WP4 – outcomes, adoption, future plans

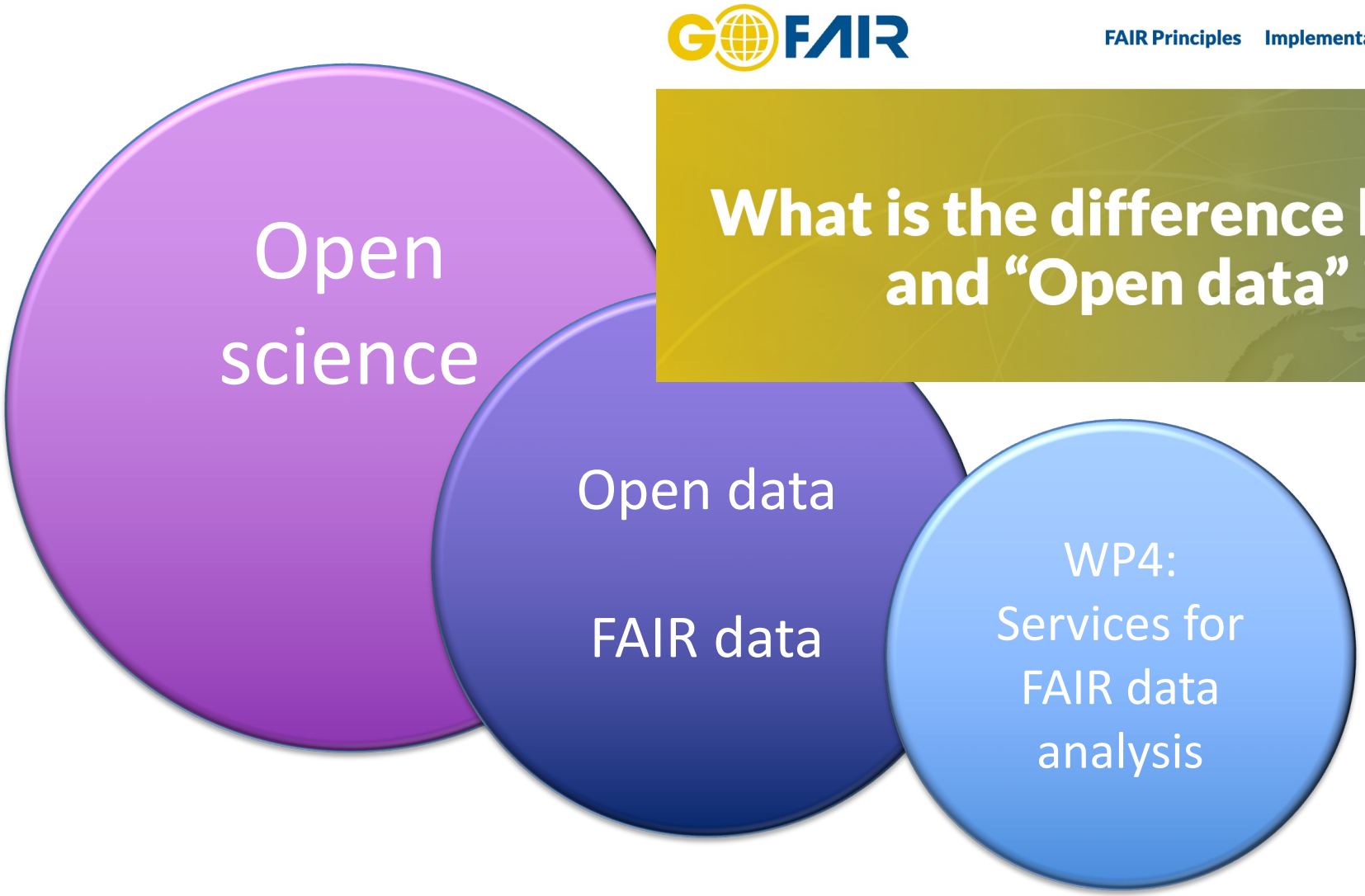
Author: Fabio Dall'Antonia

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



GO FAIR

FAIR Principles | Implementation Networks | News | Events | Resources | About GO FAIR

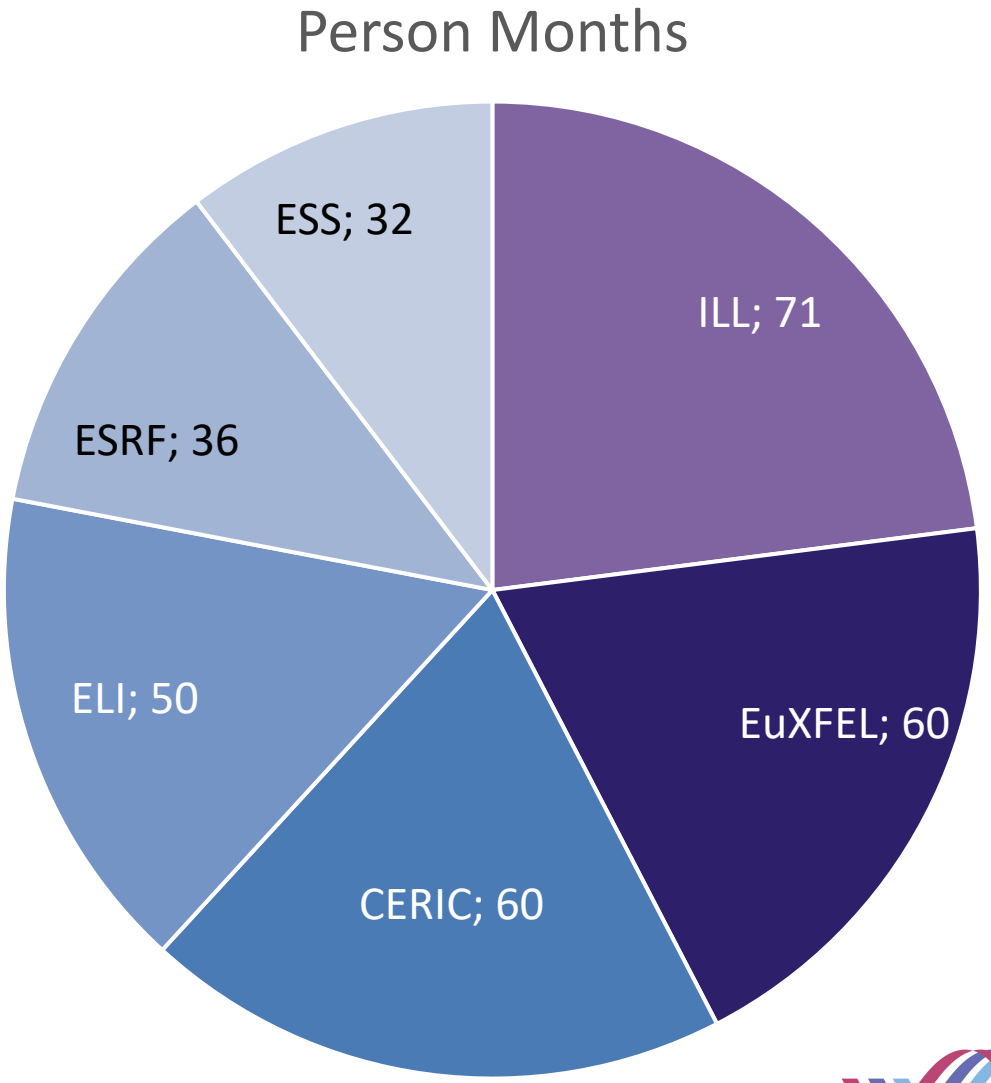
What is the difference between "FAIR data" and "Open data" if there is one?

panosc
photon and neutron open science cloud

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



panosc
photon and neutron open science cloud

WP4: Data Analysis Services

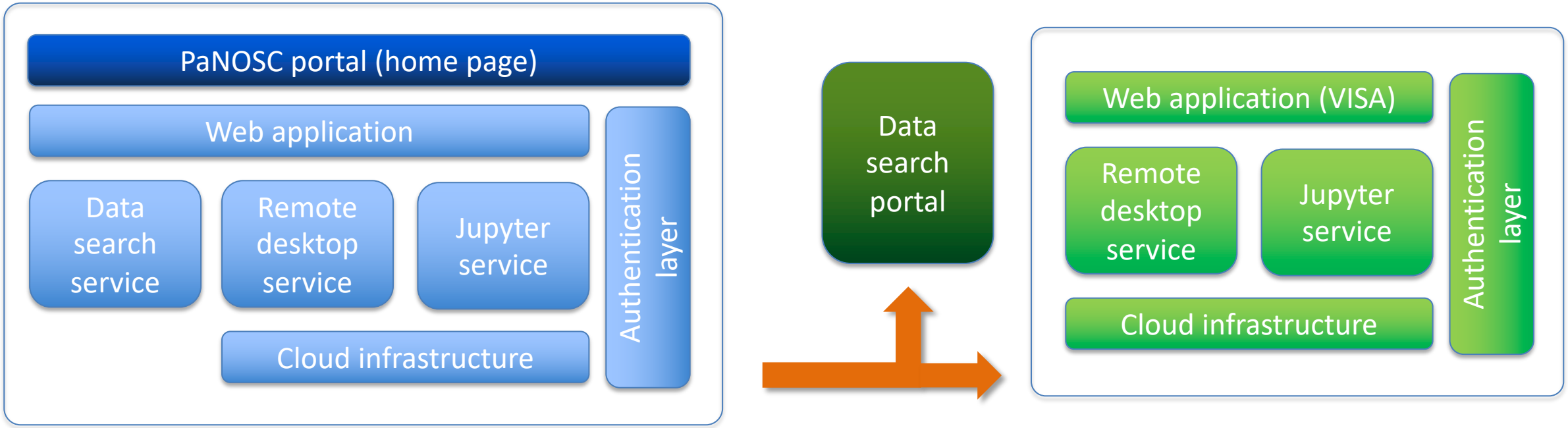
Tasks and deliverables

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

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

WP4: Challenges and opportunities

Data analysis portal: change of strategy/concept



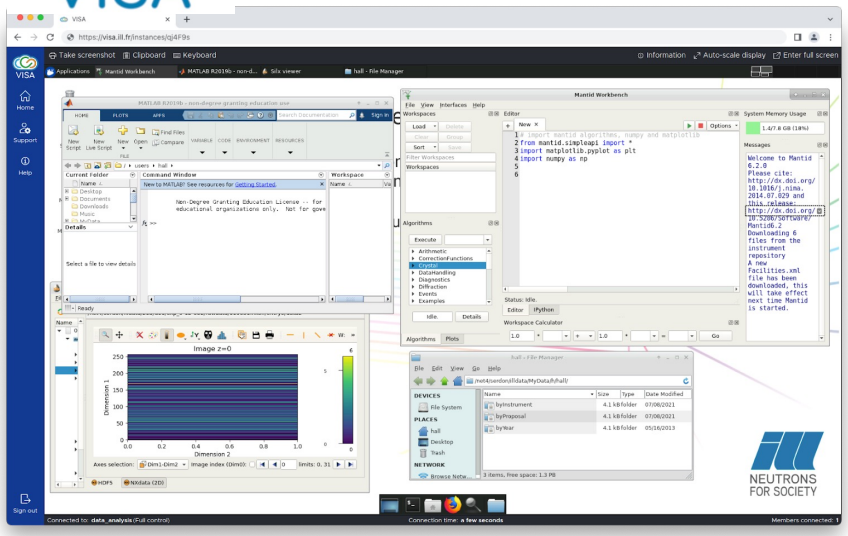
Change of work package leader

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

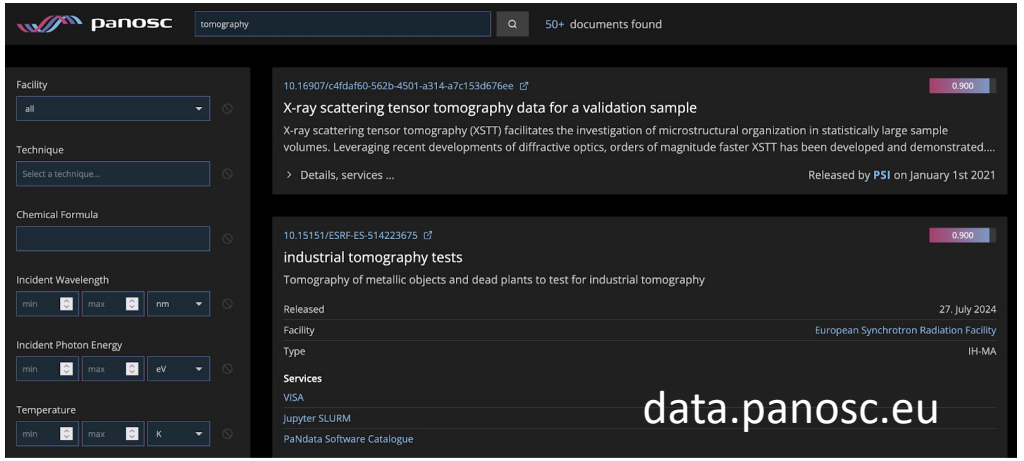
panosc
photon and neutron open science cloud

Major WP4 outcomes

VISA

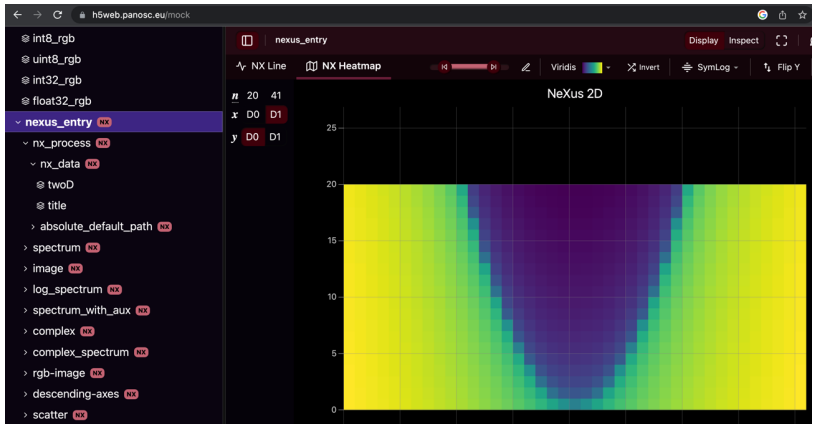


Search portal

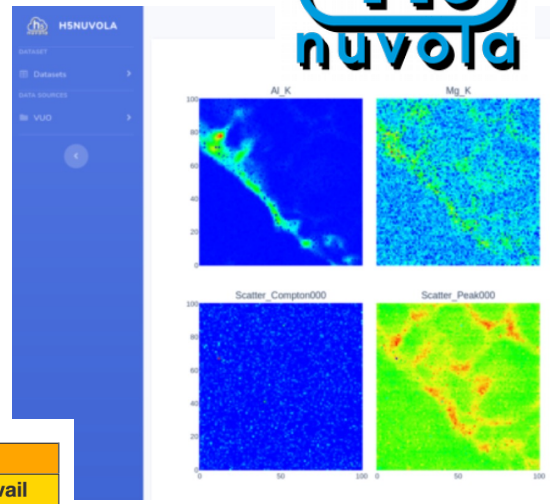


data.panosc.eu

h5web



h5 nuvola



pandata

Software catalogue

jupyter

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

panosc
photon and neutron
open science cloud

Adoption of WP4 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

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

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

photon and neutron open science cloud

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

panosc
photon and neutron open science cloud

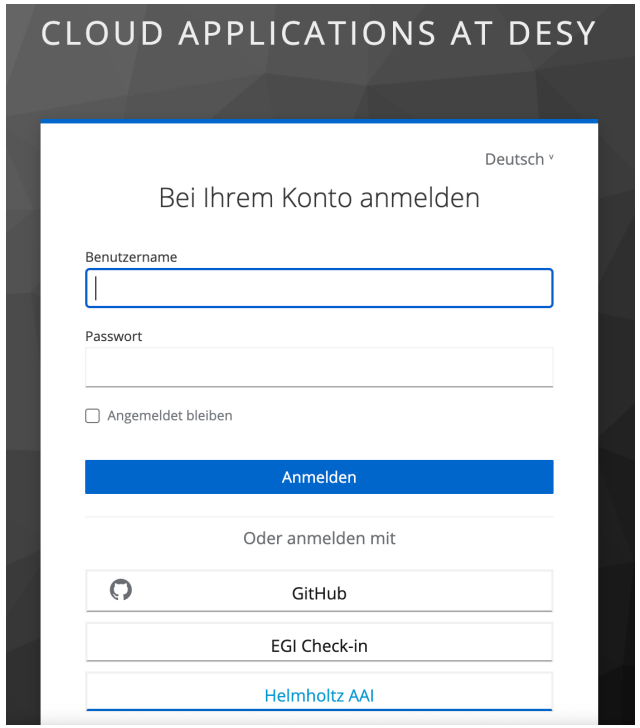
Openness: authentication vs. anonymity

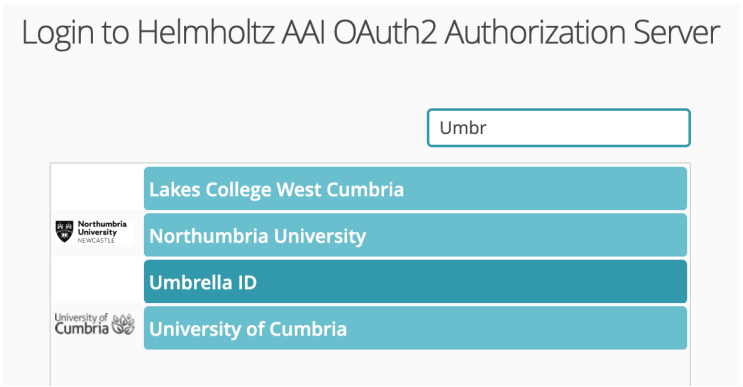
„As open as possible,
as restricted as necessary“

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)

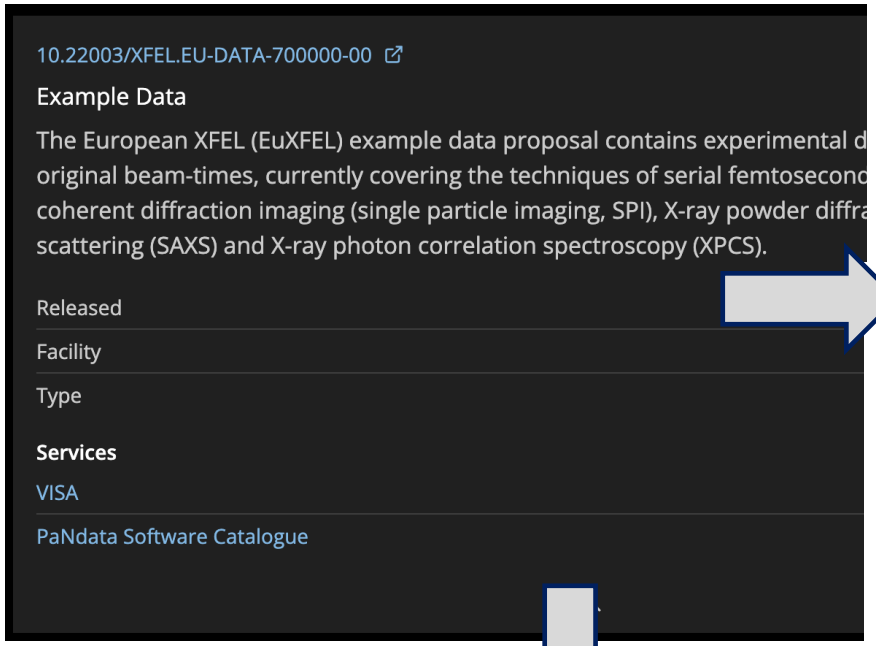
Services for analysis of open data should strive for **non-exclusivity** where ever possible, but limited resources could still raise the need to introduce a selection of users

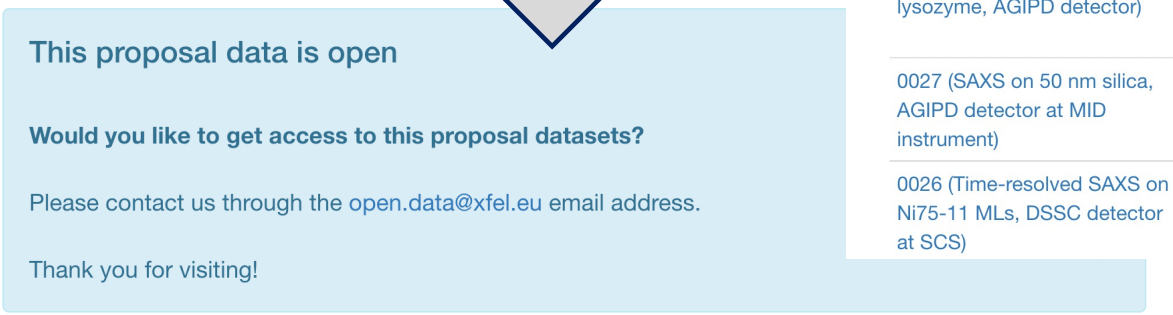




panosc
photon and neutron open science cloud

Making open data a reality: work to finish





Proposal Runs

globus

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

Off

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			
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			
0031 (SFX on Hen egg-white lysozyme, AGIPD detector)	Diffraction data	Lysozyme	serial femtosecond crystallography	2021-04-15 10:48:26 +0200	Closed	Good			
0030 (SFX on Hen egg-white lysozyme, AGIPD detector)	Diffraction	Lysozyme	serial femtosecond crystallography	2020-03-09 01:20:02 +0100	Closed	Good			
0029 (SFX on Hen egg-white lysozyme, AGIPD detector)	Diffraction	Lysozyme	serial femtosecond crystallography	2020-03-09 01:07:51 +0100	Closed	Good			
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			
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			

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

panosc
photon and neutron open science cloud

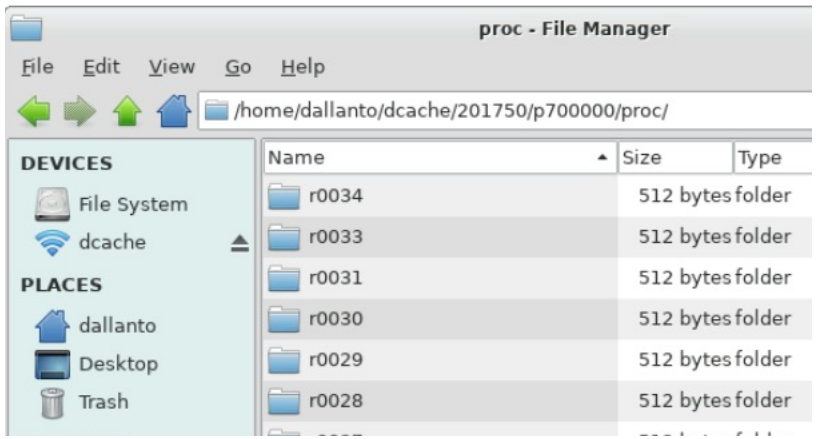
Open Data connections

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

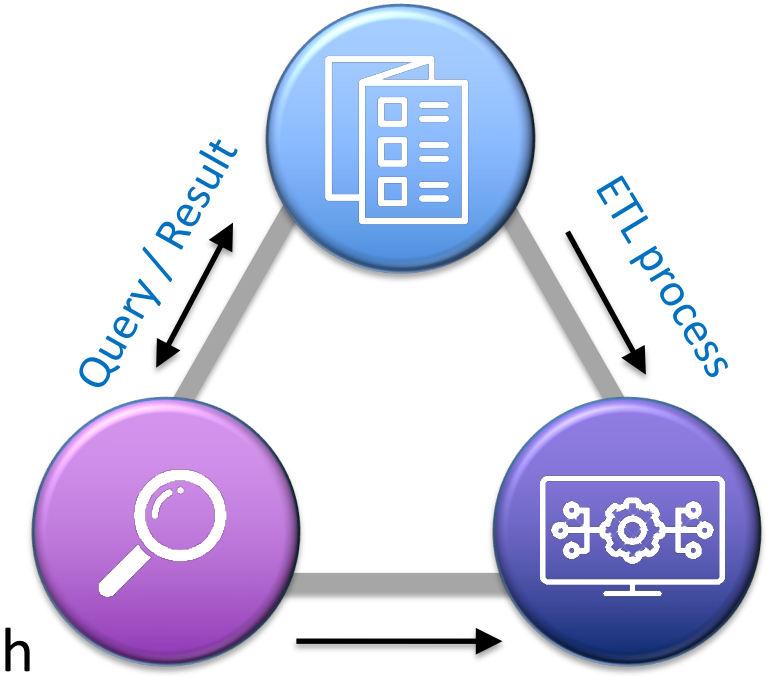
Proposal	Title	Instrument	Start Date	End Date	
p700000	SPI on sucrose solution, AGIPD detector at SPB instrument	XMPL	01 Jun 2021	01 Jun 2021	SELECT
p700000	SFX on Hen egg-white lysozyme, AGIPD detector	XMPL	15 Apr 2021	15 Apr 2021	SELECT
p700000	SAXS on vycor sample, AGIPD detector at MID instrument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

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



Data catalogues



Search portal

Service link (URL)

VISA

panosc
photon and neutron open science cloud

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

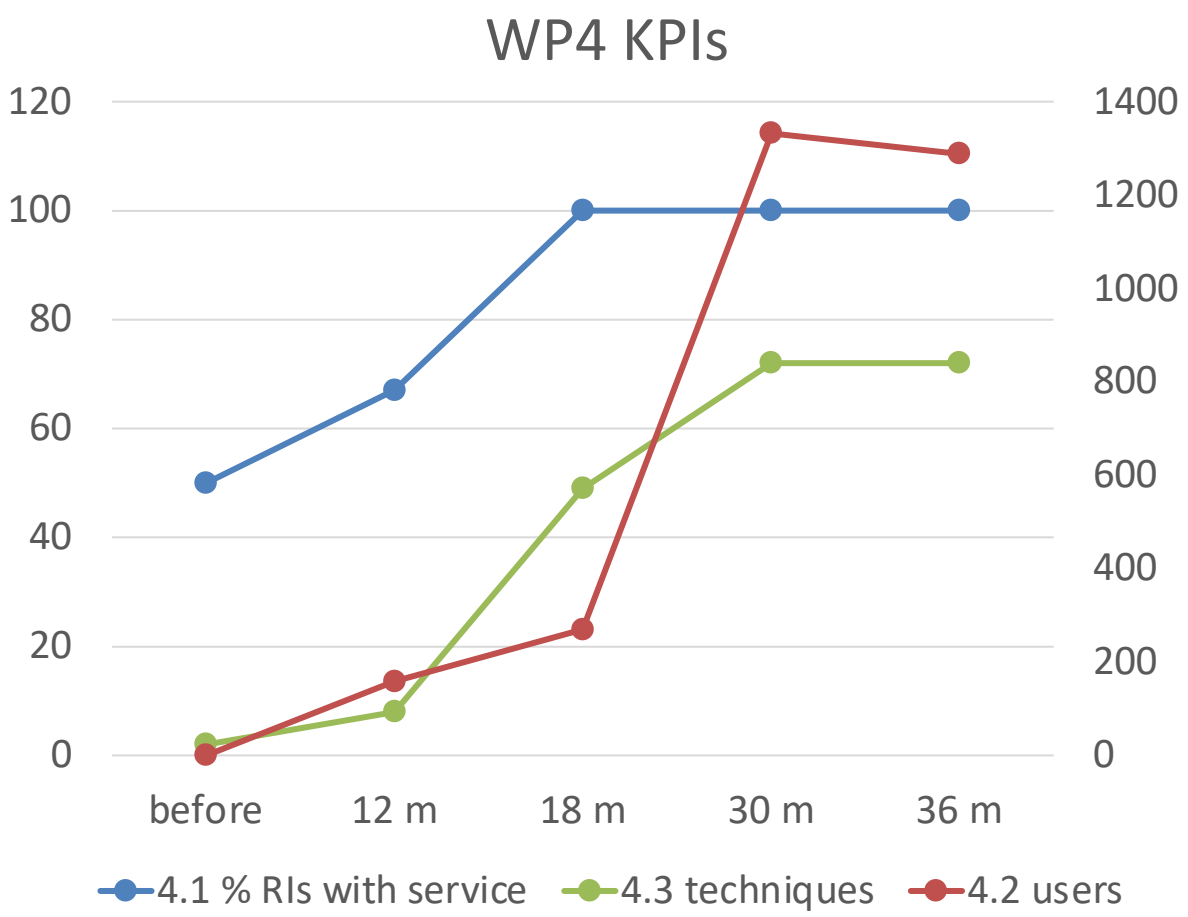
Further ideas:

- 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

panosc
photon and neutron open science cloud

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

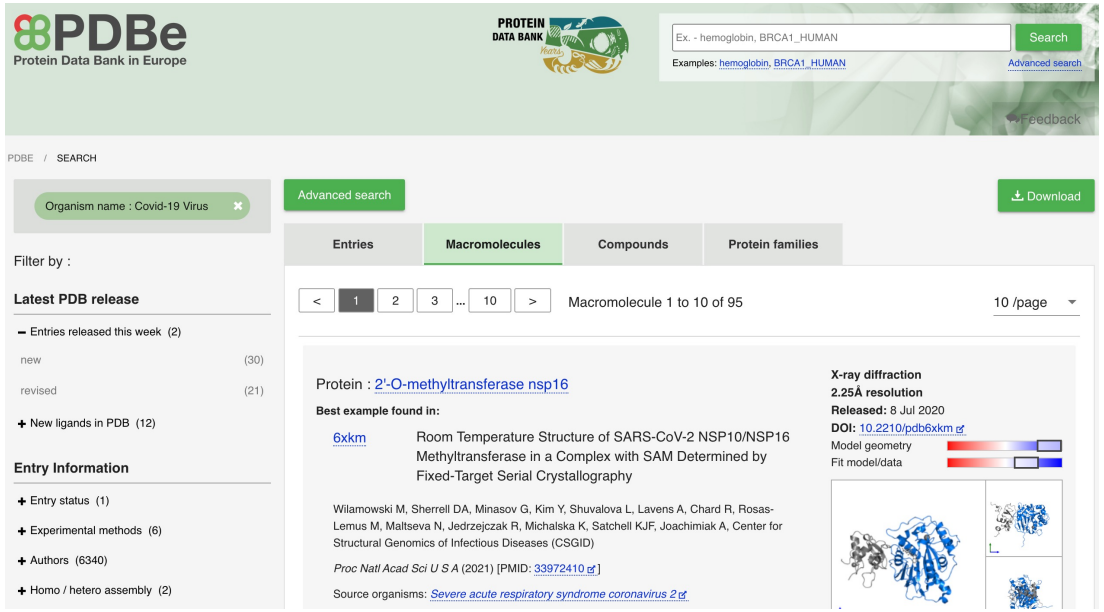
Submit your use case

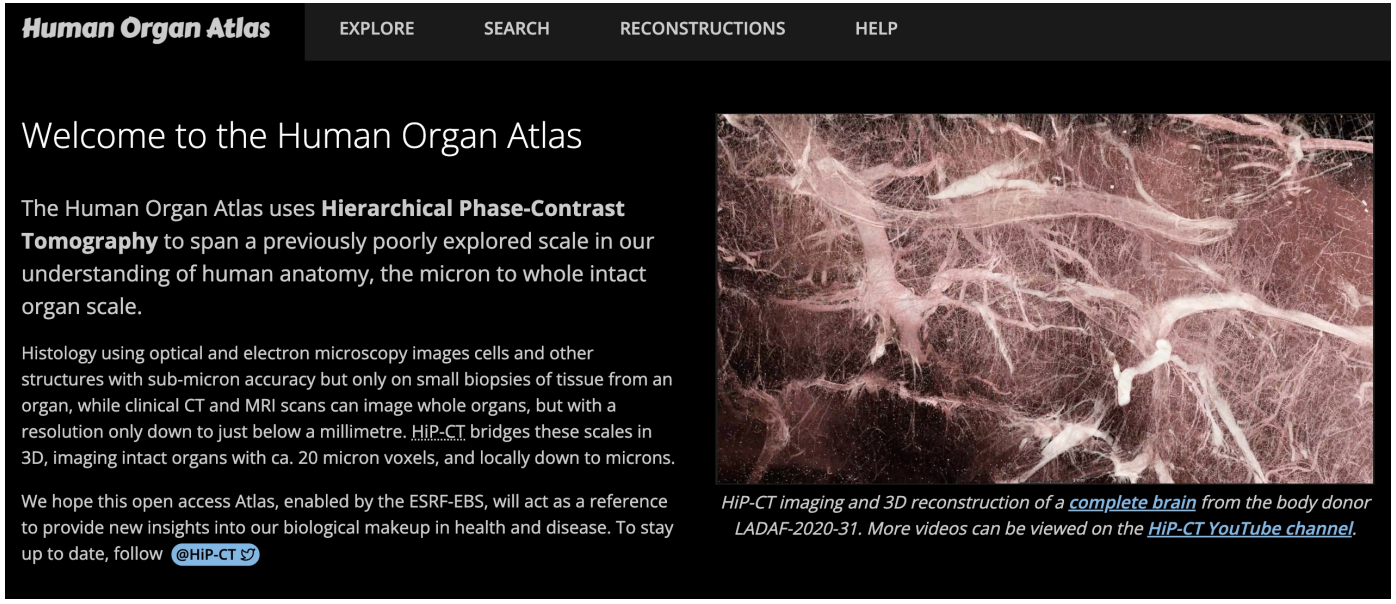
- **Use Case 29** – Run orange-pylost as a cloud service (ESRF)
- **Use Case 28** – Online visualisation, exploration and analysis of HDF5 files with h5nuvola (CERIC-ERIC)
- **Use Case 25** – WebKnossos, a web-based tool for 3D data viewing and annotation (ESRF)
- **Use Case 24** – View HDF5 files in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

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

panosc
photon and neutron open science cloud

Domain-specific open data search and exploration





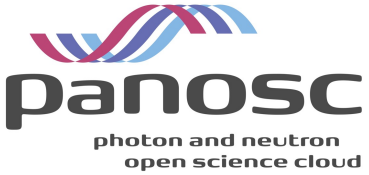
Home Mission CXI File Format Browse Data Resources

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



Thank you

fabio.dallantonia@xfel.eu

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