Use Case 10: Diffraction & Spectroscopy

Authors: J. Dallmann[1], B. Hakim[1], K.Götz[1], T.Unruh[1], C. Hövelmann[2], B. M. Murphy[2], M. Dargasz[3], C. Gutt[3], T. Schneider[4], , A. Hinderhofer[5], D. Lapkin[5], F. Schreiber[5]

[1]Friedrich-Alexander Universität Erlangen-Nürnberg [2]Christian-Albrechts-Universität (CAU) zu Kiel [3]Universität Siegen [4]European Molecular Biology Laboratory (EMBL), Hamburg [5]Eberhard-Karls Universität Tübingen.

Within Use Case 10 the ICSP@FAU group works towards an automated, generic and FAIR data workflow for research institutes handling data from various sources. Key components of the workflow are a local NOMAD Oasis[1], which is hosted by a local computer center and the utilization of the NeXus[2] standard wherever possible. We present how we utilize and customize our NOMAD Oasis to fit the FAIR requirements and support our scientific work from data collection to publication.

Key metadata requirements for biological samples studied within the Bio-SAXS part of Use Case 10 are identified by learning from the automatic data and metadata pipeline at EMBL's SAXS beamtime (P12), PETRA III and the integrated SASBDB[3] database, a federated database designed for biological SAXS and SANS data, offering curated, searchable experimental data along with relevant metadata. Yet, at most beamlines and especially lab sources no automatic metadata pipeline exists and metadata is often collected manually and separately from the raw data. Here, we present our efforts at CAU, EMBL and Uni Siegen to increase findability and accessibility by establishing data handling protocols including the creation of NeXus files and generation of persistent sample identifiers (IGSN) on an exemplary data publication of Bio-SAXS data[4] collected at BL2, DELTA.

Progress on automated feature detection in 2D image data will be presented by machine learning, focussing here on grazing incidence wide angle X-ray Scattering (GIWAXS) [5-7]. The peak detection was already demonstrated to be able to track and process scattering features during in-situ synchrotron experiments[6]. Additionally, a labelled data set for benchmarking peak detection algorithms was published[5].

References

1) https://nomad-lab.eu/nomad-lab/nomad-oasis.html
2) https://www.nexusformat.org/
3) E. Valentini et al. Nucleic Acids Research, 2015, Vol. 43, Database issue D357–D363.
4) S. C. Hövelmann et al. IUCrJ, 2024, 11, 486-493.
5) C. Völter et al. J. Appl. Cryst. (2025), in print
6) V. Starostin et al. Synchrotron Radiation News **35** (2022) 21
7) V. Starostin et al. npj Comput Mater **8** (2022) 101