CERIC

Central European
Research Infrastructure
Consortium

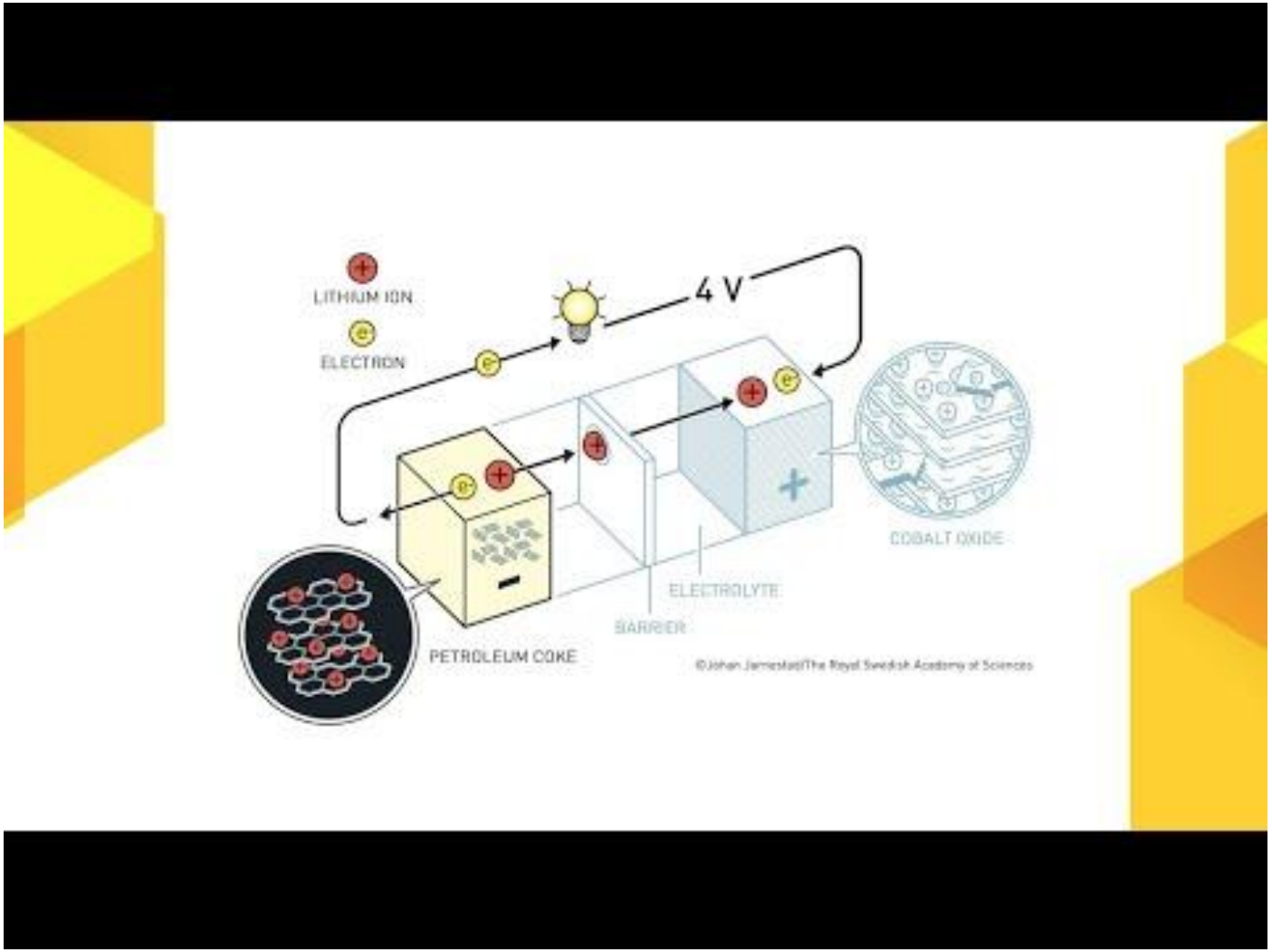
Multidisciplinary experiments at large research facilities: focus on data – answering the research question with EOSC

Aljoša Hafner
aljosa.hafner@ceric-eric.eu

Maribor, 15. 9. 2022

Table of contents

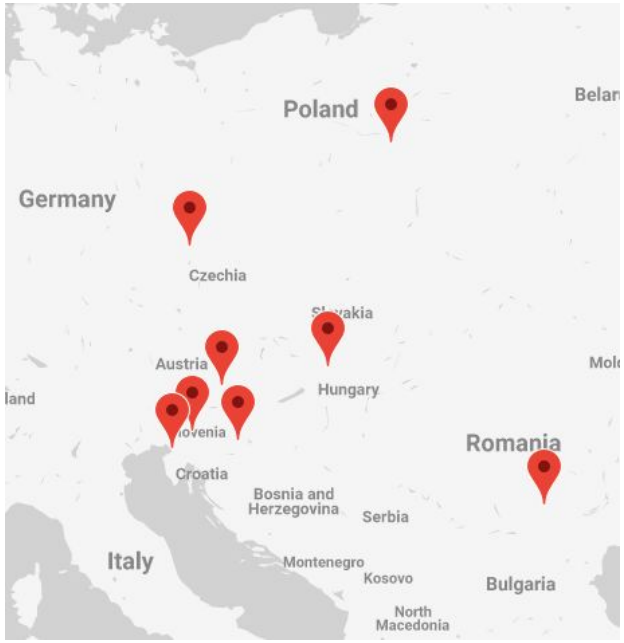
https://www.youtube.com/watch?v=iVWBx-OhU9w



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.

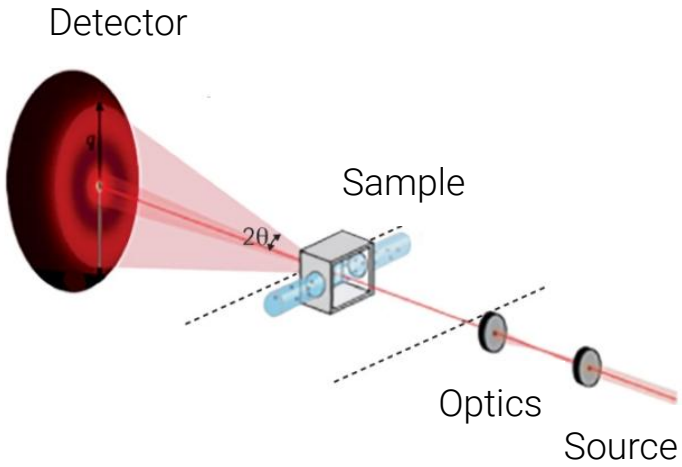
They include

- 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

Introduction (3) - Photons and neutrons?

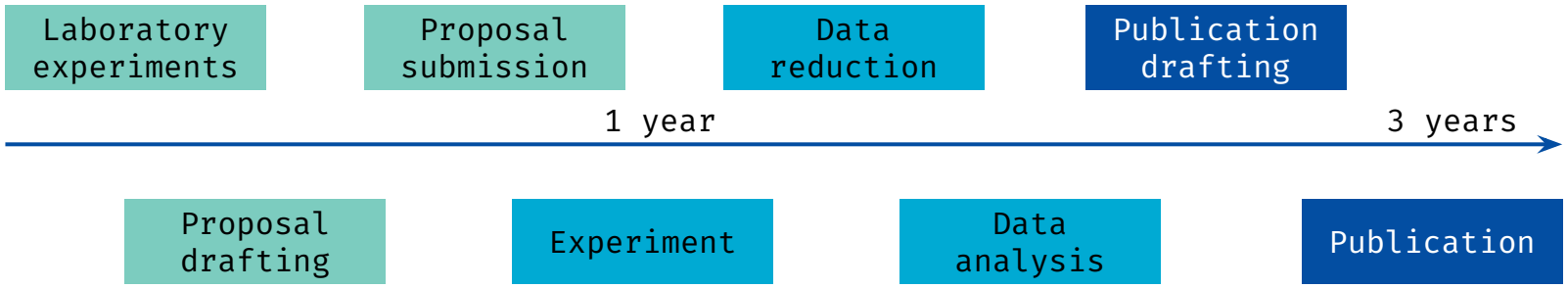
- Photon (X-ray) sources: synchrotrons, free electron lasers
- Neutron sources: reactors and spallation sources
- Applications (structure and properties of matter - static and time-resolved):
  - physics
  - chemistry
  - biology
  - material research
  - pharmacy
  - archeology
  - art restoration
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



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)

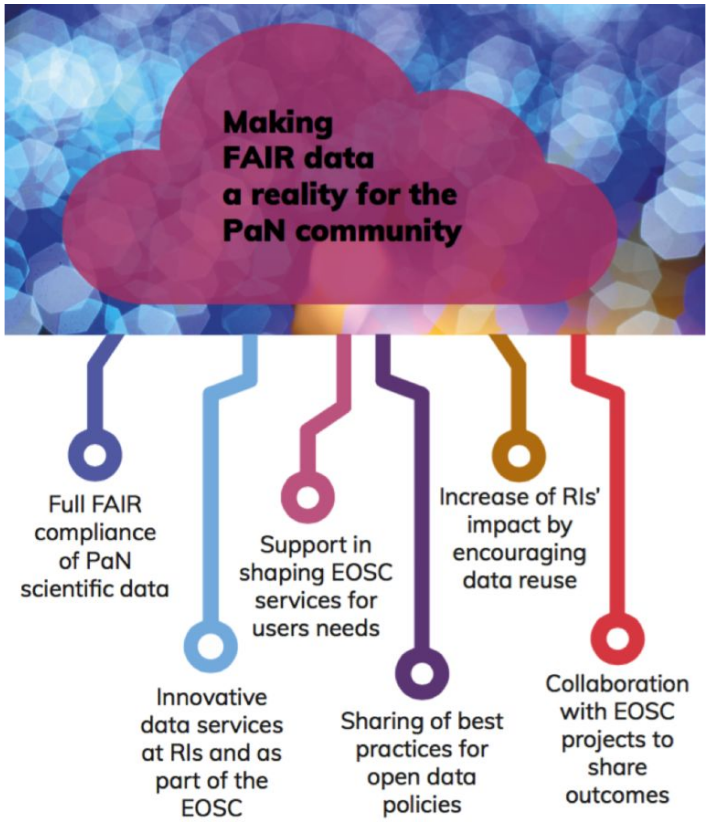
- General experimental flowchart at a *user facility*
- Infrastructure is **over-booked** and **beam time** is limited
- Highly competitive and time-consuming process
- Average time from experiment to publication is 2-3 years

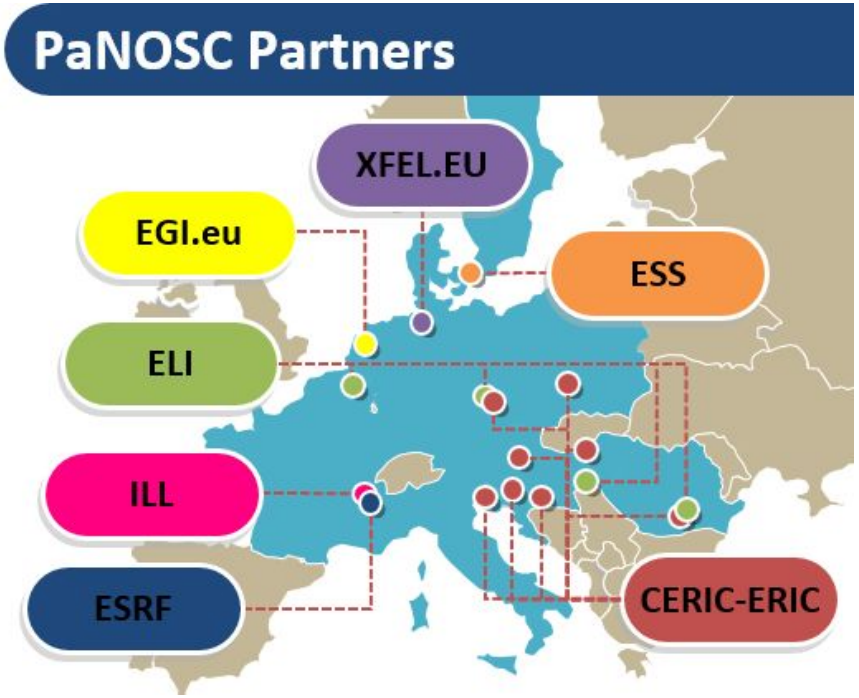


Current state (2)

- 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.panosc.eu

PaNOSC (2)

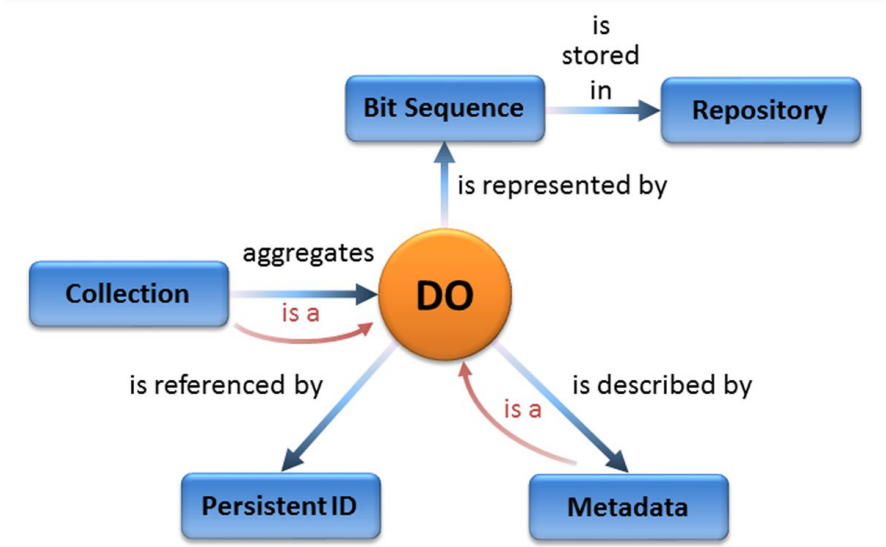
- 6 Technical Workpackages (9 in total):
  - WP2 - Data Policy and Stewardship
  - WP3 - Data Catalog Services
  - WP4 - Data Analysis Services
  - WP5 - Virtual Neutron and X-ray Laboratory
  - WP6 - EOSC integration
  - WP7 - Sustainability
  - WP8 - Staff and User Training

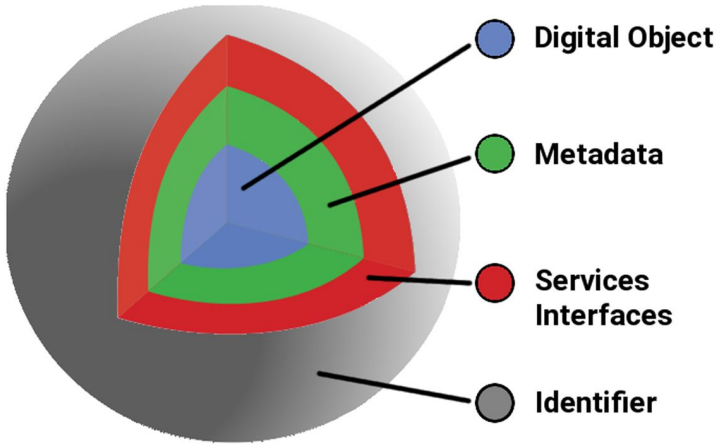
Source: www.panosc.eu

Timeline


NX
Nexus Data Format
https://www.nexusformat.org/
umbrella
https://www.umbrellaid.org/
panosc
photon and neutron
open science cloud
ExPaNDS
https://expands.eu/
https://www.panosc.eu
(1994)
Since 2002 - HDF5
2009 - 2014
Since 2011
2017 - 2021
2019 - 2022
pandata europe
http://pan-data.eu/
CALIPSOplus
http://www.calipsoplus.eu/

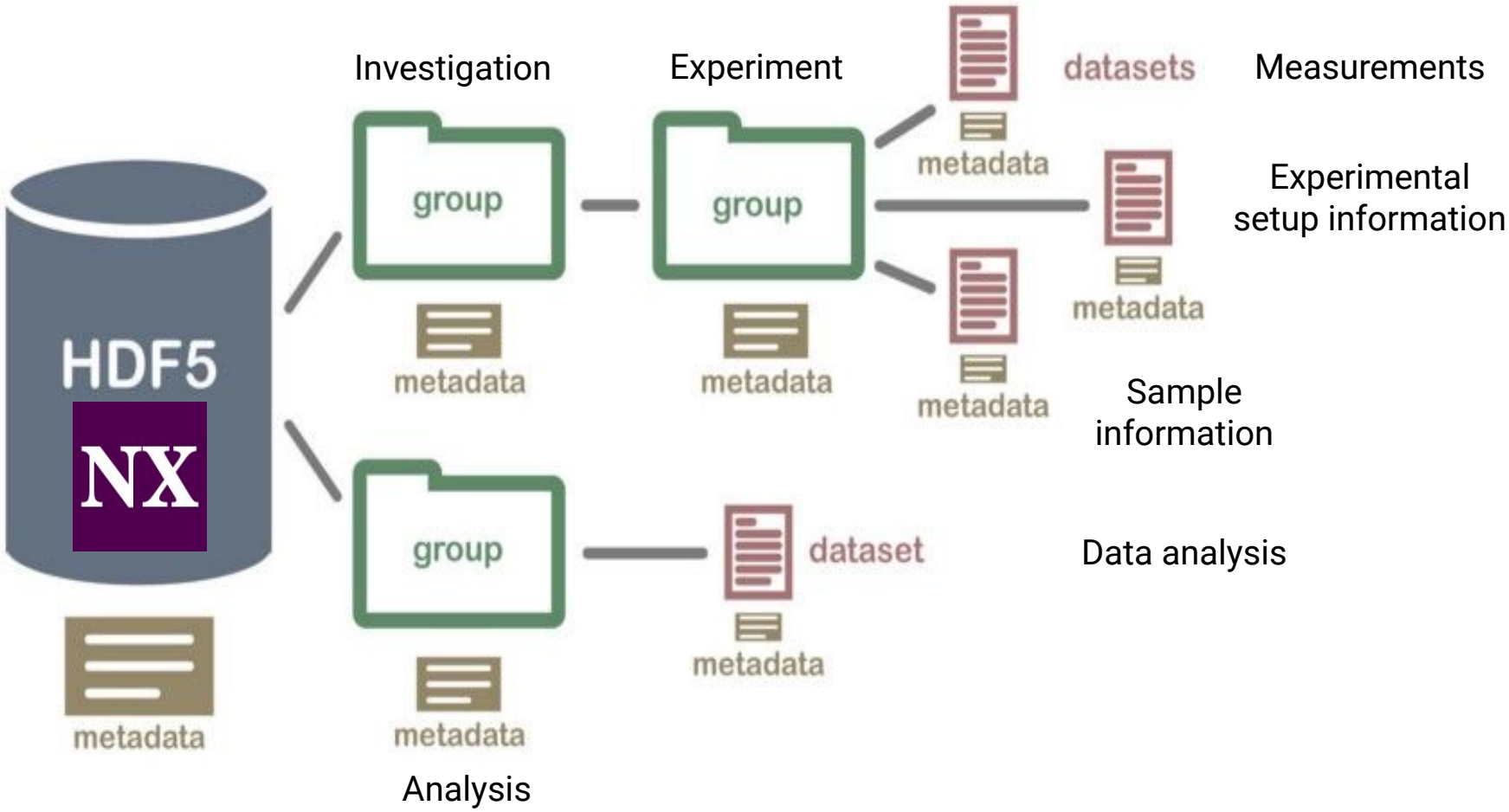
FAIR Digital Object





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

Nexus Format



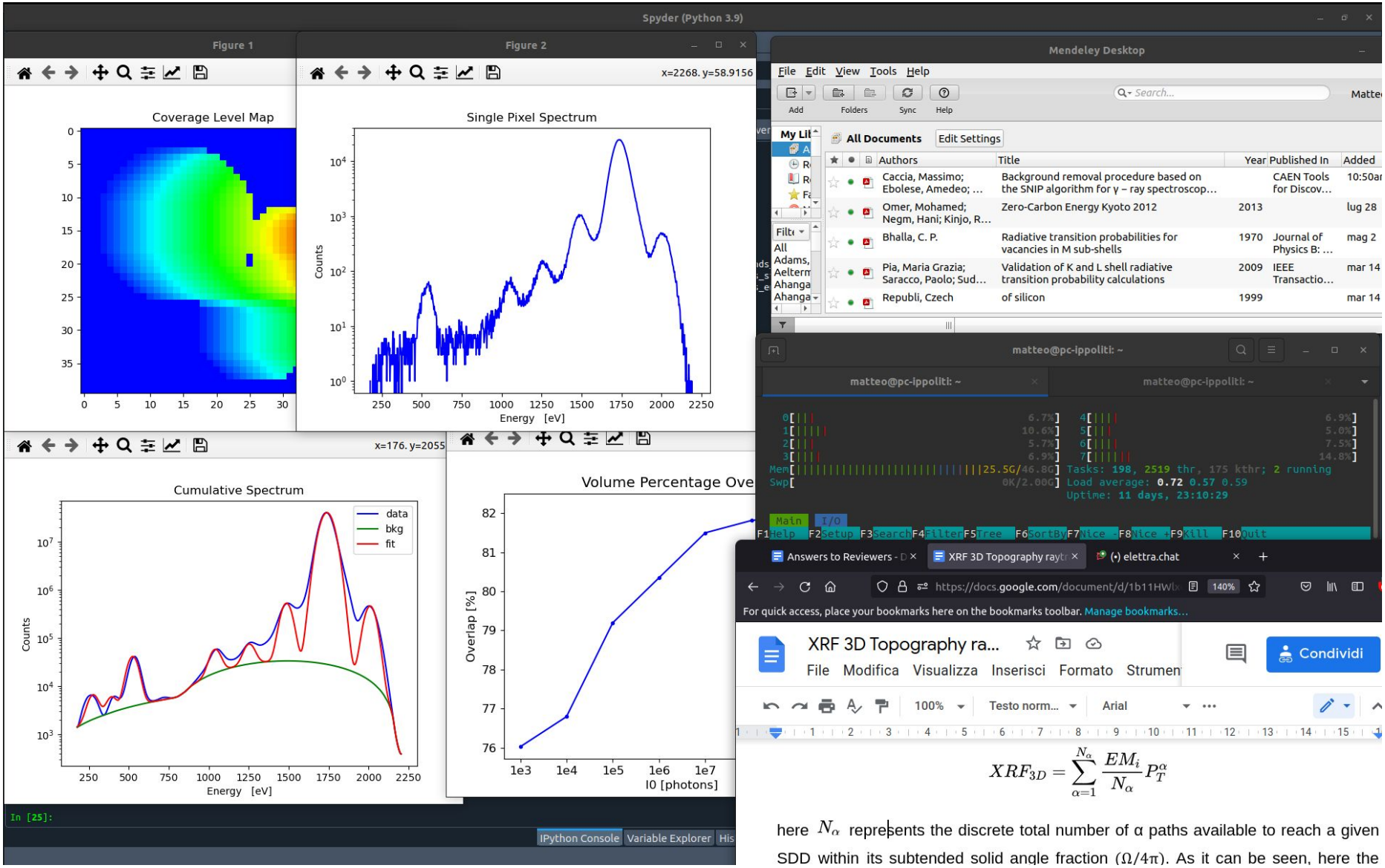
Konnecke, M. et al. (2015). J. Appl. Cryst. 48, 301-305. 10.1107/S1600576714027575

FAIR Digital Object

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

- Findable: search in catalogues, metadata and identificators (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

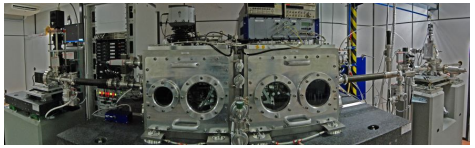
Data Analysis

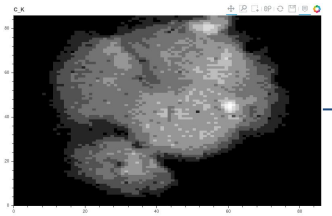


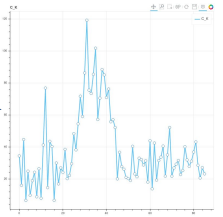
$$XRF_{3D} = \sum_{\alpha=1}^{N_\alpha} \frac{EM_i}{N_\alpha} P_T^\alpha$$

here N_α represents the discrete total number of α 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







Journal of Physics: Condensed Matter

PAPER

2D reflectometry for the investigation of polymer interfaces: off-specular neutron scattering

Aljoša Hafner[1,2], Philipp Gutfreund[3], Boris P Toperverg[3], Mark Geoghegan[4] and Michele Sferrazza[4]

Published 13 July 2021 • © 2021 IOP Publishing Ltd

Journal of Physics: Condensed Matter, Volume 33, Number 36

Intense Radiation Sources in Condensed Matter and Materials Physics

Citation Aljoša Hafner et al 2021 J. Phys.: Condens. Matter 33 364002

Acquisition

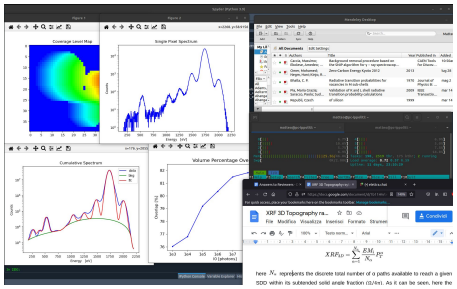
Reduction

Quality control

Analysis

Findings

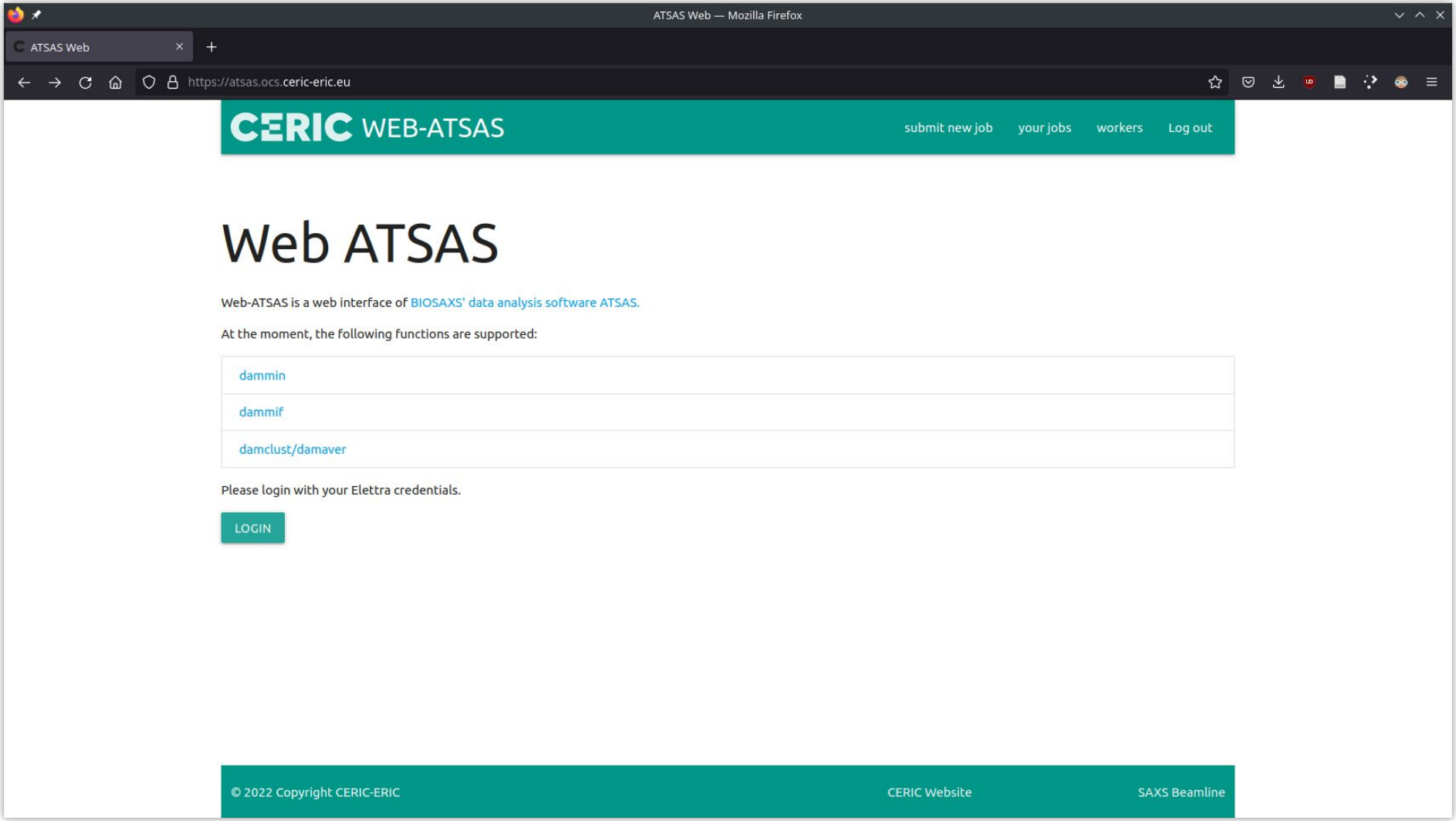
Publication



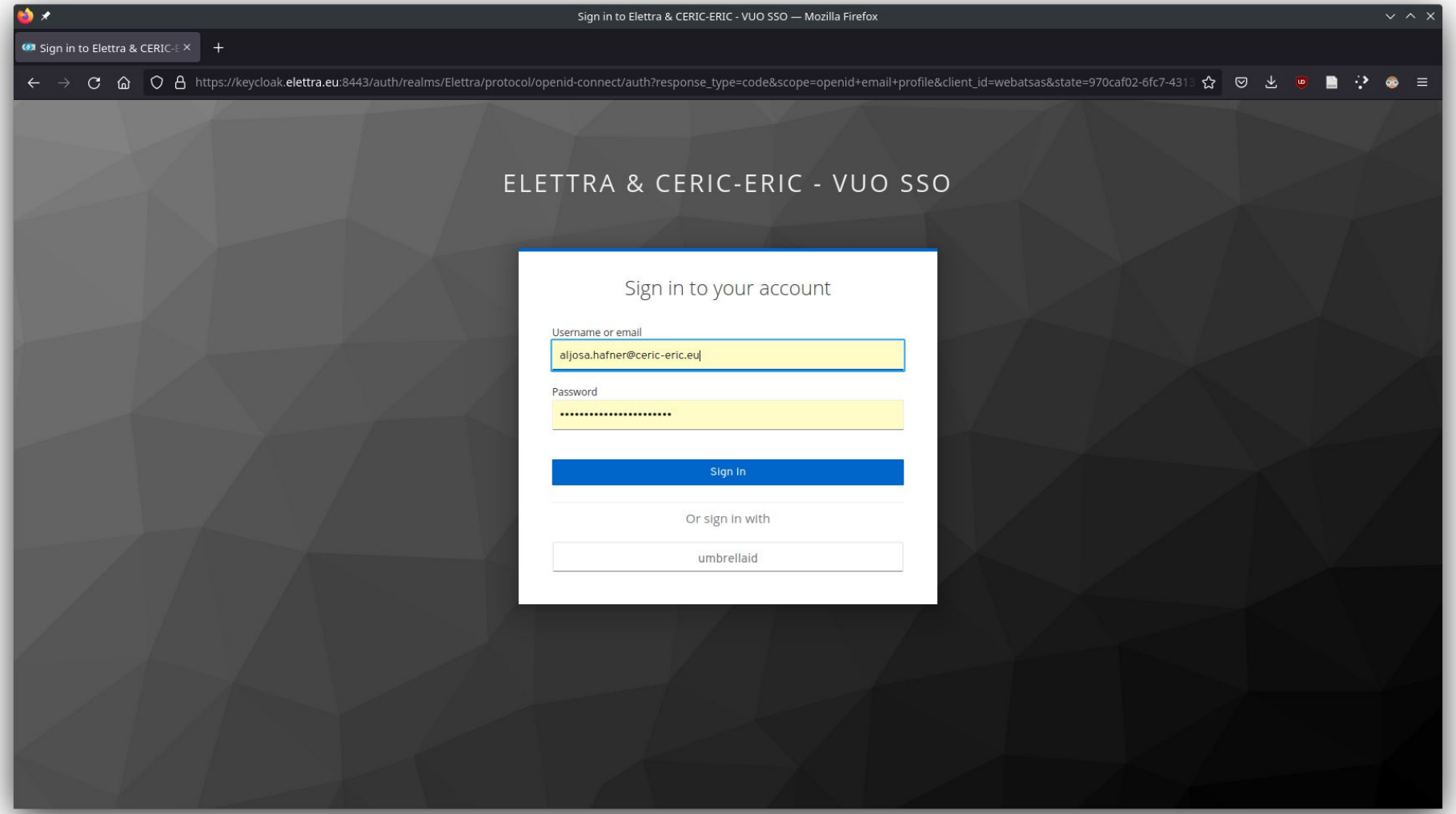
PaNOSC tools - CERIC example

- 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

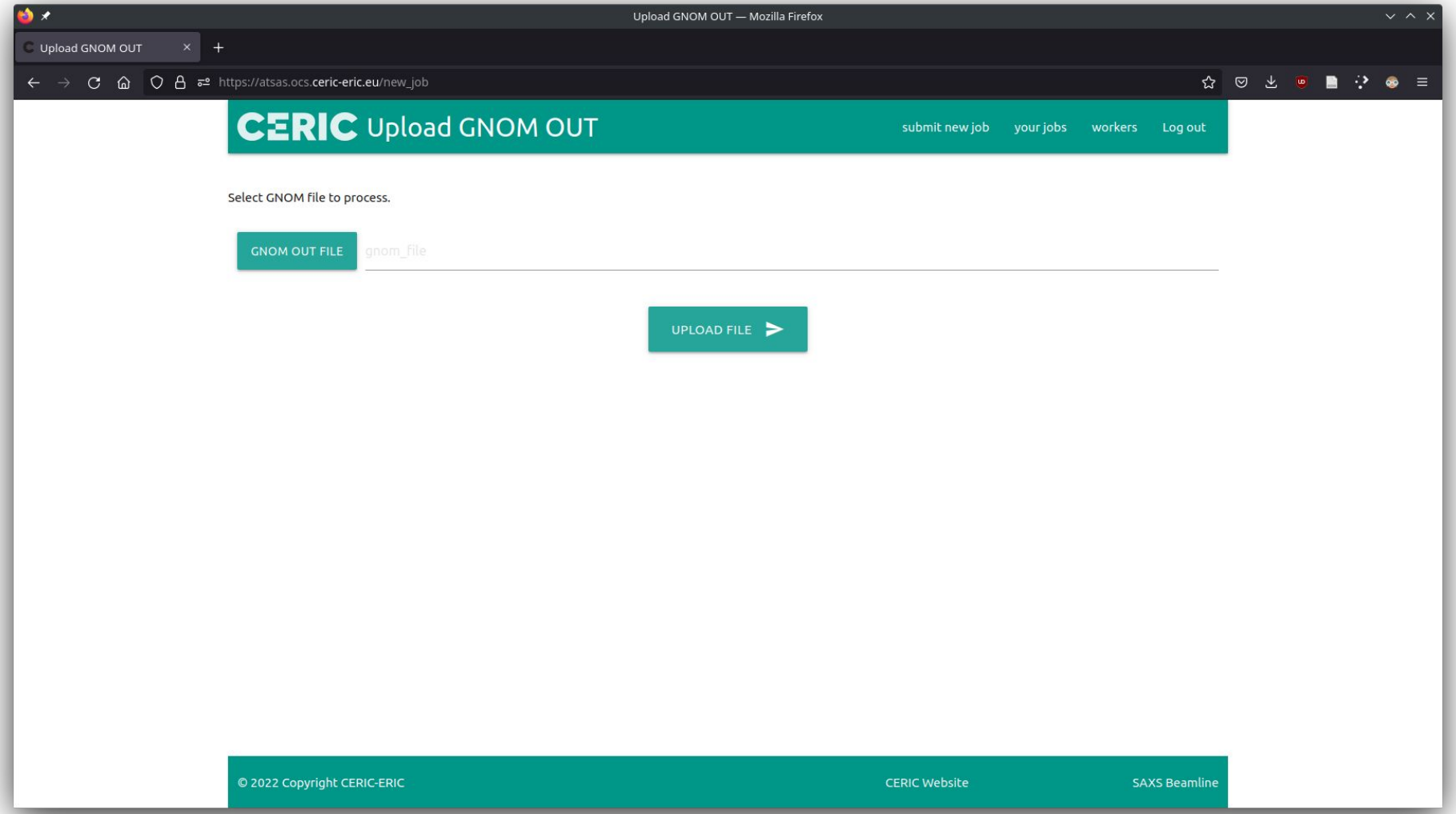
WebATSAS Software Example



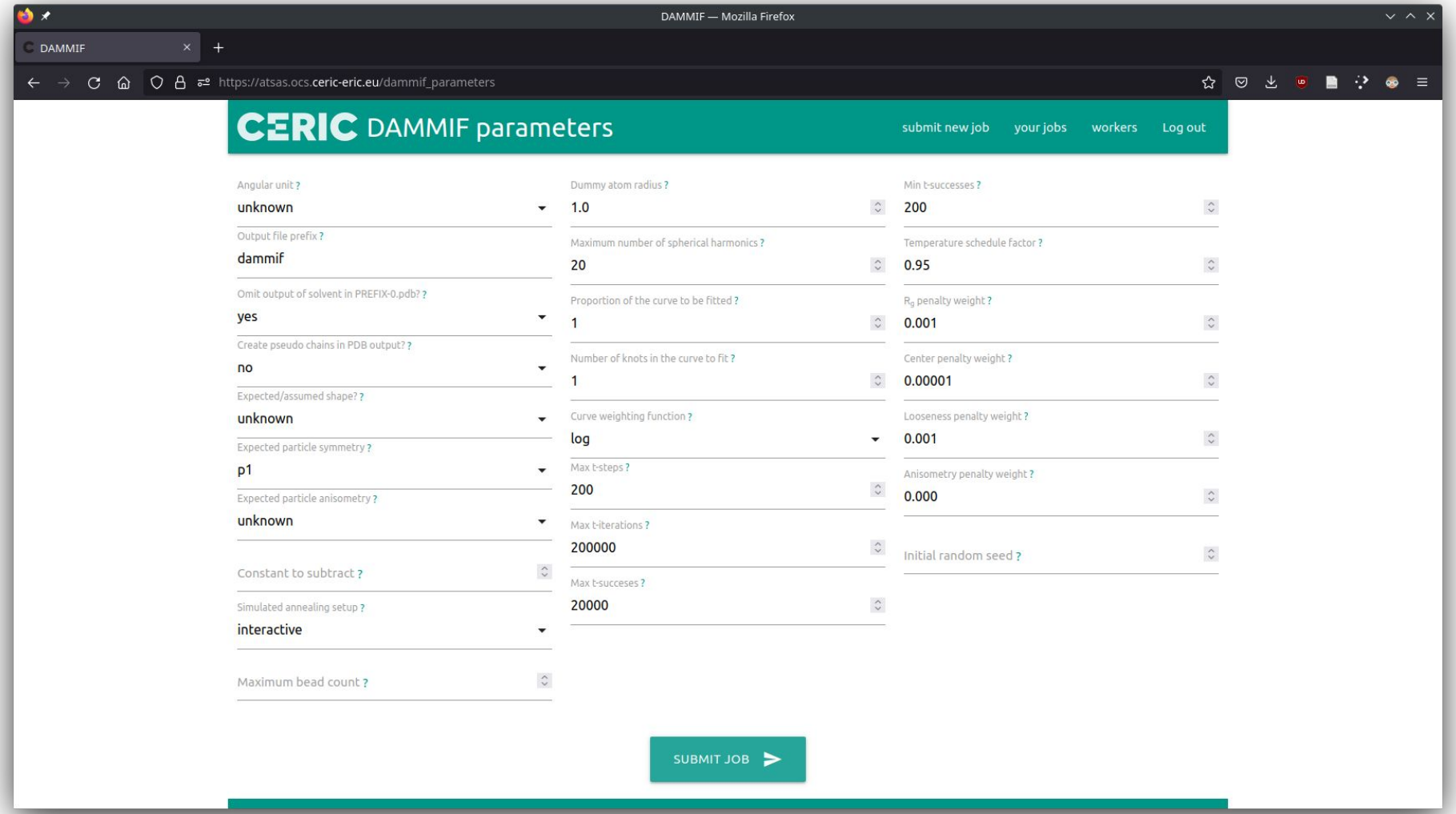
WebATSAS Software Example



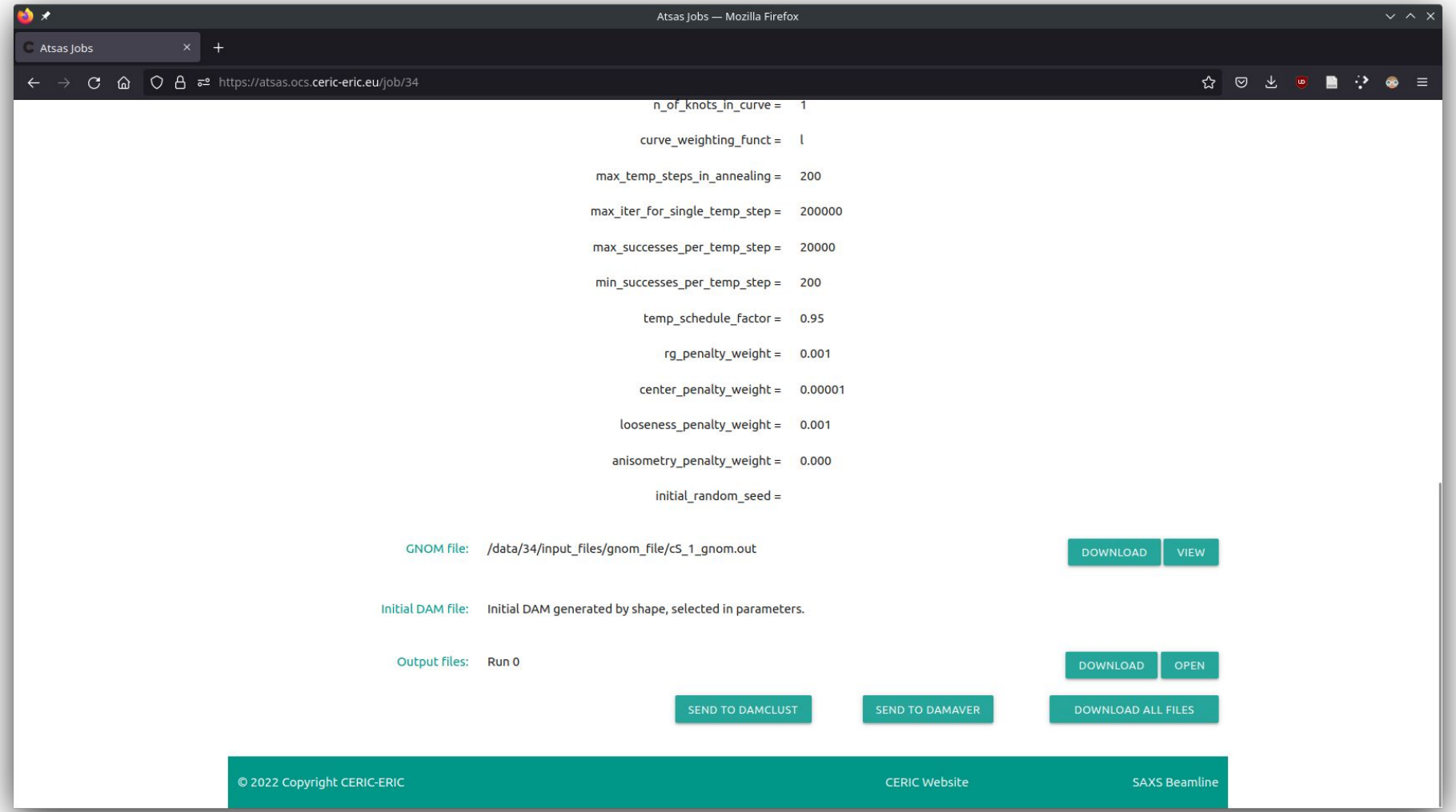
WebATSAS Software Example



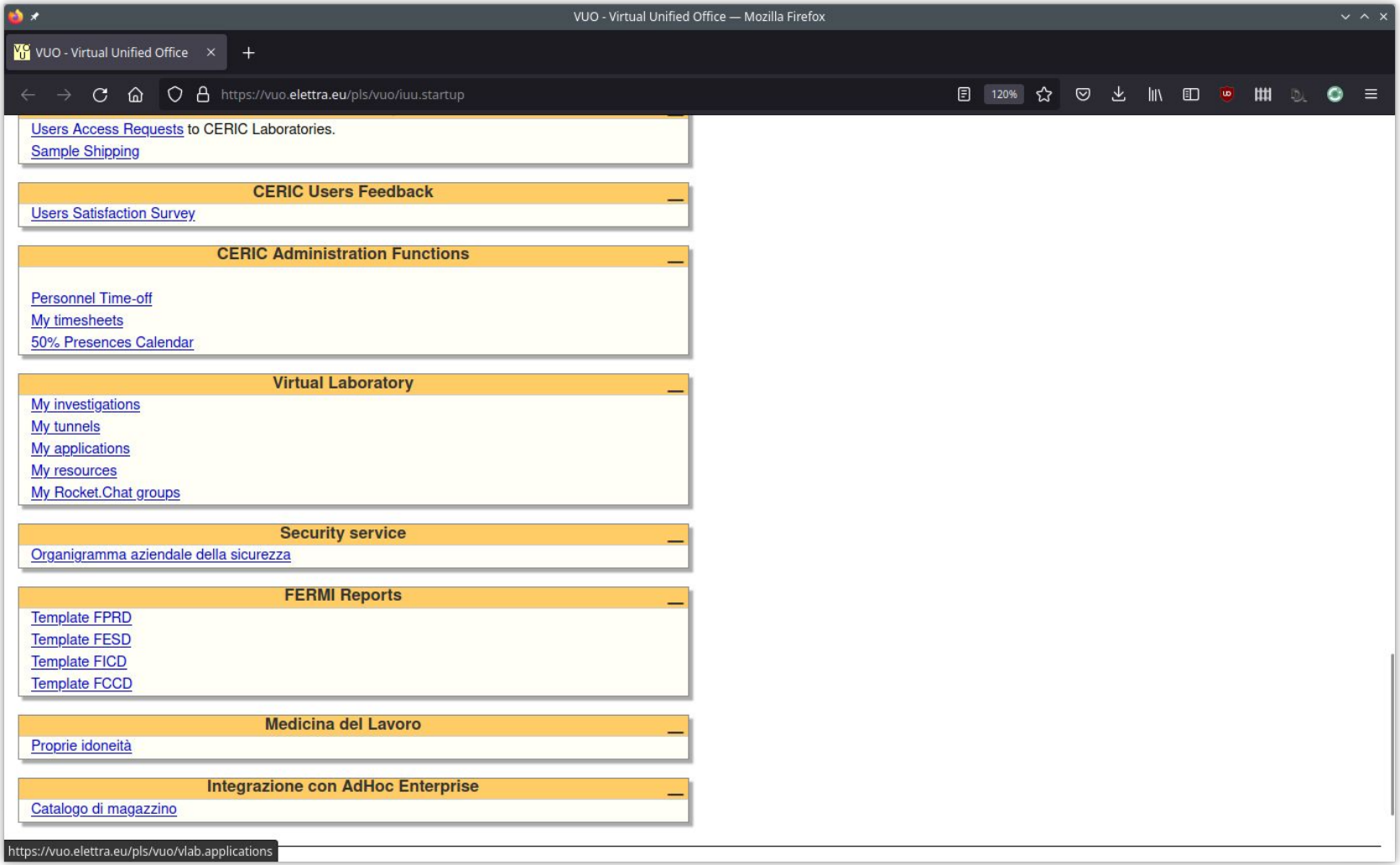
WebATSAS Software Example



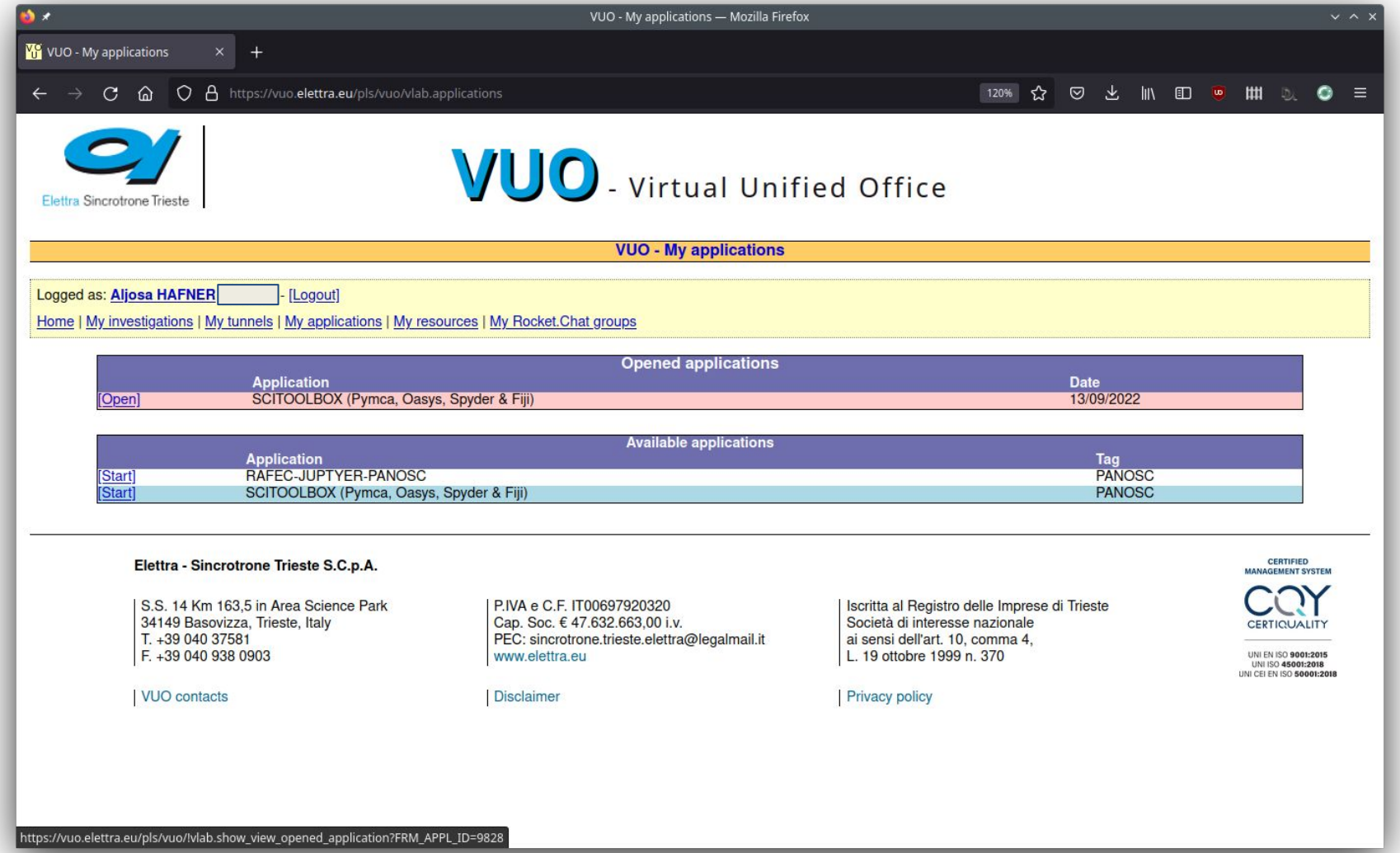
WebATSAS Software Example



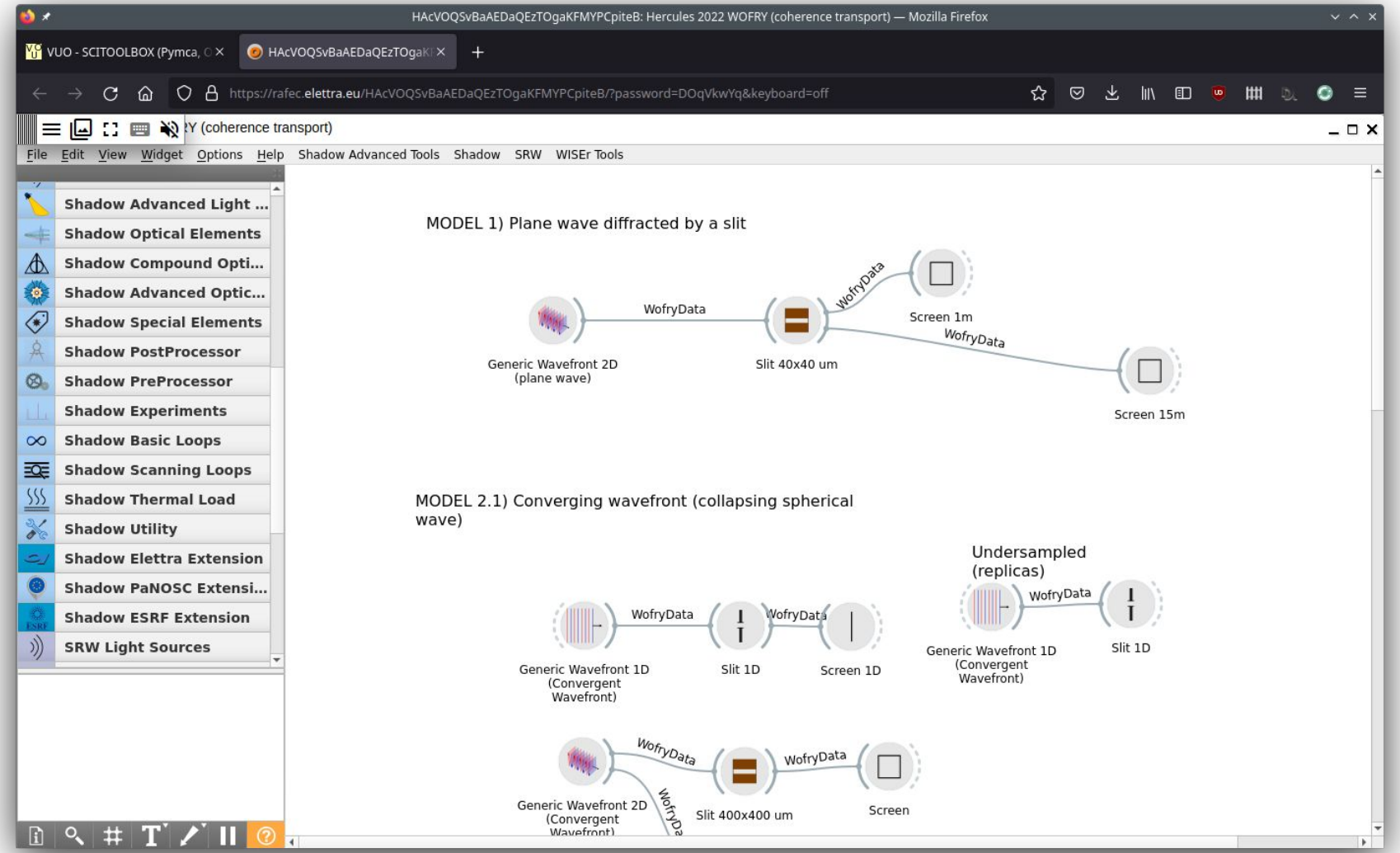
Rafec Remote Application Framework



Rafec Remote Application Framework

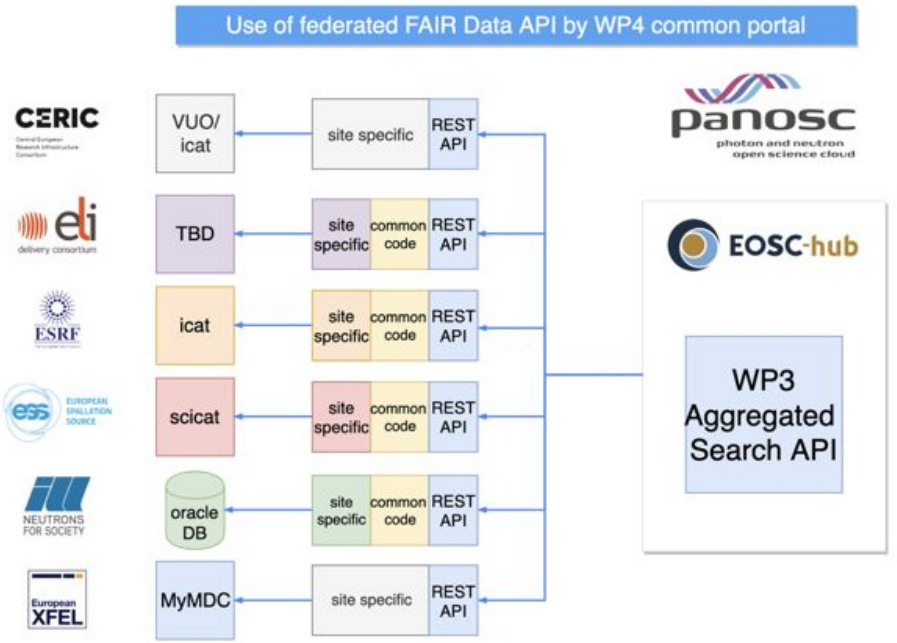


Rafec Remote Application Framework

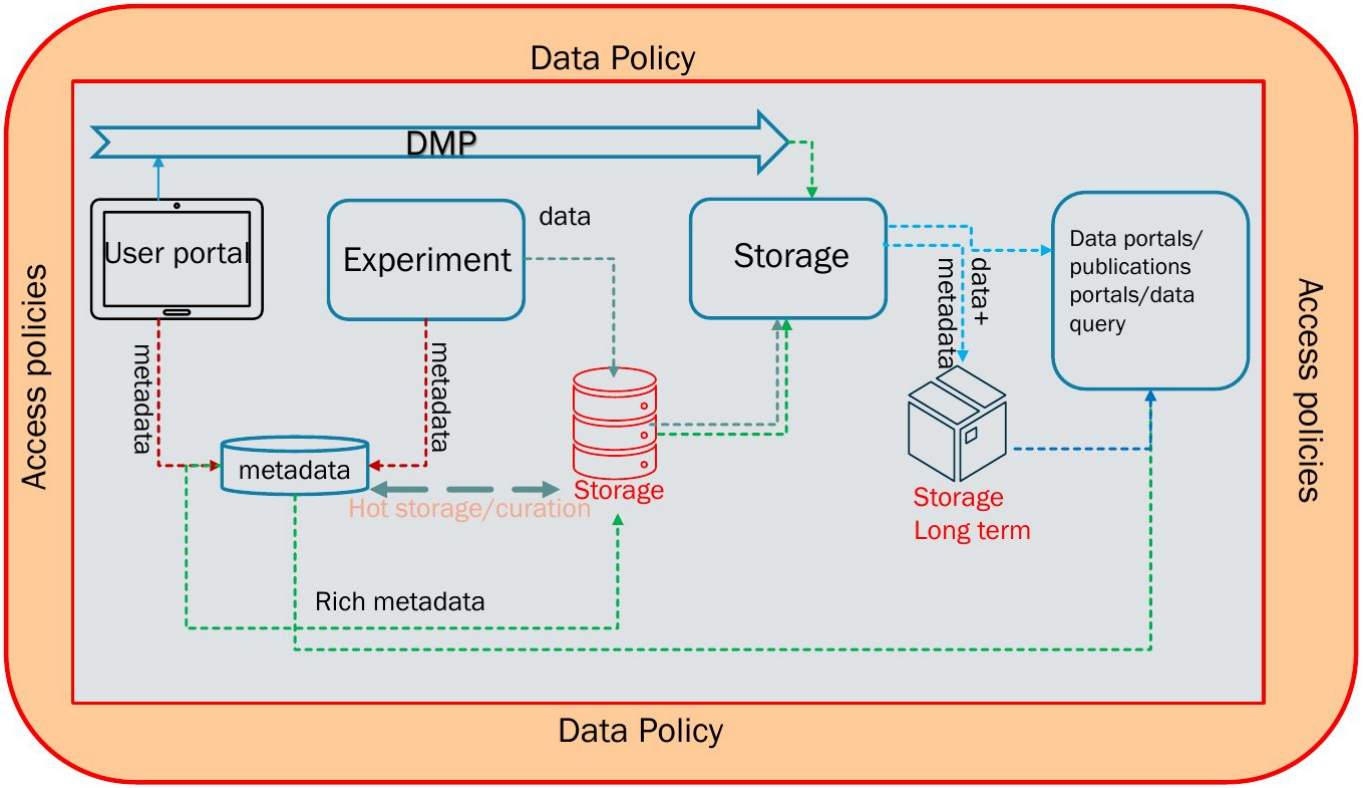


Catalogues

- 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

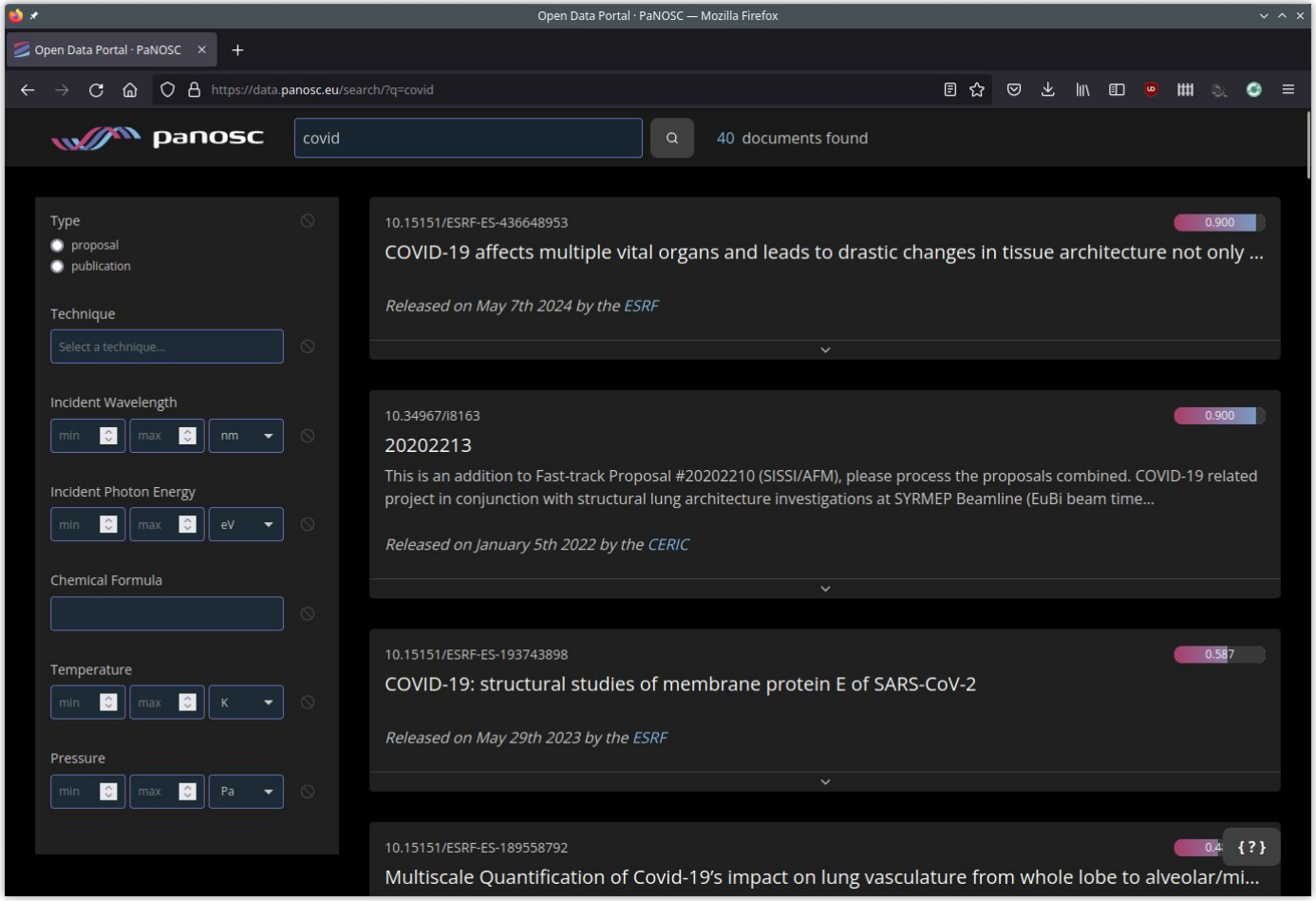


How it comes together

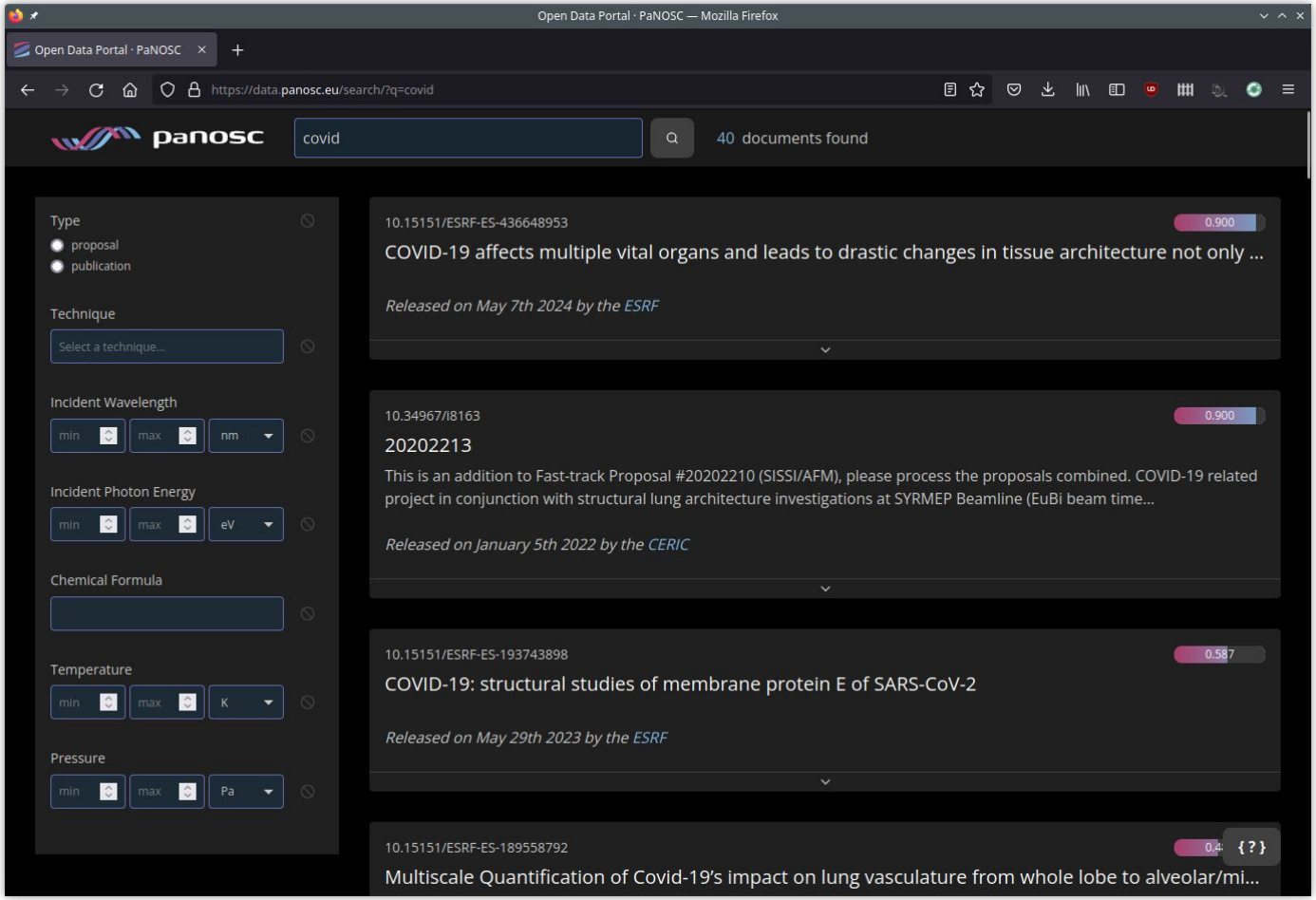


Source: ELI-ERIC;
https://www.fairsfair.eu/sites/default/files/ELI%20FAIR%20Data%20Policy.%20Guidelines%20and%20implementation%20challenges.pdf

Federated Search - data.panosc.eu



Assignment - Using the federated search



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 ; www.expands.eu ; 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://www.panosc.eu/all-use-cases/

Aljoša Hafner
aljosa.hafner@ceric-eric.eu