“HDF5 User Group European Workshop”
ITER (FRANCE)
31 May 2022

HDF5 at ESRF

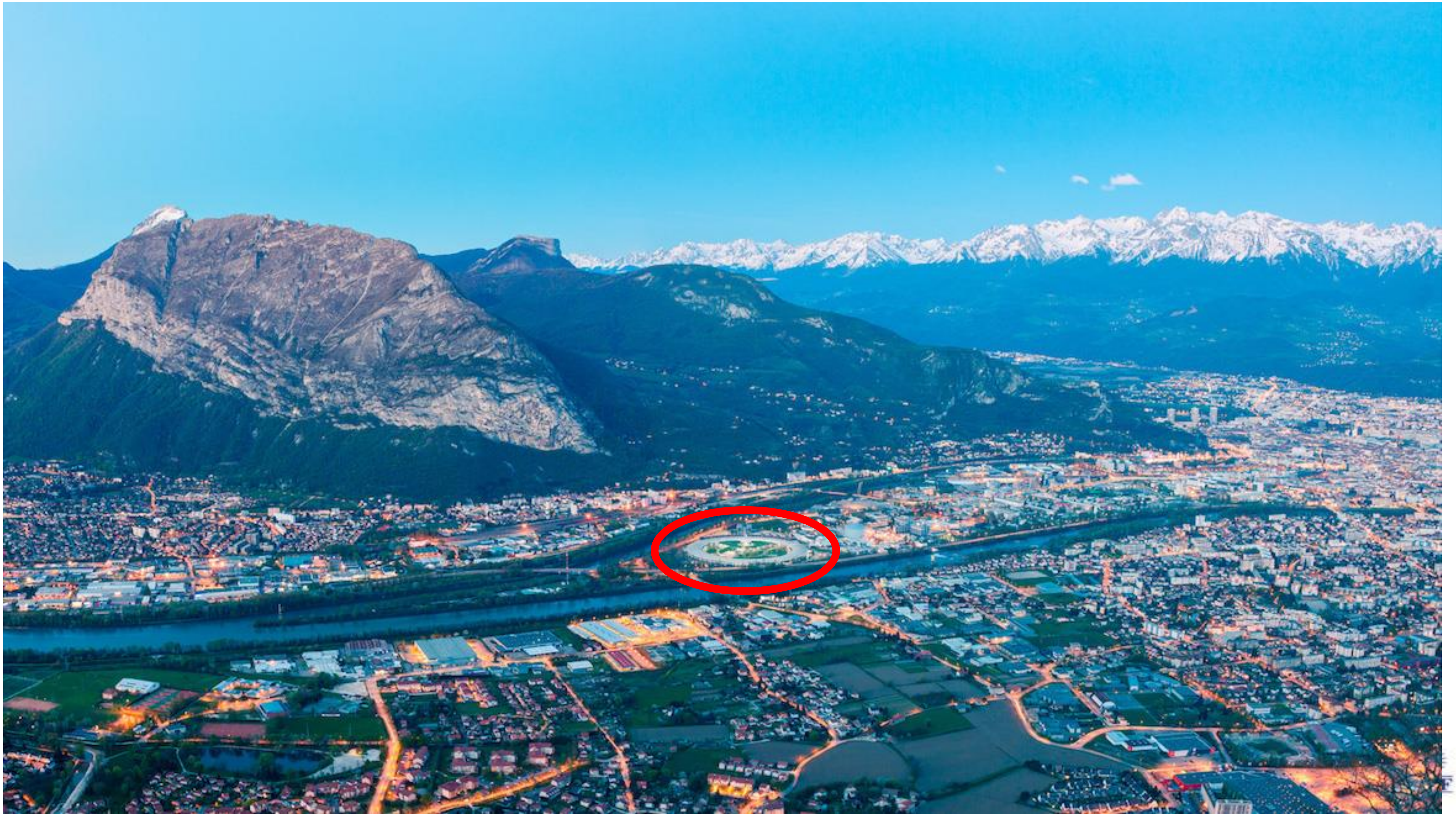
Andy Götz

ESRF, 71 avenue des Martyrs, 38000 Grenoble (France)

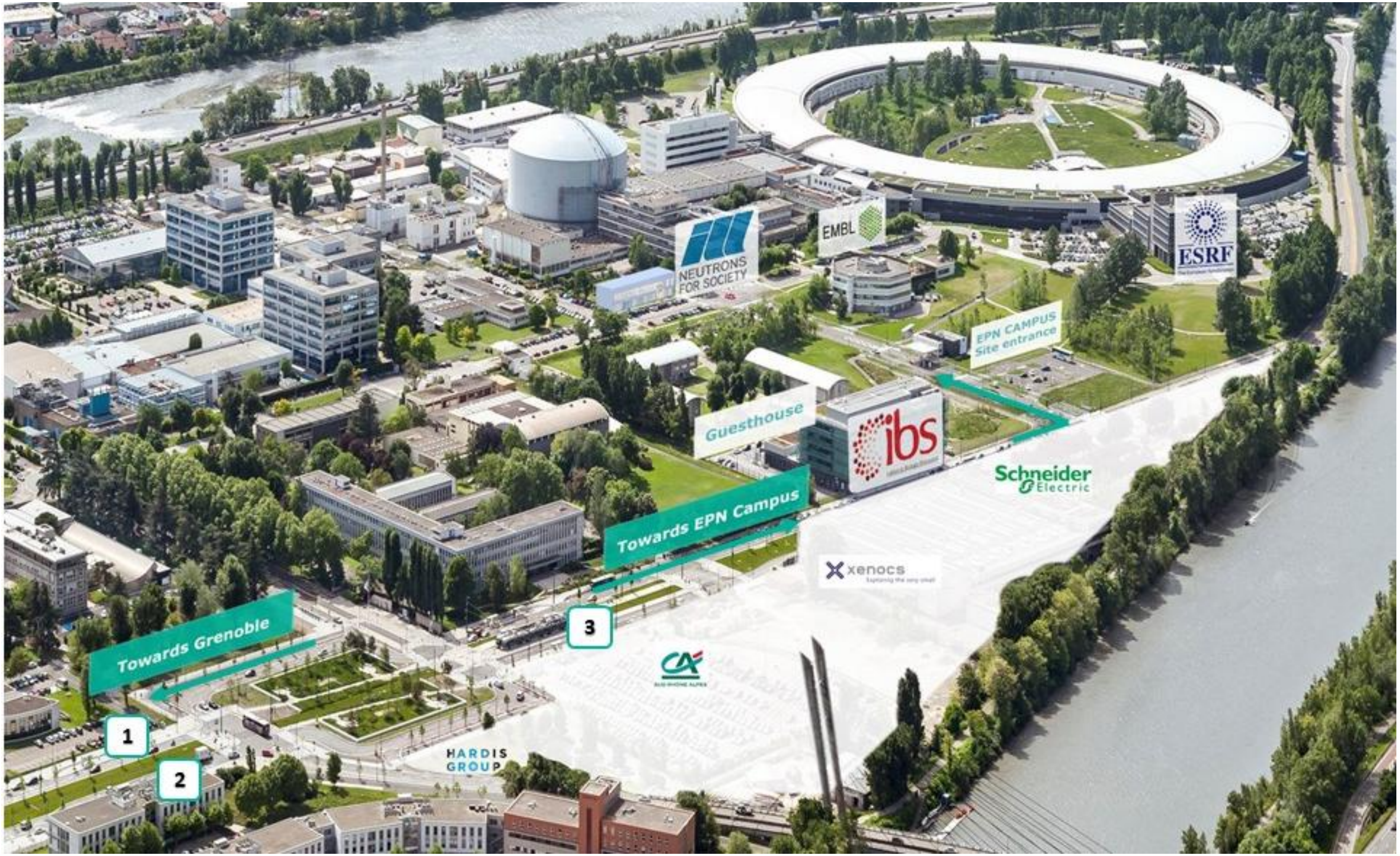
PIONEERING SYNCHROTRON SCIENCE

HDF5 - CENTRAL TO THE EBS DATA STRATEGY

ESRF – THE EUROPEAN SYNCHROTRON



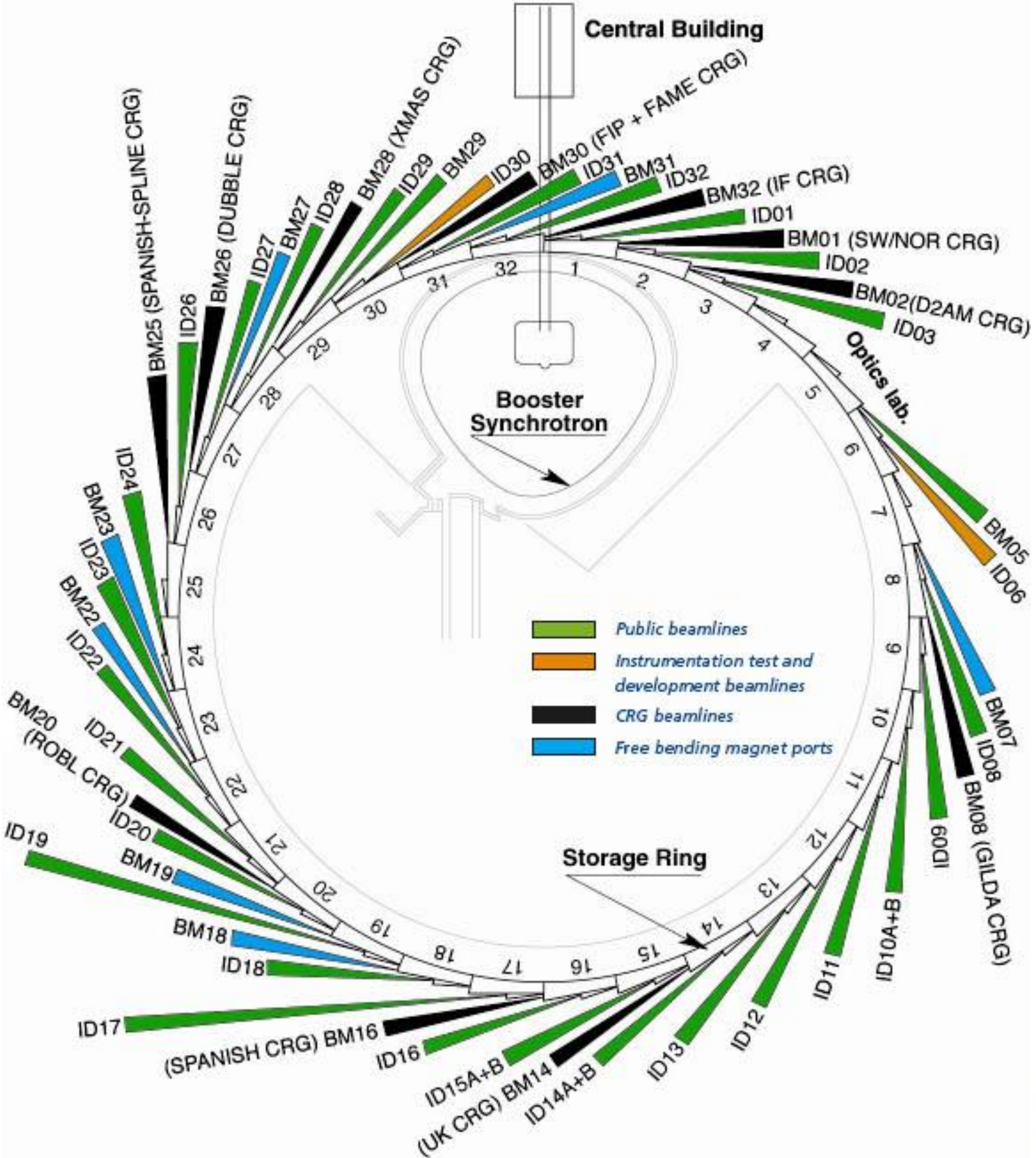
EUROPEAN PHOTON + NEUTRON CAMPUS



BACKGROUND ON ORGANISATION STRUCTURE AND MISSION

Mission

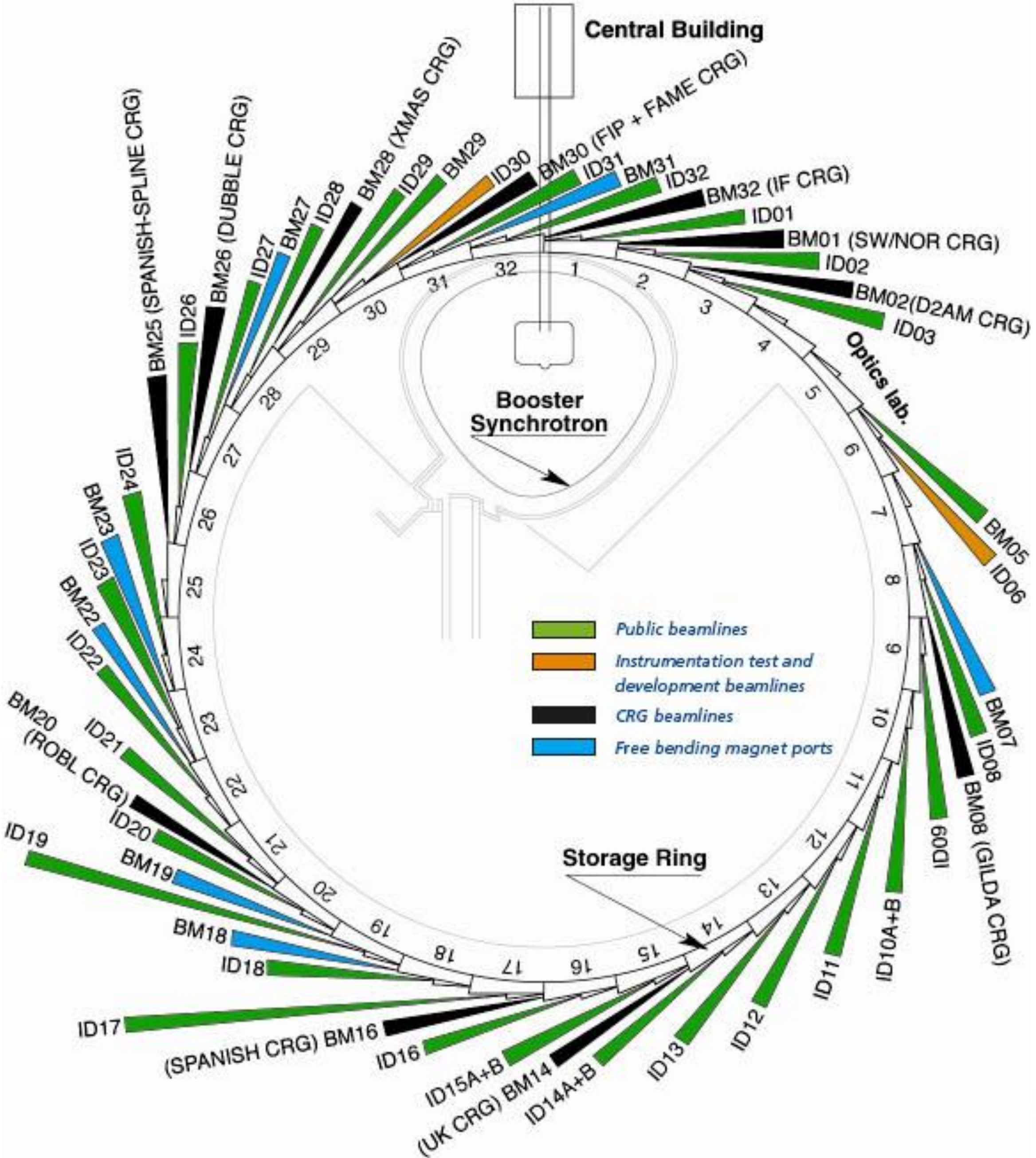
1. Produce synchrotron radiation in the hard x-ray region (10 keV – 250 keV) for doing experiments on applied science.
2. Provide visiting scientists with a hardware and software support for running experiments (free of charge for users for peer-reviewed experiments).
3. Provide users with the data from their experiments and support on how to process them.
4. Make data open and FAIR and archive them for at least 10 years



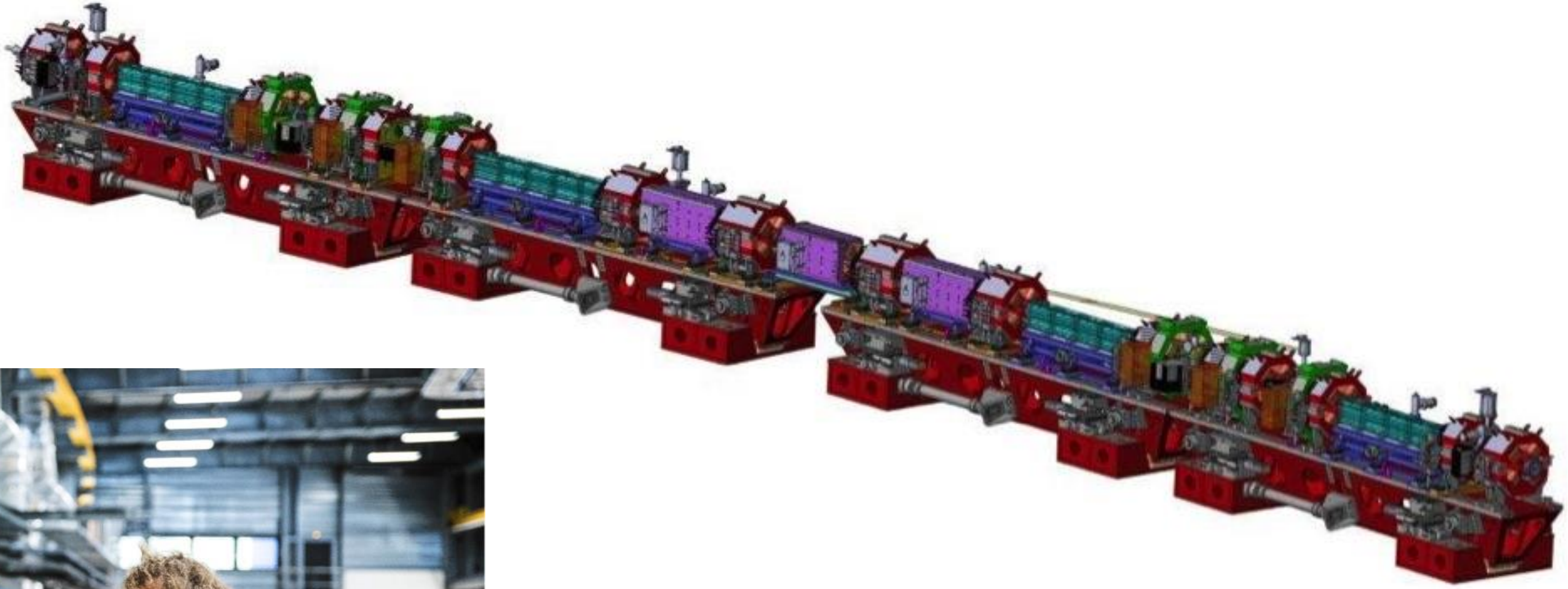
BACKGROUND ON ORGANISATION STRUCTURE AND MISSION

Experiment Categories

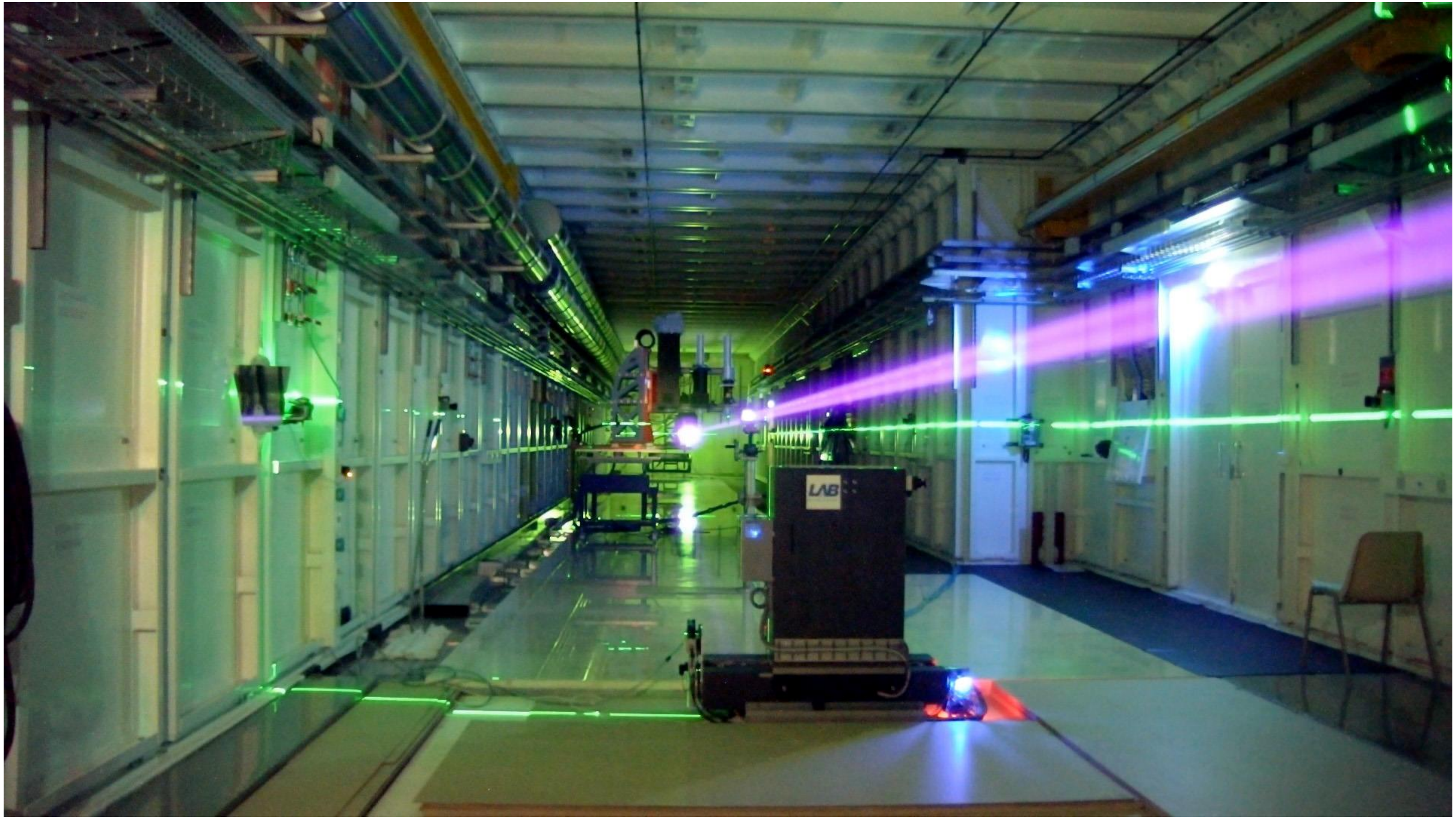
1. CH (Chemistry)
2. ES (Earth Science)
3. EV (Environment)
4. HC (Hard Condensed Matter Science)
5. HG (Cultural Heritage)
6. LS (Life Sciences)
7. MA (Applied Material Science)
8. MD (Medicine)
9. ME (Engineering)
10. MI (Methods and Instrumentation)
11. MX (Structural Biology) –
12. SC (Soft Condensed Matter Science)



NEW SOURCE EBS = 100 X MORE COHERENT BEAMS



NEW SOURCE BM18 = 100 X MORE COHERENT BEAMS



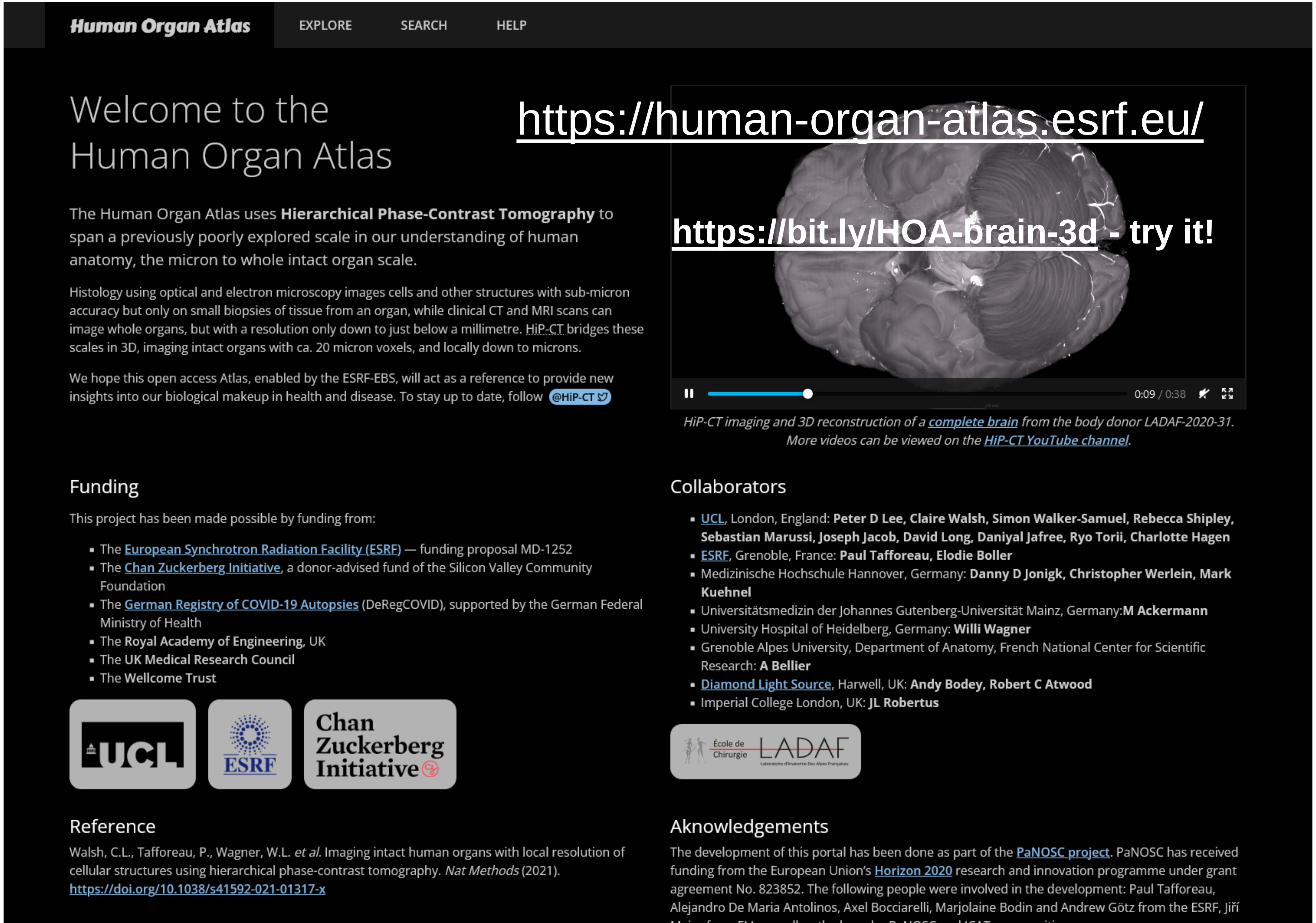
EXAMPLE OF ESRF DATA – HUMAN ORGAN ATLAS

A recent example of data from the **ESRF** is the **Human Organ Atlas** https://human-organ-atlas.esrf.eu/

The data represent the highest resolution scanning of individual human organs made possible by the new **4th generation source - EBS**

The data are being made **open** as soon as they are processed. **Over 30 groups world-wide are using the data.**

The goal is to make a complete atlas of the human body.



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

https://bit.ly/HOA-brain-3d - try it!

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.

Funding

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**

Reference

Walsh, C.L., Tafforeau, P., Wagner, W.L. *et al.* Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography. *Nat Methods* (2021). https://doi.org/10.1038/s41592-021-01317-x

Aknowledgements

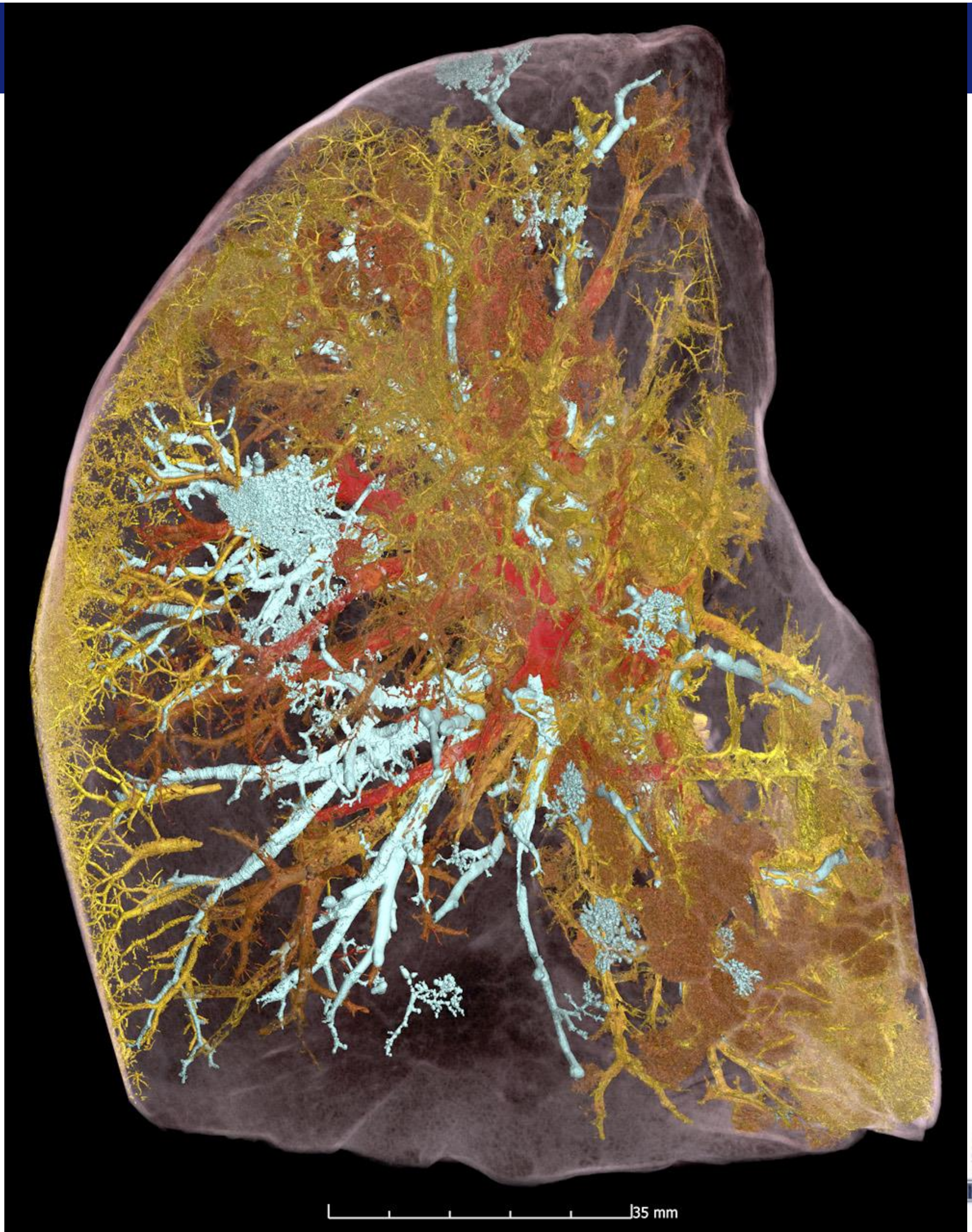
The development of this portal has been done as part of the PaNOSC project. PaNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852. The following people were involved in the development: Paul Tafforeau, Alejandro De Maria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

LUNG OF COVID19 PATIENT

National Geographic's favorite science photos in 2021

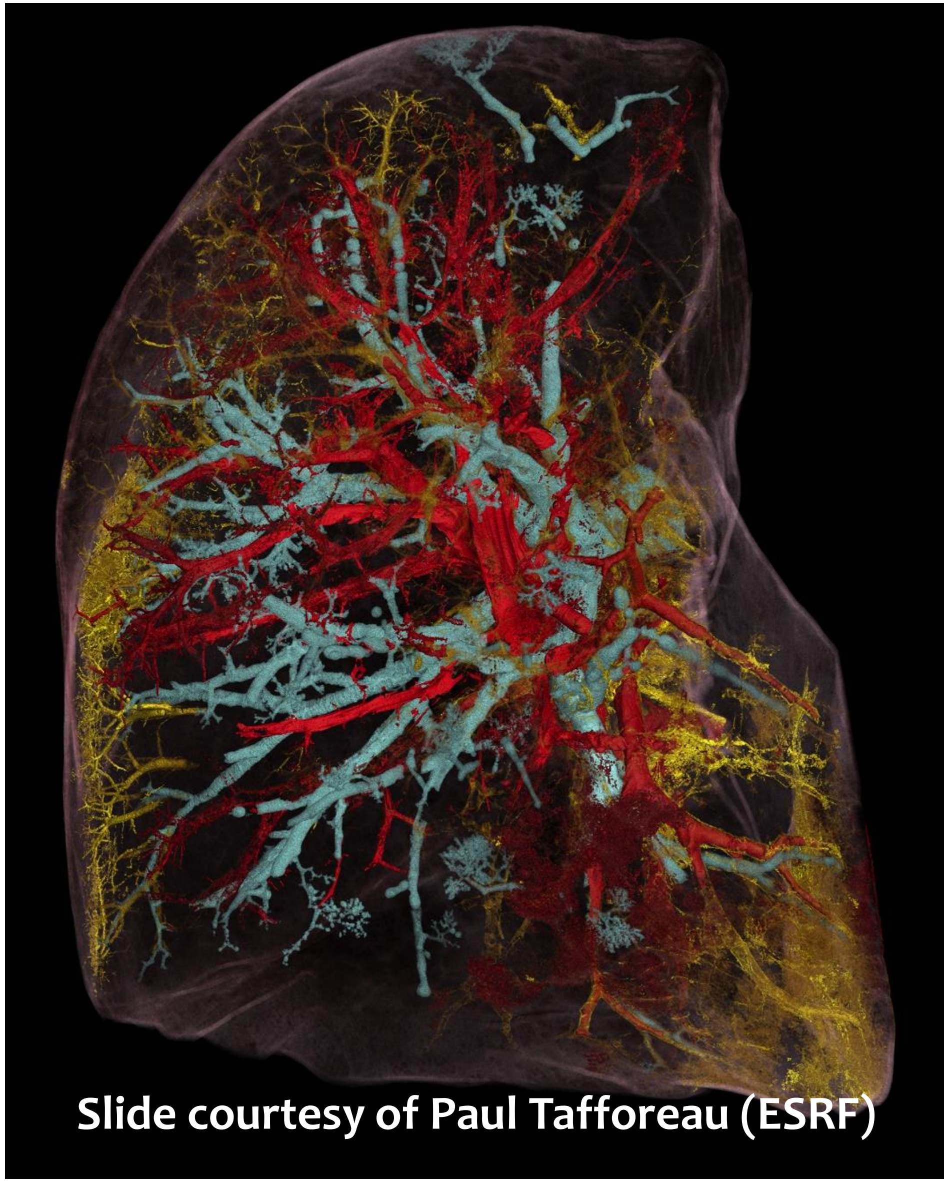
https://www.nationalgeographic.com/science/article/worlds-brightest-x-rays-reveal-covid-19-damage-to-the-body

short link: https://on.natgeo.com/3wXg3p2

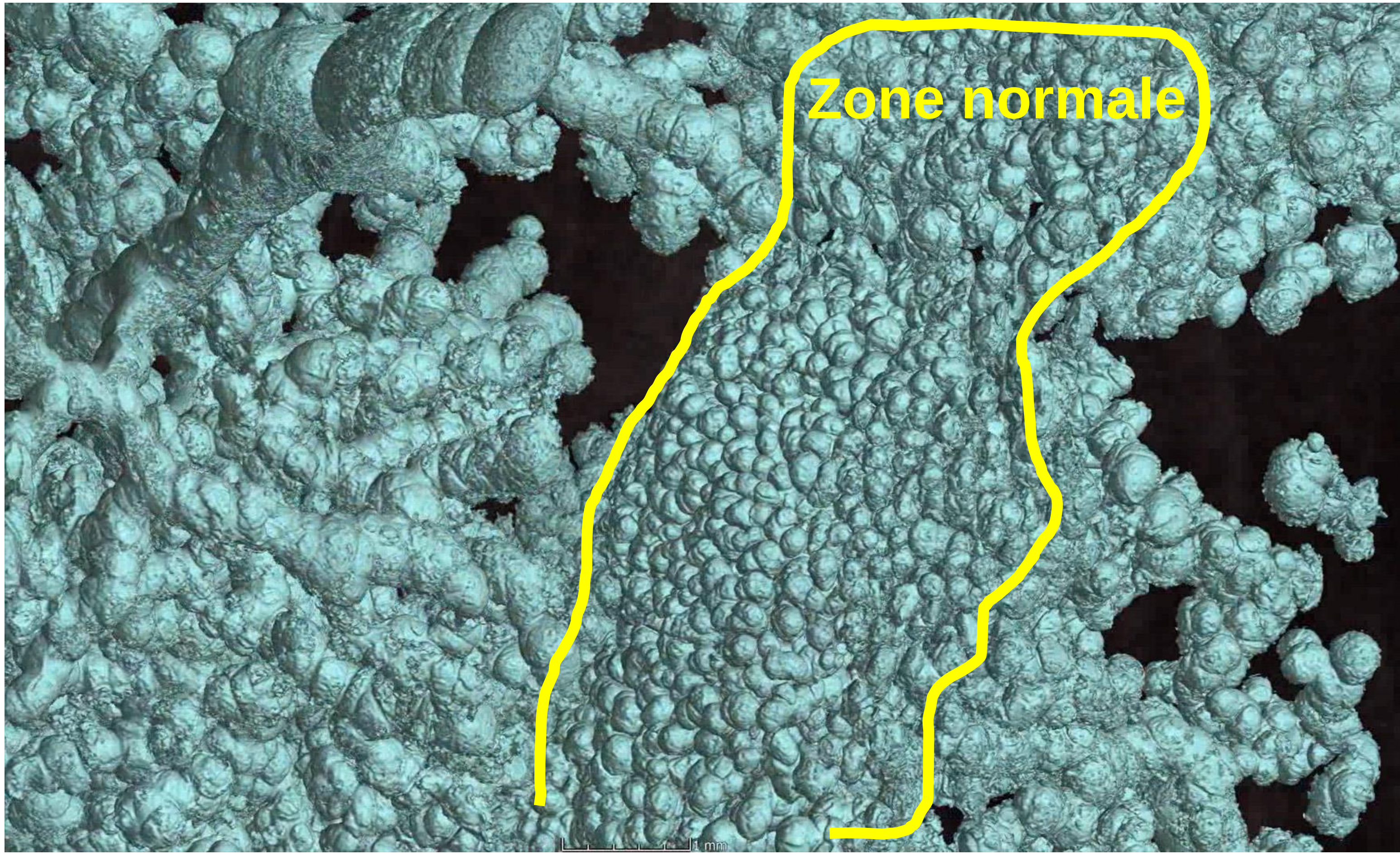


ESRF

OPEN DATA – COVID EXAMPLE



Slide courtesy of Paul Tafforeau (ESRF)



https://human-organ-atlas.esrf.eu/datasets/571998122

WHAT DO ESRF DATA LOOK LIKE ?

Data types

- Preferred format is HDF5 a hierarchical binary format for storing all data and metadata. HDF5 is used for archiving raw and processed data. We have developed tools for browsing, viewing and accessing HDF5 files.
- Additional formats are used for analysis programs e.g. tiff, cif, CSV, …

Raw Data

- 2D images from detectors (cameras) from 1 megapixel to 64+ megapixels
- 2D movies of particles (cryo-electron microscopy)
- 1D and 0D arrays (spectroscopy)

Processed Data

- 3D volumes representing models of the sample
- 3D models of electron distribution of proteins
- 2D movies of samples reactions to changes
- 2D maps of elemental distributions in samples
- 1D plots of diffraction images / spectroscopy

WHAT DATA SERVICES DOES ESRF PROVIDE?

ESRF USERS

- Experimental team who generated the data profit most from data services:
  1. *Rich metadata collected automatically + curated*
  2. *Raw data curated for (at least) 10 years*
  3. *Exclusive access for (at least) 3 years*
  4. *Efficient download of large volumes*
  5. *DOI for raw and processed data*
  6. *Searchable electronic logbook*
  7. *Data searching + viewing*

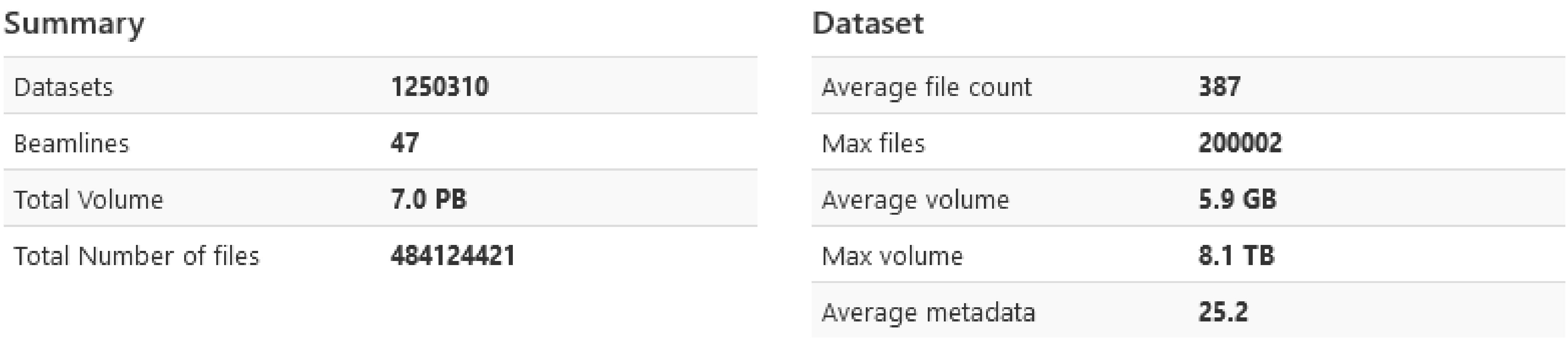
USERS of OPEN DATA

- *All the above services as soon as data are made open (after 3 years)*

OPEN DATA for AI/ML

- *The above services are available but not optimized for machines*

TOTAL CURATED DATA PRODUCTION SINCE 2015



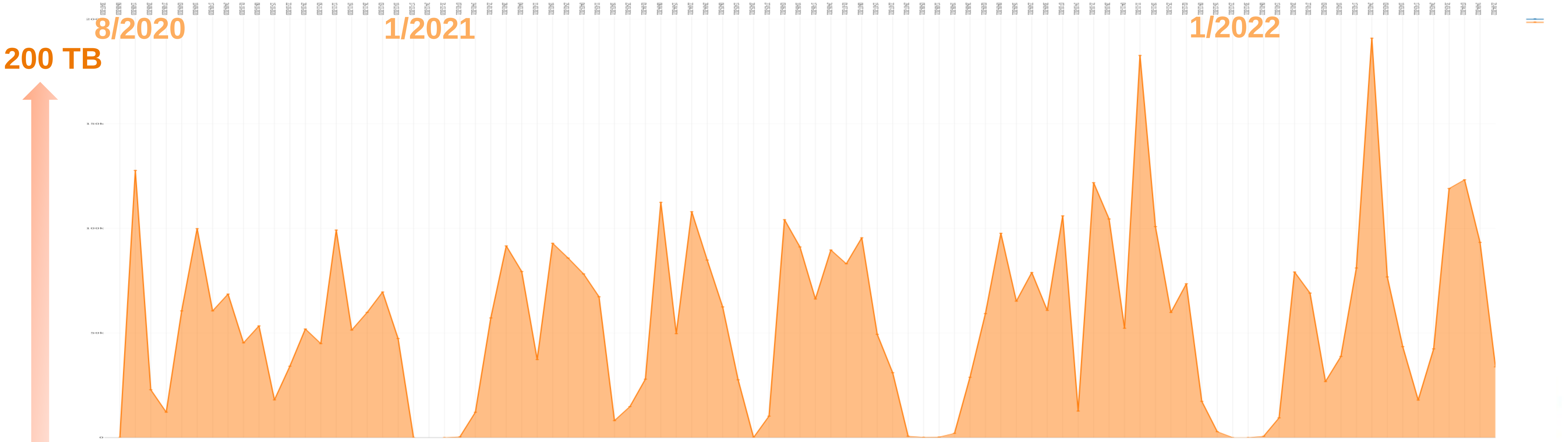
Summary

Datasets	1250310
Beamlines	47
Total Volume	7.0 PB
Total Number of files	484124421

Dataset

Average file count	387
Max files	200002
Average volume	5.9 GB
Max volume	8.1 TB
Average metadata	25.2

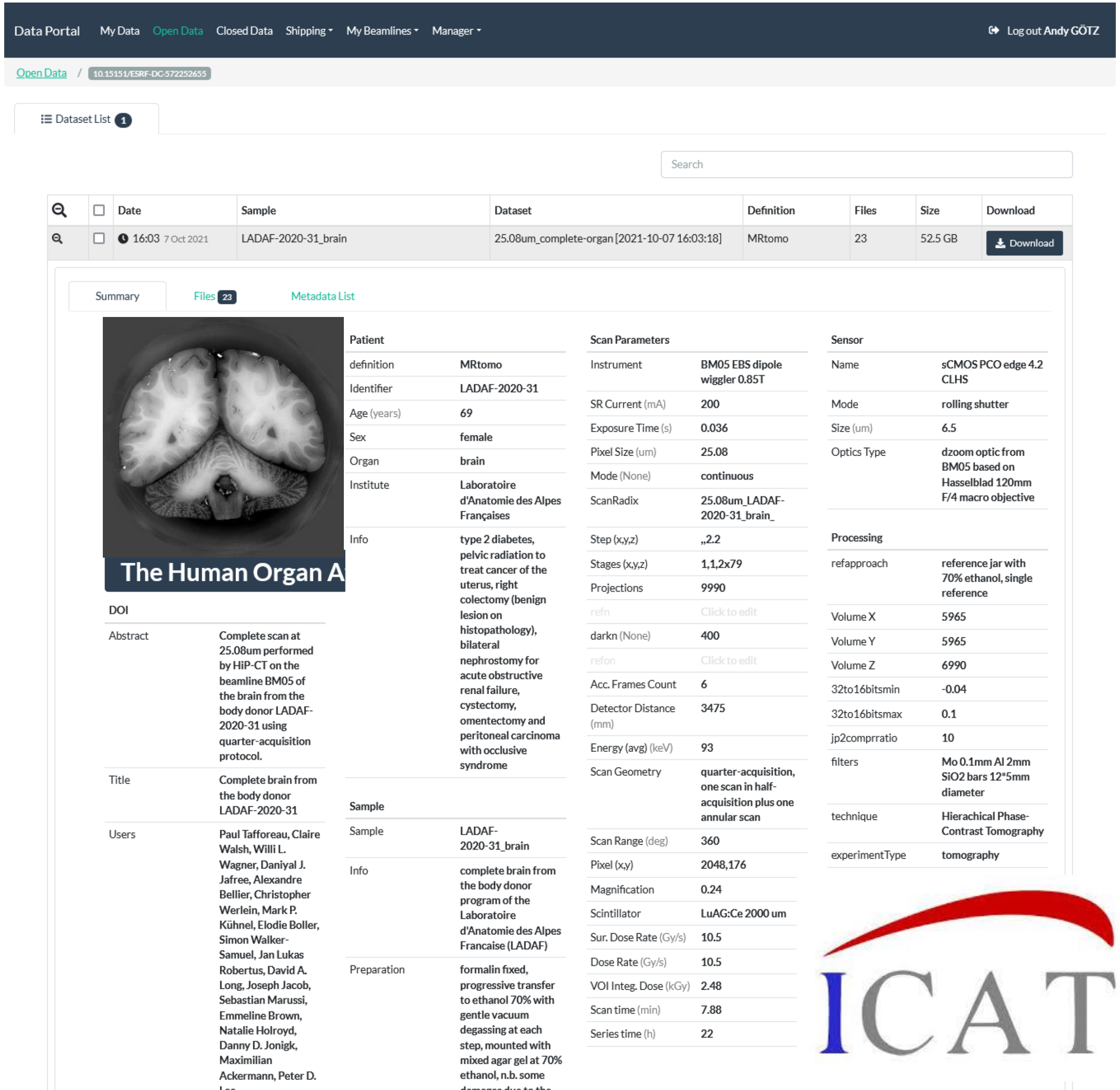
DATA CURATED WEEKLY SINCE 1/8/2020 – PEAK = 200 TB



ICAT METADATA CATALOGUE

Metadata catalogue

- **ICAT Catalogue is developed by STFC**
- **ICAT provides:**
  - Generic data model
  - Robust fine-grained user authorization
- **ESRF added:**
  - New User Interface
  - SSO login via openid
  - DOI landing page support
  - Sample shipping + tracking
  - Search based on Elasticsearch
  - E-logbook for experiments+beamlines
- **For more info: https://github.com/icatproject**



	□	Date	Sample	Dataset	Definition	Files	Size	Download
	□	16:03 7 Oct 2021	LADAF-2020-31_brain	25.08um_complete-organ [2021-10-07 16:03:18]	MRtomo	23	52.5 GB	Download

DOI	
Abstract	Complete scan at 25.08um performed by HiP-CT on the beamline BM05 of the brain from the body donor LADAF-2020-31 using quarter-acquisition protocol.
Title	Complete brain from the body donor LADAF-2020-31
Users	Paul Tafforeau, Claire Walsh, Willi L. Wagner, Daniyal J. Jafree, Alexandre Bellier, Christopher Werlein, Mark P. Kühnel, Elodie Boller, Simon Walker-Samuel, Jan Lukas Robertus, David A. Long, Joseph Jacob, Sebastian Marussi, Emmeline Brown, Natalie Holroyd, Danny D. Jonigk, Maximilian Ackermann, Peter D.

Patient	
definition	MRtomo
Identifier	LADAF-2020-31
Age (years)	69
Sex	female
Organ	brain
Institute	Laboratoire d'Anatomie des Alpes Françaises
Info	type 2 diabetes, pelvic radiation to treat cancer of the uterus, right colectomy (benign lesion on histopathology), bilateral nephrostomy for acute obstructive renal failure, cystectomy, omentectomy and peritoneal carcinoma with occlusive syndrome

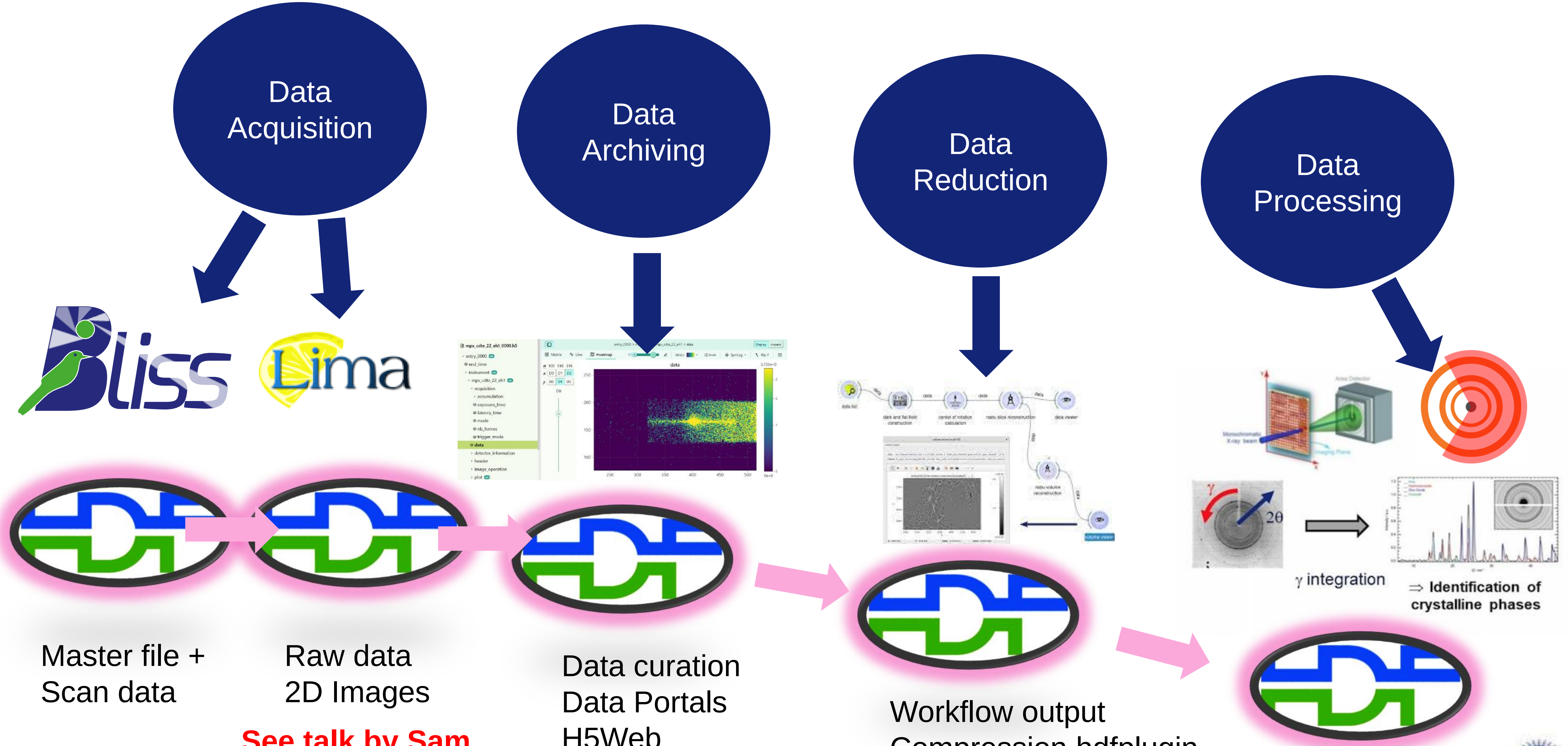
Sample	
Sample	LADAF-2020-31_brain
Info	complete brain from the body donor program of the Laboratoire d'Anatomie des Alpes Francaise (LADAF)
Preparation	formalin fixed, progressive transfer to ethanol 70% with gentle vacuum degassing at each step, mounted with mixed agar gel at 70% ethanol, n.b. some

Scan Parameters	
Instrument	BM05 EBS dipole wiggler 0.85T
SR Current (mA)	200
Exposure Time (s)	0.036
Pixel Size (um)	25.08
Mode (None)	continuous
ScanRadix	25.08um_LADAF-2020-31_brain_
Step (x,y,z)	,,2.2
Stages (x,y,z)	1,1,2x79
Projections	9990
refn	Click to edit
darkn (None)	400
refon	Click to edit
Acc. Frames Count	6
Detector Distance (mm)	3475
Energy (avg) (keV)	93
Scan Geometry	quarter-acquisition, one scan in half-acquisition plus one annular scan
Scan Range (deg)	360
Pixel (x,y)	2048,176
Magnification	0.24
Scintillator	LuAG:Ce 2000 um
Sur. Dose Rate (Gy/s)	10.5
Dose Rate (Gy/s)	10.5
VOI Integ. Dose (kGy)	2.48
Scan time (min)	7.88
Series time (h)	22

Sensor	
Name	sCMOS PCO edge 4.2 CLHS
Mode	rolling shutter
Size (um)	6.5
Optics Type	dzoom optic from BM05 based on Hasselblad 120mm F/4 macro objective

Processing	
refapproach	reference jar with 70% ethanol, single reference
Volume X	5965
Volume Y	5965
Volume Z	6990
32to16bitsmin	-0.04
32to16bitsmax	0.1
jp2comprratio	10
filters	Mo 0.1mm Al 2mm SiO2 bars 12*5mm diameter
technique	Hierachical Phase-Contrast Tomography
experimentType	tomography

HDF5/NEXUS - CENTRAL TO THE EBS DATA STRATEGY



Master file + Scan data

Raw data 2D Images

See talk by Sam

Data curation Data Portals H5Web

See talk by Loic

Workflow output Compression hdfplugin

See talk by Thomas

Analysis results

WHY DID ESRF CHOOSE HDF5?

WHY DID ESRF CHOOSE HDF5?

TO ADDRESS the FOLLOWING ISSUES

1. Reduce the number of files (pre-hdf5 we had 1 image per file)
2. Adopt the community metadata ontology (Nexus)
3. Support multiple compression schemes (not only gzip)
4. Integrate data produced by detector companies e.g. Dectris
5. Mix metadata with raw data without limitations
   - *A single master file to access all data from an experiment*
6. Use a standardized API supported for multiple languages
   - *Especially for C and Python and Matlab*
7. Efficient reading and writing performance of binary data
   - *New experiments produce more and more data (giga- to terabytes)*
8. Guaranteed to be supported for a long time (decades)

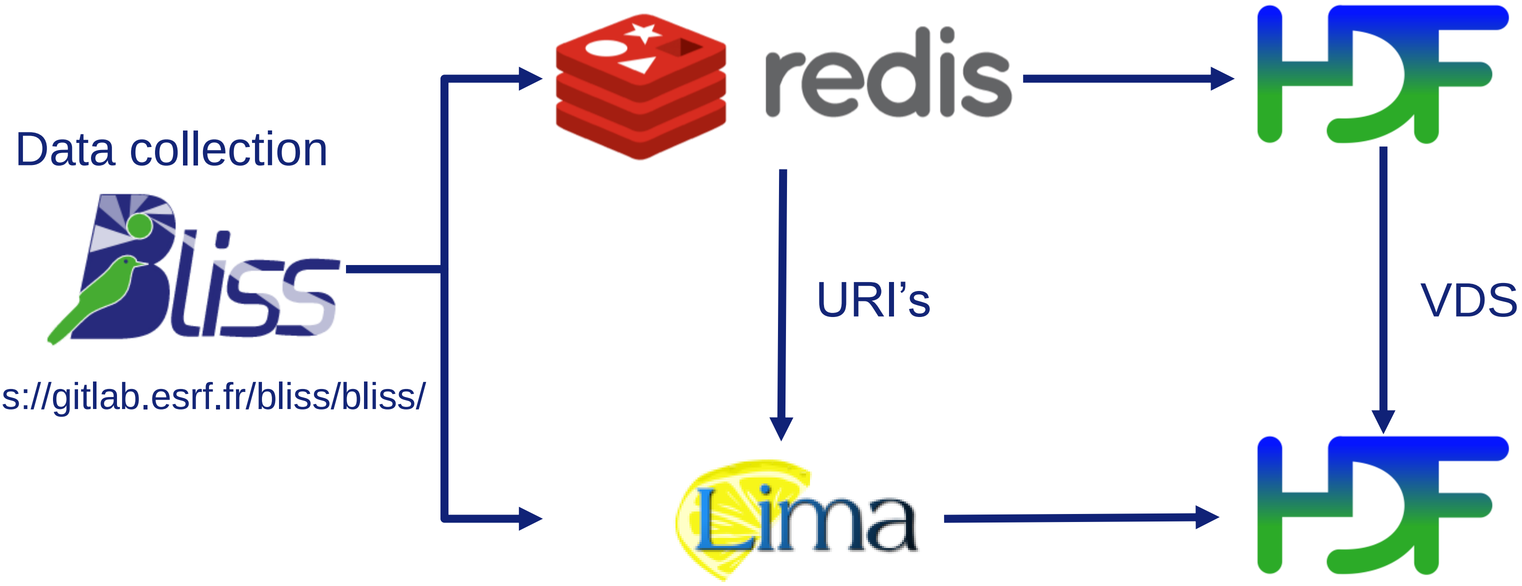
HDF5 FOR DATA ACQUISITION

Transient storage

- Metadata
- 0D/1D data
- 2D URI's

Persistent storage

- Metadata
- 0D to 3D HDF5 datasets



Data collection

https://gitlab.esrf.fr/bliss/bliss/

URI's

VDS

Transient storage

- Metadata
- 2D data

See talk on Lima/Lima2

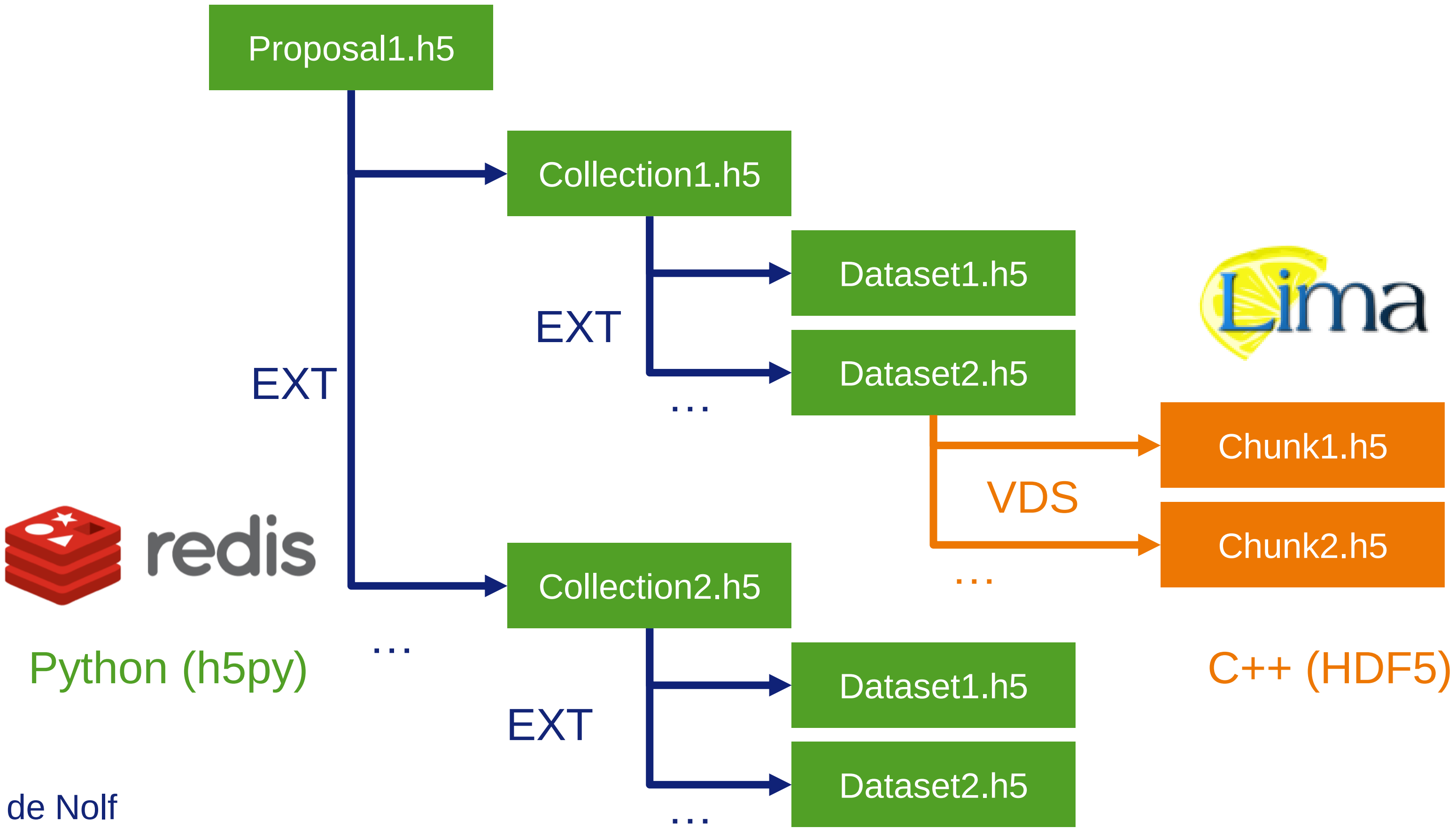
Distributed file system

- Writing: GPFS, NFS
- Reading: GPFS, NFS, SMB

Slide courtesy of Wout de Nolf

HDF5 AT THE ESRF: DATA COLLECTION

On Distributed File System



Proposal1.h5

EXT

Collection1.h5

EXT

Dataset1.h5

Dataset2.h5

…

VDS

Chunk1.h5

Chunk2.h5

…

Collection2.h5

…

EXT

Dataset1.h5

Dataset2.h5

…

redis

Python (h5py)

Lima

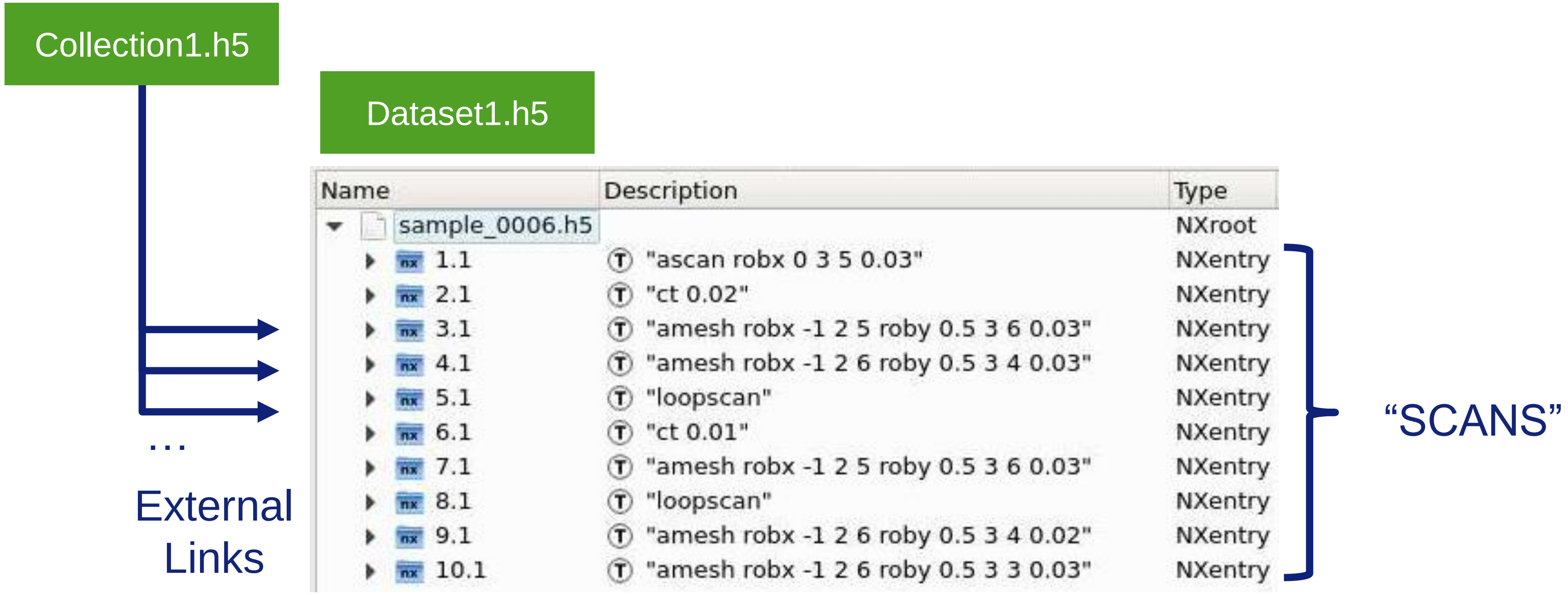
C++ (HDF5)

Slide courtesy of Wout de Nolf

HDF5 AT THE ESRF: DATA COLLECTION

HDF + NX

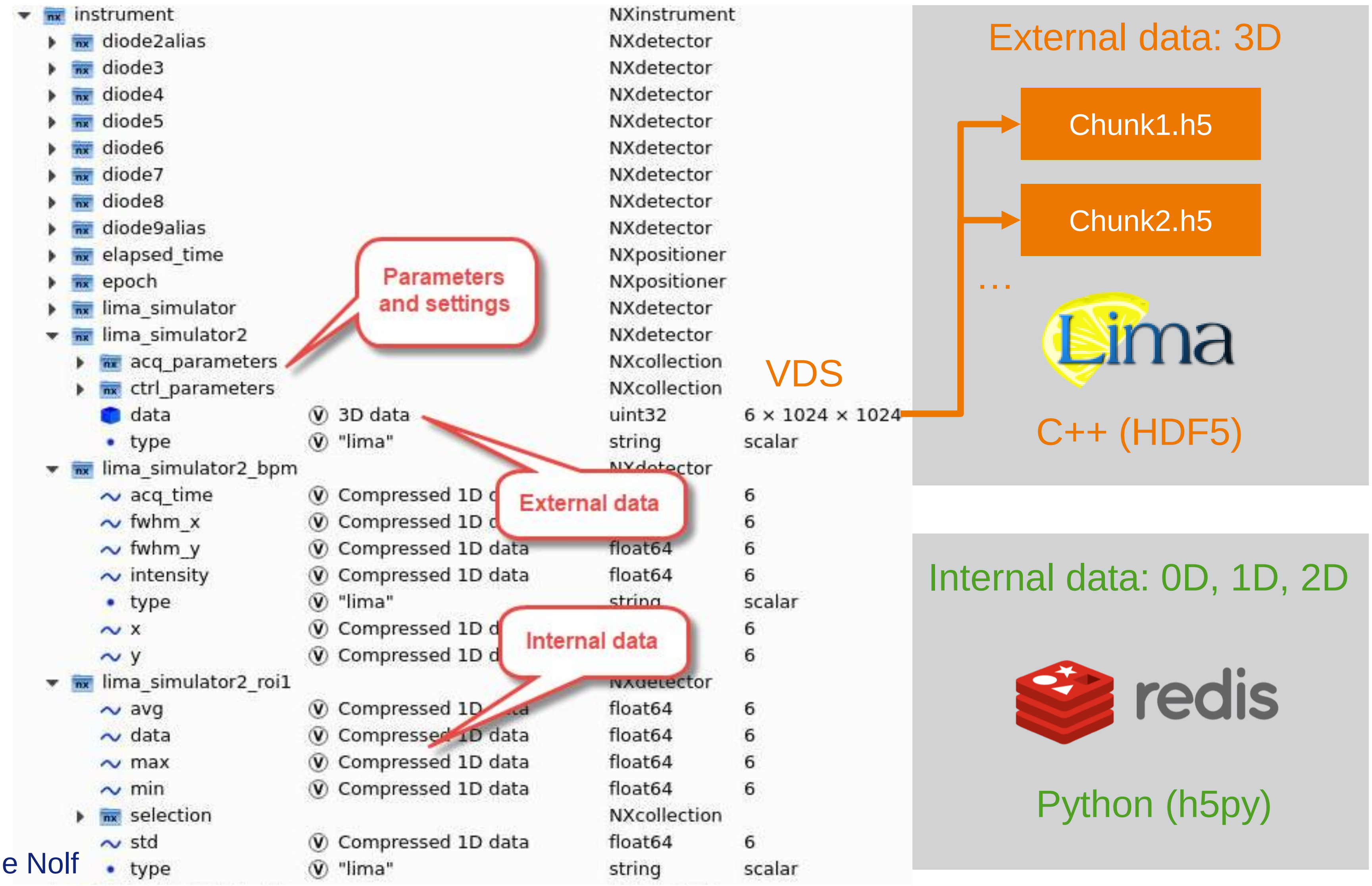
NeXus Data Format



Name	Description	Type
sample_0006.h5		NXroot
1.1	"ascan robx 0 3 5 0.03"	NXentry
2.1	"ct 0.02"	NXentry
3.1	"amesh robx -1 2 5 roby 0.5 3 6 0.03"	NXentry
4.1	"amesh robx -1 2 6 roby 0.5 3 4 0.03"	NXentry
5.1	"loopscan"	NXentry
6.1	"ct 0.01"	NXentry
7.1	"amesh robx -1 2 5 roby 0.5 3 6 0.03"	NXentry
8.1	"loopscan"	NXentry
9.1	"amesh robx -1 2 6 roby 0.5 3 4 0.02"	NXentry
10.1	"amesh robx -1 2 6 roby 0.5 3 3 0.03"	NXentry

Slide courtesy of Wout de Nolf

HDF5 AT THE ESRF: DATA COLLECTION



instrument		NXinstrument	
diode2alias		NXdetector	
diode3		NXdetector	
diode4		NXdetector	
diode5		NXdetector	
diode6		NXdetector	
diode7		NXdetector	
diode8		NXdetector	
diode9alias		NXdetector	
elapsed_time		NXpositioner	
epoch		NXpositioner	
lima_simulator		NXdetector	
lima_simulator2		NXdetector	
acq_parameters		NXcollection	
ctrl_parameters		NXcollection	
data	Ⓥ 3D data	uint32	6 × 1024 × 1024
type	Ⓥ "lima"	string	scalar
lima_simulator2_bpm		NXdetector	
acq_time	Ⓥ Compressed 1D d		6
fwhm_x	Ⓥ Compressed 1D d		6
fwhm_y	Ⓥ Compressed 1D data	float64	6
intensity	Ⓥ Compressed 1D data	float64	6
type	Ⓥ "lima"	string	scalar
x	Ⓥ Compressed 1D d		6
y	Ⓥ Compressed 1D d		6
lima_simulator2_roi1		NXdetector	
avg	Ⓥ Compressed 1D	float64	6
data	Ⓥ Compressed 1D data	float64	6
max	Ⓥ Compressed 1D data	float64	6
min	Ⓥ Compressed 1D data	float64	6
selection		NXcollection	
std	Ⓥ Compressed 1D data	float64	6
type	Ⓥ "lima"	string	scalar

Slide courtesy of Wout de Nolf

HDF5 AT THE ESRF: DATA COLLECTION

HDF5 features used for data collection

- Vanilla HDF5 (Groups, Datasets, Attributes, Softlinks)
- External Links (EXT)
- Virtual DataSets (VDS)
- Variable length data types: only for strings
- Growing datasets during acquisition
- Chunking and compression

No SWMR
No Parallel HDF5

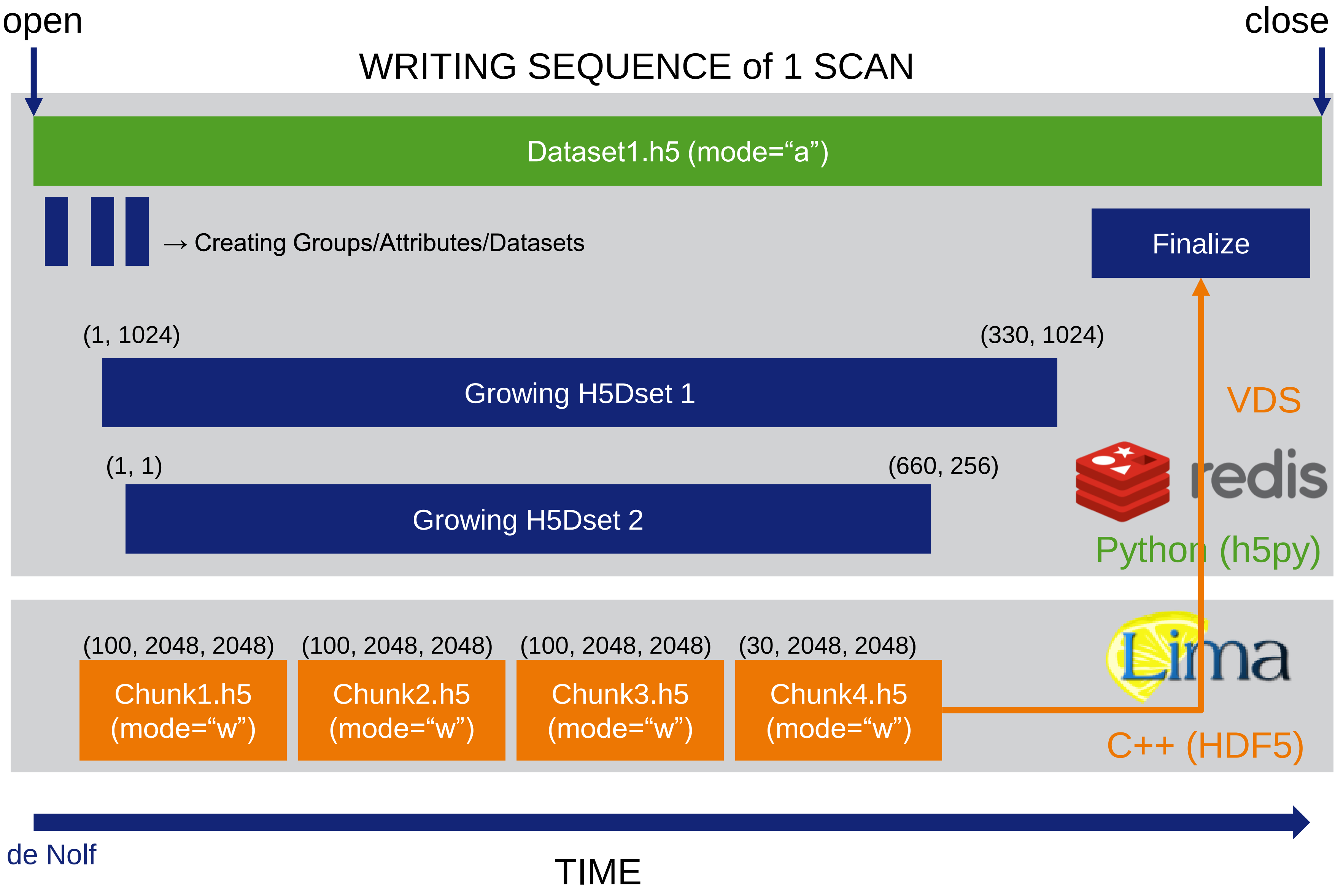
Distributed file system

- Writing: GPFS, NFS
- Reading: GPFS, NFS, SMB

No control over readers and their access mode

Slide courtesy of Wout de Nolf

HDF5 AT THE ESRF: DATA COLLECTION



open

close

WRITING SEQUENCE of 1 SCAN

Dataset1.h5 (mode="a")

→ Creating Groups/Attributes/Datasets

Finalize

(1, 1024)

(330, 1024)

Growing H5Dset 1

(1, 1)

(660, 256)

Growing H5Dset 2

VDS

redis

Python (h5py)

(100, 2048, 2048) (100, 2048, 2048) (100, 2048, 2048) (30, 2048, 2048)

Chunk1.h5 (mode="w") Chunk2.h5 (mode="w") Chunk3.h5 (mode="w") Chunk4.h5 (mode="w")

Lima

C++ (HDF5)

TIME

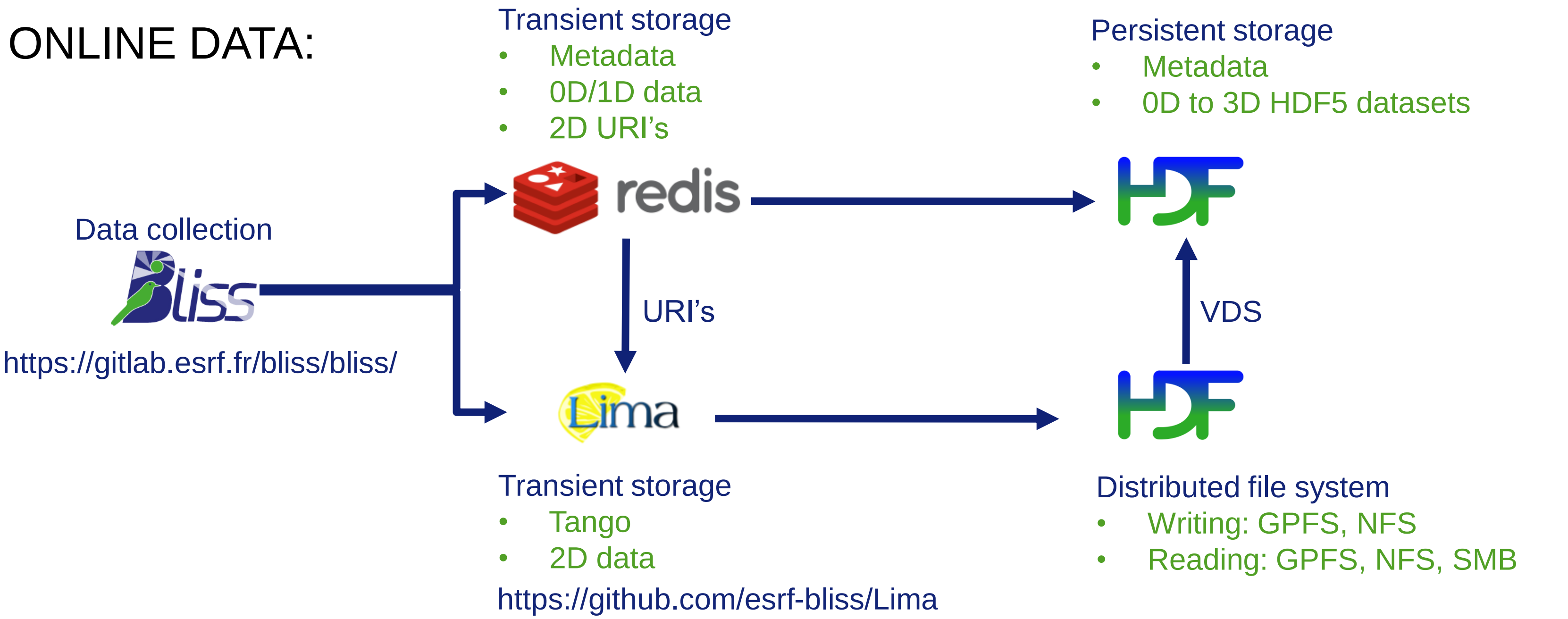
MAIN ISSUE: **close** can be hours / days after **open** BUT apps need to read data to do online processing

QUESTION: Is SWMR2 the Answer?

Slide courtesy of Wout de Nolf

ONLINE DATA ACCESS USING H5PY API

ONLINE DATA:



DATA ACCESS:

Single Python API

The API should be identical online (streaming/changing HDF5) and offline (static HDF5)

H5py inspired API: data tree with nodes
- groups have a python Mapping API
- datasets have a numpy array API
- attributes have a python Mapping API

The dynamic nature of the tree is reflected in the iterators.

Specific yield/stop conditions may need to be introduced.

Slide courtesy of Wout de Nolf

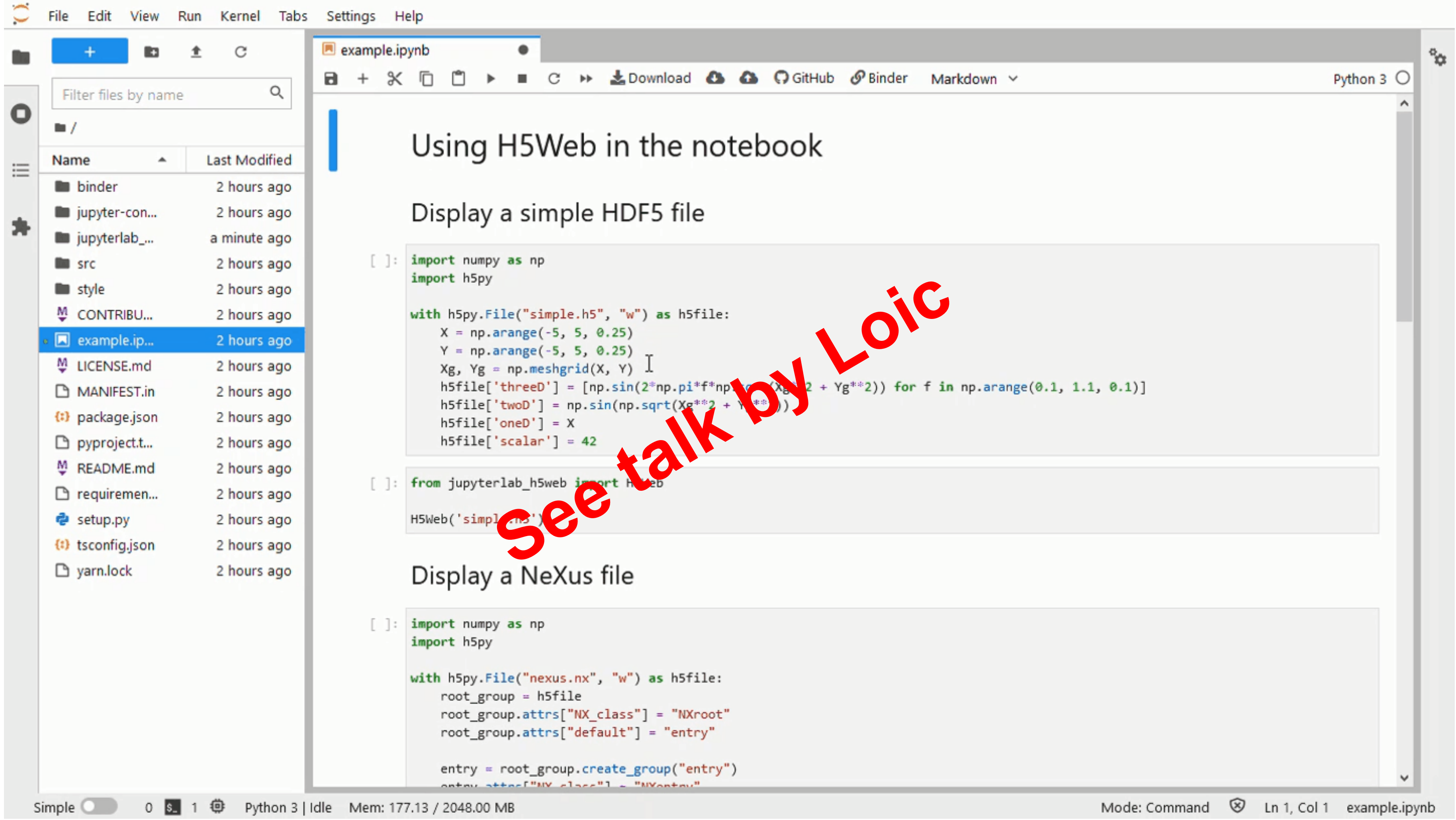
TOOLS TO VISUALIZE DATA – WEB BASED

PaNOSC developed **H5web** web-based viewer of HDF5 files and integrated it in Jupyterlab, data portals, + web applications:

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

https://h5web.panosc.eu/

Next step : 3D viewer?



Name	Last Modified
binder	2 hours ago
jupyter-con...	2 hours ago
jupyterlab_...	a minute ago
src	2 hours ago
style	2 hours ago
CONTRIBU...	2 hours ago
example.ip...	2 hours ago
LICENSE.md	2 hours ago
MANIFEST.in	2 hours ago
package.json	2 hours ago
pyproject.t...	2 hours ago
README.md	2 hours ago
requiremen...	2 hours ago
setup.py	2 hours ago
tsconfig.json	2 hours ago
yarn.lock	2 hours ago

Using H5Web in the notebook

Display a simple HDF5 file

```
[ ]: import numpy as np
     import h5py

     with h5py.File("simple.h5", "w") as h5file:
         X = np.arange(-5, 5, 0.25)
         Y = np.arange(-5, 5, 0.25)
         Xg, Yg = np.meshgrid(X, Y)
         h5file['threeD'] = [np.sin(2*np.pi*f*np.sqrt(Xg**2 + Yg**2)) for f in np.arange(0.1, 1.1, 0.1)]
         h5file['twoD'] = np.sin(np.sqrt(Xg**2 + Yg**2))
         h5file['oneD'] = X
         h5file['scalar'] = 42
```

```
[ ]: from jupyterlab_h5web import H5Web

     H5Web('simple.h5')
```

See talk by Loic

Display a NeXus file

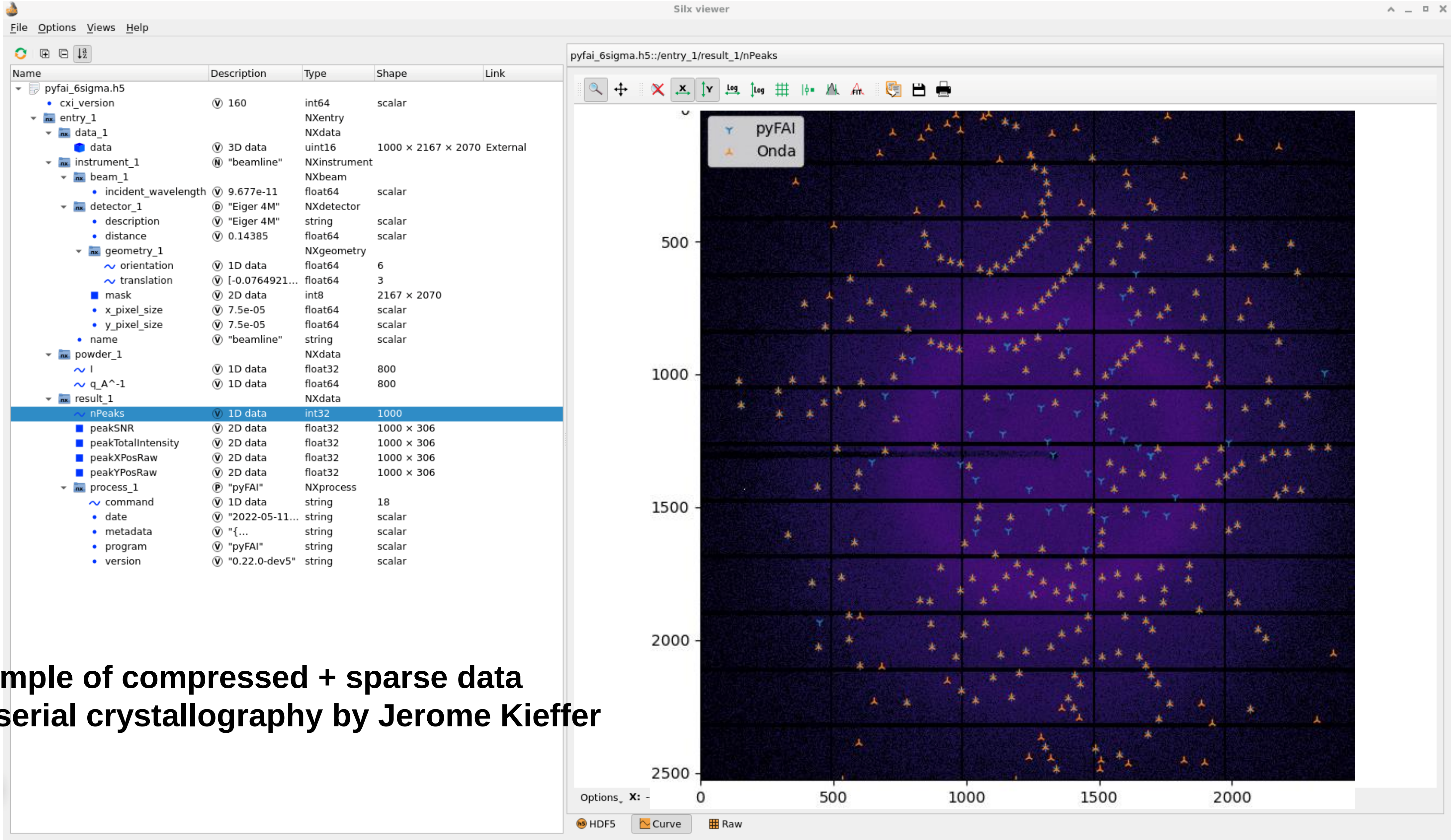
```
[ ]: import numpy as np
     import h5py

     with h5py.File("nexus.nx", "w") as h5file:
         root_group = h5file
         root_group.attrs["NX_class"] = "NXroot"
         root_group.attrs["default"] = "entry"

         entry = root_group.create_group("entry")
         entry.attrs["NX_class"] = "NXentry"
```

https://h5web.panosc.eu/

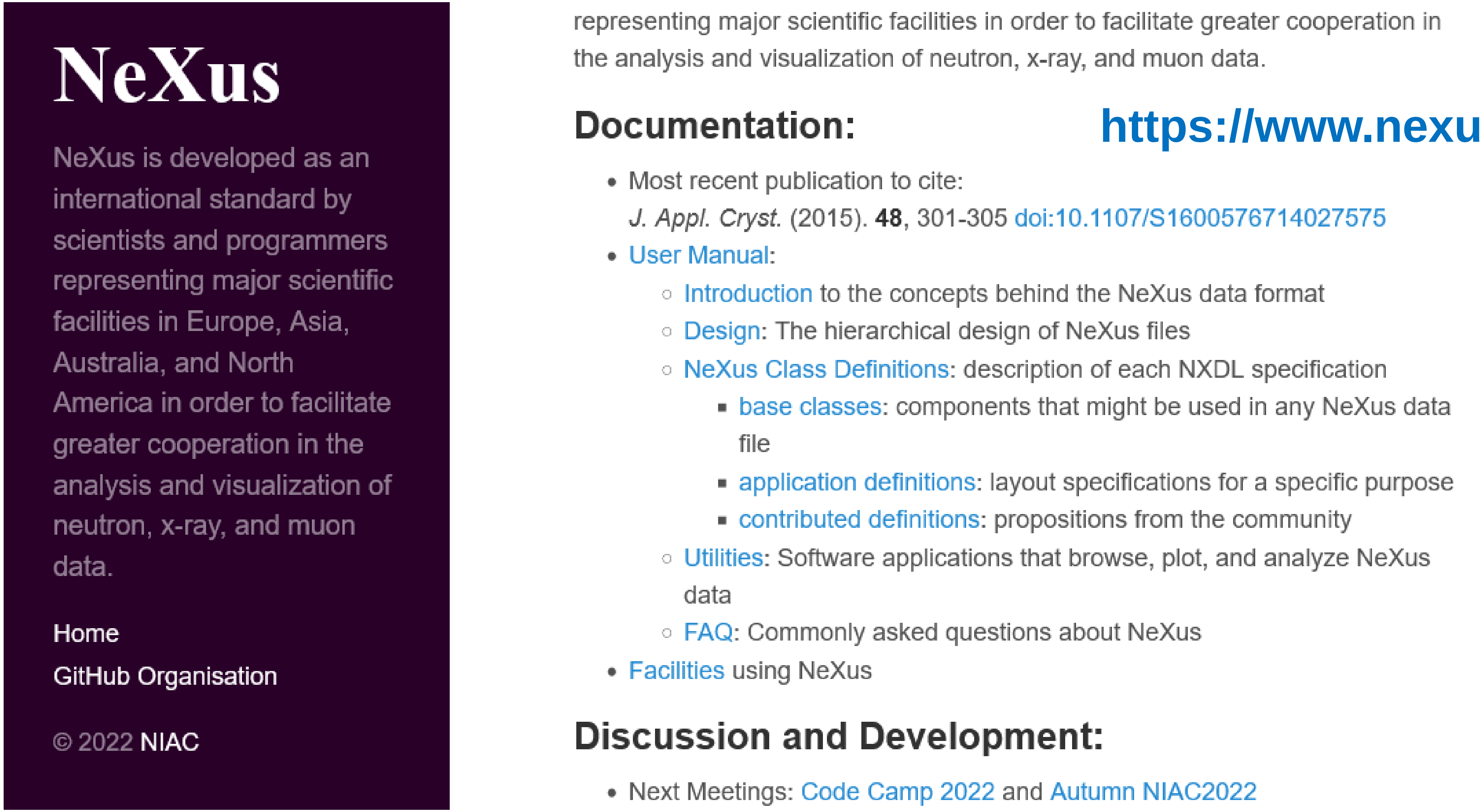
TOOLS TO VISUALIZE DATA – SILX VIEW



Name	Description	Type	Shape	Link
pyfai_6sigma.h5				
cxi_version	160	int64	scalar	
entry_1		NXentry		
data_1		NXdata		
data	3D data	uint16	1000 × 2167 × 2070	External
instrument_1	"beamline"	NXinstrument		
beam_1		NXbeam		
incident_wavelength	9.677e-11	float64	scalar	
detector_1	"Eiger 4M"	NXdetector		
description	"Eiger 4M"	string	scalar	
distance	0.14385	float64	scalar	
geometry_1		NXgeometry		
orientation	1D data	float64	6	
translation	[-0.0764921…	float64	3	
mask	2D data	int8	2167 × 2070	
x_pixel_size	7.5e-05	float64	scalar	
y_pixel_size	7.5e-05	float64	scalar	
name	"beamline"	string	scalar	
powder_1		NXdata		
I	1D data	float32	800	
q_A^-1	1D data	float64	800	
result_1		NXdata		
nPeaks	1D data	int32	1000	
peakSNR	2D data	float32	1000 × 306	
peakTotalIntensity	2D data	float32	1000 × 306	
peakXPosRaw	2D data	float32	1000 × 306	
peakYPosRaw	2D data	float32	1000 × 306	
process_1	"pyFAI"	NXprocess		
command	1D data	string	18	
date	"2022-05-11…	string	scalar	
metadata	"{…	string	scalar	
program	"pyFAI"	string	scalar	
version	"0.22.0-dev5"	string	scalar	

Example of compressed + sparse data for serial crystallography by Jerome Kieffer

NEXUS METADATA STANDARD



NeXus

NeXus is developed as an international standard by scientists and programmers representing major scientific facilities in Europe, Asia, Australia, and North America in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

Home

GitHub Organisation

© 2022 NIAC

representing major scientific facilities in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

Documentation:

https://www.nexusformat.org/

- Most recent publication to cite: *J. Appl. Cryst.* (2015). **48**, 301-305 doi:10.1107/S1600576714027575
- User Manual:
  - Introduction to the concepts behind the NeXus data format
  - Design: The hierarchical design of NeXus files
  - NeXus Class Definitions: description of each NXDL specification
    - base classes: components that might be used in any NeXus data file
    - application definitions: layout specifications for a specific purpose
    - contributed definitions: propositions from the community
  - Utilities: Software applications that browse, plot, and analyze NeXus data
  - FAQ: Commonly asked questions about NeXus
- Facilities using NeXus

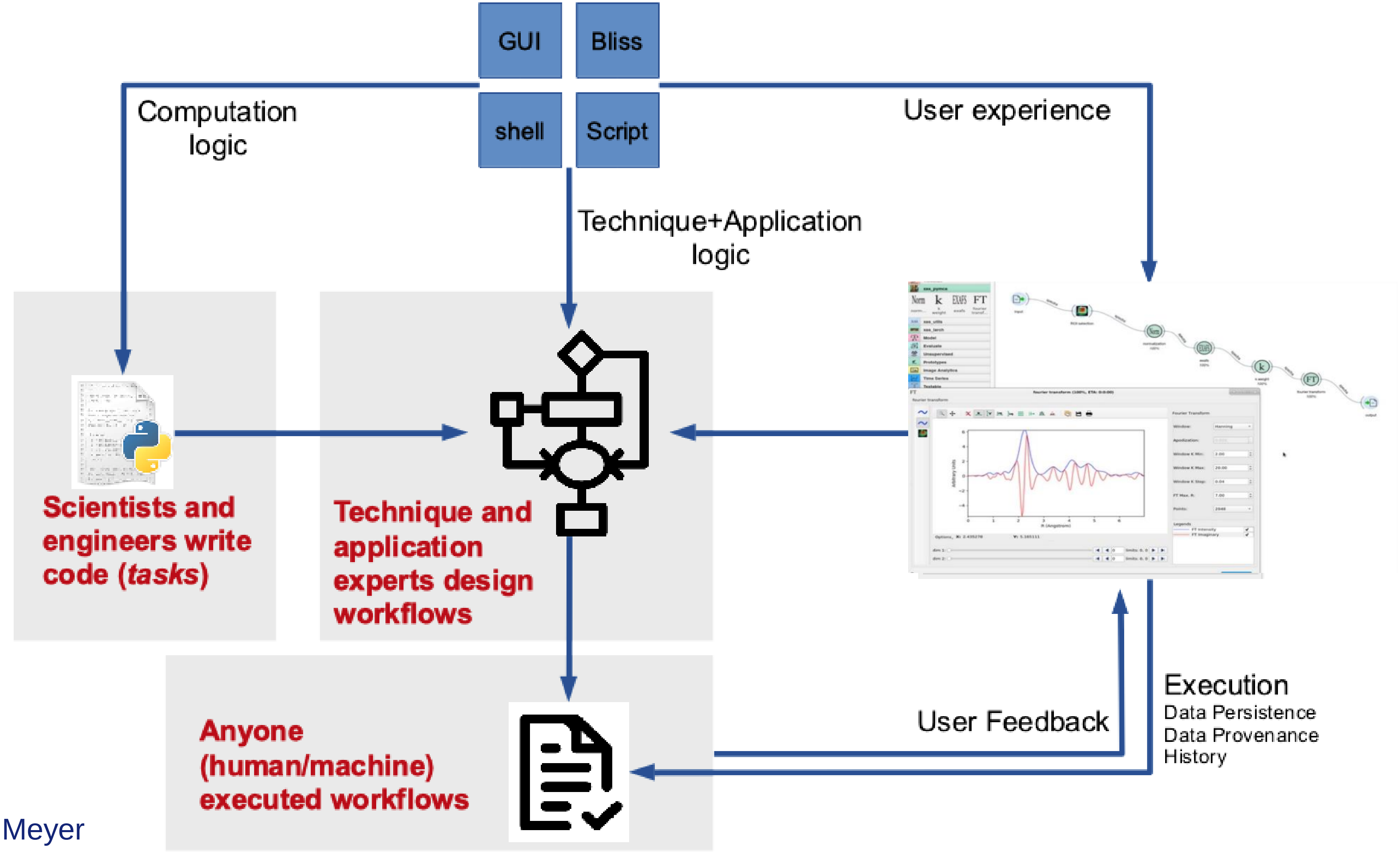
Discussion and Development:

- Next Meetings: Code Camp 2022 and Autumn NIAC2022

Community-organised metadata data standards

DATA REDUCTION AND ONLINE ANALYSIS

Data Processing Workflows



GUI

Bliss

shell

Script

Computation logic

User experience

Technique+Application logic

Scientists and engineers write code (*tasks*)

Technique and application experts design workflows

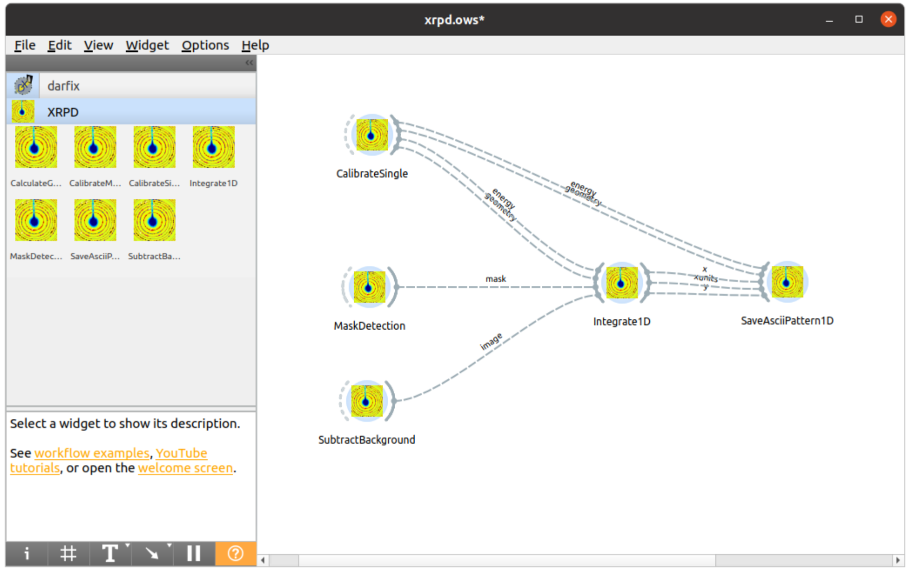
Anyone (human/machine) executed workflows

User Feedback

Execution
Data Persistence
Data Provenance
History

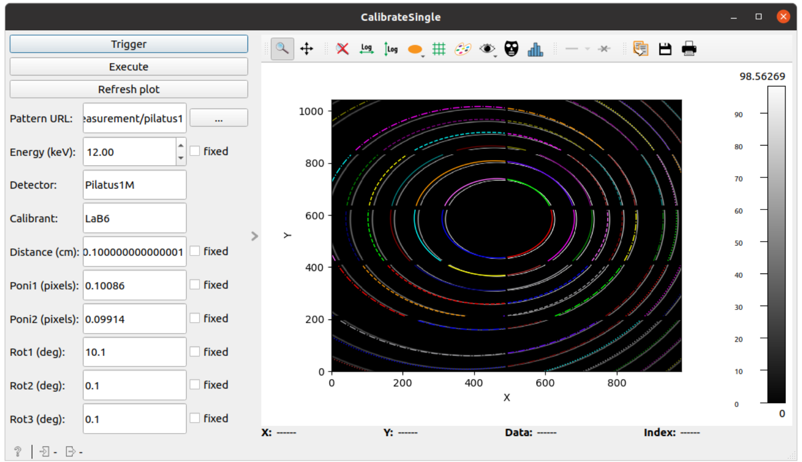
Slide courtesy of Jens Meyer

DATA REDUCTION AND ONLINE ANALYSIS



Select a widget to show its description.

See workflow examples, YouTube tutorials, or open the welcome screen.



Pattern URL: :asurement/pilatus1

Energy (keV): 12.00 □ fixed

Detector: Pilatus1M

Calibrant: LaB6

Distance (cm): 0.100000000000001 □ fixed

Poni1 (pixels): 0.10086 □ fixed

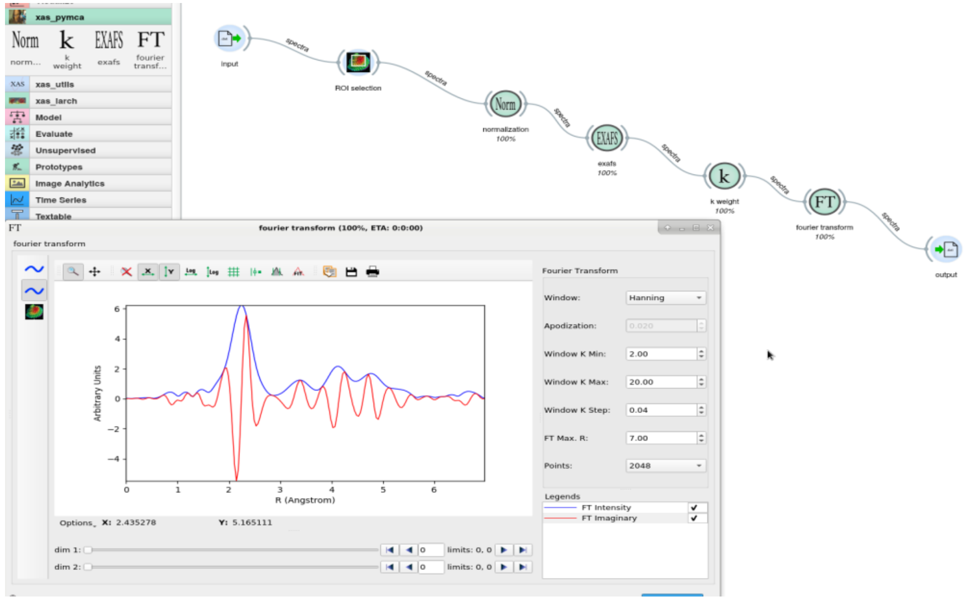
Poni2 (pixels): 0.09914 □ fixed

Rot1 (deg): 10.1 □ fixed

Rot2 (deg): 0.1 □ fixed

Rot3 (deg): 0.1 □ fixed

X: ------ Y: ------ Data: ------ Index: ------



Options_ X: 2.435278 Y: 5.165111

dim 1: 0 limits: 0, 0

dim 2: 0 limits: 0, 0

Window: Hanning

Apodization: 0.020

Window K Min: 2.00

Window K Max: 20.00

Window K Step: 0.04

FT Max. R: 7.00

Points: 2048

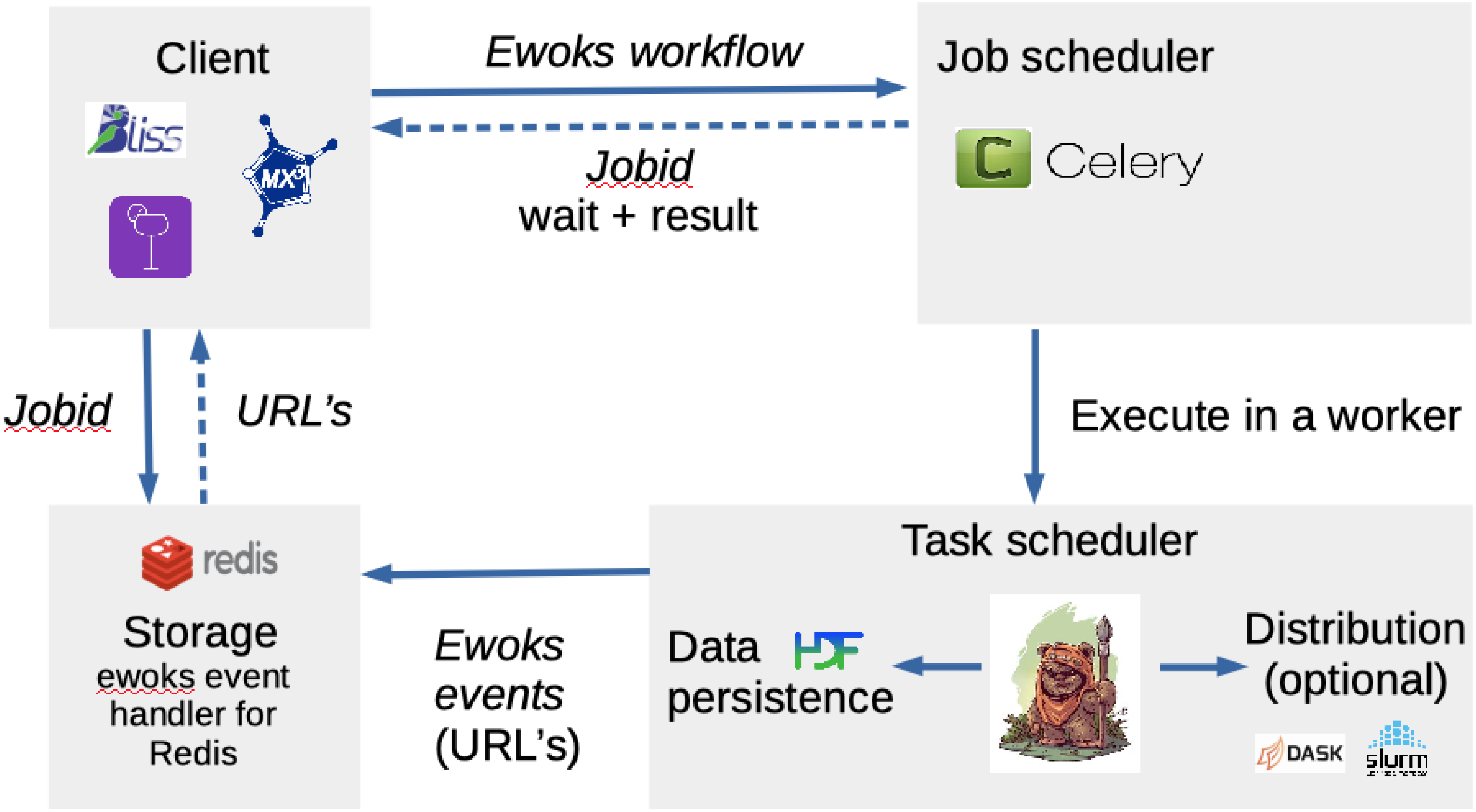
FT Intensity ✓

FT Imaginary ✓

Slide courtesy of Jens Meyer

DATA REDUCTION AND ONLINE ANALYSIS

EWoks for Online Data Processing

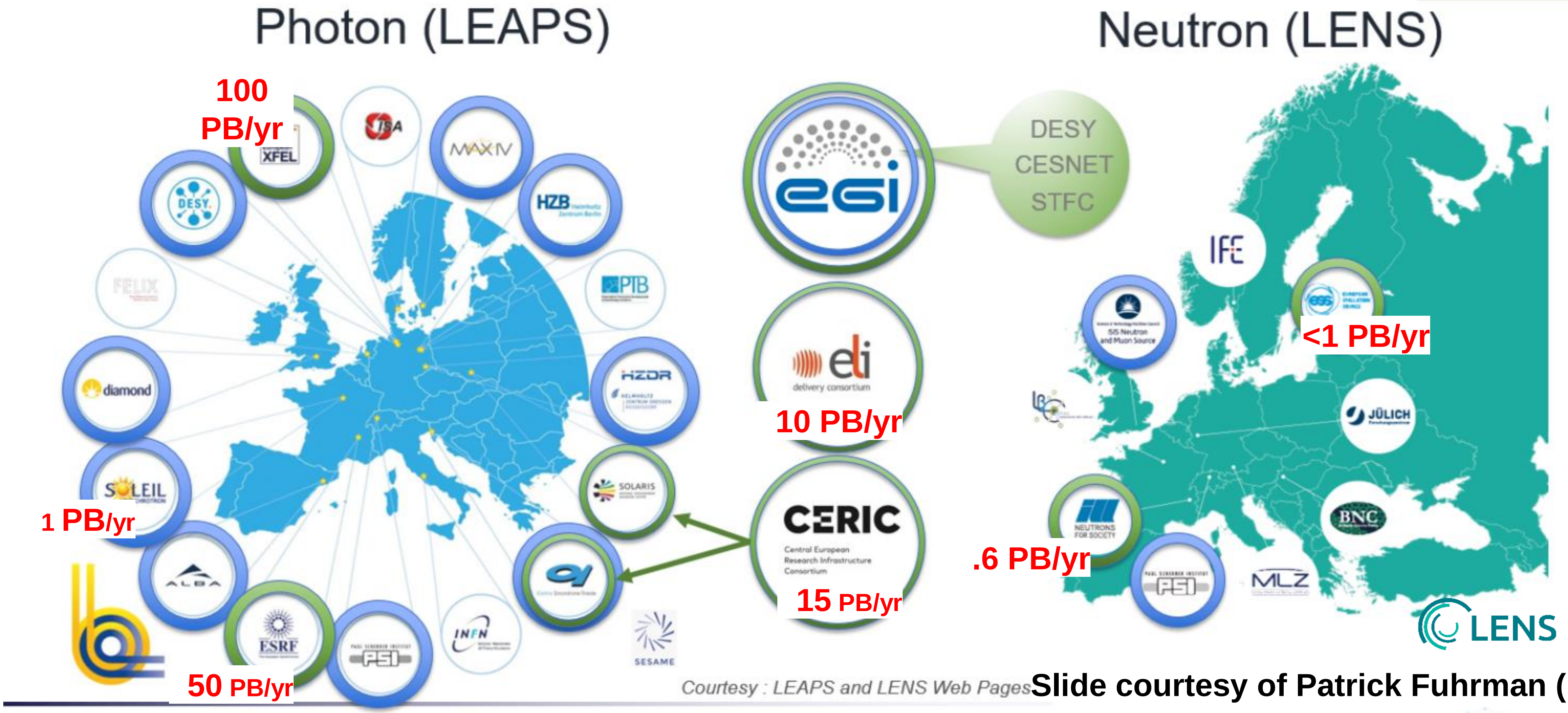


Slide courtesy of Jens Meyer

https://gitlab.esrf.fr/workflow/ewoks/ewoks

League of Photon Sources (LEAPS) and Neutrons (LENS) partners in PaNOSC and ExPaNDS

○ PaNOSC
○ ExPaNDS



100 PB/yr

10 PB/yr

<1 PB/yr

1 PB/yr

.6 PB/yr

15 PB/yr

50 PB/yr

Courtesy : LEAPS and LENS Web Pages

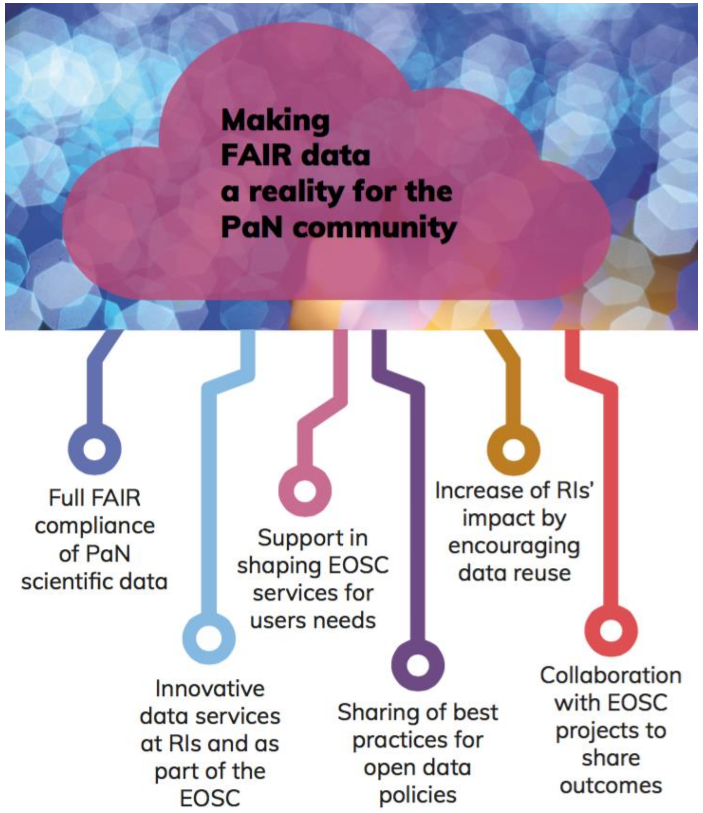
Slide courtesy of Patrick Fuhrman (DESY)

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.

EU PROJECTS PROVIDING SUPPORT FOR HDF5

PaNOSC – Photon and Neutron Open Science Cloud

- *Promoting adoption of Nexus/HDF5*
- *H5py maintenance (T.Kluywer, XFEL)*
- *H5web web viewer (A.Bocciarelli + L.Huder, ESRF)*
- *H5web in Jupyterlab (L.Huder, ESRF)*
- *HDF5 backend for OpenPMD (C.Fortmann-Grote)*


Making FAIR data a reality for the PaN community
Full FAIR compliance of PaN scientific data
Innovative data services at RIs and as part of the EOSC
Support in shaping EOSC services for users needs
Sharing of best practices for open data policies
Increase of RIs' impact by encouraging data reuse
Collaboration with EOSC projects to share outcomes

ExPaNDS is PaNOSC for national sources

- *ExPaNDS is adopting the outputs of PaNOSC*

LEAPS-INNOV WP7

- *Dedicated to data compression e.g. blosc, hdfplugin*

European Open Science Cloud should support HDF5

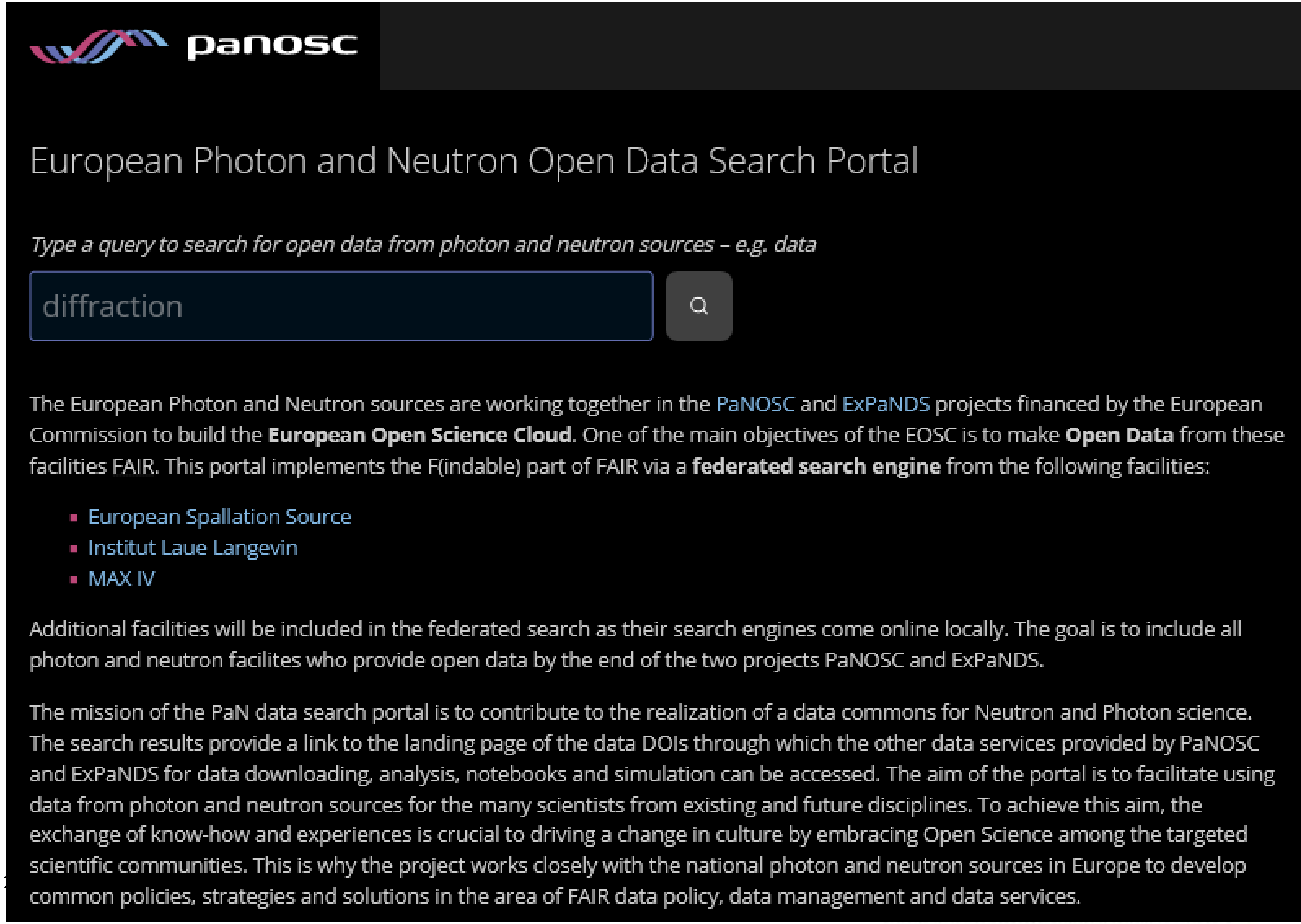
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

OPEN DATA PORTALS FOR OPEN SCIENCE

Next step is Open Data portals for FAIR Data from Photon and Neutron sources:

- Searchable
- Accessible
- Downloadable
- Re-usable

The PaN Open Data Commons will enable new user communities to access and exploit the unique data being produced at the LEAPS facilities to do new science e.g. the Human Organ Atlas is revolutionizing digital histology and medical research with high resolution 3D volumes of complete human organs.



European Photon and Neutron Open Data Search Portal

Type a query to search for open data from photon and neutron sources – e.g. data

diffraction

The European Photon and Neutron sources are working together in the PaNOSC and ExPaNDS projects financed by the European Commission to build the **European Open Science Cloud**. One of the main objectives of the EOSC is to make **Open Data** from these facilities FAIR. This portal implements the F(indable) part of FAIR via a **federated search engine** from the following facilities:

- European Spallation Source
- Institut Laue Langevin
- MAX IV

Additional facilities will be included in the federated search as their search engines come online locally. The goal is to include all photon and neutron facilites who provide open data by the end 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 data DOIs through which the other data services provided by PaNOSC and ExPaNDS for data downloading, analysis, notebooks and simulation can be accessed. The aim of the portal is to facilitate using data from photon and neutron sources for the many scientists from existing and future disciplines. To achieve this aim, the exchange of know-how and experiences is crucial to driving a change in culture by embracing Open Science among the targeted scientific communities. This is why the project works closely with the national photon and neutron sources in Europe to develop common policies, strategies and solutions in the area of FAIR data policy, data management and data services.

This project has received funding from the European Union's Horizon

SUPPORTING HDF5 @ ESRF

Not supported by common applications

- *There are hundreds of formats* out there, starting with CSV …*
- *Makes life difficult for scientists to change formats*
- *Long process which requires discussing with and helping scientists*

HDF5 to other formats

- *Developed tools like **nxtoascii** to produce CSV files (for spectroscopy)*
- *Run file conversion automatically using workflows*

Multiple Readers in any order

- *Supporting multiple readers is the main issue we face today*

Q: Would it be possible to have a file mirrored with one copy for reading only (updated regularly) and the other for writing

*http://fileformats.archiveteam.org/wiki/Scientific_Data_formats

QUESTIONS TO DISCUSS DURING WORKSHOP

1. **Will SWMR2 solve our issue?**
2. **How to address the general case of SWMR?**
3. **How to share file conversion tools from and to HDF5?**
4. **Could H5web be extended to replace HDFView?**
5. **How to include HDF5 in future EOSC projects?**

CONCLUSION

1. **HDF5 has become a first class citizen @ ESRF**
2. **EU projects help build data services and tools for HDF5**
3. **Our main issue is still concurrent access to files being written**
4. **Open data portals will help promote HDF5 further – maybe reach the goal of *one format for all*?**

PIONEERING SYNCHROTRON SCIENCE

WE ❤ DATA

THANKS to

Wout, Thomas, Loic, Axel, Samuel, Alejandro, Laurent, , Jerome, Armando, …

and all scientists who made **HDF5 a reality at ESRF** possible!

This work is supported by PaNOSC, an European project funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 870313, **in the H2020-INFRADEV-2019-2 call**.