, AGIPD detector at SPB instrument)	Single Particle Diffraction	Sucrose Solution 3% v/v	coherent diffraction imaging	2021-06-01 02:25:08 +0200	Closed	Good	<a href="#">Edit</a>	<a href="#">View</a> <a href="#">Download</a>	<a href="#">Edit</a>
0033 (SAXS on vycor sample, AGIPD detector at MID instrument)	scattering	Vycor	small angle x-ray scattering	2021-04-10 14:48:20 +0200	Closed	Good	<a href="#">Edit</a>	<a href="#">View</a> <a href="#">Download</a>	<a href="#">Edit</a>
0031 (SFX on Hen egg-white lysozyme, AGIPD detector)	Diffraction data	Lysozyme	serial femtosecond crystallography	2021-04-15 10:48:26 +0200	Closed	Good	<a href="#">Edit</a>	<a href="#">View</a> <a href="#">Download</a>	<a href="#">Edit</a>
0030 (SFX on Hen egg-white lysozyme, AGIPD detector)	Diffraction	Lysozyme	serial femtosecond crystallography	2020-03-09 01:20:02 +0100	Closed	Good	<a href="#">Edit</a>	<a href="#">View</a> <a href="#">Download</a>	<a href="#">Edit</a>
0029 (SFX on Hen egg-white lysozyme, AGIPD detector)	Diffraction	Lysozyme	serial femtosecond crystallography	2020-03-09 01:07:51 +0100	Closed	Good	<a href="#">Edit</a>	<a href="#">View</a> <a href="#">Download</a>	<a href="#">Edit</a>
0027 (SAXS on 50 nm silica, AGIPD detector at MID instrument)	scattering	Silica 50nm	small angle x-ray scattering	2019-09-21 01:12:49 +0200	Closed	Good	<a href="#">Edit</a>	<a href="#">View</a> <a href="#">Download</a>	<a href="#">Edit</a>
0026 (Time-resolved SAXS on Ni75-11 MLs, DSSC detector at SCS)	SAXS 500kHz // no pump laser	Ni75-11 MLs-b	small angle x-ray scattering	2019-08-23 07:08:02 +0200	Closed	Good	<a href="#">Edit</a>	<a href="#">View</a> <a href="#">Download</a>	<a href="#">Edit</a>



# Data catalogues

## Open Data connections

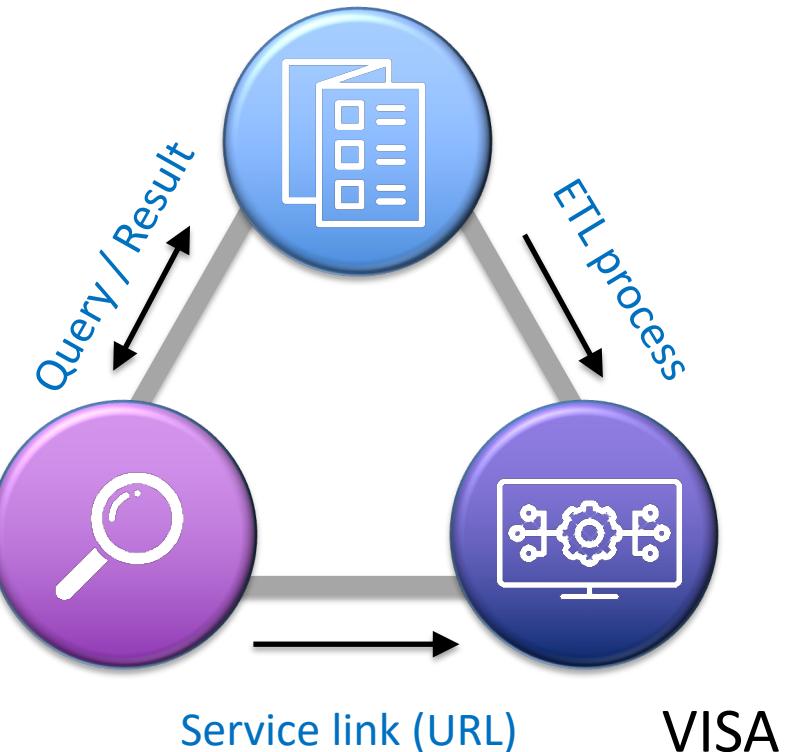
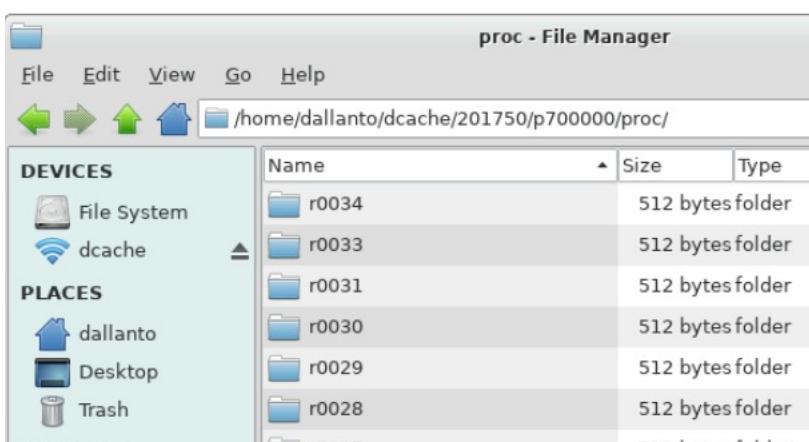
Instrument XMPL between 2017 and 2022 with open data included sort by  
date (newest first)

Proposal	Title	Instrument	Start Date	End Date	
p700000	SPI on sucrose solution, AGIPD detector at SPB instrument	XMPL	01 Jun 2021	01 Jun 2021	<button>SELECT</button>
p700000	SFX on Hen egg-white lysozyme, AGIPD detector	XMPL	15 Apr 2021	15 Apr 2021	<button>SELECT</button>
p700000	SAXS on vycor sample, AGIPD detector at MID instrument	XMPL	10 Apr 2021	10 Apr 2021	<button>SELECT</button>



→ C in.xfel.eu/metadata/proposals/30#proposal-runs

0034 (SPI on sucrose solution, AGIPD detector at SPB instrument)	Single Particle Diffraction	Sucrose Solution 3% v/v	coherent diffraction imaging
0033 (SAXS on vycor sample, AGIPD detector at MID instrument)	scattering	Vycor	small angle x-ray scattering
0031 (SFX on Hen egg-white lysozyme, AGIPD detector)	Diffraction data	Lysozyme	serial femtosecond crystallography



# VISA and open data

Very fruitful discussions and implementation of **requested features in VISA 2.4.4:** for the open data demonstrator and beyond

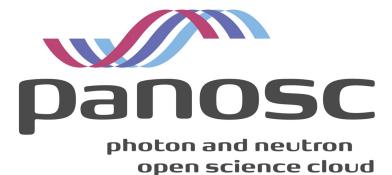
- DOIs as entry points for VISA (specific URLs from the search portal)
- Open experiment flag (by configuration and publishing date), thus user needs no association to experiment to see/access the data
- Authorisation control with instance quota

## Further ideas:

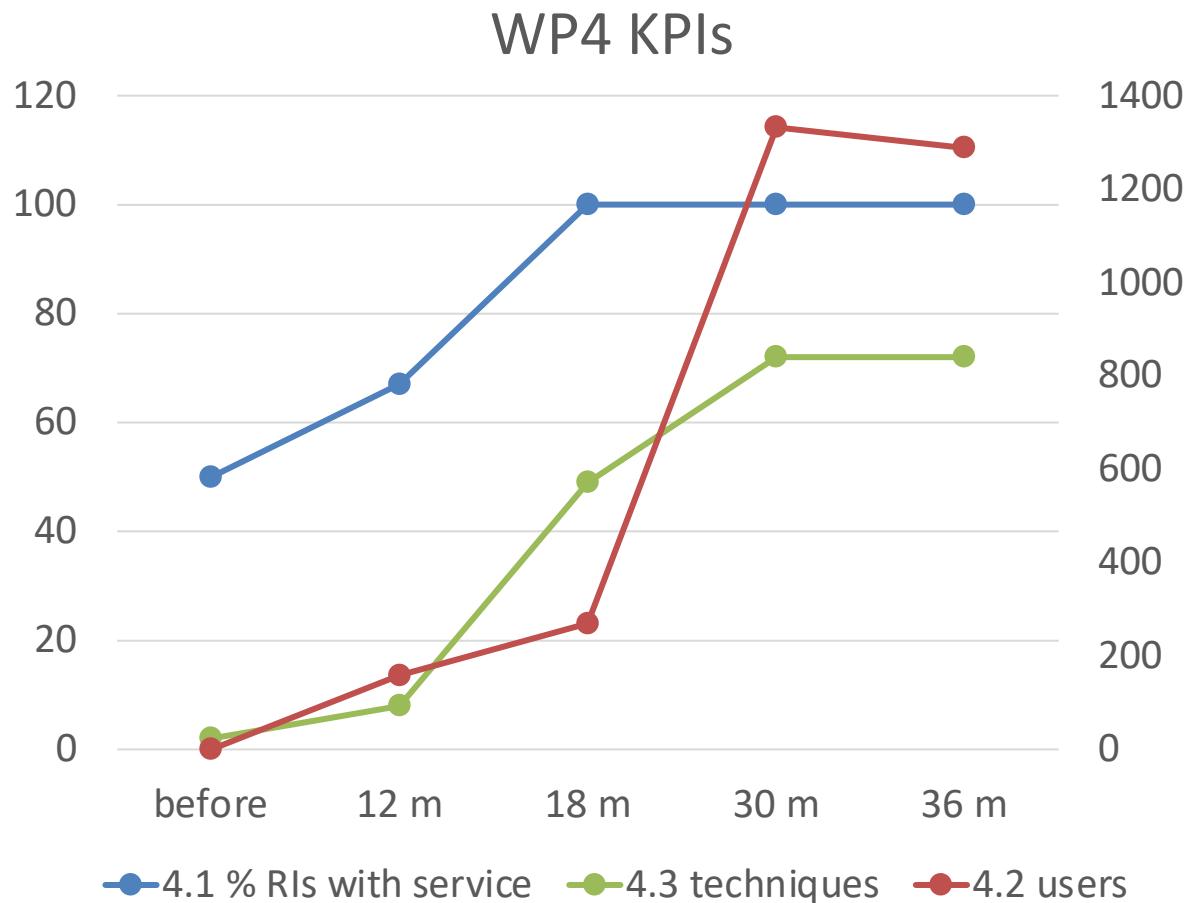
- Session collaboration tokens for shared instances in VISA, with respect to open data experiments
- Open data search in VISA



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



# Use cases and users



Partner	Use Cases Submitted	Comments
ESRF	11	
ILL	1	3 planned
ESS	2	
XFEL	7	2 planned
CERIC	8	of which 1 with ILL and EGI, 2 planned
ELI	1	1 planned
EGI	1	of which 1 with CERIC and ILL

### Data Analysis Use Cases

Submit your use case

- Use Case 29 – [Run orange-pylost as a cloud service](#) (ESRF)
- Use Case 28 – [Online visualisation, exploration and analysis of HDF5 files with h5nuvola](#) (CERIC-ERIC)
- Use Case 25 – [WebKnossos, a web-based tool for 3D data viewing and annotation](#) (ESRF)
- Use Case 24 – [View HDF5 files in ESRF Data Portal](#) (ESRF)
- Use Case 23 – [Human Organs Open Data portal](#) (ESRF)
- Use Case 22 – [BRAGGY diffraction image viewer](#) (ESRF)



# Domain-specific open data search and exploration

PDBe / SEARCH

Organism name : Covid-19 Virus

Filter by :

Latest PDB release

Entries released this week (2)

new (30)  
revised (21)  
+ New ligands in PDB (12)

Entry Information

+ Entry status (1)  
+ Experimental methods (6)  
+ Authors (6340)  
+ Homo / hetero assembly (2)

Advanced search

Entries Macromolecules Compounds Protein families

< 1 2 3 ... 10 > Macromolecule 1 to 10 of 95 10 /page

Protein : 2'-O-methyltransferase nsp16  
Best example found in:  
6xkm Room Temperature Structure of SARS-CoV-2 NSP10/NSP16 Methyltransferase in a Complex with SAM Determined by Fixed-Target Serial Crystallography  
Wilamowski M, Sherrill DA, Minasov G, Kim Y, Shuvalova L, Lavens A, Chard R, Rosas-Lemus M, Maltseva N, Jedrzejczak R, Michalska K, Satchell KJF, Joachimiak A, Center for Structural Genomics of Infectious Diseases (CSGID)  
Proc Natl Acad Sci U S A (2021) [PMID: 33972410]  
Source organisms: Severe acute respiratory syndrome coronavirus 2

X-ray diffraction  
2.25 Å resolution  
Released: 8 Jul 2020  
DOI: 10.2210/pdb6xkm  
Model geometry  
Fit model/data

Human Organ Atlas EXPLORE SEARCH RECONSTRUCTIONS HELP

Welcome to the Human Organ Atlas

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

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

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

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](#).



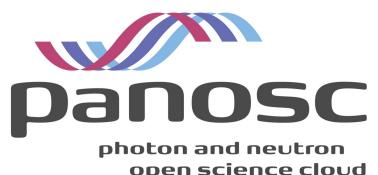
Home Mission CXI File Format Browse Data Resources

## Browse Data

- ID 1 - Single mimivirus particles intercepted and imaged with an X-ray laser
- ID 2 - Single mimivirus particles intercepted and imaged with an X-ray laser
- ID 3 - Femtosecond diffractive imaging with a soft-X-ray free-electron laser
- ID 4 - High-resolution x-ray diffraction microscopy of specifically labeled yeast cells
- ID 5 - High-resolution x-ray diffraction microscopy of specifically labeled yeast cells
- ID 6 - High-resolution x-ray diffraction microscopy of specifically labeled yeast cells
- ID 7 - High-resolution x-ray diffraction microscopy of specifically labeled yeast cells
- ID 8 - High-resolution x-ray diffraction microscopy of specifically labeled yeast cells
- ID 9 - Cryptotomography: reconstructing 3D Fourier intensities from randomly oriented single-shot diffraction patterns

## Open data collections:

- Selection – only meaningful/relevant datasets
- Preparation – pre-processing where required
- Curation





# PaNOSC Closing Event

Paving the way towards the PaN FAIR Data Commons

29-30 November 2022

Grenoble - France

## Thank you

[fabio.dallantonia@xfel.eu](mailto:fabio.dallantonia@xfel.eu)



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