

## Multidisciplinary experiments at large research facilities: focus on data – answering the research question with EOSC

Aljoša Hafner  
aljosa.hafner@ceric-eric.eu

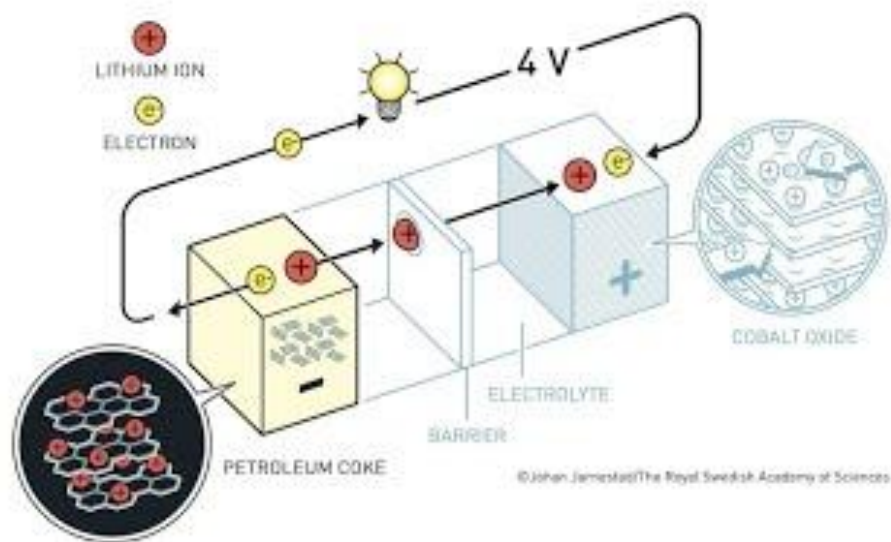
Maribor, 15. 9. 2022

# Table of contents

1. Introduction
2. Current state
3. PaNOSC
4. FAIR data and implementation
5. Data Analysis and examples
6. Catalogues and federated search
7. Assignment
8. Short quiz



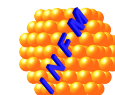
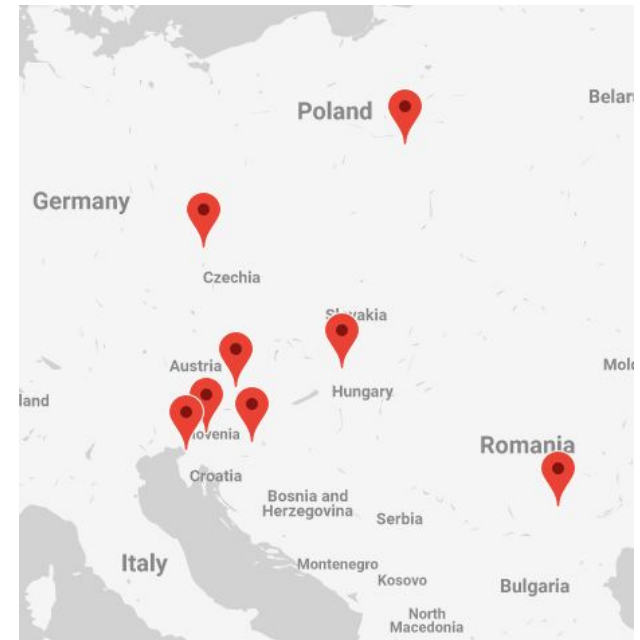
<https://www.youtube.com/watch?v=iVWBx-OhU9w>



<https://www.youtube.com/watch?v=PCVPIV3ZKaE>

# Introduction (1) - CERIC-ERIC

- Central European Research Infrastructure Consortium
- Consortium of 8 institutes from 8 countries
- User facilities
- Unique entry point for proposals and experiments
- Interdisciplinarity



# Introduction (2) - Research infrastructures

## What are Research Infrastructures?

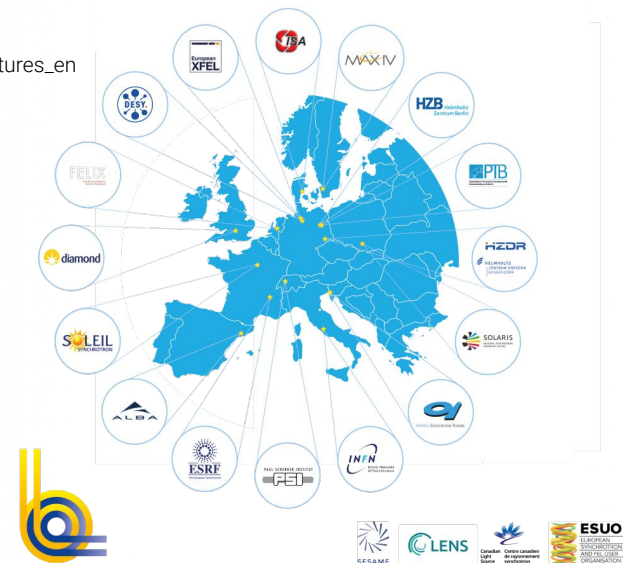
Research Infrastructures are facilities that provide resources and services for research communities to conduct research and foster innovation.

They can be used beyond research e.g. for education or public services and they may be single-sited, distributed, or virtual.

They include

- major scientific equipment or sets of instruments
- collections, archives or scientific data
- computing systems and communication networks
- any other research and innovation infrastructure of a unique nature which is open to external users

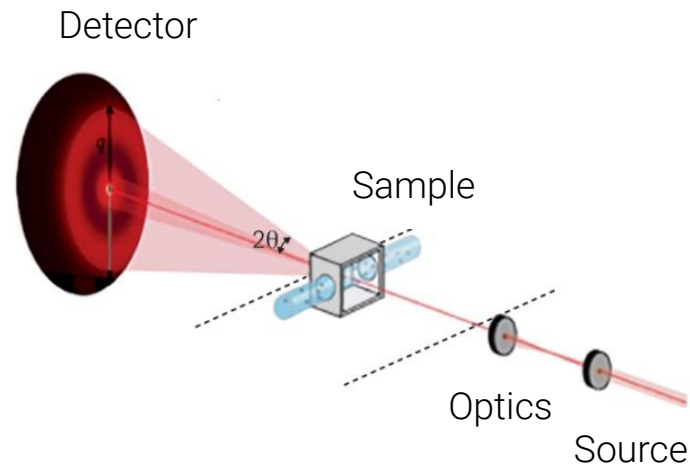
Source: [https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures_en)



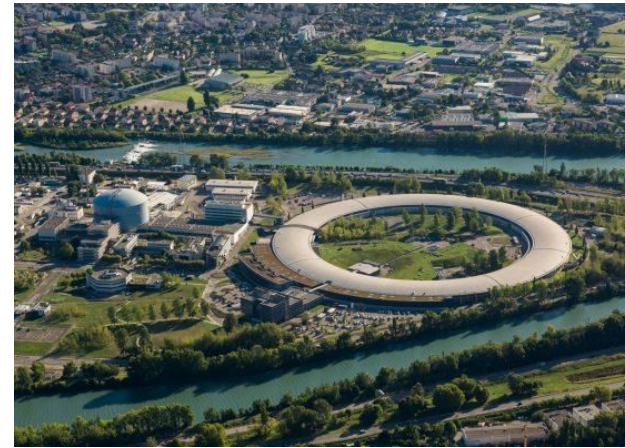


# Introduction (3) - Photons and neutrons?

- Photon (X-ray) sources: synchrotrons, free electron lasers
- Neutron sources: reactors and spallation sources
- Applications (structure and properties of matter - static and time-resolved):
  - physics
  - chemistry
  - biology
  - material research
  - pharmacy
  - archeology
  - art restoration
  - ...

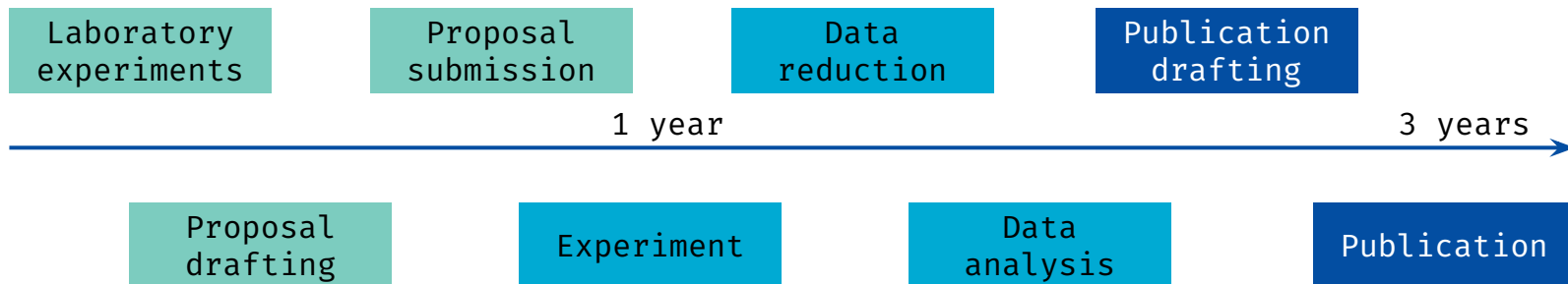


Figures: (top right) ESRF synchrotron and ILL neutron source;  
 (bottom right) Elettra synchrotron and free electron laser FERMI;  
 (top) General experimental schematic.



# Current state (1)

- General experimental flowchart at a *user facility*
- Infrastructure is **over-booked** and **beam time** is limited
- Highly competitive and time-consuming process
- Average time from experiment to publication is 2-3 years

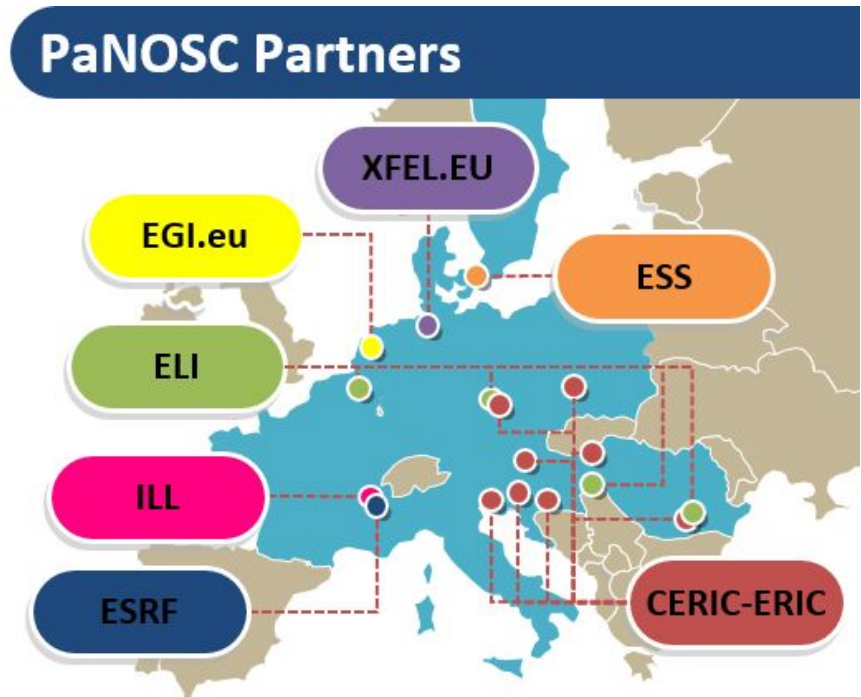
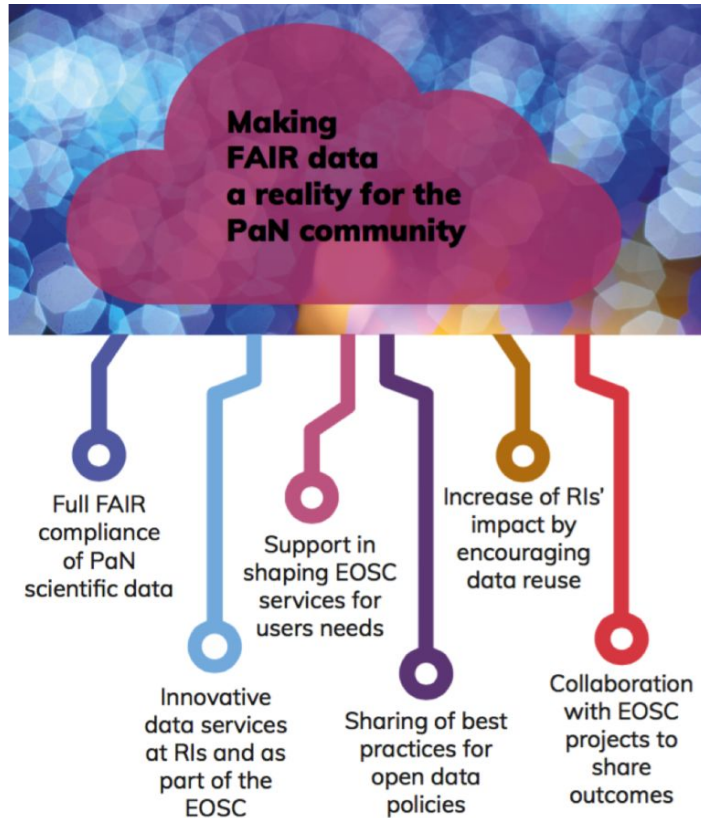




## Current state (2)

- Custom and non-standard acquisition software and tools
- AAI (Authentication and authorization infrastructure) is not unified
- Non-standard data and metadata formats
  - Software tools for reduction, analysis and simulations
- **Long time** from experiment to publication!
- Many experiments **never** produce any publication!

# PaNOSC - Photon and Neutron Open Science Cloud



Source: [www.panosoc.eu](http://www.panosoc.eu)

# PaNOSC (2)

- 6 Technical Workpackages (9 in total):
  - WP2 - Data Policy and Stewardship
  - WP3 - Data Catalog Services
  - WP4 - Data Analysis Services
  - WP5 - Virtual Neutron and X-ray Laboratory
  - WP6 - EOSC integration
  - WP7 - Sustainability
  - WP8 - Staff and User Training

# Timeline



Nexus Data Format

<https://www.nexusformat.org/>

(1994)

Since 2002 - HDF5



<https://www.umbrellaid.org/>

2009 - 2014

Since 2011

2017 - 2021

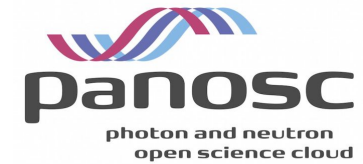
2019 - 2022



<http://pan-data.eu/>



<http://www.calipsoplus.eu/>

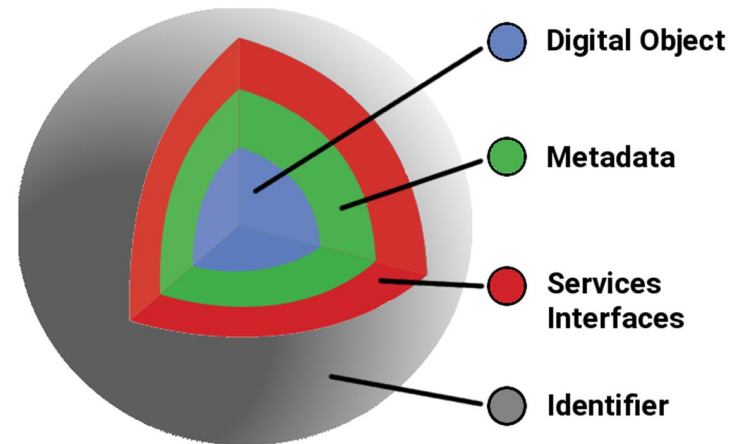
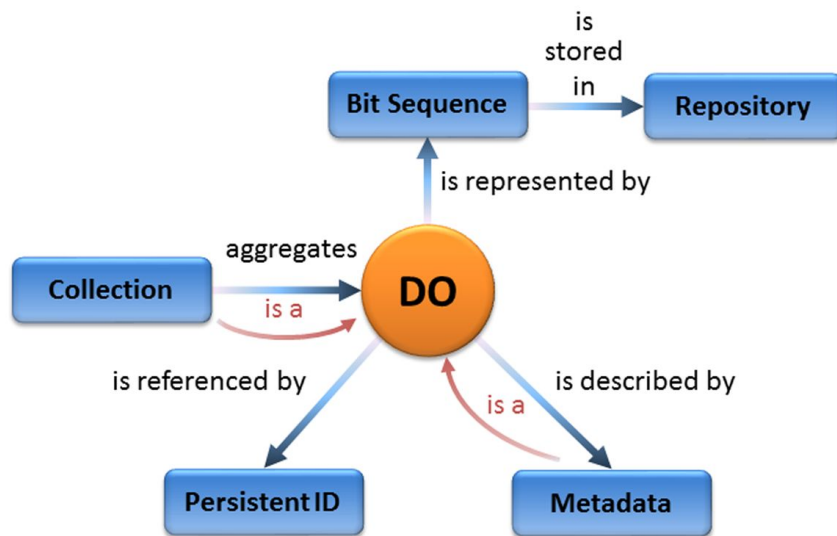


ExPaNDs

<https://expands.eu/>

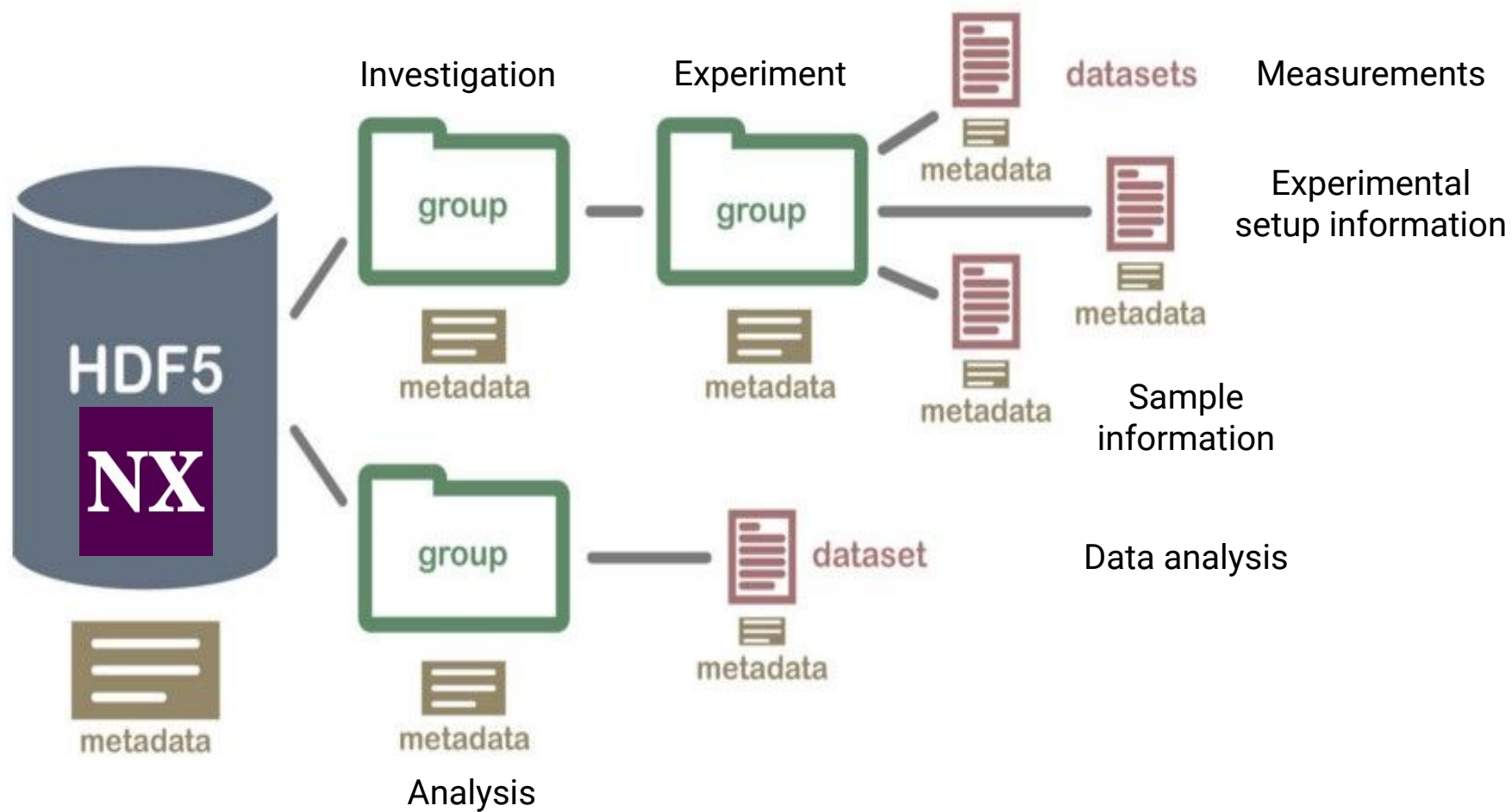
<https://www.panosc.eu>

# FAIR Digital Object



Schwardmann, U. (2020). Digital Objects – FAIR Digital Objects: Which Services Are Required?. *Data Science Journal*, 19(1), 15. DOI: [10.5334/dsj-2020-015](https://doi.org/10.5334/dsj-2020-015)

# Nexus Format



Konnecke, M. et al. (2015). J. Appl. Cryst. 48, 301-305. 10.1107/S1600576714027575

# FAIR Digital Object

## FAIR implementation (**Data management plan** and **Data policy**):

- Findable: search in catalogues, metadata and identifiers (DOI)
  - ICAT
  - Non-structured databases! (Elasticsearch, Lucene, Solr, etc.)
- Accessible: short- and long-term data storage (from a few MB to PB)
  - Online on-premise storage
  - Tape library
  - Cloud storage
- Interoperable: standard data formats (hdf, Nexus, openPMD) and metadata formats (ontology) to use in various analysis software (APIs)
- Reusable: rich and extensive metadata for further (re)use
  - Electronic logbooks



# Data Analysis

The image illustrates a data analysis workflow. On the left, four plots from the Spyder Python IDE are shown:

- Coverage Level Map:** A 2D heatmap showing the spatial distribution of data coverage, with higher values (yellow/red) concentrated in the center-right area.
- Single Pixel Spectrum:** A plot of Counts vs. Energy [eV] showing a noisy spectrum with several peaks, notably around 500, 1000, 1500, and 1750 eV.
- Cumulative Spectrum:** A plot of Counts vs. Energy [eV] showing the cumulative distribution of the data, with a blue line for 'data', a green line for 'bkg', and a red line for 'fit'.
- Volume Percentage Overlap:** A plot of Overlap [%] vs. IO [photons] on a semi-log scale, showing an increasing trend from approximately 76% at 1e3 to 82% at 1e7 photons.

On the right, a terminal window displays system statistics:

```

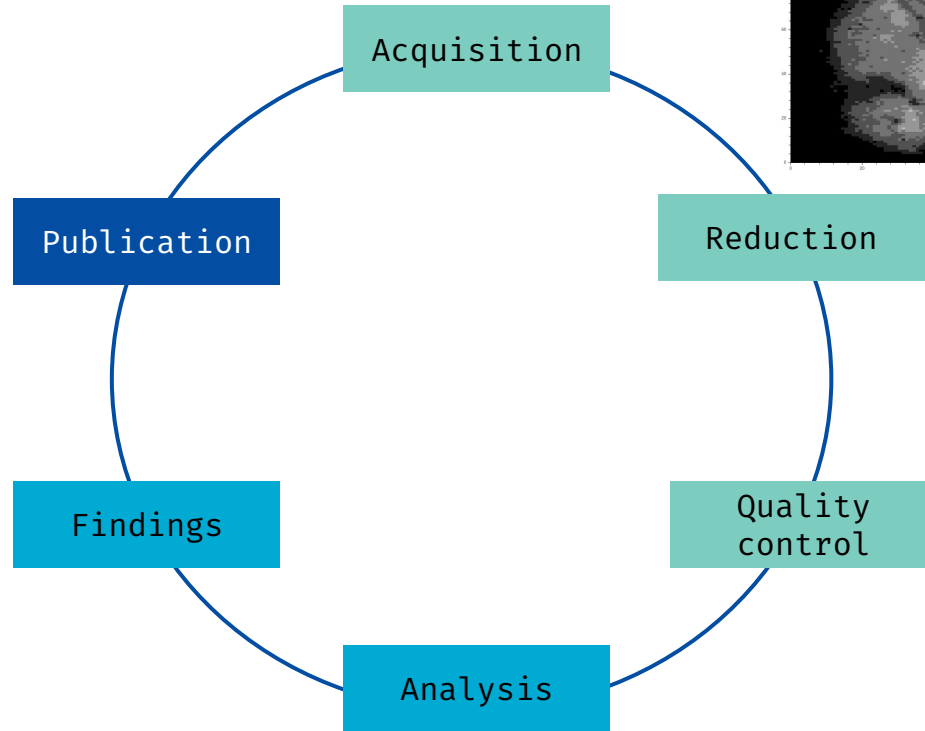
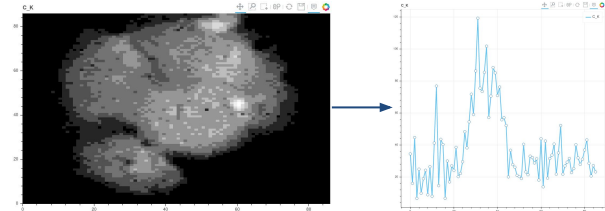
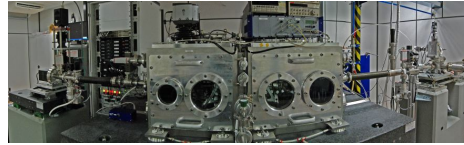
matteo@pc-ippoliti: ~
matteo@pc-ippoliti: ~
0[|||||] 6.7% 4[|||||] 6.9%
1[|||||] 10.6% 5[|||||] 5.0%
2[|||||] 5.7% 6[|||||] 7.5%
3[|||||] 6.9% 7[|||||] 14.8%
Mem [|||||] 25.5G/46.8G Tasks: 198, 2519 thr, 175 kthr; 2 running
Swp [|||||] 0K/2.00G Load average: 0.72 0.57 0.59
Uptime: 11 days, 23:10:29
  
```

Below the terminal, a Google Docs editor displays the title and abstract of the paper "XRF 3D Topography ra...". At the bottom right, the equation for XRF<sub>3D</sub> is shown:

$$XRF_{3D} = \sum_{\alpha=1}^{N_{\alpha}} \frac{EM_{\alpha}}{N_{\alpha}} P_T^{\alpha}$$

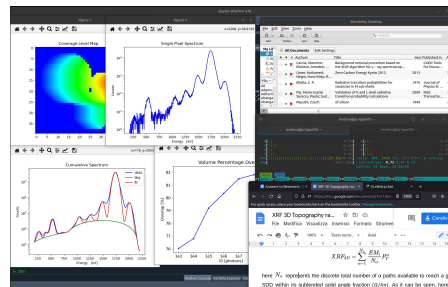
here  $N_{\alpha}$  represents the discrete total number of  $\alpha$  paths available to reach a given SDD within its subtended solid angle fraction ( $\Omega/4\pi$ ). As it can be seen, here the

# Data Analysis



Combined specular and off-specular reflectometry: elucidating the complex structure of soft buried interfaces  
 A. Hafner, P. Gutheundt, S. K. Toppewerg, V. K. R. Jeyakani, S. de Bruijn, A. Wilton, S. K. Mehta, N. Katsikoglou and M. Helsenauer  
 Neutron specular reflectometry (NSR) and off-specular scattering (OSS) are nondestructive techniques which, through illumination, grant a high contrast view among inherently identical species and therefore highly suitable for investigation of soft matter. Through combination of these two techniques, the former enabling a nearly infinite range in the vertical scale by varying surface and/or detector angles, a high-resolution depth-based resolution, and the latter quite detailed information on buried interfaces through scans in the detector plane by using varying detector positions including constant detector distance distribution of the light of scattering, which is extremely time-consuming, the method by the use of a 2D detector, makes it possible to distinguish weakly scattering secondary interfaces in soft matter by the contrast, which occurs in parallel, allowing the method to be applied to the study of buried interfaces across the full range of scattering angles. The 2D maps of intensity versus detector position are processed and analyzed, and the quantitative differentiation of the in-plane structure of long-range order, the combined roughness and bulk defects by a simple inspection of the scattering maps prior to quantitative fit.

Journal of Physics: Condensed Matter  
 PAPER  
 2D reflectometry for the investigation of polymer interfaces: off-specular neutron scattering  
 Alojza Hafner<sup>1,2</sup>, Philipp Gutheundt<sup>1</sup>, Boris P. Toppewerg<sup>1,3</sup>, Mark Georgieva<sup>1,4</sup> and Michelle Sliemazzi<sup>2\*</sup>  
 Published 11 July 2021 • © 2021 IOP Publishing Ltd  
 Journal of Physics: Condensed Matter, Volume 33, Number 28  
 https://doi.org/10.1088/1361-6480/abf800



# PaNOSC tools - CERIC example

- ATSAS ( <https://www.embl-hamburg.de/biosaxs/software.html> )
  - Developed at EMBL
  - Software for small-angle scattering data analysis from biological macromolecules
  - Made available to CERIC users as an online tool (SaaS model) - WebATSAS
- RAFEC ( <https://rafec.elettra.eu> )
  - Developed in-house
  - Allows for running GUI applications in a browser using CERIC AAI

# WebATSAS Software Example

ATSAS Web — Mozilla Firefox

ATSAS Web

https://atsas.ocs.ceric-eric.eu

**CERIC** WEB-ATSAS

submit new job your jobs workers Log out

## Web ATSAS

Web-ATSAS is a web interface of [BIOSAXS' data analysis software ATSAS](#).

At the moment, the following functions are supported:

- [dammin](#)
- [dammif](#)
- [damclust/damaver](#)

Please login with your Elettra credentials.

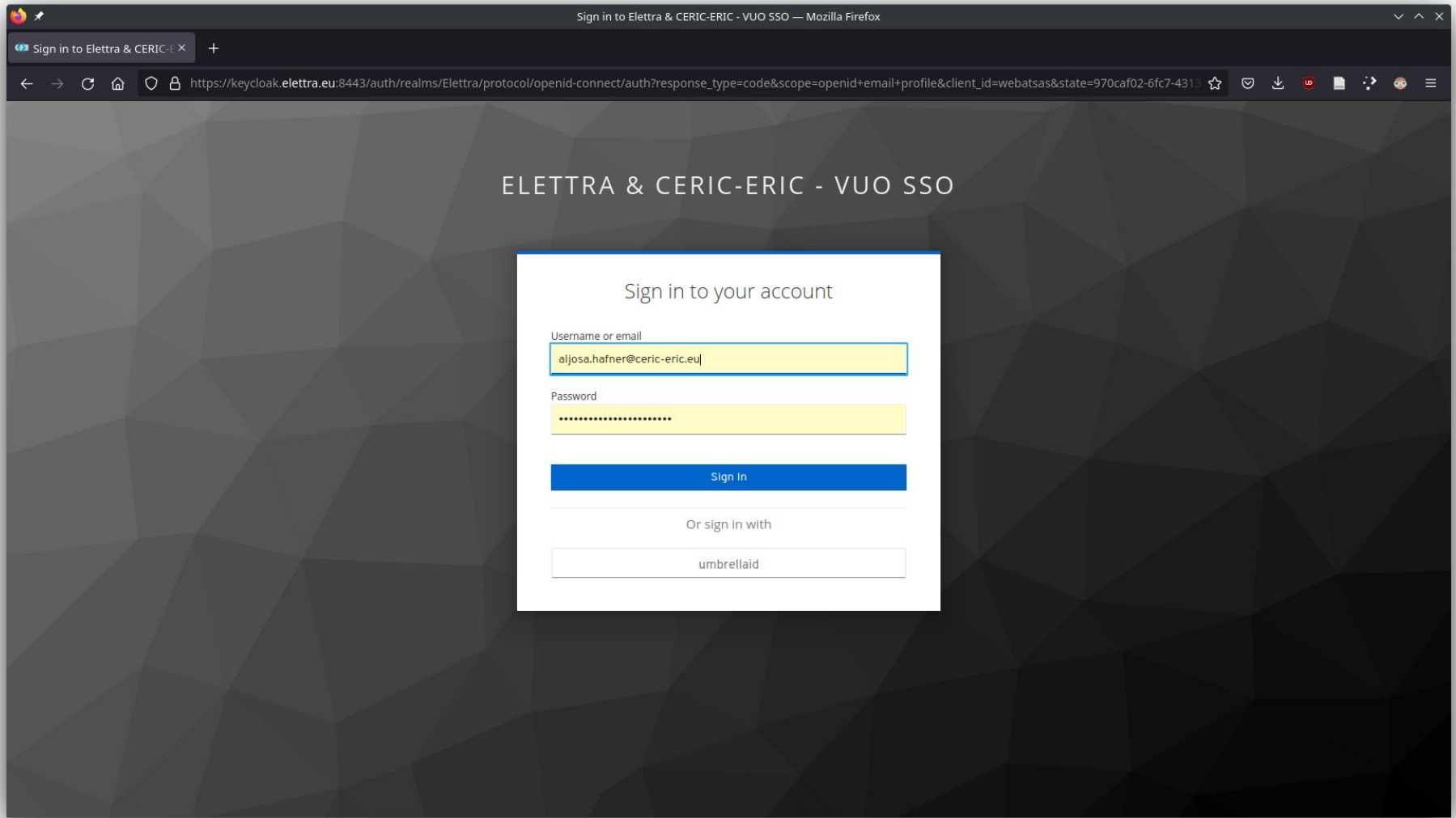
**LOGIN**

© 2022 Copyright CERIC-ERIC

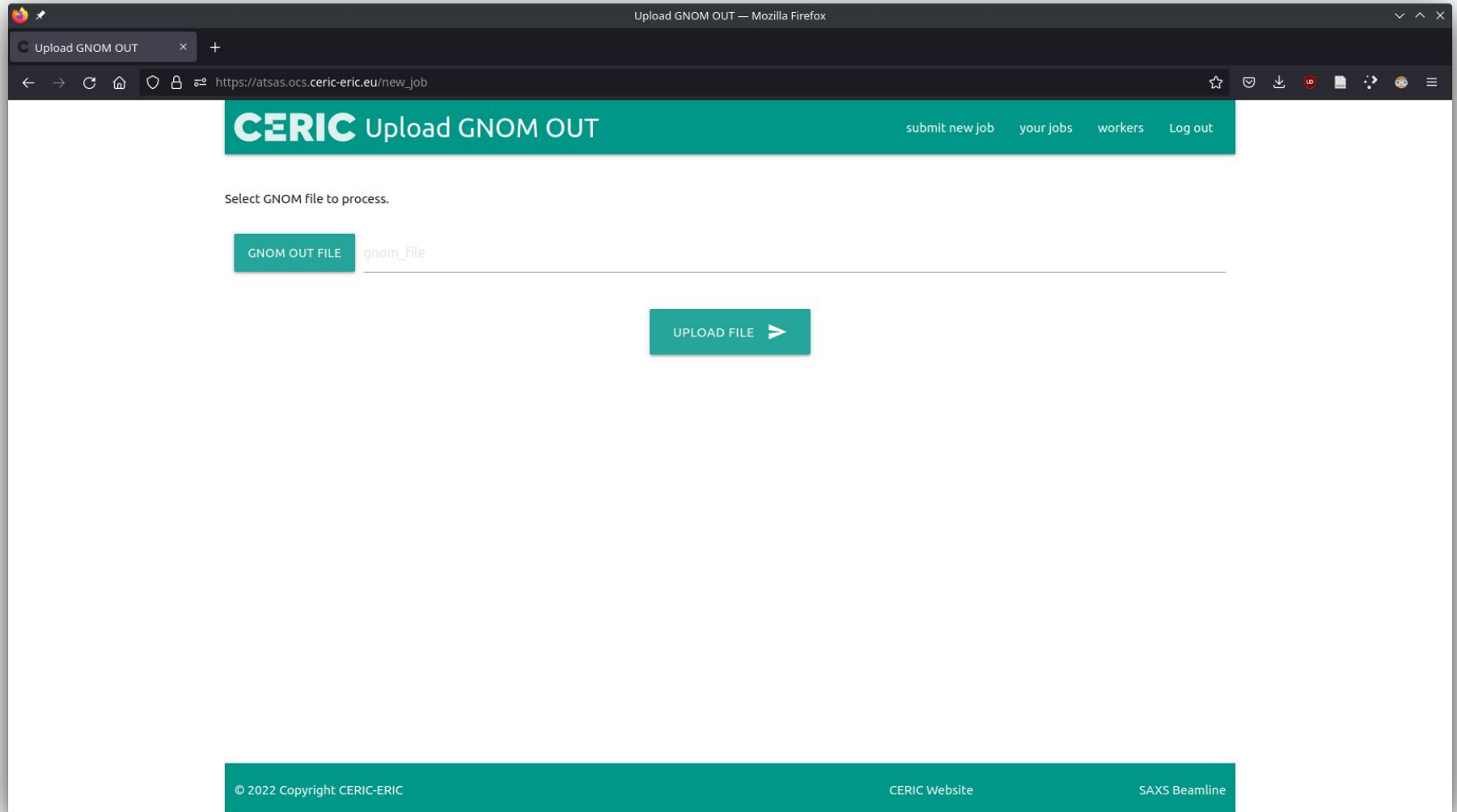
CERIC Website

SAXS Beamline

# WebATSAS Software Example



# WebATSAS Software Example



# WebATSAS Software Example

DAMMIF parameters

submit new job   your jobs   workers   Log out

Angular unit ?	unknown	Dummy atom radius ?	1.0	Min t-successes ?	200
Output file prefix ?	dammif	Maximum number of spherical harmonics ?	20	Temperature schedule factor ?	0.95
Omit output of solvent in PREFIX-0.pdb? ?	yes	Proportion of the curve to be fitted ?	1	R <sub>g</sub> penalty weight ?	0.001
Create pseudo chains in PDB output? ?	no	Number of knots in the curve to fit ?	1	Center penalty weight ?	0.00001
Expected/assumed shape? ?	unknown	Curve weighting function ?	log	Looseness penalty weight ?	0.001
Expected particle symmetry ?	p1	Max t-steps ?	200	Anisotropy penalty weight ?	0.000
Expected particle anisotropy ?	unknown	Max t-iterations ?	200000	Initial random seed ?	200000
Constant to subtract ?	200000	Max t-successes ?	20000		
Simulated annealing setup ?	interactive				
Maximum bead count ?	20000				

SUBMIT JOB >



# WebATSAS Software Example

Atsas Jobs — Mozilla Firefox

Atsas Jobs

https://atsas.ocs.ceric-eric.eu/job/34

n\_of\_knots\_in\_curve = 1

curve\_weighting\_funct = l

max\_temp\_steps\_in\_annealing = 200

max\_iter\_for\_single\_temp\_step = 200000

max\_successes\_per\_temp\_step = 20000

min\_successes\_per\_temp\_step = 200

temp\_schedule\_factor = 0.95

rg\_penalty\_weight = 0.001

center\_penalty\_weight = 0.00001

looseness\_penalty\_weight = 0.001

anisometry\_penalty\_weight = 0.000

initial\_random\_seed =

GNOM file: /data/34/input\_files/gnom\_file/cs\_1\_gnom.out

DOWNLOAD VIEW

Initial DAM file: Initial DAM generated by shape, selected in parameters.

Output files: Run 0

DOWNLOAD OPEN

SEND TO DAMCLUST SEND TO DAMAVER DOWNLOAD ALL FILES

© 2022 Copyright CERIC-ERIC CERIC Website SAXS Beamline

# Rafec Remote Application Framework

The screenshot displays the VUO - Virtual Unified Office web application interface within a Mozilla Firefox browser window. The browser's address bar shows the URL <https://vuo.elettra.eu/pls/vuo/iuu.startup>. The page content is organized into several sections, each with a yellow header and a list of links:

- Users Access Requests to CERIC Laboratories.**
  - [Sample Shipping](#)
- CERIC Users Feedback**
  - [Users Satisfaction Survey](#)
- CERIC Administration Functions**
  - [Personnel Time-off](#)
  - [My timesheets](#)
  - [50% Presences Calendar](#)
- Virtual Laboratory**
  - [My investigations](#)
  - [My tunnels](#)
  - [My applications](#)
  - [My resources](#)
  - [My Rocket.Chat groups](#)
- Security service**
  - [Organigramma aziendale della sicurezza](#)
- FERMI Reports**
  - [Template FPRD](#)
  - [Template FESD](#)
  - [Template FICD](#)
  - [Template FCCD](#)
- Medicina del Lavoro**
  - [Proprie idoneità](#)
- Integrazione con AdHoc Enterprise**
  - [Catalogo di magazzino](#)


The browser's status bar at the bottom shows the URL <https://vuo.elettra.eu/pls/vuo/vlab.applications>.

# Rafec Remote Application Framework

VUO - My applications — Mozilla Firefox

VUO - My applications x +

https://vuo.elettra.eu/pls/vuo/vlab.applications 120% ☆

 | **VUO** - Virtual Unified Office

**VUO - My applications**

Logged as: **Aljosa HAFNER**  - [\[Logout\]](#)

[Home](#) | [My investigations](#) | [My tunnels](#) | [My applications](#) | [My resources](#) | [My Rocket.Chat groups](#)

Opened applications	
Application	Date
<a href="#">[Open]</a> SCITOOLBOX (Pymca, Oasys, Spyder & Fiji)	13/09/2022


Available applications	
Application	Tag
<a href="#">[Start]</a> RAFEC-JUPTYER-PANOSC	PANOSC
<a href="#">[Start]</a> SCITOOLBOX (Pymca, Oasys, Spyder & Fiji)	PANOSC

---

**Elettra - Sincrotrone Trieste S.C.p.A.**

S.S. 14 Km 163,5 in Area Science Park 34149 Basovizza, Trieste, Italy T. +39 040 37581 F. +39 040 938 0903	P.IVA e C.F. IT00697920320 Cap. Soc. € 47.632.663,00 i.v. PEC: sincrotrone.trieste.elettra@legalmail.it <a href="http://www.elettra.eu">www.elettra.eu</a>	Iscritta al Registro delle Imprese di Trieste Società di interesse nazionale ai sensi dell'art. 10, comma 4, L. 19 ottobre 1999 n. 370
---	---	---

[VUO contacts](#) | [Disclaimer](#) | [Privacy policy](#)

  
CERTIFIED MANAGEMENT SYSTEM  
UNI EN ISO 9001:2015  
UNI ISO 45001:2018  
UNI CEI EN ISO 50001:2018

https://vuo.elettra.eu/pls/vuo/vlab.show\_view\_opened\_application?FRM\_APPL\_ID=9828

# Rafec Remote Application Framework

The screenshot shows a web browser window with the URL <https://rafec.elettra.eu/HAcVOQSVBaAEDaQEzTOgaKFMYPcpiTB/?password=DOqVkwYq&keyboard=off>. The browser title is "HAcVOQSVBaAEDaQEzTOgaKFMYPcpiTB: Hercules 2022 WOFRY (coherence transport) — Mozilla Firefox".

The application interface includes a menu on the left with the following items:

- Shadow Advanced Light ...
- Shadow Optical Elements
- Shadow Compound Opti...
- Shadow Advanced Optic...
- Shadow Special Elements
- Shadow PostProcessor
- Shadow PreProcessor
- Shadow Experiments
- Shadow Basic Loops
- Shadow Scanning Loops
- Shadow Thermal Load
- Shadow Utility
- Shadow Elettra Extension
- Shadow PaNOSC Extensi...
- Shadow ESRF Extension
- SRW Light Sources

The main content area displays two optical models:

**MODEL 1) Plane wave diffracted by a slit**

The diagram shows a "Generic Wavefront 2D (plane wave)" on the left, connected by a "WofryData" link to a "Slit 40x40 um". From the slit, two "WofryData" links lead to "Screen 1m" and "Screen 15m".

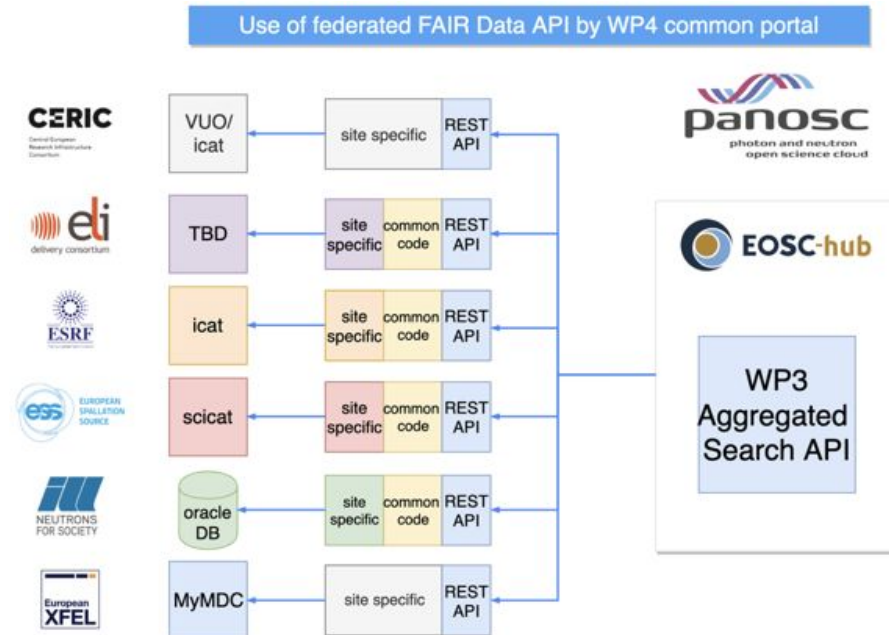
**MODEL 2.1) Converging wavefront (collapsing spherical wave)**

This model is shown in two parts. The top part shows a "Generic Wavefront 1D (Convergent Wavefront)" connected by a "WofryData" link to a "Slit 1D", which is then connected by another "WofryData" link to a "Screen 1D". To the right, a separate diagram shows "Undersampled (replicas)" with a "Generic Wavefront 1D (Convergent Wavefront)" and a "Slit 1D" connected by a "WofryData" link.

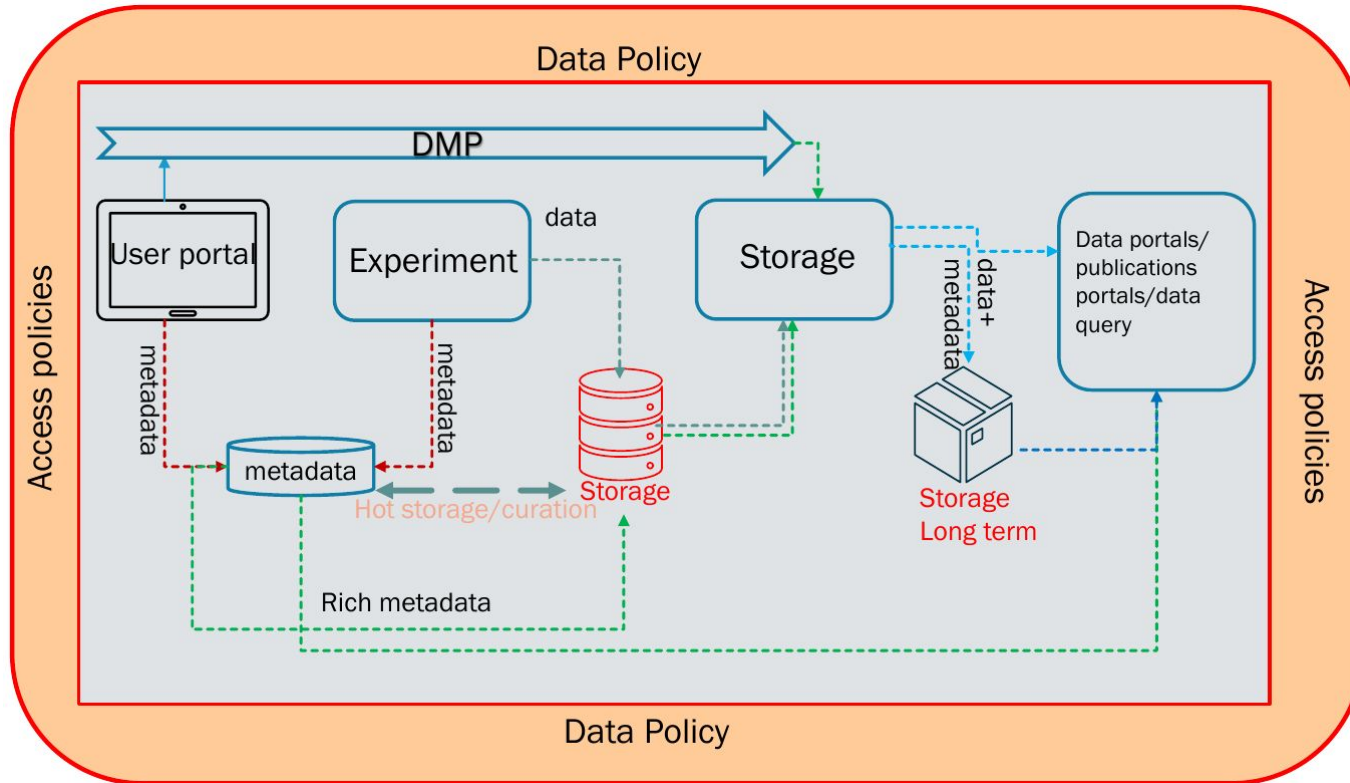
The bottom part of Model 2.1 shows a "Generic Wavefront 2D (Convergent Wavefront)" connected by a "WofryData" link to a "Slit 400x400 um", which is then connected by another "WofryData" link to a "Screen".

# Catalogues

- Metadata is recorded and searchable
- Access is governed by the **data policy**
- Data is stored according to the **data policy** and **data management plan**
- Procedures are governed by data stewardship processes



# How it comes together



Source: ELI-ERIC;

<https://www.fairsfair.eu/sites/default/files/ELI%20FAIR%20Data%20Policy.%20Guidelines%20and%20implementation%20challenges.pdf>

# Federated Search - data.panosc.eu

The screenshot displays the Open Data Portal - PaNOSC search interface. The search bar contains the term "covid", and the results show 40 documents found. The interface includes a sidebar with filters for Type (proposal, publication), Technique, Incident Wavelength, Incident Photon Energy, Chemical Formula, Temperature, and Pressure. The search results are displayed in a list format, showing document IDs, titles, and release dates.

Open Data Portal - PaNOSC — Mozilla Firefox

Open Data Portal - PaNOSC x +

https://data.panosc.eu/search/?q=covid

panosc covid 40 documents found

Type

- proposal
- publication

Technique

Select a technique...

Incident Wavelength

min max nm

Incident Photon Energy

min max eV

Chemical Formula

Temperature

min max K

Pressure

min max Pa

10.15151/ESRF-ES-436648953 0.900

COVID-19 affects multiple vital organs and leads to drastic changes in tissue architecture not only ...

Released on May 7th 2024 by the ESRF

10.34967/I8163 0.900

20202213

This is an addition to Fast-track Proposal #20202210 (SISSI/AFM), please process the proposals combined. COVID-19 related project in conjunction with structural lung architecture investigations at SYRMEP Beamline (EuBi beam time...

Released on January 5th 2022 by the CERIC

10.15151/ESRF-ES-193743898 0.587

COVID-19: structural studies of membrane protein E of SARS-CoV-2

Released on May 29th 2023 by the ESRF

10.15151/ESRF-ES-189558792 0.4 {?}

Multiscale Quantification of Covid-19's impact on lung vasculature from whole lobe to alveolar/mi...



# Assignment - Using the federated search

The screenshot shows a web browser window displaying the Open Data Portal - PaNOSC search results for the query 'covid'. The browser address bar shows the URL <https://data.panosc.eu/search/?q=covid>. The search results page features a sidebar on the left with various filters and a main content area on the right displaying search results.

**Search Results:**

- 10.15151/ESRF-ES-436648953** (Relevance: 0.900)  
COVID-19 affects multiple vital organs and leads to drastic changes in tissue architecture not only ...  
*Released on May 7th 2024 by the ESRF*
- 10.34967/I8163** (Relevance: 0.900)  
**20202213**  
This is an addition to Fast-track Proposal #20202210 (SISSI/AFM), please process the proposals combined. COVID-19 related project in conjunction with structural lung architecture investigations at SYRMEP Beamline (EuBi beam time...  
*Released on January 5th 2022 by the CERIC*
- 10.15151/ESRF-ES-193743898** (Relevance: 0.587)  
COVID-19: structural studies of membrane protein E of SARS-CoV-2  
*Released on May 29th 2023 by the ESRF*
- 10.15151/ESRF-ES-189558792** (Relevance: 0.4 {?})  
Multiscale Quantification of Covid-19's impact on lung vasculature from whole lobe to alveolar/mi...

**Filters (Left Sidebar):**

- Type:** proposal, publication
- Technique:** Select a technique...
- Incident Wavelength:** min, max, nm
- Incident Photon Energy:** min, max, eV
- Chemical Formula:**
- Temperature:** min, max, K
- Pressure:** min, max, Pa

# Assignment (1)

Explore the search portal:

- <https://data.panosc.eu>
- Create an account
  - or VUO - CERIC virtual unified office: <https://vuo.elettra.eu>
  - or UmbrellaID: <https://umbrellaid.org/euu/account/create>
- Which organizations have integrated their data into the search?
- Find a *BrightnESS* project entry
  - Where was it recorded?
  - Who is the author?
  - Which persistent identifier is used?
  - How many datasets are in the entry and what is their size?

# Assignment (2)

*manuscript\_dataset\_sample\_prep\_XRF2022* investigation:

- Which organization contributed the entry?
- Who are the authors? What is their ORCID, do they have it?
- Which persistent identifiers are used?
- What is the entry about?
- How many experiments were performed?
- Find the publication associated with the datasets

# Assignment (3)

Explore *manuscript\_dataset\_sample\_prep\_XRF2022* investigation using online analysis:

- Access the data through h5nuvola browser
- What is the dataset structure? Were there more techniques performed?
- Where are the images?
- Investigate the images (find the .h5 files and click on “image”)
  - What is the size in pixels?
  - How many hdf5 files are there?
  - Are they RAW data or reduced?

# Short quiz

Kahoot quiz

Participate through **Kahoot app** or **kahoot.it + PIN**

# Conclusion

- Research Infrastructures (RIs) offer user access based on merit
- Beamtime is limited and overbooked
- Data is managed according to data management plan and data policy
- Transparent data analysis and data reduction is difficult
- EOSC will enable easier access to
  - already acquired data
  - find existing data and authors
  - submit proposals and perform beamtime
  - computing resources

Aljoša Hafner  
aljosa.hafner@ceric-eric.eu

- Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- Schwardmann, U., 2020. Digital Objects – FAIR Digital Objects: Which Services Are Required?. *Data Science Journal*, 19(1), p.15. DOI: <http://doi.org/10.5334/dsj-2020-015>
- Konnecke, M. *et al.* (2015). *J. Appl. Cryst.* 48, 301-305. 10.1107/S1600576714027575
- Jones, Sarah. (2018, March 12). Open data, FAIR data and RDM: the ugly duckling. International Open Science Conference (OSC), Berlin. Zenodo. <https://doi.org/10.5281/zenodo.1196631>
- [www.panosc.eu](http://www.panosc.eu) ; [www.expands.eu](http://www.expands.eu) ; [www.eosc.eu](http://www.eosc.eu)
- What are Research Infrastructures?  
[https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures_en)
- <https://www.panosc.eu/all-use-cases/>

Aljoša Hafner  
[aljosa.hafner@ceric-eric.eu](mailto:aljosa.hafner@ceric-eric.eu)