The common goal of the PaNOSC and ExPaNDS projects is to make scientific FAIR data (Findable, Accessible, Interoperable and Reproducible) a reality and support the advancement of open science through the development of high-value-added services leveraging that data that will also be available from the European Open Science Cloud.

Data Analysis Services for Photon and Neutron Science

PaNOSC has been developing the **Common Portal for Data Analysis Services** to facilitate starting a data analysis session after a dataset of interest has been collected. The Portal aims to provide access to both remote desktop environments and Jupyter Notebooks, enabling users to **remotely analyse data** from PaN facilities, such as synchrotrons and neutron sources.

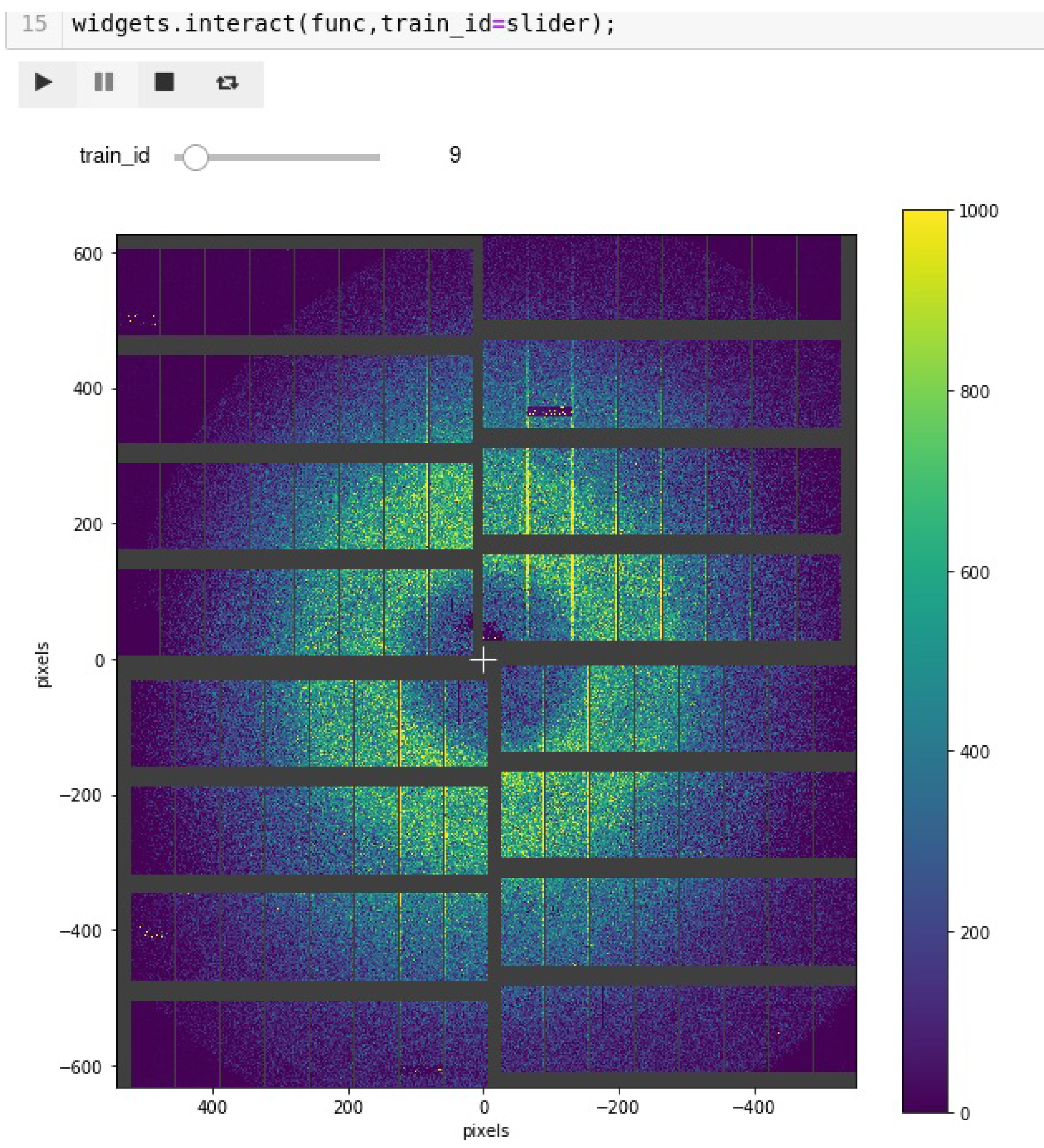
After initial deployment at facilities to provide remote analysis services to local data, the Portal will be deployed as part of the EOSC to provide federated data analysis of data across the facilities.

This poster aims to present a selection of use cases of the data analysis services developed in the projects.

Typical workflow for Common Portal usage

- User authentication using Umbrella ID
- Search for datasets (for which the user has been a proposal member, thus requiring authenticated access)
- Selection of an analysis environment (in either a remote desktop, or Jupyter Notebook environment)
- Spawn the chosen analysis environment and link it to the chosen datasets (if the analysis environment is not physically located with the data, data transfer protocols will be active)
- Access the environment via the Portal and perform data analysis

EuXFEL Data in NeXus Golden Standard



There were great efforts put in the last two decades in the research community to elaborate a common standard for high data-rate macromolecular crystallography (HDRMX). This agreed "Gold Standard" builds on the NeXus/HDF5 NXmx application definition and the International Union of Crystallo-graphy (IUCr) imgCIF/CBF dictionary, and it is compatible with major data-processing programmes and pipelines. Here we demonstrate the EuXFEL data packed into a NeXus file, which is fully compliant with the Gold Standard by design, since it is built directly from HDRMX NeXus definitions. We use open-source software developed both by community (cctbx) and in-house (extra-data).

Developers:
Herbert Bernstein, Andreas Förster, Asmit Bhowmick, Aaron Brewster, Sandor Brockhauser, Luca Gelisio, David Hall, Filip Leonarski, Valerio Mariani, Gianluca Santoni, Clemens Vonrheink, Graeme Winter, Anton Barty, Thomas Kluyver, Fabio Dall'Antonias, Yury Kirienko

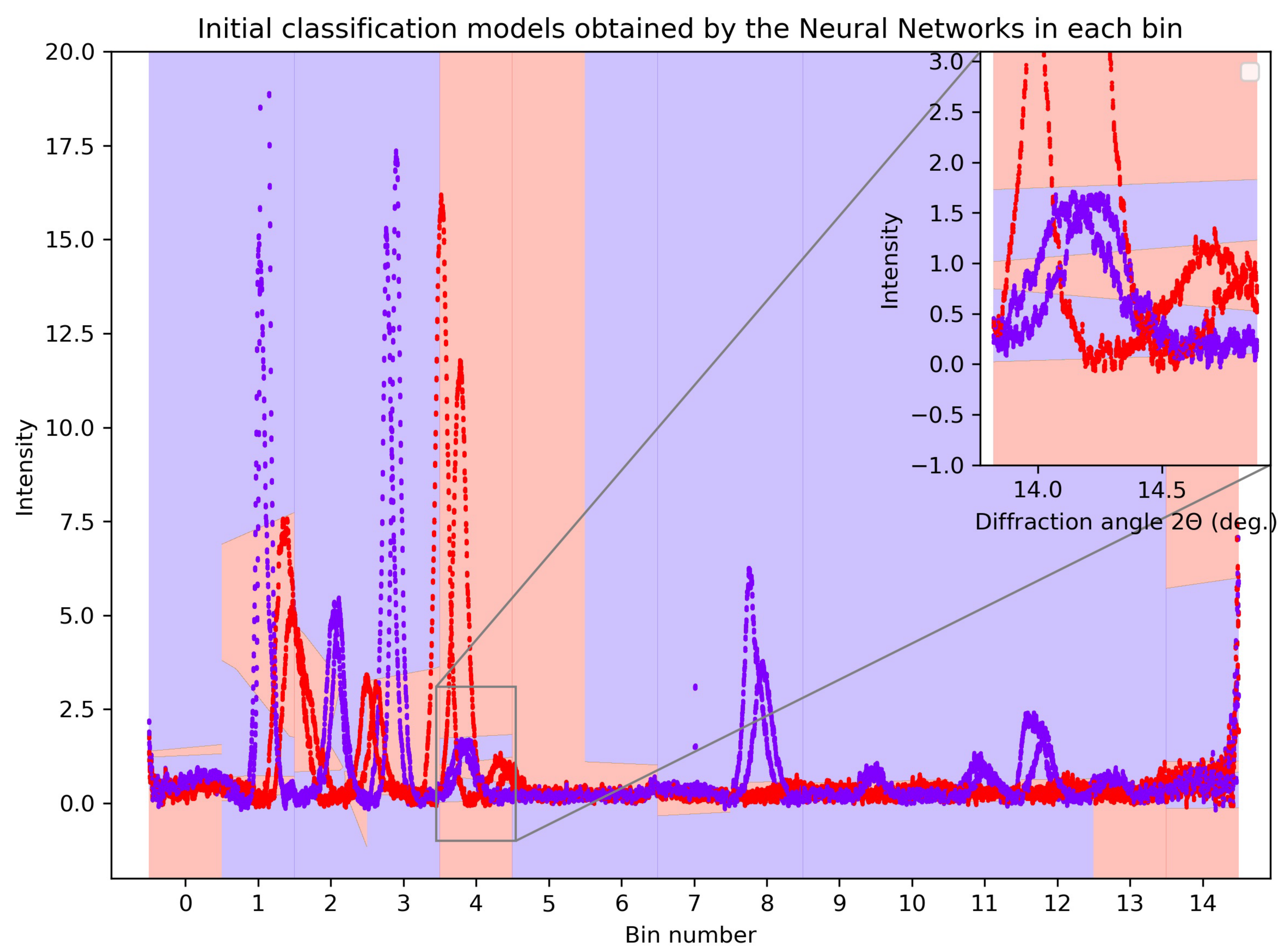
References:

[1] Bernstein, H. J., Förster, A., Bhowmick, A., Brewster, A. S., Brockhauser, S., Gelisio, L., ... & Winter, G. (2020). Gold Standard for macromolecular crystallography diffraction data. IUCrJ, 7(5), DOI: https://doi.org/10.1107/S2052252520008672
[2] CXI DB #80, DOI: https://dx.doi.org/10.11577/1464101
[3] A. Brewster, M. Wang, H. Bernstein. (2019). 68 image lysozyme dataset recorded on the Jungfrau 16M detector at SwissFEL and formatted as a NeXus file (Version 3) [Data set]. DOI: https://doi.org/10.5281/zenodo.3834335

More: https://github.com/European-XFEL-examples/panosc-nexus

Machine Learning Based Spectra Classification

At the European XFEL, X-ray pulses can be generated with only 220ns separation in time and a maximum of 27000 pulses per second. In experiments at European XFEL, spectral changes can indicate the change of the system under investigation and so the progress of the experiment. An immediate feedback on the actual status (e.g., time resolved status of the sample) would be essential to quickly judge how to proceed with the experiment. The two major spectral changes that we aim to capture are either the change of intensity distribution (e.g., drop or appearance) of peaks at certain locations, or the shift of those on the spectrum.
Machine Learning (ML) opens up new avenues for data-driven analysis in spectroscopy by offering the possibility to quickly recognize such specific changes on-the-fly during data collection. ML requires lots of data which are clearly annotated. Hence, it is important that research outputs align with the FAIR principles. In case of XFEL experiments, it is suggested to introduce NeXus data format standards in future experiments.
In this work an example is presented, of a possible use of Neural Network based ML for accurately classifying the system state if data is properly provided. A solution is shown, to automatically find the regions (or bins) with high separability where the spectra classes differ significantly. By teaching individual neural networks for each bin, and by combining them with a weighting technique, a robust classification of any new spectral curve can be quickly obtained.



Developers:
Christian Plueckthun; Zuzana Konopkova; Sandor Brockhauser; Yue Sun

References:

[1]. Pennicard, D., Smoljanin, S., Pithan, F., Sarajlic, M., Rothkirch, A., Yu, Y., Liermann, H.P., Morgenroth, W., Winkler, B., Jenei, Z. and Stawitz, H., 2018. LAMBDA 2M GaAs—A multi-megapixel hard X-ray detector for synchrotrons. Journal of Instrumentation, 13(01), p.C01026, DOI: 10.1088/1748-0221/13/01/C01026

More: https://github.com/European-XFEL-examples/panosc-ml-spectra-classification

PaNOSC and ExPaNDS have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements no. 823852 (PaNOSC) and no. 857641 (ExPaNDS).

https://expands.eu/ | https://panosc.eu
@ExPaNDs_EU / @Panosc_eu