

# Photon and Neutron reference data sets

## Document Control Information

Settings	Value
Document Identifier:	D4.2
Project Title:	ExPaNDS
Work Package:	WP4
Document Author(s):	Alun Ashton (PSI), Anton Barty (DESY), Uwe Konrad (HZDR), Franz Lang (UKRI), Znedek Matej (Max IV), Majid Ounsy (SOLEIL), Christopher Reynolds (Diamond)
Internal Reviewer(s):	Patrick Fuhrmann (DESY), Sophie Servan (DESY)
Responsible Partner:	UKRI
Doc. Issue:	1.0
Dissemination level:	Public
Date:	28/02/2021

## Abstract

This document summarises work in ExPaNDS Data Analysis Services work package (WP4) for the preparation and publication of Photon and Neutron reference data sets to be used to adapt, align and validate the project's Prototype Data Analysis Services.

## Licence

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.*

## Executive Summary

We identify, prepare and publish Photon and Neutron reference data sets, which will be used to adapt, align and validate the Prototype Data Analysis Services developed within WP4 of ExPaNDS. We target actual use cases from the facilities with a real and tangible remote data analysis need, and to adapt these workflows to the EOSC/cloud computing paradigm. To this end, WP4 has selected the following widely used techniques and workflows as candidate EOSC demonstration projects:

1. Serial crystallography (DESY)
2. Full field tomography (Max-IV, PSI)
3. Single crystal X-ray diffraction (Diamond)
4. Ptychography (Diamond, PSI, MAX IV)
5. Small angle X-ray scattering (SAXS) (Soleil)
6. Neutron diffraction (SANS and Reflectometry) and tomography (UKRI/ISIS)
7. Reflectometry (Diamond)
8. Terahertz spectroscopy (HZDR)
9. Scanning electron imaging (Diamond)

Selecting use cases from existing photon and neutron work flows aligns EOSC services with the needs of the research community as well as contributing to sustainability after the project by fostering the development of EOSC services in demand within the science user community.



# Table of Contents

<b>Executive Summary</b>	<b>2</b>
<b>Introduction</b>	<b>3</b>
Background	3
Definitions of types of data	4
<b>Reference data sets</b>	<b>5</b>
Summary of published datasets	5
Serial crystallography	6
Full field tomography	7
4D Full field tomography	9
Single-crystal X-ray diffraction	10
Ptychography, Ptychographic X-ray computed tomography	11
Kinetic SAXS, 2D Scanning SAXS	12
Neutron Small Angle Scattering	13
Neutron Imaging/Tomography	14
Neutron - Reflectometry	16
Terahertz Spectroscopy	17
Electron scanning diffraction imaging (4DSTEM)	18

## 1. Introduction

### 1.1. Background

The goal of task 4.2 and associated deliverable 4.2 is to identify publicly available datasets that can be used as example/reference data sets for the Prototype Data Analysis Services. In particular, to be used in combination with the DAaaS (Data Analysis as a Service) portal and individual DAaaS platform prototypes to inform about any necessary changes. Our collection of reference datasets needs to be representative of the different types of data sets that will be generated at the various ExPaNDS facilities.

This requires us to:

- **Identify:** Find and agree on a set of different use cases (and example data sets for each use case), for which Data Analysis Services can be validated.
- **Prepare:** The aim is to prepare the reference data sets into existing, recognisable and self-contained formats (a common file format is probably not feasible). The data sets must include all artifacts and metadata that are needed in order for someone with reasonable knowledge of the technique (such as a beamline scientist) to be able to analyse the data correctly; this can include, for example, calibration runs and information about the sample and sample environment. The metadata should be included in such a way that the data are findable (in some way that is yet to be



determined). The preference is for data sets used in common and well-established experimental techniques.

- **Publish:** The reference data needs to be open, accessible and outside any embargo period.
- **Adapt, align and validate:** Ensure that DAaaS prototypes can be used with the reference data sets (this likely will take some iterations to get right). Ideally, this will also train users on how to use the DAaaS platform with their own data. Not clear if this includes potential changes to either the (meta) data files/formats, or the DAaaS prototype, or both.

Outputs from this task will potentially feed into WP5, to be integrated in the PaN learning platform wherever possible, to accompany the training material made available for the different Data Analysis pipelines. They can also be useful for WP6 and PaNOSC WP9 as use cases to be showcased to the community.

## 1.2. Definitions of types of data

Definitions of data types are being developed as part of PaNOSC<sup>1</sup> and ExPaNDS WP2. In line with these data definitions which are being developed, we adopt the following definitions (which may see some modification as WP2 evolves):

- **Raw data** refers to data collected from facility instruments. This is the unprocessed data saved during the course of an experiment.
- **Reduced/processed/intermediate data** describes data obtained by reducing/processing raw data and may include the results of intermediate processing steps. Users should be able to take the reduced/processed data home for further analysis including the results of intermediate data processing steps.
- **Metadata** describes information describing data collected from instruments, including (but not limited to) the context and purpose of the experiment and the experimental team, the facility/beamline/instrument used, samples studied and their preparation, experimental conditions, electronic logbooks generated during the experiment and other logistical information. Metadata is used for the purposes of finding and locating the data set, for interpreting the data, and includes information necessary to identify sub-portions of the entire dataset (eg: sample changes and which portions of the data correspond to which samples).
- **Auxiliary data** refers to data that provides information regarding the experiment and its dataset but may have been collected outside the context of the experiment. This includes (but not limited to) additional data required to interpret the raw data such as encoder calibration constants, detector calibration data, files describing the experiment geometry, and measurements from reference standards necessary for calibration during the course of data processing. Auxiliary data may have been measured in separate experiments (eg: facility commissioning). Auxiliary data can also include processing scripts and software parameters used in creating data for a publication (if applicable).

<sup>1</sup> See chapter 3.1 of PaNOSC data policy framework: <https://doi.org/10.5281/zenodo.3862701>



## 2. Reference data sets

### 2.1. Summary of published datasets

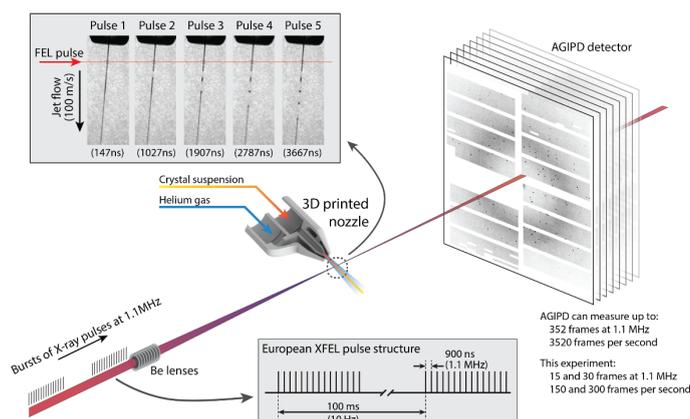
#	Technique	Lead	Published location
1	Serial crystallography	DESY	<a href="https://cxidb.org/id-98.html">https://cxidb.org/id-98.html</a>
2	Full field tomography	PSI	<a href="https://doi.psi.ch/detail/10.16907/7eb141d3-11f1-47a6-9d0e-76f8832ed1b2">https://doi.psi.ch/detail/10.16907/7eb141d3-11f1-47a6-9d0e-76f8832ed1b2</a>
3	4D Full field tomography	Max-IV	<a href="https://app.globus.org/file-manager?origin_id=e133a81a-6d04-11e5-ba46-22000b92c6ec&amp;origin_path=%2Ftomobank%2Ftomo_00080%2F">https://app.globus.org/file-manager?origin_id=e133a81a-6d04-11e5-ba46-22000b92c6ec&amp;origin_path=%2Ftomobank%2Ftomo_00080%2F</a>
4	Single-crystal X-ray diffraction	Diamond	<a href="https://zenodo.org/record/4445063#.YAVazNanxR4">https://zenodo.org/record/4445063#.YAVazNanxR4</a>
5	Ptychography, Ptychographic X-ray computed tomography	Max-IV	<a href="https://doi.org/10.5281/zenodo.3702582">https://doi.org/10.5281/zenodo.3702582</a>
6	2D Scanning SAXS	Soleil	<a href="http://doi.org/10.5281/zenodo.4553498">http://doi.org/10.5281/zenodo.4553498</a>
7	Neutron Small Angle Scattering	UKRI/ISIS	<a href="https://github.com/DAAA-reference-data/SANS">https://github.com/DAAA-reference-data/SANS</a>
8	Neutron Imaging/Tomography	UKRI/ISIS	<a href="https://doi.org/10.5281/zenodo.3672798">https://doi.org/10.5281/zenodo.3672798</a>
9	Neutron reflectometry	UKRI/ISIS	<a href="https://data.isis.stfc.ac.uk/doi/INVESTIGATION/84794745/">https://data.isis.stfc.ac.uk/doi/INVESTIGATION/84794745/</a>
10	Terahertz Spectroscopy	HZDR	<a href="https://rodare.hzdr.de/record/277">https://rodare.hzdr.de/record/277</a>
11	Electron scanning diffraction imaging (4DSTEM)	Diamond	<a href="https://dls.ltd.sharepoint.com/:u:/s/GRA0046/ER6eICBerHxDICfYCIBeUFsBakNbDsREMRobNW9VvrUXaQ?e=b7MLjY">https://dls.ltd.sharepoint.com/:u:/s/GRA0046/ER6eICBerHxDICfYCIBeUFsBakNbDsREMRobNW9VvrUXaQ?e=b7MLjY</a>



# 1. Serial crystallography

**Lead:** DESY

**EOSC motivation:** Experiments generate up to 150 TB per day of data saved at the measurement facility. Such large datasets are impractical for users to take home. Subsequent analysis needs to be performed remotely making it attractive for deployment as a cloud-like use case. Involving EOSC in the analysis and re-use of this data is an appropriate use case.



**Raw data:** Raw data as saved by the facility. File format varies according to facility and instrument - could be XTC files (LCLS), HDF5 files (Eiger) or CBF files (Pilatus).

**Reduced/processed/intermediate data:** MTZ files (<1GB) ready for structure analysis. Option of exporting reduced datasets containing only hits (<1TB) that are useful for researchers to take home for more detailed analysis or archiving. This gives a defined output for the pipeline. Will consider delivery of MTZ files through a web page or email rather than disk-based file retrieval.

**Metadata:** From the published publication for this example

**Auxiliary data:** Contained in calibration files etc saved with the data sets

**Analysis pipeline:** First perform hit finding using Cheetah (or equivalent) and then process hits through CrystFEL to produce MTZ files. Current interface would require a virtual desktop approach with access to underlying raw data and scratch space for intermediate file access. Will try to adapt to a workflow streaming data rather than using intermediate files for each step.

**Sample data set:** Currently considering: <https://cxidb.org/id-98.html>.

Raw data: 10TB. Reduced data containing hits only: 79 GB. MTZ file: <1GB.

Current data location: Processed data on CXIDB, raw data on dCache at DESY.

**Associated publication:** 10.1063/1.5124387

**Data DOI:** 10.11577/1598270



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

CXIDB ID 98	
Deposition Summary	
Depositor:	Oleksandr Yefanov
Contact:	oleks...@desy.de
Deposition date:	2020-02-07
Last modified:	2020-02-07
DOI:	10.11577/1598270
Publication Details	
Title:	Evaluation of serial crystallographic structure determination within megahertz pulse trains
Authors:	Oleksandr Yefanov et al.
Journal:	Structural Dynamics
Year:	2019
DOI:	10.1059/1.512487
Experimental Conditions	
Method:	Serial Femtosecond Crystallography
Sample:	Lysozyme
Wavelength:	1.33 Å (9.30 keV)
Lightsource:	European XFEL
Beamline:	SPB/SFX
Data Files	
Raw Data:	cheetah/ (4 KB)
Indexing:	indexing/ (4 KB)
Auxiliary Files	
Phenix Files:	phenix_0051+0151.tar.gz (262.6 MB)

Many other potential datasets at: <http://cxidb.org/>

Alternative:

<http://cxidb.org/id-32.html>

Associated publication(s): Sci Data **3**, 160021 (2016).

<https://doi.org/10.1038/sdata.2016.21>

## 2. Full field tomography

**Lead:** PSI

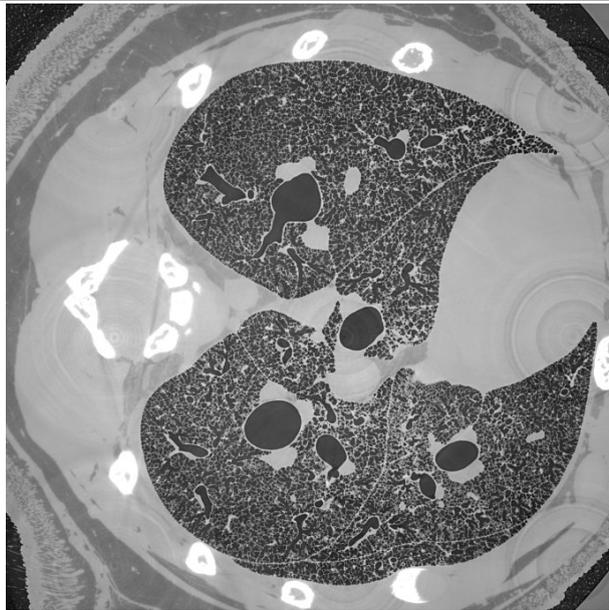
**EOSC motivation:** Full field tomography is a well established and widely used imaging technique. Synchrotron full field tomography allows imaging specimens with high resolution in space (3D) and time (4D).

Tomography datasets often present large volumes (100 GBs - few TBs) difficult to compress and transfer. The tomographic reconstruction is highly demanding on compute (GPU) and storage resources for the intermediate and/or final result. In addition, the optional image segmentation step may be demanding on computer memory.

The offline analysis (after experiment) could be performed remotely by users at home making it attractive for deployment as a cloud-like use case. Finally, this technique is applied at many facilities and in different scientific domains - therefore a portable result is more useful.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.*



Imaging of a complete intact post mortem juvenile rat lung, see sample data set below

**Raw data:**

**Reduced/processed/intermediate data:**

**Metadata:**

**Auxiliary data:**

**Analysis pipeline:**

The process of assembling the final 3D volume from individual 360-degree tomographic scans consisted of several steps. First, all individual projections were dark-field corrected using the mean of the 50 calibration dark frames. Second, a (dynamic) flat-field correction based on the principal component analysis was applied ([Van Nieuwenhove et al. 2015<sup>2</sup>](#)). Finally, tomographic reconstruction from phase-retrieved projections was performed using the gridrec algorithm ([Marone and Stampanoni 2012<sup>3</sup>](#)). More details in the associated publication below.

**Sample data sets:** Micrometer-resolution X-ray tomographic imaging of a complete intact post mortem juvenile rat lung:

<https://doi.psi.ch/detail/10.16907/7eb141d3-11f1-47a6-9d0e-76f8832ed1b2>

**Associated publication(s):**

Borisova, E., Lovric, G., Miettinen, A. et al. Micrometer-resolution X-ray tomographic full-volume reconstruction of an intact post-mortem juvenile rat lung. *Histochem Cell Biol* (2020). <https://doi.org/10.1007/s00418-020-01868-8>

<sup>2</sup> <https://doi.org/10.1364/oe.23.027975>

<sup>3</sup> <https://doi.org/10.1107/S0909049512032864>



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.*

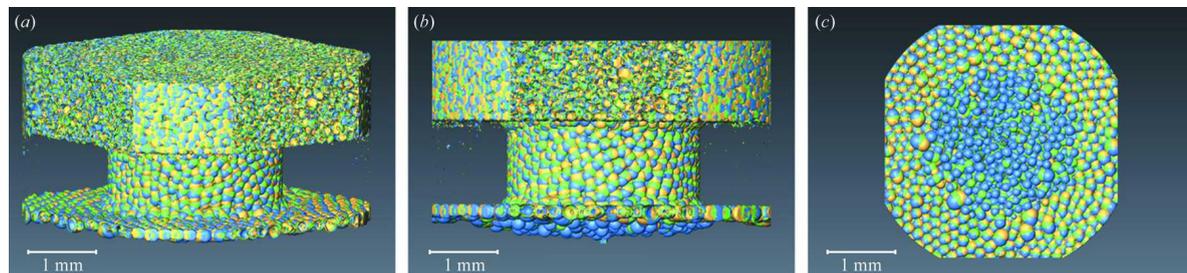
## 3. 4D Full field tomography

**Lead:** MAX IV

**EOSC motivation:** Full field tomography is a well established and widely used imaging technique. Synchrotron full field tomography allows imaging specimens with high resolution in space (3D) and time (4D).

Tomography datasets often present large volumes (100 GBs - few TBs) difficult to compress and transfer. The tomographic reconstruction is highly demanding on compute (GPU) and storage resources for the intermediate and/or final result. In addition, the optional image segmentation step may be demanding on computer memory.

The offline analysis (after experiment) could be performed remotely by users at home making it attractive for deployment as a cloud-like use case. Finally, this technique is applied at many facilities and in different scientific domains - therefore a portable result is more useful.



Volume rendering of a liquid foam flowing through a constriction<sup>4</sup>, see second publication below.

**Raw data:** SLS TOMCAT

**Reduced/processed/intermediate data:**

**Metadata:**

**Auxiliary data:**

**Analysis pipeline:** The processing pipeline starts with basic data corrections followed by tomographic reconstruction and image processing as segmentation, labeling and quantification. At PaN facilities the analysis is focused on the first initial robust steps (correction, reconstruction and segmentation) that allows assessment of the quality of experiment as well as significant reduction of data.

An advanced tomographic reconstruction algorithm (Nikitin, 2019<sup>5</sup>) that is implemented within a popular [Tomopy](#)<sup>6</sup> software is used. Image segmentation and quantification can be done with popular [scikit-image](#)<sup>7</sup> software.

For the reconstruction and segmentation steps, HPC-like data processing frameworks such as [SAVU](#)<sup>8</sup> and nowadays popular [jupyter notebooks](#)<sup>9</sup> can be used.

**Sample data sets:**

There is a large variety of specimens and imaging applications in medical sciences, material research or archeology (see deposited datasets in [Tomobank](#)<sup>10</sup>, [ESRF Paleo bank](#)<sup>11</sup> or recently deposited [PSI lung data set](#)<sup>12</sup> from dataset #2). Metal foams present one

<sup>4</sup> <https://journals.iucr.org/s/issues/2017/06/00/pp5108/pp5108fig7.html>

<sup>5</sup> <https://doi.org/10.1109/TCI.2019.2898088>

<sup>6</sup> <https://tomopy.readthedocs.io>

<sup>7</sup> <https://scikit-image.org/>

<sup>8</sup> <https://savu.readthedocs.io>

<sup>9</sup> <https://github.com/tomograms/tomography-notebooks>

<sup>10</sup> <https://tomobank.readthedocs.io>

<sup>11</sup> <http://paleo.esrf.fr>

<sup>12</sup> <https://doi.psi.ch/detail/10.16907/7eb141d3-11f1-47a6-9d0e-76f8832ed1b2>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

such widely investigated class of materials. Foams are complex cellular systems which require artifact free tomographic reconstruction for a reliable quantification of their time-dependent properties such as deformation fields of bubbles.

- Tomobank: Datasets → Dynamic → Foam data
  - o [Globus link \(APS/Petrel\)](#)<sup>13</sup>

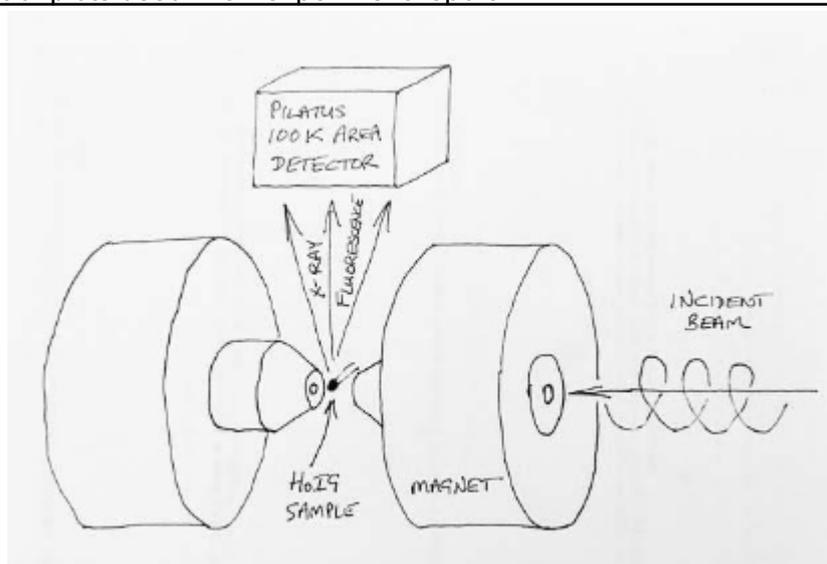
#### Associated publication(s):

- V. V. Nikitin, M. Carlsson, F. Andersson, and R. Mokso, IEEE Transactions on Computational Imaging (2019), [doi:10.1109/TCI.2019.2898088](https://doi.org/10.1109/TCI.2019.2898088)
- R. Mokso et al., Journal of Synchrotron Radiation (2017), [doi:10.1107/S1600577517013522](https://doi.org/10.1107/S1600577517013522)

## 4. Single-crystal X-ray diffraction

**Lead:** Diamond

**EOSC motivation:** Demonstrate a workflow for a novel/one-off experiment using low-level analysis tools. This example shows Diamond NeXus files from an experiment on magnetism and basic x-ray physics. The files are processed in a Jupyter notebook to produce a set of plots used in an experiment report.



Sketch of the environmental setup (not to scale) around the HoIG crystal sample (see associated publication below)

**Raw data:** NeXus files, see <https://doi.org/10.5281/zenodo.4445063>

**Reduced/processed/intermediate data:**

In Jupyter notebook: <https://doi.org/10.5281/zenodo.4444832>

**Metadata:** In NeXus files

**Auxiliary data:** None

**Analysis pipeline:** ideally a conda environment with some non-standard conda packages (for Bayesian inference and reflectometry calculations). Typical using Pandas and Matplotlib with Diamond specific additions.

**Sample data sets:** <https://zenodo.org/record/4445063#.YAVazNanxR4>

**Associated publication(s):** Collins, Stephen Patrick, Collins, Joel Thomas, & Lovesey, Stephen William. (2021, January 16). Quadrupolar X-ray Magnetic Circular Dichroism using Superchiral X-rays. Zenodo. <https://zenodo.org/record/4444832>

<sup>13</sup> [https://app.globus.org/file-manager?origin\\_id=e133a81a-6d04-11e5-ba46-22000b92c6ec&origin\\_path=%2Ftomobank%2Ftomo\\_00080%2F](https://app.globus.org/file-manager?origin_id=e133a81a-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Ftomobank%2Ftomo_00080%2F)



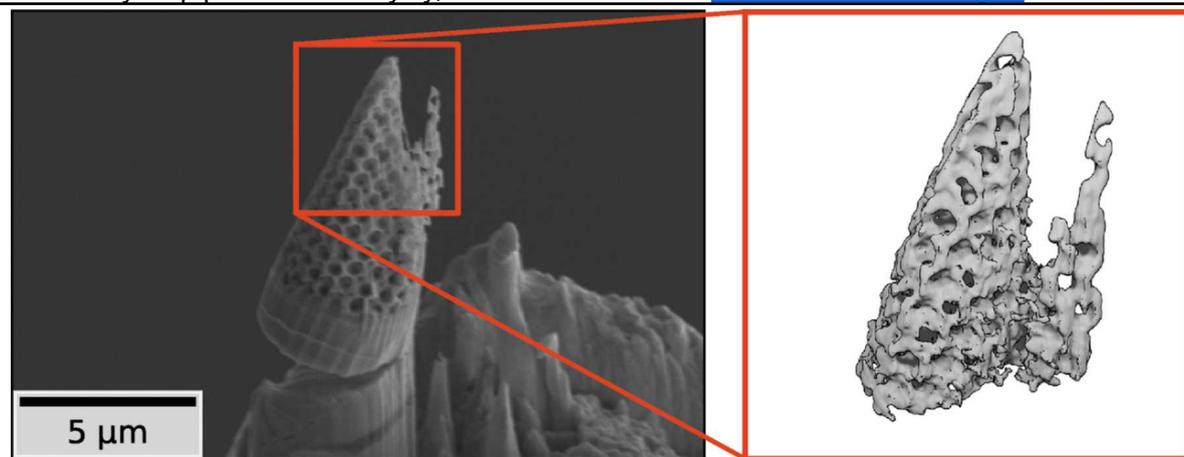
## 5. Ptychography, Ptychographic X-ray computed tomography

**Lead:** MAX IV

**EOSC motivation:** Ptychography and ptychographic X-ray computed tomography (PXCT) are well established experimental techniques that are taking advantage of the increased coherence of the 4th generation light sources.

PXCT allows reconstruction of the sample electron-density in space and imaging 3D samples with resolution down to nanometers. An experimental dataset presents 180 angular projections, each of them with 168 x 80 steps, i.e. more than 2.4 million measurements in total.

The analysis pipeline uses PtyPy, a software from the [PaN software catalogue](#).



**Raw data:**

/data/visitors/nanomax/20191087/2019120708/raw/tomo\_Ni\_552nm/ (1.5 TB, 1000 files)

**Processed data:**

/data/visitors/nanomax/20191087/2019120708/processed/tomo\_Ni\_552nm/ (765 GB, 24000 files)

**Published data:** <https://doi.org/10.5281/zenodo.3702582>

**Metadata:** (TODO) need to be integrated in MAX IV SciCat

**Auxiliary data:** reconstruction scripts available

**Analysis pipeline:** PXCT data processing pipeline is a series of three major steps: ptychographic phase reconstruction, projections alignment and tomographic volume reconstruction. For this dataset the first phase reconstruction step, using *PtyPy* software (Enders & Thibault, 2016) is applied which is a robust and well established method. The dataset can be reduced to 128 x 128 pixels field of detector view and in such a case around 5 hours are needed for the first stage of data processing and it is advised to run 50 jobs in parallel. PtyPy: [software.pan-data.eu/software/160/ptypy](http://software.pan-data.eu/software/160/ptypy)

**Sample data sets:**

Maik Kahnt, Simone Sala, Ulf Johansson, Alexander Björling, Zhimin Jiang, Sebastian Kalbfleisch, ... Karina Thånell. (2020). Raw data for "First ptychographic X-ray computed tomography experiment at the NanoMAX beamline" [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3702583>

This dataset is Pilatus detector data, Eiger2 datasets are available later; only the first stage, most robust and well established step, of the pipeline is implemented

**Associated publication(s):**

J. Appl. Cryst. (2020). 53, 1444-1451 <https://doi.org/10.1107/S160057672001211X>

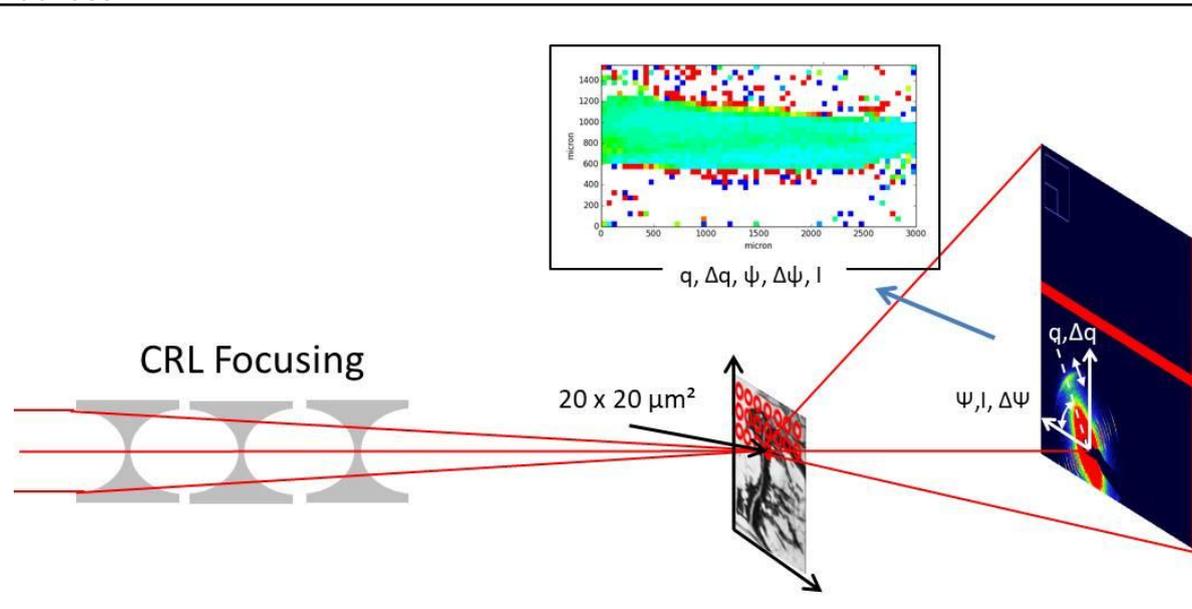


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

## 6. Kinetic SAXS, 2D Scanning SAXS

**Lead:** Soleil

**EOSC motivation:** Cartography or kinetic experiments can generate up to 3 TB per day of raw data saved at the measurement facility. Such large datasets are impractical for users to take home. Subsequent analysis needs to be performed remotely on a performant platform, making it attractive for deployment as a cloud-like use case. Even for a smaller amount of data, the possibility to perform offline analysis remotely is very attractive, in particular if mail-in and remote controlled modes are to be developed at Synchrotron facilities.



**Raw data:** Raw data as saved by the facility. File format: a series of NeXus files (HDF5), sharing the same root\_name. Each file contains the Eiger4M images and Sample and Beamline metadata of a sub-region of the total mapping region. Altogether, the list of files describes the whole mapped region.

**Reduced/processed/intermediate data:** <http://doi.org/10.5281/zenodo.4553506>

Using Foxtrot application, each of the raw file making the entire map is treated by an editable specific macro, which converts the stacks of raw images into stacks of physically meaningful parameters, according to the specific analysis that has to be performed. For example, there is a macro which fits the radially grouped scattering pattern with a Gaussian function and delivers the  $q$  position,  $\Delta q$  extension and maximum intensity of the peak. For each of the resulting parameters, the corresponding sub-maps are computed onto a normalized grid, taking into account the actual reported positions of the two scanning motors. All the sub-maps are then exported to an output NeXus file (<0.2GB), which contains also metadata related to the data reduction steps (essentially a reminder of the macro parameters). A python script then extracts the individual sub-maps as ASCII files, and a second python script concatenates the sub-maps into a final Image Ascii file, one per parameter. The latter can be readily imported by, e.g., ImageJ.

**Metadata:** All experimental metadata are in the raw data files. The analysis metadata are in the intermediate output NeXus files.

**Auxiliary data:** None.

**Analysis pipeline:** The Foxtrot Java based application is used to reduce all SAXS data.

[http://www.smallangles.net/wgwiki/images/f/f3/Abstract\\_JavierPerez.pdf](http://www.smallangles.net/wgwiki/images/f/f3/Abstract_JavierPerez.pdf)

<https://www.synchrotron-soleil.fr/en/beamlines/swing>

**Sample data sets:** One 2D Scanning SAXS: raw data files, intermediate output files, final Ascii files.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Constantin, Doru, Perez, Javier, & Ounsy, Majid. (2021). SAXS Mapping reduced data [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.4553506>

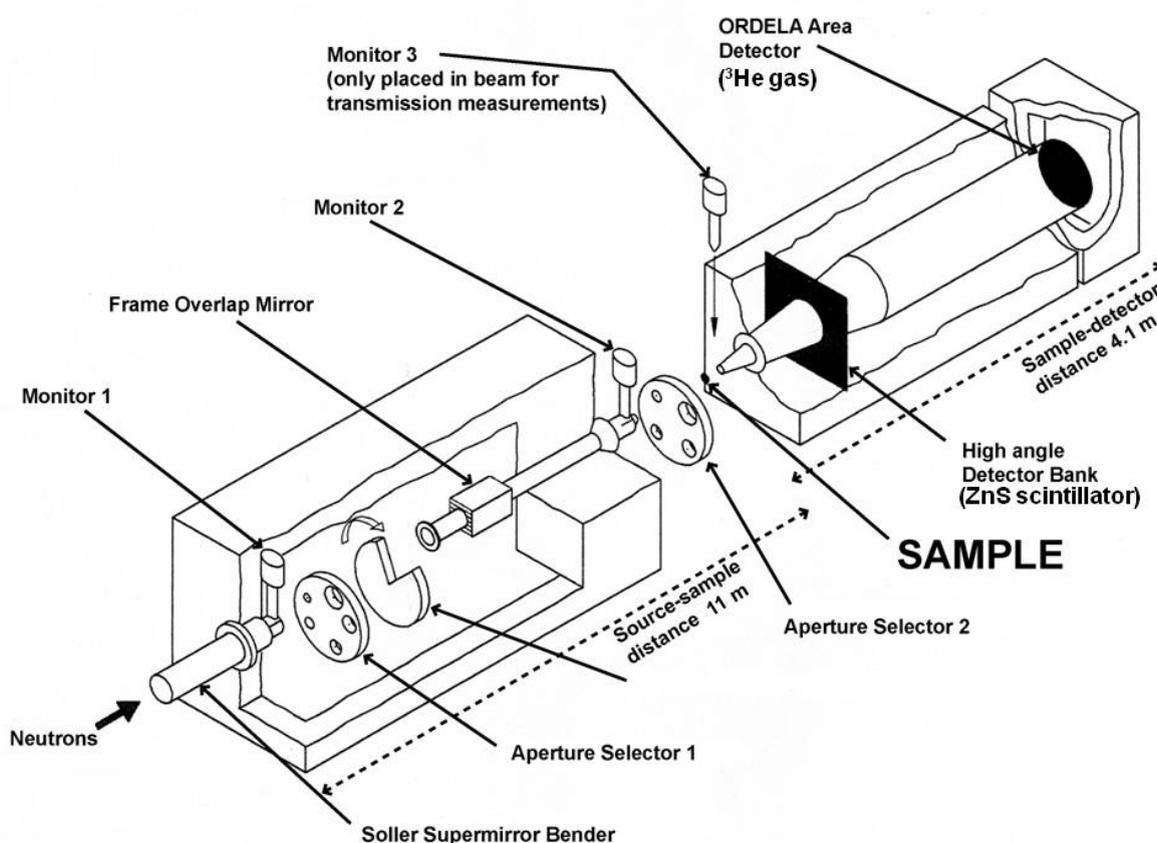
**Associated publication(s):** -

## 7. Neutron Small Angle Scattering

**Lead:** UKRI/ISIS ([isis.stfc.ac.uk](http://isis.stfc.ac.uk))

**EOSC motivation:** Use the data reduction software [Mantid](#)<sup>14</sup> (the most used data reduction software at ISIS and adopted neutron data reduction software elsewhere) on workflows to turn raw data into reduced data (which can then be used to analyse data). The specific example here is of 'protein standard' data required for calibration purposes and on the LOQ beamline.

[LOQ beamline](#)<sup>15</sup>



LOQ used to investigate the shape and size of large molecules, small particles or porous materials with dimensions in the range of 1 - 100nm. Length scales of up to 400nm can be probed in highly anisotropic systems.

**Raw data:** [https://github.com/DAAAS-reference-data/SANS/tree/main/nexus\\_files](https://github.com/DAAAS-reference-data/SANS/tree/main/nexus_files)

**Reduced/processed/intermediate data:**

[https://github.com/DAAAS-reference-data/SANS/tree/main/original\\_reduced\\_data](https://github.com/DAAAS-reference-data/SANS/tree/main/original_reduced_data)

**Metadata:** <https://github.com/DAAAS-reference-data/SANS/blob/main/README.md>

**Auxiliary data:** <https://github.com/DAAAS-reference-data/SANS/tree/main/mantid-files>

**Analysis pipeline:**

[https://github.com/DAAAS-reference-data/SANS/blob/main/ExPaNDS\\_reduction.ipynb](https://github.com/DAAAS-reference-data/SANS/blob/main/ExPaNDS_reduction.ipynb)

<sup>14</sup> <https://www.mantidproject.org/>

<sup>15</sup> <https://www.isis.stfc.ac.uk/Pages/loq.aspx>



**Sample data sets:** <https://github.com/DAAA-reference-data/SANS> with raw data files, meta data files, reduced data files, analysis pipeline via mantid files and jupyter notebook

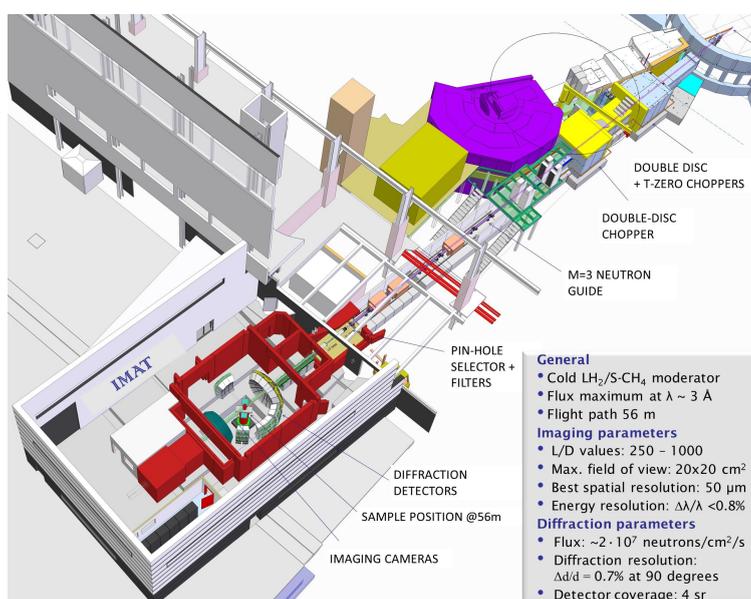
**Associated publication(s):** Biomedical Spectroscopy and Imaging **6** (2017) 123–134 123  
<https://doi.org/10.3233/BSI-170167>

## 8. Neutron Imaging/Tomography

**Lead:** UKRI/ISIS ([isis.stfc.ac.uk](http://isis.stfc.ac.uk))

**EOSC motivation:** example of neutron (rather than photon) tomography, 4D tomography data set (3D+energy), tomography workflow. Data collected on ISIS IMAT beamline. Data reduction and analysis included contributions from three neutron facilities: ISIS, [PSI](https://www.psi.ch/en/ins)<sup>16</sup> and [ESS](https://www.europeanspallationsource.se)<sup>17</sup>.

[IMAT beamline](https://www.isis.stfc.ac.uk/beamlines/imat)<sup>18</sup>



A neutron imaging and diffraction instrument for materials science, materials processing and engineering

**Raw data:** This dataset contains the wavelength-resolved neutron tomography of a contrast sample imaged with the time-of-flight (TOF) transmission imaging method at the IMAT beamline at the ISIS pulsed neutron source. The sample is made of several polycrystalline materials: nickel, iron, titanium, lead, copper and aluminium

**Reduced/processed/intermediate data:** The data are pre-processed for the detector event overlap correction, binned in the TOF axis and cropped. The data is available at <https://doi.org/10.5281/zenodo.3672798>

**Metadata:** -

**Auxiliary data:** -

**Analysis pipeline:** The datasets can be used for replicating the tomographic reconstruction as described in: J. Appl. Cryst. (2020). **53**, 188-196

<https://doi.org/10.1107/S1600576720000151>

**Sample data sets:**

Chiara, Triestino, Winfried, & Anders. (2020). Time-of-flight neutron tomography [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3672798>

<sup>16</sup> <https://www.psi.ch/en/ins>

<sup>17</sup> <https://europeanspallationsource.se>

<sup>18</sup> <https://www.isis.stfc.ac.uk/Pages/IMAT.aspx>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

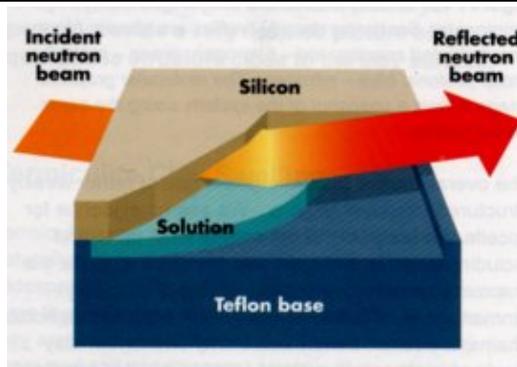
**Associated publication(s):**

J. Appl. Cryst. (2020). 53, 188-196 <https://doi.org/10.1107/S1600576720000151>

## 9. Neutron - Reflectometry

**Lead:** UKRI/ISIS ([isis.stfc.ac.uk](https://isis.stfc.ac.uk))

**EOSC motivation:** The sample is part of a published study of the determination of a thin-film structure using Neutron Reflectivity.



Schematic of Neutron Reflectivity measurement

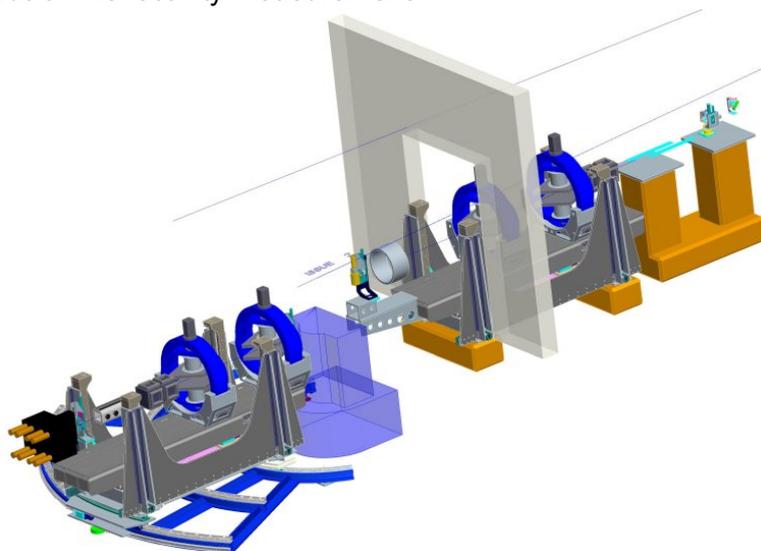


Diagram of [OFFSPEC reflectometer](#)

**Raw data:**

<https://data.isis.stfc.ac.uk/doi/INVESTIGATION/84794745/>

**Reduced/processed/intermediate data:**

[https://github.com/DAAA-reference-data/Reflectivity/blob/main/OFFSPEC\\_runs\\_46992-4/Runs\\_46992\\_46993\\_46994\\_polymer\\_on\\_Si.dat](https://github.com/DAAA-reference-data/Reflectivity/blob/main/OFFSPEC_runs_46992-4/Runs_46992_46993_46994_polymer_on_Si.dat)

**Metadata:** -**Auxiliary data:** -**Analysis pipeline:** Mantid workflows.

[https://github.com/DAAA-reference-data/Reflectivity/blob/main/OFFSPEC\\_runs\\_46992-4/RB1720415\\_run\\_46992.py](https://github.com/DAAA-reference-data/Reflectivity/blob/main/OFFSPEC_runs_46992-4/RB1720415_run_46992.py)

**Sample data sets:**

<https://data.isis.stfc.ac.uk/doi/INVESTIGATION/84794745/> with raw data files, reduced data files, analysis pipeline via Python reduction script. Data taken on OFFSPEC reflectometer.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

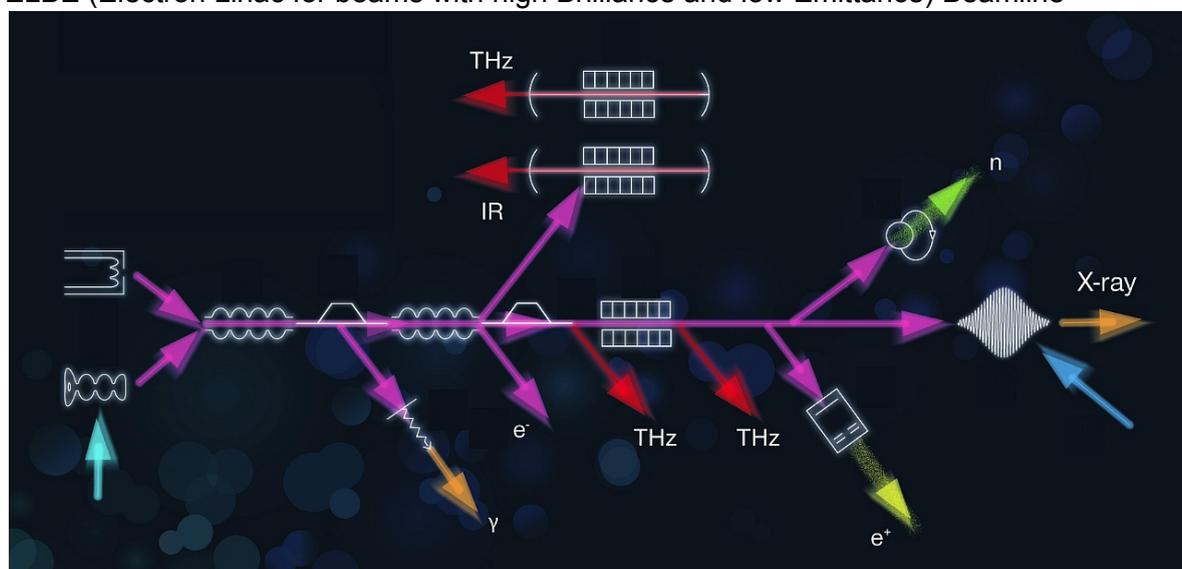
**Associated publication(s):** ACS Appl. Mater. Interfaces 2019, 11, 14, 13803–13811  
<https://doi.org/10.1021/acsami.9b02700>

## 10. Terahertz Spectroscopy

**Lead:** HZDR (<https://www.hzdr.de>)

**EOSC motivation:** Nonlinear THz Spectroscopy has developed into a prosperous field of basic research in physics, chemistry and material sciences. In high energy physics, the Higgs field couples to gauge bosons and fermions and gives mass to their elementary excitations. Experimentally, such couplings can be verified from the decay product of the Higgs boson, the scalar (amplitude) excitation of the Higgs field. We investigate the Higgs mode of several cuprate thin films using phase-resolved terahertz third harmonic generation (THG).

ELBE (Electron Linac for beams with high Brilliance and low Emittance) Beamline



**Raw data:** Phase-resolved Higgs response in superconducting cuprates

DOI: 10.14278/rodare.277, Files (2 TB)

**Reduced/processed/intermediate data:**

DOI: 10.14278/rodare.277, Data analysis.zip (1,7 MB)

**Metadata:**

DOI: 10.14278/rodare.277, Labbook-June2018 Higgs.zip, Labbook-Dec2018 Higgs.zip

**Auxiliary data:**

DOI: 10.14278/rodare.277, Readme\_Rodare.txt

**Analysis pipeline:** TELBE

<https://www.hzdr.de/db/Cms?pOid=34100&pNid=2609&pLang=en>

**Sample data sets:** <https://rodare.hzdr.de/record/277>

**Associated publication(s):**

Chu, H., Kim, M.J., Katsumi, K. *et al.* Phase-resolved Higgs response in superconducting cuprates. *Nat Commun* 11, 1793 (2020). <https://doi.org/10.1038/s41467-020-15613-1>

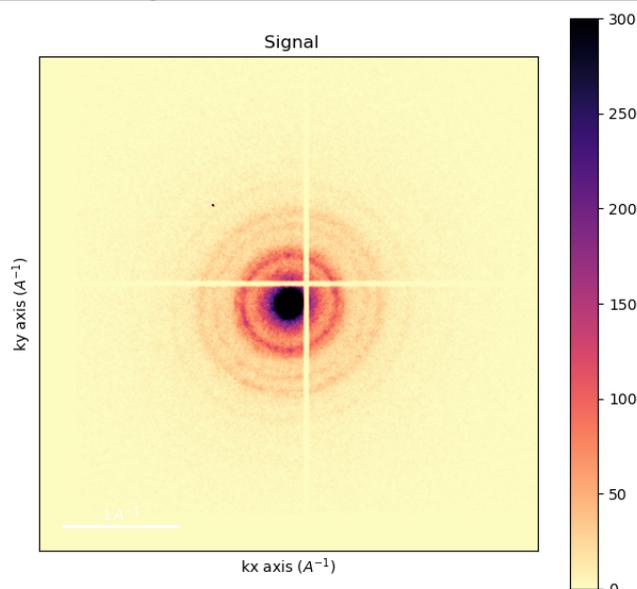


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

## 11. Electron scanning diffraction imaging (4DSTEM)

**Lead:** Diamond

**EOSC motivation:** The recent development of high speed direct electron detectors has enabled the recording of full electron diffraction patterns at each position in a scan STEM scan within a transmission electron microscope. A wide variety of analysis techniques (including ptychographic reconstructions) can then be performed on this data to determine a wealth of information about the sample being studied. The size of these data-sets (raw data > 100Gb) presents challenges for users.



One step of the diffraction roundness correction from the example SED analysis using PyXEM on Diamond Jupyterhub (see analysis pipeline below)

**Raw data:** ePSIC at Diamond

**Reduced/processed/intermediate data:**

<https://dls1td.sharepoint.com/:u:/s/GRA0046/ER6elCBerHxDICfYCIBeUFsBakNbDsREMRobNW9VvrUXaQ?e=b7MLiY>

**Metadata:** -

**Auxiliary data:** -

**Analysis pipeline:** pyxem in non-standard conda environment. Example notebook:

<https://dls1td.sharepoint.com/:u:/s/GRA0046/ER6elCBerHxDICfYCIBeUFsBakNbDsREMRobNW9VvrUXaQ?e=b7MLiY>

**Sample data sets:**

<https://dls1td.sharepoint.com/:u:/s/GRA0046/ER6elCBerHxDICfYCIBeUFsBakNbDsREMRobNW9VvrUXaQ?e=b7MLiY>

**Associated publication(s):** <https://pubs.acs.org/doi/abs/10.1021/jacs.0c04468>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.