What PaNOSC changed for European XFEL

Luis Maia

European X-Ray Free Electron Laser

30th November, 2022

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

Main Achievements

- Scientific Data Policy review
- Open Proposals metadata publicly available via OAI-PMH and Loopback APIs
- Adoption of NeXus MX
- Libraries/Frameworks for Data Analysis
- Libraries/Frameworks for simulation
- Open proposals metadata, data and computing resources available for community at large
- Dissemination of information and trainings

panosc
photon and neutron
open science cloud

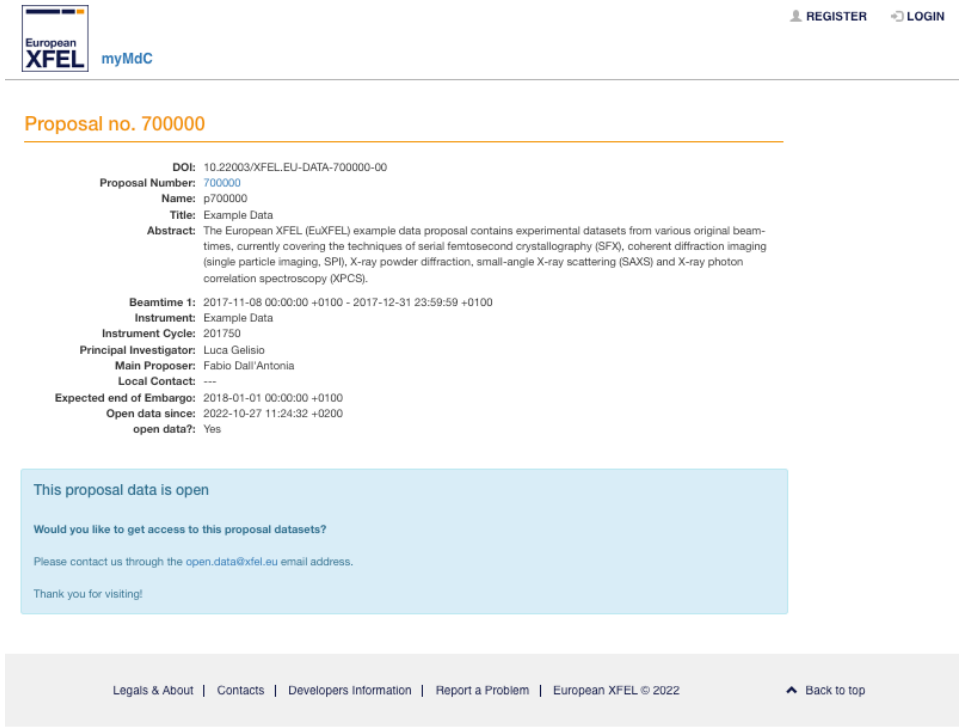
Scientific Data Policy review

European XFEL Scientific Data Policy (2017)	Updated European XFEL Scientific Data Policy
FAIR concepts not mentioned	Explicitly mention FAIR and the objectives
Data Management Plan not mentioned	Define the Data Management Plan as mandatory
Reduced data not mentioned	Explicitly mention the possibility of storing reduced data as raw data
Processed data are not included in the long-term storage	Consider storing processed data for long-term
Some particular auxiliary data were defined (geometries, calibration constants)	Define general auxiliary data term, expand on this data and introduce rules to manage it
Electronic logbook not mentioned	Explicitly mention the electronic logbook as part of the metadata capture
ORCID not mentioned	Encourage usage ORCID as a means of linking users to data DOIs

European XFEL Scientific Data Policy (2017)	Updated European XFEL Scientific Data Policy
Data format is not explicitly mentioned	Define HDF5 as the preferred data format is HDF5
Only metadata from European XFEL software accepted	Allow the possibility of metadata from non-European XFEL software especially for processed data
Persistent identifiers (PIDs) mentioned in general	Explicitly define use of DOIs as PIDs
Granularity of PID is experiment and dataset	Explicitly mention PID can refer to a bespoke collection of datasets in addition to experiment and dataset automatically generated
Personal Data Protection included but GDPR not explicitly mentioned	Make explicit reference to the GDPR regulations
Few good practices included	Expand on the “Good practices” according to the new recommendations

Metadata publicly available (1)

- **myMdC** (https://in.xfel.eu/metadata) OAI-PMH available and DOI automatic generation and publishing available



This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2022-11-23T10:38:36Z</responseDate>
  <request>https://in.xfel.eu/metadata{"verb"=>"Identify"}</request>
  <Identify>
    <repositoryName>European XFEL Proposals and Datasets Repository</repositoryName>
    <baseURL>https://in.xfel.eu/metadata</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>luis.maia@xfel.eu</adminEmail>
    <adminEmail>krzysztof.wrona@xfel.eu</adminEmail>
    <earliestDatestamp>2019-02-14T14:27:59Z</earliestDatestamp>
    <deletedRecord>no</deletedRecord>
    <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
  </Identify>
</OAI-PMH>
```

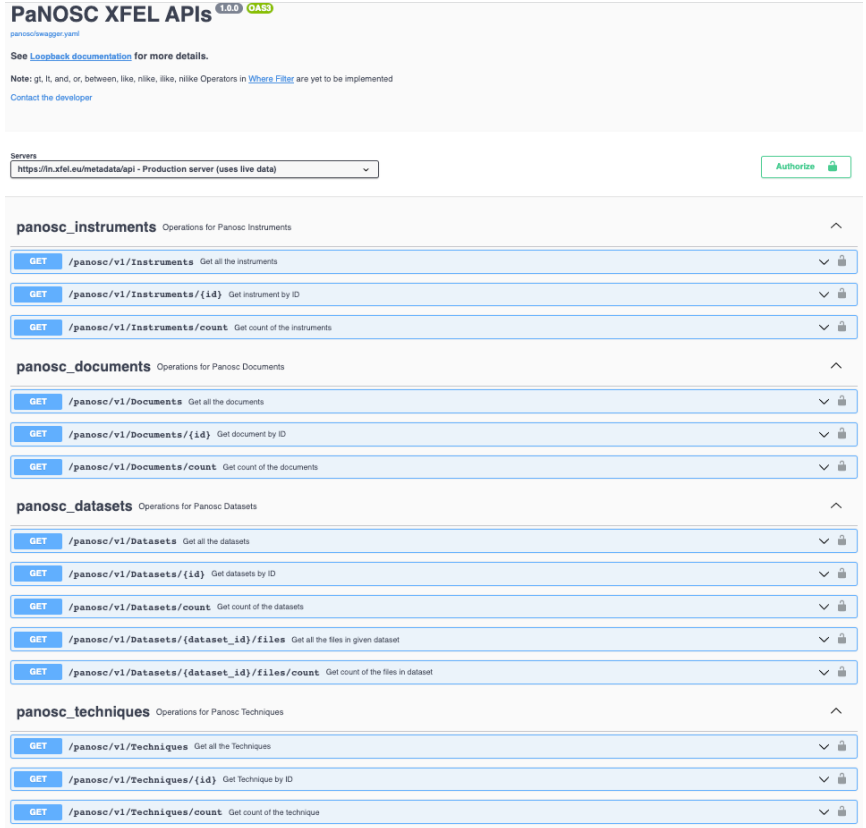

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2022-11-23T10:40:59Z</responseDate>
  <request>https://in.xfel.eu/metadata{"identifier"=>"oai:xfel.eu:30", "metadataPrefix"=>"panosc", "verb"=>"GetRecord"}</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:xfel.eu:30</identifier>
        <datestamp>2022-10-27T09:59:47Z</datestamp>
        <setSpec>openaire_data</setSpec>
      </header>
      <metadata>
        <panosc:panoscType xmlns:panosc="http://scicat.esss.se/panosc/" xsi:schemaLocation="http://scicat.esss.se/panosc https://raw.githubusercontent.com/panosc-eu/fair-data-api/master/panosc.xsd">
          <panosc:id>10.22003/XFEL.EU-DATA-700000-00</panosc:id>
          <panosc:name>Example Data</panosc:name>
          <panosc:description>The European XFEL (EuXFEL) example data proposal contains experimental datasets from various original beam-times, currently covering the techniques of serial femtosecond crystallography (SFX), coherent diffraction imaging (single particle imaging, SPI), X-ray powder diffraction, small-angle X-ray scattering (SAXS) and X-ray photon correlation spectroscopy (XPCS). </panosc:description>
          <panosc:owner>Fabio Dall'Antonia</panosc:owner>
          <panosc:contributor>Luca Gelisio</panosc:contributor>
          <panosc:embargoEndDate>2018-01-01T00:00:00+01:00</panosc:embargoEndDate>
          <panosc:startDate>2017-11-08T00:00:00+01:00</panosc:startDate>
          <panosc:path>https://in.xfel.eu/metadata/doi/10.22003%2FXFEL.EU-DATA-700000-00</panosc:path>
          <panosc:technique>coherent diffraction imaging, serial femtosecond crystallography, small angle x-ray scattering, x-ray photon correlation spectroscopy, x-ray powder diffraction</panosc:technique>
          <panosc:sampleName>Water, Lysozyme (201804 small crystals), Lithium Titanate, No Sample, Lysozyme, Silica 50nm, Vycor, Sucrose Solution 3% v/v, Potassium hexacyanoferrate(II) trihydrate, Cu foil, Xenon, 2-Co8_pt14_8fold - 30nm Pt cap, 1-Co10_Pt_6fold, Ni-20 MLs - b, Ni75-11 MLs-b, Bolometer</panosc:sampleName>
          <panosc:chemicalFormula>AGIPD, Calibration - Dark HG, Calibration - Dark LG, Calibration - Dark MG, Configuration Tests, Dark, Diffraction, Diffraction data, General, Sample, SAXS 500kHz // no pump laser, scattering, SFX Jeting, Single Particle Diffraction, Test DAQ, XGM, XGM calibration, XPCS</panosc:chemicalFormula>
          <panosc:size>9074975744</panosc:size>
          <panosc:wavelength/>
        </panosc:panoscType>
      </metadata>
    </record>
  </GetRecord>
</OAI-PMH>
```


panosc
photon and neutron open science cloud

Metadata publicly available (2)

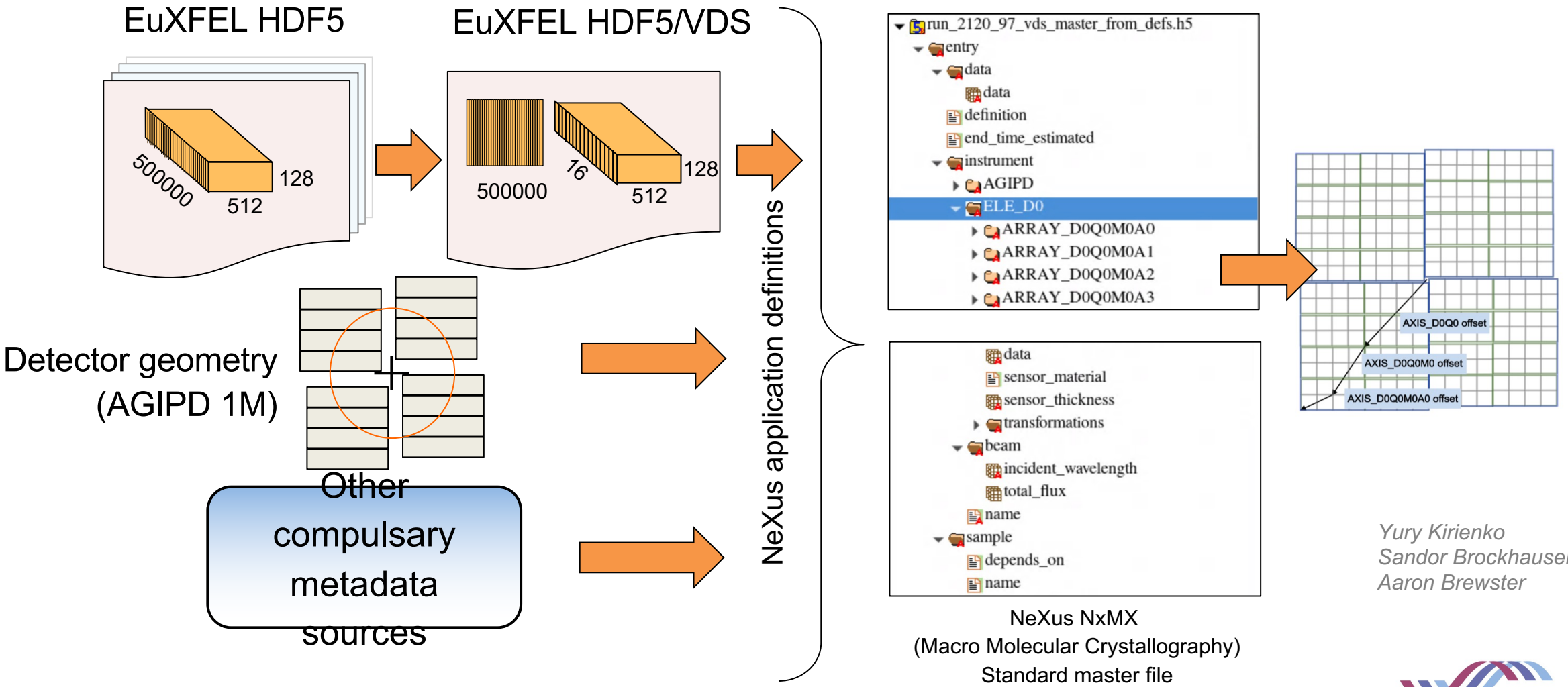
- **myMdC** RESTful and PaNOSC Loopback APIs [available]



https://in.xfel.eu/metadata/api-docs/index.html

photon and neutron open science cloud

Adoption of NeXus MX, SFX use case



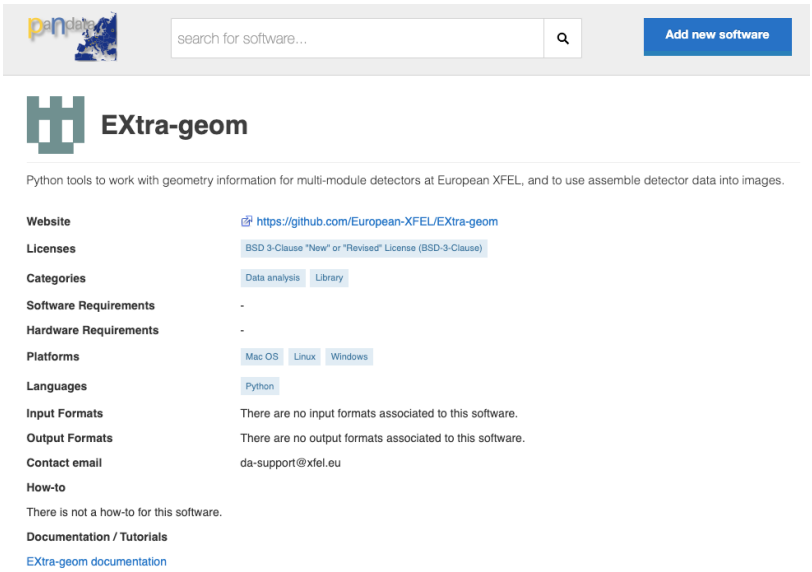
NeXus NxMX
(Macro Molecular Crystallography)
Standard master file

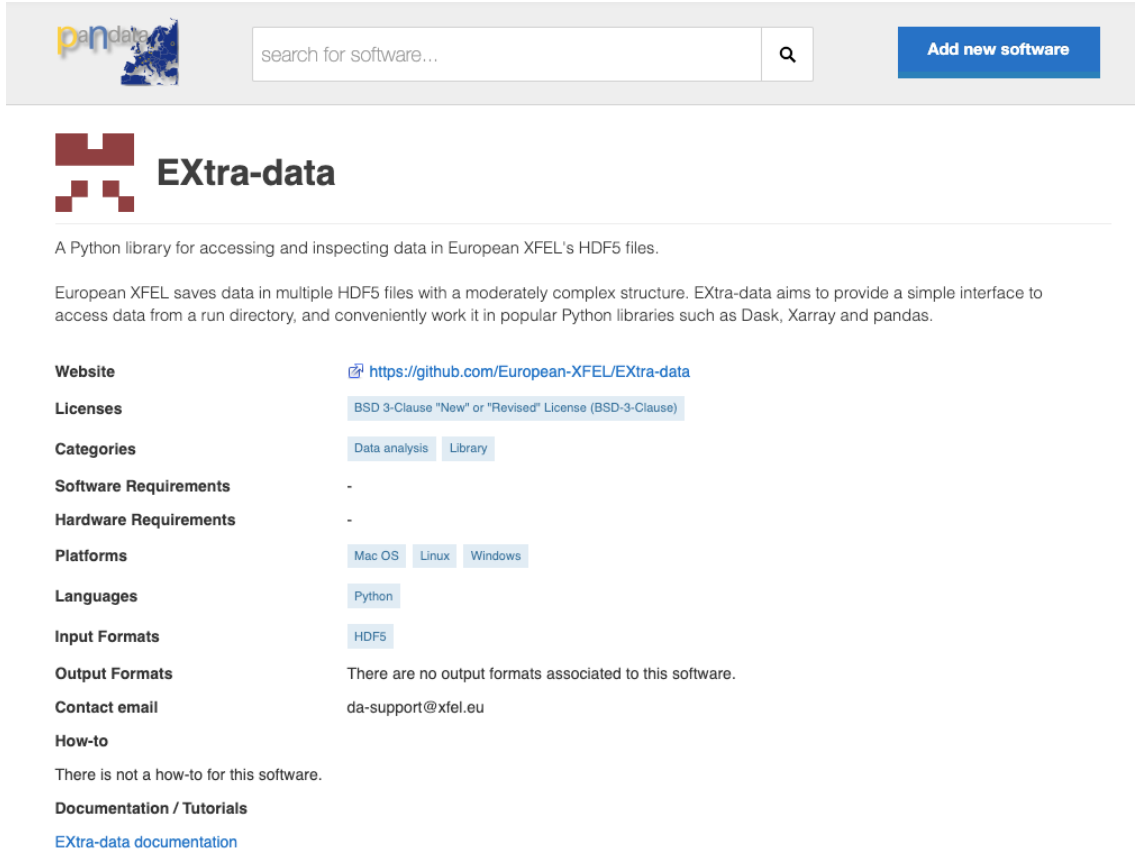
Yury Kirienko
Sandor Brockhauser
Aaron Brewster

panosc
photon and neutron
open science cloud

Libraries/Frameworks for Data Analysis

- Python packages EXtra-data and EXtra-geom



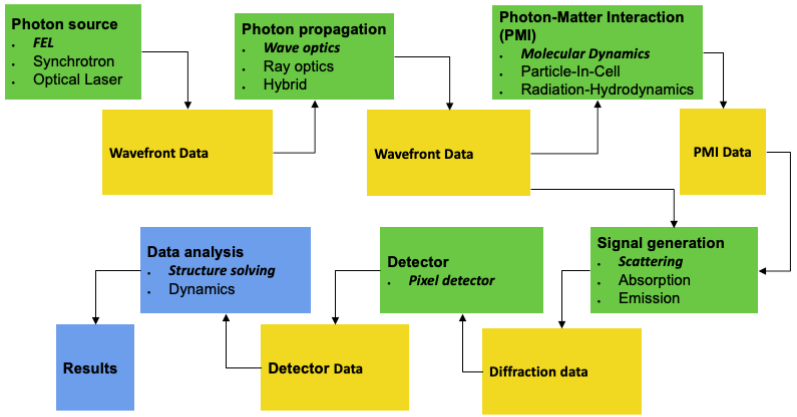


panosc
photon and neutron
open science cloud

Libraries/Frameworks for simulation

Python packages libpyvinyl and SimEx-Lite

SimEx-Lite

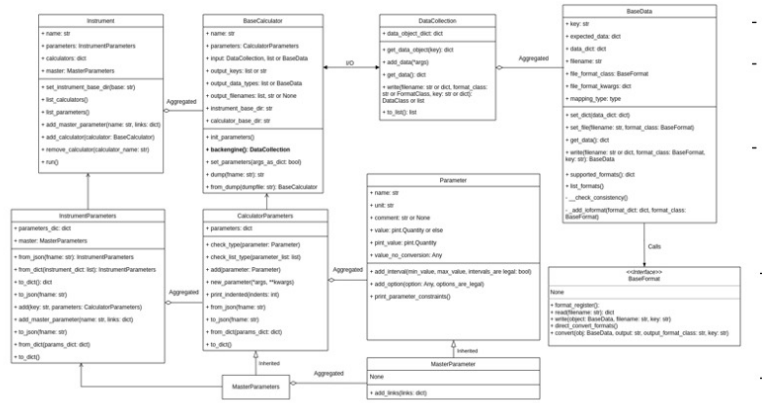


Calculators
Data interfaces

https://github.com/PaNOSC-ViNYL/SimEx-Lite

- It is the core package of the SIMEX platform providing the calculator interfaces and data APIs.
- It is built based on **libpyvinyl**
- A **calculator** can be easily constructed within SimEx-Lite with the definition of the corresponding **data class** and the **format class.**
- Users can choose which back engine software to implement to make the installation minimal per needs.

libpyvinyl

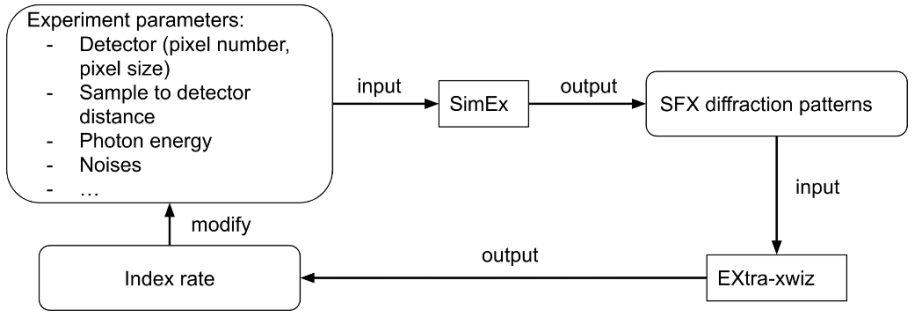


- Harmonize the user interfaces for **neutron** and **X-ray** simulation
- Base classes:
  - BaseCalculator
  - BaseData
  - BaseFormat
- Auxiliary classes:
  - Parameter
  - CalculatorParameters
  - DataCollection
  - Instrument
- It provides the developers of start-to-end simulation platforms in neutron and X-ray community with a framework to integrate various backengine software.
- It reduces the effort needed to integrate their software to a start-to-end simulation platform for simulation software developers.

https://github.com/PaNOSC-ViNYL/libpyvinyl

A simple usage example: https://github.com/PaNOSC-ViNYL/libpyvinyl/tree/master/tests/integration/plusminus

Optimize sample to detector distance in different energy densities

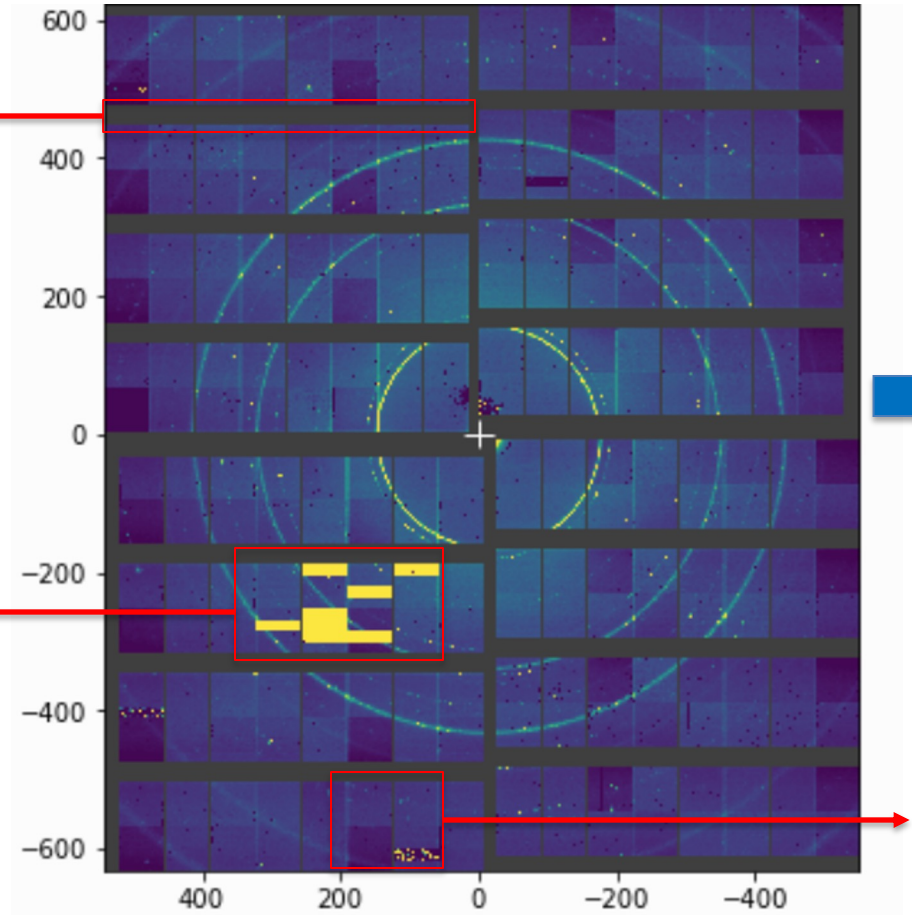


photon and neutron open science cloud

Serial Crystallography Simulation and Analysis

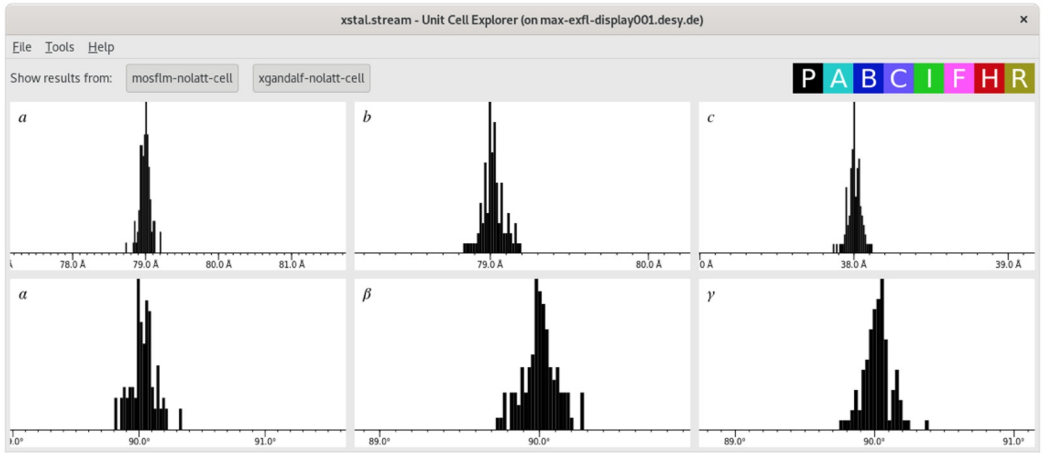
Panel gaps

Masks



Noise

Demonstrate the effects of noise, panel gaps and masks on the results of crystallography analysis with SimEx simulation diffraction pattern results



Scenarios:

New researchers want to get hands-on experience of data processing for serial crystallography.

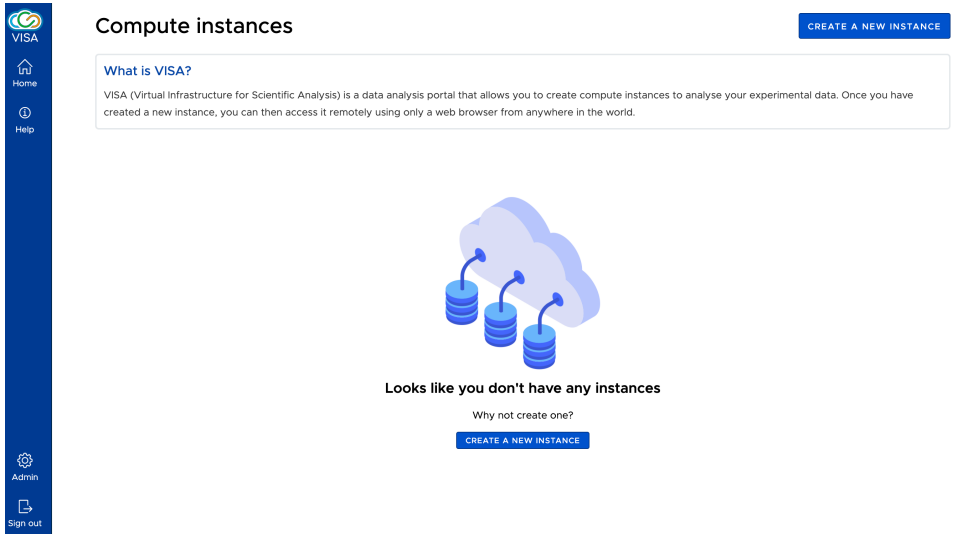
Starting from a working example of a simulation coupled to an analysis pipeline

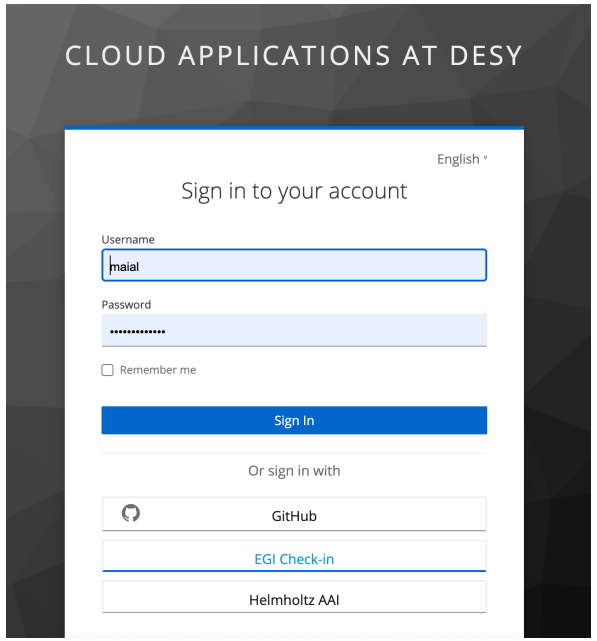
Explore how different experimental conditions and different analysis options affect the results.

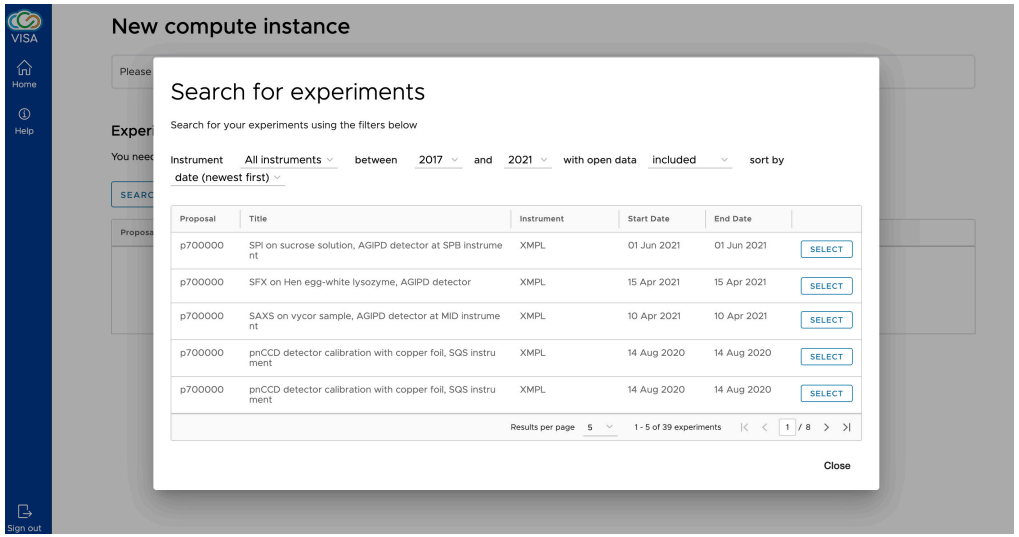
panosc
photon and neutron open science cloud

Data open to Scientific Community (1)

- Visa portal visa.xfel.eu for community at large (e.g. EGI Check-in which includes Umbrella)



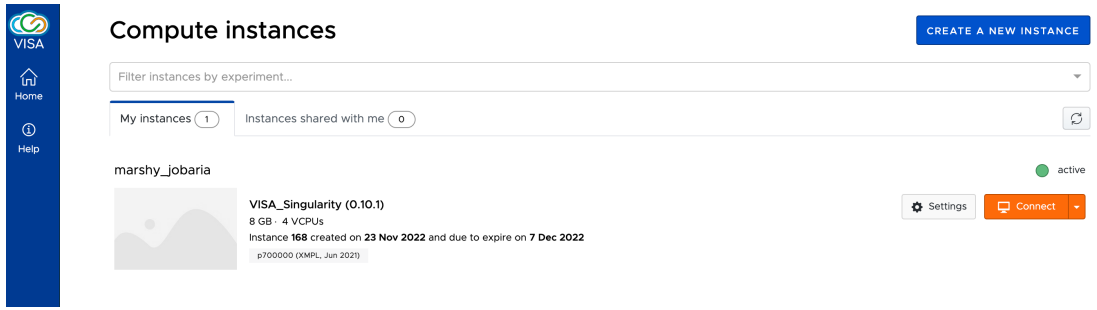


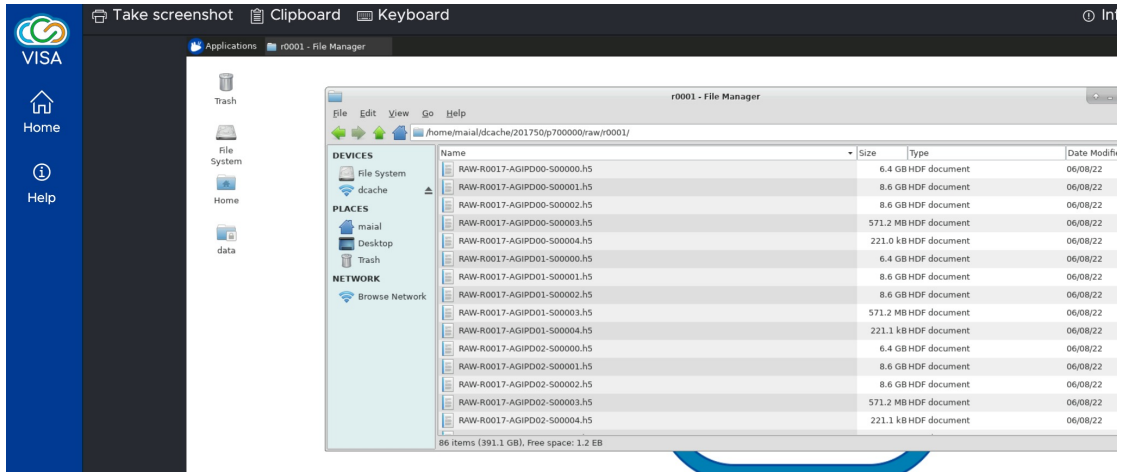


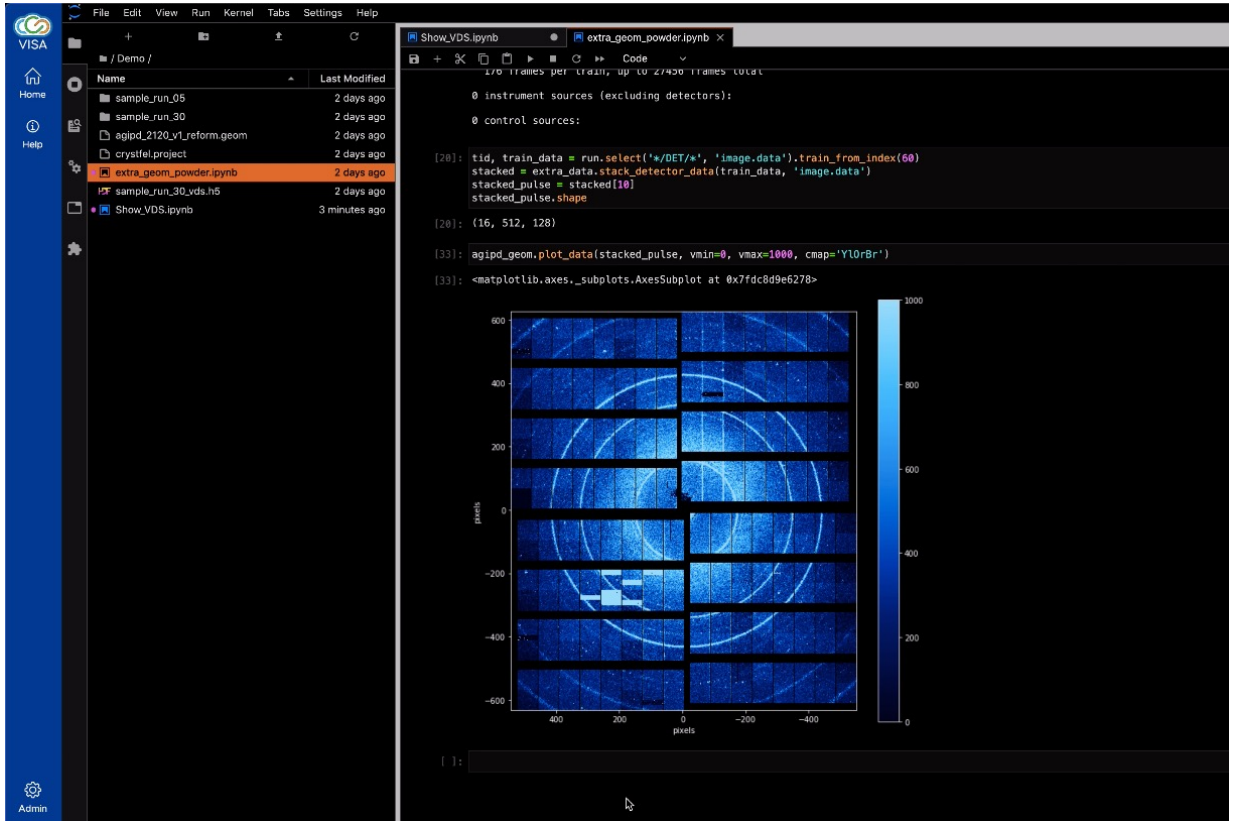
panosc
photon and neutron open science cloud

Data open to Scientific Community (2)

- Visa portal allows access to proposals data



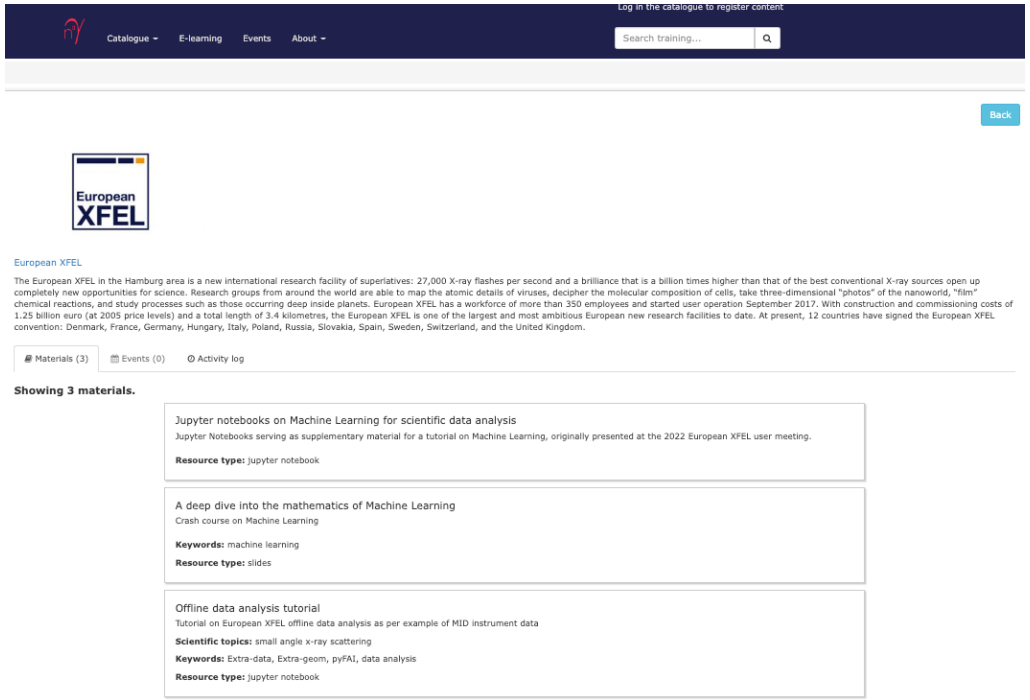


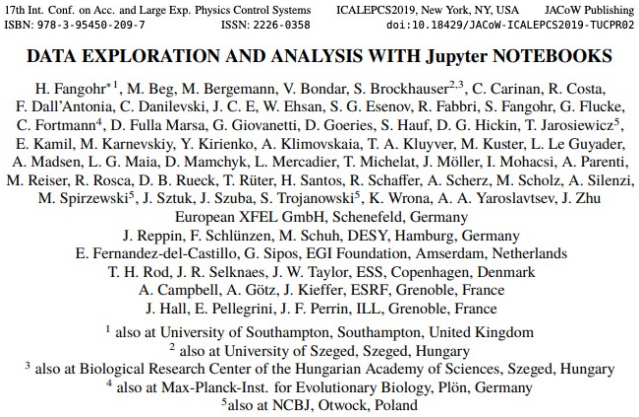


panosc
photon and neutron
open science cloud

Dissemination of information and trainings

- Production of scientific papers and trainings





panosc
photon and neutron open science cloud

Acknowledgements

European XFEL

- Carsten Fortmann-Grote
- Deike Pahl
- Fabio Dall'Antonia
- Hans Fangohr
- Ivette Bermudez Macias
- Janusz Malka
- Janusz Szuba
- Julian Hörsch
- Juncheng E
- Krzysztof Wrona
- Luca Gelisio
- Luis Maia
- Oleksii Turkot
- Robert Rosca
- Sandor Brockhauser
- Sina Nolle
- Steve Aplin
- Sudhanshu Singh
- Tarcisio Fedrizzi
- Thomas Kluyver
- Thomas Tschentscher
- Yury Kirienko
- Varun Singh

DESY.

- Humaira Abdul Salam
- Johannes Reppin
- Michael Schuh
- Patrick Fuhrmann
- Tim Wetzel

panosc
photon and neutron open science cloud

Thank you

luis.maia@xfel.eu

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